

Uber Trip Analysis (NYC 2014)

Generated on: June 11, 2025

1. Introduction

This project analyzes and forecasts Uber trip demand using 2014 data from New York City. The aim is to identify trip trends and build predictive models that can help understand future demand. We use machine learning techniques and time series decomposition to improve forecast accuracy.

2. Dataset

The dataset includes Uber pickup data from April 2014 to September 2014. Each record includes a timestamp and coordinates of the pickup location.

Files used:

- uber-raw-data-apr14.csv
- uber-raw-data-may14.csv
- uber-raw-data-jun14.csv
- uber-raw-data-jul14.csv
- uber-raw-data-aug14.csv
- uber-raw-data-sep14.csv

3. Exploratory Data Analysis (EDA)

We analyzed trip volume patterns by hour and day of the week. Key insights:

- Peak hours are between 5 PM and midnight.
- Fridays and Saturdays show the highest volume.
- Weekdays have consistent morning and evening demand peaks.

4. Forecasting Models

Uber Trip Analysis (NYC 2014)

We implemented three time series forecasting models using lag-based features:

- XGBoost Regressor
- Random Forest Regressor
- Gradient Boosting Regressor

We also built an ensemble model using weighted MAPE-based blending.

5. Model Evaluation

Model	MAPE
----- -----	
XGBoost	~8.0%
Random Forest	~9.5%
Gradient Boosting	~10.0%
Ensemble	~8.2%

XGBoost provided the lowest error, and the ensemble model slightly improved consistency.

6. Conclusion

Uber demand shows strong cyclical patterns based on hour and day of week.

Predictive modeling using time series lags and tree-based algorithms yields high forecasting accuracy.

This analysis can be useful for business planning, driver allocation, and marketing strategies.