



## Projeto Final - BIG QUERY

**Professora:** Franciane Rodrigues

**Equipe:**

1. Ana Carolina
2. Carina Mayara
3. David Ben
4. Giovanna Shiguemori
5. Marcos Neves
6. Rafael Souza

- 1) -- Pergunta de negócio 1 :  
-- Análise de Vendas por Categoria de Produto:  
-- Qual categoria de produto tem o maior volume de vendas totais?

```
SELECT categoria_produto, SUM(total_vendas) AS total_vendas
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY categoria_produto
ORDER BY total_vendas DESC
LIMIT 1;
```

#### Explicação:

- **SELECT** categoria\_produto, **SUM**(total\_vendas) AS total\_vendas: Seleciona a categoria do produto e a soma das vendas totais para cada categoria.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica o dataset e a tabela que contém os dados.
- **GROUP BY** categoria\_produto: Agrupa os dados por categoria de produto.
- **ORDER BY** total\_vendas **DESC**: Ordena os resultados em ordem decrescente de vendas totais.
- **LIMIT 1**: Retorna apenas a categoria de produto com o maior volume de vendas.

The screenshot displays the Google Cloud BigQuery web interface. On the left, the 'Explorer' pane shows the project hierarchy: 'projeto-final-424922' containing a 'Datalake' dataset with a table named 'Tabela\_original'. The main editor shows a SQL query titled 'Pergunta1' with the following text:

```
1 -- Pergunta de negócio 1 :
2 -- Qual categoria de produto tem o maior volume de vendas totais?
3 SELECT categoria_produto, SUM(total_vendas) AS total_vendas
4 FROM `projeto-final-424922.Datalake.Tabela_original`
5 GROUP BY categoria_produto
6 ORDER BY total_vendas DESC
7 LIMIT 1;
```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	categoria_produto	total_vendas
1	eletronicos	34982.41

- 2) -- Pergunta de negócio 2 :  
-- Análise de Vendas por Região:  
-- Qual região tem o maior valor total de vendas?

```
SELECT regioao, SUM(total_vendas) AS total_vendas
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY regioao
ORDER BY total_vendas DESC
LIMIT 1;
```

#### Explicação:

- **SELECT regioao, SUM(total\_vendas) AS total\_vendas:** Seleciona a região e a soma das vendas totais para cada região.
- **FROM `projeto-final-424922.Datalake.Tabela\_original`:** Especifica o dataset e a tabela que contém os dados.
- **GROUP BY regioao:** Agrupa os dados por região.
- **ORDER BY total\_vendas DESC:** Ordena os resultados em ordem decrescente de vendas totais.
- **LIMIT 1:** Retorna apenas a região com o maior valor total de vendas.

The screenshot displays the Google Cloud BigQuery console. The left sidebar shows the 'Explorer' view with a tree structure containing 'projeto-final-424922', 'Queries', 'Datalake', and 'dados\_originais'. The main panel shows a SQL query named 'Pergunta2' with the following text:

```
1 -- Pergunta de negócio 2 :
2 -- Análise de Vendas por Região:
3 --Qual região tem o maior valor total de vendas?
4
5 SELECT regioao, SUM(total_vendas) AS total_vendas
6 FROM `projeto-final-424922.Datalake.Tabela_original`
7 GROUP BY regioao
8 ORDER BY total_vendas DESC
9 LIMIT 1;
```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	regiao	total_vendas
1	América do Norte	36844.34000000...

- 3) -- Pergunta de negócio 3 :  
-- Análise de Vendas por Método de Pagamento:  
-- Qual método de pagamento é o mais utilizado?

```
SELECT metodo_pagamento, COUNT(id_transacao) AS total_transacoes
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY metodo_pagamento
ORDER BY total_transacoes DESC
LIMIT 1;
```

#### Explicação:

- **SELECT** metodo\_pagamento, **COUNT**(id\_transacao) **AS** total\_transacoes: Seleciona o método de pagamento e a contagem total de transações para cada método.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica o dataset e a tabela que contém os dados.
- **GROUP BY** metodo\_pagamento: Agrupa os dados por método de pagamento.
- **ORDER BY** total\_transacoes **DESC**: Ordena os resultados em ordem decrescente de contagem de transações.
- **LIMIT 1**: Retorna apenas o método de pagamento mais utilizado.

The screenshot shows the Google Cloud BigQuery console. The left sidebar displays the 'Explorer' view with a tree structure of resources under 'projeto-final-424922', including 'Queries' (Pergunta1, Pergunta2, Pergunta3) and 'Datalake' (Pergunta1, Pergunta2, Pergunta3, Tabela\_original, dados\_originais). The main panel shows the SQL query for 'Pergunta3' with a 'RUN' button. Below the query, the 'Query results' section is visible, showing a table with two columns: 'metodo\_pagamento' and 'total\_transacoes'. The first row shows 'Cartão de Crédito' with a count of 120.

Row	metodo_pagamento	total_transacoes
1	Cartão de Crédito	120

- 4) -- Pergunta de negócio 4 :  
-- Desempenho de Produtos Específicos:  
-- Qual produto gerou a maior receita total?

```
SELECT nome_produto, SUM(total_vendas) AS total_vendas
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY nome_produto
ORDER BY total_vendas DESC
LIMIT 1;
```

#### Explicação:

- **SELECT** nome\_produto, **SUM**(total\_vendas) **AS** total\_vendas: Seleciona o nome do produto e a soma das vendas (total\_vendas) para cada produto.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica a tabela de dados a ser usada.
- **GROUP BY** nome\_produto: Agrupa os resultados por nome de produto, para calcular a soma das vendas de cada produto individualmente.
- **ORDER BY** total\_vendas **DESC**: Ordena os resultados em ordem decrescente de total\_vendas, colocando o produto com a maior receita no topo.
- **LIMIT 1**: Retorna apenas o primeiro resultado, que é o produto com a maior receita.

The screenshot displays the Google Cloud BigQuery console. On the left, the 'Explorer' sidebar shows a project named 'projeto-final-424922' with a 'Datalake' folder containing a table named 'Tabela\_original'. The main editor shows a SQL query labeled 'Pergunta4' with the following text:

```
1 -- Pergunta de negócio 4 :
2 -- Desempenho de Produtos Específicos:
3 -- Qual produto gerou a maior receita total?
4
5 SELECT nome_produto, SUM(total_vendas) AS total_vendas
6 FROM `projeto-final-424922.Datalake.Tabela_original`
7 GROUP BY nome_produto
8 ORDER BY total_vendas DESC
9 LIMIT 1;
```

Below the query editor, the 'Query results' section is visible, showing a table with two columns: 'nome\_produto' and 'total\_vendas'. The results are as follows:

Row	nome_produto	total_vendas
1	Câmera Canon EOS R5	3899.99

- 5) -- Pergunta de negócio 5 :  
-- Desempenho de Produtos Específicos:  
-- Qual produto foi vendido em maior quantidade de unidades?

```
SELECT nome_produto, SUM(unidade_venda) AS total_unidades
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY nome_produto
ORDER BY total_unidades DESC
LIMIT 1;
```

#### Explicação:

- **SELECT** nome\_produto, **SUM**(unidade\_venda) **AS** total\_unidades: Seleciona o nome do produto e a soma das unidades vendidas (total\_unidades) para cada produto.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica a tabela de dados a ser usada.
- **GROUP BY** nome\_produto: Agrupa os resultados por nome de produto, para calcular a soma das unidades vendidas de cada produto individualmente.
- **ORDER BY** total\_unidades **DESC**: Ordena os resultados em ordem decrescente de total\_unidades, colocando o produto com a maior quantidade de unidades vendidas no topo.
- **LIMIT 1**: Retorna apenas o primeiro resultado, que é o produto com a maior quantidade de unidades vendidas.

The screenshot shows the Google Cloud BigQuery interface. The top navigation bar includes the Google Cloud logo, a project selector set to 'PROJETO-FINAL', and a search bar. The left sidebar contains an 'Explorer' panel with a search bar and a tree view of resources. Under the 'projeto-final-424922' project, there are 'Queries' (Pergunta1 through Pergunta5) and 'Datalake' (Tabela\_original). The main panel displays the query 'Pergunta5' with its SQL code and a 'Query results' section. The results are shown in a table with columns 'nome\_produto' and 'total\_unidades'.

Query results table:

Row	nome_produto	total_unidades
1	Camiseta Hanes ComfortSoft	10

- 6) -- Pergunta de negócio 6 :  
-- Tendências Temporais:  
-- Qual mês teve o maior volume de vendas?

```
SELECT EXTRACT(MONTH FROM data) AS mes, SUM(total_vendas) AS total_vendas
FROM `projeto-final-424922.Datalake.Tabela_original`
GROUP BY mes
ORDER BY total_vendas DESC
LIMIT 1;
```

#### Explicação:

- **SELECT EXTRACT(MONTH FROM data) AS mes, SUM(total\_vendas) AS total\_vendas:** Seleciona o mês da data da transação e a soma das vendas totais para cada mês.
- **FROM `projeto-final-424922.Datalake.Tabela\_original`:** Especifica o dataset e a tabela que contém os dados.
- **GROUP BY mes:** Agrupa os dados por mês.
- **ORDER BY total\_vendas DESC:** Ordena os resultados em ordem decrescente de vendas totais.
- **LIMIT 1:** Retorna apenas o mês com o maior volume de vendas.

The screenshot shows the Google Cloud BigQuery interface. The top bar includes the Google Cloud logo, a project selector set to 'PROJETO-FINAL', and a search bar. The left sidebar shows the 'Explorer' view with a list of resources under 'Datalake', including 'Pergunta1' through 'Pergunta6' and 'Tabela\_original'. The main panel displays a query named 'Pergunta6' with the following SQL code:

```
1 -- Pergunta de negócio 6 :
2 -- Tendências Temporais:
3 -- Qual mês teve o maior volume de vendas?
4
5
6 SELECT EXTRACT(MONTH FROM data) AS mes, SUM(total_vendas) AS total_vendas
7 FROM `projeto-final-424922.Datalake.Tabela_original`
8 GROUP BY mes
9 ORDER BY total_vendas DESC
10 LIMIT 1;
11
```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	mes	total_vendas
1	1	14548.31999999...

The interface also includes tabs for 'JOB INFORMATION', 'RESULTS', 'CHART', 'JSON', 'EXECUTION DETAILS', and 'EXECUTION GRAPH'. At the bottom, there is a 'Job history' section with a 'REFRESH' button.

- 7) -- Pergunta de negócio 7 :  
-- Eficiência de Preços::  
-- Qual produto tem o maior preço por unidade?

```
SELECT nome_produto, preco_unidade
FROM `projeto-final-424922.Datalake.Tabela_original`
ORDER BY preco_unidade DESC
LIMIT 1;
```

#### Explicação:

- **SELECT** nome\_produto, preco\_unidade: Seleciona o nome do produto e o preço por unidade.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica o dataset e a tabela que contém os dados.
- **ORDER BY** preco\_unidade **DESC**: Ordena os resultados em ordem decrescente de preço por unidade.
- **LIMIT 1**: Retorna apenas o produto com o maior preço por unidade.

The screenshot displays the Google Cloud BigQuery interface. On the left, the 'Explorer' pane shows a project named 'PROJETO-FINAL' with a list of resources including 'Pergunta1' through 'Pergunta7' and a 'Datalake' folder. The main editor shows a SQL query for 'Pergunta7'. The query is as follows:

```
1 -- Pergunta de negócio 7 :
2 -- Eficiência de Preços::
3 -- Qual produto tem o maior preço por unidade?
4
5 SELECT nome_produto, preco_unidade
6 FROM `projeto-final-424922.Datalake.Tabela_original`
7 ORDER BY preco_unidade DESC
8 LIMIT 1;
```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	nome_produto	preco_unidade
1	Câmera Canon EOS R5	3899.99

The interface also includes a 'Job history' section at the bottom and various utility buttons like 'RUN', 'SAVE QUERY', 'DOWNLOAD', 'SHARE', 'SCHEDULE', and 'MORE'.



- 8) -- Pergunta de negócio 8:  
-- Top 10 dos produtos mais vendidos na Europa no mês de Janeiro de 2024

```
SELECT
    nome_produto,
    SUM(total_vendas) AS total_vendas_janeiro
FROM
    `projeto-final-424922.Datalake.Tabela_original`
WHERE
    EXTRACT(MONTH FROM data) = 1 -- Janeiro
    AND EXTRACT(YEAR FROM data) = 2024 -- Ano de 2024
    AND regioao = 'Europa'
GROUP BY
    nome_produto
ORDER BY
    total_vendas_janeiro DESC
LIMIT
    10;
```

#### Explicação:

- **SELECT** nome\_produto, **SUM**(total\_vendas) **AS** total\_vendas\_janeiro: Seleciona o nome do produto e a soma das vendas como "total\_vendas\_janeiro".
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica de qual tabela e dataset os dados serão extraídos.
- **WHERE** **EXTRACT**(MONTH FROM data) = 1: Filtra os dados para incluir apenas vendas realizadas em Janeiro (mês 1).
- **AND** **EXTRACT**(YEAR FROM data) = 2024: Filtra os dados para incluir apenas vendas realizadas no ano de 2024.
- **AND** regioao = 'Europa': Filtra os dados para incluir apenas vendas na região 'Europa'.
- **GROUP BY** nome\_produto: Agrupa os resultados pelo nome do produto, permitindo a soma das vendas para cada produto.
- **ORDER BY** total\_vendas\_janeiro **DESC**: Ordena os resultados em ordem decrescente com base no total de vendas em Janeiro.
- **LIMIT** 10: Limita os resultados aos 10 primeiros, mostrando apenas os produtos mais vendidos.

The screenshot displays the Google Cloud BigQuery interface. The top navigation bar shows 'Google Cloud' and 'PROJETO-FINAL'. The left sidebar contains the 'Explorer' panel with a search bar and a list of resources including 'Pergunta6', 'Pergunta7', 'Pergunta8', 'Notebooks', 'Data canvases', 'External connect...', and 'Datalake'. The main panel shows the query editor for 'Pergunta8' with the SQL query pasted in. Below the editor, the 'Query results' section is visible, showing a table with two columns: 'nome\_produto' and 'total\_vendas\_janeiro'. The results table contains two rows of data.

Row	nome_produto	total_vendas_janeiro
1	Purificador de Ar Blueair Classi...	1199.98
2	Aspirador de Pó Dyson V11	499.99

9) -- Pergunta de negócio 9:  
-- Qual foi o dia que mais teve produtos vendidos?

```
SELECT
  data,
  SUM(unidade_venda) AS total_transacoes
FROM
  `projeto-final-424922.Datalake.Tabela_original`
GROUP BY
  data
ORDER BY
  total_transacoes DESC
LIMIT
  1;
```

#### Explicação:

- **SELECT data, SUM(unidade\_venda) AS total\_transacoes:** Esta parte seleciona duas informações:
- **data:** Recupera a data da coluna "data".
- **SUM(unidade\_venda) AS total\_transacoes:** Calcula a soma dos valores na coluna "unidade\_venda" (provavelmente representando unidades vendidas) e nomeia essa soma calculada como "total\_transacoes" (total de transações).
- **FROM `projeto-final-424922.Datalake.Tabela\_original`:** Especifica a origem dos dados, que é a tabela "Tabela\_original" localizada dentro do conjunto de dados "Datalake" do projeto "projeto-final-424922".
- **GROUP BY data:** Agrupa os resultados com base em datas únicas. Portanto, você obterá uma linha para cada data distinta em seu conjunto de dados.
- **ORDER BY total\_transacoes DESC:** Ordena os resultados em ordem decrescente com base no "total\_transacoes" calculado. O dia com o maior número de transações estará no topo.
- **LIMIT 1:** Limita a saída apenas à primeira linha após a ordenação. Basicamente, você está pedindo a data com o maior total de transações.

The screenshot displays the Google Cloud BigQuery console. The top navigation bar includes the Google Cloud logo, a project selector set to 'PROJETO-FINAL', and a search bar. The left sidebar shows the 'Explorer' view with a list of resources, including 'Pergunta8' through 'Pergunta9', under the 'Datalake' folder. The main panel shows the 'Pergunta9' query editor with the following SQL code:

```
1 -- Pergunta de negócio 9:
2 -- Qual foi o dia que mais teve produtos vendidos?
3
4 SELECT
5   data,
6   SUM(unidade_venda) AS total_transacoes
7 FROM
8   `projeto-final-424922.Datalake.Tabela_original`
9 GROUP BY
10  data
11 ORDER BY
12  total_transacoes DESC
13 LIMIT
14  1;
```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	data	total_transacoes
1	2024-03-03	10

The interface also includes various toolbars for running, saving, and sharing queries, as well as tabs for job information, results, chart, JSON, execution details, and execution graph.

10)-- Pergunta de negócio 10:

-- Em qual continente foi mais usado o método de pagamento menos usado no geral?

```
WITH MetodoPagamentoContagem AS (  
    SELECT  
        metodo_pagamento,  
        COUNT(*) AS total_transacoes  
    FROM  
        `projeto-final-424922.Datalake.Tabela_original`  
    GROUP BY  
        metodo_pagamento  
) ,  
MenosUsado AS (  
    SELECT  
        metodo_pagamento  
    FROM  
        MetodoPagamentoContagem  
    ORDER BY  
        total_transacoes ASC  
    LIMIT 1  
)  
SELECT  
    regioao,  
    COUNT(*) AS total_transacoes  
FROM  
    `projeto-final-424922.Datalake.Tabela_original`  
WHERE  
    metodo_pagamento = (SELECT metodo_pagamento FROM MenosUsado)  
GROUP BY  
    regioao  
ORDER BY  
    total_transacoes DESC  
LIMIT 1;
```

#### Explicação:

- **1. Subquery MetodoPagamentoContagem:**
- **SELECT** metodo\_pagamento, **COUNT(\*) AS** total\_transacoes: Seleciona o método de pagamento e conta a quantidade de vezes que ele aparece.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Define a tabela de onde os dados serão extraídos.
- **GROUP BY** metodo\_pagamento: Agrupa os resultados pelo método de pagamento, permitindo a contagem.
- **2. Subquery MenosUsado:**
- **SELECT** metodo\_pagamento: Seleciona apenas a coluna do método de pagamento.
- **FROM** MetodoPagamentoContagem: Utiliza os resultados da subquery anterior.
- **ORDER BY** total\_transacoes **ASC**: Ordena os métodos de pagamento em ordem crescente de acordo com a quantidade de transações.
- **LIMIT 1**: Seleciona apenas o primeiro resultado, que representa o método de pagamento menos usado.

- **3. Query Principal:**
- **SELECT** regiao, COUNT(\*) AS total\_transacoes: Seleciona a região e conta o número de transações.
- **FROM** `projeto-final-424922.Datalake.Tabela\_original`: Especifica a tabela de dados.
- **WHERE** metodo\_pagamento = (SELECT metodo\_pagamento FROM MenosUsado): Filtra as transações para incluir apenas aquelas que utilizam o método de pagamento menos usado, encontrado na subquery MenosUsado.
- **GROUP BY** regiao: Agrupa os resultados por região.
- **ORDER BY** total\_transacoes DESC: Ordena as regiões em ordem decrescente de acordo com o número de transações.
- **LIMIT 1**: Seleciona a primeira região da lista, representando aquela onde o método de pagamento menos usado é mais frequente.

The screenshot displays the Google Cloud BigQuery interface. The top navigation bar includes the Google Cloud logo, a project dropdown set to 'PROJETO-FINAL', and a search bar. The left sidebar shows the 'Explorer' view with a list of resources, including 'Pergunta1' through 'Pergunta10' and 'Datalake'. The main panel shows a SQL query for 'Pergunta10' with the following code:

```

17  metodo_pagamento FROM MenosUsado)
18  ORDER BY
19  total_transacoes ASC
20  LIMIT 1
21  )
22  SELECT
23  regiao,
24  COUNT(*) AS total_transacoes
25  FROM
26  `projeto-final-424922.Datalake.Tabela_original`
27  WHERE
28  metodo_pagamento = (SELECT metodo_pagamento FROM MenosUsado)
29  GROUP BY
30  regiao
31  ORDER BY
32  total_transacoes DESC
33  LIMIT 1;

```

Below the query editor, the 'Query results' section is visible, showing a table with the following data:

Row	regiao	total_transacoes
1	Asia	40

The interface also includes various toolbars for running, saving, and sharing queries, as well as tabs for 'RESULTS', 'CHART', 'JSON', 'EXECUTION DETAILS', and 'EXECUTION GRAPH'.