OLA Data Analysis Project Report

# Project Summary

This project analyzes a simulated OLA ride dataset using SQL. The goal was to uncover key business insights related to customer behavior, driver performance, cancellation trends, operational efficiency, and revenue optimization. The dataset includes 40,000 rows of rides from July , including detailed fields like vehicle types, booking status, ratings, pickup/drop locations, timings, and more.

1. Date

2. Time

3. Booking ID

4. Booking Status

5. Customer ID

6. Vehicle Type

- Auto

- Prime Plus

- Prime Sedan

- Mini

- Bike

- eBike

- Prime SUV 7.

Pickup Location

8. Drop Location

9. AvgVTAT (Time taken to arrive at the vehicle)

10. Avg CTAT (Time taken to arrive the Customer)

11. Cancelled Rides by Customer

12. Reason for cancelling by Customer

-Driver is not moving towards pickup location

- Driver asked to cancel AC is not working (Only for 4-wheelers)

-Change of plans

-Wrong Address

13. Cancelled Rides by Driver

-Personal & Car related issues

- Customer related issue

-The customer was coughing/sick

-More than permitted people in there

14. Incomplete Rides

15. Incomplete Rides Reason

-Customer Demand

- Vehicle Breakdown

-Other Issue

16. Booking Value

17. Ride Distance

18. Driver Ratings

19. Customer Rating

# SQL Analysis Questions and Insights

## 1. Retrieve all successful bookings

A basic filter used to isolate successful rides for focused analysis.

## 2. Find the average ride distance for each vehicle type

This helps understand how far different vehicles typically travel, informing fleet allocation.

## 3. Top 5 customers who booked the highest number of rides

Identifies highly engaged users who could be targeted for loyalty rewards.

## 4. Identify peak ride request hours for each vehicle type on weekends

Reveals temporal demand trends to guide driver availability and pricing strategies.

## 5. Find the cancellation rate by drivers per vehicle type

Highlights which vehicle categories face the most cancellations initiated by drivers.

## 6. Out of all bookings cancelled by drivers, what proportion came from each vehicle type

Gives global contribution of each vehicle type to total driver cancellations.

## 7. Detect anomalies: Bookings where VTAT or CTAT > 30 minutes

Used to spot operational inefficiencies or unusually long waits.

## 8. Find customers with the highest cancellation rate (customer-initiated)

Detects potentially unreliable users who frequently cancel rides.

## 9. Select unusually low-rated successful rides (Customer or Driver rating < 3.5)

Flags rides where either party had a poor experience, useful for quality audits.

## 10. Find average booking value per km for each vehicle type

Helps compare revenue efficiency across vehicle categories.

## 11. Detect pickup-drop pairs with high cancellation rates

Reveals problematic routes that may need review or incentives.

## 12. Calculate average VTAT and CTAT by vehicle and time of day block

Analyzes vehicle/customer response speed across different day periods.

## 13. Identify vehicle types that frequently result in incomplete rides

Pinpoints vehicle types prone to ride failures for deeper investigation.

# Power BI Insights

The Power BI section included visuals for ride volume trends, booking status breakdowns, top customers by revenue, cancellation reasons, and rating comparisons.

# Power BI Questons

# 1. Ride Volume Over Time

# 2. Booking Status Breakdown

# 3. Top 5 Vehicle Types by Ride Distance

# 4. Average Customer Ratings by Vehicle Type

# 5. cancelled Rides Reasons

# 6. Revenue by Payment Method

# 7. Top 5 Customers by Total Booking Value

# 8. Ride Distance Distribution Per Day

# 9. Driver Ratings Distribution

# 10. Customer vs. Driver Rating

Power BI Answers:

OLA Data Analyst Project Segregation of the views:

OLA Dashboard (Executive Summary Page)

Summary Cards (using Card visuals):

-Total Revenue

-Total Number of Bookings

-Successful Bookings

-Travelled

Booking Value by Vehicle Type (Bar Chart)

Count of Bookings by Pickup Location (Matrix Visual)

Filters/Slicers:

Date Range

Payment Method

1. Overall

- Ride Volume Over Time

-Booking Status Breakdown

-Average of booking value when booking status is success

2. Vehicle Type

-Top 5 Vehicle Types by Ride Distance

3. Revenue

- Revenue by Payment Method

-Top 5 Customers by Total Booking Value

-Ride Distance Distribution Per Day

4. Cancellation

- Cancelled Rides Reasons (Customer)

-cancelled Rides Reasons(Drivers)

5. Ratings

- Driver Ratings

- Customer Rating

SQL Question and Answer

# 1) retrive all successful bookings

Create view successful\_bookings as

Select \* from ola\_details where booking\_status="success"

Select\* from successfull\_bookings

# 2)Find the average ride distance for each vehicle type

select vehicle\_type,avg(ride\_distance) as avg\_ride from ola\_details group by vehicle\_type

# 3)top 5 customer who booked highest number rides

select customer\_id,count(1) as c from ola\_details group by customer\_id order by c desc limit 5

# 4) identify peak ride request hour sfor each vehivle type on weekends

with cte as (select vehicle\_type,hour(time) as hour ,count(hour(time)) as "number of time request per hour" ,rank() over(partition by vehicle\_type order by `number of time request per hour` desc) as rank from ola\_details where dayname(date) in ("sunday","saturday") group by vehicle\_type,hour )

select \* from cte where rank<4 order by vehicle\_type , rank

# 5)Find the cancellation rate by driver per vehiclet type

select vehicle\_type , count(case when booking\_status="canceled by driver" then 1 end)\*100/count(\*) as cancel\_rate\_by\_driver from ola\_details group by vehicle\_type

# 6)Out of all bookings that were cancelled by drivers, what proportion came from each vehicle type

select vehicle\_type,count(\*)\*100/(select count(\*) from ola\_details where booking\_status="canceled by driver") from ola\_details where booking\_status="canceled by driver" group by vehicle\_type

# 7)Detect anomalies :booking where v\_TAT or C\_TAT>30 minutes

select \* from ola\_details where v\_tat>=30 or c\_tat>=30

# 8)Find customer with the highest cancellation rate

select customer\_id,count(case when booking\_status="canceled by customer" then 1 end)\*100/count(\*) as cancel\_rating from ola\_details group by customer\_id order by cancel\_rating desc

#below query is slow compared to above one

create view cancel\_vs\_total\_booking as

with cancel\_booking as (select customer\_id ,count(customer\_id) as count from ola\_details where booking\_status="canceled by customer" group by customer\_id ),

total\_booking as (select customer\_id,count(customer\_id) as count from ola\_details group by customer\_id)

select cb.customer\_id,cb.count as cancel\_count,t.count as total\_count from cancel\_booking as cb join total\_booking as t on cb.customer\_id=t.customer\_id

select customer\_id,cancel\_count\*100/total\_count as cancel\_rate from cancel\_vs\_total\_booking

# 9) Setect unusually low-rated successful rides(customer\_rating<3.5 or driver\_rating<3.5)

select \* from ola\_details where booking\_status = "success" and (customer\_rating<3.5 or driver\_ratings<3.5)

# 10)Find average booking value per k for each vehicle type

select vehicle\_type,avg(booking\_value/ride\_distance) as "average price per km" from ola\_details where booking\_status="success" group by vehicle\_type

# 11)detect pickup-drop pairs with high cancellation rates

select pickup\_location,drop\_location,count(\*) as total\_bookings,count(case when booking\_status="canceled by driver" or booking\_status="canceled by customer" then 1 end) as cancelled, count(case when booking\_status="canceled by driver" or booking\_status="canceled by customer" then 1 end)\*100/count(\*) as cancel\_rate from ola\_details group by pickup\_location,drop\_location order by cancel\_rate desc

# 12)calculate avearage VTAT and CTAT by vehicle and time of day block(morning/afternoon/Evening/Night)

SELECT Vehicle\_Type,

Case

When hour(Time)<12 and hour(time)>4 then "Morning"

WHEN hour(Time)>=12 and hour(time)<17 then "Afternoon"

WHEN hour(Time)>=17 and hour(time)<20 then "Evening"

WHEN hour(Time)>=20 and hour(time)<=24 then "Night"

WHEN hour(Time)>=0 and hour(time)<=4 then "Night"

End as "day\_block",round(avg(V\_TAT),2) as avg\_VTAT,round(avg(C\_TAT),2) as avg\_CTAT from ola\_details WHERE Booking\_Status="Success" group by Vehicle\_Type,day\_block

# 13)identify vehicle types that frequently result in incomplete rides

select vehicle\_type,count(\*) from ola\_details where incomplete\_rides="yes" group by vehicle\_type

# Conclusion

This project demonstrates the use of SQL and Power BI in analyzing ride-hailing data. It covers both fundamental and advanced questions to extract insights that could benefit operations, customer experience, and profitability in real-world mobility platforms like OLA.