

Customer Satisfaction on the MTA

Santiago Arango, Tanay Shah, Alexander Zhen, Deekshita Madharam

The topic we have chosen is the MTA system in NYC, focusing on its efficiency in being run, how well it meets users' expectations (comfort, delays, and safety) as well as ridership levels. Through this research, we hope to understand what causes passengers' satisfaction or dissatisfaction with the MTA service. This is a very intriguing topic to cover because of the vast number of datasets that are available in this sector/field. Given the MTA is public transportation, any surveys, tracking methods, and related information about the transportation system are publicly accessible for analysis.

Reference papers:

- <https://journals.sagepub.com/doi/abs/10.3141/1735-16>
- <https://link.springer.com/article/10.1007/s40864-016-0042-0>

Images:



Datasets:

data.ny.gov/Transportation-Customer-Feedback
<https://catalog.data.gov/dataset/mta-subway-stations>

Dataset 1:

To delve deeper into our topic on the Metropolitan Transportation Authority system in NYC, we chose a dataset that originated from the New York State open database (data.ny.gov). Specifically, this dataset is sourced from the MTA open data program/portal and contains 600,000+ rows/observations of MTA customer feedback data submitted by MTA transit riders spanning from 2014 to 2019. Each observation contains 8 columns: Agency, Commendation or Complaint, Subject Matter, Subject Detail, Issue Detail, Year, Quarter, and lastly Branch/Line/Route (Data Dictionary is shown below):

Data Label	Data Type	Data Description
Agency	Text	The agency referenced in the e-mail; valid values are: <ul style="list-style-type: none">• Long Island Rail Road

		<ul style="list-style-type: none"> • Metro-North Railroad • NYC Buses • Subways
Commendation or Complaint	Text	Type of Issue (complaint or commendation) referenced in the email; valid values are: <ul style="list-style-type: none"> • Complaint • Commendation
Subject Matter	Text	Subject matter referenced in the email; valid values are: <ul style="list-style-type: none"> • Buses • Commendation • Complaint • Customer • Employees • Ferry Service – Hudson River • MetroCard/Tickets/E-ZPass & Tolls • MTA Agency Cars / Trucks • Policies, Rules & Regulations • Public Hearing • Schedules / Reservations • Station/Bus Stop/Facility/Structure • Telephone / Website / Mobile Apps • Trains • Travel Disruption / Trip Problem
Subject Detail	Text	The specific area of service referenced in the email
Issue Detail	Text	The feedback, pre-defined response, or category of the area of service
Year	Integer	The year that the complaint/commendation was entered
Quarter	Integer	The quarter of the year in which the issue occurred
Branch/Line/Route	Character	The branch/line/route of transit that the issue occurred on

Dataset 2:

The second dataset we plan to utilize in this analysis is a dataset derived from the federal government open database (data.gov). This dataset lists all 496 MTA subway stations with the following attributes: Station Master Reference Number (MRN), Complex MRN, GTFS Stop ID, the services that stop there, the type of structure the station is on or in, whether the station is in Manhattan's Central Business District (CBD), and their ADA-accessibility status. This subway station dataset is frequently maintained by NY Open Data and was last updated on August 23, 2024.

Analysis:

In order to begin our analysis, we must first organize and clean the data by removing empty rows and handling other missing values. If the dataset still contains an excessive number of rows, we may narrow down our focus on a specific agency (NYC buses or subway for example), a specific time period, certain types of concerns, or even a certain branch/line/route depending on the dataset's size.

A possible analysis of the dataset regarding customer feedback is creating a line graph throughout the 4 years to show the number of complaints per year and how it has changed over time. Additionally, creating multiple line graphs that overlap each other where each line graph represents a different type of complaint will give the readers a way to see what type of complaints are the most prevalent and what type of complaints have increased/decreased significantly in the past couple years.

For another type of static graph, we could track what train lines experience the most problems and can make a grouped bar graph out of that resulting graph for each train route depending on the size of the analysis. This would allow us to compare the issue frequency across different routes and discover which types of issues are most common to specific routes.

Another analysis we could perform is identifying which complaints are the most common throughout the years and breaking them down based on what time of year they are the most prevalent, based on what quarter the complaint is listed in.

As for an interactive visualization, a potential idea would be to use geospatial visualization, where the user will be able to click on each subway station on a map of New York City to see which complaint type is the most prominent at that location.

Another possible analysis would be to have an interactive dashboard where the user would be able to select a year, station, or line to observe a breakdown of customer feedback.

Group Member Duties:

As a group, we plan to meet twice a week at least to go through our ideas to ensure everyone is on the same page as to what must be done. Individually, we'll each create one visualization that is relevant to the topic (3 static and 2 interactive) and will each have our own description of the visualization, included in our final presentation. After completing our visualizations, we will all meet up to create our PowerPoint, visual demonstration of our interactive graphs, and website to connect them all. Through our weekly meetings, we will figure out the tasks for the week and work on our individual parts of the website and then push the code frequently to GitHub so everyone else can work on them as well.