

Pandas

1 - Como apagar alguns registros em um dataframe?

Considerando que queremos eliminar os registros que tenham valores negativos

```
import numpy as np
import pandas as pd
column = ['a', 'b']
df = pd.DataFrame(np.random.randint(-3, 5, (30,2)),
                  columns = column)
```

Opção 1

```
df[(df['a'] >= 0) & (df['b'] >= 0)]
```

Opção 2

```
df[(df['a'] >= 0) and (df['b'] >= 0)]
```

Recebemos dados novos e precisamos organizar alguns problemas já detectados.

2 – Corrija o nome dos municípios para ficar apenas os nomes. Sem o texto, como está aparecendo atualmente.

Como o exemplo:

- Antes: “122 – Xapuri”
- Depois “Xapuri”

```
import pandas as pd

!gdown --id 16U7YTqmKhMUmkY7bHJ8sivGhkuS7DA6k
df_info_municipio = pd.read_csv('/content/info_municipios.csv')
```

Opção 1

```
import re
def limpa_nome_municipio(nome_municipio):
    nome_municipio_limpo = re.sub(r'([\d]* [--] )(.*)', r'\2',
    nome_municipio)
```

```

        return nome_municipio_limpo
municipios_tratados =
limpa_nome_municipio(df_info_municipio['Município'])

```

Opção 2

```

import re
def limpa_nome_municipio(nome_municipio):
    nome_municipio_limpo = re.sub(r'([\d]* [--] )(.*)', r'\2',
nome_municipio)
    return nome_municipio_limpo
municipios_tratados =
df_info_municipio['Município'].apply(limpa_nome_municipio)

```

3 – Substitua a coluna de Municípios com os dados que foram tratados anteriormente:

Opção 1

```
df_info_municipio.loc[:, 'Município'] = municipios_tratados
```

Opção 2

```
df_info_municipio.iloc[:, 'Município'] = municipios_tratados
```

4 - Juntar 2 dataframes

```

import pandas as pd
!gdown --id 1wXVWazKYzinH9Fv_rtWcxaara8lmiey7
df_estado = pd.read_csv('/content/estados.csv')
!gdown --id 1Zx3N-mfRiBc9XMGFAfbGw7WrKi9Bxjls
df_municipio = pd.read_csv('/content/municipios.csv')

```

Opção 1

```
df_junto = pd.concat(df_estado, df_municipio, on="codigo_uf")
df_junto.info()
```

Opção 2

```
df_junto = pd.merge(df_estado, df_municipio, on="codigo_uf")
df_junto.info()
```

5 - Renomear colunas, seguindo o dicionário abaixo

```
de_para_colunas = { 'nome_x' : 'Estado',
                    'latitude_x': 'latitude_uf',
                    'longitude_x': 'longitude_uf',
                    'nome_y' : 'Município',
                    'latitude_y': 'latitude_cd',
                    'longitude_y': 'longitude_cd',
                    }
```

Opção 1

```
df_junto.rename(columns=de_para_colunas, inplace=True)
```

Opção 2

```
df_junto.replace(columns=de_para_colunas, inplace=True)
```

6 - Gere um dicionário com os Estados e suas Siglas

```
import pandas as pd
!gdown --id 1wXVWazKYzinH9Fv_rtWcxaara81miey7
df_estado = pd.read_csv('/content/estados.csv')
```

Opção 1

```
df_estado[['nome', 'uf']].set_index('nome').to_dict()
```

Opção 2

```
dict(zip(df_estado.nome, df_estado.uf))
```