Data Science Learning Studio

Study Case

Data Analytics at New York City Taxi and Limousine Commission (NYC TLC)

by **Adrian Maulana Muhammad**

21/12/2022

Table of Contents

Business Objectives

Business Understanding |Analysis Goals

Business Understanding

Taxi service in New York is the fourth largest transportation network in the United States. The New York City Taxi and Limousine Commission (TLC) regulates the system, which oversees Yellow Taxi, Charter Taxi, Commuter Car, Transit Vehicle, and Certain Limousin. In New York City, TLC serves several Boroughs such as the Bronx, Brooklyn, Manhattan, Queens, and Staten Islands.

Analysis Goals

Despite its large network, the current system does not serve Boroughs evenly. The goals of this project are:

1. In your opinion, what causes Yellow Taxi service to only be centered in certain Boroughs, resulting in an imbalance of supply and demand in other Boroughs?

2. The TLC wants to evenly distribute Yellow Taxi service throughout the city of New York to meet the existing demand. What should the TLC do in the short-term (6 months) to address this challenge?

Data Understanding

Datasets used in this project

The datasets used in this project:

Taxi Trip Records

This project's main data is a dataset that includes records and metadata related to picking up and dropping off passengers in taxis. To access more information about this dataset, you can visit the following link:

https://drive.google.com/drive/folders/1IBoslLMCotYTsMaK-9hhwySV38wMFlZN?usp=share_link

To learn more about the features of this dataset, you can view the data dictionary at the following link:

https://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

Taxi Zones

I will be incorporating an external dataset to aid in mapping the metadata of pick-up and drop-off locations within the Taxi Trip Records dataset. If you would like to learn more about this external dataset and how it will be used to enhance the analysis of the Taxi Trip Records, you can visit the following link for additional information.

https://catalog.data.gov/dataset/nyc-taxi-zones

NYC Borough Geo JSON

I will utilize this data to aid in creating visualizations of the pick-up and drop-off data for the New York City Taxi and Limousine Commission (TLC) in each borough. If you would like to learn more about this topic, you can visit the following link for additional information:

https://github.com/thomashikaru/nycvisualization

Exploratory Data Analysis

EDA & Statistical Analysis

Which Boroughs are Most Often Picked Up by NYC TLC?

In order to gain a better understanding of the distribution of New York City Taxi and Limousine Commission (TLC) services across the different boroughs, we will examine the proportion of passengers picked up in each borough. From our preliminary analysis, it appears that **Manhattan** has the highest proportion of passenger pick-up services, while the **Bronx** has the lowest. However, we will further investigate this finding and conduct a more in-depth analysis to confirm these results and gain a more comprehensive understanding of the distribution of TLC services in each borough.

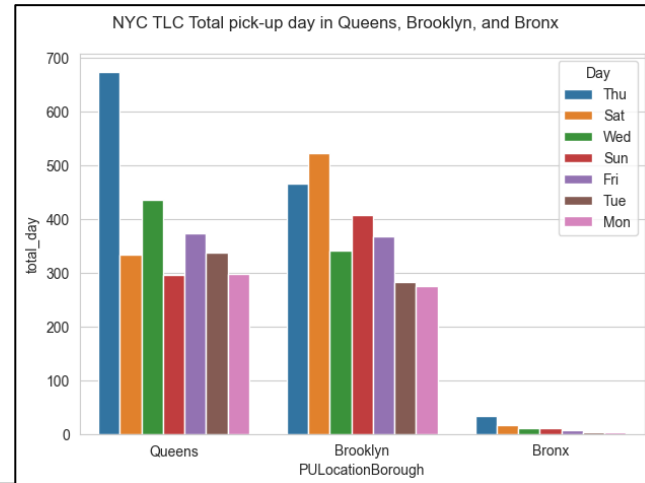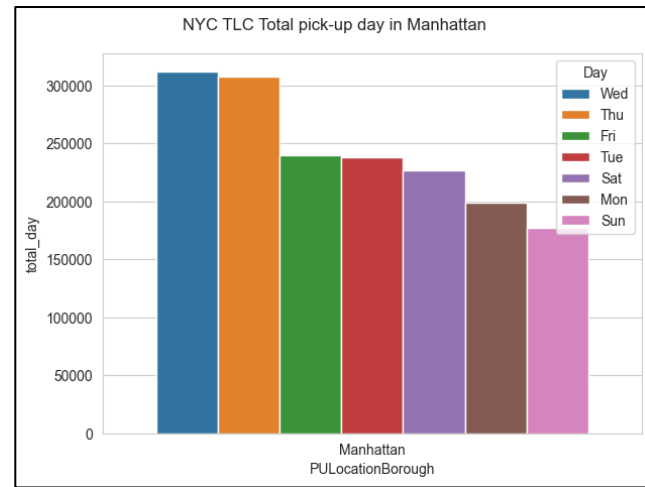Which Boroughs are Most Often Picked Up by NYC TLC?
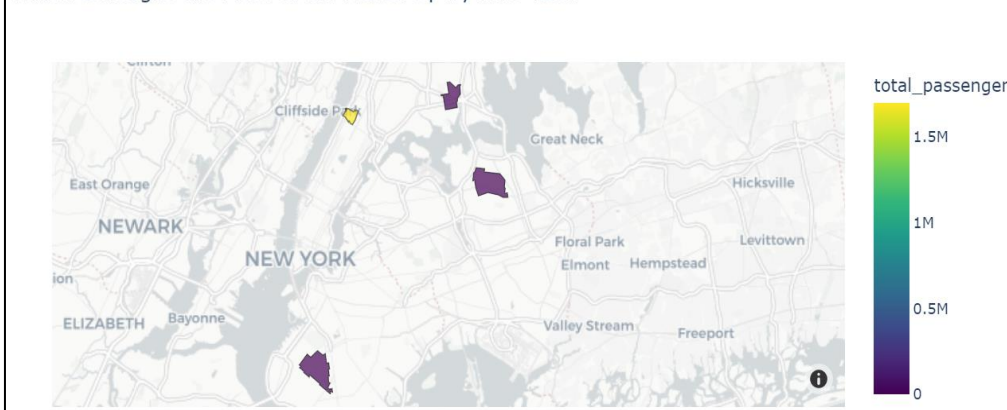
	PULocationBorough	total_passenger
0	Manhattan	1698895
1	Unknown	6375
2	Queens	2746
3	Brooklyn	2658
4	Bronx	85

On which day does the NYC TLC most often pick up passengers in each Borough?

To see the behavior of passenger pickups in each Borough, I want to see on which day the pickups are most often done.

- According to our analysis, **Wednesdays** are the most common day for pick-ups by the New York City Taxi and Limousine Commission (TLC) in **Manhattan**, while Sundays are the least common.

- According to our analysis, pick-ups by the New York City Taxi and Limousine Commission (TLC) in the **other boroughs** are more commonly observed on **Thursdays and Saturdays**, as compared to Wednesdays in Manhattan. In contrast, Mondays are the least common day for TLC pick-ups in the other boroughs.

At which time does the NYC TLC most often pick up passengers in each Borough?

To see the behavior of passenger pickups in each Borough, I want to see at which time the pickups are most often done.

- According to our analysis, the pick-up time for New York City Taxi and Limousine Commission (TLC) services in **Manhattan** appears to be more common in the **afternoon**. This finding suggests that there may be a higher demand for TLC services in Manhattan during this time of day.

- Our analysis has shown that, in contrast to Manhattan, pick-ups for New York City Taxi and Limousine Commission (TLC) services are more common in the **morning** in the **other boroughs**.
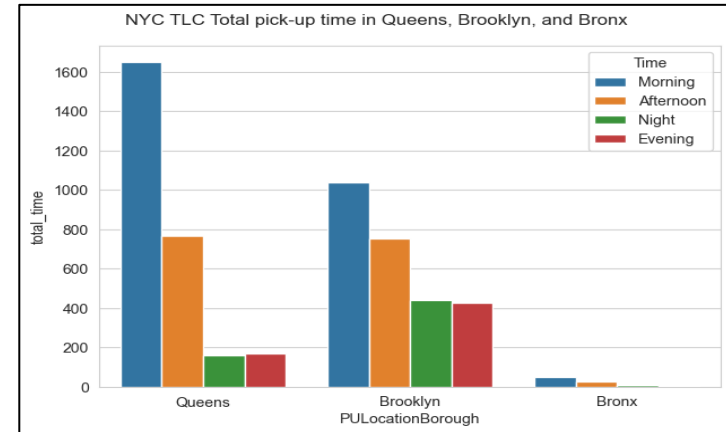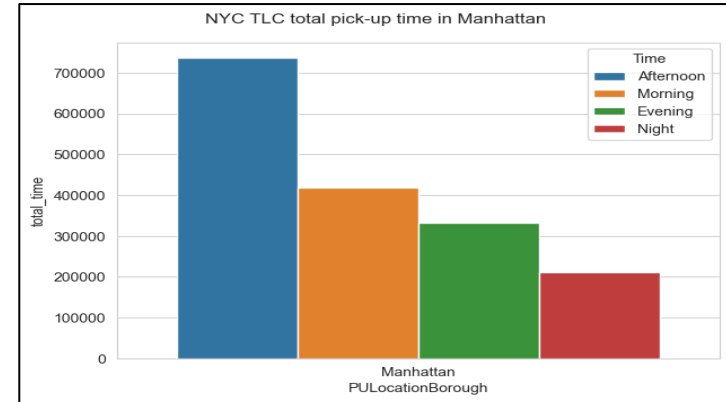
Numerical Data Distribution

- Based on the numerical data distribution visualization, it appears that the data can be divided into two main categories: columns with a single central value and columns with data distributed over a range of values. In order to gain a better understanding of the relationships between these columns, it is more useful to focus on the columns with data distributed over a range of values.

- The visualization above reveals a strong correlation between the columns. Given this finding, we can further analyze the data by selecting one of these columns and examining the average values for each borough. This will allow us to gain a deeper understanding of the relationships between these variables and how they vary across the different boroughs.

Numerical Data Distribution

Korelasi Variabel Numerik dari Dataset Customer
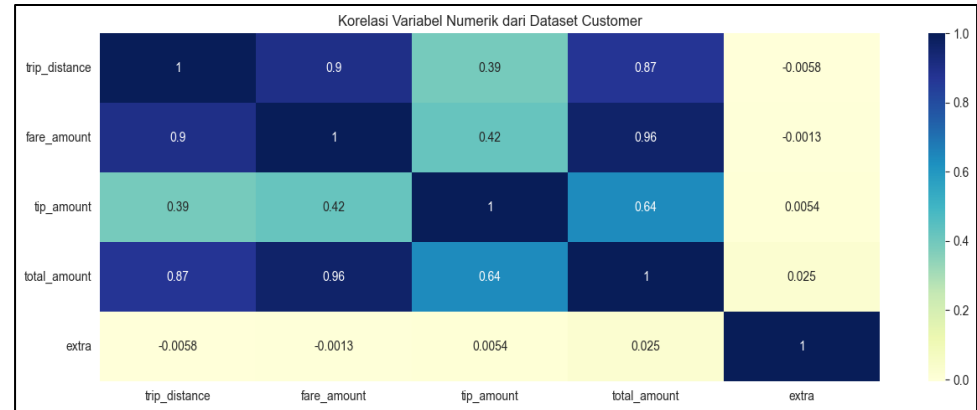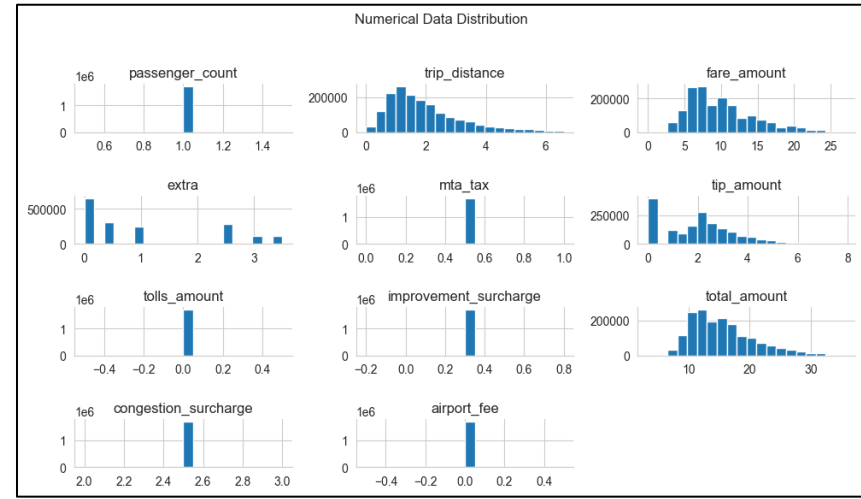
# What is the average trip distance traveled by NYC TLC in each Borough?

- The visualization before indicates a strong correlation between the trip_distance, fare_amount, and total_amount columns. Based on this finding, we can further analyze the data by selecting the **trip_distance** column and examining the average values for each borough.

- It is noteworthy that the borough with the highest rate of passenger pickups by the New York City Taxi and Limousine Commission (TLC), **Manhattan**, has the **lowest average trip distance**. On the other hand, the borough with the **highest average trip distance**, the **Bronx**, has the lowest rate of TLC passenger pickups. distance in each borough.

Average of NYC TLC Trips on each Boroughs
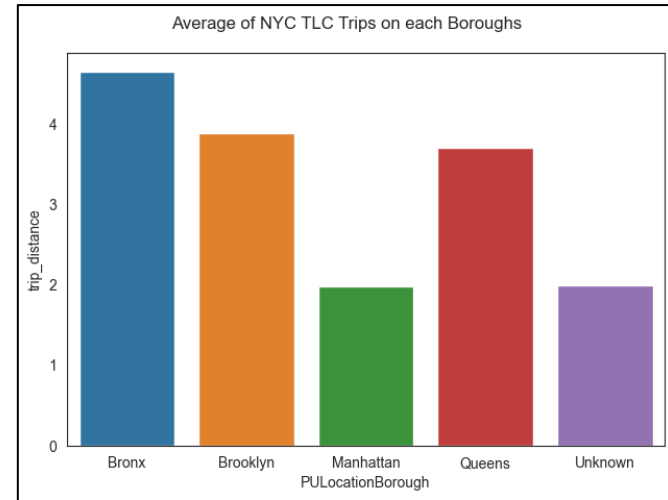
PULocationBorough	trip_distance
Bronx	4.650353
Brooklyn	3.879605
Manhattan	1.973766
Queens	3.697917
Unknown	1.987926

What is the average travel time in each Borough?

In addition to the previously mentioned finding, it is also worth noting that **Manhattan**, the borough with the highest rate of passenger pickups by the New York City Taxi and Limousine Commission (TLC), has **the lowest average travel time** for TLC services. It is possible that the demand for TLC services in Manhattan is particularly high, resulting in shorter travel times for TLC vehicles.

Average Time of NYC TLC Trips on each Boroughs
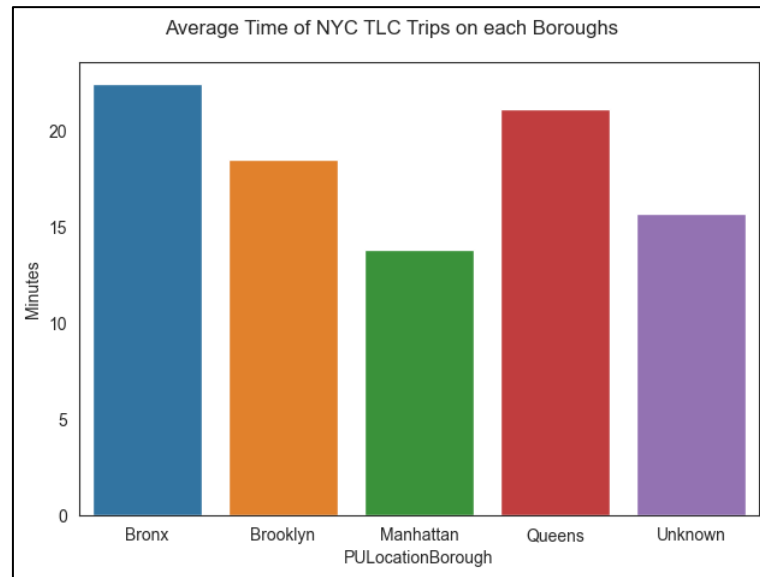
Summaries

Based on the above data analysis, several conclusions can be drawn.

Based on the above data analysis, the following conclusions can be drawn

1. The borough with the highest rate of passenger pickups by the New York City Taxi and Limousine Commission (TLC) in Manhattan, while the borough with the lowest rate is the Bronx.

2. In Manhattan, TLC pickups are most common on Wednesdays or Saturdays and least common on Sundays. In the other boroughs, TLC pickups are most common on Thursdays or Saturdays and least common on Mondays.

3. TLC pickups in Manhattan are more common in the afternoon, while pickups in the other boroughs are more common in the morning.

4. The borough with the highest rate of TLC passenger pickups, Manhattan, has the lowest average trip distance, while the borough with the highest average trip distance, Bronx, has the lowest rate of TLC passenger pickups.

5. Manhattan, the borough with the highest rate of TLC passenger pickups, also has the lowest average TLC travel time.

From these conclusions, it is apparent that the distribution of TLC services in each borough is influenced by differences in the days and times of TLC service, as well as the distance and time traveled by TLC vehicles.

Recommendations

Based on the conclusions drawn from the data analysis, the following recommendations can be made

Based on the conclusions drawn from the data analysis, the following recommendations can be made

1. To meet the demand for TLC services on Thursday and Saturday mornings in the Bronx, Brooklyn, and Queens, the TLC may consider increasing service during these times.

2. To meet the demand for TLC services on Wednesday and Saturday daytime in Manhattan, the TLC may consider increasing service during these times.

3. Given the relatively high average travel distance and travel time in the Bronx, Brooklyn, and Queens, the TLC may consider using services such as commuter cars or transit vehicles to better accommodate these needs.

4. Given the relatively low average travel distance and travel time in Manhattan, the TLC may consider using services such as yellow taxis, chartered taxis, or certain limousines to better meet the needs of passengers in this borough.

Thank You

Email

adrian.m.muhammad@gmail.com

Phone

+6285233791448

Website

github.com/adrn-mm/study_case_DSLS_2023

Linkedin

linkedin.com/in/adrn-mm/