

Abstract

The goal of this project is to study the feasibility of [Malone's Cones](#)' business plan to position a network of ice cream carts throughout Manhattan during the summer. The company wishes to maximize profit by deploying ice cream carts in heavily-trafficked areas during the times of the day with the most pedestrian activity, while also avoiding potential competitors such as brick-and-mortar ice cream stores. We inferred estimates of pedestrian traffic from MTA turnstile data identified candidate locations and screened for potential competitors to generate some preliminary recommendations. We also recommend further analysis.

Design

The company's strategy is to build brand recognition by deploying ice cream carts in areas of Manhattan that experience high pedestrian traffic. To build brand recognition, the company would prefer to deploy multiple carts in one or two selected areas rather than positioning one cart per area throughout the entire borough. The company wishes to use subway turnstile data as a means of estimating which areas grow in population throughout the day and which shrink. The company will deploy carts to areas with large and/or growing populations. We will also screen candidate areas for brick-and-mortar ice cream stores that could be competitors.

Data

We drew data for 378 MTA stations from the MTA website for May - September 2019, the warmest months of the year in which ice cream is most likely to sell well. The data set contains time-resolved cumulative counts for the number of entries and exits for each turnstile in the MTA system.

Algorithms

Data Cleaning

- Removed duplicate rows.
- Corrected outlier data caused by turnstile counters not resetting at the same times each day and other unknown sources.
- Corrected non-standardized time stamps in the data by binning times into four-hour blocks of time beginning at midnight.

Analysis

- Calculated daily entries for each block of time for each turnstile by subtracting the data from the previous block of time and aggregating daily entries across all turnstiles in each station for each day.

Tools

- SQL for importing data into Python
- Numpy, Pandas for data analysis
- Matplotlib and Seaborn for visualization

Communication

- Presentation and write-up

