

Analyzing Uber Ride Demand with Machine Learning

Rimsha Shafique
2024-MSDS-03
Presented To: Dr. Awais

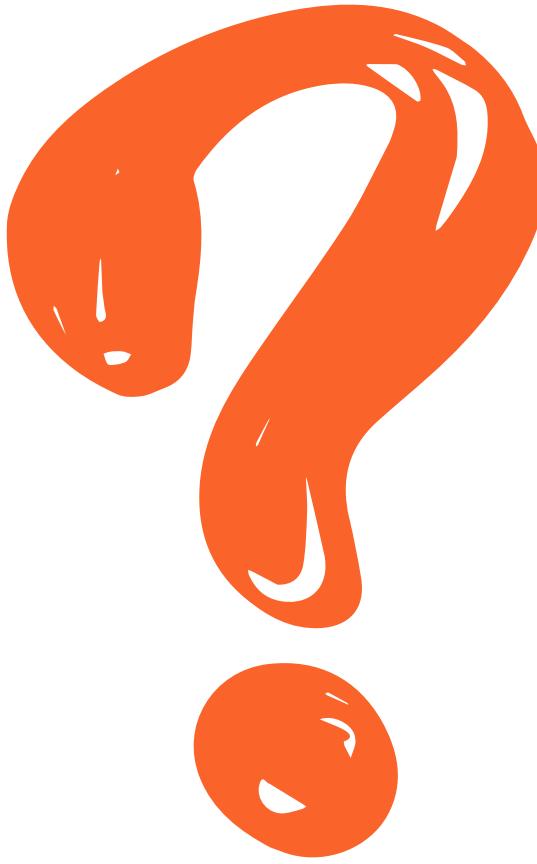


Uber Overview:

- Global leader in ride-hailing, courier, and food delivery
- Operates in 70 countries, 10,500 cities
- 150M monthly users, 6M drivers, 28M daily trips
- 47B trips since 2010, \$37.2B revenue (2023)

Problem Statement

Understanding and predicting **ride demand patterns** is crucial for optimizing Uber's operations. Accurate demand forecasting helps improve driver allocation, reduce wait times, and enhance customer satisfaction, ultimately driving operational efficiency and profitability.



Research Questions

- 01 In which category do people book Uber rides the most?
- 02 What are the most common purposes for booking rides?
- 03 How far do people typically travel using Uber?
- 04 At what times and days do people book rides most?
- 05 Can we predict high-demand periods?

DATASET



Sourced from GitHub.



7 original attributes:

Start Date, End Date, Start Location,
Stop Location, Miles, Category,
Purpose



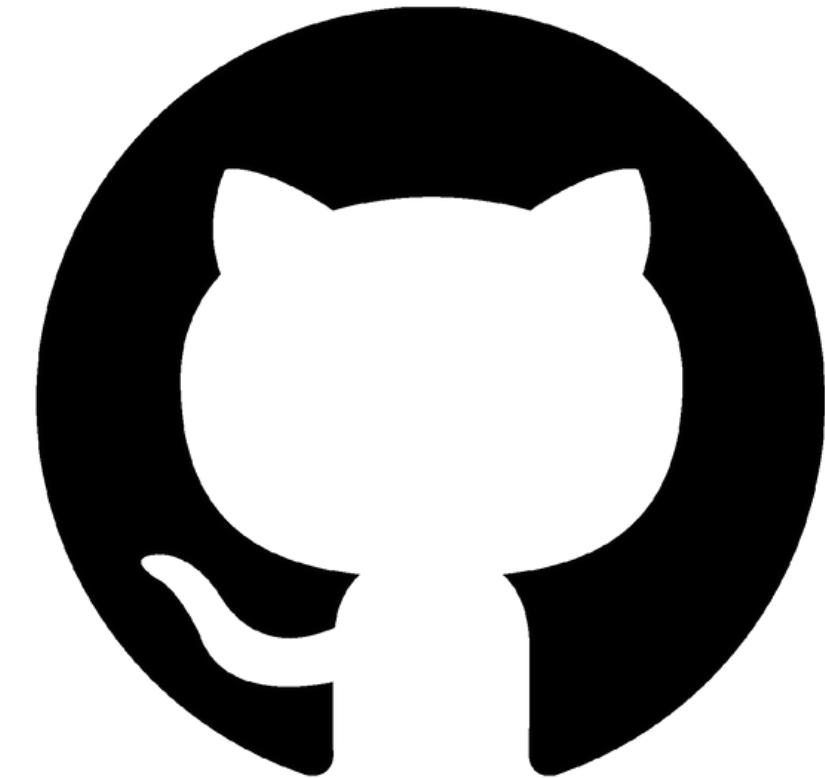
Additional attributes created during preprocessing:

Hour, Day, Month, Demand Level.



Total rides analyzed:

1156



GitHub

Data Wrangling

Missing Values

Replaced nulls in Purpose with "Not Defined."

Date Conversion

Transformed Start Date and End Date into datetime formats.

Feature Engineering

Extracted time-based features like Hour, Day, and Month

Demand Classification

Defined "Demand Level" based on ride count thresholds (e.g., 75th percentile).

Visual Analysis

Created visualizations for ride distribution by time and day.

Predictive Modeling

Selected Model

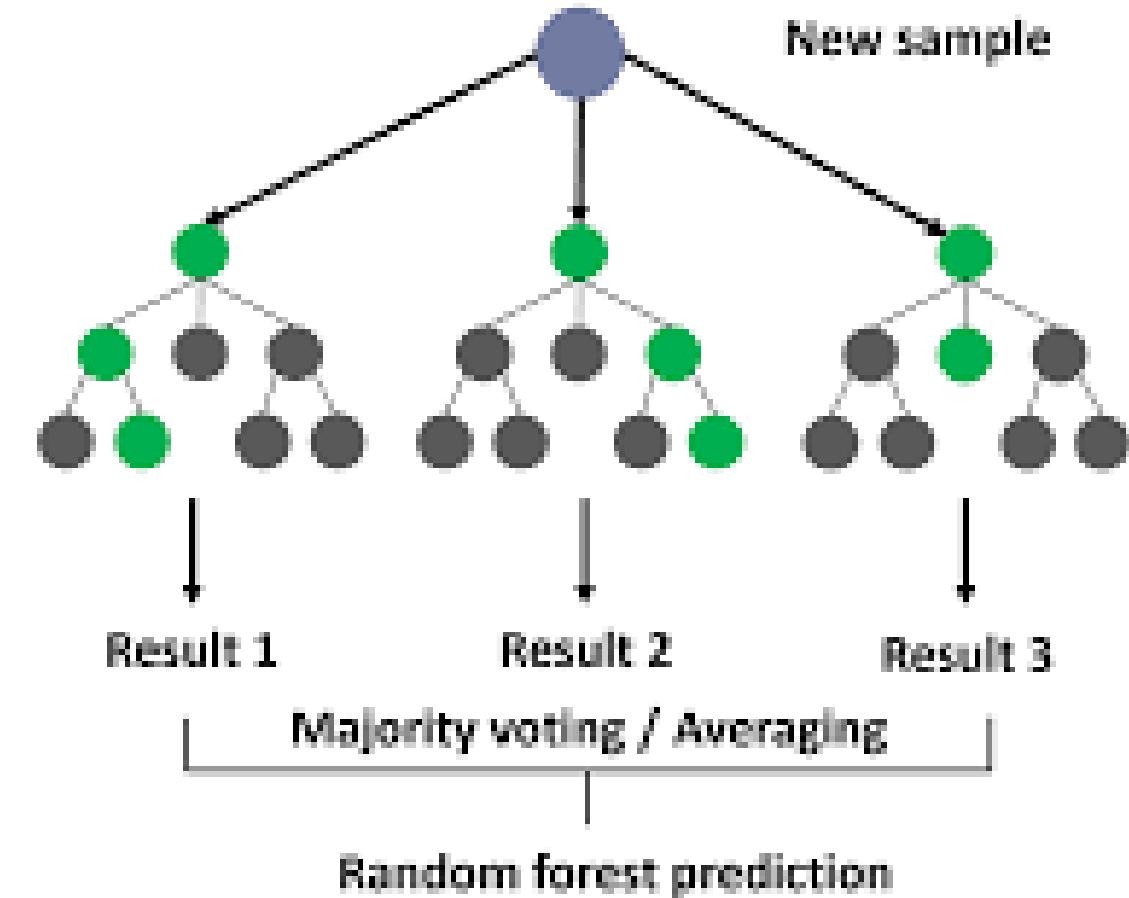
Random Forest Classifier for its robust performance on tabular data.

Data Splitting

Divided the dataset into training and testing set

Evaluation Metrics

Assessed model performance using Precision, Recall, F1-Score, and Accuracy.

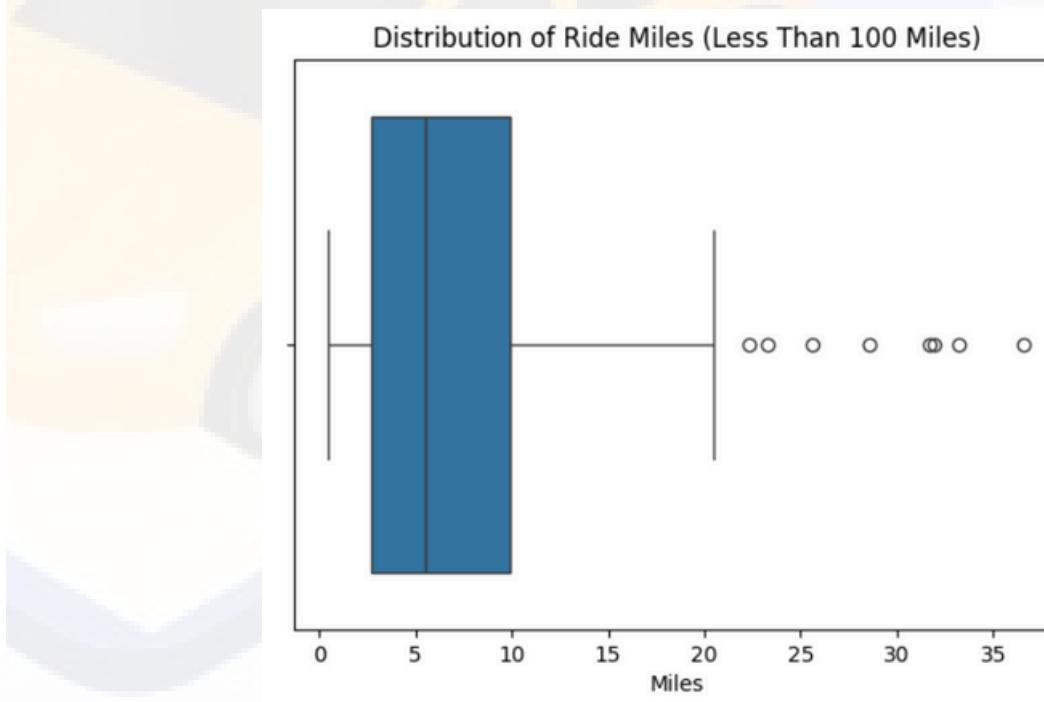
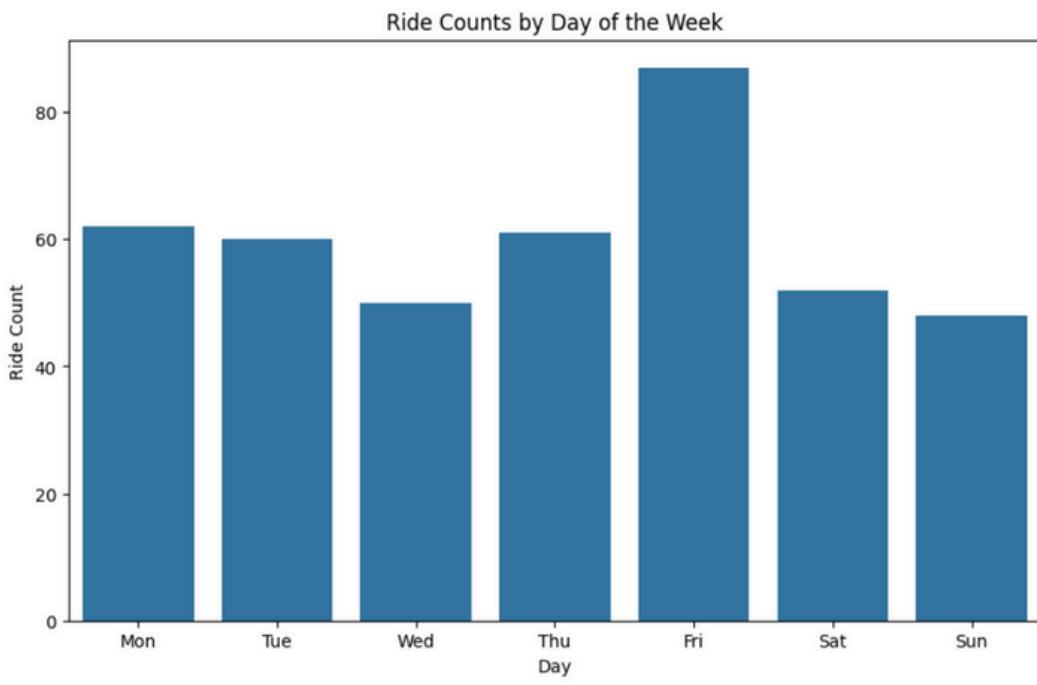
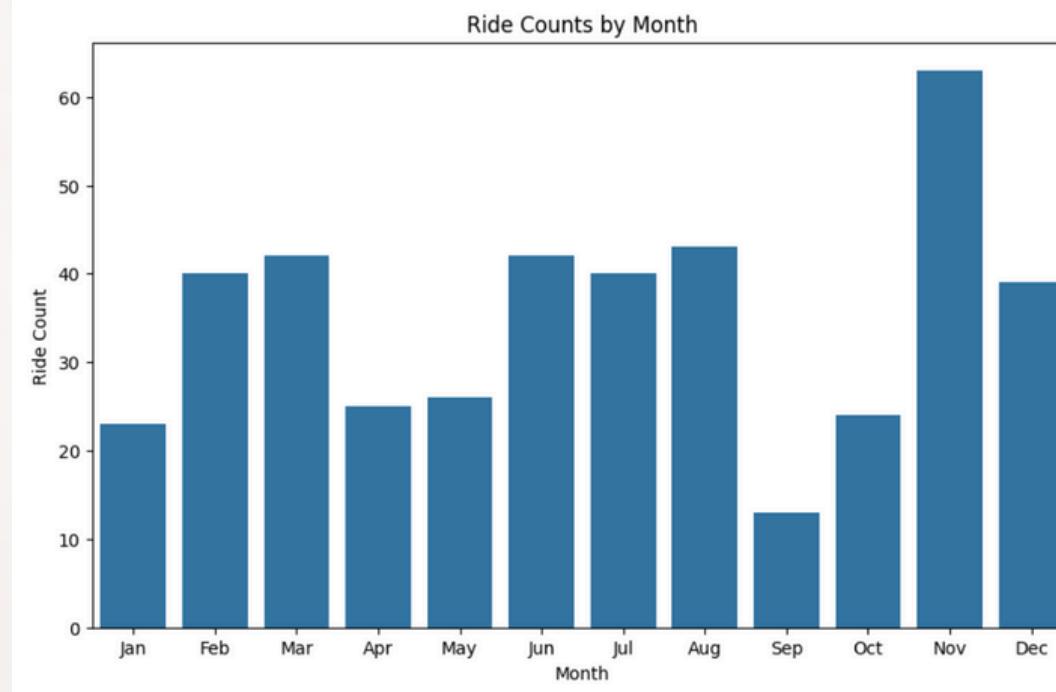
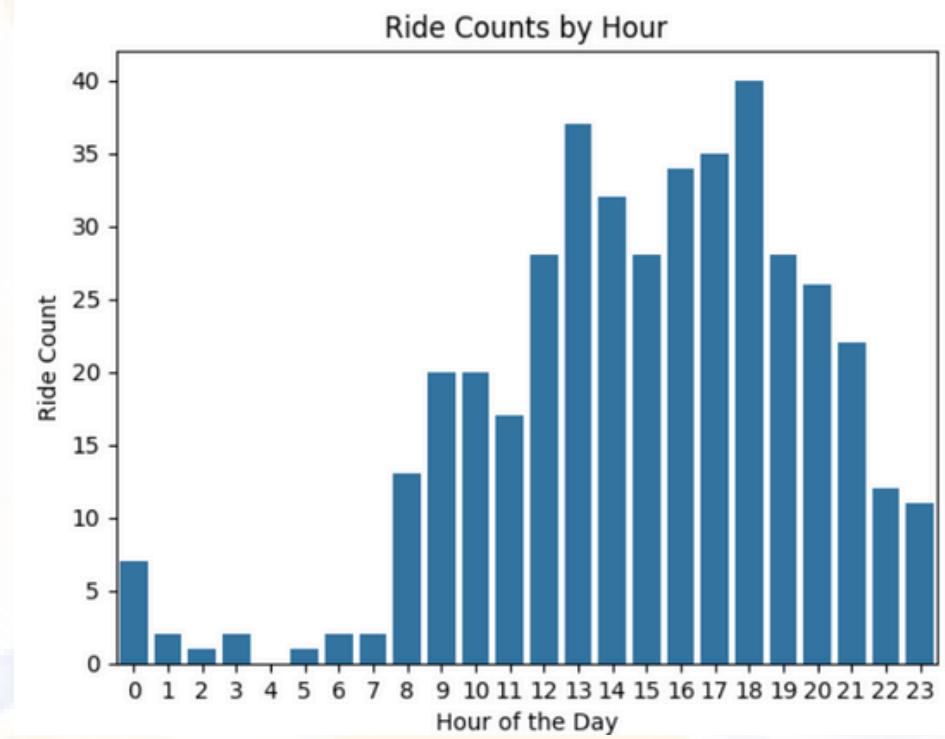
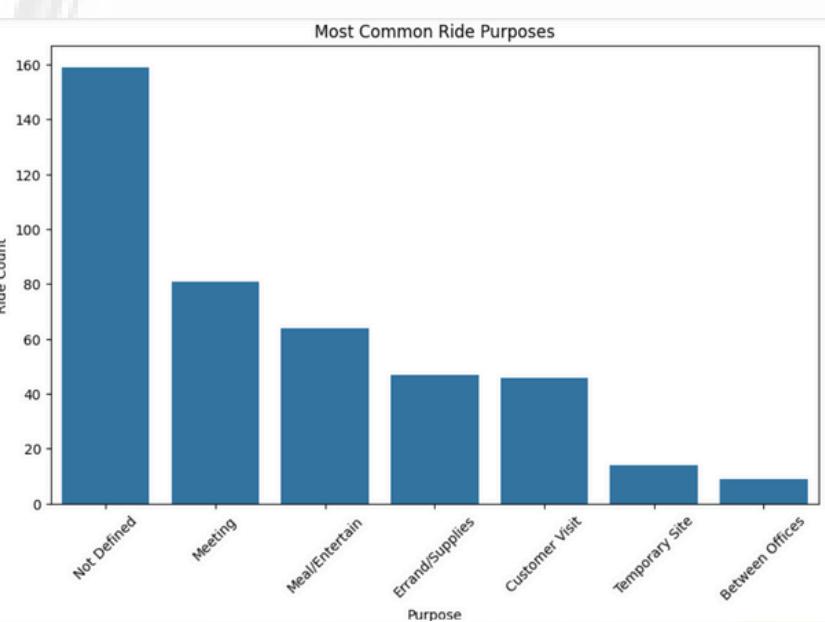
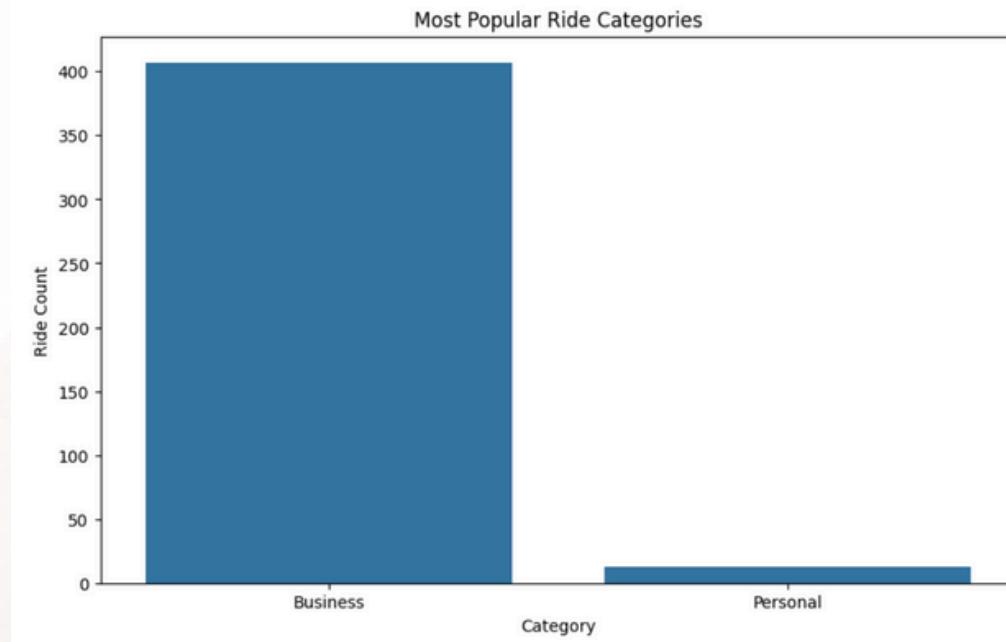


Results

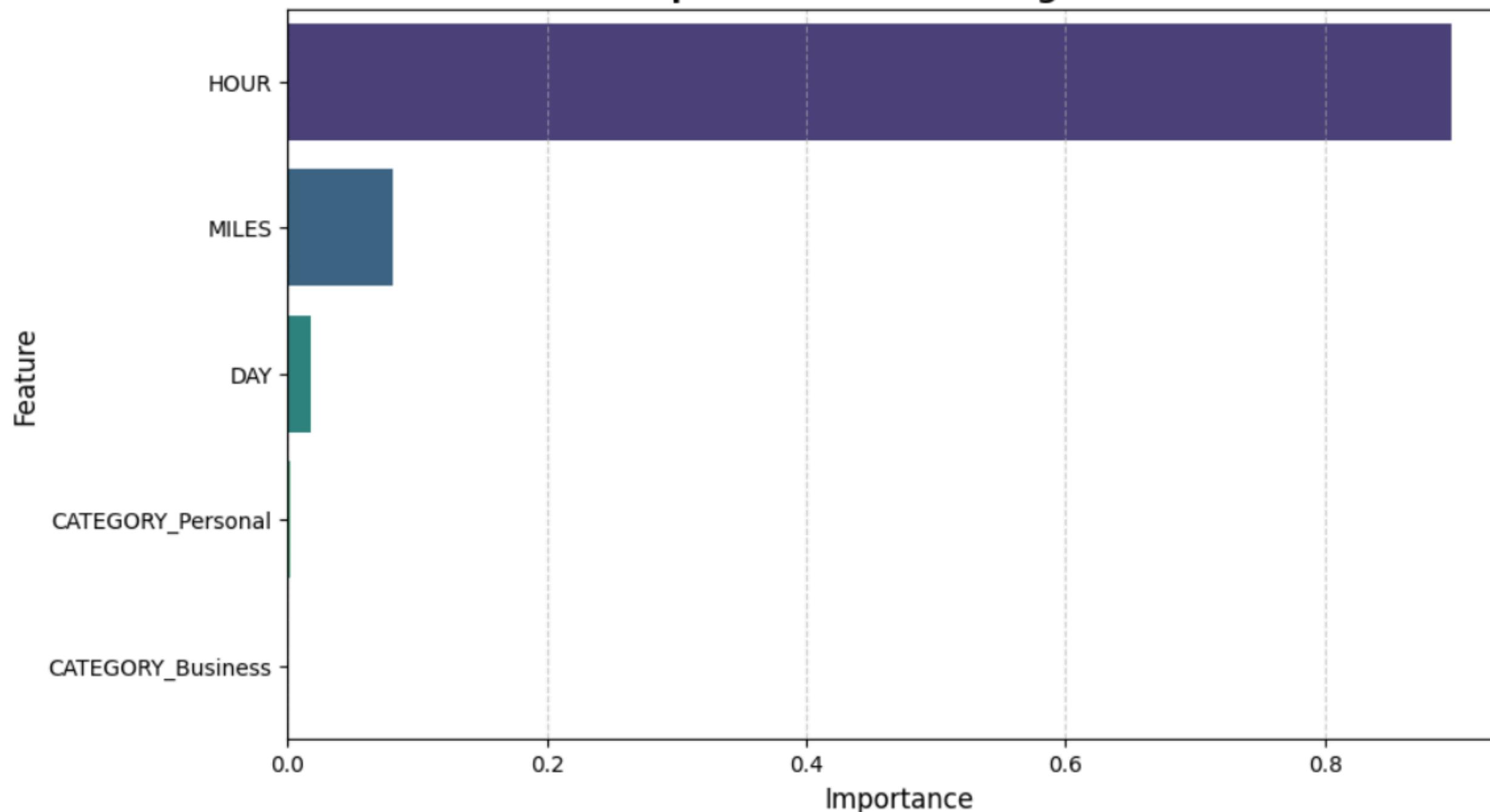


- 01 Category: Most rides are for Business purposes.
- 02 Purpose: Common purposes include Meetings
- 03 Distance: Most trips are under 10 miles.
- 04 Peak Times: Afternoon hours (2-6 PM) and Fridays see the highest demand.
- 05 High-Demand Prediction: Achieved 100% accuracy using the Random Forest model.

Visualizations



Feature Importance in Predicting Demand Level



	precision	recall	f1-score
0	1.00	1.00	1.00
1	1.00	1.00	1.00
accuracy			1.00

Conclusion

In conclusion, a Random Forest Classifier achieved 100% accuracy in predicting Uber ride demand, revealing peak demand on Fridays and afternoons, driven by business rides (especially meetings), with lower demand in the month of September. This demonstrates the potential of machine learning to optimize ride-hailing operations.

A row of yellow taxis is parked on a city street. In the background, there's a building with a sign that reads "ΜΑΤΣΟΙ ΦΑΣΙΣΤΕ".

Thank you!
