



Addressing Delays and Accessibility to Enhance MTA Subway User Experience

Group 26

Matthew Taylor, Max Drogin, Evan Beck, Kanishkha

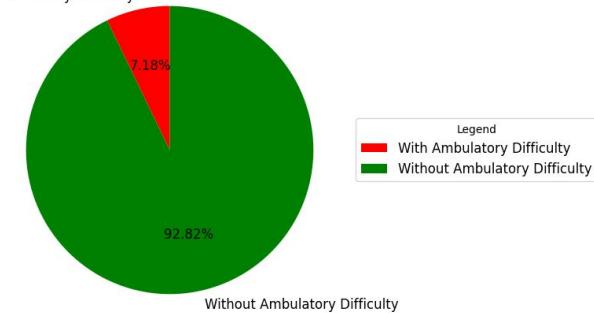


We Chose Two Pillars of Customer Experience



Distribution of Population with Ambulatory Difficulty

Ambulatory Difficulty



Delays

There were 449,652 delays in 2023, down from 498,487 in 2022, but still quite high

Accessibility

We estimate 7% of the NYC population have an ambulatory deficiency





Our Analysis Utilized Many Python Packages

Analysis Overview

- Observe patterns in overall Delays and Major Incidents
- Deep dive on issues in Delays
- Observe patterns in survey data for accessibility
- Analyze station availability rate
- Do a by Borough and by zip code visualization on accessibility
- Deep dive on ADA station and disability metrocard usage



Packages

- Data Manipulation & Analysis
 - Pandas, Numpy, Datetime
- Data Visualization
 - Matplotlib, Seaborn, Plotly, Squarify
- Geographic Visualization
 - Folium, Geopandas, Shapely
- Machine Learning
 - Sklearn
- Web Scraping and Requests
 - Request, Urllib, Webbrowser
- File Processing
 - Os, Json, Re, google.colab, IPython.Display, Kaggle



A look at Delays

Delay: A delay is normally defined as a train being more than five minutes late

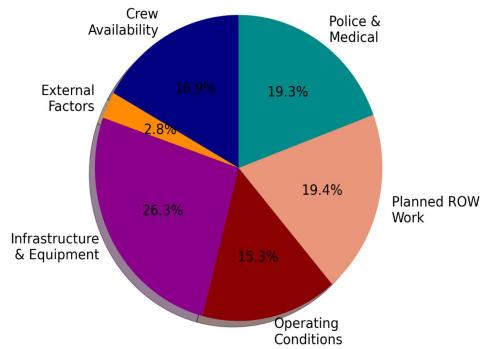
Major Incident: Incidents when 50 or more trains are delayed





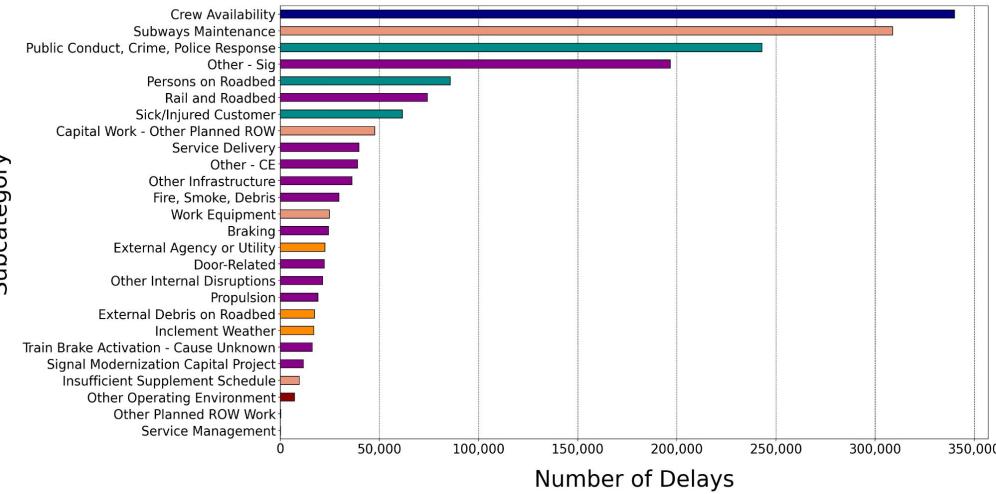
Simple Delays are Dominated by Infrastructure

Frequency of Simple Delays Per Reporting Category



- 36.3% of simple delays are due to crew availability and Planned ROW work- areas that may be able to be adjusted
- The remaining 63.7% of simple delays may be more difficult to account for

Frequency of Simple Delays Per SubCategory



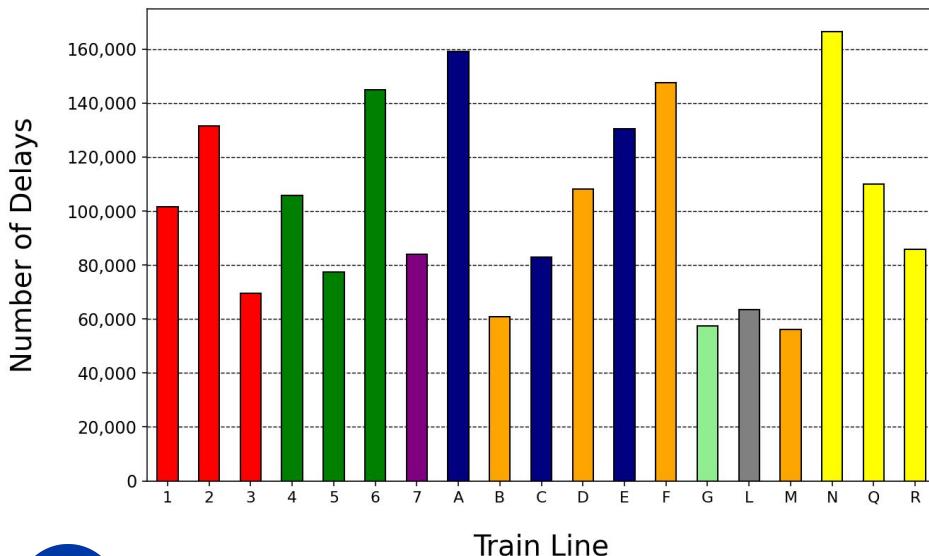
- The main corresponding subcategory under Planned ROW work is due to subway maintenance whereas crew availability's main and only subcategory is 'crew availability'



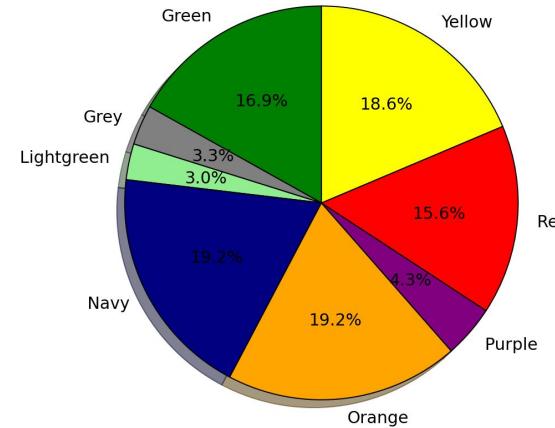


The Blue & Orange Train Lines Experience the Most Delays

Frequency of Simple Delays Per Train Line



Total Simple Delays by Train Line Color



- The N train had the most delays of any line
- The yellow, red, orange, green and blue line groups experience the most delays.
- The light green, grey and purple line groups experience the least amount of delays.

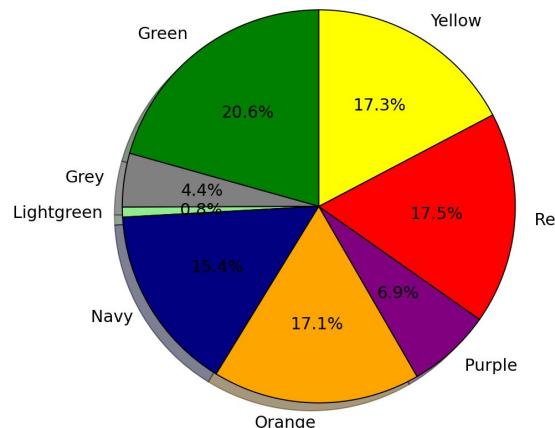
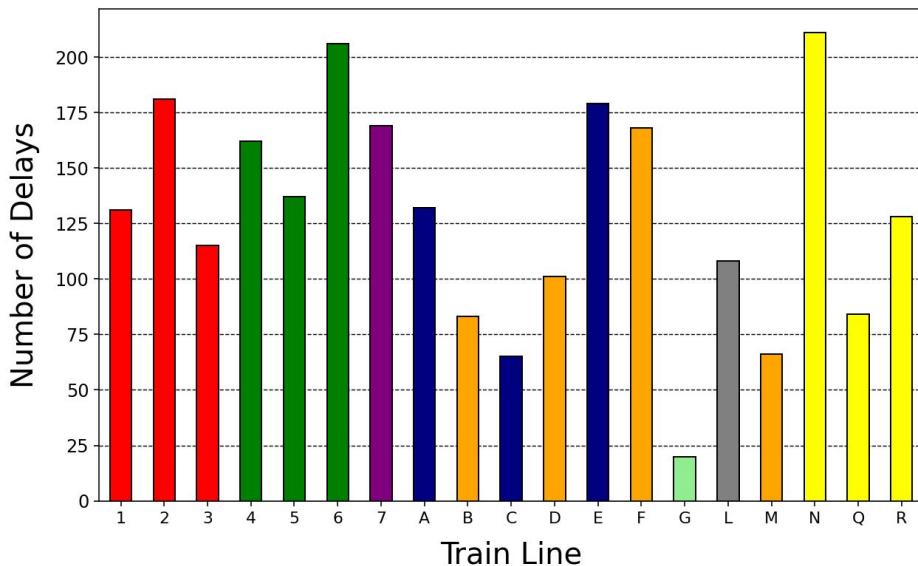




Major Incidents Affect the 456 Trains the Most

Total Major Incidents by Train Line Color

Frequency of Major Incidents Per Train Line

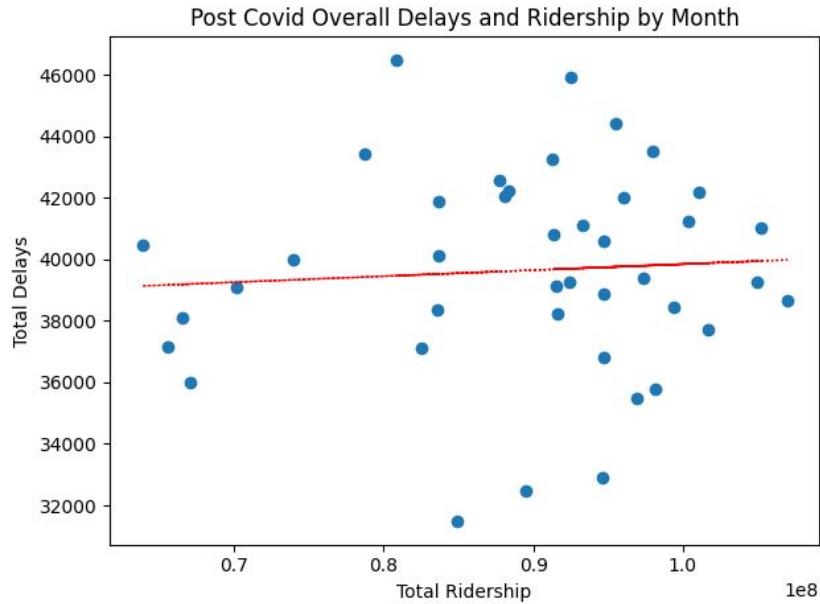


- Overall the number of major incidents is much lower than major delays
- The green line group had the most significant increase in percent of total (4%) from normal delays
- The blue lines had the most significant decrease at 4%.
- These may be indicative of where the MTA is working to prevent major delays



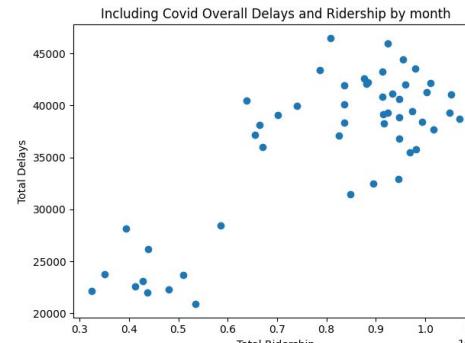


Overall Ridership and Delays Per Month do not Correlate*



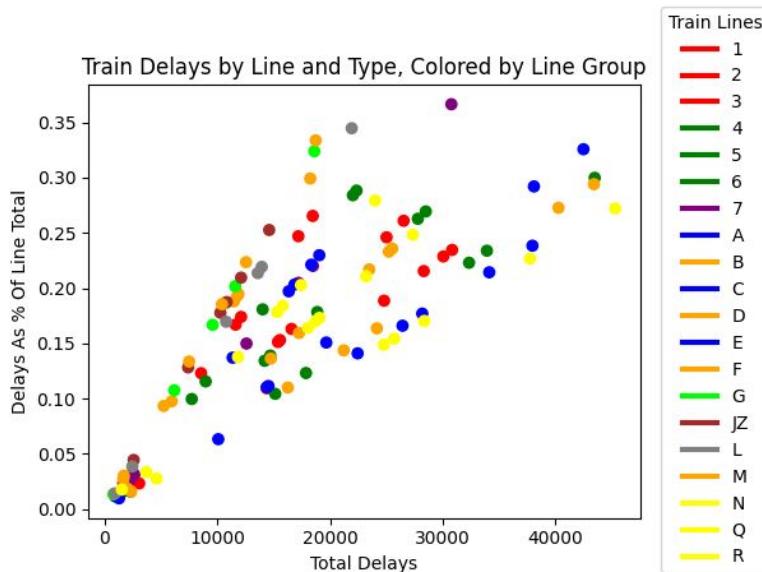
- We compared overall ridership and the number of delays seen in those months
- Using data from 6/1/2021 onwards we see an r of 0.07 which is not an indicator of linear correlation
- This implies that improving the subways which will increase riders will not increase delays

*Except when including Covid Data ($r = 0.8$)

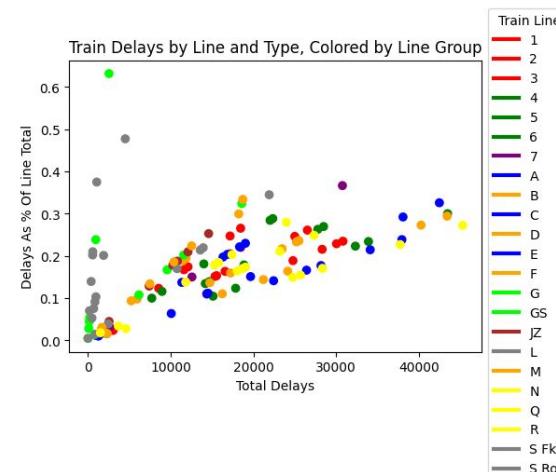




The N Still Leads Delays When Broken Out by Line and Type

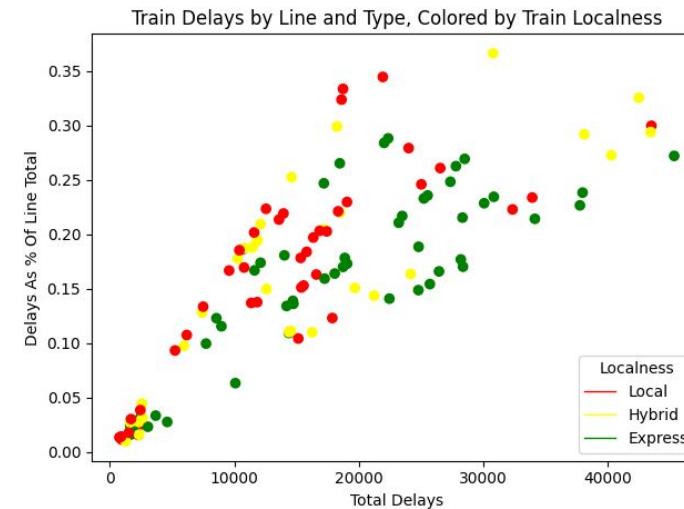
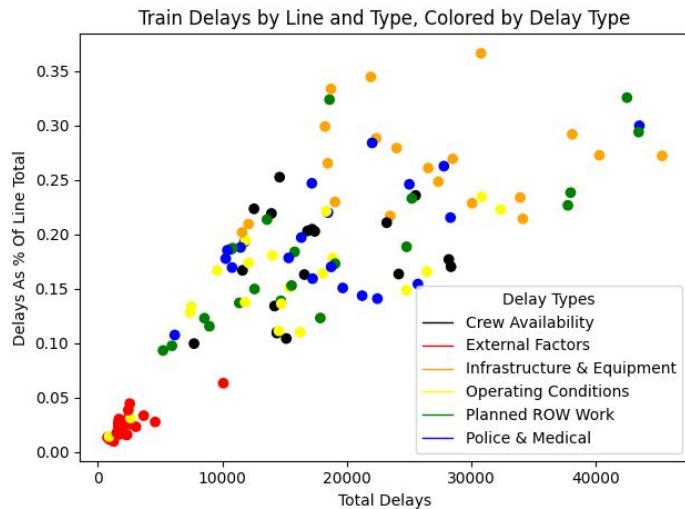


- We created a view of Delays by Delays as % of the total line's delays to look for outliers
- The 7 Train stands out as a highly delayed train with a common reason (Infra and equipment)
- We took out the GS and 2 S train types included which stood out as suffering from particular issues





Infrastructure and Equipment is a Common Problem for Trains



- We recolored the previous graph to observe patterns in delays across delay types and train types
- On the left graph we see the outliers like Police and Medical assistance on the 6 train
- On the right we see a large clumping of local trains on the top of the “cone” led by the L train



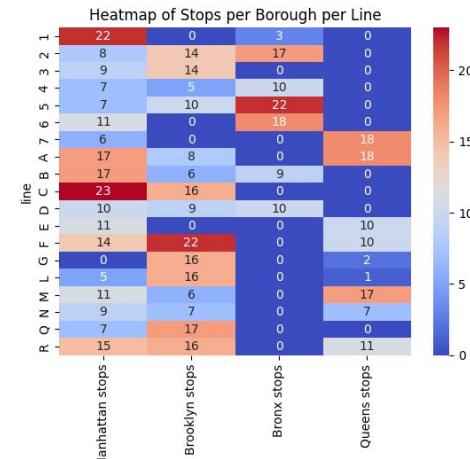
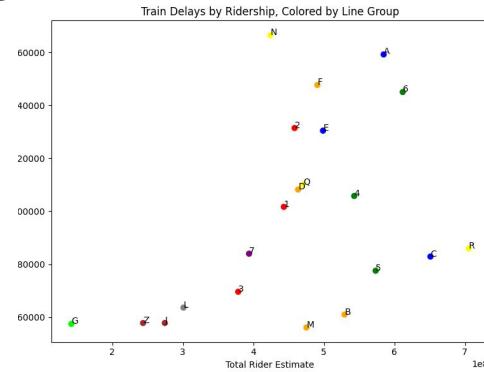
Delays: Takeaways & Recommendations

Takeaways

- The N train experienced the most delays and major incidents since 2020
- For its ridership the G sees an unexpected amount of delays
- Crew availability is the most impactful delay subcategory

Recommendations

- As an overall approach, fixing Infrastructure and Equipment as well as ensuring crew availability is a great start
- There also is an opportunity to think about individual line strategies like crime on the 6
- An investigation deeper into the properties of each line and other factors that lead to delays would be interesting (like a borough classifier)



Investigating Accessibility

Accessibility: Degree to which a product, service, resource, or environment is available to a given user.

ADA Station: A Station that abides by the American with Disabilities Act.

Ambulatory Disability: A disability that affects a person's ability to walk or move around.

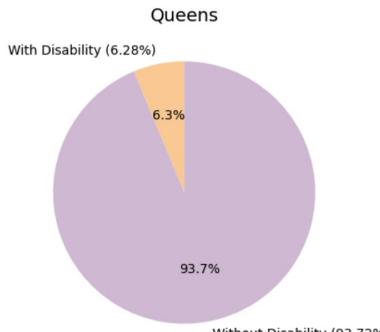
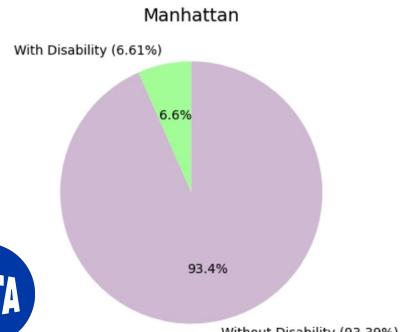
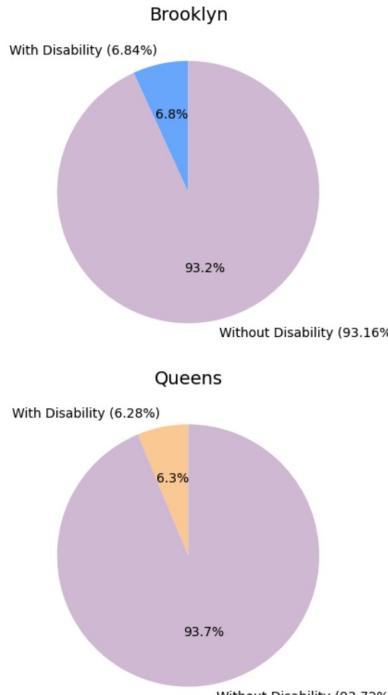
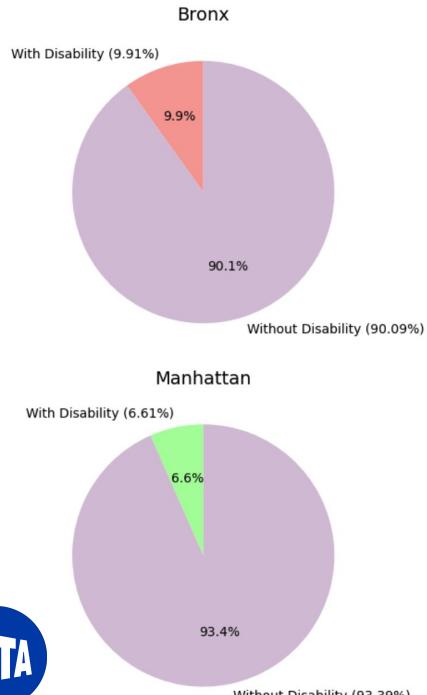
Main user groups focused on are seniors and those with disabilities





Proportion of Population with Ambulatory Difficulty is High in each Borough

Population with Ambulatory Disabilities by Borough

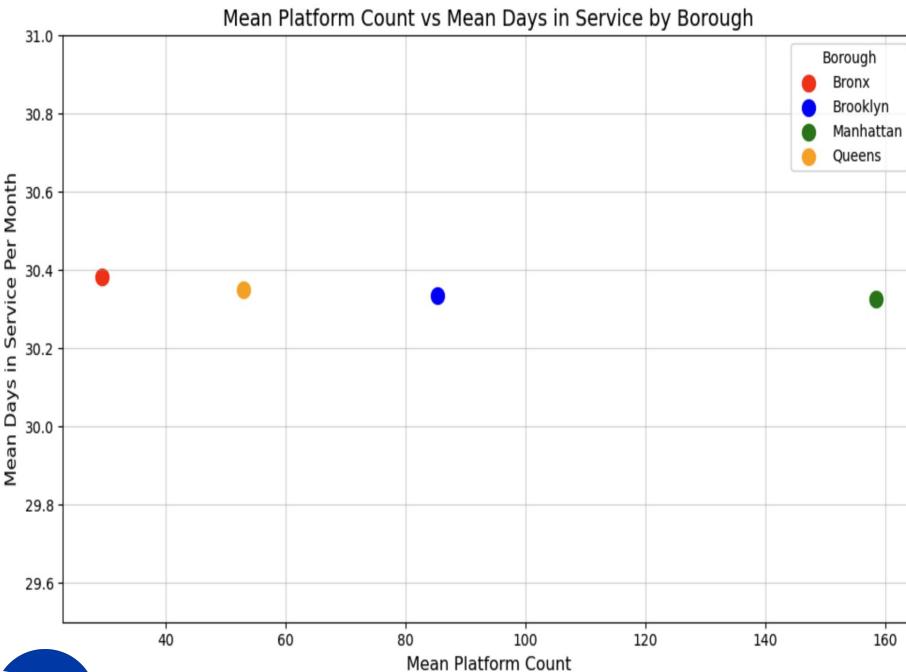


- Ambulatory Disability: disability that affects a person's ability to walk or move around
- Accessibility really does matter, and maybe even more so in certain regions
- The nearly 10% of the population of the Bronx has an Ambulatory Disability!





Availability of Accessible Stations is Good and Constant Throughout Boroughs

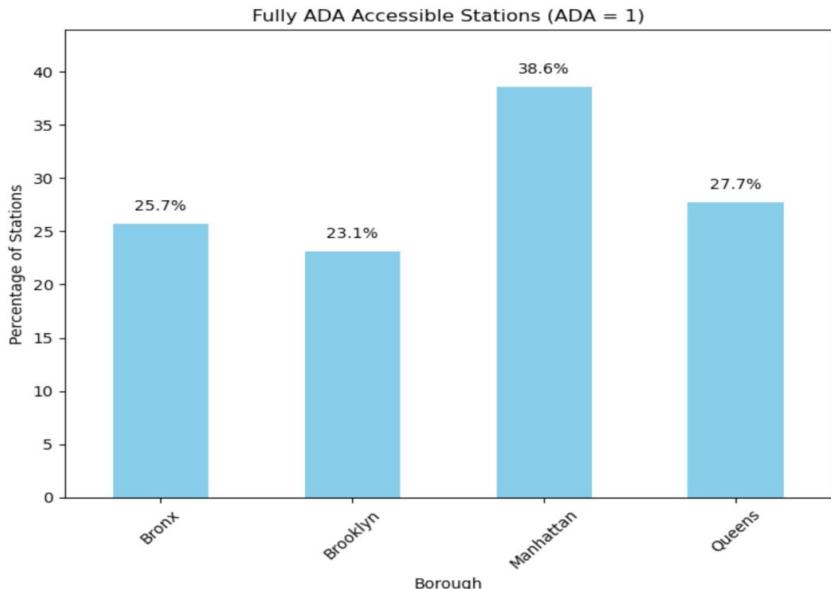


- Here we see the average number of days in a month that the **accessible** platforms were **actually made available** throughout 2022-2024 for each borough, along with their mean **platform** count.
- Shows no significant difference between borough.
- Number of days that the average accessible station in a given borough is available remains consistent.





No Borough has Even 40% of its Stations Fully ADA Compliant



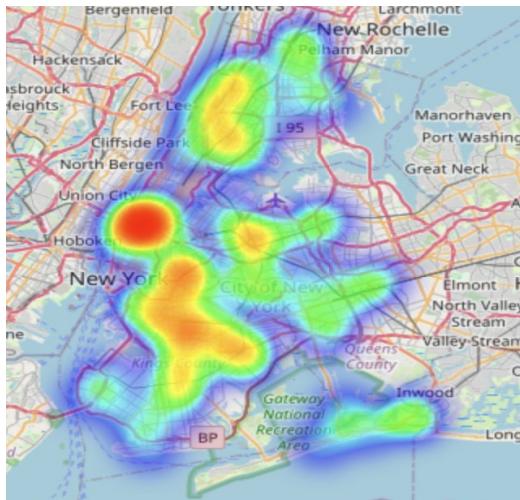
- Greatest proportion of full compliance is 38.6% and the minimum is 23.1%. No borough has even 50% of its stations fully ADA compliant.
- Out of every borough, manhattan has the greatest proportion of fully ADA accessible stations. Then Queens, then Bronx, and then Brooklyn coming in last.
- If we consider partially compliant, the proportions remain very similar.



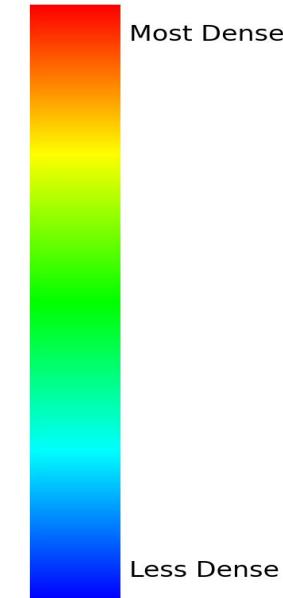
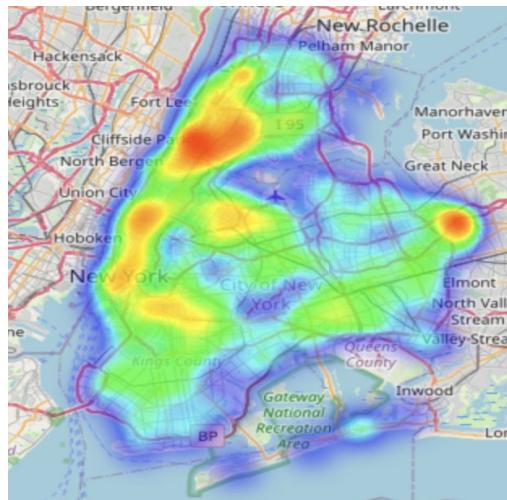


Density of ADA Stations and Density of Those Most Impacted do not Match

Density of ADA Compliant Stations



Density of Those with Ambulatory Difficulty



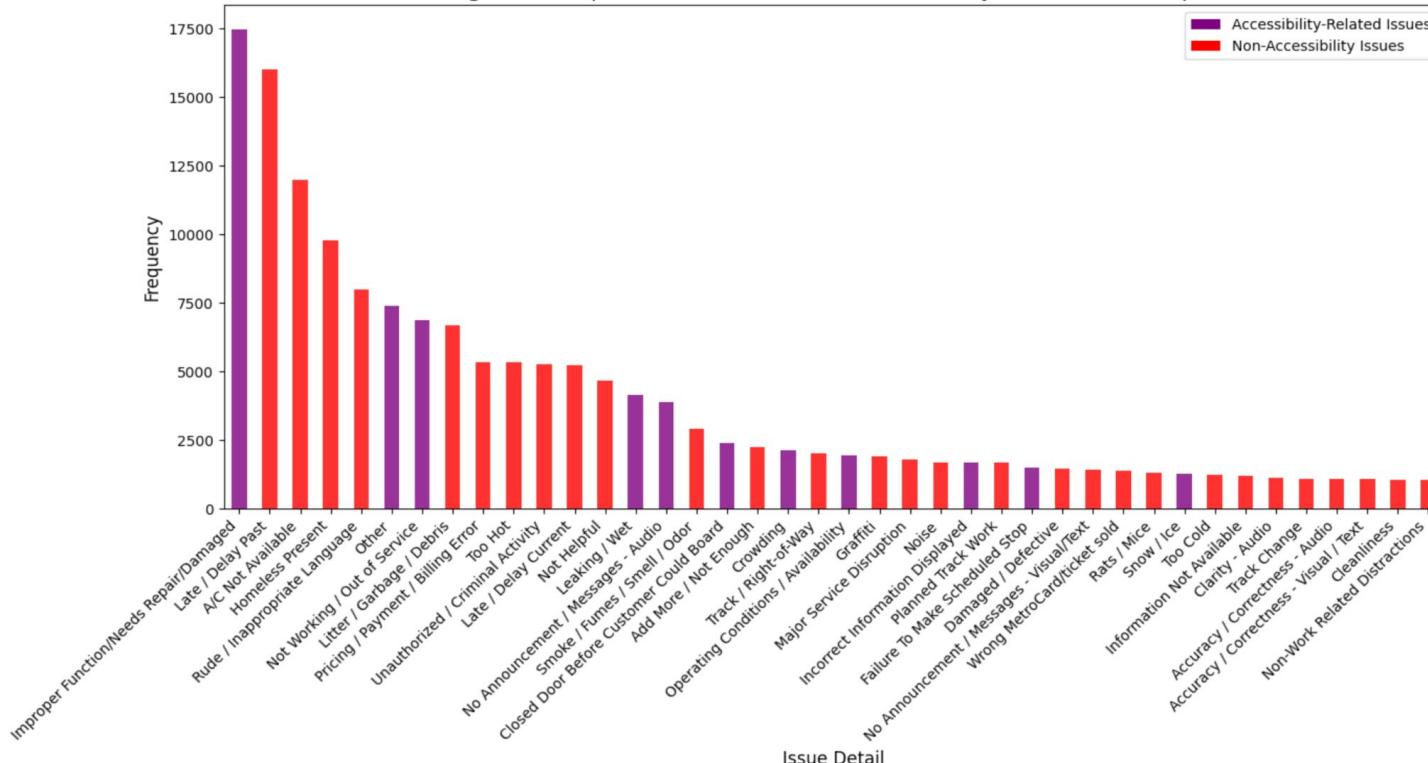
- Alignment in some areas, but not in others
- MTA can increase accessibility in the most vulnerable and underserved areas.





Despite Availability of Accessible Stations, There are Additional Barriers

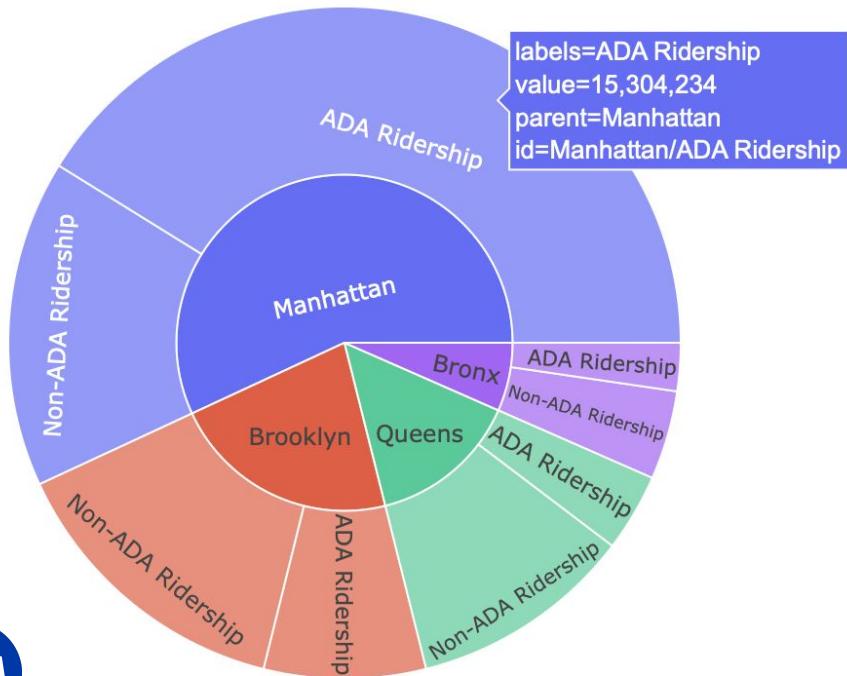
Histogram of Top 40 Issue Details from all Subway Consumer Complaints



- Many of the most prominent complaints make disabled life far more difficult.
- The same technique from previous slide with target list.
- The histogram leaves out 14.83% of the total area.



Manhattan's ADA Stations Cause it to Dominate Ridership at ADA stations

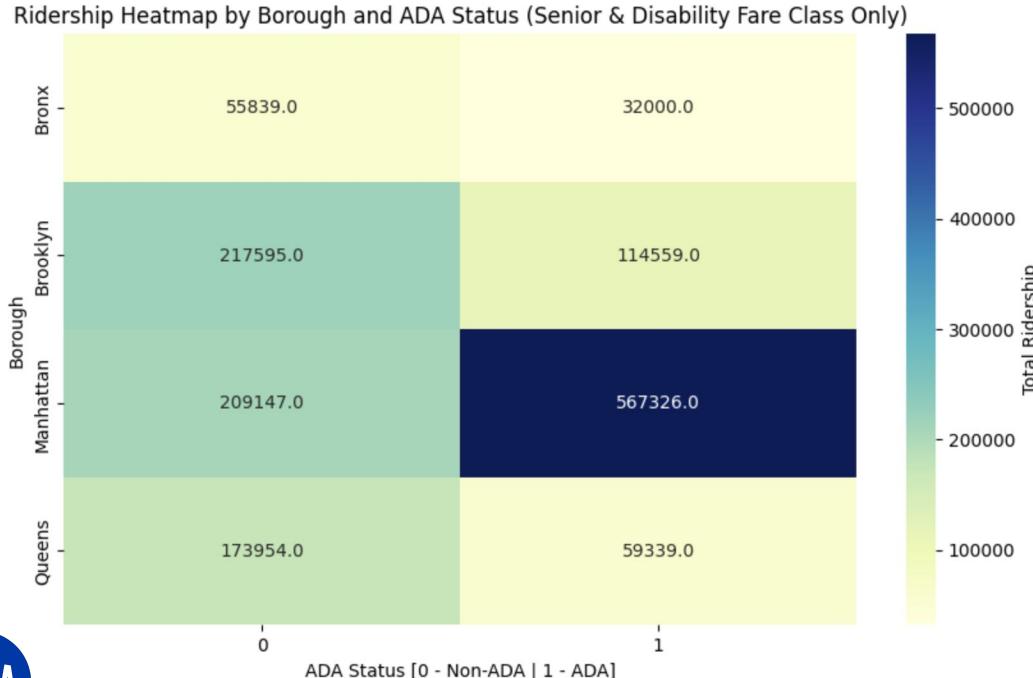


- We observe that Manhattan has a higher amount of ADA stations used by riders than all other boroughs
- Remarkably it has more ADA ridership than ridership at non-ADA stations





Brooklyn & Queens show Opportunities to Enhance Accessibility for Seniors & Disabled Riders

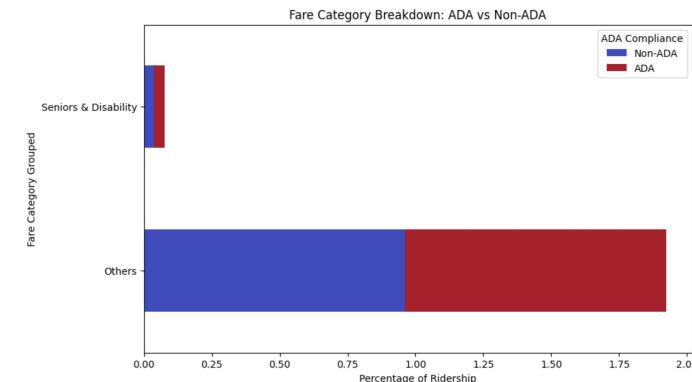
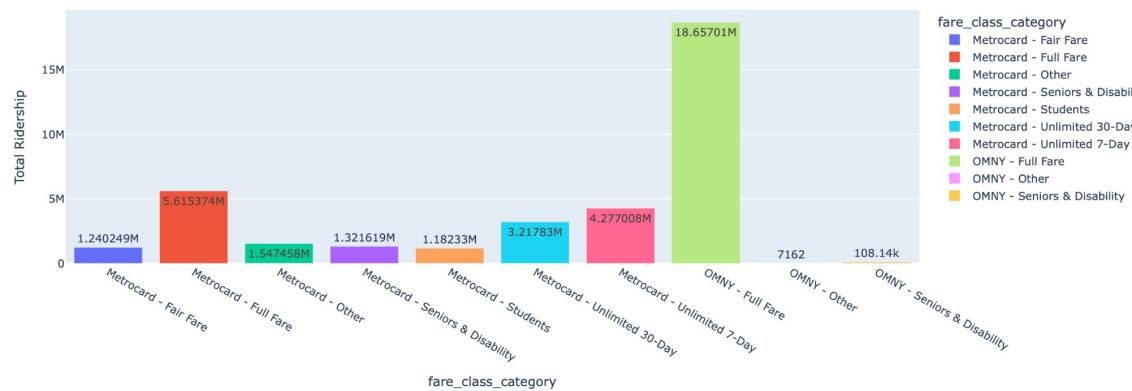


- Manhattan's ADA facilities are highly utilized by seniors and disabled individuals, highlighting its infrastructure's accessibility.
- Brooklyn and Queens may benefit from increased ADA-compliant infrastructure or outreach programs to promote accessibility.
- Bronx shows more balanced ridership, but ridership at ADA stations remains slightly lower.



A Large Amount of People Use Senior and Disability Metro Cards

Total Ridership by Fare Class Category

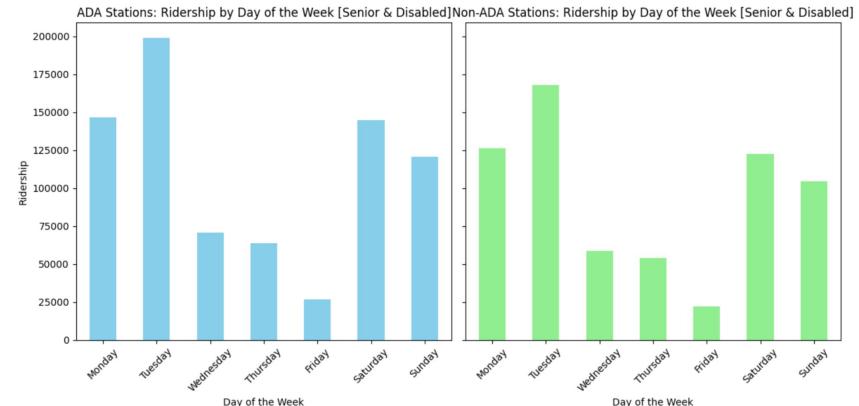
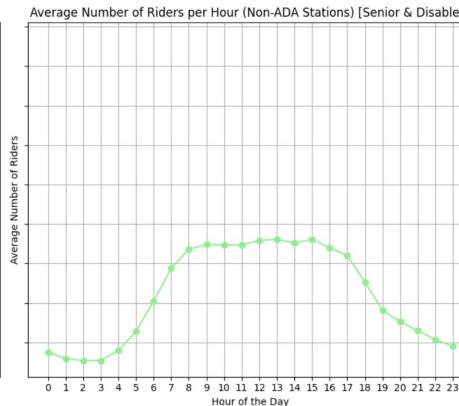
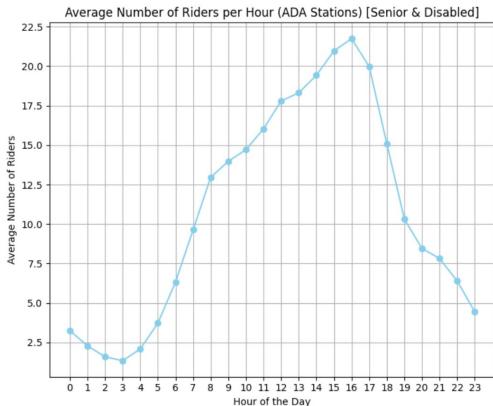


- While OMNY - Full Fare has taken over as the main source of overall ridership there were ~1.4M total Senior and Disability metrocard swipes Sep 2023 - June 2024
- Still there is a big disparity in the number of people who use a Senior & Disability card versus other card types





ADA Stations are key for work-related travel, Non-ADA Stations for non-work travel among Senior & Disabled Riders

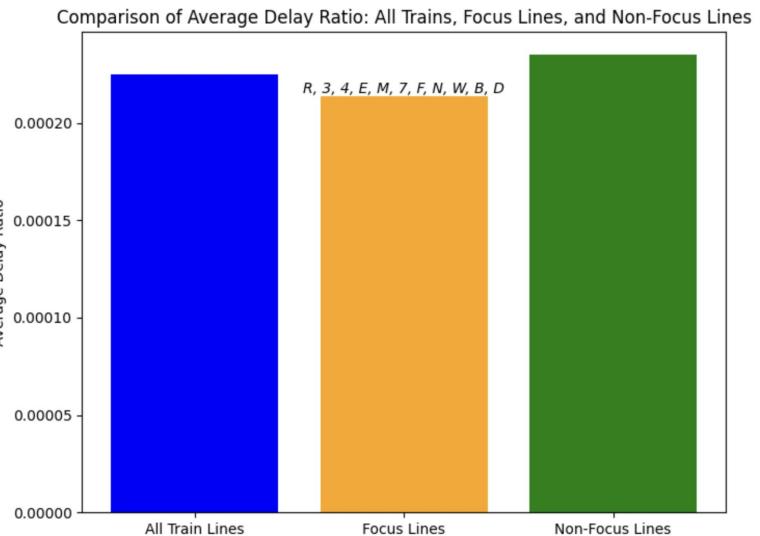
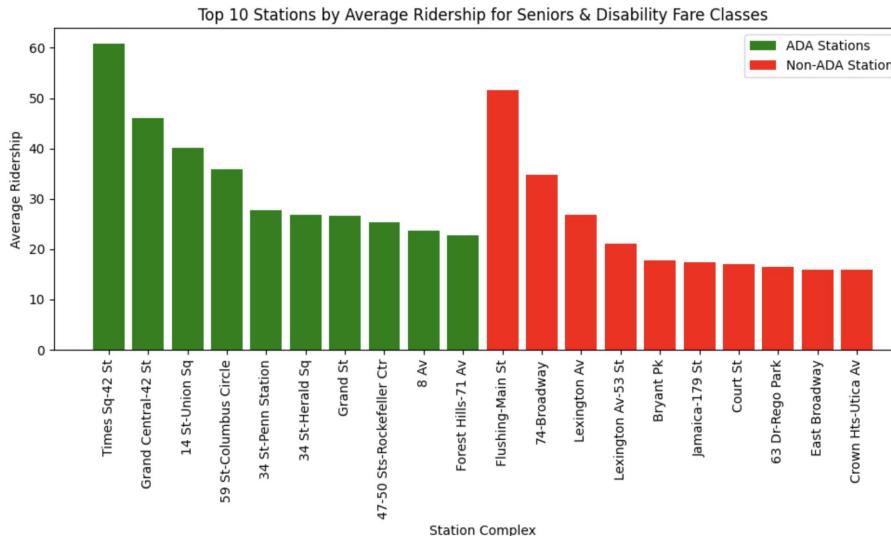


- ADA Stations: Senior & Disabled population Ridership peaks during evening commuting hours (5–6 PM), indicating heavy reliance on ADA stations for work-related travel.
- Non-ADA Stations: Senior & Disabled population Ridership is steadier throughout the day, peaking slightly from 10 AM to 3 PM, suggesting usage for non-work activities.
- ADA stations are critical during peak hours, while non-ADA stations see broader daytime usage by seniors and disabled riders.





Train Lines passing through Non-ADA stations do not contribute much to delays



- We can observe that the delays_ratio in Lines of the top 10 Non-ADA stations travelled by Senior & Disabled people is less compared to other lines and overall delay_ratio. Which means accessibility in terms of delay in these stations are pretty good.



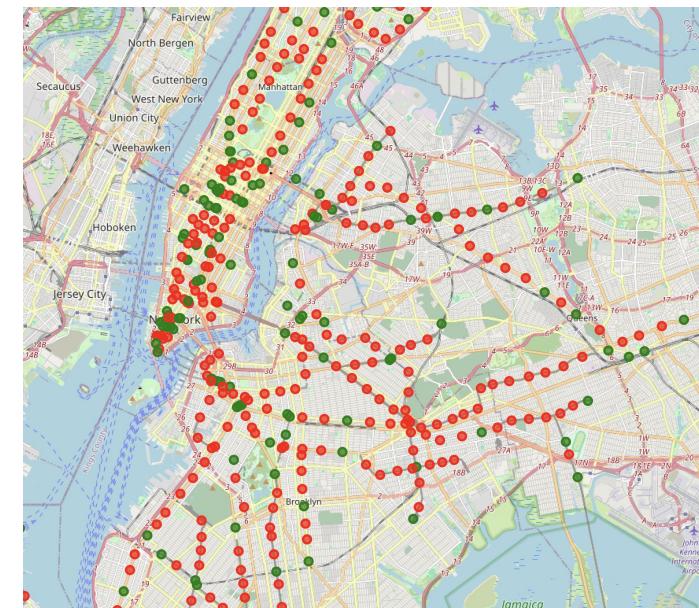
Accessibility: Takeaways & Recommendations

Takeaways

- Many New Yorkers live with ambulatory disabilities.
- While some ADA-compliant stations exist and are available, they are too few, often far from vulnerable populations, and when they are nearby, they are not always usable.
- Senior & Disabled people ridership is increasing

Recommendations

- Prioritizing and increasing ADA-compliant stations in areas with high disability density
- Champion ADA compliance throughout NYC
- Improve the functionality of stations to keep them accessible and useable



“

Thank You

Appendix

-
- References
 - Additional Figures

Data Sources Full List

Data Sources

1. NYS Transportation Data - <https://data.ny.gov/Transportation>
2. NYC Health and Safety data - <https://data.cityofnewyork.us/>
3. GitHub <https://github.com/erikgregorywebb/nyc-housing/blob/master/Data/nyc-zip-codes.csv>
4. US Census <https://data.census.gov/table/ACSST5Y2022.S1810?t=Disability&g=040XX00US18&y=2022>
5. Open NY <https://data.ny.gov/widgets/i9wp-a4ja>

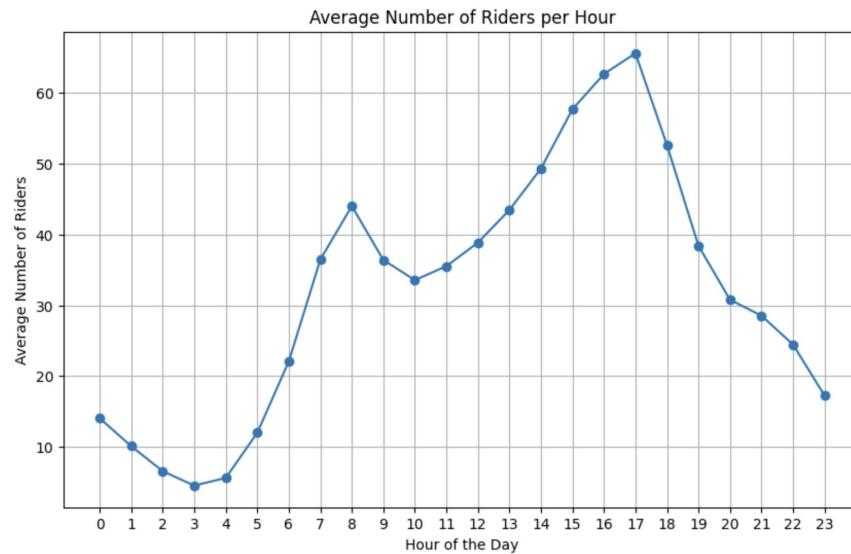
MTA Subway Datasets Used

- Hourly Ridership (1)
- MTA Feedback Data (1)
- Accessibility Data (1)
- Subway Wait Assessment (1)
- Trains Delayed (1)
- Major Incidents (1)
- Entrances and exits (1)
- MTA ADA Stations (5)

Outside Datasets Used

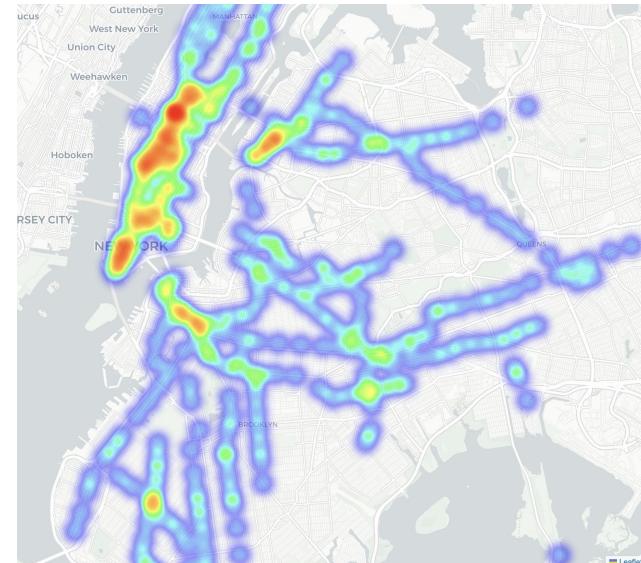
- Zip codes in NYC (3)
- Disability Characteristics/Distribution/Zip Codes (4)
- ADA Stations (5)

Ridership Peaks at 5PM and in Manhattan



We see spikes in ridership for going and coming from jobs/school

Heat Map of Ridership across NYC



There are gaps in station placement and usage especially in Queens

Delays Methods and Data



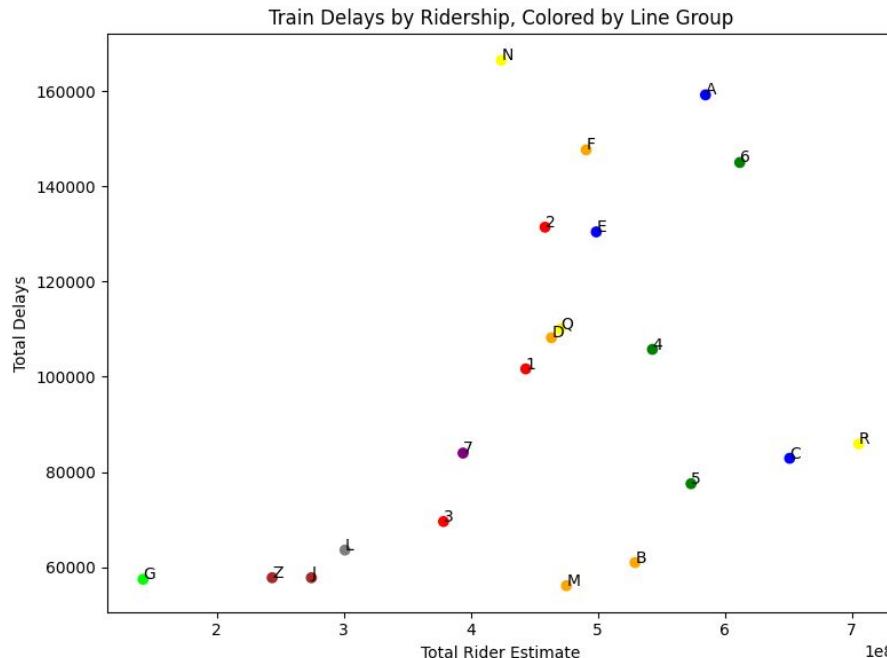
MTA Subway Datasets Used

- Trains Delayed
- Hourly Ridership
- Major Incidents
- Entrances and exits

Top Questions

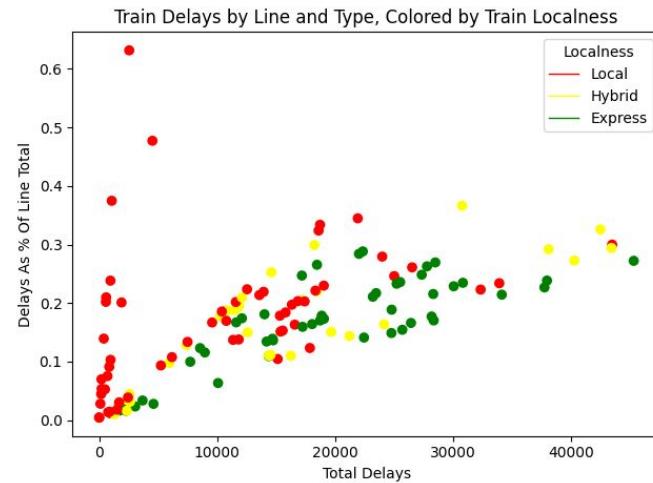
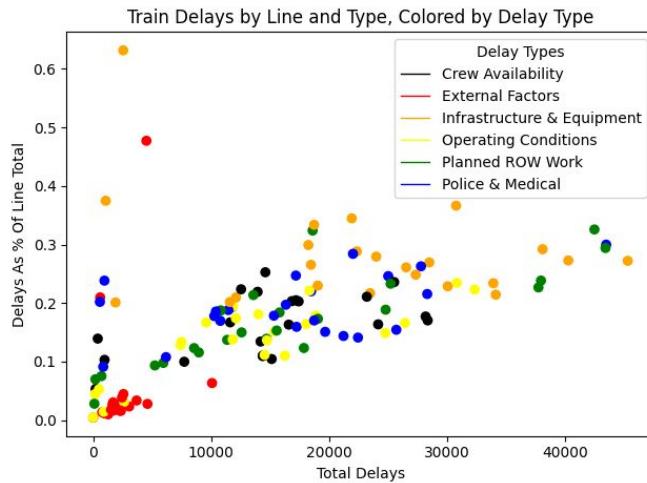
- Which Subways and Groups experience the most delays?
- Do these also experience the most major delays?
- Is there a relationship between ridership and delays?
- Do any delay categories stand out?

Broken out by Line, Ridership and Total Delays Correlate Better

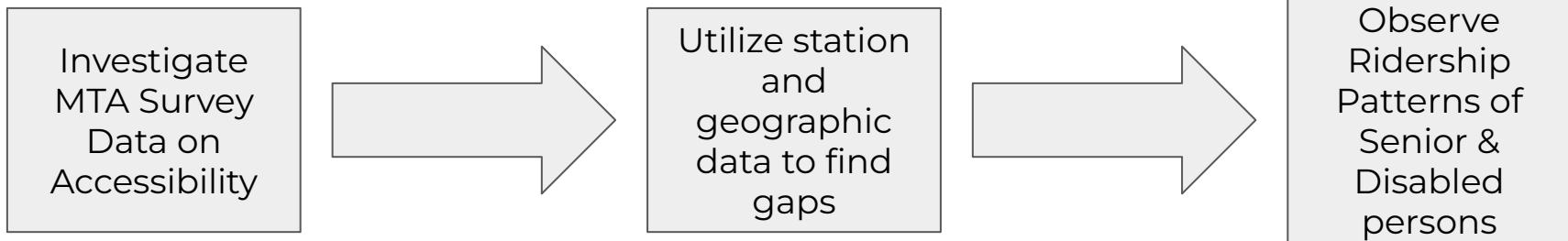


- We estimated ridership by allocating all swipes at a station to all trains at that station (biased to manhattan and longer lines)
- The linear regression on this data returned an r of 0.42
- Notice that the N stands out for most delays and the G stands out for its amount of delays despite its ridership
- We tested removing the G train and it made the correlation worse ($r = 0.34$)

Plots of Delays Including GS and S trains



Accessibility Methods and Data



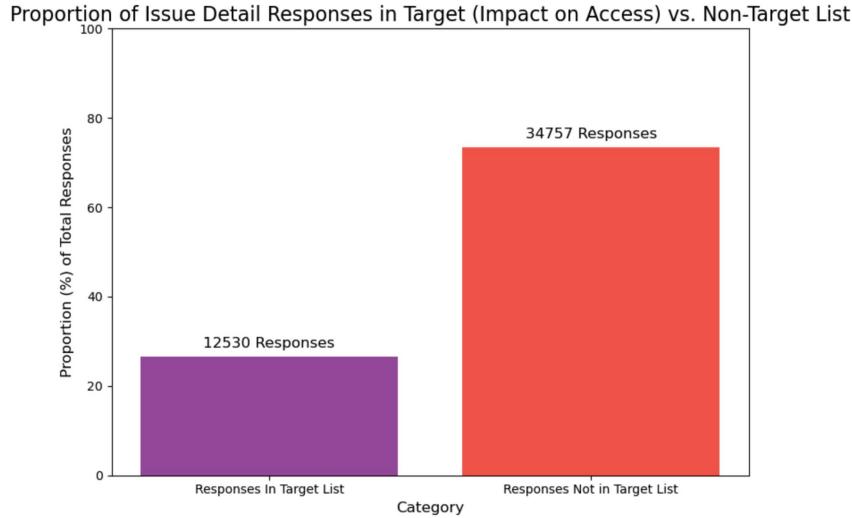
MTA Subway Datasets Used

- Hourly Ridership
- Subway Stations ADA
- MTA Feedback Data
- Accessibility Data
- And several more MTA and outside sources

Top Questions

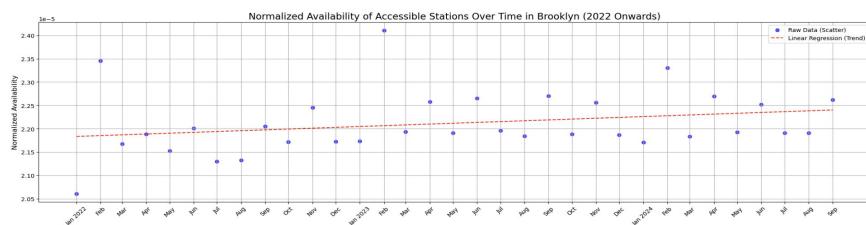
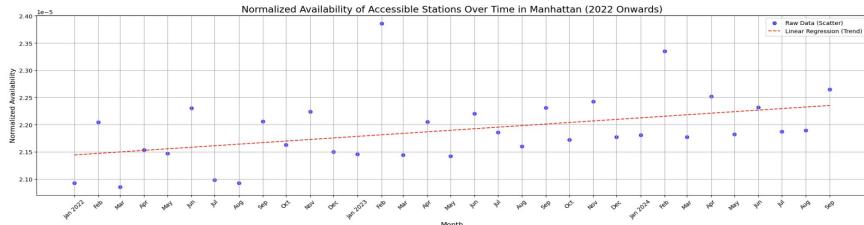
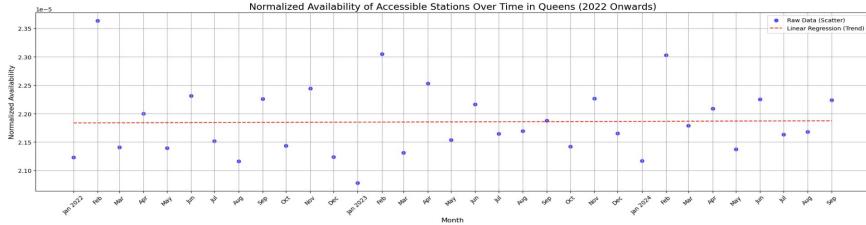
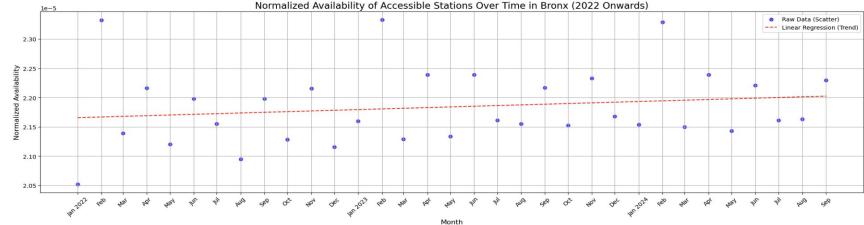
- What is the current state of accessibility in the Subway?
- Is accessibility different between boroughs?
- Are there usage difference between ADA and non-ADA stations?
- Which areas are underserved?

Feedback Data Had Many Accessibility Related Complaints



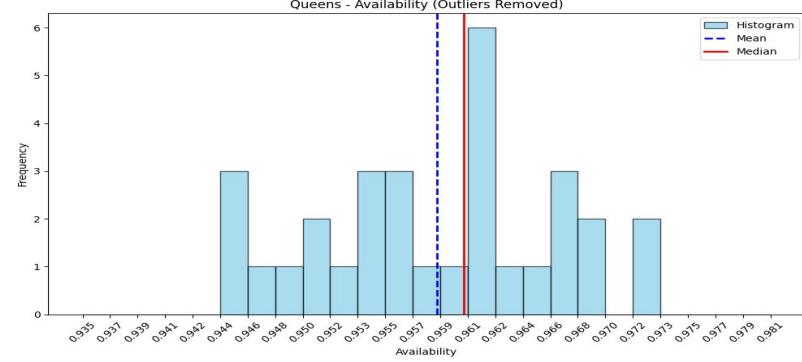
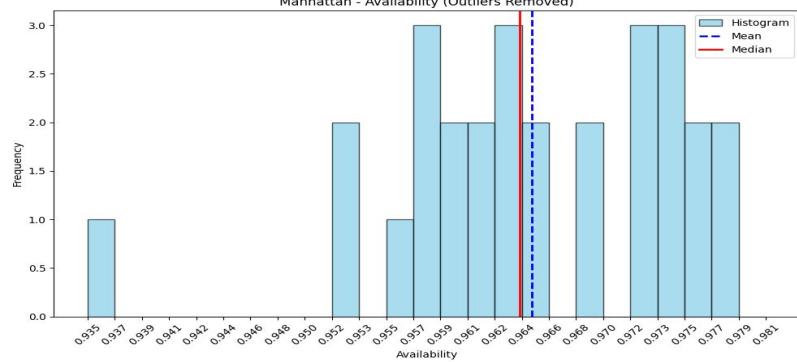
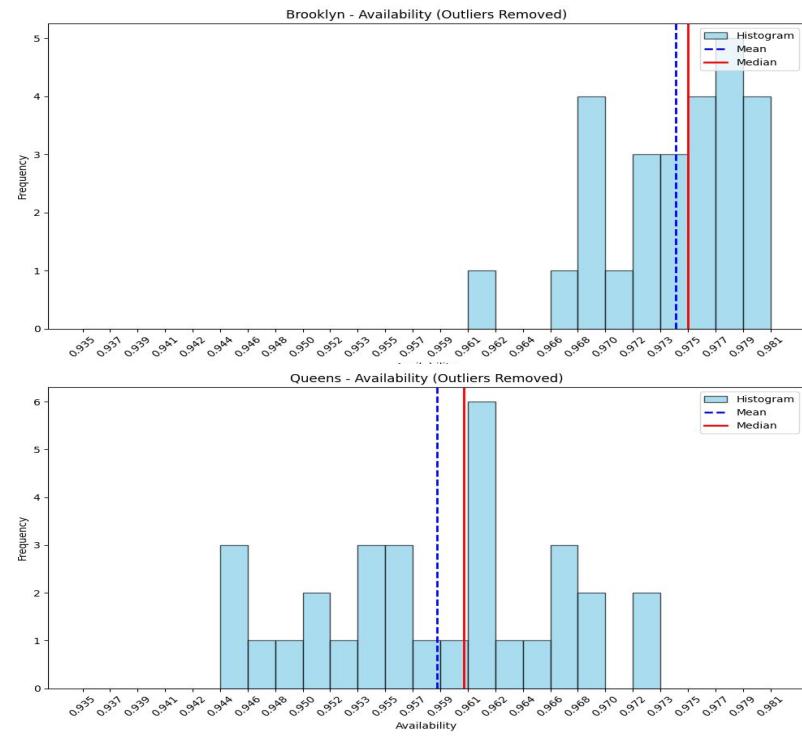
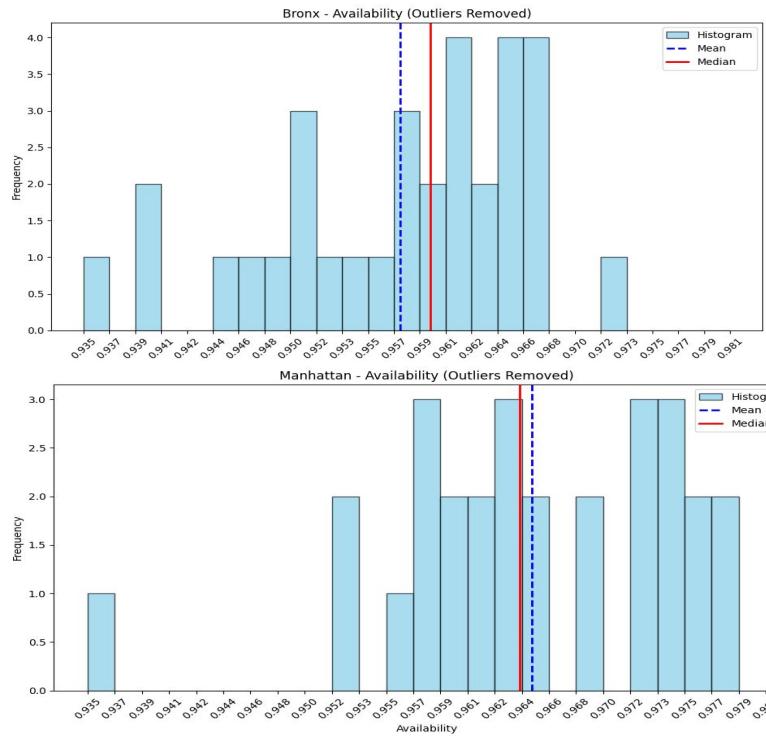
- Histogram to visualise frequency of complaints versus others. Filtered feedback for complaints that impact accessibility most based on my domain knowledge
- Includes only complaints from consumers about the operating conditions of the train and platforms
- Target list only includes complaints that are most likely to impact accessibility. For example, ['Needs Repair / Damaged / Not Working', 'AutoGate not accessible', 'Obstruction',...]
- Bias because of certain confounds. For instance, disabled individuals make up only 7% of the population of NYC. If survey was collected properly, majority of survey respondents did not have an ambulatory difficulty.
- Limited in showing importance of these issues to the disabled community, however, does well in explaining the importance of these issues to the entire consumer community
- If these issue matter to average consumer, they definitely matter to those with accessibility needs

Accessible Station Availability Rate did not Significantly Change from 2022-2024

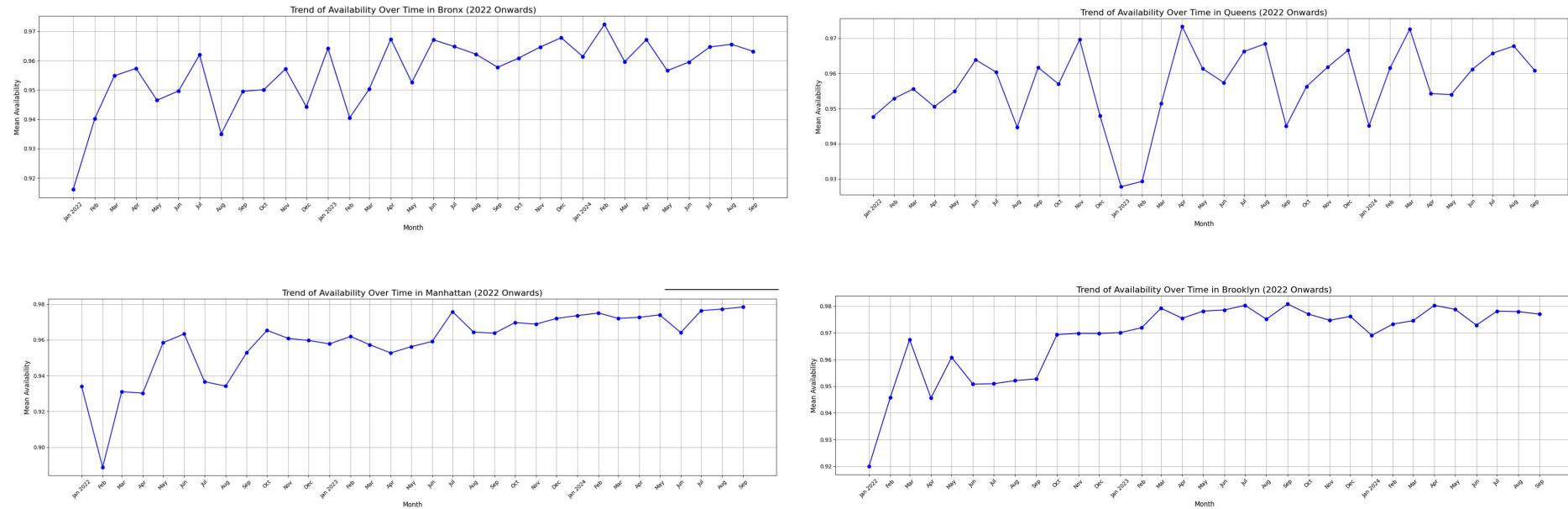


- Normalized availability remains almost constant across most boroughs
- No significant trends over time.
- Manhattan shows a slight increase ($R^2=0.1846$), while other boroughs exhibit minimal variability (R^2 ranging from 0.0003 to 0.0656).
- Availability is not decreasing, at least not within the last 2 years. Only slightly increasing.

Availability of Accessible platforms by borough



Change of Availability of Accessible stations over time

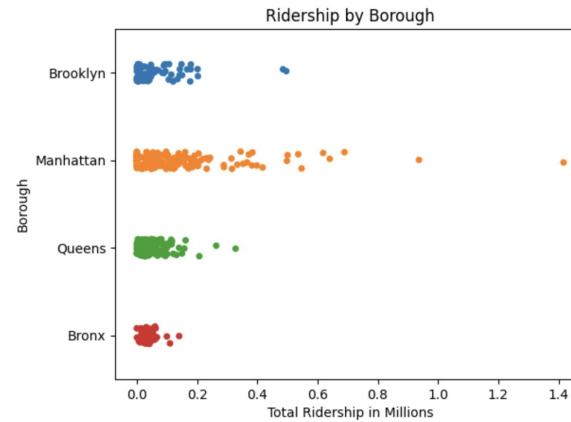
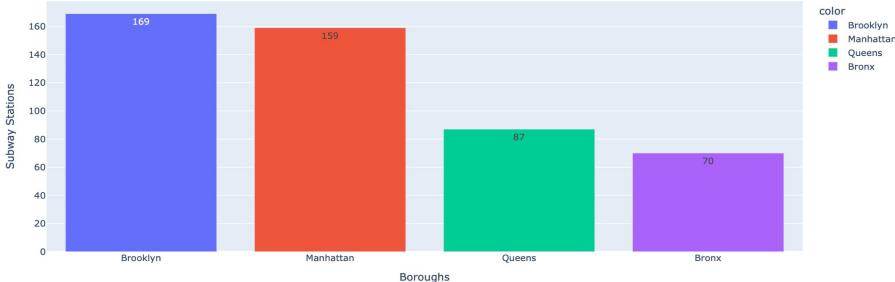


There is More Ridership in Manhattan but Fewer Stations Compared to Brooklyn

Subway Ridership by Borough

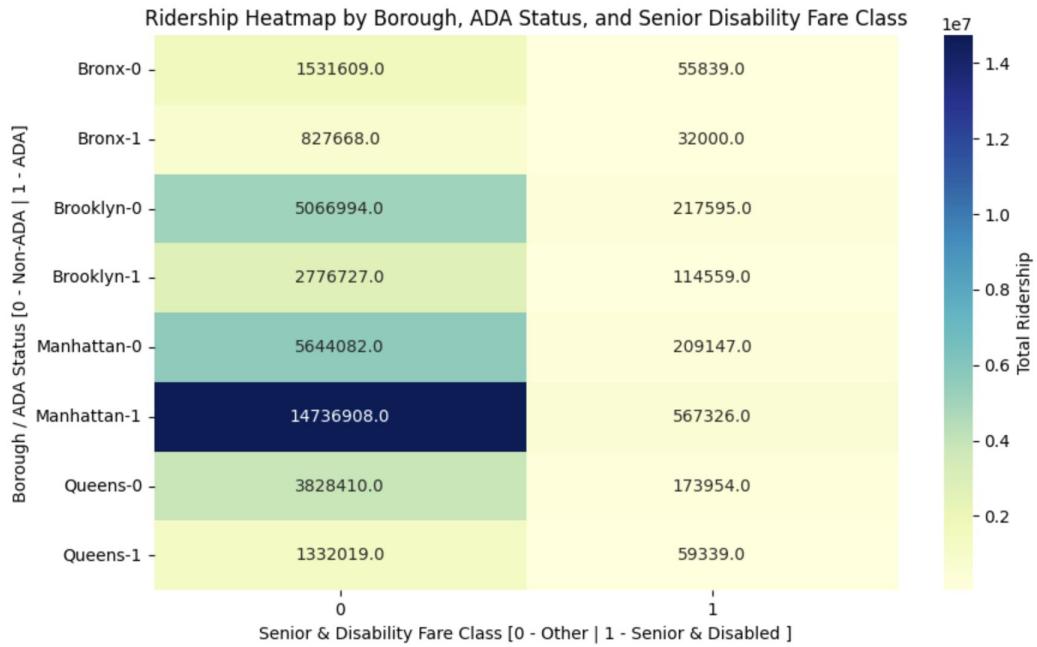


Number of Stations by Borough



- The strip plot reveals that each borough has a high concentration of stations with relatively low ridership, while a few stations stand out with significantly higher ridership.
- Notably, the Times Sq-42 St station demonstrates a remarkable disparity in ridership compared to other stations.

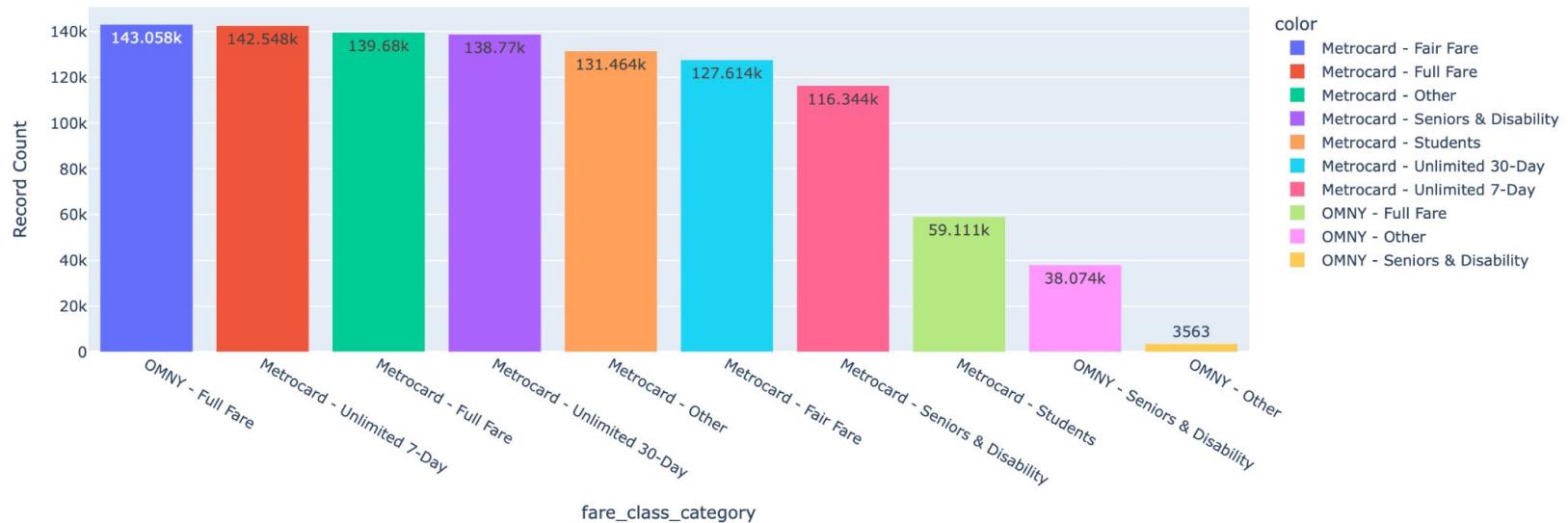
Heatmap Insights: Ridership by Borough, ADA Status and senior Disability Fare Class



- ADA-compliant ridership with senior/disabled fare class is significantly higher in Manhattan compared to other boroughs.
- Non-ADA ridership for the general population ("Other Fare Class") is relatively high across boroughs like Bronx and Manhattan.
- Brooklyn and Queens show more balanced ridership values across ADA statuses and fare classes.

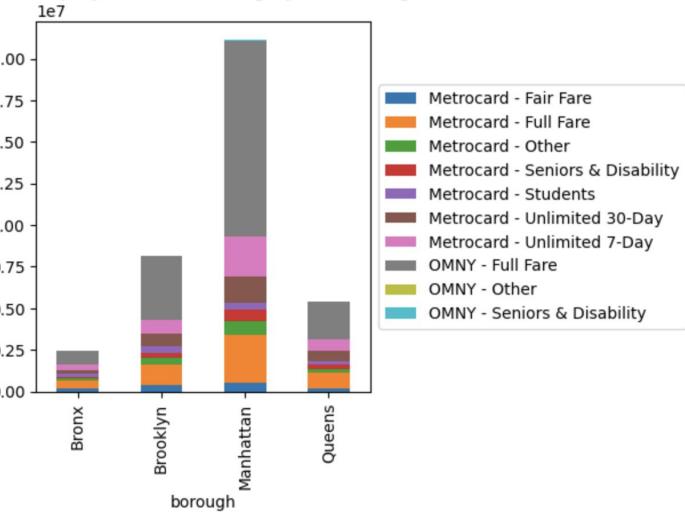
A Large Amount of PEople Use Senior and Disability Metro Cards

Number of Records by Fare Class Category

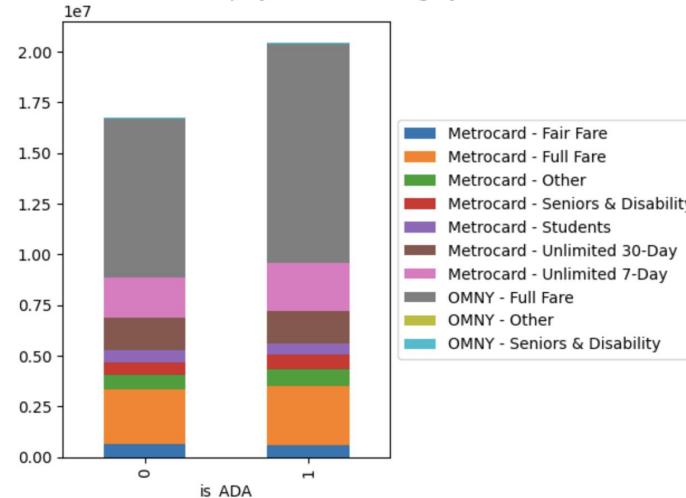


A Large Amount of People Use Senior and Disability Metro Cards

Ridership by Fare Class Category and Borough

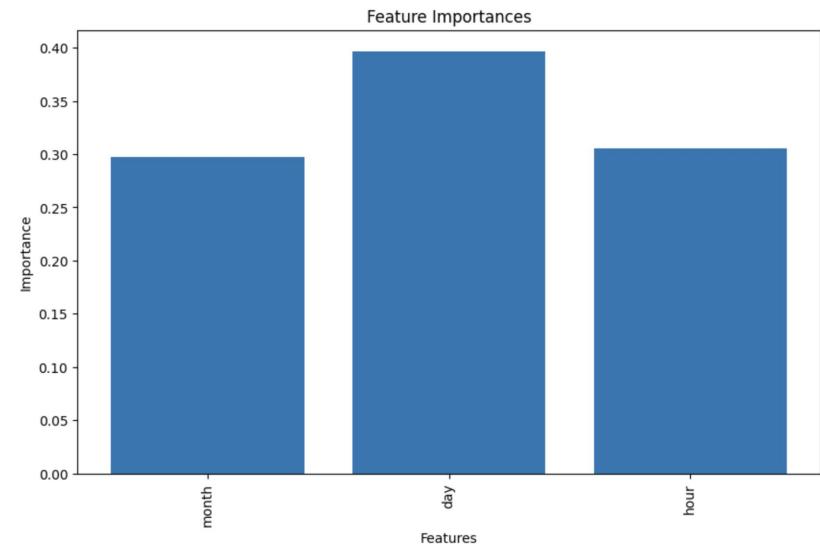
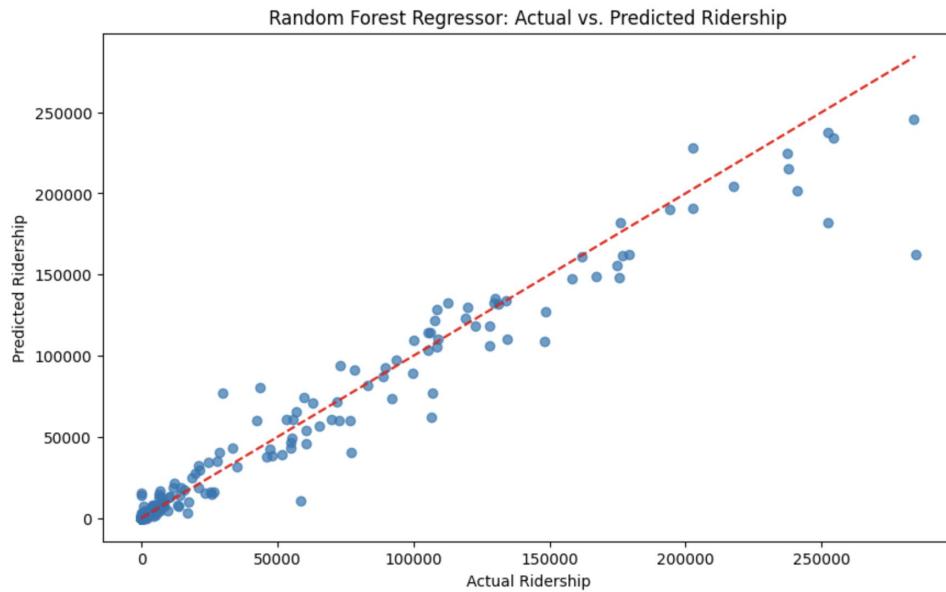


ADA vs Non-ADA Ridership by Fare Class Category



- Senior and disabled fare categories contribute more significantly to ridership in **ADA-compliant stations** compared to non-ADA stations.
- The majority of senior and disabled riders use **Metrocard Seniors & Disability** fares over other payment methods like OMNY.

Using Machine Learning (Random Forest Algorithm) for Disabled Ridership Prediction



We can see we got an impressive R^2 score of 0.96. This indicates that our model has potential to fit the data. This was only based on three factors. Unfortunately, the model has a mean absolute error of around 1797.23 riders. However if more factors were added to the model, it could potentially be improved, such as weather, unique stations, holidays, etc.

Ridership Growth between ADA vs Non-ADA station

