

Lista de Exercícios 2 (16/14/2025)

Ciência de Dados - 2025/1
Dr. Prof. Raimundo C. S. Vasconcelos
Dr. Prof. Fabiano C. Fernandes

Tales Lima de Oliveira
tales.oliveira@estudante.ifb.edu.br

Disponibilidade do Código

Os códigos-fonte e scripts desenvolvidos para esta atividade estão disponíveis publicamente e podem ser acessados através do [GitHub](#) e do [Google Colab](#).

Código 0: Caminho para os arquivos.

```
1 ##### IMPORTAR DADOS DO DRIVE #####
2 from google.colab import drive
3 drive.mount('/content/drive')
4 path_ecom = '/content/drive/MyDrive/ecommerce_purchases.csv'
5 path_conc = '/content/drive/MyDrive/conceito_enade_2021.xlsx'
6
7 ##### IMPORTAR DADOS DA PASTA LOCAL #####
8 path_ecom = 'data/ecommerce_purchases.csv'
9 path_conc = 'data/conceito_enade_2021.xlsx'
```

1. Exercício - Compras de Ecommerce

Código 1.1: Importe pandas e leia o arquivo csv Ecommerce Purchases e configure-o para um DataFrame chamado ecom.

```
1 import pandas as pd
2 df_ecom = pd.read_csv(path_ecom)
```

Código 1.2: Verifique o head do DataFrame.

```
1 df_ecom.head()
```

Código 1.3: Quantas linhas e colunas existem?

```
1 rows, columns = df_ecom.shape
2 print(rows, columns)
```

Código 1.4: Qual é o preço de compra médio?

```
1 average_price = df_ecom['Purchase Price'].mean()
2 print(f"${average_price:.2f}")
```

Código 1.5: Quais foram os preços de compra mais altos e mais baixos?

```
1 max_price = df_ecom['Purchase Price'].max()
2 min_price = df_ecom['Purchase Price'].min()
3
4 print(f"Max: ${max_price:.2f}")
5 print(f"Min: ${min_price:.2f}")
```

Código 1.6: Quantas pessoas têm Inglês (**en**) como sua língua de escolha no site?

```
1 num_english_users = (df_ecom['Language'] == 'en').sum()
2 print(num_english_users)
```

Código 1.7: Quantas pessoas têm o cargo de **Lawyer** (Advogado)?

```
1 num_lawyers = df_ecom['Job'].str.contains(r'\blawyer\b', case=False, na=False).sum()
2 # num_lawyers = df_ecom['Job'].str.contains('Lawyer').sum()
3 print(num_lawyers)
```

Código 1.8: Quantas pessoas fizeram a compra durante a AM e PM?

```
1 am_pm_counts = df_ecom['AM or PM'].value_counts()
2 print(am_pm_counts)
```

Código 1.9: Quais são os 5 títulos de trabalho mais comuns?

```
1 top_5_job = df_ecom['Job'].value_counts().head(5)
2 print(top_5_job)
```

Código 1.10: Alguém fez uma compra que veio do Lot: **90 WT**, qual foi o preço de compra para esta transação?

```
1 price_for_lot_90wt = df_ecom[df_ecom['Lot'] == '90 WT']['Purchase Price'].iloc[0]
2 print(f"${price_for_lot_90wt:.2f}")
```

Código 1.11: Qual é o email da pessoa com o seguinte número do cartão de crédito: **4926535242672853**

```
1 email_for_card = df_ecom[df_ecom['Credit Card'] == 4926535242672853]['Email'].iloc[0]
2 print(email_for_card)
```

Código 1.12: Quantas pessoas têm o American Express como seu fornecedor de cartão de crédito e fizeram uma compra acima de US \$95?

```
1 num_american_express_above_95 = df_ecom[(df_ecom['CC Provider'] == 'American Express') &
2 ↪ (df_ecom['Purchase Price'] > 95)].shape[0]
3 print(num_american_express_above_95)
```

Código 1.13: *Difícil:* Quantas pessoas tem um cartão de crédito que expira em 2025?

```
1 df_ecom['CC Exp Date'] = df_ecom['CC Exp Date'].astype(str)
2 num_expiring_2025 = df_ecom[df_ecom['CC Exp Date'].str.endswith('25')].shape[0]
3 print(num_expiring_2025)
```

Código 1.14: *Difícil:* Quais são os 5 principais provedores de e-mail?

```
1 df_ecom['Email Domain'] = df_ecom['Email'].str.split('@').str[1]
2 top_5_email_providers = df_ecom['Email Domain'].value_counts().head(5)
3 print(top_5_email_providers)
```

2. Exercício - Compras de Ecommerce

Código 2.1: Faça uma análise estatística e realize as plotagens que achar pertinente na exploração dos dados.

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
```

```
1 df_conc = pd.read_excel(path_conc)
2
3 print(df_conc.head())
4 print(df_conc.info())
5 print(df_conc.describe())
6 print(df_conc.isnull().sum())
```

Código 2.2: Construa um modelo de clusterização e analise os resultados encontrados.

```
1 #!TODO...
```