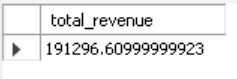
**Taxi Fare Prediction Analysis MySQL QUERY DOCUMENT**

**Taxi Fare Prediction | SUMMARY : BY Shivanand S Nashi**

**1.Total revenue generated:**

SELECT SUM(amount) AS total\_revenue FROM taxifare;



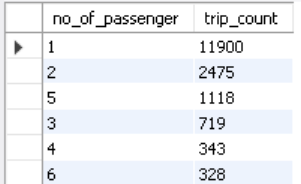
**2.Number of trips per passenger count:**

SELECT no\_of\_passenger, COUNT(\*) AS trip\_count

FROM taxifare

GROUP BY no\_of\_passenger

ORDER BY trip\_count DESC;



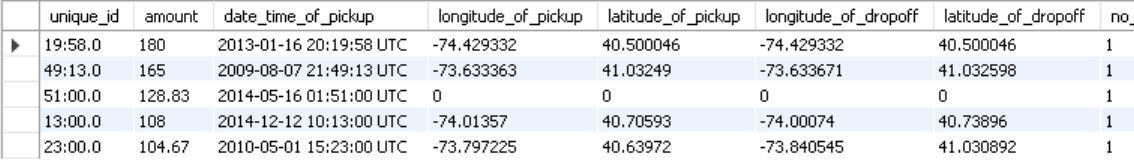
**3.Trips with the highest fare amount:**

SELECT \*

FROM taxifare

ORDER BY amount DESC

LIMIT 10;



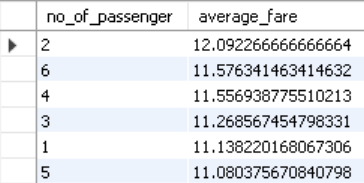
**4.Average fare amount by passenger count:**

SELECT no\_of\_passenger, AVG(amount) AS average\_fare

FROM taxifare

GROUP BY no\_of\_passenger

ORDER BY average\_fare DESC;



**5.Trips occurring during peak hours (7 AM - 9 AM and 5 PM - 7 PM):**

SELECT COUNT(\*) AS peak\_hour\_trips

FROM taxifare

WHERE (HOUR(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) BETWEEN 7 AND 9)

   OR (HOUR(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) BETWEEN 17 AND 19);

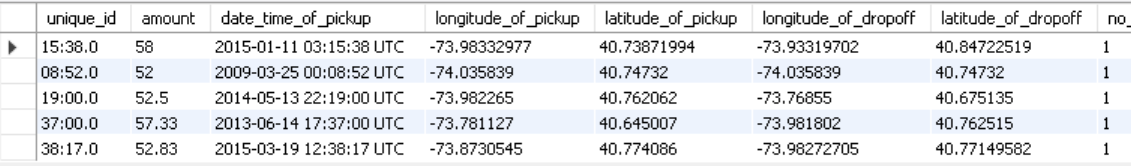


**6.Identify outliers in the fare amount (e.g., trips with fare > $500):**

SELECT \*

FROM taxifare

WHERE amount > 50;



**7.Longest trip based on geographical coordinates (Haversine distance):**

SELECT unique\_id,

       (3959 \* ACOS(COS(RADIANS(latitude\_of\_pickup))

                  \* COS(RADIANS(latitude\_of\_dropoff))

                  \* COS(RADIANS(longitude\_of\_dropoff) - RADIANS(longitude\_of\_pickup))

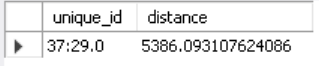
                  + SIN(RADIANS(latitude\_of\_pickup))

                  \* SIN(RADIANS(latitude\_of\_dropoff)))) AS distance

FROM taxifare

ORDER BY distance DESC

LIMIT 1;



**8.Daily revenue trends:**

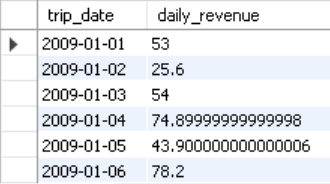
SELECT DATE(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) AS trip\_date,

       SUM(amount) AS daily\_revenue

FROM taxifare

GROUP BY trip\_date

ORDER BY trip\_date;



**9.Most frequent pickup and dropoff locations:**

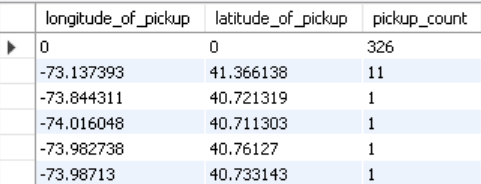
SELECT longitude\_of\_pickup, latitude\_of\_pickup, COUNT(\*) AS pickup\_count

FROM taxifare

GROUP BY longitude\_of\_pickup, latitude\_of\_pickup

ORDER BY pickup\_count DESC

LIMIT 10;



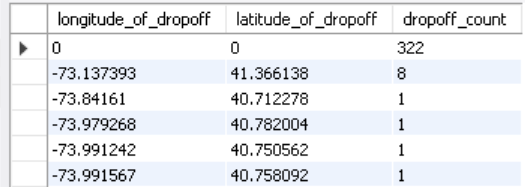
SELECT longitude\_of\_dropoff, latitude\_of\_dropoff, COUNT(\*) AS dropoff\_count

FROM taxifare

GROUP BY longitude\_of\_dropoff, latitude\_of\_dropoff

ORDER BY dropoff\_count DESC

LIMIT 10;



**10.Trips during weekends vs. weekdays:**

SELECT CASE

         WHEN WEEKDAY(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) IN (5, 6) THEN 'Weekend'

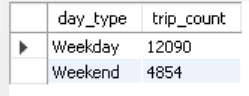
         ELSE 'Weekday'

       END AS day\_type,

       COUNT(\*) AS trip\_count

FROM taxifare

GROUP BY day\_type;



**11.Top 5 dates with the highest total revenue:**

SELECT DATE(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) AS trip\_date,

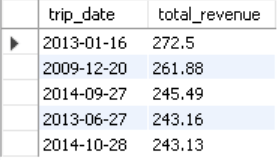
       SUM(amount) AS total\_revenue

FROM taxifare

GROUP BY trip\_date

ORDER BY total\_revenue DESC

LIMIT 5;

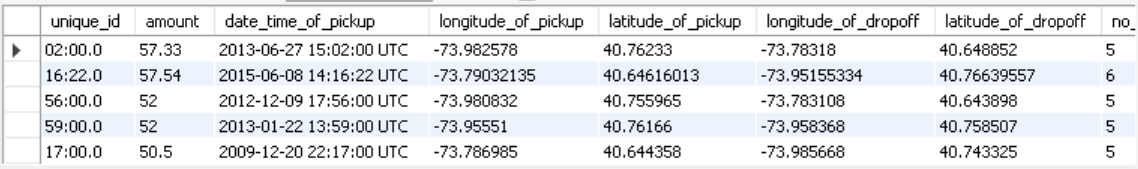


**12.Trips with more than 3 passengers and fare above $50:**

SELECT \*

FROM taxifare

WHERE no\_of\_passenger > 3 AND amount > 50;



**13.Hourly distribution of trips:**

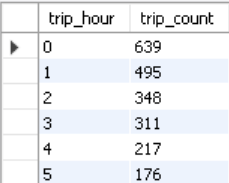
SELECT HOUR(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s')) AS trip\_hour,

       COUNT(\*) AS trip\_count

FROM taxifare

GROUP BY trip\_hour

ORDER BY trip\_hour;



**14. Average fare amount per trip distance bucket (short, medium, long):**

**# Assuming distances are pre-calculated in buckets:**

SELECT CASE

         WHEN distance <= 2 THEN 'Short (<=2 miles)'

         WHEN distance > 2 AND distance <= 5 THEN 'Medium (2-5 miles)'

         ELSE 'Long (>5 miles)'

       END AS distance\_category,

       AVG(amount) AS average\_fare

FROM (

      SELECT \*,

             (3959 \* ACOS(COS(RADIANS(latitude\_of\_pickup))

                      \* COS(RADIANS(latitude\_of\_dropoff))

                      \* COS(RADIANS(longitude\_of\_dropoff) - RADIANS(longitude\_of\_pickup))

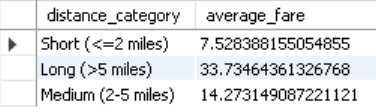
                      + SIN(RADIANS(latitude\_of\_pickup))

                      \* SIN(RADIANS(latitude\_of\_dropoff)))) AS distance

      FROM taxifare

     ) AS trips

GROUP BY distance\_category;



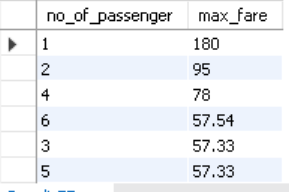
**15. Most expensive trip per passenger count:**

SELECT no\_of\_passenger, MAX(amount) AS max\_fare

FROM taxifare

GROUP BY no\_of\_passenger

ORDER BY max\_fare DESC;



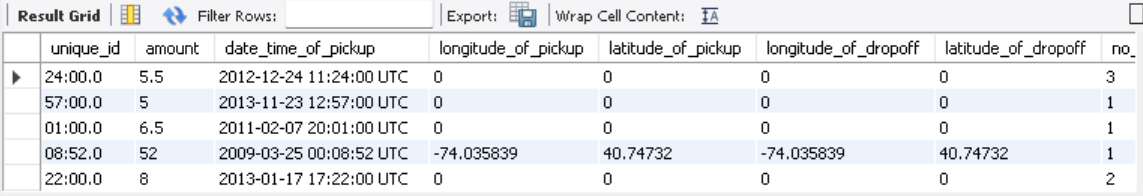
**16.Trips with identical pickup and dropoff locations:**

SELECT \*

FROM taxifare

WHERE longitude\_of\_pickup = longitude\_of\_dropoff

  AND latitude\_of\_pickup = latitude\_of\_dropoff;

****

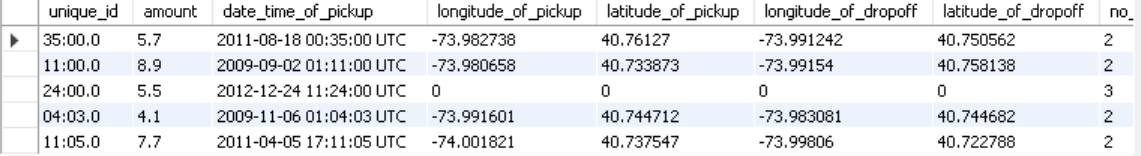
**17.Trips with multiple passengers but a fare below the average fare:**

SELECT \*

FROM taxifare

WHERE no\_of\_passenger > 1

  AND amount < (SELECT AVG(amount) FROM taxifare);



**18.Revenue generated by each unique pickup location:**

SELECT longitude\_of\_pickup, latitude\_of\_pickup,

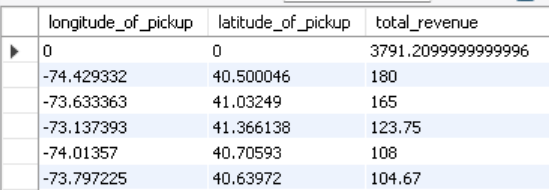
       SUM(amount) AS total\_revenue

FROM taxifare

GROUP BY longitude\_of\_pickup, latitude\_of\_pickup

ORDER BY total\_revenue DESC

LIMIT 10;



**19. Identify trips with an unusually high fare for the distance traveled:**

WITH TripDetails AS (

SELECT \*,

(3959 \* ACOS(COS(RADIANS(latitude\_of\_pickup))

\* COS(RADIANS(latitude\_of\_dropoff))

\* COS(RADIANS(longitude\_of\_dropoff) - RADIANS(longitude\_of\_pickup))

+ SIN(RADIANS(latitude\_of\_pickup))

\* SIN(RADIANS(latitude\_of\_dropoff)))) AS distance

FROM taxifare

)

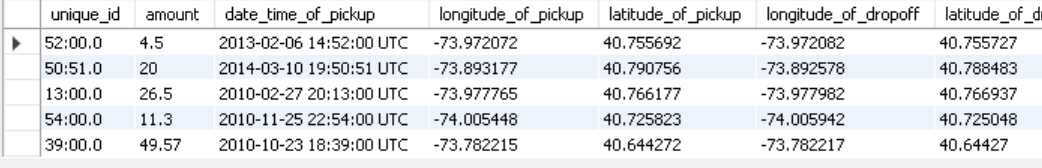
SELECT \*,

(amount / distance) AS fare\_per\_mile

FROM Tripletail’s

WHERE (amount / distance) > 100

AND distance > 0;



**20. Trips with the longest travel time:**

WITH TravelTime AS (

SELECT \*,

TIMESTAMPDIFF(MINUTE, latitude\_of\_pickup,latitude\_of\_dropoff) AS travel\_time\_minutes

FROM taxifare

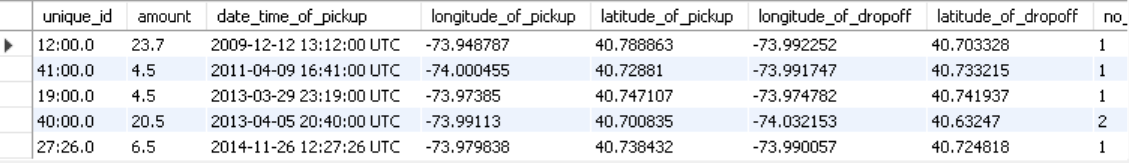
)

SELECT \*

FROM TravelTime

ORDER BY travel\_time\_minutes DESC

LIMIT 10;



**21.Monthly trends in the number of trips:**

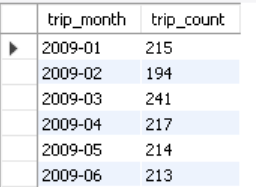
SELECT DATE\_FORMAT(STR\_TO\_DATE(date\_time\_of\_pickup, '%Y-%m-%d %H:%i:%s'), '%Y-%m') AS trip\_month,

COUNT(\*) AS trip\_count

FROM taxifare

GROUP BY trip\_month

ORDER BY trip\_month;



**22.Top 10 drop-off locations by fare amount:**

SELECT longitude\_of\_dropoff, latitude\_of\_dropoff,

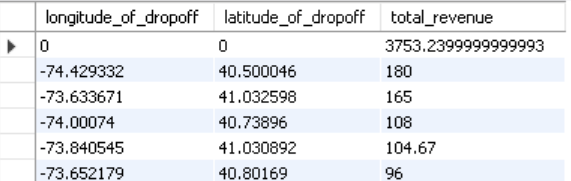
       SUM(amount) AS total\_revenue

FROM taxifare

GROUP BY longitude\_of\_dropoff, latitude\_of\_dropoff

ORDER BY total\_revenue DESC

LIMIT 10;



**23.Average distance traveled by trip based on the number of passengers:**

SELECT no\_of\_passenger,

       AVG(3959 \* ACOS(COS(RADIANS(latitude\_of\_pickup))

                     \* COS(RADIANS(latitude\_of\_dropoff))

                     \* COS(RADIANS(longitude\_of\_dropoff) - RADIANS(longitude\_of\_pickup))

                     + SIN(RADIANS(latitude\_of\_pickup))

                     \* SIN(RADIANS(latitude\_of\_dropoff)))) AS average\_distance

FROM taxi\_fare

GROUP BY no\_of\_passenger

ORDER BY average\_distance DESC;

