

Capstone course project : Battle of Neighborhoods

Project Report : By Jigisha Desai

2. Data Section:

Data section describes the sources of the data collected to solve the problem. It also indicates how the data will be used to solve the problem.

2.1 Data of Current city, The city of London:

Mr. X currently lives in the city of London. To get the details about the venues in London city, Foursquare credentials and APIs are used. These venues are then shown on the London map. The details about Foursquare APIs and how to show venues on London map are given in "Methodology section" of the report. It serves as a reference for comparison with the desired future location in Manhattan NY.

2.2 Data required to solve the problem:

In order to make a good choice of a similar apartment in Manhattan NY, we need data about subway stations, rental places and venues in Manhattan.

- **Subway Stations Data:**

To retrieve subway stations names and address (about 125 stations) in Manhattan following website is used.

https://en.wikipedia.org/wiki/Category:New_York_City_Subway_stations_in_Manhattan

To get the latitude and longitude of all the subway stations, “Nominatim” function from “geocoders” library is used.

List of the subway stations along with their latitude and longitude are saved in "subway_stations.csv" file.

- **Rental Places Data:**

To retrieve the details of rental places with the monthly rental between US\$2000 and US\$6000 following websites are used. Details like rent, address, no of beds, bathrooms etc. are retrieved.

<https://www.realtor.com/>

<http://www.rentmanhattan.com/>

<https://www.calibernyc.com/>

To get the latitude and longitude of the rental places, “**Nominatim**” function from “**geocoders**” library is used.

Details of the rental places along with their latitude and longitude are saved in "**rental_places.csv**" file.

- **Venues Details:**

To retrieve different venues in Manhattan, first we need to get the dataset of neighborhoods in Manhattan. This can be retrieved from “**newyork_data.json**” file or from the following website.

https://geo.nyu.edu/catalog/nyu_2451_34572

Now we can get the details about the venues in Manhattan using neighborhood details, Foursquare credentials and APIs.

2.3 How the data will be used to solve the problem?

- Venue details collected about the current city, London will be displayed on the map of London. Categories of the five most available venues will also be displayed. This will be helpful to compare venues in Manhattan.
- Subway stations details stored in "**subway_stations.csv**" file will be retrieved and stations will be displayed on the map of Manhattan.
- Rental places details stored in "**rental_places.csv**" file will be retrieved and places will be displayed on the map of Manhattan.

- Venue details retrieved using Foursquare credentials and APIs will be clustered using the clustering algorithm and then clusters will be displayed on the map of Manhattan.
- To explore **subway stations** and **rental places**, the map will be created to show both the details together.
- To explore **rental places and clusters of venues**, the map will be created to show both the details together.
- The final map will be created that shows details about **rental places**, **subway stations** and **clusters of venues** together. It will also provide a facility to check the distance between a subway station and a rental place or a distance between a rental place and a venue. This will help Mr. X to make a good choice of a similar apartment in Manhattan NY.
