**Atividade de Ciências de Dados**

**Aluno:** Heberth Vinícius Amarante Severo

**Descrição do problema:**

Essa é a segunda parte da nossa atividade de Web Scraping, na primeira parte fizemos a coleta bruta dos dados OVNI’s. Nessa parte da atividade foi proposto a exploração dos dados coletados na etapa anterior usando Structured Query Language(SQL).

Essa etapa da atividade foi dividida em várias partes, que estarão a seguir, em cada tópico será explicado a atividade proposta, contendo os códigos feitos no Colab e a explicação deles.

**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.

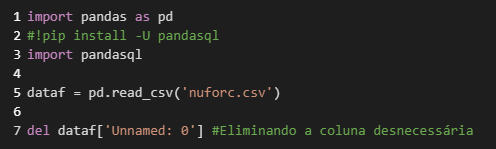


Figura 1 Importação das funções

A primeira etapa é bem simples, nela há apenas as importações que estão presentes até a terceira linha, a leitura do arquivo csv na linha cinco, esse arquivo csv foi obtido na primeira parte do projeto. Na linha sete 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 foi informado a quantidade de linhas que o arquivo csv possui, como mostra a Figura 2.



Figura 2 Quantidade de linhas.

Para saber a quantidade total de linhas foi necessário fazer a contagem através da coluna Date/Time do arquivo, pois não há dados em branco nela.

**3ª Etapa**

Nessa etapa começamos de fato a usar os comandos SQL para mostrar informações especificas, como mostra a Figura 3.

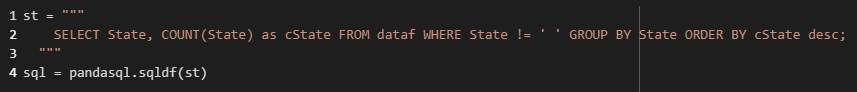


Figura 3 Fazendo a contagem de relatos por estado e colocando os dados em ordem decrescente

Esse trecho de código está fazendo a contagem dos estados, ele ignora os dados cujo não foi informado o estado durante a postagem do relato, a saída mostra a sigla dos estados e a quantidade de relatos que foram feitos em cada estado, como mostra a Figura 4.

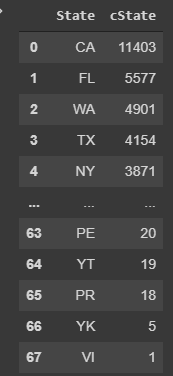


Figura Estados e quantidade de relatos

**4ª Etapa**

O arquivo csv mostra os dados de vários países ao redor do mundo, alguns deles são o Canadá e França, nessa etapa foi feita a restrição dos dados para que fosse mostrado apenas os dados relativos aos EUA, a Figura 5 mostra o trecho de código dessa etapa.

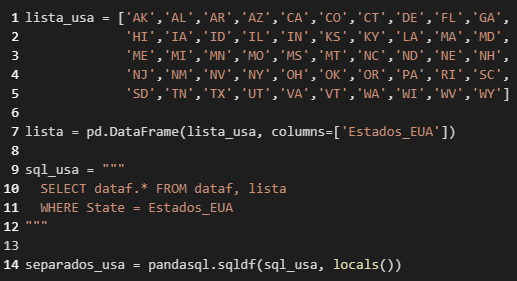


Figura Selecionando apenas dados dos EUA

Foi necessário criar uma lista com as siglas de todos os estados dos Estados Unidos para fazer uma comparação com as siglas do csv e assim retornar apenas os dados desejados, como mostra a figura 6.

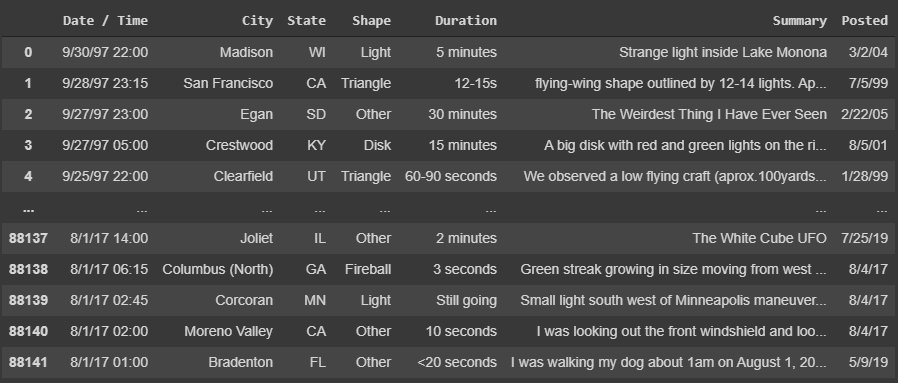


Figura Saída dos dados dos EUA

**5ª Etapa**

Tivemos um grande avanço com os dados do csv, mas podemos especificar um pouco mais os relatos, nessa etapa foi feito a contagem de relatos por cidades de acordo com a Figura 7.

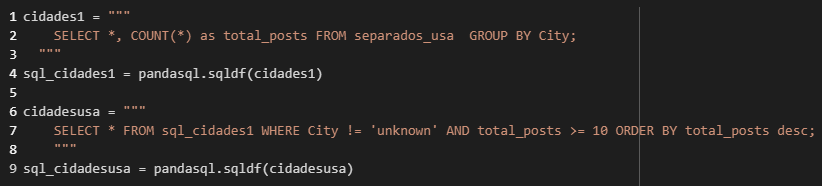


Figura Total de relatos por cidade

O primeiro bloco, que vai até a linha 4, faz a contagem dos relatos de acordo com cada cidade, o segundo bloco está fazendo duas restrições, a primeira é para que não seja mostrado os dados cujo não foi informado a cidade, a segunda restrição é para que seja mostrado apenas as cidades que tiveram no mínimo 10 relatos, com mostra a Figura 8.

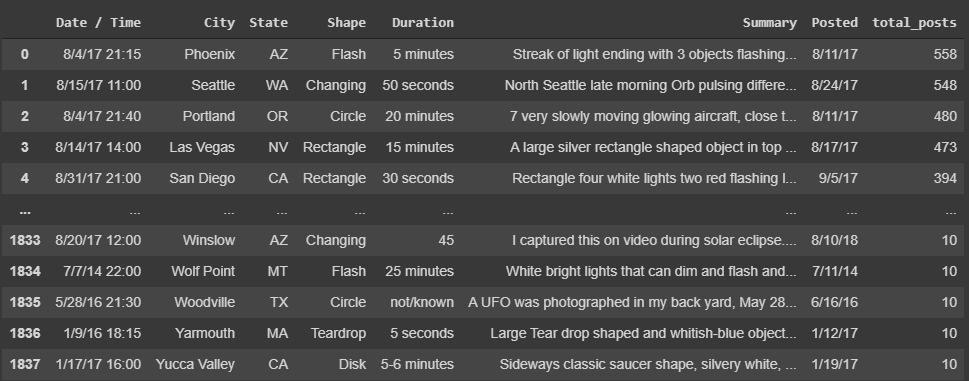


Figura 8 Cidades com no mínimo 10 relatos

**6ª Etapa - Pergunta**

Se observarmos bem a Figura 8, podemos perceber que a cidade de Phoenix é a que possui mais relatos, com o total de 558 relatos, com isso temos que responder a seguinte pergunta: Por que será que essa é a cidade que possui mais relatos?

Resposta: Em 1997 ocorreu um evento chamado Luzes de Phoenix que foram uma série de fenômenos ópticos que ocorreu no estado do Arizona cuja capital é Phoenix, em 2007 também ocorreram esses eventos. Diversas pessoas descreveram esse evento, algumas delas deram detalhes específicos sobre os acontecimentos como diferentes tipos de luzes aparecendo no céu, e esse fato pode ter contribuído com a quantidade de relatos que ocorreram na cidade que é a sexta mais populosa dos Estados Unidos.

**7ª Etapa**

Nessa etapa, é realizado uma query para identificar qual estado possui o maior número de relatos e quais cidades desses estados possuem uma quantidade de relatos maior que 10. Para isso foram realizadas duas consultas de SQL, como mostra a Figura 9.

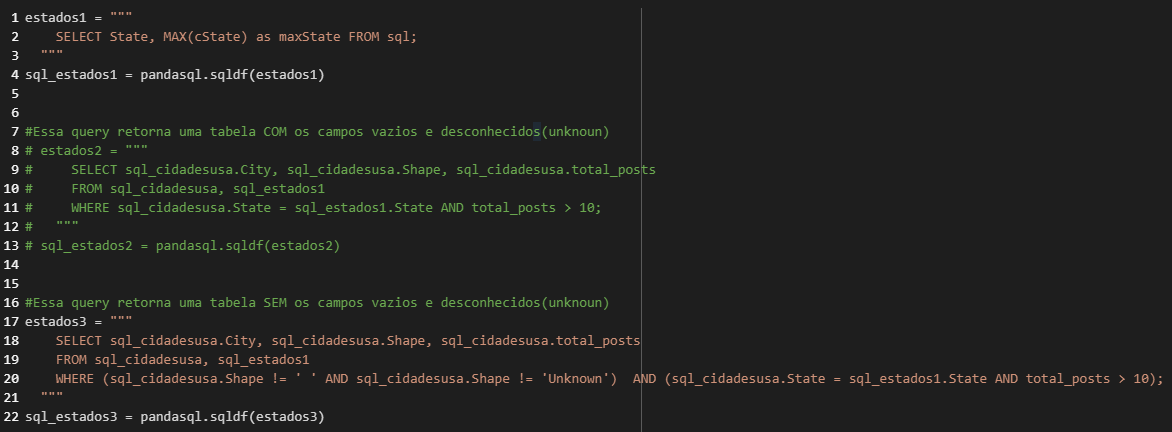


Figura 9 Consulta SQL para identificar o estado com mais relatos, e as cidades desse que possuem mais de 10 relatos

As consultas SQL retornam um data frame que contém o nome das cidades, formato do objeto e a quantidade de relatos que ocorreram nas cidades, como mostra a Figura 10.

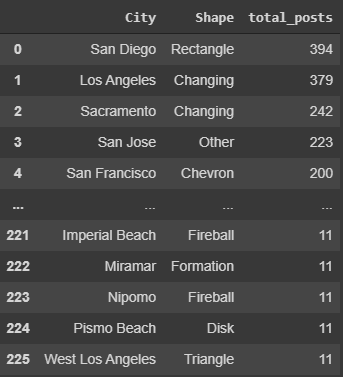


Figura 10 Resultado das consultas SQL da Figura 9

**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>

**Luzes de Phoenix:** <https://pt.wikipedia.org/wiki/Luzes_de_Phoenix>

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