

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. Desenvolvimento	5
3.1 Código implementado	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.

```
[ ] 1 !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.

```
[ ] 1  ovnis.head(5)
    2
```

	Date / Time	City	State	Shape	Duration	Summary	Posted
0	1/29/97 23:15	East Greenwich	RI	Disk	5 minutes	I witnessed a UFO which may be of the Lazar"Sp...	3/7/98
1	1/26/97 22:00	Flagstaff	AZ	Light	six minutes	It traveled at about the same speed we were go...	3/19/09
2	1/25/97 21:00	Marion	WI	Triangle	2 minutes	On a camping trip 3 triangle shaped objects we...	3/7/98
3	1/25/97 06:00	Mount Hope/Binbrook (Canada)	ON	Disk	1/2 hour	A large disk or saucer type object, approximat...	4/28/01
4	1/24/97 19:00	Alta	UT	Other	3 hours plus	A "moving star" similar to a satellite stopped...	3/7/98

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.

```
[ ] 1  # Item 2 Quantos relatos ocorreram por estado em ordem decrescente?
    2
    3  estado_desc = ovnis["State"].value_counts() # contando os relatos pela coluna state
    4  estado_desc.sort_values(ascending=False)
```

CA	11745
FL	5721
WA	4992
TX	4230
NY	3952
...	
PE	20
PR	19
YT	19
YK	5
VI	1

Name: State, Length: 68, dtype: int64

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

```
1 #Item 4
2 q = ""
3 | SELECT * from ovnis where State LIKE '%AK%' OR State LIKE '%AL%'
4 OR State LIKE '%AR%'
5 OR State LIKE '%AZ%'
6 OR State LIKE '%CA%'
7 OR State LIKE '%CO%'
8 OR State LIKE '%CT%'
9 OR State LIKE '%DE%'
10 OR State LIKE '%FL%'
11 OR State LIKE '%GA%'
12 OR State LIKE '%HI%'
13 OR State LIKE '%IA%'
14 OR State LIKE '%ID%'
15 OR State LIKE '%IL%'
16 OR State LIKE '%IN%'
17 OR State LIKE '%KS%'
18 OR State LIKE '%KY%'
19 OR State LIKE '%LA%'
20 OR State LIKE '%MA%'
21 OR State LIKE '%MD%'
22 OR State LIKE '%ME%'
23 OR State LIKE '%MI%'
24 OR State LIKE '%MN%' OR State LIKE '%MO%' OR State LIKE '%MS%' OR State LIKE '%MT%' OR State LIKE '%NC%' OR State LIKE '%ND%'
25 OR State LIKE '%NE%'
26 OR State LIKE '%NH%'
27 OR State LIKE '%NJ%'
28 OR State LIKE '%NM%'
29 OR State LIKE '%NV%'
30 OR State LIKE '%NY%'
31 OR State LIKE '%OH%'
```

Retornando a tabela.

```
41 OR State LIKE '%VA%'
42 OR State LIKE '%VT%'
43 OR State LIKE '%WA%'
44 OR State LIKE '%WI%'
45 OR State LIKE '%WV%'
46 OR State LIKE '%WY%'
47 | ""
48 estados_unidos = pandasql.sqldf(q.lower(), locals())
49 estados_unidos
50
51
```

	Date / Time	City	State	Shape	Duration	Summary	Posted
0	1/29/97 23:15	East Greenwich	RI	Disk	5 minutes	I witnessed a UFO which may be of the Lazar"Sp...	3/7/98
1	1/26/97 22:00	Flagstaff	AZ	Light	six minutes	It traveled at about the same speed we were go...	3/19/09
2	1/25/97 21:00	Marion	WI	Triangle	2 minutes	On a camping trip 3 triangle shaped objects we...	3/7/98
3	1/24/97 19:00	Alta	UT	Other	3 hours plus	A "moving star" similar to a satellite stopped...	3/7/98
4	1/23/97 18:30	North Kingstown	RI	Triangle	3 minutes	1/23/97: Huge black triangle w/ rearguard heli...	1/27/05
...
90587	12/1/17 17:00	New Rochelle	NY	Sphere	5 minutes	Red light sphere moving in vertical and horizo...	12/4/17
90588	12/1/17 17:00	Foyil	OK	Formation	All night	At dusk my wife and I noticed a star that look...	12/10/17
90589	12/1/17 04:00	Chesapeake	VA	Light	5 minutes	Light was moving at a constant speed, vanished...	12/4/17
90590	12/1/17 04:00	Boise	ID	Cigar	10 minutes	Large cigar-shaped UFO with visible cabin lights.	12/4/17
90591	12/1/17 01:00	Wasilla	AK	Flash	25 minutes	Flashing, fast moving light over Alaska.	12/4/17

90592 rows x 7 columns

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

```
[ ] 1 #Item 3 Remover possíveis campos vazios (sem estado).
    2 ovnis["State"].dropna()

0      RI
1      AZ
2      WI
3      ON
4      UT
..
102385  NY
102386  OK
102387  VA
102388  ID
102390  AK
Name: State, Length: 95084, dtype: object
```

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

```
1 # tem 5 Consulta por cidades, com o objetivo de saber quais contêm o maior número de relatos (cidades que apresentem ao menos 10 relatos).
2
3 cidade_estados_unidos = estados_unidos.City.value_counts()
4 cidade_estados_unidos[cidade_estados_unidos>=10]

Seattle      283
Phoenix      280
Portland     207
Las Vegas    182
Los Angeles  181
...
Mt. Vernon   10
Palatine     10
Stanwood     10
Charlottesville 10
Ames         10
Name: City, Length: 564, dtype: int64
```


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

```
1 # tem 5 Consulta por cidades, com o objetivo de saber quais contém o maior número de relatos (cidades que apresentem ao menos 10 relatos).
2 q = """
3     SELECT City,count(city) as 'Número Relatos' from ovnis Group by city having count(city) >=10 order by count(city) desc
4     """
5 cidade_estados_unidos = pandasql.sqldf(q.lower(), locals())
6 cidade_estados_unidos
```

	City	número relatos
0	Seattle	283
1	Phoenix	280
2	Portland	207
3	Las Vegas	182
4	Los Angeles	181
...
633	Waldorf	10
634	Watertown	10
635	Wichita Falls	10
636	Woburn	10
637	Wolverhampton (UK/England)	10

638 rows x 2 columns

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 registrou o maior número de relatos em relação as outras cidades, pois constatou um alto número de testemunha
3 # que avistaram diversosos objteos no céu dos EUA ao mesmo tempo, esse fenômeno ficou conhecida na mídia 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.

```
1 #Fazer uma query exclusiva para o estado com maior número de relatos,
2 #buscando cidades que possuam um número superior a 10 relatórios. Enfatizar a cidade, a quantidade de relatos e formato do objeto não identificado.
3
4 q = """
5     SELECT State,city,shape, count(state) as 'número relatos' from estados_unidos Group by state having count(state) >10 order by count(state)desc
6     """
7
8 cidade_estados_unidos_relatos = pandasql.sqldf(q.lower(), locals())
9 cidade_estados_unidos_relatos
```

	State	City	Shape	número relatos
0	CA	Chico	Light	4614
1	WA	Walla Walla	Unknown	2116
2	TX	Marfa	Cigar	1722
3	FL	Ocala	Light	1689
4	AZ	Payson	Disk	1516
5	NY	Lagrange	Diamond	1386
6	IL	Rural El Paso	Formation	1344
7	OH	Ashley	Formation	998
8	PA	Erie	Light	979
9	OR	Lafayette	Oval	950
10	MI	Gwinn	Triangle	892

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>