COMS4060A: Data Visualisation and Exploration - Assignment 2

Lisa Godwin 2437980 **Nihal Ranchod 2427378** **Zach Schwark 2434346**

All plots and Jupyter notebook can be found at this GitHub Link.

1 Data Cleaning

For this Assignment Data Cleaning was done in two parts:

1. Data Exploration
2. Outlier Identification and Removal

1.1 Data Exploration

The initial data exploration process provided valuable insights into the structure of the dataset by visualising the distributions of key columns. During this phase, we examined aspects such as missing values, basic statistical measures, and the data-types of features present.

```
Data columns (total 11 columns):
 #   Column              Non-Null Count    Dtype
---  ------              --------------    -----
 0   id                  1458644 non-null  object
 1   vendor_id           1458644 non-null  int64
 2   pickup_datetime     1458644 non-null  object
 3   dropoff_datetime    1458644 non-null  object
 4   passenger_count     1458644 non-null  int64
 5   pickup_longitude    1458644 non-null  float64
 6   pickup_latitude     1458644 non-null  float64
 7   dropoff_longitude   1458644 non-null  float64
 8   dropoff_latitude    1458644 non-null  float64
 9   store_and_fwd_flag  1458644 non-null  object
 10  trip_duration       1458644 non-null  int64
dtypes: float64(4), int64(3), object(4)
memory usage: 122.4+ MB

None
```

	vendor_id	passenger_count	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	trip_duration
count	1458644.000	1458644.000	1458644.000	1458644.000	1458644.000	1458644.000	1458644.000
mean	1.535	1.665	-73.973	40.751	-73.973	40.752	959.492
std	0.499	1.314	0.071	0.033	0.071	0.036	5237.432
min	1.000	0.000	-121.933	34.360	-121.933	32.181	1.000
25%	1.000	1.000	-73.992	40.737	-73.991	40.736	397.000
50%	2.000	1.000	-73.982	40.754	-73.980	40.755	662.000
75%	2.000	2.000	-73.967	40.768	-73.963	40.770	1075.000
max	2.000	9.000	-61.336	51.881	-61.336	43.921	3526282.000

Figure 1: Image displaying the data types of features along with the basic statistics of the dataset.

```
Missing values in each column:
id                    0
vendor_id             0
pickup_datetime       0
dropoff_datetime      0
passenger_count       0
pickup_longitude      0
pickup_latitude       0
dropoff_longitude     0
dropoff_latitude      0
store_and_fwd_flag    0
trip_duration         0
dtype: int64
```

Figure 2: Image displaying the number of missing values in each column.

Figure 1 presents the data types of the features in the dataset alongside the basic statistics of the initial dataset. Additionally, Figure 2 confirms that there are no missing values in the dataset.

Initial distributions of key columns: `vendor_id`, `passenger_count`, `trip_duration`, `store_and_fwd_flag`.

We analyse the following features of the dataset to gain an initial understanding of their distributions and to help identify any potential outliers or anomalies.

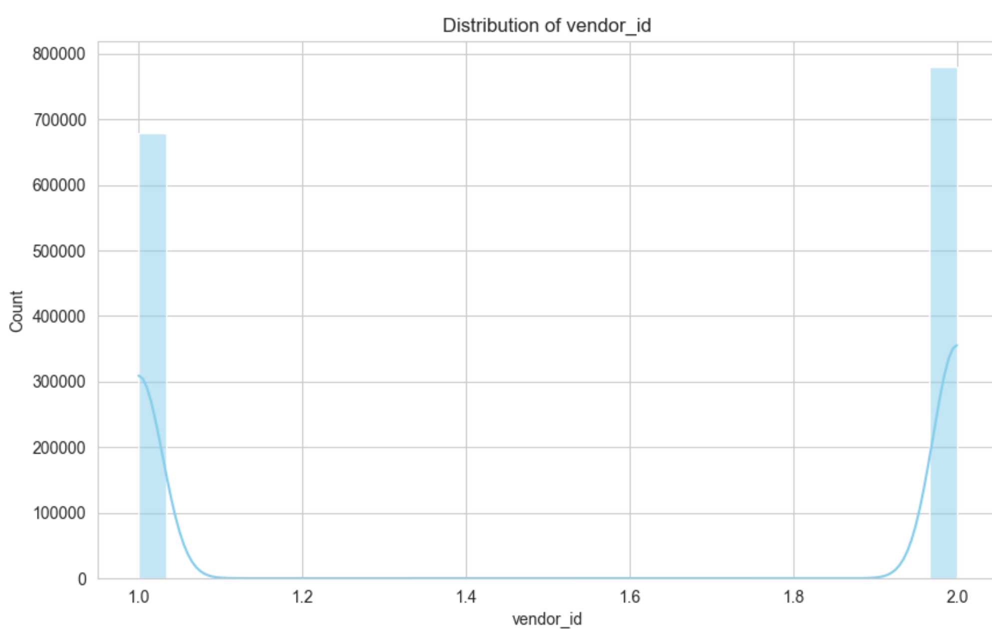

Figure 3: Distribution of `vendor_id` feature.

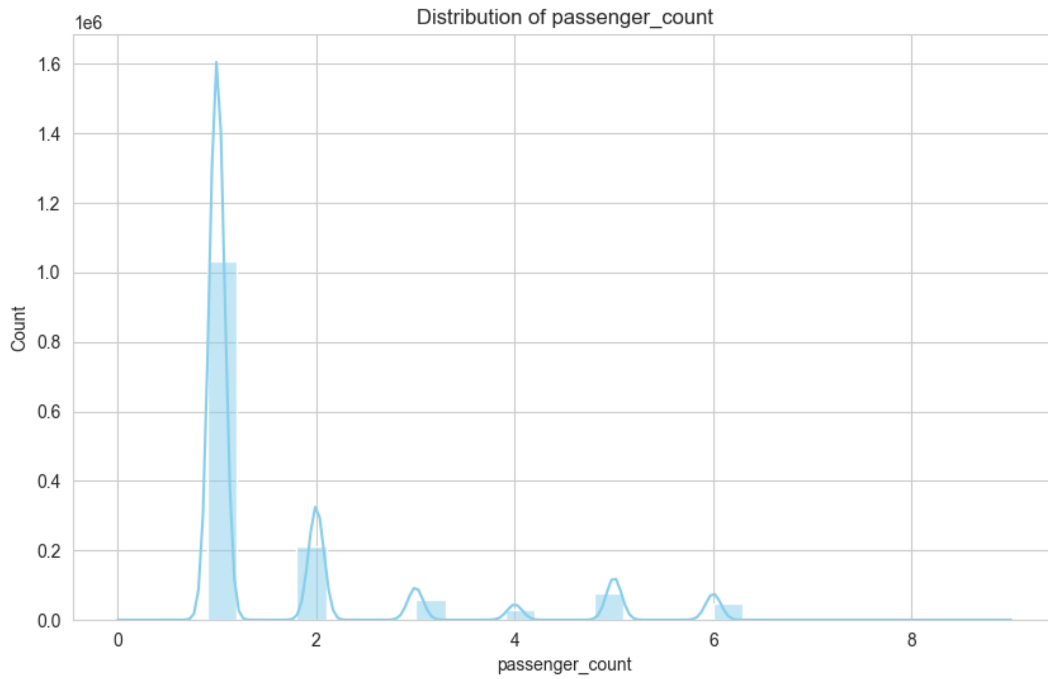

Figure 4: Distribution of `passenger_count` feature.

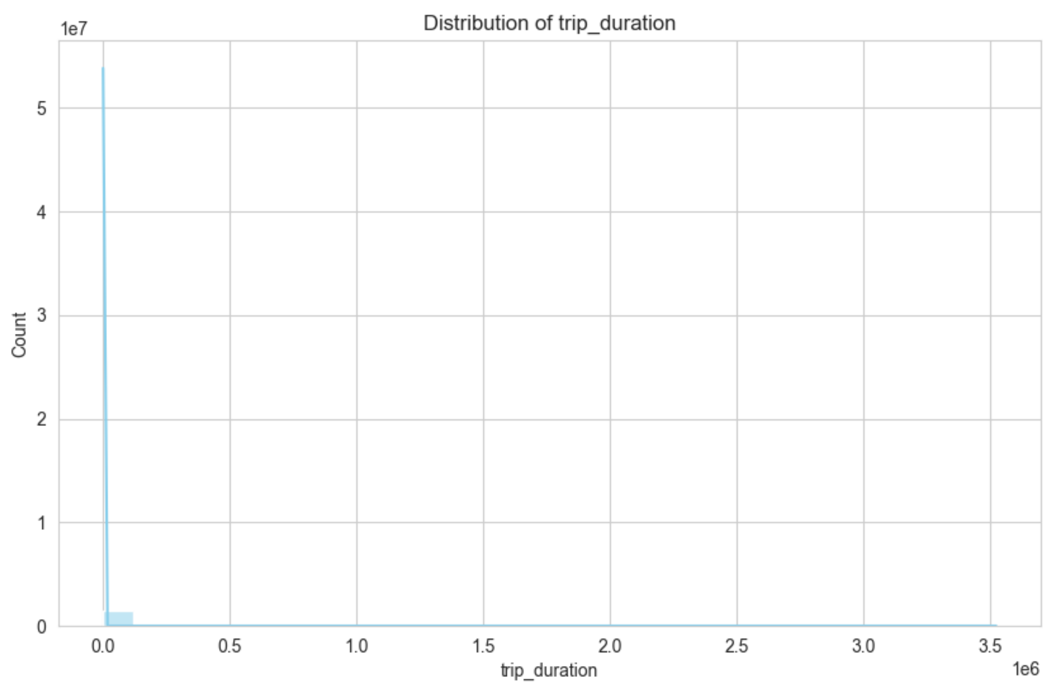

Figure 5: Distribution of `trip_duration` feature.

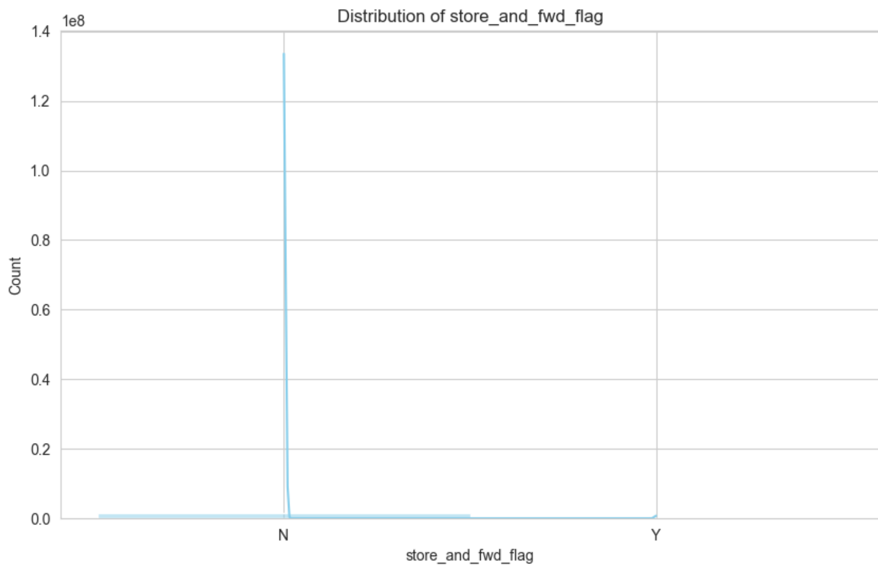

Figure 6: Distribution of `store_and_fwd_flag` feature.

1.2 Outlier Identification and Removal

1.2.1 Trip Duration & Distance

In order to get an initial idea of the outliers for trip distance and trip duration, box-plots were created to visually demonstrate this, which can be seen in Figure 7 and Figure 8 below.

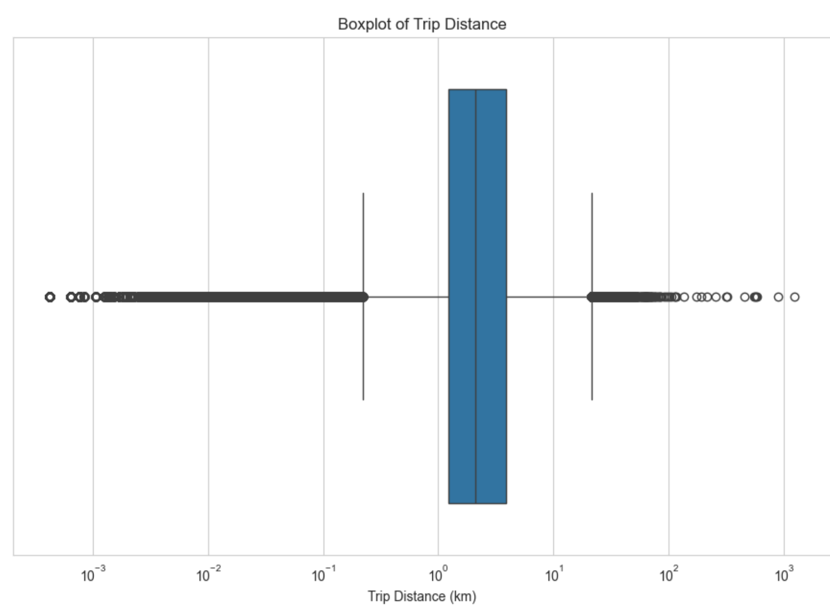

Figure 7: Boxplot of `trip_distance` feature.

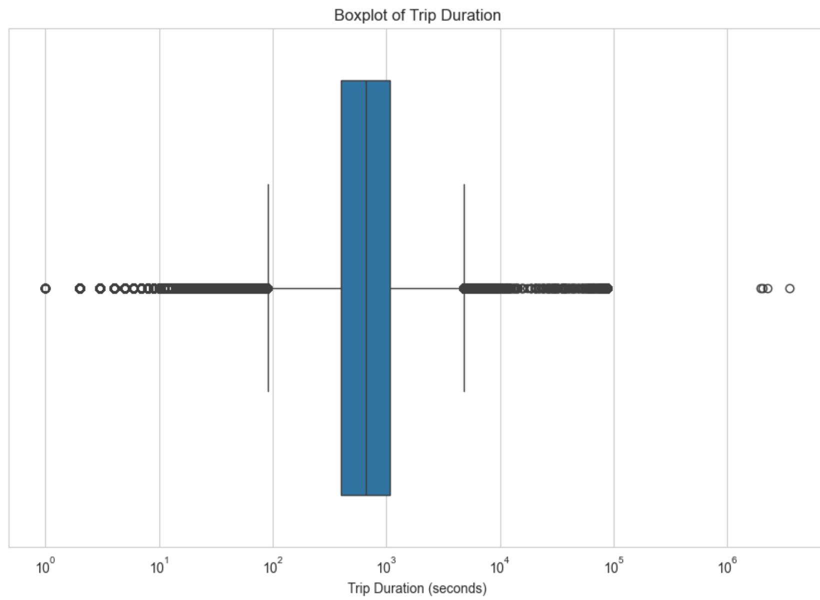

Figure 8: Boxplot of `trip_duration` feature.

Outlier Removal: The process of outlier removal was executed in several stages to ensure that the dataset is clean and accurate for further analysis. The key variable used for this process was the relationship between trip duration and trip distance, as visually represented by scatter plots, seen below. We performed multiple iterations of outlier detection and removal based on domain knowledge assumptions regarding typical trip duration for various distances. This step-by-step approach ensures that extreme outliers are systematically removed from the dataset while preserving the integrity of normal trip records for further analysis.

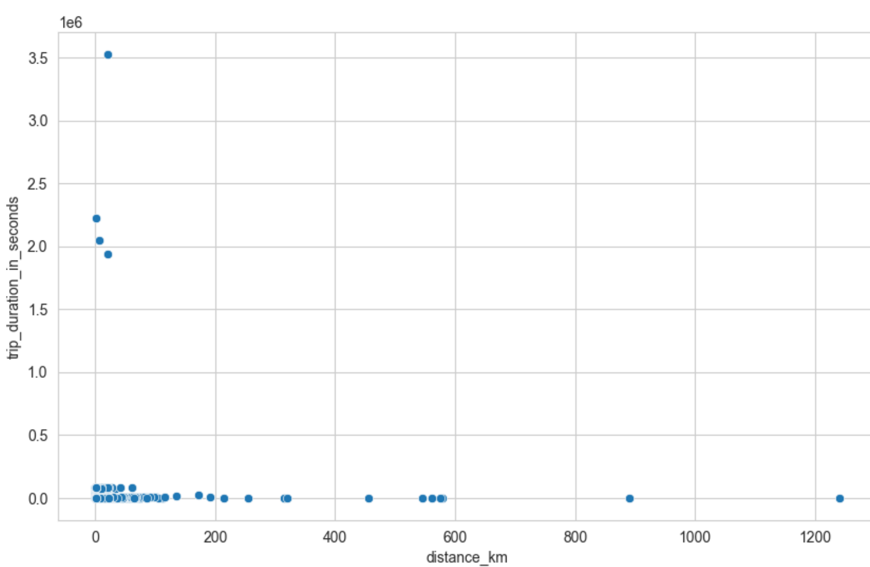

Figure 9: Initial Scatter plot of trip distance against trip duration.

From Figure 9, we can see several trips with duration's exceeding 1.5 million seconds for distances under 50 kilometres. Since it is highly unlikely that trips covering less than 25 km would take more than 8,000 seconds (approximately 2.2 hours), we have set a conservative threshold at 8,000 seconds to flag these outliers. The removal of these outliers are shown in Figure 10 below.

The same approach was applied to identify other outliers based on the trends observed in Figure 9.

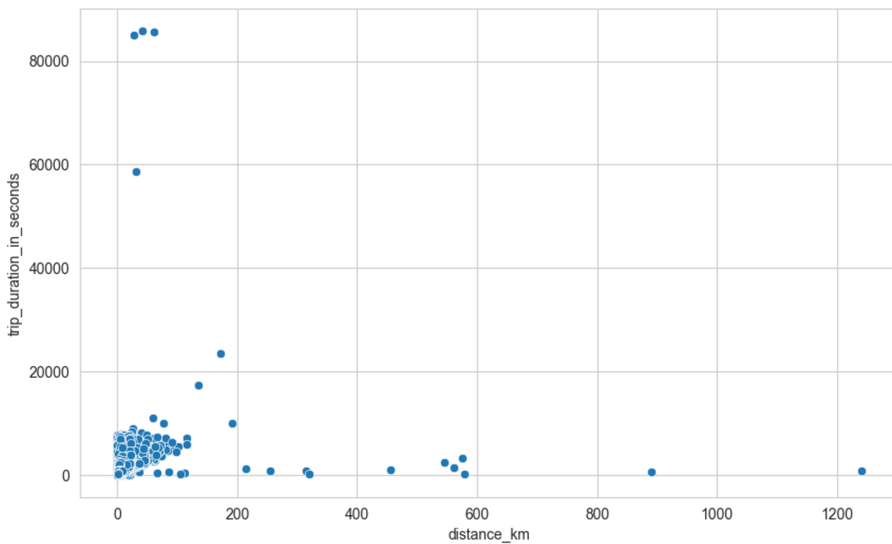

Figure 10: Scatter plot of trip distance against trip duration after removing outliers with thresholds stated above.

For trips under 200 km, trip duration's exceeding 30,000 seconds (approximately 8.3 hours) are likely outliers. The removal of these are shown in the the scatter plot below, Figure11

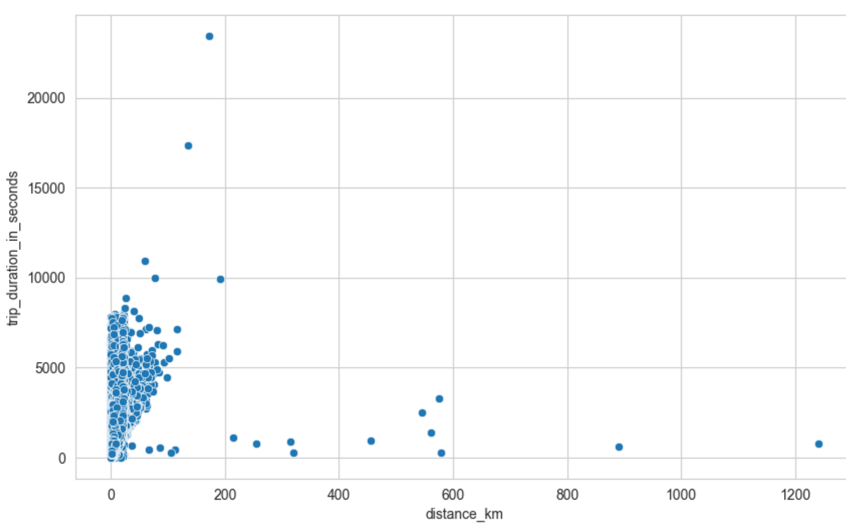

Figure 11: Scatter plot of trip distance against trip duration after removing outliers.

Based on Figure 11: Trips that covered more than 200 km but had a duration of less than 50,000 seconds (around 13.9 hours) are identified as potentially inaccurate and are removed. The removal of these outliers are shown in Figure 12 below.

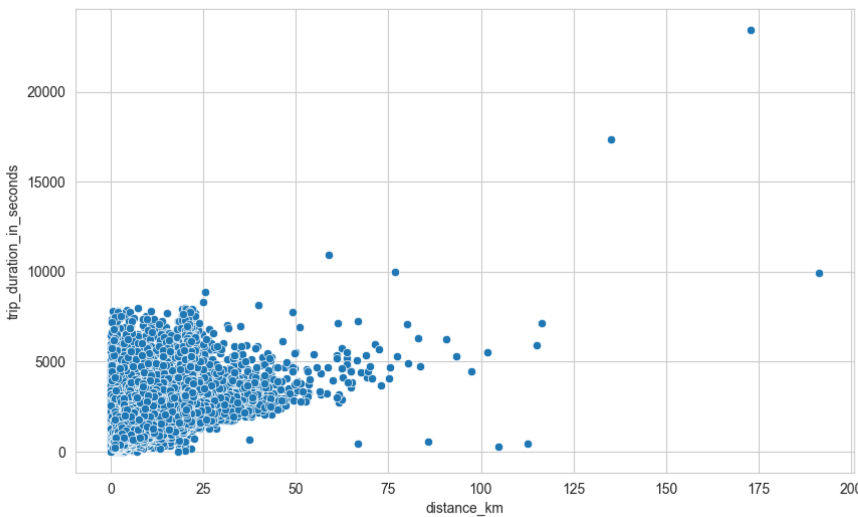

Figure 12: Scatter plot of trip distance against trip duration after removing outliers.

Based on Figure 12Trips with a distance of zero km's but a non-zero trip duration were flagged as outliers, since these records likely represent data entry errors or issues with GPS. The removal of these outliers can be seen in Figure 13 below.

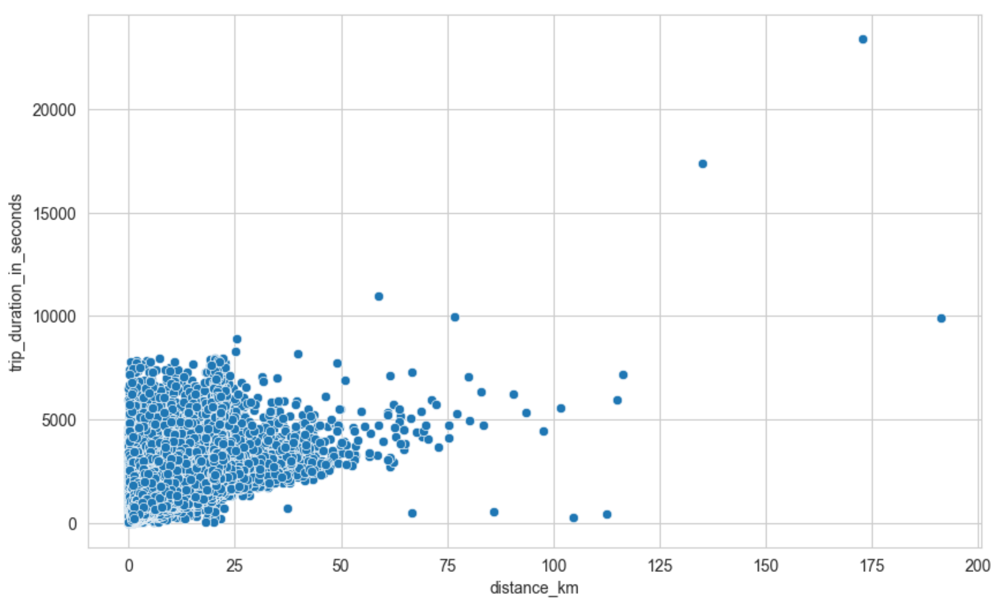

Figure 13: Scatter plot of trip distance against trip duration after removing outliers.

This is now the final scatter plot (Figure 13) with outliers removed with respect to distance and trip duration.

1.2.2 Trip Speed

Outlier Removal:

- Calculated the speed of each trip in kilometres per hour (km/h) by dividing the trip distance (in kilometres) by the trip duration (converted from seconds to hours). This metric helps identify trips with implausibly high or low speeds.
  - A box-plot was generated to visualise the distribution of trip speeds, allowing us to identify extreme outliers, Figure 14.
  - The x-axis was displayed on a logarithmic scale to capture a wide range of speeds, as some trips might exhibit unusually high or low values.
- Identify and remove extreme speed outliers, we employed the Z-score method, which measures the number of standard deviations a data point is from the mean.
  - Z-scores were calculated for the speed_kmh column using the `stats.zscore()` function. a Z-score threshold of 3 was used.

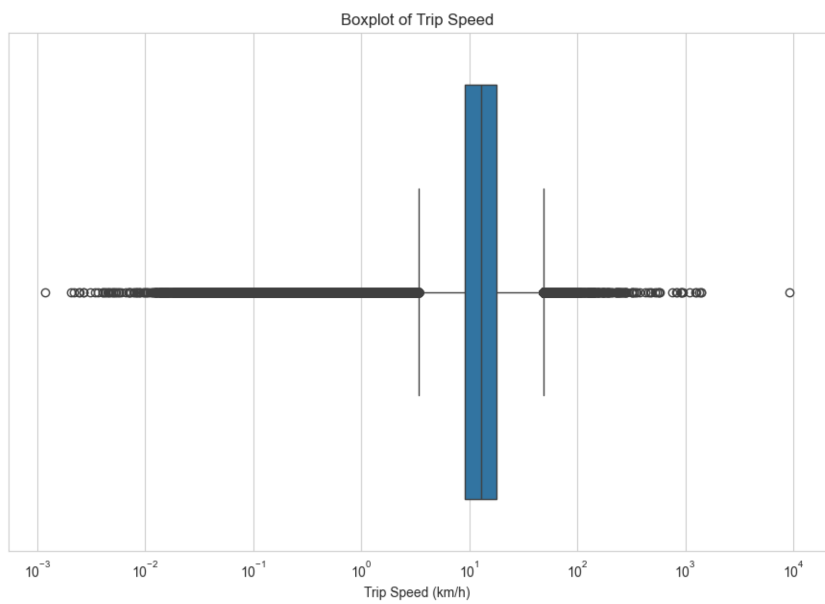

Figure 14: Boxplot of the initial trip speeds with outliers.

1.2.3 Passenger Count

Filtered out trips where the passenger count was 0, as it is not reasonable for a taxi trip to have zero passengers. Such entries were likely errors or anomalies in the data.

```
passenger_count
1    1025583
2     208702
5      77379
3      59439
6      47801
4      28156
0         49
7          2
Name: count, dtype: int64
```

(a) Passenger count statistics before outlier removal.

```
passenger_count
1    1025583
2     208702
5      77379
3      59439
6      47801
4      28156
Name: count, dtype: int64
```

(b) Passenger count statistics after outlier removal.

Figure 15: Comparison of passenger count statistics before and after outlier removal.

2 Feature Engineering

Note: Distance of trip and average speed of trip were already calculated in the Data Cleaning steps in Question 1, which can be seen in Figure 16.

	trip_distance_km	speed_kmh
0	1.498	11.849
1	1.804	9.798
2	6.381	10.815
3	1.485	12.458
4	1.188	9.830
...	...	...
1447055	1.224	5.665
1447056	6.046	33.230
1447057	7.820	36.847
1447058	1.092	10.538
1447059	1.133	20.606

1447060 rows × 2 columns

Figure 16: Image of new features: Trip Distance and Average Speed of Trip.

In the preprocessing step, several additional features were created to enhance the analysis of trip patterns. Here is an explanation of why each feature was created:

1. Day of the Week (day_of_week and day_of_week_str):
   - Purpose: Understanding the day of the week when trips occur can reveal patterns related to weekdays and weekends. This can help identify peak travel days and variations in trip volumes across different days.
   - Implementation: The day_of_week feature extracts the day of the week as an integer (0=Monday, 6=Sunday), and day_of_week_str maps these integers to their corresponding day names.
2. Hour of the Day (pickup_hour):
   - Purpose: Analysing the hour of the day when trips occur helps in understanding daily travel patterns. This can highlight rush hours, off-peak times, and other temporal trends in trip activity.
   - Implementation: The pickup_hour feature extracts the hour from the pickup date-time, providing a granular view of trip distribution throughout the day.
3. Time of Day (time_of_day):
   - Purpose: Categorising trips into different times of the day (morning, afternoon, evening, night) allows for a more intuitive understanding of travel behaviour. This can be useful for identifying periods of high and low demand.
   - Implementation: The time_of_day feature categories the pickup_hour into four time periods: Morning (5 AM - 12 PM), Afternoon (12 PM - 5 PM), Evening (5 PM - 9 PM), and Night (9 PM - 5 AM).
4. Trip Duration in Minutes (trip_duration_min):
   - Purpose: Converting trip duration from seconds to minutes makes the data more interpretable and easier to analyse. This feature is essential for understanding the length of trips and identifying patterns related to trip duration.
   - Implementation: The trip_duration_min feature is calculated by dividing the original trip duration (in seconds) by 60 to convert it to minutes.

trip_distance_km	speed_kmh	day_of_week	day_of_week_str	pickup_hour	time_of_day	trip_duration_min
1.498	11.849	0	Monday	17	Evening	7.583
1.804	9.798	6	Sunday	0	Night	11.050
6.381	10.815	1	Tuesday	11	Morning	35.400
1.485	12.458	2	Wednesday	19	Evening	7.150
1.188	9.830	5	Saturday	13	Afternoon	7.250

Figure 17: Image of new additional features created.

3 Time-based

3.1 Most Popular Day of the Week

To identify the most popular day of the week, we created a plot showing the Number of Trips per Day, as displayed in Figure 18. From the figure, it's evident that Friday is the busiest day of the week.

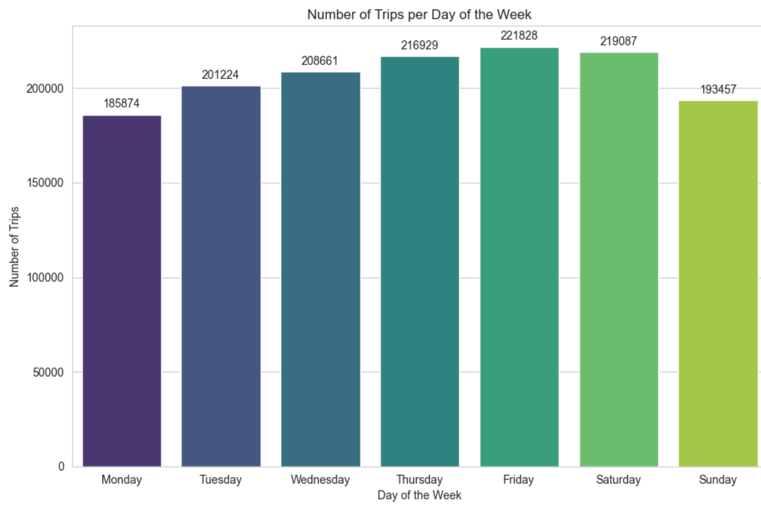

Figure 18: Barplot of the Number of Trips per Day of the Week.

3.2 Most Popular Hour of the Day for each day

Overall Distribution of Trips by Hour and Day of the Week: A count plot was created to visualise the distribution of trips by hour of the day, with different colours representing different days of the week. This plot helps in understanding the overall trip patterns across different hours and days, Figure 19.

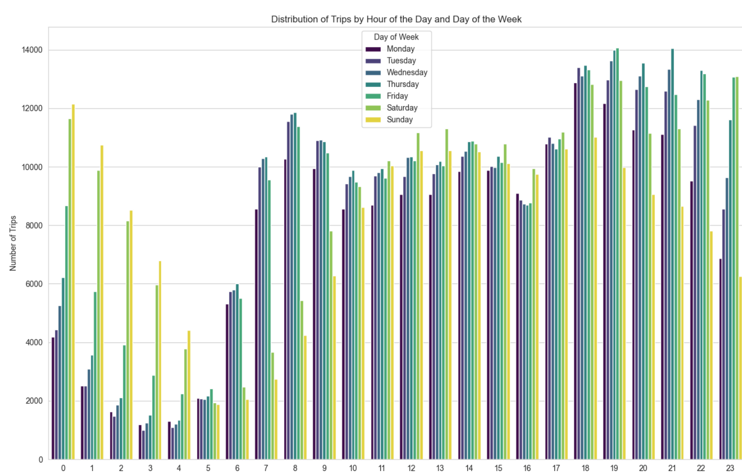

Figure 19: Count plot showing distribution of trips per hour of the day.

Individual Plots for Each Day of the Week: For each day of the week, a count plot was created to visualise the distribution of trips by hour. This helps in identifying the peak hours for each specific day.

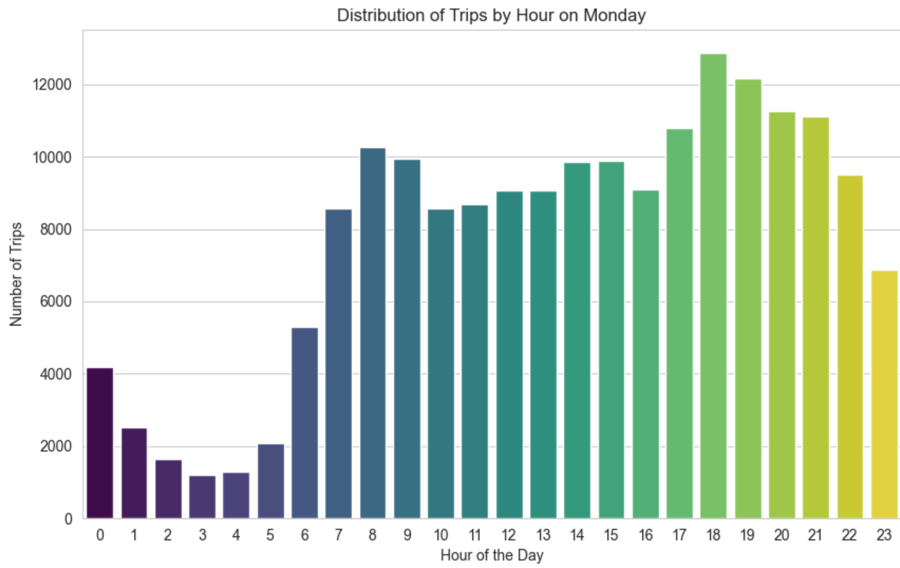

Figure 20: Trip Distribution by Hour for Monday

From Figure 20, the most popular hour on Monday is: 18.

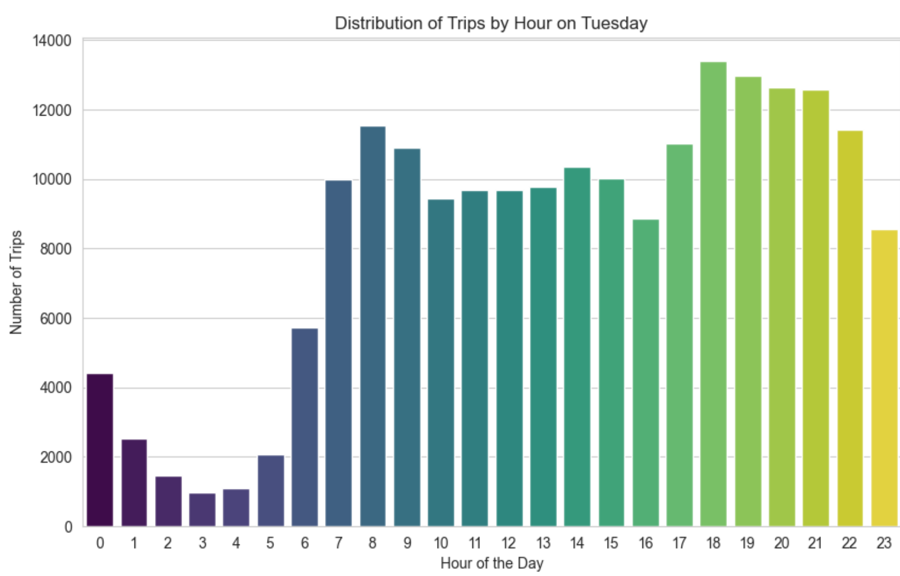

Figure 21: Trip Distribution by Hour for Tuesday

From Figure 21, the most popular hour on Tuesday is: 18.

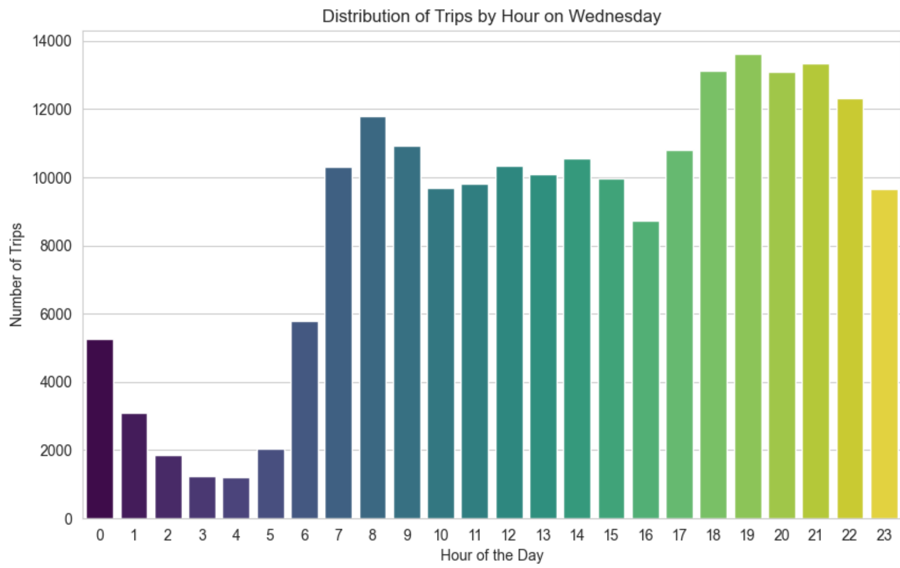

Figure 22: Trip Distribution by Hour for Wednesday

From Figure 22, the most popular hour on Wednesday is: 19.

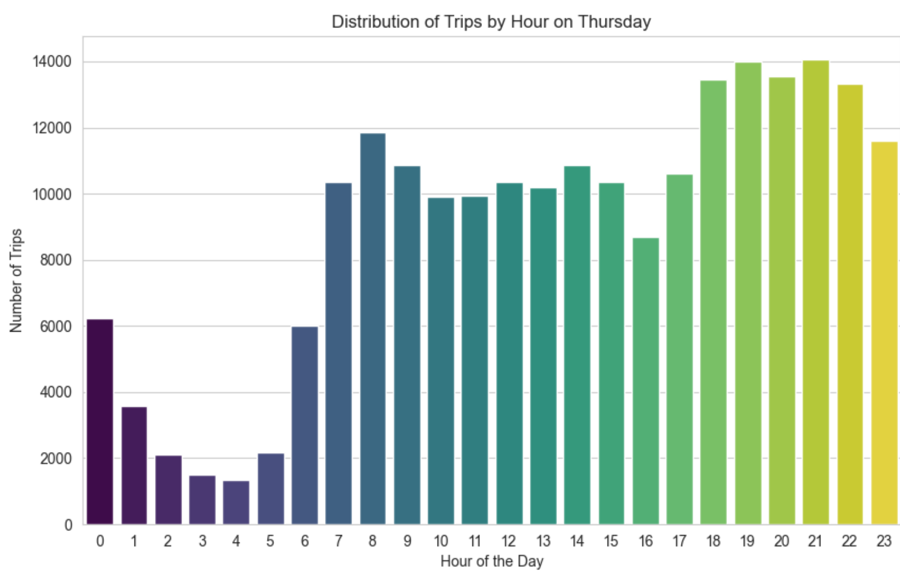

Figure 23: Trip Distribution by Hour for Thursday

From Figure 23, the most popular hour on Thursday is: 21.

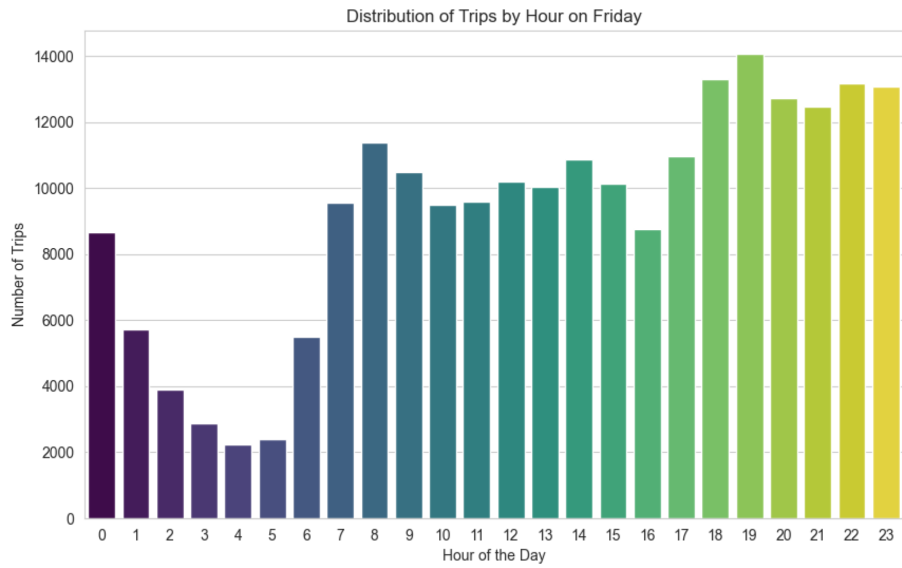

Figure 24: Trip Distribution by Hour for Friday

From Figure 24, the most popular hour on Friday is: 19.

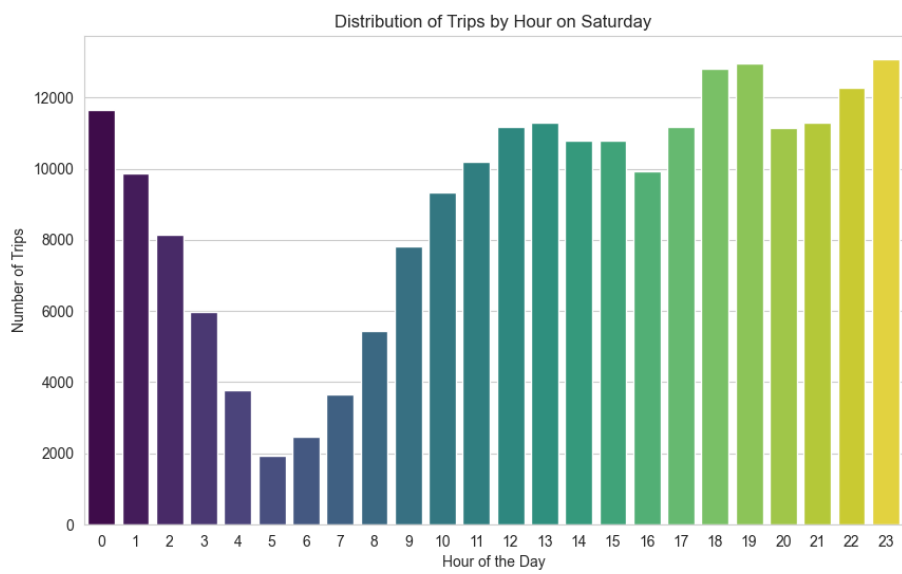

Figure 25: Trip Distribution by Hour for Saturday

From Figure 25, the most popular hour on Thursday is: 23.

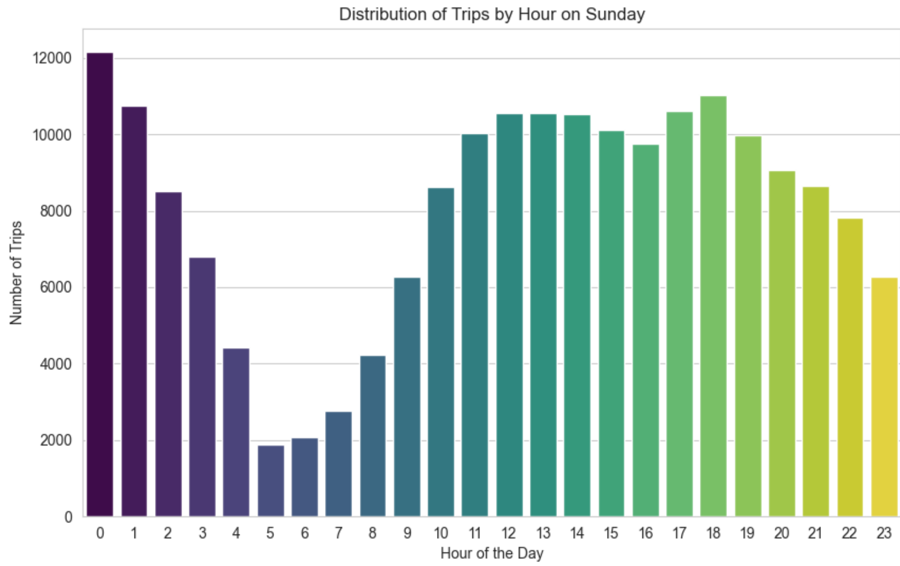

Figure 26: Trip Distribution by Hour for Sunday

From Figure 26, the most popular hour on Sunday is: 0.

3.3 Investigating the Differences between Weekdays and Weekends

Comparison of Trip Volumes: A bar plot was created to compare the total number of trips on weekdays versus weekends. This provides an initial overview of how trip volumes differ between these two categories. Figure 27.

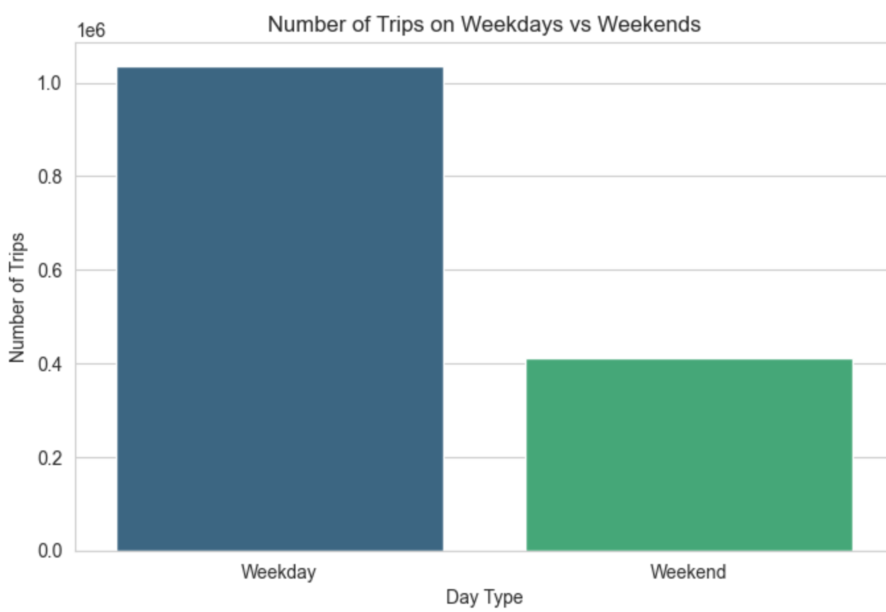

Figure 27: Trip Distribution for Weekdays vs. Weekends

Average Trip Duration: Line plots were created to compare the average trip duration by hour of the day for weekdays and weekends. This helps in identifying any differences in trip lengths between these two categories. Figure 28.

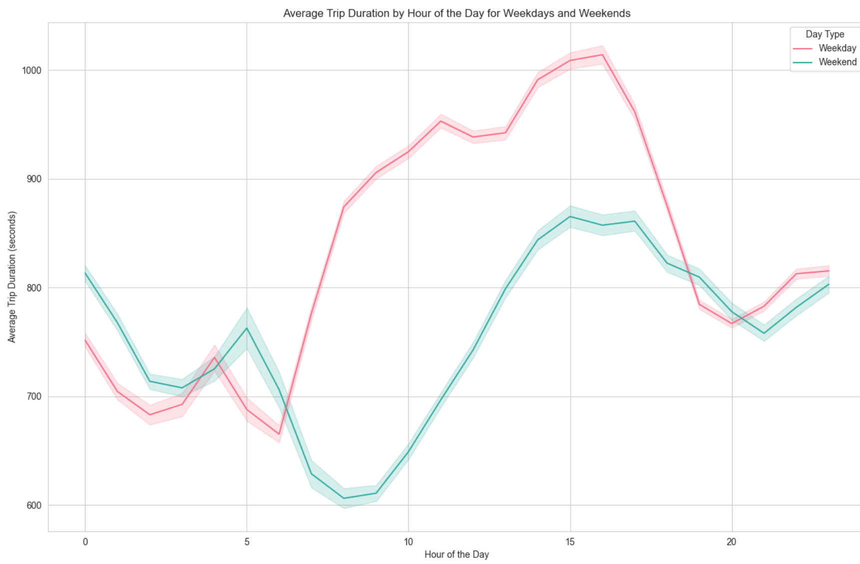

Figure 28: Average Trip Duration for Weekdays vs. Weekends

The bar plot visualises the number of trips on weekdays compared to weekends. The data shows that significantly more trips occur on weekdays than on weekends. There are several factors that could account for this difference:

1. **Work and School Commutes:** Weekdays typically have higher trip volumes due to regular work and school commutes. People travel to and from work, school, and other routine activities, contributing to the increased number of trips. The structured nature of weekdays means that more people are out and about during set times, driving up the number of trips.
2. **Business Activity:** Many businesses operate primarily during weekdays, which results in more trips for meetings, deliveries, and other business-related activities. This business activity contributes to higher trip counts on weekdays compared to weekends, when business operations are reduced.
3. **Leisure vs. Routine Travel:** Weekends are generally reserved for leisure activities, which may not require as many short or frequent trips as weekdays. People may prefer to stay home or engage in fewer activities, leading to a decrease in overall transportation demand. This contrasts with weekdays where routine activities drive consistent trip demands.
4. **Public Transportation and Ride-Hailing Services:** The availability and demand for public transportation and ride-hailing services might differ between weekdays and weekends. On weekdays, these services are often used more for commuting purposes, whereas on weekends, people might use alternative transportation options or have different travel patterns, impacting the number of trips.

3.4 Analysis in Trip Patterns for Major Holidays

Objective: Analyse how trip patterns change on major holidays compared to regular days. The holidays of interest are St. Patrick's Day, Easter, Memorial Day, Valentine's Day, and Martin Luther King Day in 2016. Specifically, we want to determine if there are significant differences in the hourly distribution of trips and cumulative trip patterns on these holidays.

1. Identify Major Holidays in the Dataset:
   - Define the dates of the major holidays in 2016 and convert these dates to date-time objects.
   - Process the dataset to include a new column, 'is_holiday', which indicates whether a trip occurred on a major holiday.
2. Overall Comparison of Trip Counts:
   - Create a bar plot to compare the total number of trips on holidays versus non-holidays.
3. Hourly Distribution of Trips on Each Holiday:
   - For each holiday, filter the dataset to include only trips that occurred on that specific holiday.
   - Calculate the date exactly one week before the holiday and ensure it is not another major holiday.
   - Filter the dataset to include trips that occurred one week before the holiday.
   - Create a combined count plot to compare the hourly distribution of trips on the holiday versus one week before.
4. Cumulative Distribution Function (CDF) Analysis:
   - Plot the Cumulative Distribution Function (CDF) for trips on each holiday and compare it with the CDF for trips on a regular day one week before.
5. Average Trip Duration Analysis:
   - Create a line plot for average trip duration by hour of the day for the holiday and one week before.
6. Holiday vs. Non-Holiday Cumulative Distribution:
   - Plot the Cumulative Distribution Function (CDF) for trips on holidays versus non-holidays.
7. Average Trip Duration on Holidays vs Non-Holidays:
   - Create a line plot for average trip duration by hour of the day for holidays and non-holidays.

This multi-faceted approach provides a comprehensive analysis of how trip patterns change on major holidays. By comparing hourly distributions and cumulative distributions, we gain valuable insights into whether holidays significantly affect travel behaviour in New York City. These insights could be useful for understanding traffic management, service demand, and urban planning during holidays.

St. Patrick’s Day vs. Regular Day One Week Before

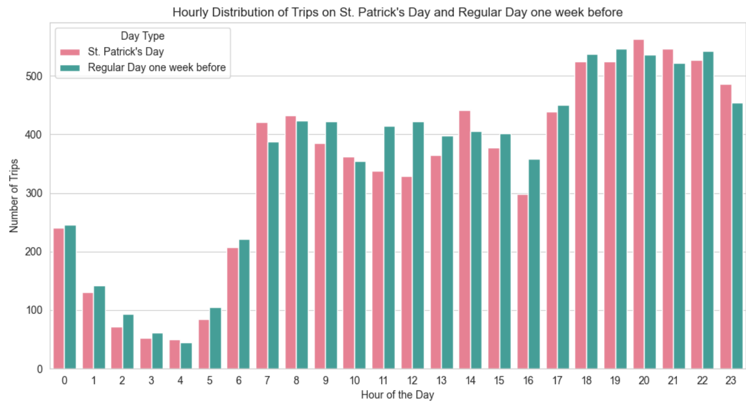

Figure 29: Count plot showing Trip Distribution for St. Patrick’s Day vs. Regular Day One Week Before

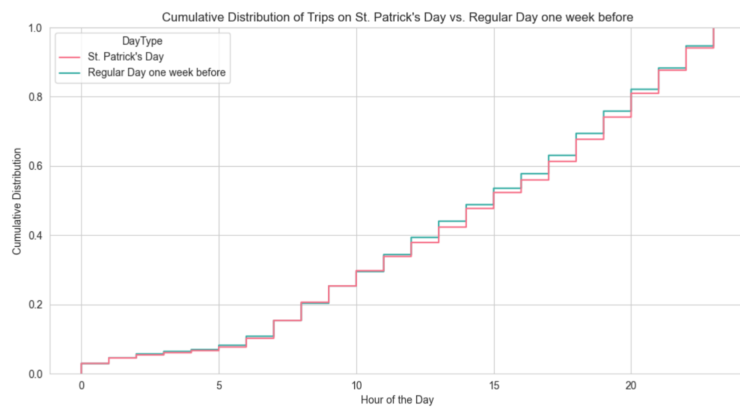

Figure 30: Cumulative Distribution showing Trip Distribution for St. Patrick’s Day vs. Regular Day One Week Before

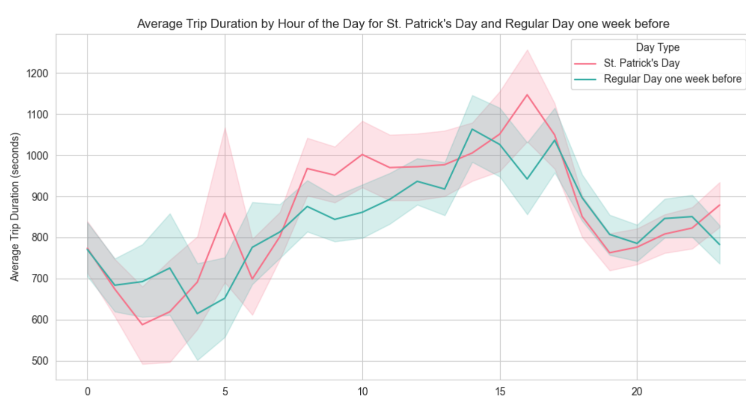

Figure 31: Average Trip Duration for St. Patrick’s Day vs. Regular Day One Week Before

Easter vs. Regular Day One Week Before

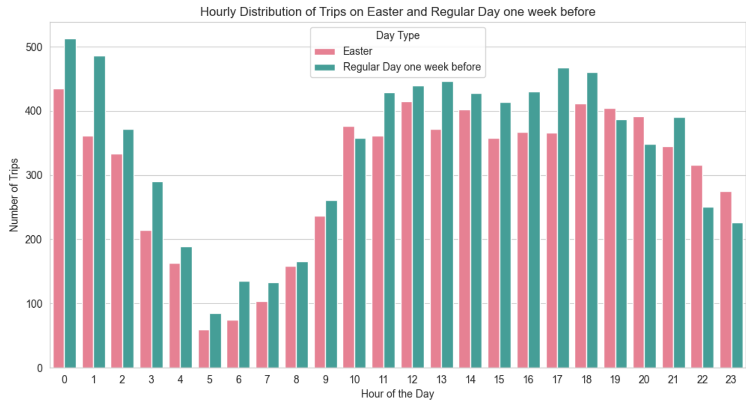

Figure 32: Count plot showing Trip Distribution for Easter vs. Regular Day One Week Before

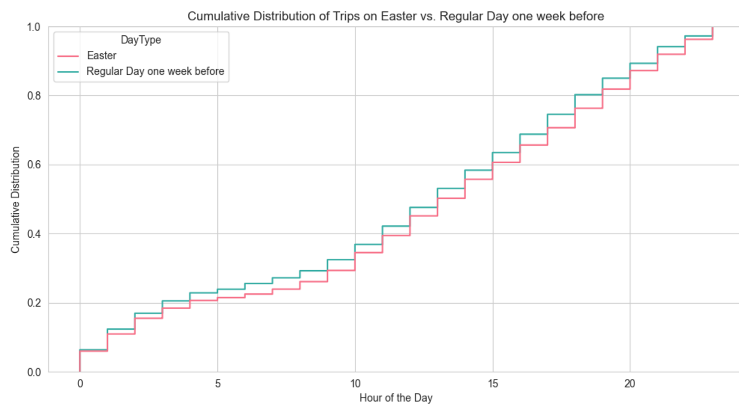

Figure 33: Cumulative Distribution showing Trip Distribution for Easter vs. Regular Day One Week Before

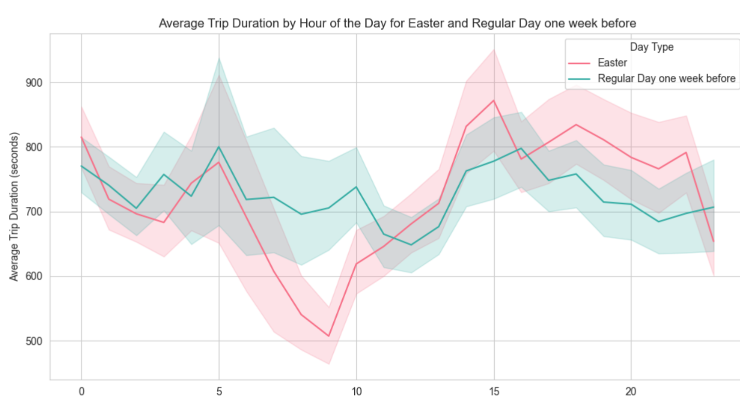

Figure 34: Average Trip Duration for Easter vs. Regular Day One Week Before

Memorial Day vs. Regular Day One Week Before

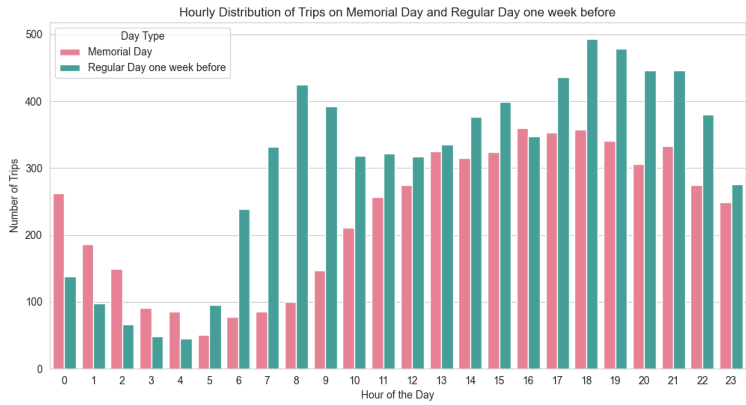

Figure 35: Count plot showing Trip Distribution for Memorial Day vs. Regular Day One Week Before

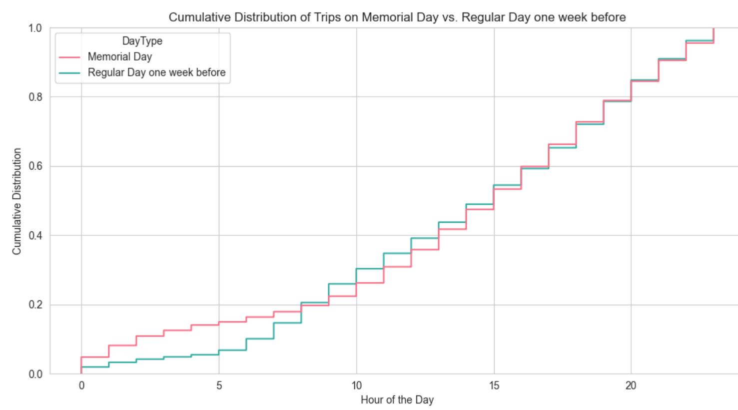

Figure 36: Cumulative Distribution showing Trip Distribution for Memorial Day vs. Regular Day One Week Before

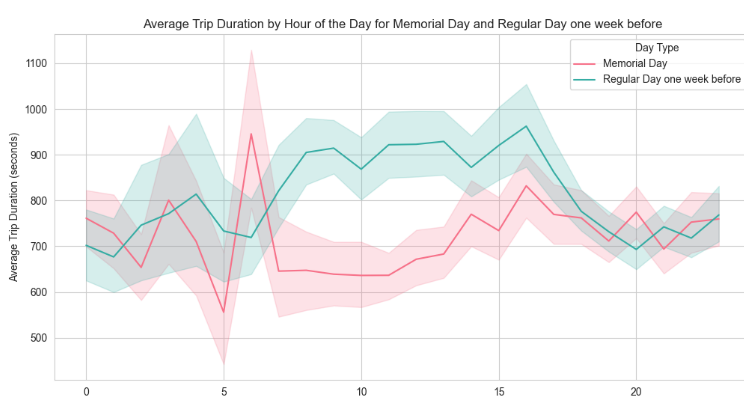

Figure 37: Average Trip Duration for Memorial Day vs. Regular Day One Week Before

Valentine's Day vs. Regular Day One Week Before

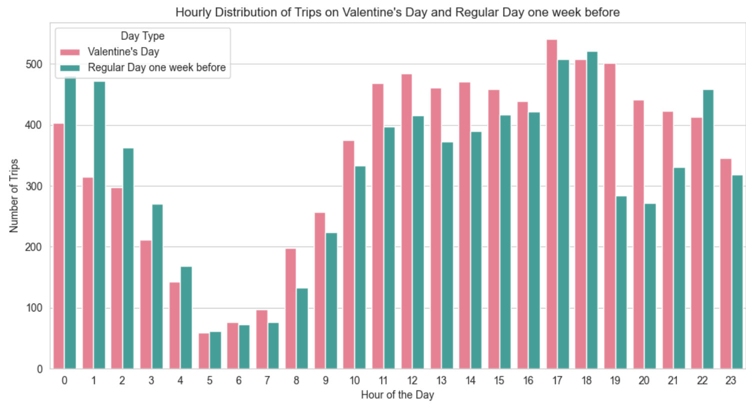

Figure 38: Count plot showing Trip Distribution for Valentine's Day vs. Regular Day One Week Before

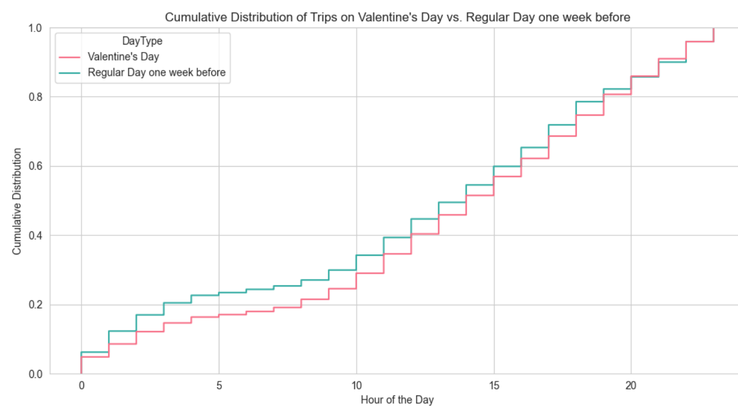

Figure 39: Cumulative Distribution showing Trip Distribution for Valentine's Day vs. Regular Day One Week Before

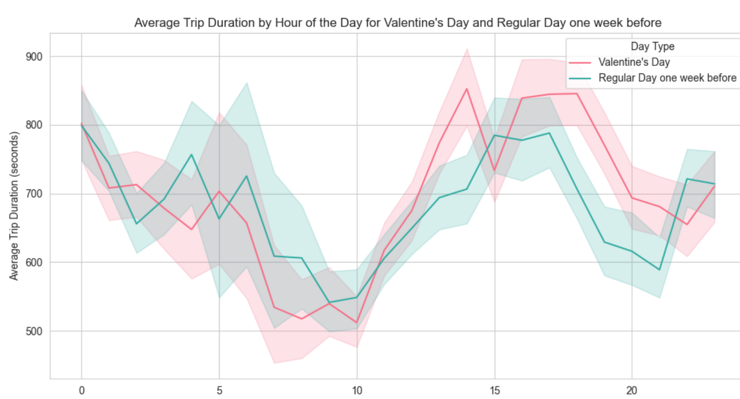

Figure 40: Average Trip Duration for Valentine's Day vs. Regular Day One Week Before

Martin Luther King Day vs. Regular Day One Week Before

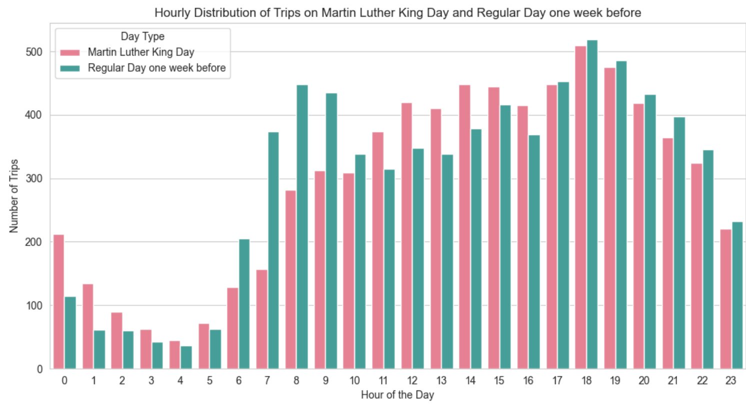

Figure 41: Count plot showing Trip Distribution for Martin Luther King Day vs. Regular Day One Week Before

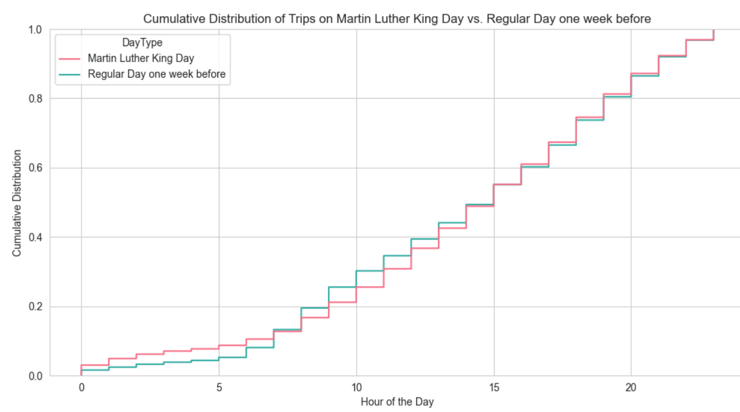

Figure 42: Cumulative Distribution showing Trip Distribution for Martin Luther King Day vs. Regular Day One Week Before

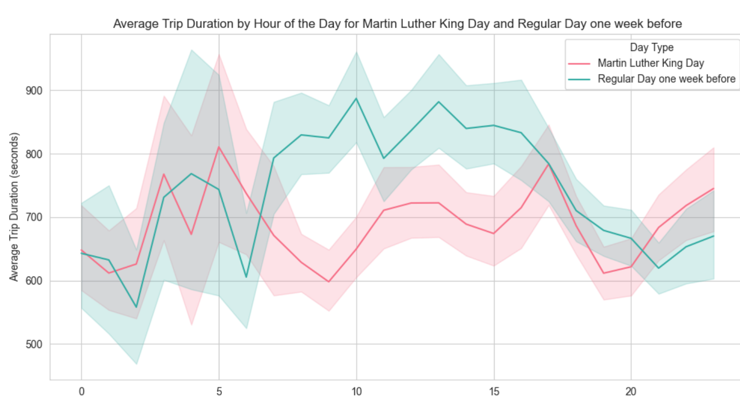

Figure 43: Average Trip Duration for Martin Luther King Day vs. Regular Day One Week Before

Holidays vs. Non-Holidays

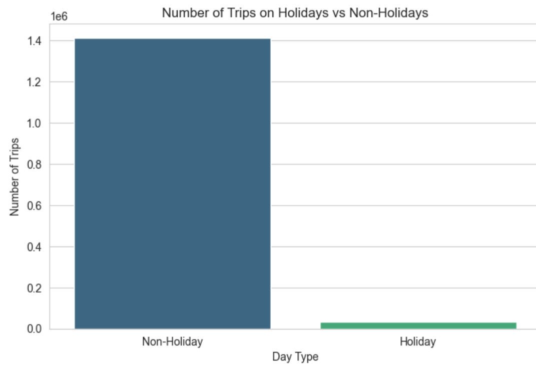

Figure 44: Trip Distribution for Holidays vs. Non-Holidays

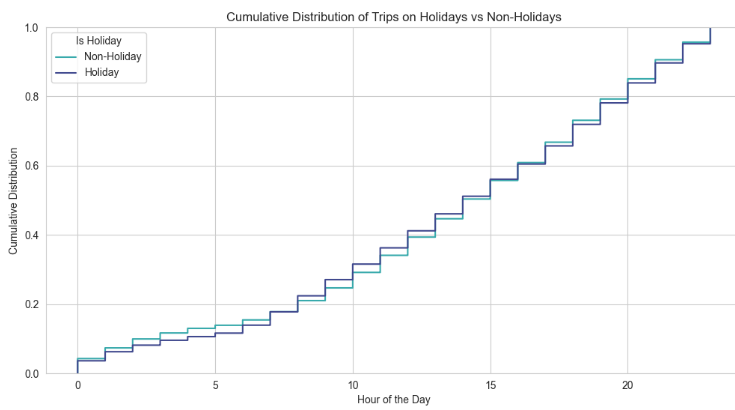

Figure 45: Cumulative Distribution showing Holidays vs. Non-Holidays

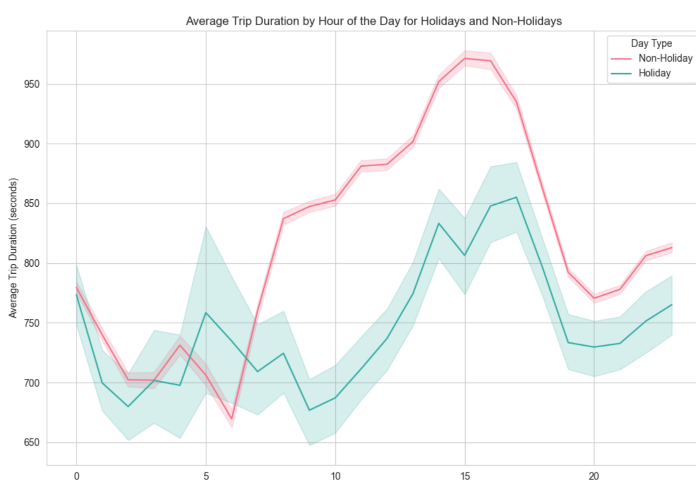

Figure 46: Average Trip Duration for Holidays vs. Non-Holidays

Analysing the Differences Between Holiday and Non-Holiday Trip Patterns

Based on the CDF plots above, we can observe several key differences in trip patterns between holidays and non-holidays:

1. Delayed Starts: The CDF for holidays consistently starts lower than the CDF for non-holidays, indicating fewer trips in the early hours on holidays. This suggests a delayed start to the day compared to regular days.
2. Extended Evening Hours: The CDF for holidays consistently ends higher than the CDF for non-holidays, indicating a higher proportion of trips occurring later in the evening on holidays. This suggests extended activities or a different pattern of evening travel
3. Mid-Day Similarity: While the early and late hours show significant differences, the lines for holidays and non-holidays tend to converge around the mid-day and afternoon hours. This suggests that trip patterns during these times are relatively similar, with less variation between holidays and non-holidays.

Overall, the analysis indicates that while the general pattern of trip accumulation over the day is similar for holidays and non-holidays, there are notable differences in the timing of trips. Holidays are characterised by delayed starts and extended evening hours, suggesting that social and leisure activities play a significant role in shaping travel behaviour on these days.

3.5 Average Trip Speed Throughout the Day

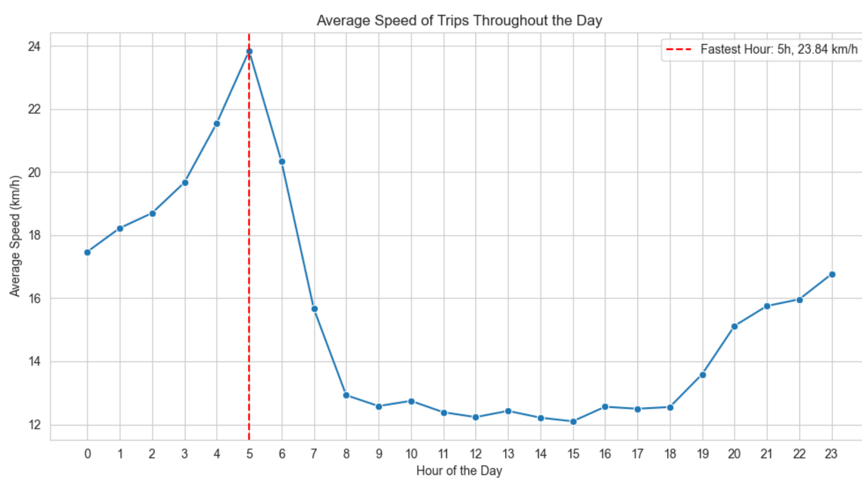

Figure 47: Average Trip Speed Throughout the Day

From figure 47 The fastest time of day is around 5:00 with an average speed of 23.84 km/h.

4 Location Clusters

4.1 Heatmaps

1. Weekday vs Weekend

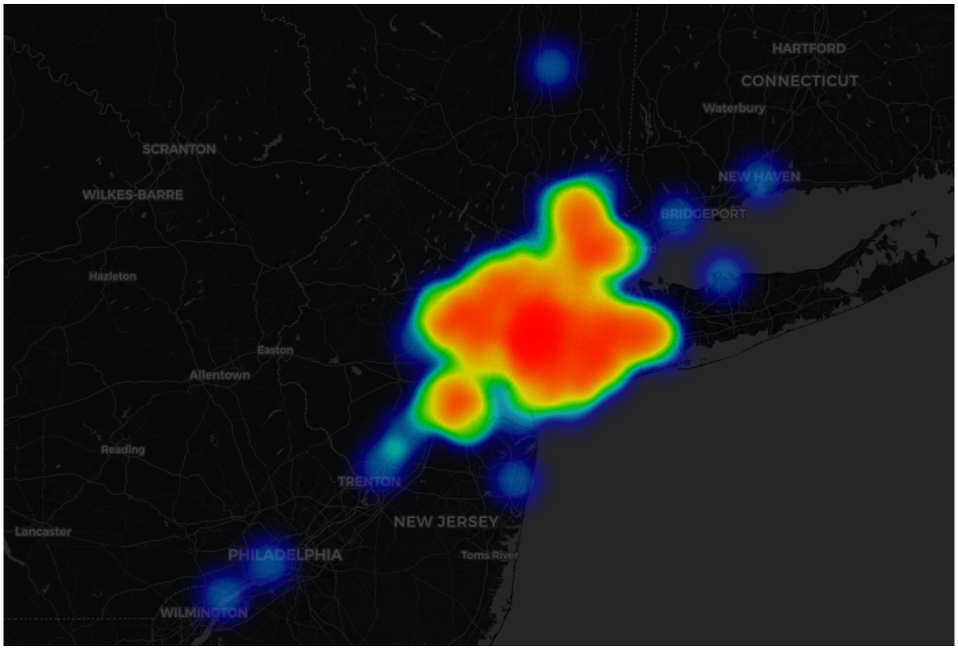

Figure 48: Heatmap of pickups during Weekdays

View this map in more detail by downloading here Weekday Pickups Heatmap.

Figure 49: Heatmap of pickups during Weekend

View this map in more detail by downloading here Weekend Pickups Heatmap.

During weekdays, the heatmap 48 reveals a much broader area of activity, extending out to places like Paterson and West Babylon, likely reflecting the commuter patterns of people travelling to work. Central New York is particularly busy, showcasing the city's weekday hustle. On weekends the activity contracts on the heatmap 49, with significantly fewer pickups in areas like Paterson and West Babylon, suggesting these spots are less frequented outside of workdays. The busiest areas on the weekend are concentrated around Central New York, though some places, like Stamford, experience a slight increase in activity.

2. Morning vs Evening

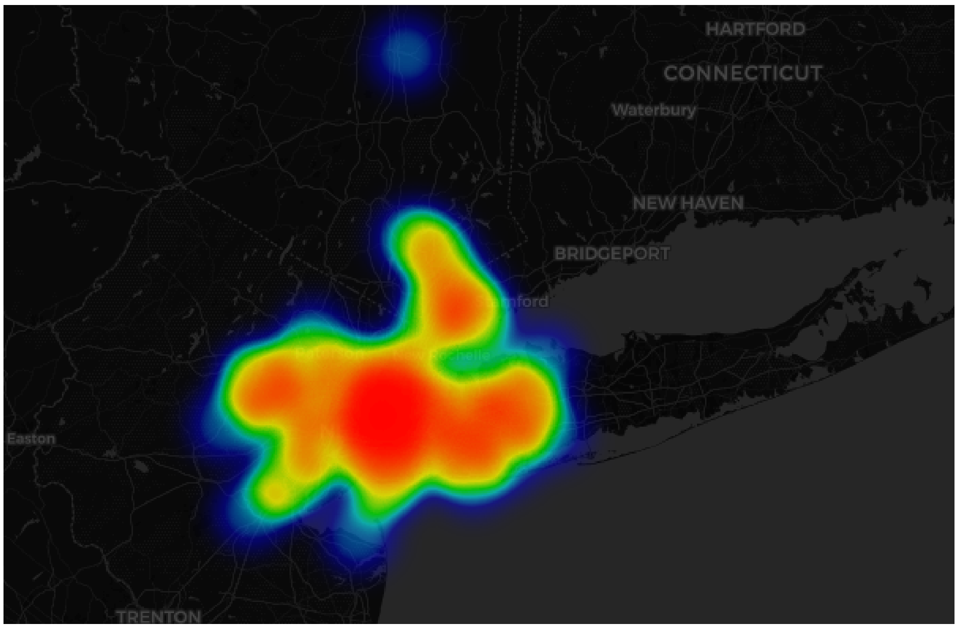

Figure 50: Heatmap of pickups during Mornings

View this map in more detail by downloading here Morning Pickups Heatmap.

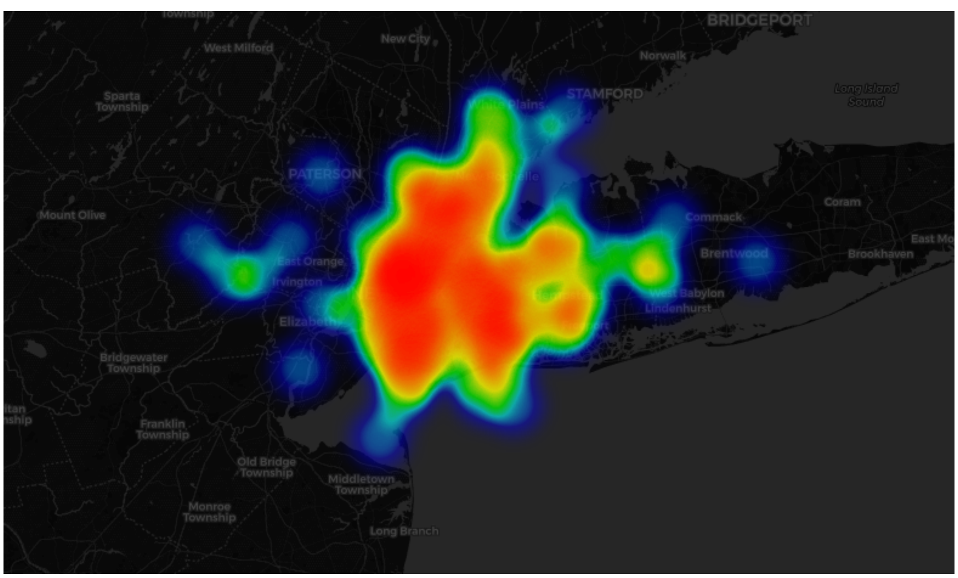

Figure 51: Heatmap of pickups during Evenings

View this map in more detail by downloading here Evening Pickups Heatmap.

In the morning (heatmap 50), Central New York is busy with activity, with pickups occurring over a slightly larger area compared to the evening, probably from people going to work. While the evening (heatmap 51) remains busy, areas like Stamford and Paterson see significantly fewer pickups, suggesting these locations may not be as popular for evening outings. One constant, however, is the area around Central Park, which remains a hotspot for pickups throughout both times of the day.

4.2 Hotspots

Number of clusters: 59

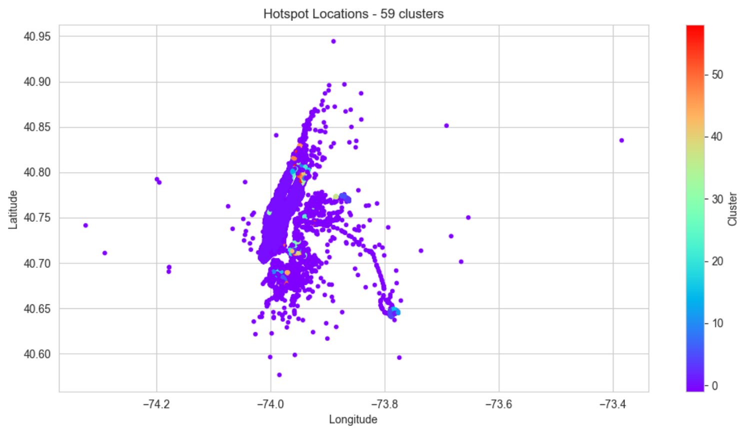

Figure 52: Hotspot locations on plot

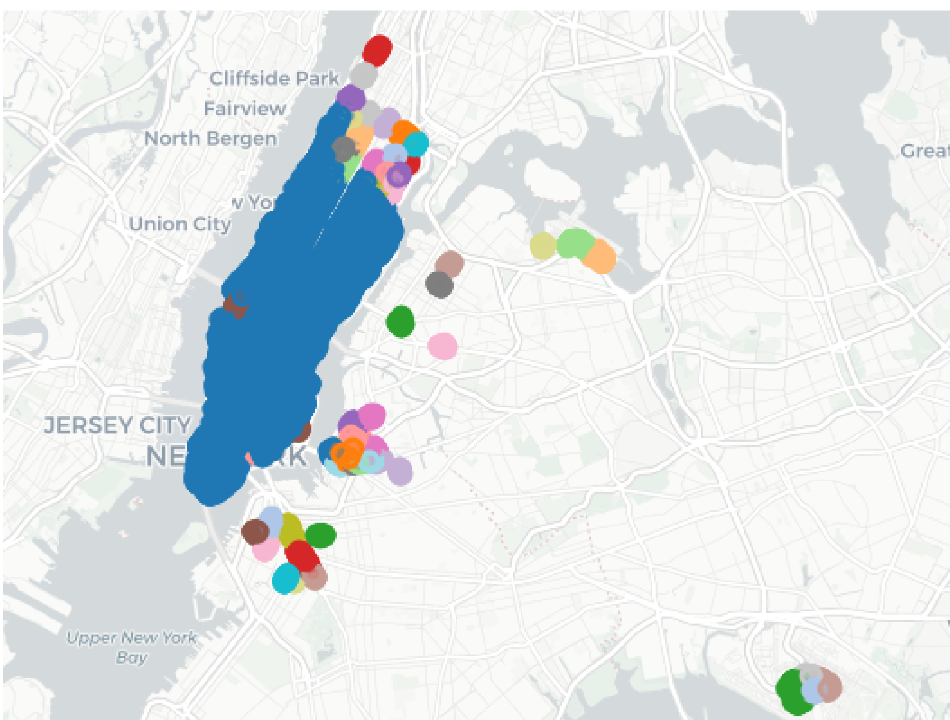

Figure 53: Hotspot locations on a map

View this map in more detail by downloading here Cluster Map.

5 Airports

JFK:

- On average it takes 49.48 minutes to get to JFK airport from the Empire State Building.
- It takes the longest to get to JFK from the Empire State building in the afternoon and the shortest at night.

Newark:

- On average it takes 36.33 minutes to get to Newark airport from the Empire State Building.
- It takes the longest to get to Newark from the Empire State Building in the evening and the shortest at night.

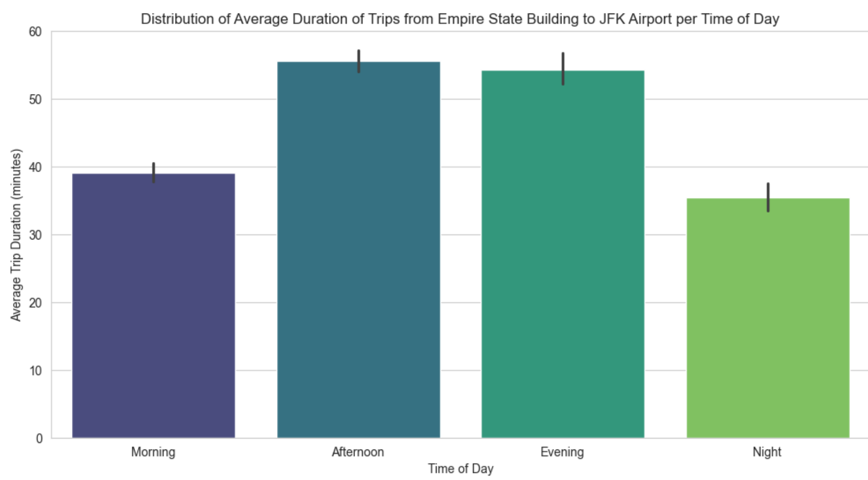

Figure 54: Distribution of Average Duration of Trips from Empire State Building to JFK Airport per Time of Day

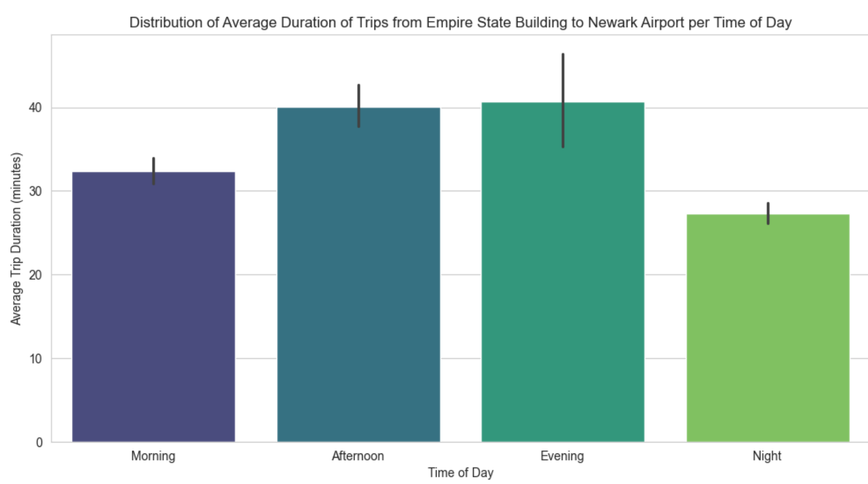

Figure 55: Distribution of Average Duration of Trips from Empire State Building to Newark Airport per Time of Day

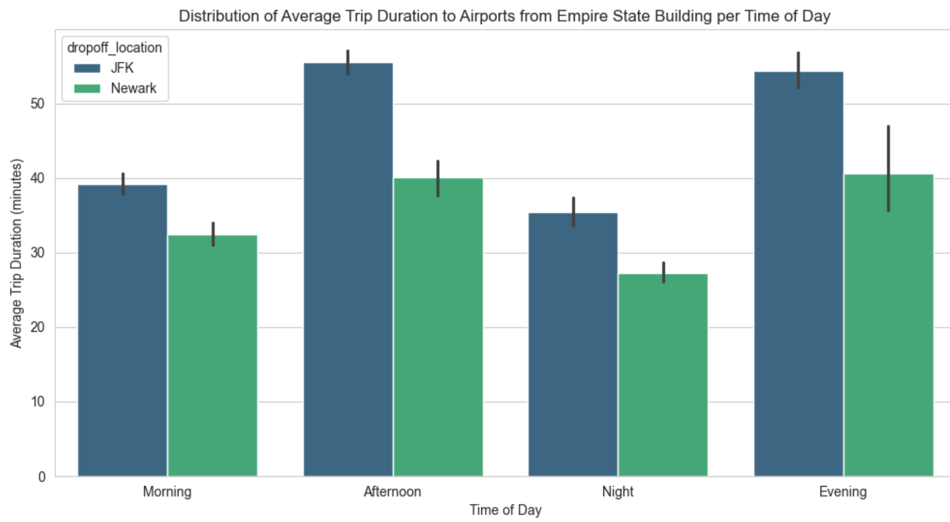

Figure 56: Distribution of Average Trip Duration to Airports from Empire State Building per Time of Day

On average it takes longer to get to JFK airport than it takes to get to Newark airport from the Empire State building. At any time of day it takes longer to get to JFK airport than it does to get to Newark Airport. The general trends of the trip duration's at the different times of days is the same for both airports, with the only slight difference being that the longest duration trip is in the afternoon for JFK and in the evening for Newark. The second longest trip duration's are in the evening and afternoon for JFK airport and Newark airport respectively. The shortest for both is at night and the second shorted is in the morning for both as well.

The haversine function was used to fine the distances between the pickup and drop-off locations and the Empire State Building and airports respectively. These distances were checked if they were smaller than the defined radius to determine if the pickup and drop-off points were located at the Empire State building and the airports. The pandas data frame was modified to only contain the trips from the Empire State Building to the different airports. The average of the different trip duration's was calculated for each time of day.

Radius motivation:
We decided to have 3 different radius values for the 3 different locations. The coordinates of the different locations is the coordinates of the centre of the location. Therefore, it would be unfair to use the same radius for the Empire State building and for JFK airport as these 2 locations are vastly different in size.

The radius values for the 3 different locations was determined by looking at the area that the location takes up and the context of the location.

The radius value for the Empire state building was decided to be 500m. This is because the Empire State Building does not take up that much area on the ground. The Empire State Building is also in the dense city, meaning having a radius that is too big will result in incorrectly classifying pickup locations that are not specifically at the Empire State building to be at the Empire State Building. We did try different values for this radius, for example we tries 100 meters but found that this would result in there being too little trips included in order to accurately assess the travel duration's for the

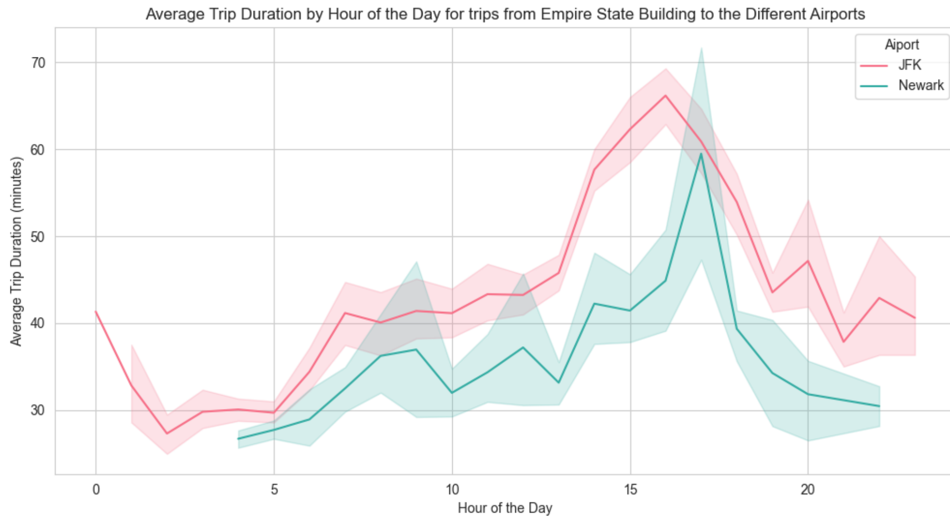

Figure 57: Average Trip Duration by Hour of the Day for trips from Empire State Building to the Different Airports

trips to the airport.

The radius values for the airports was determined by googling the surface area of the different airports. we used this surface area as the approximate circular area of the airports. A calculation was done to determine the radius of a circle with this area.

- The area of JFK airport is $21.04\,\mathrm{km}^2$. The radius of a circle that has the area of $21.04\,\mathrm{km}^2$ is 2.588km.
- The area of Newark airport is $8.20\,\mathrm{km}^2$. The radius of a circle that has the area of $8.20\,\mathrm{km}^2$ is 1.6156 km.

Therefore we decided to make the radius of the JFK airport 3.5 km and the radius of Newark airport 2.5 km. This is about 1km greater than the approximate circular radius, and this is to account for ubers/taxis not dropping off customers directly at the airport perimeter. This increase of the radius is to help increase the amount of data points available in order to accurately calculate the average duration of the trips, as explained above with the Empire State Building radius. The radius for the airports can also be increased more as the airports will be in a more open area of New York, meaning that drop-off points will be less likely to be misclassified as the airports.

6 Boroughs

6.1 Neighbourhoods for the trip start and end locations

Done in code

6.2 Chloropeth of all pickups

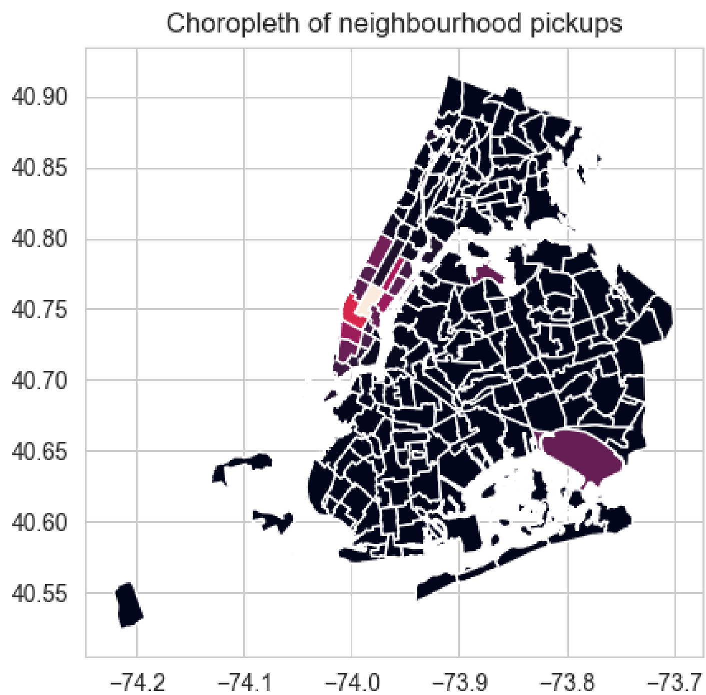

Figure 58: Choropleth of neighbourhood pickups

The chloropleths 58 and 59 are unevenly distributed as there are there are only a few neighbourhoods that have a notable number of pickups or drop-offs. Most of the neighbourhoods have around the same number of very few pickups and drop-offs. The neighbourhoods and general area of New York that have the most pickups also have the most drop-offs. The same can be said about the areas with the least amount of drop-offs and pickups. Therefore the trend is similar for the pickups and drop-offs.

There are some areas that do not that pickups at all but they do have drop-offs. The drop-offs chloropleth distribution covers a wider range of the different areas of New York compared to the pickups chloropleth.

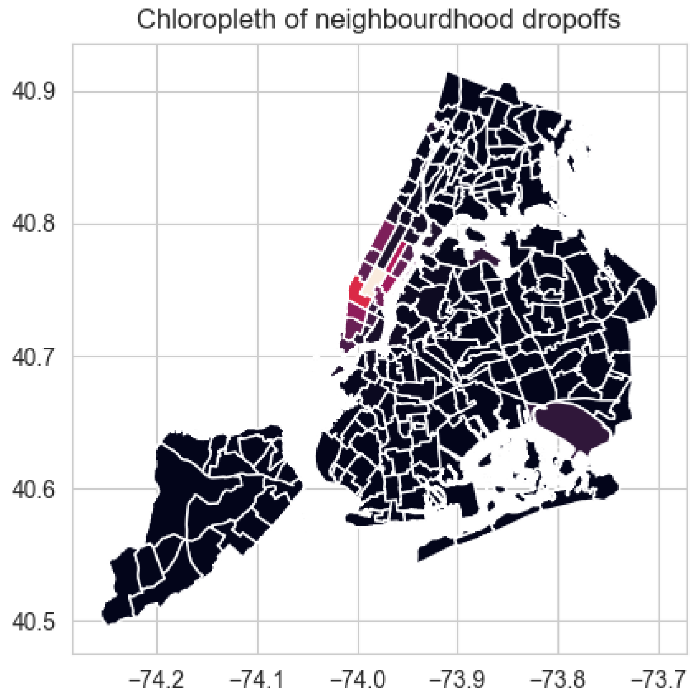

Figure 59: Average Trip Speed Throughout the Day

6.3 Quietest neighbourhoods between midnight and 5AM

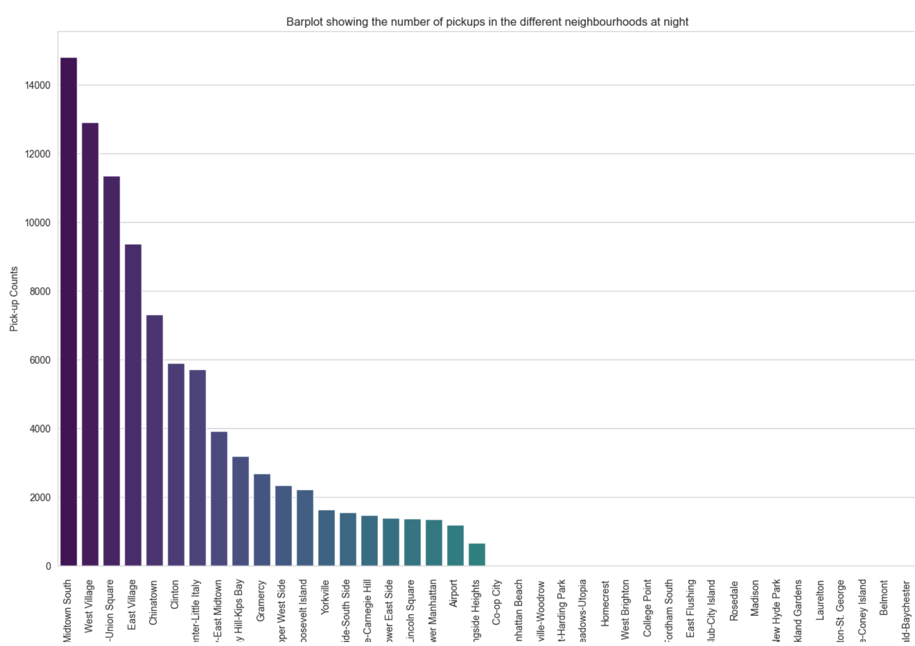

Figure 60: Barplot showing the number of pickups in the different neighbourhoods at night

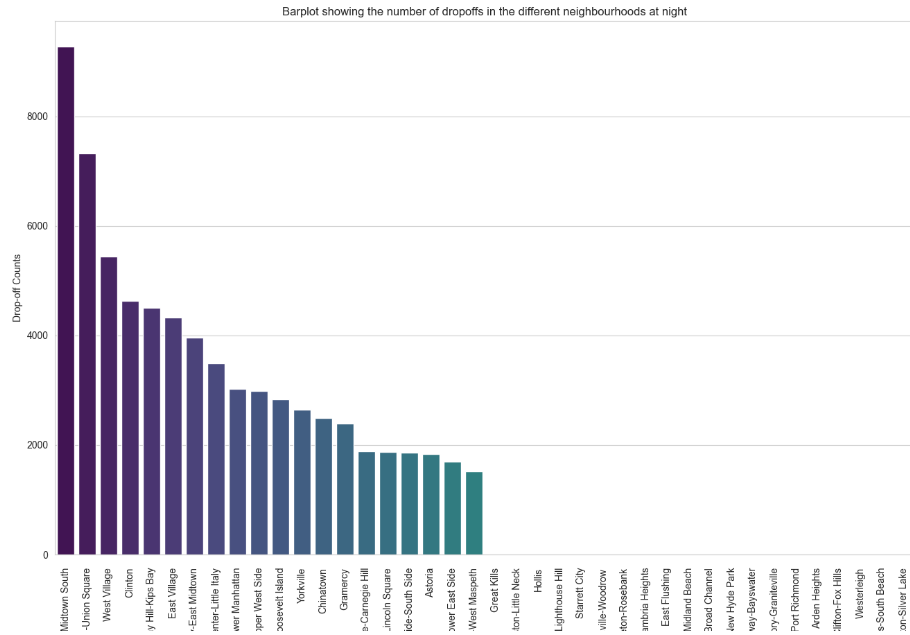

Figure 61: Barplot showing the number of drop-offs in the different neighbourhoods at night

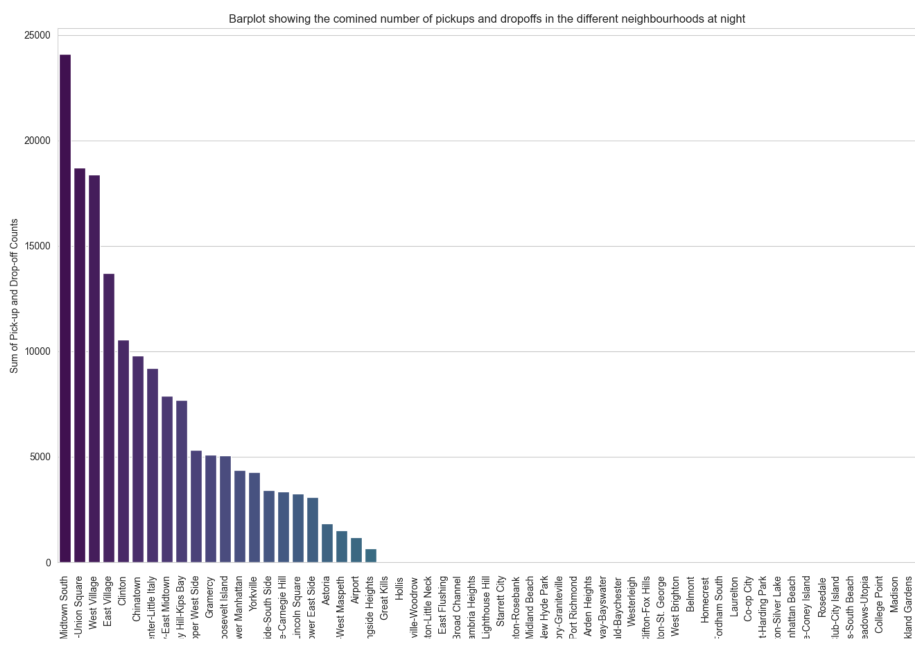

Figure 62: Barplot showing the combined number of pickups and drop-offs in the different neighbourhoods at night

See Figures 60, 61 and 62

- Oakland Gardens
- Madison

- College Point
- Fresh Meadows-Utopia
- Old Town-Dongan Hills-South Beach

6.4 Busiest neighbourhoods between midnight and 5AM

See Figures 60, 61 and 62

- Midtown-Midtown South
- Hudson Yards-Chelsea-Flatiton-Union Square
- West Village
- East Village
- Clinton
- Chinatown
- SoHo-TribeCa-Civiv Center-Little Italy