Projeto de prática Integrada de ciência de dados, inteligência artificial e machine learning



Relatório de avistamento de Objetos Voadores Não Identificados.

Sprint 3 → Análise temporal

Curso: Tecnologia em sistemas para internet

Estudantes:

Brenda Lopes Miranda Teixeira

Mateus Gomes da Silva Fonteles

Rickson Queiroz Marques de Souza

Samuel Araújo Lopes

Professores

Fábio Henrique

Diego Queiroz

Ana Régia

Brasília, agosto de 2021

Sumário

1. Objetivos 3
2. Descrição do problema 4
3. Desenvolvimento 5
4. Código implementado 6
5. Considerações Finais 12

# 1. Objetivos

A análise é a etapa do trabalho em que extraímos valor dos dados, criando hipóteses com base nos *insights* gerados pelos dados. Vamos agora criar um modelo preditivo a fim de realizar uma estimativa futura baseada nas experiências prévias armazenadas no nosso dataset.

Para alcançar este resultado, aplicamos aos nossos dados um modelo de machine learning para estimar, a partir dos dados contidos em nosso dataset, a quantidade de avistamentos que haverá no futuro.

# 2. Descrição do problema

Para realizar o trabalho proposto, vamos levantar os dados sobre a cidade de Fênix, EUA, para logo visualizá-los em forma de série temporal, com gráficos de barras e de linha, vamos construir nossos conjuntos de treinamento e de teste e investigar os parâmetros para encontrar o melhor modelo. Por último, vamos realizar uma previsão de avistamentos futuros utilizando a função SARIMAX.

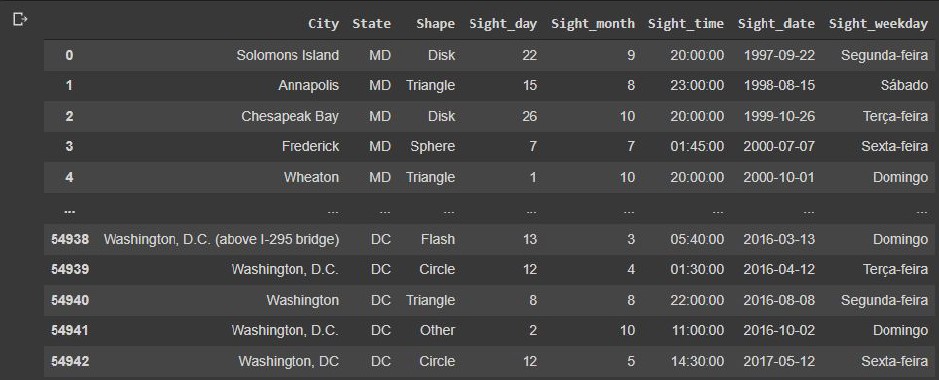
# 3. Desenvolvimento

Este trabalho está sendo desenvolvido usando um Script Python por ser uma linguagem orientada a objetos é bastante maleável, o grupo está utilizando a plataforma Google Colaboratory, assim todos podem modificar e acrescentar o código quando necessário.

# 3. Código implementado

* Importar a biblioteca Pandas
* Criar um dataframe com o arquivo .CSV

|  |
| --- |
| # Importação do panda |
| import pandas as pd |
| # Carrega seu arquivo csv |
| ovnis\_preparado = pd.read\_csv('df\_OVNI\_preparado.csv') |
| ovnis\_preparado |



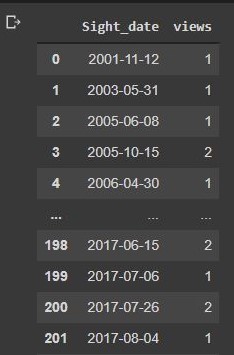
* Criar dataframe somente com os dados da cidade de Fênix e ordenar por data de avistamento:

|  |
| --- |
| #filtra a cidade dentro do csv |
| cidade\_phoenix = ovnis\_preparado[ovnis\_preparado['City']=='Phoenix'] |
| cidade\_phoenix.sort\_values(by='Sight\_date') |



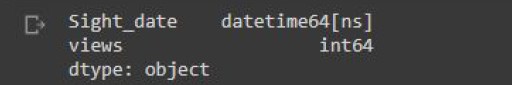
* Importar a biblioteca PandaSQL
* Construir uma query para contar os avistamentos por dia, ordenando pela data:

|  |  |
| --- | --- |
| import pandasql |  |
| # Roda o seu comando SQL e retorna um dataframe |  |
| query = ''' |  |
| SELECT Sight\_date ,Count(\*) as Views FROM cidade\_phoenix | group by |
| Sight\_day, Sight\_month order by Sight\_date |  |
| ''' |  |
| views\_phoenix= pandasql.sqldf(query.lower(), locals()) |  |
| views\_phoenix |  |

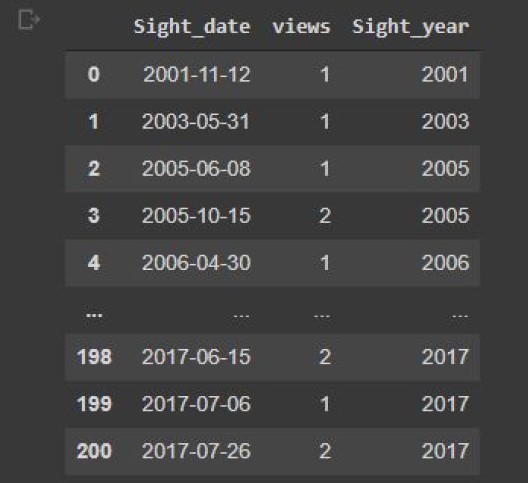


* Converter a coluna sight\_date para o tipo data
* Consultar o tipo da coluna para confirmar o sucesso da operação
* Separar o ano da coluna data e criar uma nova coluna chamada sight\_year

|  |  |
| --- | --- |
| #filtra a data e o ano |  |
| views\_phoenix['Sight\_date'] | = |
| pd.to\_datetime(views\_phoenix['Sight\_date']) |  |
| views\_phoenix.dtypes |  |



|  |
| --- |
| views\_phoenix['Sight\_year'] = views\_phoenix['Sight\_date'].dt.year |
| views\_phoenix |



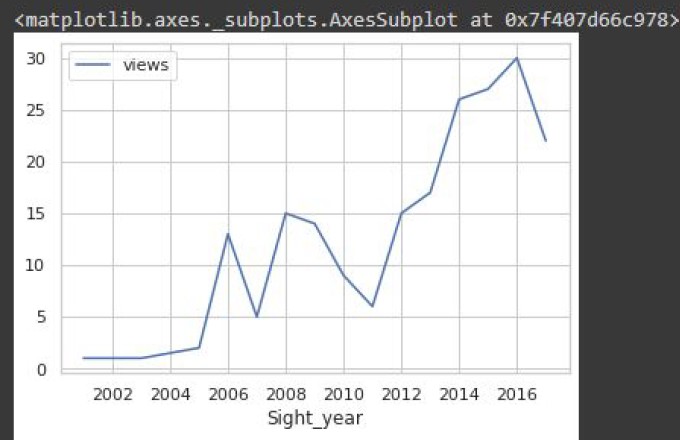
* Criar uma nova consulta SQL para ver os avistamentos por ano (na cidade de Fênix) e guardar o resultado em um novo dataframe:

|  |  |
| --- | --- |
| query = ''' |  |
| SELECT Count(\*) as views, Sight\_year FROM views\_phoenix group | by |
| Sight\_year |  |
| ''' |  |
| views\_phoenix\_per\_year= pandasql.sqldf(query.lower(), locals()) |  |
| views\_phoenix\_per\_year |  |



* Gerar um gráfico para observar a quantidade de relatos por ano:

views\_phoenix\_per\_year.plot.line(x='Sight\_year',y='views')



# 4. Considerações Finais

Nesta etapa do trabalho visualizamos as ocorrências relatadas na cidade de Fênix, consultando por data, ano e número de ocorrências, a fim de conhecer o fluxo de ocorrências nesta localidade.

Além disso, expressamos este dado também em um gráfico de linhas.