

✓ Módulo 2 - Esquentando os 11/11
motores!

✓ Módulo 3 - Primeira análise 17/17

○ Módulo 4 - Análises 15/25
estatísticas com python

✓ 4.1 - Estatística básica
VIDEO - 23 MIN

✓ Exercício de Código | Estatística
TEXT

✓ Faça você mesmo | Estatística
TEXT

○ 4.2 - Tratando valores faltantes (parte 1)
VIDEO - 9 MIN

○ 4.2 - Tratando valores faltantes (parte 2)
VIDEO - 17 MIN

○ 4.2 - Tratando valores faltantes (parte 3)
VIDEO - 5 MIN

✓ Exercício de Código | Valores faltantes
TEXT

✓ Exercício de Revisão | Tratando valores
faltantes
QUIZ - 1 PERGUNTA

✓ Faça você mesmo | Tratando valores

Exercício de Código | Valores faltantes

7 COMENTÁRIOS/DÚVIDAS

Exercício de Código

Nas últimas aulas vimos diferentes estratégias para lidar com dados faltantes, não é mesmo?
Que tal replicar cada um delas por aí?

E vamos deixar aqui o passo a passo para acompanhar e ver se seus resultados bateram com o da instrutora Carol! 🍷

Parte 1

Primeiro, vamos conferir os valores nulos:

Valores faltantes

```
1 dados.info()
```

Vamos trabalhar com a coluna de gênero:

Trabalhando coluna de gênero

```
1 dados.groupby('GÊNERO', dropna=False)['ID'].nunique()
```

```
[72] 1 dados.groupby('GÊNERO', dropna=False)['ID'].nunique()

GÊNERO
Feminino      1855
Masculino     3194
Prefiro não informar    12
NaN              9
Name: ID, dtype: int64
```

```
1 dados['GÊNERO'].fillna("Prefiro não informar")
```

Atribua o valor do **fillna** na coluna de **gênero**

```
1 dados['GÊNERO'] = dados['GÊNERO'].fillna("Prefiro não informar")
```

Por fim, acompanhe a substituição dos **nulos (NaN)** por **Prefiro não informar**

```
1 dados.groupby('GÊNERO', dropna=False)['ID'].nunique()

GÊNERO
Feminino      1855
Masculino     3194
Prefiro não informar    21
Name: ID, dtype: int64
```

Parte 2

Vamos trabalhar com a coluna de idade:

```
1 dados[dados['IDADE'].isnull()]['FAIXA IDADE'].value_counts()

FAIXA IDADE
55+      68
17-21      6
Name: count, dtype: int64
```

Faça a média das idades 17-21

```
1 dados[dados['FAIXA IDADE']=="17-21"]['IDADE'].mean()

28.285479452854796
```

Adicione a média na variável, encontre as linhas que estejam com idades de 17-21 e na coluna de idades esteja nulo:

```
[83] 1 media_17_21 = dados[dados['FAIXA IDADE']=="17-21"]['IDADE'].mean()

[84] 1 dados.loc[(dados['FAIXA IDADE']=="17-21') & (dados['IDADE'].isnull()), 'IDADE']

1925 NaN
2393 NaN
3367 NaN
4086 NaN
4177 NaN
4269 NaN
Name: IDADE, dtype: float64
```

Por fim, substitua os valores adicionando = **media_17_21**

```
[85] 1 [(dados['FAIXA IDADE']=="17-21') & (dados['IDADE'].isnull()), 'IDADE'] = media_17_21
```

Confirme que deu certo:

```
1 dados[dados['IDADE'].isnull()]['FAIXA IDADE'].value_counts()

FAIXA IDADE
55+      68
Name: count, dtype: int64
```

Verifique a média da coluna de idade 55+

```
1 media_geral = dados['IDADE'].mean()
2 media_geral

31.153517226250375
```

Localize os valores nulos para substituí-los pela média geral:

```
1 dados.loc[(dados['FAIXA_IDADE']=='55+') & (dados['IDADE'].isnull()), 'IDADE']
```

Adicione = **media_geral**

```
195] oc[(dados['FAIXA_IDADE']=='55+') & (dados['IDADE'].isnull()), 'IDADE'] = media_geral
```

Parte 3

Vamos trabalhar com a coluna de salário

Filtre a mediana e substituir nos valores nulos:

```
[181] 1 mediana_salario = dados['SALARIO'].median()
7625.5
```

```
1 dados.loc[dados['SALARIO'].isnull(), 'SALARIO'] = mediana_salario
```

faltantes

TEXT

MARCAR COMO INCOMPLETA

CONTINUAR →