

## Exercicio2

March 4, 2025

```
[1]: # Importar bibliotecas
from pyspark.sql import SparkSession

# Criar sessão do Spark
spark = SparkSession.builder.appName("Ecommerce_Analysis").getOrCreate()

# Criar dados das compras
dados = [
    (1, 'Alice Silva', 'Notebook Dell', 'Eletrônicos', 4500.0, 1, 'Cartão de_
↳Crédito'),
    (2, 'João Pedro', 'Mouse Logitech', 'Eletrônicos', 150.0, 2, 'Boleto'),
    (3, 'Maria Souza', 'Geladeira Brastemp', 'Eletrodomésticos', 3200.0, 1,
↳'Pix'),
    (4, 'Bruno Lima', 'Celular Samsung', 'Eletrônicos', 2500.0, 1, 'Cartão de_
↳Crédito'),
    (5, 'Camila Rocha', 'Cadeira Gamer', 'Móveis', 1200.0, 1, 'Boleto')
]

# Definir os nomes das colunas
colunas = ["id_compra", "cliente", "produto", "categoria", "preco",
↳"quantidade", "metodo_pagamento"]

# Criar DataFrame
df = spark.createDataFrame(dados, colunas)

# Criar uma tabela temporária
df.createOrReplaceTempView("compras")
```

```
[ ]:
```

```
[2]: # Perguntas para os alunos:
# 1. Liste todas as compras feitas por clientes que pagaram com "Cartão de_
↳Crédito".
# 2. Ordene as compras pelos preços dos produtos em ordem decrescente.
# 3. Calcule o valor total gasto por cada cliente.
# 4. Conte quantas compras foram feitas por categoria de produto.
# 5. Adicione uma nova coluna `valor_total` (preco * quantidade).
```

```
[3]: # a) Listar todas as compras feitas por clientes que pagaram com "Cartão de
      ↳Crédito"
      spark.sql("""
        SELECT * FROM compras
        WHERE metodo_pagamento = 'Cartão de Crédito'
        """).show()
```

```
+-----+-----+-----+-----+-----+-----+
-----+
|id_compra|  cliente|      produto| categoria| preco|quantidade|
metodo_pagamento|
+-----+-----+-----+-----+-----+-----+
-----+
|      1| Alice Silva| Notebook Dell| Eletrônicos|4500.0|      1|Cartão de
Crédito|
|      4| Bruno Lima| Celular Samsung| Eletrônicos|2500.0|      1|Cartão de
Crédito|
+-----+-----+-----+-----+-----+-----+
-----+
```

```
[4]: # b) Ordenar as compras pelos preços dos produtos em ordem decrescente
      spark.sql("""
        SELECT * FROM compras
        ORDER BY preco DESC
        """).show()
```

```
+-----+-----+-----+-----+-----+-----+
-----+
|id_compra|  cliente|      produto| categoria| preco|quantidade|
metodo_pagamento|
+-----+-----+-----+-----+-----+-----+
-----+
|      1| Alice Silva| Notebook Dell| Eletrônicos|4500.0|
1|Cartão de Crédito|
|      3| Maria Souza| Geladeira Brastemp| Eletrodomésticos|3200.0|      1|
Pix|
|      4| Bruno Lima| Celular Samsung| Eletrônicos|2500.0|
1|Cartão de Crédito|
|      5| Camila Rocha| Cadeira Gamer| Móveis|1200.0|      1|
Boleto|
|      2| João Pedro| Mouse Logitech| Eletrônicos| 150.0|      2|
Boleto|
+-----+-----+-----+-----+-----+-----+
-----+
```

```
[5]: # c) Calcular o valor total gasto por cada cliente
spark.sql("""
    SELECT cliente, SUM(preco * quantidade) AS valor_total
    FROM compras
    GROUP BY cliente
    """).show()
```

cliente	valor_total
Alice Silva	4500.0
João Pedro	300.0
Maria Souza	3200.0
Bruno Lima	2500.0
Camila Rocha	1200.0

```
[6]: # d) Contar quantas compras foram feitas por categoria de produto
```

```
spark.sql("""
    SELECT categoria, COUNT(*) AS total_compras
    FROM compras
    GROUP BY categoria
    """).show()
```

categoria	total_compras
Eletrônicos	3
Eletrodomésticos	1
Móveis	1

```
[7]: # e) Adicionar uma nova coluna `valor_total` (preco * quantidade)
```

```
spark.sql("""
    SELECT *, (preco * quantidade) AS valor_total
    FROM compras
    """).show()
```

id_compra	cliente	produto	categoria	preco	quantidade	valor_total
1	Alice Silva	Notebook Dell	Eletrônicos	4500.0		

