

Coursera Capstone project

Coursera IBM Data Science Certification

Abhijit R. Parab

Mar. 25th, 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



**1.1 Scenario and Background**

I am currently living in India, Goregaon suburb in Mumbai city. I also enjoy great venues and attractions, such as international cuisine, entertainment and shopping. I have an oﬀer to move to work to Manhattan NY

and I would like to move if I can find a place to live similar with similar venues.

**1.2 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.

**1.3 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.

2.0 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 in Course

Lab). 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://ww](http://www.google.com/)w[.google.com/](http://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, 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](http://www.rentmanhattan.com/index.cfm?page=search&amp;state=results) [https://ww](http://www.nestpick.com/search)w[.nestpick.com/sear](http://www.nestpick.com/search)ch? city=new-

- Place to work in Manhattan (Park Avenue and 53rd St) for reference

**2.2 Data Sources, Data Processing and Tools used**

- Singapore data and map is to be created with use of Nominatim , Foursquare and Folium mapping

- Manhattan neighborhoods were obtained from Wikipedia and organized by Neighborhoods with geodata via Nominatim 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 MH. The geolocation

(lat,long) data was found with algorithm coding and using Nominatim.

- Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed to make a selection of apartment

3.0 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:

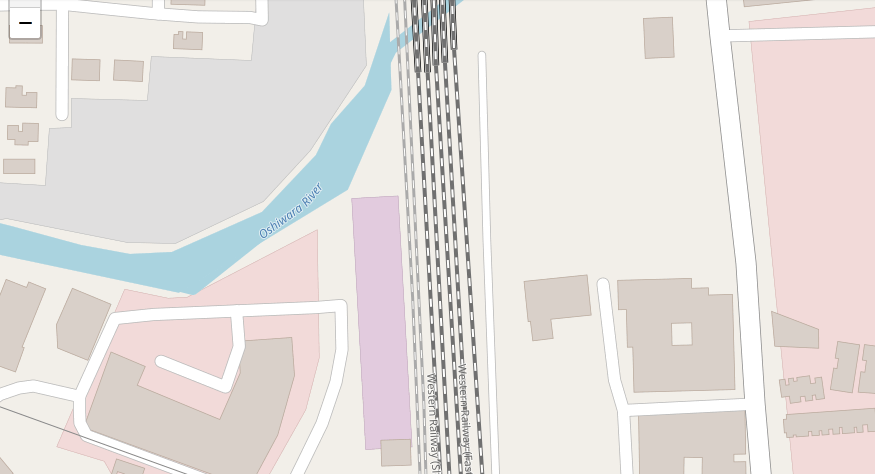
Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent 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



4.0 Execution and Results

**My place in India**





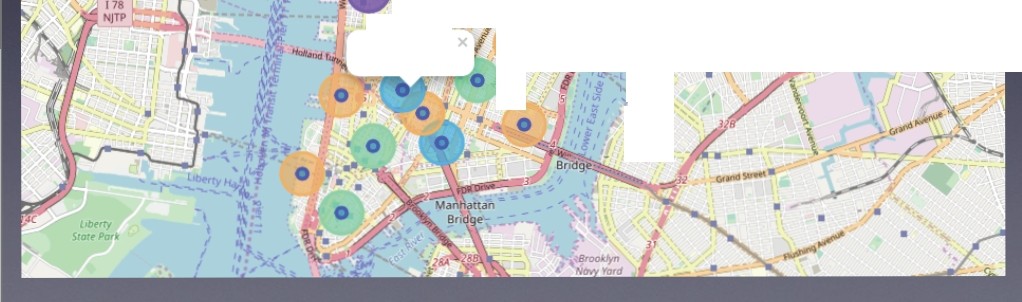


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

• •

**Hoboken**

178



•

NJTP

•

Soho Cluster 2

*.!.f*

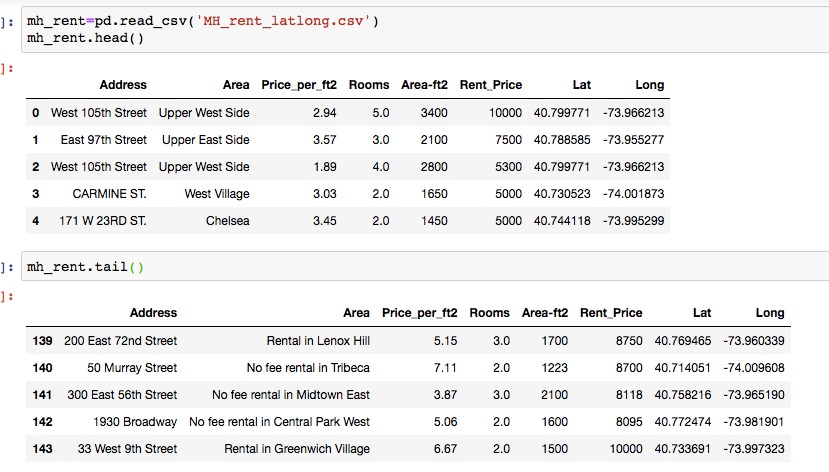
I'

**f**..

·

;;

/ .



GeoData Manhattan apts for rent

<matplotlib.axea.\_aubplota.AxaaSubplot at Oxla2415fcl8> <matplotlib.axes.\_subplota.AxeaSubplot at Oxla25dd8400>

025 0.00020

020

cus

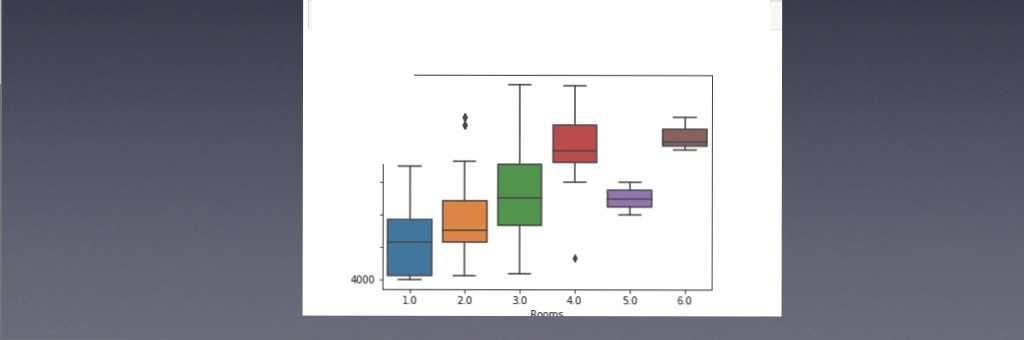
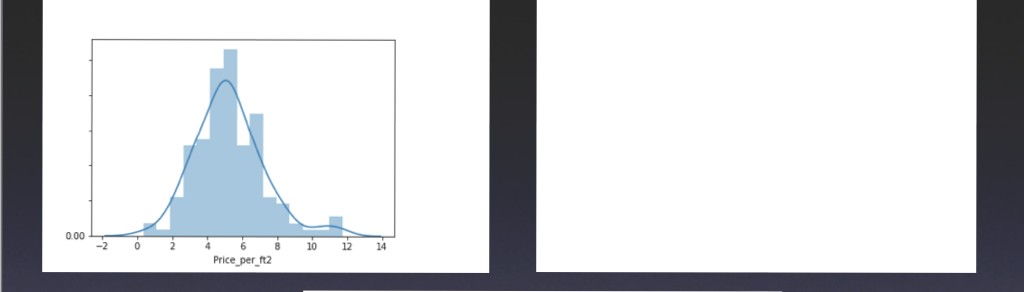
010

000015

000010

005

000005



000000 -- ---- - -A L---

0 2500 5000 1500 10000 U500 15000 11500 *20000*

**Rcnt\_Pnce**

ana.boxplot(x•'Rooms', ya 'Rent\_Price', data-mh\_rent)

<matplotlib.axes.\_subplots.AxesSubplot at Oxla25f2a2b0>

16000

14000 •

12000

**a., lOOOO**

!eooo

6000

Hoboken

.'

I

, ''



I

*I*

*I I*

'

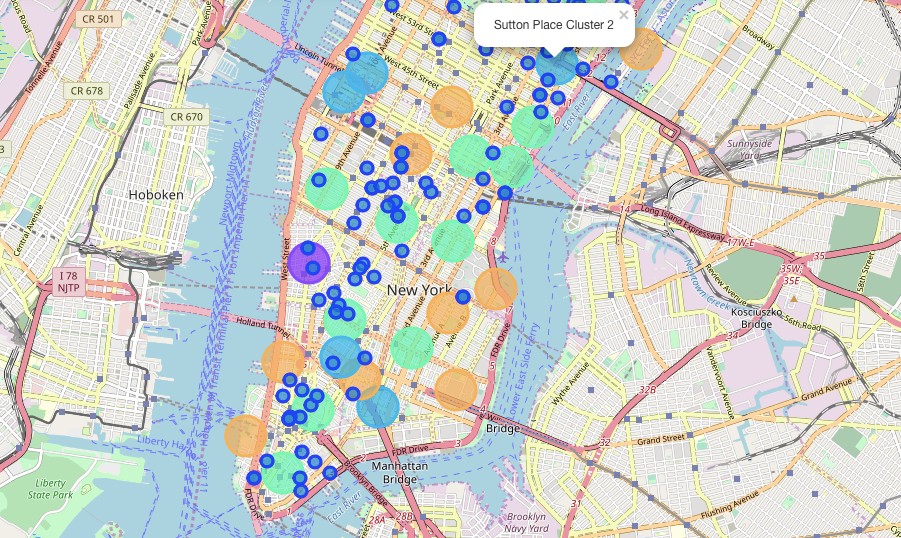
*I*

• •

·

.

•·"""' '



MH apts for rent with venue clusters

## *kk is the cluster number* to *explore*

kl< = 3

manhattan\_merged.loc(manhattan\_merged('Cluster Labels'] == kl<, manhattan\_merged.columns([l]+ list(range(S, manhattan n

3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1st Most | 2nd Most | 3rd Most | 4th Most | 5th Most | 6th Most | 7thMost | 8111 Most | 9th Most | 10th Most |
| Nelghbomood | Common | Common | Common | **Common** | Common | Common | Common | Common | Common | Common |
|  | Venue | Venue | Venue | venue | Venue | venue | Venue | Venue | Venue | venue |
| Inwood | Mexican  Restaurant | Lounge | Pizza Place | care | Wine Bar | Bakery | American  Restaurant | P811< | Frozen  Yogurt Shop | Spanish  Restaurant |
| Manhattanville | Dell/ Bodega | Italian  Restaurant | Seafood  Restaurant | Mexican  Restaurant | Sushi  Restaurant | Beer Garden | Coffee Shop | Falafel  Restaurant | Bike Trail | Other  Nightlife |

5

10 Lenox Hill

Upper West

12 Side

16 Murray Hill

Sushi

Restaurant

Italian

Restaurant

Sandwich

Place

Italian Coffee Shop

Restaurant

Bar Bakery

Hotel Japanese

Restaurant

Italian Ice Cream

Gym *I* Fitness

Center

Vegetarian *I Vegan* Restaurant

Gym *I* Fitness

Center

Pizza Place

Indian

Restaurant

Coffee Shop

Burger Joint

Coffee Shop

Salon/ Barbershop

Dell/ Bodega

Cosmetics

Shop

Burger Joint

Gym Sporting Thai

Goods Shop Restaurant

Wine Bar Mexican Sushi

Restaurant Restaurant

French Bar Italian

Restaurant Restaurant

Seafood Amellcan

17 Chelsea Coffee Shop

Restaurant Shop

Bakery

Nightclub

Theater

Art Gallery

Restaurant Restaurant

Hotel

Greenwich Italian

18 Village Restaurant

Sushi French

Restaurant Restaurant

Clothing Store

Chni ese

Restaurant

care

Indian

Restaurant

Bakery

Seafood Electronics

Restaurant Store

Gramercy Italian

27

Restaurant

Restaurant Thrift *I*

VIntage Store

Cocktail Bar

Bagel Shop

Coffee Shop

Pizza Place

Mexican Grocery Wine Shop

Restaurant Store

Financial Coffee Shop

29

District

Hotel Gym

Wine Shop Steakhouse

Bar

Italian

Restaurant

Pizza Place Pari< Gym *I* Fitness

Center

31 Noho Italian

French CocktailBar

Gilt Shop

Bookstore Grocery Store

Mexican

Hotel Sushi Coffee Shop

: c:.... ..-;o : :.- • ,.., : *ro* :· r.o



. .

sub addntss Iat long

ellek to sero outpUt;double el'<:< to hide

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0 | Dyckman Stte« Subway S!lii!Oil | 170 Nagle Ave.New York. NY 10034,USA | 40.861857 | -73.924509 |
| 1 | 57 Street Subway Slation | New Yof1(, NY 10106.USA | 40.764250 | ·73.954525 |
| 2 | Broad St | New York. NY 10005,USA | 40.730862 | -73.987156 |
| 3 | 175 Street Slation | 807 W 1nth St, New YOI1<,NY 10033,USA | 40.847991 | -73.939785 |
| 4 | 5 Av llld 53 St | New York. NY 10022.USA | 40.784250 | -73.954525 |

# *removing duplicate rows* and *creating new set mbsubl*

mhsubl-m.h.d.rop\_duplicates(subset ('lat','long'],lteep-"last").reset\_index(drop-True)

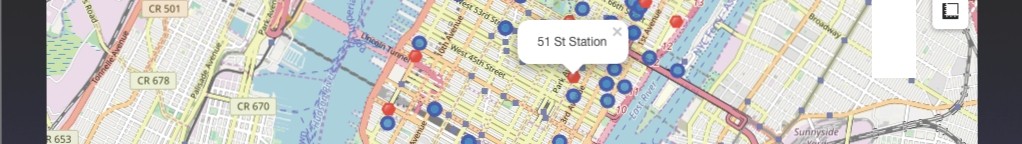
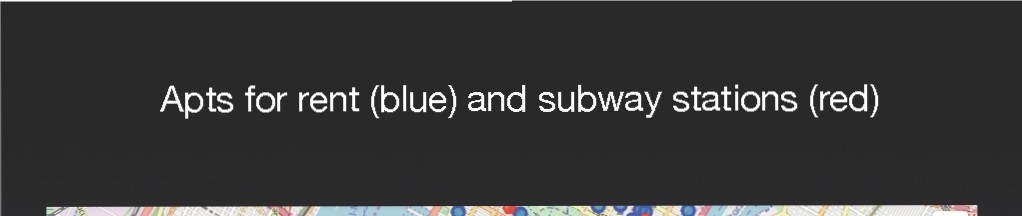
mhsubl.shape

(22, 4)

mhsubl.tail()

sub station aub\_addrna lat long

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 17 | 190 Street Subway Station | Bennett Ave,New York,NY 10040,USA | 40.858113 | -73.932983 |
| 18 | 59 SH.exmgton Av Station | E 60th St,New Y011<,NY 10065,USA | 40.762259 | -73.966271 |
| 19 | 57 Street Station | New York,NY 10019,United States | 40.764250 | -73.954525 |
| 20 | 14 Streel/8 Av | NewYOI1<,NY 10014,United States | 40.730862 | -73.987156 |
| 21 | MTA New YOI1< C.ty | 525 11th Ave,New York, NY 10018,USA | 40.759809 | -73.999282 |

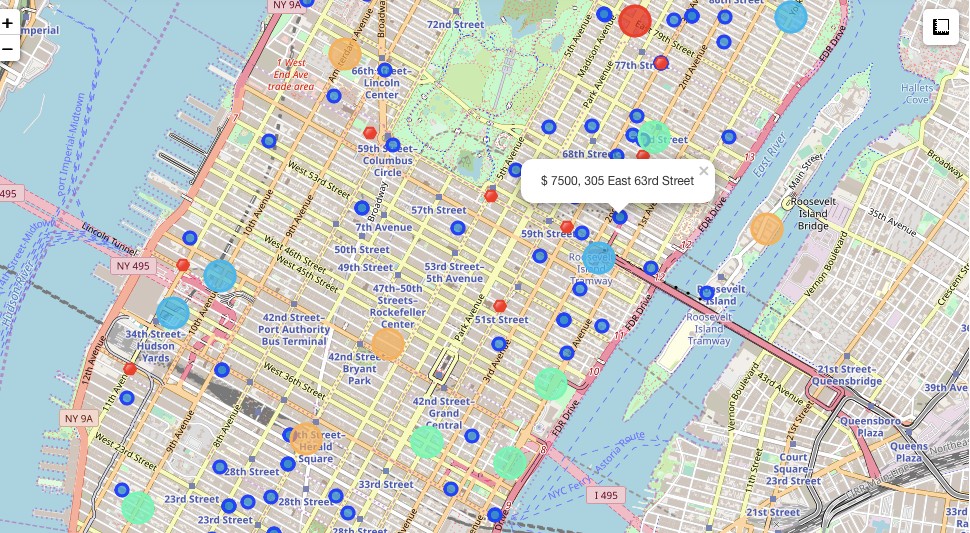


• •

•

Ho•boken

Selected Apartment!



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 station, Bubbles=Cluster of Venues



Apartment Selection

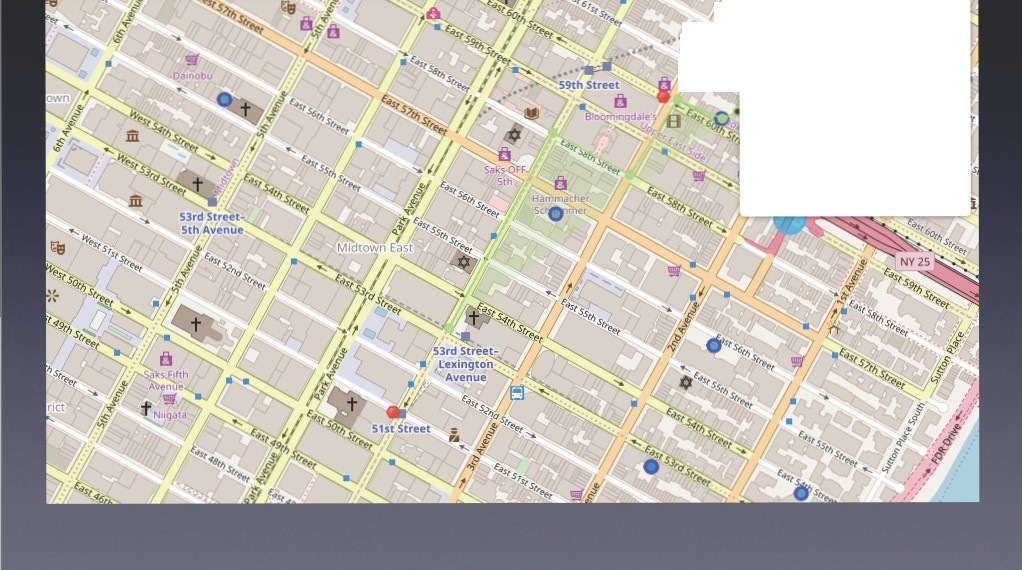
Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision.

Apartment 1 rent cost is US7500 slightly above the US7000 budget. Apt 1 is located

400 meters from subway station at 59th Street and work place ( Park Ave and 53rd) is another 600 meters way. I can walk to work place and use subway for other places around. Venues for this apt are as of Cluster 2 and it is located in a fine district in the East side of Manhattan.

Apartment 2 rent cost is US6935, just under the US7000 budget. Apt 2 is located 60 meters from subway station at Fulton Street, but I will have to ride the subway daily to work , possibly 40-60 min ride. Venues for this apt are as of Cluster 3.¶

Based on current Singapore venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that APARTMENT 1 is a better choice since the extra monthly rent is worth the conveniences it provides.



••*:f/'4,*.I easure dostances and areas

*I* ... 40" 45'43.26" N 73' 57' 52.85'W

40.762016 -73.964682

p 1? 837 Meters (05. 2 Mlies) "' 47,328SqMeters(11.70Acres)

0 Cancel 0 Fonlshmeasurement

## *kk is the cluster number* to *explore*

kl< = 2

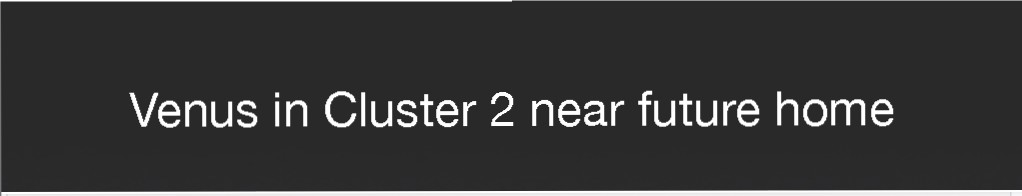
manhattan\_merged.loc(manhattan\_merged('Cluster Labels'] == kl<, manhattan\_merged.columns((l]+ list(range(S, manhattan n

Restaurant

Restaurant Restaurant Barbershop House Shop Shop

Restaurant

Restaurant Shop Restaurant Center



Italian Sushi Pizza Place Mexican Dell/ Bodega Japanese Pub

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nelghbor1>ood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most **Common** Venue | 4th Most Common venue | 5thMost Common Venue | 6thMost Common Venue | 7th Most Common venue | 8th Most Common Venue | 9th Most Common venue | 10th Most Common venue |
| 0 Marble Hill Coffee Shop | | Discount  Store | Yoga Studio | Steakhouse Supplement Tennis Shoe Store Gym Bank Seafood  Shop Stadium Restaurant | | | | | | |
| 1 Chinatown Chinese | | CocktailBar | Dim Sum  Restaurant | American Vietnamese Salon/ Noodle Bakery Bubble Tea Joe Cream | | | | | | |
| 6 CentralHarlem African | | Seafood  Restaurant | French  Restaurant | American Cosmetics Chinese Event Space Liquor Store Beer Bar Gym/Fimess | | | | | | |

9 YO<kvllle Coffee Shop Gym Bar

Restaurant Restaurant Restaurant Restaurant

Italian American Gym/

14 Clinton Theater Restaurant Coffee Shop Restaurant Fitness Hotel Wine Shop Spa Gym lndle Theater

Center

Clothing women's Men's Store Furniture *I* Italian Mediterranean Art Gallery Design Studio

23 Soho Boutique Shoe Store

Store Store Home Store Restaurant Restaurant

Morningside Coffee Shop Ameflcan Park Bookstore Pizza Place Sandwich Burger Joint Cat6 Deli *I* Tennis Court

26 Heights Restaurant Place Bodega

Gym / Italien Furniture / Indian American Sushi

34 Sutton Place Fitness Restaurant Home Store Restaurant Dessert Shop Restaurant Bakery Juice Bar Boutique Restaurant

Center

ltallen American Gym / Thai

39 Hudson Yards Coffee Shop Restaurant Hotel Theater Restaurant cate Fitness Restaurant Restaurant Gym

Center

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.

6.0 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.