Modeling Project

Name: Gareth Sun, Zimu Li, Dahong Zhang
NetID: zs389, zl898, dz285
Course Name: ORIE 5530 Modeling Under Uncertainty (2023FA)
Professor's Name: Prof. Omar El Housni

1 Warm Up Questions

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

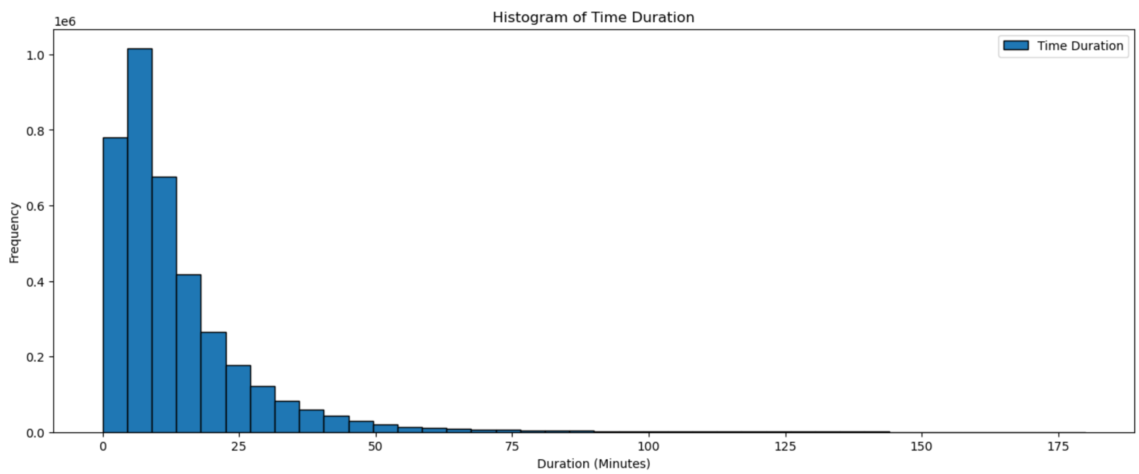

Figure 1: Histogram of ride Duration

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 min?

   Answer:

   Expected ride duration is 13.65 minutes.
   The empirical variance of ride duration 203.91 in minutes.
   The probability that a ride duration is greater than 20 min is 19.78%.

3. What is the probability that a ride duration is greater than 20 min conditioning on the fact that the user is a CitiBike member? Note that the last column gives whether the ride is for a casual client or a CitiBike member.

   Answer:

The probability that a ride duration is greater than 20 min conditioning on the fact that the user is a CitiBike member is 15.56%.

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

   Answer:

   The probability that this ride ride is more than 25min belongs to a CitiBike member is 57.88%.

5. What is the expected ride duration of an electric bike? What is the expected ride duration of a classic bike?

   Answer:

   The expected ride duration of an electric bike is 12.28 mins.

   The expected ride duration of a classic bike 13.72 mins.

6. Suppose that the duration of some ride is less than 10min. 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.

   Answer:

   The probability that this ride uses an classic bike 8.91%.

   The probability that this ride uses an electric bike 91.02%.

   The results indicate a strong preference for using classic bikes for shorter rides. For short distance riders, they would prefer to use classic bikes.

   This may be due to the availability of classic bikes and the fact that the charge for classic bikes is lower than for electric bikes. So, for those who travel short distances, they would not bother to find electric bikes. Electric bikes, on the other hand, are primarily used for long distance travel and take a longer time duration.

   However, the expectation that the time duration of classic bike use is shorter than that of electric bikes may be because a small number of classic bikes are left unlocked and used for a long time.

2 Citibike Project

2.1 Introduction

Citibike is a bike-sharing system operating in some large cities. And it displayed great benefits in reducing the traffic congestion and air pollution in New York City. However, because of the pattern of commute during different time, the availability of bike or docks may not be ensured. For example, during morning rush hours, people tend to move to the city center to work. On such condition, some popular stations in the city center may be filled with bike and people could find the available dock to return their bike.

To study and understand this question, we used the ride data of Citibike of New York City to simulate the in and out situation of some major bike stations and predict the availability of bikes.

Our project is based on the ride data of Citibike in NYC that happened in July 2023. We selected three busiest stations during that time as our research target and used discrete Markov chain to simulate the in and out of each station during commute hours of weekdays. Then, we calculated the stationary distribution of it, use it to obtain the expectation of available bikes in these stations.

2.2 Method

2.2.1 Data Clean

To simulate the availability of number of docks, we cleaned the ride data set by filtering out the rides which last for more three hours and those rides with negative time duration, which may be caused by the statistics errors. Also, since we only focus on the pattern of ride during weekdays. we also filtered out the ride data happened during the weekends.

2.2.2 Station Select

To select the busiest bike station, we calculate the total amount of in and out rides for each station. The results is shown in the following figure.

Based on the results, we select the top three stations as our research targets: W 21 St & 6 Ave, West St & Chambers St and 11 Ave & W 41 St.

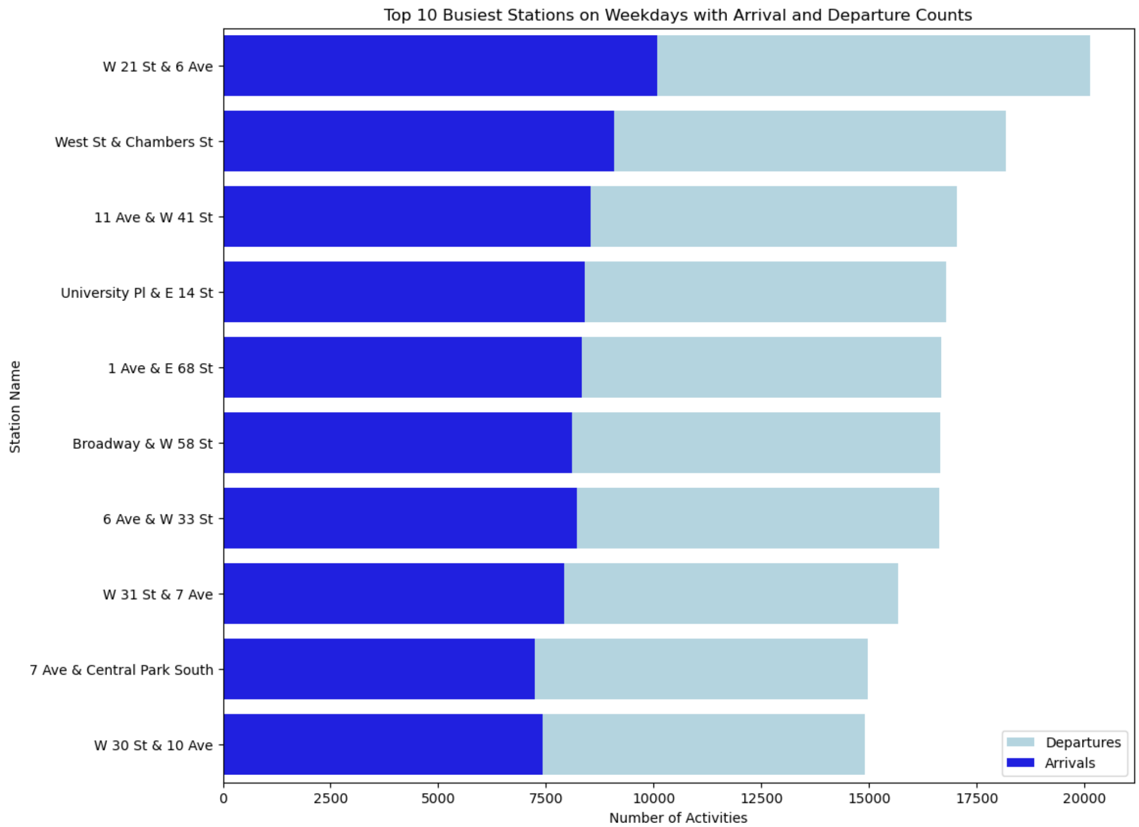

Figure 2: Top 10 Busiest Stations

2.2.3 discrete Markov Chain

To simulate the available bikes in each station, we used the discrete Markov chain. We selected two time periods for the project. The first is the 6 am to 11 am and the second one is from 4 pm to 10 pm. These two time periods cover the majority of New York City residents' commuting hours to and from work. Then, we divide each of the time periods into 10 mins' slots, and each of the time slot is considered as one step. And for each of the station, the state of the transition matrix is the number of bikes in the station.

For each station, we obtained the number of docks through the Citibike app, and used this data to determine the size of the transition matrix. The number of Docks is shown in the following:

W 21 St & 6 Ave: 52
West St & Chambers St : 101
11 Ave & W 41 St: 46

In the problem setting, we assume that in this Markov chain, if the difference between two states is the same, then the transition probability between these two states is equal.

As a result, for any state i, j there will be a path between i and j. The Markov chain is irreducible and will converge to the stationary distribution in the long run. Based on this fact, we can calculate the expectation of available bikes by multiplying the number of bikes and the stationary distribution probability.

2.2.4 Probability Distribution of Bike Number Change

Since we assume that in this Markov chain, if the difference between two states is the same, then the transition probability is equal. Then the distribution of bike number change for each station and each time period should be the same.

For each station and each time period, we obtain the bike number change situations and calculate the probability of bike change through ride data. We use the in and out ride information to calculate the bike number change for each 10-minute interval and use the frequency of such change to calculate the possibility change. We used the ride data to calculate the probability distribution. The results are shown in the following figures.

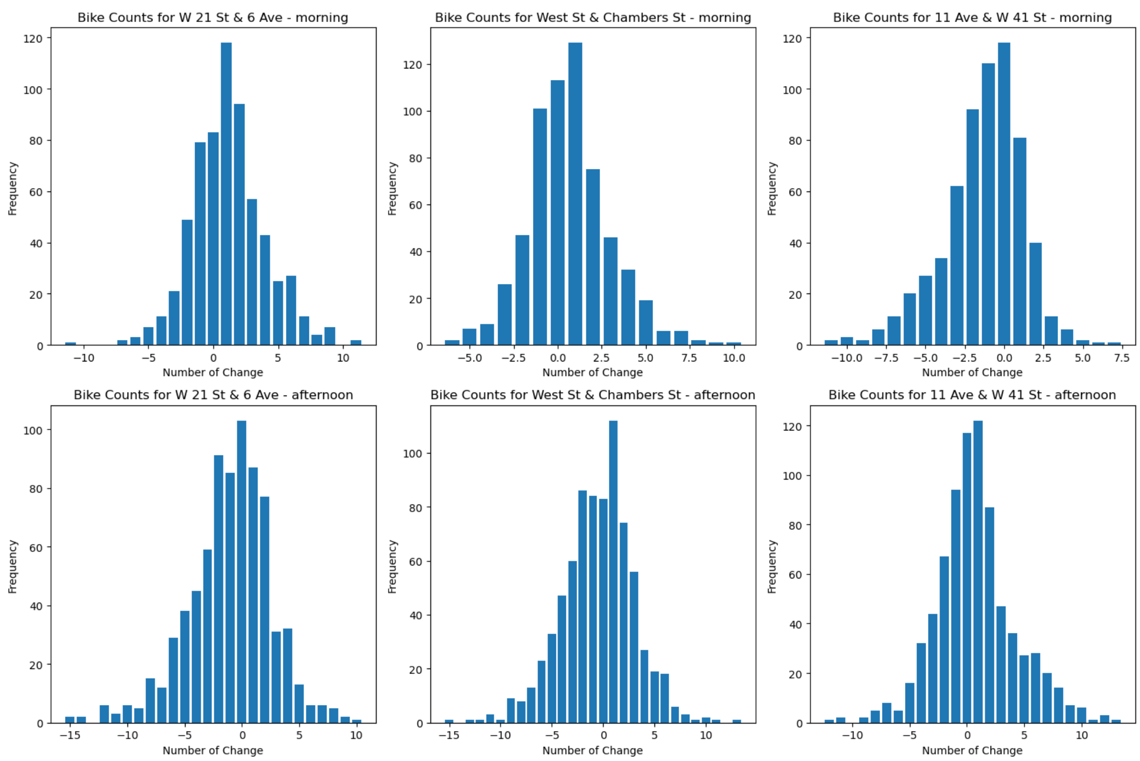

Figure 3: Net Change Frequency

So, we can calculate the distribution of bike number change through this function:

$$\text{Probability of change i} = \frac{\text{Number of intervals that bike number change i}}{\text{Total number of time intervals}}$$

2.2.5 Transition Matrix

For each station during each time period, we assume that when the difference between states is same, the probability is also the same. On such condition, we can use the probability distribution of bike number change to establish the transition matrix.

$$P(i, j) = \text{Probability of Change}(j - i)$$

But for the some states, it can't reach every kind of change in the probability distribution of change. For example, when there is only 1 bikes left in the station, the probability of change -2 should be 0 instead of what we calculated in the probability distribution.

To solve this problem, we need to calculate the change probability condition on the feasible changes:

$$P(i, j) = \frac{\text{Probability of Change}(j - i)}{\sum \text{Probability of feasible changes}}$$

Based on our probability distribution of number changes, each of the state could increase or decrease, so for any state i and j, there will be a path. As a result, our Markov chain is irreducible and we can get its stationary distribution through calculation.

2.2.6 Calculation of Stationary Distribution

Stationary Distribution represents the distribution of the number of bikes in the station in long run. Since the transition matrix is irreducible, there exists stationary distribution. We would use the following equations to calculate the stationary distribution.

$$\pi = \text{the stationary distribution vector with n states}$$

$$\pi P = \pi$$

$$\sum_{i=1}^{n} \pi_i = 1$$

$$\pi > 0$$

2.3 Result

2.3.1 Stationary Distribution

The stationary distribution is plotted in the following figures:

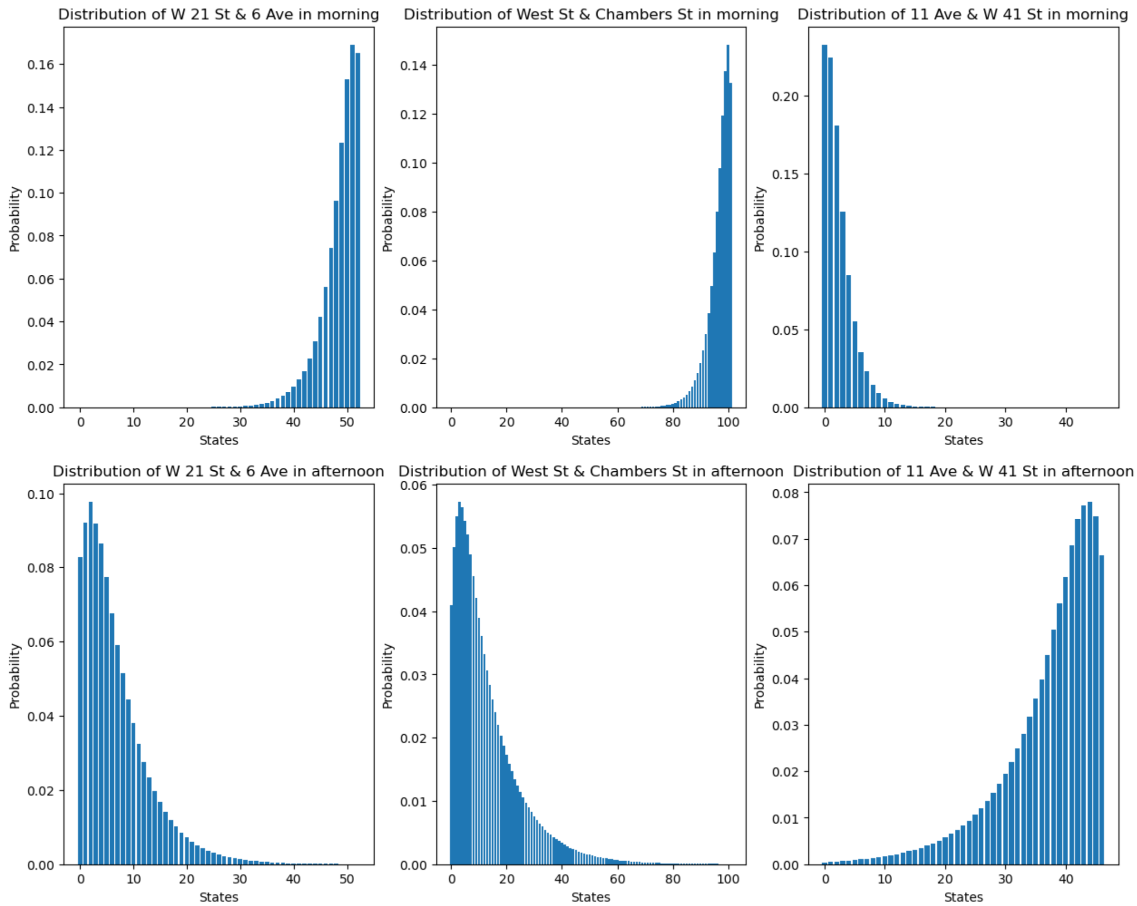

Figure 4: Stationary Distribution

2.3.2 Expectation

To calculate how many bikes are left in the station for different periods, we need to obtain the expectation of bike through the stationary distribution.

$$E(\text{bikes}) = \sum_{i}^{N} \pi(i) \times i$$

And the results is shown as following:

Table 1: Expectation of Bikes

Location	Morning	Afternoon
W 21 St & 6 Ave	48	7
West St & Chambers St	97	13
11 Ave & W 41 St	2	37

In each time period, the distribution of bicycle counts at the stations as a percentage of the total dock capacity is shown in the following figures.

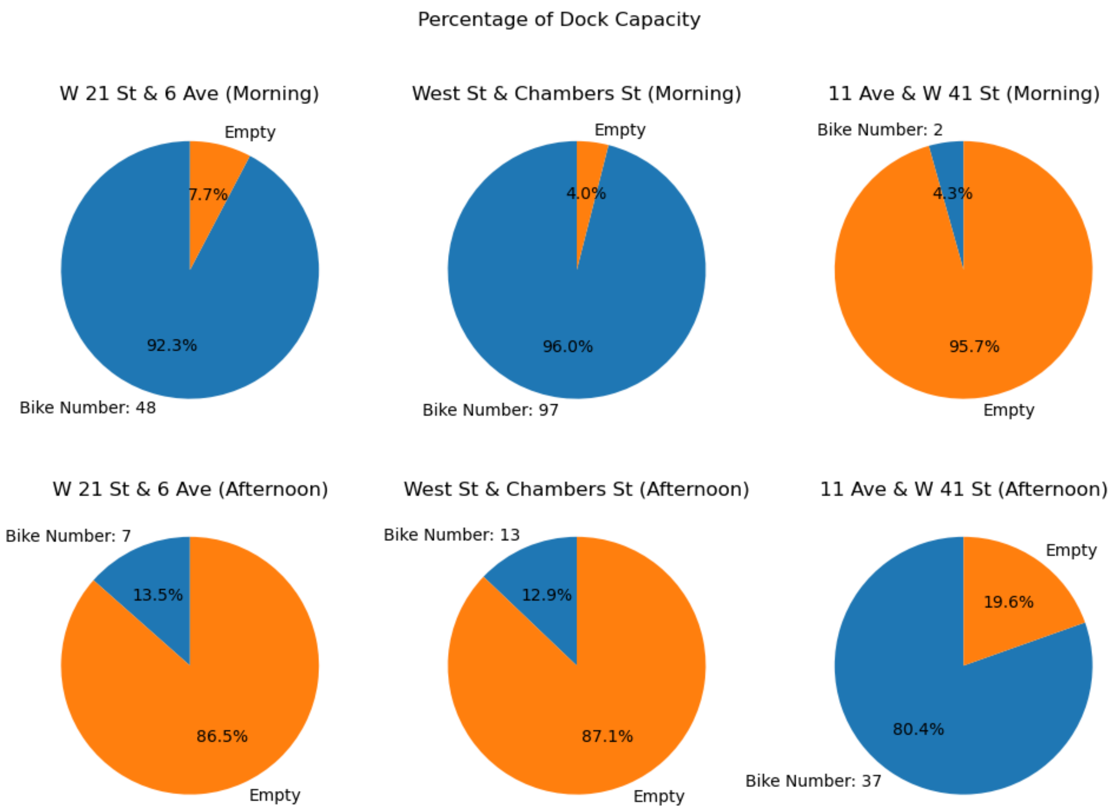

Figure 5: Percentage of Dock Capacity

2.4 Discussion

Three main NYC CitiBike stations are studied in this project to find the to estimate the steady state of the number of available bikes using Markov chains. By calculating the stationary distribution, the distinct patterns observed at each station during different times of the day can be found, which reflect the dynamic nature of urban mobility in New York City.

2.4.1 Insights and Suggestions

At W 21 St & 6 Ave station, there is a noticeable trend of many bikes in the morning but significantly fewer in the afternoon. This suggests that the station is primarily used by office workers who commute to this area in the morning and leave in the evening. To address this pattern, CitiBike should focus on redistributing bikes to this station later in the day, involving moving bikes from stations that are less busy in the afternoon, ensuring that there are enough bikes available for commuting period. Additionally, CitiBike could consider offering incentives to encourage users to return bikes to this station in the afternoon.

For the West St & Chambers St station, the pattern of being nearly full in the morning and then experiencing a decrease in the afternoon indicates it serves a dual purpose. It appears to be a popular destination in the morning and a starting point for evening commutes. This could be due to its proximity to commercial zones and tourist attractions. In response, CitiBike should implement a flexible strategy for bike rebalancing. In the mornings, bikes could be moved from this station to other locations where there is a shortage. Conversely, in the evening, ensuring enough bikes at this station would support the needs of commuters starting their journeys from there.

The contrasting bike availability at 11 Ave & W 41 St station, with many bikes being taken in the morning and returned in the evening, indicates a pattern typical of residential areas. To cater to the morning demand in this area, CitiBike should ensure a larger number of bikes are available at this station at the start of the day. This could be achieved by redistributing bikes from stations that have an excess in the morning, thereby meeting the high demand for bikes from morning commuters in this area.

2.4.2 Discussion of Existing Issues and Future Research Prospects

Our study gives us important information, but it also has some limits. First, we only look at data from July 2023. This means we don't know if the patterns we see change in other months or during special events. Second, we only study three bike stations. This might not show us what is happening in all of New York City because other stations could be different. Lastly, we don't think about things like the weather, problems with buses or trains, or new buildings being made. These things can change how people use bikes but we didn't include them in our study.

For the next steps in research, we can do many things to learn more from our study. We can look at bike data from different times of the year and from past years too. This can help us see if the way people use bikes changes with the seasons or over time. We can also study more bike stations in different parts of the city, which will give us a better idea of how all people in the city use bikes. Another good idea is to think about things like the weather, if buses and trains are working, and big events in the city. These can change how people use bikes. Making models to guess how many bikes will be needed at different times and places can also help. Lastly, we can ask people why they use CitiBike and what they like or don't like about it. This will help us understand more about why people use bikes.

2.5 Conclusion

Our analysis of the stationary distribution of bikes at three key CitiBike stations in New York City reveals distinct usage patterns and station utilization dynamics. The station at W 21 St & 6 Ave and West St & Chambers St witness a high turnover, indicating its popularity among morning commuters and afternoon starters. On the other hand, the station at 11 Ave & W 41 St experiences high demand in the morning, possibly due to nearby residential areas, with a return flow of bikes in the afternoon. These insights offer valuable information for optimizing bike logistics management. Implementing strategies like reallocating bikes, monitoring stations and promptly increasing supply at rush hour can enhance the efficiency of bike distribution, better meeting the diverse needs of commuters throughout the day.