PROBLEM STATEMENT

Q1 How many unique passengers uptill now have used the taxi service?

Q2 What is the average trip distance?

Q3 How many trips were paid by cash? [1.Credit card 2.Cash 3.No charge 4.Dispute 5.Unknown 6.Voided trip]

Q4 What is the longest trip that was paid by credit card?

Q5 What is the total fare amount (fare\_amount) collected from all trips?

Q6 How many trips were stored and forwarded (store\_and\_fwd\_flag)?

Q7 What is the total amount collected from all trips that were not stored and forwarded?

Q8 What is the average trip distance for each unique number of passengers?

Q9 Using a window function, rank the payment\_types based on the fare\_amount in descending order.

Q10 Calculate the cumulative sum of fare\_amount for each payment\_type.

Q11 Find the average trip\_distance for each payment\_type, and then rank these averages.

Q12 Classify payment types into ‘Electronic’ (Credit card) and ‘Non-Electronic’ (Cash, No charge, Dispute, Unknown, Voided trip) and find the total number of trips for each category.

Q13 How does the average trip distance vary by the hour of the day?

Q14 What is the average fare amount for each day of the week?

Q15 What is the total fare amount for each month of the year?

Q16 What is the 90th percentile of trip distance for each passenger count?

Q17 What is the average number of passengers for trips with a fare amount above the overall average fare amount?

Q18 What is the average fare amount for each month of the year, only considering trips that are shorter than the average trip distance?

Q19 Find the top 5 pickup locations (latitude and longitude) with the highest average fare amount. Exclude trips with zero passengers?

Q20 Write a SQL query to find the difference in total\_amount between each trip and its preceding trip for each payment\_type.

Q21 Write a SQL query to find the maximum fare\_amount for each hour of the day. Assume that the tpep\_pickup\_datetime is stored in the format ‘YYYY-MM-DD HH:MI:SS’.

Q22 Write a SQL query to find all taxi trips where the absolute difference between pickup\_longitude and dropoff\_longitude is greater than 1. Use the ABS function to calculate the absolute difference.

Q23 Write a SQL query to find the average total\_amount for each passenger\_count, but only include those passenger\_count groups having more than 100 trips.?