



EDA (EXPLORATORY DATA ANALYSIS)

Of MTA Turnstile data New York CITY

Mada Abudahish
September 29, 2021

I. Background

The world is experiencing a once-in-a-lifetime pandemic that necessitates controlling and limiting the capacity in public and crowded places. We keen to preserve the health and safety of citizens and residents from the risk of the spread of novel coronavirus. Our task was to use New York City MTA Turnstile data and cooperation with Ministry of Health to avoid crowded place and to provide people with vaccination centers and provide enough space in each station.

II. Question:

- Are there more MTA riders during certain days of the week? (School days vs. weekends?)
 - Are there more MTA riders during certain hours of the day? (Rush hour vs. late night?)

III. Data Description:

In this analysis we use data from MTA Turnstile and we chose to focus on first half of 2020 after lifting many of its remaining safety COVID-19 restrictions.

Data files are produced weekly, data records are collected typically every 4 hours
Variables included in initially processed data:

- **C/A** = Control Area (e.g., A002)
 - **unit** = Remote Unit for a station (e.g., R051)
 - **SCP** = Subunit Channel Position represents an specific address for a device (e.g., 02-00-00)
 - **Station** = Represents the station name the device is located at
 - **date** = Represents the date (MM-DD-YY)
 - **time** = Represents the time (hh:mm:ss) for a scheduled audit event
 - **desc** = Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)
 - **entries** = The cumulative entry register value for a device
 - **exits** = The cumulative exit register value for a device

IV.Tools:

Data:



Visualization:



Data