

TECNOLOGIA EM SISTEMAS PARA INTERNET

**Daniel Evangelista Pereira
Leonardo Campos Muniz**

**RELATÓRIO DE PRÁTICA INTEGRADA
DE
CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA**

Brasília - DF

25/03/2021

Sumário

1. Objetivos	3
2. Descrição do problema	4
3. Desenvolvimento	5
3.1 Código implementado	5
4. Considerações Finais	6
Referências	7

1. Objetivos

Este é o documento é o relatório referente a última sprint de atividades da PRÁTICA INTEGRADA DE CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA.

A equipe é composta pelos alunos Daniel Evangelista Pereira e Leonardo Campos Muniz, com a mentoria de Thiago Marinha e utilizará a metodologia de desenvolvimento ágil SCRUM, com um total de 4 sprints de uma semana, estando atualmente na quarta e última semana.

O presente relatório trata do desafio de **Análise temporal**, como a atividade já sugere será feito uma análise com base no tempo, mais especificamente será analisado da cidade de Phoenix no estados unidos usando de base o banco de dados gerado nas sprints anteriores.

2. Descrição do problema

O problema consiste em alguns fatores, o primeiro de carregar a tabela de dados filtrada posteriormente, achar a cidade de Phoenix nela e isolá-la dos outros estados agrupando ela por dia, mês e ano.

Seguindo para a próxima etapa iremos organizar essa coluna de forma ascendente mostrando da data mais antiga nas primeiras linhas para as datas mais novas nas últimas linhas indo do ano de 1997 ao de 2017.

Após toda essa coleta ser feita iremos fazer um gráfico em barras pegando um ano de exemplo, para fins de teste deste relatório iremos pegar o ano de 2017 para o experimento.

3. Desenvolvimento

Para facilitar o desenvolvimento do código e da criação dos gráficos, foram utilizadas as seguintes bibliotecas:

- Pandas: Biblioteca utilizada para fazer a carga de dados e a transformação da mesma em forma de tabela;
- Matplotlib: Biblioteca utilizada para poder fazer a plotagem de informações no mapa;

3.1 Código implementado

Segue abaixo o código desenvolvido para este desafio, levando em conta que o ano escolhido para o teste foi 2017..

```
# -*- coding: utf-8 -*-
import pandas as pd
import matplotlib.pyplot as plt

def plotar_grafic_bar_views_by_mes(ovnis, year):
    df = pd.DataFrame(ovnis['Sight_Date']).reset_index()

    df = df[df['Sight_Date'].dt.year == year]

    df = df.groupby(df['Sight_Date'].dt.month_name(),
sort=False) ['Sight_Date'].count().reset_index(name='views')

    ax = df.plot.bar(rot=0, stacked=False, x='Sight_Date',
y='views')
    ax.set_ylabel("Views")

    fig = ax.get_figure()
    fig.savefig("grafico_aparicoes_meses.jpg")

    print(df)

def plotar_grafic_line_views_by_year(ovnis):
```

```

        df = ovnis.groupby(ovnis['Sight_Date'].dt.year,
sort=False)['Sight_Date'].count().reset_index(name='views')

    df['Sight_Date'] = df['Sight_Date'].astype(str)

    ax = df.plot.line(rot=0, stacked=False, x='Sight_Date',
y='views')
    ax.set_ylabel("Views")

    fig = ax.get_figure()
    fig.savefig("grafico_aparicoes_anos.jpg")

    print(df)

def resolucao_parte_um():
    ovnis = pd.read_csv('df_OVNI_preparado.csv')

    ovnis['City'].fillna("", inplace=True)

    ovnis = ovnis[ovnis['City'].str.lower() == "phoenix"]

    ovnis[['Sight_Date']] = pd.to_datetime(ovnis['Sight_Date'],
format='%d/%m/%Y')

    ovnis.sort_values(by='Sight_Date', inplace=True)

    plotar_grafic_bar_views_by_mes(ovnis, 2017)

    plotar_grafic_line_views_by_year(ovnis)

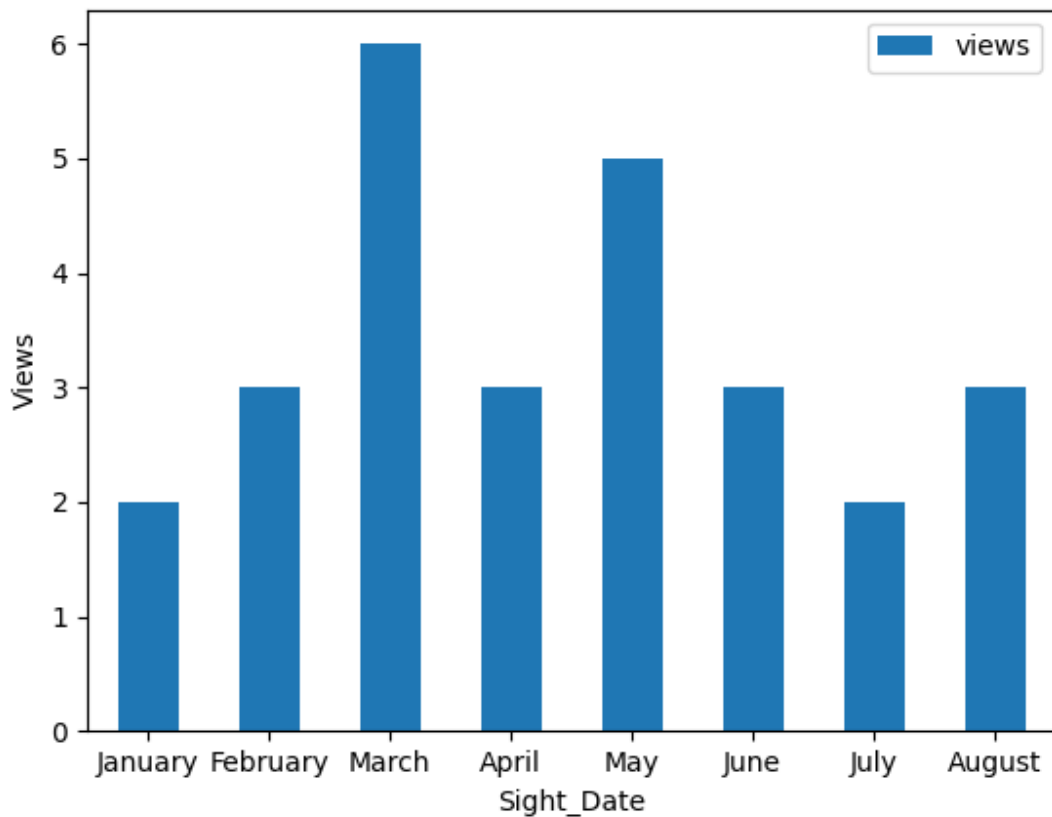
    plt.show()

def main():
    resolucao_parte_um()

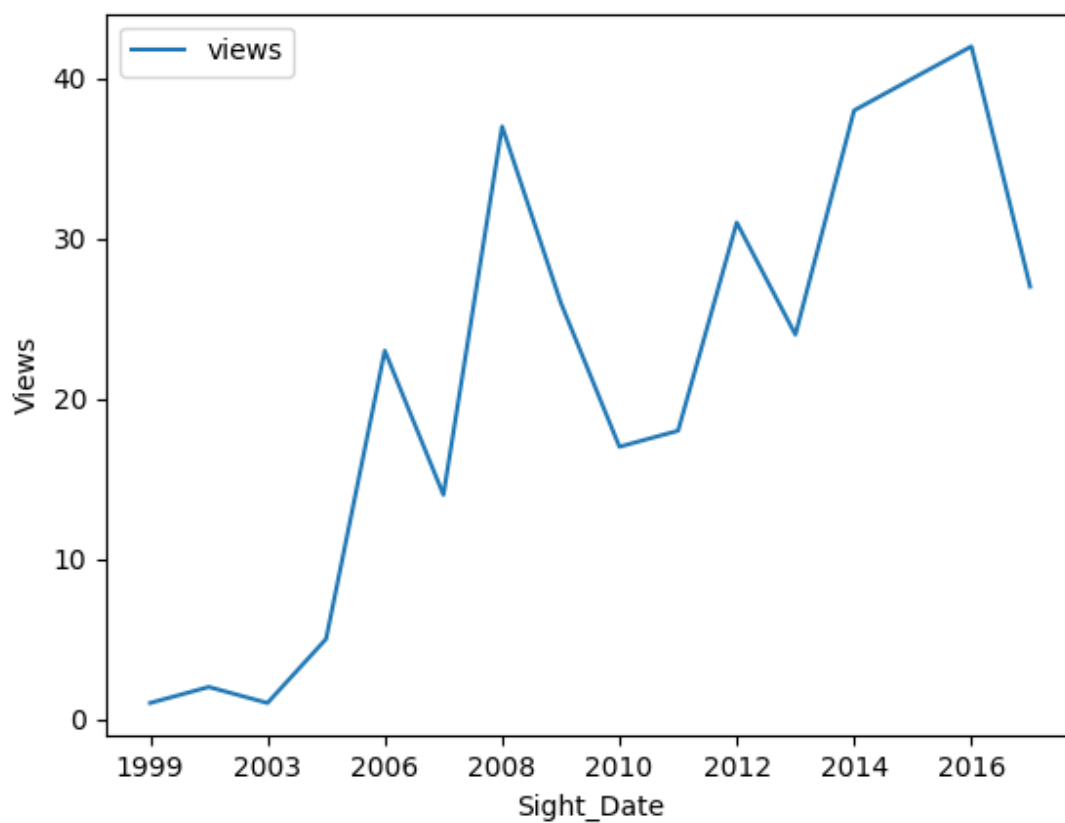
if __name__ == '__main__':
    main()

```

Como resultado do presente código ele irá gerar dois gráficos referente ao ano de 2017, o primeiro sendo um gráfico de barras.



E o segundo gráfico a ser apresentado é um gráfico em linhas.



E também a tabela das datas como essa:

	Sight_Date	views
0	January	2
1	February	3
2	March	6
3	April	3
4	May	5
5	June	3
6	July	2
7	August	3
	Sight_Date	views
0	1999	1
1	2001	2
2	2003	1
3	2005	5
4	2006	23
5	2007	14
6	2008	37
7	2009	26
8	2010	17
9	2011	18
10	2012	31
11	2013	24
12	2014	38
13	2015	40
14	2016	42
15	2017	27

4. Considerações Finais

Sobre as considerações a serem dadas neste trabalho é que por mais que as dificuldades apresentadas neste desafio já foram exploradas em trabalhos anteriores como coleta e limpeza de dados, criação de tabela e gráficos, o tempo hábil para a resolução deste em específico foi realmente curto, dos três desafios anteriores que possuíam sete dias para serem feitos este teve apenas quatro.

Porém de toda a forma não foi um trabalho impossível de ser feito, ele ilustra muito bem trabalhos que são feitos no mundo real e no mercado de trabalho como um todo, a utilização de gráficos para partes específicas de um banco de dados e até previsões com os dados já coletados é algo realmente comum como é o caso da cidade de phoenix no ano 2017 que foi mostrado.

Referências

PANDAS DOCUMENTATION. Pandas, 2021. Disponível em: <
<https://pandas.pydata.org/docs/>>. Acesso em 25/03/2021.