

Project Proposal MTA

Abstract:

The New York subway MTA turnstile data is a series of data files containing cumulative number of entries and exits by station, turnstile, date and time. Data files are produced weekly, data records are collected typically every 4 hours with some exceptions.

Question/need:

Question:

- How heavy traffic station could influence processes of people getting in and out of station?
- Do certain subway station get visited more often than others? Where are these stations located?
- Are there more MTA riders during certain days of the week? (Workdays vs. weekends?)
- Are there more MTA riders during certain hours of the day? (Rush hour vs. late night?)

Benefits:

The main benefit is more to the company however, the analysis could result in positive impact to people by improving the traffic process.

Data Description:

Dataset will use:

Data from Turnstile MTA involving the number of entries and exits at various turnstiles around each subway station in NYC .

In this analysis we use data from 2021. Data size is over 5 million.

The plan in this project is to use time interval as individual sample/unit of analysis. Furthermore, using the time interval will help in calculating traffic per turnstile for each station might be able to help MTA better design subway entrances.

Tools:

I will use python because it contains libraries (pandas, Matplotlib, NumPy ,etc.), also for exploratory analysis and data cleaning will use SQLite or PostgreSQL .