

UNIVERSIDADE DO PLANALTO CATARINENSE
PROGRAMA DE PÓS-GRADUAÇÃO EM AMBIENTE E SAÚDE

NEY KASSIANO RAMOS

GESTÃO DO CONHECIMENTO, DATA SCIENCE E DADOS
INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO

LAGES
2020

NEY KASSIANO RAMOS

**GESTÃO DO CONHECIMENTO, DATA SCIENCE E DADOS
INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO**

Dissertação apresentada para obtenção do título de Mestre no Programa de Pós-Graduação em Ambiente e Saúde da Universidade do Planalto Catarinense - UNIPLAC.

Orientadora: Profa. Dra. Cristina Keiko Yamaguchi

Coorientador: Prof. Dr. Ubirajara Maciel da Costa

LAGES

2020

Ficha Catalográfica

R175g Ramos, Ney Kassiano.
Gestão do conhecimento, Data Science e dados interlaboratoriais na geração de ativos de conhecimento/Ney Kassiano Ramos – Lages, SC, 2020.
178 p.

Dissertação (Mestrado) - Universidade do Planalto Catarinense. Programa de Pós-Graduação em Ambiente e Saúde da Universidade do Planalto Catarinense.
Orientadora: Cristina Keiko Yamaguchi.
Coorientador: Ubirajara Maciel da Costa

1. Ambiente. 2. Gestão do Conhecimento. 3. Saúde. 4. Tecnologia de Informação. I. Yamaguchi, Cristina Keiko. II. Costa, Ubirajara Maciel de. III. Título.

CDD 658.4038

Catálogo na Fonte: Biblioteca Central

PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU*
MESTRADO ACADÊMICO EM AMBIENTE E SAÚDE



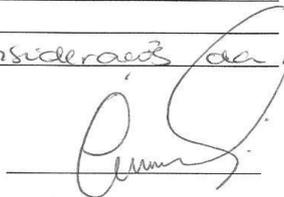
ATA DE DEFESA PÚBLICA DE DISSERTAÇÃO Nº 056

Aos seis dias do mês de março de 2020, às quatorze horas, no Auditório do MIDILages da UNIPLAC, reuniram-se as/os professoras/es: **Profa. Dra Cristina Keiko Yamaguchi** (Orientadora e Presidente da Banca Examinadora – PPGAS/UNIPLAC), **Prof. Dr. Ubirajara Maciel da Costa** (Coorientador), **Prof. Dr. Alex Paubel Junger** (Examinador Titular Externo - UFABC), **Profa. Dra. Rosana Claudio Silva Ogoshi - UNIARP** (Examinadora Suplente Externa – UCS), **Profa. Dra. Lenita Agostinnetto** (Examinadora Titular Interna - PPGAS/UNIPLAC), **Profa. Dra Lilia Aparecida Kanan** (Examinadora Suplente Interna - PPGAS/UNIPLAC), para arguir e avaliar a Dissertação intitulada **GESTÃO DO CONHECIMENTO, DATA SCIENCE E DADOS INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO** de autoria de **Ney Kassiano Ramos**. Realizada a sessão, conforme os procedimentos regimentais, a Banca Examinadora considerou a dissertação Aprovado. A versão final, incorporando correções e/ou complementos sugeridos pela Banca Examinadora, considerados pertinentes, é requisito para obtenção do Diploma de Mestre, devendo ser protocolada na Secretaria do PPGAS no prazo máximo de sessenta (60) dias a contar desta data. Nada mais havendo a tratar foi lavrada a presente ata, lida, aprovada e assinada. Recomendações da Banca:

A banca salientou a relevância do tema pesquisa.
Solicitou-se adequar as considerações da banca.

Banca Examinadora:

Profa. Dra Cristina Keiko Yamaguchi
(Orientadora e Presidente da Banca Examinadora – PPGAS/UNIPLAC)

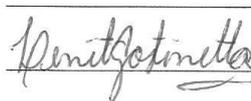


Prof. Dr. Ubirajara Maciel da Costa
(Coorientador)

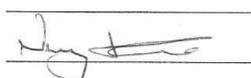
Prof. Dr. Alex Paubel Junger
(Examinador Titular Externo - UFABC)

Participação não presencial – Resolução 432/2020

Profa. Dra. Rosana Claudio Silva Ogoshi - UNIARP
(Examinadora Suplente Externa – UCS)



Profa. Dra. Lenita Agostinnetto
(Examinadora Titular Interna - PPGAS/UNIPLAC)



Profa. Dra Lilia Aparecida Kanan
(Examinadora Suplente Interna - PPGAS/UNIPLAC)

Mestrando: **Ney Kassiano Ramos**

DEDICATÓRIA

Aos meus pais, obrigado pela dedicação e pela vida

À minha esposa e filhas

À família, amigas e amigos

AGRADECIMENTOS

À minha família, pelo apoio constante, encorajamento e carinho em todos os momentos.

À minha orientadora, Prof^a Dr^a Cristina Keiko Yamaguchi, por acreditar neste projeto e na minha visão, e que com sua paciência e conhecimento o tornaram possível, tornando mais fácil o que, às vezes, parecia complicado.

Ao Prof. Dr. Ubirajara Maciel da Costa, coorientador deste trabalho, pelas ideias iniciais e pela ajuda constante.

Aos membros da banca avaliadora, que com seus direcionamentos e referências ajudaram no aprofundamento do trabalho.

Ao colegiado do PPGAS pela delicadeza e carinho no tratamento e pela oportunidade de aprendizado.

Aos colegas de mestrado pelo conhecimento e tempo compartilhado.

À UDESC e ao Centro de Ciências Agroveterinárias, CAV, na pessoa de seu Diretor Geral e Diretores Adjuntos, aos colegas de trabalho e do Setor de TI, além dos diretores dos laboratórios participantes da pesquisa, que a todos os momentos foram acessíveis ao fornecer as informações que precisávamos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela contribuição na pesquisa científica e estímulo na consolidação da pós-graduação no País.

A todos, que de alguma maneira estiveram presentes durante a construção deste trabalho, e, de maneira sutil ou ostensiva ajudaram e apoiaram, com ideias, encorajamento ou somente por estarem ali.

RESUMO

RAMOS, Ney Kassiano. **Gestão do conhecimento, Data Science e dados interlaboratoriais na geração de ativos de conhecimento.** 2020. 177 f. Dissertação (Mestrado em Ambiente e Saúde). Programa de Pós-Graduação em Ambiente e Saúde. UNIPLAC, Lages, 2020.

O uso correto das informações é um fator chave na capacidade de tomada de decisão em uma instituição. Para isso é fundamental utilizar as técnicas adequadas que habilitem o seu uso ágil e otimizado. A Gestão do Conhecimento aliada à Tecnologia da Informação emerge como uma ferramenta para administrar o que uma instituição conhece, dando origem a Memória Organizacional, que trata dos registros dos conhecimentos criados nessa instituição. A pesquisa objetivou compreender como a gestão do conhecimento, Data Science e dados interlaboratoriais que contribuem para gerar ativos de conhecimento de saúde animal. As metodologias adotadas para essa pesquisa foram: exploratória, descritiva, aplicada, documental, qualitativa com abordagem interdisciplinar. A pesquisa foi aplicada nos Laboratórios Centro de Diagnóstico Microbiológico Animal (CEDIMA), Laboratório Parasitologia e Doenças, Laboratório Patologia Animal (LAPA), Patologia Clínica Veterinária, localizado no Hospital de Clínica Veterinária (HCV), do Campus III da Universidade Estadual de Santa Catarina, especificamente no Centro de Ciências Agroveterinárias (CAV) em Lages – SC. Os laboratórios de diagnósticos veterinários da universidade de ciências agroveterinárias geram e mantém um grande volume de informações e podem se utilizar da memória organizacional para extrair o máximo de conhecimento sobre seus resultados. A correlação das informações interlaboratoriais originadas dos laboratórios, geram ativos de conhecimento, contribuindo para melhorar o entendimento sobre o meio ambiente, a saúde e seus atores: animais e humanos. Nesse contexto, a pesquisa busca mostrar como a tecnologia de *Data Science* pode usar a memória organizacional e os dados interlaboratoriais para gerar ativos de conhecimentos. Foi desenvolvida de uma prova de conceito, simulando os bancos de dados de cada laboratório e, por meio da linguagem de programação Python e do *software* Tableau, foram gerados exemplos de extração de informações úteis destas bases de dados. Os resultados demonstraram que a aplicabilidade dessa técnica em dados de saúde animal, permitem melhorar a gestão do conhecimento em laboratórios de diagnósticos veterinários da universidade, de forma a retornar o maior valor agregado após sua correlação, os chamados ativos de conhecimento, ou como são conhecidos em *Data Science*, os *insights*. As ferramentas de análise usadas, Tableau e a linguagem Python, se mostraram capazes, cada uma com sua forma de aplicação e nicho de mercado, de manipular as informações dos bancos de dados criados na procura de padrões e correlações, na fase do Data Science conhecida como exploração. A curva de aprendizado do *software* Tableau se mostrou mais suave, em comparação com o Python e suas bibliotecas associadas. Por outro lado, o poder de uma linguagem de programação dificilmente pode ser comparado ao de um programa que busca ser fácil para o usuário final: sua complexidade ainda é, infelizmente, inversamente proporcional à acessibilidade de uso. Em síntese, entende-se que a Gestão do Conhecimento, se implantada nos laboratórios como um sistema informatizado e uma metodologia de gestão (por meio da Memória Organizacional, por exemplo), se mostra útil na geração de ativos de conhecimento, na forma de *insights*, pelo meio da técnica de *Data Science*.

Palavras-chave: Gestão do Conhecimento, Data Science, Ambiente, Saúde Animal.

ABSTRACT

RAMOS, Ney Kassiano. **Knowledge Management, Data Science and Interlaboratorial Data in Knowledge Assets Generation**. 2020. 177 f. Dissertation (Master's Degree in Environment and Health). Environment and Health Postgraduate Studies Program. UNIPLAC, Lages, 2020.

The correct use of information is a key factor in an institution's decision-making ability. It is essential to use the appropriate techniques that enable its agile and optimized use to achieve this. Knowledge Management combined with Information Technology emerges as a tool to manage what an institution knows, originating what is known as Organizational Memory, which deals with the records of knowledge created in that institution. This research aimed to understand how knowledge management, data science and interlaboratory data contribute to generating animal health knowledge assets. The methodologies adopted for this research were: exploratory, descriptive, applied, documentary, qualitative with an interdisciplinary approach. This research was applied at the Laboratory of Animal Microbiological Diagnosis (CEDIMA), Laboratory of Parasitology and Diseases, Laboratory of Animal Pathology (LAPA), Veterinary Clinical Pathology, located at the Veterinary Clinic Hospital (HCV), Campus III at the Santa Catarina State University, at the Center for Agroveterinary Sciences (CAV) in Lages - SC. The veterinary diagnostic laboratories at the university generate and maintain a large volume of information and can use organizational memory to extract the maximum knowledge about their results. The correlation of the interlaboratory information originating from the laboratories, generates knowledge assets, contributing to improve the understanding about the environment, health and its actors: animals and humans. In this context, this research aims to show how Data Science technology can use organizational memory and interlaboratory data to generate knowledge assets. It was built on a proof of concept, simulating the databases of each laboratory and using the Python programming language and the Tableau software. Examples of useful information from these databases were generated. The results demonstrated that the applicability of this technique in animal health data allows to improve the knowledge management in veterinary diagnostic laboratories of the university, in order to return the highest added value after its correlation, the so-called knowledge assets, or as they are known in Data Science, the insights. The analysis tools used, Tableau and the Python language, proved capable, each with its own methods, to manipulate the information from the databases created in search of patterns and correlations, in the Data Science phase known as exploration. The learning curve of the Tableau software turned out to be smoother compared to Python and its associated libraries. On the other hand, the power of a programming language can hardly be compared to that of a program that seeks to be easy for the end user: its complexity is still, unfortunately, inversely proportional to the accessibility of use. In summary, it is understood that Knowledge Management, if implemented in laboratories as a computerized system and a management methodology (through Organizational Memory, for example), is useful in the generation of knowledge assets, in the form of insights, using the Data Science technique.

Keywords: Knowledge Management, Data Science, Environment, Animal Health.

LISTA DE FIGURAS

Figura 1 - Ilustração da formação dos dados interlaboratoriais	6
Figura 2 - Fluxograma da busca na base de dados de artigos sobre a temática proposta.....	13
Figura 3 - Os quatro modos de conversão do conhecimento e a espiral do conhecimento	24
Figura 4 - Relação interdisciplinar entre as áreas de conhecimento	37
Figura 5 - Os 3 “Vs” do <i>Big Data</i>	48
Figura 6 - Algumas áreas de pesquisa científica que utilizam <i>Data Science</i>	54
Figura 7 - Áreas de conhecimento que formam o Data Science	55
Figura 8 - Fluxo: memória organizacional, Data Science e <i>insights</i>	57
Figura 9 - Tabelas do banco de dados Cedima.....	59
Figura 10 - Amostra de registros da tabela "tipo_amostra"	62
Figura 11 - Amostra de registros da tabela "espécie"	63
Figura 12 - Amostra de registros da tabela "raça"	64
Figura 13 - Tabelas do banco de dados Parasitologia	66
Figura 14 - Tabelas do banco de dados Patologia	67
Figura 15 - Estrutura das máquinas virtuais	68
Figura 16 - Linux virtualizado - Plataforma Jupyter	69
Figura 17 - Linux virtualizado – MySQL.....	70
Figura 18 - Interface do programa Tableau: configuração de conexão.....	71
Figura 19 - Interface do programa Tableau: geração de gráfico	72
Figura 20 - Painel de controle - Jupyter	73
Figura 21 - Editor de texto – Jupyter.....	74
Figura 22 - Cedima: quantidade de exames, por tipo - Tableau.....	79
Figura 23 - Cedima: quantidade de exames, por tipo - Python	79
Figura 24 - Cedima: resultados para antibiograma - Tableau	82
Figura 25 - Cedima: relação de resultados para FeLV - Tableau.....	82
Figura 26 - Cedima: relação de espécies mais atendidas – Tableau.....	83
Figura 27 - Cedima: porcentagem de espécies mais atendidas – Tableau	83
Figura 28 - Cedima: totais de exames para gatos - Tableau.....	84
Figura 29 - Cedima: relação de exames mais comuns em gatos siameses – Tableau	84
Figura 30 - Cedima: média de peso por espécie - Tableau.....	87
Figura 31 - Cedima: distribuição de peso dos gatos - Python	88
Figura 32 - Cedima: distribuição de peso dos gatos, por sexo - Python	88
Figura 33 - Componentes básicos de um gráfico <i>boxplot</i>	89
Figura 34 - Cedima: distribuição de peso dos gatos, por idade - Python	89
Figura 35 - Cedima: distribuição de idade dos cães, por raça - Python	90
Figura 36 - Cedima: distribuição geográfica de clientes - Tableau.....	91
Figura 37 - Cedima: localização geográfica de infecções por um agente específico - Python.....	92
Figura 38 - Cedima: localização geográfica de infecções por um agente específico, com destaque - Python	92
Figura 39 - Cedima: distribuição de materiais recebidos para análise - Python	93
Figura 40 - Cedima: distribuição de pacientes por sexo – Tableau.....	94
Figura 41 - Cedima: nomes de pacientes mais comuns - Tableau	94
Figura 42 - Cedima: Arrecadação anual por tipo de exame - Tableau.....	95

Figura 43 - Parasitologia: total de exames por agente infeccioso - Tableau	97
Figura 44 - Parasitologia: total de exames por agente infeccioso - Tableau	99
Figura 45 - Parasitologia: distribuição de tipos de amostras, gráfico de pizza – Tableau	100
Figura 46 - Parasitologia: distribuição resultados de exames – Tableau.....	101
Figura 47 - Parasitologia: distribuição resultados de exames para <i>Leishmania dovani</i> – Tableau	101
Figura 48 - Parasitologia: positivos para <i>Entamoeba dispar</i> , por mês, por ano - Python	101
Figura 49 - Parasitologia: total de exames por mês, de 2015 a 2017	102
Figura 50 – Cedima x Parasitologia: <i>heatmap</i> de pesos versus idades de equinos - Python	107
Figura 51 – Cedima x Parasitologia: exames por mês, ano 2017 - Tableau	108
Figura 52 – Cedima, Parasitologia x Patologia: comparação da quantidade de exames por mês, em 2017 – Python	111

LISTA DE TABELAS

Tabela 1 - Frequência das publicações por ano	17
Tabela 2 - Cedima: Total de exames por ano - Tableau	77
Tabela 3 - Total de exames por tipo - Python.....	78
Tabela 4 - Cedima: Positivos por ano - Tableau	80
Tabela 5 - Cedima: Bactérias identificadas - Tableau	81
Tabela 6 - Cedima: Vírus identificados – Tableau	81
Tabela 7 – Cedima: totais de raças de gatos - Tableau.....	84
Tabela 8 – Cedima: Resultados de exames por raça de gatos – Tableau	85
Tabela 9 - Cedima: totais de positivos para gatos SRD por agente patogênico - Tableau	85
Tabela 10 - Cedima: totais de positivos para gatos persa por agente patogênico – Tableau.....	86
Tabela 11 - Cedima: totais de positivos para gatos siameses por agente patogênico, por ano e data de nascimento - Tableau	86
Tabela 12 - Cedima: totais de positivos para gatos persa por agente patogênico, por ano e data de nascimento - Tableau	86
Tabela 13 - Cedima: média de peso dos gatos, por ano - Tableau	87
Tabela 14 - Cedima: média de peso dos gatos, por raça e por ano - Tableau.....	87
Tabela 15 - Parasitologia: total de exames por agente infeccioso - Tableau.....	97
Tabela 16 - Parasitologia: total de exames, por ano, por agente infeccioso - Tableau.....	98
Tabela 17 - Parasitologia: total de exames positivos por agente infeccioso - Tableau	98
Tabela 18 - Parasitologia: total de exames positivos, por ano, por agente infeccioso - Tableau	99
Tabela 19 - Parasitologia: distribuição de tipos de amostras - Tableau	100
Tabela 20 - Parasitologia: distribuição de tipos de amostras, por ano - Tableau	100

LISTA DE QUADROS

Quadro 1 - Artigos relacionados à Gestão do Conhecimento e Tecnologia da Informação ...	14
Quadro 2 - Sumário sobre dados, informação e conhecimento	21
Quadro 3 - Ficha cadastral e de solicitação de exame - Cedima.....	42
Quadro 4 - Ficha de resultados de exame - Cedima - Bacteriologia.....	43
Quadro 5 - Ficha de resultados de exame - Cedima - Virologia.....	43
Quadro 6 - Ficha cadastral, de solicitação de exames e resultados - Parasitologia.....	44
Quadro 7 - Ficha de resultados de exames - Patologia	44
Quadro 8 - Ficha de solicitação de exames - Laboratório de Clínica Animal - Hemograma .	46
Quadro 9 - Tipos de dados quanto à estrutura.....	48
Quadro 10 - Definições de “ <i>Data Science</i> ”	54
Quadro 11 - Script de preenchimento da tabela amostra – Cedima	60
Quadro 12 - Tipos de exame - Cedima	61
Quadro 13 - Cedima: listagem de gatos com positivo para <i>Streptococcus pneumoniae</i> - Python.....	95
Quadro 14 - Cedima: listagem de cães positivos para qualquer agente - Python	96
Quadro 15 - Patologia: exemplo de relatório com dados e diagnóstico – Python	103
Quadro 16 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para <i>Blastocystis hominis</i> e Cedima quaisquer resultados – Python	105
Quadro 17 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para <i>Schistosoma mansoni</i> e Cedima positivos - Python	105
Quadro 18 - Cedima x Parasitologia: cães, raça perdigueiro, positivos para isolamento e identificação bacteriana, Cedima, e que constam na parasitologia, independente do resultado – Python.....	105
Quadro 19 - Cedima x Parasitologia: gatos, com 16 anos ou menos, sexo feminino, com menos de 2 quilos e da raça Havana e que foram atendidos nos dois laboratórios - Python .	106
Quadro 20 - Cedima x Parasitologia: cães, no Cedima, com 11 anos ou mais, machos, com peso maior que 1 quilo e na Parasitologia com positivo para isolamento viral e qualquer resultado de exame - Python.....	106
Quadro 21 – Cedima x Parasitologia x Patologia: cães, yorkshire, antibiograma no Cedima, quaisquer exames Parasitologia e relação com o baço, na Patologia - Python	109
Quadro 22 - Cedima x Parasitologia x Patologia: felinos, SRD, positivos para FeLV no Cedima, qualquer resultado Parasitologia e com anotações sobre os rins, na Patologia - Python.....	110
Quadro 23 – Cedima, Parasitologia x Patologia: bovinos da raça holandês, Cedima, positivos para herpes bovina do tipo 1 e 5 e Patologia Animal algum achado relacionado ao fígado - Python.....	111

SUMÁRIO

1 INTRODUÇÃO	5
1.1 PERGUNTA DE PESQUISA	7
2 OBJETIVOS	7
2.1.1 OBJETIVO GERAL.....	7
2.1.2 OBJETIVOS ESPECÍFICOS	7
3 REVISÃO DA LITERATURA	9
3.1 INTRODUÇÃO.....	9
3.1.1 Aspectos Sobre A Gestão do Conhecimento	9
3.2 Metodologia da Revisão	12
3.3 Apresentação dos Resultados	13
3.4 Resumo da Base Teórica Encontrada	19
3.4.1 O Valor estratégico do conhecimento	19
3.4.2 Dados, Informação e Conhecimento	20
3.4.3 Gestão do Conhecimento.....	21
3.4.4 Conhecimento Tácito x Explícito	22
3.4.5 Criação de Conhecimento.....	23
3.4.6 Estratégias de Gestão de Conhecimento.....	24
3.4.7 Memória Organizacional	25
3.4.8 A Tecnologia da Informação	26
3.4.9 Os Desafios no Uso da Tecnologia da Informação	28
3.4.10 A Relação entre a Gestão do Conhecimento e a Tecnologia da Informação	29
3.4.11 Os Desafios na Relação entre a GC e a TI	30
3.5 CONSIDERAÇÕES FINAIS DA REVISÃO DE LITERATURA.....	30
4 METODOLOGIA	33
4.1 CARACTERIZAÇÃO DO ESTUDO	33
4.1.1 DEFINIÇÕES	33
4.1.2 A INTERDISCIPLINARIDADE NA GESTÃO DO CONHECIMENTO, TECNOLOGIA DA INFORMAÇÃO, SAÚDE ANIMAL, HUMANA E O AMBIENTE	34
4.2 LOCAL DE ESTUDO.....	37
4.3 PARTICIPANTES DA PESQUISA.....	38
4.3.1 Centro de Diagnóstico Microbiológico Animal (CEDIMA).....	38
4.3.2 Parasitologia e Doenças.....	39
4.3.3 Patologia Animal (LAPA).....	39
4.3.4 Patologia Clínica Veterinária (localizado no Hospital de Clínica Veterinária).....	39
4.3.5 Hospital de Clínica Veterinária (HCV)	39
4.4 Critérios de Inclusão e Exclusão dos Participantes	40
4.5 Critérios de Confidencialidade/ Privacidade	40
5 COLETA E APRESENTAÇÃO DOS DADOS	41
5.1 CEDIMA	41
5.2 PARASITOLOGIA E DOENÇAS PARASITÁRIAS	43

5.3 PATOLOGIA ANIMAL	44
5.4 LABORATÓRIO DE PATOLOGIA CLÍNICA VETERINÁRIA (HCV)	45
5.5 HOSPITAL DE CLÍNICAS VETERINÁRIAS (HCV)	46
6 ANÁLISE DOS DADOS LEVANTADOS	47
6.1 DISCUSSÃO	47
6.1.1 Quanto ao volume e à estrutura dos dados	47
6.1.2 Quanto aos meios de armazenamento	49
6.1.3 Quanto à segurança.....	50
6.1.4 Quanto à Gestão Eletrônica de Registros de Saúde Animal.....	51
6.1.5 Quanto ao Compartilhamento de Informações	52
6.1.6 Quanto às Análises Individuais	52
6.2 DETERMINAÇÃO DA TECNOLOGIA DE CRUZAMENTO DE DADOS	53
6.2.1 A importância dos <i>Insights</i>	56
6.3 PROVA DE CONCEITO	58
6.3.1 Bancos de Dados	58
6.3.2 Linux e VirtualBox	68
6.3.3 Tableau, Anaconda e Jupyter Notebook.....	70
6.3.4 Python.....	74
6.4 RESULTADOS DOS TESTES DA PROVA DE CONCEITO	75
6.4.1 Cedima.....	77
6.4.2 Parasitologia	96
6.4.3 Patologia Animal	102
6.4.4 Cedima x Parasitologia – Dados Interlaboratoriais	104
6.4.5 Cedima, Parasitologia e Patologia Animal – Dados Interlaboratoriais	108
7 CONSIDERAÇÕES FINAIS.....	116
8 REFERÊNCIAS	119
9 APÊNDICE A - Códigos-fonte Cedima: SQL e Python.....	127
10 APÊNDICE B - Códigos-fonte Parasitologia: SQL e Python.....	141
11 APÊNDICE C - Códigos-fonte Patologia Animal: SQL e Python.....	147
12 APÊNDICE D - Códigos-fonte Cedima x Parasitologia: SQL e Python	152
13 APÊNDICE E - Códigos-fonte Cedima, Parasitologia e Patologia Animal: SQL e Python.....	158
14 ANEXO A - Autorização de pesquisa - Parasitologia	163
15 ANEXO B - Autorização de pesquisa - HCV	164
16 ANEXO C - Autorização de pesquisa - Patologia	165
17 ANEXO D - Autorização de pesquisa - Cedima.....	166
18 ANEXO E - Serial Software Tableau	167

1 INTRODUÇÃO

A forma como são manipuladas as informações em instituições e empresas é um fator definidor do sucesso delas. Atualmente, na assim chamada “sociedade do conhecimento”, ele é o aspecto que mais afeta a produção, ultrapassando em valor tanto o capital quanto o trabalho (DRUCKER, 1993).

Seguindo nessa premissa, busca-se trabalhar de maneira inteligente com este conhecimento e gerenciá-lo seguindo padrões funcionais, planejados previamente. Sabendo que a produtividade é a fonte de vantagem competitiva, os setores passam a se apoiar em uma melhor gerência de seus conhecimentos (DRUCKER, 1991).

Assim, a Gestão do Conhecimento se apresentou recentemente como um conjunto de técnicas e conceitos que procuram padronizar e facilitar o gerenciamento do conhecimento nas organizações. Ela facilita a conversão de recursos em competências e por isso possui um papel importante na organização e no suporte em relação ao conhecimento (CHASE, 2006).

Da mesma maneira, a Tecnologia da Informação (TI) tem um papel importante na Gestão do Conhecimento, sendo usada por pessoas e principalmente organizações para alinhar-se ao ritmo das transformações mundiais, ajudando na melhora dos processos de produção, apoiando a tomada de decisão e o relacionamento com os *stakeholders* (ROSSETTI; MORALES, 2007).

Destaca-se o uso da TI no manejo de massas de informações, sejam grandes ou pequenas, por meio de sistemas de gerenciadores de banco de dados ou em servidores de busca distribuídos, por exemplo, usando técnicas como *Big Data* ou *Data Mining*, para a criação de conhecimento (DUARTE *et al.*, 2006). A tecnologia, se aplicada de maneira coesa e distribuída, integra-se com os indivíduos e seus conhecimentos pessoais, tornando-os úteis (DUARTE *et al.*, 2006).

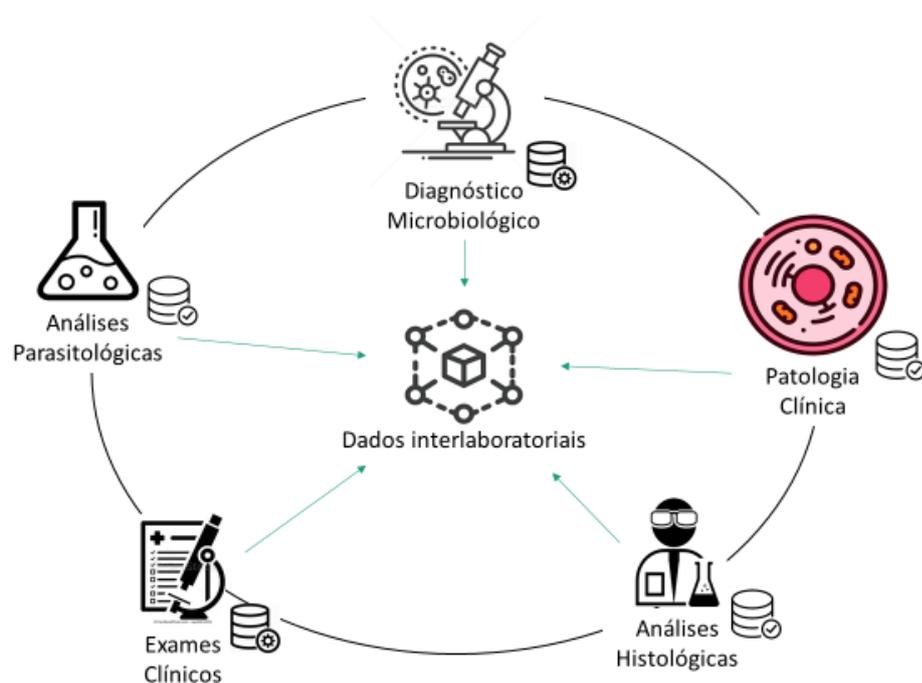
A memória organizacional é implantada em uma instituição para ser uma representação clara e duradoura do seu conjunto de conhecimentos e de suas principais informações, buscando facilitar o seu acesso, compartilhamento e reuso, pelos diversos membros da organização. Ela o faz ao facilitar a ampliação do saber, por absorver, sistematizar, transferir e reutilizar o conhecimento criado (FREIRE *et al.*, 2012).

Laboratórios ligados a análises de saúde animal, utilizados no ensino, pesquisa e extensão de uma universidade, recebem, criam e armazenam dados e informações continuamente, baseando grande parte de seu valor em sua capacidade de transformar uma visão

e medida da realidade complexa em conhecimento útil, e imediatamente disponível para tomadas de decisão e pesquisas científicas. Eles precisam se assegurar que os resultados produzidos reflitam, de forma fiel e sólida, o estado atual de um paciente, afastando a possibilidade de falhas no processo (CHAVES, 2010).

Essas instituições têm a o papel de analisar as suas amostras e o contexto de onde foram extraídas, de maneira a agregar valor às análises que precisam fazer, mas levando em conta que assim, isolado, um objeto pode perder a conexão com o seu contexto ambiental (MORIN, 2014). Assim, considera-se, a partir de agora, que os dados interlaboratoriais são informações produzidas, de diferentes laboratórios, que quando correlacionados com de outros (figura 1), tem o seu significado ampliado, possibilitando uma visão da realidade em que o indivíduo está inserido em seu meio ambiente, de maneira a expor suas possíveis relações.

Figura 1 - Ilustração da formação dos dados interlaboratoriais



Fonte: o autor (2019)

Assim, apesar de ter diferentes competências e gerar resultados distintos no primeiro momento, a interação, o cruzamento e rastreamento das informações obtidas em cada um dos laboratórios podem ganhar um significado relevante, a partir do momento que se utilizam as técnicas modernas de Gestão de Conhecimento e ferramentas de Tecnologia de Informação, criando dessa forma, os chamados Ativos de Conhecimento.

De maneira semelhante, os ativos de conhecimento (AC), também chamados de ativos intangíveis, são entendidos como recursos não palpáveis que geram valor para uma estrutura organizada e todos ligados à ela, fornecendo valor pelos privilégios que o seu controle oferece ao seu detentor (FERNANDES; FADEL, 2013). Nesse sentido, existem diferentes tipos de ativos de conhecimento dentro de uma organização, tais como, marcas, a imagem da empresa, bancos de dados, rotinas e práticas, por exemplo (CARLUCCI, 2012). Toda a massa de dados gerada, pelos resultados das análises de amostras enviadas aos laboratórios, pode ser considerada como AC, por carregarem consigo as informações de cada indivíduo que forneceu o material primário de análise, e também pela possibilidade de correlação entre as mesmas, o que refletiria em uma visão macro para o ambiente, o estado de saúde e a realidade do objeto da análise.

Por fim, o desenvolvimento de um modelo para analisar as informações de um grupo de laboratórios, seguindo um conceito pré-desenvolvido e planejado, permite gerar e armazenar novos conhecimentos, distribuí-los e utilizá-los em tudo que a organização tem à oferecer, como por exemplo, seus produtos, serviços (NONAKA; TAKEUCHI, 1997) e na pesquisa científica.

1.1 PERGUNTA DE PESQUISA

Como a Gestão de conhecimento, Data Science e os dados interlaboratoriais contribuem para gerar ativos de conhecimento de saúde animal?

2 OBJETIVOS

OBJETIVO GERAL

Compreender como a Gestão do conhecimento, Data Science e dados interlaboratoriais contribuem para gerar ativos de conhecimento de saúde animal.

OBJETIVOS ESPECÍFICOS

- 1) Realizar uma revisão bibliográfica para fundamentação dos assuntos tratados
- 2) Verificar como e quais dados são gerenciados atualmente nos laboratórios e suas especificidades.
- 3) Pesquisar a tecnologia de análise de dados disponível mais indicada para o cruzamento de dados interlaboratoriais e que viabilize a criação de ativos de conhecimento.

- 4) Propor um modelo utilizando ferramentas de TI, para a correlação dos dados, habilitando a prospecção de ativos de conhecimento de saúde animal, testando a tecnologia encontrada.

3 REVISÃO DA LITERATURA

A revisão de literatura será apresentada no formato de artigo de revisão.

3.1 INTRODUÇÃO

Para que sejam identificados os aspectos da relação que existe entre a Gestão do Conhecimento (GC) e a Tecnologia da Informação (TI), na base de conhecimento científico atual, e servir como fundamentação teórica do estudo, foi realizada uma revisão de literatura, buscando evidências por meio de métodos sistematizados de pesquisa, leitura crítica e síntese das informações, e, ao final identificando os temas que necessitam de evidência para ajudar na orientação das investigações futuras (SAMPAIO; MANCINI, 2007).

Aspectos Sobre A Gestão do Conhecimento

O conhecimento é considerado um recurso estratégico, e, como vários outros recursos, pode ser adquirido. Em contrapartida, a inteligência é mais fugidia, por ser uma característica humana e não tangível. Conhecimento e inteligência estão conectados: a inteligência é necessária para se gerar conhecimento e o conhecimento fornece a base através da qual a inteligência pode ser aplicada (PRUSAK, 1997). Este conhecimento normalmente não é gratuito, para uma empresa ou organização, precisando ser adquirido, pela contratação de pessoal treinado, ou gerado, através de sistemas de gerenciamento de transações ou de suporte à decisão, o que sempre incorre em custos (PRUSAK, 1997).

Por ser um conceito multifacetado e com vários significados possíveis, é preciso entendê-lo, esclarecendo inicialmente sua diferença em relação ao termo “informação”: informação é um fluxo de mensagens ou significados que podem incrementar, reestruturar ou mudar um conhecimento, que é criado e organizado por este fluxo das informações, ancorado pelo comprometimento e crenças de seu possuidor (NONAKA, 1994).

Nonaka e Takeuchi (1997), considerados como alguns dos mais importantes pesquisadores nesta área, dividem o conhecimento entre tácito (pessoal e de difícil formalização) e explícito, que pode ser mais facilmente processado por sistemas computacionais, e por isso mesmo transmitido e armazenado. Para que o conhecimento tácito

possa ser compartilhado, deve ser convertido em explícito, de modo que qualquer um possa entender, gerando o chamado Conhecimento Organizacional, durante esta transformação.

Eventualmente reconheceu-se o valor do conhecimento, criando a necessidade, entre os responsáveis pelas gerências das organizações, de procurar saber mais sobre ele, como é distribuído e de como é usado, especificamente quando a base das economias industrializadas mudou da simples exploração de recursos naturais para o uso também de ativos intelectuais; ao mesmo tempo, como nunca antes, o uso crescente das redes de computadores e sua presença por todo o mundo permitiu, com baixo custo, codificar, armazenar e compartilhar certos tipos de conhecimento mais facilmente (HANSEN; NOHRIA; TIERNEY, 1999). É preciso entender que foi a aplicação do conhecimento ao trabalho que criou as economias desenvolvidas atuais, ao instigar enormemente a produtividade nos últimos cem anos (DRUCKER, 1993).

As vantagens competitivas de uma empresa dependem de seu conjunto de conhecimentos, de como se aproveita dele e de sua capacidade de aprender algo novo rapidamente (PRUSAK, 1997). Como o conhecimento sempre foi um bem privado e controlado por poucos, em tempos anteriores e, quase do dia para a noite, se tornou um bem público (DRUCKER, 1993), entende-se por que hoje é tratado com um ativo nas empresas, sendo chamado de “ativo de conhecimento“ ou ”intangível” e o papel central que adquiriu.

A gestão do conhecimento não é algo novo. Através dos tempos, os trabalhadores vêm trocando ideias e técnicas em relação a suas atividades, artesãos ensinando aos seus aprendizes e donos de negócios de família passando seus conhecimentos para os descendentes, mas foi somente na década de 1990 em que as empresas começaram a tratar especificamente sobre o assunto, com o devido interesse (HANSEN; NOHRIA; TIERNEY, 1999).

A GC é a ferramenta que busca identificar e alinhar o conhecimento coletivo em uma organização, tentando melhorar a sua competitividade, juntamente com sua capacidade de inovar e de responder às mudanças (ALAVI; LEIDNER, 2001). De acordo com Davenport e Prusak (1998), a maioria dos projetos de GC tem três focos: (1) construir uma infraestrutura de conhecimento (e não somente um sistema especialista), um conjunto de interconexões pessoais, que ofereça as ferramentas e o incentivo para a colaboração e a interação entre as pessoas, (2) tornar o conhecimento acessível e destacar o papel dele na organização, através de várias técnicas, como documentos de fácil acesso e atualização, por exemplo, e (3) difundir uma cultura baseada em conhecimento e que encoraje o compartilhamento, para evitar a sua concentração em poucos indivíduos e assim o retrabalho.

Utilizar os recursos que já existem na organização para que os usuários tenham acesso e possam usufruir das melhores práticas e das lições aprendidas, sem precisar reinventar o que já havia sido criado. Agregar valor às informações, selecionando, resumindo e sintetizando-as, para permitir acesso à informação necessária: este é o tema central da Gestão do Conhecimento (PEREIRA; SILVA; PINTO, 2016).

A Memória Organizacional é, dentro da GC, o registro e a organização destes conhecimentos e experiências, gerados na instituição, que acabam por formar um banco de dados, passível de ser manipulado como um grande conjunto de arquivos e depois compartilhado pelo grupo de usuários, envolvendo aspectos técnicos, sociais, funcionais e também do ambiente de trabalho (MOLINA; VALENTIM, 2011), se tornando essencial para o aprendizado e a adaptação organizacional. Este é o mapa do seu passado institucional (PEREIRA; SILVA; PINTO, 2016).

Nesse sentido, a Tecnologia da Informação se ocupa, como uma de suas principais funções, da aquisição, armazenamento, processamento e organização de informações e dados e com a sua transmissão para o uso das pessoas e organizações (RAJAMARAN, 2018); dessa maneira, ajuda a atenuar as dificuldades que envolvem a gerência do conhecimento, evitando que seja distorcido ou explorado de maneira errada (PRUSAK, 1997). Por esse motivo, acaba por desempenhar um papel na maioria das iniciativas de GC, em menor ou em maior grau, como uma ferramenta para a sua aplicação (ALAVI; LEIDNER, 2001), principalmente por implementar a Memória Organizacional nativamente, como um de seus principais papéis, desde sua fundação.

Algumas das características mais atraentes da TI é que ela está cada vez mais transparente para o usuário, é onipresente, e já possui a capacidade de capturar o conhecimento real, complexo, e não somente dados ou palavras, simplesmente (DAVENPORT; BARTH; BEAN, 2012; PRUSAK, 1997). As redes de comunicação interativas existem graças à novas tecnologias de telecomunicação e estão colocando pessoas que trabalham com o conhecimento em contato umas com as outras, como nunca. À medida que estas tecnologias se tornem mais ricas, computadores pode tomar um papel ainda maior na habilitação do conhecimento: como agentes de transformação (PRUSAK, 1997; VENKITACHALAM; AMBROSINI, 2017). A TI não mudará somente como trabalhamos, mas o que somos (DAVENPORT; PRUSAK, 1998)

Isto posto, entende-se que o conhecimento é um recurso escasso e valioso, e que a GC, utilizando da Memória Organizacional, juntamente com a TI, ajudam a manejar, armazenar e reproduzir o conhecimento, trazendo valor, competitividade e agilidade às instituições.

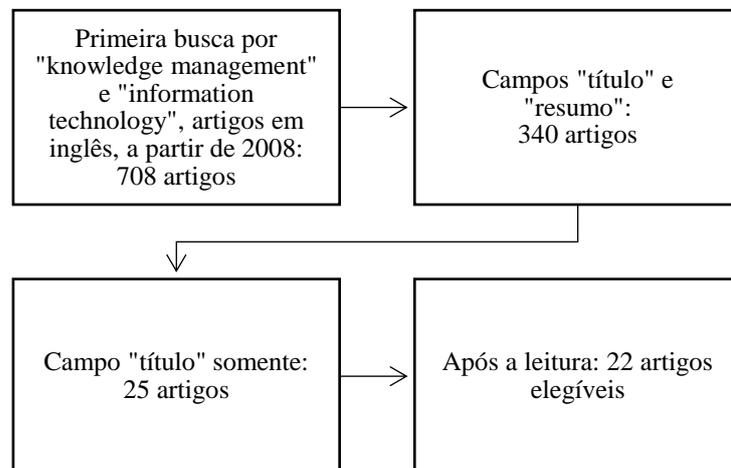
3.1 METODOLOGIA DA REVISÃO

Iniciou-se pela definição dos critérios de inclusão e exclusão dos artigos a serem pesquisados, tais como palavras-chave, campos de procura, tipos de documento, idioma e ano de publicação para limitação do alcance da procura. Como o foco do estudo foi esclarecer o papel da Tecnologia da Informação na Gestão do Conhecimento, o eixo norteador foi a procura por artigos que tratem destes assuntos. Em um segundo momento, procedeu-se a busca destas palavras-chave na base de dados de artigos *online* Scopus, e, por fim, discutiu-se a análise dos resultados na tentativa de resumir o que foi encontrado sobre a relação entre a Gestão do Conhecimento e a TI.

A plataforma Scopus, ligada à empresa Elsevier, foi a base de dados consultada por ser uma plataforma digital que fornece acesso a trabalhos científicos e resultados de pesquisas em várias áreas, como as de Ciências Humanas, Biológicas e Exatas e ter o perfil desejado de publicações. Ela permite o acesso público a publicações, páginas de conteúdo científico, patentes e anais de conferência, sendo a maior base de dados do mundo, em número de resumos e citações, com 53.000.000 de registros, mais de 20.000 títulos e 5.000 publicações, além de mais de 1.200 revistas e cerca de 500 anais de conferência. Oferece ainda resumos e citações de 130 mil livros, com periódicos revisados de editoras internacionais, como a Cambridge University Press, Springer Nature e Wiley (BRASIL, 2017; ELSEVIER, 2018).

Neste sistema foram testadas várias combinações de termos de procura, em junho de 2018, usando as palavras “*knowledge management*” (gestão de conhecimento) e “*information technology*” (tecnologia da informação). Estes termos foram utilizados, inicialmente, nos campos “título do artigo”, “resumo” e “palavras-chave”, simultaneamente, restringindo a documentos do tipo “artigo”, escritos a partir de 2008 e em publicações somente em inglês. A figura 2 apresenta um resumo dos passos seguidos na busca em base de dados.

Figura 2 - Fluxograma da busca na base de dados de artigos sobre a temática proposta



Fonte: o autor (2019)

A procura somente no título foi mantida pela grande quantidade de documentos retornados nas primeiras buscas e pela ocorrência de peças científicas que fogem grandemente do foco da pesquisa, principalmente as que não tem a TI como um dos enfoques.

Cada texto foi analisado na tentativa de identificar se relacionava ou não a Gestão do Conhecimento com a TI, o método utilizado, objetivo e resultados alcançados.

A leitura buscou aqueles que de alguma maneira tentam definir ou tratam, mesmo que de maneira indireta, do impacto da TI na Gestão do Conhecimento, descrevendo-os em vários ambientes, como o comercial, de pequenas e médias empresas, no setor de produção, de saúde, agricultura, governamental e no educacional, notadamente de nível superior. Os materiais que não mantinham essa relação com o assunto entraram no critério de exclusão e foram desconsiderados do capítulo de revisão bibliográfica e fundamentação teórica.

3.2 APRESENTAÇÃO DOS RESULTADOS

Os dados completos sobre os artigos encontrados se encontram no quadro 1, em ordem decrescente de data de publicação.

Quadro 1 - Artigos relacionados à Gestão do Conhecimento e Tecnologia da Informação

Nº	Informações bibliográficas	Objetivo	Resultados
1	FOOTE, A.; HALAWI, L. A. Knowledge management models within information technology projects. Journal of Computer Information Systems , v. 58, n. 1, p. 89–97, 2018.	Verificar o uso da Gestão do Conhecimento em gerência de projetos, na implementação de projetos de TI para fins de organização.	Revelou quatro temas em Gestão de Conhecimento, para o sucesso do projeto de TI: especificação de requerimentos, desenvolvimento de código, testes e suporte.
2	TURULJA, L.; BAJGORIĆ, N. Information Technology, Knowledge Management and Human Resource Management: Investigating Mutual Interactions towards Better Organizational Performance. VINE Journal of Information and Knowledge Management Systems , 2017.	Oferecer um framework conceitual para esclarecer a natureza dos efeitos da TI, Gestão do Conhecimento e competências de RH no desempenho organizacional empresarial.	Identificou-se a relação entre os recursos mais proeminentes nas empresas na era digital: conhecimento, RH e TI. Os resultados mostram causas e efeitos nas relações estes recursos.
3	VENKITACHALAM, K.; AMBROSINI, V. A triadic link between knowledge management, information technology and business strategies. Knowledge Management Research & Practice , v. 15, n. 2, p. 192–200, 2017.	Melhor entender como a Gestão do Conhecimento influencia a TI e vice-versa dentro do contexto da estratégia de negócios, através de uma síntese da literatura.	Desenvolveu uma representação diagramática e também articulada na forma de seis proposições conectando as três estratégias (negócios, Gestão do Conhecimento e TI).
4	MOHAMAD, A. A.; RAMAYAH, T.; LO, M. C. Knowledge management in MSC Malaysia: The role of information technology capability. International Journal of Business and Society , v. 18, n. S4, p. 651–660, 2017.	Investigar as dimensões da Gestão do Conhecimento, TI e a inovação em empresas e examinar o impacto das TI em mediar a conexão entre as dimensões da Gestão do Conhecimento.	Destacam que a conversão e a proteção do conhecimento estão relacionadas de maneira positiva e significativa à inovação na empresa. As técnicas de TI foram usadas para mediar a conexão entre a conversão de conhecimento e a proteção do conhecimento.
5	MAO, H. <i>et al.</i> Information technology resource, knowledge management capability, and competitive advantage: The moderating role of resource commitment. International Journal of Information Management , v. 36, n. 6, p. 1062–1074, 2016.	Estender a literatura sobre Gestão do Conhecimento relacionada à TI ao investigar o papel moderador no comprometimento de recursos.	Identificou dois efeitos positivos semi-atenuantes de comprometimento de recursos na relação entre TI e Gestão do Conhecimento. Especificamente, o comprometimento de recursos direta e positivamente valoriza a Gestão do Conhecimento, e fortalece os efeitos da relação entre pessoal de TI e seus recursos na Gestão do Conhecimento.
6	MAROOFI, F. Sociability Organizations Can Facilitate Knowledge Management Through Information Technology? International Business Management , v. 10, n. 13, p. 2551–2555, 2016.	Verificar se as organizações sociáveis podem facilitar sua Gestão do Conhecimento através da TI.	Sugere que as organizações sociáveis necessitam rever suas iniciativas de TI de maneira estratégica.
7	HENDRIKS, P. H. J.; LIGTHART, P. E. M.; SCHOUTETEN, R. L. J. Knowledge management, health	Fornecer uma visão em como a TI em saúde (HIT) afeta o conhecimento explícito e tácito	O conhecimento tácito das enfermeiras aparenta ser um indicador de seu

	information technology and nurses' work engagement. Health Care Management Review , v. 00, n. 3, p. 1, 2015.	em seus processos de trabalho em andamento e seu engajamento no trabalho.	engajamento no trabalho. O conhecimento explícito somente indiretamente afeta seu engajamento. O uso de HIT afeta de maneira significativa somente seu conhecimento explícito, não seu conhecimento tácito ou engajamento.
8	LIN, H. An investigation of the effects of cultural differences on physicians' perceptions of information technology acceptance as they relate to knowledge management systems. Computers in Human Behavior , v. 38, p. 368–380, 2014.	Explorar o impacto das diferenças culturais nacionais na percepção de médicos sobre a aceitação de sistemas de Gestão do Conhecimento.	Diferenças culturais mostraram ter impacto na aceitação de sistemas de Gestão do Conhecimento.
9	CHUANG, S.-H.; LIAO, C.; LIN, S. Determinants of knowledge management with information technology support impact on firm performance. Information Technology and Management , v. 14, n. 3, p. 217–230, 2013.	Melhor entender os efeitos da Gestão do Conhecimento na performance da empresa, focando nos seus efeitos na performance, com o suporte da TI.	Os resultados confirmam os efeitos dos fatores organizacionais do suporte da TI, que se mostrou crítica para melhorar a performance das empresas.
10	OKUMUS, F. Facilitating knowledge management through information technology in hospitality organizations. Journal of Hospitality and Tourism Technology , v. 4, n. 1, p. 64–80, 2013.	Discutir como organizações de hospedagem podem facilitar a Gestão do Conhecimento através de ferramentas de TI.	A Gestão do Conhecimento e a TI podem ajudar organizações de hospedagem a criar e sustentar uma vantagem competitiva.
11	HUANG, H. L. <i>et al.</i> The relationship between knowledge management strategy and information technology strategy. World Academy of Science, Engineering and Technology , v. 77, n. 5, p. 432–436, 2011.	Desenvolver um modelo de pesquisa para explicar a relação entre estratégias de Gestão do Conhecimento e de TI e seus efeitos no desempenho	Proposições e conclusões foram derivadas, e sugestões para pesquisas futuras são propostas e discutidas
12	KUO, Y.-K.; YE, K.-D. How employees' perception of information technology application and their knowledge management capacity influence organisational performance. Behaviour & Information Technology , v. 29, n. 3, p. 287–303, 2010.	Investigar como o as características de um empregado afetam sua percepção da TI, capacidade de auto percepção em Gestão do Conhecimento e percepção do desempenho organizacional.	Mostram uma correlação positiva entre a utilização de TI, Gestão do Conhecimento e desempenho organizacional.
13	SITARSKI, K. The Role of Information Technology Systems in Knowledge Management. Foundations of Management , v. 2, n. 1, p. 117–132, 2010.	Pesquisa no modelo de sistema organizacional de Gestão do Conhecimento em TI.	Foi construído de um sistema de Gestão do Conhecimento e o modelo foi verificado tendo por base a pesquisa do protótipo do sistema.
14	CHUNG, Y. C. <i>et al.</i> Relationship between organizational cultures, information technology involvement, degrees of knowledge management Implementations and performance of new product developments. Information Technology Journal , v. 9, n. 7, p. 1504–1516, 2010.	Com a indústria de alta tecnologia de Taiwan como objeto, pesquisou sobre a correlação entre as culturas organizacionais, uso de TI, graus de uso da Gestão do Conhecimento e desempenho no desenvolvimento de novos produtos.	Revelou que quanto maior o grau de envolvimento da cultura organizacional e da TI, mais significativa a influência na implementação da Gestão do Conhecimento e no desempenho no desenvolvimento de novos produtos.

15	REVILLA, E.; RODRÍGUEZ-PRADO, B.; PRIETO, I. Information technology as knowledge management enabler in product development: Empirical evidence. European Journal of Innovation Management , v. 12, n. 3, p. 346–363, 2009.	Focar em um habilitador da Gestão do Conhecimento: a TI, e explorar sua influência no desenvolvimento de produtos.	Demonstrou-se que as diferenças na configuração de TI no desenvolvimento de produtos podem levar a diferenças nos termos de aproveitamento do conhecimento e também as vantagens de uma configuração equilibrada da TI
16	CAVALCANTI, M. Knowledge management: Using information technology to obtain knowledge in a competitive environment. Corporate Ownership & Control , v. 7, n. 1 B CONT. 1, p. 232–243, 2009.	Enfatizar a importância da governança corporativa, sistemas de informação e seu conhecimento teórico e prático nos próximos anos.	Mostrou que os estudantes ainda não graduados de cursos em Administração de Empresas têm um nível baixo de conhecimento em Sistemas de Informação e Pesquisa em Marketing, enquanto estão na graduação, e este fato permite uma visão analítica dos riscos no tratamento da governança.
17	RASLI, A. M.; WAN MOHD, M. W. Project performance framework: The role of knowledge management and information technology infrastructure. Asian Journal of Business and Accounting , v. 1, n. 2, p. 39–64, 2008.	Investigar de maneira empírica se a Gestão do Conhecimento e a TI afetam a performance de projetos, e identificar os fatores e barreiras que influenciam a sua implementação, para ajudar a performance em companhias de consultoria em construção.	Propôs conceitos que interligam a Gestão do Conhecimento, TI e a o desempenho em projetos, para contribuir para o desempenho pelo aumento da qualidade de provisionamento e diminuição de custos através do gerenciamento do ciclo de vida do projeto.
18	WILD, R.; GRIGGS, K. A model of information technology opportunities for facilitating the practice of knowledge management. VINE Journal of Information and Knowledge Management Systems , v. 38, n. 4, p. 490–506, 2008.	Apresentar uma técnica de Gestão do Conhecimento e um modelo para ajudar a identificar as oportunidades nas quais a TI podem facilitar as práticas de Gestão do Conhecimento, através da síntese da literatura.	Descreve um sistema protótipo de simulação distribuída que demonstra a habilidade potencial da TI para realizar as atividades de Gestão do Conhecimento e contribuir para a sua efetiva aplicação.
19	LIN, D. <i>et al.</i> Does knowledge management matter for information technology applications in China? Asia Pacific Journal of Management , v. 25, n. 3, p. 489–507, 2008.	Explorar o papel da Gestão do Conhecimento na mediação e moderação das relações entre a TI e o desempenho organizacional, baseado nos dados coletados de 236 empresas da China.	Pelo uso de um modelo, descobriu-se que os as técnicas da Gestão do Conhecimento parcialmente mediam o impacto da performance nas técnicas de TI. Além disso, a Gestão do Conhecimento afetou a força da relação entre a TI e o desempenho organizacional.
20	TSENG, S. M. The effects of information technology on knowledge management systems. Expert Systems with Applications , v. 35, n. 1–2, p. 150–160, 2008.	Explorar o papel e o efeito da TI na implementação da Gestão do Conhecimento em empresas, usando revisão de literatura, entrevistas com experts e análises de questionários.	Discutiu como melhorar a efetividade e a eficiência da implementação de Gestão do Conhecimento através da TI mais apropriada.
21	PILLANIA, R. K. Information technology strategy for knowledge	Pesquisar sobre a TI na Gestão do Conhecimento em pequenas	Os problemas variam entre o contexto internacional e o

	management in Indian automotive components SMEs. Knowledge and Process Management , v. 15, n. 3, p. 203–210, 2008.	e médias empresas, em particular, no contexto da Índia, através de revisão da literatura.	indiano. Os maiores, na Índia, são a falta de treinamento, problemas técnicos que afetam o sucesso e falta de atualização dos usuários pela comunicação pobre, entre outros
22	CHANG, Y. H.; CHEN, J. W.; LIN, B. KMsharer: an information technology approach to enable knowledge management services. International Journal of Technology Management , v. 43, n. 1–3, p. 252–265, 2008.	Contribuir para as pesquisas sobre sistemas de Gestão do Conhecimento ao propor um ambiente de Gestão do Conhecimento que facilite o ciclo evolutivo do conhecimento, ao desenvolver um sistema de gestão de conhecimento.	O KMsharer é ferramenta resultante, utilizada para acessar, criar e armazenar informações de maneira a melhorar o compartilhamento, colaboração e a agilização da tomada de decisão estratégica.

Fonte: dados da pesquisa (2019)

Como mostra a tabela 1, nota-se, nos resultados, uma maior concentração das publicações no ano de 2008, com uma queda contínua até 2014 e um aumento lento, até 2018. A média foi 2,5 artigos publicados por ano.

Tabela 1 - Frequência das publicações por ano

Ano	Frequência
2008	6
2009	2
2010	3
2011	1
2013	2
2014	1
2015	1
2016	2
2017	3
2018	1

Fonte: O autor (2019)

Sobre as fontes de publicação, os artigos se distribuíram entre vinte e dois periódicos científicos. Destes, nove tratam de mais de um tópico em suas edições, como gestão de empresas, contabilidade, comportamento humano, inovação, saúde humana, Tecnologia da Informação, negócios e sociedade, turismo e hospitalidade, gestão de processos, ciência e engenharia, mostrando a forte presença da interdisciplinaridade neste campo.

Em relação ao interesse de pesquisa, o foco principal foi sobre o uso da GC e a TI e como a relação entre elas contribuem para outras áreas do conhecimento, totalizando doze

artigos que referem-se ao desempenho nas empresas, seja na produção, no desenvolvimento de projetos ou na área organizacional com influência positiva e em alguns casos críticos, com efeitos positivos na produtividade (CHUANG; LIAO; LIN, 2013; HUANG *et al.*, 2011; KUO; YE, 2010; LIN *et al.*, 2008; RASLI; WAN MOHD, 2008; TURULJA; BAJGORIĆ, 2017).

Em segundo, a cultura organizacional se mostrou recorrente nas citações (CAVALCANTI, 2009; CHUANG; LIAO; LIN, 2013; CHUNG *et al.*, 2010; MAO *et al.*, 2016; OKUMUS, 2013), com cinco artigos, sendo descrito como um fator determinante e de impacto positivo no uso da GC e a TI. Um ponto citado por Lin (2014), é a influência da cultura individual na aceitação do uso e da implementação de políticas de GC e TI, citando as influências de seus modelos comportamentais neste mister. Foram colocados como interferências chave também a nacionalidade, mas admitindo que o corpo de saber nesta área é ainda limitado, necessitando de mais pesquisas.

Em terceiro lugar, um total de quatro autores (CHANG; CHEN; LIN, 2008; SITARSKI, 2010; BAJGORIĆ, 2017; WILD; GRIGGS, 2008) tiveram como assunto o desenvolvimento de modelos ou *frameworks* para utilização em momentos de implantação de políticas de GC relacionadas à TI em organizações, identificando oportunidades de uso e seus efeitos. Em alguns casos (CHANG; CHEN; LIN, 2008; SITARSKI, 2010), a investigação deu origem à protótipos e produtos finalizados, que posteriormente entraram em comercialização, como o descrito por Chang, Chen e Lin (2008).

Na área da saúde, em três artigos, definiu-se que há um consenso sobre a importância do uso destas ferramentas em hospitais e sua influência no trabalho de médicos e enfermeiras (HENDRIKS; LIGTHART; SCHOUTETEN, 2015; LIN, 2014; OKUMUS, 2013) e, com a visão até mesmo sobre vantagens competitivas que oferecem, tópico que poderíamos não pensar estar associado à saúde, mas que é importante para o setor privado. Do mesmo modo, a GC e a TI demonstram afetar de maneira positiva como os profissionais de saúde se relacionam com seu conhecimento explícito, sendo a cultura pessoal um fator decisivo na aceitação destas tecnologias.

Outro tema são as pequenas e médias empresas, citadas em dois trabalhos (LIN *et al.*, 2008; PILLANIA, 2008), com atenção aos problemas na implantação de políticas de GC e TI simultaneamente, mas reforçando o seu papel chave e definidor no bom desempenho organizacional.

Em se tratando de lacunas, o resultado da busca demonstra o número reduzido de pesquisas (SITARSKI, 2010; TSENG, 2008; WILD; GRIGGS, 2008) tratando da relação entre

GC e TI de forma pura, sem que esteja ligada a algum outro assunto. Nos dias atuais, em que a TI é ubíqua e a necessidade de gerência de conhecimento é amplamente aceita e entendida como vital, é necessário que cada vez mais se fundamente esta relação através do método científico.

Nota-se também a pouca atenção em relação às ferramentas avançadas de TI para a gestão de dados e tomada de decisão, mais atuais, como a inteligência artificial, *business intelligence*, *Data Science* (ciência de dados) e toda a matriz de conhecimento que as ferramentas relacionadas à *big data* podem oferecer.

3.3 RESUMO DA BASE TEÓRICA ENCONTRADA

Em relação ao conjunto teórico que os autores apresentam, segue abaixo um resumo de suas posições sobre ao binômio GC e TI e dos outros conteúdos pertinentes que se apresentaram durante a leitura e a análise.

O Valor estratégico do conhecimento

Atualmente, entende-se que o único ativo de uma organização que pode dar a ela vantagem na competição com outras, e que é passível de ser gerado e replicado, é o conhecimento que ela possui (VENKITACHALAM; AMBROSINI, 2017). Em vista das mudanças que ocorreram nos modelos de negócios e a sua globalização, é preciso destacar as diferenças entre ativos tangíveis, como o capital físico e propriedades da empresa, por exemplo, dos intangíveis, que são as rotinas organizacionais e o conjunto de conhecimentos que ela gera e manipula (TURULJA; BAJGORIĆ, 2017).

Por isso, é hoje citado como o único recurso econômico de valor e é reconhecido como a fonte primária de vantagem competitiva de uma empresa (OKUMUS, 2013). Na economia atual, baseada em informações, o sucesso é dependente da adaptabilidade da organização e da aplicação correta de seus ativos intangíveis nos seus processos chave (KUO; YE, 2010). O conhecimento é um ativo importante para uma organização na geração de valor (MAO *et al.*, 2016), tanto que, para alguns autores, o capital de uma empresa é majoritariamente feito de conhecimento e organização (SITARSKI, 2010).

Pode-se dizer que o conhecimento possui alguns valores associados à si, como a sua durabilidade (por quanto tempo será útil), transparência (período que levará para que os

competidores alcancem este mesmo conhecimento), replicabilidade (quanto tempo levará para que os competidores possam usar este mesmo conhecimento) e capacidade de transferência (quão fácil é para que os competidores possam copiar este conhecimento) (OKUMUS, 2013). Ele pode ser explicado como um agrupamento de razão, intuição, detalhes e projetos de produtos, trocas e relações com o cliente, processos ou similares e que possibilitam ações concretas na organização (PILLANIA, 2008). É a base de sustentação sobre a qual ocorrem novas experiências e se criam ou se adquirem novas informações (SITARSKI, 2010).

Assim, a produtividade ou valor das corporações modernas se apoiam mais em suas bases intelectuais e de seus sistemas do que em seus ativos tangíveis, com raras exceções (WILD; GRIGGS, 2008).

Dados, Informação e Conhecimento

A gerência da informação começa pelos indivíduos e deve se difundir por toda a organização. Dados, que são informações atômicas, podem ser convertidos em informações complexas e também conhecimento, mas são modificados de acordo com os pontos de vista de quem os maneja e os fins para os quais se destinam (FOOTE; HALAWI, 2018).

Por isso, entende-se que os conceitos de conhecimento, informação e dados são diferentes. A informação só faz sentido quando está conectado à um processo de interpretação, o que o relaciona à mente, que lhe dá sentido, e à fatos, que o amplia e explica. O conhecimento reside no pensamento, sendo por isso um ato humano; ele pertence à um indivíduo ou grupo e neste caso circula e se modifica por ele de várias maneiras (CAVALCANTI, 2009; OKUMUS, 2013). Novos conhecimentos são criados graças aos anteriores e sobre eles se formam. Não têm sentido quando desconectados daqueles que o interpretam, ampliam e exercitam (OKUMUS, 2013).

Já a informação é a base sobre a qual se gera o conhecimento, e, por consequência, quanto mais informações temos sobre algo, proporcionalmente aumenta o conhecimento sobre ele (HENDRIKS; LIGTHART; SCHOUTETEN, 2015). O conhecimento, por outro lado, pode ser gerado pela organização, utilizando das informações que possui, ou ser adquirido de fontes externas e o seu processamento e dinamização precisa de uma combinação de sistemas humanos e automatizados, como os informáticos, tão difundidos atualmente (OKUMUS, 2013). O quadro 2 sumariza as características de cada uma destas entidades.

Quadro 2 - Sumário sobre dados, informação e conhecimento

Dados	Informação	Conhecimento
<p>Simple observações sobre o estado do mundo:</p> <ul style="list-style-type: none"> • facilmente estruturado; • facilmente obtido por máquinas; • frequentemente quantificado; • facilmente transferível. 	<p>Dados dotados de relevância e propósito:</p> <ul style="list-style-type: none"> • requer unidade de análise; • exige consenso em relação ao significado; • exige necessariamente a mediação humana. 	<p>Informação valiosa da mente humana.</p> <p>Inclui reflexão, síntese, contexto.</p> <ul style="list-style-type: none"> • de difícil estruturação; • de difícil captura em máquinas; • frequentemente tácito; • de difícil transferência.

Fonte: Vieira (2014)

3.4.3 Gestão do Conhecimento

A Gestão do Conhecimento (GC), apesar de não ter tido sempre esta designação, tem sido praticada por muito tempo em organizações que aprenderam a valorizar o compartilhamento de conhecimento. Sua definição, segundo alguns autores, ainda é incompleta, por causa das múltiplas interpretações possíveis. A GC é uma disciplina nova e está evoluindo constantemente (CHUANG; LIAO; LIN, 2013). Não há uma definição única universalmente aceita, no momento, para a GC.

Tem como uma de suas metas a criação de valor a partir dos ativos intangíveis de uma organização, coordenando e explorando os recursos ligados e originados do conhecimento com o fim de criar benefícios e vantagens competitivas. Pretende também organizar as fontes de conhecimento desde a sua formação, controlando a distribuição e a sua aplicação (MOHAMAD; RAMAYAH; LO, 2017; WILD; GRIGGS, 2008). A GC pode ser entendida como um processo sistêmico e de origem organizacional para criar, integrar, transferir e alinhar conhecimentos correlatos, de maneira que aqueles originados em uma unidade da organização sejam aplicáveis à outras, com o máximo de aproveitamento para a instituição (CHUANG; LIAO; LIN, 2013).

A GC se relaciona fortemente com a visão e a missão da organização, influenciando na tomada de decisão através da captação e interpretação de informações vitais. Ela demonstrou, através de estudos prévios, ter a capacidade de aumentar a inovação na empresa e ampliar a sua competitividade (MOHAMAD; RAMAYAH; LO, 2017).

Para isso se vale dos chamados “trabalhadores do conhecimento”, os responsáveis em interpretar os dados e informações disponíveis, ressignificando-os (WILD; GRIGGS, 2008). O

que diferencia a GC de outras técnicas organizacionais é o foco neste trabalhador do conhecimento, que é o agente de sua transformação para que gere valor (HENDRIKS; LIGTHART; SCHOUTETEN, 2015).

Conhecimento Tácito x Explícito

O conhecimento que pode ser capitalizado pelas empresas, também chamado de Conhecimento Organizacional (LIN *et al.*, 2008), pode ser dividido em explícito ou tácito. Conhecimento explícito é de fácil entendimento e pode ser organizado e armazenado em sistemas computadorizados, enquanto o conhecimento tácito se refere às informações práticas, incorporadas ao trabalhador, difíceis de serem descritas em palavras. Geralmente são informais, intuitivas, baseadas em experiências pessoais, que podem ser subjetivas e complexas para serem formalizadas e codificadas (CAVALCANTI, 2009; KUO; YE, 2010; LIN *et al.*, 2008; OKUMUS, 2013).

O conhecimento explícito pode ser expressado em uma linguagem formal e ser compartilhado na forma de dados, especificações, fórmulas científicas, manuais descritivos ou similares (CAVALCANTI, 2009; HENDRIKS; LIGTHART; SCHOUTETEN, 2015). Em contrapartida, conhecimento tácito é altamente pessoal e difícil de formalizar. É profundamente enraizado em rotinas, ações, procedimentos, ideais, valores e emoções. Conhecimento explícito e tácito não podem simplesmente ser combinados como em uma equação matemática. Possuem uma conexão mais profunda (CAVALCANTI, 2009; HENDRIKS; LIGTHART; SCHOUTETEN, 2015). Segundo Wild e Griggs (2008, p. 8) “muitos argumentam que o conhecimento tácito é informação processada pelo cérebro humano, e, por isso, poderia existir somente na mente de um indivíduo” (tradução nossa)¹.

Ou, como resume Cavalcanti (2009, p. 3) “conhecimento explícito pode ser definido como ‘saber sobre’, enquanto conhecimento tácito é ‘saber como’” (tradução nossa)².

O desafio para a maioria das organizações é capturar e disseminar o conhecimento tácito (OKUMUS, 2013; WILD; GRIGGS, 2008), o mais valioso para o cliente corporativo e que tem o maior potencial de valor agregado da instituição (WILD; GRIGGS, 2008).

¹ “Many argue that tacit knowledge is information processed by the human brain and can therefore reside only in the minds of individuals.”

² “Explicit knowledge can be defined as ‘to know about’ whereas tacit knowledge is ‘to know how’”.

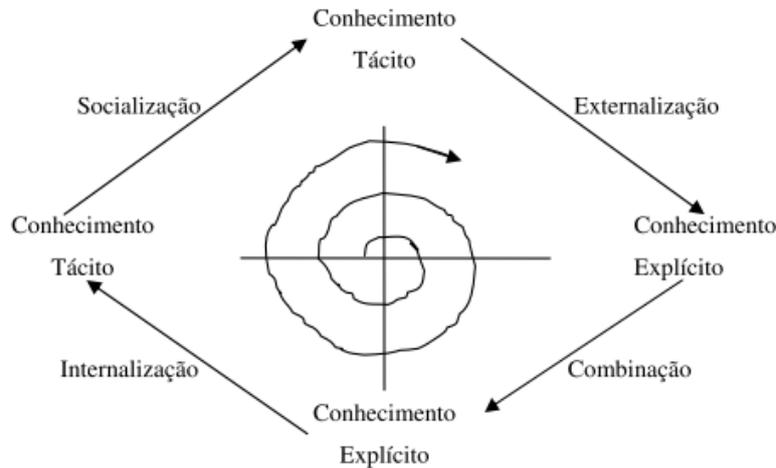
Criação de Conhecimento

O conceito de criação de conhecimento se relaciona com a geração de soluções e conceitos novos (que é um processo constante), e normalmente não pode ser planejado ou sistematizado, levando em conta que é um evento pessoal, assim como o pensamento, mas alguns pesquisadores propõem que criatividade pode sim, ser ensinada e aprendida (WILD; GRIGGS, 2008). Mas não basta ele somente ser criado. É preciso que seja capturado, relacionado com o conhecimento existente, e armazenado da maneira mais útil para a organização (WILD; GRIGGS, 2008).

Buscando entender como estes saberes circulam, Okumus (2013, p. 6) cita o trabalho seminal de Nonaka e Takeuchi (1997), que propuseram um modelo em quatro fases para a criação de conhecimento, definidos como: socialização, externalização, combinação e internalização. Socialização é, por exemplo, o método de discussão informal, levando em conta a troca de experiências dentro de um grupo, tentando chegar à troca de conhecimento tácito entre os membros.

Externalização leva em conta o uso de modelos, exemplos, desenhos ou analogias para articular o conhecimento tácito, para depois combiná-los, convertendo-os em um saber explícito (OKUMUS, 2013). Este modelo é então compartilhado via documentação física ou digital, utilizando os métodos mais comuns na organização. A internalização é o caminho inverso: o de conversão de conhecimento explícito em tácito, na mente e nas práticas do agente final deste conhecimento (OKUMUS, 2013). A figura 3 ilustra a visão de Nonaka e Takeuchi sobre a criação do conhecimento.

Figura 3 - Os quatro modos de conversão do conhecimento e a espiral do conhecimento



Fonte: Spiller (2005)

Este conceito de classificação de conhecimento é semelhante à conhecida como “Estratégia de Codificação”, citada no método desenvolvido por Hansen *et al.* (1999). Ambos visam tornar o saber acessível para toda a organização e, enquanto habilita o uso econômico da informação, evita que seus ativos tenham que ser “reinventados”, diminui custos e permite seu uso final com máxima eficiência (VENKITACHALAM; AMBROSINI, 2017).

Estratégias de Gestão de Conhecimento

Venkitachalam e Ambrosini (2017, p. 3) argumentam que a estratégia de GC que a empresa escolher deve se basear em sua estratégia competitiva, e em como o conhecimento existente nela, atualmente, cria valor para seus serviços e produtos, levando em conta também as expectativas dos clientes sobre eles.

É preciso entender primeiro as falhas relacionadas ao manejo de conhecimento na organização, se existirem, procurando apreender suas reais necessidades e qual o *know-how*³ atual, para depois implementar uma estratégia específica (VENKITACHALAM; AMBROSINI, 2017).

³ Habilidades adquiridas pelo saber prático de uma pessoa. Engloba também sua experiência em uma determinada área

Memória Organizacional

Interessantemente, os autores encontrados na pesquisa não discorrem especificamente sobre a Memória Organizacional, que é a concretização da Gestão do Conhecimento nas empresas (MORESI *et al.*, 2009), talvez pelo enfoque mais generalista que buscavam dar em seus estudos, com foco não específico na GC, mas como uma ferramenta de apoio à outra disciplina que gostariam de estudar. Por isso, aprofunda-se abaixo a definição e o papel da Memória Organizacional - que é um dos eixos deste trabalho - na GC.

Este termo é conhecido, em inglês, como “*corporate memory*” ou “*organizational memory*”, sendo sinônimos e traduzidos como, respectivamente, Memória Corporativa e Memória Organizacional, mas podendo também ser denominado de base de conhecimento empresarial, memória de grupo, mente de grupo, memória coletiva, inteligência coletiva, inteligência corporativa, base de conhecimento compartilhado, repositório corporativo, banco de dados de *know-how*, todos entendidos como tendo o mesmo significado (MOLINA; VALENTIM, 2011).

A Memória Organizacional (MO) é um dos pontos-chave dentro da Gestão do Conhecimento, se apresentando como um método que permite que aqueles que compõem a instituição criem e mantenham um repositório de seu entendimento e memórias sobre ela, compartilhando informações e conhecimento, de maneira que possam ser lembrados caso necessário e assim evitar que se percam. Esse acervo é uma ferramenta indispensável para o gerenciamento das atividades e do conhecimento dos indivíduos ou do grupo (FREIRE *et al.*, 2012).

Ainda assim, mesmo que se utilizem sistemas informáticos e dinâmicas voltadas à aprendizagem organizacional, muitas empresas não gerenciam seu conhecimento de maneira a evitar erros do passado, e com isso perdem a oportunidade de, com sua experiência prévia e registros de tentativas e acertos, acumular experiência através de sua história (SPILLER, 2005).

A MO estende o conhecimento, por organizar aquele criado pelos trabalhadores da organização, de maneira que se torne explícito, durável e de fácil acesso, melhorando o compartilhamento e reuso, com o foco principal de aumentar a competitividade da empresa/instituição (FREIRE *et al.*, 2012; MOLINA; VALENTIM, 2011; MORESI *et al.*, 2009). As principais questões que se façam sobre a instituição, seu ambiente, processos e produtos, ela, a memória organizacional, deve ser capaz de responder (MOLINA; VALENTIM, 2011).

A Tecnologia da Informação

A TI pode ser definida como a totalidade de ativos investidos em tecnologia ligada à gerência de informação, que podem ser *hardware* (equipamentos de processamento, armazenamento ou transmissão de dados), *software* (programas, aplicativos ou sistemas operacionais) ou recursos humanos (CHUNG *et al.*, 2010).

Reconhece-se hoje que os ativos tangíveis têm menos influência para as empresas, em relação ao desempenho dos negócios e à competitividade. A TI e as pessoas, com seus conhecimentos, são os ativos mais importantes da organização, com a TI como peça chave que modifica todos os aspectos dos negócios na era digital (TURULJA; BAJGORIĆ, 2017).

A utilização da Tecnologia de Informação em uma organização tem como um de seus efeitos benéficos o estímulo à sua capacidade de inovar, porque influencia positivamente na velocidade de adoção de novas técnicas e processos (MOHAMAD; RAMAYAH; LO, 2017). Empresas que tem uma política de TI bem definida são capazes de se infiltrar mais facilmente em novos segmentos de mercado, facilitar o contato com novos fornecedores e criar uma relação mais próxima com seus clientes (MOHAMAD; RAMAYAH; LO, 2017).

Há, claramente, uma relação positiva entre os recursos de TI e a inovação nas empresas (MOHAMAD; RAMAYAH; LO, 2017; TSENG, 2008). As pesquisas em sistemas de informação reconhecem sua força na geração de ativos intangíveis, dentre outros benefícios, afetando também as operações de baixo nível, otimizando a cadeia de produção, reduzindo custos, facilitando processos e elevando os níveis de performance na empresa (CHUANG; LIAO; LIN, 2013; KUO; YE, 2010; MAO *et al.*, 2016; TURULJA; BAJGORIĆ, 2017).

A TI demonstra sua real dimensão através da relação de integração que mantém com outros recursos e ativos organizacionais (LIN *et al.*, 2008). Ela se mescla à diferentes atividades dentro de cadeia produtiva, incluindo o planejamento de processos e o suporte à logística, produção e operações, como também às áreas de serviços, vendas e *marketing* (CHUANG; LIAO; LIN, 2013). De acordo com Mohamad *et al* (2017, p. 8), “os recursos de TI são um parte importante no aumento de inovação em uma empresa e mais ainda na indústria” (tradução nossa)⁴.

⁴ “*Information technology capability is an important constituent that enhances firm innovativeness and more so within an industry*”

Lin *et al.* (2008) e Kuo e Ye (2010) citam o trabalho de Bharadwaj (2000) como o modelo de referência na classificação de recursos de TI, que os distribuiu em (1) recursos tangíveis, tais como os componentes físicos da infraestrutura de TI (*hardware*), (2) recursos humanos ligados à TI, englobando tanto as habilidades técnicas quanto as de gerenciamento e (3) os recursos intangíveis viabilizados pela TI, como por exemplo, ativos de conhecimento.

Em relação à sua estratégia de uso, Venkitachalam e Ambrosini (2017) usam a classificação proposta por Weill, Subramani e Broadbent (2002), que utilizam quatro visões estratégicas para definir como investir em TI, chamados de estratégias “Inexistente”, “Utilitária”, “Dependente” e “Habilitadora”. Como segue:

- Inexistente: acontece normalmente em organizações onde a infraestrutura de TI (ITI) existe em unidades individuais, onde as decisões relacionadas à tecnologia são tomadas localmente. Por isso, acabam por não ter os benefícios que a ITI em escala pode prover, resultando em baixa eficiência e queda na sinergia entre suas unidades de negócios (VENKITACHALAM; AMBROSINI, 2017).

- Utilitária: utilizada nas organizações que desejam focar em redução de custos. É pensada como uma estratégia para ajudar a empresa a ter grande eficiência e ganhos econômicos, mas sem escopo de investimento em TI que fuja desta visão. São considerados investimentos para cobrir funções administrativas básicas somente (VENKITACHALAM; AMBROSINI, 2017).

- Dependente: busca ajudar a organização a construir uma ITI que se acople firmemente à sua estratégia de negócios, com as soluções desenvolvidas para codificar o máximo de informações geradas em sua estrutura, aumentando sua sinergia (VENKITACHALAM; AMBROSINI, 2017).

- Habilitadora: considera a TI como um recurso chave de sua estrutura organizacional, de maneira que possa prover opções estratégicas para o futuro, em vistas de abarcar mudanças táticas da organização. Deve ser entendida pela chefia como um recurso estratégico, que fornece vantagem competitiva. Vai além de codificar as informações, habilitando a comunicação máxima entre as estruturas da empresa, com uma infraestrutura de troca ampla, interligada e modernizada (VENKITACHALAM; AMBROSINI, 2017).

Para que a TI seja utilizada de maneira correta e possa fornecer o máximo de utilidade, em relação ao investimento necessário, é preciso definir qual a melhor estratégia para a sua implantação. Estudar a estratégia de uso da TI em uma organização depende do seu contexto,

que é único, em relação as outras (HUANG *et al.*, 2011; VENKITACHALAM; AMBROSINI, 2017).

Os Desafios no Uso da Tecnologia da Informação

A implantação e uso da TI de maneira eficaz nas organizações enfrenta vários desafios. A menos que a TI seja incorporada profundamente nos processos estratégicos da empresa e combinada com seus recursos de gerência, as capacidades que ela oferece não podem de maneira automática, *per se* somente, produzir efeitos positivos na performance da empresa (LIN *et al.*, 2008). A maneira como uma empresa aloca seus recursos de TI é crítica (MAO *et al.*, 2016).

Introduzir o uso e aproveitar das funções primordiais da TI, de maneira efetiva, pede o envolvimento de vários atores da organização, como seus gerentes, clientes, administradores de TI e de projetos, consultores e usuários finais, todos trabalhando cooperativamente para resolver os problemas e aproveitar as oportunidades que ela oferece (KUO; YE, 2010). Um alto grau de comprometimento com o uso dos recursos de TI reflete a convicção e o entendimento da empresa em relação à sua contribuição para a organização; por outro lado, quando este comprometimento falha, a própria relação com o gerenciamento de conhecimento fica debilitada. Ainda assim, sabe-se que mais e maiores recursos de TI somente não levam a um maior poder de controle do saber (MAO *et al.*, 2016).

Alguns estudos falharam em comprovar uma conexão positiva e significativa entre o uso de recursos de TI e a performance organizacional. Os autores argumentam que empresas com excelentes parques de TI podem erodir suas vantagens competitivas pela contínua comoditização (supervalorização) da TI (MAO *et al.*, 2016). Grandes investimentos nesta área aumentam posteriormente o custo mais do que melhoram as vantagens competitivas, comparativamente, e não necessariamente de maneira proporcional. Como a TI está cada vez mais homogênea e omnipresente nas empresas e organizações, copiar e imitar as capacidades de um competidor está mais fácil. As relações entre vantagem competitiva e conhecimentos acumulados em TI, na literatura, ainda são contraditórios (MAO *et al.*, 2016).

Nota-se também que a aceitação do uso de novas tecnologias está diretamente ligada com a noção de que esta propiciará uma diminuição nos esforços físicos e mentais do usuário, sendo influenciada, de maneira semelhante, pela cultura local do grupo em que se tenta se infiltrar. O entendimento atual de como isto se processa atualmente ainda é limitado. O corpo

atual de literatura trata de modelos de aceitação da tecnologia, mas ainda com pouco foco na influência da cultura neste fenômeno (LIN, 2014).

A Relação entre a Gestão do Conhecimento e a Tecnologia da Informação

A literatura no campo de gestão demonstra um esforço em definir o que é a GC, do que trata e de como deveria ser nas empresas a dinâmica de criação, transferência e uso do conhecimento, e como a TI pode ajudá-las a facilitar o processo de GC na criação de recursos dinâmicos (OKUMUS, 2013). É muito importante a alocação de recursos de TI para o processo de aquisição, uso e transferência do conhecimento (MAO *et al.*, 2016), visto que a TI é considerada como a grande habilitadora da GC (HUANG *et al.*, 2011; KUO; YE, 2010; LIN *et al.*, 2008; OKUMUS, 2013; PILLANIA, 2008; REVILLA; RODRÍGUEZ-PRADO; PRIETO, 2009). Dessa maneira, ela, a TI, acaba por influenciar direta e indiretamente os processos de geração e uso de conhecimento, suas fontes e também fatores ambientais da empresa (OKUMUS, 2013).

Por outro lado, assim como a TI facilita a GC em sua necessidade de transformar dados em conhecimento com valor agregado, a GC também age e influencia a TI, porque é preciso que aja conhecimento de que tratar, e, antes que os dados possam ser coletados, efetivamente, ela, a GC, determina quais dados serão (LIN *et al.*, 2008).

Escolher as estratégias corretas que correlacionarão GC e TI é crítico para a organização (HUANG *et al.*, 2011). A estratégia de GC chamada HNT, iniciais de Hansen, Nohria e Tierney (1999) e citado por Venkitachalam e Ambrosini (2017), oferece pistas de como a organização pode implementar a GC com base na infraestrutura de TI disponível, para habilitar o uso do conhecimento.

Esta técnica recomenda três pontos a serem levados em conta pela organização antes de implementar uma estratégia de GC em um ambiente com infraestrutura de TI disponível, que são (1) se a empresa normalmente utiliza conhecimento explícito ou tácito para resolver as necessidades dos clientes, (2) se o seu foco é em produtos maduros ou em inovações e (3) se ela oferece produtos ou serviços baseados em um modelo ou personalizados (VENKITACHALAM; AMBROSINI, 2017).

Entende-se que a gerência estratégica da TI é uma facilitadora no desempenho dos negócios e, quando aliada a certos aspectos da GC, ajuda na sobrevivência da instituição no ambiente altamente competitivo atual (HUANG *et al.*, 2011). O uso de soluções de TI para

atender à estratégia de GC contribui para a criação de bolsões de conhecimento corporativo através de seu mapeamento ou pela criação de redes de conhecimento (HUANG *et al.*, 2011).

Os Desafios na Relação entre a GC e a TI

Okumus (2013), Mao *et al.* (2016) e Maroofi (2016) discutem que as aplicações de TI não têm a capacidade de promover a criação, nem de facilitar o uso de conhecimento de maneira automática. Empresas e entidades manejam conhecimentos tácitos e explícitos e eles podem ser trabalhados pela TI, mas a captura do conhecimento tácito, normalmente oferecido pelas pessoas, é, algumas vezes, difícil. Em muitos ambientes, mais complexos ou dinâmicos, onde mudanças acontecem de maneira constante, as ferramentas de TI podem não ser suficientes, sozinhas, para encontrar, acumular e compartilhar o saber (OKUMUS, 2013).

A aplicabilidade da TI pode ser bastante limitada na tentativa de gerenciar informações, algumas vezes. Motores de pesquisa, também chamados de ferramentas de busca, podem produzir quantidades muito grandes de informações, algumas irrelevantes, e torná-las úteis ou ordená-las requer esforço, tempo, recursos humanos, computacionais ou estudos específicos para automatizar seu aproveitamento (OKUMUS, 2013). O uso de GC e TI não significa também, obrigatoriamente, a substituição do elemento humano nesta equação. As pessoas continuam sendo essenciais na gestão do saber nas organizações, preferencialmente organizadas em comunidades dispostas à compartilhar o conhecimento (MAO *et al.*, 2016; OKUMUS, 2013; TSENG, 2008).

Por outro lado, estas ferramentas exigem proficiência dos usuários, e, sem ela, podem ser incapazes de encontrar ou até mesmo utilizar as informações que procuram. A necessidade constante de treinamento é primordial, para que o interessado aproveite do conhecimento e não perca a confiança em sua capacidade de tomar decisões, por não ser capaz de transformar a informação em conhecimento útil, se afastando da ferramenta de TI e da GC por consequência (KUO; YE, 2010).

3.4 CONSIDERAÇÕES FINAIS DA REVISÃO DE LITERATURA

Este ensaio teve como foco identificar o estado atual dos estudos que relacionam a Gestão do Conhecimento e a Tecnologia da Informação. Foram encontrados e utilizados 22 artigos que cumpriram os critérios de inclusão.

Todos os autores reconhecem que o conhecimento é atualmente um ativo de alto valor para as organizações, é a fonte primária na sua vantagem competitiva e na geração de valor para a empresa, que o seu capital é baseado, na maioria das vezes, no que ela sabe, e não nos bens materiais que possui e que ele está diretamente ligado às pessoas e grupos, precisando de meios para circular e ser melhor aproveitado.

Salientou-se que Gestão do Conhecimento sempre existiu, relacionada à humanidade e suas práticas de subsistência, sem necessariamente ser reconhecida por este nome ou pelo valor que oferece, e que hoje é uma disciplina relativamente nova, que ainda carece de melhor definição acadêmica e que está em constante evolução. E, por gerar valor a partir dos recursos intangíveis das organizações, através da gerência coordenada do seu conjunto de conhecimentos, para melhor distribuí-lo, evitando retrabalho e melhorando sua aplicação, mostra ter efeito no desempenho e na capacidade de inovar das empresas. Várias técnicas definem qual a melhor estratégia para a implantação de uma política de GC, sempre levando em conta a visão competitiva da organização, qual a relação do conhecimento que possui com os serviços ou produtos que fornece e com os clientes.

A TI é colocada, pela maioria dos autores, como sendo de grande importância, influenciando o desempenho, competitividade, planejamento, inovação e todos os aspectos dos negócios. Quando possui uma estratégia de uso bem definida e é estratégica para a organização, vai além de classificar as informações, permitindo a comunicação total dentro de sua estrutura. Precisa ultrapassar diversos desafios em seu uso, como a necessidade de ser incorporada fortemente aos processos estratégicos e planejamentos gerenciais, pedindo comprometimento com seu uso. Não é, por si só, geradora de efeitos positivos e lucros, se não receber investimentos corretos e for difundida através de treinamentos e mudanças de paradigma. Estas relações são comprovadas por estudos na área de aceitação de tecnologia, mas carecem ainda de maior profundidade nas pesquisas.

Demonstrou-se que a TI é considerada, amplamente, como a real habilitadora da GC, com influência direta e positiva no manejo de conhecimento; é de grande valia ao ajudar na sobrevivência de uma organização em um ambiente altamente competitivo, como o atual. Ainda assim, é preciso escolher a estratégia correta para utilizar as duas, não havendo ainda uma solução genérica para as organizações.

Existem poucos estudos que tratem da relação entre GC e TI somente, explorando suas implicações. A maioria dos autores busca explicar e entender a sua influência em outras matérias, sempre como coadjuvantes em temas que consideram como relevantes e, desta

maneira, se mostra pobre a visão interdisciplinar nesta área, ainda mais quando leva-se em conta, atualmente, a emergência de realidades como o *big data*, disciplinas de manipulação de dados como *Data Science* (ciência de dados), *machine learning* (aprendizado de máquina) e a inteligência artificial.

Esta revisão bibliográfica buscou demonstrar o papel fundamental da Gestão do Conhecimento e da Tecnologia da Informação nas instituições modernas, sua importância na geração de valor por meio de ativos intangíveis e como a TI está intimamente ligada à GC, habilitando seu uso de maneira ampla e funcional, através do uso da Memória Organizacional. Esta pesquisa mostra que ainda há espaço para reflexões e aprofundamentos a serem desenvolvidos nesta área, principalmente em vista de novas tecnologias de manipulação de dados que se apresentam e pela moderna norma ISO, proposta recentemente e que destina a padronizar a Gestão do Conhecimento.

4 METODOLOGIA

4.1 CARACTERIZAÇÃO DO ESTUDO

Seguindo as orientações do Programa de Pós-Graduação, a visão interdisciplinar deve ser parte imprescindível das pesquisas e as análises delas decorrentes, devem levar em conta este processo em seu desenvolvimento. A seguir, desenvolveu-se as conexões e influências nos assuntos propostos neste trabalho.

DEFINIÇÕES

Isto posto, definiu-se que a pesquisa é, quanto aos fins da investigação, do tipo exploratória, descritiva e aplicada. A análise exploratória, segundo Vergara (2009), é utilizada quando há pouco conhecimento acumulado na área em que se busca estudar, podendo através dela emergir as hipóteses relacionadas ao fenômeno em foco, ajudando o pesquisador à ter uma aproximação com o assunto, sendo empregada nas fases iniciais de uma investigação e na reformulação do problema (VENTURA, 2007).

A modalidade de pesquisa descritiva tem o foco em descrever como um evento ocorre, tentando criar relações entre os seus componentes e retratando sua estrutura, mas sem o compromisso de explicá-lo, apesar de poder servir de base para tal (VERGARA, 2009), exigindo do pesquisador aprofundamento sobre o que deseja observar. O estudo de caso é um exemplo de pesquisa descritiva, que aqui será utilizado (SILVEIRA; CÓRDOVA, 2009).

Em relação a ser também do tipo aplicada, justifica-se por ter como objetivo, além da exploração e descrição, o foco em um problema concreto, real, tendo portanto, finalidade prática, se afastando da pesquisa exclusivamente expositiva, conjectural (VERGARA, 2009).

Quanto aos meios que utilizou, se encaixa na categoria documental e estudo de caso, para que, através destas técnicas, se consiga levantar, tratar e analisar as fontes informacionais desejadas, pelo potencial elucidativo destes e relevância para o estudo em questão (GARCIA JÚNIOR; MEDEIROS; AUGUSTA, 2017). A pesquisa documental emerge, no enfoque qualitativo, que será aprofundado logo a seguir, como uma proposta que traz subsídios importantes no estudo de alguns temas, já que os documentos compõem uma fonte rica de dados para a pesquisa (GODOY, 1995), sendo considerado como a sua variante mais antiga (GÜNTER, 2006).

O estudo de caso é delineado como a coleta, análise de dados e a observação de um comportamento em seu ambiente natural, sobre um caso ou modelo individual, buscando entender um fenômeno mais amplo (GÜNTER, 2006).

E, finalmente, quanto à análise dos dados, é de natureza qualitativa. Godoy (1995, p. 2) afirma, neste mister, que a perspectiva que define um estudo qualitativo é que “um fenômeno pode ser melhor compreendido no contexto em que ocorre e do qual é parte, devendo ser analisado numa perspectiva integrada, partindo de questões amplas que vão se aclarando no decorrer da investigação”, não se preocupando com representatividade numérica, mas, sim, com o aprofundamento da compreensão de uma comunidade ou organização, por exemplo (SILVEIRA; CÓRDOVA, 2009), nos levando a acreditar que esta abordagem é a mais indicada para este caso, tendo em conta as questões em aberto que surgem no estado atual do problema e das proposições que se pretende fazer para sua resolução.

A INTERDISCIPLINARIDADE NA GESTÃO DO CONHECIMENTO, TECNOLOGIA DA INFORMAÇÃO, SAÚDE ANIMAL, HUMANA E O AMBIENTE

A interdisciplinaridade, segundo Alvarenga *et al.* (2011), é um processo que trabalha nas divisas entre os pensamentos disciplinares, buscando a conexão entre seus conteúdos, sempre com o foco em tentar explicar e entender eventos complexos e de diferentes naturezas.

Ela ocorre realmente quando se amplia esta relação, buscando além da simples comunicação de ideias entre os interlocutores de cada disciplina, para uma integração de conceitos base e significados, metodologias ou procedimentos. Um grupo interdisciplinar se compõe, por isso, de pessoas com formação em diferentes áreas de estudos e que carregam consigo um conjunto de conhecimentos, cada qual em sua área (ALVARENGA *et al.*, 2011), fazendo, desta maneira, que a interdisciplinaridade seja um recurso de interconexão de saberes e práticas (LEFF, 2011).

Seguindo este raciocínio, este capítulo propõe destacar o valor que existe na interconexão entre a Gestão do Conhecimento e sua capacidade em gerir o saber, a TI, com as ferramentas que tem à disposição e as ciências que tratam da saúde animal, humana e do meio-ambiente.

No momento, as pesquisas que utilizam a ciência e a gestão de dados na área de saúde animal são diversificadas, com destaque no monitoramento da expansão de doenças em

territórios nacionais. Através de buscas em algumas bases de dados (Periódicos Capes, Scopus e PubMed), com os termos “*animal electronic health records*”, “*animal health data science*”, “*animal health big data*”, as expressões “*animal health surveillance*” e “*animal health system*” se apresentaram como as mais utilizadas, sobressaindo os trabalhos que tratam das tentativas de utilizar sistemas eletrônicos para gestão de registros de saúde animal.

Está claro para os autores o papel de protagonismo da fiscalização/vigilância (*surveillance*) sobre as doenças relacionadas à animais através de ferramentas informatizadas de coleta, armazenamento, análise e visualização de dados, incluindo a parte importante que a comunicação tem neste setor (BEREZOWSKI; LINDBERG; WARD, 2015); além disso, destacam que a vigilância e o monitoramento são dois componentes em um sistema nacional de gestão epidemiológica ideal, na tentativa de proteger as populações animais locais assim como para criar e preservar mercados de exportação (HUESTON, 1993).

Um exemplo de monitoramento informatizado é o Sistema Mundial de Informações de Saúde Animal (*World Animal Health Information System - WHAIS*), disponibilizado em 2005 pela Organização Mundial para a Saúde Animal (*World Organization for Animal Health - OIE*) e relançado e atualizado em 2017. A proposta é que o WHAIS seja um sistema online focado no processamento de dados de doenças em animais em tempo real, para os países membros e alguns não membros, sendo acessível pela comunidade internacional através de uma interface *web*. (ÉLOIT, 2017; OIE-WAHIS, 2019). As informações compreendem dados geoespaciais, informações genômicas ligados à epidemias, alertas via aplicações móveis e informações em estado bruto para o uso em programas de análise de dados (OIE-WAHIS, 2019).

A Organização Mundial para a Saúde Animal mantém, além disso, uma ligação com a Organização Mundial da Saúde (*World Health Organization - WHO*), através de instrumentos legais, que exige a troca de relatórios sobre certos eventos ligados a doenças e surtos, de interesse para a saúde pública, para as suas respectivas organizações. Países membros da OIE devem avisar sobre eventos epidemiológicos excepcionais envolvendo doenças listadas pela OIE, incluindo as zoonoses (CÁCERES *et al.*, 2017).

Da mesma maneira, na área de saúde humana, animal e sua relação com o meio-ambiente, há um exemplo de esforço neste sentido, que é o modelo chamado de “*One Health*” (Saúde Única), em que a visão do bem-estar se afasta do modelo hierárquico inicial, antropocêntrico, no qual só existe a saúde humana, para incluir os sistemas animais e do meio-ambiente, todos interconectados (RABINOWITZ *et al.*, 2018). Nele, entende-se que há uma interface entre as relações e doenças que afetam os seres humanos e animais e o estado do meio-

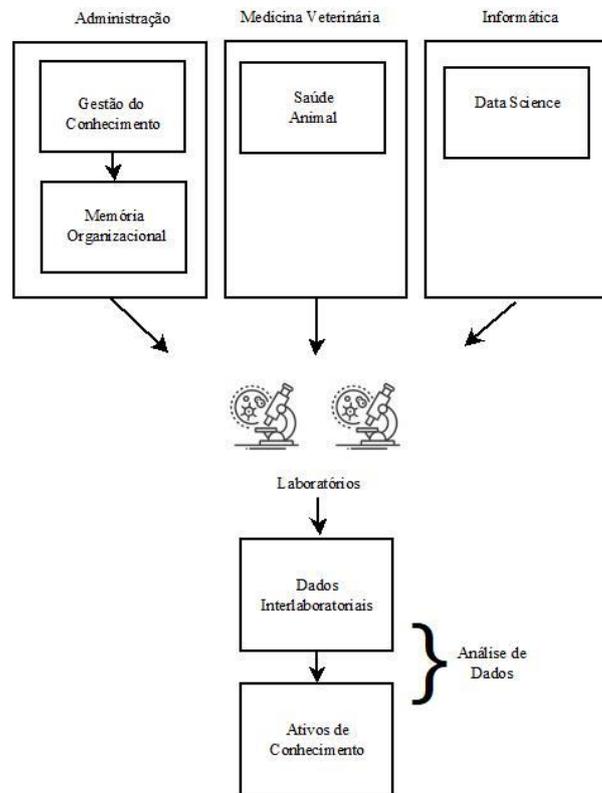
ambiente, e que somente uma abordagem interdisciplinar pode equilibrá-las, buscando uma coexistência saudável (RABINOWITZ *et al.*, 2018).

O reconhecimento da complexidade destas interações reafirma a necessidade da participação de pesquisadores de diversas áreas no esforço de responder as questões de saúde da tríade humano-animal-ambiente. Veterinários, epidemiologistas, cientistas sociais, ecologistas, economistas, médicos e políticos são alguns exemplos (MARDONES *et al.*, 2017).

Com a sociedade da Informação em pleno curso, atualmente, onde toda a realidade é medida, quantificada e ponderada, sejam nossos dados e estado pessoais ou da atmosfera, climatológicos, georreferenciais, da natureza que consumimos e aproveitamos, é preciso de alguma maneira gerir e entender estas informações, de tantas fontes e origens distintas, e buscar a correlação entre todas, já que se entende que há, inegavelmente, uma conjuntura que às interliga. Como afirma Morin (2014), toda e qualquer informação tem apenas um sentido em relação à uma situação, a um contexto.

Neste sentido, a Gestão do Conhecimento, com seu *know-how* no manejo de informações e a TI, considerada o seu grande instrumentador, emergem como um expediente técnico, trazendo, com experiência na gerência de grandes quantidades de dados a visão macro e a análise que se precisa para o entendimento da nossa complexa realidade socioambiental moderna. Esta relação é ilustrada na figura 4.

Figura 4 - Relação interdisciplinar entre as áreas de conhecimento



Fonte: o autor (2019)

Para isso são necessários cientistas e profissionais das mais diversas áreas, em apoio mútuo, para se compreender como essas relações se criam e mantêm. Nossa ciência tende a separar para entender e medir, e com olhar macro fornecido pela interrelação dos setores citados acima, pode-se alcançar novos entendimentos.

4.1 LOCAL DE ESTUDO

O local de estudo deste trabalho foi o campus III, da UDESC, a Universidade do Estado de Santa Catarina, conhecido como Centro de Ciências Agroveterinárias (CAV), em Lages, SC.

Lages é um município localizado no interior do estado de Santa Catarina, com aproximadamente 158 mil habitantes, conforme dados do IBGE de agosto de 2014, sendo parte da região serrana do estado. É o maior de Santa Catarina, em extensão territorial, com uma economia fortemente amparada pela agropecuária, agricultura, turismo rural e ênfase na indústria de papel e celulose e na indústria madeireira (PREFEITURA DE LAGES, 2018).

A UDESC, fundada em 1965, foi inicialmente conhecida como Universidade para o Desenvolvimento do Estado de Santa Catarina, tendo seu reconhecimento pelo MEC como universidade em 1990 quando do seu desligamento da então Fundação Educacional de Santa Catarina (FESC), alcançando sua autonomia pela criação da Fundação Universidade do Estado de Santa Catarina (UDESC, 2018).

É uma universidade gratuita, organizada no formato multicampi. Está presente em 9 cidades do estado, distribuído em 12 centros de ensino, oferecendo, no momento, 59 cursos de graduação e 49 de mestrado e doutorado (UDESC, 2018)

O Centro de Ciências Agroveterinárias, CAV, foi criado em 1973 e possui atualmente 4 cursos de graduação (Agronomia, Medicina Veterinária, Engenharia Ambiental e Sanitária e Engenharia Florestal) e 10 de pós-graduação, entre mestrados (Bioquímica e Biologia Molecular, Ciência Animal, Ciência do Solo, Ciências Ambientais, Engenharia Florestal e Produção Vegetal) e doutorados (Bioquímica e Biologia Molecular, Ciência Animal, Ciência do Solo e Produção Vegetal). Ocupa uma área de 79 hectares com um Hospital de Clínica Veterinária, prédios administrativos e voltados ao ensino, além de 35 laboratórios (UDESC, 2018)

4.2 PARTICIPANTES DA PESQUISA

A população alvo deste estudo foi formada pelos laboratórios de diagnósticos veterinários do Centro de Ciências Agroveterinárias (CAV), selecionados por fornecem serviços ao público interno e externo ao campus, tanto pessoas físicas quanto jurídicas, e assim tratarem com amostras originárias da região e do Hospital de Clínica Veterinária (HCV), instalado no campus.

Os laboratórios são os seguintes:

Centro de Diagnóstico Microbiológico Animal (CEDIMA)

O CEDIMA, Centro de Diagnóstico Microbiológico Animal, fundado em março de 2009 é o resultado da integração das estruturas do laboratório de Doenças Infectocontagiosas com o de Imunologia e Microbiologia, e atualmente possui as divisões de Bacteriologia e Virologia (CAV, 2018).

O CEDIMA conta com uma área construída de aproximadamente 400 m², distribuídos em um laboratório para aulas práticas e cursos (capacidade de 20 alunos),

quatro laboratórios para pesquisa, cinco salas para professores, sala para alunos de graduação e pós-graduação, sala de reuniões com capacidade para 20 pessoas, duas salas de lavagem de material e almoxarifado (CAV, 2018).

Parasitologia e Doenças

É um laboratório que fornece ao público vários serviços, incluindo cultivo e identificação de larvas de 3º estágio, pesquisa de larvas, pesquisa de ovos em urina, identificação de endo e ectoparasitas, pesquisa de hemoparasitas, pesquisa de *Giardia spp.*, oocistos, teste carrapaticida, sorologia para toxoplasmose e neosporose e orientações técnicas aos clientes (CAV, 2018).

Patologia Animal (LAPA)

O Laboratório de Patologia Animal, fundado em 1980 atualmente atende a comunidade catarinense e de estados vizinhos e está ligado ao Mestrado e Doutorado em Ciência Animal, com seus orientados.

Presta serviço de extensão fazendo o diagnóstico de doenças em várias espécies de animais, através de análise histológica (CAV, 2018).

Patologia Clínica Veterinária (localizado no Hospital de Clínica Veterinária)

Trabalha em apoio à disciplina de Análise Clínica e às atividades diárias do Hospital de Clínica Veterinária, oferecendo exames de complemento ao diagnóstico, como hemograma, urinálise, análise de líquidos corporais, citologia aspirativa, pesquisa de hemoparasitas em esfregaços sanguíneos, de mioglobina na urina e etc. (CAV, 2018).

Hospital de Clínica Veterinária (HCV)

Fundado em 1975, com o objetivo de complementar a formação dos acadêmicos do curso de Medicina Veterinária. Embora destinado ao ensino, pesquisa e extensão, é o grande prestador de serviços à comunidade por meio das áreas de Clínica Médica e Cirúrgica de Pequenos e Grandes Animais, Fisiatria, Anestesiologia, Cardiologia, Acupuntura, Medicina de Animais Silvestres e Exóticos, Diagnóstico por Imagem e Patologia Clínica. Os serviços veterinários promovem a integração da comunidade com a universidade e tem como finalidade garantir o atendimento de qualidade e acessível à população promovendo a sanidade e bem-estar animal, bem como a colaboração na saúde pública por meio da profilaxia de zoonoses e qualidade dos produtos de origem animal (CAV, 2018).

A análise da estrutura de gerenciamento de informações utilizada no HCV é importante pela relação próxima que mantêm os laboratórios do campus, pois, em vários momentos, é utilizador de seus serviços e pode ser beneficiar de uma maior conectividade com eles.

4.4 CRITÉRIOS DE INCLUSÃO E EXCLUSÃO DOS PARTICIPANTES

Critérios de Inclusão

Os critérios de inclusão utilizados para os participantes no estudo, foram:

- a) Laboratórios de diagnósticos de saúde animal, pertencentes ao CAV, que voluntariamente, através da direção geral ou de seus coordenadores, permitiram participar.
- b) Laboratórios pertencentes ao CAV que trabalham em áreas cujos resultados podem, ao ser associados, gerar novos conhecimentos ou valores derivados desta correlação.

Critérios de Exclusão

Os critérios exclusão adotados para os participantes no estudo, foram:

- a) Laboratórios de diagnósticos de saúde animal, pertencentes ao CAV que, através da direção geral ou de seus coordenadores, não permitiram em participar.
- b) Laboratórios pertencentes ao CAV que trabalham em áreas cujos resultados não podem ser associados a outros, por serem muito distintos ou não correlatos, e que não gerariam novos conhecimentos.
- c) Laboratórios que não trabalham em áreas relacionadas à saúde animal.

4.5 CRITÉRIOS DE CONFIDENCIALIDADE/ PRIVACIDADE

Foram asseguradas a confidencialidade e a privacidade de quaisquer pessoas ou entidades envolvidas no estudo, garantindo a não utilização de informações em prejuízo delas, ficando claro que em momento algum serão usados dados pessoais na pesquisa, somente a descrição dos dados e seus tipos e dos procedimentos para sua gerência.

5 COLETA E APRESENTAÇÃO DOS DADOS

De acordo com o que foi previsto nos objetivos, os laboratórios foram visitados para a análise de suas metodologias de tratamento dos dados de clientes, pacientes e de resultados dos exames. As visitas aconteceram na semana de 17 a 21 de dezembro de 2018, nos horários previamente marcados, acompanhados pelos responsáveis dos laboratórios ou seus representantes. Nestas ocasiões, foi permitido acesso aos documentos e/ou programas que utilizam internamente.

O procedimento seguido foi o de análise documental de quais são e como são coletadas, gerenciadas e armazenadas as informações que concernem às amostras levadas para análise, tais como: dados pessoais do interessado (nome, endereço, finalidade do exame e afins), dados intermediários e parciais de resultados de exames e os finais, que normalmente são arquivados e/ou devolvidos aos interessados.

Nos casos em que são mantidos em formulários de papel, foram solicitadas cópias físicas de seus modelos, sem dados pessoais de nenhum cliente, para estudo; naqueles em que o laboratório utilizar um *software* para o controle destes dados, foi solicitado acesso aos formulários digitais do mesmo e à modelagem do banco de dados usado, excluindo-se da análise quaisquer dados reais constante neles.

Esta etapa tem a função igualmente de levar ao entendimento da cultura e dos procedimentos atuais de gerência de informações de cada laboratório, normalmente desenvolvidos localmente e segundo as necessidades diárias e práticas de suas atividades.

Segue logo abaixo o resumo dos campos que cada um possui, com as suas descrições e detalhamentos. Estão listados na ordem em que aparecem no documento. Alguns campos são autoexplicativos, enquanto outros constam com uma descrição que aprofunda o seu significado, se necessário.

5.5 CEDIMA

Este laboratório possui uma ficha padrão para a solicitação dos exames (quadro 3) e dois documentos diferentes para a devolutiva de resultado, variando caso seja relacionado à uma bactéria (quadro 4) ou a um vírus (quadro 5), como descritos a seguir.

O documento é preenchido em mídia digital, usando o programa *Microsoft Word*, mantendo-se uma cópia para cada solicitação e sendo impresso se solicitado.

Quadro 3 - Ficha cadastral e de solicitação de exame - Cedima

Número	Data
Espécie	Raça
Sexo	Idade
Proprietário	Telefone
Remetente	Endereço
Cidade	Cep
Telefone	Fax
CPF/CNPJ	E-mail
Médico veterinário	Telefone
Material enviado	Exames solicitados
Histórico	Suspeita clínica

Fonte: O autor (2019)

Campos em destaque:

Número: valor sequencial que identifica de maneira única o cadastro e a solicitação do exame.

Remetente: nome do responsável pelo envio do material a ser analisado. Normalmente um profissional da medicina veterinária.

Material enviado: consta a descrição do material recebido para análise. Por exemplo: *swab*⁵ de ferida, de osso, nasal, de órgão, urina ou leite, água (para uso alimentar ou de piscina) e alimentos contaminados.

Exames solicitados: especificação do exame a ser realizado no material. Exemplo: isolamento e identificação bacteriana, antibiograma, contagem bacteriana, pesquisa de Salmonella, pesquisa de Listeria, coliformes fecais totais e termotolerantes, isolamento viral, FeLV (vírus da leucemia felina), leucose bovina etc.

Histórico: provê uma breve descrição do estado do animal, tanto atual quando pregresso, que possa ajudar na descoberta do agente patológico. É de escrita livre e depende de um relato por parte do responsável pelo material e dos conhecimentos técnicos de quem preencher o documento.

⁵ Cotonete estéril usado em coleta de material para testes microbiológicos

Quadro 4 - Ficha de resultados de exame - Cedima - Bacteriologia

Cliente	Requisitante
Endereço	Cep
Cidade	Fone
Celular	Endereço eletrônico
CPF	Método
Espécie	Material
Chegada	Resultado
Finalização da análise	

Fonte: O autor (2019)

O documento com os resultados repete alguns dados recebidos na ficha de pedido, inserindo ao final o resultado do exame.

Campos em destaque:

Finalização da análise: data em que ocorreu o fechamento da análise.

Método: técnica de análise utilizada no processo.

Ao final do documento, consta a assinatura do veterinário responsável pelo laudo.

Quadro 5 - Ficha de resultados de exame - Cedima - Virologia

Resultado de exame () sorológico () antígeno	Agente
Registro	Entrada
Proprietário	Município
Veterinário	Fone
Espécie animal	Material
Resultado	

Fonte: o autor (2019)

Campos em destaque:

Agente: o nome do vírus identificado. Exemplo: “Vírus da Leucemia Felina - FeLV”

Resultado: descrição por extenso do teste realizado e do resultado encontrado. Exemplo: “A pesquisa de antígeno foi realizada pelo teste de Imunofluorescência Indireta (VMRD-AB1-FELV)”.

5.6 PARASITOLOGIA E DOENÇAS PARASITÁRIAS

No caso deste laboratório, o resultado do exame solicitado é dado na própria ficha cadastral, entregue como uma cópia para o solicitante, em mídia impressa ou preenchida manualmente. A ficha se inicia com um cadastro padrão de pessoa física/jurídica, e os campos seguintes se destinam à detalhes relacionados à investigação do agente patológico (quadro 6).

Quadro 6 - Ficha cadastral, de solicitação de exames e resultados - Parasitologia

Responsável	RG/CPF
Telefone	E-mail
Cidade	Espécie
Raça	Idade
Sexo (macho ou fêmea)	Histórico
Data da coleta	Quantidade de amostras
Exames	Resultado
Observação	

Fonte: o autor (2019)

Campos em destaque:

Exames: específica qual se quer realizar, servindo para direcionar qual técnica será empregada, visto que o laboratório pode detectar qualquer parasita, baseado na demanda do solicitante e na suspeita clínica.

Histórico: provê uma breve descrição do estado do animal, tanto atual quanto pregresso, que possa ajudar na descoberta do parasita.

Quantidade de amostras: descreve quantas e quais amostras foram entregues para análise, com a maioria das vezes sendo fezes, seguido por sangue/soro e por exemplares do parasita.

5.7 PATOLOGIA ANIMAL

Neste caso, são utilizados documentos não padronizados para o controle da chegada de material e dos dados do requisitante. As informações básicas são anotadas e depois consolidadas em um documento com os resultados dos exames (quadro 7).

Quadro 7 - Ficha de resultados de exames - Patologia

Número do exame	Endereço
Proprietário	CPF
Telefone	Empresa/Clínica
Remetente	Telefone/Fax
Endereço	Animal
E-mail	Ficha
Identificação	Raça
Espécie	Sexo
Idade	Plantonista
Data do recebimento	Necropsia
Histórico	Histologia
Material enviado	Diagnóstico

Fonte: O autor (2019)

Campos em destaque:

Número do exame: utiliza um valor numérico único, seguido do ano corrente

Necropsia: narrativa do estado do animal pós necropsia, contendo, por exemplo, idade, estado corporal, alterações visíveis dos órgãos perante análise etc. Exemplo: “Bovino, fêmea, 5 anos, mestiça, estado corporal bom. Exame externo: nada a declarar. Exame interno: aumento de volume da região lateral esquerda da fase. Subcutâneo: edema, acentuado, difuso, em região torácica. Cavidade abdominal com grande quantidade de líquido, amarelo (ascite). Fígado diminuído de tamanho, firme, padrão lobular, levemente evidenciado. Rúmen com edema acentuado, difuso, em submucosa e conteúdo ressecado. Omaso com conteúdo ressecado. Abomaso com edema acentuado, difuso, em submucosa”.

Histórico: descrição do contexto prévio do paciente, como orientação aos exames subsequentes. Exemplo: “Animal encontrado em área de banhado, caído. Medicado com Fluxina. Morreu no dia seguinte”.

Material enviado: órgãos ou tecidos utilizados no exame. Praticamente qualquer tecido orgânico. Exemplo: músculo, pele, sistema nervoso, rim, fígado, baço, coração etc.

Histologia: análise microscópica dos tecidos de interesse à investigação. Exemplo: “Fígado: megalocitose de hepatócitos, acentuada, difusa, com proliferação de epitélio biliar associado a fibrose periportal difusa e hemorragia centrolobular acentuada, difusa”.

Diagnóstico: detalhamento final do resultado do exame. Como, por exemplo, “Intoxicação por *Senecio brasiliensis* (Maria mole)”

Os campos “histórico”, “necropsia”, “material enviado”, “histologia” e “diagnóstico” seguem o formato de texto livre, abrindo várias possibilidades de tratamento, diferenciando-os de outros que são de ordem quantitativa somente, em outros laboratórios.

Ao final seguem a identificação e assinatura do responsável pelo laboratório.

5.8 LABORATÓRIO DE PATOLOGIA CLÍNICA VETERINÁRIA (HCV)

Este laboratório utiliza uma ficha padronizada para a solicitação dos exames, enquanto o retorno dos resultados é feito através do sistema informatizado do hospital veterinário. Os campos estão descritos no quadro 8.

Quadro 8 - Ficha de solicitação de exames - Laboratório de Clínica Animal - Hemograma

Exame número	Proprietário
Ficha número	Data
Espécie	Raça
Sexo	Idade
Horário da colheita	Suspeita diagnóstica
Animal sob tratamento () Sim () Não. Qual:	História clínica resumida
Hemograma completo (sem plaquetas) ()	Reticulócitos ()
Plaquetas ()	Volume globular e Proteína plasmática total ()
Leucometria ()	Fibrinogênio ()
Observações ()	Médico(a) veterinário(a) responsável
CRMV	Assinatura

Fonte: O autor (2019)

Campos em destaque:

Exame número: identificação do exame

Ficha número: identificação da ficha utilizada

Horário da colheita: consta a hora em que o material a ser examinado foi colhido

5.9 HOSPITAL DE CLÍNICAS VETERINÁRIAS (HCV)

O HCV utiliza atualmente um sistema computadorizado, adquirido através de licitação, desenvolvido em PHP (linguagem de programação de código aberto), para controle dos dados de animais, de seus responsáveis, laudos, requisições, procedimentos etc. Foi permitido o acesso ao seu banco de dados, mantido pelo programa MySQL e que é formado por 94 tabelas em sua organização interna. As tabelas arranjam de maneira lógica as informações em um banco de dados moderno.

Utilizando o programa “MySQL Workbench”⁶ versão 8.0 CE, foi realizada a engenharia reversa do banco de dados, o que nos permitiu visualizar as tabelas e suas relações, proporcionando um panorama de como os dados são manejados no hospital veterinário.

Não há, atualmente, conexões automatizadas para a troca de informações entre os laboratórios citados acima e o HCV. Se for necessário que um material seja analisado, deve seguir as vias normais de aceite e cadastro, sem que haja troca de informações e por consequência, gerando retrabalho.

⁶ Disponível em <https://www.mysql.com/products/workbench>

6 ANÁLISE DOS DADOS LEVANTADOS

Este capítulo aprofunda a discussão sobre o que foi encontrado, para que seja possível alcançar o que o proposto nos objetivos específicos, com base nos fundamentos teóricos de Gestão do Conhecimento e da TI, como segue abaixo.

6.1 DISCUSSÃO

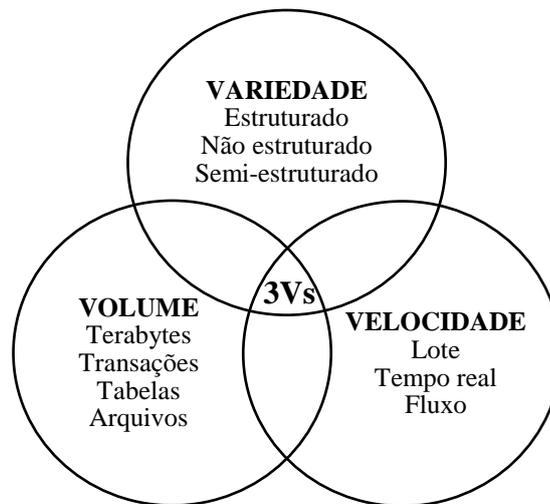
6.1.1 Quanto ao volume e à estrutura dos dados

Em relação ao seu volume, os dados armazenados nos laboratórios se apresentam em grande quantidade, principalmente o legado, que é mantido em papel, mas o fluxo diário de criação de novas informações é médio ou baixo, segundo o que foi constatado.

Desta maneira, estas bases de dados não se encaixam no que poderia ser classificado como “*Big Data*”, um paradigma de tratamento de dados que cresce cada vez mais, levado pelo custo decrescente das tecnologias de armazenamento e de processamento de informações (ERICKSON; ROTHBERG, 2014).

Para ser considerado como *big data*, os dados precisam ser gerados em altas velocidades, em grande volume e com grande variedade de tipos, segundo a definição mais aceita (DEWANGAN; THOMBRE; PATEL, 2016; LANEY, 2001), em agrupamentos com tamanhos acima da habilidade que programas e ferramentas comuns tem para capturar, gerenciar e processar, em tempo hábil (CHEN *et al.*, 2013). Na figura 5 estão ilustrados os chamados 3 “Vs” do *Big Data* e sua relação.

Figura 5 - Os 3 “Vs” do *Big Data*



Fonte: adaptado de Russom (2011)

Volume se refere à quantidade de dados que uma organização coleta, gera e processa, o que pode variar grandemente dependendo da sua área de aplicação (GANDOMI; HAIDER, 2015). A linha de corte utilizada atualmente, segundo Lee (2017) para que o tamanho de uma base seja considerada *big data*, é a de 1 *terabyte* (mil *gigabytes*⁷), mas esta definição é relativa e pode variar segundo muitos fatores, como o tipo de dados. À medida que no futuro as capacidades de armazenamento se ampliem e fiquem mais baratas, o que é hoje *big data* será considerado comum (GANDOMI; HAIDER, 2015).

Velocidade está relacionada ao ritmo em que é gerada e processada a informação. O ideal, atualmente, é que o seu processamento aconteça em tempo real, mas isto depende de equipamentos de alto desempenho e dos tipos de dados a serem controlados (LEE, 2017).

A sua variedade está ligada aos tipos de informações. Hoje as técnicas permitem o gerenciamento de vários tipos de dados, classificados como consta no quadro 9.

Quadro 9 - Tipos de dados quanto à estrutura

Tipo de dados	Descrição
Não estruturados	Não têm um formato que possibilita computação eficiente. Exemplo: Fotos, textos, áudio e vídeo.
Semiestruturados	Possuem algum tipo de organização. Exemplo: o formato XBRL (<i>Extensible Business Reporting Language</i>), membro da família das linguagens baseadas em XML ⁸ (<i>Extensible Markup Language</i>), desenvolvida para a troca eletrônica de informações comerciais.

⁷ Um gigabyte corresponde a um bilhão de bytes

⁸ XML é uma linguagem de marcação utilizada para a criação de documentos organizados hierarquicamente

Estruturados	É predefinido e pode ser encontrado normalmente em bases de dados tradicionais, como dados de transações.
--------------	---

Fonte: adaptado de Lee (2017)

A empresa IBM adicionou veracidade como outra dimensão à ser levada em conta e a SAS a qualidades de variabilidade e complexidade (GANDOMI; HAIDER, 2015; LEE, 2017). Veracidade se refere ao nível de confiabilidade ou incerteza que se tem em um dado, pela sua possível incompletude, subjetividade, inacurácia e/ou inconsistência. A variabilidade trata das alterações que ocorrem nas taxas de entrada de dados com o passar do tempo, às vezes imprevisíveis, e a complexidade se refere a um número grande de fontes de informações, o que pode tornar difícil armazenar, normalizar e processar dados que, por causa disso, podem ser muito heterogêneos (LEE, 2017). E, por fim, a empresa de tecnologia Oracle introduziu a qualidade chamada de valor, segundo a qual os dados presentes em uma base *big data* normalmente possuem valor baixo, em relação à sua densidade. O que significa que a informação em formato original precisa ser analisada primeiro, em grandes volumes, para dela poder se obter valor (GANDOMI; HAIDER, 2015). Temos assim, modernamente, então, 8 características que definem uma massa de dados como sendo *big data*.

Mas esta não é só gerada e utilizada por grandes empresas do varejo, companhias de internet, bancos ou instituições financeiras (pelo uso ubíquo que fazem de ferramentas de TI), como se poderia esperar. Vários tipos de empresas e entidades de pequeno e médio porte geram grandes quantidades de dados estruturados e não estruturados, impulsionadas pelo acesso à tecnologia e pela sua participação em redes sociais (MELLO; LEITE; MARTINS, 2014).

Dessa maneira, segundo as classificações descritas acima, é possível afirmar que os laboratórios de análises de saúde animal, deste campus, trabalham, nativamente, com dados estruturados, por desde o momento de sua aquisição serem claramente identificados e armazenados de maneira ordenada, mesmo em mídia de papel. Em uma possível futura implementação informatizada destes sistemas, a estrutura das informações se manteria com características semelhantes, adicionando as melhorias que a TI oferece no trato de grandes massas de dados.

6.1.2 Quanto aos meios de armazenamento

Os laboratórios visitados seguem um padrão no manuseio dos dados, ao utilizarem formulários em papel para o cadastro de clientes e pacientes e para a devolutiva de resultados.

Neles, o formulário digital, cópia fiel da versão física, em formato *Microsoft Word*, é mantido em computador, ajudando na replicação, se for necessário, e servindo como uma cópia de segurança das informações, mas não é formalizada como tal. Na Gestão do Conhecimento, esta base de dados poderia fazer parte da Memória Organizacional, (um constructo maior), do laboratório, que, quando implantada, provê as capacidades de adquirir, representar, armazenar e recuperar o conhecimento (PEREIRA, 2013). O ideal, nestes casos, é a utilização de sistemas informáticos que implementem esse conjunto de recursos para facilitar o acesso aos seus conteúdos (MOLINA; VALENTIM, 2011).

Neste sentido, é preciso reforçar o papel chave que a TI desempenha em relação à Gestão do Conhecimento, pela sua capacidade de apoiar a comunicação e a colaboração, habilitando o aprendizado colaborativo e o compartilhamento de conhecimento (ALAVI; LEIDNER, 1999).

A implantação de sistemas individuais informatizados para a gerência das informações em cada laboratório supriria algumas lacunas, que são detalhadas a seguir.

6.1.3 Quanto à segurança

A segurança da informação em TI está intimamente ligada aos termos confidencialidade, integridade e disponibilidade, a tríade que descreve os elementos fundamentais da segurança de dados desde que foi cunhada, por Saltzer e Schoeder, no artigo seminal “*The Protection of Information in Computer Systems*”, de 1975 (KUMAR; RAJ; JELCIANA, 2018).

Neste paradigma, confidencialidade está relacionada à utilização de práticas em TI que evitem a revelação de uma informação, ou parte dela, para alguém ou um sistema que não tenha a autorização adequada. Para alcançar este fim, a TI se utiliza de várias ferramentas, que implementam os conceitos de criptografia, autenticação, autorização, controle de acesso e até mesmo a segurança física dos dados digitais, por exemplo (GOODRICH; TAMASSIA, 2013).

Integridade prevê que o sistema seja capaz de garantir que a informação não foi alterada, quando armazenada ou em trânsito. A criptografia e o controle de autenticação, juntamente com técnicas que promovam a redundância dos dados ajudam na manutenção da sua integridade (GOODRICH; TAMASSIA, 2013; KUMAR; RAJ; JELCIANA, 2018).

E, finalmente, a disponibilidade se refere à possibilidade de, sempre que for necessário modificar ou ter acesso à uma informação, ela estará disponível para este fim (COSS; SAMONAS, 2014), evitando a chamada negação de uso (SALTZER; SCHROEDER, 1975).

Exemplo: programas e equipamentos responsáveis por cópias de segurança de um sistema (*backups*).

Atualmente é muito difícil garantir quaisquer um dos termos citados acima, nos laboratórios, mesmo se utilizando de controle de acesso físico aos documentos impressos ou listas de permissão de manuseio deles. Nas versões digitais, armazenadas em formato *Microsoft Word*, pouco se pode fazer neste sentido, levando em conta que este programa não foi desenvolvido para este tipo de controle.

6.1.4 Quanto à Gestão Eletrônica de Registros de Saúde Animal

A ideia de se utilizar tecnologias computacionais para melhor gerenciar dados laboratoriais é antiga, sendo citada pelo Dr. Vannevar Bush, já em 1945, no artigo “*As We May Think*”, em que descreve o uso de um dispositivo, então imaginário, capaz de armazenar e recuperar rapidamente dados médicos (PARK *et al.*, 2012). Este era o primeiro esboço do que, posteriormente, seriam os primeiros LIS (*Laboratory Information Systems*) ou Sistemas de Informações Laboratoriais que, atualmente, são montados através de programas de computador interligados e infraestruturas de equipamentos dedicados, de maneira a dar suporte às necessidades de processamento de informações em laboratórios.

Um LIS está presente em todas as fases dos exames, incluindo a entrada do pedido e das amostras, processamento e controle das amostras, suporte à análise e interpretação dos resultados e a criação dos relatórios, além de sua distribuição. Além disso, deve fornecer relatórios gerenciais, entre outros, que o laboratório precisa para manter suas operações e melhorar constantemente a qualidade dos seus serviços (HENRICKS, 2016)

É impensável que estes níveis de análise de informações, monitoramento, extração de padrões e troca de relatórios possam ser alcançados passando ao largo da utilização das mais modernas técnicas de análise fornecidas pela TI. Por isso considera-se, atualmente, que os registros eletrônicos de saúde animal, gerados por laboratórios, são indispensáveis para o bem-estar dos animais, da saúde humana e por consequência, do meio ambiente em que convivemos, sem falar dos impactos econômicos que as doenças relacionadas à animais desencadeiam, se não forem monitoradas, evitadas ou pelo menos atenuadas.

6.1.5 Quanto ao Compartilhamento de Informações

Com o estabelecimento global da Internet e das tecnologias que nela se apoiam, como as páginas *web*, o e-mail e redes sociais, é possível o seu uso para o acesso controlado e prático à resultados parciais ou finais de exames para clientes ou profissionais da área de saúde animal, para confecção de laudos ou emissão de receituário para tratamento do animal e geração de estatísticas ou materiais de estudo para alunos e professores dos cursos alinhados à estes assuntos, do campus em questão.

Em casos de convênios com órgãos municipais ou estaduais de controle de zoonoses, pesquisas, extensão e desenvolvimento agropecuário etc., seria possível o compartilhamento segmentado de dados através de canais criptografados e autorizados previamente, para geração e troca de relatórios e sumarização de ocorrências.

6.1.6 Quanto às Análises Individuais

Laboratórios com sistemas informatizados permitiriam análises mais rápidas da sua massa de informações. À medida que uma organização se expande e seu método de gerenciamento o acompanha, se tornando mais complexo, acaba por gerar uma quantidade cada vez maior de dados, a maioria brutos. Em tempos em que toda esta cadeia de conhecimento está informatizada, estes dados podem ser usados para retornar informações úteis e conhecimento para a organização, tais como modelos de padrões, tendências ou de comportamento, que podem não ser facilmente discerníveis e, depois de revelados, ajudam em decisões táticas e operacionais (BARLAS; LANNING; HEAVEY, 2012).

Os padrões encontrados, originados da área de saúde animal de cada laboratório, forneceria uma visão do estado de saúde dos pacientes e do seu meio-ambiente, além de proporcionar, aos gestores de cada laboratório, material decisório de como ele pode ser aperfeiçoado, como as grandes empresas fazem hoje, com seus negócios. O conhecimento extraído destas fontes pode descrever o que aconteceu, explicar porque algo aconteceu, e prever o que pode ou é provável que aconteça (BAŠKARADA; KORONIOS, 2017).

E, no momento em que cada laboratório tiver suas análises individualizadas, estas poderiam ser cruzadas com as de outros, ampliando as possibilidades de geração de conhecimento.

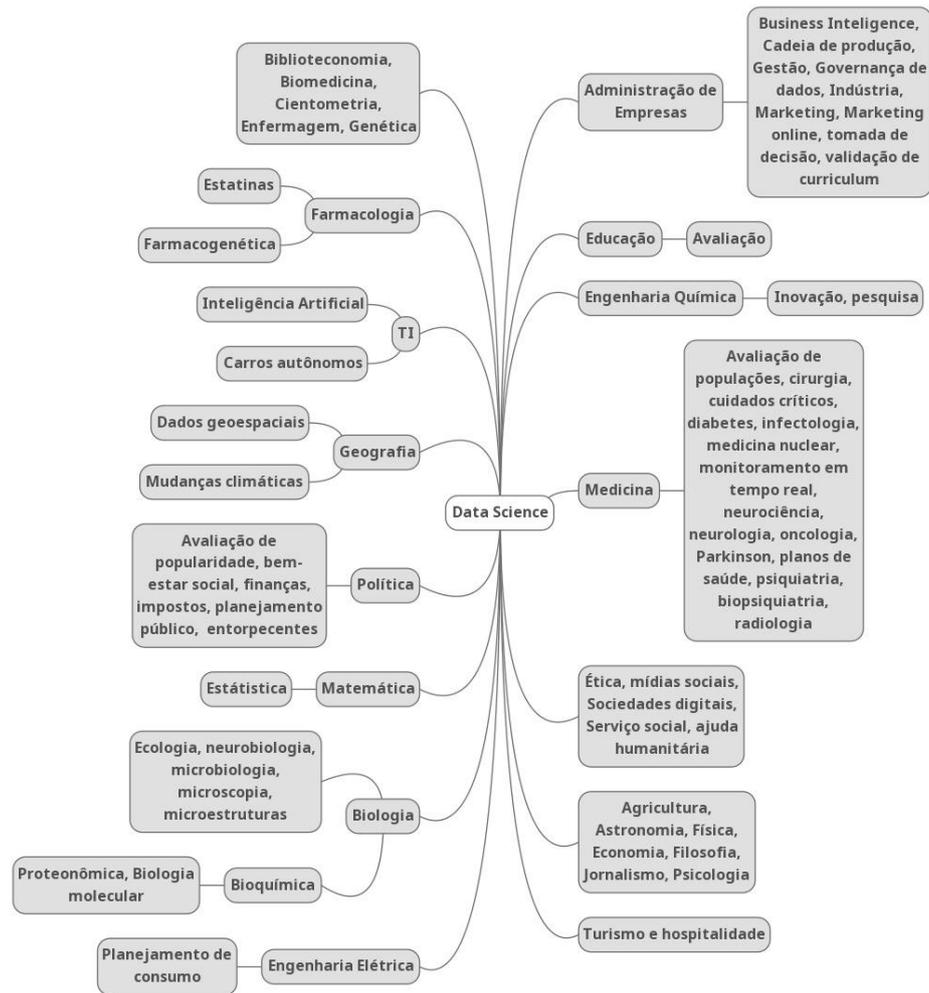
Isto já acontece na medicina humana, onde se entende que a digitalização dos sistemas de saúde (*health-care systems*) mudou a maneira que se pratica a medicina e se conduz pesquisas clínicas. A disseminação dos prontuários médicos eletrônicos (*eletronic health records*) tem levado à pesquisas em *big data*, trazendo a ciência de dados (*data science*) para os hospitais (SANCHEZ-PINTO; LUO; CHURPEK, 2018).

6.2 DETERMINAÇÃO DA TECNOLOGIA DE CRUZAMENTO DE DADOS

Durante o levantamento bibliográfico realizado para fundamentar este capítulo, os termos “*big data*”, “*data analytics*” e “*data science*” foram os que mais se apresentaram quando o assunto tratado era a análise de grandes massas de dados, buscando a extração de ativos de conhecimento, seja de bases únicas ou da correlação entre várias, o que aqui é identificado como dados interlaboratoriais.

A disciplina de *Data Science*, que engloba as outras citadas acima, aparece normalmente ligada à vários tipos de pesquisas científicas. Uma busca na base de dados Periódicos Capes pelo termo “*data science*”, em 2019, retornou aproximadamente 53.000 artigos, publicados nos últimos 5 anos, todos em inglês. Ao serem feitas as análises dos 160 resultados mais relevantes, segundo a interface do sistema, foi elaborada a figura 6, um resumo dos campos de estudo que citam e utilizam *Data Science* em seus trabalhos.

Figura 6 - Algumas áreas de pesquisa científica que utilizam *Data Science*



Fonte: Periódicos Capes (2019). Adaptado pelo autor

Este termo foi utilizado pela primeira vez por Peter Naur, em 1974, no livro “Concise Survey of Computer Methods”, para descrever a ciência de lidar com os dados, “uma vez que tenham sido definidos, enquanto a relação dos dados com o que eles representam é delegado à outros campos ou ciências” (NAUR, 1974).

Esta expressão tem várias definições, atualmente, a depender de cada autor. O quadro 10 relaciona algumas:

Quadro 10 - Definições de “*Data Science*”

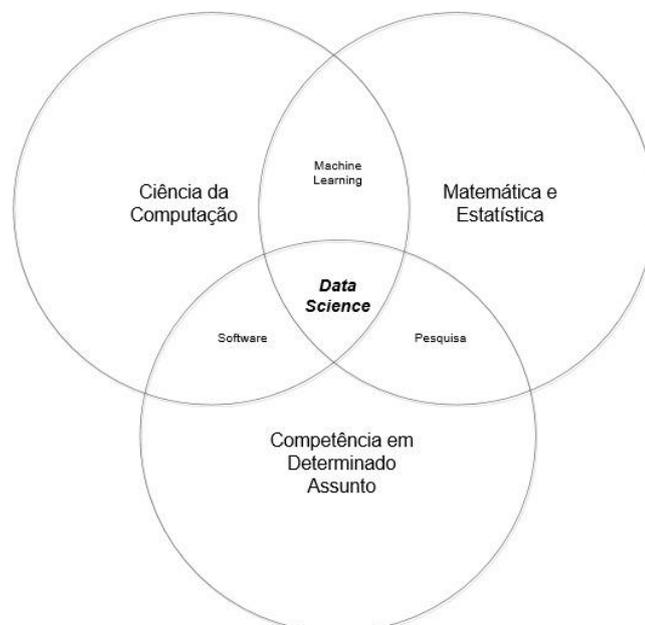
Davenport e Patil (2012, tradução nossa)	<i>Data Science</i> é uma disciplina aplicada emergente focada em facilitar as tomadas de decisão nas organizações, através do
--	--

	desenvolvimento de modelos estatísticos que extraem conhecimento de dados brutos.
Provost e Fawcett (2013, tradução nossa)	<i>Data Science</i> é um conjunto de princípios fundamentais que suportam e guiam a extração de informação e conhecimento a partir dos dados. O conceito relacionado mais próximo de <i>data science</i> é <i>data mining</i> (mineração de dados).
Foreman (2014, tradução nossa)	<i>Data Science</i> é a transformação de dados usando matemática e estatística em <i>insights</i> valiosos, decisões e produtos
Cielen, Meysman e Ali (2016, tradução nossa)	<i>Data Science</i> é uma extensão evolucionária da estatística, com capacidades de lidar com quantidades massivas de dados produzidos atualmente. Ela adiciona métodos da Ciência da Computação ao repertório da estatística.
Blei e Smyth (2017, tradução nossa)	<i>Data Science</i> foca em explorar o dilúvio moderno de dados em busca de predição, exploração, entendimento e intervenção. Enfatiza o valor e a necessidade de aproximação e simplificação. Valoriza a comunicação eficaz dos resultados das análises e o entendimento do mundo que extraímos dele.

Fonte: dados da pesquisa (2019)

Neste sentido, esta ciência se mostra como um campo amplamente multidisciplinar, por unificar diversas áreas, como estatística, análise de dados e ferramentas computacionais de apoio, *machine learning* (aprendizado de máquina), matemática, inteligência artificial e *business intelligence* (inteligência de negócios), entre outros, como ilustrado na figura 7.

Figura 7 - Áreas de conhecimento que formam o Data Science



Fonte: adaptado de Schutt e O'Neil, (2013)

Por ser uma ferramenta bastante difundida (CIELEN; MEYSMAN; ALI, 2016), é usada para fins comerciais e não comerciais, na tentativa de ampliar o entendimento na área em que

é aplicada. Como a maioria das organizações tendem a gerar grandes volumes de dados, atualmente, é preciso transformar esta quantidade de informações em conhecimento útil, o que por consequência traz vantagens competitivas para empresas e aceleração em pesquisas científicas.

Sendo assim, as técnicas aplicadas de *Data Science* podem contribuir para a Gestão do Conhecimento, o desenvolvimento da memória organizacional e a criação de ativos intangíveis, que em *Data Science* são semelhantes aos “*insights*”⁹.

O *Data Science* é aplicado em fases. Estes passos não se processam necessariamente de forma linear, mas na realidade, de maneira interativa, com o pesquisador retornando à fases anteriores gerando conhecimento de maneira incremental, o que pode inclusive dar origem à novas perguntas (CIELEN; MEYSMAN; ALI, 2016)

É necessário ter em mente que cada uma destas fases é aplicada de maneira diferente e variam em quantidade, dependendo do autor, mas mantêm, em última instância, o foco na geração e aplicação do conhecimento, além da geração de ativos de conhecimento (Gestão do Conhecimento) ou dos “*insights*” (*Data Science*), que é a compreensão ou clareza súbita sobre um determinado assunto, que pode posteriormente ser usado para ajudar em uma tomada de decisão. Interessantemente, este modelo replica, de maneira análoga, o próprio ciclo de vida dos dados, como apresentado por Lazer *et al.* (2009, tradução nossa):

Como organismos biológicos, dados tem um ciclo de vida, desde o nascimento, passando por uma vida ativa, até a ‘imortalidade’ ou alguma forma de expiração. E como um organismo vivo e inteligente, eles sobrevivem em um ambiente que provê tanto suporte físico quanto significado existencial.

6.2.1 A importância dos *Insights*

Como a partir deste momento estão sendo associados os conceitos de ativos de conhecimento com o de *insights*, é necessário aprofundar os seus significados e a sua importância.

A expressão ativos de conhecimento tem, como sendo uma de suas definições, o conjunto do *know-how* das equipes de uma organização e as suas rotinas, ligadas a cada uma, todos colocados como recursos estratégicos não palpáveis e fontes de criação de valor para a

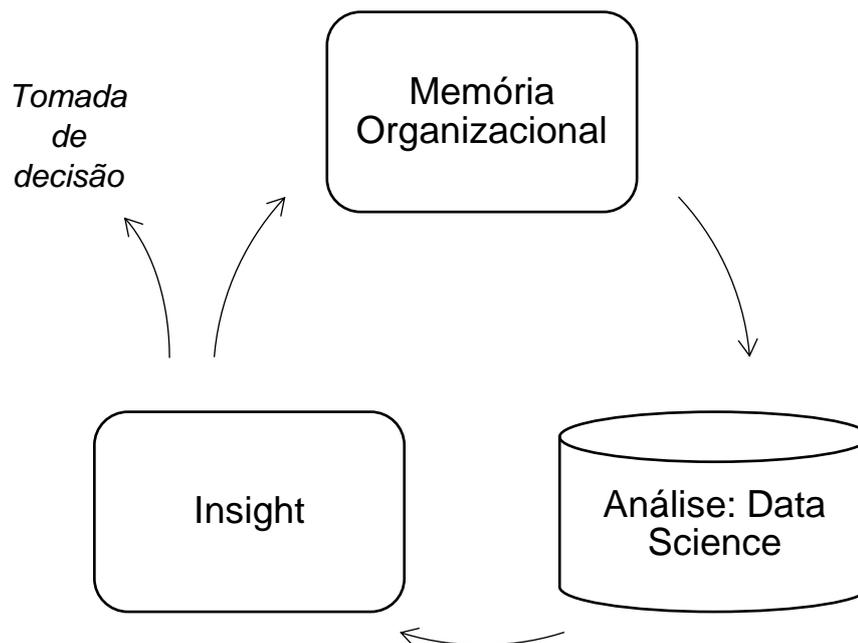
⁹ Em *Data Science*, é a conclusão ou compreensão súbita acerca de algo, baseado na análise de dados relacionados

organização, que assim acabam por fornecer valor ao seu titular (FERNANDES; FADEL, 2013; SCHIUMA; CARLUCCI; SOLE, 2012; YANG; FANG; LIN, 2010).

De maneira semelhante, o termo *insight* nasce em *Data Science* ligado ao sentido original que a expressão possui em inglês: uma série de sinônimos para valor, informação, predições úteis, análises, dados-chave, modelos de comportamento, etc., sempre quando estes surgem a partir da análise de uma massa de dados, buscando responder a uma pergunta prévia ou tentando trazer o entendimento para uma situação (BLEI; SMYTH, 2017; DAVENPORT; PATIL, 2012; MATHUR; PUROHIT, 2017).

Isto posto, poder-se-ia dizer que os ativos de conhecimento de uma organização dão origem a *insights*, quando são analisados, e englobam aqueles que já conhece. São a fonte destes e imediatamente, ao originá-los, os absorve, se ampliam e se retroalimentam, como ilustrado na figura 8.

Figura 8 - Fluxo: memória organizacional, Data Science e *insights*



Fonte: o autor (2019)

A Gestão do Conhecimento pode, então, dessa maneira, se utilizar de *Data Science* para produzir os ativos de conhecimento, na forma de *insights*, subprodutos deste constructo maior, para o qual acabam por retornar.

Baseado nisto, esta tecnologia foi a escolhida como a mais indicada para a análise dos dados interlaboratoriais, e aplicada através da prova de conceito.

6.3 PROVA DE CONCEITO

Buscando demonstrar de maneira prática como os princípios do *Data Science* podem ser aplicados sobre dados interlaboratoriais, utilizou-se o método conhecido como “Prova de Conceito” (*Proof of Concept*).

Esta técnica é descrita como sendo um tipo de protótipo, ainda em fase inicial, empregado quando se deseja testar alguns aspectos de um *design* final pretendido, sem a intenção de simular a aparência, materiais ou processo de desenvolvimento final da solução. Estas simulações podem ser usados para “provar” o potencial de um método ou sua viabilidade, demonstrando o que vai ou não funcionar, apontando em que direção testes e desenvolvimentos são necessários (CARSTEN, 1989).

Para a modelagem deste tipo de prova de conceito estão disponíveis vários programas, alguns gratuitos e liberados sob licenças de código-fonte aberto (*open source*¹⁰) e outros comerciais. Sobre as escolhas feitas, será discorrido a seguir.

6.3.1 Bancos de Dados

Para realizar os testes, foram projetados e implementados os bancos de dados dos laboratórios Cedima, Parasitologia e Patologia Animal, usando como ponto de partida os achados do levantamento documental. O laboratório de Patologia Clínica não foi incluído nos experimentos por já possuir uma parte de seu uso implementada no sistema interno do hospital veterinário e compartilhar de várias características com os citados anteriormente. Apesar disso, sugestões de correlação interlaboratorial com seus dados são feitas ao final deste capítulo.

O *software* MySQL foi o gerenciador de banco de dados escolhido, um dos mais populares do mundo (SOLIDIT, 2019), pela facilidade de uso, por trabalhar com a linguagem de manipulação de dados SQL, amplamente suportada por ferramentas de *design* de bancos de dados, e pela integração com a linguagem Python e a plataforma Jupyter, que também serão utilizados.

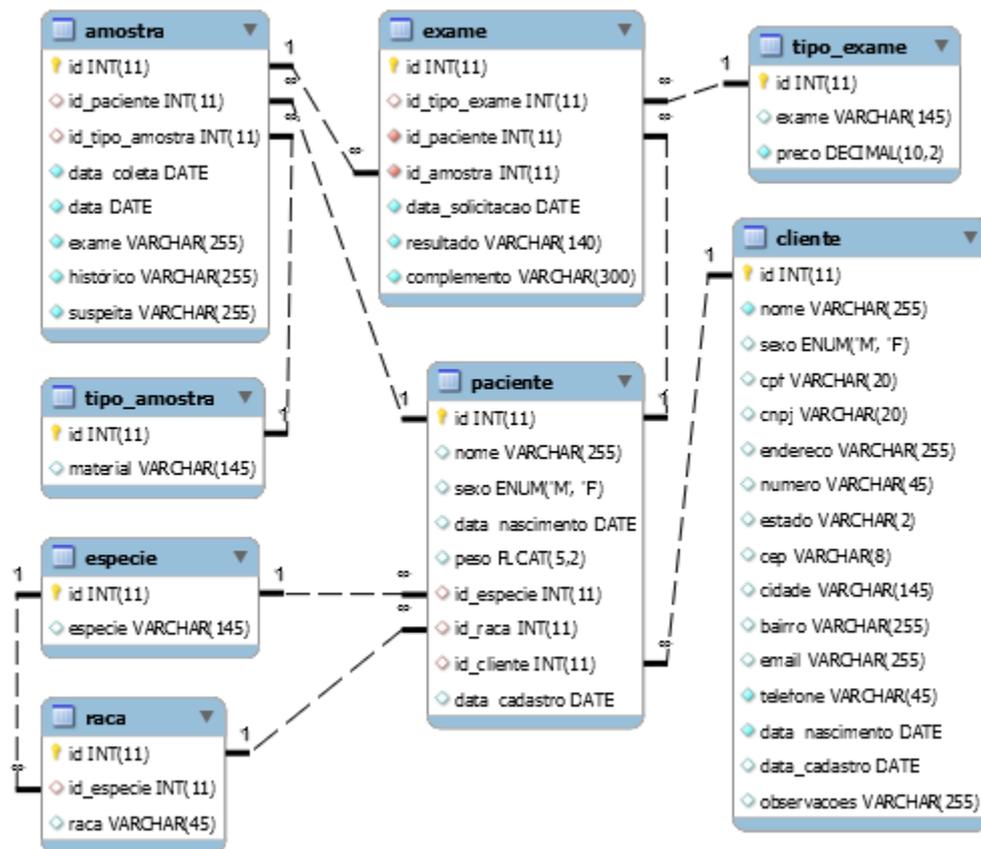
Os bancos de dados foram criados com base nos campos dos formulários reais e na lógica de funcionamento de cada laboratório, com as adaptações necessárias e a adição de novos

¹⁰ Um programa de computador com o código fonte licenciado e liberado para o estudo e modificação para qualquer pessoa e para qualquer finalidade, além de outras garantias

campos, que pudessem ser úteis nas simulações. Seguindo o que se espera de uma prova de conceito, são versões simplificadas em relação àquelas que poderiam estar em uso ou em produção. Não se propõem a ser o projeto ideal nem o mais prático. Sua função é demonstrar a aplicabilidade das técnicas propostas.

Segue abaixo o modelo de entidade relacionamento e a descrição de cada um dos bancos de dados criados, gerados pelo programa MySQL Workbench. A figura 9 é específica para o Cedima. Os códigos SQL usados na criação de cada um estão disponíveis nos apêndices, referenciados individualmente.

Figura 9 - Tabelas do banco de dados Cedima



Fonte: dados da pesquisa (2019)

Atualmente, os laboratórios que foram simulados não possuem registros disponíveis em mídia digital que possam ser usados para análise, como proposto. Por esta razão, os bancos de dados foram preenchidos com dados fictícios, usando informações geradas de maneira

randômica através da a linguagem Python, na interface Jupyter. Eis aí a importância dos códigos para a geração de dados aleatórios.

Abaixo segue um exemplo de *script*¹¹ em Python que populou a tabela “amostras” da base de dados “Cedima” (quadro 11). As linhas que iniciam com o símbolo “#” são comentários, não sendo interpretadas pelo Python como um comando válido.

Quadro 11 - Script de preenchimento da tabela amostra – Cedima

```
# script para a inserção de amostras na tabela "amostra" - Cedima

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='cedima',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de opções de tipos de amostras para exame
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    lista_amostras = { "14": 10, "13": 10, "12": 9, "11": 9, "10": 8, "9": 8, "8": 7, "7": 6, "6": 6, "5": 6, "4": 5, "3": 4, "2": 3, "1": 2}

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

            x=1
            while(x <= total_inserir):

                # retorna um dos itens da lista, levando em conta o peso
                tipo_amostra=random.choice([k for k in lista_amostras for dummy in range(lista_amostras[k])])

                # data coleta
                data_coleta = fake.date_between(start_date="-5y", end_date="now")

                # sql
                sql = "INSERT INTO `amostra` (`id_tipo_amostra`, `id_paciente`, `data_coleta`, `data`)
                      VALUES (%s, %s, %s, %s)"

                # sql passado com os parâmetros acima
                cursor.execute(sql, (tipo_amostra, x, data_coleta, data_coleta))
                connection.commit()

                # atualiza a barra de progresso
                time.sleep(0.1)
                bar.update(x)

                x+=1

# eof
finally:
    cursor.close()
    connection.close()
```

Fonte: o autor (2019)

¹¹ Lista de comandos que descreve uma sequência de ações a serem realizadas por um sistema computadorizado

Quatro informações são necessárias em cada tupla¹², nesta tabela: a identificação do tipo de amostra, a identificação do paciente sendo testado, a data da coleta da amostra e a data da criação do registro no sistema. Para cada uma destas informações, o código se encarrega de gerar, aleatoriamente, um dado para ser inserido na tabela.

No caso do tipo de amostra, foi utilizado o método da linguagem Python chamado *random.choice*, que permite que um determinado valor seja escolhido em uma lista, de maneira randômica e, se for necessário, é possível determinar um “peso” para cada item da lista, fazendo com que a escolha aconteça baseada neste valor, privilegiando algum (PYTHON, 2019). Esta técnica foi usada para que a distribuição dos valores neste campo fosse mais próxima da realidade, já que a tendência seria que entre os 14 tipos de exames possíveis as inserções se dividiriam de maneira igualitária, se assim não fosse designado.

A data da amostra é também gerada aleatoriamente, contando da data atual até os últimos 5 anos. Para fins de simplificação, a data da criação do registro é a mesma da data da coleta.

A variável “total_inserir” controla quantos registros serão adicionados à tabela. Para os testes desta prova de conceito, foram usados 10.000 registros.

No quadro 12 está uma listagem dos serviços que o laboratório realiza, utilizado como um dos parâmetros dentro do código.

Quadro 12 - Tipos de exame - Cedima

Tipo	Descrição
1	Isolamento e identificação bacteriana
2	Antibiograma
3	Cultura e antibiograma
4	Contagem bacteriana
5	Pesquisa de <i>Salmonella</i>
6	Pesquisa de <i>Listeria</i>
7	Pesquisa de <i>Staphylococcus coagulase</i>
8	Pesquisa de anaeróbicos
9	Coliformes totais e termotolerantes
10	Isolamento viral
11	BHV-1 e 5
12	FeLV (imunofluorescência direta)
13	Leucose bovina (imunodifusão em gel de ágar)
14	Vacina papilomatose
15	FIV (PCR)
16	BVD (PCR)
17	BHV-1 (imunofluorescência direta)

¹² Em bancos de dados, é sinônimo para uma linha da tabela

Fonte: dados da pesquisa (2019)

O campo “id_tipo_exame” foi preenchido com um rol de nomes de vírus e bactérias, como consta no código do *script*, com seus determinados pesos, que afetam a sua escolha randômica, visando a simulação das ocorrências como se fossem casos reais, que de outra maneira seriam distribuídas de maneira uniforme.

Utilizando a biblioteca Faker (ZANINOTTO, 2019), do Python, foram gerados os resultados “positivo” ou “negativo”, com uma chance de 65% de preferência para os “positivos”.

O código se encarrega de adaptar os valores dependendo do resultado para o tipo de exame e o resultado, retornando um vírus ou uma bactéria, para o registro.

A tabela “tipo_amostra” contém os tipos de amostras e suas identificações únicas, com as quais o laboratório trabalha, como listados na figura 10. Estas informações constam na documentação do laboratório, à qual foi permitido o acesso.

Figura 10 - Amostra de registros da tabela "tipo_amostra"

id	material
1	material de aborto
2	fezes
3	secreção nasal
4	órquão
5	urina
6	soro
7	sanque com anticoagulante
8	secreção auricular
9	secreção anal
10	secreção oral
11	secreção ocular
12	secreção vulvar
13	secreção prepucial
14	outros

Fonte: dados da pesquisa (2019)

Estes dados são utilizados diretamente pela tabela “exame”, que os armazena.

A tabela “espécie” (figura 11), é composta de nomes de espécies de animais que mais comumente são atendidos pelo laboratório, com uma identificação para cada um, e conta no momento com 53 registros. Esses valores são os mesmos que constam no sistema de gerenciamento do HCV, e por ser bastante completa e pensando nos testes em que há correlação de informações entre estes, foram adicionados nesta base também.

Figura 11 - Amostra de registros da tabela "espécie"

id	especie
10	FELINA
21	CANINA
38	SUINA
41	BOVINA
53	OVINA
62	CAPRINA
77	ROEDORES
85	GALINÁCEOS
96	EQUINOS
906	MUAR
912	CUNICOLA

Fonte: dados da pesquisa (2019)

A tabela “raça” (figura 12) lista as raças dos animais mais comuns (265 registros, no total), presentes na tabela “espécie”. Através do programa que usará o banco de dados é possível se aproveitar das relações entre as tabelas, chamadas de relacionamentos. Esta técnica é uma das características que definem os sistemas gerenciadores de bancos de dados de terceira geração, como o MySQL (HOGAN, 2018).

Figura 12 - Amostra de registros da tabela "raça"

id	id_especie	raca
100	21	S.R.D
115	21	POODLE
123	21	DOBERMANN
131	21	FILA BRASILEIRA
146	21	COCKER SPANIEL
152	21	DOGO ARGENTINO
168	21	SHEEP DOG
174	21	ROTTWEILLER
187	21	BOXER
191	21	SCHNAUZER
201	21	PASTOR ALEMAO
216	21	AKITA
222	21	COLLIE

Fonte: dados da pesquisa (2019)

A tabela “cliente” mantém os cadastros padrão dos usuários do laboratório, constando com nome, endereço, CNPJ ou CPF e outras informações pessoais pertinente. Para os fins das simulações desta prova de conceito, somente alguns destes campos foram preenchidos, pois poderiam ser analisados e retornariam *insights* de interesse. Mesmo assim, a adição de vários registros na tabela é necessária para manter a integridade referencial das outras.

Campos como data de nascimento, para cálculo de idade dos clientes, endereço, bairro e estado, para geração de listagens com localização de origem ou mapas, são os mais úteis, neste caso.

A biblioteca Faker é capaz de gerar informações aleatórias de endereço, oferecendo a possibilidade de configuração de localização para o Brasil com o comando:

```

# importa a biblioteca
from faker import Faker

# configura a localização para Português Brasil
fake = Faker('pt_BR')

# gera quatro registros
for _ in range(4):
    print (fake.address(), '\n')

# Saída:

Jardim Freitas, 30
Jardim Do Vale
51813748 Gonçalves de Costela / RN

Feira de Pereira, 566
Conjunto Providencia
05890-380 Oliveira / CE

Pátio Otávio Gomes, 8
Vista Alegre
20733187 das Neves do Sul / MA

Morro da Cruz, 8
Santa Maria
67163135 Aragão / ES

```

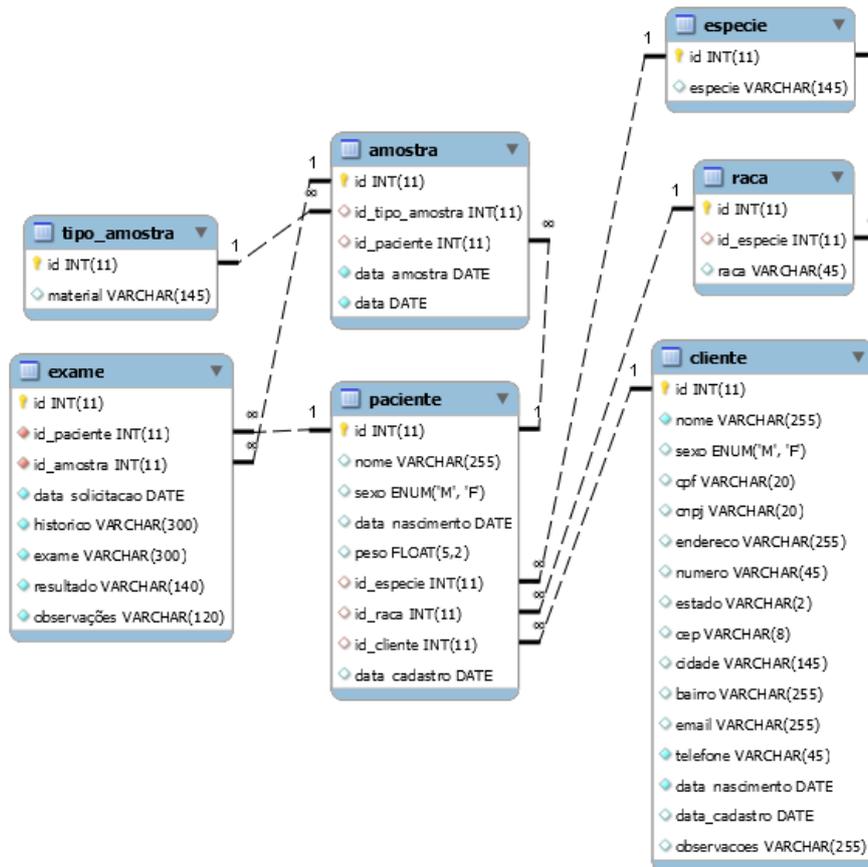
A partir deste ponto ele passa a gerar endereços locais, mas o alcance é nacional, fugindo do escopo de nossas necessidades. Sendo assim, nos testes em que estas informações são necessárias, elas foram adicionadas manualmente ou de maneira não aleatória.

Por fim, a tabela “paciente” armazena as informações do animal cujo material será analisado. Como citado anteriormente, alguns campos foram adicionados além daqueles que o laboratório usava, como peso, óbito e data do óbito, podendo ser úteis nas simulações.

A estrutura do banco de dados do laboratório Parasitologia (figura 13) se assemelha bastante com a anterior, repetindo alguns mecanismos de gerenciamento das informações. Os pontos em que se diferenciam é que a quantidade de tipos de amostras é em número menor e não há uma tabela que lista os tipos de exames, já que o laboratório trabalha especificamente com a detecção de qualquer espécie de parasita, além de alguns campos que possuem nomes diferentes.

Dessa maneira, para cada base de dados citada há uma série de *scripts* para o seu preenchimento. Estes estão disponíveis nos apêndices.

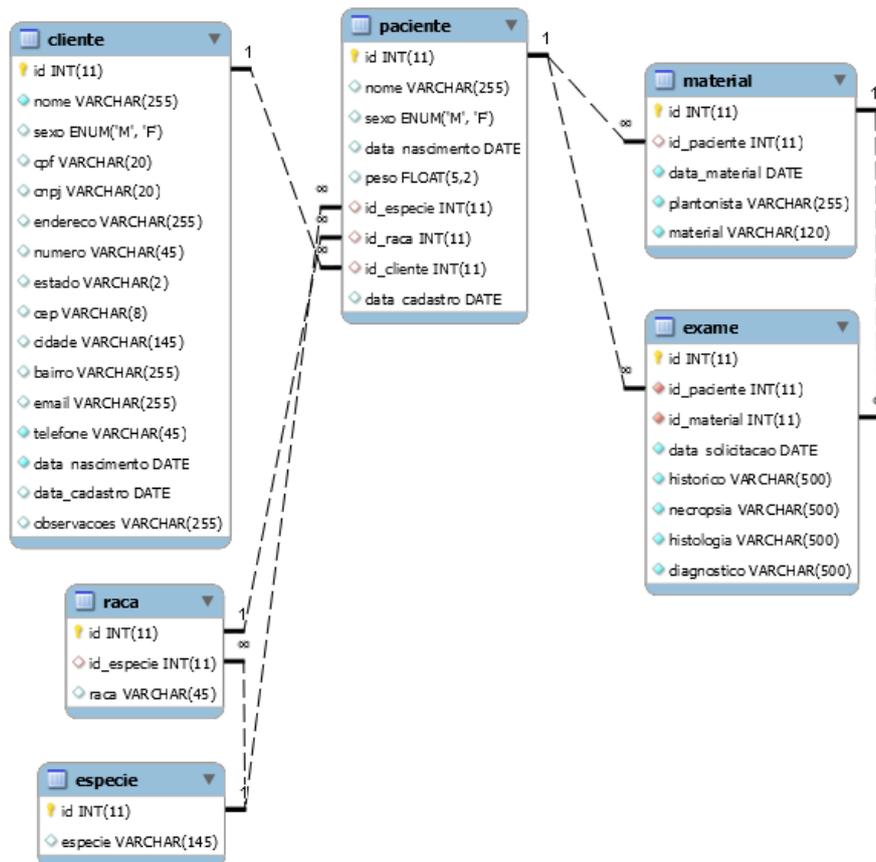
Figura 13 - Tabelas do banco de dados Parasitologia



Fonte: o autor (2019)

No caso da Patologia Animal, em particular (figura 14), as informações que são armazenadas na tabela “exame” seguem um padrão diferenciado. Como consta nas descrições dos documentos físicos deste laboratório, os campos “histórico”, “necropsia”, “histologia” e “diagnóstico” são preenchidos com dados qualitativos, em oposição à maioria dos outros dos outros laboratórios, onde o “positivo” ou “negativo” ou o nome de um vírus ou parasita são a maioria. O mesmo acontece na tabela “material”, que, no campo de mesmo nome, pode descrever quaisquer tecidos biológicos de um animal, em várias quantidades, como amostra.

Figura 14 - Tabelas do banco de dados Patologia



Fonte: o autor (2019)

Esta característica altera a maneira como as informações podem ser tratadas, para a extração das informações na forma de relatórios ou estatísticas ou na geração dos dados aleatórios para os testes.

Os *scripts* encarregados destas funções se utilizam, para o preenchimento dos campos considerados como qualitativos, de listas de termos relacionados à disciplina de Patologia Animal e que são comumente utilizados em laudos técnicos, que são aleatoriamente selecionados para o preenchimento dos campos correspondentes. Dessa maneira é possível realizar os testes que a prova de conceito prevê.

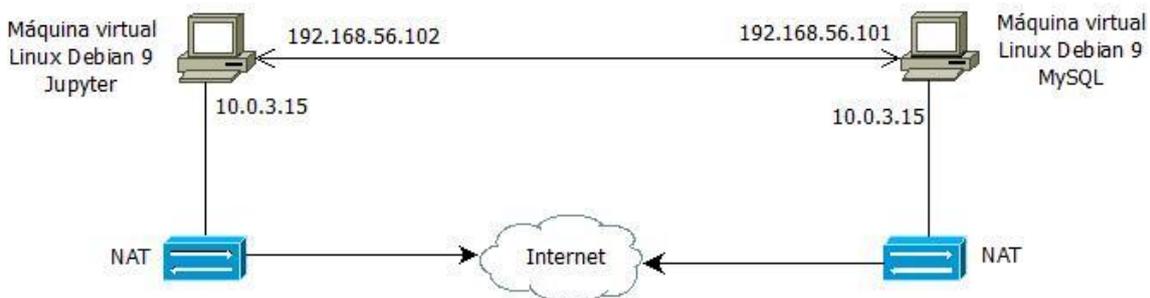
6.3.2 Linux e VirtualBox

Como sistema operacional¹³ usado pelos programas da prova de conceito, foi escolhida a distribuição Linux Debian¹⁴, na versão 9.

Para que vários servidores possam ser utilizados simultaneamente, simulando uma situação real de acesso cliente-servidor, foi usado o programa de virtualização VirtualBox¹⁵, recentemente adquirido pela empresa Oracle. Esta é uma solução de virtualização liberada como *open source* (de código aberto), sob os termos da GPL¹⁶ versão 2 (ORACLE, 2019). A arquitetura da solução de virtualização utilizada está demonstrada na figura 15.

Figura 15 - Estrutura das máquinas virtuais

PC Windows 10 + VirtualBox



Fonte: o autor (2019)

Abaixo seguem as imagens das configurações de cada máquina virtual criada para a simulação (figura 16 e 17).

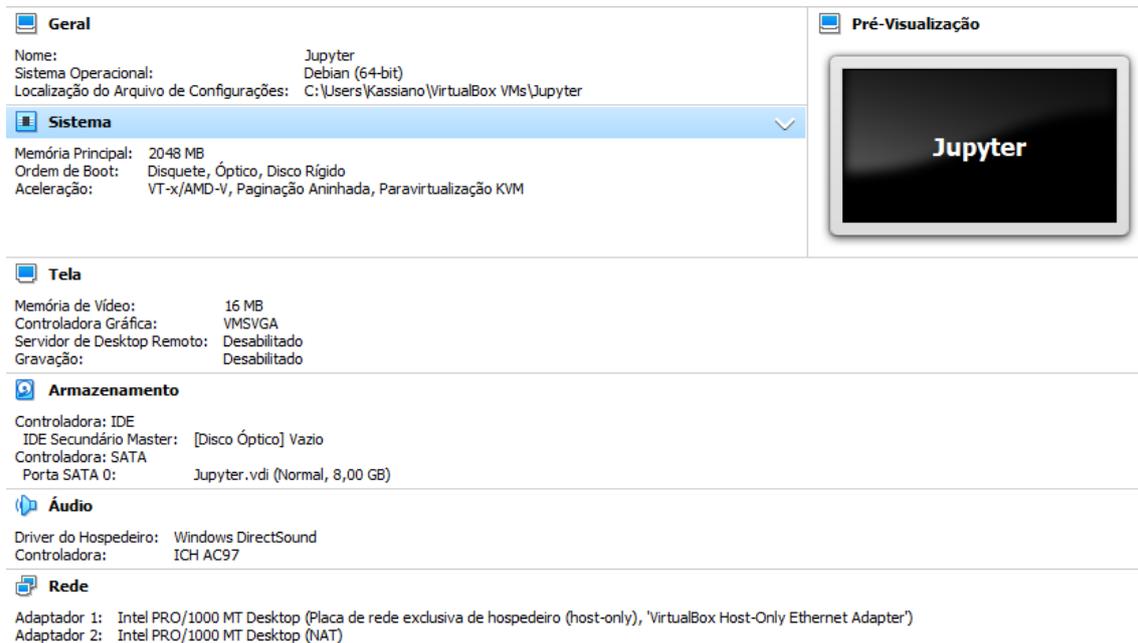
¹³ Programa com a função de gerenciar os recursos de um sistema, sendo a interface entre o usuário e o computador

¹⁴ Sistema operacional livre baseado em Linux, disponível em www.debian.org

¹⁵ Programa de virtualização para servidores, desktops e sistemas embarcados, disponível em www.virtualbox.org

¹⁶ Designa a licença de *software* desenvolvida por Richard Stallman, da *Free Software Foundation*, muito utilizada por projetos de *software* livre e de código aberto

Figura 16 - Linux virtualizado - Plataforma Jupyter

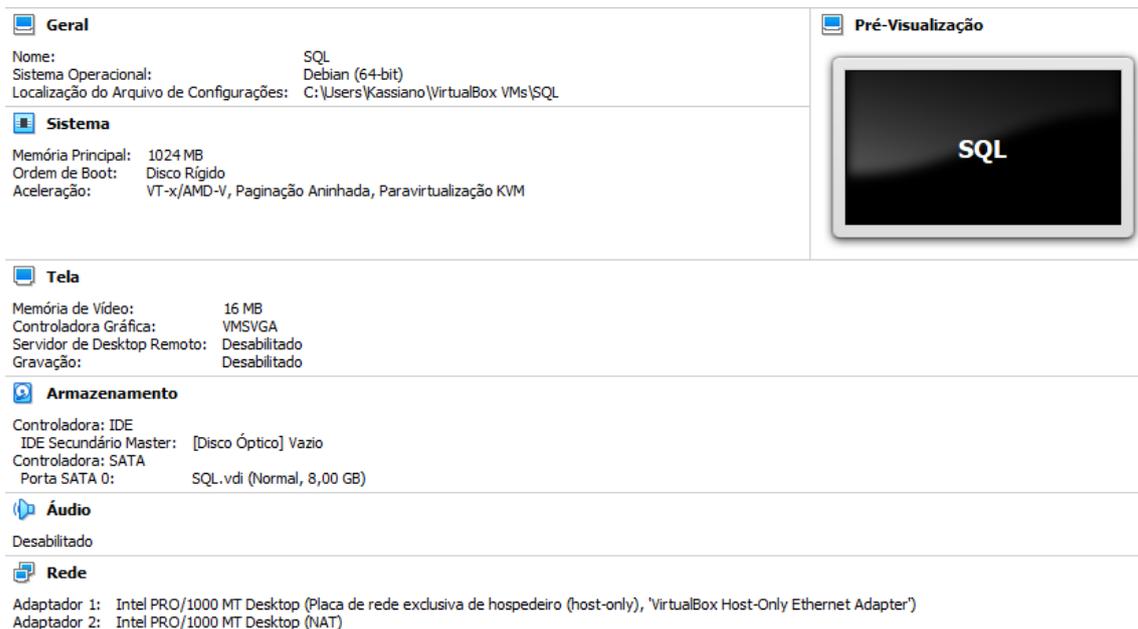


Fonte: o autor (2019)

A máquina virtual que hospedou a plataforma Jupyter (explicada com mais vagar logo abaixo) possui duas placas de rede virtuais, uma exclusiva para comunicação com a rede local e, através dela, com a outra máquina virtual (no caso, o servidor MySQL) e o plataforma Windows que a suporta, e outra placa de rede em modo NAT (*Network Address Translation*)¹⁷, para instalação de programas remotos ou acesso à dados na internet.

¹⁷ Técnica que permite que um equipamento em uma rede interna tenha acesso à internet, reescrevendo os endereços IP de origem ao passar por um roteador ou firewall

Figura 17 - Linux virtualizado – MySQL



Fonte: o autor (2019)

De maneira semelhante, a distribuição Linux utilizada pelo servidor de banco de dados MySQL tem duas placas de rede virtuais, uma para permitir o acesso à plataforma Jupyter a partir do computador Windows hospedeiro, onde está instalado. A outra é disponibilizada para acesso à rede externa, no caso, a internet.

6.3.3 Tableau, Anaconda e Jupyter Notebook

Para os testes, foram utilizados o programa Tableau e a plataforma Anaconda através da interface Jupyter.

O Tableau¹⁸, desenvolvido pela empresa *Tableau Software*, é um programa com 10 anos no mercado voltado para aplicações em *big data* e assemelhados, permitindo a análise de dados, geração de relatórios e gráficos através de interface gráfica, sem a necessidade de conhecimentos em programação (MURRAY, 2016). Este tipo de programa foi usado na tentativa de demonstrar a possibilidade da criação de informações úteis através das chamadas ferramentas de BI (*Business Intelligence*), que buscam diminuir a dependência na organização de especialistas em análise e nas ferramentas semelhantes, por meio do uso de uma interface simples (figuras 18 e 19), em que os campos que se deseja analisar são dispostos pela tela e o

¹⁸ Disponível em <https://www.tableau.com>

resultado do cruzamento das informações aparece imediatamente, calculados a partir da base de dados escolhida. Para os fins desta dissertação foi utilizada uma licença educacional fornecida pela empresa Tableau, por um ano, liberando o uso completo do programa na versão profissional.

Figura 18 - Interface do programa Tableau: configuração de conexão

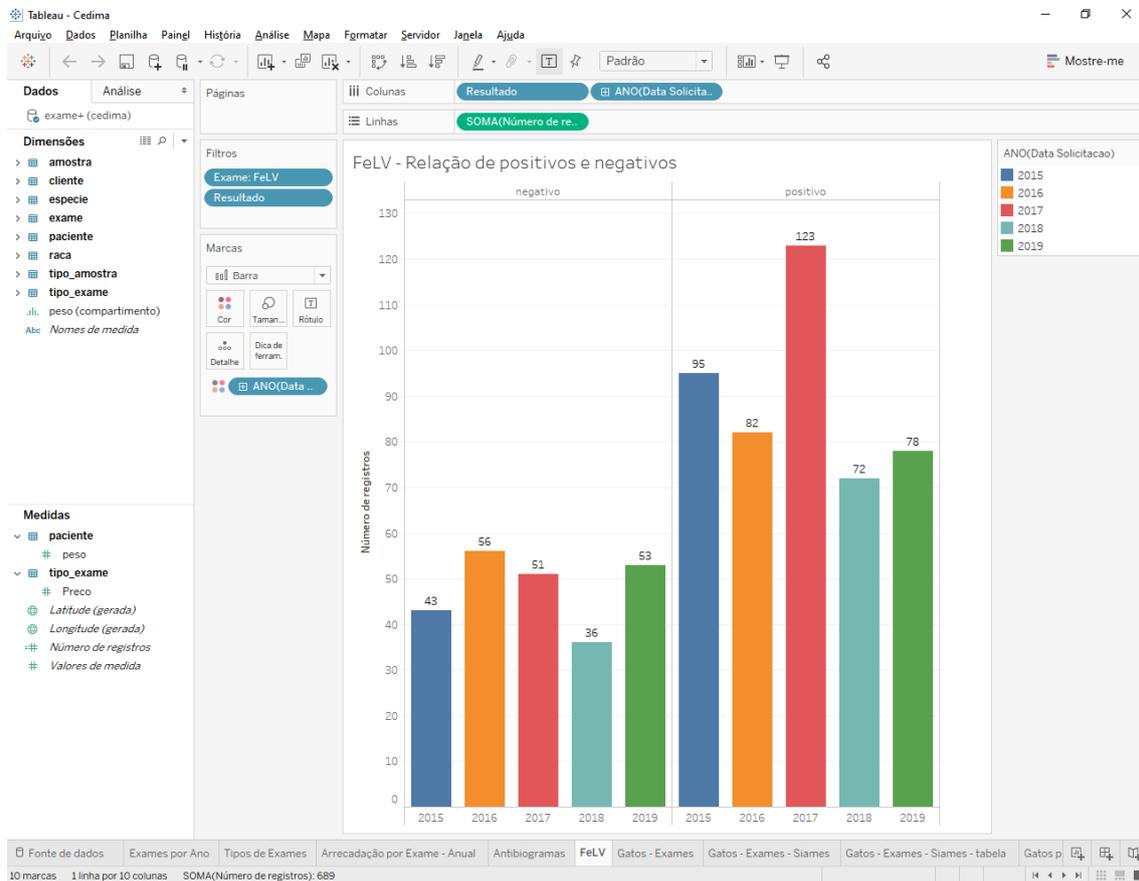
The screenshot shows the Tableau interface for configuring a data connection. The main window is titled 'exame+ (cedima)'. The interface includes a sidebar with 'Conexões', 'Banco de dados', and 'Tabela'. The main area shows a data flow diagram with tables 'exame', 'amostra', 'paciente', 'tipo_exame', 'tipo_amostra', 'cliente', 'especie', and 'raca'. Below the diagram is a table with columns for 'cliente', 'especie', 'exame', and 'amostra'.

cliente	cliente	Abc cliente	# especie	Abc especie	# exame	# exame	# exame	# exame
Data Nascimento (...)	Data Cadastro (CLI...	Observacoes	Id (Especie)	Especie	Id	Id Tipo Exame	Id Paciente	Id Amostra
23/06/1970	11/06/2019	nulo	21	CANINA	23	1	23	
11/09/1984	03/09/2019	nulo	10	FELINA	31	1	31	
05/03/1998	28/02/2017	nulo	41	BOVINA	40	1	40	
13/06/1981	04/06/2017	nulo	21	CANINA	55	1	55	
28/06/1994	22/06/2018	nulo	96	EQUINOS	60	1	60	
13/03/1973	02/03/2019	nulo	1035	PRIMATA	143	1	143	
12/07/1980	04/07/2015	nulo	10	FELINA	194	1	194	
27/01/1998	22/01/2019	nulo	62	CAPRINA	202	1	202	
01/01/1976	21/12/2018	nulo	10	FELINA	211	1	211	
09/05/1974	28/04/2016	nulo	62	CAPRINA	234	1	234	
13/05/1998	08/05/2017	nulo	21	CANINA	246	1	246	
04/03/1989	26/02/2015	nulo	21	CANINA	249	1	249	
21/09/1985	14/09/2015	nulo	951	PSITACIDEO	264	1	264	
05/08/1967	23/07/2016	nulo	62	CAPRINA	265	1	265	

Fonte: o autor (2019)

O mercado oferece uma grande variedade de programas nesta área, indo de soluções gratuitas, como o IBM Watson Analytics, Microsoft Power BI, RapidMiner Studio Free, a pagas, como o Databox, Qlik e outras de código-fonte aberto, como o KNIME, Knowage e Metabase.

Figura 19 - Interface do programa Tableau: geração de gráfico



Fonte: o autor (2019)

Anaconda é uma das plataformas para aplicação de *data science* mais populares do mercado (KADIYALA; KUMAR, 2017). É um produto de distribuição gratuita baseada nas linguagens Python e R, buscando ser um pacote (*stack*) completo de desenvolvimento em Python. Para isto, inclui um interpretador completo e as bibliotecas padrão, assim como os módulos mais comuns utilizados em análise de dados (SHEPPARD, 2019).

Os procedimentos de instalação foram os seguintes:

1. No site www.anaconda.com/distribution estão disponíveis os *links* para *download* das versões para Windows, macOS e Linux. Ao acessar a versão para Linux é disponibilizado dois *links* com a opção de versão 3.7 e 2.7 do Anaconda. Foi utilizada a opção “64-bit (x86)”, que

permite a descarga do arquivo “Anaconda3-2019.10-Linux-x86_64.sh”, que é um *shell script* responsável pela instalação do programa.

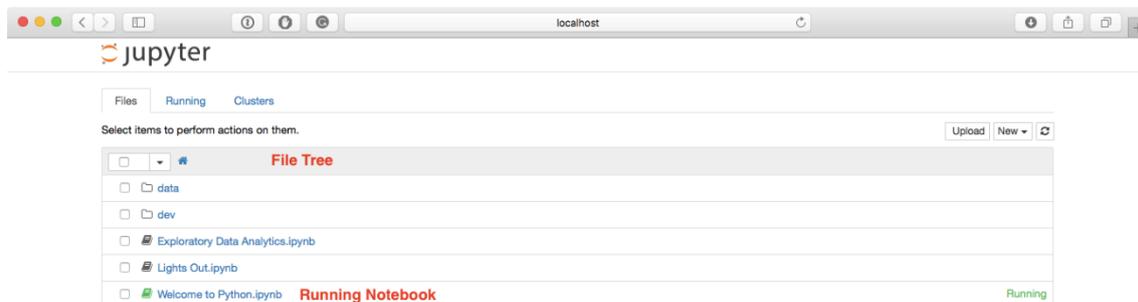
2. No terminal *shell* do Linux o *script* foi tornado executável com o comando “`chmod +x Anaconda3-2019.10-Linux-x86_64.sh`” e em seguida executado com o comando “`sh Anaconda3-2019.10-Linux-x86_64.sh`”, o que dá início ao processo de instalação.

3. Após o aceite da licença de usuário final do programa, um assistente procede com o *download* dos arquivos necessários até o término do processo.

Sobre a plataforma Anaconda, e como interface de desenvolvimento, foi utilizado o Jupyter Notebook (disponível na instalação padrão do sistema Anaconda), um projeto sem fins lucrativos e de código-fonte aberto, nascido do projeto *IPython*, em 2014, com a intenção de dar suporte à aplicações em *data science* e à computação científica, para várias linguagens de programação (PROJECTJUPYTER, 2019).

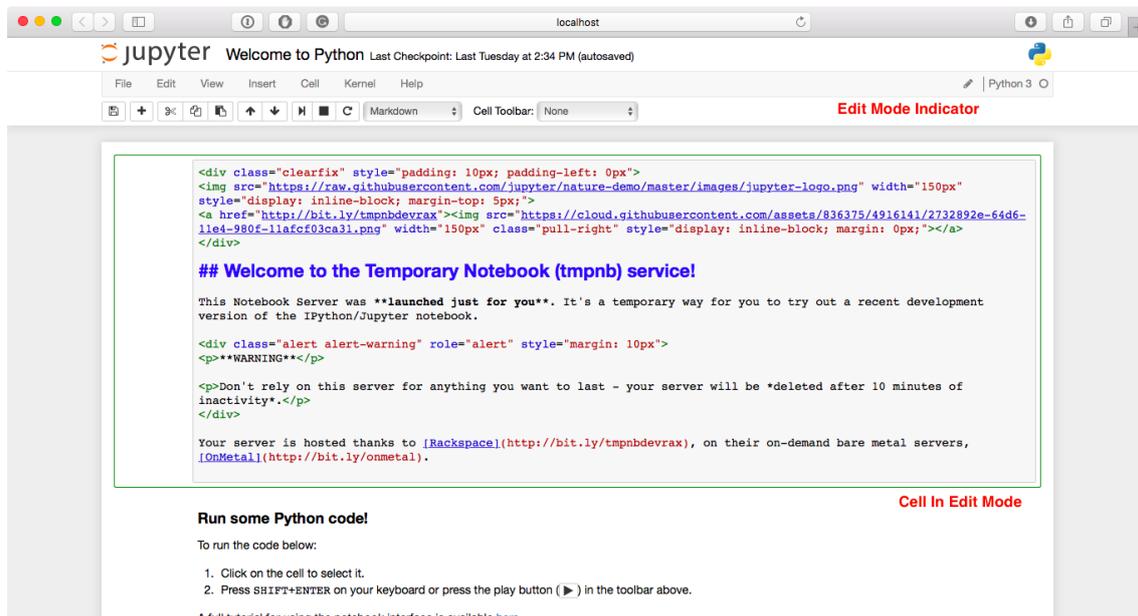
Ele é acessado via navegador *web*, onde disponibiliza um *Notebook* (caderno), que é a sua interface principal, dividida em um painel de controle (*dashboard*) (figura 20) e um editor de texto (figura 21).

Figura 20 - Painel de controle - Jupyter



Fonte: extraído de JupyterTeam (2015)

Figura 21 - Editor de texto – Jupyter



Fonte: extraído de JupyterTeam (2015)

Em seu editor de texto, que é dividido em células, é possível trabalhar com várias linguagens de programação e executá-las diretamente na interface (sendo o Python a padrão), com os resultados sendo mostrados logo abaixo.

Para esta prova de conceito foram criados vários *notebooks*, um para cada laboratório, contendo os seus *scripts* de preenchimento dos bancos de dados e os de análise e exploração de dados.

6.3.4 Python

A linguagem Python foi criada por Guido van Rossum em dezembro de 1989, na Holanda, para servir como um sucessor da linguagem de programação ABC, utilizada no sistema operacional Amoeba, e foi lançada com uma licença própria, bastante flexível, compatível com a licença GPL. Python suporta orientação a objetos, programação estruturada ou procedural, é altamente legível e disponibiliza vários tipos de estruturas de dados de alto nível. É multiplataforma e bastante extensível, através de módulos de terceiro disponibilizados pela comunidade (BASSI, 2017; KADIYALA; KUMAR, 2017; PYTHONBRASIL, 2017).

É bastante utilizada em *Data Science*, com uma comunidade muito grande de cientistas e pesquisadores em sua base de usuários, e a sua adoção só tem crescido desde o início dos anos 2000, incluindo entre aqueles que trabalham com análise, manipulação e visualização de dados, seja na indústria ou na pesquisa. (MCKINNEY, 2013).

Por todo o exposto acima, ser usada nativamente na plataforma Jupyter e Anaconda, ter uma ampla gama de bibliotecas à disposição e farta documentação, foi escolhida para os testes em nossa prova de conceito.

6.4 RESULTADOS DOS TESTES DA PROVA DE CONCEITO

Buscando demonstrar a aplicabilidade da ciência de dados sobre as informações disponíveis nos laboratórios, uma série de gráficos, relatórios e resumos foram gerados, através do *software* Tableau e posteriormente pelo Python, usando das informações das bases de dados de teste.

Para chegar aos resultados, seguimos o processo ou *framework* padrão utilizado na aplicação de *Data Science* em projetos, uma abordagem que ajuda a maximizar as chances de sucesso e a diminuir custos, apesar de não ser uma solução generalizável para todos os problemas (CIELEN; MEYSMAN; ALI, 2016). Foi usada a versão típica que consiste em seis passos, descritos a seguir, formalizados segundo Schutt e O'Neil (2013) e Cielen, Meysman e Ali (2016):

- 1) Meta da pesquisa: a função deste passo é definir quais as questões que se buscam responder com o projeto, além de como e por quê.

Normalmente a pesquisa se foca nas questões que a empresa ou organização precisa equacionar para resolver um problema ou para lançar um novo produto, por exemplo. “Quem são nossos clientes”, “porque compram nossos produtos ou serviços”, “como prever se eles serão adquiridos” e similares.

- 2) Recuperação dos dados: permite que os dados estejam disponíveis para análise, o que inclui encontrar as informações necessárias e conseguir acesso através do seu proprietário. Normalmente estão em formato bruto e algumas vezes desconhecido, e é preciso que sejam tratados antes de serem aproveitáveis.

- 3) Preparação: inclui transformar de um formato cru (*raw*) em outro que possa ser usado pelos modelos. É preciso corrigir os diferentes tipos de erros nos dados, combinando-os de várias fontes, em alguns casos.

Em relação aos itens 2 e 4, os dados inseridos nos bancos de dados foram previamente desenhados para corresponder ao melhor cenário possível de utilização de um programa de gerenciamento de informações no laboratório, e por isso estão totalmente acessíveis, em um formato ideal para o uso e sem erros. Normalmente é preciso corrigir erros de digitação, espaços em branco redundantes, valores faltantes, atípicos e etc. (CIELEN; MEYSMAN; ALI, 2016).

- 4) Exploração: é neste passo que este capítulo foca, buscando padrões, correlações ou desvios, usando técnicas visuais ou descritivas, como gráficos ou tabelas, para um maior entendimento dos dados e a interrelação entre as variáveis. No caso da prova de conceito, entende-se que esta se mostra suficiente para comprovar os pontos levantados anteriormente.
- 5) Modelagem ou análise em profundidade: esta fase usa técnicas importadas da estatística, aprendizado de máquina (*machine learning*) e mineração de dados (*data mining*). A construção de modelos e a sua comparação permite alcançar o diagnóstico avançado e a previsão que se busca.
- 6) Apresentação dos resultados: neste momento tudo o que foi descoberto precisa ser apresentado aos interessados, e é preciso que a execução dos modelos e análises seja automatizada, visto que os *insights* encontrados serão explicados várias vezes para uma série de audiências diferentes. Materiais como planilhas, apresentações, painéis ou *dashboards* estão entre os mais usados.

Os resultados dos testes com os programas citados são apresentados logo abaixo, com as explicações de possíveis cenários de uso em situações reais, caso possível.

6.4.1 Cedima

Os gráficos e tabelas a seguir são o resultado do cruzamento das informações constantes na base de dados de testes do Cedima. Para tal foram inseridos dez mil registros nas tabelas de clientes, pacientes, amostras e exames e cada um se propõe a ser uma fonte de informação que possa gerar ativos de conhecimento, ou *insights*, segundo a visão do *Data Science*.

Inicialmente, a tabela 2 sumariza a quantidade de exames por ano realizados no laboratório, a partir de 2015, através do Tableau. A tabela 3 filtra os resultados por tipo de exame, criado pelo Jupyter.

Tabela 2 - Cedima: Total de exames por ano - Tableau

2015	2.002
2016	2.019
2017	2.005
2018	1.933
2019	2.034
2020	7
Total geral	10.000

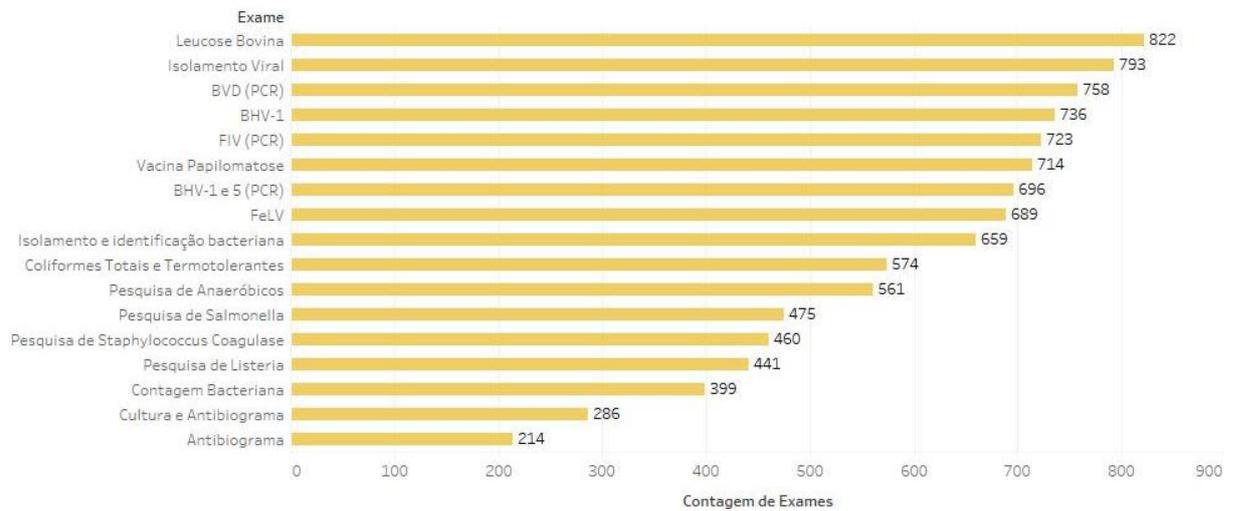
Fonte: dados da pesquisa (2019)

Tabela 3 - Total de exames por tipo - Python

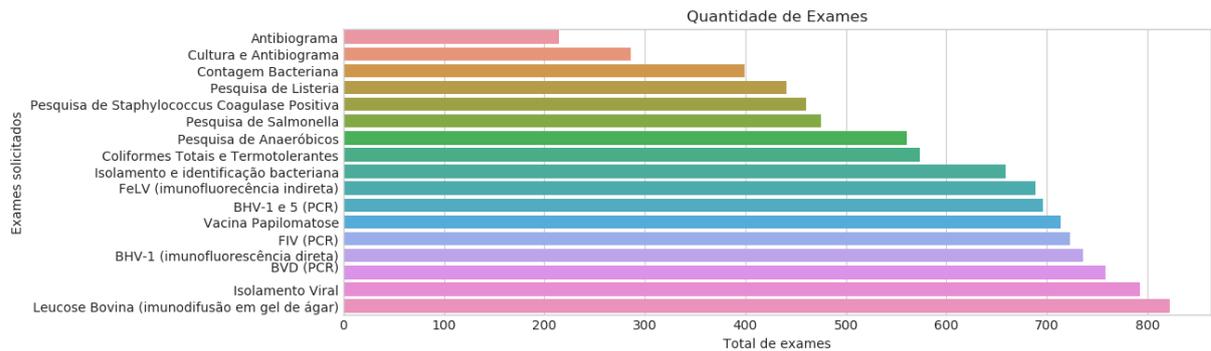
Exame	Total
Antibiograma	214
Cultura e Antibiograma	286
Contagem Bacteriana	399
Pesquisa de Listeria	441
Pesquisa de Staphylococcus Coagulase Positiva	480
Pesquisa de Salmonella	475
Pesquisa de Anaeróbicos	581
Coliformes Totais e Termotolerantes	574
Isolamento e identificação bacteriana	659
FeLV (imunofluorescência indireta)	689
BHV-1 e 5 (PCR)	696
Vacina Papilomatose	714
FIV (PCR)	723
BHV-1 (imunofluorescência direta)	736
BVD (PCR)	758
Isolamento Viral	793
Leucose Bovina (imunodifusão em gel de ágar)	822

Fonte: dados da pesquisa (2019)

O gráfico da figura 22, gerado pelo Tableau, demonstra os mesmos resultados de maneira visual, enquanto que o de número 23, criado pelo Jupyter, também o faz. Todos os dados são contados a partir de 2015, como sendo o ano com o primeiro registro.

Figura 22 - Cedima: quantidade de exames, por tipo - Tableau

Fonte: dados da pesquisa (2019)

Figura 23 - Cedima: quantidade de exames, por tipo - Python

Fonte: dados da pesquisa (2019)

As tabelas abaixo resumem todos os agentes patogênicos identificados como positivos pelo laboratório, por ano, de maneira geral (tabela 4), somente as bactérias (tabela 5) e somente os vírus (tabela 6).

Os nomes dos micro-organismos utilizados são reais, enquanto as quantidades foram geradas aleatoriamente, servindo para a demonstração da prova de conceito.

Tabela 4 - Cedima: Positivos por ano - Tableau

complemento	Ano de Data Solicitacao				
	2015	2016	2017	2018	2019
Aphthovirus	11	13	5	3	3
Avian avulavirus	3	7	3	13	6
Avihepatovirus	4	10	8	6	8
Canarypox	10	11	11	12	16
Canine parvovirus	12	12	9	7	5
Cardiovirus	9	4	5	4	5
Chlamydia psittaci	4	4	6	6	5
Clostridium botulinum	10	4	5	8	6
Clostridium chavoiei	3	2	3	1	3
Clostridium piliforme	1	4	3	2	2
Clostridium tetani	4	7	8	6	9
Erbovirus	4	6	5	4	6
Escherichia coli		1			2
Herpesvirus Suídeo	8	8	8	7	7
Histophilus somni	5	6	5	5	2
Kobuvirus	9	9	5	8	6
Leptospira interrogans	10	15	7	9	10
Lyssavirus	11	12	12	11	19
Parechovirus	1	4	4	4	6
Pasteurella haemolytica	8	3	5	2	8
Pasteurella multocida	2	5	5	6	6
Porcine circovirus	18	13	9	11	8
Rickettsia rickettsii	13	7	7	3	5
Salmonella bongori	8	5	9	3	7
Salmonella enterica	2	6	2	3	6
Sapelovirus	3	2		2	
Senecavirus	4	3		2	1
Staphylococcus aureus	1	1			1
Streptococcus pneumoniae	6	7	3	9	9
Teschovirus	2	1	2		
Tremovirus			1	2	3
Vibrio cholerae	13	12	10	11	14
Total geral	199	204	165	170	194

Fonte: dados da pesquisa (2019)

Tabela 5 - Cedima: Bactérias identificadas - Tableau

	2015	2016	2017	2018	2019
<i>Leptospira interrogans</i>	10	15	7	9	10
<i>Vibrio cholerae</i>	13	12	10	11	14
<i>Salmonella enterica</i>	2	6	2	3	6
<i>Clostridium tetani</i>	4	7	8	6	9
<i>Streptococcus pneumoniae</i>	6	7	3	9	9
<i>Rickettsia rickettsii</i>	13	7	7	3	5
<i>Chlamydia psittaci</i>	4	4	6	6	5
<i>Clostridium botulinum</i>	10	4	5	8	6
<i>Salmonella bongori</i>	8	5	9	3	7
<i>Histophilus somni</i>	5	6	5	5	2
<i>Pasteurella multocida</i>	2	5	5	6	6
<i>Clostridium piliforme</i>	1	4	3	2	2
<i>Pasteurella haemolytica</i>	8	3	5	2	8
<i>Clostridium chavoiei</i>	3	2	3	1	3
<i>Staphylococcus aureus</i>	1	1			1
<i>Escherichia coli</i>		1			2
Total geral	132	132	125	125	145

Fonte: dados da pesquisa (2019)

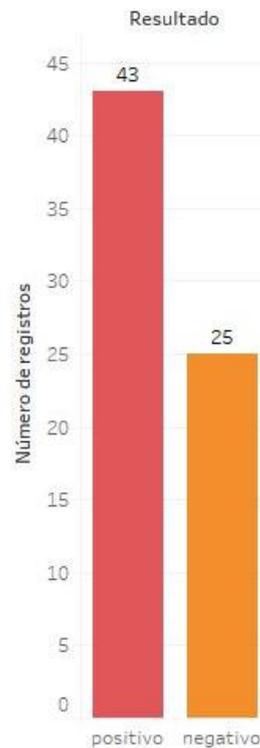
Tabela 6 - Cedima: Vírus identificados – Tableau

	2015	2016	2017	2018	2019
Canarypox	10	11	11	12	16
Herpesvírus Suídeo	8	8	8	7	7
Aphthovirus	11	13	5	3	3
Canine parvovirus	12	12	9	7	5
Lyssavirus	11	12	12	11	19
Porcine circovirus	18	13	9	11	8
Avihepatovirus	4	10	8	6	8
Kobuvirus	9	9	5	8	6
Erbovirus	4	6	5	4	6
Avian avulavirus	3	7	3	13	6
Cardiovirus	9	4	5	4	5
Parechovirus	1	4	4	4	6
Sapelovirus	3	2		2	
Senecavirus	4	3		2	1
Tremovirus			1	2	3
Teschovirus	2	1	2		
Total geral	168	165	134	162	164

Fonte: dados da pesquisa (2019)

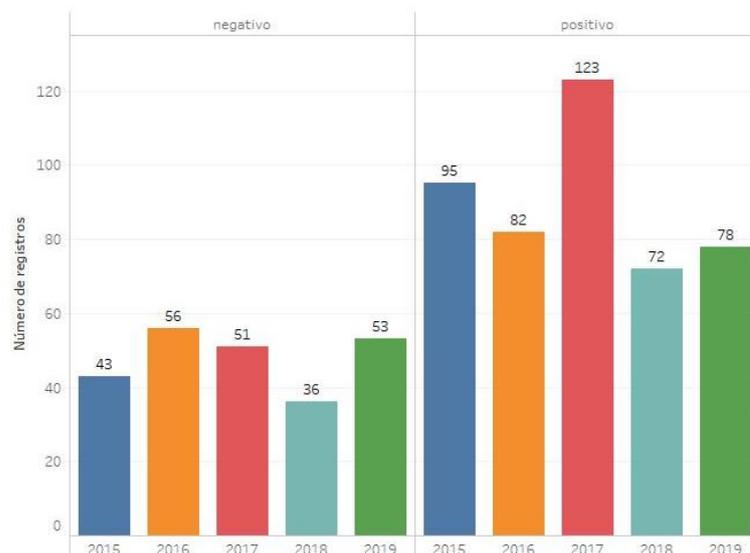
Em relação às estatísticas de resultados de exames, o gráfico da figura 24 demonstra as relacionadas a testes de antibiogramas e o seguinte, figura 25, aprofunda os detalhes, sobre a FeLV (Leucemia Felina), comparando positivos e negativos por ano.

Figura 24 - Cedima: resultados para antibiograma - Tableau



Fonte: dados da pesquisa (2019)

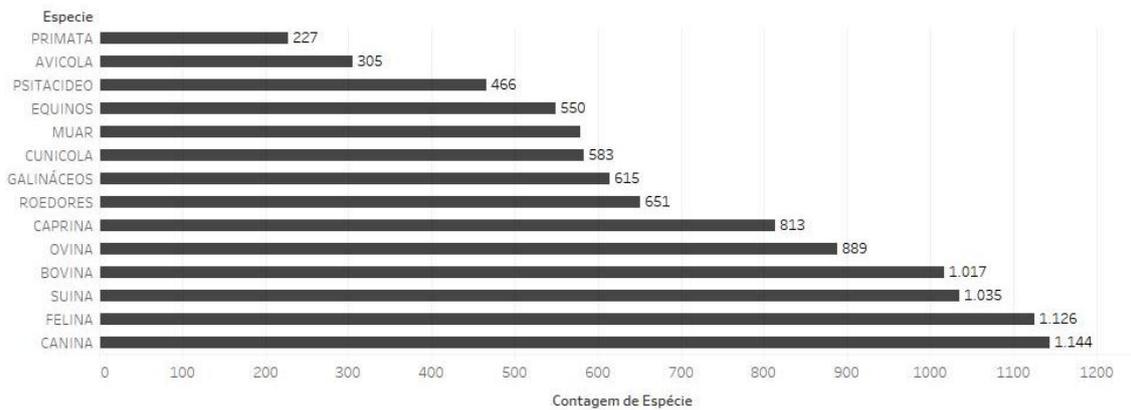
Figura 25 - Cedima: relação de resultados para FeLV - Tableau



Fonte: dados da pesquisa (2019)

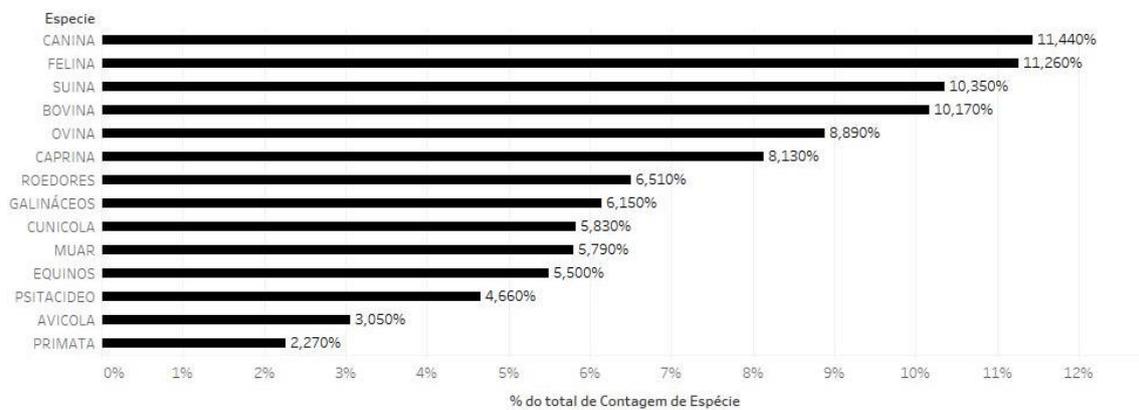
A figura 26 mostra um exemplo que resume as espécies de animais mais comuns nos atendimentos do laboratório e a figura 27 o faz em porcentagens.

Figura 26 - Cedima: relação de espécies mais atendidas – Tableau



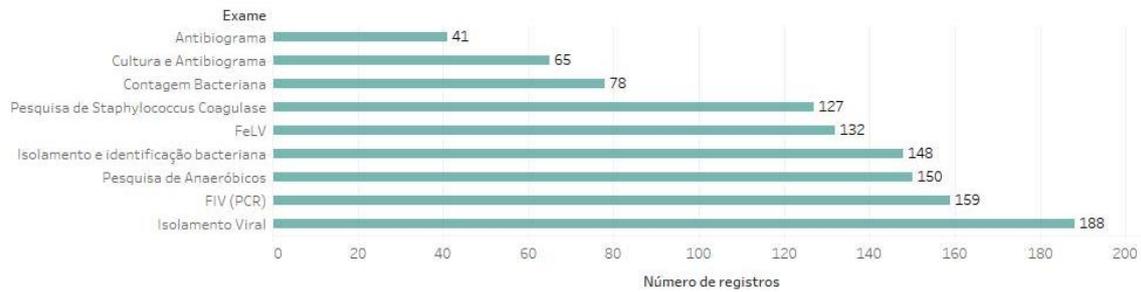
Fonte: dados da pesquisa (2019)

Figura 27 - Cedima: porcentagem de espécies mais atendidas – Tableau

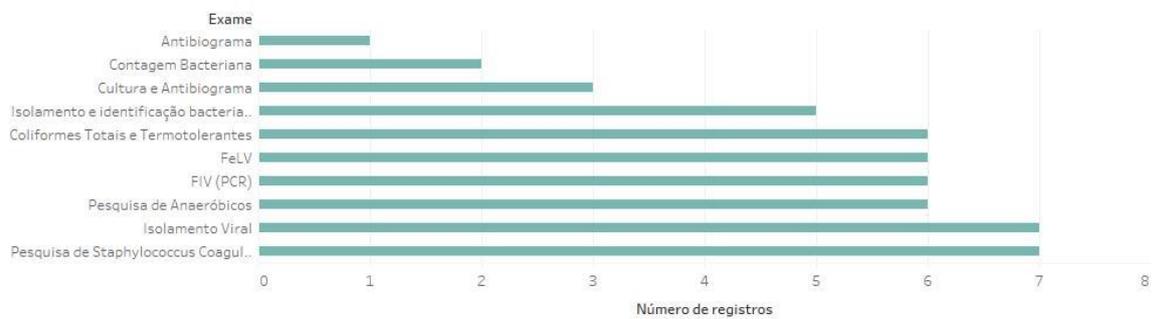


Fonte: dados da pesquisa (2019)

Na sequência, a partir do gráfico de barras da figura de número 28, a intenção é demonstrar as capacidades de filtragem na criação de informações a partir da base. Neste exemplo, ele sumariza somente os exames que foram realizados em gatos, de quaisquer raças. O seguinte (figura 29) é gerado somente com os dados de gatos siameses e a tabela 7 sumariza o total de gatos por raça.

Figura 28 - Cedima: totais de exames para gatos - Tableau

Fonte: dados da pesquisa (2019)

Figura 29 - Cedima: relação de exames mais comuns em gatos siameses – Tableau

Fonte: dados da pesquisa (2019)

Tabela 7 – Cedima: totais de raças de gatos - Tableau

Raca	Total
KHAO MANEE	84
CHARTREUX	83
AZUL RUSSO	83
PERSA	82
AMERICAN SHORTHAIR	81
SIAMES	80
PUMA CONDOR	75
GATO DO MATO	75
SRD	74
GATO MARACAJA	74
HIMALAIA	73
BIRMA	73
HAVANA	70
MAINE COON	64
RAGDOLL	55
Total geral	1.126

Fonte: dados da pesquisa (2019)

Da mesma maneira, é possível resumir, por raças dos felinos, todos os resultados positivos e negativos e seus totais gerais (tabela 8).

Tabela 8 – Cedima: Resultados de exames por raça de gatos – Tableau

Raça	Resultado		Total geral
	negativo	positivo	
AMERICAN SHORTHAIR	22	59	81
AZUL RUSSO	29	54	83
BIRMA	22	51	73
CHARTREUX	31	52	83
GATO DO MATO	24	51	75
GATO MARACAJA	26	48	74
HAVANA	25	45	70
HIMALAIA	25	48	73
KHAO MANEE	33	51	84
MAINE COON	19	45	64
PERSA	27	55	82
PUMA CONDOR	29	46	75
RAGDOLL	13	42	55
SIAMES	32	48	80
SRD	27	47	74
Total geral	384	742	1.126

Fonte: dados da pesquisa (2019)

A partir da tabela 9 demonstra-se como os filtros podem ser cada vez mais específicos. Por exemplo: o pesquisador pode precisar de uma listagem com a quantidade de resultados positivos de exames realizados em gatos sem raça definida. Na tabela 10, segue-se a mesma premissa, mas somente para gatos persa. A tabela 11 separa os resultados por ano para siameses e a 12 para persas. Os registros contam sempre a partir de 2015.

Tabela 9 - Cedima: totais de positivos para gatos SRD por agente patogênico - Tableau

Raça	complemento	Resultado
		positivo
SRD	Avian avulavirus	2
	Canarypox	1
	Kobuvirus	1
	Pasteurella haemolytica	1
	Porcine circovirus	1
	Rickettsia rickettsii	1

Fonte: dados da pesquisa (2019)

Tabela 10 - Cedima: totais de positivos para gatos persa por agente patogênico – Tableau

Raca	complemento	Resultado
		positivo
PERSA	Aphthovirus	1
	Chlamydia psittaci	2
	Rickettsia rickettsii	1

Fonte: dados da pesquisa (2019)

Tabela 11 - Cedima: totais de positivos para gatos siameses por agente patogênico, por ano e data de nascimento - Tableau

Raca	complemento	Ano de Data Nascimento	Ano de Data Solicitacao	Resultado
				positivo
SIAMES	Aphthovirus	2007	2016	1
		2008	2016	1
	Canarypox	2010	2019	1
	Clostridium tetani	2008	2018	1
	Lyssavirus	2012	2016	1
	Parechovirus	2016	2017	1
	Staphylococcus aureus	2013	2016	1

Fonte: dados da pesquisa (2019)

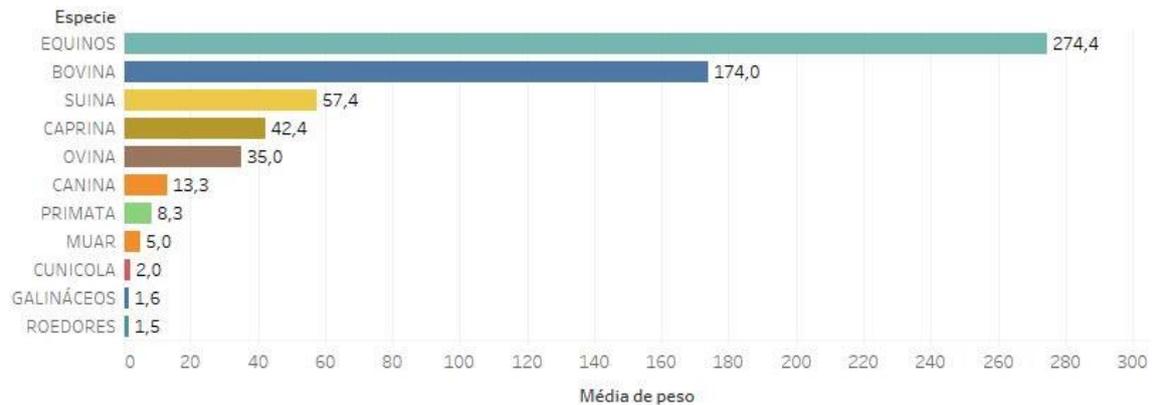
Tabela 12 - Cedima: totais de positivos para gatos persa por agente patogênico, por ano e data de nascimento - Tableau

Raca	complemento	Ano de Data Solicitacao	Ano de Data Nascimento	Resultado
				positivo
PERSA	Aphthovirus	2016	2010	1
	Chlamydia psittaci	2017	2005	1
			2012	1
	Rickettsia rickettsii	2015	2014	1

Fonte: dados da pesquisa (2019)

Com os dados dos pacientes, como o peso, por exemplo, é possível conseguir várias informações, como a média geral por espécie, por ano (figura 30), a média somente dos felinos, por ano (tabela 13) e a sua média de peso, por ano, separada por raça (tabela 14).

Figura 30 - Cedima: média de peso por espécie - Tableau



Fonte: dados da pesquisa (2019)

Tabela 13 - Cedima: média de peso dos gatos, por ano - Tableau

2015	4,207
2016	4,037
2017	4,128
2018	4,117
2019	4,195
2020	6,570

Fonte: dados da pesquisa (2019)

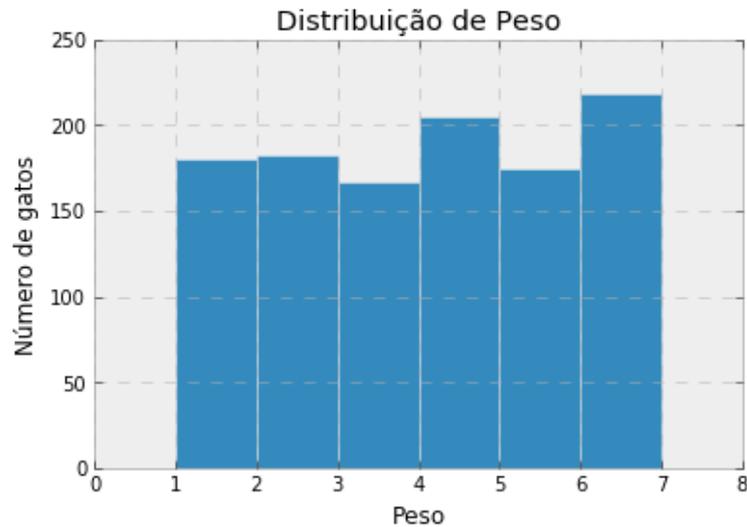
Tabela 14 - Cedima: média de peso dos gatos, por raça e por ano - Tableau

	Raca														
	AMERICAN SHORTHAIR	AZUL RUSSO	BIRMA	CHARTREUX	GATO DO MATO	GATO MARACAJA	HAVANA	HIMALAIA	KHAO MANEE	MAINE COON	PERSA	PUMA CONDOR	RAGDOLL	SIAMES	SRD
2015	3,841	4,108	3,710	4,543	4,020	4,372	4,724	4,547	4,351	4,511	3,958	3,790	4,661	4,341	3,879
2016	4,457	3,558	3,998	3,605	3,499	4,022	3,609	3,753	3,974	4,692	4,403	4,264	4,625	4,058	4,524
2017	4,107	3,132	3,966	4,017	3,961	4,628	3,860	3,973	4,240	4,045	4,409	4,468	4,495	4,825	3,421
2018	4,038	3,708	3,806	4,365	4,573	4,213	3,882	3,812	4,032	4,184	4,366	4,146	3,984	4,274	4,136
2019	3,877	3,926	4,359	4,756	4,749	4,239	4,139	3,596	4,071	3,679	3,725	4,439	4,746	4,271	4,871

Fonte: dados da pesquisa (2019)

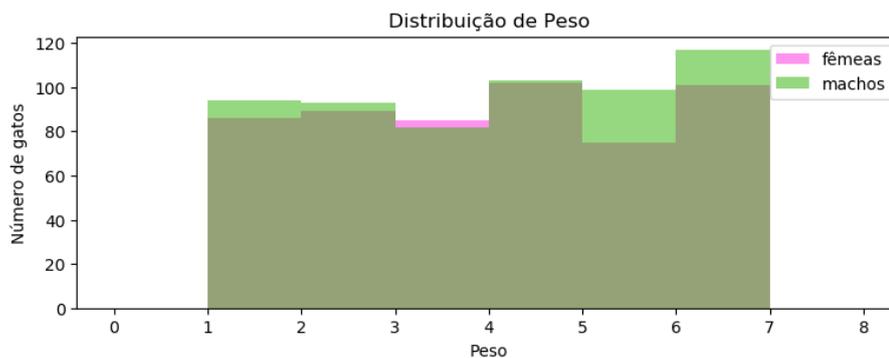
Nesta seção, busca-se demonstrar as capacidades da linguagem Python, juntamente com a plataforma Jupyter, para a criação de gráficos de barra. A figura 31 demonstra a distribuição de peso dos felinos, enquanto a 32 o faz comparando os sexos de cada animal.

Figura 31 - Cedima: distribuição de peso dos gatos - Python



Fonte: dados da pesquisa (2019)

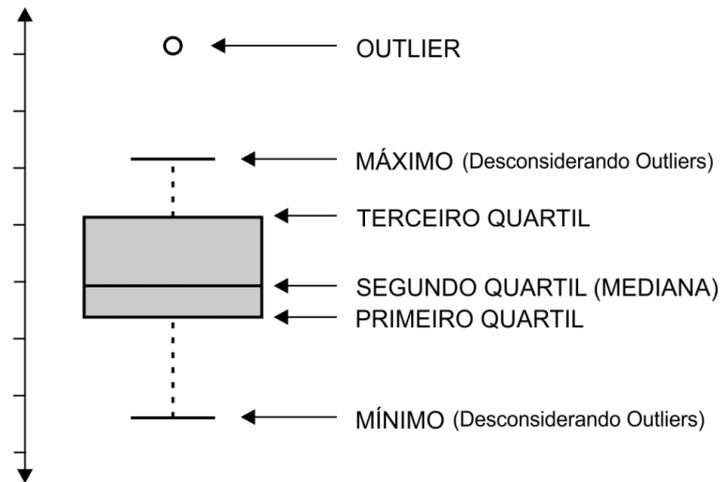
Figura 32 - Cedima: distribuição de peso dos gatos, por sexo - Python



Fonte: dados da pesquisa (2019)

Com a plataforma Jupyter é possível a criação de gráficos do tipo *boxplot*, também conhecido como diagrama de caixa ou diagrama esquemático, que sumariza e interpreta dados visualmente, auxiliando na representação da variação numérica de um conjunto de dados por meio de quartis e seus valores atípicos, ou ocorrências incomuns (KENDRICK, 1989). Os componentes de um diagrama de caixa estão ilustrados na figura 33.

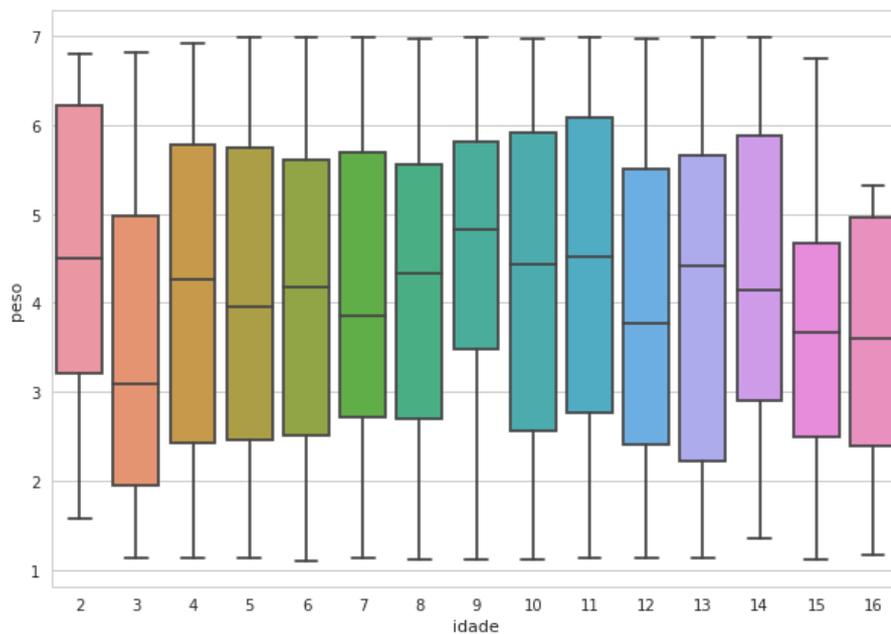
Figura 33 - Componentes básicos de um gráfico *boxplot*



Fonte: Oliveira (2019)

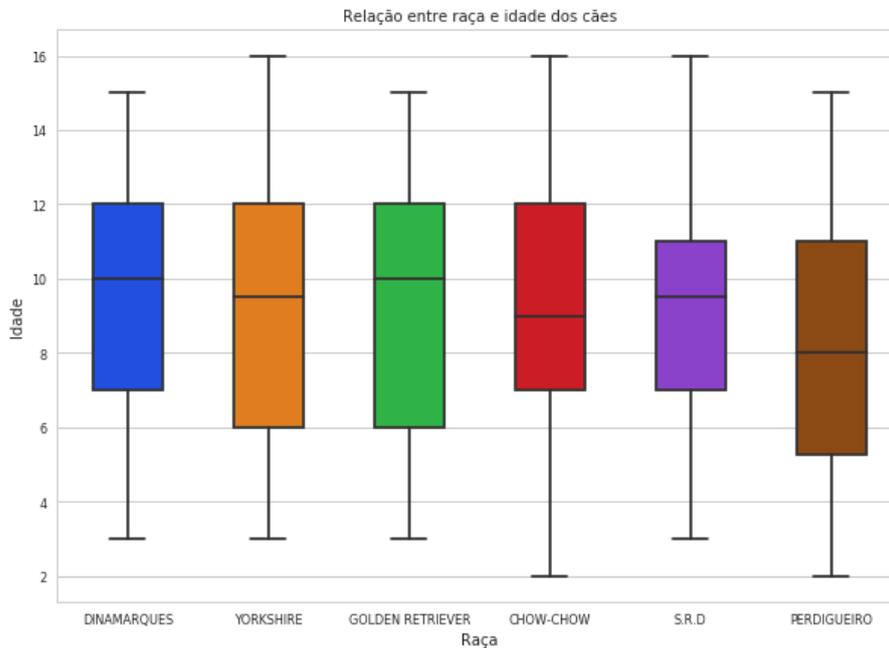
Seguem exemplos de gráficos *boxplot* que usam os valores de peso dos animais para a geração de informações, nas figuras 34 e 35.

Figura 34 - Cedima: distribuição de peso dos gatos, por idade - Python



Fonte: dados da pesquisa (2019)

Figura 35 - Cedima: distribuição de idade dos cães, por raça - Python

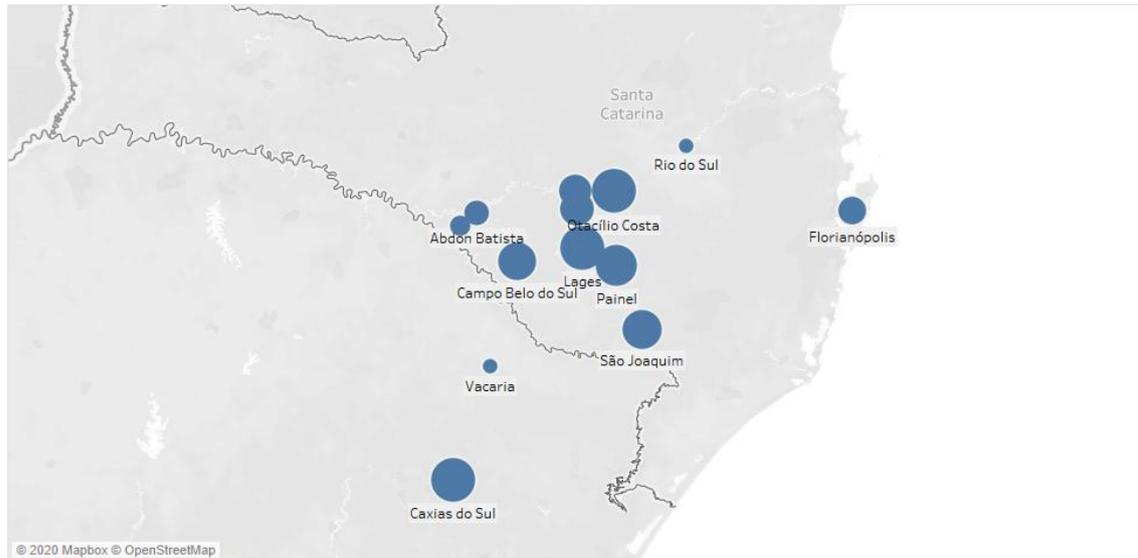


Fonte: dados da pesquisa (2019)

Tanto o Tableau quanto o Jupyter oferecem a possibilidade de gerar mapas com marcações geográficas, utilizando as informações correlatas do banco de dados, como endereço, bairro ou nome da cidade do cliente ou paciente.

O gráfico constante da figura 36 é criado automaticamente, pelo Tableau, baseado na contagem de cidades de origem dos clientes. O tamanho do indicador da cidade corresponde à sua quantidade relativa de clientes.

Figura 36 - Cedima: distribuição geográfica de clientes - Tableau



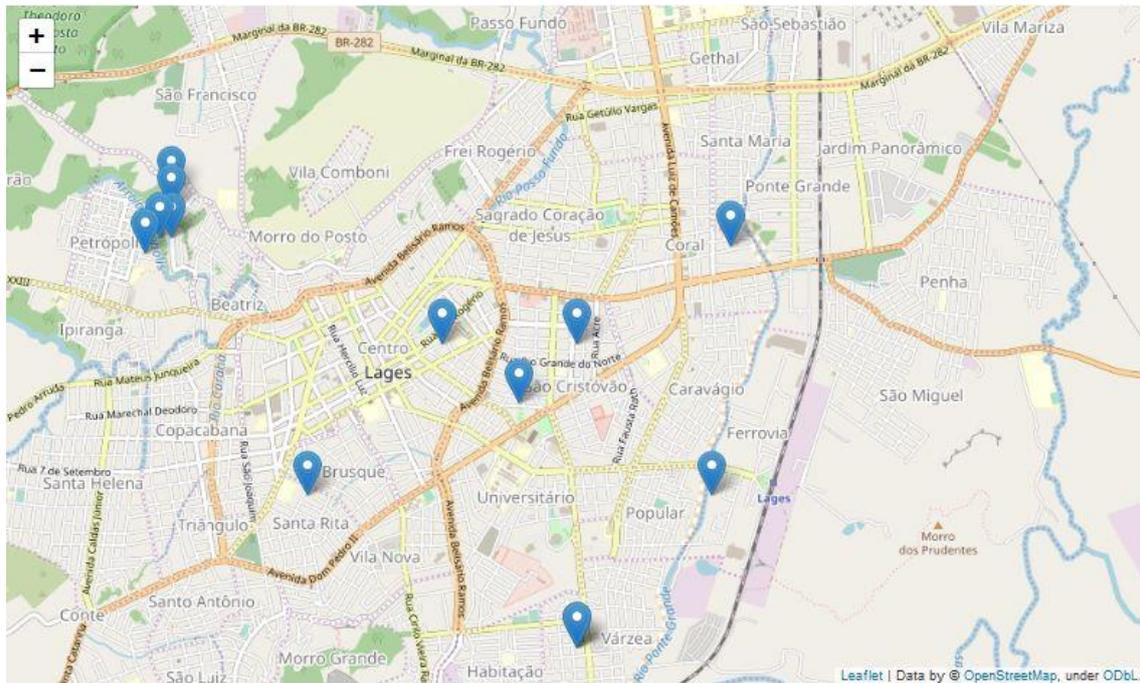
Fonte: dados da pesquisa (2019)

A linguagem Python pode usar a biblioteca folium (FOLIUM, 2019) para isto, se baseando em algum dado básico, como o nome da cidade ou o endereço para conseguir a latitude e longitude de uma marcação. No caso desta prova de conceito, seria necessária a utilização de uma API (*Application Programming Interface*) de um serviço externo, como por exemplo o *Google Maps*, que é pago, ou o *OpenStreetmap*, por exemplo, que oferece o serviço de geocodificação (conversão de nome de cidade ou endereço em longitude e latitude) gratuitamente.

Para os exemplos abaixo feitos com o Python, as localizações foram adicionadas ao *script* manualmente, para ilustrar o resultado final, que seria o mesmo.

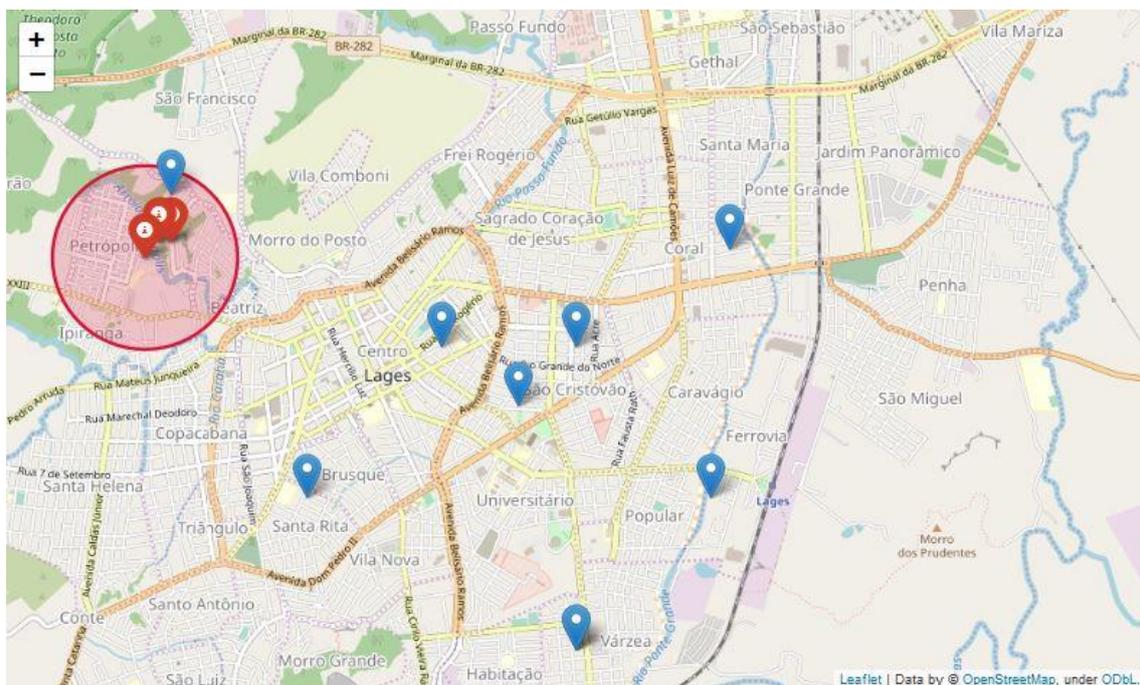
Eles retratam como seriam assinaladas, em um mapa, a localização de animais infectados com um determinado patógeno (figura 37). A figura 38 é um exemplo de marcação automática de casos simultâneos com uma distância mínima entre si, gerando um alarme ou aviso útil em pesquisa ou para o estudo da infectologia.

Figura 37 - Cedima: localização geográfica de infecções por um agente específico - Python



Fonte: dados da pesquisa (2019)

Figura 38 - Cedima: localização geográfica de infecções por um agente específico, com destaque - Python

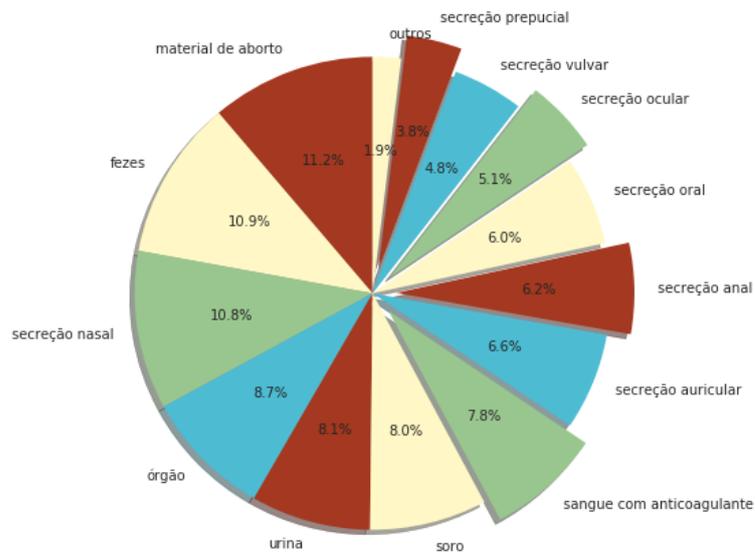


Fonte: dados da pesquisa (2019)

Outras possibilidades incluem o uso de informações de precipitação de chuvas e de variação de temperaturas em determinadas épocas do ano, em comparação com os números de positivos para certos agentes.

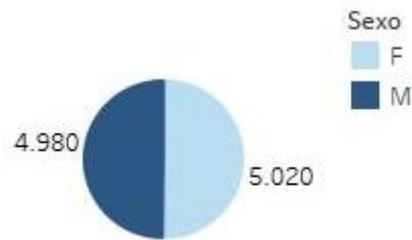
Ainda no Jupyter, a figura 39 é um gráfico do tipo pizza, ilustrando a divisão dos tipos de materiais que o laboratório recebe como fonte de seus exames, criado pelo Python.

Figura 39 - Cedima: distribuição de materiais recebidos para análise - Python



Fonte: dados da pesquisa (2019)

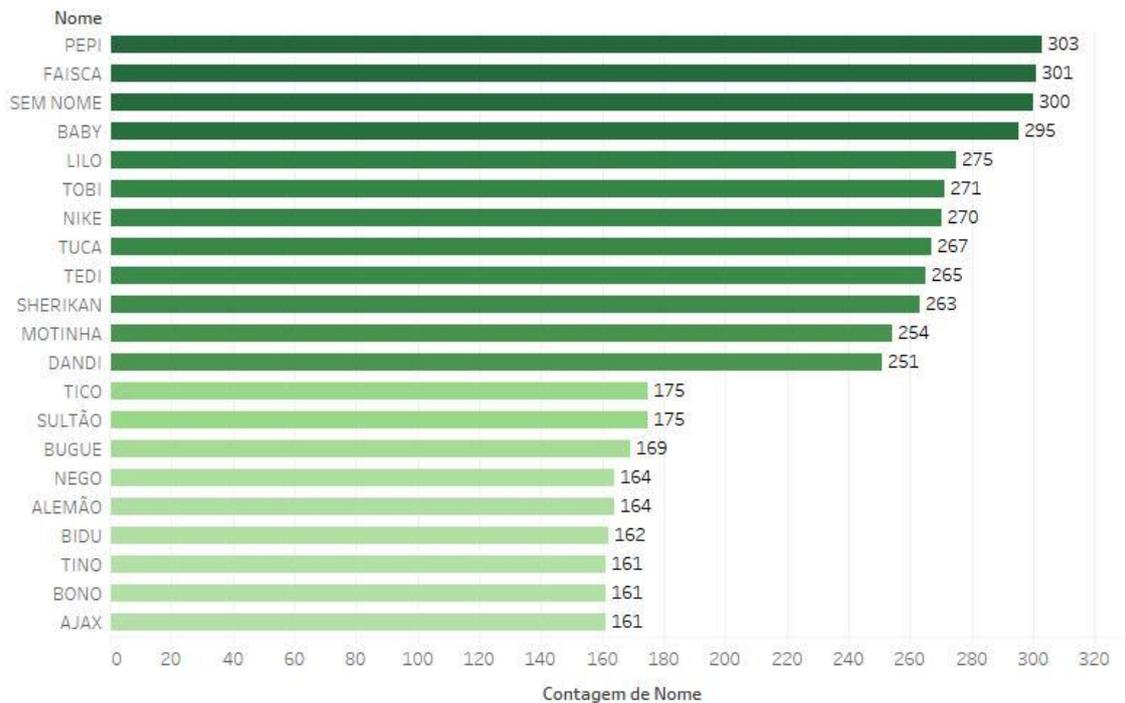
Alguns resumos podem ser conseguidos facilmente, tal como o sexo predominante entre os pacientes (figura 40) e quais os seus nomes mais comuns (figura 41). Resta ao pesquisador interessado decidir se tais dados são úteis ou não em pesquisas ou na tomada de decisão.

Figura 40 - Cedima: distribuição de pacientes por sexo – Tableau

Fonte: dados da pesquisa (2019)

Figura 41 - Cedima: nomes de pacientes mais comuns - Tableau

Nomes de Animais

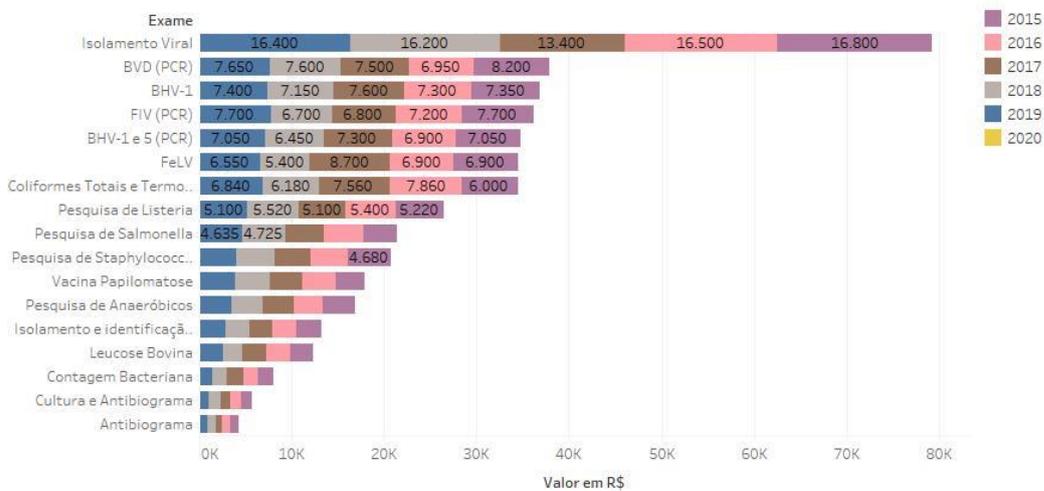


Fonte: dados da pesquisa (2019)

E, apesar dos laboratórios do campus não visarem o lucro em suas atividades, usando a frequência dos exames e o seu valor unitário, a figura 42 demonstra a arrecadação anual por tipo de exame.

Figura 42 - Cedima: Arrecadação anual por tipo de exame - Tableau

Arrecadação por Exame - Anual



Fonte: dados da pesquisa (2019)

Uma das vantagens do uso de uma linguagem de programação como o Python sobre ferramentas automatizadas de análise de dados é a customização. Através do código-fonte de programação pode-se chegar a um controle fino na criação inclusive de relatórios, na forma de tabelas e quadros. O quadro 13 é resultado de uma pesquisa na base de dados do Cedima por gatos com resultados positivos para “*Streptococcus pneumoniae*”, retornando vários campos como nome, sexo, idade (calculado a partir da data de nascimento), raça e etc.

Quadro 13 - Cedima: listagem de gatos com positivo para *Streptococcus pneumoniae* - Python

Nome	Sexo	Idade	Peso	Resultado	Complemento	Espécie	Raça	Material	Nome Cliente	Cidade	Data Exame
PITT	F	6	2.79	positivo	Streptococcus pneumoniae	FELINA	GATO MARACAJA	secreção oral	Francisco Monteiro	Campo Belo do Sul	2019-05-03
SCUBI	M	9	6.91	positivo	Streptococcus pneumoniae	FELINA	MAINE COON	secreção vulvar	Stella Rezende	Lages	2018-10-08
SNOOPY	F	8	2.61	positivo	Streptococcus pneumoniae	FELINA	GATO DO MATO	urina	Henrique da Costa	São Joaquim	2018-07-23

Fonte: dados da pesquisa (2019)

Para se conseguir este resultado lança-se mão tanto das capacidades da linguagem SQL, usada para as consultas nos bancos de dados quanto da biblioteca pandas (PANDAS, 2019), extensamente usada pelo Python para manipulação de grandes massas de dados em projetos de *Data Science*. Ela introduziu o conceito de *dataframes* ao Python (um tipo de tabela de dados

mantida em memória), semelhante ao existente no ambiente de computação estatística conhecido como R (CIELEN; MEYSMAN; ALI, 2016)

Em cada exemplo abaixo uma técnica foi usada, procurando demonstrar as capacidades de cada técnica. Os códigos correspondentes a cada um estão disponíveis nos apêndices.

O quadro 14 lista cães com resultados positivos para qualquer agente juntamente com os campos de interesse. Espera-se que através de relatórios deste estilo um pesquisador possa comprovar uma tese que empiricamente deduz, ou que o resultado o leve a entender, ou se aprofundar na pesquisa de um assunto que ainda não observava.

Quadro 14 - Cedima: listagem de cães positivos para qualquer agente - Python

Nome	Sexo	Data Nascimento	Idade	Peso	Resultado	Complemento	Espécie	Raça	Material	Nome Cliente	Cidade	Data Exame
KIKA	F	2016-01-22	3	8.8	positivo	Porcine circovirus	CANINA	URRADOR AMERICANO	órgão	Vitor Gabriel Silva	Painel	2019-12-10
BOBI	F	2015-10-23	4	17.5	positivo	Avian avulavirus	CANINA	PASTOR DA MANTIQUEIRA	outros	Alana da Cruz	Otacílio Costa	2017-02-24
PEQUENA	F	2016-02-28	3	4.67	positivo	Salmonella bongori	CANINA	OLD ENGLISH SHEEP DOG	secreção ocular	Vitória Fernandes	Otacílio Costa	2015-09-22
JUPIRA	F	2015-11-06	4	16.9	positivo	Avian avulavirus	CANINA	HUSKY SIBERIANO	secreção oral	Olivia Cardoso	São Joaquim	2016-08-10
SAPECA	F	2016-04-23	3	18.46	positivo	Lyssavirus	CANINA	GALGO	secreção vulvar	Vitor Gabriel da Cunha	Capão Alto	2019-07-02
BELLA	F	2017-08-05	2	12.97	positivo	Tremovirus	CANINA	HUSKY SIBERIANO	secreção vulvar	Levi da Mota	São Joaquim	2019-03-30
SEM NOME	F	2016-10-13	3	13.65	positivo	Canarypox	CANINA	GOLDEN RETRIEVER	secreção vulvar	Theo da Rosa	Ponte Alta	2018-04-02
LILO	F	2015-05-21	4	10.28	positivo	Salmonella bongori	CANINA	PERDIGUEIRO	soro	Gabrielly Pires	Ponte Alta	2015-06-01

Fonte: dados da pesquisa (2019)

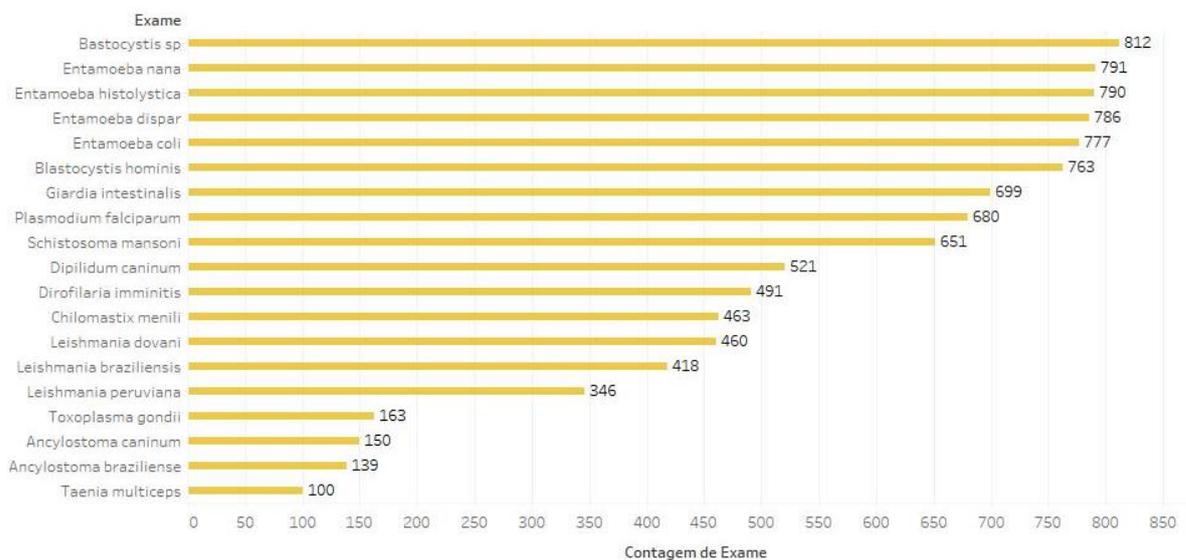
6.4.2 Parasitologia

O laboratório Parasitologia se assemelha muito, no formato de tratamento das informações, ao Cedima. Pode-se notar pelos projetos de banco de dados em que compartilham de várias semelhanças. Nesta seção, são demonstrados alguns relatórios específicos para este laboratório, evitando repetir os que já foram criados, usando os dez mil registros que foram inseridos em sua base de dados. As tabelas de 15 a 20 e figuras de 43 a 49 são os resultados destes testes.

Tabela 15 - Parasitologia: total de exames por agente infeccioso - Tableau

Bastocystis sp	812
Entamoeba nana	791
Entamoeba histolytica	790
Entamoeba dispar	786
Entamoeba coli	777
Blastocystis hominis	763
Giardia intestinalis	699
Plasmodium falciparum	680
Schistosoma mansoni	651
Dipilidium caninum	521
Dirofilaria immitis	491
Chilomastix menili	463
Leishmania dovani	460
Leishmania braziliensis	418
Leishmania peruviana	346
Toxoplasma gondii	163
Ancylostoma caninum	150
Ancylostoma braziliense	139
Taenia multiceps	100
Total geral	10.000

Fonte: dados da pesquisa (2019)

Figura 43 - Parasitologia: total de exames por agente infeccioso - Tableau

Fonte: dados da pesquisa (2019)

Tabela 16 - Parasitologia: total de exames, por ano, por agente infeccioso - Tableau

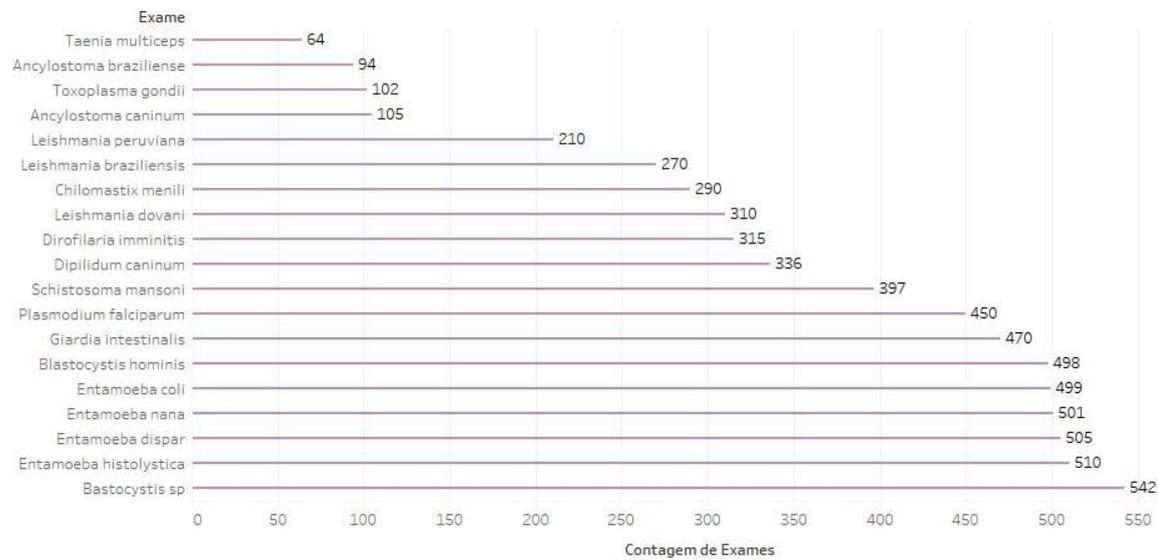
	2015	2016	2017	2018	2019	2020	Total geral
Bastocystis sp	155	170	174	144	168	1	812
Entamoeba nana	159	149	168	167	148		791
Entamoeba histolytica	163	144	163	160	156	4	790
Entamoeba dispar	166	145	152	150	172	1	786
Entamoeba coli	158	150	169	157	143		777
Blastocystis hominis	140	177	154	138	154		763
Giardia intestinalis	145	144	150	125	135		699
Plasmodium falciparum	141	142	133	133	131		680
Schistosoma mansoni	119	110	140	142	139	1	651
Dipilidium caninum	105	113	94	102	106	1	521
Dirofilaria immitis	88	94	101	102	106		491
Chilomastix menili	88	110	81	81	103		463
Leishmania dovani	86	94	79	94	107		460
Leishmania braziliensis	83	89	71	92	83		418
Leishmania peruviana	64	72	72	69	68	1	346
Toxoplasma gondii	33	27	32	39	32		163
Ancylostoma caninum	34	28	38	27	23		150
Ancylostoma braziliense	24	27	33	31	23	1	139
Taenia multiceps	21	20	23	18	18		100
Total geral	1.972	2.005	2.027	1.971	2.015	10	10.000

Fonte: dados da pesquisa (2019)

Tabela 17 - Parasitologia: total de exames positivos por agente infeccioso - Tableau

Bastocystis sp	542
Entamoeba histolytica	510
Entamoeba dispar	505
Entamoeba nana	501
Entamoeba coli	499
Blastocystis hominis	498
Giardia intestinalis	470
Plasmodium falciparum	450
Schistosoma mansoni	397
Dipilidium caninum	336
Dirofilaria immitis	315
Leishmania dovani	310
Chilomastix menili	290
Leishmania braziliensis	270
Leishmania peruviana	210
Ancylostoma caninum	105
Toxoplasma gondii	102
Ancylostoma braziliense	94
Taenia multiceps	64
Total geral	6.468

Fonte: dados da pesquisa (2019)

Figura 44 - Parasitologia: total de exames por agente infeccioso - Tableau

Fonte: dados da pesquisa (2019)

Tabela 18 - Parasitologia: total de exames positivos, por ano, por agente infeccioso - Tableau

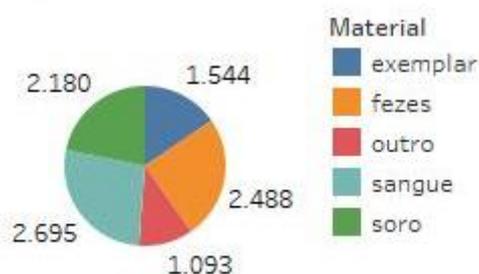
	2015	2016	2017	2018	2019	2020	Total geral
Bastocystis sp	92	109	124	98	118	1	542
Entamoeba histolytica	108	99	99	101	100	3	510
Entamoeba dispar	106	95	90	102	112		505
Entamoeba nana	101	98	111	106	85		501
Entamoeba coli	104	96	111	97	91		499
Blastocystis hominis	94	114	105	93	92		498
Giardia intestinalis	93	96	102	89	90		470
Plasmodium falciparum	99	83	89	89	90		450
Schistosoma mansoni	75	73	80	79	89	1	397
Dipilidium caninum	63	78	62	65	67	1	336
Dirofilaria immitis	57	61	58	69	70		315
Leishmania dovani	56	61	47	68	78		310
Chilomastix menili	57	60	55	53	65		290
Leishmania braziliensis	50	59	52	52	57		270
Leishmania peruviana	41	49	37	46	37		210
Ancylostoma caninum	24	21	22	20	18		105
Toxoplasma gondii	24	14	20	26	18		102
Ancylostoma braziliense	18	19	23	20	14		94
Taenia multiceps	13	14	16	10	11		64
Total geral	1.275	1.299	1.303	1.283	1.302	6	6.468

Fonte: dados da pesquisa (2019)

Tabela 19 - Parasitologia: distribuição de tipos de amostras - Tableau

Material	
sangue	2.695
fezes	2.488
soro	2.180
exemplar	1.544
outro	1.093
Total geral	10.000

Fonte: dados da pesquisa (2019)

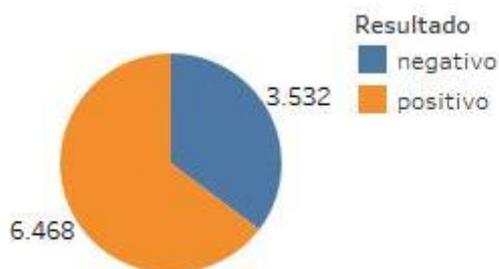
Figura 45 - Parasitologia: distribuição de tipos de amostras, gráfico de pizza – Tableau

Fonte: dados da pesquisa (2019)

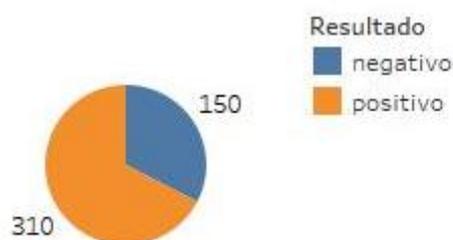
Tabela 20 - Parasitologia: distribuição de tipos de amostras, por ano - Tableau

Material	Data						Total geral
	2015	2016	2017	2018	2019	2020	
sangue	543	551	530	549	521	1	2.695
fezes	478	517	491	470	528	4	2.488
soro	412	424	457	446	438	3	2.180
exemplar	315	319	300	309	299	2	1.544
outro	216	220	215	227	214	1	1.093
Total geral	1.964	2.031	1.993	2.001	2.000	11	10.000

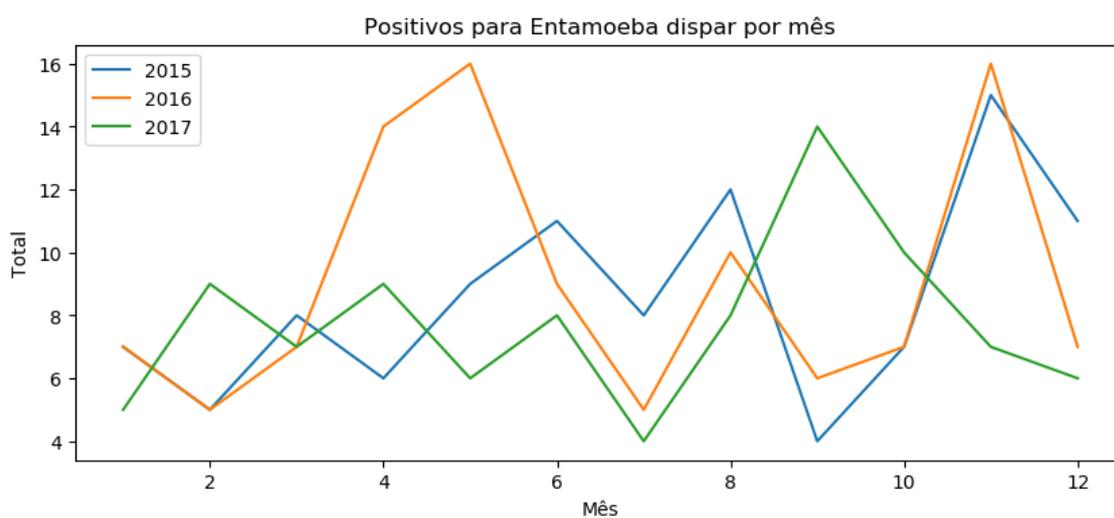
Fonte: dados da pesquisa (2019)

Figura 46 - Parasitologia: distribuição resultados de exames – Tableau

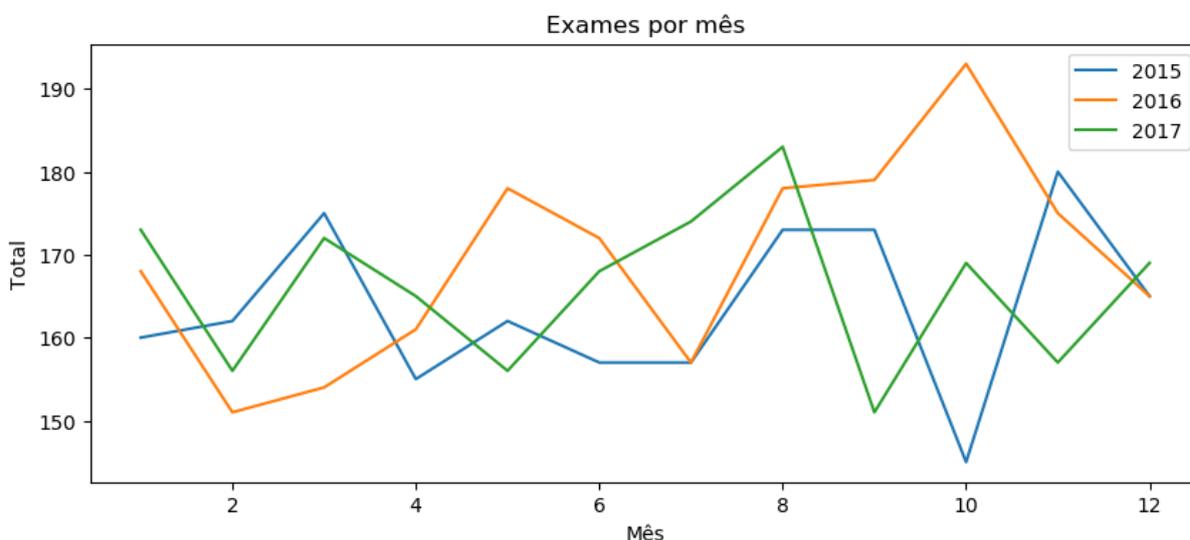
Fonte: dados da pesquisa (2019)

Figura 47 - Parasitologia: distribuição resultados de exames para *Leishmania dovani* – Tableau

Fonte: dados da pesquisa (2019)

Figura 48 - Parasitologia: positivos para *Entamoeba dispar*, por mês, por ano - Python

Fonte: dados da pesquisa (2019)

Figura 49 - Parasitologia: total de exames por mês, de 2015 a 2017

Fonte: dados da pesquisa (2019)

6.4.3 Patologia Animal

Os dados que este laboratório movimenta podem ser divididos em dois tipos: quantitativos, como os de cadastro pessoal do cliente e do paciente (no sentido de que são mais facilmente analisados como informações tangíveis, mensuráveis) e qualitativos, resultantes da análise prévia do material recebido e do diagnóstico, cujo produto normalmente são vários textos descritivos. Este tipo de informação, em *data science*, pode ser analisado pela tecnologia de aprendizado de máquina.

Machine learning, ou aprendizado de máquina, um campo de estudos altamente relacionado à matemática e à estatística, e considerado uma subárea da inteligência artificial, é definido por Mike Roberts, editor do livro “*Introducing Data Science*” de Cielen, Meysman e Ali (2016, tradução nossa), como “o processo pelo qual um computador consegue trabalhar cada vez com mais precisão, à medida que coleta e aprende com os dados que recebe”.

Sendo assim, a premissa é que um algoritmo, ou programa de computador, seja capaz de “aprender”, como se estivesse sendo programado através de exemplos, em comparação com a programação tradicional, que define seu comportamento explicitamente, desde o início. Esta definição é bastante abrangente e pode cobrir vários tipos de programas, mas considera-se que esta conceituação é reconhecida como válida (BEAM; KOHANE, 2018).

Exemplificando, no caso do laboratório de Patologia Animal, um modelo básico do uso desta tecnologia: o sistema baseado em *machine learning* seria alimentado com a maior quantidade possível de exames anteriores, usando as informações dos campos como “necropsia”, “histórico”, “material enviado” e “histologia”, além do diagnóstico decidido por um humano, na busca por padrões, treinando o sistema.

O quadro 15 mostra o formato destas informações na base de dados que foi criada para a prova de conceito. Na tabela “exame” os campos qualitativos possuem dados reais, mas randômicos, inseridos e mesclados para cada tupla de maneira aleatória, em um total de dez mil registros.

Quadro 15 - Patologia: exemplo de relatório com dados e diagnóstico – Python

nome	sexo	idade	peso	especie	raca	data_material	historico	necropsia	histologia	diagnostico
SEM NOME	F	5	14.52	CANINA	LHASA APSO	2015-05-19	ruptura da fascia muscular abdominal, o cadáver encontrava-se em bom estado nutricional, criado com ração, a proprietária observou diarreia sangüinolenta, o animal apresentou sangramento pelo nariz e ponta das orelhas	figado: exibindo hemorragias no parênquima, base: áreas de hemorragia subcapsulares multifocais, vesiga: encontrada retraída, esteatose difusa moderada, vesiga: sem conteúdo urinoso	estômago: sem alterações microscópicas, rim: espessamento da cápsula de Bowman, cerebelo: sem alterações microscópicas, base: rarefação linfoide moderada, figado: focos de hepatite portal mista de grau leve	doença periodontal
ALEMÃO	M	5	2.98	CUNICOLA	SRD	2016-07-25	não era vacinado	presença de vermes no intestino	figado: colestase moderada	Eriçiose
CHITARA	F	8	17.15	CANINA	POLICIAL	2019-12-05	mucosas se apresentavam normocoradas, tetraparesia espástica dos membros do lado direito	figado: mostrou-se evidencição de sua lobulação e de tonalidade difusamente pálido-amarelada, hemorragia pulmonar	coração: focos de miocardite mista discreta, rim: espessamento da cápsula de Bowman	câncer
TUCA	F	8	29.19	CAPRINA	PARDO ALPINO	2018-08-04	criado com alimentação caseira	figado: exibindo hemorragias no parênquima	rim: pigmentos acastanhados intra-tubulares	leptospirose
ALEMOA	F	11	1.17	ROEDORES	LEBRE	2019-11-27	estava se alimentando e tomando água, o animal apresentou sangramento pelo nariz e ponta das orelhas, hemograma não apresentou alterações	hematoma na parte cervical esquerda, figado: exibindo hemorragias no parênquima, ossos: vértebras da região lombo-sacra apresentaram fratura extensa	coração: focos de miocardite mista discreta, figado: deposição de tecido conjuntivo na região periportal, cerebelo: sem alterações microscópicas	leptospirose

Fonte: dados da pesquisa (2019)

Dessa maneira, com as múltiplas iterações o modelo criado pode ser capaz de alcançar níveis excepcionais de acurácia, como o descrito por McKinney *et al.* (2020), que descreve em um estudo um algoritmo treinado com imagens de raio-x de aproximadamente 29 mil mulheres, em que foi capaz de trabalhar com a mesma eficiência de seis radiologistas, na análise das imagens em busca de sinais de câncer de mama, com menos falsos positivos (1,2 %) e falsos negativos (2,7 %). Os autores deixam claro que um humano deverá ser o encarregado pelo diagnóstico final, principalmente quando não houver consenso entre o resultado automatizado e o de um médico treinado. Mesmo assim, a combinação de trabalho máquina/humano economiza 88% do tempo da segunda leitura.

Este uso do *machine learning* está mais próximo ao apoio no funcionamento diário do laboratório do que para a geração de *insights* ou ativos de conhecimento, mas a base de raciocínio é semelhante. Com alterações no modelo que se usa, e através do treinamento do algoritmo, muda-se a resposta que se deseja encontrar. Isto levaria a análise dos dois tipos de dados, simultaneamente, se desejado. Para isso, o Python possui bibliotecas altamente especializadas para a aplicação em *machine learning*, prontas para uso (CIELEN; MEYSMAN; ALI, 2016).

Na área de dados quantitativos, as análises deste laboratório seriam análogas as demonstradas anteriormente, para os laboratórios Cedima e Parasitologia.

6.4.4 Cedima x Parasitologia – Dados Interlaboratoriais

Esta seção demonstra algumas possibilidades de correlação de dados entre os laboratórios Cedima e Parasitologia, tentando comprovar o valor dos dados interlaboratoriais.

Para que isso seja possível, alguns pressupostos foram assumidos como verdadeiros: os clientes e pacientes possuem um cadastro único, com um identificador universal, compartilhado entre os sistemas dos laboratórios que permite o acesso às suas informações básicas, como nome, data de nascimento e etc., evitando duplicidades e permitindo uma busca centralizada. Entende-se também que os códigos para espécie e raça usados para classificar os animais são universais.

Para esta aplicação, o Python se mostrou mais indicado, pela capacidade maior que uma linguagem de programação tem de aceitar comandos e controles mais granulados em relação a um programa que usa uma interface “*point-and-click*”, que apesar de ser mais fácil para o usuário final, limita as possibilidades de customização de comandos e de gráficos e tabelas que pode gerar. Aqui, nota-se que quanto mais controle o pesquisador quiser ter sobre os resultados mais complicadas serão as ferramentas que terá que utilizar, a ponto de que em algum momento terá que recorrer a um especialista na área ou buscar treinamentos específicos, fora de seu campo de conhecimento inicial.

Os quadros 16, 17 e 18 mostram correlações básicas entre as bases de dados, demonstrando as possibilidades de pesquisa que um pesquisador teria em mãos.

Quadro 16 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para *Blastocystis hominis* e Cedima quaisquer resultados – Python

Nome	Sexo	Raça	Idade	Peso	Parasitologia: Data Exame	Parasitologia: Resultado	Parasitologia: Exame	Cedima: Data Exame	Cedima: Exame	Cedima: Resultado	Cedima: complemento
LUKE	M	S.R.D	12	5	2016-02-19	positivo	Blastocystis hominis	2018-08-11	Isolamento e identificação bacteriana	positivo	Streptococcus pneumoniae
DUQUE	M	S.R.D	8	24.19	2015-04-25	positivo	Blastocystis hominis	2017-02-19	Isolamento e identificação bacteriana	positivo	Leptospira interrogans
XUXA	F	S.R.D	4	21.97	2019-10-16	positivo	Blastocystis hominis	2018-06-24	Pesquisa de Anaeróbicos	negativo	
SEM NOME	M	S.R.D	12	7.91	2016-12-31	positivo	Blastocystis hominis	2019-11-13	Isolamento e identificação bacteriana	negativo	Streptococcus pneumoniae

Fonte: dados da pesquisa (2019)

Quadro 17 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para *Schistosoma mansoni* e Cedima positivos - Python

Nome	Sexo	Idade	Peso	Raça	Parasitologia: exame	Parasitologia: resultado	Parasitologia: Data Exame	Cedima: Data Exame	Cedima: Exame	Cedima: Resultado
NEGO	M	9	16.66	S.R.D	Schistosoma mansoni	positivo	2016-01-05	2018-10-08	Pesquisa de Anaeróbicos	positivo

Fonte: dados da pesquisa (2019)

Quadro 18 - Cedima x Parasitologia: cães, raça perdigueiro, positivos para isolamento e identificação bacteriana, Cedima, e que constam na parasitologia, independente do resultado – Python

Nome	Sexo	Idade	Peso	Raça	Parasitologia: exame	Parasitologia: resultado	Parasitologia: Data Exame	Cedima: Data Exame	Cedima: exame	Cedima: Resultado	Complemento
LILLO	F	4	10.28	PERDIGUEIRO	Blastocystis sp	positivo	2018-09-16	2015-06-01	Isolamento e identificação bacteriana	positivo	Salmonella bongori

Fonte: dados da pesquisa (2019)

O quadro 19 foi gerada usando não somente, para a seleção dos registros, os comandos SQL do banco de dados, mas também o método *query* da biblioteca pandas, disponível na linguagem python. Ele permite que os registros de um *dataframe* sejam filtrados pelos seus cabeçalhos. O código usado foi:

```
dfJoin.query('Idade >= 16 and Sexo == "M" and Especie == "FELINA" and Peso > 2 and Raca == "HAVANA")
```

Desta maneira, a decisão de filtragem dos dados acontece em um nível mais próximo do pesquisador, que pode trabalhar com todos os registros da base de dados em memória, sem precisar manipular o banco de dados, que já foi lido. O quadro 20 é outro exemplo.

Quadro 19 - Cedima x Parasitologia: gatos, com 16 anos ou menos, sexo feminino, com menos de 2 quilos e da raça Havana e que foram atendidos nos dois laboratórios - Python

Nome	Sexo	Nascimento	Idade	Especie	Raca	Peso	Parasitologia Data do Exame	Parasitologia Resultado	Cedima Data do Exame	Cedima Resultado	Complemento	Cedima Exame
FAISCA	M	2003-08-21	16	FELINA	HAVANA	5	2015-08-04	positivo	2019-05-22	negativo		Vacina Papilomatose
BUGUE	M	2003-02-20	16	FELINA	HAVANA	4.35	2019-11-04	negativo	2017-02-13	positivo		BHV-1 e 5 (PCR)

Fonte: dados da pesquisa (2019)

Quadro 20 - Cedima x Parasitologia: cães, no Cedima, com 11 anos ou mais, machos, com peso maior que 1 quilo e na Parasitologia com positivo para isolamento viral e qualquer resultado de exame - Python

Nome	Sexo	Idade	Especie	Raca	Peso	Parasitologia Data do Exame	Parasitologia Resultado	Exame	Cedima Data do Exame	Cedima Resultado	Agente	Cedima Exame
SULTÃO	M	11	CANINA	BORDER COLLIE	7.41	2016-08-08	negativo	Schistosoma mansoni	2018-02-16	positivo	Avian avulavirus	Isolamento Viral
BONO	M	13	CANINA	VEADEIRO	19.74	2015-10-10	positivo	Plasmodium falciparum	2016-01-17	positivo	Porcine circovirus	Isolamento Viral
NEGO	M	13	CANINA	FOX TERRIER	16.29	2016-08-17	positivo	Blastocystis sp	2019-02-20	positivo	Herpesvírus Suídeo	Isolamento Viral
BUGUE	M	12	CANINA	URRADOR AMERICANO	20.7	2018-05-17	negativo	Leishmania peruviana	2019-01-24	positivo	Canine parvovirus	Isolamento Viral
NIKE	M	13	CANINA	LULU DA POMERANIA	19.7	2015-10-03	positivo	Chilomastix menili	2016-11-20	positivo	Erbovirus	Isolamento Viral
FILHOTE	M	13	CANINA	GALGO	15.27	2015-11-14	positivo	Chilomastix menili	2015-08-04	positivo	Lyssavirus	Isolamento Viral
BABY	M	11	CANINA	RED HILLER	16.51	2019-01-22	positivo	Dirofilaria immitis	2016-09-13	positivo	Canarypox	Isolamento Viral
PEPI	M	13	CANINA	SAO BERNARDO	23.81	2017-02-19	positivo	Leishmania braziliensis	2016-08-20	positivo	Canine parvovirus	Isolamento Viral
FARRAPO	M	13	CANINA	GOLDEN RETRIEVER	5	2016-06-21	negativo	Entamoeba coli	2015-07-24	positivo	Sapelovirus	Isolamento Viral
NIKE	M	11	CANINA	SAO BERNARDO	12.97	2016-11-14	positivo	Blastocystis hominis	2017-08-28	positivo	Parechovirus	Isolamento Viral
FALCÃO	M	13	CANINA	BORDER COLLIE	11.74	2016-08-26	positivo	Ancylostoma caninum	2015-08-12	positivo	Lyssavirus	Isolamento Viral
TEDI	M	12	CANINA	RED HILLER	7.82	2018-07-28	positivo	Chilomastix menili	2015-05-31	positivo	Porcine circovirus	Isolamento Viral
BRUNO	M	13	CANINA	CIMARROM	5	2016-09-30	positivo	Giardia intestinalis	2016-07-24	positivo	Kobuvirus	Isolamento Viral
LILO	M	13	CANINA	LHASA APSO	19.1	2018-01-24	positivo	Plasmodium falciparum	2016-01-14	positivo	Senecavirus	Isolamento Viral
AJAX	M	13	CANINA	URRADOR AMERICANO	12.52	2019-03-13	negativo	Schistosoma mansoni	2017-03-21	positivo	Aphthovirus	Isolamento Viral
PERIGO	M	12	CANINA	BORDER COLLIE	18.45	2015-10-09	negativo	Leishmania peruviana	2018-04-07	positivo	Kobuvirus	Isolamento Viral

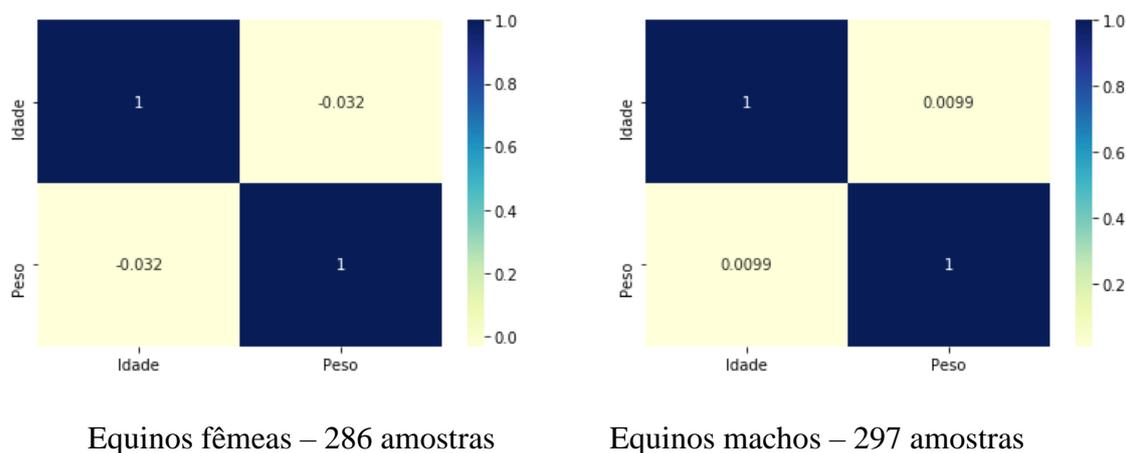
Fonte: dados da pesquisa (2019)

Método *query* usado:

```
Idade >= 11 and Sexo == "M" and Espécie == "CANINA" and Cedima_Resultado == "positivo" and Peso > 1 and Cedima_Exame == "Isolamento Viral"
```

Seguindo com as possibilidades do Python, um *heatmap* (mapa de calor) é uma maneira prática de realizar análise exploratória em dados numéricos, podendo revelar alguns de seus padrões. O método *heatmap* da biblioteca pandas correlaciona as informações que recebe e as devolve como um gráfico de cores. Na figura 50, o peso e a idade dos equinos atendidos pelos laboratórios Cedima e Parasitologia são comparados na tentativa de verificar se os seu valores possuem alguma relação.

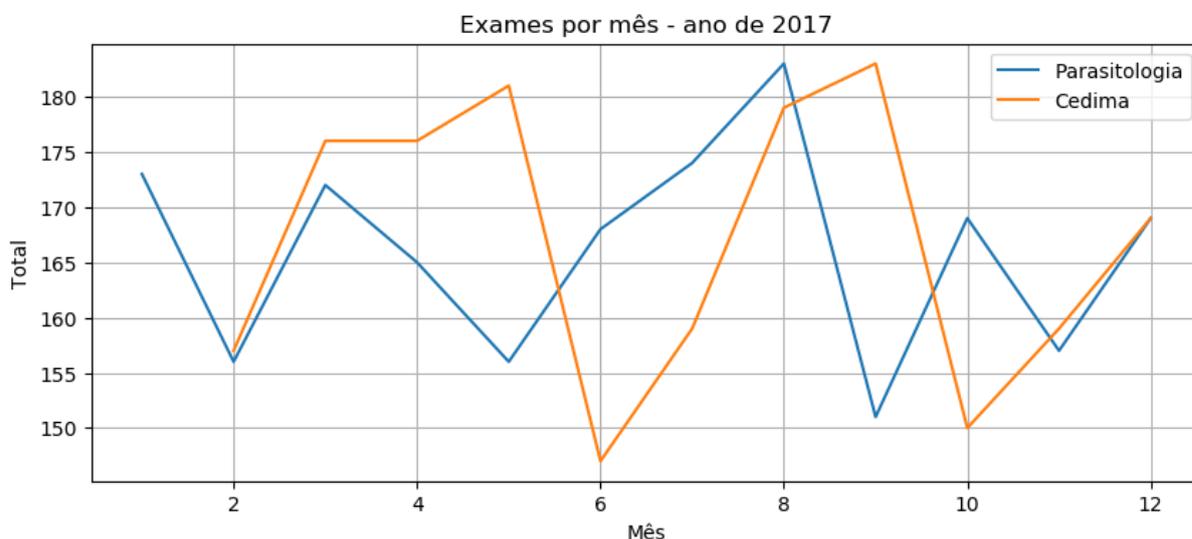
Figura 50 – Cedima x Parasitologia: *heatmap* de pesos versus idades de equinos - Python



Fonte: dados da pesquisa (2019)

Quanto mais próximo de “1” o valor na caixa de referência for, maior a correlação entre os valores que representa, e, quanto mais próximo de “-1”, menor a relação entre as variáveis, segundo o padrão seguido por esta biblioteca. Neste exemplo, o resultado demonstra que para os machos há uma relação maior entre a idade e o peso, neste grupo amostral. O método de mapa de calor pode utilizar várias informações simultaneamente e ser usado também em mapas geográficos.

E, finalmente, a figura 51 ilustra uma comparação da quantidade de exames realizados em cada laboratório, por mês, no ano de 2017. Com um *dataframe* para cada laboratório carregado na memória é possível fazer a contagem de registros por ano e por mês para a geração do gráfico.

Figura 51 – Cedima x Parasitologia: exames por mês, ano 2017 - Tableau

Fonte: dados da pesquisa (2019)

6.4.5 Cedima, Parasitologia e Patologia Animal – Dados Interlaboratoriais

Nesta sessão tentou-se demonstrar as possibilidades de uso dos dados interlaboratoriais, correlacionando as informações dos três bancos de dados, criados para esta prova de conceito.

Notou-se, durante os testes, que ao se usar os três laboratórios, as possibilidades de geração de *insights* se amplia, de maneira proporcionalmente direta à complexidade para se achar as correlações, principalmente pelo grande número de variáveis disponíveis para análise.

Um exemplo é a resultante da pesquisa demonstrada no quadro 21. Ela visa gerar um relatório que relacione o máximo possível de informações, como os “cães, da raça yorkshire, antibiograma no Cedima, quaisquer exames na Parasitologia e com alguma ocorrência relacionada ao baço, na Patologia”. Neste caso o pesquisador poderia estar tentando confirmar uma relação entre problemas deste órgão com infecções de um determinado agente.

Quadro 21 – Cedima x Parasitologia x Patologia: cães, yorkshire, antibiograma no Cedima, quaisquer exames Parasitologia e relação com o baço, na Patologia - Python

Nome	Sexo	Especie	Raca	Parasitologia Exame	Parasitologia Resultado	Cedima Exame	Cedima Resultado	Complemento	Patologia histórico	Patologia Necropsia	Patologia Histologia	Patologia diagnóstico
TICO	M	CANINA	YORKSHIRE	Schistosoma mansoni	positivo	Antibiograma	positivo	Clostridium chauvoei	<p>mucoas se apresentavam normocoradas, na região da cabeça, ao corte, apresenta área avermelhada e crepitante, criado com alimentação caseira, o cadáver encontrava-se em bom estado nutricional</p>	<p>bexiga: encontrada retraída, baco: áreas de hemorragia subcapsulares multifocais, rins: ao corte dos dois antímeros, as regiões córtico-medulares e medulares apresentavam tonalidade levemente avermelhada, pulmões: levemente avermelhado</p>	<p>fígado: vacuolização discreta de hepatócitos, cérebro: sem alterações microscópicas, megalocitose de hepatócitos, estômago: sem alterações microscópicas</p>	alergia alimentar

Fonte: dados da pesquisa (2019)

No próximo exemplo (quadro 22), tenta-se criar uma relação entre a FeLV, a leucemia felina, infecção comum em gatos, com possíveis efeitos nos rins do animal, usando o campo “histologia” do laboratório de Patologia Animal.

Quadro 22 - Cedima x Parasitologia x Patologia: felinos, SRD, positivos para FeLV no Cedima, qualquer resultado Parasitologia e com anotações sobre os rins, na Patologia - Python

Nome	Sexo	Idade	Especie	Raca	Parasitologia Exame	Parasitologia Resultado	Cedima Exame	Cedima Resultado	Patologia histórico	Patologia Necropsia	Patologia Histologia	Patologia diagnóstico
PRETA	F	9	FELINA	SRD	Giardia Intestinalis	negativo	FeLV (Imunofluorescência Indireta)	positivo	na região da cabeça, ao corte, apresenta área avermelhada e crepitante	cavidade peritoneal: presença de moderada quantidade (100 mililitros) de conteúdo sangüinolento	rim: pigmentos acastanhados intra-tubulares	doença periodental
NIKE	M	8	FELINA	SRD	Dipylidium caninum	negativo	FeLV (Imunofluorescência Indireta)	positivo	áreas de tricotomia nos membros anteriores, orifícios naturais: as mucosas mostraram-se esmaecidas (pálidas), hemograma não apresentou alterações, foi encaminhado ao hospital, lesões circunscritas avermelhadas e com ausência de pêlo na região da cabeça e das patas	bexiga: sem conteúdo urinoso, bazo: com extensa ruptura de sua extremidade caudal. Vasos dilatados e congestionados no intestino, pleura: conteúdo coágulo cruótrico, vesícula biliar: distendida conteúdo fluido biliar de aspecto seroso, amarelo-amarronzado	estômago: sem alterações microscópicas, hemorragia centrolobular difusa, rim: pigmentos acastanhados intra-tubulares, fígado: colestase moderada	TRAUMATISMO causado PROVAVELMENTE por choque de grande impacto, resultando em perda sangüinea significativa e aguda
BRUNO	M	11	FELINA	SRD	Bastocystis sp	positivo	FeLV (Imunofluorescência Indireta)	positivo	na região da cabeça, ao corte, apresenta área avermelhada e crepitante, presença de sangue na cavidade nasal	hemorragia ileo-ceco acerosa, sutúdes subcutâneas e musculares	rim: nefrite intersticial mononuclear moderada, fígado: deposição de tecido conjuntivo na região periportal	eutanásia

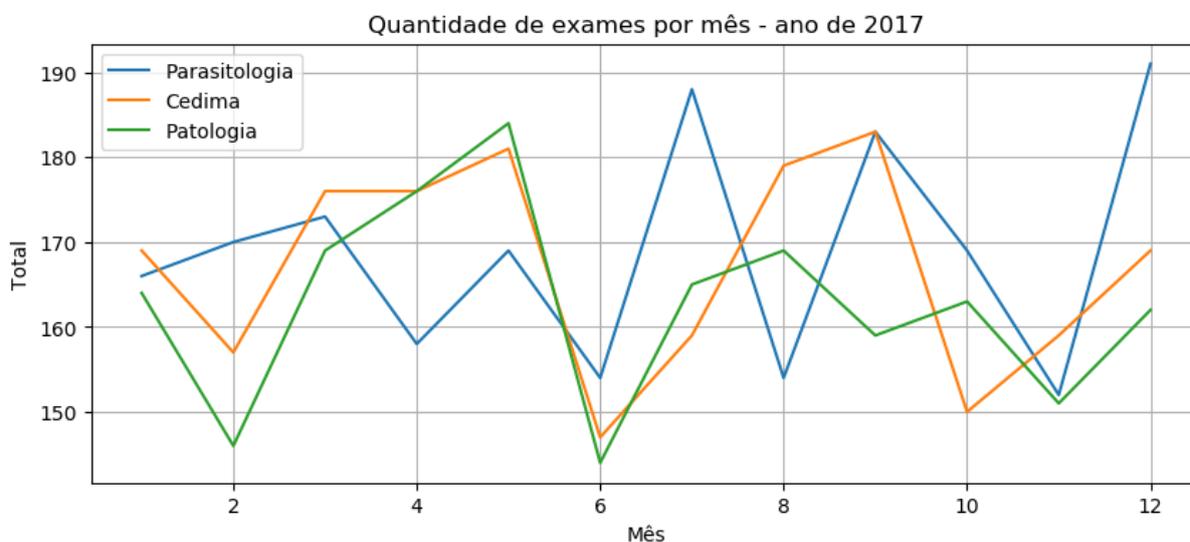
Fonte: dados da pesquisa (2019)

Dependendo do resultado de uma procura, nas bases de dados, neste formato, poderia ser possível confirmar uma suspeita ou uma noção empírica que um pesquisador já possuía ou afastar esta possibilidade, caso as evidências sejam fortes o suficiente.

Ao expandir este raciocínio, poderiam ser incluídos em relatórios assim os dados do laboratório de Patologia Clínica, do Hospital Veterinário, usando, em um caso análogo, as informações resultantes dos hemogramas dos felinos, tentando visualizar alguma conexão entre as doenças.

De maneira semelhante, a figura 52 é um gráfico que compara o número de exames que os três laboratórios tiveram no ano de 2017. Tal informação pode ser de grande valia para a administração do campus ou do laboratório na tomada de decisão, mas, como será discutido logo adiante, não pode ser, por si só, uma fonte isolada, deslocada do âmbito de cada um deles.

Figura 52 – Cedima, Parasitologia x Patologia: comparação da quantidade de exames por mês, em 2017 – Python



Fonte: dados da pesquisa (2019)

E, por fim, um exemplo (quadro 23) que foca em bovinos da raça holandês, que passaram pela Parasitologia, com resultados positivos para o vírus da herpes bovina do tipo 1 e 5 e que na Patologia Animal a histologia teve algum achado relacionado ao fígado do animal

Quadro 23 – Cedima, Parasitologia x Patologia: bovinos da raça holandês, Cedima, positivos para herpes bovina do tipo 1 e 5 e Patologia Animal algum achado relacionado ao fígado - Python

Nome	Sexo	Idade	Especie	Raca	Parasitologia Exame	Parasitologia Resultado	Cedima Exame	Cedima Resultado	Patologia histórico	Patologia Necropsia	Patologia Histologia	Patologia diagnóstico
SAM	F	8	BOVINA	HOLANDES	Entamoeba dispar	negativo	BHV-1 e 5 (PCR)	positivo	não era vacinado, hemograma não apresentou alterações, lesões circunscritas avermelhadas e com ausência de pêlo na região da cabeça e das patas, tecido subcutâneo e musculatura: hemorragia do tipo sufusão, foi encaminhado ao hospital	cavidade peritoneal: presença de moderada quantidade (100 mililitros) de conteúdo sanguinolento, rins: bilateralmente, revelaram superfíoles discretamente avermelhadas, fígado: mostrou-se evidencição de sua lobulação e de tonalidade difusamente pálido-amarelada, baço: com extensa ruptura de sua extremidade caudal, pleura: contendo coágulo crúrtico	megalocitose de hepatócitos, pulmão: enfisema crônico moderado em bordos, rim: pigmentos acastanhados intra-tubulares, fígado: vacuolização discreta de hepatócitos, cerebelo: sem alterações microscópicas	TRAUMATISMO causado PROVAVELMENTE por choque de grande impacto, resultando em perda sanguínea significativa e aguda

Fonte: dados da pesquisa (2019)

O objetivo da prova de conceito foi verificar a aplicabilidade do *Data Science* em uma base de dados formada por informações interlaboratoriais relacionadas à saúde animal, na tentativa de extrair ativos de conhecimento, ou *insights*, segundo esta técnica.

Para tanto, a Gestão do Conhecimento se mostrou como um norteador, revelando a importância das informações (PRUSAK, 1997), do seu trato correto (DAVENPORT; PRUSAK, 1998) e definindo a memória organizacional como o ponto-chave de sua aplicabilidade (MOLINA; VALENTIM, 2011), o que notou-se não ser utilizado atualmente nos laboratórios.

Ao englobar a memória organizacional, os ativos de conhecimento, ou intelectuais, de uma organização ressaltam em valor (FERNANDES; FADEL, 2013), razão pela qual buscou-se ferramentas que fossem capazes de encontrá-los, em meio à massas de informações que só tendem a crescer, em um laboratório informatizado.

Desta maneira, a técnica de *Data Science* surge, na atualidade, como uma ferramenta amplamente testada e disponível, por meio de várias implementações de ferramentas e tecnologias, tanto comerciais quanto gratuitas (CIELEN; MEYSMAN; ALI, 2016).

Notou-se que um *insight* é um constructo assemelhado ao ativo de conhecimento, mas menor, mais veloz, fugidio, carente de análise profunda e criteriosa, que, ao ser confirmado, realimenta e valoriza a massa de ativos intangíveis da instituição, e a partir do qual, imediatamente, novos *insights* podem ser conseguidos, reiniciando o ciclo da informação.

As ferramentas de análise usadas, Tableau e a linguagem Python, se mostraram capazes, cada uma com sua forma de aplicação e nicho de mercado, de manipular as informações dos bancos de dados criados na procura de padrões e correlações, na fase do *Data Science* conhecida como exploração. A curva de aprendizado do *software* Tableau se mostrou mais suave, em comparação com o Python e suas bibliotecas associadas. Por outro lado, o poder de uma linguagem de programação dificilmente pode ser comparado ao de um programa que busca ser fácil para o usuário final: sua complexidade ainda é, infelizmente, inversamente proporcional à acessibilidade de uso.

Em síntese, entende-se que a Gestão do Conhecimento, se implantada nos laboratórios como um sistema informatizado e uma metodologia de gestão (por meio da Memória Organizacional, por exemplo), se mostra útil na geração de ativos de conhecimento, na forma de *insights*, pelo meio da técnica de *Data Science*.

Entretanto, faz-se necessário anotar algumas observações que surgiram durante a pesquisa, em relação à validade das amostras, confidencialidade e ética, correlações espúrias e correlação versus causalidade, gerando uma crítica em relação ao alcance destes métodos e de sua aparente ubiquidade de uso.

Validade das Amostras

Para que uma pesquisa quantitativa baseada em uma população seja considerada válida, é preciso que a amostra estudada seja considerada válida. Esta amostragem deve seguir regras claras, para que o pesquisador possa, posteriormente, generalizar as suas conclusões para toda uma população (FONSECA, 2002).

Os dados disponibilizados pelos laboratórios são representativos da sua região de abrangência, limitado à região sul e ao planalto catarinense, em grande parte. Com isso em mente, é necessário ressaltar as limitações nas amostras e por consequência das conclusões que derivarem. Esta particularidade não impede que gerem *insights*, ideias e propostas para trabalhos mais amplos, que podem até mesmo confirmar uma visão inicial que carecia de maior fundamentação amostral.

Confidencialidade e ética

A confidencialidade e a ética no trato com os dados precisa ser levado em conta, em vista das consequências que pode trazer o acesso massivo e direto à grandes massas de dados das pessoas, com informações particulares e dos animais pelos quais são responsáveis (SCHUTT; O'NEIL, 2013)

Neste sentido, foram previstos mecanismos que ofereçam camadas de proteção ao acesso aos dados. Neste sentido, um dos princípios da segurança de informação é que um usuário só tenha acesso somente ao que lhe compete (KUMAR; RAJ; JELCIANA, 2018), respeitando a confidencialidade e ao mesmo tempo fazendo pleno uso do conhecimento disponível para as pesquisas, segundo o que foi autorizado pelo seu titular ou guardião.

No que tange à ética, a bioética, uma de suas áreas, preve que o método científico siga regras claras, visando humanizar a sua aplicação tanto para humanos quanto para animais, e, por isso, a sua importância extrapola a pesquisa e entra no campo da filosofia, chegando até a ética e a moral (FRANCO *et al.*, 2014). Com isso em conta, entende-se que há a possibilidade

de que as informações geográficas de infecção na cidade e região, números de prevalência em raças, espécies ou outros dados sensíveis possam, se utilizados de maneira errônea, levar à decisões que ferem os pilares do bom trato aos animais e aos seus responsáveis. A massa de informações e os resultados de suas análises não devem nublar o valor da vida humana e animal.

É necessário tato neste mister, pois, como alertam Schutt e O’Neil (2013), os modelos e algoritmos não só são capazes de prever o futuro, mas de gerá-lo.

Correlações Espúrias: correlação x causalidade

Em contraponto à visão quase universal da importância e validade do *Data Science*, alguns autores tecem críticas à aproximação quase mítica tecida em alguns momentos.

Neste sentido, Calude e Longo (2017), criticam a ideia de que as conclusões tiradas por algoritmos poderosos o suficiente, ao explorar bases de dados gigantescas, sejam aceitas prontamente. As correlações e padrões seriam alheias a significados, conteúdos e valores, visto que são conceitos difíceis de serem incluídos nos parâmetros de análise computacional.

O poder e a generalização desta técnica se baseiam nos tamanhos imensos das bases de dados: quanto maior, mais poderoso e eficaz é o método. Consequentemente, não haveria necessidade de teorizar, entender ou criticar as correlações encontradas. Nenhuma semântica ou análise causal é necessária. Como questionam Calude e Longo (2017): “com dados, suficiente poder computacional e algoritmos estatísticos, padrões serão encontrados. Mas estes padrões são de interesse? Nem todos serão, já que padrões espúrios superam e muito os válidos”. Estes padrões são fruto do acaso, e não apresentam relação entre si, de maneira direta ou indireta.

Neste momento, a expressão originária da Estatística, segundo a qual “correlação não implica causalidade” (LUIZ; STRUCHINER, 2002; SCHUTT; O’NEIL, 2013), deve ser lembrada: causalidade é um conceito que define que há uma relação entre duas variáveis, na medida que uma é consequência da outra. Como seres humanos, intuitivamente concluímos, às vezes de maneira errada, que acontecimento correlacionados são causa um de outro, mesmo sem comprovação. No caso de dados extraídos de bases de dados gigantescas, que podem dar origem à milhares de correlações e padrões, é necessário cuidado nas interpretações.

A solução se baseou em aprofundar o significado de uma correlação, evitando conclusões apressadas. Foi necessário procurar outros fatores subjacentes, causas ocultas, testá-las e só então derivá-las (SINGH, 2018)

Isto posto, é preciso entender o real valor e papel do *Data Science*, como uma técnica auxiliar na tomada de decisão, pois, como afirma Han (2015), em relação à grandes bases de dados, “mais informações nem sempre levam a decisões melhores. Por vezes, menor quantidade de informação pode levar a mais clareza. Mais dados não são o mesmo que ‘mais verdade’ ou ‘mais sentido’”.

6 CONSIDERAÇÕES FINAIS

É possível concluir que a utilização correta das tecnologias relacionadas à gestão de informação, nas áreas de saúde animal, pode influenciar benéficamente as suas pesquisas e a gerência cotidiana dos laboratórios associados.

O *Data Science* e as ferramentas correlatas se apresentam como uma alternativa disponível e comprovadamente eficientes pelo corpo bibliográfico atual, abundante de exemplos e ensaios aplicados por vários tipos de entidades, o que se comprova pelo número expressivo de soluções em *software* disponíveis para implementação.

Assim, entende-se que este tipo de solução para a região, desde que implementada e funcional, pode fomentar a conectividade com outros órgãos, principalmente públicos, municipais e estaduais, ampliando as possibilidades de análise. Atualmente pouco se armazena, calcula e analisa neste setor em nosso País. Por consequência, pouco se sabe e se compartilha, dificultando a criação de conhecimento e também a fiscalização.

Na instituição de ensino, onde a saúde animal é o foco principal do curso de Medicina Veterinária, a manutenção de bancos de dados de longa data relacionados aos pacientes, seus dados, resultados de exames e prontuários de atendimento no hospital de clínica veterinária é imprescindível para a pesquisa. Estudantes da graduação e de pós-graduação, com acesso a esses dados, podem submeter trabalhos e artigos mais rapidamente, desenvolvendo pesquisas em menor tempo e com informações mais confiáveis, levando a um maior impacto no mundo acadêmico, com retorno para o campus local e a Universidade. Além disso, por ser o HCV um hospital escola, as informações relacionadas aos animais são usadas em sala de aula, incrementando a qualidade do ensino.

E, claramente, a informatização dos processos de gestão nos laboratórios trazem os ganhos esperados, desde a velocidade no trato das informações, possibilidade de replicação, compartilhamento, segurança no armazenamento, controle de acesso e a restauração à um estado funcional, em caso de desastre.

Além disso, as informações geradas são de grande valia para a gestão do campus ou do laboratório, para a correta tomada de decisão de forma conjunta e otimizada da Universidade.

E, em momentos de surtos e epidemias que o mundo vive atualmente, as conexões e correlações entre as doenças complexas, poderão proporcionar novos ativos de conhecimento com grande número de variáveis disponíveis para análise por meio de evidências fortes entre seres humanos e animais.

Isto posto, pode-se dizer que os objetivos específicos propostos foram alcançados da seguinte forma:

O objetivo 1: Verificar como e quais dados são gerenciados atualmente nos laboratórios e suas especificidades. Esse objetivo foi alcançado por meio das visitas aos laboratórios e da análise dos métodos utilizados para o manejo das suas informações, por meio de documentos impressos, editores de texto ou sistemas específicos, desenvolvidos para estes fins. Após a sua análise foi possível identificar a natureza dos dados manejados no dia-dia.

O objetivo 2: Pesquisar a tecnologia de análise de dados disponível mais indicada para o cruzamento de dados interlaboratoriais e que viabilize a criação de ativos de conhecimento: Durante o período de pesquisa bibliográfica, necessária para cumprir o objetivo 1, o termo *Data Science*, se destacou como o mais conhecido e difundido para a análise de informações e geração de derivados. Após aprofundamento e busca de fundamentação, essa noção de concretizou, baseada principalmente pela quantidade profusa de pesquisa disponibilizada nos últimos anos sobre esse assunto e as áreas de pesquisa que se utilizam dela atualmente.

Objetivo 3: Propor um modelo utilizando ferramentas de TI, para a correlação dos dados, habilitando a prospecção de ativos de conhecimento de saúde animal, testando a tecnologia encontrada. Para cumprir este objetivo, foi criado um protótipo conhecido como prova de conceito, para o qual foram projetados e criados os bancos de dados dos laboratórios. Estes foram preenchidos com informações aleatórias e neles, através da linguagem de programação Python e do software Tableau, foram testados os princípios de exploração de dados proposto pelo Data Science.

Para estudos futuros, algumas sugestões relacionados ao tema são apontados a seguir:

- A inclusão, nos estudos, de outros laboratórios do campus, de áreas de pesquisa semelhantes, ampliando a disponibilidade de variáveis a serem analisadas e com isso a amplitude das correlações possíveis.

- Aprofundar os estudos sobre as ferramentas de *software* disponíveis para a implantação de uma solução deste tipo. Programas como o Apache Hadoop, Spark, Storm, ou o Hydra, usadas em processamento distribuído de bancos de dados e aplicações ou o uso de outras linguagens de programação e análise, como o R, Julia, MatLab, Scala e até mesmo C++.

- Identificar como o conhecimento é percebido, entendido, processado e utilizado por alunos, técnicos e professores na Instituição, visto que ela é, tipicamente, uma entidade gestora do conhecimento.

- Aplicar um estudo amplo nesse mesmo raciocínio em laboratórios relacionados à saúde humana. Qual o impacto que teria um banco de dados, onde estejam à disposição informações consolidadas, com resultados dos exames de laboratórios de toda uma população humana, com dados pessoais como idade, peso, altura, sexo e endereço associados a tipos sanguíneos, resultados de hemogramas, testes de vírus, bactérias, parasitas, e etc.? Os ganhos compensariam os possíveis desvios ou usos discriminatórios das informações? Temos uma legislação atualizada que preveja este caso ou permita sua implantação? Importante, neste caso, aprofundar as implicações éticas e de confidencialidade dos dados.

- Estudar a correlação entre bancos de dados humanos e animais. Com a ascensão de doenças de fundo zoonótico, todas as informações disponíveis sobre estas relações devem ser analisadas, para entender o que está acontecendo e prever futuros incidentes, antecipando-os, se possível.

Por fim, entende-se que nenhuma das possibilidades de análises foram esgotadas, visto que as técnicas possíveis dentro do *Data Science* são inúmeras, novas possibilidades surgem continuamente na comunidade científica e comercialmente e, para os fins da prova de conceito, as utilizadas se mostraram eficientes para alcançar os objetivos propostos.

7 REFERÊNCIAS

- ALAVI, M.; LEIDNER, D. Knowledge Management Systems: Issues, Challenges, and Benefits. **Communications of the Association for Information Systems**, v. 1, p. 38, 1999.
- ALAVI, M.; LEIDNER, D. E. Review: Knowledge Management and Knowledge Systems: Conceptual Foundations and Research Issues. **MIS Quarterly**, v. 25, n. 1, p. 107–136, 2001.
- ALVARENGA, A. T. DE et al. Histórico, fundamentos filosóficos e teórico-metodológicos da interdisciplinaridade. In: PHILIPPI JR., A.; SILVA NETO, A. J. (Eds.). . **Interdisciplinaridade em ciência, tecnologia & inovação**. 1. ed. Barueri - SP: Editora Manole Ltda., 2011. p. 350.
- BARLAS, P.; LANNING, I.; HEAVEY, C. A survey of open source data science tools. **International Journal of Intelligent Computing and Cybernetics**, v. 8, n. 10, p. 232–261, 2012.
- BAŠKARADA, S.; KORONIOS, A. Unicorn data scientist: the rarest of breeds. **Program**, v. 51, n. 1, p. 65–74, 2017.
- BASSI, S. **Python for Bioinformatics**. Boca Raton, FL: Chapman & Hall, 2017.
- BEAM, A. L.; KOHANE, I. S. Big data and machine learning in health care. **JAMA - Journal of the American Medical Association**, v. 319, n. 13, p. 1317–1318, 2018.
- BEREZOWSKI, J.; LINDBERG, A.; WARD, M. Surveillance against the odds: Addressing the challenges of animal health surveillance. **Preventive Veterinary Medicine**, v. 120, n. 1, p. 1–3, 2015.
- BHARADWAJ, A. S. A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. **MIS Quarterly**, v. 24, n. 1, p. 169, 2000.
- BLEI, D. M.; SMYTH, P. Science and data science. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 33, p. 8689–8692, 2017.
- BRASIL. **Capes disponibiliza plataforma científica com mais de 60 milhões de registros**. Disponível em: <www.brasil.gov.br/editoria/educacao-e-ciencia/2017/08/capes-disponibiliza-plataforma-cientifica-com-mais-de-60-milhoes-de-registros>. Acesso em: 23 dez. 2017.
- CÁCERES, P. et al. The World Organisation for Animal Health and the World Health Organization: Intergovernmental disease information and reporting systems and their role in early warning. **OIE Revue Scientifique et Technique**, v. 36, n. 2, p. 539–548, 2017.
- CALUDE, C. S.; LONGO, G. The Deluge of Spurious Correlations in Big Data. **Foundations of Science**, v. 22, n. 3, p. 595–612, 2017.
- CARLUCCI, D. Assessing the links between knowledge assets and value creation in organisations. **Measuring Business Excellence**, v. 16, n. 2, p. 70–82, 2012.
- CARSTEN, B. C. Carsten's Corner. **Power Conversion & Intelligent motion**, n. 38, 1989.

CAV. **UDESC CAV - Centro de Ciências Agroveterinárias**. Disponível em: <<http://www.cav.udesc.br>>. Acesso em: 3 jul. 2018.

CAVALCANTI, M. Knowledge management: Using information technology to obtain knowledge in a competitive environment. **Corporate Ownership & Control**, v. 7, n. 1, p. 232–243, 2009.

CHANG, Y. H.; CHEN, J. W.; LIN, B. KMsharer: an information technology approach to enable knowledge management services. **International Journal of Technology Management**, v. 43, n. 1–3, p. 252–265, 2008.

CHASE, R. Knowledge management, innovation and firm performance. **Journal of Knowledge Management**, v. 9, n. 1, p. 101–115, 2006.

CHAVES, C. Controle de qualidade no laboratório de análises clínicas. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, v. 46, n. 5, p. 1, 2010.

CHEN, J. et al. Big data challenge: A data management perspective. **Frontiers of Computer Science**, v. 7, n. 2, p. 157–164, 2013.

CHUANG, S.-H.; LIAO, C.; LIN, S. Determinants of knowledge management with information technology support impact on firm performance. **Information Technology and Management**, v. 14, n. 3, p. 217–230, 2013.

CHUNG, Y. C. et al. Relationship between organizational cultures, information technology involvement, degrees of knowledge management implementations and performance of new product developments. **Information Technology Journal**, v. 9, n. 7, p. 1504–1516, 2010.

CIELEN, D.; MEYSMAN, A. D. B.; ALI, M. **Introducing Data Science**. New York: Manning Publications Co., 2016.

COSS, D.; SAMONAS, S. The CIA Strikes Back: Redefining Confidentiality, Integrity and Availability in Security. **Journal of Information System Security**, v. 10, n. 3, p. 21–45, 2014.

DAVENPORT, T. H.; BARTH, P.; BEAN, R. How “big data” is different. **MIT Sloan Management Review**, v. 54, n. 1, p. 21–25, 2012.

DAVENPORT, T. H.; PATIL, D. J. **Data Scientist: The Sexiest Job of the 21st Century**. Disponível em: <<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>>. Acesso em: 5 dez. 2019.

DAVENPORT, T. H.; PRUSAK, L. Working Knowledge How Organization Manage What They Know. **Harvard Business School Press**, n. January 1998, p. 1–15, 1998.

DEWANGAN, R. R.; THOMBRE, D.; PATEL, C. Big data technology in health and biomedical research: A literature review. **International Journal of Database Theory and Application**, v. 9, n. 11, p. 175–184, 2016.

DRUCKER, P. The new productivity challenge. **Harvard Business Review**, v. 1, n. novembro-dezembro, p. 69–79, 1991.

DRUCKER, P. The Rise of The knowledge Society. **The Wilson Quarterly**, v. 17, n. 2, p. 52–71, 1993.

DUARTE, E. N. et al. Vantagens do uso de tecnologias para criação, armazenamento e disseminação do conhecimento em bibliotecas universitárias. **Transinformação**, v. 18, n. 2, p. 131–141, 2006.

ÉLOIT, M. Survey on updating OIE World Animal Health Information System. **Veterinary Record**, v. 180, n. 15, p. 386, 2017.

ELSEVIER. **Scopus**. Disponível em: <<https://www.elsevier.com/pt-br/solutions/scopus>>. Acesso em: 28 ago. 2018.

ERICKSON, S.; ROTHBERG, H. Big data and knowledge management: Establishing a conceptual foundation. **The Electric Journal of Knowledge Management**, v. 12, n. 2, p. 115–123, 2014.

FERNANDES, A.; FADEL, B. O capital intelectual a partir dos ativos do conhecimento aplicados ao desenvolvimento das micro e pequenas empresas: perspectivas teóricas. In: **Na vanguarda do conhecimento: diálogos e debates**. 1. ed. Franca (SP): Uni-FACEF, 2013. p. 70–82.

FOLIUM. **Folium**. Disponível em: <<https://python-visualization.github.io/folium/>>. Acesso em: 3 jan. 2020.

FONSECA, J. J. S. DA. **Metodologia da Pesquisa Científica**. Fortaleza: UECE - Universidade Estadual do Ceará, 2002.

FOOTE, A.; HALAWI, L. A. Knowledge management models within information technology projects. **Journal of Computer Information Systems**, v. 58, n. 1, p. 89–97, 2018.

FOREMAN, J. W. **Data Smart Using Data Science to Transform Information into Insight**. Indianapolis: John Wiley & Sons, Inc., 2014.

FRANCO, A. L. et al. Pesquisas em animais: Uma reflexão bioética. **Acta Bioethica**, v. 20, n. 2, p. 247–253, 2014.

FREIRE, P. DE S. et al. Memória organizacional e seu papel na gestão do conhecimento. **Revista de Ciências da Administração**, v. 14, n. 33, p. 41–51, 2012.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, n. 2, p. 137–144, 2015.

GARCIA JÚNIOR, E. F.; MEDEIROS, S.; AUGUSTA, C. Análise documental: uma metodologia da pesquisa para a Ciência da Informação. **Núcleo de Artes Midiáticas - NAMID/UFPB**, v. 7, n. 34, p. 138–150, 2017.

GODOY, A. S. Pesquisa Qualitativa Tipos Fundamentais. **Revista de Administração de empresas**, v. 35, n. 3, p. 20–29, 1995.

GOODRICH, M. T.; TAMASSIA, R. **Introdução à Segurança de Computadores**. Porto Alegre: Bookman, 2013.

GÜNTER, H. Pesquisa Qualitativa Versus Pesquisa Quantitativa: Esta É a Questão? **Psicologia: Teoria e Pesquisa**, v. 22, n. 2, p. 201–210, 2006.

HAN, B.-C. **The Transparency Society**. California: Stanford University Press, 2015.

HANSEN, M. T.; NOHRIA, M.; TIERNEY, T. What's your strategy for managing knowledge? **Harvard Business Review**, v. 77, n. 2, p. 106, 1999.

HENDRIKS, P. H. J.; LIGTHART, P. E. M.; SCHOUTETEN, R. L. J. Knowledge management, health information technology and nurses' work engagement. **Health Care Management Review**, v. 00, n. 3, p. 1, 2015.

HENRICKS, W. H. Laboratory Information Systems. **Clinics in Laboratory Medicine**, v. 36, n. 1, p. 1–11, 2016.

HOGAN, R. **A Practical Guide to Database Design**. Boca Raton, FL: Chapman and Hall/CRC, 2018.

HUANG, H. L. et al. The relationship between knowledge management strategy and information technology strategy. **World Academy of Science, Engineering and Technology**, v. 77, n. 5, p. 432–436, 2011.

HUESTON, W. D. Assessment of national systems for the surveillance and monitoring of animal health. **Revue scientifique et technique**, v. 12, p. 1187–1196, 1993.

JUPYTERTEAM. **User interface components**. Disponível em: <https://jupyter-notebook.readthedocs.io/en/stable/ui_components.html>.

KADIYALA, A.; KUMAR, A. Applications of Python to evaluate environmental data science problems. **Environmental Progress and Sustainable Energy**, v. 36, n. 6, p. 1580–1586, 2017.

KENDRICK, J. S. The box plot: A simple visual method to interpret data. **Academia and Clinic**, v. 10, n. July 1989, p. 916–921, 1989.

KUMAR, P. R.; RAJ, P. H.; JELCIANA, P. Exploring Data Security Issues and Solutions in Cloud Computing. **Procedia Computer Science**, v. 125, n. 2009, p. 691–697, 2018.

KUO, Y.-K.; YE, K.-D. How employees' perception of information technology application and their knowledge management capacity influence organisational performance. **Behaviour & Information Technology**, v. 29, n. 3, p. 287–303, 2010.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. **Application Delivery Strategies**, v. 6, n. 70, p. 1–4, 2001.

LAZER, D. et al. Social science: Computational social science. **Science**, v. 323, n. 5915, p. 721–723, 2009.

LEE, I. Big data: Dimensions, evolution, impacts, and challenges. **Business Horizons**, v. 60, n. 3, p. 293–303, 2017.

LEFF, E. Complexidade, interdisciplinaridade e saber ambiental. **Olhar de Professor**, v. 14,

n. 2, p. 309–335, 2011.

LIN, D. et al. Does knowledge management matter for information technology applications in China? **Asia Pacific Journal of Management**, v. 25, n. 3, p. 489–507, 2008.

LIN, H. An investigation of the effects of cultural differences on physicians' perceptions of information technology acceptance as they relate to knowledge management systems. **Computers in Human Behavior**, v. 38, p. 368–380, 2014.

LUIZ, R. R.; STRUCHINER, C. J. **Inferência causal em epidemiologia: o modelo de respostas potenciais**. Rio de Janeiro - RJ: Editora FIOCRUZ, 2002.

MAO, H. et al. Information technology resource, knowledge management capability, and competitive advantage: The moderating role of resource commitment. **International Journal of Information Management**, v. 36, n. 6, p. 1062–1074, 2016.

MARDONES, F. O. et al. Veterinary epidemiology: Forging a path toward one health. **Preventive Veterinary Medicine**, v. 137, p. 147–150, 2017.

MAROOFI, F. Sociability Organizations Can Facilitate Knowledge Management Through Information Technology? **International Business Management**, v. 10, n. 13, p. 2551–2555, 2016.

MATHUR, N.; PUROHIT, R. Issues and Challenges in Convergence of Big Data, Cloud and Data Science. **International Journal of Computer Applications**, v. 160, n. 9, p. 7–12, 2017.

MCKINNEY, S. M. et al. International evaluation of an AI system for breast cancer screening. **Nature**, v. 577, n. 7788, p. 89–94, 2020.

MCKINNEY, W. **Python for Data Analysis**. Sebastopol, CA: O'Reilly, 2013. v. 53

MELLO, R.; LEITE, L. R.; MARTINS, R. A. Is big data the next big thing in performance measurement systems? **IIE Annual Conference and Expo 2014**, n. September, p. 1837–1846, 2014.

MOHAMAD, A. A.; RAMAYAH, T.; LO, M. C. Knowledge management in MSC Malaysia: The role of information technology capability. **International Journal of Business and Society**, v. 18, n. S4, p. 651–660, 2017.

MOLINA, L. G.; VALENTIM, M. L. P. Memória organizacional, memória corporativa e memória institucional: discussões conceituais e terminológicas. **Revista EDICIC - Asociación de Educación e Investigación en Ciencia de la Información de Iberoamérica y el Caribe**, v. 1, n. 1, p. 262–276, 2011.

MORESI, E. A. D. et al. Memória organizacional de comunidades de prática como fator de vantagem competitiva das organizações. **Organizações & Sociedade**, v. 16, n. 50, p. 479–495, 2009.

MORIN, E. Complex Thinking for a Complex World - About Reductionism, Disjunction and Systemism. **Systema**, v. 2, n. 1, p. 14–22, 2014.

MURRAY, D. G. **Tableau Your Data!: Fast and Easy Visual Analysis with Tableau**

Software. 2. ed. Indianapolis - IN: Wiley, 2016.

NAUR, P. **Concise survey of computer methods**. New York: Petrocelli Books, 1974.

NONAKA, I. A dynamic theory of organizational knowledge creation. **Organization Science**, v. 5, n. 1, p. 14–37, 1994.

NONAKA, I.; TAKEUCHI, H. **Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação**. 16. ed. Rio de Janeiro (RJ): Campus, 1997.

OIE-WAHIS. **The OIE-WAHIS project**. Disponível em: <<https://www.oie.int/en/animal-health-in-the-world/the-oie-wahis-project/>>. Acesso em: 9 dez. 2019.

OKUMUS, F. Facilitating knowledge management through information technology in hospitality organizations. **Journal of Hospitality and Tourism Technology**, v. 4, n. 1, p. 64–80, 2013.

OLIVEIRA, B. **Boxplot: Como Interpretar?** Disponível em: <<https://operdata.com.br/blog/como-interpretar-um-boxplot/>>. Acesso em: 25 jan. 2020.

ORACLE. **VirtualBox**. Disponível em: <<https://www.virtualbox.org/>>. Acesso em: 15 dez. 2019.

PANDAS. **The pandas project**. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 10 jan. 2020.

PARK, S. L. et al. Anatomic Pathology Laboratory Information Systems: A Review. **Advanced Anatomic Pathology**, v. 19, n. 2, p. 81–96, 2012.

PEREIRA, C. DE S. Memória organizacional: conceito e práticas em construção. **37 Encontro da ANPAD - EnANPAD**, p. 1–15, 2013.

PEREIRA, M. O. F.; SILVA, H. DE F. N.; PINTO, J. S. DE P. A Memória organizacional nos processos de gestão do conhecimento: um estudo na Universidade Federal do Paraná. **Informação & Informação**, v. 21, n. 1, p. 348, 2016.

PILLANIA, R. K. Information technology strategy for knowledge management in Indian automotive components SMEs. **Knowledge and Process Management**, v. 15, n. 3, p. 203–210, 2008.

PREFEITURA DE LAGES. **Conheça Lages**. Disponível em: <<http://www.lages.sc.gov.br/>>. Acesso em: 3 jul. 2018.

PROJECTJUPYTER. **About Us**. Disponível em: <<https://jupyter.org/about>>. Acesso em: 16 dez. 2019.

PROVOST, F.; FAWCETT, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. **Big Data**, v. 1, n. 1, p. 51–59, 2013.

PRUSAK, L. **Knowledge in Organizations (Resources for the Knowledge-Based Economy)**. 1. ed. Newton, MA: Butterworth-Heinemann, 1997.

PYTHON. **random - Generate pseudo-random numbers**. Disponível em: <<https://docs.python.org/3/library/random.html>>. Acesso em: 17 dez. 2019.

PYTHONBRASIL. **PerguntasFrequentes / SobrePython**. Disponível em: <<https://wiki.python.org.br/PerguntasFrequentes/SobrePython>>. Acesso em: 15 dez. 2019.

RABINOWITZ, P. M. et al. A planetary vision for one health. **BMJ global health**, v. 3, n. 5, p. 6, 2018.

RAJAMARAN, V. **Introduction to Information Technology**. 3. ed. Patparganj, Delhi: PHI Learning Pvt. Ltd., 2018.

RASLI, A. M.; WAN MOHD, M. W. Project performance framework: The role of knowledge management and information technology infrastructure. **Asian Journal of Business and Accounting**, v. 1, n. 2, p. 39–64, 2008.

REVILLA, E.; RODRÍGUEZ-PRADO, B.; PRIETO, I. Information technology as knowledge management enabler in product development: Empirical evidence. **European Journal of Innovation Management**, v. 12, n. 3, p. 346–363, 2009.

ROSSETTI, A.; MORALES, A. B. T. O papel da tecnologia da informação na gestão do conhecimento. **Ciência da Informação**, v. 36, n. 1, 5 dez. 2007.

RUSSOM, P. BIG DATA ANALYTICS - TDWI BEST PRACTICES REPORT Introduction to Big Data Analytics. **TDWI best practices report, fourth quarter**, v. 19, n. 4, p. 1–34, 2011.

SALTZER, J. H.; SCHROEDER, M. D. The Protection of Information in Computer Systems. **Proceedings of The IEEE**, v. 63, n. 9, p. 1278–1308, 1975.

SAMPAIO, R.; MANCINI, M. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia**, v. 11, n. 1, p. 83–89, 2007.

SANCHEZ-PINTO, L. N.; LUO, Y.; CHURPEK, M. M. Big Data and Data Science in Critical Care. **Chest**, v. 154, n. 5, p. 1239–1248, 2018.

SCHIUMA, G.; CARLUCCI, D.; SOLE, F. Applying a systems thinking framework to assess knowledge assets dynamics for business performance improvement. **Expert Systems with Applications**, v. 39, n. 9, p. 8044–8050, 2012.

SCHUTT, R.; O'NEIL, C. **Doing Data Science: Straight Talk**. Sebastopol, CA: O'Reilly, 2013.

SHEPPARD, K. Introduction to Python for Econometrics, Statistics and Data Analysis. **Report**, p. 363, 2019.

SILVEIRA, D. T.; CÓRDOVA, F. P. Unidade 2 - A Pesquisa Científica. In: **Métodos de Pesquisa**. Porto Alegre: Editora da UFRGS, 2009. p. 31–42.

SINGH, S. **Why correlation does not imply causation?** Disponível em: <<https://towardsdatascience.com/why-correlation-does-not-imply-causation-5b99790df07e>>. Acesso em: 2 fev. 2020.

SITARSKI, K. The Role of Information Technology Systems in Knowledge Management. **Foundations of Management**, v. 2, n. 1, p. 117–132, 2010.

SOLIDIT. **DB-Engines**. Disponível em: <<https://db-engines.com/en/ranking>>. Acesso em: 15 dez. 2019.

SPILLER, A. R. **Memória organizacional e reutilização do conhecimento contábil-financeiro : estudo de caso em uma indústria multinacional eletroeletrônica no Brasil**. São Paulo: Centro Universitário Álvares Penteado - UNIFECAP, 2005.

TSENG, S. M. The effects of information technology on knowledge management systems. **Expert Systems with Applications**, v. 35, n. 1–2, p. 150–160, 2008.

TURULJA, L.; BAJGORIĆ, N. Information Technology, Knowledge Management and Human Resource Management : Investigating Mutual Interactions towards Better Organizational Performance. **VINE Journal of Information and Knowledge Management Systems**, 2017.

UDESC. **UDESC - Universidade do Estado de Santa Catarina**. Disponível em: <<http://www.udesc.br>>. Acesso em: 3 jul. 2018.

VENKITACHALAM, K.; AMBROSINI, V. A triadic link between knowledge management, information technology and business strategies. **Knowledge Management Research & Practice**, v. 15, n. 2, p. 192–200, 2017.

VENTURA, M. O Estudo de Caso como Modalidade de Pesquisa. **Sociedade de Cardiologia do Estado do Rio de Janeiro - SOCERJ**, v. 20, n. 5, p. 383–386, 2007.

VERGARA, S. C. **Projetos e Relatórios de Pesquisa em Administração**. 2. ed. São Paulo: Atlas, 2009.

VIEIRA, E. A percepção da informação e da sua relevância no cenário institucional: sob a perspectiva de gestores e líderes. **Cadernos EBAPE**, v. 12, n. 8, p. 533–552, 2014.

WEILL, P.; SUBRAMANI, M.; BROADBENT, M. Building IT Infrastructure for Strategic Agility. **MIT Sloan Management Review**, n. January 2003, p. 57–66, 2002.

WILD, R.; GRIGGS, K. A model of information technology opportunities for facilitating the practice of knowledge management. **VINE Journal of Information and Knowledge Management Systems**, v. 38, n. 4, p. 490–506, 2008.

YANG, C. W.; FANG, S. C.; LIN, J. L. Organisational knowledge creation strategies: A conceptual framework. **International Journal of Information Management**, v. 30, n. 3, p. 231–238, 2010.

ZANINOTTO, F. **Faker 3.0.0**. Disponível em: <<https://pypi.org/project/Faker>>. Acesso em: 17 dez. 2019.

8 APÊNDICE A - CÓDIGOS-FONTE CEDIMA: SQL E PYTHON

SQL 1 - Criação do banco de dados “Cedima”

```

SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8mb4 */;

CREATE DATABASE IF NOT EXISTS `cedima` DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_general_ci;
USE `cedima`;

CREATE TABLE `amostra` (
  `id` int(11) NOT NULL,
  `id_paciente` int(11) DEFAULT NULL,
  `id_tipo_amostra` int(11) DEFAULT NULL,
  `data_coleta` date NOT NULL,
  `data` date NOT NULL,
  `exame` varchar(255) NOT NULL,
  `histórico` varchar(255) NOT NULL,
  `suspeita` varchar(255) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `cliente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) NOT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `cpf` varchar(20) DEFAULT NULL,
  `cnpj` varchar(20) DEFAULT NULL,
  `endereco` varchar(255) DEFAULT NULL,
  `numero` varchar(45) DEFAULT NULL,
  `estado` varchar(2) DEFAULT NULL,
  `cep` varchar(8) DEFAULT NULL,
  `cidade` varchar(145) DEFAULT NULL,
  `bairro` varchar(255) DEFAULT NULL,
  `email` varchar(255) DEFAULT NULL,
  `telefone` varchar(45) NOT NULL,
  `data_nascimento` date NOT NULL,
  `data_cadastro` date DEFAULT NULL,
  `observacoes` varchar(255) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `especie` (
  `id` int(11) NOT NULL,
  `especie` varchar(145) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

INSERT INTO `especie` (`id`, `especie`) VALUES
(10, 'FELINA'), (21, 'CANINA'), (38, 'SUINA'), (41, 'BOVINA'), (53, 'OVINA'), (62, 'CAPRINA'), (77, 'ROEDORES'), (85, 'GALINÁCEOS'), (96,
'EQUINOS'), (906, 'MUAR'), (912, 'CUNICOLA'), (951, 'PSITACIDEO'), (975, 'AVICOLA'), (1035, 'PRIMATA'), (1228, 'FELIDAE'), (1301, 'MARRECO'),
(1362, 'CERVIDEO'), (1398, 'CHELONIA'), (1443, 'ASININO'), (1451, 'QUELONIO'), (1468, 'FAISAO'), (1521, 'CHINCHILA'), (1608, 'ALOATA'), (1716,
'CAMELUS BACTRIANO'), (1737, 'CRISTATUS'), (1831, 'ANFIBIO'), (1851, 'TUCANO'), (1895, 'GALICTIS VITTATA'), (1956, 'EDENTATA'), (2001,
'COLUMBINA'), (2024, 'ANDORINHA'), (2036, 'TITO ALBA'), (2045, 'PORCO ESPINHO'), (2057, 'DASYPUS'), (2073, 'TYTO ALBA'), (2098,
'SPIZACTUS'), (2241, 'AVE'), (2311, 'REPTIL'), (2442, 'MAZAMA GOUAZOUBIRO'), (2478, 'L. TIGRINUS'), (2490, 'RANPHOSTUS DICOLARUS'),
(2512, 'CERDOCIYON THOUS'), (2531, 'TAMANDUA TETRADACTYLA'), (2551, 'MELOPSITTACUS UNDULATUS'), (2586, 'MEPHITIS'), (2666,
'RATO RATTUS'), (2809, 'GAMBA'), (2820, 'COELHO'), (2821, 'SEM CLASSIFICAÇÃO'), (2822, 'CROCODILIANOS'), (2824, 'NÃO DEFINIDO'), (2825,
'DIDELPHIS'), (2826, 'TATU');

CREATE TABLE `exame` (
  `id` int(11) NOT NULL,
  `id_tipo_exame` int(11),
  `id_paciente` int(11) NOT NULL,
  `id_amostra` int(11) NOT NULL,
  `data_solicitacao` date NOT NULL,
  `resultado` varchar(140) NOT NULL,
  `complemento` varchar(300) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `paciente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) DEFAULT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `data_nascimento` date DEFAULT NULL,
  `peso` float(5,2) DEFAULT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `id_raca` int(11) DEFAULT NULL,
  `id_cliente` int(11) DEFAULT NULL,
  `data_cadastro` date DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

```
CREATE TABLE `raca` (
  `id` int(11) NOT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `raca` varchar(45) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

```
INSERT INTO `raca` (`id`, `id_especie`, `raca`) VALUES
```

```
(100, 21, 'S.R.D'), (115, 21, 'POODLE'), (123, 21, 'DOBERMANN'), (131, 21, 'FILA BRASILEIRA'), (146, 21, 'COCKER SPANIEL'), (152, 21, 'DOGO ARGENTINO'), (168, 21, 'SHEEP DOG'), (174, 21, 'ROTTWEILLER'), (187, 21, 'BOXER'), (191, 21, 'SCHNAUZER'), (201, 21, 'PASTOR ALEMAO'), (216, 21, 'AKITA'), (222, 21, 'COLLIE'), (238, 21, 'DALMATA'), (244, 21, 'PINSCHER'), (257, 21, 'SETTER IRLANDES'), (261, 21, 'WEIMARANER'), (270, 21, 'BASSET HOUND'), (285, 21, 'POINTER'), (293, 21, 'FOX PAULISTINHA'), (308, 21, 'AFGHAN HOUND'), (314, 21, 'DOG ALEMÃO'), (327, 21, 'DACHSHUND'), (331, 10, 'PERSA'), (340, 10, 'SIAMES'), (355, 10, 'HAVANA'), (363, 10, 'BIRMA'), (371, 96, 'CRIOULO'), (386, 96, 'PSI'), (392, 96, 'ARABE'), (401, 96, 'MANGA LARGA'), (410, 96, 'QUARTO DE MILHA'), (425, 96, 'APALOOSA'), (433, 96, 'BRETAO'), (441, 96, 'PERCHERON'), (456, 906, 'PEGA'), (462, 41, 'HOLANDES'), (478, 41, 'JERSEY'), (484, 41, 'GUERNSEY'), (497, 41, 'NORMANDO'), (503, 41, 'SIMENTAL'), (511, 41, 'PARDO SUICO'), (526, 41, 'CHAROLÉS'), (532, 41, 'CANCHIM'), (548, 41, 'ABERDEEN ANGUS'), (554, 41, 'NELORE'), (567, 41, 'TABAPUA'), (571, 41, 'GUZERA'), (580, 41, 'CHIINANINA'), (595, 41, 'MARCHIGIANA'), (602, 41, 'BLOND D'AQUITAINE'), (618, 41, 'DEVON'), (624, 41, 'RED POOL'), (637, 41, 'SANTA GERTRUDES'), (641, 41, 'PITANGUEIRAS'), (650, 41, 'FLAMENGA'), (665, 53, 'HAMPSHIRE DOWN'), (673, 53, 'CORRIEDALE'), (681, 53, 'TEXEL'), (696, 53, 'SUFFOLK'), (707, 53, 'ILLE DE FRANCE'), (711, 53, 'MERINO'), (720, 62, 'ANGLO NUBIANA'), (735, 62, 'SAANEN'), (743, 62, 'PARDO ALPINO'), (751, 62, 'TOGUEMBURGER'), (766, 38, 'LANDRACE'), (772, 38, 'LARGE WHITE'), (788, 38, 'DUROC'), (794, 38, 'DUROC WHITE'), (805, 38, 'MACAU'), (813, 77, 'GIGANTE DE FLANDRES'), (821, 2820, 'CALIFORNIA'), (842, 2820, 'ANGORA'), (858, 2820, 'NOVA ZELANDIA'), (864, 2311, 'TARTARUGA DO DESERTO'), (877, 21, 'PEQUINES'), (881, 21, 'PASTOR BELGA'), (890, 21, 'BERNESSE'), (928, 21, 'FOX'), (934, 21, 'LABRADOR'), (947, 21, 'HUSKY SIBERIANO'), (960, 2241, 'CURICACA'), (983, 21, 'FOX TERRIER'), (991, 21, 'VEADEIRO'), (1000, 21, 'LULU DA POMERANIA'), (1014, 21, 'CHIHUAHUA'), (1026, 21, 'BULLDOGUE'), (1047, 1608, 'BUGIO'), (1052, 21, 'BULLMASTIFF'), (1063, 53, 'ZEBU'), (1071, 21, 'LHASA APPO'), (1088, 53, 'LIMOUSINE'), (1091, 1035, 'SAGUI'), (1105, 2241, 'CANÁRIO'), (1117, 2241, 'PERIQUITO'), (1122, 21, 'SAO BERNARDO'), (1133, 2241, 'PAPAGAIO'), (1141, 21, 'MASTIM NAPOLITANO'), (1158, 2241, 'JOAO DE BARRO'), (1161, 975, 'VANELLUSC'), (1170, 96, 'PONEI'), (1184, 21, 'BEAGLE'), (1196, 21, 'URRADOR AMERICANO'), (1203, 21, 'OLD ENGLISH SHEEP DOG'), (1211, 21, 'DINAMARQUES'), (1231, 1228, 'PANTERA TIGER TIGER'), (1240, 21, 'CHOW-CHOW'), (1254, 21, 'GALGO'), (1266, 21, 'BORDER COLLIE'), (1275, 96, 'INGLES'), (1287, 2241, 'TUCANO'), (1292, 21, 'PERDIGUEIRO'), (1310, 1301, 'PEQUIN'), (1324, 21, 'POLICIAL'), (1336, 21, 'YORKSHIRE'), (1345, 21, 'GOLDEN RETRIEVER'), (1357, 21, 'SAMOIEDA'), (1373, 1035, 'MACACO-PREGO'), (1381, 21, 'PIT BULL'), (1406, 21, 'STAFFORDSHIRE'), (1415, 21, 'FREEZE'), (1427, 2241, 'FAISAO'), (1432, 21, 'SHIH-TZU'), (1471, 2311, 'TIGRE D'AGUA'), (1480, 2311, 'QUELONIO'), (1494, 21, 'PASTOR ALEMAO BRANCO'), (1502, 1228, 'TIGRE SIBERIANO'), (1513, 10, 'HIMALAIA'), (1538, 21, 'POINTER INGLES'), (1541, 21, 'MALTES'), (1550, 21, 'BULLTERRIER'), (1564, 21, 'SHARPEI'), (1576, 77, 'CHINCHILA'), (1585, 21, 'KUVASZ'), (1597, 21, 'SCOTTISH TERRIER'), (1611, 21, 'TERRA NOVA'), (1620, 2241, 'SIREMA'), (1634, 21, 'PUG'), (1646, 41, 'GEULBEVI'), (1655, 21, 'BICHON FRISE'), (1667, 21, 'KOMONDOR'), (1672, 96, 'ARDINES'), (1683, 77, 'HAMISTER'), (1691, 21, 'AMERICAN STAFFORDSHIRE TERRIER'), (1704, 21, 'BRACO ITALIANO'), (1725, 1716, 'CAMELO'), (1742, 2241, 'PAVO CRISTATUS'), (1753, 41, 'RED ANGUS'), (1761, 2241, 'SABIA'), (1778, 41, 'HEREFORD'), (1781, 1362, 'VEADO'), (1790, 41, 'MINIJERSEY'), (1807, 1716, 'LHAMA'), (1812, 77, 'PORCO DA INDIA'), (1823, 2241, 'CALUPSITA'), (1848, 2241, 'CORUJA'), (1860, 2241, 'CARDIAL'), (1874, 10, 'GATO DO MATO'), (1886, 1895, 'MUSTELIDEO'), (1901, 21, 'BIANCO'), (1918, 21, 'CHIMARRON'), (1921, 1228, 'LEAO BAI0'), (1930, 2241, 'GAVIAO'), (1944, 975, 'ARARA'), (1965, 1956, 'TAMANDUA'), (1977, 2241, 'PAVAO'), (1982, 2241, 'PAVAO'), (1993, 1451, 'CAGADO'), (2010, 2241, 'POMBA'), (2062, 2057, 'TATU MOLITA'), (2081, 2241, 'SUINDARA'), (2106, 2098, 'GAVIAO PENACHO'), (2115, 21, 'AUSTRALIAN CATTLE DOG - RED HEELER'), (2132, 10, 'PUMA CONDOR'), (2143, 21, 'DOGUE DE BORDEAUX'), (2151, 21, 'AUSTRALIAN CATTLE DOG - BLUE HEELER'), (2168, 96, 'BRASILEIRO DE HIPISMO'), (2171, 21, 'AIRIEDALE CATTLE DOG'), (2180, 21, 'GARGO BIANCO'), (2194, 21, 'WEST HIGHLAND WHITE TERRIER'), (2202, 21, 'AUSTRALIAN CATTLE DOG - RED HEELER'), (2213, 21, 'AUSTRALIAN CATTLE DOG - BLUE HEELER'), (2221, 21, 'SERRA DA ESTRELA'), (2238, 21, 'AMERICAN B. TERRIER'), (2250, 53, 'TECKEL'), (2264, 21, 'AUSTRALIAN CATTLE DOG - BLUE HEELER'), (2276, 21, 'SPITZ ALEMÃO'), (2285, 21, 'DOG CAMPEIRO'), (2297, 96, 'PAINT HORSE'), (2308, 41, 'POLI HEREFORD'), (2320, 96, 'HPB'), (2334, 21, 'AMERICAN PIT BULL'), (2346, 21, 'AMERICA'), (2355, 2241, 'GARCA PARDA'), (2372, 10, 'AMERICAN SHORTHAIIR'), (2383, 10, 'RAGDOLL'), (2391, 21, 'PASTORE DELLA BRIE'), (2404, 10, 'CHARTREUX'), (2437, 2241, 'PSITTACIDAE'), (2453, 2442, 'VEADO CATINGUEIRO'), (2461, 2241, 'CATURRITA'), (2481, 10, 'GATO MARACAIA'), (2507, 2241, 'TUCANO DO BICO VERDE'), (2523, 2512, 'GRAXAIN'), (2548, 2241, 'COLAPTES CAMPESTRES'), (2560, 21, 'PASTOR CANADENSE'), (2574, 62, 'MESTICO'), (2595, 21, 'DOGO ARGENTINO'), (2618, 77, 'CAVIA PORCELLUS'), (2621, 21, 'CANE CORSO'), (2630, 21, 'PASTOR DA MANTIQUEIRA'), (2644, 96, 'CAMPOLINA'), (2655, 21, 'RED HILLER'), (2677, 21, 'CIMARRON'), (2688, 10, 'KHAO MANEE'), (2699, 96, 'PAINT HORSE'), (2710, 21, 'AMERICAN BULLY'), (2721, 53, 'LACAUNE'), (2732, 21, 'CAMPEIRO'), (2743, 2821, 'PATO'), (2754, 21, 'RED NOOSE'), (2765, 10, 'AZUL RUSSO'), (2776, 96, 'PURO SANGUE INGLES'), (2787, 41, 'ANGUS'), (2798, 2241, 'SARACURA'), (2831, 21, 'BENJI ROAD'), (2842, 21, 'LEAO DA RODESIA'), (2843, 21, 'COCKER'), (2844, 2822, 'JACARÉ'), (2845, 2824, 'S.R.D'), (2847, 10, 'SRD'), (2848, NULL, 'BLUE HILLER'), (2849, NULL, 'BLUE HILLER'), (2850, 96, 'MESTIÇO'), (2851, 53, 'CRIOLA'), (2852, 951, 'MAITACA'), (2853, 951, 'ARARA-AZUL'), (2854, 951, 'PAPAGAIO'), (2855, 951, 'PIRIQUITO'), (2856, 951, 'SABIÁ'), (2857, 951, 'ARARA-VERMELHA'), (2858, 951, 'TIRIBA'), (2859, 77, 'CAPIVARA'), (2860, 2241, 'ANDORINHA'), (2861, 2820, 'MINI LION HEDE'), (2862, 21, 'RHODESIAN RIDGEBACK (LEÃO DA RODÉSIA)'), (2863, 1035, 'BUGIO'), (2864, 41, 'CARACÚ'), (2865, 41, 'CRIOULA LAGEANA'), (2867, 912, 'SRD'), (2868, 2826, 'CANASTRA'), (2869, 2826, 'GALINHA'), (2870, 41, 'MESTIÇO'), (2871, 85, 'POLACA'), (2872, 21, 'PASTOR MAREMANO ABRUZÉS'), (2873, 2241, 'BACURAU'), (2874, 2241, 'QUERO - QUERO'), (2875, 10, 'MAINE COON'), (2876, 2045, 'OURIÇO'), (2877, 2241, 'SANHAÇO CINZENTO - TANGARÁ SAYACA'), (2878, 2241, 'BEIJA FLOR'), (2879, 2241, 'CANARIO'), (2880, 77, 'LEBRE'), (2881, 96, 'CAMPEIRO');
```

```
CREATE TABLE `tipo_amostra` (
  `id` int(11) NOT NULL,
  `material` varchar(145) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

```
INSERT INTO `tipo_amostra` (`id`, `material`) VALUES
```

```
(2, 'fezes'),
(1, 'material de aborto'),
(4, 'órgão'),
(14, 'outros'),
(7, 'sangue com anticoagulante'),
(9, 'secreção anal'),
(8, 'secreção auricular'),
(3, 'secreção nasal'),
(11, 'secreção ocular'),
(10, 'secreção oral'),
(13, 'secreção prepucial'),
(12, 'secreção vulvar'),
(6, 'soro'),
(5, 'urina');
```

```
CREATE TABLE `tipo_exame` (
  `id` int(11) NOT NULL,
  `exame` varchar(145) DEFAULT NULL,
```

```

`preco` decimal(10,2) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

INSERT INTO `tipo_exame` (`id`, `exame`, `preco`) VALUES
(1, 'Isolamento e identificação bacteriana', '20.00'),
(2, 'Antibiograma', '20.00'),
(3, 'Cultura e Antibiograma', '20.00'),
(4, 'Contagem Bacteriana', '20.00'),
(5, 'Pesquisa de Salmonella', '45.00'),
(6, 'Pesquisa de Listeria', '60.00'),
(7, 'Pesquisa de Staphylococcus Coagulase Positiva', '45.00'),
(8, 'Pesquisa de Anaeróbicos', '30.00'),
(9, 'Coliformes Totais e Termotolerantes', '60.00'),
(10, 'Isolamento Viral', '100.00'),
(11, 'BHV-1 e 5 (PCR)', '50.00'),
(12, 'FeLV (imunofluorescência indireta)', '50.00'),
(13, 'Leucose Bovina (imunodifusão em gel de ágar)', '15.00'),
(14, 'Vacina Papilomatose', '25.00'),
(15, 'FIV (PCR)', '50.00'),
(16, 'BVD (PCR)\n', '50.00'),
(17, 'BHV-1 (imunofluorescência direta)', '50.00');

ALTER TABLE `amostra`
  ADD PRIMARY KEY (`id`),
  ADD KEY `id_paciente` (`id_paciente`),
  ADD KEY `amostra` (`id_tipo_amostra`);

ALTER TABLE `cliente`
  ADD PRIMARY KEY (`id`),
  ADD KEY `proprietario` (`id`);

ALTER TABLE `especie`
  ADD PRIMARY KEY (`id`);

ALTER TABLE `exame`
  ADD PRIMARY KEY (`id`),
  ADD KEY `material` (`id_tipo_exame`),
  ADD KEY `id_paciente` (`id_paciente`),
  ADD KEY `id_amostra` (`id_amostra`);

ALTER TABLE `paciente`
  ADD PRIMARY KEY (`id`),
  ADD KEY `id_especie` (`id_especie`),
  ADD KEY `id_raca` (`id_raca`),
  ADD KEY `paciente` (`id`),
  ADD KEY `FK_paciente_cliente` (`id_cliente`);

ALTER TABLE `raca`
  ADD PRIMARY KEY (`id`),
  ADD KEY `fk_raca_especie_idx` (`id_especie`);

ALTER TABLE `tipo_amostra`
  ADD PRIMARY KEY (`id`),
  ADD KEY `material` (`material`);

ALTER TABLE `tipo_exame`
  ADD PRIMARY KEY (`id`),
  ADD KEY `material` (`exame`);

ALTER TABLE `amostra`
  ADD CONSTRAINT `amostra_ibfk_1` FOREIGN KEY (`id_tipo_amostra`) REFERENCES `tipo_amostra` (`id`),
  ADD CONSTRAINT `amostra_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`);

ALTER TABLE `exame`
  ADD CONSTRAINT `exame_ibfk_1` FOREIGN KEY (`id_tipo_exame`) REFERENCES `tipo_exame` (`id`),
  ADD CONSTRAINT `exame_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`),
  ADD CONSTRAINT `exame_ibfk_3` FOREIGN KEY (`id_amostra`) REFERENCES `amostra` (`id`);

ALTER TABLE `paciente`
  ADD CONSTRAINT `FK_paciente_cliente` FOREIGN KEY (`id_cliente`) REFERENCES `cliente` (`id`),
  ADD CONSTRAINT `FK_paciente_especie` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`),
  ADD CONSTRAINT `FK_paciente_raca` FOREIGN KEY (`id_raca`) REFERENCES `raca` (`id`);

ALTER TABLE `raca`
  ADD CONSTRAINT `raca_ibfk_1` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`);

/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;

```

Script 1 - Inserção de registros na tabela "cliente" - Cedima

```

# mysql
import pymysql
# random.choice
import random
# sleep
import time
# datas
import datetime
# barra de progresso
import progressbar
# Faker
from faker import Faker
# configura a localização para Português Brasil
fake = Faker('pt_BR')

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='cedima',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de opções de cidades
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    lista_cidades = { "Lages": 10, "Otacílio Costa": 10, "Caxias do Sul": 9, "Painel": 9, "São Joaquim": 8,
                     "Capão Alto": 8, "Campo Belo do Sul": 7, "Bocaina do Sul": 6, "Palmeira": 6, "Correia Pinto": 6, "Ponte Alta": 5, "Urupema": 4, "Florianópolis": 4,
                     "Porto Alegre": 3, "Abdon Batista": 3, "Anita Garibaldi": 2, "Campos Novos": 2, "Vacaria": 1, "Rio do Sul": 1, "Criciúma": 1}

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

            x=1
            # para inserção de "total" amostras
            while(x <= total_inserir):

                # retorna um dos itens da lista, levando em conta o peso
                cidade=random.choice([k for k in lista_cidades for dummy in range(lista_cidades[k])])

                # data de nascimento e cadastro
                cadastro = fake.date_between(start_date="-5y", end_date="now")
                delta = datetime.timedelta(days=365*random.randint(18,50))
                nascimento = cadastro-delta

                # gera um nome fictício, seguindo o padrão pt_BR
                # foi utilizado estes dois métodos porque o fake.name() adiciona
                # um título ao nome, de tempos em tempos
                nome = fake.first_name() + " " + fake.last_name()

                # sql
                sql = "INSERT INTO `cliente` (`nome`, `cidade`, `data_nascimento`,`data_cadastro`) \
                    VALUES (%s,%s,%s,%s)"

                # sql passado com os parâmetros acima
                cursor.execute(sql, (nome, cidade, nascimento, cadastro))
                connection.commit()

                # atualiza a barra de progresso
                time.sleep(0.1)
                bar.update(x)

                # atualiza contador
                x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Script 2 - Inserção de registros na tabela “paciente” - Cedima

```

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar
# Faker
from faker import Faker
# configura a localização para Português Brasil
fake = Faker('pt_BR')

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='cedima',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    lista_nomes_f = ["DIANA","DILA","PITT","SACHA","PRETA","SEM NOME","BELLA","SHERIKAN","MOTINHA","NIKE",
                    "FOFA","BONECA","XUXA","GROW","WOODY","BOBI","VIOLA","TOBI","CHITARA","SAPECA","JUPIRA",
                    "SAM","PEQUENA","SNOOPY","TUCA","LADY","LILO","PEPI","RAISA","KIKA","MAGALI","BABY",
                    "DANDI","TEDI","FAISCA","BABI","DANKA","MIMOSA","FLOR","BINGA","ALEMOA","TUTI"]
    lista_nomes_m = ["BRINQUEDO","FARRAPO","FILHOTE","FALCÃO","SEM NOME","SHERIKAN","MOTINHA","NIKE","TINO","TOBI",
                    "PERIGO","DUQUE","SCUBI","LUKE","JOE","TUCA","LILO","PEPI","NEGO","BABY","DANDI","TEDI",
                    "SULTÃO","FAISCA","AJAX","TICO","BONO","BIDU","KING","BUGUE","ALEMÃO","BRUNO"]

    # lista de opções de cidades
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    # 10 - felina, 21 - canina, 38 - suína, 41 - bovina, 53 - ovina, 62 - caprina, 77 - roedores,
    # 85 - galináceos, 96 - equinos, 906 - muar, 912 - cunicola, 951 - psitacideo, 975 - avicola, 1035 - primata

    lista_especies = {"10": 10, "21": 10, "38": 9, "41": 9, "53": 8, "62": 7, "77": 6, "85": 6,
                     "96": 5, "906": 5, "912": 5, "951": 4, "975": 3, "1035": 2}

    lista_raca_bovina = ["1646","637","624","602","595","641","650","1753","1778","1790","2787","511","580",
                        "2865","484","497","503","567","526","532","548","554","2308","2870","2864","618",
                        "462","478","571"]

    lista_raca_canina = ["100","1292","1345","1211","1240","1254","1336","1324","1266","2391","991","1203",
                        "1196","983","1000","1014","1026","947","1052","1071","2677","2655","2630","1122"]

    lista_raca_felina = ["2847","2404","1513","2765","2481","331","340","355","2688",
                        "363","2875","2383","1874","2372","2132"]

    lista_raca_equinos = ["1672","2644","410","2881","2320","1275","2776","1170","2297","441","433","2699",
                        "386","371","392","2168","401","2850","425"]

    lista_raca_ovina = ["2851","1088","1063","673","2721","2250","711","665","707","696","681"]

    lista_raca_psitacideo = ["2856","2855","2858","2857","2854","2853","2852"]

    lista_raca_roedores = ["1576","1812","2880","1683","2859","2618","813"]

    lista_raca_caprina = ["2574","743","720","735","751"]

    lista_raca_suina = ["788","805","794","766","772"]

    lista_raca_primata = ["2863","1091","1373"]

    lista_raca_avicola = ["1944","1161"]

    lista_raca_cunicola = ["2867"]

    lista_raca_galinaceos = ["2871"]

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

            x=1
            # para inserção de "total" amostras

```

```

while(x <= total_inserir):

    # sexo
    valor = random.randint(0,1)
    if valor == 0:
        sexo = "F"
        nome = random.choice(lista_nomes_f)
    else:
        sexo = "M"
        nome = random.choice(lista_nomes_m)

    # data de nascimento e cadastro
    cadastro = fake.date_between(start_date="-5y", end_date="now")
    delta = datetime.timedelta(days=365*random.randint(2,12))
    nascimento = cadastro-delta

    # retorna um dos itens da lista, levando em conta o peso
    especie=random.choice([k for k in lista_especies for dummy in range(lista_especies[k])])

    #####
    # caso a raça seja "10 - gato"
    if especie == "10":
        # retorna uma raça da lista acima
        raca = random.choice(lista_raca_felina)
        # designa um peso, entre 1 e 15 quilos
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=7)

    # "21 - cão"
    elif especie == "21":
        raca = random.choice(lista_raca_canina)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=25)

    # "38 - suína"
    elif especie == "38":
        raca = random.choice(lista_raca_suina)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=15, max_value=100)

    # "41 - bovina"
    elif especie == "41":
        raca = random.choice(lista_raca_bovina)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=50, max_value=300)

    # "53 - ovina"
    elif especie == "53":
        raca = random.choice(lista_raca_ovina)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=20, max_value=50)

    # "62 - caprina"
    elif especie == "62":
        raca = random.choice(lista_raca_caprina)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=15, max_value=70)

    # "77 - roedores"
    elif especie == "77":
        raca = random.choice(lista_raca_roedores)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=2)

    # "85 - galinaceos"
    elif especie == "85":
        raca = random.choice(lista_raca_galinaceos)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=2)

    # "96 - equinos"
    elif especie == "96":
        raca = random.choice(lista_raca_equinos)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=40, max_value=500)

    # "912 - cunicola"
    elif especie == "912":
        raca = random.choice(lista_raca_cunicola)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=3)

    # "951 - psitacideo"
    elif especie == "951":
        raca = random.choice(lista_raca_psitacideo)
        peso = random.random()

    # "975 - avicola"
    elif especie == "975":
        raca = random.choice(lista_raca_avicola)
        peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=2)

    # "1035 - primata"

```

```

elif especie == "1035":
    raca = random.choice(lista_raca_primata)
    peso = fake.pyfloat(left_digits=None, right_digits=3, positive=True, min_value=1, max_value=15)

else:
    peso = 5

# sql
sql = "INSERT INTO `paciente` (`nome`,`sexo`,`peso`,`id_especie`,`id_raca`,`id_cliente`,`
`data_nascimento`,`data_cadastro`) VALUES (%s,%s,%s,%s,%s,%s,%s,%s)"

# sql passado com os parâmetros acima
cursor.execute(sql, (nome, sexo, peso, especie, raca, x, nascimento, cadastro))
connection.commit()

# atualiza a barra de progresso
time.sleep(0.1)
bar.update(x)

x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Script 3 - Inserção de registros na tabela "amostra" - Cedima

```

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='cedima',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de opções de tipos de amostras para exame
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    lista_amostras = {"14": 10, "13": 10, "12": 9, "11": 9, "10": 8, "9": 8, "8": 7, "7": 6, "6": 6, "5": 6,
                      "4": 5, "3": 4, "2": 3, "1": 2}

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

            x=1
            while(x <= total_inserir):

                # retorna um dos itens da lista, levando em conta o peso
                tipo_amostra=random.choice([k for k in lista_amostras for dummy in range(lista_amostras[k])])

                # data coleta
                data_coleta = fake.date_between(start_date="-5y", end_date="now")

                # sql
                sql = "INSERT INTO `amostra` (`id_tipo_amostra`,`id_paciente`,`data_coleta`,`data`) \
VALUES (%s, %s, %s, %s)"

                # sql passado com os parâmetros acima
                cursor.execute(sql, (tipo_amostra, x, data_coleta, data_coleta))
                connection.commit()

                # atualiza a barra de progresso
                time.sleep(0.1)
                bar.update(x)

                x+=1

```

```
# eof
finally:
    cursor.close()
    connection.close()
```

Script 4 - Inserção de registros na tabela "exame" - Cedima

```
# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar
# faker
from faker import Faker
# configura o faker para localização pt_BR
fake = Faker('pt_BR')

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='cedima',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de opções de tipos de amostras para exame
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    lista_tipo_exame = { "17": 10, "16": 10, "15": 10, "14": 10, "13": 10, "12": 9, "11": 9, "10": 10,
                        "9": 8, "8": 7, "7": 6, "6": 6, "5": 6, "4": 5, "3": 4, "2": 3, "1": 9 }

    # tipo 1 - isolamento e identificação bacteriana - complemento -> nome da bactéria
    # tipo 2 - antibiograma
    # tipo 3 - cultura e antibiograma
    # tipo 4 - contagem bacteriana
    # tipo 5 - pesquisa de salmonella - complemento -> vazio
    # tipo 6 - pesquisa de listeria - complemento -> vazio
    # tipo 7 - pesquisa de staphylococcus coagulase positiva - complemento -> vazio
    # tipo 8 - pesquisa de anaeróbicos
    # tipo 9 - coliformes totais e termotolerantes
    # tipo 10 - isolamento viral
    # tipo 11 - BHV-1 e 5
    # tipo 12 - felv (imunofluorescência indireta)
    # tipo 13 - leucose bovina (imunodifusão em gel de ágar)
    # tipo 14 - vacina papilomatose
    # tipo 15 - fiv(pcr)
    # tipo 16 - bvd (pcr)
    # tipo 17 - bhv-1 (imunofluorescência direta)

    lista_bacterias = {"Leptospira interrogans" : 10,"Vibrio cholerae" : 10,"Rickettsia rickettsii" : 8,
                      "Clostridium botulinum" : 8,"Streptococcus pneumoniae" : 7,"Salmonella enterica" : 6,
                      "Salmonella bongori" : 6,"Clostridium tetani" : 6,"Chlamydia psittaci" : 5,
                      "Pasteurella multocida" : 5,"Pasteurella haemolytica" : 4,"Histophilus somni" : 4,
                      "Clostridium chavoei" : 2,"Clostridium piliforme" : 2,"Escherichia coli" : 1,
                      "Staphylococcus aureus" : 1 }

    lista_virus = { "Lyssavirus" : 10,"Canarypox" : 10,"Porcine circovirus" : 9,"Canine parvovirus" : 9,
                   "Herpesvírus Suídeo" : 8,"Avian avulavirus" : 7,"Aphthovirus" : 7,"Avihepatovirus" : 6,
                   "Cardiovirus" : 5,"Erbovirus" : 5,"Kobuvirus" : 5,"Parechovirus" : 3,"Sapelovirus" : 2,
                   "Senecavirus" : 2,"Teschovirus" : 1,"Tremovirus" : 1 }

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

            x=1
            while(x <= total_inserir):

                # retorna um dos itens da lista, levando em conta o peso
                tipo_exame=random.choice([k for k in lista_tipo_exame for dummy in range(lista_tipo_exame[k])])

                # data exame
```

```

data_exame = fake.date_between(start_date="-5y", end_date="now")

# geracao de resultado (positivo/negativo), com chance de 65% para positivos
resultado = fake.boolean(chance_of_getting_true=65)

# Para resultado "True", insere "positivo" na base
if resultado == True:
    resultado = "positivo"
else:
    resultado = "negativo"

# caso o tipo de exame seja "1 - isolamento e identificação bacteriana", retorna um nome
# aleatório da lista acima, baseado em seu peso
if tipo_exame == "1" and resultado == "positivo":
    complemento = random.choice([k for k in lista_bacterias for dummy in range(lista_bacterias[k])])
else:
    # caso o tipo de exame seja "10 - isolamento vital", retorna um nome
    # aleatório da lista acima, baseado em seu peso
    if tipo_exame == "10" and resultado == "positivo":
        complemento = random.choice([k for k in lista_virus for dummy in range(lista_virus[k])])
    else:
        complemento = ""

# montagem do comando sql
sql = "INSERT INTO `exame` (`id_tipo_exame`, `id_paciente`, `id_amostra`, `data_solicitacao`, \
    `resultado`, `complemento`) VALUES (%s, %s, %s, %s, %s, %s)"

# sql passado com os parâmetros acima
cursor.execute(sql, (tipo_exame, x, x, data_exame, resultado, complemento))
connection.commit()

# atualiza a barra de progresso
time.sleep(0.1)
bar.update(x)

x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

SQL padrão - usado pela maioria dos *scripts* a seguir

```

import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="cedima",
    charset='utf8mb4')

df = pd.read_sql_query("SELECT \
    amostra.data, exame.resultado, exame.data_solicitacao, tipo_exame.preco, \
    tipo_exame.exame, \
    paciente.nome, \
    paciente.sexo, \
    paciente.data_nascimento, CURDATE(), TIMESTAMPDIFF(YEAR, paciente.data_nascimento, CURDATE()) AS 'idade', \
    paciente.peso, \
    paciente.data_cadastro, \
    especie.especie, \
    tipo_amostra.material, \
    raca.raca, \
    cliente.nome, \
    cliente.sexo, \
    cliente.estado, \
    cliente.cidade, \
    cliente.bairro, \
    cliente.data_nascimento, \
    cliente.data_cadastro \
FROM paciente \
INNER JOIN raca \
    ON paciente.id_raca = raca.id \
INNER JOIN cliente \

```

```

    ON paciente.id_cliente = cliente.id \
INNER JOIN amostra \
    ON amostra.id_paciente = paciente.id \
LEFT JOIN especie \
    ON paciente.id_especie = especie.id \
    AND raca.id_especie = especie.id \
INNER JOIN tipo_amostra \
    ON amostra.id_tipo_amostra = tipo_amostra.id \
INNER JOIN exame \
    ON exame.id_paciente = paciente.id \
    AND exame.id_amostra = amostra.id \
INNER JOIN tipo_exame \
    ON exame.id_tipo_exame = tipo_exame.id", \
conn)

```

Tabela 3 - Cedima: total de exames por tipo - Python

```

dfCedima = pd.read_sql_query("SELECT tipo_exame.exame 'Exame', COUNT(exame.id) 'Total' \
    FROM exame INNER JOIN tipo_exame ON exame.id_tipo_exame = tipo_exame.id \
    GROUP BY exame \
    ", \
    conn)

dfCedima.sort_values(by=['Total']).style.hide_index()

```

Figura 23 - Cedima: quantidade de exames, por tipo - Python

```

import seaborn as sns
from matplotlib import pyplot as plt
import numpy as np

# ordena as linhas pela coluna total
dfCedima = dfCedima.sort_values(by='Total')

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('seaborn-whitegrid')
plt.style.use('default')

# tamanho da figura (largura, altura)
plt.figure(figsize=(12,4))

plt.title('Quantidade de Exames')
sns.barplot(x='Total', y='Exame', data=dfCedima)

plt.xlabel('Total de exames')
plt.ylabel('Exames solicitados')

plt.show()

```

Figura 31 - Cedima: distribuição de peso dos gatos - Python

```

dfGatos = pd.read_sql_query("SELECT \
    paciente.nome, \
    paciente.sexo, \
    paciente.data_nascimento, CURDATE(), TIMESTAMPDIFF(YEAR, data_nascimento, CURDATE()) AS 'idade', \
    paciente.peso, \
    raca.raca, \
    amostra.data, \
    especie.especie, \
    tipo_exame.exame, \
    tipo_exame.preco, \
    tipo_amostra.material, \
    exame.resultado \
FROM paciente \
INNER JOIN raca \
    ON paciente.id_raca = raca.id \
INNER JOIN amostra \
    ON amostra.id_paciente = paciente.id \
LEFT JOIN especie \
    ON paciente.id_especie = especie.id \
    AND raca.id_especie = especie.id \
INNER JOIN tipo_amostra \
    ON amostra.id_tipo_amostra = tipo_amostra.id \
INNER JOIN exame \

```

```

ON exame.id_paciente = paciente.id \
AND exame.id_amostra = amostra.id \
INNER JOIN tipo_exame \
ON exame.id_tipo_exame = tipo_exame.id WHERE paciente.id_especie = '10', \
conn)

%matplotlib inline

peso = dfGatos['peso']

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('bmh')

plt.figure(figsize=(6, 4))
plt.hist(peso, bins=range(0,9,1))
plt.title('Distribuição de Peso')
plt.xlabel('Peso')
plt.ylabel('Número de gatos')

```

Figura 32 - Cedima: distribuição de peso dos gatos, por sexo - Python

```

%matplotlib notebook

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('bmh')

sns.distplot(dfGatos['peso'], color='blue', kde=False)
plt.title('Distribuição de peso dos gatos')

```

Figura 34 - Cedima: distribuição de peso dos gatos, por idade - Python

```

# boxplot
# Tipo caixa. Mostra a caixa, que é a amplitude da variável, do dado. Muito útil para ver anomalias (outliners).
# A linha do meio é a mediana. Os dados que ultrapassam são mostrados acima e abaixo. O pontinhos são dados que estão
# distantes da maioria dos valores

%matplotlib inline

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('seaborn-whitegrid')

plt.figure(figsize=(10,7))

sns.boxplot(x='idade', y='peso', data=dfGatos)

```

Figura 35 - Cedima: distribuição de idade dos cães, por raça - Python

```

%matplotlib inline

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('fast')
plt.figure(figsize=(10,7))

# utilizando filtragem das raças
filtroRacas = ['PERDIGUEIRO', 'S.R.D', 'GOLDEN RETRIEVER', 'DINAMARQUES', 'CHOW-CHOW', 'YORKSHIRE']
dfRacas = dfCaes[dfCaes.raca.isin(filtroRacas)]

# Método do Pandas para boxplot. Este não tem as opções de cores, tamanho, etc
#dfRacas.boxplot(by='raca', column=['idade'], fontsize=8)

# Método usando Seaborn. É o melhor

```

```
# opções de palheta de cores: deep, muted, bright, pastel e dark
# width especifica a largura das caixas

# IMPORTANTE
# Bottom black horizontal line of box plot is minimum value
# First black horizontal line of rectangle shape of box plot is First quartile or 25%
# Second black horizontal line of rectangle shape of box plot is Second quartile or 50% or median.
# Third black horizontal line of rectangle shape of box plot is third quartile or 75%
# Top black horizontal line of rectangle shape of box plot is maximum value.
# Small diamond shape of box plot is outlier data or erroneous data.

bplot = sns.boxplot(x='raca', y='idade', data=dfRacas, width=0.5, palette="bright")
bplot.axes.set_title("Relação entre raça e idade dos cães", fontsize=10)
bplot.set_xlabel("Raça", fontsize=10)
bplot.set_ylabel("Idade", fontsize=10)
bplot.tick_params(labelsize=8)
```

Figura 37 - Cedima: localização geográfica de infecções por um agente específico - Jupiter

```
# mapas
import folium

# localização importada do google maps ou openmaps
lages = folium.Map(location=[-27.8148,-50.3119],zoom_start=14)

# clicar no mapa retorna posição atual
lages.add_child(folium.LatLngPopup())

# insere o marcador
# formato
# 27.8048 - vertical, latitude
# 50.3119 - horizontal, longitude

folium.Marker(location=[-27.8083,-50.3006]).add_to(lages)
folium.Marker(location=[-27.8148,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8187,-50.3162]).add_to(lages)
folium.Marker(location=[-27.8148,-50.3219]).add_to(lages)
folium.Marker(location=[-27.8248,-50.3319]).add_to(lages)
folium.Marker(location=[-27.8248,-50.3019]).add_to(lages)
folium.Marker(location=[-27.8348,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8348,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8048,-50.3419]).add_to(lages)
folium.Marker(location=[-27.8058,-50.3419]).add_to(lages)
folium.Marker(location=[-27.8078,-50.3419]).add_to(lages)
folium.Marker(location=[-27.8078,-50.3429]).add_to(lages)
folium.Marker(location=[-27.8088,-50.3439]).add_to(lages)

# invoca o mapa
lages
```

Figura 38 - Cedima: localização geográfica de infecções por um agente específico, com destaque - Python

```
# mapas
import folium

# localização importada do google maps ou openmaps
# opções de tiles: 'Stamen Toner', 'Stamen Terrain' e 'OpenStreetMap' (padrão)
lages = folium.Map(
    location=[-27.8148,-50.3119],
    tiles='OpenStreetMap',
    zoom_start=14
)

lages.add_child(folium.LatLngPopup())

folium.Marker(location=[-27.8083,-50.3006]).add_to(lages)
folium.Marker(location=[-27.8148,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8187,-50.3162]).add_to(lages)
folium.Marker(location=[-27.8148,-50.3219]).add_to(lages)
folium.Marker(location=[-27.8248,-50.3319]).add_to(lages)
folium.Marker(location=[-27.8248,-50.3019]).add_to(lages)
folium.Marker(location=[-27.8348,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8348,-50.3119]).add_to(lages)
folium.Marker(location=[-27.8048,-50.3419]).add_to(lages)
folium.Marker(location=[-27.8078,-50.3419], tooltip='Vírus', icon=folium.Icon(color='red')).add_to(lages)
folium.Marker(location=[-27.8078,-50.3429], tooltip='Vírus', icon=folium.Icon(color='red')).add_to(lages)
folium.Marker(location=[-27.8088,-50.3439], tooltip='Vírus', icon=folium.Icon(color='red')).add_to(lages)
```

```
# adiciona um marcador circular vermelho
folium.CircleMarker(
    radius=80,
    location=[-27.8088,-50.3439],
    popup='Região',
    color='crimson',
    fill=True
).add_to(lages)

# invoca o mapa
lages
```

Figura 39 - Distribuição de materiais recebidos para análise - Python

```
## %matplotlib inline

import seaborn as sns
from matplotlib import pyplot as plt

# Diagrama de Pizza para uma coluna

# Calculos
optv = dfCaes['material']
optv_sum = optv.value_counts()
total = dfCaes['material'].count()
optv_freq = (optv_sum / total) * 100

topicos = ['material de aborto', 'fezes', 'secreção nasal', 'órgão', 'urina', 'soro', 'sangue com anticoagulante',
'sекреção auricular', 'secreção anal', 'secreção oral', 'secreção ocular', 'secreção vulvar',
'secreção prepucial', 'outros']

plt.figure(figsize=(8, 8))
labels = list(topicos)
colors = ['#A43820', '#FF8C6', '#99C68E', '#4DBCD3']
explode = (0, 0, .0, 0, 0, 0, .1, 0, .1, 0, .1, 0, .1, 0, .1, 0)
# Separa uma das fatias de acordo com indice e o valor dado(.1)

# Sobre o gráfico de pizza:
# x: The wedge sizes.
# autopct: Se não for "nenhum", é uma string ou função usado para rotular as cunhas com seu valor numérico.
# colors: Uma seqüência de args de matplotlib por meio da qual o gráfico de pizza será alternado.
# explode: Se não for None, é uma matriz len(x) que especifica a fração do raio com a qual deslocar cada cunha.
# shadow: Se True desenha uma sombra sob o gráfico.

#plt.pie(optv_freq, labels=labels, autopct='%1.1f%%', colors=colors, shadow=True, explode=explode)
plt.pie(optv_freq, labels=labels, autopct='%1.1f%%',
        colors=colors, shadow=True, startangle=90, explode=explode)
plt.show()
```

Quadro 13 - Cedima: listagem de gatos com positivo para Streptococcus pneumoniae - Python

```
import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="cedima",
    charset='utf8mb4')

dfli = pd.read_sql_query("SELECT \
paciente.nome 'Nome', paciente.sexo 'Sexo', paciente.data_nascimento 'Data Nascimento', CURDATE(), \
TIMESTAMPDIFF(YEAR, paciente.data_nascimento, CURDATE()) AS 'Idade', paciente.peso 'Peso', \
exame.resultado Resultado, exame.complemento 'Complemento', especie.especie 'Espécie', raca.raca 'Raça', \
tipo_amostra.material 'Material', cliente.nome 'Nome Cliente', cliente.cidade 'Cidade', \
exame.data_solicitacao 'Data Exame'\
FROM paciente \
INNER JOIN raca \
ON paciente.id_raca = raca.id \
INNER JOIN cliente \
ON paciente.id_cliente = cliente.id \
```

```

INNER JOIN amostra \
  ON amostra.id_paciente = paciente.id \
LEFT JOIN especie \
  ON paciente.id_especie = especie.id \
  AND raca.id_especie = especie.id \
INNER JOIN tipo_amostra \
  ON amostra.id_tipo_amostra = tipo_amostra.id \
INNER JOIN exame \
  ON exame.id_paciente = paciente.id \
  AND exame.id_amostra = amostra.id \
  AND exame.resultado = 'positivo' \
  AND especie.especie = 'felina' \
  AND exame.complemento = 'Streptococcus pneumoniae' \
  AND exame.id_tipo_exame = '1',
conn)

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'Data Nascimento']

dfli.sort_values(by='Nome').style.hide_columns(cols_ocultar).hide_index()

```

Quadro 14 - Cedima: listagem de cães positivos para qualquer agente - Python

```

# função de estilo
def color_negative_red(val):
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()']

# pandas query
df = dfli.query('Idade < 5 and Sexo == "F" and Espécie == "CANINA" and Complemento != ""')

df.style.hide_columns(cols_ocultar).hide_index().applymap(color_negative_red)

```

9 APÊNDICE B - CÓDIGOS-FONTE PARASITOLOGIA: SQL E PYTHON

SQL 1 - Criação do banco de dados “Parasitologia”

```

SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8mb4 */;

CREATE DATABASE IF NOT EXISTS `parasitologia` DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_general_ci;
USE `parasitologia`;

CREATE TABLE `amostra` (
  `id` int(11) NOT NULL,
  `id_tipo_amostra` int(11) DEFAULT NULL,
  `id_paciente` int(11) DEFAULT NULL,
  `data_amostra` date NOT NULL,
  `data` date NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `cliente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) NOT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `cpf` varchar(20) DEFAULT NULL,
  `cnpj` varchar(20) DEFAULT NULL,
  `endereco` varchar(255) DEFAULT NULL,
  `numero` varchar(45) DEFAULT NULL,
  `estado` varchar(2) DEFAULT NULL,
  `cep` varchar(8) DEFAULT NULL,
  `cidade` varchar(145) DEFAULT NULL,
  `bairro` varchar(255) DEFAULT NULL,
  `email` varchar(255) DEFAULT NULL,
  `telefone` varchar(45) NOT NULL,
  `data_nascimento` date NOT NULL,
  `data_cadastro` date DEFAULT NULL,
  `observacoes` varchar(255) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `especie` (
  `id` int(11) NOT NULL,
  `especie` varchar(145) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

-- dados da tabela "especie" idênticos ao da base Cedima. Suprimidos por questão de espaço

CREATE TABLE `exame` (
  `id` int(11) NOT NULL,
  `id_paciente` int(11) NOT NULL,
  `id_amostra` int(11) NOT NULL,
  `data_solicitacao` date NOT NULL,
  `historico` varchar(300) NOT NULL,
  `exame` varchar(300) NOT NULL,
  `resultado` varchar(140) NOT NULL,
  `observações` varchar(120) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `paciente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) DEFAULT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `data_nascimento` date DEFAULT NULL,
  `peso` float(5,2) DEFAULT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `id_raca` int(11) DEFAULT NULL,
  `id_cliente` int(11) DEFAULT NULL,
  `data_cadastro` date DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `raca` (
  `id` int(11) NOT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `raca` varchar(45) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

-- dados da tabela "raca" idênticos ao da base Cedima. Suprimidos por questão de espaço

CREATE TABLE `tipo_amostra` (
  `id` int(11) NOT NULL,
  `material` varchar(145) DEFAULT NULL

```

```

) ENGINE=InnoDB DEFAULT CHARSET=latin1;

INSERT INTO `tipo_amostra` (`id`, `material`) VALUES
(4, 'exemplar'),
(2, 'fezes'),
(5, 'outro'),
(1, 'sangue'),
(3, 'soro');

ALTER TABLE `amostra`
  ADD PRIMARY KEY (`id`),
  ADD KEY `id_paciente` (`id_paciente`),
  ADD KEY `amostra` (`id_tipo_amostra`);

ALTER TABLE `cliente`
  ADD PRIMARY KEY (`id`);

ALTER TABLE `especie`
  ADD PRIMARY KEY (`id`);

ALTER TABLE `exame`
  ADD PRIMARY KEY (`id`),
  ADD KEY `id_paciente` (`id_paciente`),
  ADD KEY `id_amostra` (`id_amostra`);

ALTER TABLE `paciente`
  ADD PRIMARY KEY (`id`),
  ADD KEY `id_especie` (`id_especie`),
  ADD KEY `id_raca` (`id_raca`),
  ADD KEY `paciente` (`id`),
  ADD KEY `FK_paciente_cliente` (`id_cliente`);

ALTER TABLE `raca`
  ADD PRIMARY KEY (`id`),
  ADD KEY `fk_raca_especie_idx` (`id_especie`);

ALTER TABLE `tipo_amostra`
  ADD PRIMARY KEY (`id`),
  ADD KEY `material` (`material`);

ALTER TABLE `amostra`
  ADD CONSTRAINT `amostra_ibfk_1` FOREIGN KEY (`id_tipo_amostra`) REFERENCES `tipo_amostra` (`id`),
  ADD CONSTRAINT `amostra_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`);

ALTER TABLE `exame`
  ADD CONSTRAINT `exame_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`),
  ADD CONSTRAINT `exame_ibfk_3` FOREIGN KEY (`id_amostra`) REFERENCES `amostra` (`id`);

ALTER TABLE `paciente`
  ADD CONSTRAINT `paciente_ibfk_1` FOREIGN KEY (`id_cliente`) REFERENCES `cliente` (`id`),
  ADD CONSTRAINT `paciente_ibfk_2` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`),
  ADD CONSTRAINT `paciente_ibfk_3` FOREIGN KEY (`id_raca`) REFERENCES `raca` (`id`);

ALTER TABLE `raca`
  ADD CONSTRAINT `raca_ibfk_1` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`);

/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;

```

Script 1 - Inserção de registros na tabela "amostra" - Parasitologia

```

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar
# faker
from faker import Faker
fake = Faker()

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='parasitologia',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
  # lista de opções de tipos de amostras para exame

```

```

# o valor ao lado do id é o peso, utilizado pelo comando random.choice
# 1 - sangue, 2 - fezes, 3 - soro, 4 - exemplar, 5 - outro
lista_amostras = { 1: 10, 2: 9, 3: 8, 4: 6, 5: 4 }

# total de registros a serem inseridos
total_inserir = 10000

with connection.cursor() as cursor:

    # barra de progresso
    # valor deve ser igual ao do range do contador
    with progressbar.ProgressBar(max_value=total_inserir) as bar:

        x=1
        while(x <= total_inserir):

            # retorna um dos itens da lista, levando em conta o peso
            tipo_amostra=random.choice([k for k in lista_amostras for dummy in range(lista_amostras[k])])

            # data coleta
            data_coleta = fake.date_between(start_date="-5y", end_date="now")

            # sql
            sql = "INSERT INTO `amostra` (`id_tipo_amostra`, `id_paciente`, `data_amostra`, `data`) \
                VALUES (%s, %s, %s, %s)"

            # sql passado com os parâmetros acima
            cursor.execute(sql, (tipo_amostra, x, data_coleta, data_coleta))
            connection.commit()

            # atualiza a barra de progresso
            time.sleep(0.1)
            bar.update(x)

            x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Script 2 - Inserção de registros na tabela "exame" - Parasitologia

```

# mysql
import pymysql
# random.choice
import random
# datas
from datetime import datetime, timedelta, date
# sleep
import time
# barra de progresso
import progressbar
# faker
from faker import Faker
# configura o faker para localização pt_BR
fake = Faker('pt_BR')

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='parasitologia',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de opções de tipos de amostras para exame
    # o valor ao lado do id é o peso, utilizado pelo comando random.choice
    lista_parasitas = {"Bastocystis sp": 10,"Entamoeba coli": 10,"Entamoeba histolytica": 10,
                      "Entamoeba dispar": 10,"Entamoeba nana": 10,"Blastocystis hominis": 9,
                      "Giardia intestinalis": 9,"Plasmodium falciparum": 8, "Schistosoma mansoni": 8,
                      "Dipilidium caninum": 7,"Dirofilaria immitis": 6,"Chilomastix menili": 6,
                      "Leishmania dovani": 6,"Leishmania braziliensis": 5,"Leishmania peruviana": 4,
                      "Toxoplasma gondii": 2,"Ancylostoma braziliense": 2,"Ancylostoma caninum": 2,
                      "Taenia multiceps": 1}

    # total de registros a serem inseridos
    total_inserir = 10000

    with connection.cursor() as cursor:

        # barra de progresso
        # valor deve ser igual ao do range do contador
        with progressbar.ProgressBar(max_value=total_inserir) as bar:

```

```

x=1
while(x <= total_inserir):

    # data exame
    data_exame = fake.date_between(start_date="-5y", end_date="now")

    # histórico aleatório
    historico = fake.text()

    # geracao de resultado (positivo/negativo), com chance de 65% para positivos
    resultado = fake.boolean(chance_of_getting_true=65)

    # caso True, insere "positivo" na base
    if resultado == True:
        resultado = "positivo"
    else:
        resultado = "negativo"

    # retorna um dos itens da lista, levando em conta o peso
    parasita=random.choice([k for k in lista_parasitas for dummy in range(lista_parasitas[k])])

    # montagem do comando sql
    sql = "INSERT INTO `exame` (`id_paciente`, `id_amostra`, `data_solicitacao`, \
        `historico`, `exame`, `resultado`) \
        VALUES (%s, %s, %s, %s, %s, %s)"

    # sql passado com os parâmetros acima
    cursor.execute(sql, (x, x, data_exame, historico, parasita, resultado))
    connection.commit()

    # atualiza a barra de progresso
    time.sleep(0.1)
    bar.update(x)

    x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Python e SQL padrão - usado pela maioria dos *scripts* a seguir

```

import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="parasitologia",
    charset='utf8mb4')

dfParasitologia = pd.read_sql_query("SELECT \
    amostra.data_amostra, \
    exame.resultado, \
    exame.data_solicitacao, \
    exame.exame, \
    exame.historico, \
    paciente.nome, \
    paciente.sexo, \
    paciente.data_nascimento, CURDATE(), TIMESTAMPDIFF(YEAR, paciente.data_nascimento, CURDATE()) AS 'idade', \
    paciente.peso, \
    paciente.data_cadastro, \
    especie.especie, \
    tipo_amostra.material, \
    raca.raca, \
    cliente.id, \
    cliente.nome, \
    cliente.sexo, \
    cliente.estado, \
    cliente.cidade, \
    cliente.bairro, \
    cliente.data_nascimento, \

```

```

cliente.data_cadastro \
FROM paciente \
INNER JOIN raca \
  ON paciente.id_raca = raca.id \
INNER JOIN cliente \
  ON paciente.id_cliente = cliente.id \
INNER JOIN amostra \
  ON amostra.id_paciente = paciente.id \
LEFT JOIN especie \
  ON paciente.id_especie = especie.id \
  AND raca.id_especie = especie.id \
INNER JOIN tipo_amostra \
  ON amostra.id_tipo_amostra = tipo_amostra.id \
INNER JOIN exame \
  ON exame.id_paciente = paciente.id \
  AND exame.id_amostra = amostra.id", \
conn)

```

Figura 48 - Parasitologia: positivos para Entamoeba dispar, por mês, por ano - Python

```

import janitor

dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2015-01-01"
end = "2015-12-31"

df15 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df15 = df15.query('exame == "Entamoeba dispar" and resultado == "positivo")

df15.index=df15['data_amostra']

```

```

import janitor

dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2016-01-01"
end = "2016-12-31"

df16 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df16 = df16.query('exame == "Entamoeba dispar" and resultado == "positivo")

df16.index=df16['data_amostra']

```

```

import janitor

dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2017-01-01"
end = "2017-12-31"

df17 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df17 = df17.query('exame == "Entamoeba dispar" and resultado == "positivo")

df17.index=df17['data_amostra']

```

```

from matplotlib import pyplot as plt

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('default')

plt.figure(figsize=(10,4))
plt.title('Positivos para Entamoeba dispar por mês')
plt.xlabel('Mês')
plt.ylabel('Total')

```

```
plt.plot(df15.groupby([df15.index.month])[ 'exame'].count(),label='2015')
plt.plot(df16.groupby([df16.index.month])[ 'exame'].count(),label='2016')
plt.plot(df17.groupby([df17.index.month])[ 'exame'].count(),label='2017')
plt.legend()

# ocultar ou mostrar grades
# plt.grid()
plt.show
```

Figura 49 - Parasitologia: total de exames por mês, de 2015 a 2017

```
dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2015-01-01"
end = "2015-12-31"

df15 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df15.index=df15['data_amostra']
```

```
dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2016-01-01"
end = "2016-12-31"

df16 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df16.index=df16['data_amostra']
```

```
dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2017-01-01"
end = "2017-12-31"

df17 = dfJoin.filter_date('data_amostra', start=start, end=end)

# pandas query
df17.index=df17['data_amostra']
```

```
from matplotlib import pyplot as plt

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#   seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#   seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#   seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('default')

plt.figure(figsize=(10,4))
plt.title('Exames por mês')
plt.xlabel('Mês')
plt.ylabel('Total')

plt.plot(df15.groupby([df15.index.month])[ 'exame'].count(),label='2015')
plt.plot(df16.groupby([df16.index.month])[ 'exame'].count(),label='2016')
plt.plot(df17.groupby([df17.index.month])[ 'exame'].count(),label='2017')
plt.legend()

# plt.grid()
plt.show
```

10 APÊNDICE C - CÓDIGOS-FONTE PATOLOGIA ANIMAL: SQL E PYTHON

```

SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8mb4 */;

CREATE DATABASE IF NOT EXISTS `parasitologia` DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_general_ci;
USE `parasitologia`;

CREATE TABLE `amostra` (
  `id` int(11) NOT NULL,
  `id_tipo_amostra` int(11) DEFAULT NULL,
  `id_paciente` int(11) DEFAULT NULL,
  `data_amostra` date NOT NULL,
  `data` date NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `cliente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) NOT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `cpf` varchar(20) DEFAULT NULL,
  `cnpj` varchar(20) DEFAULT NULL,
  `endereco` varchar(255) DEFAULT NULL,
  `numero` varchar(45) DEFAULT NULL,
  `estado` varchar(2) DEFAULT NULL,
  `cep` varchar(8) DEFAULT NULL,
  `cidade` varchar(145) DEFAULT NULL,
  `bairro` varchar(255) DEFAULT NULL,
  `email` varchar(255) DEFAULT NULL,
  `telefone` varchar(45) NOT NULL,
  `data_nascimento` date NOT NULL,
  `data_cadastro` date DEFAULT NULL,
  `observacoes` varchar(255) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `especie` (
  `id` int(11) NOT NULL,
  `especie` varchar(145) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

-- dados da tabela "especie" idênticos ao da base Cedima. Suprimidos por questão de espaço

CREATE TABLE `exame` (
  `id` int(11) NOT NULL,
  `id_paciente` int(11) NOT NULL,
  `id_amostra` int(11) NOT NULL,
  `data_solicitacao` date NOT NULL,
  `historico` varchar(300) NOT NULL,
  `exame` varchar(300) NOT NULL,
  `resultado` varchar(140) NOT NULL,
  `observações` varchar(120) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `paciente` (
  `id` int(11) NOT NULL,
  `nome` varchar(255) DEFAULT NULL,
  `sexo` enum('M','F') DEFAULT NULL,
  `data_nascimento` date DEFAULT NULL,
  `peso` float(5,2) DEFAULT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `id_raca` int(11) DEFAULT NULL,
  `id_cliente` int(11) DEFAULT NULL,
  `data_cadastro` date DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

CREATE TABLE `raca` (
  `id` int(11) NOT NULL,
  `id_especie` int(11) DEFAULT NULL,
  `raca` varchar(45) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

-- dados da tabela "raca" idênticos ao da base Cedima. Suprimidos por questão de espaço

CREATE TABLE `tipo_amostra` (
  `id` int(11) NOT NULL,
  `material` varchar(145) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

```

INSERT INTO `tipo_amostra` (`id`,`material`) VALUES
(4,'exemplar'),
(2,'fezes'),
(5,'outro'),
(1,'sangue'),
(3,'soro');

ALTER TABLE `amostra`
ADD PRIMARY KEY (`id`),
ADD KEY `id_paciente` (`id_paciente`),
ADD KEY `amostra` (`id_tipo_amostra`);

ALTER TABLE `cliente`
ADD PRIMARY KEY (`id`);

ALTER TABLE `especie`
ADD PRIMARY KEY (`id`);

ALTER TABLE `exame`
ADD PRIMARY KEY (`id`),
ADD KEY `id_paciente` (`id_paciente`),
ADD KEY `id_amostra` (`id_amostra`);

ALTER TABLE `paciente`
ADD PRIMARY KEY (`id`),
ADD KEY `id_especie` (`id_especie`),
ADD KEY `id_raca` (`id_raca`),
ADD KEY `paciente` (`id`),
ADD KEY `FK_paciente_cliente` (`id_cliente`);

ALTER TABLE `raca`
ADD PRIMARY KEY (`id`),
ADD KEY `fk_raca_especie_id` (`id_especie`);

ALTER TABLE `tipo_amostra`
ADD PRIMARY KEY (`id`),
ADD KEY `material` (`material`);

ALTER TABLE `amostra`
ADD CONSTRAINT `amostra_ibfk_1` FOREIGN KEY (`id_tipo_amostra`) REFERENCES `tipo_amostra` (`id`),
ADD CONSTRAINT `amostra_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`);

ALTER TABLE `exame`
ADD CONSTRAINT `exame_ibfk_2` FOREIGN KEY (`id_paciente`) REFERENCES `paciente` (`id`),
ADD CONSTRAINT `exame_ibfk_3` FOREIGN KEY (`id_amostra`) REFERENCES `amostra` (`id`);

ALTER TABLE `paciente`
ADD CONSTRAINT `paciente_ibfk_1` FOREIGN KEY (`id_cliente`) REFERENCES `cliente` (`id`),
ADD CONSTRAINT `paciente_ibfk_2` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`),
ADD CONSTRAINT `paciente_ibfk_3` FOREIGN KEY (`id_raca`) REFERENCES `raca` (`id`);

ALTER TABLE `raca`
ADD CONSTRAINT `raca_ibfk_1` FOREIGN KEY (`id_especie`) REFERENCES `especie` (`id`);

/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;

```

Script 1 - Inserção de registros na tabela "material" - Patologia

```

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='patologia',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    # lista de materiais possíveis. O lab analisa qualquer tecido
    lista_materiais = { "ESTÔMAGO", "CÉREBRO", "PULMÃO ESQUERDO", "PULMÃO DIREITO", "RIM DIREITO", "RIM ESQUERDO",
                       "FÊMUR ESQUERDO", "RÁDIO DIREITO", "RÁDIO ESQUERDO", "OLHO ESQUERDO", "OLHO DIREITO",
                       "FÍGADO", "CORÇÃO", "BAÇO", "PÂNCREAS", "RÍMEM", "MÚSCULO", "SISTEMA NERVOSO CENTRAL",
                       "ABOMASO", "INTESTINO GROSSO", "INTESTINO DELGADO", "TRAQUÉIA", "GENITAIS", "GÔNADAS",
                       "LÍNGUA", "ORELHA ESQUERDA", "ORELHA DIREITA", "ESÔFAGO", "BEXIGA", "SUPRA-RENAIS", "TIREÓIDE",

```

```

"MEMBRO INFERIOR ESQUERDO", "MEMBRO INFERIOR DIREITO", "MEMBRO SUPERIOR ESQUERDO",
"MEMBRO SUPERIOR DIREITO" }

# total de registros a serem inseridos
total_inserir = 10000

with connection.cursor() as cursor:

    # barra de progresso
    # valor deve ser igual ao do range do contador
    with progressbar.ProgressBar(max_value=total_inserir) as bar:

        x=1
        # para inserção de "total" amostras
        while(x <= total_inserir):

            # valor randomico entre 1 e 7
            # quantidade de materiais recebidos para exame
            quantidade = random.randint(1,7)

            # retorna a quantidade acima
            material = (str(random.sample(lista_materiais,k=quantidade)).strip("[]")).replace("'", "")

            # data material
            data_material = fake.date_between(start_date="-5y", end_date="now")

            # sql
            sql = "INSERT INTO `material` (`material`, `id_paciente`, `data_material`) \
                VALUES (%s, %s, %s)"

            # sql passado com os parâmetros acima
            cursor.execute(sql, (material, x, data_material))
            connection.commit()

            # atualiza a barra de progresso
            time.sleep(0.1)
            bar.update(x)

            x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Script 2 - Inserção de registros na tabela "exame" - Patologia

```

# mysql
import pymysql
# random.choice
import random
# datas
import datetime
# sleep
import time
# barra de progresso
import progressbar
# faker
from faker import Faker
# configura o faker para localização pt_BR
fake = Faker('pt_BR')

# conexão ao mysql
connection = pymysql.connect(host='192.168.56.101',user='root',password='root',
                             db='patologia',charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:

    lista_historico = {"o animal apresentou sangramento pelo nariz e ponta das orelhas",
                       "a proprietária observou diarreia sanguinolenta","o animal não respondeu ao tratamento",
                       "foi encaminhado ao hospital", "não era vacinado","estava se alimentando e tomando água",
                       "criado com alimentação caseira","criado com ração",
                       "animal em bom estado geral e bom estado de conservação",
                       "lesões circunstritas avermelhadas e com ausência de pêlo na região da cabeça e das patas",
                       "na região da cabeça, ao corte, apresenta área avermelhada e crepitante",
                       "o cadáver encontrava-se em bom estado nutricional",
                       "orifícios naturais: as mucosas mostraram-se esmaecidas (pálidas)",
                       "tecido subcutâneo e musculatura: hemorragia do tipo sufusão",
                       "tecido subcutâneo: hemorragia músculo obliquo abdominal direito",

```

```
"ruptura da fascia muscular abdominal", "cadáver sem congelamento prévio",
"tetraparesia espástica dos membros do lado direito", "hemograma não apresentou alterações",
"exame sorológico para toxoplasmose não obteve resultados", "seborréia seca",
"áreas de tricotomia nos membros anteriores", "mucosas se apresentavam normocoradas",
"presença de sangue na cavidade nasal" }
```

```
lista_necropsia = { "estado corporal bom",
"cavidade peritoneal: presença de moderada quantidade (100 mililitros) de conteúdo sanguinolento",
"cavidade peritoneal: contendo coágulo cruóric", "baço: apresentou-se levemente aumentado de volume",
"cérebro apresenta coloração avermelhada", "hipostáse na região do plexo braquial do membro torácico esquerdo",
"hematoma na parte cervical esquerda", "Edema na parte interna próximo ao membro inferior esquerdo",
"baso crepitante ao corte", "Vasos dilatados e congestos no intestino", "presença de vermes no intestino",
"hemorragia ileo-ceco acerosa", "musculatura básica corporal apresenta áreas com coloração pálida",
"baço: com extensa ruptura de sua extremidade caudal", "baco: áreas de hemorragia subcapsulares multifocais",
"baço: ao corte, deixou fluir discreta quantidade de sangue",
"fígado e vesícula biliar: o fígado mostrou-se discretamente aumentado de volume",
"fígado: com bordas suavemente abauladas, de consistência levemente friável",
"fígado: com superfície externa discretamente irregular",
"fígado: mostrou-se evidenciação de sua lobulação e de tonalidade difusamente pálido-amarelada",
"fígado: exibindo hemorragias no parênquima",
"vesícula biliar: distendida contendo fluido biliar de aspecto seroso, amarelo-amarronzado",
"rins: bilateralmente, revelaram superfícies discretamente avermelhadas",
"rins: hemorragias petequiais de distribuição aleatória",
"rins: ao corte dos dois antímeros, as regiões córtico-medulares e medulares apresentavam tonalidade levemente avermelhada",
"bexiga: encontrada retraída", "bexiga: sem conteúdo urinoso", "pleura: pequena quantidade (50 mililitros) de conteúdo sanguinolento",
"pleura: contendo coágulo cruóric",
"pulmões: lobos pulmonares com superfícies externas, em quase toda a extensão, de aspecto liso, úmido, de tonalidade difusa",
"pulmões: levemente avermelhado", "pulmões: todos os lobos apresentavam áreas hemorragias",
"pulmões: ao corte, sob pressão manual, deixou fluir discreta quantidade de conteúdo espumoso de tonalidade avermelhada",
"ossos: vértebras da região lombo-sacra apresentaram fratura extensa", "sufusões subcutâneas e musculares",
"hemoperitônio", "ruptura esplênica", "esteatose difusa moderada", "hemorragias hepáticas multifocais",
"hemotórax", "hemorragia pulmonar", "congestão e edema pulmonar" }
```

```
lista_histologia = { "megalocitose de hepatócitos", "fibrose periportal difusa", "hemorragia centrolobular",
"baço: rarefação linfóide moderada", "baço: com infiltrado de histiócitos",
"baço: hemossiderose discreta",
"coração: raros vacúolos parasitóforos entre cardiomiócitos que são característicos do protozoário Rangelia vitalii",
"coração: focos de miocardite mista discreta",
"fígado: deposição de tecido conjuntivo na região periportal", "fígado: focos de hepatite portal mista de grau leve",
"fígado: colestase moderada", "fígado: vacuolização discreta de hepatócitos",
"rim: nefrite intersticial mononuclear moderada", "rim: pigmentos acastanhados intra-tubulares",
"rim: espessamento da cápsula de Bowman", "cérebro: edema perivascular discreto",
"pulmão: enfisema crônico moderado em bordos",
"adrenal: sem alterações microscópicas", "cerebelo: sem alterações microscópicas", "intestino grosso: sem alterações microscópicas",
"intestino delgado: sem alterações microscópicas", "estômago: sem alterações microscópicas" }
```

```
lista_diagnostico = { "TRAUMATISMO causado PROVAVELMENTE por choque de grande impacto, resultando em perda sanguínea significativa e aguda",
"choque hipovolêmico", "Rangeliose canina", "eutanasia", "cinomose", "doença periodontal", "parvovirose",
"câncer", "leptospirose", "hepatite", "Erlíchiose", "Insuficiência renal", "Otite",
"alergia alimentar", "Raiva" }
```

```
# total de registros a serem inseridos
total_inserir = 10000
```

```
with connection.cursor() as cursor:
```

```
# barra de progresso
# valor deve ser igual ao do range do contador
with progressbar.ProgressBar(max_value=total_inserir) as bar:
```

```
x=1
# para inserção de "total" amostras
while(x <= total_inserir):

# valor randômico entre 1 e 5
# quantidade de de frases retornadas da lista
quantidade = random.randint(1,5)

# retorna a quantidade acima
# removendo colchetes e apóstrofos
historico = (str(random.sample(lista_historico,k=quantidade)).strip('[]')).replace("'", "")
necropsia = (str(random.sample(lista_necropsia,k=quantidade)).strip('[]')).replace("'", "")
histologia = (str(random.sample(lista_histologia,k=quantidade)).strip('[]')).replace("'", "")
diagnostico = (str(random.sample(lista_diagnostico,k=1)).strip('[]')).replace("'", "")
```

```
# diagnóstico: frase aleatória
#diagnostico = fake.sentence()
```

```
# data exame
data_exame = fake.date_between(start_date="-5y", end_date="now")
```

```
# montagem do comando sql
sql = "INSERT INTO `exame` (`id_animal`, `id_material`, `data_solicitacao`, `historico`, \
```

```

        `necropsia`, `histologia`, `diagnostico`) VALUES (%s, %s, %s, %s, %s, %s, %s)"

# sql passado com os parâmetros acima
cursor.execute(sql, (x, x, data_exame, historico, necropsia, histologia, diagnostico))
connection.commit()

# atualiza a barra de progresso
time.sleep(0.1)
bar.update(x)

x+=1

# eof
finally:
    cursor.close()
    connection.close()

```

Python e SQL padrão - usado pela maioria dos *scripts* a seguir

```

import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="patologia",
    charset='utf8mb4')

df = pd.read_sql_query("SELECT \
    paciente.nome, paciente.sexo, paciente.data_nascimento, CURDATE(), TIMESTAMPDIFF(YEAR, paciente.data_nascimento, \
    CURDATE()) AS 'idade', paciente.peso, paciente.data_cadastro, especie.especie, raca.raca, \
    material.data_material, exame.historico, exame.data_solicitacao, exame.necropsia, \
    exame.histologia, exame.diagnostico \
FROM paciente \
INNER JOIN raca \
    ON paciente.id_raca = raca.id \
INNER JOIN cliente \
    ON paciente.id_cliente = cliente.id \
INNER JOIN material \
    ON material.id_paciente = paciente.id \
LEFT JOIN especie \
    ON paciente.id_especie = especie.id \
    AND raca.id_especie = especie.id \
INNER JOIN exame \
    ON exame.id_paciente = paciente.id \
    AND exame.id_material = material.id \
", \
conn)

```

Quadro 15 - Patologia: exemplo de relatório com dados e diagnóstico – Python

```

# lista de colunas a ocultar
cols_ocultar = ['CURDATE()', 'data_nascimento', 'data_cadastro', 'data_solicitacao']

# pandas query
df.head(5).style.hide_columns(cols_ocultar).hide_index()

```

11 APÊNDICE D - CÓDIGOS-FONTE CEDIMA X PARASITOLOGIA: SQL E PYTHON

Quadro 16 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para *Blastocystis hominis* e Cedima quaisquer resultados – Python

```

import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("SELECT \
    cedimaPaciente.nome 'Nome', cedimaPaciente.sexo 'Sexo', \
    cedimaRaca.raca 'Raça', \
    cedimaPaciente.data_nascimento 'Nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, CURDATE()) 'Idade', \
    cedimaPaciente.peso 'Peso', \
    parasitologiaExame.data_solicitacao 'Parasitologia: Data Exame', \
    parasitologiaExame.resultado 'Parasitologia: Resultado', \
    parasitologiaExame.exame 'Parasitologia: Exame', \
    cedimaExame.data_solicitacao 'Cedima: Data Exame', \
    cedimaTipoExame.exame 'Cedima: Exame', \
    cedimaExame.resultado 'Cedima: Resultado', \
    cedimaExame.complemento 'Cedima: complemento' \
FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
WHERE cedimaRaca.id = 100 \
AND parasitologiaExame.exame = 'Blastocystis hominis' \
AND parasitologiaExame.resultado = 'positivo' \
", \
conn)

# renderizar o resultado ocultando uma coluna e destacando os maiores valores
# Styler.apply(func, axis=0) para valores em colunas
# Styler.apply(func, axis=1) para valores em linhas
# Styler.apply(func, axis=None) para valores globais

# função de estilo. Busca o valor "positivo" e colore em vermelho
def color_negative_red(val):
    """
    Takes a scalar and returns a string with
    the css property ``color: red`` for negative
    strings, black otherwise.
    """
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'Nascimento']

dfJoin.drop(cols_ocultar, axis=1).style.highlight_max(axis=0, subset=['Peso', 'Idade']).hide_index().applymap(color_negative_red)
dfJoin.drop(cols_ocultar, axis=1).style.applymap(color_negative_red).hide_index()

```

Quadro 17 - Cedima x Parasitologia: cães S.R.D. Parasitologia positivos para *Schistosoma mansoni* e Cedima positivos - Python

```

import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("""SELECT \
    cedimaPaciente.nome 'Nome', cedimaPaciente.Sexo, cedimaPaciente.data_nascimento, \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, CURDATE()) 'Idade', \
    cedimaPaciente.peso 'Peso', \
    cedimaRaca.raca 'Raça', \
    parasitologiaExame.exame 'Parasitologia: exame', \
    parasitologiaExame.resultado 'Parasitologia: resultado', \
    parasitologiaExame.data_solicitacao 'Parasitologia: Data Exame', \
    parasitologiaExame.data_solicitacao 'Parasitologia: Data Coleta', \
    cedimaExame.data_solicitacao 'Cedima: Data Exame', \
    cedimaTipoExame.exame 'Cedima: Exame', \
    cedimaExame.resultado 'Cedima: Resultado' \
    FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
    WHERE cedimaRaca.id = 100 \
    AND parasitologiaExame.exame = 'Schistosoma mansoni' \
    AND parasitologiaExame.resultado = 'positivo' \
    AND cedimaExame.resultado = 'positivo' \
    ", \
    conn)

# renderizar o resultado, ocultando uma coluna, usando o método "hide_columns"
# hide_index() -> oculta o índice
# sort_values() -> ordena pela coluna selecionada

# função de estilo
def color_negative_red(val):
    """
    Recebe um valor escalar e retorna uma string com
    a propriedade css "color: red" para true
    e preto, caso contrário
    """
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'data_nascimento', 'Parasitologia: Data Coleta']

dfJoin.sort_values(by='Cedima: Exame').style.hide_columns(cols_ocultar).hide_index().applymap(color_negative_red)

```

Quadro 18 - Cedima x Parasitologia: cães, raça perdigueiro, positivos para isolamento e identificação bacteriana, Cedima, e que constam na parasitologia, independente do resultado –

Python

```
import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("""SELECT \
    cedimaPaciente.nome 'Nome', cedimaPaciente.Sexo, \
    cedimaPaciente.data_nascimento 'Data de nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, CURDATE()) 'Idade', \
    cedimaPaciente.peso 'Peso', \
    cedimaRaca.raca 'Raça', \
    parasitologiaExame.exame 'Parasitologia: exame', \
    parasitologiaExame.resultado 'Parasitologia: resultado', \
    parasitologiaExame.data_solicitacao 'Parasitologia: Data Exame', \
    parasitologiaExame.data_solicitacao 'Parasitologia: Data coleta', \
    cedimaExame.data_solicitacao 'Cedima: Data Exame', \
    cedimaTipoExame.exame 'Cedima: exame', \
    cedimaExame.resultado 'Cedima: Resultado', \
    cedimaExame.complemento 'Complemento' \
    FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
    INNER JOIN cedima.especie cedimaEspecie \
    ON cedimaEspecie.id = cedimaRaca.id_especie \
    AND cedimaEspecie.especie = 'canina' \
    AND cedimaRaca.raca = 'perdigueiro' \
    AND cedimaTipoExame.exame = 'Isolamento e Identificação Bacteriana' \
    AND cedimaExame.resultado = 'positivo' \
    AND parasitologiaExame.resultado = 'positivo' \
    ", \
    conn)

# renderizar o resultado, ocultando uma coluna, usando o método "hide_columns"
# hide_index() -> oculta o índice
# sort_values() -> ordena pela coluna selecionada

# função de estilo
def color_negative_red(val):
    """
    Recebe um valor escalar e retorna uma string com
    a propriedade css "color: red" para true
    e preto, caso contrário
    """
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'Data de nascimento', 'Parasitologia: Data coleta']

dfJoin.sort_values(by='Parasitologia: exame').style.hide_columns(cols_ocultar).hide_index().applymap(color_negative_red)
```

Quadro 19 - Cedima x Parasitologia: gatos, com 16 anos ou menos, sexo feminino, com menos de 2 quilos e da raça Havana e que foram atendidos nos dois laboratórios - Python

```

import os
import pymysql
import pandas as pd
import time

from datetime import datetime

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("""SELECT \
    cedimaPaciente.nome AS 'Nome', \
    cedimaPaciente.sexo AS 'Sexo', cedimaPaciente.data_nascimento AS 'Nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, \
    CURDATE()) AS 'Idade', \
    cedimaEspecie.especie 'Espécie', \
    cedimaRaca.raca 'Raça', \
    cedimaPaciente.peso 'Peso', \
    parasitologiaExame.data_solicitacao 'Parasitologia Data do Exame', \
    parasitologiaExame.resultado 'Parasitologia Resultado', \
    parasitologiaExame.exame 'Exame', \
    cedimaExame.data_solicitacao 'Cedima Data do Exame', \
    cedimaExame.resultado AS 'Cedima Resultado', \
    cedimaExame.complemento AS 'Agente', \
    cedimaTipoExame.exame AS 'Cedima Exame' \
    FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.especie cedimaEspecie \
    ON cedimaEspecie.id = cedimaRaca.id_especie \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
    ", \
    conn)

# função de estilo
def color_negative_red(val):
    """
    Recebe um valor escalar e retorna uma string com
    a propriedade css `color: red` para true
    e preto, caso contrário
    """
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'Nascimento', 'Parasitologia: Data coleta']

# replacing blank spaces with '_'
dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

# pandas query
df = dfJoin.query('Idade >= 16 and Sexo == "M" and Espécie == "FELINA" and and Peso > 2 and RACA == "HAVANA"')

# replacing '_' with blank spaces
df.columns =[column.replace("_", " ") for column in df.columns]

df.style.hide_columns(cols_ocultar).hide_index().applymap(color_negative_red)

```

Quadro 20 - Cedima x Parasitologia: cães, no Cedima, com 11 anos ou mais, machos, com peso maior que 1 quilo e na Parasitologia com positivo para isolamento viral e qualquer resultado de exame

```
# função de estilo
def color_negative_red(val):
    """
    Recebe um valor escalar e retorna uma string com
    a propriedade css `color: red` para true
    e preto, caso contrário
    """
    color = 'red' if val == 'positivo' else 'black'
    return 'color: %s' % color

# lista de colunas à ocultar
cols_ocultar = ['CURDATE()', 'Nascimento', 'Parasitologia: Data coleta']

# replacing blank spaces with '_'
dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

# pandas query
df = dfJoin.query('Idade >= 11 and Sexo == "M" and Espécie == "CANINA" and Cedima_Resultado == "positivo" \
    and Peso > 1 and Cedima_Exame == "Isolamento Viral")

# replacing '_' with blank spaces
df.columns =[column.replace("_", " ") for column in df.columns]

df.style.hide_columns(cols_ocultar).hide_index().applymap(color_negative_red)
```

Figura 50 - Cedima x Parasitologia: *heatmap* de pesos versus idades de equinos - Python

```
# pandas query
df = dfJoin.query('Sexo == "M" and Espécie == "EQUINOS")

%matplotlib inline

import seaborn as sns
from matplotlib import pyplot as plt

correlacoes = df[['Idade', 'Peso']].corr()
sns.heatmap(correlacoes, annot=True, cmap="YlGnBu")
```

```
# pandas query
df = dfJoin.query('Sexo == "F" and Espécie == "EQUINOS")

%matplotlib inline

import seaborn as sns
from matplotlib import pyplot as plt

correlacoes = df[['Idade', 'Peso']].corr()
sns.heatmap(correlacoes, annot=True, cmap="YlGnBu")
```

Figura 51 - Cedima x Parasitologia: exames por mês, ano 2017

```
import janitor

dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2017-01-01"
end = "2017-12-31"

dfParasitologia = dfJoin.filter_date('data_amostra', start=start, end=end)

dfParasitologia.index=dfParasitologia['data_amostra']
```

```
import janitor

dfJoin.columns =[column.replace(" ", "_") for column in dfJoin.columns]

start = "2017-02-01"
```

```
end = "2017-12-31"
```

```
dfCedima = dfJoin.filter_date('data_solicitacao', start=start, end=end)
dfCedima.index=dfCedima['data_solicitacao']
```

```
from matplotlib import pyplot as plt
```

```
# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#         seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#         seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#         seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('default')

plt.figure(figsize=(10,4))
plt.title('Exames por mês - ano de 2017')
plt.xlabel('Mês')
plt.ylabel('Total')

plt.plot(dfParasitologia.groupby([dfParasitologia.index.month])[ 'exame'].count(),label='Parasitologia')
plt.plot(dfCedima.groupby([dfCedima.index.month])[ 'exame'].count(),label='Cedima')
plt.legend()
plt.grid()
plt.show
```

12 APÊNDICE E - CÓDIGOS-FONTE CEDIMA, PARASITOLOGIA E PATOLOGIA ANIMAL: SQL E PYTHON

Quadro 21 - Cedima x Parasitologia x Patologia: cães, yorkshire, antibiograma no Cedima, quaisquer exames Parasitologia e relação com o baço, na Patologia

```
import os
import pymysql
import pandas as pd
import time

from datetime import datetime

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("SELECT \
    cedimaPaciente.nome AS 'Nome', \
    cedimaPaciente.sexo AS 'Sexo', cedimaPaciente.data_nascimento AS 'Nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, \
    CURDATE()) AS 'Idade', \
    cedimaEspecie.especie as 'Especie', \
    cedimaRaca.raca as 'Raca', \
    cedimaPaciente.peso as 'Peso', \
    parasitologiaExame.data_solicitacao AS 'Parasitologia Data do Exame', \
    parasitologiaExame.exame AS 'Parasitologia Exame', \
    parasitologiaExame.resultado AS 'Parasitologia Resultado', \
    cedimaExame.data_solicitacao AS 'Cedima Data do Exame', \
    cedimaTipoExame.exame AS 'Cedima Exame', \
    cedimaExame.resultado AS 'Cedima Resultado', \
    cedimaExame.complemento AS 'Complemento', \
    patologiaExame.data_solicitacao AS 'Patologia Data', \
    patologiaExame.historico AS 'Patologia histórico', \
    patologiaExame.necropsia AS 'Patologia Necropsia', \
    patologiaExame.histologia AS 'Patologia Histologia', \
    patologiaExame.diagnostico AS 'Patologia diagnóstico' \
FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.especie cedimaEspecie \
    ON cedimaEspecie.id = cedimaRaca.id_especie \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
    INNER JOIN \
    patologia.paciente patologiaPaciente \
    ON cedimaPaciente.id = patologiaPaciente.id \
    INNER JOIN \
    patologia.exame patologiaExame \
    ON patologiaPaciente.id = patologiaExame.id_paciente \
WHERE patologiaExame.necropsia LIKE '%baço%' \
    and cedimaEspecie.especie = 'CANINA' \
    and cedimaRaca.raca = 'YORKSHIRE' \
    and cedimaExame.id_tipo_exame = 2 \
", \
conn)

# renderizar o resultado ocultando uma coluna
cols_ocultar = ['CURDATE()', 'Nascimento', 'Peso', 'Parasitologia Data do Exame', 'Idade', 'Cedima Data do Exame',
    'Patologia Data']
```

```
dfJoin.drop(cols_ocultar, axis=1).style.hide_index()
```

Quadro 22 - Cedima x Parasitologia x Patologia: felinos, SRD, positivos para FeLV no Cedima, qualquer resultado Parasitologia e com anotações sobre os rins, na Patologia

```
import os
import pymysql
import pandas as pd
import time

from datetime import datetime

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("SELECT \
    cedimaPaciente.nome AS 'Nome', \
    cedimaPaciente.sexo AS 'Sexo', cedimaPaciente.data_nascimento AS 'Nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, \
    CURDATE()) AS 'Idade', \
    cedimaEspecie.especie as 'Especie', \
    cedimaRaca.raca as 'Raca', \
    cedimaPaciente.peso as 'Peso', \
    parasitologiaExame.data_solicitacao AS 'Parasitologia Data do Exame', \
    parasitologiaExame.exame AS 'Parasitologia Exame', \
    parasitologiaExame.resultado AS 'Parasitologia Resultado', \
    cedimaExame.data_solicitacao AS 'Cedima Data do Exame', \
    cedimaTipoExame.exame AS 'Cedima Exame', \
    cedimaExame.resultado AS 'Cedima Resultado', \
    cedimaExame.complemento AS 'Complemento', \
    patologiaExame.data_solicitacao AS 'Patologia Data', \
    patologiaExame.historico AS 'Patologia histórico', \
    patologiaExame.necropsia AS 'Patologia Necropsia', \
    patologiaExame.histologia AS 'Patologia Histologia', \
    patologiaExame.diagnostico AS 'Patologia diagnóstico' \
    FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.especie cedimaEspecie \
    ON cedimaEspecie.id = cedimaRaca.id_especie \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \
    INNER JOIN cedima.tipo_exame cedimaTipoExame \
    ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
    INNER JOIN \
    patologia.paciente patologiaPaciente \
    ON cedimaPaciente.id = patologiaPaciente.id \
    INNER JOIN \
    patologia.exame patologiaExame \
    ON patologiaPaciente.id = patologiaExame.id_paciente \
    WHERE patologiaExame.histologia LIKE "%rim%" \
    and cedimaEspecie.especie = 'felina' \
    and cedimaRaca.raca = 'SRD' \
    and cedimaExame.id_tipo_exame = 12 \
    ", \
    conn)

# renderizar o resultado ocultando uma coluna
cols_ocultar = ['CURDATE()', 'Nascimento', 'Peso', 'Parasitologia Data do Exame', 'Idade', 'Cedima Data do Exame',
    'Patologia Data']

dfJoin.drop(cols_ocultar, axis=1).style.hide_index()
```

Figura 52 - Cedima, Parasitologia x Patologia: comparação da quantidade de exames por mês, em 2017 – Python

```
import os
import pymysql
import pandas as pd
import janitor

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="cedima",
    charset='utf8mb4')

dfCedima = pd.read_sql_query("SELECT exame.* FROM exame", conn)

start = "2017-01-01"
end = "2017-12-31"

dfCedima = dfCedima.filter_date('data_solicitacao', start=start, end=end)
dfCedima.index=dfCedima['data_solicitacao']
```

```
import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="parasitologia",
    charset='utf8mb4')

dfParasitologia = pd.read_sql_query("SELECT exame.* FROM exame", conn)

start = "2017-01-01"
end = "2017-12-31"

dfParasitologia = dfParasitologia.filter_date('data_solicitacao', start=start, end=end)
dfParasitologia.index=dfParasitologia['data_solicitacao']
```

```
import os
import pymysql
import pandas as pd

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    db="patologia",
    charset='utf8mb4')

dfPatologia = pd.read_sql_query("SELECT exame.* FROM exame", conn)

start = "2017-01-01"
end = "2017-12-31"

dfPatologia = dfPatologia.filter_date('data_solicitacao', start=start, end=end)
dfPatologia.index=dfPatologia['data_solicitacao']
```

```

from matplotlib import pyplot as plt

# definir um estilo para o plot
# opções: default, classic, _classic_test, bmh, dark_background, fast, fivethirtyeight, ggplot, grayscale, seaborn,
#         seaborn-bright, seaborn-colorblind, seaborn-dark, seaborn-dark-palette, seaborn-dark-grid, seaborn-deep
#         seaborn-muted, seaborn-notebook, seaborn-paper, seaborn-pastel, seaborn-poster, seaborn-talk, seaborn-ticks
#         seaborn-white, seaborn-whitegrid, seaborn-colorblind10
plt.style.use('default')

plt.figure(figsize=(10,4))
plt.title('Quantidade de exames por mês - ano de 2017')
plt.xlabel('Mês')
plt.ylabel('Total')

plt.plot(dfParasitologia.groupby([dfParasitologia.index.month])[ 'resultado' ].count(),label='Parasitologia')
plt.plot(dfCedima.groupby([dfCedima.index.month])[ 'resultado' ].count(),label='Cedima')
plt.plot(dfPatologia.groupby([dfPatologia.index.month])[ 'historico' ].count(),label='Patologia')
plt.legend()
plt.grid()
plt.show

```

Quadro 23 - Cedima, Parasitologia x Patologia: bovinos da raça holandês, Cedima, positivos para herpes bovina do tipo 1 e 5 e Patologia Animal algum achado relacionado ao fígado -

Python

```

import os
import pymysql
import pandas as pd
import time

from datetime import datetime

port = os.getenv('MYSQL_PORT')
user = os.getenv('MYSQL_USER')
password = os.getenv('MYSQL_PASSWORD')
database = os.getenv('MYSQL_DATABASE')

conn = pymysql.connect(
    host="192.168.56.101",
    port=int(3306),
    user="root",
    passwd="root",
    charset='utf8mb4')

dfJoin = pd.read_sql_query("SELECT \
    cedimaPaciente.nome AS 'Nome', \
    cedimaPaciente.sexo AS 'Sexo', cedimaPaciente.data_nascimento AS 'Nascimento', \
    CURDATE(), TIMESTAMPDIFF(YEAR, cedimaPaciente.data_nascimento, \
    CURDATE()) AS 'Idade', \
    cedimaEspecie.especie as 'Especie', \
    cedimaRaca.raca as 'Raca', \
    cedimaPaciente.peso as 'Peso', \
    parasitologiaExame.data_solicitacao AS 'Parasitologia Data do Exame', \
    parasitologiaExame.exame AS 'Parasitologia Exame', \
    parasitologiaExame.resultado AS 'Parasitologia Resultado', \
    cedimaExame.data_solicitacao AS 'Cedima Data do Exame', \
    cedimaTipoExame.exame AS 'Cedima Exame', \
    cedimaExame.resultado AS 'Cedima Resultado', \
    cedimaExame.complemento AS 'Complemento', \
    patologiaExame.data_solicitacao AS 'Patologia Data', \
    patologiaExame.historico AS 'Patologia histórico', \
    patologiaExame.necropsia AS 'Patologia Necropsia', \
    patologiaExame.histologia AS 'Patologia Histologia', \
    patologiaExame.diagnostico AS 'Patologia diagnóstico' \
    FROM cedima.paciente cedimaPaciente \
    INNER JOIN \
    parasitologia.paciente parasitologiaPaciente \
    ON cedimaPaciente.id = parasitologiaPaciente.id \
    INNER JOIN \
    parasitologia.exame parasitologiaExame \
    ON parasitologiaPaciente.id = parasitologiaExame.id_paciente \
    INNER JOIN cedima.raca cedimaRaca \
    ON cedimaPaciente.id_raca = cedimaRaca.id \
    INNER JOIN cedima.especie cedimaEspecie \
    ON cedimaEspecie.id = cedimaRaca.id_especie \
    INNER JOIN cedima.exame cedimaExame \
    ON cedimaPaciente.id = cedimaExame.id_paciente \

```

```
INNER JOIN cedima.tipo_exame cedimaTipoExame \
  ON cedimaExame.id_tipo_exame = cedimaTipoExame.id \
INNER JOIN \
patologia.paciente patologiaPaciente \
  ON cedimaPaciente.id = patologiaPaciente.id \
INNER JOIN \
patologia.exame patologiaExame \
  ON patologiaPaciente.id = patologiaExame.id_paciente \
WHERE patologiaExame.histologia LIKE '%figado%' \
  and cedimaEspecie.especie = 'bovina' \
  and cedimaExame.id_tipo_exame = 11 \
  and cedimaRaca.raca = 'holandes' \
", \
conn)
```

renderizar o resultado ocultando uma coluna

```
cols_ocultar = ['CURDATE()', 'Nascimento', 'Peso', 'Parasitologia Data do Exame', 'Cedima Data do Exame',  
              'Patologia Data', 'Complemento']
```

```
dfJoin.drop(cols_ocultar, axis=1).style.hide_index()
```

13 ANEXO A - AUTORIZAÇÃO DE PESQUISA - PARASITOLOGIA

UNIVERSIDADE DO PLANALTO CATARINENSE

CARTA DE AUTORIZAÇÃO

Eu, Anderson Barbosa de Moura, Coordenador do Laboratório de Parasitologia e Doenças Parasitárias, portador do CPF nº [REDACTED], tenho ciência e autorizo a realização da pesquisa intitulada **MEMÓRIA ORGANIZACIONAL E OS DADOS INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO**, sob responsabilidade do pesquisador **NEY KASSIANO RAMOS** no Laboratório de Parasitologia e Doenças Parasitárias (CAV-UDESC).

Para isto, serão disponibilizados ao pesquisador acesso aos modelos de documentos utilizados no laboratório, relacionados ao cadastro de usuários, amostras e/ou de resultados e aos seus correspondentes, caso se utilize um *software* para tal, e ao seu modelo de banco de dados, com o objetivo de verificar como e quais dados são gerenciados atualmente no laboratório e suas especificidades.

O pesquisador acima qualificado se compromete a:

- 1- Obedecer às disposições éticas de proteger os dados da pesquisa;
- 2- Responsabilizar-se pelos procedimentos que assegurem a confidencialidade e a privacidade de quaisquer pessoas ou entidades envolvidas no estudo;
- 3- Não utilizar dados pessoais no estudo;
- 4- Utilizar somente a descrição dos dados e seus tipos e dos procedimentos para sua gerência.

Esta instituição está ciente de suas responsabilidades como instituição coparticipante da presente pesquisa e de seu compromisso no resguardo da segurança das informações, dispondo de infraestrutura necessária para a garantia de tal segurança.

Lages, 18 de Fevereiro de 2019.

Anderson Barbosa de Moura
Coordenador Laboratório
de Parasitologia

14 ANEXO B - AUTORIZAÇÃO DE PESQUISA - HCV

UNIVERSIDADE DO PLANALTO CATARINENSE

CARTA DE AUTORIZAÇÃO

Eu, Letícia Andreza Yonezawa, Coordenadora do Hospital de Clínicas Veterinárias, portadora do CPF nº [REDACTED], tenho ciência e autorizo a realização da pesquisa intitulada **MEMÓRIA ORGANIZACIONAL E OS DADOS INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO**, sob responsabilidade do pesquisador **NEY KASSIANO RAMOS** no Hospital de Clínica Veterinária (CAV-UDESC).

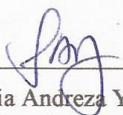
Para isto, serão disponibilizados ao pesquisador acesso aos modelos de documentos utilizados no hospital, relacionados ao cadastro de usuários, amostras e/ou de resultados e aos seus correspondentes, caso se utilize um *software* para tal, e ao seu modelo de banco de dados, com o objetivo de verificar como e quais dados são gerenciados atualmente no hospital e suas especificidades.

O pesquisador acima qualificado se compromete a:

- 1- Obedecer às disposições éticas de proteger os dados da pesquisa;
- 2- Responsabilizar-se pelos procedimentos que assegurem a confidencialidade e a privacidade de quaisquer pessoas ou entidades envolvidas no estudo;
- 3- Não utilizar dados pessoais no estudo;
- 4- Utilizar somente a descrição dos dados e seus tipos e dos procedimentos para sua gerência.

Esta instituição está ciente de suas responsabilidades como instituição coparticipante da presente pesquisa e de seu compromisso no resguardo da segurança das informações, dispondo de infraestrutura necessária para a garantia de tal segurança.

Lages, 26 de Fevereiro de 2019



Letícia Andreza Yonezawa
Coordenadora do Hospital de Clínica
Veterinária CAV-UDESC

15 ANEXO C - AUTORIZAÇÃO DE PESQUISA - PATOLOGIA

UNIVERSIDADE DO PLANALTO CATARINENSE

CARTA DE AUTORIZAÇÃO

Eu, Aldo Gava, Coordenador do Laboratório de Patologia Animal, portador do CPF nº [REDACTED], tenho ciência e autorizo a realização da pesquisa intitulada **MEMÓRIA ORGANIZACIONAL E OS DADOS INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO**, sob responsabilidade do pesquisador **NEY KASSIANO RAMOS** no Laboratório de Patologia Animal (CAV-UDESC).

Para isto, serão disponibilizados ao pesquisador acesso aos modelos de documentos utilizados no laboratório, relacionados ao cadastro de usuários, amostras e/ou de resultados e aos seus correspondentes, caso se utilize um *software* para tal, e ao seu modelo de banco de dados, com o objetivo de verificar como e quais dados são gerenciados atualmente no laboratório e suas especificidades.

O pesquisador acima qualificado se compromete a:

- 1- Obedecer às disposições éticas de proteger os dados da pesquisa;
- 2- Responsabilizar-se pelos procedimentos que assegurem a confidencialidade e a privacidade de quaisquer pessoas ou entidades envolvidas no estudo;
- 3- Não utilizar dados pessoais no estudo;
- 4- Utilizar somente a descrição dos dados e seus tipos e dos procedimentos para sua gerência.

Esta instituição está ciente de suas responsabilidades como instituição coparticipante da presente pesquisa e de seu compromisso no resguardo da segurança das informações, dispondo de infraestrutura necessária para a garantia de tal segurança.

Lages, 20 de Dezembro de 2018.

Aldo Gava
Coordenador Laboratório
de Patologia Animal

16 ANEXO D - AUTORIZAÇÃO DE PESQUISA - CEDIMA

UNIVERSIDADE DO PLANALTO CATARINENSE

CARTA DE AUTORIZAÇÃO

Eu, Ubirajara Maciel da Costa, Coordenador do Centro de Diagnóstico Microbiológico Animal (CEDIMA), portador do CPF nº [REDACTED], tenho ciência e autorizo a realização da pesquisa intitulada **MEMÓRIA ORGANIZACIONAL E OS DADOS INTERLABORATORIAIS NA GERAÇÃO DE ATIVOS DE CONHECIMENTO**, sob responsabilidade do pesquisador **NEY KASSIANO RAMOS** no laboratório CEDIMA (CAV-UDESC).

Para isto, serão disponibilizados ao pesquisador acesso aos modelos de documentos utilizados no laboratório, relacionados ao cadastro de usuários, amostras e/ou de resultados e aos seus correspondentes, caso se utilize um *software* para tal, e ao seu modelo de banco de dados, com o objetivo de verificar como e quais dados são gerenciados atualmente no laboratório e suas especificidades.

O pesquisador acima qualificado se compromete a:

- 1- Obedecer às disposições éticas de proteger os dados da pesquisa;
- 2- Responsabilizar-se pelos procedimentos que assegurem a confidencialidade e a privacidade de quaisquer pessoas ou entidades envolvidas no estudo;
- 3- Não utilizar dados pessoais no estudo;
- 4- Utilizar somente a descrição dos dados e seus tipos e dos procedimentos para sua gerência.

Esta instituição está ciente de suas responsabilidades como instituição coparticipante da presente pesquisa e de seu compromisso no resguardo da segurança das informações, dispondo de infraestrutura necessária para a garantia de tal segurança.

Lages, 20 de Dezembro de 2018.

Ubirajara Maciel da Costa
Coordenador CEDIMA

17 ANEXO E - Serial Software Tableau



Hi Ney,

Please retain this email for your records. You will need the product key and instructions enclosed to install Tableau Desktop and Tableau Prep.

Welcome to Tableau for Students! Your academic license now includes Tableau Desktop and Tableau's newest product, **Tableau Prep**. Tableau Prep makes it easy for you to confidently and visually clean, shape and model data for quick analysis.

The product key below can be used to activate Tableau Desktop and Tableau Prep on two separate machines, Windows or Mac. This key will expire in one year.

- [Download Tableau Desktop](#)
- [Download Tableau Prep](#)
- Activate with your product key: **[REDACTED]**
- If you're receiving the error "the product key is invalid" visit the [knowledge base page](#) to resolve your issue
- Already have a copy of Tableau Desktop or Tableau Prep installed? Update your license in the application: *Help Menu -> Manage Product Keys*

This key is for your personal use only. Please do not share it.

Once Tableau is installed, get started with [free training videos](#) and visit the dedicated [student resource page](#). Need additional help? Check out the [FAQ's](#) or [submit a case](#) for installation support.

Best,

Tableau Academic Team
Tableau Software

e: tft@tableau.com