

Pandas: Agrupamento e Ordenação

Professor: **Nator Junior Carvalho da Costa**

October 8, 2022

- 1 Revisão apply e map;
- 2 Atividade Prática.
- 3 Função **groupby**;
- 4 Atividade Prática;
- 5 Ordenação;



- A função **apply** serve pra aplicar mudanças em Series e DataFrames

```
1 import pandas as pd
2 df = pd.DataFrame({'X': [1, 2, 3,],
3                    'Y': [4, 1, 8]})
4 modified_df=df.apply(sum)
5 -> output
6 X      6
7 Y     13
8 dtype: int64
```



```
1 import pandas as pd
2 df = pd.DataFrame({'X': [1, 2, 3],
3                     'Y': [4, 1, 8]})
4 modified_df=df.apply(sum,axis=1)
5 -> output
6 0      5
7 1      3
8 2     11
9 dtype: int64
```



```
1 import pandas as pd
2 def soma(x):
3     return x.iloc[0]+x.iloc[1]
4 df = pd.DataFrame({'X': [1, 2, 3,],
5                     'Y': [4, 1, 8]})
6 modified_df=df.X.apply(soma,axis=1)
7 -> output
8 0      5
9 1      3
10 2     11
11 dtype: int64
```



```
1 import pandas as pd
2 def elevQuad(x):
3     return pow(x,2)
4 df = pd.DataFrame({'X': [1, 2, 3,],
5                     'Y': [4, 1, 8]})
6 modified_df=df.X.map(elevQuad)
7 -> output
8 0    1
9 1    4
10 2    9
11 Name: X, dtype: int64
```



Para essa atividade considere a base de dados a seguir:

https://raw.githubusercontent.com/natorjunior/pandas/main/Aula-02/ibge_populacao.csv

- 1 Descobrir as 10 maiores cidades do Ceará no ano de 2019;
- 2 Descobrir as 10 maiores cidades do Piauí;
- 3 Plotar a curva de crescimento populacional de Fortaleza;
- 4 Plotar a curva de crescimento populacional de Sobral;
- 5 Substituir os códigos (2304400 ← Fortaleza, 2312908 ← Sobral) do IBGE da coluna **id_municipio** pelo nome usando **apply** ou **map**;

obs: Para plotar as curvas de crescimento populacional use a LIB **plotly**



- `sum();`
- `max();`
- `min();`
- `count();`
- `get_group();`
- `mean();`
- `describe();`



```
1 import pandas as pd
2 df = pd.DataFrame({'ano': [2010, 2010, 2011, 2012, 2012],
3                       'idade': [10, 16, 20, 25, 31]})
4 df.groupby('ano').mean()
```

Out[78]:

idade	
ano	
2010	13
2011	20
2012	28



```
1 import pandas as pd
2 df = pd.DataFrame({'ano': [2010, 2010, 2011, 2012, 2012],
3                       'idade': [10, 16, 20, 25, 31]})
4 df.groupby('ano').max()
```

Out[79]:

idade	
ano	
2010	16
2011	20
2012	31



```
1 import pandas as pd
2 df = pd.DataFrame({'ano': [2010, 2010, 2011, 2012, 2012],
3                       'idade': [10, 16, 20, 25, 31]})
4 df.groupby('ano').min()
```

Out[88]:

idade	
ano	
2010	10
2011	20
2012	25



```
1 import pandas as pd
2 df = pd.DataFrame({'ano': [2010, 2010, 2011, 2012, 2012],
3                       'idade': [10, 16, 20, 25, 31]})
4 df.groupby('ano').describe()
```

Out[80]:

	idade							
	count	mean	std	min	25%	50%	75%	max
ano								
2010	2.0	13.0	4.242641	10.0	11.5	13.0	14.5	16.0
2011	1.0	20.0	NaN	20.0	20.0	20.0	20.0	20.0
2012	2.0	28.0	4.242641	25.0	26.5	28.0	29.5	31.0

```
1 import pandas as pd
2 df = pd.DataFrame({
3     'col1': ['g', 't', 'n', 'w', 'n', 'g'],
4     'col2': [5, 2, 5, 1, 3, 6],
5     'col3': [0, 7, 2, 8, 1, 2],
6 })
7 print(df.sort_values(by=['col1']))
```

Out[85]:

	col1	col2	col3
0	g	5	0
5	g	6	2
2	n	5	2
4	n	3	1
1	t	2	7

```
1 import pandas as pd
2 df = pd.DataFrame({
3     'col1': ['g', 't', 'n', 'w', 'n', 'g'],
4     'col2': [5, 2, 5, 1, 3, 6],
5     'col3': [0, 7, 2, 8, 1, 2],
6 })
7 print(df.sort_values(by=['col1', 'col2']))
```

Out[92]:

	col1	col2	col3
0	g	5	0
5	g	6	2
4	n	3	1
2	n	5	2
1	t	2	7

```
1 import pandas as pd
2 df = pd.DataFrame({
3     'col1': ['g', 't', 'n', 'w', 'n', 'g'],
4     'col2': [5, 2, 5, 1, 3, 6],
5     'col3': [0, 7, 2, 8, 1, 2],
6 })
7 df.sort_values(by=['col1', 'col2'], ascending=False)
8
```

Out[93]:

	col1	col2	col3
3	w	1	8
1	t	2	7
2	n	5	2
4	n	3	1



Para essa atividade considere a base de dados a seguir:

https://raw.githubusercontent.com/natorjunior/pandas/main/Aula-02/microdados_violencia_reduzido.csv

- 1 Separar apenas os dados da UF 23;
- 2 Quantificar ocorrências separando por sexo;
- 3 Descobrir a média de idade das ocorrências por meio de arma de fogo;
- 4 Plotar as ocorrências por arma de fogo por ano (Gráfico de barras);

obs: Para plotar os gráficos use a LIB **plotly**

Obrigado!

Pandas: Agrupamento e Ordenação

Professor: **Nator Junior Carvalho da Costa**

October 8, 2022