Lab 12/09

Gabriel Farias Ramalho
TADS020

```
[1]  import pandas as pd
     import numpy as np
     ✓ 2.1s                                                    Python
```

```
[2]  meusDados = np.matrix('1, 2, 3; 4, 5, 6; 7, 8, 9; 10, 11, 12')
     meusDados
     ✓ 0.0s                                                    Python
```

```
matrix([[ 1,  2,  3],
        [ 4,  5,  6],
        [ 7,  8,  9],
        [10, 11, 12]])
```

```
[3]  meuDf = pd.DataFrame(meusDados)
     type(meuDf)
     ✓ 0.0s                                                    Python
```

```
pandas.core.frame.DataFrame
```

```
[4]  indice = ['Linha' + str(i) for i in range(4)]
     coluna = ['coluna'+ str(i) for i in range(3)]
     ✓ 0.0s                                                    Python
```

```
[5]  df2 = pd.DataFrame(data = meusDados, index = indice, columns = coluna)
     df2
     ✓ 0.0s                                                    Python
```

	coluna0	coluna1	coluna2
Linha0	1	2	3
Linha1	4	5	6
Linha2	7	8	9
Linha3	10	11	12

```
df2 = pd.DataFrame (meusDados, indice, coluna)
df2
```

[6] ✓ 0.0s Python

	coluna0	coluna1	coluna2
Linha0	1	2	3
Linha1	4	5	6
Linha2	7	8	9
Linha3	10	11	12

```
df3 = {
    'colunas0' : {'linha0' : 1, 'linhas1' : 4, 'linhas2' : 7},
    'colunas1' : {'linha0' : 2, 'linhas1' : 5, 'linhas2' : 8},
    'colunas2' : {'linha0' : 3, 'linhas1' : 6, 'linhas2' : 9},
}
df3
```

[7] ✓ 0.0s Python

```
{'colunas0': {'linha0': 1, 'linhas1': 4, 'linhas2': 7},
 'colunas1': {'linha0': 2, 'linhas1': 5, 'linhas2': 8},
 'colunas2': {'linha0': 3, 'linhas1': 6, 'linhas2': 9}}
```

```
df3 = pd.DataFrame(df3)
df3
```

[8] ✓ 0.0s Python

	colunas0	colunas1	colunas2
linha0	1	2	3
linhas1	4	5	6
linhas2	7	8	9

```
df4 = pd.concat([meuDf, df2, df3])
df4
```

[9] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	NaN	NaN	NaN	NaN	NaN	NaN
1	4.0	5.0	6.0	NaN	NaN	NaN	NaN	NaN	NaN
2	7.0	8.0	9.0	NaN	NaN	NaN	NaN	NaN	NaN
3	10.0	11.0	12.0	NaN	NaN	NaN	NaN	NaN	NaN
Linha0	NaN	NaN	NaN	1.0	2.0	3.0	NaN	NaN	NaN
Linha1	NaN	NaN	NaN	4.0	5.0	6.0	NaN	NaN	NaN
Linha2	NaN	NaN	NaN	7.0	8.0	9.0	NaN	NaN	NaN
Linha3	NaN	NaN	NaN	10.0	11.0	12.0	NaN	NaN	NaN
linha0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	2.0	3.0
linhas1	NaN	NaN	NaN	NaN	NaN	NaN	4.0	5.0	6.0
linhas2	NaN	NaN	NaN	NaN	NaN	NaN	7.0	8.0	9.0

```
[10]  df4.isnull()
      ✓ 0.0s                                                                    Python
```

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	False	False	False	True	True	True	True	True	True
1	False	False	False	True	True	True	True	True	True
2	False	False	False	True	True	True	True	True	True
3	False	False	False	True	True	True	True	True	True
Linha0	True	True	True	False	False	False	True	True	True
Linha1	True	True	True	False	False	False	True	True	True
Linha2	True	True	True	False	False	False	True	True	True
Linha3	True	True	True	False	False	False	True	True	True
linha0	True	True	True	True	True	True	False	False	False
linhas1	True	True	True	True	True	True	False	False	False
linhas2	True	True	True	True	True	True	False	False	False

```
[11]  df4.isnull().sum()
      ✓ 0.0s                                                                    Python
```

```
0           7
1           7
2           7
coluna0     7
coluna1     7
coluna2     7
colunas0    8
colunas1    8
colunas2    8
dtype: int64
```

```
[12]  df4.isnull().sum().sum()
      ✓ 0.0s                                                                    Python
```

```
66
```

```
df4.fillna
```

[13] ✓ 0.0s Python

```
<bound method DataFrame.fillna of            0     1     2  coluna0  coluna1  coluna2  colunas0  colunas1  \
0          1.0   2.0   3.0      NaN      NaN      NaN       NaN       NaN
1          4.0   5.0   6.0      NaN      NaN      NaN       NaN       NaN
2          7.0   8.0   9.0      NaN      NaN      NaN       NaN       NaN
3         10.0  11.0  12.0      NaN      NaN      NaN       NaN       NaN
Linha0     NaN   NaN   NaN      1.0      2.0      3.0       NaN       NaN
Linha1     NaN   NaN   NaN      4.0      5.0      6.0       NaN       NaN
Linha2     NaN   NaN   NaN      7.0      8.0      9.0       NaN       NaN
Linha3     NaN   NaN   NaN     10.0     11.0     12.0       NaN       NaN
linha0     NaN   NaN   NaN      NaN      NaN      NaN       1.0       2.0
linhas1    NaN   NaN   NaN      NaN      NaN      NaN       4.0       5.0
linhas2    NaN   NaN   NaN      NaN      NaN      NaN       7.0       8.0

          colunas2
0              NaN
1              NaN
2              NaN
3              NaN
Linha0         NaN
Linha1         NaN
Linha2         NaN
Linha3         NaN
linha0         3.0
linhas1        6.0
linhas2        9.0  >
```

```
df5 = df4.fillna(df4.mean())
df5
```

[14] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df8 = df5.copy()
```

[15] ✓ 0.0s Python

```
df4.dropna(inplace = True)
df4
```

[16] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2

```
df5
```

[17] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df5.dropna(subset = [1], inplace = True)
df5
```

[18] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df6 = df5.drop(columns = ['coluna0'])
df6
```

[19] ✓ 0.0s Python

	0	1	2	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	6.5	7.5	7.0	8.0	9.0

```
df7 = df6.drop(index = [2])
df7
```

[20] ✓ 0.0s Python

	0	1	2	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	6.5	7.5	7.0	8.0	9.0

```
df5.coluna2.unique()
```

[21] ✓ 0.0s Python

```
array([ 7.5,  3. ,  6. ,  9. , 12. ])
```

```
df5.columns.unique()
```

[22] ✓ 0.0s Python

```
Index([0, 1, 2, 'coluna0', 'coluna1', 'coluna2', 'colunas0', 'colunas1',
       'colunas2'],
      dtype='object')
```

```
df5.index.unique()
```

[23] ✓ 0.0s Python

```
Index([       0,        1,        2,        3,  'Linha0',  'Linha1',
       'Linha2',  'Linha3',  'linha0', 'linhas1', 'linhas2'],
      dtype='object')
```

```
df8.sort_values(by = 'coluna1')
```

[24] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0

```
df8.rename(columns = {'Coluna1':'COLUNA1'}, inplace = True)
df8
```

[25] ✓ 0.0s Python

	0	1	2	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df8.rename(columns = {0:'janeiro', 1:'fevereiro', 2:'março',
                      'Coluna 0': 'abril', 'Coluna1':'maio', 'Coluna2':'junho',
                      'Coluna0':'julho', 'COLUNA1':'agosto', 'Coluna2':'setembro'}, inplace = True)
df8
```

[26] ✓ 0.0s Python

	janeiro	fevereiro	março	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df8.dropna(subset = ['fevereiro'], inplace = True)
df8
```

[27] ✓ 0.0s Python

	janeiro	fevereiro	março	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df9 = df8.rename(index = {0:'Elon Musk', 1:'Bill Gates',
                          2:'Bernard Arnault', 3:'Mark Zuckerberg'})
df9
```

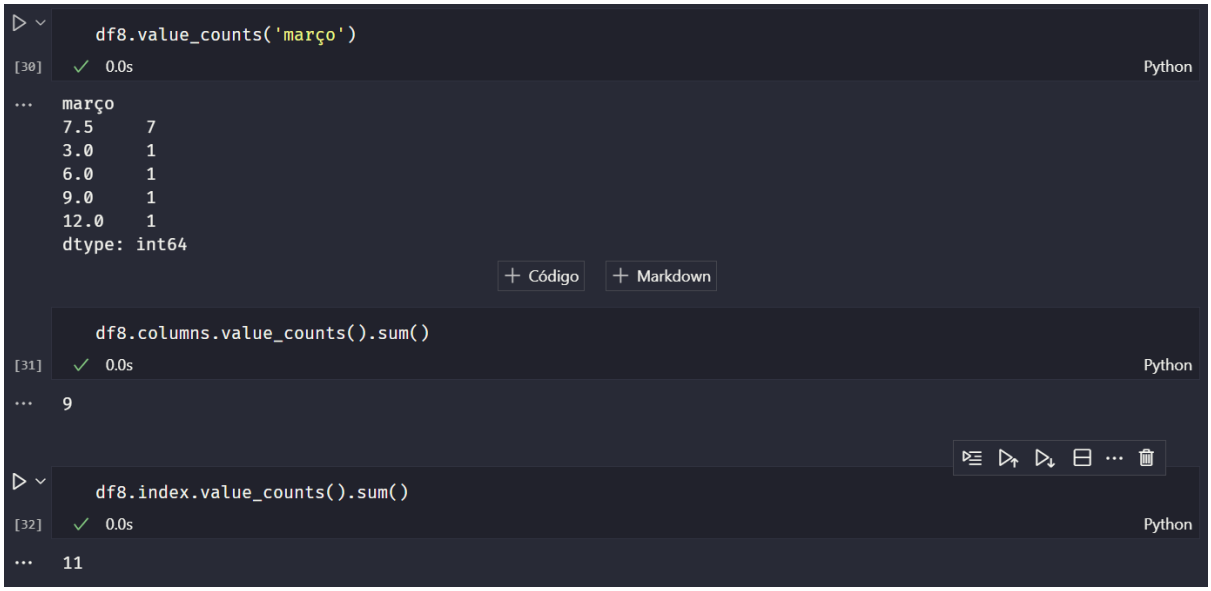
[28] ✓ 0.0s Python

	janeiro	fevereiro	março	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
Elon Musk	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
Bill Gates	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
Bernard Arnault	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
Mark Zuckerberg	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0

```
df8.reset_index()
```

[29] ✓ 0.0s Python

	index	janeiro	fevereiro	março	coluna0	coluna1	coluna2	colunas0	colunas1	colunas2
0	0	1.0	2.0	3.0	5.5	6.5	7.5	4.0	5.0	6.0
1	1	4.0	5.0	6.0	5.5	6.5	7.5	4.0	5.0	6.0
2	2	7.0	8.0	9.0	5.5	6.5	7.5	4.0	5.0	6.0
3	3	10.0	11.0	12.0	5.5	6.5	7.5	4.0	5.0	6.0
4	Linha0	5.5	6.5	7.5	1.0	2.0	3.0	4.0	5.0	6.0
5	Linha1	5.5	6.5	7.5	4.0	5.0	6.0	4.0	5.0	6.0
6	Linha2	5.5	6.5	7.5	7.0	8.0	9.0	4.0	5.0	6.0
7	Linha3	5.5	6.5	7.5	10.0	11.0	12.0	4.0	5.0	6.0
8	linha0	5.5	6.5	7.5	5.5	6.5	7.5	1.0	2.0	3.0
9	linhas1	5.5	6.5	7.5	5.5	6.5	7.5	4.0	5.0	6.0
10	linhas2	5.5	6.5	7.5	5.5	6.5	7.5	7.0	8.0	9.0


df8.value_counts('março')
[30]
0.0s
Python
março
7.5 7
3.0 1
6.0 1
9.0 1
12.0 1
dtype: int64
+ Código
+ Markdown
df8.columns.value_counts().sum()
[31]
0.0s
Python
9
df8.index.value_counts().sum()
[32]
0.0s
Python
11