Simulating a Ride-Sharing Service Dispatch System

Term Project Assignment

Aaliya Merchant, Nikhil Prabhu, Shiraz Rehmani

MSDS 460: Decision Analytics

December 8th, 2024

Contents

1 Abstract

This paper presents the design, implementation, and analysis of a simulation model for ride-sharing services, focusing on the dynamics of customer requests, driver availability, and ride cancellations. Using the SimPy library, a discrete-event simulation was developed to model ride requests, cancellations, and completed rides under real-world conditions, including weather impacts and surge pricing. The simulation evaluates subsets of 100, 200, and 300 rides to examine scalability and performance. Results highlight the effects of weather on cancellations, surge pricing on customer costs, and driver availability on ride completion rates. These insights provide valuable guidance for optimizing resource allocation, enhancing customer satisfaction, and improving operational efficiency. This paper discusses the methodology, results, and recommendations for future improvements in system design.

Keywords: Driver Availability, Ride Cancellations, Discrete-event Simulation, Weather Impact.

2 Introduction

Ride-sharing services such as Uber and Lyft have become integral to urban transportation, offering convenience and affordability to users while providing flexible income opportunities for drivers. However, managing the complexities of ride requests, cancellations, driver availability, and ride durations presents a significant challenge for these platforms. Understanding how external factors, such as weather conditions, time of day, and surge pricing, influence operational dynamics is critical for improving resource allocation, customer satisfaction, and pricing strategies. These factors play a crucial role in determining both the rider and driver experience, which in turn impacts the efficiency of the system.

This paper explores the use of simulation modeling to analyze the operational processes of a ride-sharing service. The SimPy discrete-event simulation framework is employed to model customer requests, driver availability, ride durations, cancellations, and dynamic pricing adjustments. By incorporating real-world factors such as weather, peak-hour demand, and ride distance, the model simulates operational outcomes under varying conditions. The system's scalability and performance are evaluated by testing subsets of 1,000, 100,000, and 400,000, and 600,000 rides, providing insights into how the platform performs under different workloads.

The primary objective of this research is to develop an analytical tool that enhances decision-making for ride-sharing platforms. By simulating a range of scenarios, this study offers actionable insights into optimizing driver allocation, understanding cancellation trends, and adapting to the impacts of weather-related disruptions and pricing fluctuations. These findings can be applied to improve operational strategies and maximize platform efficiency.

3 Literature Review

Ride-sharing businesses such as Uber and Lyft have revolutionized urban mobility, offering convenience and opportunities while also introducing complexities in operations and decision-making. Simulation can offer a direct and cost-effective way to evaluate different scenarios and features of this complex ecosystem. Tenenboim et al. (2023) utilized simulations to study and emphasize the critical role of understanding driver behavior to optimize ride-sharing models. Their study utilized vehicle trace data from San Francisco to evaluate factors such as driver shift frequency, duration, and locations. These variables were chosen as they significantly affect ride-hailing efficiency, environmental impacts, and service quality. Simulations offered the ability to model these variables to produce real-world insights for policymakers looking to improve transportation systems.

Similarly, Kucharski and Cats (2022) explored the complexities of ride-sharing systems by developing MaaSSim, a simulation framework that models interactions between drivers, and customers. Their research highlights the challenges of balancing profitability with driver incentives and affordable customer pricing. In another study, Guasch et al. (2014) investigated ridesharing systems as an environmentally sustainable solution, emphasizing their potential to reduce costs and congestion. Their simulation demonstrated the effectiveness of real-time matching in enhancing ride quality, particularly in areas with limited public transportation. These studies show how simulations can solve complex challenges in the ride-sharing ecosystem.

While these studies provide critical insights into the behavioral aspects of ride-sharing systems, few have integrated the external variables—weather, ride cancellations, and surge pricing—into a single, comprehensive simulation. This paper aims to bridge that gap by using SimPy to model these factors and understand their impact on platform performance.

4 Exploratory Data Analysis (EDA)

The exploratory data analysis (EDA) focuses on several key metrics related to Uber and Lyft rides, s uch as pricing, ride types, distances, and route distributions. The distribution of cab types (Uber vs. Lyft) reveals a balanced presence of both services in the dataset, as shown in Figure 1. Ride types, categorized by products like Economy, Premium, and Standard, show varying usage patterns across both platforms, as depicted in Figure 2. Uber's average price is $15.80, while Lyft's average price is slightly higher at $17.35. Notably, significant price variations emerge when breaking down by ride type, with Premium and Premium Extra Room rides being the most expensive across both platforms. Additionally, combining the averages for both services provides a comprehensive overview of pricing trends for different ride categories.

In terms of ride distance, both Uber and Lyft share an average distance of 2.19 miles, indicating similar travel patterns, as shown in Figure 4. The price distribution for both platforms, illustrated through histograms in Figure 3, reveals notable differences, with higher prices found in the Premium and Premium Extra Room categories. Route analysis, represented by heatmaps in Figure 5, displays the frequency of rides between different source and destination pairs, highlighting key routes with higher demand for both services.

Correlation analysis between various ride metrics, such as the number of rides, surge multiplier, and distance, reveals weak to moderate relationships as shown in Table 2. For example, a slight negative correlation is observed between the surge multiplier and average distance for Uber rides. This suggests that surge pricing tends to occur more frequently on shorter trips, although this relationship is not strong enough to definitively establish causality.

Interestingly, when analyzing the correlation between weather-related variables (such as temperature, humidity, precipIntensity, and windSpeed) and ride characteristics, the results show very weak or near-zero correlations as shown in Figure 6. For instance, the correlation between price and weather factors such as precipIntensity and windSpeed is almost negligible, with values close to zero (0.000166 and 0.000990, respectively). Despite these weak correlations, we decided to incorporate weather conditions into our simulation modeling, as weather events can still influence real-world ride availability and pricing. This inclusion reflects the belief that, in practice, weather disruptions like rain or fog may lead to adjustments in pricing or cancellations, even if the statistical relationships are not strong.

Overall, this EDA provides valuable insights into the distribution, pricing, and other key aspects of Uber and Lyft rides, which will serve as the foundation for further analysis and simulation modeling in the study. By incorporating these findings into the simulation, we can explore how different operational factors influence the system's behavior and performance.

5 Research Design and Methodology

This study simulates ride-sharing events to analyze factors affecting ride outcomes, such as cancellations, price adjustments, and ride durations, for Uber and Lyft. Using historical data and an event-driven simulation framework, the model incorporates key variables like driver availability, demand fluctuations, and external disruptions, including weather.

The simulation is built using SimPy, a Python-based discrete-event library, which models customer requests, driver responses, ride completions, cancellations, and dynamic pricing. To assess scalability, the system is tested with different ride subsets (1,000, 100,000, 400,000 and 600,000).

Combining statistical analysis and simulation, the study evaluates how changes in key variables influence system behavior, providing insights into operational dynamics and guiding potential improvements in efficiency, customer satisfaction, and cost-effectiveness.

5.1 Research Design

The primary dataset consists of historical ride-sharing data, containing attributes such as ride details, weather conditions, and geospatial information. Ride-specific attributes include cab type (`cab_type`), product ID (`product_id`), source and destination locations, distance (`distance`), and initial price (`price`). Weather-related features include temperature (`temperature`), precipitation intensity (`precipIntensity`), humidity (`humidity`), and short summary conditions (`short_summary`), such as "Rain" or "Foggy." Geospatial data captures latitude and longitude for ride origins and destinations.

To evaluate scalability and variability in outcomes, the simulation is tested on three subsets of data: 100, 200, and 300 rides. This approach enables comparative analysis of the simulation's robustness under varying workload conditions.

5.2 Simulation Design

The simulation is implemented using the SimPy framework, which provides a process-based discrete-event simulation structure. Each ride request is treated as an independent process that progresses through several key stages:

5.2.1 Driver Availability Check

The availability of drivers is determined probabilistically based on the `cab_type` and `product_id`. Standard services such as UberX and Lyft exhibit higher availability rates than premium services like Uber Lux or Lyft Lux. Availability rates are set at 90% for Uber and 85% for Lyft, with reductions applied for premium rides. Random delays ranging from one to five minutes are introduced to simulate real-world variability in driver allocation.

5.2.2 Price Adjustment

Ride prices are dynamically adjusted based on ride attributes, weather conditions, and demand factors. The price adjustment incorporates the following components:

- **Weather Conditions:** Prices are increased by multipliers for adverse weather, such as +50% for rain, +30% for light rain, and +20% for foggy conditions.
- **Ride Type:** Luxury rides, such as Lyft Lux, incur a 50% premium, while shared rides, such as Lyft Line, receive a 25% discount.
- **Surge Multiplier:** Demand surges during peak hours are modeled through a surge multiplier applied to the base price.

5.2.3 Cancellation Logic

Cancellations are probabilistically modeled using factors such as ride distance, peak hours, and weather conditions. Shorter rides (less than 2 miles) increase the cancellation chance by 10%, while rides during rush hours (7–10 AM and 5–8 PM) add a 5% increase. Adverse weather, such as rain or high precipitation intensity, further elevates the cancellation probability.

While the exploratory data analysis (EDA) revealed weak correlations between weather variables (such as precipitation and wind speed) and key metrics like price and distance, we still decided to incorporate weather effects into the simulation. This decision was driven by the understanding that weather can still influence real-world ride availability and pricing, even if these effects are not captured by basic statistical correlations. Weather-related factors, such as rain, fog, or light drizzle, are accounted for with multipliers that adjust ride prices and cancellation probabilities accordingly, reflecting realistic conditions that could affect ride behavior. By including these weather effects, we aim to simulate real-world dynamics more accurately.

5.2.4 Ride Completion

For rides that are not canceled, the simulation calculates the expected duration based on an average travel speed of 30 mph. The formula used for duration estimation is $(\texttt{distance}/30) \times 60$, representing the ride duration in minutes. The process concludes by logging the ride as completed with detailed metadata, including the final price and duration.

5.3 Simulation Parameters

Key parameters and configurations are defined to reflect real-world scenarios.

- **Driver Availability:** Uber exhibits a base availability of 90%, while Lyft is set at 85%. Premium rides reduce these values to 80% and 75%, respectively.
- **Price Multipliers:** Clear weather has a multiplier of 1.0, rain increases it to 1.5, and foggy conditions result in 1.2. Ride types such as Lyft Line and Lyft Lux apply multipliers of 0.75 and 1.5, respectively.
- **Cancellation Probability:** The base cancellation rate is 10%, adjusted dynamically for distance, time of day, and weather conditions.

5.4 Testing Design

The simulation is executed on three subsets of the data, comprising 100, 200, and 300 rides. These subsets facilitate a comparative evaluation of system behavior, scalability, and performance across different workload sizes.

5.5 Results Interpretation

The simulation generates a detailed event log that captures ride outcomes, including cancellations, price adjustments, and durations. Key metrics derived from the results include:

- **Cancellation Rates:** Overall and segmented by ride type and cab type.
- **Price Adjustments:** Mean adjusted prices for Uber and Lyft rides across various conditions.
- **Driver Availability:** Proportion of rides affected by driver unavailability.
- **Ride Durations:** Distribution and average duration of completed rides.

6 Results and Interpretation

In the initial phase of the simulation, only the first 1000 rides were utilized to model aggregated trends in ride cancellations and completions over 24 hours, segmented into 5-hour intervals. In the resulting visualization, the red line represents ride cancellations, while the blue line corresponds to completed rides. The cancellation rate remains relatively stable throughout the day, with minor fluctuations, though a noticeable peak occurs at the 15-hour mark as seen in Figure 8. This spike could be attributed to factors such as increased traffic, adverse weather conditions, or operational challenges that impact ride availability. In contrast, the ride completions exhibit a more dynamic pattern. A peak is observed at the beginning, probably reflecting a higher demand for transportation during the beginning hours. As the day progresses, the number of completed rides gradually decreases, with a major decrease from the 15th hour mark to the 20th hour mark. This may indicate a reduction in both demand and driver availability as the day approaches.

The second simulation, based on 100,000 rides, revealed insightful trends in ride cancellations and completion. The highest number of completed rides was approximately 32 which occurred during the 5-hour block, gradually decreasing to 20 by the 20-hour block as seen in Figure 12. In contrast, cancellations peaked during the fifth-hour block, with 7 rides canceled and the 15-hour block, with 9 rides cancelled. Following these peaks, cancellations declined steadily, reaching their lowest point (5 rides) around the 10th and 20th hour block. These results highlight a period of high operational efficiency at the beginning and at the 15th-hour blocks, characterized by a high completion rate and minimal disruptions. However, the upward trend in cancellations during the later hours, particularly in the 15th-hour block, suggests potential challenges such as driver fatigue or increased traffic congestion. These findings emphasize the need for targeted strategies to address late-hour inefficiencies and maintain consistent service quality.

The third simulation, based on 400,000 rides, demonstrated that the highest number of completed rides occurred at the 15th-hour mark, with approximately 30 rides completed. This reflects a consistent pattern of operational efficiency, with a steady increase from 25 completions of the ride in the previous 5-hour block as seen in Figure 16. However, in the 20th-hour block, completions decreased slightly to around 25. In contrast, ride cancellations began at approximately 13 events in the initial time block but decreased significantly to about 5 cancellations by the 5th-hour mark. Notably, during the 15th to 20th-hour blocks, there was a gradual rise in cancellations, which may indicate late-hour challenges such as increased traffic congestion or potential driver fatigue, ultimately impacting service reliability.

The third simulation, based on 600,000 rides, demonstrated that the highest number of completed rides occurred at the 15th-hour mark, with over 30 rides completed as seen in Figure 20. This reflects a consistent pattern of operational efficiency, with a steady increase from 25 completions of the ride in the previous 5-hour block. However, in the 20th-hour block, completions decreased slightly to 25. In contrast, ride cancellations began at a little more than 5 events in the initial time block but increased to about 9 cancellations by the 5th-hour mark. Between the 5-hour mark 20-hour mark we can see a decline in amount of cancellations, with the complpetion rate increasing during the 10th-hour mark. Notably, during the 15th to 20th-hour blocks, there was a gradual rise in cancellations, which may indicate late-hour challenges such as increased traffic congestion or potential driver fatigue, ultimately impacting service reliability.

7 Discussion

Throughout the simulations, many different patterns were observed to shed light on the dynamics of ride-hailing operations for both Uber and Lyft. From the simulations, midday looked to be the most efficient (the least amount of cancellations), it generally peaked during those times and early evening hours. This was specifically seen during the 10th to 15th-hour block. This can be attributed to the high customer demand and operational readiness during peak hours. With the peaks in completions, there conversely was a spike in cancellations, which were found to be the most during the early morning hours or late nights. It can be specifically seen that the 20th-hour block, is the highest for cancellations. These are attributed to factors such as driver fatigue or traffic congestion. That being said, the supply and demand fluctuate as these underscore the complexities of creating such a balance, of high surge prices based on external variables. Furthermore, external factors such as price adjustments and adverse weather conditions can also play a role in either more completions or cancellations. The role is significant as it will take a lot to maintain operational profitability, which influences user behavior. These insights suggest that time-sensitive operational strategies, such as reallocating drivers during peak hours, and implementing mitigation measures for late-night cancellations, could significantly enhance service reliability and customer satisfaction.

8 Conclusion

In conclusion, the simulation provides an overview of the dynamics of ride-sharing services. These simulations have highlighted key patterns through both completions and cancellations over time. As seen through the simulations, it concisely peaked more often during the 10th to 15th-hour block, which indicates a high demand and operational efficiency. About the cancellations, it can be concluded that the 15th-hour mark suggested many challenges as many of them were high during those periods of time. This can be a sign that many more rides were taken during, which results in many more completions and cancellations. These findings emphasize the importance of targeted strategies to optimize driver allocation. With the addition of weather and surge pricing into operational planning, ride-sharing platforms can help enhance customer satisfaction and improve overall system performance.

For the future work, incorporating more granular real world data such as driver fatigue, detailed traffic patterns, or customer wait could help create a more refined simulation. Additionally, with the use of machine learning algorithms it can help create dynamic diver allocation and real time price adjustments which will also help with predictive capabilities of the model. Furthermore, by utilizing customer satisfaction rating and loyalty metrics could also provide a holistic view of the ride-sharing ecosystem. By utilizing these areas, ride-sharing platforms can achieve better quality for both drivers and passengers, improved resource optimization, and a more accurate forecasting.

References

Kucharski, R., & Cats, O. (2022). MaaSSim: An agent-based simulation framework for analysis of mobility as a service. *Transportation Research Part C: Emerging Technologies, 140*, 103707. `https://doi.org/10.1371/journal.pone.0269682`

Guasch, A., Figueras, J., Fonseca i Casas, P., Montañola-Sales, C., & Casanovas-Garcia, J. (2014). Simulation analysis of a dynamic ridesharing model. *Proceedings of the 2014 Winter Simulation Conference*, 1965–1977. `https://ieeexplore.ieee.org/document/7020043`

Tenenboim, E., Gerber, A., & Abernethy, J. (2023). Understanding driver participation in ride-hailing services through vehicle trace data. *Transportation Research Part A: Policy and Practice, 170*, 103512. `https://rosap.ntl.bts.gov/view/dot/72993/dot_72993_DS1.pdf`

Mohan, S. (2020). Rideshare dataset. *Kaggle*. `https://www.kaggle.com/datasets/mohan28169/rideshare-kagglecsv`

9 Appendix

9.1 EDA Results

Table 1: Average Prices for Lyft, Uber, and Combined Services

Category	**Lyft (\$)**	**Uber (\$)**	**Combined (\$)**
Economy	6.03	8.75	7.44
Premium	20.42	20.52	20.45
Standard	9.61	9.77	9.72
Premium Extra Room	32.32	30.29	31.27
Standard Extra Room	15.31	15.68	15.50

Table 2: Correlations for Lyft and Uber

	Lyft			Uber		
	`num_rides`	`avg_surge_multiplier`	`avg_distance`	`num_rides`	`avg_surge_multiplier`	`avg_distance`
`num_rides`	1	0.11016	0.0676291	1	nan	-0.118804
`avg_surge_multiplier`	0.11016	1	-0.0928518	nan	nan	nan
`avg_distance`	0.0676291	-0.0928518	1	-0.118804	nan	1

	price	distance	temperature	humidity	precipIntensity	windSpeed
price	1.000000	0.345061	-0.000084	-0.001238	0.000166	0.000990
distance	0.345061	1.000000	-0.002884	-0.003901	-0.000256	0.002277
temperature	-0.000084	-0.002884	1.000000	0.313853	0.182724	0.058655
humidity	-0.001238	-0.003901	0.313853	1.000000	0.417558	-0.207223
precipIntensity	0.000166	-0.000256	0.182724	0.417558	1.000000	0.307369
windSpeed	0.000990	0.002277	0.058655	-0.207223	0.307369	1.000000

Table 3: Correlation Matrix of Variables

9.2 EDA Visuals

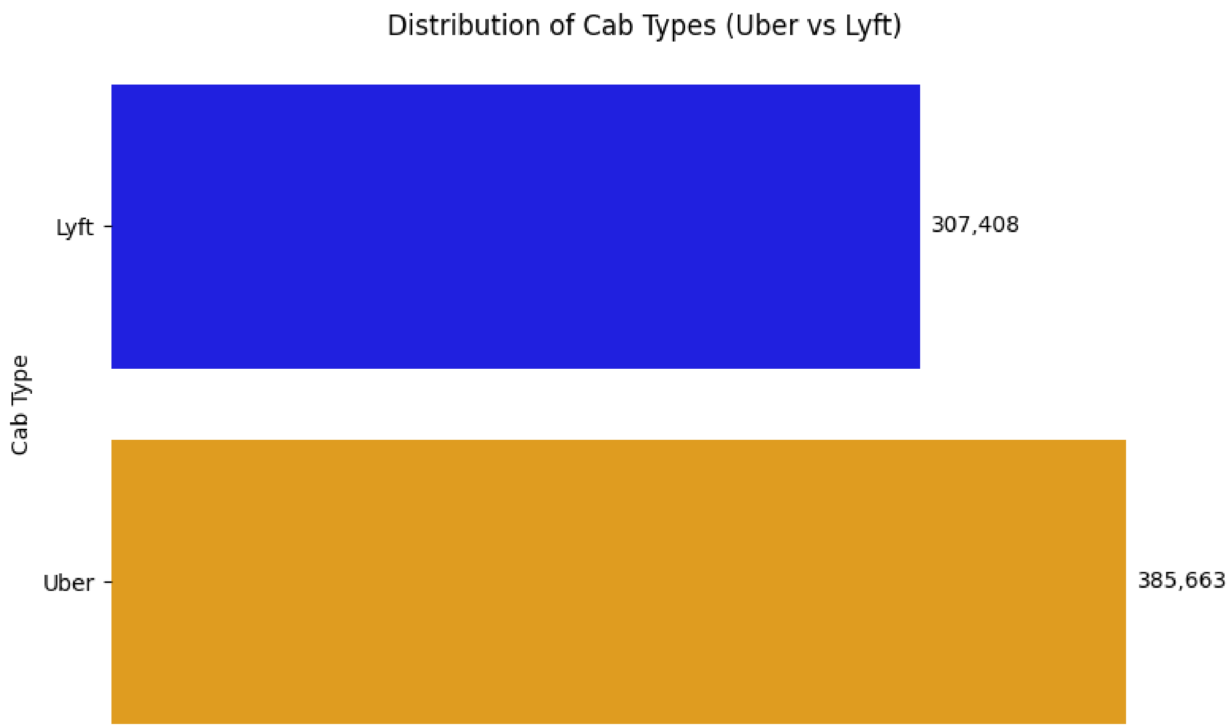

Figure 1: Distribution of Cab Types (Uber vs Lyft)

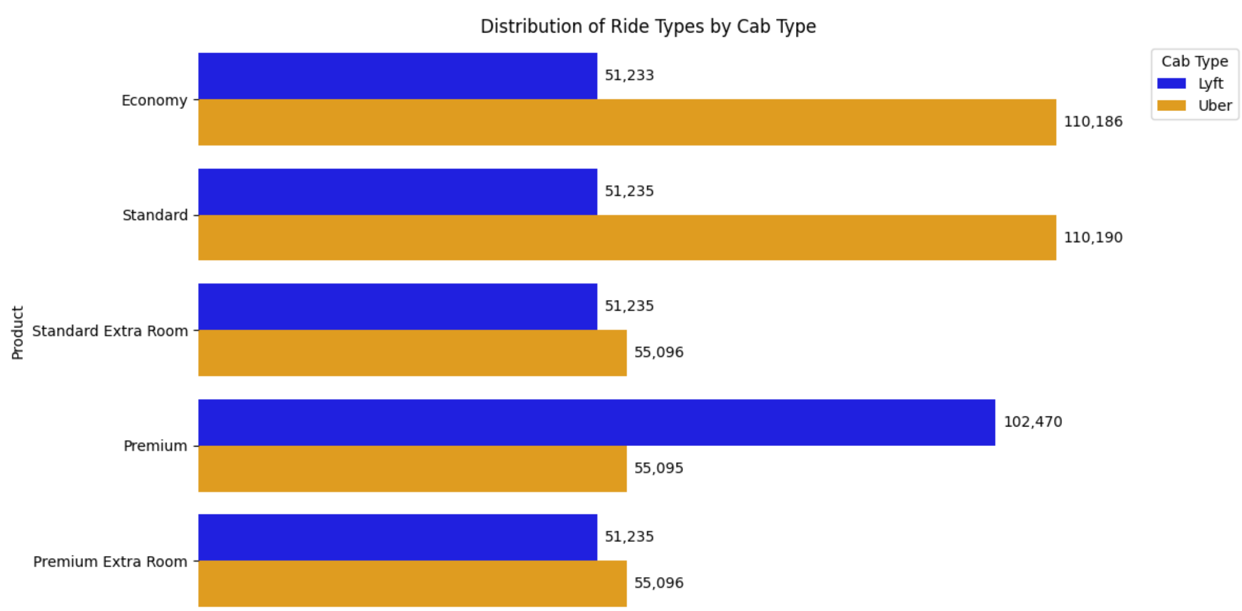

Figure 2: Distribution of Ride Types by Cab Type

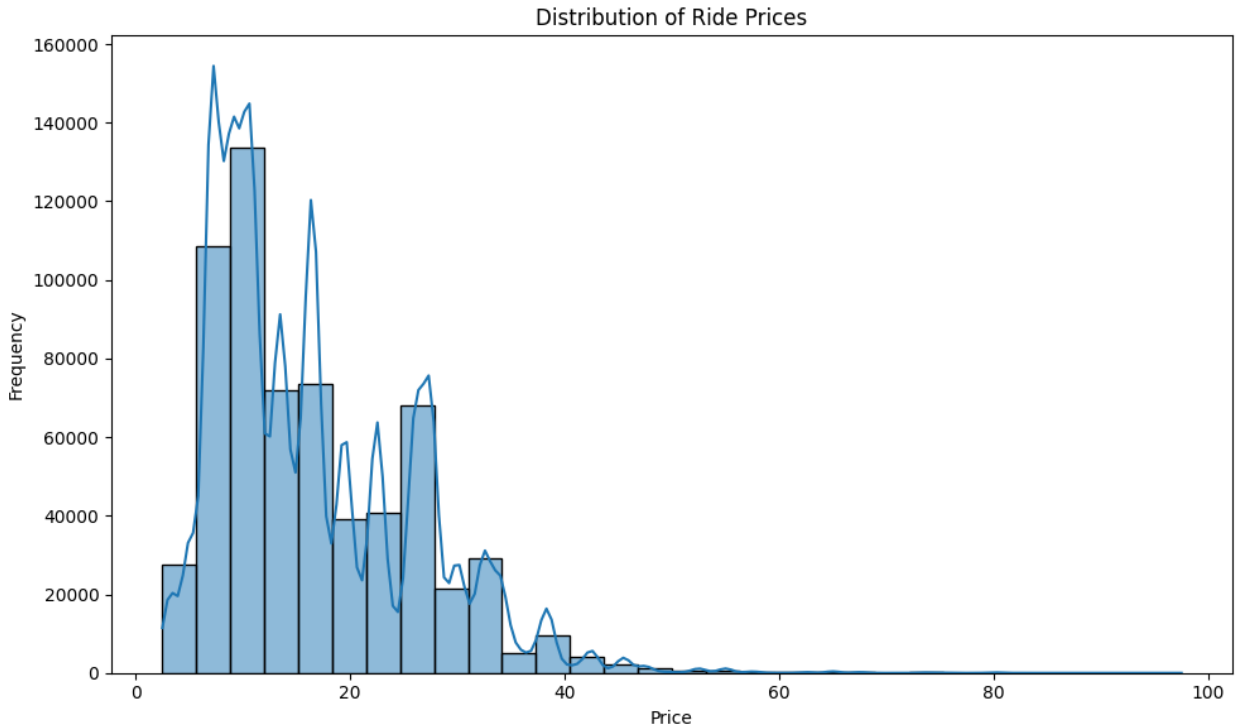

Figure 3: Price Distribution by Cab Type

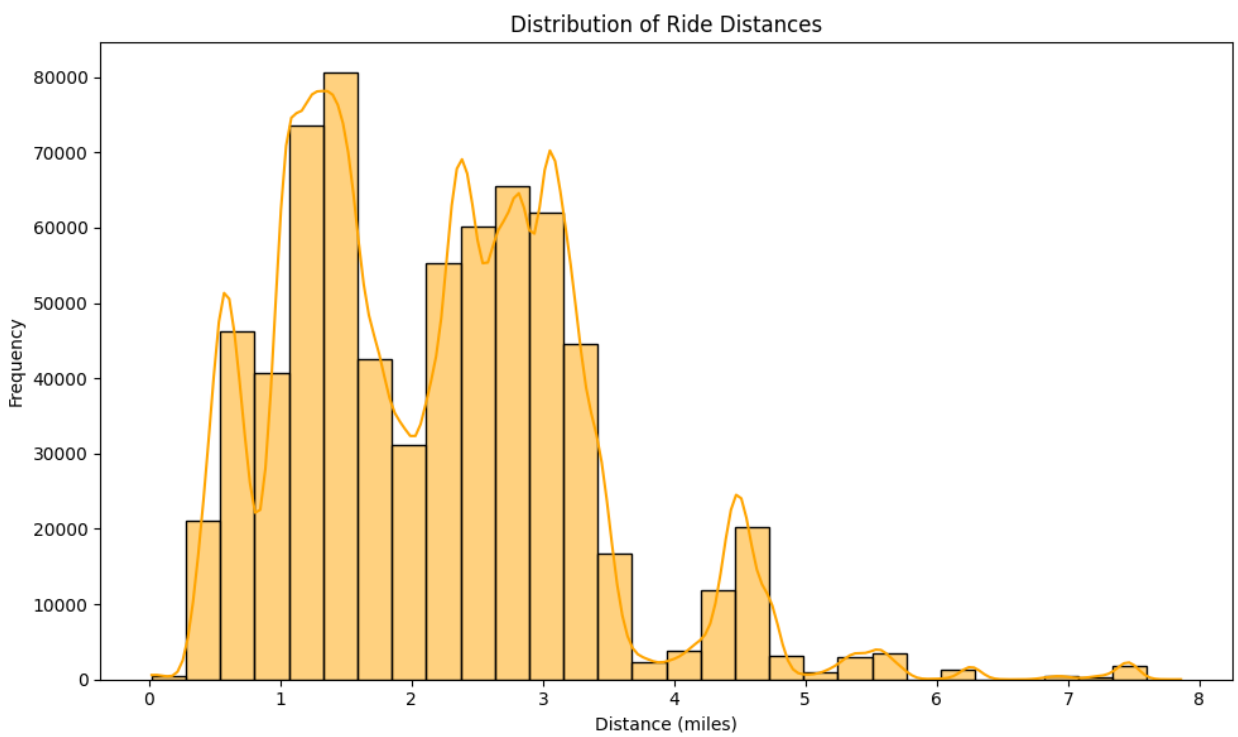

Figure 4: Distribution of Ride Distances

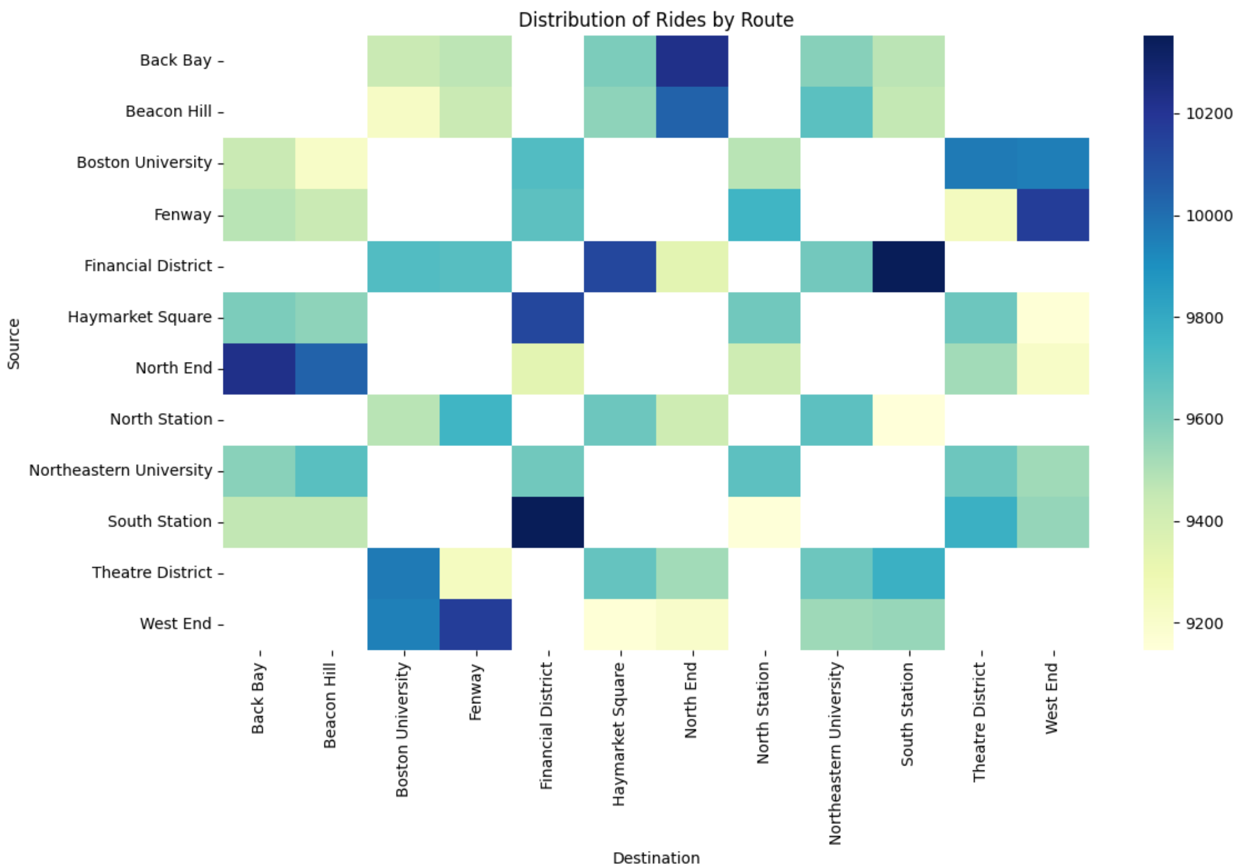

Figure 5: Distribution of Rides by Route

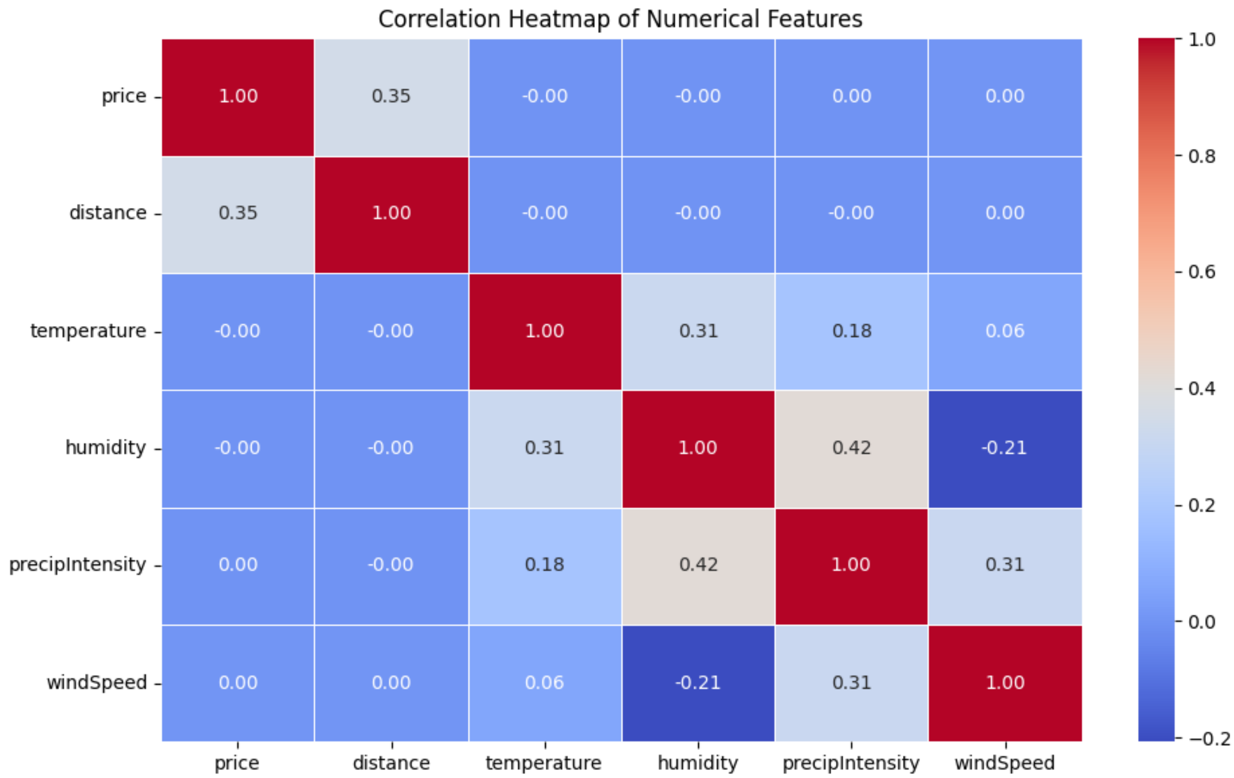

Figure 6: Correlation Heatmap of Numerical Features

9.3 Simulation 1,000

Output:

```
Overall Cancellation Rate: 9.64%

Average Price by Ride Type and Cab Type:
cab_type  product
Lyft      Economy                  4.31
          Premium                 31.19
          Premium Extra Room      50.36
          Standard                 9.75
          Standard Extra Room     15.27
Uber      Economy                  3.49
          Premium                 39.75
          Premium Extra Room      58.21
          Standard                12.39
          Standard Extra Room     20.91
Name: price, dtype: float64
20.531790633608818

Driver Unavailability Rate: 10.19%

Completion Rate by Cab Type:
cab_type
Lyft    0.317568
Uber    0.376744
```

```
dtype: float64

Average Ride Duration for Completed Rides: 4.61 minutes
Completion Rate by Cab Type:
cab_type
Lyft    0.344444
Uber    0.344444
dtype: float64

Average Ride Duration for Completed Rides: 4.04 minutes
```

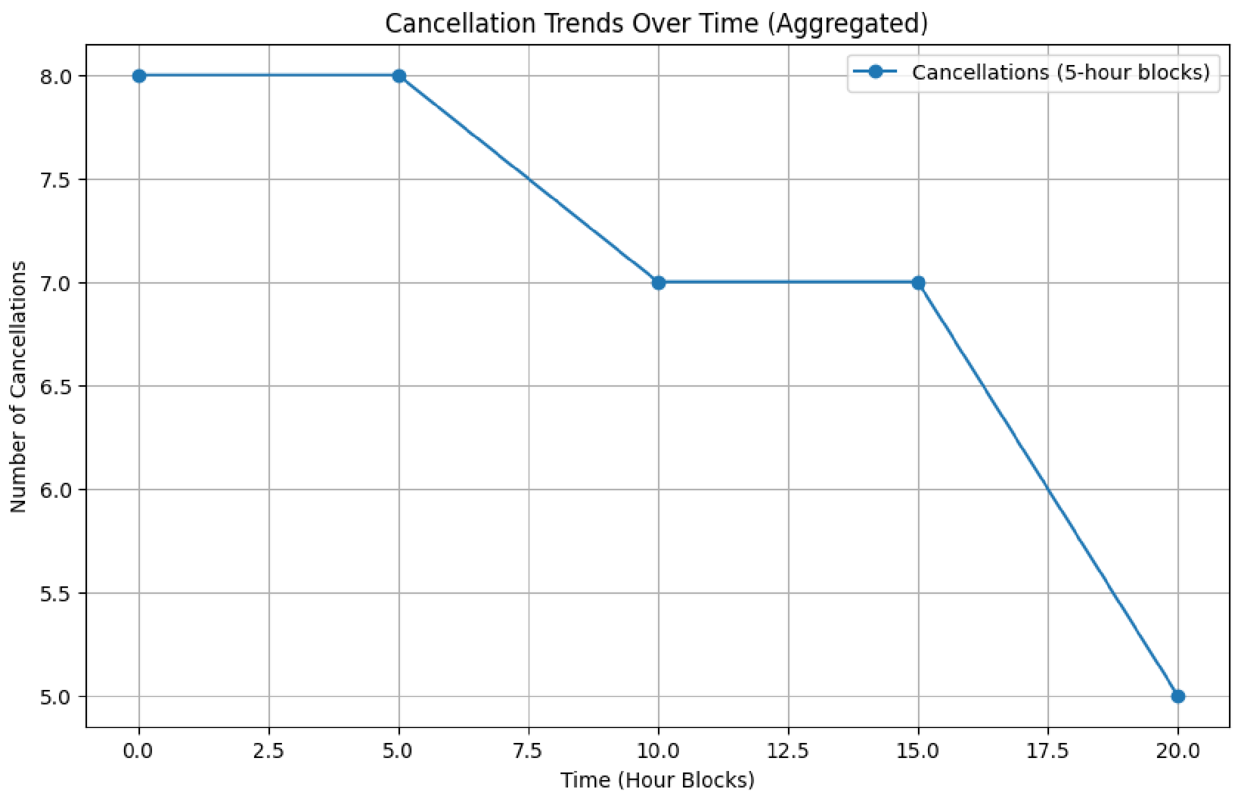

Figure 7: Aggregated Cancellation Trends Over Time: Simulation of 1,000 Rides

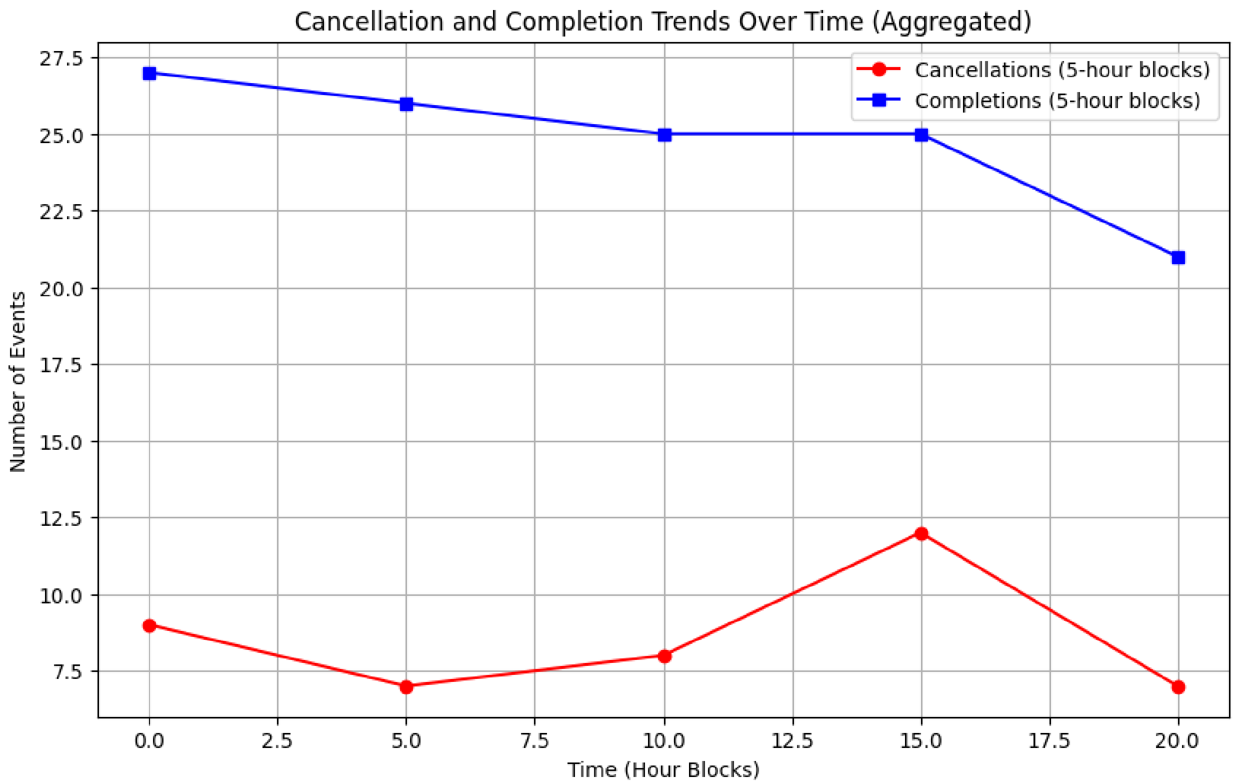

Figure 8: Aggregated Cancellation and Completion Trends Over Time: Simulation of 1,000 Rides

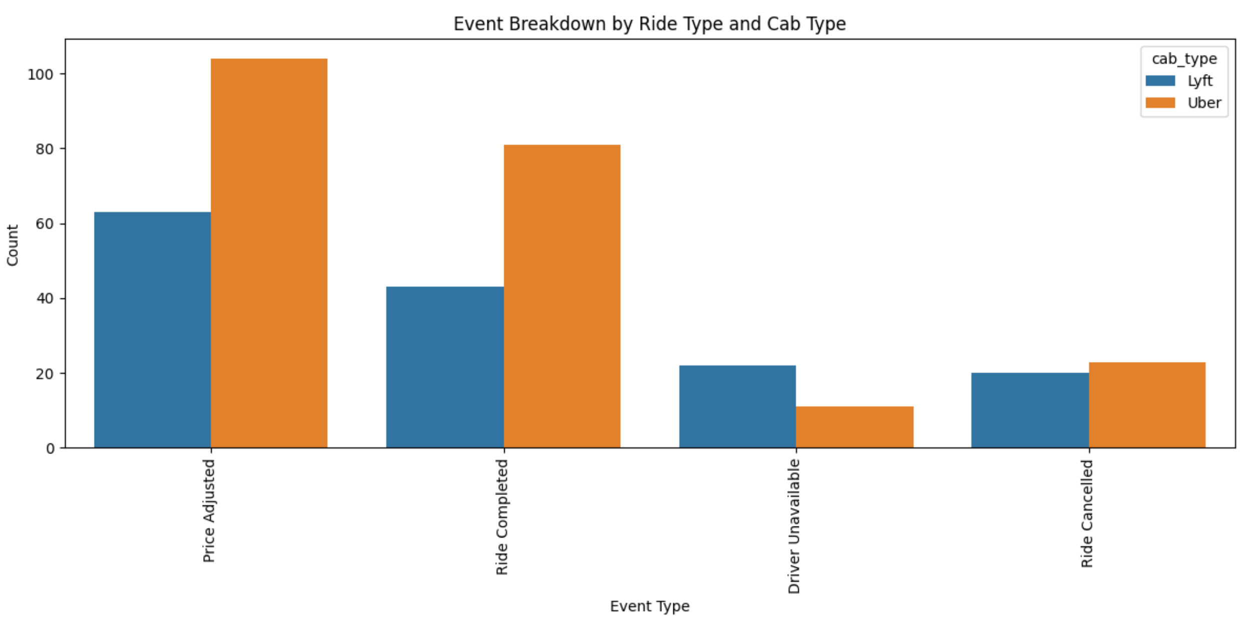

Figure 9: Event Breakdown by Ride and Cab Type: Simulation of 1,000 Rides

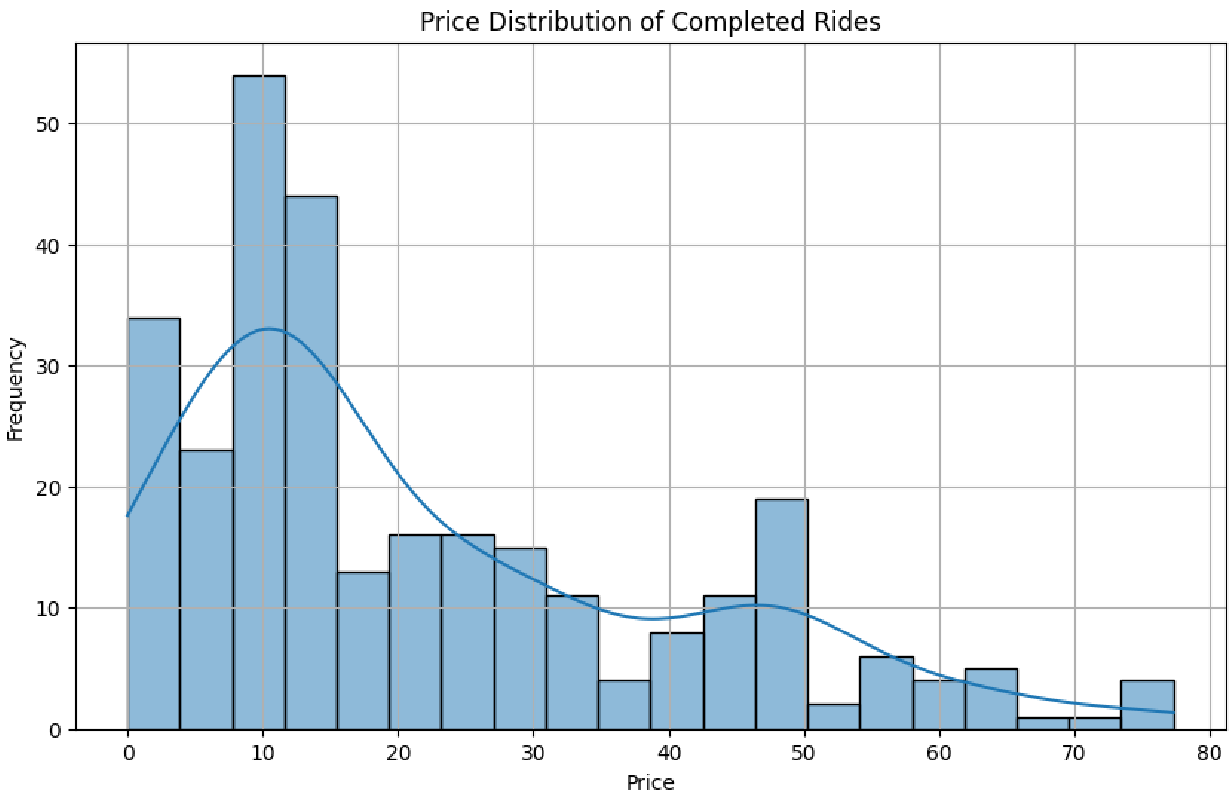

Figure 10: Price Distribution of Completed Rides: Simulation of 1,000 Rides

9.4 Simulation 100,000

Output:

```
Overall Cancellation Rate: 8.85%

Average Price by Ride Type and Cab Type:
cab_type  product
Lyft      Economy                 4.28
          Premium                30.37
          Premium Extra Room     51.95
          Standard                9.48
          Standard Extra Room    14.60
Uber      Economy                 3.80
          Premium                41.74
          Premium Extra Room     56.04
          Standard               12.88
          Standard Extra Room    20.47
Name: price, dtype: float64
21.075764075067028

Driver Unavailability Rate: 7.24%

Completion Rate by Cab Type:
cab_type
Lyft    0.350318
Uber    0.393519
dtype: float64

Average Ride Duration for Completed Rides: 4.50 minutes
```

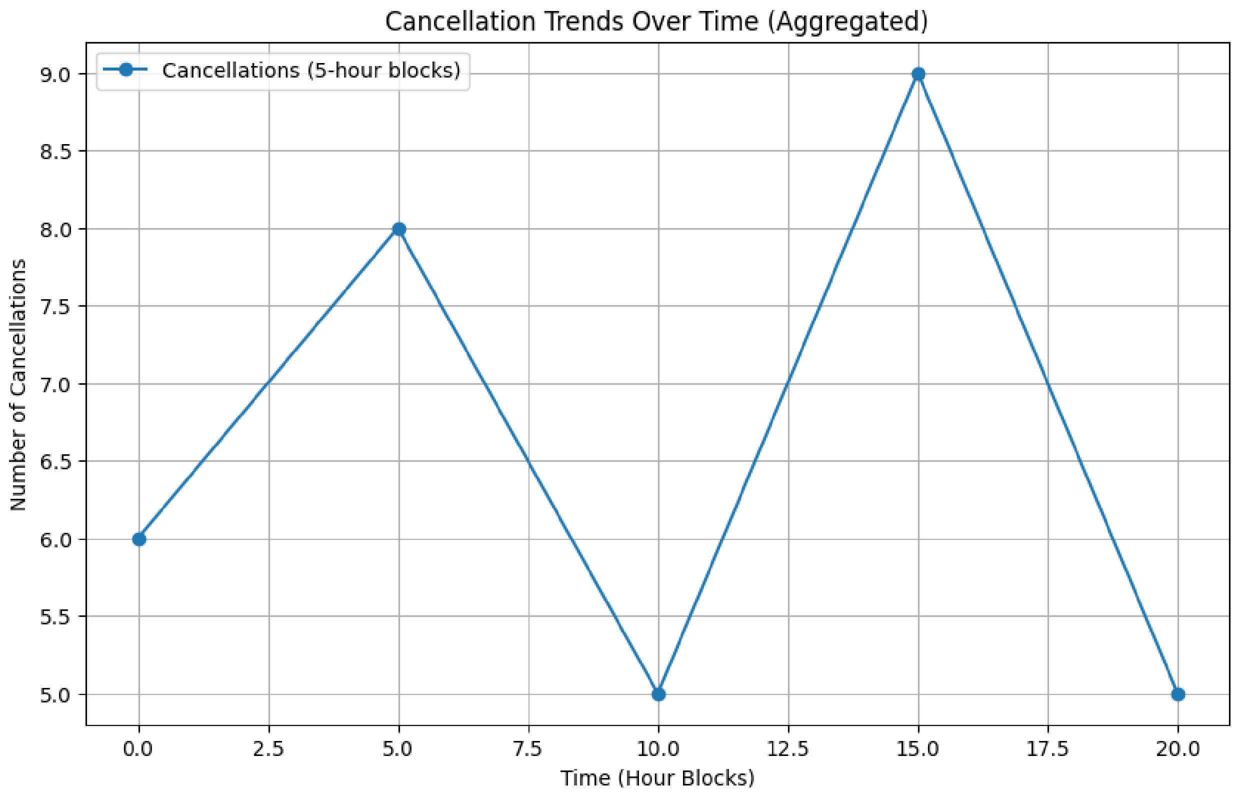

Figure 11: Aggregated Cancellation Trends Over Time: Simulation of 100,000 Rides

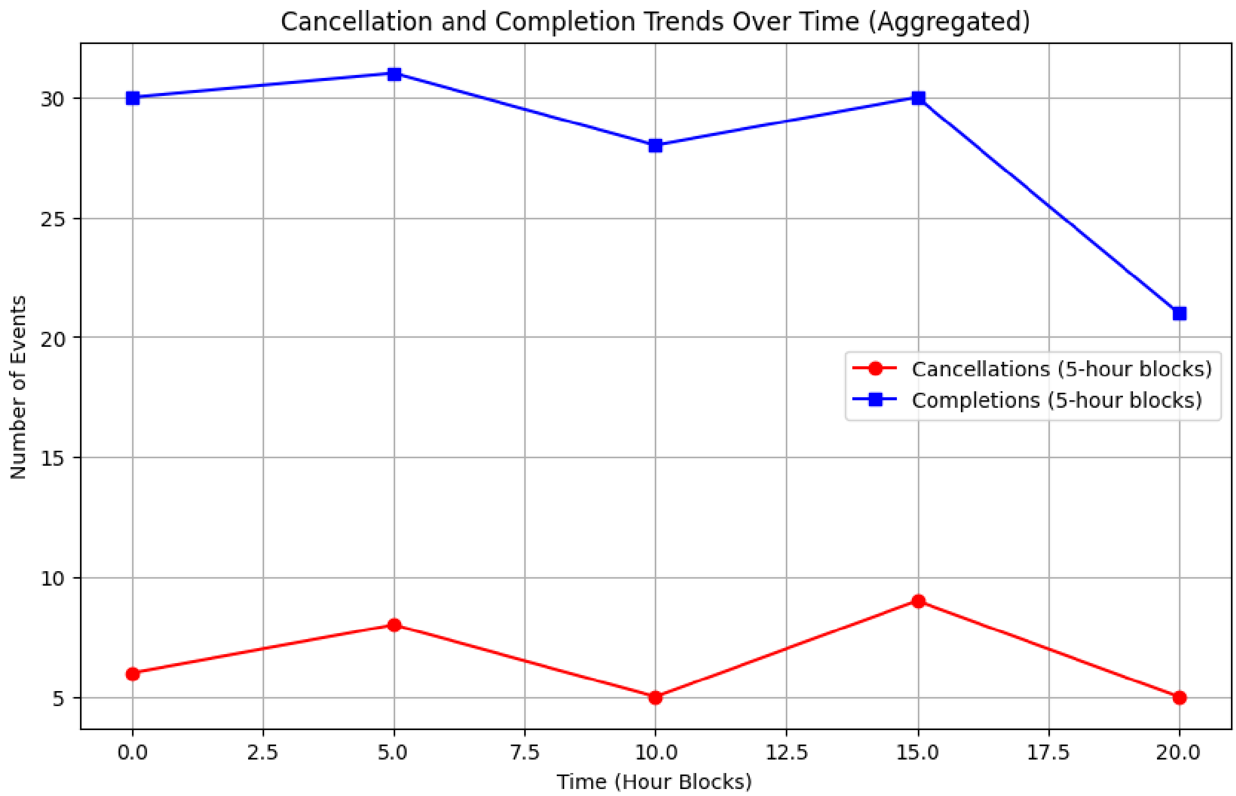

Figure 12: Aggregated Cancellation and Completion Trends Over Time: Simulation of 100,000 Rides

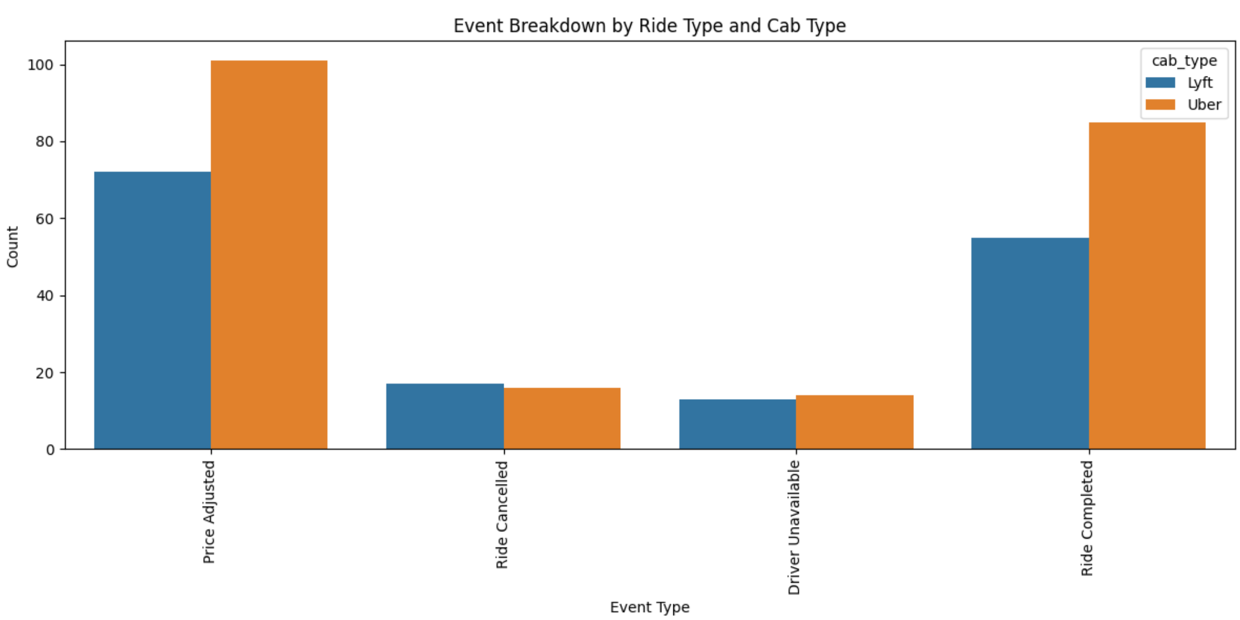

Figure 13: Event Breakdown by Ride and Cab Type: Simulation of 100,000 Rides

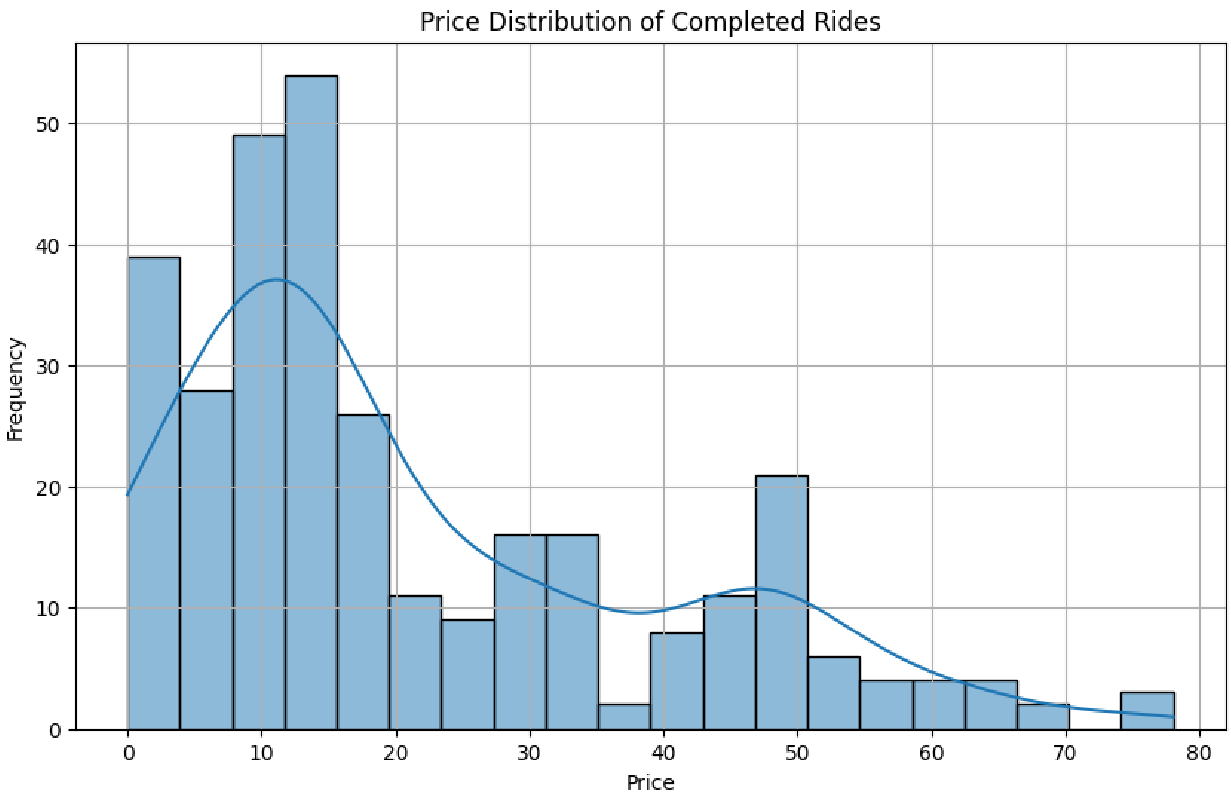

Figure 14: Price Distribution of Completed Rides: Simulation of 100,000 Rides

9.5 Simulation 400,000

Output:

```
Overall Cancellation Rate: 11.48%

Average Price by Ride Type and Cab Type:
cab_type  product
Lyft      Economy                 4.44
          Premium                30.14
          Premium Extra Room     51.72
          Standard                9.71
          Standard Extra Room    14.49
Uber      Economy                 3.89
          Premium                41.18
          Premium Extra Room     56.19
          Standard               12.35
          Standard Extra Room    20.55
Name: price, dtype: float64
20.740382513029366

Driver Unavailability Rate: 8.38%

Completion Rate by Cab Type:
cab_type
Lyft    0.358447
Uber    0.399276
dtype: float64

Average Ride Duration for Completed Rides: 4.60 minutes
```

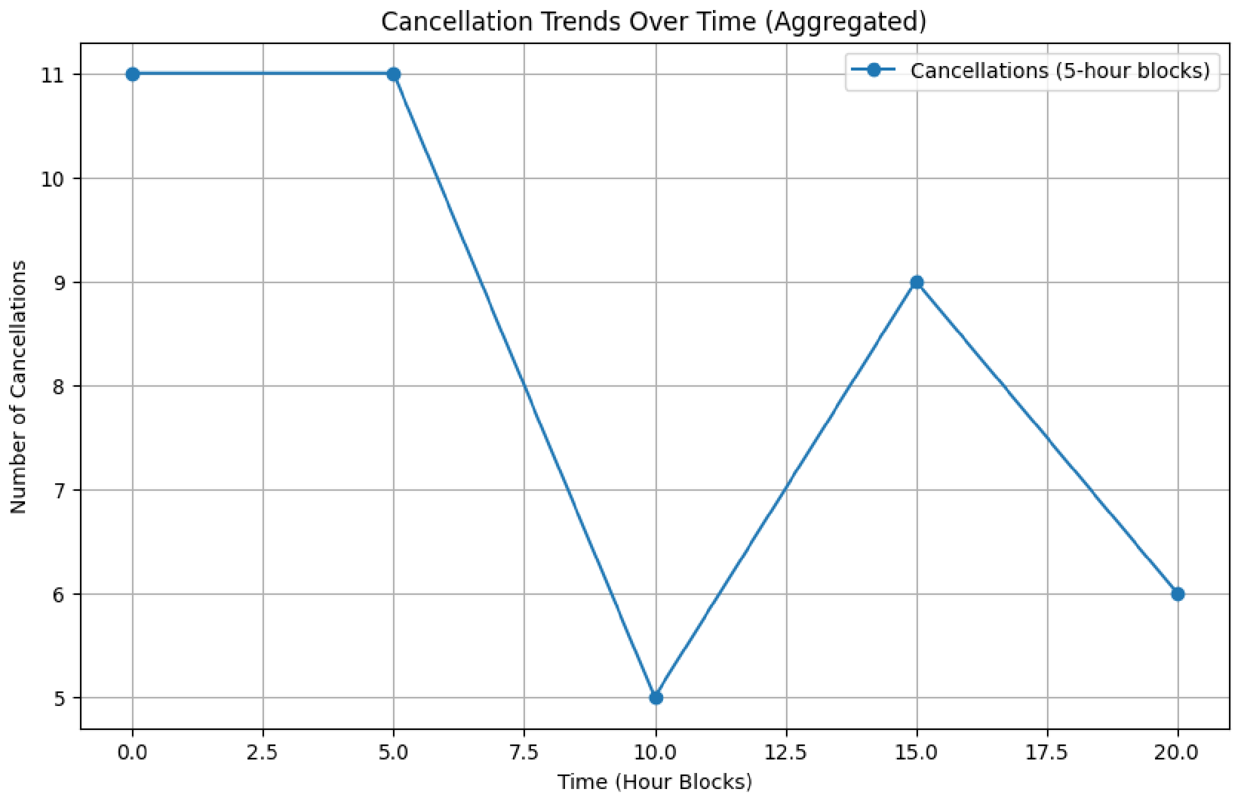

Figure 15: Aggregated Cancellation Trends Over Time: Simulation of 400,000 Rides

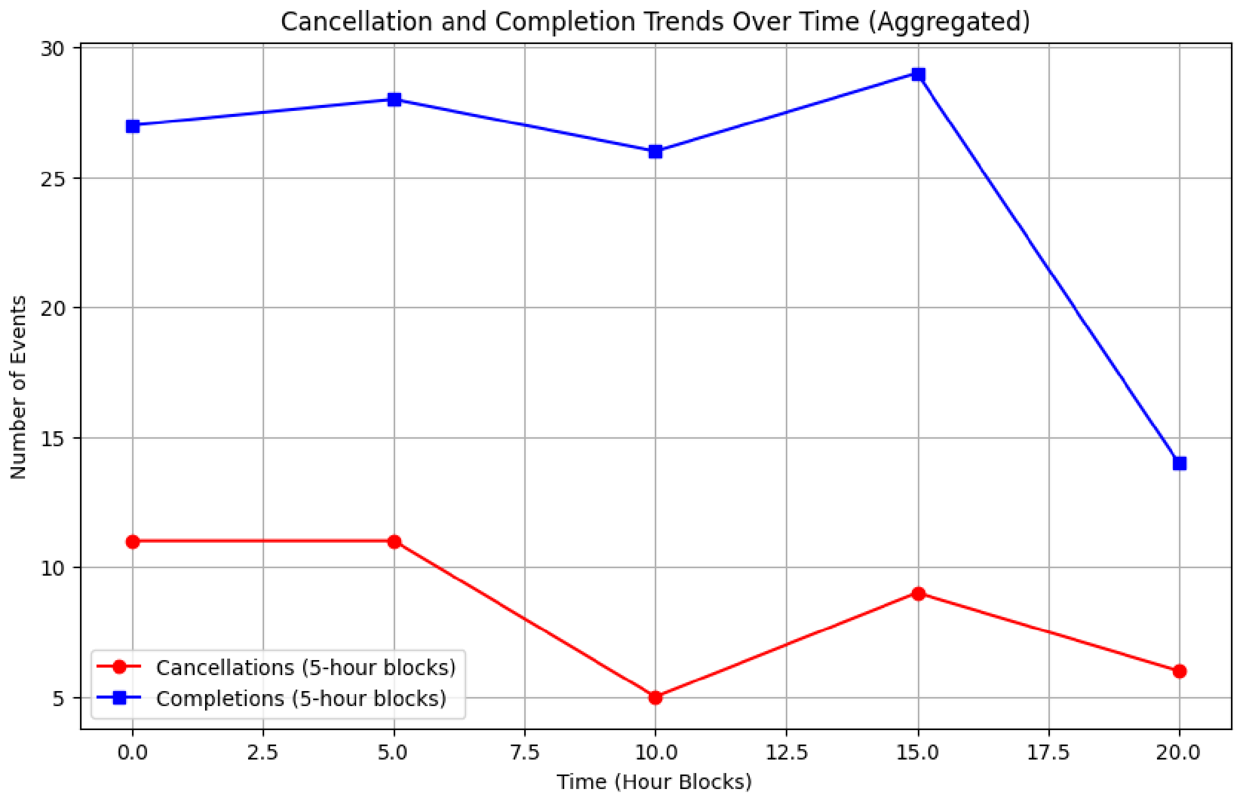

Figure 16: Aggregated Cancellation and Completion Trends Over Time: Simulation of 400,000 Rides

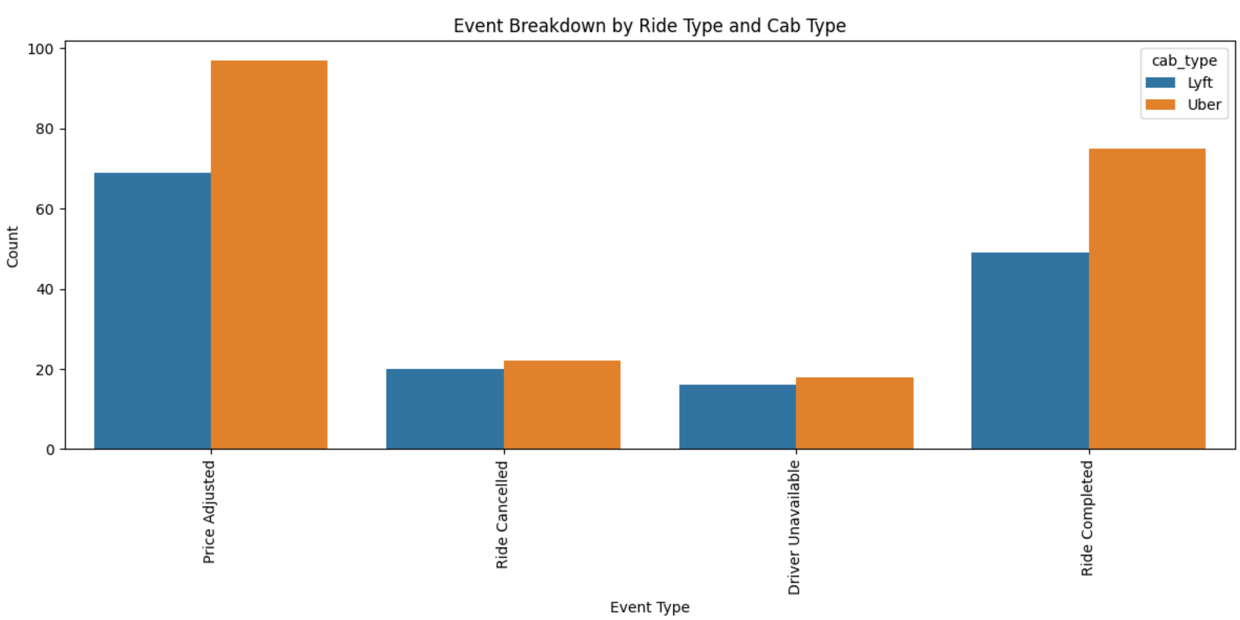

Figure 17: Event Breakdown by Ride and Cab Type: Simulation of 400,000 Rides

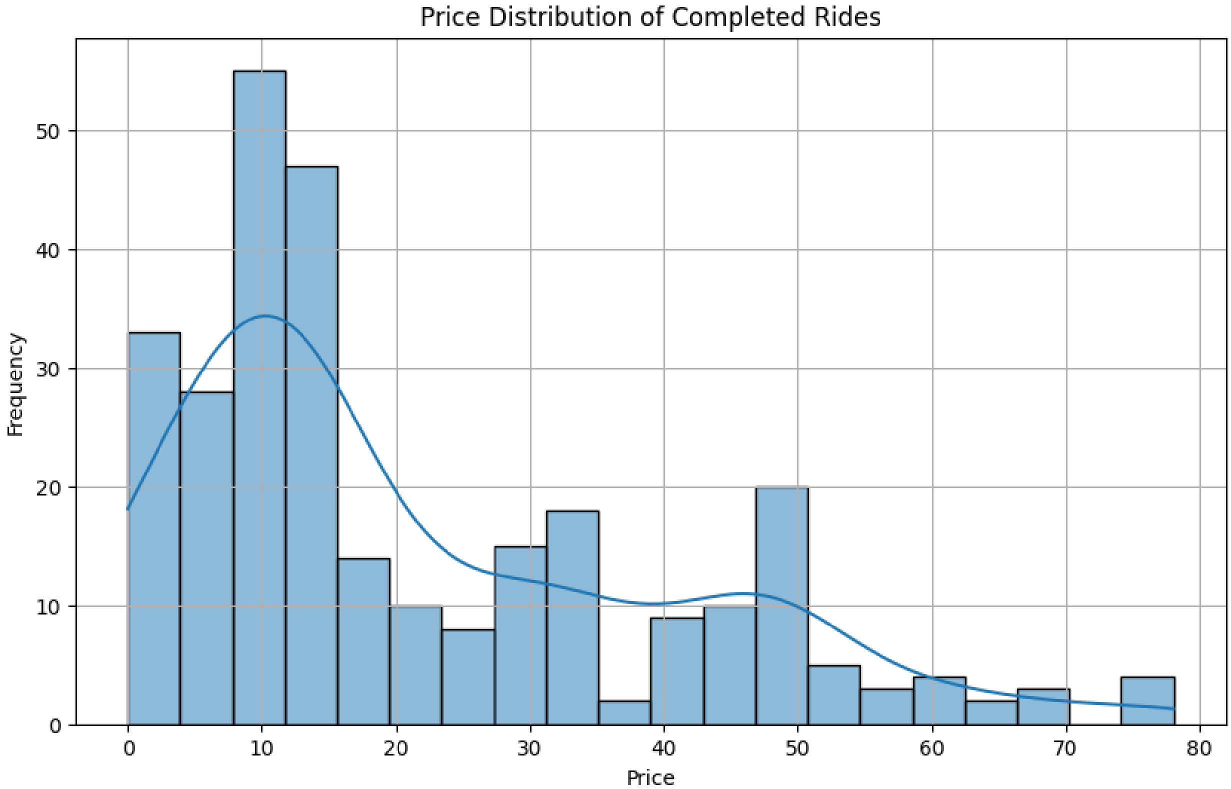

Figure 18: Price Distribution of Completed Rides: Simulation of 400,000 Rides

9.6 Simulation 600,000

Output:

```
Overall Cancellation Rate: 8.77%

Average Price by Ride Type and Cab Type:
cab_type  product
Lyft      Economy                 4.33
          Premium                31.01
          Premium Extra Room     49.30
          Standard                9.76
          Standard Extra Room    14.26
Uber      Economy                 3.71
          Premium                41.58
          Premium Extra Room     56.33
          Standard               12.17
          Standard Extra Room    20.35
Name: price, dtype: float64
20.82772602739726

Driver Unavailability Rate: 9.59%

Completion Rate by Cab Type:
cab_type
Lyft    0.340000
Uber    0.381395
dtype: float64

Average Ride Duration for Completed Rides: 4.45 minutes
```

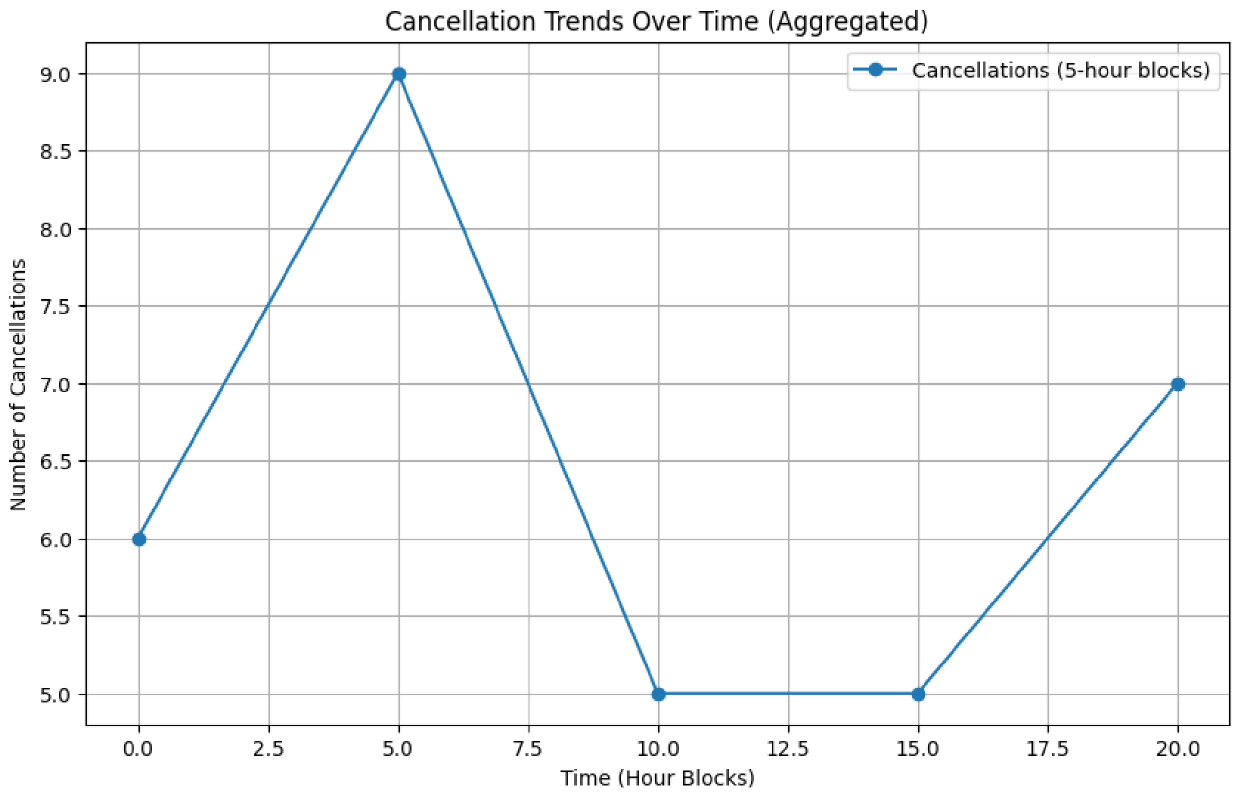

Figure 19: Aggregated Cancellation Trends Over Time: Simulation of 600,000 Rides

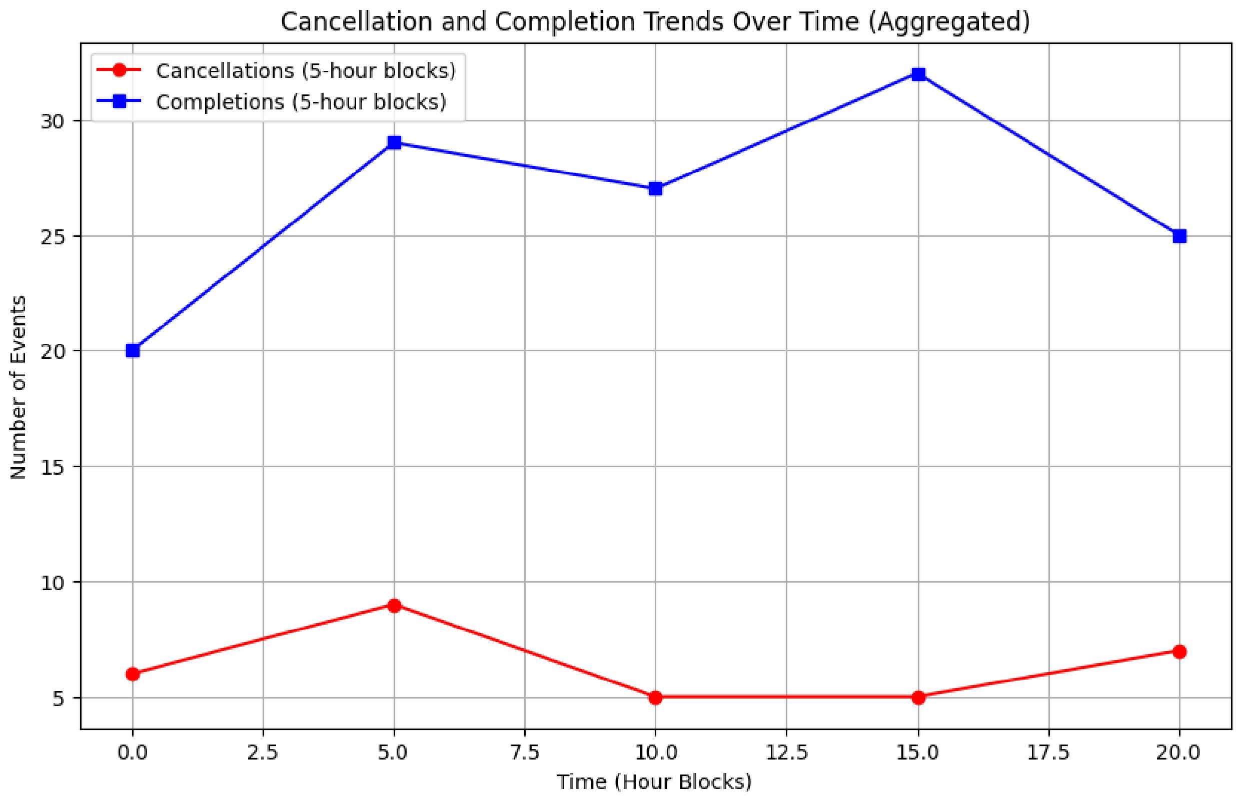

Figure 20: Aggregated Cancellation and Completion Trends Over Time: Simulation of 600,000 Rides

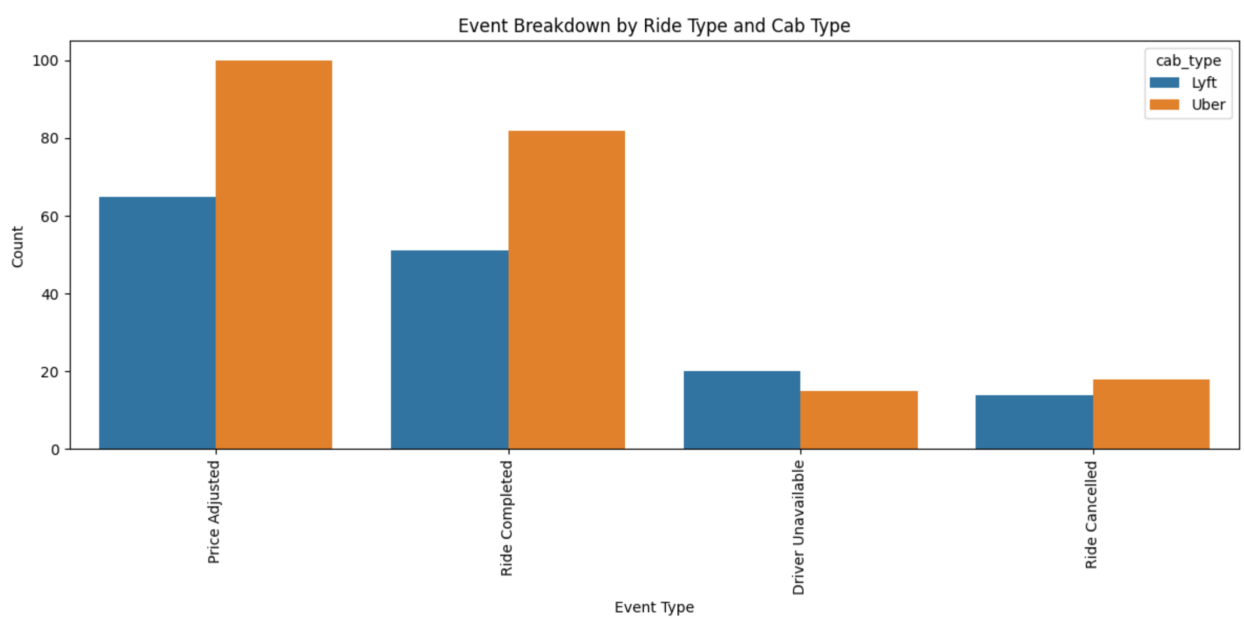

Figure 21: Event Breakdown by Ride and Cab Type: Simulation of 600,000 Rides

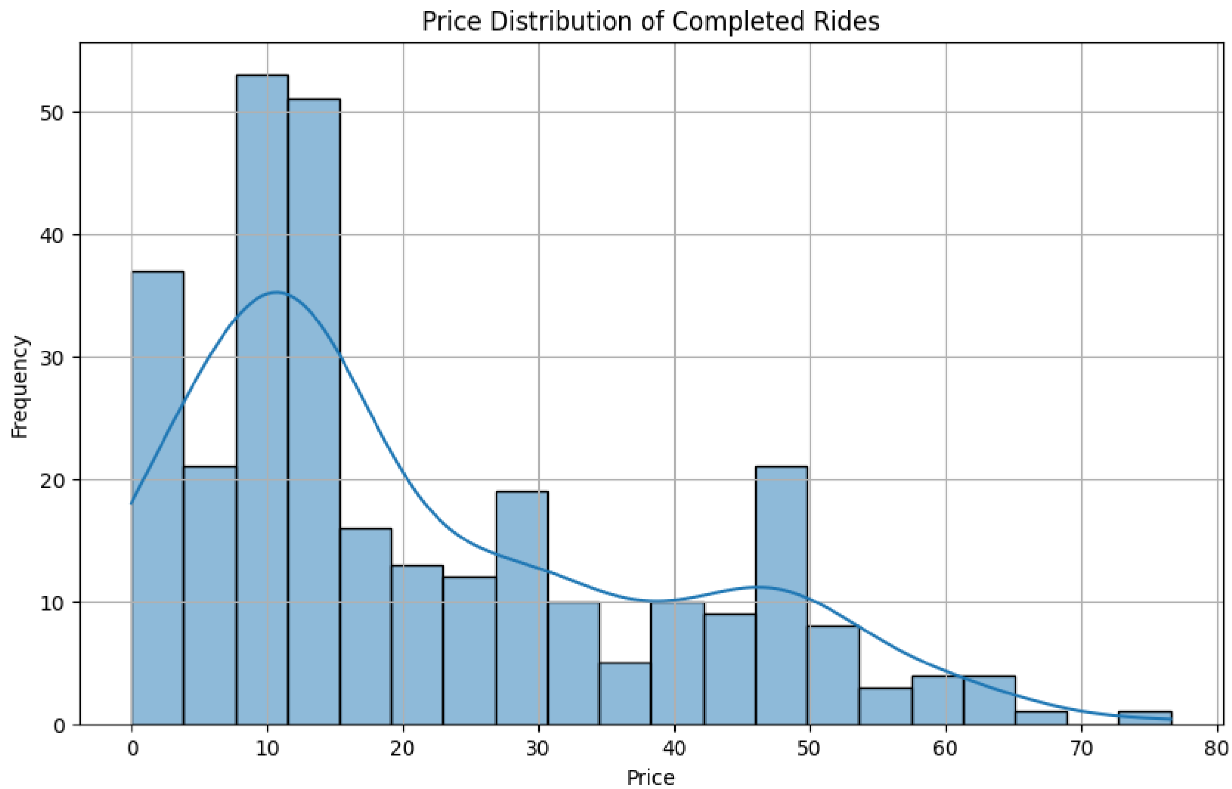

Figure 22: Price Distribution of Completed Rides: Simulation of 600,000 Rides

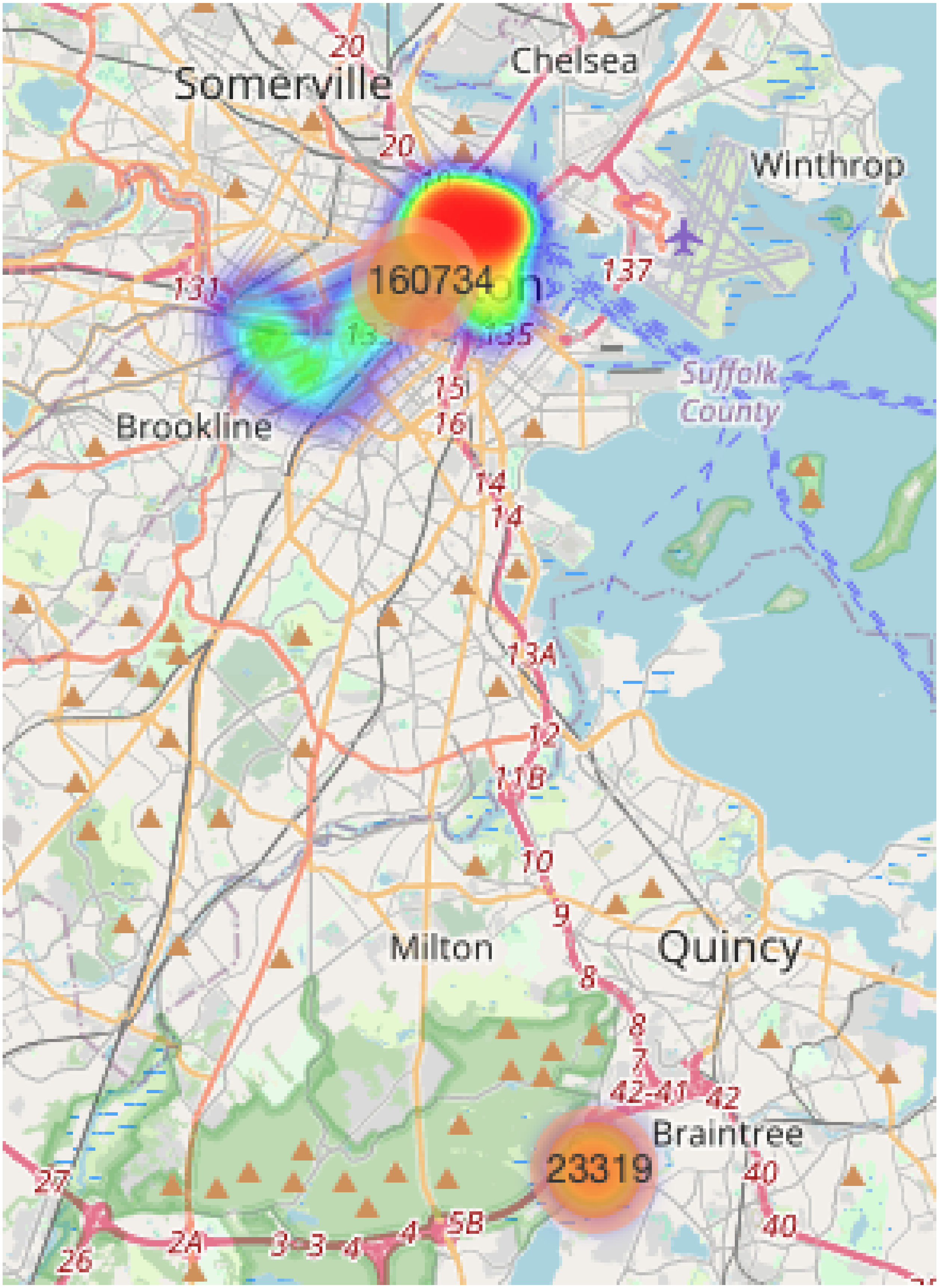

Figure 23: Folium Map of 100,000 Rides

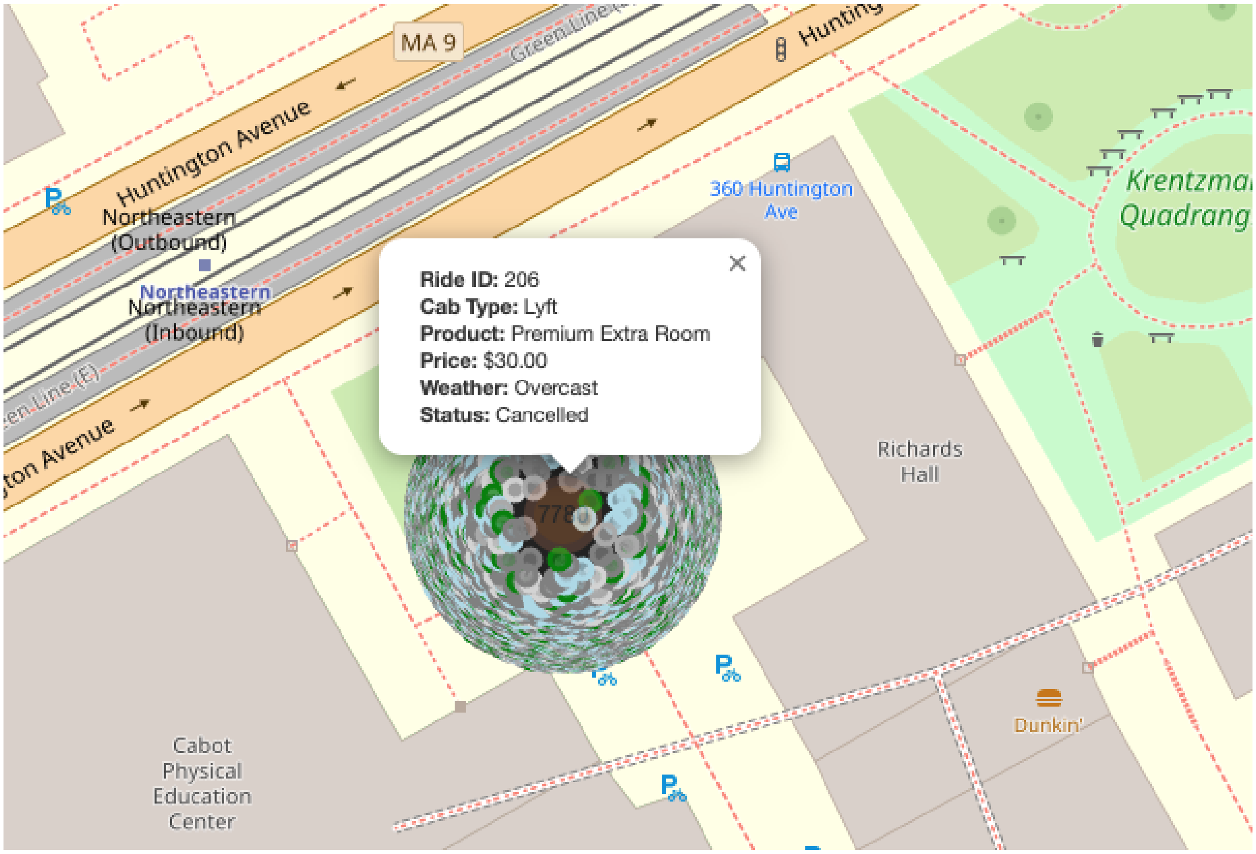

Figure 24: Northeastern Folium Map of 100,000 Rides