Coursera Capstone project

Pranit Shinde

[Course title]

Index

The Battle of Neighbourhoods (Week 1)

1. Introduction Section: Discussion of the business problem and the interested audience in this project

Background:

I am a computer scientist residing in Singapore city centre. I currently stay within walking distance of the metro station Downtown Telok Ayer MRT and I enjoy several facilities and venues in the city, including numerous foreign cousin restaurants , cafes, snack shops and entertainment. I was given a wonderful opportunity to work with a leading organization in Manhattan, NY. I am very motivated and would like to use this chance to continue my research in Coursera to address the related questions that emerged. The main problem is: How can I find in Singapore a comfortable and pleasant location close to mine now? I will definitely use usable real estate applications and Youtube, but the aim is to utilize the acquired resources through the course to adapt them to myself. To assess and analyze the rental choices in Manhattan NY, I have to put down some ground, so the Manhattan apartment will satisfy the following demands:

- Apartment will consist of two or three bedrooms
- The target position is adjacent to the Manhattan subway station and within a distance of 1.0 mile (1.6 km)
- Cost will not surpass $7,000 a month
- Top facilities shall be identical to the new residence in the chosen neighbourhood
- Wanted to include locations like coffee shops, Asian Thai bars, wine stores, gymnasiums and grocery shops
- I also provided a chart of the places around the new residence in Singapore as a guide.

Company issue:

The challenge is to find a suitable rental apartment in Manhattan NY that suits place, price and venue demands. The data needed to overcome this problem is listed in section 2, below.

Audience involved

I think that is a real obstacle for those traveling to many big cities in the US, the EU or Asia with legitimate concerns. As appropriate, the same approach can be implemented according to demands. This case also extends to everyone interested in exploring whether to start or locate a new business in any area. Eventually, it may also function as a successful realistic exercise for Data Science skills growth

2. Data Section: Description of the data that will be used to solve the problem and the sources

To address the problems of the question the following data are required:

- Map of Manhattan Boroughs and Districts with their geodata (latitude and longitude)
- Directory of Manhattan Metro Stations with their website
- Directory of rental apartments in Manhattan city with their addresses and costs
- Preferably a list of rented apartments with additional details including size, address, location, # of beds, etc.
- Venues (which may be clustered) for every Manhattan neighbourhood
- Metro station positions, as required

When to use the data to address a question

They would use the details as follows:

- Using Foursquare and geopy data to chart top 10 locations for all Manhattan communities, divided into categories (as per LAB course)
- Using foursquare and geopy data to chart the position of metro stations, independently and on top of the clustered chart above, to classify places and facilities close each metro station or to discover each subway position independently
- Using Foursquare and geopy data to chart the site, in any manner, of rental places linked to subway locations.
- For eg, build a chart showing the average rental price per square foot, around 1.0-mile (1.6 km) distance from each subway station-or similar metrics. I'll be able to point to the popups easily to learn the relative price per subway sector.
- Using Geopy-distance and Nominatim, addresses from the rental locations will be translated to geodata(lat, long).

Processing these DATA would require the main questions to be addressed in order to take a decision:

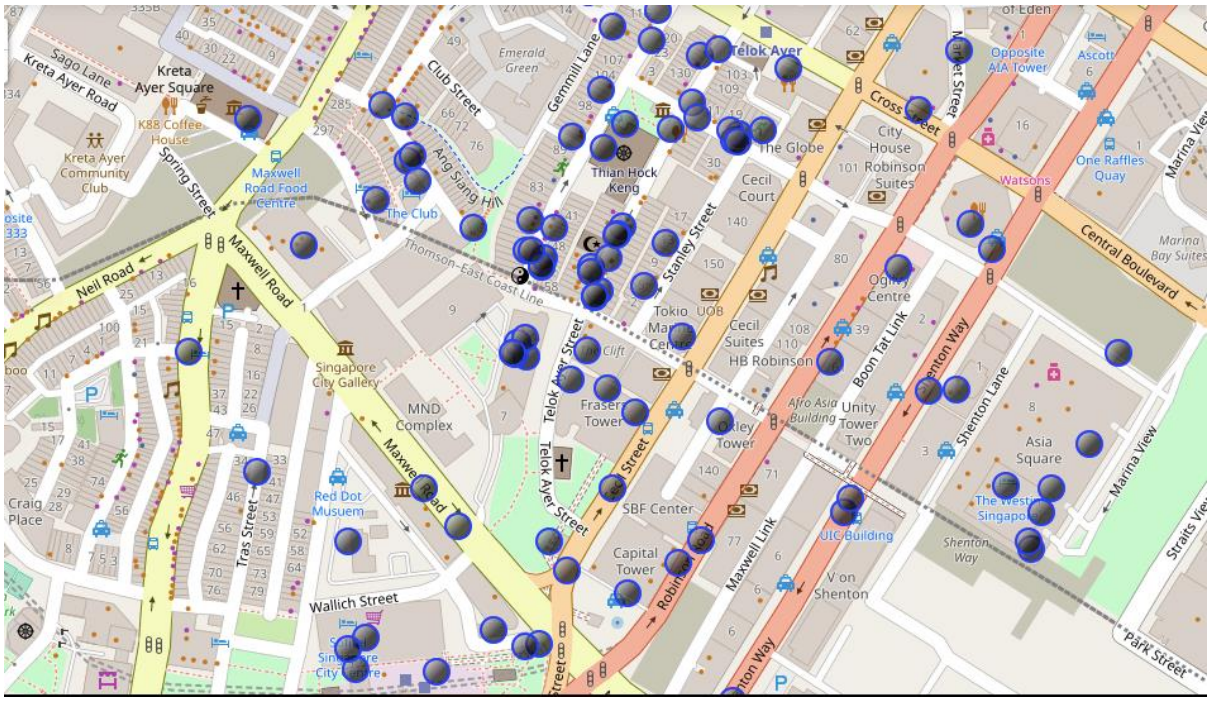
1. Which are the rental costs (per square ft) at a mile radius from each metro station?
2. Which is Manhattan's best-priced rental area that fits defined criteria?
3. What is the gap from the place of work (Park Ave and 53 rd St) and what is the potential future home?
4. What are the safest cities to stay in for the two? Why will rates match?
5. Where will Downtown communities and across the subway stations allocate locations?
6. Are there tradeoffs between size and place, and price?
7. What other important statistical results of the real estate and the general knowledge

Methodology:

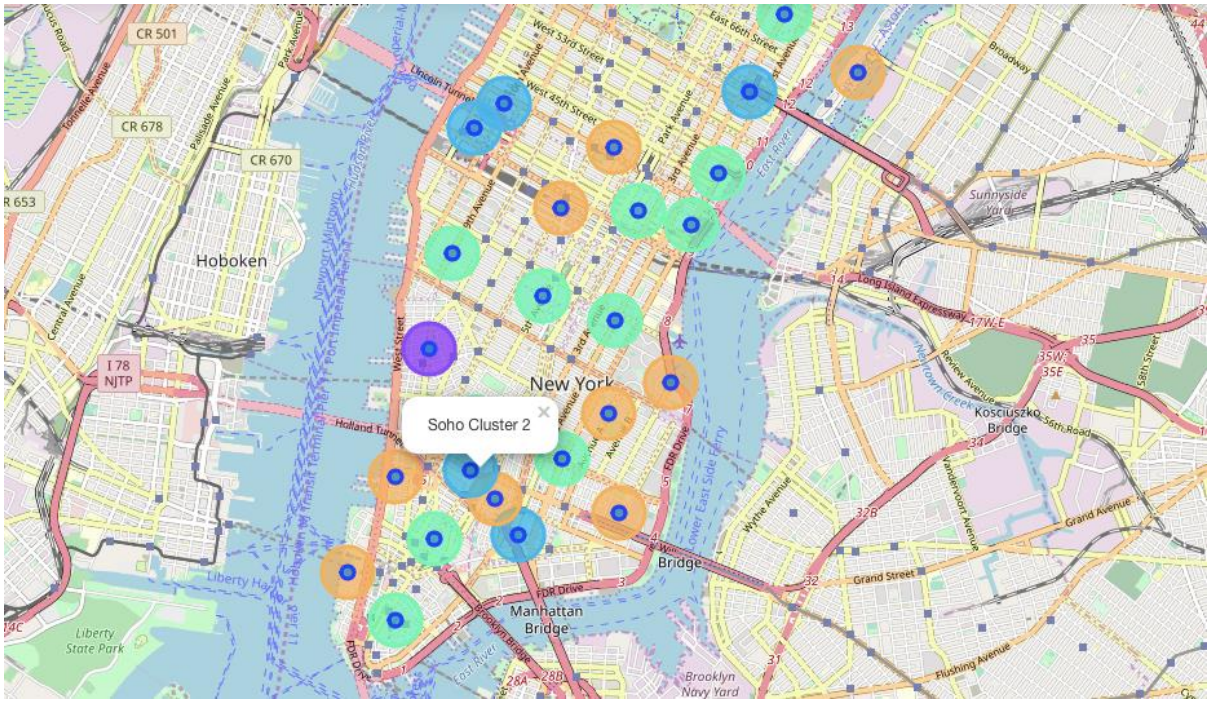
The Strategy to find the answer:
The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the 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 simply 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.

Execution and Results

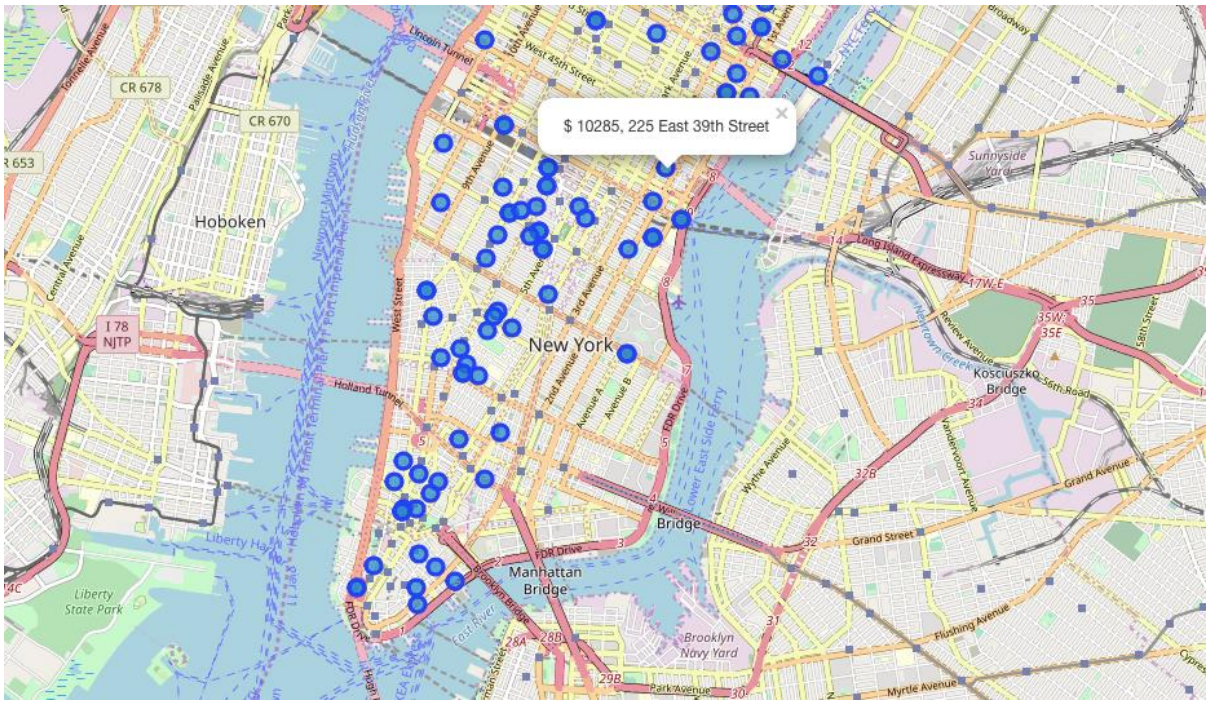
Current residence Neighborhood in Singapore



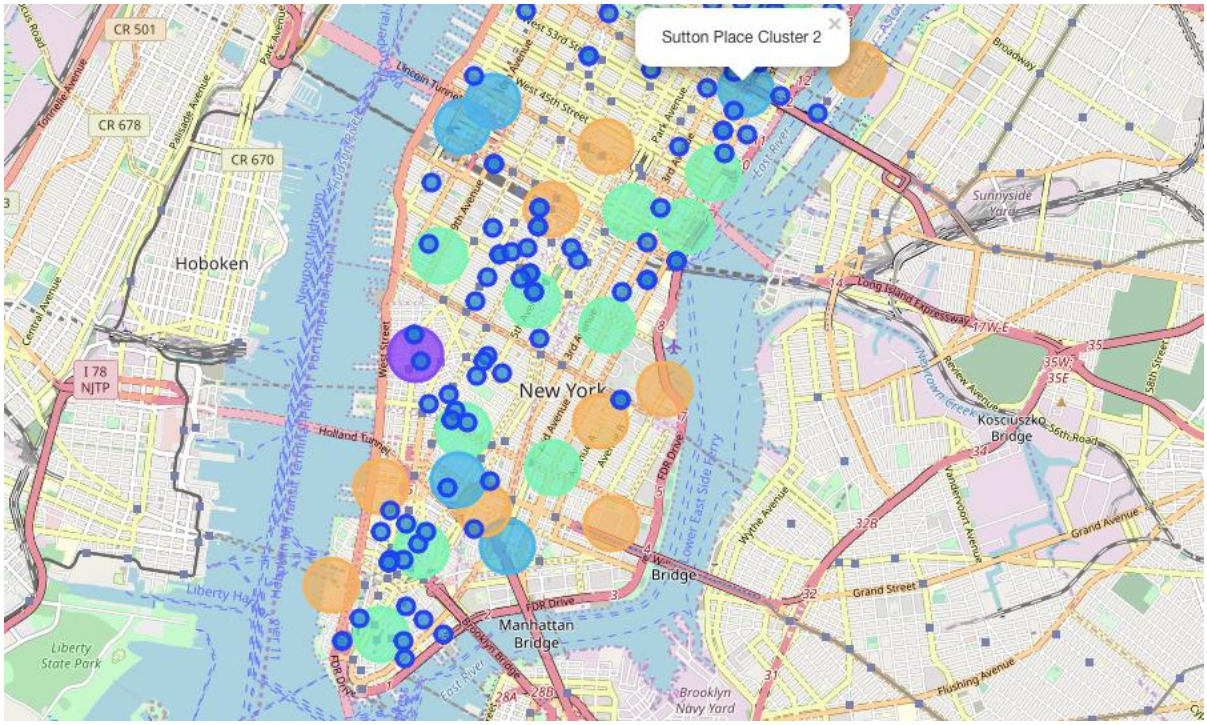
Manhattan Map - Neighborhoods and Cluster of Venues



Apartments for Rent in MH

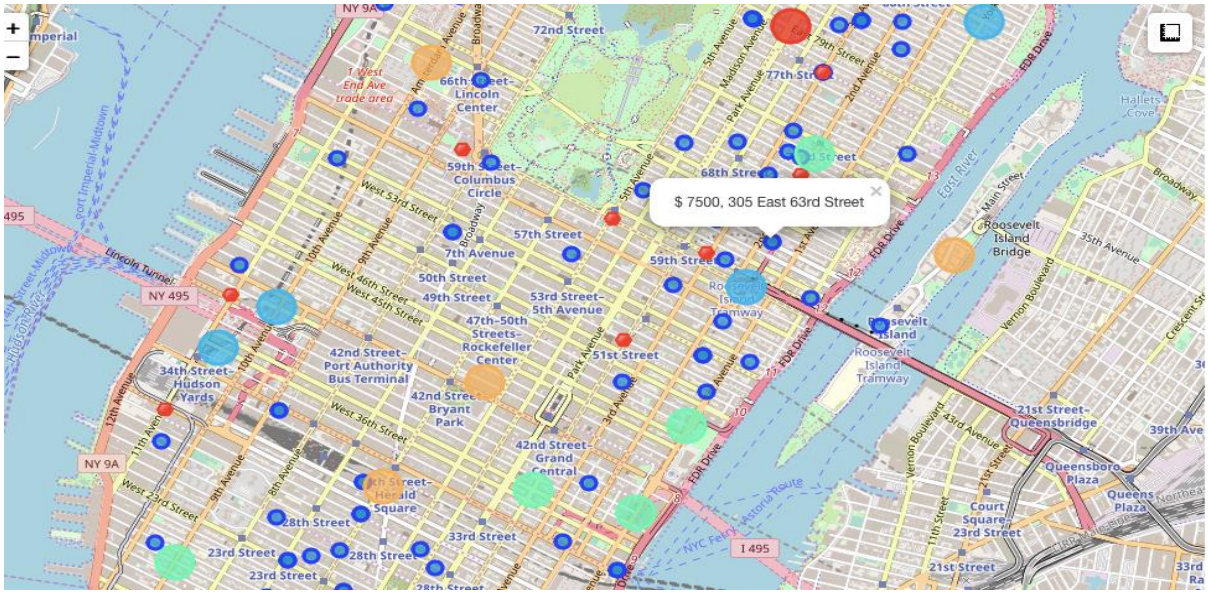


MH apts for rent with venue clusters



The ONE consolidated map shows all information for decision:
Apartments address, price, neighborhood, cluster of venues and subway station nearby.

Blue dots=apts , Red dots=Subway



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.

• One must keep abreast of new tools for DS that continue to appear for application in several business fields.