**Coursera Capstone - REPORT**

**Content**

1. Introduction Section :

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

1.2 Problem to be resolved

1.3 Audience for this project.

1. 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

1. Methodology section :

3.1 Process steps and strategy to resolve the problem

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

1. Results section

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

1. Discussion section

Elaboration and discussion on any observations and/or recommendations for improvement.

1. Conclusion section

Desicison taken and Report Conclusion.

# 1. Introduction Section :

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

### 1.1 Scenario and Background

I am a data scientist currently residing in Downtown Singapore. I currently live within walking distance to Downtown "Telok Ayer MRT metro station" therefore I have access to good public transportation to work. Likewise, I enjoy many ammenities in the neighborhood , such as international cousine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Manhattan. Therefore, I decided to apply the learned skills during the Coursera course to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it rewarding doing it myself with learned tools.

### 1.2 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 find a renta unit subject to the following conditions:

* Apartment with min 2 bedrooms with monthly rent not to exceed US$7000/month
* Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan
* Area with ammenities and venues similar to the ones described for current location ( See item 2.1)

**1.3 Interested Audience**

I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

**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 'Mccallum Street' in Downtonw Singapore. I use Foursquare to identify the venues around the area of residence which are then shown in the Singapore map shown in methodology and execution in section 3.0 . 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 ( latitud and longitud. 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 ammenities 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 dataframe 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 becasues 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 complied 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 algorythm using Nominatim. Details will be shown in the execution of methodolody 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,

271

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 algorythm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorythm coding may be shown in 'markdown' mode becasues it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longiude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorythms 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 ammenities 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 radious 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.

### The analysis and the stragegy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed : location near a subway, rental price and similar venues to Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

* what is the cost of available rental places that meet the demands?
* what is the cost of rent around a mile radius from each subway metro station?
* what is the area of Manhattan with best rental pricing that meets criteria established?
* What is the distance from work place ( Park Ave and 53 rd St) and the tentative future rental home?
* What are the venues of the two best places to live? How the prices compare?
* How venues distribute among Manhattan neighborhoods and around metro stations?
* Are there tradeoffs between size and price and location?
* Any other interesting statistical data findings of the real estate and overall data.