Treino funcionarios

September 24, 2023

[1]:
```
import pandas as pd
import numpy as np
```

[2]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\55219\anaconda3\lib\site-
packages (3.7.0)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: packaging>=20.0 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: numpy>=1.20 in c:\users\55219\anaconda3\lib\site-
packages (from matplotlib) (1.23.5)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\55219\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: six>=1.5 in c:\users\55219\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
import matplotlib.pyplot as plt
```

[4]:
```
df = pd.read_excel('Base_funcionarios_treino.xlsx')
```

[5]:
```
df.head()
```

[5]:

	ID_Funcionario	Funcionarios	Salarios	Sexo	Idade
0	746	Cleber	9824	M	48
1	341	Junior	3276	M	21

```
2            1742     Matheus       4162    M     34
3             307       Miguel      2737    M     38
4             418       Arthur      8225    M     49
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19 entries, 0 to 18
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ID_Funcionario  19 non-null     int64
 1   Funcionarios    19 non-null     object
 2   Salarios        19 non-null     int64
 3   Sexo            19 non-null     object
 4   Idade           19 non-null     int64
dtypes: int64(3), object(2)
memory usage: 888.0+ bytes
```

[7]:
```
df.describe()
```

[7]:

	ID_Funcionario	Salarios	Idade
count	19.000000	19.000000	19.000000
mean	747.631579	5709.894737	38.315789
std	471.210759	2609.997060	13.618014
min	173.000000	2737.000000	21.000000
25%	379.500000	3413.500000	25.000000
50%	669.000000	5451.000000	34.000000
75%	959.000000	8026.000000	48.500000
max	1742.000000	9927.000000	61.000000

[8]:
```
idade_salarios = df[['Idade','Salarios']]
display(idade_salarios)
```

	Idade	Salarios
0	48	9824
1	21	3276
2	34	4162
3	38	2737
4	49	8225
5	32	6239
6	58	3551
7	25	8378
8	53	9926
9	34	5841
10	30	6365
11	24	4205
12	21	7827

```
13      61      5451
14      25      3948
15      48      2875
16      24      9927
17      46      2808
18      57      2923
```

[9]:
```
df_sexo = df.groupby('Sexo').mean()
df_sexo = df_sexo[['Salarios','Idade']]
display(df_sexo)
```

```
C:\Users\55219\AppData\Local\Temp\ipykernel_15132\3792877398.py:1:
FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is
deprecated. In a future version, numeric_only will default to False. Either
specify numeric_only or select only columns which should be valid for the
function.
  df_sexo = df.groupby('Sexo').mean()
```

Sexo	Salarios	Idade
F	4655.333333	43.500000
M	6196.615385	35.923077

[10]:
```
df2 = df.rename(columns={'Idade':'Age','Salarios':'Salary'})
```

[11]:
```
display(df2)
```

	ID_Funcionario	Funcionarios	Salary	Sexo	Age
0	746	Cleber	9824	M	48
1	341	Junior	3276	M	21
2	1742	Matheus	4162	M	34
3	307	Miguel	2737	M	38
4	418	Arthur	8225	M	49
5	1349	Gael	6239	M	32
6	315	Théo	3551	M	58
7	1604	Heitor	8378	M	25
8	598	Ravi	9926	M	53
9	674	Davi	5841	M	34
10	849	Bernardo	6365	M	30
11	220	Noah	4205	M	24
12	1069	Gabriel	7827	M	21
13	173	Helena	5451	F	61
14	1355	Alice	3948	F	25
15	570	Laura	2875	F	48
16	775	Maria Alice	9927	F	24
17	669	Sophia	2808	F	46
18	431	Manuela	2923	F	57

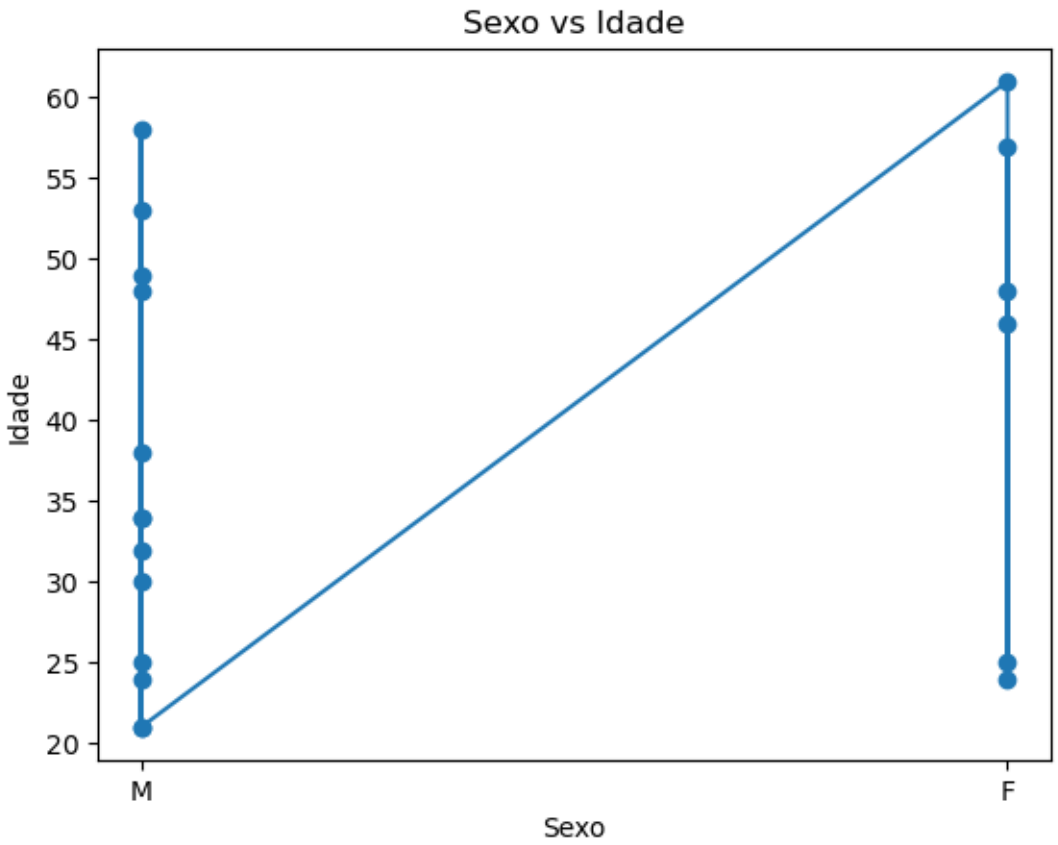
[12]:
```
display(df2['Age'])
```

```
0     48
1     21
2     34
3     38
4     49
5     32
6     58
7     25
8     53
9     34
10    30
11    24
12    21
13    61
14    25
15    48
16    24
17    46
18    57
Name: Age, dtype: int64
```

[13]:
```
import plotly.express as px

grafico = px.bar(df_sexo, x=df_sexo.index, y='Salarios')  #usar o index pq a␣
 ↪coluna sexo virou um índice depois do groupby. Se não tivesse o grupoby era␣
 ↪só passar o nome da coluna
grafico.show()
```

[14]:
```
plt.title('Sexo vs Idade')
plt.xlabel('Sexo')
plt.ylabel('Idade')
plt.plot(df['Sexo'],df['Idade'], marker ='o')
plt.show()
```


Sexo vs Idade
Idade
60
55
50
45
40
35
30
25
20
M
F
Sexo