Coursera Capstone Project

## The Battle of Neighborhoods - Final Report

### Contents

#### 1.Introduction Section :

1.1 Discussion of the "background situation" leading to the problem at hand:

1.2 Problem to be resolved

1.3 Audience for this project.

#### 2.Data Section:

2.1 Data of Current Situation (current residence place)

2.2 Data required to resolve the problem

2.3 Data sources and data manipulation

#### 3.Methodology section :

3.1 Process steps and strategy to resolve the problem

3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis.

#### 4.Results section

Discussion of the results and how they help to take a decision.

#### 5.Conclusion section

Decision taken and Report Conclusion.

## *1. Introduction Section :*

### Discussion of the business problem and the audience who would be interested in this project.

### *Scenario and Background*

I am an analyst residing in the city of Pune in the country India. The area I live in is a 2 minute ride from the airport, is a 5 minute walk from the largest mall in the city, has more than 7 bakeries in a 100 meter square radius, consists of 4 night clubs and many other amenities such as various international cuisine restaurants, cafes, park, gyms, shops and repair shops. I am currently working with a leading consulting firm. This firm has offered me an opportunity to work for a project in New York City for a period of 1 and a half year. Here I wish to compare the two cities and find myself a place to stay that could be as similar as the area I currently reside in. The apartment in Manhattan must meet the following demands

### *Problem to be resolved:*

The challenge to resolve is being able to find a rental apartment unit in Manhattan NY that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to rent a unit subject to the following conditions:

• 1 bedroom apartment.

• desired location is near a metro station in the Manhattan area and within 1.0 mile (1.6 km) radius

• price of rent not exceed $7,000 per month

• top amenities in the selected neighborhood shall be similar to current residence

• desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym and food shops

### *Interested Audience*

This project could be helpful to anyone who is moving to another country for work or for study. The same methodology can be applied in accordance to demands as applicable. This case is also applicable for anyone interested in exploring starting or locating a new business in any city.

# *2. Data Section:*

## Description of the data and its sources that will be used to solve the problem

### 2.1 *Data of Current Situation*

I currently reside in the neighborhood of 'Vimannagar' in Pune, India. I will use Foursquare data to identify the venues around the area of residence which are then shown in the map of Pune city (methodology section). It serves as a reference for comparison with the desired future location in Manhattan NY.

### *2.2 Data Required to resolve the problem*

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required: List/Information on neighborhoods form Manhattan with their Geodata ( latitude and longitude. List/Information about the subway metro stations in Manhattan with geodata. Listed apartments for rent in Manhattan area with descriptions ( how many beds, price, location, address) Venues and amenities in the Manhattan neighborhoods (e.g. top 10) 2.3 sources and manipulation The list of Manhattan neighborhoods is worked out during Lab exercise during the course. A csv file was created which will be read in order to create a data frame and its mapping. The csv file 'mh\_neigh\_data.csv' has the following below data structure. The file will be directly read to the Jupiter Notebook for convenience and space savings. The clustering of neighborhoods and mapping will be shown however. An algorithm was used to determine the geodata from Nominatim . The actual algorithm coding may be shown in 'markdown' mode because it takes time to run.

mh\_neigh\_data.tail():

Borough Neighborhood Latitude Longitude

35 Manhattan Turtle Bay 40.752042 -73.967708

36 Manhattan Tudor City 40.746917 -73.971219

37 Manhattan Stuyvesant Town 40.731000 -73.974052

38 Manhattan Flatiron 40.739673 -73.990947

39 Manhattan Hudson Yards 40.756658 -74.000111

A list of Manhattan subway metro stops was compiled in Numbers (Apple excel) and it was complemeted with wikipedia data ( <https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan>) and information from NY Transit authority and Google maps (<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>) for a final consolidated list of subway stops names and their address. The geolocation was obtained via an algorithm using Nominatim. Details will be shown in the execution of methodology in section 3.0. The subway csv file is ''MH\_subway.csv'" and the data structure is:

mhsub.tail(): sub\_station sub\_address lat long

17 190 Street Subway Station Bennett Ave, New York, NY 10040, USA 40.858113 -73.932983

18 59 St-Lexington Av Station E 60th St, New York, NY 10065, USA 40.762259 -73.966271

19 57 Street Station New York, NY 10019, United States 40.764250 -73.954525

20 14 Street / 8 Av New York, NY 10014, United States 40.730862 -73.987156

21 MTA New York City 525 11th Ave, New York, NY 10018, USA 40.759809 -73.999282

A list of places for rent was collected by web-browsing real estate companies in Manhattan : <http://www.rentmanhattan.com/index.cfm?page=search&state=results><https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list> <https://www.realtor.com/apartments/Manhattan_NY> A csv file was compiled with the rental place that indicated: areas of Manhattan, address, number of beds, area and monthly rental price. The csv file "nnnn.csv" had the following below structure. An algorithm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longitude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorithms used are shown in the execution of data in section 3.0 "Great circle" function from geolocator was used to calculate distances between two points , as in the case to calculate average rent price for units around each subway station and at 1.6 km radius. Foursquare is used to find the avenues at Manhattan neighborhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

### *2.4 How the data will be used to solve the problem*

The data will be used as follows: Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups ( as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. create a map that depicts, for instance, the average rental price per square ft, around a radius of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the relative price per subway area. Addresses from rental locations will be converted to geodata( lat, long) using Geopy-distance and Nominatim. Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc.

### *2.5 Mapping of Data*

The following maps were created to facilitate the analysis and the choice of the palace to live. Manhattan map of Neighborhoods Manhattan subway metro locations Manhattan map of places for rent Manhattan map of clustered venues and neighborhoods Combined maps of Manhattan rent places with subway locations Combined maps of Manhattan rent places with subway locations and venues clusters.

# *3. Methodology section:*

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

1. City of Pune is analyzed as it would be the reference point in selecting a neighborhood in New York.
2. Coordinates of Pune are found using geolocator and the neighborhood ‘VimanNagar’ is mapped for reference using Foursquare data.
3. Venues are classified using Foursquare data.
4. Manhattan Neighborhood data is analyzed and mapped.
5. Manhattan Boroughs are clustered.
6. A map of Manhattan is shown with top 10 clusters using Folium.
7. A map of Manhattan is shown with places for rent.
8. Statistics of rent prices are shown with the help of matpotlib using box plots and graphs.
9. Clustering is done using the method of k- means clustering.
10. Sub way locations are mapped.
11. Finally a consolidated map is shown consisting of rental places, subway stations and cluster of venues.
12. Apartment is selected basis the above requirements.

*6.0 Conclusions.*

* Built useful models to analyze, cluster, cleanse and choose suitable living options.
* Used geo graphs and other forms of visualizations to analyze large amounts of data.
* Compared two cities that are located in 2 different continents and are culturally and infrastructure vise very different. Despite this difference we were able to find similarities in the neighborhoods.
* Analyzed geographical, physical and financial data within the same project.
* Project is useful to all sorts of people and not limited to professionals.
* Project won’t get outdated as people will always immigrate to a different city.

*Thank You.*