Advertising NYC's Electronics Drop-off Boxes To MTA Riders

Maira Asmat

Abstract

The goal of this project was to use MTA Turnstile data along with data on the locations of electronics drop-off boxes in New York City to hypothetically help the city advertise these boxes to areas of high traffic. I analyzed the data in SQL to find the zip codes where these boxes were least likely to overflow by calculating the ratio of daily MTA exits and entries to the number of bins in each zip code.

Design

I based the goal of this project around a hypothetical need the city of New York posed to me, which was effectively advertising the location of these drop-off boxes. It is illegal to dispose of electronics outside of these boxes, and it also harmful to the environment and public health. And as the boxes were only created in 2015, their advertisement is still necessary. The MTA data is a reliable way to determine the movement patterns of NYC, and thus it can be used by the city to extrapolate how many people are within a zip code. As these boxes could potentially overflow, it is necessary to take into account how many people could potentially visit the number of drop-boxes in one zip code, so that ratio must be part of the final recommendation.

Data

11 weeks of turnstile data was obtained from the MTA website, and this contained the names of the stations along with the amount of people that entered and exited turnstiles within a 4-hour period. The 4-hour periods were aggregated by day and by station to find the total amount of entries and exits from a station within a singular day. The drop-box data contained the specific addresses/zip codes of the boxes, and each row of data described a single box. The MTA Zip code data simply contained all stations and their corresponding zip codes, and this was used with the drop-box data to determine how many of each were in each NYC zip code.

Algorithms

The MTA data was uploaded and cleaned in Python using Pandas. Data cleaning included deleting duplicate datasets, dealing with null values, aggregating the data to fit the given prompt, and creating visualizations to better address the prompt. The methodology for the Exploratory Data Analysis was borrowed largely from class lectures and collaboration with cohort members (who are properly credited).

Tools

- Numpy and Pandas for Exploratory Data Analysis
- SQL for additional data analysis
- Matplotlib for data visualizations

Communication

All visuals are embedded within the slides, which are attached in the GitHub repository and will be presented.