



ISSN: 2447-5580

Disponível em: <http://periodicos.ufes.br/BJPE/index>



Brazilian Journal of
Production Engineering

BJPE - Revista Brasileira de Engenharia de Produção



Campus São Mateus
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

ARTIGO ORIGINAL

OPEN ACCESS

PERFIL DAS OCORRÊNCIAS POLICIAIS DA CIDADE DE ITAMARAJU-BA UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

PROFILE OF ITAMARAJU-BA CITY POLICY OCCURRENCES USING DATA MINING TECHNIQUES

Raphael Silva Souza¹, Wilian Hiroshi Hisatugu², Silvia das Dores Rissino^{3*}

^{1,2,3} Departamento de Computação e Eletrônica (DCEL), Centro Universitário Norte do Espírito Santo da Universidade Federal do Espírito Santo, Rodovia BR 101 Norte, Km. 60, Bairro Litorâneo, CEP 29932-540, São Mateus.

¹ raphasilvasouza96@gmail.com ² wilian.hisatugu@ufes.br ^{3*} srissino@gmail.com

ARTIGO INFO.

Recebido em: 27.08.2019

Aprovado em: 07.09.2019

Disponibilizado em: 20.09.2019

PALAVRAS-CHAVE:

Descoberta de Conhecimento; Ocorrência Policial; Mineração de Dados; Mapa de Calor; Correlações nas Ocorrências.

KEYWORDS:

Discovery of knowledge; Police report; Data mining; Heat map; Correlations in Occurrences.

*Autor Correspondente: Rissino, S.das D.

RESUMO

Este trabalho analisa os dados de ocorrências policiais da Cidade de Itamaraju-BA do período de 2016-2018. O objetivo é descobrir o perfil das ocorrências policiais registradas no período de 2016 a 2018 utilizando técnicas de mineração de dados implementadas com Software R. No pré-processamento realiza-se a identificação e tratamento de inconsistências, além das adequações necessárias no conjunto de dados para análise. Na etapa de

mineração de dados utiliza-se o algoritmo Apriori para identificação das correlações entre dois ou mais tipos de ocorrências o que possibilita a descoberta do perfil das ocorrências policiais. No Pós-processamento utiliza-se a API do Google Maps para construção de um mapa de calor das ocorrências. O perfil das ocorrências é exibido através de gráficos que mostram a grande incidência de violência doméstica na cidade.

ABSTRACT

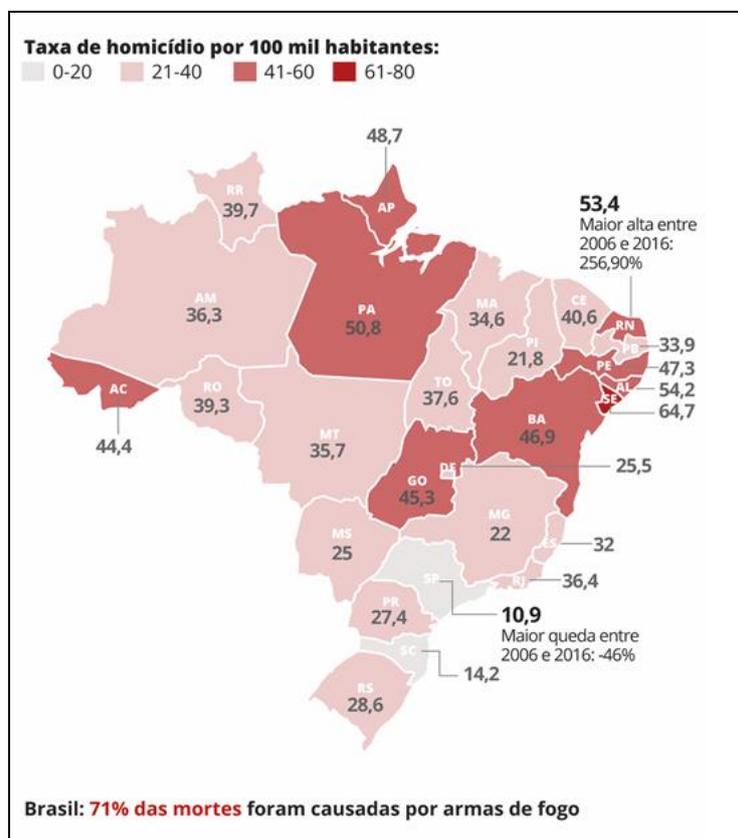
This paper analyzes data on police occurrences of the city of Itamaraju-BA from 2016-2018. The objective is to find out the profile of police occurrences recorded in the period from 2016 to 2018 using data mining techniques implemented with Software R. Preprocessing identifies and handles inconsistencies, as well as the necessary adjustments to the data set analyze. In the finger mining stage, the Apriori algorithm is used to identify correlations between two or more types of occurrences, which enable the discovery of the profile of police occurrences. Post-processing uses the Google Maps API to build a heat map of occurrences. The profile of the occurrences is displayed through graphs that show the high incidence of domestic violence in the city.



INTRODUÇÃO

O número de crimes realizados no mundo inteiro é alarmante e o Brasil é um dos países que contribuem com grande quantidade de ocorrências todos os anos. A Figura 1 apresenta a taxa de homicídio por 100 mil habitantes em cada estado do Brasil.

Figura 1. Mapa de Homicídios no Brasil em 2018.



Fonte – IPEA, 2018.

No ano de 2015, cerca de 59 mil vidas foram perdidas devido a homicídios registrados no país inteiro (Deursen, 2018). Esse número é mais alarmante se comparado com outros países como os Estados Unidos que tem uma população maior, mas mesmo assim possui um número de mortes menor por homicídios, aproximadamente 15,7 mil homicídios em 2015 (Deursen, 2018).

Em 10 anos, a taxa de homicídios na Bahia a cada 100 mil habitantes cresceu 97,8%. É o que aponta o Atlas da Violência, pesquisa do Instituto de Pesquisa Econômica Aplicada (IPEA) e do Fórum Brasileiro de Segurança Pública, chegando a taxa de 46,9 mortes a cada 100 mil habitantes. Um resultado muito elevado se comparado a o índice que a Organização das Nações Unidas (ONU) considera aceitável, que é a quantia de 10 mortes por 100 mil habitantes (IPEA, 2018). Nos últimos anos, a Ciência de Dados tem sido cada vez incorporada a metodologias de combate ao crime, através de técnicas e de tecnologia da informação e comunicação (TIC). As TICs possuem recursos que podem auxiliar a compreender padrões e tendências dos crimes ocorridos a partir dos dados de ocorrências policiais. Dentre estes, se destaca o processo de Descoberta de Conhecimento em Bases de Dados (KDD, do inglês Knowledge Discovery in



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Databases). O KDD, de acordo com Fayyad, Piatetsky-Shapiro & Smyth (1996), é todo o processo de preparação, seleção, limpeza, incorporação de conhecimento prévio apropriado e a interpretação corretas dos dados, essa prática está sendo chamada de *policciamento preditivo*.

O policiamento preditivo oferece resultados promissores para os tomadores de decisão e responsáveis por aplicar a lei. A inteligência preditiva pode ajudar a formular estratégias policiais mais eficazes.

As entidades policiais necessitam usar a tecnologia para melhorar o monitoramento e as suas ações para a manutenção da ordem pública. Conseqüentemente, há uma demanda elevada de analistas de dados de ocorrências, que sejam hábeis em usar a Ciência de Dados para criar produtos de informação descritivos e preditivos para apoiar os tomadores de decisão (Matos, s.d).

Neste contexto, este trabalho visa analisar os dados de ocorrências policia do município de Itamaraju e apresentar conhecimento útil aos tomadores de decisões de estratégia policial (Neto, 2017).

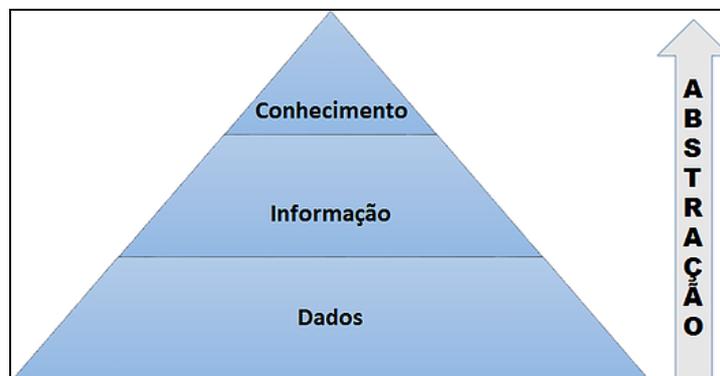
2. DESCOBERTA DE CONHECIMENTO EM BASE DADOS (KDD)

A análise de grandes quantidades de dados, pelo homem, é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação (Goldschmdit & Passos, 2005).

Para atender a este novo contexto, é necessário conhecer o Processo de KDD que é a Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases – KDD*), que vem despertando grande interesse junto às comunidades científica e industrial. A expressão Mineração de Dados, mais popular, é, na realidade, uma das etapas da Descoberta de Conhecimento em Bases de Dados.

Para proporcionar um melhor entendimento do problema, é importante destacar as diferenças e a hierarquia entre dado, informação e conhecimento, conforme ilustra a Figura 2 (Goldschmdit & Passos, 2005).

Figura 2. Hierarquia entre dados, informação e conhecimento.



Fonte – Elias, s.d.



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

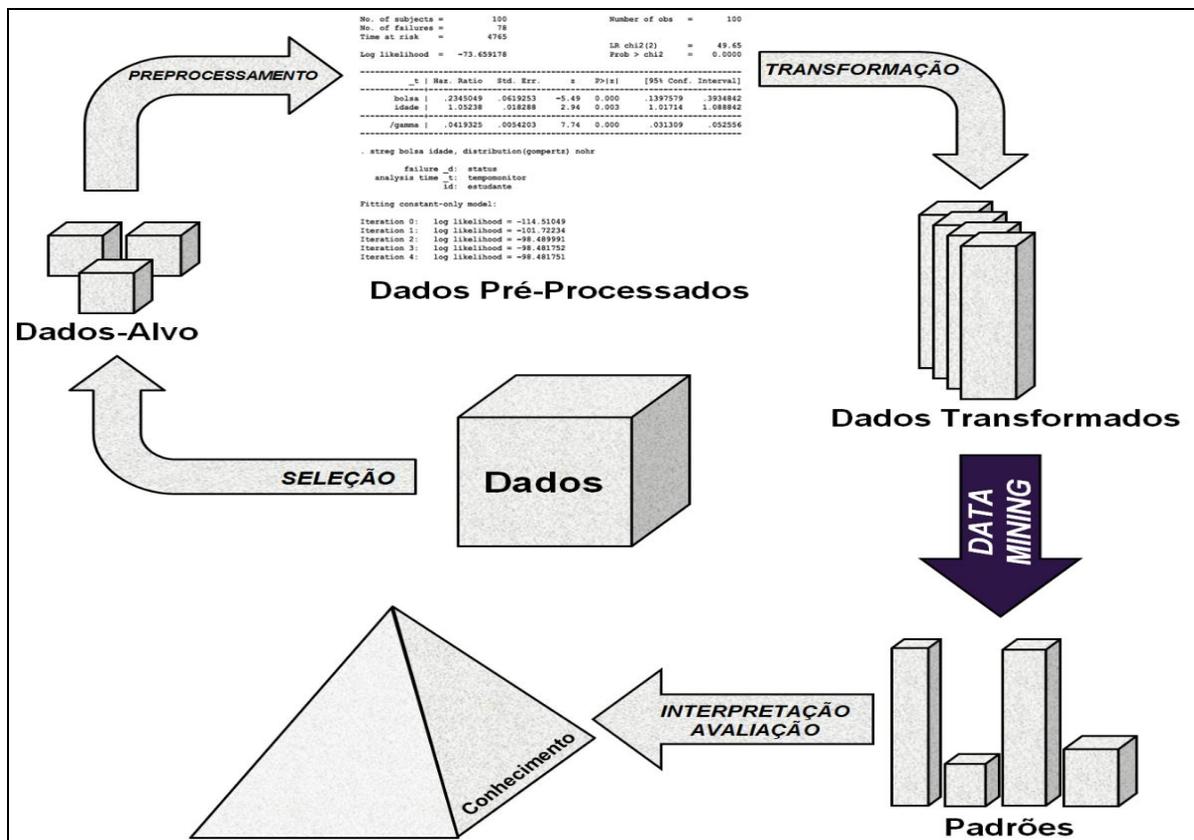
Os dados, na base da pirâmide, podem ser interpretados como itens elementares, captados e armazenados por recursos da Tecnologia da Informação.

As informações representam os dados processados, com significados e contextos bem definidos. Diversos recursos da Tecnologia da Informação são utilizados para facilmente processar dados e obter informações.

No topo da pirâmide está o conhecimento, padrão ou conjunto de padrões cuja formulação pode envolver e relacionar dados e informações (Goldschmdit & Passos, 2005).

O KDD é composto por 3 etapas principais: pré-processamento, mineração de dados e pós-processamento. A etapa de pré-processamento compreende as funções relacionadas à captação, à organização e ao tratamento dos dados. A etapa de pré-processamento tem como objetivo a preparação dos dados para os algoritmos da etapa seguinte, a Mineração de Dados. Durante a etapa de Mineração de Dados, é realizada a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD. A etapa de pós-processamento abrange o tratamento do conhecimento obtido na Mineração de Dados. Tal tratamento, nem sempre necessário, tem como objetivo viabilizar a avaliação da utilidade do conhecimento descoberto (Goldschmdit & Passos, 2005). A Figura 3 apresenta uma visão geral das etapas do KDD com as respectivas atividades.

Figura 3. Uma visão geral das etapas do KDD.



Fonte – Adaptado de (Fayyad & Piatetsky-Shapiro, 1996).



2.1 Pré-Processamento

Um dos principais obstáculos para MD (Mineração de Dados) são dados de má qualidade. Quando os dados são precários o produto de qualquer tarefa de MD também é precário.

- Muitos algoritmos não processam dados com ausência de valores de atributos, outros não trabalham com valores contínuos, outros não aceitam dados categóricos ou binários. Para resolver estes problemas é necessário efetuar um pré-processamento, que pode ser realizado manualmente ou de forma automática (Romão, 2002). Para a eficiente aplicação das técnicas de MD é necessário realizar uma preparação dos dados, conhecida como pré-processamento, que inclui as seguintes etapas (Wang & Sundaresh, 1998):
- Integração dos dados: remover inconsistências nos nomes ou em valores de atributos de diferentes origens;
- Limpeza dos dados: detectar e corrigir erros nos dados, substituir valores perdidos e outros;
- Conversão de dados nominais, ou em forma de códigos, para números inteiros;
- Redução do domínio (valores possíveis) para reduzir a distribuição dos valores no espaço de valores originalmente possíveis;
- Construir ou derivar novos atributos;
- Discretização: transformar atributos contínuos em categóricos, quando o algoritmo de MD não trabalha com atributos contínuos ou para melhorar a compreensão do conhecimento descoberto;
- Seleção de atributos: escolher atributos relevantes para a tarefa em questão.

2.2 Mineração de dados

Na Mineração de Dados, são definidos as técnicas e os algoritmos a serem utilizados no problema em questão. Redes Neurais (Haykin, 1999), Algoritmos Genéticos (Davis, 1990), Modelos Estatísticos e Probabilísticos (Michie, et al., 1994) são exemplos de técnicas que podem ser utilizadas na etapa de Mineração de Dados. A escolha da técnica depende, muitas vezes, do tipo de tarefa de KDD a ser realizada. A seguir algumas tarefas de KDD encontram-se comentadas (Goldschmdit & Passos, 2005).

- **Descoberta de Associação:** Abrange a busca por itens que frequentemente ocorram de forma simultânea em transações do banco de dados (Goldschmdit & Passos, 2005).
- **Classificação:** Consiste em descobrir uma função que mapeie um conjunto de registros em um conjunto de rótulos categóricos predefinidos, denominados classes. Uma vez descoberta, tal função pode ser aplicada a novos registros de forma a prever a classe em que tais registros se enquadram (Goldschmdit & Passos, 2005).
- **Regressão:** Compreende a busca por uma função que mapeie os registros de um banco de dados em valores reais. Esta tarefa é similar à tarefa de classificação, sendo restrita apenas a atributos numéricos (Goldschmdit & Passos, 2005).
- **Clusterização:** Utilizada para separar os registros de uma base de dados em subconjuntos ou clusters, de tal forma que os elementos de um cluster compartilhem de propriedades comuns que os distingam de elementos em outros clusters. O objetivo nessa tarefa é maximizar similaridade intracluster e minimizar similaridade intercluster. Diferente da



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

tarefa de classificação, que tem rótulos predefinidos, a clusterização precisa automaticamente identificar os grupos de dados aos quais o usuário deverá atribuir rótulos (Goldschmdit & Passos, 2005).

- **Sumarização:** Essa tarefa, muito comum em KDD, consiste em procurar identificar e indicar características comuns entre conjuntos de dados (Weiss& Indurkhya, 1998) (Goldschmdit & Passos, 2005).

A tabela 1 abaixo mostra algumas atividades da MD (mineração de dados) relacionados com alguns métodos/algoritmos.

Tabela 1. Algoritmos relacionados a sua atividade na mineração de dados.

Atividades MD	Métodos/algoritmos
Descoberta de Associações	Apriori, GSP, DHP, Basic, DIC, ASCX-2P
Classificação	Redes Neurais, Algoritmos Genéticos, Lógica Indutiva, Classificadores Bayesianos
Regressão	Métodos da estatística, Fuzzy e de Redes Neurais.
Clusterização	K-Means, KModes, K-Prototypes, K-Medoids e Kohonem.
Sumarização	Lógica Indutiva e Algoritmos Genéticos

Fonte: Adaptado de (Rissino & Lambert-Torres, 2009)

O analista de dados deve escolher qual desses métodos/algoritmos responde melhor ao problema de pesquisa. Cada um desses métodos exige que os dados estejam em um dado formato, de maneira que, possam ser moldados através da fase de pré-processamento.

3. Contextualização do Ambiente de Dados

O conjunto de dados (*dataset*) utilizado nesta análise é proveniente da 43ª Companhia Independente da Polícia Militar de Itamaraju/BA. O município de Itamarajú localiza-se no sul do estado da Bahia, no Brasil. Com 2.580 km² de área, sua população, conforme estimativas do IBGE de 2018 eram de 64.521 habitantes (IBGE, 2019).

Este trabalho foi realizado através de uma parceria realizada ente o Centro Universitário Norte do Espírito Santo - CEUNES/UFES a 43ª Companhia Independente da Polícia Militar de Itamaraju/BA. Esta parceria teve como único objetivo a análise de dados de ocorrência policial. Os dados foram solicitados através do Ofício 002/2019/DG/CEUNES/UFES da Direção do Centro Universitário Norte do Espírito Santo em primeiro de fevereiro de 2019.

A 43ª Companhia Independente da Polícia Militar de Itamaraju atendeu à solicitação e entregou os dados de 2016, 2017 e 2018 de forma anonimizada, isto é, sem identificação das vítimas e infratores das ocorrências, para que fossem realizadas as análises e a descoberta de conhecimento implícito entre os dados.



4. Ferramentas e Técnicas de Mineração

4.1 Formato do arquivo disponibilizado

A 43ª Companhia Independente da Polícia Militar de Itamaraju disponibilizou os dados em planilhas no formato CSV (Comma-Separated Values).

4.2 Software R

R é 'GNU S', isto é, linguagem e ambiente disponíveis gratuitamente para computação estatística e gráfica, o qual fornece uma ampla variedade de técnicas estatísticas e gráficas: modelagem linear e não linear, testes estatísticos, análise de séries temporais, classificação, agrupamento, entre outras (R Cran, s.d).

O R é também altamente expansível com o uso dos pacotes, que são bibliotecas para funções específicas ou áreas de estudo específicas. Um conjunto de pacotes é incluído com a instalação do software, além de outros disponíveis na rede de distribuição do R (em inglês CRAN) (Silva & Diniz, 2009). Neste trabalho utiliza-se o Software R na versão 3.5.3, R Studio na versão 1.1.463 e as bibliotecas *ggplot2*, *stringr*, *arules*.

4.3 Algoritmo Apriori

O Apriori é um algoritmo clássico de Mineração de Regras de Associação que utiliza um *hash* sobre uma árvore para coletar informações em uma base de dados ou dataset (Agrawal, 1993).

Diversos algoritmos tais como GSP, DHP, Partition, DIC, Eclat, MaxEclat, Clique e MaxClique foram inspirados no funcionamento do Apriori e se baseiam no princípio da antimonotonicidade do suporte. Segundo este princípio um k-itemset somente pode ser frequente se todos os seus (k-1)-itemsets forem frequentes. Logo, a combinação de itemsets para gerar um novo itemset somente ocorre quando estes são frequentes (Goldschmidt & Passos, 2005). Os algoritmos relacionados podem ser decompostos basicamente em duas etapas:

- a) Encontrar todos os conjuntos de itens frequentes (que satisfazem à condição de suporte mínimo).
- b) Gerar as regras de associação (que satisfazem à condição de confiança mínima), a partir do conjunto de itens frequentes.

Como a tarefa do item (a) demanda maior custo computacional e uma vez gerado todos os conjuntos de itens frequentes, a tarefa (b) se torna mais imediata, esforços de otimização têm sido concentrados na etapa (a).

5. Procedimentos Metodológicos do Trabalho

5.1 Obtenção dos dados

Na etapa inicial, foi realizada a solicitação dos dados, conforme descrito na seção 3 deste trabalho. Sendo, também, necessário realizar reuniões com os oficiais da 43ª Companhia



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Independente da Polícia Militar de Itamaraju, para entender como os dados eram coletados e enviados para digitação.

5.2 Pré-processamento

Com a base de dados (*dataset*) disponibilizada, esta foi carregada no software R, para iniciar a adequação dos conjuntos de dados e posterior análise.

Como os dados são inseridos de forma manual, pelo o responsável de transcrever o boletim de ocorrência, na base de dados da 43ª Companhia Independente da Polícia Militar, o qual apresenta várias inconsistências como, por exemplo, um mesmo tipo de ocorrências ser coletada maneiras diferentes. Para solucionar esse problema foi realizada uma codificação nos campos ocorrência, bairro, sexo, envolvido (V - Vítima, C - Condutor (para ocorrências de trânsito), A - Autor) e dia, consequentemente gerando uma padronização dos dados. A Figura 4 apresenta a base de dados de 2018 antes do pré-processamento.

Figura 4. Base de dados de março de 2018 antes do pré-processamento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DATA	DIA	HORA	RUA	BAIRRO	OCORRÊNCIA	RESULTAD	ENVOLVIE	SEXO	ENV	ENVOLVIE	SEXO	ENV	ENVOLVIDO 3
2	01/mar	QUI	09:00	ESTRADA - BR 101		VEÍC. RECUPERAD	1	VEÍC. RECUPERADO						
3	01/mar	QUI	?	RUA DOM	CENTRO	INFRAÇÃO DE TRÂ	1	AP. MOT	M					
4	01/mar	QUI	19:37	AVENIDA	BELA VIST	ACIDENTE DE TRÂN	CONFEC.	I	C	M	V	M		
5	01/mar	QUI	19:37	AVENIDA	BELA VIST	ACIDENTE DE TRÂN	1	AP. MOT	C	M	V	M		
6	02/mar	SEX	11:25	RUA MEN	BNH	OUTROS	RESOLV.	NA	M	T	M	T	F	
7	02/mar	SEX	15:10	RUA CARA	SÃO DOM	HOMICÍDIO	EVADIU-SIV	M	M					
8	02/mar	SEX	15:10	RUA CARA	SÃO DOM	HOMICÍDIO	01 AP DE CV	M	M					
9	02/mar	SEX	15:10	RUA CARA	SÃO DOM	HOMICÍDIO	01 AP DE CV	M	M					
10	02/mar	SEX	15:10	RUA CARA	SÃO DOM	HOMICÍDIO	01 AP DE TV	M	M					
11	02/mar	SEX	15:11	RUA JOSÉ	CENTRO	ACIDENTE DE TRÂN	CONFEC	D	C	M	C	M		
12	02/mar	SEX	15:40	RUA JOSÉ	CENTRO	INFRAÇÃO DE TRÂ	01	AP DE VA	M					
13	02/mar	SEX	16:45	RUA BONI	CENTRO	INFRAÇÃO DE TRÂ	1	AP. MOT	F					
14	03/mar	SAB	12:20	AVENIDA	CENTRO	INFRAÇÃO DE TRÂ	1	AP. MOT	M					
15	03/mar	SAB	16:20	RUA JOSE	CENTRO	ACIDENTE DE TRÂN	CONFEC	D	F	C	F			
16	03/mar	SAB	01:00	RUA AGUA	CORUJÃO	ROUBO A PESSOA	02	PRISÃO	F	A	M	V	F	
17	03/mar	SAB	14:25	TRAVESSA	31 DE MA	DANO	01	COND IV	M	A	M			
18	04/mar	DOM	13:10	RUA BENE	NOVO PR	DESENTENDIMENT	RESOLV.	NA	F	V	M			
19	04/mar	DOM	17:53	RUA VISCO	SANTO AN	INFRAÇÃO DE TRÂ	1	AP. MOT	M					
20	04/mar	DOM	21:40	AVENIDA	CENTRO	INFRAÇÃO DE TRÂ	1	AP. MOT	M					

Fonte - Autores, 2019

A Figura 5 apresenta a base de dados de 2018 depois do pré-processamento, isto é, após a codificação nos campos ocorrência, bairro, sexo, envolvido (V - Vítima, C - Condutor (para ocorrências de trânsito), A - Autor) e dia, consequentemente gerando uma padronização dos dados.

Figura 5. Base de dados de março de 2018 depois do pré-processamento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DATA	DIA	HORA	RUA	BAIRRO	OCORRÊN	RESULTAD	ENVOLVIE	SEXO	ENV	ENVOLVIE	SEXO	ENV	ENVOLVIDO.3
2	01/mar	5	09:00	ESTRADA	24	49	1	VEÍC. RECUPERADO						
3	01/mar	5	?	RUA DOM	1	6	1	AP. MOT	2	1				
4	01/mar	5	19:37	AVENIDA	20	12	CONFEC.	I	4	1	1	1		
5	01/mar	5	19:37	AVENIDA	20	12	1	AP. MOT	4	1	1	1		
6	02/mar	6	11:25	RUA MEN	38	99	RESOLV. N	2	1	3	1	3	0	
7	02/mar	6	15:10	RUA CARA	8	30	EVADIU-SI	1	1					
8	02/mar	6	15:10	RUA CARA	8	30	01 AP DE C	1	1					
9	02/mar	6	15:10	RUA CARA	8	30	01 AP DE C	1	1					
10	02/mar	6	15:10	RUA CARA	8	30	01 AP DE T	1	1					
11	02/mar	6	15:11	RUA JOSÉ	1	1	CONFEC D	4	1	4	1			
12	02/mar	6	15:40	RUA JOSÉ	1	6	01 AP DE V	2	1					
13	02/mar	6	16:45	RUA BONI	1	6	1	AP. MOT	2	0				
14	03/mar	7	12:20	AVENIDA	1	6	1	AP. MOT	2	1				
15	03/mar	7	16:20	RUA JOSE	1	1	CONFEC D	4	0	4	0			
16	03/mar	7	01:00	RUA AGUA	17	87	02	PRISÃO	2	0	2	1	1	0
17	03/mar	7	14:25	TRAVESSA	21	39	01	COND I	1	1	2	1		
18	04/mar	1	13:10	RUA BENE	5	24	RESOLV. N	2	0	1	1			
19	04/mar	1	17:53	RUA VISCO	13	6	1	AP. MOT	2	1				
20	04/mar	1	21:40	AVENIDA	1	6	1	AP. MOT	2	1				

Fonte - Autores, 2019



5.3 Regra de Associação (Apriori)

O Algoritmo Apriori foi aplicado nos dados da base de ocorrência criminais de Itamaraju-Ba, após os devidos ajustes nos parâmetros do Software R, os quais foram configurados com um support => 0.0011 e confidence => 0.3, após a execução do algoritmo apriori obteve um número alto de resultados – regras de associação.

A Tabela 2 apresenta as regras de associação descoberta na aplicação do Algoritmo Apriori sobre o conjunto de dados (dataset).

Tabela 2: Regras de associação geradas.

Número	CONDIÇÕES
1	{DOM, ROUBO A TRANSEUNTE} => {CENTRO} [support=0.003016591 confidence=0.8333333 lift=3.107424 count=10]
2	{AVENIDA VITÓRIA, POSSE DE ENTORPECENTES, VÁRZEA ALEGRE} => {19:00} [support=0.001508296 confidence=1.0000000 lift=82.875000 count=5]
3	{CRISTO REDENTOR, TER, TRÁFICO DE DROGAS} => {RUA ITAPEBI} [support=0.001809955 confidence=0.5454545 lift=48.869779 count=6]
4	{20:30, ROUBO A TRANSEUNTE} => {CENTRO} [support=0.001809955 confidence=0.6000000 lift=2.237345 count=6]
5	{LIBERDADE, VIOLÊNCIA DOMÉSTICA} => {SAB} [support=0.002714932 confidence=0.5000000 lift=2.717213 count=9]
6	{CORUJÃO, LESÃO CORPORAL} => {SAB} [support=0.001508296 confidence=0.8333333 lift=4.528689 count=5]
7	{SANTO ANTONIO, TRÁFICO DE DROGAS} => {QUI} [support=0.002111614 confidence=0.7777778 lift=5.382742 count=7]
8	{ROUBO A POSTO DE COMBUSTÍVEL} => {SANTO ANTONIO} [support=0.001809955 confidence=0.4285714 lift=11.645199 count=6]
9	{ROUBO A POSTO DE COMBUSTÍVEL, SANTO ANTONIO} => {SAB} [support=0.001206637 confidence=0.6666667 lift= 3.622951 count=4]
10	{CENTRO, ROUBO DE MOTO} => {SEG} [support=0.001508296 confidence=0.5000000 lift=3.946429 count=5]
11	{ROUBO A COMÉRCIO, SAB} => {CENTRO} [support=0.001508296 confidence=0.4545455 lift=1.694959 count=5]
12	{DOM, TENTATIVA DE HOMICÍDIO} => {NOVO PRADO} [support=0.001809955 confidence=0.3750000 lift=4.346591 count=6]
13	{MAROTINHO, VIOLÊNCIA DOMÉSTICA} => {DOM} [support=0.001508296 confidence=0.3571429 lift=2.188408 count=5]

Fonte - Autores, 2019

A Tabela 3 apresenta as regras apresentadas na Tabela 2, mas em formato de texto livre.



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Tabela 3: Descrição das regras de associação geradas em formato texto.

Número Regra	Descrição
1	Se o dia for DOMINGO e a ocorrência for ROUBO A TRANSEUNTE, então com uma confiança de 83,33% será no bairro CENTRO
2	Se for no endereço for AVENIDA VITÓRIA no bairro VÁRZEA ALEGRE e a ocorrência for POSSE DE ENTORPECENTES, então com uma confiança de 100% o horário será 19:00.
3	Se o dia for TERÇA-FEIRA no bairro CRISTO REDENTOR e a ocorrência for TRÁFICO DE DROGAS, então com uma confiança de 54,54% será na RUA ITAPEBI.
4	Se o horário for 20:30 e a ocorrência for ROUBO A TRANSEUNTE, então com uma confiança de 60% será no bairro CENTRO
5	Se o bairro for LIBERDADE e a ocorrência for VIOLÊNCIA DOMÉSTICA então com uma confiança de 50% o dia será SÁBADO.
6	Se o bairro for CORUJÃO e a ocorrência for LESÃO CORPORAL, então com uma confiança de 83,33% será no SÁBADO.
7	Se o bairro for SANTO ANTONIO e a ocorrência for TRÁFICO DE DROGAS, então com uma confiança de 77,77% será na QUINTA-FEIRA.
8	Se a ocorrência for ROUBO A POSTO DE COMBUSTÍVEL, então com uma confiança de 42,85% será no bairro SANTO ANTONIO.
9	Se a ocorrência for ROUBO A POSTO DE COMBUSTÍVEL e o bairro for SANTO ANTONIO, então com uma confiança de 66,66% será no SÁBADO.
10	Se o bairro for CENTRO e a ocorrência for ROUBO DE MOTO, então com uma confiança de 50% será na SEGUNDA-FEIRA.
11	Se a ocorrência for ROUBO A COMÉRCIO e o dia for SAB, então com uma confiança de 45,45% será no bairro CENTRO.
12	Se o dia for DOMINGO e a ocorrência for TENTATIVA DE HOMICÍDIO, então com uma confiança de 37,5% será no bairro NOVO PRADO.
13	Se o bairro for MAROTINHO e a ocorrência for VIOLÊNCIA DOMÉSTICA, então com uma confiança de 37,71% será no DOMINGO.

Fonte - Autores, 2019

5.4 Georreferenciamento das localidades das ocorrências criminais

Georreferenciamento ou georreferenciação de uma imagem ou um mapa ou qualquer outra forma de informação geográfica é tornar suas coordenadas conhecidas num dado sistema de referência (MundoGeo, 2012).

Neste trabalho, este processo inicia-se com a obtenção das coordenadas das ocorrências criminais, o que necessitou realizar o seguinte procedimento: construir dois códigos, o primeiro código em node.js (é um interpretador, com código aberto, em JavaScript de modo assíncrono e orientado a eventos), que captura o conteúdo das colunas rua e bairro e armazena em um arquivo.txt. O segundo código construído lê o arquivo.txt gerado da etapa anterior e com auxílio da Geocoding API do Google, no qual o endereço é convertido em coordenadas



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

com latitudes e longitudes, na terceira etapa do código são lidas as coordenadas de latitudes e longitudes e enviadas para Heatmaps API, a qual gera o mapa de calor das ocorrências na cidade. A Figura 6 apresenta o mapa com todas as ocorrências da cidade excluindo as ocorrências de trânsito, as quais são: (Acidente de trânsito sem vítimas, acidente de trânsito com vítimas, crime de trânsito, infração de trânsito e condução de veículo adulterado).

Figura 6. Mapa de calor de ocorrências em Itamaraju de 2016 a 2018.



Fonte - Autores, 2019

A Figura 7 apresenta o mapa de calor com as ocorrências contra a vida as quais são: (homicídio, duplo homicídio, tentativa de homicídio e latrocínio).

Figura 7. Mapa de calor de ocorrências contra a vida em Itamaraju de 2016 a 2018.



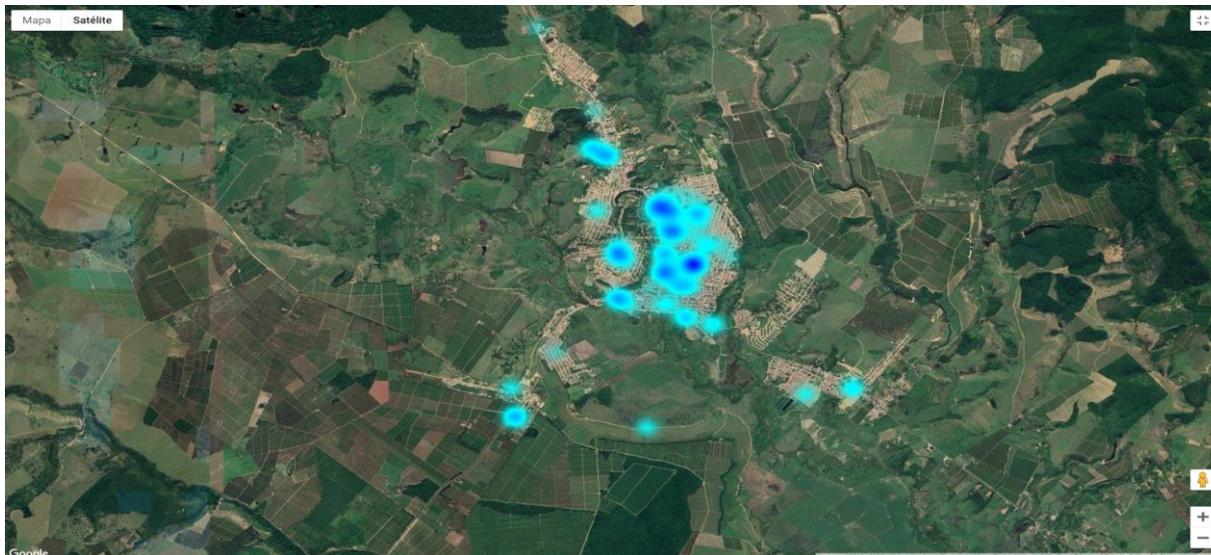
Fonte - Autores, 2019



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

A Figura 8, apresenta o mapa de calor das ocorrências contra o patrimônio que agrupa todos os tipos de roubos e furtos e suas respectivas tentativas.

Figura 8. Mapa de calor de ocorrências contra o patrimônio em Itamaraju de 2016 a 2018.

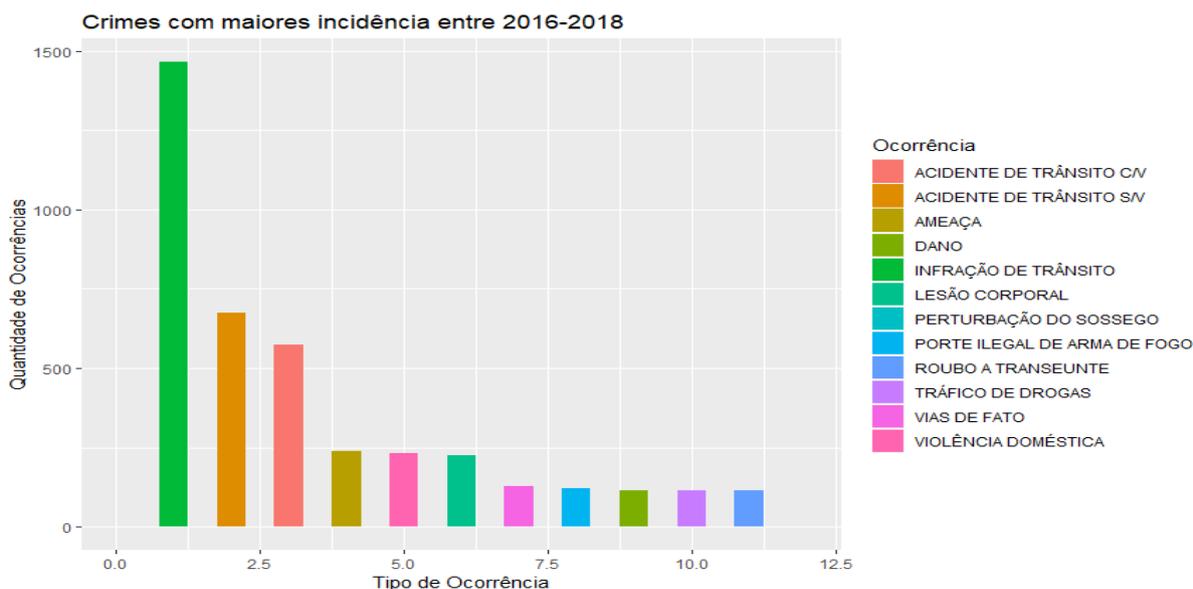


Fonte - Autores, 2019

6. PÓS-PROCESSAMENTO

Com o objetivo de apresentar os resultados, isto é, transformar o conjunto de dados iniciais em conhecimento útil, são gerados os infográficos, os quais exibirão o perfil das ocorrências do município de Itamaraju no período de 2016 até 2018. Os gráficos gerados e apresentados nas Figuras 9, 10, 11, 12 e 13 foram criados com o software R e a biblioteca “ggplot2”.

Figura 9. As 10 ocorrências com mais incidência em Itamaraju no período de 2016 até 2018.



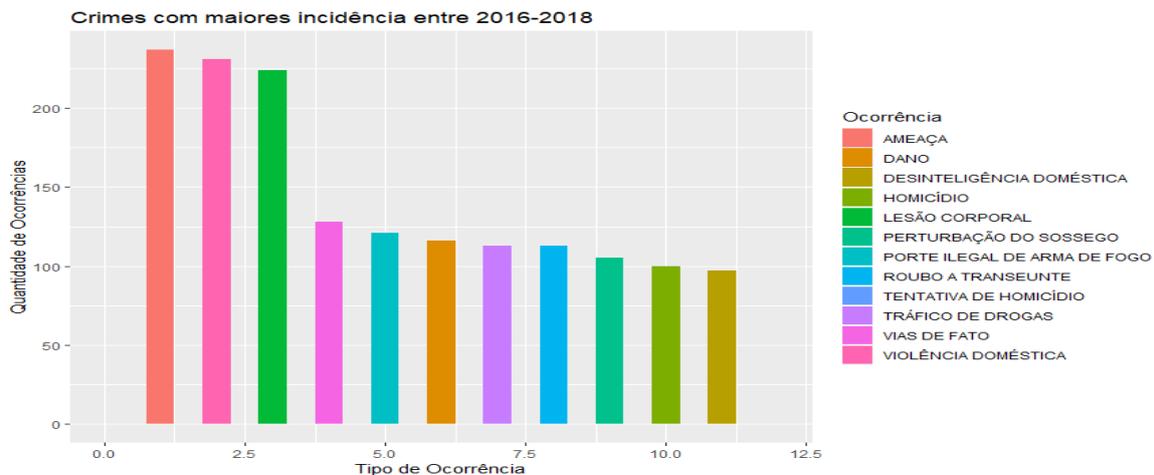
Fonte - Autores, 2019



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Com o resultado da figura 9 foi identificado que aproximadamente 50% das ocorrências da base de dados são relacionadas a trânsito, visto que o do objetivo do trabalho é identificar o perfil das criminais, por esse motivo foram retirados da base de dados as ocorrências "ACIDENTE DE TRÂNSITO S/V", "ACIDENTE DE TRÂNSITO C/V", "CRIME DE TRÂNSITO", "INFRAÇÃO DE TRÂNSITO" e "CONDUÇÃO DE VEÍCULO ADULTERADO". Após esse processo foi gerada o Gráfico da Figura 10, onde apresenta-se as ocorrências criminais com maiores incidências na cidade de Itamaraju no período de 2016 até 2018.

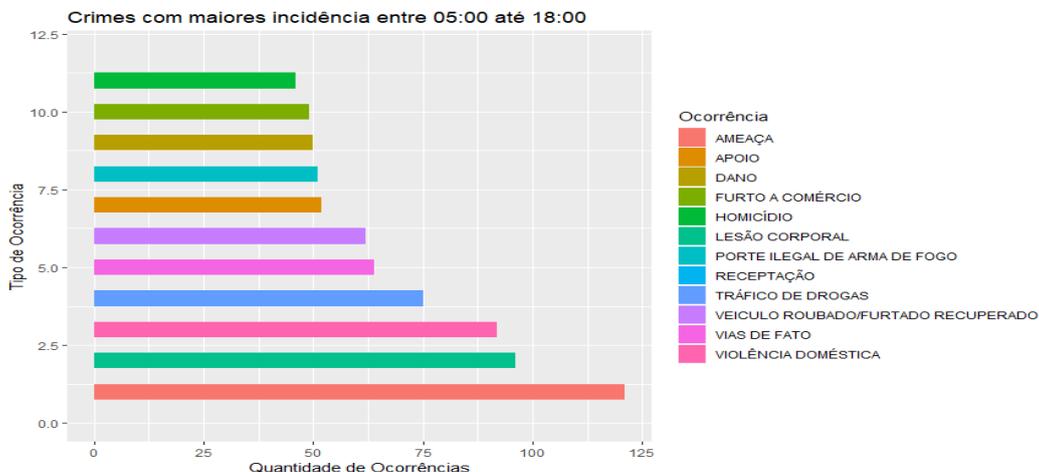
Figura 10. Ocorrências criminais com maiores incidências em Itamaraju no período de 2016 até 2018.



Fonte - Autores, 2019

Na Figura 10 identifica-se que a ocorrência ‘Violência Doméstica’ apresenta um índice muito alto em relação aos outros tipos. Nas Figuras 11,12 e 13 foram analisadas as ocorrências com maiores incidências durante os seguintes horários: dia (05:00 até as 18:00); noite (18:00 até as 22:00) e madrugada (22:00 até as 05:00).

Figura 11. Ocorrências criminais com maiores incidências em Itamaraju no período de 2016 até 2018 entre o horário das 05:00 até as 18:00.

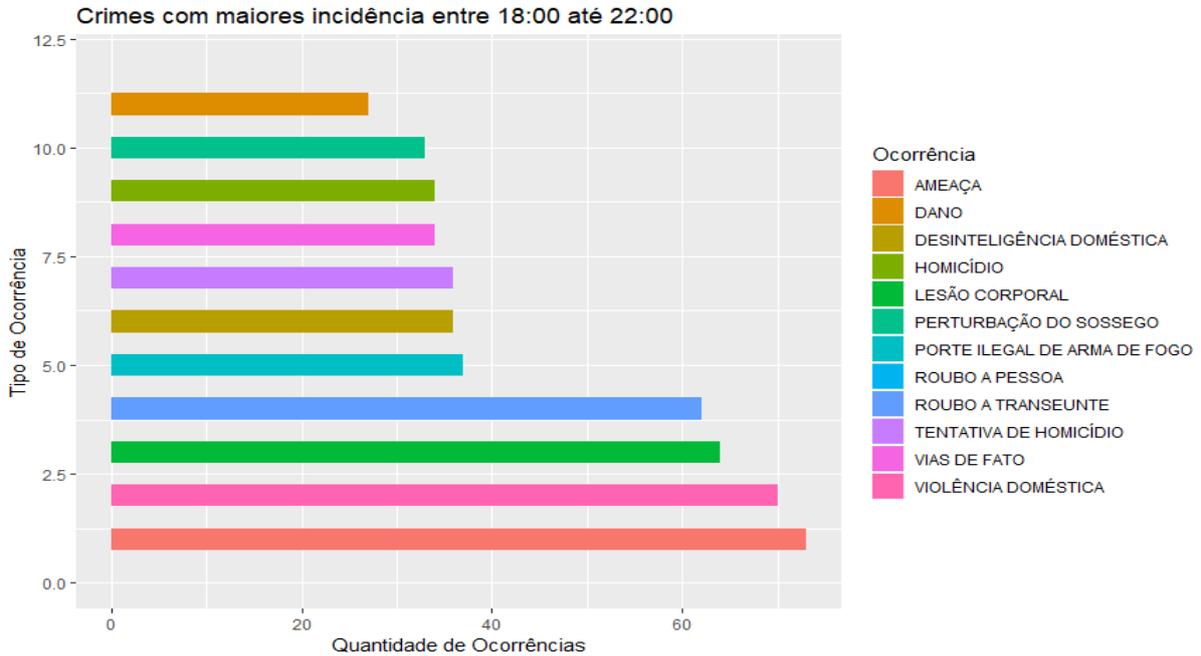


Fonte - Autores, 2019



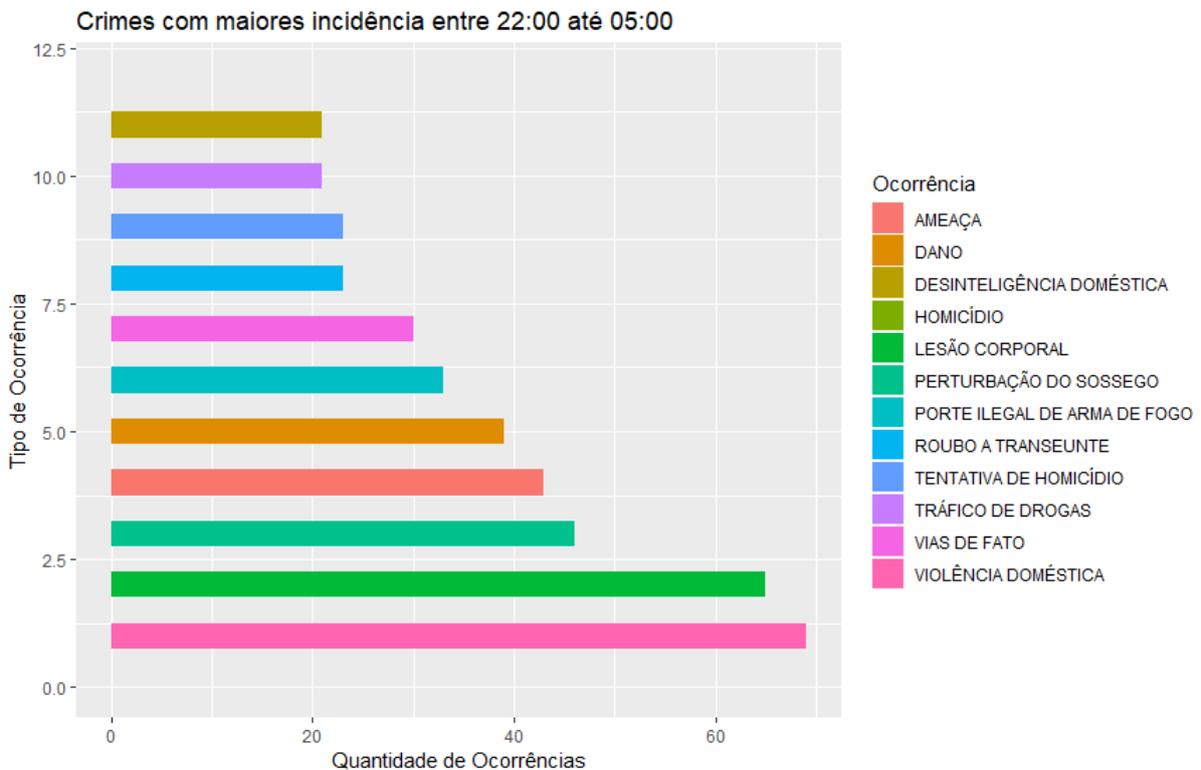
Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Figura 12. Ocorrências criminais com maiores incidências em Itamaraju no período de 2016 até 2018 entre o horário das 18:00 até as 22:00.



Fonte - Autores, 2019

Figura 13. Ocorrências criminais com maiores incidências em Itamaraju no período de 2016 até 2018 entre o horário das 22:00 até as 05:00.



Fonte - Próprio autor, 2019



Conclusão

Este trabalho aplicou o processo de descoberta de conhecimento (KDD) sobre o conjunto de dados disponibilizado pela 43ª Companhia Independente da Polícia Militar resultando a descoberta do perfil das ocorrências no Município de Itamaraju/BA.

O processo de KDD, com ênfase na etapa de mineração dados e aplicação do Algoritmo Apriori resultou na descoberta das regras de associação das ocorrências. A utilização do Apriori possibilitou identificar uma quantidade significativa da ocorrência “Violência doméstica”, o que indica a necessidade do uso outras técnicas de mineração de dados e inteligência artificial que possibilite uma avaliação sobre o crescimento desta ocorrência e a realização de um estudo mais detalhado sobre o CVLI (crimes violentos letais intencionais). Este conhecimento descoberto dificilmente seria identificadas sem o uso de técnicas computacionais, resultando no perfil das ocorrências da cidade de Itamaraju/BA.

O uso do georreferenciamento nas ocorrências criminais possibilitou a construção do mapa de calor, onde a intensidade da cor indica a localização aproximada com maior incidência criminal, proporcionando a tomada de decisão através do conhecimento disponibilizado com auxílio da inteligência geográfica.

Outras análises podem ser realizadas neste dataset, principalmente se houver possibilidade de aumentar a série histórica para um intervalo de tempo maior, o que resultará na predição de crimes de forma mais acurada.

A violência apresenta-se de várias formas e em vários locais, não importa se a cidade é pequena com menos de 100 mil habitantes ou megametrópoles com mais de um milhão de habitantes. Neste trabalho, a cidade de Itamaraju-Ba, serviu como objeto de estudo, possibilitando mostrar que ferramentas computacionais são poderosas aliadas a prevenção da violência, pois possibilita informações para o policiamento, que pode ser preditivo, através do conhecimento extraído dos dados das ocorrências.

Referências

- Davis, L. (1991). *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold.
- Deursen, F. (2018). O Brasil tem mais assassinatos do que todos esses países somados. Retrieved Maio 30, 2019 from <https://super.abril.com.br/blog/contaoutra/o-brasil-tem-mais-assassinatos-do-que-todos-estes-paises-somados/>.
- Elias, Diego. Dados VS Informação: Qual a diferença? [S. l.]. Retrieved Maio 27, 2019 from <https://www.binap pratica.com.br/dados-x-informacao>.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.
- Goldschmidt, R., & Passos, E. (2005). *Data mining: um guia prático*. Rio de Janeiro: Elsevier Editora Ltda.



Citação (APA): Souza, R.S., Hisatugu, W.H. & Rissino, S.das D. (2019). Perfil das ocorrências policiais da cidade de Itamaraju-BA utilizando técnicas de mineração de dados. *Brazilian Journal of Production Engineering*, 5(4), 66-81.

Haykin, S. (1999). *A Comprehensive Foundation: Neural Networks*.

Instituto Brasileiro de Geografia e Estatística. (2010). Itamaraju. Retrieved Maio 30, 2019 from <https://cidades.ibge.gov.br/brasil/ba/itamaraju/panorama>.

Cerqueira, D. C., Lima, R. S. D., Bueno, S., Neme, C., Ferreira, H., Coelho, D., ... & Reis, M. (2018). Atlas da violência 2018. Retrieved Junho 05, 2019 from http://www.ipea.gov.br/portal/images/stories/PDFs/relatorio_institucional/180604_atlas_da_violencia_2018.pdf.

Matos, David. Cientista de Dados na Análise de Crimes. [S. l.]. Retrieved Maio 29, 2019 from <http://www.cienciaedados.com/cientista-de-dados-na-analise-de-crimes>.

Michie, D & J. Spiegelhalter, D & Taylor, Charles. (1999). *Machine Learning, Neural and Statistical Classification*. Upper Sadler River: Ellis Horwood.

MundoGeo. (2012). Mapeamento de crimes e policiamento comunitário. Retrieved Maio 30, 2019 from

<https://mundogeo.com/2002/07/01/mapeamento-de-crimes-e-policiamento-comunitario/>.

Neto, S. D. (2017). Mineração de dados de ocorrências criminais para identificação de zonas de alta criminalidade em fortaleza e região metropolitana. Monografia (Graduação Sistemas de Informação), Universidade Federal do Ceará, Quixadá, CE, Brasil.

R Cran. The Comprehensive R Archive Network. [S. l.]. Retrieved Maio 28, 2019 from <https://cran.r-project.org/>.

Rissino, S., & Lambert-Torres, G. (2009). Rough set theory—fundamental concepts, principals, data extraction, and applications. In *Data mining and knowledge discovery in real life applications*. IntechOpen. Retrieved Maio 30, 2019 from https://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/rough_set_theory_fundamental_concepts__principals__data_extraction__and_applications.

Romão, W. (2002). Descoberta de conhecimento relevante em banco de dados sobre ciência e tecnologia. Tese de pós-graduação (pós-graduação em engenharia de produção), Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

Silva, Bruno Fontana da; Diniz, Jean; Bortoluzzi, Matias Américo. Minicurso de Estatística Básica: Introdução ao Software R. 04/2009. Retrieved Junho 05, 2019 from http://www.uft.edu.br/engambiental/prof/catalunha/arquivos/r/r_bruno.pdf.

