Unioeste - Universidade Estadual do Oeste do Paraná
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
Curso de Licenciatura em Matemática

Descoberta de conhecimento em base de dados – Estudo de caso SAMU

Guilherme Vieira Bochi

CASCAVEL
2020

Guilherme Vieira Bochi

Descoberta de conhecimento em base de dados – Estudo de caso SAMU

Monografia apresentada como requisito parcial para obtenção do grau de Licenciado em Matemática, do Centro de Ciências Exatas e Tecnológicas da Universidade Estadual do Oeste do Paraná - Campus de Cascavel

Orientador: Prof[a]. Rosangela Villwock

CASCAVEL
2020

Guilherme Vieira Bochi

Descoberta de conhecimento em base de dados – Estudo de caso SAMU

Monografia apresentada como requisito parcial para obtenção do Título de Licenciado em Matemática, pela Universidade Estadual do Oeste do Paraná, Campus de Cascavel, aprovada pela Comissão formada pelos professores:

_____

Profª. Rosangela Villwock (Orientadora)
Colegiado de Matemática, UNIOESTE

_____

Profª. Simone Aparecida Miloca
Colegiado de Matemática, UNIOESTE

_____

Profº. Amarildo Vicente
Colegiado de Matemática, UNIOESTE

_____

Jader Felipe Burg
Graduado em Matemática (Unioeste/Cvel)
Encarregado de setor de Estatística - CONSAMU

Cascavel, 02 de julho de 2021

”Ninguém vai bater mais forte do que a vida. Não importa como você bate e sim o quanto aguenta apanhar e continuar lutando; o quanto pode suportar e seguir em frente. É assim que se ganha.”

Rocky Balboa

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

x	Um objeto
C_i	Grupo de índice i
c_i	Centroide do grupo C_i
K	Número de grupos
m_i	Número de objetos no grupo de índice i

Sumário

Resumo

Este trabalho consiste em aplicar o processo KDD (Knowledge Discovery in Database), conhecido como Descoberta de Conhecimento em Base de Dados, na base de registros de ocorrências do SAMU (Serviço de Atendimento Móvel às Urgências) visando buscar os perfis dos pacientes atendidos. As características foram observadas nas ocorrências referentes aos meses de Março à Agosto do ano de 2020. Por meio da tarefa de Agrupamento de Dados foram gerados grupos de indivíduos que foram atendidos pelo SAMU. A idade dos indivíduos mostrou-se como um dos fatores importantes para a divisão dos grupos. Foi possível observar que o agravo possui uma relação com cada um dos grupos formados.

Palavras-chave: Mineração de Dados; K-means; Agrupamento de dados.

Capítulo 1

Introdução

O Ministério da Saúde lançou, em 2003, a Política Nacional de Urgência e Emergência com o intuito de estruturar e organizar a rede de urgência e emergência no país, com atenção primária constituída pelas unidades básicas de saúde e Equipes de Saúde da Família, atenção em nível intermediário a encargo do SAMU 192 (Serviço de Atendimento Móvel às Urgências) e das Unidades de Pronto Atendimento (UPA 24H) e o atendimento de média e alta complexidade feito nos hospitais (CONSAMU, Documento eletrônico).

O Consórcio de Saúde dos Municípios do Oeste do Paraná - CONSAMU é um consórcio sem fins lucrativos sediado na cidade de Cascavel. Este tem as finalidades de executar ações e serviços na área de regulação das urgências, transporte de pacientes, atendimentos pré-hospitalares e gestão hospitalar (CONSAMU, Documento eletrônico).

O CONSAMU contempla as 10ª e 20ª Regionais de Saúde, que abrange em 43 municípios, atendendo a aproximadamente 950 mil pessoas. Além de contar com o SAMU 192 e a UPA 24H, o consórcio tem o apoio de motolâncias para que o socorro chegue de forma mais rápida ao local em que a vítima se encontra, o serviço de helicóptero que tem como objetivo transportar pacientes em situações que não podem ser feitas por meio terrestre, além de 27 Unidades de Suporte Básico e 8 de Unidades de Suporte Avançada.

Para contactar o SAMU pode-se discar para o número 192, que é destinado ao serviço de situações de emergências médicas, em que neste atendimento são feitas perguntas para a pessoa que entra em contata, devendo ela responder de forma clara e objetiva, dando indicações do local e os acontecimentos. Todas as ocorrências atendidas pelo SAMU são registradas e ficam arquivadas em seu banco de dados.

Segundo Fayyad et al. (1996), o processo KDD é um processo não trivial de descoberta de padrões válidos, novos, úteis e acessíveis. A principal vantagem do processo de descoberta é que não são necessárias hipóteses, sendo que o conhecimento é extraído dos dados sem conhecimento prévio.

Desta forma, o tema em estudo nesse projeto será Aplicações em Knowledge Discovery in Databases (KDD) ou Descoberta de Conhecimento em Base de Dados. O problema a ser resolvido neste trabalho será explorar a base de dados fornecida pelo SAMU por meio do processo KDD, com a finalidade de buscar um algum tipo de padrão gerado pelo processo e por fim encontrar um tipo de conhecimento.

A partir das ocorrências registradas pelo SAMU, esse trabalho visou extrair conhecimento com o propósito de ajudar a entidade, servindo de subsídio para tomada de decisões mais acertadas pela equipe responsável, sendo estas decisões relacionadas as características dos indivíduos e as ocorrências.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo principal deste trabalho foi buscar uma relação entre o perfil dos grupos formados com seus agravos, através do algoritmo de Agrupamento de Dados.

1.1.2 Objetivos Específicos

- Explorar as ocorrências notificadas;
- Avaliar o perfil de cada grupo formado;
- Relacionar as características dos grupos com os tipos de agravo;

1.2 Justificativa

Em grandes bases de dados existem diversas informações, aquelas que são úteis pode-se chamar de conhecimento, mas extraí-las se torna uma tarefa difícil, sendo que, geralmente, as informações não são facilmente identificadas. Desta forma, grandes bancos de dados não

são abordados por analistas humanos, pois demandaria um grande tempo para que fosse feita a análise dos dados, podendo assim faltar detalhes neles, o que poderia gerar confusão na interpretação dos resultados obtidos na extração do conhecimento.

Um conhecimento corresponde, em geral, a um padrão ou diversos padrões, que podem ser descritos a partir de dados, oportunizando o estabelecimento de uma relação entre os dados e as informações relevantes, apoiando à tomada de decisões.

1.3 Estrutura do Trabalho

Para um melhor entendimento e localização do leitor, a monografia estará estruturada como segue.

No Capítulo 1 foram apresentados tema, estrutura do campo de pesquisa, problemas, objetivos, justificativa e a estrutura em que o trabalho será elaborado.

No Capítulo 2 foram apresentados os tópicos: Knowlegde Discovery in Databate (KDD); Mineração de Dados (MD) e Agrupamento de dados com a utilização do K-means.

No Capítulo 3 foram apresentados dados sobre a implementação para a construção dos grupos para o problema, bem como detalhes do tratamento e pré-processamento da base de dados.

No Capítulo 4 foram apresentados os resultados das variáveis escolhidas no trabalho, tanto de forma geral quanto individual, bem como o número de grupos gerados na implementação e seus resultados, onde também se encontra as características dos grupos por agravo.

No Capítulo 5 foram apresentadas as considerações finais dos resultados obtidos.

Capítulo 2

Referencial Teórico

2.1 Knowlegde Discovery in Databate(KDD)

As etapas do processo KDD, segundo Fayyad et al. (1996) e conforme a Figura 2.1, são: seleção, pré-processamento, formatação, mineração de dados e interpretação.

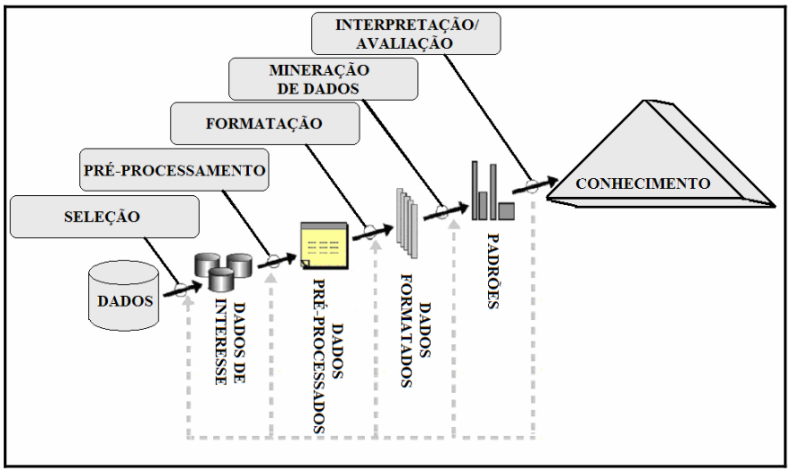

Figura 2.1: Esquema do KDD(Adaptado de FAYYAD et al., 1996)

Na etapa de seleção os dados de interesse são selecionados e nela o especialista de domínio do problema aponta as variáveis de interesse. Esta fase compreende, em essência, a identificação de quais informações, devem ser efetivamente consideradas durante o processo de KDD (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

A etapa do pré-processamento é responsável pela forma como os dados serão representados durante o processo do KDD (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Dentro desta etapa pode existir dados inconsistentes, que possuem algum tipo de discrepância entre os demais, sendo assim necessário efetuar uma limpeza dentro de sua base.

Na etapa da formatação temos a agregação de mais informações, para uma descoberta de conhecimento mais eficaz. Nesta etapa também é utilizada a transformação dos dados. Tal ajuste faz-se necessário para evitar que alguns atributos (padronização/normalização) influenciem de forma tendenciosa em determinados métodos de Mineração de Dados (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Na quarta etapa temos a principal etapa do processo, a MD. Dentre as tarefas de MD estão a Associação, Classificação, Clustering(Agrupamento), Regressão, Sumarização, entre outras. Durante está etapa é realizada a busca efetiva por conhecimentos úteis no contexto de aplicação do KDD (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Por fim, a quinta etapa consiste na avaliação do conhecimento adquirido. Também nesta etapa ocorre a interpretação do conhecimento obtido. Esta etapa envolve a visualização dos padrões extraídos ou visualização pelos dados fornecidos dos modelos extraídos (FAYYAD; SHAPIRO; SMYTH, 1996).

Nos remetendo à definição de Fayyad et al. (1996), vemos que o KDD é um processo interativo e iterativo. Goldschmidt et al. (2015) vê o processo *interativo* como a necessidade de um ser humano atuando nos processos do KDD, ou seja, o especialista de domínio analisa e interpreta os resultados que são obtidos pelo processo. Já o *iterativo* se refere ao fato de que repetições buscam por resultados satisfatórios, utilizando técnicas de refinamento, por exemplo.

2.2 Mineração de Dados

Porém, para alguns autores os termos KDD e MD são confundidos. Conforme Han et al. (2012) mineração de dados é um processo de interessantes descobrimentos de padrões vindo de grandes quantidades de dados. Segundo Tan et al. (2009) a mineração de dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados.

Para Freitas (2002) o termo mineração de dados refere-se à etapa principal de um processo mais amplo. Segundo Goldschmidt et al. (2015) a execução da etapa de Mineração de Dados compreende a aplicação de algoritmos sobre os dados procurando abstrair conhecimento.

Dentre as várias tarefas de MD, neste trabalho apenas as mais utilizadas serão abordadas, são elas: Agrupamento, Associação e Classificação.

No **agrupamento**, também chamado de *Cluterização* o objetivo é definir grupos de modo a

maximizar a similaridade dentro do grupo e minimizar a similaridade entre os grupos. Segundo Goldschmidt el al.(2015) o agrupamento é utilizado para separar os registros de uma base de dados em subconjuntos ou grupos, de tal forma que os elementos de um grupo compartilhem propriedades comuns que os distinguem de elementos de outros grupos.

Na **associação**, busca-se encontrar conjuntos de itens que ocorrem simultaneamente e de forma frequente em um banco de dados. O objetivo é produzir regras de dependência que irão predizer a ocorrência de um atributo baseado na ocorrência de outros (TAN; STEINBACH; KUMAR, 2009).

A **classificação**, segundo Goldschmidt el al.(2015), consiste em descobrir uma função que mapeie um conjunto de registros em um conjunto de rótulos categóricos predefinidos, denominados classes. Segundo Tan et al.(2009) o objetivo da classificação é encontrar um modelo de predição da classe como função dos outros atributos.

A tarefa a ser utilizada neste trabalho é o de agrupamento de dados.

2.3 Agrupamento de dados

Para execução da tarefa de agrupamento de dados existem diversas classes de algoritmos. Neste tópico serão abordados agrupamentos hierárquicos e por particionamento.

2.3.1 Agrupamento Hierárquico

Segundo Han et al.(2012) o método hierárquico cria uma decomposição (ou composição) hierárquica de um dado conjunto de objetos. Em sua abordagem aglomerativa cada registro começa como grupos individuais e, em cada etapa, os pares de grupos mais próximos são fundidos (TAN; STEINBACH; KUMAR, 2009). Na sua abordagem divisiva todos os registros iniciam num único grupo e, a cada etapa, divide-se um grupo até que restem apenas grupos de registros individuais (TAN; STEINBACH; KUMAR, 2009).

Um algoritmo básico de agrupamento hierárquico aglomerativo pode ser observado na Figura 2.2.

1: Calcule a matriz de proximidade.
2: **repita**
3: Funda os dois grupos mais próximos.
4: Atualize a matriz de proximidade para refletir a proximidade entre o novo grupo e os grupos originais.
5: **até que** Reste apenas um grupo.

Figura 2.2: Algoritmo de agrupamento hierárquico aglomerativo básico (adaptado de Tan et al.2009).

A matriz de proximidade pode ser encontrada com o nome de matriz de distância. A matriz é triangular inferior e nela se apresentam os valores para uma medida de dissimilaridade (ou distância) entre os registros.

Segundo Tan et al.(2009) e conforme a Figura 2.3, o dendrograma e a forma aninhada mostram tanto os relacionamentos grupo-subgrupo quanto a ordem na qual os grupos são fundidos (visão aglomerativa) ou divididos (visão divisiva).

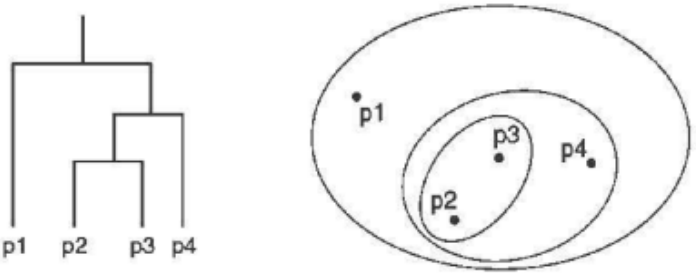

Figura 2.3: Agrupamento hierárquico em dendrograma e forma aninhada (TAN et al., 2009).

2.3.2 Agrupamento por particionamento

Segundo Goldschmidt et al. (2015) os agrupamentos por particionamento dividem o conjunto de dados em K grupos, inicialmente estes algoritmos escolhem K objetos como sendo os centros de K grupos.

Ainda para Goldschmidt et al. (2015) os métodos de clusterização por particionamento formam K grupos tão compactos e separados quanto se queira. O algoritmo mais comum dentre os de particionamento é o *K-means*.

No K-means, primeiramente é escolhido K, definido como parâmetro pelo usuário, sendo

ele o número de grupos desejados.

Conforme Goldschmidt et al. (2015) o algoritmo *K-means* é inicializado com os *K* centroides em posições aleatórias, os quais são atualizados de forma iterativa. Cada centroide representa um grupo e cada registro deve ser incluído ao grupo do centroide cuja distância ao registro seja a mínima. Na Figura 2.4, pode-se ver um algoritmo básico *K-means*.

1: Informar *K* quantidade de grupos.
2: Inicializar os *K* centroides de modo aleatório. Cada centroide é associado a um grupo.
3: Calcular distância de cada objeto até os centroides.
4: Atribuir cada objeto ao grupo cuja distância ao centroide associado seja mínima.
5: Atualizar os centroides.
6: Volte ao passo 3 até que não haja mudanças nos grupos ou, de forma que equivalente, que o centroide não sofra mudanças.
7: Exibir os clusters.

Figura 2.4: Pseudo código do *K-means*.

No exemplo da Figura 2.5 *K* foi inicializado com 3 grupos. Pelo fato dos centroides serem iniciados de forma aleatória, vemos que eles se encontram dispersos. Posteriormente, tem-se a rotulação dos objetos que se encontram mais próximos a cada centroide.

O centroide de cada grupo é então atualizado baseado nos pontos atribuídos ao grupo e na Figura 2.6 é possível observar o movimento dos centroides a cada iteração.

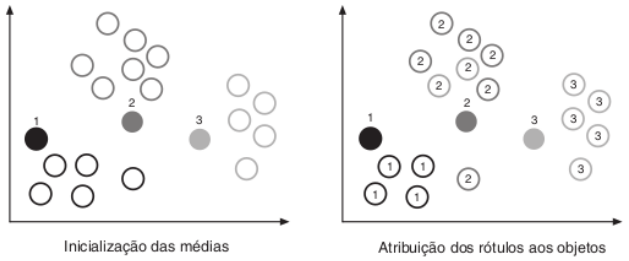

Figura 2.5: Ilustração dos primeiros subsequentes do algoritmo *K-means* (GOLSCHMIDT et al., 2015).

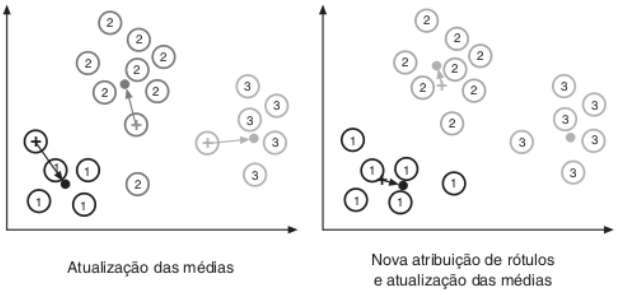

Figura 2.6: Ilustração dos passos subsequentes do algoritmo *K-means* (GOLSCHMIDT et al., 2015).

Quando os centroides não se movem mais o algoritmo deve parar e apresentam-se os grupos. O algoritmo *K-means* não garante convergir para um ótimo global e muitas vezes termina em um ótimo local (HAN; KAMBER; PEI, 2012). Tal fato pode ser dado pela ocorrência de que os centroides são iniciados de forma aleatória, sendo assim necessário realizar ajustes de parâmetros no algoritmo e executá-lo *n* vezes, tomando o melhor resultado destas execuções.

Conforme Tan et al. (2009), pode-se avaliar a qualidade do agrupamento calculando a Soma do Quadrado do Erro (SSE - Sum of Square Error) que é apresentada na Equação 2.1.

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} (c_i - x)^2 \tag{2.1}$$

onde, *x* é um objeto que deve pertencer ao grupo C_i e c_i é o centroide deste grupo.

Como visto anteriormente o centroide é o vetor de médias, o centro de um grupo. Sua formulação é dada da seguinte maneira:

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x \tag{2.2}$$

onde, m_i é o número de objetos no grupo de índice *i*.

A seguir, um exemplo numérico da aplicação do *K-means*, disponível em Johnson; Wichern (1998), que é dado na Tabela 2.1.

Tabela 2.1: Base de dados (retirado de Johnson; Wichern (1998)) exemplo.

	Variáveis	
Objetos	x_1	x_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Para efeitos de exemplo será utilizado $K = 2$ e os centroides escolhidos foram $c_1 = (2, 2)$ e $c_2 = (-1, -2)$. Calcula-se então a Distância Euclidiana de cada objeto que para cada centroide, os resultados são apresentados na Tabela 2.2.

Tabela 2.2: Distâncias iteração 1 - exemplo.

Objetos	c_1	c_2
A	10	61
B	10	9
C	17	4
D	41	4

Observando a menor distância para os centroides, o grupo c_1 é formado pelo objeto A e os demais objetos formam um segundo grupos c_2. A partir disto extraí-se as coordenadas dos centroides atualizados.

$$c_1 = (5, 3)$$
$$c_2 = (\frac{-1 + 1 - 3}{3}, \frac{1 - 2 - 2}{3}) = (-1, -1)$$

Como houve alteração nos centroides, inicia-se a segunda iteração.

Tabela 2.3: Distâncias iteração 2 - exemplo.

Objetos	c_1	c_2
A	0	52
B	40	4
C	41	5
D	89	5

Após calcular as distâncias aos centroides (Tabela 2.3), novamente vê-se que o objeto A fica no grupo c_1 e os demais formam grupo c_2. Logo o algoritmo nos aponta no exemplo que os grupos formados são $c_1 = \{A\}$ e $c_2 = \{BCD\}$. Os centroides portanto não mudam, pois continuam nos mesmos grupos

Como os grupos foram encontrados deve-se calcular a Soma do Quadrado do Erro (SQE).

$$\begin{aligned} SQE = (5-5)^2 + (3-3)^2 + (-1-(-1))^2 + (1-(-1))^2 + (1-(-1))^2 + \\ (-2-(-1))^2 + (-3-(-1))^2 + (-2-(-1))^2 \\ SQE = 0 + 0 + 0 + 2^2 + 2^2 + (-1)^2 + (-2)^2 + (-1)^2 \\ SQE = 4 + 4 + 1 + 4 + 1 = 14 \end{aligned}$$

Em Johnson; Wichern (1998) pode-se observar todos os possíveis grupos formados e o respectivo SQE. O menor valor do SQE obtido para $c_1 = \{A\}$ é 0, pois é sua distância até ele mesmo e para o $c_2 = \{BCD\}$ tem como total 14 devido a soma da distância dos três objetos.

Capítulo 3

Metodologia

Como mencionado anteriormente o SAMU disponibilizou diversas planilhas de registros das ocorrências referentes aos meses de Abril à Agosto do ano de 2020. Como estes arquivos possuíam informações duplicadas foi necessário um tratamento inicial, dando o devido tratamento em tais ocorrências.

Dada a duplicação de ocorrência, o que era identificado pela sua numeração, caso esta fosse relativa ao mesmo paciente, informações foram adicionadas a uma das linhas de ocorrência e a outra linha foi removida. Isto ocorre, por exemplo, quando uma segunda unidade móvel é acionada.

Após este tratamento, as ocorrências compuseram a base de dados a ser explorada neste trabalho, totalizando 19481 registros de ocorrências.

Para cada registro são apresentados os valores de 15 variáveis em geral, mas para a utilização neste trabalho foram escolhidas 7 variáveis, as quais são: A Data corresponde ao dia do mês em que a ocorrência se deu, já o Dia refere-se ao dia da semana em que houve a ocorrência. O Acionamento corresponde ao horário em que o recurso foi acionado pra prestar o atendimento. O prefixo corresponde a unidade móvel enviada até ó local da ocorrência. A Idade identifica qual a idade do paciente que foi assistido, o Sexo corresponde ao sexo do paciente assistido e o Encaminhamento corresponde ao encaminhamento dado ao paciente. Sendo eles: Liberado no local, Unidade de Pronto Atendimento(UPA), Hospital, Pronto Atendimento(PA), Óbito e outros.

Visto a duplicação, quatro novas variáveis foram inseridas na base dados para a eliminação da duplicata, as quais foram: Necessidade de USA foi marcado o número 1 como verdadeiro nesta variável quando existe a necessidade de uma USA para o atendimento; QTA USA corres-

ponde ao código de aborto da unidade. É marcado o número 1 como verdadeiro nesta variável quando existe a remoção da USA da ocorrência; Para a Necessidade de USB é marcado o número 1 como verdadeiro nesta variável quando existe a necessidade de uma USB para o atendimento; Para a Moto foi marcado o número 1 como verdadeiro nesta variável quando existe a necessidade de uma moto para o atendimento.

A partir disto, foi realizado a parte da Seleção dos dados que foram utilizados no trabalho. O Pré-processamento que se deu a partir das transformações que foram realizadas nos dados e escolha de variáveis, e a formatação que se deu pela inclusão de mais informações nos dados pela adição de quatro variáveis, as quais foram apresentadas no parágrafo anterior. No que se refere na Mineração de Dados foi utilizada a tarefa de agrupamento, adotando o algoritmo *K-means*, para posteriormente analisar os padrões encontrados e assim chegar em um conhecimento.

Dada a escolha do algoritmo *K-means* foi necessário fazer a transformação de variáveis categóricas em numéricas.

Para a variável Dia utilizou-se o valor 1 para descrever a Segunda-feira, 2 para a Terça-feira e assim por diante, até 7 que se refere ao Domingo. Para a variável Acionamento, os horários foram representados por um número real entre 0 e 1.

Para a variável Prefixo utilizou-se o valor 1 para Unidade de Suporte Básico (USB), 2 para Unidade de Suporte Avançada (USA) e 3 para a utilização da moto.

Para a variável Sexo utilizou-se 0 para sexo não informado na ocorrência, 1 para pessoas do sexo feminino e 2 para pessoas do sexo masculino.

Para a variável Encaminhamento utilizou-se o número 1 para a ocorrência com liberação no local; 2 para encaminhamento até uma Unidade de Pronto Atendimento (UPA); 3 para encaminhamento hospitalar; 4 para encaminhamento até um Pronto Atendimento (PA); 5 para óbito e 6 para outros tipos de encaminhamentos.

Além disso, para a variável idade utilizou-se 0 para idades que não foram informadas ou identificadas na ocorrência. Para crianças com menos de um ano de idade, atribui-se os seguintes valores: 100 (para crianças com menos de um mês de vida); 101 (para crianças com um mês); 102 (para crianças com dois meses); e assim por diante até 111 (para crianças com 11 meses).

3.1 Implementação da solução para o problema

Para a implementação do algoritmo *K-means* foi utilizado o software Python. Python é uma linguagem de programação de propósito geral, o que significa que ela pode ser empregada nos mais diferentes tipos de projeto, variando desde aplicações Web até sistemas de inteligência artificial (CORRÊA, 2019).

O Python é um programa de linguagem livre e de acesso a vários pacotes incorporados, bem como pacotes que podem ser adicionados. Existem muitos pacotes voltados para o meio científico. Conforme Corrêa(2019), 'NumPy'(manipulação de vetores e matrizes); 'SciPy'(integração e cálculo numérico); 'pandas'(manipulação de DataFrames); 'Matplotlib'(geração de gráficos) e 'scikit-learn'(algoritmo de mineração de dados e aprendizado de máquinas), estão entre estes.

O pacote Pandas trabalha com arquivos em forma de planilhas, mais especificamente com o arquivo CSV para leitura e edição, o pacote Matplotlib é responsável pela criação dos gráficos. O pacote scikit-learn é utilizado em *Machine Learning*. O módulo de clustering é o mais utilizado neste trabalho. Este módulo contém algoritmos necessários para a realização da tarefa de agrupamento de dados. Do pacote scikit-learn é importado a biblioteca *cluster*, a qual trás a função *KMeans* que foi o algoritmo utilizado neste trabalho. Para uso desta função são utilizados alguns parâmetros, dentre eles:

- n_cluster: Número máximo de grupos à ser gerado. Este parâmetro deve ser inserido pelo usuário.

- init: Este é método escolhido para a inicialização do *KMeans*. 'k-means++', o qual trás um caminho mais rápido de convergência outro parâmetro é o 'random' que inicia com centroides aleatórios, sendo que o parâmetro default é 'k-means++'.

- n_init: Número de vezes que o algoritmo k-means será executado com diferentes centroides. Por default é realizado 10 vezes este processo.

- max_iter: Número máximo de iterações a se realizar em cada execução. Por default são realizados 300 iterações.

3.2 Definição do número de grupos

O algoritmo de agrupamento aplicado foi o *K-means*. Neste trabalho apenas o primeiro parâmetro foi indicado, sendo que os demais foram executados pelo default do KMeans.

Segundo Tan, Steinbach e Kumar (2009), encontra-se o número natural de grupos no conjunto de dados procurando o número de grupos no qual exista um joelho, pico ou queda no desenho da medida de avaliação desenhada contra o número de grupos. Sendo que este desenho aponta para o número ideal de grupos, *K*, que deve ser utilizado.

Segundo Costa e Villwock (2015) ao fazer o gráfico do SQE para cada valor k de grupos, pode-se observar um joelho (cotovelo), o que indicaria uma redução significativa no SQE. A cada aumento de *K* haverá decréscimos no SQE até que em um certo valor *K* a diminuição será tão pequena que não será vantajoso aumentar o número de grupos.

O gráfico da Figura 3.1 mostra a variação do SQE para valores de *K* de 2 a 20. Não foi observado joelho (ou cotovelo), mas conclui-se que o número de grupos ideal foi $K = 7$. Para escolha de *K* foi observada a taxa de decréscimo do SQE, sendo que para este caso o aumento do número de grupos (*K* = 7) não apresentou relevante queda no SQE.

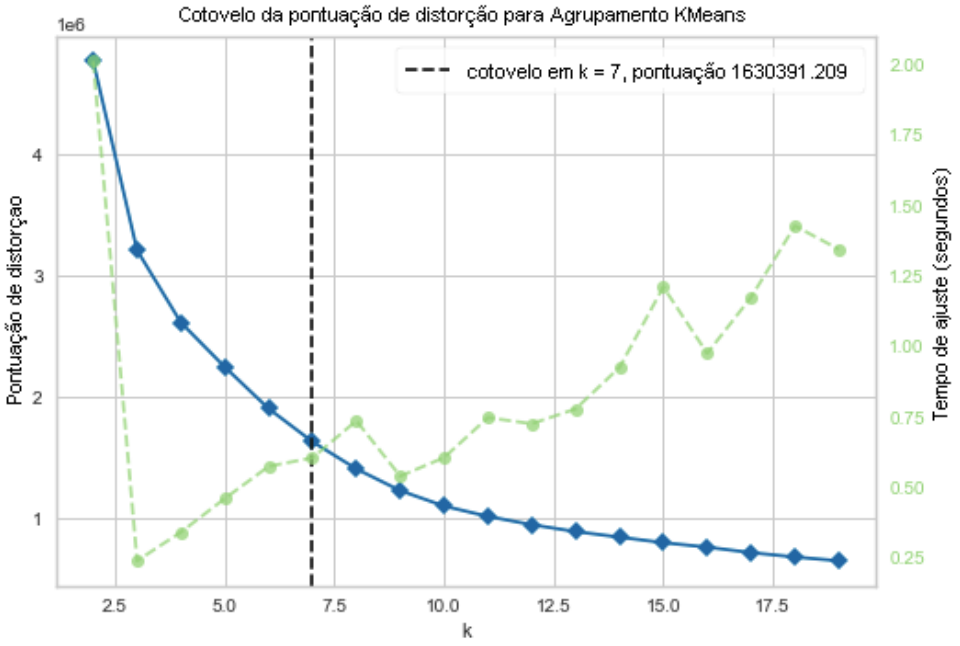

Figura 3.1: Gráfico SQE para *K* grupos.

Capítulo 4

Resultados e discussões

4.1 Resultados Preliminares

Na Tabela 4.1 é apresentado uma estatística geral de todas as variáveis que foram utilizadas no trabalho, que servem de base para visualizar os dados que possuem ou não algum tipo de influência dentro da base de dados, por exemplo o ACIONAMENTO teve média de 0,58 que se encontra relativamente próximo da média de seu mínimo e máximo.

Com relação à variável DATA não existe uma prevalência de dia do mês dentre as ocorrências do SAMU. O mesmo pode ser observado para a variável DIA (ver Tabela 4.2). Pode-se observar que os percentuais de atendimentos por dia da semana não possuem uma prevalência (ou concentração).

Tabela 4.1: Estatísticas das ocorrências do SAMU no período de Março a Agosto.

Variáveis	Média	Desvio padrão	Máximo	Mínimo	Moda
DATA	15,75	8,85	31	1	27
DIA	4,04	2,01	7	1	5
ACIONAMENTO	0,58	0,34	1	0	0,8
PREFIXO	1,17	0,40	3	1	1
IDADE	49,14	24,14	111	0	23
SEXO	1,50	0,52	2	0	2
ENCAMINHAMENTO	2,92	1,41	6	1	3
NECESSIDADE_DE_USA	0,04	0,19	1	0	0
QTA_USA	0,01	0,11	1	0	0
NECESSIDADE_DE_USB	0,02	0,14	1	0	0
MOTO	0,01	0,07	1	0	0

As médias da tabela serão utilizadas para efeito de comparação com os diferentes grupos formados.

Tabela 4.2: Ocorrências atendidas pelo SAMU por dia da semana no período de Março a Agosto.

Dias da semana	Número de ocorrências	% total
Segunda-feira	2836	14,56
Terça-feira	2648	13,59
Quarta-feira	2606	13,38
Quinta-feira	2693	13,82
Sexta-feira	2978	15,29
Sábado	2898	14,88
Domingo	2822	14,49
TOTAL	19481	100

Para variável Acionamento foram construídos quatro intervalos de 6 horas cada um, sendo madrugada (00:00 à 05:59), manhã (06:00 à 11:59), tarde (12:00 à 17:59) e noite (18:00 à 23:59). Pode-se observar na Tabela 4.3 que os períodos da tarde e noite são as que possuem um maior número de ocorrências atendidas pelo SAMU, somando mais de 60% das ocorrências.

Tabela 4.3: Ocorrências atendidas pelo SAMU por faixa de horário no período de Março a Agosto.

Horário	Número de ocorrências	% total
00:00 ⊢ 06:00	2513	12,90
06:00 ⊢ 12:00	4544	23,33
12:00 ⊢ 18:00	6283	32,25
18:00 ⊢ 00:00	6141	31,52
TOTAL	19481	100

Na Tabela 4.4 pode-se notar que a grande maioria das ocorrências são assistidas por uma Unidade de Suporte Básica(USB), vendo que em poucos casos é necessário a utilização de uma Unidade Avançada. Já a moto é acionada para poucas ocorrências.

Tabela 4.4: Ocorrências atendidas pelo SAMU por prefixo no período de Março a Agosto.

Prefixos	Número de ocorrências	% total
USB	16367	84,0
USA	2908	14,9
MOTO	206	1,1
TOTAL	19481	100

Para a variável Idade foram criadas algumas faixas etárias para uma melhor visualização dos dados. Pode-se observar que as faixas menor de um ano e de 91 à menor que 100 são aqueles

que possuem um menor registro de ocorrências. As demais faixas possuem números maiores de ocorrências, em especial a faixa dos 21 aos 30 anos de idade, que possui o maior índice de ocorrências registradas. Na sequência, em ordem decrescente de número de ocorrências, mas também em alto número, as faixas dos 31 aos 40 anos, 51 aos 60 anos, 41 aos 50 anos, 61 aos 70 anos e os posteriores.

Tabela 4.5: Ocorrências atendidas pelo SAMU por faixa de idade no período de Março a Agosto.

Intervalos de idades	**Soma das idades**
Não informado	313
Menos de 1 ano	173
1 – 10	511
11 – 20	1526
21 – 30	2964
31 – 40	2552
41 – 50	2371
51 – 60	2376
61 – 70	2363
71 – 80	2303
81 – 90	1662
91 – 99	367
TOTAL	19481

Com relação à variável Sexo não existe uma prevalência do sexo do indivíduo dentre as ocorrências do SAMU. Com relação ao encaminhamento, na Tabela 4.6 observa-se uma prevalência de encaminhamentos feitos para hospitais, totalizando 36% da ocorrências. Em segundo lugar vem a UPA (Unidade de Pronto Atendimento) para quais são encaminhados aproximadamente 30% das ocorrências. No que se refere ao parâmetro ”Outros”estão: não localizado, evadiu-se, recusa ao atendimento, transportado por meios próprios dentre outros.

Tabela 4.6: Encaminhamentos das ocorrências registradas pelo SAMU no período de Março a Agosto.

Encaminhamento	**Número de ocorrências**	**% do total**
Liberado no local	2433	12,49
UPA	5835	29,95
Hospital	7014	36
Pronto Atendimento	1456	7,47
Óbito	555	2,85
Outros	2188	11,23
TOTAL	19481	100

4.2 Resultados do Agrupamento de Dados

Na Tabela 4.7 estão apresentados os números e percentuais de registros de ocorrências em cada um dos grupos, neste caso os grupos formados tem percentuais próximos de ocorrência. A Tabela 4.8 apresenta os centroides dos sete grupos formados a partir das variáveis escolhidas, na qual observa-se pelos centroides das variáveis aquelas que possuem maior ou menor distinção entre os grupos. Por exemplo pode-se observar que a variável Dia possui pouca variação entre os grupos, enquanto a variável Idade possui uma maior variação entre os grupos.

Tabela 4.7: Ocorrências atendidas pelo SAMU por grupo.

Grupos	Registros	Relação ao todo(%)
0	3054	15,68
1	2437	12,51
2	3037	15,59
3	2451	12,58
4	2889	14,83
5	2660	13,65
6	2953	15,16

Tabela 4.8: Centroides dos grupos formados pela aplicação *K-means*.

Variáveis	Grupo 0	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
Data	8,87	12,15	8,22	9,99	23,06	23,99	23,97
Dia	3,99	4,19	4,08	4,03	3,97	3,93	4,1
Acionamento	0,58	0,6	0,58	0,57	0,59	0,57	0,57
Prefixo	1,18	1,16	1,15	1,2	1,15	1,2	1,15
Idade	60,72	13,45	35,97	84,29	49,9	75,42	26,59
Sexo	1,56	1,35	1,52	1,46	1,59	1,51	1,46
Encaminhamento	2,82	3,18	2,93	2,91	2,82	2,9	2,93
Necessidade Usa	0,03	0,03	0,03	0,05	0,04	0,05	0,03
QTA USA	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Necessidade USB	0,02	0,02	0,02	0,02	0,02	0,02	0,03
Moto	0,01	0	0	0,01	0,01	0,01	0,01

Em relação a variável Data pode-se notar que existem distinções entre alguns grupos. Pela Tabela 4.9 pode-se observar a porcentagem de ocorrências em cada intervalo (geral e por grupos), relativamente atendidas pelo SAMU.

Pode-se observar que nos grupos 0, 1, 2 e 3 existe uma concentração das ocorrências na primeira quinzena dos meses. Já nos últimos três grupos, a concentração das ocorrências encontra-se ao final dos meses. Observando a porcentagem geral pode-se observar que não há concentração de ocorrências relativamente ao dia do mês.

Tabela 4.9: Distribuição das ocorrências (em %) por intervalo de datas (por grupo e geral).

Datas	Geral	G0	G1	G2	G3	G4	G5	G6
1 – 5	16,1	30,1	21,0	33,2	28,5	–	–	–
6 – 10	16,8	31,8	24,0	33,9	27,9	–	–	–
11 – 15	16,8	26,4	25,6	25,3	25,9	8,0	3,7	3,7
16 – 20	16,0	11,0	14,8	5,4	13,5	24,4	20,2	21,1
21 – 25	15,7	0,7	8,0	–	3,4	30,7	34,4	32,4
26 – 31	18,6	–	6,6	–	0,9	37,0	41,7	42,8

Relativamente à variável Dia, assim como nos resultados preliminares, pois o dia da semana não tem influências sobre o número de ocorrências dentro dos grupos. Observando a Tabela 4.10, o percentual de ocorrências por grupo por dia da semana não difere de maneira relevante entre os grupos.

Tabela 4.10: Distribuição das ocorrências (em %) por dias da semana (por grupo e geral).

Dias da semana	Geral	G0	G1	G2	G3	G4	G5	G6
Segunda-feira	14,6	14,5	13,0	15,1	14,1	15,6	14,9	14,5
Terça-feira	13,6	13,0	13,0	12,6	13,7	14,1	15,2	13,7
Quarta-feira	13,4	14,4	13,3	12,4	13,2	13,7	14,3	12,5
Quinta-feira	13,8	14,6	11,9	13,1	15,2	14,0	14,3	13,6
Sexta-feira	15,3	17,0	16,0	16,5	15,7	13,6	14,2	14,1
Sábado	14,9	13,5	17,3	16,2	14,0	14,6	12,8	15,7
Domingo	14,5	13,1	15,6	14,1	14,2	14,4	14,3	15,9

Para a variável Acionamento pode-se observar que o horário de acionamento das ocorrências possuem o atendimento nos períodos da tarde e noite. Na Tabela 4.11 pode-se observar o número de ocorrências realizadas por grupos por faixas de horas. Porém, pode-se observar que no grupo 1 tem-se uma maior incidência (comparada à geral) à noite, enquanto que à tarde houve uma menor. Já para os grupos 3 e 5 houve uma maior incidência no horário da manhã, e no da noite houve uma menor incidência (comparada à geral). No que se refere ao grupo 6 houve uma maior incidência (em relação ao geral) na madrugada.

Tabela 4.11: Distribuição das ocorrências (em %) por intervalos de horários (por grupo e geral).

Acionamento	Geral	G0	G1	G2	G3	G4	G5	G6
00:00 ⊢ 06:00	12,9	10,8	14,9	14,6	9,7	11,9	10,3	17,5*
06:00 ⊢ 12:00	23,3	25,2	18,7	20,8	29,2*	21,8	27,7*	20,4
12:00 ⊢ 18:00	32,3	34,3	27,90**	30,5	34,8	34,9	35,2	28,2
18:00 ⊢ 00:00	31,5	29,7	38,4*	34,1	26,3**	31,4	26,7**	33,9

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

Para a variável Prefixo, unidade móvel utilizada na ocorrência pode-se observar a prevalência (em todos os grupos e no geral) no uso de Unidade de Suporte Básica (USB). Isso se dá pela frota de unidades móveis que estão disponíveis, composta por 27 unidades móveis básicas enquanto a Unidade de Suporte Avançada (USA) é de 8. No grupo 2 pode-se observar que existe uma maior incidência em relação ao geral para USB e Moto, enquanto que para USA existe uma menor incidência comparado com o geral. Para o grupo 3 observar-se que existe uma maior incidência em relação ao geral de USA e uma menor incidência no que se refere à USB. Já no grupo 4 existe uma menor incidência de USA em relação geral. No grupo 5 existe uma maior incidência em relação ao geral de USA. Já no grupo 6 observa-se uma maior incidência em relação ao geral de USB e uma menor incidência para USA em relação ao geral.

Tabela 4.12: Distribuição das ocorrências (em %) por unidade móvel enviada (por grupo e geral).

Prefixos	Geral	G0	G1	G2	G3	G4	G5	G6
USB	84,0	82,7	85,1	86,0*	80,7**	85,8	81,2	86,0*
USA	14,9	16,2	14,1	12,7**	18,4*	12,9**	18,0*	12,9**
MOTO	1,1	1,1	0,8	1,4*	0,9	1,2	0,8	1,2

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

A variável Idade é aquela que possui maior distinção entre todos os grupos. Na Tabela 4.13 pode-se observar as porcentagens de ocorrências (geral e dos grupos).

Tabela 4.13: Distribuição das ocorrências (em %) por faixas etárias (por grupo e geral).

Faixas de idades	Geral	G0	G1	G2	G3	G4	G5	G6
Não informado	1,61	-	1,61	-	-	-	-	-
Menos de 1 ano	0,89	-	-	-	6,97*	-	0,08	-
1 – 10	2,62	-	20,76*	-	0,04	-	-	0,14
11 – 20	7,83	-	42,22*	-	-	-	-	16,83*
21 – 30	15,21	-	24,17	24,66	-	-	-	55,06*
31 – 40	13,10	-	-	49,32	-	7,89	-	27,97*
41 – 50	12,17	6,39	-	26,01	-	47,98*	-	-
51 – 60	12,20	41,00*	-	-	-	38,91*	-	-
61 – 70	12,13	46,95*	-	-	-	5,23	29,25*	-
71 – 80	11,82	5,66	-	-	36,51*	-	46,02*	-
81 – 90	8,53	-	-	-	44,37*	-	21,80*	-
91 – 99	1,88	-	-	-	12,10	-	2,86	-

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

O grupo 0 é aquele em que os atendimentos se concentram em pessoas de 51 a 70 anos.

O grupo 1 é aquele em que o atendimento se concentra (em maioria) para crianças e adolescentes são atendidos, o qual também possui todas as pessoas que não informaram a idade.

O grupo 3 é aquele em que os atendimentos se concentram em pessoas de 71 a 90 anos. Neste grupo também são atendidos a maior parte dos bebês, acredita-se que os bebês foram atribuídos a esse grupo devido à forma como foram representados.

O grupo 4 é aquele em que os atendimentos se concentram em pessoas de 41 a 60 anos.

O grupo 5 é aquele em que os atendimentos se concentram em pessoas de 61 a 90 anos.

O grupo 6 é aquele em que os atendimentos se concentram nos jovens de 11 a 40 anos.

Em relação à variável Sexo pode-se observar na Tabela 4.14 as variações nos grupos em relação ao sexo de maior atendimento, sendo no geral a maioria masculina, mas com pequena vantagem.

Tabela 4.14: Distribuição das ocorrências (em %) por sexo (por grupo e geral).

Sexo	Geral	G0	G1	G2	G3	G4	G5	G6
Não informado	1,08	0,10	7,22	0,20	0,53	0,07	0,19	0,20
Feminino	48,08	43,84	50,76	47,55	53,00	41,23	48,42	53,13
Masculino	50,83	56,06	42,02	52,26	46,47	58,71*	51,39	46,66

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

Nos grupos 1, 3 e 6 têm-se uma maior incidência de atendimentos para pessoas do sexo feminino. Nos grupos 0, 2, 4 e 5 têm-se uma maior incidência de atendimentos de pessoas do sexo masculino, sendo que o grupo 4 apresenta a maior diferença entre os sexos.

Com relação a variável Encaminhamento, tem-se que sua prevalência é dada pelos hospitais em quase todos os grupos. Na Tabela 4.15 pode-se observar os percentuais de ocorrências por encaminhamento (no geral e em cada grupo).

Tabela 4.15: Distribuição das ocorrências (em %) por encaminhamento (por grupo e geral).

Encaminhamento	Geral	G0	G1	G2	G3	G4	G5	G6
Liberado no local	12,49	12,8	10,9	13,0	13,2	13,5	11,6	12,2
UPA	29,95	34,1*	24,5**	29,5	27,7	34,2*	30,1	28,2
Hospital	36,00	34,1	38,2	35,6	37,1	32,1**	37,4	38,4
Pronto Atendimento	7,47	6,3	7,0	7,8	7,8	7,9	7,0	8,6
Óbito	2,85	3,2	0,9**	1,4**	6,1*	2,1	6,1*	0,7**
Outros	11,23	9,6	18,5*	12,7	8,1**	10,3	7,9**	12,0

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

No grupo 0 pode-se observar que houve em parte encaminhamento para UPAs e hospitalar, mas com uma maior incidência (em relação ao geral) de encaminhamento para uma UPA.

No grupo 1 houve menor incidência (em relação ao geral) de encaminhamentos para UPAs e para Óbitos, tendo uma maior incidência (em relação ao geral) para os ”Outros” tipos de encaminhamento.

Nos grupos 2 e 6 (assim como no grupo 1) tiveram uma menor incidência (em relação ao geral) de Óbitos.

Nos grupos 3 e 5 pode-se observar que houve uma maior incidência (em relação ao geral) em Óbitos.

No grupo 4 houve uma menor incidência (em relação ao geral) para o hospital, apresentando uma maior incidência (em relação ao geral) para encaminhamentos para UPAs.

Na variável de necessidade USA tem-se que sua prevalência é dada por 0, ou seja, não houve necessidade de encaminhar uma USA para apoio no local (Tabela 4.16). O percentual de necessidade de apoio é pequeno.

Tabela 4.16: Distribuição das ocorrências (em %) por necessidade de USA (por grupo e geral).

Necessidade USA	Geral	G0	G1	G2	G3	G4	G5	G6
Não (0)	96,3	96,9	97,3	96,9	94,9	95,9	95,3	96,8
Sim (1)	3,7	3,1	2,7**	3,1	5,1*	4,1	4,7*	3,2

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

O grupo 1 apresenta uma menor incidência (em relação ao geral) na necessidade do apoio de uma USA.

Nos grupos 3 e 5 vê-se que a necessidade de apoio de uma USA possui uma maior incidência (em relação ao geral).

Para a variável de QTA USA (aborte desta unidade móvel) pode-se observar que o percentual desta variável é muito pequena e não difere de forma relevante entre os grupos (Tabela 4.17).

Tabela 4.17: Distribuição das ocorrências (em %) por aborte de USA (por grupo e geral).

QTA USA	Geral	G0	G1	G2	G3	G4	G5	G6
Não (0)	98,8	98,8	98,9	98,6	98,8	98,7	98,6	99,0
Sim (1)	1,2	1,2	1,1	1,4	1,2	1,3	1,4	1,0

Na variável necessidade USB um baixo percentual da necessidade de encaminhar uma segunda USB para apoio no local (Tabela 4.18). No grupo 3 pode-se observar que houve uma maior incidência (em relação ao geral) no não apoio de uma de USB, entretanto quanto a necessidade de apoio de uma USB houve uma menor incidência (em relação ao geral). Já no grupo 6 houve uma maior incidência necessidade de apoio de uma USB (em relação ao geral).

Tabela 4.18: Distribuição das ocorrências (em %) por necessidade de USB (por grupo e geral).

Necessidade USB	Geral	G0	G1	G2	G3	G4	G5	G6
Não (0)	97,9	98,1	98,3	97,7	98,4*	97,8	98,0	97,4
Sim (1)	2,1	1,9	1,7	2,3	1,6**	2,2	2,0	2,6*

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

Para a variável Moto a prevalência é de não acionamento (Tabela 4.19) não houve distinção relevante nos percentuais nos diferentes grupos.

Tabela 4.19: Distribuição das ocorrências (em %) por necessidade de moto (por grupo e geral).

Moto	Geral	G0	G1	G2	G3	G4	G5	G6
Não (0)	99,5	99,5	99,7	99,5	99,5	99,4	99,5	99,2
Sim (1)	0,5	0,5	0,3	0,5	0,5	0,6	0,5	0,8*

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

4.2.1 Perfil de cada grupo

De forma resumida, abaixo o perfil de cada grupo.

No Grupo 0 (G0) tem-se 3054 ocorrências, neste grupo há concentração de ocorrências em datas da primeira quinzena dos meses. Este grupo possui um maior acionamento nos períodos da tarde e da noite, nele há concentração de ocorrências na faixa de idade dos 51 aos 70 anos, que contabiliza 86,95% do grupo. A maioria das ocorrências são para o sexo masculino. Neste grupo o encaminhamento para UPA é maior que no geral.

Já o Grupo 1 (G1) tem-se 2437 ocorrências. Neste grupo houve concentração de ocorrências na primeira quinzena dos meses. Neste grupo o acionamento tem maior incidência no período da noite e menor incidência que o geral na parte da tarde. Este grupo atendeu pessoas com idades de 1 a 30 anos, sendo que a maioria das ocorrências em adolescentes. A maioria das pessoas atendidas nesse grupo são do sexo feminino. Neste grupo houve menor incidência que no geral para encaminhamento para UPAs e Óbitos. Neste grupo também a necessidade de apoio de uma USA é menos que no geral.

Para o Grupo 2 (G2) tem-se 3037 ocorrências. Neste grupo há concentração de ocorrências em datas da primeira dezena dos meses. Para o acionamento suas concentrações se encontram nos períodos da tarde e da noite, somando mais de 64% das ocorrências. Neste grupo o atendimento foi dedicado à pessoas dos 21 aos 50 anos, sendo a maior concentração de ocorrências para faixa de 31 a 40 anos, totalizando 49,32% das ocorrências. Neste grupo a maioria das pessoas atendidas é do sexo masculino. Nele houve uma menor incidência em relação ao geral em Óbitos. Neste grupo houve uma maior incidência no uso de USB em relação ao geral e uma menor incidência que no geral no uso de USA.

No Grupo 3 (G3) tem-se 2451 ocorrências. Neste grupo a concentração de ocorrências se dá na primeira quinzena dos meses. Referente ao horário de acionamento pode-se observar que existe um maior incidência nos períodos da manhã e da tarde, além de menor incidência

em relação ao geral. Este grupo possui maior incidência que no geral para o uso da USA e menor para USB. Neste grupo as ocorrências atendem pessoas (em maioria) dos 71 aos 90 anos. Grande parte dos menores de 1 ano de idade atendidos estão neste grupo. Neste grupo os atendimentos formados (em maioria) às pessoas do sexo feminino. Neste grupo tem-se maior incidência que no geral em óbitos, ainda no grupo é apresentado uma maior incidência que no geral da necessidade de apoio de USA e menor na necessidade de USB.

O Grupo 4 (G4) possui 2889 ocorrências. Neste grupo a concentração de ocorrências na última quinzena dos meses. Neste grupo o horário de acionamento se concentra a tarde e noite (com soma superior a 65% das ocorrências). Neste grupo houve uma menor incidência que no geral no uso de USA. Neste grupo a idade com maior concentração de ocorrências é de 41 a 60 anos, seguido pela faixa dos 51 aos 60. Neste grupo a maioria das pessoas atendidas é do sexo masculino, sendo neste a maior distinção entre os sexos. Neste grupo houve uma maior incidência que no qual para o encaminhamento para uma UPA e menor incidência que no geral para hospitalar.

Para o Grupo 5 (G5) tem-se 2660 ocorrências. Neste grupo a concentração de ocorrências na última quinzena dos meses. Neste grupo existe uma maior incidência de ocorrências no período da manhã e uma menor incidência que no geral no período da noite. Nele existe uma utilização mais elevada que no geral de USA. Neste grupo a maior parte dos atendimentos ocorre na faixa dos 61 aos 90 anos. Neste grupo a maioria dos atendimentos foi dado para o sexo masculino. Nele houve uma maior incidência nos óbitos que no geral, da mesma forma houve maior necessidade de apoio de USA que no geral.

No Grupo 6 (G6) tem-se 2953 ocorrências. Neste grupo a concentração de ocorrências na última quinzena dos meses. Neste grupo houve uma concentração nos períodos da tarde e noite, mas nele houve uma grande incidência (maior que no geral) no período da madrugada. Neste grupo há uma maior incidência no uso de USB que no geral e uma menor incidência de USA que no geral. Neste grupo a maior concentração de ocorrências foi para pessoas na faixa dos 21 aos 30 anos, sendo 55,06% das ocorrências. Neste grupo há maior incidência de atendimentos para pessoas do sexo feminino. Neste grupo houve uma menor incidência de óbitos que no geral.

4.3 Relacionando agravos aos grupos

Esta característica tem como significado visualizar qual o tipo de injúria sofrida pelo indivíduo, que pela base de dados é dada pelo agravo. Este por sua vez possui 7 tipos de informações, que são definidas pelas seguintes características:

- Acidentes de transito;
- Causas Externas;
- Clínico;
- Clínico pediátrico;
- Clínico obstétrico;
- Psiquiátrico;
- Não informado.

Alguns exemplos de causas externas são: queda de mesmo nível, agressão corporal, transferência e outros. A partir destes 7 tipos de agravos que existem na base de dados, pode-se fazer uma relação com os grupos que foram criados, o qual pode ser observado na Tabela 4.20.

Tabela 4.20: Distribuição das ocorrências (em %) por agravos (por grupo e geral).

Agravo	Geral	G0	G1	G2	G3	G4	G5	G6
Acidente de trânsito	6,83	4,55	13,71*	9,42*	0,08**	6,16	1,50**	11,92*
Causas Externas	7,61	6,32	10,75*	8,56	6,65	7,75	5,38	8,06
Clínico	72,13	85,53	49,65**	59,40**	92,21*	79,96	91,95*	47,71**
Clínico pediátrico	0,01	-	0,04	-	0,04	-	-	-
Clínico obstétrico	7,18	-	16,09*	12,35*	0,20**	0,83**	-	20,35*
Psiquiátrico	6,20	0,07**	9,48*	10,27*	0,82**	5,30	1,17**	11,95*
Não informado	0,04	3,54*	0,29*	-	-	-	-	-

* Houve maior incidência que no geral.

** Houve menor incidência que no geral.

Em geral pode-se observar que sua prevalência se encontra nos estados de agravo clínico (72,13%).

Nos diferentes grupos existem diversas diferenças nos tipos de agravos que foram registrados, isto se dá principalmente, pela faixa etária atendida que os grupos possuem.

No que se refere ao agravo ”acidente de trânsito”as maiores incidências (em relação ao geral) foram observadas nos grupos 1, 2 e 6. Acredita-se que isto se deve ao fato de que nestes grupos a maior parte dos atendimentos seja de adultos jovens. Corroborando com isto está o fato de que nos grupos 3 e 5 a incidência deste agravo seja menor em relação ao geral (nestes grupos os atendimentos são destinados a idosos.)

No que se refere ao agravo ”causas externas”, são diversas características que o compõem, sendo elas por exemplo: quedas, afogamentos, engasgos, dentre outros. Observa-se que o grupo 1 possui maior incidência (em relação ao geral) de atendimento para este agravo, sendo que este agravo deu-se em um grupo em que os atendimentos foram em maior quantidade para adolescentes e crianças, enquanto que nos demais grupos sua incidência permaneceu próxima da geral.

No que se refere ao agravo ”clínico”as maiores incidências (em relação ao geral) foram observadas nos grupos 3 e 5. Acredita-se que isto se deve ao fato de que são atendidos idosos de 60 a 90 anos e também pelas crianças com menos de 1 ano de idade. Já nos grupos 1, 2 e 6 existe uma menor incidência (em relação ao geral) neste tipo de agravo, nestes grupos os atendimentos são destinados às crianças, adolescentes e jovens adultos.

Para o agravo ”clínico pediátrico”houve incidência nos grupos 1 e 3 que possuem crianças, sendo no grupo 1 de crianças de 1 a 10 anos e no grupo 3 por crianças menores de 1 ano de idade.

No que se refere ao agravo ”clínico obstétrico”as maiores incidências (em relação ao geral) foram observadas nos grupos 1, 2 e 6. Acredita-se que isto se deve ao fato de que os os atendimentos nestes grupos são formados, por uma maioria, de adolescentes e jovens adultos. No grupo 4 existe uma menor incidência (em relação ao geral), devido à faixa etária atendida neste grupo. Nos grupos 0 e 5 não houve incidência, também devido à faixa etária atendida neste grupo. No grupo 3 (formado por idosos e bebês) houve baixa incidência, relativa ao atendimento à mãe já em trabalho de parto.

Para o agravo ”psiquiátrico”as maiores incidências (em relação ao geral) foram observadas nos grupos 1, 2 e 6. Estes grupos possuem atendimentos, em sua maioria, na faixa dos 11 aos 40 anos. Os grupos 0, 3 e 5 atendem a faixas etárias bem distintas, onde supõe-se incidência menor (em relação ao geral) deste agravo.

Para o agravo ”não informado”as incidências se encontram nos grupos 0 e 1, enquanto que nos demais não foi constatado. Este tipo de agravo se dá por exemplo: Caído em via pública, mal estar geral, dentre outros. Nestes grupos foram observados que os atendimentos se deram em maioria para faixa de 51 a 70 anos (no caso do grupo 0) e na faixa dos 11 aos 20 anos (para o grupo 1).

Capítulo 5

Considerações Finais

Discando o número 192, o cidadão estará ligando para uma central de regulação que conta com profissionais de saúde e médicos treinados para dar orientações de primeiros socorros por telefone. São estes profissionais que definem o tipo de atendimento, ambulância e equipe adequados a cada caso. Em alguns casos basta uma orientação por telefone. O SAMU sempre está preparado para chegar ao local em que paciente se encontra. A equipe presta atendimento no menor tempo possível, já no local (ainda fora do ambiente hospitalar) salvando vidas e diminuindo sequelas.

O problema resolvido neste trabalho foi a exploração da base de dados fornecida pelo SAMU por meio do processo do KDD, mais especificamente por agrupamento de dados. Busca-se com este processo descobrir padrões novos, válidos, úteis e acessíveis.

A partir dos resultados obtidos pode-se observar que a idade foi um fator de influência para a construção dos grupos, pois nela constatou-se as maiores diferenças entre grupos.

Fatores como data, dia da semana e sexo não possuem influências no que se refere aos atendimentos realizados pelo SAMU.

A partir da relação entre grupo e agravo observou-se que as características dos incidentes que houve estão relacionadas às faixas etárias que compõem os grupos, sendo que o tipo de agravo não foi relacionado como uma variável neste trabalho, o que pode ter levado a ter essa distribuição entre os grupos.

Espera-se que os resultados apresentados neste trabalho possam servir de auxílio para o CONSAMU, como base de novas elaborações ou estruturação, para uma reavaliação em maiores e futuras bases de dados, podendo assim encontrar novos padrões.

Referências Bibliográficas

CONSAMU. Documento eletrônico. [Acesso em: 14 mai 2020]. Disponível em: ⟨https://www.consamu.com.br/⟩.

CORRÊA, E. *Meu Primeiro Livro de Python*. 1. ed. EBOOK: Publicação digitalizada, 2019.

COSTA, C. M.; VILLWOCK, R. Utilização da soma do quadrado do erro na determinação do número de grupos no agrupamento de dados. Anais da XXIX Semana Acadêmica de Matemática, Cascavel, 2015.

FAYYAD, U.; SHAPIRO, G. P.; SMYTH, P. From data mining to knowledge discovery in databases. In: *Al Magazine*. California: AAAI, 1996.

FREITAS, A. A. *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. New York: Springer, 2002.

GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. *Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações*. Rio de Janeiro: Elsevier, 2015.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining Concepts and Techniques*. 3. ed. USA: Elsevier, 2012.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 6. ed. New Jersey: Pretice Hall, 1998.

SCIKIT-LEARN. Documento eletrônico. [Acesso em: 17 jun 2020]. Disponível em: ⟨https://scikit-learn.org/stable/index.html⟩.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introdução ao Data Mining*. Rio de Janeiro: Ciência Moderna, 2009.