SQL practical test

date 17/06/2022

Klayton Crul Correa

E-mail: klayton12341@live.com

Linkedin: https://www.linkedin.com/in/klayton-crul

Github: https://github.com/klaytoncrul/data-science-projects

Telefone: (45)998023698

Objetive:

Monk Racoon data Science internship program

Database:

tlc_yellow_trips_2018

Data source:https://docs.google.com/spreadsheets/d/18V9jsGMn1KSb7vsEN4Fh0CQNIKweE5ctFrmGDEeWG_k/edit#gid=460814509

Proposed exercise:

Using the Google BigQuery SQL documentation and basing on the schema of the table above and the data present in the “TABLE” tab of the provided worksheet, create the SQL queries corresponding to the following questions:

Questions:

1) What was the revenue of each type of payment on March 15, 2018?

2) Assume valid taxi rides have 1 to 5 passengers. How many rides are made with each number of passengers?

3) Considering only the races that had tolls (tolls), what is the average amount paid in tolls per race?

4) What time did most races start?

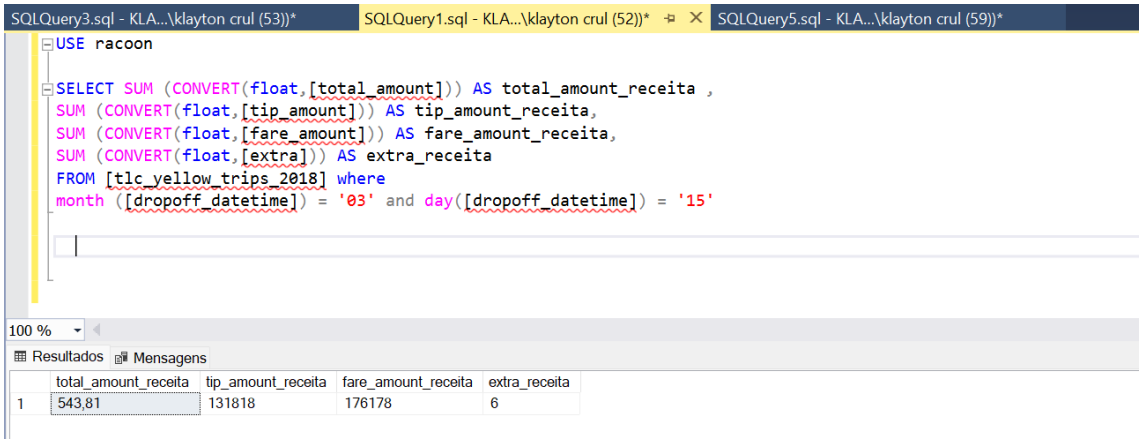
1) What was the revenue of each type of payment on March 15, 2018?total_amount_receita foi de **543,81**

tip_amount_receita was of **131818**
fare_amount_receita was of **176818**
extra_receita was of **6**

Code:

```
USE racoon

SELECT SUM (CONVERT(float,[total_amount])) AS total_amount_receita ,
SUM (CONVERT(float,[tip_amount])) AS tip_amount_receita,
SUM (CONVERT(float,[fare_amount])) AS fare_amount_receita,
SUM (CONVERT(float,[extra])) AS extra_receita
FROM [tlc_yellow_trips_2018] where
month ([dropoff_datetime]) = '03' and day([dropoff_datetime]) = '15'
```



2) Assume valid taxi rides have 1 to 5 passengers. How many rides are made with each number of passengers?

Number of rides per passenger:

1 passenger = **1590**

2 passengers = **795**

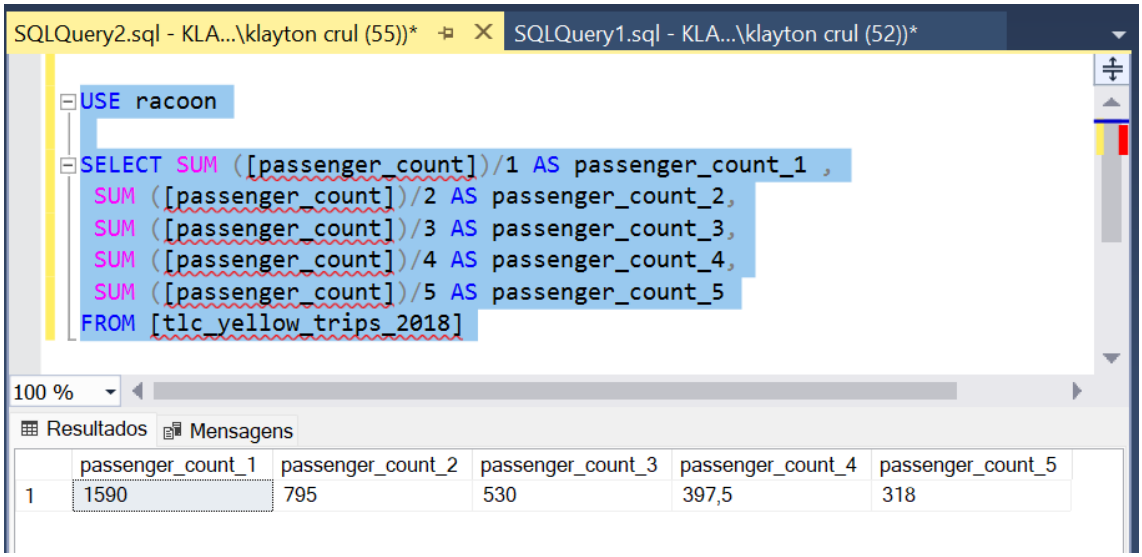
3 passengers = **530**

4 passengers = **397.5**

5 passengers = **318**

Code:

```
USE racoon
SELECT SUM ([passenger_count])/1 AS passenger_count_1 ,
 SUM ([passenger_count])/2 AS passenger_count_2,
 SUM ([passenger_count])/3 AS passenger_count_3,
 SUM ([passenger_count])/4 AS passenger_count_4,
 SUM ([passenger_count])/5 AS passenger_count_5
FROM [tlc_yellow_trips_2018]
```

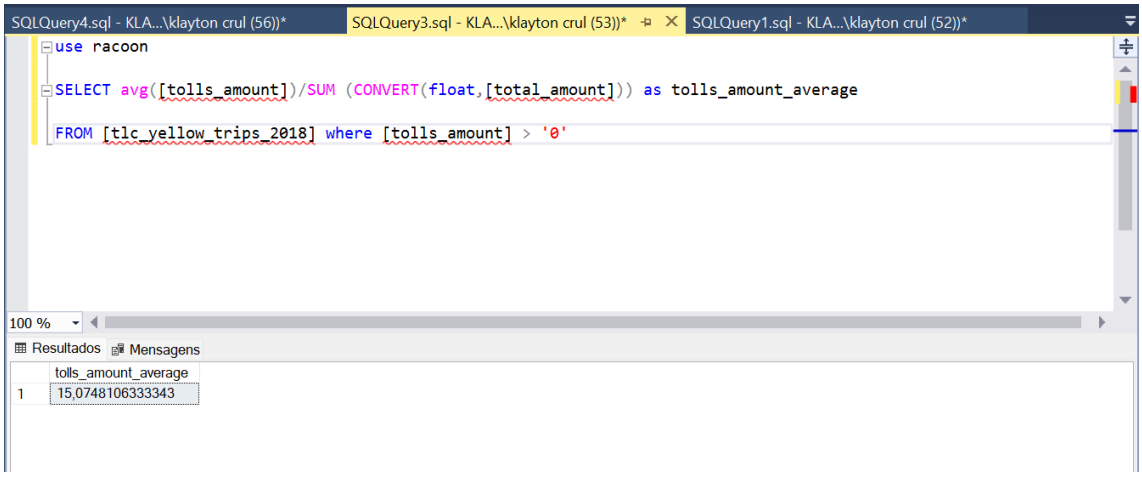


3) Considering only the races that had tolls (tolls), what is the average amount paid in tolls per race?

The average amount paid in tolls for rides is **15.0748106333343**

Code:

```
use racoon
SELECT avg([tolls_amount])/SUM (CONVERT(float,[total_amount])) as
tolls_amount_average
FROM [tlc_yellow_trips_2018] where [tolls_amount] > '0'
```

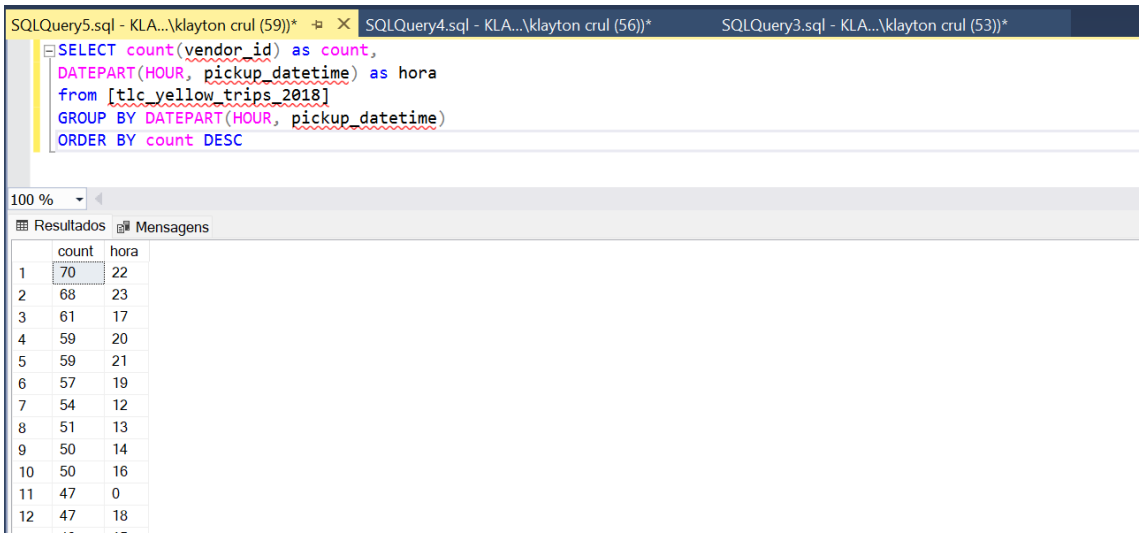


4) What time did most races start?

According to the data, the time when most races take place is **22:00** with **70** races.

Code:

```
SELECT count(vendor_id) as count,
DATEPART(HOUR, pickup_datetime) as hora
from [tlc_yellow_trips_2018]
GROUP BY DATEPART(HOUR, pickup_datetime)
ORDER BY count DESC
```



	count	hora
1	70	22
2	68	23
3	61	17
4	59	20
5	59	21
6	57	19
7	54	12
8	51	13
9	50	14
10	50	16
11	47	0
12	47	18