

Case SQL - Media.Monks - Data Science PSEL - Gustavo Rocha Barbosa - 2024

Projeto em SQL Perguntas:

- 1) Qual foi a receita de cada tipo de pagamento no dia 15 de Março de 2018?
- 2) Considere que corridas de táxi válidas tenham de 1 a 5 passageiros. Qual a quantidade de corridas feitas com cada número de passageiros, faturamento médio de cada corrida e faturamento médio por passageiro?
- 3) Qual a hora que mais começaram corridas?
- 4) Considerando apenas as corridas que houveram pedágios (tolls) e que transportaram até 3 passageiros, qual a média do valor pago em pedágios por corrida?

Antes, mesmo sabendo que Não Deveria entrar no BQ, realizei as consultas por lá, pois minha expertise estava em SQL do PostgreSQL, sendo assim, tive que estudar a sintaxe adequada para o exercício e foi pertinente conectar os dados no BQ para o estudo.

Tipo do SQL: SQL Standard BigQuery

Nome da Tabela: tlc_yellow_trips_2018

Nome do Conjunto de Dados: testesql

Nome do Projeto: projeto-sql-423612

Respostas:

Número 1)

-- Primeiro jeito usando SUBQUERY - o Código fica mais fácil de ser entendido e daria para usar o HAVING também, mas optei por SUBQUERY:

```
SELECT
    ROUND(SUM(total_amount), 2) AS soma_receita,
    dia_mes_ano,
    payment_type
FROM (
    SELECT
        total_amount,
        FORMAT_DATE('%d/%m/%y', DATE(pickup_datetime)) AS dia_mes_ano,
        payment_type
    FROM
        `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
)
WHERE dia_mes_ano = '15/03/18'
GROUP BY payment_type, dia_mes_ano
LIMIT 6; -- Limite de 6 linhas, pois isso garante que aparece 6 tipos de pagamentos existentes!
```

-- Segundo jeito sem usar SUBQUERY:

```
SELECT
    ROUND(SUM(total_amount), 2) AS soma_receita,
    FORMAT_DATE('%d/%m/%y', DATE(pickup_datetime)) AS dia_mes_ano,
    payment_type
FROM
    `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
WHERE FORMAT_DATE('%d/%m/%y', DATE(pickup_datetime)) = '15/03/18'
GROUP BY payment_type, dia_mes_ano
LIMIT 6;
```

Aqui eu fiz de duas formas, portanto são duas Queries para responder a mesma pergunta!

Logo a resposta da Query é:

Resultados da consulta

| INFORMAÇÕES DO JOB | | RESULTADOS | GRÁFICO | JSON |
|--------------------|--------------|-------------|--------------|------|
| Linha | soma_receita | dia_mes_ano | payment_type | |
| 1 | 23.8 | 15/03/18 | 2 | |
| 2 | 635.61 | 15/03/18 | 1 | |

Número 2)

```
SELECT
  passenger_count,
  COUNT(passenger_count) AS contagem_corridas,
  ROUND(SUM(total_amount), 2) AS faturamento_medio_corrida,
  ROUND(SUM(total_amount)/COUNT(passenger_count), 2) AS
faturamento_medio_por_passageiro
FROM
  `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
WHERE passenger_count >= 1 AND passenger_count <= 5
GROUP BY passenger_count
ORDER BY passenger_count
LIMIT 5;
```

Resultado:

Resultados da consulta

| INFORMAÇÕES DO JOB | | RESULTADOS | GRÁFICO | JSON | DETALHES DA EXECUÇÃO |
|--------------------|-----------------|-------------------|-------------------|-------------------|----------------------|
| Linha | passenger_count | contagem_corridas | faturamento_medio | faturamento_medio | |
| 1 | 1 | 716 | 28519.28 | 39.83 | |
| 2 | 2 | 151 | 5957.23 | 39.45 | |
| 3 | 3 | 32 | 1108.89 | 34.65 | |
| 4 | 4 | 25 | 918.51 | 36.74 | |
| 5 | 5 | 50 | 1898.75 | 37.97 | |

Número 3)

```
SELECT
  FORMAT_DATETIME('%H', DATETIME(pickup_datetime)) AS hora,
  COUNT(pickup_datetime) AS contagem_horas
FROM
  `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
GROUP BY hora
ORDER BY contagem_horas DESC
LIMIT 1;
```

Resultado:

Resultados da consulta

| INFORMAÇÕES DO JOB | | RESULTADOS | GRÁFICO | JSON |
|--------------------|------|----------------|---------|------|
| Linha | hora | contagem_horas | | |
| 1 | 22 | 70 | | |

Número 4)

-- Primeira QUERY com um tipo de interpretação do enunciado:

```
SELECT
  ROUND(SUM(tolls_amount), 2) AS soma_receita,
  COUNT(tolls_amount) AS contagem_pedagios,
  ROUND(SUM(tolls_amount)/COUNT(tolls_amount), 2) AS media_valor_pedagios
FROM
  `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
```

-- Minha interpretação: Até 3 passageiros inclui a condição de 0 passageiros também, por isso coloquei <= 3 (no WHERE abaixo) para o número de passageiros.

```
WHERE passenger_count <= 3 AND tolls_amount > 0
LIMIT 1;
```

-- Segunda QUERY, pois o enunciado pode ter duas interpretações: considerando 0 passageiros e considerando 1 passageiro como mínimo:

```
SELECT
  ROUND(SUM(tolls_amount), 2) AS soma_receita,
  COUNT(tolls_amount) AS contagem_pedagios,
  ROUND(SUM(tolls_amount)/COUNT(tolls_amount), 2) AS media_valor_pedagios
FROM
  `projeto-sql-423612.testesql.tlc_yellow_trips_2018`
-- Considerando agora que a condição seja de ter pelo menos 1 passageiro até 3 passageiros.
WHERE passenger_count >= 1 AND passenger_count <= 3 AND tolls_amount > 0
LIMIT 1;
```

Aqui eu fiz duas queries, pois o enunciado pode ter ambiguidade e explico no código essas ambiguidades!

Resultado da primeira query da questão 4):

Resultados da consulta

| INFORMAÇÕES DO JOB | | RESULTADOS | GRÁFICO | JSON |
|--------------------|----------------|-------------------|----------------------|------|
| Linha | soma_receita ▼ | contagem_pedagios | media_valor_pedagico | |
| 1 | 1804.88 | 268 | 6.73 | |

Resultado da segunda query da questão 4):

Resultados da consulta

| INFORMAÇÕES DO JOB | | RESULTADOS | GRÁFICO | JSON |
|--------------------|----------------|-------------------|----------------------|------|
| Linha | soma_receita ▼ | contagem_pedagios | media_valor_pedagico | |
| 1 | 1787.6 | 265 | 6.75 | |