Relatório UFO Data – Parte 2

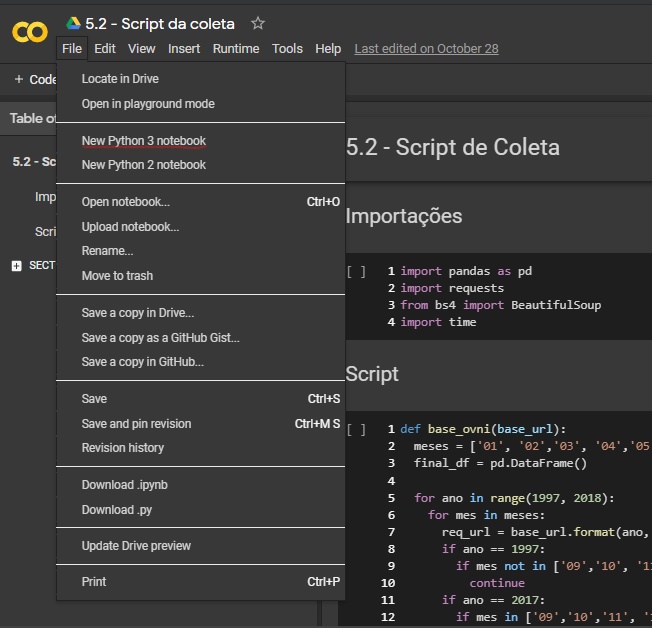
Levi Alves de Freitas Junior

**Problema**

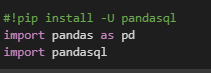
Criar uma análise utilizando banco de dados, utilizando a base de dados requisitada no primeiro relatório, e utilizar o pandasql para a exploração dos dados. Vamos utilizar a base gerada no primeiro relatório e explora-la.

Para a primeira parte vamos criar um novo notebook no colab research. Siga as instruções a seguir e veja as imagens ilustrativas:

1 – Logo após terminar as instruções da primeira etapa de análise dos relatos dos ovni’s, crie um novo notebook no colab, acessando no menu a opção File e logo após vá em New Python 3 notebook.



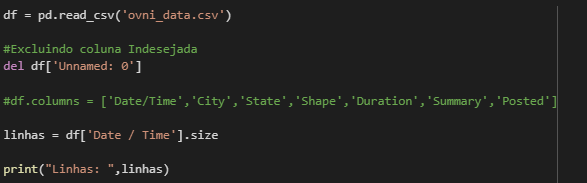
2 – Após criar um novo notebook, podemos começar a criar nosso código e inicialmente precisamos importar as bibliotecas pandas e para usar o banco de dados vamos importar o pandasql.



Primeiro selecione o comando - !pip install -U pandasql para instalar o pandassql no colab, logo após inicialize a célula no colab para importar o pandas e o pandas sql,

Obs. O import do pandasql pode não funcionar antes de utilizar o comando pip para utilizá-lo.

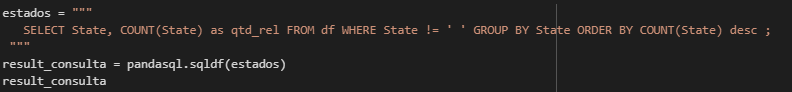
3 – Agora vamos importar a base de dados



Deletamos com o comando ‘del’ a coluna não nomeada pois não será necessária, então vamos imprimir a quantidade de linhas do nosso dataframe.

O resultado deve ser 99705 linhas.

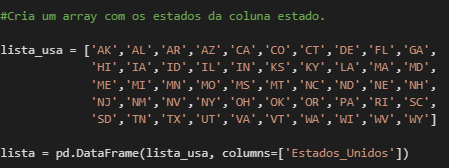
4 – Vamos organizar os dados e remover campos em branco



Fazemos uma pesquisa utilizando o pandasql, os estados com mais relatos de aparições de ovni’s e organizamos em ordem decrescente.

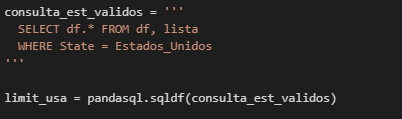
O estado com mais relato é CA – Califórnia com 11403 relatos.

5 – Neste momento vamos analisar somente os casos relatados nos Estados Unidos, com o seguinte código vamos comparar com a base de dados para filtrar e retornar os relatos.

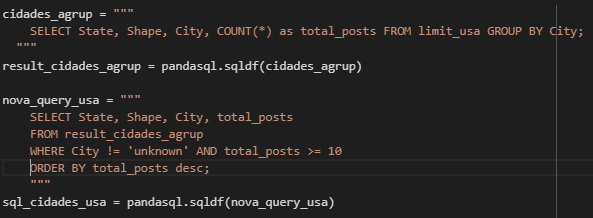


Criamos uma lista com as iniciais do estado e armazenamos esta lista à variável lista\_usa e logo após criamos uma coluna de um dataframe e armazenamos na variável lista.

Com a coluna ‘Estados\_Unidos’ do dataframe criada podemos agora comparar o dataframe original – ‘df’ com a coluna gerada e retorna somente os dados dos estados que constam na lista\_usa. Então fazemos a seguinte query.

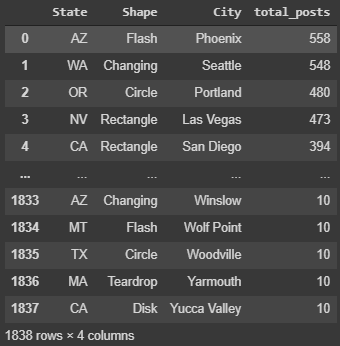


6 – Consulta e filtragem das cidades com 10 relatos ou mais



Fazemos a primeira query para agrupar as cidades e criamos uma subquery para tirar os campos não nomeados and que o total de posts (relatos) seja igual ou maior que 10.

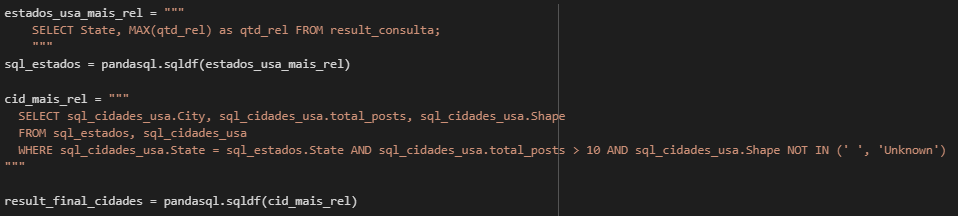
O resultado desta query deve ser como no modelo abaixo.



7 – Porque esta é a cidade que mais tem relatos.

**Phoenix** - No dia 13 de março de 1997, o céu do estado do Arizona e de Nevada, encheram de luzes, milhares de pessoas presenciaram o ocorrido, as pessoas afirmaram ver algo em formato triangular vagando pelo céu na cidade de Phoenix, capital do Arizona e cidade com mais relatos na análise de todos os casos de aparecimento de OVNI’s, o próprio governador presenciou e afirmou que seria um “objeto de outro mundo". Milhares de pessoas registraram essas aparições isso explica ser um dos locais com mais relatos.

8 – Busca nos estados com maior número de relato, juntamente buscando as cidades que tenham um número de relatos superior a 10 relatórios.



O resultado deve ser o seguinte rodando result\_final\_cidades.

