

Uber Trip Data Analysis

Presented By

- Nallamalli Vishnu Vikas
- Shaheen Chirakula
- Sai Varun Nimmagadda

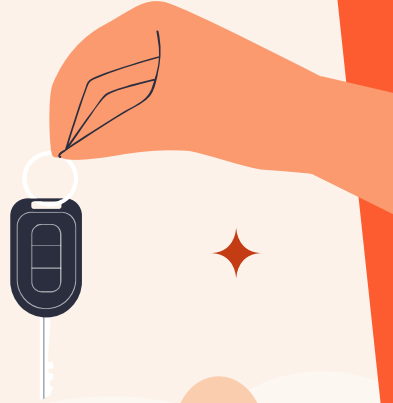
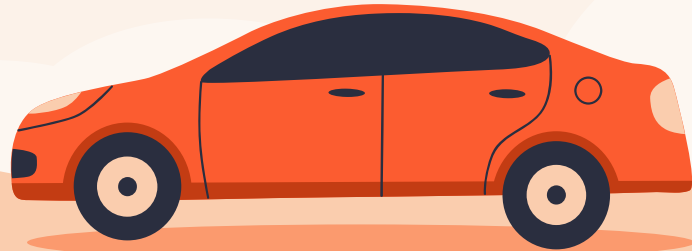


Table of contents

01

Introduction

02

Dataset

03

Framework

04

Sample dataset

05

**Expected
Results**

06

**Scope Of
Analysis**

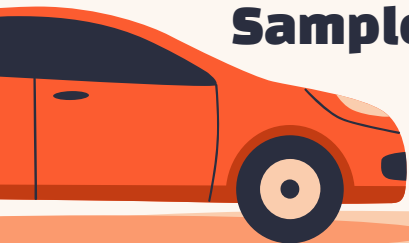


Table of contents

07

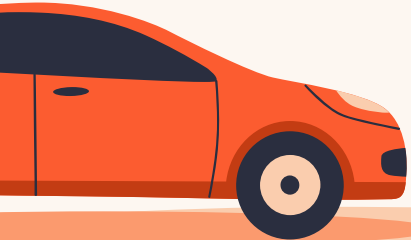
**Outcomes Of
Analysis**

08

**Benefits of
Anlaysis**

09

References



An illustration featuring two hands. The hand on the left is light orange and holds a dark blue car key. The hand on the right is dark brown and is reaching out. In the center, a white square with an orange border contains the number '01' in orange. The background is a light beige color with stylized white clouds, small orange stars, and a landscape at the bottom with rolling hills and two orange trees.

01

Introduction

- ✦ • In New York City, all taxi vehicles are managed by TLC (Taxi and Limousine Commission) established in 1971.
- ✦ • TLC regulates New York City's Medallion (Yellow) taxi cabs, for-hire vehicles (community-based liveries, black cars, and luxury limousines), commuter vans, and paratransit vehicles.
- Over 200,000 (2 Lakhs) TLC licensed vehicles complete approximately 1,000,000 (1 Million) trips each day.
- High-volume-for-hire vehicle bases(HVFH) are companies that dispatch 10,000+ trips per day.
- We have selected UBER for our analysis which is also an HVFH company.



An illustration featuring two hands, one light orange and one brown, exchanging a dark blue car key. In the center, a white square with an orange border contains the number '02' in orange. The background is a light beige landscape with stylized trees, clouds, and small orange stars.

02

Dataset



•Data set contains trips made by UBER in 2021 in NYC.

•Size is 3.6 GB.

•Contains 17.45 crore entries of data with 24 columns.

•Columns include details like Pickup Time, Drop Time, Trip Miles, and Trip Time.



An illustration featuring two hands. The hand on the left is light orange and holds a dark blue car key. The hand on the right is dark brown and is reaching out. In the center, a white square with an orange border contains the number '03' in orange. The background is a light beige color with stylized white clouds, small orange stars, and a landscape at the bottom with two orange trees and rolling hills.

03

Framework

Organizational chart

**'Pyspark' in
Databricks**

**Import the
dataset using
"Pyspark"**

Data Cleaning

Visualization



databricks



An illustration featuring two hands, one light orange and one brown, exchanging a dark blue car key. In the center, a white square with an orange border contains the number '05' in orange. The background is a light beige color with stylized white clouds, small orange four-pointed stars, and a landscape at the bottom with two orange trees and rolling hills. The entire scene is framed by orange borders on the left and right sides.

05

Sample dataset

| | hvfhs_license_num | dispatching_base_num | originating_base_num | request_datetime | on_scene_datetime | pickup_datetime | dropoff_datetime | PULocation |
|----------|-------------------|----------------------|----------------------|------------------------|------------------------|------------------------|------------------------|------------|
| 0 | HV0003 | B02682 | B02682 | 2021-01-01 00:28:09 | 2021-01-01 00:31:42 | 2021-01-01 00:33:44 | 2021-01-01 00:49:07 | |
| 1 | HV0003 | B02682 | B02682 | 2021-01-01 00:45:56 | 2021-01-01 00:55:19 | 2021-01-01 00:55:19 | 2021-01-01 01:18:21 | |
| 2 | HV0003 | B02764 | B02764 | 2021-01-01 00:21:15 | 2021-01-01 00:22:41 | 2021-01-01 00:23:56 | 2021-01-01 00:38:05 | |
| 3 | HV0003 | B02764 | B02764 | 2021-01-01 00:39:12 | 2021-01-01 00:42:37 | 2021-01-01 00:42:51 | 2021-01-01 00:45:50 | |
| 4 | HV0003 | B02764 | B02764 | 2021-01-01 00:46:11 | 2021-01-01 00:47:17 | 2021-01-01 00:48:14 | 2021-01-01 01:08:42 | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| 11908463 | HV0003 | B02765 | B02765 | 2021-01-31 23:13:51 | 2021-01-31 23:25:03 | 2021-01-31 23:25:40 | 2021-01-31 23:40:10 | |
| 11908464 | HV0003 | B02872 | B02872 | 2021-01-31 23:23:56 | 2021-01-31 23:29:03 | 2021-01-31 23:29:31 | 2021-01-31 23:47:44 | |
| 11908465 | HV0003 | B02872 | B02872 | 2021-01-31 23:42:53 | 2021-01-31 23:49:23 | 2021-01-31 23:49:32 | 2021-02-01 00:04:36 | |
| 11908466 | HV0003 | B02764 | B02764 | 2021-01-31 23:04:32 | 2021-01-31 23:09:13 | 2021-01-31 23:09:29 | 2021-01-31 23:27:46 | |
| 11908467 | HV0003 | B02764 | B02764 | 2021-01-31 23:22:20 | 2021-01-31 23:28:33 | 2021-01-31 23:28:33 | 2021-01-31 23:56:36 | |

11908468 rows × 24 columns


An illustration featuring two hands. The hand on the left is light orange and holds a dark blue car key. The hand on the right is dark brown and is reaching out. In the center, a white square with an orange border contains the number '06' in orange. The background is a light beige color with stylized white clouds, small orange stars, and a landscape at the bottom with orange hills and two stylized orange trees.

06

Expected Results



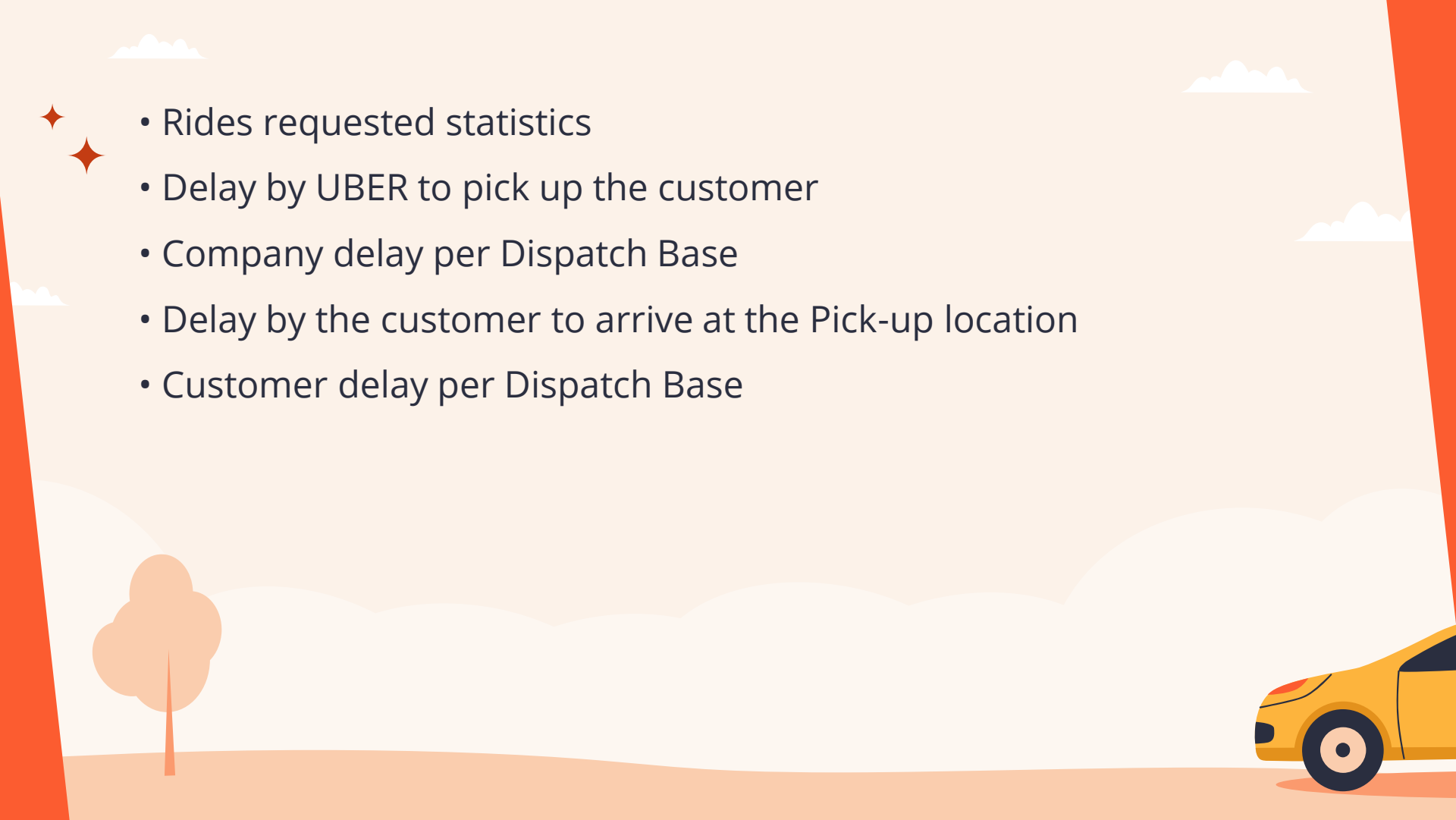
EXPECTATIONS

1. What time of the day do users request most taxis?
 2. Average distance traveled by taxi.
 3. Taxi zone where more taxis are requested.
 4. Percentage of wheelchair-accessible vehicles requested.
 5. Insights about shared rides.
- 

An illustration featuring two hands, one light orange and one brown, exchanging a dark blue car key. In the center, a white square with an orange border contains the number '07' in orange. The background is a light beige color with stylized white clouds, small orange stars, and two orange trees on a light orange ground. The entire scene is framed by orange borders on the left and right sides.

07

Scope Of Analysis

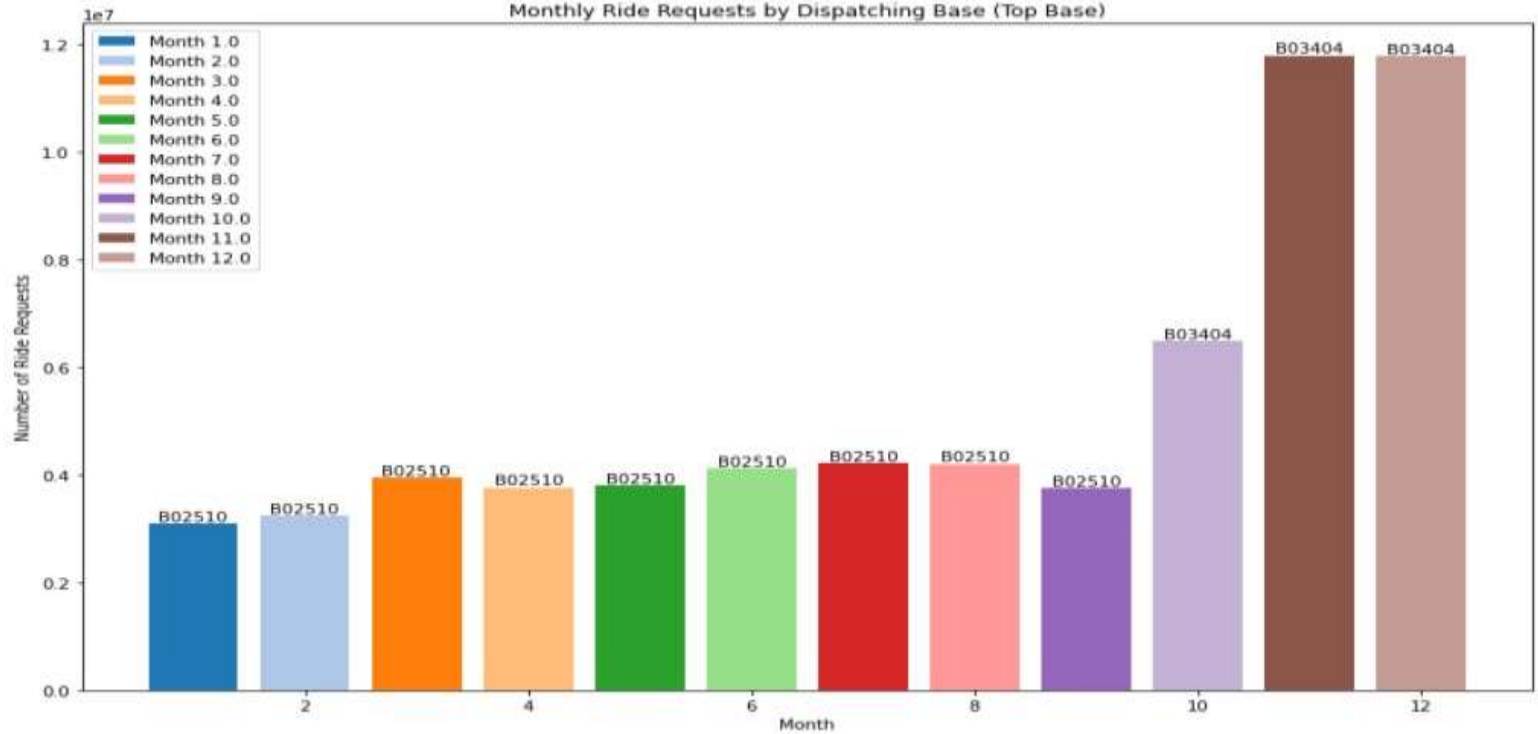
- 
- Rides requested statistics
 - Delay by UBER to pick up the customer
 - Company delay per Dispatch Base
 - Delay by the customer to arrive at the Pick-up location
 - Customer delay per Dispatch Base



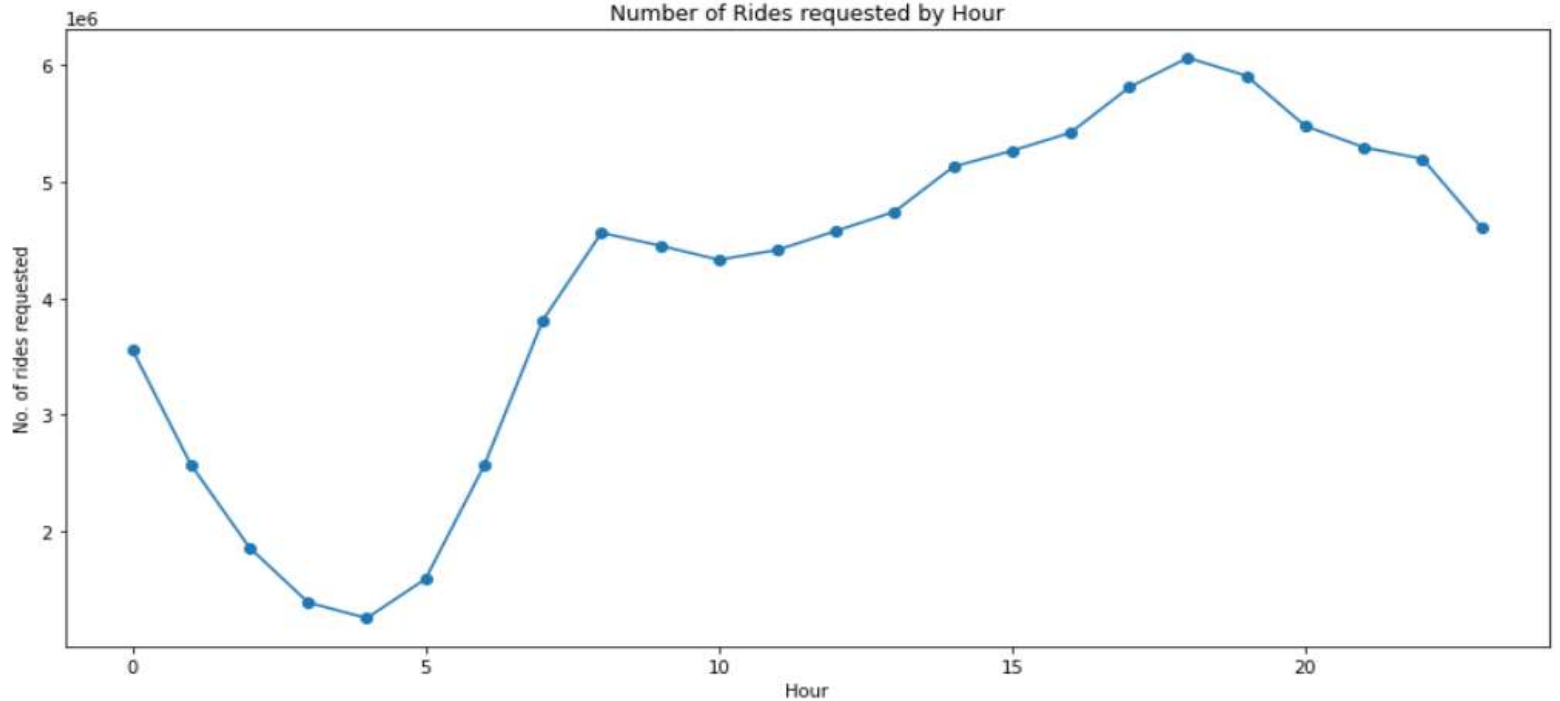
08

Outcomes Of Analysis

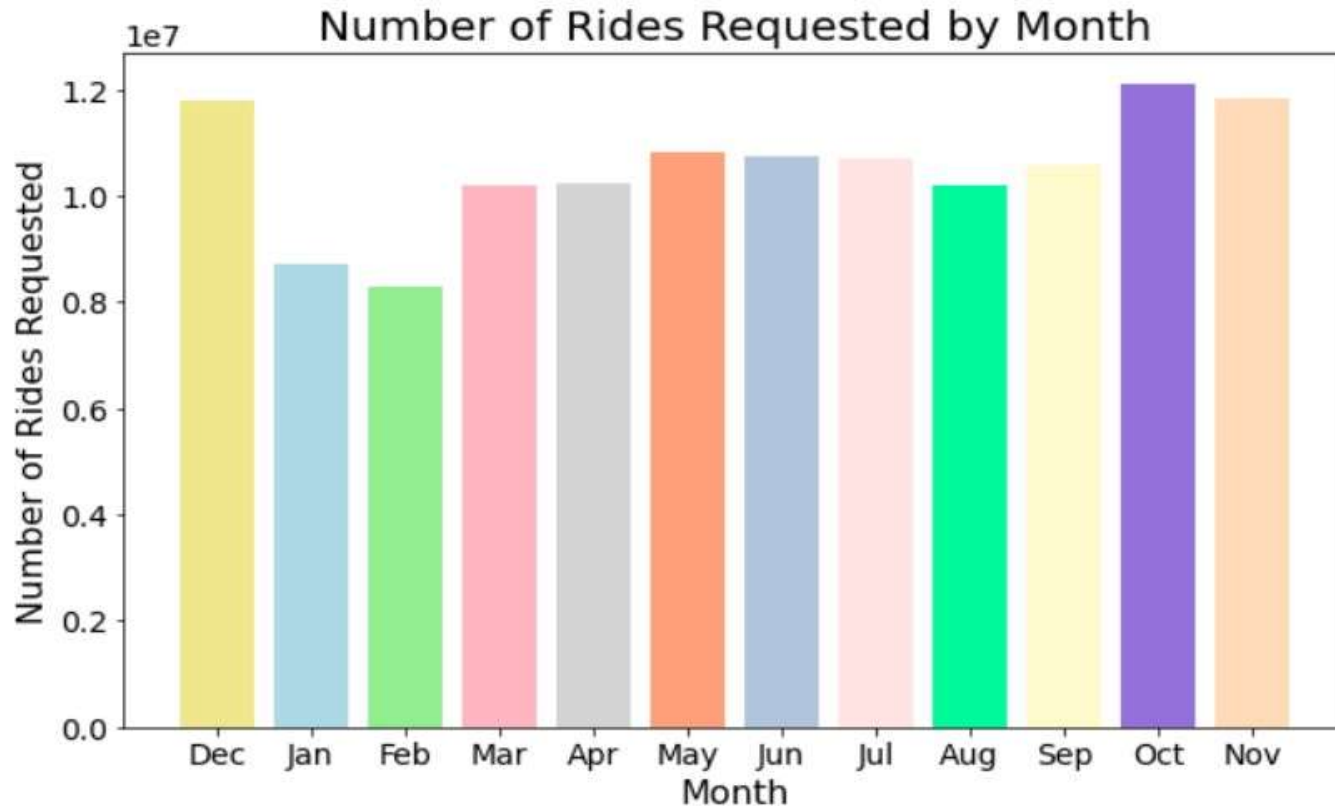
Monthly Top Dispatching Base as Per Ride Requested



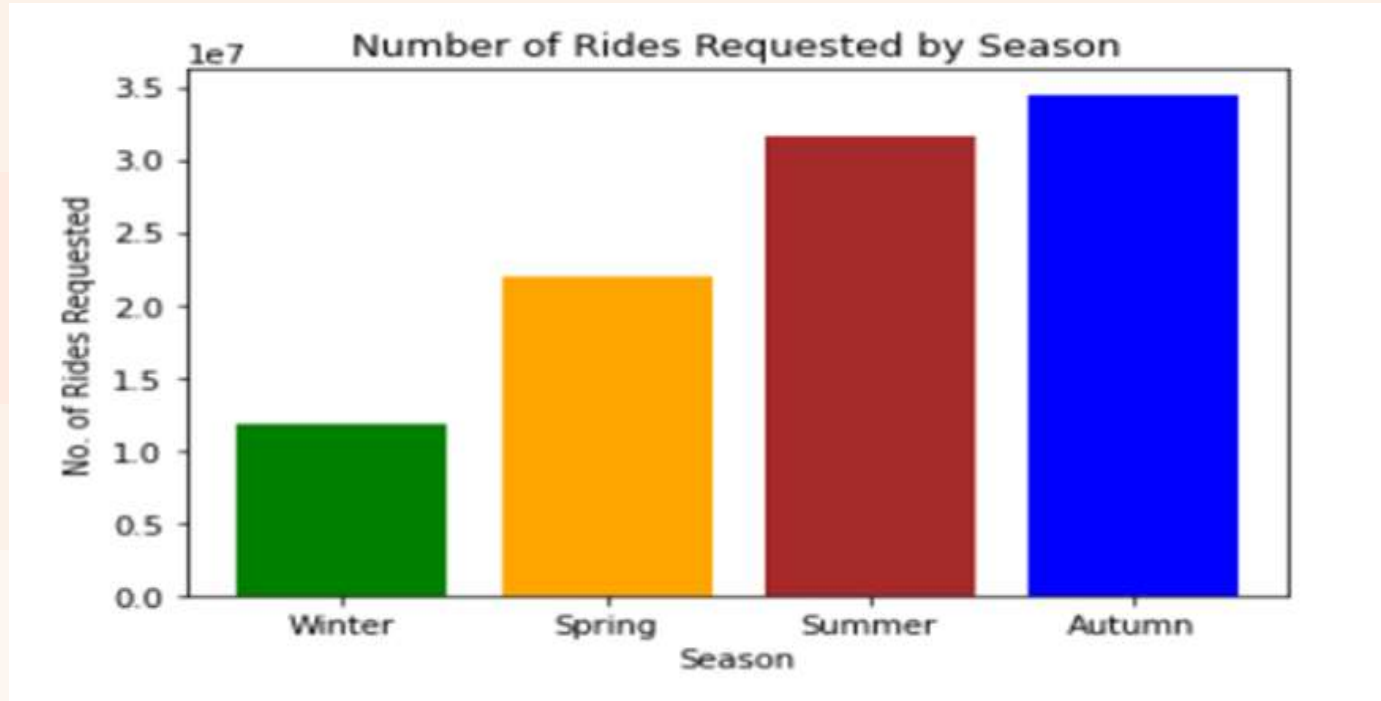
No. of Rides Requested Per Hour



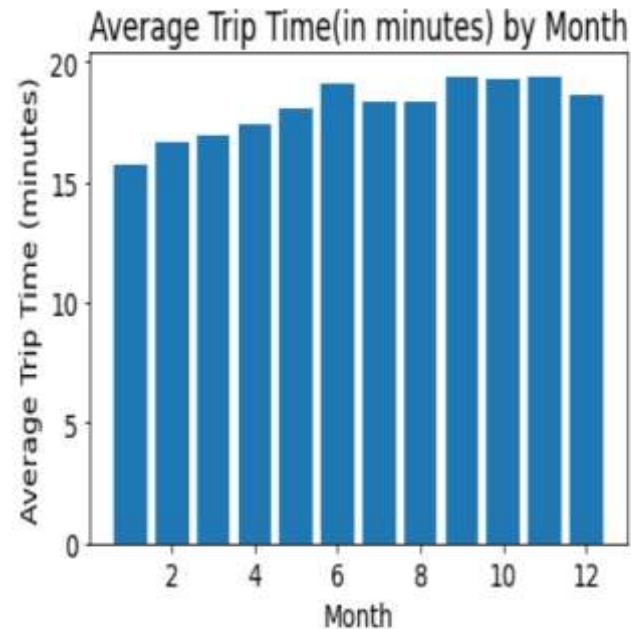
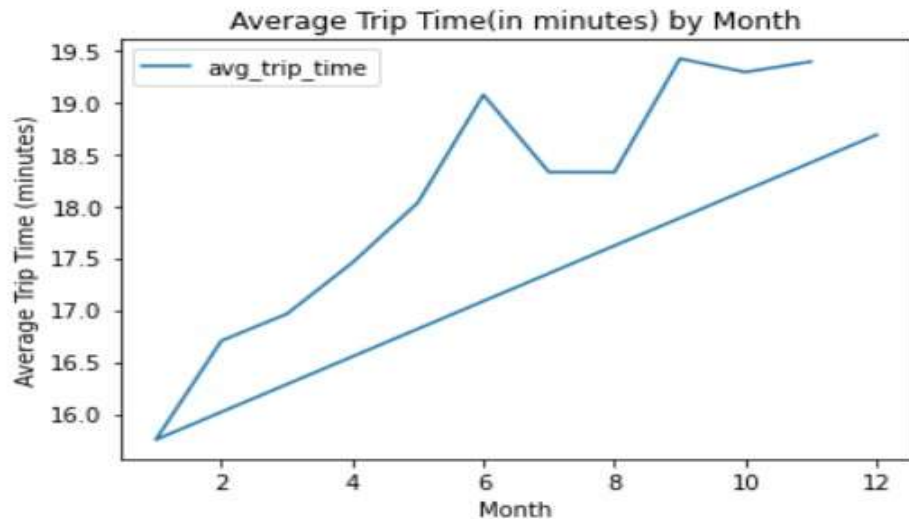
No. of Rides Requested Per Month



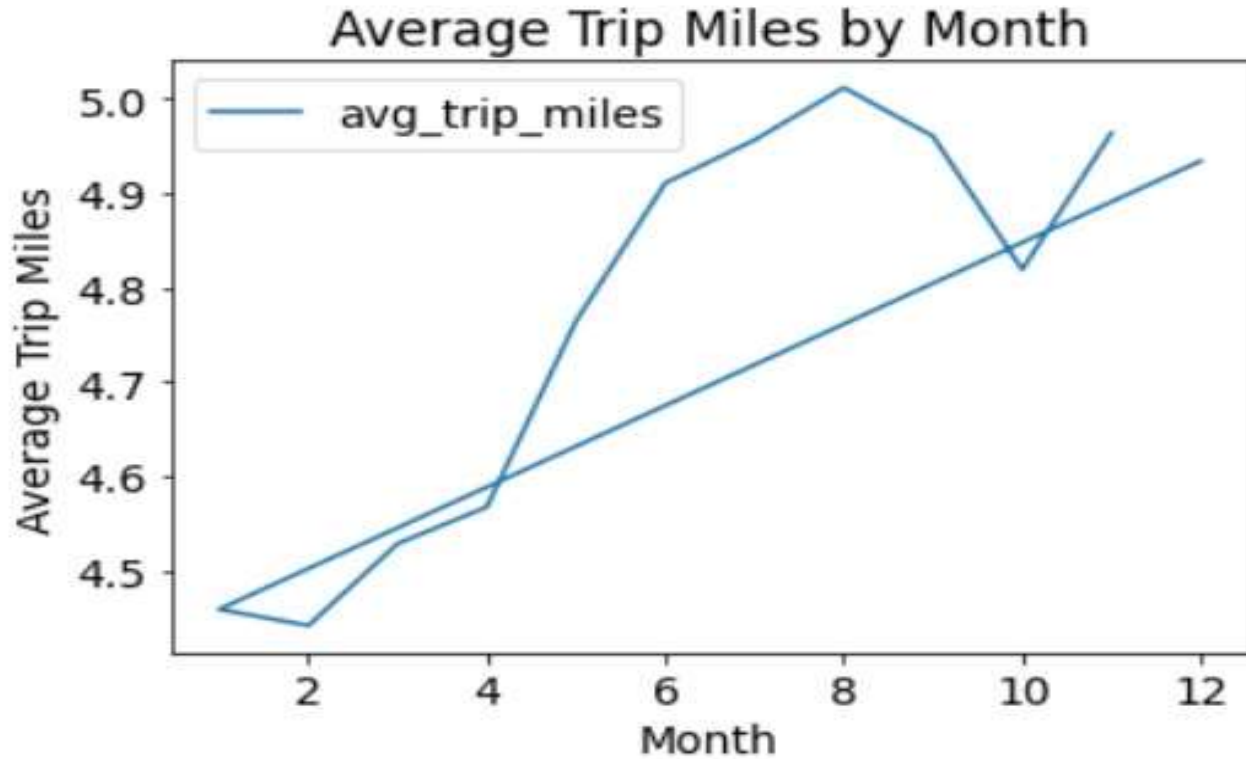
No. of Rides Requested Per Season



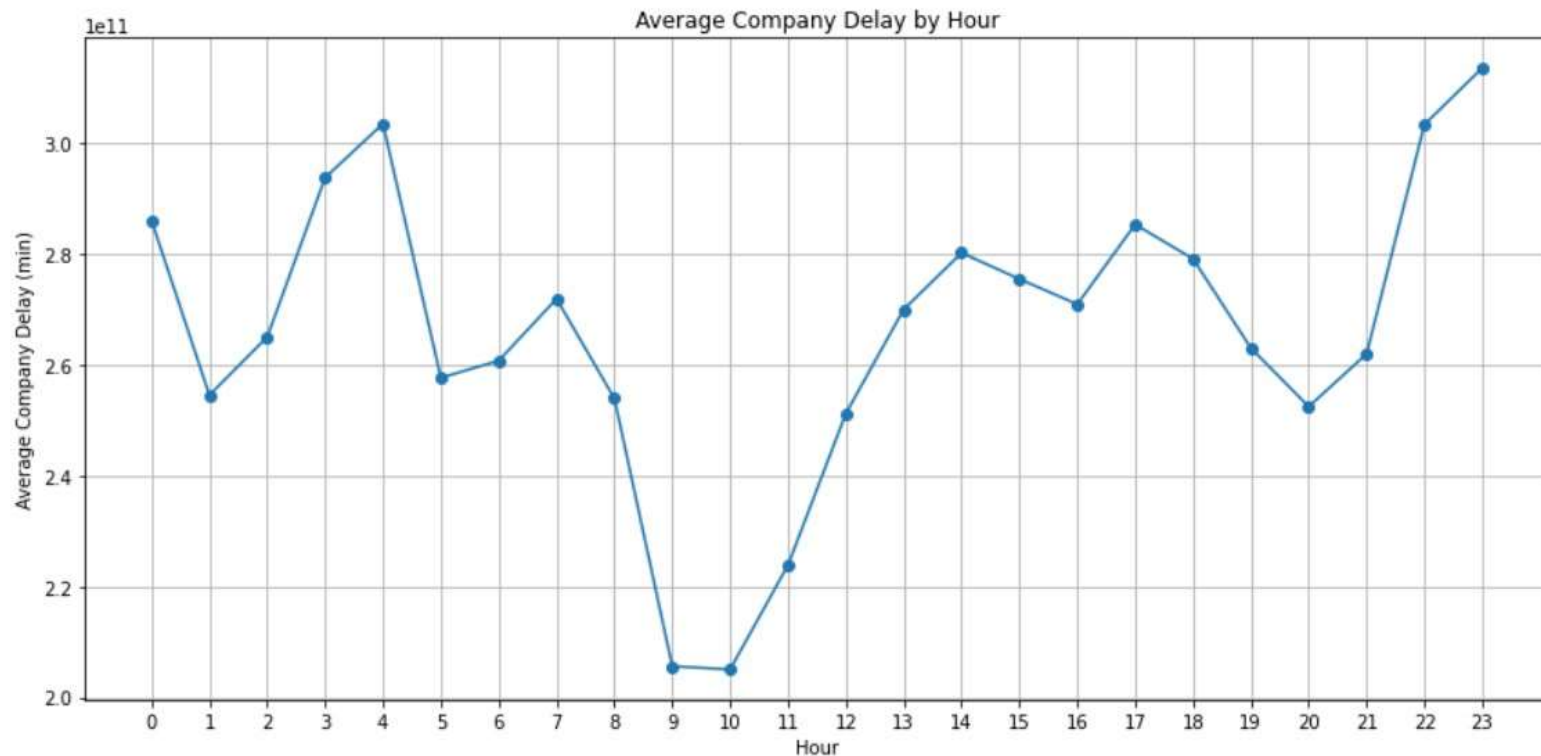
Average Trip Time(in minutes) Per Month Average



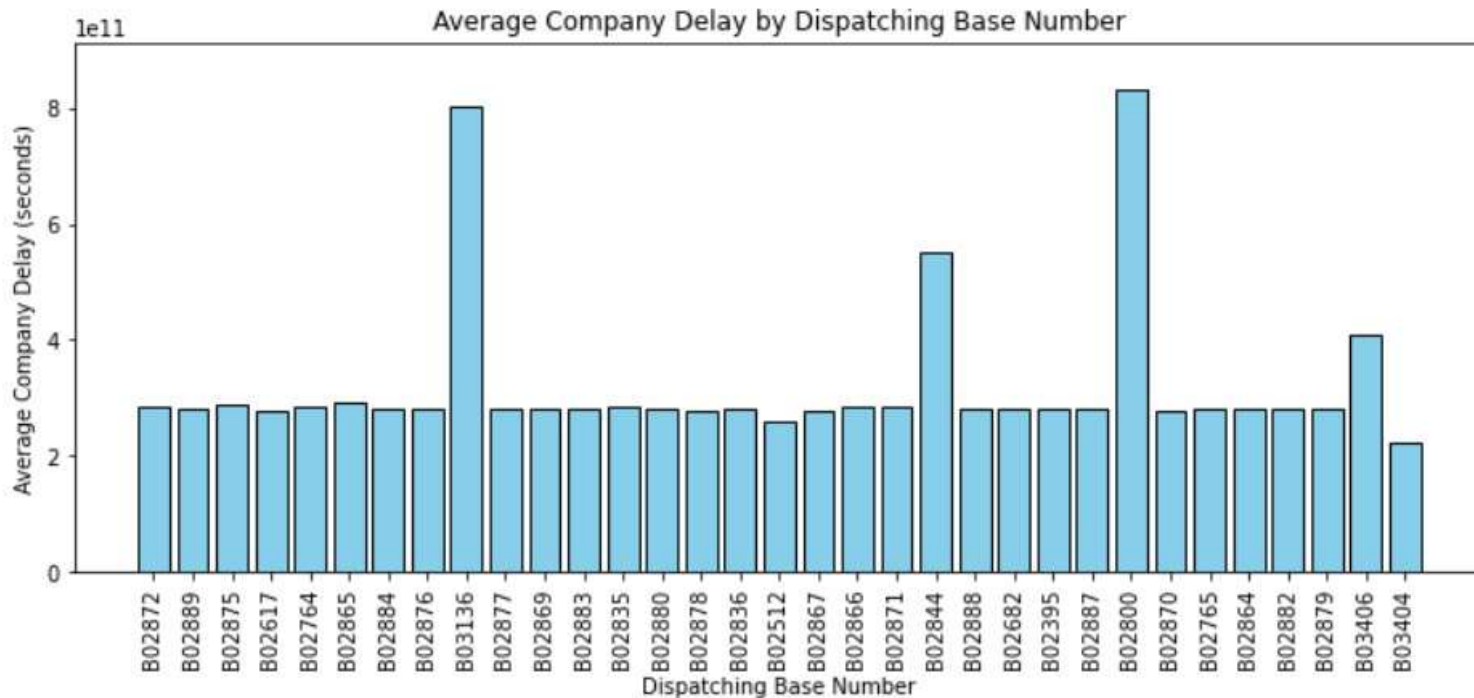
Average Trip Miles Per Month



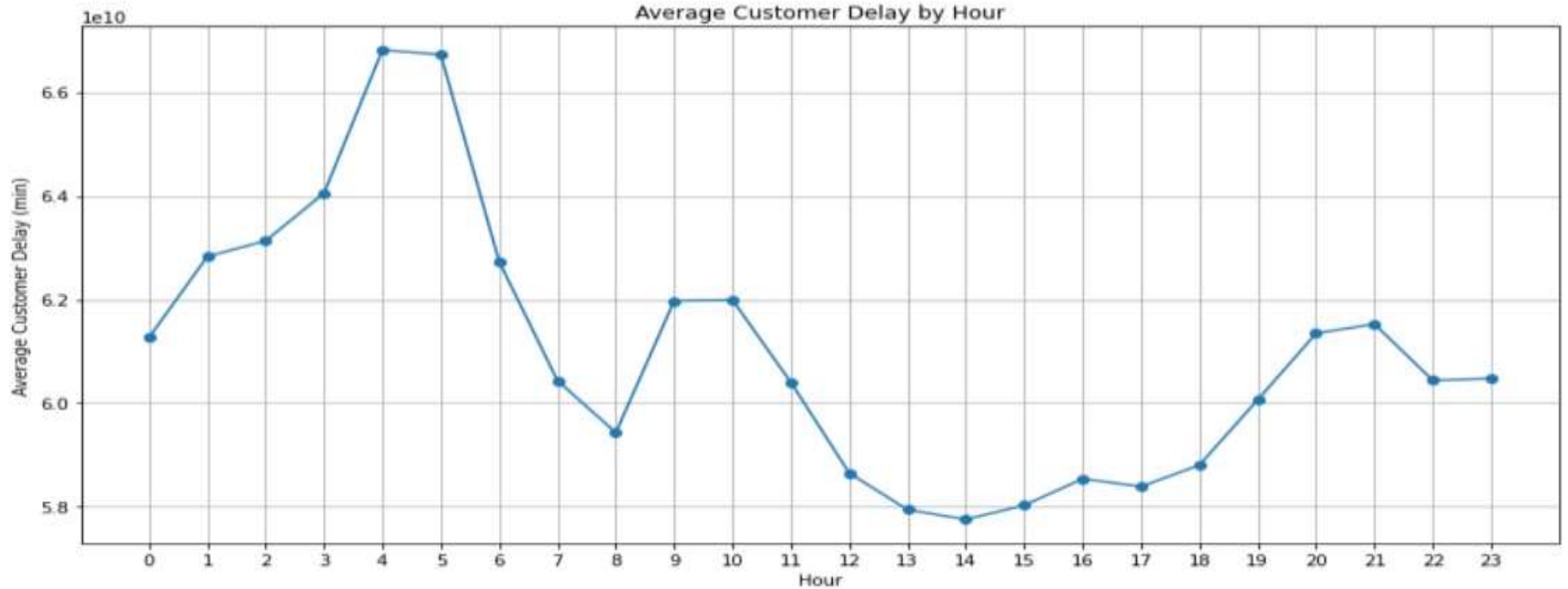
Average Company Delay Per Hour



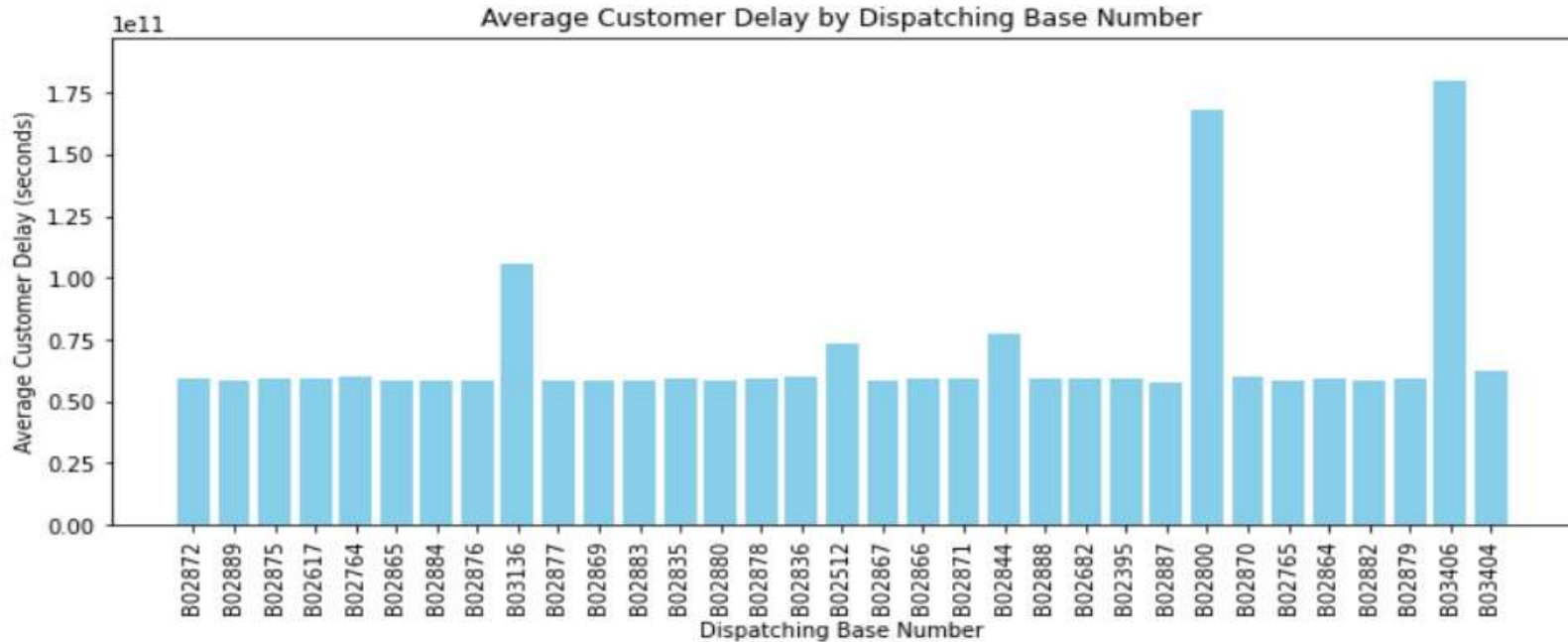
Average Company Delay Per Dispatching Base



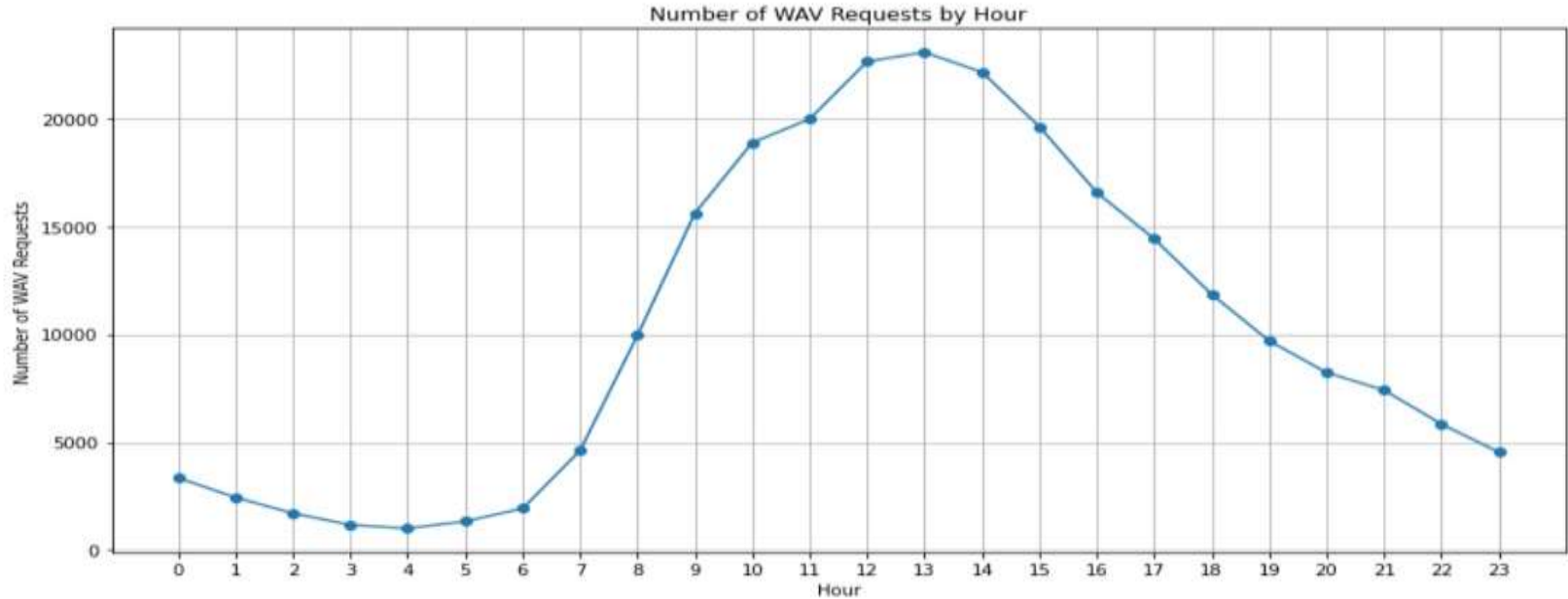
Average Customer Delay Per Hour



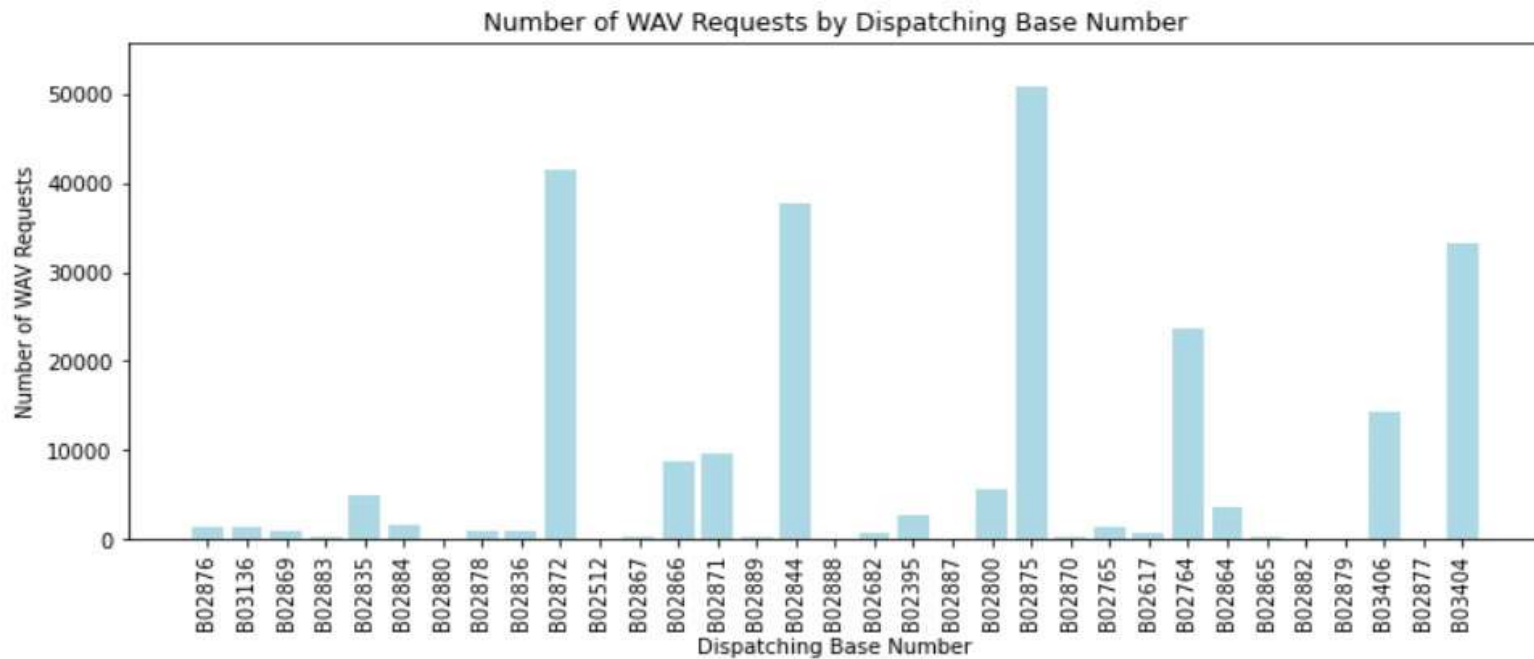
Average Customer Delay Per Dispatching Base



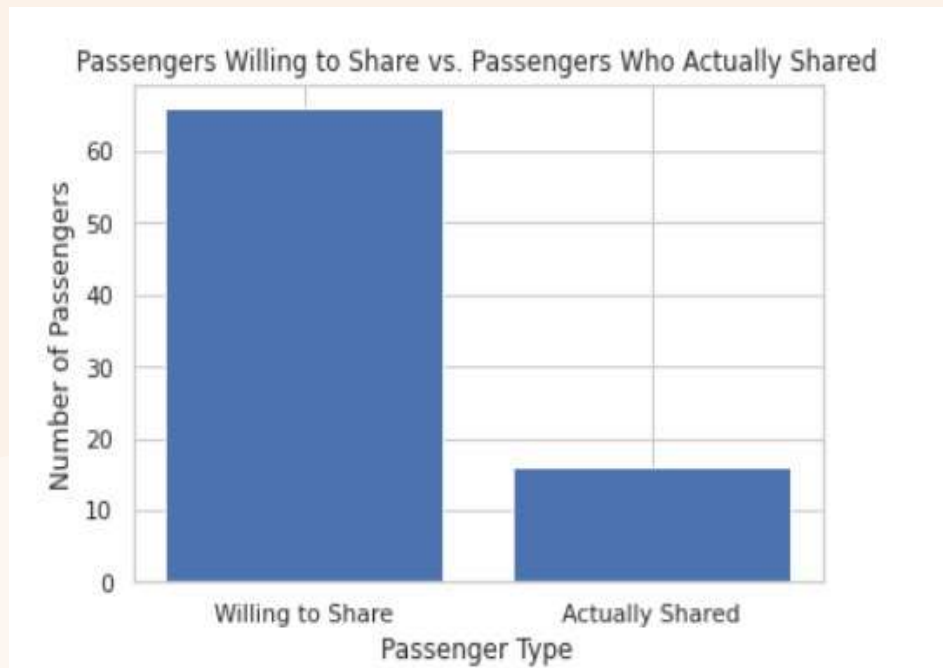
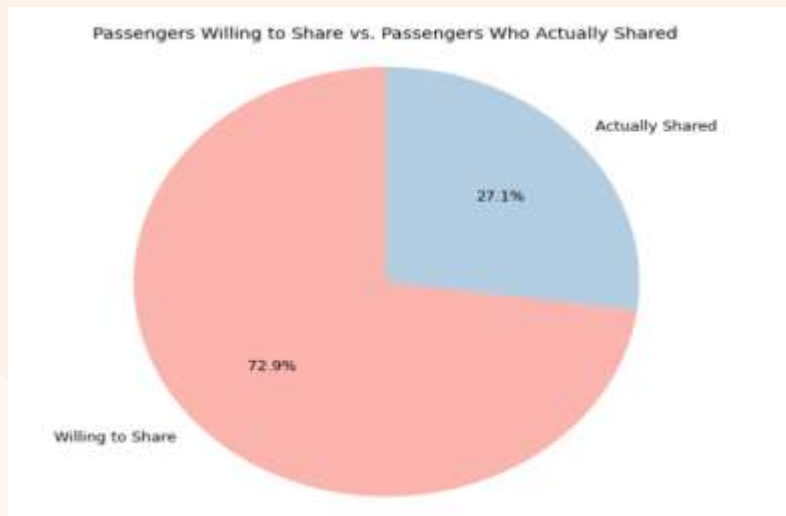
No. of Wheel-Chair Accessible Vehicles Requested Per Hour



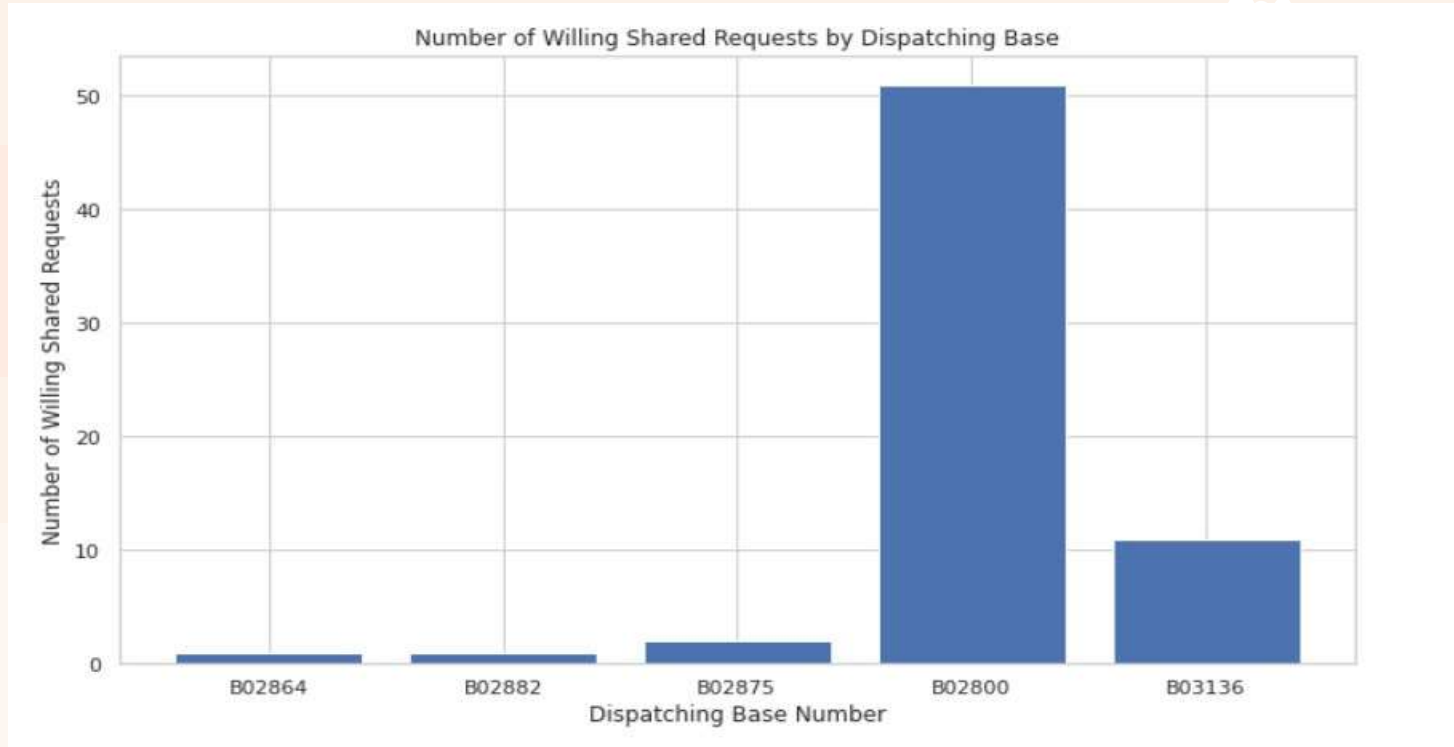
No. of Wheel-Chair Accessible Vehicles Requested Per Dispatching Base



Passengers Willing to Share Ride Vs. Who Actually Shared



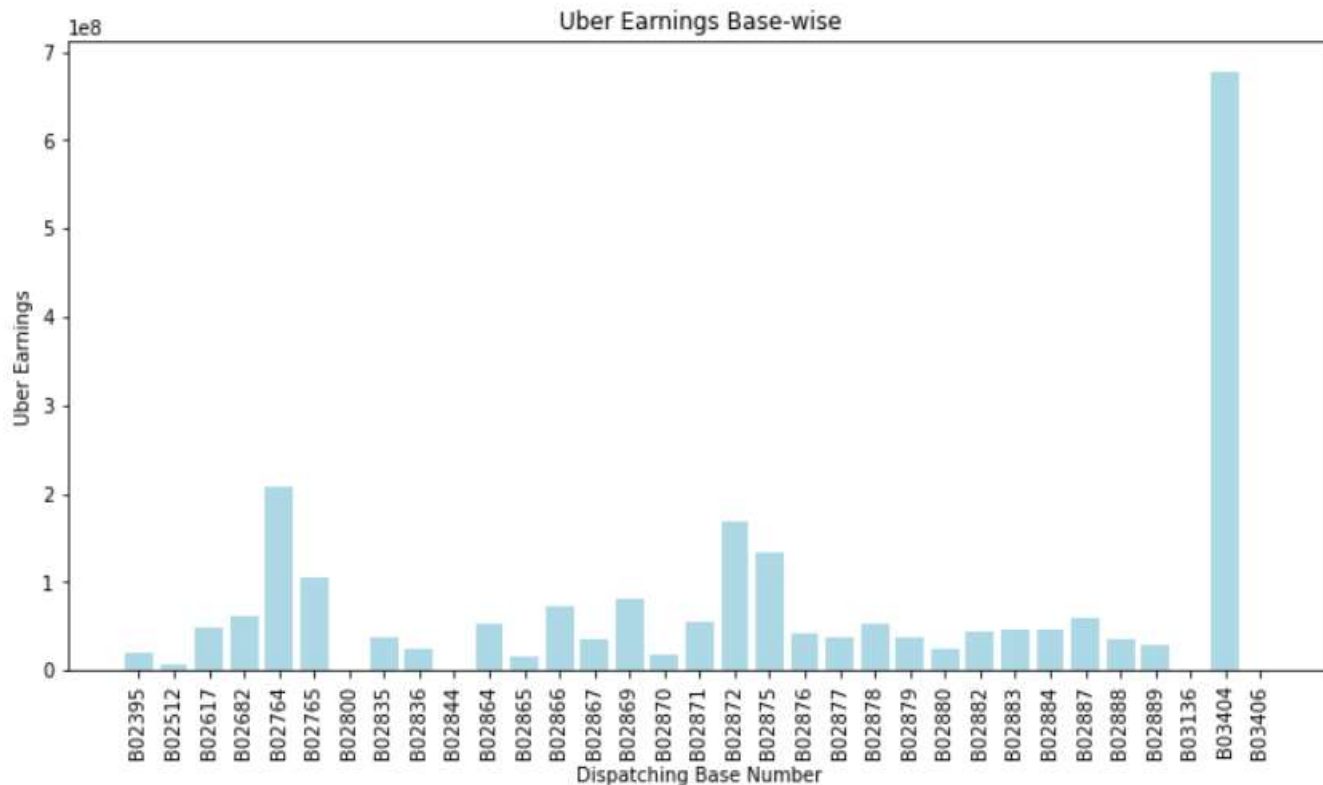
No. of Willing Shared Requests by Dispatching Base



UBER Earnings Per Month



UBER Earnings Per Dispatching Base





08

Benefits Of Analysis



Analyze ride statics hourly, monthly, and season-wise.



Analyzing Company Delay



Positioning of Wheelchair Accessible Vehicles



Analyzing customer delay



Planning the Maintenance of vehicles



Position its vehicles at the correct Dispatching Base during busy hours



09

References

- Uber NYC for-hire vehicles trip data (2021). (2023, February 2).
- Get a Vehicle License - TLC. (n.d.). <https://www.nyc.gov/site/tlc/vehicles/get-a-vehicle-license.page>
- Earn Money by Driving or Get a Ride Now | Uber Ireland. (n.d.). Uber. <https://www.uber.com/>
- What is data warehousing on Databricks? (n.d.). Databricks on AWS. <https://docs.databricks.com/sql/index.html>
- PySpark Overview — PySpark 3.4.0 documentation. (n.d.). <https://spark.apache.org/docs/latest/api/python/>

Link for notebook:

<https://community.cloud.databricks.com/?o=2546614800643647#notebook/485734581331326>



Thank You!

Any questions?

