Coursera Capstone Project

The Battle of Neighbourhoods Final (Week 1 & Week 2)

Philip Shield
26 May 2021

1. Introduction

Description of the Problem and Background

Scenario:

Let us say I am a data scientist residing in Toronto. A business firm in Manhattan, NY hire me for data analysis to open a restaurant. Both NY and Toronto cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighbourhoods of the two cities and determine how similar or dissimilar they are. In order to make a comparison and evaluation of the rental options in Manhattan NY, I must set some basis, therefore the property in Manhattan must meet the following demands:

- desired location is in the Manhattan area, New York.
- should be near to the metro stations.
- price of rent does not exceed $7,000 per month.
- desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym, and food shops.

Business Problem:

The challenge is if someone is looking to open a restaurant in Manhattan area, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

2. Data Section:

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

Description of the Data:

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

- List of Boroughs and neighbourhoods of Manhattan with their geodata (latitude and longitude)
- List of property for rent in Manhattan area with their addresses and price
- Venues for each Manhattan neighbourhood (than can be clustered)
- List Subway metro stations

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 and identify the venues and ammenities near each metro station
- Use Foursquare and geopy data to map the location of rental places
- create a map, the average rental price per square ft
- Addresses from rental locations will be converted to geodata(lat, long) using Geopy-distance and Nominatim.

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)?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- What are the venues of the two best places to open a restaurant? How the prices compare?
- How venues distribute among Manhattan neighbourhoods and around metro stations?

3. Methodology section:

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The analysis and the strategy:

The strategy is based on mapping the above-described data, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price and similar venues to Toronto. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow to answer the key questions to decide:

- what is the cost of available rental places that meet the demands?
- what is the cost of rent around a mile radius from each subway metro station?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- How venues distribute among Manhattan neighbourhoods and around metro stations?
- Are there trade-offs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.

METHODOLOY EXECUTION - Mapping Data

Reference of venues around current residence in Toronto for comparison to Manhattan place

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Nathan Phillips Square	Plaza	43.652270	-79.383516
2	LUSH	Cosmetics Shop	43.653557	-79.380400
3	Indigo	Bookstore	43.653515	-79.380696
4	Old City Hall	Monument / Landmark	43.652009	-79.381744
5	CF Toronto Eaton Centre	Shopping Mall	43.654447	-79.380952
6	M Square Coffee Co	Coffee Shop	43.651218	-79.383555
7	Crepe Delicious	Fast Food Restaurant	43.654536	-79.380889
8	Poke Guys	Poke Place	43.654895	-79.385052
9	Chatime 日出茶太	Bubble Tea Shop	43.655542	-79.384684

Map of Toronto with venues near CBD place - for reference
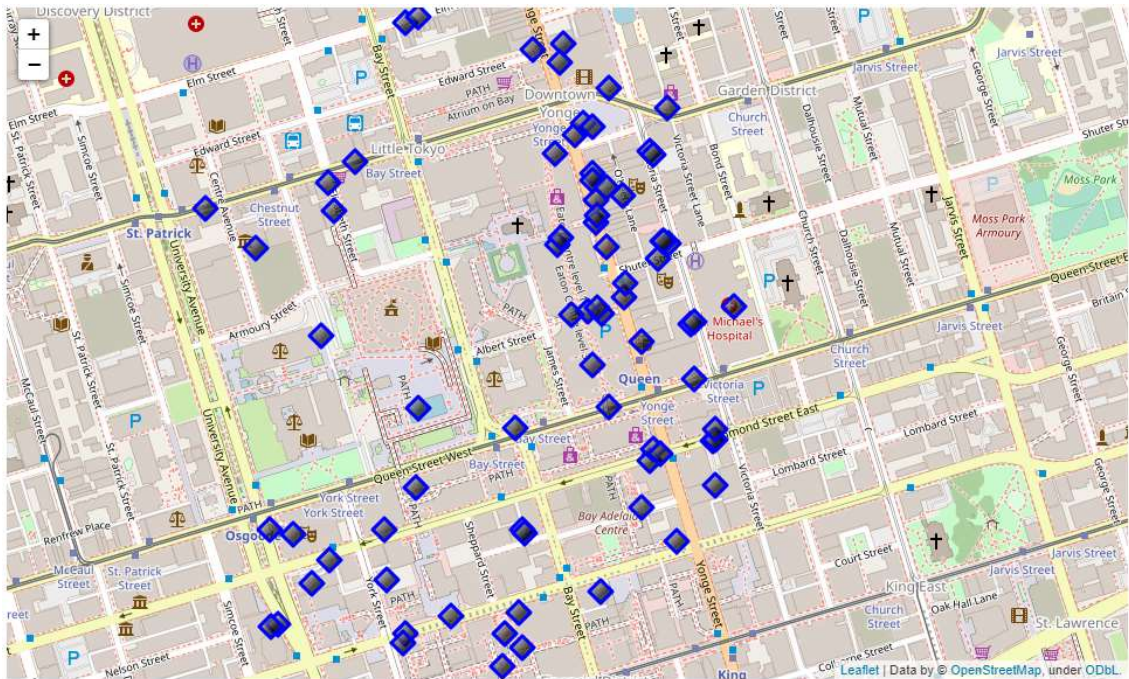
MANHATTAN NEIGHBORHOODS - DATA AND MAPPING

1) Build top 10 Cluster dataframe for near by restaurents, café, Gym, Yoga studio etc. using Fousquare API

Used csv file with cluster neighborhood data was produced with Foursquare during course lab work and create a map.

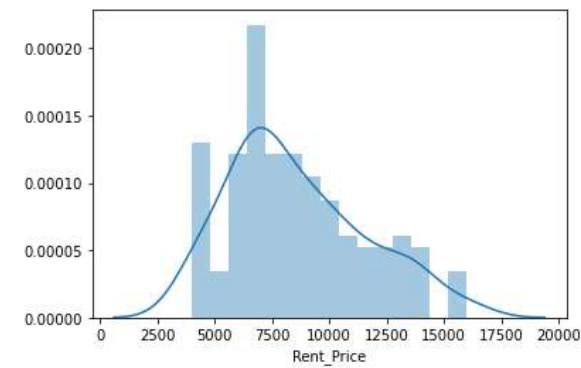|	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place	Park	Gym	Latin American Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Mexican Restaurant	Coffee Shop	Café	Deli / Bodega	Pizza Place	Liquor Store	Indian Restaurant	Sushi Restaurant

Manhattan Rental Data

Extract rental data from web scrapping

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5.0	3400	10000	40.799771	-73.966213
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	40.788585	-73.955277
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	40.799771	-73.966213
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	40.730523	-74.001873
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	40.744118	-73.995299

Price Analysis and Findings

- **Rental price mean value is USD7000 per month as per below graph**

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fea6ecb2a10>
```
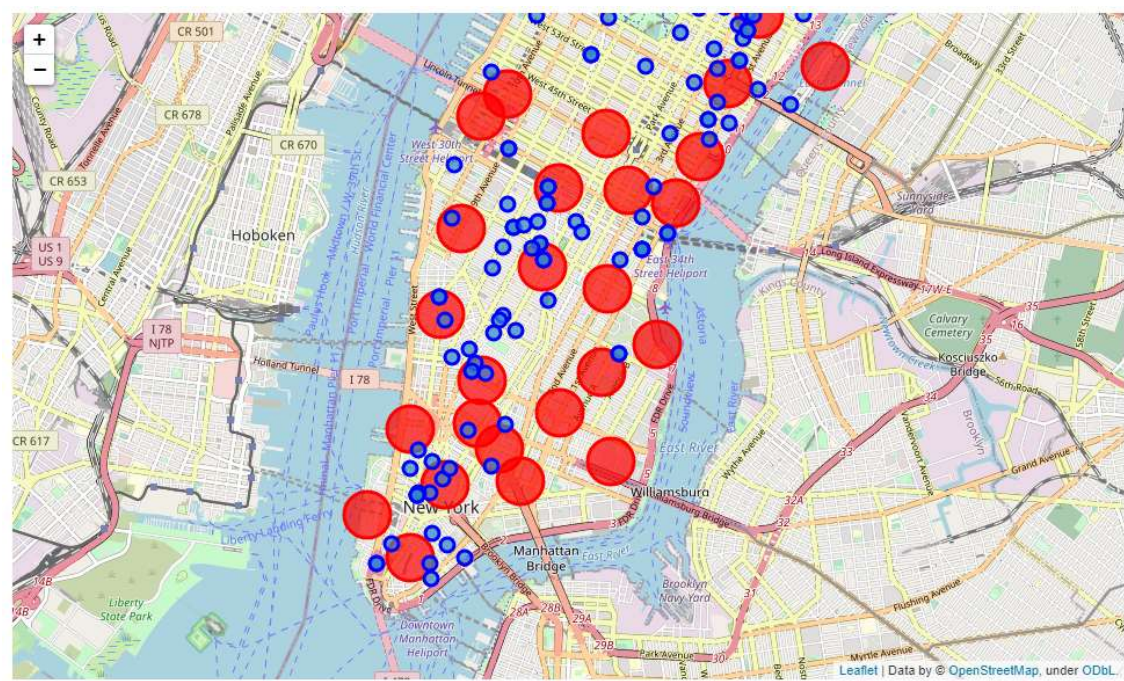

Manhattan Rental Data Map for price and geo location

**Merged Venue Clusters Map and Rental Data Map**

Filtered the cluster numbers to explore.

```
## kk is the cluster number to explore
kk = 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant	Pizza Place	Park	Gym / Fitness Center
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	French Restaurant
36	Tudor City	Café	Park	Pizza Place	Mexican Restaurant	Greek Restaurant	Sushi Restaurant	Hotel	Deli / Bodega	Diner	Dog Run
38	Flatiron	Italian Restaurant	American Restaurant	Gym	Gym / Fitness Center	Yoga Studio	Vegetarian / Vegan Restaurant	Bakery	Clothing Store	Cosmetics Shop	Cycle Studio

Get Manhattan Subway Metro Station List and Mapped
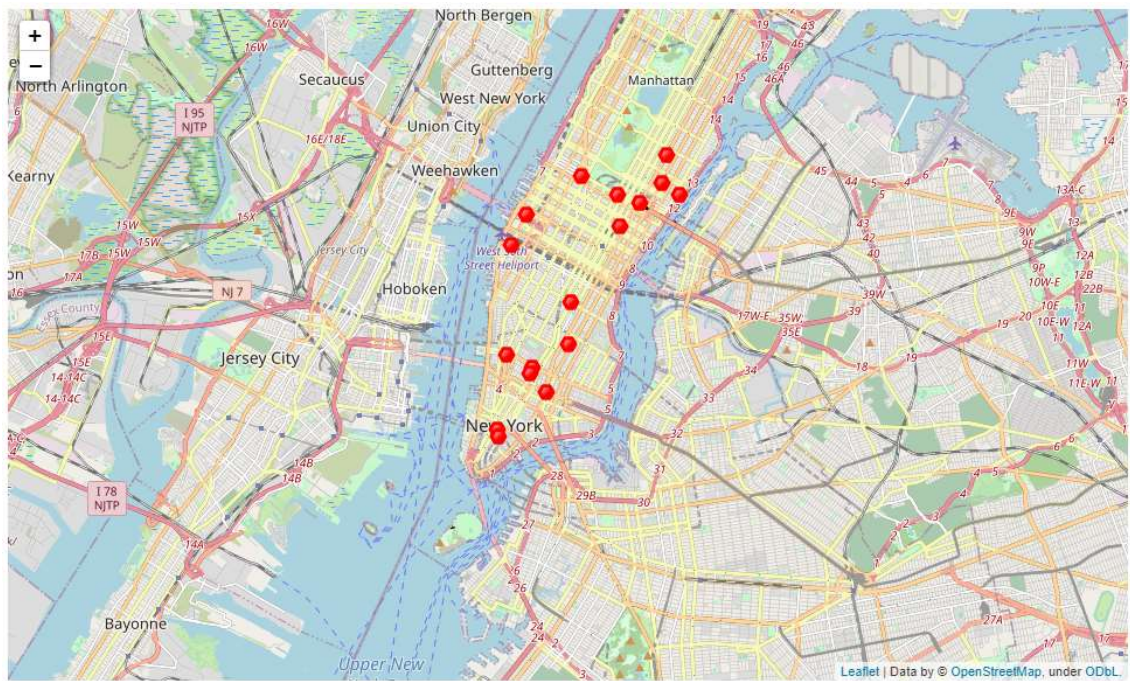
4.0 Results

Now all maps are consolidated into one for final decision making

Property Selection

Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision.

Cluster 2 Sutton Place is the best choice for the rental for the requirements and budget

305 East 63rd Street rental cost is US7500 slightly above the US7000 budget but located 400 meters from "59 St-Lexington Av subway station" and very near to cluster 2, Sutton Place.

150 East 57th Street rent cost is USD5950, under the US7000 budget. The same distance from "59 St-Lexington Av Subway Station" but it is a bit more far from cluster 2, Sutton Place.

Based on all information 305 East 63rd Street is the best choice as proximity to Cluster 2 will make more traffic to the shop.

5. 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. Conclusion

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.