Choose the best place in manhattan

for Italian restaurant

1) Introduction/Business Problem

a)background

An investor wants to open a luxurious Italian restaurant, meaning that is a restaurant that provides high-quality services For him , he wants to invest in New york city, and Manhaten borough.He expressed his desire to use data science to choose the right place for this purpose

The question is :
what is the right neighborhood that has Advantages for receive this project ?

Starting from this question is definitely a difficult matter without trying to divide it and know what these characteristics are and how we can define them , knowning that the data will only be from the Internet.

we will try to divide this complex question into several simple questions , through wih we can at least lay the foundation for this construction :

1)what are the neighborhoods located in a borough of Manhatten?
konwing all neighborhood is important for studying the area and this is an important question among the questions , because one of these elements will be chosen, and therefore we my say tha the target !!!!

2)what is the socio-economic classification of these neighborhood?
this is a question that refers directly to an important aspect, which is the individual income and standard of living for individuals , because we are talking about a luxury restaurant and this is important.

3)how is the competitions betwen restaurant ?
based on the presence of the same classfication of restaurants and the same category and the restaurant in general

b)why is research in this way important for the project ?

there are many projects that are conducted by study and does not rely on data science in their studies , perhaps this due to the reliance on a small number of the data and its generalisation, wich results in a lot of errors. of course it cannot be said that it is not useful but it is not sufficient today the economy has devloped and taken a turn and relianceon data science will be useful in
determining the target group and all the basic elements affect profit.A restaurant project is also project that needs studing and adopting a large amount of data analyzing and woking on it that help thosse who approach the real picture and of course , additional profit and success in business

2)A description of the data and how it will be used to solve the problem

a)choice of data source :

As for the data ,it must be related to the questions raised in the introduction and thus it willl be relied upon in searshing for the necessary data features: variables and e target.

1)what are the neighborhoods located in a borough of Manhatten?

to answer this question we will search on a ibm storage , it contains New York City data wich will refer us to neighborhoods within Manhattan , and location

2)what is the socio-economic classification of these neighborhood?

to answer this question we will search in kaggls or scraping from here for knowing income and population in new york in general , manhattan in spetial

3) how is the competitions betwen restaurant ?

here we will be fetching data from forsquare api and chose all

maybe other features will hep me to more indurstand the good places

The choice of data in order to srudy places and how many potential customers are and what are the most imortant features of the places that have it ,perhaps real questions to answer what is the right one , so I must know everyting that needs to be taken into account and therefore in the midst of work and analysis will apear other feature that i should focus on , thiw was only preliminary choice, not a final one

so the target is : "Neighberhood"

variable choce all after feature ingenieuring i will dertermine what are the ncessary variables that have an impact on the neighborhood

3)Methodology

We will rely on the Forsquire platform to collect data about all restaurants in Manhattan, then we will know the features of each neighborhood and through it we will try to classify by kmeans algorith , and then repeat the process in favor of neighborhoods over another. We will repeat the process, but this time , only choose Italian restaurants and reclassify them to see which neighborhood is known for Italian restaurants, so my choice falls on one neighborhood. Then I will try to inventory the number of restaurants to get a clear picture, and to closely discover the competitors.

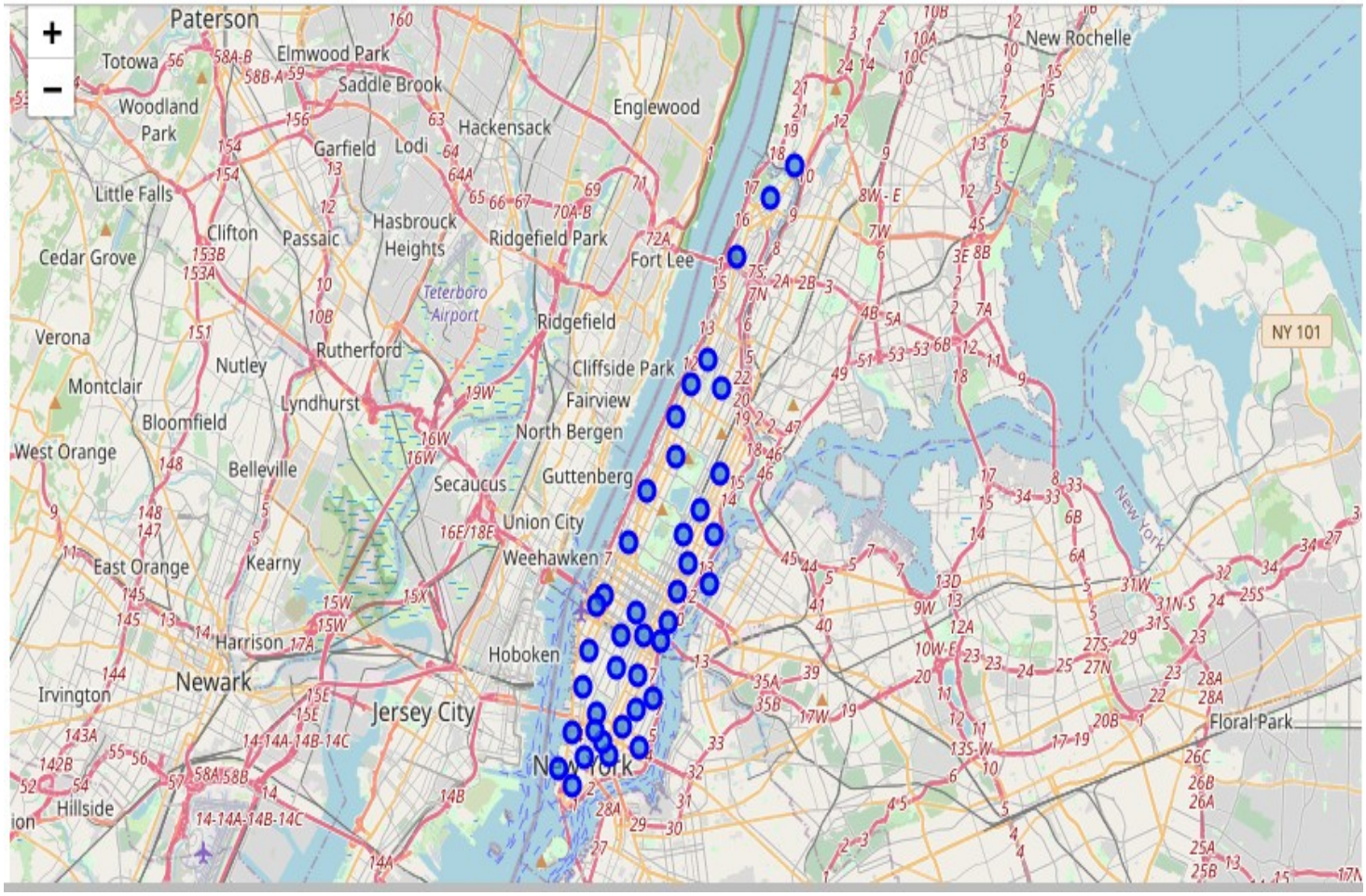
4). Exploratory Data Analysis

a) Discus detail about neighborhoods

total of neighborhoods is 40 :

(Marble Hill,Chinatown, Washington Heights,Inwood,Hamilton Heights,Manhattanville
Central Harlem ,East Harlem ,Upper East Side ,Yorkville, Lenox Hill Roosevelt Island ,
Upper West Side , Lincoln Square, Clinton , Midtown, Murray Hill
Chelsea, Greenwich Village,East Village , Lower East Side ,Tribeca,Little Italy
Soho, West Village ,Manhattan Valley,Morningside Heights, Gramercy,Battery Park
City ,Financial District, Carnegie Hill,Noho,Civic Center,Midtown South,Sutton
Place,Turtle Bay,Tudor City,Stuyvesant Town,Flatiron,Hudson Yards).

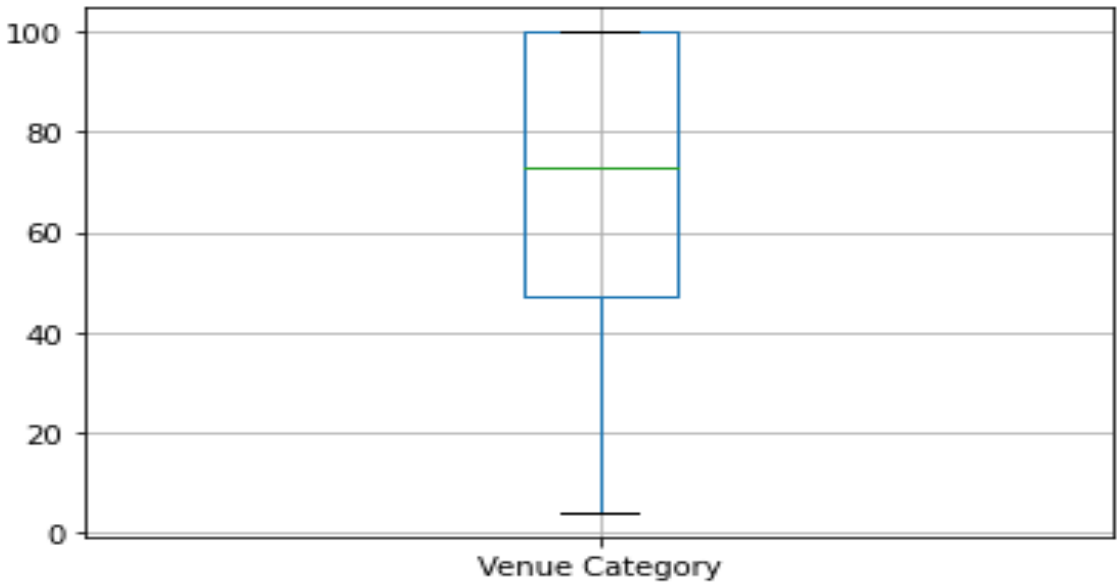
So all these neighborhoods have the potential to be the place for the restaurant
and the search will begin from them.



figue1 ; map of manhattan .

b) statiscs descriptive of Restaurant

we’re trying to find out the number of restaurants in Manhattan and determine wich neighborhoods have the largest number of restaurant and also the fewest neighborhoods that have a small number of restayrants



figue2 : distribution of the restaurants in all Neighborhoods

statistic descriptive of number of restaurants :
the largest number in one neighberhood is 100 restaurants.
The smallest number in one neighberhood is 4 restaurants.
The mean is 70 restaurants in one neighberhood
_ list the number of restaurants in each neighberhoods
the neighberhood has the largest number of restaurants is :

```
:  count     40.00000
   mean      69.55000
   std       28.11875
   min        4.00000
   25%       47.00000
   50%       73.00000
   75%      100.00000
   max      100.00000
```

neighberhoods	Number of restraurants
East Village	100
Financial District	100
Little Italy	100
West Village	100
Midtown	100
Greenwich Village	100
Midtown South	100
Flatiron	100
Lenox Hill	100
Clinton	100
Noho	100
Chinatown	100
Soho	100

the neighberhood has the fewest number of restaurants is :

neighberhoods	Number of restaurants
Battery Park City	29
Marble Hill	15
Roosevelt Island	13
Stuyvesant Town	4

c)Analyse of category every restaurant :

Top 5 neighberhood and frenquency

this is a sample of the categorical restaurant in each neighborhood . Since we are looking fo the right neighborhood for the restaurant , each neighberhood must know what kind of retaurants required in it , in order to searsh for the neighborhood in which we can invest it the lanch of the

```
----Washington Heights----
                venue  freq
0       Deli / Bodega  0.10
1         Pizza Place  0.10
2  Chinese Restaurant  0.09
3             Bakery  0.07
4               Café  0.07


----West Village----
                      venue  freq
0       Italian Restaurant  0.16
1      American Restaurant  0.11
2  New American Restaurant  0.06
3        French Restaurant  0.05
4                Restaurant  0.04


----Yorkville----
                venue  freq
0  Italian Restaurant  0.14
1         Pizza Place  0.11
2       Deli / Bodega  0.08
3    Sushi Restaurant  0.07
4  Chinese Restaurant  0.05
```

```
----Battery Park City----
            venue  freq
0     Pizza Place  0.14
1  Sandwich Place  0.10
2    Burger Joint  0.07
3      Food Court  0.07
4       BBQ Joint  0.07


----Carnegie Hill----
                venue  freq
0               Café  0.12
1         Pizza Place  0.10
2  Italian Restaurant  0.09
3    Sushi Restaurant  0.09
4             Bakery  0.07


----Central Harlem----
                venue  freq
0  Fried Chicken Joint  0.09
1       Deli / Bodega  0.09
2   African Restaurant  0.07
3   Chinese Restaurant  0.07
4           Restaurant  0.07
```

For exemple in Yorkville the most popular restaurant are Italian .the least common restaurants out
of the five are Chinese restaurants .

5) Clustring modele

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of explatory analysis, and a common technique for statistic data analysis, used in many fields.

Cluster analysis itself is not one specific algorithme, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their understanding of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distance between cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including parameters such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It is often necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

a) K-means

***k*-means clustering** is a method of vector quantization, originally from signal processing, that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

b) application of K-means

we apply k-means on our problem to divide the neighberhoods into groups, and like that we will know the relation of the Italian restaurants with each group, then we will choose the

neighborhoods, then we will redo on a new group. but we will choose the group by the highest frequency of Italian restaurants in each neighborhoods .

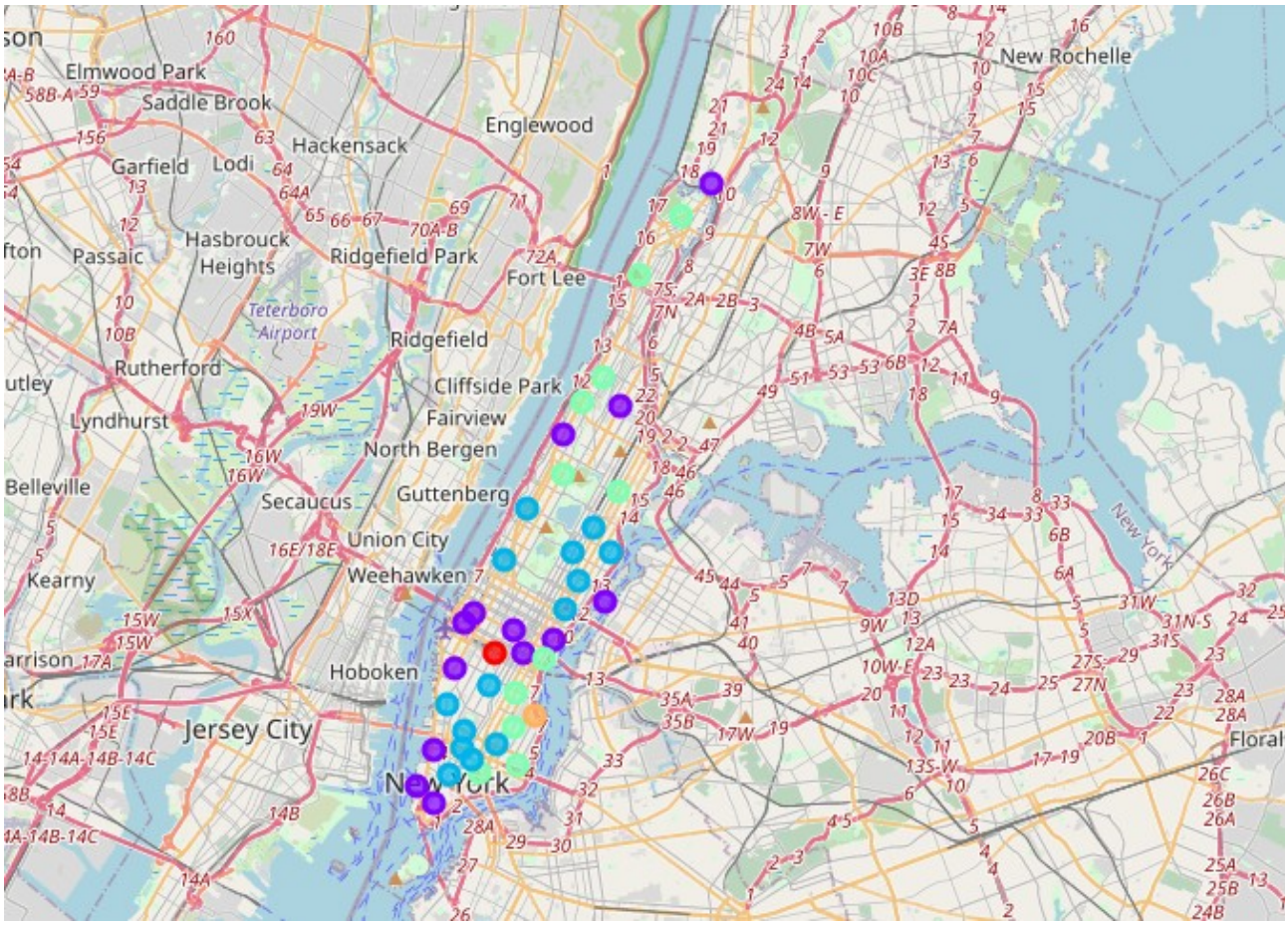
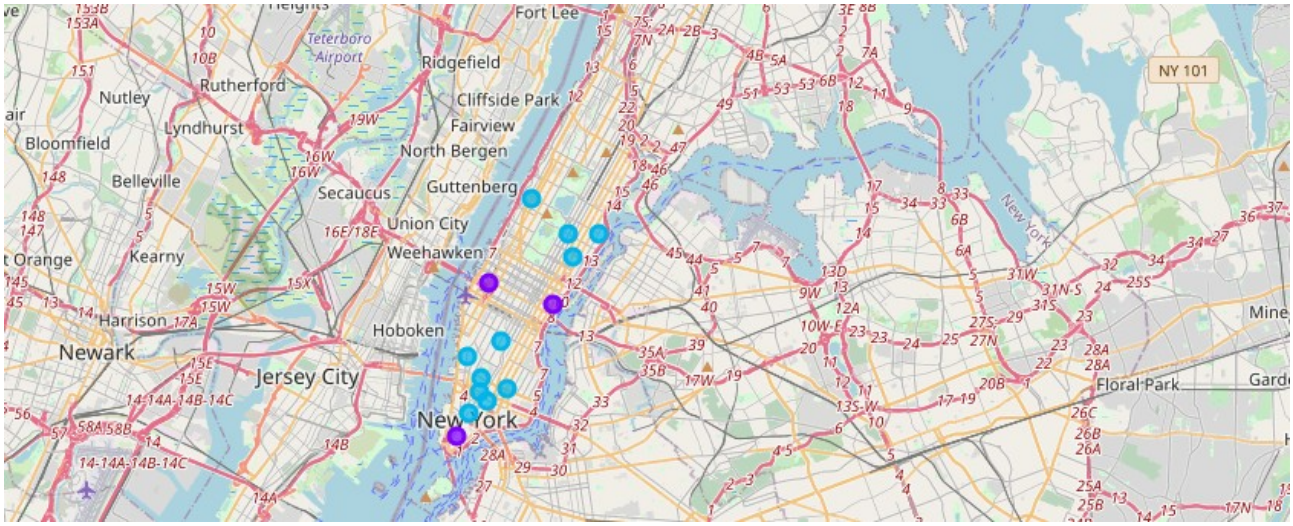


Figure3 : cluster of neighberhoods .

now we will qualify all the neighborhoods that have 1st Most Common Venue is :



Restaurant italian .

c) explory data analyse and recieve new data

now That we have chosen of the neighborhoods in two cluster (2,3) ,we will reiew each neighborhood and group of restaurants .

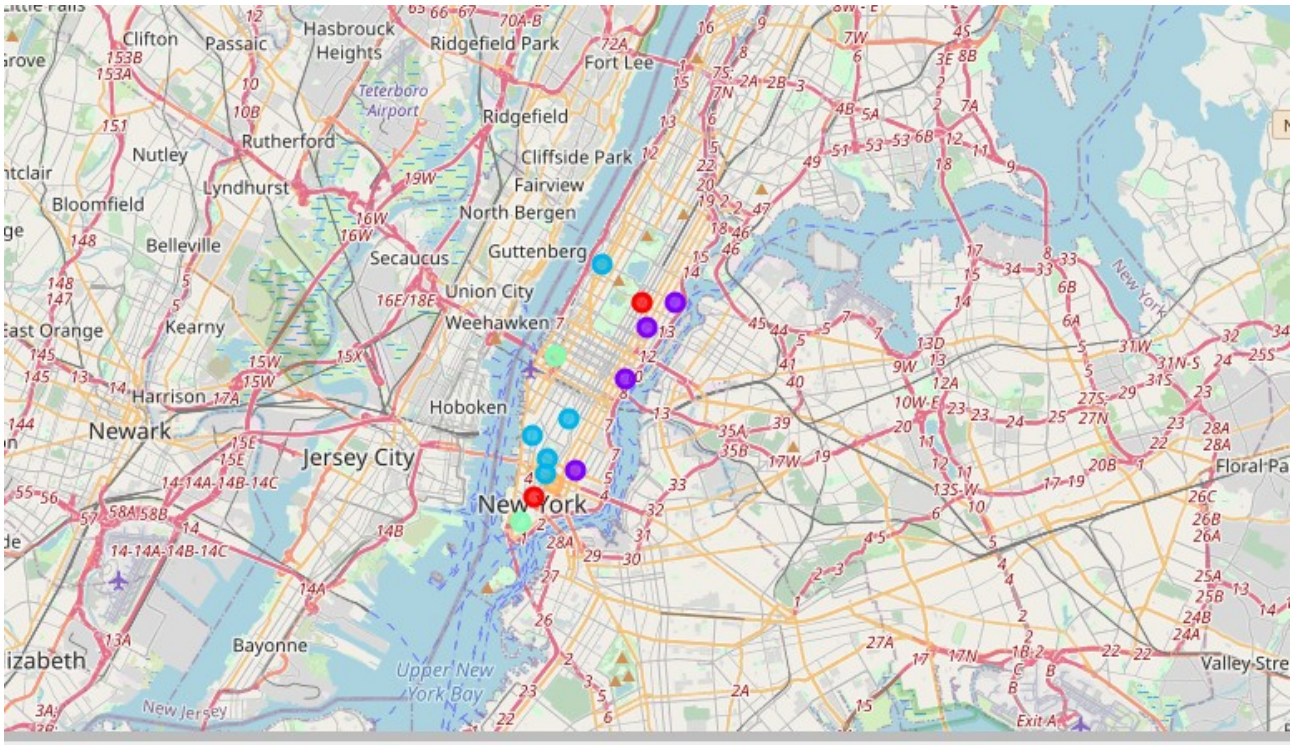
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Upper East Side	Italian Restaurant	American Restaurant	Pizza Place	French Restaurant
Yorkville	Italian Restaurant	Pizza Place	Deli / Bodega	Sushi Restaurant
Lenox Hill	Italian Restaurant	Sushi Restaurant	Pizza Place	Deli / Bodega
Upper West Side	Italian Restaurant	Mediterranean Restaurant	Bakery	Restaurant
Clinton	Italian Restaurant	Deli / Bodega	Restaurant	American Restaurant
Greenwich Village	Italian Restaurant	Sushi Restaurant	Japanese Restaurant	American Restaurant
Little Italy	Italian Restaurant	Chinese Restaurant	Bakery	Café
Soho	Italian Restaurant	Café	French Restaurant	American Restaurant
West Village	Italian Restaurant	American Restaurant	New American Restaurant	French Restaurant
Financial District	Italian Restaurant	Sandwich Place	Pizza Place	American Restaurant
Noho	Italian Restaurant	Japanese Restaurant	Pizza Place	Mexican Restaurant
Civic Center	Italian Restaurant	American Restaurant	French Restaurant	Sandwich Place
Turtle Bay	Italian Restaurant	Café	Deli / Bodega	Pizza Place
Flatiron	Italian Restaurant	American Restaurant	Café	Japanese Restaurant

calcule frequency in every neighborhoos , but wu put just a few exemple :

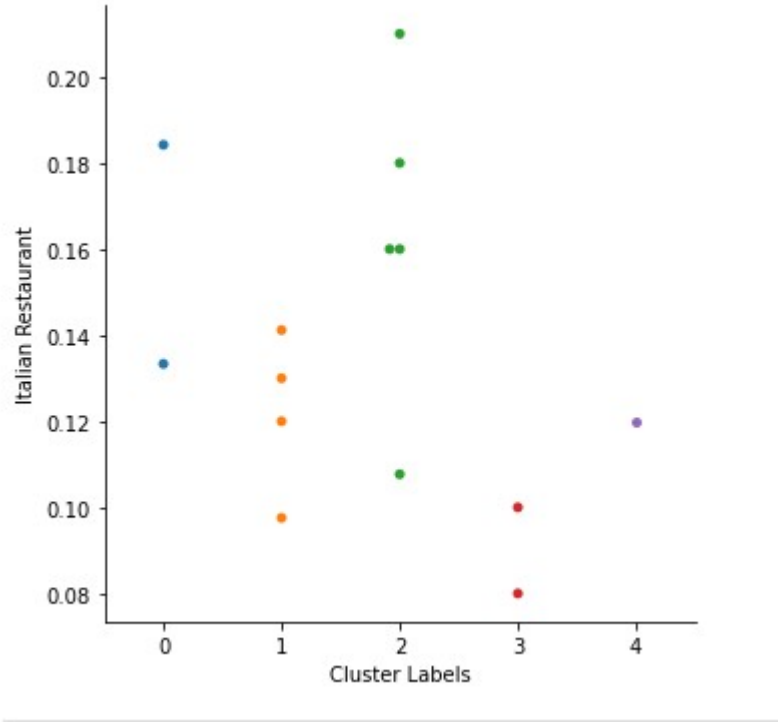
```
----Flatiron----
                       venue  freq
0        Italian Restaurant  0.16
1       American Restaurant  0.07
2                     Café  0.06
3  Mediterranean Restaurant  0.05
4       Japanese Restaurant  0.05


----Greenwich Village----
                  venue  freq
0   Italian Restaurant  0.21
1     Sushi Restaurant  0.06
2  American Restaurant  0.05
3  Japanese Restaurant  0.05
4                 Café  0.05
```

d) Reapeat kmeans

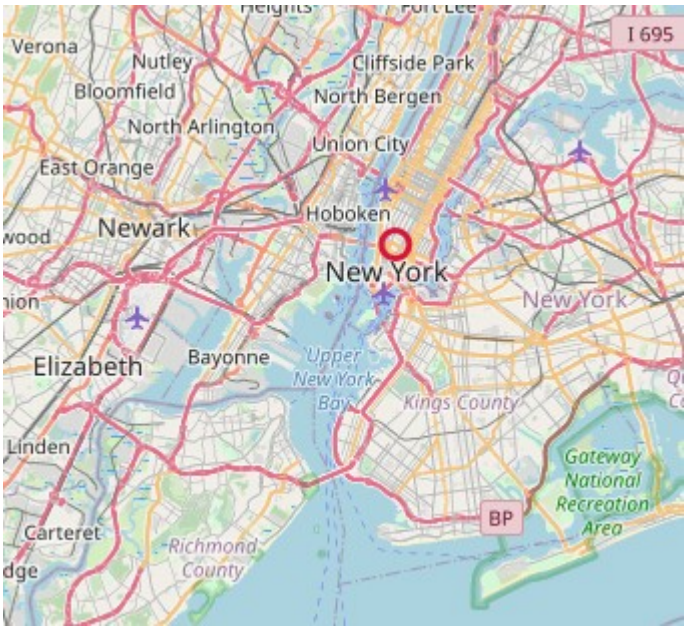
As we said earlier, after choosing the neighborhoods, we will return to dividing agian , except for groups and we will choose that we deem the most appropriate , of course this selection will be based on most Italian restaurants .



now that we have divided the neighborhoods again we will choose the group that italian restaurants frequents :



i choce the Cluster label = 2 and the hight frequency .




```
----Greenwich Village----
                  venue  freq
0   Italian Restaurant  0.21
1     Sushi Restaurant  0.06
2  American Restaurant  0.05
3  Japanese Restaurant  0.05
4                Café  0.05
```

6) conclusion :

In this study we tried to rely only on data from Forsquire and I think that this is not enough of course, to find out the best place in which the Italian restaurant could be, why? This is a good question, I must answer it, because the location of the restaurant can be expected, but a person's taste in choosing the restaurant he wants to go to is not necessarily a neighborhood known for Italian restaurants. But at least the risk of investment will be lower in that area, given that it contains restaurants of the same type, and therefore Italian restaurant visitors will be curious, knowing the new restaurant.