

# **Project 4: SQL Analysis of Uber's Operational Data**

**By Vyankat Jadhav Sir**

## Problem Statement:

Uber, a global ride-sharing company, operates in many cities worldwide and relies heavily on data to make better decisions. However, Uber faces challenges in managing its rides, payments, drivers, and city-specific issues. To stay competitive and expand into new markets, Uber needs to analyze its data in-depth to address problems like revenue leaks, driver performance, and high cancellation rates. This project will use SQL to explore Uber's operational data and uncover key insights to improve performance and efficiency across different cities.

## Objective

The main goals of this SQL project are to:

1. Analyze Uber's ride data to assess performance in various cities.
2. Study financial data by examining fare trends, cancellations, and payment methods.
3. Evaluate driver performance based on ride counts, ratings, and earnings.
4. Investigate the impact of dynamic pricing and cancellations on revenue.
5. Propose operational improvements using SQL queries and analysis.
6. Implement SQL-based solutions to ensure data integrity and improve query performance.

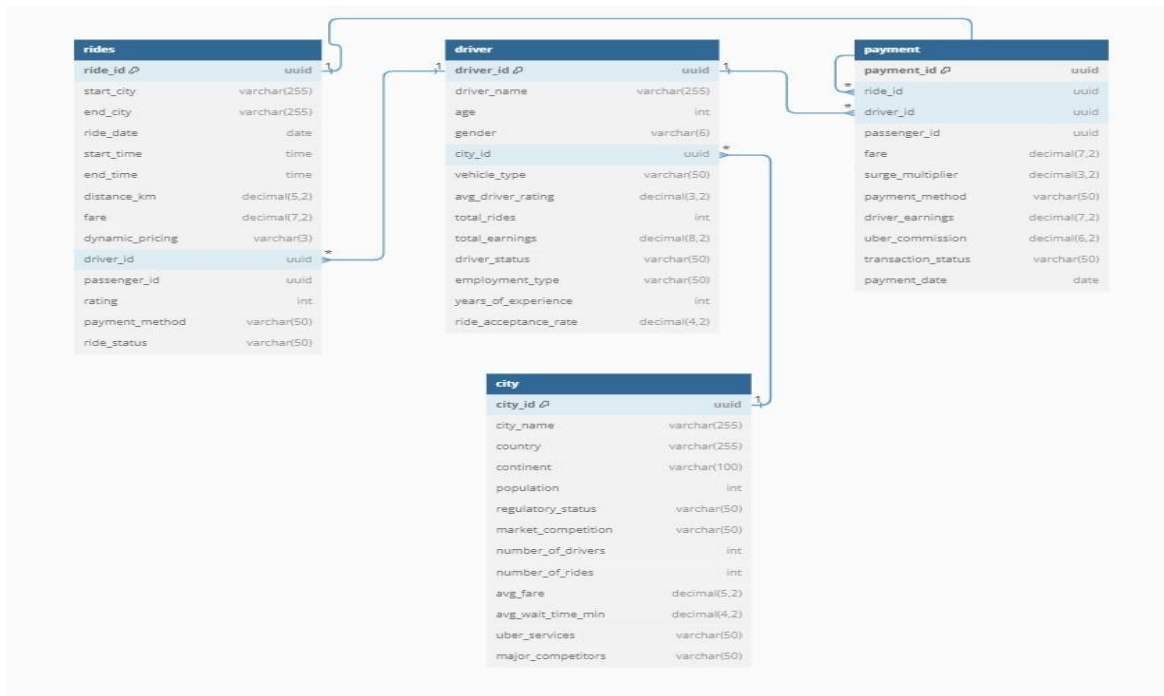
## Data Cleaning

Before analysis, the dataset will undergo a thorough cleaning process to address issues such as:

- Handling missing values in critical columns (e.g., fare, ride\_id, population).
- Resolving duplicate records.
- Ensuring data consistency across tables (e.g., matching driver ratings with actual ride data).

**Dataset-**  [Uber Dataset](#)

## Data Flow Diagram (DFD)



## Capstone Project Questions (SQL-Based)

### City-Level Performance Optimization

Which are the top 3 cities where Uber should focus more on driver recruitment based on key metrics such as demand high cancellation rates and driver ratings?

### Revenue Leakage Analysis

How can you detect rides with fare discrepancies or those marked as "completed" without any corresponding payment?

### Cancellation Analysis

What are the cancellation patterns across cities and ride categories? How do these patterns correlate with revenue from completed rides?

### Cancellation Patterns by Time of Day

Analyze the cancellation patterns based on different times of day. Which hours have the highest cancellation rates, and what is their impact on revenue?

**Seasonal Fare Variations**

How do fare amounts vary across different seasons? Identify any significant trends or anomalies in fare distributions.

**Average Ride Duration by City**

What is the average ride duration for each city? How does this relate to customer satisfaction?

**Index for Ride Date Performance Improvement**

How can query performance be improved when filtering rides by date?

**View for Average Fare by City**

How can you quickly access information on average fares for each city?

**Trigger for Ride Status Change Logging**

How can you track changes in ride statuses for auditing purposes?

**View for Driver Performance Metrics**

What performance metrics can be summarized to assess driver efficiency?

**Index on Payment Method for Faster Querying**

How can you optimize query performance for payment-related queries?