

### Exercício de Código

Refaça os tratamentos de valores discrepantes que foram realizados pela instrutora Carol nas últimas aulas!

### Parte 1:

```
✓ [1241] 1 import matplotlib.pyplot as plt
```

Acompanhe os outliers da coluna salário

E os limites:

Verifique a quantidade de pessoas por faixa salarial:

Vamos calcular a média e o desvio padrão:

Calcule o limite:

## Parte 2:

Vamos verificar a faixa salarial:

```
1 dados[dados['SALARIO']>limite_superior]['FAIXA SALARIAL'].value_counts()
```

FAIXA SALARIAL

Acima de R\$ 40.001/mês	19
de R\$ 30.001/mês a R\$ 40.000/mês	3
Name: count, dtype: int64	

Vamos filtrar determinados médias salariais e desconsiderar os outliers

```
media_30_40 = dados[dados['FAIXA SALARIAL'] == 'de R$ 30.001/mês a R$ 40.000/mês' & (dados['SALARIO'] > limite_superior)]
```

Em seguida, localize os outliers:

```
dados.loc[(dados['FAIXA SALARIAL'] == 'de R$ 30.001/mês a R$ 40.000/mês') & (dados['SALARIO'] > limite_superior), 'SALARIO'] = media_30_40
```

Confirme se deu certo:

```
[144]: 1 dados[dados['SALARIO'] > limite_superior]['FAIXA SALARIAL'].value_counts()

FAIXA SALARIAL
Acima de R$ 40.001/mês    19
Name: count, dtype: int64
```

Faça a média de salários acima de R\$ 40.000:

```
[145]: 1 media_40 = dados[dados['FAIXA SALARIAL'] == 'Acima de R$ 40.001/mês' & (dados['SALARIO'] > limite_superior)]['SALARIO'].mean()
      2 media_40

53127.846153846156

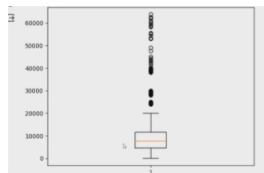
[146]: 1 dados.loc[(dados['FAIXA SALARIAL'] == 'Acima de R$ 40.001/mês') & (dados['SALARIO'] > limite_superior), 'SALARIO'] = media_40

[147]: 1 dados[dados['SALARIO'] > limite_superior]['FAIXA SALARIAL'].value_counts()

Series([], Name: count, dtype: int64)
```

Verifique o resultado:

```
1 plt.boxplot(dados['SALARIO'])
```



discrepantes

QUIZ - 1 PERGUNTA

4.4 - Intervalo de confiança e distribuição

MARCAR COMO INCOMPLETA

CONTINUAR →