# New York City Taxi & Limousine Commission Analysis

Martin Liang

## 1 INTRODUCTION

The transportation business is an integral part of our daily lives, facilitating the movement of people, goods, and services from one location to another. One of the most popular transportation options is the taxi service, and in this data analysis, the focus will be on the Yellow Medallion Taxicabs in New York City (NYC).

The dataset was sourced from the NYC Taxi & Limousine Commission (TLC) official website. The dataset contains several variables used to assess a completed trip such as pick-up, drop-off dates and trip distances

The report will include sections covering the introduction, data analysis, regression analysis, discussion, and conclusion, as well as references to the sources of information or data. The data analysis section will entail an extensive examination of the gathered data. The regression analysis section will entail the formulation of a regression model for forecasting the total fare for a taxi ride. The modeling process and outcomes will be discussed in detail. Finally, the discussion and conclusion section will summarize the analysis.

The aim of this task is to analyze the Yellow Medallion Taxicabs data set and answer some business questions related to daily cab activities. We will create a regression model to predict the total amount paid by the passengers after a given trip, given the trip information in the data set. The data set was sourced from the NYC Taxi & Limousine Commission (TLC) official website. These are the following questions that will be answered through data analysis;

- What is the average demand for taxis on the days of the week (i.e., daily trend)? Which of the days has the highest and which lowest demand?
- Which time of the day /(morning, afternoon, evening, and night) is likely a peak period for the operation of the taxi from the data?
- On average, how much revenue was generated on the weekdays and weekends for the business for the period covered in the data set?

## 2 DATA ANALYSIS

### 2.1 Question 1: What is the average demand for taxis on the days of the week (i.e., daily trend)? Which of the days has the highest and which lowest demand?

Knowing the average demand for taxis on each day of the week can help taxi companies to allocate their resources more efficiently.[1] Scheduling more drivers and taxis during days when demand is higher, such as weekdays, and reducing the number of available taxis during days when demand is lower, such as weekends. This can result in optimized resource utilization and reduced operating costs for the company. [1]

Based on the results presented in Figure 1, The analysis indicates that the day with the highest demand is Friday, while the day with the lowest demand is Monday. This observation is further supported by the data presented in Table 1, which reveals that the demand for taxis gradually increases from Monday to Friday, culminating in its peak value on Friday, and subsequently declines over the weekend. These findings suggest that the demand for taxis is heavily influenced by the day of the week, with a discernible pattern of higher demand on weekdays as compared to weekends.

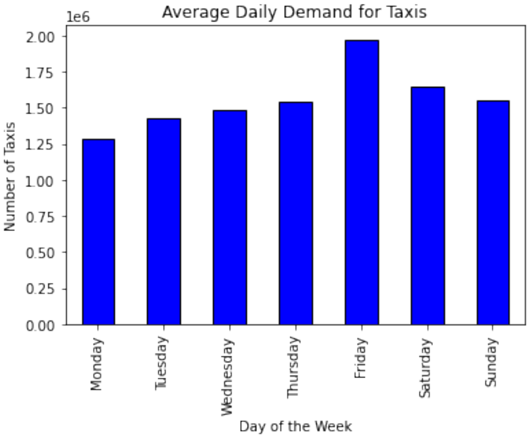

**Figure 1: Average Daily Demands on Taxi.**

| Day of the Week | Number of Taxis |
|---|---|
| Monday | 1282390 |
| Tuesday | 1423483 |
| Wednesday | 1484891 |
| Thursday | 1544347 |
| Friday | 1972597 |
| Saturday | 1644057 |
| Sunday | 1555093 |

**Table 1: More Accurate Numbers**

### 2.2 Question 2: Which time of the day (morning, afternoon, evening, and night) is likely a peak period for the operation of the taxi from the data?

The identification of peak periods for taxi services is of critical importance for the efficient and effective operations of taxi companies.[3] This information is necessary for ensuring that adequate resources are available to meet the demands of the customers during times of high demand. [3] The information is particularly valuable for taxi companies who can use it to optimize operations, reduce wait times, and increase customer satisfaction. Knowing the peak periods for taxi services enables companies to allocate resources, such as vehicles and drivers, more efficiently. By increasing the number of available taxis during peak hours, the company can reduce wait times and improve service quality.

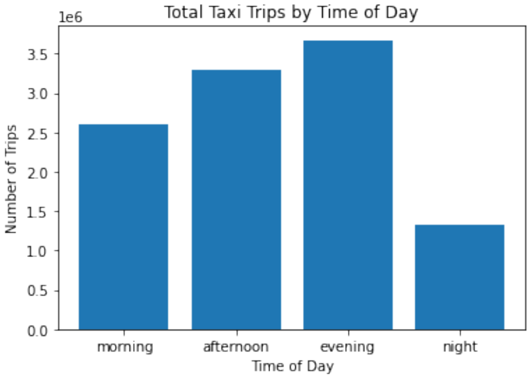

**Figure 2: Taxi peak period.**

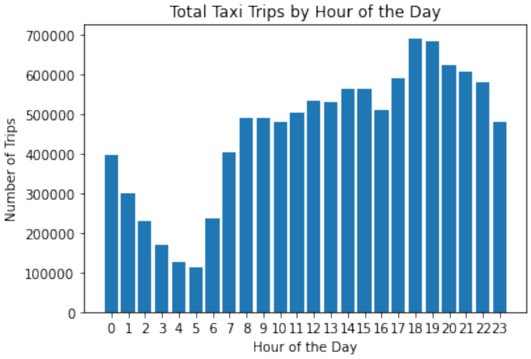

**Figure 3: Greater details of statistic of peak period.**

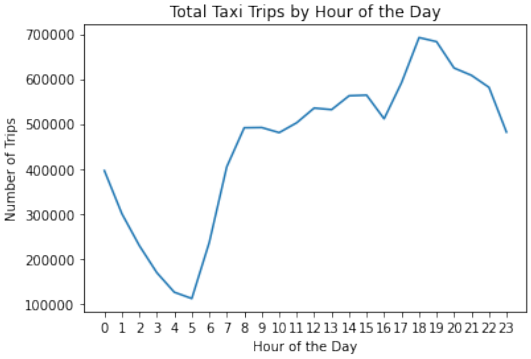

**Figure 4: line chart of peak period .**

The results depicted in Figure 2 suggest that there is a clear pattern in the demand for taxi services in the observed area. Specifically, the data shows that the evening hours are the busiest, indicating that customers are more likely to use taxis during this time. Conversely, the night hours exhibit the lowest number of trips, suggesting that taxi services are less frequently utilized during this period. Furthermore, the data indicates that afternoon hours are the second busiest time, while morning hours have the least demand for taxi services.

In order to investigate the peak hours for taxi operations, we analyzed the data represented in Figure 3 and Figure 4. The graph displays an average gain of 500,000 from noon to evening, with 18:00 being the most popular hour. Conversely, 5:00 at night has the lowest number of taxi trips. To further breakdown the time periods, a period is considered the following in European time measure: morning to be from 6:00 to 12:00, afternoon from 12:00 to 18:00, evening from 18:00 to 0:00, and night from 0:00 to 6:00. The data indicates that the evening is the most popular period for taxi trips, followed by the afternoon, while the morning is the third most popular. The night has the least amount of taxi trips.

- Morning; 6:00
- Afternoon; 12:00 - 18:00
- Evening; 18:00 - 00:00
- Night; 00:00 - 06:00

By understanding the patterns of taxi usage, companies can adjust their operations and allocate resources in a way that aligns with the demand for their services. For instance, companies may choose to increase the number of available taxis during peak hours in the evening to better meet customer demand, while reducing the number of available taxis during low-demand periods at night.

## 2.3 Question 3: On average, how much revenue was generated on the weekdays and weekends for the business for the period covered in the data set?

By collecting and analyzing data on revenue generated on each day of the week over a specified time period, taxi companies can gain valuable insights into revenue patterns and make informed business decisions. The revenue generated on weekdays and weekends is a significant indicator of customer behavior and demand patterns. By understanding these patterns, taxi companies can adjust their pricing strategies and promotional activities to maximize revenue and improve profitability.

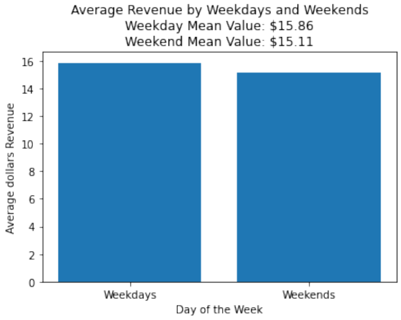

**Figure 5: Mean value of weekdays and weekends.**

Based on the data presented in Figure 5, the mean value generated on weekdays is 15.86 dollars, which is 0.75 dollars higher than the mean value generated on weekends, which is 15.11 dollars. Furthermore, upon analyzing Figure 6, it can be observed that Saturday

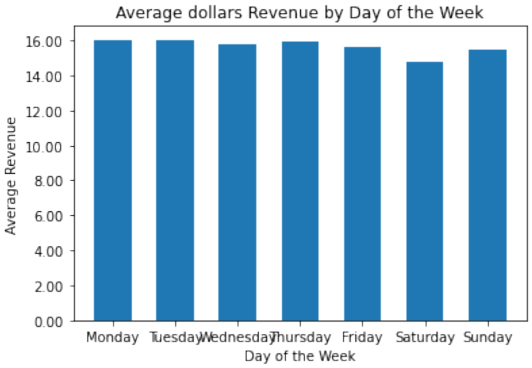

**Figure 6: Average revenue .**

is the day of the week with the lowest mean value generated, while the rest of the days in a week earn an average of 15 dollars. These findings may be a significant difference in the earnings generated on weekdays and weekends and that Saturday may be a day of lower business activity compared to other days of the week.

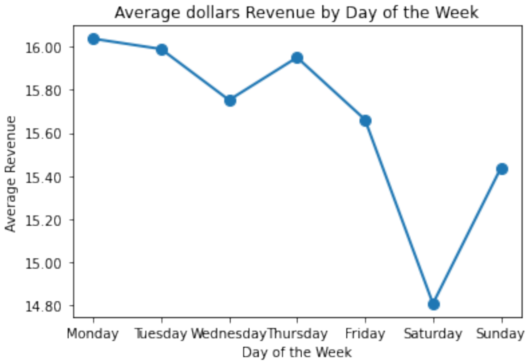

**Figure 7: Line chart of revenue .**

According to the Figure 7 line chart, it can be observed that the mean revenue value drops from 15.60 dollars on Friday to 14.80 dollars on Saturday, which indicates that Saturday is the least profitable day of the week. The mean revenue value then increases back to 15.40 dollars on Sunday. This suggests that there is a difference of 0.80 dollars in mean value lost from Friday to Saturday. Therefore, the data support the conclusion that Saturday is the least profitable day of the week in terms of mean revenue generated.

## 3 REGRESSION ANALYSIS

Figure 8 depicts a comparison between the actual amount paid and the amount predicted by a model for total taxi fares. The model was developed using two distinct data sets and aimed to forecast the total amount paid for taxi rides. The scatter plot was chosen as the visualization method due to its simplicity in its comparison with the relationship between current amount and predicted amounts.

The RMSE is 1.2213569520991874. The acronym stands for Root Mean Square Error. The RMSE is the square root of the average of the squared differences between the actual and predicted values. It measures the average magnitude of the errors in a set of predictions, with higher values indicating a higher error. [2]

An R$\hat{2}$ score of 0.9914 indicates that the model can explain with about 99.14% of accuracy. High accuracy is a sign that it provides a good fit for the data and has strong predictive power.

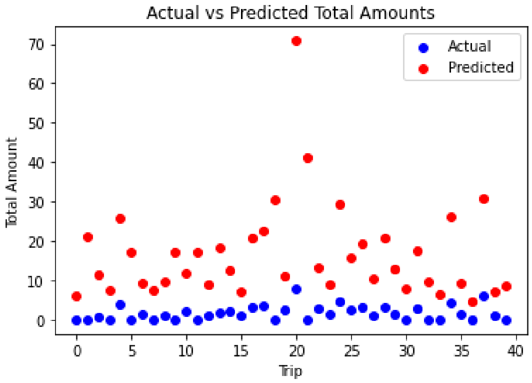

**Figure 8: Model to predict taxi fares.**

## 4 DISCUSSION

Based on the results of the predictive model, specifically the Root Mean Square Error and R$\hat{2}$ score, the model demonstrated good performance in accurately predicting outcomes. It should be noted, however, that the model was trained and tested on a specific data set, and it may not perform as accurately on other data sets. Additionally, the predictor variables utilized in the model may not be the most relevant or appropriate for different data sets, highlighting the importance of carefully selecting and refining variables for different modeling tasks. As with any modeling approach, alternative methods such as Logistic Regression and Decision Trees could also be explored as potential alternatives or complements to this approach.

## 5 CONCLUSION

The data analysis section for the taxi company in this report provides valuable insights into optimizing operational resources by understanding patterns of demand for taxi services. The analysis highlights peak times and profitable days of the week, enabling the company to allocate resources effectively. Additionally, the report compares revenue generated on weekdays and weekends, providing useful information for future business decisions. Overall, this data analysis is a valuable contribution to the field of transportation management and can inform decision-making processes for taxi companies seeking to improve efficiency and profitability.

## REFERENCES

[1] Xianlei Dong, Min Zhang, Shuang Zhang, Xinyi Shen, and Beibei Hu. 2019. The analysis of urban taxi operation efficiency based on GPS trajectory big data. *Physica A* 528, 121456 (Aug. 2019), 121456.

[2] Adair Ribeiro, Jr, https://www.allankardec.online/, Carlos Bastos, and Luciana Farias. 2023. Uma revisão na história da 5a edição de A Gênese Parte III – A atuação de Amélie Boudet no pós-Kardec, a denúncia da adulteração da obra e sua repercussão na França e no Brasil. *J. Estud. Espíritas* (March 2023).

[3] Josep Maria Salanova, Miquel Estrada, Georgia Aifadopoulou, and Evangelos Mitsakis. 2011. A review of the modeling of taxi services. *Procedia Soc. Behav. Sci.* 20 (2011), 150–161.