

# Análise das Solicitações 156 da Cidade do Recife

Lucas Alves Rufino (lar) e Rodrigo de Lima Oliveira (rlo)

*Universidade Federal de Pernambuco*

---

## Abstract

Este projeto tem o intuito de realizar uma análise sobre os dados das solicitações 156 da cidade do Recife sobre serviços cotidianos fornecidos pela prefeitura, associando informações como dados de localização, temporal e socioeconômicos, a fim de promover a identificação de características que descrevem os dados, assim como a retirada de *insights*.

*Keywords:* Solicitações 156, Recife, Emlurb, Dados Públicos, localização, socioeconômico, temporal, pluviométrico.

---

## 1. Introdução

Com o objetivo de realizar uma análise sobre os dados das solicitações 156 da cidade do Recife, sobre os serviços cotidianos fornecidos pela prefeitura. Este relatório descreve as etapas de coleta, integração de dados de diversas fontes, pré-processamento das bases, aplicação de análises exploratória sobre hipóteses, assim como detecção de padrões nos dados.

Os dados que estão sendo estudados, traz uma informação que é utilizada para detectar necessidades da população, e fornecer melhores serviços a todos os cidadãos. O resultado da análise sobre os dados das solicitações 156 do Recife, é uma informação de interesse dos cidadãos, assim como do estado.

## 2. Metodologia

Os dados que foram analisados tratam sobre registros das solicitações e serviços do dia atual, em tempo real. Informações sobre demandas de serviços de arborização, drenagem, iluminação, limpeza, pavimentação, tapa buracos, entre outros serviços de reparação da cidade, solicitados pelos cidadãos via atendimento online ou pelo telefone 156 (links na última seção).

Os dados do *dataset* são atualizados periodicamente, foram contemplados dados até dia 11 de novembro de 2018. Foram consideradas informações a partir de 2012 quando o conteúdo começaram a ser coletado.

O *dataset* possui os seguintes atributos: código do grupo de serviço, descrição do grupo de serviço, código do serviço, descrição do serviço, descrição da situação atual, endereço (logradouro, número, bairro e RPA) e datas (data de cadastro e data da última movimentação). Além dos dados fornecidos no arquivo disponibilizado, o conjunto foi enriquecido com informações locais (latitude, longitude, áreas urbanas), temporais (períodos sazonais, eventos relevantes) e sociais (dados socioeconômicos por bairro do Recife).

O projeto foi realizado com auxílio de ferramentas como sistema de versão, jupyter notebook, python 3 e outras tecnologia. Bibliotecas de python como Request, APIs de localização, Pandas, BeautifulSoup, Holoviews, Matplotlib, Scipy, Scikit-Learn, Seaborn, Numpy e outras foram amplamente utilizadas durante o projeto. O projeto foi pensado e estruturado segundo o padrão Cookiecutter [1]. A ferramenta de visualização Tableau também foi utilizada para plotagens e *dashboard* de apresentação.

### 3. Bases de dados e coletas

A coleta de dados se deu inicialmente pelo site de dados abertos da prefeitura do Recife [2], as informações já foram apresentadas em formato CSV, estruturados por ano, com seus respectivos dicionários de descrição.

O dados socioeconômicos foram obtidos através de *crawler* sobre o site da prefeitura do Recife sobre os perfis dos diferentes bairros da cidade [3]. O site apresenta grande irregularidade de obtenção de informação, sendo por partes complexa a retirada de dados, contudo a mesma se mostrou possível. O *crawler* foi necessário devido a não disponibilidade de informações no site de origem dos dados, no caso, o CENSO do IBGE.

Os dados de localização foram obtidos através do *plug-in* para a plataforma de tabelas da Google chamada Geocode Cells [4]. Para retirada de dados, basta selecionar a célula com o endereço e fazer a requisição, o resultado é a latitude e longitude na célula vizinha ao endereço, o dado por si só já estava estruturado em planilha. Outras APIs como geopy e google maps foram utilizadas mas não davam liberdade para a obtenção de dados em larga escala. Cerca de 60.000 endereços foram crawlados repetitivo ao ano de 2018.

Sobre as informações temporais, foram obtidas através de dois mecanismos, API da plataforma calendário [5]. Os dados foram obtidos em formato JSON, sendo facilmente integrável. Sobre os dados de monitoramento pluviométrico, os mesmo foram obtidos no site da APAC [6]. As informações de dados pluviometricos estavam registrados em um planilha HTML facilmente lida pela biblioteca Pandas.

#### 4. Pré-processamento dos dados

Todas as bases trabalhadas possuíam um padrão bem regular para a execução do projeto, em sua maioria, o trabalho com relação ao pré-processamento ficou essencialmente sobre a tipagem dos dados, isso é, discriminar informações categóricas, numéricas, datas e afins. Outra modificação ficou a cargo da padronização de nomes utilizados diferentemente ao longo do ano em que as informações eram coletadas, como o nome dos grupos de serviço que passavam a ter nome diferente ou apresentavam uma representação com caracteres especiais que precisava ser substituída. Quanto aos dados pluviométricos, um importante pré-processamento que foi realizado, é a modificação dos grupos de coleta em cada endereço para a média de todos os grupos em um dia, isso porque seria muito complexo lidar com o dataset pluviométrico a nível de onde a coleta foi realizada e não sobre um valor médio diário de nível pluviométrico. Os dados processados foram salvos em formato .h5 (hdf) de forma a preservar a forma após o pré-processamento, garantindo a tipagem das colunas.

#### 5. Integração dos dados

Os dados foram integrados à base de dados das solicitações através das principais colunas de indexação de cada *dataset* externo, as principais colunas utilizadas são:

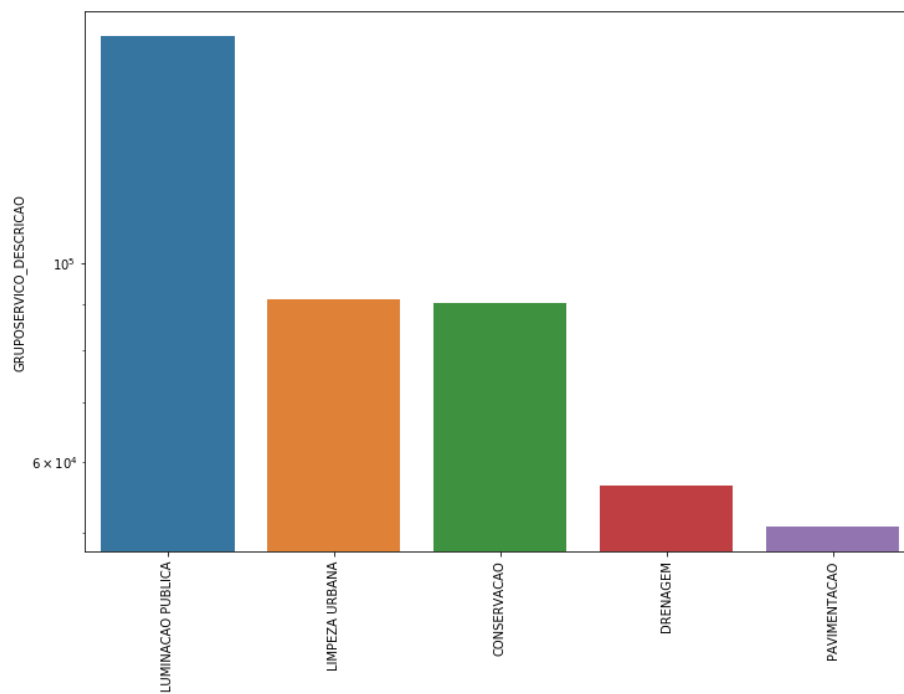
- dataset socioeconômico: merge pelo bairro
- dataset de localização: merge pelo bairro, logradouro e número
- dataset de datas e pluviométrico: merge pela data de demanda.

## 6. Avaliação de hipóteses

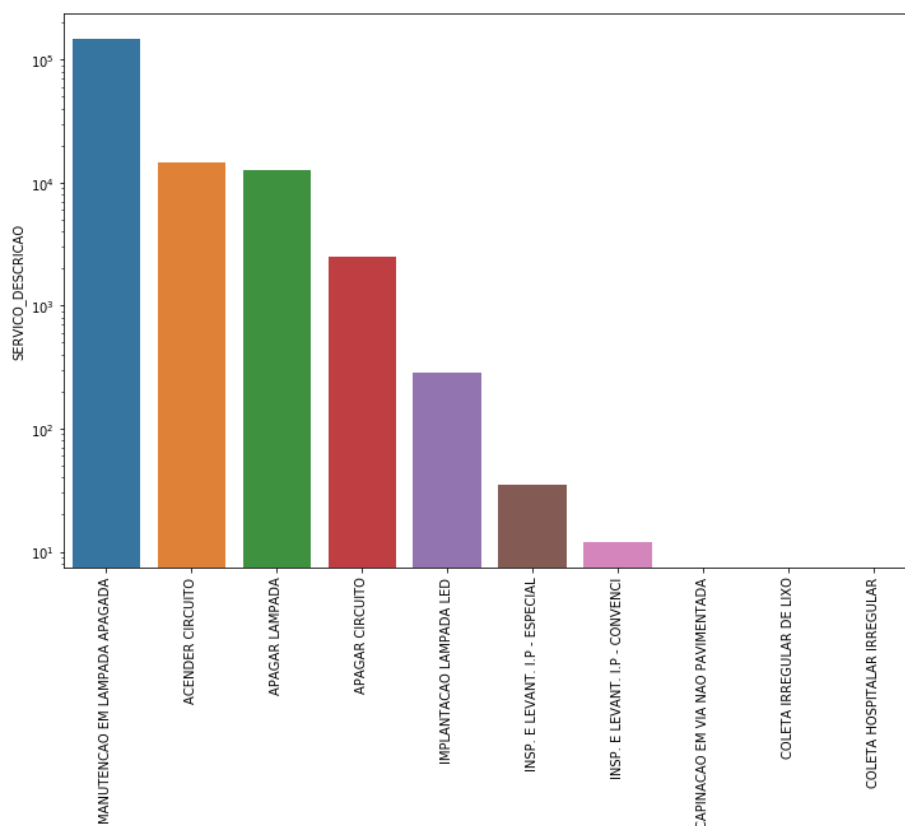
Sobre a análise foram formulados seis hipóteses iniciais a fim de compreender os dados que estavam sendo estudados, de forma a ter um caminho estruturado de exploração, com relação a proposta original existiu a modificação das hipóteses assim que eram retirados dos dados informações relevantes que indicavam um melhor caminho para exploração ou a demonstração de uma análise mais relevante.

### 6.1. *Hipótese 1: Análise das solicitações mais frequentes*

Para responder essa questão foram utilizados uma serie de procediemntos de contagem a fim de evidênciar serviços que eram utilizados com maior frequência da base de dados, tanto quando ao macrogrupo de serviços fornecidos, quanto ao serviço fornecido especificamente.



No gráfico acima tem-se a visualização dos top 5 grupos de serviços mais frequentes no *dataset*, observe que o número de requisições relacionada a iluminação são bem maiores com relação aos grupos de limpeza urbana e de conservação. A escala logaritmica evidência esse grau de maior frequência.

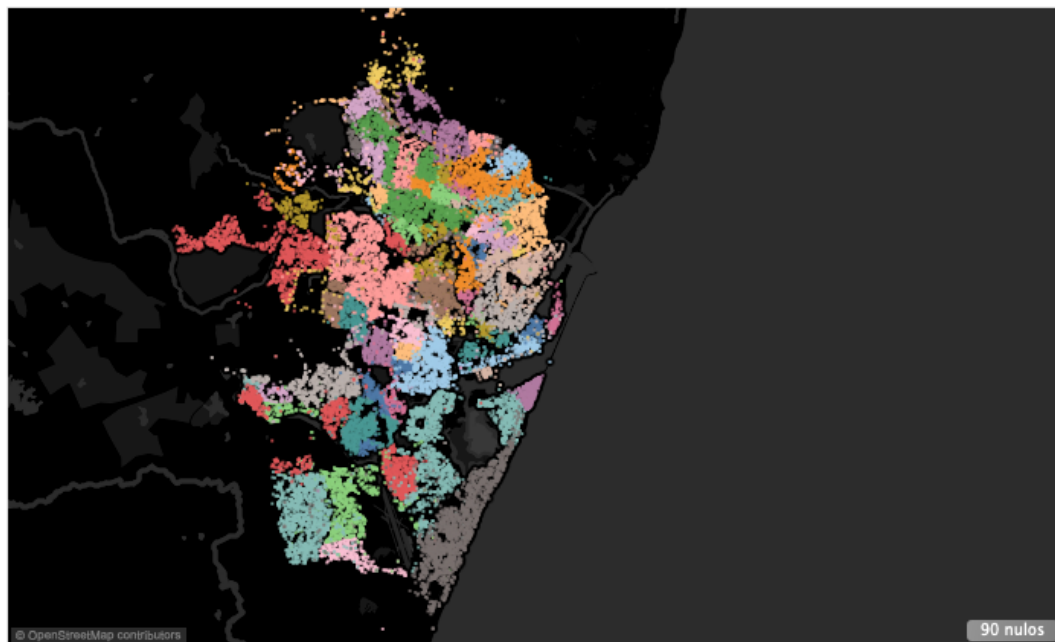


Se analisar mais a fundo os serviços sobre iluminação pública, pode-se observar que o principal serviço entre eles é o serviço de manutenção de lâmpadas apagadas, seguida dos pedidos para religarem a rede elétrica e apagar determinada lâmpada de uma região. Esses dados podem ser vistos no gráfico acima.

Os dados também foram estudados de forma a analisar o serviço mais frequente por bairro, no qual ficou claro que iluminação pública esta presente em quase todos eles como o mais frequente. Caso o serviço de iluminação não seja o mais pedido, o de limpeza pública ganha na grande maioria dos casos.

### *6.2. Hipótese 2: Análise das solicitações sobre disposição gráfica*

A fim de compreender como é o comportamento da disposição gráfica dos dados de solicitação na cidade do Recife, foram obtidos dados sobre latitude e longitude gráfica de forma a evidenciar essa característica, foram obtidos 60.000 dados de localização respectiva a base de 2018, não foram obtidos de outros anos devido o elevado numero de requisições necessárias para a obtenção.



Através da plataforma tableau foi possível visualização a disposição gráfica de pontos no espaço, no caso da visualização acima, temos a disposição de pontos de solicitação na cidade com a segmentação por bairros observe que os pontos parecem apresentar comportamento quase uniforme quanto ao número de serviços em toda a cidade.



Observando especificamente uma região, percebe-se que também podemos observar que a distribuição de tipos de serviço realizados na cidade também parece ter uma determinada uniformidade, além disso, uma característica importante é que a maioria dos dados apresentam uma rasterização sobre ruas, devido ao caráter do endereço, sendo um comportamento já esperado.

### 6.3. Hipótese 3: Análise das solicitações sobre tempo de conclusão

## 7. Referências

- [1] <https://drivendata.github.io/cookiecutter-data-science/>



- [2] <http://dados.recife.pe.gov.br/dataset/central-de-atendimento-de-servicos-da-emlurb-156>
- [3] <http://www2.recife.pe.gov.br/servico/perfil-dos-bairros>
- [4] <https://chrome.google.com/webstore/detail/geocode-cells/pkocmaboheckp-kcbnnlghnfcjjikmfc>
- [5] [http://www.calendario.com.br/api\\_feriados\\_municipais\\_estaduais\\_nacionais.php](http://www.calendario.com.br/api_feriados_municipais_estaduais_nacionais.php)
- [6] <http://www.apac.pe.gov.br/meteorologia/monitoramento-pluvio.php>