**SQL practical test**

**date 17/06/2022**

**Klayton Crul Correa**

E-mail: [klayton12341@live.com](mailto: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 ,

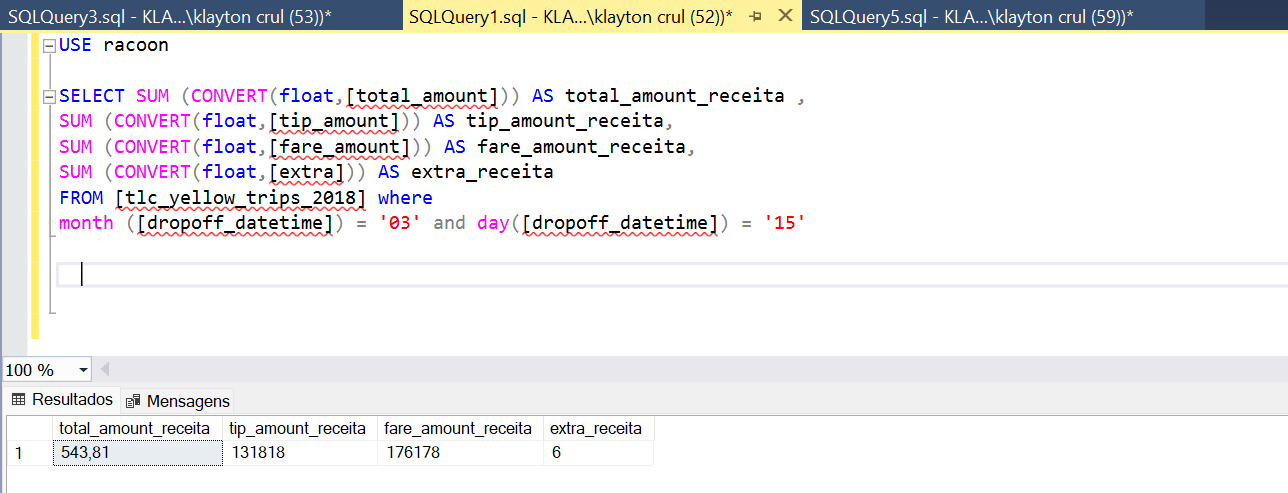
SUM (CONVERT(float,[tip\_amount])) AS tip\_amount\_receita,

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 ,

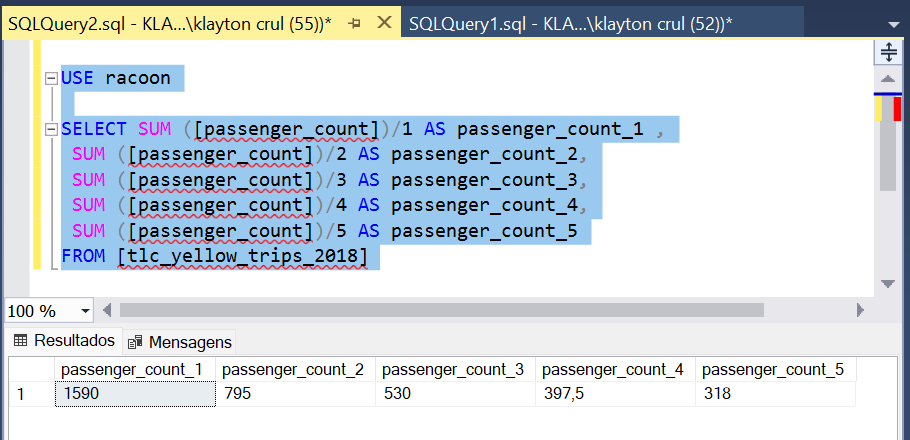
SUM ([passenger\_count])/2 AS passenger\_count\_2,

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?**

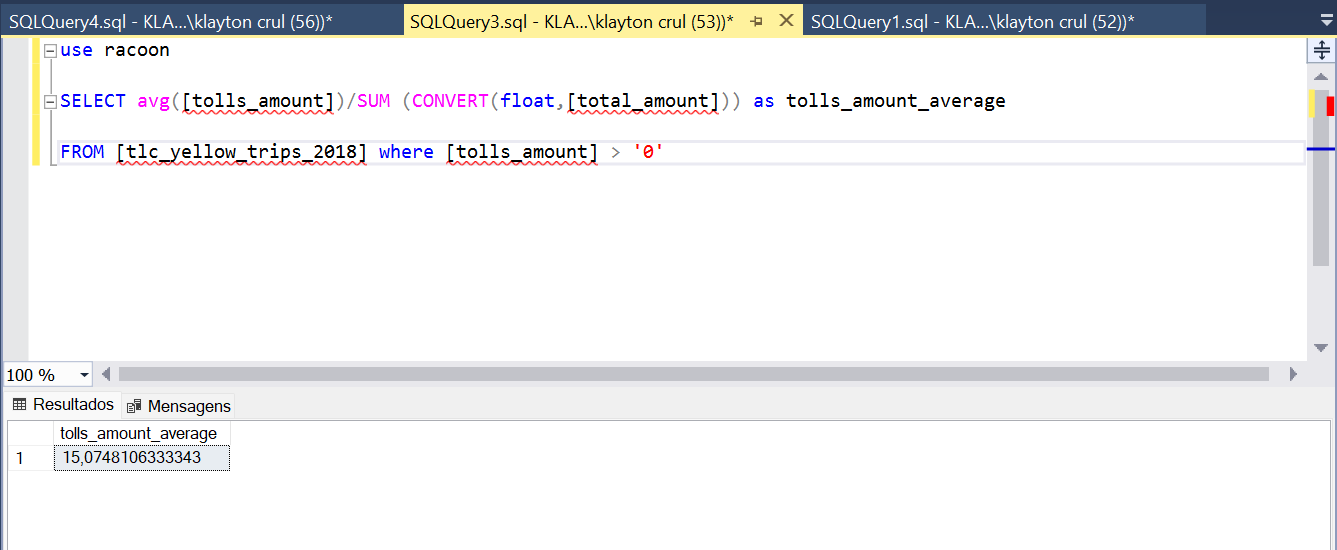
The average amount paid in tolls for rides is **15.0748106333343**

**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

