



# Coursera Capstone project

Coursera IBM Data Science Certification

Kshitij Sharma

April, 2020

# 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.

### 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.

### 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 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->
- 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