ITC 686 – BIGDATA ANALYTICS

Analysis of NYC FHVs and HVs Dataset

Team Members

Siva Venkata Chaitanya Nannapaneni(NANNA2S)

Ajay Adakula(ADAKU1A)

Kyathi Pasham(PASHA1K)

Professor

Dr. Lisa Gandy

**Introduction:**

The New York City For-Hire Vehicle (FHV) and High-Volume (HV) data contain records of taxi rides that have taken place in the city over the years. The dataset contains a wealth of information that can be analyzed to understand trends and patterns in taxi usage in New York City. In this report, we will be exploring the dataset to analyze the number of trips, taxi rides, tips, and average tips per trip by pickup and drop-off zones.

**Data Cleaning:**

Before any analysis can be performed, it is important to ensure that the dataset is free of any null or empty values. Therefore, the first step of this project would be to clean the dataset by removing any such values that may exist.

**Dataset Creation:**

The dataset contains files for each month, and thus, the next step would be to combine all these files to create one big dataset that can be used for analysis.

Adding new columns like date, weekday, time of the day required for additional analysis.

**Analysis:**

The analysis would be carried out using Python programming language and various libraries such as Pandas, Matplotlib, and Seaborn. The following analysis will be carried out.

1. Number of trips on each day of the week.

Using the dataset, we will analyze the number of trips taken on each day of the week. This will help us understand the pattern of taxi usage in the city and the days when the demand for taxis is the highest.

A picture containing chart

Description automatically generated

1. On Monday, As time progress, number of trips are increasing.
2. Highest number of trips are taken on Friday and Saturday Evenings

2. *Average tips for trip per zone*:

We will also analyze the number of taxi rides at different times of the day on each day of the week. This analysis will provide insights into the peak hours of taxi usage in the city.

Chart, bar chart

Description automatically generated

Trips to Newark Airport, LaGuardia Airport and FK Airport has highest tips per trip. Trips to residential areas like Columbia Street has lowest tips per trip.

1. *Tips throughout different times of day on each day of week:*

Using the dataset, we will analyze the tips given by passengers at different times of the day on each day of the week. This analysis will help us understand the tipping behavior of passengers and the times of the day when they are more likely to tip.

Treemap chart

Description automatically generated with medium confidence

Most tips are calculated at early in the mornings and late at nights. Even during mid days during office hours collected max tips.

1. *Number of taxi rides of times on each day of the week:*

We will analyze the average tips per trip by pickup and drop-off zones. This analysis will help us understand which zones have the highest average tips and which zones have the lowest average tips.

Chart, bar chart

Description automatically generated

Friday Evenings and Saturday Evenings have maximum taxi rides. Saturday and Sunday nights have next highest taxi rides. Least number of trips are seen on Sunday Mornings.

5. *Average tips for trip by pickup and dropoff Zones:*

Using the dataset, we will analyze the average tips per trip for each zone. This analysis will help us understand the zones where passengers are more likely to tip generously.

Calendar

Description automatically generated with medium confidence

From the maps, staten island, have average tips given for both pickup and drop off zones. Highest tips at pick up location is Queens. Least tips are seen on Bronx.

Conclusion:

In conclusion, this project aims to explore the NYC FHVs and HVs dataset to understand the patterns and trends of taxi usage in the city. By analyzing the number of trips, taxi rides, tips, and average tips per trip by pickup and drop-off zones, we will gain insights into the behavior of passengers and the areas of the city with the highest and lowest taxi usage. This information can be used by taxi companies and the city administration to optimize taxi services and better serve the passengers.