

**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**

25/10/2020

**Brasília - DF**

**2020**

# Sumário

<b>1. Objetivos</b>	<b>3</b>
<b>2. Descrição do problema</b>	<b>4</b>
<b>3. Desenvolvimento</b>	<b>5</b>
3.1 Código implementado	5
<b>4. Considerações Finais</b>	<b>9</b>
<b>Referências</b>	<b>10</b>

# 1. Objetivos

Neste último sprint foi requisitado a recuperação dos de visualização da cidade de Phoenix e realizar suas observações como, ordenar e agrupar por mês, dia e ano. Visualizar os dados em forma de série temporal, gerando um gráfico de linha, tendo a evolução do número de observações ao longo tempo em anos.

Além disso, realizar a construção dos conjuntos separando em treinamentos e teste, separando 70% das observações para o treinamento e 30% para teste. Realizar a investigação dos parâmetros e discriminar o melhor modelo, utilizando a função fit para ajustar o modelo e medir sua qualidade. Após isso, realizar a previsão do melhor modelo utilizando a função forecast

## 2. Descrição do problema

Nesta etapa o problema é realizar a separação dos dados de treinamento e testes e a realização da predição futura

## 3. Desenvolvimento

As tecnologias utilizadas para a elaboração para esta terceira fase do projeto, será a linguagem python e algumas bibliotecas para ajudar no desenvolvimento, com a biblioteca pandas, statsmodel e pandasql, por meio do notebook e o ambiente de desenvolvimento Google Colab.

### 3.1 Código implementado

#### 5.12 - Análise temporal

Realizando a importação da biblioteca pandas. No segundo bloco atribuindo os dados do arquivo csv que será lido pela função read\_csv do pandas e criando um dataframe a partir desses dados.

```
[74] 1 import pandas as pd
```

```
1 ovnis_preparado = pd.read_csv('df_OVNI_preparado.csv')
2 ovnis_preparado
```

	City	State	Shape	Sight_day	Sight_month	Sight_time	Sight_date	Sight_weekday
0	Solomons Island	MD	Disk	22	9	20:00:00	1997-09-22	Segunda-feira
1	Annapolis	MD	Triangle	15	8	23:00:00	1998-08-15	Sábado
2	Chesapeake Bay	MD	Disk	26	10	20:00:00	1999-10-26	Terça-feira
3	Frederick	MD	Sphere	7	7	01:45:00	2000-07-07	Sexta-feira
4	Wheaton	MD	Triangle	1	10	20:00:00	2000-10-01	Domingo
...	...	...	...	...	...	...	...	...
54938	Washington, D.C. (above I-295 bridge)	DC	Flash	13	3	05:40:00	2016-03-13	Domingo
54939	Washington, D.C.	DC	Circle	12	4	01:30:00	2016-04-12	Terça-feira
54940	Washington	DC	Triangle	8	8	22:00:00	2016-08-08	Segunda-feira
54941	Washington, D.C.	DC	Other	2	10	11:00:00	2016-10-02	Domingo
54942	Washington, DC	DC	Circle	12	5	14:30:00	2017-05-12	Sexta-feira

Coletando do Dataframe apenas as ocorrências na cidade de Phoenix e atribuindo a um novo dataframe chamado cidade\_phoenix e após isso é realizado a ordenação dos valores pela coluna Sight\_date

```
1 cidade_phoenix = ovnis_preparado[ovnis_preparado['City']=='Phoenix']
2 cidade_phoenix.sort_values(by='Sight_date')
```

	City	State	Shape	Sight_day	Sight_month	Sight_time	Sight_date	Sight_weekday
24043	Phoenix	AZ	Disk	12	6	02:30:00	1999-06-12	Sábado
24061	Phoenix	AZ	Changing	12	11	23:33:00	2001-11-12	Segunda-feira
24073	Phoenix	AZ	Cigar	31	5	13:00:00	2003-05-31	Sábado
24090	Phoenix	AZ	Light	21	2	18:00:00	2005-02-21	Segunda-feira
24091	Phoenix	AZ	Light	20	3	03:30:00	2005-03-20	Domingo
...	...	...	...	...	...	...	...	...
26019	Phoenix	AZ	Other	15	6	15:35:00	2017-06-15	Quinta-feira
26035	Phoenix	AZ	Oval	6	7	21:25:00	2017-07-06	Quinta-feira
26031	Phoenix	AZ	Fireball	26	7	04:20:00	2017-07-26	Quarta-feira
26050	Phoenix	AZ	Flash	4	8	21:15:00	2017-08-04	Sexta-feira
26048	Phoenix	AZ	Light	14	8	00:20:00	2017-08-14	Segunda-feira

305 rows x 8 columns

Realizando a instalação da biblioteca pandasql para realizar as consultas por meio de queries em SQL

```
1 !pip install pandasql
2 import pandasql
```

Utilizando a biblioteca pandasql e construindo uma query para realizar a consutla que foi requisitada, que no caso é agrupar por dia as visualizações e ordená-los pela data

```
1 # Executa o seu comando SQL e retorna um dataframe
2 query = '''
3 | SELECT Sight_date ,Count(*) as Views FROM cidade_phoenix group by Sight_date, Sight_month order by Sight_date
4 | '''
5 views_phoenix= pandasql.sqldf(query.lower(), locals())
6 views_phoenix
```

	Sight_date	views
0	2001-11-12	1
1	2003-05-31	1
2	2005-06-08	1
3	2005-10-15	2
4	2006-04-30	1
...	...	...
198	2017-06-15	2
199	2017-07-06	1
200	2017-07-26	2
201	2017-08-04	1

Utilizando a função do pandas `to_datetime` para converter a coluna `sight_date` para o tipo data para poder realizar a separação do ano e consultar qual foi a quantidade de ocorrências por ano que aconteceram na cidade de Phoenix. Em seguida é realizada a consulta para conferir os tipos das colunas, para ter a certeza que o tipo dela foi alterada

```
[79] 1 views_phoenix['Sight_date'] = pd.to_datetime(views_phoenix['Sight_date'])
      2 views_phoenix.dtypes
```

```
↳ Sight_date    datetime64[ns]
   views        int64
   dtype: object
```

Separando o ano da coluna data e criando uma nova coluna com o nome de `Sight_year`.

```
[80] 1 views_phoenix['Sight_year'] = views_phoenix['Sight_date'].dt.year
      2 views_phoenix
```

```
↳
```

	Sight_date	views	Sight_year
0	2001-11-12	1	2001
1	2003-05-31	1	2003
2	2005-06-08	1	2005
3	2005-10-15	2	2005
4	2006-04-30	1	2006
...	...	...	...
198	2017-06-15	2	2017
199	2017-07-06	1	2017
200	2017-07-26	2	2017
201	2017-08-04	1	2017
202	2017-08-14	3	2017

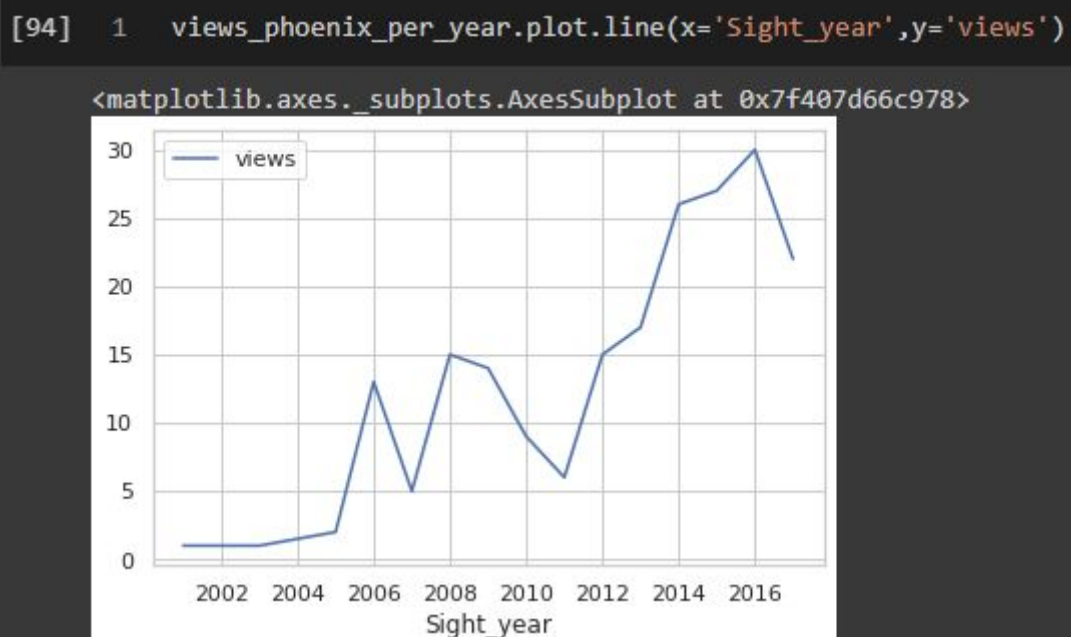
203 rows x 3 columns

Utilizando a linguagem SQL para realizar o agrupamento e contagem das visualizações pelo ano. E atribuindo o resultado da consulta a um novo dataframe nomeado `views_phoenix_per_year`.

```
1 # Executa o seu comando SQL e retorna um dataframe
2 query = '''
3 | SELECT Count(*) as views, Sight_year FROM views_phoenix group by Sight_year
4 | '''
5 views_phoenix_per_year= pandasql.sqldf(query.lower(), locals())
6 views_phoenix_per_year
```

	views	Sight_year
0	1	2001
1	1	2003
2	2	2005
3	13	2006
4	5	2007
5	15	2008
6	14	2009
7	9	2010
8	6	2011
9	15	2012
10	17	2013

Gerando o gráfico para a observar as ocorrências que são relatadas por ano.





## 4. Considerações Finais

Nesta última etapa do projeto foi realizado a visualização das ocorrências acontecidas na cidade de Phoenix e foram realizados as consultas por ano para conhecer o fluxo de observações que ocorrem por ano na região.

Além disso a realização da separação dos dados de treinamento e testes deveriam ser feitos para a conclusão desta etapa.

## Referências