

Resolução do Exercício da aula Data Warehouse

Janaína Scher de Oliveira

1- Projeto e dataset criados e upload de um conjunto de dados próprio realizado

2- Realize consultas no BigQuery para responder às seguintes perguntas

1- Projeto e dataset criados e upload de um conjunto de dados próprio realizado

The screenshot shows the Google Cloud BigQuery console interface. At the top, the Google Cloud logo is on the left, a dropdown menu shows 'engenharia-de-dados', and a search bar on the right says 'Pesquise (/) recursos, doc'. Below this is the 'Explorer' panel on the left, which contains a search bar 'Digite para pesquisar' and a list of resources. The resource 'aula_07_data_warehouse' is selected and highlighted. The main panel on the right displays the 'aula_07_data_warehouse' dataset details under the heading 'Informações do conjunto de dados'. The details include a table with various attributes and their values.

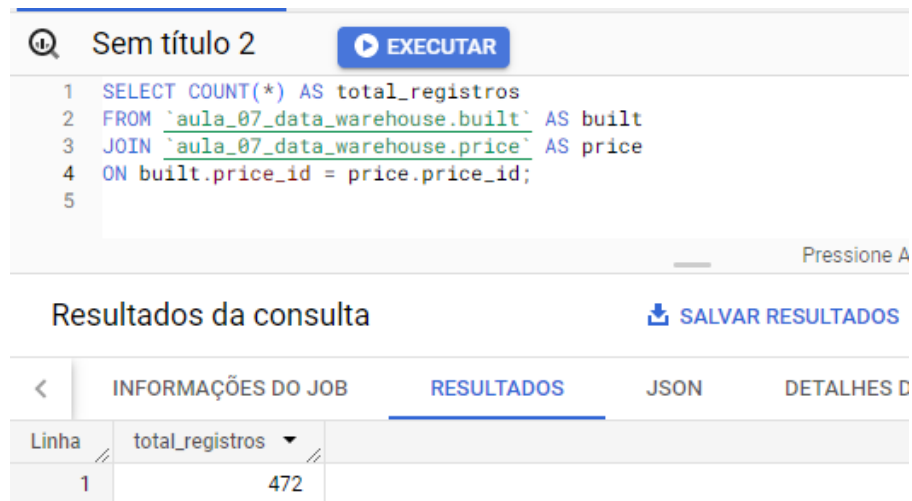
Informações do conjunto de dados	
ID do conjunto de dados	engenharia-de-dados-387814.aula_07_data_warehouse
Criado	25 de mai. de 2023, 13:55:49 UTC-3
Validade da tabela padrão	Nunca
Última modificação	25 de mai. de 2023, 13:55:49 UTC-3
Local dos dados	US
Descrição	
Compilação padrão	
Modo de arredondamento padrão	ROUNDING_MODE_UNSPECIFIED
Não diferencia maiúsculas e minúsculas	false
Rótulos	
Tags	

2- Realize consultas no BigQuery para responder às seguintes perguntas

Realizei as consultas abaixo juntando duas tabelas CSVs, a “built” e a “price”.

▼ Qual é a contagem total de registros na tabela?

```
SELECT COUNT(*) AS total_registros
FROM `aula_07_data_warehouse.built` AS built
JOIN `aula_07_data_warehouse.price` AS price
ON built.price_id = price.price_id;
```



The screenshot shows the BigQuery console. At the top, there's a search bar with "Sem título 2" and an "EXECUTAR" button. Below it, the SQL query is displayed in a text area. The results section shows a table with one row and one column, indicating 472 total records.

```
1 SELECT COUNT(*) AS total_registros
2 FROM `aula_07_data_warehouse.built` AS built
3 JOIN `aula_07_data_warehouse.price` AS price
4 ON built.price_id = price.price_id;
5
```

Resultados da consulta [SALVAR RESULTADOS](#)

Linha	total_registros
1	472

▼ Quais são as colunas presentes na tabela?

```
SELECT column_name
FROM `aula_07_data_warehouse.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name IN ('built', 'price')
```

Dez Primeiros...

EXECUTAR

```

1 SELECT column_name
2 FROM `aula_07_data_warehouse.INFORMATION_SCHEMA.COLUMNS`
3 WHERE table_name IN (('built', 'price'))

```

Pres

Resultados da consulta

SALVAR RESULT.

<

INFORMAÇÕES DO JOB

RESULTADOS

JSON

DET

Linha	column_name
1	price_id
2	year_built
3	location
4	price_id
5	bedroom
6	space
7	room
8	lot
9	tax
10	bathroom
11	garage
12	condition

▼ Faça uma consulta que retorne os 10 primeiros registros da tabela.

```

SELECT built.*, price.* EXCEPT (price_id)
FROM `aula_07_data_warehouse.built` AS built
JOIN `aula_07_data_warehouse.price` AS price
ON built.price_id = price.price_id
LIMIT 10

```

Sem título 2

EXECUTAR

SALVAR

COMPARTILHAR

PROGRAMAÇÃO

MAIS

```
1 SELECT built.*, price.* EXCEPT (price_id)
2   FROM `aula_07_data_warehouse.built` AS built
3   JOIN `aula_07_data_warehouse.price` AS price
4   ON built.price_id = price.price_id
5   LIMIT 10
6
```

Resultados da consulta

INFORMAÇÕES DO JOB

RESULTADOS

JSON

DETALHES DA EXECUÇÃO

GRÁFICO DE EXECUÇÃO

PRÉ-VISUALIZAÇÃO

Linha	price_id	year_built	location	bedroom	space	room	lot	tax	bathroom	garage	condition
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	53	1995	Chicago	2	967	5	39	652	1.5	0	0
3	53	1995	Chicago	2	636	6	30	553	1	2	1
4	53	1995	Chicago	3	673	6	30	589	1	2	1
5	53	1995	Chicago	3	716	6	30	585	1	2	1
6	60	1995	Chicago	2	782	5	25	834	1	0	0
7	60	1995	Chicago	2	828	5	35	913	1.5	1	0
8	70	1995	Chicago	3	2261	6	29	2700	1	2	0
9	70	1995	Chicago	3	2230	6	29	2727	1	2	0
10	70	1995	Chicago	2	2191	6	29	2688	1	2	0

▼ Execute uma consulta que agregue os dados de uma coluna numérica, fornecendo a média, o valor mínimo e o valor máximo.

```

SELECT
  AVG(CAST(price.price_id AS FLOAT64)) AS media,
  MIN(CAST(price.price_id AS FLOAT64)) AS minimum,
  MAX(CAST(price.price_id AS FLOAT64)) AS maximum
FROM `aula_07_data_warehouse.built` AS built
JOIN `aula_07_data_warehouse.price` AS price
ON built.price_id = price.price_id
WHERE price.price_id != 'NA' AND SAFE_CAST(price.price_id AS FLOAT64) IS NOT NULL

```

Como a coluna `price` é do tipo `STRING` em vez de `FLOAT64`, precisei converter a coluna para um tipo numérico antes de calcular a média, o valor mínimo e o valor máximo usando `CAST()`.

Sem título 2
EXECUTAR

```

1 SELECT
2   AVG(CAST(price.price_id AS FLOAT64)) AS media,
3   MIN(CAST(price.price_id AS FLOAT64)) AS minimum,
4   MAX(CAST(price.price_id AS FLOAT64)) AS maximum
5 FROM `aula_07_data_warehouse.built` AS built
6 JOIN `aula_07_data_warehouse.price` AS price
7 ON built.price_id = price.price_id
8  WHERE price.price_id != 'NA' AND SAFE_CAST(price.price_id AS FLOAT64) IS NOT NULL
9
10
11

```

Pressione Alt+F1 para consultar as opções de ace

Resultados da consulta
SALVAR RESULTADOS
EXPLORAR DADOS

<	INFORMAÇÕES DO JOB	RESULTADOS	JSON	DETALHES DA EXECUÇÃO	GR
Linha	media	minimum	maximum		
1	56.09787234042...	32.0	88.0		

▼ **Faça uma consulta que filtre os registros com base em uma condição específica.**

```

SELECT built.*, price.* EXCEPT (price_id)
FROM `aula_07_data_warehouse.built` AS built
JOIN `aula_07_data_warehouse.price` AS price
ON built.price_id = price.price_id
WHERE REGEXP_CONTAINS(price.price_id, r'^-[0-9]+(\.[0-9]+)?$') AND CAST(price.price_id AS FLOAT64) > 85

```

Como a coluna `price` é do tipo `STRING` em vez de `INT64`, e também contém valores não numéricos, como “NA”, precisei ajustar a consulta para lidar com esses valores não numéricos antes de aplicar a condição.

Sem título 2

EXECUTAR

SALVAR

COMPARTILHAR

PROGRAMAÇÃO

MAIS

```

1 SELECT built.*, price.* EXCEPT (price_id)
2 FROM `aula_07_data_warehouse.built` AS built
3 JOIN `aula_07_data_warehouse.price` AS price
4 ON built.price_id = price.price_id
5 WHERE REGEXP_CONTAINS(price.price_id, r'^-[0-9]+(\.[0-9]+)?$') AND CAST(price.price_id AS FLOAT64) > 85
6

```

Resultados da consulta

INFORMAÇÕES DO JOB

RESULTADOS

JSON

DETALHES DA EXECUÇÃO

GRÁFICO DE EXECUÇÃO

PRÉ-VISUALIZAÇÃO

Linha	price_id	year_built	location	bedroom	space	room	lot	tax	bathroom	garage	condition
1	88	2005	Chicago	8	2278	12	50	1183	3	2	0
2	88	2005	Chicago	8	2228	12	50	1208	3	2	0
3	88	2005	Chicago	7	2277	12	50	1248	3	2	0
4	88	2005	Chicago	8	2278	12	50	1183	3	2	0
5	88	2005	Chicago	8	2228	12	50	1208	3	2	0
6	88	2005	Chicago	7	2277	12	50	1248	3	2	0