Data Exploration Using SQL

Q What is the total number of completed trips?

```
SELECT COUNT(1) as "Total Trips" FROM trips_detail
WHERE end_ride = 1;
```

otp_entered
983

Q How many drivers are there?

```
SELECT COUNT(DISTINCT driverid) AS "Total Drivers" FROM trips;
```

Total Drivers
30

Q What is the total driver's earnings during a day?

```
SELECT SUM(fare) AS total_earnings FROM trips;
```

total_earnings
751343

Q What is the total number of searches that took place?

```
SELECT COUNT(*) Searches FROM trips_detail
WHERE searches = 1;
```

Searches
2161

Q What is the total number of searches which got estimate?

```
SELECT SUM(searches_got_estimate) searches FROM trips_detail;
```

searches
1758

Q What is the total of searches for quotes?

SELECT SUM(searches_for_quotes) searches FROM trips_detail;

searches
1455

Q What is the total number of searches that got quotes ?

SELECT SUM(searches_got_quotes) searches FROM trips_detail;

searches
1277

Q What is the number of times a cancellation was made by drivers?

SELECT (COUNT(tripid) - SUM(driver_not_cancelled)) driver_cancelled_trips FROM trips_detail;

driver_cancelled_trips
1021

Q What is the number of times OTP entered?

SELECT SUM(otp_entered) AS otp_entered FROM trips_detail;

otp_entered
983

Q What is the average distance travelled per trip?

SELECT AVG(distance) AS avgerage_distance FROM trips;

avgerage_distance
14.3927

Q What is the average fare per trip?

SELECT AVG(fare) avg_fare FROM trips;

avg_fare
764.3367

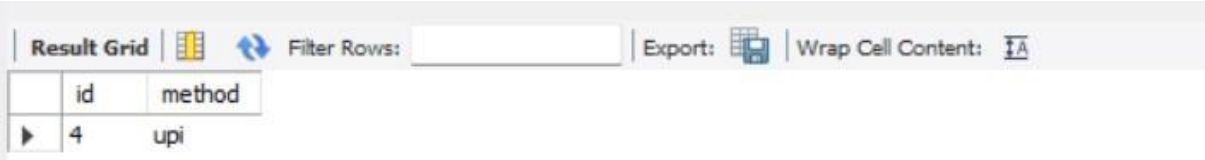
Q What is the total distance travelled during a day?

```
SELECT SUM(distance) distance_travelled FROM trips;
```

distance_travelled
14148

Q Which is the most preferred payment method?

```
SELECT * FROM payment
WHERE  id = (SELECT faremethod FROM trips
             GROUP BY 1 ORDER BY COUNT(tripid) DESC
             LIMIT 1);
```



id	method
4	upi

Q Through which mode of payment , the highest fare was paid?

```
SELECT a.method , b.fare FROM payment a JOIN
(SELECT *  FROM trips WHERE fare = (SELECT MAX(fare) FROM trips)) b
 ON a.id = b.faremethod;
```

method	fare
upi	1500
cash	1500

Q Through which method, total fare paid was highest in a day?

```
SELECT method FROM payment
WHERE id = (SELECT faremethod FROM trips

GROUP BY faremethod ORDER BY SUM(fare) DESC LIMIT 1);
```

method
upi

Q Which two locations had most trips?

```
SELECT loc_from , loc_to
FROM (SELECT loc_from , loc_to , COUNT(DISTINCT tripid),
                    RANK() OVER(ORDER BY COUNT(DISTINCT tripid) DESC ) rno
                    FROM trips
                    GROUP BY 1,2 ) a
WHERE rno = 1;
```

	loc_from	loc_to
▶	16	21
	35	5

Q Who are top 5 earning drivers?

```
SELECT driverid,rno  FROM (
SELECT * ,DENSE_RANK() OVER(ORDER BY fare DESC ) rno
FROM( SELECT driverid, SUM(fare) fare FROM trips GROUP BY 1 ) a ) b
WHERE rno < 6;
```

	driverid	rno
▶	12	1
	8	2
	21	3
	24	4
	30	5

Q For what duration had most of the trips occurred during a day?

```
SELECT duration duration
WHERE id = ( SELECT b.duration
                FROM    (         SELECT *,
                          DENSE_RANK() OVER(ORDER BY cnt DESC) rnk
                          FROM ( SELECT duration, COUNT(DISTINCT tripid) cnt
                                    FROM trips  GROUP BY 1 ) a
                             )
                 b  WHERE rnk <2 );
```

	duration
▶	0-1

Q Which driver and customer pair had most trips together?

```
SELECT * FROM
( SELECT * ,
        DENSE_RANK() OVER(ORDER BY cnt desc ) rnk
 FROM (         SELECT driverid, custid, COUNT(DISTINCT tripid) cnt
        FROM trips
        GROUP BY 1, 2 ) a) b
WHERE rnk < 2;
```

	driverid	custid	cnt	rnk
▶	17	96	4	1
	28	15	4	1

Q What is the conversion from search to estimate rate?

```
SELECT SUM(searches_got_estimate)*100.00/SUM(searches) AS percentage FROM
trips_detail;
```

	percentage
▶	81.351226

Q What is conversion from the estimate to search for quote rates?

```
SELECT SUM(searches_for_quotes)*100.00/SUM(earches_got_estimate) AS percentage
FROM trips_detail;
```

	percentage
▶	82.764505

Q What is the quote acceptance rate?

```
SELECT SUM(searches_got_quotes)*100.00/SUM(searches_for_quotes) AS percentage
FROM trips_detail;
```

	percentage
▶	87.766323

Q What is the booking cancellation rate?

```
SELECT SUM(driver_not_cancelled)*100.00/SUM(searches_got_quotes) AS percentage
FROM trips_detail;
```

	percentage
▶	89.271731

Q What is conversion rate?

```
SELECT SUM(end_ride)*100.00/SUM(searches) AS percentage FROM trips_detail;
```

percentage
45.488200

Q Which area got highest trips in a single duration?

```
SELECT * FROM
( SELECT * ,
        DENSE_RANK() OVER(PARTITION BY duration ORDER BY cnt DESC) rnk
FROM  ( SELECT duration, loc_from, COUNT(DISTINCT tripid) cnt
         FROM trips
        GROUP BY 1, 2 ) a
) b
WHERE rnk = 1;
```

duration	loc_from
6-7	19

Q Which duration got the highest number of trips in each of the location present?

```
SELECT  * FROM
( SELECT * ,
          DENSE_RANK() OVER(PARTITION BY loc_from ORDER BY cnt DESC) rnk
 FROM ( SELECT duration, loc_from, COUNT(DISTINCT tripid) cnt
        FROM trips
        GROUP BY 1, 2 ) a
 ) b
 WHERE rnk = 1;
```

duration	loc_from	cnt	rnk
14	1	3	1
7	2	4	1
18	3	4	1
4	4	3	1
8	4	3	1
23	4	3	1
2	5	3	1
14	6	4	1
9	7	3	1
14	8	3	1
15	8	3	1

Q Which area has highest aggregated fare prices compared to others?

```
SELECT * FROM
( SELECT * ,
RANK() OVER(ORDER BY fare DESC) rnk
FROM  ( SELECT loc_from , SUM(fare) fare
 FROM trips GROUP BY 1) a
 ) b
WHERE rnk <2;
```

	loc_from	fare	rnk
▶	6	30295	1

Q Which area has highest number of cancellations by drivers?

```
SELECT * FROM
( SELECT loc_from ,
(COUNT(1) - SUM(driver_not_cancelled)) AS driver_cancellation ,
        DENSE_RANK() OVER( ORDER BY (COUNT(1) - SUM(driver_not_cancelled)) DESC) rnk
FROM trips_detail
GROUP BY 1) a
WHERE rnk < 2;
```

	loc_from	cancelled	rnk
▶	1	43	1

Q Which area got highest number of cancellations by customer?

```
SELECT * FROM (
SELECT loc_from , (COUNT(1) - SUM(customer_not_cancelled)) customer_cancellation ,
        dense_rank() OVER( ORDER BY (count(1) - SUM(customer_not_cancelled)) DESC) rnk
FROM trips_detail
GROUP BY 1) a
WHERE rnk < 2;
```

	loc_from	customer_cancellation	rnk
▶	4	40	1

Q Which duration got the highest fares ?

```
WITH highest_fare AS (
        SELECT * FROM(
        SELECT * , RANK() OVER(ORDER BY fare DESC) rnk FROM
        (SELECT duration , SUM(fare) fare FROM trips GROUP BY 1) a ) b
        WHERE rnk <2 )
SELECT d.duration, h.fare FROM highest_fare h
JOIN duration d ON h.duration = d.id
```

	duration	fare
▶	0-1	45019