

**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**

19/08/2020

**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) **10**

[**Referências**](#_5hyd2pq8xv2g) **11**

# 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 fase anterior por meio de web scraping, utilizando a linguagem python e gerando um arquivo Comma-separated values(CSV). Que será utilizado para a exploração. E o código estará armazenado no repositório do Github com a utilização do versionamento de código do git.

# 2. Descrição do problema

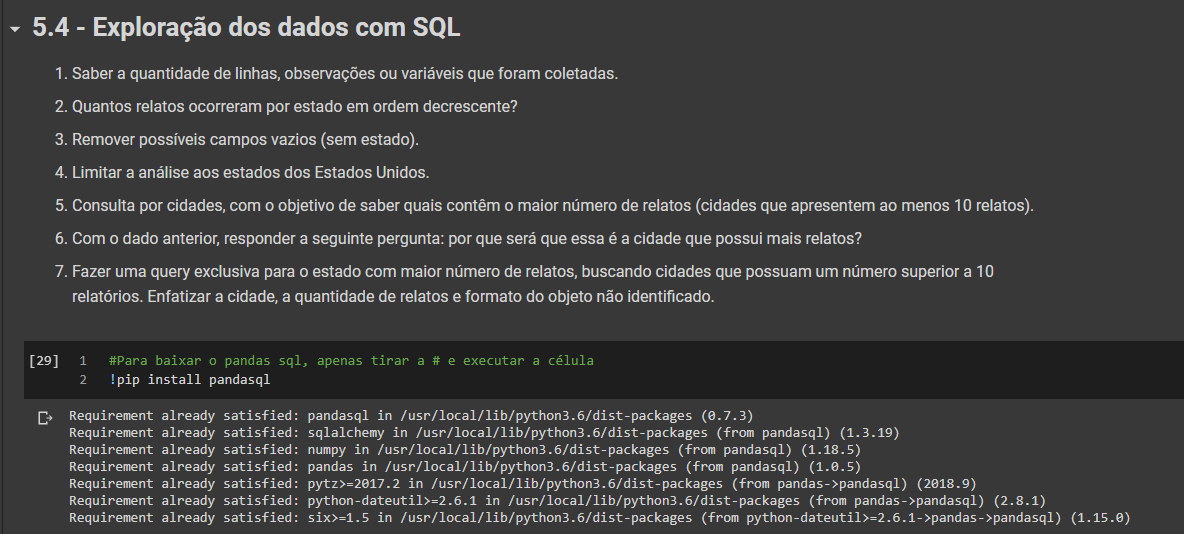
Ao iniciar o desenvolvimento do projeto podemos destacar a forma da coleta dos dados e como estão distribuídos no site e a partir desta análise inicial, foi visto que havia uma grande quantidade de dados em diversas páginas que precisar ser extraídas do site para realizar a coleta de dados e armazenar em DataFrames para que se possa gerar um arquivo csv para começar o estudo dos dados na primeira etapa do projeto. E a partir dos dados que foram gerados e armazenados no csv será feita a exploração de dados com a utilização da linguagem sql para a melhor filtragem dos dados.

# 3. Desenvolvimento

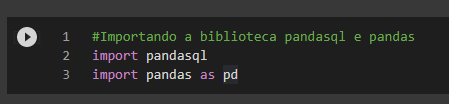
As tecnologias utilizadas para a elaboração da primeira parte do projeto na fase de coleta de dados consiste na linguagem de programação python com o auxílio de suas bibliotecas, como o requests para fazer a requisição do site e retornar o status code, a biblioteca BeautifulSoup para a leitura das páginas em Hyper Text Markup Language(HTML) e para o armazenamento de dados foi utilizada a biblioteca pandas, e o pandasql para que seja possível executar códigos SQL para fazer consultas no dataframe de ovnis, por meio do notebook e o ambiente de desenvolvimento Google Colab.

## 3.1 Código implementado

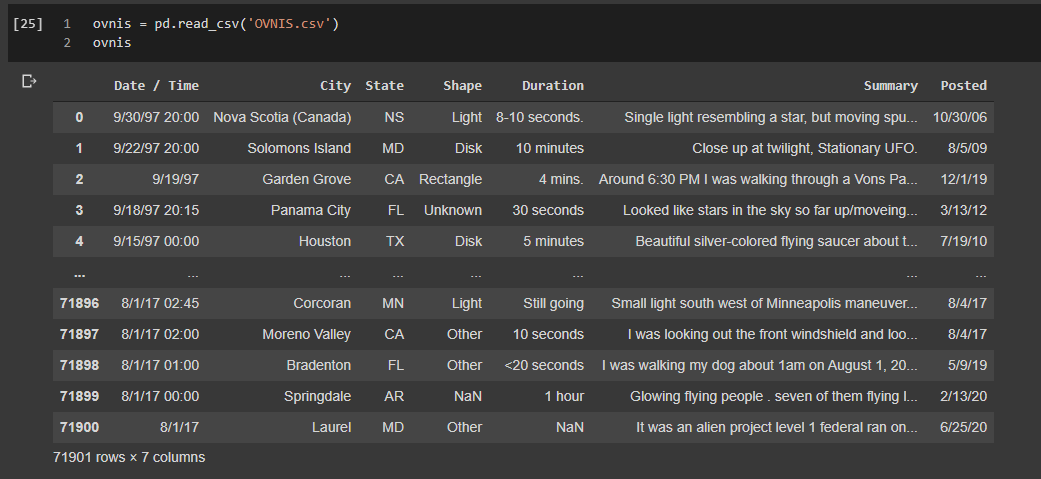
5.4 - Exploração de dados com SQL



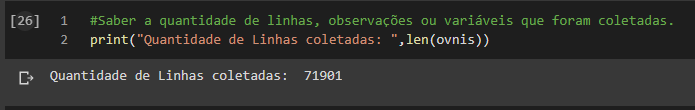
Baixando a biblioteca pandasql para a utilização de códigos SQL na exploração de dados



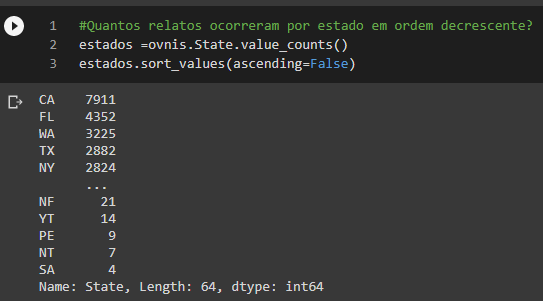
Importação das bibliotecas pandasql e pandas para exploração de dados com sql e a leitura e geração de arquivos csv, respectivamente.



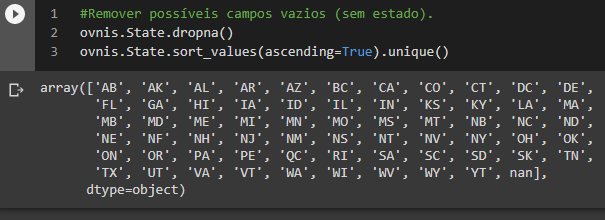
Leitura do arquivo OVNIS csv e armazenando na variável ovnis



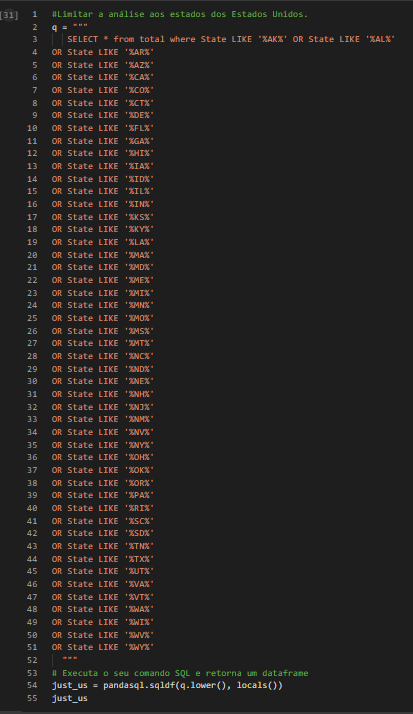
Imprimindo a quantidade de linhas que foram coletadas na atividade anterior



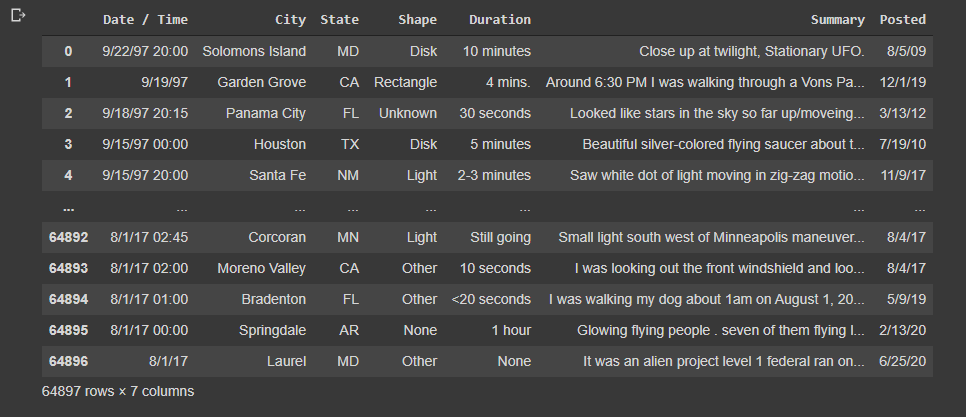
Declarando a variável estados e contando os estados por valores, ou seja, por siglas que apresentam na base de dados e depois organizando as siglas por ordem crescente.



Removendo os possíveis campos vazios da base de dados.



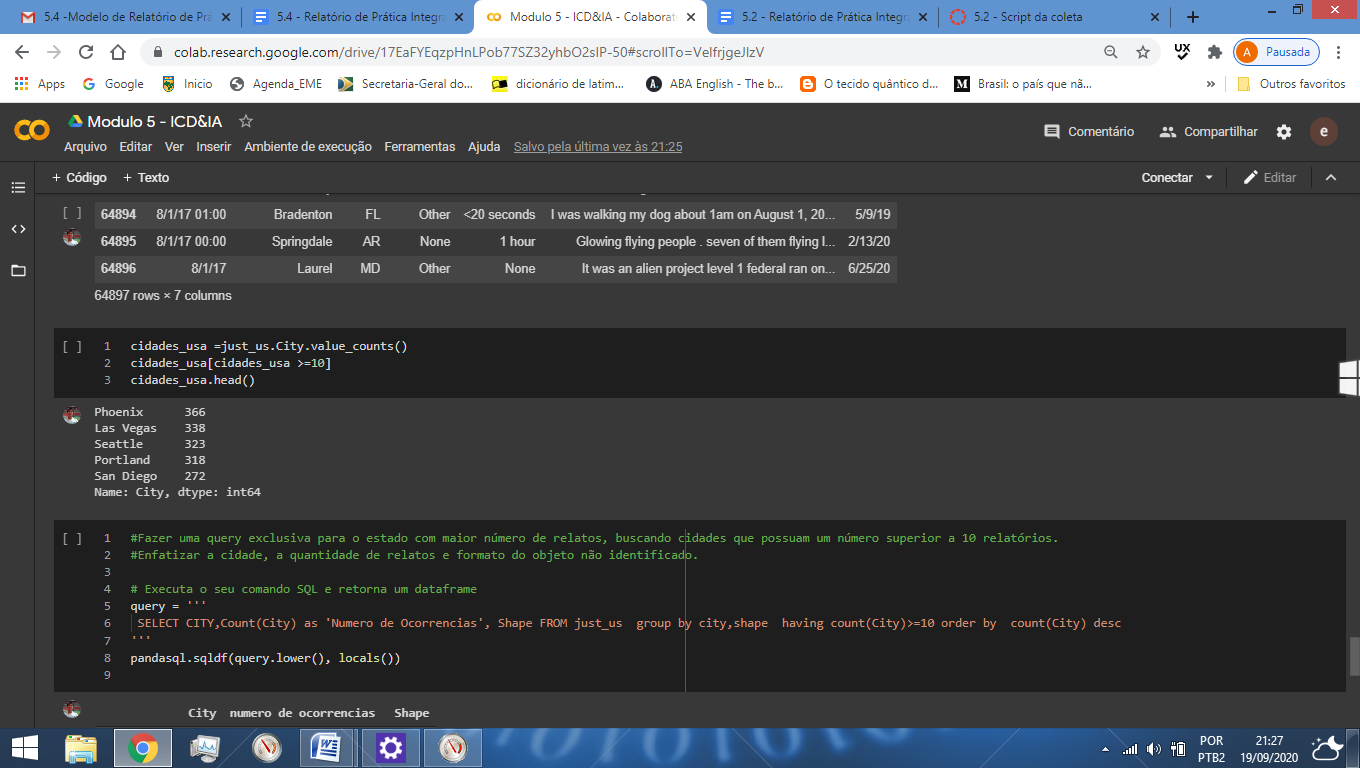
Limitando os estados para apenas estados norte americanos e executando o comando sql.



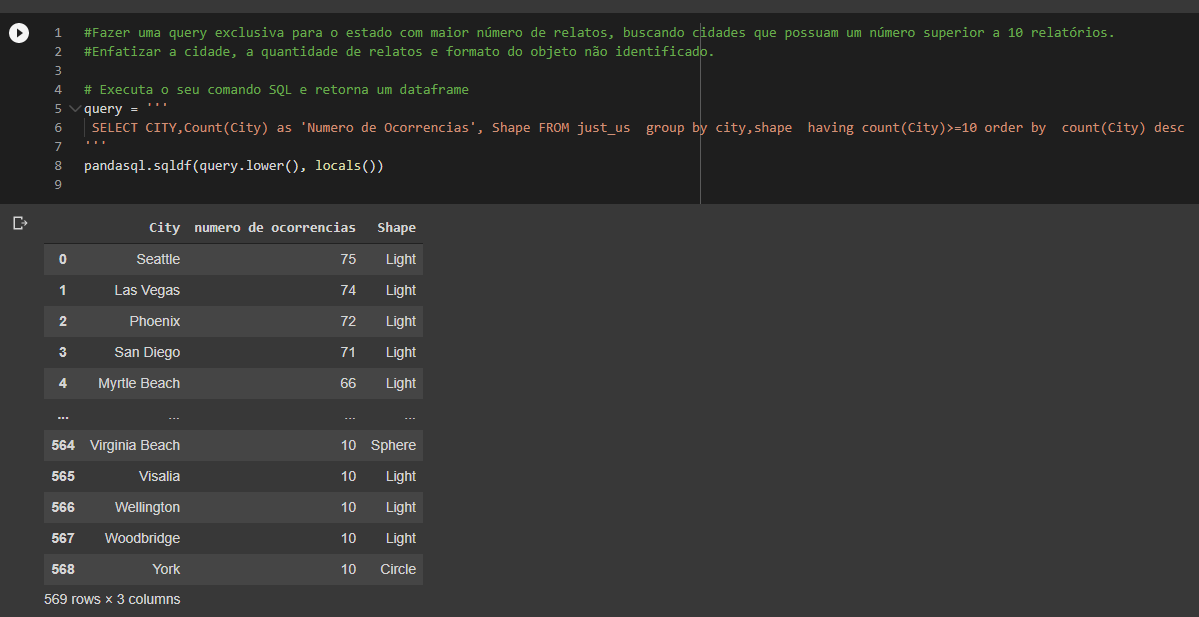
Resultado da consulta feita para filtrar apenas estados norte americanos.



Gerando um arquivo csv dos dados que foram gerados anteriormente, com a filtragem de apenas estados americanos.



Contando a quantidade de caso por cidades dos estados unidos, e mostrando apenas os estados que têm mais de 10 ocorrências. E mostrando apenas os 5 mais recorrentes.



Identificando os formatos dos objetos não identificados por cidade e quantidade de ocorrência do formato que foi relatado.

# 4. Considerações Finais

Nesta parte de exploração de dados, houve estudos das bibliotecas utilizadas e o modo que as consultas foram realizadas por meio da linguagem de consulta estruturada. No processo de desenvolvimento houve alguns empecilhos, como a forma de executar alguns comandos SQL para que se obtenha o exato resultado e a remoção de dados nulos, houve uma pequena dificuldade, pois acredita-se que alguns campos não estejam nulos, mas está escrito com sem valor.

# Referências

ANDRADE, Andrew. Análise Exploratória de Dados.**Escola de Dados**.[s.d.]. Disponível em:<https://escoladedados.org/tutoriais/analise-exploratoria-de-dados/>.Acesso em: 13, Setembro de 2020.

GEROLA, Letícia. 5 pythons hacks para a exploração do dados.**Medium**, 2020. Disponivelem:<https://medium.com/joguei-os-dados/4-python-hacks-para-explora%C3%A7%C3%A3o-de-dados-8c89931b6d1f>. Acesso em: 13, Setembro de 2020.

CARACIOLO, Marcel. **Introdução a análise exploratória com python e pandas**. 2013.(1h32m25s).Disponivel em: <https://www.youtube.com/watch?v=vIJwq6QjZL8>. Acesso em: 14, Setembro de 2020.