

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

21/03/2021

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 à terceira e penúltima sprint de atividades da PRÁTICA INTEGRADA DE CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA, sendo o trabalho apresentado aqui relativo ao segundo dos dois desafios presentes na semana.

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 terceira semana.

O presente relatório trata do desafio de **Acréscimo de variáveis**, utilizando da tabela gerada no desafio 5.7 o objetivo aqui é refinar ainda mais a pesquisa tentando achar quais os dias da semana teve mais ocorrência de OVNI's e também quais meses também.

2. Descrição do problema

O problema do desafio da semana se dá por alguns fatores, como a tabela já tinha sido limpa e gerada no desafio anterior boa parte do processo já estava feito, agora é preciso pegar a coluna Date / Time, onde estão anotados quando e em que horas aconteceu o evento e separar em duas colunas, uma em específico para a data e outra para a hora.

Fazendo isso seguiremos para a segunda parte do desafio, traduzir a data para dias da semana como segunda e terça feira, e por fim separar mais duas colunas para os meses e os dias.

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;
- Datetime: Biblioteca utilizada para fins de fuso horário;

3.1 Código implementado

Segue abaixo o código utilizado para realizar o desafio 5.8

```
# -*- coding: utf-8 -*-  
import pandas as pd  
import datetime  
  
def traduzir_dia_da_semana(dia):  
    dias = {  
        'Monday': 'Segunda-Feira',  
        'Tuesday': 'Terça-Feira',  
        'Wednesday': 'Quarta-Feira',  
        'Thursday': 'Quinta-Feira',  
        'Friday': 'Sexta-Feira',  
        'Saturday': 'Sábado',  
        'Sunday': 'Domingo'  
    }  
  
    if dia not in dias.keys():  
        return dia  
  
    return dias[dia.strip()]  
  
def main():  
    ovnis = pd.read_csv('df_OVNI_limpo.csv')  
    ovnis.columns = [col.strip().replace(' ', '').replace(r'/', '  
'') for col in ovnis.columns]
```

```

# separando date e time em duas colunas
ovnis[['Sight_Date', 'Sighth_Time']] =
ovnis['DateTime'].str.split(' ', 1, expand=True)

# remover colunas desnecessarias
ovnis.drop(['DateTime', 'row', 'Unnamed:0'], axis='columns',
inplace=True)

# converte a cluna para o tipo data
ovnis['Sight_Date'] = pd.to_datetime(ovnis['Sight_Date'])

# separa a coluna dos dias
ovnis[['Sight_Day']] = ovnis['Sight_Date'].dt.strftime("%d")

# separa a coluna dos meses
ovnis[['Sight_Month']] = ovnis['Sight_Date'].dt.strftime("%m")

# criar coluna com nome dos dias da semana
ovnis[['Sight_WeekDate']] = ovnis['Sight_Date'].dt.strftime("%A")

# traduzir os dias da semana para pt-br
ovnis['Sight_WeekDate'] = ovnis['Sight_WeekDate'].apply(lambda
dia: traduzir_dia_da_semana(dia))

# converter data para o padaro brasileiro
ovnis['Sight_Date'] =
ovnis['Sight_Date'].dt.strftime('%d/%m/%Y')

ovnis.to_csv('df_OVNI_preparado.csv', mode='a', header=True)

print(ovnis)

if __name__ == '__main__':
    main()

```

Com a execução do código após ser implementado os filtros é gerado em um arquivo csv a tabela que segue abaixo.

	State	Shape	Sight_Date	Sigth_Time	Sight_Day	Sight_Month	Sight_WeekDate
0	MD	Disk	22/09/1997	20:00	22	09	Segunda-Feira
1	CA	Rectangle	19/09/1997	None	19	09	Sexta-Feira
2	FL	Unknown	18/09/1997	20:15	18	09	Quinta-Feira
3	TX	Disk	15/09/1997	00:00	15	09	Segunda-Feira
4	NM	Light	15/09/1997	20:00	15	09	Segunda-Feira
...
61667	GA	Fireball	01/08/2017	06:15	01	08	Terça-Feira
61668	MN	Light	01/08/2017	02:45	01	08	Terça-Feira
61669	CA	Other	01/08/2017	02:00	01	08	Terça-Feira
61670	FL	Other	01/08/2017	01:00	01	08	Terça-Feira
61671	MD	Other	01/08/2017	None	01	08	Terça-Feira

[61672 rows x 7 columns]

4. Considerações Finais

As considerações para se dar ao trabalho proposto são bem simples, com a limpeza já efetuada no desafio anterior e com a grande ajuda da biblioteca `datetime`, que foi indispensável, o maior problema encontrado foi fazer o programa saber qual era aquele dia da semana com base na data, por sorte a biblioteca já vem com muitas funcionalidades que ajudaram a fazer isso.

Logo após as outras partes se tornaram mais simples como dividir a data por dia, por mês, por ano e depois fazer a tradução dos dias da semana para português brasileiro e converter a data do padrão americano para o padrão BR.

Referências

PANDAS DOCUMENTATION. Pandas, 2021. Disponível em: <
<https://pandas.pydata.org/docs/>>. Acesso em 21/03/2021.