

**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**

01/10/2020

**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>12</b>
<b>Referências</b>	<b>13</b>

# 1. Objetivos

O Objetivo desta sprint é acrescentar novas variáveis para que os dados fiquem de uma forma mais acessível para a análise, e por isso será destrinchado a coluna Date/time para que seja mais fácil a análise com o desagrupamento do dia e da hora.

## 2. Descrição do problema

Desagrupar a data da hora e depois separar o dia por mês, dia, e dia da semana.

### 3. Desenvolvimento

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

#### 3.1 Código implementado

##### 5.8 Acréscimo de variáveis

Realizando a importação da biblioteca e lendo o csv df\_OVNI\_limpo que foi gerado na fase anterior que está sendo atribuindo a variável df\_OVNI\_limpo que será o Dataframe que será trabalhado

##### 5.8 - Acréscimo de variáveis

1. Carregar o seu arquivo df\_OVNI\_limpo.csv (arquivo gerado após a limpeza de dados efetuada na atividade 5.7) em um dataframe;
2. Dividir o conteúdo da coluna Date / Time em duas novas colunas no mesmo dataframe e deletar a coluna Date / Time .
3. Fazer o mesmo procedimento para dias da semana. Será que existe um dia da semana com mais ocorrências de relatórios para OVNI's? Para descobrir isso, você deve criar uma nova coluna chamada weekdays.
4. Separar as variáveis mês (Month) e dia (Day). Desse modo, será possível refinar as pesquisas.
5. Por fim, salvar o dataframe resultante em um arquivo .csv com o nome: 'df\_OVNI\_preparado'.



```
1 #Importação da biblioteca Pandas
2 import pandas as pd
```

[5]

```
1 #Lendo o arquivo df_OVNI_limpo.csv e atribuindo a um dataframe
2 df_OVNI_limpo = pd.read_csv('df_OVNI_limpo.csv')
3 df_OVNI_limpo
```



	Date / Time	City	State	Shape
0	9/22/97 20:00	Solomons Island	MD	Disk
1	9/19/97	Garden Grove	CA	Rectangle
2	9/15/97 00:00	Houston	TX	Disk
3	9/15/97 20:00	Santa Fe	NM	Light
4	9/15/97 20:00	Kent	WA	Sphere
...	...	...	...	...

Convertendo Date / Time para o tipo datetime. Utilizando a função do pandas to\_datetime.

```
1 #Convertendo a coluna Date/time para o formato datetime utilizando a biblioteca pandas e a função to_datetime
2 df_OVNI_limpo['Date / Time'] = pd.to_datetime(df_OVNI_limpo['Date / Time'])
3 df_OVNI_limpo
```

	Date / Time	City	State	Shape
0	1997-09-22 20:00:00	Solomons Island	MD	Disk
1	1998-08-15 23:00:00	Annapolis	MD	Triangle
2	1999-10-26 20:00:00	Chesapeake Bay	MD	Disk
3	2000-07-07 01:45:00	Frederick	MD	Sphere
4	2000-10-01 20:00:00	Wheaton	MD	Triangle
...	...	...	...	...
54938	2016-03-13 05:40:00	Washington, D.C. (above I-295 bridge)	DC	Flash
54939	2016-04-12 01:30:00	Washington, D.C.	DC	Circle
54940	2016-08-08 22:00:00	Washington	DC	Triangle
54941	2016-10-02 11:00:00	Washington, D.C.	DC	Other
54942	2017-05-12 14:30:00	Washington, DC	DC	Circle

54943 rows x 4 columns

Verificando o tipos das colunas do dataframe . Utilizando o dtypes.

```
1 #Verificando os tipos das colunas
2 df_OVNI_limpo.dtypes
```

Date / Time	datetime64[ns]
City	object
State	object
Shape	object
dtype:	object

Declarando uma nova coluna chamada Sight\_day e coletando o dia a partir da coluna Date/time.

```
[47] 1 #Separando o dia do Date/time.
      2 df_OVNI_limpo['Sight_day'] =df_OVNI_limpo['Date / Time'].dt.day
      3 df_OVNI_limpo
```

	Date / Time	City	State	Shape	Sight_day
0	1997-09-22 20:00:00	Solomons Island	MD	Disk	22
1	1998-08-15 23:00:00	Annapolis	MD	Triangle	15
2	1999-10-26 20:00:00	Chesapeake Bay	MD	Disk	26
3	2000-07-07 01:45:00	Frederick	MD	Sphere	7
4	2000-10-01 20:00:00	Wheaton	MD	Triangle	1
...	...	...	...	...	...
54938	2016-03-13 05:40:00	Washington, D.C. (above I-295 bridge)	DC	Flash	13
54939	2016-04-12 01:30:00	Washington, D.C.	DC	Circle	12
54940	2016-08-08 22:00:00	Washington	DC	Triangle	8
54941	2016-10-02 11:00:00	Washington, D.C.	DC	Other	2

Declarando uma nova coluna chamada sight\_month e atribuindo o valor da coluna Date/time coletando apenas o mês da ocorrência

```
[48] 1 # Coletando o mes da coluna Date/time e colocando em uma nova coluna
      2 df_OVNI_limpo['Sight_month'] =df_OVNI_limpo['Date / Time'].dt.month
      3 df_OVNI_limpo
```

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

54943 rows x 6 columns

Declarando uma nova coluna chamada sight\_time e coletando a hora a partir da coluna date /time

```
[49] 1 df_OVNI_limpo['Sight_time'] =df_OVNI_limpo['Date / Time'].dt.time
      2 df_OVNI_limpo
```

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

54943 rows x 7 columns

Coletando a data da coluna Date/time e atribuindo este valor a uma nova coluna chamada sight\_date

```
1 df_OVNI_limpo['Sight_date'] =df_OVNI_limpo['Date / Time'].dt.date
2 df_OVNI_limpo
```

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

54943 rows x 8 columns



Coletando o dia da semana que ocorreu a ocorrência, criando uma nova coluna e atribuindo o valor de 0 a 6 indicando o dia da semana do fato, onde 0 é segunda feira, 1 é terça feira, 2 é quarta-feira, 3 é quinta-feira, 4- Sexta-feira, 5- Sábado e 6 - Domingo

```
[51] 1 df_OVNI_limpo['Sight_weekday'] = df_OVNI_limpo['Date / Time'].dt.weekday
      2 df_OVNI_limpo
```

	Date / Time	City	State	Shape	Sight_day	Sight_month	Sight_time	Sight_date	Sight_weekday
0	1997-09-22 20:00:00	Solomons Island	MD	Disk	22	9	20:00:00	1997-09-22	0
1	1998-08-15 23:00:00	Annapolis	MD	Triangle	15	8	23:00:00	1998-08-15	5
2	1999-10-26 20:00:00	Chesapeake Bay	MD	Disk	26	10	20:00:00	1999-10-26	1
3	2000-07-07 01:45:00	Frederick	MD	Sphere	7	7	01:45:00	2000-07-07	4
4	2000-10-01 20:00:00	Wheaton	MD	Triangle	1	10	20:00:00	2000-10-01	6
...	...	...	...	...	...	...	...	...	...
54938	2016-03-13 05:40:00	Washington, D.C. (above I-295 bridge)	DC	Flash	13	3	05:40:00	2016-03-13	6
54939	2016-04-12 01:30:00	Washington, D.C.	DC	Circle	12	4	01:30:00	2016-04-12	1
54940	2016-08-08 22:00:00	Washington	DC	Triangle	8	8	22:00:00	2016-08-08	0
54941	2016-10-02 11:00:00	Washington, D.C.	DC	Other	2	10	11:00:00	2016-10-02	6
54942	2017-05-12 14:30:00	Washington, DC	DC	Circle	12	5	14:30:00	2017-05-12	4

54943 rows x 9 columns

Para ficar mais fácil a interpretação do dia da semana, foi criado um dicionário atribuindo os valores do dia da semana por extenso e foi colocado na coluna de sight\_weekday utilizando a função map que percorreu os valores e atribui os valores por extenso.

```
1 dayOfWeek={0:'Segunda-feira', 1:'Terça-feira', 2:'Quarta-feira', 3:'Quinta-feira', 4:'Sexta-feira', 5:'Sábado', 6:'Domingo'}
2 df_OVNI_limpo['Sight_weekday'] = df_OVNI_limpo['Date / Time'].dt.dayofweek.map(dayOfWeek)
3 df_OVNI_limpo
```

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

54943 rows x 9 columns

E após colocado os nomes dos dias da semana, foi possível saber qual foi o dia da semana que houve mais ocorrências

```
1 df_OVNI_limp0['Sight_weekday'].value_counts()
```

Sábado	10725
Sexta-feira	8152
Domingo	7822
Quarta-feira	7405
Quinta-feira	7302
Terça-feira	6885
Segunda-feira	6652

Name: Sight\_weekday, dtype: int64

Excluindo a coluna Date/time

```
1 df_OVNI_limp0.drop(['Date / Time'],axis=1,inplace=True)
2 df_OVNI_limp0
```

	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

54943 rows x 8 columns

Utilizando a função `to_csv` para gerar um novo csv com o dataframe dos ovnis já preparados com o dia, mês, hora, data e dia da semana.

```
1 df_OVNI_limpo.to_csv('df_OVNI_preparado.csv',index=False)
2 df_OVNI_preparado= pd.read_csv('df_OVNI_preparado.csv')
3 df_OVNI_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	

54943 rows x 8 columns

## 4. Considerações Finais

Após feita a etapa de acréscimo a variáveis pode-se concluir que para futuras análises o dataframe ficou mais dividido para a coleta de informações.

## Referências

Little Bob Tables, Num day to Name day With Pandas.Stack overflow. 1º de Setembro de 2018. Disponível em <<https://stackoverflow.com/questions/29096381/num-day-to-name-day-with-pandas/>> Acesso em 30 de Setembro de 2020.