

## 2.EnableHiveSupport\_ecommerce

March 4, 2025

```
[5]: # 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")

# 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
""")

# 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")

```

```

[6]: # Perguntas para os alunos:
# 1. Liste todas as compras feitas por clientes que compraram mais de uma
    ↪ unidade de um produto.
# 2. Ordene os produtos pelo nome em ordem alfabética.
# 3. Calcule o total arrecadado por categoria de produto.
# 4. Identifique qual método de pagamento foi mais utilizado.
# 5. Filtre as compras onde o valor total (preço * quantidade) seja superior a
    ↪ 2000.

# 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()

# b) Ordenar os produtos pelo nome em ordem alfabética
spark.sql("""
    SELECT * FROM compras
    ORDER BY produto ASC
    """).show()

# c) Calcular o total arrecadado por categoria de produto
spark.sql("""
    SELECT categoria, SUM(preco * quantidade) AS total_arrecadado
    FROM compras
    GROUP BY categoria
    """).show()

# d) Identificar qual método de pagamento foi mais utilizado
spark.sql("""
    SELECT metodo_pagamento, COUNT(*) AS total_uso
    FROM compras
    GROUP BY metodo_pagamento
    """).show()

```

```

ORDER BY total_uso DESC
LIMIT 1
""").show()

# e) Filtrar as compras onde o valor total (preço * quantidade) seja superior a
↳ 2000
spark.sql("""
SELECT *, (preço * quantidade) AS valor_total
FROM compras
WHERE (preço * quantidade) > 2000
""").show()

```

```

+-----+-----+-----+-----+-----+-----+
--+
|id_compra| cliente|      produto|
categoria|preço|quantidade|metodo_pagamento|
+-----+-----+-----+-----+-----+-----+
--+
|          2|João Pedro|Mouse Logitech|Eletrônicos|150.0|          2|
Boleto|
+-----+-----+-----+-----+-----+-----+
--+

```

```

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

```

```

+-----+-----+
|      categoria|total_arrecadado|
+-----+-----+
|      Eletrônicos|          7300.0|
|Eletrodomésticos|          3200.0|

```

	Móveis	1200.0
--	--------	--------

  

metodo_pagamento	total_uso
Cartão de Crédito	2

  

id_compra	cliente	produto	categoria	preco	quantidade
1	4  Bruno Lima	Celular Samsung	Eletrônicos	2500.0	
	Cartão de Crédito	2500.0			
1	1  Alice Silva	Notebook Dell	Eletrônicos	4500.0	
	Cartão de Crédito	4500.0			
	3  Maria Souza	Geladeira Brastemp	Eletrodomésticos	3200.0	1
	Pix	3200.0			

```
[ ]: # TRABALHAO -----

# 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).

# 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()

# b) Ordenar as compras pelos preços dos produtos em ordem decrescente
spark.sql("""
    SELECT * FROM compras
```

```

        ORDER BY preco DESC
    """).show()

# 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()

# d) Contar quantas compras foram feitas por categoria de produto
spark.sql("""
    SELECT categoria, COUNT(*) AS total_compras
    FROM compras
    GROUP BY categoria
""").show()

# e) Adicionar uma nova coluna `valor_total` (preco * quantidade)
spark.sql("""
    SELECT *, (preco * quantidade) AS valor_total
    FROM compras
""").show()

```