UBER SUPPLY-DEMAND GAP ANALYSIS

By Anish Chakravorty

INTRODUCTION

In this data-driven case study, we perform an in-depth analysis of Uber's ride request data to uncover the root causes behind service inefficiencies, particularly focusing on the gap between customer demand and available supply. With ride-hailing being a fast-paced, real-time logistics challenge, understanding patterns in trip cancellations and unfulfilled requests is critical for optimizing fleet utilization and enhancing customer experience. The dataset, consisting of over 6,700 records, includes detailed ride-level data such as timestamps, driver IDs, trip statuses, pickup locations, and derived time features.

The project combines analysis through Python, SQL, and Excel Pivot Tables to generate actionable insights. Using exploratory data analysis (EDA), we uncover temporal patterns of ride failures, identify peak demand windows, and assess how effectively Uber's driver supply matches user expectations. This blended technical approach enables a comprehensive investigation into Uber's operational dynamics, ultimately supporting strategic recommendations to close the supply-demand gap.

COMPANY OVERVIEW

Uber Technologies Inc. is a global ride-hailing platform that connects passengers with nearby drivers through its mobile app. Since its founding in 2009, Uber has revolutionized urban transportation by offering on-demand mobility solutions. Operating in over 10,000 cities worldwide, the company's success is heavily reliant on its ability to fulfill user requests efficiently, especially during high-demand periods like rush hours or airport transfers.

As a platform business, Uber faces the continual challenge of balancing supply (drivers) and demand (riders) in real-time. Fluctuations in demand, driver availability, and traffic conditions can result in unmet ride requests, either due to cancellations or "No Cars Available" scenarios. This analysis seeks to support Uber's operations and strategy teams by providing data-backed insights that can improve rider satisfaction, reduce trip failures, and increase driver utilization.

BUSINESS TASK

As a data analyst, my goal in this project is to examine Uber's ride request data to:

1. Understand the root causes of ride failures — specifically cancellations and lack of available cars.
2. Identify when and where demand exceeds supply, using time-based and location-based segmentation.
3. Quantify the supply-demand gap, and visualize patterns across various request time slots.
4. Assess driver activity and utilization, especially during peak demand periods.
5. Provide business recommendations to reduce trip failures, improve driver allocation, and enhance the rider experience.

The outcome of this project is intended to support Uber's data operations and driver management strategies by highlighting critical problem areas and offering practical solutions based on real-world ride behavior.

DATA OVERVIEW

- Original Dataset
  - Source: CSV file (uber_cleaned_data.csv)
  - Total Rows: 6,745
  - Original Columns:
  - Request id, Pickup point, Driver id, Status, Request timestamp, Drop timestamp

These 6 columns formed the raw dataset collected from Uber's ride request logs.

DATA TRANSFORMATION

Initial Data Issues Addressed:

- Inconsistent timestamp formats
- Mixed data types (floats for IDs, etc.)
- Blank/“null” string entries instead of true SQL/Excel NULLs

Final Dataset Result:

- Rows: 6,745
- Columns Increased To: 15
- Transformation Layers: Time-based features, binning, and cleaning for Pivot and SQL analysis

KEY TRANSFORMATIONS PERFORMED

Transformation	Description	Resulting Column(s)
🕒 **Timestamp to Date & Hour**	Extracted date and hour from Request timestamp & Drop timestamp	Request date, Request hour, Drop date, Drop hour
📆 **Weekday Mapping**	Mapped dates to day of week	Request day, Drop day
🕓 **Time Slot Binning**	Converted hour into time-of-day categories like Morning, Evening, etc.	Request hour timeslot, Drop hour timeslot
⏱️ **Trip Duration Calculation**	Calculated duration between request and drop times	Trip duration_min, Trip duration_hr
📊 **Trip Duration Binning**	Grouped durations into predefined categories	Trip duration slot
❌ **Null Handling**	Replaced "null"/' ' strings with Excel-valid blanks and Python NaN	Accurate NULLs retained for logical filtering

AD-HOC-ANALYSIS QUESTIONS

1. What is the total number of ride requests made during the observed period?
2. How are the ride requests distributed across different trip statuses (Completed, Cancelled, No Cars Available)?
3. Which pickup point (City or Airport) receives the highest number of ride requests?
4. How are the different trip statuses distributed across each pickup point?
5. What is the distribution of ride requests by each hour of the day?
6. Which hours of the day show the highest number of cancellations or 'No Cars Available' incidents?
7. How do ride requests vary across different days of the week?
8. How many trips were successfully completed compared to those that were unfulfilled (i.e., cancelled or no cars available)?
9. What is the average trip duration for trips that were completed?
10. Which 5 hours of the day have the highest number of completed trips?
11. What is the overall supply-demand gap, and how significant is the portion of demand that went unfulfilled?
12. Which drivers had the highest utilization rates based on the number of completed trips?
13. How does ride request performance vary across different time-of-day segments (request hour timeslots)?
14. Which pickup point has the highest cancellation rate, and how does it compare to the other?

1. WHAT IS THE TOTAL NUMBER OF RIDE REQUESTS MADE DURING THE OBSERVED PERIOD?

```
SELECT COUNT(*) AS total_requests
FROM  uber_cleaned_data_clean;
```

	total_requests
▶	6745

2. HOW ARE THE RIDE REQUESTS DISTRIBUTED ACROSS DIFFERENT TRIP STATUSES (COMPLETED, CANCELLED, NO CARS AVAILABLE)?

```
SELECT status, COUNT(*) AS count
FROM uber_cleaned_data_clean
GROUP BY status
ORDER BY count DESC;
```

status	count
Trip Completed	2831
No Cars Available	2650
Cancelled	1264

3. WHICH PICKUP POINT (CITY OR AIRPORT) RECEIVES THE HIGHEST NUMBER OF RIDE REQUESTS?

```
SELECT `pickup point`, COUNT(*) AS count
FROM uber_cleaned_data_clean
GROUP BY `pickup point`
ORDER BY count DESC;
```

pickup point	count
City	3507
Airport	3238

4. HOW ARE THE DIFFERENT TRIP STATUSES DISTRIBUTED ACROSS EACH PICKUP POINT?

```
SELECT `pickup point`, status, COUNT(*) AS count
FROM uber_cleaned_data_clean
GROUP BY `pickup point`, status
ORDER BY `pickup point`, count DESC;
```

pickup point	status	count
Airport	No Cars Available	1713
Airport	Trip Completed	1327
Airport	Cancelled	198
City	Trip Completed	1504
City	Cancelled	1066
City	No Cars Available	937

5.WHAT IS THE DISTRIBUTION OF RIDE REQUESTS BY EACH HOUR OF THE DAY?

```
SELECT `request hour`, COUNT(*) AS total_requests
FROM uber_cleaned_data_clean
GROUP BY `request hour`
ORDER BY `request hour`;
```

request hour	total_requests
0	99
1	85
2	99
3	92
4	203
5	445
6	398
7	406
8	423
9	431
10	243
11	171
12	184
13	160
14	136
15	171
16	159
17	418
18	510
19	473
20	492
21	449
22	304
23	194

6. WHICH HOURS OF THE DAY SHOW THE HIGHEST NUMBER OF CANCELLATIONS OR 'NO CARS AVAILABLE' INCIDENTS?

```sql
SELECT `request hour`, status, COUNT(*) AS count
FROM uber_cleaned_data_clean
WHERE status != 'Trip Completed'
GROUP BY `request hour`, status
ORDER BY `request hour`;
```

request hour	status	count
0	Cancelled	3
0	No Cars Available	56
1	Cancelled	4
1	No Cars Available	56
2	Cancelled	5
2	No Cars Available	57
3	Cancelled	2
3	No Cars Available	56
4	Cancelled	51
4	No Cars Available	74
5	Cancelled	176
5	No Cars Available	84
6	Cancelled	145
6	No Cars Available	86
7	Cancelled	169
7	No Cars Available	63
8	Cancelled	178
8	No Cars Available	90
9	Cancelled	175
9	No Cars Available	83
10	Cancelled	62
10	No Cars Available	65
11	Cancelled	15
11	No Cars Available	41
12	Cancelled	19
13	Cancelled	18
13	No Cars Available	53
14	Cancelled	11
14	No Cars Available	37
15	Cancelled	21
15	No Cars Available	48
16	Cancelled	22
16	No Cars Available	46
17	Cancelled	35
17	No Cars Available	232
18	Cancelled	24
18	No Cars Available	322
19	Cancelled	24
19	No Cars Available	283
20	Cancelled	41
20	No Cars Available	290
21	Cancelled	42
21	No Cars Available	265
22	Cancelled	12
22	No Cars Available	138
23	Cancelled	10
23	No Cars Available	81

7. HOW DO RIDE REQUESTS VARY ACROSS DIFFERENT DAYS OF THE WEEK?

```
SELECT `request day`, status, COUNT(*) AS count
FROM uber_cleaned_data_clean
GROUP BY `request day`, status
ORDER BY FIELD(`request day`, 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday');
```

request day	status	count
Monday	Trip Completed	601
Monday	Cancelled	262
Monday	No Cars Available	504
Tuesday	Trip Completed	562
Tuesday	Cancelled	240
Tuesday	No Cars Available	505
Wednesday	Trip Completed	577
Wednesday	Cancelled	270
Wednesday	No Cars Available	490
Thursday	Trip Completed	530
Thursday	Cancelled	252
Thursday	No Cars Available	571
Friday	Trip Completed	561
Friday	Cancelled	240
Friday	No Cars Available	580

8. HOW MANY TRIPS WERE SUCCESSFULLY COMPLETED COMPARED TO THOSE THAT WERE UNFULFILLED (I.E., CANCELLED OR NO CARS AVAILABLE)?

```sql
SELECT
  CASE WHEN `Drop date` IS NOT NULL THEN 'Completed' ELSE 'Not Completed' END AS trip_status,
  COUNT(*) AS total
FROM uber_cleaned_data_clean
GROUP BY trip_status;
```

trip_status	total
Completed	2831
Not Completed	3914

9. WHAT IS THE AVERAGE TRIP DURATION FOR TRIPS THAT WERE COMPLETED?

```sql
SELECT
  ROUND(AVG(`Trip duration_min`), 2) AS avg_duration_min,
  ROUND(AVG(`Trip duration_hr`), 2) AS avg_duration_hr
FROM uber_cleaned_data_clean
WHERE status = 'Trip Completed';
```

avg_duration_min	avg_duration_hr
52.41	0.87

10. WHICH 5 HOURS OF THE DAY HAVE THE HIGHEST NUMBER OF COMPLETED TRIPS?

```
SELECT `request hour`, COUNT(*) AS no_car_requests
FROM uber_cleaned_data_clean
WHERE status = 'Trip Completed'
GROUP BY `request hour`
ORDER BY no_car_requests DESC
LIMIT 5;
```

request hour	no_car_requests
5	185
7	174
9	173
6	167
19	166

11.WHAT IS THE OVERALL SUPPLY-DEMAND GAP, AND HOW SIGNIFICANT IS THE PORTION OF DEMAND THAT WENT UNFULFILLED?

```
SELECT
  COUNT(*) AS total_requests,
  SUM(CASE WHEN status = 'Trip Completed' THEN 1 ELSE 0 END) AS trips_completed,
  SUM(CASE WHEN status != 'Trip Completed' THEN 1 ELSE 0 END) AS supply_gap,
  ROUND(SUM(CASE WHEN status != 'Trip Completed' THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS gap_percentage
FROM uber_cleaned_data_clean;
```

total_requests	trips_completed	supply_gap	gap_percentage
6745	2831	3914	58.03

12.WHICH DRIVERS HAD THE HIGHEST UTILIZATION RATES BASED ON THE NUMBER OF COMPLETED TRIPS?

```
SELECT
  `Driver id`,
  COUNT(*) AS total_assigned,
  SUM(CASE WHEN status = 'Trip Completed' THEN 1 ELSE 0 END) AS completed_trips,
  ROUND(SUM(CASE WHEN status = 'Trip Completed' THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS utilization_rate
FROM uber_cleaned_data_clean
WHERE `Driver id` IS NOT NULL
GROUP BY `Driver id`
ORDER BY utilization_rate DESC
LIMIT 20;
```

Driver id	total_assigned	completed_trips	utilization_rate
75	11	11	100.00
42	7	7	100.00
55	10	10	100.00
11	8	8	100.00
12	12	12	100.00
23	15	14	93.33
188	14	13	92.86
273	13	12	92.31
46	12	11	91.67
18	12	11	91.67
162	11	10	90.91
26	10	9	90.00
156	10	9	90.00
252	9	8	88.89
91	9	8	88.89
184	17	15	88.24
41	8	7	87.50
118	16	14	87.50
208	8	7	87.50
48	15	13	86.67

13. HOW DOES RIDE REQUEST PERFORMANCE VARY ACROSS DIFFERENT TIME-OF-DAY SEGMENTS (REQUEST HOUR TIMESLOTS)?

```
SELECT
  `Request hour timeslot`,
  COUNT(*) AS total_requests,
  SUM(CASE WHEN status = 'Trip Completed' THEN 1 ELSE 0 END) AS completed_trips,
  SUM(CASE WHEN status = 'Cancelled' THEN 1 ELSE 0 END) AS cancelled,
  SUM(CASE WHEN status = 'No Cars Available' THEN 1 ELSE 0 END) AS no_cars,
  ROUND(SUM(CASE WHEN status != 'Trip Completed' THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS failure_rate
FROM uber_cleaned_data_clean
GROUP BY `Request hour timeslot`
ORDER BY total_requests DESC;
```

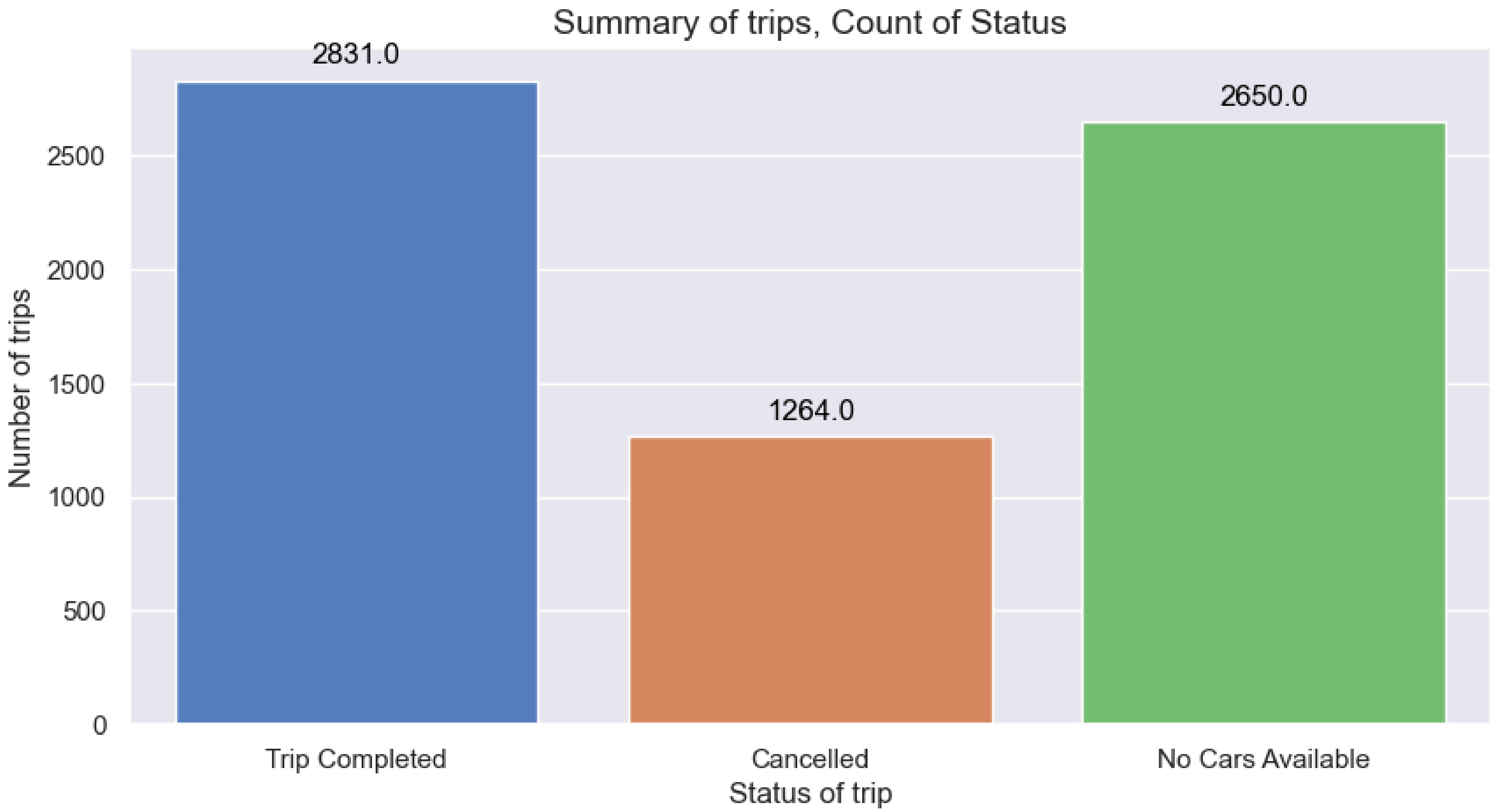
Request hour timeslot	total_requests	completed_trips	cancelled	no_cars	failure_rate
Evening	1893	642	124	1127	66.09
Early morning	1672	681	668	323	59.27
Late morning	1029	525	271	233	48.98
Night	947	399	64	484	57.87
Afternoon	626	370	72	184	40.89
Late night	394	149	58	187	62.18

14. WHICH PICKUP POINT HAS THE HIGHEST CANCELLATION RATE, AND HOW DOES IT COMPARE TO THE OTHER?

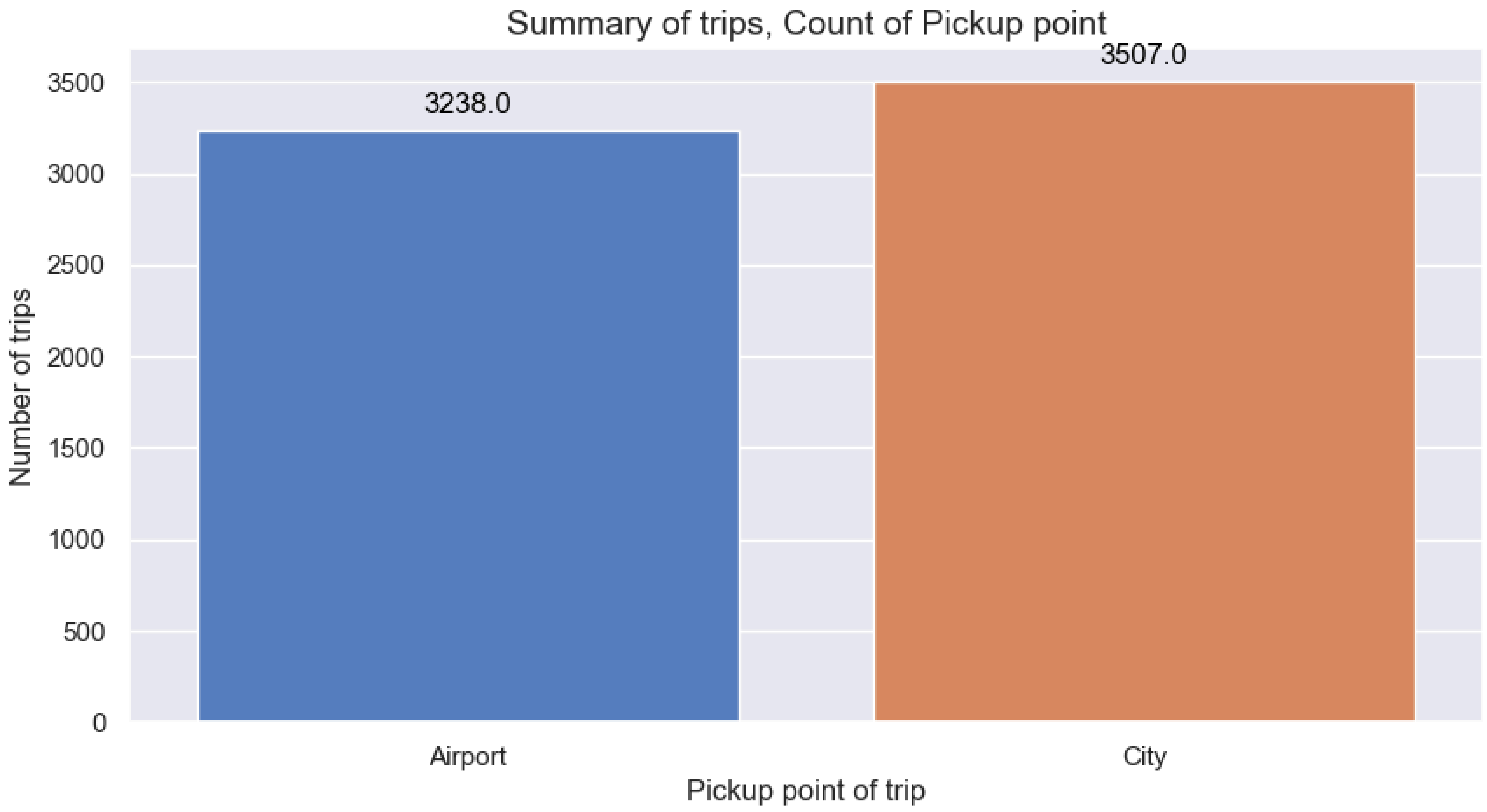
```sql
SELECT
  `Pickup point`,
  COUNT(*) AS total_requests,
  SUM(CASE WHEN status = 'Cancelled' THEN 1 ELSE 0 END) AS cancellations,
  ROUND(SUM(CASE WHEN status = 'Cancelled' THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS cancellation_rate
FROM uber_cleaned_data_clean
GROUP BY `Pickup point`
ORDER BY cancellation_rate DESC;
```

Pickup point	total_requests	cancellations	cancellation_rate
City	3507	1066	30.40
Airport	3238	198	6.11

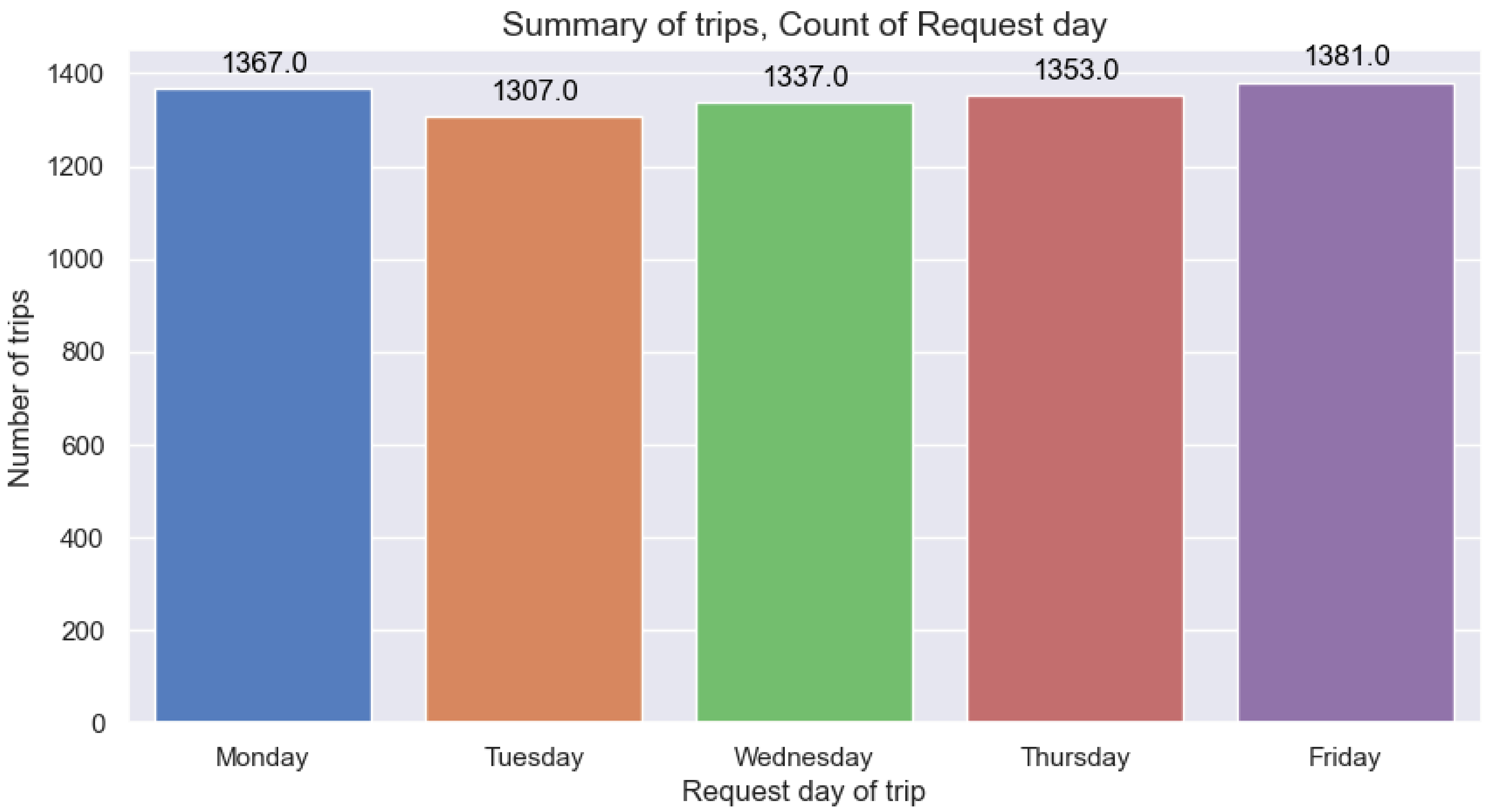
COUNTPLOT FOR STATUS



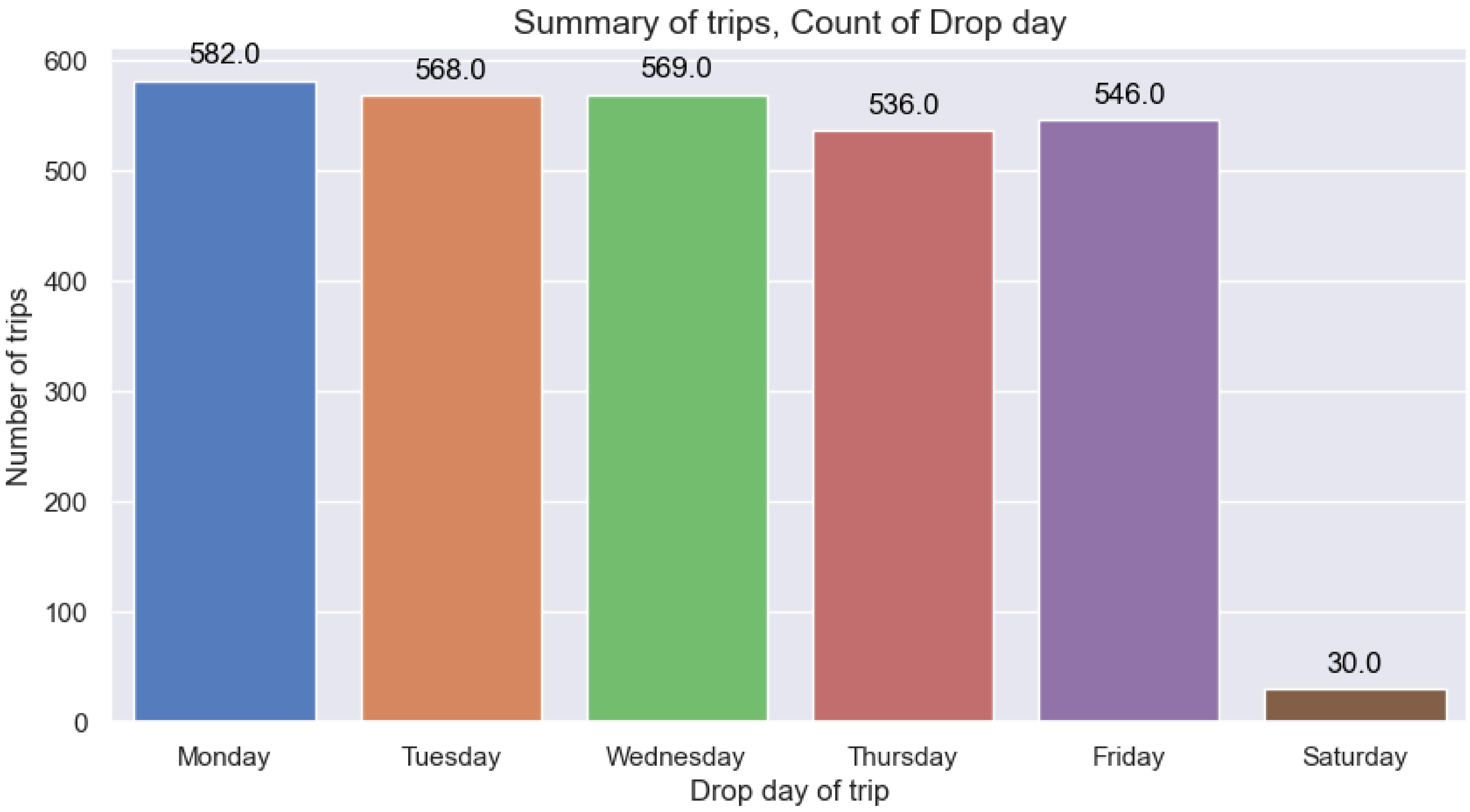
COUNTPLOT FOR PICKUP POINT



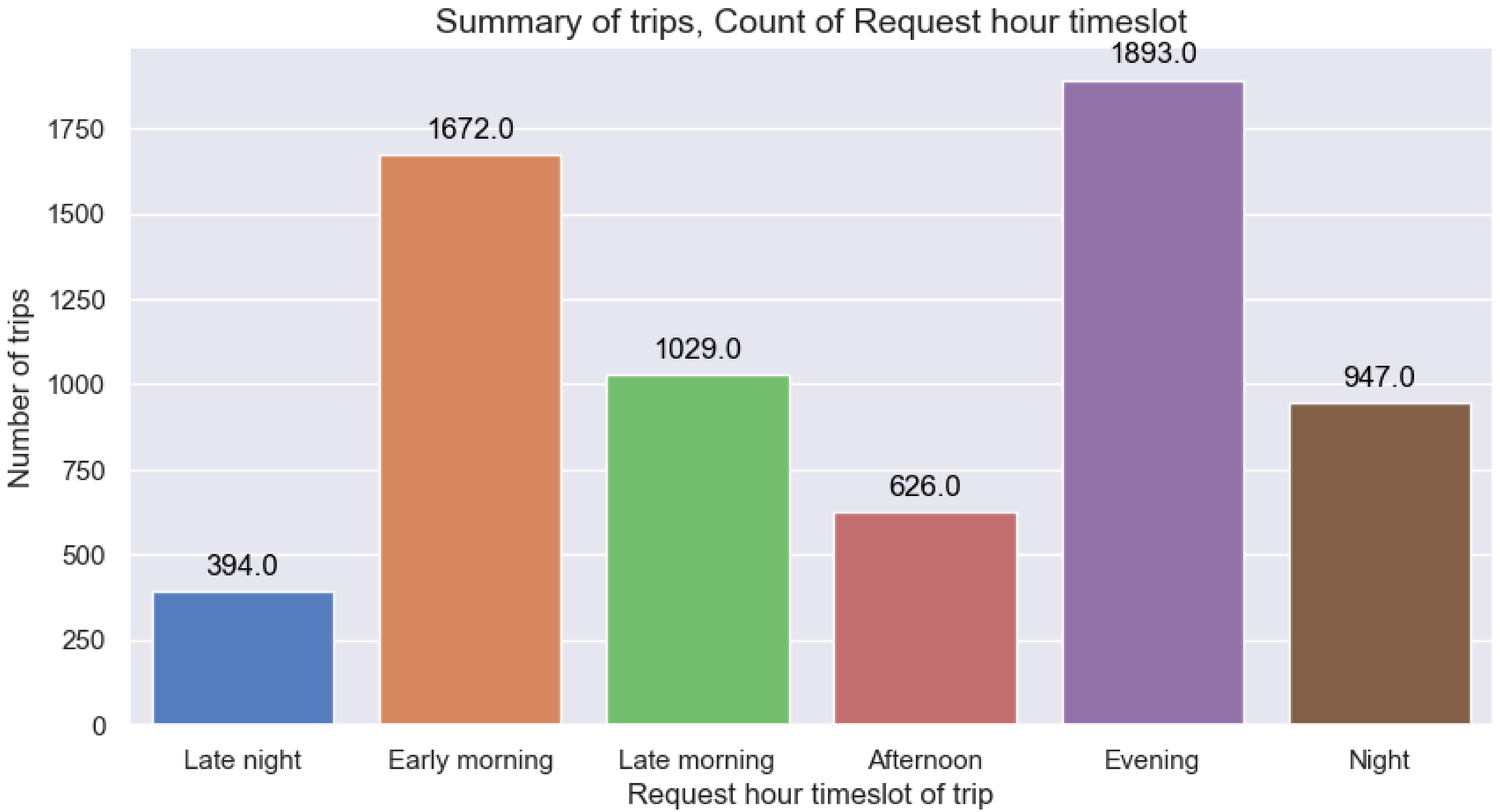
COUNTPLOT FOR REQUEST DAY



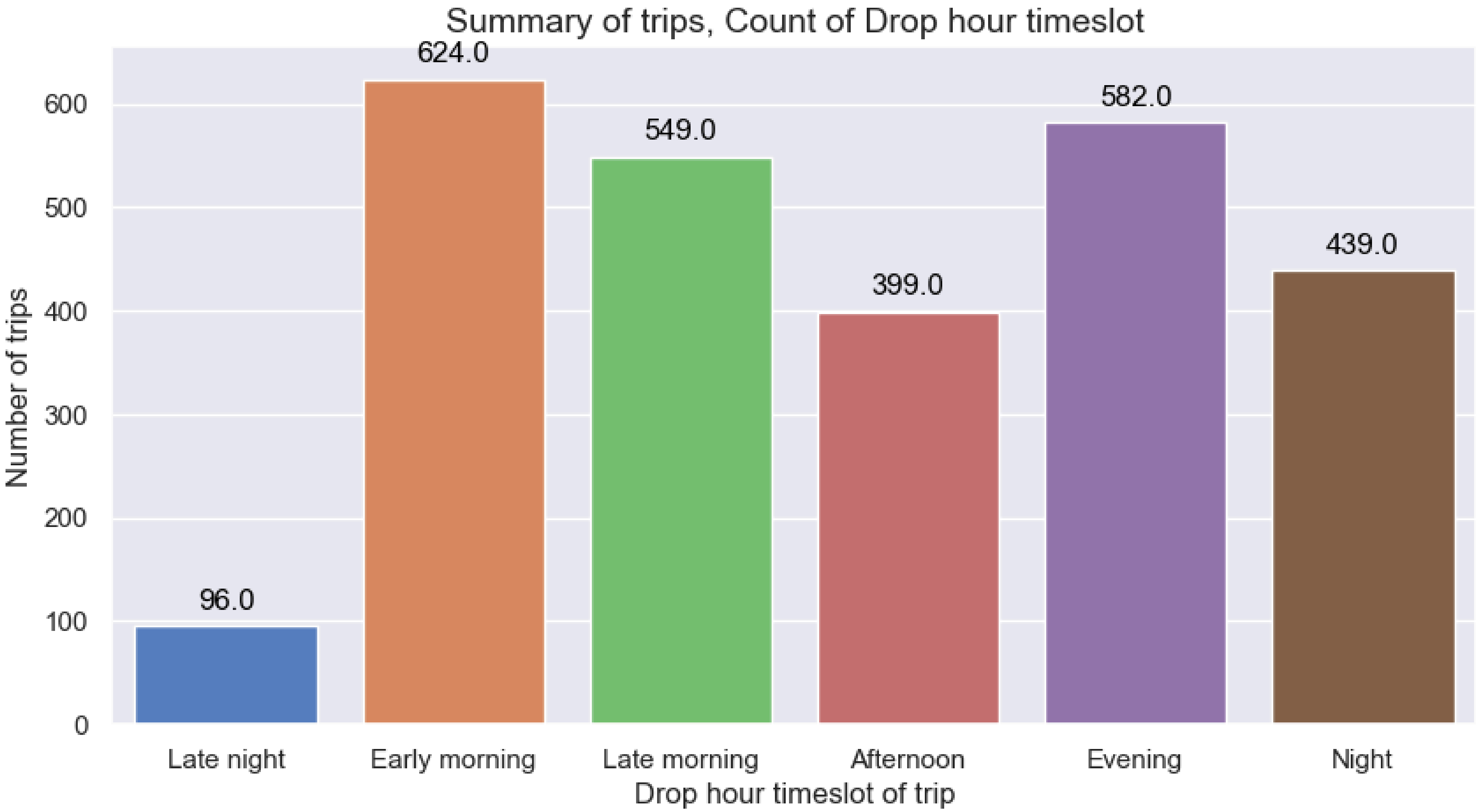
COUNTPLOT FOR DROP DAY



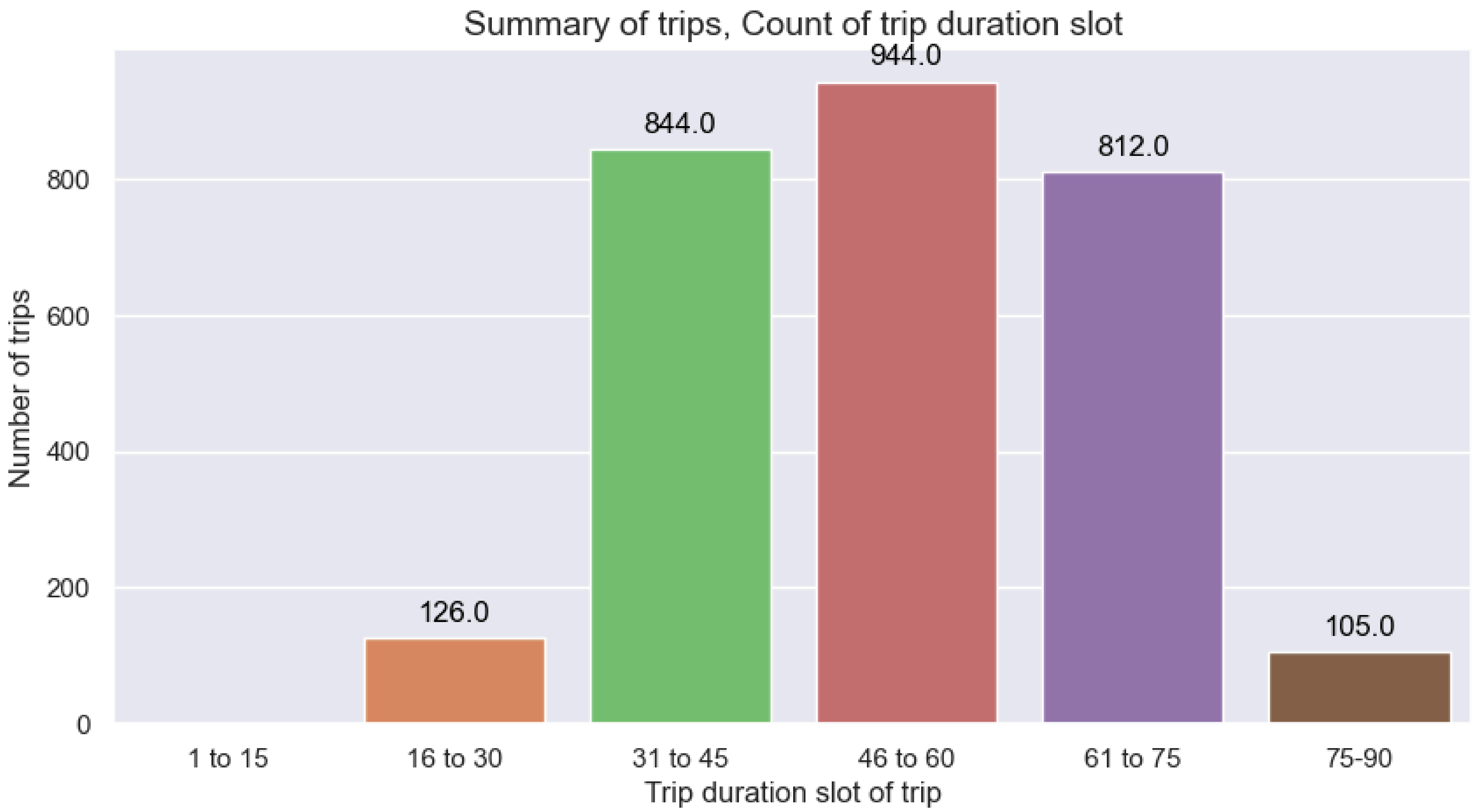
COUNTPLOT FOR REQUEST HOUR TIMESLOT



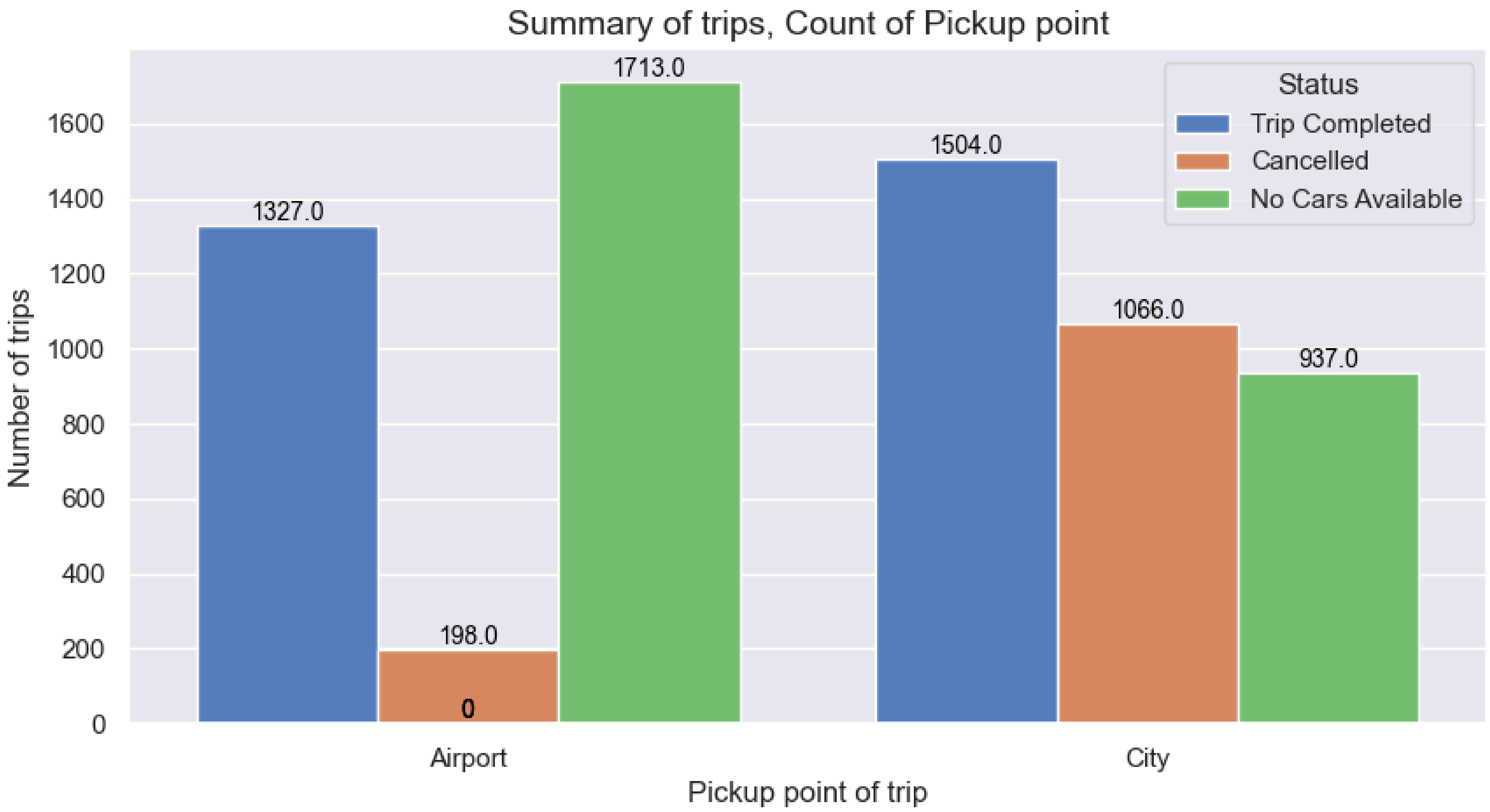
COUNTPLOT FOR DROP HOUR TIMESLOT



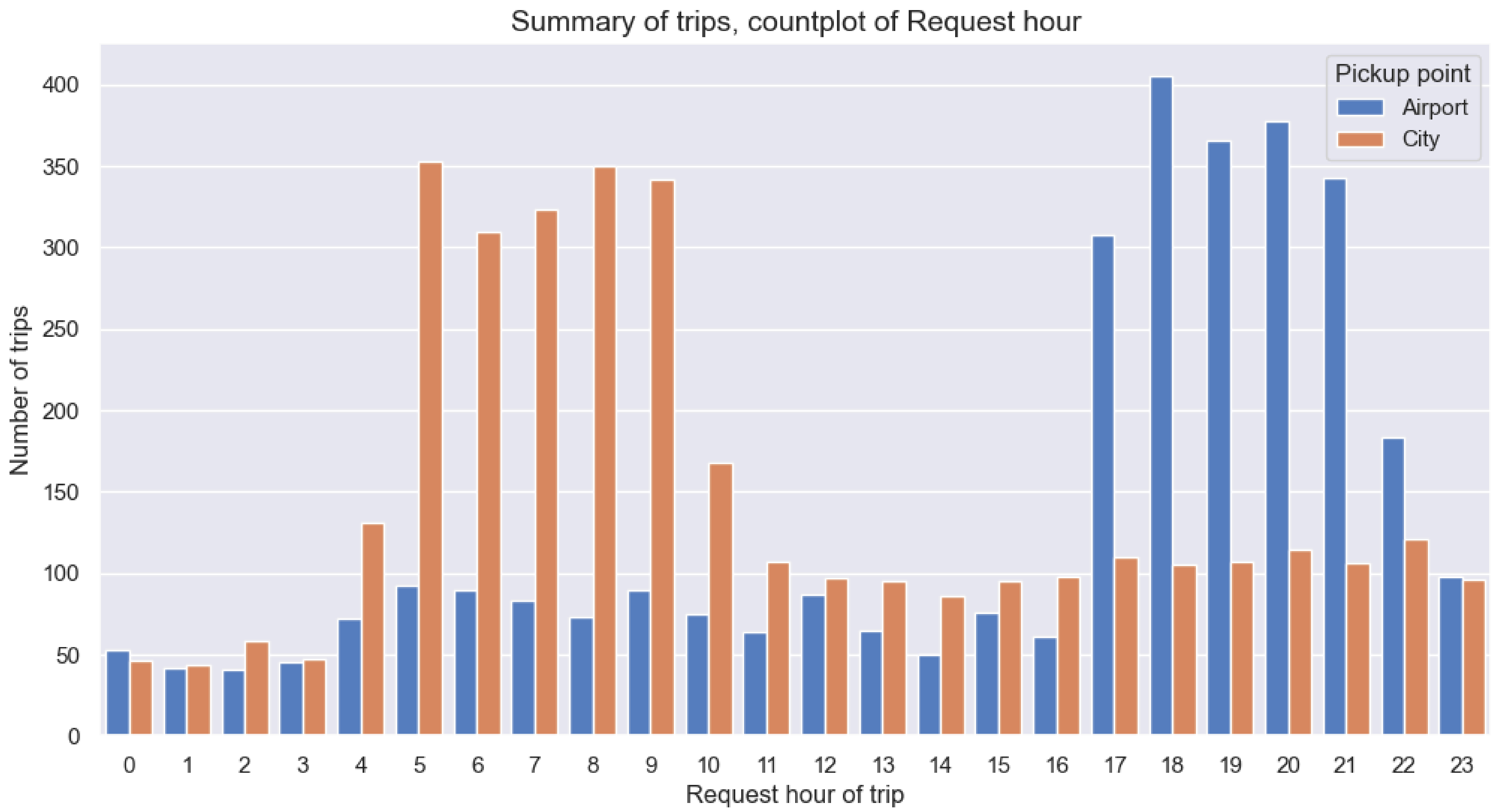
COUNTPLOT FOR TRIP DURATION SLOT



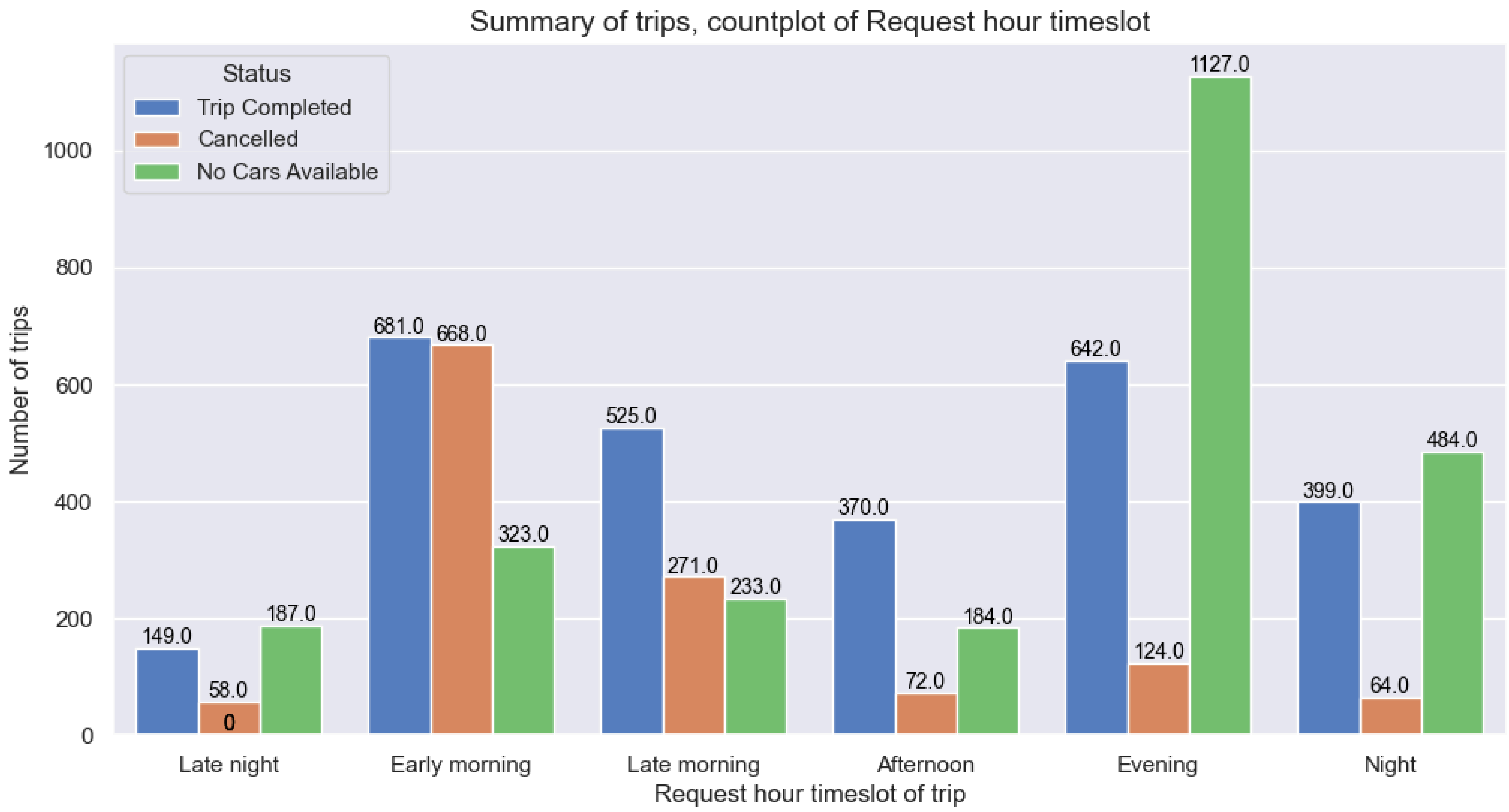
COUNTPLOT FOR PICKUP POINT BASED ON STATUS



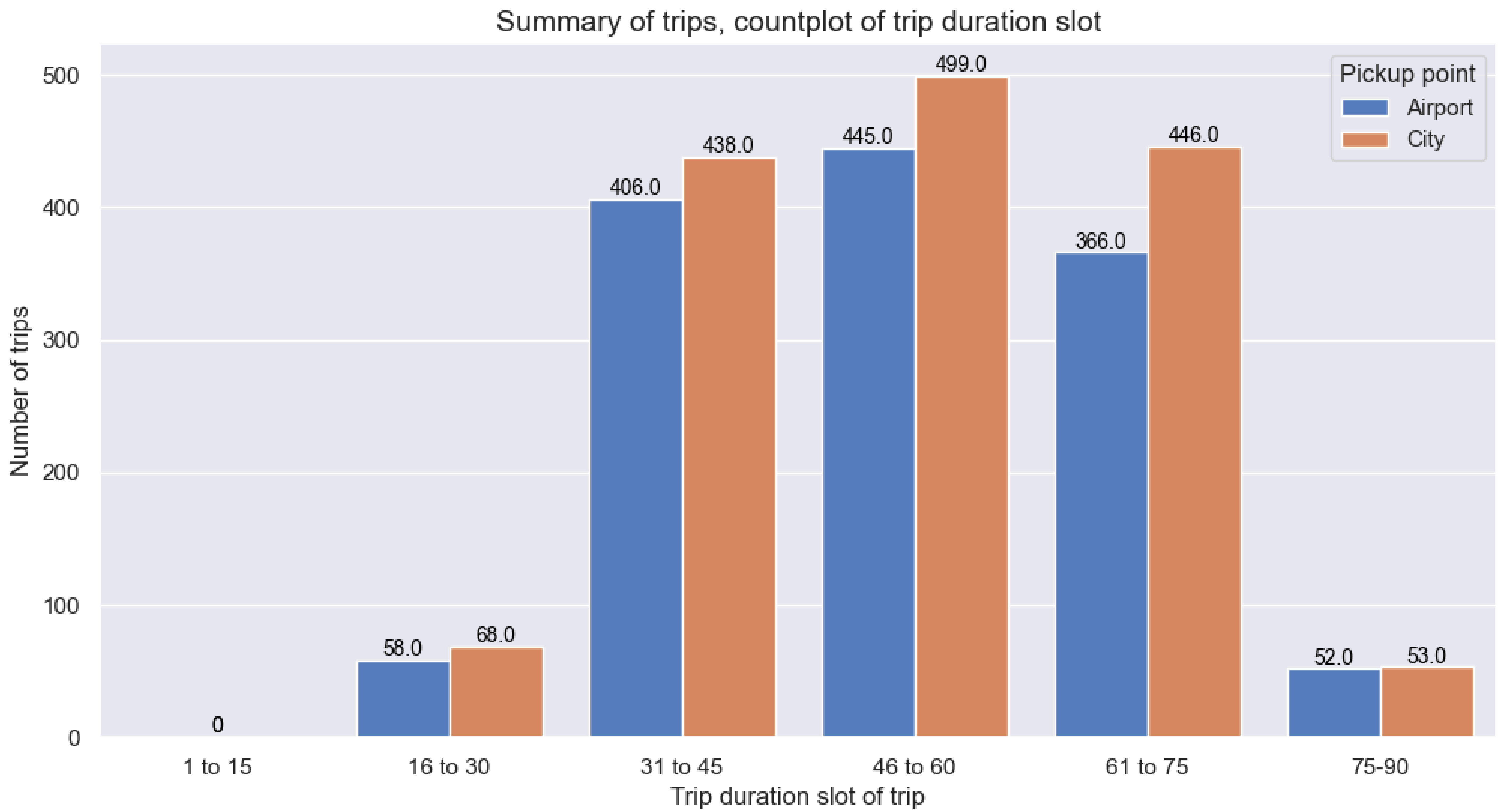
REQUEST HOUR BASED ON PICKUP POINT



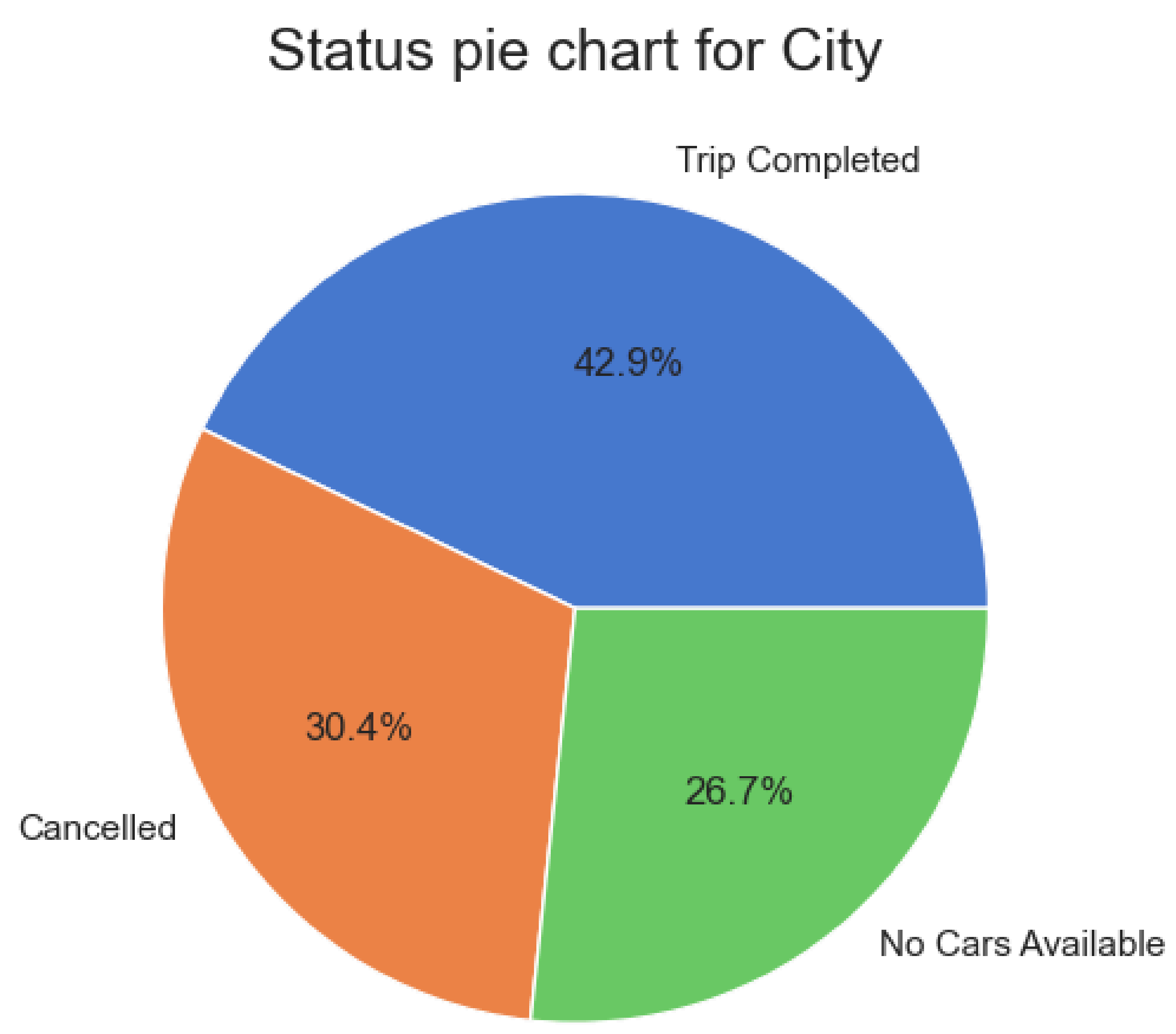
REQUEST HOUR TIMESLOT BASED ON STATUS

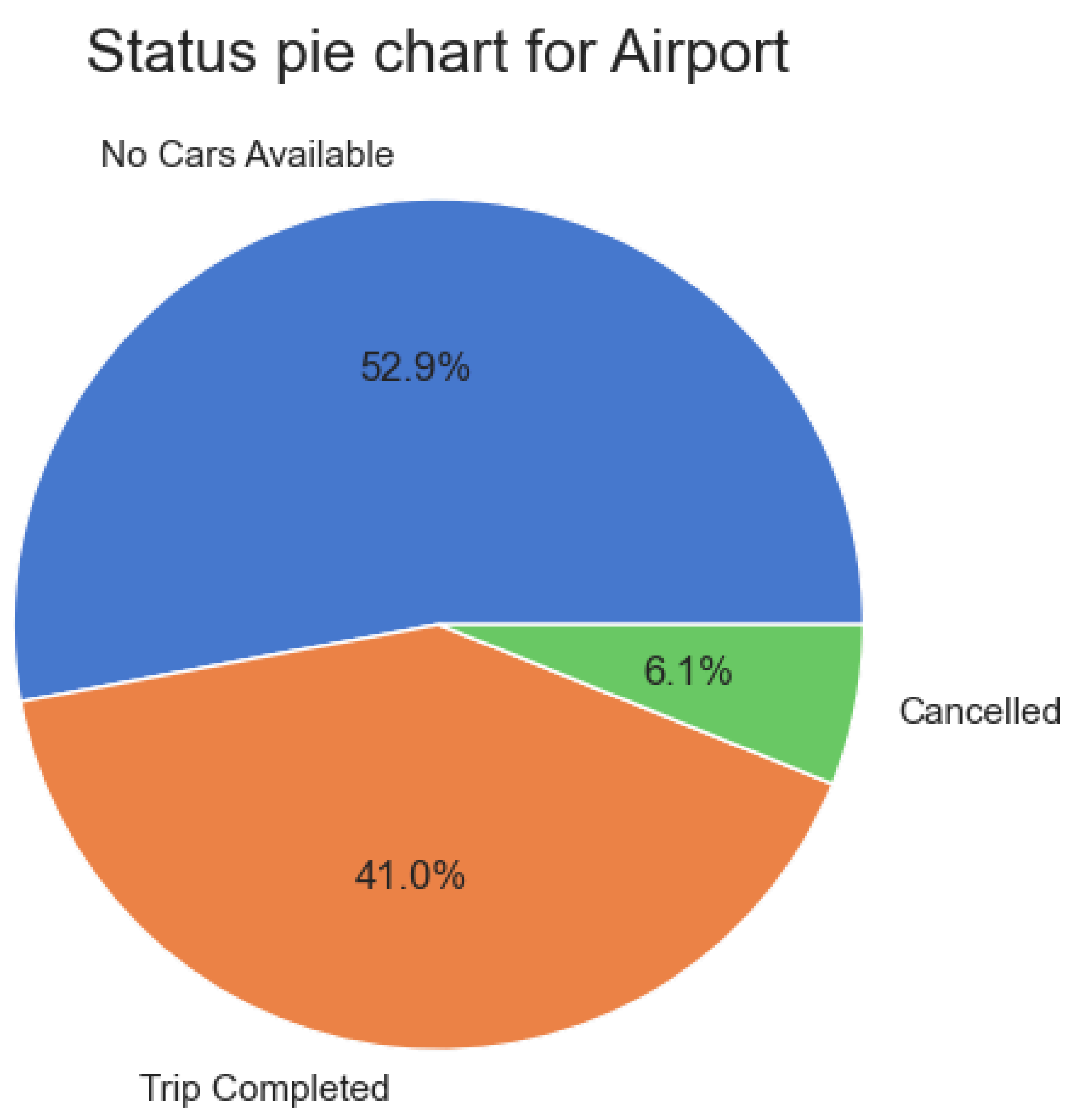


TRIP DURATION SLOT BASED ON PICKUP POINT

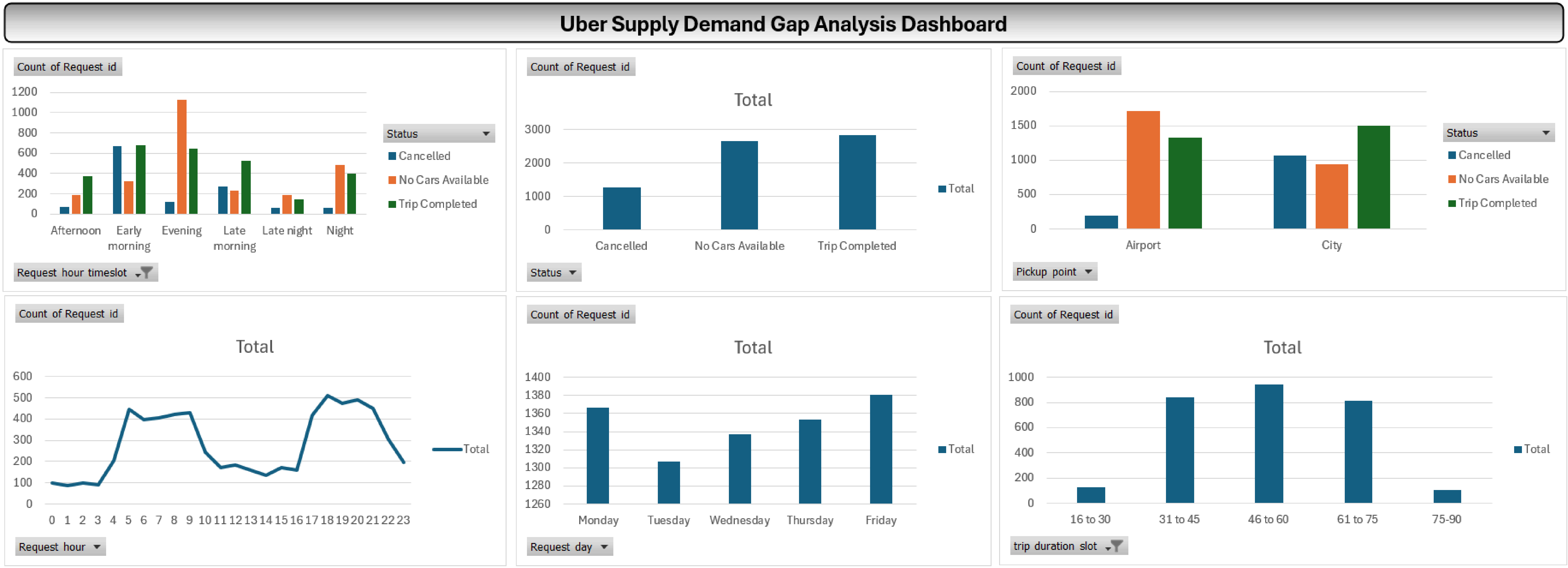


STATUS BASED ON PICKUP POINT IN PIE CHART



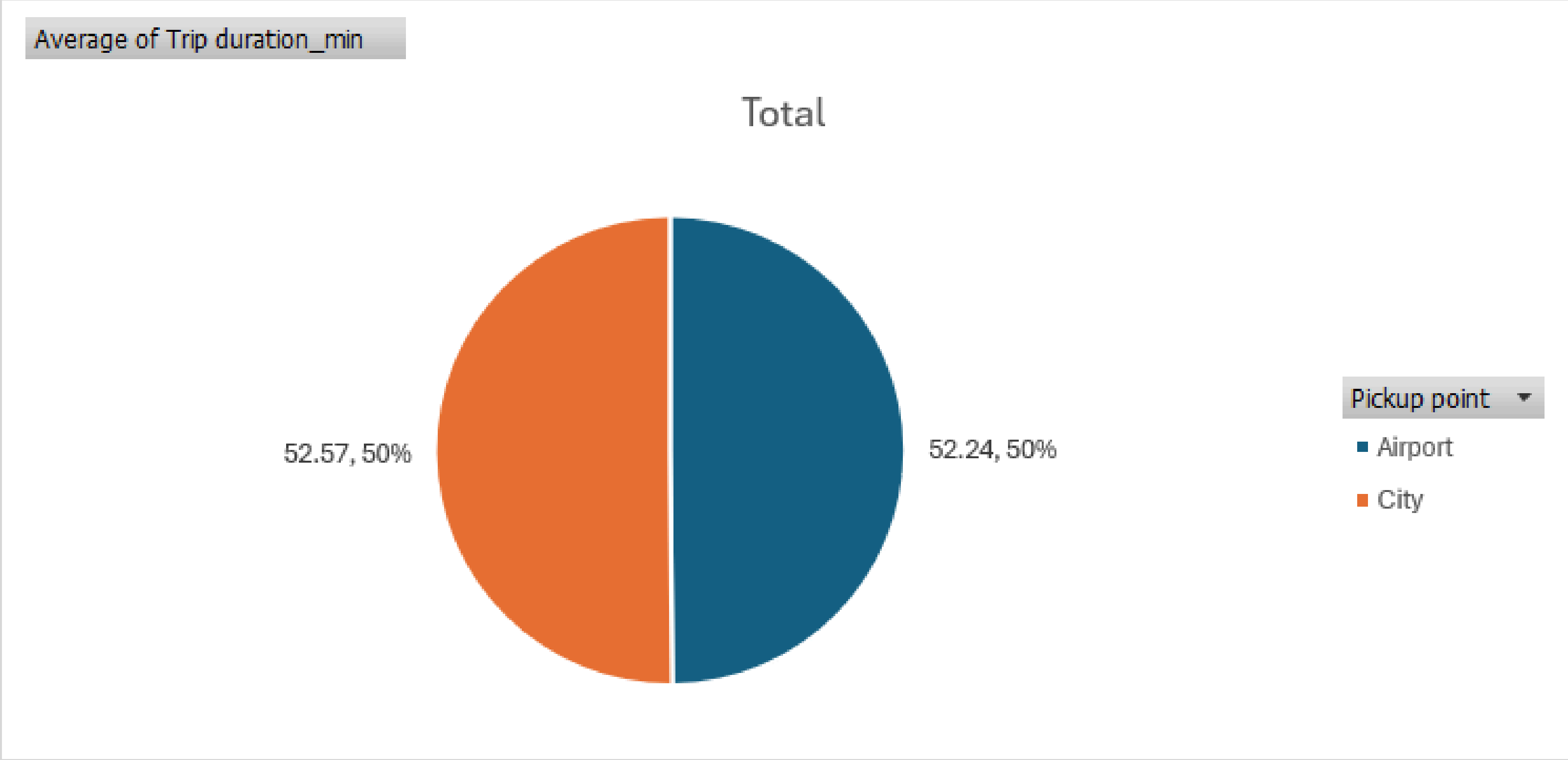


DASHBOARD ON EXCEL



DASHBOARD ON EXCEL

Uber Supply Demand Gap Analysis Dashboard



Count of Request id	Column Labels					
Row Labels	Afternoon	Early mor	Evening	Late morning	Late night	Night
Monday	127	339	385	227	70	185
Tuesday	128	312	376	217	73	175
Wednesday	112	343	349	192	88	213
Thursday	116	338	412	193	86	164
Friday	143	340	371	200	77	210

Count of Request id	Column Labels				
Row Labels	Monday	Tuesday	Wednesday	Thursday	Friday
0	16	16	23	23	21
1	18	10	17	21	19
2	14	16	24	24	21
3	17	14	24	18	19
4	39	43	40	44	37
5	95	76	87	91	96
6	87	84	86	75	66
7	76	74	88	75	93
8	81	78	82	97	85
9	98	92	77	88	76
10	43	49	56	37	58
11	39	42	25	36	29
12	47	34	34	32	37
13	32	27	26	41	34
14	26	32	26	16	36
15	29	38	33	32	39
16	40	31	27	27	34
17	83	85	87	90	73
18	106	99	86	112	107
19	96	97	91	104	85
20	100	95	85	106	106
21	85	101	87	91	85
22	60	47	78	44	75
23	40	27	48	29	50

INSIGHTS

1. High Volume of Unfulfilled Requests
    - A significant number of ride requests were either cancelled by users or marked as 'No Cars Available', particularly during peak demand hours.
2. Demand Exceeds Supply During Rush Hours
    - Most cancellations and unfulfilled trips occurred during Morning Rush (5 AM–9 AM) and Evening Rush (5 PM–9 PM).
    - Driver availability during these time slots is not sufficient to meet demand.
3. City Has Higher Request Volume
    - The City pickup point recorded more requests than the Airport.
    - However, the Airport showed a higher rate of cancellations and no car availability — highlighting a key supply issue in that location.
4. Time Slot Impact
    - The Request Hour Timeslot analysis revealed that Morning Rush has the most service failures, making it the most problematic period operationally.
    - Midday and Late Night have the highest trip success rates but relatively low volume.

INSIGHTS

5. Day-wise Patterns
    - Requests were evenly distributed throughout the week with minor peaks on weekdays.
    - Weekend demand didn't spike significantly, suggesting business-driven commuter patterns.

6. Average Trip Duration
    - Completed trips had a reasonable duration distribution, with most trips falling into the 15–30 minute slot, helping gauge fleet rotation needs.

7. Driver Utilization & Distribution Gaps
    - Drivers were not uniformly available throughout time slots and locations.
    - Some drivers had low utilization, indicating a potential mismatch in driver scheduling.

RECOMMENDATIONS

1. Dynamic Driver Scheduling
    - Increase driver incentives and availability during rush hours, especially 5 AM–9 AM and 5 PM–9 PM.
    - Deploy more drivers to Airport during peak periods, as this area shows high rejection/failure rates.
2. Predictive Allocation System
    - Use historical request patterns to implement predictive driver dispatching — pre-positioning drivers at known hotspots before demand peaks.
3. Optimize Driver-Partner Incentives
    - Launch time-slot-based bonus schemes to encourage driver logins during high-demand periods like early morning and late evening.

RECOMMENDATIONS

4. Improve Rider Communication
    - Notify riders during high failure times and suggest alternative time slots or nearby pickup points with higher fulfillment likelihood.

5. Monitor & Balance Demand Across Locations
    - Real-time dashboards can help track location-based gaps and reassign drivers dynamically.

6. Feedback & Engagement Loop
    - Collect cancellation feedback to refine driver-rider matching algorithms and flag any quality-of-service issues.

CONCLUSION

The Uber Supply-Demand Gap Analysis reveals a clear misalignment between ride request patterns and driver availability, particularly during high-demand hours and at the Airport pickup point. Through a combination of Python-based EDA, SQL ad-hoc queries, and Excel dashboards, we identified the specific time slots and locations that suffer the most from unmet demand.

Implementing targeted operational changes — such as predictive driver allocation, location-specific incentives, and enhanced real-time monitoring — can significantly reduce trip cancellations and improve customer satisfaction. By bridging the supply-demand gap using data, Uber can ensure better fleet efficiency and strengthen its market position in high-volume urban corridors.

THANK YOU

By Anish Chakravorty