ANOINTED UBONGABASI AKAI

NYC TAXI TRAVEL ANALYSIS (PYTHON).

THE NEW YORK TAXI ANALYSIS

INTRODUCTION

Project Introduction:

The purpose of the New York City dataset is to analyze patterns, trends, and behaviors in taxi transportation within a city (typically New York City) over the year 2017. This analysis is targeted to reveal key trends in seasons, demand analysis it will also serve as a valuable resource for understanding urban mobility patterns and improving transportation systems through data-driven insights.

Background Of the Project:

This project focuses on analyzing the Taxi Trips in 2018, which contains detailed records of taxi trips conducted throughout the year 2018. By examining time-based attributes such as pickup times, hours of operation, and seasonal variations, the project aims to uncover patterns in taxi demand, identify peak travel periods, and assess the influence of seasons on taxi usage. These insights can provide valuable information for taxi companies, policymakers, and city officials to optimize services, allocate resources efficiently, and improve the overall transportation infrastructure.

Objectives of the Project

1. To analyze trends based on seasons and time and provide insight on favorable travel times and busiest times, and seasons.
2. To support operational planning for taxi services
3. To provide insights for urban mobility planning

METHODOLOGY: DATA SOURCES AND PREPROCESSING METHODS

DATA SOURCES

This dataset is typically made available by municipal transportation authorities or open data platforms that track all sales-related activities, interactions, and outcomes. The data to be used includes Taxi Trips, which include all the important details to be used in this analysis, like the pick up and drop off date and time, trip duration and distance, fare amount and payment type, and pick up and drop off locations. And also the Taxi Calendar and zones.

DATA EXPLORATION AND PREPROCESSING METHODS

Before analyzing the data, it was necessary to explore and preprocess it to ensure accuracy, consistency, and readiness for analysis. This involved the following key steps:

1. Data Cleaning:

- **Removing duplicates:** Ensured that duplicate entries (e.g., travels ID entered multiple times) were removed to avoid inflating metrics. No duplicate were found here.

- **Handling missing values:** In these project it is ideal that all missing box are left blank or with null values or 0 because adding the average, mean or median to it will alter the values and the outcome of the analysis.

2. **Data Standardization and Relevant Keys - Standardizing formats:** Ensured uniformity in data entry, such as date formats of the close date column (e.g., converting all dates to a common format), currency values, and product names. Foreign keys was not used here because none was indicated.

3. Filtering Relevant Data: Few Columns were dropped and others were renamed.

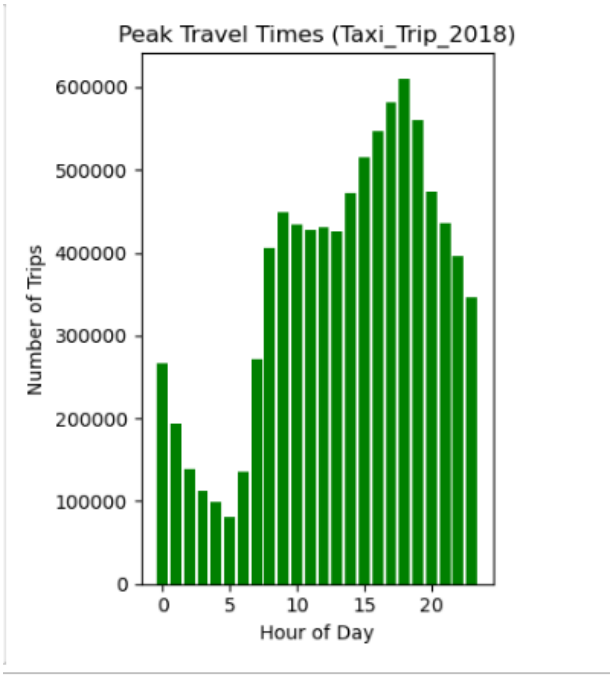
4. **Metrics and Data Summary :**

Relevant Metrics were checked to have an overview of the taxi trips. Measure like the sum of trip distances, sum of passenger count, the total revenue, average fare, and sum of tip amount. This was done in preparation for the data for deeper analysis.

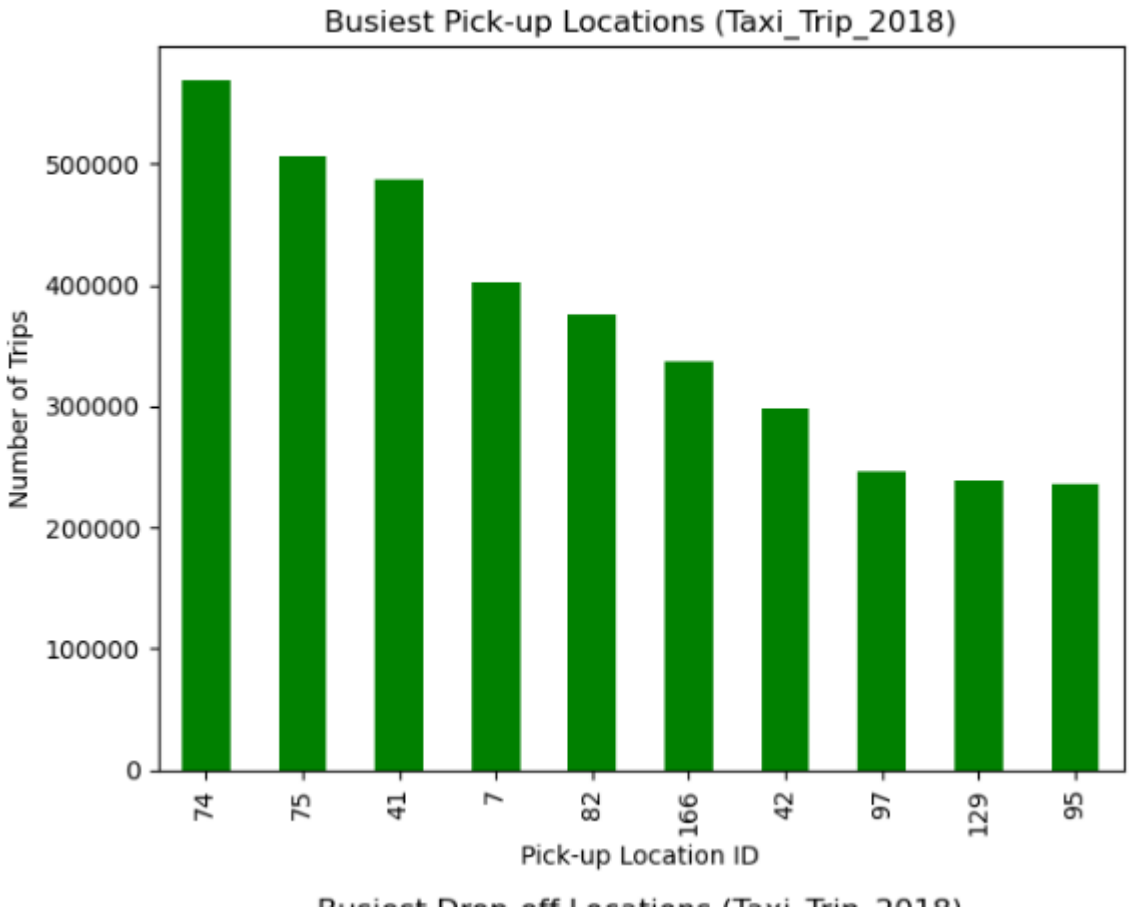
TRAVEL PATTERN ANALYSIS

Peak Travelled Times:

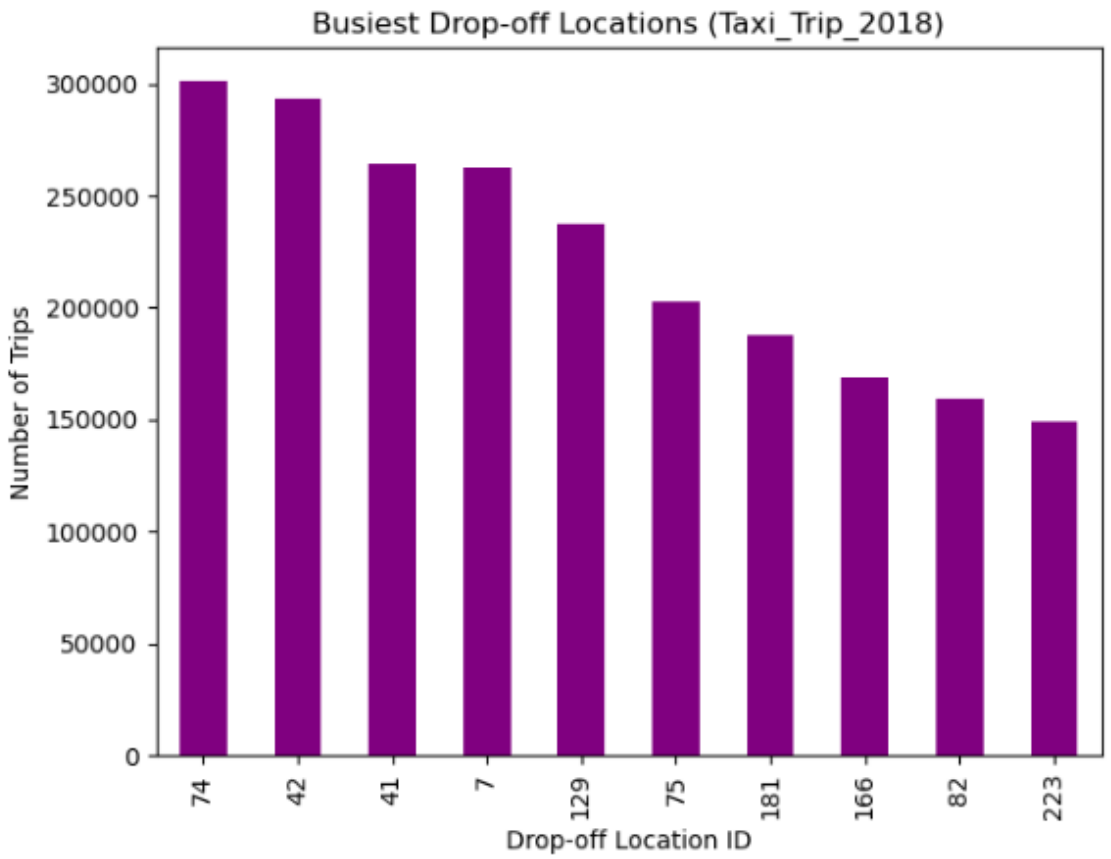
The most traveled hour and busiest hours are 18 (610,245 trips), 17 (581,546), and 19 (559,289), while the least busy are 4 (89,927), 3 (112,135), and 2 (135,943). Peak activity occurs around evening rush hours (17-19), with the lowest in early morning (2-4).



Busiest Pick-up and drop-off Location:



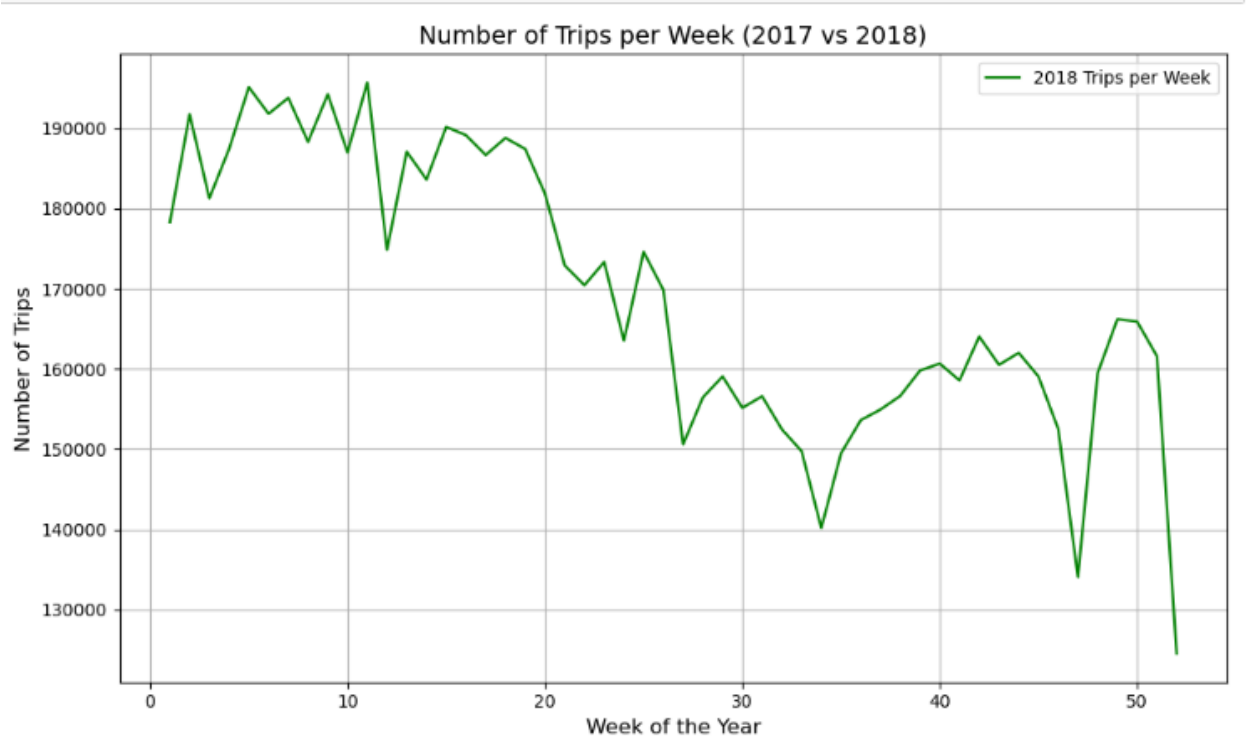
The histogram shows the taxi pick-up frequency by location ID in 2018. Location 74 has the most trips, around 550,000, followed by 75 and 41, each over 450,000. Locations 7, 82, and 166 have around 400,000 trips, while 42, 97, 127, and 95 range from 250,000 to 200,000. Location 74 is the busiest, with a clear drop-off in trips across other locations.



The histogram shows drop-off trip counts by location ID in 2018. Location 74 leads with around 30,000 trips, followed by 42 and 41 with about 27,500 and 25,000 trips. Locations 7, 29, 75, 181, 166, 82, 28, and 223 range from 20,000 to 15,000 trips, with a gradual decline. Location 74 is the busiest drop-off point.

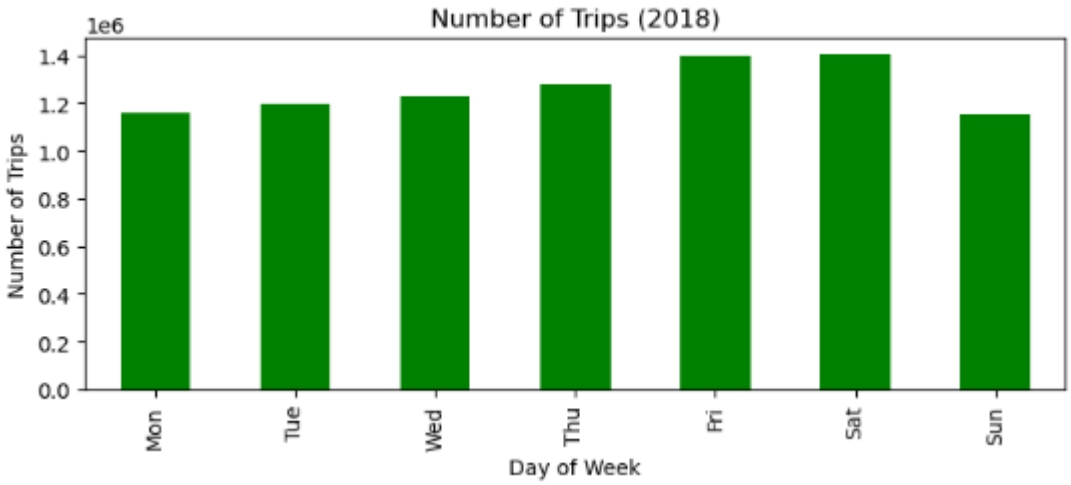
Number of trips per week:

The line graph shows 2018 taxi trips per week, ranging from 1.9M to 1.3M. Trips start high at 1.9M, fluctuate early, then decline steadily to 1.5M by week 30, dip to 1.3M around week 40, briefly rise to 1.6M, and drop sharply to 1.3M by year-end. Overall, trips trend downward with notable dips mid-to-late year. (Note: 2017 data is not shown.)



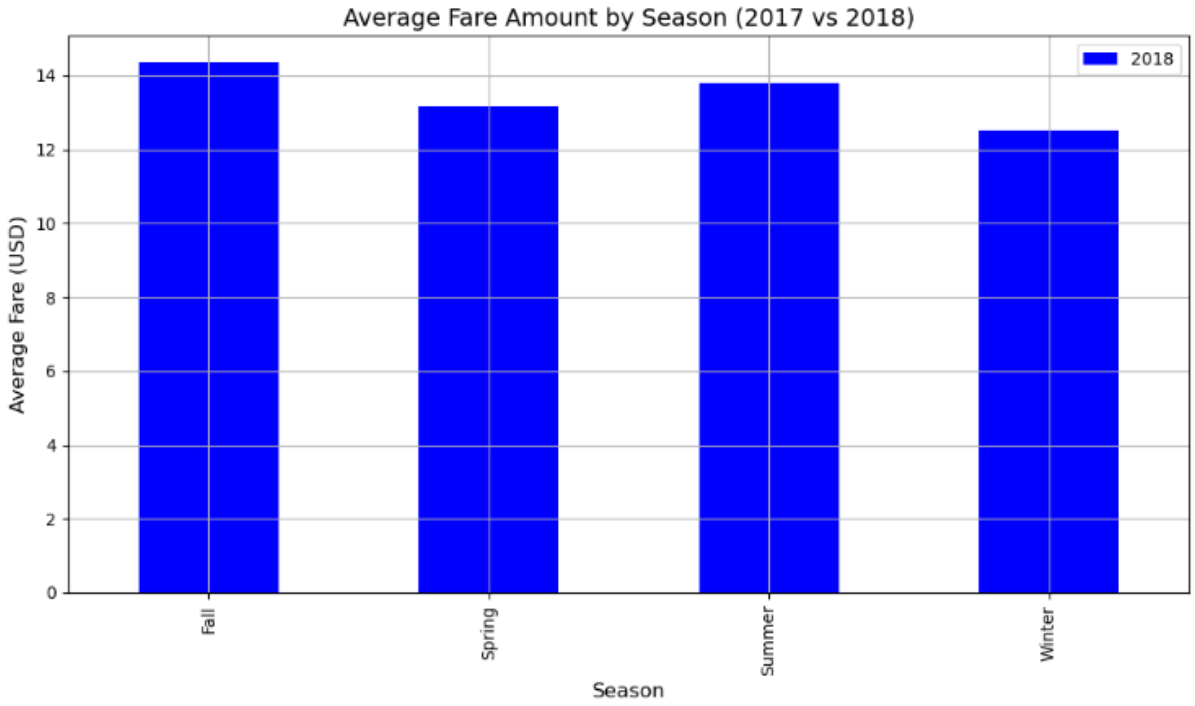
Number Of Trips on a particular Day of the Week:

The histogram shows weekly taxi trip counts by day in 2018. Friday and Saturday are the busiest, each with around 1.3 million trips, followed by Thursday and Wednesday at about 1.2 million. Monday, Tuesday, and Sunday have slightly fewer trips, around 1.1 million each. Trip volume peaks mid-to-late week, with a slight dip on weekends and early weekdays**.**



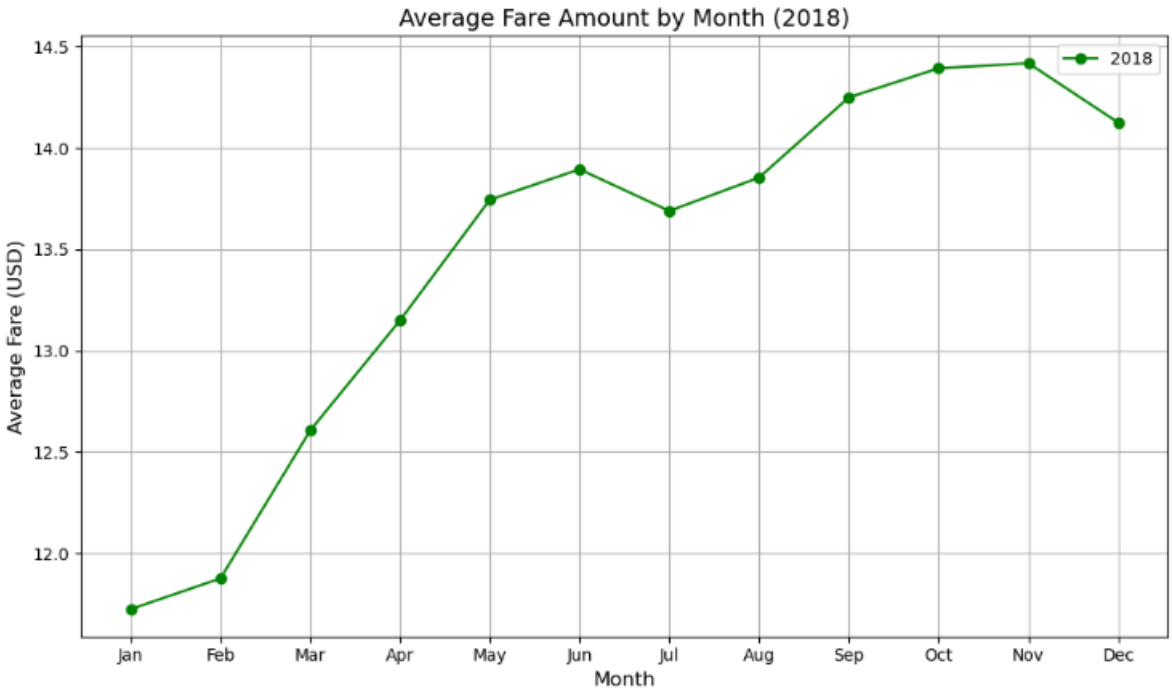
FARE STRUCTURE ANALYSIS

Seasonal Fare Trends:



The histogram shows 2018 taxi fares by season. Fall and Summer lead with around $14, Spring follows at about $13, and Winter is lowest at approximately $12. Fares are highest in Fall and Summer, with a slight decline in Winter. (Note: 2017 data is not shown.)

Seasonal Trends of Fares by Months



The line graph shows monthly taxi fare trends in 2018. Fares start at $12 in January, rise steadily to $14 by May, peak at $14.5 in October, and slightly dip to $14.3 by December. Fares generally increase throughout the year, with the highest in late fall.