

Universidade Federal do Pampa - UNIPAMPA
Engenharia de Software

Iderli Pereira de Souza Filho
Luiz Guilherme Fritsch

RELATÓRIO DE INTRODUÇÃO A ANÁLISE DE DADOS:

Alegrete
2018

Iderli Pereira de Souza Filho
Luiz Guilherme Fritsch

RELATÓRIO DE INTRODUÇÃO A ANÁLISE DE DADOS

Relatório final do primeiro
trabalho de Introdução a análise de dados
utilizando datasets abertos na web para
analisar os dados

Alegrete
2018

Resumo

Nos dias atuais cada vez mais é dada importância para que os dados a respeito de prefeituras, educação, política, sejam públicos. Porém, apenas disponibilizar não é o bastante, o maior ganho em se ter esses dados é as informações que podem ser retiradas com o estudo deles. Com isso, o objetivo deste trabalho é fazer uma análise de uma base de dados pública, para isso foi escolhida a base de dados de imóveis do Patrimônio público da prefeitura municipal de Alegrete.

Palavras-chave: Análise de Dados. Prefeitura Municipal, Patrimônio público, Alegrete.

LISTA DE ILUSTRAÇÕES

Figura 1 -Dataset sem limpeza.....	8
Figura 2- Mudança nas colunas do dataset.....	8
Figura 3 - Gráfico 01- Imóveis por departamento (Rural, Urbano).....	11
Figura 4 - Gráfico 02 - Imóveis por localização.....	12
Figura 5 - Gráfico 03 - Imóveis adquiridos por mandato de prefeitos.....	12

LISTA DE ABREVIATURAS

ES	Engenharia de Software
CSV	Formato de arquivo tabular (excel)

SUMÁRIO

1. Apresentação geral do dataset.....	6
2. Desenvolvimento.....	7
2.1 Preparação dos dados.....	7
2.2 Elaboração e discussão dos resultado das questões.....	7
3. Considerações finais.....	8
4. Referências.....	9

Apresentação geral do dataset

A prefeitura Municipal de Alegrete atendendo a Lei Federal N° 12.527/11 e o Decreto Municipal N° 222/14 disponibilizou dados públicos referentes a seus patrimônios, projetos, contratos, etc. Um dos dados disponibilizados pela prefeitura é o de Imóveis do Patrimônio público.

O dataset de imóveis é dividido em localização do imóvel, descrição, data de aquisição, departamento e observações. O primeiro trabalho que nos deparamos ao verificar o dataset foi com a necessidade realizar uma limpeza nos dados, após esta tarefa, foi feito um brainstorm para decidir possíveis questões a serem verificadas neste dataset.

Um cuidado que tivemos foi o de separar questões relevantes e não só perguntas sem um propósito, para isso fizemos questionamentos sobre onde esses resultados poderiam ser aplicado e quem pode estar interessado nestas questões, com isso determinamos 10 questões para esta análise de dados.

Desenvolvimento - Preparação dos dados

Para a preparação dos dados na limpeza, fizemos uso da linguagem python e de suas ferramentas.

Primeiramente avaliamos o csv que nos foi passado, tinha algumas inconsistências que nos impedia de simplesmente criar um dataframe com todos os dados, inconsistências como a coluna de localização, que não estava definida como as demais, seus dados estavam em destaque como se fosse uma localização contendo vários imóveis, como pode ser observado na figura 1.

ÁREA DO ANTIGO AEROPORTO			
TERRENO MATRÍCULA 24575	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Uma faixa de terreno, situac
TERRENO MATRÍCULA 25468	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25469	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25470	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25471	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25472	25/01/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25473	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25474	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Um terreno, situado, zona le
TERRENO MATRÍCULA 25476	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Uma faixa de terreno, situac
TERRENO MATRÍCULA 25477	5/1/2007	CADASTRO DE IMÓVEIS URBANOS-PATRIMÔNIO	Uma faixa de terreno, situac
10 imóveis(s) na divisão ÁREA DO ANTIGO AEROPORTO			

Figura 1. Dataset sem limpeza

Para resolvermos este problema, criamos uma função que torna os dados conforme podem ser observados na figura 2.

```
Bairro
Descrição; DataDeAquisição; Departamento; Observações
Descrição; DataDeAquisição; Departamento; Observações
Descrição; DataDeAquisição; Departamento; Observações

Assim:
Bairro; Descrição; DataDeAquisição; Departamento; Observações
Bairro; Descrição; DataDeAquisição; Departamento; Observações
Bairro; Descrição; DataDeAquisição; Departamento; Observações
```

Figura 2. Mudança nas colunas do dataset

Primeiramente lemos linha por linha do csv, cada linha vem no formato de uma lista, mas como a coluna '*observação*' é muito grande e contém diversos caracteres, ela não era um só elemento na lista então ocupava várias posições logo, transformamos esta lista toda em uma única String e criamos de novo uma lista, mas desta vez usando como critério para o separador o “;”.

A partir de agora temos uma lista correta com os dados, falta só resolver o problema da coluna '*localizacao*'. Para resolver este problema, quando ele lia uma lista que tinha tamanho 1 e que não continha a string “*na divis*” nós armazenamos esta string, caso contrário, caso fosse uma lista de tamanho 1 e tinha aquela string dentro, nós sabíamos que aquela string era a última linha, a linha que continha quantas vezes o bairro aparecia ali. No final da iteração, simplesmente criamos um vetor '*localizacao*' e iteramos a quantidade de vez que o bairro aparecia, adicionando o nome do bairro. Pronto! Agora temos como criar nosso dataframe para trabalharmos.

As bibliotecas utilizadas nesta tarefa foram: pandas, numpy, io, csv e datetime.

Desenvolvimento - Elaboração e discussão dos resultados

Uma questão que foi observada e caracterizada como importante foi a de imóveis por departamento, nela o objetivo é verificar quantos imóveis estão cadastrados em localidade rural e quantos estão em localidade Urbana.

Com esta questão podemos observar qual é o foco de localidade em patrimonio do municipio de Alegrete. Os dados nos mostraram que 183 imóveis estão localizados em área urbana e 64 em área rural. Fazendo uma análise dos dados gerados podemos observar que mesmo o Alegrete tendo uma alta extensão em perímetro rural, os imóveis em localidade urbana superam em muitos os rurais, isso pode nos mostrar que o foco de imóveis do município é em extensão urbana.

```
CADASTRO DE IMVEIS URBANOS-PATRIMNIO    183  
CADASTRO IMVEIS RURAIS                  64  
Name: departamento, dtype: int64
```

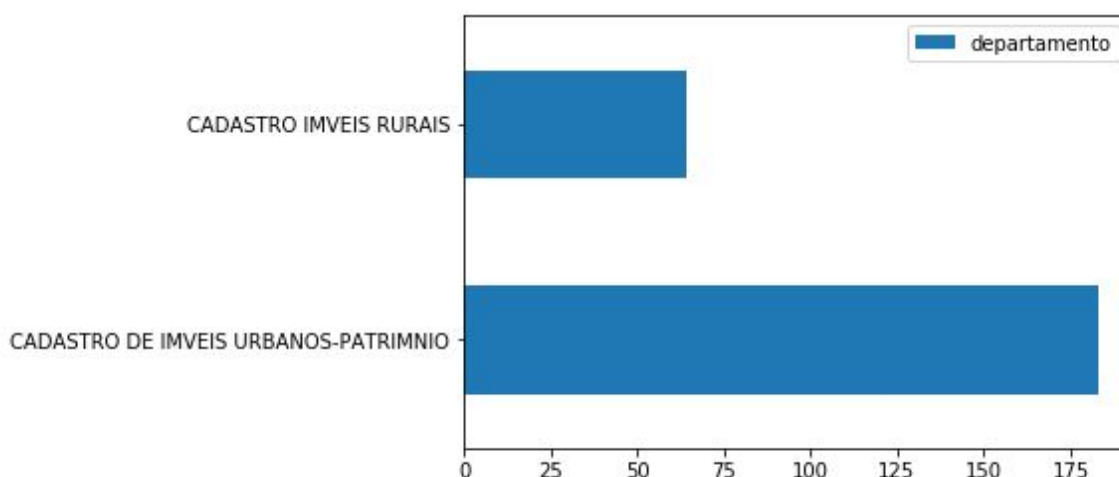


Figura 3. Imóveis por departamento (Rural, Urbano)

Outra questão abordada é a de quantos imóveis existem por localidade (bairro, vila, perímetro, setor), com esses resultados podemos estimar em quais regiões da cidade a prefeitura municipal tem mais interesse. Por exemplo, o setor

N15 (nomenclatura dada pelo dataset) é o que tem mais imóveis do patrimônio municipal, com isso é possível demonstrar um interesse da prefeitura por esta região, em contrapartida as localidades, bairro pedreiras, setor N°18, setor N°19. bairro josé de abreu, Capão do Angico e setor N°22, só contém um imóvel da prefeitura municipal.

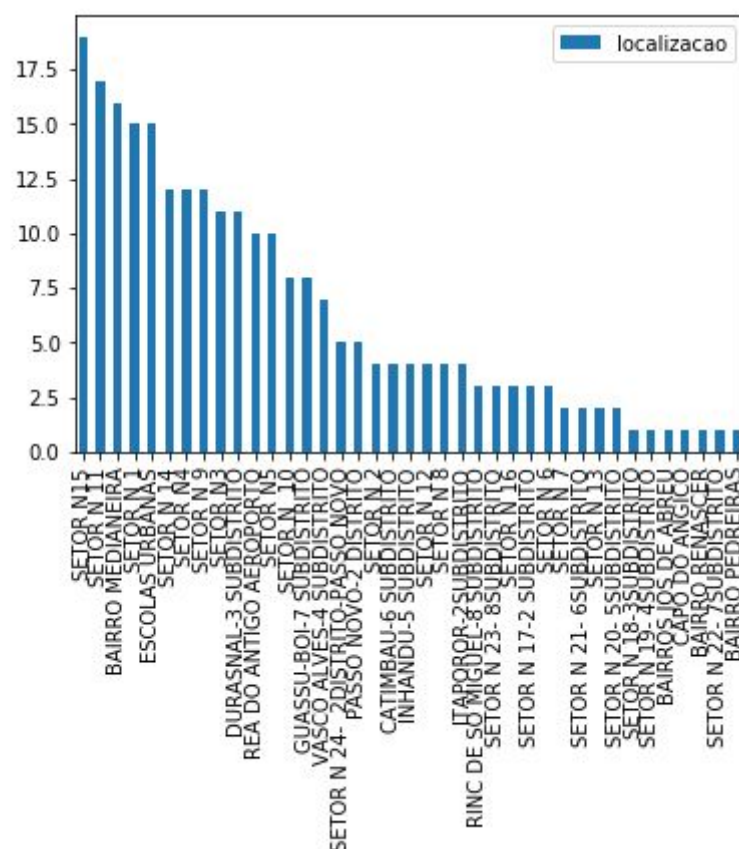


Figura 4. Imóveis por localização

Mais uma questão vista foi a de imóveis adquiridos por data, com esses resultados podemos observar a quantidade de imóveis adquiridos por mandato dos prefeitos de Alegrete, com isso um cidadão pode fazer cobranças ao prefeito ou verificar os investimentos da prefeitura.

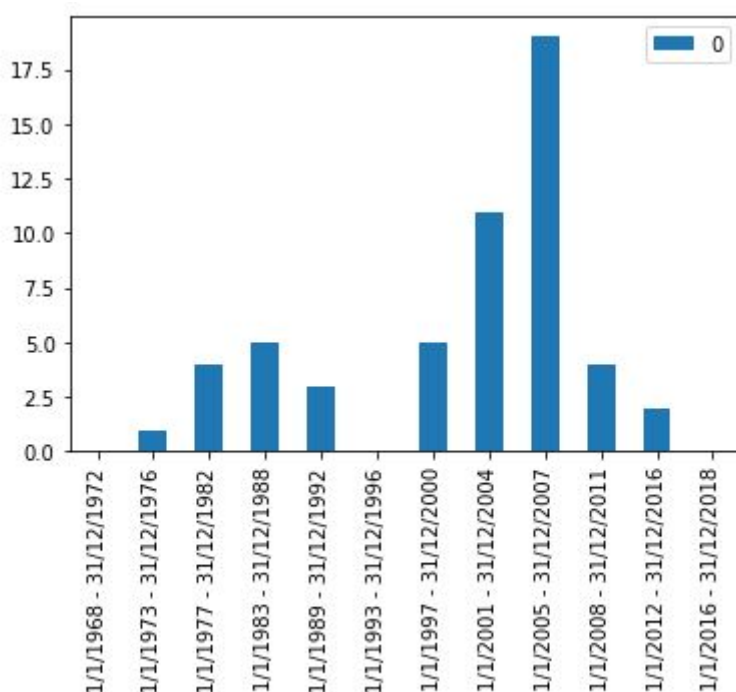


Figura 5 - Gráfico 03 - Imóveis adquiridos por mandato de prefeitos

As demais perguntas se encontram dentro do script ou do IPYthonNotebook.

Considerações finais

Neste trabalho podemos observar que o município de Alegrete possui vários imóveis tanto rurais como urbanos, sendo que a maioria foi adquirido em 2007. Podemos observar também como os datasets governamentais são desorganizados e sujos.

Como dificuldades para realizar o trabalho, podemos citar que o enunciado estava muito amplo, e, tivemos pouco tempo para conversar com a professora, levando a tomar decisões por conta própria, como deixar todas as data de aquisição que estavam erradas (11/11/1111) dentro do dataframe. Outra dificuldade encontrada foi que o nosso dataset estava muito errado e sem descrição de algumas coisas como as localizações, que tinha dado como setor, que não diz nada sobre o que é.

Referências

PREFEITURA MUNICIPAL DE ALEGRETE. Disponível 04/06/2018 - em:
<<http://www.alegrete.rs.gov.br/sic/acessoainformacao/?imoveis>>