

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

1. Objetivos	3
2. Descrição do problema	4
3. Desenvolvimento3.1 Código implementado	5
4. Considerações Finais	12
Referências	13

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.

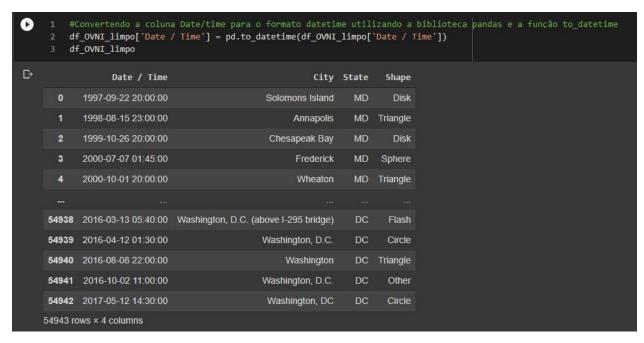
3.1 Código implementado

5.8 Acréscimo de variáveis

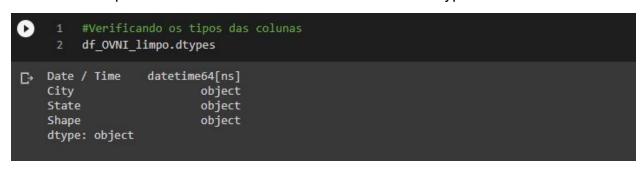
Realizando a importação da biblioteca e lendo o csv df_OVNI_limpo que foi gerado na fase anterior que está sendo atribuindo a variável df_OVNI_limpo que será o Dataframe que será trabalhado



Convertendo Date / Time para o tipo datetime. Utilizando a função do pandas to_datetime.



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



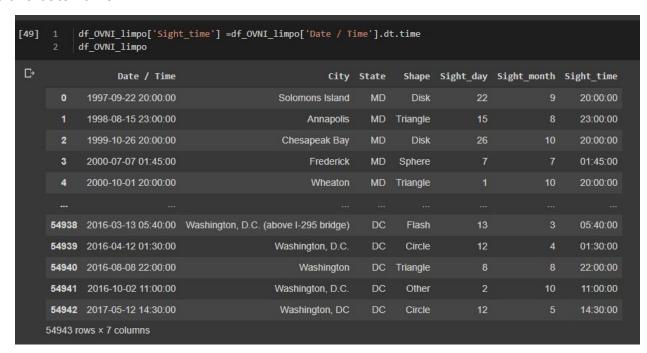
Declarando uma nova coluna chamada Sight_day e coletando o dia a partir da coluna Date/time.

[47]	2 d	2 df_OVNI_limpo['Sight_day'] =df_OVNI_limpo['Date / Time'].dt.day								
C÷		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	Chesapeak 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				

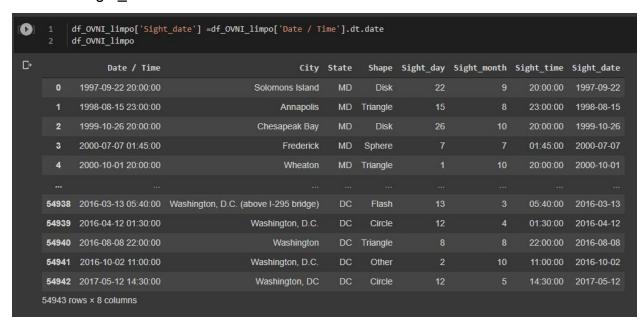
Declarando uma nova coluna chamada sight_month e atribuindo o valor da coluna Date/time coletando apenas o mês da ocorrência

[48]	1 2 3	<pre># Coletando o mes da coluna Date/time e colocando em uma nova coluna df_OVNI_limpo['Sight_month'] =df_OVNI_limpo['Date / Time'].dt.month df_OVNI_limpo</pre>								
C →		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	Chesapeak 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			
	5493	8 2016-03-13 05:40:00	Washington, D.C. (above I-295 bridge)	DC	Flash	13	3			
	5493	9 2016-04-12 01:30:00	Washington, D.C.	DC	Circle	12	4			
	5494	0 2016-08-08 22:00:00	Washington	DC	Triangle	8	8			
	5494	1 2016-10-02 11:00:00	Washington, D.C.	DC	Other	2	10			
	5494	2 2017-05-12 14:30:00	Washington, DC	DC	Circle	12	5			
	54943 rows × 6 columns									

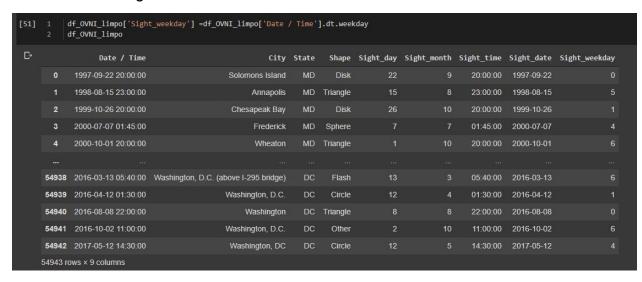
Declarando uma nova coluna chamada sight_time e coletando a hora a partir da coluna date /time



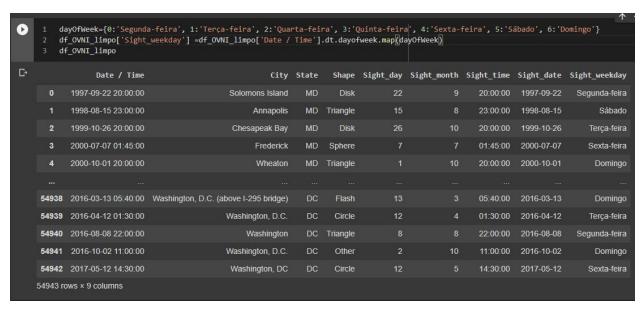
Coletando a data da coluna Date/time e atribundo este valor a uma nova coluna chamada sight_date



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



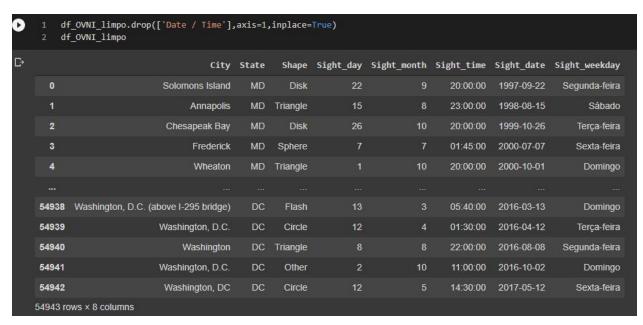
Para ficar mais fácil a interpretação do dia da semana, foi criado um dicionário atribuindo os valores do dia da semana por extenso e foi colocado na coluna de sight_weekday utilizando a função map que percorreu os valores e atribui os valores por extenso.



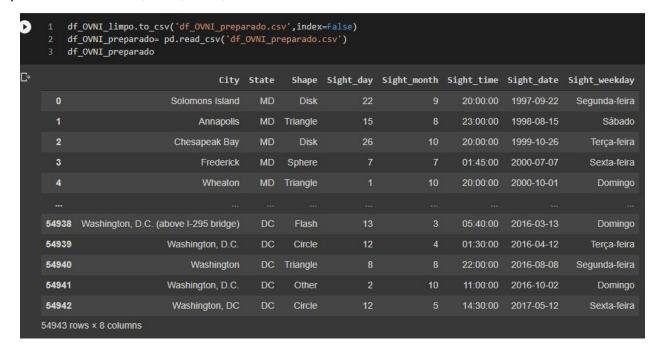
E após colocado os nomes dos dias da semana, foi possível saber qual foi o dia da semana que houve mais ocorrências



Excluindo a coluna Date/time



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.



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. 1° 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.