Conclusion/Remarks:

1.1 After loading data rows 3041714 , columns 19

Conclusion/Remarks:

1.1 Sample Data created with 5% data from each hour from 12 months file

Conclusion/Remarks:

2.1 Checking missing value percentage

Conclusion/Remarks:

2.1.1 Fix the index and drop unnecessary columns: Index reset done, but will drop column later after analysis

Conclusion/Remarks:

2.1.2 So as there are 1747917 rows missing in airport_fee and 213357 rows in Airport_fee, so we can combine

Conclusion/Remarks:

2.1.2 For After combining , Airport_fee missing value percentage reduced from 11.25% to 3.42%

Conclusion/Remarks:

2.1.2 Dropped column airport_fee

Conclusion/Remarks:

2.1.3 Checking the distribution of RatecodeID of the negative amount value sets., Mainly it is 1 and 2

Conclusion/Remarks:

2.1.3 Detail checking the distribution of RatecodeID of the negative amount value sets., Mainly it is 1 and 2

Conclusion/Remarks:

2.2.2 Handling missing values in passerger count: Here from the distribution we can see around 75% is for passerger count 1. So taking the mode to impute NaN values here

Conclusion/Remarks:

2.2.3 Handle missing values in RatecodeID: As the mode of RatecodeID is 1 so imputing with it

Conclusion/Remarks:

2.2.4 Impute NaN in congestion_surcharge: As there are 3 discrete values and mode is 2.5, so imputing with it

Conclusion/Remarks:

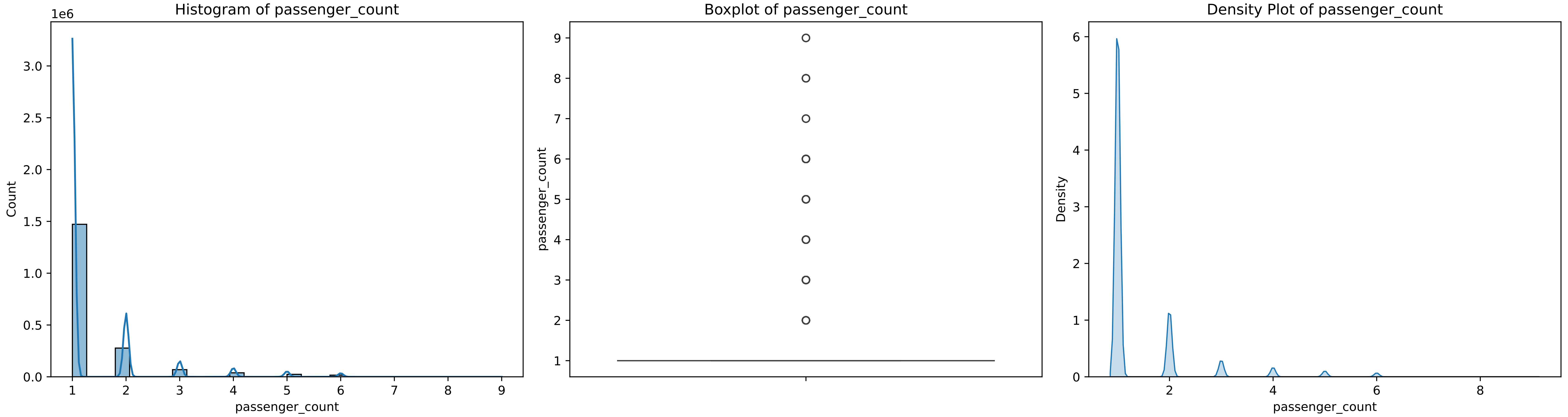
2.2.4 Impute NaN in store_and_fwd_flag: As there are 2 discrete values and mode is N, so imputing with it

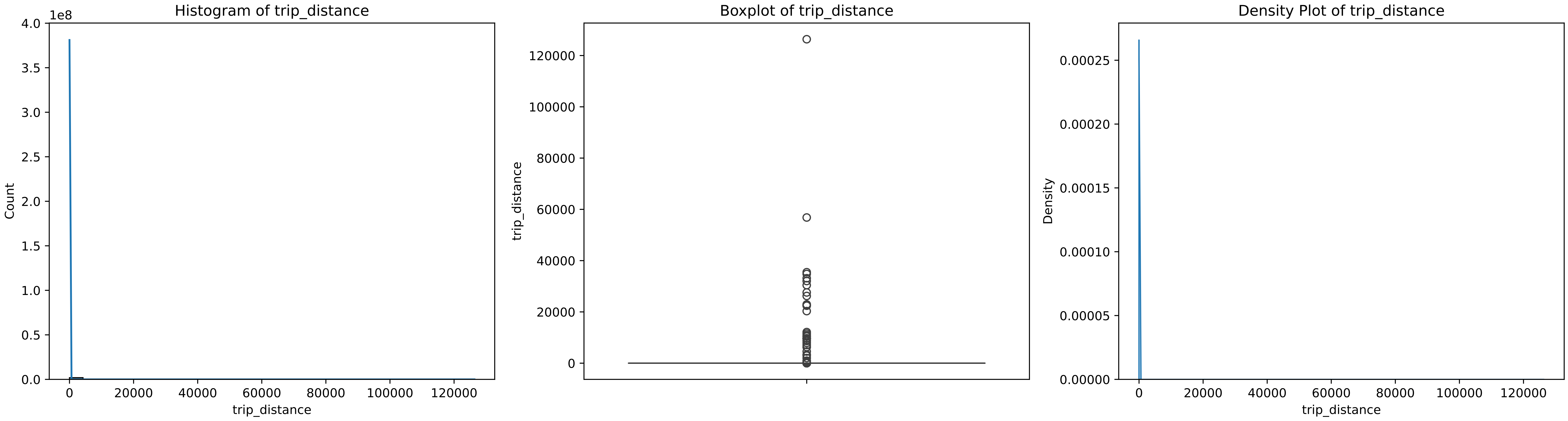
Conclusion/Remarks:

2.2.4 Impute NaN in airport_fe: As there are 4 discrete values and mode is 0, so imputing with it

Conclusion/Remarks:

2.3 Checking the Outlier count, percentage and Skewness

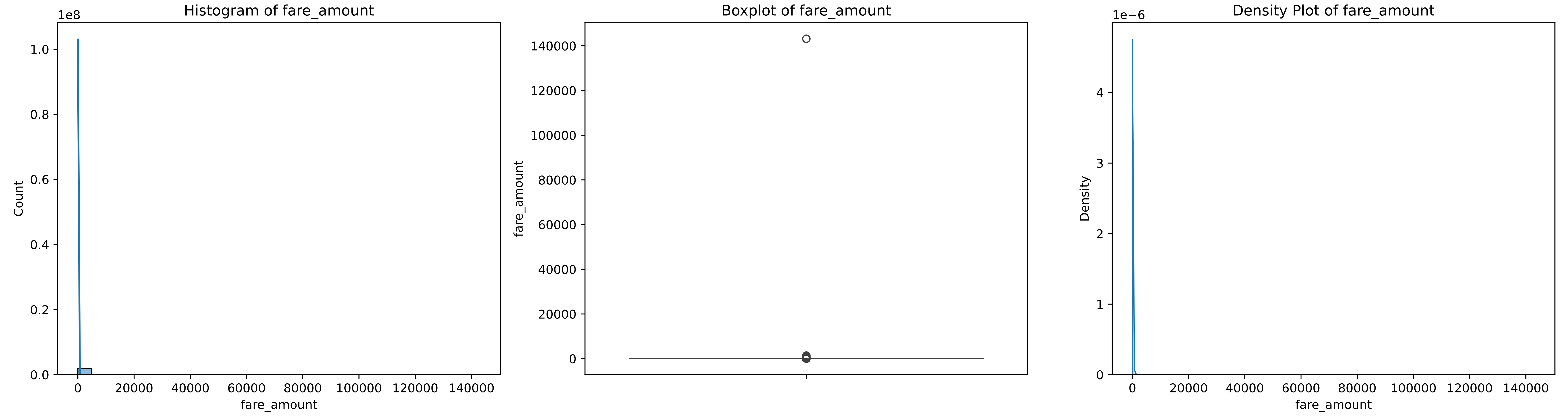

Histogram of passenger_count
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
1
2
3
4
5
6
7
8
9
passenger_count
Boxplot of passenger_count
9
8
7
6
5
4
3
2
1
passenger_count
Density Plot of passenger_count
6
5
4
3
2
1
0
Density
2
4
6
8
passenger_count

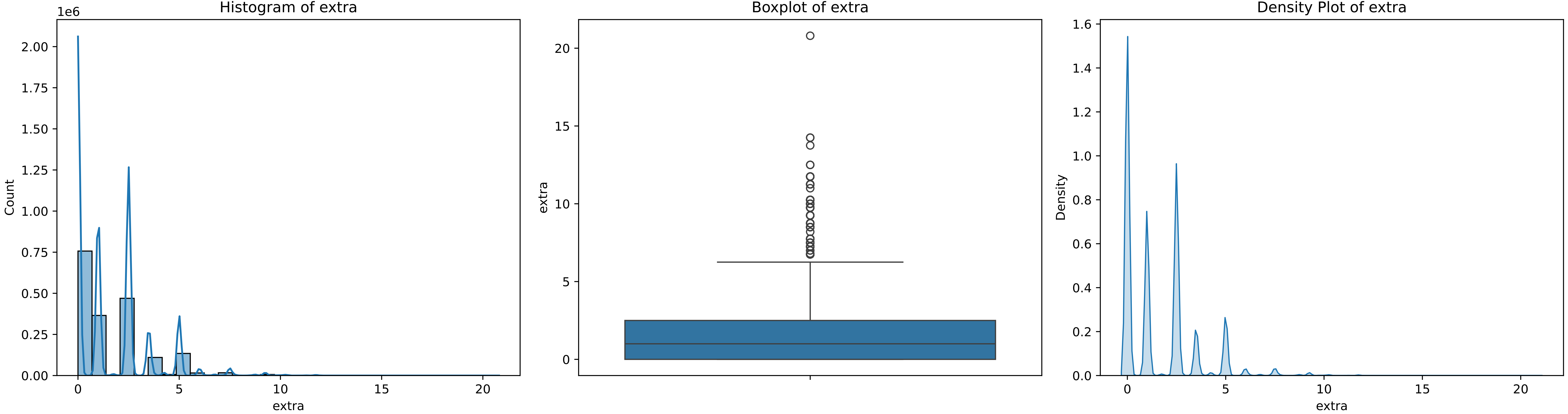


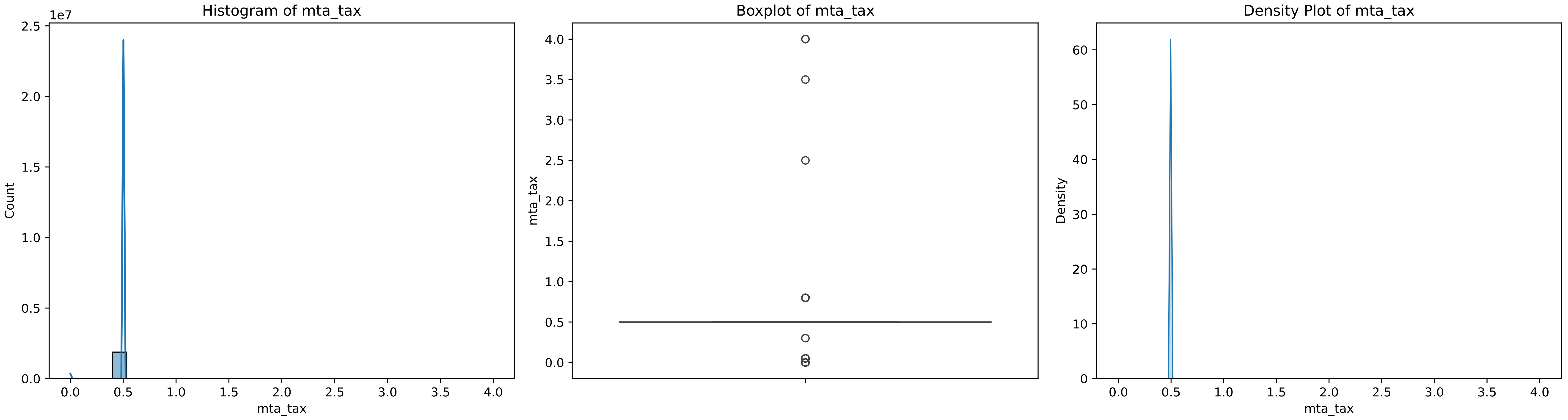
Histogram of trip_distance

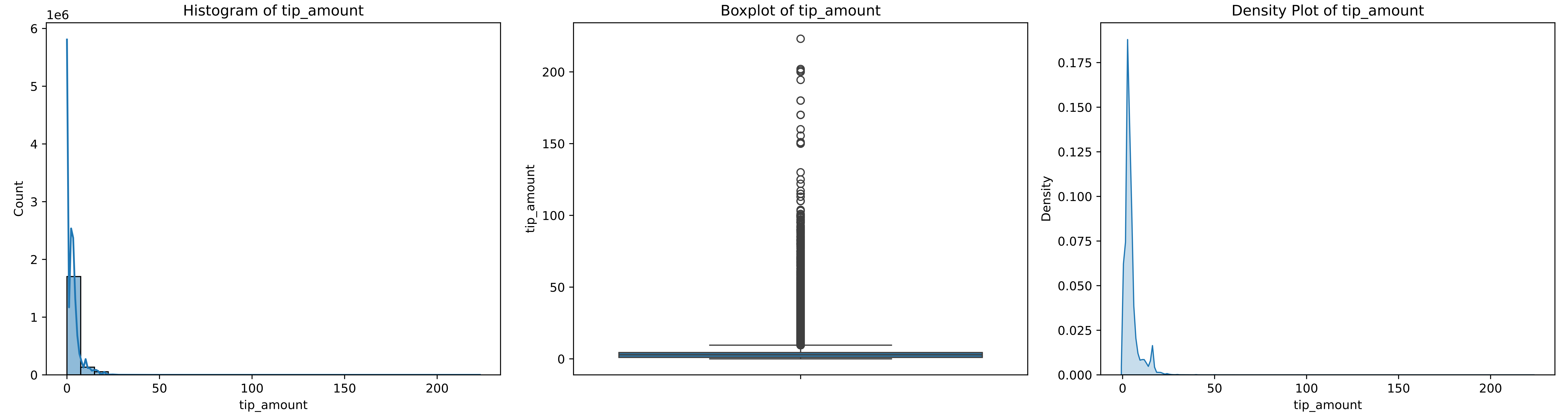
Boxplot of trip_distance

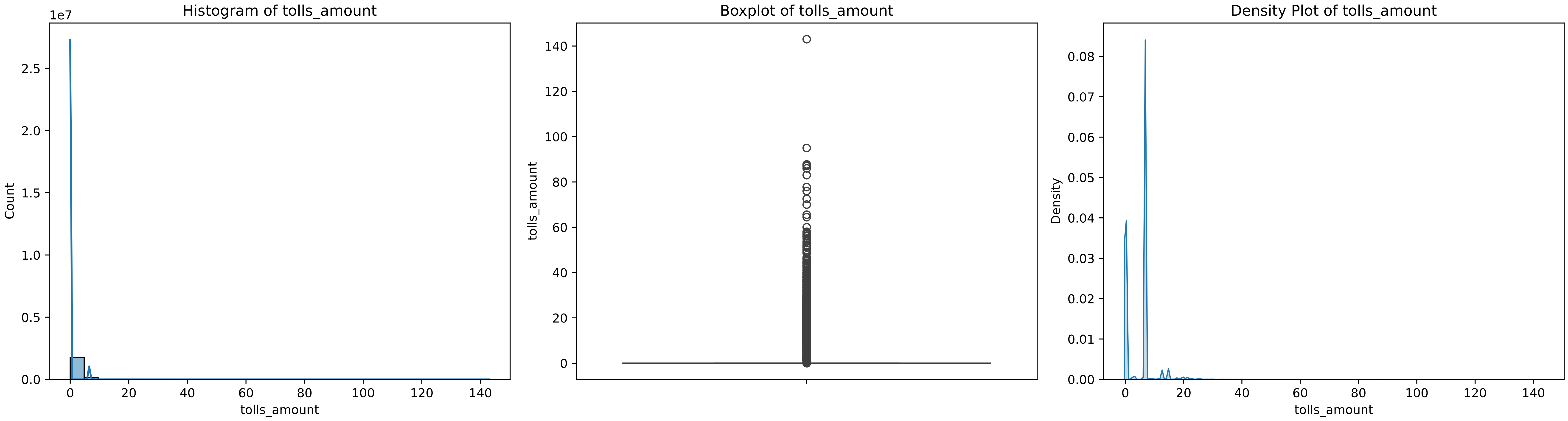
Density Plot of trip_distance


Histogram of fare_amount
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Count
0
20000
40000
60000
80000
100000
120000
140000
fare_amount
Boxplot of fare_amount
140000
120000
100000
80000
60000
40000
20000
0
fare_amount
Density Plot of fare_amount
1e−6
4
3
2
1
0
Density
0
20000
40000
60000
80000
100000
120000
140000
fare_amount

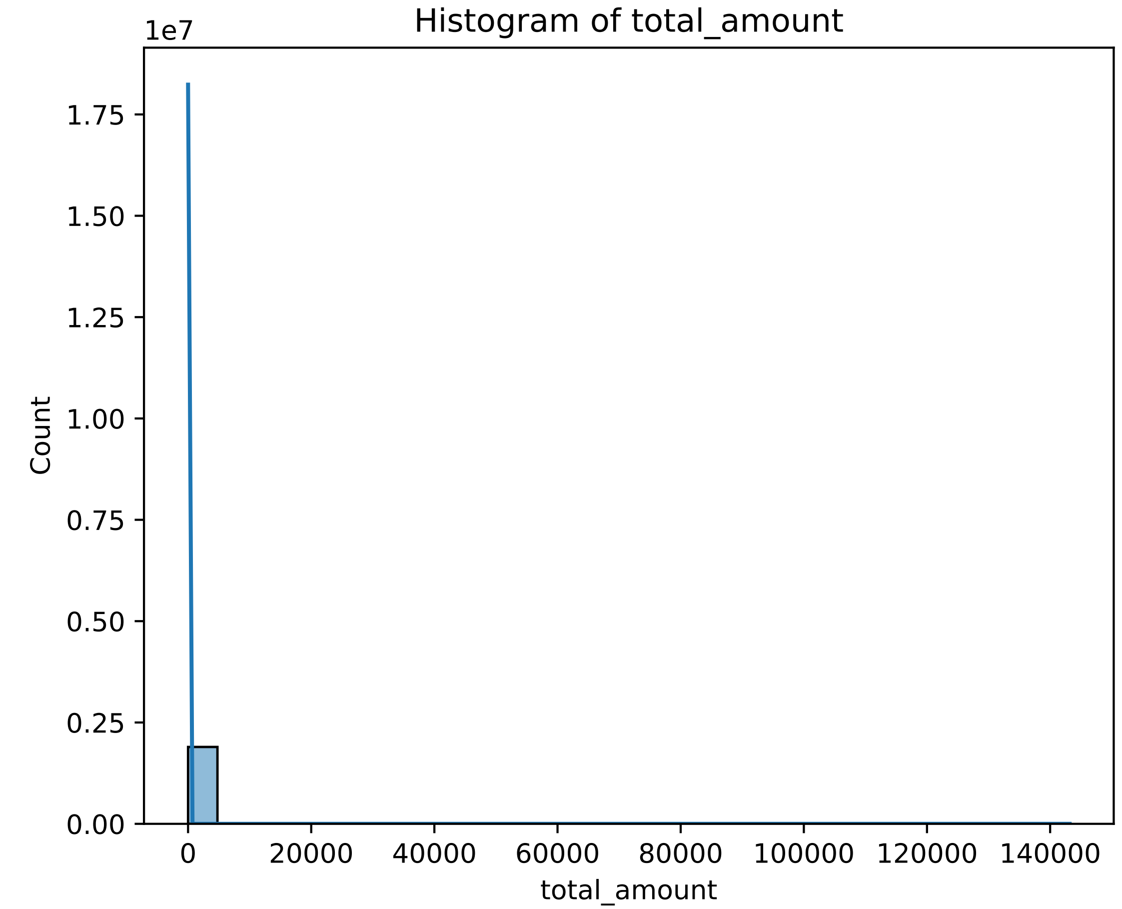

Histogram of extra
1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0
5
10
15
20
extra
Boxplot of extra
20
15
10
5
0
extra
Density Plot of extra
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Density
0
5
10
15
20
extra

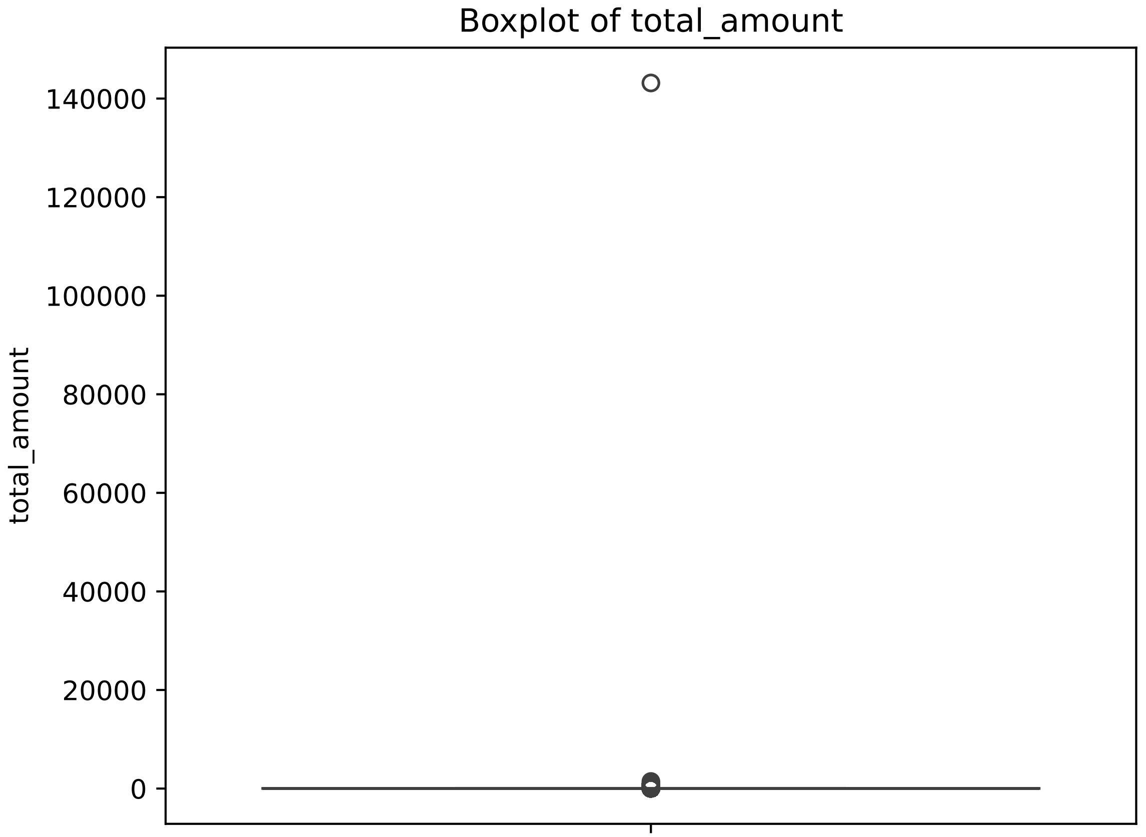

Histogram of mta_tax
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Count
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
mta_tax
Boxplot of mta_tax
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
mta_tax
Density Plot of mta_tax
60
50
40
30
20
10
0
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
mta_tax

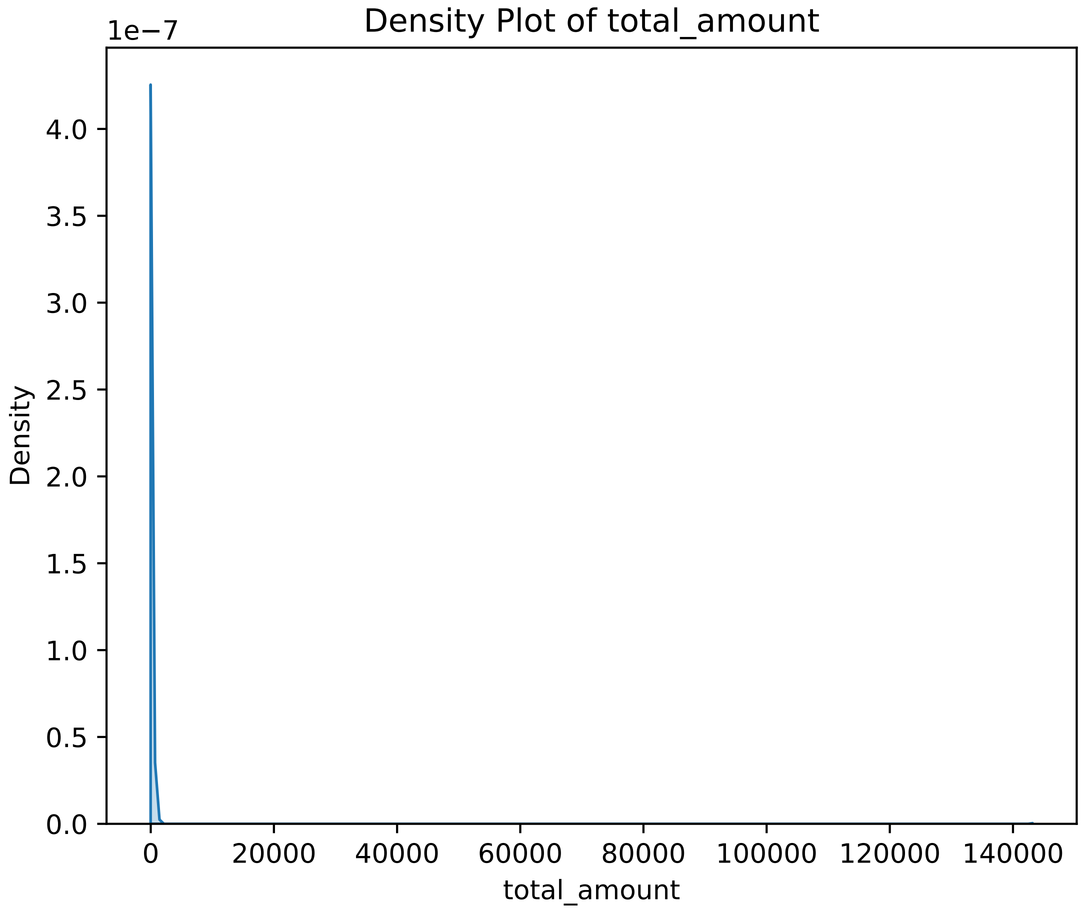

Histogram of tip_amount
1e6
6
5
4
3
2
1
0
Count
0
50
100
150
200
tip_amount
Boxplot of tip_amount
200
150
100
50
0
tip_amount
Density Plot of tip_amount
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Density
0
50
100
150
200
tip_amount

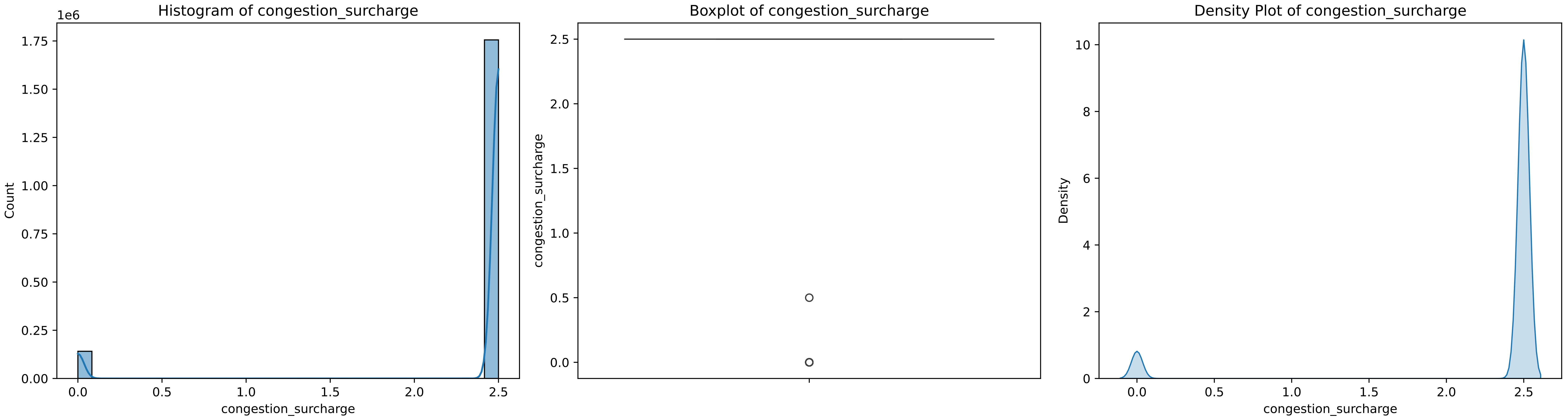

Histogram of tolls_amount
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Count
0
20
40
60
80
100
120
140
tolls_amount
Boxplot of tolls_amount
140
120
100
80
60
40
20
0
tolls_amount
Density Plot of tolls_amount
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Density
0
20
40
60
80
100
120
140
tolls_amount

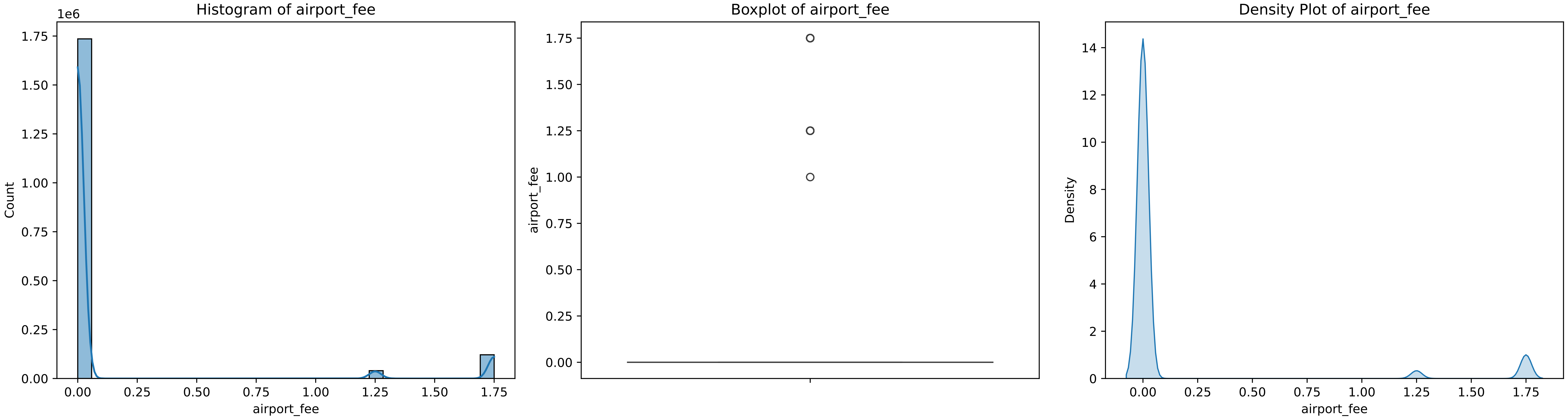

Histogram of improvement_surcharge
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Count
0.0
0.2
0.4
0.6
0.8
1.0
improvement_surcharge
Boxplot of improvement_surcharge
1.0
0.8
0.6
0.4
0.2
0.0
improvement_surcharge
Density Plot of improvement_surcharge
250
200
150
100
50
0
Density
0.0
0.2
0.4
0.6
0.8
1.0
improvement_surcharge


Histogram of total_amount
1e7
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0
20000
40000
60000
80000
100000
120000
140000
total_amount


Boxplot of total_amount
140000
120000
100000
80000
60000
40000
20000
0
total_amount


Density Plot of total_amount
1e−7
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Density
0
20000
40000
60000
80000
100000
120000
140000
total_amount


Histogram of congestion_surcharge
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0.0
0.5
1.0
1.5
2.0
2.5
congestion_surcharge
Boxplot of congestion_surcharge
2.5
2.0
1.5
1.0
0.5
0.0
congestion_surcharge
Density Plot of congestion_surcharge
10
8
6
4
2
0
Density
0.0
0.5
1.0
1.5
2.0
2.5
congestion_surcharge


Histogram of airport_fee
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
airport_fee
Boxplot of airport_fee
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
airport_fee
Density Plot of airport_fee
14
12
10
8
6
4
2
0
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
airport_fee

Conclusion/Remarks:

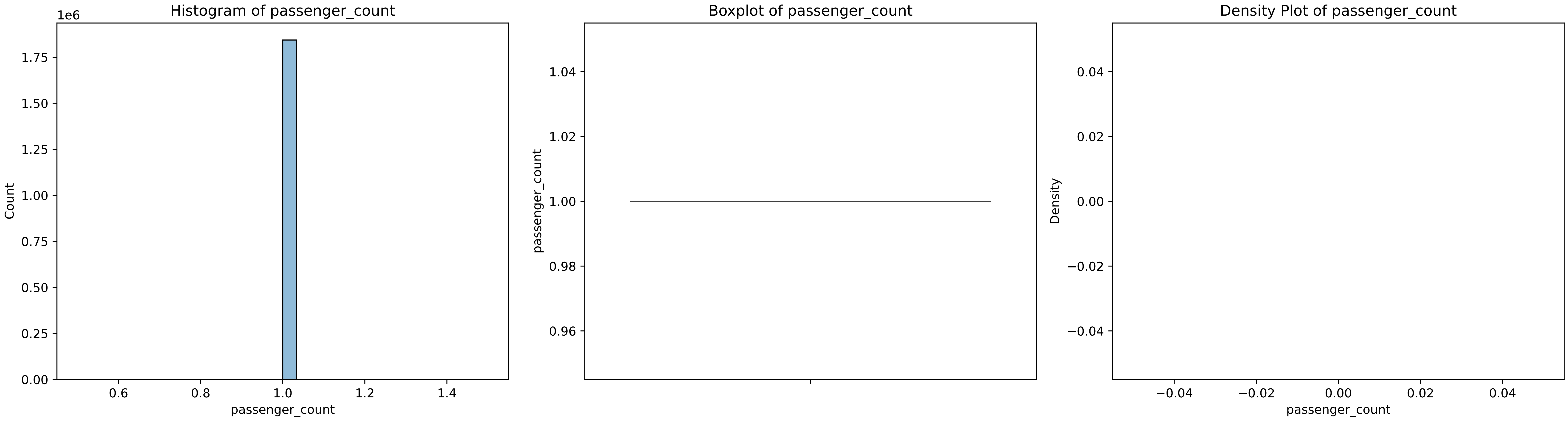
2.3 Analysing outliers for numerical values Only in Histogram,Box and KDE Plot. And there are lot of outliers in long skew. So dropping of 5% extreme data

Conclusion/Remarks:

2.3 Removes entire rows where certain columns have outliers. Drops rows only if the percentage of outliers in a column is below a given threshold (5%). Keeps track of outlier percentages before and after removal

Conclusion/Remarks:

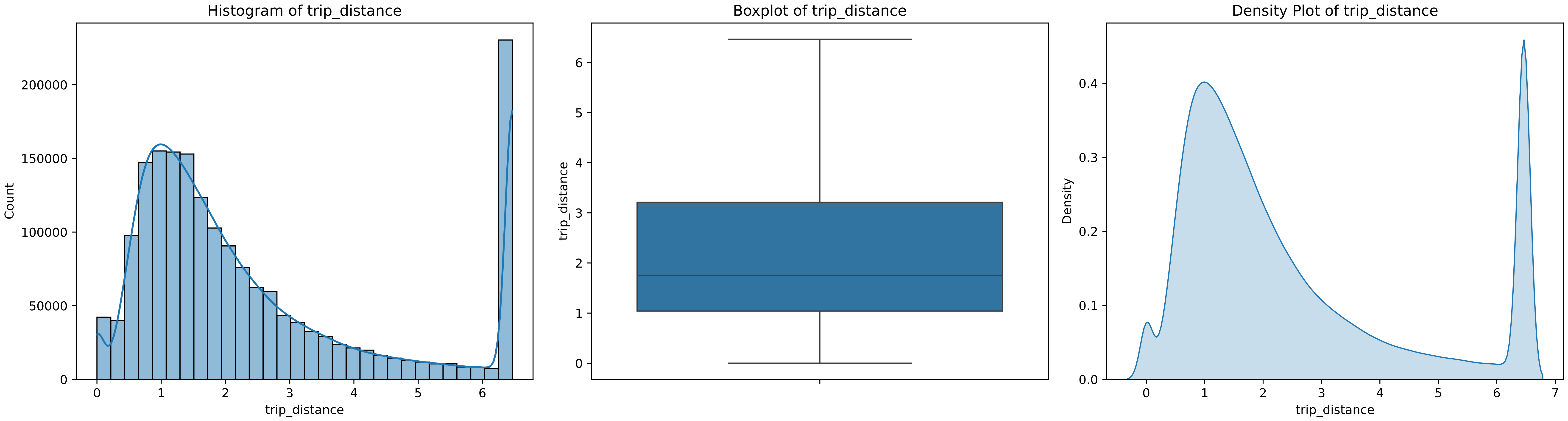
2.3 Analysing the Skewness by Clipping data outside lower_bound = Q1 - 1.5 * IQR, upper_bound = Q3 + 1.5 * IQR

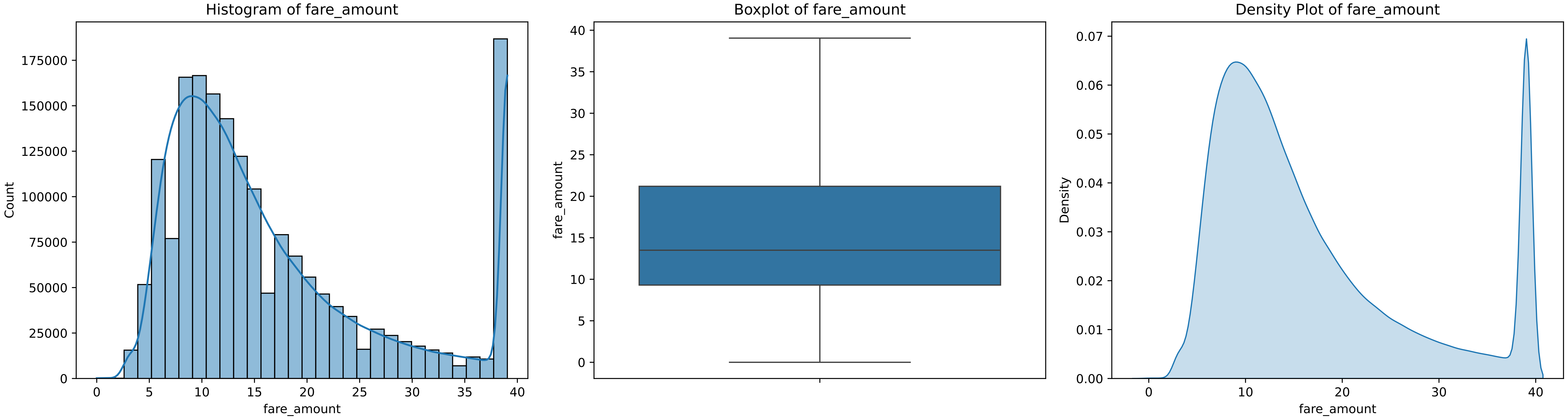


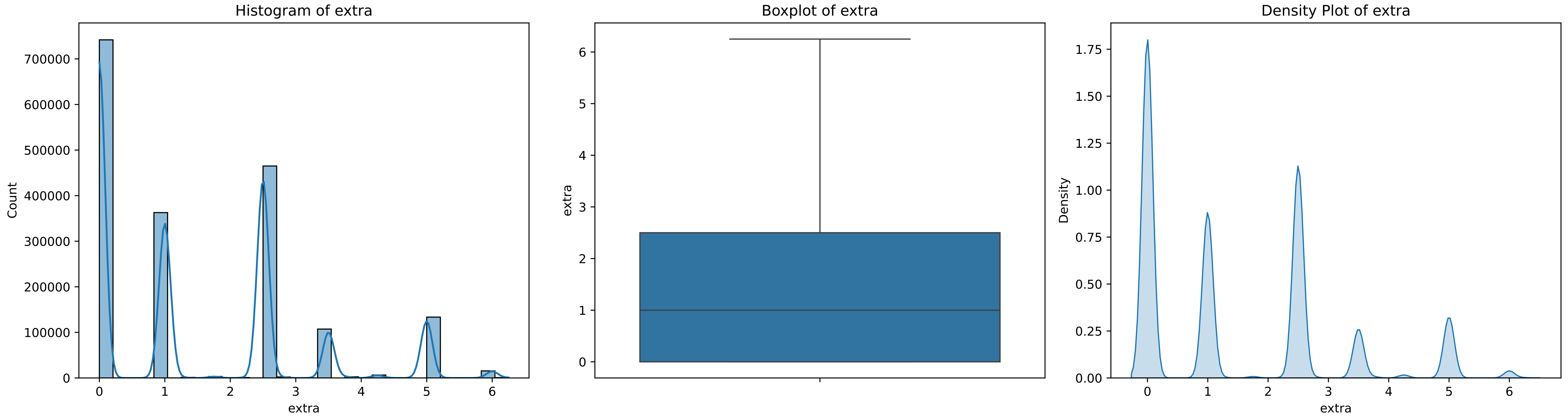
Histogram of passenger_count

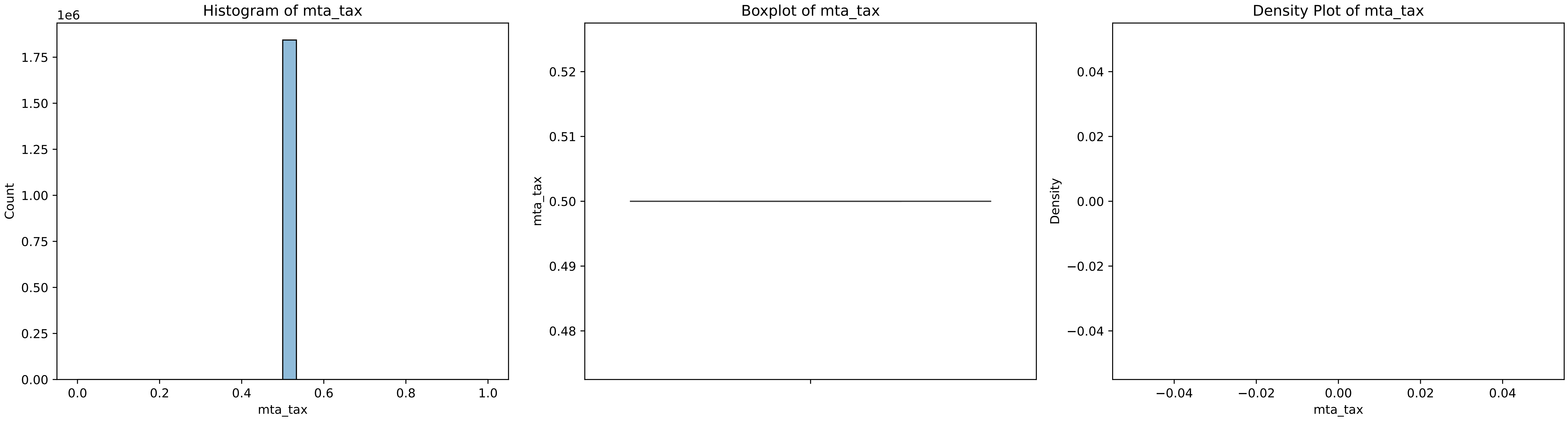
Boxplot of passenger_count

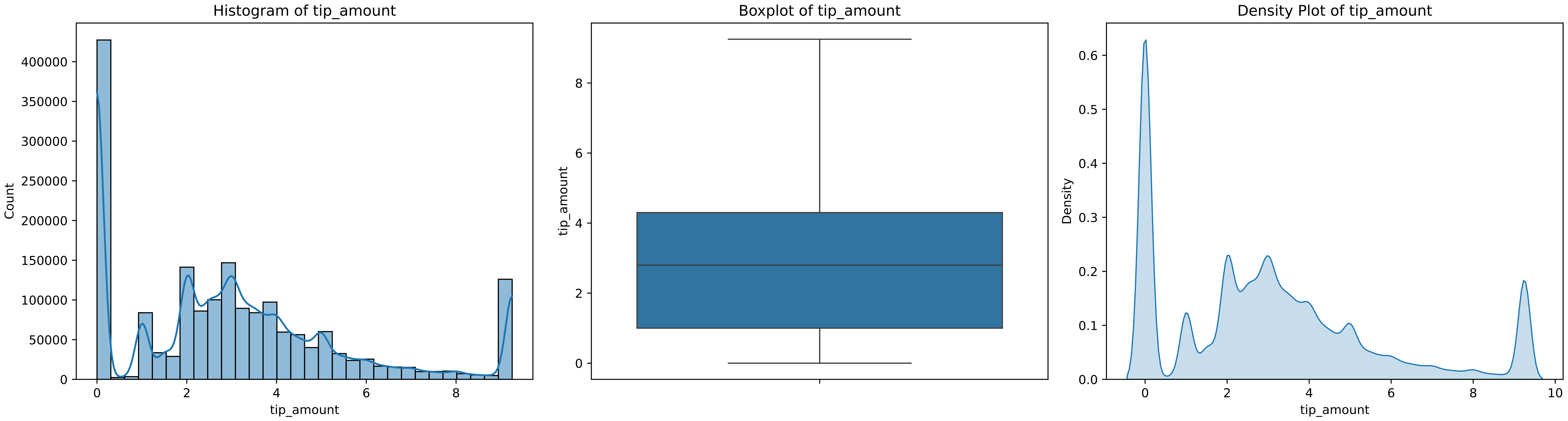
Density Plot of passenger_count

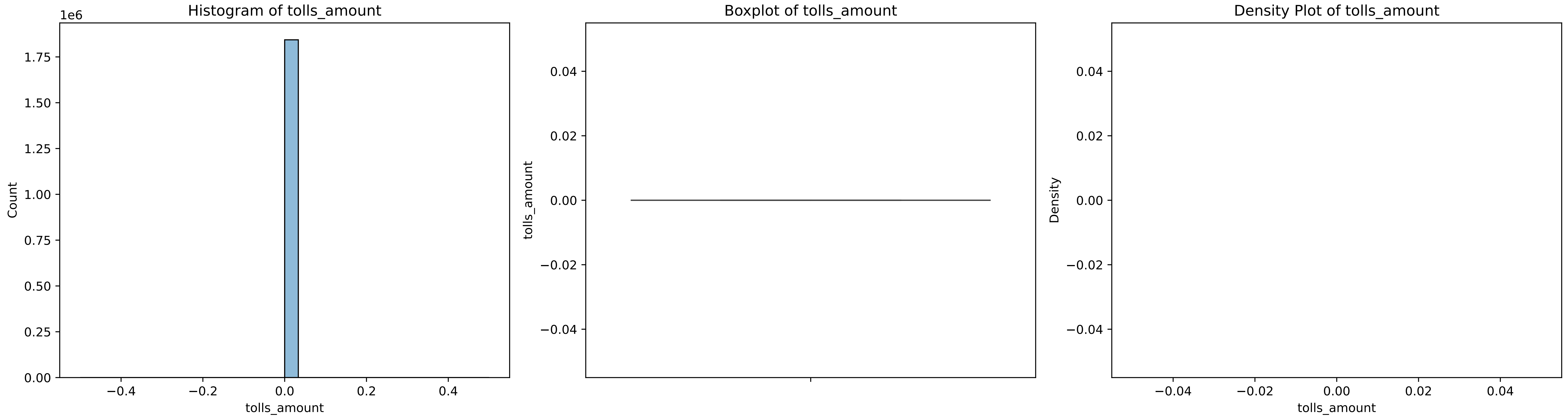

Histogram of trip_distance
Count
200000
150000
100000
50000
0
0
1
2
3
4
5
6
trip_distance
Boxplot of trip_distance
trip_distance
6
5
4
3
2
1
0
Density Plot of trip_distance
Density
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
7
trip_distance

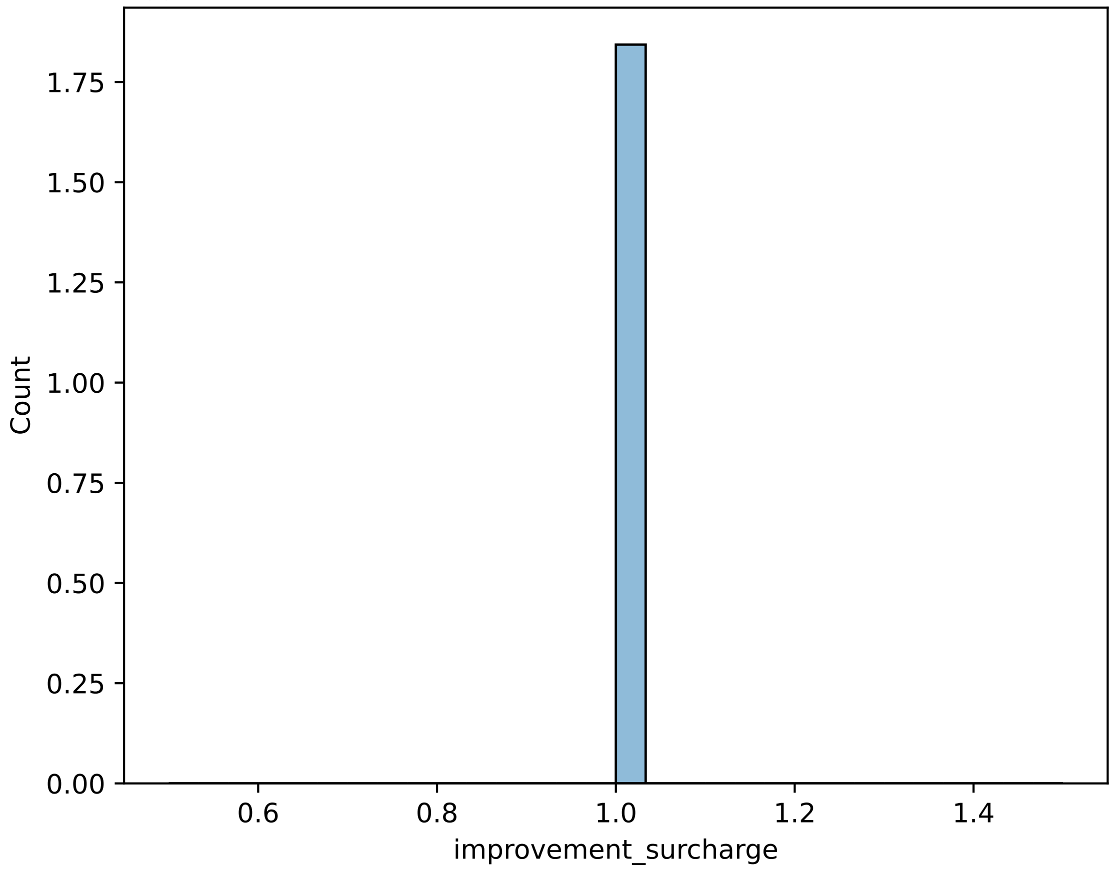

Histogram of fare_amount
Count
0
25000
50000
75000
100000
125000
150000
175000
0
5
10
15
20
25
30
35
40
fare_amount
Boxplot of fare_amount
fare_amount
0
5
10
15
20
25
30
35
40
Density Plot of fare_amount
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0
10
20
30
40
fare_amount


Histogram of extra
Count
0
100000
200000
300000
400000
500000
600000
700000
0
1
2
3
4
5
6
extra
Boxplot of extra
extra
0
1
2
3
4
5
6
Density Plot of extra
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0
1
2
3
4
5
6
extra


Histogram of mta_tax
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0.0
0.2
0.4
0.6
0.8
1.0
mta_tax
Boxplot of mta_tax
0.52
0.51
0.50
0.49
0.48
mta_tax
Density Plot of mta_tax
0.04
0.02
0.00
−0.02
−0.04
Density
−0.04
−0.02
0.00
0.02
0.04
mta_tax

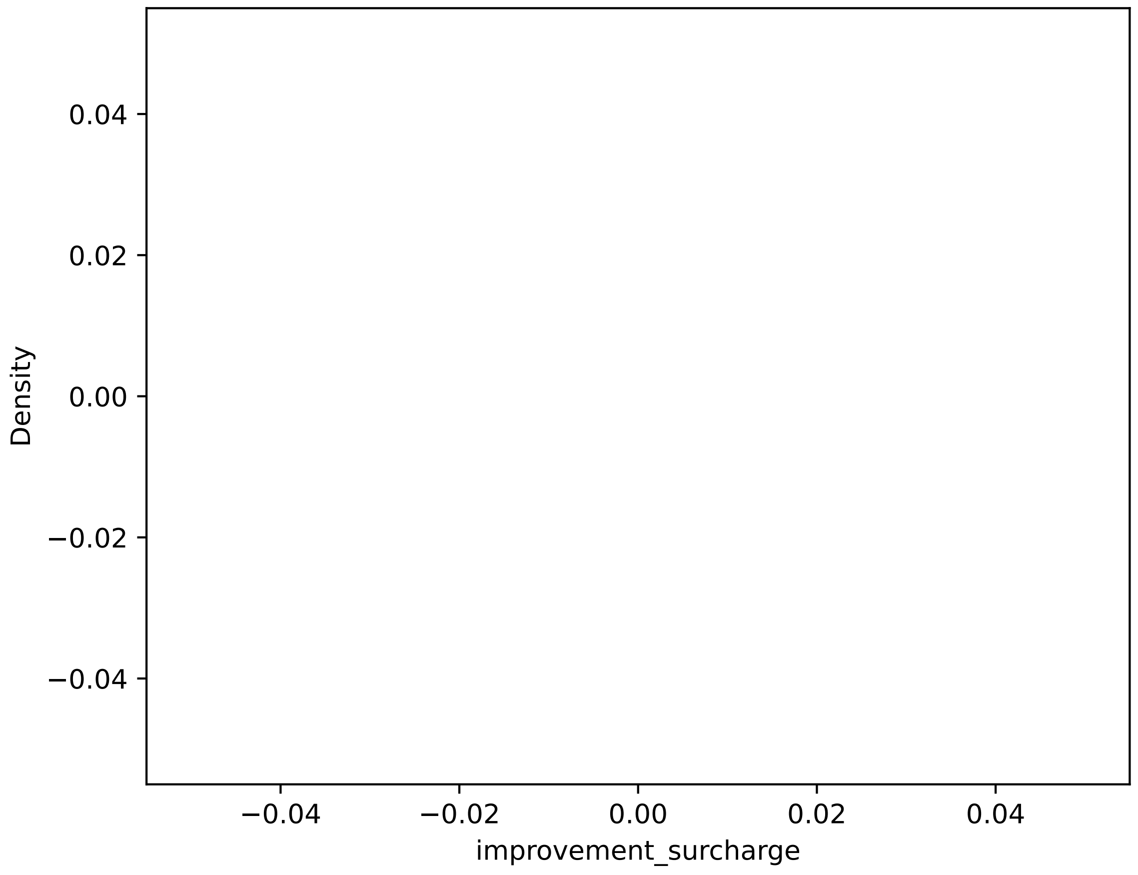

Histogram of tip_amount
Count
0
50000
100000
150000
200000
250000
300000
350000
400000
0
2
4
6
8
tip_amount
Boxplot of tip_amount
tip_amount
0
2
4
6
8
Density Plot of tip_amount
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
2
4
6
8
10
tip_amount


Histogram of tolls_amount
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
−0.4
−0.2
0.0
0.2
0.4
tolls_amount
Boxplot of tolls_amount
0.04
0.02
0.00
−0.02
−0.04
tolls_amount
Density Plot of tolls_amount
0.04
0.02
0.00
−0.02
−0.04
Density
−0.04
−0.02
0.00
0.02
0.04
tolls_amount

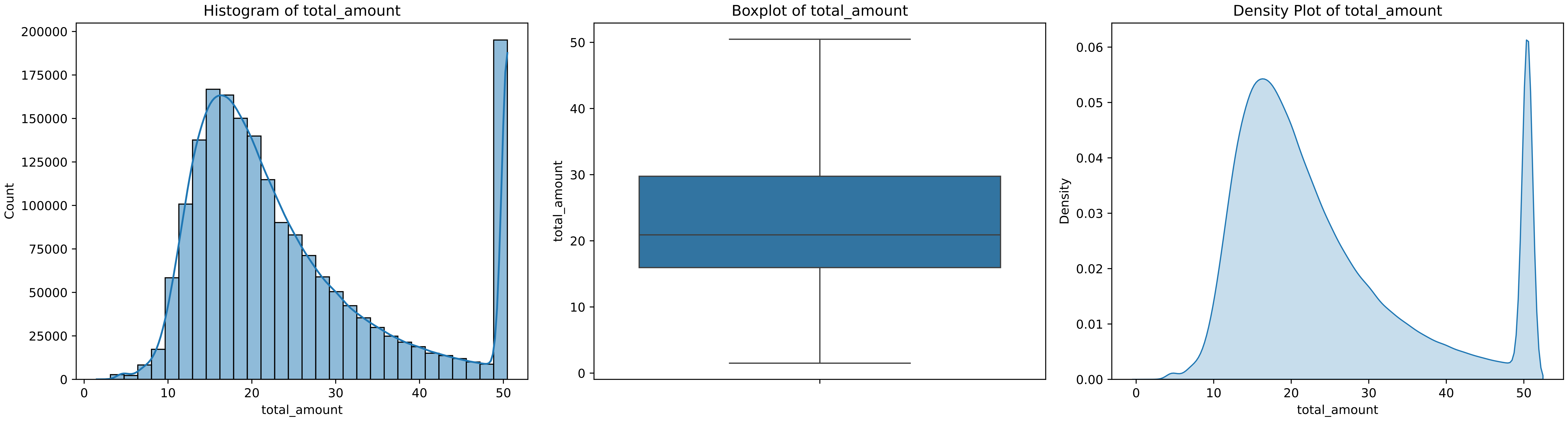


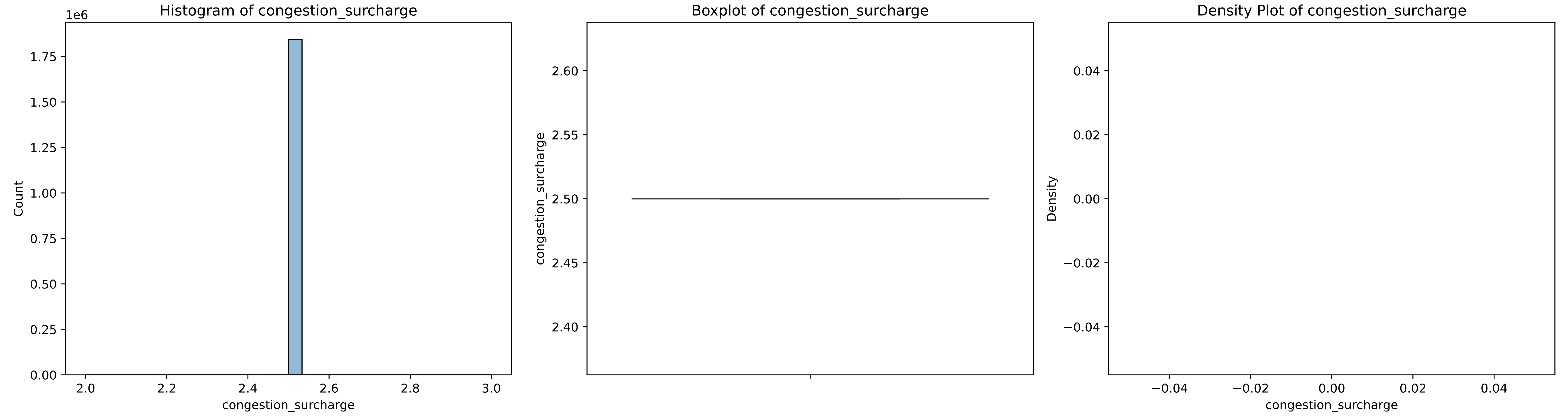
Histogram of improvement_surcharge

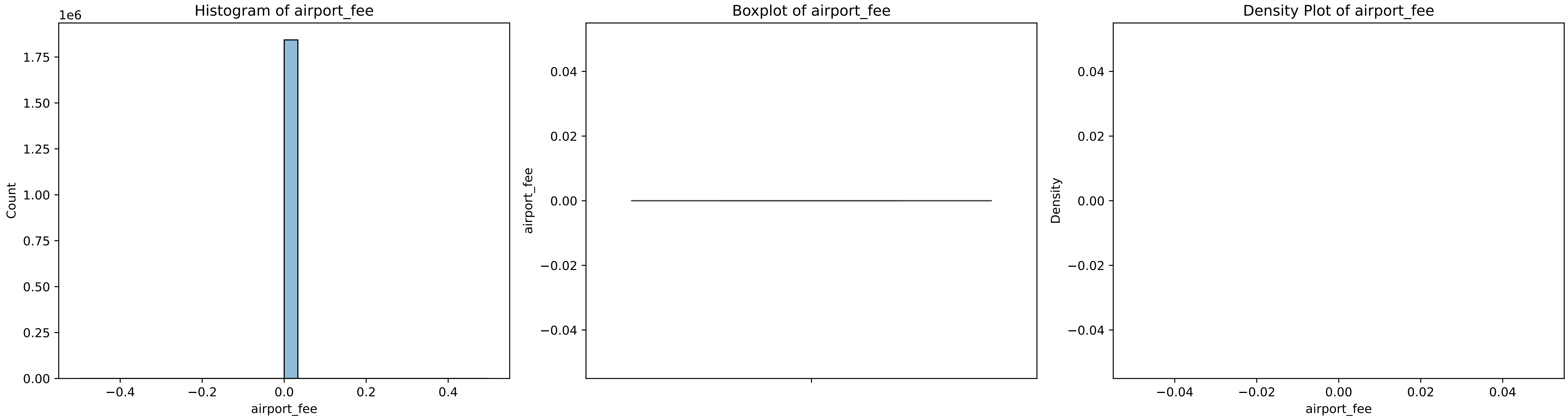
Boxplot of improvement_surcharge



Density Plot of improvement_surcharge

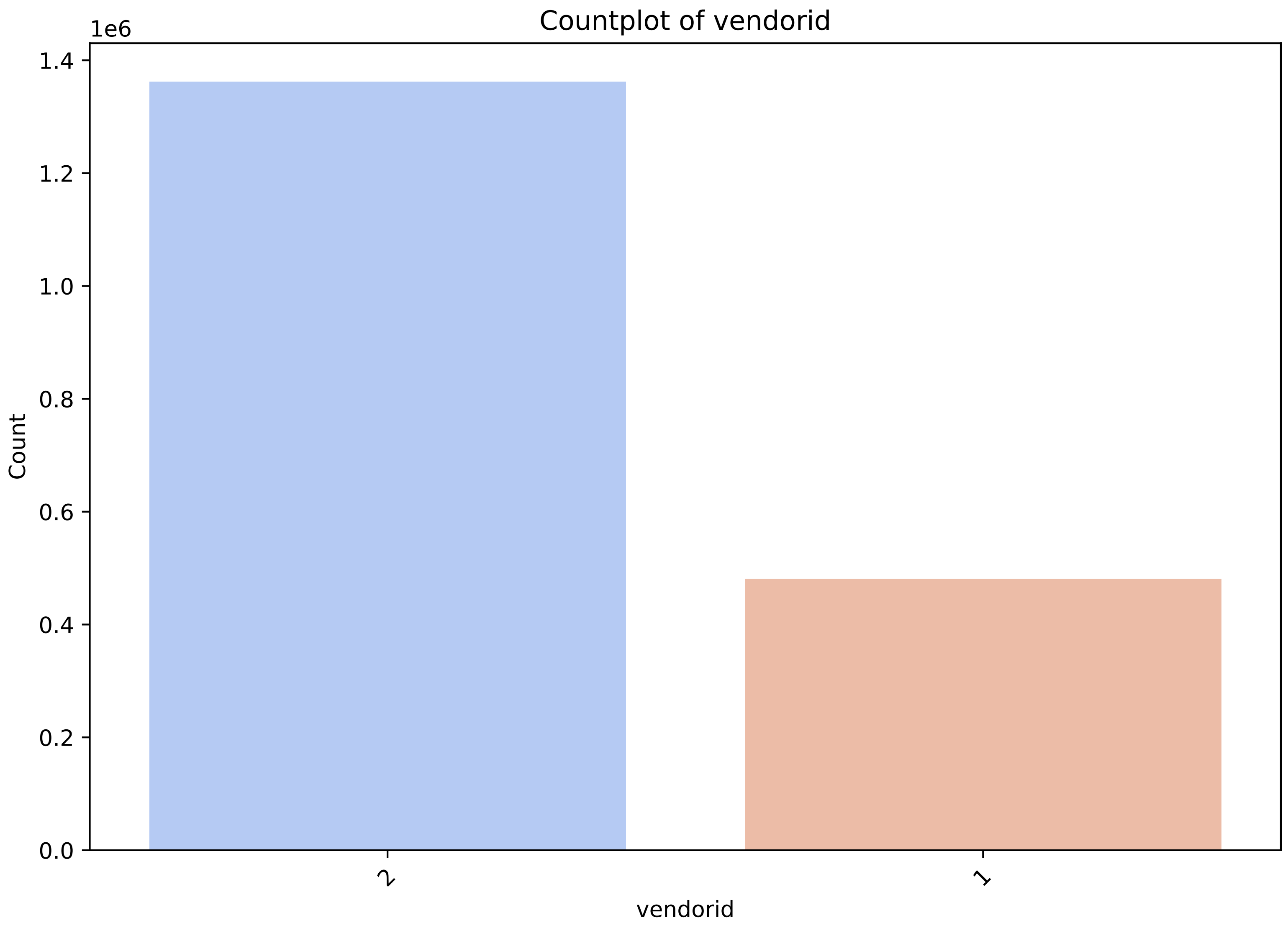

Histogram of total_amount
Count
200000
175000
150000
125000
100000
75000
50000
25000
0
0
10
20
30
40
50
total_amount
Boxplot of total_amount
total_amount
50
40
30
20
10
0
Density Plot of total_amount
Density
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40
50
total_amount

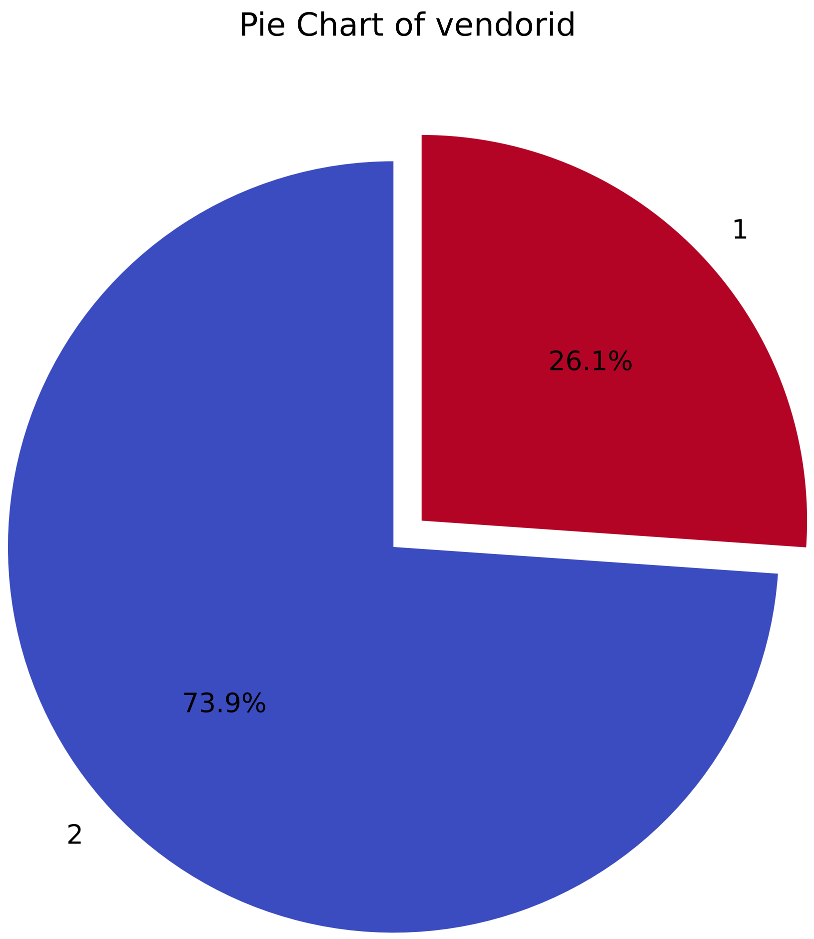

Histogram of congestion_surcharge
1e6
Count
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2.0
2.2
2.4
2.6
2.8
3.0
congestion_surcharge
Boxplot of congestion_surcharge
congestion_surcharge
2.60
2.55
2.50
2.45
2.40
Density Plot of congestion_surcharge
Density
0.04
0.02
0.00
−0.02
−0.04
−0.04
−0.02
0.00
0.02
0.04
congestion_surcharge

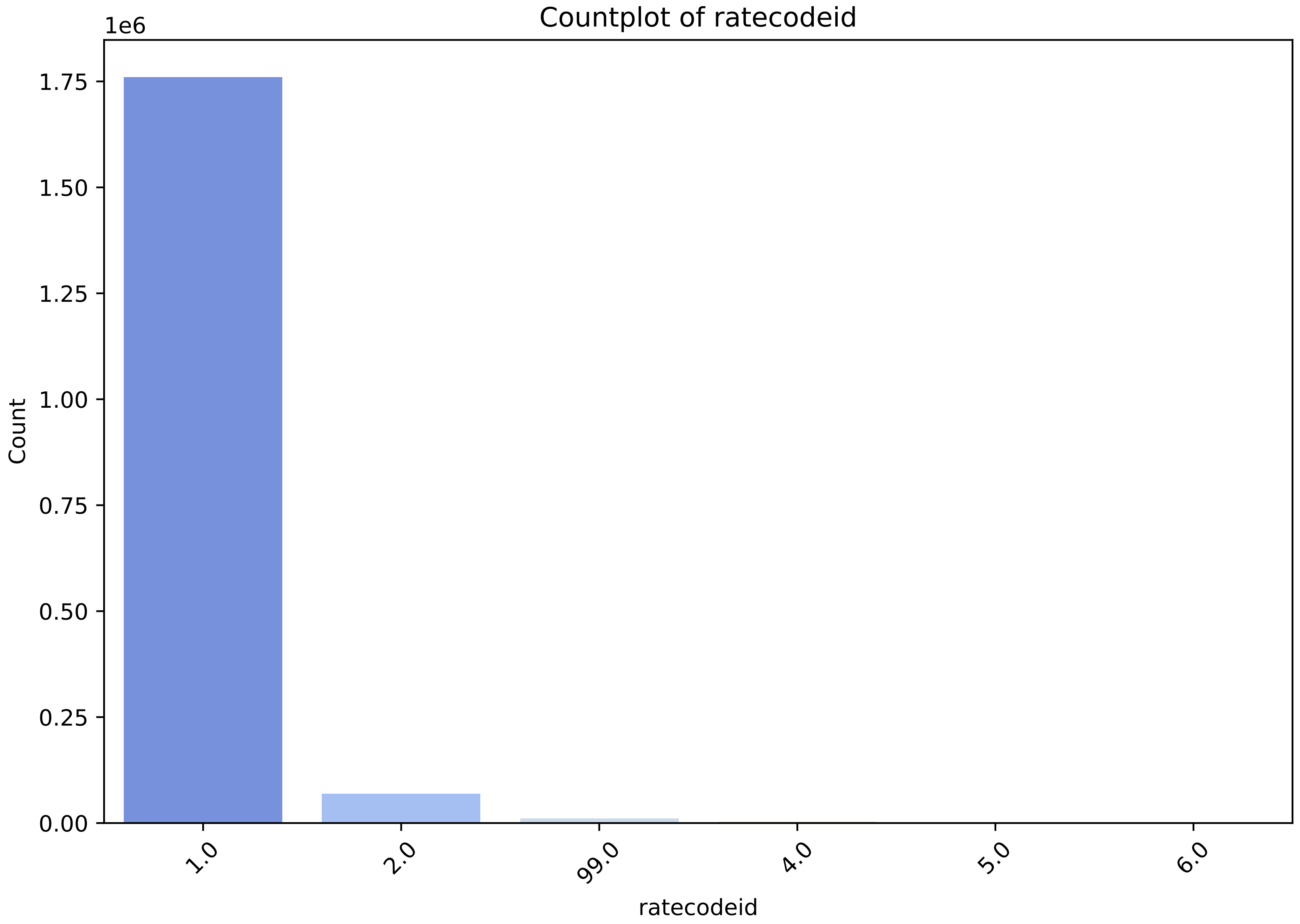

Histogram of airport_fee
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
−0.4
−0.2
0.0
0.2
0.4
airport_fee
Boxplot of airport_fee
0.04
0.02
0.00
−0.02
−0.04
airport_fee
Density Plot of airport_fee
0.04
0.02
0.00
−0.02
−0.04
Density
−0.04
−0.02
0.00
0.02
0.04
airport_fee

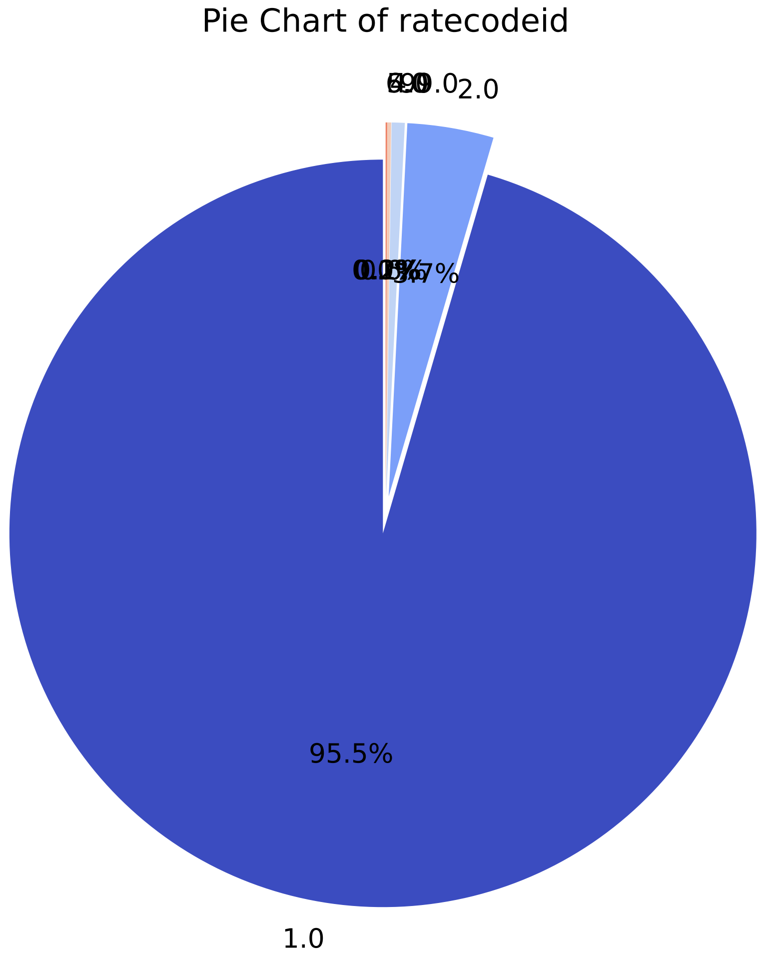
Conclusion/Remarks:

2.3 Analysing the Outlier and Skewness After Clipping data

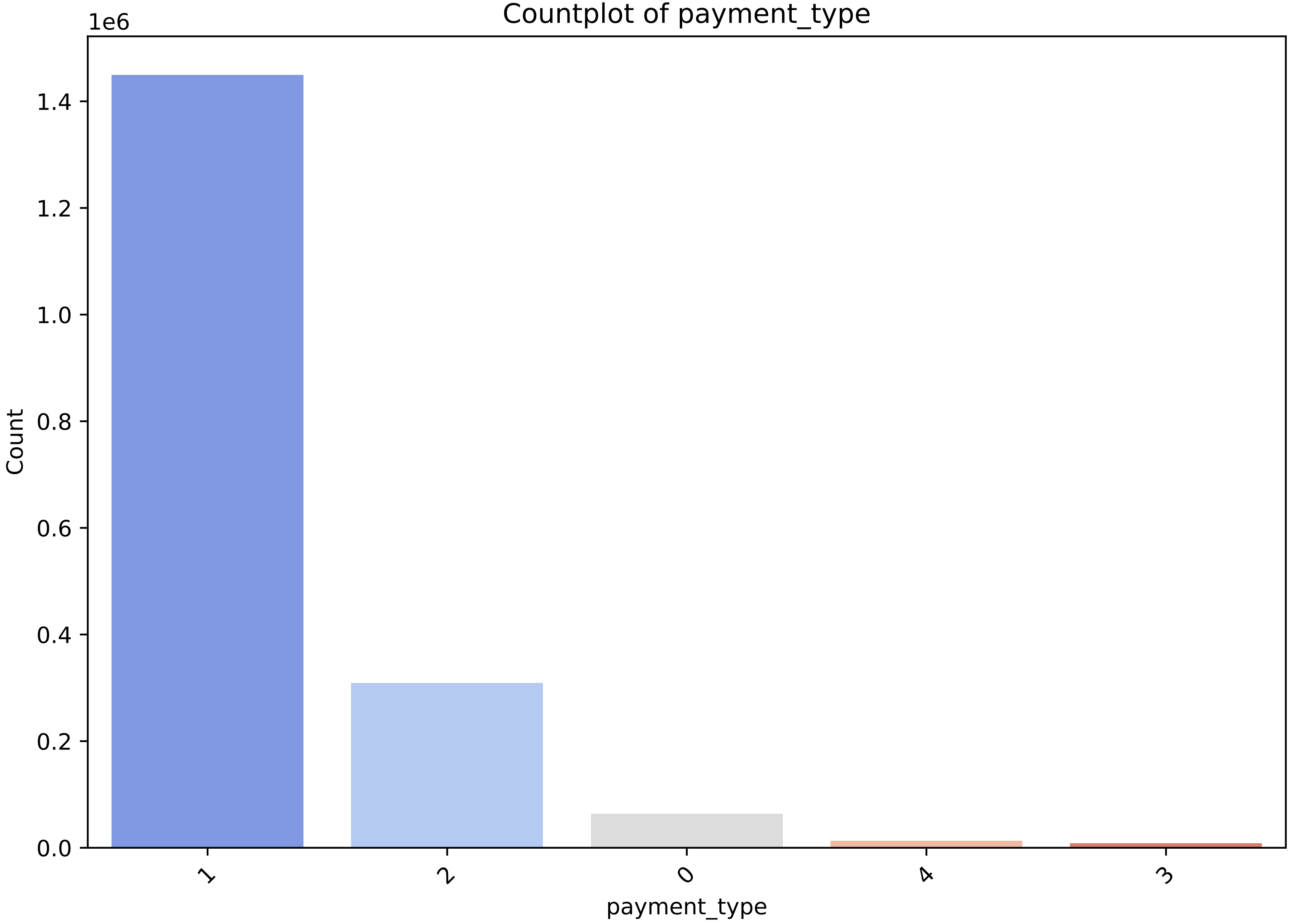

Countplot of vendorid
1e6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Count
2
1
vendorid


Pie Chart of vendorid
1
26.1%
73.9%
2

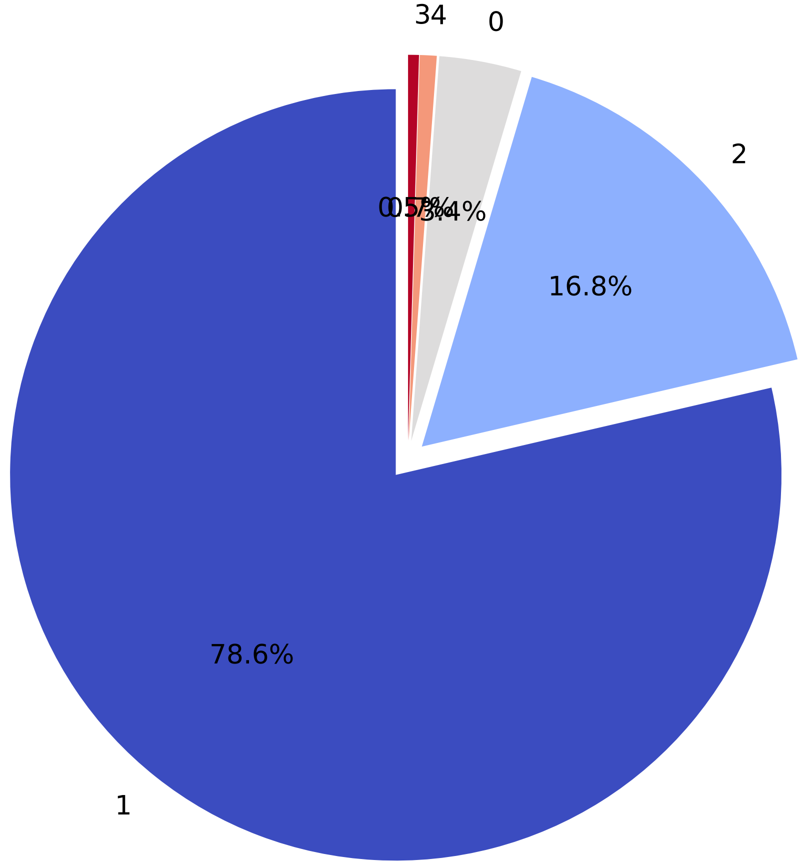

Countplot of ratecodeid
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
1.0
2.0
99.0
4.0
5.0
6.0
ratecodeid


Pie Chart of ratecodeid
2.0
3.7%
95.5%
1.0

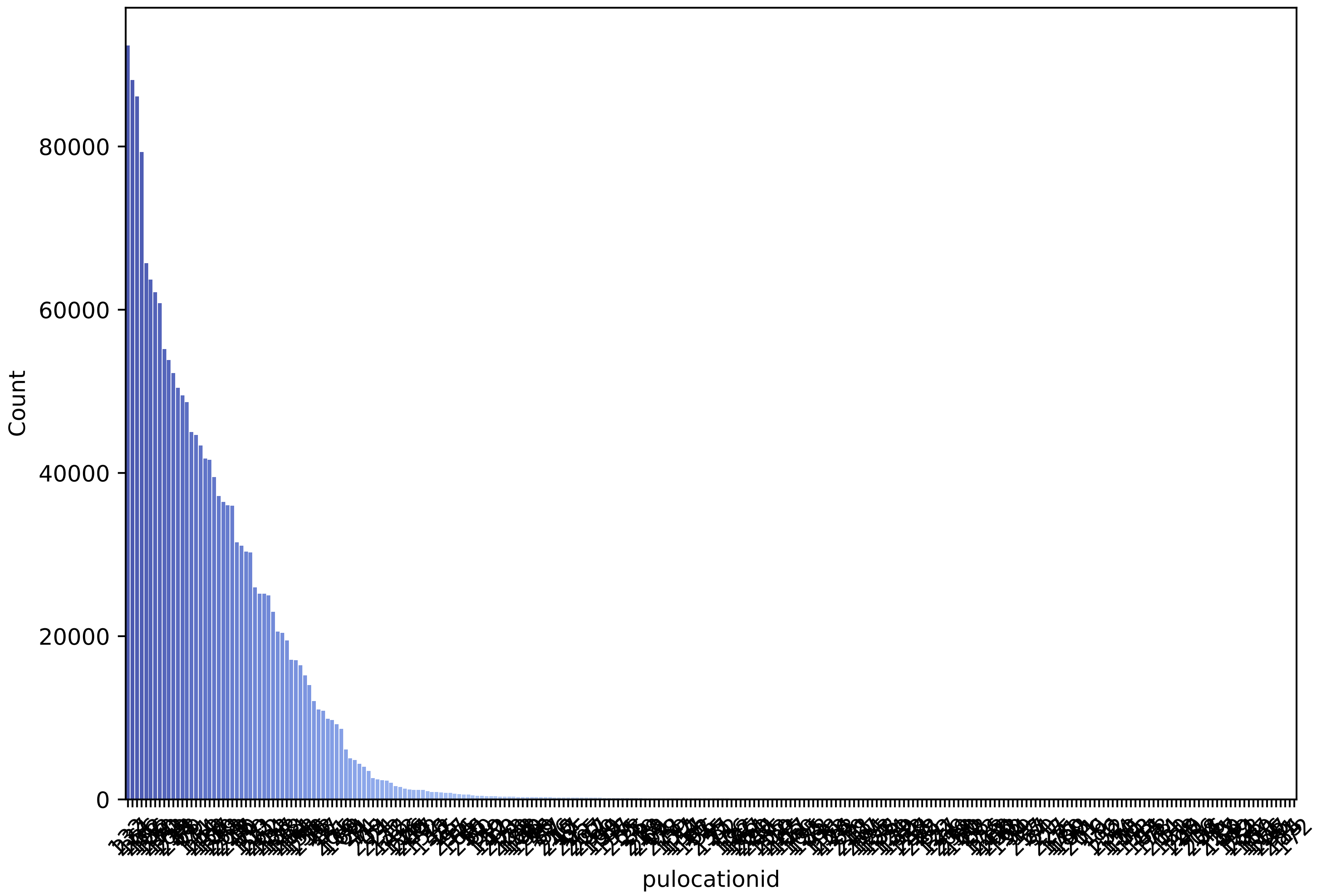
Countplot of payment_type



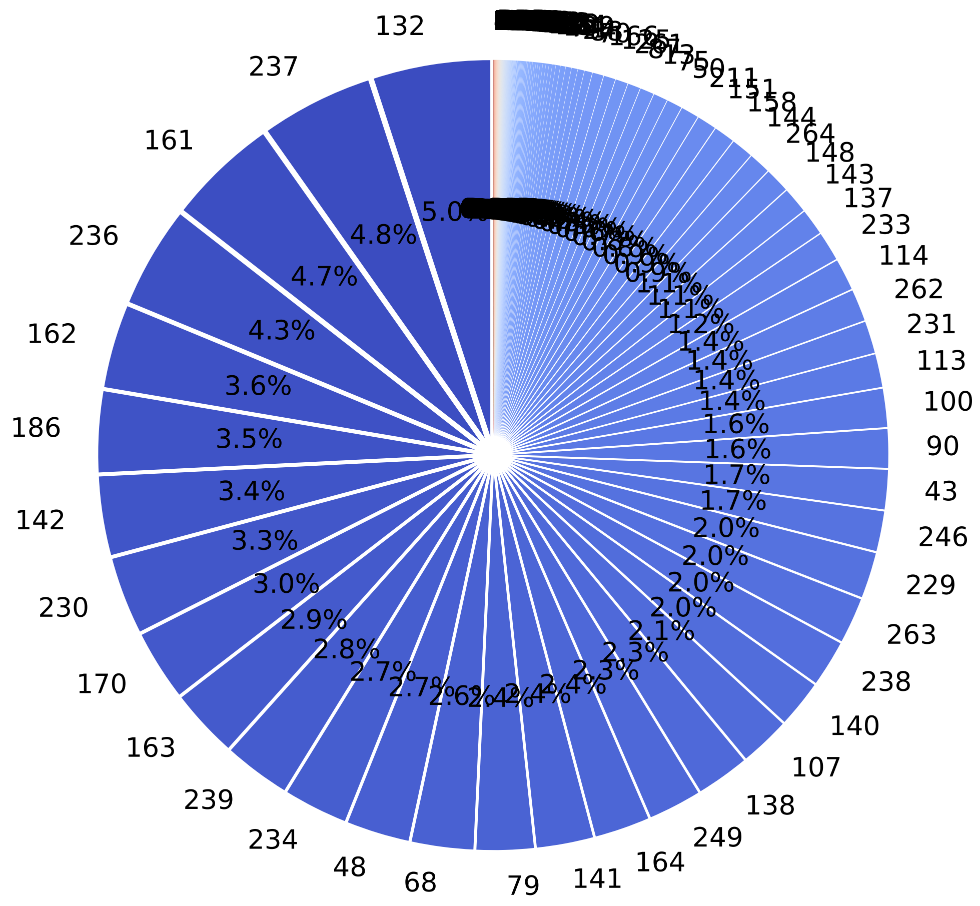
Pie Chart of payment_type

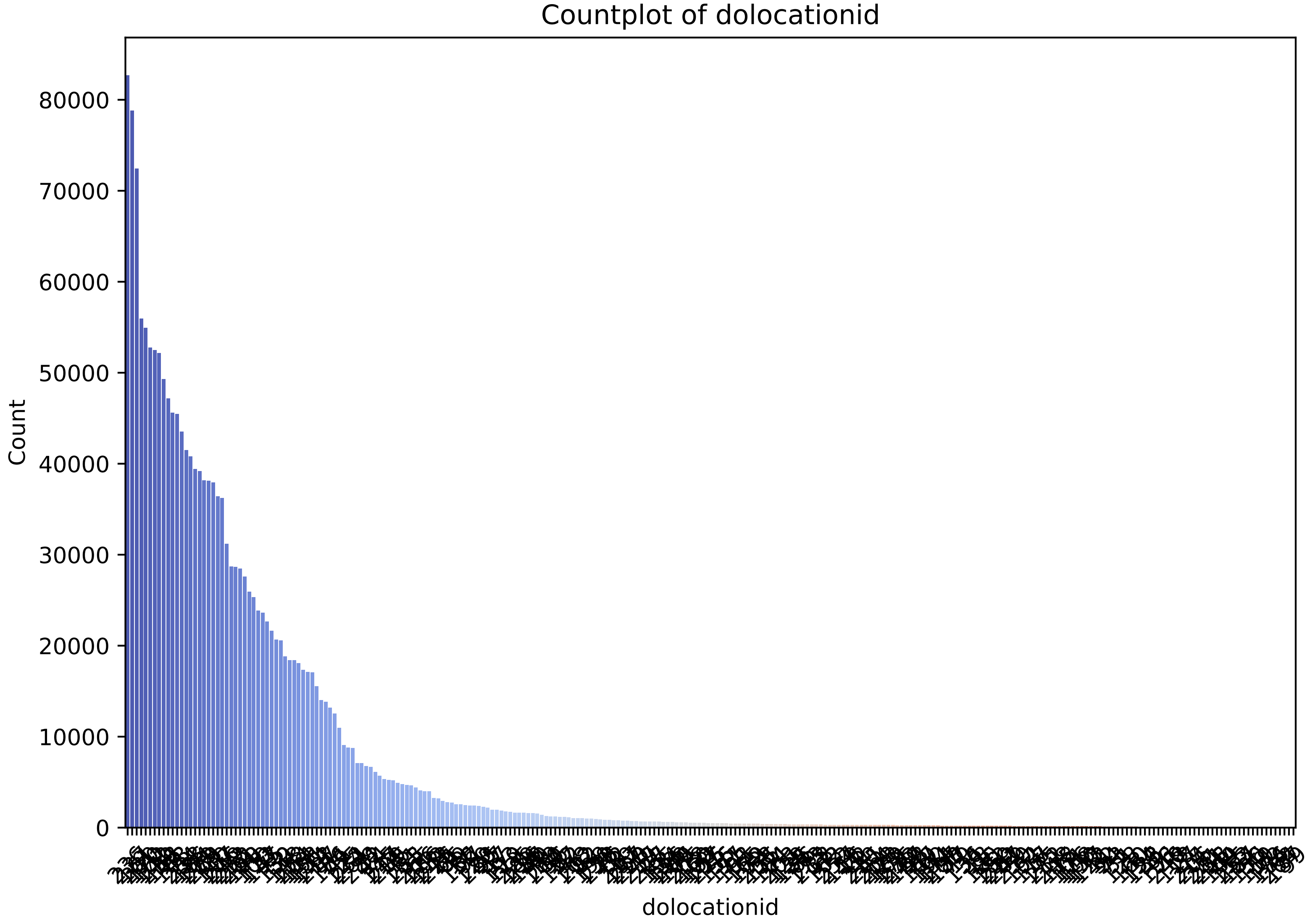


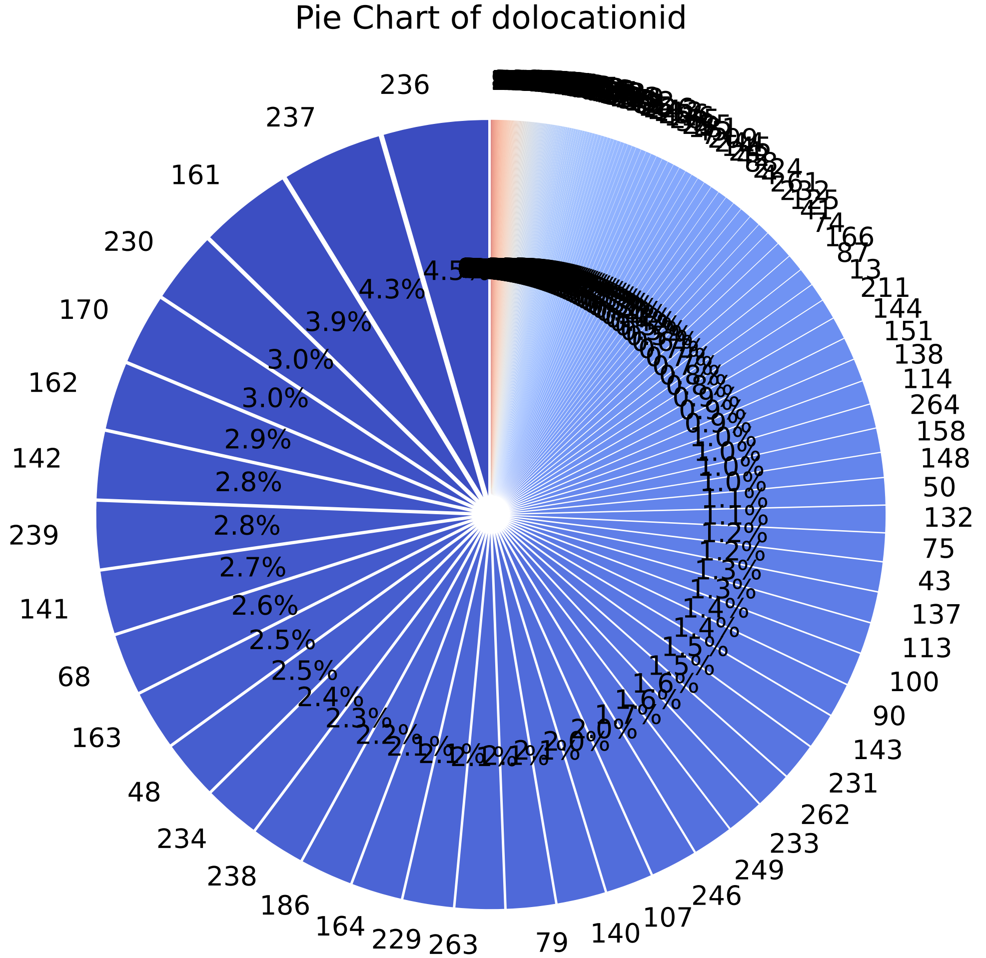
Countplot of pulocationid

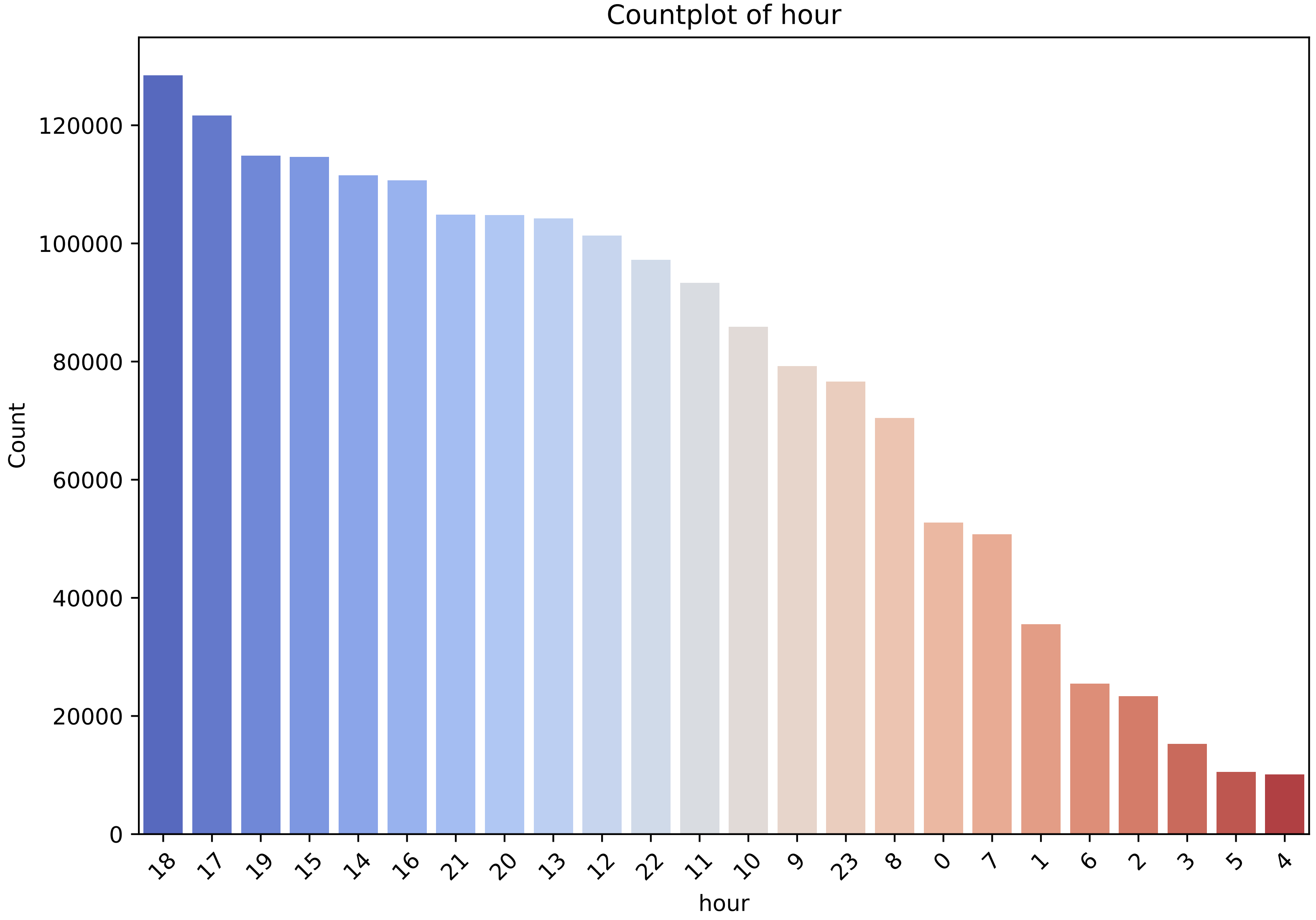


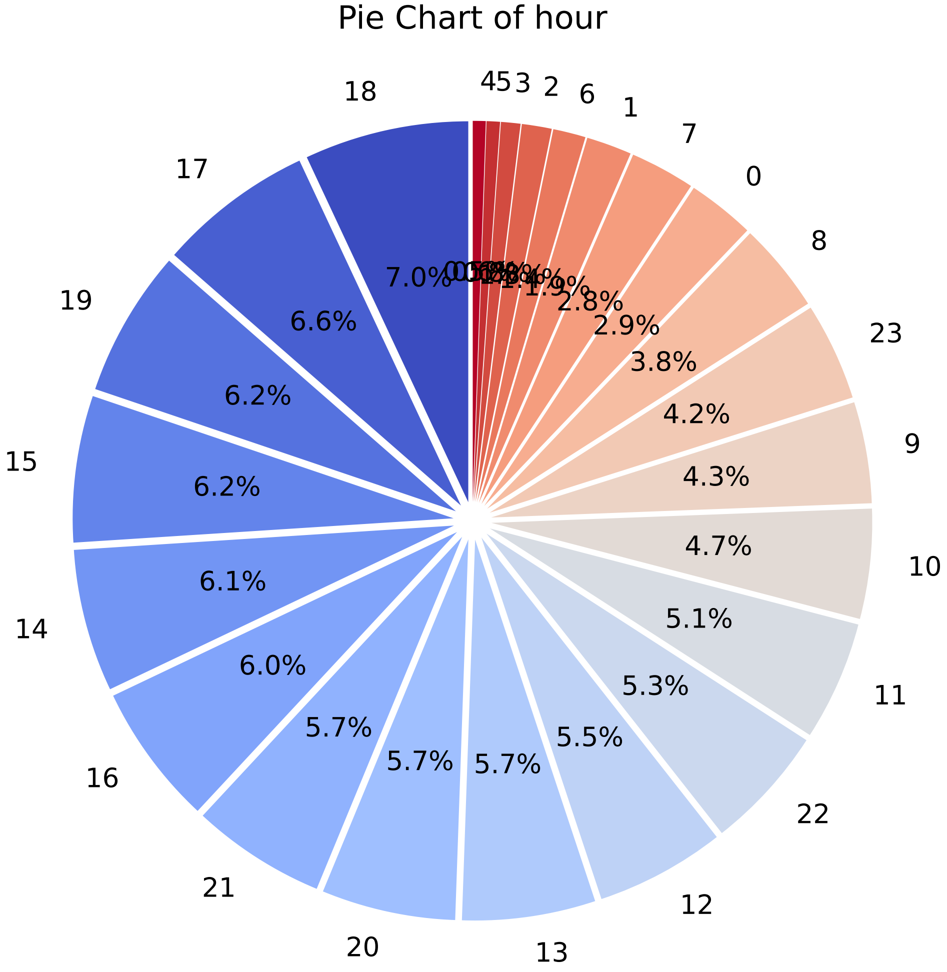
Pie Chart of pulocationid




Countplot of dolocationid
Count
80000
70000
60000
50000
40000
30000
20000
10000
0
dolocationid

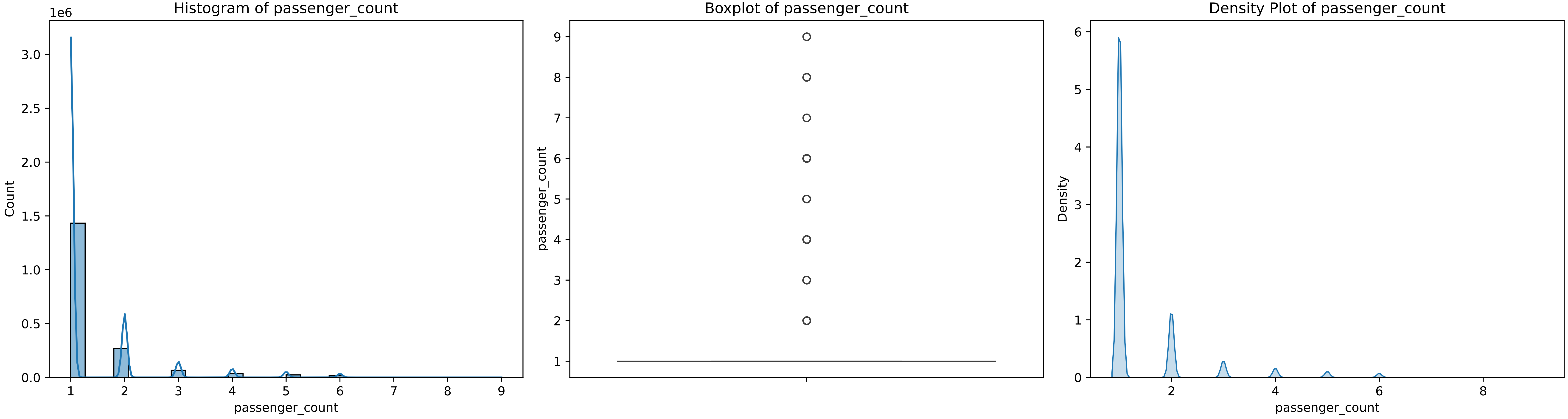

Pie Chart of dolocationid
236
237
161
230
170
162
142
239
141
68
163
48
234
238
186
164
229
263
79
140
107
246
249
233
262
231
143
90
100
113
137
43
75
132
50
148
158
264
114
138
151
144
211
13
87
166
74
4.5%
4.3%
3.9%
3.0%
3.0%
2.9%
2.8%
2.8%
2.7%
2.6%
2.5%
2.5%
2.4%
2.3%
2.2%
1.7%
1.6%
1.6%
1.5%
1.5%
1.4%
1.4%
1.3%
1.3%
1.2%
1.2%
1.1%
1.0%
1.0%


Countplot of hour
Count
hour
0
20000
40000
60000
80000
100000
120000
18
17
19
15
14
16
21
20
13
12
22
11
10
9
23
8
0
7
1
6
2
3
5
4


Pie Chart of hour
18
7.0%
17
6.6%
19
6.2%
15
6.2%
14
6.1%
16
6.0%
21
5.7%
20
5.7%
13
5.7%
12
5.5%
22
5.3%
11
5.1%
10
4.7%
9
4.3%
23
4.2%
8
3.8%
0
2.9%
7
2.8%
1
1.9%
6
2
3
5
4

Conclusion/Remarks:

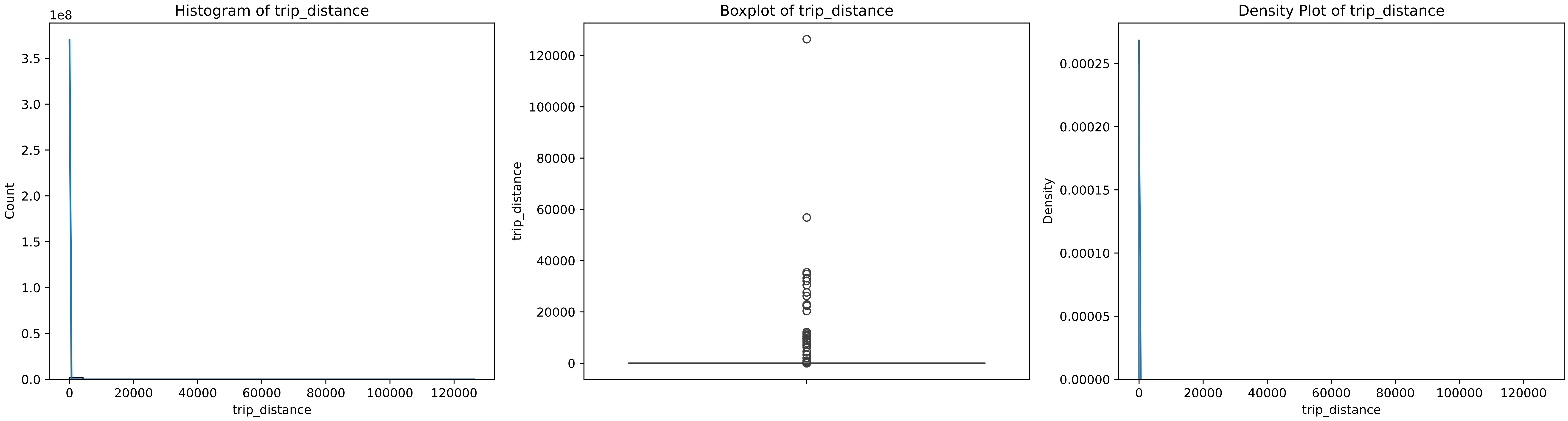
2.3 Analysing the cardinal type data like vendor, ratecode,payment type, piclup location, drop location, hourly pick up etc.

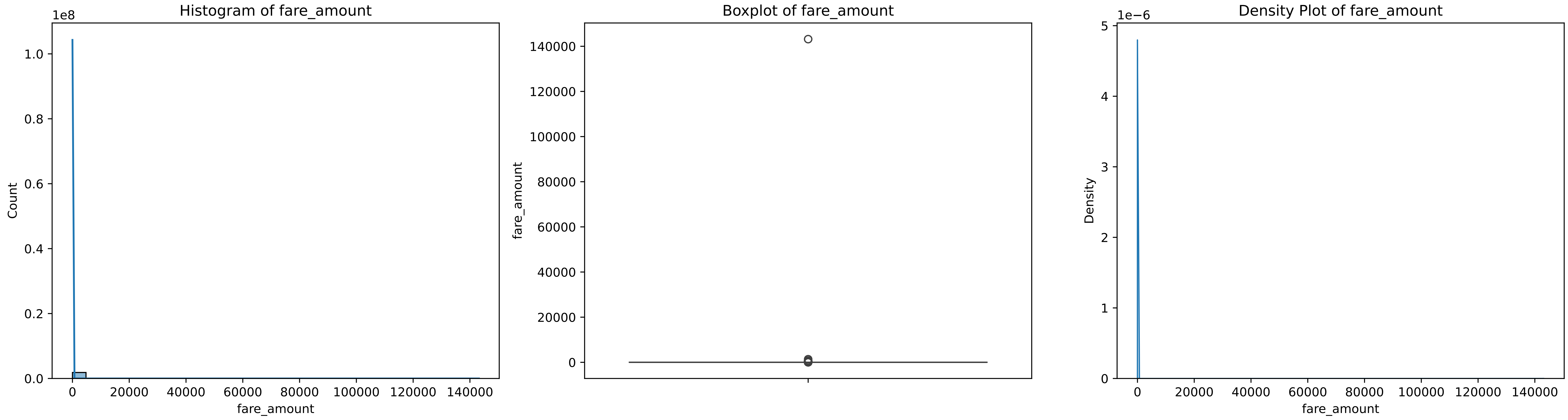


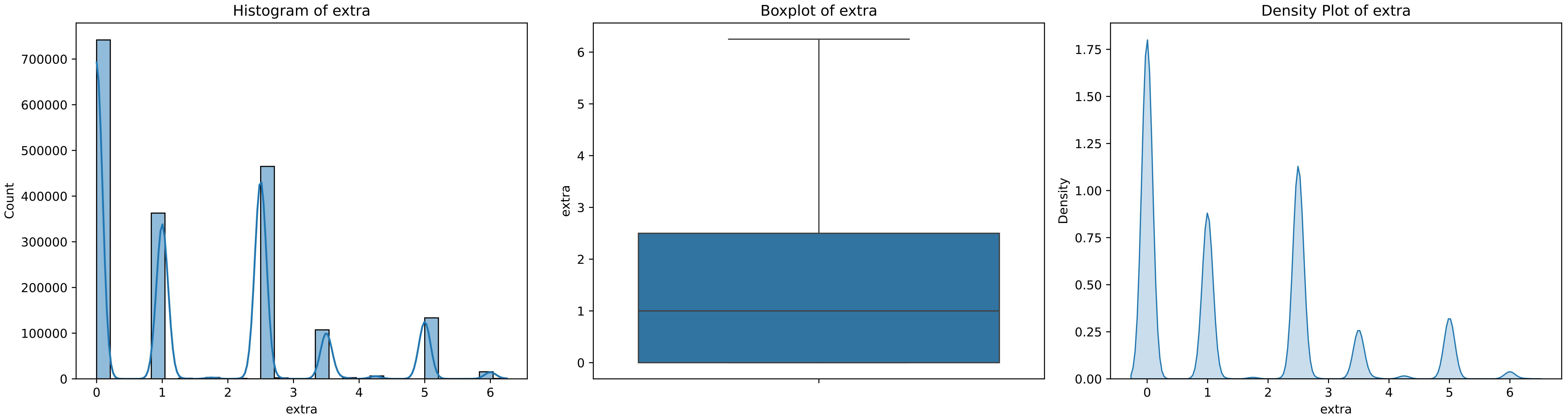
Histogram of passenger_count

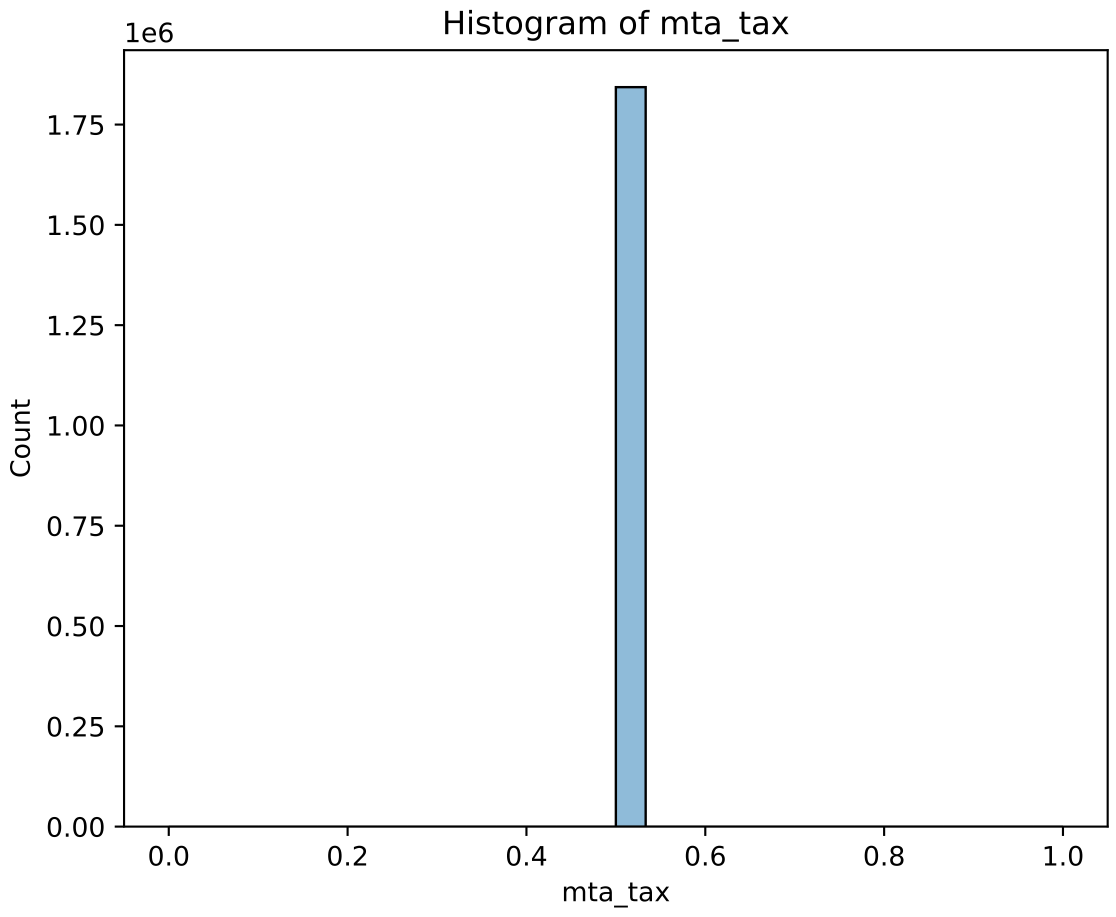
Boxplot of passenger_count

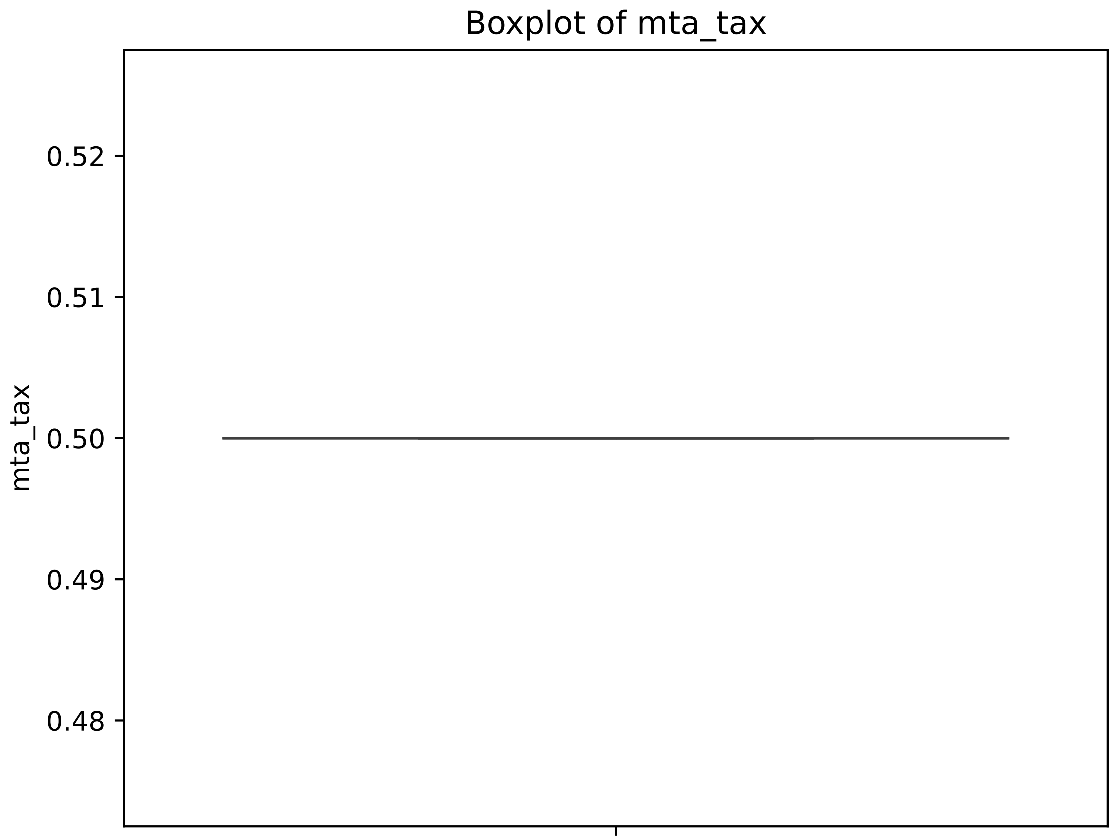
Density Plot of passenger_count

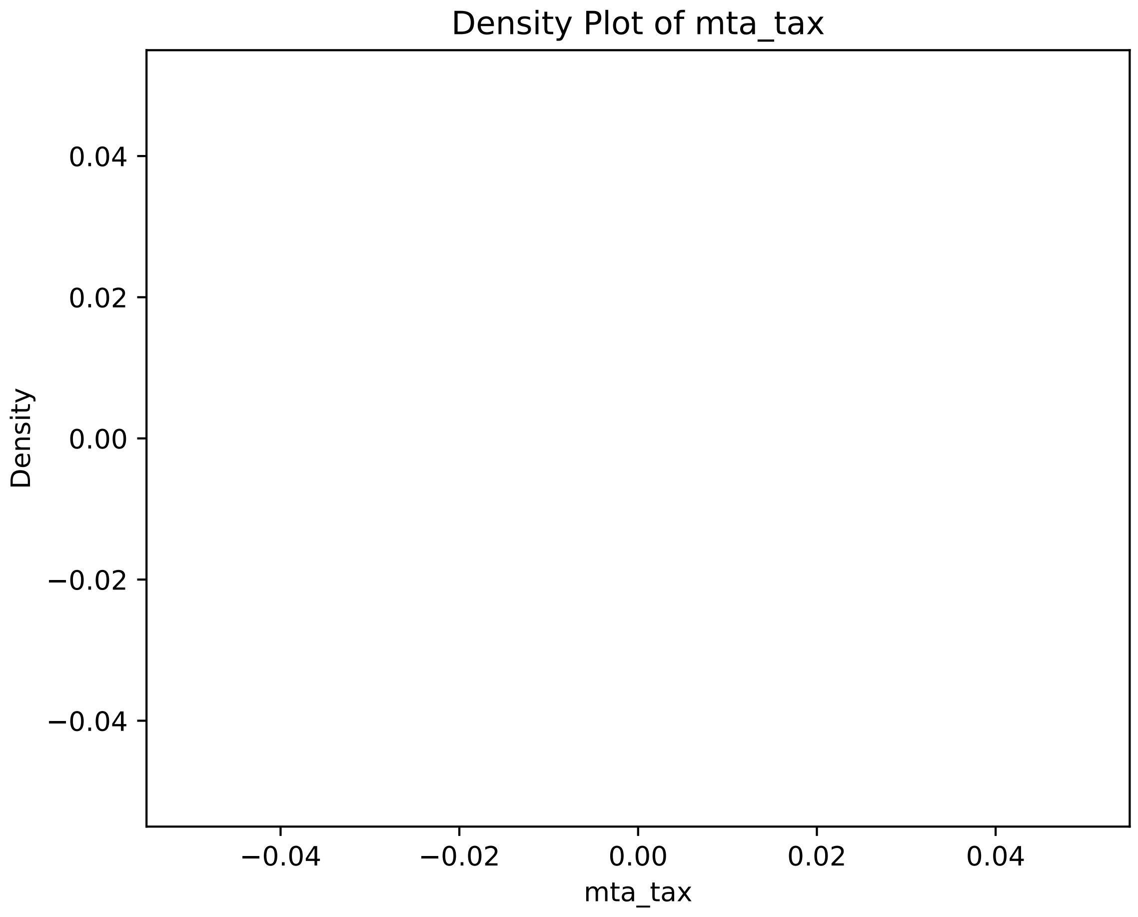

Histogram of trip_distance
1e8
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
0
20000
40000
60000
80000
100000
120000
trip_distance
Boxplot of trip_distance
120000
100000
80000
60000
40000
20000
0
trip_distance
Density Plot of trip_distance
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
Density
0
20000
40000
60000
80000
100000
120000
trip_distance

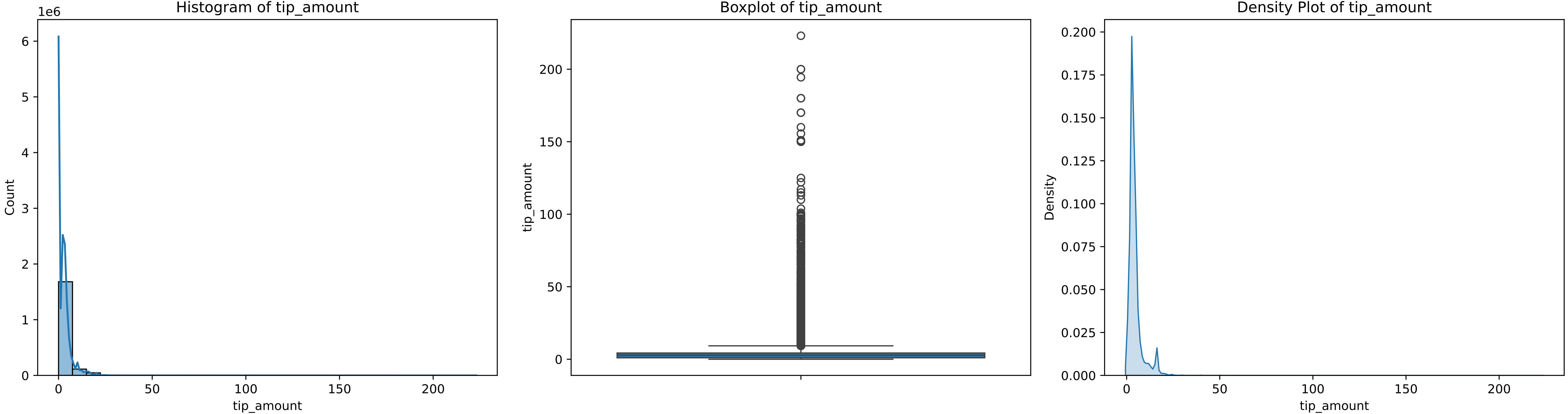

Histogram of fare_amount
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Count
0
20000
40000
60000
80000
100000
120000
140000
fare_amount
Boxplot of fare_amount
140000
120000
100000
80000
60000
40000
20000
0
fare_amount
Density Plot of fare_amount
1e−6
5
4
3
2
1
0
Density
0
20000
40000
60000
80000
100000
120000
140000
fare_amount

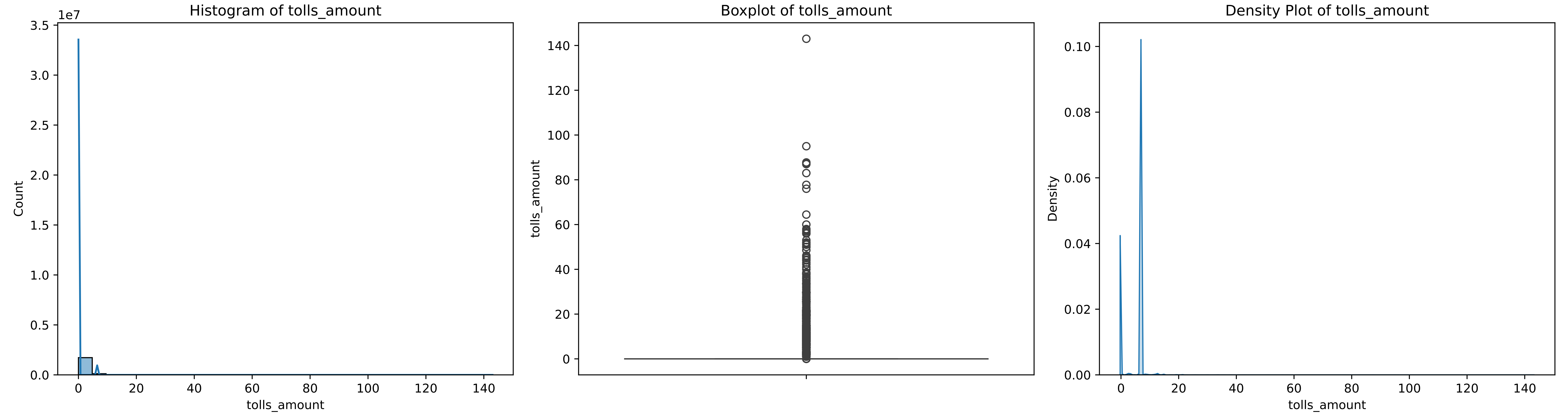

Histogram of extra
Count
700000
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
6
extra
Boxplot of extra
extra
6
5
4
3
2
1
0
Density Plot of extra
Density
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
6
extra

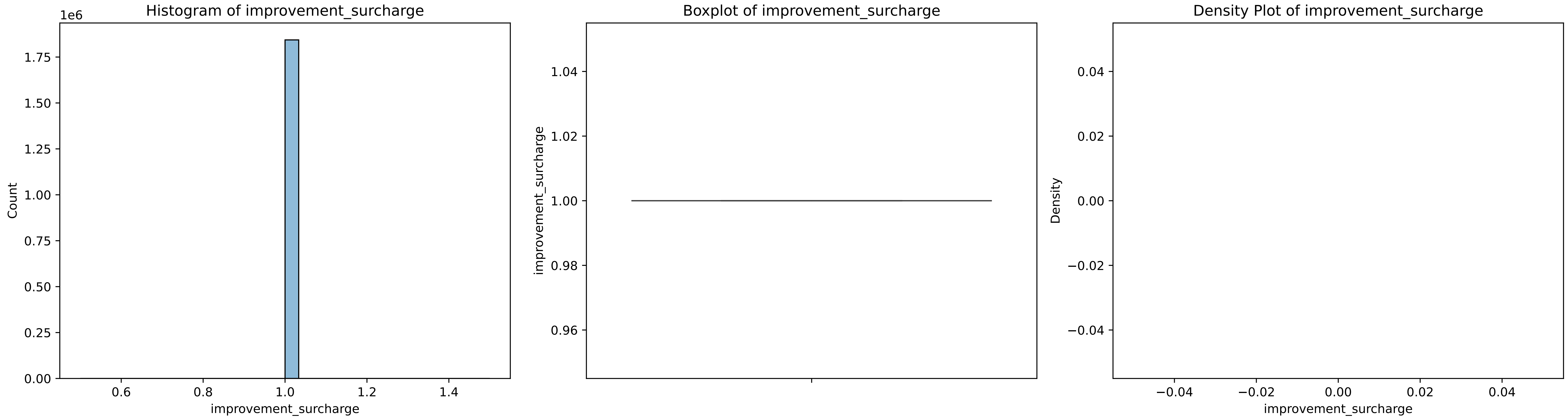

Histogram of mta_tax
1e6
Count
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.2
0.4
0.6
0.8
1.0
mta_tax

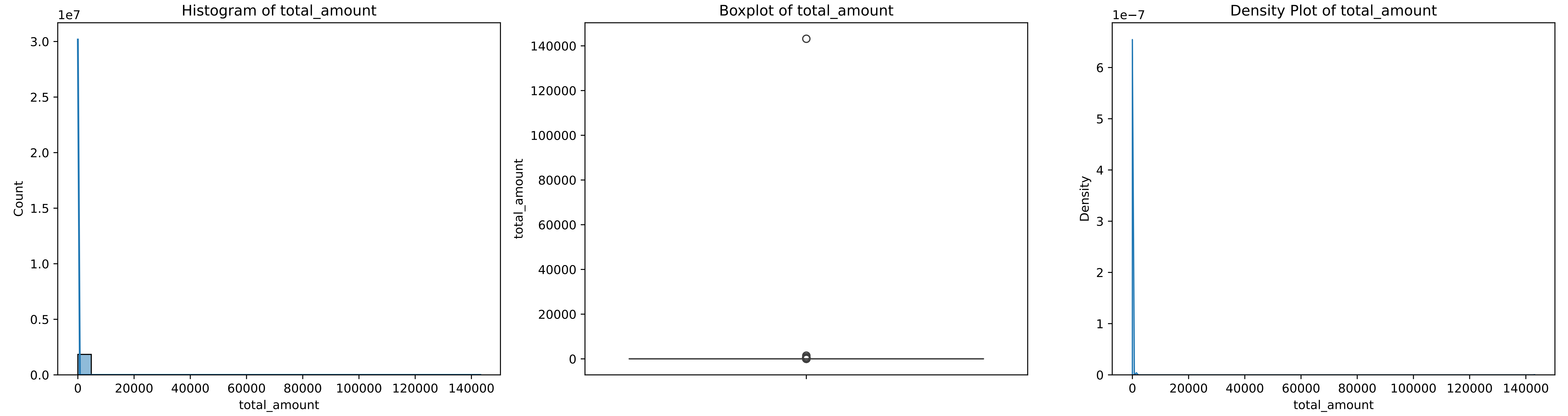

Boxplot of mta_tax
mta_tax
0.48
0.49
0.50
0.51
0.52

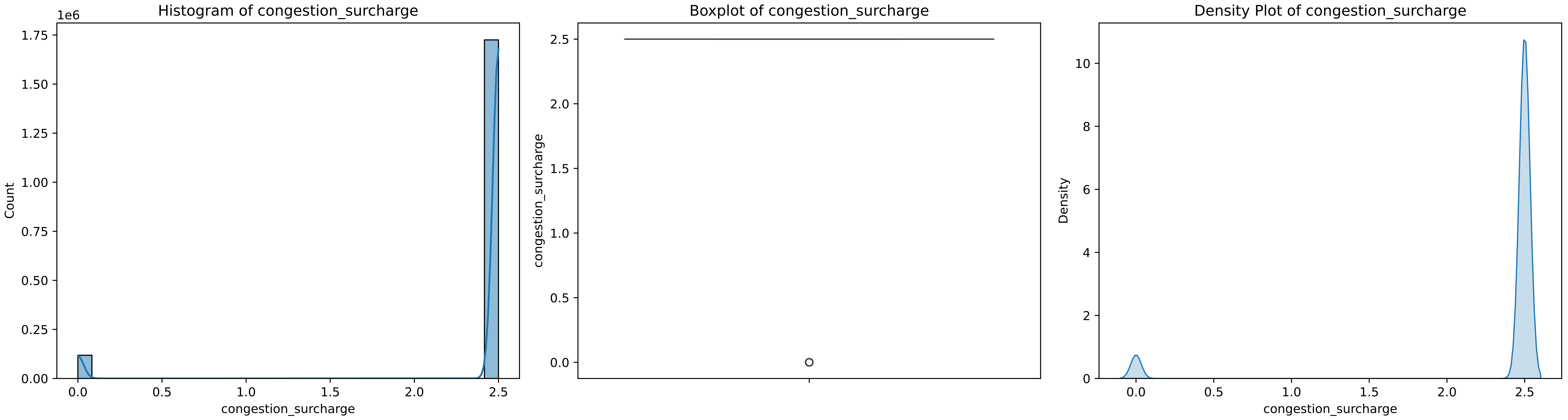

Density Plot of mta_tax
Density
−0.04
−0.02
0.00
0.02
0.04
−0.04
−0.02
0.00
0.02
0.04
mta_tax

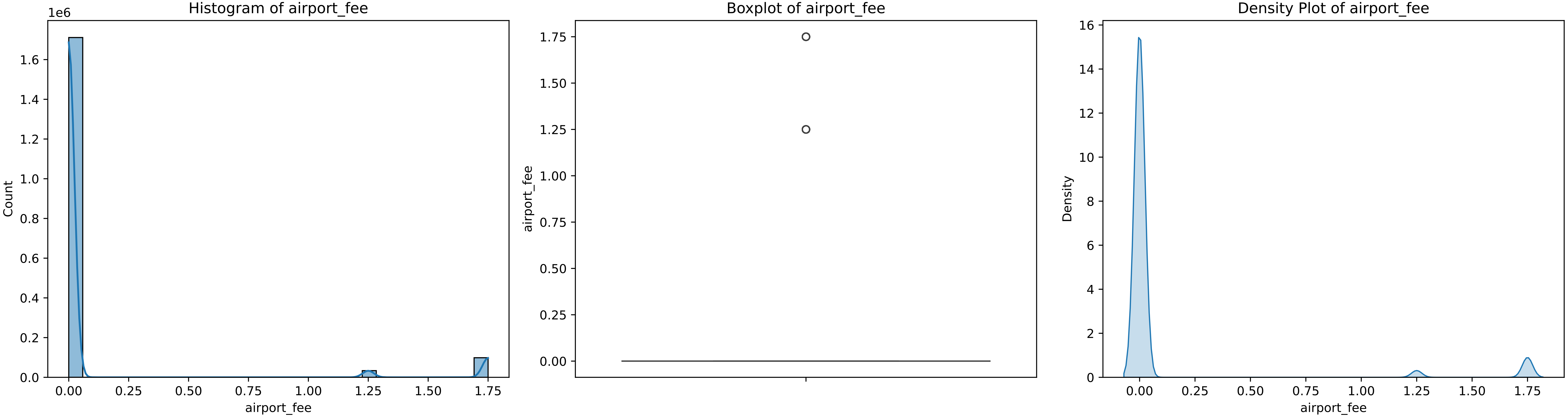

Histogram of tip_amount
1e6
Count
6
5
4
3
2
1
0
0
50
100
150
200
tip_amount
Boxplot of tip_amount
tip_amount
200
150
100
50
0
Density Plot of tip_amount
Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
50
100
150
200
tip_amount


Histogram of tolls_amount
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
0
20
40
60
80
100
120
140
tolls_amount
Boxplot of tolls_amount
140
120
100
80
60
40
20
0
tolls_amount
Density Plot of tolls_amount
0.10
0.08
0.06
0.04
0.02
0.00
Density
0
20
40
60
80
100
120
140
tolls_amount


Histogram of improvement_surcharge
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0.6
0.8
1.0
1.2
1.4
improvement_surcharge
Boxplot of improvement_surcharge
1.04
1.02
1.00
0.98
0.96
improvement_surcharge
Density Plot of improvement_surcharge
0.04
0.02
0.00
−0.02
−0.04
Density
−0.04
−0.02
0.00
0.02
0.04
improvement_surcharge


Histogram of total_amount
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
0
20000
40000
60000
80000
100000
120000
140000
total_amount
Boxplot of total_amount
140000
120000
100000
80000
60000
40000
20000
0
total_amount
Density Plot of total_amount
1e−7
6
5
4
3
2
1
0
Density
0
20000
40000
60000
80000
100000
120000
140000
total_amount


Histogram of congestion_surcharge
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
0.0
0.5
1.0
1.5
2.0
2.5
congestion_surcharge
Boxplot of congestion_surcharge
2.5
2.0
1.5
1.0
0.5
0.0
congestion_surcharge
Density Plot of congestion_surcharge
10
8
6
4
2
0
Density
0.0
0.5
1.0
1.5
2.0
2.5
congestion_surcharge


Histogram of airport_fee

Boxplot of airport_fee

Density Plot of airport_fee

Conclusion/Remarks:

2.3 Analysing the data in Hystogram, Box plot and KDE plot. It looks much better

Conclusion/Remarks:

2.3.1 Remoded records with trip distance 0 or greater than 300. Also if fare amount is 0 but pick up and drop location is different, then it is problematic data

Conclusion/Remarks:

2.3.1 Analysing trip_distance for rows where Payment type =0 (not defined) or 3 (no charge)
1> trip_distance=0 and total_amount<> 0 then no meaning of payment_type.Impute payment_type = 4 (Dispute) 2> trip_distance<> 0 and total_amount<> 0 and tip_amount = 0 then Impute payment_type = 2 (Cash) 3> trip_distance<> 0 and total_amount<> 0 and tip_amount <> 0 then Impute payment_type = 1 (Credit card) from Mode) - drop these records

Conclusion/Remarks:

2.3.1 Marking the trips as disputed(4) where trip_distance = 0 and total_amount != 0

Conclusion/Remarks:

2.3.1 Imputing the payment_type as Cash(2) where tip_amount = 0 and total_amount != 0, as trip_amount is not applicable in Cash Payment

Conclusion/Remarks:

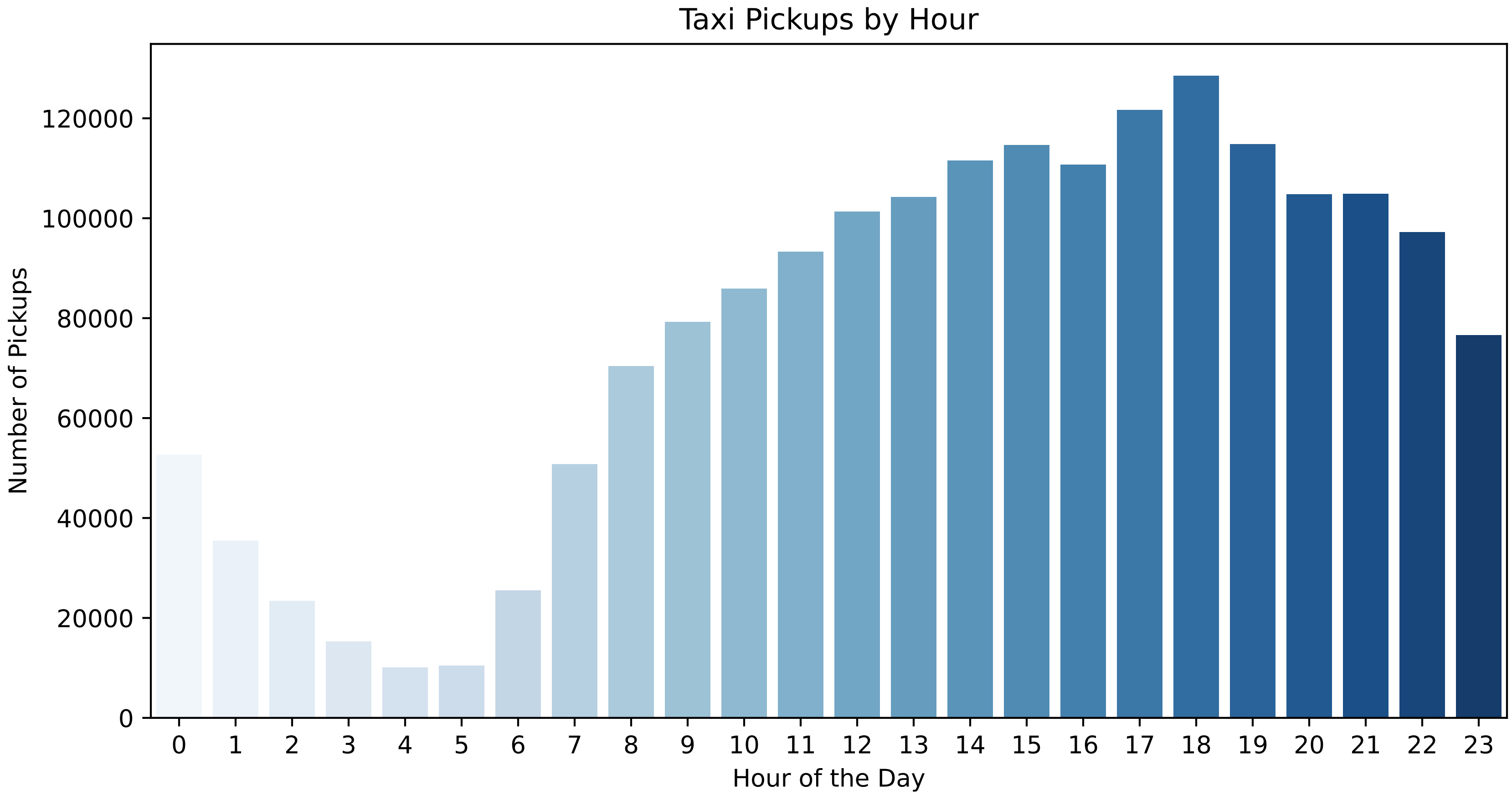
2.3.1 Imputing the payment_type as Credit Card(1) where tip_amount != 0 and total_amount != 0

Conclusion/Remarks:

2.3.1 Dropping some records based on above 3 conditions, passenger count > 6 and

Conclusion/Remarks:

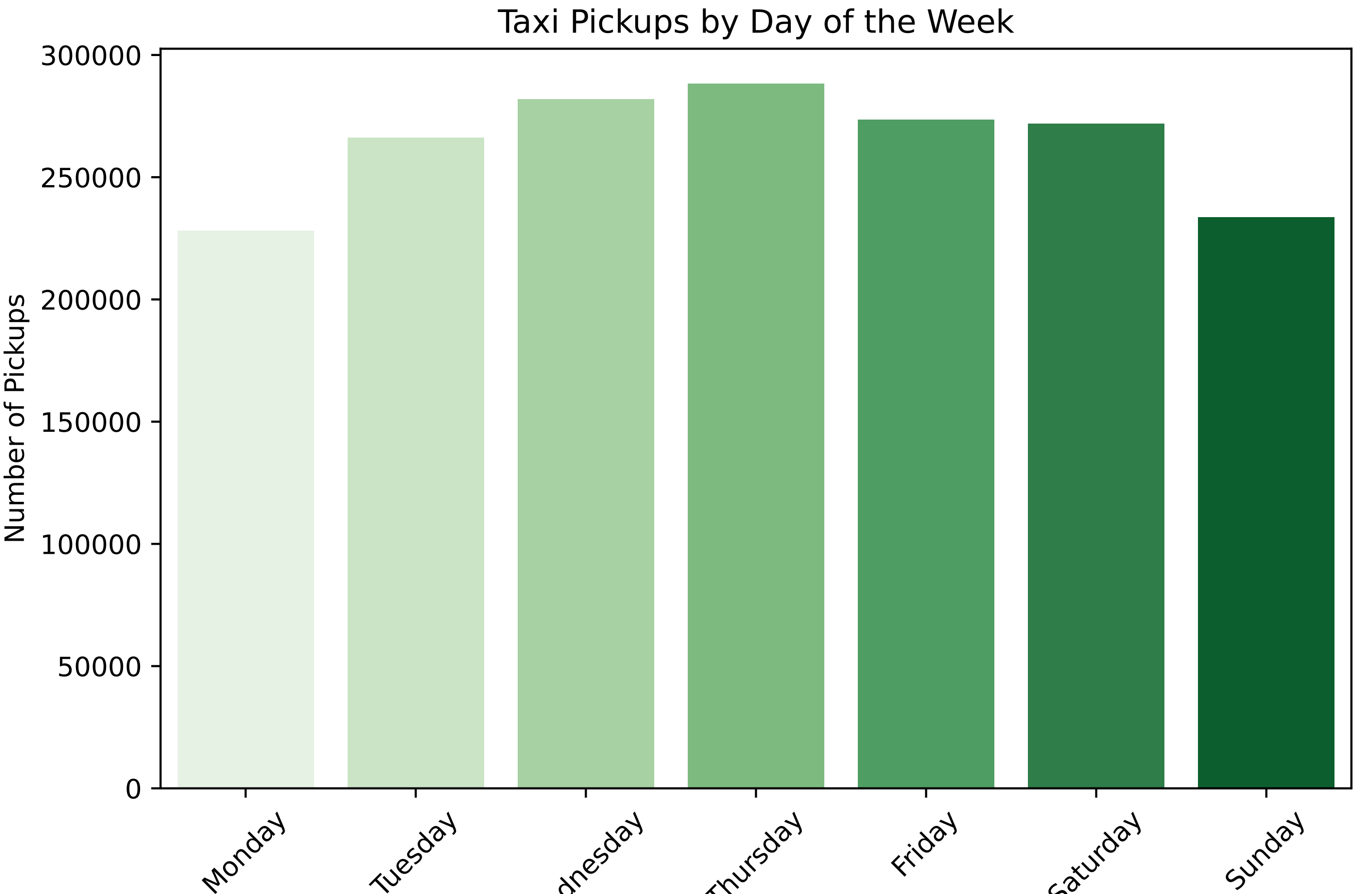
3.1.2 Analysing pickup count by month, day and year wise. in 2nd June 18:00 PM was the maximum pickup



Taxi Pickups by Hour
Number of Pickups
0
20000
40000
60000
80000
100000
120000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Hour of the Day

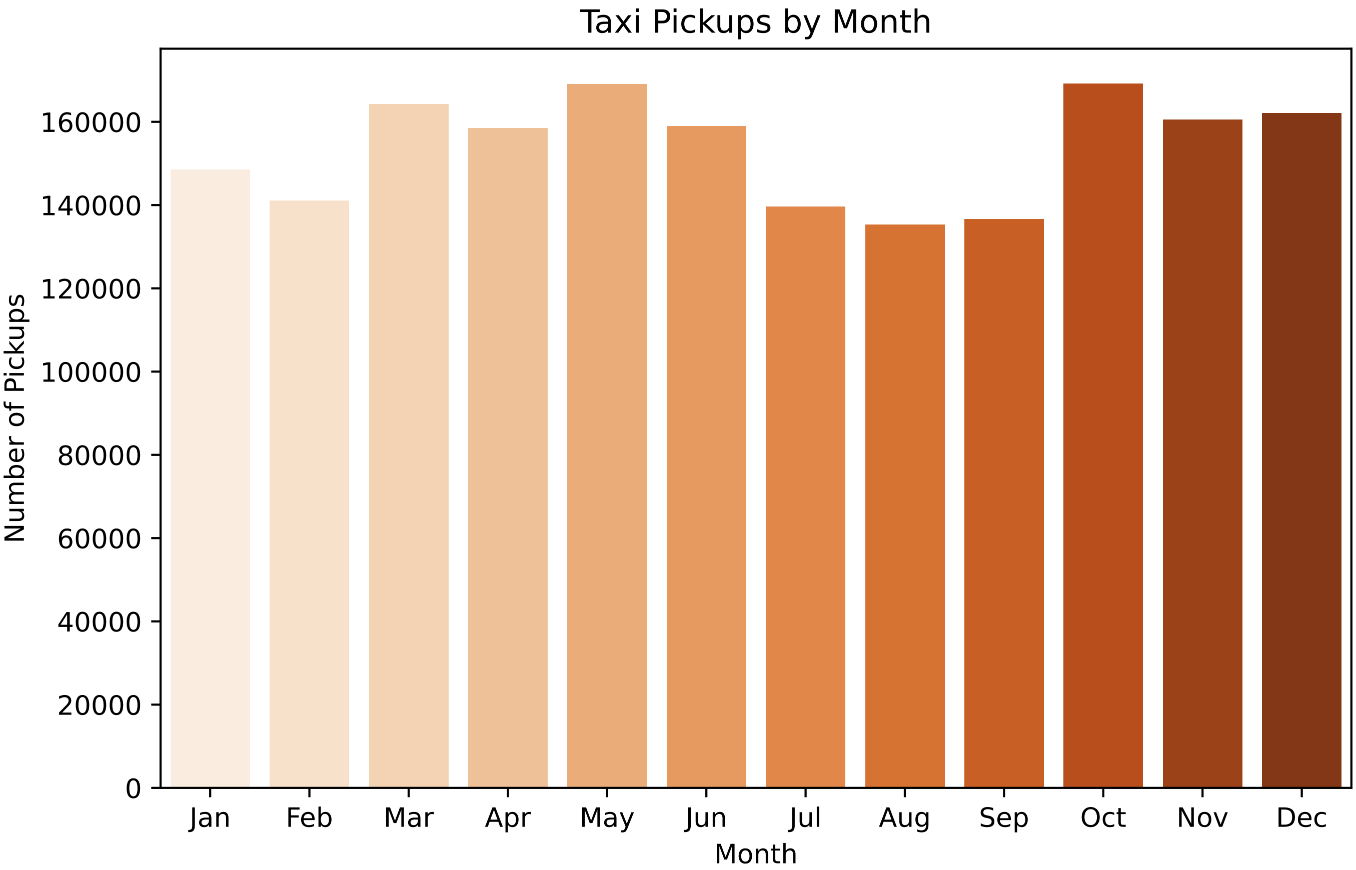
Conclusion/Remarks:

3.1.2 Analysing pickup Hour Wise. Around 18:00 PM is maximum pickup


Taxi Pickups by Day of the Week
Number of Pickups
300000
250000
200000
150000
100000
50000
0
Monday
Tuesday
dnesday
Thursday
Friday
Saturday
Sunday

Conclusion/Remarks:

3.1.2 Analysing pickup Day Wise. Around Thursday is maximum pickup


Taxi Pickups by Month
Number of Pickups
0
20000
40000
60000
80000
100000
120000
140000
160000
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Conclusion/Remarks:

3.1.2 Analysing pickup Month Wise. In the month of October is maximum pickup

Conclusion/Remarks:

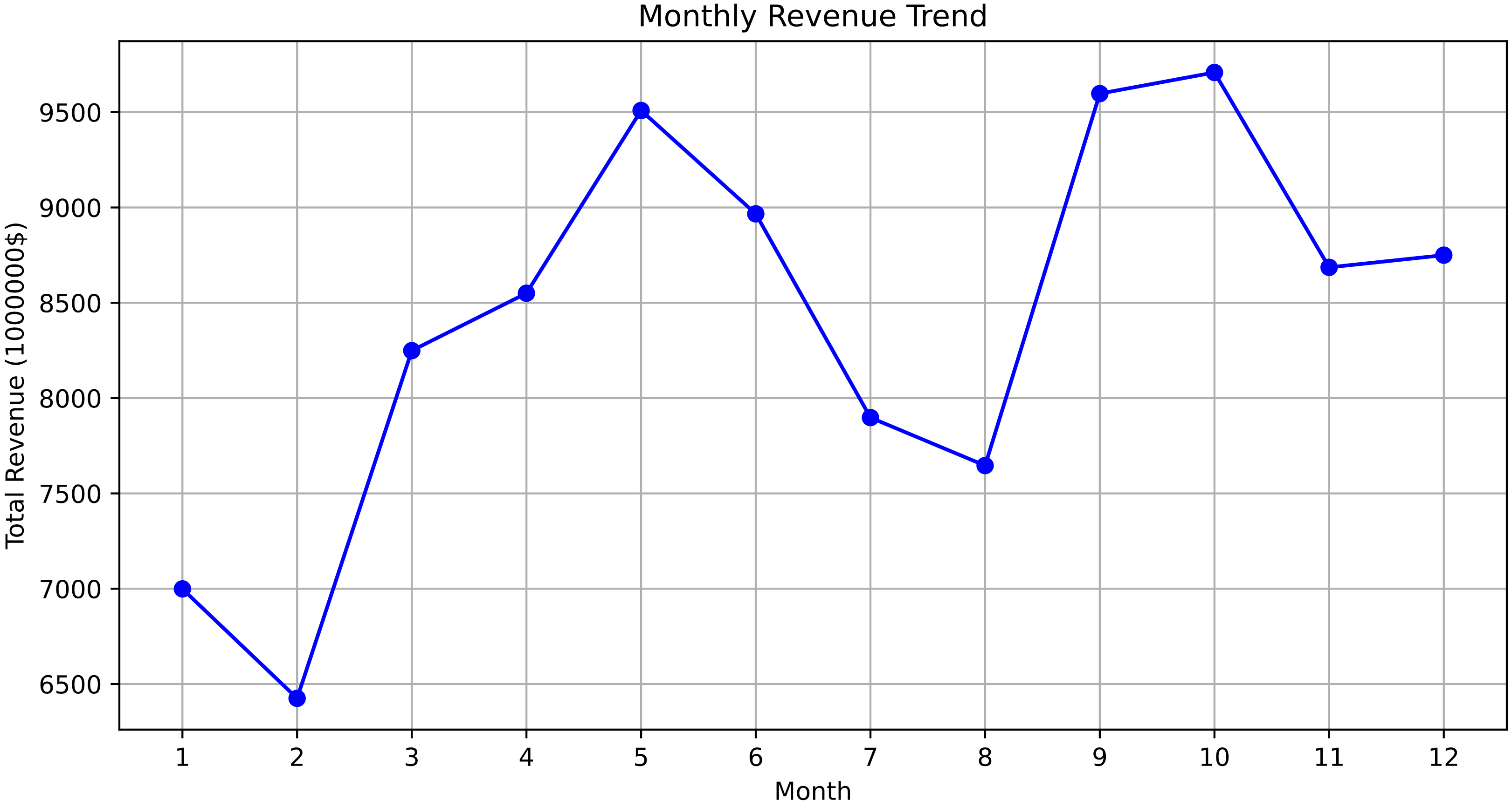
3.1.3 Analysing Negative value count for fare_amount(1284311), tip_amount(1143079), total_amount(1304872), trip_distance(1447231)

Conclusion/Remarks:

3.1.3 First checking the record counts for the amount >= 0

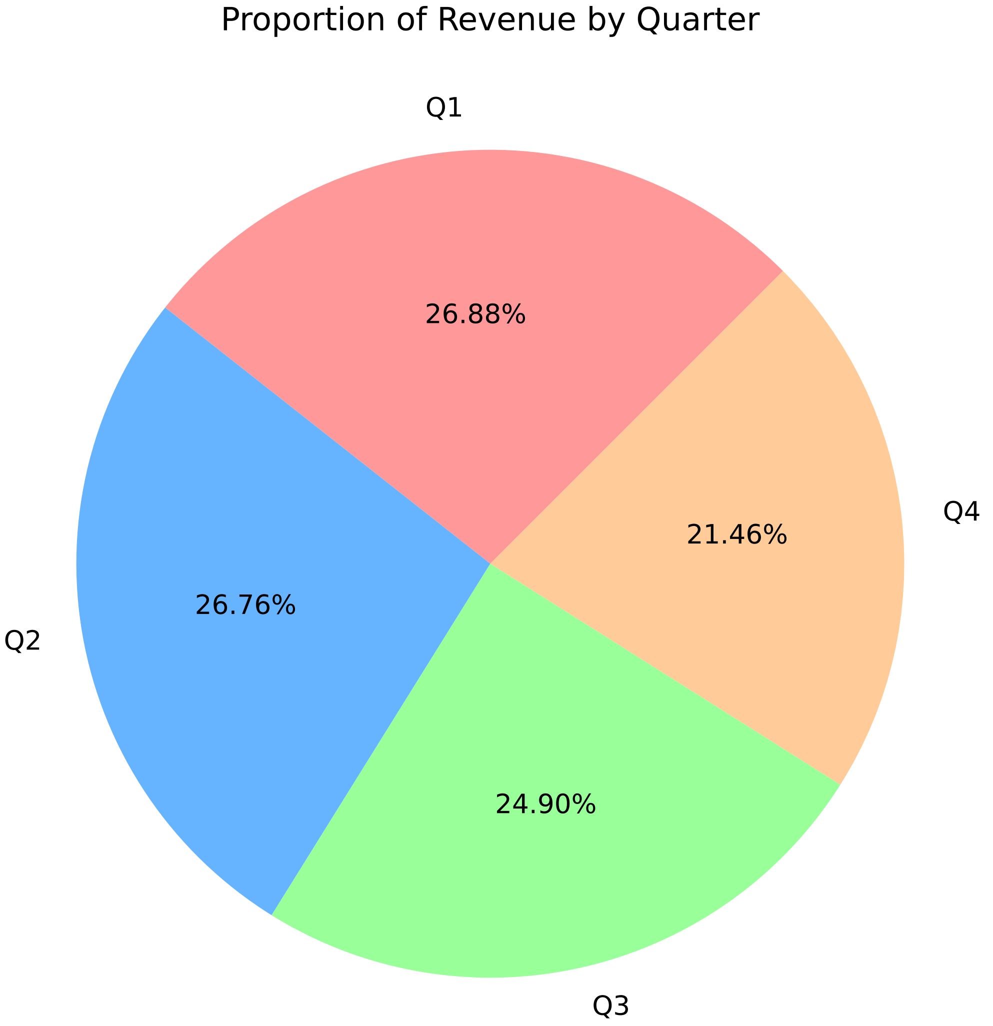
Conclusion/Remarks:

3.1.3 Cleaned the data, Only for the amount > 0 retained


Monthly Revenue Trend
Total Revenue (1000000$)
9500
9000
8500
8000
7500
7000
6500
1
2
3
4
5
6
7
8
9
10
11
12
Month

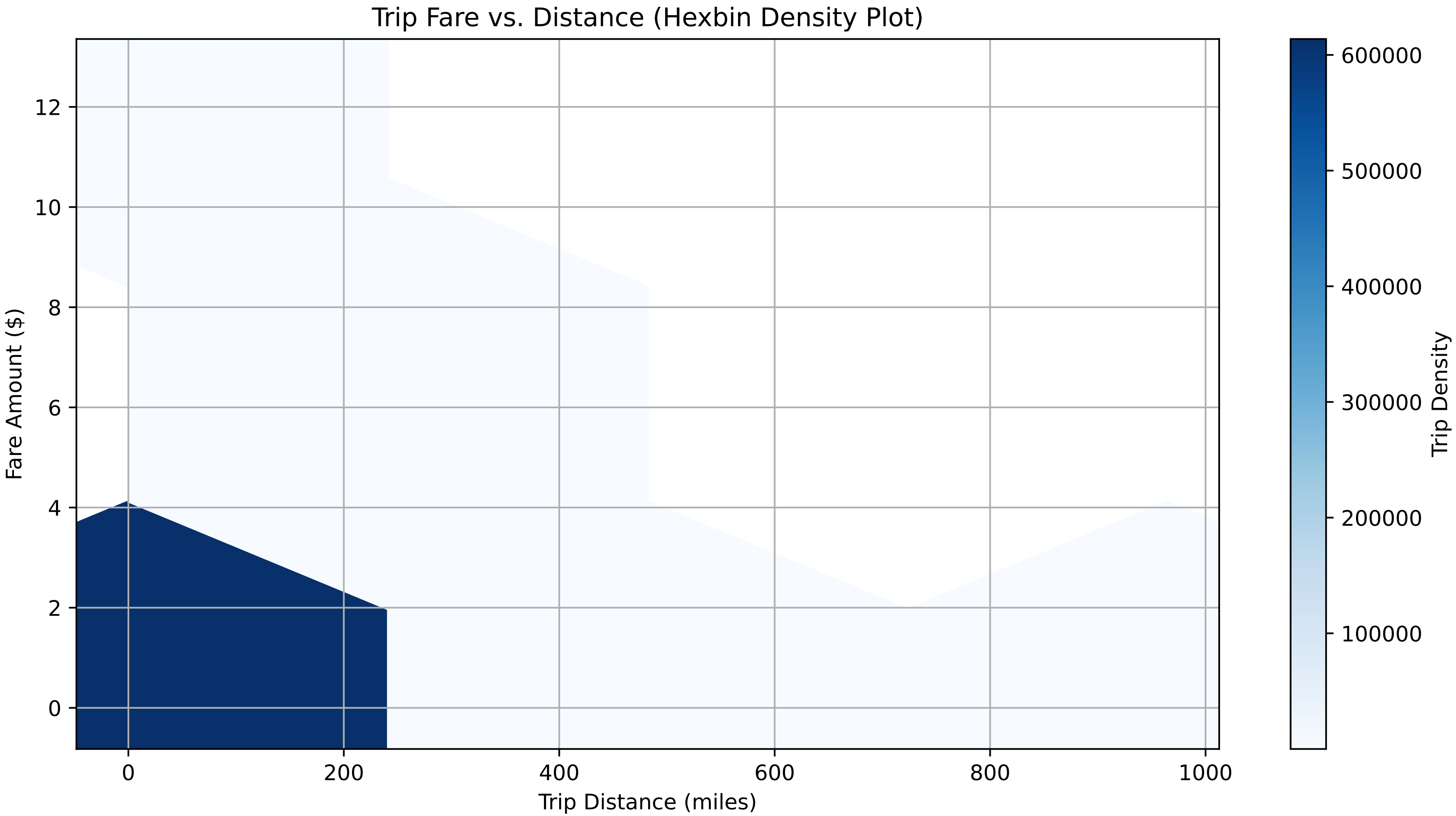
Conclusion/Remarks:

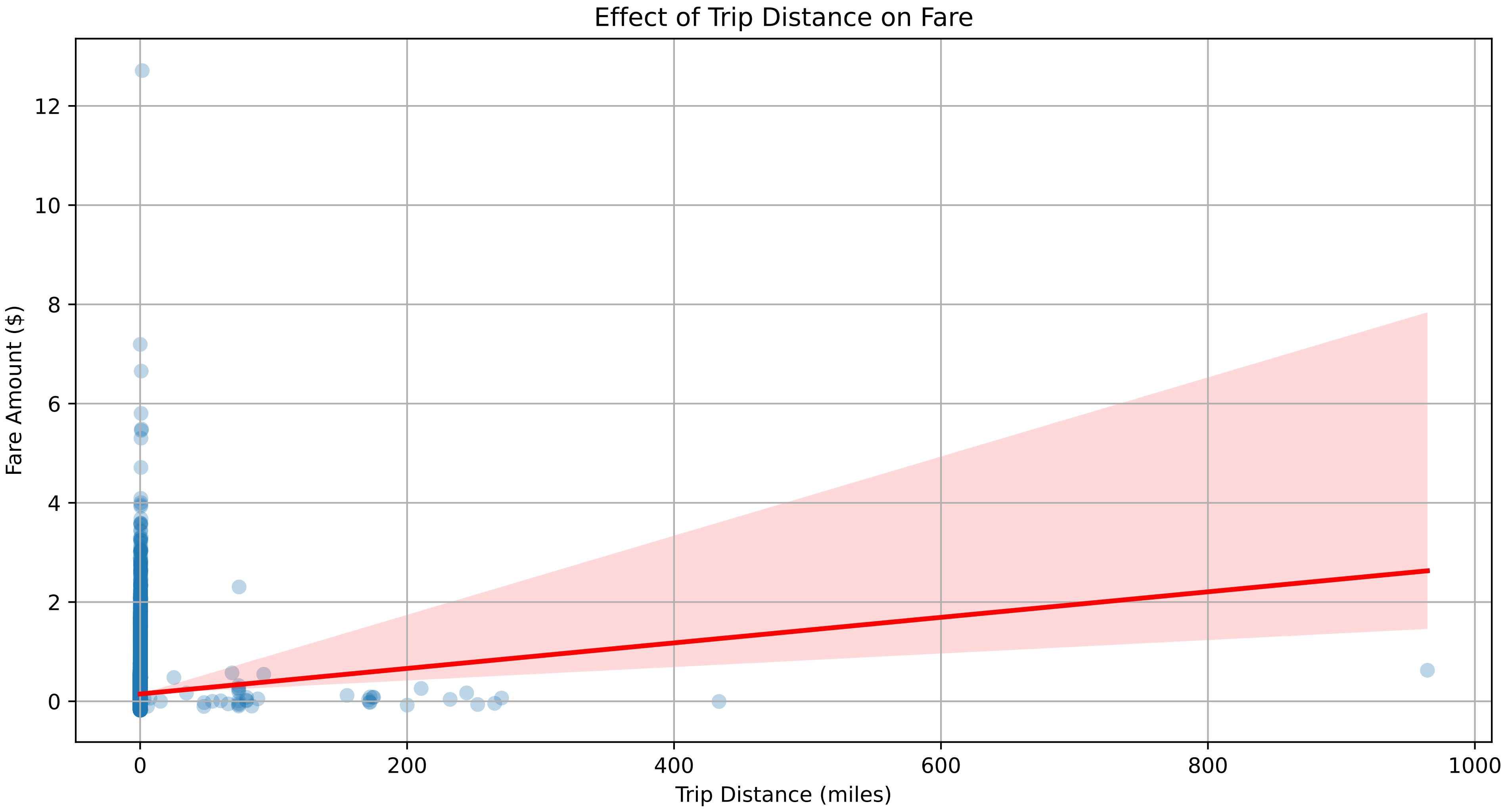
3.1.4 Analysing monthly revenue. October is the heighest with count 4765013


Proportion of Revenue by Quarter
Q1
26.88%
Q4
21.46%
Q2
26.76%
24.90%
Q3

Conclusion/Remarks:

3.1.5 Analysing Quarterly revenue. Q2 is the heighest with 26.960258% of yearly revenue


Trip Fare vs. Distance (Hexbin Density Plot)
Fare Amount ($)
12
10
8
6
4
2
0
0
200
400
600
800
1000
Trip Distance (miles)
600000
500000
400000
300000
200000
100000
Trip Density


Effect of Trip Distance on Fare
Fare Amount ($)
Trip Distance (miles)
0
2
4
6
8
10
12
0
200
400
600
800
1000

Conclusion/Remarks:

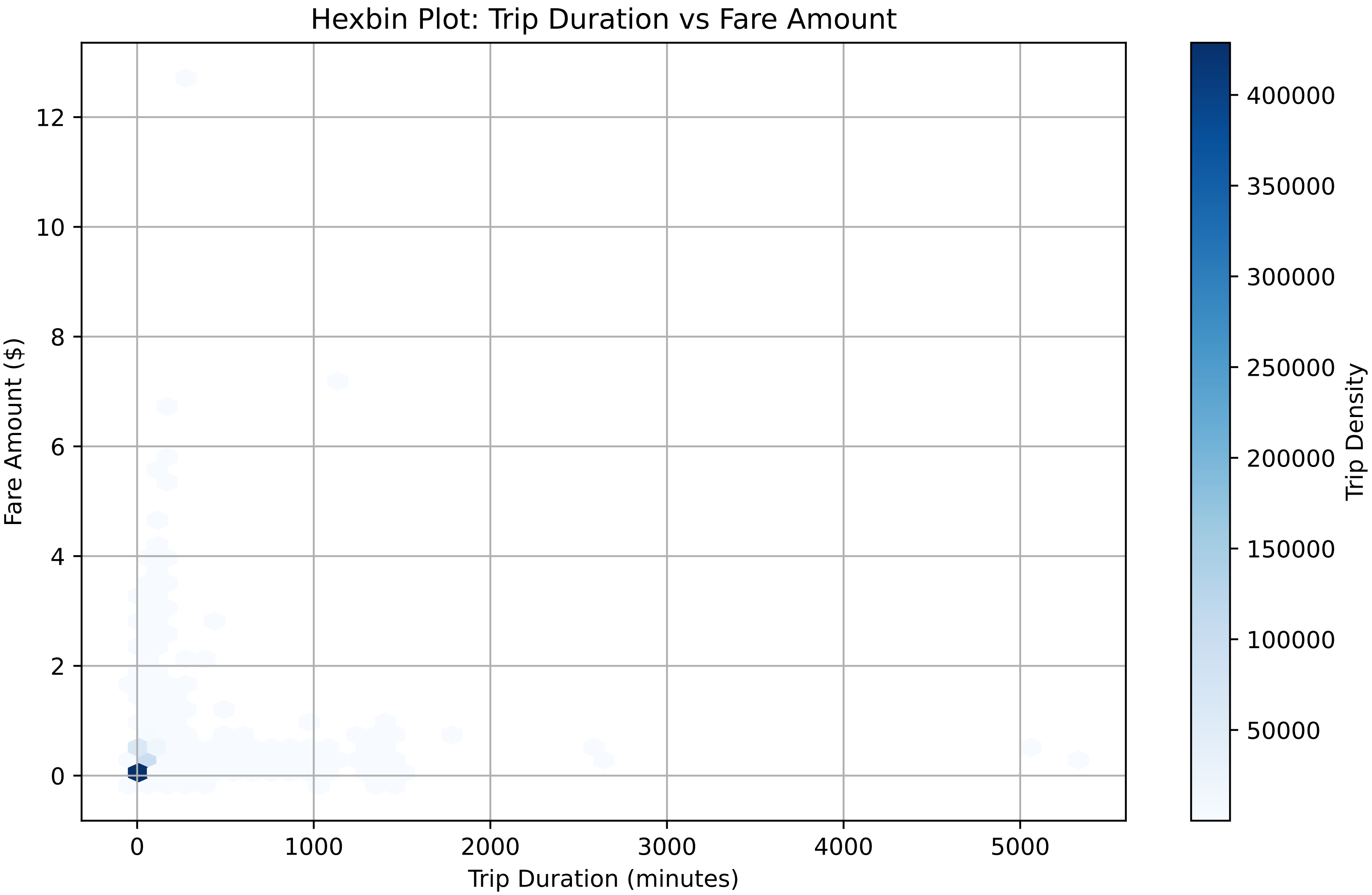
3.1.6 Analysing Showing Trip Distance Vs. Fare Amount. It is a line with small slope

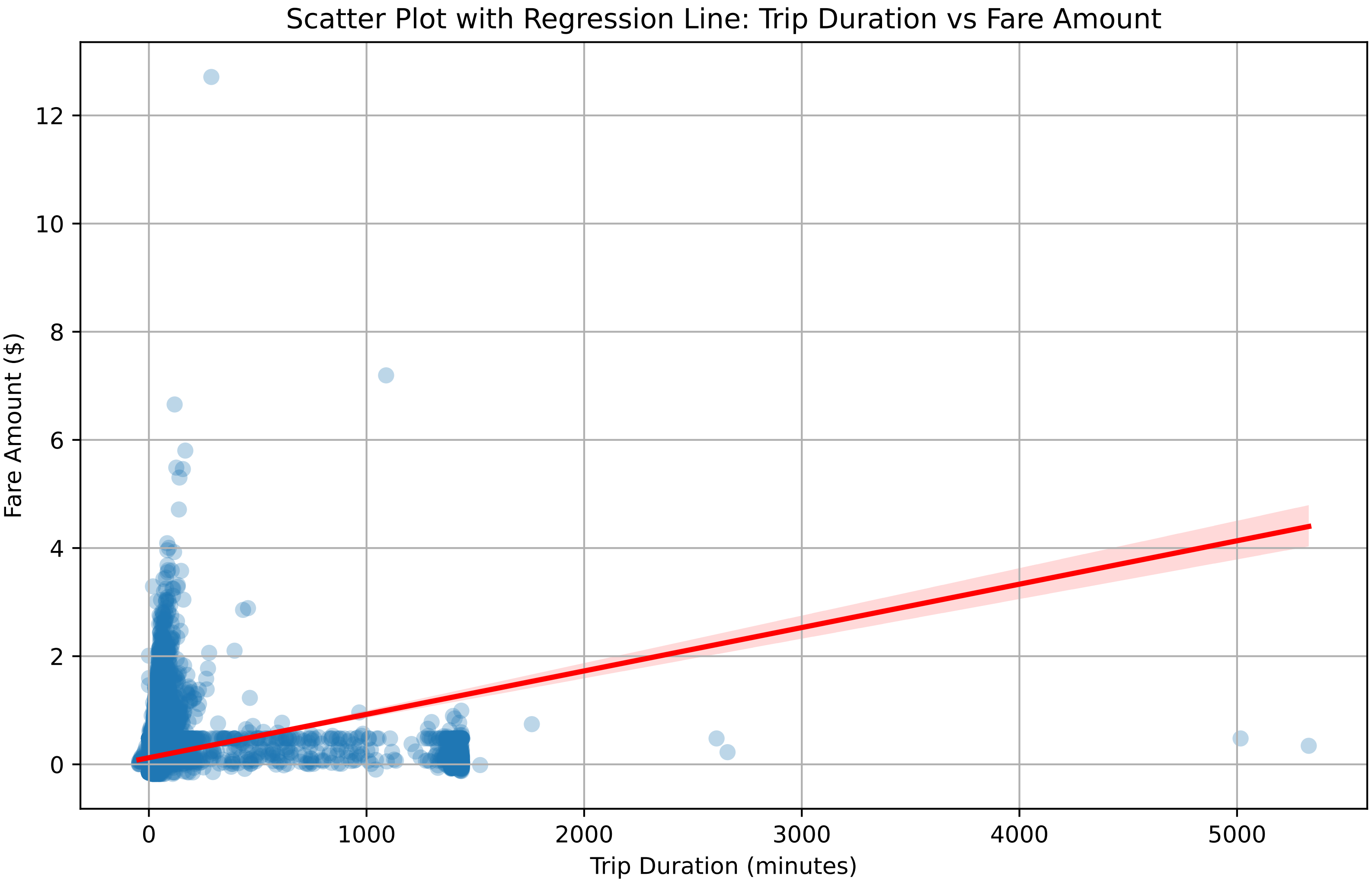
Conclusion/Remarks:

3.1.7 Added new column trip_duration = (drop time - pick time)/60

Conclusion/Remarks:

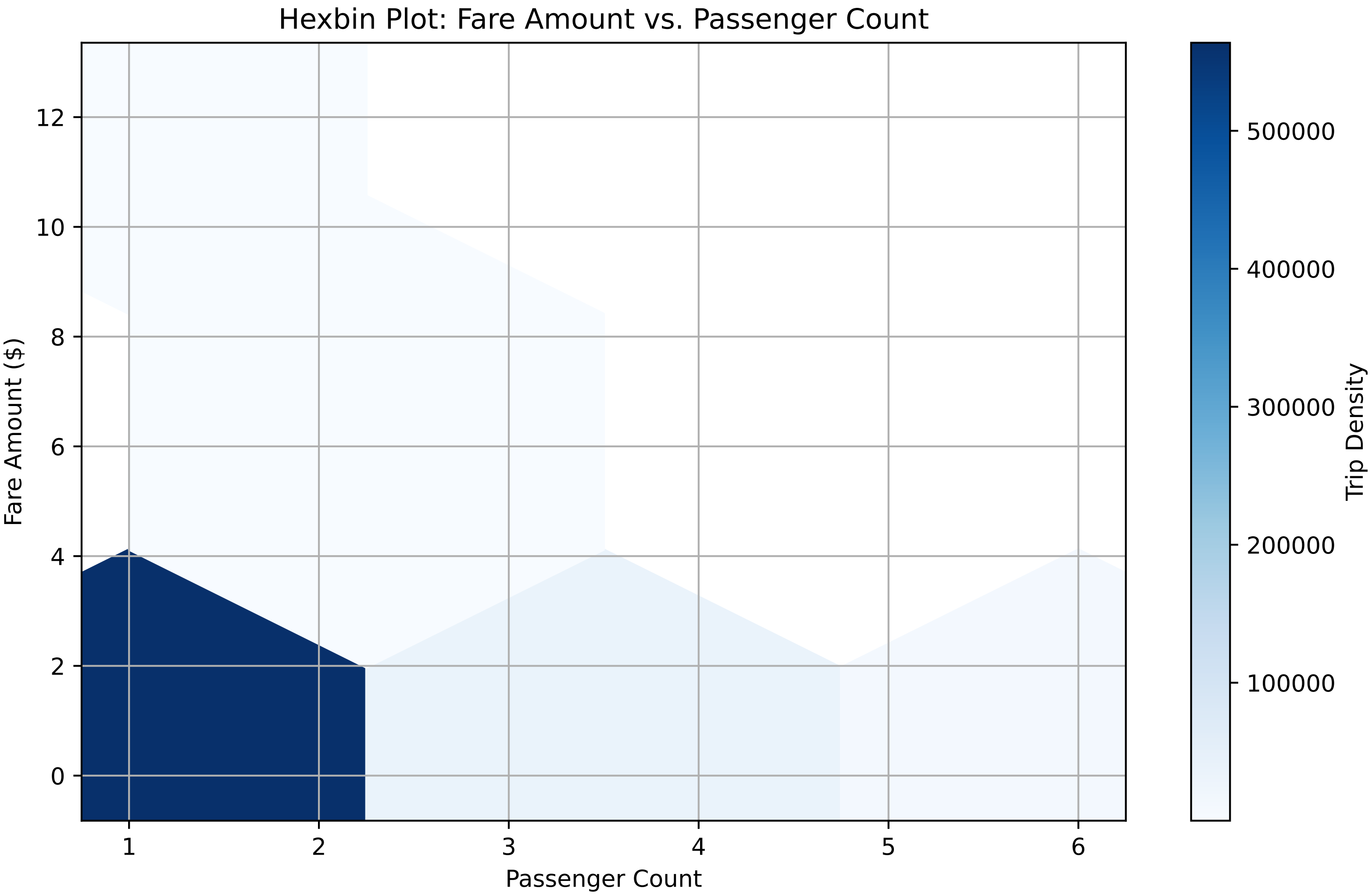
3.1.7 Visualization 1> fare_amount and trip duration 2> fare_amount and passenger_count 3> tip_amount and trip_distance By Scatter, Line and Regression Plot

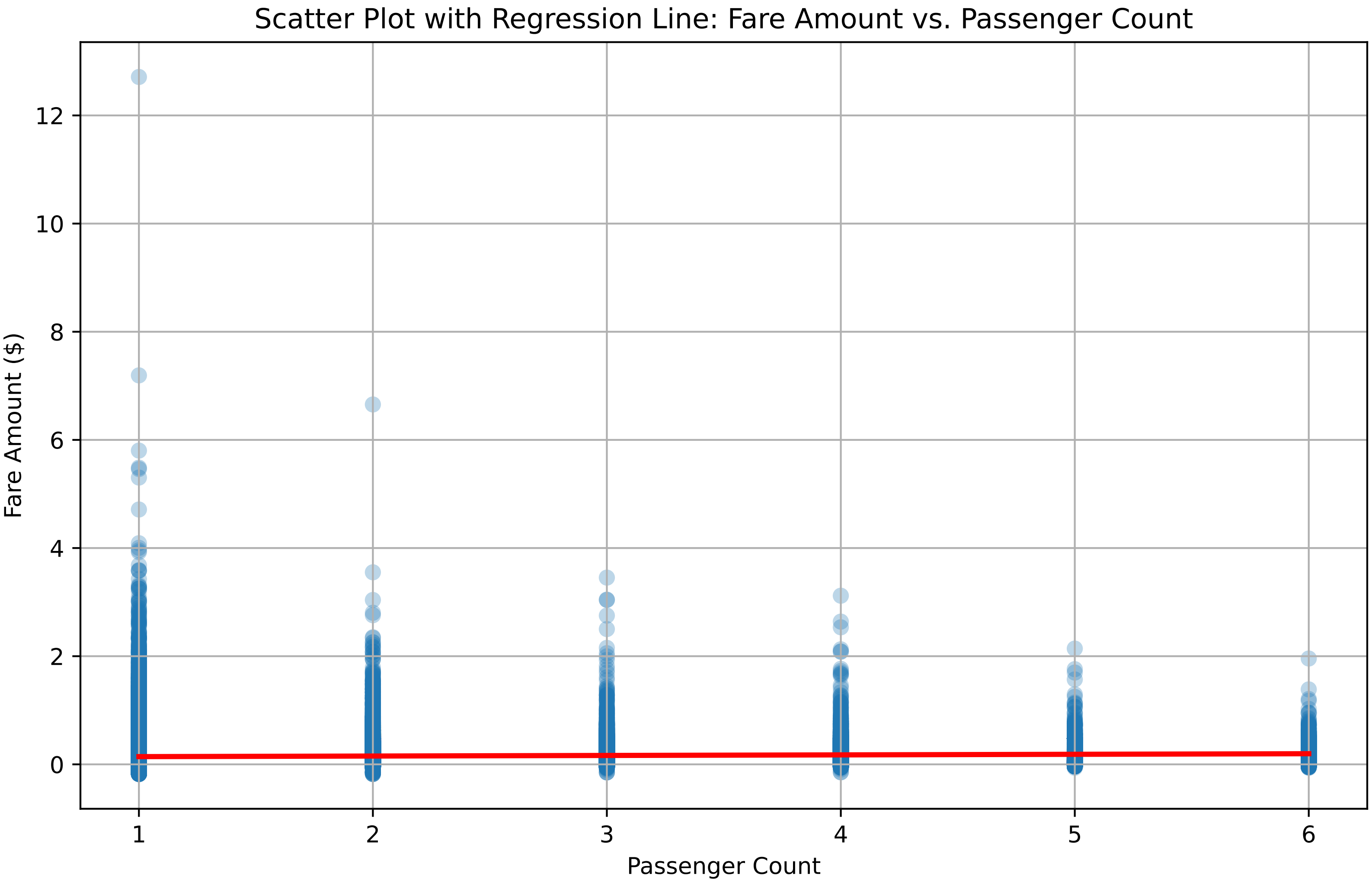

Hexbin Plot: Trip Duration vs Fare Amount
Fare Amount ($)
0
2
4
6
8
10
12
0
1000
2000
3000
4000
5000
Trip Duration (minutes)
Trip Density
50000
100000
150000
200000
250000
300000
350000
400000


Scatter Plot with Regression Line: Trip Duration vs Fare Amount
Fare Amount ($)
12
10
8
6
4
2
0
0
1000
2000
3000
4000
5000
Trip Duration (minutes)

Conclusion/Remarks:

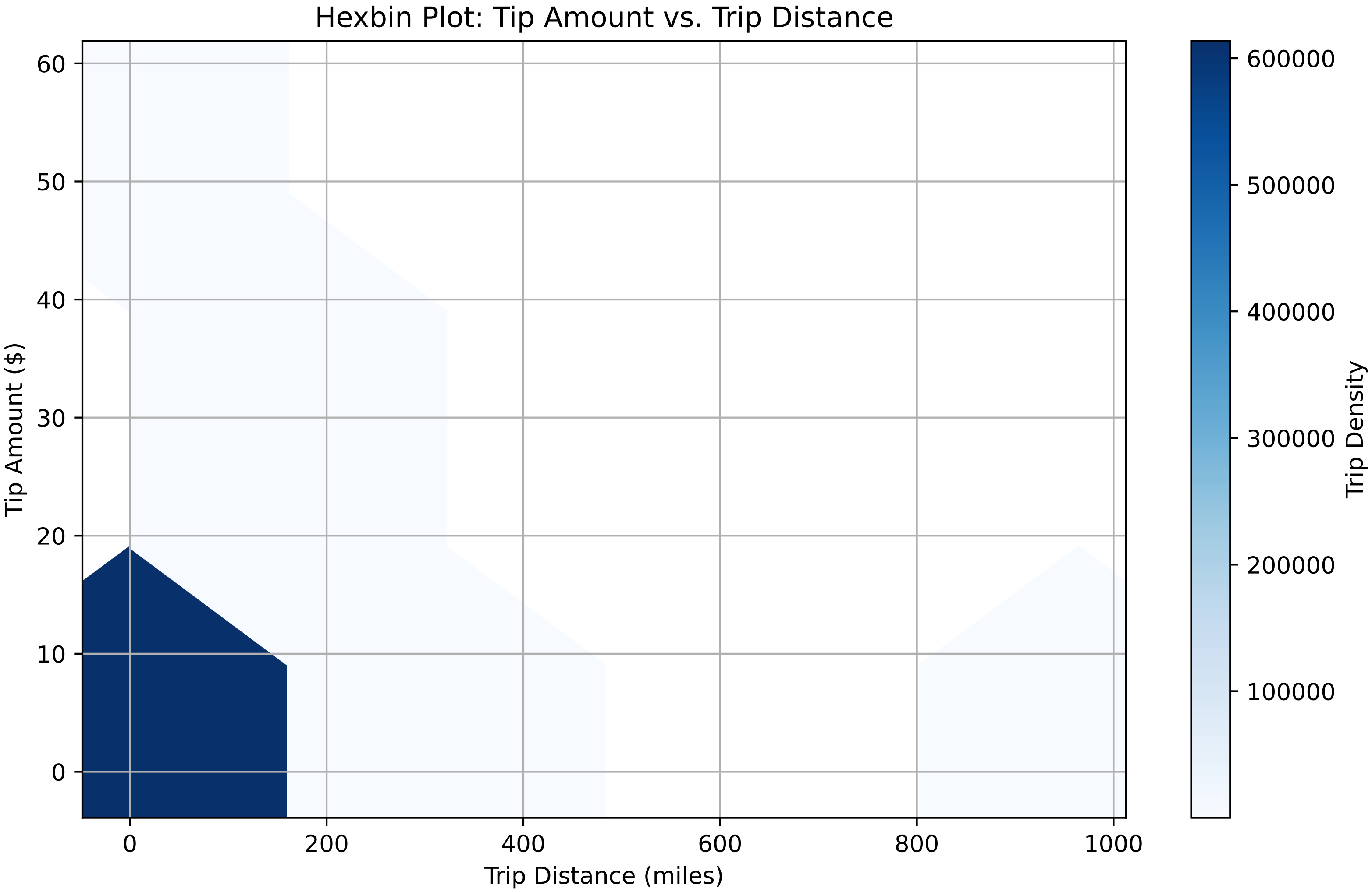
3.1.6 Visualization 1> trip_duration and fare_amount. It is interestingly fare_amount does not increase with trip_duration

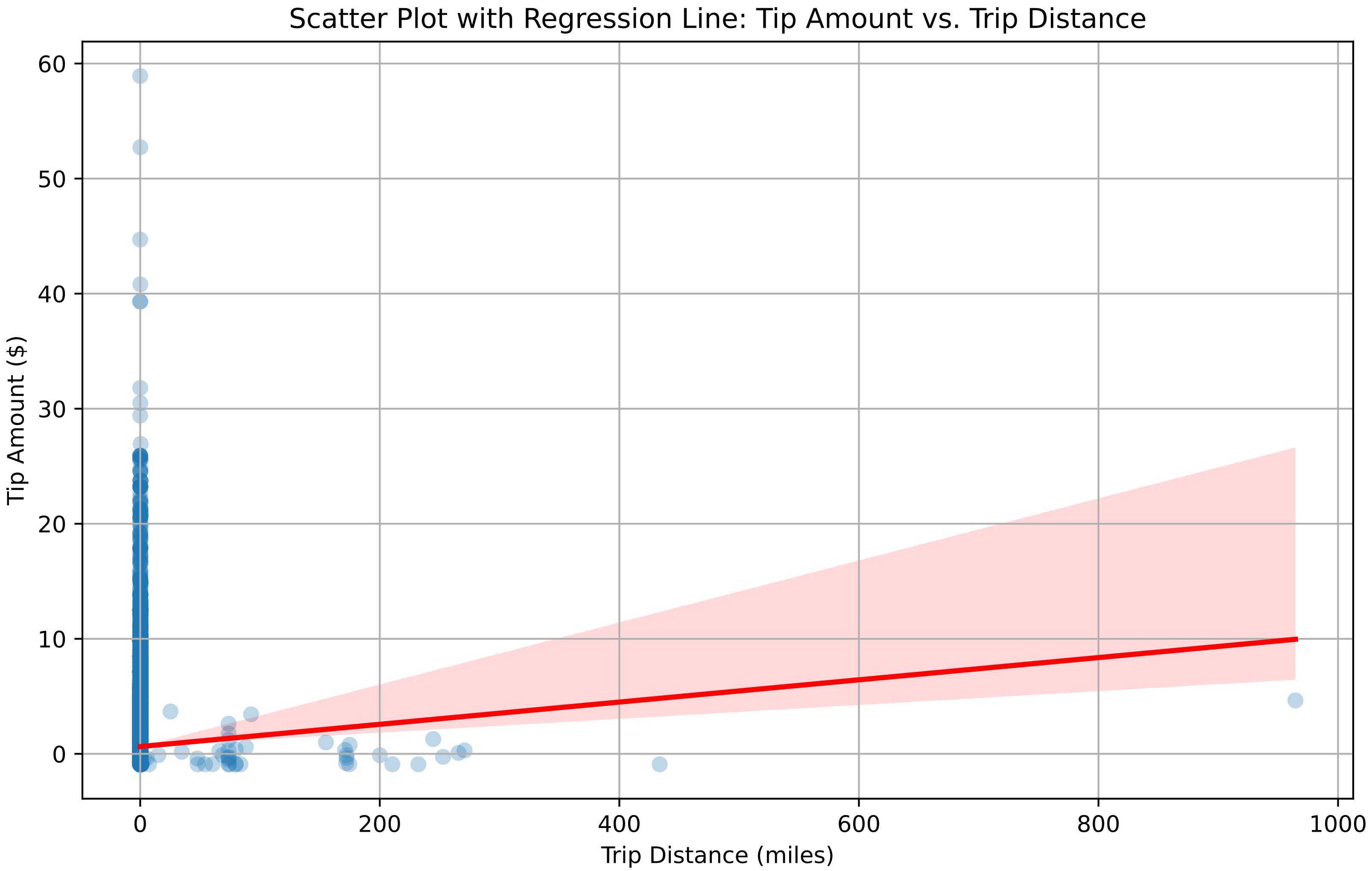

Hexbin Plot: Fare Amount vs. Passenger Count
Fare Amount ($)
12
10
8
6
4
2
0
1
2
3
4
5
6
Passenger Count
Trip Density
500000
400000
300000
200000
100000


Scatter Plot with Regression Line: Fare Amount vs. Passenger Count
Fare Amount ($)
0
2
4
6
8
10
12
Passenger Count
1
2
3
4
5
6

Conclusion/Remarks:

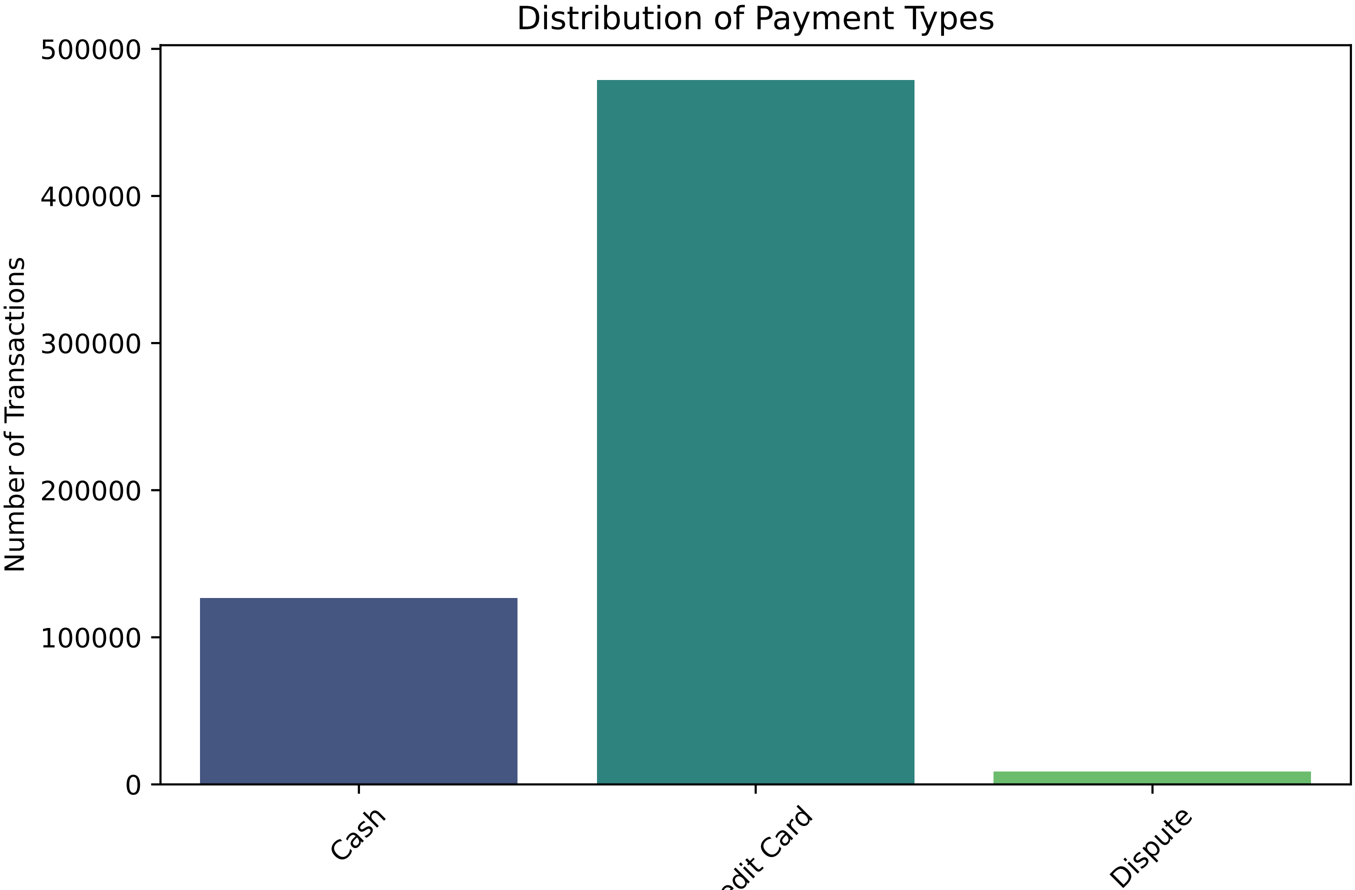
3.1.6 Visualization 2> fare_amount and passenger_count. It is interestingly does not increase with passerger count


Hexbin Plot: Tip Amount vs. Trip Distance
Tip Amount ($)
60
50
40
30
20
10
0
0
200
400
600
800
1000
Trip Distance (miles)
Trip Density
600000
500000
400000
300000
200000
100000


Scatter Plot with Regression Line: Tip Amount vs. Trip Distance
Tip Amount ($)
60
50
40
30
20
10
0
0
200
400
600
800
1000
Trip Distance (miles)

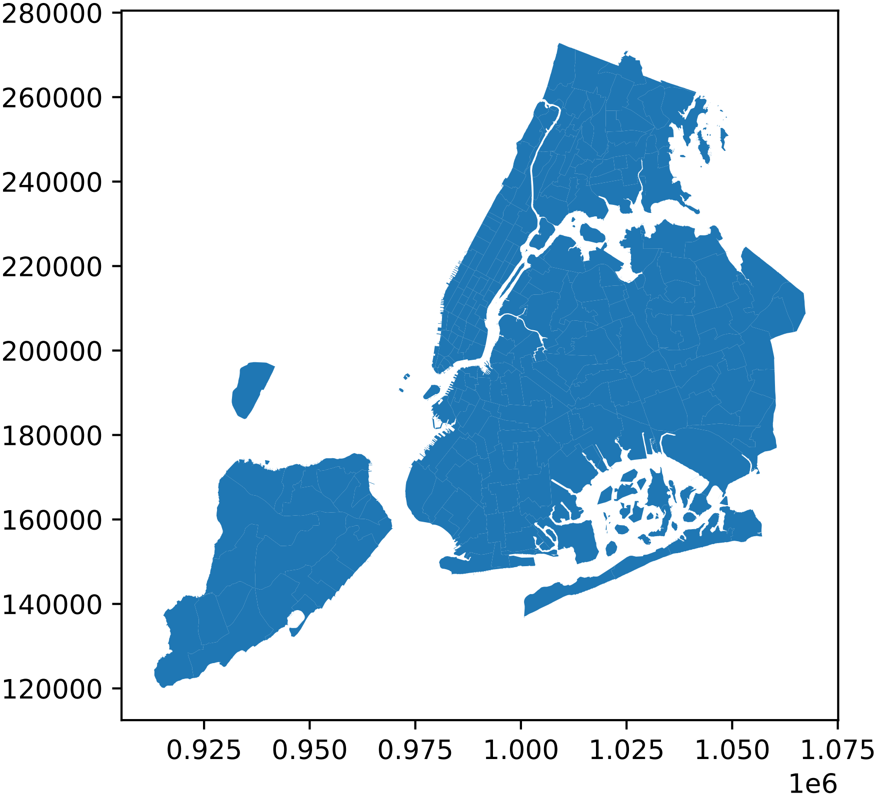
Conclusion/Remarks:

3.1.6 Visualization 3> tip_amount and trip_distance. Interestingly the top fares for very close distance


Distribution of Payment Types
Number of Transactions
500000
400000
300000
200000
100000
0
Cash
edit Card
Dispute

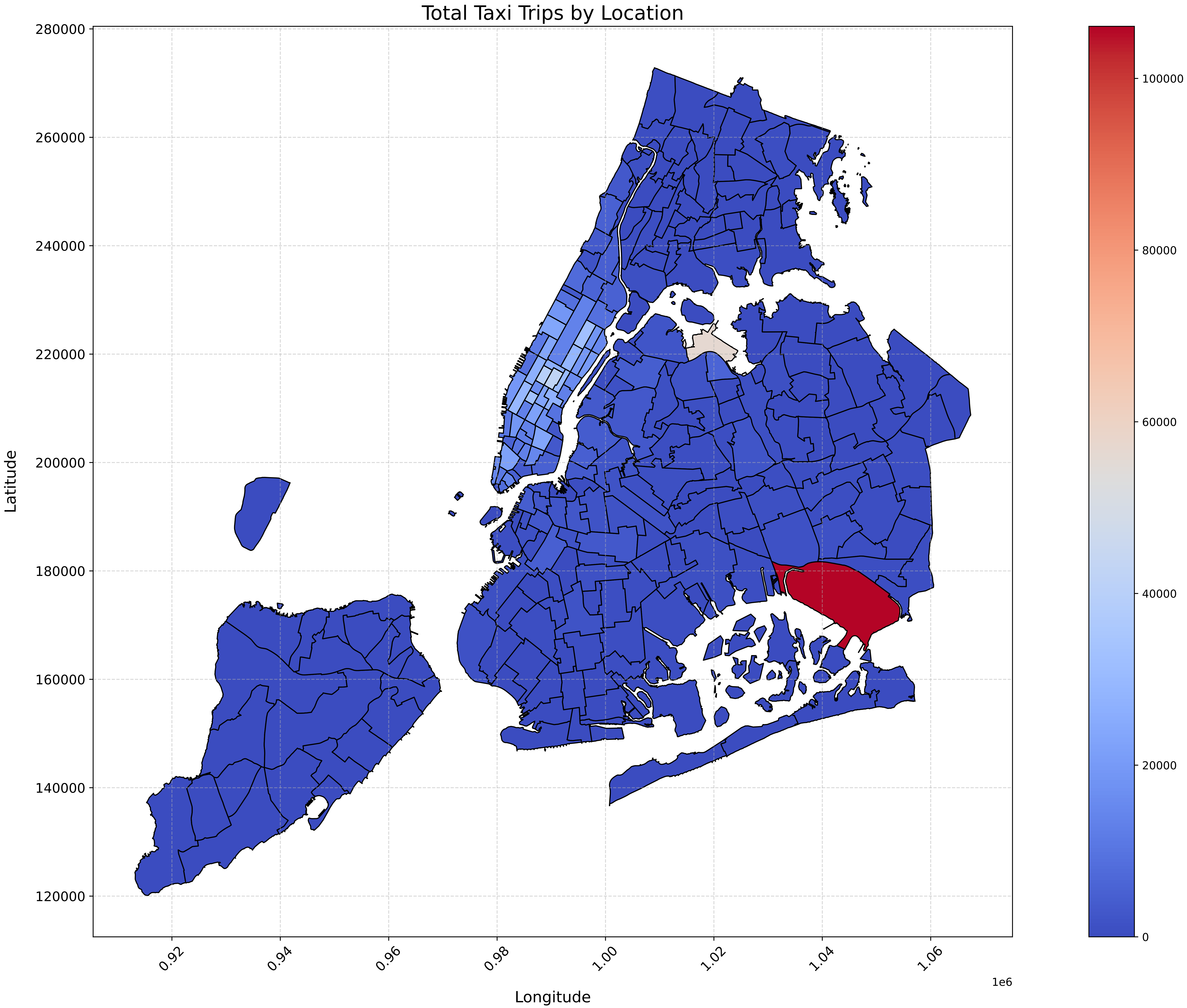
Conclusion/Remarks:

3.1.6 Visualization: Bar Plot of Payment Type Distribution -> Credit Card is of Most use


280000
260000
240000
220000
200000
180000
160000
140000
120000
0.925
0.950
0.975
1.000
1.025
1.050
1.075
1e6

Conclusion/Remarks:

3.1.11 Group data by location IDs to find the total number of trips per location ID. It is showing top 10 in total trip count



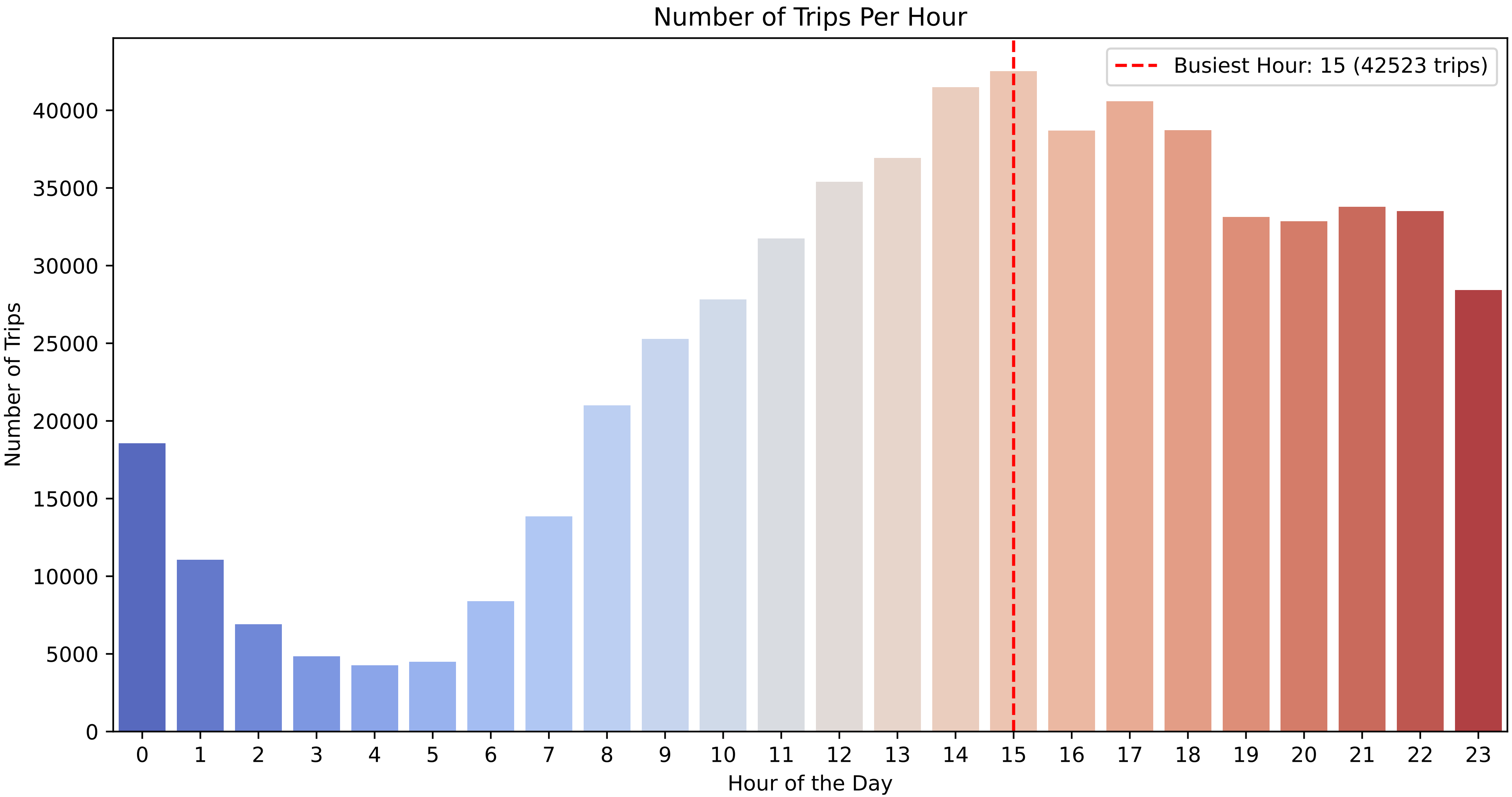
Total Taxi Trips by Location

Conclusion/Remarks:

3.1.13 Plot a color-coded map showing zone-wise trips

Conclusion/Remarks:

3.2.1 Showing slow routes overall and hour basis. We can increate the surcharge in those routes and hours to make more revenue.

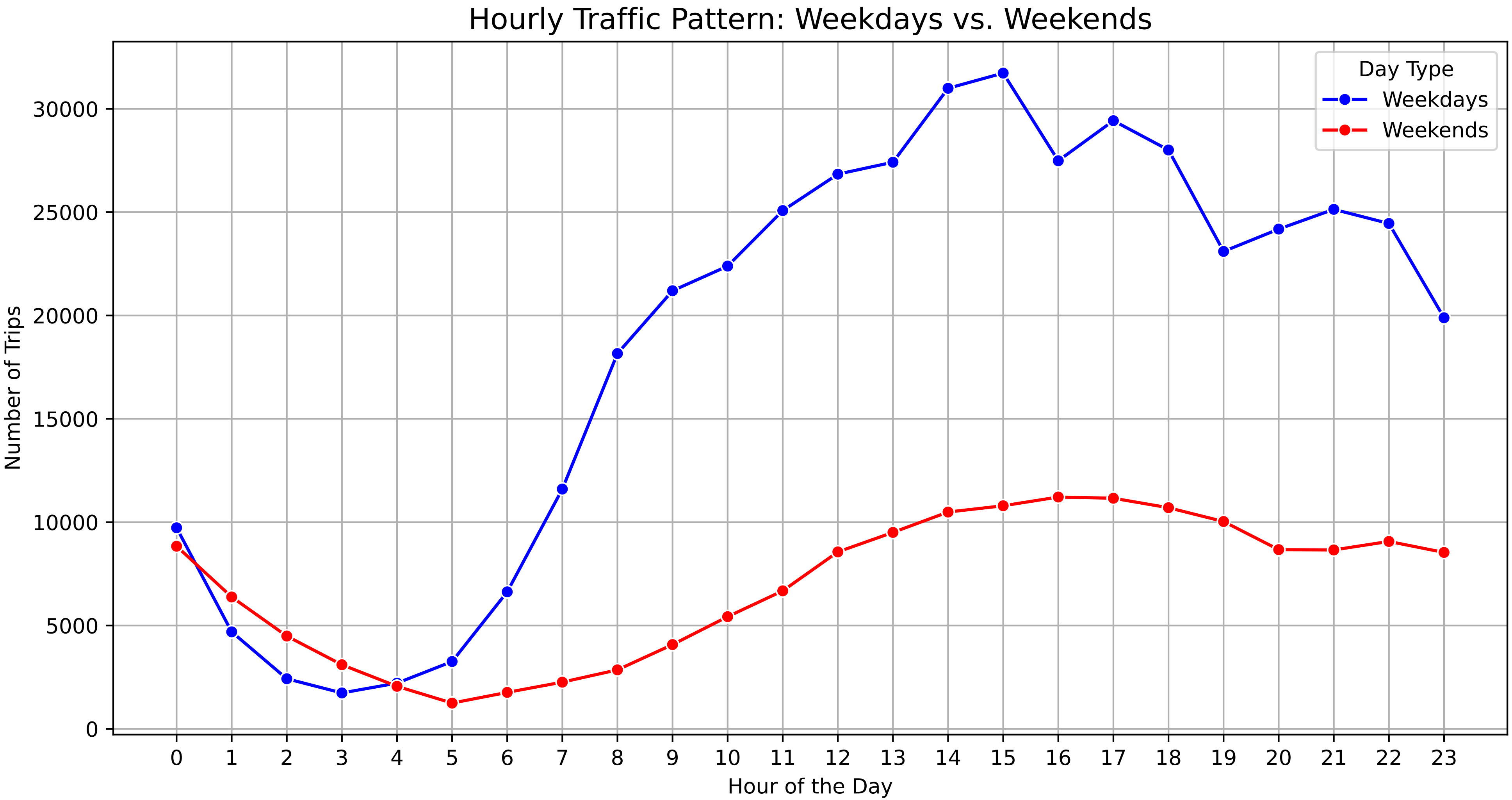

Number of Trips Per Hour
Busiest Hour: 15 (42523 trips)
Number of Trips
0
5000
10000
15000
20000
25000
30000
35000
40000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Hour of the Day

Conclusion/Remarks:

3.2.2 Inference from the Hourly Traffic Patterns

Weekday Trends (Blue Line):
- Morning rush between 7 AM - 10 AM due to office commuters.
- Increase in trips from noon to evening suggests people heading for lunch, meetings, or errands.
- Peak from 5 PM - 9 PM aligns with office closure and evening social activities.
- Drop after 10 PM as fewer people travel late at night.

Weekend Trends (Red Line):
- Fewer trips overall compared to weekdays.


Hourly Traffic Pattern: Weekdays vs. Weekends
Day Type
Weekdays
Weekends
Number of Trips
0
5000
10000
15000
20000
25000
30000
Hour of the Day
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

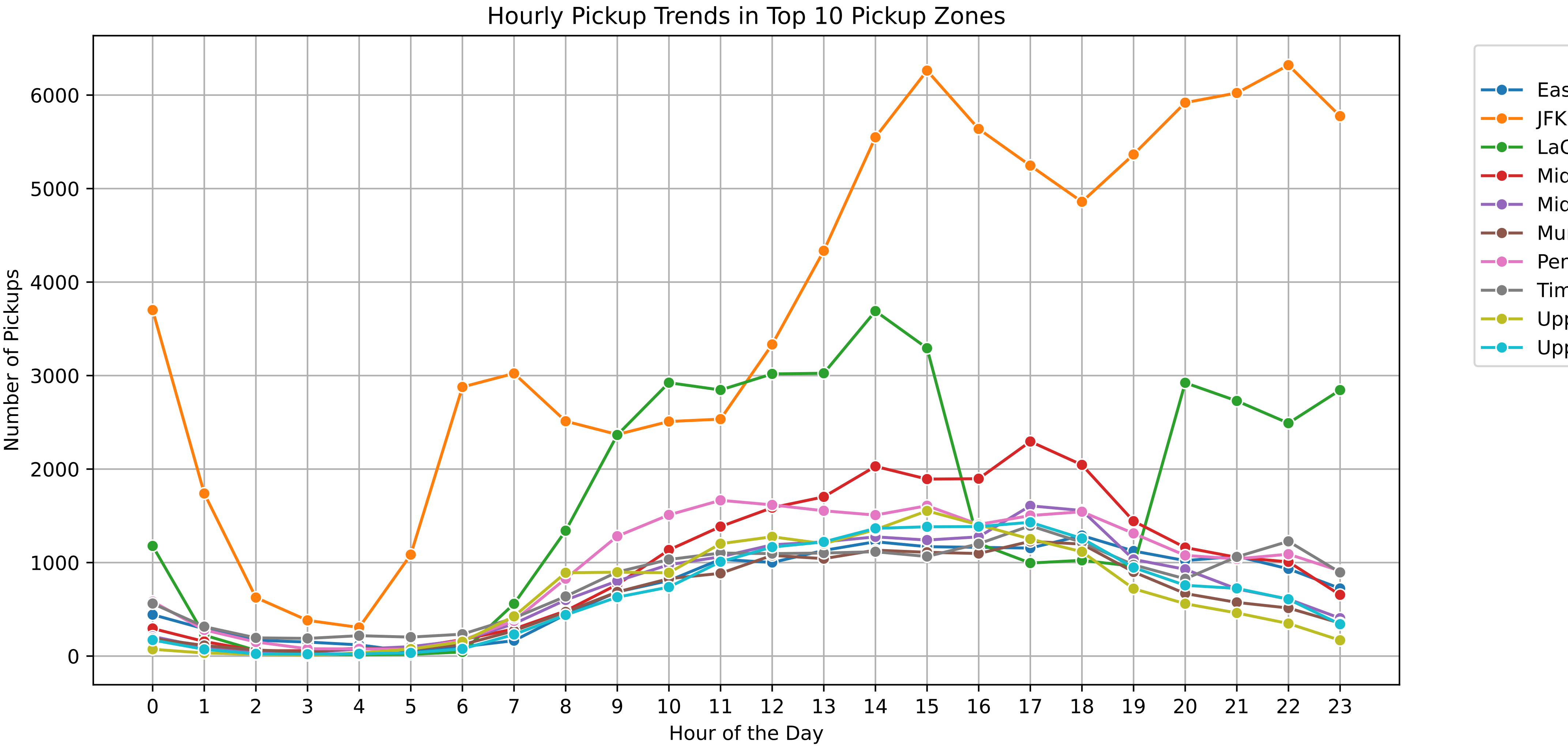
Conclusion/Remarks:

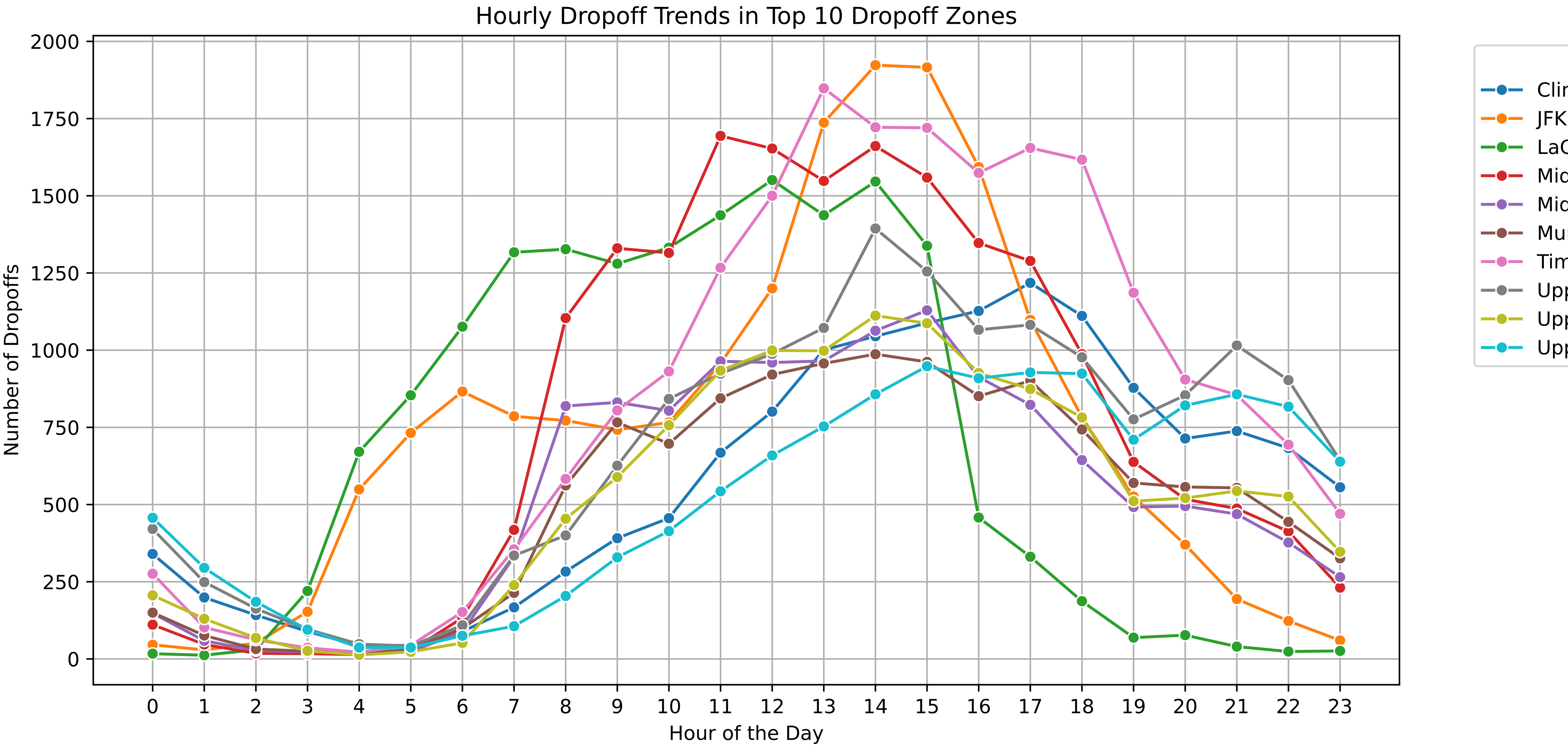
Inference from the Hourly Traffic Patterns

Weekday Trends (Blue Line)

There's a morning rush between 7 AM - 10 AM, likely due to office commuters.
A steady increase in trips from noon to evening suggests people heading for lunch, meetings, or errands.
The peak occurs around 5 PM - 9 PM, which aligns with office closure and evening social activities.
A drop after 10 PM as fewer people travel late at night.

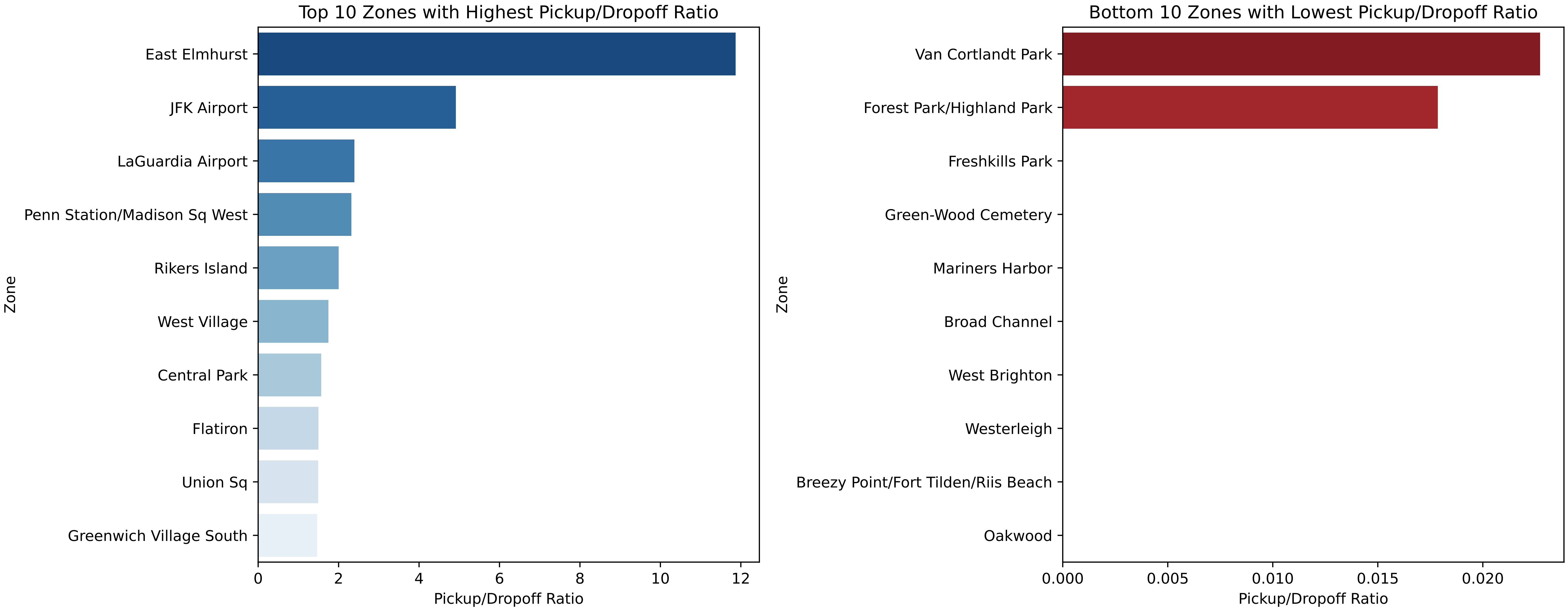
Weekend Trends (Red Line)


Hourly Pickup Trends in Top 10 Pickup Zones
Number of Pickups
0
1000
2000
3000
4000
5000
6000
Hour of the Day
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Eas
JFK
LaG
Mid
Mid
Mu
Pen
Tim
Upp
Upp


Hourly Dropoff Trends in Top 10 Dropoff Zones
Number of Dropoffs
Hour of the Day
0
250
500
750
1000
1250
1500
1750
2000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Clin
JFK
LaG
Mid
Mid
Mu
Tim
Upp
Upp
Upp

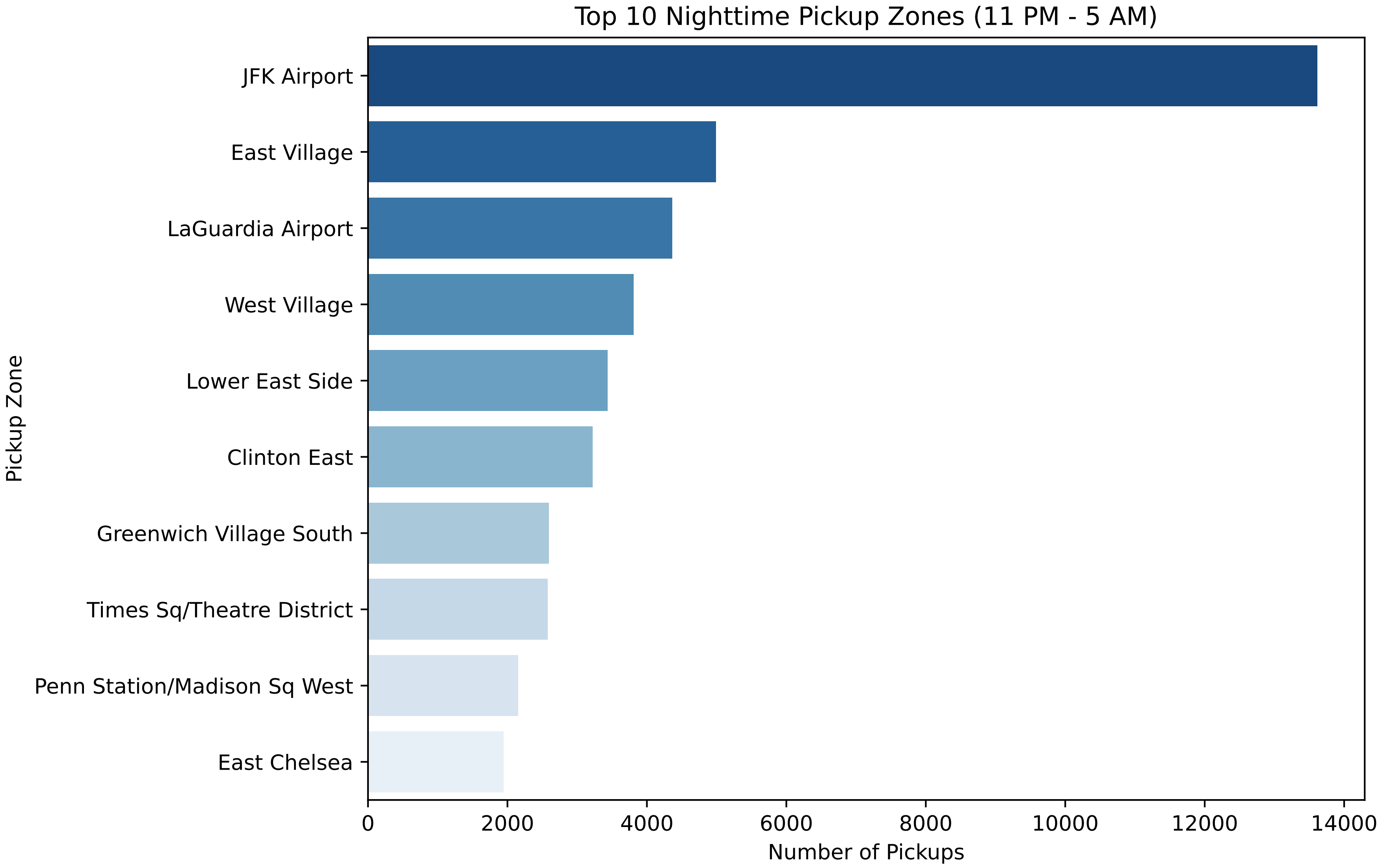
Conclusion/Remarks:

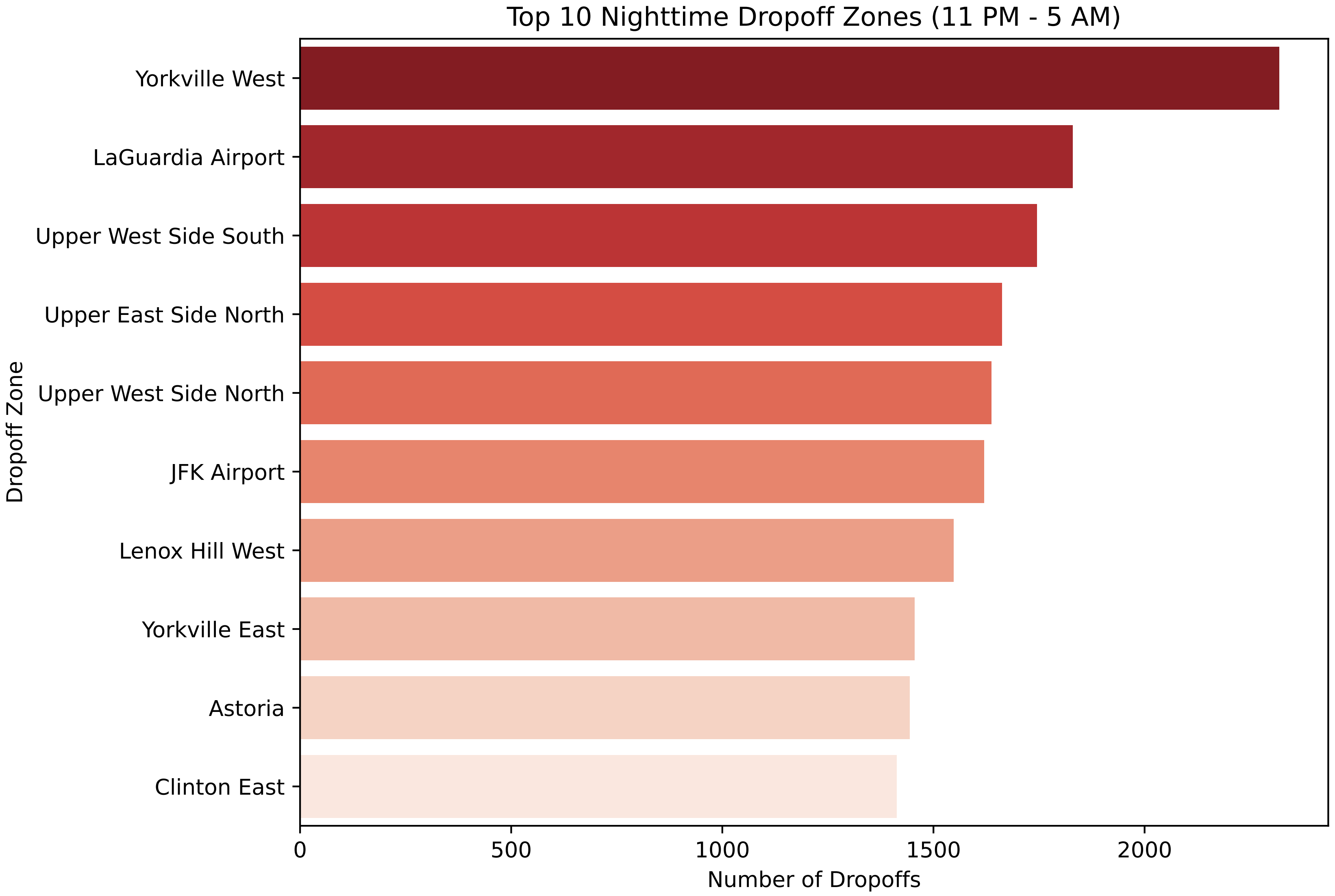
3.2.5 Showing Top 10 Pickup and Dropoff Zones, Throughout the Day, Hour Wise in Line plot.
Aroung 15:00 PM both are high


Top 10 Zones with Highest Pickup/Dropoff Ratio
East Elmhurst
JFK Airport
LaGuardia Airport
Penn Station/Madison Sq West
Rikers Island
West Village
Central Park
Flatiron
Union Sq
Greenwich Village South
Zone
0
2
4
6
8
10
12
Pickup/Dropoff Ratio
Bottom 10 Zones with Lowest Pickup/Dropoff Ratio
Van Cortlandt Park
Forest Park/Highland Park
Freshkills Park
Green-Wood Cemetery
Mariners Harbor
Broad Channel
West Brighton
Westerleigh
Breezy Point/Fort Tilden/Riis Beach
Oakwood
Zone
0.000
0.005
0.010
0.015
0.020
Pickup/Dropoff Ratio

Conclusion/Remarks:

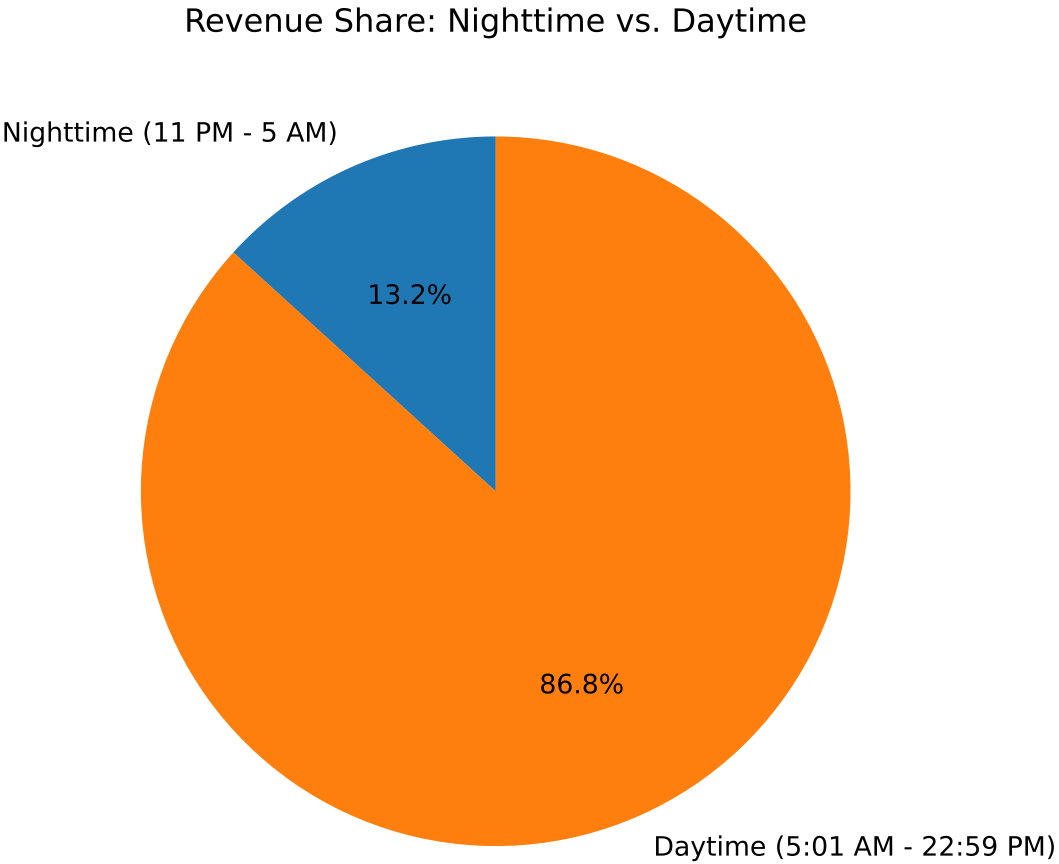
3.2.6 Showing Top & Botth 10 Pickup and Dropoff Ratio. East Elmhurstis the heighest and Forest Park/Highland Park is the lowest


Top 10 Nighttime Pickup Zones (11 PM - 5 AM)
Pickup Zone
JFK Airport
East Village
LaGuardia Airport
West Village
Lower East Side
Clinton East
Greenwich Village South
Times Sq/Theatre District
Penn Station/Madison Sq West
East Chelsea
0
2000
4000
6000
8000
10000
12000
14000
Number of Pickups


Top 10 Nighttime Dropoff Zones (11 PM - 5 AM)
Dropoff Zone
Yorkville West
LaGuardia Airport
Upper West Side South
Upper East Side North
Upper West Side North
JFK Airport
Lenox Hill West
Yorkville East
Astoria
Clinton East
0
500
1000
1500
2000
Number of Dropoffs

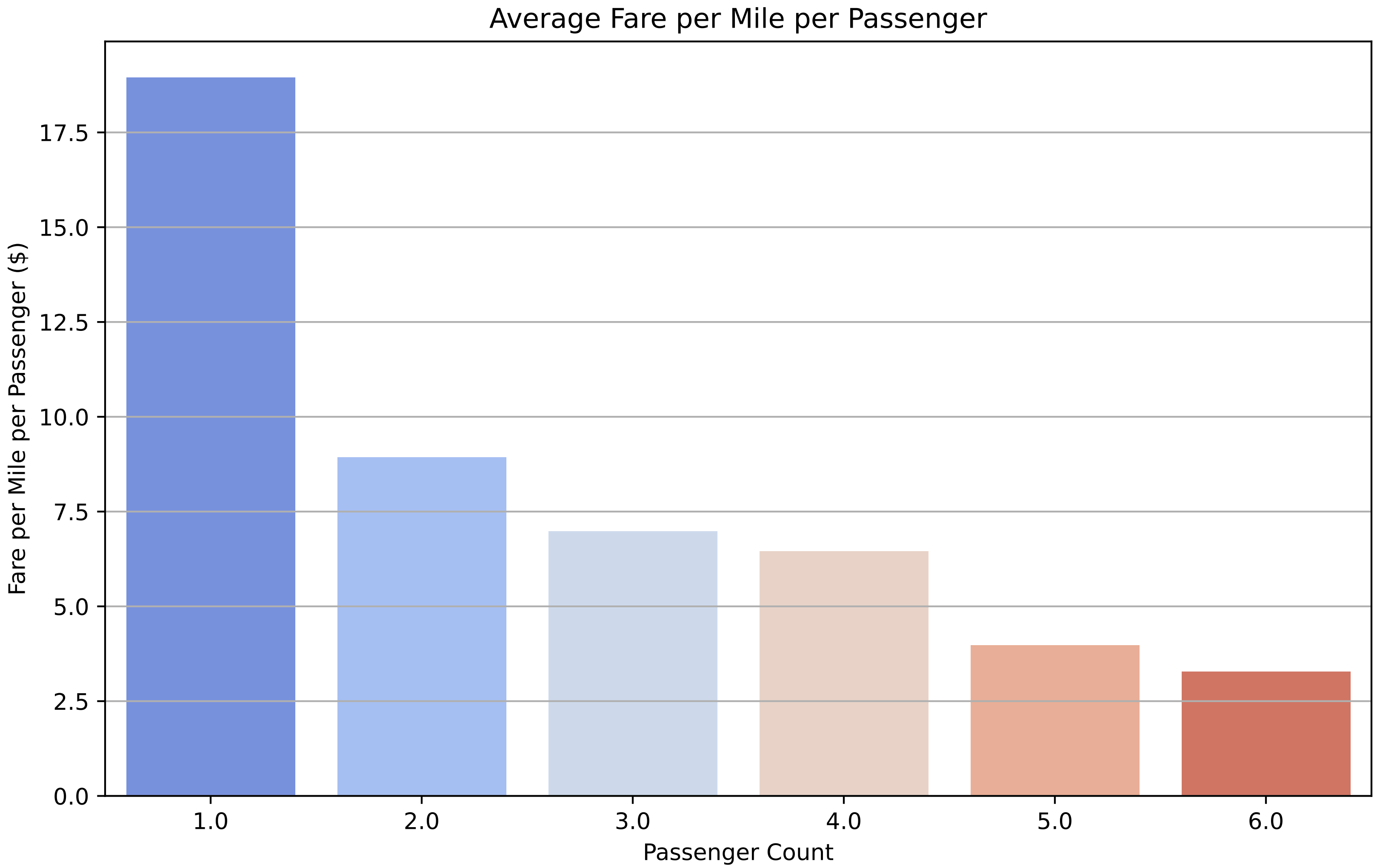
Conclusion/Remarks:

3.2.7 Showing Zones with high Pickup and Dropoff Between 11:00-17:00. JFK Airport the heighest Pickup and Yorkville West is the heighest Dropoff


Revenue Share: Nighttime vs. Daytime
Nighttime (11 PM - 5 AM)
13.2%
86.8%
Daytime (5:01 AM - 22:59 PM)

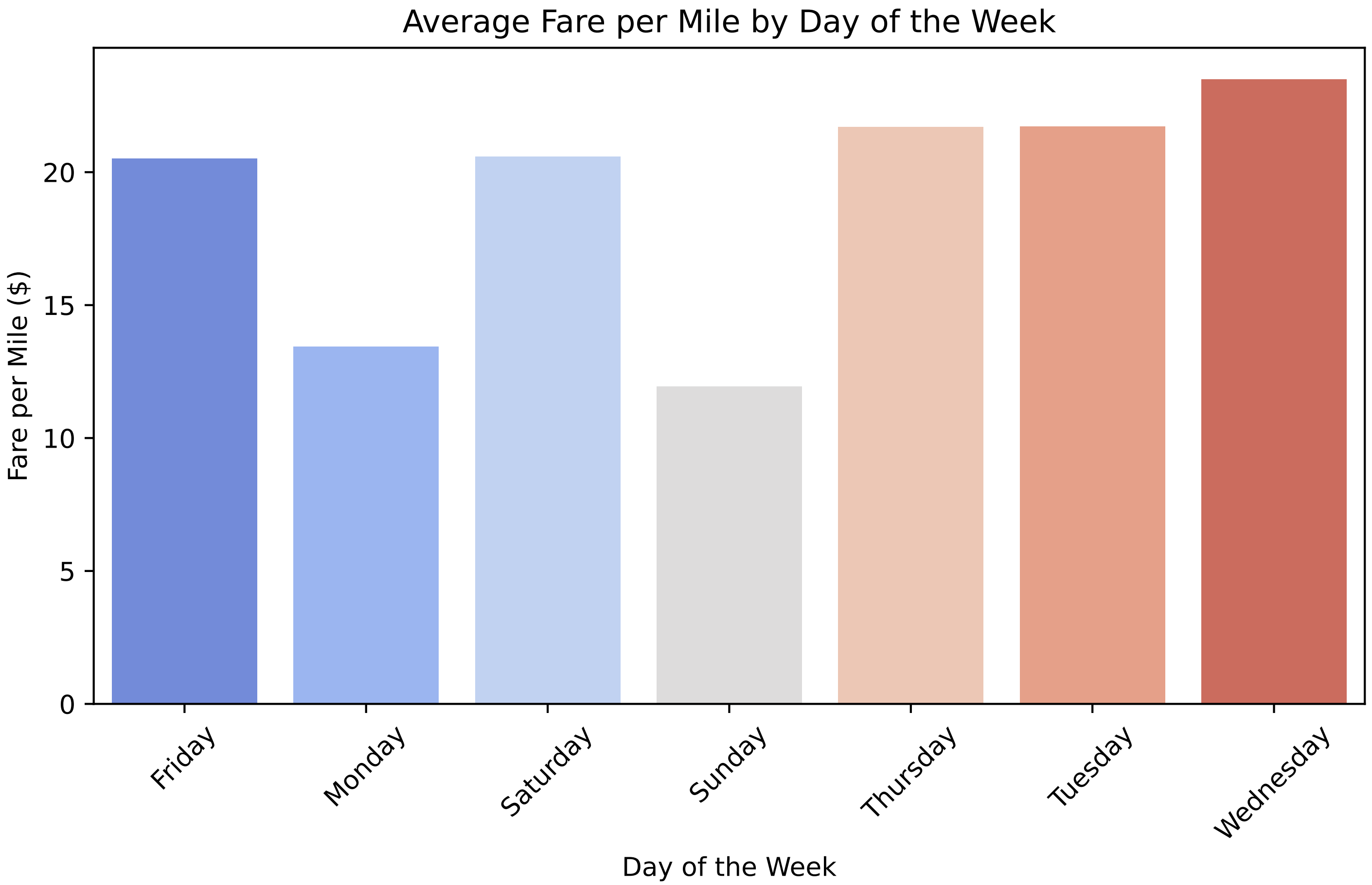
Conclusion/Remarks:

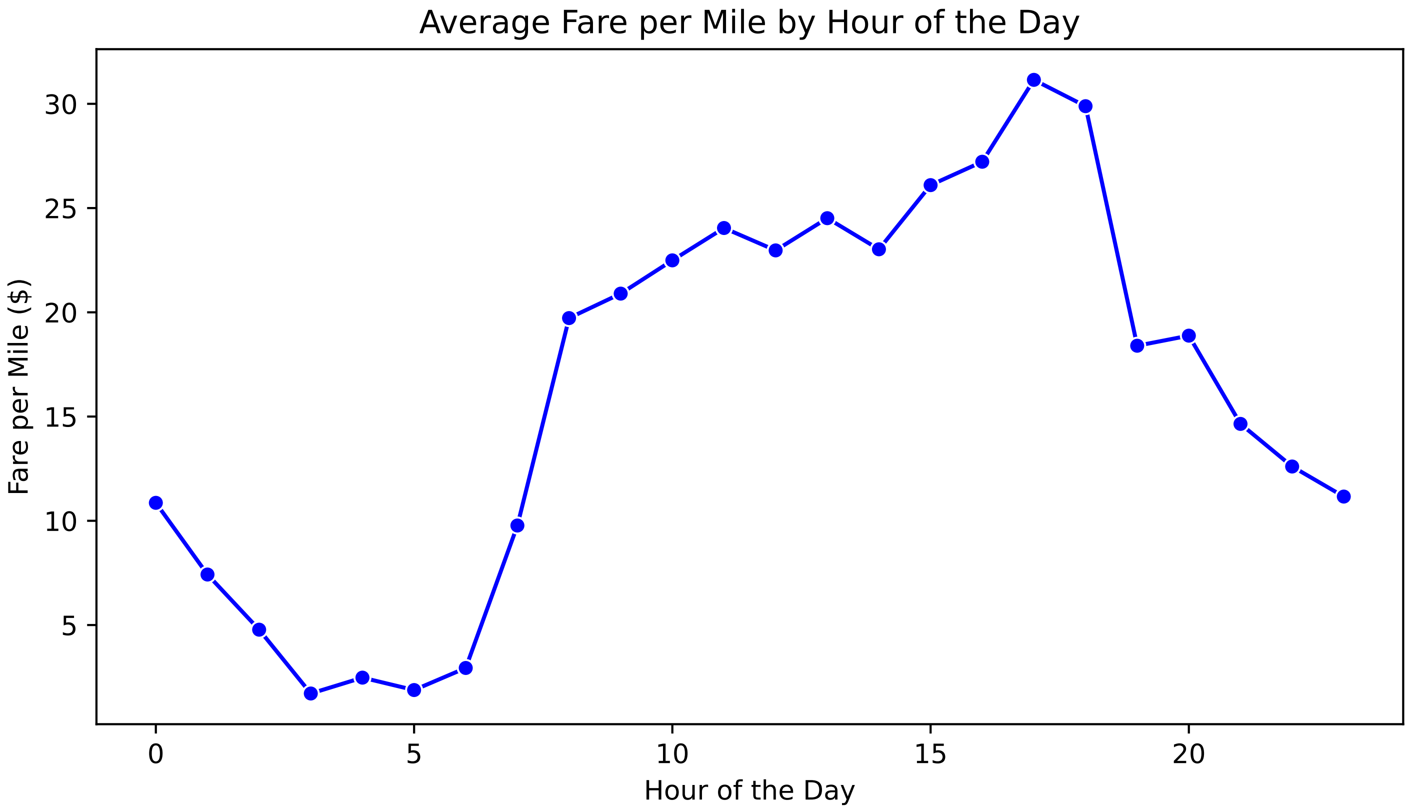
3.2.8 Showing the revenue share for nighttime and daytime hours. 86.8% in day time


Average Fare per Mile per Passenger
Fare per Mile per Passenger ($)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
1.0
2.0
3.0
4.0
5.0
6.0
Passenger Count

Conclusion/Remarks:

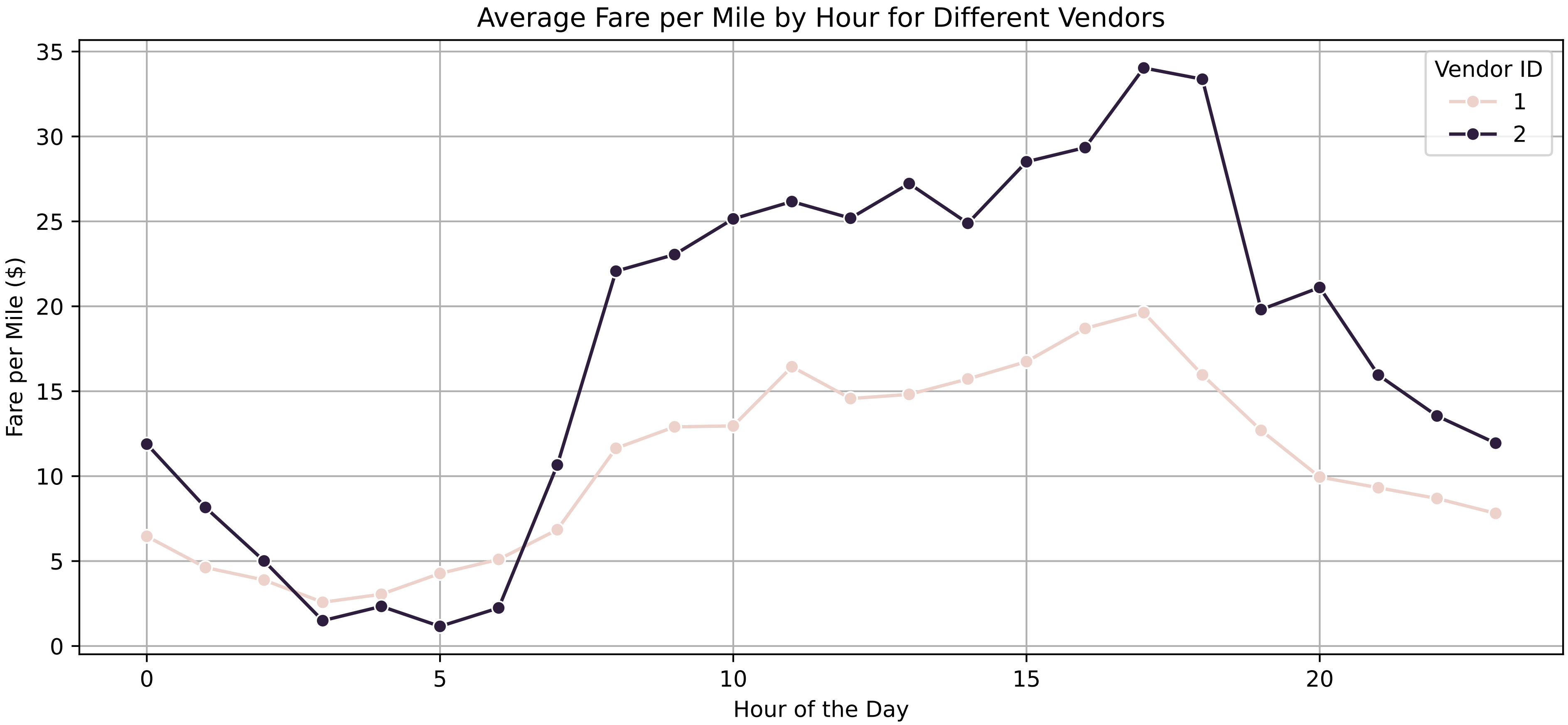
3.2.9 Showing the different passengercount and fare per mile

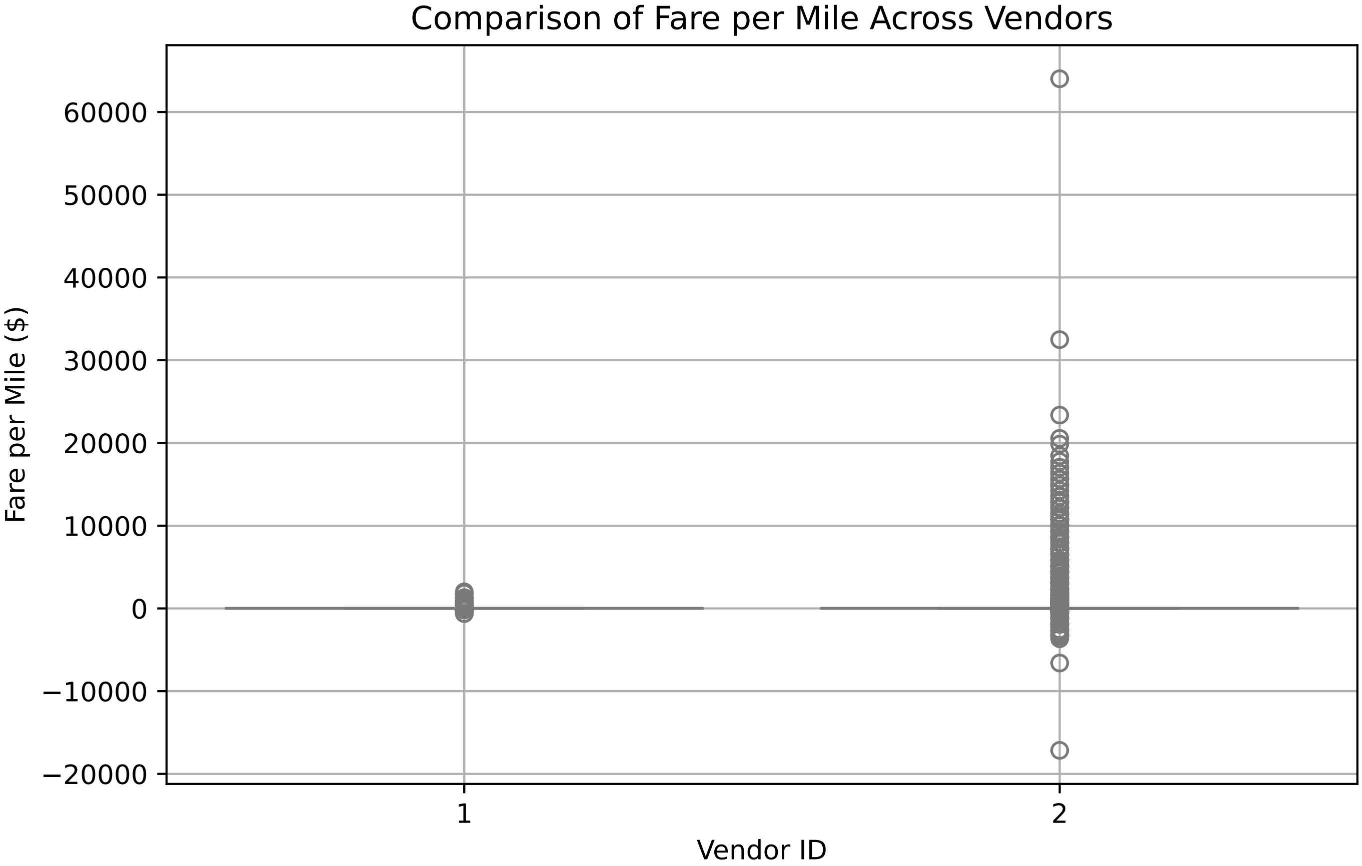

Average Fare per Mile by Day of the Week


Average Fare per Mile by Hour of the Day

Conclusion/Remarks:

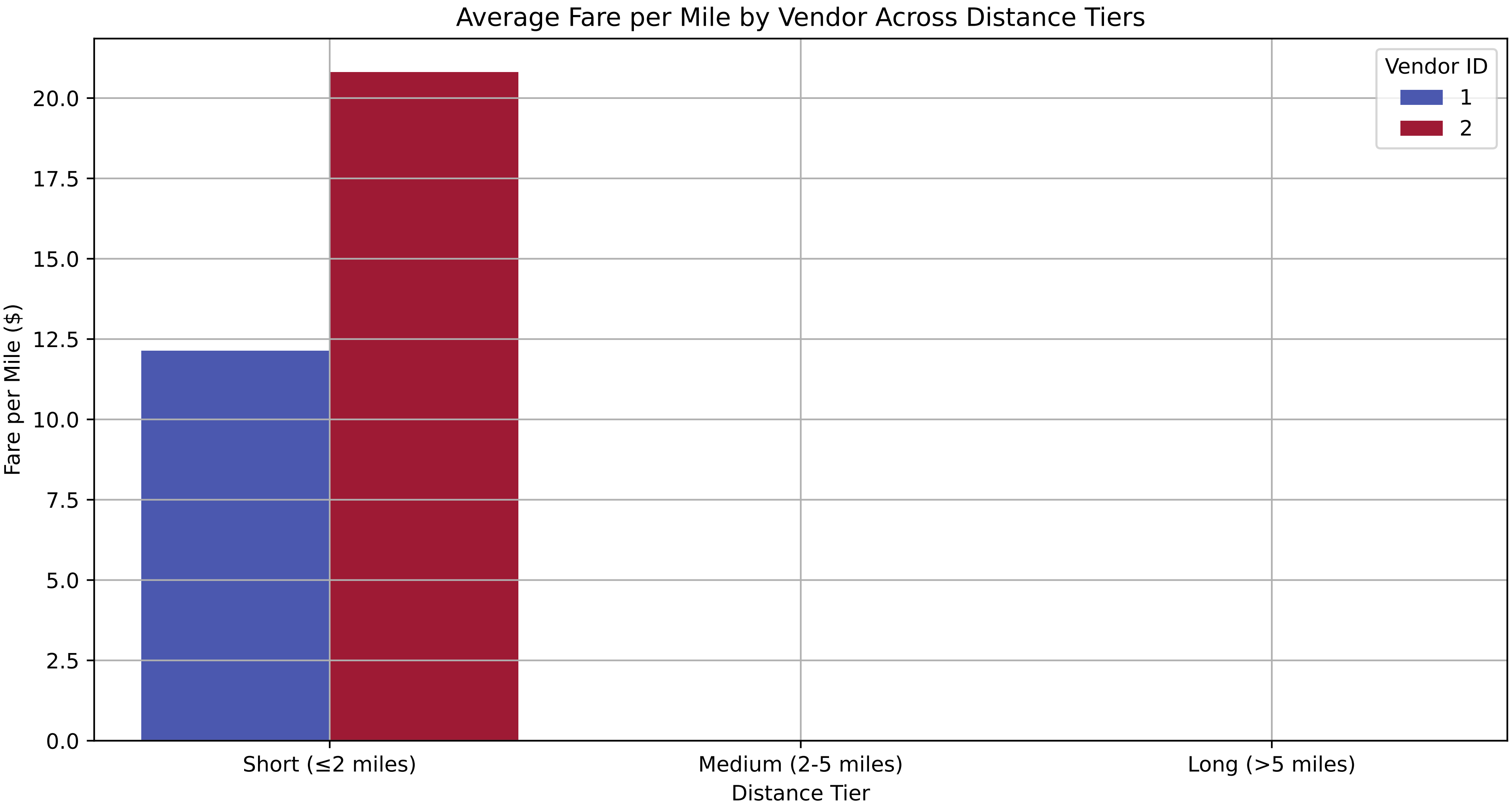
3.2.10 Showing the average fare per mile by hours of the day and by days of the week

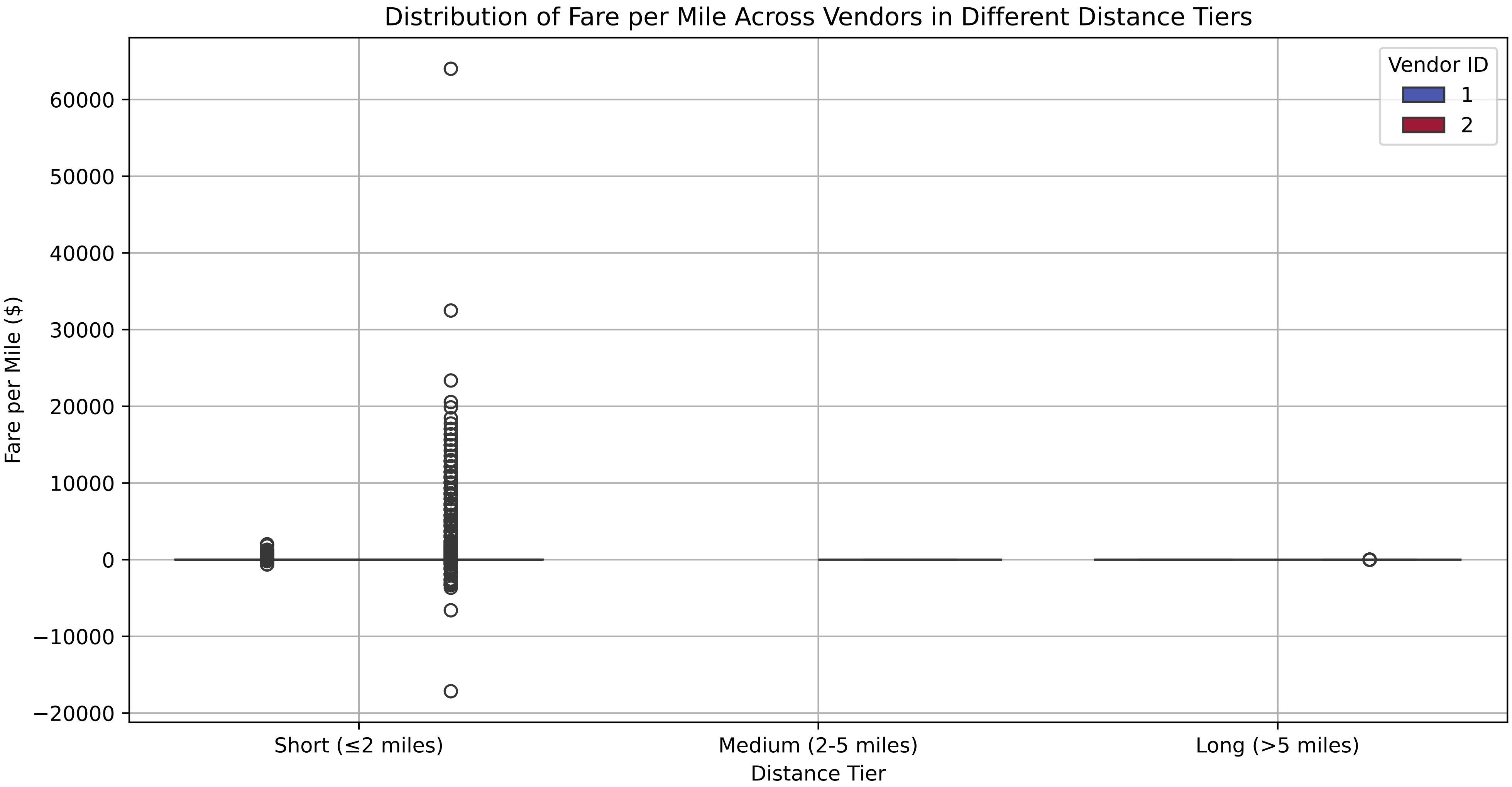

Average Fare per Mile by Hour for Different Vendors
Fare per Mile ($)
Hour of the Day
Vendor ID
1
2
0
5
10
15
20
25
30
35
0
5
10
15
20


Comparison of Fare per Mile Across Vendors
Fare per Mile ($)
60000
50000
40000
30000
20000
10000
0
−10000
−20000
1
2
Vendor ID

Conclusion/Remarks:

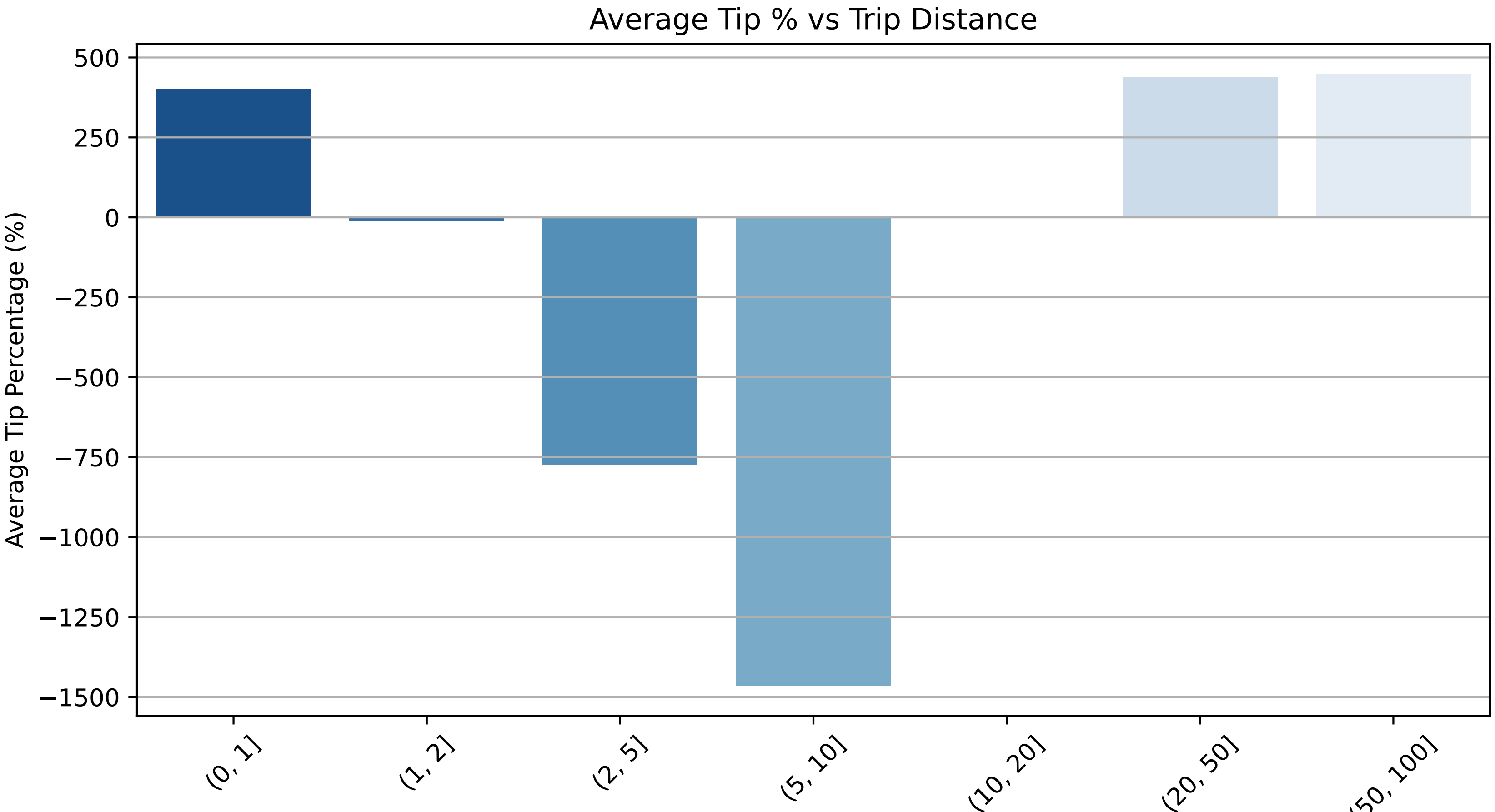
3.2.11 Showing Fare per mile rate for differenmt vendors, throughout the day. Vendor 1 is cheaper.

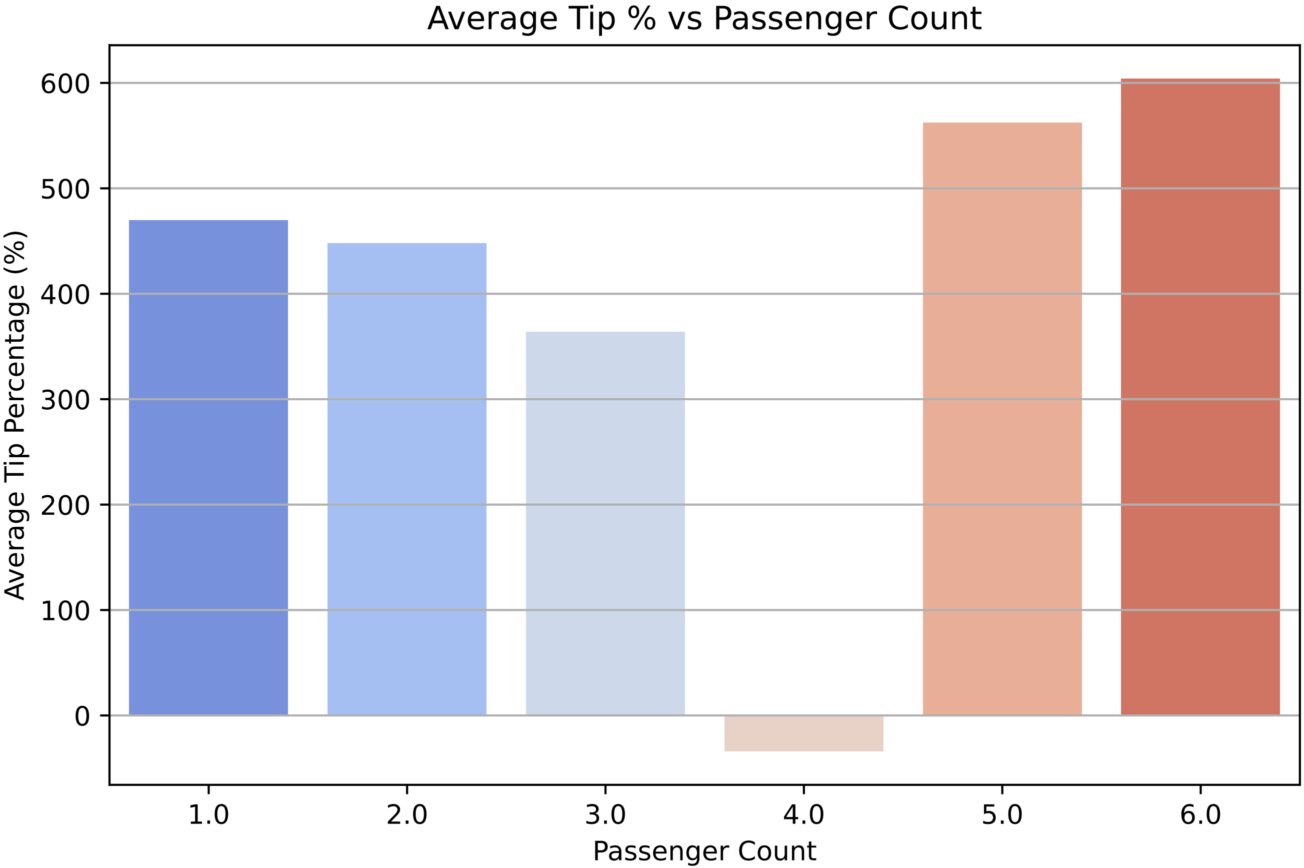

Average Fare per Mile by Vendor Across Distance Tiers
Fare per Mile ($)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Short (≤2 miles)
Medium (2-5 miles)
Long (>5 miles)
Distance Tier
Vendor ID
1
2

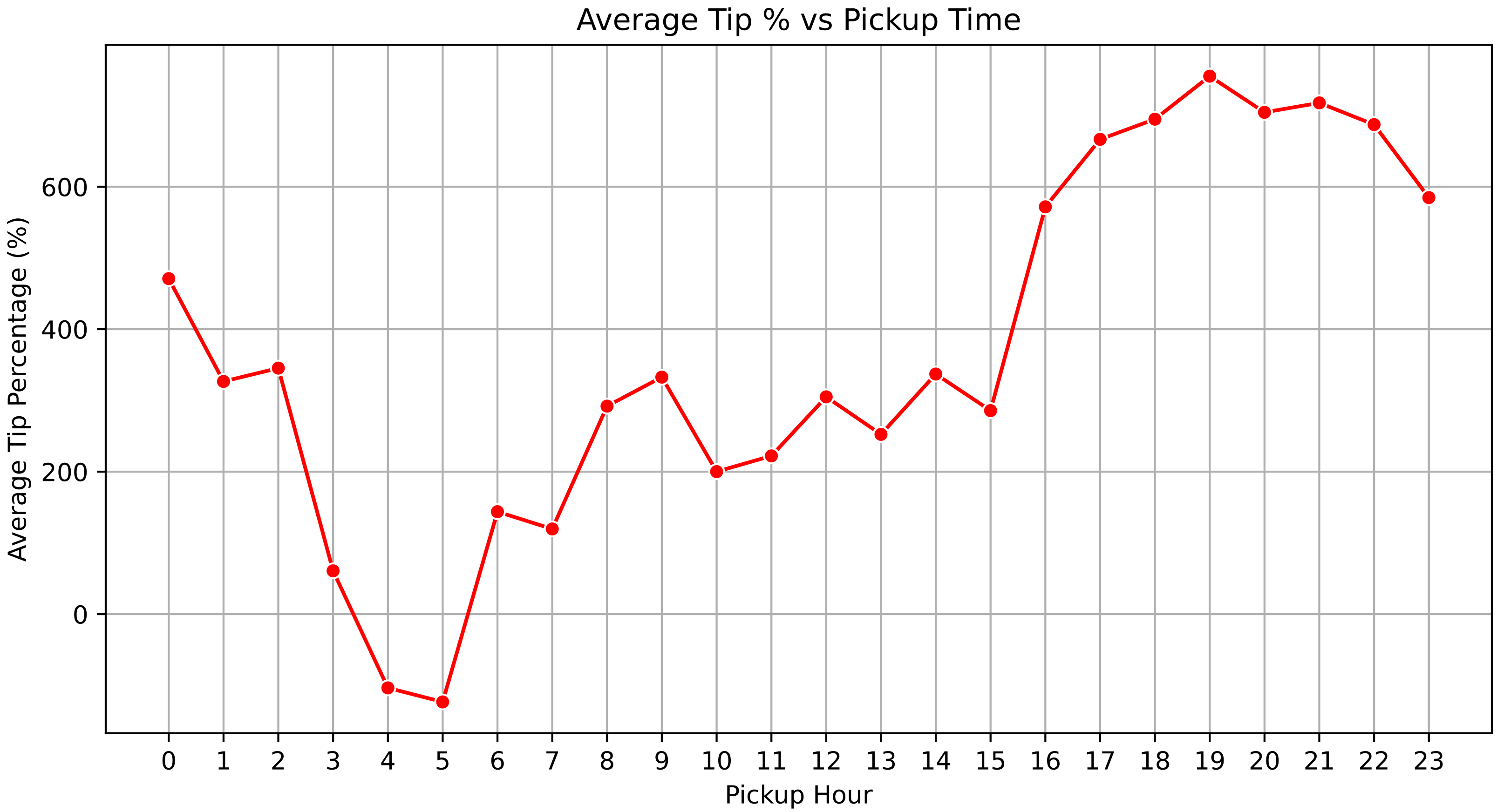

Distribution of Fare per Mile Across Vendors in Different Distance Tiers
Vendor ID
1
2
Fare per Mile ($)
60000
50000
40000
30000
20000
10000
0
−10000
−20000
Short (≤2 miles)
Medium (2-5 miles)
Long (>5 miles)
Distance Tier

Conclusion/Remarks:

3.2.12 Showing the comparison of fare rates of the different vendors in a tiered fashion. Here is also vendor 1 is cheaper


Average Tip % vs Trip Distance
Average Tip Percentage (%)
500
250
0
−250
−500
−750
−1000
−1250
−1500
(0, 1]
(1, 2]
(2, 5]
(5, 10]
(10, 20]
(20, 50]
(50, 100]


Average Tip % vs Passenger Count
Average Tip Percentage (%)
600
500
400
300
200
100
0
1.0
2.0
3.0
4.0
5.0
6.0
Passenger Count

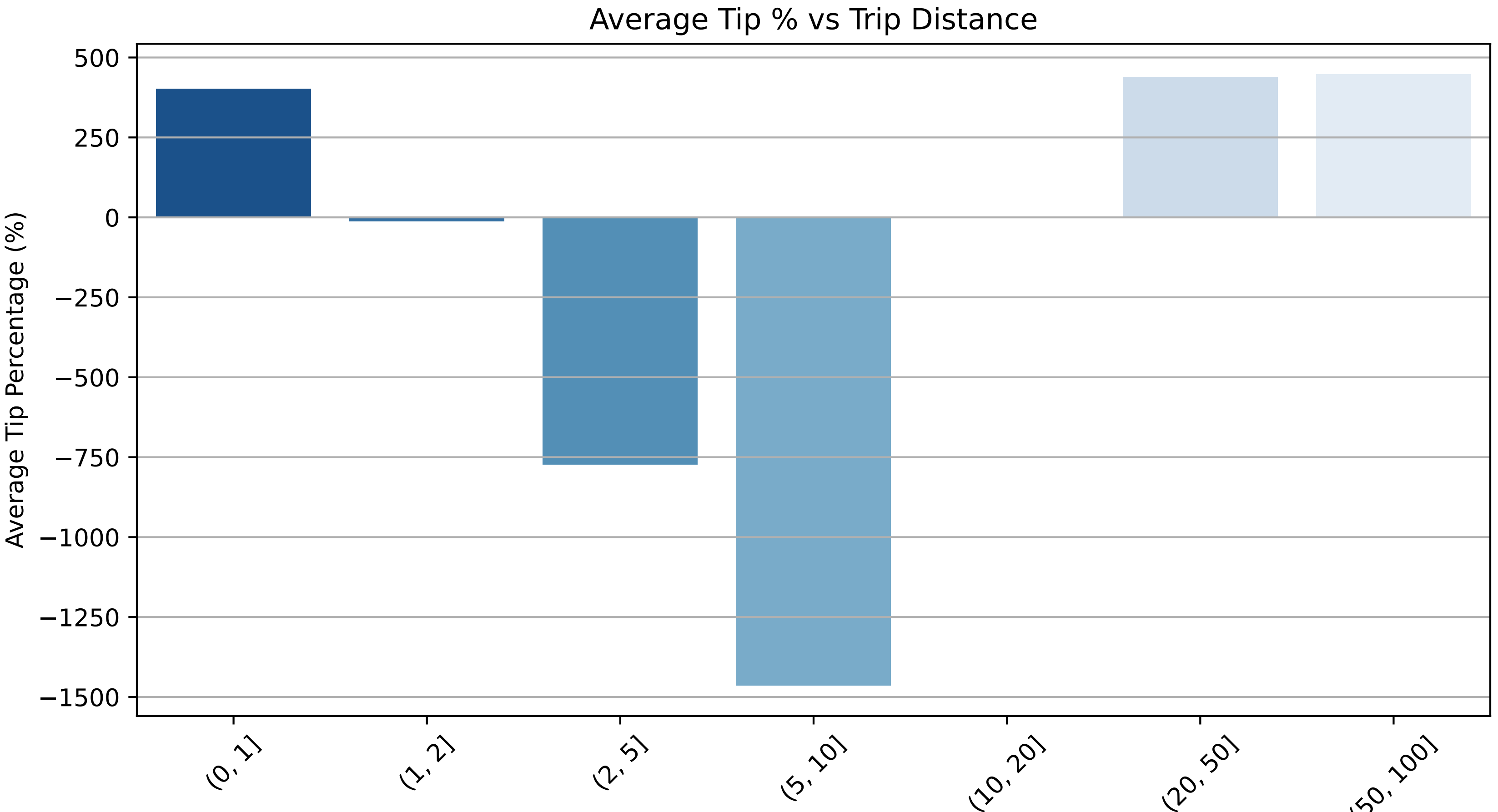

Average Tip % vs Pickup Time
Average Tip Percentage (%)
600
400
200
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Pickup Hour

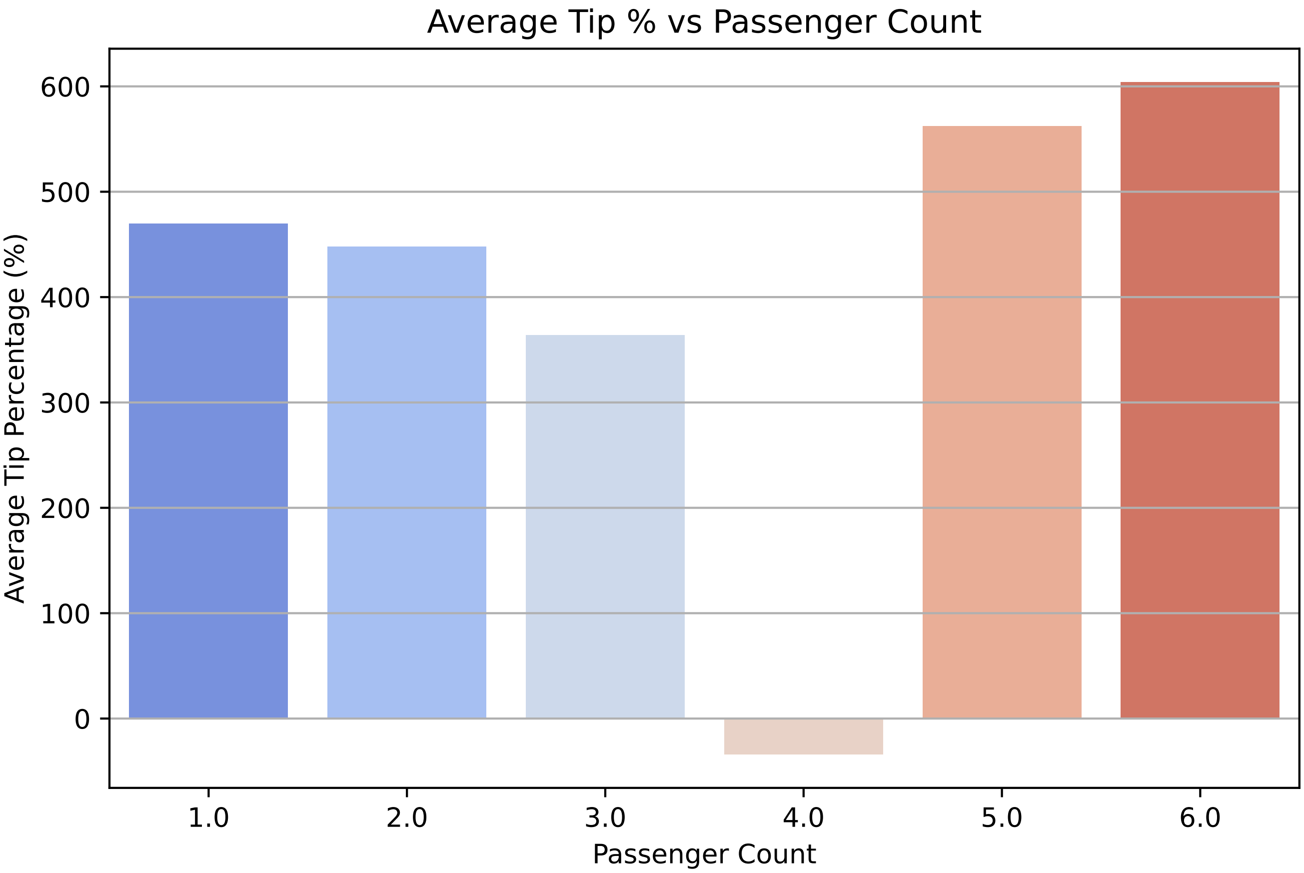
Conclusion/Remarks:

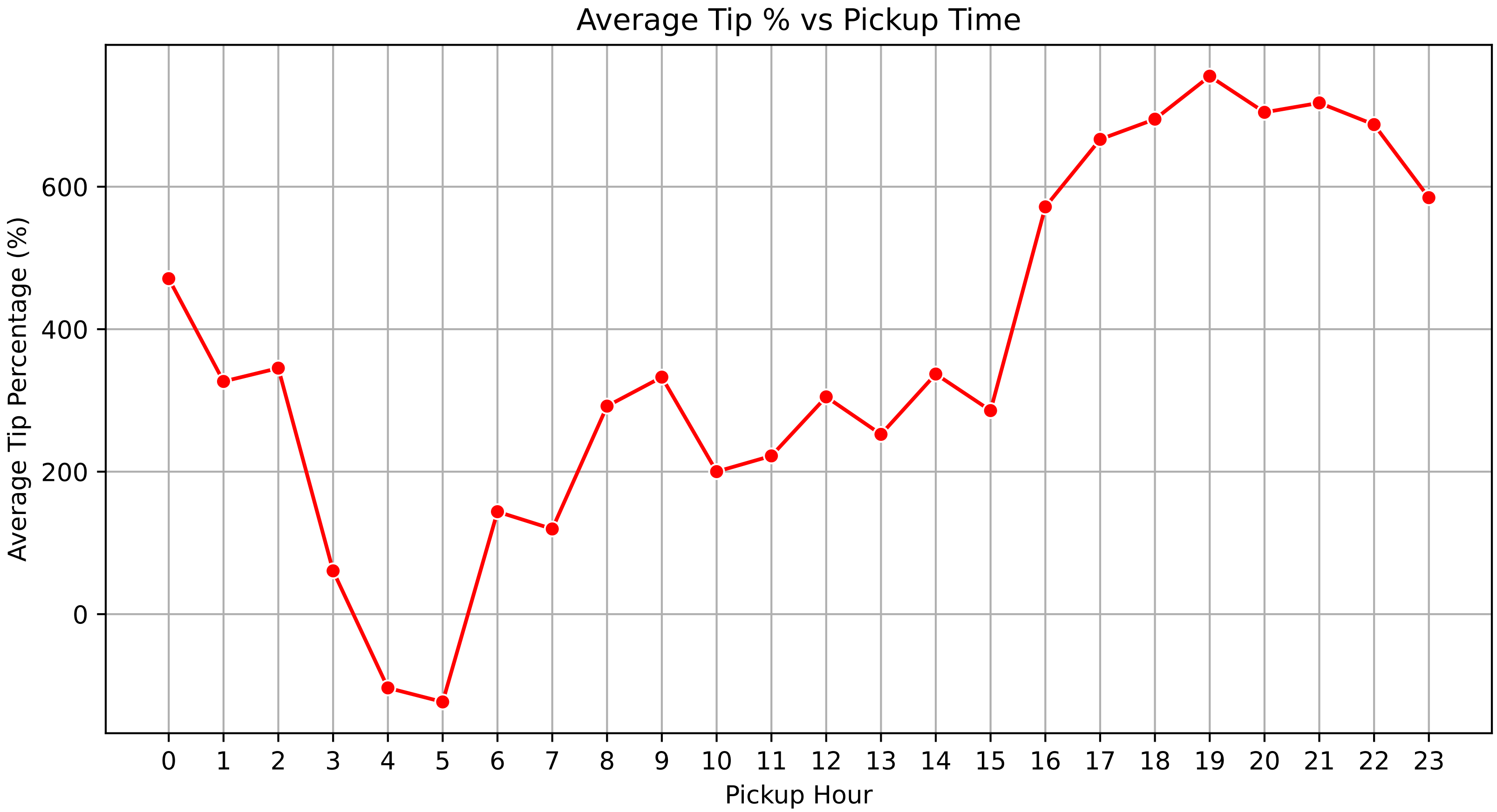
3.2.13 Showing analysis average tip percentages based on trip distances, passenger counts and time of pickup. Tips depends on factors like, Trip distance- between 2 to 20 mile very few tips, Hours - Between 3:30 to 5:00 AM no trips almostn

But I think there should be another factor which is ratecode


Average Tip % vs Rate Code ID
Average Tip Percentage (%)
0
20
40
60
80
100
120
140
160
1
2
3
4
5
6
99
Rate Code ID

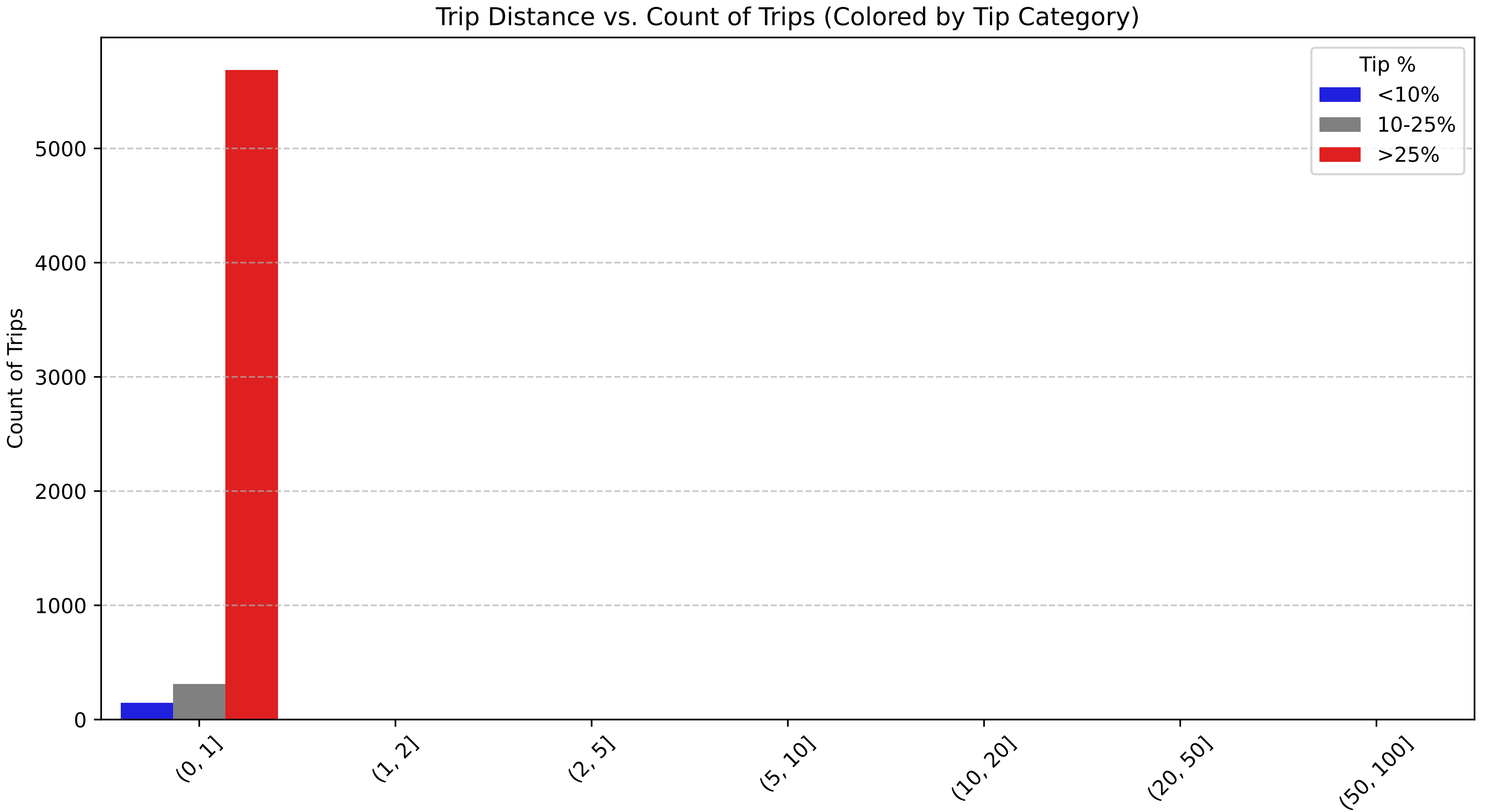

Average Tip % vs Trip Distance
Average Tip Percentage (%)
500
250
0
−250
−500
−750
−1000
−1250
−1500
(0, 1]
(1, 2]
(2, 5]
(5, 10]
(10, 20]
(20, 50]
(50, 100]

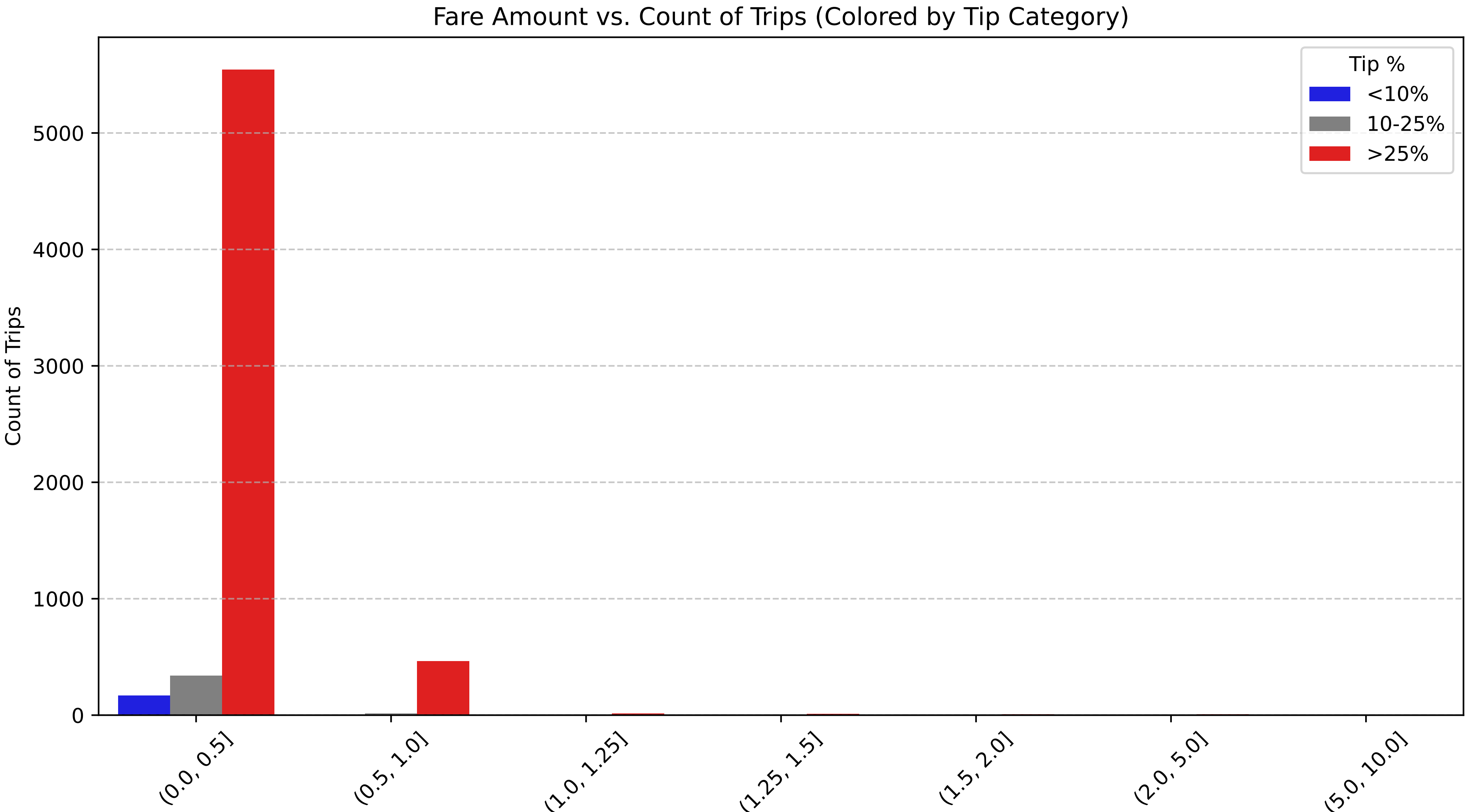

Average Tip % vs Passenger Count
Average Tip Percentage (%)
600
500
400
300
200
100
0
1.0
2.0
3.0
4.0
5.0
6.0
Passenger Count

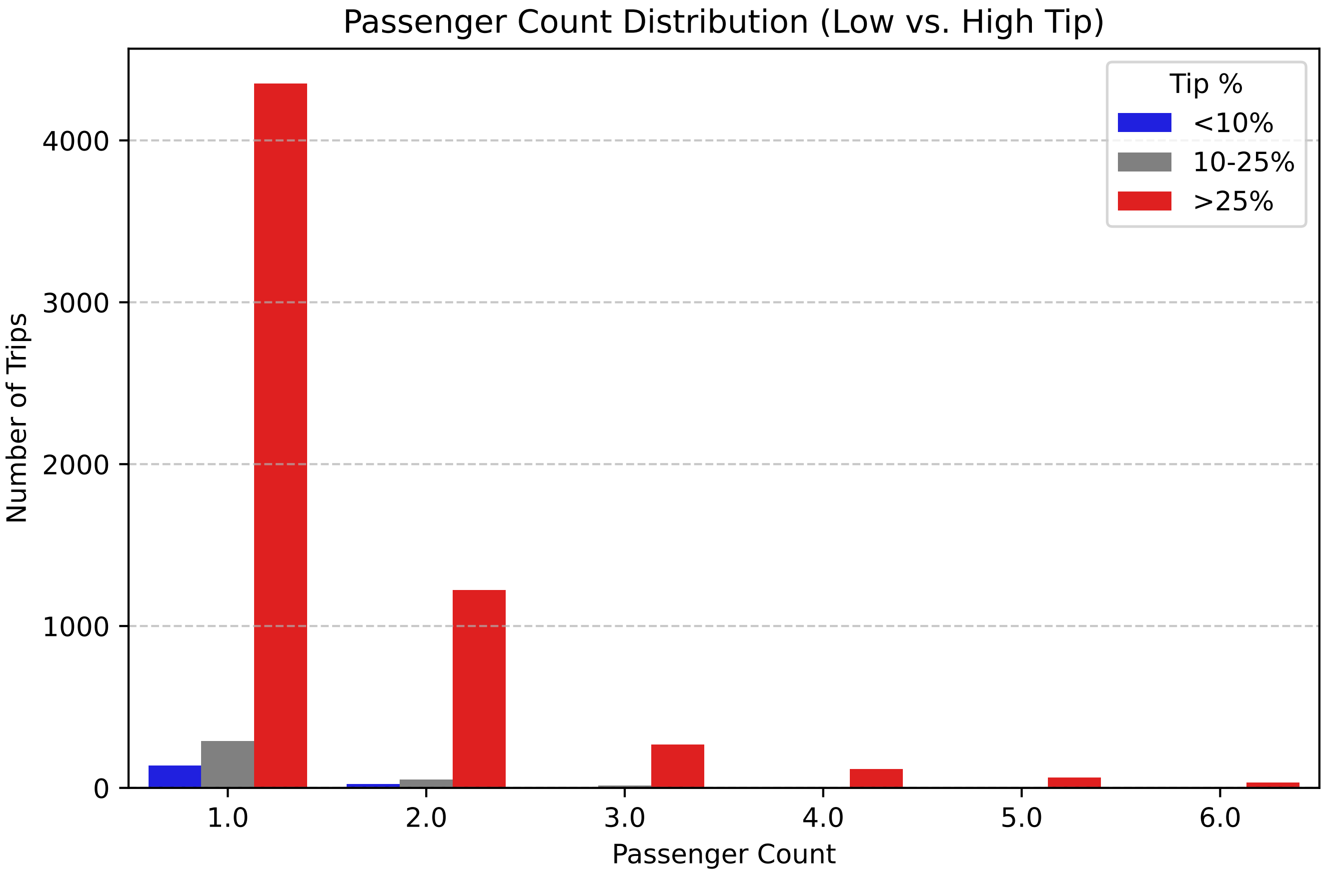

Average Tip % vs Pickup Time
Average Tip Percentage (%)
600
400
200
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Pickup Hour

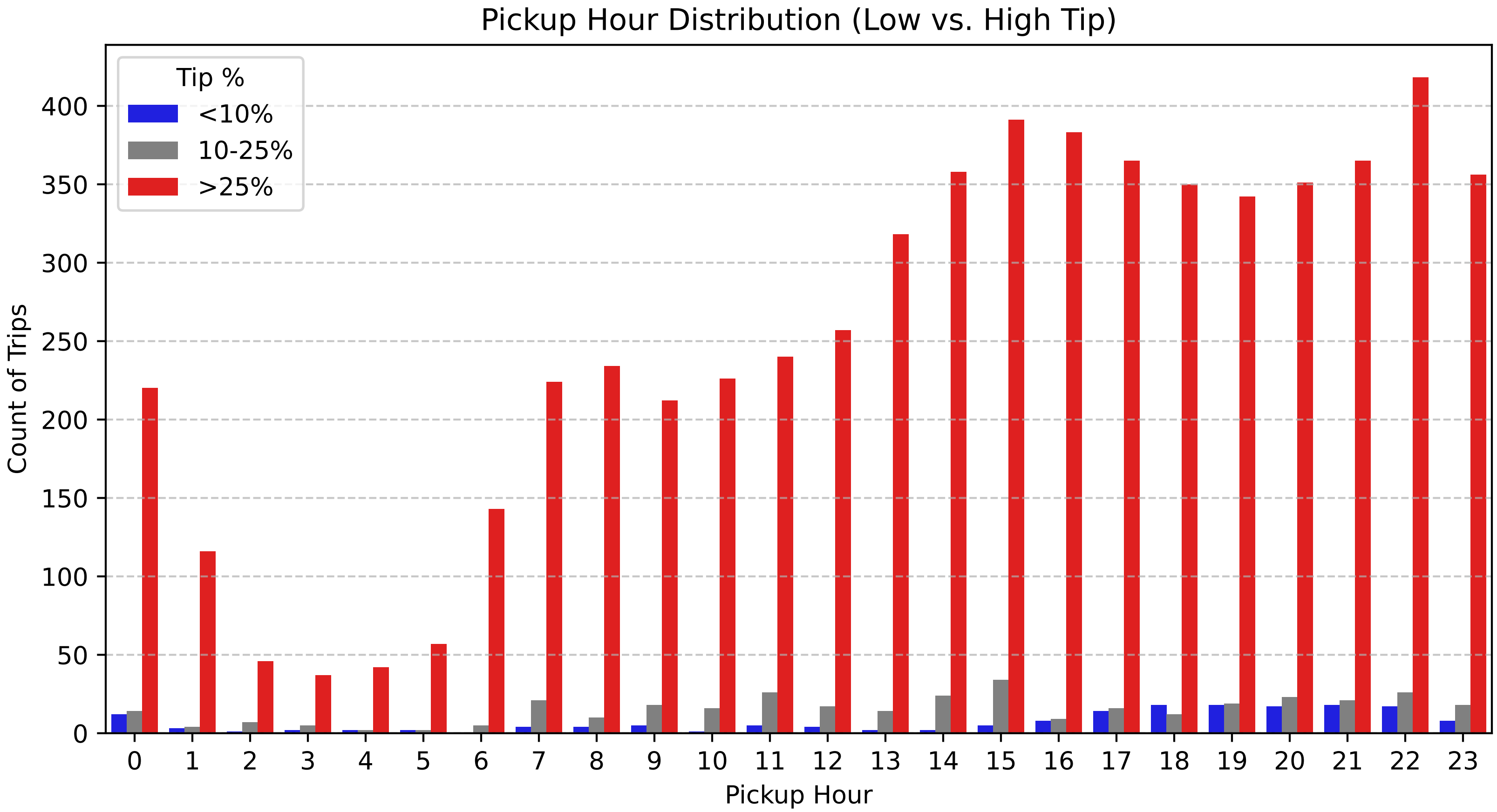
Conclusion/Remarks:

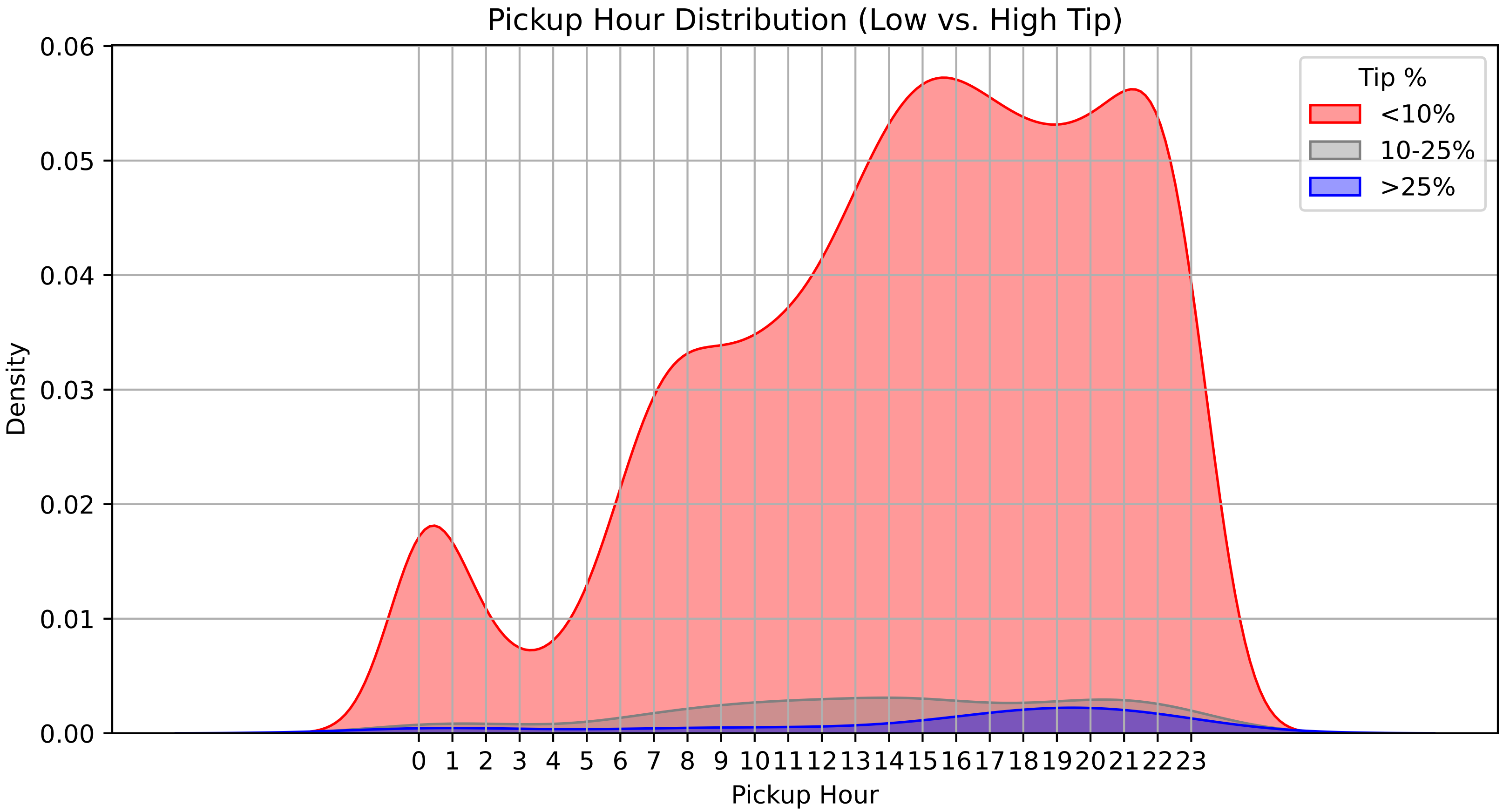
3.2.13 After analysing, it is clear that average tip is maximum for retecode 5, which is Negotiated fare


Trip Distance vs. Count of Trips (Colored by Tip Category)
Count of Trips
0
1000
2000
3000
4000
5000
(0, 1]
(1, 2]
(2, 5]
(5, 10]
(10, 20]
(20, 50]
(50, 100]
Tip %
<10%
10-25%
>25%

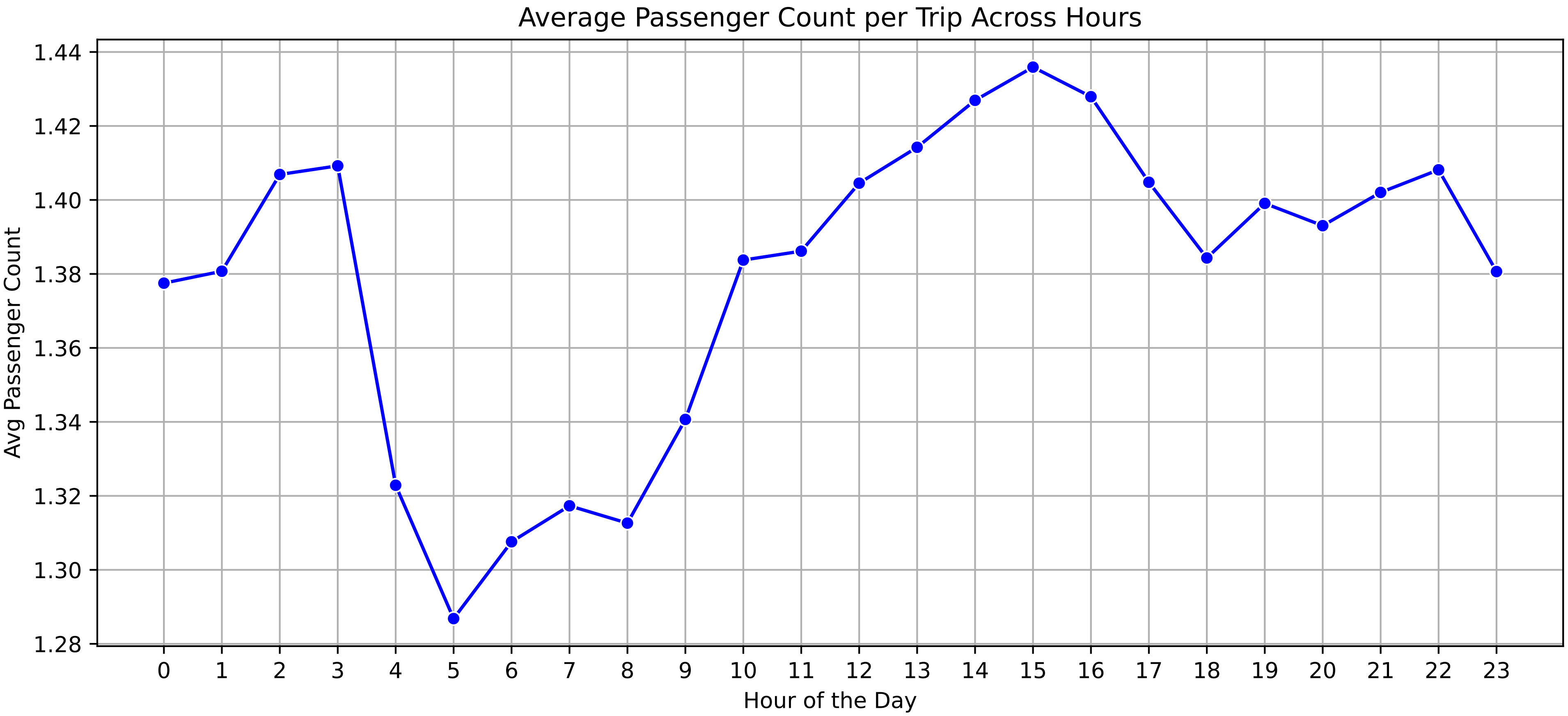

Fare Amount vs. Count of Trips (Colored by Tip Category)
Count of Trips
0
1000
2000
3000
4000
5000
(0.0, 0.5]
(0.5, 1.0]
(1.0, 1.25]
(1.25, 1.5]
(1.5, 2.0]
(2.0, 5.0]
(5.0, 10.0]
Tip %
<10%
10-25%
>25%


Passenger Count Distribution (Low vs. High Tip)
Tip %
<10%
10-25%
>25%
Number of Trips
0
1000
2000
3000
4000
1.0
2.0
3.0
4.0
5.0
6.0
Passenger Count

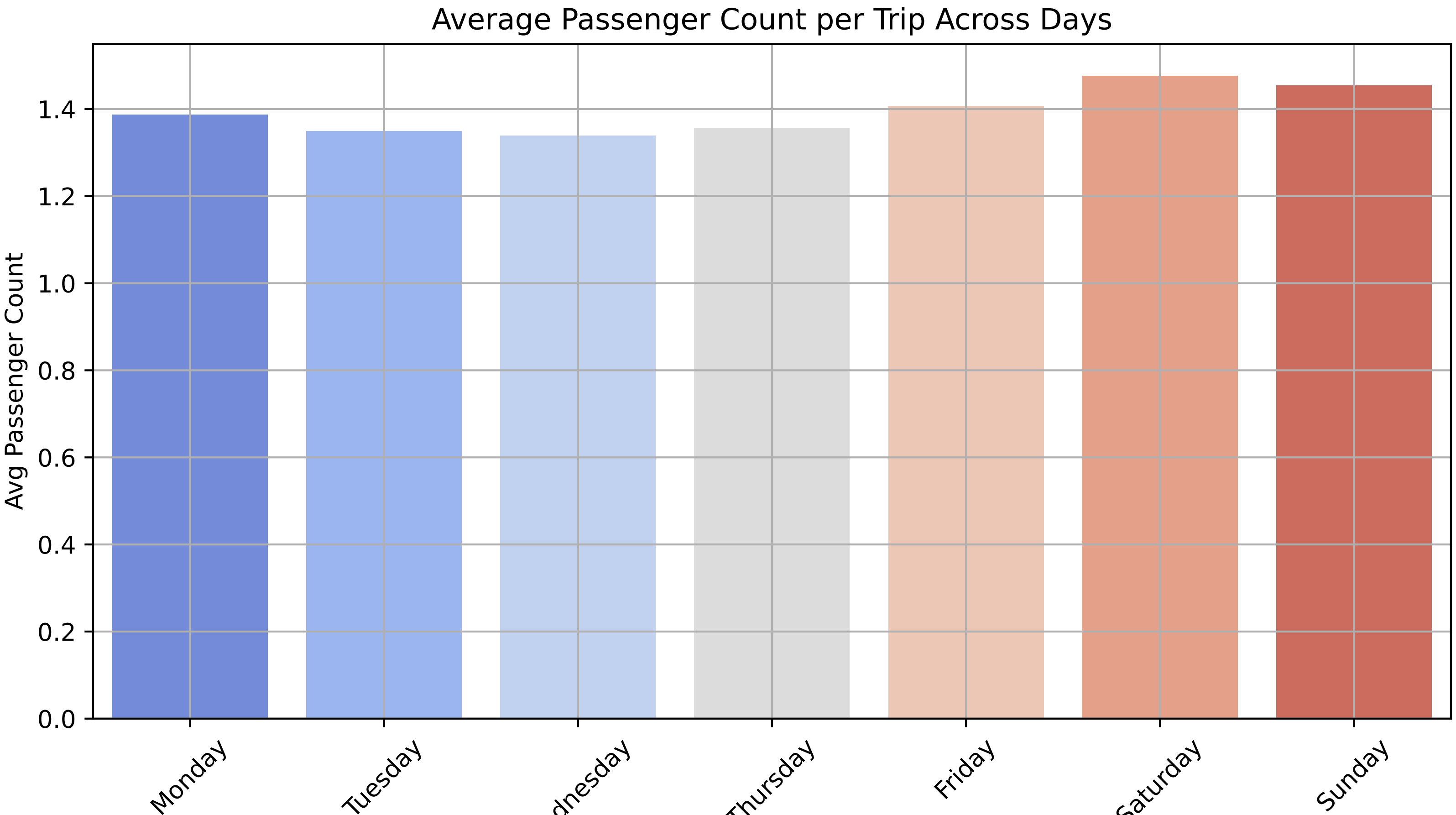

Pickup Hour Distribution (Low vs. High Tip)
Tip %
<10%
10-25%
>25%
Count of Trips
0
50
100
150
200
250
300
350
400
Pickup Hour
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23



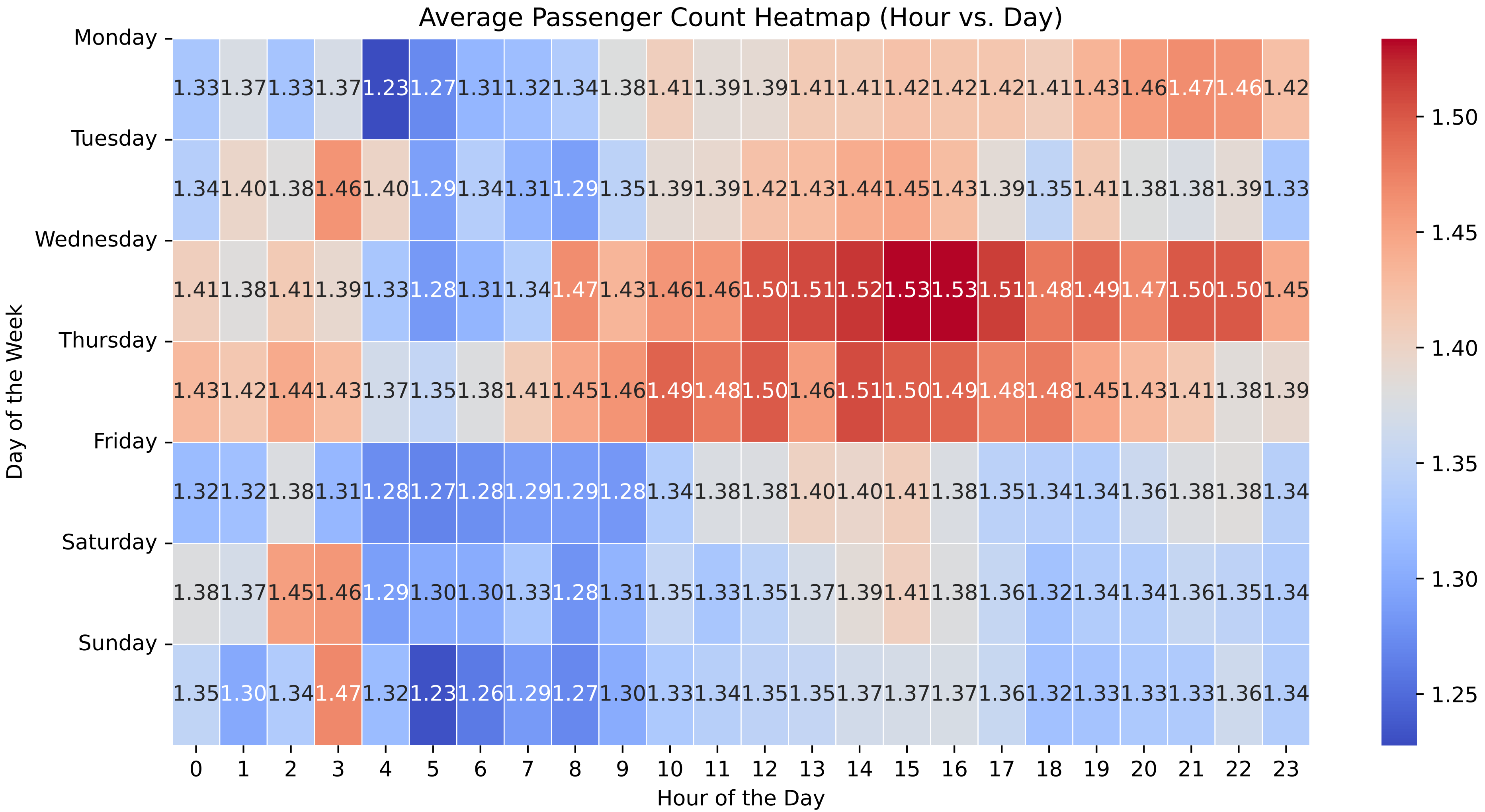
Pickup Hour Distribution (Low vs. High Tip)
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Pickup Hour
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Tip %
<10%
10-25%
>25%



Average Passenger Count per Trip Across Hours
Avg Passenger Count
1.44
1.42
1.40
1.38
1.36
1.34
1.32
1.30
1.28
Hour of the Day
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23



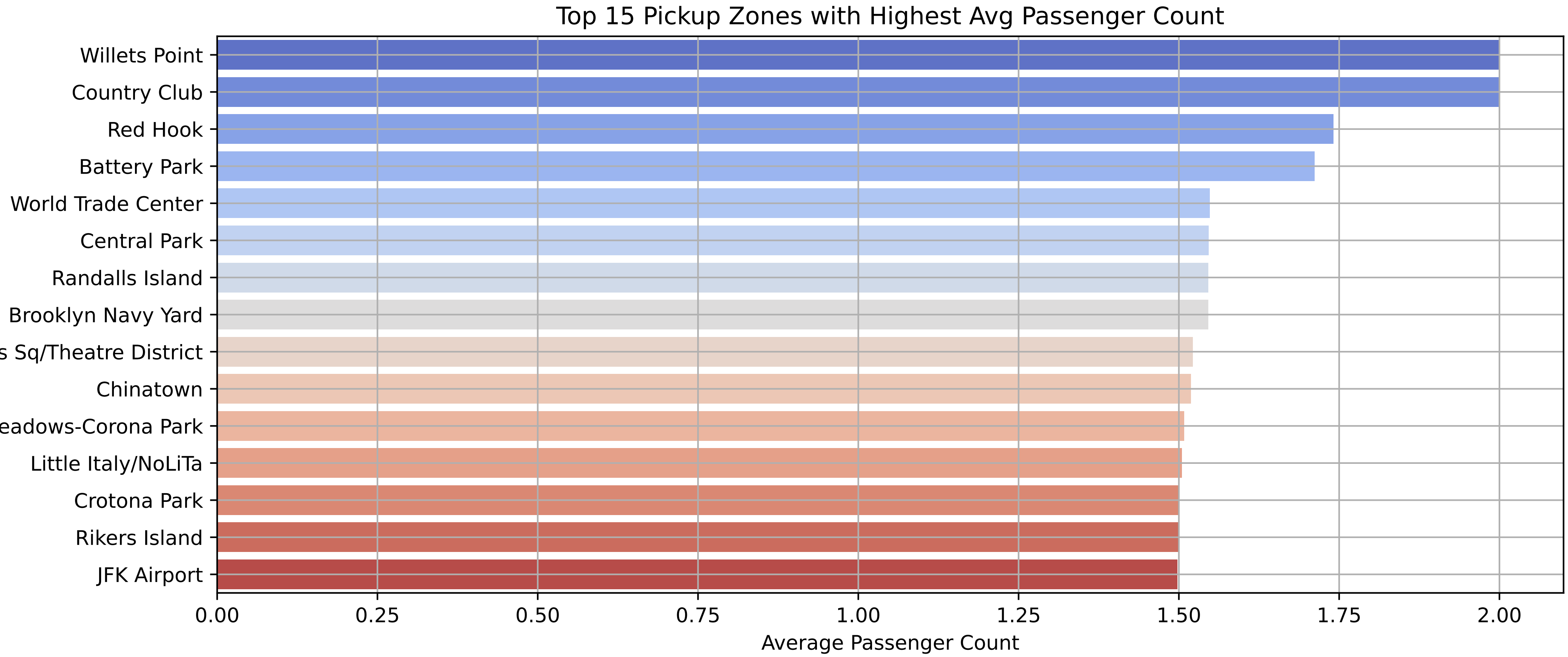
Average Passenger Count per Trip Across Days

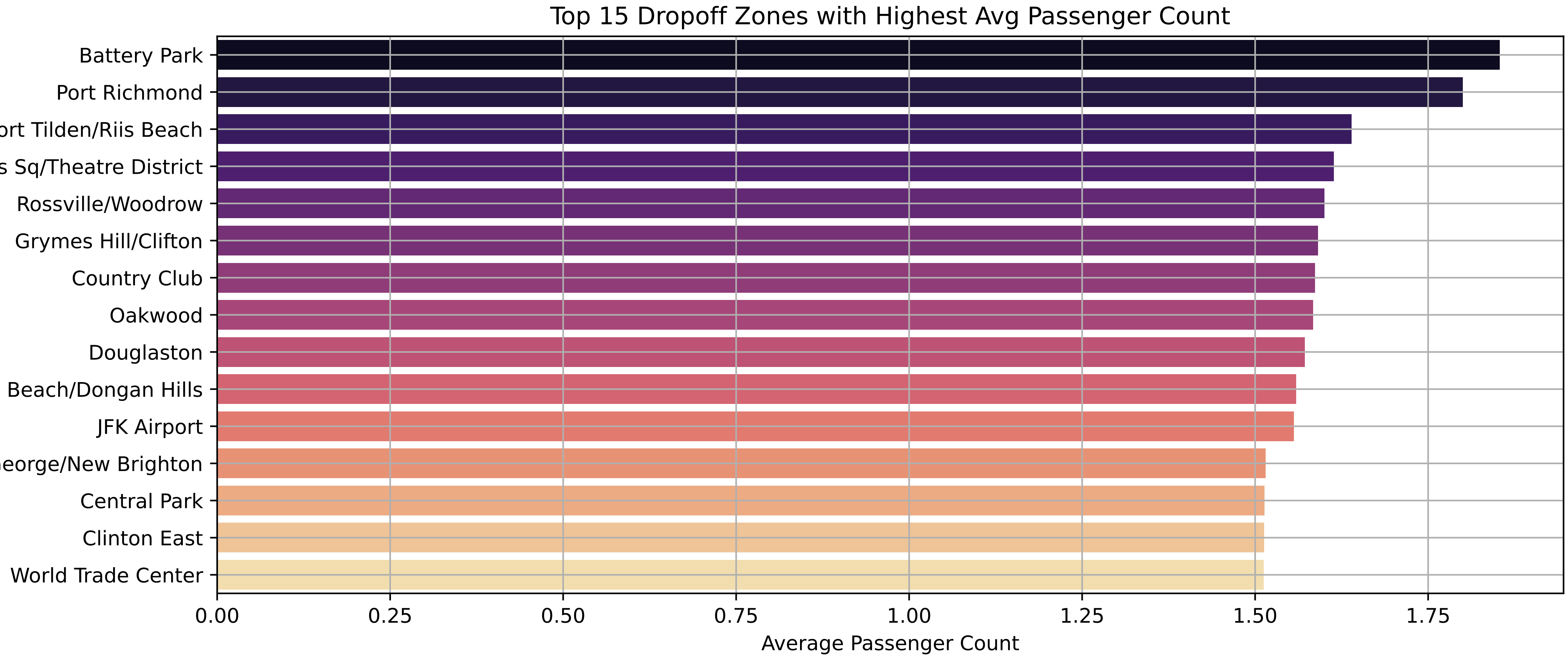

Average Passenger Count Heatmap (Hour vs. Day)
Day of the Week
Monday
1.33 1.37 1.33 1.37 1.23 1.27 1.31 1.32 1.34 1.38 1.41 1.39 1.39 1.41 1.41 1.42 1.42 1.42 1.41 1.43 1.46 1.47 1.46 1.42
Tuesday
1.34 1.40 1.38 1.46 1.40 1.29 1.34 1.31 1.29 1.35 1.39 1.39 1.42 1.43 1.44 1.45 1.43 1.39 1.35 1.41 1.38 1.38 1.39 1.33
Wednesday
1.41 1.38 1.41 1.39 1.33 1.28 1.31 1.34 1.47 1.43 1.46 1.46 1.50 1.51 1.52 1.53 1.53 1.51 1.48 1.49 1.47 1.50 1.50 1.45
Thursday
1.43 1.42 1.44 1.43 1.37 1.35 1.38 1.41 1.45 1.46 1.49 1.48 1.50 1.46 1.51 1.50 1.49 1.48 1.48 1.45 1.43 1.41 1.38 1.39
Friday
1.32 1.32 1.38 1.31 1.28 1.27 1.28 1.29 1.29 1.28 1.34 1.38 1.38 1.40 1.40 1.41 1.38 1.35 1.34 1.34 1.36 1.38 1.38 1.34
Saturday
1.38 1.37 1.45 1.46 1.29 1.30 1.30 1.33 1.28 1.31 1.35 1.33 1.35 1.37 1.39 1.41 1.38 1.36 1.32 1.34 1.34 1.36 1.35 1.34
Sunday
1.35 1.30 1.34 1.47 1.32 1.23 1.26 1.29 1.27 1.30 1.33 1.34 1.35 1.35 1.37 1.37 1.37 1.36 1.32 1.33 1.33 1.33 1.36 1.34
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of the Day
1.50
1.45
1.40
1.35
1.30
1.25

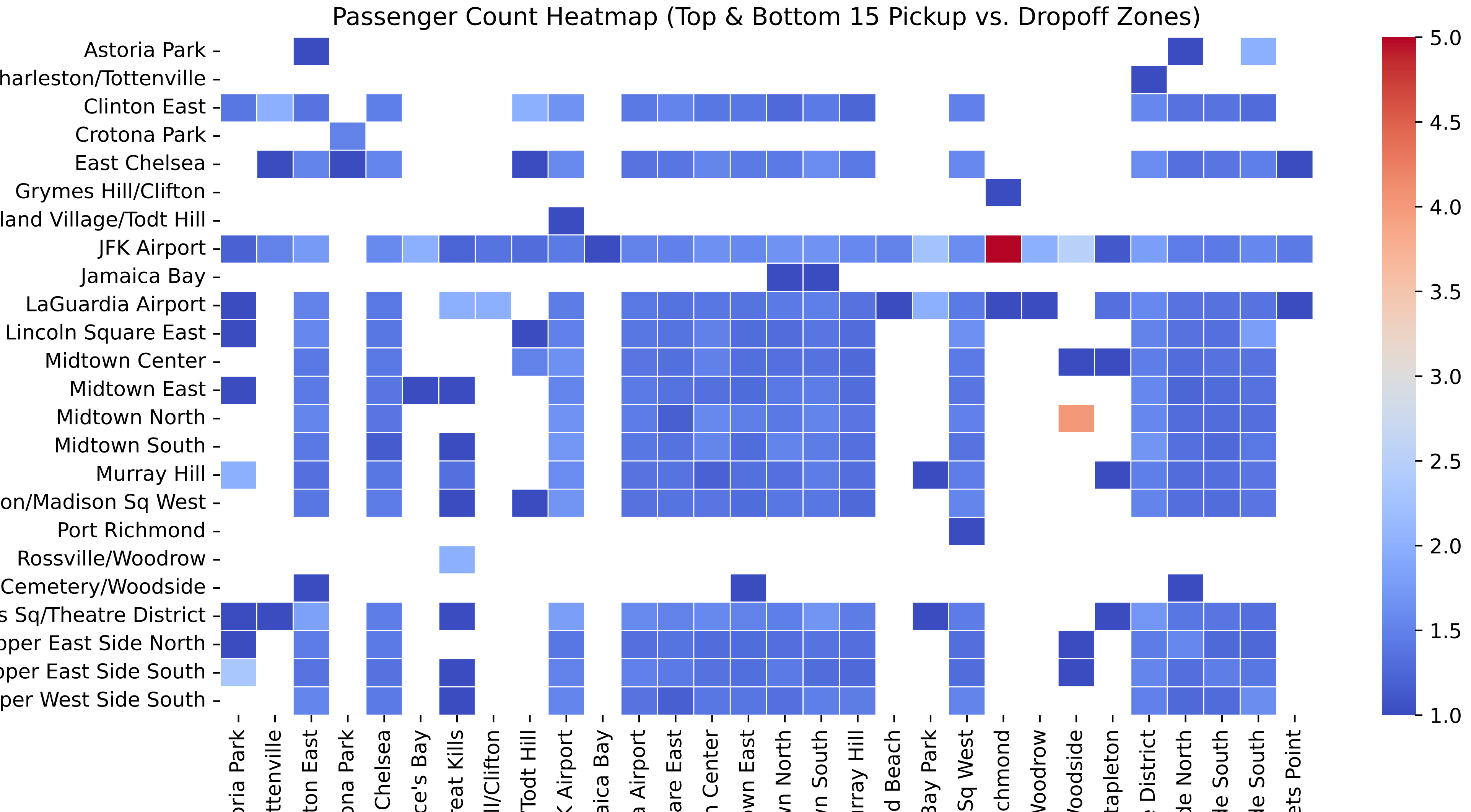
Conclusion/Remarks:

3.2.14 Showing the analysis the variation of passenger count across hours and days of the week

Average passenger count is heighest at 15:00 & 16:00 wifh value 1.53 PM
Average passenger count is heighest on SWeekends


Top 15 Pickup Zones with Highest Avg Passenger Count
Willets Point
Country Club
Red Hook
Battery Park
World Trade Center
Central Park
Randalls Island
Brooklyn Navy Yard
s Sq/Theatre District
Chinatown
eadows-Corona Park
Little Italy/NoLiTa
Crotona Park
Rikers Island
JFK Airport
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Average Passenger Count

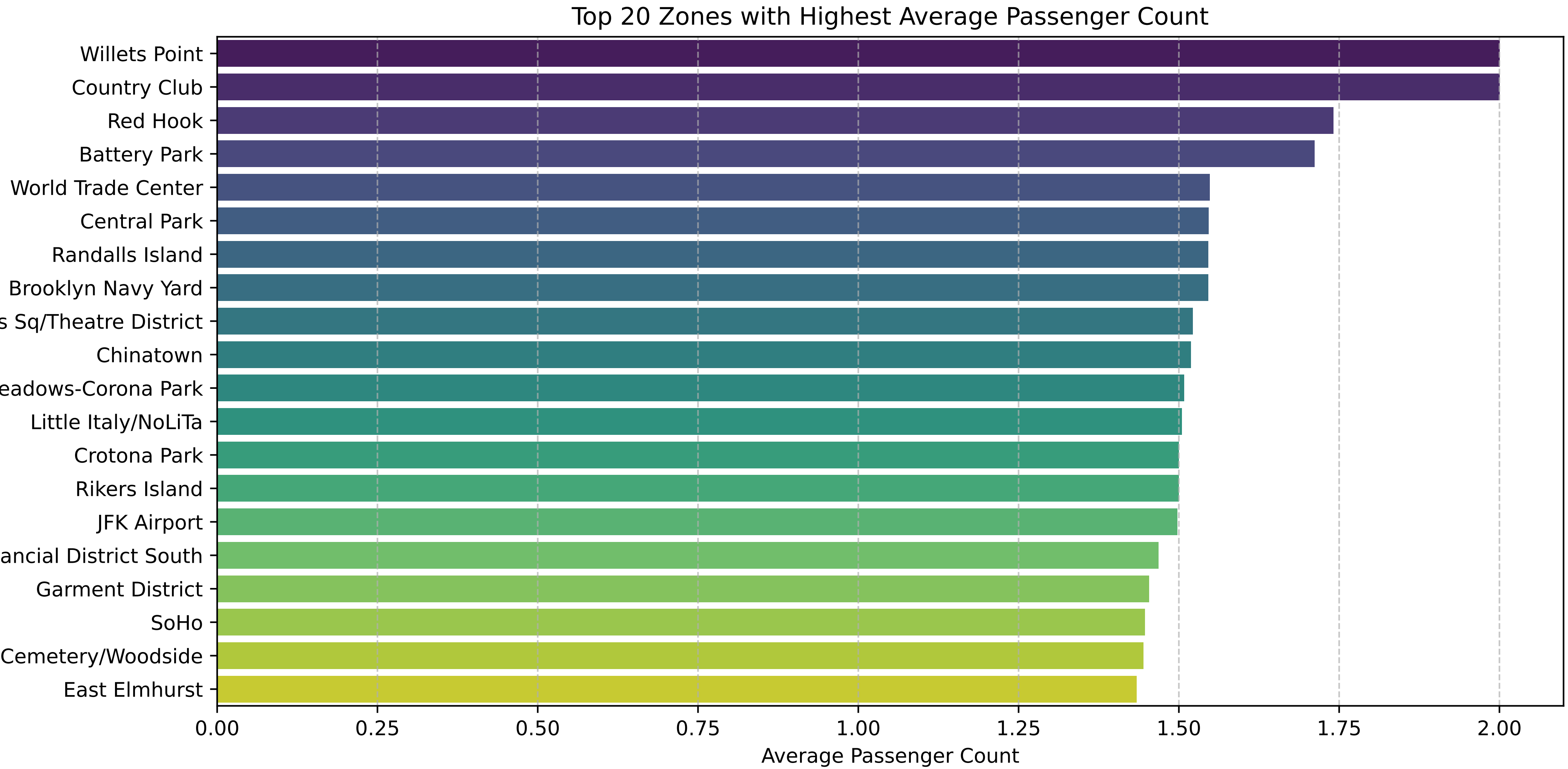
Top 15 Dropoff Zones with Highest Avg Passenger Count
Battery Park
Port Richmond
ort Tilden/Riis Beach
s Sq/Theatre District
Rossville/Woodrow
Grymes Hill/Clifton
Country Club
Oakwood
Douglaston
Beach/Dongan Hills
JFK Airport
eorge/New Brighton
Central Park
Clinton East
World Trade Center
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Average Passenger Count


Passenger Count Heatmap (Top & Bottom 15 Pickup vs. Dropoff Zones)
Astoria Park
harleston/Tottenville
Clinton East
Crotona Park
East Chelsea
Grymes Hill/Clifton
land Village/Todt Hill
JFK Airport
Jamaica Bay
LaGuardia Airport
Lincoln Square East
Midtown Center
Midtown East
Midtown North
Midtown South
Murray Hill
on/Madison Sq West
Port Richmond
Rossville/Woodrow
Cemetery/Woodside
s Sq/Theatre District
pper East Side North
pper East Side South
per West Side South
oria Park
ttenville
ton East
ona Park
Chelsea
ce's Bay
eat Kills
ll/Clifton
Todt Hill
K Airport
aica Bay
a Airport
are East
n Center
wn East
n North
n South
rray Hill
d Beach
Bay Park
Sq West
chmond
Woodrow
oodside
tapleton
District
de North
de South
de South
ets Point
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0

Conclusion/Remarks:

3.2.15 Showing the analysis the variation of passenger counts across zones

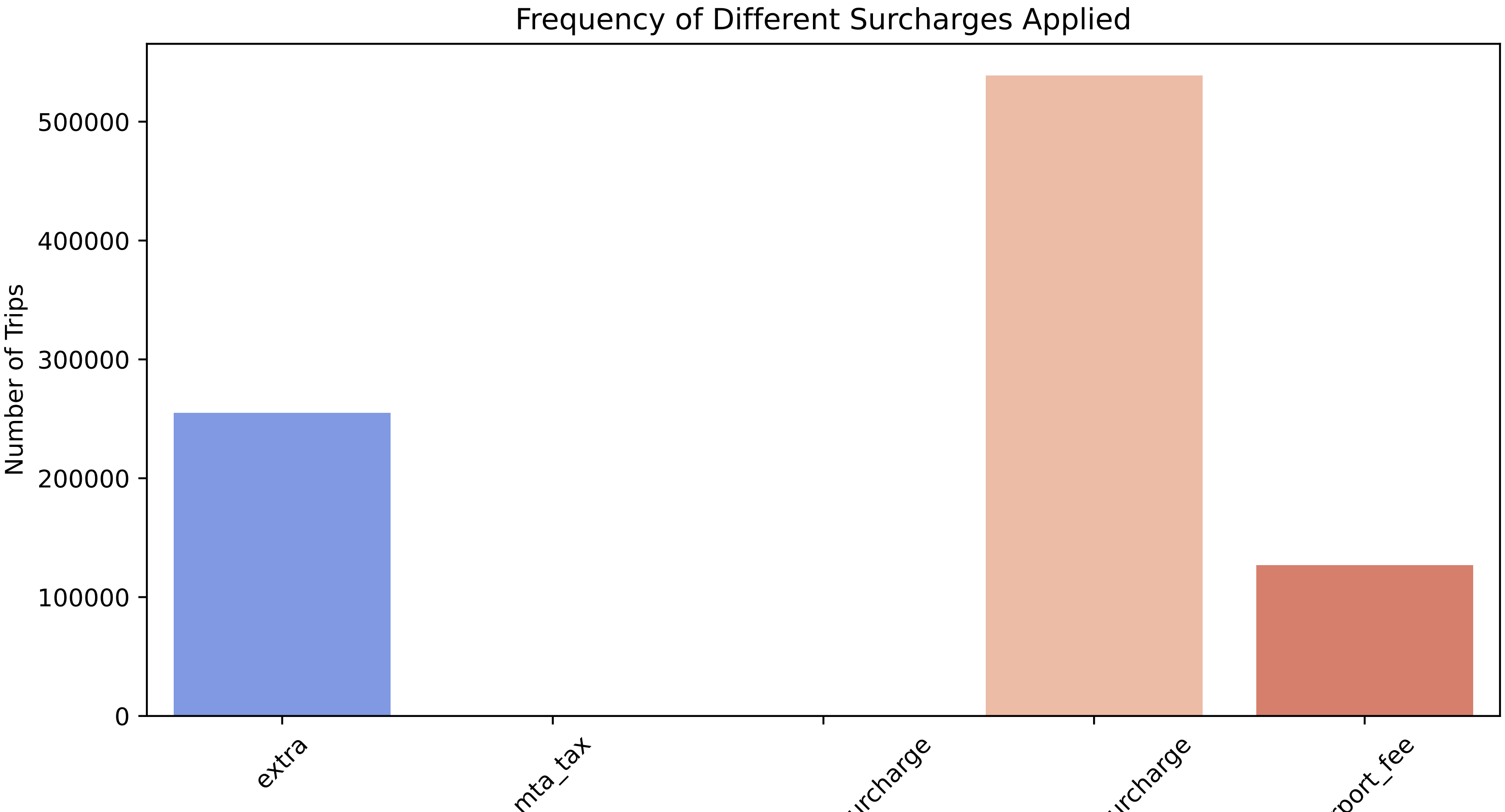
Willets Point and Country Club has the heighest pickups
Battery park has the heighest Dropoff
Point Richmond has the heighest in total Pickup and Dropoff together


Top 20 Zones with Highest Average Passenger Count
Willets Point
Country Club
Red Hook
Battery Park
World Trade Center
Central Park
Randalls Island
Brooklyn Navy Yard
s Sq/Theatre District
Chinatown
eadows-Corona Park
Little Italy/NoLiTa
Crotona Park
Rikers Island
JFK Airport
ancial District South
Garment District
SoHo
Cemetery/Woodside
East Elmhurst
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Average Passenger Count

Conclusion/Remarks:

3.2.15 Showing the analysis the average passenger counts per trip across pickup zones

Willets Point is the heighest with value 2

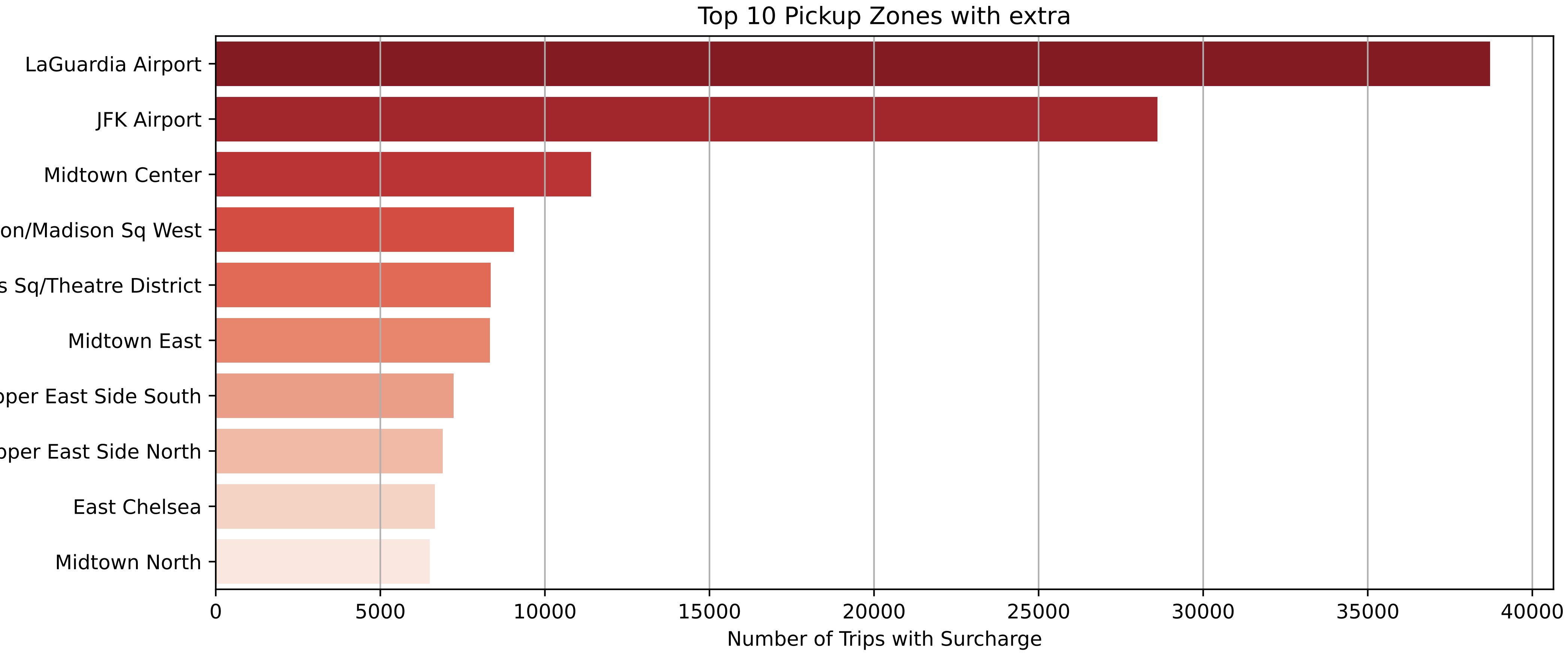


Frequency of Different Surcharges Applied

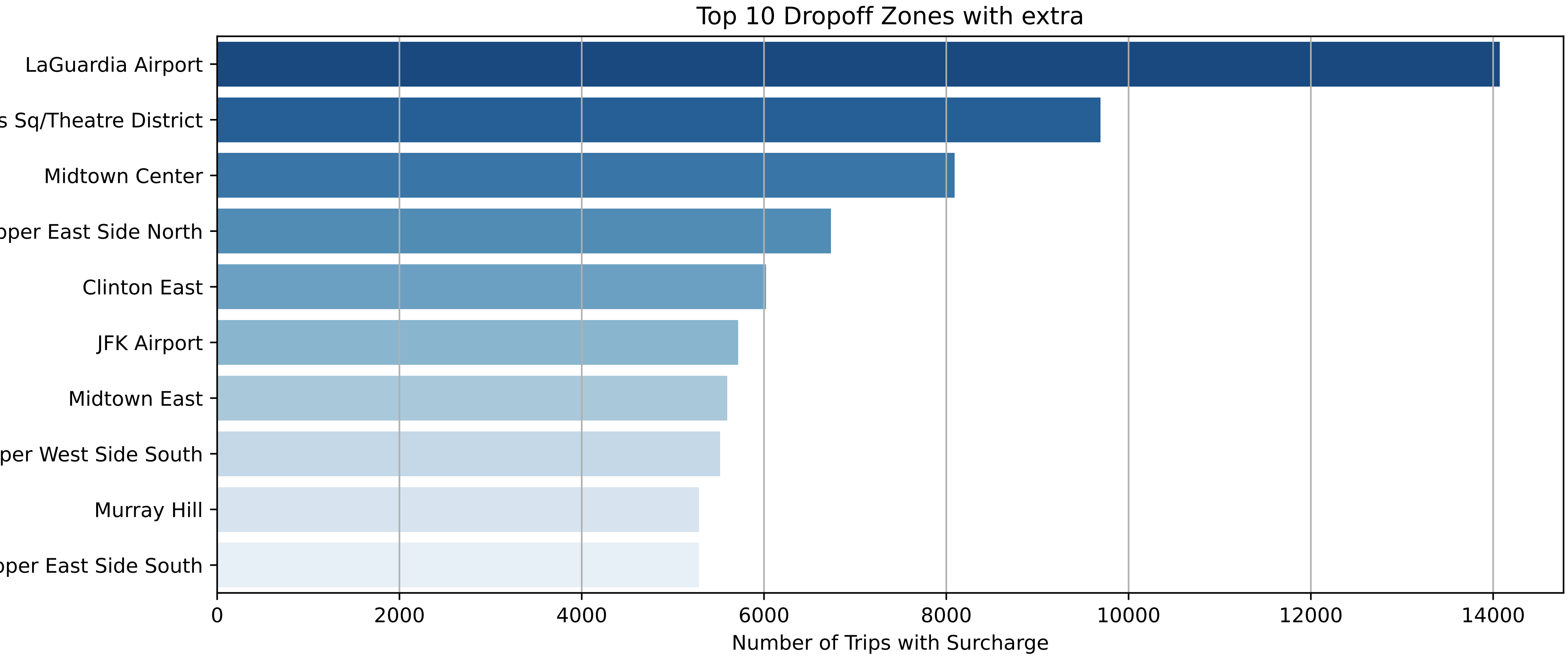
Conclusion/Remarks:

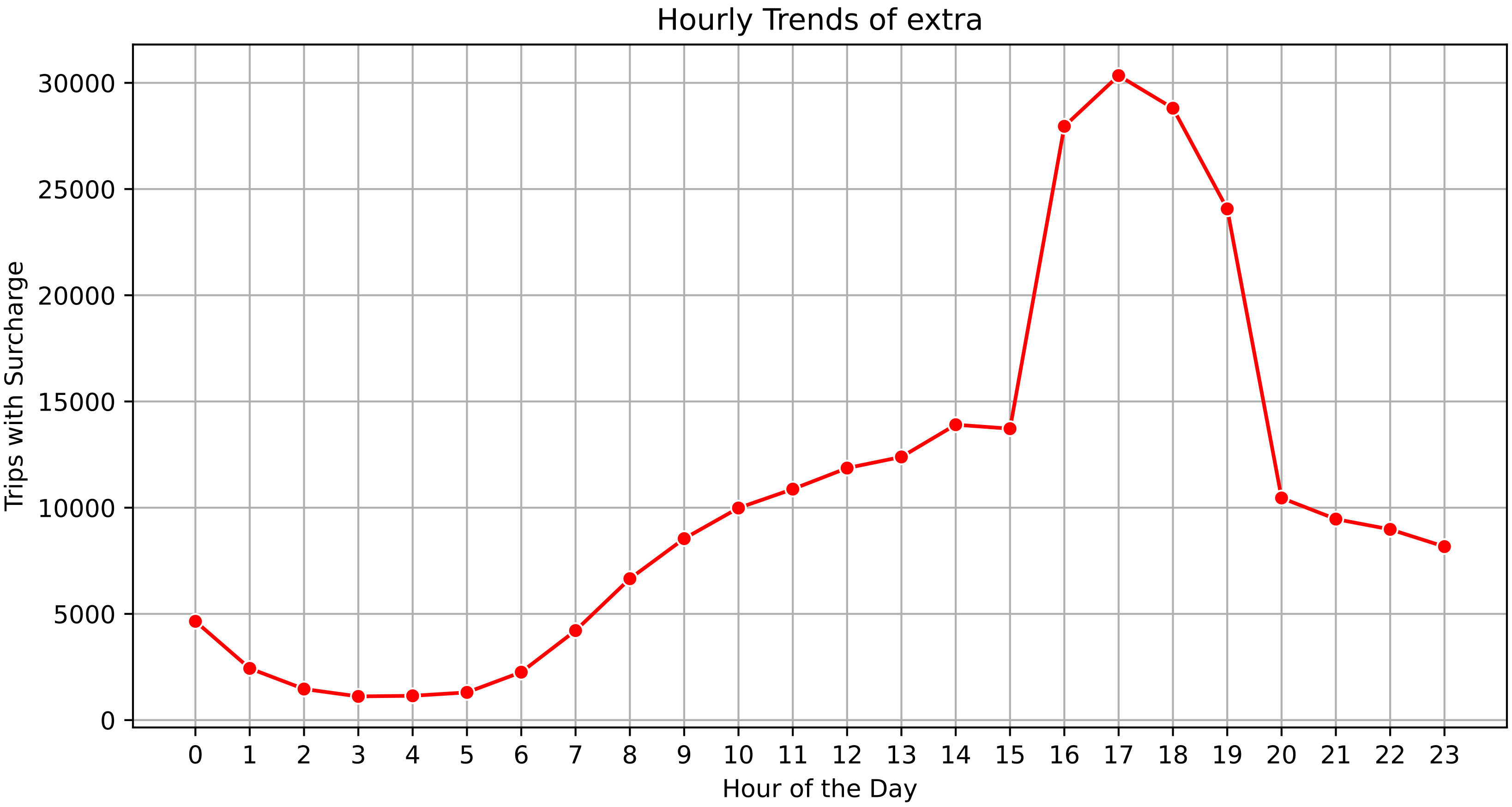
3.2.16 Showing the analysis of surcharges applied more frequently

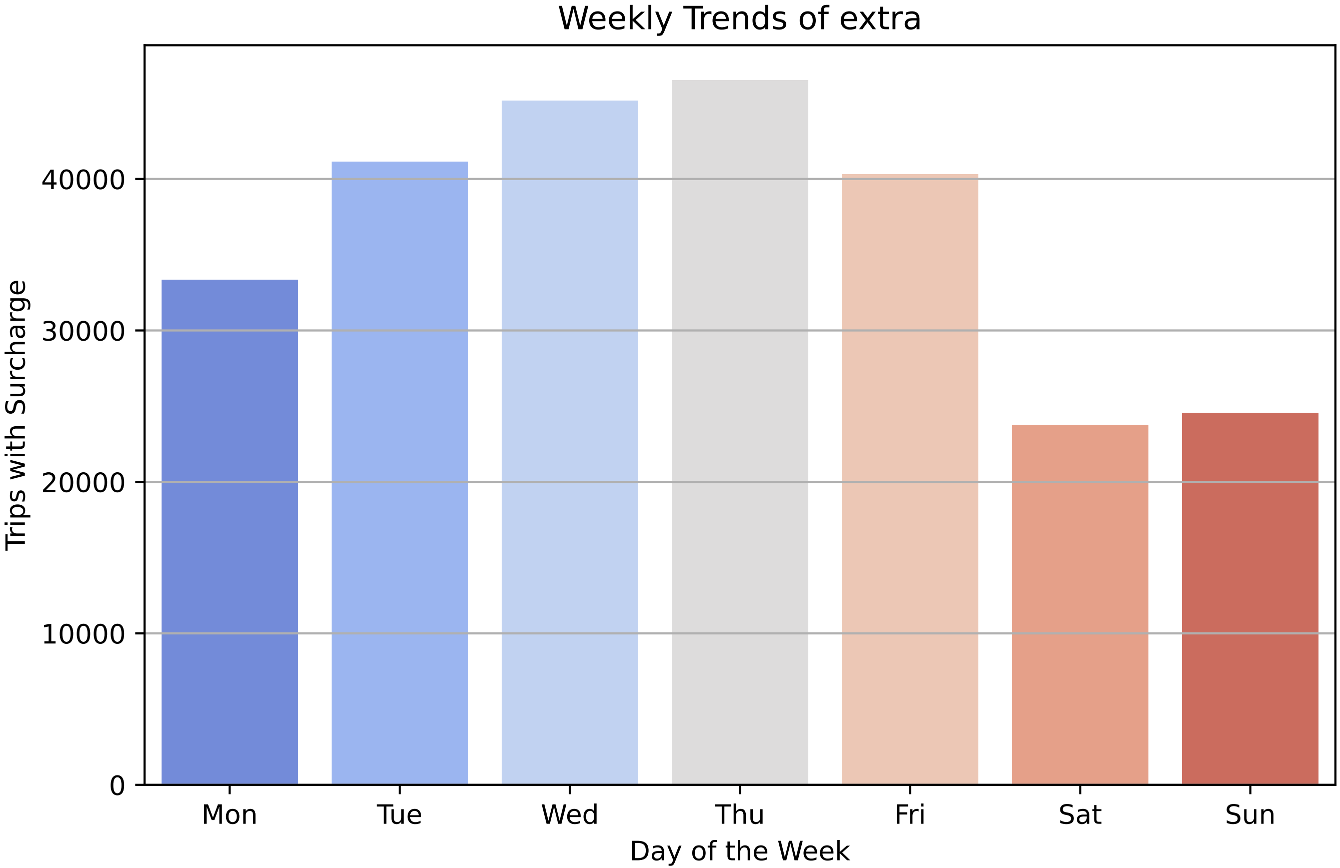
congestion_surcharge is applied maximum

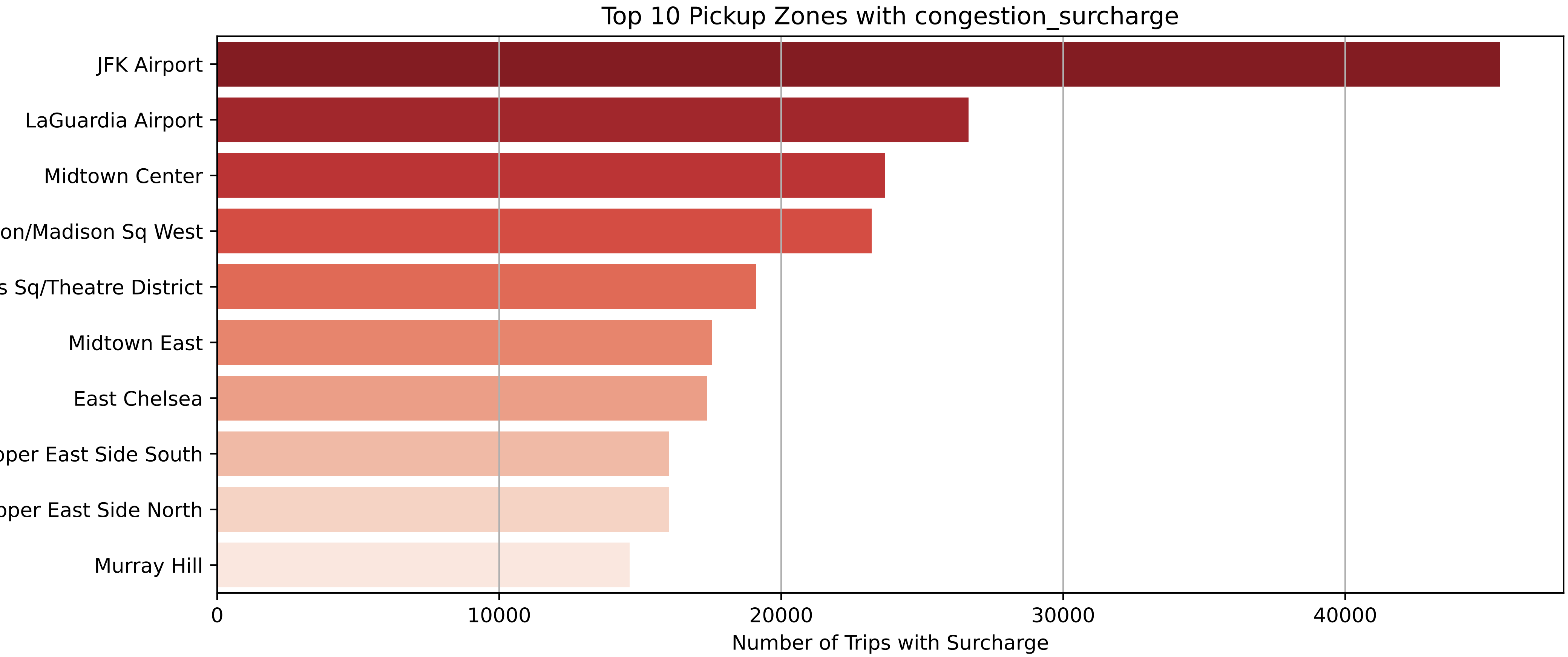


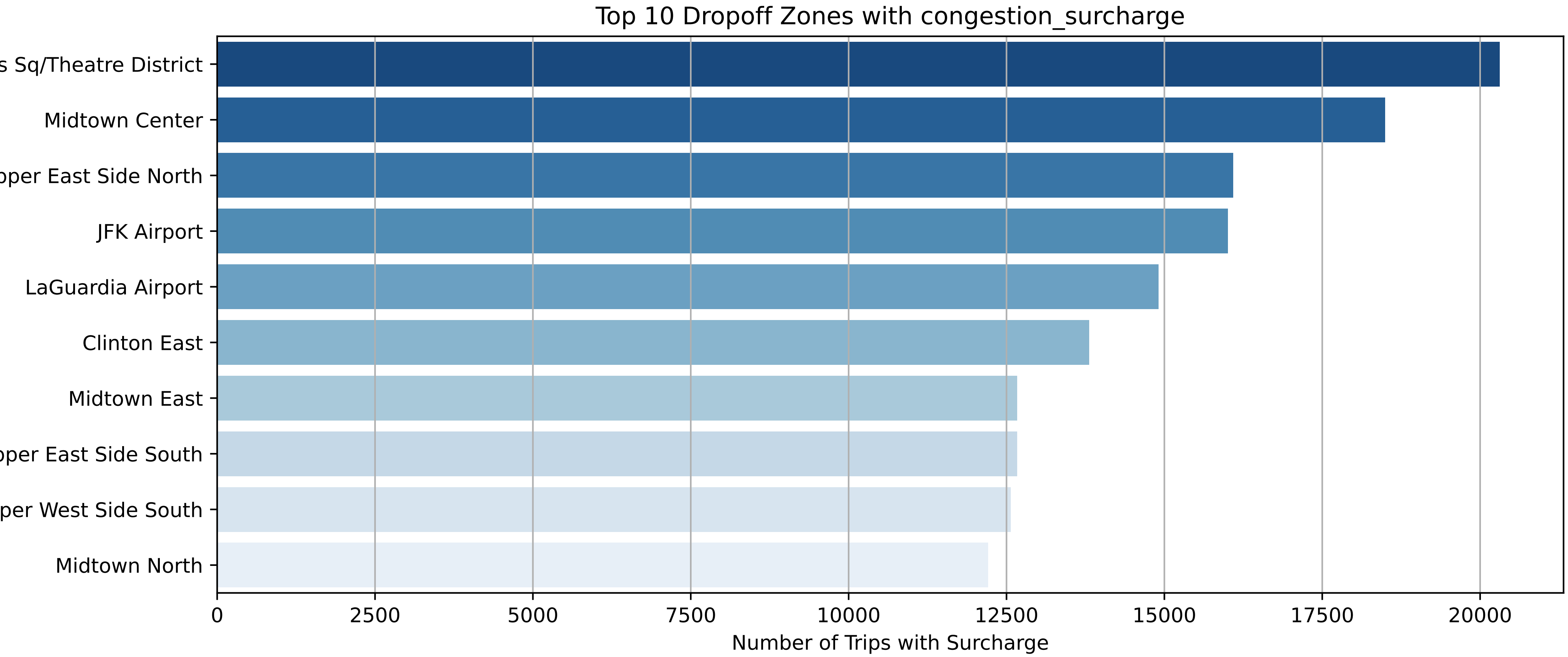
Top 10 Pickup Zones with extra
LaGuardia Airport
JFK Airport
Midtown Center
on/Madison Sq West
s Sq/Theatre District
Midtown East
pper East Side South
pper East Side North
East Chelsea
Midtown North
0
5000
10000
15000
20000
25000
30000
35000
40000
Number of Trips with Surcharge


Top 10 Dropoff Zones with extra
LaGuardia Airport
s Sq/Theatre District
Midtown Center
per East Side North
Clinton East
JFK Airport
Midtown East
per West Side South
Murray Hill
per East Side South
0
2000
4000
6000
8000
10000
12000
14000
Number of Trips with Surcharge

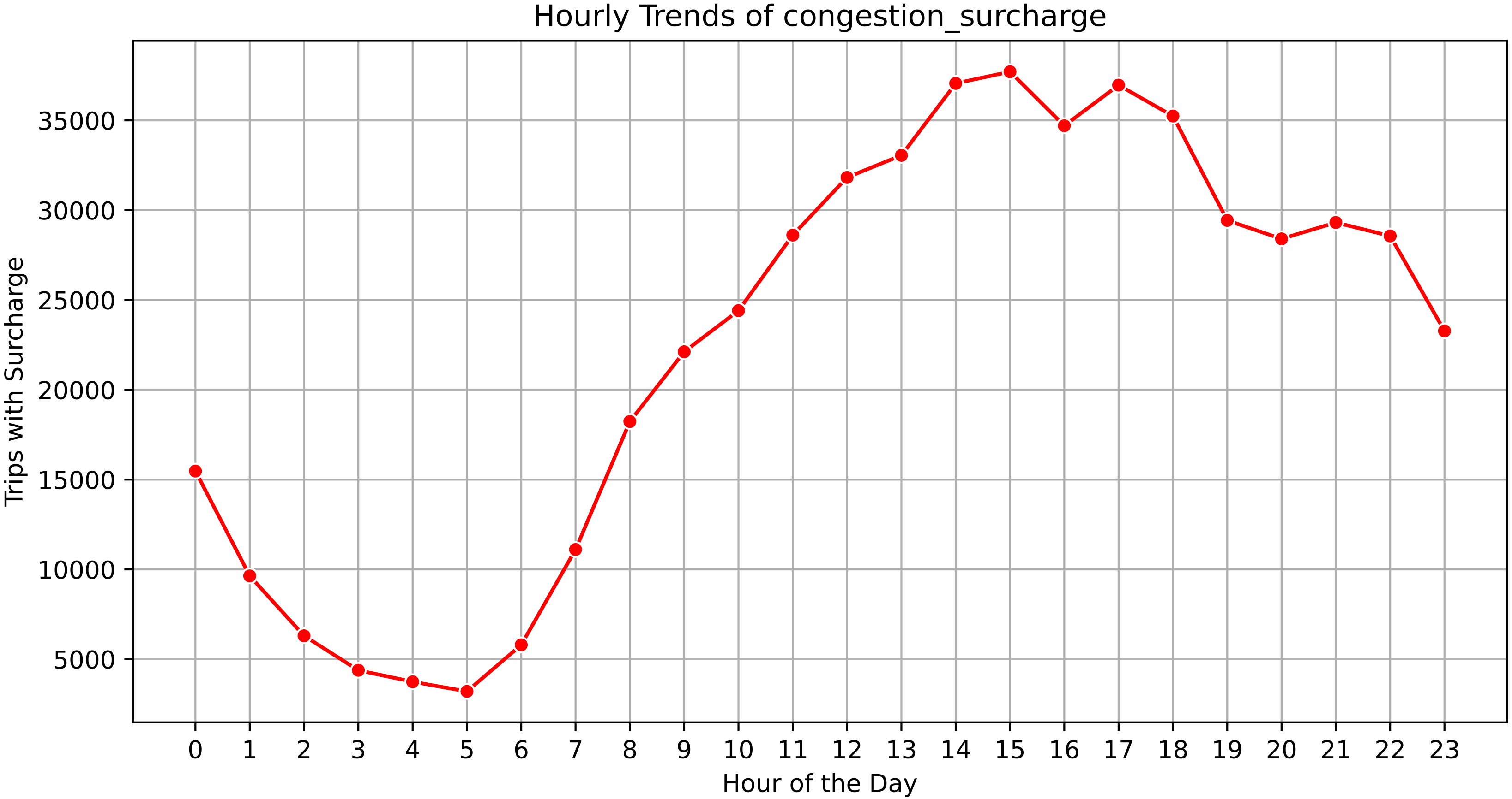

Hourly Trends of extra
Trips with Surcharge
0
5000
10000
15000
20000
25000
30000
Hour of the Day
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

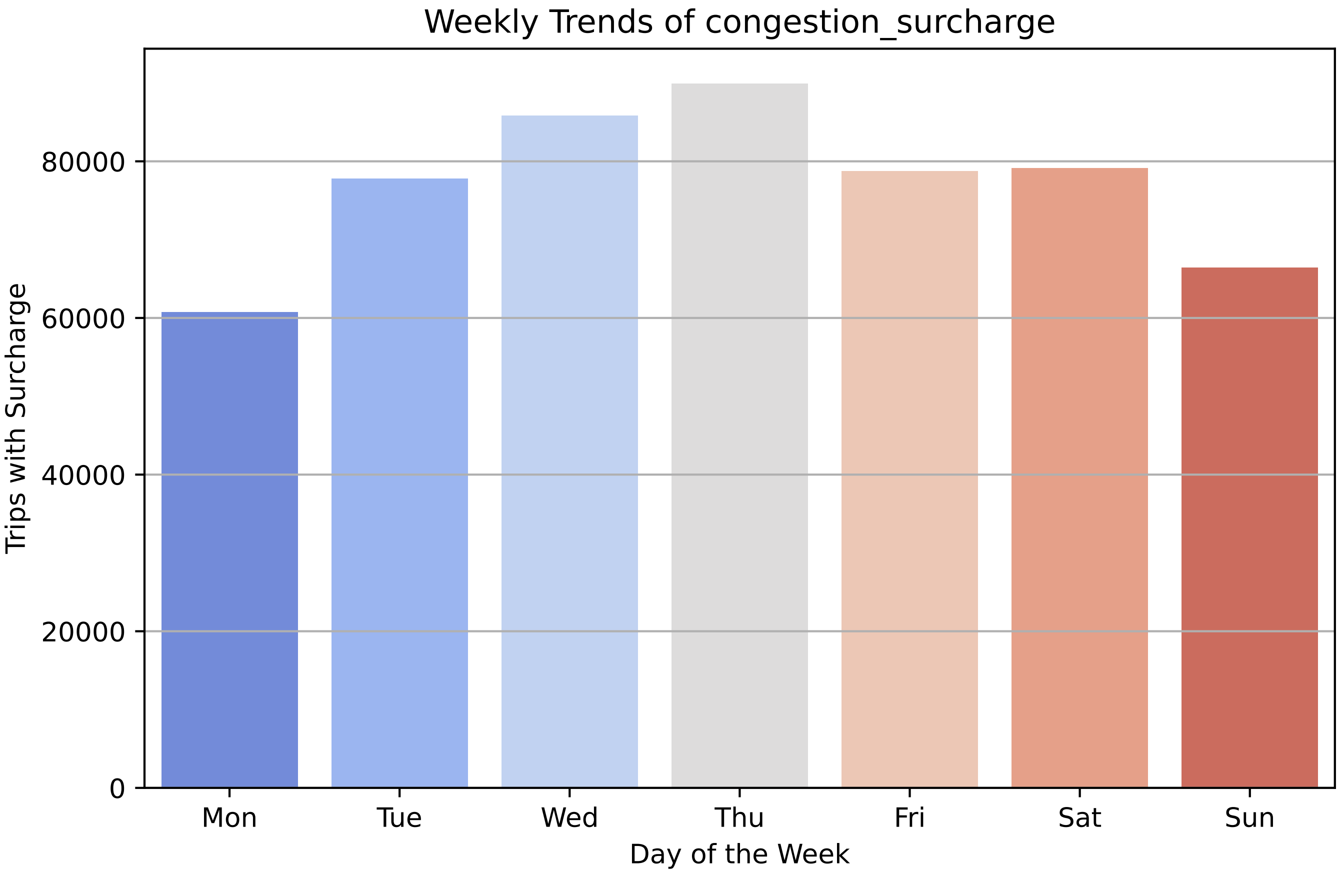

Weekly Trends of extra
Trips with Surcharge
0
10000
20000
30000
40000
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of the Week

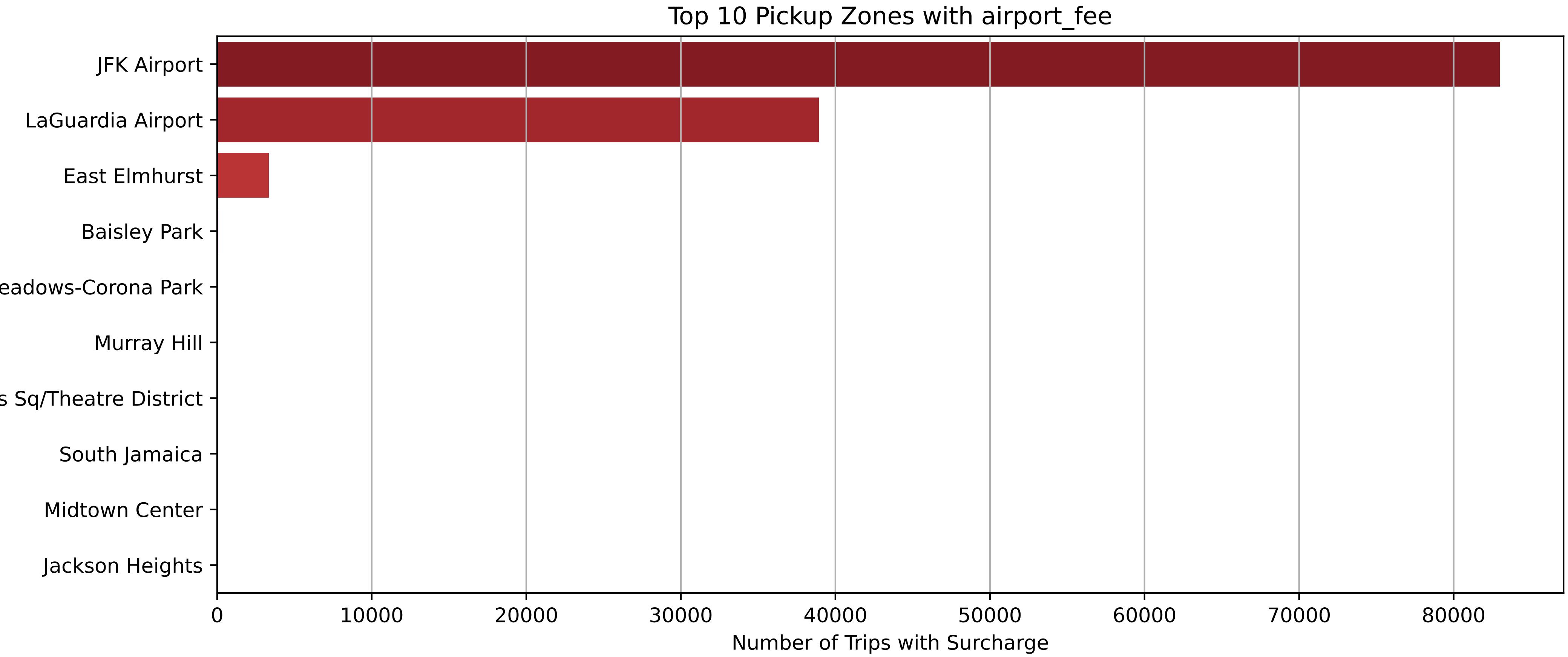

Top 10 Pickup Zones with congestion_surcharge
JFK Airport
LaGuardia Airport
Midtown Center
on/Madison Sq West
s Sq/Theatre District
Midtown East
East Chelsea
pper East Side South
pper East Side North
Murray Hill
0
10000
20000
30000
40000
Number of Trips with Surcharge

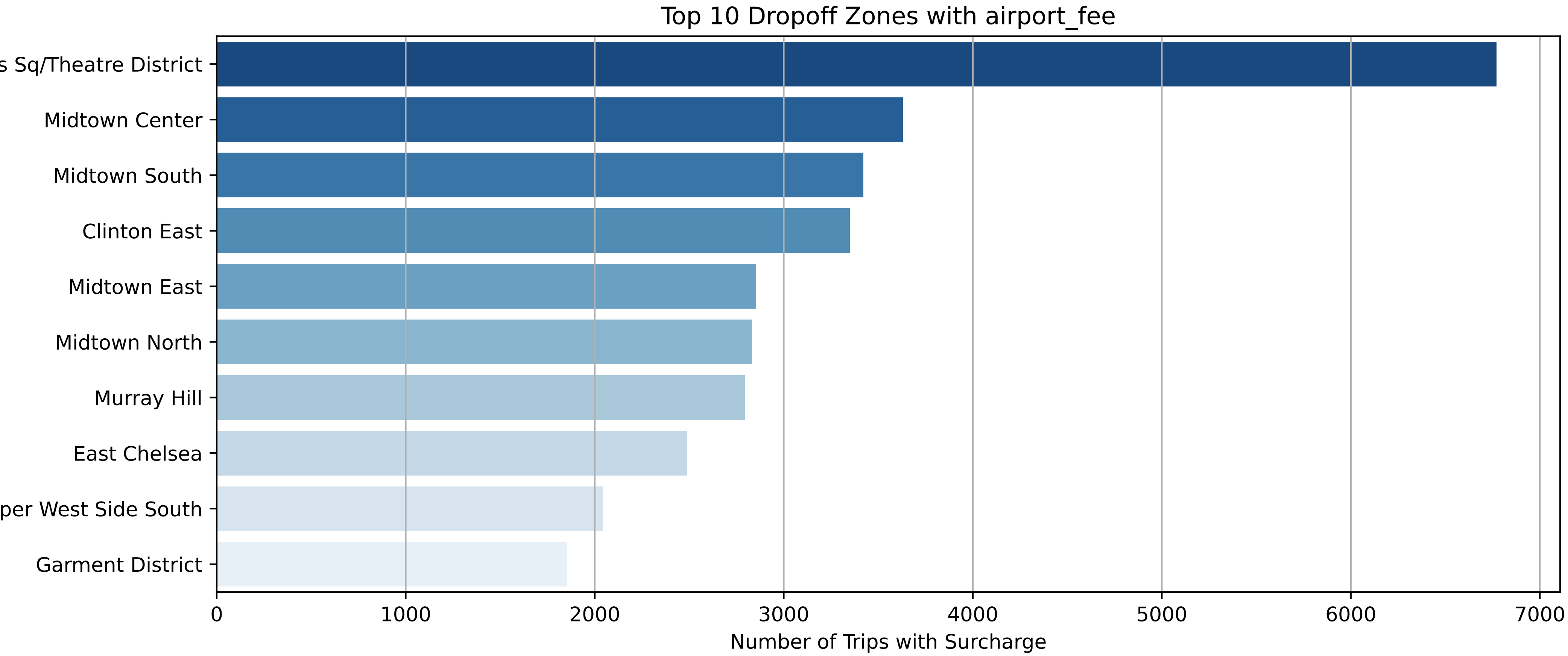


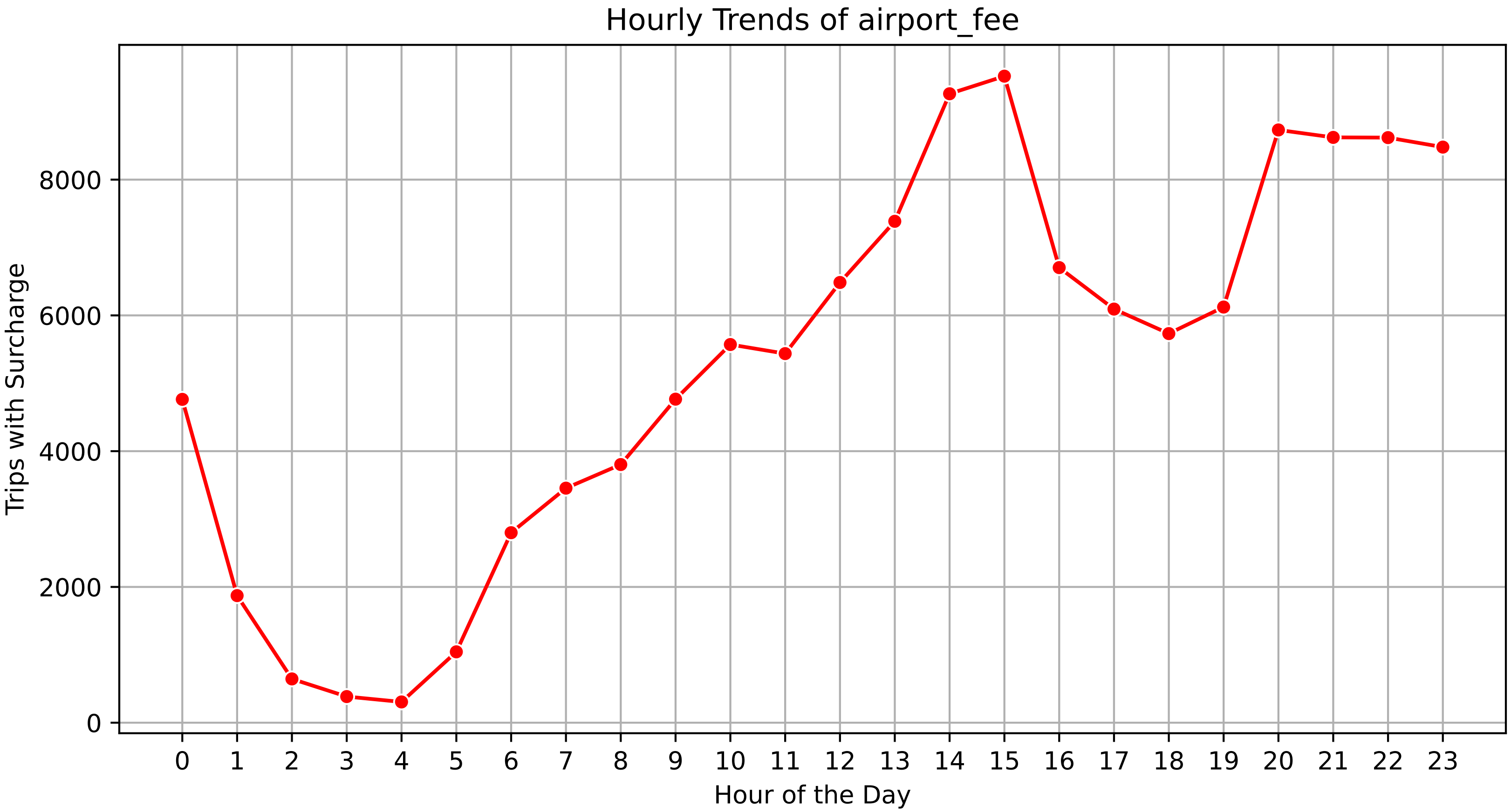
Top 10 Dropoff Zones with congestion_surcharge
s Sq/Theatre District
Midtown Center
pper East Side North
JFK Airport
LaGuardia Airport
Clinton East
Midtown East
pper East Side South
per West Side South
Midtown North
0
2500
5000
7500
10000
12500
15000
17500
20000
Number of Trips with Surcharge

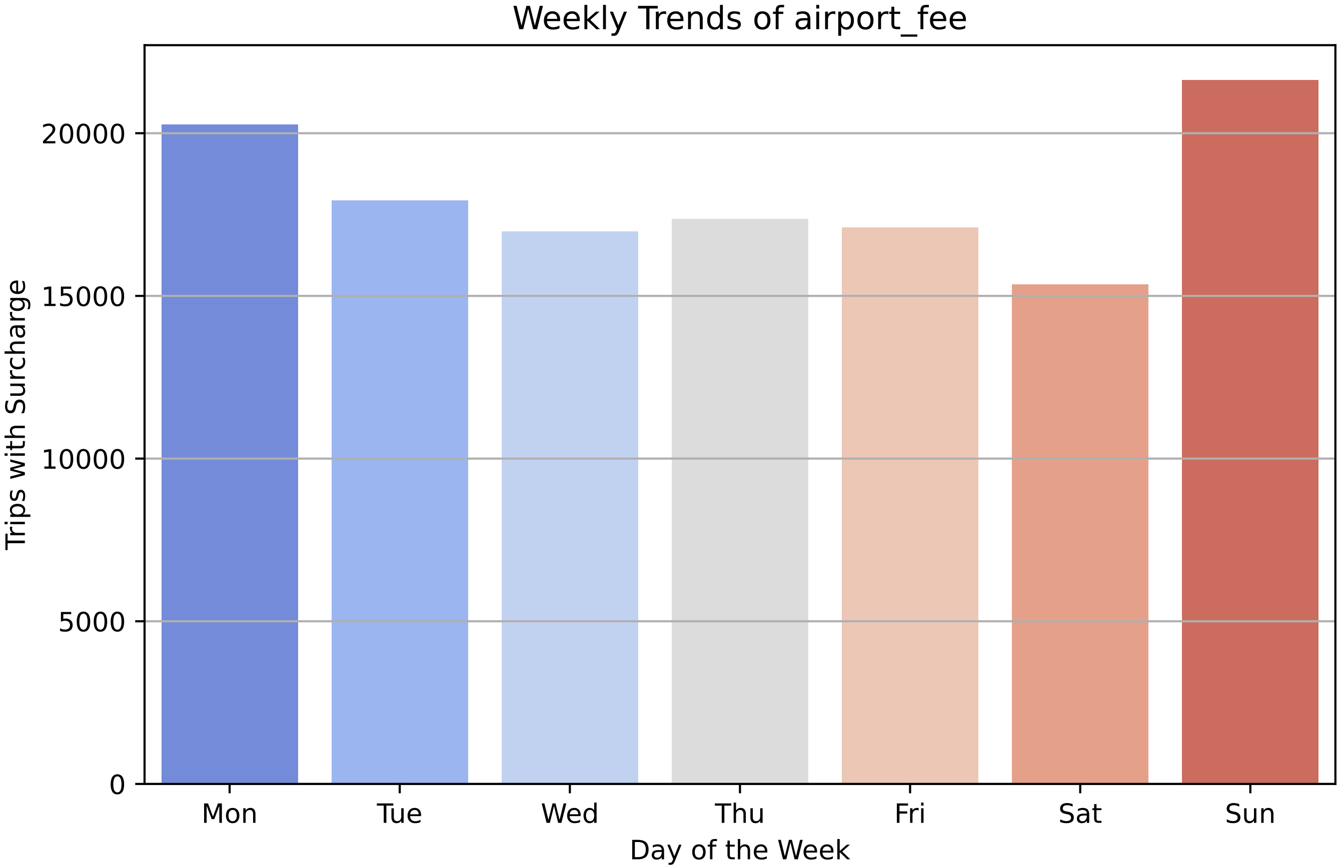

Hourly Trends of congestion_surcharge
Trips with Surcharge
35000
30000
25000
20000
15000
10000
5000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Hour of the Day


Weekly Trends of congestion_surcharge
Trips with Surcharge
0
20000
40000
60000
80000
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of the Week


Top 10 Pickup Zones with airport_fee
JFK Airport
LaGuardia Airport
East Elmhurst
Baisley Park
eadows-Corona Park
Murray Hill
s Sq/Theatre District
South Jamaica
Midtown Center
Jackson Heights
0
10000
20000
30000
40000
50000
60000
70000
80000
Number of Trips with Surcharge


Top 10 Dropoff Zones with airport_fee
s Sq/Theatre District
Midtown Center
Midtown South
Clinton East
Midtown East
Midtown North
Murray Hill
East Chelsea
per West Side South
Garment District
0
1000
2000
3000
4000
5000
6000
7000
Number of Trips with Surcharge


Hourly Trends of airport_fee
Trips with Surcharge
0
2000
4000
6000
8000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Hour of the Day

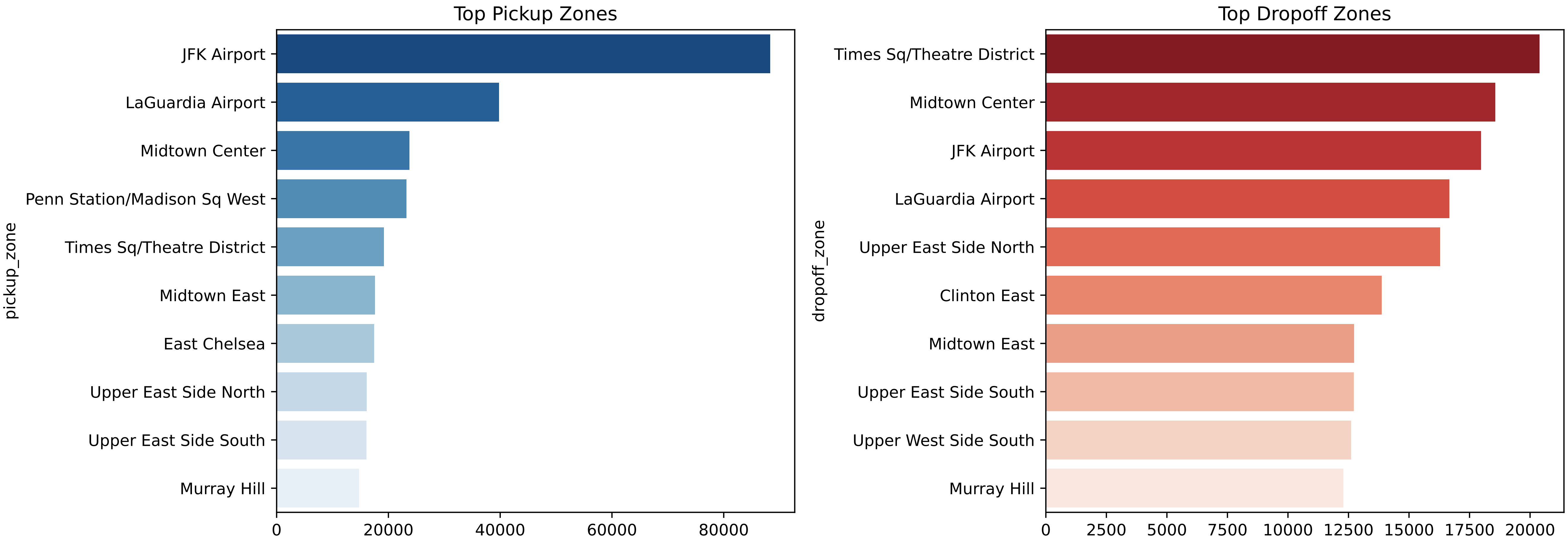

Weekly Trends of airport_fee
Trips with Surcharge
20000
15000
10000
5000
0
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of the Week

Conclusion/Remarks:

3.2.16 Showing the analysis of Zone wise surcharges applied more frequently

LaGuardia Airport is Pickup Zone where maximum surcharge type extra applied
LaGuardia Airport is Dropoff Zone where maximum surcharge type extra applied
The maximum surcharge type extra applied at around 17:00 PM
The maximum surcharge type extra applied at on Thursday

JFK Airport is Pickup Zone where maximum surcharge type congestion_surcharge applied
Times Sq/Theatre District is Dropoff Zone where maximum surcharge type congestion_surcharge applied
The maximum surcharge type congestion_surcharge applied at around 15:00 PM

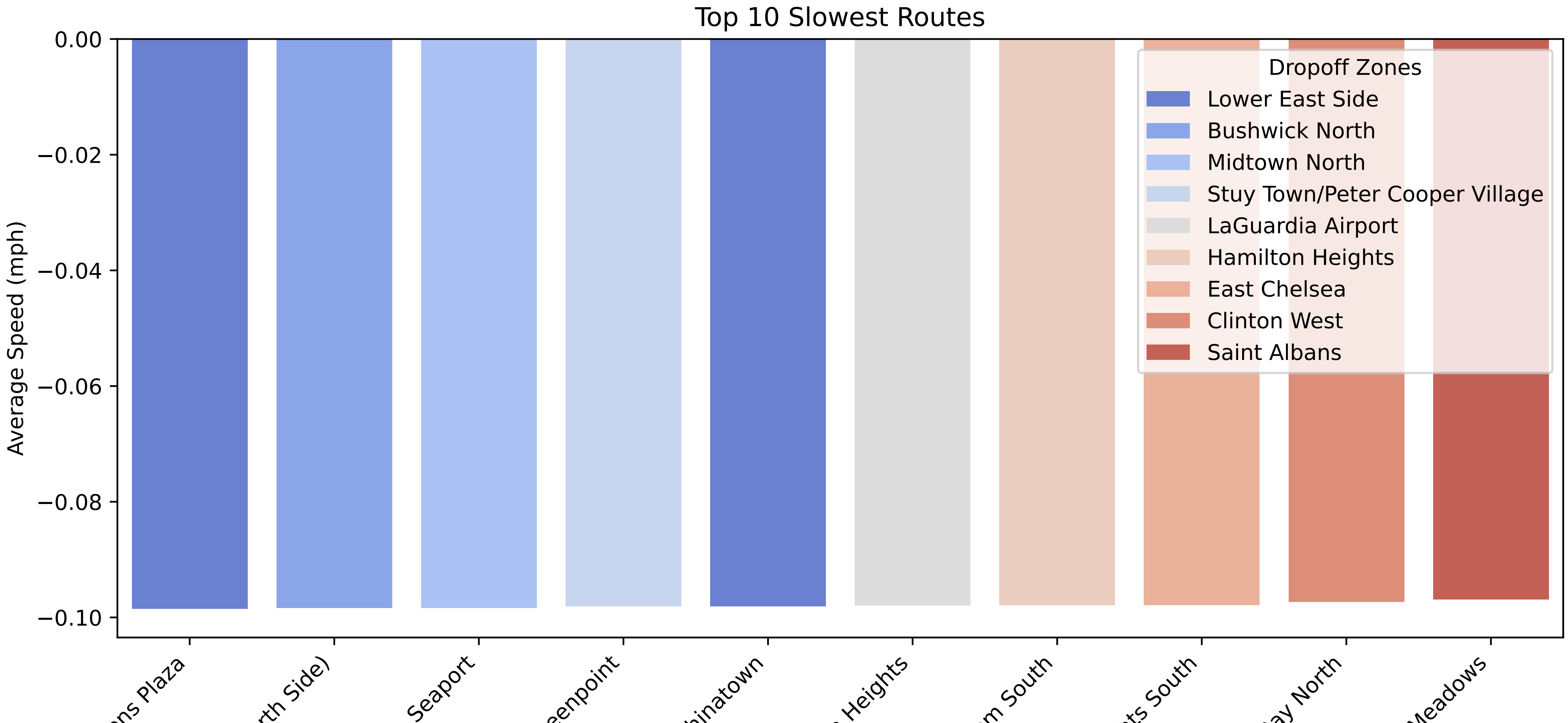

Top Pickup Zones
pickup_zone
JFK Airport
LaGuardia Airport
Midtown Center
Penn Station/Madison Sq West
Times Sq/Theatre District
Midtown East
East Chelsea
Upper East Side North
Upper East Side South
Murray Hill
0
20000
40000
60000
80000
Top Dropoff Zones
dropoff_zone
Times Sq/Theatre District
Midtown Center
JFK Airport
LaGuardia Airport
Upper East Side North
Clinton East
Midtown East
Upper East Side South
Upper West Side South
Murray Hill
0
2500
5000
7500
10000
12500
15000
17500
20000

Conclusion/Remarks:

4.1.1 Identify and Optimize Routes with High Demand

Insights from df_merged:
Analyze the busiest pickup/dropoff zones by hour, day, and season.
Identify peak demand locations (e.g., airports, entertainment districts, business hubs).

Optimization Strategy:
Dynamic Routing System: Implement real-time route optimization to direct available cabs to high-demand zones.
Pre-positioning Strategy: During peak hours, dispatch cabs ahead of demand to hot zones to reduce wait times.

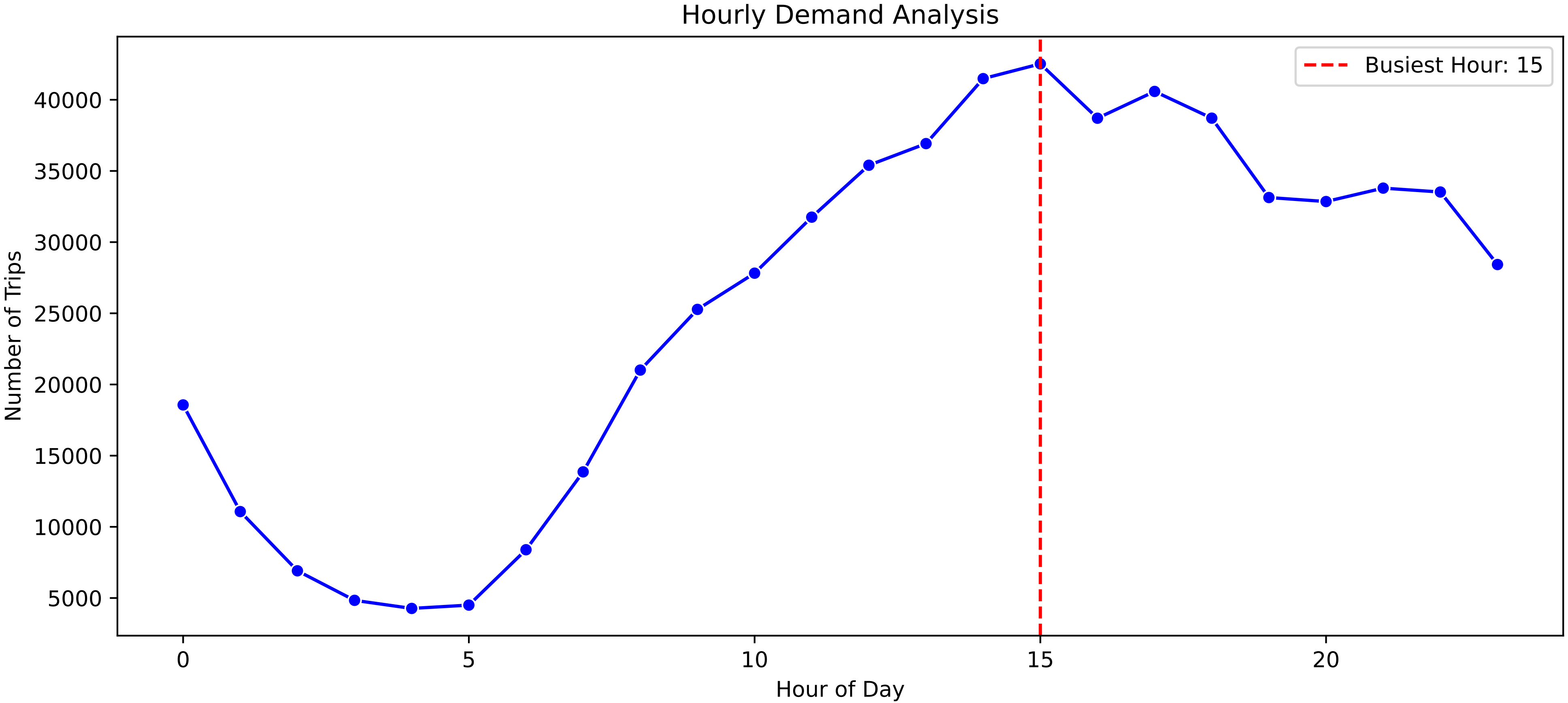

Top 10 Slowest Routes
Average Speed (mph)
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
Dropoff Zones
Lower East Side
Bushwick North
Midtown North
Stuy Town/Peter Cooper Village
LaGuardia Airport
Hamilton Heights
East Chelsea
Clinton West
Saint Albans
ns Plaza
rth Side)
Seaport
eenpoint
hinatown
Heights
m South
ts South
ay North
Meadows

Conclusion/Remarks:

4.1.1 Reduce Idle Time by Identifying Low-Demand Areas

Insights from df_merged:
Identify zones with low trip frequency but high supply.
Detect times when cabs wait too long for trips in specific zones.

Optimization Strategy:
  Reallocate Vehicles Dynamically: Move cabs from low-demand to high-demand areas.
  Surge Pricing Awareness: If demand is low in an area, reduce prices or offer discounts to encourage trips.


Hourly Demand Analysis
Busiest Hour: 15
Number of Trips
40000
35000
30000
25000
20000
15000
10000
5000
0
5
10
15
20
Hour of Day

Conclusion/Remarks:

4.1.1 Optimize Dispatching for Night vs. Daytime Hours

Insights from df_merged:
Night hours (11 PM - 5 AM) may have fewer but longer trips (e.g., airport rides).
Daytime hours may have shorter, frequent rides (e.g., commuters).

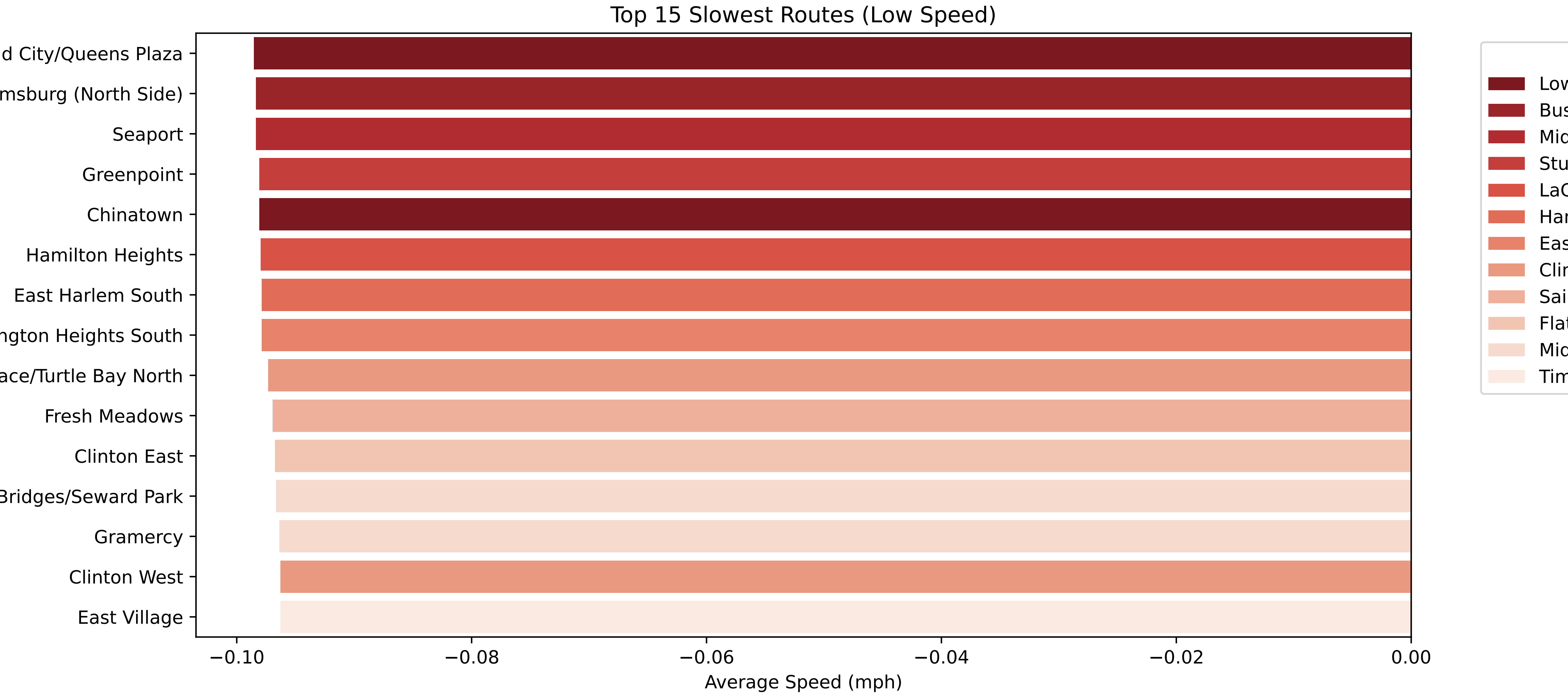
Optimization Strategy:
- Adjust Fleet Size Based on Hourly Trends:
  - Night: Allocate more cars near airports, railway stations, or nightlife hubs.
  - Day: Focus on business districts and commuter-heavy zones.
- Offer Nighttime Incentives: Encourage drivers to operate during low-driver hours with

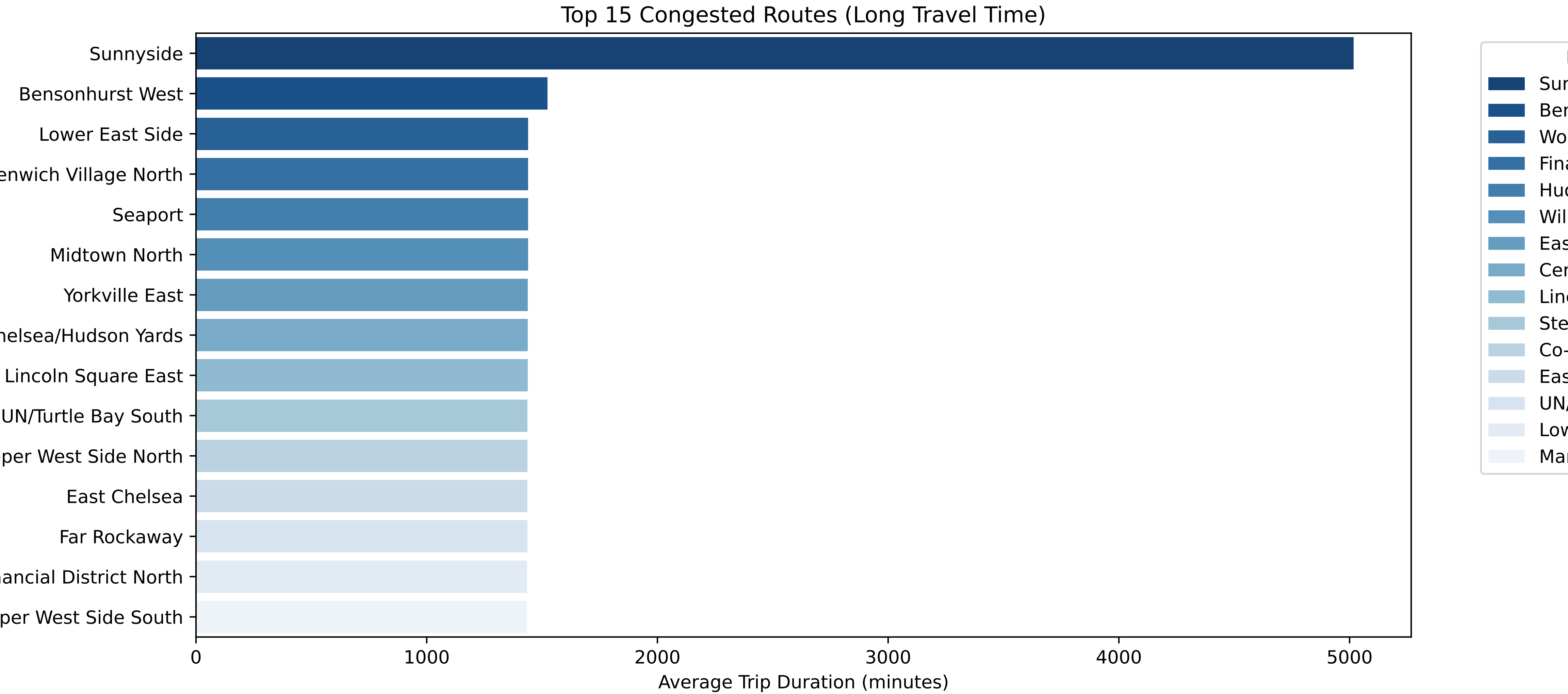
Conclusion/Remarks:

4.1.1 Improve Route Efficiency for Slow Routes

Insights from df_merged:
- Calculate average speed per route and hour to find slow zones.
- Identify roads with frequent congestion or delays.

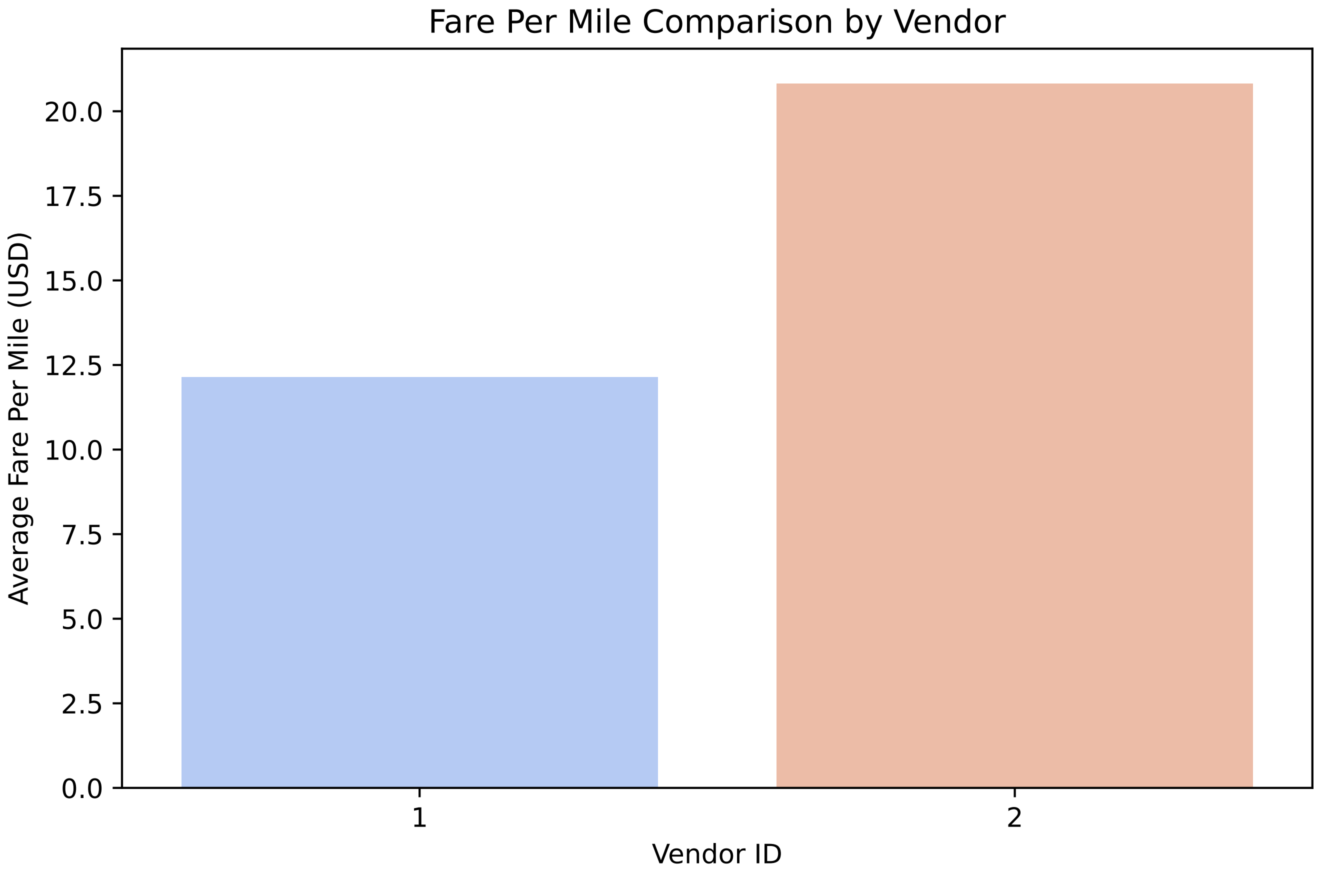
Optimization Strategy:
- Navigation & Traffic Data Integration:
  - Use real-time traffic data to reroute cabs away from slow zones.
  - Implement AI-based predictive routing based on historical traffic trends.


Top 15 Slowest Routes (Low Speed)
d City/Queens Plaza
msburg (North Side)
Seaport
Greenpoint
Chinatown
Hamilton Heights
East Harlem South
ngton Heights South
ace/Turtle Bay North
Fresh Meadows
Clinton East
Bridges/Seward Park
Gramercy
Clinton West
East Village
−0.10
−0.08
−0.06
−0.04
−0.02
0.00
Average Speed (mph)
Low
Bus
Mic
Stu
LaG
Har
Eas
Clin
Sai
Flat
Mic
Tim


Top 15 Congested Routes (Long Travel Time)
Sunnyside
Bensonhurst West
Lower East Side
enwich Village North
Seaport
Midtown North
Yorkville East
helsea/Hudson Yards
Lincoln Square East
UN/Turtle Bay South
per West Side North
East Chelsea
Far Rockaway
ancial District North
per West Side South
0
1000
2000
3000
4000
5000
Average Trip Duration (minutes)
Sur
Ben
Wo
Fina
Huc
Wil
Eas
Cer
Linc
Ste
Co-
Eas
UN/
Low
Mar

Conclusion/Remarks:

4.1.1 SHowing Top 15 Slowest Routes and Congested Routes to implement strategy to generate more revenue


Fare Per Mile Comparison by Vendor
Average Fare Per Mile (USD)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
1
2
Vendor ID

Conclusion/Remarks:

4.1.1 Increase Revenue Per Trip by Optimizing Fare & Tip Collection

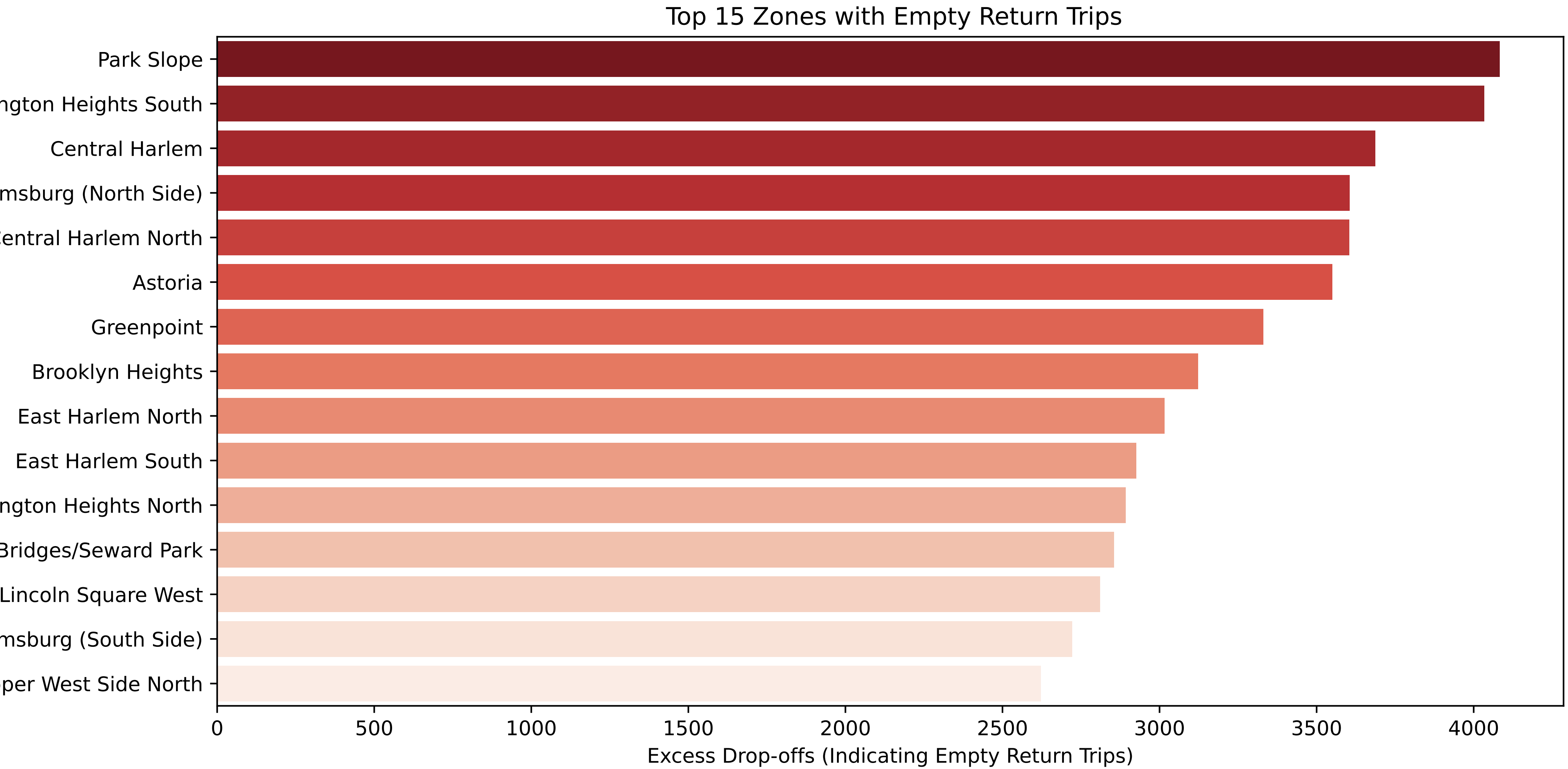
Insights from df_merged:

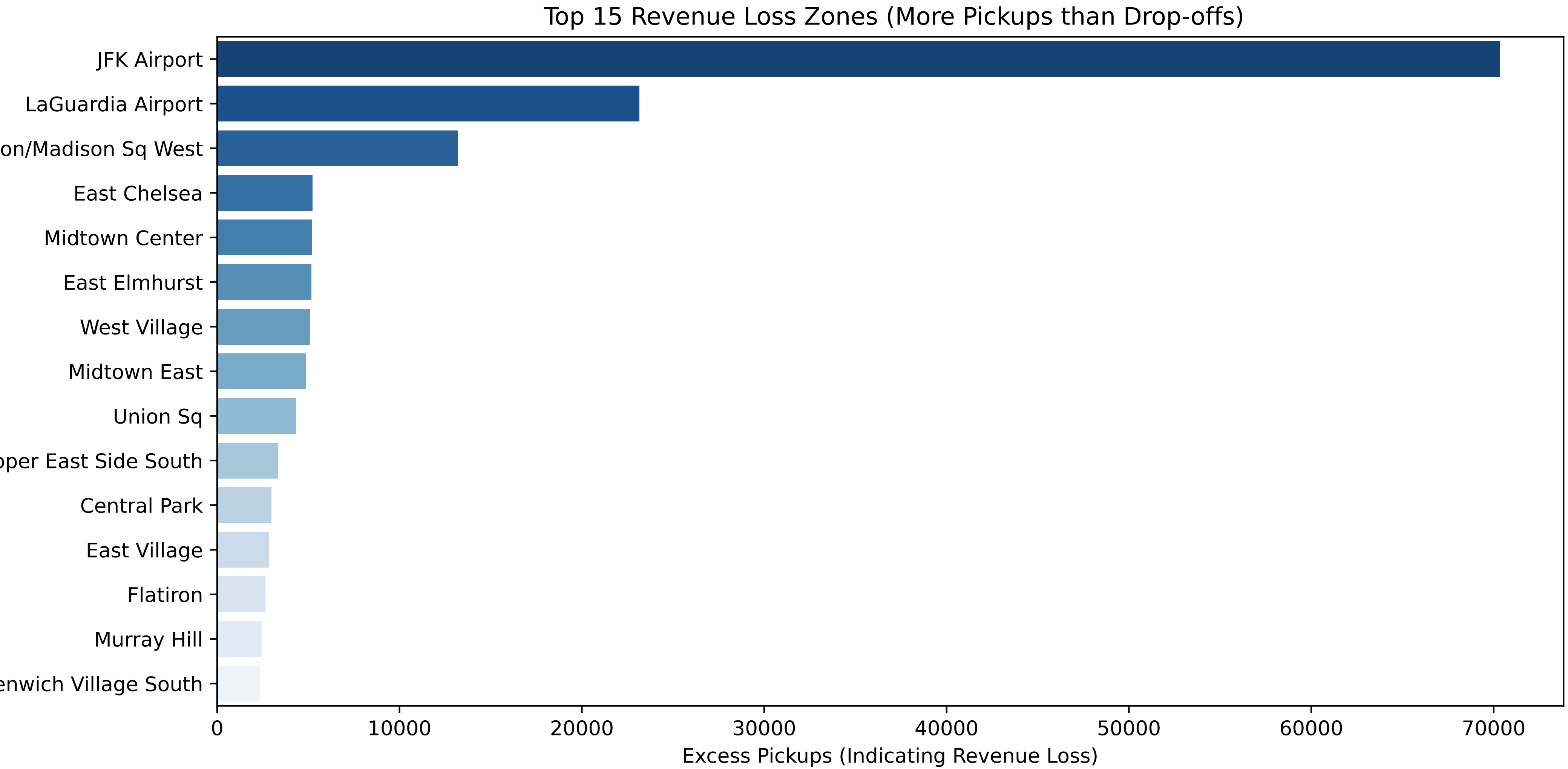
- Long-distance trips → Lower fare per mile.
- Short trips → Higher fare per mile, but more frequent.
- Tips depend on passenger count, trip distance, and pickup time.

Optimization Strategy:

Dynamic Pricing Based on Demand:

- Lower fares in high-competition areas.
- Increase fares for peak hours and low-supply areas.

Top 15 Zones with Empty Return Trips
Park Slope
ngton Heights South
Central Harlem
msburg (North Side)
entral Harlem North
Astoria
Greenpoint
Brooklyn Heights
East Harlem North
East Harlem South
ngton Heights North
Bridges/Seward Park
Lincoln Square West
msburg (South Side)
per West Side North
0
500
1000
1500
2000
2500
3000
3500
4000
Excess Drop-offs (Indicating Empty Return Trips)

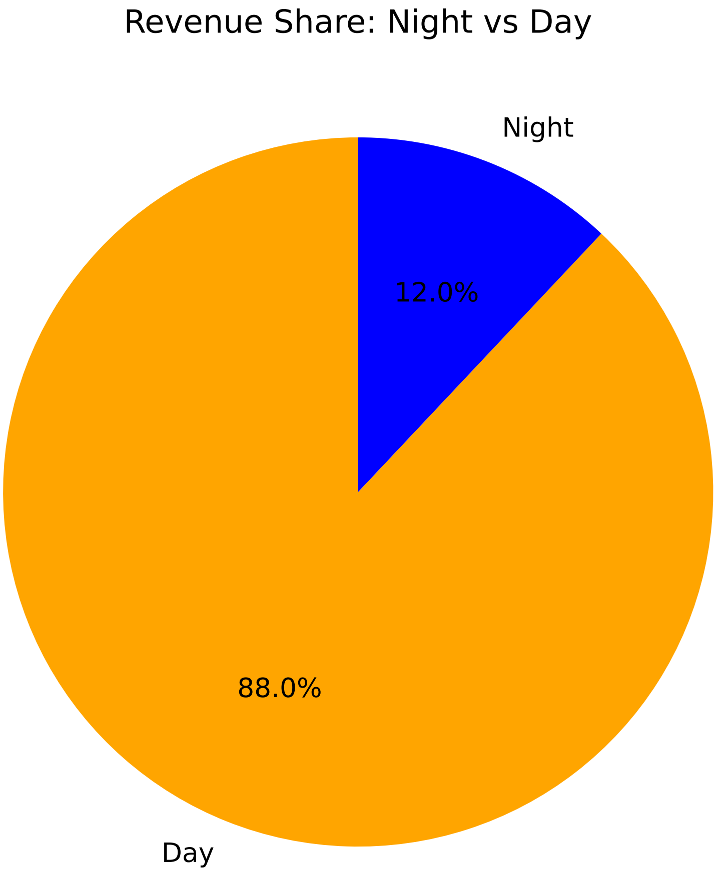

Top 15 Revenue Loss Zones (More Pickups than Drop-offs)
JFK Airport
LaGuardia Airport
on/Madison Sq West
East Chelsea
Midtown Center
East Elmhurst
West Village
Midtown East
Union Sq
pper East Side South
Central Park
East Village
Flatiron
Murray Hill
enwich Village South
0
10000
20000
30000
40000
50000
60000
70000
Excess Pickups (Indicating Revenue Loss)

Conclusion/Remarks:

4.1.1 Balance Supply and Demand to Reduce Empty Trips

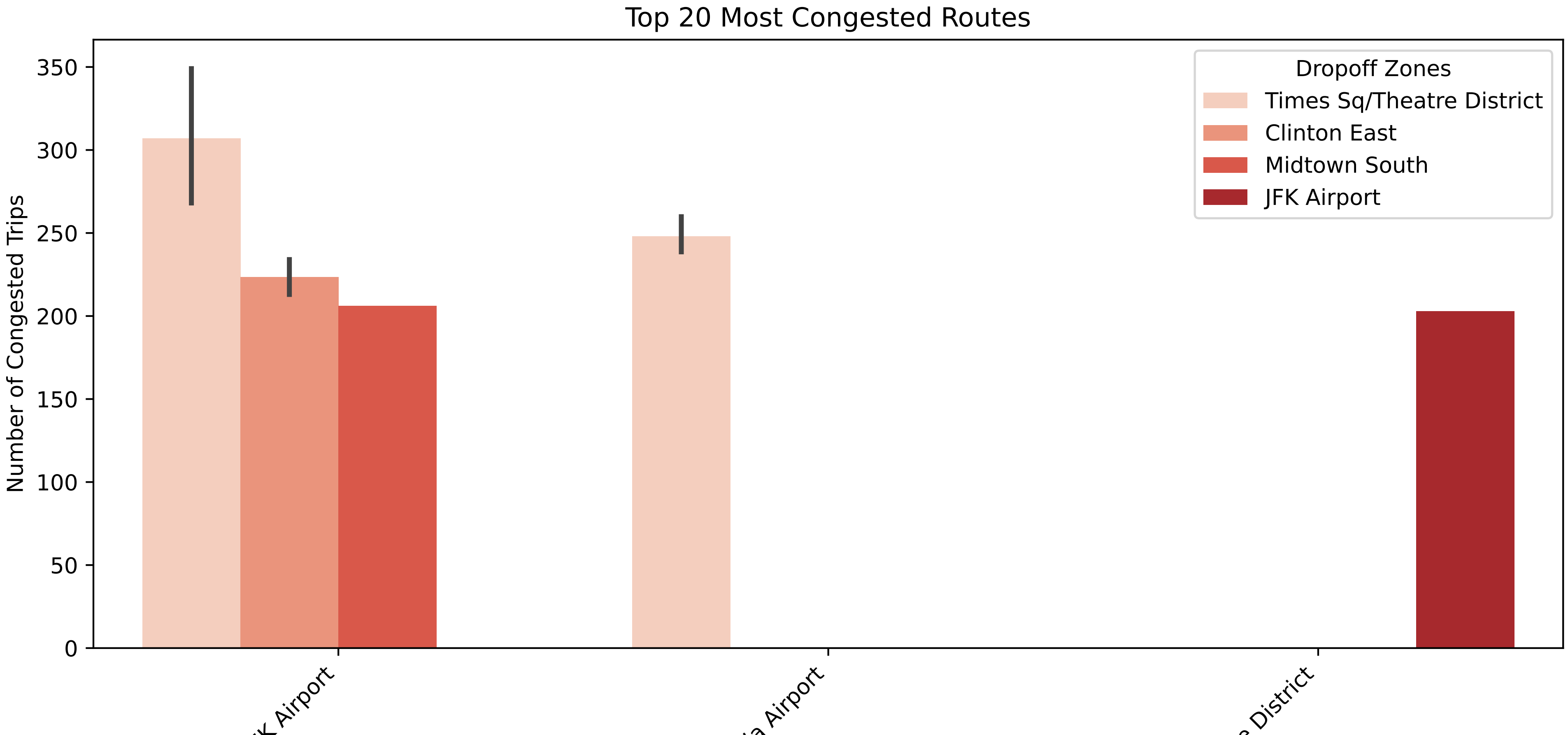
Insights from df_merged:
  - Identify times and areas where drop-offs happen, but pickups are low → leads to empty trips.
  - Zones with many outgoing trips but fewer incoming trips cause revenue loss.

Optimization Strategy:
  Backhaul Optimization:
    - Encourage drivers to pick up passengers from low-demand return zones.


Revenue Share: Night vs Day
Night
12.0%
88.0%
Day

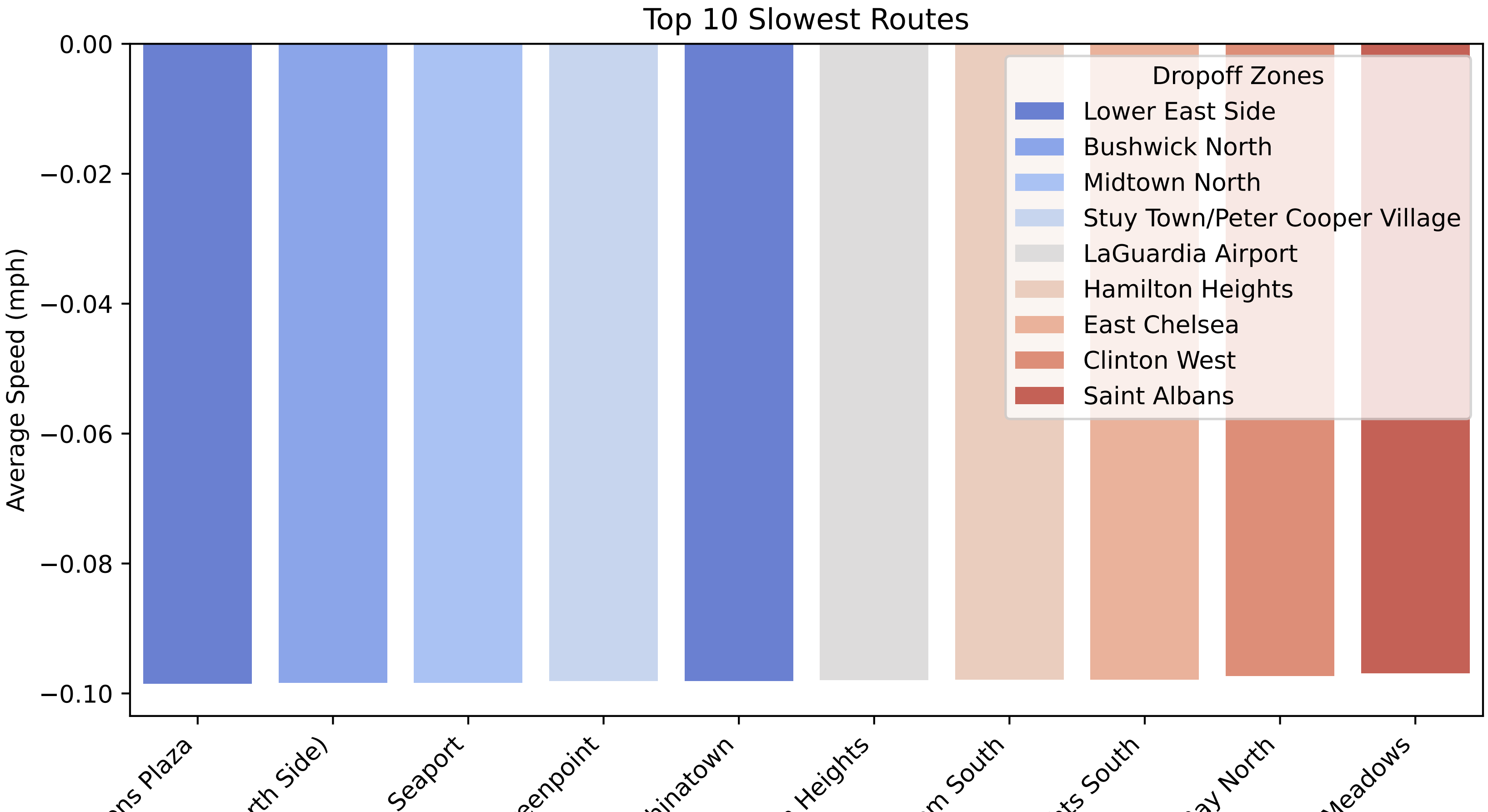
Conclusion/Remarks:

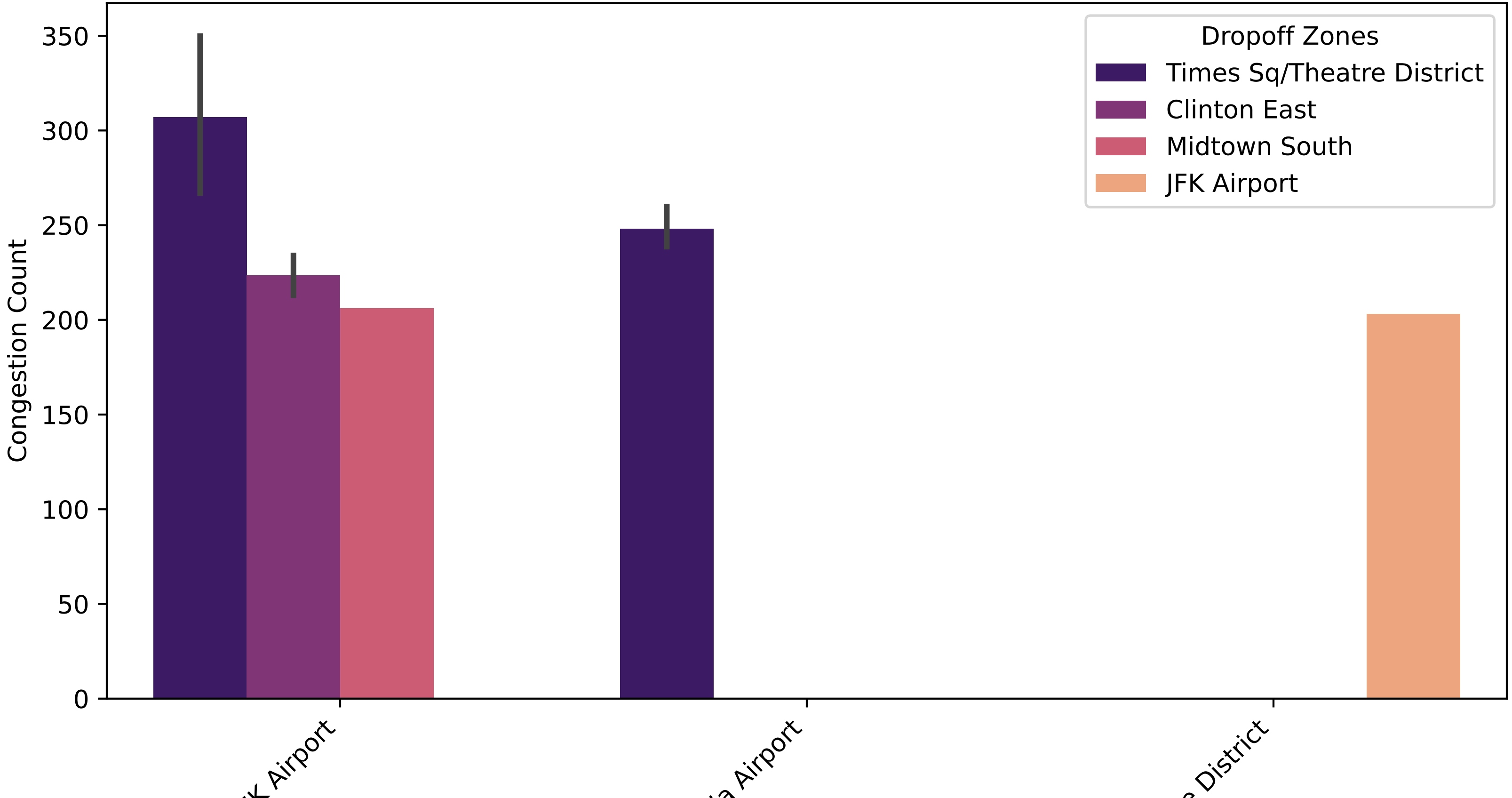
4.1.1 Compare nigh vs Day time Revenue share


Top 20 Most Congested Routes
Number of Congested Trips
0
50
100
150
200
250
300
350
Dropoff Zones
Times Sq/Theatre District
Clinton East
Midtown South
JFK Airport
K Airport
a Airport
e District

Conclusion/Remarks:

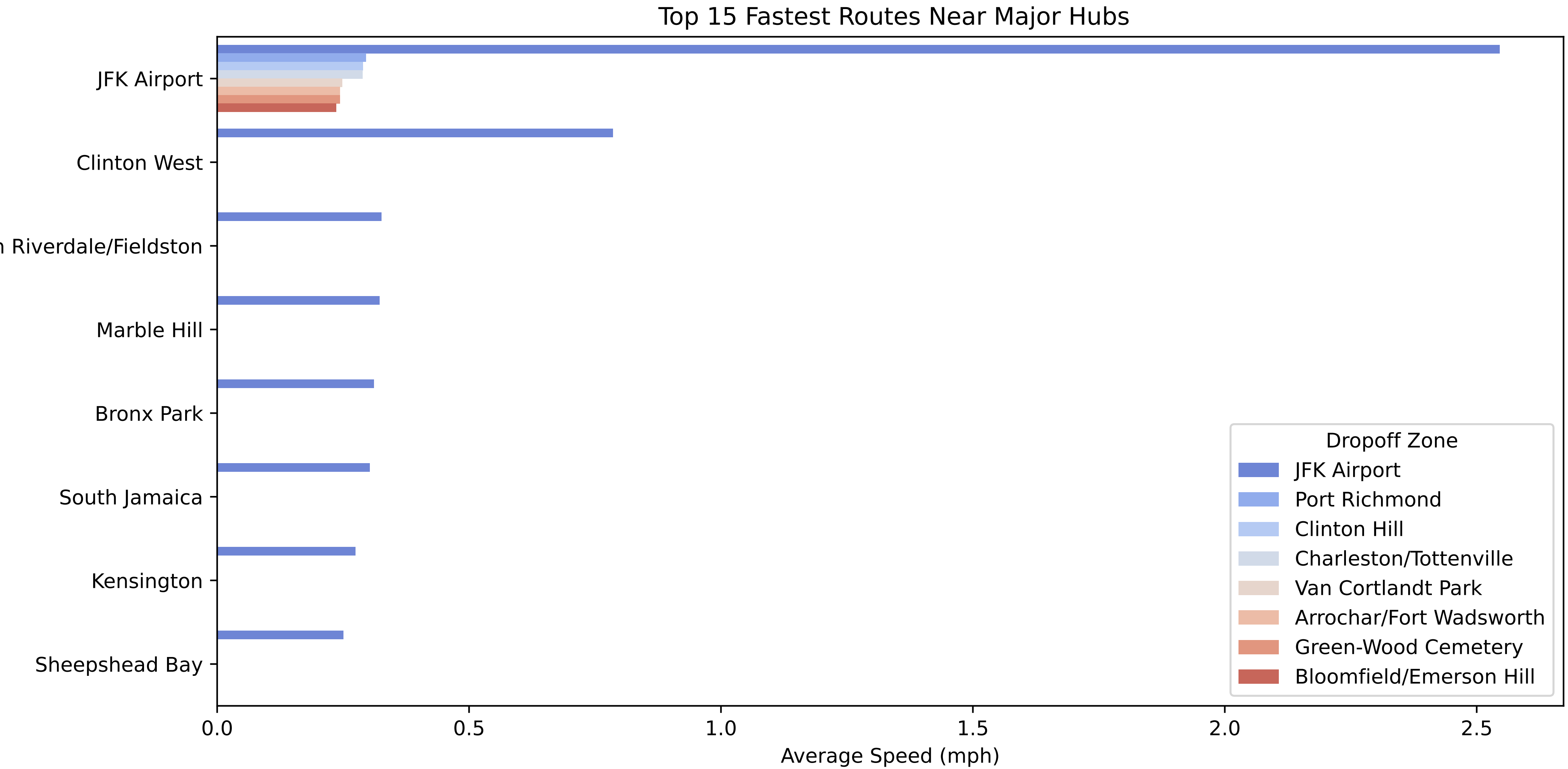
4.1.1 Compare Routes with High Congestion Pickup Zone and corresponding Dropoff Zones. JFK Airport is the most congested

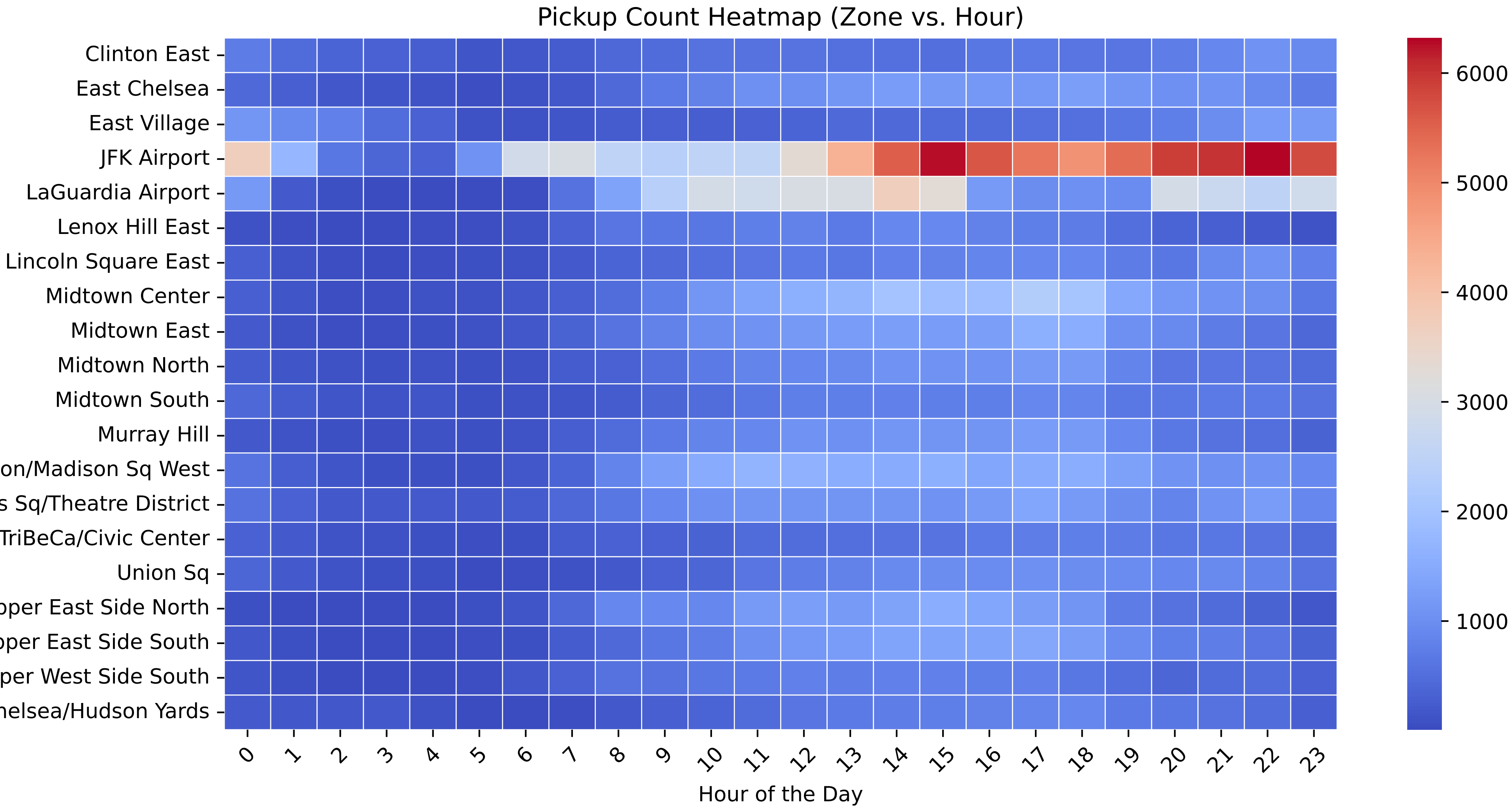

Top 10 Slowest Routes
Average Speed (mph)
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
Dropoff Zones
Lower East Side
Bushwick North
Midtown North
Stuy Town/Peter Cooper Village
LaGuardia Airport
Hamilton Heights
East Chelsea
Clinton West
Saint Albans
ns Plaza
rth Side)
Seaport
eenpoint
hinatown
Heights
m South
ts South
ay North
Meadows

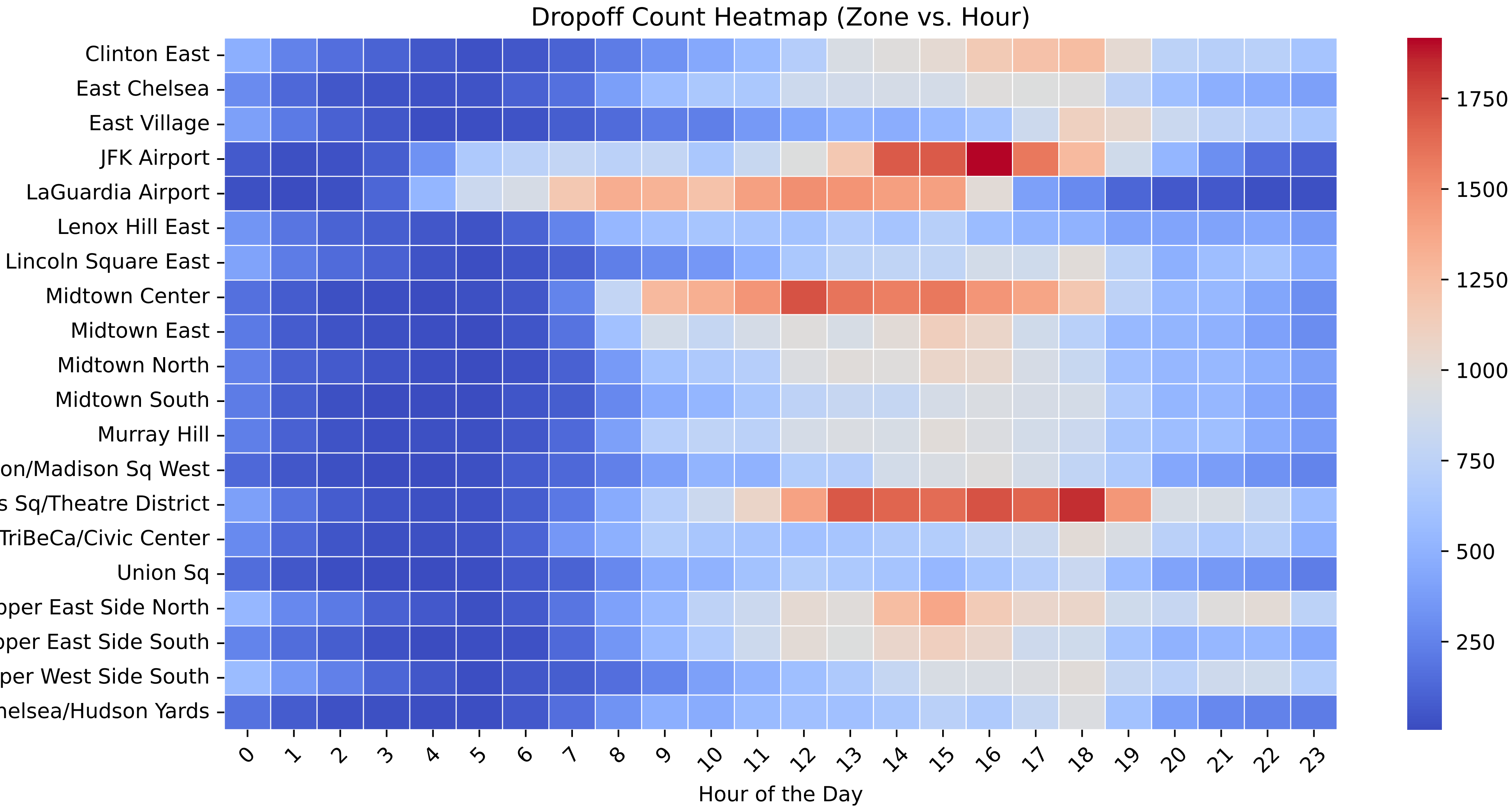

Top 10 Most Congested Routes

Conclusion/Remarks:

4.1.1 Compare the Top 10 slowest route with corresponding Dropoff


Top 15 Fastest Routes Near Major Hubs
JFK Airport
Clinton West
Riverdale/Fieldston
Marble Hill
Bronx Park
South Jamaica
Kensington
Sheepshead Bay
0.0
0.5
1.0
1.5
2.0
2.5
Average Speed (mph)
Dropoff Zone
JFK Airport
Port Richmond
Clinton Hill
Charleston/Tottenville
Van Cortlandt Park
Arrochar/Fort Wadsworth
Green-Wood Cemetery
Bloomfield/Emerson Hill

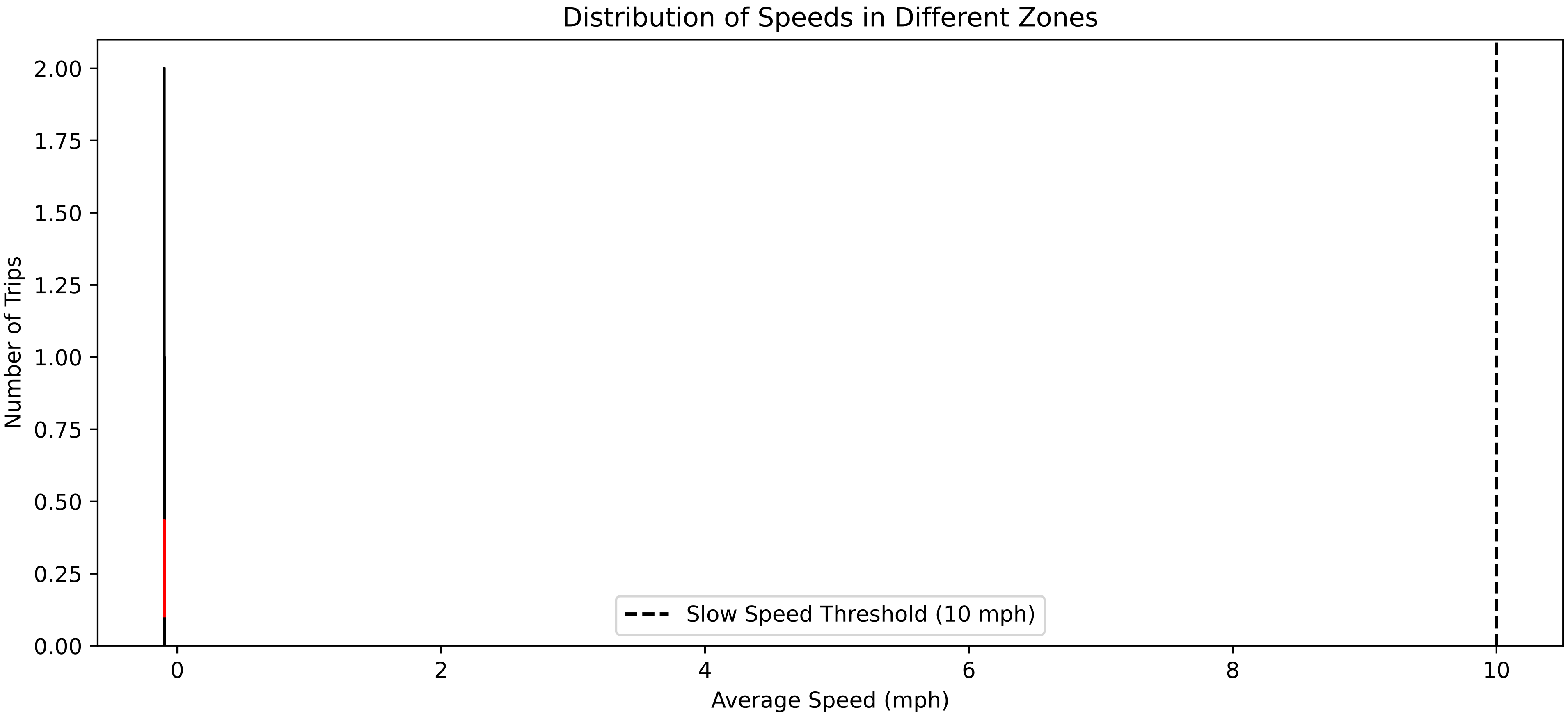

Pickup Count Heatmap (Zone vs. Hour)
Clinton East
East Chelsea
East Village
JFK Airport
LaGuardia Airport
Lenox Hill East
Lincoln Square East
Midtown Center
Midtown East
Midtown North
Midtown South
Murray Hill
on/Madison Sq West
s Sq/Theatre District
TriBeCa/Civic Center
Union Sq
pper East Side North
pper East Side South
per West Side South
helsea/Hudson Yards
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Hour of the Day
6000
5000
4000
3000
2000
1000



Dropoff Count Heatmap (Zone vs. Hour)

Conclusion/Remarks:

4.1.1 Compare Pickup vs Dropoff vs Hour in Heatmap. JFK airport Pickup 15:00 and 22:00 and Dropoff at 16:00


Distribution of Speeds in Different Zones
Number of Trips
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10
Average Speed (mph)
Slow Speed Threshold (10 mph)

Conclusion/Remarks:

4.1.1 Showing Slow Route Speeds

Conclusion/Remarks:

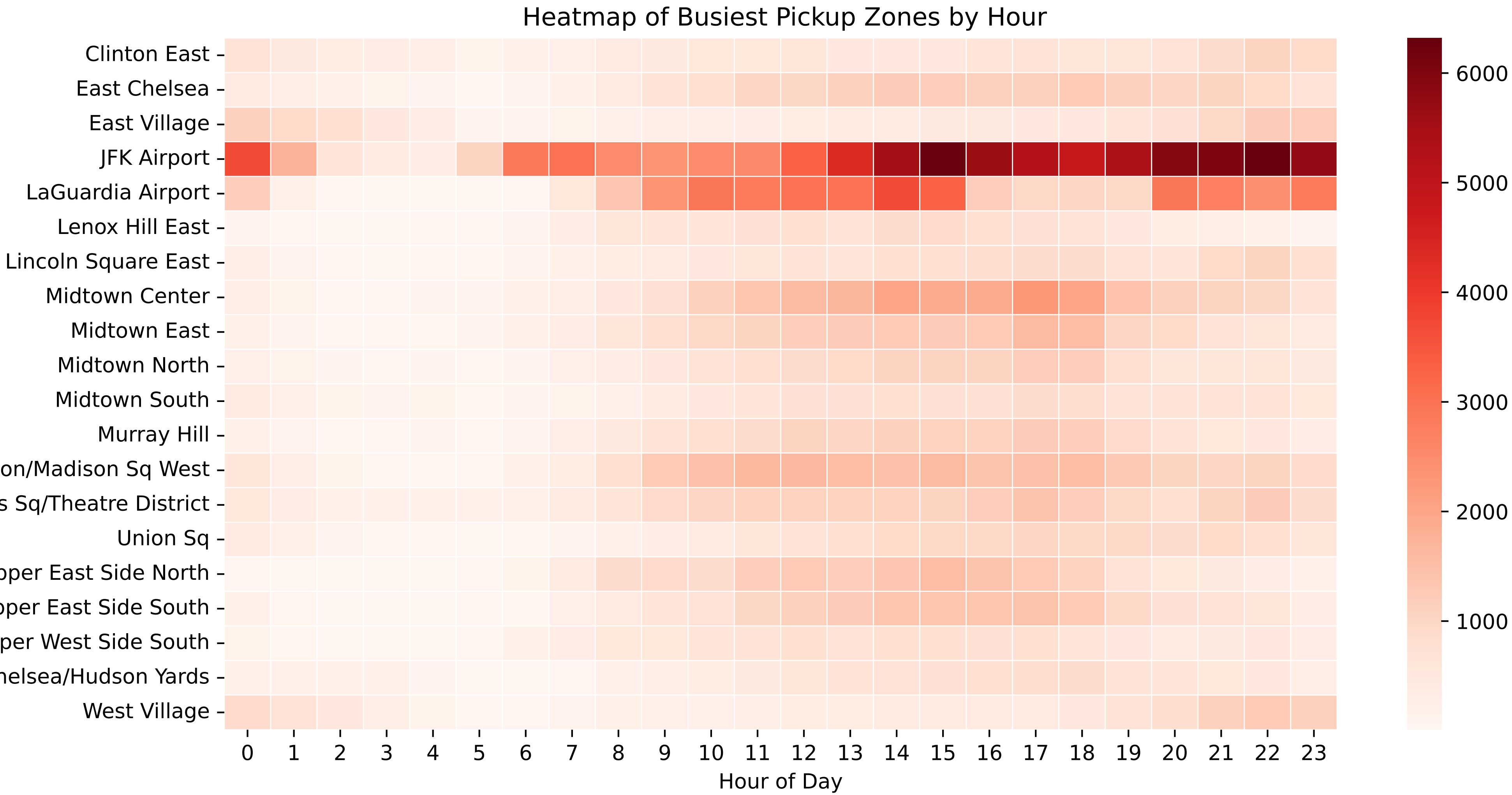
4.1.2 Strategic Cab Positioning to Maximize Revenue & Reduce Empty Trips

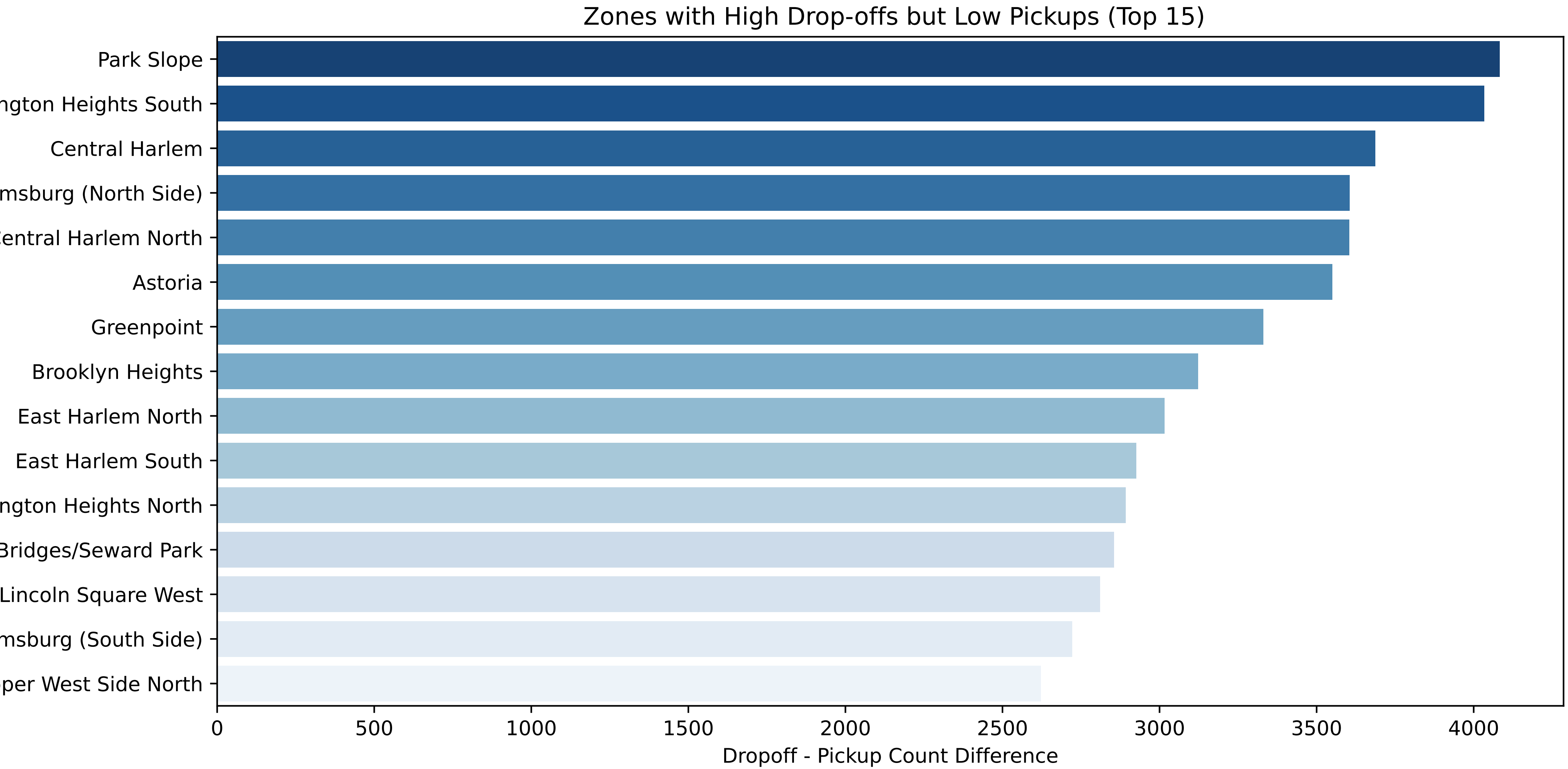
By analyzing trip data across time, days, and months, we can strategically position cabs:

1 Busiest Zones & Peak Demand:
   - Deploy more cabs where demand is high.

2 Imbalanced Zones (High Drop-offs, Low Pickups):
   - Prevent empty trips by repositioning cabs.

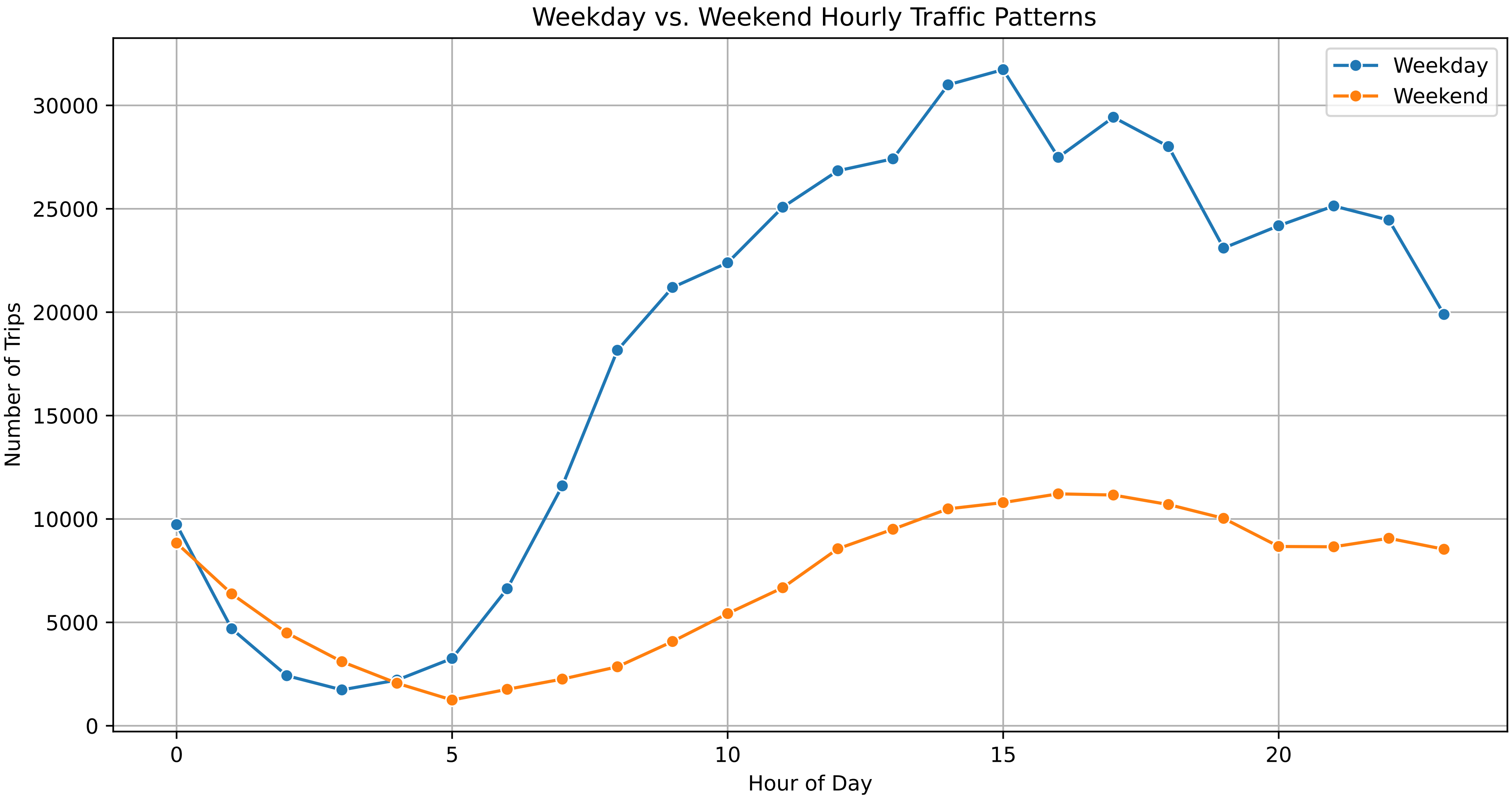
3 Time-Based Demand Shifts:


Heatmap of Busiest Pickup Zones by Hour
Clinton East
East Chelsea
East Village
JFK Airport
LaGuardia Airport
Lenox Hill East
Lincoln Square East
Midtown Center
Midtown East
Midtown North
Midtown South
Murray Hill
on/Madison Sq West
s Sq/Theatre District
Union Sq
pper East Side North
pper East Side South
per West Side South
helsea/Hudson Yards
West Village
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day
6000
5000
4000
3000
2000
1000

Zones with High Drop-offs but Low Pickups (Top 15)
Park Slope
ngton Heights South
Central Harlem
msburg (North Side)
entral Harlem North
Astoria
Greenpoint
Brooklyn Heights
East Harlem North
East Harlem South
ngton Heights North
Bridges/Seward Park
Lincoln Square West
msburg (South Side)
per West Side North
0
500
1000
1500
2000
2500
3000
3500
4000
Dropoff - Pickup Count Difference

Conclusion/Remarks:

4.1.2 Showing Zones where Low Pickup but High Dropoff. From this we can plan strategically


Weekday vs. Weekend Hourly Traffic Patterns
Weekday
Weekend
Number of Trips
0
5000
10000
15000
20000
25000
30000
Hour of Day
0
5
10
15
20

Conclusion/Remarks:

4.1.2 Showing Weekend vs. Weekday Demand Distribution

Conclusion/Remarks:

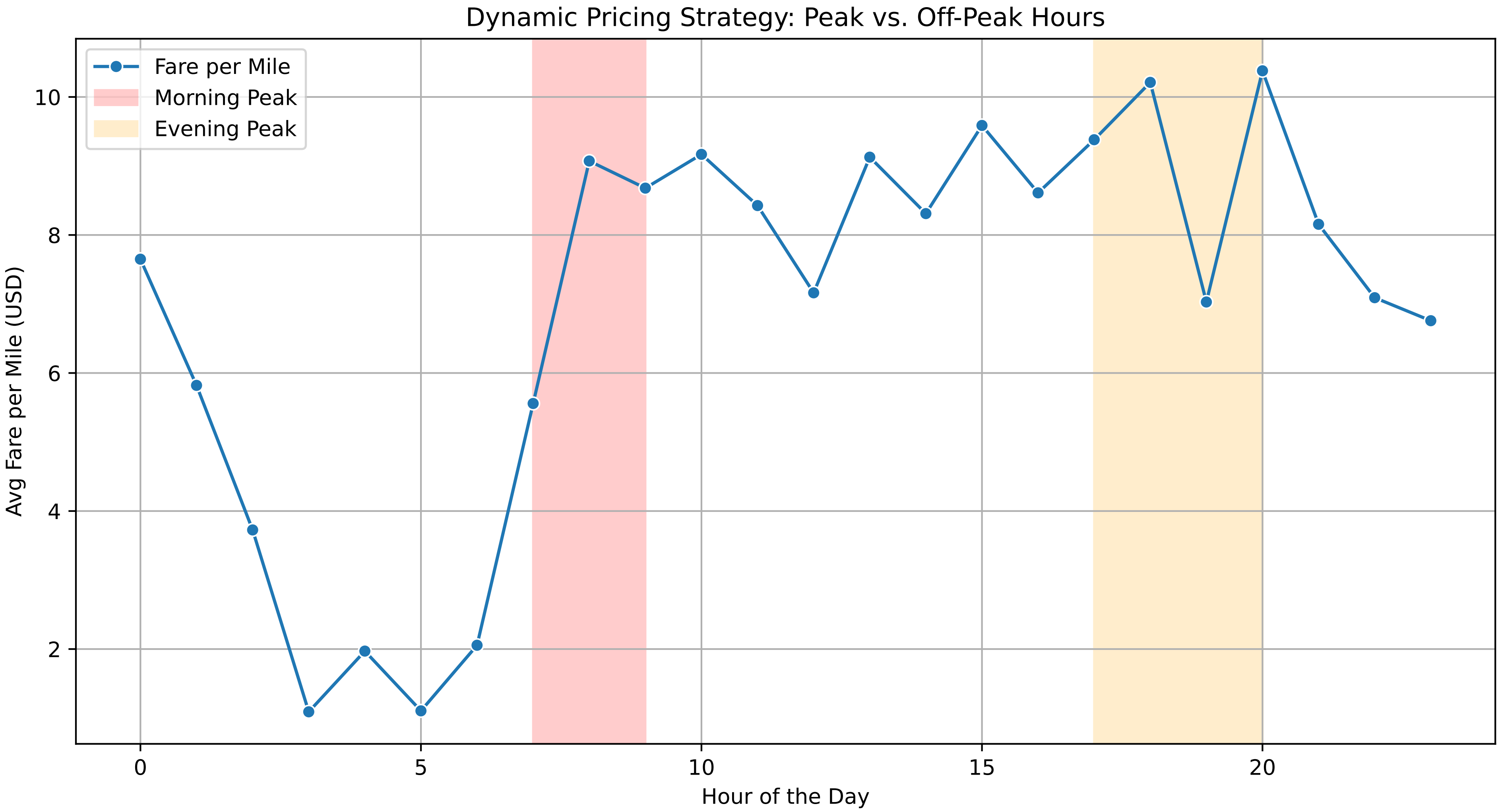
4.1.3 To Optimize Pricing While Staying Competitive

We analyze fare trends, demand patterns, and vendor competition:

1 Demand-Based Surge Pricing:
   - Adjust fares based on peak vs. off-peak hours.

2 Distance-Based Tiered Pricing:
   - Optimize fare per mile based on trip length.

3 Competitive Vendor Pricing:


Dynamic Pricing Strategy: Peak vs. Off-Peak Hours
Fare per Mile
Morning Peak
Evening Peak
Avg Fare per Mile (USD)
2
4
6
8
10
0
5
10
15
20
Hour of the Day

Conclusion/Remarks:

4.1.3 Showing plot to Identify peak demand hours where higher fares can be applied

Conclusion/Remarks:

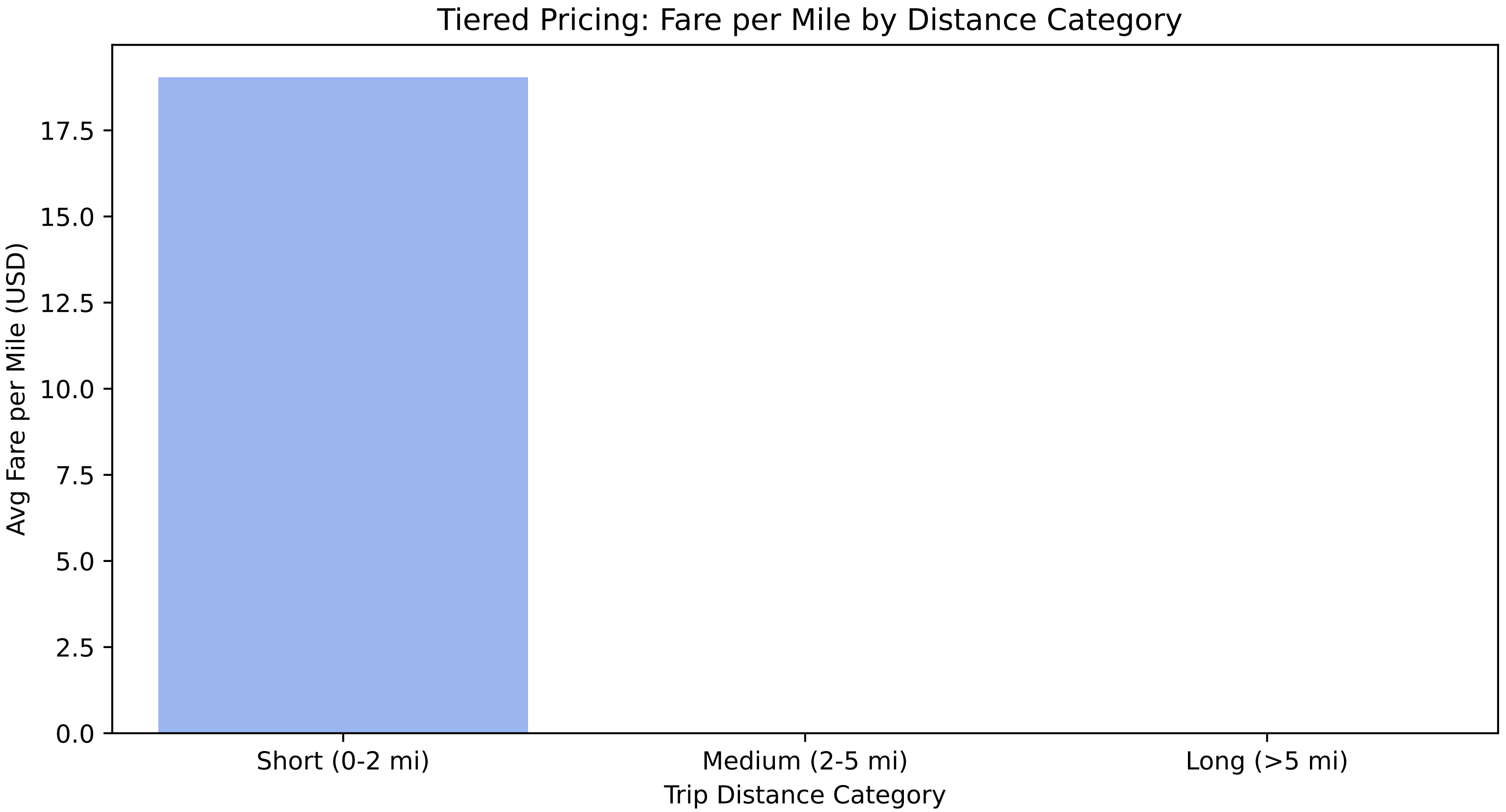
4.1.3 Distance-Based Tiered Pricing Strategy

Adjust fares based on trip length to optimize revenue:

Short trips (<2 miles):
  - Slightly higher fare per mile to compensate for short distances.

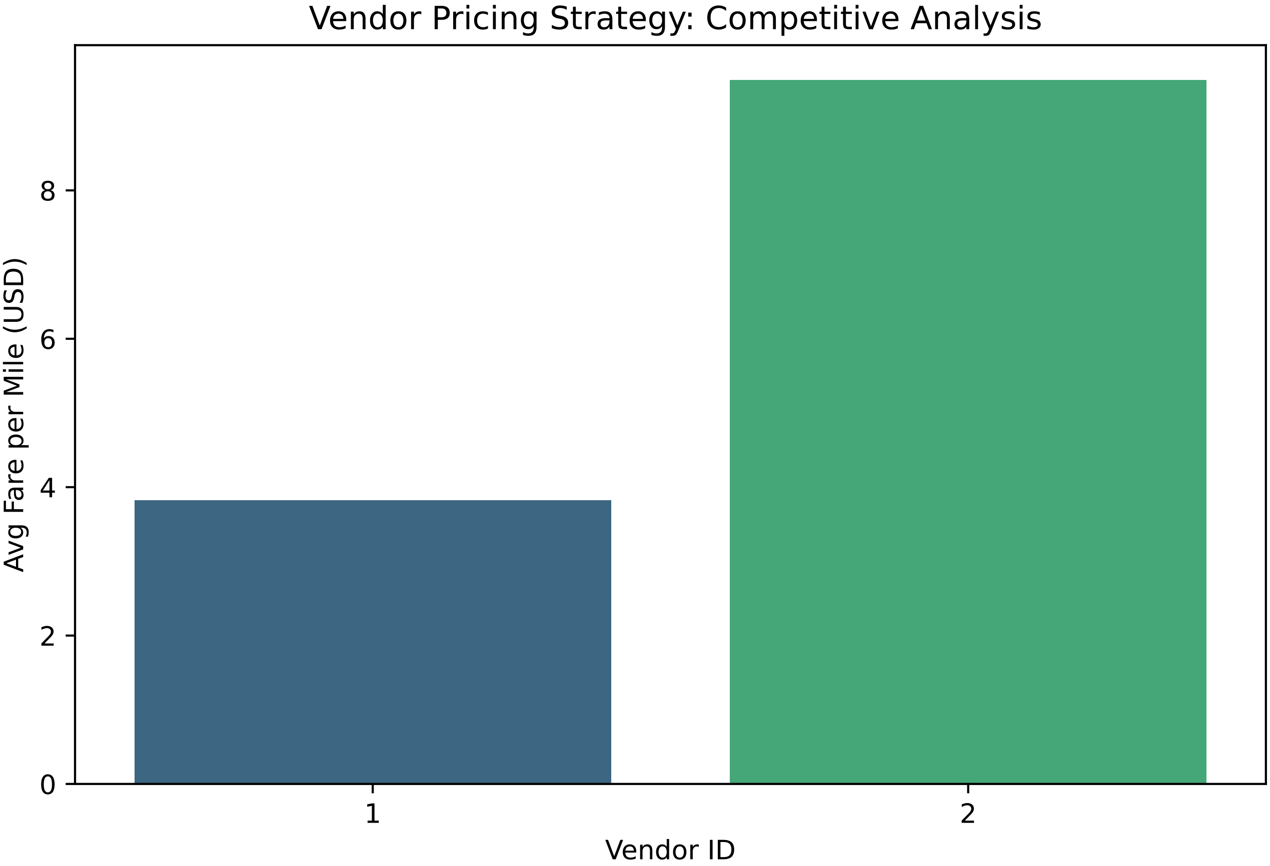
Mid-range trips (2-5 miles):
  - Standard fare per mile for balanced pricing.

Long trips (>5 miles):


Tiered Pricing: Fare per Mile by Distance Category
Avg Fare per Mile (USD)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Short (0-2 mi)
Medium (2-5 mi)
Long (>5 mi)
Trip Distance Category

Conclusion/Remarks:

4.1.3 Showing Tiered Pricing: Fare per Mile by Distance Category


Vendor Pricing Strategy: Competitive Analysis
Avg Fare per Mile (USD)
0
2
4
6
8
1
2
Vendor ID

Conclusion/Remarks:

4.1.3 Showing the comparison fare per mile for different vendors and adjust pricing to stay competitive

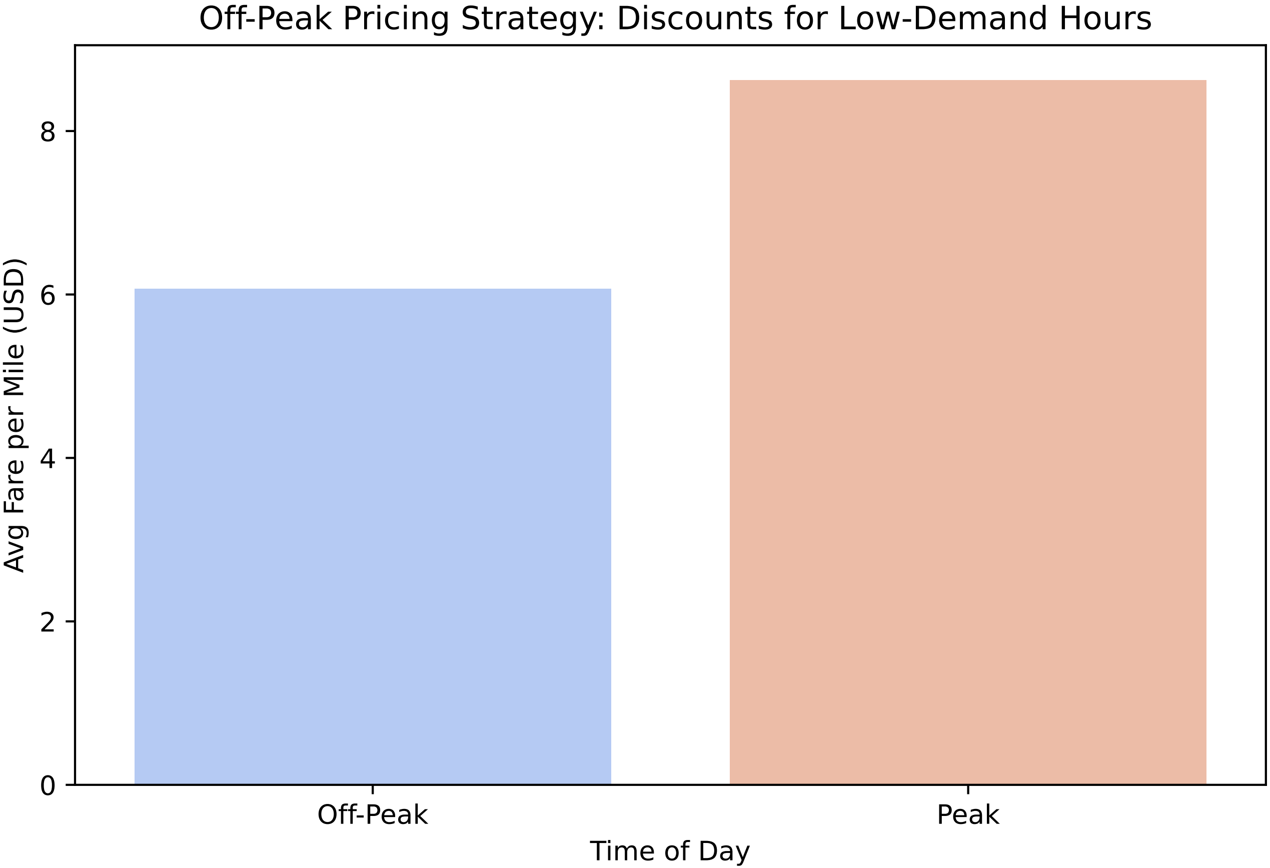
Conclusion/Remarks:

Time-Based Discounts for Off-Peak Hours

Offer discounts between 10 PM - 5 AM to boost nighttime rides.

Justification:

- Encourages more passengers to travel during low-demand hours.
- Helps drivers fill empty rides, reducing idle time.
- Improves overall revenue by maintaining a steady stream of trips.


Off-Peak Pricing Strategy: Discounts for Low-Demand Hours
Avg Fare per Mile (USD)
0
2
4
6
8
Off-Peak
Peak
Time of Day

Conclusion/Remarks:

4.1.3 Showing the Off-Peak Pricing Strategy: Discounts for Low-Demand Hours