**Project 1 (Big-Data)**

**Querying using Hive on Yellow Taxi data**

[**https://www.kaggle.com/datasets/elemento/nyc-yellow-taxi-trip-data**](https://www.kaggle.com/datasets/elemento/nyc-yellow-taxi-trip-data)

**Problem statement:**

In this case study, we are giving a real-world example of how to use HIVE on top of the HADOOP for different exploratory data analysis. In here, we have a predefined dataset (2018\_Yellow\_Taxi\_Trip\_Data.csv) having more than 15 columns and more than 100000 records in it. The dataset has different attributes like

1. vendor\_id string,
2. pickup\_datetime string,
3. dropoff\_datetime string,
4. passenger\_count int,
5. trip\_distance DECIMAL(9,6),
6. pickup\_longitude DECIMAL(9,6),
7. pickup\_latitude DECIMAL(9,6),
8. rate\_code int,
9. store\_and\_fwd\_flag string,
10. dropoff\_longitude DECIMAL(9,6),
11. dropoff\_latitude DECIMAL(9,6),
12. payment\_type string,
13. fare\_amount DECIMAL(9,6),
14. extra DECIMAL(9,6),
15. mta\_tax DECIMAL(9,6),
16. tip\_amount DECIMAL(9,6),
17. tolls\_amount DECIMAL(9,6),
18. total\_amount DECIMAL(9,6),
19. trip\_time\_in\_secs int

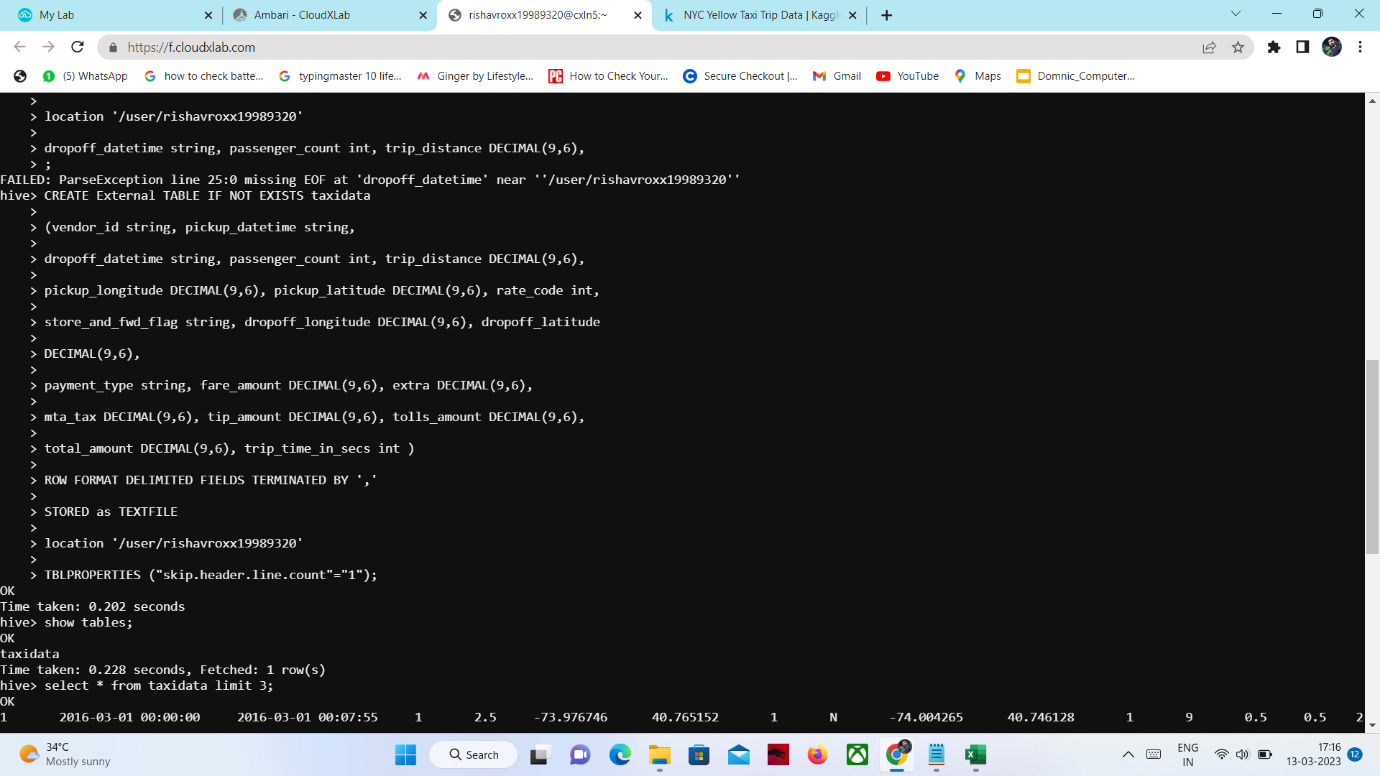
**Perform taxi trip analysis by solving the questions below:**

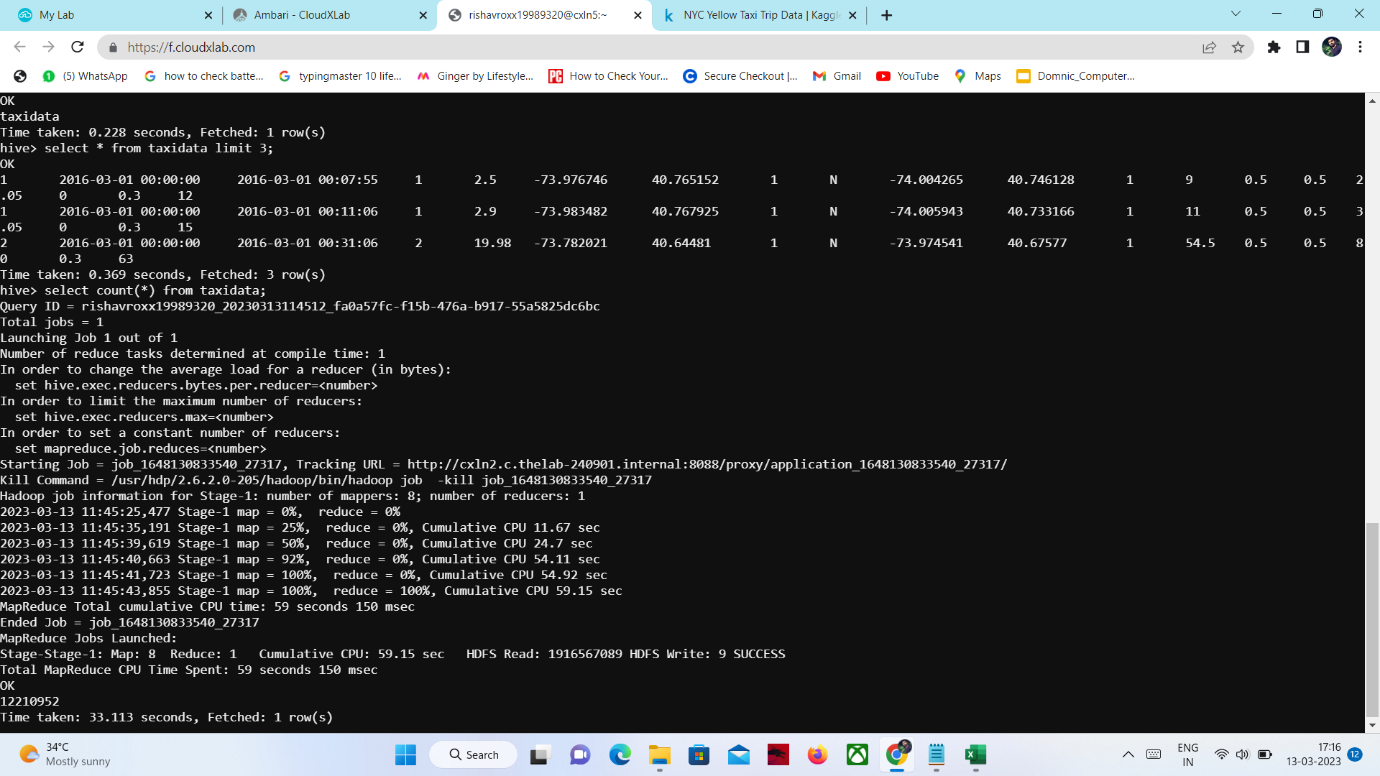
1. What is the total Number of trips ( equal to the number of rows)?
2. What is the total revenue generated by all the trips? The fare is stored in the column total\_amount.
3. What fraction of the total is paid for tolls? The toll is stored in tolls\_amount.
4. What fraction of it is driver tips? The tip is stored in tip\_amount.
5. What is the average trip amount?
6. What is the average distance of the trips? Distance is stored in the column trip\_distance.
7. How many different payment types are used?
8. For each payment type, display the following details:

* Average fare generated
* Average tip
* Average tax – tax is stored in column mta\_tax

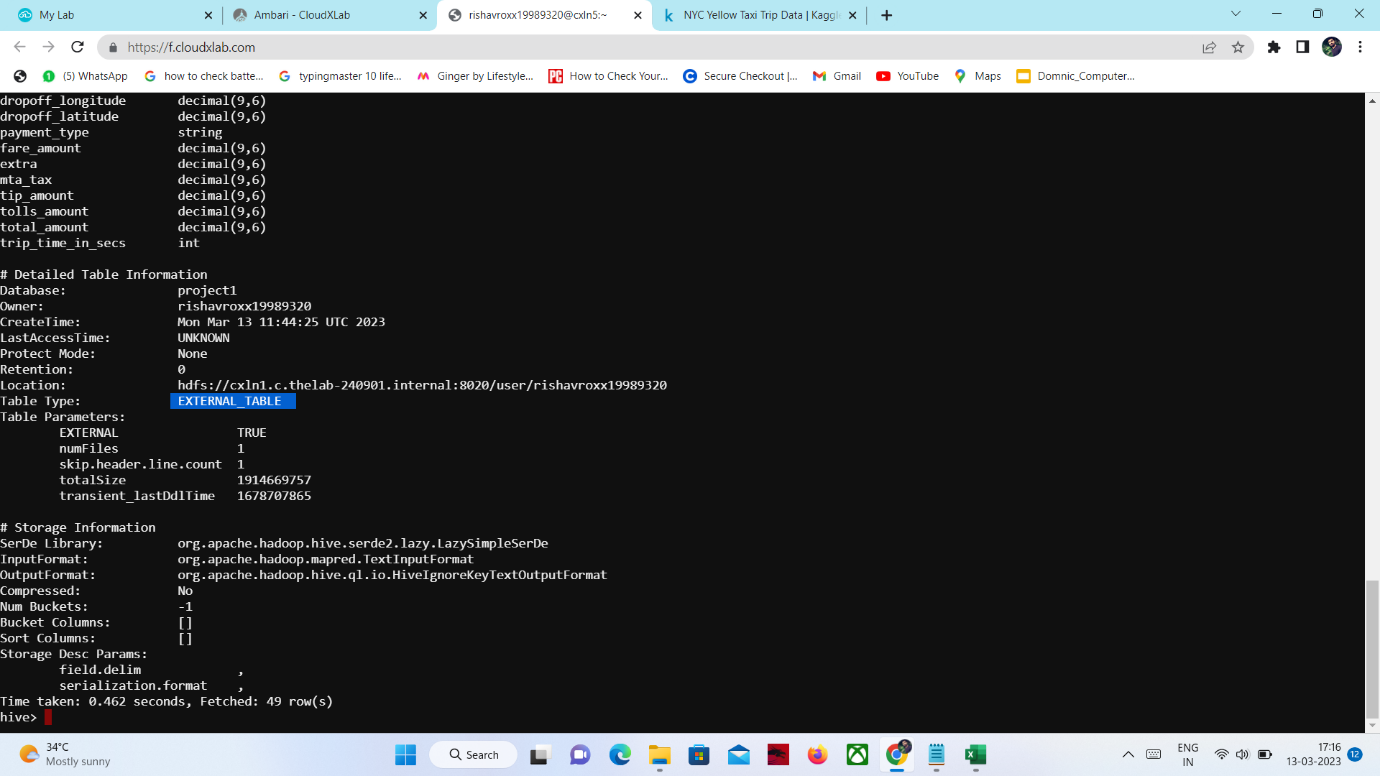
1. On average which hour of the day generates the highest revenue?

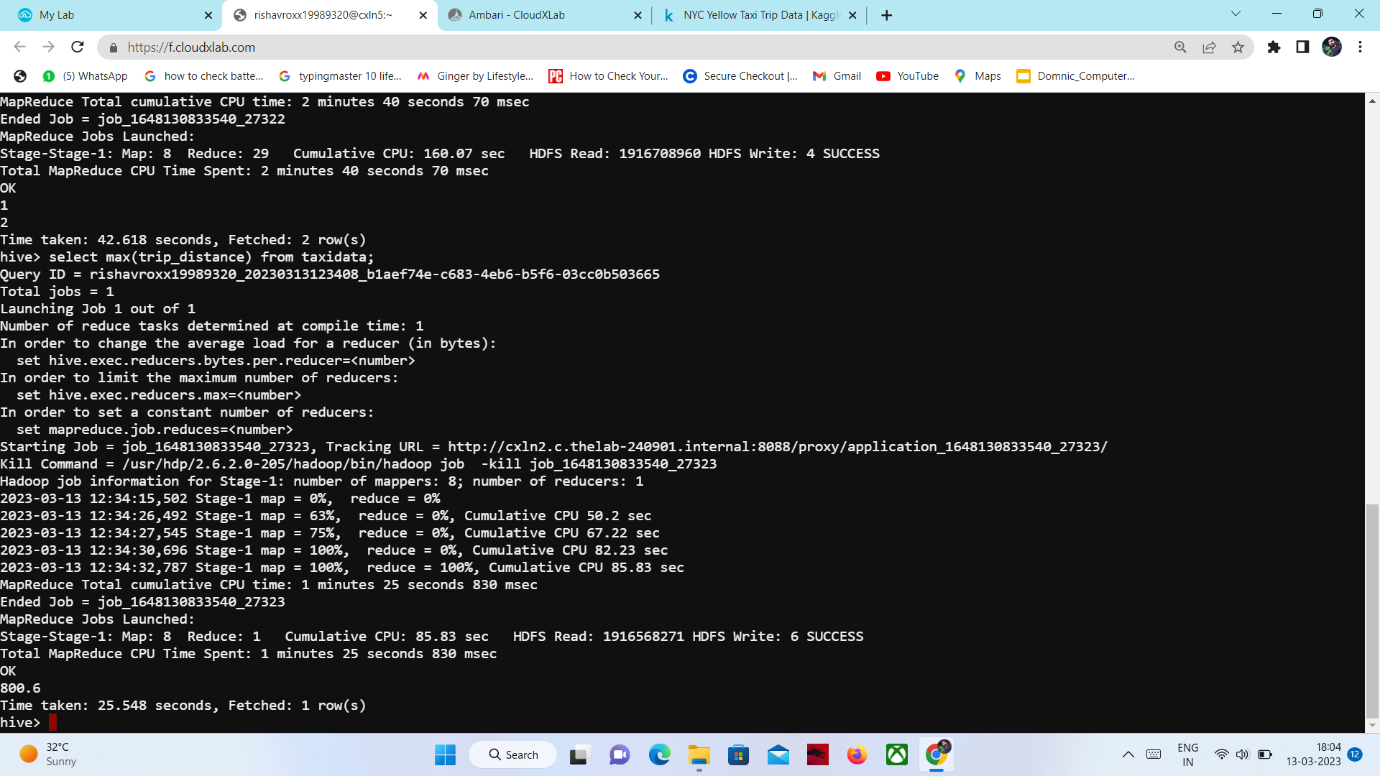
Q1) Creating table:-



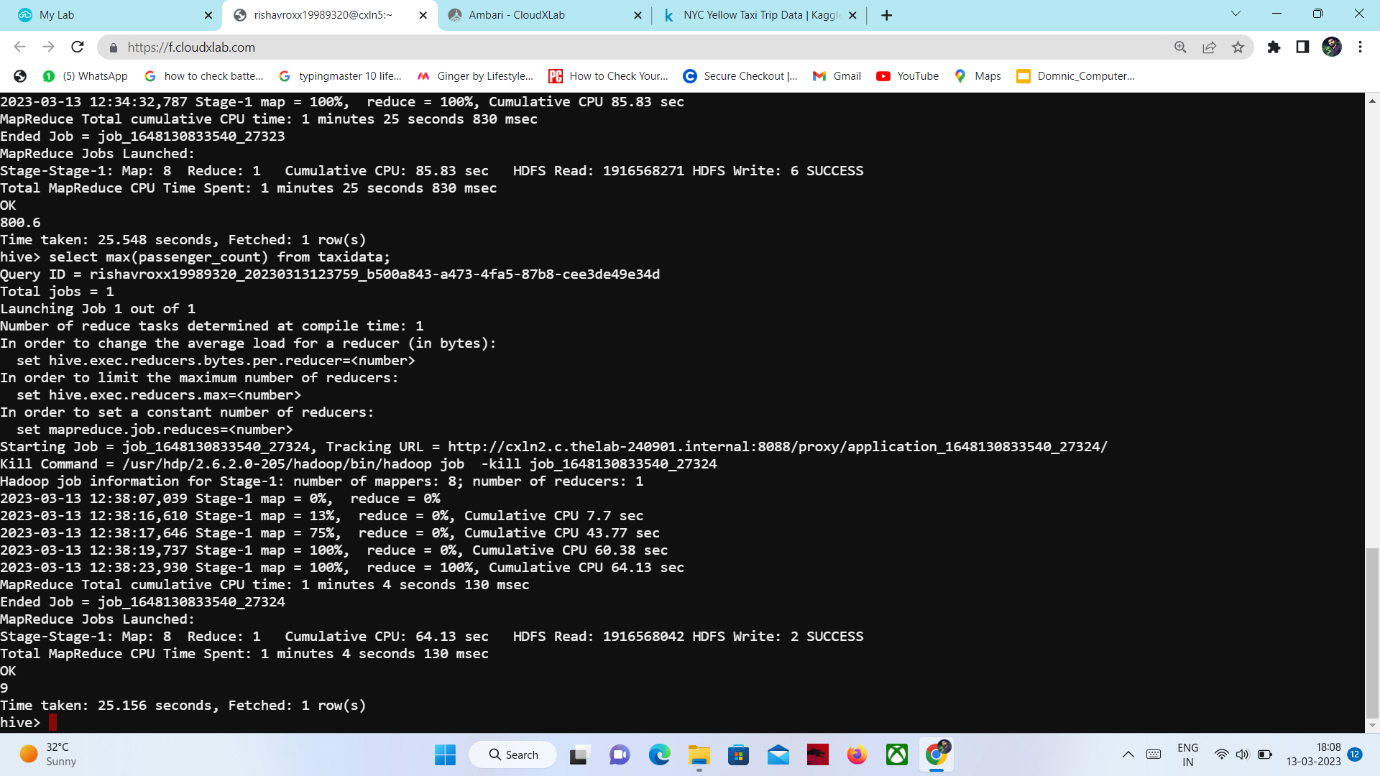
Q.2) Finding no of Records:

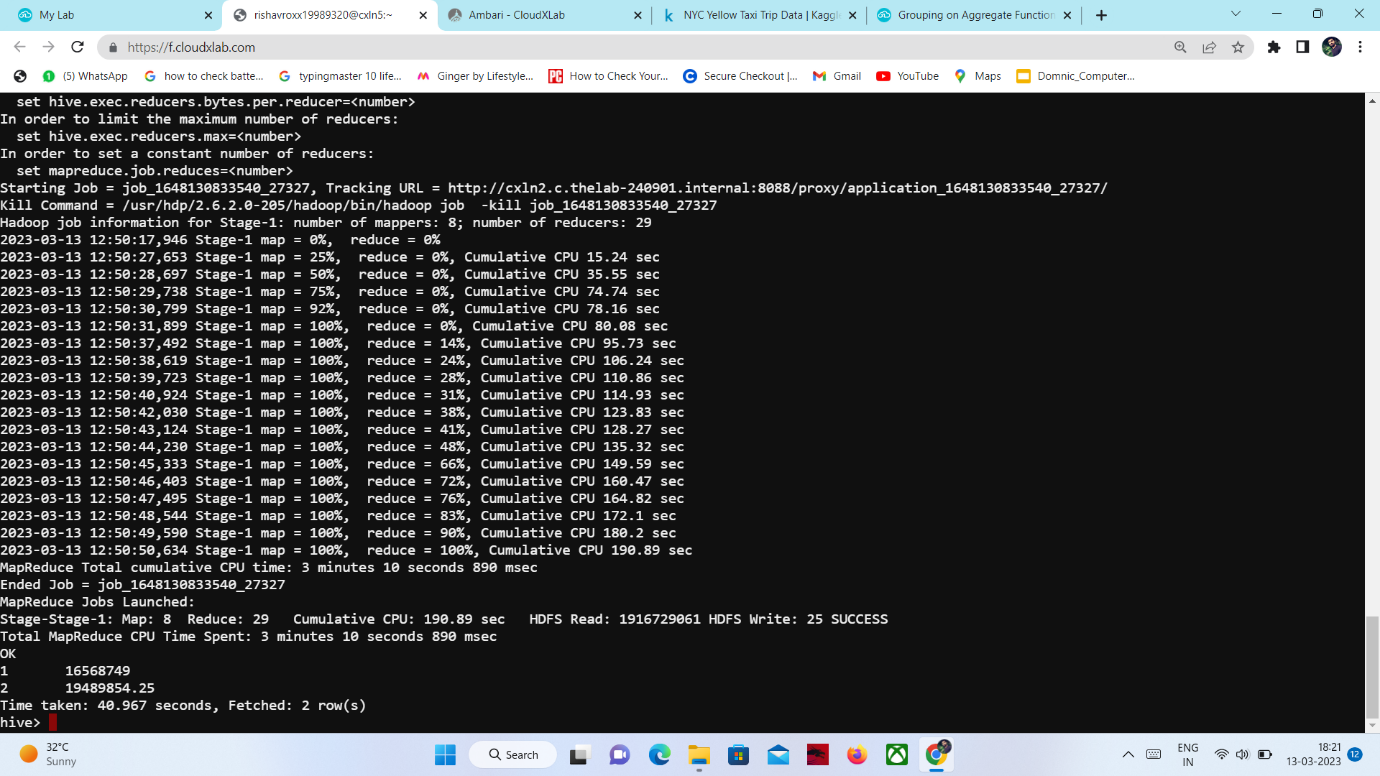
Q3. Info about Table: -



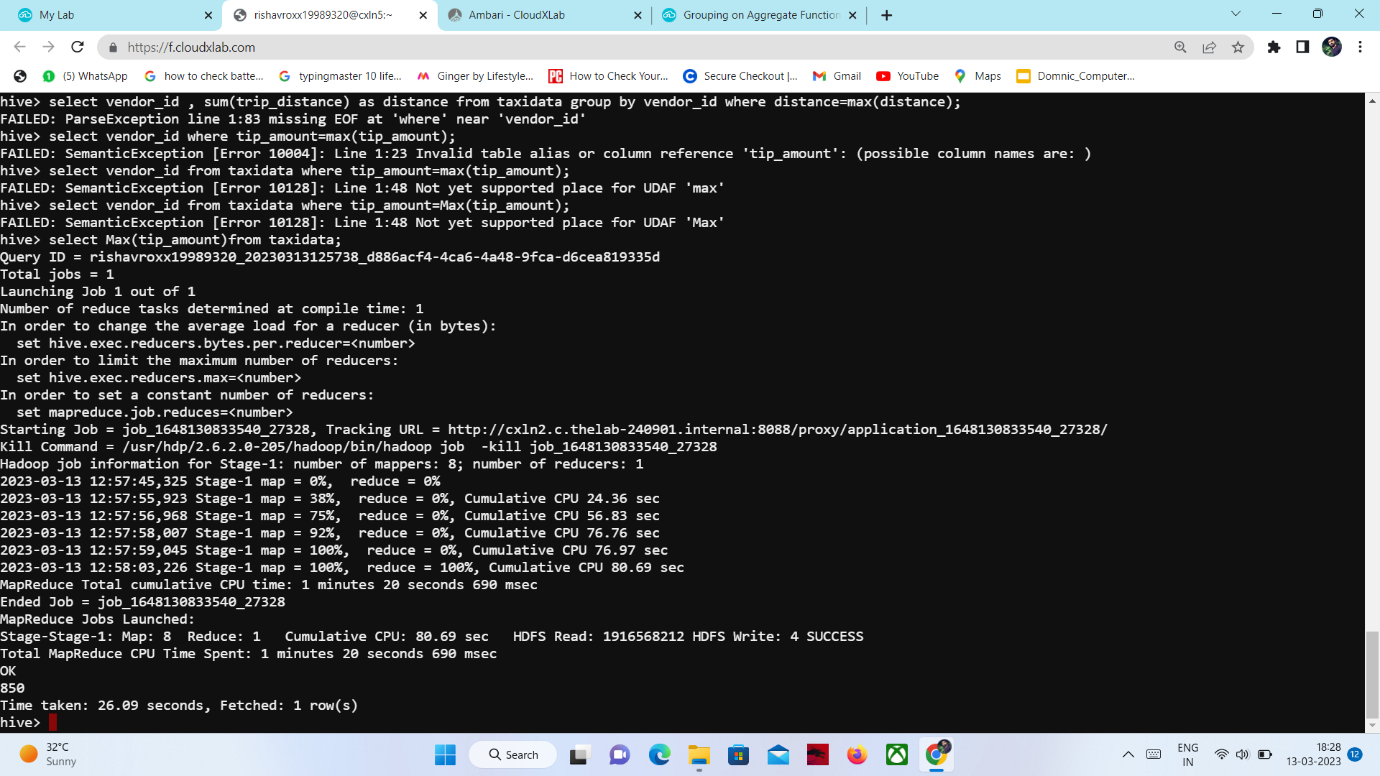
Q4) Listing the Distinct Vendors: -

Q5) Maximum Passengers hold by a vendor:-

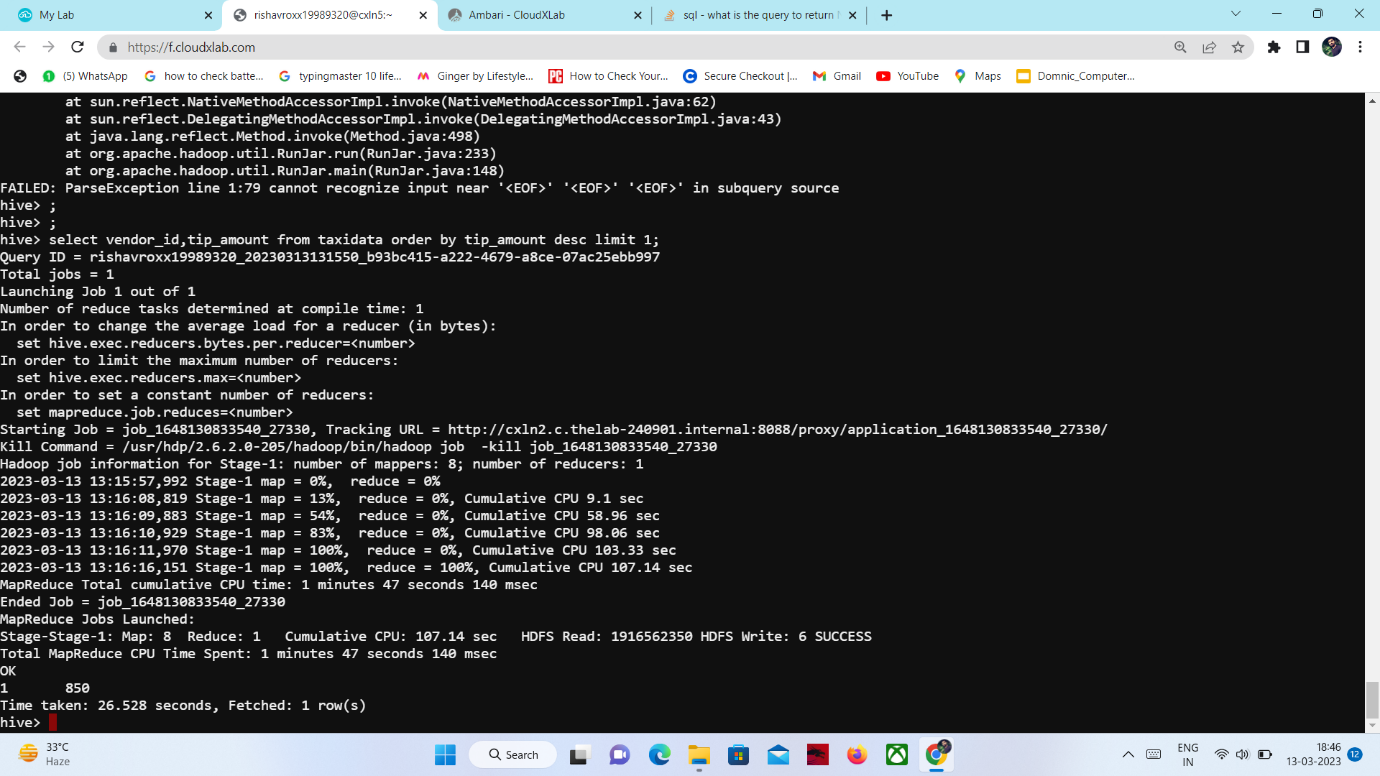


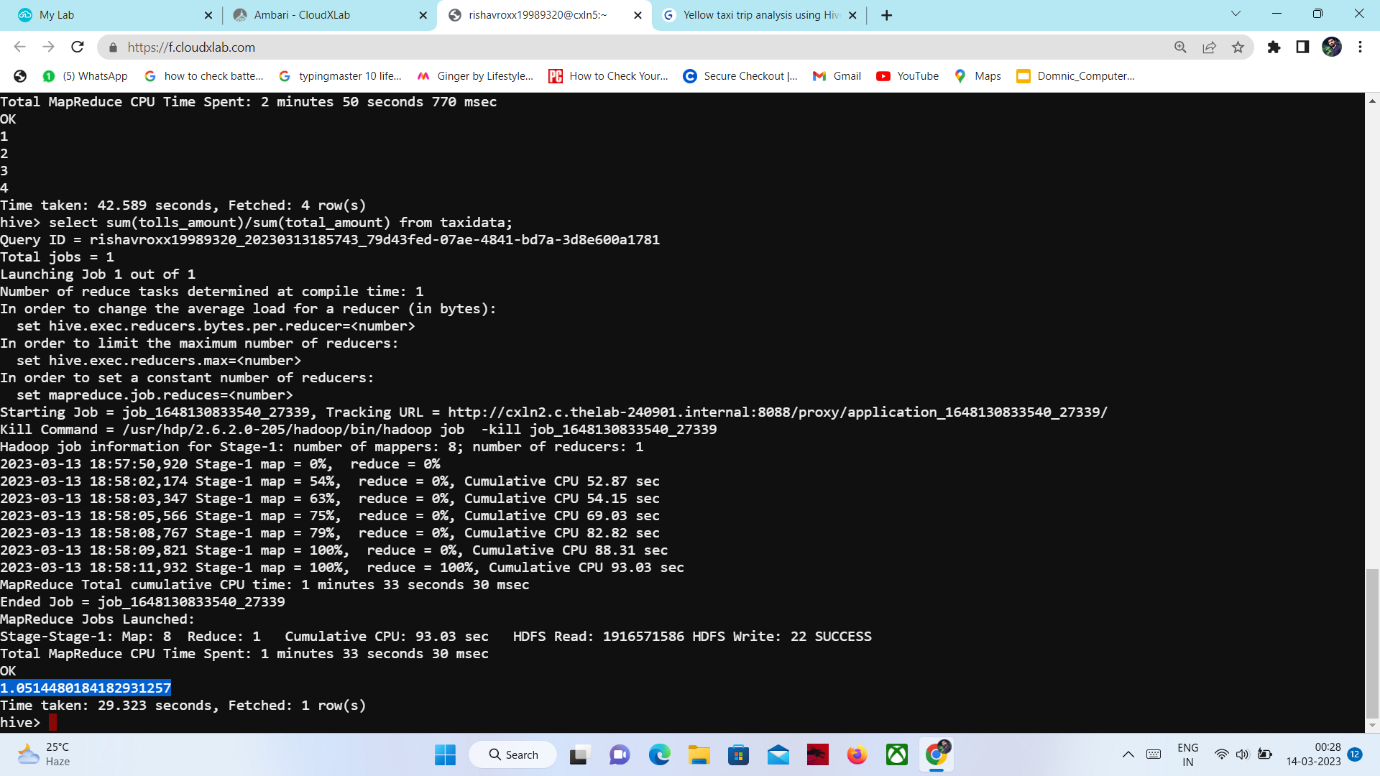
Q6) How many times each vendor carry their passengers: -

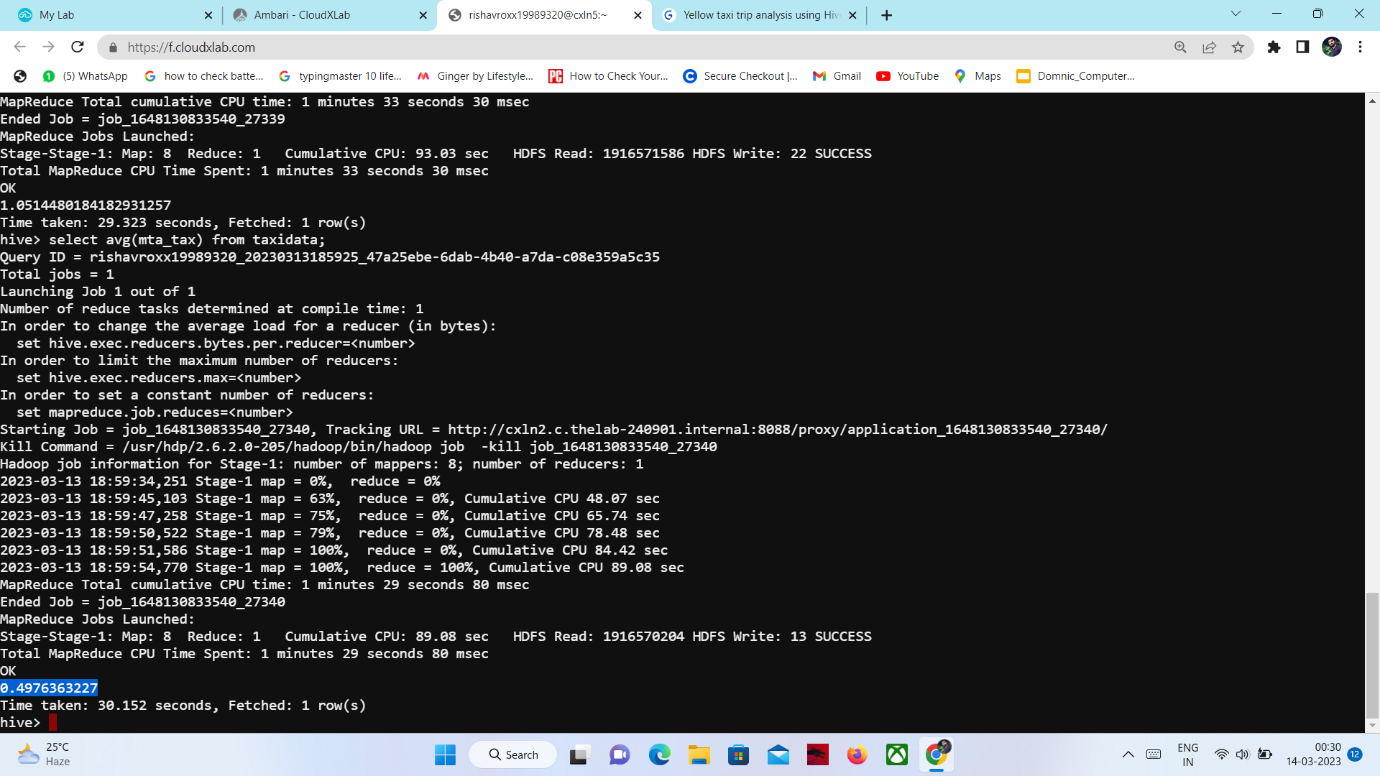
Q7) Maximum tip to the vendor ever: -

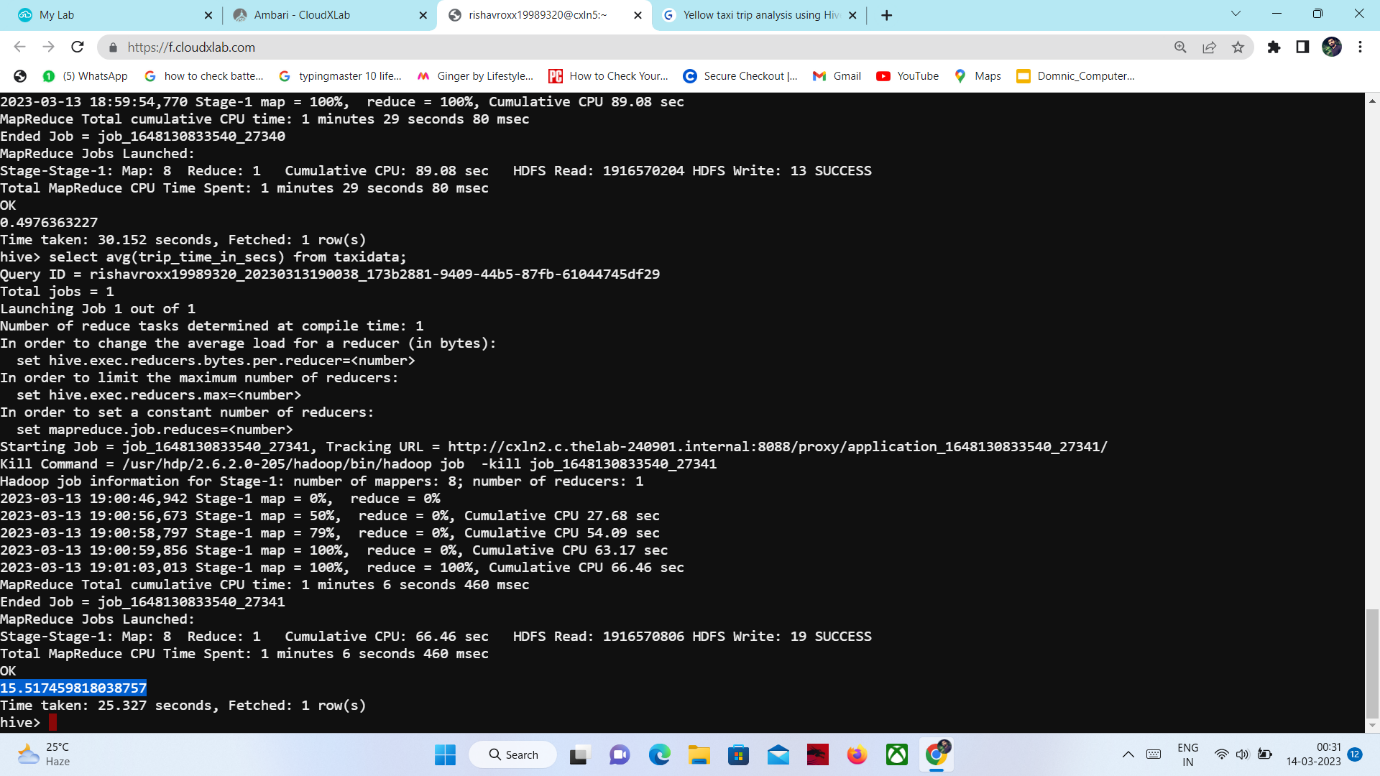
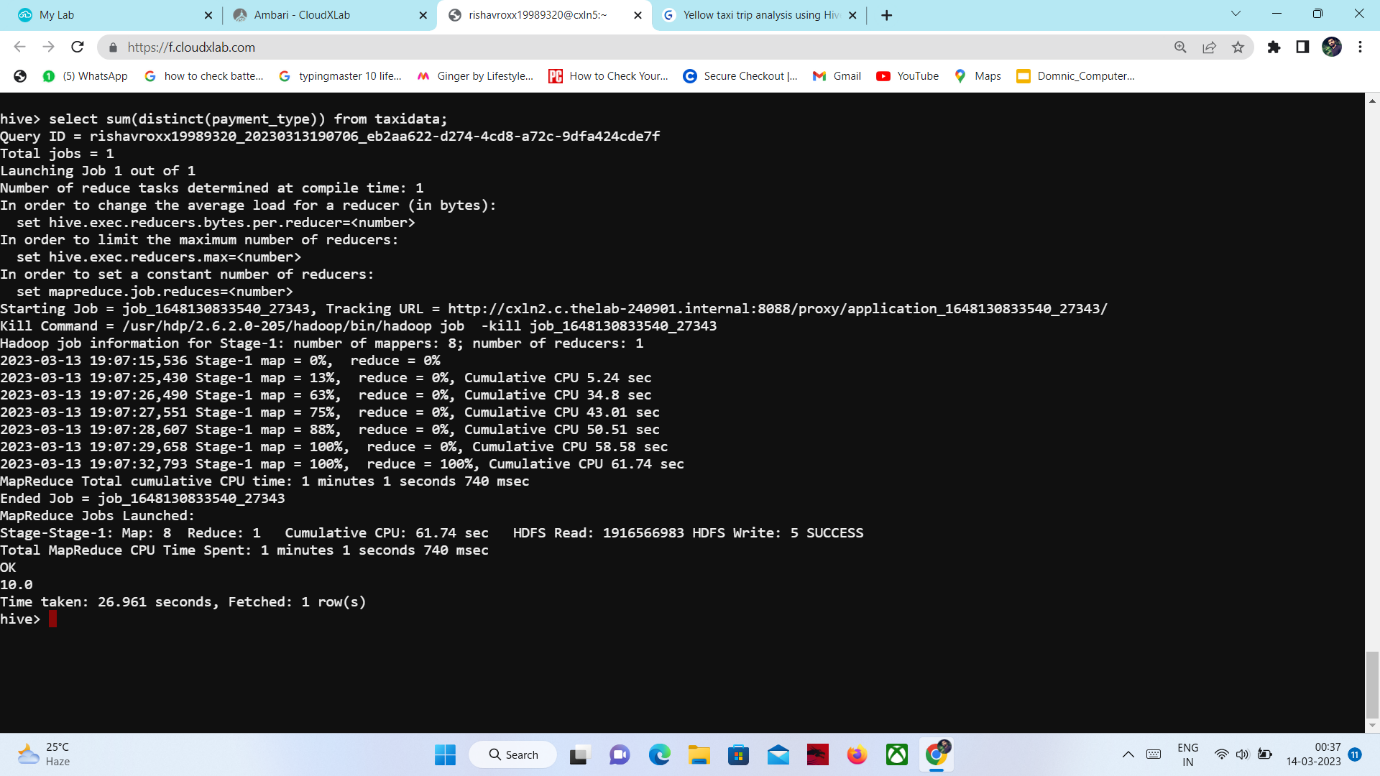


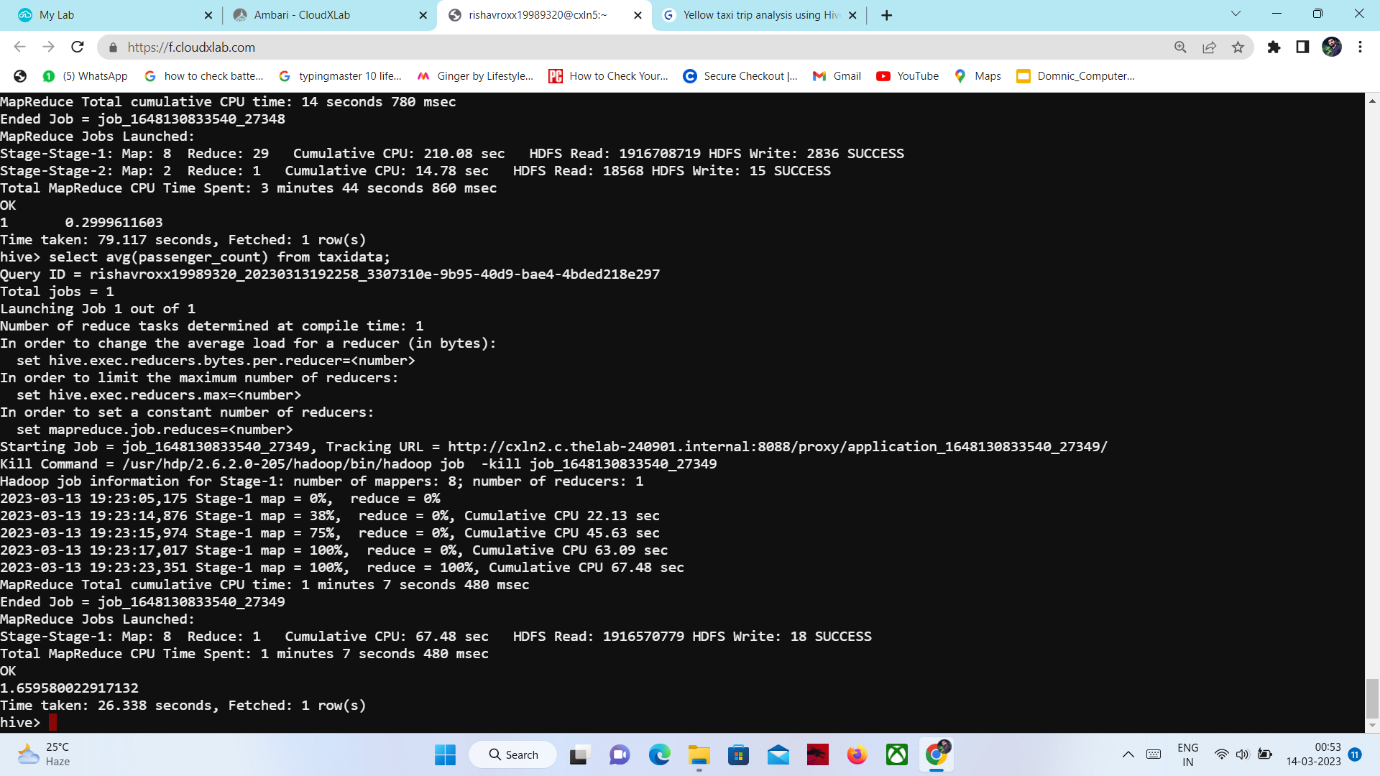
Q8) which vendor have been given maximum tip and how much:-

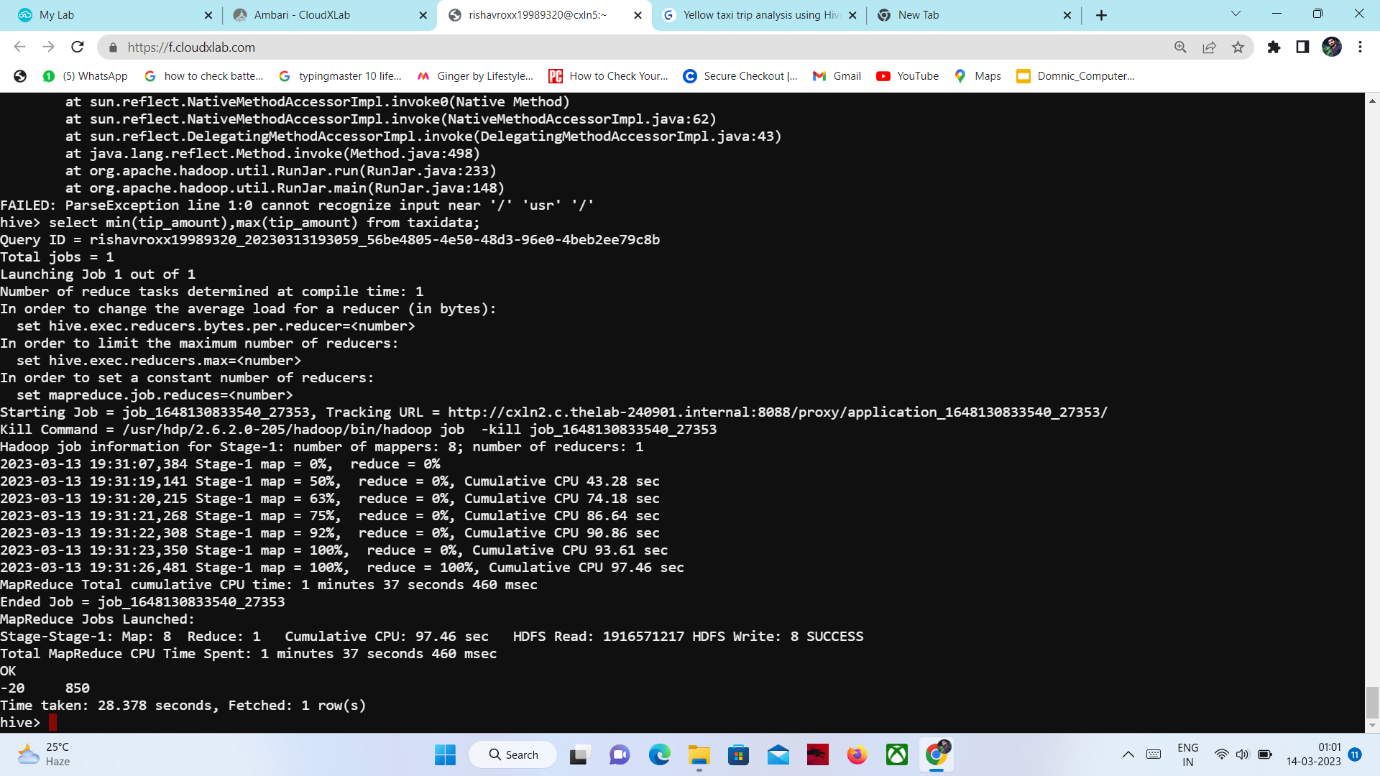


Q9) Fraction to tolls amount w.r.t to total amount:-

Q10) On an average how much mta\_tax has been given:-

Q11) Average trip time :-Q12) How many different payment types are used:-

Q13) Average passenger count:

Q14) Maximum time spent by passenger:-

Q15)Distinct Rate Code: