

TECNOLOGIA EM SISTEMAS PARA INTERNET

Emanuelle Ferreira da Silva Neander Wendel Nobre Teixeira Lucas Matheus Silva

RELATÓRIO DE PRÁTICA INTEGRADA DE CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

Brasília - DF

25/09/2020

Sumário

1. Objetivos	3
2. Descrição do problema	4
3.1 Código implementado	5 5
4. Considerações Finais	9
Referências	10

1. Objetivos

O projeto na fase da exploração de dados, tem como objetivo analisar os dados que foram coletados na fase anterior, por meio de consultas feitas com a linguagem SQL.

2. Descrição do problema

O problema desta etapa foi a análise de forma precisa dos dados a serem coletados, pois muitas colunas dos dados podem aparecer de forma nula ou vazia, e por isso a remoção dos dados que foram rotulados como nulos, podem apresentar algum problema.

3. Desenvolvimento

O desenvolvimento desta etapa consiste em na exploração dos dados que foram coletados na etapa anterior. Utilizando o ambiente de desenvolvimento Google Colab e as bibliotecas pandas e pandasql para sua investigação de dados.

3.1 Código implementado

Instalando o pandas sql por meio do comando pip install pandasql.

```
Requirement already satisfied: pandasql in /usr/local/lib/python3.6/dist-packages (0.7.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.6/dist-packages (from pandasql) (1.0.5)
Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.6/dist-packages (from pandasql) (1.3.19)
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from pandasql) (1.18.5)
Requirement already satisfied: python-dateutil>=2.6.1 in /usr/local/lib/python3.6/dist-packages (from pandas->pandasql) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages (from pandas->pandasql) (2018.9)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.6/dist-packages (from python-dateutil>=2.6.1->pandas->pandasql) (1.15.0)
```

Importando as bibliotecas que serão utilizadas para o desenvolvimento dessa etapa do projeto como as biblioteca pandas que servirá para armazenar,organiza organiza código e salvar o documento e a biblioteca pandasql para a utilização de algumas funções.

```
[ ] 1 import pandas as pd
2 import pandasql
```

Lendo o arquivo ovnis.csv e atribuindo a uma variável chamada ovnis.

```
[ ] 1 ovnis = pd.read_csv("OVNIS.csv")
```

Lendo as 5 primeiros resultados por meio da função head.



Utilizando a função len para saber a quantidade de linhas.

```
1 # 5.4 Item 1 Saber a quantidade de linhas, observações ou variáveis que foram coletadas.
2 3 len(ovnis)

102391
```

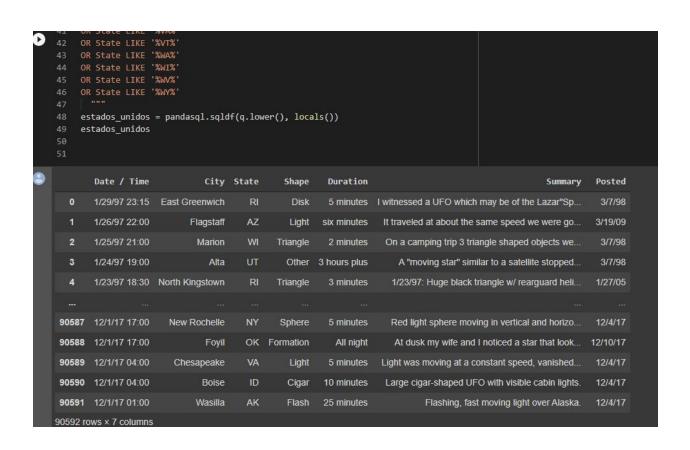
Criando uma variável chamada estado_desc que está armazenado o df ovnis, que por meio da função value_counts mostra a quantidade de relatos que se encontra na tabela state.

```
# Item 2 Quantos relatos ocorreram por estado em ordem decrescente?
     estado_desc = ovnis["State"].value_counts() # contando os relatos pela coluna state
     estado_desc.sort_values(ascending=False)
CA
      11745
FL
       5721
       4992
WA
       4230
NY
       3952
PE
         20
PR
         19
         19
YT
VI
Name: State, Length: 68, dtype: int64
```

Criando uma query para limitar e analisar os estados dos estados unidos.

```
q =
  SELECT * from ovnis where State LIKE '%AK%' OR State LIKE '%AL%'
OR State LIKE '%AR%'
OR State LIKE '%AZ%'
OR State LIKE '%CA%'
OR State LIKE '%CO%'
OR State LIKE '%DE%'
OR State LIKE '%FL%'
OR State LIKE
OR State LIKE '%IA%'
OR State LIKE '%IL%'
OR State LIKE '%MN%' OR State LIKE '%MO%' OR State LIKE '%MS%' OR State LIKE '%MT%' OR State LIKE '%MC%' OR State LIKE '%MD%'
OR State LIKE
OR State LIKE
OR State LIKE
               '%NM%
OR State LIKE
OR State LIKE
OR State LIKE
```

Retornando a tabela.

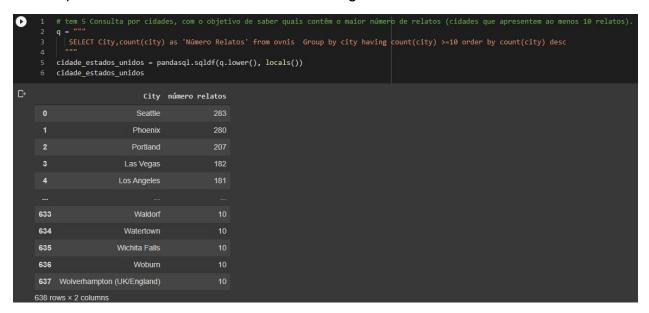


Removendo os campos vazios da tabela state por meio da função dropna().

```
#Item 3 Remover possíveis campos vazios (sem estado).
     ovnis["State"].dropna()
          RI
1
          AZ
2
          WI
          ON
          UT
102385
         NY
102386
         OK
          VA
102387
102388
          ID
102390
          AK
```

Contando o número de cidade que possuem um número maior ou que apresentem pelo menos 10 relatos por meio da tabela city.

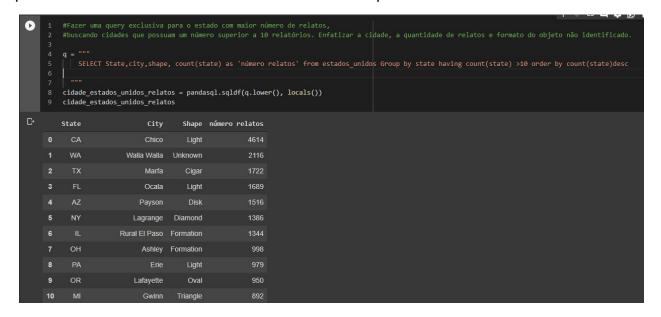
Criando uma query e consultando a tabela cidade com o objetivo de saber qual a cidade que contém o maior número de casos registrados.



Por que será que a cidade de Phoenix possui mais relatos?

```
[] 1 #Com o dado anterior, responder a seguinte pergunta: por que será que essa é a cidade que possui mais relatos?
2 # A cidade de Phoenix resgistrou o maior número de relatos em relação as outras cidades, pois constatou um alto número de testemunha
3 # que avistaram diversoso objteos no céu dos EUA ao mesmo tempo, esse fenômeno ficou conhecida na midia social como luzes de phoenix.
```

Criando uma query que está selecionando a tabela, state, city e shape, que está contando os relatos por meio da tabela state da variável estados unidos que agrupa pelo estado tendo os números de relatos maiores que 10.



4. Considerações Finais

Podemos concluir que por meio da exploração dos dados apresentando, o estado que mostrou o maior número de casos registrados phoenix, chico, para chegar no resultados, ao longo do caminho tivemos vários problema em relação a construção do sql e na análise, pois houve dúvidas se a consulta se apresentada foi que foi requisitada no canvas.

Referências

Zingano Jr. Arthur. De uma galáxia muito distante? Ovni volta a ser testemunhado em Phoenix. Blasting News Brasil. 25 de Dezembro de 2017. Disponível em < https://br.blastingnews.com/mundo/2017/12/de-uma-galaxia-muito-distante-ovni-volta-a-ser-testemunhado-em-phoenix-002251639.html/>. Acesso em 25 de setembro de 2020

Rigue, André. Luzes de Phoenix completam 20 anos de mistério.Band, Uol notícias. 13 de Março de 2017. Seção Notícias. Disponível em < https://noticias.band.uol.com.br/noticias/100000848525/luzes-de-phoenix-completam-20 -anos-de-misterio.html/> Acesso em 25 de Setembro de 2020

https://colab.research.google.com/drive/1kNU5KhNR4NB-asYukHI5JC3LJPWr1rva?authuser=1