DOI: 10.5748/9788599693148-15CONTECSI/PS-5851

LARGE SCALE ANALYSIS OF VEHICLES ROBBERY OF THE CITY OF SÃO PAULO USING K-MEASURES AND APRIORI

Fabio Silva Lopes, 0000-0001-8274-7682, (Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil) - megaflopes@gmail.com
Mario Leandro Pires Toledo, 0000-0001-9893-110X, (Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil) - mariotoledo12@gmail.com
Ricardo Kanazawa, 0000-0002-7251-6140, (Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil) - ricardokanazawa@gmail.com
Rosinei Cristiano Pereira, 0000-0003-3894-8537, (Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil) - rosinei@gmail.com

This work presents a research's result that had the objective of analyzing the data of car robbery in São Paulo in a certain set of days, in order to identify a pattern among the robberies events presents in the data sample. For this, data were obtained from the São Paulo State Government's Transparency Portal and then the K-Means and Apriori algorithms were applied to classify and group them.

Keywords: *Big Data, K-Means,* Apriori, *Car Robbery*

ANÁLISE EM LARGA ESCALA DOS ROUBOS DE VEÍCULOS DA CIDADE DE SÃO PAULO UTILIZANDO K-MÉDIAS E APRIORI

Este trabalho apresenta o resultado de uma pesquisa que teve como objetivo analisar os dados dos eventos de roubos de carros em São Paulo em um determinado intervalo de dias, a fim de identificar um padrão entre os roubos presentes na amostra de dados. Para isto, foram utilizados dados do Portal da Transparência do Governo do Estado de São Paulo e aplicados os algoritmos de K-Médias e *Apriori* para a classificação e agrupamento dos dados.

Palavras-chave: *Big Data, K-Means,* Apriori, Roubo de carros

1. Introdução

De acordo com o Fórum Brasileiro de Segurança Pública (2016), mais de 509 mil de carros foram roubados ou furtados no Brasil no ano de 2015, resultando em uma média de 1 caso de furto ou roubo por minuto. Em valores absolutos, o Estado de São Paulo aparece como o maior estado brasileiro em números de roubos e furtos de carros, contando com 189 mil casos em 2015.

Um estudo do Banco Interamericano de Desenvolvimento (2017) afirma que a violência e a criminalização são as principais barreiras do desenvolvimento da América Latina, com custos anuais equivalentes a 10,5% do PIB do Brasil. Para Madalozzo e Furtado (2011), o entendimento do que é um crime é essencial para que melhores práticas de combate e prevenção sejam implementadas para que haja redução das ocorrências e bem estar dos indivíduos. Nesse aspecto, o crime pode ser examinado de formas diferentes, considerando todas as modalidades de crime, incluindo de furto e roubo.

A fim de disponibilizar publicamente dados de servidores públicos estaduais, o Governo do Estado de São Paulo (n.d.) possibilita visualizar dados de roubos e furtos de veículos filtrados por períodos de ano e mês através do Portal da Transparência, permitindo a exportação dos mesmos em um formato de planilha contendo registros dos Boletins de Ocorrência dos casos.

Segundo o manual da Secretaria da Segurança Pública (2005), disponibilizado para avaliar corretamente a evolução da criminalidade e a atuação da Polícia, é preciso ter o entendimento correto da natureza dos crimes e para que estes façam parte das estatísticas oficiais, é preciso que três etapas sejam cumpridas: o crime deve ser detectado, notificado às autoridades policiais e por último registrado no boletim de ocorrência. O manual menciona ainda que pesquisas de vitimização realizadas no Brasil sugerem, em média, que apenas um terço dos crimes ocorridos são registrados. Todavia, de acordo com a UNICRI - Instituto Europeu de Criminologia da ONU, a taxa de notificação para roubos de carros no Brasil com mais de 100 mil habitantes possui um percentual significativo, conforme demonstrado pela Tabela 1.

Tabela 1 - Taxa de notificação de cidades com mais de 100 mil habitantes.

Descrição	**Inglat.**	**Finlândia**	**Esp.**	**Itália**	**C. Rica**	**Brasil**	**Argen.**
Roubo de carro	93,9	100	80,9	94,9	73,7	91,9	90,3
Furto de dentro do carro	74,3	5	29,2	40,1	22,1	18,3	53,8
Vandalismo no	35,5	6,1	18,4	14,9	18,2	0,9	18,8

carro							
Roubo de moto	93,5	5,7	85,4	76,4	91,7	65	79,5
Roubo de bicicleta	74,6	4,6	40,9	27,5	35,7	7,1	41,4
Arrombamento	94,6	5	70,8	65,5	50,8	38,4	68,9
Tentativa de arrombamento	55,2	2,2	22,5	20,9	22,5	19,3	40,9
Assalto	52,1	8,6	32,1	37,5	27,6	19,1	42
Ofensas sexuais	16,4	1,2	3,6	4,3	9,3	9,8	43
Agressão / ameaça	41,7	4,4	24,4	25,4	29,9	11,5	34,4

Fonte: UNICRI / ILANUD

2. Objetivo

Dado o cenário apresentado, o objetivo deste trabalho é analisar os dados de roubos de veículos da cidade de São Paulo a fim de identificar um padrão entre os roubos realizados dentro de um determinado período de dias. Para tal, são utilizadas técnicas de processamento de dados em larga escala em dados públicos de roubos de veículos coletados através do Portal da Transparência do Governo do Estado de São Paulo (n.d.) no mês de janeiro do ano de 2017.

Através de uma análise investigativa realizada nos dados obtidos, a hipótese formulada por este trabalho é que existe uma homogeneidade de grupos de variáveis dentro do conjunto estudado considerando o bairro, período do dia e dia da semana da ocorrência. Para identificar esta hipótese, o algoritmo de K-Médias (Macqueen, 1967) são utilizados para o agrupamento dos registros e, em seguida, o algoritmo de classificação de Apriori (Agrawal e Srikant, 1994) é utilizado para identificar a relação entre os grupos identificados.

3. Metodologia

A linguagem R (Ihaka e Gentleman, 1996) foi utilizada para permitir um melhor tratamento nos dados coletados juntamente com a ferramenta *RStudio*[1]. Os itens a seguir

[1] https://www.rstudio.com/

apresentam, passo a passo, os métodos utilizados com exemplos de aplicação na linguagem escolhida.

3.1 Coleta

Os dados coletados são de fontes secundárias disponíveis no Portal da Transparência do Governo do Estado de São Paulo (n.d.) e pelo Sistema Atlas Municipal (n.d.). A primeira amostra foi obtida aplicando-se os seguintes filtros:

- **Opção**: Roubo de Veículos
- **Departamento**: DECAP
- **Seccional**: Todas
- **Período**: 01/01/2017 - 31/01/2017

O site disponibiliza a exportação de dados para formato excel (.xls) e foi necessário converter o arquivo exportado para o formato .csv com encoding *UTF-8* para melhor interpretação da ferramenta *RStudio* e reconhecimento de caracteres com acentuação.

A amostra complementar contendo os dados de IDH do ano de 2000 dos distritos da cidade de São Paulo foi extraída manualmente do Sistema Atlas Municipal (n.d.)

Tabela 2 - Dados de roubos de carros importados para a ferramenta RStudio

NR BO	**LOGRADOURO**	**BAIRRO**	**CEP**	**CIDADE**
NA	RUA GUILHOBEL	SAUDE	4304020	S.PAULO
NA	RUA ITAJUIBE	ITAIM PAULISTA	8120470	S.PAULO
NA	RUA ITAJUIBE	ITAIM PAULISTA	8120470	S.PAULO
NA	RUA ITAJUIBE	ITAIM PAULISTA	8120470	S.PAULO
NA	RUA ITAJUIBE	ITAIM PAULISTA	8120470	S.PAULO
NA	AVENIDA COMENDADOR FEIZ ZARZUR	PIRITUBA	2942000	S.PAULO

NA	RUA LUIS CARLOS GENTILE DE LAET	MANDAQUI	2378000	S.PAULO
NA	RUA LUIS CARLOS GENTILE DE LAET	MANDAQUI	2378000	S.PAULO
NA	RUA LEONARDO VILAS BOAS	SAO LUCAS	3240000	S.PAULO
NA	RUA CHAFARIZ DE PEDRA	IGUATEMI	8341120	S.PAULO

Fonte: Elaborado pelos autores

3.2 Análise Exploratória

Com os dados importados para a ferramenta *RStudio*, foi realizado uma análise exploratória a fim de identificar maiores detalhes sobre os dados a serem trabalhados, em conjunto foi utilizada a ferramenta *OpenRefine*[2] para obter uma melhor visualização dos dados inicialmente.

3.3 Estatística Descritiva

A amostra selecionada compreende um conjunto de 3514 elementos com 51 variáveis (colunas), que representam o registro de boletins de ocorrência dos roubos de veículos.

As 51 colunas estão distribuídas em variáveis quantitativas e qualitativas conforme mostrado na tabela abaixo (Tabela 3).

Tabela 3 - Exemplo parcial de identificação das variáveis

Coluna	**Tipo do dado**	**Tipo da variável**
CORCUTIS	character	Não há dados para assumir o tipo de variável

[2] http://openrefine.org/

BO_AUTORIA	character	Qualitativa Nominal
FLAGRANTE	character	Qualitativa Nominal
LOGRADOURO	integer	Qualitativa Nominal
NUMERO	integer	Qualitativa Nominal
BAIRRO	character	Qualitativa Nominal
CIDADE	character	Qualitativa Nominal
UF	character	Qualitativa Nominal
DESCRICAOLOCAL	character	Qualitativa Nominal
EXAME	character	Qualitativa Nominal

Fonte: Elaborado pelos autores

Dado o objetivo do projeto em identificar uma relação entre as variáveis e os roubos de veículos na amostra selecionada, as seguintes colunas foram identificadas como as mais importantes para os resultados da pesquisa:

- **Data de Ocorrência**: Data do registro da ocorrência, no formato de data DD/MM/YYYY
- **Período**: Período da ocorrência, variando entre MANHÃ, TARDE e NOITE
- **Bairro**: Texto identificando o bairro da ocorrência

3.4 Pré-Processamento

A etapa de pré-processamento é necessária para consolidar e transformar informações em um formato compreensível para os algoritmos durante a análise dos dados.

3.4.1 Remoção de placas repetidas

Foram identificados 280 registros de placas repetidas e registros sem placas que foram removidos da amostragem.

3.4.2 Extração de colunas

Com base na Análise Exploratória, as colunas com maior importância na base foram mantidas, enquanto as colunas com informações irrelevantes foram isoladas em um outro conjunto de dados.

Tabela 4 - Exemplo parcial das informações extraídas

DATA OCORRENCIA	**PERIODO OCORRENCIA**	**BAIRRO**
31/12/2016	A NOITE	SAUDE
31/12/2016	A NOITE	ITAIM PAULISTA
31/12/2016	A NOITE	ITAIM PAULISTA
31/12/2016	A NOITE	ITAIM PAULISTA
31/12/2016	A NOITE	ITAIM PAULISTA
01/01/2017	DE MADRUGADA	PIRITUBA
31/12/2016	A NOITE	MANDAQUI
01/01/2017	DE MADRUGADA	SAO LUCAS
01/01/2017	DE MADRUGADA	IGUATEMI
01/01/2017	PELA MANHÃ	SOCORRO

Fonte: Elaborado pelos autores

3.4.3 Ajustes dos caracteres

Todos os valores texto foram transformados para caixa alta e foram removidos caracteres com acentuação a fim de evitar coerção de dados.

Tabela 5 - Exemplo parcial das informações extraídas

DATA OCORRENCIA	**PERIODO OCORRENCIA**	**BAIRRO**
31/12/2016	A NOITE	SAUDE
31/12/2016	A NOITE	ITAIM PAULISTA
31/12/2016	A NOITE	ITAIM PAULISTA

31/12/2016	A NOITE	ITAIM PAULISTA
31/12/2016	A NOITE	ITAIM PAULISTA
01/01/2017	DE MADRUGADA	PIRITUBA
31/12/2016	A NOITE	MANDAQUI
01/01/2017	DE MADRUGADA	SAO LUCAS
01/01/2017	DE MADRUGADA	IGUATEMI
01/01/2017	PELA MANHÃ	SOCORRO

Fonte: Elaborado pelos autores

3.4.4 Valores Nulos

Para garantir a conformidade dos resultados dos algoritmos, foi necessário efetuar a limpeza dos dados verificando a existência de algum valor nulo em todos os registros do conjunto, descartando registros com uma única variável nula.

Ao final deste processo, a base contou com 3040 registros, uma diferença de 199 registros da base original.

- **A NOITE**: 87 registros excluídos
- **A TARDE**: 39 registros excluídos
- **DE MADRUGADA**: 31 registros excluídos
- **PELA MANHÃ**: 37 registros excluídos

3.4.5 Informações inconsistentes

Durante a análise exploratória, foram identificados registros com a data da ocorrência fora do período estudado. Desta forma, foram filtrados os registros cuja a data de ocorrência era superior à 1 de janeiro de 2017.

3.4.6 Conversão de variáveis

Através da análise exploratória, foram identificadas colunas importantes para o resultado do processamento dos algoritmos. Considerando a diferença entre os tipos das variáveis, e a fim da execução de algoritmos de classificação, todas as variáveis foram convertidas para quantitativas discretas, conforme descritas nos itens abaixo.

3.4.6.1 Bairro

Os dados da variável Bairro foram classificados utilizando os dados de Índice de Desenvolvimento Humano do IBGE[3], classificando os bairros com IDH "Muito Elevado", "Elevado" e "Médio" de acordo como apresentado o site Atlas Brasil[4]:

- **IDH Muito Elevado**: a partir de 0,800
- **IDH Elevado**: Entre 0,700 e 0,799
- **IDH Médio**: Entre 0 e 0,699

Tabela 5 - Exemplo parcial das informações após classificação por IDH

DATA OCORRENCIA	**PERIODO OCORRENCIA**	**BAIRRO**	**BAIRRO IDH**
31/12/2016	A NOITE	SAUDE	MUITO ELEVADO
31/12/2016	A NOITE	ITAIM PAULISTA	MEDIO
31/12/2016	A NOITE	ITAIM PAULISTA	MEDIO
31/12/2016	A NOITE	ITAIM PAULISTA	MEDIO
31/12/2016	A NOITE	ITAIM PAULISTA	MEDIO
01/01/2017	DE MADRUGADA	PIRITUBA	ELEVADO
31/12/2016	A NOITE	MANDAQUI	ELEVADO
01/01/2017	DE MADRUGADA	SAO LUCAS	ELEVADO
01/01/2017	DE MADRUGADA	IGUATEMI	MEDIO
01/01/2017	PELA MANHÃ	SOCORRO	ELEVADO

Fonte: Elaborado pelos autores

3.4.6.2 Data de Ocorrência

[3] https://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=355030&search=sao-paulo|sao-paulo
[4] http://www.atlasbrasil.org.br/2013/pt/o_atlas/idhm/

A coluna data de ocorrência foi classificada pelo seu dia da semana, variando de segunda-feira à domingo.

3.4.7 Transformação

Para que os algoritmos de agrupamento tenham um resultado eficiente, foi necessário a transformação das variáveis selecionadas de qualitativas nominais para qualitativas discretas.

Tabela 6 - Exemplo parcial das informações após transformação dos dados

DIA_SEMANA	PERIODO_OCORRENCIA	IDH
6	3	1
6	3	3
6	3	3
6	3	3
6	3	3
7	4	2
6	3	2
7	4	2
7	4	3
7	1	2

Fonte: Elaborado pelos autores

3.4.7.1 Normalização

Com a classificação das variáveis, os dados foram então normalizados utilizando o método de Minmax (Hazewinkel, 1994) para cada coluna.

Tabela 7 - Exemplo parcial das informações normalizadas utilizando Minmax.

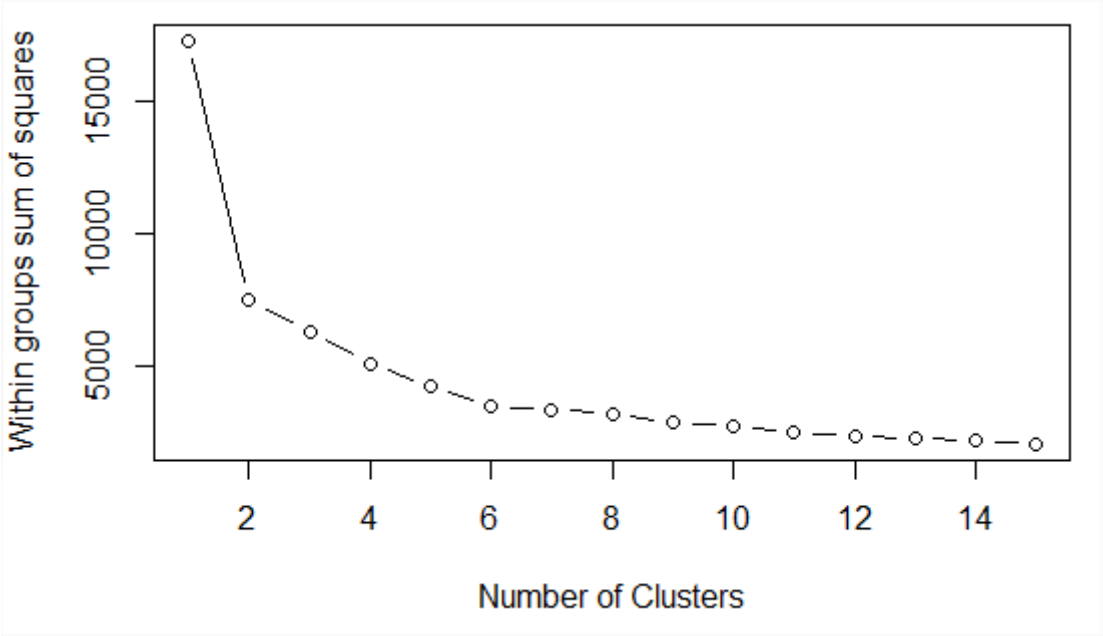
DIA_SEMANA	**PERIODO_OCORRENCIA**	**IDH**
0.8333333	0.50	0.0000000
0.8333333	0.50	0.6666667
0.8333333	0.50	0.6666667
0.8333333	0.50	0.6666667
0.8333333	0.50	0.6666667
1.0000000	0.75	0.3333333
0.8333333	0.50	0.3333333
1.0000000	0.75	0.3333333
1.0000000	0.75	0.6666667
1.0000000	0.00	0.3333333

Fonte: Elaborado pelos autores

3.5 Agrupamento

Após a normalização dos dados, foi possível utilizar os algoritmos de agrupamento na amostra selecionada. Utilizando o algoritmo de k-partições, foi possível identificar um k único representando o número de clusters da amostra no maior ponto de curvatura da parábola (Figura 1).

Figura 1 - Plot da soma dos quadrados por grupo (clusters)



Fonte: Elaborado pelos autores

Utilizando k = 4, foi possível utilizar o algoritmo de k-médias, identificando assim os clusters e seus respectivos registros. Em seguida, os respectivos grupos foram identificados nos registros da amostra, conforme Tabela 8.

Tabela 8 - Informações parciais da clusterização

BAIRRO	BAIRRO IDH	DIA SEMANA	DIA SEMANA C	PERIODO OCORRENCIA C	BAIRRO IDH_C	CLUSTER
SAUDE	MUITO ELEVADO	sábado	6	3	1	4
ITAIM PAULISTA	MEDIO	sábado	6	3	3	4
ITAIM PAULISTA	MEDIO	sábado	6	3	3	4
ITAIM PAULISTA	MEDIO	sábado	6	3	3	4
ITAIM PAULISTA	MEDIO	sábado	6	3	3	4
PIRITUBA	ELEVADO	domingo	7	4	2	4
MANDAQUI	ELEVADO	sábado	6	3	2	4

SAO LUCAS	ELEVADO	domingo	7	4	2	4
IGUATEMI	MEDIO	domingo	7	4	3	4
SOCORRO	ELEVADO	domingo	7	1	2	3

Fonte: Elaborado pelos autores

3.6 Classificação

Com os clusters identificados, foi necessário o uso do algoritmo de Apriori para identificar as variáveis de maior ocorrência dentro de cada cluster, identificando assim sua homogeneidade.

Foram realizadas 3 interações do algoritmo, utilizando o parâmetro de suporte com os valores de 15%, 20% e 25%.

Tabela 9 - Exemplo de regras geradas utilizando suporte de 15%

ID	***Rules***	***Suport***	***Confidence***	***Lift***
1	{sábado , PELA MANHÃ , ELEVADO} => {1}	0.01513158	1	2.892483
2	{segunda-feira , A NOITE , MEDIO} => {4}	0.01513158	1	3.932730
3	{segunda-feira , PELA MANHÃ , MEDIO} => {3}	0.01546053	1	4.222222
4	{quinta-feira , PELA MANHÃ , ELEVADO} => {3}	0.01546053	1	4.222222
5	{quarta-feira , A TARDE , ELEVADO} => {3}	0.01578947	1	4.222222
6	{sábado , DE MADRUGADA , ELEVADO} => {2}	0.01578947	1	6.092184
7	{segunda-feira , PELA MANHÃ , ELEVADO} => {3}	0.01611842	1	4.222222
8	{domingo , A TARDE , ELEVADO} => {1}	0.01644737	1	2.892483
9	{terça-feira , A TARDE , ELEVADO} => {3}	0.01644737	1	4.222222

10	{domingo , DE MADRUGADA , MEDIO} => {1}	0.01644737	1	2.892483

Fonte: Elaborado pelos autores

Tabela 10 - Exemplo de regras geradas utilizando suporte de 20%

ID	***Rules***	***Suport***	***Confidence***	***Lift***
1	{terça-feira , A NOITE , MEDIO} => {4}	0.02171053	1	3.932730
2	{terça-feira , A NOITE , ELEVADO} => {4}	0.02335526	1	3.932730
3	{segunda-feira , A NOITE , ELEVADO} => {4}	0.02467105	1	3.932730
4	{sexta-feira , A NOITE , MEDIO} => {1}	0.02565789	1	2.892483
5	{domingo , A NOITE , MEDIO} => {1}	0.02565789	1	2.892483
6	{quinta-feira , A NOITE , ELEVADO} => {2}	0.02631579	1	6.092184
7	{domingo , A NOITE , ELEVADO} => {1}	0.02960526	1	2.892483
8	{sábado , A NOITE , ELEVADO} => {1}	0.03092105	1	2.892483
9	{sábado , A NOITE , MEDIO} => {1}	0.03421053	1	2.892483
10	{quarta-feira , A NOITE , ELEVADO} => {4}	0.03421053	1	3.932730

Fonte: Elaborado pelos autores

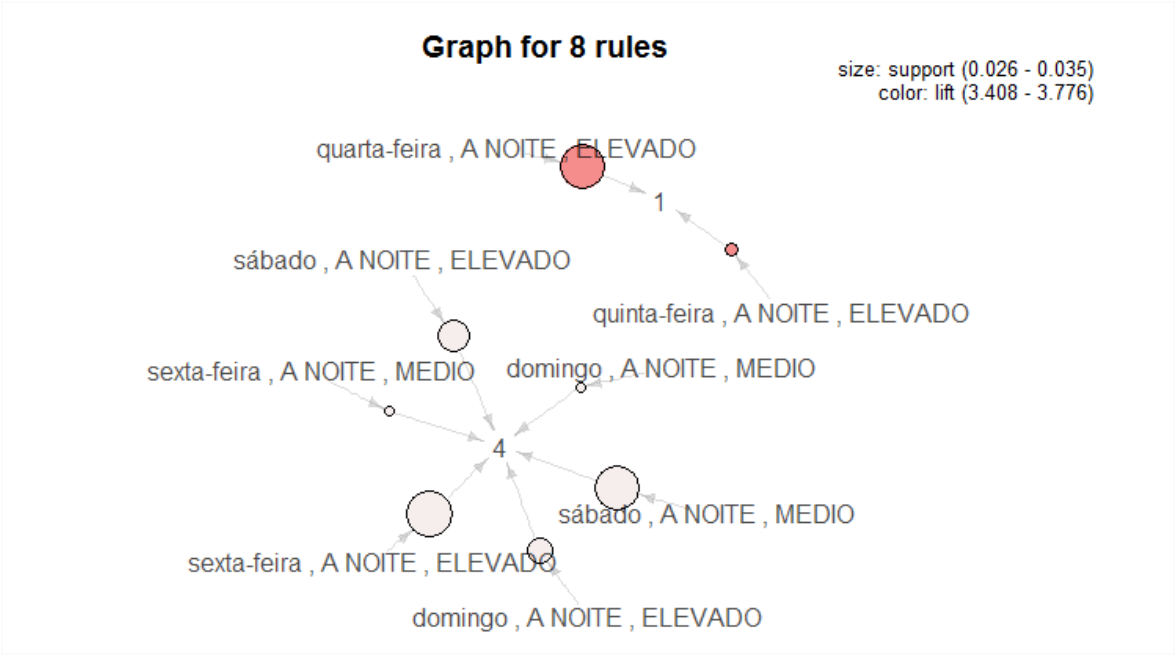
Tabela 11 - Exemplo de regras geradas utilizando suporte de 25%

ID	***Rules***	***Suport***	***Confidence***	***Lift***
1	{sexta-feira , A NOITE , MEDIO} => {1}	0.02565789	1	2.892483

2	{domingo , A NOITE , MEDIO} => {1}	0.02565789	1	2.892483
3	{quinta-feira , A NOITE , ELEVADO} => {2}	0.02631579	1	6.092184
4	{domingo , A NOITE , ELEVADO} => {1}	0.02960526	1	2.892483
5	{sábado , A NOITE , ELEVADO} => {1}	0.03092105	1	2.892483
6	{sábado , A NOITE , MEDIO} => {1}	0.03421053	1	2.892483
7	{quarta-feira , A NOITE , ELEVADO} => {4}	0.03421053	1	3.932730
8	{sexta-feira , A NOITE , ELEVADO} => {2}	0.03453947	1	6.092184
NA	NA	NA	NA	NA
NA.1	NA	NA	NA	NA

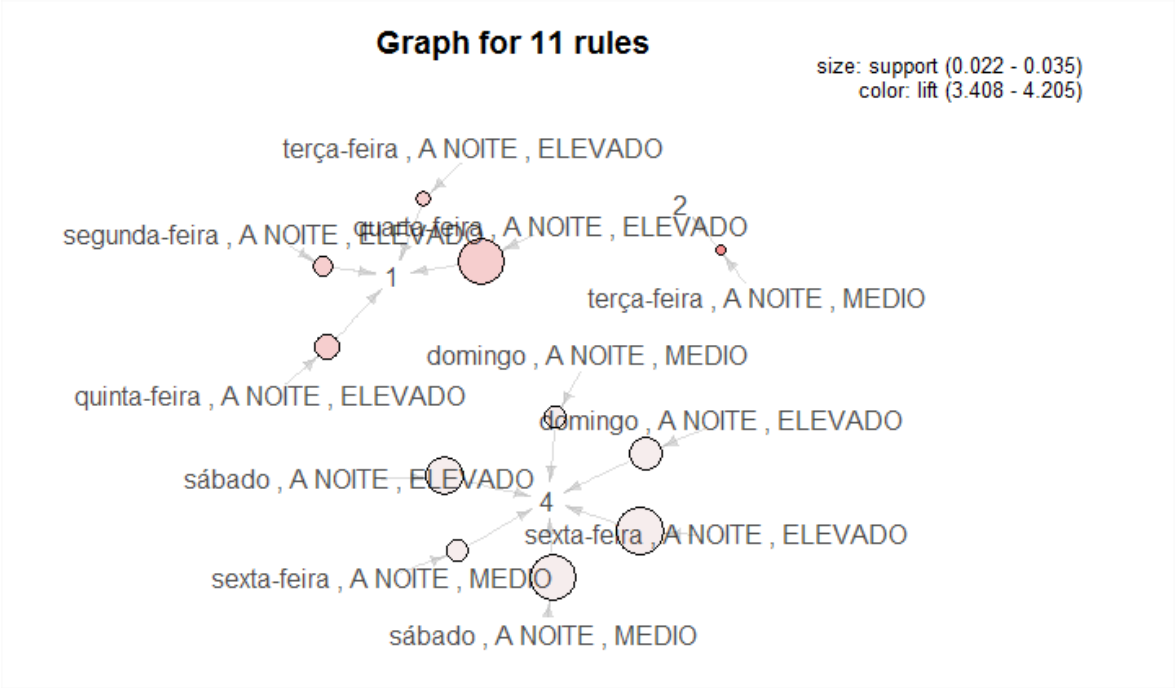
Fonte: Elaborado pelos autores

Figura 2 - Plot das regras com 15% de suporte



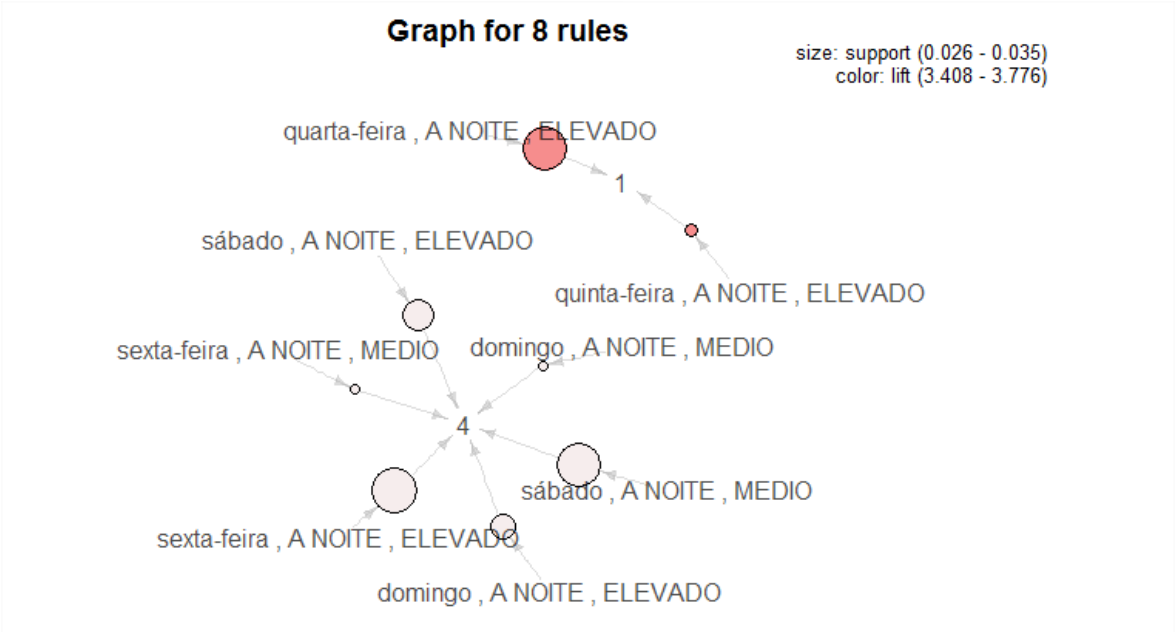
Fonte: Elaborado pelos autores

Figura 3 - Plot das regras com 20% de suporte



Fonte: Elaborado pelos autores

Figura 4 - Plot das regras com 25% de suporte



Fonte: Elaborado pelos autores

4. Conclusão

Este trabalho propôs o uso de técnicas de análise de dados em larga escala para identificar uma homogeneidade entre variáveis de roubos de veículos na cidade São Paulo. Para tal, foram coletados dados do Governo do Estado de São Paulo (n.d.) e realizado a limpeza e transformação dos dados para, em seguida, utilizar o algoritmo de k-médias para agrupá-los em k-clusters e o algoritmo de Apriori para classificar cada cluster pela maior relação de variáveis.

Utilizando os gráficos formados a partir do agrupamento pelo algoritmo de Apriori, foi possível identificar a seguinte relação homogênea dentro dos 4 clusters obtidos:

- No primeiro cluster, foi possível identificar uma maior densidade de casos ocorridos de sexta-feira a domingo à noite com predominância nos bairros de IDH classificados como Médio;
- No segundo cluster, foi possível identificar uma densidade grande nas noites de quinta-feira e sexta-feira em bairros de IDH classificados como elevado;
- No terceiro cluster, foi possível identificar uma densidade baixa de casos ocorridos durante a semana entre os períodos da manhã e tarde, com predominância entre bairros de IDH classificados como Elevado;
- No quarto cluster, foi possível identificar uma densidade alta de casos ocorridos durante as quartas no período na noite entre bairros de IDH classificados como Elevado;

A partir dos resultados obtidos, este trabalho sugere uma ação da polícia do estado de São Paulo de remanejamento de seus recursos da seguinte forma:

- Remanejar o contingente para priorizar os bairros de IDH classificados como elevado durante os dias de semana, nos períodos de manhã e tarde;
- Aumentar a contingência da polícia em bairros de IDH classificados como elevado entre as quintas-feiras e sextas-feiras à noite, e nos sábados de madrugada; também elevar o contingente em bairros de IDH classificados como elevado durante as quartas-feiras à noite;

5. Trabalhos futuros

A partir dos resultados encontrados, sugere-se os seguintes trabalhos futuros para este projeto:

- Adicionar o modelo veicular no modelo a fim de identificar uma relação de furtos com com a faixa do valor do automóvel;
- Abranger o período para outros meses ainda na cidade de São Paulo e verificar se os resultados são alterados de acordo com o mês;
- Identificar uma possível relação de geolocalização entre casos ocorridos no mesmo dia a fim de verificar casos relacionados (exemplo: identificar quadrilhas de roubo e receptação de veículos);
- Adicionar variáveis de geo-localização no modelo e relacionar com estabelecimentos de grande circulação nas proximidades, como shoppings, universidades, supermercados, estádios e avenidas com maior circulação;

- Relacionar a evolução dos casos de roubo de veículos com casos de homicídios a partir da geo-localização e período da ocorrência
- Identificar a proporção de flagrantes nas ocorrências de roubos e furtos
- Identificar a proporção de apreensão de entorpecentes ou armas de fogo e verificar se casos de roubo ocorrem com maior frequência quando há a relação das duas variáveis;

6. Referências

Agrawal, Rakesh; Srikant, Ramakrishnan, *Fast algorithms for mining association rules*, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, 1994.

Banco Interamericano de Desenvolvimento, *Crime acarreta custos sociais, públicos e privados na América Latina e Caribe: estudo do BID*, Disponível em: https://publications.iadb.org/bitstream/handle/11319/8133/Os-custos-do-crime-e-da-violencia_novas-evidencias-e-constatacoes-na-America-Latina-e-Caribe.pdf?sequence=9 , 2017.

Fórum Brasileiro de Segurança Pública, *10º anuário brasileiro da segurança pública*, Disponível em: http://www.forumseguranca.org.br/publicacoes/10o-anuario-brasileiro-de-seguranca-publica/, 2016.

Governo do Estado de São Paulo, *Portal da transparência de segurança Pública*, http://www.ssp.sp.gov.br/transparenciassp/.

Ihaka, Ross; Gentleman, Robert, *R: A language for data analysis and graphics*, The Journal of Computational and Graphical Statistics, 1996.

Macqueen, J.; *Some methods for classification and analysis of mulivariate observations*, In: L.M. LeCam, J. Neyman (eds.): Proc. 5th Berkely Symp. Math. Statist. Probab. 1965/66. Univ. of California Press, Berkely, vol. I, 281- 297, 1967.

Madalozzo, R.; Furtado, G. M., *Um estudo sobre a vitimização para a cidade de São Paulo*, Revista de Economia Política, vol. 31, nº 1 (121), pp. 160-180, 2011.

Hazewinkel, M., Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic, 1994;

Secretaria da segurança pública, *Estatística de criminalidade manual de interpretação*, Disponível em: http://www.ssp.sp.gov.br/Estatistica/download/manual.pdf, 2005.

Sistema Atlas Municipal. n.d. "Sistema Atlas Municipal." http://atlasmunicipal.prefeitura.sp.gov.br (http://atlasmunicipal.prefeitura.sp.gov.br).