

Exercício de Código | Tratando valores discrepantes

7 COMENTÁRIOS/DÚVIDAS

Exercício de Código

Vamos colocar em prática? 🚀

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

Caso ajude, deixamos algumas dicas com passo a passo por aqui:

Parte 1:

Importe a biblioteca matplotlib:

```
[124] 1 import matplotlib.pyplot as plt
      2 plt.boxplot(lista_idades)
```

Acompanhe os outliers da coluna salário

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

Aplique a regra dos quartis:

```
[127] 1 Q1 = dados['SALARIO'].quantile(0.25)
      2 Q1
4751.5
[128] 1 Q3 = dados['SALARIO'].quantile(0.75)
      2 Q3
11794.5
      1 IQR = Q3-Q1
      2 IQR
```

E os limites:

```
[130] 1 lim_superior = Q3 + (1.5*IQR)
      2 lim_superior
22359.0
      1 lim_inferior = Q1 - (1.5*IQR)
      2 lim_inferior
-5813.0
```

Verifique a quantidade de pessoas por faixa salarial:

```
[131] 1 dados['FAIXA SALARIAL'].value_counts()
FAIXA SALARIAL
de R$ 8.001/mês a R$ 12.000/mês    752
de R$ 4.001/mês a R$ 6.000/mês    584
de R$ 6.001/mês a R$ 8.000/mês    527
de R$ 12.001/mês a R$ 16.000/mês   437
de R$ 3.001/mês a R$ 4.000/mês    303
de R$ 2.001/mês a R$ 3.000/mês    285
de R$ 16.001/mês a R$ 20.000/mês   229
de R$ 1.001/mês a R$ 2.000/mês    216
de R$ 20.001/mês a R$ 25.000/mês   134
de R$ 25.001/mês a R$ 30.000/mês    77
de R$ 30.001/mês a R$ 40.000/mês    70
Acima de R$ 40.001/mês             45
Menos de R$ 1.000/mês              35
Name: count, dtype: int64
```

Vamos calcular a média e o desvio padrão:

```
[134] 1 media_salario = dados['SALARIO'].mean()
      2 media_salario
10517.53348162023
      1 desvio_salario = dados['SALARIO'].std()
      2 desvio_salario
18096.21098911858
```

Calcule o limite:

```
[136] 1 limite_superior = media_salario + (3*desvio_salario)
      2 limite_superior
64806.166448975964
```

Parte 2:

Vamos verificar a faixa salarial:

```
[137] 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]: 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]: media_40 = dados[(dados['FAIXA SALARIAL'] == 'Acima de R$ 40.001/mês') & (dados['SALARIO'] > limite_superior)]['SALARIO'].mean()
media_40
53127.846153846156

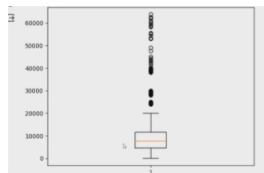
[146]: dados.loc[(dados['FAIXA SALARIAL'] == 'Acima de R$ 40.001/mês') & (dados['SALARIO'] > limite_superior), 'SALARIO'] = media_40

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

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

Verifique o resultado:

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



VIDEO - 11 MIN

Exercício de Código | Intervalo de
confiança e distribuição amostral

COMPLETE E CONTINUE SEU CURSO! →