OPERATIONS ANALYTICS FINAL REPORT

GROUP 5

Anirudh Nair - axn230010
Ankit Patange - axp230040
Jayasree Lakshmi Narayanan -jxl230012
Mahadevan Ramanan - mxr230040
Sai Lakshmi Harini Chundi -sxc23002
Shreya Shetty - sxs230104

Introduction

The project aimed to analyze NYC taxi data, focusing on demand forecasting, price prediction, and operational optimization. This report explores the methodologies employed, challenges encountered, and insights derived, along with the practical applications of each type of analysis.

Dataset Overview

The analysis utilized the TLC Trip Record Data for green taxis, covering four months (May 2024 – August 2024). This dataset captured critical trip details, such as:

- **Temporal Data**: Pick-up and drop-off dates and times.
- **Spatial Data**: Locations of pick-ups and drop-offs.
- **Trip Metrics**: Distance, passenger count, and fare breakdowns.

Key areas analyzed include Northern Manhattan, Brooklyn, Queens, and the Bronx, offering a diverse urban dataset

1. Descriptive and Exploratory Analysis

Objectives:

- To identify temporal and geographical trends in taxi demand and pricing.
- To uncover patterns in fare dynamics and trip characteristics.

Key Findings:

1. **Time Trends**:
    - Peak fares occurred during early morning hours, influenced by limited taxi availability and nightlife demand.
    - Weekend fares showed a slight uptick compared to weekdays.
    - Demand followed a U-shaped pattern: high during rush hours, low during mid-day, and peaking again in the early evening.
2. **Geo Trends**:
    - **High-Fare Zones**: Areas like Schuylerville/Edgewater Park and Newark Airport exhibited high average fares due to long-distance trips.
    - **Moderate Fare Zones**: Locations like Downtown Brooklyn and Central Harlem showed steady demand.
    - **Short-Distance Trips**: Zones such as East Harlem North had lower average fares but higher trip volumes.
3. **Variable Relationships**:
    - **Distance and Fare:** Strong positive correlation ($R^2 = 0.82$), highlighting distance as the primary driver of fares.
    - **Duration and Fare:** Weaker correlation ($R^2 = 0.16$), indicating minimal impact of trip time on fare variability.

- **Temporal Factors:** Pickup hour and day of the week showed weaker but meaningful correlations, reflecting demand behavior and aiding predictive modeling.

Geospatial heatmaps highlighted high-demand zones, including Manhattan neighborhoods (East and Central Harlem) and airports (JFK, LaGuardia). Longer trips from areas like Downtown Brooklyn and Jamaica were associated with higher fares.The strong correlation between trip distance and fare highlighted the significance of trip length in determining fare variability. Meanwhile, temporal variables such as pickup hour and day of the week offered additional insights into behavioral patterns, informing the development of predictive models.

Operational Challenges Addressed: Descriptive analysis highlighted inefficiencies in resource allocation, especially during peak hours, and emphasized the need for dynamic fleet management.

2. Demand Forecasting

Techniques Used:

1. **Random Forest Regression**:
   - **Goal**: Predict hourly pick-up demand for Downtown Brooklyn/MetroTech.
   - **Features**: Pickup location (PULocationID), pickup hour, and day of the week.
   - **Performance**:
     - In-sample R^2: 98.9%
     - Out-of-sample R^2: 93.48%

These findings emphasize prioritizing high-demand zones, aligning resources with peak periods, and leveraging predictive tools to optimize operations and revenue

2. **SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors)**:
   - **Goal**: Forecast demand across multiple zones.
   - **Advantages**: Handled recurring seasonal patterns (e.g., daily or weekly demand trends). Scaled across various locations for broad applicability.

Insights: Peak Demand: Identified between 5 PM and 7 PM, with unmet demand reaching up to 2 rides/hour. Fleet Allocation: Recommended deploying additional taxis during peak hours to address unmet demand.

Operational Challenges Addressed: Forecasting models helped predict service bottlenecks, enabling better resource distribution and minimizing revenue losses.

3. Price Prediction

Technique Used: Employed a regression-based model to predict fare amounts based on key variables like distance, passenger count, and trip type.

Performance Metrics:

- R^2 (Train): 78%
- R^2 (Out-of-sample): 77.82%

Insights: Fare variation was largely explained by trip distance.Outliers in the data (e.g., $600 fares for short distances) highlighted the need for robust data cleaning and anomaly detection.

Operational Challenges Addressed: The pricing model offered a reliable framework for fare estimation, improving customer transparency and helping drivers plan optimal routes.

4. Operational Insights and Recommendations

Fleet Optimization:

1. **Immediate Actions**:
   - Add 2 taxis at 5 PM and 1 taxi at 6 PM to meet peak demand.
   - Strategically position taxis near high-demand zones like airports and major transit hubs.
2. **Long-Term Strategy**:
   - Use dynamic scheduling to adjust fleet sizes based on demand forecasts.
   - Integrate real-time demand data to enhance operational efficiency.

Revenue Maximization: Focus on high-fare zones during off-peak hours to optimize earnings. Balance fleet distribution between short-trip and high-revenue zones.
Customer Experience: Address unmet demand to reduce wait times and enhance satisfaction. Use demand predictions to deploy promotional offers during low-demand periods.

Conclusion

This comprehensive analysis of NYC green taxi data combined descriptive trends, predictive modeling, and operational insights to optimize fleet operations. Key takeaways include:

1. Demand forecasting models provided actionable insights for resource allocation.
2. Price prediction tools enhanced transparency and operational planning.
3. Geo- and time-based analysis highlighted revenue-maximizing strategies.

By leveraging these findings, NYC taxi operators can achieve enhanced service quality, improved profitability, and greater operational efficiency.