New York City’s MTA: Why It’s Bad (and why it’s gotten worse)

Team Name: DaXuan Baba
Team Members: Cristian Statescu, David Wang, Wilmer Chang
Course Instructor: 王轩

May 22, 2024

Contents

Abstract

This research project investigates the dynamics of Metropolitan Transportation Authority (MTA) ridership in New York City, focusing on daily patterns and broader trends over several years. By visualizing extensive datasets, we identify key ridership behaviors and fluctuations. The study extends to analyzing the operations of the New York Police Department (NYPD) within subway transit districts, specifically examining fare evasion incidents and arrests. This analysis is contextualized by comparing the MTA's budget and enforcement practices with those of other major transit systems throughout the world, including Seoul's Metro and Hong Kong's MTR. Through this comprehensive approach, we aim to provide insights into the effectiveness of fare enforcement policies and their impact on both ridership and transit system economics. The findings offer valuable implications for policy adjustments and resource allocation within urban transit systems.

1 Introduction

The MTA subway system provides service to over 8 million New Yorkers with its 472 stations, each in a different state of maintenance and has a different capacity. As New York City evolves, the MTA continuously develops projects to improve the system throughout the city in terms of safety capabilities, reliability, and efficiency. Through ongoing projects and initiatives, the MTA strives to meet the ever-changing demands of the city's dynamic landscape. Central to these efforts is the utilization of data, which serves as a cornerstone for stakeholders and project managers, informing policy-making, guiding project development, and facilitating resource allocation.

This project aims to offer valuable insights to our readers by addressing the claims of a 13% increase in subway crime since 2023, with assaults specifically rising by 11% within the transit system. Through meticulous data exploration and visualization techniques, we aim to shed light on the correlation, if any, between crime rates and ridership numbers.[1] By a comprehensive analysis, we will not only clarify the trends in subway crime but also provide a deeper understanding of the factors influencing these patterns.

2 Data Collection and Preprocessing

In order to begin our project, data needs to be selected and processed. By accessing the NYC Open Data, we're able to access relevant datasets from the various departments that exist throughout the city that are relevant to our study. The following datasets we utilize for this project are:

- MTA Open Data
  - MTA Subway Hourly Ridership Beginning February 2022
  - MTA Turnstile Usage Data
- New York Police Department
  - Subway Transit Districts
  - Subway Fare Evasion

2.1 MTA Subway Hourly Ridership Dataset

The MTA Subway Hourly Ridership dataset consists of estimates of subway ridership on an hourly basis by subway station complexes across the city and the type of fare payment that was used. As this dataset dates only back to February 2022, we also access the dataset published by the MTA on turnstile usage in the years before February 2022. Since the MTA Subway Hourly Ridership is currently maintained by the MTA Open Data team, our focus was to aggregate our data for 2023 and obtain a station complex dictionary to be able to use the data in conjecture with the turnstile usage data to have the ability to comprehensively temporally analyze the ridership data for patterns and points of interest.

```
df = pd.read_csv('data/MTA_Subway_Hourly_Ridership__Beginning_February_2022_3.20.2024.csv', low_memory=False)
df['transit_timestamp'] = pd.to_datetime(df['transit_timestamp'])
df_ridership = df.groupby([df['transit_timestamp'].dt.date,'station_complex_id','station_complex','borough','Georeference']
                          ).agg({'ridership': 'sum', 'transfers': 'sum'}).reset_index()
df_2023_ridership = df_ridership[df_ridership['transit_timestamp'].dt.year == 2023]
df_2023_ridership.to_csv('data/ridership_daily_station_2023', index=False)
```

Figure 1: Code snippet to process MTA Subway Hourly Ridership data.

2.2 NYPD Subway Transit Districts

Upon researching for data for transit districts we discovered that the data doesn't exist in a form that is easily accessible using the current tools we are using, as well as the formatting is not manipulable. This data can be seen outlined in the map in Figure 2. However, as the transit districts are mapped on an MTA Subway Map, we are able to associate each subway station in our station dictionary to be able to utilize in the future.

Figure 2: MTA Subway Transit Districts

2.3 NYPD Subway Fare Evasions and Arrests

Like the data from the NYPD on the Subway Transit Districts, the data in regards to the NYPD reports on subway fare evasions and arrests isn't in a data format that is relatively ready for visualizations with the rest of the data that we currently have. The data exists in the form of pivot tables in Excel. However, it is as simple as transitioning the columns and their data from the pivot tables to be consolidated into our existing data files as columns in our dataset. After doing that organization, we result in a dataset with a data dictionary that can be seen in Table 1.

Data Label	**Data Type**	**Data Description**
Quarter	DATE	Denotes quarter and year that the data was collected
Transit District	INTEGER	Denotes the transit district in which the data was collected
Gender	TEXT, INTEGER	Count based on Gender
Race	TEXT, INTEGER	Count based on Race
Age	TEXT, INTEGER	Count based on Age

Table 1: Data Labels, Types, and Descriptions

2.4 MTA Turnstile Usage Data

The MTA Turnstile Usage Data was using out of date labeling system that was then abandoned after 2022. This dataset also did not specify the ridership and transfer number during each specified recording time. There were no variables /columns denoting station complex as well. Therefore, we needed to clean the datasets and convert the formatting into usable format. First, we created unique ID & Use Crosswalk tables to get from old format/ names to up to date format/names as well additional georeference points and geographic information. We also identified and removed empty values as they are just very small porttion of the entire dataset. Next, we calculate the ridership and transfer count for each row with constraints set to eliminate outliers/extreme numbers from system counting fault. Finally we aggregate the daily turnstiles counts into each daily station counts and then daily station complex counts.

```
#Create special name to use crosswalk table
# creating a new column "unique_ID" based on the station, line name and division
turnstiles["unique_ID"] = turnstiles[["Station", "Line Name", "Division"]].apply("-".join, axis=1)
```

Figure 3: Code to get special name for each station.

	station_complex_id	Date	ridership	net_exits	station_complex	Georeference	borough
0	A002	2019-01-01	12897	14333	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
1	A002	2019-01-02	37302	37973	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
2	A002	2019-01-03	40823	40826	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
3	A002	2019-01-04	40207	41009	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
4	A002	2019-01-05	17778	19431	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
5	A002	2019-01-06	14638	16490	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
6	A002	2019-01-07	39281	40364	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
7	A002	2019-01-08	40687	41082	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
8	A002	2019-01-09	41877	42428	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
9	A002	2019-01-10	41695	43215	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M
10	A002	2019-01-11	40534	41233	Lexington Av (N,R,W)/59 St (4,5,6)	POINT (-73.967255 40.76266)	M

Figure 4: Example of the post-clean mta turnstiles usage data.

2.5 External Research

We needed to do some additional research about other Metro systems. As such, the group went online and began researching in order to make our own simple data sets. We looked up different Metro Systems and found different metrics regarding them. These metrics were:

- Length of track per Metro System
- Annual Budget for each Metro System for 2024
- Average Cost per mile

The way that we decided for the cities was by looking up major metro systems in America, Europe, and finally, Asia. That led us to gathering data for DC Metro, Denver RTD, BART (San Fransico), Tokyo Subway, Seoul Metro, Berlin U Bahn, Munich U Bahn, Dublin Luas, and MTR (Hong Kong Metro). Once we finished looking for the information for these Metro Systems, we made an Excel file for them (sometimes doing a bit more work - like in order to the find the average cost per mile for NYC, we took the average travel distance for a NYC resident and divided \$2.90 by the total average distance for that visual), and then plotted them out. Those can be found within our GitHub Repository. In the report, as an example of the data which helped us in the right direction in terms of research, we wanted to show the annual

System	Average Cost per Mile (USD)
NYC MTA	0.53
DC Metro	0.33
Denver RTD	0.25
BART	0.25
Tokyo Subway	0.20
Seoul Metro	0.15
Berlin U-Bahn	0.30
Munich U-Bah	0.35
Dublin Luas	0.25
MTR (HK)	0.20

Figure 5: Data for the Average Cost per Mile for Different Metro Systems

We noticed that MTA's annual budget for 2024 was incredibly high comparing to other Metro Systems. [2][3][4][5][6][7][8][9][10][11]

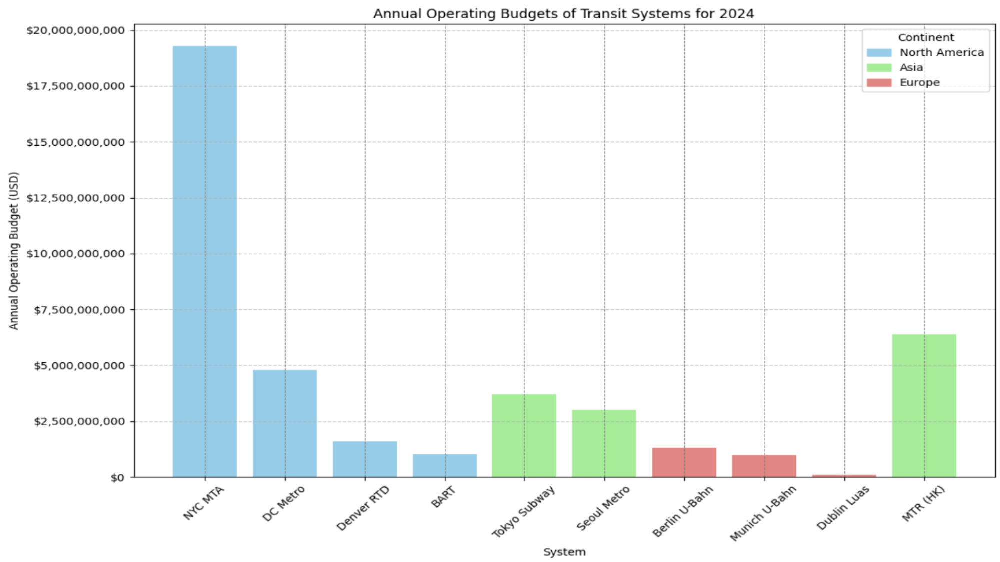

Figure 6: Transit Systems' Budget for 2024

Hence, with this reference, we went ahead and then looked further into the MTA's annual budget from 2017 up until now by looking at MTA's financial documents, which led to us collecting more information about fare-evasions arrests and summonses from NYPD's data. [12] Those figures again can be found within our GitHub Repository under their respective IPYNB files.

NYPD's data was constructed in a way such that it was a copy and pasted Pivot-table (created in Excel). So, to make the data easy to graph out, we aggregated many of their spreadsheets from different quarters of different years into one massive, clean Excel file, which we later used to create more visuals and see some interesting events which led to the increase of fare evasion tickets and summonses.

Subway Fare Evasion
(21 NYCRR Section 1050.4)
Summons Report
4th Quarter 2023

Station
Summons by Gender

Borough/Station/Line	FEMALE	MALE	UNKNOWN	UNSPECIFIEI	Grand Total
BRONX	**332**	**2919**	**7**	**0**	**3258**
138 ST.-GRAND CONCOURSE	**2**	**15**	**0**	**0**	**17**
4,5	2	15	0	0	17
149 ST.-GRAND CONCOURSE	**12**	**76**	**0**	**0**	**88**
2,5	0	8	0	0	8
4	12	68	0	0	80
161 ST.-YANKEE STADIUM	**39**	**506**	**1**	**0**	**546**
4	9	184	1	0	194
B,D	30	322	0	0	352
167 STREET	**5**	**23**	**0**	**0**	**28**
4	2	13	0	0	15
B,D	3	10	0	0	13
170 STREET	**7**	**40**	**0**	**0**	**47**
4	7	32	0	0	39
B,D	0	8	0	0	8
174-175 STREETS	**0**	**9**	**0**	**0**	**9**
B,D	0	9	0	0	9
176 STREET	**5**	**78**	**0**	**0**	**83**
4	5	78	0	0	83
182-183 STREETS	**1**	**5**	**0**	**0**	**6**
B,D	1	5	0	0	6
183 STREET	**4**	**19**	**1**	**0**	**24**
4	4	19	1	0	24
205 ST.-NORWOOD	**1**	**4**	**0**	**0**	**5**
D	1	4	0	0	5
219 STREET	**0**	**5**	**0**	**0**	**5**
2,5	0	5	0	0	5
225 STREET	**1**	**4**	**0**	**0**	**5**
2.5	1	4	0	0	5

Figure 7: The fare evasion summonses data for the last quarter of 2023, before cleaning/aggregation

W35

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Quarter	Quarter Start Date	Quarter End Date	Year	Transit Distict	Female	Male	Unknown-Gender	American Indian	Asian/Pac Isl	Black	Hispanic	Unknown-Race	White	10 - 17	18 - 24	25 - 40	41 - 59	60+	Unknown-Age	Grand Total
2	1	1/1/18	3/31/18	2018	1	507	1204	1	12	124	572	472	90	442	79	792	627	191	20	3	1,712
3	1	1/1/18	3/31/18	2018	2	376	666	4	6	134	294	221	92	299	41	403	433	146	20	3	1,046
4	1	1/1/18	3/31/18	2018	3	138	336	2	1	24	178	182	11	80	55	193	152	69	6	1	476
5	1	1/1/18	3/31/18	2018	4	353	906	3	11	93	423	322	55	358	49	516	517	168	11	1	1,262
6	1	1/1/18	3/31/18	2018	11	121	507	3	2	13	258	319	16	23	52	275	222	80	2	0	631
7	1	1/1/18	3/31/18	2018	12	107	421	1	0	8	243	250	8	20	57	228	175	64	5	0	529
8	1	1/1/18	3/31/18	2018	20	264	657	1	16	120	244	296	176	70	68	372	309	161	10	2	922
9	1	1/1/18	3/31/18	2018	23	118	283	2	4	31	220	87	37	24	55	181	98	66	3	0	403
10	1	1/1/18	3/31/18	2018	30	371	648	0	3	86	438	216	67	209	102	393	374	135	14	1	1,019
11	1	1/1/18	3/31/18	2018	32	142	390	0	2	8	372	77	16	57	75	245	152	56	4	0	532
12	1	1/1/18	3/31/18	2018	33	108	411	0	1	22	232	127	14	123	35	226	191	60	7	0	519
13	1	1/1/18	3/31/18	2018	34	251	478	2	3	89	121	203	43	272	69	253	263	141	4	1	731
14	2	4/1/18	6/30/18	2018	1	464	1068	2	7	93	509	449	96	380	63	706	545	199	20	1	1,534
15	2	4/1/18	6/30/18	2018	2	345	635	2	8	105	287	197	76	309	40	429	345	143	25	0	982
16	2	4/1/18	6/30/18	2018	3	141	389	6	1	21	197	213	14	90	44	198	192	83	17	2	536
17	2	4/1/18	6/30/18	2018	4	385	808	5	3	67	369	340	45	374	46	488	456	186	21	1	1,198
18	2	4/1/18	6/30/18	2018	11	137	654	3	0	10	317	413	16	38	50	305	263	158	17	1	794
19	2	4/1/18	6/30/18	2018	12	189	574	0	2	10	380	336	11	24	89	286	235	143	9	1	763
20	2	4/1/18	6/30/18	2018	20	290	707	2	15	132	290	311	170	81	72	408	328	169	17	5	999
21	2	4/1/18	6/30/18	2018	23	81	245	0	3	20	168	72	39	24	43	128	94	55	5	1	326
22	2	4/1/18	6/30/18	2018	30	389	745	1	2	92	529	229	58	225	85	470	400	169	8	3	1,135
23	2	4/1/18	6/30/18	2018	32	102	217	4	1	13	218	43	8	40	34	130	119	36	3	1	323
24	2	4/1/18	6/30/18	2018	33	123	371	0	0	16	183	129	18	148	19	225	192	48	10	0	494
25	2	4/1/18	6/30/18	2018	34	234	540	3	2	71	173	199	45	287	82	287	255	144	9	0	777
26	3	7/1/18	9/30/18	2018	1	656	1,536	1	8	158	743	644	144	496	118	1,032	803	215	23	2	2,193
27	3	7/1/18	9/30/18	2018	2	469	865	1	13	131	429	286	110	366	66	591	482	173	23	0	1,335
28	3	7/1/18	9/30/18	2018	3	241	602	1	0	28	325	342	18	131	86	360	266	113	18	1	844
29	3	7/1/18	9/30/18	2018	4	584	1,129	4	8	125	559	420	95	510	66	804	622	195	26	4	1,717
30	3	7/1/18	9/30/18	2018	11	204	890	3	1	17	476	554	12	37	89	472	368	154	13	1	1,097
31	3	7/1/18	9/30/18	2018	12	528	1,669	1	10	45	1,125	899	29	90	195	906	748	318	29	2	2,198
32	3	7/1/18	9/30/18	2018	20	256	638	2	13	113	237	323	123	87	71	393	291	128	11	2	896
33	3	7/1/18	9/30/18	2018	23	84	244	1	1	24	174	76	38	16	32	113	103	67	12	2	329
34	3	7/1/18	9/30/18	2018	30	362	829	1	2	52	597	282	54	205	109	507	391	172	13	0	1,192
35	3	7/1/18	9/30/18	2018	32	171	394	2	1	6	428	77	6	49	89	272	138	55	12	1	567
36	3	7/1/18	9/30/18	2018	33	209	601	2	1	37	350	185	36	203	56	373	304	69	9	1	812
37	3	7/1/18	9/30/18	2018	34	215	568	1	1	59	185	213	51	275	99	299	264	110	10	2	784
38	4	10/1/18	12/31/18	2018	1	822	1,904	5	18	162	907	788	147	709	96	1,261	997	324	51	2	2,731
39	4	10/1/18	12/31/18	2018	2	614	1,250	4	13	218	551	447	151	488	75	743	757	246	45	2	1,868
40	4	10/1/18	12/31/18	2018	3	275	799	1	3	37	444	397	29	165	77	427	382	155	33	1	1,075
41	4	10/1/18	12/31/18	2018	4	643	1,312	4	17	141	589	557	86	569	44	854	751	279	25	6	1,959

Figure 8: Fare Evasion Summonses after cleaning/aggregation

We ended up also doing our own pivot-tables on this data (see PIVOT marked Excel files in the GitHub Repository), and used them as a reference to inspire some of our other visuals.

NOTE: These visuals created from the data in this section were also visuals we used in our research, hence, they are also a ”Result”.

3 Results and Discussion

We began our exploration of our dataset with a visualization using D3 to visualize the stations that saw the most ridership throughout the system daily. By using our Observable notebook and using the D3 library, we were able to create a dynamic visual that allow the user to select the date range that the user wishes to see. A frame of this can be seen in Figure 9.

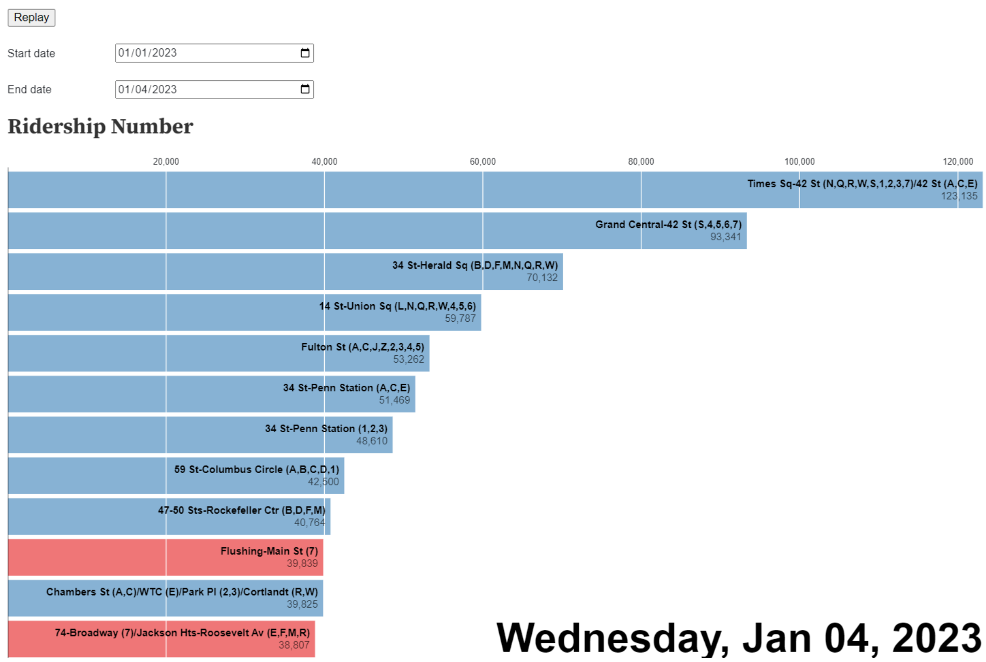

Figure 9: MTA Subway Ridership on Jan 4th, 2023

Looking into how the summonses issued are distributed in different transit districts, which can be seen in Figure 10. Unfortunately, we don't have a shape file to visualize the number in their exact transit district areas, here we are looking at how they are distributed at each subway station on the MTA subway line. We can see that summonses issued located in the East of Manhattan are more than in any other areas, followed by the Bronx and Queens districts, and lastly least in Brooklyn.

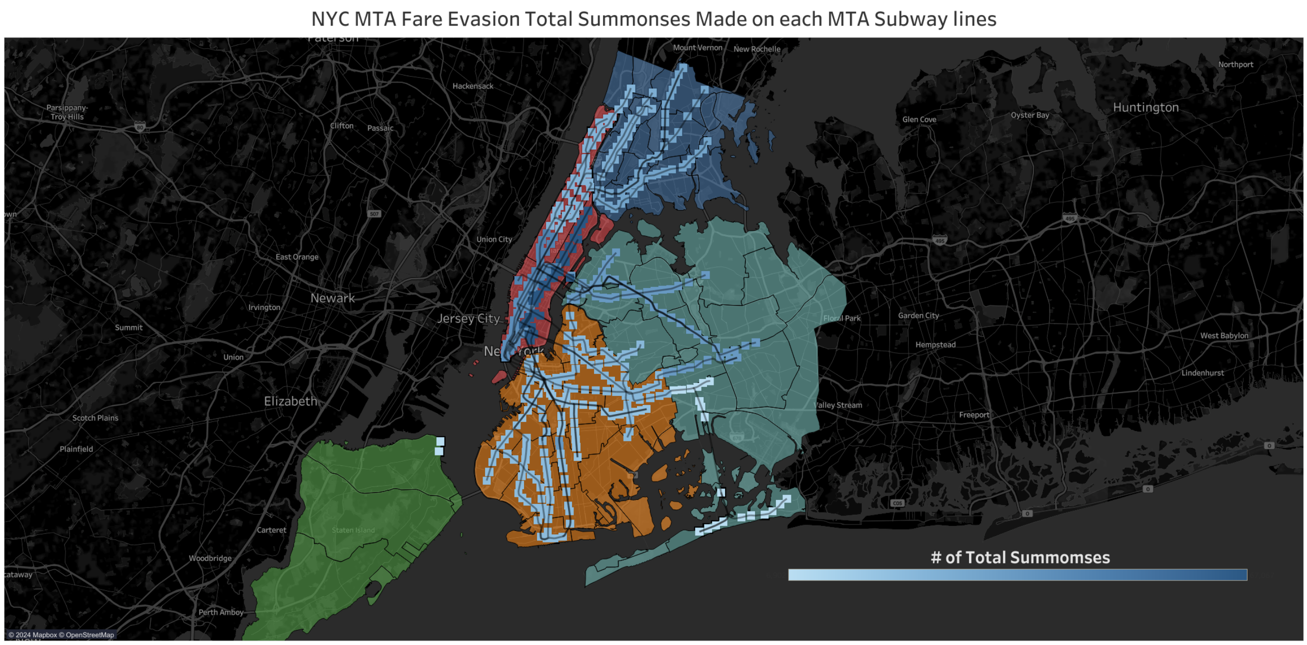

Figure 10: Distribution of Summonses issued in different boroughs

Looking at the temporal analysis of arrests made and summons issued numbers, before going into 2020, we are already seeing a steady decreasing trend and then a steep decrease in the arrest made number. Starting in 2021, we see an increasing trend with the number of arrests made number going up ever since.

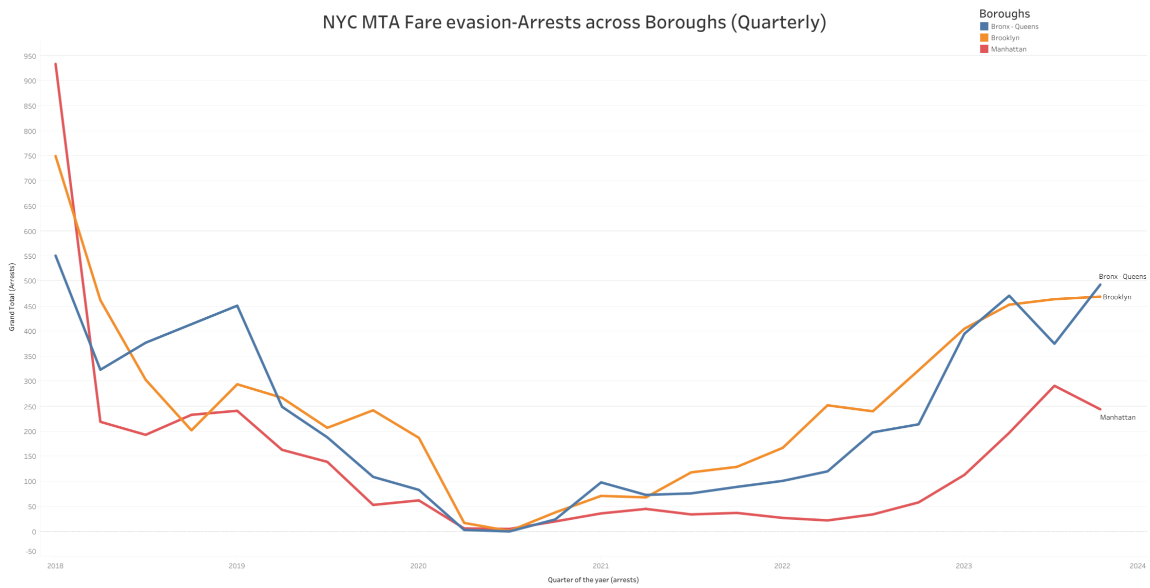

Figure 11: Arrests made in different boroughs over time (2018 - 2023)

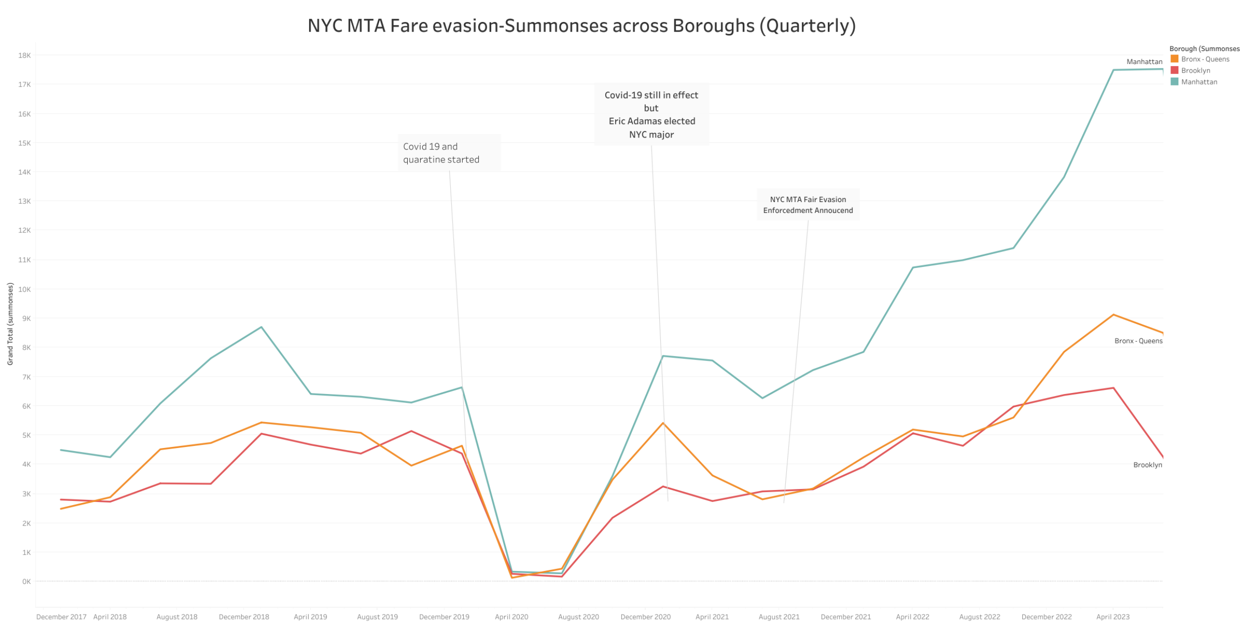

Figure 12: Summonses issued in different boroughs over time(2018 - 2023)

4 Conclusion

What we gathered from our data is that overall, more of the money that MTA has allocated needs to be spent on betting the stations overall. Rather than spending money on security guards, money should instead be spent on mimicking other Metro Systems around the world, like Seoul's, Tokyo's, Hong Kong's Metro Systems. Better infrastructure, such as platform screen doors (PSDs) to prevent any sort of access to the track before a train is at the station would help prevent not only suicides, but shovings, as well as track fires from littering.[13] Additionally, it would help combat Subway Surfing, which has been on the rise as of late.[14][15]

Figure 13: Tokyo PSDs

Additionally, in terms of public safety, providing better homeless aid would help with the increased unsafety in the subways. Many homeless individuals have mental health issues (a large number of the homeless population suffers from mental health conditions, with about 30% having severe mental illness and many also facing substance use issues).[16][17] If the city were to put more work into assisting homeless individuals and providing mental health aid, the issue of homelessness would slowly start to get solved, and the safety on MTA's trains will improve.

In all, a better allocation of resources into the right improvements would go a long way in bettering NYC's MTA and bringing them up to the standard that other Metro Agencies in other cities around the world (and also in the USA) have in our current day.

5 Future Work

The following are some ideas we have for how we'll take this project into the future:

- Creating a shapefile to present the Transit District jurisdiction efficiently.
- Create a comprehensive dashboard to allow user to explore, visualize and follow the research.
- Further research what other countries do and think of concrete plans to better allocate MTA's budget

6 Contribution

Task	Contributor
Project Research, Proposal, Progress Report, Final Report, Presentation, General Visuals	Cristian, David, Wilmer
• MTA Open Data Ridership Data Aggregation and Processing • NYPD Transit Bureau District Data Processing • NYPD Fare Evasion Arrest and Summons Aggregation • Bar Chart Race (Ridership Data)	David
• Turnstile Data Cleaning and Processing • Turnstile Data Station Dictionary • NYPD 3D Density Map Arrests and Summons Aggregation • Tableau Dashboard	Wilmer
• Other cities' data visualization • Other Metro Systems' Comparison Research • Bettering MTA Research • Turnstile Data Cleaning for D3 (long/lat fixes) • Other city's metro systems research • Data cleaning for other cities	Cristian

References

[1] A. Siff and N. N. Y. Staff, "NY Gov Announces New Subway Safety Push, Deploys National Guard Amid Crime Spike," 2024.

[2] B. A. R. Transit, "BART Financials," https://www.bart.gov/about/financials, 2024.

[3] Washington Metropolitan Area Transit Authority, "Fiscal Year 2024 Budget: Proposed Fare and Service Changes," https://www.wmata.com/initiatives/budget/fy2024-budget.cfm, 2024.

[4] T. Metro, "Tokyo Metro," https://www.tokyometro.jp/en/index.html, 2024.

[5] Metropolitan Transportation Authority, "Financial Information and Budget Statements," https://new.mta.info/transparency/financial-information/financial-and-budget-statements, 2024.

[6] U. P. News, "Seoul Metropolitan Subway," https://english.seoul.go.kr/city-hall/budget/, 2024.

[7] R. Denver, "RTD Facts and Figures," https://www.rtd-denver.com/open-records/reports-and-policies/facts-figures, 2024.

[8] Luas.ie, "Luas," https://www.luas.ie/, 2024.

[9] Wikipedia, "MTR," https://en.wikipedia.org/wiki/MTR, 2024.

[10] Berlin.de, "Subway (U-Bahn)," https://www.berlin.de/en/public-transportation/1742343-2913840-underground-subway.en.html, 2024.

[11] MVG.DE, "Munich Metro Website," https://www.mvg.de/, 2024.

[12] NYPD, "Subway Evasion Reports," https://www.nyc.gov/site/nypd/stats/reports-analysis/subway-fare-evasion.page, 2024.

[13] CDC, "Platform gates and doors," https://trespasstoolkit.fra.dot.gov/eLib/Details/L00028#:~:text=Platform%20screen%20doors%20are%20effective,entry%20to%20the%20train%20tracks., 2023.

[14] J. MARTINEZ and K. HONAN, "City Rolls Out Anti-Subway Surfing Campaign as Deaths Spike," 2023.

[15] MTA, "MTA Releases Track Trespassing Task Force Report," https://new.mta.info/press-release/mta-releases-track-trespassing-task-force-report, 2022.

[16] H. E. Marano, "The Complex Link Between Homelessness and Mental Health," 2021.

[17] CDC, "Homelessness and Health," https://www.cdc.gov/orr/homelessness/index.html, 2022.