

## 4.SQL\_ecommerce

March 4, 2025

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

# Criar sessão do Spark com suporte ao Hive
spark = SparkSession.builder \
    .appName("Ecommerce_Analysis") \
    .enableHiveSupport() \
    .getOrCreate()

# Criar o banco de dados se não existir
spark.sql("CREATE DATABASE IF NOT EXISTS ecommerce")

# Selecionar o banco de dados
spark.sql("USE ecommerce")
```

[1]: DataFrame[]

```
[2]: # Criar a tabela permanente no Hive
spark.sql("""
    CREATE TABLE IF NOT EXISTS compras (
        id_compra INT,
        cliente STRING,
        produto STRING,
        categoria STRING,
        preco DOUBLE,
        quantidade INT,
        metodo_pagamento STRING
    ) USING PARQUET
""")
```

[2]: DataFrame[]

```
[ ]: # 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)

# Gravar os dados na tabela Hive
df.write.mode("overwrite").saveAsTable("ecommerce.compras")

```

```

[ ]: # a) Listar todas as compras feitas por clientes que compraram mais de uma
      unidade de um produto
spark.sql("""
    SELECT * FROM compras
    WHERE quantidade > 1
""").show()

```