Uber Trips Data Analysis (2009–2015)

Data Analytics & Machine Learning Project

- Basem Hamada Ebrahim
- Mohamed Nabil Elawad
- Mohannad Ashraf Abd Albadea
- Khaled Ashraf Abd Shahed
- Mahmoud Sobhy Hamed Atiaa

Project Overview

This project analyzes Uber trip data from 2009 to 2015 to uncover patterns, understand ride behavior, and build predictive models. It combines data cleaning, exploratory analysis, dashboard development, and machine learning.

Project Objectives

Identify Key Trends

Analyze Uber trip patterns across years, months, and days.

Explore Trip Factors

Understand how weather, distance, and other factors influence ride patterns.

Predict Trip Fare

Build a model to estimate trip fare based on multiple variables.

Interactive Dashboard

Develop a Power BI dashboard for quick visualization of insights.

Dataset Description

- Trip timestamps (day, month, week, year)
- Distance traveled
- Weather conditions
- Car condition indicators
- Road type
- Passenger count
- Pickup & drop-off details
- Fare amount (target variable)

Team Roles

Basem Hamada Ebrahim

Machine Learning Model Development

Mohamed Nabil Elawad

Presentation, Documentation, & Model Evaluation

Mohannad Ashraf Abd Albadea

Data Analysis, Data Visualization & Dashboard Design

Khaled Ashraf Abd Shahed

Data Analysis, Insight Extraction & SQL

Mahmoud Sobhy Hamed Atiaa

Data Cleaning & Feature Engineering

Data Analysis Steps

01

Data Preprocessing

Removed missing, duplicated, and inconsistent values.

02

Feature Engineering

Converted timestamps into analytical features (year, month, day, week).

03

Categorization

Mapped weather and road condition categories.

04

Variable Creation

Engineered new variables like trip speed and time-of-day class.

05

Statistical Analysis

Conducted analysis to identify correlations and ride demand levels.

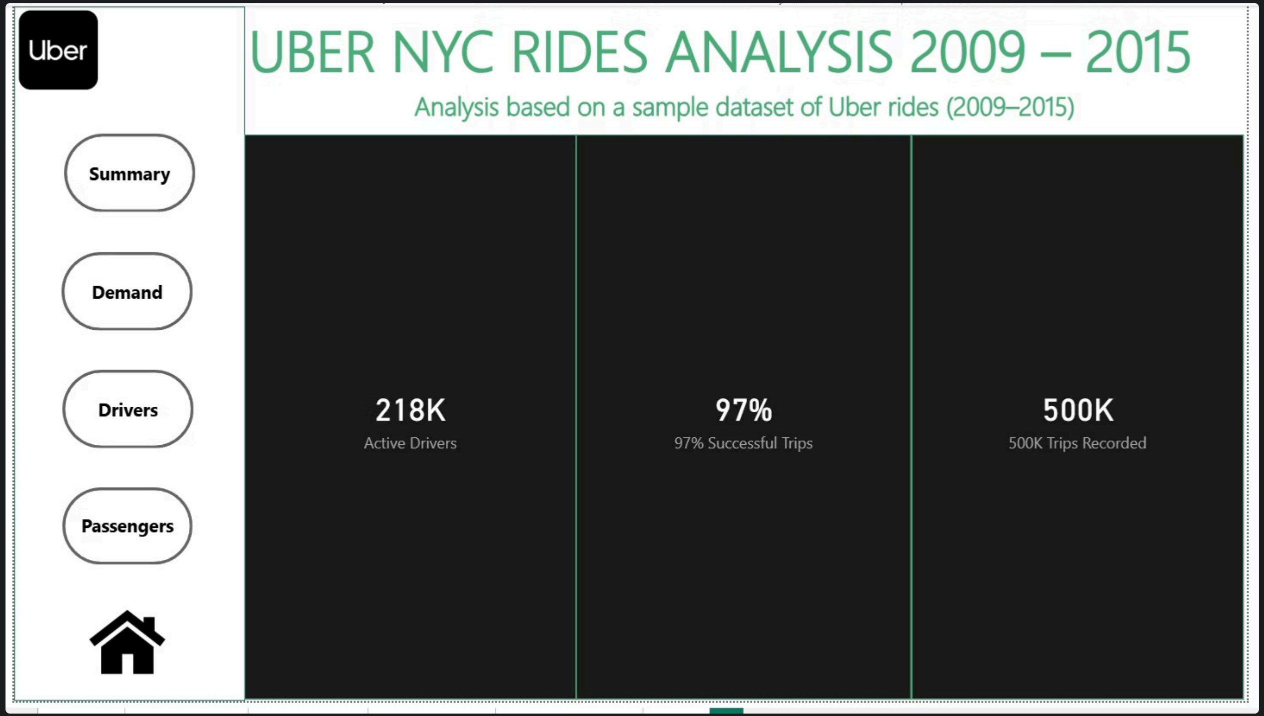
Power BI Dashboard

Our interactive dashboard provides:

- Trip volume trends over time
- Distribution of rides by distance, time, weather, and road type
- Passenger behavior insights
- Revenue patterns across months and years
- Filters for custom exploration

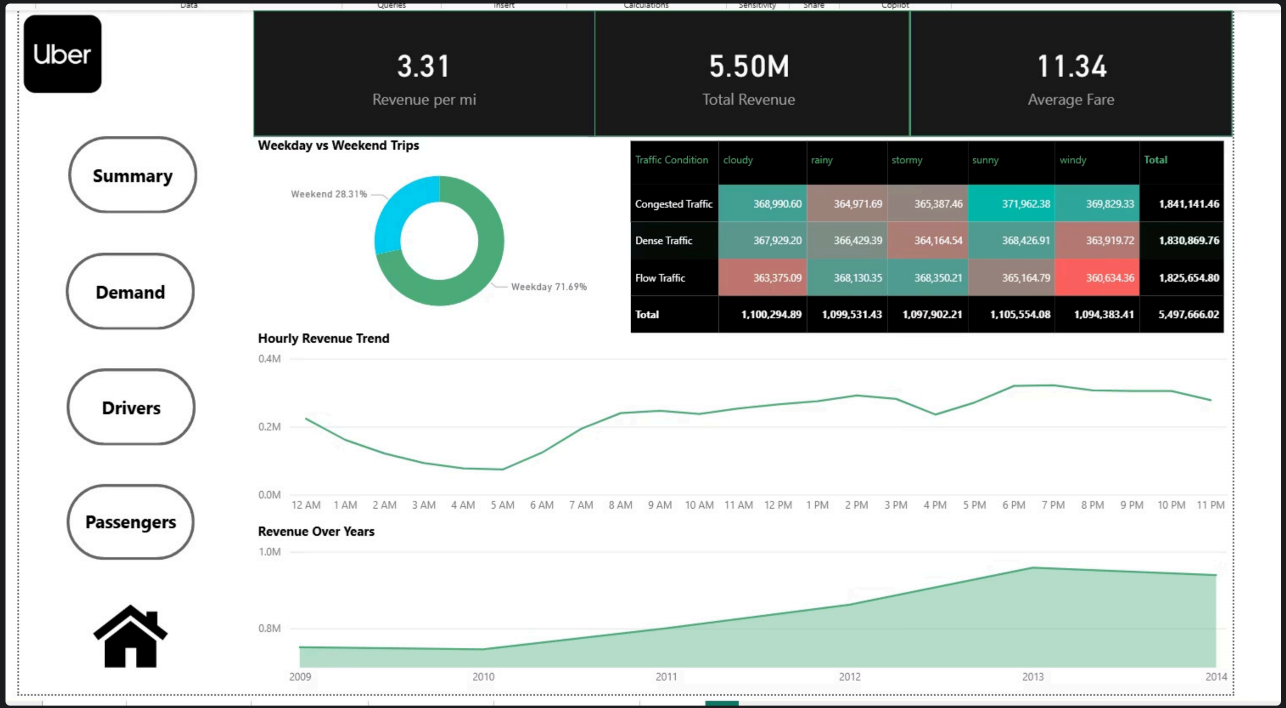
This dashboard enables stakeholders to understand ride dynamics quickly and clearly.

Key Dashboard Insight: Trip Volume Trends



Our Power BI dashboard prominently displays the evolution of Uber trip volume. The chart illustrates a consistent year-over-year growth in ridership from 2009 to 2015, with noticeable seasonal fluctuations. This insight is critical for understanding demand patterns and informing future service adjustments.

Key Dashboard Insight: Monthly Demand Patterns



Our interactive dashboard reveals monthly patterns in Uber trip demand, highlighting seasonal fluctuations and peak ridership periods. These insights support better resource allocation, improved driver availability, and more efficient operational planning. Understanding monthly demand also enables proactive decision-making for marketing and service improvements.

Machine Learning Model

We developed a predictive model to estimate Uber trip fare using factors such as distance, weather, car condition, road type, passenger count, and time variables.

Key Factors

Distance, weather, car condition, road type, passenger count, and time variables.

Increased Accuracy

Clean input features significantly improved model accuracy.

Fare Pricing Insights

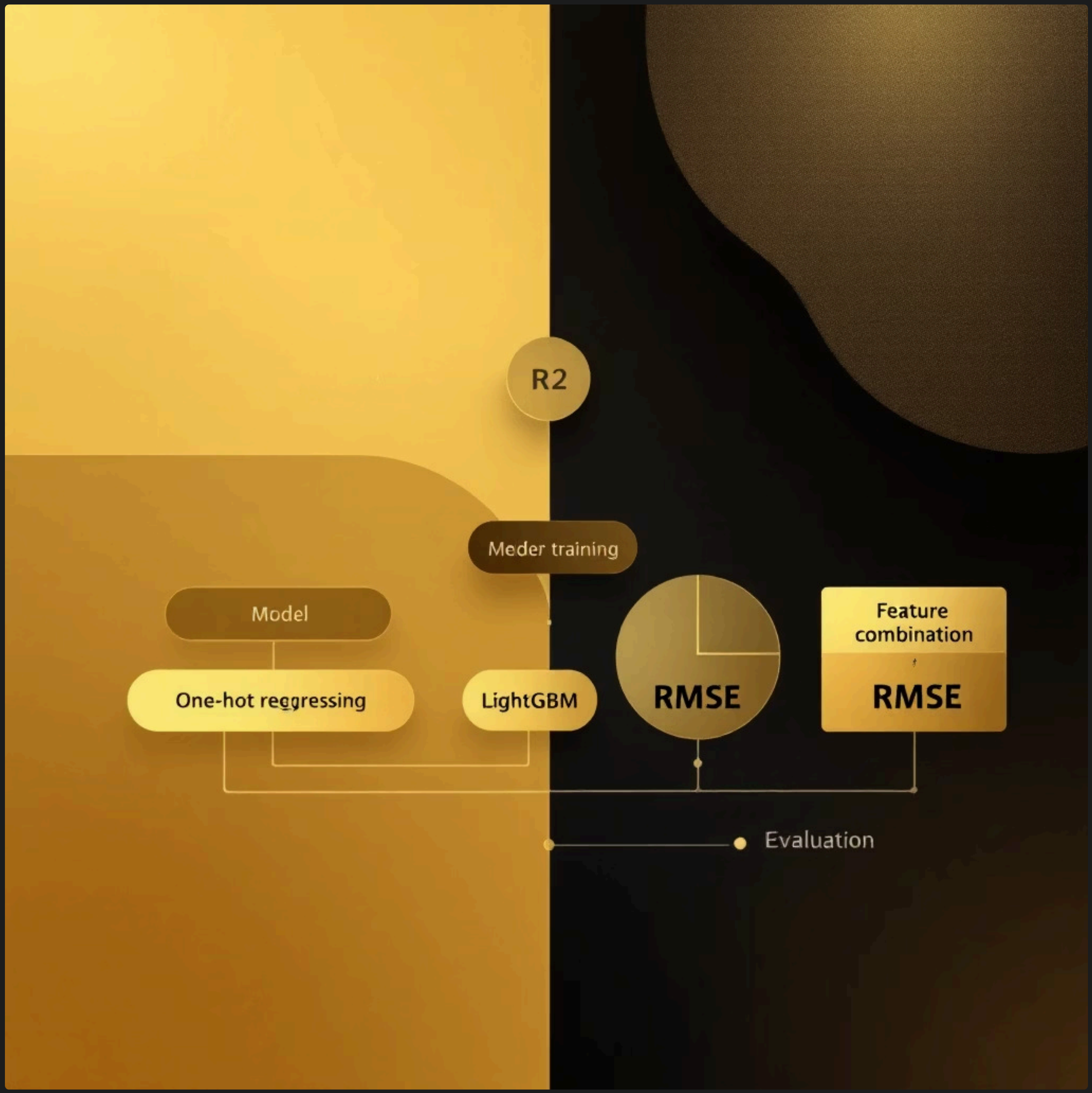
Model explains key contributors to fare pricing.

Optimization Tool

Useful for pricing optimization and cost prediction.

Data Preprocessing & Model Training

This section prepares the dataset for machine learning by applying one-hot encoding to categorical features and combining them with numerical variables. After preprocessing, the LightGBM regression model is trained to predict fare amounts. Model performance is evaluated using RMSE and R^2, followed by generating an Actual vs. Predicted visualization.



Our Forest Model & Prediction Interface

This part introduces the our Forest regression model used to predict fare values based on processed input features. It also showcases the interactive Gradio interface built to allow real-time fare prediction using user-selected conditions such as weather, traffic, and trip coordinates.

SQL Analysis

Database Queries & Data Extraction

SQL Code

This SQL script cleans the dataset, removes invalid trips, and generates key insights such as trip volume, revenue, average fares, distance patterns, driver performance, and time-based trends. It also includes filters for specific trips, locations, and drivers to support deeper analysis.

SQLQuery2.sq...\khale (54))

```sql

--Trip count and fare statistics under each weather condition

SELECT
    Weather,
    COUNT(*) AS trips,
    AVG(fare_amount) AS avg_fare,
    AVG(fare_amount / distance) AS avg_fare_per_km
FROM clean_trips
GROUP BY Weather
ORDER BY trips DESC;

--Trip volume and fare metrics grouped by traffic condition

SELECT
    Traffic_Condition,
    COUNT(*) AS trips,
    AVG(fare_amount) AS avg_fare,
    AVG(fare_amount / distance) AS avg_fare_per_km
FROM clean_trips
GROUP BY Traffic_Condition
ORDER BY trips DESC;

--Average fare per km for each weather + traffic condition combination

SELECT
    Weather,
    Traffic_Condition,
    COUNT(*) AS trips,
    AVG(fare_amount / distance) AS avg_fare_per_km
FROM clean_trips
GROUP BY Weather, Traffic_Condition
ORDER BY Weather, Traffic_Condition;
```

Key Findings

→ **Distance is Key**

Distance remains the strongest predictor of trip fare.

→ **Weather Impact**

Bad weather and poor road conditions increase average fare.

→ **Peak Demand**

Higher demand appears during morning and evening peaks.

→ **Seasonal Spikes**

Certain months and holidays show significant fare spikes.

→ **Reliable Predictions**

Machine learning model shows consistent and reliable predictions.

Future Development Ideas

- Integrate real-time APIs for live price prediction.
- Create driver performance monitoring dashboards.
- Expand model to predict trip duration and demand forecasting.
- Deploy the ML model as a web app or mobile interface.
- Add customer segmentation for marketing optimization.
- Develop predictive alerts for surge pricing and peak demand.


THANK YOU!