


Python - Pandas



Pandas: trabalhando com Dataframes/Series

O método Describe: exibe uma estatística descritiva do dataframe ou de colunas específicas.

Ex: `df.describe()`



	id	quantidade
count	549.000000	549.000000
mean	274.000000	6.045537
std	158.626921	3.201238
min	0.000000	1.000000
25%	137.000000	3.000000
50%	274.000000	6.000000
75%	411.000000	9.000000
max	548.000000	11.000000

Nota-se que a função *describe*, por padrão, escolhe apenas os dados numéricos,

Pandas: trabalhando com Dataframes/Series

Uma coluna em um dataframe pode ser obtida como uma Series, seja usando uma notação do tipo dicionário ou por meio de atributo. Ex: `df['produto']` ou `df.produto`

```
0      toalha
1      toalha
2      toalha
3      toalha
4      toalha
...
544  quebra_cabeca
545  quebra_cabeca
546  quebra_cabeca
547  quebra_cabeca
548  quebra_cabeca
Name: produto, Length: 549, dtype: object
```

```
type(df['produto'])
```

```
pandas.core.series.Series
```

Pandas: trabalhando com Dataframes/Series

Se quisermos, também é possível selecionar uma única coluna e obter como retorno um dataframe. Para isso, ao invés de passarmos um único par de colchetes, passaremos dois.

```
df[['produto']]
```



	produto
0	toalha
1	toalha
2	toalha
3	toalha
4	toalha
...	...
544	quebra_cabeca
545	quebra_cabeca
546	quebra_cabeca
547	quebra_cabeca
548	quebra_cabeca

549 rows × 1 columns

Pandas: trabalhando com Dataframes/Series

A seleção de linhas mais simples, que já conhecemos das listas, tuplas e arrays Numpy, funciona basicamente da mesma forma. Sintaxe:

`dataset[i:j]`

onde i representa a linha inicial e j representa a j-1 linha do dataset.

```
df[0:3]
```

	id	data_aq	produto	quantidade	valor UN	Total	setor
0	0	01/01/2019	toalha	6	R\$ 35,00	R\$ 210,00	Mesa_banho
1	1	02/01/2019	toalha	6	R\$ 35,00	R\$ 210,00	NaN
2	2	03/01/2019	toalha	2	R\$ 35,00	R\$ 70,00	mesa_banho

Pandas: trabalhando com Dataframes/Series

Existem outras formas mais elegantes de fazermos seleções, conhecidas como **.loc** e **.iloc**

O **.loc** seleciona um grupo de linhas e colunas segundo os rótulos ou uma matriz booleana. Para selecionarmos um rótulo, simplesmente o passaremos entre colchetes. Ex:

```
df.loc[[rótulos de indice],[colunas]]
```