

▼ Importando a biblioteca

```
import pandas as pd
```

▼ Importando um arquivo csv

```
california_h_df = pd.read_csv('./sample_data/california_housing_train.csv')
```

texto em itálico### Tipos de Visualizações

- print
- display

```
print(california_h_df)
# display(california_h_df)
```

▼ Dados simples do Data Frame

- head (mostra por default as 5 primeiras linhas)
- shape (mostra a quantidade de linhas e colunas)
- describe (exibe um resumo de uma serie de informações da tabela)

```
display(california_h_df.head())
# display(california_h_df.shape)
# display(california_h_df.describe())
```

▼ Pegar somente uma coluna

```
total_rooms_sr = california_h_df['total_rooms']
display(total_rooms_sr)
```

▼ Pegar mais de uma coluna

```
filtros = california_h_df[['total_rooms', 'total_bedrooms']]
display(filtros)
```

▼ Método .loc

- Pega 1 unica linha

- Pega 1 linha com uma condicional
- Pega linhas e colunas especificas
- Pega 1 valor especifico

```
# pega pelo indice
# display(california_h_df.loc[1])
# display(california_h_df.loc[1:5])

# pega somente linhas que passar por uma condição
# display(california_h_df.loc[california_h_df['total_rooms'] >= 12000])

# pega varias linhas e colunas usando uma condição
# display(california_h_df.loc[california_h_df['total_rooms'] >= 12000,["total_rooms"]])

# pega a coluna de uma linha especifica
# display(california_h_df.loc[1, 'total_rooms'])
```

▼ Adicionando uma coluna nova

```
# a partir de uma coluna existente
# california_h_df['Nova Coluna'] = california_h_df['total_rooms']/2
# display(california_h_df)

# criando uma coluna com valor fixo
# california_h_df.loc[:, "Outra Nova Coluna"] = 10
# display(california_h_df)
```

▼ Apagar Colunas

```
# Excluir Coluna
california_h_df = california_h_df.drop("Outra Nova Coluna", axis=1)
display(california_h_df)
```

▼ Apagar Linhas

```
# Excluir Coluna
california_h_df = california_h_df.drop(1, axis=0)
display(california_h_df)
```

▼ Deletar valores vazios

- Deletar linhas/colunas vazias
- Deletar linhas que possuem valores vazios

- Preencher valores vazios (média e ultimo valor)

```
# deleta linhas que tem todos os valores varios
# california_h_df = california_h_df.dropna(how="all", axis=0)
# deleta colunas que tem todos os valores varios
# california_h_df = california_h_df.dropna(how="all", axis=1)

# deletar linhas que possuem pelo menos um valor vazio
# california_h_df = california_h_df.dropna()

# preenche os valores vazio com a media da coluna
# california_h_df['total_rooms'] = california_h_df['total_rooms'].fillna(california_h_df['
# preenche com o valor da ultima linha acima que contem valor
# california_h_df = california_h_df.ffill()
```

▼ Calculando Indicadores

- Groupby
- Value Counts

```
# conta o numero de vezes que tal valor aparece
# total_rooms = california_h_df['total_rooms'].value_counts()
# display(total_rooms)

# agrupa os dados por uma condição
total_rooms_media = california_h_df.groupby('total_rooms').sum()
display(total_rooms_media)
```

▼ Importando biblioteca de grafico

```
import matplotlib.pyplot as plt
```

▼ Gerando gráfico

```
#display(total_rooms_media['latitude'])
plt.plot(total_rooms_media['median_income'])
```

Produtos pagos do Colab - Cancelar contratos

