

**TECNOLOGIA EM SISTEMAS PARA INTERNET**

**Daniel Evangelista Pereira**

**Ribson Coelho Cardoch Valdés**

**Douglas Seidi Shibata**

**RELATÓRIO DE PRÁTICA INTEGRADA**

**DE**

**CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL**

03/10/2020

**Brasília - DF**

**2020**

**Sumário**

[**1. Objetivos**](#_r578hjw9oxco) **3**

[**2. Descrição do problema**](#_jsmvxwf8smme) **4**

[**3. Desenvolvimento**](#_ybi6lms96bcw) **5**

[3.1 Código implementado](#_tcrscypicfq5) 5

[**4. Considerações Finais**](#_sp9srgapjf6s) **15**

[**Referências**](#_5hyd2pq8xv2g) **16**

# 1. Objetivos

Está etapa do projeto tem como finalidade a exploração dos dados utilizando a linguagem SQL no script do notebook em python, que foram coletados na primeira fase. E a partir desta exploração será gerados mapas e gráficos para a melhor visualização dos dados coletados.

# 2. Descrição do problema

Nesta Etapa foi feita a exploração dos dados e gerados gráficos e mapas para a melhor visualização.

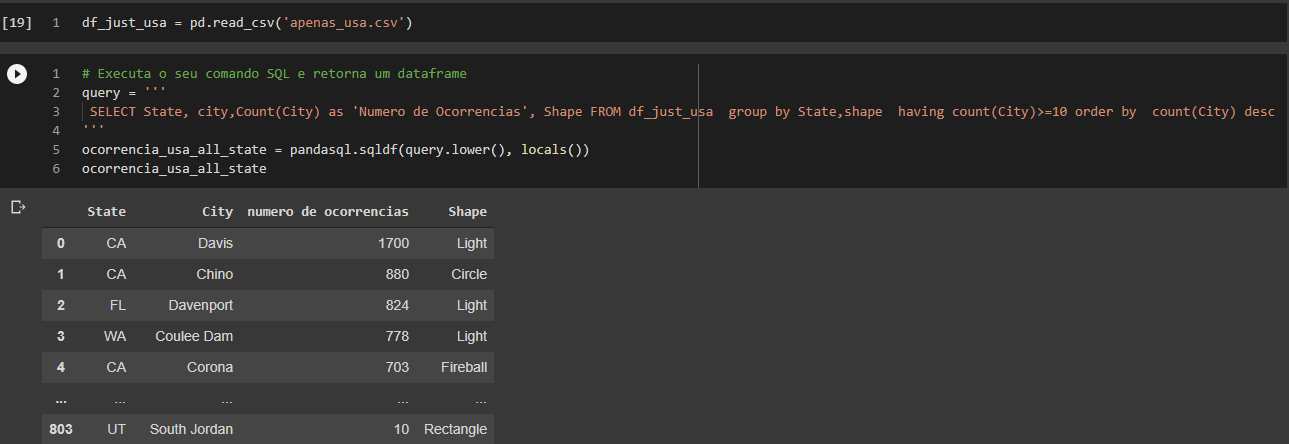
# 3. Desenvolvimento

As tecnologias utilizadas para a elaboração desta segunda fase do projeto, será a linguagem python e algumas bibliotecas para ajudar no desenvolvimento, com a biblioteca matplotlib e seaborn para a geração de gráficos, por meio do notebook e o ambiente de desenvolvimento Google Colab.

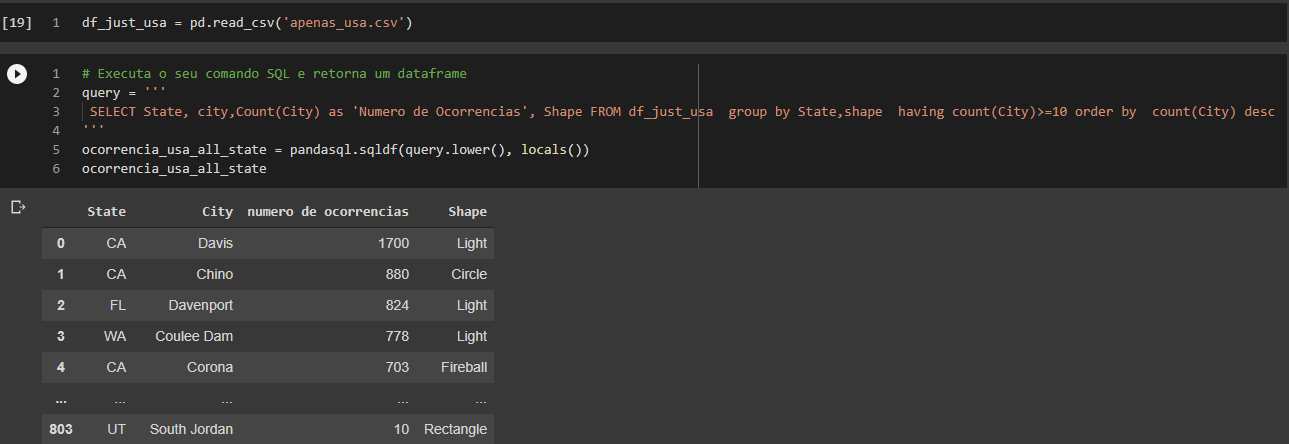
## 3.1 Código implementado

5.5 - Exploração com gráficos e mapas

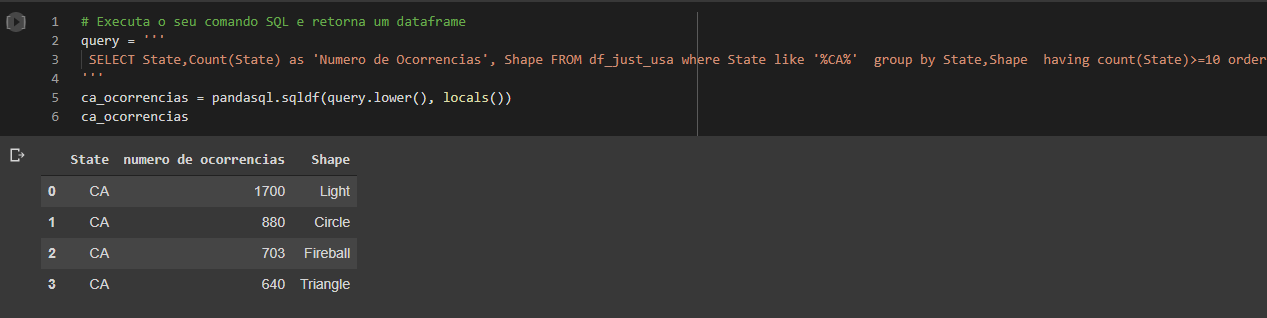
Realizando a leitura do arquivo com apenas dados dos estados unidos, e atribuindo a variável df\_just\_usa que será o dataframe para realizar a análise.



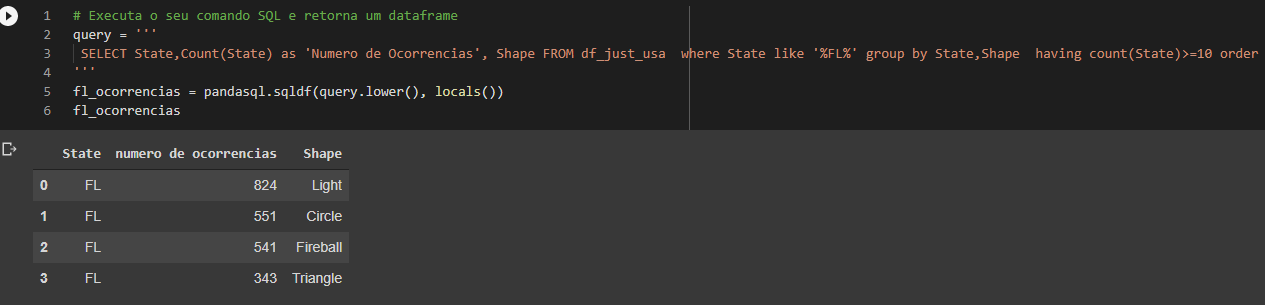
Realizando consulta utlizando a query do SQL para ter a informação do número de ocorrências, agrupando por estado e formato para saber a quantidade de casos por estados e quais formas foram mais recorrentes naquele estado.



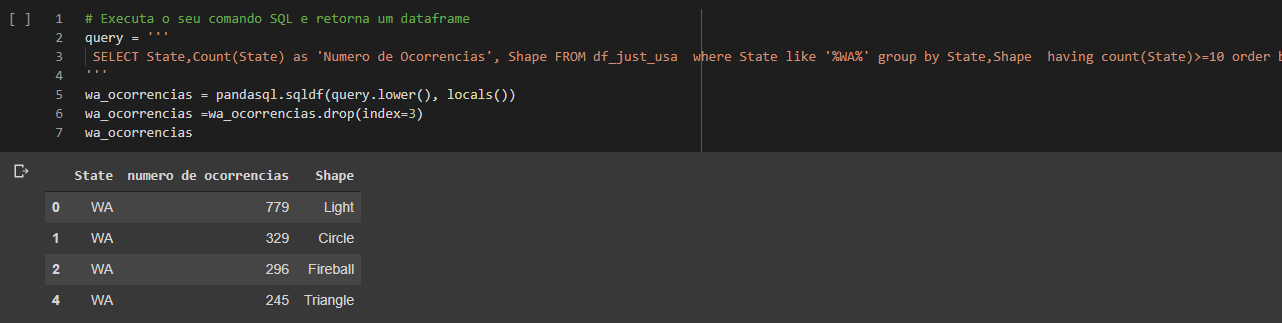
Dividindo pelo estado com o maior número de ocorrência, que neste caso é o estado da Califórnia e mostrando os formatos mais recorrentes. E atribuindo a um novo dataframe chamado ca\_ocorrencias



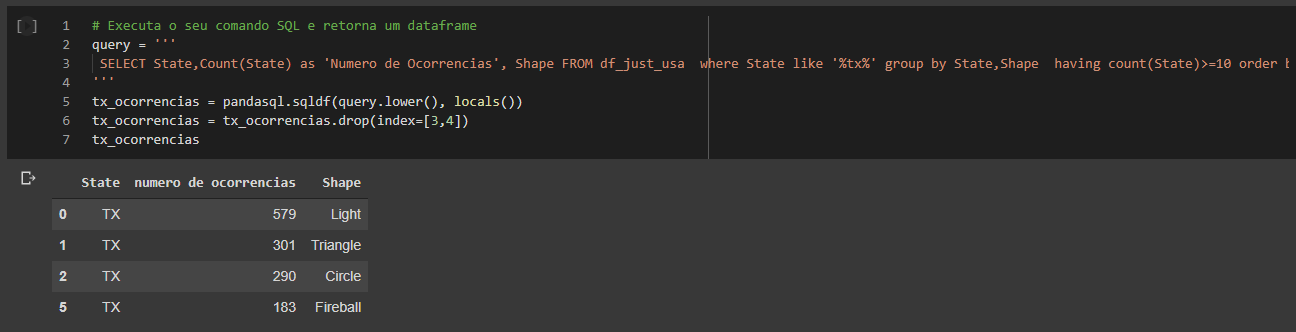
Dividindo pelo estado com o maior número de ocorrência, que neste caso é o estado da Flórida e mostrando os formatos mais recorrentes. E atribuindo a um novo dataframe chamado fl\_ocorrencias



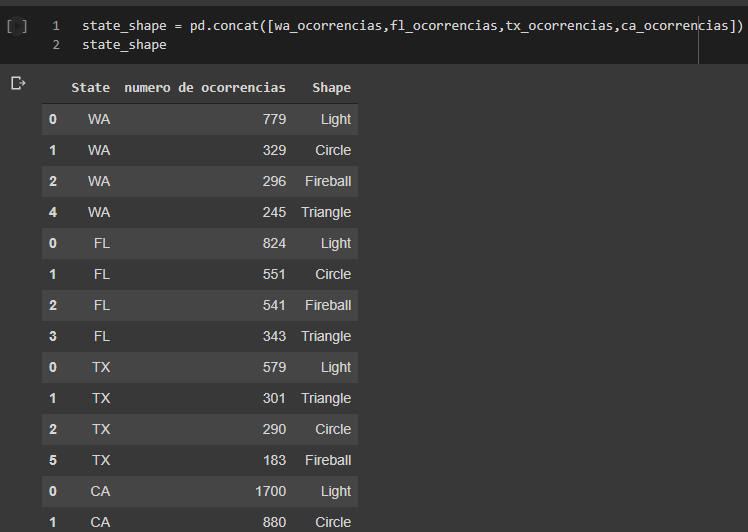
Dividindo pelo estado com o maior número de ocorrência, que neste caso é o estado de Washington e mostrando os formatos mais recorrentes. E atribuindo a um novo dataframe chamado wa\_ocorrencias



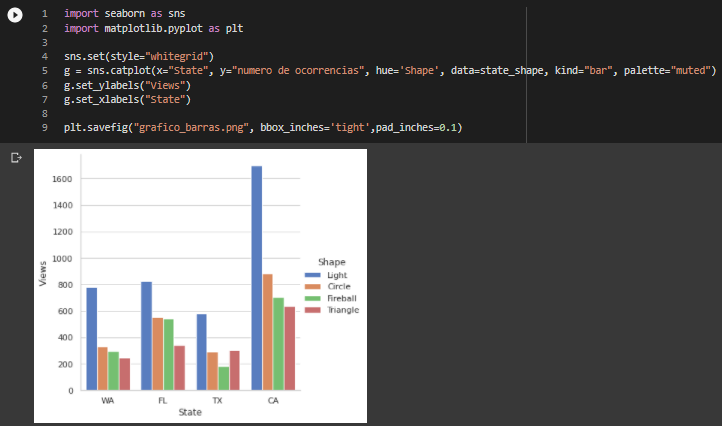
Dividindo pelo estado com o maior número de ocorrência, que neste caso é o estado do Texas e mostrando os formatos mais recorrentes. E atribuindo a um novo dataframe chamado tx\_ocorrencias



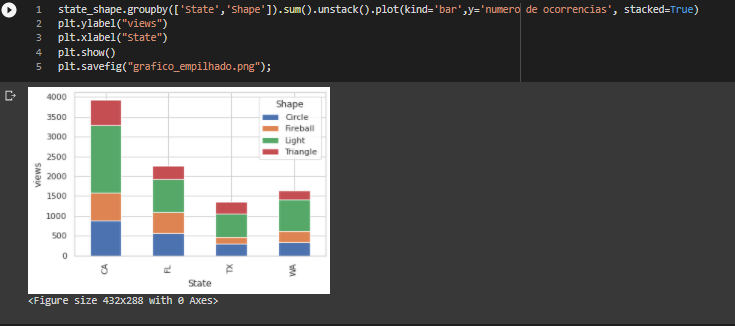
Juntando os Dataframes e atribuindo a um novo dataframe, com os 4 estados mais recorrentes e suas formas.



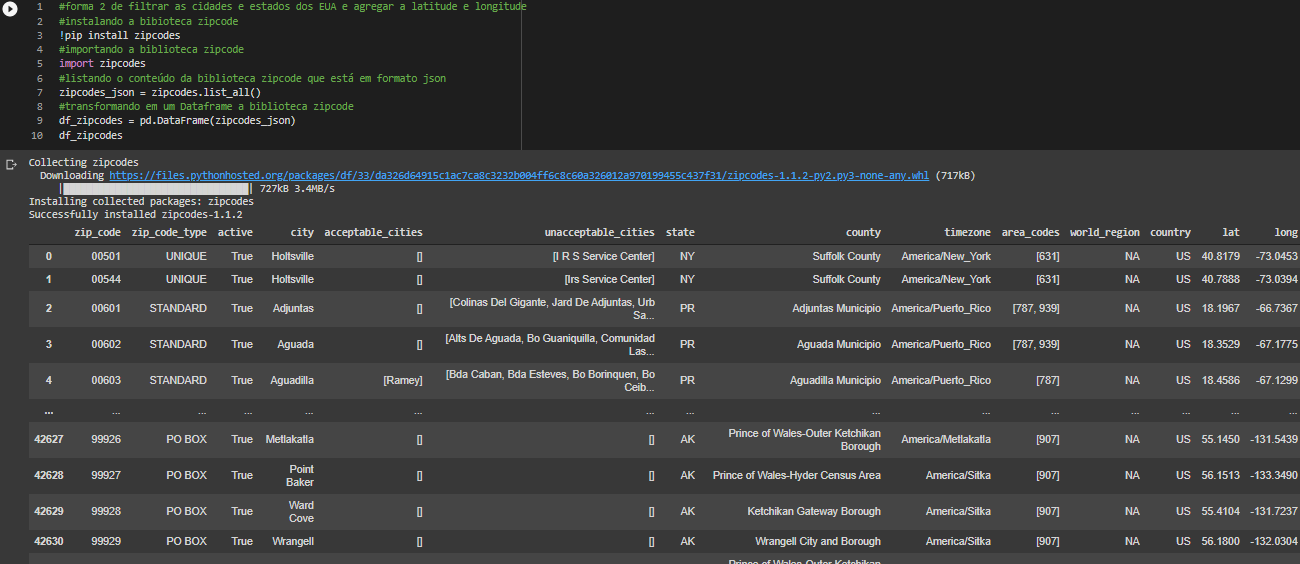
Gerando o gráfico de barras mostrando os estados da califórnia, washington, Flórida e Texas, com as visualização no eixo y e os estados no eixo x. E depois exportando o gráfico.



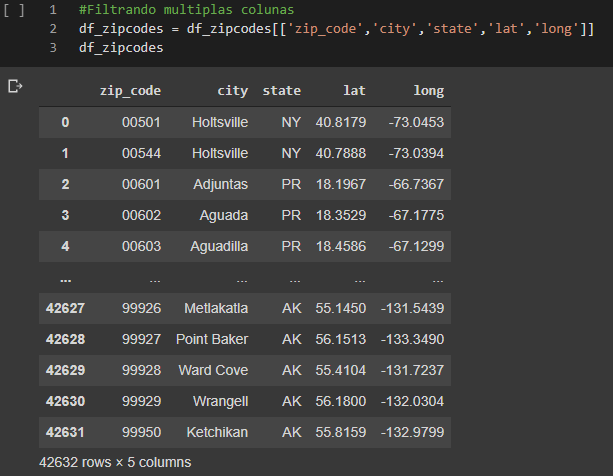
Agrupando o dataframe por state e shape e somando para poder empilhar os dados e mostrar no gráfico.



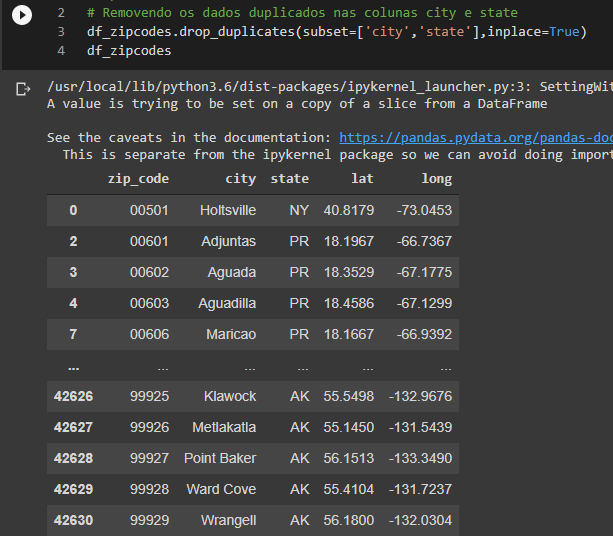
Utilizando a biblioteca zipcode. E coletando os dados que a biblioteca possui e atribuindo este valor para a variável zipcode\_json e depois transformando a resposta em json para um dataframe para facilitar a análise de dados.



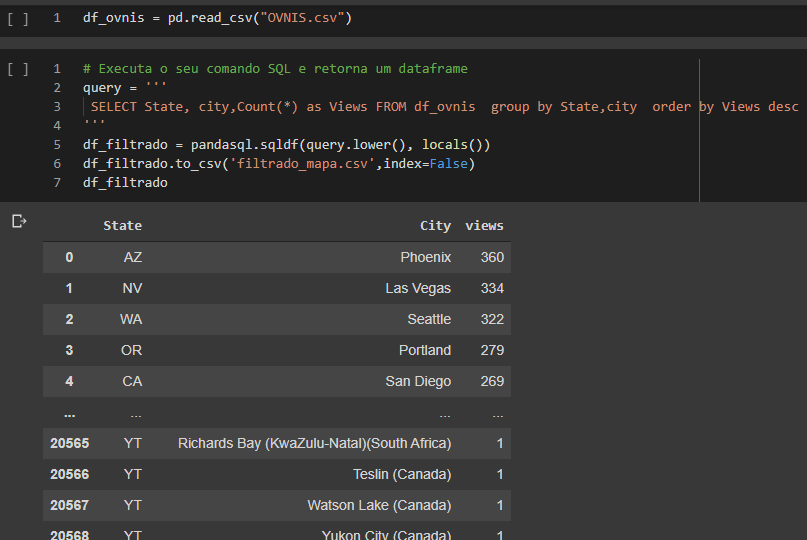
Filtrando as colunas do dataframe zipcode que foi criado para apenas as colunas que iremos utilizar para a análise



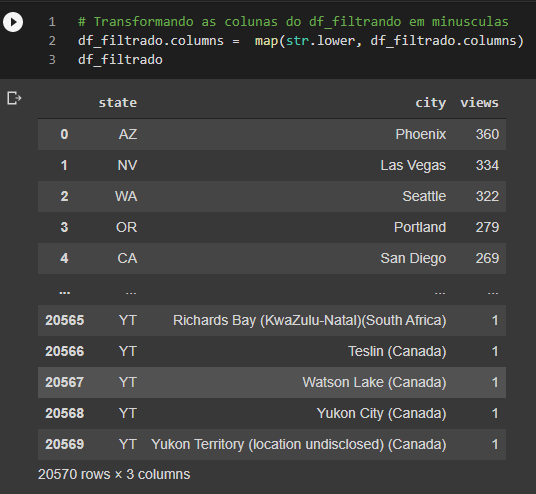
Removendo valores duplicados nas colunas city e state do dataframe df\_zipcode.



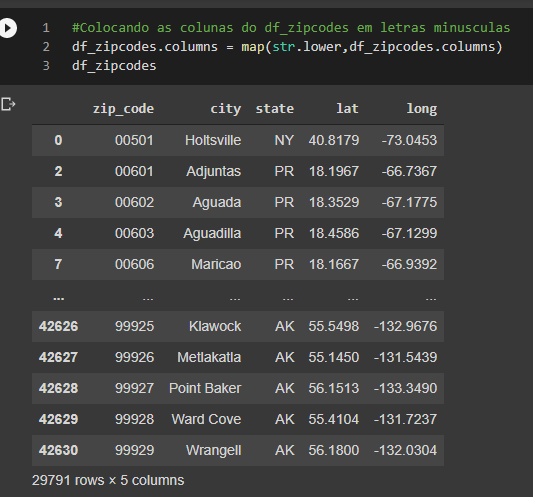
Lendo o arquivo ovnis.csv e atribuindo a variável df\_ovnis. E no bloco seguinte realizando a consulta da query, que irá filtrar os dados agrupados por state e city.

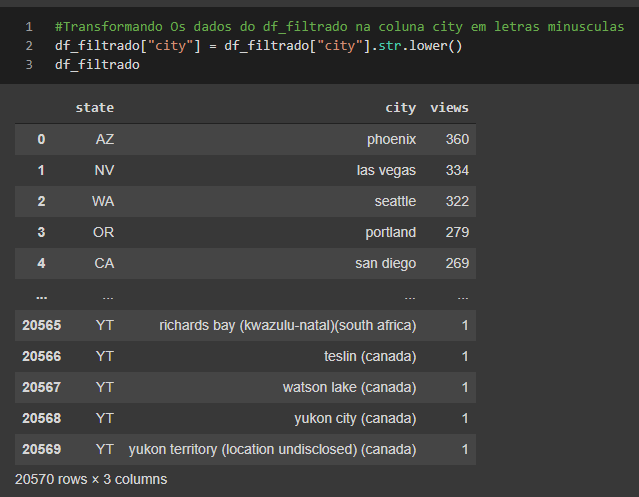


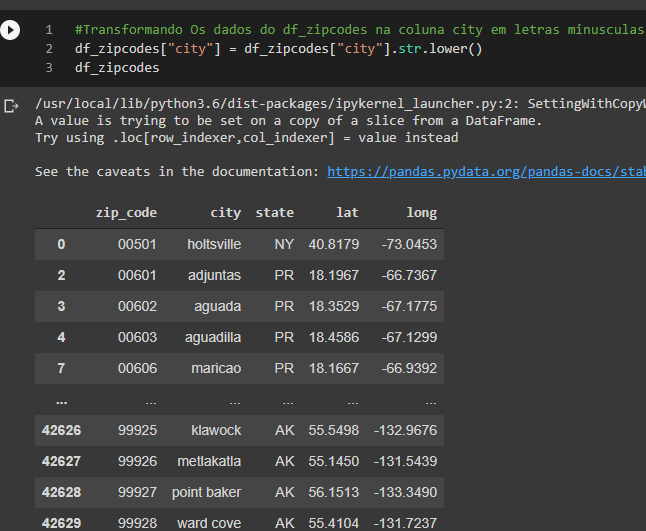
Transformando o nome das colunas para letras minusculas para que sejam padronizadas para que depois possa se juntar como dataframe do zipcodes

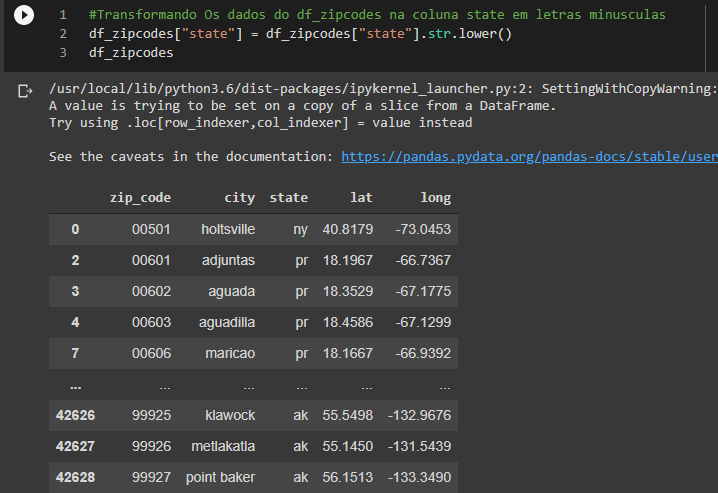


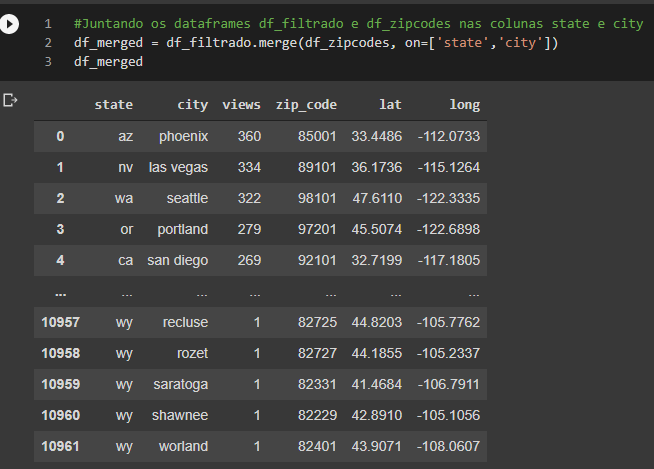
Transformando as colunas para letras minúsculas no dataframse df\_zipcode.









Juntando os dataframes a partir da coluna city e state. Utilizando a função merge do pandas.

# 4. Considerações Finais

Nesta parte de exploração de dados, houve estudos das bibliotecas utilizadas para que se fosse possível gerar os gráficos e o mapa. No processo de desenvolvimento houve alguns empecilhos, como a execução dos gráficos e do mapa. Porém com os estudos e leitura da documentação, foi possível gerá-los

# Referências

Raymond. Pandas DataFrame Plot - Bar Chart. Março de 2020. Kontext. Disponível em <<https://kontext.tech/column/code-snippets/399/pandas-dataframe-plot-bar-chart/>> Acesso em 27 de Setembro de 2020

Albon,Chris. Lower Case Column Names In Pandas Dataframe. Chris Albon. 20 de dezembro de 2017. Disponível em<<https://chrisalbon.com/python/data_wrangling/pandas_lowercase_column_names/> >Acesso em 01 de Outubro de 2020

Kite your programming copilot. How to make a pandas DataFrame string column lowercase in Python. Disponível em <<https://www.kite.com/python/answers/how-to-make-a-pandas-dataframe-string-column-lowercase-in-python#:~:text=Use%20str.,a%20DataFrame%20string%20column%20lowercase&text=lower()%20to%20make%20all,%5B%22first_column%22%5D%20lowercase>/> Acesso em 01 de Outubro de 2020.