Coursera Applied Data Science Capstone Course by IBM
Capstone Project: Exploring Neighborhoods Around Subway Stations With Largest Ridership Drops During the COVID-19 Pandemic

Phyo Shwe Yee Win
April 19, 2020

I.   Introduction

New York City is the epicenter of the coronavirus pandemic in the United States. As the number of COVID-19 cases grew in NYC throughout March 2020, subway ridership plunged as businesses implemented remote working policies, and non-essential businesses were shuttered. In this project, I will utilize the subway ridership data from the NYC Metropolitan Transportation Authority ("MTA") along with the Foursquare API to identify trends in subway ridership through March 2020, which stations experienced the greatest decline in passengers, as well as what type of venues are most common near the subway stations that showed the greatest drop in ridership in March.

Based on the results, we can infer what type of businesses would have experienced the largest drop in the number of customers/visitors. This analysis can be used by the New York State and New York City governments in planning support programs for businesses hardest hit by the pandemic. We can also potentially infer what type of venues people would most likely visit if the New York State government were to relax the social distancing measures in the future and subway ridership increases. This can provide them insights on how to start relaxing the New York State on PAUSE order when the time is right.

II.   Data

First, I will utilize the subway turnstile data from the MTA.[1] This dataset includes the cumulative number of entries and exits for subway turnstile registers at each NYC subway station at various points in time, normally recorded every four hours. Each subway station has one or multiple control areas (shown in column 'C/A'), one or multiple remote units (in column 'UNIT'), and one or multiple devices (in column 'SCP', short for Subunit Channel Position, which is a specific address for a device).

---

[1] Available at http://web.mta.info/developers/turnstile.html.

For example, the cumulative number of entries and exits at every 4-hour interval for device no. 02-00-00 at control area A033 and remote unit R170 in the 14th street Union Square station on March 1, 2010 is shown in Table 1 below:

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
4709	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	00:00:00	REGULAR	326919	995477
4710	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	04:00:00	REGULAR	326932	995492
4711	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	08:00:00	REGULAR	326937	995516
4712	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	12:00:00	REGULAR	326950	995668
4713	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	16:00:00	REGULAR	327012	995928
4714	A033	R170	02-00-00	14 ST-UNION SQ	LNQR456W	BMT	03/01/2020	20:00:00	REGULAR	327107	996156

Table 1. Sample MTA subway turnstile data for 14th St-Union Square station

Next, I will obtain the geographical coordinates for the identified stations using the NYC subway station geolocation data.[2] I can obtain a GeoJSON file containing the names of the subway stations in New York City, as well as their latitude and longitudes, and reformat the file into a Pandas dataframe in the format shown below.

Station	Line	Latitude	Longitude
Grand Central - 42nd St	4-5-6-6 Express	40.751807	-73.976713
Times Sq - 42nd St	N-Q-R-W	40.754612	-73.986768
86th St	1-2	40.788644	-73.976218
23rd St	4-6-6 Express	40.739864	-73.986599
59th St - Columbus Circle	1-2	40.768247	-73.981929
Union Sq - 14th St	4-5-6-6 Express	40.734673	-73.989951
Fulton St	G	40.687119	-73.975375
Herald Sq - 34th St	N-Q-R-W	40.749645	-73.987937
34th St - Penn Station	1-2-3	40.750373	-73.991057
42nd St - Port Authority Bus Term	A-C-E	40.757308	-73.989735

Table 2. Train lines and coordinates of sample subway stations

Finally, with the coordinates of the subway stations, I will use the FourSquare API to obtain the most popular venues around the identified stations. FourSquare is a location data platform that provides tools for exploring places around a given set of geographical coordinates. Using the API, I can obtain the venues, venue categories and their coordinates around subway stations. Some example venues are shown below.

---

[2] Available from the NYC Open Data website at
https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Grand Central - 42nd St	40.751807	-73.976713	Equinox East 44th Street	40.752687	-73.975323	Gym
Grand Central - 42nd St	40.751807	-73.976713	Num Pang Sandwich Shop	40.750522	-73.976124	Sandwich Place
Grand Central - 42nd St	40.751807	-73.976713	Grand Central Market	40.752321	-73.976086	Gourmet Shop
Grand Central - 42nd St	40.751807	-73.976713	Grand Central Terminal	40.752809	-73.977014	Train Station
Grand Central - 42nd St	40.751807	-73.976713	Meyers Bageri	40.752476	-73.977518	Sandwich Place

Table 3. Sample venues around the 42nd Street - Grand Central station

III.   Methodology

First, I manipulate the MTA subway turnstile dataset to calculate the daily number of entries for each station in the subway system. The dataset provides the cumulative entry register value at each turnstile at four hour intervals. I calculate the daily entries for each turnstile as the minimum number of entry register value for the following day minus the minimum number of entry register value for the day. I then sum up the daily entries for all the turnstiles in a station to calculate the daily entries for the station. Using the daily entries data, I conduct some exploratory data analysis to identify the ridership trends throughout the month of March.

Next, I identify the top five stations that experienced the largest drop in average weekday subway entries in the first week of March (2-6), to the last full week of March (23-27).

Finally, I explore the neighborhoods around these stations using the FourSquare API. I obtained the venues within a 500 meter radius of the coordinates of the station and calculated the frequency of each venue category, which is then used to identify the top ten most common venue categories around each station. This data is used to analyze whether the areas around the stations share particular characteristics, such as having mostly restaurants, fitness studios or coffee shops.

IV.   Results

Figure 1 below plots the total number of entries to the entire MTA system from February 29, 2020 to April 2, 2020. We can see that subway ridership has been in steep decline since the beginning of March, when there were more than 5 million daily entries on weekdays. In the last full week of March, ridership has declined to less than 800,000 daily weekday entries. We can also see that weekends generally have fewer passengers than weekdays.
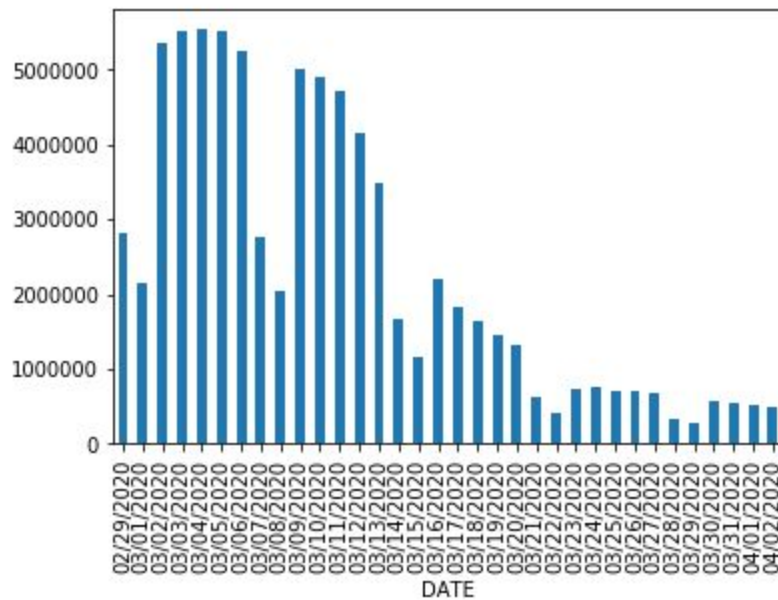
Figure 1. Daily total number of entries to the NYC MTA system

The next step is to identify which stations are generally the busiest. I calculated the average daily weekday ridership in the first week of March, and ranked the stations on this metric. Table 4 shows the top five stations with the highest average weekday ridership from March 2 (Monday) to March 6 (Friday).

STATION	Rank	Weekday Avg
34 ST-PENN STA	1	145647.2
GRD CNTRL-42 ST	2	130866.2
23 ST	3	114271.8
34 ST-HERALD SQ	4	94588.4
FULTON ST	5	93530.8

Table 4. Stations with highest average weekday ridership from March 2-6, 2020

The top five stations are all located in Manhattan. They consist of three stations serving midtown Manhattan where a large number of offices are (34th Street - Penn Station, Grand Central - 42nd Street and 34th Street - Herald Square), one serving Flatiron district (23rd Street), and one serving lower Manhattan (Fulton Street). I visualized the trend in entries at these stations throughout March in Figure 2. All of them showed the same trend - the daily entries started visibly declining during the second week of March, and by the end of the month, the stations are only serving a small fraction of the population.
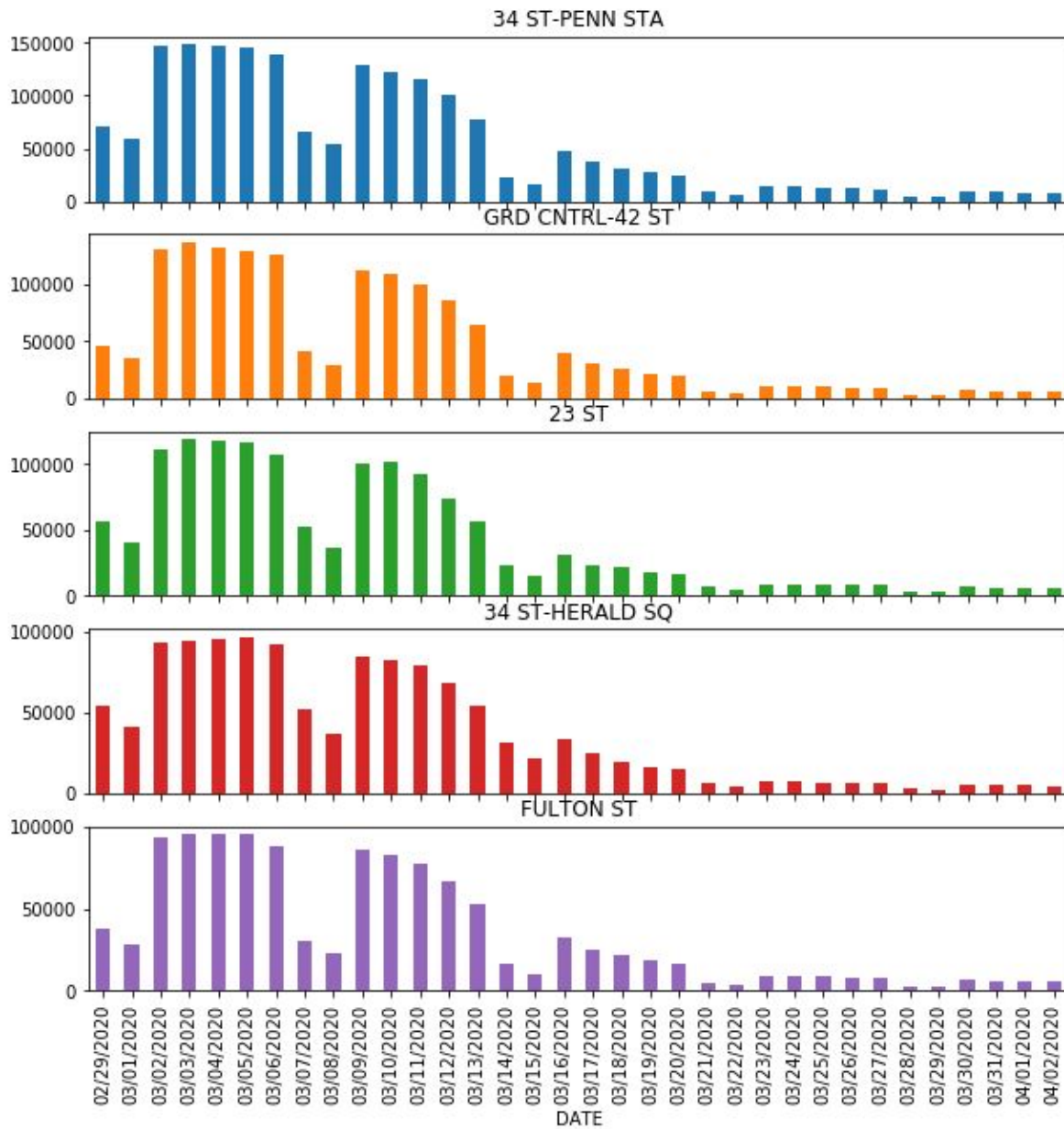
Figure 2. Daily entries to the top five busiest stations

Next, I analyzed the percentage change and absolute change in weekday ridership between the first week of March and the last week of March. I focused on weekdays because we can see a clear trend that weekday ridership is higher than weekend ridership and weekday travel is generally more non-discretionary than weekend ridership. As we can see in Table 5 below, the five busiest stations are also among the stations with the largest drop in ridership.

STATION	1stWeek	LastWeek	PctChg	AbsChg
34 ST-PENN STA	728236.0	64118.0	-91.195437	-664118.0
GRD CNTRL-42 ST	654331.0	51610.0	-92.112555	-602721.0
23 ST	571359.0	43281.0	-92.424903	-528078.0
34 ST-HERALD SQ	472942.0	35025.0	-92.594229	-437917.0
FULTON ST	467654.0	43085.0	-90.786992	-424569.0
14 ST-UNION SQ	461940.0	46627.0	-89.906265	-415313.0
TIMES SQ-42 ST	428444.0	32065.0	-92.515941	-396379.0
42 ST-PORT AUTH	414639.0	46386.0	-88.812919	-368253.0
86 ST	391973.0	47556.0	-87.867532	-344417.0
59 ST COLUMBUS	341180.0	33257.0	-90.252359	-307923.0

Table 5. Stations with largest absolute drop in entries between the first and last week of March

In the final step, I obtained the most common venues around these subway stations using the FourSquare API shown in Table 6 below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	23rd St	Indian Restaurant	Bar	Cosmetics Shop	Wine Shop	American Restaurant	Japanese Restaurant	New American Restaurant	Cheese Shop	Coffee Shop	Café
1	34th St - Penn Station	Korean Restaurant	Music Venue	Hotel	Donut Shop	Burger Joint	Pub	Deli / Bodega	Fast Food Restaurant	Lounge	Sushi Restaurant
2	42nd St - Port Authority Bus Term	Theater	Hotel	Italian Restaurant	Coffee Shop	Sandwich Place	Bakery	Furniture / Home Store	Seafood Restaurant	Gym / Fitness Center	Gym
3	59th St - Columbus Circle	Concert Hall	Gym	Jazz Club	Theater	Spa	Juice Bar	Bakery	Hotel	Grocery Store	French Restaurant
4	86th St	Italian Restaurant	Pizza Place	Bakery	Bagel Shop	Gym	Thai Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Dessert Shop
5	Fulton St	Italian Restaurant	Coffee Shop	Lounge	Cosmetics Shop	Flower Shop	Burger Joint	Performing Arts Venue	Opera House	Theater	Playground
6	Grand Central - 42nd St	Sandwich Place	Hotel	Coffee Shop	Japanese Restaurant	Deli / Bodega	Clothing Store	Grocery Store	Salon / Barbershop	Liquor Store	Steakhouse
7	Herald Sq - 34th St	Korean Restaurant	Hotel	Burger Joint	Japanese Restaurant	Dessert Shop	Hotel Bar	Gym / Fitness Center	New American Restaurant	Donut Shop	Cosmetics Shop
8	Times Sq - 42nd St	Theater	Hotel	Bakery	Burger Joint	Cycle Studio	Bookstore	Cuban Restaurant	Plaza	Deli / Bodega	Japanese Restaurant
9	Union Sq - 14th St	American Restaurant	Café	Ice Cream Shop	Park	Cosmetics Shop	Gym / Fitness Center	Bookstore	Japanese Restaurant	Mexican Restaurant	Mediterranean Restaurant

Table 6. Most common venues around subway stations with largest drop in entries

6

V.   Discussion

New Yorkers have to rely on the subway system to get to most places. Usage of the NYC subway system has dramatically decreased throughout March 2020, on a system-wide basis as well as at the station level for the top five busiest stations. The use of the subway system is now only for essential workers who still depend on it to get to their workplaces. The decline in passengers and revenue will cause significant financial burden for the MTA. The United States government has provided $3.8 billion to the MTA through the COVID-19 stimulus bill. The MTA Chairman has said that it would need another $3.9 billion as a result of the losses.[3] This is not surprising given the rock bottom ridership levels. Even after NYC recovers from the immediate crisis, people would be wary of being in enclosed spaces with multiple passengers for a prolonged period of time potentially in crowded conditions. Therefore, subway ridership is not expected to recover to pre-COVID-19 levels anytime soon.

We also see that stations that experience the largest drop in entries are in Manhattan. Half of them (the 34th street stations and the 42nd street stations) are in midtown, which has plenty of offices with the jobs that allow remote work. The most common venues around these subway stations include cafes and restaurants, places for tourists such as hotels and theaters, as well as places that cater more to locals such as gyms and fitness studios. Around 34th street Koreatown, we see that Korean restaurants are the most popular. Around the 42nd street stations, theaters are the most popular since this is where the broadway theaters are. All these businesses would experience a significant decline in customers as reflected in the decline in usage of the subway stations which provide access to these businesses.

VI.   Conclusion

The analysis shows that the NYC subway transit system serves a large number of passengers, with more than 5 million daily riders on weekdays during regular times. As the COVID-19 pandemic became more severe, there has been a steep decline in ridership of the subway, with weekday ridership falling by more than 90% across the system.

For the top ten subway stations that experienced the largest absolute drop in number of entries in March 2020, the surrounding neighborhoods largely consist of cafes, restaurants and bars, as well as fitness studios and gyms. We can also see that hotels and theaters are in the top ten venues for half of the neighborhoods. From these results, we can expect that the travel and entertainment, restaurants and fitness industries are going to be the economically hardest hit industries in NYC. Government support will be needed to keep the local businesses in these industries alive.

For future analysis, it would be interesting to see how the other modes of public transportation in NYC such as bike sharing (Citi Bike) and ride sharing (Uber, Lyft and Via) fared during the coronavirus pandemic. People might prefer these modes of transportation to being in an enclosed subway car with multiple other passengers for a prolonged period of time, so they might experience less of a decline. Given that there are many essential workers in the lower income group e.g. grocery store workers, food delivery workers, and home health aides, and they tend to live in the outer boroughs of NYC and commute to Manhattan, these may not be a viable alternative mode of transportation for them. It would be interesting to see which of these hypotheses the data supports.

---

[3] New York Post, "MTA expects low end-of-year ridership, wants another coronavirus bailout" dated April 16, 2020, available at
https://nypost.com/2020/04/16/mta-expects-low-end-of-year-ridership-wants-another-bailout/