Estudo da Correlação entre os Chamados da Empresa de Manutenção e Limpeza Urbana e os Bairros da Cidade do Recife

Gabriel Barbosa, Luís Santos, Thiago Casa Nova

Centro de Informática - Universidade Federal de Pernambuco

Avenida Jornalista Aníbal Fernandes, s/n - Cidade Universitária, Recife - PE

Abstract

Este documento tem como objetivo relatar o processo realizado durante o desenvolvimento de um estudo com o objetivo de relacionar caracteristicas dos bairros da cidade do Recife com os chamados da Empresa de Manutenção e Limpeza Urbana (EMLURB), tema proposto para a cadeira de Introdução à Ciência dos Dados ministrada pelo Professor Renato Vimieiro.

Keywords: Recife, EMLURB, infraestrutura, atendimento, bairro, perfil socioeconômico, região político administrativa (RPA), microrregião

---

1. Hipótese

Inicialmente queremos saber se existem diferenças na resolução de problemas em bairros com diferentes perfis socioeconômicos? Se existem, são recorrentes? Nosso objetivo com este projeto é mapear os problemas e suas características, bem como suas correlações com os diferentes perfis socioeconômicos dos bairros do Recife.

2. Coleta de Dados

Parte dos dados já estavam disponibilizados em forma de *CSV*, no caso, as demandas de atendimento à EMLURB. Para fazermos a análise desejada, precisamos também de informações descritivas de cada bairro para dividí-los em classes socioeconômicas diferentes. Como esses dados não estavam disponíveis em CSV, utilizamos BeautifulSoup para coletá-los diretamente das páginas web correspondentes, exceto em casos limítrofes de formatações de página problemáticas, caso em qual os dados foram postos manualmente.

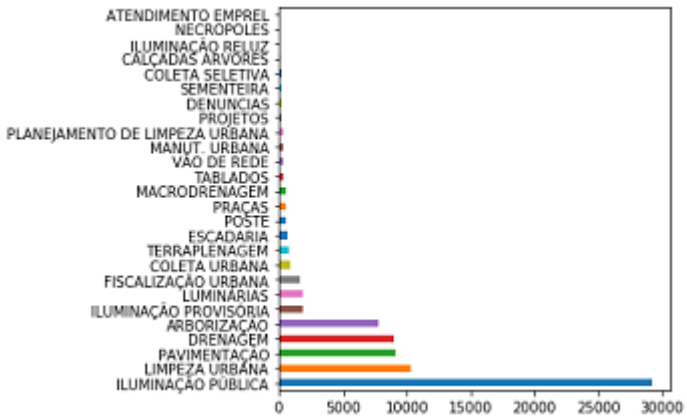
3. Análise de Dados

Dos dados socioeconômicos [1], extraimos informações como população por faixa etária, cor de pele e taxa de alfabetização. Organizamos esses dados em um *Data Frame*.

Em seguida utilizamos os dados de chamados à EMLURB [2] na análise exploratória, começamos por ajustar as colunas de datas para o formato correto, em seguida subtrair as datas dos chamados iniciais das datas de resolução, ou atualização, dos chamados, com isso já era possível ter valores mais fáceis de utilizar. Em seguida plotamos diversos gráficos e dados para termos um panorama mais geral da situação, como a quantidade de serviços e bairros é muito grande, plotamos de início por **Região Político Administrativa (RPA)**, que é um aglomerado de bairros vizinhos que são agrupados para facilitar decisões administrativas, e por Grupo de Serviço, que é um rótulo do *data frame* que agrupa serviços que pertencem a mesma área.

Após a plotagem dos gráficos podemos observar as primeiras características dos dados. O problema mais abundante é de iluminação pública, mais especificamente a troca de lâmpadas defeituosas, sendo dominante em todas as RPA's, chegando a muitas vezes ter o dobro de ocorrências dos problemas que ocupam a segunda posição.

Analisaremos nos dados de demandas de serviços: quais são os tipos de serviços mais solicitados, que lugares pedem mais atendimento e a quantidade de problemas resolvidos por completo, na cidade como um todo e por subrregião da cidade.



Outra característica interessante é que Boa Viagem é o bairro com mais chamados, todavia também é um dos maiores em área, o que nos levou a verificar a relação entre quantidade de chamados e área do bairro. Foi encontrado que com exceção de um bairro, Ibura, todos os demais possuem uma relação **Quantidade de chamados/Área** entre **0.507519** e **15.357143**, com exceção de um bairro que teve uma proporção superior a 115, sendo considerado como um outlier, a média dos bairros é de **7.639661**. Entre os 10 menores bairros **7 ficaram acima da média**

nome_bairro	qtd_chamados	area_hectare	chamados_p_area
TOTO	215.0	14	15.357143
TORREAO	214.0	16	13.375000
IBURA	2188.0	19	115.157895
PONTO DE PARADA	91.0	20	4.550000
JAQUEIRA	265.0	24	11.041667
ROSARINHO	305.0	25	12.200000
ALTO DO MANDU	233.0	25	9.320000
ILHA DO LEITE	337.0	26	12.961538
MANGABEIRA	215.0	29	7.413793
HIPODROMO	216.0	30	7.200000

E entre os maiores, **apenas 1 ficou acima da média**

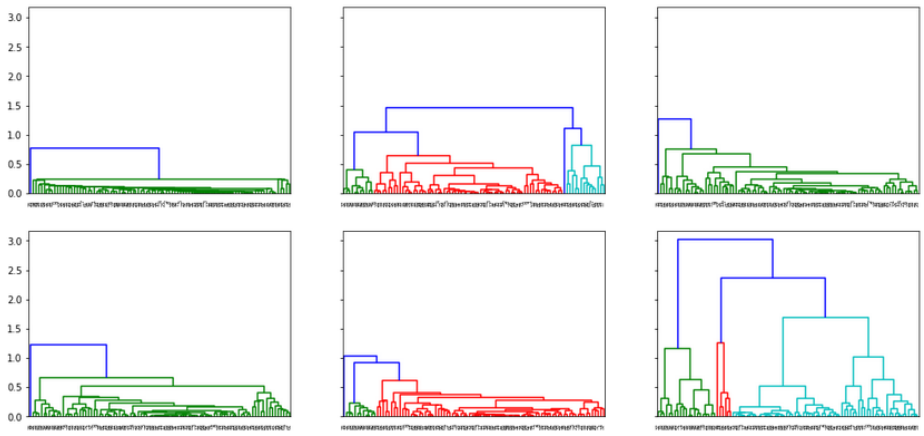
PASSARINHO	818.0	406	2.014778
COHAB	3354.0	426	7.873239
IPUTINGA	1942.0	434	4.474654
BARRO	1136.0	454	2.502203
DOIS IRMAOS	412.0	585	0.704274
GUABIRABA	474.0	617	0.768233
PINA	1373.0	629	2.182830
IMBIRIBEIRA	3031.0	666	4.551051
BOA VIAGEM	4668.0	753	6.199203
CURADO	405.0	798	0.507519

Com isso é possível afirmar que bairros menores realizam, proporcionalmente, mais chamados para EMLURB em comparação a bairros maiores.

Ao aumentarmos a granularidade da análise investigamos as ruas, nas quais pode-se um possível ruído na base, sendo ele a Rua Projetada, uma vez essa é a rua que mais possui chamados, porém quando verificamos essa rua no *data frame* nos deparamos com algo incomum, a rua está presente em diversos bairros em diversas regiões espalhadas pela cidade, podendo ser desde um nome padrão para a rua ou até mesmo, realmente, um nome muito comum para ruas na cidade do Recife. Depois da Rua Projetada podemos ver Avenidas e Ruas bastante conhecidas por serem muito grandes e movimentadas, o que faz sentido com o fato de possuírem diversos chamados na EMLURB.

4. Clustering

Para dividirmos os bairros em classes socioeconômicas, dentre as informações tidas, as mais pertinentes são a área do bairro, o número de habitantes e o rendimento médio dos domicílios. Já que o número ideal de *clusters* era desconhecido, utilizamos um algoritmo de *clustering* hierárquico juntamente a dendrogramas, a fim de determinarmos um número de *clusters* que pudesse balancear as características de cada bairro. Existem seis métodos diferentes que podem ser utilizados para o algoritmo de classificação hierárquica, estes são o **single link**, **complete link**, **group average**, **centroid**, **median** e **ward's method**. Dendrogramas foram plotados para cada um dos métodos, e esses são os respectivos resultados:



Podemos observar que, dos métodos, o que obteve o resultado mais balanceado foi o **Ward's Method**. Realizando um corte na altura 1,5, são obtidos quatro *clusters* de tamanhos 28, 21, 39 e 6, respectivamente. As seguintes tabelas descrevem cada *cluster* em função de certo aspecto:

area_hectare :

	count	mean	std	min	25%	50%	75%	max
cluster								
0	28.0	248.892857	109.034259	19.0	169.5	231.0	329.50	454.0
1	21.0	78.904762	54.146011	16.0	44.0	56.0	102.00	188.0
2	39.0	55.641026	24.854828	14.0	36.0	52.0	69.50	127.0
3	6.0	674.666667	83.514470	585.0	620.0	647.5	731.25	798.0

populacao :

	count	mean	std	min	25%	50%	75%	max
cluster								
0	28.0	28481.714286	18046.440944	602.0	17790.0	30459.5	34725.25	70453.0
1	21.0	8845.428571	8271.436711	72.0	2658.0	5917.0	14124.00	29180.0
2	39.0	8412.000000	4043.594258	285.0	6308.5	8480.0	10996.50	18334.0
3	6.0	37654.000000	44997.184107	2566.0	8852.0	22797.0	43678.00	122922.0

rend_medio :

	count	mean	std	min	25%	50%	75%	max
cluster								
0	28.0	1616.980714	550.576593	567.00	1181.8625	1550.57	2046.1425	2812.73
1	21.0	7153.118571	2354.858100	3618.45	5115.0600	7106.75	9040.7600	11339.79
2	39.0	1598.199231	751.772189	705.83	1049.0800	1296.05	2052.0900	3747.16
3	6.0	2662.498333	2235.829807	1159.26	1396.2950	2022.27	2362.2325	7108.00

A partir dessas tabelas, podemos caracterizar cada *cluster* como segue:

- O primeiro cluster tem bairros com área razoável, população grande e rendimento entre pequeno e razoável.
- O segundo cluster tem bairros de área pequena, população razoável e rendimento médio muito alto.
- O terceiro cluster tem bairros de área pequena, população razoável e rendimento médio razoável.
- O quarto cluster tem bairros de área grande, população razoável e rendimento médio entre razoável e alto.

Feita a divisão, resta verificar, para cada par de classes socioeconômicas estabelecidas, se é possível afirmar que existe diferença estatisticamente significativa entre os tempos de atendimento. Para tal, será utilizado o teste de Mann-Whitney, que tem base no teste de Wilcoxon, mas permite que as amostras comparadas tenham números de exemplos diferentes, que é o caso desses dados pelo número diferente de bairros por clusters e mesmo de pedidos por bairro. São selecionados todos os pedidos já atendidos pela EMLURB, e então subtrai-se a data da demanda de serviço da data de resolução, obtendo-se o tempo de espera para cada pedido. Seleciona-se então o tempo de espera para cada classe socioeconômica, e verifica-se, para cada par de classes (x, y), se x ¡ y. Para rejeitar a hipótese nula, necessita-se de p-values pequenos. Estes são os p-values resultantes para cada par:

- (0, 1): 1,0.

- (0, 2): 1,0.
- (0, 3): 0,9.
- (1, 2): 0,7.
- (1, 3): 0,36e-9
- (2, 3): 3e-11

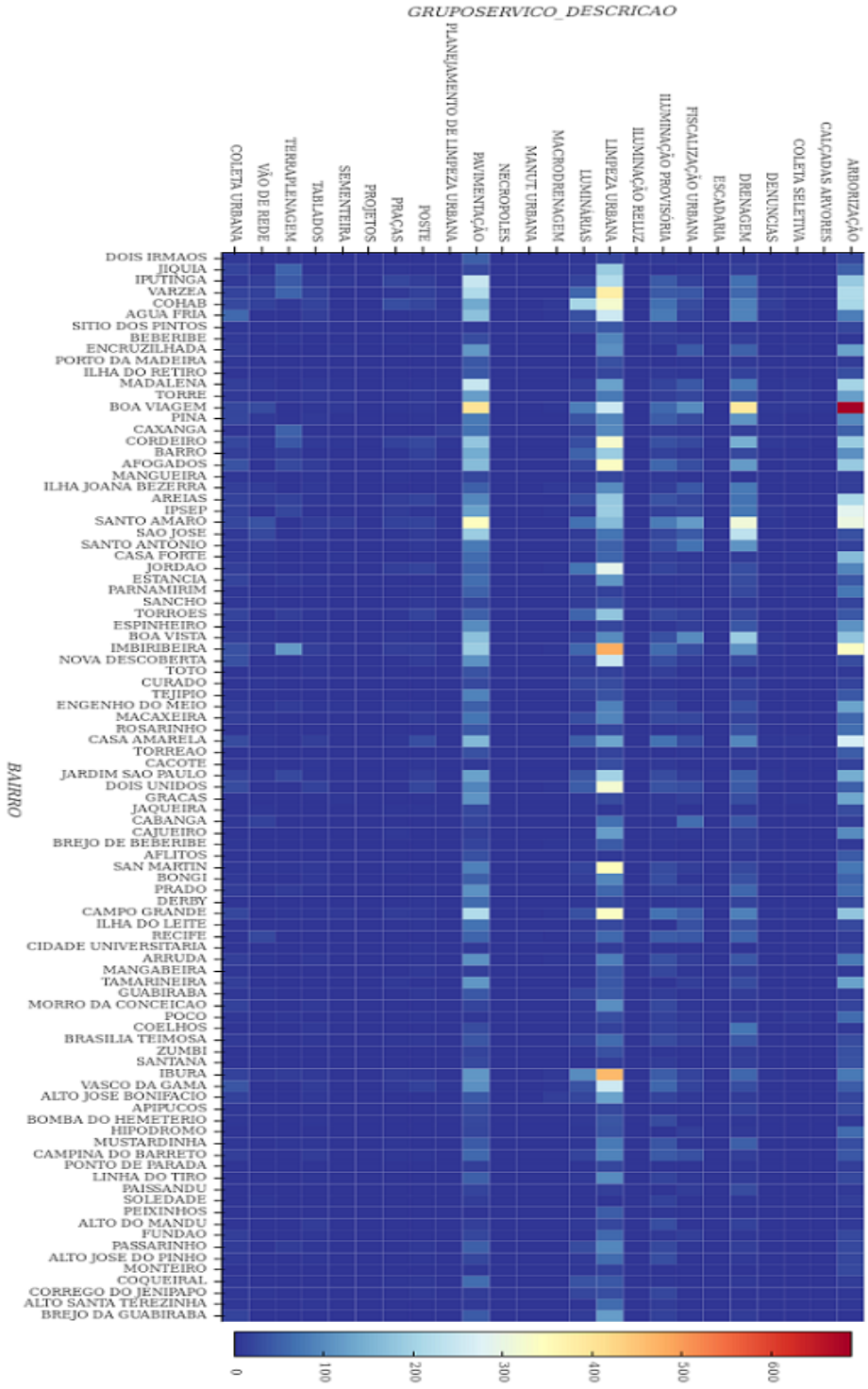
Pode-se rejeitar a hipótese nula somente para os pares (1, 3) e (2, 3), afirmando-se que as classes socioeconômicas 1 e 2 têm tempos de espera menores do que a classe 3. Para investigar mais profundamente, a hipótese alternativa será invertida, e agora, para cada par, será testado se x ¿ y. Estes são os p-values resultantes:

- (0, 1): 1,8e-17
- (0, 2): 5e-24
- (0, 3); 0,07
- (1, 2): 0,29
- (1, 3): 0,99
- (2, 3): 0,99

Pode-se rejeitar a hipótese nula somente para os pares (0, 1) e (0, 2), afirmando-se que as classes socioeconômicas 1 e 2 têm tempos de espera menores do que a classe 0. Nada pode-se afirmar sobre os pares (0, 3) e (1, 2). Como a característica dos bairros de classe 1 e de classe 2 são áreas pequenas, e dos bairros de classe 0 e classe 3 são áreas grandes, pode-se dizer que bairros maiores esperam mais por atendimentos da EMLURB.

5. Heat Map

Por fim, foi criado um Heat Map para representar a quantidade de pedidos por bairro e grupo de serviço, com exceção de serviços de iluminação pública, que dominam os dados em quantidade.


GRUPOSERVICO_DESCRICAO
COLETA URBANA
VÃO DE REDE
TERRAPLENAGEM
TABLADOS
SEMENTEIRA
PROJETOS
PRAÇAS
POSTE
PLANEJAMENTO DE LIMPEZA URBANA
PAVIMENTAÇÃO
NECROPOLES
MANUT URBANA
MACRODRENAGEM
LUMINÁRIAS
LIMPEZA URBANA
ILUMINAÇÃO RELUZ
ILUMINAÇÃO PROVISÓRIA
FISCALIZAÇÃO URBANA
ESCADARIA
DRENAGEM
DENUNCIAS
COLETA SELETIVA
CALÇADAS ARVORES
ARBORIZAÇÃO
BAIRRO
DOIS IRMAOS
JIQUIA
IPUTINGA
VARZEA
COHAB
AGUA FRIA
SITIO DOS PINTOS
BEBERIBE
ENCRUZILHADA
PORTO DA MADEIRA
ILHA DO RETIRO
MADALENA
TORRE
BOA VIAGEM
PINA
CAXANGA
CORDEIRO
BARRO
AFOGADOS
MANGUEIRA
ILHA JOANA BEZERRA
AREIAS
IPSEP
SANTO AMARO
SAO JOSE
SANTO ANTÔNIO
CASA FORTE
JORDAO
ESTANCIA
PARNAMIRIM
SANCHO
TORROES
ESPINHEIRO
BOA VISTA
IMBIRIBEIRA
NOVA DESCOBERTA
TOTO
CURADO
TEJIPIO
ENGENHO DO MEIO
MACAXEIRA
ROSARINHO
CASA AMARELA
TORREAO
CACOTE
JARDIM SAO PAULO
DOIS UNIDOS
GRACAS
JAQUEIRA
CABANGA
CAJUEIRO
BREJO DE BEBERIBE
AFLITOS
SAN MARTIN
BONGI
PRADO
DERBY
CAMPO GRANDE
ILHA DO LEITE
RECIFE
CIDADE UNIVERSITARIA
ARRUDA
MANGABEIRA
TAMARINEIRA
GUABIRABA
MORRO DA CONCEICAO
POCO
COELHOS
BRASILIA TEIMOSA
ZUMBI
SANTANA
IBURA
VASCO DA GAMA
ALTO JOSE BONIFACIO
APIPUCOS
BOMBA DO HEMETERIO
HIPODROMO
MUSTARDINHA
CAMPINA DO BARRETO
PONTO DE PARADA
LINHA DO TIRO
PAISSANDU
SOLEDADE
PEIXINHOS
ALTO DO MANDU
FUNDAO
PASSARINHO
ALTO JOSE DO PINHO
MONTEIRO
COQUEIRAL
CORREGO DO JENIPAPO
ALTO SANTA TEREZINHA
BREJO DA GUABIRABA
0
100
200
300
400
500
600

6. Análise de Resultados

A hipótese inicial proposta foi de verificar se bairros com indicadores socioeconômicos possuem algum tipo de preferência na resolução de seus chamados, com uma análise inicial logo foi possível verificar que essa hipótese não condizia com a realidade. Ao analisar os dados de outra maneira, já com o apoio dos clusters gerados verificamos a relação de proximidade entre bairros de tamanhos semelhantes, pequenos com pequenos e grandes com grandes, cruzando isso com os dados percebemos que os maiores bairros geram proporcionalmente menos chamados que a média da cidade e que os bairros menores geram proporcionalmente mais chamados que a média da cidade.

Em posse de ambas as informações podemos inferir que bairros pequenos geram chamados mais frequentemente, com isso o tempo de resposta da EMLURB diminui, uma vez que esses problemas já são recorrentes, enquanto nos bairros maiores os problemas são mais dispersos o que acaba gerando um tempo maior de resposta para resolução dos mesmos.

Referências

[1] Prefeitura do Recife (2010), Perfil dos bairros (http://www2.recife.pe.gov.br/servico/perfil-dos-bairros).

[2] Prefeitura do Recife (2018), Central de atendimento de serviços da EMLURB (http://dados.recife.pe.gov.br/dataset/central-de-atendimento-de-servicos-da-emlurb-156).