

**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**

29/09/2020

**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) **11**

[**Referências**](#_5hyd2pq8xv2g) **12**

# 1. Objetivos

Realizar a limpeza dos dados que se apresentam nulos, em branco ou possuem os valores Unknown. Para que se possa realizar uma exploração mais precisa de dados. Carregar os dados do csv Ovnis para um dataframe, remover as variáveis irrelevantes para análise, manter os formatos mais populares que contêm mais de 1000 ocorrências e salvar em um novo dataframe e converte-lo para csv.

# 2. Descrição do problema

Realizar a limpeza dos dados que são irrelevantes para a análise. E delimitar os dados da coluna dos states para apenas os estados dos norte americanos.

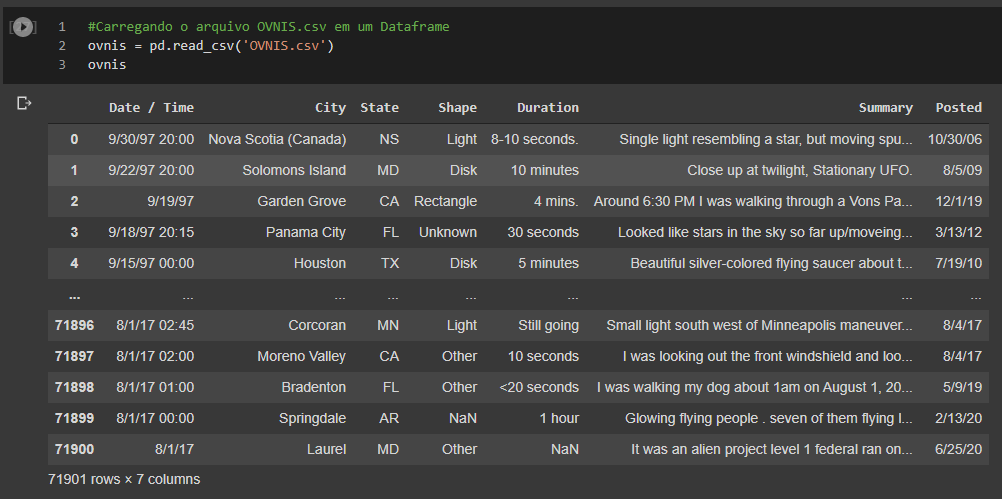
# 3. Desenvolvimento

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 pandas e pandasql para a análise dos dados e remoção dos dados irrelevantes, por meio do notebook e o ambiente de desenvolvimento Google Colab.

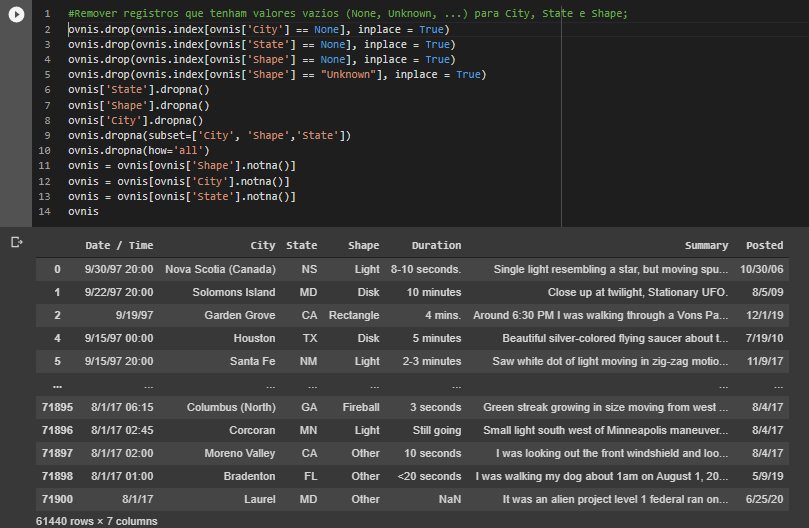
## 3.1 Código implementado

5.7 Limpeza dos dados

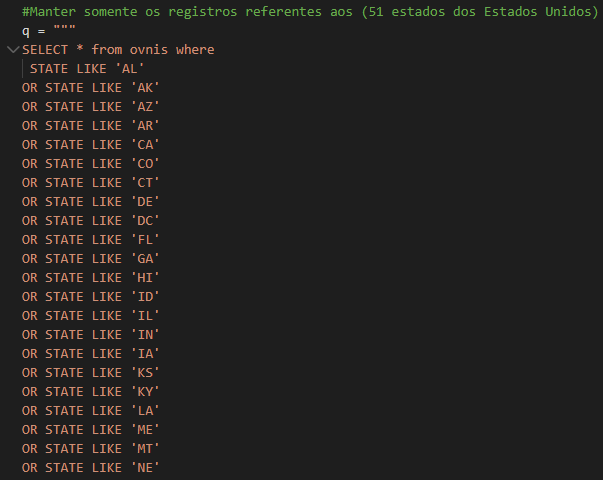
Carregando o arquivo Ovnis com a biblioteca pandas e atribuindo a uma variável.

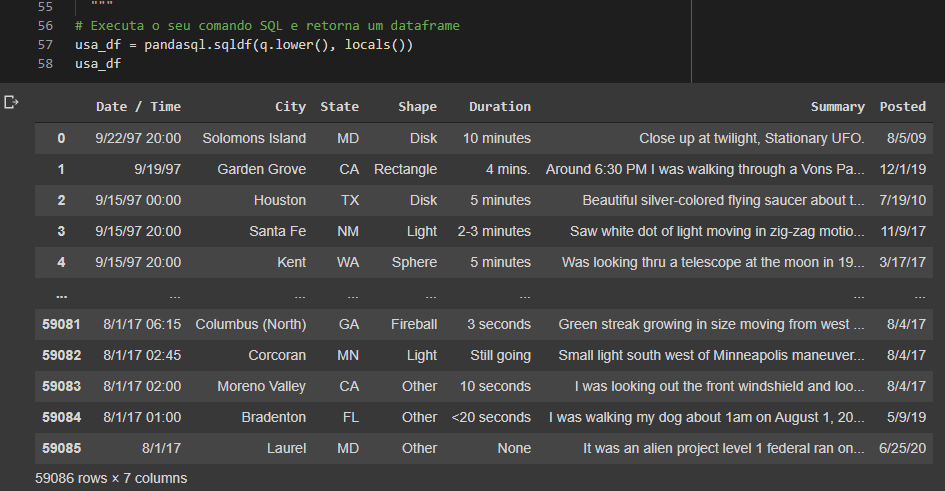


Remoção dos valores que se apresentam nulos, vazios ou desconhecidos nas colunas City, State e Shape. Utilizando as funções drop e especificando o index e a coluna para ser removida. Depois utilizando a função dropna para remover os valores nulos nas colunas.

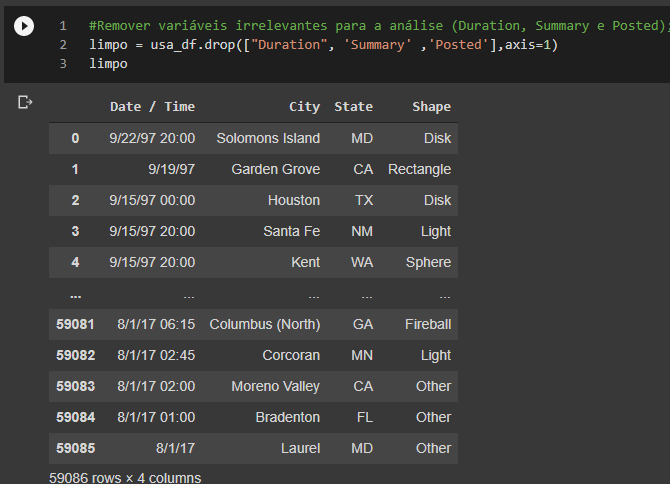


Executando a query para limitar apenas aos estados dos estados Unidos.

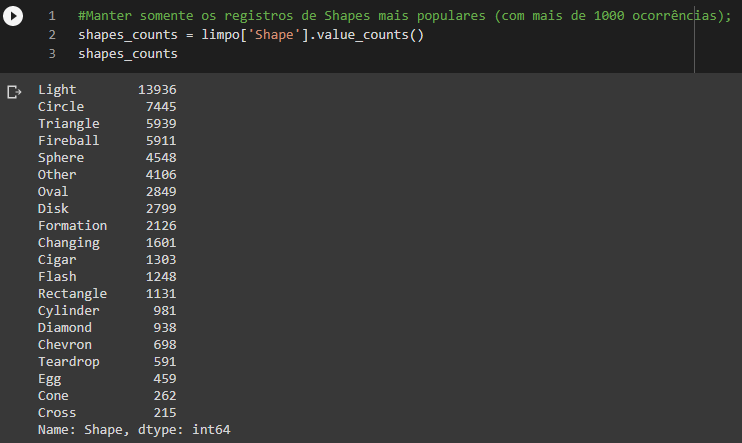




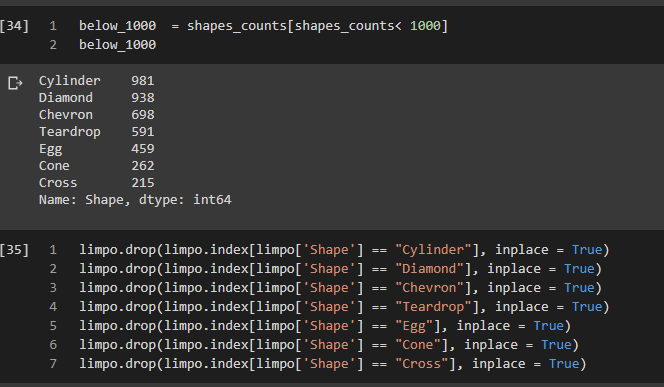
Removendo as colunas irrelevantes para a análise utilizando a função drop e atribuindo a uma nova variável que irá receber o novo dataframe limpo.



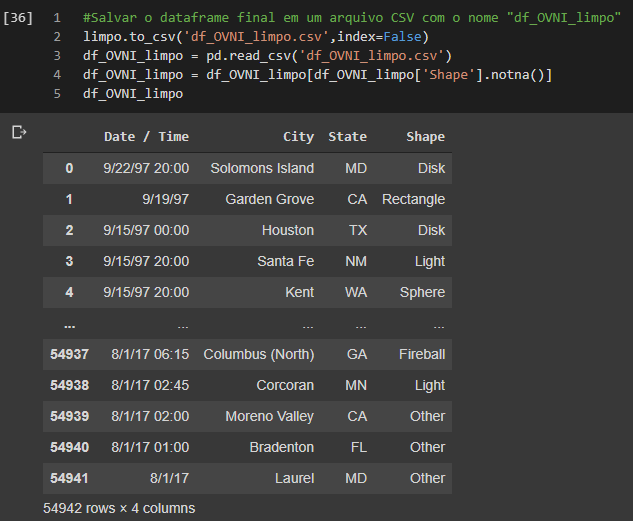
Verificando os formatos mais populares para que sejam removidos os que apresentam menos de 1000 ocorrências.



Verificando os valores abaixo de 1000 ocorrências e depois excluindo estes valores do dataframe.



Gerando um csv a partir da limpeza dos dados que foram solicitados. E depois lendo o csv gerado e mostrando na tela



# 4. Considerações Finais

Nesta parte do script foi realizado a remoção dos dados que não são essenciais para a continuidade do projeto e por não apresentar um padrão de dados nulos, ou irrelevantes, a remoção acabou sendo um pouco difícil, mas foi obtido o resultado esperado.

# Referências

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