Impact of COVID-19 on Human Mobility in NYC

Student Presenters:
Ivan Miller
Yinzi Xiang
Tyron Samaroo
Satesh Ramnath
Professor: Ronak Etemadpour
Visual Analytics DSE I2700

Impact of COVID-19 on Human Mobility in NYC

Ivan Miller, Tyron Samaroo, Satesh Ramnath, Yinzi Xiang

Abstract – In this project we wanted to perform visual analysis of the associations between recorded COVID-19 cases as reported by the NYC Department of Health and Mental Hygiene and the data on subway entries collected by the Metropolitan Transportation Authority and Citi Bike ridership as measurements of human mobility in New York City during the pandemic as well as one full year prior (2019). We chose Python as the main programming language for the project using the Pandas library to perform data cleaning and exploratory data analysis for each dataset and packages such as Matplotlib and Plotly for visualizations. By analyzing subway and Citi Bike ridership over comparable periods of time both before and during the pandemic, as well as visualizing all three datasets in one figure, we concluded that COVID-19 virus must have had a significant impact on human mobility in New York City.

Index Terms – COVID-19, bike sharing, Citi Bike, public transport, subway, MTA, New York City, time-series analysis, data visualization

I. Introduction

New York was one of the first cities to face the outbreak of COVID-19 virus at the beginning of the pandemic. The City with its densely populated neighborhoods and a vast network of public transportation was hit much harder than other places and it quickly became an epicenter of the pandemic in the United States. Several days before WHO declared COVID-19 a global pandemic on March 11, 2020 an unprecedented decrease in subway turnstile entries was recorded in the data collected by the MTA, and the downward trend continued up until mid April 2020 when the first wave of COVID-19 cases had finally plateaued. To assess the impact of pandemic on human mobility in our project, however, we wanted to look beyond the data on the subway only, and use subway usage data in conjunction with the data on public bicycle sharing system Citi Bike as alternative mode of transportation.

From March 1, 2020 when the first COVID-19 case was reported in New York State through the peak of the first wave of infections that occurred around mid April, the use of the subway system has dropped dramatically. As the infection rates subsided several weeks after the city came to almost a complete stop as a result of social distancing measures as well as closure of non-essential businesses, subway ridership has been on an upward trend. By mid-October, a third wave of COVID-19 hit New York and its transportation system with subway stations located throughout the four out of five boroughs of the City: Manhattan, Brooklyn, Queens, and Bronx. It is worth noting though that the use of public transportation in New York City is still susceptible to fluctuations triggered by waves of COVID-19 virus and it has a long road to recover to its pre-pandemic levels as hybrid or fully remote work became a norm for many of the workers.

We started the work on this project that was named impact of COVID-19 on human mobility in NYC with a review of several relevant papers published in recent years and selected three datasets to work with: data on daily COVID-19 cases collected by the NYC Department of Health, data on subway entry and exit swipes provided by MTA, and the data on New York's public bike sharing system Citi Bike in an attempt to establish a connection between different modes of transportation. The following sections lay out the background of the research, data preprocessing and analyses along with the work on the visualizations and the tools we selected to work with.

II. Data Specifications

A. *MTA*

The MTA turnstile data is publicly available on various platforms. However, the data isn't trivially layed out for immediate visualization. On a weekly basis, MTA posts CSV files containing entry data, bus fare and various other forms of ticketing and entry on all of their platforms. Specifically for our use case, we focused on NYC subways so looking at the turnstile entry data suffices.The raw MTA data is odometer-like so the number of entries is a historical value which requires some data cleanup; additionally boroughs, and unique stations identifiers aren't taken into account for each turnstile. Chris Whong, an

Outreach Engineer at Qri.io built a data pipeline for MTA's turnstile entry data, which contains the fields of interest to us; station name, borough, properly aggregated count of entry data per station, etc [6,9]. With the data aggregated to this point, the only form of data transformations that needed to be done were across grouping MTA data by station and by data to get uniquely aggregated values for specific timeline visualizations and geo-mapped averages of MTA turnstile entry data.

B. Citi Bike

Citi Bike data is publicly available online and includes much useful information about riders [7]. Information ranged from 2013 to present-day 2021. Citi Bike provides Citi bikes trip histories consisting of the start time, end time, trips taken, and type of memberships. For the purpose of this project, we focus on data from 2019 and onwards to understand the impact of COVID-19. Citi Bike has its data broken up into months. We collect data between January 2019 and November 2021. In 2021 Citi Bike changed its data format so an adjustment had to be made to collect the aggregated information we wanted to collect. The format before 2021 has 11 columns and the format for 2021 has 13 columns After collecting and cleaning the data there were about 36 million data points. One of the challenges was making sure we collected the right information. To better understand the data, it was grouped by day. Now that the data was grouped by day we were able to analyze the number of trips taken a day, average trip duration, subscribers, and customers.

C. COVID-19

Publicly available data was used to obtain information on registered COVID-19 cases, hospitalizations, and deaths, including breakout of metrics of interest by all five boroughs of New York City [8]. Selected dataset contained one categorical and 66 numerical columns. Each row in the dataset represents a summary of metrics for a given date, with the first record dating back to February 2020. The data is being actively updated on a daily basis by the NYC Department of Health and Mental Hygiene, however, it was limited to make it comparable with MTA and CIti Bike data that is not being updated as frequently. It was not necessary to keep all of the original columns to achieve the goals of the project and 29 numerical columns were removed to improve computational efficiency during the initial part of the analysis. Consequently datasets powering visualizations on the Streamlit dashboard that feature COVID-19 data were further reduced to improve computational efficiency and loading speed.

III. Background and Related Work

A. Citi Bike

This study focuses on using available data on NYC subway and bike-share systems to understand the impact of COVID-19 in New York City. COVID-19 impacted many aspects of society which led to damages to the economic and urban systems of our society. Given the nature of a virus like COVID-19, there is a rise in fear of overcrowding with the current use of subway transportation, making bike sharing a capable solution for transportation.

New York City experienced its first COVID-19 case on March 2nd and ten days later declared a State of Emergency that implemented measures to reduce the spread of the virus. Limiting space occupation, flexible scheduling for city workers, and stay-at-home orders were used to help accomplish this. But even with these measures implemented cases still increased. Given the increase of cases and measures implementing the subway system was heavily impacted. Facing a reduction in demand, COVID-19 cases among its workforce the subway system had no choice but to reduce its operating capacity. With the subway system facing such a dilemma, a program for free trips via bike sharing was introduced. The study plans to understand the shift between subway to bike-sharing transportation given the impact of the COVID-19 outbreak.

City Bike (Bike sharing system) launched in NYC in May 2013 and averages more than one million trips every month. It has over 14,500 bicycles and 890 stations in NYC. Citi Bike has open data available for its station trips history. This data included the start, end time, station used, type of ushered, gender, and year of birth. New York Subway launched in 1904 has an average of 5.5 million weekday trips. It has 472 stations and is available 24 hours a day. There is subway data publicly available but only provides data of the number of entrances and exits per station every four hour. Data for the subway(MTA Turnstile) and bike sharing system(Citi Bike System Data - NYC) was collected for February and March 2019, 2020. The data for COVID-19 cases was retrieved from New York City Health Database (COVID-19) Data for trips that were less than one minute or more than 6 hours were excluded because they were considered outliers.

Impact of COVID-19 on the two modes of transportation was evaluated through a combination of statistical and GIS methods. According to the study Mann-Whitnes U Test was used to look at the difference between the two independent groups by using rank. Ordinary Least Square(OLS) regression was also used to determine the relationship between the dependent variable COVID-19 and the two independent variables. OLS was not used to generate a prediction, instead, it was used to analyze COVID-19 impact on the subway and bike-sharing system. Also, a network subset was designed using ArcGIS which used a custom buffer to provide a more reliable analysis.

In conclusion, as time went on and COVID-19 was on the rise both the subway and bike sharing system were impacted. In the given time frame trips were reduced from an average of 5.5 million to 500,000 making that a total of 90% reduction of usage. The average Citi Bike daily ridership also faces a reduction from peak usage of 60,000 trips to 15,000 trips; it faced a total of 71% reduction of usage. Noticeably, for Citi Bike daily ridership increased when subway ridership decreased. Even though COVID-19 cases increased there was an average 44% increase in trip duration. To better understand this change, daily trip duration in the months of February 2020 and March 2019 was explored. Then, statistical testing was done to determine if COVID-19 impact on City Bike trips in March 2020 would be statistically significantly different from the control months(February 2020 and March 2019). The Study proved that COVID-19 had an impact on both the New York City subway and bike-sharing system. Even though both systems faced a decrease there was statistical evidence that the bike-sharing system is a suitable solution for transportation during a pandemic such as COVID-19.

B. Transportation Impacts

The COVID-19 pandemic and the shutdown of non-essential business transformed mobility in, through and around New York City. This report provides a detailed analysis of the way in which the transportation systems in New York City and the surrounding region were affected by the pandemic and curtailed economic activity through May 31, 2020.

Prior to the COVID-19 pandemic, New York City Transit carried 5.5 million subway riders on a typical weekday, or 2.5 times the total ridership of all other U.S. subway systems combined. On April 12th, 2020, subway ridership had dropped 96% to 213,424 its lowest point during the pandemic, and likely the lowest number in 100 years.This paper is designed to compare panel participants' behavior in spring 2019 to their behavior after the COVID-19 outbreak from spring 2020 through fall 2020, through this to understand the impacts of COVID-9 on city wide travel behavior. According to the data which was collected from the panel participants' behavior. This paper special analyzes the people's multiple behavior change after COVID-19 compared before COVID-19. Especially people's mobility. Inside the paper, the authors analyze the data and detail how the people choose public transportation. Like after the COVID-19, 76% of Manhattan residents prefer to walk to run an errand, while 74% of Staten Island residents used a household vehicle. We can see COVID-19 changing people's mobility and behavior a lot.

According to the above scholars' reports for the impact of COVID-19 on transportation in NYC and our life environment, we can see that there are a lot of things that changed during the COVID-19 pandemic including transportation. Especially in NYC, most of the residents rely on public transportation for working, going outside and running errands etc. So we are very curious how the COVID-19 impacts the public transportation in NYC? Did the vaccination rate change the pattern of NYC people's transportation? After more and more people get vaccinated, do the daily COVID-19 cases affecting the transportation are the same as before the people get vaccinated? Let's explore and get our answer.

C. Correlation Turnstile Entries with COVID-19

The research paper *Correlation of Subway Turnstile Entries and COVID-19 Incidence and Deaths in New York City* focused primarily on observing a correlation between the COVID-19 death cases and the amount of turnstile entries in the NYC subways system. The datasets used in this paper are the NYC Department of Health (DOH) dataset and the MTA's

Turnstile Usage dataset. The rationale behind analyzing the relationship between turnstile usage and the spread of COVID-19 is due to the fact that NYC subways are an enclosed environment and is one of the main sources of transportation in New York City. This paper states that while there is a pattern observed, there is no direct correlation between riding in the NYC subway system and getting infected with COVID-19. This follows the main idea that correlation does not imply causation; while there may be an observable relationship between taking the subway and COVID-19 deaths/cases following the same pattern does not mean someone will get COVID-19.

A correlation between two different Feature sets was compared against COVID-19 cases and COVID-19 deaths. The number of deaths is more correlated than the number of cases; this relationship is observed to be stronger since there were significantly higher numbers of cases than deaths during the timeline at the beginning of the pandemic. Hence, since there were such a large number of cases in relationship to deaths, there is less variance which makes the correlation stronger. Furthermore, a similar regression analysis was broken down further by each borough. The conclusion drawn on each borough did not show any significance of one borough having a stronger correlation than the rest, however the general idea of each borough having a correlation between COVID-19 cases and deaths and turnstile activity was observed. The accuracy of all the regressions observed were measured using R-Squared and MSE; these metrics are commonly utilized to observe how strong a regression is and how much it varies. Due to a large amount of reported cases at different boroughs in the beginning of the pandemic there were drastic differences in the variances on the number of cases in each borough, which explains why each borough's correlation varied differently. However, the general trend of each borough compared to the trend of New York City as a whole still applied which proves there was a significant correlation.

D. Socioeconomic Disparities in Subway Use and COVID-19

In the paper titled "Socioeconomic Disparities in Subway Use and COVID-19 Outcomes in New York City" that was published in the American Journal of Epidemiology in December 2020 the researchers used three different sources of data to to explore the associations between human mobility, sociodemographic factors, and COVID-19 incidence and to assess whether the impact of COVID-19 was significantly different between boroughs. Weekly subway ridership data from MTA and daily COVID-19 cases in 4 NYC boroughs (Staten Island was excluded as there's no subway there) were used in the research. The data was aggregated on borough and ZIP code (ZCTA) level and only 124 out 214 NYC Zip codes were used for the final analyses, as those that had no subway stations or population recorded (according to the American Community Survey data) were filtered out.

The researchers used R to work with the MTA data , in particular, ggmap package was utilized to geocode stations by borough and change in use of subway in each individual ZIP code was calculated based on the following formula using standardized data: (weekly ridership after lockdown – weekly pre-lockdown ridership) / (weekly pre-shutdown standard deviation). Several metrics were derived from the NYC Department of Health's COVID-19 dataset: log of the cumulative case counts was used to assess the timing of the exponential growth for the number of cases reported. The researchers also utilized three additional metrics: 1)rate of positive COVID-19 cases per 100,000 population; 2) percentage of positive tests out of the total number of tests; 3) the rate of total tests per 100,000 population. Demographic data from the US Census Bureau (population size, income and education levels, age, race etc) was used to derive socioeconomic insights.

The study reported on substantial social distancing inequities throughout NYC neighborhoods.
The ZCTAs with the highest median income tended to have the greatest decrease in mobility while the communities with smaller decreases in mobility were more likely to be socioeconomically disadvantaged, have higher population of people of color, and have a higher share of essential workers, indicating that the ability to shelter in place for marginalized communities was lower. However, those associations did not remain when the analyses were adjusted for the percentage of the population employed in essential services, suggesting that disparities in mobility reductions were driven by essential work and reduced ability to social distance because of that.Also, the research determined that an increase in median individual income of $10,000 was associated with approximately a 9% decrease in COVID-19 rates. Additionally, an increase of 10% of the ZCTA

population determined to be an essential services worker was associated with a nearly 2-fold increase in the COVID-19 positivity rates per 100,000 population.

IV. Dataset That Has Been Used

A. MTA

1) Geomap

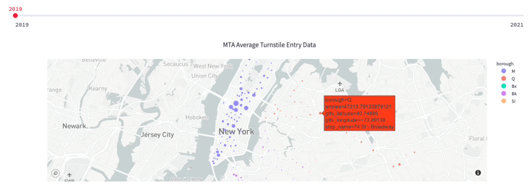

Figure 1. MTA Geomap

Figure 1 is a Scatterplot Mapbox designed using Plotly. Each scattered point represents an MTA train station and its entry data for a given year. At the top of the plot is a slider that is default located at year 2019; it can be switched to years 2020 and 2021 to update the respective data points for that year. The points are categorized by borough through color, as can be seen by the legend on the upper right of the visualization. When a user utilizes the dashboard and hovers over a particular point, a tooltip is presented with the stop name, its latitude/longitude, the borough, and average number of entries.

2) Normalize MTA Data

We collected the 2019, 2020, 2021 three years MTA subway dataset. After cleaning the data, this dataset has 450,242 rows and 12 columns. The columns include stop station name, daytime routes, division, line, New York City borough, structure, longitude, latitude, complex ID, date, entries and exits. In our whole project, we mainly used the borough, longitude, latitude, date, and entries for the subway analysis during the pandemic.

B. Citi Bike

Figure 2 below shows total ridership by year for both MTA and Citi Bike. Before the pandemic MTA averaged about 1.5B trips a year. We see that the number of trips made dropped from 1.7B to 617M trips in 2020 and also dropped to 528M in 2021. Citi Bike ridership is a lot less in retrospect to the MTA. Cibibike averaged about 20M riders a year before the pandemic, this trend did not actually change but increased during the pandemic.

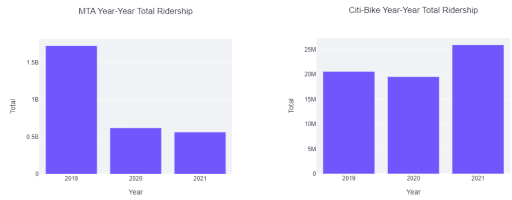

Figure 2. Year-Year Analysis on Citi Bike and MTA Turnstile Entries

Figure 3 below is a visualization of the Citbike data with key indicators; trips, and the number of subscribers and customers by day . This visual is a bit noisy so a 30-day rolling average was used to improve interoperability as shown on the right. We can see that Citi Bike data has seasonality. This makes sense because people tend to ride bikes during warmer weather. If we focus on the warmer months we see that the number of trips taken increased in 2020 and significantly increased in 2021.

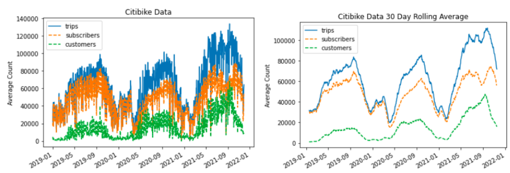

Figure 3. Rolling Average on Citi Bike Data

C. COVID

1) Time Series

While working with the COVID-19 data we needed to visually represent changes in data over time and the below time series chart visualizing confirmed cases by borough helped us identify four periods of exponential growth (so-called waves), as well as assess the difference in the severity of pandemic by borough (see Fig. 4). With different colors representing different boroughs we can see, for example, that during the second and third waves of COVID-19 from November 2020 through April 2021 lines representing Brooklyn and Queens were moving alongside each other while other boroughs showed different trends and in general had much lower numbers of reported cases.

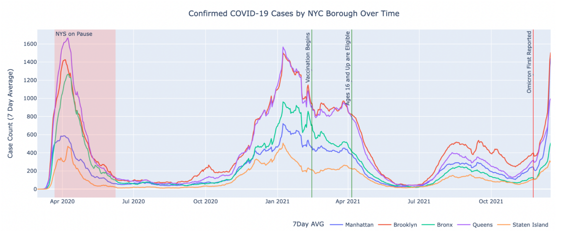

Figure 4. Time series chart visualizing 7-day moving average of cases by borough. Vertical lines represent meaningful events of interest.

After COVID-19, CDC, governors and news media are telling people to keep social distance, wear face masks and get vaccines to keep safe from COVID-19. As an ordinary person, sometimes I would doubt whether those mandates really work. As a data science student, I collected the data and analyzed it. Let the data tell us the truth.

2) Normalized Rolling Average of Cases

To address the difference in scale of the data across MTA, Citi Bike, and COVID-19 datasets, all data was transformed into a range [0,1] using MinMaxScaler from the scikit-learn package. In addition, 7-, 14-, and 30-day aggregation windows were explored while trying to achieve the best results when plotting time series charts with all three datasets plotted in one figure in order to find the right balance between smoothing over minor fluctuations while preserving the level of detail necessary to identify trends in data. We concluded that a 30-day rolling average was the option best suited for the task (see Fig. 5 for a comparison).

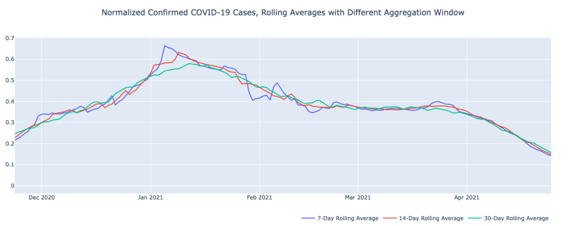

Figure 5. Moving average of reported COVID-19 cases with different aggregating windows shown, normalized data. The 30-day moving average was selected for visualization on the Streamlit.IO dashboard together with MTA and Citi Bike data.

3) Stacked Bar Chart for Monthly Cases by Borough

We utilized the Plotly package in Python for most of the visualizations because of its wide range of functions and the fact that plots could be interacted with and changed by users directly on the dashboard without any coding. See Figure 6 for an example of a stacked bar chart representing registered COVID-19 cases with the data aggregated by months and ability to single out just one borough or explore any combination of boroughs which was extremely useful for comparisons during the research stage of the project.

Figure 6. Monthly COVID-19 cases by NYC borough: A) Manhattan, B) Brooklyn, C) Bronx, D) Stacked bar chart visualizing all boroughs, E) Queens, F) Staten Island, G) "Zoom in" feature of Plotly library highlighting COVID-19 cases by borough from November 2020 through April 2021

V. Contributions

All team-members collaboratively worked on the principal component analysis performed on the MTA dataset.

Tyron

Tyron focused on working on the Citi Bike dataset. This included performing EDA and producing visualization related to Citi Bike relating to trips and the type of customers. Additionally, Tyron collaborated in putting together the visual dashboard utilizing Streamlit.IO

Satesh

Satesh focused primarily on working on the MTA dataset. This included performing EDA and producing visualizations related to MTA ridership data. Additionally, Satesh collaborated in putting together the visual dashboard utilizing Streamlit.IO.

Yinzi

Mainly working on the PCA/ Viz on the line plot with Critical Events, collecting some related dataset for events and mandates analysis. EDA on Citi Bike dataset , MTA dataset and COVID-19 dataset to get know all the dataset and combine the mandated information to analyze what's the relationship between COVID-19 , subway, Citi Bike and the NYC

COVID-19 related mandates. I made the visualization and tried to let the data tell us the story about the relationship between the actions or mandates the government took and COVID-19 situation.

D. Ivan

Ivan mainly worked with the COVID-19 data, from the research of multiple publicly available datasets, data cleaning, and exploratory data analysis to work on the visualizations and assistance in their implementation on the Streamlit.IO dashboard.

VI. Methods

A. Libraries

Python was the primary programming language of choice to design the dashboard. Noteworthy libraries that were used throughout the course of this project were Matplotlib, Pyplot, Seaborn, and Plotly. Plotly stood out as a major component for this project as it provides multiple levels of interactivity with the user and displayed visualizations on the dashboard.

B. Visualization Tool: Streamlit.IO

Streamlit.IO is an open source Python library that allows for a feasible platform to host a plethora of Python library visualizations. Through connecting a Python script alongside dependencies with a Github account, Streamlit.IO is able to generate an application for a specified project.

C. Principal Component Analysis

Principal Component Analysis, also known as PCA, is a tool commonly used for dimension reduction. PCA is commonly used in EDA to observe which features contribute to the most variance in a dataset. Utilizing this, features that tend to contribute less to variance can be removed from a dataset to reduce computation; ideally the higher dimension data you have the better choice it is to utilize PCA since it would cut down on complex computation.

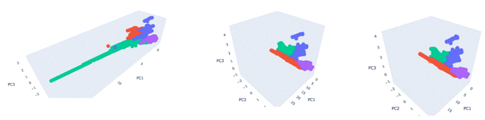

Figure 7. 3D Principal Component Analysis with borough color coded points

From a three dimensional perspective on the MTA dataset from figure 7 above, we can see the variation across each principal component. Each point is color coded by borough, and of all the three principal components, latitude and longitude were taken into consideration. From that point, we can see the variation in those components shrunk as we moved from 2019 to 2020 (left to right). Additionally, it appears that the principal components take on a similar shape in 2021 as we still are in the pandemic and that may be an explanation for low variance in the data.

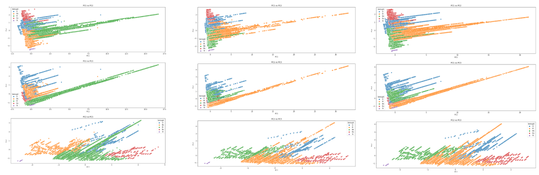

Figure 8. 2D Principal Component Analysis with borough color coded points

From a similar lens from right to left, figure 8 above represents the 3 principal components plotted in 2-D axis. Each row represents one form of plotting two principal components against each-other to get a solid image of the relationship of the two. Moving from left to right across each column represents each year; 2019, 2020 and 2021 respectively.

D. Combining Visualization of MTA, Citi Bike, COVID-19 Data

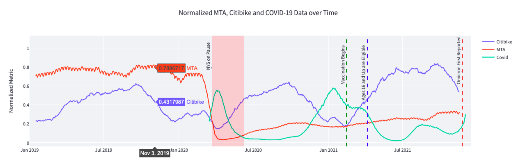

Figure 9. MTA, Citi Bike and COVID-19 Normalized Comparison Timeline

Figure 9 above shows the combined Normalized data on MTA turnstile entries, Citi Bike ridership and COVID-19 cases over the course of 2019-2021. Normalizing the three datasets were pivotal to display the three of these metrics over time as they all had different scales respectively. Utilizing scikit-learn's min-max scaler allowed these three datasets to be normalized to the degree of comparing the overall trend as seen in the figure above. Observing the trends between these three modes of data was important to gain insight in how the behavior of one may have impacted the other and we can see that there are some relationships between them:

As shown on the visualization, subway entries dropped rapidly after a state of emergency was declared in New York State on March 7, 2020 while the COVID-19 cases continued to grow for several weeks. At the same time, Citi Bike ridership followed its usual seasonal trend around the same time and later increased as people shifted from the subway to alternative modes of transportation.

2) In the summer of 2020 with mask wearing mandates and social distancing rules in effect COVID-19 rates remained steadily low even before the vaccine was developed.
3) After vaccines became available in February 2021 the second and the most severe at the time wave of COVID-19 cases in New York began its steady decline through the summer.

VII. Conclusions and Observations

Through the usage of our data and visual analysis, we were able to observe the general trends in MTA turnstile usage and their relationship to COVID-19 cases. Since COVID-19 pandemic began in March 2020, New York's subway system faced a massive reduction in transit entries, however, after it took a dip in the beginning of the pandemic the amount of subway activity has slowly been increasing. Another noteworthy point is subway entries have not in any way resumed activity like pre-pandemic. As for COVID cases and a relationship with MTA like the beginning of the pandemic, the trend seems to come in waves, however cases have been drastically lower upon introduction of vaccines. With the new variant we can see there is the beginning of a new increase in cases yet again.

After analyzing the timeline of social distancing, face mask wearing mandates, and a vaccine rollout we may conclude that those events have strong relationship and impact COVID-19 rates and the usage of transportation options in New York that we covered in this project, namely subway and Citi Bike. As we can see from the previous events and COVID-19 cases analysis, while the NYS on Pause Program was in effect and all non-essential businesses faced mandatory closures, the subway turnstile usage decreased and consequently COVID-19 cases went down as well. Interestingly, Citi Bike ridership was not affected by the initial hit of the pandemic and instead showed its usual for spring seasonal increase in usage right around the time of the first wave, the trend that was later strengthened by growth in membership.

To better understand and compare MTA and Citi Bike data and their relationship with COVID-19 rates, a normalized method was used to scale the data from all three sources in a range between zero and one. This allowed us to represent the datasets that originally had different scales in one visualization and made it easier to perform visual analyses and identify trends over the years thereby allowing us to observe key events such as virus breakout phases, social-distancing restrictions, and vaccination efforts with Citi Bike and MTA ridership. During the years 2019 and 2020 Citi Bike followed a similar pattern when the number of monthly trips remained relatively constant on a year over year basis. We did note that seasonality has an impact on the public bike sharing system which makes sense given the mode of transportation. If we look at when NYS went on pause Citi Bike was not impacted. Looking at MTA data it was significantly impacted by the NYS on Pause Program when the subway ridership dropped significantly and since then has been on a path to recovery but has not yet reached its pre-pandemic form.

VIII. Scope & Limitations

Limitations in the MTA dataset include having access to ride duration information. Since the only current metric used to analyze MTA subway ridership is turnstile entry, there wasn't a figure to measure the length of NYC commuters average commute time. It would've been interesting to observe if COVID-19 had some sort of correlation on trip durations using public transportation.

The main limitation of the COVID-19 dataset was a lack of data on case counts broken out by ZIP codes. The only data provided by the NYC Department of Health and Mental Hygiene is ratios of cases per 100,000 population which made it incompatible for a comparison with the ridership data from MTA and Citi Bike. On top of that, any ZIP code level COVID-19 data did not have latitude and longitude coordinates available which also prevented us from showing COVID-19 cases on the map alongside with the ridership data.

IX. References

[1] J. F. Teixeira and M. Lopes, "The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike" *Transportation Research Interdisciplinary Perspectives*, 08-Jul-2020. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S2590198220300774. [Accessed: 20-Dec-2021].

[2] K. T. L. Sy, M. E. Martinez, B. Rader, and L. F. White, "Socioeconomic disparities in subway use and COVID-19 outcomes in New York City," American Journal of Epidemiology, Volume 190, Issue 7, July 2021 Available: https://academic.oup.com/aje/article/190/7/1234/6054607?login=true. [Accessed: 20-Dec-2021].

[3] S. Fathi-Kazerooni, R. Rojas-Cessa, Z. Dong, and V. Umpaichitra, "Correlation of subway turnstile entries and covid-19 incidence and deaths in New York City," *Infectious Disease Modelling*, 04-Dec-2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468042720300762. [Accessed: 20-Dec-2021].

[4] S. M. Kaufman, M. L. Moss, K. B. McGuinness, N. R. Cowan, C. E. Rudner, S. I. Schwartz, M. Joshi, D. L. Schack, J. Rogers, O. Limone, J. Malloy, J. Koh, K. Rivard, R. Wise, Y. Chudnaya, W. Farrell, and H. L. Wu, "Transportation during coronavirus in New York City," *TRID*, 30-Jun-2020. [Online]. Available: https://trid.trb.org/view/1728199. [Accessed: 20-Dec-2021].

[5] "Citywide mobility survey: Transportation impacts of covid-19." [Online]. Available: https://www1.nyc.gov/html/dot/downloads/pdf/2020_cms_covid_october_summary_report.pdf. [Accessed: 20-Dec-2021].

[6] C. Whong, "Taming the MTA's unruly turnstile data," *Medium*, 19-Jun-2020. [Online]. Available: https://medium.com/qri-io/taming-the-mtas-unruly-turnstile-data-c945f5f96ba0. [Accessed: 21-Dec-2021].

[7] I. Motivate International, "Citi Bike System Data," *Citi Bike NYC*. [Online]. Available: https://ride.citibikenyc.com/system-data. [Accessed: 21-Dec-2021].

[8] Nychealth, "Coronavirus-Data/trends at master · nychealth/coronavirus-data," *GitHub*. [Online]. Available: https://github.com/nychealth/coronavirus-data/tree/master/trends. [Accessed: 21-Dec-2021].

[9] C. Whong, "Taming the MTA's unruly turnstile data," *Medium*, 19-Jun-2020. [Online]. Available: https://medium.com/qri-io/taming-the-mtas-unruly-turnstile-data-c945f5f96ba0. [Accessed: 21-Dec-2021].

X. Project Resource Links

[1] Streamlite-IO Dashboard https://share.streamlit.io/sateshr/mta-cov19-viz/main/st-app.py

[2] [Github Source Code for Streamlit(private repo] https://github.com/sateshr/MTA-COV19-Viz

[3] [Github Dataset repo] https://github.com/sateshr/VisualAnalytics

[4] [Github Dataset EDA Jupyter Notebooks] https://github.com/yxiang001/Notebooks-part2-for-final-project

[5] [Github visualization Jupyter Notebooks] https://github.com/yxiang001/Final-project-for-data-visualization