

Coursera Capstone project

Coursera IBM Data Science Certification

Sachin Yadav

## Report Content

1. Introduction Section :

* The “business problem” to be solved by this project and who may be interested

1. Data Section:

* Describe Data requirements and Sources needed to solve the problem

1. 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.

1. Results section:

* Discussion of the results and finding of answer

1. Discussion section:

* Discussion of observations noted and any recommendations

1. Conclusion section:

* Answer chosen and conclusions.

## Introduction

###### Scenario and Background

I am currently living in Singapore, within walking distance to Downtown "Telok Ayer MRT metro station" . 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.

###### 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.

## Data Section

* 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://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](https://www.google.com/maps/search/manhattan%2Bsubway%2Bmetro%2Bstations/%4040.7837297%2C-74.1033043%2C11z/data%3D!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&state=results) [https://www.nestpick.com/search? city=new-](https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list)
* Place to work in Manhattan (Park Avenue and 53rd St) for reference
  1. **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.

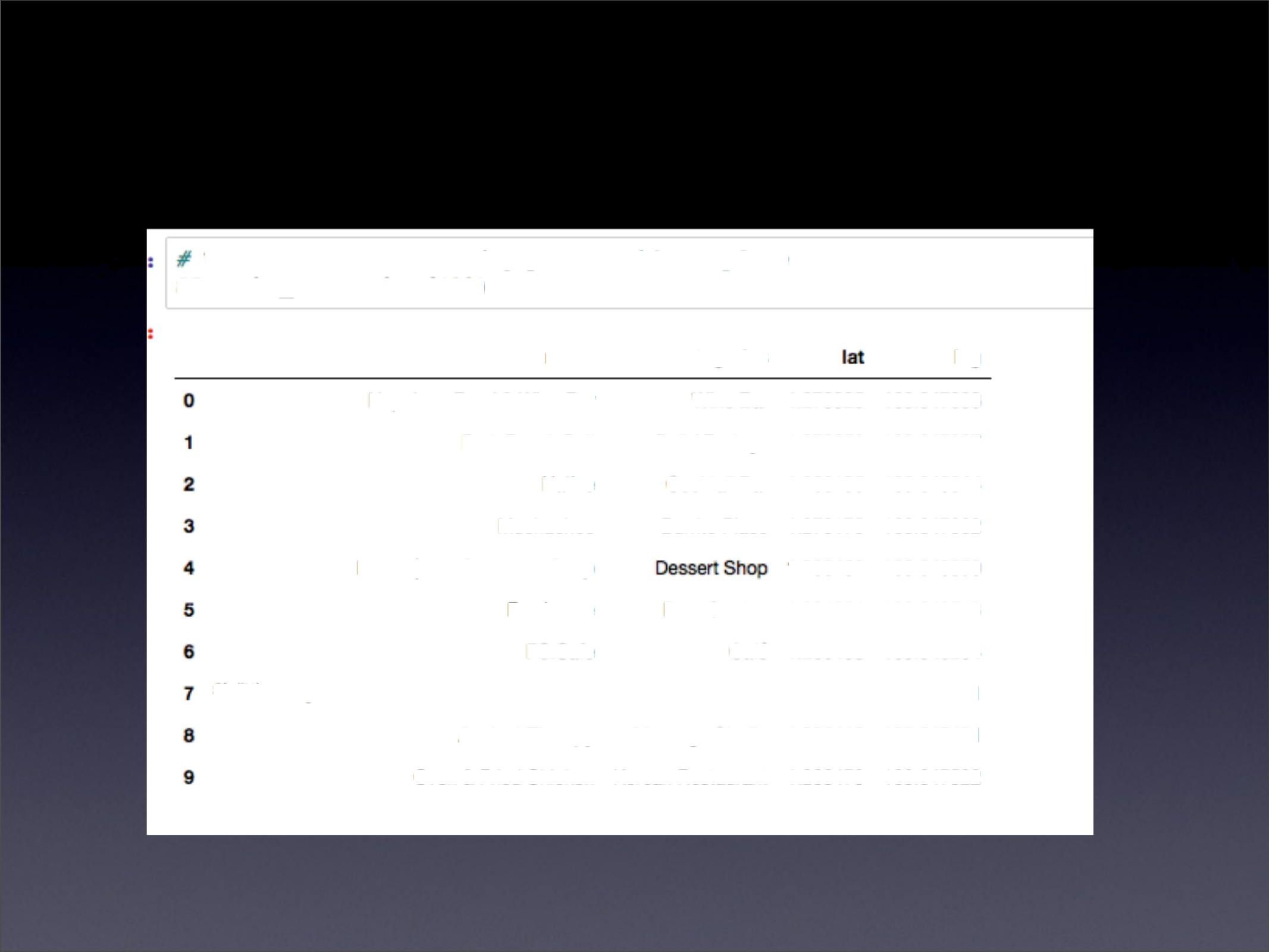
Geopy\_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data.

Maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy



## 4.0 Execution and Results

#### Current residence Neighborhood in Singapore

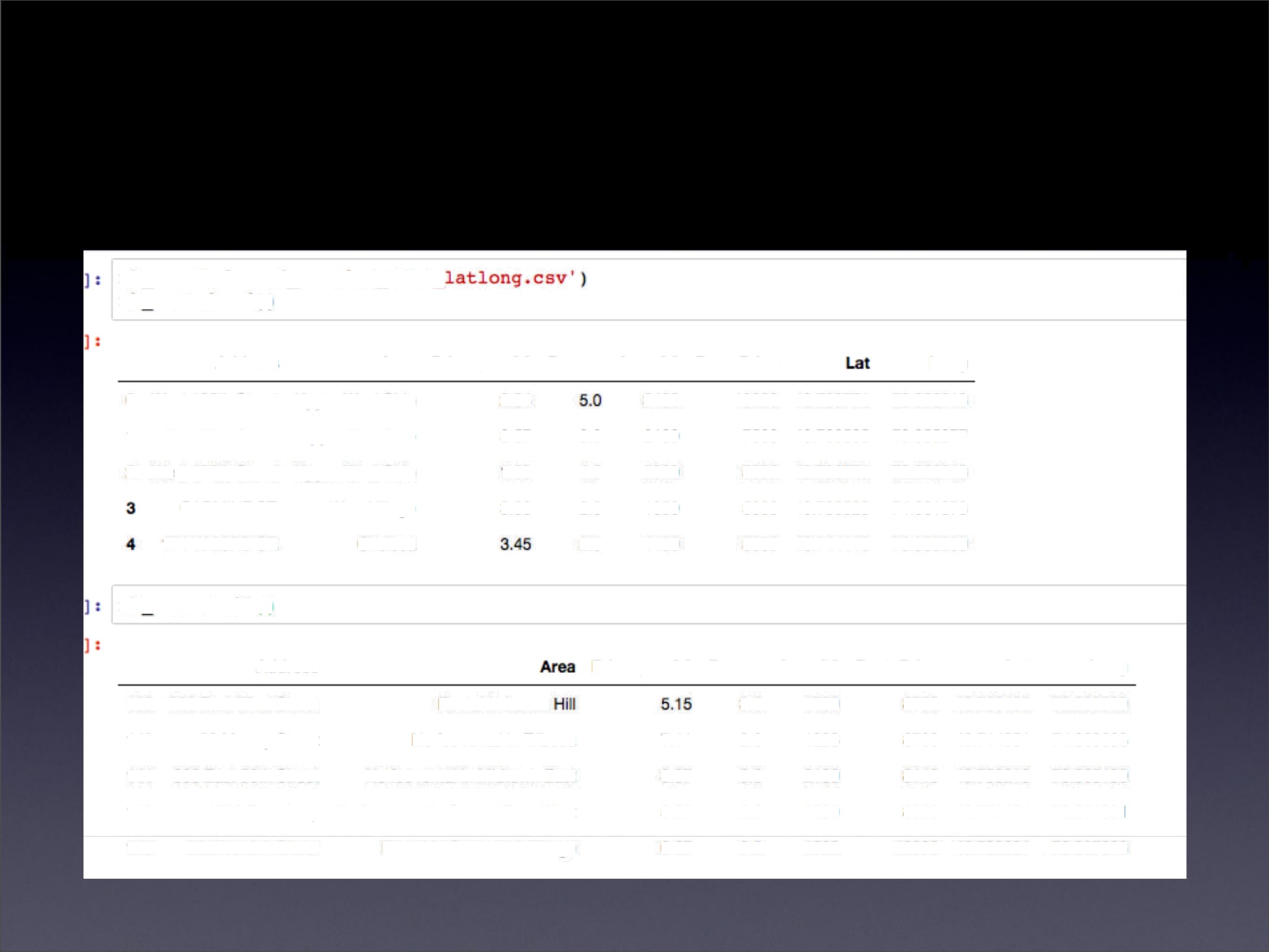
Venues around Neighborhood in

Uenuea *neav* current *S:ingnpove veii?dence plxce*

S0nearby\_venuea.head(10)

|  |  |  |  |
| --- | --- | --- | --- |
| name | cmxae |  | |
| Napoleon Fmd & Rne Bar | Whe Ba | 1.274025 | 10.847333 |
| Pak Bench I%II | Dai / Bodega | 1.274872 | 10.847287 |
| Native | CocDgl Ba | 1.280135 | 10.846844 |
| Muchachos | Burñto Race | 1.279175 | 10.847082 |
| Mau's | The Chocda@ SI&p |  | 1.280462 | 10.846950 |
| Ffeel1ouse | Beer Garden | 1.281254 | 10.848513 |
| m.Ca» | Café | 1.280468 | 10.840264 |
| &/1l°4 Wang Das Bak Kofaan BBO Rwisurant | Koaan Restaurant | 1.281345 | 10.847551 |
|  | Massage Studb | 1.280413 | 10.847481 |
| Oven & Fried Chkken | Korean Restaurant | 1.280479 | 10.847522 |

#### Manhattan Map - Neighborhoods and Cluster of Venues



GeoData Manhattan apts for rent

mh\_rent-pd.read\_csw('Ma\_rent\_

& xent.hea4( )

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Addre¥a | Area | Pdoe\_per\_O | Rooms | Area-fD | Rem\_PrIce | Long |
| 0 West 105% Stfeat | Upper Wast SDe | 2.44 |  | 340 | 1AQ0 40.746771 | -73.066213 |
| 1 East97th Street | UpperEast Side | 3.57 | 3.0 | 21QI | 7500 40.788585 | -73.955277 |
| 2 ¥\bO 105Q Stfeat | Upper West 5k¥e | 1.86 | 4.0 | 28tXl | 5300 40.746771 | -73.466213 |
| CARMINE ST. | V&st Vllage | 3.03 | 2.0 | 1650 | 5000 40.7305Z3 | -74.fXl1873 |
| 171 W23RDST. | Chelsaa |  | 2.0 | 1450 | 5000 40.744118 | -73.D95290 |

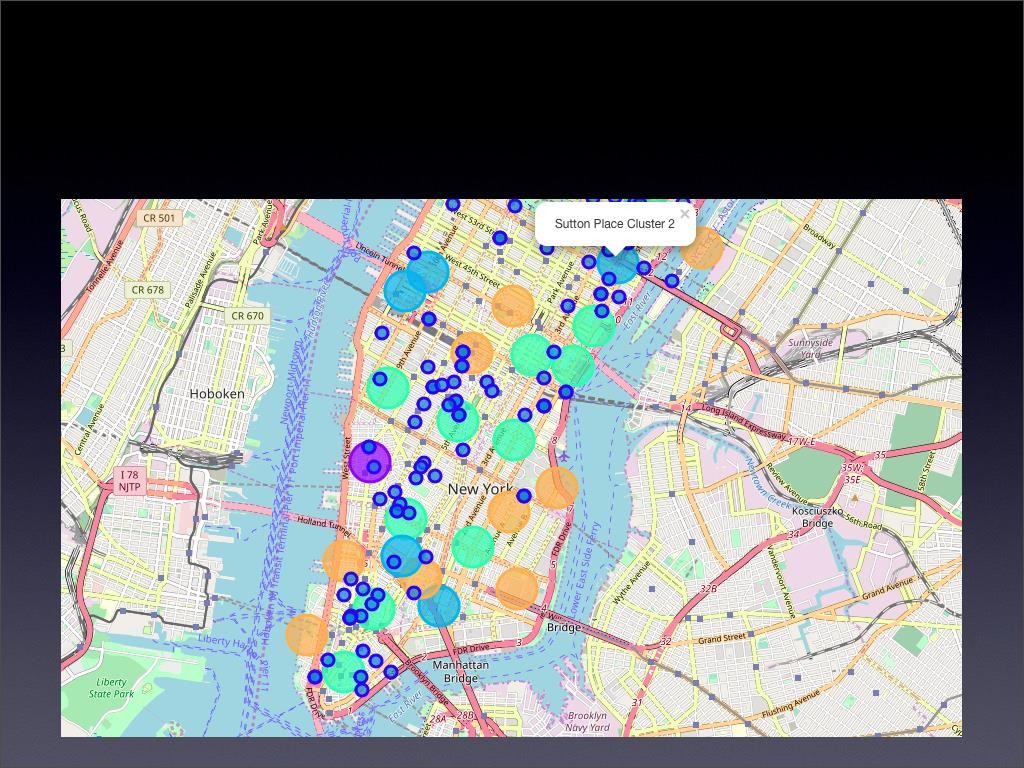
mh rent.tail(1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Addreea | Pf1ce„per\_fD | Rooma | Are&-fD | RefN\_Pdce | L»i | Long |
| 1A | 2IAt East 72nd Sttaet | Rana In Lef\aX | 3.0 | 1700 | 87D | 40.7@465 | -73.960039 |
| 140 | 50 Murray Stfeet | No fee fental In Tl1beca 7.11 | 2.0 | 1223 | 87Q\ | 40.714051 | -74.O0B608 |
| UI | 3IXI East 56th Street | No fee ‹ernal in MODWf\ East 3.87 | 3.0 | 2100 | 8118 | 40.zsa21a | -73.9651N |
| 14Z | 1930 Braadway | No fee rental In Central Park \/\ st 5.06 | 2.0 | 1600 | 8095 | 40.772474 | -73.981901 |
| 1N | 33 West Oth Street | RentB in Greenwich 1Ilage 6.67 | 2.0 | 1500 | 1IXXX\ | 40.735601 | -73.9973Z3 |

#### Rental Price Statistics MH Apartments

Budget US7000/month is around the mean

## Apartments for Rent in MH



MH apts for rent with venue clusters

### Venues of cluster 3

manhattan\_mergRd.loc[manhattan\_mRrged['CluStRr Labels' ; == kk, manhattan\_mRrged.columns[[1] + list(range(5, manhattan\_n

Bakery

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1stMost | 2ndMost | 3rd Mast | Cth Most | Cth Mast | Bth Most | 7th Mast | 8th Mast | 9th Most | 10th Mast |
| **Nelghborhood** | **Common** | **Common** | **Common** | **Common** | **Common** | **Common** | **Common** | **Common** | **Common** | **Common** |
|  | Ve.nul2 | Venue. | Venue | Venue | Venue | Venue | Venue | Venue | Venue | Venue |

Inwood

MRxican Restaurant

Lounge Pizza PITCH Café Wind Bar

American Restaurant

Park

Frazen Spanish

Yogun Shop Resaurant

Bike Trail

s Manhartanville

DCII / Italian Seafood Mexican Sushi Bodega Restaurant Resiaurant Restaurant Resiaurant

Beer GardRn CDffeR Shop

Falafel Restaurant

OthRr NighlIife

1o LRnox Hill

Sushi Italian

Restaurant Resmurant

CDffeR Shop

Dym / Fitness

Center

PiZZA PlacR BurgRr Joint

Deli / Bodeoa

Sporting Thai Goeds Shop Resaurant

1s Upper

West Italian

Side Restaurant

Bar

BakRry

Vegetarian /

Vegan Restaurant

Indian Resiaurant

Called Shop

Cosmetics

Shop

Wind Bar

Mexican Sushi

Restaurant Restaurant

1. Murray Hill

Sandwich

Place

Hotel

Japanese• Dym / Fitness Restaurant Center

CDfteR Shop

Salon / BarbRf5hop

Burger Joint

French Resaurant

Bar

Italian

Restaurant

1. ChRlsea Coffg•0 Shop

Italian let CfOdM

Restaurant Shop

Restaurant

Bakery Nightclub

Theater An GaIIRD/

AmeflC an Restaurant Restaurant

GreRnTch Italian Sushi French

1. Village Restaurant Resmurant Restaurant

Clothing Store ChinesR

Indian Restaurant

BakRry

Seafood ElOCtrDnics

Resmurant Store

27 Grarnercy

Italian Restaurant

Restaurant

Thrift /

Cocktail Bar Bagel Shop Called Shop Pizza Place

Mexican Grocery

RestaUrant StDfe

Café

Wine Shop

financial District

Coffee Shop

Hotel

Gym Wind Shop SteakhoUSe• Bar

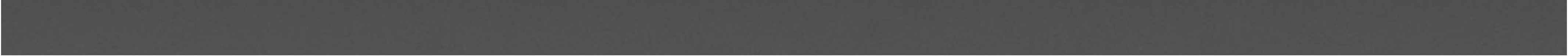
lalian Restaurant

Pizza Place•

Park

/ Fitness CRntRr

31 NDho



liaIian French

Cocktail Bar Gift Shop BoOkstOrR Grocery Store

Mexican

Rr me i i re nt

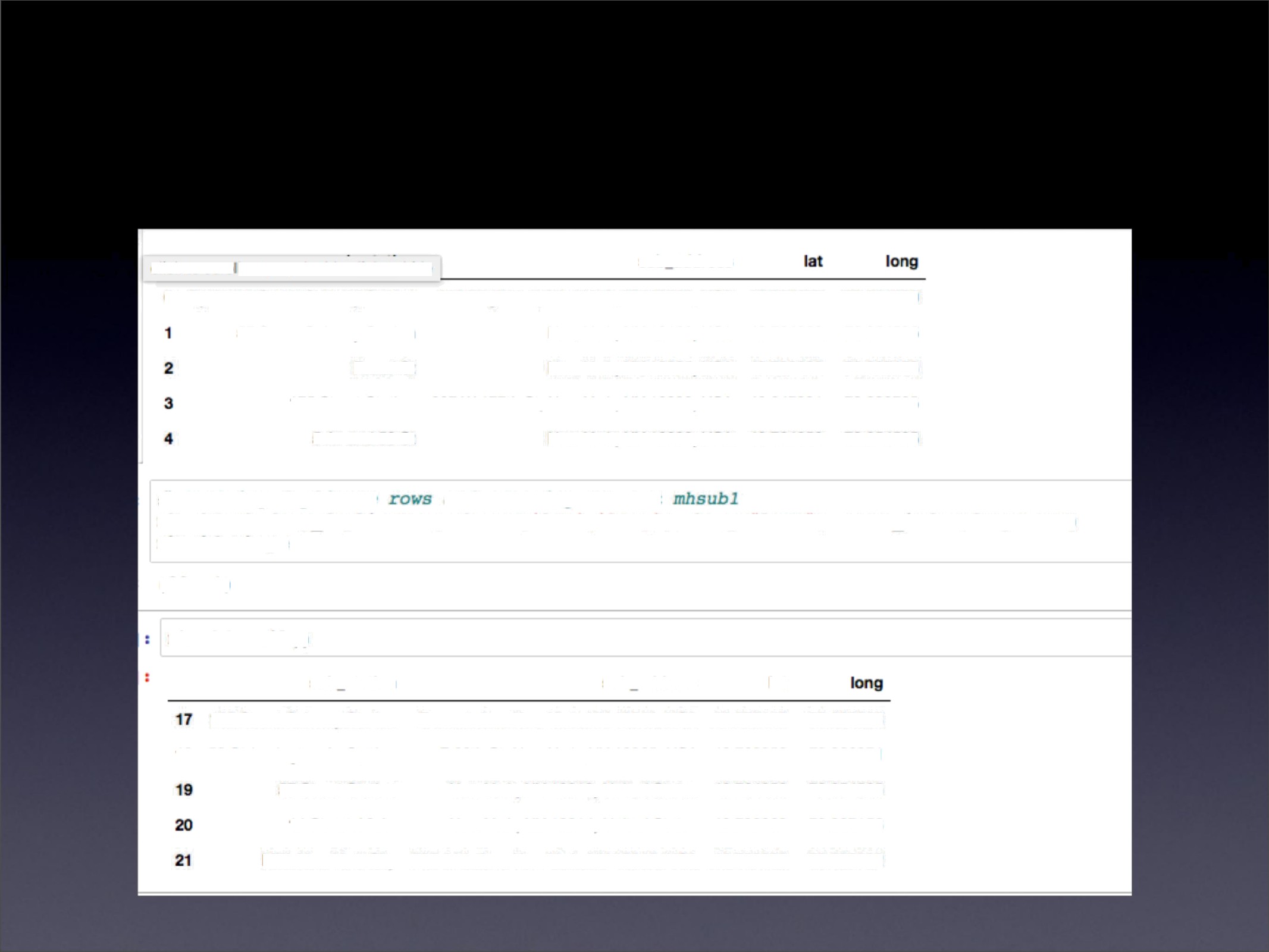
Hotel

Sushi

RosToi iro nT

Coffee Shop

## Manhattan subway stations geodata



click to scro I output; double click to hide sub address

0 Dyckman Street Subway SatDn 170 Nmb Ave, New @rk, NY 10034, USA 40,861&7 -73.624506

57 Street Subway Station New 'Ybrk, NY 10106, USA 40.7642D -73.954525

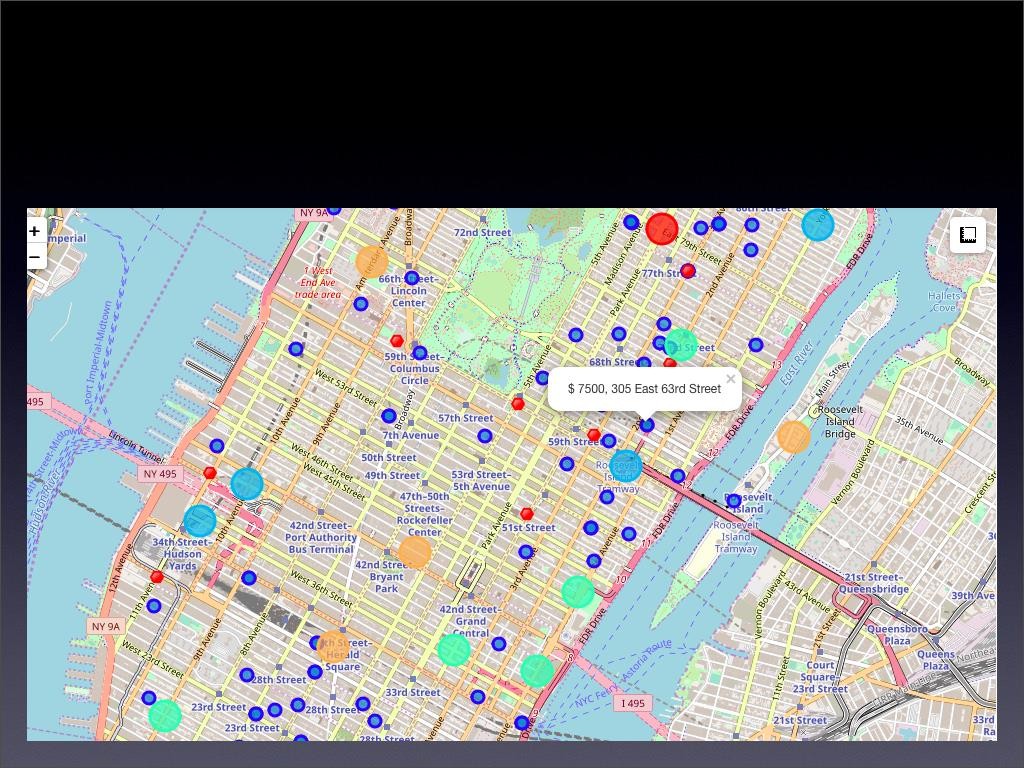
|  |  |  |  |
| --- | --- | --- | --- |
| &oad St | New \brk, NY 10005, USA | 40.730862 | -73.967156 |
| 175 Sbeet Station | 807 W 177th St. New York, NY 100Z3, USA | 40.847991 | -73.039785 |
| 5 Av and 53 St | Naw Ynrk, NY 10022, USA | 40.764250 | -73.954525 |

/ *removing dup1?ca ie and* cceaCJnp new ret

mbsubl-mR.drop\_duplicates(subset-('lat','long'], keep•“last“).reset\_index(drop•True) (22, 4)

|  |  |  |  |
| --- | --- | --- | --- |
| aub ewsofi | aub address | lat |  |
| 140 Stfeet Subway Station | Bennett Ave, New Yolk, NY 1Q\40, USA | 40.858113 | -73.932983 |
| 16 59 St-Lexington Av Sstion | E 60G St, New York, NY 10065, USA | 40.762259 | -73.966271 |
| 57 Streat Station | New York, NY 10019. Omited SBtes | 40.764250 | -73.954525 |
| t4 Street / 8 Av | Naw York, NY 10014, United Staas | 40.730862 | -73.987156 |
| MTA N«>/ Y«/¥ CII¿ | 525 11\A Aw, N¥I< Y+¥, NY 1QI18, USA | 40.759800 | -73.996282 |

##### Apts for rent (blue) and subway stations (red)

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.

## I will walk to work

#### Walk from home to work is less than 1 km!

Venus in Cluster 2 near future home

kX = 2

manhattan\_mergRd.loc[manhattan\_mRzged['CluStRl Labels } kK, manhattan\_mRrged.columns[[1] + list(range(5, manhattan\_n

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1st Most | 2nd Most | 3rd Most | Cth Most | Sth Most | bth Mast | 7th Most | 8th Most | 9th Most | 10th Most |
| MelghbDFhood | Common | Common | Common | Common | Common | **Common** | **ComI¥lDn** | **Common** | **Common** | **ComI¥lDn** |
|  | Ve.nul2 | Venue | Ve.nul2 | Venue | Venue | Venue | Venue | Ve.nue | Venue | Venue |

1. Marble Hill CoWee Shop

Discount

StDre

Yoga Studio Steakhouse Supplement Te•nnis

Shoe Store

Chafood Resaurant

1. Chinatown

Chinese Restaurant

Cocktail Bar

Dim Sum American Vietnamese Salon / Restaurant Restaurant Restaurant Barbershop

NoodlR HouSR

Bakery

Bubble Tea IcR Cream

Shop Shop

Bank

6 Central Harlem

African SeafDDd French American Cosmetics Chinese Restaurant Resmurant Restaurant Restaurant Shop Re.StBMfdnt

EvRnt SpacR Liquor Store BeRr Bar

Fitness Ce.nte.r

/DfkviIIe CoW l•9 Shop

Gym

Bar

Italian Sushi

Restaurant Restaurani

Pizza Place

Mexican Restaurant

Deli / Bodega

JapanesR Restaurant

Pub

14 ClintDn

Theater

I\aIian

Ressurant

Coffee Shop

American Restaurant

Gym / Fitness Center

Hotel Wine Shop

soa

Gym Indie TheatRf

Dho

Clothing

StDre

Boutique

Women's

Store

ShDe Store Men’s Store

Furniture / Italian Medite•rranean HDMI Store Restaurant Restaurani

Art GallRry Design StudiD

Morningside

Heights

CoWee Shop

/tMRf1CaN Resmurant

Park Bookstore Pizza Place

Sandwich

Place

Burger Joint Cake

Oeli / Bo deea

Tennis Court

34 Sutton Place

Gym */* Fitness Center

Italian Furniture / Indian Restaurant Home Store Restaurant

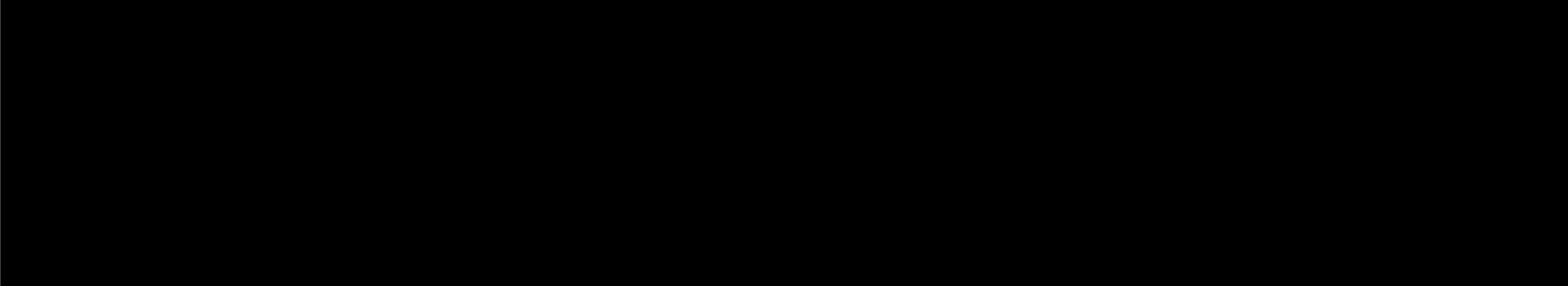
Dessert Shop

American ReS@gOpt

Bake.ry Juice Bar Boutique.

Sushi Restaurant

39 Hudson Yards CoWee Shop



Shop Stadium

Imlian

Resmurant

Hotel Theater

Ameican Restaurant

Café

Fitness CRntRf

Thai Restaurant

Resaurant

# 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.

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.