



Blood Donation

Fahad Alharbi

Abstract

The goal of this project was to use the available data to discover the busiest station in New York City to place a mobile health vehicle for blood donation near crowded stations and donate blood from people, through data set from the MTA dataset at the Subway station, through SQL, Python and use Python library such as pandas and NumPy and matplotlib and display it as a plot to see the busiest station and analyze the data and make use of it through the available information, and accordingly we can identified the busiest station out of more than 300 stations.

Design

This project was to help the need for blood donors. Through the available data, I was able to find out the busiest station and its solution during a period of 3 months, and the busiest days of the week through behavior for station visitors and placing a blood donation mobile health vehicle near the busiest station.

DATA

The data set contains more than 3 million rows and 11 columns, which is the entry and exit information for each turnstile in the station, distributed over days and hours so that it is calculated every four hours per day the data set was analyzed daily and every four hours

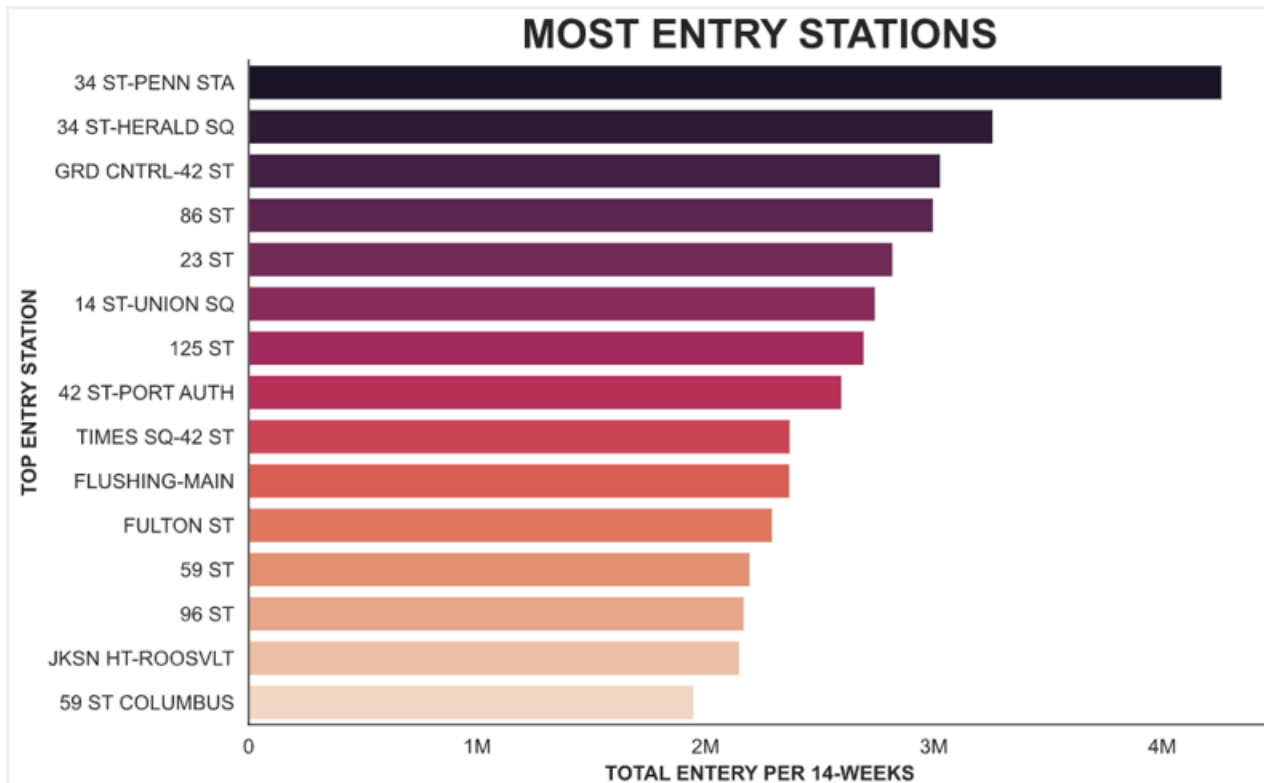
Algorithm

Analyzing the data by summing the days, weeks and days and knowing which station is the busiest, handling errors with the data such as outlier Through the analysis I concluded that greater than 5000 entries per day is a reasonable number to extract valid data

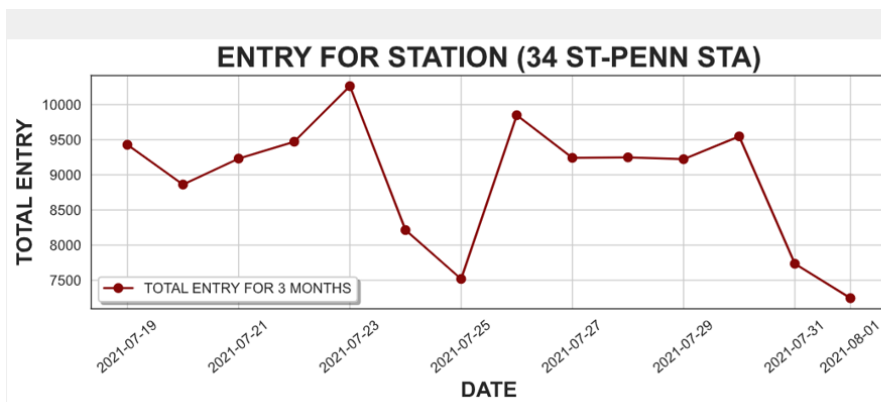
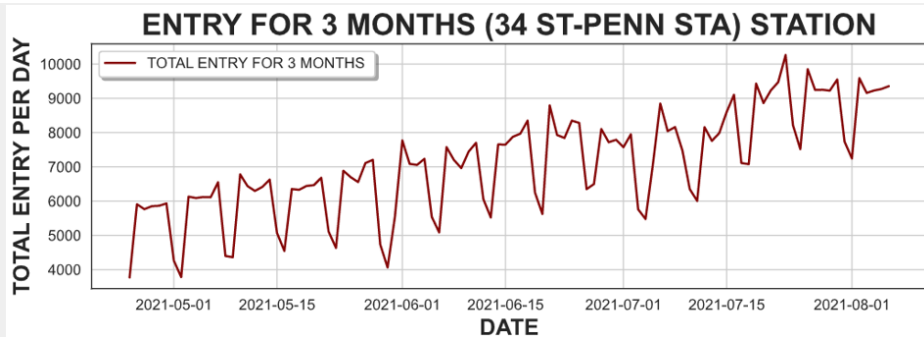
Tools

Technologies: SQLite ,python, Jupyter notebook.
Libraries: Numpy, Pandas, Matplotlib, Seaborn.

Communication



➤ In this chart, we can know for 3 months each day which days have the most entry to the station



➤ If we go deeper, and choose two weeks at random to know the days and see them more clearly, we find that the days of the week are more crowded than the weekend