Transportation Around New York City: Assessing the Use of the Citi Bike System Alongside the Taxi System

Sally Stahl
sastahl28@gmail.com
Department of Computer Science
Furman University
sally.stahl@furman.edu

Abstract

New York City has an extensive taxi and bike share system in place. Increasing the usage of bikes in this large city would help to reduce carbon emissions and could help to reduce the negative environmental impacts that this city causes. This project explores the relationship between these two transportation systems. By looking at the use of the patterns of use of the bike share system in 2021, and comparing it to the behavior seen in the taxi system in the same year, we can assess ways that these two systems are currently being used. Once we understand this, then there is opportunity to find areas where a taxi ride could be substituted with a bike and improve the system in place.

1 Introduction

New York City is the most densely populated city in the United States and the population is still increasing [1]. The city is known worldwide as a hotspot for opportunity and a popular tourist location year round. With so many people existing in one area, there is a cost that the environment must pay. In 2016 New York City did a study called The Inventory of Greenhouse Gas Emissions, where they quantified how much CO2 was being released and what the source was. From that study it was found that 60 million ton of carbon dioxide was released that year. Of that, "30% of citywide emissions come from the transportation sector, and 95% of that comes from on-road vehicles." [2] It is evident that the transportation around the city has a negative impact on the environment since such a large portion of emissions are coming from vehicles in the city. While the "total emissions of the city has been decreasing in the past 5 years," [3] it is important to note that the city has a goal of "reducing emissions from all sectors by 2050." [4] In order for this goal to happen, there must be changes to the current system. There needs to be less of a reliance on on-road vehicles and we must turn to other forms of transportation that are zero-emission, such as bikes.

This projects explores the possibility and probability of decreasing the use of taxis and increasing the usage of the bike share system in an effort to reduce the CO2 emissions released by the city, more specifically the transportation sector. This project investigates the interaction between the bike share system and the taxi system to identify the most popular locations for bike trips and taxi trips. I am interested to see if there is overlap in popular taxi trip routes and if there are bike stations that exist there already. The motivation for this project comes from an interest in reducing human impact on the environment. New York City has such a dense population and it has the benefit of being laid out in a way that important locations are close to each other. There is so much potential to decrease the environmental impact from this city. Increasing the reliance on bikes and decreasing the use of taxis would be beneficial for the environment and would decrease the emissions released by on-road vehicles.

The data looks at the taxi trips and bike rides collected from the 2021 time frame. This data is used to compare popular locations of the taxi trips and the bike share system and visualizing the popularity rates of both of these transportation systems. After calculating the distance of trips of the bike rides, the average distance of a bike find is found and compared to the distances of the collection of taxi

trips to investigate how many taxi rides were taken that are of equal distance or less than an average bike ride.

2 Related Work

Looking at New York City, you can tell that the city is reliant on transportation of all forms; personal cars, rideshare system, bike, subways, buses, and many types. The city has been analyzed in many different ways but this project is interested in the transportation, emissions, and bikeshare systems in this city and others like it.

There is no lack of research that proves that there are major pollutants coming from the transportation sector in New York City and this specific sector is responsible for 90% of total emissions. There have been in-depth looks into many of the major countries in the world to assess their city emissions, as well as some specific cities. [5] [2]. There are evident side effects that come from this behavior and on the use of cars alone there has been attempts at optimizing the flow of traffic in New York City. Using taxis, there has been done and finding problematic areas of flow in the city in order to increase the number of people who can efficiently travel around the busy city. [6]

When looking at the environmental impact of New York City emissions, there is a wide variety of areas that have been researched already about this. There is evidence of this poor air quality effecting children in schools between the hours in the day where they are outside and this has put them at greater risk for elevated airway inflammation. [7] The increase of black carbon that is being inhaled is proven and When traveling to and from school and other times spent outside, this is negatively impacting their health.

When looking at the pollution in the city itself, researchers identified other problematic particles that are appearing in the air in the city, in addition to carbon. There are increased levels of copper, iron, and calcium found on the road as a result of different of the cars such as brake wear. [8] [9] It is important that there is research to other contributing factors in addition to carbon.

There is obvious motivation to cut down on the use of cars, but how can this be done? The answer lies in a system that is already in place in New York City; bikes. This is an area of study that is being viewed as an impressive alternative to cars around dense areas. There has been evident proof that bikes can be an efficient replacement for cars and there have been studies done on cities around the world such as London, Brisbane, Washington D.C. and Melbourne, Australia. [10] In 2012, there was an estimated value of 115,826 km less car driving in Melbourne due to the replacement of bikes in the city center. This rate is even higher in London, England where there was an estimated amount of 632,840 km saved from the bikeshare system. This speaks volumes to the efficiency of the bikeshare system being put in cities with high population density and locations that are in close proximity to each other. This will motivate further research into areas like this and bikes around city areas.

It must be understood though, that the bikeshare system in place is not perfect and there are problems with the distribution of bikes around the cities as well as the unsafe routes that the bikers must take to get around the city as well. [11] [10] These are two areas that need solutions and are being looked into, especially in NYC. There needs to be a better understanding of the spatiotemporal patterns of the city to provide us with some perspectives and insights for improving urban planning, transport efficiency, and energy conservation and emission reduction. [12] Understanding how the city functions is providing some information on how to better fit the city for transportation.

One of these was looking at the popular destination dropoff points of bikes. [13] It was stations like these that saw problems such as having too many bikes being docked in some locations and other stations that did not have enough bikes to meet the demand. This problem motivated new programs to work on bike distribution and then triggered research into these programs. There has been a system tested out to promote even distribution of bikes around the city called the Bike Angels system. [14] This is a program that has taken initiative to optimize the bikeshare system and is taking steps to allow more people to use these opportunities. When studied, there was a notable success of having people relocate bikes, and finding the threshold of a reward that will ensure that people are actually doing what needs to be done to make the system work.

This project should highlight the possibility of increasing the use of bikes around New York City and see how big of a change can happen with just choosing to grab a bike instead of hailing a taxi.

The city is motivated to make changes as a whole, and this is just one of many ways to contribute to the goal of reducing emissions. Research has been done about New York City, bikeshare systems around the world, and the carbon emissions in various cities. This project finds the intersection point between these three topics.

3 Data

Two databases were used for the project; one that tracks the taxi rides in the New York City area and one that tracks the rides taken using the bike share system in the New York City area. Each were chosen using the year 2021 to compare data from the same year. A one-year time frame was chosen in order to account for all seasons of the year and get a larger set of taxi and bike rides. Both data sets were grouped by month so there are 12 individual sets for each month for bikes and the taxis. The sets start in January 2021 and end in December of 2021.

3.1 Taxi Data

The first data set is found on the New York City Taxi and Limousine Commission (TLC) website [15]. TLC has been receiving taxi trip data since 2009, but this data wasn't publicly available until 2015 and was released through the Open Data portal. This data set contains information on trips taken in green and yellow taxis since January of 2009. The data is organized by month and each file contains information about all taxi trips taken in that month. From all 12 datasets from 2021, there were about 31,000,000 trips that were tracked. Each trip has the pick up and drop off date and time, we well as the total distance of the trip. In addition, the data provides the zone that the taxi trip started and ended in. The city is divided into different zones and these zones are roughly based on NYC Department of City Planning's Neighborhood Tabulation Areas (NTAs) and are meant to approximate neighborhoods within the city. The data shows what neighborhood the passenger/s were picked up and dropped off in. This data set does not include the exact latitude and longitude points of where the trip started and ended. In addition, the datetime of starting and ending the trip, how many passengers were in the car, the distance of the trip, and a lot of information about the cost of the ride with information about extra fares, taxes, and toll costs. There were a few miscellaneous rows that included information from different years that had to be filtered out, but the dataset was easy to work with after removing those rows.

Taxi Ride Data					
PU Datetime	DO Datetime	Passenger Count	Trip Distance	PU Location	DO Location
2022-01-01 00:35:40	2022-01-01 00:53:29	2.0	3.80	142	236
2022-01-01 00:33:43	2022-01-01 00:42:07	1.0	2.10	236	42
2022-01-01 00:53:21	2022-01-01 01:02:19	1.0	0.97	166	166
2022-01-01 00:25:21	2022-01-01 00:35:23	1.0	1.09	114	68
2022-01-01 00:36:48	2022-01-01 01:14:20	1.0	4.30	68	163

Table 1: Taxi Ride Data Table

With the total distance of the trips provided, it was easy to find the total miles driven by the taxis each day by summing the total distance and plotting it below. This helps to understand the trends in popularity in taxi rides during the year. As you can see below, it is evenly spread with some specific out liar dates. There is slight increase during May and June and another slight increase in November and December. In all though, there is a consistent total distance driven each day in 2021.

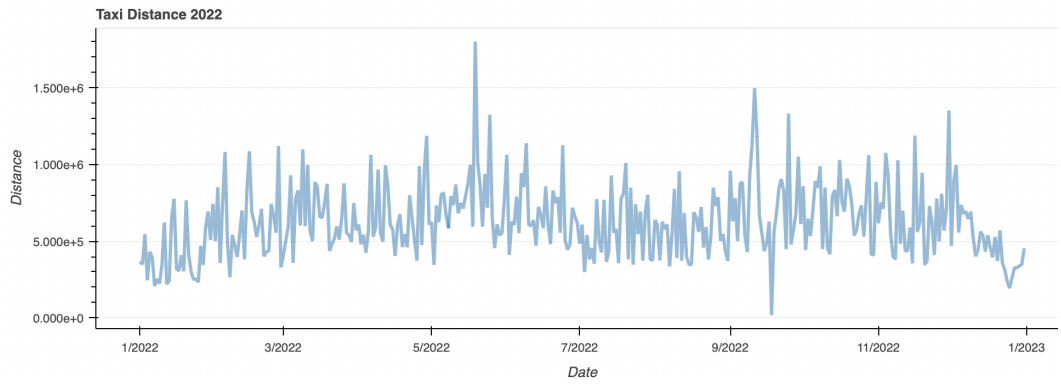

Figure 1: The total miles travelled in each day in 2021 plotted for the whole year to notice overall usage trends.

3.2 Bike Data

The second data set used is from the Citi Bikeshare system website, specifically for the New York City area [16]. The data was started to be collected in July, 2013 and is organized by month up to August, 2023. Each files contains every instance of a bike being taken out of a docking station and returned. This database contain the latitude and longitude of the starting and ending station. There were about 27,600,000 rides that were tracked in 2021. The fields for this table included the type of bike, the starting and ending datetime, the start and end station name and ID, if the user is a member of the Citi Bike system, and the starting and ending latitude and longitude points of the station. The fields that were mainly used in this project were the starting and ending station name and the geographical coordinates of the stations. There were rows where the starting and ending location were the same. The data has already had removed any instance of a "mistake use" where the bike was taken out of the dock and returned by accident or was put back in the same location within a short time frame. If the bike was taken out, ridden for miles, and then returned to the same place, then there is no way to differentiate this between a mistake use because the distance of the trip is not recorded. Thus, the rows that had the same starting and ending station had to be removed since they did not provide information about the actual distance of the trips taken.

Bike Ride Data					
Rideable Type	Started at	Ended at	Start Station	Start Lat	Start Long
electric bike	2022-01-26 18:50:39	2022-01-26 18:51:53	12 St & Sinatra Dr N	40.750604	-74.024020
classic bike	2022-01-28 13:14:07	2022-01-28 13:20:23	Essex Light Rail	40.712774	-74.036486
electric bike	2022-10-18 09:12:12	2022-10-18 09:18:57	McGinley Square	40.725340	-74.067622
classic bike	2022-10-02 21:44:01	2022-10-02 21:45:57	Christ Hospital	40.734786	-74.050444

Table 2: Ride Ride Data Table

Another problem that had to be accounted for was the geographic range that was recorded. This dataset also included trips that started and ended in New Jersey. A popular bike trip is going over the George Washington Bridge to cross between states. Many bike rides were crossing the state line where it would start in New Jersey and end in New York or start in New York and end in New Jersey. For the sake of this project, New Jersey was not of interest so points that started and in New Jersey were removed. The two charts below show the points before and after the locations that fall exclusively in New Jersey territory were excluded.

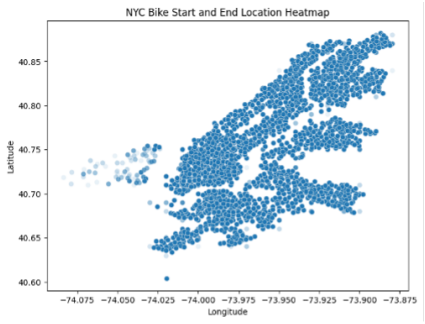

Figure 2: All bike trips taken in 2021 including points in New Jersey.

Figure 3: All bike trips, excluding points in New Jersey, on map of New York City

4 Methods

Changes were made to the data since there was the differing information that was being provided. In order to work with the data, the information had to be compatible to compare these systems. The goal was to do some comparative images between the taxi and bike datasets to get a better understanding of the current uses of the system and identify ways to improve it. But to do this, some columns had to be added and some calculations needed to be done using the information that was provided.

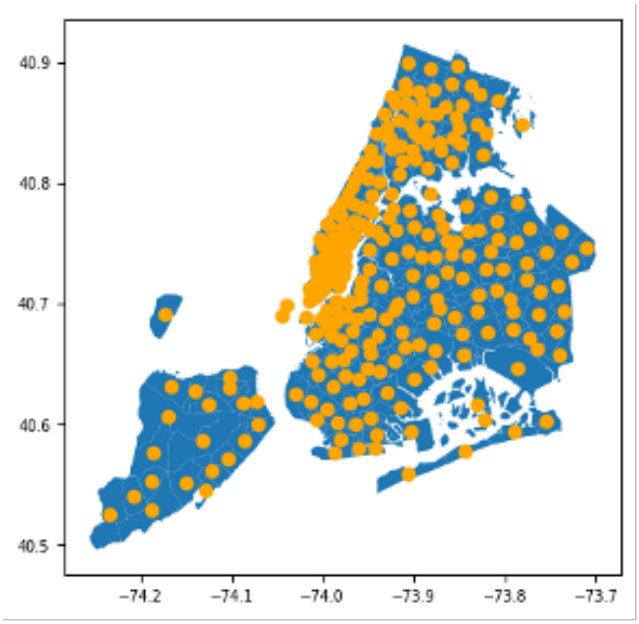

Figure 4: Taxi Zones of NYC With the center point that were used to assign latitude and longitude locations with all taxi trips based on the zone location.

The taxi data provides the distance of the trips and the areas that the trips started and ended, but there are no coordinate points for the zones. Using a GeoJson file of the taxi zones in NYC, the center of each of the zones was found with the corresponding geographical coordinates to make a rough estimation as to where the trips started and ended.

The goal was to have the latitude and longitude points to be able to map these along with the bike share starting and stopping points. From here, the top ten most popular pick up and drop off pairs were mapped for the bike share system and the taxi system, resulting in Figure 5. It should be noted that within the top 10 most popular taxi trips there was a path that started at Midtown Center and ended at the John F. Kennedy Airport. This point was excluded since biking to the airport is not a possibility.

The bike starting points are in blue and the stopping points are the red dots on the map. The green dots show the start of the taxi trips and the yellow points are the ending taxi trips. Evidently, the areas that taxis are very popular does not overlap with the places that the bike share system is heavily used. While there are some longer trips that are being used regularly by the taxis, there are also very short trips that occur frequently that could easily be replaced with a bike. This showed that the places that the bike share system is being utilized heavily is different than the areas that the taxi system is being used heavily and there are areas of improvement.

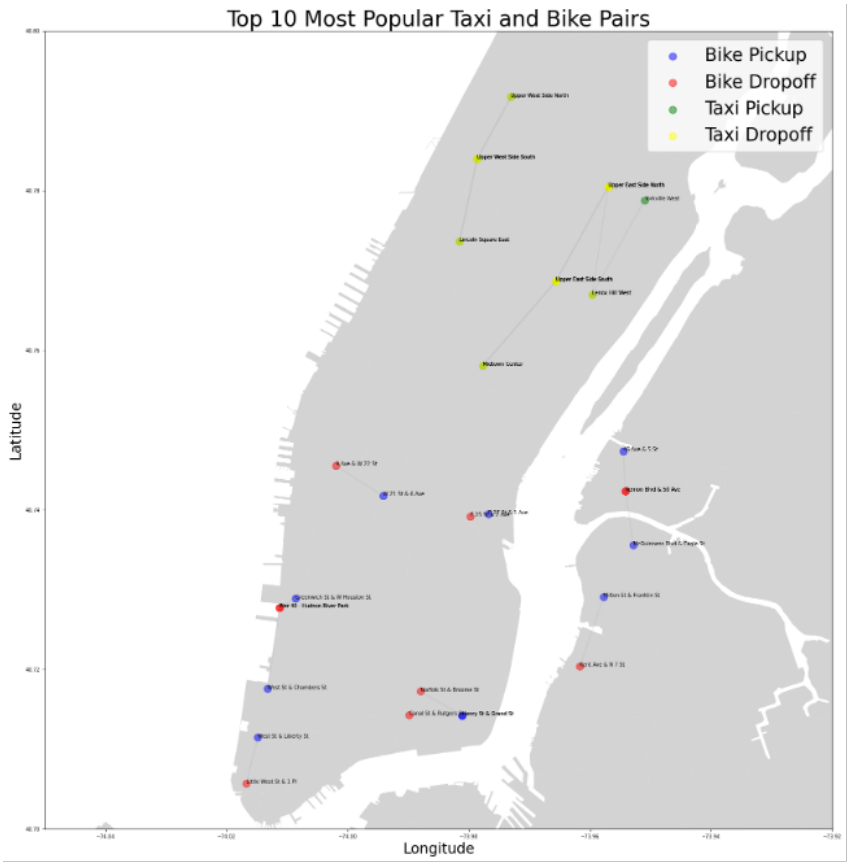

Figure 5: Map of top 10 popular bike start and stop points, and connected by a line, mapped on New York City.

The next problem that had to be solved was the fact that the bike data did not track how far the bike ride was. With the starting and ending geographical coordinates, the straight line distance between the points could not be used because this would not have resulted in an accurate distance of the ride. Instead, the Manhattan distance was used. This is the distance between two coordinates that is measured along the right angels between the points. In other words, given a right triangle, the distance along the hypotenuse is not used, but instead are using the distance along the other two edges that make up the triangle is the distance between the points. This would gives a more accurate distance to account for the layout of the city which is organized in a block system.

The distance needed to be in miles to be able to compare this with the distance provided in the taxi data. To do this, a method was used that took the Manhattan Distance and converted it to be miles using conversion factors that roughly accounted for how many miles per degree of latitude and

longitude. These conversion values are rough estimations, so these distances are not exactly accurate. But this method was able to produce a distance, in miles, that could be assigned to each trip on the bike that could be worked with for now. A new column was added to the bike dataframe that had the rough estimations of the distances of each trip. Once this point was reached, data frames then had comparable information about location and distances of both taxi trips and the bike rides.

5 Results

The first thing that was looked into was the popularity of these systems throughout the one year time period. By looking at the total users of bikes and the number of passengers in taxis each month, Figure 6 plots this data.

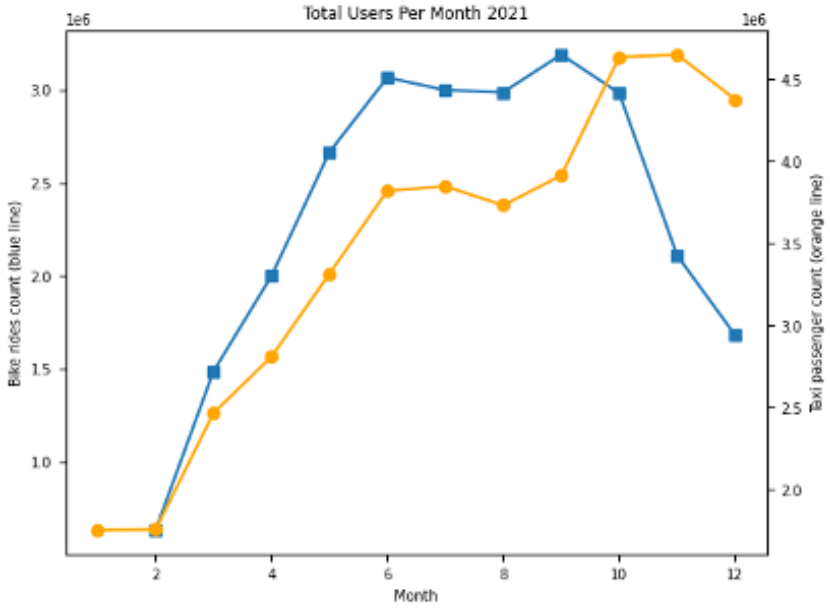

Figure 6: Total tike and taxi usage per month throughout 2021

At the beginning of the year, there is lower quantities of users for both taxis and of bikes. The following months show consistent increase in both taxi users and bike users, with bike being the majority. The rates of bikes being used follow suit with the patterns of warm weather, as the temperature increase, so does the users of the bike share system. The outdoor conditions should be considered when inspecting the usage of bikes in the city. The numerical gap in the total taxi users and bike users increases during June, July, August, and September shows that the low levels of bike usage in the winter months could be due to the weather. This could also be attributed to the high rates of tourism that comes with the late spring and summer months. There are more people who are travelling around the city and seeing similar attractions and so there is more traffic going through the city. These people who are visiting could be opting to use a different form of transportation other than a car. These trends flip in October when the weather can start to become unpredictable. It can be extremely cold during October, November, and December. It makes sense that the people are opting to use a taxi as compared to a bike during the cold weather. This sharp decrease shows that there is good use of the bike system during the months where it is easy to be outside and riding a bike.

When looking at the most popular starting and ending locations for the taxi system, there was not much overlap between these areas where the bikeshare system was popular. The taxis are quite popular in the Manhattan area and the bikes seem to be a key transportation option on the lower side of Manhattan. This showed that there are places where the bike share system is being utilized, but this also showed that there are many places where the bike share system is not as popular and its use should continue to be encouraged.

Using the distances of the bike rides, the average distance of a bike ride was found. This number ended up being 1.557 miles which rounded down to be 1.5 miles. This distance provides a reasonable distance of a taxi trip that could have been replaced with a bike instead. Using the 1.5 mile average, the total amount taxi trips that happened in 2021 that were a mile and a half or shorter in distance were found. In total, there were 11,699,913 taxi rides that were a mile and a half or shorter in distance. Using this, the total miles that were driven from these rides was summed, which was 11,082,139, with the idea that this is a total number of miles that could have been ridden on a bike instead of using a taxi. The next step was to calculate the total amount of emissions that could have been saved by replacing and they way to calculate this is with the following formula.

$$\text{CO2 Emissions (kg)} = \left(\frac{TotalMiles}{FuelEfficiency(MPG)}\right) * CO_2/mile$$

$$\text{CO2 Emissions for 2021 (kg)} = \left(\frac{11{,}082{,}139}{27(MPG)}\right) * 2.3kg/mile = 429{,}639 \text{ kg}$$

The NYC Taxis in 2021 could be one of five models of cars, each with a similar fuel economy [17]. We will use the value of the fuel economy of 27 mpg and an emissions rate of 2.3 kg/m which is the most common value of the five models. With these values plugged into the equation, we get a total value of roughly 429,639 kg or approximately 429 metric tons of CO2 that was released as a result of these short taxi trips. With so many of these short trips being taken, you can see how much can be saved by taking a bike ride instead of getting in a cab. By encouraging the use of the bike share system, there could be an evident decrease in the total emissions from just one year.

Using these short taxi trips, the top 10 most popular starting and end points were plotted, all of which fell into the Manhattan area. These were mapped along with all of the bike stations that fell within the area and this figure is shown in Figure 7.

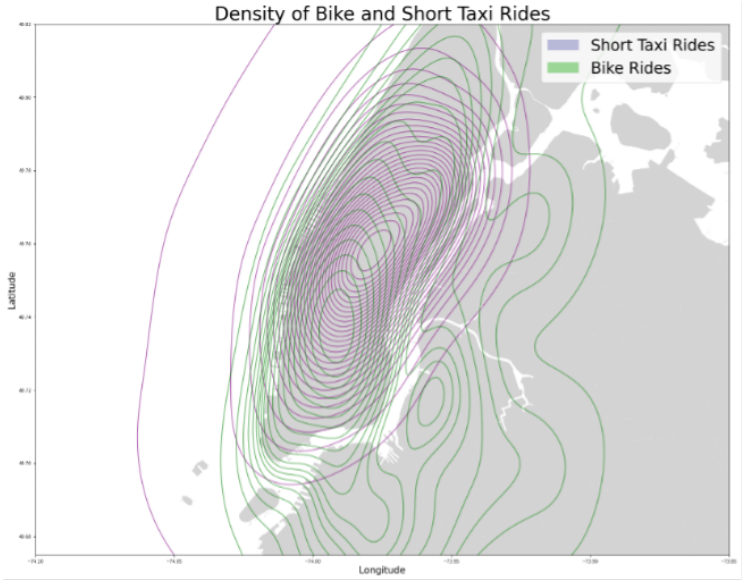

Figure 7: Overall density of bike trips taken and taxi trips that were a mile long or shorter.

When looking at the location of the green circles, there is a high density of trips taken in southern part of the Manhattan borough and there are less stations that can be found in the northeast. There is also quite a few that can be found on the very northwest part of Queens and the northern spot in Brooklyn. When looking at the plotting of the short taxi trips, these trips are centralized a little further north in Manhattan. These taxi trips are happening parallel to Central Park. Central Park is a popular tourist spot and the taxi passengers around here are being pickup at one end of the park and being dropped off at the other end.

This can be interpreted in multiple ways. We can acknowledge the points where all these short taxi rides are happening and see that there are bike share stations that exist all over this part of the city. This is bad as it is showing that this part of the city is relying on the on-road vehicles and instead

of walking through or around the park, people are opting to hail a taxi. But a positive note is that there are a lot of other places around the city where the bikes are being heavily used. Seeing so many bike rides happening where there are less short taxi trips being taken could indicate that people are deciding to check out a bike as a way to see the city and get around more efficiently. The locations of these short taxi trips is very central to this one spot and there are very few places where there are other short taxi trips being taken. This is especially promising since this is a place where there are bike stations already.

Lastly, the total number of bike and taxi rides that have happened in each year from 2014-2021 has been plotted below.

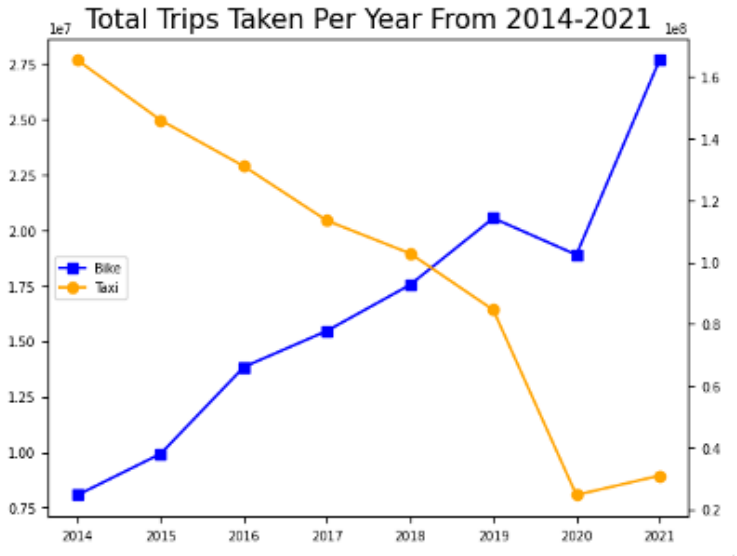

Figure 8: The total number of bike trips and taxi trips that happened each year to show annual trends of users.

The past 8 years have shown some promising trends. There is a steady decrease in taxi rides and a constant increase in bike rides, and a sharp increase seen from 2020 to 2021. As the bike share system is constantly being improved, it is good to see that this system is being utilized. In 2020, there was a sharp decrease in rides due to the pandemic. In 2021, we can see this number start to increase again. The goal would be to keep this number low and not have this rise again to be where it was in 2019. But the decrease in taxi rides and increase in bike rides is showing that the city is seeing some change in terms of users opting for the different transportation systems.

6 Future Work

The main obstacle was that datasets were not completely compatible. The taxi data did not include the exact start and end latitude and longitude points. While there were estimations made about the starting and stopping locations using the center of these sectors, it it not perfect. This would needs to modified by the data collectors though. With the bike data, the opposite problem. The starting and stopping location were given, but the trip distance was not. This caused problem for multiple reasons. Since some bike trips started and ended in the same location, some models excluded trips like this since it gives little insight to start and end at the same spot and not have a value for how far the user travelled, if any. In addition, there had to be estimations as to how far the trip was based on the geographical coordinates given. The distances of the trips are not exactly accurate as a result of this.

This project can be expanded in a few different ways. There is an opportunity to include more years of both taxi and bike data and see if there are changes that can be found over time. Looking at these long-term changes could encourage analysis on how to optimize the bike share system in the city. In addition, there are other, smaller, bikeshare systems in place around New York City and other rideshare systems that could also be analyzed. Lastly, there should be research done on the area

around Central Park to identify ways to encourage the use of bikes and make it safe for those who are wanting to use it. This could be a way to increase the usage of bikes in this part of the city by making sure it is safe for people to be riding bikes.

References

[1] Nigel Harris. Some trends in the evolution of big cities: studies of the usa and india. *Habitat International*, 8(1):7–28, 1984.

[2] The City of New York. Inventory of new york city greenhouse gas emissions in 2016. pages 7–28, 2017.

[3] Daniel L Goldberg, Zifeng Lu, Tomohiro Oda, Lok N Lamsal, Fei Liu, Debora Griffin, Chris A McLinden, Nickolay A Krotkov, Bryan N Duncan, and David G Streets. Exploiting omi no2 satellite observations to infer fossil-fuel co2 emissions from us megacities. *Science of The Total Environment*, 695:133805, 2019.

[4] Umakanta Mishra, Jei-Zheng Wu, and Biswajit Sarkar. A sustainable production-inventory model for a controllable carbon emissions rate under shortages. *Journal of Cleaner Production*, 256:120268, 2020.

[5] Abhinav Pandey, Govind Pandey, and Rajeev Kumar Mishra. Tailpipe emission from petrol driven passenger cars. *Transportation Research Part D: Transport and Environment*, 44:14–29, 2016.

[6] Wahyu Andy Prastyabudi. Finding frequent routes from taxi trips with time windows: Nyc case. In *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, pages 1–5, 2019.

[7] Kyung Hwa Jung, Kathleen E Goodwin, Matthew S Perzanowski, Steven N Chillrud, Frederica P Perera, Rachel L Miller, and Stephanie Lovinsky-Desir. Personal exposure to black carbon at school and levels of fractional exhaled nitric oxide in new york city. *Environmental Health Perspectives*, 129(9):097005, 2021.

[8] Ramona Lall and George D Thurston. Identifying and quantifying transported vs. local sources of new york city pm2. 5 fine particulate matter air pollution. *Atmospheric Environment*, 40:333–346, 2006.

[9] Masha Pitiranggon, Sarah Johnson, Jay Haney, Holger Eisl, and Kazuhiko Ito. Long-term trends in local and transported pm2. 5 pollution in new york city. *Atmospheric Environment*, 248:118238, 2021.

[10] Elliot Fishman, Simon Washington, and Narelle Haworth. Bike share's impact on car use: Evidence from the united states, great britain, and australia. *Transportation research Part D: transport and environment*, 31:13–20, 2014.

[11] Sabreena Anowar, Naveen Eluru, and Marianne Hatzopoulou. Quantifying the value of a clean ride: How far would you bicycle to avoid exposure to traffic-related air pollution? *Transportation research part A: policy and practice*, 105:66–78, 2017.

[12] Chen Xie, Dexin Yu, Xiaoyu Zheng, Zhuorui Wang, and Zhongtai Jiang. Revealing spatiotemporal travel demand and community structure characteristics with taxi trip data: A case study of new york city. *PLoS one*, 16(11):e0259694, 2021.

[13] Andreas Keler, Jukka M Krisp, and Linfang Ding. Extracting commuter-specific destination hotspots from trip destination data–comparing the boro taxi service with citi bike in nyc. *Geo-spatial Information Science*, 23(2):141–152, 2020.

[14] Hangil Chung, Daniel Freund, and David B Shmoys. Bike angels: An analysis of citi bike's incentive program. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 1–9, 2018.

[15] `https://s3.amazonaws.com/tripdata/index.html`. [Online; accessed 14-Sep-2023].

[16] https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page. [Online; accessed 14-Sep-2023].

[17] https://www.nyc.gov/site/tlc/businesses/yellow_cab_hackup.page. [Online; accessed 11-NOV-2023].