CORNELL TECH

Modeling Under Uncertainty

Instructor: Dr. Omar El Housni

Final Project

March 3, 2025

Team Members:

Dennis Chen

Rishab Jain

Jet Semrick

Date: December 3, 2024

Contents

1 Introduction

This project aims to model and estimate the steady-state availability of bikes at CitiBike stations in New York City using Discrete Markov Chains. CitiBike is a vital component of NYC's transportation system, serving a wide range of users from commuters to recreational riders. Understanding bike availability patterns at key stations can help improve operational efficiency and improve user satisfaction. This study focuses on three distinct stations to capture varying neighborhood characteristics and usage patterns.

The W 21 St & 6 Ave station in Chelsea reflects a mix of residential, commercial, and tourist activity, showcasing balanced usage and high daily turnover. The West St & Chambers St station in the Financial District represents a professional area with significant weekday commuter inflows and outflows. Meanwhile, the 1 Ave & E 68 St station in the Upper East Side highlights a more residential area, contrasting commuter and recreational usage patterns. These stations were chosen to represent a diverse range of CitiBike usage scenarios.

To conduct the analysis, the day is split into morning and evening blocks, further discretized into 5- or 10-minute intervals. By examining weekday bike inflows and outflows, each station's capacity and state dynamics are identified. States range from 0 (no bikes available) to the maximum dock capacity. Transition probability matrices are calculated for the morning and evening blocks based on a month of data, capturing the probabilities of moving between different bike availability states.

Finally, stationary distributions are computed for each station to determine the long-term probabilities of bike availability in the morning and evening. By focusing on high-turnover stations, this study provides valuable insights into how bike availability stabilizes over time. The findings can support strategies for better demand management during peak hours, operational planning, and enhanced user experiences for CitiBike riders.

2 Methodology

This study will estimate the steady-state distribution of available bikes at selected CitiBike stations using discrete-time Markov chains. We will discuss collecting and preprocessing data. We will model the system as a Markov chain, estimate transition probability matrices, and finally analyze the stationary distributions

2.1 Data Collection and Preprocessing

Data for this analysis were sourced from publicly available CitiBike trip records for the month of July 2024. Three popular stations in New York City were selected based on their high volume of bike activity and diverse location contexts:

- **Station A:** W 21 St & 6 Ave - This station is located in Chelsea, surrounded by buildings, cafes, and restaurants, and is also close to offices. It is likely busiest during weekday mornings (7–10 AM) and evenings (4–7 PM) due to commuter traffic, often resulting in significant drops in bike availability. Midday usage is moderate as locals and tourists explore Chelsea, while late evenings and early mornings see minimal activity. On weekends, the station might attract more recreational riders visiting nearby attractions like the High Line and Chelsea Market, with weather playing a key role in influencing usage patterns.

- **Station B:** West St & Chambers St - The West St and Chambers St station is located near the Hudson River Greenway and close to Battery Park City, making it a prime spot for both commuters and recreational riders. During weekday mornings (7–10 AM) and evenings (4–7 PM), the station is likely busy with office commuters heading to and from nearby financial and residential areas. Midday usage might increase due to tourists exploring attractions like the World Trade Center or the waterfront. On weekends, the station could see heavy recreational use from riders enjoying the Hudson River Greenway or nearby parks, with pleasant weather likely boosting activity.

- **Station C:** 1 Ave & E 68 St - The 1 Ave and E 68th St station is located on the Upper East Side, a primarily residential neighborhood with nearby hospitals, schools, and parks. It is likely busiest during weekday mornings (7–10 AM) as residents commute to work and schools and during afternoons (3–6 PM) when schools let out and locals run errands. Midday usage may be moderate with hospital visitors and nearby residents using bikes for short trips. Weekend activity might be quieter but could increase with recreational riders heading to nearby attractions like Central Park or the East River Greenway.

To study the variations in bike availability, weekdays were divided into two time blocks: **Morning (6:00 AM to 12:00 PM)** and **Evening (4:00 PM to 10:00 PM)**. Each block was further discretized into 10-minute intervals, yielding 36 time steps per block. Weekends were excluded to ensure consistent weekday usage patterns and to avoid confounding effects from recreational weekend activity.

Data preprocessing involved cleaning the trip records by removing anomalies such as trips with negative durations, durations exceeding three hours, and outliers in recorded trip times. Only trips starting or ending at the selected stations during the specified time blocks were retained for analysis. The station capacities, representing the maximum number of available docks, were determined as follows: **70 docks for W 21 St & 6 Ave**, **111 docks for West St & Chambers St**, and **35 docks for 1 Ave & E 68 St**. These values were obtained from official CitiBike records and supplemented with observational data. Additional images are provided below to demonstrate the station selection procedures.

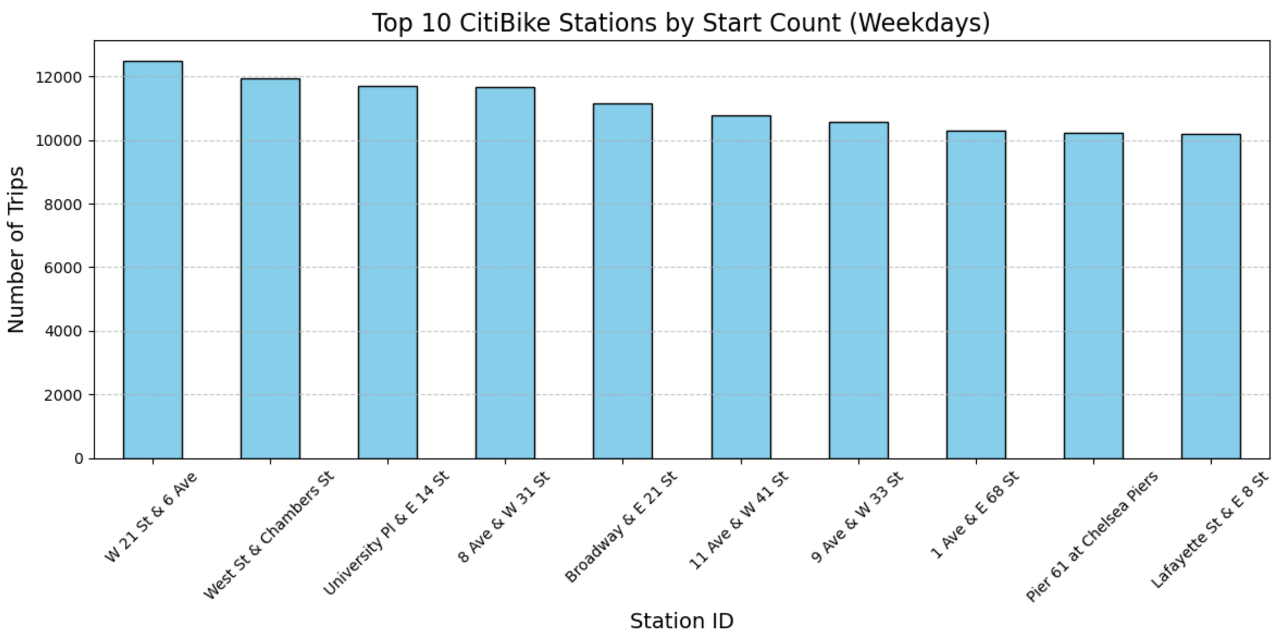

Figure 1: Top 10 stations by start count

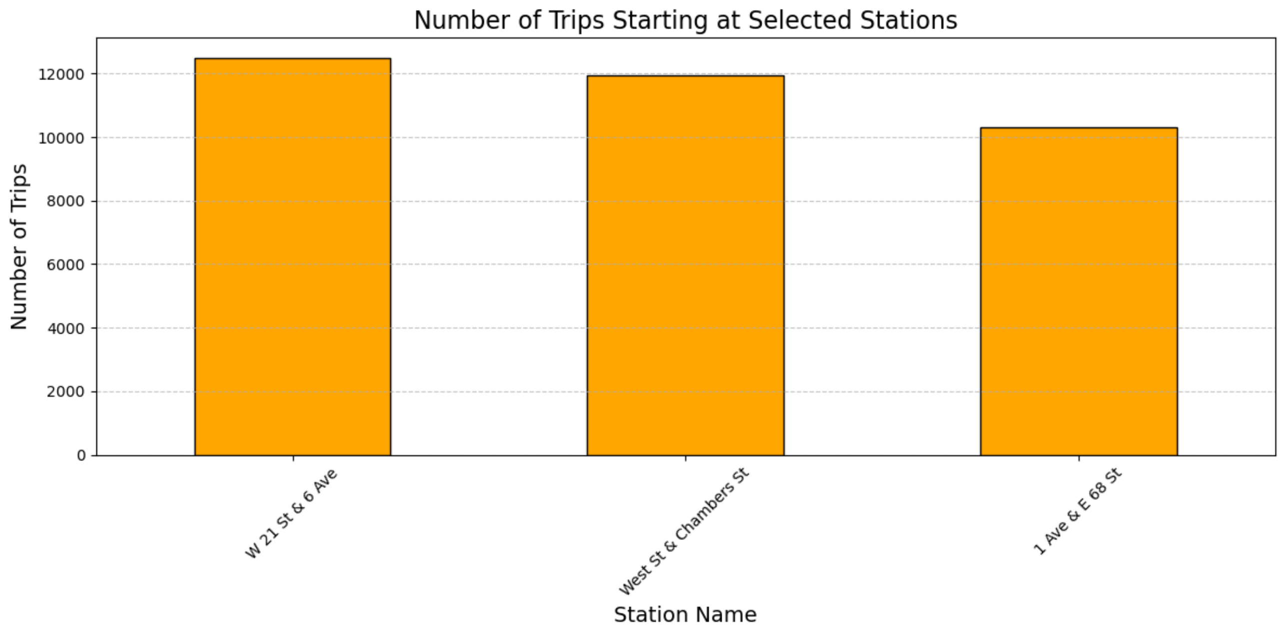

Figure 2: Selected 3 stations' activity levels

2.2 Modeling

2.2.1 Markov Chain Formulation

The availability of bikes at each CitiBike station was modeled as a discrete-time Markov chain, where the states represent the number of available bikes at any given time. The state space S for each station was defined as $S = 0, 1, 2, \ldots, C$, where C represents the station's maximum dock capacity. $X_t isthenumberofavailablebikesattimet.This$ $[P(X_{t+1} = j | X_t = i, X_{t-1}, \ldots, X_0) = P(X_{t+1} = j | X_t = i).]$

2.2.2 Transition Probability Matrices

For each station and time block (morning and evening), a transition probability matrix P was calculated, where each element P_{ij} denotes the probability of transitioning from state i to state j in one time step. The transition probabilities were estimated using relative frequencies: [$P_{ij} = \frac{N_{ij}}{N_i}$,]

where N_{ij} is the number of observed transitions from state i to j, and N_i is the total number of transitions starting from state i.

2.2.3 Accounting for Latent Demand

To ensure realistic modeling, we addressed latent demand, which arises when stations are either full or empty:

Full Stations (State C): When a station is at full capacity, additional bike arrivals cannot be docked and thus go unrecorded. This unmet demand was estimated as:

[latentdemandfull = $\max(0, netflow - capacity)$]

where net flow represents the difference between arrivals and departures during a time interval.

Empty Stations (State 0): Similarly, when a station is empty, additional bike departures cannot occur. The unmet demand in this scenario was calculated as:

[latentdemandempty = $\max(0, -netflow)$.]

2.2.4 Stationary Distributions

Finally, the stationary distributions, π, representing the long-term probabilities of the system being in each state, were computed. These distributions satisfy the equation: [$\pi = \pi P$,]

with the normalization condition:

$\sum_{i \in S} \pi_i = 1$.]

3 Results

Our results will include stationary distributions for our selected stations, using bar charts for clarity.

The X-axis represents the station states, indicating the number of available bikes, ranging from 0 to C. A state of 0 signifies that the station is completely empty, with no bikes available, while C corresponds to full capacity, where all docks are occupied, leaving no empty spaces.

The Y-axis illustrates the stationary probabilities of the system being in each state, displayed on a logarithmic scale. This logarithmic representation is used to effectively capture and present probabilities across a wide range, from extremely small to relatively larger values.

3.1 W 21 St 6 Ave

The graphs below show the stationary distributions for the morning and evening for the W 21 St 6 Ave station.

3.1.1 Stationary Distribution

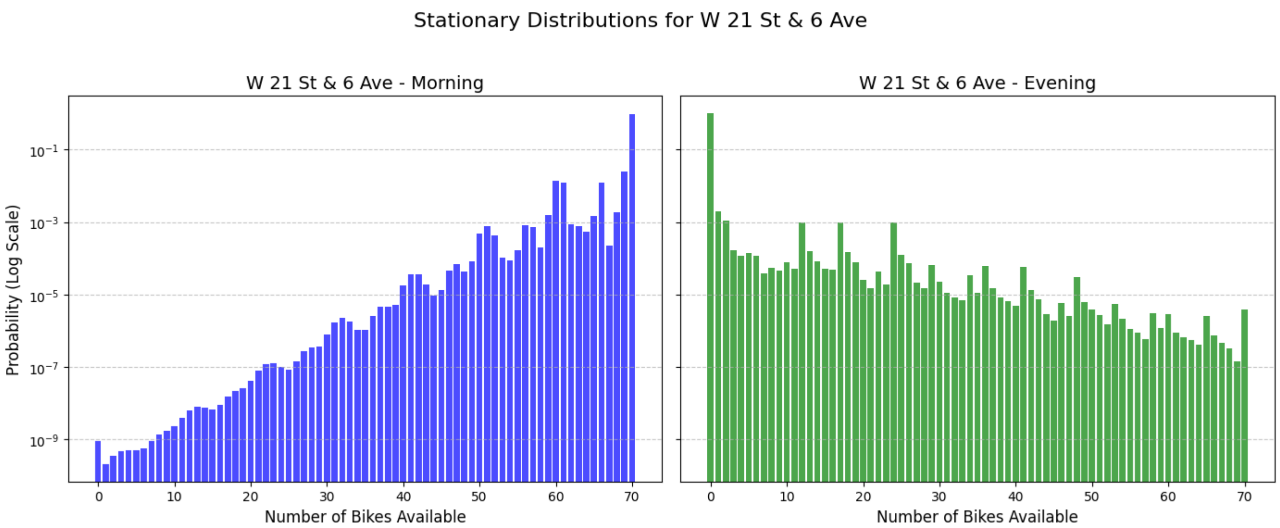

Figure 3: Stationary distributions for the number of bikes available at W 21 St & 6 Ave during the morning and evening hours.

3.1.2 Analysis

This station is generally full in the morning and empty in the evening indicated by peaks at state 70 for the morning and state 0 for the evening. The results are consistent with previous analysis that bikers will commute to work in the morning and remove bikes to return home in the evening. In comparison to the evening distribution, the morning distribution has a steeper slope indicating riders arrive rapidly in the morning and gradually depart in the afternoon.

3.2 West St Chambers St

The graphs below show the stationary distributions for the morning and evening for West St Chambers St.

3.2.1 Stationary Distribution

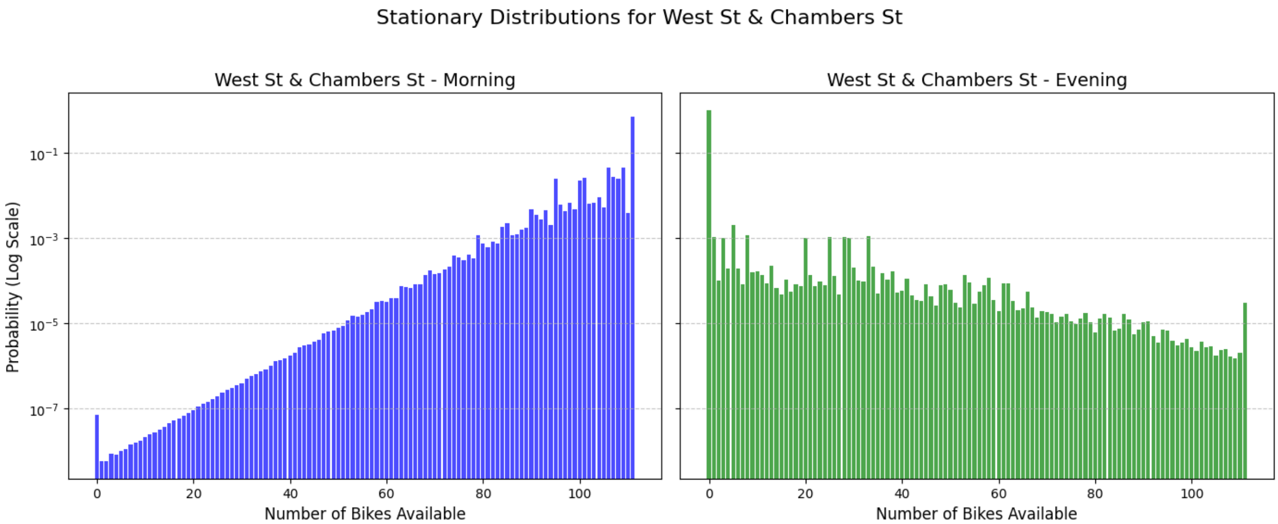

Figure 4: Stationary distributions for the number of bikes available at West St Chambers St during the morning and evening hours.

3.2.2 Analysis

Similar to the W 21 St 6 Ave station, this station is generally full in the morning and empty in the evening. The results are consistent with previous analysis that this station primarily services commuters and tourists. The slopes of these graphs are also similar to the previous station. However, this station is more likely to service tourists as it is closer to Battery Park and the World Trade Center.

3.3 1 Ave E 68 St

The graphs below show the stationary distributions for the morning and evening for 1 Ave E 68 St.

3.3.1 Stationary Distribution

3.3.2 Analysis

In this station, both graphs have a more consistent range of probabilities with peaks at full in the morning and empty in the evening. Unlike the past two stations, the slope of the graphs are far less steep which indicates the rate bikes leave or enter the station is gradual over each period of time analyzed.

Located in the Upper East Side, this station is likely less busy and faces less peak pressure since it is located in a primarily residential area. Since there are also hospitals and schools we still see the similar behaviors of the other stations just at a less pronounced level.

4 Conclusion

After analyzing three distinct CitiBike stations in New York City, we have shown there are consistent usage patterns across stations. The primary theme regardless of location is that stations in commercial areas are

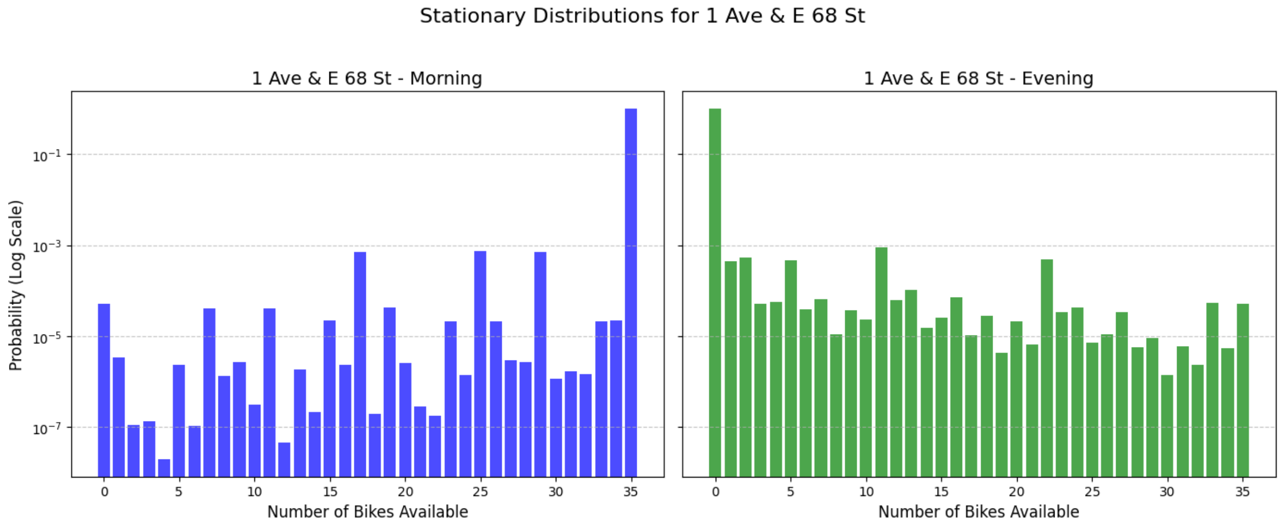

Figure 5: Stationary distributions for the number of bikes available at 1 Ave E 68 St during the morning and evening hours.

more likely to be full of bikes in the morning and empty in the evening as commuters return to their place of residence.

In busier commercial areas such as W 21 St 6 Ave and West St Chambers St, there are steep peaks in the morning and evening indicating these stations spend a significant amount of time in extreme states. In contrast, the calmer station at 1 Ave E 68 St has a more consistent stationary distribution and is more likely to not be full or empty at any given time.

The general insight that is revealed through analysis is that in the evening bikes should be redistributed to avoid shortages in the morning when commuters need to ride to work. This ensures there is never a disastrous imbalance of bikes that prevents the system from functioning properly.

5 Appendix

5.1 Warm Up Questions

5.1.1 Ride Duration and Histogram

Understanding the distribution of ride durations is key to identifying typical usage patterns.

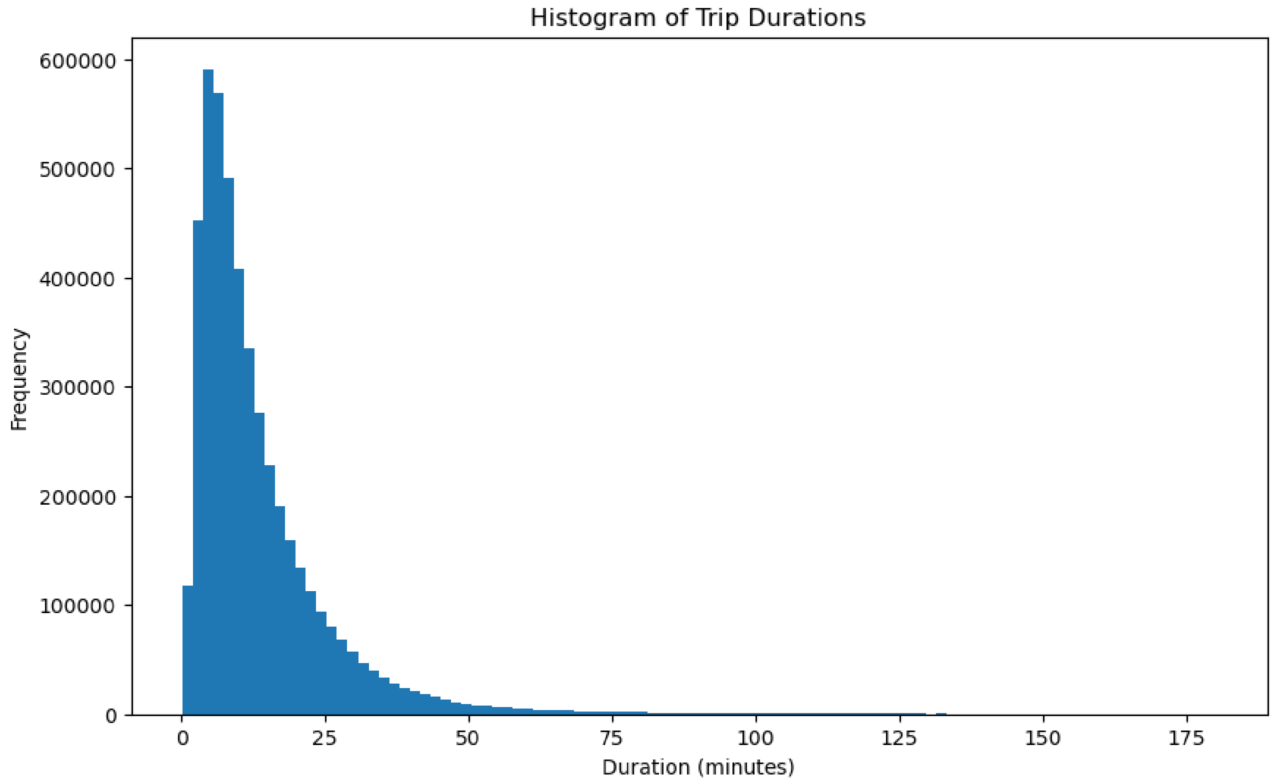

Figure 6: Histrogram Warm Up 1

5.1.2 Expected Ride Duration and Variance

Calculating the average ride duration and variance provides insights into the central tendency. Based on our analysis, the expected ride duration is 13.42 minutes, with an empirical variance of 172.15. The probability of a ride greater than 20 minutes is 0.19, suggesting most rides are near the average.

5.1.3 Conditional Probability for CitiBike Members

Examining the likelihood of longer rides conditioned on membership type allows us to understand how user preferences and membership status influence ride duration patterns. The probability of a ride duration greater than 20 minutes given the user is a CitiBike member is 15 percent.

5.1.4 Reverse Conditioning for CitiBike Members

By evaluating the probability that a longer ride belongs to a CitiBike member, we gain insights into how membership demographics correlate with ride duration, helping to determine whether casual clients or members are more likely to take longer trips. The probability that the ride belongs to a CitiBike member given the ride duration is greater than 25 minutes is 57 percent.

5.1.5 Expected Duration by Bike Type

Comparing the expected ride durations for electric bikes and classic bikes highlights how the choice of bike type affects trip behavior. This analysis can reveal whether users prefer electric bikes for longer commutes or if classic bikes are favored for shorter, leisurely rides. The expected ride duration for electric bikes is approximately 13.78 minutes, while for classic bikes, it is slightly lower at around 12.74 minutes.

5.1.6 Short Ride and Bike Type Probability

Analyzing the probability of short rides being taken on electric versus classic bikes offers insights into user preferences for quick trips. This can also help identify patterns in how the bike-sharing service is utilized for short-distance travel. The results indicate that for rides with a duration of less than 10 minutes, there is a 63 percent probability that the bike is electric and 37 percent chance that the bike is classic. There could be a few reasons for the large difference in probabilities.

1) Electric bikes may be more popular for shorter rides. The elctric bikes might be more for the intent of commuting to a place quickly and efficiently. The extra speed would allow users to reach a destination faster who have limited time. On the contrary, the classic bike may be used for leisure purposes. Thus, riders may tend to take the classic bike to enjoy a longer ride.

2) Since the electric bikes are faster, riders may just be getting to destinations at a faster pace then classic bikers.

3) The pricing of the bikes may play a role as well. The electric bikes are more expensive, incentivizing users to use them for shorter trips.

5.1.7 Evening Transition Matrix for Each Station

```
Evening Transition Matrix for W 21 St & 6 Ave:
[[9.94418605e-01 1.86046512e-03 9.30232558e-04 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [8.33333333e-01 0.00000000e+00 5.55555556e-02 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [8.33333333e-01 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 ...
 [8.33333333e-02 0.00000000e+00 2.77777778e-02 ... 0.00000000e+00
  5.55555556e-02 1.11111111e-01]
 [8.33333333e-02 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 1.66666667e-01]
 [8.33333333e-02 0.00000000e+00 0.00000000e+00 ... 2.77777778e-02
  0.00000000e+00 1.66666667e-01]]
```

Figure 7: Transition Matrix 1

```
Evening Transition Matrix for West St & Chambers St:
[[9.90723562e-01 9.27643785e-04 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [7.22222222e-01 0.00000000e+00 2.77777778e-02 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [7.22222222e-01 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 ...
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  2.77777778e-02 2.50000000e-01]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 2.77777778e-01]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 2.77777778e-01]]
```

Figure 8: Transition Matrix 2

```
Evening Transition Matrix for 1 Ave & E 68 St:
[[9.97513469e-01 4.14421881e-04 4.14421881e-04 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [8.05555556e-01 2.77777778e-02 2.77777778e-02 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [8.05555556e-01 0.00000000e+00 2.77777778e-02 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 ...
 [5.00000000e-01 0.00000000e+00 0.00000000e+00 ... 2.77777778e-02
  2.77777778e-02 1.38888889e-01]
 [5.00000000e-01 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  2.77777778e-02 1.66666667e-01]
 [5.00000000e-01 0.00000000e+00 0.00000000e+00 ... 2.77777778e-02
  0.00000000e+00 1.94444444e-01]]
```

Figure 9: Transition Matrix 3

5.1.8 Morning Transition Matrix for Each Station

```
Morning Transition Matrix for W 21 St & 6 Ave:
[[0.44642857 0.         0.01785714 ... 0.         0.         0.08928571]
 [0.13888889 0.         0.         ... 0.02777778 0.         0.13888889]
 [0.08333333 0.05555556 0.         ... 0.         0.02777778 0.13888889]
 ...
 [0.         0.         0.         ... 0.         0.         0.86111111]
 [0.         0.         0.         ... 0.05555556 0.         0.86111111]
 [0.         0.         0.         ... 0.         0.02666667 0.93333333]]

Morning Transition Matrix for West St & Chambers St:
[[0.70588235 0.01176471 0.         ... 0.         0.         0.        ]
 [0.25       0.05555556 0.02777778 ... 0.         0.         0.        ]
 [0.25       0.         0.05555556 ... 0.         0.         0.        ]
 ...
 [0.         0.         0.         ... 0.05555556 0.02777778 0.66666667]
 [0.         0.         0.         ... 0.         0.05555556 0.69444444]
 [0.         0.         0.         ... 0.05555556 0.         0.75      ]]

Morning Transition Matrix for 1 Ave & E 68 St:
[[0.50746269 0.         0.         ... 0.         0.         0.40298507]
 [0.08333333 0.         0.         ... 0.02777778 0.         0.75      ]
 [0.08333333 0.         0.         ... 0.         0.02777778 0.75      ]
 ...
 [0.         0.         0.         ... 0.         0.         0.91666667]
 [0.         0.         0.         ... 0.         0.         0.91666667]
 [0.         0.         0.         ... 0.         0.         0.9978602 ]]
```

Figure 10: Morning Transition Matrix

6 Code

The following is the code for the ORIE 5530 Final Project: