Abstract

As our planet undergoes climate change, New York's transportation sector is one of the largest contributors to carbon emissions. Central Bikes has therefore contributed to the launch of the "How green is cycling" campaign. Compared to any other mode of transportation, cycling has the least carbon footprint and it makes sense from a climate perspective for as many journeys as possible to be made by bicycle. This project will assist in achieving the success of Central Bikes campaign by distributing bikes at the five busiest stations. The project aims to analyze the MTA NYC Turnstiles data to find the top stations for bike distribution and to raise customer awareness of the benefits of cycling on a sustainable future.

Design

Central Bikes is one of New York City's largest bike rental companies, and is the official bike rental provider of NYC Parks. Daily bicycle rentals are offered to New York City residents and visitors, as well as tours of Central Park, Brooklyn Bridge, and more. As part of our mission to reduce carbon emissions, Central Bikes periodically engages in campaigns to raise public awareness about climate change and promote sustainable changes that help our planet to avoid destruction.

Data

The New York subway MTA turnstile data is a series of data files containing cumulative number of entries and exits by station, turnstile, date and time. Data files are produced weekly, data records are collected typically every 4 hours. The dataset contains 2,722,610 observations with 13 features for each, 2 of which are cumulative. Our Daily Traffic feature includes almost all Entries and Exits by aggregating their difference.

Algorithms

We plan to analyze the MTA turnstile data retrieved between June and August 2021 in order to identify the five most heavily used stations and the best days of the week to distribute Central Bikes.

Based on our study of the data set, we made the following observations:
1.Turnstile count is recorded in a cumulative manner, indicating that data recorded increases constantly over time.
2. Data set does not include day of the week (e.g. Saturday).
Therefore, we created new columns for these fields.

Traffic	EXIST + ENTRIES
Daily_Traffic	Traffic - Previous Traffic
DayOfWeek	e.g. Monday

We cleaned up our dataset by removing any duplicates and null values. Then, we saved the data frame on a database. The dataset were retrieved from dataset and used to plot the graphs by aggregating by stations, daily traffic and day of the week.

Based on the Daily Traffic of ridership in New York City, we have found out the top 5 busiest stations with the highest number of traffic. And we realized that more human traffic took place during weekdays over weekends.

Tools

- Jupyter Notebook will be used to create and document live code and visualization
- Sqlite3 for creating the database
- Pandas for data manipulation
- Matplotlib and Seaborn for plotting

Communication

Figure 1 represent top 5 stations based on high number of traffic

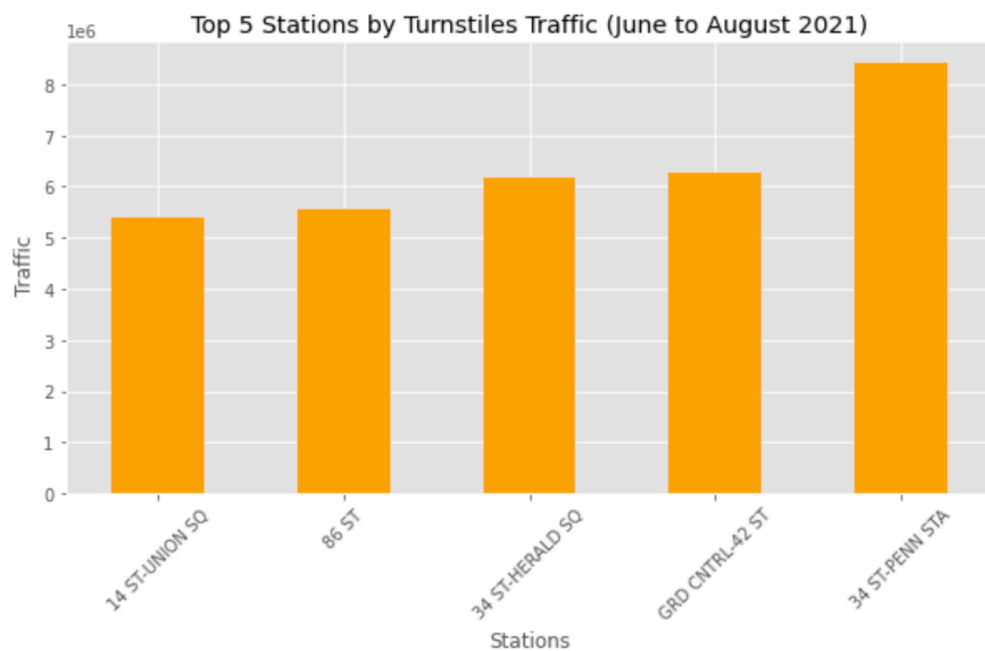Top 5 Stations by Turnstiles Traffic (June to August 2021)

Figure 2 represent high number of traffic based on days of the week