Data Science Capstone

Kevin Winston
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

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 (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan • Venues and amenities as in my current residence.

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. Data Section

Description of the data:

location data of the city

Information regarding the amenities of the various locations in the city

The property prices of various locations within the city

The accessibility of various locations in the city

How will the data be used to solve the problem:

The data will be used as follows

- Use Foursquare and geopy data to map the venues for the city and cluster it in groups
- Use foursquare and geopy to map the various ammenities on top of the clustered map inorder to be able to easily explore various locations
- Use the data concerning property prices of the various locations to map it linking it to the foursquare data.
- To analyse various trends and habits of the prices and features concering the ammenities of each location.
- The required data will be searched for in open research channels and data sources

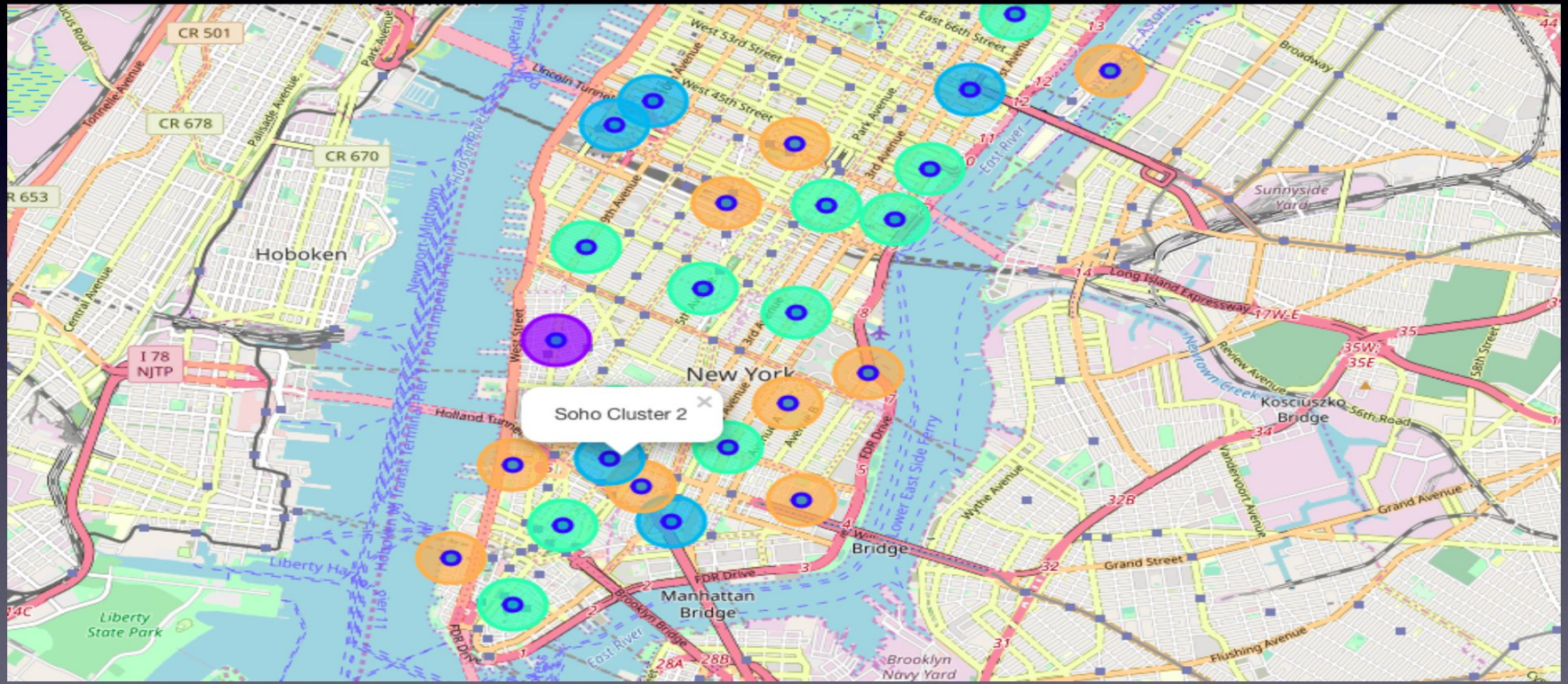
3. Methodology Section

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

- what is the cost of available rental places that meet the demands?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data

4. Results and Execution

Manhattan Map - Neighborhoods and Cluster of Venues



GeoData Manhattan apts for rent

```
]: mh_rent=pd.read_csv('MH_rent_latlong.csv')
   mh_rent.head()
```

]:

	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

```
]: mh_rent.tail()
```

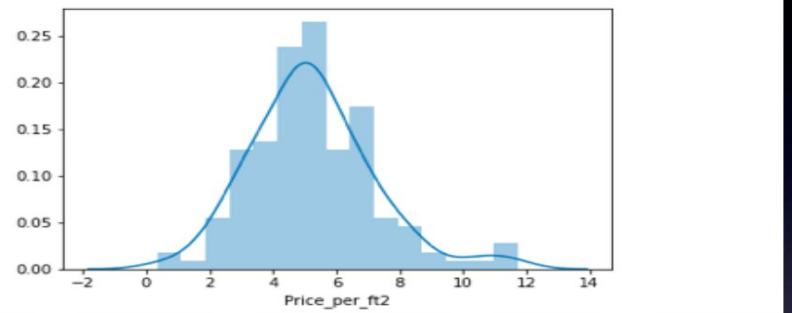
]:

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	8118	40.758216	-73.965190
142	1930 Broadway	No fee rental in Central Park West	5.06	2.0	1600	8095	40.772474	-73.981901
143	33 West 9th Street	Rental in Greenwich Village	6.67	2.0	1500	10000	40.733691	-73.997323

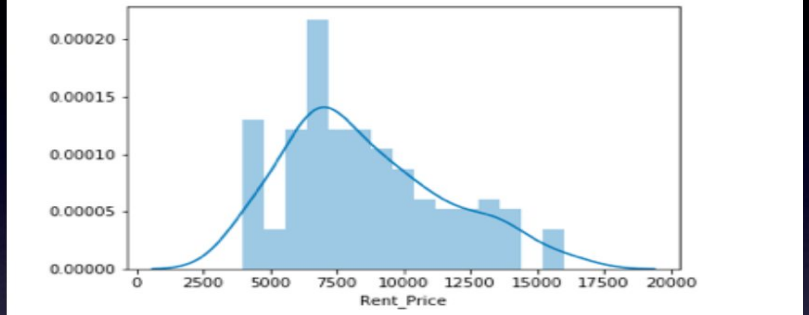
Rental Price Statistics MH Apartments

Budget US7000/month is around the mean

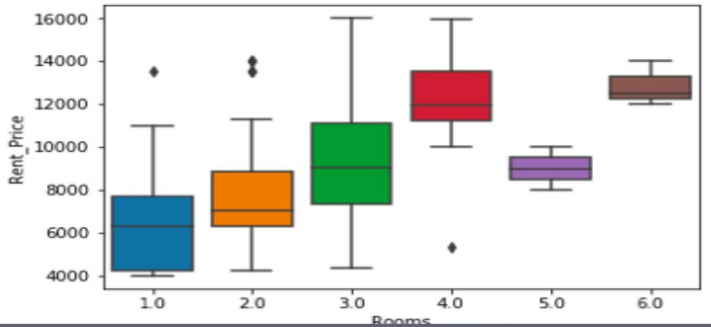
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2415fc18>
```



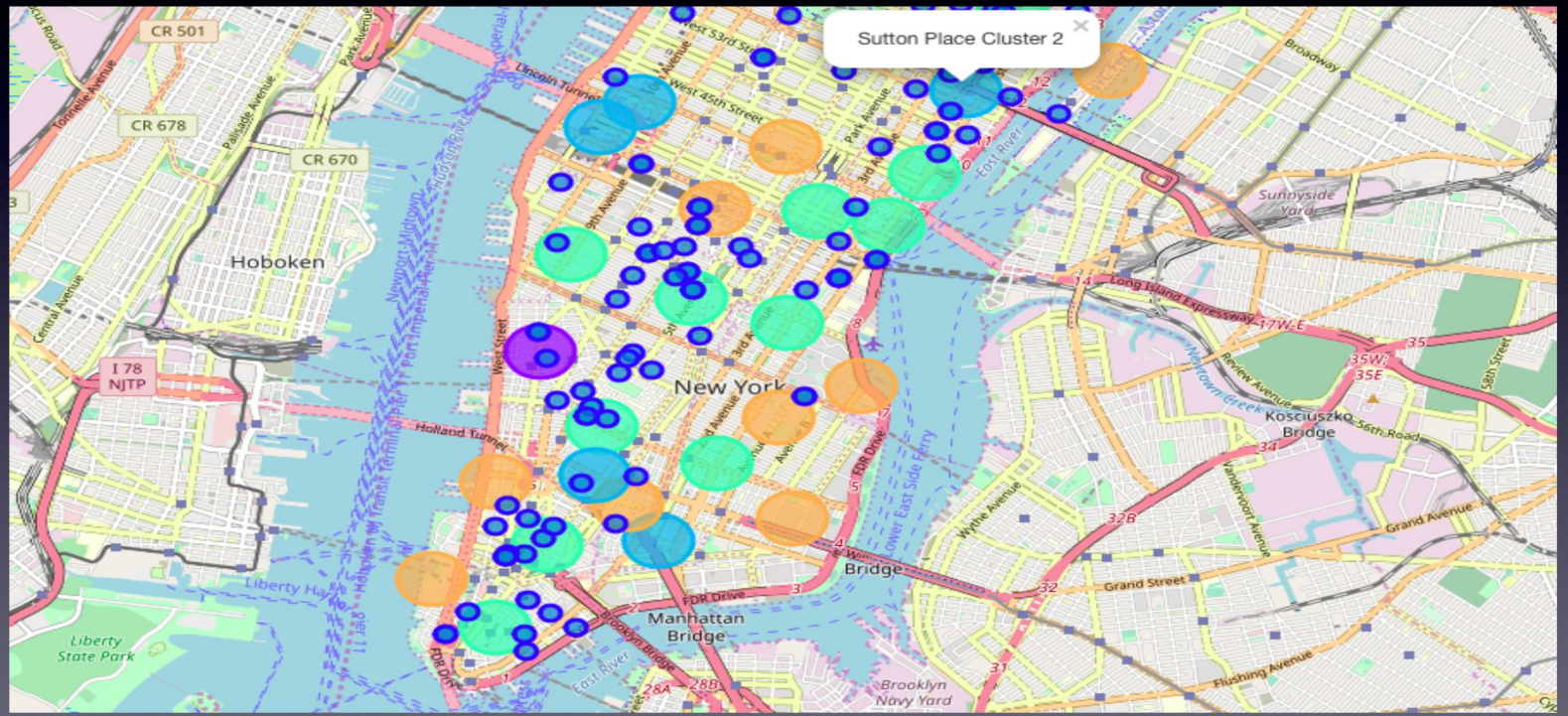

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```




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



MH apts for rent with venue clusters



5.0 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