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

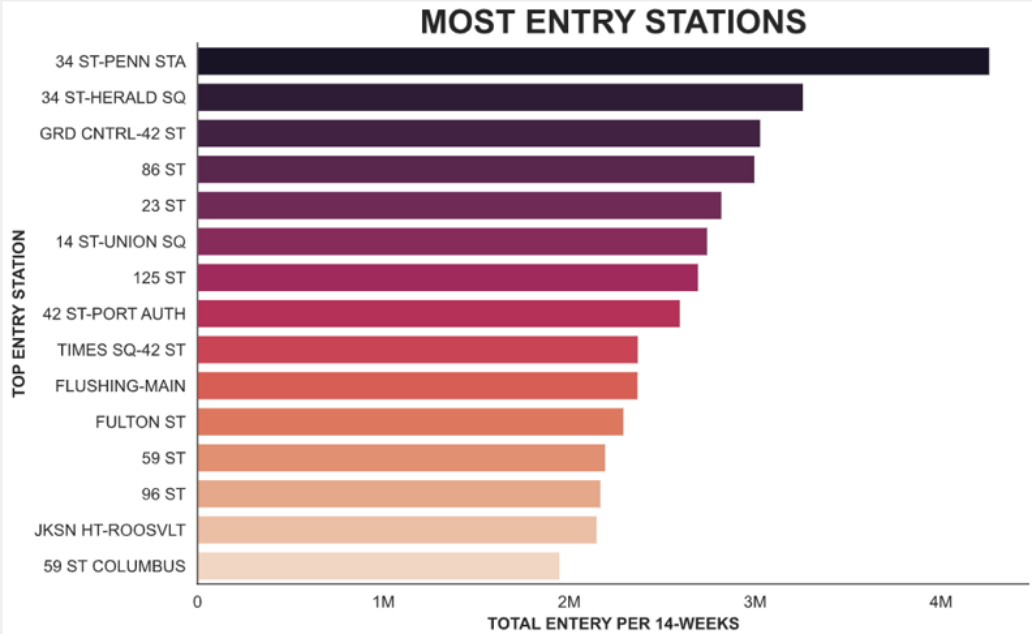
Algorithm

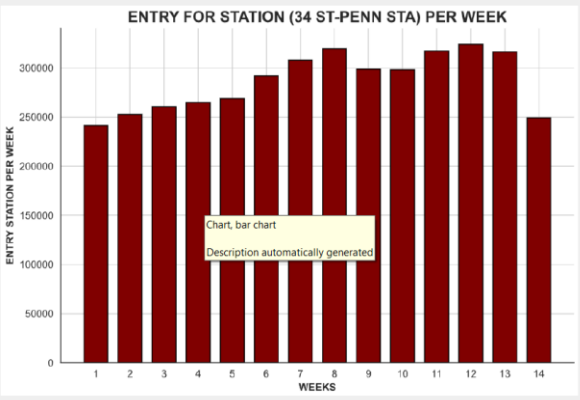
I analyze the data by collecting days, weeks and months and knowing the busiest station by using the python libraries pandas to convert data to dataframe and sqlalchemy for readers from SQL and used seaborn and matplotlib to draw charts to find the difference between days, weeks and months and know the busiest times, and processing wrong data such as external data through analysis, I concluded that more than 5000 entries per day is a reasonable number to extract correct data because the numbers that reset entry counter, I set the data as less than 5 K entry per day we found the number of rows = 458182 and when we put less than 10 K we found the number of rows= 458185 and this means that the difference is 3 rows

Tools

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

Communication

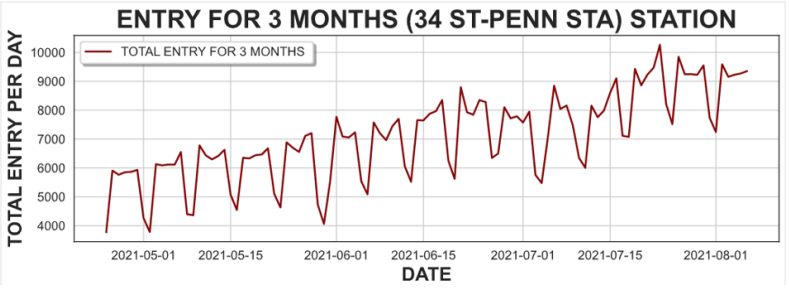


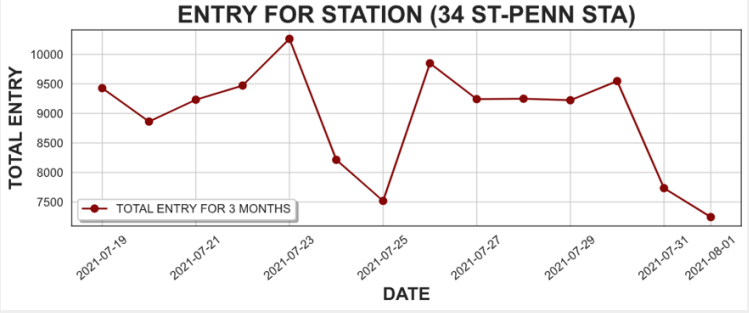


Observations

- Through this chart, we can know the sum of the entry per week for the most visited station over a period of 3 months

- 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