

Relatório - (Acréscimo de Variáveis)

Para esta atividade o problema é fazer uma separação das variáveis criando colunas com dados separados como (dia, mês e ano). Devido aos dados limpos da atividade “5.7 Limpeza de dados”

1ª PARTE:

Como start para o projeto foi importado as bibliotecas necessária. No nosso caso foi somente necessário importar a biblioteca pandas para a organização dos dados e o datetime para converter os argumentos de data e hora.

(* O DATETIME SERÁ USADO MAIS PARA FRENTE).

```
1 import pandas as pd
2 import datetime
```

Em seguida vamos carregar o DataFrame, nomeado por “df_OVNI_limpo.csv”

```
1 #1 Carregando DataFrame limpo
2
3 df = pd.read_csv('df_OVNI_limpo.csv')
```

Saída:

	Unnamed: 0	Date / Time	City	State	Shape
0	0	1/29/97 23:15	East Greenwich	RI	Disk
1	1	1/26/97 22:00	Flagstaff	AZ	Light
2	2	1/25/97 21:00	Marion	WI	Triangle
3	4	1/24/97 19:00	Alta	UT	Other
4	5	1/23/97 18:30	North Kingstown	RI	Triangle
...
80041	102385	12/1/17 17:00	New Rochelle	NY	Sphere
80042	102386	12/1/17 17:00	Foyil	OK	Formation
80043	102387	12/1/17 04:00	Chesapeake	VA	Light
80044	102388	12/1/17 04:00	Boise	ID	Cigar
80045	102390	12/1/17 01:00	Wasilla	AK	Flash

2ª PARTE:

Em seguida estou selecionando a coluna 'Date / Time', acessando a função 'str' e Split para separar a coluna, com a separação realizada renomeio as coluna 'Sight_Date', 'Sight_Time' com a Data e Hora devidamente separadas.

Nas linhas de códigos seguintes foi realizado uma exclusão da coluna "Date / Time"

```
1 #2 Separar a coluna 'data' e 'hora' do mesmo DataFrame
2
3 # Dividir o conteúdo da coluna Date / Time em duas novas colunas no mesmo dataframe e deletar a coluna Date / Time .
4
5 df1 = df.join(df['Date / Time'].str.split(' ', 1, expand=True).rename(columns={0:'Sight_Date', 1:'Sight_Time'}))
6 df1
7 # Excluindo a coluna [Date / Time]
8 df1.drop(['Date / Time'], inplace=True, axis= 1)
9 df1
```

Vamos ver como ficou? Saída:

	Unnamed: 0	City	State	Shape	Sight_Date	Sight_Time
0	0	East Greenwich	RI	Disk	1/29/97	23:15
1	1	Flagstaff	AZ	Light	1/26/97	22:00
2	2	Marion	WI	Triangle	1/25/97	21:00
3	4	Alta	UT	Other	1/24/97	19:00
4	5	North Kingstown	RI	Triangle	1/23/97	18:30
...
80041	102385	New Rochelle	NY	Sphere	12/1/17	17:00
80042	102386	Foyil	OK	Formation	12/1/17	17:00
80043	102387	Chesapeake	VA	Light	12/1/17	04:00
80044	102388	Boise	ID	Cigar	12/1/17	04:00
80045	102390	Wasilla	AK	Flash	12/1/17	01:00

SUPIMPA deu certo.

3ª PARTE:

Até agora todas as nossas tentativas de fazer a separação dos dados estão dando certo.

Nessa terceira parte foi criada uma nova coluna chamada `sight_weekday` Com os dias da semana, e aqui usamos a propriedade do DateTime para a criação da coluna com os dias da semana

```

1 #3 Separando dias e colunas
2 df1['sight_weekday'] = pd.to_datetime(df1['Sight_Date'])
3 df1['sight_weekday'] = df1['sight_weekday'].dt.weekday_name
4
5 weekday_ptbr = {
6     'Monday': 'Segunda-feira',
7     'Tuesday': 'Terça-feira',
8     'Wednesday': 'Quarta-feira',
9     'Thursday': 'Quinta-feira',
10    'Friday': 'Sexta-feira',
11    'Saturday': 'Sábado',
12    'Sunday': 'Domingo'
13 }
14
15 df1 = df1.replace({'sight_weekday': weekday_ptbr})
16
17 df1.head()

```

SAÍDA:

	Unnamed: 0	City	State	Shape	Sight_Date	Sight_Time	sight_weekday
0	0	East Greenwich	RI	Disk	1/29/97	23:15	Quarta-feira
1	1	Flagstaff	AZ	Light	1/26/97	22:00	Domingo
2	2	Marion	WI	Triangle	1/25/97	21:00	Sábado
3	4	Alta	UT	Other	1/24/97	19:00	Sexta-feira
4	5	North Kingstown	RI	Triangle	1/23/97	18:30	Quinta-feira

Observe que agora temos uma separação nítida da Data, Hora e os Dias da Semana.

4ª PARTE:

Para termos uma visibilidade melhor, vamos criar duas novas colunas especificando elas como DIA e MÊS.

```

1 #4 Separação de variáveis mes e dia
2
3 df1['sight_Day']=pd.to_datetime(df1['Sight_Date']).dt.day
4 df1['sight_Month']=pd.to_datetime(df1['Sight_Date']).dt.month

```

Vamos visualizar apertando os comandos CTRL + SHIFT + ENTER para rodar do DataFrame.

```
6 df1
```

SAÍDA:

	Unnamed: 0	City	State	Shape	Sight_Date	Sight_Time	sight_weekday	sight_Day	sight_Month
0	0	East Greenwich	RI	Disk	1/29/97	23:15	Quarta-feira	29	1
1	1	Flagstaff	AZ	Light	1/26/97	22:00	Domingo	26	1
2	2	Marion	WI	Triangle	1/25/97	21:00	Sábado	25	1
3	4	Alta	UT	Other	1/24/97	19:00	Sexta-feira	24	1
4	5	North Kingstown	RI	Triangle	1/23/97	18:30	Quinta-feira	23	1
...
80041	102385	New Rochelle	NY	Sphere	12/1/17	17:00	Sexta-feira	1	12
80042	102386	Foyil	OK	Formation	12/1/17	17:00	Sexta-feira	1	12
80043	102387	Chesapeake	VA	Light	12/1/17	04:00	Sexta-feira	1	12
80044	102388	Boise	ID	Cigar	12/1/17	04:00	Sexta-feira	1	12
80045	102390	Wasilla	AK	Flash	12/1/17	01:00	Sexta-feira	1	12

80046 rows x 9 columns


Observe que foi criada duas novas colunas uma armazenando o DIA do mês e outra armazenando o MÊS.

5ª PARTE:

Para finalizar este modulo, vamos criar um novo “ DataFrame – df ” com as alterações feitas no “ df ”

```
1 #5 Criando novo arquivo (.csv) com as alterações feitas
2 df1.to_csv('df_OVNI_preparado.csv')
```

Vamos ver se foi criado a DataFrame?

 df_OVNI_limpo.csv

Pronto foi criado com Sucesso!!!

...

Pronto! Ai está o DataFrame final com a limpeza realizada e os Dados devidamente separados!!!

LINK do projeto: <https://github.com/samuelflopes/DataScience.git>

FIM