

Determining Locations for NYC Ballot Drop Boxes

Chan, Natural

9 November, 2021

Abstract

The goal of this project is to determine convenient to place absentee ballot drop boxes by looking at the Metropolitan Transportation Authority (MTA) public turnstile data. We primarily look at the early voting period that begins 10 days prior to election day in order to determine the stations that see the most foot traffic. From there, we identify the closest government owned entity (public libraries, parks, schools, etc.) where ballot drop boxes could be placed.

1 Design

In January of 2021, the New York State Senate passed legislation S.492 to authorize the Board of Elections to establish absentee ballot drop boxes. Using the MTA public turnstile data, we want to identify the stations with the highest foot traffic which will help minimize the number of ballot drop boxes needed and increase the voter turnout without requiring individuals to go far out of their way to vote.

2 Data

The MTA turnstile data is stored in separate weeks so we need to determine which range of dates will be most effective for our analysis. Election Day occurs on the first Tuesday that follows the first Monday of November, but we are mostly interested in the early voting period, which begins 10 days before Election Day in New York. We specifically look at the election years: 2016, 2018, and 2020 to get an idea on how foot traffic changes between Presidential and Midterm Elections as well as pandemic and non-pandemic years. Each row of data contains information about each unique turnstile along with the corresponding station, date, time, cumulative entries, and cumulative exits.

3 Algorithms

We first obtain the data we wish to work on from the MTA web page. Since Election Day can occur from the 2nd to the 8th of November, we choose to obtain the data for October and November in the years 2016, 2018 and 2020. Next, we remove any duplicate values and

then split the data into their respective years since the data will be originally be in a single table. The ultimate goal is to determine the stations with the highest foot traffic each year so we change the cumulative values of entries and exits to daily values and then take the sum of the daily entries and exits to get daily foot traffic at each station. Lastly, we remove any dates that do not fall within the 10 day early voting period and plot the top 10 stations for each year.

4 Tools

- Converting data from MTA web page to SQL database
- Querying database into a Pandas data frame using SQLAlchemy
- Pandas for exploratory data analysis
- Seaborn and Matplotlib for visualizations
- Plotly for table visualizations

5 Communication

Presentation slides and plots are located in my github repository.

