EBA 5005 Practice Module in Specialized Predictive modelling & Forecasting

Proposal for Optimizing Taxi Economics: A Data-Driven Approach to Pricing and Resource Management in NYC

Date of Proposal

27 Feb 2025

Team Name or Group No

Group 1

Team Members

Ashish Nagineni (A0297393L)

Koh Zhen Wen Ian (A0033871H)

Li Yingzhen (A0297776B)

Mohammed Muzaffar Ali(A0297946A)

Zhu Wenqing (A0269734N)

1. Introduction

The taxi industry has undergone significant transformation in recent years due to the rise of rideshare services and evolving consumer preferences. The NYC Taxi and Limousine Commission (TLC) has engaged us as consultants to assess the current state of the taxi industry and recommend data-driven solutions to address the emerging challenges.

This project addresses the challenges of unoptimized pricing and driver allocation, leading to demand-supply imbalances. We aim to provide a comprehensive overview of the current state and future trajectory of New York City's taxi industry, while also offering data-driven solutions to optimize supply-demand dynamics, improve operational efficiency, and enhance the overall customer experience.

2. Industry Overview

New York City's taxi industry is a complex system comprising yellow medallion taxis, green borough taxis, and the dominant for-hire vehicle (FHV) sector, primarily rideshare companies like Uber and Lyft. Each segment plays a distinct role and faces unique challenges.

The industry faces several key challenges. The rise of FHVs has created intense competition, leading to a sharp decline in yellow cab ridership and a collapse in medallion values, triggering a medallion debt crisis for many owners. The COVID-19 pandemic further compounded these issues, drastically reducing ridership and revenue, forcing some drivers into delivery services.

The NYC taxi industry has undergone a dramatic transformation. While challenges remain, the industry is adapting through innovation, technology, and policy adjustments. The future will likely involve a mix of traditional taxis, rideshare services, and other options, creating a dynamic urban transportation ecosystem.

3. Business Problem & Objectives

Key Business Problem:

- Inefficiencies in supply-demand matching in NYC taxi industry
- Unoptimized pricing and driver allocation causes demand-supply imbalances.
- Demand elasticity aids dynamic pricing, while forecasting and simulation can improve fleet distribution and wait times.

A. Pricing Optimisation:

Business Objectives:

- Increase Revenue: By strategically adjusting prices based on demand elasticity and market dynamics, we aim to maximize overall revenue generation.

Technical Objectives:

- Find the demand elasticity to understand how price impacts demand.
- Perform linear programming to optimize the price for maximum revenue.

B. Driver Resource Optimal Allocation:

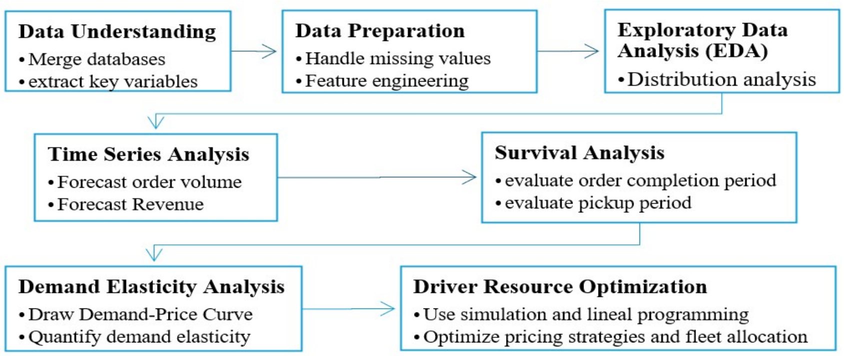
Business Objectives:

- Improve Supply-Demand Matching: Better matching of taxi availability with predicted demand, minimizing wait times for customers and reducing the number of idle taxis on the road.
- Optimized Dispatching: Streamlining taxi fleet dispatch based on demand forecasts, ensuring more efficient routes and minimizing unnecessary taxi congestion in low-demand areas.
- Fleet Size Optimization: Fine-tune the total number of taxis and drivers in the fleet to match demand, ensuring optimal resource utilization and minimizing inefficiencies.

Technical Objectives:

- Use time series machine learning algorithms to predict future demand, thus allowing us to allocate resources for demand based on location, time and season.
- Perform survival analysis to estimate the wait times and related probabilities. This allows us to understand current situation and improve upon it.
- Finally, we will perform optimisation and simulation. By utilizing Monte Carlo simulation, we can optimise the driver allocation based on time and location.

4. Project Design



5. Scope of work

The dataset used in this project is sourced from NYC government website. The datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP).[1]

The dataset provides information about taxi trips, with each entry containing price, time and location data of each taxi ride. Dataset is large (~4GB per year), historical and structured.

Potential data preparation and transformation that would be performed are as follows:

- Union and joining of tables
- Removing fields which are irrelevant
- Data cleaning and normalisation
- Data aggregation and binning
- Preparing data dictionary
- Preparing entity relational diagram

A. Pricing Optimisation

To enhance revenue generation, a robust pricing optimization model will be developed. This model will leverage a dual analytical approach to achieve its objectives.

Methodology

The pricing optimization strategy will be built on two key analytical techniques:

- **Demand Elasticity:** Analyse the relationship between order volume and price per kilometre to understand how changes in price impact demand.
- **Linear Programming:** Optimize the taxi price, maximising revenue.
- ✓ **Evaluation criteria:** Expected increase in revenue in across board and per geographical zone after price optimization

B. Driver Allocation Optimisation

Enhancing efficiency and profitability through demand forecasting, wait time prediction, and optimized driver allocation using linear programming and Monte Carlo simulation.

Methodology

1) Forecasting Order Volume, Demand & Taxi Revenue

- **Time Series Forecasting:** Utilizing statistical models like ARIMAX, RNN and LSTM networks to predict future demand patterns based on historical data.
- ✓ **Evaluation criteria:** Accuracy metrics (e.g. MAE, RMSE).

2) Wait time Prediction

- **Survival analysis:** Utilize Kaplan-Meier estimation, Cox Proportional Hazards Model, and Accelerated Failure Time (AFT) models to predict time to order completion and evaluate influencing factors. Estimate the probability of an order being completed within a given time frame and identifying risks leading to delays.
- ✓ **Evaluation criteria:** Accuracy metrics (e.g. C-Index).

3) Optimisation and Simulation

- **Simulation**: Use Monte Carlo simulation to model passenger demand across regions, accounting for other factors (demand forecasting, order time predictions and driver locations).
- **Linear Programming:** Optimize driver allocation through linear programming based on regional demand fluctuations and fleets constraints to maximize profit.
- ✓ **Evaluation criteria:** % increase in revenue, minimised wait time after optimization.

6. Key Deliverables

A. Pricing Optimisation:

Pricing optimization model – Maximise Revenue by adjusting Taxi fare pricing

- Demand elasticity analysis to determine the relationship between order volume and price per kilometer.
- Linear programming model for optimizing price adjustments based on demand forecasts.

B. Driver Allocation Optimisation:

1. Time Series Predictive Model - Forecast future taxi order volumes and demand across various time periods and locations.
    - Forecasting model results, segmented by time of day and location.
    - Linear programming optimization model

2. Survival analysis model - Predict order completion efficiency, considering location, time periods, and other operational factors.
    - Survival analysis results showing order completion probabilities across various conditions (e.g., time of day, route types).
    - Efficiency prediction metrics, such as expected completion time and probability of successful completion.
3. Driver Resource Optimization
    - Demand and Supply model to determine optimal taxi supply.
    - Visualization dashboard with geographical segmentation: A dynamic dashboard that visually represents geographical demand patterns and demonstrate simulation.

7. Effort Estimates and Timeline

Task	WEEK1	WEEK2	WEEK3	WEEK4	WEEK5	WEEK6	WEEK7	WEEK8
	Mar3-9	Mar10-16	Mar17-23	Mar24-30	Mar31-37	Apr7-13	Apr14-20	Apr21-23
Data Understanding								
Data Preparation								
Exploratory Data Analysis								
Time Series Analysis								
Survival Analysis								
Demand Elasticity Analysis								
Driver Resource Optimization								

8. References

1. Taxicab & Livery Passenger Enhancement Programs, NYC Taxi and Limousine Commission (TLC) Trip Record Data, https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page