**Introduction**

A data scientist currently residing in Downtown Singapore. He currently live within walking distance to Downtown "Telok Ayer MRT metro station" therefore He has access to good public transportation to work. Likewise, He enjoy many amenities in the neighbourhood, such as international cuisine restaurants, cafes, food shops and entertainment. He has been offered a great opportunity to work in Manhattan, NY. Although, He is very excited about it, He is a bit stress toward the process to secure a comparable place to live in Manhattan. Therefore, He decided to apply the learned skills during the Coursera course to explore ways to make sure his decision is factual and rewarding.

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, we want to find a rent 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 amenities and venues similar to the ones described for the location

**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 Four Square 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.

**Data Section**

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required: List/Information on neighbourhoods form Manhattan with their Geo data (latitude and longitude. List/Information about the subway metro stations in Manhattan with geo data. Listed apartments for rent in Manhattan area with descriptions (how many beds, price, location, address) Venues and amenities in the Manhattan neighbourhoods (e.g. top 10)

**Sources**

The list of Manhattan neighbourhoods 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 neighbourhoods and mapping will be shown however. An algorithm was used to determine the geo data. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run.

mh\_neigh\_data.tail():

Borough Neighbourhood 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 complemente 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 geo location 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 geo data 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 geo locator 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 neighbourhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

**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 neighbourhoods 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 geo data( 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.

**Mapping of Data**

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