### Comparing the impact of COVID-19 on MTA ridership and NYC business activity

By: Alexander Amado

### Abstract

The goal of this study was to understand the relationship between MTA turnstile activity and business activity in NYC. I worked with data from the Metropolitan Transportation Authority, Department of Consumer Affairs, and Department of Health and Mental Hygiene. After cleaning and studying the data, I built visuals using matplotlib and seaborn to help show the changes in activity during the pre and post pandemic time periods.

### Design

This project idea originated from a Bloomberg article I read on trends in return to office in NYC. Bloomberg created a Pret Index to serve as a proxy for business activity. I was interested in studying business activity as well, and wondered how it would relate to turnstile activity. I downloaded turnstile data, COVID-19 case count data, and legally operating business data into pandas data frames. I cleaned and analyzed the data and then represented my findings visually.

### Data

The MTA data set included 2,707,183 rows of data at the turnstile level and 15 columns representing times of day, stations, lines, entries and exits. The operating businesses dataset consisted of 273,125 rows of data and 35 columns that included items like license status, creation date, industry, and license type. Some of the individual features could be grouped up to industry level or status for a broad analysis of trends. The COVID dataset included 739 rows and 62 columns. The columns of interests were date and case count. Other columns like hospitalization rate, and trends by borough were not a focus of my study.

### Tools

* Pandas for data manipulation
* Matplotlib and Seaborn for plotting
* SQLite, DB Browser for SQLite, and SQLAlchemy for data queries