**Atividade de Ciências de Dados**

**Aluno:** Heberth Vinícius Amarante Severo

**Descrição do problema:**

Essa é a terceira parte da nossa atividade de Web Scraping, nela vamos explorar mais um pouco os dados sobre os Estados Unidos.

Em cada etapa será explicado o que foi feito, com todas as imagens necessárias para o entendimento da atividade.

**1ª Etapa**

Para o início do arquivo no Colab é necessário fazer a importação de algumas funções do python, a Figura 1 mostra todas as funções utilizadas para a nossa atividade.

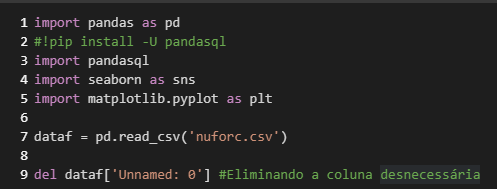


Figura 1 Importação das funções

A primeira etapa é bem simples, nela há apenas as importações que estão presentes até a quinta linha, a leitura do arquivo csv na linha sete, esse arquivo csv foi obtido na primeira parte do projeto. Na linha nove foi necessário deletar uma coluna do nosso csv, pois ela foi gerada automaticamente e não faz parte dos dados coletados.

**2ª Etapa**

Nessa etapa da atividade é necessário selecionar os dados apenas dos Estados Unidos já que queremos utilizar os dados dele, a primeira parte foi criar um data frame apenas com os dados dos estados, cuja a sigla seja dos EUA, como mostra a Figura 2.

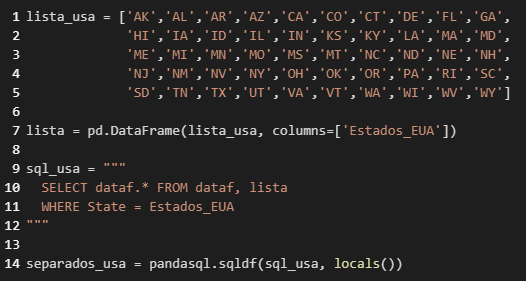


Figura 2 Estados dos Estados Unidos.

Para criar os gráficos que desejamos temos que fazer algumas buscas no nosso dataset para filtrar os dados até chegarmos ao que desejamos.

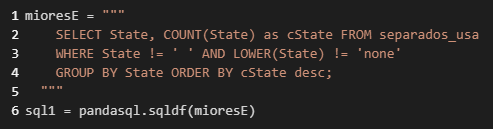


Figura 3 Fazendo a contagem dos relatos de cada estado

A Figura 3 é uma pesquisa no arquivo de dados, que irá retornar a quantidade de relatos que aconteceram em cada estado dos Estados Unidos.

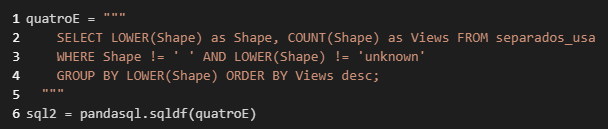


Figura 4 Fazendo a contagem dos formatos que foram relatados

A Figura 4 é uma pesquisa para saber quais os formatos que foram relatados e para saber a quantidade que cada formato apareceu no dataset.

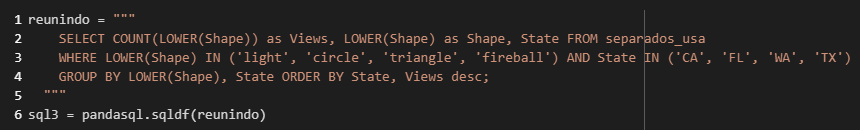


Figura 5 Estados com mais relatos, formato e quantidade de vezes que cada formato apareceu

A Figura 5 foi feita para reunir os dados obtidos nas três figuras anteriores, a Figura 6 mostra a saída dessa query.

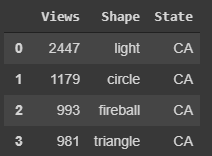


Figura 6 Resultado da query feita na figura 5

**3ª Etapa**

Nessa etapa iremos fazer dois gráficos a partir dos dados obtidos na Figura 6. O primeiro gráfico é bem simples de se compreender, nele há os quatro estados mais citados e os formatos nesses quatros estados. O segundo gráfico é um pouco diferente do primeiro, mesmo ele retornando as mesmas informações, mas ele é diferente pois é um gráfico empilhado.



Figura 7 Código para a criação do primeiro gráfico

A Figura 7 mostra o trecho de código para a criação do primeiro gráfico, a primeira linha é uma paleta de cores para o gráfico que foi feita manualmente, e a segunda linha é a criação do gráfico, podemos ver o gráfico na Figura 8.

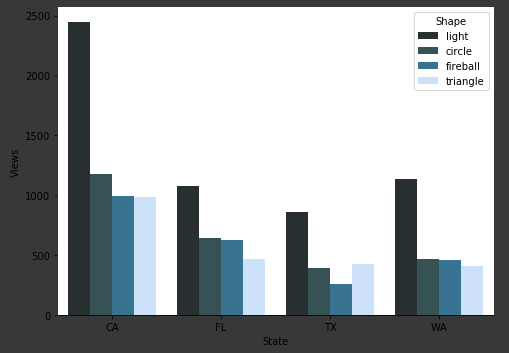


Figura 8 Primeiro gráfico

A Figura 9 mostra o trecho de código para a criação do gráfico empilhado como podemos ver a seguir.



Figura 9 Código para o gráfico empilhado

Para a criação do primeiro gráfico foi usado o seaborn, para a criação do segundo gráfico foi usado o matplotlib, a Figura 10 mostra o gráfico empilhado.

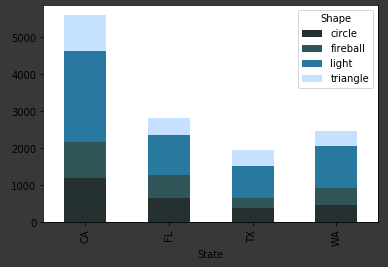


Figura 10 Gráfico empilhado

**4ª Etapa**

Essa etapa é a criação do mapa dos Estados Unidos mostrando os relatos de acordo com cada cidade.

A criação de mapas é um pouco diferente do que vimos até o momento, é necessário fazer alguns imports além do que já fizemos no começo dessa atividade, como mostra a Figura 11.

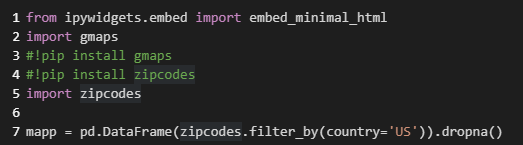


Figura 11 Import para criação do mapa

Além dos imports, na linha sete está sendo feito um data frame usando o zipcodes para obter os dados das cidades para fazer o mapa.

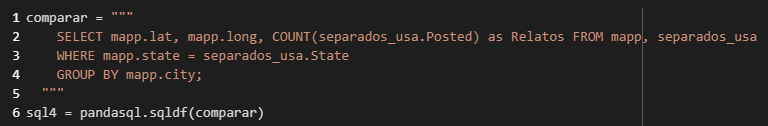


Figura 12 Pesquisa para receber os dados para o gráfico

Com os dados da Figura 12 podemos fazer o gráfico, porém há mais um trecho de código, como mostra a Figura 13.

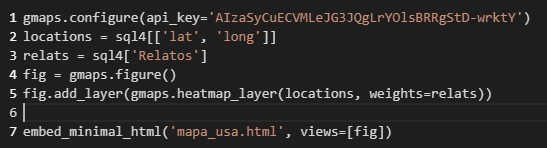


Figura 13 Criando o gráfico

A primeira linha é para colocar a chave gerado pelo Maps para que possamos criar o nosso mapa, a segunda linha passamos as colunas que desejamos, no nosso caso são as colunas de latitude e longitude, a linha sete é a criação do mapa em um arquivo html, já que a plataforma que estamos usando não permite a criação do mapa nela mesma.

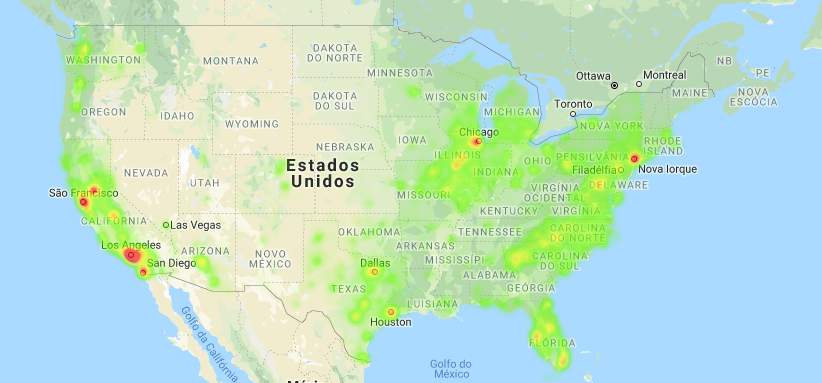


Figura 14 Gráfico dos Estados Unidos

**Observação**

Este relatório estará no GitHub, cujo link está a seguir: <https://github.com/heberth-hvas/icd_5>

**Referências**

**Google Colab:** <https://colab.research.google.com/notebooks/welcome.ipynb>

**Documentação Pandas:** <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.append.html>

**Google Maps:** <https://www.google.com.br/maps/@39.3383087,-97.4941402,4.78z?hl=pt-BR&authuser=0>

**Siglas dos estados dos EUA:** <https://www.geografiaopinativa.com.br/2017/01/siglas-estados-unidos.html>

**Gráfico Empilhado:** <https://living-sun.com/pt/python/726321-plotting-multiple-stacked-bar-graph-given-a-pandas-dataframe-in-python-python-pandas.html>

**Gráficos catplot:** <https://seaborn.pydata.org/generated/seaborn.catplot.html>