# SQL Project | Uber Rides Data Analysis

**Tools Used :** PostgreSQL / MySQL, SQL
**Domain :** Ride-Hailing & Transportation Analytics
**Company Context :** Uber

---

## Project Summary

This project analysis Uber ride booking data to understand revenue performance, cancellations, customer behavior, vehicle efficiency, and location demand.
The goal is to simulate how a real ride-hailing company uses data to make business decisions and reduce revenue loss.

## Project Overview

### Trip Information

- BookingID
- TripDate
- TripTime
- BookingStatus

### Customer & Driver Information

- CustomerID
- CustomerCancelledRides
- CustomerCancelReason
- DriverCancelledRides
- DriverCancelReason

### Vehicle & Location

- VehicleType
- PickupLocation
- DropLocation

### Financial & Ratings

- Amount
- RideDistanceKM
- PaymentMethod
- DriverRating
- CustomerRating

### Incomplete Rides

- IncompleteRides
- IncompleteReason

## Key Business KPIs

- Total Bookings
- Completed vs Cancelled vs Incomplete Rides
- Total & Average Revenue
- Total & Average Ride Distance
- Average Customer & Driver Ratings

---

## Create Table:-

```
CREATE TABLE Uber (

    BookingID               VARCHAR(20) PRIMARY KEY,

    TripDate                DATE,

    TripTime                TIME,

    BookingStatus            VARCHAR(30),

    CustomerID              VARCHAR(20),

    VehicleType              VARCHAR(15),

    PickupLocation           VARCHAR(100),

    DropLocation             VARCHAR(100),

    CustomerCancelledRides INT,

    CustomerCancelReason    VARCHAR(100),

    DriverCancelledRides    INT,

    DriverCancelReason      VARCHAR(100),

    IncompleteRides          INT,

    IncompleteReason         VARCHAR(100),

    Amount                   DECIMAL(10,2),

    RideDistanceKM           DECIMAL(8,2),

    DriverRating             DECIMAL(2,1),

    CustomerRating           DECIMAL(2,1),

    PaymentMethod            VARCHAR(20)
);
```

---

## Overall Business Performance (Key Business KPIs)

```sql
select
count(*) as Total_Booking,
sum(case when BookingStatus = 'Completed' then 1 else 0 end ) as
Complete_Booking,
sum(case when BookingStatus like 'Canc%' then 1 else 0 end ) as
Cancelled_Booking,
to_char(sum(amount), '"RS "999,999,999.00') as Total_Revenue,
to_char(avg(amount), '"RS"999,999,999.00') as Avg_Revenue,
concat(round(sum(ridedistanceKM),2),' Km') as Total_Distance,
concat(round(avg(ridedistanceKM),2),' Km') as Avg_Distance
from uber;
```

**Result:-**

Data Output   Messages   Notifications

Showing rows: 1 to 1   Page No:

| | total_booking bigint | complete_booking bigint | cancelled_booking bigint | total_revenue text | avg_revenue text | total_distance text | avg_distance text |
|---|---|---|---|---|---|---|---|
| 1 | 148767 | 91950 | 37427 | RS  51,288,283.00 | RS      508.32 | 2485817.43 Km | 24.64 Km |

## 1.Business Performance

**How many total rides were booked and how much revenue was generated?**

```sql
Select
count(*) as Total_Rides,
to_char(sum(amount), '"RS "999,999,999.00') as Total_Revenue
From uber;
```

**Result:-**

Data Output   Messages   Notifications

| | total_rides bigint | total_revenue text |
|---|---|---|
| 1 | 148767 | RS  51,288,283.00 |

**What percentage of rides were completed, cancelled, and incomplete?**

```sql
Select
Concat(round(sum(case when Bookingstatus = 'Completed' then 1 else 0 end)
* 100.00 / count(*) ,2),' %') as Completed_Rides,

 Concat(round(sum(case when Bookingstatus like 'Cance%' then 1 else 0 end)
* 100.00 / count(*) ,2),' %') as Cancelled_Rides,

Concat(round(sum(case when Bookingstatus in('Incomplete','No Driver
Found') then 1 else 0 end) * 100.00 / count(*) ,2),' %') as
Incomplete_Rides
```

```
from uber;
```
Result:-



## What percentage of rides were completed, cancelled by Customer & Driver,noDriver Found and incomplete?

```
Select
Bookingstatus,Count(*) as TotalRides,
Concat(Round(count(*) * 100 / Sum(count(*)) over(),2),' %') as Percentage
From uber
Group by Bookingstatus
Order by totalRides Desc;
```

Result:-



## --What is the average fare and average distance per ride?

```
select
concat('Rs ',round(avg(amount),2)) as Average_Fare ,
concat(round(Avg(RideDistanceKM),2),' Km') as Average_Distance
from uber
where bookingStatus='Completed';
```

Result:-

## What are the average driver and customer ratings?

```
Select
round(AVG(DriverRating),2) AS Average_Driver_Rating,
round(AVG(CustomerRating),2) AS Avgerage_Customer_Rating
FROM Uber;
```

**Result:-**

Data Output   Messages   Notifications

| | average_driver_rating numeric | avgerage_customer_rating numeric |
|---|---|---|
| 1 | 4.23 | 4.40 |

# 2.Time Analysis

## What are the peak booking hours?

```
select
    to_char(triptime, 'hh12 AM')as PeakHour,
    count(*) as bookingCount
From uber
group by PeakHour
order by BookingCount Desc,peakhour;
```

**Result:-**

Data Output   Messages   Notifications

| | peakhour text | bookingcount bigint |
|---|---|---|
| 1 | 06 PM | 12301 |
| 2 | 07 PM | 10960 |
| 3 | 05 PM | 10955 |
| 4 | 04 PM | 9571 |
| 5 | 08 PM | 9534 |
| 6 | 10 AM | 9505 |
| 7 | 11 AM | 8319 |
| 8 | 09 AM | 8158 |
| 9 | 03 PM | 8142 |
| 10 | 09 PM | 8024 |
| 11 | 02 PM | 6973 |
| 12 | 12 PM | 6941 |
| 13 | 08 AM | 6805 |
| 14 | 01 PM | 5429 |
| 15 | 07 AM | 5409 |
| 16 | 10 PM | 5389 |
| 17 | 06 AM | 4124 |
| 18 | 05 AM | 2769 |
| 19 | 11 PM | 2740 |
| 20 | 03 AM | 1370 |
| 21 | 12 AM | 1366 |
| 22 | 01 AM | 1350 |
| 23 | 02 AM | 1325 |
| 24 | 04 AM | 1308 |

## Which time slot (Morning, Afternoon, Evening, Night) generates the most rides?

```
select

case
when extract(Hour from Triptime) between 5 and 8 then 'Early Morning 5 AM
- 8 AM'
when extract(Hour from Triptime) between 9 and 11 then 'Late Morning 9 AM
- 11 AM'
when extract(Hour from Triptime) between 12 and 16 then 'Afternoon 12 PM -
4 PM'
when extract(Hour from Triptime) between 17 and 20 then 'Evening 5 PM - 8
PM'
else 'Night 9 PM - 4 AM' End as TimeSlot,

Count(*) as Booking_Count
From uber
group by Timeslot
Order by Booking_Count Desc;
```

**Result:-**

Data Output  Messages  Notifications

| | timeslot text | booking_count bigint |
|---|---|---|
| 1 | Evening 5 PM - 8 PM | 43750 |
| 2 | Afternoon 12 PM - 4 PM | 37056 |
| 3 | Late Morning 9 AM - 11 AM | 25982 |
| 4 | Night 9 PM - 4 AM | 22872 |
| 5 | Early Morning 5 AM - 8 AM | 19107 |

## 3.Vehicle Performance

### Which vehicle type is most used?

```
select
vehicleType,
Count(*) as Total_Rides
From uber
group by vehicletype
order by total_rides desc;
```

**Result:-**

Data Output  Messages  Notificati

| | vehicletype character varying | total_rides bigint |
|---|---|---|
| 1 | Auto | 37089 |
| 2 | Go Mini | 29545 |
| 3 | Go Sedan | 26940 |
| 4 | Bike | 22330 |
| 5 | Premier Sedan | 17979 |
| 6 | eBike | 10469 |
| 7 | Uber XL | 4415 |

## Which vehicle type generates the highest revenue and what % of total revenue does it contribute?

```
select
vehicletype,
round(sum(amount),2) as Revenue,
Concat(round(100.00 * sum(amount) / sum(sum(amount)) over(),2),' %') as
percentage
from uber
where bookingStatus ='Completed'
Group by vehicletype
order by Revenue Desc;
```

**Result:-**

| | vehicletype<br>character varying | revenue<br>numeric | percentage<br>text |
|---|---|---|---|
| 1 | Auto | 11589350.00 | 24.80 % |
| 2 | Go Mini | 9302615.00 | 19.91 % |
| 3 | Go Sedan | 8448170.00 | 18.08 % |
| 4 | Bike | 7062195.00 | 15.11 % |
| 5 | Premier Sedan | 5679261.00 | 12.15 % |
| 6 | eBike | 3260566.00 | 6.98 % |
| 7 | Uber XL | 1389371.00 | 2.97 % |

## What is the average revenue per km by vehicle type?

```
Select
Vehicletype,
concat('Rs ',round(sum(amount)/ sum(RidedistanceKM),2) ) as
Avg_Revenue_per_KM
from uber
where bookingstatus = 'Completed'
group by vehicletype;
```

**Result:-**

| | vehicletype<br>character varying | avg_revenue_per_km<br>text |
|---|---|---|
| 1 | Auto | Rs 19.50 |
| 2 | Bike | Rs 19.56 |
| 3 | eBike | Rs 19.11 |
| 4 | Go Mini | Rs 19.52 |
| 5 | Go Sedan | Rs 19.70 |
| 6 | Premier Sedan | Rs 19.63 |
| 7 | Uber XL | Rs 19.58 |

# 4.Customer Behavior

## What percentage of customers have cancelled at least one ride?

select

count(distinct customerID) as Total_Customer,

count(distinct CustomerID) filter(where CustomerCancelledRides = 1) as Customer_Cancelled_Rides,

concat(Round(count(distinct CustomerID) filter(where CustomerCancelledRides = 1) * 100.00 / count(distinct customerID),2),' %')  as CancelationRate

from uber;

**Result:-**

| | total_customer bigint | customer_cancelled_rides bigint | cancelationrate text |
|---|---|---|---|
| 1 | 103462 | 2181 | 2.11 % |

## Who are the top 10 highest-spending customers?

Select
CustomerID,to_char(sum(Amount),'"Rs"999,999,999.00') as revenue
from uber
where Bookingstatus = 'Completed'
group by CustomerId
order by revenue Desc limit 10 ;

**Result:-**

| | customerid character vary | revenue text |
|---|---|---|
| 1 | C7828101 | Rs 7,683.00 |
| 2 | C7186567 | Rs 6,101.00 |
| 3 | C2536937 | Rs 6,019.00 |
| 4 | C9494011 | Rs 5,966.00 |
| 5 | C3446144 | Rs 5,589.00 |
| 6 | C9610969 | Rs 5,494.00 |
| 7 | C5235759 | Rs 5,379.00 |
| 8 | C1578062 | Rs 5,368.00 |
| 9 | C3780888 | Rs 5,323.00 |
| 10 | C3791237 | Rs 5,292.00 |

## 5.Driver Reliability

**What are the top driver cancellation reasons?**

```
select
DrivercancelReason as Driver_cancel_Reason,Count(*) AS Total_Cancellations
from uber
where drivercancelledRides = 1
group by DrivercancelReason
order by Total_Cancellations desc;
```

**Result:-**

Data Output   Messages   Notifications

| | driver_cancel_reason<br>character varying (100) | total_cancellations<br>bigint |
|---|---|---|
| 1 | Customer related issue | 6823 |
| 2 | The customer was coughing/sick | 6732 |
| 3 | Personal & Car related issues | 6700 |
| 4 | More than permitted people in there | 6676 |

## 6.Location Intelligence

**Which pickup locations generate the most completed rides?**

```
select
pickuplocation,
sum(case when bookingstatus = 'Completed' then 1 else 0  end ) as
Complete_rides
from uber
group by pickuplocation
order by Complete_rides desc;
```

**Result:-**

Data Output   Messages   Notifications

| | pickuplocation<br>character varying ( | complete_rides<br>bigint |
|---|---|---|
| 1 | Khandsa | 596 |
| 2 | Barakhamba R... | 589 |
| 3 | Subhash Chowk | 574 |
| 4 | Madipur | 572 |
| 5 | Mehrauli | 568 |
| 6 | Kanhaiya Nagar | 566 |
| 7 | Badarpur | 564 |
| 8 | Ashok Park Ma... | 562 |
| 9 | Lok Kalyan Marg | 560 |
| 10 | Dwarka Sector ... | 558 |
| 11 | Jahangirpuri | 556 |
| 12 | Saket | 555 |
| 13 | AIIMS | 555 |
| 14 | Panchsheel Pa... | 554 |
| 15 | Malviya Nagar | 554 |

## Which pickup-drop pairs generate the highest average fare?

```
select
pickuplocation,droplocation,
round(avg(amount),2) as avg_trip_amount
from uber
where bookingStatus = 'Completed'
group by pickuplocation,droplocation
order by avg_trip_amount Desc;
```

**Result:-**

Data Output    Messages    Notifications

| | pickuplocation character varying ( | droplocation character varying ( | avg_trip_amount numeric |
|---|---|---|---|
| 1 | Sultanpur | Lal Quila | 3753.00 |
| 2 | Old Gurgaon | Greater Kailash | 3683.00 |
| 3 | Shahdara | Tagore Garden | 3587.00 |
| 4 | Basai Dhankot | New Delhi Rail… | 3504.00 |
| 5 | Shivaji Park | Laxmi Nagar | 3303.00 |
| 6 | Kadarpur | Peeragarhi | 3292.00 |
| 7 | Vasant Kunj | Greater Kailash | 3175.00 |
| 8 | Kaushambi | Rohini | 3081.00 |
| 9 | Panchsheel Pa… | GTB Nagar | 3002.00 |
| 10 | Central Secret… | Vaishali | 2819.00 |
| 11 | Karol Bagh | Janakpuri | 2493.00 |
| 12 | Aya Nagar | Old Gurgaon | 2449.00 |
| 13 | Sarojini Nagar | Ashok Vihar | 2414.00 |
| 14 | Karkarduma | Delhi Gate | 2413.00 |
| 15 | IFFCO Chowk | Govindpuri | 2395.00 |
| 16 | IIT Delhi | Pulbangash | 2394.00 |
| 17 | Ashram | New Colony | 2372.00 |
| 18 | Bhiwadi | Sonipat | 2370.00 |
| 19 | Tilak Nagar | Pragati Maidan | 2364.00 |
| 20 | Rithala | Paschim Vihar | 2347.00 |
| 21 | Tagore Garden | INA Market | 2321.00 |
| 22 | Tughlakabad | Palam Vihar | 2309.00 |
| 23 | Rohini West | INA Market | 2308.00 |
| 24 | Punjabi Bagh | Arjangarh | 2275.00 |
| 25 | Maidan Garhi | IGI Airport | 2265.00 |
| 26 | Welcome | Dilshad Garden | 2257.00 |
| 27 | Vishwavidyalaya | Maidan Garhi | 2251.00 |
| 28 | South Extension | New Colony | 2230.00 |
| 29 | Rohini | GTB Nagar | 2211.00 |
| 30 | Welcome | Jama Masjid | 2200.00 |
| 31 | Adarsh Nagar | Gurgaon Sector… | 2197.00 |
| 32 | Badshahpur | Samaypur Badli | 2196.00 |

# 7.Revenue & Payments

## How is revenue distributed across payment methods?

```
Select
paymentMethod,
to_char(sum(Amount),'"Rs "999,999,999.00') as Revenue
From uber
where bookingStatus = 'Completed'
group by paymentmethod
order by revenue Desc;
```

## Result:-

| paymentmethod character varying (20) | revenue text |
|---|---|
| 1 | UPI | Rs  20,986,521.00 |
| 2 | Cash | Rs  11,631,648.00 |
| 3 | Uber Wallet | Rs  5,623,581.00 |
| 4 | Credit Card | Rs  4,709,814.00 |
| 5 | Debit Card | Rs  3,779,964.00 |

# 8.Risk & Loss Analysis

## What are the main reasons for incomplete rides and their percentages?

```
select
incompleteReason, count(*) as total_Reason,
concat(round(count(*) * 100.00 / sum(count(*)) over() ,2),' %') as
Percentage
from uber
where incompleterides = 1
group by incompleteReason ;
```

## Result:-

| incompletereason character varying (100) | total_reason bigint | percentage text |
|---|---|---|
| 1 | Customer Demand | 3023 | 33.78 % |
| 2 | Other Issue | 2932 | 32.77 % |
| 3 | Vehicle Breakdown | 2993 | 33.45 % |

## Analysis Sections Covered

1. Business Performance Analysis
2. Time-Based Demand Analysis
3. Vehicle Performance & Revenue Contribution
4. Customer Behavior & Spending Patterns
5. Driver Reliability & Cancellations
6. Location Intelligence
7. Revenue & Payment Method Analysis
8. Risk & Loss Analysis

---

## Key Insights

- A significant percentage of revenue loss comes from customer and driver cancellations
- Peak demand occurs during evening and night hours
- Certain vehicle types contribute disproportionately higher revenue
- Top pickup-drop pairs have significantly higher average fares
- Incomplete rides highlight operational gaps such as driver unavailability

---

## Conclusion

This project demonstrates the ability to apply SQL for real-world business problem solving, convert raw ride data into actionable insights, and communicate findings in a structured analytical format.
It reflects skills required for Data Analyst / Business Analyst roles in transportation, logistics, and marketplace companies.

---