

# Data Science Capstone

Omkar Shinde

08-Apr-2019

# Report Content

1. Introduction Section : - The “business problem” to be solved by this project and who may be interested
2. Data Section: - Describe Data requirements and Sources needed to solve the problem
3. Methodology section: - Main component of the report - Execute data processing, describe/discuss any exploratory data analysis and/or inferential statistical testing performed, and/or machine learnings used.
4. Results section: - Discussion of the results and finding of answer
5. Discussion section: - Discussion of observations noted and any recommendations
6. Conclusion section: - Answer chosen and conclusions.

# 1.0 Introduction

1.1 Scenario and Background I am currently living in Singapore, within walking distance to Downtown "Telok Ayer MRT metro station" . I also enjoy great venues and attractions, such as international cuisine, entertainment and shopping. I have an offer to move to work to Manhattan NY and I would like to move if I can find a place to live similar with similar venues.

1.2 Problem to be resolved: How to find an apartment in Manhattan with the following conditions: • Apartment with min 2 bedrooms • Monthly rent not to exceed US\$7000/month • Located within walking distance ( $\leq 1.0$  mile, 1.6 km) from a subway metro station in Manhattan • Venues and amenities as in my current residence

1.3 Interested Audience I believe the methodology, tools and strategy used in this project is relevant for a person or entity considering moving to a major city in US, Europe or Asia. Europe, US or Asia, Likewise, it can be helpful approach to explore the opening of a new business. 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 for a person developing Data Science skills. 2.0 Data .

## 2.0 Data Section

2.1 Data Requirements - Geodata for current residence in Singapore with venues established using Foursquare. - List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare (as in Course Lab).  
[https://en.wikipedia.org/wiki/List\\_of\\_Manhattan\\_neighborhoods#Midtown\\_neighborhoods](https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods) - List of subway metro stations in Manhattan with addresses and geo data (lat,long): [https://en.wikipedia.org/wiki/List\\_of\\_New\\_York\\_City\\_Subway\\_stations\\_in\\_Manhattan](https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan)), (<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>) - List of apartments for rent in Manhattan area with information on neighborhood location, address, number of beds, area size, monthly rent price and complemented with geo data via Nominatim. <http://www.rentmanhattan.com/index.cfm?page=search&state=results>  
<https://www.nestpick.com/search?city=new-york> - Place to work in Manhattan (Park Avenue and 53rd St) for reference  
2.2 Data Sources, Data Processing and Tools used - Singapore data and map is to be created with use of Nominatim, Foursquare and Folium mapping - Manhattan neighborhoods were obtained from Wikipedia and organized by Neighborhoods with geodata via Nominatim for mapping with Folium. - List of Subway stations was obtained via Wikipedia, NY Transit web site and Google map, - List of apartments for rent was consolidated from web-scraping real estate sites for MH. The geolocation (lat,long) data was found with algorithm coding and using Nominatim. - Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed to make a selection of apartment  
3.0 Methodology

## 3.0 Methodology

The Strategy to find the answer: The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location from a subway station and from work place. A measurement tool icon will also be provided. The popups on the map items will display rent price, location and cluster of venues applicable. The Tools: Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simplify the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent listed. Geopy\_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data. Maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy

# Apartment Selection

Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision. Apartment 1 rent cost is US\$7500 slightly above the US\$7000 budget. Apt 1 is located 400 meters from subway station at 59th Street and work place ( Park Ave and 53rd) is another 600 meters away. I can walk to work place and use subway for other places around. Venues for this apt are as of Cluster 2 and it is located in a fine district in the East side of Manhattan.

Apartment 2 rent cost is US\$6935, just under the US\$7000 budget. Apt 2 is located 60 meters from subway station at Fulton Street, but I will have to ride the subway daily to work , possibly 40-60 min ride. Venues for this apt are as of Cluster

3. Based on current Singapore venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that APARTMENT 1 is a better choice since the extra monthly rent is worth the conveniences it provides.

# 5.0 Discussion

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course

- I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.
- I have created a good project that I can present as an example to show my potential.
- I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

# 6.0 Conclusions

- I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.
- This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.
- The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.