**Cab Booking System for Data Analysis:**

The cab booking industry has grown significantly, providing convenient transportation services to customers while offering earning opportunities for drivers. To ensure smooth operations, improve customer satisfaction, and enhance business performance, cab companies rely on data analysis. The Cab Booking System consists of multiple entities such as Customers, Drivers, Cabs, Bookings, Trip Details, and Feedback, all of which generate valuable data. By analysing booking trends, driver performance, customer behaviour, and revenue metrics, companies can optimize their services, improve resource allocation, and maximize profitability. This case study focuses on designing a relational database schema for a cab booking system and using SQL queries to extract insights from the stored data.

**Objective:**

A cab service company wants to enhance its operations by analysing customer bookings, driver performance, and trip details. The company aims to: - Monitor ongoing and completed bookings. - Track customer preferences and behaviour. - Evaluate driver efficiency and performance. - Analyse revenue trends based on fares and trip details. - Identify operational bottlenecks and areas for improvement. By structuring and querying the database, stakeholders can make data-driven decisions to improve the efficiency of the cab booking system.

Problem Statement:

Customer and Booking Analysis

1. Identify customers who have completed the most bookings. What insights can you draw about their behaviour?

Ans: SELECT c.\*, COUNT(\*) AS total\_bookings FROM bookings b

join customers c on c.customer\_id=b.customer\_id

WHERE status='Completed'

GROUP BY customer\_id

ORDER BY total\_bookings DESC

LIMIT 1;



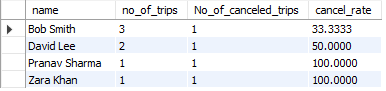
2. Find customers who have cancelled more than 30% of their total bookings. What could be the reason for frequent cancellations?

Ans: SELECT c.name,count(\*)as no\_of\_trips, sum(status='Cancelled')as No\_of\_canceled\_trips, SUM(status='Cancelled')/COUNT(\*)\*100 AS cancel\_rate FROM bookings b

join customers c on c.customer\_id=b.customer\_id

GROUP BY b.customer\_id

HAVING cancel\_rate > 30;



The output shows that many of the customers who have high cancelation rate are of the people who have just used it once and those David Lee had cancel the ride because of a delay at arrival, and Bob Smith was cancelled by the driver.

3. Determine the busiest day of the week for bookings. How can the company optimize cab availability on peak days?

Ans: SELECT DAYNAME(booking\_time) AS day, COUNT(\*) AS total FROM bookings

GROUP BY day

ORDER BY total DESC

LIMIT 1;



The busiest day in the week is Tuesday so the optimize the cab routes.

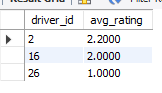
Driver Performance & Efficiency

1. Identify drivers who have received an average rating below 3.0 in the past three months. What strategies can be implemented to improve their performance?

Ans: SELECT driver\_id, AVG(rating) AS avg\_rating FROM feedback

GROUP BY driver\_id

HAVING avg\_rating < 3.0;



As I have the data of only 2 months the driver id’s 16 and 26 have only one ride and driver id 2 has many bad comments so if he could reach his pick up location faster.

2. Find the top 5 drivers who have completed the longest trips in terms of distance. What does this say about their working patterns?

Ans: SELECT b.driver\_id,d.name, SUM(distance\_km) AS total\_distance FROM bookings b

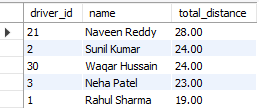
join drivers d on d.driver\_id=b.driver\_id

WHERE status='Completed'

GROUP BY b.driver\_id

ORDER BY total\_distance DESC

LIMIT 5;



Compared to the other drivers these 5 have the most distance covered. That they covered lots of distance in a ride.

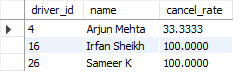
3. Identify drivers with a high percentage of cancelled trips. Could this indicate driver unreliability?

Ans: SELECT b.driver\_id,d.name, SUM(status='Cancelled')/COUNT(\*)\*100 AS cancel\_rate FROM bookings b

join drivers d on d.driver\_id=b.driver\_id

GROUP BY driver\_id

HAVING cancel\_rate > 25;



The output shows that driers with id 4,16,26 are not very reliable.

Revenue & Business Metrics

1. Calculate the total revenue generated by completed bookings in the last 6 months. How has the revenue trend changed over time?

Ans: SELECT SUM(fare) AS total\_revenue FROM bookings

WHERE status='Completed' AND booking\_time >= DATE\_SUB(CURDATE(), INTERVAL 6 MONTH);



The total revenue accumulated in the past 2 months is five thousand seven hundred and fifteen rupees.

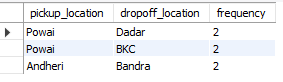
2. Identify the top 3 most frequently travelled routes based on Pickup Location and Drop off Location. Should the company allocate more cabs to these routes?

Ans: SELECT pickup\_location, dropoff\_location, COUNT(\*) AS frequency FROM bookings

GROUP BY pickup\_location, dropoff\_location

ORDER BY frequency DESC

LIMIT 3;



The most popular routes are Powai-Dadar, Powai-BKC, Andheri-Bandra. As there are only 2 pickups so I don’t believe that allocating more cabs are necessary.

3. Determine if higher-rated drivers tend to complete more trips and earn higher fares. Is there a direct correlation between driver ratings and earnings?

Ans: SELECT d.driver\_id, AVG(f.rating) AS avg\_rating, COUNT(b.booking\_id) AS total\_trips, SUM(b.fare) AS total\_earnings FROM drivers d

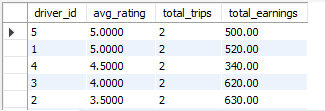
JOIN bookings b ON d.driver\_id=b.driver\_id

JOIN feedback f ON b.booking\_id=f.booking\_id

WHERE b.status='Completed'

GROUP BY d.driver\_id

order by total\_trips DESC,avg\_rating DESC;



As we can see that the total earnings has no corelation with the average rating as drivers with id number of 3 and 2 has higher earning even though their rating is at 4 and 3.5

Operational Efficiency & Optimization

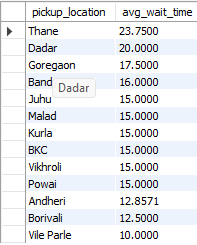
1. Analyse the average waiting time (difference between booking time and trip start time) for different pickup locations. How can this be optimized to reduce delays?

Ans: SELECT pickup\_location, AVG(TIMESTAMPDIFF(MINUTE, booking\_time, trip\_start\_time)) AS avg\_wait\_time FROM bookings

WHERE status='Completed'

GROUP BY pickup\_location

Order by avg\_wait\_time;



The highest average wait time is of 23 and 20 minutes for Thane and Dadar respectfully.

Leaving that the average wait time is between 10-17 minutes.

2. Identify the most common reasons for trip cancellations from customer feedback. What actions can be taken to reduce cancellations?

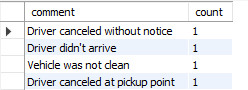
Ans: SELECT comment, COUNT(\*) AS count FROM feedback f

JOIN bookings b ON f.booking\_id=b.booking\_id

WHERE b.status='Cancelled'

GROUP BY comment

ORDER BY count DESC ;



The greatest number of the cancellation is done by the drivers. So, to reduce the number of cancellations we have to train the driver to ignore the message if they are unable to make it or it seems to much time will be taken. And as for the vehicle being messy we need to train the drivers to clean up the vehicle after every ride.

3. Find out whether shorter trips (low-distance) contribute significantly to revenue. Should the company encourage more short-distance rides?

Ans: SELECT CASE WHEN distance\_km <= 8 THEN 'Short' ELSE 'Long' END AS trip\_type, SUM(fare) AS total\_revenue FROM bookings

WHERE status='Completed'

GROUP BY trip\_type;



As the result show the revenue of short distance is contributing around 26% in the overall contribution to the revenue. So, it would not be a good move to promote short-distance ride

Comparative & Predictive Analysis

1. Compare the revenue generated from 'Sedan' and 'SUV' cabs. Should the company invest more in a particular vehicle type?

Ans: SELECT c.vehicle\_type, SUM(b.fare) AS total\_revenue,sum(b.fare)/(select sum(b.fare) from bookings b )\*100 as total\_percentage

FROM bookings b

JOIN cabs c ON b.cab\_id = c.cab\_id

WHERE b.status = 'Completed' and c.vehicle\_type in("Sedan","SUV")

GROUP BY c.vehicle\_type;



As per the outcome the SUV is generating about 34% of revenue so the company has to invest in SUV’s.

2. Predict which customers are likely to stop using the service based on their last booking date and frequency of rides. How can customer retention be improved?

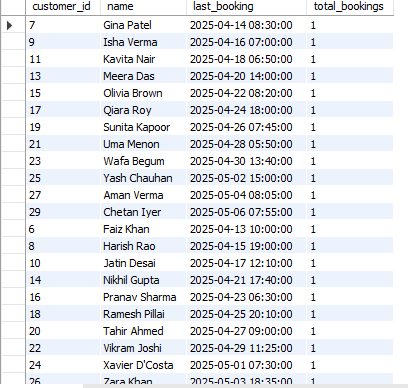
Ans: SELECT b.customer\_id,c.name, MAX(b.booking\_time) AS last\_booking, COUNT(\*) AS total\_bookings FROM bookings b

join customers c on c.customer\_id=b.customer\_id

GROUP BY customer\_id

HAVING DATEDIFF(CURDATE(), last\_booking) > 60

order by total\_bookings;



According to the output the almost all of the customers except for customers with id number 1,2,3,4,5,12 are the people who are using it the most.

3. Analyse whether weekend bookings differ significantly from weekday bookings. Should the company introduce dynamic pricing based on demand?

Ans: SELECT CASE WHEN DAYOFWEEK(booking\_time) IN (1,7) THEN 'Weekend' ELSE 'Weekday' END AS day\_type, COUNT(\*) AS total\_bookings FROM bookings

GROUP BY day\_type;



As per the output we should do dynamic pricing for the weekdays.