

PRES

ENTATION

NEW YORK TAXI OPERATION

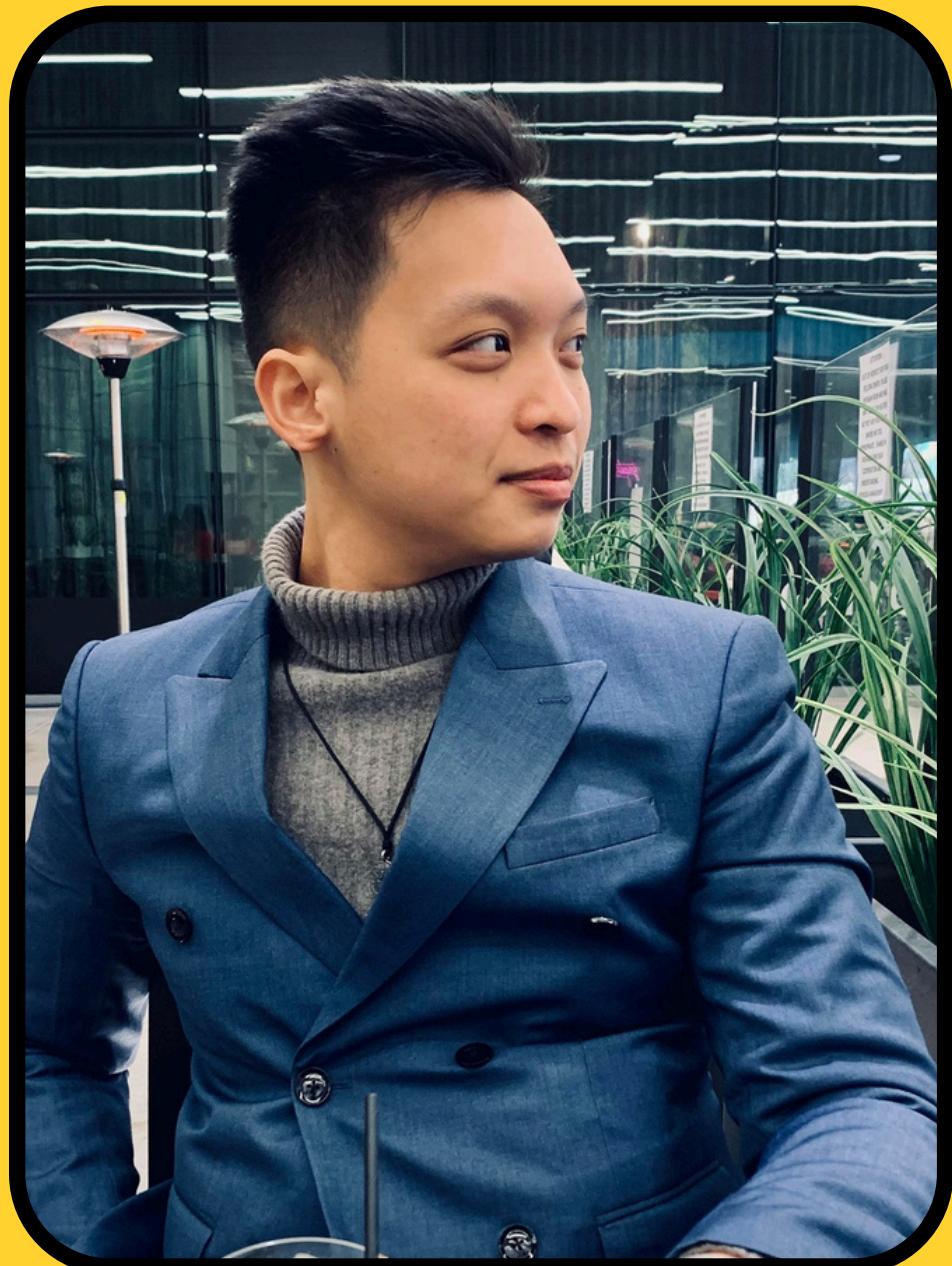
Long Pham - Data Analytics Initial Report

CONTENTS

1. Introduction
2. Data Cleaning and Processing
3. Features with strong relationships
4. Neighborhoods that love Taxi
5. Improving Operations:
 - a. Traffic Jams?
 - b. What time and where do New Yorkers need Taxi?
 - c. What day to increase cabs?
A hint about holidays
6. Model with 20% tips:
7. What's next?
8. Question and Answers



1. INTRODUCTION



LONG PHAM

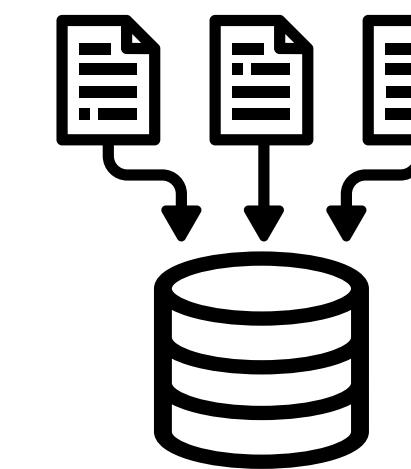
DATA ANALYTICS

- Data Analytics
- MBA specialized Marketing
- Post-Graduate Data Analytics at Langara (Current)
- Previous work experience as a contractor at Microsoft



2. DATA CLEANING AND PROCESSING

- 2015 data was merged, cleaned, and processed.
- Unfamiliar inputs (outliers) were removed for more accurate analysis and predictions.
- There was no missing values, therefore no imputations needed for the datasets.



3. FEATURES WITH STRONG RELATIONSHIPS

Technical Information

- 1 Payment type - Drop-off borough p value < 0.01
 - 2 Payment type - Passenger Count p value < 0.01
 - 3 Tip amount - Drop-off borough p value < 0.01
-  Highest average tip for drop-off: Staten Island at \$7.21



4. NEIGHBORHOODS THAT LOVE TAXI

- **Chelsea and Clinton:** Nearly 300,000 trips in 2015

PLACE MORE CABS HERE?



- **Gramercy Park and Murray Hill:** Around 250,000 trips
- **Upper West Side:** Around 102,000 trips
- **Greenwich Village and Soho:** 99,000 trips
- **Lower Manhattan:** Nearly 54,000 trips



5. IMPROVING OPERATION (TRAFFIC JAMS?)

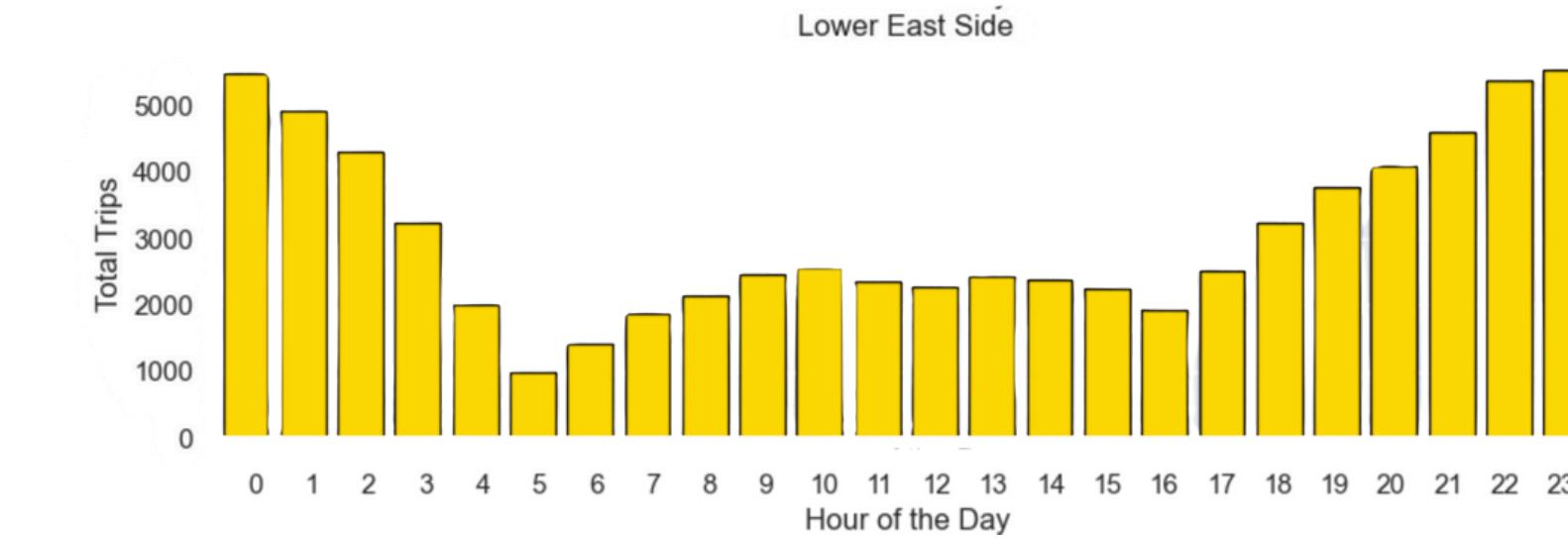
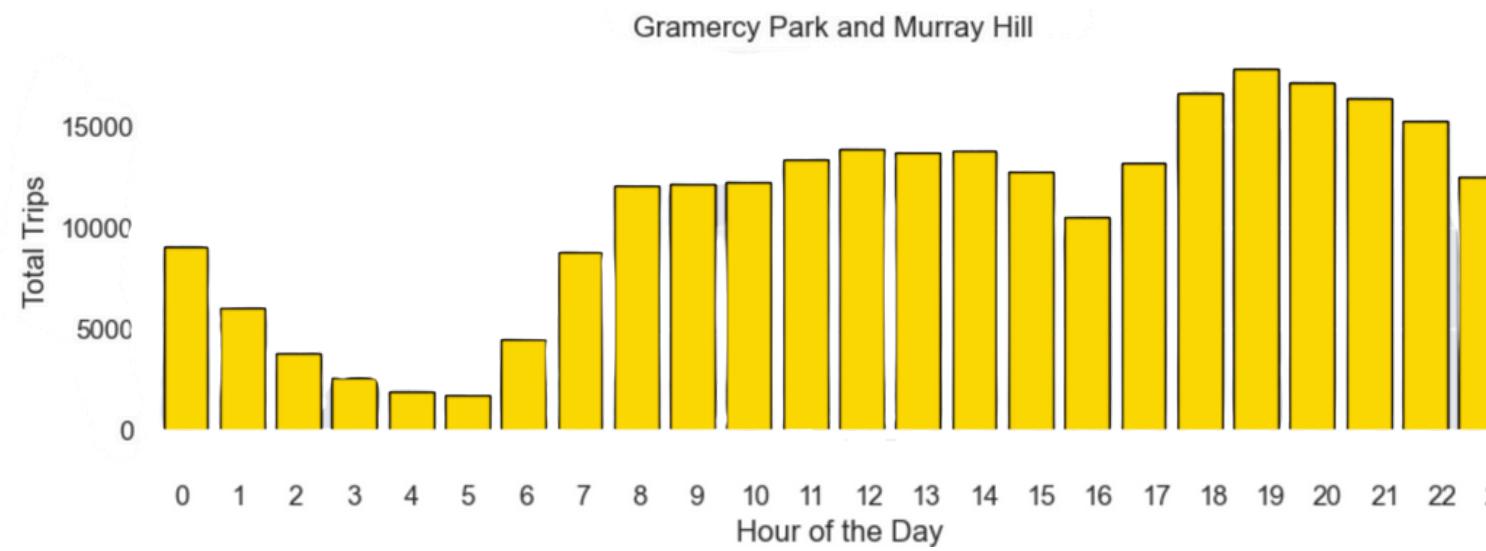
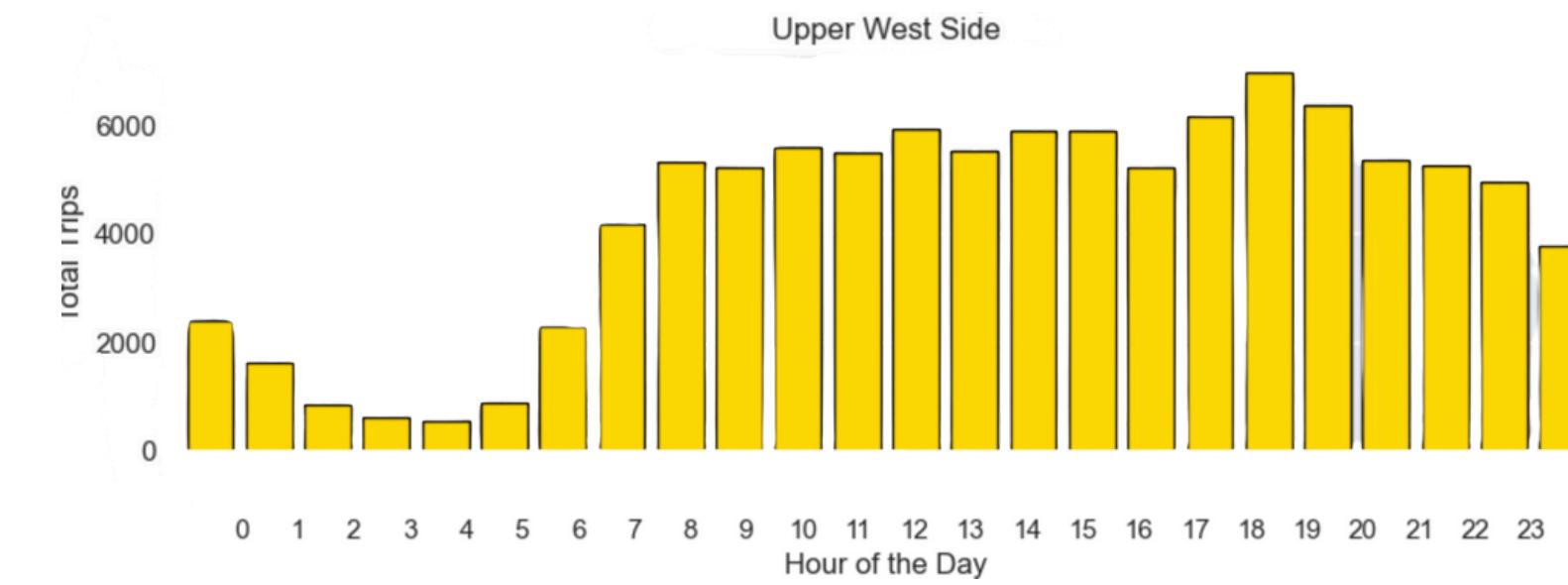
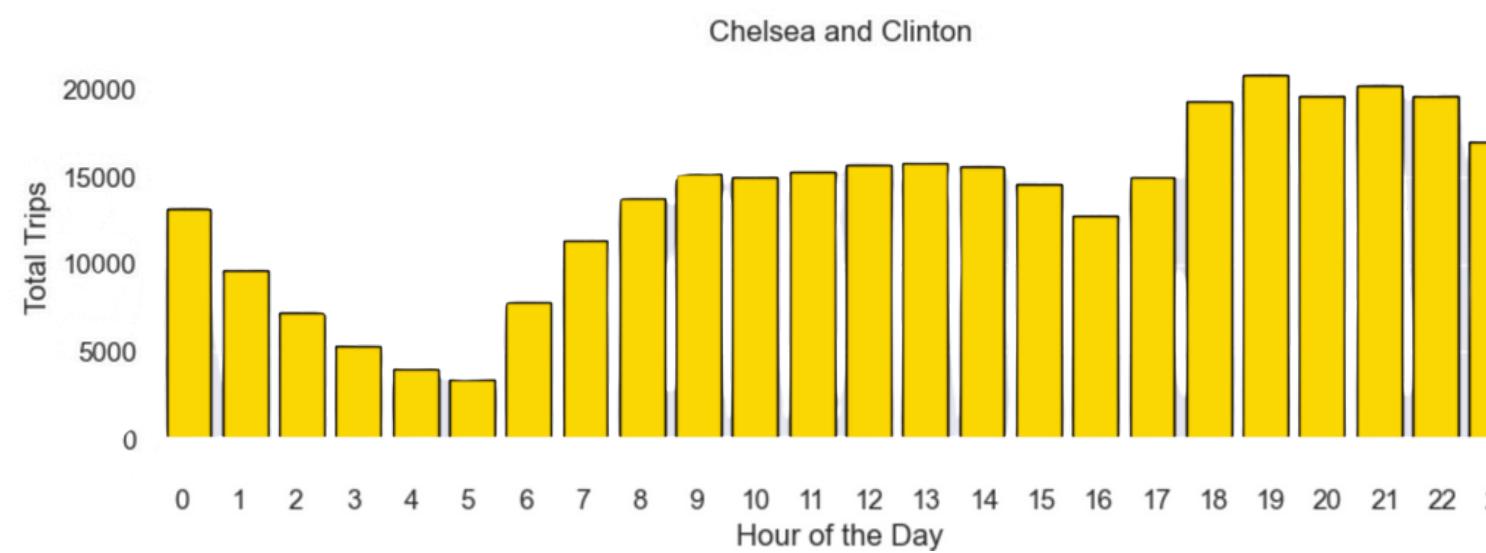
MOST HEAVY TRAFFIC NEIGHBORHOODS

Neighborhood Names	Average Duration per Trip (Mins)
Rockaways	28.01
Canarsie and Flatlands	14.96
Central Queens	12.22
Northeast Queens	11.64



5. IMPROVING OPERATIONS

What time and where do New Yorkers need Taxi?



New York never sleeps!
Each areas have very different active times

5. IMPROVING OPERATIONS

What day to increase cabs? A hint about holidays

Example Holiday Names	Days of the week	Average Trip	Holiday Trip
New Years Day	Thursday	3386	2888
Independence Day (observed)	Saturday	3570	3844
Memorial Day	Monday	2942	2232

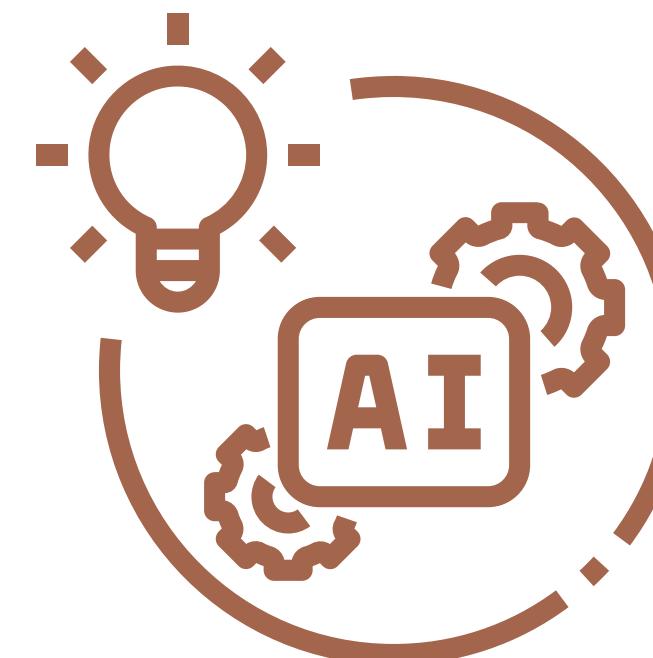


6. MACHINE LEARNING MODEL

To predict if a trip will receive over 20% tips or not

Random Forrest Model was trained and used, because:

- Fast and efficient
- It avoids Bias and reduces Biases and Errors
- Highly stable and can work with different types of data, especially good with classification



6. MACHINE LEARNING MODEL (CONTINUED)



Overall Analysis:

- Model has a high accuracy of 81% overall.
- Performs very well in identifying cases of tips > 20%.
- Have misses some cases where tip is low



Top Features importance:

- Payment Type: 45.4%
- Dropoff_latitude: 6.38%
- Pickup_longitude: 6.37%
- Dropoff_longitude: 6.31%
- Pickup_minutes (At what time of the day): 6.15%



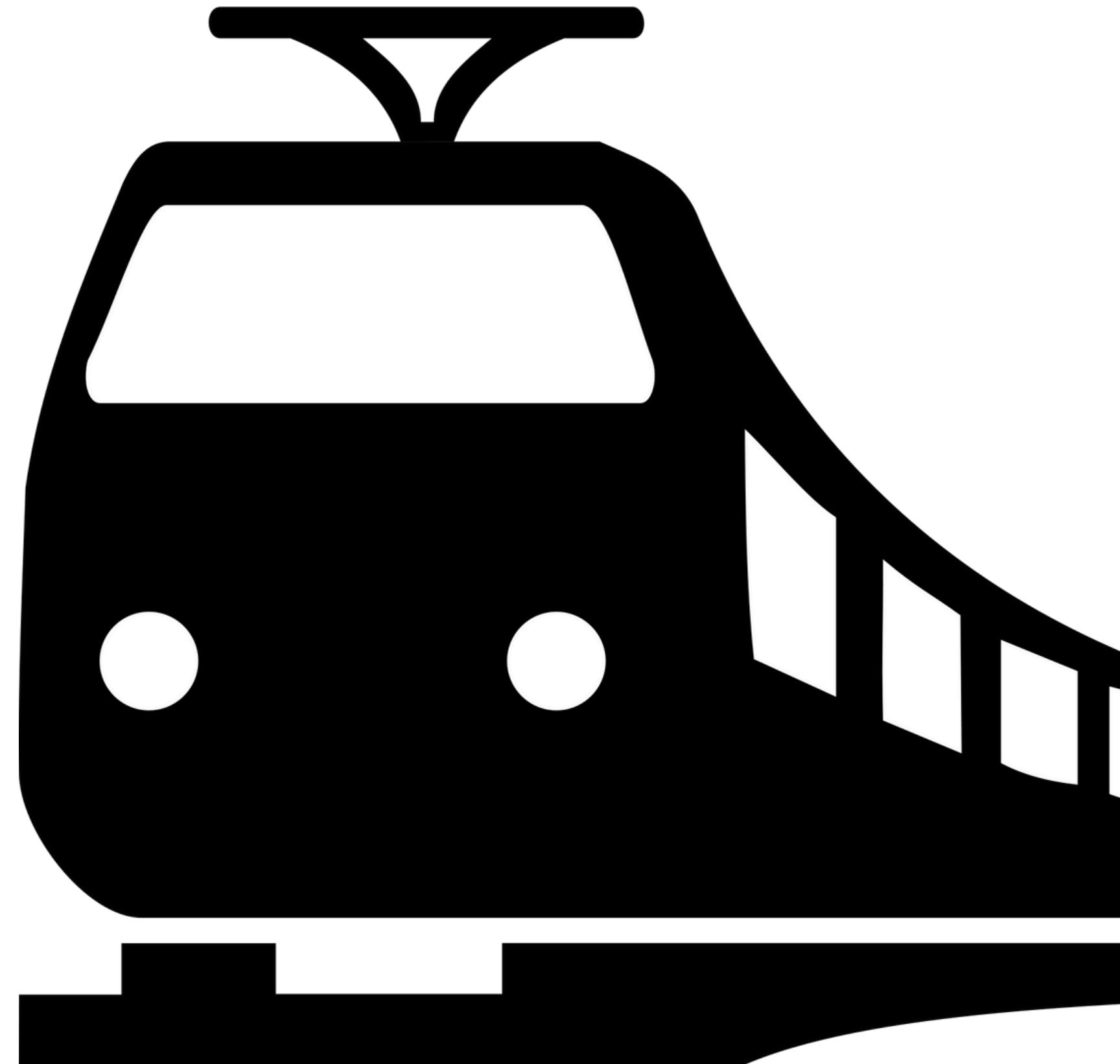
7. WHAT'S NEXT? - FINAL REPORT

Topics are included in the final report

- More strong pair relationship features
- Operation Analysis based on weather conditions.
- Detail Analysis of Tip Amount per Neighborhood
- Revenue report per Seasons
- Comprehensive report data and tables for all above discussion
- Fine-tuned and Accuracy Boost for Machine Learning Model
- Adjust to keep improving the Model's results

COMING

SOON



**THANK YOU FOR
YOUR TIME**

