# Exploratory Data Analysis Project Proposal

Josh Jingtian Wang

1/9/2022

**Question/Need:**

The COVID-19 pandemic has been devastating to many sectors of the society – public transportation being of them. In this project, the Metropolitan Transportation Authority (MTA) has asked us to investigate when and where they can reduce the number of trains to reduce their budget. I will explore whether the pandemic, the rise of for-hire vehicles (Uber, Lyft, etc) and weather have potentially played a part in the fluctuation of MTA ridership. The MTA will then use my findings to make changes in their train schedules.

**Data Description:**

The primary dataset is the MTA turnstile data of New York City: <http://web.mta.info/developers/turnstile.html>

I will also download the NYC weather data from the National Oceanic and Atmospheric Administration (NOAA) website: <https://www.ncdc.noaa.gov/cdo-web/search>

Rideshare services data will come from the NYC Taxi & Limousine Commission (TLC): <https://www1.nyc.gov/site/tlc/about/aggregated-reports.page>

The rows will be the data (day or month)

The columns will be MTA turnstile entrances, Uber/Lyft trips provided, rain/snowfall.

**Tools:**

Data retrieval will be done via SQLite. Data cleaning will be done via Python’s pandas module. Data visualization will be done with Python’s matplotlib and seaborn modules.

**MVP Goal:**

I will plot 2019’s MTA turnstile entrances, Uber/Lyft trips data vs time.