

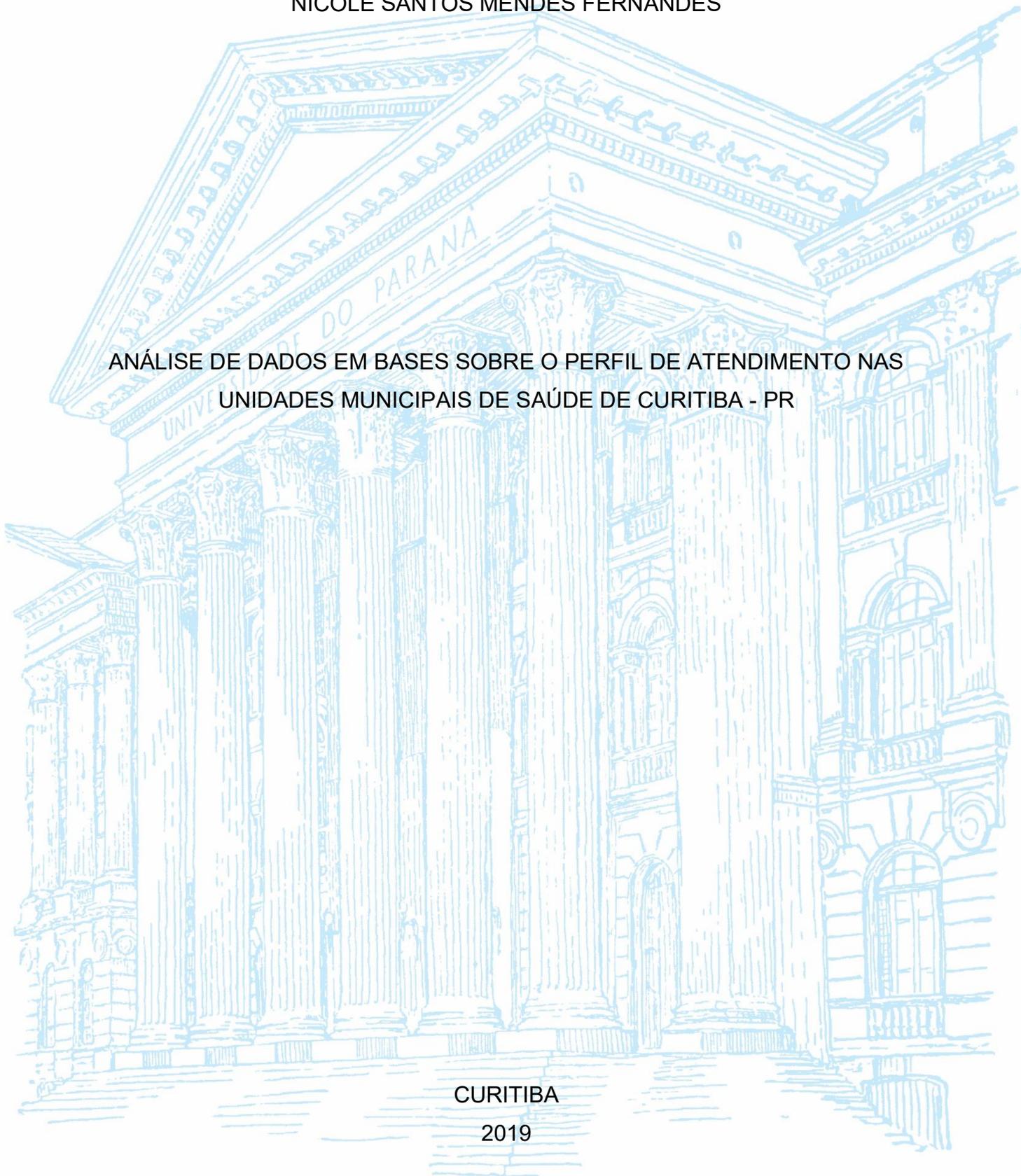
UNIVERSIDADE FEDERAL DO PARANÁ

NICOLE SANTOS MENDES FERNANDES

ANÁLISE DE DADOS EM BASES SOBRE O PERFIL DE ATENDIMENTO NAS
UNIDADES MUNICIPAIS DE SAÚDE DE CURITIBA - PR

CURITIBA

2019



NICOLE SANTOS MENDES FERNANDES

ANÁLISE DE DADOS EM BASES SOBRE O PERFIL DE ATENDIMENTO NAS
UNIDADES MUNICIPAIS DE SAÚDE DE CURITIBA - PR

Trabalho de Conclusão de Curso
apresentado ao curso de Graduação em
Gestão da Informação, Setor de Ciências
Sociais Aplicadas, Universidade Federal do
Paraná, como requisito parcial à obtenção do
título de Bacharel em Gestão da Informação.

Orientadora: Profa. Dra. Denise Fukumi
Tsunoda

CURITIBA

2019

AGRADECIMENTOS

Em todos os momentos, mesmo que fisicamente me sinta desta forma, nunca estou sozinha, pois sempre há a presença de Deus, portanto, agradeço à Ele por todas as bênçãos e desafios recebidos ao longo desta jornada.

Aos meus pais pelo apoio, por acreditarem em meu potencial e pelo suporte a cada decisão; sou muito grata por ser filha de vocês.

À professora Denise pela paciência, orientação e inspiração, não apenas durante as disciplinas de Trabalho de Conclusão I e II, mas em toda a graduação. Obrigada por tudo.

Aos meus amigos e colegas pelos momentos de descontração e companheirismo ao longo dos períodos alegres e difíceis na Universidade. A presença de vocês enriquece a minha vida. Por fim, gratidão à UFPR por ter me proporcionando uma educação de qualidade e a oportunidade de cursar o ensino superior.

“If you can dream it, you can do it”

- Walt Disney

RESUMO

Apresenta pesquisa explicativa sobre análise de dados em base do Sistema E-Saúde. Visa identificar padrões que permitam a descoberta de conhecimento em base sobre o perfil de atendimento dos profissionais médicos da rede municipal de saúde de Curitiba. Objetiva classificar os atendimentos pelos capítulos da Classificação Internacional de Doenças (CID) e utilizar o algoritmo Apriori para a geração de regras de associação, J48 (C4.5) para árvores de decisão e Decision Table para regras de classificação. Expõe comparativo entre os meses de janeiro a setembro de 2018. Elucida os motivos pela escolha da melhor heurística, com os resultados relacionados à descoberta de padrões por meio das regras de associação e regras de classificação com a Classificação Internacional de Doenças (CID) como atributo meta. Conclui com indicativo para trabalhos futuros e sugestões de outras ferramentas para a Mineração de Dados.

Palavras-chave: Sistema E-Saúde. Análise de Dados. Descoberta de Conhecimento em Bases de Dados. Perfil de Atendimento dos Profissionais Médicos. Rede Municipal de Saúde de Curitiba.

ABSTRACT

It presents explanatory research on data analysis based on the E-Health System. It aims to identify patterns that allow the discovery of knowledge based on the profile of care of the medical professionals of the municipal health network of Curitiba. The objective of this study was to classify care by the chapters of the International Classification of Diseases (ICD) and to use the Apriori algorithm for the generation of association rules, J48 (C4.5) for decision trees and Decision Table for classification rules. It exposes comparative between the months of January to September of 2018. It elucidates the reasons for the choice of the best heuristic, with the results related to the discovery of standards through the rules of association and classification rules with the International Classification of Diseases (ICD) as an attribute goal. Concludes with indicative for future work and suggestions for other tools for Data Mining.

Keywords: E-Saude system. Data Analysis. Knowledge discovery in Databases. Medical professionals ' attendance profile. Curitiba Municipal Health Network.

LISTA DE FIGURAS

FIGURA 1 - PESQUISA EM PORTUGUÊS NA WEB OF SCIENCE (ISI Web of Knowledge)	19
FIGURA 2 - PESQUISA EM INGLÊS NA WEB OF SCIENCE (ISI Web of Knowledge)	20
FIGURA 3 - OS “NÍVEIS HIERÁRQUICOS” DA INFORMAÇÃO.....	25
FIGURA 4 - MODELO PROPOSTO PARA REPRESENTAR O FLUXO DA INFORMAÇÃO NAS ORGANIZAÇÕES	27
FIGURA 5 - O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS.....	34
FIGURA 6 - ÁRVORE DE DECISÃO PARA DADOS DE CLASSIFICAÇÃO DA GRADUAÇÃO	38
FIGURA 7 - EXEMPLO DE REDE NEURAL	39
FIGURA 8 - EXEMPLO DE CLUSTERIZAÇÃO (AGRUPAMENTO) DE DADOS	40
FIGURA 9 – EXTRATO DA TABELA DE INDICADORES - TIPO DE ESTABELECIMENTO CENTRO DE SAUDE/UNIDADE BASICA	56
FIGURA 10 – EXTRATO DA TABELA DE INDICADORES - FLUXO DE CLIENTELA - ATENDIMENTO DE DEMANDA ESPONTANEA - ESTADO - PR	56
FIGURA 11 - ESTATÍSTICAS DA BASE 1 DO MÊS DE FEVEREIRO/2018 ...	62
FIGURA 12 - EXEMPLO DE BOXPLOT.....	68
FIGURA 13 - BOXPLOT ATRIBUTO 'comodos' 01FEV18.....	69
FIGURA 14 - BOXPLOT ATRIBUTO 'prescritaCuritibana' 01FEV18	69
FIGURA 15 - BOXPLOT ATRIBUTO 'dispensadaCuritibana' 01FEV18.....	70
FIGURA 16 - BOXPLOT ATRIBUTO 'medicamentoNPadronizado' 01FEV18 .	70
FIGURA 19 - DASHBOARD CONSULTA (jan-set/2018).....	105
FIGURA 20 - DASHBOARD INTERNAMENTO (jan-set/2018)	106
FIGURA 21 - DASHBOARD CARACTERÍSTICAS SÓCIO ECONÔMICAS (jan-set/2018)	108
FIGURA 22 - VISÃO INICIAL SOFTWARE WEKA	112
FIGURA 23 - HISTOGRAMA DOS DADOS NO WEKA	112

FIGURA 24 - PARÂMETROS <i>DEFAULT</i> APRIORI.....	114
FIGURA 25 - RESULTADOS APRIORI - <i>DEFAULT</i> (jan/2018).....	116
FIGURA 26 - RESULTADOS APRIORI EXPERIMENTO 1 (jan/2018).....	117
FIGURA 27 - REGRAS APRIORI EXPERIMENTO 1 (jan/2018).....	117
FIGURA 28 - ATRIBUTOS REMOVIDOS PARA EXPERIMENTO 2.....	118
FIGURA 29 - RESULTADOS DO EXPERIMENTO 2 APRIORI (jan/2018)	118
FIGURA 30 - RESULTADOS DO EXPERIMENTO 3 APRIORI (jan/2018)	120
FIGURA 31 - RESULTADOS DO EXPERIMENTO 3 APRIORI (fev/2018)	120
FIGURA 32 - RESULTADOS DO EXPERIMENTO 3 APRIORI (mar/2018)...	121
FIGURA 33 - RESULTADOS DO EXPERIMENTO 3 APRIORI (set/2018)	122
FIGURA 34 - PARÂMETROS <i>DEFAULT</i> J48.....	123
FIGURA 35 - RESULTADOS DO EXPERIMENTO <i>DEFAULT</i> J48 (jan/2018)	126
FIGURA 36 - ÁRVORE DE DECISÃO <i>DEFAULT</i> J48 (jan/2018)	127
FIGURA 37 - RESULTADOS DO EXPERIMENTO 2 J48 (jan/2018)	128
FIGURA 38 - RESULTADOS DO EXPERIMENTO 3 J48 (jan/2018)	128
FIGURA 39 - RESULTADOS DO EXPERIMENTO 4 J48 (jan/2018)	128
FIGURA 40 - ACURÁCIA DETALHADA POR CLASSE 4-FOLD J48 (jun/2018)	131
FIGURA 41 - MATRIZ DE CONFUSÃO 4-FOLD J48 (jun/2018).....	132
FIGURA 42 - PARÂMETROS <i>DEFAULT</i> ALGORITMO DECISION TABLE ..	133
FIGURA 43 - RESULTADOS EXPERIMENTO <i>DEFAULT</i> DECISION TABLE (jan/2018).....	134
FIGURA 44 - RESULTADOS EXPERIMENTO 1 DECISION TABLE (jan/2018)	135
FIGURA 45 - REGRAS EXPERIMENTO 1 DECISION TABLE (jan/2018).....	135
FIGURA 46 - RESULTADOS EXPERIMENTO 2 DECISION TABLE (jan/2018)	136
FIGURA 47 - RESULTADOS EXPERIMENTO 3 DECISION TABLE (jan/2018)	136
FIGURA 48 - RESULTADOS EXPERIMENTO 1 DECISION TABLE (fev/2018)	137
FIGURA 49 - RESULTADOS EXPERIMENTO 2 DECISION TABLE (fev/2018)	138

FIGURA 50 - RESULTADOS EXPERIMENTO 3 DECISION TABLE (fev/2018)	
.....	138
FIGURA 51 - ACURÁCIA DETALHADA POR CLASSE 4-FOLD DECISION TABLE (jul/2018)	
.....	140
FIGURA 52 - MATRIZ DE CONFUSÃO 4-FOLD DECISION TABLE (jul/2018)	
.....	141
FIGURA 53 - REGRAS EXPERIMENTO 4-FOLD DECISION TABLE (jul/2018)	
.....	141
FIGURA 54 - RESULTADO 4-FOLD J48 (jun/2018)	
.....	143
FIGURA 55 - ÁRVORE J48 (jun/2018)	
.....	144
.....	155

LISTA DE QUADROS

QUADRO 1 - PESQUISAS EM PORTUGUÊS E INGLÊS NA BIBLIOTECA BRASILEIRA DE TESES E DISSERTAÇÕES (BDTD)	20
QUADRO 2 - O VALOR DA INFORMAÇÃO PARA AS ORGANIZAÇÕES	26
QUADRO 3 - VANTAGENS E DESVANTAGENS DOS MODELOS CLÁSSICOS DE RECUPERAÇÃO DA INFORMAÇÃO	30
QUADRO 4 - DADOS DE CLASSIFICAÇÃO DE GRADUAÇÃO	37
QUADRO 5 - DICIONÁRIO DE DADOS: SISTEMA E-SAÚDE MÉDICOS	45
QUADRO 6 - NOMENCLATURA BASES DE DADOS	50
QUADRO 7 - ATRIBUTOS NOMINAIS	52
QUADRO 8 - AJUSTE DO ATRIBUTO 'codTipoUnid'	55
QUADRO 9 - PADRONIZAÇÃO DO ATRIBUTO 'tratamentoDomicilio'	56
QUADRO 10 - PADRONIZAÇÃO DO ATRIBUTO 'abastecimento'	56
QUADRO 11 - PADRONIZAÇÃO DO ATRIBUTO 'tipoHabitacao'	57
QUADRO 12 - PADRONIZAÇÃO DO ATRIBUTO 'destinoLixo'	57
QUADRO 13 - PADRONIZAÇÃO DO ATRIBUTO 'fezesUrina'	58
QUADRO 14 - PADRONIZAÇÃO DO ATRIBUTO 'grupoComunitario'	58
QUADRO 15 - PADRONIZAÇÃO DO ATRIBUTO 'casoDoenca'	58
QUADRO 16 – MORBIDADE HOSPITALAR DO SUS CID-10 CCAPÍTULOS.	59
QUADRO 17 - REPETIÇÃO DOS MESES DE ATENDIMENTO	71
QUADRO 18 - DEFINIÇÃO DAS BASES ANALISADAS E MINERADAS	72

LISTA DE TABELAS

TABELA 1 - REGISTROS ANTES E DEPOIS DO FILTRO "CURITIBA".....	52
TABELA 2 - ESTATÍSTICAS (parte um).....	63
TABELA 2 - ESTATÍSTICAS (parte dois).....	64
TABELA 2 - ESTATÍSTICAS (parte três).....	65
TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (jan-mar/2018).....	74
TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (abr-jun/2018).....	74
TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (jul-set/2018).....	75
TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (jan-mar/2018).....	75
TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (abr-jun2018).....	76
TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (jul-set/2018).....	76
TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (jan-mar/2018).....	76
TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (abr-jun/2018).....	77
TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (jul-set/2018).....	77
TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (jan-mar/2018).....	78
TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (abr-jun/2018).....	78
TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (jul-set/2018).....	79
TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (jan-mar/2018).....	79
TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (abr-jun/2018).....	80
TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (jul-set/2018).....	80
TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (jan-mar/2018).....	81
TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (abr-jun/2018).....	81
TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (jul-set/2018).....	81
TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (jan-mar/2018) ..	82
TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (abr-jun/2018) ...	82
TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (jul-set/2018).....	82
TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (jan-mar/2018).....	82
TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (abr-jun/2018).....	83
TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (jul-set/2018).....	83
TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (jan-mar/2018).....	84
TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (abr-jun/2018).....	84

TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (jul-set/2018)	85
TABELA 12 - FREQUÊNCIA ATRIBUTO 'capCID' (jan-mar/2018).....	85
TABELA 26 - FREQUÊNCIA ATRIBUTO 'capCID' (abr-jun/2018).....	86
TABELA 26 - FREQUÊNCIA ATRIBUTO 'capCID' (jul-set/2018).....	86
TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (jan-mar/2018).....	87
TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (abr-jun/2018).....	87
TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (jul-set/2018).....	87
TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtlInternamento' (jan-mar/2018).....	87
TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtlInternamento' (abr-jun/2018).....	88
TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtlInternamento' (jul-set/2018).....	88
TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (jan-mar/2018).....	88
TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (abr-jun/2018).....	89
TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (jul-set/2018).....	89
TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (jan-mar/2018).....	90
TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (abr-jun/2018).....	90
TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (jul-set/2018).....	91
TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (jan-mar/2018).....	91
TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (abr-jun/2018).....	92
TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (jul-set/2018).....	92
TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (jan-mar/2018).....	93
TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (abr-jun/2018).....	94
TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (jul-set/2018).....	94
TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (jan-mar/2018).....	95
TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (abr-jun/2018).....	95
TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (jul-set/2018).....	95
TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (jan-mar/2018).....	96
TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (abr-jun/2018).....	96
TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (jul-set/2018).....	96
TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (jan-mar/2018).....	97
TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (abr-jun/2018).....	97

TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (jul-set/2018).....	97
TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (jan-mar/2018).....	98
TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (abr-jun/2018).....	98
TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (jul-set/2018)	98
TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (jan-mar/2018).....	99
TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (abr-jun/2018).....	99
TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (jul-set/2018)	100
TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (jan-mar/2018)	100
TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (abr-jun/2018)	101
TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (jul-set/2018).....	101
TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (jan-mar/2018)	102
TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (abr-jun/2018).	102
TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (jul-set/2018) ..	102
TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (jan-mar/2018).....	103
TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (abr-jun/2018).....	103
TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (jul-set/18)	103
TABELA 27 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORTIMO J48 (jan/2018).....	129
TABELA 28 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO J48 (fev/2018).....	129
TABELA 29 - COMPARATIVO DOS RESULTADOS ENTRE OS MESES ANALISADOS COM O ALGORITMO J48	130
TABELA 30 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO DECISION TABLE (jan/2018).....	137
TABELA 31 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO DECISION TABLE (fev/2018).....	139
TABELA 32 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS ENTRE OS MESES ANALISADOS COM O ALGORITMO DECISION TABLE	139

SUMÁRIO

1 INTRODUÇÃO	15
1.1 DEFINIÇÃO DE PROBLEMA.....	16
1.2 OBJETIVOS.....	18
1.3 JUSTIFICATIVA.....	19
1.4 DELIMITAÇÃO DA PESQUISA	22
1.5 ESTRUTURA DO DOCUMENTO	23
2 REVISÃO DE LITERATURA	24
2.1 GESTÃO DA INFORMAÇÃO	24
2.2 RECUPERAÇÃO DA INFORMAÇÃO	29
2.3 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (<i>KNOWLEDGE DISCOVERY IN DATABASES - KDD</i>).....	31
2.4 ANÁLISE DE DADOS	35
2.5 MINERAÇÃO DE DADOS.....	35
2.6 SISTEMA ÚNICO DE SAÚDE	41
3 METODOLOGIA	43
3.1 CARACTERIZAÇÃO DA PESQUISA.....	43
3.2 AMBIENTE DE PESQUISA	44
3.3 DICIONÁRIO DE DADOS	44
3.4 COMITÊ DE ÉTICA	48
4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS	50
4.1 PRÉ-PROCESSAMENTO.....	50
4.1.1 Tratamento das Bases de Dados	52
4.1.2 Estatística Descritiva.....	61
4.2 MINERAÇÃO DE DADOS.....	111
4.2.1 Apriori	113
4.2.2 J48.....	122
4.2.3 Decision Table.....	133
4.3 PÓS-PROCESSAMENTO	143

5 CONSIDERAÇÕES FINAIS	146
REFERÊNCIAS	150
APÊNDICE A - Amostra de outros gráficos do tipo Boxplot (2FEV18)	154
APÊNDICE B - Códigos para atributos nominais	156
APÊNDICE C - Lista de bairros encontrados nas bases	168
APÊNDICE D - Portal de Dados Abertos	170
APÊNDICE E - TRECHO DO RESULTADO DO ALGORITMO J48 (JUN2018) 171	
APÊNDICE F - TRECHO DO RESULTADO DO ALGORITMO DECISION TABLE (JUL2018)	184

1 INTRODUÇÃO

Desde os primórdios, a saúde é tema central na vida das pessoas; inicialmente por um motivo óbvio, uma vez que os ancestrais tinham como foco um objetivo primário, a sobrevivência; esta era essencial diante das adversidades e obstáculos do passado. Ao trazer (para a discussão) e tratar a definição de saúde, no decorrer dos anos, refere-se à um conceito que sofreu alterações, pois como explica Scliar (2007) ele segue a conjuntura social, econômica, política e cultural e, dessa forma, difere para cada situação e realidade. Atualmente, a promoção da saúde, prevenção de doenças, busca pela qualidade de vida e bem-estar, relacionam-se ao conceito supracitado.

No Brasil, além de soluções privadas, a saúde conta, conforme explica o Ministério da Saúde, com “um dos maiores e mais complexos sistemas de saúde pública do mundo[...]. Engloba a atenção básica, média e alta complexidades, os serviços urgência e emergência, a atenção hospitalar, as ações e serviços das vigilâncias epidemiológica, sanitária e ambiental e assistência farmacêutica”. Refere-se à uma estrutura de Gestão que abrange a União, Estados e Distrito Federal, e Municípios, cujo atendimento é integral, universal e gratuito para toda população do país.

A garantia desta operação requer investimentos, de modo que a promoção da saúde no país seja fomentada e aprimorada; apesar da consciência sobre a realidade do sistema, ainda há esforços a serem feitos para a evolução do mesmo, dentre os quais, este trabalho objetiva contribuir, ainda que minimamente, para uma mudança positiva no cenário por meio da análise, mineração e entendimento dos dados para a descoberta de conhecimento relevante.

Acerca dos dados, Beal (2004) explica que podem ser compreendidos, basicamente, como registros primários e sem atribuição de significado. Conforme expõe Rezende (2003), o volume de dados armazenados no atual contexto mundial possui grandes proporções que ainda estão em crescimento.

Já as informações referem-se, segundo Beal (2004), ao fornecimento de propósito aos dados, às relações entre eles; e por fim, o conhecimento diz respeito à combinação das informações, portanto, há um vínculo explícito entre

os conceitos. Sobre ele Rezende (2003) revela que extrair conhecimento a partir dos dados exige o desenvolvimento de novas ferramentas e técnicas.

Acerca desta extração, um processo que auxilia a ação é o Processo de Descoberta de Conhecimento em Bases de Dados, mais conhecido como Knowledge Discovery in Databases (KDD); esta metodologia consiste na aplicação de determinadas etapas para a identificação de conhecimento a partir dos dados em uma base. O número de etapas varia de acordo com a literatura, entretanto, é possível identificar três momentos importantes: pré-processamento, mineração de dados e pós-processamento, dentre os quais a mineração é a fase primordial; sua importância também é abordada por Han, Kamber e Pei (2011) no sentido de ser a evolução natural da tecnologia da informação. A análise revela-se fundamental ao processo de compreensão e claro entendimento sobre os dados.

Portanto, diante deste contexto, da relevância da análise e mineração de dados, da descoberta de conhecimento e do progresso da própria saúde, o presente trabalho tem como objetivo relacionar a todos os conceitos por meio da aplicação de métodos e técnicas de mineração para identificar padrões em uma base de dados sobre o perfil de atendimento dos profissionais médicos, ou seja, registros acerca do atendimento de pacientes por médicos na rede municipal de saúde.

1.1 DEFINIÇÃO DE PROBLEMA

Ao longo dos anos, torna-se visível o avanço da tecnologia nas mais diversas áreas. Acerca deste progresso, Coeli (2010, p.335) complementa com a afirmação de que “os expressivos desenvolvimento e difusão das tecnologias de informação observados nas últimas décadas impactaram de forma marcante em todos os ramos da atividade humana, incluindo a pesquisa em saúde”.

Os procedimentos e processos de trabalho no âmbito médico sofreram diversas mudanças ao decorrer dos avanços e irão passar por outras alterações, afinal, a evolução ainda ocorre. Um aspecto notável e significativo à saúde, além dos desenvolvimentos específicos de medicamentos e tratamentos, refere-se à

adoção de sistemas informatizados para o armazenamento de dados e informações de caráter significativo a pesquisas, aos profissionais da saúde, pacientes e demais interessados.

Um exemplo desta realidade é o sistema “E-saúde”, uma ferramenta desenvolvida pelo Departamento de Informática do SUS (DATASUS) com o objetivo ser um canal de comunicação entre o Ministério da Saúde e o cidadão, por meio da divulgação online de informações de saúde de caráter pessoal e exclusivo para cada cidadão brasileiro; conforme divulgação do Ministério, a aplicação permite o acesso aos dados do cartão nacional de saúde, lista de medicamentos obtidos nas unidades de saúde, acompanhamento do cartão de vacinação, lista de exames, entre outras informações; segundo Ricardo Barros, ministro da saúde durante o lançamento do e-saúde

a informatização é uma das prioridades dessa gestão. Precisamos usar a tecnologia para integrar os dados da saúde, promover a correta aplicação dos recursos públicos, aprimorar o planejamento das ações e, principalmente, ampliar o acesso e a qualidade da assistência prestada à população, tornando o atendimento mais eficiente¹.
(Departamento de Informática do SUS - DATASUS, 2017)

Como revelado, trata-se de um sistema que visa facilitar a comunicação e melhorar processos. Porém, é importante ressaltar que além destas questões explícitas, o E-saúde, de acordo com a Prefeitura de Curitiba, também torna possível o registro dos atendimentos prestados pela Secretaria Municipal de Saúde dentro de sua concentração, composta por: Unidades Básicas de Saúde, Unidades de Pronto Atendimento, Centros de Especialidades Médicas e Odontológicas, entre outras. Este armazenamento converte-se em bases de dados, cuja disponibilidade é feita no Portal da Prefeitura de Curitiba, opção ‘Portal Dados Abertos e Curitiba’, em contato com o Núcleo de Informação e Tecnologia da Secretaria de Saúde de Curitiba, foi explicado que o fornecimento deste sistema é feito pelo Instituto das Cidades Inteligentes (ICI).

De modo a relacionar tecnologia e saúde, o presente trabalho dispõe-se à aplicar a análise de dados, com foco na mineração, em bases do supracitado

¹ Fonte: <http://datasus.saude.gov.br/noticias/atualizacoes/1094-ministerio-lanca-aplicativo-para-ampliar-o-acesso-da-populacao-as-informacoes-de-saude>

sistema, e como pergunta da pesquisa, responsável pela orientação dos encaminhamentos metodológicos, há uma questão referente à natureza dos dados da base(s) selecionada(s); esta pode ser definida da seguinte forma: **como identificar padrões no perfil de atendimento dos profissionais médicos da rede de saúde em Curitiba - PR por meio da análise de dados e aplicação de métodos de mineração de dados?**

1.2 OBJETIVOS

Esta seção visa apresentar os objetivos traçados para o desenvolvimento do trabalho; os objetivos a serem alcançados dividem-se em: geral e específicos.

O objetivo geral consiste em analisar os dados e aplicar métodos de Mineração de Dados a fim de identificar padrões que possibilitem a descoberta de conhecimento em base de dados do sistema “E-saúde” sobre o “Perfil de atendimento nas Unidades Municipais de Saúde de Curitiba”.

Os objetivos específicos deste trabalho são:

- analisar os dados por meio da estatística descritiva;
- criar visualizações de dados sobre os conjuntos de dados do “Perfil de atendimento nas Unidades Municipais”;
- selecionar algoritmo(s) para aplicar nas bases de dados supracitadas para identificação de padrões;
- comparar resultados da mineração de dados no espectro temporal de nove (9) meses;
- classificar os atendimentos com base na Classificação Internacional de Doenças (CID).

1.3 JUSTIFICATIVA

A saúde é um tema de grande relevância, afinal, conforme Paim et al. (2015, p.11) explicam “é uma questão que tem ocupado o centro das atenções de muitas pessoas, governos, empresas e comunidades. Além de corresponder a um estado da vida, isto é, a um modo de levar a vida, a saúde compõe um setor da economia no qual se produzem bens e serviços”, ou seja, trata-se de uma área essencial à sociedade, portanto, discussões a respeito são importantes para que novos conhecimentos sejam descobertos e auxiliem no desenvolvimento desta esfera.

A partir do anseio de contribuir para a área, este trabalho visa identificar padrões em base de dados do Sistema E-saúde; base sobre o perfil de atendimento dos profissionais médicos da rede municipal de saúde. O sistema viabiliza o registro dos atendimentos prestados pela Secretaria Municipal de Saúde de Curitiba em sua rede de atenção².

A fim de identificar trabalhos semelhantes a este, no dia 05 de abril de 2019, foram realizadas pesquisas na Web of Science (ISI Web of Knowledge) e na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Foi utilizado o operador booleano AND (E) e termos relacionados ao tema. Os resultados da pesquisa na Web of Science estão presentes na FIGURA 1 e na FIGURA 2:

FIGURA 1 - PESQUISA EM PORTUGUÊS NA WEB OF SCIENCE (ISI Web of Knowledge)

Histórico de pesquisa:			Editar resultados	Combinar resultados	Excluir resultados
Resultados	Resultados			AND OR	Selecionar tudo
		<input type="button" value="Salvar histórico/Criar alerta"/> <input type="button" value="Abrir histórico salvo"/>		<input type="button" value="Combinar"/>	<input type="button" value="Excluir"/>
# 4	0	(TS=("Análise de Dados" AND sus AND "sistema e-saúde")) AND IDIOMA: (Portuguese) AND TIPOS DE DOCUMENTO: (Article) <small>Índices=SCI-EXPANDED, SSCI, ABHCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado=2000-2019</small>	Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 3	0	(TS=("Análise de Dados" AND sus AND curitiba)) AND IDIOMA: (Portuguese) AND TIPOS DE DOCUMENTO: (Article) <small>Índices=SCI-EXPANDED, SSCI, ABHCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado=2000-2019</small>	Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 2	0	(TS=("Análise de Dados" AND sus AND "atendimento médico")) AND IDIOMA: (Portuguese) AND TIPOS DE DOCUMENTO: (Article) <small>Índices=SCI-EXPANDED, SSCI, ABHCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado=2000-2019</small>	Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 1	0	(TS=("Análise de Dados" AND sus)) AND IDIOMA: (Portuguese) AND TIPOS DE DOCUMENTO: (Article) <small>Índices=SCI-EXPANDED, SSCI, ABHCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado=2000-2019</small>	Editar	<input type="checkbox"/>	<input type="checkbox"/>
				<input type="button" value="Combinar"/>	<input type="button" value="Excluir"/>

FONTE: Web of Science - ISI Web of knowledge (2019).

² Informações do Portal da Prefeitura de Curitiba . Fonte: <http://www.curitiba.pr.gov.br/dadosabertos/consulta/?grupo=1>.

A FIGURA 2 traz os resultados encontrados na língua inglesa:

FIGURA 2 - PESQUISA EM INGLÊS NA WEB OF SCIENCE (ISI Web of Knowledge)

Histórico de pesquisa:

Resultados	Resultados	Salvar histórico/Criar alerta	Abrir histórico salvo	Editar resultados	Combinar resultados AND OR Combinar	Excluir resultados Selecionar tudo Excluir
# 4	0	(TS=("data analysis" AND sus AND "e-saúde system")) AND IDIOMA: (English) AND TIPOS DE DOCUMENTO: (Article) Índices-SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado-2000-2019		Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 3	0	(TS=("data analysis" AND sus AND curitiba)) AND IDIOMA: (English) AND TIPOS DE DOCUMENTO: (Article) Índices-SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado-2000-2019		Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 2	0	(TS=("data analysis" AND sus AND "medical care")) AND IDIOMA: (English) AND TIPOS DE DOCUMENTO: (Article) Índices-SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado-2000-2019		Editar	<input type="checkbox"/>	<input type="checkbox"/>
# 1	32	(TS=("data analysis" AND sus)) AND IDIOMA: (English) AND TIPOS DE DOCUMENTO: (Article) Índices-SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Tempo estipulado-2000-2019		Editar	<input type="checkbox"/>	<input type="checkbox"/>

FONTE: Web of Science - ISI Web of knowledge (2019).

Destaca-se que TS representa a busca pelos termos no título, resumo, palavras-chave e *keywords plus*. A partir dos resultados identificados, é possível perceber apenas 32 correspondências em inglês na Web of Science, em português não há resultados. Analisando, brevemente, os artigos encontrados, constatou-se que não se referem ao tema deste trabalho. Na base BDTD, os resultados estão presentes no QUADRO 1:

QUADRO 1 - PESQUISAS EM PORTUGUÊS E INGLÊS NA BIBLIOTECA BRASILEIRA DE TESES E DISSERTAÇÕES (BDTD)

CRITÉRIO	TERMOS	TIPO DE DOCUMENTO	ANO DE DEFESA	RESULTADOS
Português	Todos os campos "Análise de dados" AND sus	dissertação ou tese	2000-2019	536
	Todos os campos "Análise de dados" AND sus AND "atendimento médico"	dissertação ou tese	2000-2019	3
	Todos os campos "Análise de dados" AND sus AND Curitiba	dissertação ou tese	2000-2019	10
	Todos os campos "Análise de dados" AND sus AND "sistema e-saúde"	dissertação ou tese	2000-2019	66
Inglês	Todos os campos "data analysis" AND sus	dissertação ou tese	2000-2019	0

	"data analysis" AND sus AND "medical care"	dissertação ou tese	2000-2019	0
	Todos os campos "data analysis" AND sus AND curitiba	dissertação ou tese	2000-2019	0
	Todos os campos "data analysis" AND sus AND "e-saúde system"	dissertação ou tese	2000-2019	0

FONTE: Adaptado da Biblioteca Brasileira de Bases e Dissertações (2019).

Na primeira pesquisa foram identificados muitos resultados devido à natureza geral dos termos, quando refinado e especificado, a busca retorna menos resultados. No caso da última pesquisa com o termo "Sistema E-Saúde", percebeu-se que as dissertações e/ou teses encontradas referiam-se à "Sistema de Saúde", de modo que se torna possível perceber que a exata correspondência ao tema deste trabalho não foi encontrada.

Com base nas informações do quadro, fica claro que o tema de perfil de atendimento médico, do Sistema Único de Saúde (SUS), atrelado à análise de dados é pouco explorado, o que complementa a motivação para o desenvolvimento deste trabalho.

Além das questões relacionadas à temática, de modo geral, durante a graduação em Gestão da Informação, somos apresentados à diversas disciplinas, destas, destacam-se as que possibilitaram o desenvolvimento deste trabalho como a Mineração de Dados, Banco de Dados, I e II, Metodologia e Técnicas de Pesquisa, Necessidades e Usos de Informação. Informação e Estratégia, Métodos Quantitativos de Análise, Gestão de Projetos e Tópicos Especiais e Aprendizagem de Máquina e Reconhecimento de Padrões, portanto, este trabalho reflete na aplicação de conteúdos estudados em sala em uma área essencial, que carece de soluções e/ou propostas que fomentem suas práticas.

No âmbito pessoal a saúde sempre foi tema presente, uma vez que situações particulares permitiram um vínculo estreito com as práticas desta

esfera, bem como admiração por seus resultados, desta forma, espera-se retornar, ainda que minimamente, todos os benefícios proporcionados por ela.

1.4 DELIMITAÇÃO DA PESQUISA

Este trabalho de pesquisa tem como objeto principal as bases de dados cedidas pela iniciativa da Prefeitura de Curitiba denominada Portal de Dados Abertos de Curitiba; trata-se de uma ação relacionada à transparência e cidadania participativa que disponibiliza documentos, informações e dados governamentais de domínio público para livre uso pela sociedade, além de garantir o acesso aos dados primários com possibilidade de reutilização para a produção de novas informações e aplicações digitais³. Ressalta-se que o foco deste trabalho é o Município de Curitiba e a descoberta de padrões gerais, ou seja, não objetiva a identificação de pacientes.

As bases trazem dados com foco no perfil de atendimento dos profissionais médicos da rede municipal de saúde; estão presentes dados gerais de caracterização do atendimento, tais como: data do atendimento; código do tipo de unidade; tipo de unidade; código da unidade e descrição da unidade.

Também apresenta dados relacionados aos procedimentos e diagnósticos realizados durante o atendimento, são eles: código do procedimento; descrição do procedimento; código do CID; descrição do CID; solicitação de exames; quantidade prescrita Farmácia Curitibana; quantidade dispensada Farmácia Curitibana; quantidade de medicamento não padronizado; encaminhamento para atendimento especialista; desencadeou internamento; data do internamento; estabelecimento solicitante; estabelecimento destino e CID do internamento.

Ainda há dados sobre os profissionais como: código do CBO; descrição CBO e área de atuação; e dados sobre os pacientes: data de nascimento; sexo; tratamento no domicílio; abastecimento; energia elétrica; tipo de habitação;

³ <http://www.curitiba.pr.gov.br/dadosabertos/>

destino lixo; fezes/urina; cômodos; em caso de doença; grupo comunitário; meio de comunicação; meio de transporte; município e bairro do paciente.

1.5 ESTRUTURA DO DOCUMENTO

A organização deste trabalho de conclusão de curso é composta por cinco seções e visa destacar a trajetória da pesquisa para o alcance de resultados que satisfaçam os propósitos traçados.

A primeira seção compreende a introdução e apresenta as definições iniciais relacionadas ao problema de pesquisa; objetivos, geral e específicos; justificativa para a seleção do tema e delimitação da pesquisa.

A segunda seção refere-se à elucidação de conceitos essenciais à compreensão do tema, ou seja, trata-se da revisão de literatura definida para auxiliar no entendimento deste trabalho; as definições presentes tratam da Gestão da Informação; recuperação da informação; processo de Descoberta de Conhecimento em Bases de Dados; Mineração de Dados e Sistema Único de Saúde (SUS).

A terceira seção indica a metodologia e representa as informações acerca de ambiente de pesquisa, caracterização da pesquisa, dicionário de dados e comitê de ética.

A quarta seção refere-se aos procedimentos metodológicos, no caso, à aplicação das etapas de KDD aos conjuntos de dados, compreendendo desde o pré-processamento com a estatística descrita, tratamento das bases de dados e visualização de dados; mineração de dados; e pós-processamento.

A seção final abrange as considerações finais, com discussão dos resultados identificados e desenvolvimento de trabalhos futuros.

2 REVISÃO DE LITERATURA

Esta seção tem como objetivo apresentar uma discussão acerca de conceitos relevantes à compreensão do presente trabalho de conclusão de curso. Constituem a revisão de literatura abordagens sobre a Gestão da Informação, recuperação da informação, processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases - KDD*), Análise de Dados, Mineração de Dados e Sistema Único de Saúde (SUS).

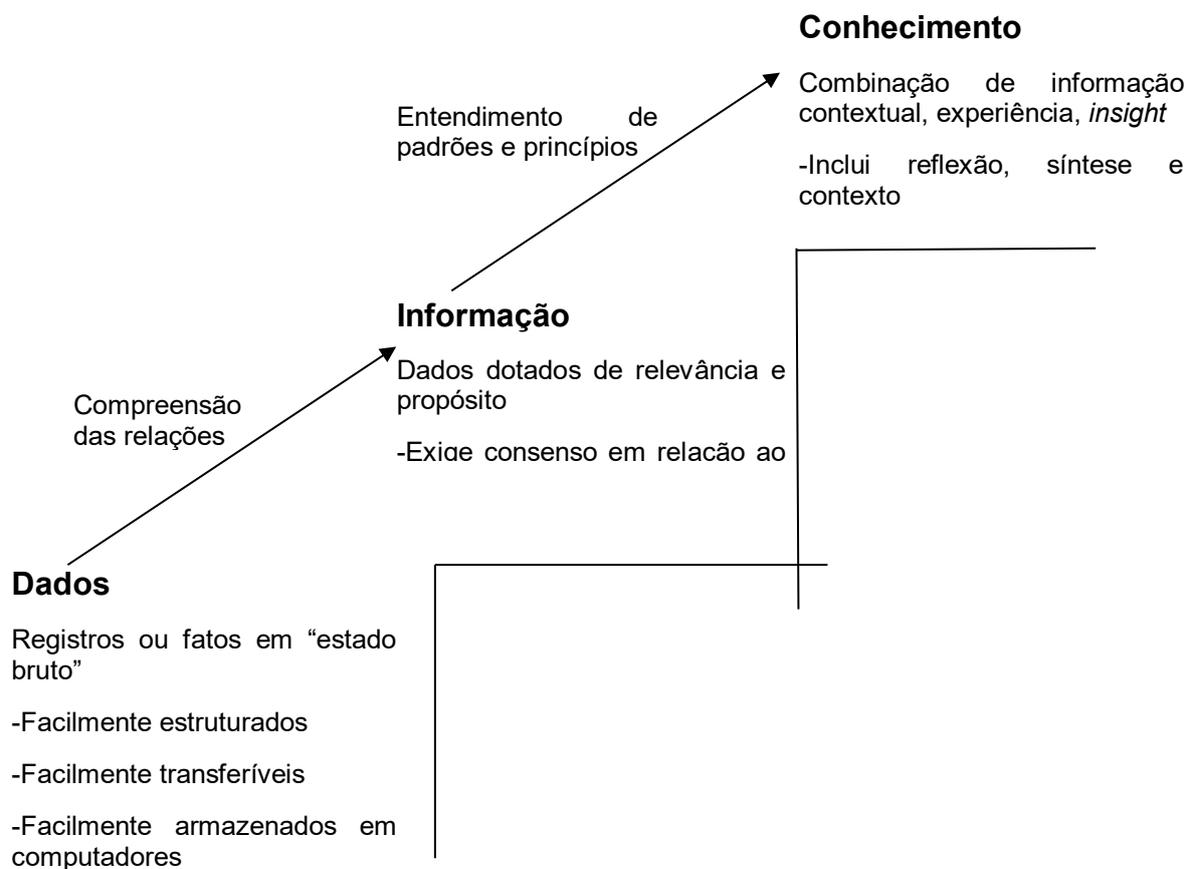
2.1 GESTÃO DA INFORMAÇÃO

Atribuir significado e contexto, pertinentes, aos dados revela a informação e a partir da definição de Beal é possível perceber a relação entre os conceitos, uma vez que

dados podem ser entendidos como registros ou fatos em sua forma primária, não necessariamente físicos - uma imagem guardada na memória também é um dado. Quando esses registros ou fatos são combinados de forma significativa, eles se transformam numa *informação*. (BEAL, 2004, p.12, grifo do autor)

Esta apresenta grande relevância para as pessoas e empresas, uma vez que possibilita, por meio da interpretação e compreensão, a descoberta de conhecimento, cuja importância para a sociedade, segundo Moreira (2015, p. 09), “[...] não é recente pois, mesmo de maneira informal, esse ativo é usado, como diferencial, desde a pré-história quando os homens inventaram uma maneira de registrar nas paredes das cavernas símbolos do que aprendiam”. A composição dos “níveis hierárquicos” da informação está presente na FIGURA 3.

FIGURA 3 - OS "NÍVEIS HIERÁRQUICOS" DA INFORMAÇÃO



FONTE: Adaptado de Beal (2004).

A partir do exposto na figura acima, fica explícita a relação de ordem e sequência entre os conceitos; no primeiro "degrau", estão os dados, basicamente "registros em estado bruto", após compreender as relações e combiná-los de modo significativo, com relevância e propósito, revela a informação; esta, quando há entendimento de padrões e princípios corresponde ao conhecimento.

Além da descoberta de conhecimento, outro processo significativo para as organizações é a tomada de decisão, esta ocorre com o auxílio das informações de modo a atender os pressupostos de unidade, confidencialidade e agilidade, a fim de garantir que estejam de acordo com os objetivos organizacionais (MOREIRA, 2015); de modo geral, é perceptível o valor das

informações para as empresas, pois constituem, segundo McGee e Prusak (1994 apud Moreira, 2015) “um importante insumo estratégico, capaz de influenciar o negócio da empresa, tornando-se cada vez mais a base para a competição”.

A supracitada afirmação, está em concordância com o explicado por Barbosa, afinal, ele expõe que

à medida que os ambientes profissionais e de negócios se tornam mais complexos e mutantes, a informação se transforma, indiscutivelmente, em uma arma capaz de garantir a devida antecipação e análise de tendências, bem como a capacidade de adaptação, de aprendizagem e de inovação. (BARBOSA, 2008, p.02)

Em suma, a informação fornece vantagem competitiva e subsídio para as empresas desenvolverem seus negócios de forma a acompanhar e prever situações de provável ocorrência no mercado. O QUADRO 2 apresenta uma síntese de parte do conteúdo presente no capítulo 02 do livro “Gestão Estratégica da Informação” de Adriana Beal (2004), denominado “O valor da Informação para as Organizações”:

QUADRO 2 - O VALOR DA INFORMAÇÃO PARA AS ORGANIZAÇÕES

FATOR DE APOIO À DECISÃO	“A informação possibilita a redução da incerteza na tomada de decisão, permitindo que escolhas sejam feitas com menor risco e no momento adequado.”
FATOR DE PRODUÇÃO	“A informação é elemento importante para se criar e introduzir no mercado produtos (bens ou serviços) de maior valor adicionado.”
FATOR DE SINERGIA	“O desempenho de uma organização está condicionado à qualidade das ligações e relações entre as unidades organizacionais, e estas por sua vez dependem da qualidade do fluxo informacional existente para proporcionar o intercâmbio de ideias e informações.”
FATOR DETERMINANTE DE COMPORTAMENTO	“A informação exerce influência sobre o comportamento dos indivíduos e dos grupos, dentro e fora das organizações [...]”

FONTE: Adaptado de Beal (2004).

O quadro 2 revela o valor aplicado à diferentes contextos, ou seja, a observação da importância da informação em diversas aplicações, o que

corroborar com as afirmações anteriormente apresentadas sobre a relevância da informação para as empresas.

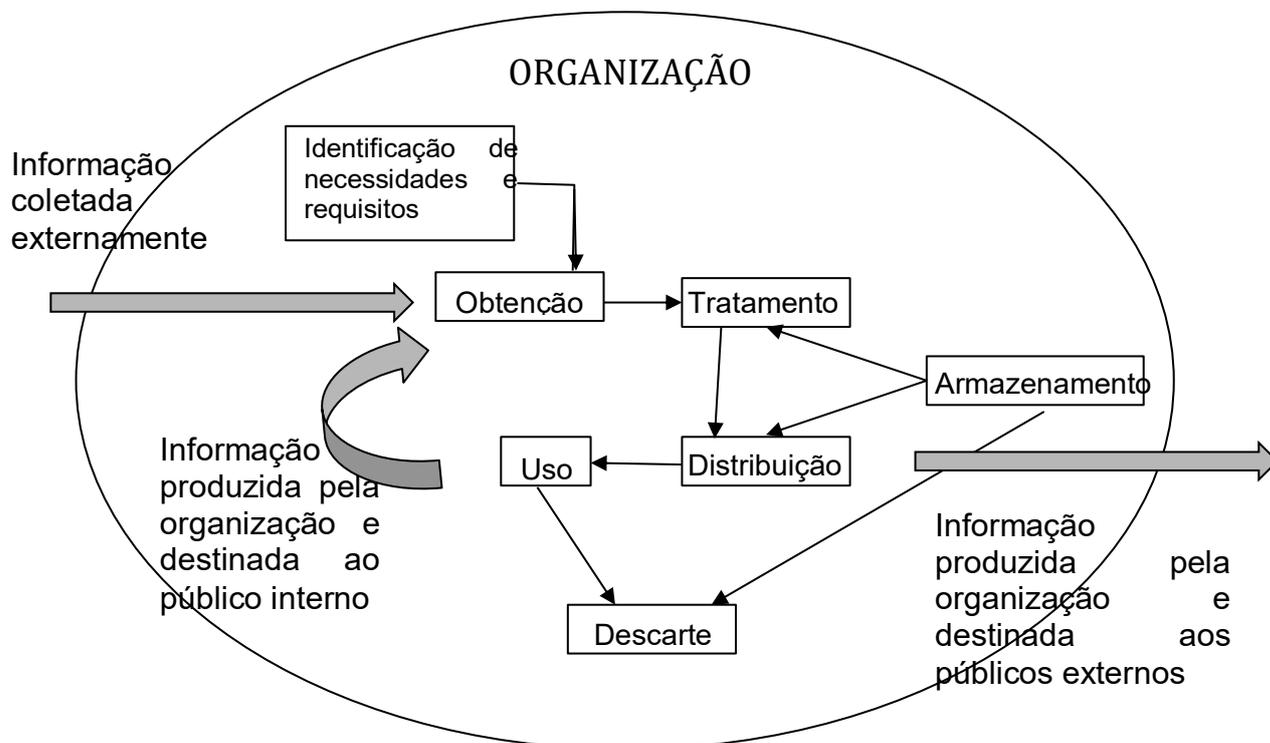
Segundo Vital, Floriani e Varvakis,

a gestão da informação requer o estabelecimento de processos, etapas ou fluxos sistematizados e estruturados, associado às pessoas responsáveis por sua condução, para que se obtenham os resultados almejados (VITAL; FLORIANI; VARVAKIS, 2010, p.85-86)

O que significa que fluxos informacionais são relevantes para se fazer uso do ativo essencial às organizações, a informação; ainda, de acordo com Vital, Floriani e Varvakis (2010, p.86), “os fluxos de informação permitem o estabelecimento das etapas de obtenção, tratamento, armazenamento, distribuição, disseminação e uso da informação no contexto organizacional”.

Representando importante função do fluxo informacional, há os subsídios aos gestores para a execução do processo de tomada de decisão (VITAL; FLORIANI; VARVAKIS, 2010). Acerca do fluxo de informação, há o modelo proposto por Beal (2004), na FIGURA 4:

FIGURA 4 - MODELO PROPOSTO PARA REPRESENTAR O FLUXO DA INFORMAÇÃO NAS ORGANIZAÇÕES



FONTE: Adaptado de Beal (2004).

O modelo apresentado na FIGURA 4 tem como etapa inicial a identificação das necessidades e requisitos, este, segundo Beal (2004, p.30) revela que “identificar as necessidades de informação dos grupos e indivíduos que integram a organização e de seus públicos externos é um passo fundamental para que possam ser desenvolvidos produtos informacionais orientados especificamente para cada grupo e necessidade”; a próxima etapa visa obter as informações que atendam às necessidades anteriormente levantadas.

Os processos tais como: organizar, formatar, estruturar, classificar, analisar, sintetizar e apresentar a informação de modo a torná-la acessível e de fácil localização constituem a etapa de tratamento (BEAL, 2004). A etapa seguinte, de distribuição refere-se à entrega da informação para quem dela necessita. O uso da informação é uma etapa essencial e, segundo Beal (2004, p. 31) “[...] possibilita a combinação de informações e o surgimento de novos conhecimentos, que podem voltar a alimentar o ciclo da informação corporativo, num processo contínuo de aprendizado e crescimento”, em suma, é uma etapa relevante para a descoberta de conhecimento, atividade de extrema importância para a Gestão da Informação.

O armazenamento refere-se à garantia da conservação de dados e informações, a fim de permitir seu uso e reuso na organização (BEAL, 2004, p.31). Por fim, há o descarte da informação. Este processo ocorre quando a informação se torna obsoleta ou inútil para a empresa, desta forma deverá ser descartada, em um processo que obedeça à normas legais, políticas operacionais e exigências internas (BEAL, 2004, p. 32).

De modo geral, o fluxo da informação possibilita a visualização da trajetória da informação em uma organização, desde o momento de levantamento das necessidades dos usuários até o armazenamento ou descarte, portanto, caracteriza-se como elemento indispensável à compreensão e entendimento do valor das informações relacionadas ao seu uso para suporte em tomada de decisão, dos processos de trabalho de um gestor da informação e de sua relevância para as empresas.

2.2 RECUPERAÇÃO DA INFORMAÇÃO

A palavra “recuperação” remete ao conceito de “localização” de algo, no caso, de informação. Mooers (1951 apud Saracevic, 1996, p.44) responsável por cunhar o termo Recuperação da Informação (RI), revela que ele “engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação”, ou seja, é um termo relacionado à definição das informações para posterior busca.

Segundo Ferneda (2003, p.14) “o processo de recuperação de informação consiste em identificar, no conjunto de documentos (*corpus*) de um sistema, quais atendem à necessidade de informação do usuário”, em um paralelo com o fluxo da informação, a etapa de obtenção refere-se à essa busca informacional que atenda às necessidades levantadas na etapa anterior, de “identificação das necessidades e requisitos”.

Acerca da recuperação, Saracevic (1996, p.44) expõe que “considerando o problema da informação conforme definido, isto é, a explosão informacional, a recuperação da informação tornou-se uma solução bem sucedida encontrada pela CI [...]”, basicamente, com o grande ápice de informações a nível mundial, as técnicas de recuperação auxiliam no processo de busca e localização das informações úteis e precisas às necessidades dos usuários, o que corrobora com o paralelo supracitado.

Cardoso (2004, p. 01) explica que a “recuperação de informação é uma subárea da ciência da computação que estuda o armazenamento e recuperação automática de documentos que são objetos de dados, geralmente textos”. O processo de recuperação da informação utiliza modelos para realizar a busca; estes modelos, denominados “modelos clássicos”, consistem em: booleano, vetorial e probabilístico. Abaixo, estão as definições desses modelos:

- **MODELO BOOLEANO:** “Dada uma consulta Q e um conjunto de documentos considerados relevantes para a Q, o índice atribuído aos documentos deve indicar qual documento é mais relevante que outro, estabelecendo uma ordem de relevância. Esses índices são calculados com base na comparação entre a consulta e os documentos. No modelo booleano os documentos recuperados são aqueles que contêm os termos que satisfazem a expressão lógica da consulta. Uma

consulta é considerada como uma expressão booleana convencional formada com os conectivos lógicos AND, OR e NOT [...]”;

- **MODELO VETORIAL:** “O modelo de espaço vetorial, ou simplesmente modelo vetorial, representa documentos e consultas como vetores de termos. Termos são ocorrências únicas nos documentos. Os documentos devolvidos como resultado para uma consulta são representados similarmente, ou seja, o vetor resultado para uma consulta é montado através de um cálculo de similaridade [...]”;

- **MODELO PROBABILÍSTICO:** “O modelo probabilístico descreve documentos considerando pesos binários que representam a presença ou ausência de termos. O vetor resultado gerado pelo modelo tem como base o cálculo da probabilidade de que um documento seja relevante para uma consulta [...] O modelo probabilístico é baseado no princípio probabilístico de ordenação (*Probability Ranking Principle*), que estabelece que este modelo pode ser usado de forma ótima. Este princípio é baseado na hipótese de que a relevância de um documento para uma determinada consulta é independente de outros documentos [...]”. (CARDOSO, 2004, p.03)

Em suma, esses modelos auxiliam no processo de RI no sentido de mostrarem estratégias de busca de documentos relevantes para uma consulta, uma *query* (CARDOSO, 2004, p.02); cada modelo possui sua particularidade de forma a auxiliar o usuário na localização da informação que necessita. Barros Correia (2018, p. 58) listou algumas vantagens e desvantagens dos modelos clássicos, apresentadas abaixo, no QUADRO 3:

QUADRO 3 - VANTAGENS E DESVANTAGENS DOS MODELOS CLÁSSICOS DE RECUPERAÇÃO DA INFORMAÇÃO

MODELO	VANTAGENS	DESVANTAGENS
BOOLEANO	<ul style="list-style-type: none"> • Expressividade completa se o utilizador souber exatamente o que quer; • É facilmente programável e exato; • Fácil de entender e implementar; • Modelo simples baseado em teoria bem definida. 	<ul style="list-style-type: none"> • Pessoas lidam com conhecimento parcial; • Saída pode ser nula ou haver <i>overload</i> (problema que dá um falso senso de segurança); • Não permite ordenação de documentos recuperados; saída desordenada.
VETORIAL	<ul style="list-style-type: none"> • Atribuir pesos aos termos melhora o desempenho; • É uma estratégia de encontro parcial (função de similaridade), que é melhor que a exatidão do modelo booleano; • Co-seno ordena, os documentos são ordenados de acordo com seu grau de similaridade com a consulta. 	<ul style="list-style-type: none"> • Ausência de ortogonalidade entre os termos, isto poderia encontrar relações entre termos que aparentemente não tem nada em comum; • É um modelo generalizado; • Um documento relevante pode não conter termos de consulta.

PROBABILÍSTICO	<ul style="list-style-type: none"> • Medida de similaridade: os documentos são retornados em ordem decrescente do seu grau de semelhança. 	<ul style="list-style-type: none"> • Necessidade de separar os documentos relevantes a priori
-----------------------	--	--

FONTE: Adaptado de Barros Correia (2018).

O quadro revela pontos a serem considerados no momento de selecionar um modelo para a Recuperação da Informação, afinal, um modelo inadequado às necessidades dos usuários resulta em informações inúteis e/ou imprecisas que podem interferir, inclusive, de forma negativa em uma pesquisa. Nas bases de dados, Barros Correia (2018, p.69) explica que “[...] o uso dos recursos como ferramentas de busca e operadores booleanos, são instrumentos indispensáveis no momento da pesquisa, tanto na pesquisa livre, como na estruturada, seja ela realizada em bases de dados ou na Internet [...]”. A aplicação dos modelos para recuperar a informação é importante, entretanto, é necessário haver um planejamento das estratégias que serão adotadas na supracitada aplicação para que os resultados sejam efetivamente utilizados e atinjam ou superem as expectativas dos usuários.

2.3 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (*KNOWLEDGE DISCOVERY IN DATABASES - KDD*)

As atividades relacionadas ao processamento, tratamento, organização, aplicação de métodos e ferramentas, são algumas das etapas de um processo denominado *Knowledge Discovery in Databases* (KDD) ou Descoberta de Conhecimento em Bases de Dados. Os autores Maimon e Rokach (2010, p.01, tradução nossa) explicam, que o KDD “[...] é uma análise automática, exploratória e modelagem de grandes repositórios de dados. O KDD é o processo organizado de identificar padrões válidos, novos, úteis e compreensíveis a partir de conjuntos de dados grandes e complexos”, ou seja, trata-se de um processo com atividades e etapas específicas que visam, por meio de análise, descobrir, identificar conhecimento a partir de dados.

Ainda de acordo com os autores, o processo de KDD é composto por nove (9) etapas; o início consiste na determinação dos objetivos do processo, enquanto a finalização ocorre com a implementação do conhecimento

descoberto (MAIMON; ROKACH, 2010, p. 02, tradução nossa). A grosso modo, a primeira etapa consiste em compreender o domínio da aplicação e na preparação para entender as diversas decisões, relacionadas à transformação, algoritmos, representação e outros aspectos, além de que os responsáveis pelo KDD necessitam assimilar e definir os objetivos do usuário final e do ambiente no qual a descoberta de conhecimento ocorrerá (MAIMON; ROKACH, 2010, p. 02, tradução nossa).

Na segunda etapa,

[...] Tendo definido os objetivos, os dados que serão usados para a descoberta do conhecimento devem ser determinados. Isso inclui descobrir quais dados estão disponíveis, obter dados adicionais necessários e, então, integrar todos os dados para a descoberta do conhecimento em um conjunto de dados, incluindo os atributos que serão considerados para o processo. Esse processo é muito importante porque a Mineração de Dados aprende e descobre os dados disponíveis. Esta é a base de evidências para a construção dos modelos. Se alguns atributos importantes estiverem faltando, todo o estudo poderá falhar. A partir do sucesso do processo, é bom considerar o maior número possível de atributos nesse estágio. Por outro lado, coletar, organizar e operar repositórios de dados complexos é caro, e há uma compensação com a oportunidade de entender melhor os fenômenos. Essa troca representa um ponto em que o aspecto interativo e iterativo do KDD está ocorrendo. Começa com o melhor conjunto de dados disponível e depois expande e observa o efeito em termos de descoberta e modelagem de conhecimento. (MAIMON; ROKACH, 2010, p.02-03, tradução nossa).

Basicamente, a terceira etapa corresponde ao pré-processamento e limpeza, nela ocorre o aprimoramento da confiabilidade dos dados, bem como limpeza de dados, manipulação de valores omissos, remoção de ruído ou outliers (MAIMON; ROKACH, 2010, p. 03, tradução nossa), em suma, tratam-se dos primeiros procedimentos para o ajuste da base de dados. Chakrabarti et al. (2008, p.78, tradução nossa), reforçam a relevância da etapa três “o pré-processamento de dados é um passo importante no processo de descoberta de conhecimento, porque as decisões de qualidade devem ser baseadas em dados de qualidade”. A quarta etapa é a de

Transformação de dados. Nesta etapa, a geração de dados melhores para a mineração de dados é preparada e desenvolvida. Os métodos aqui incluem redução de dimensão (como seleção e extração de recurso, e amostragem de registro) e transformação de atributos (como discretização de atributos numéricos e transformação funcional). Esta etapa é frequentemente crucial para o sucesso de todo o projeto KDD,

mas é geralmente muito específico do projeto. (MAIMON; ROKACH, 2010, p.04, tradução nossa).

A partir do processo acima, o foco recai sobre a Mineração de Dados. A etapa cinco representa a seleção da tarefa de Mineração de Dados apropriada, podendo ser, por exemplo, de classificação, regressão, clustering, escolha essa que irá depender dos objetivos do KDD e das etapas anteriores (MAIMON; ROKACH, 2010, p. 04, tradução nossa). Na próxima etapa há a escolha do algoritmo de Mineração de Dados que representa as táticas, no caso, a opção por um método específico para pesquisa de padrões (MAIMON; ROKACH, 2010, p. 04, tradução nossa), é importante ressaltar que os algoritmos possuem características, parâmetros específicos, portanto, trata-se de uma seleção determinante para o processo de KDD.

A escolha incorreta do algoritmo poderá resultar em insatisfação e retrabalho, por exemplo. A etapa sete representa o emprego efetivo do algoritmo da Mineração de Dados, nesta, é possível haver necessidade de diversas aplicações para que um resultado suficiente seja obtido (MAIMON; ROKACH, 2010, p. 04, tradução nossa), como mencionado, os algoritmos possuem parâmetros, estes permitem ajustes para a realização dos experimentos até o resultado final.

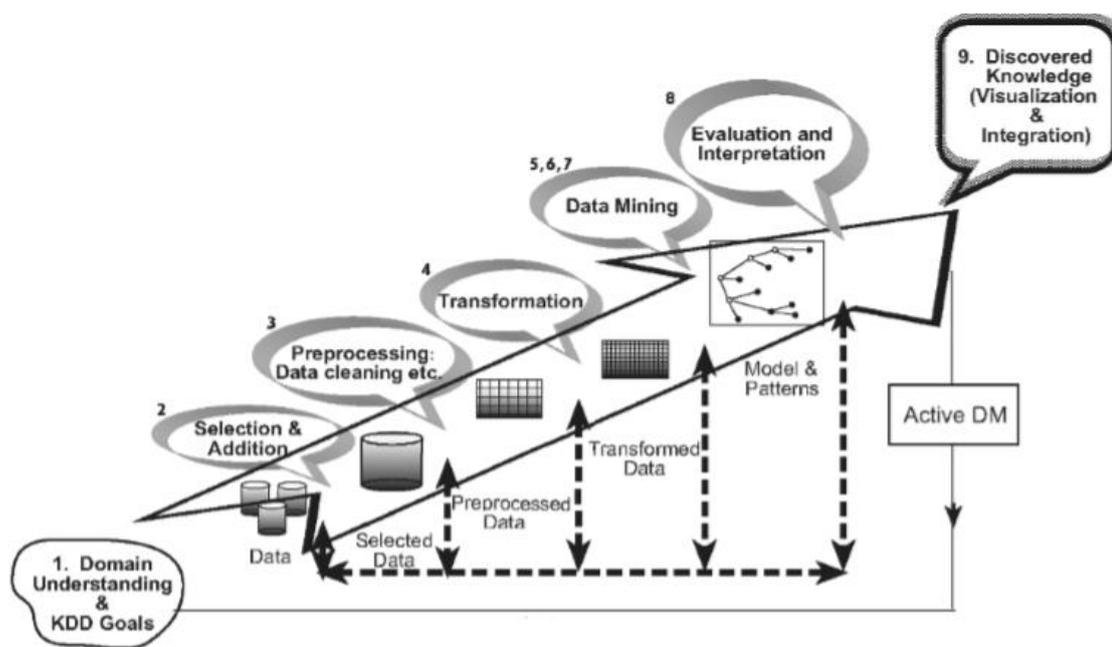
Na penúltima etapa

[...] avaliamos e interpretamos os padrões minerados (regras, confiabilidade, etc.) com relação aos objetivos definidos na primeira etapa. Aqui, consideramos as etapas de pré-processamento com relação ao seu efeito nos resultados do algoritmo de mineração de dados (por exemplo, adicionando recursos na etapa 4 e repetindo a partir dele). Esta etapa enfoca a compreensibilidade e utilidade do modelo induzido. Neste estágio, o conhecimento descoberto também é documentado para uso posterior. (MAIMON; ROKACH, 2010, p.05, tradução nossa).

O uso do conhecimento descoberto caracteriza a última etapa do processo de KDD; nela, o conhecimento poderá ser integrado em outro sistema para futura ação, ele torna-se ativo, no sentido de poder realizar alterações neste sistema e medir os efeitos (MAIMON; ROKACH, 2010, p. 05, tradução nossa), ou seja, a etapa nove representa o uso concreto do conhecimento descoberto,

este poderá ser utilizado em tomada de decisão, por exemplo. Abaixo, a FIGURA 5 apresenta, de forma visual, as etapas anteriormente explicadas.

FIGURA 5 - O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS



FONTE: Maimon e Rokach (2010).

Na FIGURA 5, o processo tem início com o primeiro estágio, que corresponde à compreensão do domínio e metas do KDD; em seguida há a seleção e adição de dados; a terceira etapa consiste no pré-processamento e limpeza de dados; a etapa quatro refere-se à transformação dos dados; os estágios cinco, seis e sete correspondem à Mineração de Dados; a etapa seguinte é de avaliação e interpretação dos padrões e modelos encontrados; por fim, há a integração e visualização do conhecimento descoberto.

É interessante ressaltar que o processo parte de um início de organização dos dados para posterior descoberta de conhecimento, ou seja, há uma preocupação com o tratamento dos dados, com a preparação da base antes de começar a Mineração de Dados, pois é por meio dela que padrões serão identificados e o conhecimento será descoberto. Portanto, é fundamental que o conjunto de dados seja completamente compreendido, o que acontecerá mediante Análise dos Dados.

2.4 ANÁLISE DE DADOS

Conforme anteriormente mencionado, os dados podem ser entendidos como unidades que não possuem significado, isoladamente e sem a atribuição de relevância, entretanto, a sua análise é de extrema importância, pois, a partir dela, ocorrerá o início do entendimento do que os dados representam e, portanto, o começo do que conhecemos como informação.

A análise de dados envolve a compreensão do conjunto de dados, ou seja,

[...] é o processo de formação de sentido além dos dados, e esta formação se dá consolidando, limitando e interpretando o que as pessoas disseram e o que o pesquisador viu e leu, isto é, o processo de formação de significado. A análise dos dados é um processo complexo que envolve retrocessos entre dados pouco concretos e conceitos abstratos, entre raciocínio indutivo e dedutivo, entre descrição e interpretação. Estes significados ou entendimentos constituem a constatação de um estudo. (TEIXEIRA, 2003, p.191-192).

Portanto, é possível entender que representa uma fase fundamental ao processo de descoberta de conhecimento em bases de dados, intrinsecamente relacionado à etapa de pré-processamento. Para a presente pesquisa, são aplicados cálculos da estatística descritiva, como medidas de tendência central e dispersão, bem como técnicas de visualização de dados com a ferramenta da Microsoft, o Power BI, de modo a conhecer os dados antes da Mineração deles.

2.5 MINERAÇÃO DE DADOS

O processo de Descoberta de Conhecimento em Bases de Dados, como anteriormente revelado, possui grande relevância como um todo, entretanto, dentre as suas fases destaca-se a abordagem prática e essencial da etapa de Mineração de Dados; esta, segundo Han, Kamber e Pei (2011, p. 08, tradução nossa) trata-se “do processo de descobrir padrões e conhecimentos interessantes a partir de grande quantidade de dados”, ou seja, é a fase responsável por identificar modelos nas bases.

Assim como a tecnologia em geral, o emprego da mineração pode ocorrer em quaisquer tipos de dados, desde que eles sejam significativos para o objetivo da aplicação (HAN; KAMBER; PEI; 2011, tradução nossa), portanto, é

importante destacar que as metas para a mineração, assim como para o KDD, devem ser bem definidas, explícitas, de modo a não permitir ambiguidades e/ou incorreto entendimento. Ainda sobre as aplicações, Bramer (2007, p.03, tradução nossa) revela uma lista dos possíveis usos da mineração de dados:

- análise de compostos orgânicos;
- abstração automática;
- detecção de fraude de cartão de crédito;
- previsão de carga elétrica;
- previsão financeira;
- diagnóstico médico;
- prevendo participação de audiências de televisão;
- design de produto;
- avaliação imobiliária;
- marketing direcionado;
- otimização da usina termoelétrica;
- análise de risco tóxico;
- previsão de tempo.

A partir da lista acima, é possível perceber os diferentes usos da mineração, bem como a presença desta área em diversos aspectos da sociedade. Bramer (2007, p.04, tradução nossa) explica que há tipos principais de aplicações (tarefas): classificação, previsão numérica, associação e agrupamento. O autor ainda revela que há dados rotulados, aqueles onde existe um atributo especificamente designado e o objetivo está em utilizar os dados fornecidos para prever o valor do atributo em instâncias que ainda não foram vistas, nesta situação, mineração de dados que faz uso dos dados rotulados é denominada como aprendizagem supervisionada, enquanto que os dados sem nenhum atributo específico são chamados de sem rótulos e a mineração neste caso é conhecida como aprendizado não supervisionado (BRAMER; 2007, p.04, tradução nossa). Nas aplicações, a classificação e a previsão numérica compõem o aprendizado supervisionado, a medida que as regras de associação e clusterização (agrupamento), fazem parte do aprendizado não supervisionado.

A classificação, de acordo com o autor, é uma das aplicações mais comuns em mineração; um exemplo elucidado por Bramer (2007, p.05) refere-se à necessidade de classificar pacientes médicos de acordo com o risco (alto, médio ou baixo) de contrair determinada doença.

O QUADRO 4, adaptado de Bramer (2007, p.05, tradução nossa), traz um exemplo do uso da classificação para um conjunto de dados que contém notas de alunos em cinco disciplinas diferentes e as classificações gerais na graduação. O objetivo é identificar uma forma de prever a classificação para outros alunos que irão receber apenas seus “perfis”.

QUADRO 4 - DADOS DE CLASSIFICAÇÃO DE GRADUAÇÃO

SoftEng	ARIN	HCI	CSA	Project	Class
A	B	A	B	B	Second
A	B	B	B	B	Second
B	A	A	B	A	Second
A	A	A	A	B	First
A	A	B	B	A	First
B	A	A	B	B	Second
.....
A	A	B	A	B	First

FONTE: Adaptado de Bramer (2007).

O autor explica que há diversas formas de resolver essa questão e cita três formas:

Vizinho mais próximo de correspondência: Esse método baseia-se na identificação (digamos) dos cinco exemplos que estão "mais próximos" em algum sentido de um não-classificado. Se os cinco vizinhos mais próximos tiverem notas Segundo, Primeiro, Segundo, Segundo e Segundo, poderíamos razoavelmente concluir que a nova instância deveria ser classificada como "Segundo"

Regras de classificação: Procuramos regras que podemos usar para prever a classificação de uma instância invisível, por exemplo:

SE SoftEng = A E Project = A ENTÃO Class = First

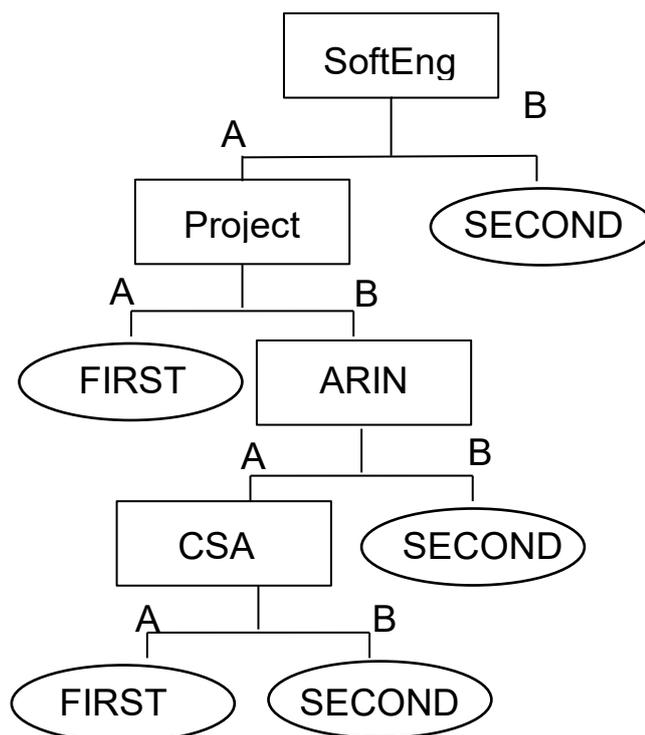
SE SoftEng = A E Projeto = B E ARIN = B ENTÃO Classe = Segundo

SE SoftEng = B ENTÃO Classe = Segundo

Árvore de Classificação: Uma maneira de gerar regras de classificação é através de uma estrutura intermediária de árvore, chamada de árvore de classificação ou árvore de decisão (BRAMER; 2007, p.05 - 06, tradução nossa, grifo do autor).

Para exemplificar a árvore de decisão, a FIGURA 6, adaptada de Bramer (2007, p.06) representa a árvore para os dados de classificação da graduação:

FIGURA 6 - ÁRVORE DE DECISÃO PARA DADOS DE CLASSIFICAÇÃO DA GRADUAÇÃO

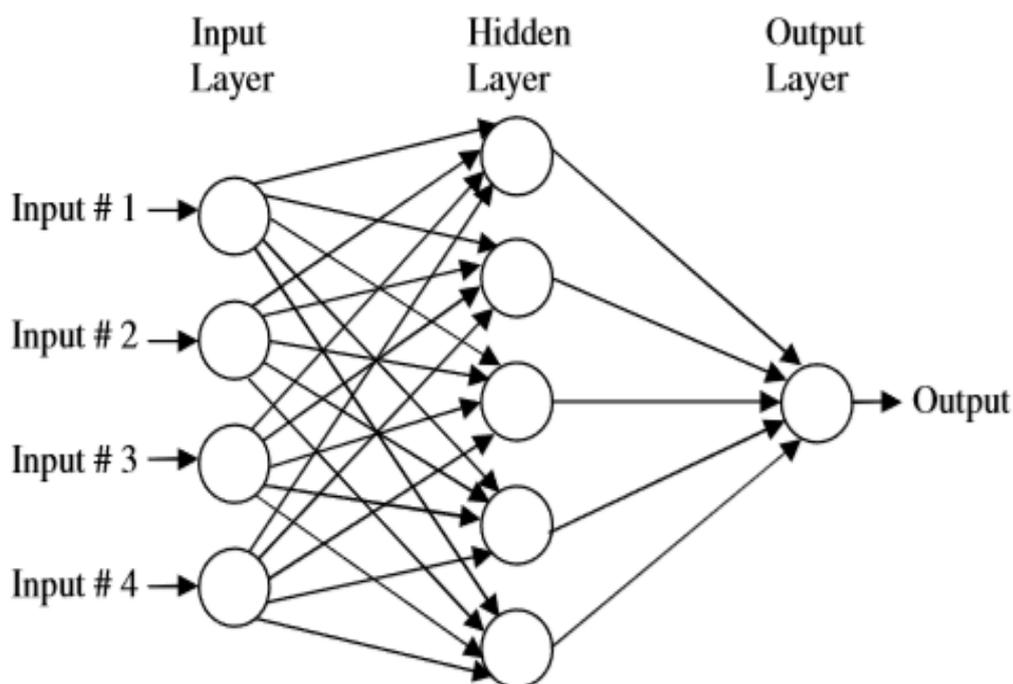


FONTE: Adaptado de Bramer (2007).

De modo geral, é possível inferir que há uma categoria principal, definida por 'SoftEng' e a partir dela, considerando as possibilidades de notas, A e B, há a classificação dos dados, no caso, se o estudante de 'SoftEng' obtiver nota B, será classificado como 'SECOND'; se o estudante obtiver nota A nas disciplinas de 'SoftEng' e 'Project', então será classificado como 'FIRST' e assim por diante, até atingir as últimas "folhas" da árvore. É importante ressaltar que para a realização completa deste processo, faz-se necessário realizar cálculos e demais procedimentos específicos. Outra aplicação da mineração poderá ocorrer por meio da previsão numérica ou regressão; esta refere-se ao desejo de, segundo Bramer (2007, p.06, tradução nossa) "[...] prever um valor numérico,

como os lucros de uma empresa ou o preço de uma ação”, diferentemente da classificação, onde o valor, objeto da previsão, é um rótulo. De acordo com o autor, uma forma de descobrir os lucros de uma organização, por exemplo, é por meio do uso de redes neurais (modelagem baseada no modelo de um neurônio humano), técnica presente na FIGURA 7:

FIGURA 7 - EXEMPLO DE REDE NEURAL

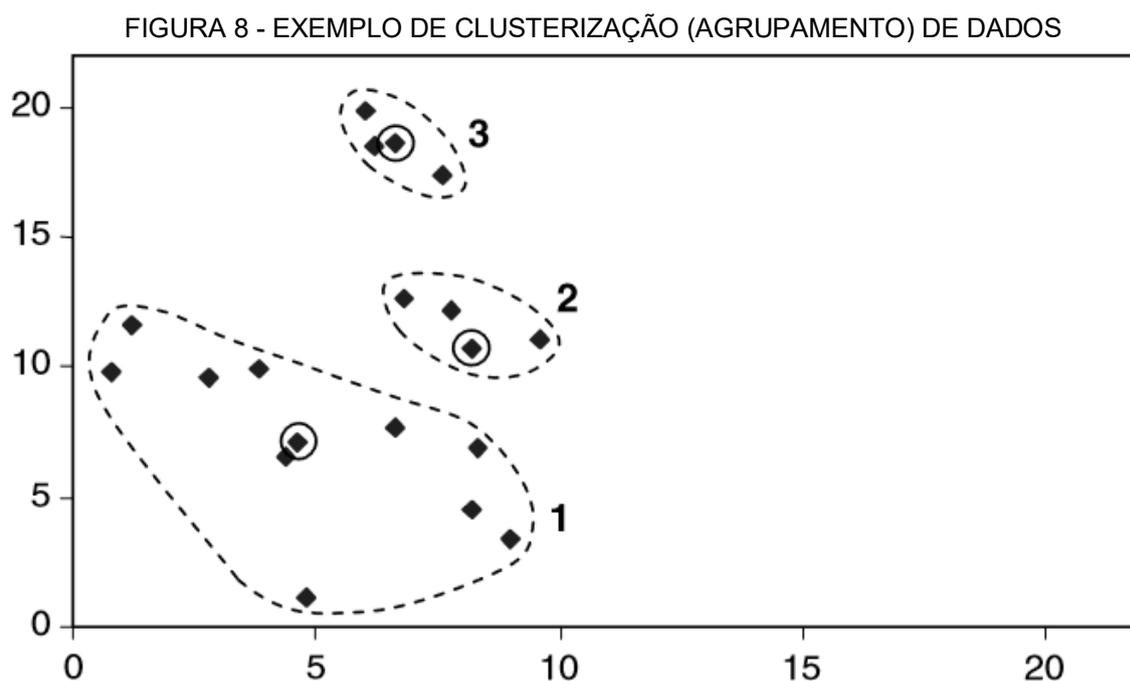


FONTE: Elaborado por Bramer (2007).

A FIGURA 7 apresenta a estrutura de uma rede neural, com as suas camadas: entrada, escondida (do meio) e saída. Além das aprendizagens supervisionadas, como supracitado, também há aplicações de aprendizagem não supervisionada e a primeira refere-se às regras de associação. Conforme explicado por Bramer (2007, p. 07, tradução nossa) “às vezes, desejamos usar um conjunto de treinamento para encontrar qualquer relação existente entre os valores das variáveis, geralmente na forma de regras conhecidas como regras de associação”, entretanto, o autor destaca que grande parte das regras não são relevantes e é por este motivo que elas apresentam informações adicionais que auxiliam na identificação de sua confiabilidade, um exemplo é a probabilidade

associada à regra, ou seja, uma porcentagem que revela a possibilidade da formulação da regra estar correta.

Por fim, há os algoritmos de clusterização ou agrupamento, estes são responsáveis por analisar os dados a fim de detectar grupos de itens semelhantes, um exemplo é a junção de clientes de acordo com renda, idade, entre outros (BRAMER; 2007, p.08). A clusterização poderá ser verificada na FIGURA 8:



FONTE: Elaborado por Bramer (2007).

O exemplo acima relaciona-se a uma aplicação de diagnóstico de falhas elétricas que podem ser agrupadas conforme os valores de variáveis específicas (BRAMER; 2007, p.08). A partir da figura é possível perceber a junção de determinados dados, conforme a marcação de uma variável distintiva e a divisão destes dados em três grupos principais, o que representa a clusterização dos dados, a ligação entre eles por meio de determinados aspectos.

Em suma, a mineração de dados é um área de estudo com grande importância, onde suas aplicações dividem-se por meio dos tipos de dados: rotulados (aprendizagem supervisionada) ou não (aprendizagem não supervisionada), além da segmentação em outros quatro tipos principais

baseadas na categoria da aprendizagem; na supervisionada encontram-se a classificação e regressão, enquanto que sem a supervisão estão as regras de associação e clusterização.

É relevante destacar que a seleção do tipo da aplicação irá variar conforme os dados e objetivos da mineração. Acerca das funcionalidades desta área, seus usos foram listados, o que possibilita reconhecer sua relevância para os diferentes campos da sociedade, dentre os quais evidencia-se a saúde, objeto deste trabalho e de discussão dos próximos tópicos.

2.6 SISTEMA ÚNICO DE SAÚDE

O Sistema Único de Saúde (SUS), basicamente, é um sistema adotado no Brasil para prestação de serviços de saúde à população, de forma gratuita; conforme explicação do Ministério da Saúde (2002), antes do SUS, ele próprio, por meio do auxílio dos estados e municípios, era o responsável por grande parte do desenvolvimento das ações relacionadas à promoção da saúde e prevenção de doenças, tais como: campanhas de vacinação e controle de epidemias; iniciativas de caráter universal, sem discriminação em relação aos beneficiados.

Um importante marco na história da saúde no Brasil ocorreu com a Constituição Federal de 1988, nela o Art. 196, revela que a “saúde é direito de todos e dever do Estado”; este trecho é essencial à compreensão da relevância do tema para o governo e população.

O Ministério da Saúde expõe que “no período anterior a CF-88, o sistema público de saúde prestava assistência apenas aos trabalhadores vinculados à Previdência Social, aproximadamente 30 milhões de pessoas com acesso aos serviços hospitalares, cabendo o atendimento aos demais cidadãos às entidades filantrópicas”, o que significa que o sistema de saúde não conseguia atender à toda a população, o que representa um problema, pois sem este suporte aos cidadãos, a promoção da saúde no país é comprometida de forma negativa.

A fim de alterar esta realidade, há o processo de criação do SUS; sobre ele,

Um passo significativo na direção do cumprimento da determinação constitucional de construção do Sistema Único de Saúde foi a publicação do decreto n.º 99.060, de 7 de março de 1990, que transferiu o INAMPS do Ministério da Previdência para o Ministério da Saúde. Esse fato, portanto, foi anterior à promulgação da Lei 8.080, que só veio a ocorrer em setembro do mesmo ano. A Lei 8.080 instituiu o Sistema Único de Saúde, com comando único em cada esfera de governo e definiu o Ministério da Saúde como gestor no âmbito da União. A Lei, no seu Capítulo II – Dos Princípios e Diretrizes, Art. 7º, estabelece entre os princípios do SUS a “universalidade de acesso aos serviços de saúde em todos os níveis de assistência”. Isso se constituiu numa grande alteração da situação até então vigente. O Brasil passou a contar com um sistema público de saúde único e universal. (MINISTÉRIO DA SAÚDE, 2002)

Diante dos aspectos e informações apresentadas, identifica-se a extrema importância que este sistema tem para toda a população, uma vez que possibilita o acesso à saúde básica e tratamento das mais diversas enfermidades; em suma, motivação que representa e complementa uma, dentre as diversas, justificativas pelas quais a escolha do tema foi feita.

3 METODOLOGIA

O objetivo desta seção é apresentar elementos iniciais ao desenvolvimento do projeto, tais como: caracterização da pesquisa; ambiente de pesquisa; dicionário de dados; e comitê de ética. Destaca-se que para a execução do trabalho serão utilizadas ferramentas (R, Excel e Power BI) na análise com estatística descritiva e para a mineração de dados o software Weka; esclarecimentos e modo de aplicação encontram-se na seção de experimentos e resultados.

3.1 CARACTERIZAÇÃO DA PESQUISA

Conforme explica Gil (2002, p. 41) “é sabido que toda e qualquer classificação se faz mediante algum critério. Com relação às pesquisas, é usual a classificação com base em seus objetivos gerais”, a partir deste raciocínio, o autor revela que é possível agrupar as pesquisas em três grupos: exploratórias, descritivas e explicativas.

A pesquisa exploratória visa possibilitar maior familiaridade com o problema para que se torne mais claro ou para a concepção de hipóteses, em suma, propõe-se a aprimorar ideias ou descobrir intuições (GIL; 2002, p. 41). Já as pesquisas descritivas, segundo Gil (2002, p. 42) “têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre a variáveis”; esta pesquisa relaciona-se à descoberta de associações entre elementos.

Por fim, há a pesquisa explicativa que se preocupa com a identificação de fatores que determinam ou contribuem o acontecimento de fenômenos, ela explica o motivo das coisas (GIL; 2002, p. 42). Este trabalho de pesquisa classifica-se como pesquisa explicativa, pois o objetivo é a identificação de padrões na base de dados selecionada, portanto, refere-se à evidenciação de possíveis associações buscando explicar determinadas questões relacionadas ao perfil de atendimento.

3.2 AMBIENTE DE PESQUISA

Esta pesquisa é realizada nas bases disponibilizadas pelo Portal da Prefeitura de Curitiba, por meio da iniciativa designada Portal de Dados Abertos e Curitiba. Como anteriormente comentado, trata-se de uma ação para o acesso à documentos, informações e dados governamentais de domínio público.

Este material é para o livre uso da sociedade, de forma a garantir acesso à dados primários, a fim de que ocorra reutilização e produção de novas informações e aplicações digitais.

No Portal, é elucidado que as bases são de diversos órgãos do Governo Municipal de Curitiba e ficam disponível para download, juntamente com o dicionário de dados que auxilia no entendimento dos campos base. É importante ressaltar que “a cada mês, novas bases serão disponibilizadas em formato aberto para uso e edição irrestrita de seus usuários, sem a necessidade de assinatura de termo ou identificação pessoal, com ou sem finalidade comercial”⁴.

Ainda há possibilidade de solicitar bases que não estejam disponíveis por meio do endereço de e-mail específico; além de informações de contato e de últimas atualizações, de modo a deixar explícito o espectro temporal e responsáveis pela base, por exemplo.

3.3 DICIONÁRIO DE DADOS

Segundo Silberschatz, Korth e Sudarshan (2006, p. 14), um dicionário de dados é aquele que “[...] armazena metadados sobre a estrutura do banco de dados, em especial o esquema do banco de dados”, ou seja, mantém dados sobre os dados de forma a auxiliar na compreensão e entendimento da base. O QUADRO 5 apresenta o dicionário de dados da base do Sistema E-Saúde referente ao perfil de atendimento dos profissionais médicos da rede de saúde do Município de Curitiba.

⁴ Fonte: <http://www.curitiba.pr.gov.br/dadosabertos/>

QUADRO 5 - DICIONÁRIO DE DADOS: SISTEMA E-SAÚDE MÉDICOS

ID	NOME DO CAMPO	DESCRIÇÃO	TIPO	TAMANHO
01	Data do atendimento*	Data de realização do atendimento	DATE	
02	Data de Nascimento*	Data de nascimento do paciente	DATE	
03	Sexo*	Sexo do paciente	VARCHAR2	1
04	Código do Tipo de Unidade*	Código do Tipo de Unidade de Atendimento	NUMBER	5
05	Tipo de Unidade*	Tipo de Unidade de Atendimento	VARCHAR2	50
06	Código da Unidade*	Código da Unidade de Atendimento	VARCHAR2	150
07	Descrição da Unidade*	Descrição da Unidade de Atendimento	VARCHAR2	80
08	Código do Procedimento*	Código do procedimento realizado	VARCHAR2	12
09	Descrição do Procedimento*	Descrição do procedimento realizado	VARCHAR2	255
10	Código do CBO*	Código da ocupação do profissional	VARCHAR2	8
11	Descrição do CBO*	Descrição da ocupação do profissional	VARCHAR2	200
12	Código do CID*	Código do diagnóstico	VARCHAR2	4
13	Descrição do CID*	Descrição do diagnóstico	VARCHAR2	150
14	Solicitação de Exames*	Indica se ocorreu solicitação de exames	VARCHAR2	3
15	Quantidade Prescrita Farmácia Curitiba*	Quantidade de medicamentos prescritos na Farmácia Curitiba	NUMBER	10
16	Quantidade Dispensada Farmácia Curitiba*	Quantidade de medicamentos dispensados na Farmácia Curitiba	NUMBER	10
17	Quantidade de Medicamento Não Padronizado*	Quantidade de medicamentos não padronizado	NUMBER	10
18	Encaminhamento para Atendimento Especialista*	Indica se houve encaminhamento para atendimento de especialista	VARCHAR2	3
19	Área de Atuação	Área de atuação	VARCHAR2	255
20	Desencadeou Internamento	Indica se desencadeou internamento	VARCHAR2	3
21	Data do Internamento	Data do internamento do paciente	DATE	

22	Estabelecimento Solicitante	Estabelecimento que solicitou o internamento	VARCHAR2	80
23	Estabelecimento Destino	Estabelecimento que houve a internação	VARCHAR2	80
24	CID do Internamento	Código do diagnóstico do internamento	VARCHAR2	4
25	Tratamento no Domicílio	Tipo de tratamento de água no domicílio	VARCHAR2	30
26	Abastecimento	Tipo de abastecimento de água no domicílio	VARCHAR2	40
27	Energia Elétrica	Indica se há energia elétrica no domicílio	VARCHAR2	3
28	Tipo de Habitação	Tipo de habitação no domicílio	VARCHAR2	60
29	Destino Lixo	Destino do lixo no domicílio	VARCHAR2	30
30	Fezes/Urina	Destino das fezes/urina no domicílio	VARCHAR2	30
31	Cômodos	Quantidade de cômodos no domicílio	NUMBER	5
32	Em Caso de Doença	Serviços procurados em caso de doença	VARCHAR2	40
33	Grupo Comunitário	Grupo comunitário em que o paciente participa	VARCHAR2	40
34	Meio de Comunicação	Meios de comunicação utilizados no domicílio	VARCHAR2	40
35	Meio de Transporte	Meios de transporte utilizados no domicílio	VARCHAR2	40
36	Município	Município do paciente	VARCHAR2	50
37	Bairro	Bairro do paciente	VARCHAR2	72
38	Nacionalidade	Nacionalidade do paciente (Brasileira, Naturalizado, Estrangeiro e Não informado)	VARCHAR2	20

FONTE: Adaptado do Dicionário de Dados do Sistema E-Saúde Médicos (2018).

Na coluna 'ID' há o número de identificação do atributo, do campo; em seguida, a coluna 'NOME DO CAMPO' apresenta a denominação do campo utilizada na base; a coluna 'DESCRIÇÃO' explica o campo, no caso, auxilia o preenchimento do mesmo por meio de elucidação do que deverá estar no campo

e a quem se refere (paciente ou profissional); a coluna tipo indica o tipo do dado, este poderá ser do tipo 'DATE', 'NUMBER' ou 'VARCHAR2'.

Conforme explicam Silberschatz, Korth e Sudarshan (2006, p. 79) o tipo *date* refere-se à “uma data de calendário contendo um ano (de quatro dígitos), mês e dia do mês”, é interessante ressaltar que apesar da definição deste tipo no dicionário, a base apresenta dados no tipo *timestamp* que de acordo com os supracitados autores (grifo dos autores) é “uma combinação de **date** e **time**”; um exemplo de *timestamp* é '2001-04-25 10:29:01:45' (SILBERSCHATZ; KORTH; SUDARSHAN, 2006).

O tipo *number*, segundo explicação presente no manual da Oracle (s/d, tradução nossa), “armazena números fixos e de ponto flutuante”; enquanto o tipo *varchar2*, ainda de acordo com o manual “[...] armazena cadeias de caracteres de comprimento variável. Ao criar uma tabela com uma coluna VARCHAR2, você especifica um comprimento máximo de cadeia (em bytes ou caracteres) entre 1 e 4000 bytes para a coluna VARCHAR2”.

A penúltima coluna representa o tamanho máximo do preenchimento do campo. Por fim, destaca-se a obrigatoriedade dos campos, em contato com o Núcleo de Informação e Tecnologia da Secretaria de Saúde de Curitiba, eles explicaram quais eram obrigatórios e quais possuem dependência de outros campos e/ou situações; o campo de área de atuação relaciona-se ao encaminhamento para atendimento especialista, o campo data do internamento, estabelecimento solicitante, estabelecimento destino, CID do internamento referem-se ao desencadeou internamento. Já os campos mais relacionados ao perfil do paciente são solicitados em casos de cadastro definitivo e aplicam-se aos campos de tratamento no domicílio, abastecimento, energia elétrica, tipo de habitação, destino lixo, fezes/urina, cômodos, em caso de doença, grupo comunitário, meio de comunicação, meio de transporte, município e bairro.

Os demais campos obrigatórios, são os iniciais na tabela (apresentam um asterisco*): data do atendimento, data de nascimento, sexo, código do tipo de unidade, tipo de unidade, código da unidade, descrição da unidade, código do procedimento, descrição do procedimento, código do CBO, descrição do CBO, código do CID, descrição do CID, solicitação de exames, encaminhamento para

especialista, quantidade prescrita Farmácia Curitibana, quantidade dispensada Farmácia Curitibana e quantidade de medicamento não padronizado. Sobre a Nacionalidade não foi questionado, uma vez que não está presente nas bases analisadas.

Em suma, trata-se de trinta e oito (38) campos para preenchimento com dados relacionados ao atendimento prestado, mais especificamente sobre os procedimentos realizados, pacientes e profissionais de saúde.

3.4 COMITÊ DE ÉTICA

Esta pesquisa tem como objeto bases da área da saúde, sobre o perfil de atendimento dos profissionais médicos nas unidades municipais. A base tem origem no Portal da Prefeitura de Curitiba e é resultado de uma iniciativa denominada Dados Abertos Curitiba, cuja finalidade, como anteriormente mencionada, está pautada na transparência e cidadania, por meio da divulgação de documentos, informações e dados governamentais de domínio público. Em contato com o Núcleo de Informação e Tecnologia da Secretaria de Saúde de Curitiba e com o Departamento de Atenção Primária, o repasse foi de que o Sistema E-Saúde relaciona-se ao DATASUS por meio de controle de faturamento e que se trata de um sistema próprio de Curitiba, fornecido pelo ICI, conforme explicado no tópico 1.1.

Dentre os objetivos específicos deste trabalho, há a proposição de comparação entre os resultados (identificação dos padrões, por exemplo) no espectro temporal de nove (9) meses. Inicialmente, seria necessário realizar uma solicitação para acessar todas as bases requeridas na pesquisa, entretanto, esta ação não será precisa, uma vez que todas as bases foram disponibilizadas via histórico diretamente no Portal de Dados Abertos⁵.

Além da disponibilização, é importante destacar que o artigo 1, parágrafo único, da Resolução nº 510, de 7 de Abril de 2016 revela que

⁵ <http://dadosabertos.c3sl.ufpr.br/curitiba/SESPAMedicoUnidadesMunicipaisDeSaude/>

Parágrafo único. Não serão registradas nem avaliadas pelo sistema CEP/CONEP:

I - pesquisa de opinião pública com participantes não identificados;

II - pesquisa que utilize informações de acesso público, nos termos da Lei n o 12.527, de 18 de novembro de 2011;

III - pesquisa que utilize informações de domínio público;

IV - pesquisa censitária;

V - pesquisa com bancos de dados, cujas informações são agregadas, sem possibilidade de identificação individual; e

VI - pesquisa realizada exclusivamente com textos científicos para revisão da literatura científica;

VII - pesquisa que objetiva o aprofundamento teórico de situações que emergem espontânea e contingencialmente na prática profissional, desde que não revelem dados que possam identificar o sujeito; e

III - atividade realizada com o intuito exclusivamente de educação, ensino ou treinamento sem finalidade de pesquisa científica, de alunos de graduação, de curso técnico, ou de profissionais em especialização.

§ 1 o Não se enquadram no inciso antecedente os Trabalhos de Conclusão de Curso, monografias e similares, devendo-se, nestes casos, apresentar o protocolo de pesquisa ao sistema CEP/CONEP;

§ 2 o Caso, durante o planejamento ou a execução da atividade de educação, ensino ou treinamento surja a intenção de incorporação dos resultados dessas atividades em um projeto de pesquisa, dever-se-á, de forma obrigatória, apresentar o protocolo de pesquisa ao sistema CEP/CONEP.

A partir do exposto, esta pesquisa caracteriza-se nos incisos II e III pois utilizará de dados públicos fornecidos com o respaldo da Lei de Acesso à Informação que visa a transparência, conforme anteriormente mencionado.

4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

É relevante mencionar que a seção anterior se refere aos encaminhamentos metodológicos, portanto, trata-se dos esclarecimentos iniciais, enquanto que esta seção, abordará, de fato, aspectos do desenvolvimento prático do projeto, com tópicos que contemplam as três principais fases da Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD): pré-processamento, mineração de dados e pós-processamento, portanto, inclui as ferramentas utilizadas, estatísticas elaboradas, aplicação do algoritmo e dos métodos selecionados, bem como discussão acerca dos resultados encontrados.

4.1 PRÉ-PROCESSAMENTO

Conforme elucidado, o pré-processamento representa o tratamento das bases antes da mineração de dados, portanto, nesta etapa foram desenvolvidas ações visando identificar *outliers*, registros faltantes e demais anomalias nos conjuntos de dados.

Ao todo, foram utilizadas quatorze (14) bases de dados que variam de fevereiro de 2018 a novembro de 2018, totalizando dez (10) meses de análise; para cada mês há um (1) conjunto de dados, entretanto, para os meses de fevereiro e junho, há três (3). No QUADRO 6 abaixo, está a nomenclatura adotada para este trabalho:

QUADRO 6 - NOMENCLATURA BASES DE DADOS

NOMENCLATURA IMPORTAÇÃO	CORRESPONDÊNCIA NO TRABALHO
2018-02-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	01FEV18 (primeiro conjunto)
2018-02-15_Sistema_E-Saude_Medicos_-_Base_de_Dados	02FEV18 (segundo conjunto)
2018-02-19_Sistema_E-Saude_Medicos_-_Base_de_Dados	03FEV18 (terceiro conjunto)
2018-03-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	MAR18

2018-04-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	ABR18
2018-05-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	MAI18
2018-06-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	01JUN18 (primeiro conjunto)
2018-06-18_Sistema_E-Saude_Medicos_-_Base_de_Dados	02JUN18 (segundo conjunto)
2018-06-20_Sistema_E-Saude_Medicos_-_Base_de_Dados	03JUN18 (terceiro conjunto)
2018-07-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	JUL18
2018-08-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	AGO18
2018-09-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	SET18
2018-10-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	OUT18
2018-11-10_Sistema_E-Saude_Medicos_-_Base_de_Dados	NOV18

FONTE: Elaborado pela autora (2019).

É necessário evidenciar que em todos os conjuntos o atributo “Nacionalidade” não está presente, portanto, não será considerado nas visualizações geradas pelo Power BI, nas aplicações do algoritmo durante a mineração de dados e em análises posteriores.

Outra questão refere-se à delimitação deste projeto, no caso, o atributo “município” apresenta registros de Curitiba e Região Metropolitana, sendo que foi definido que apenas os registros da Capital seriam utilizados para fins de análise; abaixo na TABELA 1, estão os números antes e depois deste filtro:

TABELA 1 - REGISTROS ANTES E DEPOIS DO FILTRO "CURITIBA"

ID	Base	Número de registros	
		Antes do filtro "Curitiba"	Depois do filtro "Curitiba"
1	2018-02-10	821.155	793.197
2	2018-02-15	855.938	827.046
3	2018-02-19	855.938	827.046
4	2018-03-10	783.128	751.177
5	2018-04-10	854.950	827.211
6	2018-05-10	919.260	889.321
7	2018-06-10	855.938	827.046
8	2018-06-18	919.262	889.343
9	2018-06-20	1.007.496	973.869
10	2018-07-10	919.262	889.343
11	2018-08-13	987.442	953.662
12	2018-09-10	982.643	950.805
13	2018-10-10	982.643	950.805
14	2018-11-10	962.710	932.558

FONTE: Elaborado pela autora (2019).

Além do filtro, informamos que nos casos das datas de atendimento e nascimento, apenas o mês e o ano serão utilizados, dia e horários, descartados, uma vez que não será necessário neste trabalho. Sugere-se que tais valores sejam tratados e analisados em trabalhos futuros.

4.1.1 Tratamento das Bases de Dados

As bases de dados sobre o Perfil de Atendimento Médico nas Unidades Municipais de Curitiba apresentam um total de 38 atributos, ressaltando que nas bases analisadas não há "Nacionalidade", portanto 37 atributos. Para a posterior mineração, tornou-se necessário realizar o tratamento dos atributos nominais representados abaixo:

QUADRO 7 - ATRIBUTOS NOMINAIS

Número	Atributo	Id
1	Data do Atendimento	dtAtendimento
2	Data de Nascimento	dtNascimento
3	Sexo	Sexo
4	Código do Tipo de Unidade	codTipoUnid

5	Tipo de Unidade	tipoUnid
6	Código da Unidade	codUnid
7	Descrição da Unidade	descricaoUnid
8	Código do Procedimento	codProced
9	Descrição do Procedimento	descricaoProced
10	Código do CBO	codCBO
11	Descrição do CBO	descricaoCBO
12	Código do CID	codCID
13	Descrição do CID	descricaoCID
14	Solicitação de Exames	solicitacaoExames
15	Encaminhamento para Atendimento Especialista	encamAtendEspecialista
16	Área de Atuação	areaAtuacao
17	Desencadeou Internamento	desencadeouInternamento
18	Data do Internamento	dtlInternamento
19	Estabelecimento Solicitante	estabSolicitante
20	Estabelecimento Destino	estabDestino
21	CID do Internamento	cidInternamento
22	Tratamento no Domicílio	tratamentoDomicilio
23	Abastecimento	Abastecimento
24	Energia Elétrica	energiaEletrica
25	Tipo de Habitação	tipoHabitacao
26	Destino Lixo	destinoLixo
27	Fezes/Urina	fezesUrina
28	Em Caso de Doença	casoDoenca
29	Grupo Comunitário	grupoComunitario
30	Município	Município
31	Bairro	Bairro

Fonte: Elaborado pela autora (2019).

É relevante mencionar que os atributos de “Meio de Comunicação” e “Meio de Transporte” não serão tratados e analisados, pois não fazem parte dos objetivos deste trabalho.

Em seguida, teve início a aplicação de algumas das principais tarefas de pré-processamento: limpeza, redução e transformação.

Limpeza: para imputação de valores ausentes, remoção de ruídos e correção de inconsistências;

Redução: para reduzir a dimensão da base de dados, por exemplo, agrupando ou eliminando atributos redundantes, ou para reduzir a quantidade de objetos da base, sumarizando os dados;

Transformação: para padronizar e deixar dados em um formato passível de aplicação das diferentes técnicas de mineração; (CASTRO; FERRARI, 2016, p.35)

Em um primeiro momento, buscou-se adequar o formato dos atributos do tipo data, em nosso cenário, representados por “dtAtendimento”, “dtNascimento” e “dtInternamento”; a atualização do formato resultou em manter apenas mês e ano.

Ainda em “dtNascimento”, para uma melhor visualização dos dados, optou-se por adicionar uma coluna denominada idade, em seguida, utilizando como base o Censo Demográfico do IBGE, foram delimitadas as seguintes faixas etárias:

- 0 a 4 anos;
- 5 a 9 anos;
- 10 a 14 anos;
- 15 a 19 anos;
- 20 a 24 anos;
- 25 a 29 anos;
- 30 a 34 anos;
- 35 a 39 anos;
- 40 a 44 anos;
- 45 a 49 anos;
- 50 a 54 anos;
- 55 a 59 anos;
- 60 a 64 anos;
- 65 a 69 anos;
- 70 a 74 anos;

- 75 a 79 anos;
- 80 a 84 anos;
- 85 a 89 anos;
- 90 a 94 anos;
- 95 a 99 anos;
- Mais de 100 anos.

Desta forma, o atributo de “dtNascimento” foi substituído por “faixaEtaria”; tornou-se necessário corrigir também algumas inconsistências referentes à nascimentos em 1899, uma vez que a aplicação da expressão para identificar a idade retornou um erro (#VALOR), nestes casos isolados, foram ajustados manualmente para corresponder à opção “Mais de 100 anos”. Em “codTipoUnid”, os registros eram divididos em três opções: 1, 2 e 3 que representam o “Tipo de Unidade”, portanto, os valores foram transformados para corresponder à: “BASICO”, “UPA” e “SIACE”:

QUADRO 8 - AJUSTE DO ATRIBUTO ‘codTipoUnid’

codTipoUnid	tipoUnid	codTipoUnid - ajustado
1	BASICO	B
2	UPA	U
3	SIACE	S

Fonte: Elaborado pela autora (2019).

Os atributos descartados devido à redundância, uma vez que já possuem códigos específicos, foram: “Tipo de Unidade”, “Descrição da Unidade”, “Descrição do Procedimento”, “Descrição do CBO”, “Descrição do CID” e “Município”, sendo este último descartado pois o foco deste trabalho, como mencionado, é Curitiba.

Os atributos “areaAtuacao”, “estabSolicitante” e “estabDestino” foram atualizados por meio da tarefa de transformação, no caso, os registros foram ajustados para corresponderem à códigos menores, sem espaçamento e caracteres especiais, relacionados às opções disponíveis. É possível identificar os códigos no Apêndice B

É necessário destacar que em alguns casos nos nomes dos estabelecimentos há a presença da sigla PSF (Programa Saúde da Família) e em outros não.

FIGURA 9 – EXTRATO DA TABELA DE INDICADORES - TIPO DE ESTABELECIMENTO
CENTRO DE SAÚDE/UNIDADE BÁSICA

0017094	UMS CAMARGO	76417005000186
0016888	UMS CAMPINA DO STOQUETRA	76417005000186

Fonte: Cadastro Nacional de Estabelecimentos de Saúde (s.d)

Pesquisando estes casos, foi possível identificar no Cadastro Nacional de Estabelecimentos de Saúde (CNES) que o código do estabelecimento é o mesmo, conforme exemplos nas FIGURAS 9 e 10.

FIGURA 10 – EXTRATO DA TABELA DE INDICADORES - FLUXO DE CLIENTELA -
ATENDIMENTO DE DEMANDA ESPONTANEA - ESTADO - PR

0017094	UMS CAMARGO PSF	CURITIBA
0000000	UMS CAMARGO PORTUARI PSF	CURITIBA

Fonte: Cadastro Nacional de Estabelecimentos de Saúde (s.d)

Por esta equivalência, foi considerado o mesmo código de tratamento da base para ambos. Para os atributos referentes ao aspectos do perfil dos pacientes, como tipo de tratamento no domicílio (tratamento da água), abastecimento (referente à água), energia elétrica (existência ou não de energia na residência do paciente), tipo de habitação, destino lixo e fezes/urina (destino destes resíduos na habitação do paciente) foram adotados códigos com letras, representando cada opção disponível, variando de “A” à “H”, conforme apresentado nos QUADROS 9 à 14:

QUADRO 9 - PADRONIZAÇÃO DO ATRIBUTO ‘tratamentoDomicilio’

tratamentoDomicilio	CÓDIGO
CLORACAO	A
FERVURA	B
FILTRACAO	C
SEM TRATAMENTO	D

Fonte: Elaborado pela autora (2019).

Na padronização do atributo abastecimento de água também foi utilizado código alfabético:

QUADRO 10 - PADRONIZAÇÃO DO ATRIBUTO ‘abastecimento’

abastecimento	CÓDIGO
CARRO PIPA	A

CISTERNA	B
OUTROS	C
POÇO ARTESIANO	D
REDE PÚBLICA	E

Fonte: Elaborado pela autora (2019).

O tipo de habitação refere-se à moradia do paciente:

QUADRO 11 - PADRONIZAÇÃO DO ATRIBUTO 'tipoHabitação'

tipoHabitação	CÓDIGO
MADEIRA APARELHADA	A
MATERIAL APROVEITADO	B
OUTRO MATERIAL	C
PALHA	D
TAIPA COM REVESTIMENTO	E
TAIPA SEM REVESTIMENTO	F
TIJOLO ALVENARIA SEM REVESTIMENTO	G
TIJOLO/ALVENARIA COM REVESTIMENTO	H

Fonte: Elaborado pela autora (2019).

O atributo relacionado ao destino do lixo no domicílio do paciente tem código de A a D:

QUADRO 12 - PADRONIZAÇÃO DO ATRIBUTO 'destinoLixo'

destinoLixo	CÓDIGO
CÉU ABERTO	A
COLETADO	B
OUTROS	C
QUEIMADO/ENTERRADO	D

Fonte: Elaborado pela autora (2019).

O atributo de destino de fezes e urina conta com seis (6) opções de A a F:

QUADRO 13 - PADRONIZAÇÃO DO ATRIBUTO 'fezesUrina'

fezesUrina	CÓDIGO
CÉU ABERTO	A
DIRETO PARA RIO, LAGO OU MAR	B
FOSSA RUDIMENTAR	C
FOSSA SÉPTICA	D
OUTROS	E
SISTEMA DE ESGOTO	F

Fonte: Elaborado pela autora (2019).

O grupo comunitário foi padronizado conforme quadro abaixo:

QUADRO 14 - PADRONIZAÇÃO DO ATRIBUTO 'grupoComunitario'

grupoComunitario	CÓDIGO
ASSOCIACAO	A
COOPERATIVA	B
GRUPO RELIGIOSO	C
NAO PARTICIPA	D
OUTROS	E

Fonte: Elaborado pela autora (2019).

Para o atributo casoDoenca, que representa os serviços procurados em situações de enfermidades, em cada instância/registo, havia várias possibilidades de combinação entre as opções disponíveis (benzedeira, farmácia, hospital, outros e unidade de saúde), desta forma, foram definidos códigos que representassem as possíveis junções, conforme o QUADRO 15:

QUADRO 15 - PADRONIZAÇÃO DO ATRIBUTO 'casoDoenca'

casoDoenca	CÓDIGO
BENZEDEIRA	BZ
FARMÁCIA	FM

HOSPITAL	HS
OUTROS	OT
UNIDADE DE SAÚDE	US

Fonte: Elaborado pela autora (2019).

Para o atributo bairro, foram mantidas as opções presentes em cada registro, pois não apresentaram caracteres especiais e discrepância entre fontes. Os bairros estão listados no apêndice C. O único ajuste refere-se à presença de duas opções com o mesmo significado “NAO INFORMADO” e “BAIRRO NAO INFORMADO”, para o segundo caso, houve uma alteração que correspondesse à primeira opção, decisão tomada com base no maior número de valores de cada registro.

Por fim, para os atributos relacionados à Classificação Internacional de Doenças (CID), foram adotadas classes que representam os vinte e dois (22) capítulos; conforme QUADRO 16:

QUADRO 16 – MORBIDADE HOSPITALAR DO SUS CID-10 CCAPÍTULOS

Capítulo	Descrição	Códigos da CID-10 (faixas)
I	Algumas doenças infecciosas e parasitárias	A00-B99
II	Neoplasmas [tumores]	C00-D48
III	Doenças do sangue e dos órgãos hematopoiéticos e alguns transtornos imunitários	D50-D89
IV	Doenças endócrinas, nutricionais e metabólicas	E00-E90
V	Transtornos mentais e comportamentais	F00-F99
VI	Doenças do sistema nervoso	G00-G99
VII	Doenças do olho e anexos	H00-H59
VIII	Doenças do ouvido e da apófise mastoide	H60-H95
IX	Doenças do aparelho circulatório	I00-I99
X	Doenças do aparelho respiratório	J00-J99
XI	Doenças do aparelho digestivo	K00-K93
XII	Doenças da pele e do tecido subcutâneo	L00-L99

XIII	Doenças do sistema osteomuscular e do tecido conjuntivo	M00-M99
XIV	Doenças do aparelho geniturinário	N00-N99
XV	Gravidez, parto e puerpério	O00-O99
XVI	Algumas afecções originadas no período perinatal	P00-P96
XVII	Malformações congênicas, deformidades e anomalias cromossômicas	Q00-Q99
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	R00-R99
XIX	Lesões, envenenamentos e algumas outras conseqüências de causas externas	S00-T98
XX	Causas externas de morbidade e de mortalidade	V01-Y98
XXI	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde	Z00-Z99
**	CID 10ª Revisão não disponível ou não preenchido ou inválido	U99, em branco ou inválido

Fonte: Adaptado de Ministério da Saúde – DATASUS (s.d).

Para este processo de aplicação das faixas, foi necessário considerar a hierarquia do CID, listada na página do Centro Colaborador da OMS para a Classificação de Doenças em Português – CBCD:

- CID-10-CAPÍTULOS;
- CID-10-GRUPOS;
- CID-10-CATEGORIAS;
- CID-10-SUBCATEGORIAS.

A partir dela procurou-se identificar nas bases se o atributo de código do CID atendia aos critérios de formato e padrão para serem considerados nas faixas. Ao realizar esta análise, observou-se que a única diferença é que as bases do perfil de atendimento chegam ao nível de subcategoria (formato com quatro caracteres alfanuméricos), entretanto, por estarmos considerando o

tratamento para capítulos, foram utilizadas apenas as categorias (formato com três caracteres alfanuméricos), conforme exemplo abaixo:

- CID-10-CAPÍTULOS (formato: Z00-Z99);
- CID-10-CATEGORIAS (formato: A00);
- CID-10-SUBCATEGORIAS (formato: A000)

Ou seja, apesar das bases apresentarem os códigos até o nível de subcategoria, para o tratamento, foram consideradas apenas as categorias, que posteriormente tornaram-se capítulos. Enfatiza-se a necessidade desta ação, pois ao definir o código do CID como atributo meta, havia muitas opções, mais de quatro mil (4.000) apenas na base de 1FEV18, deste modo o processamento da mineração no software ficou comprometido, o que demandou a criação destas classes a partir dos capítulos mencionados.

4.1.2 Estatística Descritiva

Os objetivos da estatística descritiva, conforme revela Martins (2006, p.25) “[...] envolvem organização, sumarização e descrição de dados quantitativos ou qualitativos”. Para este projeto, foram calculadas medidas de tendência central e de dispersão, de modo a identificar valores que caracterizem e auxiliem no entendimento dos conjuntos de dados, no que se refere aos atributos numéricos (quantidade de medicamento prescrito, quantidade de medicamento dispensado, quantidade de medicamento não padronizado e cômodos no domicílio do paciente).

Para realizar as estatísticas necessárias, foram utilizadas duas ferramentas: Excel e RStudio. A primeira, basicamente, relaciona-se à criação de planilhas eletrônicas, bem como a elaboração de cálculos. Enquanto a linguagem R

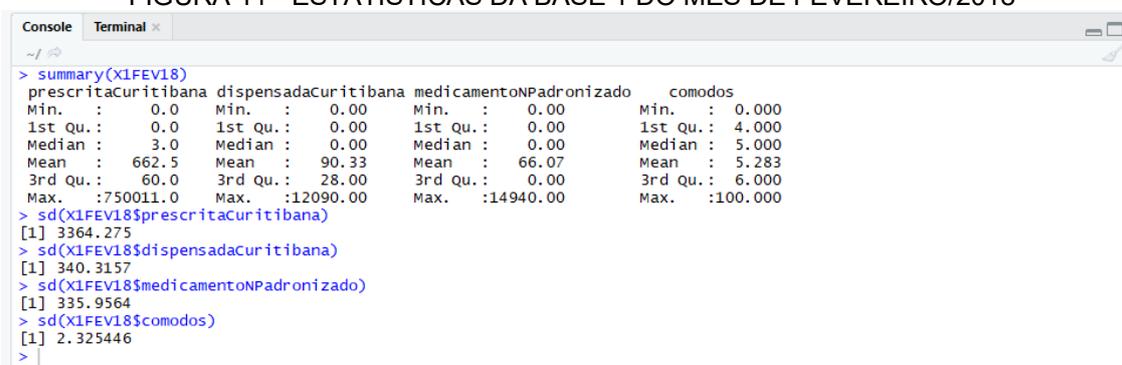
[...] é um software livre com uma linguagem dirigida à análise gráfica e estatística, semelhante à linguagem de computação S. A linguagem R, assim como a S, é designada como uma verdadeira linguagem de computação, permitindo aos usuários adicionar novas funções e implementar novas rotinas. Na verdade, o sistema R é um ambiente dentro do quais técnicas estatísticas podem ser implementadas. (PETERNELLI et al., 2009, p.850)

O que representa a possibilidade de aplicação de pacotes e comandos para um retorno estatístico, enquanto o RStudio⁶ “[...] is an *integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management*”.

Com o Excel foi possível realizar a extração inicial dos atributos numéricos (medicamentos prescritos na Farmácia Curitibana, medicamentos dispensados na Farmácia Curitibana, medicamentos não padronizados e número de cômodos na residência do paciente) após o filtro de município “Curitiba”. É importante ressaltar que para os casos de fevereiro e junho, em que o número de bases é maior, foi utilizado um número para identificar a qual base refere-se o cálculo (variando de 1 a 3).

Após a extração, foi realizado o *upload*, separadamente, dos arquivos no RStudio, para o cálculo das medidas anteriormente mencionadas; a FIGURA 11 revela um exemplo da saída no software estatístico:

FIGURA 11 - ESTATÍSTICAS DA BASE 1 DO MÊS DE FEVEREIRO/2018



```

Console Terminal x
~/
> summary(X1FEV18)
prescritaCuritibana dispensadaCuritibana medicamentoNPadronizado comodos
Min. : 0.0 Min. : 0.00 Min. : 0.00 Min. : 0.000
1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 4.000
Median : 3.0 Median : 0.00 Median : 0.00 Median : 5.000
Mean : 662.5 Mean : 90.33 Mean : 66.07 Mean : 5.283
3rd Qu.: 60.0 3rd Qu.: 28.00 3rd Qu.: 0.00 3rd Qu.: 6.000
Max. :750011.0 Max. :12090.00 Max. :14940.00 Max. :100.000
> sd(X1FEV18$prescritaCuritibana)
[1] 3364.275
> sd(X1FEV18$dispensadaCuritibana)
[1] 340.3157
> sd(X1FEV18$medicamentoNPadronizado)
[1] 335.9564
> sd(X1FEV18$comodos)
[1] 2.325446
>

```

FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

As demais saídas, com duas casas decimais, encontram-se reproduzidas na TABELA 2, que foi dividida em três partes: a primeira de 1FEV18 a ABR18; a segunda de MAI18 a JUL18; e a terceira de AGO18 a NOV18:

⁶ Fonte: <https://www.rstudio.com/products/rstudio/>

TABELA 2 - ESTATÍSTICAS (parte um)⁷

BASE	ATRIBUTO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	FIQ	OUTLIER BAIXO	OUTLIER ALTO	DESVIO PADRÃO
01 FEV-18	prescritaCuritiba	0,00	0,00	3,00	662,50	60,00	750011,00	60,00	-90,00	150,00	3364,28
	dispensadaCuritiba	0,00	0,00	0,00	90,33	28,00	12090,00	28,00	-42,00	70,00	340,32
	medicamentoNPadronizado	0,00	0,00	0,00	66,07	0,00	14940,00	0,00	0,00	0,00	335,96
	comodos	0,00	4,00	5,00	5,28	6,00	100,00	2,00	1,00	9,00	2,33
02 FEV-18	prescritaCuritiba	0,00	0,00	3,00	5258,80	58,00	200000021,00	58,00	-87,00	145,00	3110149,85
	dispensadaCuritiba	0,00	0,00	0,00	60,58	21,00	8440,00	21,00	-31,50	52,50	208,42
	medicamentoNPadronizado	0,00	0,00	0,00	63,99	0,00	99999,00	0,00	0,00	0,00	353,86
	comodos	-4,00	4,00	5,00	5,36	6,00	3000,00	2,00	1,00	9,00	7,50
03 FEV-18	prescritaCuritiba	0,00	0,00	3,00	5258,80	58,00	200000021,00	58,00	-87,00	145,00	3110149,85
	dispensadaCuritiba	0,00	0,00	0,00	60,58	21,00	8440,00	21,00	-31,50	52,50	208,42
	medicamentoNPadronizado	0,00	0,00	0,00	63,99	0,00	99999,00	0,00	0,00	0,00	353,86
	comodos	-4,00	4,00	5,00	5,36	6,00	3000,00	2,00	1,00	9,00	7,50
MAR-18	prescritaCuritiba	0,00	0,00	4,00	165328,61	61,00	119999999988,00	61,00	-91,50	152,50	137946224,42
	dispensadaCuritiba	0,00	0,00	0,00	88,85	30,00	12090,00	30,00	-45,00	75,00	322,34
	medicamentoNPadronizado	0,00	0,00	0,00	69,64	1,00	99999,00	1,00	-1,50	2,50	365,52
	comodos	-5,00	4,00	5,00	5,40	6,00	1500,00	2,00	1,00	9,00	11,00
ABR-18	prescritaCuritiba	0,00	0,00	4,00	146779,38	61,00	119999999988,00	61,00	-91,50	152,50	131941193,99
	dispensadaCuritiba	0,00	0,00	0,00	82,19	30,00	12090,00	30,00	-45,00	75,00	287,86
	medicamentoNPadronizado	0,00	0,00	0,00	69,26	1,00	20500,00	1,00	-1,50	2,50	344,37
	comodos	-5,00	4,00	5,00	5,39	6,00	1500,00	2,00	1,00	9,00	10,06

FONTE: Elaborado pela autora (2019).

⁷ Tabela continua.

TABELA 2 - ESTATÍSTICAS (parte dois)⁸

BASE	ATRIBUTO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	FIQ	OUTLIER BAIXO	OUTLIER ALTO	DESVIO PADRÃO
MAI-18	prescritaCuritiba	0,00	0,00	3,00	1774,16	60,00	800000094,00	60,00	-90,00	150,00	875861,64
	dispensadaCuritiba	0,00	0,00	0,00	70,77	23,00	9030,00	23,00	-34,50	57,50	246,84
	medicamentoNPadronizado comodos	0,00	0,00	0,00	66,46	0,00	20500,00	0,00	0,00	0,00	335,82
01 JUN-18	prescritaCuritiba	0,00	0,00	5,00	5,38	6,00	5555,00	2,00	1,00	9,00	10,97
	dispensadaCuritiba	0,00	0,00	3,00	5258,80	58,00	200000021,00	58,00	-87,00	145,00	3110149,85
	medicamentoNPadronizado comodos	0,00	0,00	0,00	60,58	21,00	8440,00	21,00	-31,50	52,50	208,42
02 JUN-18	prescritaCuritiba	0,00	0,00	5,00	63,99	0,00	99999,00	0,00	0,00	0,00	353,86
	dispensadaCuritiba	-4,00	4,00	5,00	5,36	6,00	3000,00	2,00	1,00	9,00	7,50
	medicamentoNPadronizado comodos	0,00	0,00	3,00	1683,68	60,00	800000094,00	60,00	-90,00	150,00	875711,18
03 JUN-18	prescritaCuritiba	0,00	0,00	0,00	62,54	22,00	9030,00	22,00	-33,00	55,00	212,32
	dispensadaCuritiba	0,00	0,00	0,00	66,46	0,00	20500,00	0,00	0,00	0,00	335,81
	medicamentoNPadronizado comodos	0,00	4,00	5,00	5,38	6,00	5555,00	2,00	1,00	9,00	11,09
JUN-18	prescritaCuritiba	0,00	0,00	3,00	1542,33	54,00	800000094,00	54,00	-81,00	135,00	836942,76
	dispensadaCuritiba	0,00	0,00	0,00	55,15	20,00	9030,00	20,00	-30,00	50,00	190,28
	medicamentoNPadronizado comodos	0,00	0,00	0,00	62,24	0,00	199998,00	0,00	0,00	0,00	428,45
JUL-18	prescritaCuritiba	-6,00	4,00	5,00	5,50	6,00	5555,00	2,00	1,00	9,00	80,81
	dispensadaCuritiba	0,00	0,00	3,00	1683,68	60,00	800000094,00	60,00	-90,00	150,00	875711,18
	medicamentoNPadronizado comodos	0,00	0,00	0,00	62,54	22,00	9030,00	22,00	-33,00	55,00	212,32
JUL-18	prescritaCuritiba	0,00	0,00	0,00	66,46	0,00	20500,00	0,00	0,00	0,00	335,81
	dispensadaCuritiba	0,00	4,00	5,00	5,38	6,00	5555,00	2,00	1,00	9,00	11,09
	medicamentoNPadronizado comodos	0,00	0,00	0,00	66,46	0,00	20500,00	0,00	0,00	0,00	335,81

FONTE: Elaborado pela autora (2019).

⁸ Tabela continua.

TABELA 2 - ESTATÍSTICAS (parte três)

BASE	ATRIBUTO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	FIQ	OUTLIER BAIXO	OUTLIER ALTO	DESVIO PADRÃO
AGO-18	prescritaCuritiba	0,00	0,00	4,00	464,33	58,00	7200048,00	58,00	-87,00	145,00	24519,82
	dispensadaCuritiba	0,00	0,00	0,00	48,10	20,00	20616,00	20,00	-30,00	50,00	165,19
	medicamentoNPadronizado	0,00	0,00	0,00	63,02	0,00	667500,00	0,00	0,00	0,00	1130,54
	comodos	-6,00	4,00	5,00	5,46	6,00	55555,00	2,00	1,00	9,00	81,28
SET-18	prescritaCuritiba	0,00	0,00	4,00	450,64	60,00	7200048,00	60,00	-90,00	150,00	23549,05
	dispensadaCuritiba	0,00	0,00	0,00	47,64	21,00	7330,00	21,00	-31,50	52,50	155,10
	medicamentoNPadronizado	0,00	0,00	0,00	65,53	0,00	667500,00	0,00	0,00	0,00	1137,62
	comodos	-4,00	4,00	5,00	5,32	6,00	1500,00	2,00	1,00	9,00	5,82
OUT-18	prescritaCuritiba	0,00	0,00	4,00	450,64	60,00	7200048,00	60,00	-90,00	150,00	23549,05
	dispensadaCuritiba	0,00	0,00	0,00	47,64	21,00	7330,00	21,00	-31,50	52,50	155,10
	medicamentoNPadronizado	0,00	0,00	0,00	65,53	0,00	667500,00	0,00	0,00	0,00	1137,62
	comodos	-4,00	4,00	5,00	5,32	6,00	1500,00	2,00	1,00	9,00	5,82
NOV-18	prescritaCuritiba	0,00	0,00	4,00	1514,11	60,00	300300012,00	60,00	-90,00	150,00	549069,12
	dispensadaCuritiba	0,00	0,00	0,00	49,27	22,00	18056,00	22,00	-33,00	55,00	161,13
	medicamentoNPadronizado	0,00	0,00	0,00	65,38	0,00	360120,00	0,00	0,00	0,00	648,08
	comodos	-4,00	4,00	5,00	5,33	6,00	1500,00	2,00	1,00	9,00	6,46

FONTE: Elaborado pela autora (2019).

Conforme mencionado, para compreender o conjunto de dados, uma das estratégias adotadas, foi a aplicação de cálculos estatísticos com o RStudio e Excel.

As medidas representam:

- Mínimo: o menor valor encontrado na base;
- Primeiro Quartil (1º Quartil): segundo Martins (2006, p. 42) “os quartis dividem um conjunto de dados em quatro partes iguais”, desta forma, o 1º quartil representa 25% dos elementos;
- Mediana: de acordo com Martins (2006, p. 40) “é o valor que divide a amostra, ou população, em duas partes iguais”, também é o valor do segundo quartil;
- Terceiro Quartil (3º Quartil): representa 75% dos elementos, conforme Martins (2006, p.43) “ $Q_3 = 3^\circ$ Quartil, deixa 75% dos elementos”;
- Média ($\bar{X} = \frac{\sum x}{n}$): refere-se à soma de todos os dados do conjunto dividida pelo total de elementos;
- Máximo: o maior valor encontrado na base;
- Faixa interquartil (FIQ ou Amplitude Interquartil): calculada pela diferença entre o terceiro e o primeiro quartil;
- Desvio padrão: entendido como

[...] uma medida de dispersão e o seu valor reflete a variabilidade das observações em relação à média.

[...] Se existir uma grande dispersão das observações a variância é grande. Se os valores de cada uma das observações forem próximos da média a variância é pequena. (LUNET et al., 2006, p.55)

O desvio padrão populacional tem como símbolo o σ (sigma). Acerca dos resultados, o menor valor encontrado em “mínimo” foi -6 nos meses de junho e agosto, em ambos os casos para o atributo “comodos”, ou seja, paciente(s) com número negativo de cômodos em sua residência.

O primeiro quartil (Q_1) resultou em 0 para os atributos de “prescritaCuritibana”, “dispensadaCuritibana” e “medicamentoNPadronizado” e em 4 para “comodos”, ou seja, 25% do conjunto de dados para cada mês, nos

primeiros três (03) atributos tem os valores menores ou iguais a 0 e menores ou iguais a 4 no caso dos cômodos.

A mediana, também representa o Q_2 , e significa que metade das observações estão acima e metade abaixo do valor encontrado, portanto, divide o conjunto dados. Nos meses de fevereiro, março, junho e julho a mediana foi 3 para o atributo “prescritaCuritiba” e 4 nos meses de março, abril, agosto, setembro, outubro e novembro. O valor 5 de mediana foi identificado em todos os meses no atributo “comodos”, para os demais atributos (“dispensadaCuritiba” e “medicamentoNPadronizado”) o valor encontrado foi 0, em todos os meses.

A média, ou seja, soma de todos os dados divididos pelo total de observações, depende do mês, de modo geral, o maior valor da média é 165.328,61 no mês de março, atributo “prescritaCuritiba” e o menor valor é 5,28 no mês de fevereiro (primeiro conjunto) no atributo “comodos”.

O terceiro quartil (Q_3), assim como o Q_1 relaciona-se aos dados serem menores ou iguais a determinada porcentagem, neste caso, a 75%. O maior valor identificado foi 61, encontrado nos meses de março e abril para o atributo “prescritaCuritiba” e o menor valor foi 0 identificado em todos meses, menos em março e abril para o atributo “medicamentoNPadronizado”

O máximo representa o maior valor encontrado no conjunto de dados, no caso, 119.999.999.988 identificado no atributo “prescritaCuritiba” nos meses de março e abril.

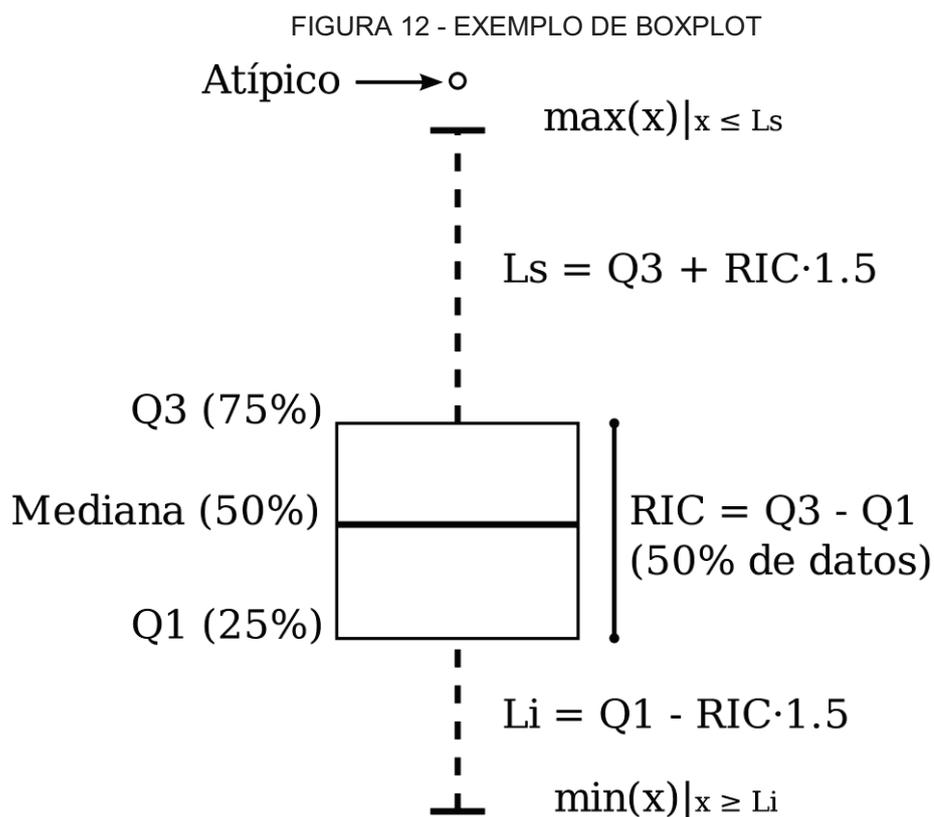
A faixa interquartil (FIQ) é o valor do terceiro menos o primeiro quartil. Alguns destaques nesta medida é que o valor máximo foi 61 em março e abril para “prescritaCuritiba” e 2, em todos os meses, para o atributo “comodos”.

Quanto às medidas de *outliers* baixos e altos, os valores expressivos estão em março e abril, -91,50 e 152,50, respectivamente para o atributo “prescritaCuritiba”.

Por fim, o desvio padrão apresentou valores elevados para os atributos “prescritaCuritiba”, “dispensadaCuritiba” e “medicamentoNPadronizado” em todos os meses, chegando ao valor máximo de 137.946.224,42 no mês de

março para os medicamentos prescritos. O menor valor e, portanto, significando maior homogeneidade amostral é identificado no atributo “comodos” (2,33) em fevereiro (primeiro conjunto).

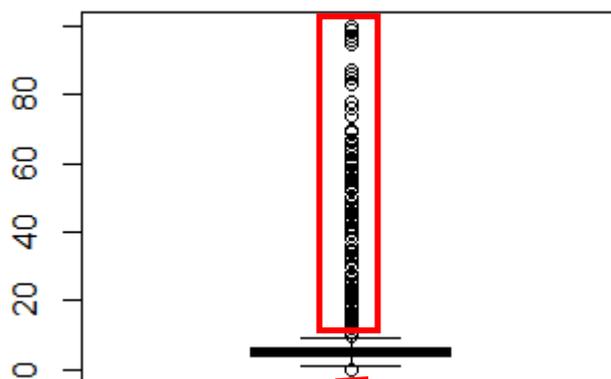
Outro recurso utilizado para compreender o conjunto de dados, foi a elaboração de gráficos do tipo Boxplot (ou Diagrama da Caixa), conforme exemplo na FIGURA 12:



FONTE: Wikimedia Commons (2012).

No Boxplot foram encontrados mínimo, máximo, média, Q_1 , Q_3 , mediana, limite superior, limite inferior e atípicos (*outliers*). Foram gerados Boxplots para todos os conjuntos de dados e atributos numéricos. Para o primeiro conjunto do mês de fevereiro, as FIGURAS 13 a 16 apresentam esses gráficos:

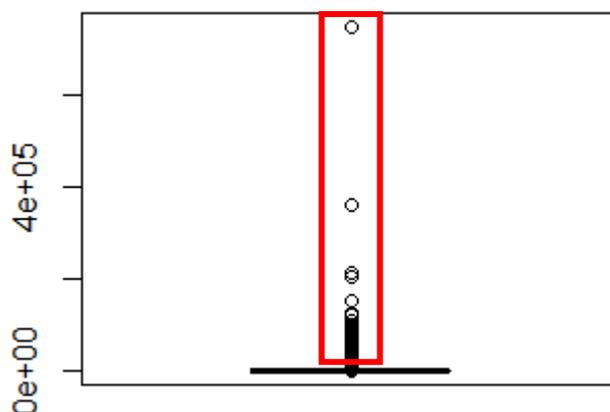
FIGURA 13 - BOXPLOT ATRIBUTO 'comodos' 01FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

O boxplot da FIGURA 13 deixa bem demarcado os valores de primeiro e terceiro quartil representados pelo início e término da caixa, bem como os valores máximo e mínimo identificados, o que possibilita a verificação do que é *outlier*, ou seja, dos valores acima do máximo e mínimo (em destaque na cor vermelha), que destoam dos demais resultados identificados no conjunto de dados, portanto, a visualização com o boxplot permite essa verificação e entendimento de que o conjunto de dados apresenta um grande número de outliers. A FIGURA 14 apresenta o resultado para o atributo de medicamentos prescritos na Farmácia Curitiba:

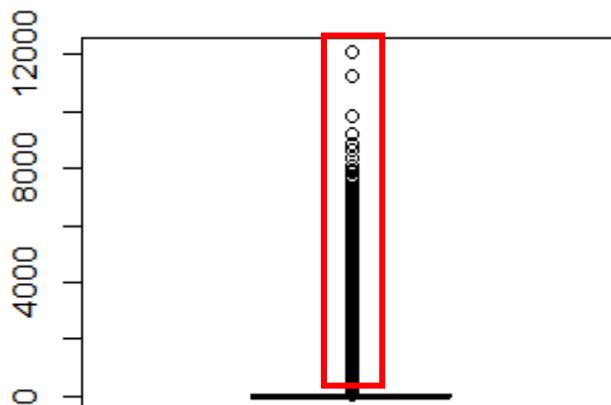
FIGURA 14 - BOXPLOT ATRIBUTO 'prescritaCuritiba' 01FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

Na figura 15 há o gráfico de boxplot para os medicamentos dispensados na Farmácia Curitiba:

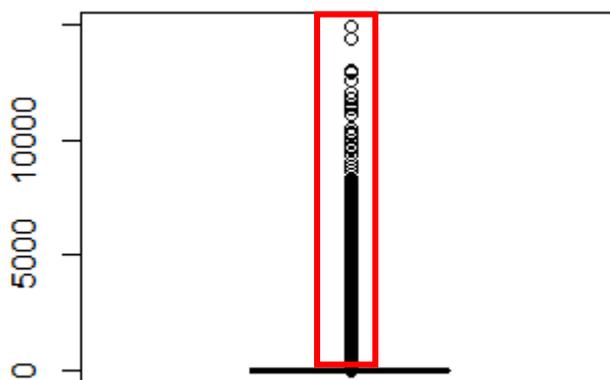
FIGURA 15 - BOXPLOT ATRIBUTO 'dispensadaCuritiba' 01FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

O boxplot sobre medicamentos não padronizados pode ser verificado na FIGURA 16 e conforme a caixa na cor vermelha, o atributo também apresenta grande quantidade de outliers:

FIGURA 16 - BOXPLOT ATRIBUTO 'medicamentoNPadronizado' 01FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

Alguns outros boxplots encontram-se no apêndice A. Após a elaboração de todos os gráficos, ficou claro que há muitos *outliers* e que os atributos numéricos não apresentam valores significativos para este trabalho, como número de cômodos negativos e/ou valores elevados para medicamentos prescritos, sendo necessário ressaltar que são valores absolutos, portanto, os

atributos “prescritaCuritibana”, “dispensadaCuritibana”, “medicamentoNPadronizado” e “comodos”, foram descartados de outras análises e Mineração de Dados, uma vez que também não fazem parte do foco deste trabalho.

Outro ponto a ser destacado é que durante as análises estatísticas, foram observadas bases idênticas em diferentes meses, em vista disso, também foram descartadas. Ressaltamos que as bases semelhantes foram identificadas por meio do tamanho do arquivo e resultados das estatísticas, sinais que possibilitaram em posterior análise, constatar, de fato, a correspondência entre as bases:

- 15 fevereiro de 2018 e 10 de junho de 2018;
- 10 de setembro de 2018 e 10 de outubro de 2018;
- 18 de junho de 2018 e 10 de julho de 2018.

O descarte não ocorreu para todas as bases, mas sim entre as iguais, no caso, foram preservadas as bases de fevereiro, setembro e julho; o critério para seleção contou com a opção por meses com apenas um conjunto de dados, no caso de outubro, esta opção mostrou-se inviável, portanto, para não perder a sequência das bases, definiu-se o trabalho com apenas uma base de cada mês, terminando em setembro. Neste momento, passaram-se para oito (8) bases.

Destaca-se que durante a análise dos atributos numéricos, percebeu-se que para o atributo “Data de Atendimento” havia recorrência do mesmo mês em diferentes bases:

QUADRO 17 - REPETIÇÃO DOS MESES DE ATENDIMENTO

BASE	DATA DE ATENDIMENTO (MÊS)
1FEV18	NOV17
	DEZ17
	JAN18
	FEV18
MAR18	DEZ17
	JAN18
	FEV18

ABR18	JAN18
	FEV18
	MAR18
MAI18	FEV18
	MAR18
	ABR18
	MAI18
1JUN18	MAR18
	ABR18
	MAI18
JUL18	FEV18
	MAR18
	ABR18
AGO18	MAI18
	JUN18
	JUL18
SET18	JUN18
	JUL18
	AGO18
	SET18

Fonte: Elaborado pela autora (2019).

A partir deste cenário, optou-se por filtrar apenas um mês de cada base, de modo a formar uma sequência de meses analisados, conforme quadro abaixo com os elementos selecionados na cor verde:

QUADRO 18 - DEFINIÇÃO DAS BASES ANALISADAS E MINERADAS

BASE	DATA DE ATENDIMENTO (MÊS)
1FEV18	NOV17
	DEZ17
	JAN18
	FEV18
MAR18	DEZ17

	JAN18
	FEV18
	JAN18
ABR18	FEV18
	MAR18
MAI18	FEV18
	MAR18
	ABR18
	MAI18
1JUN18	MAR18
	ABR18
	MAI18
JUL18	FEV18
	MAR18
	ABR18
AGO18	MAI18
AGO18	JUN18
AGO18	JUL18
SET18	JUN18
SET18	JUL18
SET18	AGO18
SET18	SET18

Fonte: Elaborado pela autora (2019).

Ao final, passaram pela segunda parte da análise, com atributos nominais, e pela mineração de dados, um total de nove (9) bases, variando de janeiro de 2018 a setembro de 2018.

Os atributos remanescentes totalizam vinte e cinco (25) campos e são do tipo nominal, segundo, Guimarães (2008, p.23, grifo do autor), “As variáveis qualitativas apresentam como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado[...] *variável qualitativa categórica ou nominal*, para a qual não existe nenhuma ordenação nas possíveis realizações[...]”.

Para estes atributos foram calculadas as frequências absolutas e relativas de acordo com o mês (janeiro a setembro), sendo que a cada trimestre está destacado o maior valor identificado com uma caixa na cor verde. Iniciando com atributos relacionados à consulta, na TABELA 3 estão os resultados para o atributo “faixaEtaria”, que conta com vinte e uma (21) opções:

TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
20-24	22082	8,27%	20-24	20012	7,86%	0-4	27458	8,99%
25-29	20942	7,84%	25-29	18952	7,45%	20-24	23714	7,76%
50-54	19521	7,31%	50-54	18182	7,14%	25-29	22106	7,24%
55-59	19460	7,29%	55-59	18116	7,12%	55-59	20641	6,76%
60-64	18543	6,94%	0-4	18081	7,10%	50-54	20281	6,64%
30-34	17596	6,59%	60-64	17744	6,97%	60-64	19385	6,35%
0-4	17290	6,47%	30-34	16376	6,43%	30-34	18996	6,22%
45-49	17169	6,43%	45-49	16111	6,33%	45-49	18708	6,12%
35-39	17012	6,37%	65-69	15969	6,27%	5-9	18196	5,96%
65-69	16867	6,32%	35-39	15913	6,25%	35-39	18084	5,92%
40-44	16542	6,19%	40-44	15875	6,24%	65-69	17693	5,79%
70-74	13356	5,00%	70-74	13159	5,17%	40-44	17471	5,72%
5-9	10322	3,86%	5-9	11118	4,37%	15-19	14236	4,66%
15-19	10117	3,79%	15-19	10303	4,05%	70-74	14050	4,60%
75-79	9835	3,68%	75-79	9221	3,62%	10-14	12233	4,00%
10-14	8121	3,04%	10-14	8277	3,25%	75-79	10069	3,30%
80-84	6759	2,53%	80-84	6173	2,42%	80-84	6732	2,20%
85-89	3637	1,36%	85-89	3276	1,29%	85-89	3649	1,19%
90-94	1459	0,55%	90-94	1312	0,52%	90-94	1413	0,46%
95-99	377	0,14%	95-99	314	0,12%	95-99	318	0,10%
Mais de 100 anos	86	0,03%	Mais de 100 anos	75	0,03%	Mais de 100 anos	49	0,02%

Fonte: Elaborado pela autora (2019).

Para uma melhor visualização, a tabela foi dividida em três (3) partes contemplando três (3) meses cada.

TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
0-4	32891	9,99%	0-4	33668	9,93%	0-4	29284	9,27%
20-24	23956	7,28%	5-9	28551	8,42%	5-9	23532	7,45%
5-9	23241	7,06%	20-24	23970	7,07%	20-24	23275	7,36%
25-29	22756	6,91%	25-29	22382	6,60%	25-29	21683	6,86%
50-54	21599	6,56%	50-54	21606	6,37%	50-54	20533	6,50%
55-59	21319	6,48%	55-59	21523	6,35%	55-59	20352	6,44%
60-64	20609	6,26%	60-64	20932	6,17%	60-64	19581	6,20%
45-49	19628	5,96%	35-39	19485	5,74%	35-39	18581	5,88%
30-34	19568	5,94%	30-34	19449	5,73%	45-49	18523	5,86%
35-39	19068	5,79%	45-49	19315	5,69%	30-34	18490	5,85%
40-44	18870	5,73%	65-69	18845	5,56%	40-44	18079	5,72%
65-69	18860	5,73%	40-44	18725	5,52%	65-69	18072	5,72%
15-19	15041	4,57%	10-14	16466	4,85%	10-14	14869	4,70%
70-74	14820	4,50%	15-19	15313	4,51%	70-74	14412	4,56%
10-14	13913	4,23%	70-74	15219	4,49%	15-19	14324	4,53%
75-79	10577	3,21%	75-79	10876	3,21%	75-79	10051	3,18%
80-84	6890	2,09%	80-84	7189	2,12%	80-84	6839	2,16%
85-89	3721	1,13%	85-89	3729	1,10%	85-89	3664	1,16%
90-94	1388	0,42%	90-94	1516	0,45%	90-94	1497	0,47%
95-99	374	0,11%	95-99	373	0,11%	95-99	344	0,11%
Mais de 100 anos	64	0,02%	Mais de 100 anos	74	0,02%	Mais de 100 anos	71	0,02%

Fonte: Elaborado pela autora (2019).

Abaixo está a tabela referente à frequência da faixaEtaria de julho a setembro de 2018:

TABELA 3 - FREQUÊNCIA ATRIBUTO 'faixaEtaria' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa	faixaEtaria	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
0-4	25789	8,64%	0-4	28254	8,40%	40-44	13	16,46%
10-14	9682	3,24%	20-24	24596	7,32%	30-34	11	13,92%
15-19	12290	4,12%	55-59	23509	6,99%	35-39	9	11,39%
20-24	22602	7,57%	25-29	23414	6,97%	20-24	8	10,13%
25-29	21099	7,07%	50-54	23252	6,92%	15-19	7	8,86%
30-34	18500	6,20%	60-64	22844	6,80%	25-29	6	7,59%
35-39	18107	6,07%	45-49	21005	6,25%	50-54	6	7,59%
40-44	18184	6,09%	30-34	20402	6,07%	0-4	4	5,06%
45-49	19022	6,37%	65-69	20358	6,06%	10-14	3	3,80%
50-54	20917	7,01%	40-44	20217	6,01%	55-59	3	3,80%
55-59	21686	7,26%	35-39	20135	5,99%	5-9	3	3,80%
5-9	14542	4,87%	5-9	17335	5,16%	65-69	3	3,80%
60-64	20146	6,75%	70-74	16695	4,97%	60-64	2	2,53%
65-69	18173	6,09%	15-19	15416	4,59%	45-49	1	1,27%
70-74	14602	4,89%	10-14	12876	3,83%			
75-79	10504	3,52%	75-79	11924	3,55%			
80-84	7011	2,35%	80-84	7721	2,30%			
85-89	3887	1,30%	85-89	4242	1,26%			
90-94	1411	0,47%	90-94	1581	0,47%			
95-99	324	0,11%	95-99	336	0,10%			
Mais de 100 anos	54	0,02%	Mais de 100 anos	49	0,01%			

Fonte: Elaborado pela autora (2019).

A faixa etária predominante entre os meses é a de 0 a 4 anos, tendo sido a que apresenta o maior valor de frequência absoluta e relativa nos meses de março, abril, maio, julho e agosto.

O atributo sexo apresentou, em todos os meses, a opção de feminino (F) com as porcentagens de frequência relativa em primeiro lugar, com uma média de 63,72%.

TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
sexo	Frequência Absoluta	Frequência Relativa	sexo	Frequência Absoluta	Frequência Relativa	sexo	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
F	173343	64,90%	F	165614	65,06%	F	195701	64,06%
M	93750	35,10%	M	88945	34,94%	M	109781	35,94%

Fonte: Elaborado pela autora (2019).

Os resultados para os meses de abril a junho de 2018 estão abaixo:

TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (abr-jun2018)

abr/18			mai/18			jun/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
sexo			sexo			sexo		
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
F	210858	64,06%	F	213981	63,08%	F	199757	63,20%
M	118295	35,94%	M	125225	36,92%	M	116299	36,80%

Fonte: Elaborado pela autora (2019).

Entre julho a setembro de 2018, a maior frequência refere-se ao sexo feminino em agosto:

TABELA 4 - FREQUÊNCIA ATRIBUTO 'sexo' (jul-set/2018)

jul/18			ago/18			set/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
sexo			sexo			sexo		
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
F	188832	63,25%	F	214638	63,85%	F	49	62,03%
M	109700	36,75%	M	121523	36,15%	M	30	37,97%

Fonte: Elaborado pela autora (2019).

Conforme mencionado, para o atributo referente ao código do tipo da unidade, foram adotados códigos, representados pela letra inicial de cada opção: básico (B), upa (U) e siace (S). Conforme o Portal de Dados Abertos do Governo, “as Unidades Básicas de Saúde (UBS) são a porta de entrada preferencial do Sistema Único de Saúde (SUS). O objetivo desses postos é atender até 80% dos problemas de saúde da população, sem que haja a necessidade de encaminhamento para hospitais⁹”.

TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (jan-mar/2018)

jan/18			fev/18			mar/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
codTipoUnid			codTipoUnid			codTipoUnid		
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
B	178425	66,80%	B	172634	67,82%	B	200123	65,51%
U	79518	29,77%	U	71328	28,02%	U	92570	30,30%
S	9150	3,43%	S	10597	4,16%	S	12789	4,19%

Fonte: Elaborado pela autora (2019).

⁹ Fonte: <http://dados.gov.br/dataset/unidades-basicas-de-saude-ubs>

As Unidades de Pronto Atendimento (UPAs), segunda a Secretaria Municipal de Saúde¹⁰ são responsáveis pelo atendimento de casos de urgência e emergência médica vinte e quatro (24) horas.

TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
codTipoUnid	Frequência Absoluta	Frequência Relativa	codTipoUnid	Frequência Absoluta	Frequência Relativa	codTipoUnid	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
B	215827	65,57%	B	218312	64,36%	B	200165	63,33%
U	99559	30,25%	U	106165	31,30%	U	101163	32,01%
S	13767	4,18%	S	14729	4,34%	S	14728	4,66%

Fonte: Elaborado pela autora (2019).

Por fim, há o Sistema Integrado de Atendimento de Consultas Especializadas (SIACE); com 65,46% as Unidades Básicas apresentam a maior média de frequência relativa; exceto no mês de setembro em que os dados se referem apenas ao UPA.

TABELA 5 - FREQUÊNCIA ATRIBUTO 'codTipoUnid' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
codTipoUnid	Frequência Absoluta	Frequência Relativa	codTipoUnid	Frequência Absoluta	Frequência Relativa	codTipoUnid	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
B	189116	63,35%	B	224984	66,93%	U	79	100,00%
U	93160	31,21%	U	91077	27,09%			
S	16256	5,45%	S	20100	5,98%			

Fonte: Elaborado pela autora (2019).

O atributo código da unidade representa a identificação da unidade em que o paciente foi atendido, é possível verificar a qual unidade refere-se consultando-se, além da própria base, no Cadastro Nacional dos Estabelecimentos de Saúde - CNES. Para a visualização dos dados, foram separados apenas os dez (10) primeiros valores de frequência absoluta e relativa nos casos dos atributos: codUnid, codProced, codCBO, areaAtuacao, estabSolicitante, estabDestino, capIntCID, casoDoenca, bairro e capCID.

¹⁰Fonte: <http://www.saude.curitiba.pr.gov.br/noticias/23-noticias-principal/314-upa-da-matriz-comeca-a-funcionar-dentro-do-hospital-de-clinicas.html>

TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
2639556	14462	5,41%	2639556	12676	4,98%	2639556	16726	5,48%
5323436	10844	4,06%	5323436	10034	3,94%	5323436	12661	4,14%
2438917	10059	3,77%	2639467	9295	3,65%	2639467	12266	4,02%
2639467	10012	3,75%	2438917	8606	3,38%	2438917	11585	3,79%
17396	8893	3,33%	17396	7862	3,09%	17396	9940	3,25%
9214097	8538	3,20%	9214097	7771	3,05%	9214097	9919	3,25%
2639548	8460	3,17%	2639548	7699	3,02%	2639548	9805	3,21%
3827836	8250	3,09%	3827836	7385	2,90%	3827836	9668	3,16%
2639572	4894	1,83%	2639572	4323	1,70%	2639572	5033	1,65%
16659	2966	1,11%	3270785	3695	1,45%	3270785	4359	1,43%

Fonte: Elaborado pela autora (2019).

Para o código da unidade de abril a junho de 2018 a maior frequência é de maio:

TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
2639556	18186	5,53%	2639556	19314	5,69%	2639556	17330	5,48%
5323436	13702	4,16%	5323436	14256	4,20%	5323436	13426	4,25%
2639467	12595	3,83%	2438917	14128	4,17%	2438917	12977	4,11%
2438917	12134	3,69%	2639467	13299	3,92%	2639467	12871	4,07%
9214097	11201	3,40%	9214097	12092	3,56%	9214097	12145	3,84%
2639548	10676	3,24%	2639548	11260	3,32%	17396	10945	3,46%
17396	10634	3,23%	17396	10914	3,22%	2639548	10865	3,44%
3827836	10431	3,17%	3827836	10902	3,21%	3827836	10604	3,36%
2639572	5315	1,61%	3270785	5050	1,49%	2639572	4601	1,46%
3270785	4706	1,43%	2639572	5010	1,48%	3270785	3851	1,22%

Fonte: Elaborado pela autora (2019).

Para julho e setembro, os resultados estão abaixo:

TABELA 6 - FREQUÊNCIA ATRIBUTO 'codUnid' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa	codUnid	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
2639556	16540	5,54%	2639556	15303	4,55%	5323495	23	29,11%
5323436	13556	4,54%	5323436	13168	3,92%	5323436	20	25,32%
2639467	11675	3,91%	2438917	10603	3,15%	17396	15	18,99%
2438917	11566	3,87%	9214097	10427	3,10%	2639548	6	7,59%
9214097	10769	3,61%	2639467	10343	3,08%	2639556	6	7,59%
17396	9850	3,30%	17396	9113	2,71%	9214097	6	7,59%
3827836	9764	3,27%	3827836	8669	2,58%	2639467	3	3,80%
2639548	9440	3,16%	2639548	8627	2,57%			
6388671	5141	1,72%	6388671	5427	1,61%			
2639572	5051	1,69%	3270785	5034	1,50%			

Fonte: Elaborado pela autora (2019).

De janeiro a agosto, a unidade com os maiores valores é o UPA do Sítio Cercado, em segundo lugar é o UPA Cajuru.

Já o código do procedimento 301010064 é o predominante até o mês de agosto e conforme a sua descrição significa a consulta médica em atenção básica; em setembro é o 301060096 ou atendimento médico em Unidade de Pronto Atendimento.

TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
301010064	159319	59,65%	301010064	151958	59,69%	301010064	175344	57,40%
301060096	60533	22,66%	301060096	54200	21,29%	301060096	72531	23,74%
301060029	18985	7,11%	301060029	17128	6,73%	301060029	20039	6,56%
301010110	9454	3,54%	301010072	10255	4,03%	301010072	11756	3,85%
301010072	9061	3,39%	301010110	9072	3,56%	301010110	9901	3,24%
301010080	4624	1,73%	301010080	5861	2,30%	301010080	7259	2,38%
301019940	1265	0,47%	301019940	1357	0,53%	301019940	2392	0,78%
301010137	1047	0,39%	301019860	1337	0,53%	301019860	2020	0,66%
301010129	965	0,36%	301010137	1101	0,43%	301010137	1370	0,45%
301060037	696	0,26%	301060037	838	0,33%	301060037	1189	0,39%

Fonte: Elaborado pela autora (2019).

Entre abril e junho a maior frequência está em abril:

TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (abr-jun/2018)

abr/18			mai/18			jun/18		
Atributo/Mês			Atributo/Mês			Atributo/Mês		
codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
301010064	190133	57,76%	301010064	192656	56,80%	301010064	175608	55,56%
301060096	79451	24,14%	301060096	85749	25,28%	301060096	81208	25,69%
301060029	20108	6,11%	301060029	20416	6,02%	301060029	19955	6,31%
301010072	12568	3,82%	301010072	13191	3,89%	301010072	13689	4,33%
301010110	9857	2,99%	301010110	9477	2,79%	301010110	9211	2,91%
301010080	7852	2,39%	301010080	7845	2,31%	301010080	7485	2,37%
301019940	2554	0,78%	301019860	2884	0,85%	301019860	2868	0,91%
301019860	2278	0,69%	301019940	2813	0,83%	301019940	2125	0,67%
301010137	1472	0,45%	301010137	1462	0,43%	301010137	1357	0,43%
301060037	1050	0,32%	301060037	1114	0,33%	301060037	1100	0,35%

Fonte: Elaborado pela autora (2019).

Entre julho e setembro de 2018, a maior frequência é referente ao procedimento de “Atendimento Médico em Unidade de Pronto Atendimento”:

TABELA 7 - FREQUÊNCIA ATRIBUTO 'codProced' (jul-set/2018)

jul/18			ago/18			set/18		
Atributo/Mês			Atributo/Mês			Atributo/Mês		
codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa	codProced	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
301010064	166718	55,85%	301010064	198778	59,13%	301060096	65	82,28%
301060096	72677	24,34%	301060096	69249	20,60%	301060029	14	17,72%
301060029	20483	6,86%	301060029	21828	6,49%			
301010072	14426	4,83%	301010072	17827	5,30%			
301010110	9199	3,08%	301010110	10262	3,05%			
301010080	6346	2,13%	301010080	7767	2,31%			
301019860	2489	0,83%	301019860	3333	0,99%			
301019940	2438	0,82%	301019940	3298	0,98%			
301010137	1336	0,45%	301010137	1511	0,45%			
301060037	993	0,33%	301010129	1040	0,31%			

Fonte: Elaborado pela autora (2019).

O próximo atributo relaciona-se à Classificação Brasileira de Ocupações (CBO), ou seja, à ocupação do profissional que atende o paciente.

TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
225125	114554	42,89%	225125	106333	41,77%	225125	126447	41,39%
225142	109686	41,07%	225142	103078	40,49%	225142	117723	38,54%
225124	17576	6,58%	225124	18938	7,44%	225124	28111	9,20%
225250	10298	3,86%	225250	11392	4,48%	225250	14020	4,59%
225170	7475	2,80%	225170	6303	2,48%	225170	8127	2,66%
225133	1378	0,52%	225133	1462	0,57%	225133	2058	0,67%
225112	1078	0,40%	225112	1295	0,51%	225112	1929	0,63%
225120	844	0,32%	225120	1184	0,47%	225120	1355	0,44%
225105	743	0,28%	225155	724	0,28%	225155	954	0,31%
225135	510	0,19%	225105	620	0,24%	225105	745	0,24%

Fonte: Elaborado pela autora (2019).

Em todos os meses a maior frequência é de 225125 e conforme a descrição, trata-se do médico clínico.

TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
225125	134304	40,80%	225125	140963	41,56%	225125	133463	42,23%
225142	126695	38,49%	225142	127021	37,45%	225142	117497	37,18%
225124	31832	9,67%	225124	36074	10,63%	225124	30966	9,80%
225250	15388	4,68%	225250	13372	3,94%	225250	13325	4,22%
225170	9316	2,83%	225170	9669	2,85%	225170	8589	2,72%
225133	2058	0,63%	225112	1933	0,57%	225120	2084	0,66%
225112	1933	0,59%	225120	1892	0,56%	225112	1803	0,57%
225120	1597	0,49%	225133	1731	0,51%	225133	1512	0,48%
225155	981	0,30%	225105	980	0,29%	225155	1201	0,38%
225105	932	0,28%	225155	978	0,29%	225105	705	0,22%

Fonte: Elaborado pela autora (2019).

A segunda opção com maior frequência é o código 225142 e refere-se ao médico da estratégia de saúde da família.

TABELA 8 - FREQUÊNCIA ATRIBUTO 'codCBO' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa	codCBO	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
225125	123662	41,42%	225125	136104	40,49%	225125	79	100,00%
225142	115707	38,76%	225142	128332	38,18%			
225124	25671	8,60%	225124	30613	9,11%			
225250	11071	3,71%	225250	14080	4,19%			
225170	9064	3,04%	225170	9626	2,86%			
225120	2726	0,91%	225120	3401	1,01%			
225112	1578	0,53%	225112	2685	0,80%			
225155	1539	0,52%	225155	1727	0,51%			
225133	1143	0,38%	225133	1621	0,48%			
225225	899	0,30%	225285	1004	0,30%			

Fonte: Elaborado pela autora (2019).

A TABELA 9 apresenta os resultados para a solicitação de exames, sendo a opção “não” predominante em todos os meses analisados.

TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
Nao	181911	68,11%	Nao	175060	68,77%	Nao	213851	70,00%
Sim	85182	31,89%	Sim	79499	31,23%	Sim	91631	30,00%

Fonte: Elaborado pela autora (2019).

A maior frequência está do segundo trimestre de 2018 está em junho:

TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
Nao	233741	71,01%	Nao	245432	72,35%	Nao	230184	72,83%
Sim	95412	28,99%	Sim	93774	27,65%	Sim	85872	27,17%

Fonte: Elaborado pela autora (2019).

Os resultados entre julho e setembro da frequência podem ser verificados abaixo:

TABELA 9 - FREQUÊNCIA ATRIBUTO 'solicitacaoExames' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa	solicitacaoExa mes	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
Nao	213639	71,56%	Nao	236709	70,42%	Nao	62	78,48%
Sim	84893	28,44%	Sim	99452	29,58%	Sim	17	21,52%

Fonte: Elaborado pela autora (2019).

Caso haja solicitação para que o paciente se consulte com profissional de determinada especialidade, isto será indicado no atributo de encaminhamento para atendimento especialista e em todos os meses as frequências absolutas e relativas da opção “não” foi maior, o que pode ser verificado nos resultados presentes na TABELA 10:

TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
Nao	221118	82,79%	Nao	211268	82,99%	Nao	255651	83,69%
Sim	45975	17,21%	Sim	43291	17,01%	Sim	49831	16,31%

Fonte: Elaborado pela autora (2019).

No segundo trimestre, os resultados foram:

TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
Nao	275682	83,75%	Nao	287726	84,82%	Nao	270095	85,46%
Sim	53471	16,25%	Sim	51480	15,18%	Sim	45961	14,54%

Fonte: Elaborado pela autora (2019).

Entre julho e setembro, a maior frequência manteve-se no não encaminhamento para atendimento especialista:

TABELA 10 - FREQUÊNCIA ATRIBUTO 'encamAtendEspecialista' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa	encamAtend Especialista	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
Nao	251487	84,24%	Nao	278007	82,70%	Nao	79	100,00%
Sim	47045	15,76%	Sim	58154	17,30%			

Fonte: Elaborado pela autora (2019).

A área de atuação do profissional especialista apresentou como resultados, em primeiro lugar OFMG (Oftalmologia Geral); em segundo OTIG (Otorrinolaringologia Geral); e em terceiro DERG (Dermatologia Geral), sendo que essa posição só é alterada nos meses de junho, com AVAD (Avaliação Complementar do Aparelho Digestivo) e em agosto com AVPS (Avaliação Psicologia - NASF). Destaca-se que, devido ao tamanho das tabelas e quantidade de opções disponíveis, apenas as dez (10) áreas de atuação com os maiores resultados de frequências, foram selecionadas e os códigos podem ser consultados no apêndice B.

TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa
TOTAL	26652	100,00%	TOTAL	25316	100,00%	TOTAL	29246	100,00%
OFMG	3327	12,48%	OFMG	3178	12,55%	OFMG	3782	12,93%
OTIG	2051	7,70%	DERG	1861	7,35%	OTIG	2116	7,24%
DERG	1912	7,17%	OTIG	1796	7,09%	DERG	1902	6,50%
CARG	1903	7,14%	AVCR	1176	4,65%	AVPS	1659	5,67%
CIRG	1241	4,66%	AVPS	1110	4,38%	AVCR	1555	5,32%
AVPS	1043	3,91%	CIRG	1077	4,25%	CIRG	1214	4,15%
URGG	965	3,62%	URGG	884	3,49%	URGG	958	3,28%
OREG	818	3,07%	OREG	804	3,18%	AVNE	843	2,88%
CIVG	785	2,95%	CIVG	711	2,81%	OREG	835	2,86%
CANG	735	2,76%	PROG	674	2,66%	AVPQ	748	2,56%

Fonte: Elaborado pela autora (2019).

A continuação da TABELA 11 apresenta os resultados entre abril e junho de 2018:

TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa
TOTAL	31647	100,00%	TOTAL	30278	100,00%	TOTAL	28311	100,00%
OFMG	3838	12,13%	OFMG	3588	11,85%	OFMG	3257	11,50%
OTIG	2165	6,84%	OTIG	2307	7,62%	OTIG	2024	7,15%
DERG	2060	6,51%	DERG	2075	6,85%	AVPS	1711	6,04%
AVPS	1911	6,04%	AVPS	1849	6,11%	DERG	1659	5,86%
AVCR	1666	5,26%	AVCR	1582	5,22%	AVCR	1462	5,16%
CIRG	1174	3,71%	CIRG	1119	3,70%	AVAD	1116	3,94%
URGG	916	2,89%	AVPQ	961	3,17%	CIRG	923	3,26%
AVNE	895	2,83%	URGG	921	3,04%	URGG	841	2,97%
AVPQ	868	2,74%	AVNE	898	2,97%	AVPQ	807	2,85%
CANG	801	2,53%	CANG	738	2,44%	AVNE	803	2,84%

Fonte: Elaborado pela autora (2019).

No terceiro trimestre de 2018, os resultados são os seguintes:

TABELA 11 - FREQUÊNCIA ATRIBUTO 'areaAtuacao' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa	areaAtuacao	Frequência Absoluta	Frequência Relativa
TOTAL	29329	100,00%	TOTAL	36825	100,00%	TOTAL	0	0,00%
OFMG	3139	10,70%	OFMG	4060	11,03%	-	-	-
OTIG	2158	7,36%	OTIG	2527	6,86%			
DERG	1742	5,94%	AVAD	2296	6,23%			
AVAD	1656	5,65%	DERG	2250	6,11%			
AVPS	1580	5,39%	AVPS	2110	5,73%			
AVCR	1502	5,12%	AVCR	1943	5,28%			
CIRG	970	3,31%	CIRG	1219	3,31%			
AVPQ	849	2,89%	URGG	1050	2,85%			
AVNE	844	2,88%	AVNE	955	2,59%			
URGG	840	2,86%	AVPQ	886	2,41%			

Fonte: Elaborado pela autora (2019).

Por fim, refere-se aos capítulos do CID, campo de preenchimento obrigatório nas bases.

TABELA 12 - FREQUÊNCIA ATRIBUTO 'capCID' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
Z00-Z99	99467	37,24%	Z00-Z99	96221	37,80%	Z00-Z99	111767	36,59%
R00-R99	35010	13,11%	R00-R99	31111	12,22%	R00-R99	38609	12,64%
M00-M99	18730	7,01%	J00-J99	19334	7,60%	J00-J99	32118	10,51%
J00-J99	18152	6,80%	M00-M99	17295	6,79%	M00-M99	19133	6,26%
I00-I99	14428	5,40%	I00-I99	14217	5,58%	I00-I99	15097	4,94%
N00-N99	13512	5,06%	N00-N99	12265	4,82%	N00-N99	13738	4,50%
A00-B99	11430	4,28%	F00-F99	11468	4,51%	A00-B99	13172	4,31%
F00-F99	11053	4,14%	E00-E90	10598	4,16%	F00-F99	12918	4,23%
E00-E90	10358	3,88%	A00-B99	9752	3,83%	E00-E90	11611	3,80%
K00-K93	7409	2,77%	K00-K93	7051	2,77%	K00-K93	8534	2,79%

Fonte: Elaborado pela autora (2019).

De modo geral, os destaques estão nos capítulos:

- Capítulo XXI - Fatores que influenciam o estado de saúde e o contato com os serviços de saúde (Z00-Z99);
- Capítulo XVIII - Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte (R00-R99);
- Capítulo XIII - Doenças do sistema osteomuscular e do tecido conjuntivo (M00-M99);
- Capítulo X - Doenças do aparelho respiratório (J00-J99).

TABELA 26 - FREQUÊNCIA ATRIBUTO 'capCID' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
Z00-Z99	118225	35,92%	Z00-Z99	119419	35,21%	Z00-Z99	109501	34,65%
J00-J99	44184	13,42%	J00-J99	61978	18,27%	J00-J99	64156	20,30%
R00-R99	38738	11,77%	R00-R99	38785	11,43%	R00-R99	36130	11,43%
M00-M99	19396	5,89%	M00-M99	18051	5,32%	M00-M99	15717	4,97%
I00-I99	16516	5,02%	I00-I99	16275	4,80%	I00-I99	15617	4,94%
N00-N99	14378	4,37%	F00-F99	13492	3,98%	F00-F99	12101	3,83%
F00-F99	13808	4,20%	N00-N99	13304	3,92%	N00-N99	11926	3,77%
A00-B99	13277	4,03%	E00-E90	12205	3,60%	E00-E90	11295	3,57%
E00-E90	12493	3,80%	A00-B99	10429	3,07%	A00-B99	8615	2,73%
K00-K93	8653	2,63%	K00-K93	7555	2,23%	K00-K93	6834	2,16%

Fonte: Elaborado pela autora (2019).

Em termos mais específicos, algumas das categorias doenças desses capítulos são: “cirurgia profilática (Z40)”, “resultado do parto (Z37)”; “aumento da glicemia (R73)”; “mal-estar, fadiga (R53)”; “escoliose (M41)”; “poliartrose (M15)”; “bronquite aguda (J20)”; “asma (J45)”.

TABELA 26 - FREQUÊNCIA ATRIBUTO 'capCID' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa	capCID	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
Z00-Z99	106732	35,75%	Z00-Z99	123544	36,75%	J00-J99	18	22,78%
J00-J99	45542	15,26%	J00-J99	43419	12,92%	R00-R99	16	20,25%
R00-R99	34409	11,53%	R00-R99	37196	11,06%	Z00-Z99	13	16,46%
M00-M99	18013	6,03%	M00-M99	20188	6,01%	M00-M99	9	11,39%
I00-I99	16049	5,38%	I00-I99	19315	5,75%	S00-T98	6	7,59%
N00-N99	12418	4,16%	F00-F99	15145	4,51%	A00-B99	6	7,59%
F00-F99	12043	4,03%	N00-N99	14480	4,31%	K00-K93	3	3,80%
E00-E90	11624	3,89%	E00-E90	13825	4,11%	L00-L99	2	2,53%
A00-B99	8718	2,92%	A00-B99	10334	3,07%	F00-F99	2	2,53%
K00-K93	7677	2,57%	K00-K93	8966	2,67%	N00-N99	2	2,53%

Fonte: Elaborado pela autora (2019).

A partir da TABELA 13 há a apresentação dos resultados das frequências relacionadas ao internamento do paciente, iniciando com o atributo 'desencadeouInternamento', sendo que a opção referente ao “não internamento” foi a predominante em todos os meses analisados.

TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
Nao	265090	99,25%	Nao	252707	99,27%	Nao	303296	99,28%
Sim	2003	0,75%	Sim	1852	0,73%	Sim	2186	0,72%

Fonte: Elaborado pela autora (2019).

De abril a junho de 2018:

TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
Nao	327017	99,35%	Nao	336993	99,35%	Nao	313804	99,29%
Sim	2136	0,65%	Sim	2213	0,65%	Sim	2252	0,71%

Fonte: Elaborado pela autora (2019).

A terceira parte da TABELA 13 apresenta os resultados de julho a setembro de 2018:

TABELA 13 - FREQUÊNCIA ATRIBUTO 'desencadeouInternamento' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa	desencadeouInternamento	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
Nao	296489	99,32%	Nao	334114	99,39%	Nao	79	100,00%
Sim	2043	0,68%	Sim	2047	0,61%			

Fonte: Elaborado pela autora (2019).

A data de internamento é complementar ao atributo anterior e representa o mês em que o paciente foi internado; de modo geral, a maior frequência ocorreu no mês de agosto com 98,13% internamentos. Em setembro, não houve dados referentes à esta data.

TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtInternamento' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
dtInternamento	Frequência Absoluta	Frequência Relativa	dtInternamento	Frequência Absoluta	Frequência Relativa	dtInternamento	Frequência Absoluta	Frequência Relativa
TOTAL	1961	100,00%	TOTAL	1810	100,00%	TOTAL	2160	100,00%
jan	1809	92,25%	fev	1650	91,16%	mar	2019	93,47%
fev	138	7,04%	mar	146	8,07%	abr	131	6,06%
mar	9	0,46%	abr	7	0,39%	mai	6	0,28%
mai	4	0,20%	mai	5	0,28%	jun	4	0,19%
abr	1	0,05%	jun	2	0,11%			

Fonte: Elaborado pela autora (2019).

De abril a junho os seguintes resultados foram encontrados:

TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtInternamento' (abr-jun/2018)

abr/18			mai/18			jun/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
TOTAL	2120	100,00%	TOTAL	2179	100,00%	TOTAL	2228	100,00%
abr	1999	94,29%	mai	2057	94,40%	jun	2123	95,29%
mai	113	5,33%	jun	122	5,60%	jul	105	4,71%
jun	8	0,38%						

Fonte: Elaborado pela autora (2019).

Por fim, entre julho e setembro, a maior frequência de internamentos ocorreu em agosto:

TABELA 14 - FREQUÊNCIA ATRIBUTO 'dtInternamento' (jul-set/2018)

jul/18			ago/18			set/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
TOTAL	2005	100,00%	TOTAL	1983	100,00%	TOTAL	0	0,00%
jul	1932	96,36%	ago	1946	98,13%	-	-	-
ago	73	3,64%	set	37	1,87%			

Fonte: Elaborado pela autora (2019).

Ainda relacionado ao internamento, as tabelas 15 e 16 apresentam os estabelecimentos solicitantes e de destino do paciente. Para a primeira atributo, as unidades com os maiores valores de frequência foram UPA Fazendinha, UPA Boa Vista, UPA Cajuru e UPA Boqueirão.

TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (jan-mar/2018)

jan/18			fev/18			mar/18		
Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa	Atributo/Mês	Frequência Absoluta	Frequência Relativa
TOTAL	2003	100,00%	TOTAL	1852	100,00%	TOTAL	2186	100,00%
UPAFAZEN	347	17,32%	UPABOAVI	291	15,71%	UPAFAZEN	376	17,20%
UPABOAVI	304	15,18%	UPAFAZEN	287	15,50%	UPABOAVI	315	14,41%
UPACAJUR	246	12,28%	UPACAJUR	237	12,80%	UPABOQUE	257	11,76%
UPAPINHE	216	10,78%	UPAPINHE	198	10,69%	UPACAJUR	244	11,16%
UPABOQUE	201	10,03%	UPABOQUE	197	10,64%	UPACAMCO	238	10,89%
UPASITCE	199	9,94%	UPACAMCO	193	10,42%	UPAPINHE	232	10,61%
UPACAMCO	196	9,79%	UPASITCE	164	8,86%	UPASITCE	217	9,93%
UPATATUQ	142	7,09%	UPATATUQ	134	7,24%	UPATATUQ	138	6,31%
UMSTRIPS	9	0,45%	UMSOSTER	12	0,65%	UMSOSTER	14	0,64%
UMSIRMTA	8	0,40%	UMSCAJUR	8	0,43%	UMSTRIPS	13	0,59%

Fonte: Elaborado pela autora (2019).

Em abril, o estabelecimento com maior frequência de solicitação foi a UPA Boa Vista, enquanto em maio e junho foi a UPA Fazendinha:

TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
estabSolicitante	Frequência Absoluta	Frequência Relativa	estabSolicitante	Frequência Absoluta	Frequência Relativa	estabSolicitante	Frequência Absoluta	Frequência Relativa
TOTAL	2136	100,00%	TOTAL	2213	100,00%	TOTAL	2252	100,00%
UPABOAVI	338	15,82%	UPAFAZEN	327	14,78%	UPAFAZEN	426	18,92%
UPAFAZEN	299	14,00%	UPABOAVI	322	14,55%	UPACAJUR	313	13,90%
UPABOQUE	263	12,31%	UPABOQUE	272	12,29%	UPABOAVI	280	12,43%
UPASITCE	230	10,77%	UPACAJUR	268	12,11%	UPABOQUE	266	11,81%
UPACAJUR	229	10,72%	UPACAMCO	265	11,97%	UPACAMCO	225	9,99%
UPAPINHE	223	10,44%	UPASITCE	240	10,85%	UPASITCE	212	9,41%
UPACAMCO	220	10,30%	UPAPINHE	214	9,67%	UPAPINHE	211	9,37%
UPATATUQ	162	7,58%	UPATATUQ	143	6,46%	UPATATUQ	141	6,26%
UMSCAJUR	11	0,51%	UMSCAPAN	10	0,45%	UMSCAJUR	14	0,62%
UMSOSTER	8	0,37%	UMSOUPVA	10	0,45%	UMSOSTER	8	0,36%

Fonte: Elaborado pela autora (2019).

Em julho e agosto o estabelecimento com o maior número de solicitações foi a UPA Fazendinha; em setembro, não houve requisições:

TABELA 15 - FREQUÊNCIA ATRIBUTO 'estabSolicitante' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
estabSolicitante	Frequência Absoluta	Frequência Relativa	estabSolicitante	Frequência Absoluta	Frequência Relativa	estabSolicitante	Frequência Absoluta	Frequência Relativa
TOTAL	2043	100,00%	TOTAL	2047	100,00%	TOTAL	0	0,00%
UPAFAZEN	391	19,14%	UPAFAZEN	306	14,95%	-	-	-
UPABOAVI	263	12,87%	UPACAJUR	286	13,97%			
UPACAJUR	246	12,04%	UPABOAVI	285	13,92%			
UPABOQUE	230	11,26%	UPABOQUE	242	11,82%			
UPAPINHE	208	10,18%	UPAPINHE	221	10,80%			
UPACAMCO	202	9,89%	UPASITCE	182	8,89%			
UPASITCE	186	9,10%	UPACAMCO	166	8,11%			
UPATATUQ	145	7,10%	UPATATUQ	141	6,89%			
UMSOSTER	16	0,78%	UPACIDID	38	1,86%			
UMSRIOBO	11	0,54%	UMSOSTER	11	0,54%			

Fonte: Elaborado pela autora (2019).

Já para os dez (10) estabelecimentos de destino selecionados, a opção com maior frequência absoluta e relativa é HOSPNSRO (Hospital Nossa Senhora do Rocio - Campo Largo) em todos os meses analisados e com dados. Nos meses de janeiro, março, abril, junho, julho e agosto, a opção é CLIDRHRO (Clínica Dr. Helio Rotenberg), enquanto em fevereiro e março é INSTMASO (IMS ou Instituto Madalena Sofia); demais códigos estão disponíveis no apêndice B.

TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa
TOTAL	2003	100,00%	TOTAL	1852	100,00%	TOTAL	2186	100,00%
HOSPNSRO	988	49,33%	HOSPNSRO	926	50,00%	HOSPNSRO	1088	49,77%
CLIDRHRO	189	9,44%	INSTMASO	177	9,56%	CLIDRHRO	203	9,29%
HOSPCLIN	160	7,99%	CLIDRHRO	163	8,80%	INSTMASO	190	8,69%
INSTMASO	150	7,49%	HOSPCLIN	109	5,89%	HOSPCLIN	182	8,33%
HOSPIDZA	112	5,59%	HOSPUNEC	88	4,75%	HOSPUNEC	95	4,35%
HOSPUNEC	73	3,64%	HOSPIDZA	83	4,48%	HOSPIDZA	86	3,93%
HOSPSCCU	54	2,70%	HOSANTCC	45	2,43%	HOSPMEDE	49	2,24%
CRVBFEP	35	1,75%	HOSPIPPR	35	1,89%	HOSPEPBR	37	1,69%
HOSPIPPR	34	1,70%	CRVBFEP	35	1,89%	HOSANTCC	37	1,69%
HOSPUNCA	33	1,65%	HOSPTRAB	33	1,78%	HOSPIPPR	35	1,60%

Fonte: Elaborado pela autora (2019).

No segundo trimestre os resultados foram:

TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa
TOTAL	2136	100,00%	TOTAL	2213	100,00%	TOTAL	2252	100,00%
HOSPNSRO	997	46,68%	HOSPNSRO	940	42,48%	HOSPNSRO	1070	47,51%
CLIDRHRO	211	9,88%	INSTMASO	208	9,40%	CLIDRHRO	206	9,15%
INSTMASO	191	8,94%	CLIDRHRO	205	9,26%	HOSPSMSF	165	7,33%
HOSPCLIN	144	6,74%	HOSPUNEC	172	7,77%	HOSPCLIN	146	6,48%
HOSPUNEC	139	6,51%	HOSPCLIN	143	6,46%	HOSPUNEC	142	6,31%
HOSPVCIC	74	3,46%	HOSPIDZA	91	4,11%	HOSPIDZA	111	4,93%
HOSPIDZA	68	3,18%	HOSPVCIC	87	3,93%	HOSPVCIC	69	3,06%
HOSPMEDE	56	2,62%	HOSPMEDE	71	3,21%	HOSPMEDE	65	2,89%
HOSANTCC	44	2,06%	HOSANTCC	42	1,90%	FUNHOPIN	47	2,09%
CRVBFEP	36	1,69%	HOSPWALM	40	1,81%	HOSPWALM	39	1,73%

Fonte: Elaborado pela autora (2019).

Enquanto no terceiro trimestre:

TABELA 16 - FREQUÊNCIA ATRIBUTO 'estabDestino' (jul-set/2018)

Atributo/Mês		jul/18		Atributo/Mês		ago/18		Atributo/Mês		set/18	
estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa	estabDestino	Frequência Absoluta	Frequência Relativa			
TOTAL	2043	100,00%	TOTAL	2047	100,00%	TOTAL	0	0,00%			
HOSPNSRO	845	41,36%	HOSPNSRO	800	39,08%	-	-	-			
CLIDRHRO	208	10,18%	CLIDRHRO	233	11,38%						
HOSPSMSF	179	8,76%	INSTMASO	199	9,72%						
HOSPCLIN	166	8,13%	HOSPIDZA	123	6,01%						
HOSPUNEC	131	6,41%	HOSPCLIN	113	5,52%						
HOSPIDZA	123	6,02%	HOSPUNEC	113	5,52%						
HOSPVVIC	63	3,08%	HOSPVVIC	73	3,57%						
HOSPMEDE	59	2,89%	HOSPMEDE	73	3,57%						
HOSANTCC	46	2,25%	HOSANTCC	60	2,93%						
HOSPEPBR	41	2,01%	HOSPIPPR	44	2,15%						

Fonte: Elaborado pela autora (2019).

Para situações de internamento, também há um CID associado e da mesma forma que o atributo do CID de atendimento, este foi definido por capítulos. O “Capítulo XVIII - Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte” corresponde à R00-R99, sendo a maior frequência absoluta e relativa.

Os outros maiores resultados das frequências absolutas e relativas foram: “Capítulo X - Doenças do aparelho respiratório” ou J00-J99; “Capítulo XI - Doenças do aparelho digestivo” ou K00-k93; “Capítulo IX - Doenças do aparelho circulatório” ou I00-I99; e “Capítulo V - Transtornos mentais e comportamentais” ou F00-F99.

TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (jan-mar/2018)

Atributo/Mês		jan/18		Atributo/Mês		fev/18		Atributo/Mês		mar/18	
capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa			
TOTAL	2003	100,00%	TOTAL	1852	100,00%	TOTAL	2186	100,00%			
R00-R99	577	28,81%	R00-R99	596	32,18%	R00-R99	631	28,87%			
I00-I99	283	14,13%	I00-I99	228	12,31%	I00-I99	272	12,44%			
K00-K93	243	12,13%	K00-K93	216	11,66%	K00-K93	248	11,34%			
F00-F99	221	11,03%	F00-F99	194	10,48%	F00-F99	246	11,25%			
N00-N99	170	8,49%	N00-N99	183	9,88%	N00-N99	210	9,61%			
J00-J99	166	8,29%	J00-J99	127	6,86%	J00-J99	195	8,92%			
L00-L99	100	4,99%	L00-L99	78	4,21%	L00-L99	89	4,07%			
A00-B99	54	2,70%	A00-B99	61	3,29%	A00-B99	74	3,39%			
D50-D89	37	1,85%	S00-T98	39	2,11%	E00-E90	38	1,74%			
S00-T98	36	1,80%	E00-E90	33	1,78%	S00-T98	36	1,65%			

Fonte: Elaborado pela autora (2019).

Nos meses de abril a junho, o resultado foi semelhante:

TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa
TOTAL	2136	100,00%	TOTAL	2213	100,00%	TOTAL	2252	100,00%
R00-R99	684	32,02%	R00-R99	596	26,93%	R00-R99	624	27,71%
J00-J99	268	12,55%	J00-J99	365	16,49%	J00-J99	492	21,85%
I00-I99	264	12,36%	I00-I99	262	11,84%	I00-I99	259	11,50%
F00-F99	234	10,96%	F00-F99	243	10,98%	F00-F99	247	10,97%
K00-K93	216	10,11%	K00-K93	236	10,66%	K00-K93	182	8,08%
N00-N99	148	6,93%	N00-N99	182	8,22%	N00-N99	179	7,95%
L00-L99	79	3,70%	L00-L99	75	3,39%	L00-L99	66	2,93%
A00-B99	74	3,46%	A00-B99	62	2,80%	A00-B99	50	2,22%
E00-E90	34	1,59%	D50-D89	35	1,58%	E00-E90	41	1,82%
D50-D89	24	1,12%	M00-M99	32	1,45%	D50-D89	23	1,02%

Fonte: Elaborado pela autora (2019).

No terceiro trimestre, não houve resultados para análise em setembro:

TABELA 17 - FREQUÊNCIA ATRIBUTO 'capIntCID' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa	capIntCID	Frequência Absoluta	Frequência Relativa
TOTAL	2043	100,00%	TOTAL	2047	100,00%	TOTAL	0	0,00%
R00-R99	603	29,52%	R00-R99	582	28,43%	-	-	-
J00-J99	321	15,71%	J00-J99	290	14,17%			
F00-F99	252	12,33%	F00-F99	265	12,95%			
I00-I99	225	11,01%	I00-I99	241	11,77%			
K00-K93	199	9,74%	N00-N99	222	10,85%			
N00-N99	140	6,85%	K00-K93	180	8,79%			
L00-L99	80	3,92%	L00-L99	62	3,03%			
A00-B99	39	1,91%	E00-E90	42	2,05%			
E00-E90	36	1,76%	M00-M99	35	1,71%			
M00-M99	34	1,66%	A00-B99	33	1,61%			

Fonte: Elaborado pela autora (2019).

A partir da TABELA 18 até a 26 temos os campos obrigatórios em caso de cadastro definitivo do paciente, relacionados aos aspectos sócio econômicos, e exceto no atributo de “em caso de doença”, os demais apresentam opções com códigos alfabéticos de A à H ou “sim” ou “não”.

O tipo de tratamento da água no domicílio do paciente tem como valores “sem tratamento”, seguido por “filtração” ou “cloração”. É importante destacar que estes resultados são corroborados pela notícia da Agência Nacional de Águas (ANA) em 2019 que revela uma realidade em que “o Brasil ainda precisa

avançar muito para oferecer à totalidade da sua população um serviço sanitário adequado, já que somente cerca de 3 mil municípios brasileiros, o equivalente a 55,2% do total, contam com coleta e tratamento de esgoto”, ainda expõem que em 2007 foi criada a Lei nº 11.445 responsável por estabelecer diretrizes nacionais para o saneamento básico no país, sendo que o art. 52 estabelece a criação do Plano Nacional de Saneamento Básico (PLANSAB)¹¹, o que demonstra a preocupação com este aspecto na vida dos brasileiros.

Especificamente no município analisado, a Agência de Notícias do Paraná (AEN), publicou em 2019 que “na véspera de completar seu 326º aniversário, Curitiba foi a única capital do Brasil a ser premiada nesta quinta-feira (28) com a mais alta pontuação do II Ranking Abes da Universalização do Saneamento[...] O ranking é feito com base em cinco critérios: abastecimento de água, coleta de esgoto, tratamento de esgoto, coleta de resíduos sólidos e destinação adequada de resíduos sólidos.¹²”, sendo ABES a Agência Brasileira de Engenharia Sanitária; destaca-se que este é apenas um resultado relacionado ao Saneamento e que as frequências na base, representadas por consultas em determinadas unidades e regiões revelam um cenário mais alinhado ao restante do país.

TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa
TOTAL	241258	100,00%	TOTAL	231532	100,00%	TOTAL	278286	100,00%
D	133469	55,32%	D	127464	55,05%	D	153837	55,28%
C	54010	22,39%	C	52013	22,46%	C	62558	22,48%
A	52088	21,59%	A	50547	21,83%	A	60184	21,63%
B	1691	0,70%	B	1508	0,65%	B	1707	0,61%

Fonte: Elaborado pela autora (2019).

Entre abril e junho, os resultados foram semelhantes, com a opção D apresentando maior frequência:

¹¹ Fonte: <https://www.ana.gov.br/noticias-antigas/plano-nacional-de-saneamento-ba-sico-o-que-falta.2019-03-15.4665581007>

¹² Fonte: <http://www.aen.pr.gov.br/modules/noticias/article.php?storyid=101643>

TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa
TOTAL	300338	100,00%	TOTAL	309159	100,00%	TOTAL	287118	100,00%
D	165839	55,22%	D	170721	55,22%	D	158847	55,32%
C	67522	22,48%	C	69315	22,42%	C	63496	22,11%
A	65068	21,66%	A	67121	21,71%	A	62850	21,89%
B	1909	0,64%	B	2002	0,65%	B	1925	0,67%

Fonte: Elaborado pela autora (2019).

No terceiro trimestre de 2018, os resultados foram os seguintes:

TABELA 18 - FREQUÊNCIA ATRIBUTO 'tratamentoDomicilio' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa	tratamentoDomicilio	Frequência Absoluta	Frequência Relativa
TOTAL	270746	100,00%	TOTAL	306599	100,00%	TOTAL	62	100,00%
D	149877	55,36%	D	168905	55,09%	D	44	70,97%
C	60622	22,39%	C	68936	22,48%	A	12	19,35%
A	58477	21,60%	A	66852	21,80%	C	5	8,06%
B	1770	0,65%	B	1906	0,62%	B	1	1,61%

Fonte: Elaborado pela autora (2019).

Já o abastecimento de água tem como maior frequência a rede pública em todos os meses analisados em Curitiba. Conforme o Plano Municipal de Abastecimento de Água (2017, p.24) “o sistema de abastecimento de água do município de Curitiba é denominado Sistema de Abastecimento Integrado de Curitiba (SAIC), que atende, além da capital, também os municípios de: São José dos Pinhais, Fazenda Rio Grande, Araucária, Campo Magro, Almirante Tamandaré, Colombo, Campina Grande do Sul, Quatro Barras, Pinhais e Piraquara”. No Plano Diretor SAIC: Sistema de Abastecimento de Água Integrado de Curitiba e Região Metropolitana (2013) citado pelo referido Plano Municipal (2017), Curitiba usa em média 66% do volume de água produzido pelo SAIC.

TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa
TOTAL	241275	100,00%	TOTAL	231553	100,00%	TOTAL	278298	100,00%
E	237365	98,38%	E	227959	98,45%	E	273977	98,45%
C	3301	1,37%	C	3100	1,34%	C	3721	1,34%
D	528	0,22%	D	431	0,19%	D	542	0,19%
B	45	0,02%	B	38	0,02%	A	29	0,01%
A	36	0,01%	A	25	0,01%	B	29	0,01%

Fonte: Elaborado pela autora (2019).

No segundo trimestre os resultados foram semelhantes quanto ao tipo de abastecimento de água:

TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa
TOTAL	300359	100,00%	TOTAL	309199	100,00%	TOTAL	287147	100,00%
E	295673	98,44%	E	304390	98,44%	E	282663	98,44%
C	4010	1,34%	C	4137	1,34%	C	3823	1,33%
D	616	0,21%	D	591	0,19%	D	591	0,21%
B	31	0,01%	A	48	0,02%	B	40	0,01%
A	29	0,01%	B	33	0,01%	A	30	0,01%

Fonte: Elaborado pela autora (2019).

Por fim, os resultados de julho a setembro:

TABELA 19 - FREQUÊNCIA ATRIBUTO 'abastecimento' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa	abastecimento	Frequência Absoluta	Frequência Relativa
TOTAL	270777	100,00%	TOTAL	306606	100,00%	TOTAL	62	100,00%
E	266612	98,46%	E	301929	98,47%	E	61	98,39%
C	3544	1,31%	C	3918	1,28%	C	1	1,61%
D	563	0,21%	D	670	0,22%			
A	32	0,01%	A	45	0,01%			
B	26	0,01%	B	44	0,01%			

Fonte: Elaborado pela autora (2019).

Em Curitiba, a energia elétrica é fornecida pela Companhia Paranaense de Energia Elétrica (COPEL) e em todos os meses analisados, a opção que revela a presença da energia ("sim") prevaleceu.

TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
Sim	250211	93,68%	Sim	239125	93,94%	Sim	287261	94,04%
Nao	16882	6,32%	Nao	15434	6,06%	Nao	18221	5,96%

Fonte: Elaborado pela autora (2019).

Entre abril e junho, a opção “Sim” continuou predominante:

TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
Sim	309078	93,90%	Sim	318661	93,94%	Sim	297029	93,98%
Nao	20075	6,10%	Nao	20545	6,06%	Nao	19027	6,02%

Fonte: Elaborado pela autora (2019).

A terceira parte dos meses analisados quando à presença de energia elétrica obteve os seguintes resultados:

TABELA 20 - FREQUÊNCIA ATRIBUTO 'energiaEletrica' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa	energiaEletrica	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
Sim	279926	93,77%	Sim	315836	93,95%	Sim	65	82,28%
Nao	18606	6,23%	Nao	20325	6,05%	Nao	14	17,72%

Fonte: Elaborado pela autora (2019).

Quanto ao tipo de habitação do paciente, nos meses analisados os valores de frequência absoluta e relativa apontaram para as opções “Tijolo/Alvenaria com revestimento”; “Madeira Aparelhada”; “Tijolo/Alvenaria sem revestimento”. Apenas no mês de setembro essa configuração mudou e a frequência da opção de “Outro Material” ficou em segundo lugar.

TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa
TOTAL	241279	100,00%	TOTAL	231550	100,00%	TOTAL	278301	100,00%
H	210317	87,17%	H	202776	87,57%	H	243448	87,48%
A	22424	9,29%	A	20664	8,92%	A	25187	9,05%
G	3730	1,55%	G	3588	1,55%	G	4401	1,58%
C	3415	1,42%	C	3291	1,42%	C	3712	1,33%
B	717	0,30%	B	634	0,27%	B	737	0,26%
E	514	0,21%	E	443	0,19%	E	604	0,22%
F	103	0,04%	F	122	0,05%	F	152	0,05%
D	59	0,02%	D	32	0,01%	D	60	0,02%

Fonte: Elaborado pela autora (2019).

Os resultados entre abril e junho foram:

TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa
TOTAL	300361	100,00%	TOTAL	309196	100,00%	TOTAL	287148	100,00%
H	262036	87,24%	H	269841	87,27%	H	250945	87,39%
A	27688	9,22%	A	28285	9,15%	A	25774	8,98%
G	4869	1,62%	G	4945	1,60%	G	4695	1,64%
C	4086	1,36%	C	4379	1,42%	C	4111	1,43%
B	876	0,29%	B	807	0,26%	B	793	0,28%
E	602	0,20%	E	699	0,23%	E	646	0,22%
F	145	0,05%	F	154	0,05%	F	126	0,04%
D	59	0,02%	D	86	0,03%	D	58	0,02%

Fonte: Elaborado pela autora (2019).

Por fim, o terceiro trimestre de 2018 apresentou as seguintes frequências:

TABELA 21 - FREQUÊNCIA ATRIBUTO 'tipoHabitacao' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa	tipoHabitacao	Frequência Absoluta	Frequência Relativa
TOTAL	270779	100,00%	TOTAL	306616	100,00%	TOTAL	62	100,00%
H	236729	87,43%	H	268211	87,47%	H	50	80,65%
A	24470	9,04%	A	27094	8,84%	A	7	11,29%
G	4413	1,63%	G	5317	1,73%	C	3	4,84%
C	3671	1,36%	C	4276	1,39%	G	2	3,23%
B	801	0,30%	B	891	0,29%			
E	526	0,19%	E	640	0,21%			
F	118	0,04%	F	112	0,04%			
D	51	0,02%	D	75	0,02%			

Fonte: Elaborado pela autora (2019).

O atributo na TABELA 21 apresenta os resultados referente ao destino do lixo na residência do paciente, sendo a opção com média de frequência relativa

de 98,85%, lixo coletado. Em Curitiba, a coleta é realizada pela Prefeitura e “Sua operação conta com cerca de 2 mil funcionários que fazem a coleta do lixo de 100% dos domicílios da cidade. Todos os meses, são coletadas cerca de 50 mil toneladas de resíduos em todos os serviços de limpeza urbana (ou 41.500 toneladas na Coleta Domiciliar)¹³”

TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa
TOTAL	241276	100,00%	TOTAL	231554	100,00%	TOTAL	278299	100,00%
B	238489	98,84%	B	228984	98,89%	B	275282	98,92%
C	2018	0,84%	C	1860	0,80%	C	2246	0,81%
D	452	0,19%	D	403	0,17%	D	412	0,15%
A	317	0,13%	A	307	0,13%	A	359	0,13%

Fonte: Elaborado pela autora (2019).

Entre abril e junho, os resultados foram os seguintes:

TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa
TOTAL	300360	100,00%	TOTAL	309200	100,00%	TOTAL	287147	100,00%
B	297178	98,94%	B	305916	98,94%	B	284098	98,94%
C	2339	0,78%	C	2439	0,79%	C	2292	0,80%
D	487	0,16%	D	477	0,15%	D	435	0,15%
A	356	0,12%	A	368	0,12%	A	322	0,11%

Fonte: Elaborado pela autora (2019).

No terceiro trimestre, a opção B continuou com as maiores frequências:

TABELA 22 - FREQUÊNCIA ATRIBUTO 'destinoLixo' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa	destinoLixo	Frequência Absoluta	Frequência Relativa
TOTAL	270778	100,00%	TOTAL	306614	100,00%	TOTAL	62	100,00%
B	267839	98,91%	B	303266	98,91%	B	61	98,39%
C	2089	0,77%	C	2413	0,79%	A	1	1,61%
D	465	0,17%	D	528	0,17%			
A	385	0,14%	A	407	0,13%			

Fonte: Elaborado pela autora (2019).

¹³ Fonte: <https://coletalixo.curitiba.pr.gov.br/>

A TABELA 23 traz os resultados das frequências para o atributo de destino das fezes/urina no domicílio do paciente; de janeiro a setembro, as opções ficaram na ordem de “sistema de esgoto” (média 92,34%), “fossa séptica” (média 5,33%) e “céu aberto” (1,25%), destaca-se que a presença, ainda que mínima do descarte destes resíduos a céu aberto representa um alerta, pois é uma questão de saúde pública e responsabilidade ambiental, conforme explicado pela Companhia de Saneamento Paranaense (SANEPAR) “a falta de tratamento do esgoto e de condições adequadas de saneamento contribuem para a proliferação de doenças parasitárias e infecciosas¹⁴”.

TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa
TOTAL	241277	100,00%	TOTAL	231550	100,00%	TOTAL	278300	100,00%
F	222475	92,21%	F	214063	92,45%	F	256811	92,28%
D	12831	5,32%	D	11803	5,10%	D	14638	5,26%
A	2935	1,22%	A	2812	1,21%	A	3398	1,22%
E	2844	1,18%	E	2714	1,17%	E	3268	1,17%
C	110	0,05%	C	102	0,04%	C	100	0,04%
B	82	0,03%	B	56	0,02%	B	85	0,03%

Fonte: Elaborado pela autora (2019).

Quanto ao tratamento de fezes e urina no domicílio do paciente, para os atendimentos realizados entre abril e junho, os resultados foram os seguintes:

TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa
TOTAL	300362	100,00%	TOTAL	309200	100,00%	TOTAL	287148	100,00%
F	277187	92,28%	F	285908	92,47%	F	265865	92,59%
D	15792	5,26%	D	15969	5,16%	D	14413	5,02%
A	3594	1,20%	A	3627	1,17%	A	3385	1,18%
E	3535	1,18%	E	3436	1,11%	E	3265	1,14%
C	136	0,05%	C	152	0,05%	B	123	0,04%
B	118	0,04%	B	108	0,03%	C	97	0,03%

Fonte: Elaborado pela autora (2019).

¹⁴ Fonte: <https://site.sanepar.com.br/a-sanepar/servicos/esgoto>

Já entre julho e setembro de 2018 as frequências foram:

TABELA 23 - FREQUÊNCIA ATRIBUTO 'fezesUrina' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa	fezesUrina	Frequência Absoluta	Frequência Relativa
TOTAL	270779	99,92%	TOTAL	306617	100,00%	TOTAL	62	100,00%
F	250242	92,42%	F	283345	92,41%	F	57	91,94%
D	14138	5,22%	D	16005	5,22%	D	4	6,45%
A	3206	1,18%	A	3726	1,22%	A	1	1,61%
E	2965	1,09%	E	3283	1,07%			
C	137	0,05%	C	131	0,04%			
B	91	0,03%	B	127	0,04%			

Fonte: Elaborado pela autora (2019).

Ao ser acometido por uma doença, o paciente dispõe de diversas possibilidades de procura por tratamento e é a respeito disso que a tabela apresenta os resultados das frequências dos serviços buscados. Em primeiro lugar, em todos os meses analisados, os pacientes vão até uma Unidade de Saúde (US), no segundo e terceiro lugares, a variação ocorre entre Unidades e Hospitais (HS).

TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa
TOTAL	241263	100,00%	TOTAL	231529	100,00%	TOTAL	278277	100,00%
US	119030	49,34%	US	114036	49,25%	US	136191	48,94%
USHS	40940	16,97%	USHS	39004	16,85%	USHS	46854	16,84%
HSUS	38622	16,01%	HSUS	37427	16,17%	HSUS	45760	16,44%
HS	15612	6,47%	HS	14954	6,46%	HS	18002	6,47%
OT	9868	4,09%	OT	9682	4,18%	OT	11826	4,25%
FMUSHS	1232	0,51%	FMHSUS	1178	0,51%	FMUSHS	1455	0,52%
HSFMUS	1206	0,50%	FMUSHS	1153	0,50%	FMHSUS	1450	0,52%
OTUS	1169	0,48%	USOT	1121	0,48%	OTUS	1362	0,49%
FMHSUS	1145	0,47%	OTUS	1080	0,47%	USFMHS	1309	0,47%
USOT	1118	0,46%	USFMHS	1071	0,46%	HSFMUS	1247	0,45%

Fonte: Elaborado pela autora (2019).

Nas consultas entre os meses de abril a junho, a frequência predominante US, o que representa Unidade de Saúde, ou seja, em caso de doença, o paciente a procura:

TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa
TOTAL	300341	100,00%	TOTAL	309181	100,00%	TOTAL	287137	100,00%
US	148146	49,33%	US	151748	49,08%	US	141677	49,34%
USHS	50748	16,90%	HSUS	54767	17,71%	HSUS	50975	17,75%
HSUS	48537	16,16%	USHS	47161	15,25%	USHS	43451	15,13%
HS	18894	6,29%	HS	19993	6,47%	HS	18504	6,44%
OT	12709	4,23%	OT	13232	4,28%	OT	11891	4,14%
FMUSHS	1484	0,49%	HSUSFM	1890	0,61%	HSUSFM	1688	0,59%
OTUS	1460	0,49%	USOT	1594	0,52%	USOT	1567	0,55%
USFMHS	1442	0,48%	USFMHS	1578	0,51%	USFMHS	1526	0,53%
FMHSUS	1423	0,47%	USHSFM	1468	0,47%	FMHSUS	1351	0,47%
USOT	1407	0,47%	OTUS	1434	0,46%	HSFMUS	1317	0,46%

Fonte: Elaborado pela autora (2019).

Por fim, entre julho e setembro, os resultados foram:

TABELA 24 - FREQUÊNCIA ATRIBUTO 'casoDoenca' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa	casoDoenca	Frequência Absoluta	Frequência Relativa
TOTAL	270761	100,00%	TOTAL	306599	100,00%	TOTAL	62	100,00%
US	133356	49,25%	US	151428	49,39%	US	41	66,13%
HSUS	47752	17,64%	HSUS	54565	17,80%	HSUS	11	17,74%
USHS	40996	15,14%	USHS	46287	15,10%	USHS	4	6,45%
HS	17381	6,42%	HS	20019	6,53%	HS	3	4,84%
OT	11104	4,10%	OT	12587	4,11%	HSFMOT	1	1,61%
HSUSFM	1754	0,65%	HSUSFM	1756	0,57%	HSFMUS	1	1,61%
USOT	1431	0,53%	USOT	1600	0,52%	OT	1	1,61%
USFMHS	1334	0,49%	FMHSUS	1395	0,45%			
FMHSUS	1307	0,48%	USFMHS	1393	0,45%			
USHSFM	1268	0,47%	USHSFM	1390	0,45%			

Fonte: Elaborado pela autora (2019).

Quanto aos grupos comunitários dos quais os pacientes podem fazer parte, pelos valores identificados nas frequências, constatou-se que não há participação para os meses e situação analisados; a opção de “outros” ocupa a segunda posição, não sendo encontrada a composição deste grupo; “grupo religioso” a terceira; “cooperativa” a quarta; e por fim, “associação”.

TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa
TOTAL	241239	100,00%	TOTAL	231507	100,00%	TOTAL	278246	100,00%
D	169244	70,16%	D	162140	70,04%	D	195809	70,37%
E	39187	16,24%	E	37882	16,36%	E	45414	16,32%
C	23669	9,81%	C	22885	9,89%	C	26466	9,51%
B	6634	2,75%	B	6227	2,69%	B	7527	2,71%
A	2505	1,04%	A	2373	1,03%	A	3030	1,09%

Fonte: Elaborado pela autora (2019).

Entre abril e junho, a opção D continuou com as maiores frequências:

TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa
TOTAL	300312	100,00%	TOTAL	309154	100,00%	TOTAL	287104	100,00%
D	211670	70,48%	D	217568	70,38%	D	203129	70,75%
E	48839	16,26%	E	50157	16,22%	E	45797	15,95%
C	28907	9,63%	C	29870	9,66%	C	27432	9,55%
B	7908	2,63%	B	8396	2,72%	B	7907	2,75%
A	2988	0,99%	A	3163	1,02%	A	2839	0,99%

Fonte: Elaborado pela autora (2019).

Entre julho e setembro, os seguintes resultados foram obtidos:

TABELA 25 - FREQUÊNCIA ATRIBUTO 'grupoComunitario' (jul-set/2018)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa	grupoComunitario	Frequência Absoluta	Frequência Relativa
TOTAL	270737	100,00%	TOTAL	306553	100,00%	TOTAL	62	100,00%
D	191269	70,65%	D	216797	70,72%	D	53	85,48%
E	43795	16,18%	E	49590	16,18%	E	5	8,06%
C	25337	9,36%	C	28735	9,37%	C	2	3,23%
B	7504	2,77%	B	8352	2,72%	B	1	1,61%
A	2832	1,05%	A	3079	1,00%	A	1	1,61%

Fonte: Elaborado pela autora (2019).

Os pacientes atendidos nos meses analisados (janeiro a setembro) residem, predominantemente, nos bairros Cidade Industrial, Sítio Cercado, Cajuru, Uberaba e Tatuquara, conforme TABELA 26:

TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (jan-mar/2018)

Atributo/Mês	jan/18		Atributo/Mês	fev/18		Atributo/Mês	mar/18	
bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa
TOTAL	267093	100,00%	TOTAL	254559	100,00%	TOTAL	305482	100,00%
CIDADE INDUSTRIAL	37193	13,93%	CIDADE INDUSTRIAL	36441	14,32%	CIDADE INDUSTRIAL	42014	13,75%
SITIO CERCADO	21457	8,03%	SITIO CERCADO	20919	8,22%	SITIO CERCADO	24228	7,93%
CAJURU	17952	6,72%	CAJURU	15902	6,25%	CAJURU	18890	6,18%
TATUQUARA	13736	5,14%	UBERABA	12296	4,83%	UBERABA	15230	4,99%
UBERABA	12644	4,73%	TATUQUARA	12288	4,83%	TATUQUARA	15080	4,94%
BOQUEIRAO	12327	4,62%	BOQUEIRAO	11521	4,53%	BOQUEIRAO	14038	4,60%
PINHEIRINHO	9459	3,54%	PINHEIRINHO	9218	3,62%	PINHEIRINHO	11266	3,69%
ALTO BOQUEIRAO	9328	3,49%	ALTO BOQUEIRAO	8791	3,45%	ALTO BOQUEIRAO	10107	3,31%
XAXIM	8461	3,17%	XAXIM	7729	3,04%	XAXIM	9873	3,23%
CAMPO DE SANTANA	8268	3,10%	BAIRRO ALTO	6854	2,69%	BAIRRO ALTO	9417	3,08%

Fonte: Elaborado pela autora (2019).

É necessário destacar que estão na TABELA 26 apenas as dez (10) maiores frequências:

TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (abr-jun/2018)

Atributo/Mês	abr/18		Atributo/Mês	mai/18		Atributo/Mês	jun/18	
bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa
TOTAL	329153	100,00%	TOTAL	339206	100,00%	TOTAL	316056	100,00%
CIDADE INDUSTRIAL	44651	13,57%	CIDADE INDUSTRIAL	45984	13,56%	CIDADE INDUSTRIAL	42387	13,41%
SITIO CERCADO	26747	8,13%	SITIO CERCADO	26967	7,95%	SITIO CERCADO	25081	7,94%
CAJURU	21005	6,38%	CAJURU	22667	6,68%	CAJURU	20110	6,36%
UBERABA	15884	4,83%	UBERABA	16233	4,79%	TATUQUARA	16013	5,07%
BOQUEIRAO	15399	4,68%	BOQUEIRAO	16222	4,78%	UBERABA	15249	4,82%
TATUQUARA	15248	4,63%	TATUQUARA	15690	4,63%	BOQUEIRAO	14804	4,68%
PINHEIRINHO	13005	3,95%	PINHEIRINHO	13349	3,94%	PINHEIRINHO	11482	3,63%
ALTO BOQUEIRAO	11067	3,36%	ALTO BOQUEIRAO	12147	3,58%	ALTO BOQUEIRAO	10790	3,41%
XAXIM	10552	3,21%	XAXIM	11077	3,27%	XAXIM	9696	3,07%
BAIRRO ALTO	10122	3,08%	BAIRRO ALTO	10177	3,00%	CAMPO DE SANTANA	9299	2,94%

Fonte: Elaborado pela autora (2019).

Entre julho e setembro, os resultados foram:

TABELA 26 - FREQUÊNCIA ATRIBUTO 'bairro' (jul-set/18)

Atributo/Mês	jul/18		Atributo/Mês	ago/18		Atributo/Mês	set/18	
bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa	bairro	Frequência Absoluta	Frequência Relativa
TOTAL	298532	100,00%	TOTAL	336161	100,00%	TOTAL	79	100,00%
CIDADE INDUSTRIAL	39334	13,18%	CIDADE INDUSTRIAL	49177	14,63%	CIDADE INDUSTRIAL	16	20,25%
SITIO CERCADO	25059	8,39%	SITIO CERCADO	25872	7,70%	CAJURU	9	11,39%
CAJURU	19385	6,49%	CAJURU	21820	6,49%	UBERABA	5	6,33%
TATUQUARA	15093	5,06%	UBERABA	16788	4,99%	SITIO CERCADO	5	6,33%
UBERABA	14037	4,70%	BOQUEIRAO	15576	4,63%	CAMPO DE SANTANA	5	6,33%
BOQUEIRAO	13902	4,66%	TATUQUARA	15476	4,60%	AUGUSTA	4	5,06%
ALTO BOQUEIRAO	11090	3,71%	PINHEIRINHO	12147	3,61%	NAO INFORMADO	4	5,06%
PINHEIRINHO	10788	3,61%	ALTO BOQUEIRAO	11417	3,40%	BACACHERI	3	3,80%
XAXIM	9982	3,34%	XAXIM	10596	3,15%	TATUQUARA	2	2,53%
CAMPO DE SANTANA	9279	3,11%	BAIRRO ALTO	9971	2,97%	SANTA CANDIDA	2	2,53%

Fonte: Elaborado pela autora (2019).

Para a visualização dos dados, a ferramenta da Microsoft Power BI, foi utilizada. Trata-se, de acordo com seu site de “[...] uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente

envolventes e interativas. Os dados podem estar em uma planilha do Excel ou em uma coleção de data warehouses híbridos locais ou baseados na nuvem¹⁵.

Foram elaborados três (3) relatórios referentes à consulta, internamento e características sócio econômicos:

¹⁵ Fonte: <https://docs.microsoft.com/pt-br/power-bi/power-bi-overview>

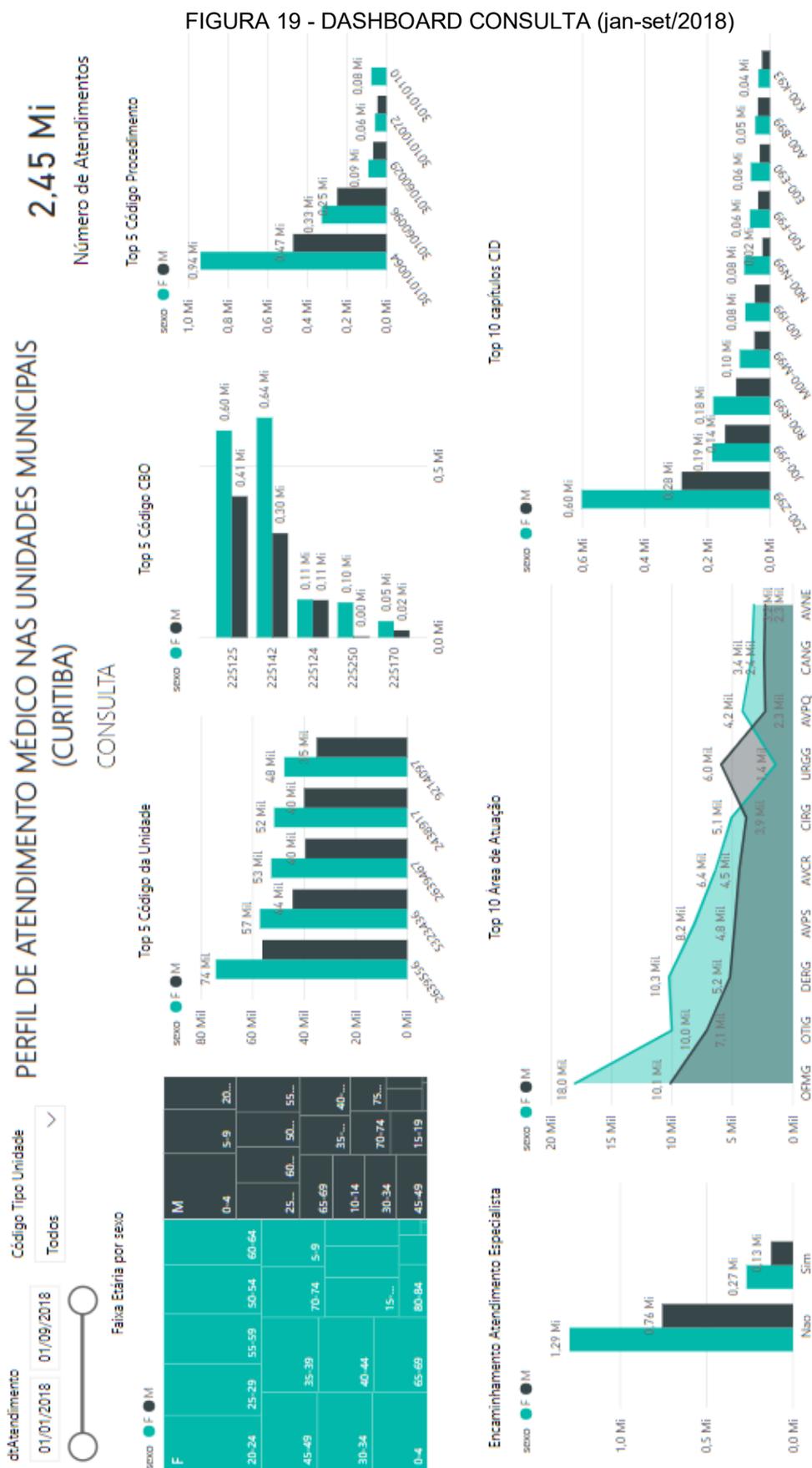
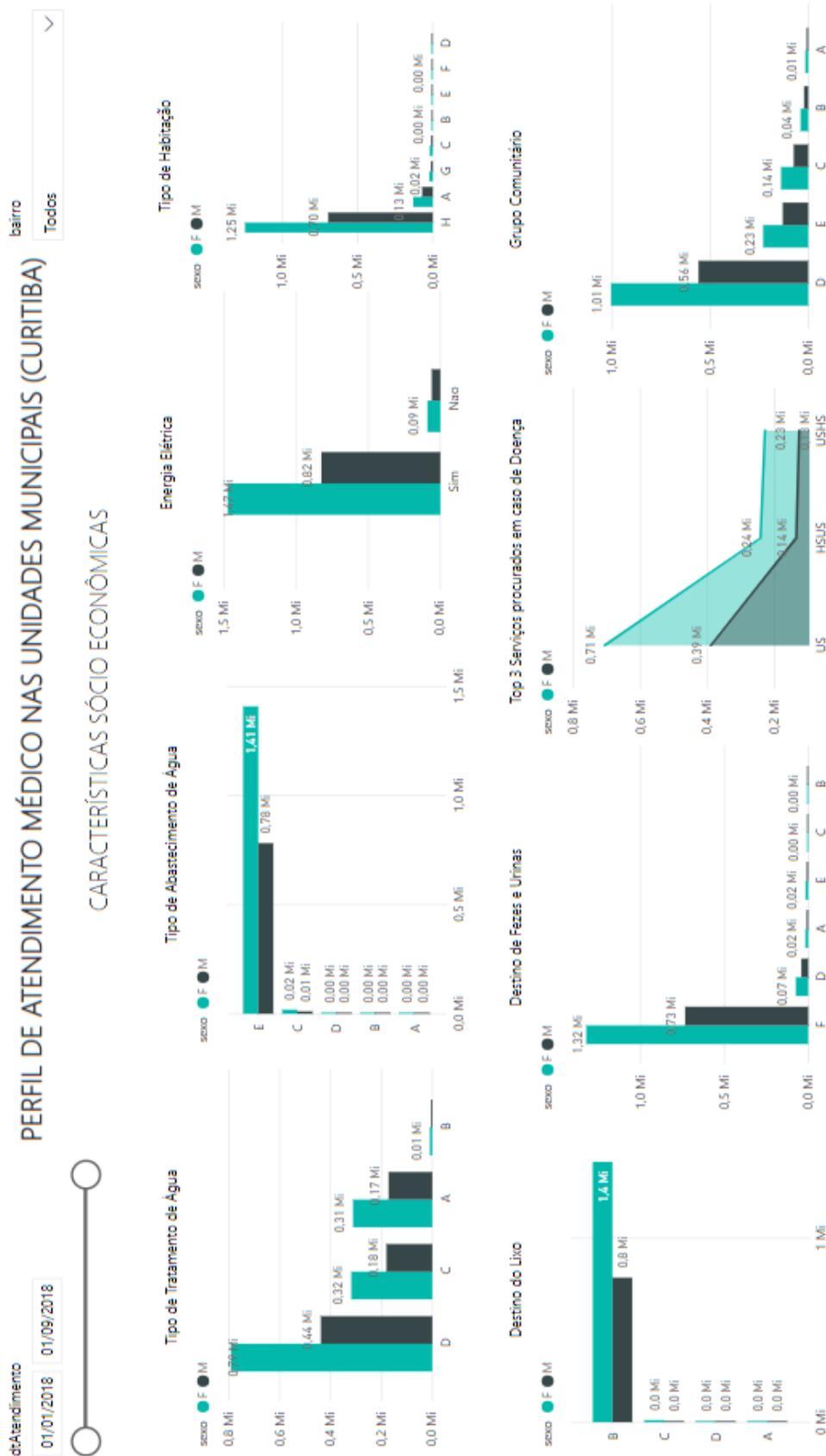


FIGURA 19 - DASHBOARD CONSULTA (jan-set/2018)

FONTE: Elaborado pela autora com a ferramenta Power BI (2019).

Foram criados gráficos de colunas e barras clusterizadas, de área, treemap e cartão de linha múltipla. Uma variação em relação ao cálculo das frequências, é que na visualização foi utilizado filtro para delimitar os meses, desta forma, não foi necessário construir um relatório para cada mês, pois com o filtro, conseguimos identificar os valores de janeiro a setembro de 2018. Outro ponto divergente é o comparativo entre pacientes do sexo masculino e feminino, no caso, os resultados referem-se a cada um deles, possibilitando a análise pela ótica do gênero, por exemplo.

FIGURA 21 - DASHBOARD CARACTERÍSTICAS SÓCIO ECONÔMICAS (jan-set/2018)



FONTE: Elaborado pela autora com a ferramenta Power BI (2019).

Se faz necessário evidenciar que a escolha do atributo de “capCID” como meta para as classificações, foi estabelecida pelo critério de que a Classificação Internacional de Doenças (CID) é o resultado do atendimento e a principal descoberta para o tratamento da enfermidade.

Acerca dos dashboards, conforme mencionado, foram separados em três categorias, onde há a distribuição do número total de atendimento, 2,45 milhões nos meses analisados, sendo estas: consulta, internamento e aspectos sócio econômicos; a primeira representa o atendimento inicial, com os atributos de faixa etária, código da unidade, código do CBO, código do procedimento, se houve ou não encaminhamento para atendimento especialista, área de atuação e capítulo do CID, sendo que houve ordenação por ranking, variando dos cinco (5) aos dez (10) primeiros. A partir da visualização separada por gênero, podemos identificar que houve mais atendimentos para pacientes do sexo feminino, ressalvas para a área de atuação em que na especialidade de Urologia Geral, o cenário foi invertido.

Com estes gráficos, é possível identificar e compreender o conjunto de dados como um todo, permitindo uma análise de para quais áreas de atuação ocorre o maior número de encaminhamentos, no caso: oftalmologia geral, otorrinolaringologia geral e dermatologia geral, dessa forma, medidas poderão ser tomadas de modo a aumentar o número de profissionais destas áreas em caso de demanda, por exemplo. Quanto ao CID, a predominância é do capítulo Z00-Z99 (fatores que influenciam o estado de saúde e o contato com os serviços de saúde); seguido por J00-J99 (doenças do aparelho respiratório); e R00-R99 (sintomas, sinais e achados anormais de exames cínicos e de laboratório, não classificados em outra parte) o que também representa uma possibilidade de desenvolvimento de políticas públicas e demais investigações sobre as enfermidades relacionadas à estes capítulos.

Quanto ao internamento, a predominância é da sua não ocorrência, em ambos os sexos. Para os estabelecimentos solicitantes e de destino, foi elaborado ranking dos dez (10) principais, o destaque dos três (3) primeiros solicitantes está no UPA Fazendinha, UPA Boa Vista e UPA Boqueirão, enquanto os de destino são Hospital Nossa Senhora do Rocio - Campo Largo, Clínica Dr. Helio Rotenberg e Hospital de Clínicas. Novamente, a partir dos

gráficos e entendimento dos dados, há a possibilidade de identificar os estabelecimentos que mais recebem pacientes e, se necessário e possível, destinar investimentos para estes hospitais e clínicas. Acerca do CID de internamento, a diferença é o surgimento, no início do ranking, do capítulo I00-I99 (doenças do aparelho circulatório).

Por fim, estão presentes os dados sobre os aspectos sócio econômicos, com a opção de filtro por bairro e análise por gênero; trata-se de um dashboard que permite a visualização dos atributos de preenchimento obrigatório quando é um cadastro definitivo do paciente. Ressalta-se que no gráfico de serviços procurados em caso de doença, a Unidade de Saúde ficou em primeiro lugar, portanto, representa uma alta procura o que reforça a necessidade de aprimoramento e fomento deste tipo de atendimento em virtude da demanda. Em conclusão, reforça-se a relevância da necessidade de análise dos dados e gráficos elaborados, pois, ações e políticas públicas poderão ser tomadas a partir da compreensão do cenário no atendimento do Sistema Único de Saúde em Curitiba (nos meses de janeiro a setembro de 2018).

4.2 MINERAÇÃO DE DADOS

Após o pré-processamento e análise estatística dos atributos numéricos e nominais, a próxima etapa consiste na Mineração de Dados, nela serão descobertos os padrões nas bases de dados analisadas, entre janeiro e setembro de 2018.

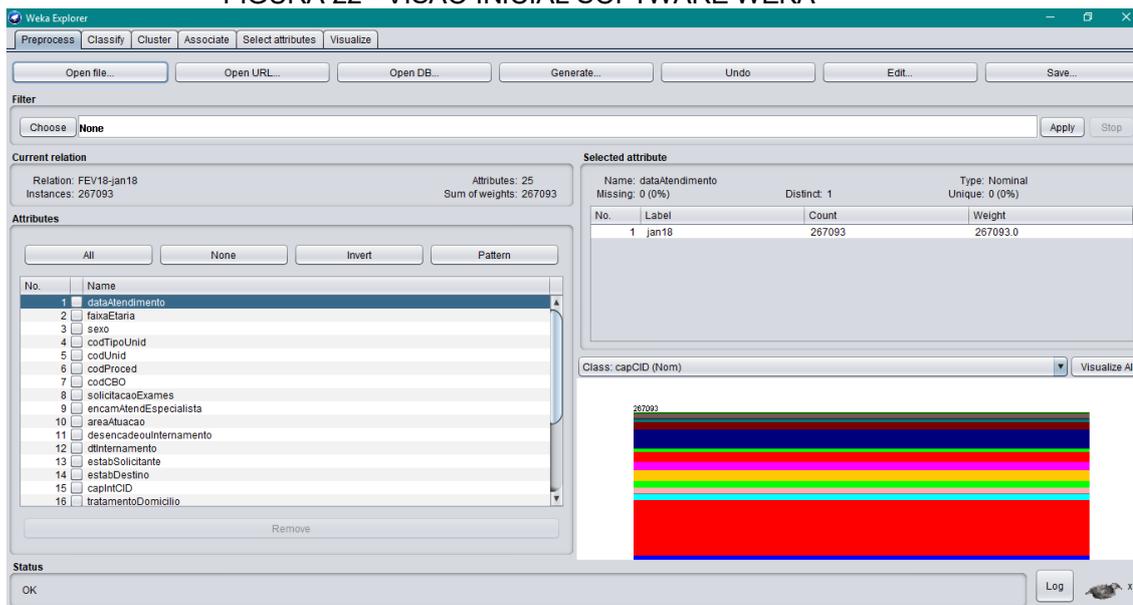
O software selecionado para a Mineração é o Weka (versão 3.8.3) e, conforme sua descrição, (tradução nossa) “[...] é uma coleção de algoritmos de *machine learning* para tarefas de Mineração de Dados. Contém ferramentas para preparação dos dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização”, desenvolvido na Universidade de Waikato na Nova Zelândia. A decisão pelo Weka foi tomada pelo grau de familiaridade da autora com a ferramenta; outros softwares/tools para mineração serão brevemente descritos na seção “trabalhos futuros”. Após essa definição, os experimentos podem ser executados.

Os downloads das bases no Portal de Dados Abertos foram de arquivos no formato *.csv* (*Comma-separated values*), para o tratamento, os arquivos foram transformados em *.xlsx* (Microsoft Excel), em seguida, novamente para *.csv*, entretanto, para o funcionamento Weka, o arquivo precisa estar no formato *.arff* (*Attribute-Relation File Format*) e para isso é fundamental abrir o arquivo *.csv* em leitor de texto (notepad, wordpad, word, por exemplo) e alterar o ponto e vírgula (;) para vírgula (,), depois carregar o arquivo no Weka e salvar em *.arff*.

Para os experimentos, também foi necessário acessar o *.arff* e adicionar manualmente as opções que o software ainda estava reconhecendo como numéricas (código da unidade, código do procedimento, código do CBO), somente ao concluir esta ação a base poderá passar pela Mineração.

A tela inicial do Weka apresenta as opções para acessar a base, remover atributos que julgar dispensáveis e visualização dos dados e sua tipologia, conforme a FIGURA 22:

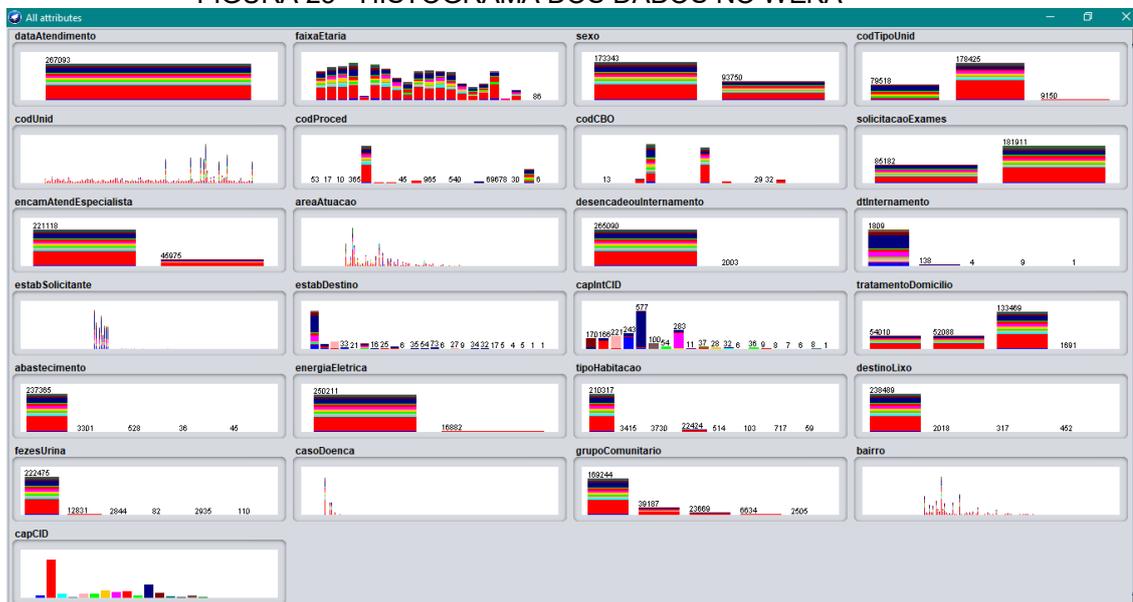
FIGURA 22 - VISÃO INICIAL SOFTWARE WEKA



Fonte: Software Weka - versão 3.8.3 (2019).

A FIGURA 23 apresenta a visualização dos dados com o Weka, no caso, os atributos estão divididos de acordo com as cores do atributo meta (capCID):

FIGURA 23 - HISTOGRAMA DOS DADOS NO WEKA



Fonte: Software Weka - versão 3.8.3 (2019).

Antes da aplicação dos algoritmos, é importante fazer uma seleção de quais podem ser aplicados e para isso, se faz necessário entender alguns conceitos fundamentais, como as chamadas “Tarefas de Mineração de Dados” que são sucintamente explicadas nos tópicos abaixo:

Análise descritiva de dados - [...] uso de ferramentas capazes de medir, explorar e descrever características intrínsecas aos dados.

Predição: classificação e estimação - *predição* é terminologia usada para se referir à construção e ao uso de um modelo para avaliar a classe de um objeto não rotulado ou para estimar o valor de um ou mais atributos de dado objeto. [...] classificação e estimação constituem os dois principais tipos de problemas de predição, sendo que a classificação é usada para prever *valores discretos*, ao passo que a estimação é usada para prever *valores contínuos*.

Análise de grupos - [...] processo de separar (particionar ou segmentar) um conjunto de objetos em *grupos* (do inglês *clusters*) de objetos similares.

Associação - [...] corresponde à descoberta de regras de associação que apresentam valores de atributos que ocorrem concomitantemente em uma base de dados.

Detecção de anomalias - uma base de dados pode conter objetos que não seguem o comportamento ou não possuem a característica comum dos dados ou de um modelo que os represente. Esses dados são conhecidos como *anomalias* ou *valores discrepantes (outliers)*.

(CASTRO;FERRARI, 2016, p.7-10, grifo do autor)

Portanto, trata-se de decisão extremamente relevante, uma vez que representa o início da Mineração, o entendimento básico sobre ela; para este trabalho serão utilizados os métodos de associação e classificação. Destaca-se que o atributo de data de atendimento foi removido, pois não é pertinente, uma vez que as bases já representam esta data.

4.2.1 Apriori

Conforme mencionado, serão geradas Regras de Associação e para isto, algoritmo selecionado foi o Apriori; de acordo com Castro e Ferrari (2016, p.247, grifo do autor) “[...] é o método mais conhecido para a mineração de regras de associação e emprega busca em profundidade e gera conjuntos de itens candidatos de k elementos a partir de conjuntos de itens com $k-1$ elementos”.

FIGURA 24 - PARÂMETROS *DEFAULT* APRIORI

Fonte: Software Weka - versão 3.8.3 (2019).

Os parâmetros acima foram mantidos para o primeiro experimento e no Weka significam:

- minMetric - pontuação métrica mínima. Considere apenas regras com pontuações mais altas que esse valor;
- verbose - Se habilitado, o algoritmo será executado no modo detalhado;
- numRules - Número de regras para encontrar;
- lowerBoundMinSupport - Limite inferior para suporte mínimo;
- classIndex - Índice do atributo de classe. Se definido como -1, o último atributo é considerado atributo de classe;

- outputItemSets - Se ativada, os conjuntos de itens também são enviados;
- car - Se habilitado, as regras de associação de classe são extraídas em vez das regras de associação (gerais);
- doNotCheckCapabilities - Se definido, os recursos do associador não são verificados antes que o associador seja criado (use com cuidado para reduzir o tempo de execução);
- removeAllMissingCols - Remove colunas com todos os valores ausentes;
- nível de significância - Nível de significância. Teste de significância (somente métrica de confiança);
- treatZeroAsMissing - Se ativado, zero (ou seja, o primeiro valor de um valor nominal) é tratado da mesma maneira que um valor ausente;
- delta - Diminui iterativamente o suporte por este fator. Reduz o suporte até que o suporte mínimo seja atingido ou o número necessário de regras tenha sido gerado;
- metricType - Defina o tipo de métrica pelo qual classificar as regras. Confiança é a proporção dos exemplos abrangidos pela premissa que também são abrangidos pela consequência (as regras de associação de classe só podem ser extraídas usando confiança). O aumento é a confiança dividida pela proporção de todos os exemplos cobertos pela consequência. Esta é uma medida da importância da associação que é independente do suporte. A alavancagem é a proporção de exemplos adicionais cobertos tanto pela premissa quanto pela consequência acima daquelas esperadas se a premissa e a consequência fossem independentes uma da outra. O número total de exemplos que isso representa é apresentado entre parênteses após a alavancagem. A condenação é outra medida de afastamento da independência. A convicção é dada por $P(\text{premissa} | \text{Consequência}) / P(\text{premissa}, \text{Consequência})$;
- upperBoundMinSupport - Limite superior para suporte mínimo. Comece a diminuir iterativamente o suporte mínimo deste valor.

As seguintes regras foram descobertas para os atendimentos realizados no mês de janeiro de 2018:

FIGURA 25 - RESULTADOS APRIORI - *DEFAULT* (jan/2018)

```
Apriori
=====

Minimum support: 0.85 (227029 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1

Best rules found:

1. abastecimento=E 237365 ==> destinoLixo=B 236232 <conf:(1)> lift:(1.11) lev:(0.09) [24287] conv:(22.42)
2. desencadeouInternamento=Nao abastecimento=E 235703 ==> destinoLixo=B 234577 <conf:(1)> lift:(1.11) lev:(0.09) [24116] conv:(22.4)
3. abastecimento=E 237365 ==> desencadeouInternamento=Nao 235703 <conf:(0.99)> lift:(1) lev:(0) [118] conv:(1.07)
4. abastecimento=E destinoLixo=B 236232 ==> desencadeouInternamento=Nao 234577 <conf:(0.99)> lift:(1) lev:(0) [116] conv:(1.07)
5. destinoLixo=B 238489 ==> desencadeouInternamento=Nao 236817 <conf:(0.99)> lift:(1) lev:(0) [116] conv:(1.07)
6. energiaEletrica=Sim 250211 ==> desencadeouInternamento=Nao 248388 <conf:(0.99)> lift:(1) lev:(0) [53] conv:(1.03)
7. desencadeouInternamento=Nao destinoLixo=B 236817 ==> abastecimento=E 234577 <conf:(0.99)> lift:(1.11) lev:(0.09) [24118] conv:(11.76)
8. destinoLixo=B 238489 ==> abastecimento=E 236232 <conf:(0.99)> lift:(1.11) lev:(0.09) [24287] conv:(11.76)
9. abastecimento=E 237365 ==> desencadeouInternamento=Nao destinoLixo=B 234577 <conf:(0.99)> lift:(1.11) lev:(0.09) [24118] conv:(9.65)
10. destinoLixo=B 238489 ==> desencadeouInternamento=Nao abastecimento=E 234577 <conf:(0.98)> lift:(1.11) lev:(0.09) [24116] conv:(7.16)
```

Fonte: Software Weka - versão 3.8.3 (2019).

As associações referem-se aos atributos de abastecimento de água, destino do lixo e necessidade de internamento do paciente, nenhuma relacionada ao capítulo do CID, uma vez que apesar de definido como meta, ainda não está passando pela classificação. A leitura da regra é feita da seguinte forma:

$$A \rightarrow C$$

Um exemplo a partir do resultado pode ser entendido por: se o destinoLixo for B (coletado), então abastecimento será E (rede pública). Um destaque neste resultado está no valor do suporte, este segundo Castro e Ferrari (2016, p.241) representa a significância estatística e a sua definição revela que

O suporte, ou cobertura, de uma regra é uma medida muito importante, pois regras com valores muito baixos de suporte ocorrem apenas ocasionalmente. Regras com baixo suporte também são de pouco interesse sob a perspectiva do negócio, pois não faz muito sentido promover itens que os clientes compram pouco em conjunto. Por essa razão, o suporte normalmente é usado para eliminar regras pouco interessantes. O suporte de uma regra de associação, $A \rightarrow C$, indica a frequência de ocorrência da regra, ou seja, a probabilidade de essa regra ser encontrada no conjunto total de transações da base. (CASTRO; FERRARI, 2016, p.241)

No experimento *default* ele ficou em 0,85. Em seguida, a confiança foi mantida, enquanto o número de regras foi ajustado para 10.000:

FIGURA 26 - RESULTADOS APRIORI EXPERIMENTO 1 (jan/2018)

```
Apriori
=====

Minimum support: 0.45 (120192 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11
```

Fonte: Software Weka - versão 3.8.3 (2019).

A partir da figura acima, é possível verificar que houve uma queda no suporte para 0,45, quanto às regras:

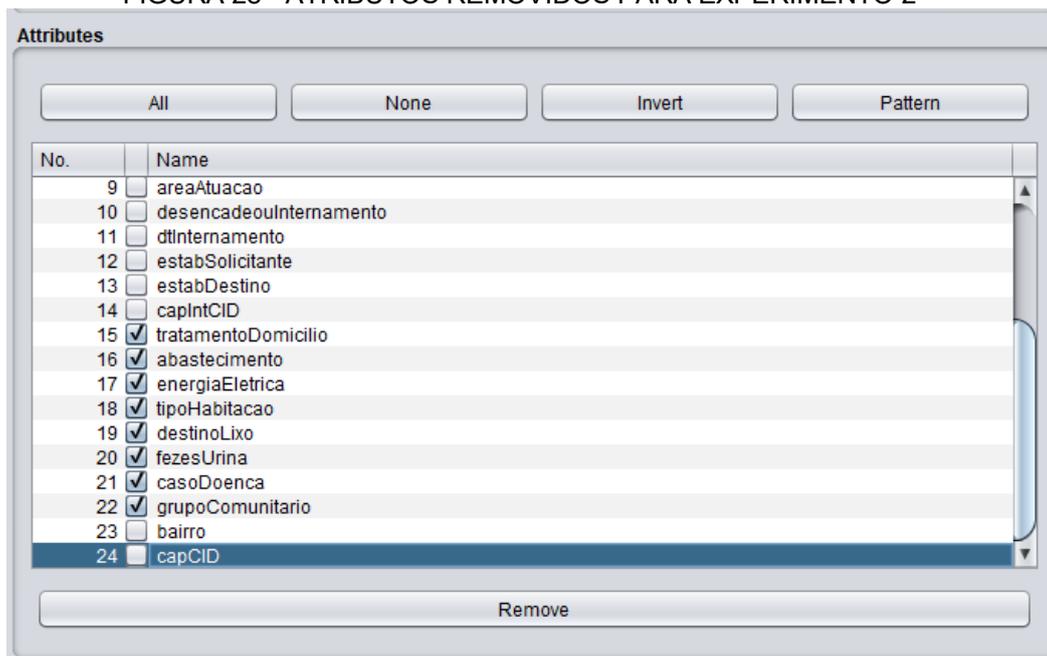
FIGURA 27 - REGRAS APRIORI EXPERIMENTO 1 (jan/2018)

```
9990. codTipoUnid=B abastecimento=E energiaEletrica=Sim 155460 ==> fezesUrina=F 143931 <conf:(0.93)> lift:(1.11) lev:(0.05) [14440] conv:(2.25)
9991. dataAtendimento=jan18 codTipoUnid=B abastecimento=E energiaEletrica=Sim 155460 ==> fezesUrina=F 143931 <conf:(0.93)> lift:(1.11) lev:(0.05) [14440] conv:(2.25)
9992. codTipoUnid=B abastecimento=E energiaEletrica=Sim 155460 ==> dataAtendimento=jan18 fezesUrina=F 143931 <conf:(0.93)> lift:(1.11) lev:(0.05) [14440] conv:(2.25)
9993. desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 224099 ==> destinoLixo=B fezesUrina=F 207471 <conf:(0.93)> lift:(1.11) lev:(0.08) [21324] conv:(2.28)
9994. dataAtendimento=jan18 desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 224099 ==> destinoLixo=B fezesUrina=F 207471 <conf:(0.93)> lift:(1.11) lev:(0.08) [21324] conv
9995. desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 224099 ==> dataAtendimento=jan18 destinoLixo=B fezesUrina=F 207471 <conf:(0.93)> lift:(1.11) lev:(0.08) [21324] conv
9996. solicitacaoExames=Nao abastecimento=E energiaEletrica=Sim 152808 ==> desencadeouInternamento=Nao destinoLixo=B fezesUrina=F 141461 <conf:(0.93)> lift:(1.12) lev:(0.06) [15433] conv
9997. dataAtendimento=jan18 solicitacaoExames=Nao abastecimento=E energiaEletrica=Sim 152808 ==> desencadeouInternamento=Nao destinoLixo=B fezesUrina=F 141461 <conf:(0.93)> lift:(1.12) 1
9998. solicitacaoExames=Nao abastecimento=E energiaEletrica=Sim 152808 ==> dataAtendimento=jan18 desencadeouInternamento=Nao destinoLixo=B fezesUrina=F 141461 <conf:(0.93)> lift:(1.12) 1
9999. tratamentoDomicilio=D abastecimento=E destinoLixo=B 131010 ==> fezesUrina=F 121281 <conf:(0.93)> lift:(1.11) lev:(0.05) [12156] conv:(2.25)
10000. dataAtendimento=jan18 tratamentoDomicilio=D abastecimento=E destinoLixo=B 131010 ==> fezesUrina=F 121281 <conf:(0.93)> lift:(1.11) lev:(0.05) [12156] conv:(2.25)
```

Fonte: Software Weka - versão 3.8.3 (2019).

Foram identificadas 10.000 regras, o que dificulta a análise. Para o segundo experimento, optou-se por manter apenas os atributos que são de preenchimento obrigatório, independente do cadastro definitivo, desta forma foram removidos os atributos selecionados na FIGURA 28, restando 16.

FIGURA 28 - ATRIBUTOS REMOVIDOS PARA EXPERIMENTO 2



Fonte: Software Weka - versão 3.8.3 (2019).

Quanto aos parâmetros, a confiança foi alterada para 1,0 e o número de regras para 15. Os resultados do experimento 2 estão na FIGURA 29:

FIGURA 29 - RESULTADOS DO EXPERIMENTO 2 APRIORI (jan/2018)

```

Apriori
*****

Minimum support: 0.3 (80128 instances)
Minimum metric <confidence>: 1
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 27
Size of set of large itemsets L(3): 26
Size of set of large itemsets L(4): 10
Size of set of large itemsets L(5): 1

Best rules found:

1. codProced=301010064 159319 ==> codTipoUnid=B 159319 <conf: (1)> lift: (1.5) lev: (0.2) [52889] conv: (52889.81)
2. codProced=301010064 desencadeouInternamento=Nao 159182 ==> codTipoUnid=B 159182 <conf: (1)> lift: (1.5) lev: (0.2) [52844] conv: (52844.33)
3. codProced=301010064 encamAtendEspecialista=Nao 117733 ==> codTipoUnid=B 117733 <conf: (1)> lift: (1.5) lev: (0.15) [39084] conv: (39084.33)
4. codProced=301010064 encamAtendEspecialista=Nao desencadeouInternamento=Nao 117602 ==> codTipoUnid=B 117602 <conf: (1)> lift: (1.5) lev: (0.15) [39040] conv: (39040.84)
5. codCBO=225142 109686 ==> codTipoUnid=B 109686 <conf: (1)> lift: (1.5) lev: (0.14) [36412] conv: (36412.93)
6. codCBO=225142 desencadeouInternamento=Nao 109577 ==> codTipoUnid=B 109577 <conf: (1)> lift: (1.5) lev: (0.14) [36376] conv: (36376.74)
7. sexo=F codProced=301010064 106969 ==> codTipoUnid=B 106969 <conf: (1)> lift: (1.5) lev: (0.13) [35510] conv: (35510.95)
8. sexo=F codProced=301010064 desencadeouInternamento=Nao 106916 ==> codTipoUnid=B 106916 <conf: (1)> lift: (1.5) lev: (0.13) [35493] conv: (35493.36)
9. codProced=301010064 solicitacaoExames=Nao 99484 ==> codTipoUnid=B 99484 <conf: (1)> lift: (1.5) lev: (0.12) [33026] conv: (33026.13)
10. codProced=301010064 solicitacaoExames=Nao desencadeouInternamento=Nao 99367 ==> codTipoUnid=B 99367 <conf: (1)> lift: (1.5) lev: (0.12) [32987] conv: (32987.29)
11. codProced=301010064 codCBO=225142 96696 ==> codTipoUnid=B 96696 <conf: (1)> lift: (1.5) lev: (0.12) [32100] conv: (32100.58)
12. codProced=301010064 codCBO=225142 desencadeouInternamento=Nao 96597 ==> codTipoUnid=B 96597 <conf: (1)> lift: (1.5) lev: (0.12) [32067] conv: (32067.72)
13. codProced=301010064 solicitacaoExames=Nao encamAtendEspecialista=Nao 84535 ==> codTipoUnid=B 84535 <conf: (1)> lift: (1.5) lev: (0.11) [28063] conv: (28063.44)
14. codProced=301010064 solicitacaoExames=Nao encamAtendEspecialista=Nao desencadeouInternamento=Nao 84421 ==> codTipoUnid=B 84421 <conf: (1)> lift: (1.5) lev: (0.1) [28025] conv: (28025.6)
15. codCBO=225142 encamAtendEspecialista=Nao 83750 ==> codTipoUnid=B 83750 <conf: (1)> lift: (1.5) lev: (0.1) [27802] conv: (27802.84)

```

Fonte: Software Weka - versão 3.8.3 (2019).

Para melhor visualização dos resultados:

1. codProced=301010064 → codTipoUnid=B
2. codProced=301010064 desencadeouInternamento=Nao → codTipoUnid=B
3. codProced=301010064 encamAtendEspecialista=Nao → codTipoUnid=B

4. codProced=301010064 encamAtendEspecialista=Nao
desencadeouInternamento=Nao → codTipoUnid=B
5. codCBO=225141 → codTipoUnid=B
6. codCBO=225142 desencadeouInternamento=Nao → codTipoUnid=B
7. sexo=F codProced=301010064 → codTipoUnid=B
8. sexo=F codProced=301010064 desencadeouInternamento=Nao →
codTipoUnid=B
9. codProced=301010064 solicitacaoExames=Nao → codTipoUnid=B
10. codProced=301010064 solicitacaoExames=Nao desencadeouInternamento=Nao
→ codTipoUnid=B
11. codProced=301010064 codCBO=225142 → codTipoUnid=B
12. codProced=301010064 codCBO=225142 desencadeouInternamento=Nao →
codTipoUnid=B
13. codProced=301010064 solicitacaoExames=Nao encamAtendEspecialista=Nao
14. codProced=301010064 solicitacaoExames=Nao encamAtendEspecialista=Nao
desencadeouInternamento=Nao → codTipoUnid=B
15. codCBO=225142 encamAtendEspecialista=Nao → codTipoUnid=B

Mesmo com a remoção dos atributos que não seriam tão interessantes para a análise, nenhuma regra relacionada ao capítulo do CID foi identificada, apesar de não ser um método de classificação, seria relevante já identificar uma regra de associação relacionada ao meta, entretanto, destacam-se que as descobertas referem-se aos procedimentos executados durante o atendimento (codProced), código do tipo da unidade, se ela é básica, UPA ou SIACE, sexo do paciente, solicitação de exame, encaminhamento para atendimento especialista e se houve internamento ou não; os dois experimentos tiveram um tempo de processamento rápido, questão de segundos. Outro importante aspecto, foi a queda no valor do suporte para 0,30 e conforme mencionado, valores baixos são de pouco interesse.

Para o experimento 3, manteve-se as 15 regras e ajuste no nível de confiança para 0,99:

FIGURA 30 - RESULTADOS DO EXPERIMENTO 3 APRIORI (jan/2018)

```

Apriori
=====

Minimum support: 0.8 (213674 instances)
Minimum metric <confidence>: 0.99
Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 7

Size of set of large itemsets L(4): 2

Best rules found:

1. abastecimento=E fezesUrina=F 220546 ==> destinoLixo=B 219986 <conf:(1)> lift:(1.12) lev:(0.09) [23059] conv:(42.1)
2. desencadeouInternamento=Nao abastecimento=E fezesUrina=F 218982 ==> destinoLixo=B 218424 <conf:(1)> lift:(1.12) lev:(0.09) [22893] conv:(41.95)
3. fezesUrina=F 222475 ==> destinoLixo=B 221859 <conf:(1)> lift:(1.12) lev:(0.09) [23209] conv:(38.62)
4. desencadeouInternamento=Nao fezesUrina=F 220897 ==> destinoLixo=B 220283 <conf:(1)> lift:(1.12) lev:(0.09) [23042] conv:(38.47)
5. abastecimento=E energiaEletrica=Sim 225682 ==> destinoLixo=B 224670 <conf:(1)> lift:(1.11) lev:(0.09) [23157] conv:(23.86)
6. desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 224099 ==> destinoLixo=B 223093 <conf:(1)> lift:(1.11) lev:(0.09) [22993] conv:(23.83)
7. abastecimento=E 237365 ==> destinoLixo=B 236232 <conf:(1)> lift:(1.11) lev:(0.09) [24287] conv:(22.42)
8. desencadeouInternamento=Nao abastecimento=E 235703 ==> destinoLixo=B 234577 <conf:(1)> lift:(1.11) lev:(0.09) [24116] conv:(22.4)
9. abastecimento=E 237365 ==> desencadeouInternamento=Nao 235703 <conf:(0.99)> lift:(1) lev:(0) [118] conv:(1.07)
10. abastecimento=E destinoLixo=B 236232 ==> desencadeouInternamento=Nao 234577 <conf:(0.99)> lift:(1) lev:(0) [116] conv:(1.07)
11. destinoLixo=B 238489 ==> desencadeouInternamento=Nao 236817 <conf:(0.99)> lift:(1) lev:(0) [116] conv:(1.07)
12. abastecimento=E energiaEletrica=Sim 225682 ==> desencadeouInternamento=Nao 224099 <conf:(0.99)> lift:(1) lev:(0) [109] conv:(1.07)
13. energiaEletrica=Sim destinoLixo=B 226704 ==> desencadeouInternamento=Nao 225113 <conf:(0.99)> lift:(1) lev:(0) [109] conv:(1.07)
14. abastecimento=E energiaEletrica=Sim destinoLixo=B 224670 ==> desencadeouInternamento=Nao 223093 <conf:(0.99)> lift:(1) lev:(0) [107] conv:(1.07)
15. abastecimento=E fezesUrina=F 220546 ==> desencadeouInternamento=Nao 218982 <conf:(0.99)> lift:(1) lev:(0) [89] conv:(1.06)

```

Fonte: Software Weka - versão 3.8.3 (2019).

Os parâmetros do experimento 3 aplicados ao mês de fevereiro retornou resultados semelhantes, com alteração na ordem das regras, como pode ser constatado na FIGURA 31:

FIGURA 31 - RESULTADOS DO EXPERIMENTO 3 APRIORI (fev/2018)

```

Apriori
=====

Minimum support: 0.8 (203647 instances)
Minimum metric <confidence>: 0.99
Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 7

Size of set of large itemsets L(4): 2

Best rules found:

1. desencadeouInternamento=Nao abastecimento=E fezesUrina=F 210744 ==> destinoLixo=B 210203 <conf:(1)> lift:(1.11) lev:(0.08) [20632] conv:(39.06)
2. abastecimento=E fezesUrina=F 212207 ==> destinoLixo=B 211662 <conf:(1)> lift:(1.11) lev:(0.08) [20774] conv:(39.05)
3. fezesUrina=F 214063 ==> destinoLixo=B 213447 <conf:(1)> lift:(1.11) lev:(0.08) [20890] conv:(34.86)
4. desencadeouInternamento=Nao fezesUrina=F 212588 ==> destinoLixo=B 211976 <conf:(1)> lift:(1.11) lev:(0.08) [20746] conv:(34.84)
5. desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 215273 ==> destinoLixo=B 214294 <conf:(1)> lift:(1.11) lev:(0.08) [20649] conv:(22.07)
6. abastecimento=E energiaEletrica=Sim 216776 ==> destinoLixo=B 215785 <conf:(1)> lift:(1.11) lev:(0.08) [20788] conv:(21.95)
7. desencadeouInternamento=Nao abastecimento=E 226387 ==> destinoLixo=B 225306 <conf:(1)> lift:(1.11) lev:(0.09) [21663] conv:(21.02)
8. abastecimento=E 227959 ==> destinoLixo=B 226866 <conf:(1)> lift:(1.11) lev:(0.09) [21809] conv:(20.93)
9. destinoLixo=B 228984 ==> desencadeouInternamento=Nao 227411 <conf:(0.99)> lift:(1) lev:(0) [92] conv:(1.06)
10. abastecimento=E destinoLixo=B 226866 ==> desencadeouInternamento=Nao 225306 <conf:(0.99)> lift:(1) lev:(0) [90] conv:(1.06)
11. fezesUrina=F 214063 ==> desencadeouInternamento=Nao 212588 <conf:(0.99)> lift:(1) lev:(0) [82] conv:(1.06)
12. destinoLixo=B fezesUrina=F 213447 ==> desencadeouInternamento=Nao 211976 <conf:(0.99)> lift:(1) lev:(0) [81] conv:(1.05)
13. abastecimento=E destinoLixo=B fezesUrina=F 211662 ==> desencadeouInternamento=Nao 210203 <conf:(0.99)> lift:(1) lev:(0) [80] conv:(1.05)
14. abastecimento=E fezesUrina=F 212207 ==> desencadeouInternamento=Nao 210744 <conf:(0.99)> lift:(1) lev:(0) [80] conv:(1.05)
15. abastecimento=E 227959 ==> desencadeouInternamento=Nao 226387 <conf:(0.99)> lift:(1) lev:(0) [86] conv:(1.05)

```

Fonte: Software Weka - versão 3.8.3 (2019).

Em março, os resultados das regras de associação também foram similares.

FIGURA 32 - RESULTADOS DO EXPERIMENTO 3 APRIORI (mar/2018)

```
Apriori
=====

Minimum support: 0.8 (244386 instances)
Minimum metric <confidence>: 0.99
Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6
Size of set of large itemsets L(2): 10
Size of set of large itemsets L(3): 7
Size of set of large itemsets L(4): 2

Best rules found:

1. desencadeouInternamento=Nao abastecimento=E fezesUrina=F 252878 ==> destinoLixo=B 252220 <conf:(1)> lift:(1.11) lev:(0.08) [24341] conv:(37.94)
2. abastecimento=E fezesUrina=F 254573 ==> destinoLixo=B 253909 <conf:(1)> lift:(1.11) lev:(0.08) [24503] conv:(37.85)
3. desencadeouInternamento=Nao fezesUrina=F 255102 ==> destinoLixo=B 254406 <conf:(1)> lift:(1.11) lev:(0.08) [24523] conv:(36.18)
4. fezesUrina=F 256811 ==> destinoLixo=B 256107 <conf:(1)> lift:(1.11) lev:(0.08) [24684] conv:(36.01)
5. desencadeouInternamento=Nao abastecimento=E energiaEletrica=Sim 258833 ==> destinoLixo=B 257674 <conf:(1)> lift:(1.1) lev:(0.08) [24429] conv:(22.06)
6. abastecimento=E energiaEletrica=Sim 260588 ==> destinoLixo=B 259418 <conf:(1)> lift:(1.1) lev:(0.08) [24591] conv:(22)
7. desencadeouInternamento=Nao abastecimento=E 272138 ==> destinoLixo=B 270857 <conf:(1)> lift:(1.1) lev:(0.08) [25622] conv:(20.99)
8. abastecimento=E 273977 ==> destinoLixo=B 272685 <conf:(1)> lift:(1.1) lev:(0.08) [25793] conv:(20.95)
9. destinoLixo=B fezesUrina=F 256107 ==> desencadeouInternamento=Nao 254406 <conf:(0.99)> lift:(1) lev:(0) [131] conv:(1.08)
10. abastecimento=E destinoLixo=B fezesUrina=F 253909 ==> desencadeouInternamento=Nao 252220 <conf:(0.99)> lift:(1) lev:(0) [127] conv:(1.08)
11. fezesUrina=F 256811 ==> desencadeouInternamento=Nao 255102 <conf:(0.99)> lift:(1) lev:(0) [128] conv:(1.07)
12. abastecimento=E fezesUrina=F 254573 ==> desencadeouInternamento=Nao 252878 <conf:(0.99)> lift:(1) lev:(0) [126] conv:(1.07)
13. destinoLixo=B 275282 ==> desencadeouInternamento=Nao 273440 <conf:(0.99)> lift:(1) lev:(0) [127] conv:(1.07)
14. abastecimento=E destinoLixo=B 272685 ==> desencadeouInternamento=Nao 270857 <conf:(0.99)> lift:(1) lev:(0) [123] conv:(1.07)
15. abastecimento=E 273977 ==> desencadeouInternamento=Nao 272138 <conf:(0.99)> lift:(1) lev:(0) [121] conv:(1.07)
```

Fonte: Software Weka - versão 3.8.3 (2019).

A partir dos resultados, ficou claro que as regras descobertas tratam de atributos específicos, não tendo variação ao longo desses três (3) meses analisados. O experimento não foi realizado nos demais meses até setembro, pois de abril à agosto, o número de instâncias sofre alteração não significativa, o que não é possível aplicar ao mês de setembro, pois este apresenta um total de setenta e nove (79) instâncias, um valor consideravelmente abaixo das demais, que apresentam média de 305.780,3 instâncias, por este motivo, o experimento também foi realizado no mês de setembro:

FIGURA 33 - RESULTADOS DO EXPERIMENTO 3 APRIORI (set/2018)

```

Apriori
=====

Minimum support: 0.95 (75 instances)
Minimum metric <confidence>: 0.99
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 6
Size of set of large itemsets L(3): 4
Size of set of large itemsets L(4): 1

Best rules found:

1. codCBO=225125 79 ==> codTipoUnid=U 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. codTipoUnid=U 79 ==> codCBO=225125 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. encamAtendEspecialista=Nao 79 ==> codTipoUnid=U 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. codTipoUnid=U 79 ==> encamAtendEspecialista=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. desencadeouuInternamento=Nao 79 ==> codTipoUnid=U 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. codTipoUnid=U 79 ==> desencadeouuInternamento=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. encamAtendEspecialista=Nao 79 ==> codCBO=225125 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. codCBO=225125 79 ==> encamAtendEspecialista=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. desencadeouuInternamento=Nao 79 ==> codCBO=225125 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. codCBO=225125 79 ==> desencadeouuInternamento=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
11. desencadeouuInternamento=Nao 79 ==> encamAtendEspecialista=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
12. encamAtendEspecialista=Nao 79 ==> desencadeouuInternamento=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
13. codCBO=225125 encamAtendEspecialista=Nao 79 ==> codTipoUnid=U 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
14. codTipoUnid=U encamAtendEspecialista=Nao 79 ==> codCBO=225125 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
15. codTipoUnid=U codCBO=225125 79 ==> encamAtendEspecialista=Nao 79 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

```

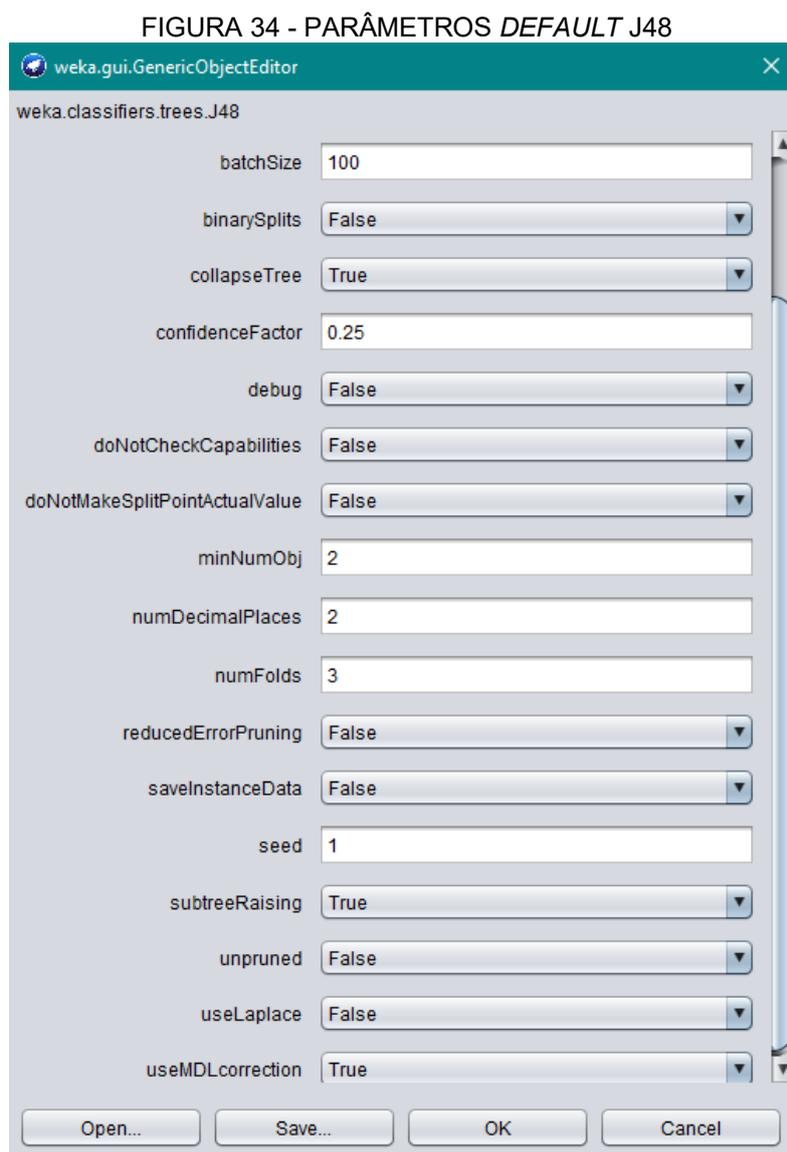
Fonte: Software Weka - versão 3.8.3 (2019).

Em setembro a ênfase está no suporte, 0,95 e na mudança de atributos, com o não aparecimento daqueles preenchidos apenas quando o cadastro é definitivo (abastecimento, energia elétrica, tratamento da água, entre outros). Quanto ao atributo do código CBO, refere-se ao médico clínico, portanto, uma das regras revela que se o atendimento for realizado por este profissional, não haverá internamento. De modo geral, é possível concluir que sim, há regras de associação descobertas nas bases, embora que não relacionadas com o atributo meta e semelhantes nos meses em que o algoritmo foi aplicado.

4.2.2 J48

O primeiro algoritmo de classificação é da heurística de *trees*, ou seja, árvore de decisão e trata-se do J48, de acordo com a própria definição do Weka “classe para gerar uma árvore de decisão C4.5 podada ou não”. Segundo Castro

e Ferrari (2016, p.170, grifo do autor) “uma *árvore de decisão (decision tree)* é uma estrutura em forma de árvore na qual cada *nó interno* corresponde a um teste de um atributo, cada *ramo* representa um resultado do teste e os *nós folha* representam classes ou distribuições de classes. O nó mais elevado da árvore é conhecido como *nós raiz*, e cada caminho da raiz até um nó da folha corresponde a uma *regra de classificação*”.



Fonte: Software Weka - versão 3.8.3 (2019).

Estas opções representam diferentes possibilidades de ajustes nos parâmetros antes da aplicação do algoritmo:

- seed - A semente usada para randomizar os dados quando a poda com erro reduzido é usada;

- unpruned - Se a poda é executada;
- confidenceFactor - O fator de confiança usado para a poda (valores menores requerem mais poda);
- numFolds - Determina a quantidade de dados usados para a poda com erro reduzido. Uma dobra é usada para a poda, o restante para o crescimento da árvore;
- numDecimalPlaces - O número de casas decimais a serem usadas para a saída de números no modelo;
- batchSize - O número preferencial de instâncias a serem processadas se a previsão em lote estiver sendo executada. Mais ou menos instâncias podem ser fornecidas, mas isso dá às implementações a chance de especificar um tamanho de lote preferencial;
- reducedErrorPruning - Se a poda com erro reduzido é usada em vez da poda do C.4.5;
- useLaplace - Se as contagens de folhas são suavizadas com base no Laplace;
- doNotMakeSplitPointActualValue - Se verdadeiro, o ponto de divisão não será realocado para um valor de dados real. Isso pode resultar em acelerações substanciais de grandes conjuntos de dados com atributos numéricos;
- debug - Se definido como true, o classificador pode gerar informações adicionais para o console;
- subtreeRaising - Se considerar a subtree aumentando a operação durante a poda;
- saveInstanceData - Se deseja salvar os dados de treinamento para visualização;
- binarySplits - Para usar divisões binárias em atributos nominais ao construir as árvores;
- doNotCheckCapabilities - Se definido, os recursos do classificador não serão verificados antes que o classificador seja criado (use com cuidado para reduzir o tempo de execução);
- minNumObj - O número mínimo de instâncias por folha;

- useMDLcorrection - Se a correção de MDL é usada ao encontrar divisões em atributos numéricos;
- collapseTree - Se as partes são removidas, isso não reduz o erro de treinamento.

Além das elucidações acerca dos parâmetros, Martinez, Casal e Janeiro (2009, p.5-6), explicam alguns conceitos básicos para a interpretação dos resultados e entendimento das métricas:

- Kappa Statistic: Índice que compara o valor encontrado nas observações com aquele que se pode esperar do acaso. É o valor calculado dos resultados encontrados nas observações e relatado como um decimal (0 a 1). Quanto menor o valor de Kappa, menor a confiança de observação, o valor 1 implica a correlação perfeita;
- Mean Absolute Error: média da diferença entre os valores atuais e os preditos em todos os casos, é a média do erro da predição;
- True Positives (TP): são os valores classificados verdadeiramente positivos;
- False Positives (FP): são os falsos positivos, são os dados classificados erroneamente como positivos pelo classificador;
- Precision (Precisão): É o valor da predição positiva (número de casos positivos por total de casos cobertos), muito influenciada pela especificidade e pouco pela sensibilidade. Sensibilidade é o número de casos positivos que são verdadeiramente positivos e especificidade é o número de casos negativos que são verdadeiramente negativos;
- Recall (Cobertura): É o valor da cobertura de casos muito influenciada pela sensibilidade e pouco pela especificidade. É calculada por número de casos cobertos pelo número total de casos aplicáveis;
- F-measure: Usada para medir o desempenho, pois combina valores de cobertura e precisão de uma regra numa única fórmula $[2 * \text{Prec} * \text{Rec} / (\text{Prec} + \text{Rec})]$;
- Root Relative Squared Error: reduz o quadrado do erro relativo na mesma dimensão da quantidade sendo predita incluindo raiz quadrada. Assim como a raiz quadrada do erro significativo (root mean-squared

error), este exagera nos casos em que o erro da predição foi significativamente maior do que o erro significativo;

- Relative Absolute Error: É o erro total absoluto. Em todas as mensurações de erro, valores mais baixos significam maior precisão do modelo, com o valor próximo de zero temos o modelo estatisticamente perfeito
- Root Mean-Squared Error: usado para medir o sucesso de uma predição numérica. Este valor é calculado pela média da raiz quadrada da diferença entre o valor calculado e o valor correto. O root mean-squared error é simplesmente a raiz quadrada do mean-squared-error (dá o valor do erro entre os valores atuais e os valores preditos);

O experimento com os parâmetros *default* apresentou os seguintes resultados:

FIGURA 35 - RESULTADOS DO EXPERIMENTO *DEFAULT* J48 (jan/2018)

```

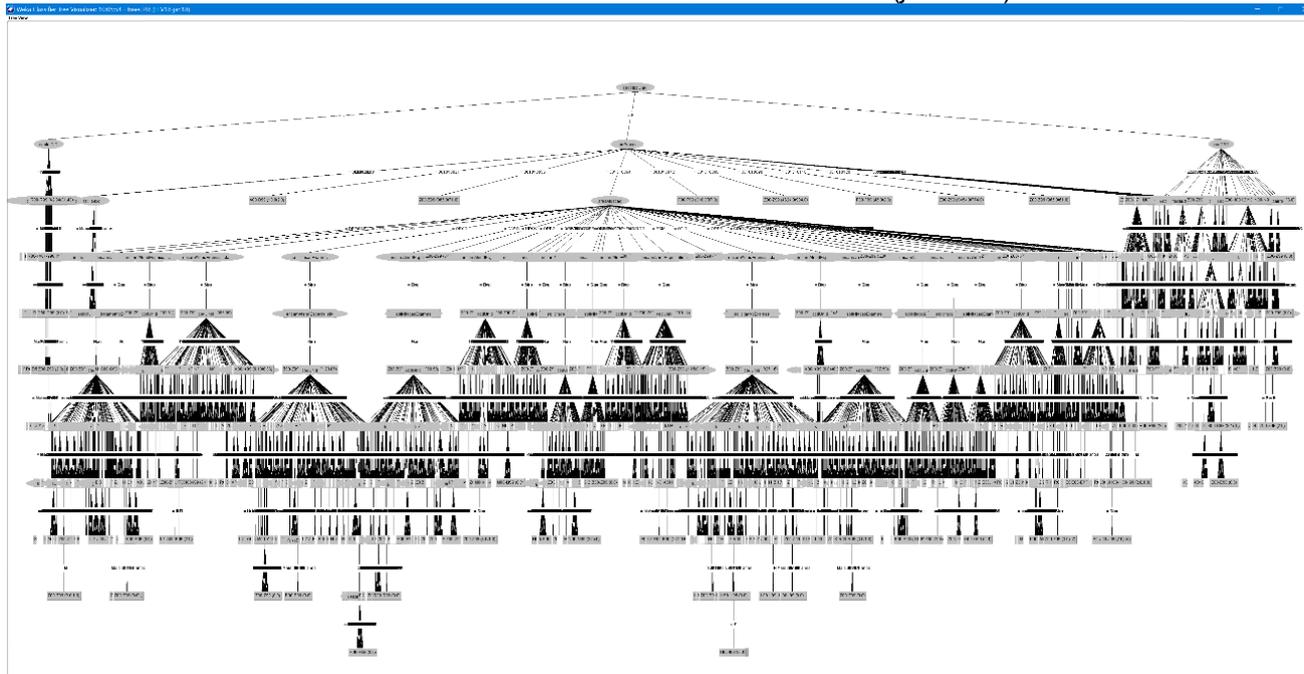
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      122974           46.0416 %
Incorrectly Classified Instances    144119           53.9584 %
Kappa statistic                     0.2504
Mean absolute error                 0.0679
Root mean squared error             0.1846
Relative absolute error             86.7423 %
Root relative squared error         93.3435 %
Total Number of Instances          267093

```

Fonte: Software Weka - versão 3.8.3 (2019).

Quanto à árvore de decisão, esta não apresenta boa visualização, como é possível ser verificado na FIGURA 36:

FIGURA 36 - ÁRVORE DE DECISÃO *DEFAULT* J48 (jan/2018)

Fonte: Software Weka - versão 3.8.3 (2019).

A partir do próximo experimento, optou-se por diminuir o número de validações cruzadas devido ao esgotamento da memória e processamento nos computadores utilizados; mesmo com o incremento de memória, quando há um elevado número de validações o software apresenta travamentos e logo na primeira tentativa de experimento após o *default*, em um computador com especificações mais modestas, ao alterar o parâmetro para 5-Fold o experimento não foi concluído por falta de memória, por este motivo, validação foi reduzida para 2-Fold e 4-Fold.

O experimento 2 foi realizado alterando a quantidade de *cross-validation*, inicialmente passando para 2-fold e alterando *reduceErrorPruning* para "TRUE", os resultados estão na FIGURA 37.

FIGURA 37 - RESULTADOS DO EXPERIMENTO 2 J48 (jan/2018)

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      122820           45.984 %
Incorrectly Classified Instances    144273           54.016 %
Kappa statistic                     0.2557
Mean absolute error                 0.0661
Root mean squared error             0.1834
Relative absolute error             84.5425 %
Root relative squared error         92.7205 %
Total Number of Instances          267093

```

Fonte: Software Weka - versão 3.8.3 (2019).

O terceiro experimento foi realizado alterando apenas o número de folds na validação cruzada para 4:

FIGURA 38 - RESULTADOS DO EXPERIMENTO 3 J48 (jan/2018)

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      124710           46.6916 %
Incorrectly Classified Instances    142383           53.3084 %
Kappa statistic                     0.2711
Mean absolute error                 0.065
Root mean squared error             0.182
Relative absolute error             83.0337 %
Root relative squared error         92.0404 %
Total Number of Instances          267093

```

Fonte: Software Weka - versão 3.8.3 (2019).

Por fim, o último experimento foi com 66% de percentage split, valor que representa o conjunto de treinamento, enquanto o restante (aproximadamente 34%) é destinado ao conjunto de testes:

FIGURA 39 - RESULTADOS DO EXPERIMENTO 4 J48 (jan/2018)

```

=== Summary ===

Correctly Classified Instances      42168            46.4344 %
Incorrectly Classified Instances    48644            53.5656 %
Kappa statistic                     0.2655
Mean absolute error                 0.0655
Root mean squared error             0.1825
Relative absolute error             83.647 %
Root relative squared error         92.2484 %
Total Number of Instances          90812

```

Fonte: Software Weka - versão 3.8.3 (2019).

Abaixo está TABELA 27 com a comparação entre os resultados dos 4 experimentos:

TABELA 27 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO J48 (jan/2018)

Sumário/Experimento	Default	2-Fold	4-Fold	66% split
Correctly Classified Instances	46,0416%	45,9840%	46,6916%	46,4344%
Incorrectly Classified Instances	53,9584%	54,0160%	53,3084%	53,5656%
Kappa Statistic	0,2504	0,2557	0,2711	0,2655
Mean absolute error	0,0679	0,0661	0,065	0,0655
Root mean squared error	0,1846	0,1834	0,182	0,1825
Relative absolute error	86,7423%	84,5425%	83,0337%	83,6470%
Root relative squared error	93,3435%	92,7205%	92,0404%	92,2484%
Tempo para construir o modelo (seconds)	110.62	61.66	142,34	11.44

Fonte: Elaborado pela autora a partir do software Weka - versão 3.8.3 (2019).

Acerca ao valor de instâncias classificadas corretamente, o melhor refere-se ao experimento 3, com 4-folds; quanto ao Kappa Statistic, em que o mais próximo de 1 indica a correlação perfeita, também é o índice mais alto no 4-fold; se o mean absolute error e o root mean squared error é baixo, significa menor erro na predição, como é caso do 4-fold;

Quanto ao relative absolute error e root relative squared error, os menores valores são aqueles que apresentam o menor erro, neste caso, o 4-fold. Por fim, no último valor analisado, o experimento com % de split é o mais rápido para processamento da base no Weka. Para o mês de fevereiro, optou-se por descartar os parâmetros *default* durante os experimentos, uma vez que os resultados não são satisfatórios, bem como a memória e capacidade de processamento do software não são altos; os demais experimentos foram mantidos:

TABELA 28 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO J48 (fev/2018)

Sumário/Experimento	Default	2-Fold	4-Fold	66% split
Correctly Classified Instances	-	46,7982%	47,5446%	47,3368%
Incorrectly Classified Instances	-	53,2018%	52,4554%	52,6632%
Kappa Statistic	-	0,2641	0,2794	0,2766
Mean absolute error	-	0,0652	0,0643	0,0645
Root mean squared error	-	0,1825	0,1811	0,1815
Relative absolute error	-	83,6163%	82,4225%	82,7837%
Root relative squared error	-	92,4430%	91,6952%	91,9546%
Tempo para construir o modelo (seconds)	-	131,31	142,26	10,11

Fonte: Elaborado pela autora a partir do software Weka - versão 3.8.3 (2019).

Assim como em janeiro, os resultados de fevereiro foram melhores para o experimento 3. A partir dos conjuntos de testes com parâmetros, é importante destacar que, além dos resultados apresentados, é necessário verificar se será adotada a validação cruzada ou particionamento da base e conforme explicam Castro e Ferrari (2016, p.157) “o objetivo da validação cruzada é, de forma sistemática, particionar a base de dados em conjuntos de treinamento e teste de modo que os dados de treinamento sejam usados para ajustar os parâmetros livres do modelo e os dados de teste sejam usados para fornecer uma estimativa de como o modelo vai generalizar para dados não usados no treinamento.[...] de maneira que permita que todos os dados sejam usados para treinamento e teste, mas em diferentes rodadas”, enquanto na opção percentage split, há apenas um conjunto de treinamento e um de teste, ambos variando de acordo com porcentagem adotada (66% é o default), dessa forma, foi considerado para os demais experimentos os parâmetros definidos no experimento 3 (reduceErrorPruning = TRUE e 4-Fold).

Após a aplicação do algoritmo em todos os meses analisados (janeiro a setembro de 2018), os resultados estão presentes na tabela abaixo:

TABELA 29 - COMPARATIVO DOS RESULTADOS ENTRE OS MESES ANALISADOS COM O ALGORITMO J48

Sumário/Mês	jan/18	fev/18	mar/18	abr/18	mai/18	jun/18	jul/18	ago/18	set/18
Correctly Classified Instances	46,6916%	47,5446%	46,6587%	47,4156%	49,1474%	50,1493%	48,6393%	48,7061%	18,9873%
Incorrectly Classified Instances	53,3084%	52,4554%	53,3413%	52,5844%	50,8526%	49,8507%	51,3607%	51,2939%	81,0127%
Kappa Statistica	0,2711	0,2794	0,2772	0,2946	0,3144	0,3252	0,3089	0,3053	-0,01
Mean absolute error	0,065	0,0643	0,0647	0,0612	0,06	0,0596	0,0632	0,0628	0,1408
Root mean squared error	0,182	0,1811	0,1816	0,1768	0,1749	0,1742	0,1796	0,1792	0,3072
Relative absolute error	83,0337%	82,4225%	82,4891%	81,7069%	80,7807%	80,4431%	80,8241%	80,3680%	97,7186%
Root relative squared error	92,0404%	91,6952%	91,7363%	91,3161%	90,7503%	90,5496%	90,7802%	90,6409%	114,9071%
Tempo para construir o modelo (seconds)	142,34	142,26	154,67	174,2	167,13	175,61	143,2	182,89	0,02

Fonte: Elaborado pela autora a partir do software Weka - versão 3.8.3 (2019).

A partir do exposto acima, é possível identificar em quais meses a classificação, com atributo meta capCID, obteve os melhores valores; ênfase para os meses de maio, junho, julho e agosto em que os resultados foram os mais satisfatórios, considerando os critérios já mencionados de instâncias classificadas corretamente, kappa statistic, mean absolute error, entre outros.

Analisando especificamente a acurácia por classe no mês de junho:

FIGURA 40 - ACURÁCIA DETALHADA POR CLASSE 4-FOLD J48 (jun/2018)

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,021	0,001	0,324	0,021	0,039	0,076	0,732	0,081	A00-B99
	0,837	0,383	0,537	0,837	0,654	0,434	0,817	0,720	Z00-Z99
	0,275	0,054	0,398	0,275	0,325	0,261	0,763	0,325	R00-R99
	0,117	0,007	0,470	0,117	0,187	0,216	0,790	0,224	M00-M99
	0,684	0,212	0,451	0,684	0,543	0,411	0,816	0,498	J00-J99
	0,038	0,001	0,350	0,038	0,069	0,112	0,766	0,081	S00-T98
	0,131	0,003	0,584	0,131	0,214	0,265	0,850	0,260	E00-E90
	0,142	0,002	0,620	0,142	0,231	0,291	0,718	0,177	L00-L99
	0,142	0,003	0,540	0,142	0,225	0,269	0,733	0,174	K00-K93
	0,200	0,004	0,671	0,200	0,308	0,354	0,817	0,294	F00-F99
	0,111	0,003	0,562	0,111	0,186	0,239	0,776	0,211	N00-N99
	0,163	0,006	0,587	0,163	0,255	0,292	0,844	0,308	I00-I99
	0,175	0,001	0,748	0,175	0,284	0,360	0,775	0,220	H00-H59
	0,005	0,000	0,211	0,005	0,009	0,031	0,772	0,017	V01-Y98
	0,126	0,001	0,518	0,126	0,203	0,252	0,759	0,149	G00-G99
	0,019	0,001	0,347	0,019	0,035	0,078	0,751	0,065	H60-H95
	0,101	0,000	0,453	0,101	0,166	0,213	0,757	0,138	C00-D48
	0,056	0,000	0,250	0,056	0,091	0,118	0,668	0,062	Q00-Q99
	0,039	0,000	0,522	0,039	0,072	0,142	0,847	0,107	O00-O99
	0,233	0,000	0,747	0,233	0,355	0,417	0,791	0,242	D50-D89
	0,035	0,000	0,600	0,035	0,066	0,145	0,831	0,077	P00-P96
	0,000	0,000	?	0,000	?	?	0,486	0,000	U04-U99
Weighted Avg.	0,501	0,183	?	0,501	?	?	0,801	0,458	

Fonte: Software Weka - versão 3.8.3 (2019).

O valor mais alto e neste caso o melhor para o *True Positive* (TP) está na classe Z00-Z99 com 83%, o que representa, conforme Rampao (2016, p.106) “uma probabilidade mais alta de um verdadeiro positivo. [...] Com isso, é possível identificar que o padrão detectado acerta mais vezes [...]” no caso desta classe. Este capítulo do CID refere-se aos “Fatores que influenciam o estado de saúde e o contato com os serviços de saúde”, sendo algumas de suas subcategorias: Z461 - Colocação e ajustamento de aparelho auditivo; Z480 - Cuidados a curativos e suturas cirúrgicas; Z503 - Reabilitação de toxicodependentes; Z52 - Doador de córnea; Z730 - Esgotamento, entre outros.

Segundo Rampao (2016, p.106) “Os resultados obtidos em *False Positive* (FP) indicam os dados classificados erroneamente como positivos”, sendo 38% também no capítulo Z00-Z99.

Quanto ao *Precision*, o maior valor é, aproximadamente, 75% para o capítulo H00-H59 ou Capítulo VII - Doenças do olho e anexos. Algumas de suas subcategorias são: H442 - Miopia degenerativa; H356 - Hemorragia retiniana; H401 - Glaucoma primário de ângulo aberto, entre outros.

Rampao (2016, p.106) explica que “O resultado do *recall* (cobertura) é equivalente aos valores obtidos no verdadeiro positivo (TP), considerando que corresponde ao valor da cobertura de casos”, ainda de acordo com a autora, valores de *F-Measure* relacionam-se ao desempenho, sendo que em junho, o melhor resultado é de 65% para Z00-Z99 e conforme o manual do Weka elaborado por Bouckaert et al. (2016, p.22) é “[...] uma medida combinada para precisão e cobertura”. Quanto aos resultados de *MCC*, *ROC Area* e *PRC*, conforme a autora, são relacionados ao desempenho, no caso, os melhores são 43% (Z00-Z99); 85% (E00-E90); 72% (Z00-Z99), respectivamente.

Por fim, temos a matriz de confusão:

FIGURA 41 - MATRIZ DE CONFUSÃO 4-FOLD J48 (jun/2018)

```

=== Confusion Matrix ===
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  r  s  t  u  v  <-- classified as
179 2980 739 41 4470 17 14 65 29 20 25 24 2 1 4 2 1 1 1 0 0 0 | a = A00-B99
95 91648 1572 480 13382 41 286 104 215 497 350 489 97 1 134 34 30 19 10 17 0 0 0 | b = Z00-Z99
45 9849 9932 352 14766 86 104 46 176 184 172 290 9 4 53 21 17 8 0 15 1 0 0 | c = R00-R99
18 7393 1634 1835 4423 28 85 19 29 46 56 118 2 4 12 7 5 2 1 0 0 0 0 | d = M00-M99
77 15093 4152 258 43881 70 85 49 28 97 86 164 22 3 32 52 6 1 0 0 0 0 0 | e = J00-J99
6 959 771 72 2251 165 5 43 4 9 8 14 2 1 4 3 2 1 1 0 0 0 0 0 | f = S00-T98
8 8556 357 176 231 5 1477 18 40 40 48 279 23 0 13 4 7 4 3 6 0 0 0 0 | g = E00-E90
30 2381 255 50 1755 10 21 771 79 9 18 44 3 0 4 8 6 0 0 1 0 0 0 0 | h = L00-L99
20 2499 927 75 2094 5 45 28 969 31 41 81 2 1 4 4 6 2 0 0 0 0 0 0 | i = K00-K93
13 7259 892 122 1025 10 90 15 24 2420 58 91 3 0 63 6 1 0 5 3 1 0 0 0 | j = F00-F99
16 5761 1477 128 2984 8 53 15 35 33 1326 63 4 0 7 1 12 1 1 1 0 0 0 0 | k = N00-N99
19 10372 1272 172 772 18 153 25 88 56 64 2539 20 0 28 13 0 4 0 2 0 0 0 0 | l = I00-I99
5 979 100 26 1558 1 29 5 4 6 9 38 589 0 6 1 0 1 0 0 0 0 0 0 0 | m = H00-H99
1 137 166 14 497 1 0 1 1 6 0 2 1 4 0 0 0 0 0 0 0 0 0 0 0 | n = V01-V98
4 1654 380 51 514 4 21 5 7 125 14 44 5 0 409 3 2 1 0 0 0 0 0 0 | o = G00-G99
8 1643 165 25 2571 1 9 5 3 14 6 13 1 0 4 85 2 0 0 0 0 0 0 0 0 | p = H60-H95
3 424 46 11 35 1 24 29 47 4 69 12 1 0 6 1 81 0 0 5 0 0 0 0 | q = C00-D48
2 146 27 14 37 0 6 0 8 1 3 6 1 0 4 0 0 15 0 0 0 0 0 0 0 | r = Q00-Q99
2 525 29 0 31 0 3 0 0 0 3 2 0 0 1 0 0 0 0 24 0 0 0 0 0 | s = O00-O99
1 362 31 4 47 0 9 0 7 8 3 12 0 0 2 0 1 0 0 148 0 0 0 0 0 | t = D50-D89
0 56 4 0 10 0 11 0 2 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 | u = P00-P96
0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | v = U04-U99

```

Fonte: Software Weka - versão 3.8.3 (2019).

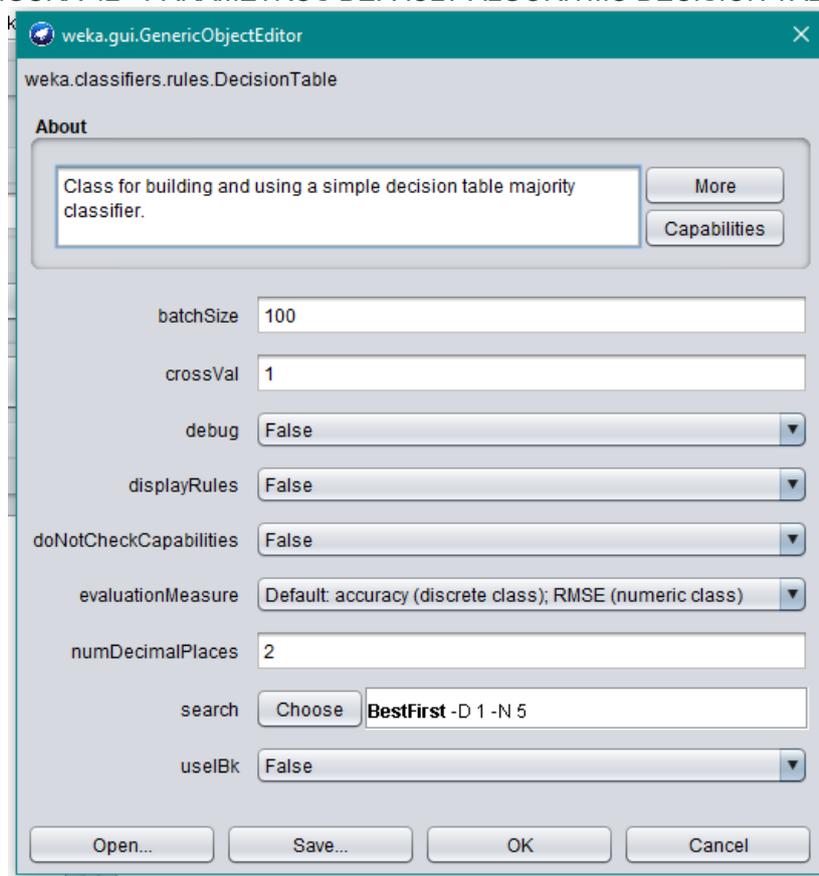
Como explica Rampao (2016, p.108), a matriz “possui como objetivo mostrar o número de previsões corretas em relação às esperadas para cada regra, sendo que na diagonal principal é possível identificar as instâncias que foram corretamente classificadas”, no caso acima, o maior número de instâncias classificadas erroneamente está na classe Z00-Z99, com um valor de 15.093 registros.

4.2.3 Decision Table

O Weka define o Decision Table como “Classe para construção e uso de um classificador de maioria de tabela de decisão simples”, trata-se de um algoritmo para a geração de regras.

Os parâmetros *default* são:

FIGURA 42 - PARÂMETROS *DEFAULT* ALGORITMO DECISION TABLE



Fonte: Software Weka - versão 3.8.3 (2019).

Além dessas configurações iniciais, o padrão também contempla 10 validações cruzadas. As opções são definidas da seguinte forma:

- numDecimalPlaces - O número de casas decimais a serem usadas para a saída de números no modelo;
- batchSize - O número preferencial de instâncias a serem processadas se a previsão em lote estiver sendo executada. Mais ou menos instâncias podem ser fornecidas, mas isso dá às implementações a chance de especificar um tamanho de lote preferencial;

- debug - Se definido como true, o classificador pode gerar informações adicionais para o console;
- doNotCheckCapabilities - Se definido, os recursos do classificador não serão verificados antes que o classificador seja criado (use com cuidado para reduzir o tempo de execução);
- evaluationMeasure - A medida usada para avaliar o desempenho das combinações de atributos usadas na tabela de decisão;
- search - O método de pesquisa usado para encontrar boas combinações de atributos para a tabela de decisão;
- displayRules - Define se as regras devem ser impressas;
- useIBk - Define se o IBk deve ser usado em vez da classe majoritária;
- crossVal - Define o número de dobras para validação cruzada (1 = deixa um para fora).

Os resultados mantendo os parâmetros no padrão foram:

FIGURA 43 - RESULTADOS EXPERIMENTO *DEFAULT* DECISION TABLE (jan/2018)

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      122008           45.68 %
Incorrectly Classified Instances    145085           54.32 %
Kappa statistic                    0.2622
Mean absolute error                 0.0712
Root mean squared error            0.1856
Relative absolute error             90.9765 %
Root relative squared error         93.8227 %
Total Number of Instances          267093

```

Fonte: Software Weka - versão 3.8.3 (2019).

No experimento *default* as regras não são apresentadas, o que impossibilita a análise além dos resultados numéricos obtidos, previamente informados na FIGURA 43.

Para uma melhor comparação, foram mantidos os critérios dos experimentos, ou seja, 2-Fold, 4-Fold e 66% *percentage split*. No experimento 1, também para a base de janeiro de 2018, o número de validações cruzadas foi alterado (para 2-Fold), assim como `displayRules` para TRUE e `evaluationMeasure` para Accuracy (discrete class):

FIGURA 44 - RESULTADOS EXPERIMENTO 1 DECISION TABLE (jan/2018)

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      120297      45.0394 %
Incorrectly Classified Instances    146796      54.9606 %
Kappa statistic                    0.2533
Mean absolute error                0.0723
Root mean squared error            0.1869
Relative absolute error            92.3801 %
Root relative squared error        94.4772 %
Total Number of Instances          267093

```

Fonte: Software Weka - versão 3.8.3 (2019).

Nesse experimento já foi possível verificar as regras, conforme FIGURA 45:

FIGURA 45 - REGRAS EXPERIMENTO 1 DECISION TABLE (jan/2018)

```

Rules:
-----
codUnid      codProced      codCBO      solicitacaoExames      encamAtendEspecialista      areaAtuacao      capIntCID      capCID
-----
2639394      301010110      225250      Nao                    Sim                          OBTH              ?              K00-K93
15466        301010072      225127      Nao                    Sim                          PHHP              ?              K00-K93
5587956      301010072      225285      Nao                    Sim                          URGI              ?              K00-K93
5587956      301010072      225250      Nao                    Sim                          OBTT              ?              Z00-Z99
17299        301010080      225124      Nao                    Sim                          PETR              ?              K00-K93
16969        301010064      225142      Nao                    Sim                          CILP              ?              K00-K93
3062224      301010064      225142      Nao                    Sim                          CILP              ?              M00-M99
3270785      301010072      225112      Nao                    Sim                          NELN              ?              K00-K93
5587956      301010072      225120      Nao                    Sim                          CICC              ?              I00-I99
5586372      301010072      225120      Nao                    Sim                          CICC              ?              I00-I99
5587956      301010072      225285      Nao                    Sim                          URGA              ?              N00-N99
3270785      301019940      225136      Nao                    Sim                          ACUM              ?              Z00-Z99
16675        301010129      225250      Nao                    Sim                          GENC              ?              K00-K93
17035        301010080      225124      Nao                    Sim                          AVUP              ?              Z00-Z99
17043        301010064      225124      Nao                    Sim                          AVUP              ?              K00-K93
16780        301010064      225124      Nao                    Sim                          AVUP              ?              K00-K93
17361        301010064      225124      Sim                    Sim                          AVUP              ?              Z00-Z99
17493        301010064      225124      Nao                    Sim                          ODOF              ?              Q00-Q99
16985        301010064      225142      Nao                    Sim                          GENC              ?              Z00-Z99
3298949      301010064      225142      Nao                    Sim                          GENC              ?              G00-G99

```

Fonte: Software Weka - versão 3.8.3 (2019).

Acima estão informadas apenas algumas regras (ponto de interrogação para vazio), cuja interpretação poderá ser: o paciente atendido pela unidade UMS Xapinhãl PSF (Sítio Cercado), que passou pelo procedimento 301010110 (consulta pré-natal), foi consultado pelo profissional identificado no CBO 225250 (Médico Ginecologista e Obstetra), não teve solicitação de exames, mas foi encaminhado para atendimento especialista com área de atuação em Obstetrícia - Hepatite, tem a classificação do CID no capítulo K00-K93 (Doenças do aparelho digestivo) e assim por diante.

No experimento 2, foram 4 validações realizadas (displayRules=TRUE e evaluationMeasure=Accuracy (discrete class)):

FIGURA 46 - RESULTADOS EXPERIMENTO 2 DECISION TABLE (jan/2018)
 === Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	121516	45.4958 %
Incorrectly Classified Instances	145577	54.5042 %
Kappa statistic	0.2595	
Mean absolute error	0.0714	
Root mean squared error	0.1858	
Relative absolute error	91.2357 %	
Root relative squared error	93.9576 %	
Total Number of Instances	267093	

Fonte: Software Weka - versão 3.8.3 (2019).

Por fim, o experimento foi com a alteração de validação cruzada para *percentage split*:

FIGURA 47 - RESULTADOS EXPERIMENTO 3 DECISION TABLE (jan/2018)
 === Evaluation on test split ===

Time taken to test model on test split: 0.76 seconds

=== Summary ===

Correctly Classified Instances	41277	45.4532 %
Incorrectly Classified Instances	49535	54.5468 %
Kappa statistic	0.2604	
Mean absolute error	0.0716	
Root mean squared error	0.186	
Relative absolute error	91.4974 %	
Root relative squared error	94.0197 %	
Total Number of Instances	90812	

Fonte: Software Weka - versão 3.8.3 (2019).

Para o mês de janeiro o comparativo entre os experimentos pode ser verificado na TABELA 30:

TABELA 30 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO DECISION TABLE (jan/2018)

Sumário/Experimento	Default	2-Fold	4-Fold	66% split
Correctly Classified Instances	45,6800%	45,0394%	45,4958%	45,4532%
Incorrectly Classified Instances	54,3200%	54,9606%	54,5042%	54,5468%
Kappa Statistic	0,2622	0,2533	0,2595	0,2604
Mean absolute error	0,0712	0,0723	0,0714	0,0716
Root mean squared error	0,1856	0,1869	0,1858	0,186
Relative absolute error	90,9765%	92,3801%	91,2357%	91,4974%
Root relative squared error	93,8227%	94,4772%	93,9576%	94,0197%
Número de regras	9777	9777	9777	9777
Tempo para construir o modelo (seconds)	166,37	337,84	369,6	354,44

Fonte: Software Weka - versão 3.8.3 (2019).

Apesar do melhor resultado no quesito instâncias classificadas corretamente, o experimento *default* tem como parâmetro `displayRules=FALSE`, ou seja, não apresenta regras, o que impossibilita uma melhor análise dos resultados, portanto, em fevereiro não será realizado. Quanto aos demais resultados, eles não sofreram muitas alterações, mas o experimento com 4-Fold se sobressaiu em relação aos outros (2-Fold e 66% *percentage split*).

Para fevereiro de 2018, os experimentos 1 a 3 foram novamente executados. Quando há 2 validações cruzadas os resultados são:

FIGURA 48 - RESULTADOS EXPERIMENTO 1 DECISION TABLE (fev/2018)

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      115703          45.4523 %
Incorrectly Classified Instances    138856          54.5477 %
Kappa statistic                    0.2535
Mean absolute error                 0.0721
Root mean squared error             0.1864
Relative absolute error              92.4495 %
Root relative squared error         94.3906 %
Total Number of Instances          254559

```

Fonte: Software Weka - versão 3.8.3 (2019).

No experimento 2 (4-Fold) os resultados foram:

FIGURA 49 - RESULTADOS EXPERIMENTO 2 DECISION TABLE (fev/2018)
Time taken to build model: 700.32 seconds

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      117008          45.965 %
Incorrectly Classified Instances    137551          54.035 %
Kappa statistic                     0.2613
Mean absolute error                 0.071
Root mean squared error             0.1852
Relative absolute error             91.0455 %
Root relative squared error         93.7859 %
Total Number of Instances          254559

```

Fonte: Software Weka - versão 3.8.3 (2019).

Por fim, no experimento 3, ao particionar a base de fevereiro de 2018, com 66% *percentage split*, os valores encontrados são os seguintes:

FIGURA 50 - RESULTADOS EXPERIMENTO 3 DECISION TABLE (fev/2018)
=== Summary ===

```

Correctly Classified Instances      39642          45.8024 %
Incorrectly Classified Instances    46908          54.1976 %
Kappa statistic                     0.2579
Mean absolute error                 0.0713
Root mean squared error             0.1856
Relative absolute error             91.4651 %
Root relative squared error         94.0095 %
Total Number of Instances          86550

```

Fonte: Software Weka - versão 3.8.3 (2019).

Ao comparar os resultados para o segundo mês, temos que a melhor porcentagem de instâncias foi observada no experimento 2, com 4-Fold; a respeito do Kappa Statistic, o maior valor, positivo para essa estatística, também está neste experimento, assim como as demais métricas:

TABELA 31 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS COM O ALGORITMO DECISION TABLE (fev/2018)

Sumário/Experimento	Default	2-Fold	4-Fold	66% split
Correctly Classified Instances	-	45,4523%	45,9650%	45,8024%
Incorrectly Classified Instances	-	54,5477%	54,0350%	54,1976%
Kappa Statística	-	0,2535	0,2613	0,2579
Mean absolute error	-	0,0721	0,071	0,0713
Root mean squared error	-	0,1864	0,1852	0,1856
Relative absolute error	-	92,4495%	91,0455%	91,4651%
Root relative squared error	-	94,3906%	93,7859%	94,0095%
Número de regras	-	25799	25799	25799
Tempo para construir o modelo (seconds)	-	560,54	700,32	663,53

Fonte: Software Weka - versão 3.8.3 (2019).

Os valores que indicam menor erro na predição são o mean absolute error e o root mean squared error, ambos com os resultados promissores quando o experimento foi realizado com 4 validações cruzadas. A respeito do relative absolute error e root relative squared error, as menores porcentagens representam resultados satisfatórios, o que também foi observado no experimento com 4-Fold. O número de regras é invariável e sobre o tempo de construção do modelo, quando são apenas 2 validações, a execução é mais rápida.

Em virtude dos melhores resultados atrelados às 4 validações, essa configuração foi aplicada também aos demais meses, de março a setembro de 2018, conforme comparativo abaixo:

TABELA 32 - COMPARATIVO DOS RESULTADOS DOS EXPERIMENTOS ENTRE OS MESES ANALISADOS COM O ALGORITMO DECISION TABLE

Sumário/Mês	jan/18	fev/18	mar/18	abr/18	mai/18	jun/18	jul/18	ago/18	set/18
Correctly Classified Instances	45,4958%	45,9650%	44,0638%	44,5261%	46,6330%	47,2688%	47,8026%	47,2866%	18,9873%
Incorrectly Classified Instances	54,5042%	54,0350%	55,9362%	55,4739%	53,3670%	52,7312%	52,1974%	52,7134%	81,0127%
Kappa Statística	0,2595	0,2613	0,2618	0,2713	0,291	0,3055	0,2994	0,2906	0,0133
Mean absolute error	0,0714	0,071	0,0769	0,0729	0,0718	0,0719	0,0704	0,0704	0,1483
Root mean squared error	0,1858	0,1852	0,1912	0,1862	0,1845	0,1845	0,1843	0,1842	0,2725
Relative absolute error	91,2357%	91,0455%	98,0335%	97,2923%	96,6309%	97,0888%	90,0171%	90,1295%	102,9472%
Root relative squared error	93,9576%	93,7859%	96,5540%	96,1977%	95,6916%	95,9031%	93,1691%	93,1691%	101,9213%
Número de regras	9777	25799	27840	29446	34760	33466	33370	35443	14
Tempo para construir o modelo (seconds)	369,6	700,32	824,71	778,73	1069,7	820,62	602,73	857,98	0,06

Fonte: Software Weka - versão 3.8.3 (2019).

Dentre estes meses, os melhores resultados estão em julho de 2018, portanto, analisando a acurácia por classe, temos:

FIGURA 51 - ACURÁCIA DETALHADA POR CLASSE 4-FOLD DECISION TABLE (jul/2018)
 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,867	0,397	0,548	0,867	0,672	0,454	0,793	0,696	Z00-Z99
	0,031	0,001	0,303	0,031	0,056	0,093	0,653	0,046	H60-H95
	0,568	0,190	0,351	0,568	0,434	0,316	0,783	0,365	J00-J99
	0,008	0,000	0,293	0,008	0,015	0,046	0,740	0,067	S00-T98
	0,081	0,001	0,511	0,081	0,141	0,199	0,668	0,107	L00-L99
	0,130	0,002	0,458	0,130	0,202	0,239	0,703	0,115	G00-G99
	0,149	0,008	0,418	0,149	0,220	0,232	0,825	0,225	E00-E90
	0,314	0,072	0,363	0,314	0,337	0,258	0,764	0,314	R00-R99
	0,109	0,011	0,392	0,109	0,171	0,182	0,750	0,184	M00-M99
	0,108	0,003	0,616	0,108	0,184	0,247	0,738	0,186	N00-N99
	0,133	0,001	0,677	0,133	0,222	0,297	0,723	0,161	H00-H59
	0,201	0,004	0,564	0,201	0,296	0,327	0,744	0,207	K00-K93
	0,006	0,000	0,311	0,006	0,013	0,041	0,707	0,069	A00-B99
	0,183	0,004	0,646	0,183	0,285	0,331	0,787	0,252	F00-F99
	0,172	0,007	0,572	0,172	0,264	0,294	0,801	0,277	I00-I99
	0,030	0,000	0,500	0,030	0,057	0,123	0,787	0,054	O00-O99
	0,053	0,000	0,389	0,053	0,093	0,143	0,759	0,050	Q00-Q99
	0,001	0,000	0,250	0,001	0,002	0,016	0,617	0,005	V00-Y98
	0,285	0,000	0,686	0,285	0,403	0,442	0,771	0,226	D50-D89
	0,120	0,000	0,508	0,120	0,194	0,246	0,809	0,125	C00-D48
	0,055	0,000	0,368	0,055	0,095	0,142	0,751	0,044	P00-P96
Weighted Avg.	0,478	0,181	0,475	0,478	0,416	0,324	0,773	0,409	

Fonte: Software Weka - versão 3.8.3 (2019).

Durante a análise da acurácia por classe, serão comentados apenas os resultados mais expressivos para cada métrica. Em *True Positive*, a maior porcentagem está na classe Z00-Z99, com 86%; ainda nesta classe, os falsos positivos, representantes dos dados classificados erroneamente, são 39%.

Quanto à precisão o melhor resultado encontra-se na classe D50-D89 (Capítulo III - Doenças do sangue e dos órgãos hematopoiéticos e alguns transtornos imunitários), são subcategorias deste capítulo: D532 (Anemia escorbútica), D573 (Estigma falciforme), D630 (Anemia em neoplasias), D70 (Agranulocitose), D891 (Crioglobulinemia), entre outras. O recall (ou cobertura) corresponde ao *True Positive*. As métricas relacionadas ao desempenho, F-Measure, MCC e PRC, tiveram os resultados de 67%, 45% e 69%, respectivamente, na classe Z00-Z99. Em relação ao ROC Area, também de desempenho, a maior porcentagem corresponde à classe C00-D48 (Capítulo II - Neoplasias [tumores]).

A respeito da matriz de confusão na FIGURA 52:

FIGURA 52 - MATRIZ DE CONFUSÃO 4-FOLD DECISION TABLE (jul/2018)

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	<-- classified as
92548	87	8562	18	130	126	585	1823	793	376	79	418	36	445	615	17	6	0	29	33	6	a = Z00-Z99
1801	128	1859	4	1	3	33	243	21	0	6	5	0	9	31	0	0	0	0	0	0	b = H00-H95
13851	62	25885	11	31	14	188	4933	279	28	12	20	15	99	105	0	2	0	0	5	2	c = J00-J99
1083	7	2625	41	18	4	24	1318	54	6	0	0	1	7	22	0	0	0	0	0	0	d = S00-T98
2669	3	1761	16	451	1	48	343	47	12	7	74	9	24	61	0	0	0	0	8	0	e = L00-L99
1594	6	627	1	3	448	51	438	55	2	4	9	0	158	55	0	0	0	0	2	1	f = G00-G99
8118	14	394	1	31	28	1730	482	277	53	24	67	5	72	312	0	0	0	15	1	0	g = E00-E90
9492	47	12143	10	53	74	252	10807	596	144	14	223	16	119	384	0	4	1	21	8	1	h = R00-R99
7175	10	5894	16	31	25	295	2340	1963	29	13	26	0	47	144	0	1	0	1	3	0	i = M00-M99
5322	4	3295	5	22	11	149	1773	330	1341	2	45	4	29	66	0	1	0	2	17	0	j = N00-N99
1270	5	1511	0	5	16	38	79	20	4	457	0	5	7	18	0	0	0	2	2	0	k = H00-H59
2446	9	2191	0	14	14	71	1212	77	7	5	1544	2	18	60	0	1	0	2	4	0	l = K00-K93
3284	5	4175	3	30	1	60	939	55	18	7	27	56	17	35	0	0	0	2	4	0	m = A00-B99
6869	14	1227	4	16	133	165	1063	154	36	11	23	13	2201	102	0	4	2	2	3	1	n = F00-F99
9609	13	994	5	20	74	355	1588	262	50	22	152	11	116	2759	0	3	0	13	3	0	o = I00-I99
471	0	30	0	0	2	34	1	3	0	0	1	0	1	17	0	0	0	0	1	0	p = O00-O99
145	4	11	2	3	3	8	13	10	7	4	5	2	7	26	0	14	0	0	0	0	q = Q00-Q99
150	0	580	0	0	0	1	268	0	0	3	0	0	6	0	0	0	1	0	0	0	r = V01-V98
375	0	31	0	2	3	24	30	13	0	1	14	2	5	17	0	0	0	208	4	0	s = D50-D89
432	4	33	3	21	1	36	33	5	61	2	84	1	5	8	0	0	0	6	100	0	t = C00-D48
56	0	16	0	0	0	20	6	0	0	2	4	1	14	2	0	0	0	0	0	7	u = P00-P96

Fonte: Software Weka - versão 3.8.3 (2019).

A maior quantidade de erros na classificação das instâncias também está na classe Z00-Z99 com um total de 13.851 inconsistências, neste caso, elas foram classificadas como a classe J00-J99.

Em relação a algumas das regras encontradas neste mês:

FIGURA 53 - REGRAS EXPERIMENTO 4-FOLD DECISION TABLE (jul/2018)

Rules:

faixaEtaria	codUnid	codProced	codCBO	solicitacaoExames	encamAtendEspecialista	areaAtuacao	capIntCID	capCID
35-39	3912698	301010064	225142	Nao	Sim	ODOR	?	Z00-Z99
45-49	6388671	301010072	225225	Nao	Sim	CRCP	?	Z00-Z99
70-74	5484456	301019940	225133	Nao	Sim	PSIP	?	Z00-Z99
60-64	6388671	301010072	225200	Sim	Sim	CIOM	?	R00-R99
50-54	6388671	301010072	225120	Sim	Sim	PNHP	?	Z00-Z99
0-4	6724299	301010064	225124	Sim	Sim	GACH	?	Z00-Z99
85-89	6388671	301010072	225155	Nao	Sim	CRCP	?	Z00-Z99
30-34	6388671	301010072	225225	Nao	Sim	CIOM	?	Z00-Z99
70-74	6388671	301010072	225225	Nao	Sim	CIOM	?	Z00-Z99
15-19	17531	301019940	225133	Nao	Sim	PSIA	?	Z00-Z99
35-39	5587956	301010072	225250	Nao	Sim	OBTH	?	Z00-Z99
20-24	5587956	301010072	225250	Nao	Sim	OBTH	?	Z00-Z99
60-64	3270785	301010072	225285	Nao	Sim	URGA	?	N00-N99
65-69	6388671	301010072	225120	Nao	Sim	PNIP	?	Z00-Z99
50-54	17477	301019940	225133	Nao	Sim	PSIP	?	Z00-Z99
20-24	17434	301019940	225133	Nao	Sim	PSIP	?	Z00-Z99
55-59	3270785	301010072	225155	Nao	Sim	CIOM	?	Z00-Z99
60-64	2639513	301010072	225125	Nao	Sim	HEPT	?	Z00-Z99
10-14	5586372	301019940	225133	Nao	Sim	PSII	?	Z00-Z99
20-24	17345	301010110	225250	Nao	Sim	OBTH	?	Z00-Z99
15-19	5506115	301019940	225133	Nao	Sim	PSII	?	Z00-Z99
10-14	5506115	301019940	225133	Nao	Sim	PSII	?	Z00-Z99
15-19	5484456	301019940	225133	Nao	Sim	PSII	?	Z00-Z99
15-19	3912698	301019940	225133	Nao	Sim	PSII	?	Z00-Z99
10-14	5506115	301010072	225133	Nao	Sim	PSII	?	Z00-Z99
0-4	17299	301010080	225124	Nao	Sim	PETR	?	Z00-Z99
55-59	17108	301010072	225133	Nao	Sim	PSIP	?	Z00-Z99

Fonte: Software Weka - versão 3.8.3 (2019).

A primeira regra pode ser interpretada da seguinte forma (capIntCID vazio): paciente na faixa etária de 35 a 39 anos, que busca atendimento na unidade 3912698 (UMS Mãe Curitibana), passa pelo procedimento 301010064 (Consulta Médica em Atenção Básica) com o profissional 225142 (Médico da Estratégia de Saúde da Família), não tem exames solicitados, é encaminhado para atendimento especialista (Odontologia - Cirurgia Ortognática e

Reconstrução), tende a ter o resultado de seu atendimento, o CID, no capítulo Z00-Z99 (Capítulo XXI - Fatores que influenciam o estado de saúde e o contato com os serviços de saúde). A interpretação pode ser realizada para as demais regras.

4.3 PÓS-PROCESSAMENTO

Conforme explicado anteriormente, basicamente, temos três macro etapas no processo de descoberta de conhecimento em bases de dados: pré-processamento, mineração de dados e pós-processamento, cujo objetivo, a grosso modo, a partir das análises, está no uso do conhecimento descoberto.

Este trabalho realizou análises e mineração em nove bases de dados, representando o período de janeiro a setembro de 2018 utilizando regras de associação, árvores de decisão e regras de classificação a partir do atributo meta definido como capCID.

Foi possível verificar que os resultados, apesar de superiores com o uso do algoritmo J48 tem a visualização prejudicada devido ao grande número de nós e folhas, impossibilitando completa análise, enquanto as regras permitem clareza.

No trecho da árvore abaixo, referente ao experimento com 4 validações no mês de junho, é possível verificar como seria a visualização e os resultados neste tipo de heurística. Na FIGURA 54, o trecho do resultado, com a folha em destaque:

FIGURA 54 - RESULTADO 4-FOLD J48 (jun/2018)

```

J48 pruned tree
-----
codTipoUnid = U
capIntCID = J00-J99
|
|   codProced = 101040024: J00-J99 (0.0)
|   codProced = 301010013: J00-J99 (0.0)
|   codProced = 301010021: J00-J99 (0.0)
|   codProced = 301010056: J00-J99 (0.0)
|   codProced = 301010064: J00-J99 (0.0)
|   codProced = 301010072: J00-J99 (0.0)
|   codProced = 301010080: J00-J99 (0.0)
|   codProced = 301010099: J00-J99 (0.0)
|   codProced = 301010110: J00-J99 (0.0)
|   codProced = 301010129: J00-J99 (0.0)
|   codProced = 301010137: J00-J99 (0.0)
|   codProced = 301019860: J00-J99 (0.0)
|   codProced = 301019940: J00-J99 (0.0)
|   codProced = 301019967: J00-J99 (0.0)
|   codProced = 301050023: J00-J99 (0.0)
|   codProced = 301060029
|   |
|   |   desencadeouInternamento = Nao
|   |   |
|   |   |   solicitacaoExames = Sim
|   |   |   |
|   |   |   |   faixaEtaria = 35-39
|   |   |   |   |
|   |   |   |   |   bairro = CIDADE INDUSTRIAL: R00-R99 (8.97/5.98)

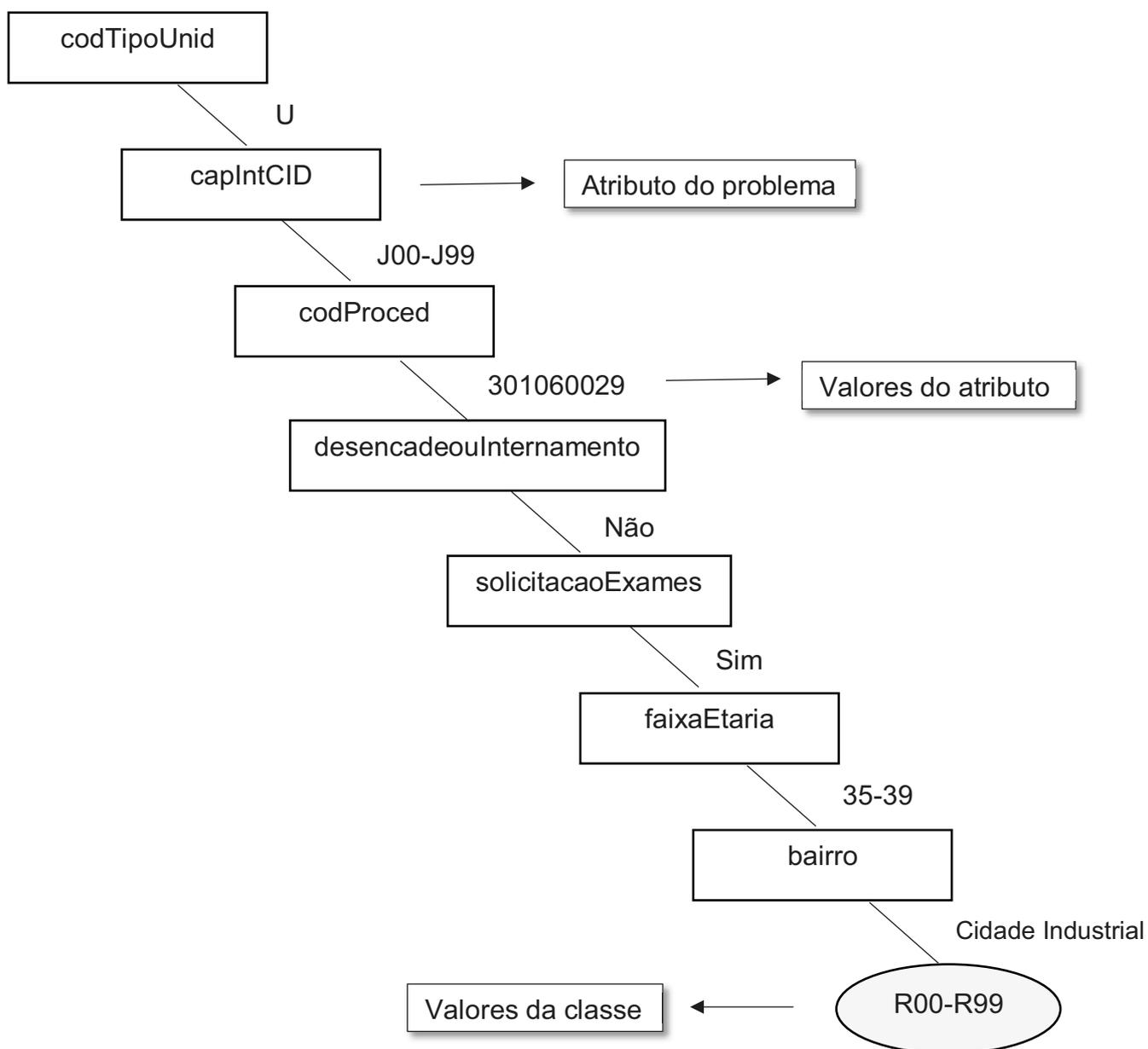
```

Fonte: Software Weka - versão 3.8.3 (2019).

Em seguida há um exemplo de como seria a árvore com os destaques explicados por Castro e Ferrari (2016, p.170, grifo do autor) em que “[...] cada *nó interno* corresponde a um teste em um atributo, cada *ramo* representa um resultado do teste e os *nós folhas* representam classes ou distribuições de

classes. O nó mais elevado da árvore é conhecido como *raiz*, e cada caminho da raiz até um nó folha corresponde à uma *regra de classificação*”.

FIGURA 55 - ÁRVORE J48 (jun/2018)



Fonte: Elaborado pela autora com base nos resultados do Software Weka - versão 3.8.3 e em Castro e Ferrari (2016).

Neste trecho, é possível observar com as regras geradas até a classe, esta noção de trajetória entre os resultados assemelha-se aos descobertos no Decision Table previamente exemplificados.

De modo geral, quando as regras estão explícitas, a interpretação ocorre de forma mais simplificada, portanto, mesmo com os valores satisfatórios do J48, para estas bases, a experiência com o Decision Table provou ser mais assertiva, principalmente na visualização e análise.

5 CONSIDERAÇÕES FINAIS

Conforme intensamente mencionado, a saúde faz parte de alguns dos temas centrais da vida humana, portanto, possui grande relevância; partindo deste princípio e de um tema específico da área, como o atendimento em unidades de saúde, atrelado aos conteúdos da graduação, serviram como base para o desenvolvimento deste trabalho.

A fim de considerar a pergunta de pesquisa (como identificar padrões no perfil de atendimento dos profissionais médicos da rede de saúde em Curitiba - PR por meio da análise de dados e aplicação de métodos de mineração de dados?) respondida e o objetivo geral (analisar os dados e aplicar métodos de Mineração de Dados a fim de identificar padrões que possibilitem a descoberta de conhecimento em base de dados do sistema “E-saúde” sobre o “Perfil de atendimento nas Unidades Municipais de Saúde de Curitiba”) atingido, é imprescindível analisar também os resultados relacionados aos objetivos específicos.

O primeiro objetivo específico consistiu na análise dos dados com a estatística descrita, o que foi realizado por meio da separação entre os atributos numéricos e nominais, de modo a aplicar cálculos específicos para cada um deles, os resultados foram apresentados e discutidos na seção de experimentos e resultados. Destaca-se que ao proceder com a estatística, é possível compreender os conjuntos de dados no âmbito de métricas com significados específicos (mínimo, máximo, desvio padrão, média, mediana, frequência absoluta, frequência relativa, entre outros), bem como a identificação de *outliers* e anomalias nas bases, neste âmbito, na análise dos atributos numéricos, foi indicada a presença de um grande número de valores discrepantes, portanto, foram descartados. Já nos atributos nominais, o destaque de resultados está na forte presença dos capítulos XXI - Fatores que influenciam o estado de saúde e o conta com os serviços de saúde (Z00-Z99; XVIII - sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte (R00-R99); e X - doenças do aparelho respiratório (J00-J99).

O segundo objetivo teve como meta a elaboração de visualizações sobre os conjuntos de dados, o que pode ser verificado nos dashboard elaborados na

ferramenta Power BI; este objetivo está muito atrelado ao anterior, pois foram utilizadas as frequências absolutas (contagens) dos atributos numéricos de todos os meses analisados, além da adição da visão dos resultados por sexo, ou seja, possibilidade de verificar dados sobre consultas, internamentos e aspectos sócio econômicos por gênero, o que fomenta novas oportunidades de análises e pode ser considerada uma indicação para trabalhos futuros.

O terceiro objetivo foi a seleção de algoritmos de mineração a fim de identificar padrões na base. Para este objetivo a compreensão dos dados é fundamental, assim como as possibilidades que os diferentes algoritmos representam, seja na visualização dos resultados, seja nas heurísticas escolhidas. Para este trabalho, optou por aplicar algoritmos de Regras de Associação (Apriori), Regras de Classificação (Decision Table) e Árvore de Decisão (J48).

Conforme presente na seção de experimentos e resultados, o Apriori gerou diversas regras, entretanto, foram pouco relevantes para o foco da pesquisa, mas, para os aspectos sócio econômicos, por exemplo, foram fundamentais, uma vez que as regras se referem, principalmente a eles. Um exemplo de regra, encontrada no experimento três (3) na base referente ao mês de março de 2018: se o abastecimento de água no domicílio do paciente for rede pública (opção E) então não desencadeará internamento (regra número 15).

Quanto aos algoritmos de classificação, ambos possuem características positivas e negativas para este trabalho, o J48, apesar de ter resultados mais expressivos, possui árvore que não permite completa análise, por outro lado, o Decision Table gerou regras de fácil interpretação, entretanto, muitas regras. Há um equilíbrio entre ambos. Um exemplo de regra de classificação, também mencionada na seção de experimentos e resultados: paciente na faixa etária de 35 a 39 anos, que busca atendimento na unidade 3912698 (UMS Mãe Curitibana), passa pelo procedimento 301010064 (Consulta Médica em Atenção Básica) com o profissional 225142 (Médico da Estratégia de Saúde da Família), não tem exames solicitados, é encaminhado para atendimento especialista (Odontologia - Cirurgia Ortognática e Reconstrução), tende a ter o resultado de seu atendimento, o CID, no capítulo Z00-Z99 (Capítulo XXI - Fatores que influenciam o estado de saúde e o contato com os serviços de saúde)

O quarto objetivo específico foi a comparação entre os resultados da mineração no espectro temporal de nove meses, o que também pode ser verificado na seção de experimentos e resultados, afinal, foram realizados experimentos com características semelhantes, de modo a verificar com clareza as divergências e similaridades entre os algoritmos de classificação. Os resultados satisfatórios, para o J48 e Decision Table, foram junho e julho de 2018, respectivamente.

Por fim, o último objetivo foi a classificação dos atendimentos com a base no CID, mais especificamente em seus capítulos e de acordo com o que foi informado ao final da análise descritiva, o atributo meta reflete o resultado do atendimento, portanto, 'capCID' foi definido para este papel. A classificação ocorreu com o J48 e Decision Table.

Quanto à tecnologia utilizada, conforme relatado, a escolha ocorreu por familiaridade com a ferramenta, entretanto, aconselha-se a pesquisa e uso de outros softwares, como a própria Linguagem R, Python, ferramentas para aplicação do algoritmo de colônia de formigas, entre outras, de modo a fomentar novas formas de descobertas de conhecimento.

Ao encerrar este trabalho, destaca-se que os objetivos foram atingidos, pois padrões, regras de classificação e comparativo entre os resultados dos meses foram realizados. A experiência com este desenvolvimento foi, no mínimo, desafiadora, o que possibilitou enriquecimento no aprendizado e utilização dos conhecimentos adquiridos durante a graduação.

Por ser tema relevante e com vasto assunto, no âmbito de trabalhos futuros, indica-se a continuação das análises em todas as bases, posterior comparação em diferentes anos, adição dos atributos numéricos relacionados à quantidade de medicamentos, maior proximidade com a Secretaria Municipal de Saúde, Instituto das Cidades Inteligentes e demais órgãos e departamentos relacionados, bem como a comparação e acréscimo de estudos sobre os números mundiais, nacionais, estaduais acerca da saúde, como forma de parâmetro para o estudo do perfil de atendimento.

Por fim, não há a pretensão de indicar a melhor forma de atuação da Secretaria Municipal de Saúde de Curitiba em sua rede de atenção, entretanto,

sugere-se que a partir dos resultados, ocorra verificação em profunda análise, de quais podem ser relevantes à elaboração de políticas públicas, uma vez que a partir deles torna-se possível identificar padrões no perfil de atendimento, o que possibilita conhecer a relevância no reconhecimento de quais pontos necessitam de aprimoramento para os pacientes, como os resultados identificados na visualização dos dados, que apontaram para predominância de doenças do aparelho respiratório, circulatório, fatores que influenciam o estado de saúde e o contato com serviços de saúde (também predominante na mineração), bem como sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte, portanto, a partir destes e de outros dados, o cenário poderá ser compreendido e até mesmo, ser um indicativo de quais áreas e/ou estabelecimentos a demanda é maior e necessita de mais investimentos. Diretamente com a mineração, ao encontrar um padrão perfil de atendimento permite-se verificar quais aspectos (como faixa etária, unidade consultada, profissional que realizou o atendimento, etc) influem sobre a Classificação Internacional de Doenças (CID), ou seja, quais atributos podem impactar na determinação da enfermidade, representando outras possibilidades de tomada de decisão na área da saúde em Curitiba.

REFERÊNCIAS

BARBOSA, R. R. Gestão da informação e do conhecimento: origens, polêmicas e perspectivas.. **Informação & Informação**, [S.l.], v. 13, n. 1esp, p. 1-25, dez. 2008. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1843>>. Acesso em: 30 mar. 2018.

BARROS CORREIA, M. F. de. **Recuperação de Informação**. 2018. Disponível em: <<https://oer.avu.org/handle/123456789/653>>. Acesso em: 21 maio 2018.

BEAL, A. **Gestão Estratégica da Informação**: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações. São Paulo: Atlas, 2004. 137 p.

BOUCKAERT, Remco R. et al. **WEKA Manual for Version 3-8-1**. New Zealand: S.i, 2016. Disponível em: <<https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/Curso18-19/WekaManual-3-8-1.pdf>>. Acesso em: 15 jun. 2019.

BRAMER, M. **Principles of data mining**. London: Springer, 2007. 342 p.

BRASIL. MINISTÉRIO DA SAÚDE. **Sistema Único de Saúde**. Disponível em: <<http://portalms.saude.gov.br/index.php/sistema-unico-de-saude/sistema-unico-de-saude>>. Acesso em: 09 jun. 2018.

BRASIL. DEPARTAMENTO DE INFORMÁTICA DO SUS. **Ministério lança aplicativo para ampliar o acesso da população às informações de saúde**. 2017. Disponível em: <<http://datasus.saude.gov.br/noticias/atualizacoes/1094-ministerio-lanca-aplicativo-para-ampliar-o-acesso-da-populacao-as-informacoes-de-saude>>. Acesso em: 20 jun. 2019.

CADASTRO Nacional de Estabelecimentos de Saúde. Disponível em: <<http://cnes.datasus.gov.br/pages/consultas.jsp>>. Acesso em: 15 jun. 2019.

CARDOSO, O. N. P. Recuperação de Informação. **INFOCOMP**, [S.l.], v. 2, n. 1, p. 33-38, nov. 2004. Disponível em: <<http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/46>>. Acesso em: 21 maio 2018.

CASTRO, L. N. de; FERRARI, D. G. **Introdução à Mineração de Dados**: Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016. 351 p.

CHAKRABARTI, S. et al. **Data mining**: know it all. Morgan Kaufmann, 2008.

COELI, C. M. Sistemas de Informação em Saúde e uso de dados secundários na pesquisa e avaliação em saúde. **Cad. Saúde Colet**, v. 18, n. 3, p. 335-6, 2010. Disponível

em:<http://www.cadernos.iesc.ufrj.br/cadernos/images/csc/2010_3/artigos/CSCv18n3_pag335-6.pdf>. Acesso em: 07 jun. 2018.

FERNEDA, E. **Recuperação de Informação**: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 147 f. Tese (Doutorado) - Curso de Ciências da Computação, Universidade de São Paulo, São Paulo, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230/en.php>>. Acesso em: 21 maio 2018.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

GUIMARÃES, P. R. B. **Métodos quantitativos estatísticos**. Curitiba: Iesde Brasil SA, v. 1, p. 252, 2008. Disponível em: <http://arquivostp.s3.amazonaws.com/qcursos/livro/LIVRO_metodos_quantitativos_estatisticos_paulo.pdf>. Acesso em: 15 jun.2018.

HAN, J.; PEI J.; KAMBER, M.; **Data mining**: concepts and techniques. [S.l]: Elsevier, 2011. 740 p.

LUNET, N., SEVERO, M., BARROS, H.. Desvio padrão ou erro padrão? Arquivos de medicina. Notas metodológicas. Serviços de Higiene e Epidemiologia da Faculdade de Medicina da Universidade do Porto. Vol. 20, N. 1/2. **ArquiMed**, 2006. Disponível em:<<https://docplayer.com.br/20871237-Desvio-padrao-ou-erro-padrao.html>>. Acesso em: 15 jun. 2018.

MAIMON, O.; ROKACH, L. **Data Mining and Knowledge Discovery Handbook**.2. ed. New York: Springer, 2010. 1306 p.

MARTINS, G. A. **Estatística geral e aplicada**. São Paulo: Atlas, 2006.

MARTINEZ, L.; CASAL, R.; JANEIRO, J. **Sistemas de apoio à decisão clínica**. Porto: Faculdade de Medicina do Porto, 2009. 16 p. Disponível em:<https://www.dcc.fc.up.pt/~ines/aulas/0910/MIM/trabs_ano_anterior/MIM3_SADC_2009%20Mar27%202106.doc>. Acesso em: 10 jun.2018.

MINISTÉRIO DA SAÚDE. Resolução n. 510 de 7 de abril de 2016. Acesso em: 15 jun. 2018. Disponível em: <http://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.htm>.

MOREIRA, C. B. **Gestão da informação**. 2015. Disponível em: <<http://repositorio.unicentro.br:8080/jspui/bitstream/123456789/962/5/Gest%C3%A3o%20da%20informa%C3%A7%C3%A3o.pdf>>. Acesso em: 30 mar. 2018.

ORACLE Data Types. Disponível em: <https://docs.oracle.com/cd/B28359_01/server.111/b28318/datatype.htm#CNCPT012>. Acesso em: 13 jun. 2018.

PAIM, J. S. et al. **O que é o SUS**. Rio de Janeiro: Fiocruz, 2015. 93 p. (Coleção Temas em Saúde). Disponível em: <<http://www.livrosinterativoseditora.fiocruz.br/sus/4/>>. Acesso em: 21 maio 2018.

Peternelli, L. & F., Fábio & Rocha, Rodrigo & Barros, Willian & Barbosa, Marcio. (2009). **Análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R**. *Bragantia*. 68 Disponível em: <https://www.researchgate.net/publication/250024070_Analise_dos_coeficientes_de_endogamia_e_de_parentesco_para_qualquer_nivel_de_ploidia_usando_o_pacote_estatistico_R>. Acesso em: 02 abr. 2019

PORTAL da Prefeitura de Curitiba: Dados Abertos Curitiba. Dados Abertos Curitiba. Disponível em: <<http://www.curitiba.pr.gov.br/dadosabertos/>>. Acesso em: 11 jun. 2018.

RAMPÃO, T. de S. **Mineração de Dados em Bases Jurídicas**: Um estudo de caso. 2016. 159 f. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2006. Disponível em: <<https://acervodigital.ufpr.br/handle/1884/45564>>. Acesso em: 15 jun. 2019.

REZENDE, S. O. et al. Mineração de dados. **Sistemas inteligentes: fundamentos e aplicações**, v. 1, p. 307-335, 2003. Disponível em: <http://www.addlabs.uff.br/enia_site/dw/mineracaodedados.pdf>. Acesso em: 13 jun. 2018.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p.41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 21 maio 2018.

SCLIAR, M. História do conceito de saúde. **Physis: Revista de saúde coletiva**, v. 17, p. 29-41, 2007. Disponível em: <<https://www.scielo.org/pdf/physis/2007.v17n1/29-41/pt>>. Acesso em: 13 jun. 2018.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de Banco de Dados**. 5. ed. Rio de Janeiro: Elsevier, 2006. 781 p.

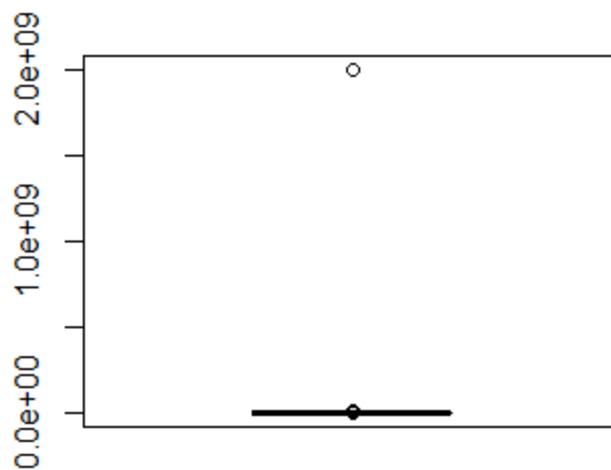
TEIXEIRA. B., E., A Análise de Dados na Pesquisa Científica. Importância e desafios em estudos organizacionais. **Desenvolvimento em Questão**, 2003, 1 (julho-dezembro). Acesso em: 15 de jun. 2018. Disponível em: <<http://www.redalyc.org/articulo.oa?id=75210209>> ISSN 1678-4855

VITAL, L. P.; FLORIANI, V. M.; VARVAKIS, G. Gerenciamento do fluxo de informação como suporte ao processo de tomada de decisão: revisão. **Informação & Informação**, Londrina, v. 15, n. 1, p.85-103, jul. 2010. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/5335>>. Acesso em: 14 maio 2018.

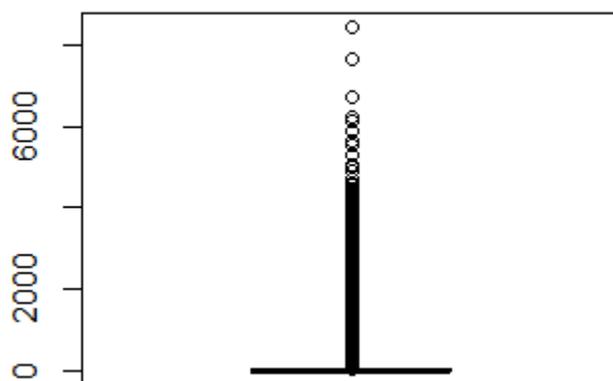
APÊNDICE A - Amostra de outros gráficos do tipo Boxplot (2FEV18)

BOXPLOT ATRIBUTO 'prescritaCuritiba' 02FEV18



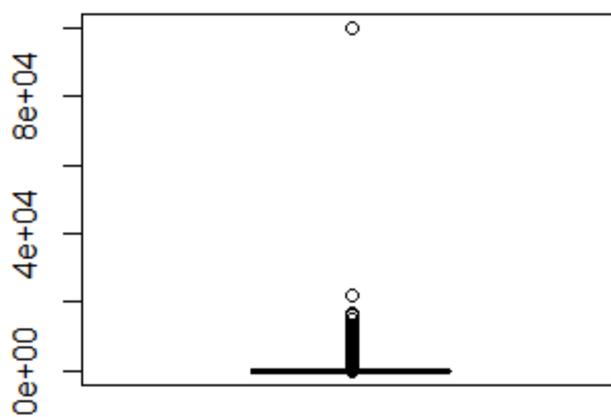
FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

BOXPLOT ATRIBUTO "DISPENSADACURITIBANA" 02FEV18



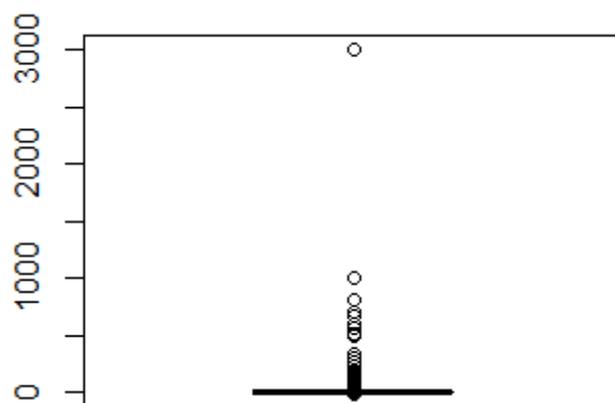
FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

BOXPLOT ATRIBUTO "MEDICAMENTONPADRONIZADO" 02FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

BOXPLOT ATRIBUTO "COMODOS" 02FEV18



FONTE: Elaborado pela autora no RStudio Versão 1.1.442 (2019).

APÊNDICE B - Códigos para atributos nominais

Códigos relacionados ao pré-processamento da base:

ÁREA DE ATUAÇÃO	CÓDIGO
ACUPUNTURA - MEDICO	ACUM
ALERGIA E IMUNOLOGIA	ALEI
AVALIAÇÃO CARDIOLOGIA	AVCR
AVALIAÇÃO COMPLEMENTAR DO APARELHO DIGESTIVO	AVAD
AVALIAÇÃO ENDOCRINOLOGIA PEDIÁTRICA	AVEP
AVALIAÇÃO MUSCULOESQUELÉTICA	AVML
AVALIAÇÃO NEUROLOGIA	AVNE
AVALIAÇÃO PARA CIRURGIA PLÁSTICA REPARADORA	AVCP
AVALIAÇÃO PEDIATRICA DO APARELHO DIGESTIVO	AVPA
AVALIAÇÃO PNEUMOLOGIA	AVPN
AVALIAÇÃO PSICOLOGIA - NASF	AVPS
AVALIAÇÃO PSIQUIATRIA - NASF	AVPQ
AVALIAÇÃO REUMATOLOGIA	AVRE
AVALIAÇÃO UROLOGIA PEDIÁTRICA	AVUP
CANCEROLOGIA GERAL	CANG
CANCEROLOGIA PEDIATRICA GERAL	CANP
CARDIOLOGIA - ARRITMIA	CARA
CARDIOLOGIA - INSUFICIÊNCIA CARDÍACA CONGESTIVA	CRIC
CARDIOLOGIA - MARCAPASSO	CRMR
CARDIOLOGIA - PRÉ OPERATÓRIO	CRPO
CARDIOLOGIA - PRÉ TRANSPLANTE	CRPT
CARDIOLOGIA - VALVULOPATIA	CARV
CARDIOLOGIA GERAL	CARG
CARDIOLOGIA PEDIATRICA	CARP
CIRURGIA CARDIOVASCULAR - CIRURGIA CARDÍACA	CICC
CIRURGIA DE CABEÇA E PESCOÇO GERAL	CRCP
CIRURGIA GERAL	CIRG
CIRURGIA GERAL - APARELHO DIGESTIVO	CIAD
CIRURGIA GERAL - MUTIRAO CIRURGIA DE HERNIA INGUINAL	CIHI
CIRURGIA GERAL - OBESIDADE MÓRBIDA	CIOM
CIRURGIA GERAL - PEQUENAS CIRURGIAS DE PELE	CICP
CIRURGIA GINECOLÓGICA	CIGI
CIRURGIA PEDIÁTRICA - CIRURGIA UROLÓGICA - FIMOSE	CIUF
CIRURGIA PEDIÁTRICA GERAL	CIPG
CIRURGIA PLÁSTICA - LIPODISTROFIA	CILP
CIRURGIA TORÁCICA GERAL	CITG

CIRURGIA VASCULAR - VARIZES	CIVV
CIRURGIA VASCULAR GERAL	CIVG
CLINICA GERAL - PARKINSON	CLGP
CONSULTA DE ENFERMAGEM AO OSTOMIZADO	CEFO
DERMATOLOGIA - HANSENOLOGIA	DRHN
DERMATOLOGIA GERAL	DERG
DOENÇAS RESPIRATÓRIAS	DORP
ENDOCRINOLOGIA - PEDIATRIA	EDPE
ENDOCRINOLOGIA GERAL	EDGE
GASTROENTEROLOGIA - COLESTASE NEONATAL	GACN
GASTROENTEROLOGIA - DOENÇAS DO FÍGADO	GADF
GASTROENTEROLOGIA GERAL	GAGE
GENETICA CLÍNICA	GENC
GERIATRIA GERAL	GERG
GINECOLOGIA - ANTICONCEPÇÃO	GINA
GINECOLOGIA - AVALIAÇÃO PARA COLPOSCOPIA	GINC
GINECOLOGIA - AVALIAÇÃO PARA LAQUEADURA	GINL
GINECOLOGIA - INCONTINÊNCIA URINÁRIA	GINU
GINECOLOGIA - INFERTILIDADE	GINI
GINECOLOGIA - PATOLOGIA CERVICAL	GINP
GINECOLOGIA - SEXOLOGIA	GINs
HEMATOLOGIA GERAL	HEMG
HEPATITES VIRAIS	HPVR
HEPATOLOGIA PRÉ-TRANSPLANTE	HEPT
HIV/AIDS	AIDS
HIV/HEPATITES/SIFILIS/TOXO	HHST
HOMEOPATIA GERAL	HOMG
INFECTOLOGIA - TUBERCULOSE	INFT
INFECTOLOGIA GERAL	INFG
MASTOLOGIA - BIRADS 0 E 4	MASB
MASTOLOGIA GERAL	MASG
MEDICINA DO ADOLESCENTE	MEDA
MEDICINA DO TRABALHO	MEDT
MEDICINA INTERNA - CLÍNICA GERAL	MEDG
NEFROLOGIA GERAL	NEFG
NEUROCIRURGIA GERAL	NEUG
NEUROLOGIA - PEDIATRIA EPILEPSIA	NELE
NEUROLOGIA - PEDIATRIA NEUROMUSCULAR	NELN
NEUROLOGIA - RECÉM NATO RISCO	NELR
NEUROLOGIA GERAL	NELG
NEUROLOGIA PEDIATRICA	NELP
NUTRIÇÃO - ATENÇÃO ESPECIALIZADA	NTAE
NUTRIÇÃO - LINHA DA OBESIDADE	NTLO
OBSTETRÍCIA - AVALIAÇÃO DE RISCO	OBTA

OBSTETRÍCIA - DOENÇA TROFOBLÁSTICA	OBDT
OBSTETRÍCIA - HEPATITE	OBTH
OBSTETRÍCIA - IMUNODEFICIÊNCIA	OBTI
OBSTETRÍCIA - RISCO	OBTR
OBSTETRÍCIA - TOXOPLASMOSE	OBTT
OBSTETRÍCIA GERAL	OBTG
ODONTOLOGIA - AMIGO ESPECIAL	ODOE
ODONTOLOGIA - AVALIAÇÃO DO FISSURADO E ANOMALIAS CRÂNIO FACIAIS	ODOF
ODONTOLOGIA - CIRURGIA BUCO MAXILO FACIAL	ODOB
ODONTOLOGIA - CIRURGIA ORTOGNÁTICA E RECONSTRUÇÃO	ODOR
ODONTOLOGIA - ONCOLOGIA	ODOO
OFTALMOLOGIA - AVALIAÇÃO TRANSPLANTE DE Córnea	OFTC
OFTALMOLOGIA - CATARATA	OFMC
OFTALMOLOGIA GERAL	OFMG
ORTOPEDIA - CIRURGIA DA MÃO	OREM
ORTOPEDIA - COLUNA	OREC
ORTOPEDIA - JOELHO	OREJ
ORTOPEDIA - MEMBRO SUPERIOR	ORES
ORTOPEDIA - PÉ E TORNOZELO	ORET
ORTOPEDIA - QUADRIL	OREQ
ORTOPEDIA - TRAUMA	OREU
ORTOPEDIA - TUMOR	ORTU
ORTOPEDIA GERAL	OREG
OTORRINO - DEFICIÊNCIA AUDITIVA	OTIA
OTORRINO - DEFICIÊNCIA AUDITIVA - ALTA COMPLEXIDADE	OTIC
OTORRINOLARINGOLOGIA - MUTIRÃO DE CIRURGIA DE OTORRINO	OTIO
OTORRINOLARINGOLOGIA GERAL	OTIG
PEDIATRIA - PUERICULTURA RISCO	PETR
PEDIATRIA - SÍNDROME DE DOWN	PETD
PNEUMOLOGIA - DOENÇA INTERSTICIAL PULMONAR	PNIP
PNEUMOLOGIA - HIPERTENSÃO PULMONAR	PNHP
PNEUMOLOGIA - OXIGENOTERAPIA	PNOT
PNEUMOLOGIA - TUBERCULOSE	PNTB
PNEUMOLOGIA GERAL	PNOG
PNEUMOLOGIA PEDIÁTRICA	PNLP
PROCTOLOGIA GERAL	PROG
PSICOLOGIA - PROCESSO TRANSEXUALIZADOR	PSLT
PSICOLOGIA GERAL	PSLG
PSIQUIATRIA - DEPENDÊNCIA QUÍMICA	PSIQ
PSIQUIATRIA - DEPENDÊNCIA QUÍMICA ADOLESCENTE	PSIA

PSIQUIATRIA - INFANTIL	PSII
PSIQUIATRIA - PSICOSES	PSIP
PSIQUIATRIA GERAL	PSIG
REUMATOLOGIA - PEDIATRIA	RETP
REUMATOLOGIA GERAL	RETG
UROLOGIA - ANDROLOGIA	URGA
UROLOGIA - INFERTILIDADE	URGI
UROLOGIA - LITÍASE URINÁRIA	URGU
UROLOGIA - NEOPLASIAS	URGN
UROLOGIA - VASECTOMIA	URGV
UROLOGIA GERAL	URGG

Códigos relacionados ao tratamento da base:

ESTABELECIMENTO SOLICITANTE	CÓDIGO
CENTRO DE ESPECIALIDADES SALGADO FILHO	CENESPSF
CENTRO DE ESPECIALIDADES VILA HAUER	CENESPVH
CLINICA DR. HELIO ROTENBERG	CLIDRHRO
CENTRO DE ESPECIALIDADES MEDICAS MATRIZ	CENESPMM
HOSPITAL SANTA CASA DE CURITIBA	HOSANTCC
HOSPITAL SAO VICENTE CIC	HOSPVVIC
UMS ABAETE	UMSABAET
UMS ABRANCHES	UMSABRAN
UMS ALVORADA PSF	UMSALVOR
UMS ATENAS	UMSATENA
UMS ATUBA	UMSATUBA
UMS AUGUSTA PSF	UMSAUGUS
UMS AURORA	UMSAUROR
UMS BAIRRO ALTO PSF	UMSBAALT
UMS BAIRRO NOVO PSF	UMSBANOV
UMS BARIGUI PSF	UMSBARIG
UMS BARREIRINHA	UMSBARRE
UMS BOM PASTOR PSF	UMSBOMPA
UMS BUTIATUVINHA PSF	UMSBUTIA
UMS CAIUA PSF	UMSCAIUA
UMS CAJURU	UMSCAJUR
UMS CAMARGO PSF	UMSCAMAR
UMS CAMPINA DO SIQUEIRA	UMSCASIQ
UMS CAMPO ALEGRE	UMSCAMAL
UMS CANDIDO PORTINARI PSF	UMSCANPT
UMS CAPANEMA PSF	UMSCAPAN
UMS CAXIMBA PSF	UMSCAXIM
UMS CONCORDIA PSF	UMSCONCO
UMS COQUEIROS PSF	UMSCOQUE

UMS DOM BOSCO PSF	UMSDOMBO
UMS ERICO VERISSIMO PSF	UMSERVER
UMS ESMERALDA	UMSESMER
UMS ESTRELA PSF	UMSESTRE
UMS EUCALIPTOS PSF	UMSEUCAL
UMS FANNY LINDOIA	UMSFANLI
UMS FERNANDO DE NORONHA	UMSFERNO
UMS IPIRANGA	UMSIPIRA
UMS IRACEMA	UMSIRACE
UMS IRMA TEREZA ARAUJO PSF	UMSIRMTA
UMS JARDIM GABINETO PSF	UMSJARGA
UMS JARDIM PARANAENSE PSF	UMSJARPA
UMS JOAO CANDIDO PSF	UMSJOACA
UMS LOTIGUACU PSF	UMSLOTIG
UMS MAE CURITIBANA PSF	UMSMAECU
UMS MARIA ANGELICA PSF	UMSMARIA
UMS MENONITAS	UMSMENON
UMS MONTEIRO LOBATO PSF	UMSMONLO
UMS MORADIAS BELEM PSF	UMSMORBE
UMS MORADIAS DA ORDEM PSF	UMSMOROR
UMS MORADIAS SANTA RITA PSF	UMSMORSR
UMS NOSSA SENHORA APARECIDA PSF	UMSNSEAP
UMS NOSSA SENHORA DA LUZ PSF	UMSNSELU
UMS NOSSA SENHORA DO SAGRADO CORACAO	UMSNSESC
UMS OSTERNACK PSF	UMSOSTER
UMS OSWALDO CRUZ PSF	UMSOSWCR
UMS OUVIDOR PARDINHO PSF	UMSOUVPA
UMS PALMEIRAS PSF	UMSPALME
UMS PANTANAL PSF	UMSPANTA
UMS PARIGOT DE SOUZA PSF	UMSPARSO
UMS PAROLIN PSF	UMSPAROL
UMS PARQUE INDUSTRIAL	UMSPARIN
UMS PILARZINHO	UMSPILAR
UMS POMPEIA PSF	UMSPOMPE
UMS RIO BONITO PSF	UMSRIOBO
UMS SABARA PSF	UMSSABAR
UMS SALGADO FILHO	UMSSALFI
UMS SALVADOR ALLENDE PSF	UMSSALAL
UMS SAMBAQUI PSF	UMSSAMBA
UMS SANTA CANDIDA	UMSSANCD
UMS SANTA EFIGENIA	UMSSANEF
UMS SANTA FELICIDADE	UMSSANFL
UMS SANTA QUITERIA II	UMSSQUII
UMS SANTA QUITERIA PSF	UMSSQUIP

UMS SANTOS ANDRADE PSF	UMSSANAN
UMS SAO DOMINGOS PSF	UMSSAODO
UMS SAO JOAO DEL REY PSF	UMSSAORE
UMS SAO JOSE PSF	UMSSAOJS
UMS SAO MIGUEL	UMSSAOMI
UMS SAO PAULO PSF	UMSSAOPA
UMS SAO PEDRO	UMSSAOPE
UMS SOLITUDE PSF	UMSSOLIT
UMS TAIZ VIVIANE MACHADO PSF	UMSTAVMA
UMS TANCREDO NEVES	UMSTANCR
UMS TAPAJOS	UMSTAPAJ
UMS TARUMA PSF	UMSTARUM
UMS TINGUI	UMSTINGU
UMS TRINDADE II PSF	UMSTRIII
UMS TRINDADE PSF	UMSTRIPS
UMS UBERABA DE CIMA	UMSUBECI
UMS UMBARA PSF	UMSUMBAP
UMS UNIAO DAS VILAS	UMSUNVIL
UMS VILA CLARICE	UMSVICLA
UMS VILA DIANA	UMSVDIAN
UMS VILA ESPERANCA PSF	UMSVILES
UMS VILA FELIZ	UMSVIFEL
UMS VILA GUAIRA	UMSVIGUA
UMS VILA HAUER	UMSVIHAU
UMS VILA LEAO	UMSVILEA
UMS VILA MACHADO	UMSVIMAC
UMS VILA SANDRA PSF	UMSVILSA
UMS VILA VERDE PSF	UMSVILVE
UMS VISITACAO	UMSVISIT
UMS VISTA ALEGRE	UMSVISTA
UMS VITORIA REGIA PSF	UMSVITRE
UMS WALDEMAR MONASTIER PSF	UMSWALMO
UMS XAPINHAL PSF	UMSXAPIN
UMS XAXIM	UMSXAXIM
UPA BOA VISTA	UPABOAVI
UPA BOQUEIRAO	UPABOQUE
UPA CAJURU	UPACAJUR
UPA CAMPO COMPRIDO	UPACAMCO
UPA FAZENDINHA	UPAFAZEN
UPA PINHEIRINHO	UPAPINHE
UPA SITIO CERCADO	UPASITCE
UPA TATUQUARA	UPATATUQ

Códigos relacionados ao tratamento da base:

ESTABELECIMENTO DESTINO	CÓDIGO
CENTRO MEDICO COMUNITARIO BAIRRO NOVO	CEMECOBN
CLINICA DR. HELIO ROTENBERG	CLIDRHRO
CRUZ VERMELHA BRASILEIRA - FILIAL DO ESTADO DO PARANA	CRVBFEP
FUNDAÇÃO HOSPITALAR PINHAIS	FUNHOPIN
HOSPITAL ANGELINA CARON - CAMPINA GRANDE DO SUL	HOSPANCA
HOSPITAL DA POLICIA MILITAR DO PARANÁ	HOSPMPR
HOSPITAL DE CLINICAS	HOSPCLIN
HOSPITAL DE DERMATOLOGIA SANITARIA - PIRAQUARA	HOSPDERM
HOSPITAL DO IDOSO ZILDA ARNS	HOSPIDZA
HOSPITAL DO TRABALHADOR	HOSPTRAB
HOSPITAL E MATERNIDADE SANTA MADALENA SOFIA	HOSPMAMS
HOSPITAL ERASTO GAERTNER	HOSPERGA
HOSPITAL ESPIRITA DE PSIQUIATRIA BOM RETIRO	HOSPEPBR
HOSPITAL INFANTIL PEQUENO PRINCIPE	HOSPIPPR
HOSPITAL INFANTIL WALDEMAR MONASTIER - CAMPO LARGO	HOSPWALM
HOSPITAL MATERNIDADE ALTO MARACANA - COLOMBO	HOSPMALT
HOSPITAL MENINO DEUS	HOSPMEDE
HOSPITAL MUNICIPAL DE ARAUCARIA	HOSPMUAR
HOSPITAL NOSSA SENHORA DO ROCIO - CAMPO LARGO	HOSPNSRO
HOSPITAL NOSSA SRA. DAS GRACAS - MATERNIDADE MATER DEI	HOSPMTMD
HOSPITAL OSVALDO CRUZ	HOSPOSCR
HOSPITAL REGIONAL DA LAPA SAO SEBASTIAO	HOSPRELS
HOSPITAL SANTA CASA DE CURITIBA	HOSANTCC
HOSPITAL SAO JOSE - SAO JOSE DOS PINHAIS	HOSPSJPI
HOSPITAL SAO LUCAS - CAMPO LARGO	HOSPSLUC
HOSPITAL SAO VICENTE	HOSPSAVI
HOSPITAL UNIVERSITARIO CAJURU	HOSPUNCA
HOSPITAL UNIVERSITARIO EVANGELICO DE CURITIBA - HUEC	HOSPUNEC
HOSPITAL VITOR DO AMARAL	HOSPVIAM
IMS	INSTMASO
OFTALMOCLINICA CURITIBA	OFTALMCU
UNIICA - UNIDADE INTERMEDIARIA DE CRISE E APOIO A VIDA	UNDITCVI

Códigos relacionados ao tratamento da base:

DESCRIÇÃO DA UNIDADE	CÓDIGO DA UNIDADE
CENTRO DE ESPECIALIDADE MEDICAS MATRIZ	15466
UMS CAPANEMA PSF	15474
UMS AURORA	16624
UMS BARIGUI PSF	16632
UMS ESTRELA PSF	16659
UMS FANNY LINDOIA	16667
UMS PAROLIN PSF	16675
UMS SABARA PSF	16683
UMS SANTA AMELIA	16691
UMS SANTA QUITERIA PSF	16705
UMS SÃO MIGUEL	16713
UMS TANCREDO NEVES	16721
UMS VILA CLARICE	16748
UMS VILA FELIZ	16756
UMS VILA GUAIRA	16764
UMS VILA LEO	16772
UMS UNIAO DAS VILAS	16780
UMS SANTOS ANDRADE PSF	16799
UMS BOM PASTOR PSF	16802
UMS JARIM GABINETO PSF	16810
UMS BUTIATUVINHA PSF	16829
UMS PINHEIROS PSF	16837
UMS SÃO BRAZ	16845
UMS SÃO JOSE PSF	16853
UMS NOVA ORLEANS	16861
UMS CAMPINA DO SIQUEIRA	16888
UMS AUGUSTA PSF	16896
UMS ATENAS	16918
UMS SANTA FELICIDADE	16926
UMS BAIRRO NOVO PSF	16934
UMS OSTERNACK PSF	16942
UMS CAXIMBA PSF	16950
UMS NOSSA SENHORA APARECIDA PSF	16969
UMS JOAO CANDIDO PSF	16977
UMS DOM BOSCO PSF	16985
UMS SÃO JOAO DEL REY PSF	16993
UMS SALVADOR ALLENDE PSF	17000
UMS PALMEIRAS PSF	17019
UMS PARIGOT DE SOUZA PSF	17027
UMS IRACENA	17035

UMS TRINDADE PSF	17043
UMS SOLITUDE PSF	17051
UMS SÃO DOMINGOS PSF	17078
UMS LOTIGUACU PSF	17086
UMS CAMARGO PSF	17094
UMS CAJURU	17108
UMS SÃO PAULO PSF	17116
UMS ESMERALDA	17124
UMS EUCALIPTOS PSF	17132
UMS MORADIAS BELEM PSF	17140
UMS IRMA TEREZA ARAUJO PSF	17159
UMS TAPAJOS	17167
UMS SÃO PEDRO	17175
UMS XAXIM	17183
UMS ERICO VERISSIMO PSF	17191
UMS WALDEMAR MONASTIER PSF	17205
UMS JARDIM PARANAENSE PSF	17213
UMS VILA HAUER	17221
UMS VISITACAO	17248
UMS CONCORDIA PSF	17256
UMS NOSSA SENHORA DO SAGRADO CORACAO	17264
UMS NOSSA SENHORA DA LUZ	17272
UMS MARIA ANGELICA PSF	17280
UMS IPIRANFA	17299
UMS OSWALDO CRUZ PSF	17302
UMS POMPEIA PSF	17310
UMS VILA MACHADO	17329
UMS PARQUE INDUSTRIAL	17337
UMS VILA VERDE PSF	17345
UMS MORADIAS SANTA RITA PSF	17353
UMS MORADIAS DA ORDEM PSF	17361
UMS TAIZ VIVIANE MACHADO PSF	17388
UPA BOA VISTA	17396
UMA VISTA ALEGRE	17418
UMS VILA DIANA	17426
UMS VILA ESPERANCA PSF	17434
UMS TINGUI	17442
UMS TARUMA PSF	17450
UMS SANTA EFIGENIA	17469
UMS SANTA CANDIDA	17477
UMS PILARZINHO	17485
UMS FERNANDO DE NORONHA	17493
UMS BARREIRINHA	17507
UMS BACACHERI	17515

UMS ATUBA	17523
UMS ABAETE	17531
UMS VILA LEONICE PSF	17558
UPA BOQUEIRAO	2438917
UMS BAIRRO ALTO PSF	2639378
UMS XPINHAL PSF	2639394
UMS UBERABA DE CIMA	2639408
UMS CANDIDO PORTINARI PSF	2639424
UMS CAIUA PSF	2639440
UPA FAZENDINHA	2639467
UMS COA CENTRO DE ORIENTACAO E ACONSELHAMENTO	2639513
UPA CAMPO COMPRIDO	2639548
UPA SITIO CERCADO	2639556
UMS OUVIDOR PARDINHO PSF	2639572
UMS SALGADO FILHO	2715899
CENTRO DE ESPECIALIDADES SANTA FELICIDADE	3270785
UMS UMBARA II PSF	3298949
UMS TRINDADE II PSF	3359646
UMS PANTANAL PSF	3507915
UMS DE ORTESE E PROTESE	3563413
UPA PINHEIRINHO	3827836
UMS UMBARA PSF	3862224
UMS MAE CURITIBANA PSF	3912698
UMS RIO BONITO PSF	5142725
UMS MONTEIRO LOBATO	5196809
UPA CAJURU	5323436
UMS VILA SANDRA PSF	5406617
UMS VITORIA REGIA PSF	5406624
UMS SAMBAQUI PSF	5446392
UMS SANTA QUITERIA II	5484456
UMS ALVORADA PSF	5506115
CENTRO DE ESPECIALIDADES OUVIDOR PARDINHO	5583225
CENTRO DE ESPECIALIDADES SALGADO FILHO	5584787
CENTRO DE ESPECIALIDADES VILA HAUER	5586372
CENTRO DE ESPECIALIDADES MAE CURITIBANA	5587956
UMS HIGIENOPOLIS	5701104
UMS MENONITAS	5709512
UMS MEDIANEIRA	5932238
UMS ABRANCHES	6724299
UMS COQUEIROS PSF	7686625
UMS CAMPO ALEGRE	9110402

UPA TATUQUARA	9214097
UMS JARDIM ALIANCA PSF	9381015

Códigos relacionados ao tratamento da base:

DESCRIÇÃO DO PROCEDIMENTO	CÓDIGO DO PROCEDIMENTO
AVALIACAO ANTROPOMETRICA	101040024
TONOMETRIA	211060259
CONSULTA AO PACIENTE CURADO DE TUBERCULOSE (TRATAMENTO SUPERVISIONADO)	301010013
CONSULTA C/ IDENTIFICACAO DE CASOS NOVOS DE TUBERCULOSE	301010021
CONSULTA MEDICA EM SAUDE DO TRABALHADOR	301010056
CONSULTA MEDICA EM ATENCAO BASICA	301010064
CONSULTA MEDICA EM ATENCAO ESPECIALIZADA	301010072
CONSULTA P/ ACOMPANHAMENTO DE CRESCIMENTO E DESENVOLVIMENTO (PUERICULTURA)	301010080
CONSULTA PARA AVALIACAO CLINICA DO FUMANTE	301010099
CONSULTA PRE-NATAL	301010110
CONSULTA PUERPERAL	301010129
CONSULTA/ATENDIMENTO DOMICILIAR	301010137
RESPOSTA TELERREGULAÇÃO/TELECONSULTORIA	301019860
TELERREGULAÇÃO/TELECONSULTORIA	301019940
CONSULTORIA EM NEUROLOGIA	301019967
ATENDIMENTO DE URGENCIA C/ OBSERVACAO ATE 24 HORAS EM ATENCAO ESPECIALIZADA	301060029
ATENDIMENTO DE URGENCIA EM ATENCAO BASICA	301060037
ATENDIMENTO DE URGENCIA EM ATENCAO BASICA COM OBSERVACAO ATE 8 HORAS	301060045
ATENDIMENTO DE URGENCIA EM ATENCAO BASICA COM REMOCAO	301060053
ATENDIMENTO MEDICO EM UNIDADE DE PRONTO ATENDIMENTO	301060096
ABORDAGEM COGNITIVA COMPORTAMENTAL DO FUMANTE (POR ATENDIMENTO / PACIENTE)	301080011

ASSISTENCIA AO PARTO SEM DISTOCIA	310010012
-----------------------------------	-----------

Códigos relacionados ao tratamento da base:

DESCRIÇÃO CBO	CÓDIGO CBO
MEDICO ACUPUNTURISTA	225105
MEDICO NEFROLOGISTA	225109
MEDICO NEUROLOGISTA	225112
MEDICO CARDIOLOGISTA	225120
MEDICO PEDIATRA	225124
MEDICO CLINICO	225125
MEDICO PNEUMOLOGISTA	225127
MEDICO PSIQUIATRA	225133
MEDICO DERMATOLOGISTA	225135
MEDICO REUMATOLOGISTA	225136
MEDICO DA ESTRATEGIA DE SAUDE DA FAMILIA	225142
MEDICO ENDOCRINOLOGISTA E METABOLOGISTA	225155
MEDICO GENERALISTA	225170
MEDICO HOMEOPATA	225195
MEDICO EM CIRURGIA VASCULAR	225203
MEDICO CIRURGIAO GERAL	225225
MEDICO CIRURGIAO PEDIATRICO	225230
MEDICO GINECOLOGISTA E OBSTETRA	225250
MEDICO MASTOLOGISTA	225255
MEDICO OFTALMOLOGISTA	225265
MEDICO UROLOGISTA	225285
MEDICO RESIDENTE	2231F9

APÊNDICE C - Lista de bairros encontrados nas bases

- ABRANCHES
- AGUA VERDE
- AHU
- ALTO BOQUEIRAO
- ALTO DA GLORIA
- ALTO DA RUA XV
- ATUBA
- AUGUSTA
- BACACHERI
- BAIRRO ALTO
- BAIRRO NOVO A
- BAIXADA
- BARREIRINHA
- BATEL
- BIGORRILHO
- BOA VISTA
- BOM RETIRO
- BOQUEIRAO
- BUTIATUVINHA
- CABRAL
- CACHOEIRA
- CAJURU
- CAMPINA DO SIQUEIRA
- CAMPO COMPRIDO
- CAMPO DE SANTANA
- CAPAO DA IMBUIA
- CAPAO RASO
- CASCATINHA
- CAXIMBA
- CENTRO
- CENTRO CIVICO
- CIC
- CIDADE INDUSTRIAL
- COLOMBO
- CRISTO REI
- FANNY
- FAZENDINHA
- GANCHINHO
- GUABIROTUBA
- GUAIRA
- HAUER
- HUGO LANGE
- JARDIM BOTANICO
- JARDIM DAS AMERICAS
- JARDIM SOCIAL
- JUVEVE
- LAMENHA PEQUENA
- LINDOIA
- MERCES
- MOSSUNGUE
- NAO INFORMADO
- NOVO MUNDO
- ORLEANS

- OUTROS
- PAROLIN
- PILARZINHO
- PINHEIRINHO
- PORTAO
- PRADO VELHO
- REBOUCAS
- RIVIERA
- SANTA CANDIDA
- SANTA FELICIDADE
- SANTA MARIA
- SANTA QUITERIA
- SANTO INACIO
- SAO BRAZ
- SAO FRANCISCO
- SAO JOAO
- SAO LOURENCO
- SAO MIGUEL
- SEMINARIO
- SITIO CERCADO
- SOLITUDE
- TABOAO
- TARUMA
- TATUQUARA
- TINGUI
- UBERABA
- UMBARA
- VILA IZABEL
- VILA NOVA
- VISTA ALEGRE
- XAXIM
- ZONA RURAL

APÊNDICE D - Portal de Dados Abertos

Tela inicial do Portal de Dados Abertos da Prefeitura de Curitiba (acesso em 26 de abril de 2019)

Portal da Prefeitura de Curitiba

CURITIBA
CIDADÃO
EMPRESA
SERVIDOR
INVESTIDOR
TURISTA
NOTÍCIAS

Dados Abertos Curitiba

Como parte de uma iniciativa muito maior de transparência e cidadania participativa, o Portal de Dados Abertos e Curitiba é uma ação da Prefeitura Municipal de Curitiba de disponibilização, através da internet, de documentos, informações e dados governamentais de domínio público para a livre utilização pela sociedade, garantindo à mesma, acesso aos dados primários, de forma que possam ser reutilizados produzindo novas informações e aplicações digitais para a sociedade.

Nesta primeira versão do Portal de Dados Abertos, são disponibilizadas bases de dados dos diversos órgãos do Governo Municipal de Curitiba para download. A cada mês, novas bases serão disponibilizadas em formato aberto para uso e edição irrestrita de seus usuários, sem a necessidade de assinatura de termo ou de identificação pessoal, com ou sem finalidade comercial.

Caso haja interesse por qualquer base que não esteja já disponível e que pelas legislações nacional e municipal possa ser disponibilizada, solicite-a através do e-mail: dadosabertos@sit.curitiba.pr.gov.br.

🔍

CONSULTAR

BASE DE DADOS

ENTRE EM CONTATO

SOLICITAÇÃO DE INFORMAÇÕES ADICIONAIS

Como usar o Dados Abertos Curitiba?

- 1 Conheça as bases de dados que estão catalogadas. Cada base vem acompanhada de breve descrição, detalhes de conteúdos e informações de formato, assim como contato do responsável por essa base na Prefeitura Municipal de Curitiba. Quaisquer ajustes ou solicitações para essas bases, poderão ser encaminhadas para o responsável por ela.
- 2 Todas as bases estarão disponíveis para download, bastando clicar sobre o formato desejado.
- 3 Você poderá se cadastrar para dar publicidade a produtos e estudos desenvolvidos por você, utilizando as bases do Portal de Dados Abertos de Curitiba. Para isso, entre em contato com dadosabertos@sit.curitiba.pr.gov.br.

Importante: Todos os dados disponibilizados pela Prefeitura Municipal de Curitiba através de seu Portal de Dados Abertos estão de acordo com as legislações nacional e municipal, assim como os princípios jurídicos aplicáveis. Caso observe algum dado em desacordo com esses aspectos, entre em contato com o administrador geral do Portal de Dados Abertos dadosabertos@sit.curitiba.pr.gov.br.

Dúvidas

- [O que são Dados Abertos?](#) ▾
- [O que é a Infraestrutura Nacional de Dados Abertos - INDA?](#) ▾
- [O que é um Mashup?](#) ▾
- [O que são metadados?](#) ▾
- [Qual a diferença deste portal para os portais de transparência?](#) ▾
- [O que são conjuntos de dados e recursos?](#) ▾
- [Posso utilizar as bases de dados para qualquer finalidade?](#) ▾
- [Como denuncio o uso das bases de dados para atividades ilícitas?](#) ▾
- [Posso cobrar por produtos, serviços ou estudos que desenvolver utilizando as bases de dados?](#) ▾
- [É cobrada alguma taxa ou licença para uso das bases de dados?](#) ▾
- [Preciso me identificar para ter acesso às bases?](#) ▾

Biblioteca

- Manual dos Dados Abertos para Desenvolvedores
- Cartilha para Desenvolvedores
- Política de Dados Abertos do Município de Curitiba
- Lei de Acesso à Informação
- Regras de Utilização

APÊNDICE E - TRECHO DO RESULTADO DO ALGORITMO J48 (JUN2018)

==== Run information ====

Scheme: weka.classifiers.trees.J48 -R -N 3 -Q 1 -M 2

Relation: AGO18-jun18

Instances: 316056

Attributes: 25

dtAtendimento

faixaEtaria

sexo

codTipoUnid

codUnid

codProced

codCBO

solicitacaoExames

encamAtendEspecialista

areaAtuacao

desencadeouInternamento

dtInternamento

estabSolicitante

estabDestino

capIntCID

tratamentoDomicilio

abastecimento

energiaEletrica

tipoHabitaçao

destinoLixo

fezesUrina

casoDoenca

grupoComunitario

bairro

capCID

Test mode: 4-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

codTipoUnid = U

| capIntCID = J00-J99

| | codProced = 101040024: J00-J99 (0.0)

| | codProced = 301010013: J00-J99 (0.0)

| | codProced = 301010021: J00-J99 (0.0)

| | codProced = 301010056: J00-J99 (0.0)

| | codProced = 301010064: J00-J99 (0.0)

| | codProced = 301010072: J00-J99 (0.0)

| | codProced = 301010080: J00-J99 (0.0)

| | codProced = 301010099: J00-J99 (0.0)

| | codProced = 301010110: J00-J99 (0.0)

| | codProced = 301010129: J00-J99 (0.0)

| | codProced = 301010137: J00-J99 (0.0)
 | | codProced = 301019860: J00-J99 (0.0)
 | | codProced = 301019940: J00-J99 (0.0)
 | | codProced = 301019967: J00-J99 (0.0)
 | | codProced = 301050023: J00-J99 (0.0)
 | | codProced = 301060029
 | | | desencadeouInternamento = Nao
 | | | | solicitacaoExames = Sim
 | | | | | faixaEtaria = 35-39
 | | | | | bairro = CIDADE INDUSTRIAL: R00-R99 (8.97/5.98)
 | | | | | bairro = BAIRRO ALTO: R00-R99 (1.74/0.5)
 | | | | | bairro = PINHEIRINHO : R00-R99 (2.49/1.99)
 | | | | | bairro = NOVO MUNDO: R00-R99 (1.49/0.75)
 | | | | | bairro = TATUQUARA : R00-R99 (3.99/3.24)
 | | | | | bairro = CACHOEIRA : Z00-Z99 (0.25)
 | | | | | bairro = NAO INFORMADO: R00-R99 (1.0)
 | | | | | bairro = UMBARA: R00-R99 (0.75)
 | | | | | bairro = SANTA FELICIDADE: R00-R99 (2.24/1.0)
 | | | | | bairro = PORTAO : N00-N99 (1.0/0.5)
 | | | | | bairro = XAXIM : R00-R99 (2.49/1.74)
 | | | | | bairro = BOQUEIRAO : R00-R99 (2.99/1.25)
 | | | | | bairro = UBERABA: R00-R99 (3.74/1.49)
 | | | | | bairro = CAPAO RASO: R00-R99 (0.5/0.25)
 | | | | | bairro = VISTA ALEGRE : J00-J99 (0.25)
 | | | | | bairro = HUGO LANGE : R00-R99 (0.0)
 | | | | | bairro = ABRANCHES : G00-G99 (0.25)

| | | | | bairro = SITIO CERCADO: R00-R99 (8.22/4.48)
 | | | | | bairro = SANTA CANDIDA : R00-R99 (1.25/0.5)
 | | | | | bairro = PILARZINHO: F00-F99 (1.74/1.0)
 | | | | | bairro = CAPAO DA IMBUIA: L00-L99 (1.25/0.75)
 | | | | | bairro = AHU : R00-R99 (0.0)
 | | | | | bairro = GUAIRA: J00-J99 (0.5/0.25)
 | | | | | bairro = FAZENDINHA: J00-J99 (0.75/0.25)
 | | | | | bairro = SAO BRAZ : R00-R99 (1.49/0.75)
 | | | | | bairro = CAMPO DE SANTANA : R00-R99 (3.74/1.0)
 | | | | | bairro = ALTO BOQUEIRAO: R00-R99 (1.74/0.5)
 | | | | | bairro = PAROLIN : N00-N99 (0.25)
 | | | | | bairro = BARREIRINHA : R00-R99 (1.74/0.25)
 | | | | | bairro = CAXIMBA : J00-J99 (0.25)
 | | | | | bairro = REBOUCAS : R00-R99 (0.25)
 | | | | | bairro = CAJURU
 | | | | | | sexo = F: J00-J99 (4.98/3.24)
 | | | | | | sexo = M: R00-R99 (4.23/2.99)
 | | | | | bairro = TARUMA : R00-R99 (0.0)
 | | | | | bairro = CAMPO COMPRIDO : N00-N99 (0.25)
 | | | | | bairro = GUABIROTUBA: R00-R99 (0.25)
 | | | | | bairro = ATUBA : R00-R99 (0.0)
 | | | | | bairro = CENTRO : R00-R99 (0.75/0.5)
 | | | | | bairro = AUGUSTA : R00-R99 (0.5/0.25)
 | | | | | bairro = PRADO VELHO : R00-R99 (0.0)
 | | | | | bairro = JUVEVE : R00-R99 (0.0)
 | | | | | bairro = LINDOIA : J00-J99 (0.25)

	bairro = CAMPINA DO SIQUEIRA	: R00-R99 (0.75)
	bairro = TINGUI	: I00-I99 (0.25)
	bairro = BIGORRILHO:	J00-J99 (0.75/0.25)
	bairro = BUTIATUVINHA:	R00-R99 (0.5/0.25)
	bairro = BACACHERI:	R00-R99 (0.75/0.5)
	bairro = SAO FRANCISCO	: F00-F99 (0.25)
	bairro = JARDIM BOTANICO	: R00-R99 (0.0)
	bairro = HAUER	: R00-R99 (0.75)
	bairro = BOA VISTA	: R00-R99 (0.75/0.25)
	bairro = JARDIM DAS AMERICAS	: J00-J99 (0.25)
	bairro = SANTO INACIO	: F00-F99 (0.25)
	bairro = CASCATINHA	: R00-R99 (0.0)
	bairro = SAO JOAO	: R00-R99 (0.25)
	bairro = VILA IZABEL	: R00-R99 (0.0)
	bairro = SANTA QUITERIA	: R00-R99 (0.5/0.25)
	bairro = AGUA VERDE:	F00-F99 (0.25)
	bairro = GANCHINHO	: R00-R99 (0.5/0.25)
	bairro = SAO LOURENCO	: R00-R99 (0.0)
	bairro = MOSSUNGUE	: R00-R99 (0.0)
	bairro = CABRAL	: F00-F99 (0.25)
	bairro = CENTRO CIVICO	: R00-R99 (0.0)
	bairro = MERCES	: J00-J99 (0.25)
	bairro = FANNY	: J00-J99 (0.25)
	bairro = SEMINARIO	: R00-R99 (0.25)
	bairro = BOM RETIRO	: R00-R99 (0.0)
	bairro = ORLEANS	: R00-R99 (0.5/0.25)

	bairro = ALTO DA GLORIA	: R00-R99 (0.25)
	bairro = NAO INFORMADO	: R00-R99 (0.0)
	bairro = CRISTO REI	: R00-R99 (0.75)
	bairro = BATEL	: R00-R99 (0.0)
	bairro = SAO MIGUEL	: R00-R99 (0.0)
	bairro = TABOAO	: R00-R99 (0.0)
	bairro = ALTO DA RUA XV	: Z00-Z99 (0.5/0.25)
	bairro = JARDIM SOCIAL	: R00-R99 (0.0)
	bairro = LAMENHA PEQUENA	: R00-R99 (0.0)
	bairro = RIVIERA	: R00-R99 (0.0)
	bairro = SANTA MARIA	: R00-R99 (0.0)
	bairro = OUTROS	: R00-R99 (0.0)
	bairro = SOLITUDE	: R00-R99 (0.0)
	bairro = ZONA RURAL	: R00-R99 (0.0)
	bairro = CIC	: R00-R99 (0.0)
	faixaEtaria = 5-9	
	bairro = CIDADE INDUSTRIAL	
	codUnid = 15466: J00-J99 (0.0)	
	codUnid = 15474: J00-J99 (0.0)	
	codUnid = 16624: J00-J99 (0.0)	
	codUnid = 16632: J00-J99 (0.0)	
	codUnid = 16659: J00-J99 (0.0)	
	codUnid = 16667: J00-J99 (0.0)	
	codUnid = 16675: J00-J99 (0.0)	
	codUnid = 16683: J00-J99 (0.0)	
	codUnid = 16691: J00-J99 (0.0)	

| | | | | | | codUnid = 16705: J00-J99 (0.0)
| | | | | | | codUnid = 16713: J00-J99 (0.0)
| | | | | | | codUnid = 16721: J00-J99 (0.0)
| | | | | | | codUnid = 16748: J00-J99 (0.0)
| | | | | | | codUnid = 16756: J00-J99 (0.0)
| | | | | | | codUnid = 16764: J00-J99 (0.0)
| | | | | | | codUnid = 16772: J00-J99 (0.0)
| | | | | | | codUnid = 16780: J00-J99 (0.0)
| | | | | | | codUnid = 16799: J00-J99 (0.0)
| | | | | | | codUnid = 16802: J00-J99 (0.0)
| | | | | | | codUnid = 16810: J00-J99 (0.0)
| | | | | | | codUnid = 16829: J00-J99 (0.0)
| | | | | | | codUnid = 16837: J00-J99 (0.0)
| | | | | | | codUnid = 16845: J00-J99 (0.0)
| | | | | | | codUnid = 16853: J00-J99 (0.0)
| | | | | | | codUnid = 16861: J00-J99 (0.0)
| | | | | | | codUnid = 16888: J00-J99 (0.0)
| | | | | | | codUnid = 16896: J00-J99 (0.0)
| | | | | | | codUnid = 16918: J00-J99 (0.0)
| | | | | | | codUnid = 16926: J00-J99 (0.0)
| | | | | | | codUnid = 16934: J00-J99 (0.0)
| | | | | | | codUnid = 16942: J00-J99 (0.0)
| | | | | | | codUnid = 16950: J00-J99 (0.0)
| | | | | | | codUnid = 16969: J00-J99 (0.0)
| | | | | | | codUnid = 16977: J00-J99 (0.0)
| | | | | | | codUnid = 16985: J00-J99 (0.0)

| | | | | | | codUnid = 16993: J00-J99 (0.0)
| | | | | | | codUnid = 17000: J00-J99 (0.0)
| | | | | | | codUnid = 17019: J00-J99 (0.0)
| | | | | | | codUnid = 17027: J00-J99 (0.0)
| | | | | | | codUnid = 17035: J00-J99 (0.0)
| | | | | | | codUnid = 17043: J00-J99 (0.0)
| | | | | | | codUnid = 17051: J00-J99 (0.0)
| | | | | | | codUnid = 17078: J00-J99 (0.0)
| | | | | | | codUnid = 17086: J00-J99 (0.0)
| | | | | | | codUnid = 17094: J00-J99 (0.0)
| | | | | | | codUnid = 17108: J00-J99 (0.0)
| | | | | | | codUnid = 17116: J00-J99 (0.0)
| | | | | | | codUnid = 17124: J00-J99 (0.0)
| | | | | | | codUnid = 17132: J00-J99 (0.0)
| | | | | | | codUnid = 17140: J00-J99 (0.0)
| | | | | | | codUnid = 17159: J00-J99 (0.0)
| | | | | | | codUnid = 17167: J00-J99 (0.0)
| | | | | | | codUnid = 17175: J00-J99 (0.0)
| | | | | | | codUnid = 17183: J00-J99 (0.0)
| | | | | | | codUnid = 17191: J00-J99 (0.0)
| | | | | | | codUnid = 17205: J00-J99 (0.0)
| | | | | | | codUnid = 17213: J00-J99 (0.0)
| | | | | | | codUnid = 17221: J00-J99 (0.0)
| | | | | | | codUnid = 17248: J00-J99 (0.0)
| | | | | | | codUnid = 17256: J00-J99 (0.0)
| | | | | | | codUnid = 17264: J00-J99 (0.0)

| | | | | | | codUnid = 17272: J00-J99 (0.0)
| | | | | | | codUnid = 17280: J00-J99 (0.0)
| | | | | | | codUnid = 17299: J00-J99 (0.0)
| | | | | | | codUnid = 17302: J00-J99 (0.0)
| | | | | | | codUnid = 17310: J00-J99 (0.0)
| | | | | | | codUnid = 17329: J00-J99 (0.0)
| | | | | | | codUnid = 17337: J00-J99 (0.0)
| | | | | | | codUnid = 17345: J00-J99 (0.0)
| | | | | | | codUnid = 17353: J00-J99 (0.0)
| | | | | | | codUnid = 17361: J00-J99 (0.0)
| | | | | | | codUnid = 17388: J00-J99 (0.0)
| | | | | | | codUnid = 17396: J00-J99 (0.0)
| | | | | | | codUnid = 17418: J00-J99 (0.0)
| | | | | | | codUnid = 17426: J00-J99 (0.0)
| | | | | | | codUnid = 17434: J00-J99 (0.0)
| | | | | | | codUnid = 17442: J00-J99 (0.0)
| | | | | | | codUnid = 17450: J00-J99 (0.0)
| | | | | | | codUnid = 17469: J00-J99 (0.0)
| | | | | | | codUnid = 17477: J00-J99 (0.0)
| | | | | | | codUnid = 17485: J00-J99 (0.0)
| | | | | | | codUnid = 17493: J00-J99 (0.0)
| | | | | | | codUnid = 17507: J00-J99 (0.0)
| | | | | | | codUnid = 17515: J00-J99 (0.0)
| | | | | | | codUnid = 17523: J00-J99 (0.0)
| | | | | | | codUnid = 17531: J00-J99 (0.0)
| | | | | | | codUnid = 17558: J00-J99 (0.0)

| | | | | | | | | | codCBO = 225203: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225225: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225230: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225250: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225255: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225265: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225280: J00-J99 (0.0)
| | | | | | | | | | codCBO = 225285: J00-J99 (0.0)
| | | | | | | | | | codCBO = 2231F9: J00-J99 (0.0)
| | | | | | | | | | codUnid = 2639513: J00-J99 (0.0)
| | | | | | | | | | codUnid = 2639548: J00-J99 (2.24/0.5)
| | | | | | | | | | codUnid = 2639556: J00-J99 (0.0)
| | | | | | | | | | codUnid = 2639572: J00-J99 (0.0)
| | | | | | | | | | codUnid = 2715899: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3270785: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3298949: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3359646: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3507815: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3827836: R00-R99 (1.0/0.25)
| | | | | | | | | | codUnid = 3862224: J00-J99 (0.0)
| | | | | | | | | | codUnid = 3912698: J00-J99 (0.0)
| | | | | | | | | | codUnid = 5142725: J00-J99 (0.0)
| | | | | | | | | | codUnid = 5196809: J00-J99 (0.0)
| | | | | | | | | | codUnid = 5323436: J00-J99 (0.0)
| | | | | | | | | | codUnid = 5406617: J00-J99 (0.0)
| | | | | | | | | | codUnid = 5406625: J00-J99 (0.0)

| | | | | | | codUnid = 5446392: J00-J99 (0.0)

| | | | | | | codUnid = 5484456: J00-J99 (0.0)

| | | | | | | codUnid = 5506115: J00-J99 (0.0)

| | | | | | | codUnid = 5583225: J00-J99 (0.0)

| | | | | | | codUnid = 5584787: J00-J99 (0.0)

| | | | | | | codUnid = 5586372: J00-J99 (0.0)

| | | | | | | codUnid = 5587956: J00-J99 (0.0)

| | | | | | | codUnid = 5701104: J00-J99 (0.0)

| | | | | | | codUnid = 5709512: J00-J99 (0.0)

| | | | | | | codUnid = 5932238: J00-J99 (0.0)

| | | | | | | codUnid = 6388671: J00-J99 (0.0)

| | | | | | | codUnid = 6724299: J00-J99 (0.0)

| | | | | | | codUnid = 7686625: J00-J99 (0.0)

| | | | | | | codUnid = 9110402: J00-J99 (0.0)

| | | | | | | codUnid = 9214097: N00-N99 (0.25)

| | | | | | | codUnid = 9381015: J00-J99 (0.0)

| | | | | | | bairro = BAIRRO ALTO: R00-R99 (2.74/0.75)

| | | | | | | bairro = PINHEIRINHO : R00-R99 (2.74/1.25)

| | | | | | | bairro = NOVO MUNDO: R00-R99 (1.99/0.75)

| | | | | | | bairro = TATUQUARA : J00-J99 (5.23/1.49)

| | | | | | | bairro = CACHOEIRA : R00-R99 (0.25)

| | | | | | | bairro = NAO INFORMADO: J00-J99 (0.25)

| | | | | | | bairro = UMBARA: R00-R99 (1.25/0.75)

| | | | | | | bairro = SANTA FELICIDADE: J00-J99 (1.25/0.75)

| | | | | | | bairro = PORTAO : J00-J99 (0.75/0.25)

| | | | | | | bairro = XAXIM : J00-J99 (3.99/1.99)

| | | | | bairro = BOQUEIRAO : J00-J99 (4.23/1.49)

| | | | | bairro = UBERABA: J00-J99 (4.73/2.24)

| | | | | bairro = CAPAO RASO: R00-R99 (2.99/1.49)

| | | | | bairro = VISTA ALEGRE : A00-B99 (0.25)

| | | | | bairro = HUGO LANGE : J00-J99 (0.0)

| | | | | bairro = ABRANCHES : J00-J99 (0.25)

| | | | | bairro = SITIO CERCADO: J00-J99 (6.48/3.49)

| | | | | bairro = SANTA CANDIDA : R00-R99 (3.24/0.75)

| | | | | bairro = PILARZINHO: J00-J99 (0.75/0.25)

| | | | | bairro = CAPAO DA IMBUIA: J00-J99 (0.75/0.25)

| | | | | bairro = AHU : A00-B99 (0.5/0.25)

| | | | | bairro = GUAIRA: R00-R99 (1.25/0.5)

| | | | | bairro = FAZENDINHA: R00-R99 (1.99/1.0)

| | | | | bairro = SAO BRAZ : J00-J99 (0.0)

| | | | | bairro = CAMPO DE SANTANA : J00-J99 (2.24/0.75)

| | | | | bairro = ALTO BOQUEIRAO: R00-R99 (2.99/1.25)

| | | | | bairro = PAROLIN : J00-J99 (0.0)

APÊNDICE F - TRECHO DO RESULTADO DO ALGORITMO DECISION TABLE (JUL2018)

==== Run information ====

Scheme: weka.classifiers.rules.DecisionTable -X 1 -E acc -R -S
"weka.attributeSelection.BestFirst -D 1 -N 5"

Relation: AGO18-jul18

Instances: 298532

Attributes: 25

dtAtendimento

faixaEtaria

sexo

codTipoUnid

codUnid

codProced

codCBO

solicitacaoExames

encamAtendEspecialista

areaAtuacao

desencadeouInternamento

dtInternamento

estabSolicitante

estabDestino

capIntCID

tratamentoDomicilio

abastecimento

energiaEletrica

tipoHabitacao

destinoLixo

fezesUrina

casoDoenca

grupoComunitario

bairro

capCID

Test mode: 4-fold cross-validation

=== Classifier model (full training set) ===

Decision Table:

Number of training instances: 298532

Number of Rules : 33370

Non matches covered by Majority class.

Best first.

Start set: no attributes

Search direction: forward

Stale search after 5 node expansions

Total number of subsets evaluated: 239

Merit of best subset found: 48.027

Evaluation (for feature selection): CV (leave one out)

Feature set: 2,5,6,7,8,9,10,15,25

```

Rules:
=====
faixaEtaria      codUnid      codProced      codCBO      solicitacaoExames      encamAtendEspecialista      areaAtuacao      capIntCID      capCID
=====
35-39            3912698      301010064      225142      Nao                Sim                ODOR                ?                Z00-299
6388671          301010072      301010072      225225      Nao                Sim                CRCP                ?                Z00-299
45-49            5484456      301019940      225133      Nao                Sim                PSIP                ?                Z00-299
70-74            6388671      301010072      225280      Sim                Sim                CIOM                ?                R00-R99
60-64            6388671      301010072      225120      Sim                Sim                PNHP                ?                Z00-299
50-54            6388671      301010064      225124      Sim                Sim                GACN                ?                Z00-299
0-4              6724299      301010072      225155      Nao                Sim                CRCP                ?                Z00-299
85-89            6388671      301010072      225225      Nao                Sim                CIOM                ?                Z00-299
30-34            6388671      301010072      225225      Nao                Sim                CIOM                ?                Z00-299
70-74            6388671      301010072      225133      Nao                Sim                PSIA                ?                Z00-299
15-19            17531        301019940      225133      Nao                Sim                OBTH                ?                Z00-299
35-39            5587956      301010072      225250      Nao                Sim                OBTH                ?                Z00-299
20-24            5587956      301010072      225250      Nao                Sim                OBTH                ?                Z00-299
60-64            3270785      301010072      225285      Nao                Sim                URGA                ?                N00-N99
65-69            6388671      301010072      225120      Nao                Sim                PNIP                ?                Z00-299
50-54            17477        301019940      225133      Nao                Sim                PSIP                ?                Z00-299
20-24            17434        301019940      225133      Nao                Sim                PSIP                ?                Z00-299
55-59            3270785      301010072      225155      Nao                Sim                CIOM                ?                Z00-299
60-64            2639513      301010072      225125      Nao                Sim                HEPT                ?                Z00-299
10-14            5586372      301019940      225133      Nao                Sim                PSII                ?                Z00-299
20-24            17345        301010110      225250      Nao                Sim                OBTH                ?                Z00-299
15-19            5506115      301019940      225133      Nao                Sim                PSII                ?                Z00-299
10-14            5506115      301019940      225133      Nao                Sim                PSII                ?                Z00-299
15-19            5484456      301019940      225133      Nao                Sim                PSII                ?                Z00-299
15-19            3912698      301019940      225133      Nao                Sim                PSII                ?                Z00-299
10-14            5506115      301010072      225133      Nao                Sim                PSII                ?                Z00-299
0-4              17299        301010080      225124      Nao                Sim                PETR                ?                Z00-299
55-59            17108        301010110      225133      Nao                Sim                OBTH                ?                Z00-299
35-39            17299        301010110      225133      Nao                Sim                OBTH                ?                Z00-299
15-19            2639408      301010064      225125      Nao                Sim                CAMP                ?                Z00-299
70-74            6388671      301010072      225285      Nao                Sim                PSII                ?                R00-R99
15-19            3359646      301010072      225133      Nao                Sim                GINS                ?                Z00-299
55-59            6388671      301010072      225285      Nao                Sim                URGN                ?                Z00-299
20-24            2639378      301010110      225142      Nao                Sim                OBTH                ?                Z00-299
60-64            6388671      301010072      225285      Nao                Sim                URGN                ?                Z00-299
75-79            6388671      301010072      225285      Nao                Sim                URGN                ?                C00-D48
70-74            6388671      301010072      225285      Nao                Sim                URGN                ?                Z00-299
80-84            6388671      301010072      225285      Nao                Sim                URGN                ?                Z00-299
15-19            17477        301019940      225133      Nao                Sim                PSII                ?                Z00-299

```