

New York Attractions' Subway Traffic Analysis

Abstract

The goal of this project was to analyze near attractions' specific subway stations' traffic of New York City in order to help **NYC & Company** base their pricings of attraction's ticketing and city pass on the analysis findings. According to a throughout search online on New York's most visited attractions, the analyzed attractions are following: Central Park, Statue of Liberty, Rockefeller Center & Top of the Rock Observation Deck, Metropolitan Museum of Art, Broadway and the Theater District, Empire State Building, and The Times Square the data worked on was provided by The Metropolitan Transportation Authority which is the MTA Turnstile Data and the stations near the specified attraction's information was gathered using google maps along with the longitude and latitude of each attraction's station. After exploring, cleaning and analyzing the data, plots and graphs were plotted to visualize and communicate the analysis findings.

NYC & Co.

NYC & Company is the official destination marketing organization (DMO) and convention and visitor's bureau (CVB) of New York City. The company's mission is to maximize travel and tourism opportunities throughout New York City and spread the dynamic image of New York City around the world. NYC & Company provides resources for visitors to discover everything they need to know about what to do and see in NYC. NYC & Company's marketing efforts position the city to attract business and leisure travelers.

As for the current situation NYC visitors dropped due to covid 19 pandemic, but NYC & Company forecast for 2021 will bring travel back to an estimated 36.4 million visitors. The pace of recovery will depend upon a strong leisure travel response to lockdown with the pace of vaccine production and distribution. The projection for domestic travel brings the city back to peak levels by 2023.

Design

This project data which is **MTA turnstile data** is provided by The Metropolitan Transportation Authority which reflects the patterns of transit traffic in New York City. Other data are collected manually through research about New York landmarks and attractions with their near stations in order to capture the traffic in areas where the attractions are. Capturing the traffic around those stations where the landmarks are will enable NYC & Company to decide their pricings and availability of ticketing and scheduling based on the findings.

Data

MTA data shows the entry/exit register values for one turnstile at control area through a specific time. The applied MTA dataset contains 378 stations with 2887519 rows only for the period of late September 2019 to early January 2020 and the reason behind choosing this data over the most recent data because of the exceptional period the world has been through which is not stable enough to make decisions based on. The other data which was collected through google maps and inserted into a newly created database containing 7 attractions around New York with their specified stations and longitude and latitude making up 66 rows of data.

Algorithms

1. Downloaded MTA turnstile data files into Jupiter notebook then Loaded the files into a pandas Data Frame.
2. Explored and got to know the data while understanding how to use it and deciding which attributes should be focused on during the analysis, and if external data is needed or.
3. Searched for external data and gathered information regarding most visited attractions and landmark in New York then which stations are near those attractions along with their longitude and latitude from google maps.
4. Cleaned the data and filtered it then joined it with the new collected data to get only the stations designated for those attractions.
5. Mapping latitude and longitude coordinates so nearby continuous values would also be close in reality
6. Selecting subsets of the total daily traffic per each attraction in order to visualize our findings through different types of plots (per timeline, per attraction , per traffic percentage etc....

Tools

- Numpy and Pandas for data manipulation
- Google maps to extract exact locations of stations and attractions
- SQLite and SQLAlchemy to create and insert database into python and join it with MTA data
- Matplotlib and Seaborn for plotting

Communication

In addition to the slides and visuals presented, everything is submitted through the submission form.