NYC CitiBike Project

Modeling Under Uncertainty

December 4, 2024

Alex Bean	Shreeya Indap	Carlijn Smeets
atb95	si223	cas572

1 Introduction

The availability of bikes at CitiBike stations is critical to the reliability and efficiency of New York City's bike-sharing system. Understanding the patterns of bike availability at specific stations can enhance user experience and optimize system operations. This project aims to estimate the steady-state distribution of available bikes at selected CitiBike stations in New York using Markov chains. In this study, we focus on three Manhattan CitiBike stations: W 21 St & 6 Ave, E 40th St & Park Ave, and West St & Chambers St. These stations are carefully chosen based on their distinct characteristics and commuter behaviors. Situated in diverse areas of Manhattan, these stations reflect unique usage patterns. By analyzing the stationary distributions of bike availability during the morning and evening hours, we aim to identify trends and insights into bike-sharing dynamics.

To estimate steady-state availability, we model the number of bikes at each station as a Discrete Time Markov Chain. The state space ranges from 0 (completely empty) to the station's maximum capacity (fully occupied). Using ride data from July 2024, we focus on weekdays to capture typical commuting patterns, dividing the morning and evening blocks into discrete intervals of 5 or 10 minutes. Transition probability matrices for each block are estimated based on bike arrivals and departures, enabling us to compute stationary distributions that represent the long-term probabilities of bike availability.

The structure of this report is organized as follows. First, Section 2 outlines our modeling approach, detailing the process of deriving the transition probability matrix and calculating the stationary distribution from the data. In Section 3, we discuss the findings, while Section 4 concludes the study with key insights and business implications.

2 Modeling Approach

In this section, we outline our modeling approach. Section 2.1 details the data preprocessing steps and the rationale behind selecting the three stations analyzed in this study. Subsequently, we provide a comprehensive explanation of our methodology for estimating the steady-state distribution of available bikes at CitiBike stations using Markov Chains.

2.1 Data Preprocessing and Station Analysis

In this study, we utilize a CitiBike dataset containing ride information for July 2024. The dataset comprises 4,722,895 rides and 14 features, including ride ID, bike type (classic or electric), start and end times, start and end station names and IDs, and additional details such as membership status. As the dataset was pre-cleaned before retrieval, we only make minimal adjustments to enhance its usability for our analysis.

First of all, we introduce a new feature, 'duration', calculated as the difference between the start and end times of each trip. Using this feature, we exclude all rides lasting more than three hours, as these trips are likely due to undocked bikes and could distort the stationary distribution analysis. Additionally, we remove entries with negative ride durations, which are likely the result of docking or undocking errors. To gain a preliminary understanding of the ride data, we solve the warm-up questions outlined in the project description.

The warm-up questions provide a foundational understanding of CitiBike ride patterns in July 2024. To narrow the focus of our study to three specific NYC CitiBike stations, we perform an additional analysis to identify suitable candidates. By analyzing trips starting and ending at each station, we create a top 25 list of stations with the highest activity, ensuring our final selections have ample data for analysis.

To analyze station activity, we track the number of available bikes at each station, decreasing this count for starting trips and increasing it for ending trips. As the CitiBike dataset does not include station capacity information, we reference the capacities listed on Google Maps for each station. To initialize the available bikes variable, we assume that every station begins at full capacity on July 1st. In Section 3.5, we conduct a sensitivity analysis to evaluate whether altering this initial assumption affects the stationary distribution of a station.

Lastly, as this study focuses on the steady-state distribution during weekdays, we filter the data to include only weekday rides and analyze the weekly patterns of various stations to identify those with distinct usage trends. Once the stations are finalized, we proceed to the steady-state estimation approach, detailed in Section 2.2.

2.2 Steady-State of CitiBikes using Markov Chains

To estimate the steady-state distribution of available bikes at a CitiBike station, we adopt a Discrete Time Markov Chain (DTMC)-based approach. Let us first define some notation. Consider N stations under analysis, where each station $i \in N$ has a capacity of c_i bikes. According to the Markov property, the number of available bikes at station i at time period $t+1$, denoted as $a_{i,t+1}$, depends solely on the number of available bikes during the previous time period, $a_{i,t}$, and is independent of any other time periods.

Recognizing that the transition patterns in the morning and evening may differ significantly, we divide the data into two time blocks: morning (7:00 AM to 11:59 AM) and evening (12:00 PM to 7:59 PM). We discretize time by aggregating all activity into 5-minute intervals to balance granularity and computational feasibility. We think 5-minute intervals provide sufficient resolution to capture dynamic changes in bike availability. To determine the stationary distribution, we first estimate the transition probability matrix P for each time block.

As discussed in Section 2.1, we calculate the number of available bikes at a station by updating it based on ride activity. Specifically, the number of available bikes at station i at time point t, $a_{i,t}$, is adjusted by accounting for the bikes departing from and arriving at the station during each time interval. By tracking the transitions from $a_{i,t}$ to $a_{i,t+1}$ and counting the frequency of each transition, we can estimate the transition probabilities for the Markov Chain. These probabilities represent the likelihood of the system moving from one state (a specific number of available bikes) to another in a 5-minute time period, forming the basis of the transition probability matrix.

More specifically, we model the DTMC for a station i with a state space $S = \{0, \ldots, c_i\}$, where each state represents the number of available bikes at the station. To estimate the transition probabilities, we track the number of transitions between states k and l within the state space S, where k and l represent two specific states. The count of these transitions over the entire time horizon is denoted as w_{kl}. The transition probability p_{kl}, representing the likelihood of transitioning from state k to state l, is then calculated as

$$p_{kl} = \frac{w_{kl}}{\sum_{o \in S} w_{ko}}, \quad \forall k, l \in S, \tag{1}$$

where, the numerator w_{kl} is the observed count of transitions from k to l, and the denominator

$\sum_{o \in S} w_{ko}$ is the total count of all transitions originating from state k. This formulation ensures that the transition probabilities for any given state k sum to 1, maintaining the fundamental property of a Markov Chain.

Finally, to get the stationary distribution of station i, $\pi_i = [\pi_0, \ldots, \pi_{c_i}]$, we only need to use the transition matrix P_i that we found for this station. The stationary distribution of station i can be found by solving the following system of equations.

$$\begin{aligned} \pi_i P_i &= \pi_i, \\ \sum_{k=0}^{c_i} \pi_{ik} &= 1, \\ \pi_{ik} &\geq 0, \quad \forall k \in \{0, 1, \ldots, c_i\}, \end{aligned}$$

which we can repeat for both the morning and evening blocks to get the desired stationary distributions.

3 Results

In this section, we present our results. We begin by exploring the dataset through the warm-up questions outlined in Section 3.1. Next, in Section 3.3, we detail our station selection process and analyze their overall activity patterns. Building on these insights, we proceed to the steady-state results in Section 3.4 and conclude with a sensitivity analysis of station capacities in Section 3.5.

3.1 Warm-Up Questions

In this section, we present the results of the warm-up questions. These questions serve to provide an initial understanding of the data.

3.1.1 Question 1: Using the start time and end time, compute the duration of each ride in minutes and plot the histogram of ride durations.

To analyze the ride durations, the difference between the start time (`started_at`) and end time (`ended_at`) was computed in minutes. As discussed in Section 2, any rides lasting more than 3 hours or with negative durations were removed from the dataset. In Figure 1 and 2, we present two histograms of ride durations. These figures reveal a right-skewed distribution, with most rides lasting less than 25 minutes. This indicates that the majority of rides are short trips, likely for commuting or errands.

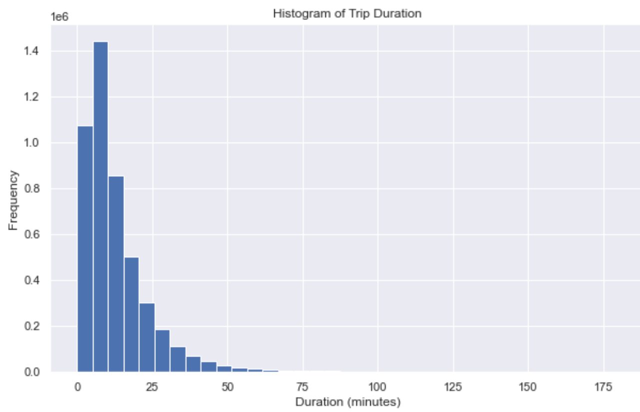


Figure 1: Histogram of the ride duration in minutes

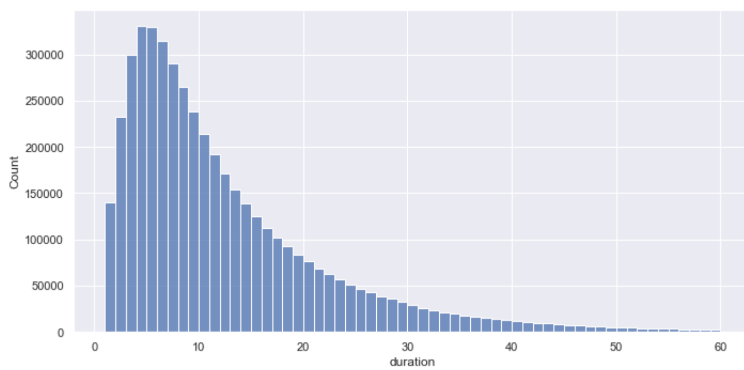


Figure 2: Histogram of ride durations (in minutes), scaled to the range of 1–60 minutes

3.1.2 Question 2: What is the expected ride duration (i.e., the average ride duration)? What is the empirical variance of ride duration? What is the probability that a ride duration is greater than 20 minutes?

- **Expected Ride Duration:** The average ride duration was calculated to be approximately $13.42 \approx 14$ minutes.

- **Variance of Ride Duration:** The variance of ride duration is approximately 172.15 minutes, showing moderate variability in ride times.

- **Probability of Ride Duration** $>$ 20 **minutes:** About 18.78% of rides had durations greater than 20 minutes.

3.2 Question 3: What is the probability that a ride duration is greater than 20 minutes conditioning on the fact that the user is a CitiBike member?

The conditional probability that a ride duration exceeds 20 minutes, given that the user is a CitiBike member, was calculated to be around 14.93%. CitiBike members appear less likely to take longer rides compared to casual users. This probability was calculated by using Bayes' Rule.

3.2.1 Question 4: Suppose that the duration of some ride is more than 25 minutes. What is the probability that this ride belongs to a CitiBike member?

Given a ride lasting more than 25 minutes, the probability of it belonging to a CitiBike member is 57.41%. This indicates that CitiBike members are more likely to take longer trips. Again, we used Bayes' Rule to calculate this probability.

3.2.2 Question 5: What is the expected ride duration of an electric bike? What is the expected ride duration of a classic bike?

- **Electric Bikes:** The expected ride duration was $13.78 \approx 14$ minutes.
- **Classic Bikes:** The expected ride duration was $12.74 \approx 13$ minutes.

Therefore, electric bikes have a marginally higher average ride duration, likely due to their ability to cover greater distances in the same time or their appeal for slightly longer commutes.

3.2.3 Question 6: Suppose that the duration of some ride is less than 10 minutes. What is the probability that this ride uses an electric bike? What is the probability that this ride uses a classic bike? Comment on the results.

- **Probability of Electric Bike Usage:** 63.33%.

- **Probability of Classic Bike Usage:** 36.67%.

The data reveals that electric bikes account for 63.33% of rides, while classic bikes account for 36.67%. This indicates that short rides are predominantly associated with electric bikes. This trend can be attributed to the convenience they provide, requiring less physical effort and making them ideal for short commutes. Additionally, CitiBike likely deploys electric bikes strategically in high-demand areas to optimize availability and turnover. Membership incentives, such as discounts on electric bike usage, may further encourage their adoption.

3.3 Station Analyses

To select the three stations for this study, we first plot the top 25 stations with the highest activity in July 2024, as shown in Figure 3. After examining the activity patterns of various stations, we chose to perform the steady-state study on 'W 21 St & 6 Ave', 'E 40th St & Park Ave', and 'West St & Chambers St'. While many of the busiest stations exhibited similar patterns with peaks during morning and evening hours, these three stations were selected for their distinct activity dynamics around these peak times. In Sections 3.3.1, 3.3.2, and 3.3.3, we discuss the activity patterns of each station individually and highlight the key differences between them.

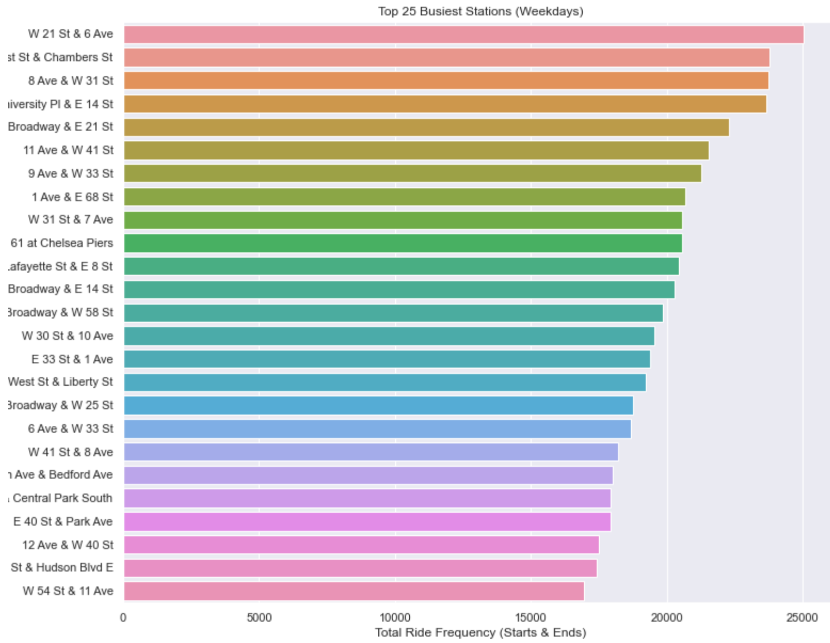


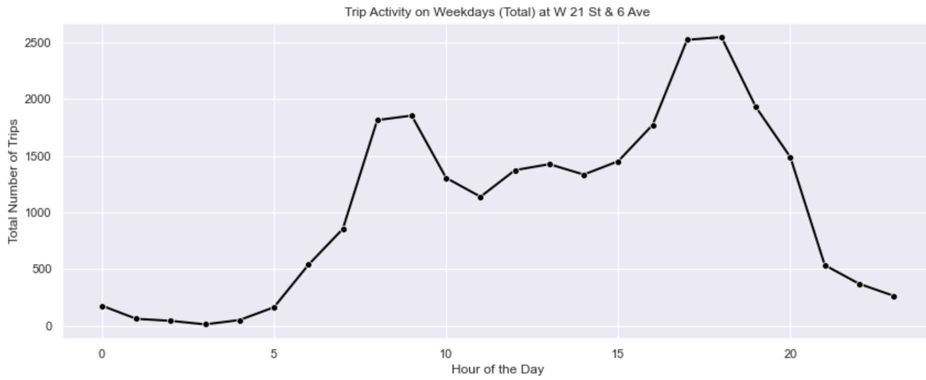
Figure 3: Top 25 stations with the most activity in July 2024

3.3.1 W 21 St & 6 Ave

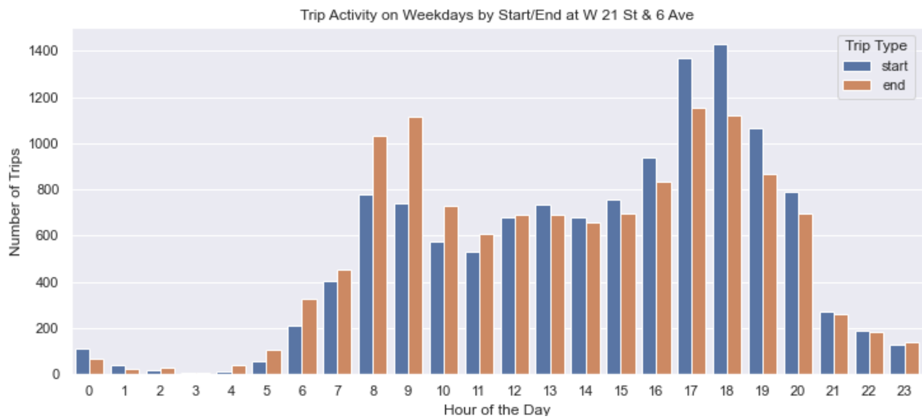
The CitiBike station at W 21 St & 6 Ave ranked as the busiest station in July 2024. It is situated in Manhattan's Flatiron District, a vibrant area known for its mix of residential, commercial, and cultural establishments. This neighborhood features a high concentration of office buildings, retail shops, and restaurants, making it a central hub of activity for both commuters and visitors.

Given its central location, this station likely experiences significant activity during weekday morning and evening rush hours, driven by commuters traveling to and from nearby workplaces. Additionally, mid-day activity is expected to be steady, as tourists and shoppers utilize CitiBikes to explore the area. The presence of residential buildings in the vicinity may also contribute to moderate activity during non-peak hours.

In Figure 4a, we display the total ride activity pattern throughout the day at this station, while Figure 4b illustrates the breakdown between starting and ending rides. These figures confirm the expected activity pattern, with clear peaks during the morning and evening rush hours. Additionally, the data reveals that activity remains steady throughout the day, gradually building toward the next peak period.



(a) Total ride activity pattern



(b) Division of starting and ending rides

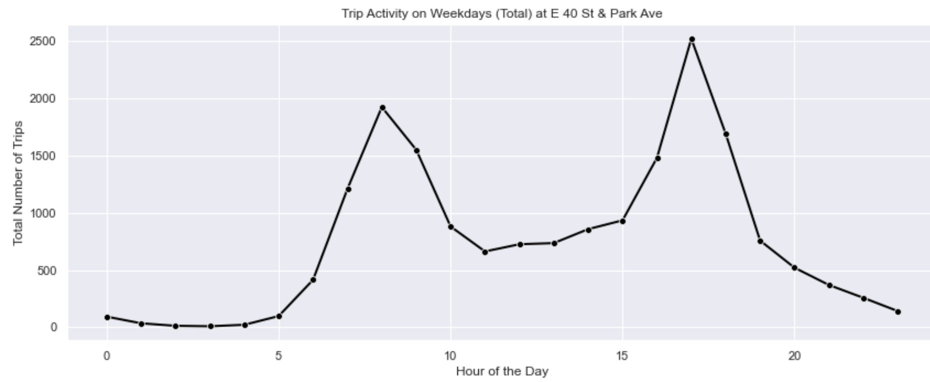
Figure 4: Activity patterns of W 21 St & 6 Ave

Interestingly, Figure 4b shows that morning rush hour is characterized by more ending trips, while evening rush hour sees more starting trips. This pattern aligns with commuter behavior: in the morning, many riders likely use CitiBikes to travel to workplaces or destinations near this station, resulting in a higher number of bike drop-offs. Conversely, in the evening, these riders use the station to begin their return trips home or to other locations, contributing to the higher number of bike pickups. This reflects the station's role as a transit hub within the neighborhood's dynamic activity flow.

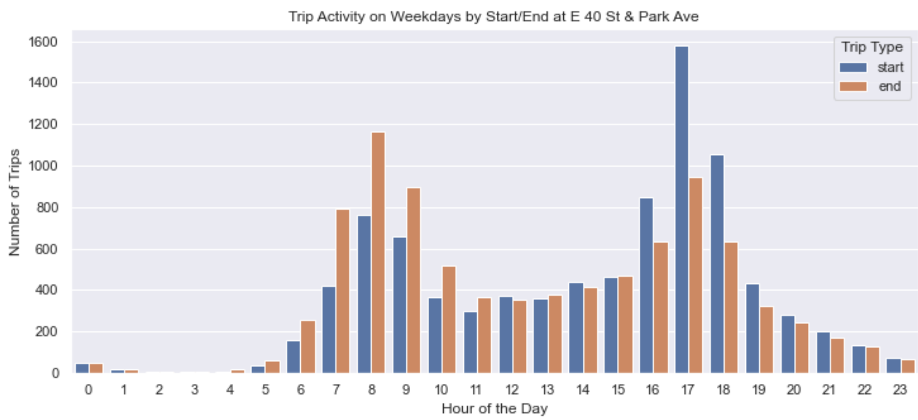
3.3.2 E 40th St & Park Ave

The CitiBike station at E 40th St & Park Ave ranks as the 22nd busiest station in our top 25 for July 2024. Situated in Midtown Manhattan, this station is located in one of the most iconic commercial districts of New York City, known for its proximity to Grand Central Terminal and the vibrant activity of corporate offices. Due to its location, we expect a high volume of commuters using this station, particularly during morning and evening rush hours, as it serves as a key transit point for individuals traveling to and from work.

Figures 5a and 5b illustrate the total activity and the breakdown of starting and ending rides for this station. The activity patterns at the E 40th St & Park Ave station exhibit distinct peaks during the morning and evening rush hours, reflecting its role as a major commuting hub in Midtown Manhattan. Unlike W 21 St & 6 Ave, with steadier activity throughout the day, this station experiences a noticeable drop in activity during non-peak hours, likely due to its proximity to corporate offices. This suggests that the station's usage is predominantly tied to commuter traffic rather than leisure activity. This observation is further supported by the breakdown of ending and starting trips during rush hours.



(a) Total ride activity pattern



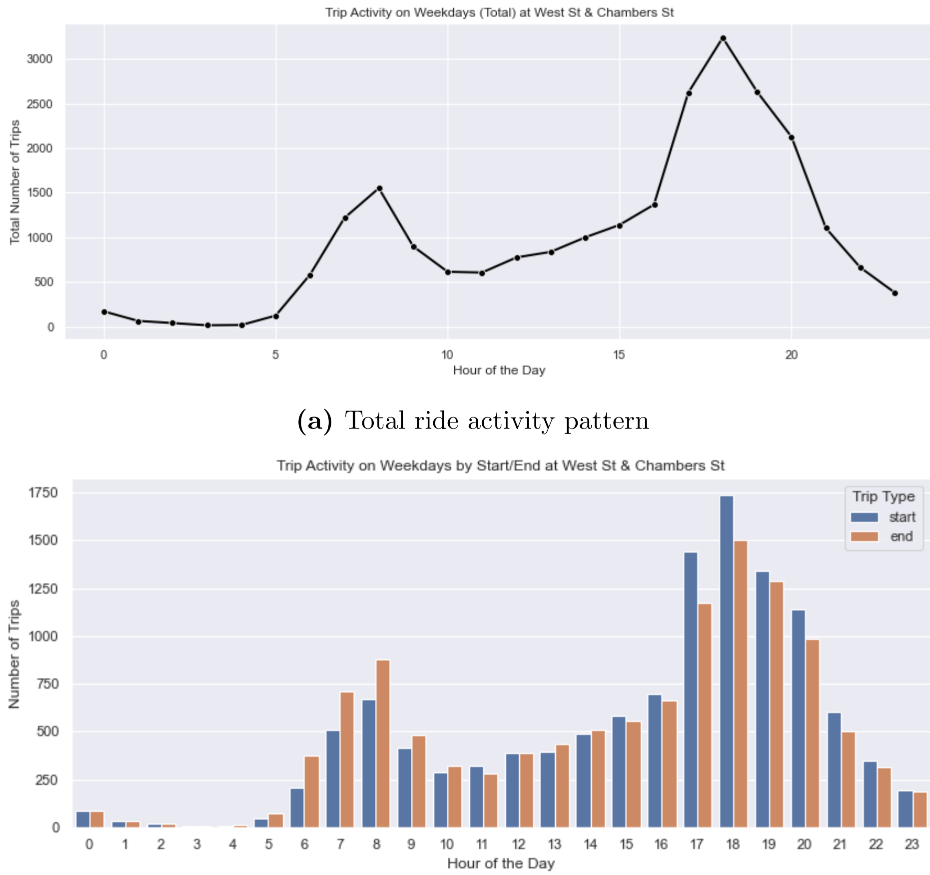
(b) Division of starting and ending rides

Figure 5: Activity patterns of E 40th St & Park Ave

3.3.3 West St & Chambers St

The West St & Chambers St station ranks as the second busiest station in our top 25 for July 2024. Located in the Battery Park City neighborhood of Lower Manhattan, this station is surrounded by a mix of residential complexes and key attractions such as the World Trade Center and Hudson River Park. This area is characterized by many residents and visitors frequenting the nearby parks and waterfront. Therefore, we anticipate peaks in the afternoon and evening, driven by a combination of commuters returning home and visitors taking advantage of the cooler periods of the day in July.

Figures 6a and 6b confirm that the highest activity levels occur during the evening hours. Notably, unlike the other two stations, the evening peak at West St & Chambers St is more prolonged, extending beyond the typical 5–6 PM rush hour observed at E 40th St & Park Ave, before gradually returning to average ride levels.



(a) Total ride activity pattern

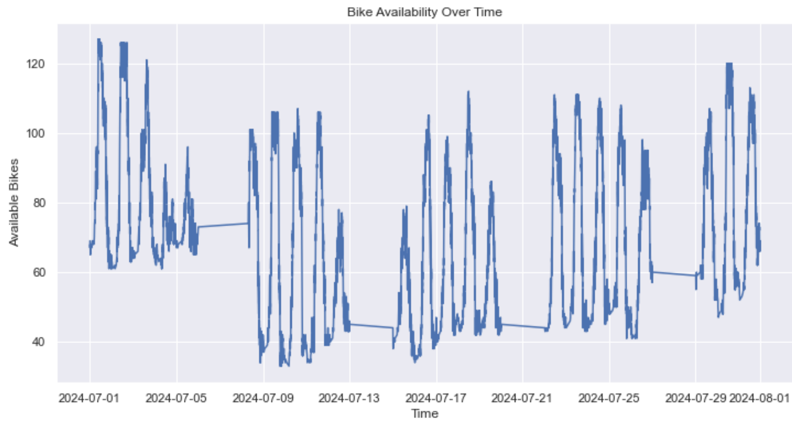
(b) Division of starting and ending rides

Figure 6: Activity patterns of West St & Chambers St

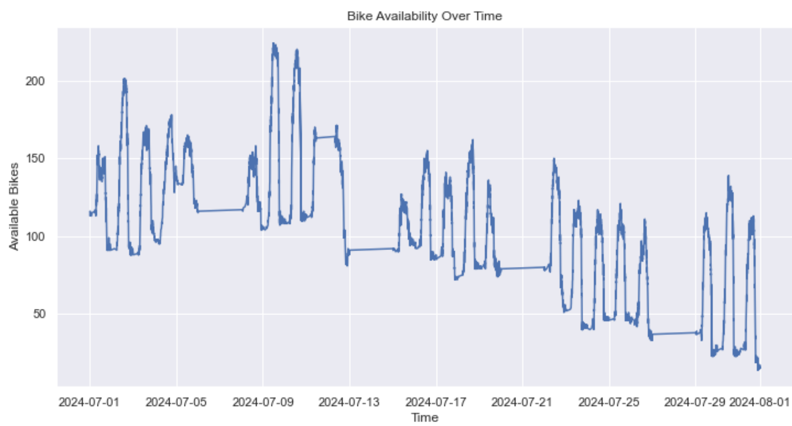
Overall, while most of the top stations experience peaks in activity during the morning and evening hours, the three selected stations exhibit unique patterns around these peaks and are located in distinct areas of Manhattan.

3.4 Stationary Distributions

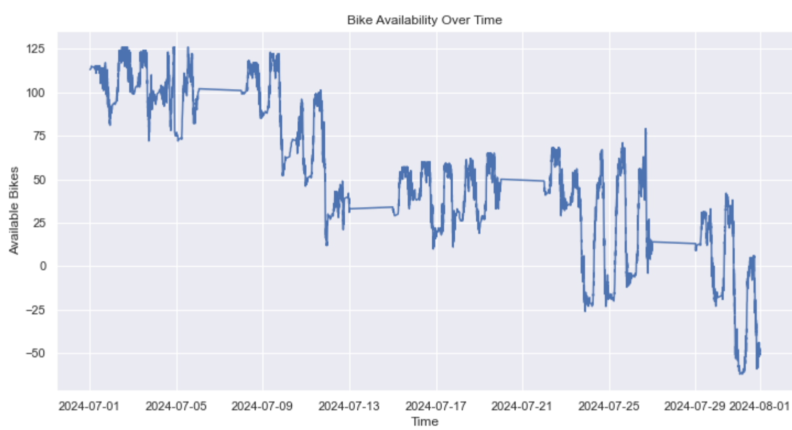
As outlined in Section 2, determining the stationary distribution π_i of a station requires knowing the station capacities c_i to construct the transition matrix P_i. We obtained the station capacities from Google Maps, with W 21 St & 6 Ave, E 40th St & Park Ave, and West St & Chambers St having capacities of 70, 114, and 112 docks, respectively. Using these capacities, we updated the bike availability based on the ride activity in the dataset, resulting in the bike availability plots shown in Figure 7.



(a) Bike availability W 21 St & 6 Ave



(b) Bike availability E 40th St & Park Ave



(c) Bike availability West St & Chambers St

Figure 7: Bike availability for each of the three stations in July 2024.

In Figure 6, we observe that bike availability can fall outside the expected range $[0, c_i]$ for a station i. Upon careful analysis, we determined this is likely due to CitiBike's rebalancing efforts, a process where bikes are redistributed to align availability with customer demand or achieve station capacity targets. To address this issue in our analysis, we constrained the bike availability to remain within the range $[0, c_i]$.

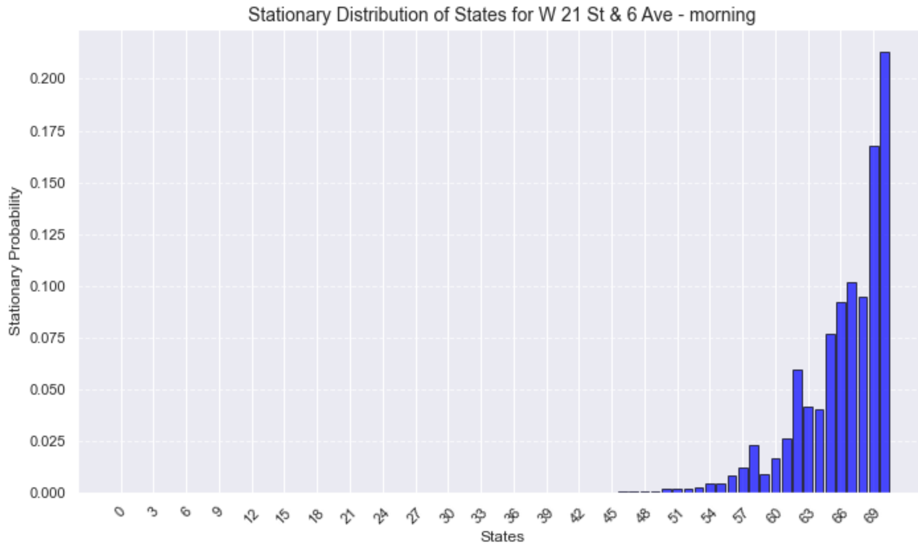
After consolidating the data into 5-minute intervals and creating separate data frames for morning and evening periods, we applied the steady-state approach outlined in Section 2. The results for each station are detailed in Sections 3.4.1, 3.4.2, and 3.4.3.

3.4.1 Steady-State Estimation - W 21 St & 6 Ave

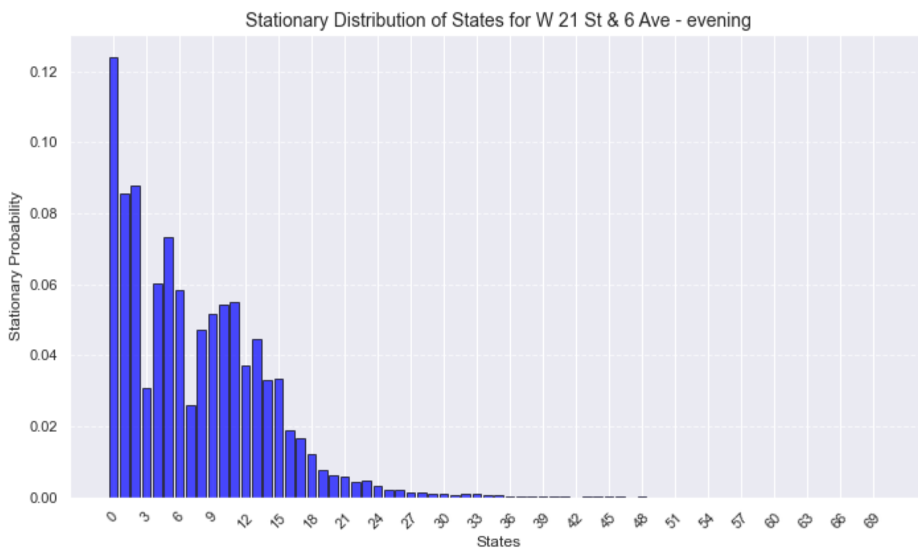
For W 21 St & 6 Ave, the stationary distributions for the morning and evening periods are shown in Figures 8a and 8b, respectively. These distributions are visualized by plotting the steady-state probability of each state in the state space $\{0, \ldots, c_i\}$. Each probability represents the estimated long-term likelihood of having a specific number of available bikes at the station.

As illustrated in Figure 8a, the W 21 St & 6 Ave station consistently shows a high number of available bikes, often nearing its full capacity, between 7:00 AM and 11:59 PM. This observation is supported by Figure 4b, which highlights that during this time, more bikes are being dropped off than taken away. During the evening hours (12:00 PM - 7:59 PM), the station exhibits the opposite pattern, with a significantly higher likelihood of having very few available bikes. The probability of finding more than 25% of the station's capacity in bikes is notably low. This trend is corroborated by Figure 4b, which shows that during this time, more bikes are taken out than dropped off.

These results align closely with the expectations based on the station's location in Manhattan's bustling Flatiron District. The high probability of finding the station near full capacity during the morning hours reflects the area's role as a hub for commuters heading to nearby office buildings and workplaces. Similarly, the likelihood of the station being nearly empty during the evening hours is consistent with workers leaving the area. While the likelihood of the station being nearly empty during the evening hours is high, the probabilities are somewhat more spread around the first 20 states, confirming the steady mid-day usage likely driven by tourists, shoppers, and local residents exploring the area.



(a) Steady-state probabilities in the morning



(b) Steady-state probabilities in the evening

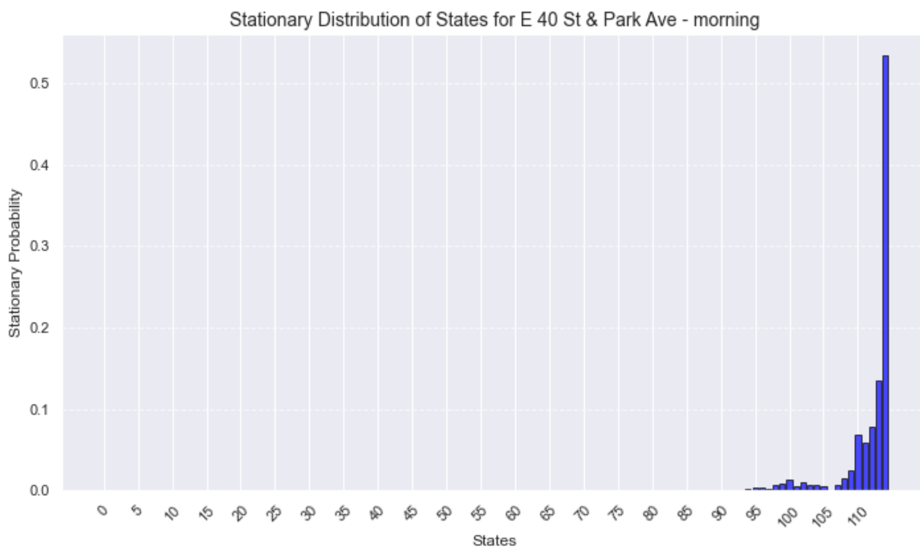
Figure 8: Steady-state probabilities (morning and evening) of station W 21 St & 6 Ave

3.4.2 Steady-State Estimation - E 40th St & Park Ave

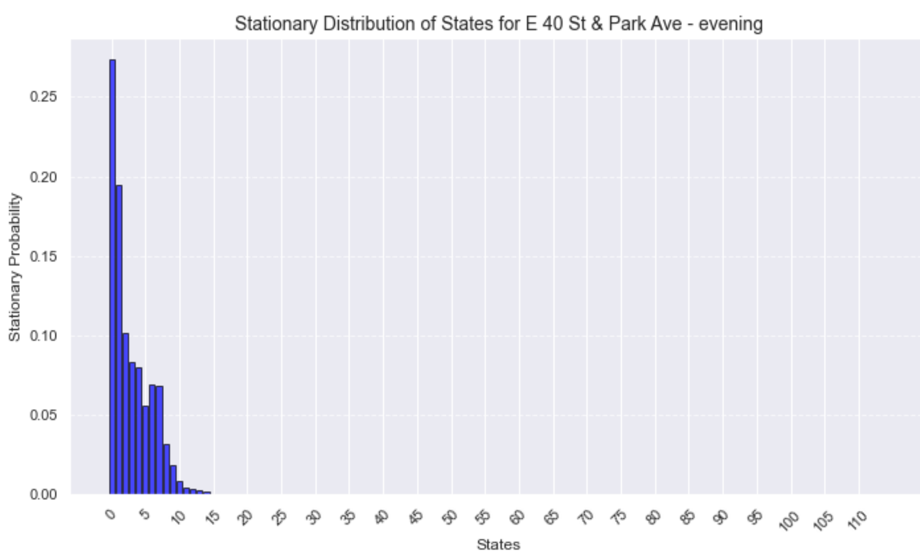
The CitiBike station at E 40th St & Park Ave displays a steady-state pattern similar to that described in Section 3.4.1. As shown in Figure 9a, there is more than a 50% chance of the station being full during the morning hours, while in the evening hours, the likelihood of having no bikes available exceeds 25%. These observations align with the trends illustrated in Figure 5b, where more bikes are dropped off in the morning and fewer in the evening.

Compared to W 21 St & 6 Ave, the likelihood of the station being full in the morning and empty in the evening is significantly higher.

The activity patterns observed in Figure 5a explain this steady-state behavior. The clear spikes in morning and evening rush hour activity, coupled with the drop during mid-day, result in the steady-state probabilities concentrating at full availability in the morning and zero bikes in the evening. This reflects the station's heavy reliance on commuter traffic, with little mid-day activity. This is consistent with the station's proximity to Grand Central Station.



(a) Steady-state probabilities in the morning

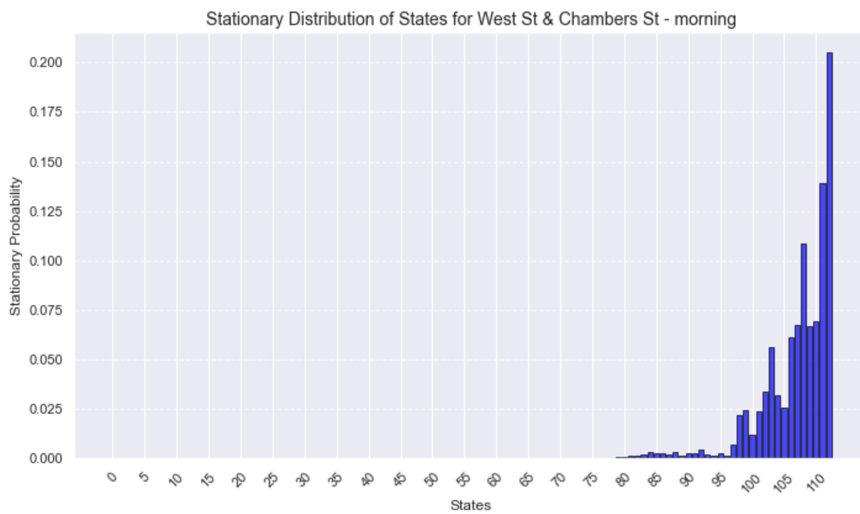


(b) Steady-state probabilities in the evening

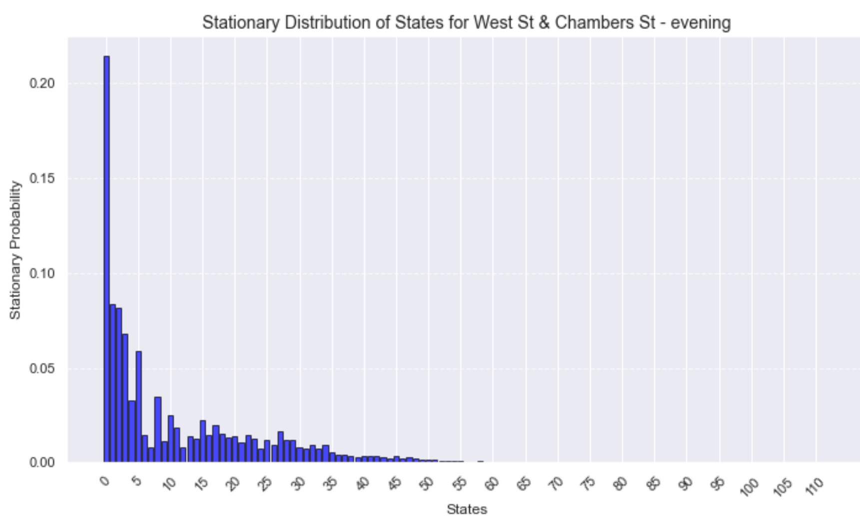
Figure 9: Steady-state probabilities (morning and evening) of station E 40th St & Park Ave

3.4.3 Steady-State Estimation - West St & Chambers St

Similar to the previous two stations, Figure 10 shows that the probability of the station being full in the morning and empty in the evening is the highest, at approximately 20%. For intermediate states, the steady-state probability is zero. This behavior is consistent with the station's location in Battery Park City, a residential and recreational hub. The high probability of the station being empty in the evening aligns with the prolonged evening peak, driven by residents returning home and visitors concluding their day at nearby attractions like Hudson River Park.



(a) Steady-state probabilities in the morning



(b) Steady-state probabilities in the evening

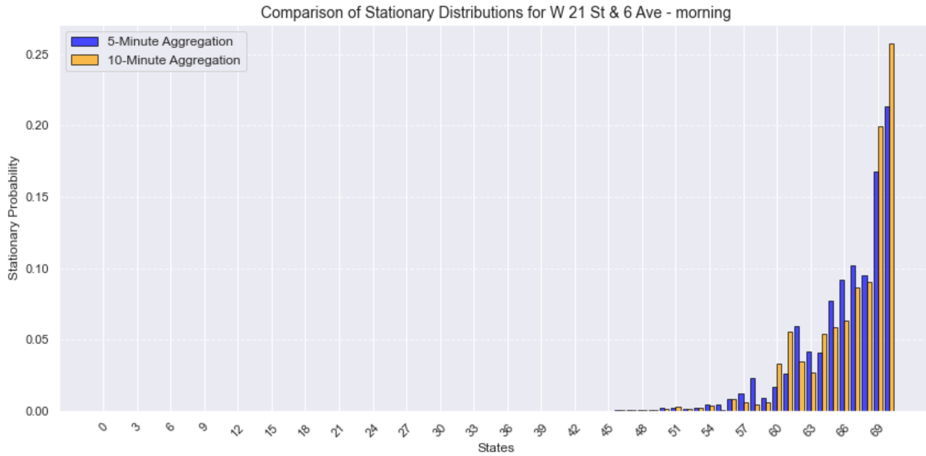
Figure 10: Steady-state probabilities (morning and evening) of station West St & Chambers St

3.5 Sensitivity Analysis

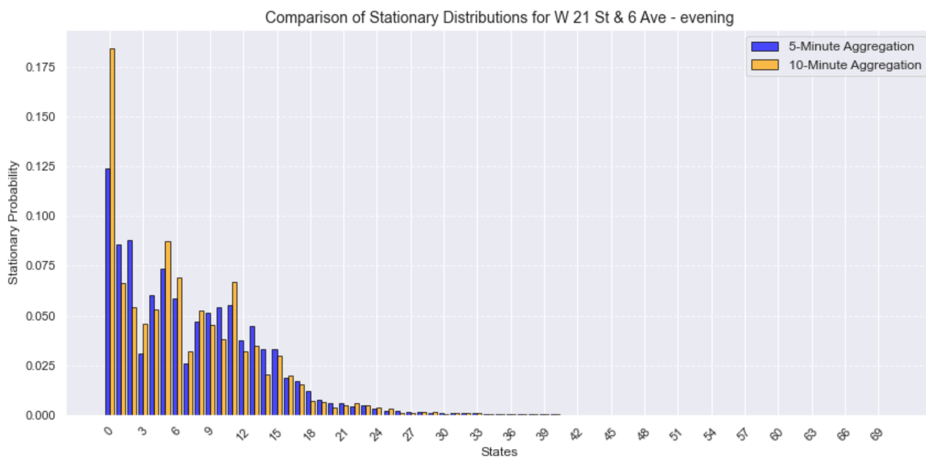
As outlined in Section 2, this research relies on several assumptions that could influence the stationary distribution. Firstly, we discretize time by aggregating ride data into 5-minute intervals. Secondly, we assume that each station begins with full bike availability on July 1st. To assess the impact of these assumptions on the stationary distributions, we conduct a sensitivity analysis. Section 3.5.1 examines the effect of changing the time aggregation from 5-minute to 10-minute intervals, while Section 3.5.2 explores the influence of varying initial bike availability, ranging from 50% to 100% of a station's total capacity c_i.

3.5.1 Time Aggregation Intervals (5 vs. 10 minutes)

Figures 11, 12, and 13 show the steady-state distributions for the morning and evening blocks, comparing results based on 5-minute and 10-minute time aggregations. Overall, the steady-state patterns for each station remain consistent across both the morning and evening time blocks. However, when using a 10-minute aggregation, the patterns become more pronounced, with peaks being more distinct compared to the 5-minute aggregation. This is likely because the larger time intervals reduce short-term fluctuations and better capture the overall trends in bike availability, making the steady-state behavior clearer and more interpretable. Overall, we can conclude that the 5-minute aggregation effectively captured the stationary behavior.

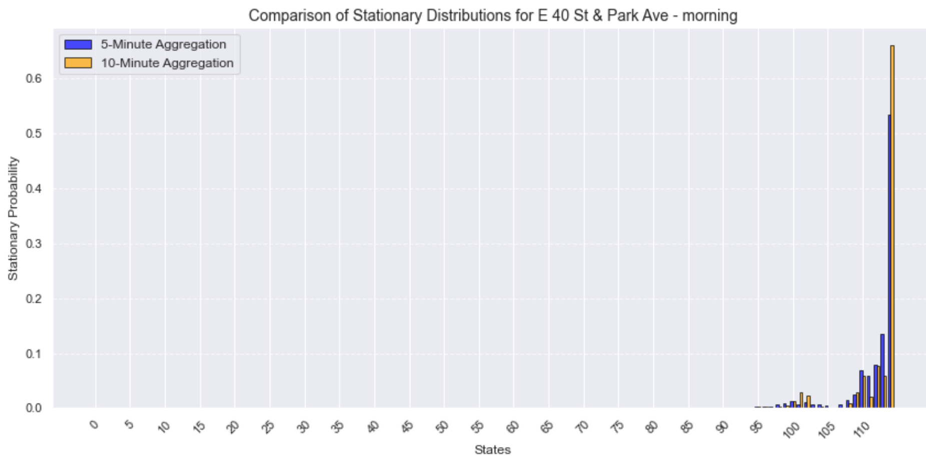


(a) Steady-state probabilities in the morning

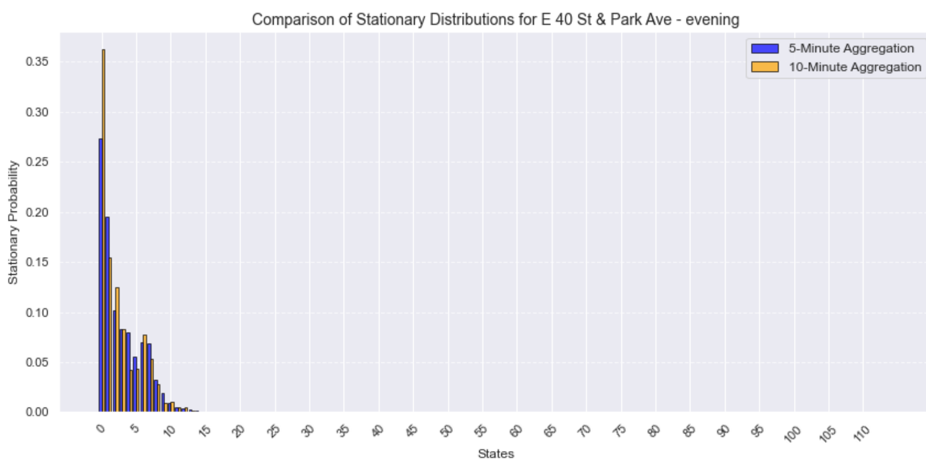


(b) Steady-state probabilities in the evening

Figure 11: Steady-state probabilities (morning and evening) of station W 21 St & 6 Ave using 5 vs. 10-minute aggregation

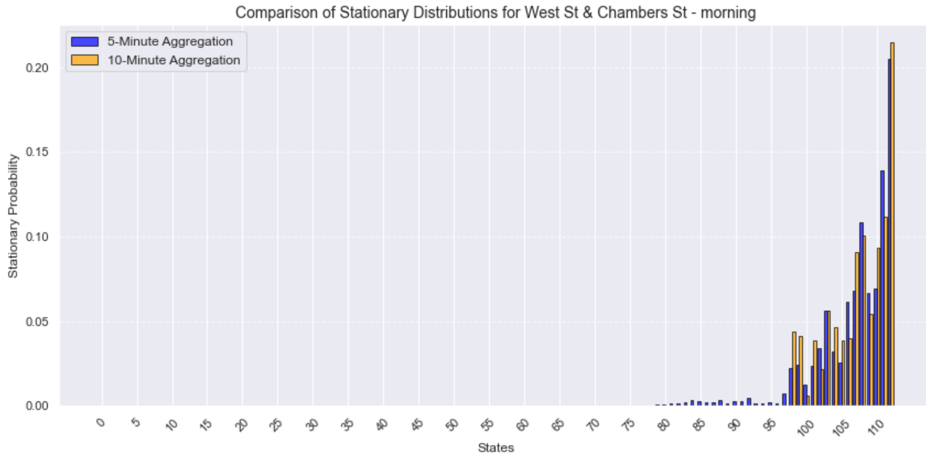


(a) Steady-state probabilities in the morning

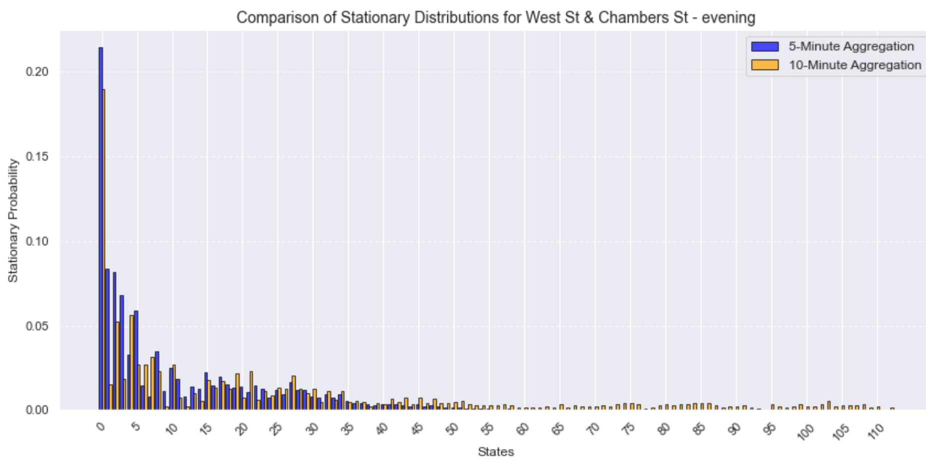


(b) Steady-state probabilities in the evening

Figure 12: Steady-state probabilities (morning and evening) of station E 40 St & Park Ave using 5 vs. 10-minute aggregation



(a) Steady-state probabilities in the morning



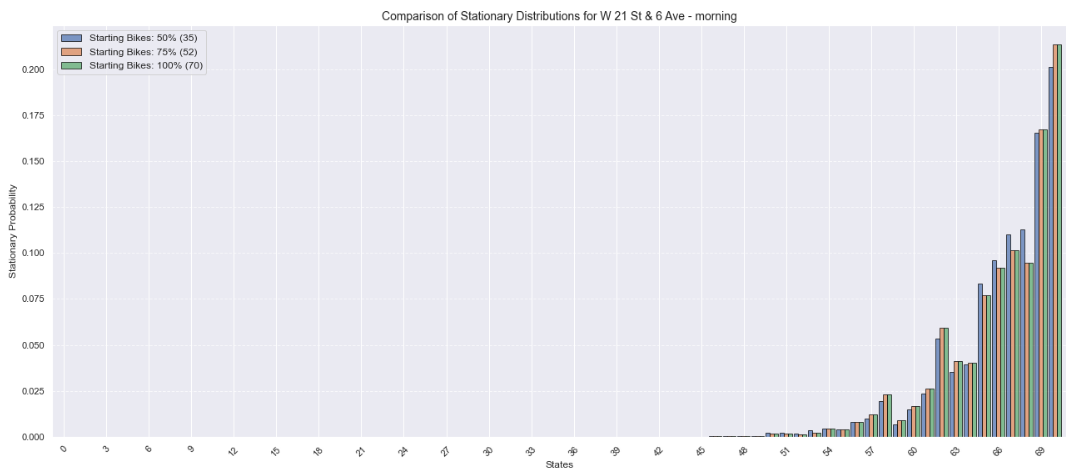
(b) Steady-state probabilities in the evening

Figure 13: Steady-state probabilities (morning and evening) of station West St & Chambers St using 5 vs. 10-minute aggregation

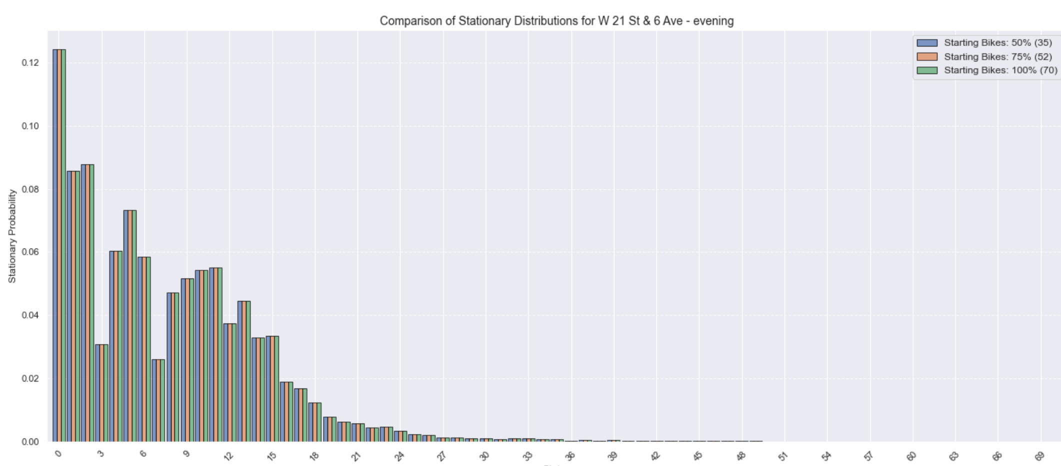
3.5.2 Available Bike State Initialization

We present the sensitivity analysis of the initial number of available bikes at each station on July 1st in Figures 14, 15, and 16, where we test starting capacities set to 50%, 75%, and 100% of the full dock capacity. Similar to the results discussed in Section 3.5.1, the overall patterns of the stationary distributions remain consistent. The only exception is the morning steady-state distribution for the West St & Chambers St station.

The observed shift in the morning steady-state distribution for the 50% starting capacity, where the stationary probabilities peak around 85 instead of 112, can be attributed to the station's bike activity dynamics in July 2024. During this period, the data reveals an overall downward trend in bike availability. Starting the month with a lower initial capacity makes it significantly harder for the station to reach higher states later on, as the reduced starting point limits its ability to recover and climb to higher availability levels compared to when it begins with a higher capacity. This effect is particularly pronounced for this station, as it even fell far below state 0 during the month, at times reaching deficits of more than -50 bikes. This further emphasizes the frequency of being in a low state and illustrates the challenge of operating with lower starting capacity. As a result, it is expected that the long-term behavior no longer centers around full bike availability.

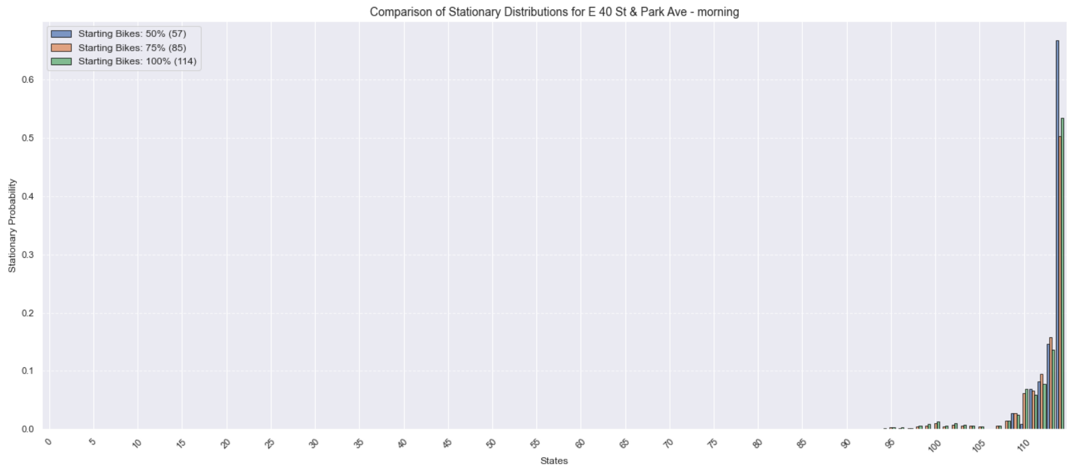


(a) Steady-state probabilities in the morning

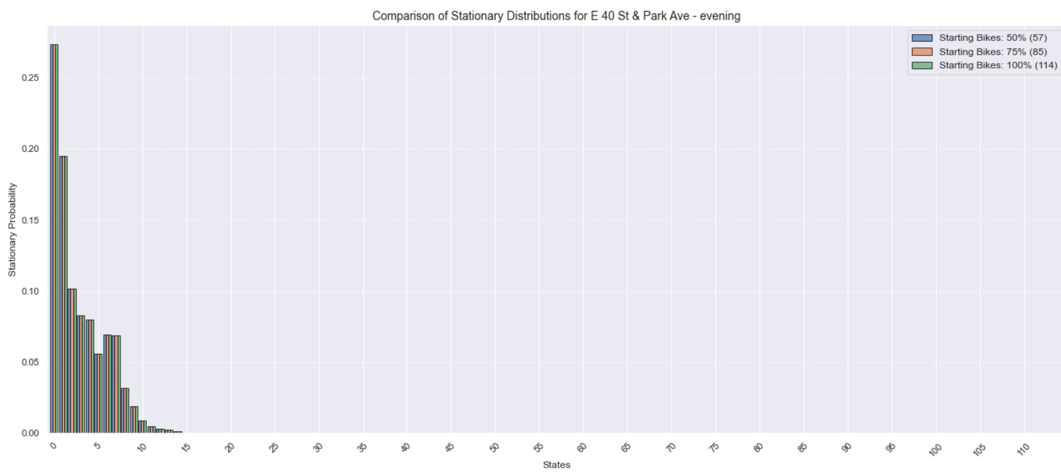


(b) Steady-state probabilities in the evening

Figure 14: Steady-state probabilities (morning and evening) of station W 21 St & 6 Ave using different starting capacities

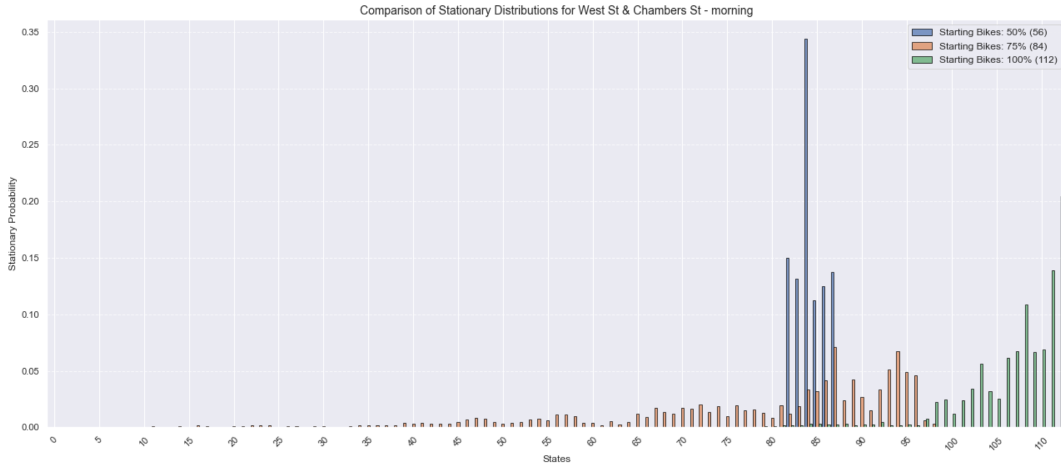


(a) Steady-state probabilities in the morning

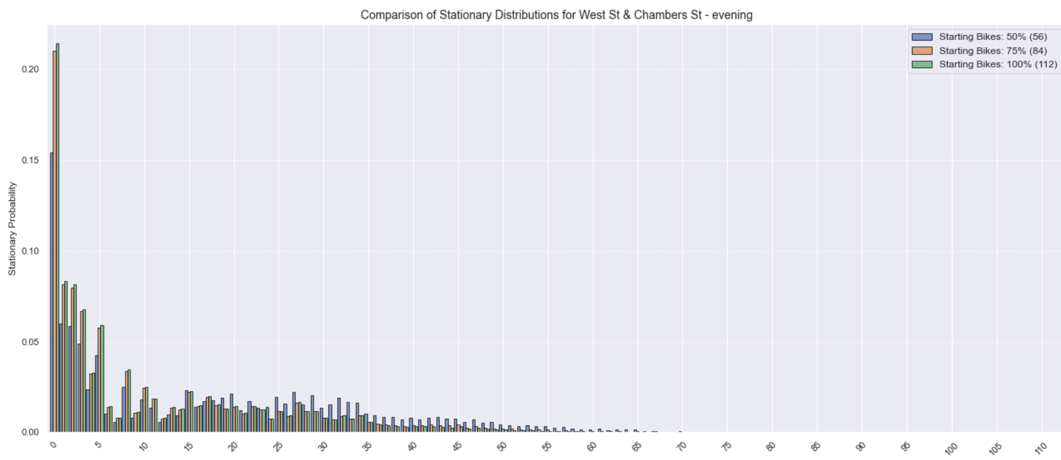


(b) Steady-state probabilities in the evening

Figure 15: Steady-state probabilities (morning and evening) of station E 40 St & Park Ave using different starting capacities



(a) Steady-state probabilities in the morning



(b) Steady-state probabilities in the evening

Figure 16: Steady-state probabilities (morning and evening) of station West St & Chambers St using different starting capacities

4 Conclusion

In this study, we estimated the steady-state distribution of bike availability at three popular CitiBike stations in Manhattan using Discrete Time Markov Chain modeling. By using ride data from July 2024 and dividing it into morning and evening blocks, we effectively estimated bike availability levels at each station and gained valuable insights into their usage patterns.

Our findings reveal several key insights. First, the steady-state distributions for all stations showed significant variation between the morning and evening blocks, highlighting distinct commuter behaviors and activity patterns during these periods. In the morning, the steady-state bike availability tends to approach full dock capacity, whereas in the evening, it is often close to zero. These patterns also underscore the impact of commuting rush hours, which were evident across all three stations. Nevertheless, each station exhibited distinct impacts of rush hours and unique patterns throughout the day, reflecting variations in their location and the specific dynamics of commuter activity.

Our study has valuable implications for CitiBike operations. Understanding the steady-state distribution of bike availability can help optimize bike redistribution strategies, particularly for stations that experience significant demand fluctuations during the day. By using the stationary distribution, CitiBike can proactively deploy or redistribute bikes to prevent shortages or overflows. This could enhance rebalancing efficiency, reduce user wait times, and improve service during peak periods. For the three investigated stations, it may be beneficial to focus redistribution efforts during evening hours, as they are more likely to experience low bike availability during this time period. For stations with higher bike availability in the evening, CitiBike could explore reallocating bikes from those locations to the three investigated stations. This approach would maximize system efficiency and better meet commuter needs.

CitiBike's rebalancing efforts, where bikes are actively moved between stations, form a key limitation of our study. Based on the ride data, we observed times where station availabilities dropped below zero or exceeded their maximum dock capacities. Due to the lack of detailed information about these rebalancing operations, we adjusted our model to keep the availability within the feasible range. These manual adjustments may in general cause the stationary distribution to skew more toward the extreme states of the range. Therefore, examining how results might differ with detailed data on rebalancing activities could offer a more accurate representation of the system's dynamics.

In summary, our research offers meaningful insights into the steady-state behavior of three CitiBike stations. While certain assumptions were necessary, our sensitivity analysis demonstrated that the results remain robust across a range of conditions. For future research, we recommend accounting for uncertainties in ride dynamics, such as those arising from events, weather conditions, and holidays. This could involve utilizing a stochastic transition probability matrix. Additionally, extending the study to include all CitiBike stations and investigating weekend usage patterns could provide CitiBike with deeper insights into station dynamics.