Battle of Neighborhoods
An Apartment in Manhattan

Deepak Singh

Data Science Project

Coursera Capstone

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.

Introduction

Discussion of the business problem and the audience who would be interested in this project.

Description of the Problem and Background

Scenario:

- I am a data scientist residing in Downtown Los Angeles.

- I currently live within walking distance to Downtown metro station and I enjoy many amenities and venues in the area, such as various international cousins restaurants, cafes, food shops and entertainment.

- I have been offered a great opportunity to work for a leader firm in Manhattan, NY. I am very excited and I want to use this opportunity to practice my learning in Coursera in order to answer relevant questions arisen. The key question is :

How can I find a convenient and enjoyable place similar to mine now in Los Angeles?

Certainly, I can 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:

1. Apartment must be 2 or 3 bedrooms

2. desired location is near a metro station in the Manhattan area and within 1.0 mile (1.6 km) radius

3. price of rent not exceed 7000 dollars per month

4. Top amenities in the selected neighborhood shall be similar to current residence.

5. Desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym and food.

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 Lab). List of Manhattan neighborhoods - Wikipedia

- 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!4b1)

- List of apartments for rent in Manhattan area with information on neighborhood location, add beds, area size, monthly rent price and complemented with geo data via Nomination http://www.rentmanhattan.com/index.cfm?page=search&state=results

Data Sources, Data Processing and Tools used

- Manhattan neighborhoods were obtained from Wikipedia and organized by neighborhoods nomination 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 Manhattan. The (latitude, longitude) data was found with algorithm coding and using Nomination.

Methodology

The Strategy to find the answer:

- The strategy is based on mapping the described data in section 2.0, in order choice of at least two candidate places for rent. The information will be consol MAP where one can see the details of the apartment, the cluster of venues in and the relative location from a subway station and from work place. A mea will also be provided. The popups on the map items will display rent price, lo cluster of venues applicable.
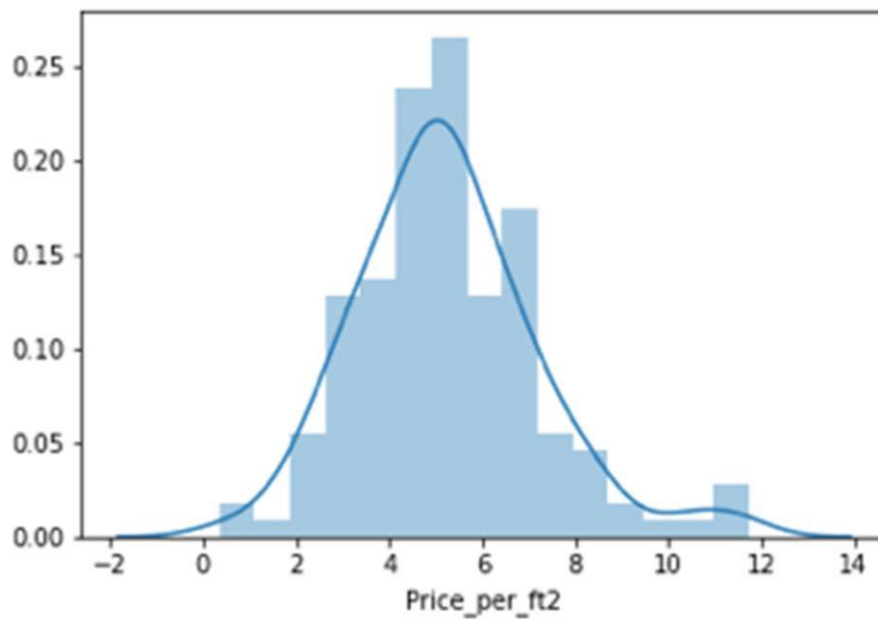
The Tools:

- Web-scraping of sites is used to consolidate data-frame information which files for convenience and to simply the report. Geodata was obtained by coding to use Nomination to get latitude and longitude of subway stations and also units) the apartments for rent listed. Geopy_distance and Nomination were used 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.
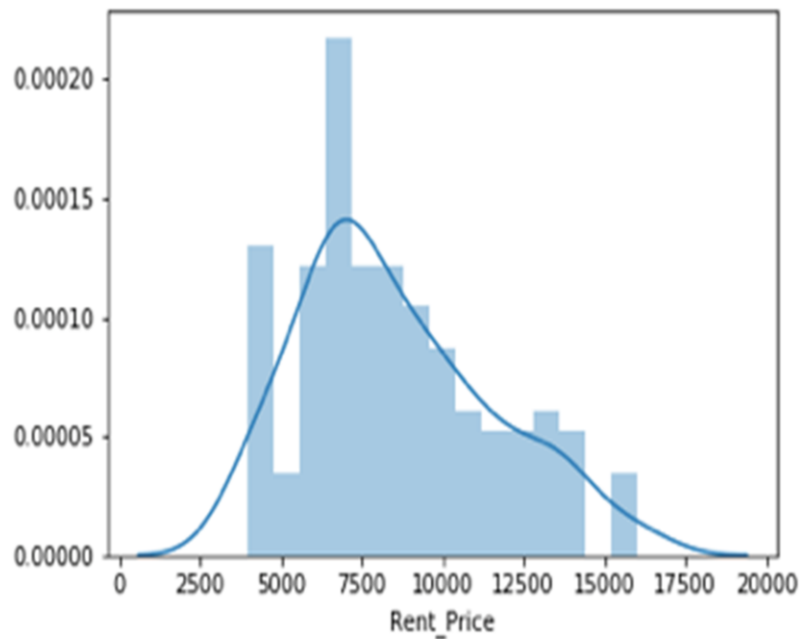
Execution and Results

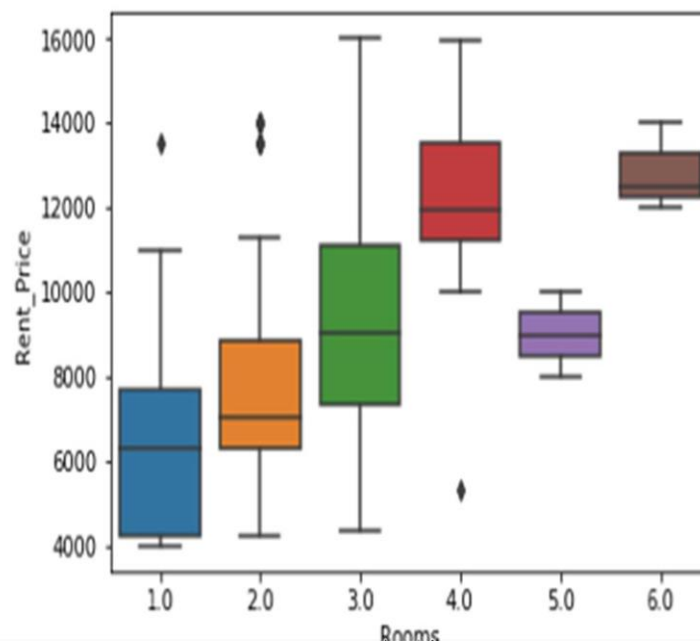## Rental Price Statistics Manhattan Apartments

`<matplotlib.axes._subplots.AxesSubplot at 0x1a2415fc18>`

<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>


```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```
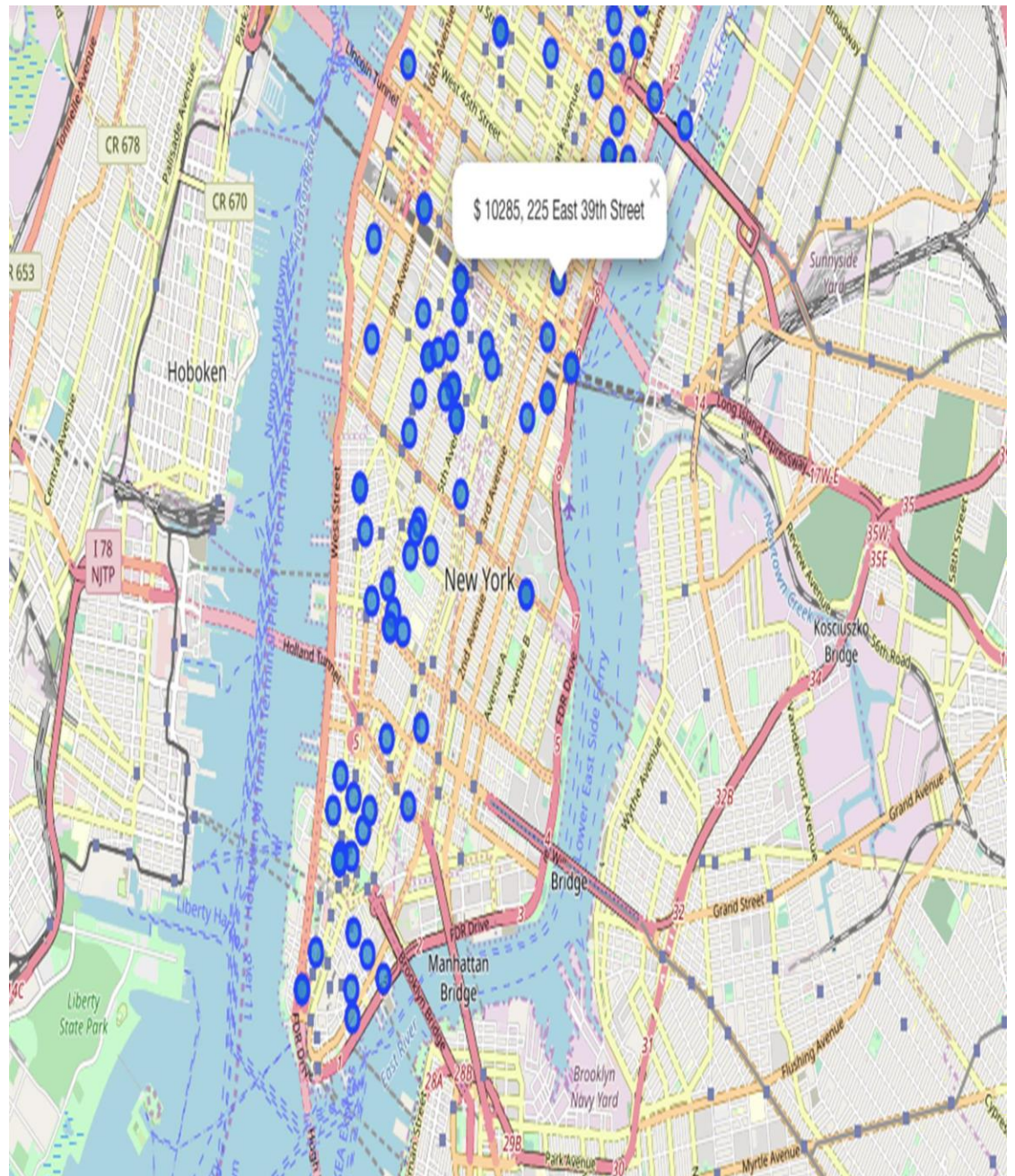
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>

Apartments for Rent in Manhatten

Venues of cluster 3

```
## 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_m
```

	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

Apartment Selection

- Using the "one map" above, I was able to explore all possibilities since the the information needed for a good decision.
- Apartment 1 rent cost is US7500 slightly above the US7000 budget. Apt 1 400 meters from subway station at 59th Street and work place (Park Ave and 600 meters way. I can walk to work place and use subway for other places a this apt are as of Cluster 2 and it is located in a fine district in the East side
- Apartment 2 rent cost is US6935, just under the US7000 budget. Apt 2 is from subway station at Fulton Street, but I will have to ride the subway dail possibly 40-60 min ride. Venues for this apt are as of Cluster 3.¶

Discussion

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

- I feel this Capstone project presented me a opportunity to practice and apply the Data tools and methodologies learned.

- I have created a good project that I can present example to show my potential.

- I feel I have acquired a good starting point to kick start my career again.

Conclusions

- I feel rewarded with the efforts and time, I believe this course with all the topics covered is of appreciation.

- This project has shown me a practical application a real situation that has impacting personal and impact using Data Science tools.

- The mapping with Folium is a very powerful te consolidate information and make the analysis a thoroughly and with confidence. I would recommend use in similar situations.