Capstone project

Coursera IBM Data Science Certification

Seyed Jamal Haddadi

Report Contents

1- Introduction:

➢ The "business problem" to be solved by this project and who may be interested.

➢ Interested Audience

2- Data:

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

➢ How the data will be used to solve the problem

1- Introduction

My name is Jamal, and I am currently a data scientist in Singapore. live within walking distance to Downtown Telok Ayer MRT metro station and I enjoy many amenities and venues in the area, such as various international cuisine restaurants, cafes, food shops and entertainment. The goal of this project is to use data analysis from web sources to find a convenient and enjoyable place similar to mine now in Singapore. In fact, I want to use available real estate apps and Google but the idea is to use and apply myself the learned tools during the course. In order to make a comparison and evaluation of the rental options in Manhattan NY, I must set some basis, therefore the apartment in Manhattan must meet the following demands:

- apartment must be 2 or 3 bedrooms
- 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

- as a reference, I have included a map of venues near current residence in Singapore.

I believe this is a relevant challenge with valid questions for anyone looking for a convenient and enjoyable place in Singapore. This case is also applicable for anyone interested in exploring starting or locating a new business in any city. Lastly, it can also serve as a good practical exercise to develop Data Science skills.

1-1 Business Problem

The challenge is to find a suitable apartment for rent in Manhattan NY that complies with the demands on location, price and venues. The data required to resolve this challenge is described in the following section 2, below.

1-2 Interested Audience

I believe this is a relevant challenge with valid questions for anyone moving to other large city in US, EU or Asia. 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. Lastly, it can also serve as a good practical exercise to develop Data Science skills.

2- Data

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

2-2 Data Description:

The following data is required to answer the issues of the problem:

- List of Boroughs and neighborhoods of Manhattan with their geodata (latitude and longitude)
- List of Subway metro stations in Manhattan with their address location
- List of apartments for rent in Manhattan area with their addresses and price
- Preferably, a list of apartments for rent with additional information, such as price, address, area, # of beds, etc.
- Venues for each Manhattan neighborhood (than can be clustered)
- Venues for subway metro stations, as needed

2-3 Data Requirements

- Geodata for current residence in Singapore with venues established using Foursquare.
- List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare.
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://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,74.1033043,11z/data=!3m1!4b)](https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan)

- 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 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 Nominate.
- 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.

The processing of these DATA will allow to answer the key questions to make a decision:

- what is the cost of rent (per square ft) 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 53rd St) and the tentative future 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.