Project 3 Proposal: Understanding and Forecasting NYC Taxi and Ridehailing Trip Patterns

Group Members:

Mauricio Monje

Sebastian Mejia

Aditya Dwivedi

Our goal is to analyze transportation patterns using open NYC taxi tata

Research Question

How have taxi and ridehailing usage patterns in New York City changed in recent years? We wish to explore whether we can find trends to build reliable models that predict future demand across different service types (Yellow Taxi vs Uber/Lyft), and identify shifts in usage patterns like changes in trip frequency, vehicle activity, or payment method?

Understanding how people use transportation services over time helps city planners, transit agencies, and ride platforms adapt to changing rider needs. Our group will explore how ride volumes, service types, and operational metrics have shifted, and we will build models to forecast future usage based on historical trends.

Related Research Summaries

1. **Safikhani et al. (2017). "Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling"**
   Compared different time series models (VAR, STAR, LASSO-STAR) for forecasting Uber demand in NYC. Found that spatial-temporal models gave much better accuracy, especially during busy times.
2. **Correa, Xie & Ozbay (2021). "Exploring the Taxi and Uber Demand in New York City"**
   Analyzed demand trends for both taxis and Uber using 2014–2015 data. Found Uber trips increased rapidly while taxi trips declined. Showed spatial models helped explain demand using land use and income data.
3. **Toman et al. (2020). "Spatiotemporal Analysis of Ridesourcing and Taxi Demand by Taxi Zones in New York City"**
   Modeled demand per taxi zone and showed that spatial correlation was still strong in residuals. Suggested future work should include spatial effects when forecasting.
4. **Acharya & Quadrifoglio (2024). "Forecasting Ride-Hailing Demand in Urban Areas"**
   Built deep learning models (CNN-LSTM + XGBoost) for forecasting Yellow Cab trips. Improved accuracy by grouping time series into similar clusters and using more features.
5. **Geary (2017). "Estimating Demand for Taxis at LaGuardia Airport"**
   Forecasted taxi pick-ups at the airport using flight data, weather, and time info. LSTM model performed best, showing deep learning works well for location-specific taxi demand.

Planned Experiment

We will use NYC taxi and ridehailing data from 2010 to the present to explore how usage has changed over time. The dataset includes monthly summaries like average daily trips, active vehicles, fare totals, credit card usage, and more.

We plan to:

- Plot trends in daily trips by service type (Yellow, Green, Uber/Lyft).
- Compare changes in trip volume, driver participation, and payment method.
- Use time-series models (like Prophet or ARIMA) to forecast trip demand.
- Explore whether adding clustering or deep learning improves predictions.
- Compare performance across service types and time periods.

Dataset(s) to Be Used

- Monthly Aggregated TLC Trip Data: Official CSV files from the NYC Taxi & Limousine Commission's Aggregated Reports page. Includes monthly metrics on trip volume, fares, vehicle activity, and more: https://www.nyc.gov/site/tlc/about/aggregated-reports.page

Expected Contribution

This project will show how NYC taxi and ridehailing usage has shifted in recent years. We will:

- Forecast future trip volumes using time-series and machine learning models.
- Compare travel and service trends across Yellow Taxis, Green Taxis, and FHVs.
- Test whether adding clustering or spatial features improves prediction.
- Share visualizations that highlight long-term shifts in NYC ride behavior.