Report Capstone Project

BY

MOLINA Jérémy

2019

Table des matières

Table des matières

I.   Introduction

1.   Description of the problem

We can assume that we are contacted by a French investor wishing to open a business. In particular, a restaurant in New york more precisely in manhattan, which aims to offer 100% food for 100% "made in france", since the country's current policy allows a certain advantage in terms of taxes and taxes for investors favouring "made in france" France."

However, this investor, not being from new york, therefore does not know or placed his establishment. Moreover being aware that the economic competition is tough in this city, he decides to take all these chances on his side and ask us what is the best location for his restaurant in this city.

Therefore, we will aim to find the best location for this restaurant in manhattan.

2.   Background Discussion

New york is undoubtedly one of the most touristic cities in the world, thanks to its emblematic monuments, such as the Empire state bulding or statue of liberty. Consequently its economy is based essentially on the tourist influx, my client would like therefore that one places his restaurant also in an area with high tourist potential.

3.   Audience

find a good location for the restaurant would allow on the one hand to have benefits for the investor but also to allow the potential customer (tourists) to easily find food made in France.

4.   Data

- For this project data available on the Foursquare's API will be used. Because The data used includes information about different venues and their neighborhoods.

- One of the technique which will be used is KMeans to group neighborhoods with similar objects.

- I will use a json file for the map of New york and Manhattan.

- finally i use Foursquare to explore the neighbourhoods and determine the most popular venues per neighbourhoods.

II.   Methodology

1.   Data Exploration

a.   *Geopy and Map of New york*

So after transform the panda into dataframe, I used geopy to get the latitude and longitude of new york, thereby allowing me to get geospatial data. Thanks to that, I was able to generate a map of New York showing all Borough. This allowed me to have an overview before focusing on Manhattan. This overview in reality allowed me to locate Manhattan within New York, and to understand the potential strategic issues of Manhattan, such as its economic

and tourist positioning.
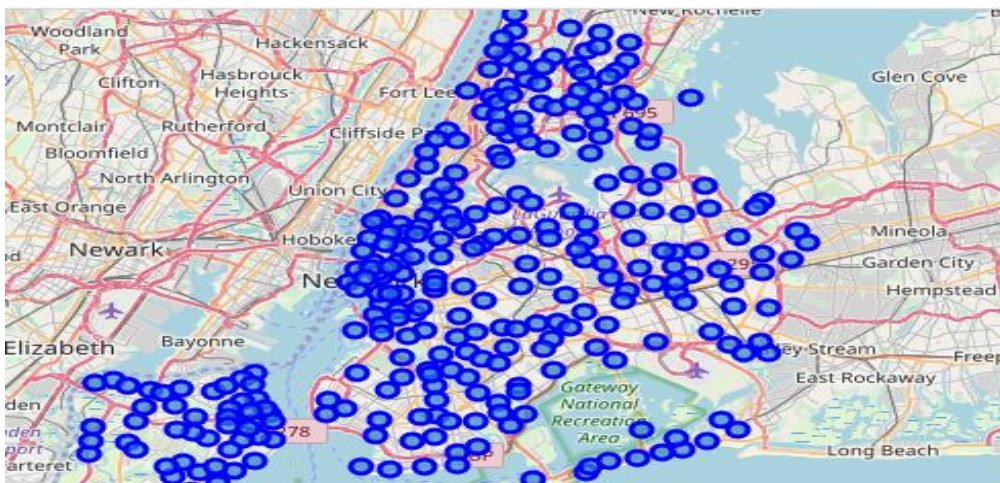
After this step, I decided to get geospatial data from Manhattan, allowing me to directly view all the neighborhoods of Manhattan, again for a better understanding of the study I generate a map so that it is more readable and easier located the different places of Manhattan.
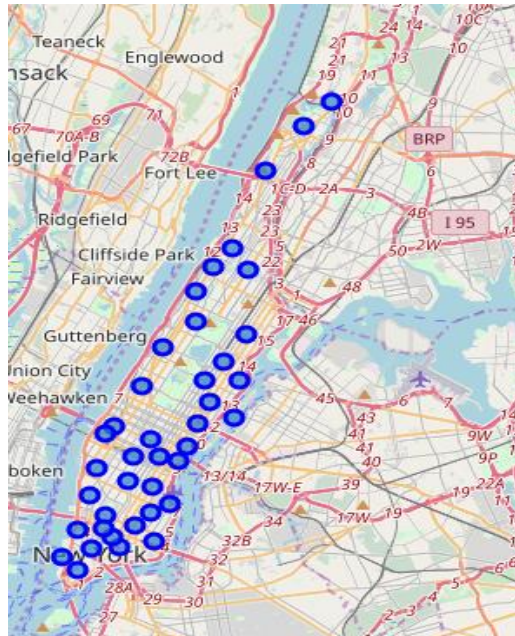
c. *Explore and analyse neighborhood*

Now that we have done all these visualization tasks, we can better understand the problem as well as the client's request by using the visualization of the two maps, which allow us a better geographical understanding of a city that is not ours. -familière. We are going to attack one of the most important tasks that is to explore and analyze neigborhood.

The screenshot below is an example of the table that allows statistical representation of the presence of types of businesses within each Manhattan neighborhood. This allows on the one hand to better understand the distribution of each type of business within Manhattan and their potential economic importance. But also to attribute via hypotheses the type of client and their demands for each neighborhood. In fact, the greater the statistical representation of a type of trade, the more we can assume that the population around them is interested in the products and services offered by these high-rate statistical businesses. This distribution of the businesses explicitly shown by these rates is therefore very important in order to better

understand the expectations of potential consumers and customers of the future restaurant of our investor.

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant	Austrian Restaurant	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Bakery	Ban
0	Battery Park City	0.000000	0.00	0.00	0.00	0.010000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.01	0.01	0.00	0.00	0.000000	0.020000	0.00	0.000000	0.010000	0.00000
1	Carnegie Hill	0.000000	0.00	0.00	0.00	0.010000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.01	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.010000	0.020000	0.00000
2	Central Harlem	0.000000	0.00	0.00	0.06383	0.042553	0.00	0.00	0.00	0.000000	0.000000	0.042553	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.021277	0.00	0.021277	0.021277	0.00000
3	Chelsea	0.000000	0.00	0.00	0.00000	0.030000	0.00	0.01	0.00	0.000000	0.000000	0.030000	0.00	0.000000	0.010000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.040000	0.00000
4	Chinatown	0.000000	0.00	0.00	0.00000	0.040000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.020000	0.00	0.00	0.00	0.01	0.000000	0.000000	0.00	0.000000	0.020000	0.00000
5	Civic Center	0.000000	0.00	0.00	0.00000	0.030000	0.00	0.01	0.00	0.000000	0.000000	0.020000	0.00	0.000000	0.010000	0.00	0.00	0.01	0.00	0.000000	0.000000	0.01	0.010000	0.030000	0.00000
6	Clinton	0.000000	0.00	0.00	0.00000	0.040000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.010000	0.00000
7	East Harlem	0.000000	0.00	0.00	0.00000	0.000000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.090909	0.02272
8	East Village	0.000000	0.00	0.00	0.00000	0.020000	0.00	0.01	0.00	0.020000	0.010000	0.010000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.020000	0.010000	0.00000
9	Financial District	0.010000	0.00	0.00	0.00000	0.030000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.010000	0.010000	0.00000
10	Flatiron	0.000000	0.00	0.00	0.00000	0.040000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.020000	0.00000
11	Gramercy	0.000000	0.00	0.00	0.00000	0.040000	0.00	0.00	0.01	0.000000	0.000000	0.010000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.040000	0.010000	0.00000
12	Greenwich Village	0.000000	0.00	0.00	0.00000	0.020000	0.00	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.010000	0.010000	0.00000
13	Hamilton Heights	0.000000	0.00	0.00	0.00000	0.000000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.032258	0.01612
14	Hudson Yards	0.000000	0.00	0.00	0.00000	0.066667	0.00	0.00	0.00	0.000000	0.000000	0.013333	0.00	0.000000	0.013333	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.00000
15	Inwood	0.000000	0.00	0.00	0.00000	0.034483	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.051724	0.00000
16	Lenox Hill	0.000000	0.00	0.01	0.00000	0.000000	0.00	0.00	0.00	0.000000	0.000000	0.020000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.010000	0.00	0.010000	0.020000	0.00000
17	Lincoln Square	0.000000	0.00	0.00	0.00000	0.020000	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.01	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.020000	0.00000
18	Little Italy	0.000000	0.00	0.00	0.00000	0.010000	0.01	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.000000	0.000000	0.00	0.00	0.01	0.00	0.000000	0.000000	0.00	0.000000	0.050000	0.00000
19	Lower East Side	0.000000	0.00	0.00	0.00000	0.000000	0.00	0.00	0.00	0.000000	0.016667	0.033333	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.016667	0.000000	0.00	0.016667	0.033333	0.00000
20	Manhattan Valley	0.000000	0.00	0.00	0.00000	0.016949	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.016949	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.016949	0.00000
21	Manhattanville	0.000000	0.00	0.00	0.00000	0.022727	0.00	0.00	0.00	0.000000	0.000000	0.022727	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.022727	0.00	0.000000	0.000000	0.02272
22	Marble Hill	0.000000	0.00	0.00	0.00000	0.038462	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.038462	0.03846
23	Midtown	0.000000	0.00	0.00	0.00000	0.040000	0.00	0.00	0.00	0.000000	0.000000	0.010000	0.00	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.030000	0.00000

After having generated this statistical table showing all the shops for each neighborhood, it was necessary to obtain data that were both more precise, but also simpler to process, which would make it possible to add the attractiveness of a type of trade to the neighborhood. within a neighborhood. That's why I chose to get the top 5 comings in the neigborhoods, it seems to me a good compromise between readability of the data and representativeness of the reality of the data. It also allowed me to identify the frequency with which the population goes into a type of restaurant within a venue, this frequency being the representation of the attractiveness force. You can see an example of this data in the screenshot below.

```
----East Village----
                 venue  freq
0                  Bar   0.06
1             Wine Bar   0.05
2  Mexican Restaurant   0.04
3  Chinese Restaurant   0.04
4       Ice Cream Shop   0.04


----Financial District----
               venue  freq
0        Coffee Shop   0.07
1         Steakhouse   0.04
2          Wine Shop   0.04
3                Gym   0.04
4        Pizza Place   0.03
```

d. Explore data with K-mean and cluster analyse

K-mean clustering helped identify the best cluster to open our client's restaurant. We have 5 cluster showing for each neighborhood, shops or tourist places with the most client.

For the first cluster :

:[44]:	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
11	Roosevelt Island	Park	Sandwich Place	Coffee Shop	Deli / Bodega	Pizza Place	Greek Restaurant	Dry Cleaner	Bus Stop	Baseball Field	Liquor Store
26	Morningside Heights	Park	Coffee Shop	Bookstore	American Restaurant	Food Truck	New American Restaurant	Burger Joint	Deli / Bodega	Tennis Court	Outdoor Sculpture
28	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Gym	Wine Shop	Clothing Store	Italian Restaurant	Women's Store	BBQ Joint

we realize that there is no French restaurant which could be a good news, but we can notice that it is mainly fast food that have a great success more than restaurants.

For the second cluster :

	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
13	Lincoln Square	Theater	Gym / Fitness Center	Café	Plaza	Italian Restaurant	Concert Hall	French Restaurant	Performing Arts Venue	Indie Movie Theater	Park
14	Clinton	Theater	Italian Restaurant	Gym / Fitness Center	American Restaurant	Hotel	Coffee Shop	Sandwich Place	Wine Shop	Spa	New American Restaurant

However, for this one, on the contrary, there are too many restaurants, two Italians, a Frenchman and American restaurants. However one can see a point of interest, indeed, we can see that the French restaurant is in 7th position. We can make two assumptions, either

the restaurant is not good enough to bring in more customers and in this case, it will be easier for our client to settle there, since there are still potential customers, except that the restaurant is not good enough to attract more, on the contrary the residents around this avenue are not interested in this type of restaurant.

For the third cluster :

	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
0	Marble Hill	Coffee Shop	Discount Store	Sandwich Place	Yoga Studio	Tennis Stadium	Supplement Shop	Steakhouse	Spa	Seafood Restaurant	Clothing Store
1	Chinatown	Chinese Restaurant	American Restaurant	Cocktail Bar	Salon / Barbershop	Spa	Bubble Tea Shop	Dumpling Restaurant	Vietnamese Restaurant	Ice Cream Shop	Hotpot Restaurant
6	Central Harlem	African Restaurant	Public Art	Art Gallery	Seafood Restaurant	Chinese Restaurant	Gym / Fitness Center	French Restaurant	American Restaurant	Cosmetics Shop	Liquor Store
8	Upper East Side	Italian Restaurant	Exhibit	Coffee Shop	Juice Bar	Bakery	Art Gallery	Gym / Fitness Center	French Restaurant	Spa	Hotel
9	Yorkville	Italian Restaurant	Coffee Shop	Bar	Gym	Pizza Place	Sushi Restaurant	Deli / Bodega	Wine Shop	Mexican Restaurant	Diner
10	Lenox Hill	Coffee Shop	Italian Restaurant	Sushi Restaurant	Pizza Place	Gym	Café	Cosmetics Shop	Burger Joint	Sporting Goods Shop	Gym / Fitness Center
12	Upper West Side	Italian Restaurant	Wine Bar	Bar	Coffee Shop	Bakery	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Yoga Studio	Pub
15	Midtown	Coffee Shop	Hotel	Cocktail Bar	American Restaurant	Clothing Store	Theater	Steakhouse	Sporting Goods Shop	Japanese Restaurant	Spa
16	Murray Hill	Coffee Shop	Hotel	Sandwich Place	Japanese Restaurant	Gym	Italian Restaurant	French Restaurant	Cocktail Bar	Bagel Shop	Bar
17	Chelsea	Coffee Shop	Ice Cream Shop	Italian Restaurant	Bakery	Nightclub	Theater	Seafood Restaurant	American Restaurant	Hotel	Art Gallery
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	Clothing Store	French Restaurant	Café	Seafood Restaurant	Indian Restaurant	Ice Cream Shop	Dessert Shop	Cocktail Bar
19	East Village	Bar	Wine Bar	Mexican Restaurant	Chinese Restaurant	Ice Cream Shop	Pizza Place	Cocktail Bar	Ramen Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant
20	Lower East Side	Coffee Shop	Pizza Place	Ramen Restaurant	Café	Cocktail Bar	Sandwich Place	Chinese Restaurant	Art Gallery	Bakery	Park
21	Tribeca	Café	Spa	Italian Restaurant	Park	American Restaurant	Boutique	Greek Restaurant	Coffee Shop	Wine Bar	Wine Shop
22	Little Italy	Bakery	Sandwich Place	Salon / Barbershop	Café	Italian Restaurant	Bubble Tea Shop	Hotel	Mediterranean Restaurant	Seafood Restaurant	Clothing Store
23	Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Art Gallery	Italian Restaurant	Sporting Goods Shop	Bakery	Men's Store	Mediterranean Restaurant
24	West Village	Italian Restaurant	New American Restaurant	Cosmetics Shop	Park	Jazz Club	Wine Bar	American Restaurant	Bakery	Coffee Shop	Gastropub
27	Gramercy	Italian Restaurant	Pizza Place	American Restaurant	Bagel Shop	Bar	Cocktail Bar	Hotel	Thai Restaurant	Mexican Restaurant	Thrift / Vintage Store
29	Financial District	Coffee Shop	Wine Shop	Gym	Steakhouse	Cocktail Bar	Bar	Pizza Place	Italian Restaurant	Café	Juice Bar
30	Carnegie Hill	Coffee Shop	Pizza Place	Café	Yoga Studio	Bookstore	Wine Shop	Cosmetics Shop	French Restaurant	Bar	Japanese Restaurant
31	Noho	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Bookstore	Rock Club	Mexican Restaurant	Sushi Restaurant	Boutique
32	Civic Center	Italian Restaurant	Gym / Fitness Center	Sandwich Place	French Restaurant	Coffee Shop	Yoga Studio	Spa	Park	Hotel	American Restaurant
33	Midtown South	Korean Restaurant	Hotel	Hotel Bar	Japanese Restaurant	Coffee Shop	Cosmetics Shop	Cocktail Bar	American Restaurant	Gym / Fitness Center	Yoga Studio

You can see that there are four French restaurants in all, all placed between 7th and 8th position. Thus even if we can see that the populations in these Neighborhoods have the area interested in French food, we can still assume that there are no positions in favour of opening our client's restaurant, simply because there is too much competition.

Cluster four :

	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
37	Stuyvesant Town	Bar	Park	Playground	Pet Service	Farmers Market	Baseball Field	Fountain	Harbor / Marina	Cocktail Bar	Coffee Shop

We can see that this is not conducive to the creation of a restaurant simply because there are no customers. We can see that there are no restaurants represented in the Cluste

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
2	Washington Heights	Café	Mobile Phone Shop	Bakery	Deli / Bodega	Spanish Restaurant	Latin American Restaurant	New American Restaurant	Sandwich Place	Tapas Restaurant	Mexican Restaurant
3	Inwood	Mexican Restaurant	Café	Lounge	Bakery	Pizza Place	Park	Frozen Yogurt Shop	Chinese Restaurant	Deli / Bodega	American Restaurant
4	Hamilton Heights	Deli / Bodega	Café	Mexican Restaurant	Pizza Place	Chinese Restaurant	Coffee Shop	Sushi Restaurant	Caribbean Restaurant	School	Bakery
5	Manhattanville	Italian Restaurant	Mexican Restaurant	Deli / Bodega	Park	Coffee Shop	Seafood Restaurant	Beer Garden	Bike Trail	Lounge	Sushi Restaurant
7	East Harlem	Mexican Restaurant	Deli / Bodega	Bakery	Latin American Restaurant	Thai Restaurant	Convenience Store	Café	Gas Station	Taco Place	Steakhouse
25	Manhattan Valley	Indian Restaurant	Coffee Shop	Pizza Place	Yoga Studio	Mexican Restaurant	Café	Bar	Thai Restaurant	Deli / Bodega	Szechuan Restaurant
36	Tudor City	Park	Mexican Restaurant	Café	Greek Restaurant	Asian Restaurant	Deli / Bodega	Pizza Place	Hotel	Dog Run	Spa

The cluster 5 is the most favorable, indeed we can see that there are many restaurants offering foreign, and European food: Spanish, Italian, Mexican, Indian, Latin, but no French restaurant, as a result we can see that there is no competition offering my client's food. The fact that there are many foreign restaurants, allows us to see that the population of the area enjoys non-local dishes.

We can therefore offer the customer to create restaurants in Tudor City; Manhattan Valley; East Harlem; Manhattanville; Hamilton Heights; Inwood or Washington Heights

III.   Conclusion

If we did a cluster analysis, we can see that the most favorable locations for the restaurant to be the most profitable are: in Tudor City; Manhattan Valley; East Harlem; Manhattanville; Hamilton Heights; Inwood or Washington Heights.

However, we can be even more precise, indeed, among all these comings, one must be more attractive than the others for that, we can analyze the frequencies of the venues show previously, in order to find the one having the strongest economic attractiveness.

The catches below show the attractiveness of each venues :

```
    ----Washington Heights----              ----Inwood----
                    venue  freq                        venue  freq
0               Café  0.06          0  Mexican Restaurant  0.07
1             Bakery  0.05          1                Café  0.07
2   Mobile Phone Shop  0.05          2             Bakery  0.05
3  Spanish Restaurant  0.04          3         Pizza Place  0.05
4        Deli / Bodega  0.04          4              Lounge  0.05


    ----Hamilton Heights----              ----Tudor City----
                    venue  freq                        venue  freq
0        Deli / Bodega  0.06          0  Mexican Restaurant  0.06
1  Mexican Restaurant  0.06          1                Park  0.06
2         Pizza Place  0.06          2   Greek Restaurant  0.05
3                Café  0.06          3                Café  0.05
4          Coffee Shop  0.05          4         Pizza Place  0.04


    ----Manhattan Valley----                ----East Harlem----
                  venue  freq                          venue  freq
0  Indian Restaurant  0.05          0      Mexican Restaurant  0.14
1        Coffee Shop  0.05          1           Deli / Bodega  0.09
2        Pizza Place  0.05          2                  Bakery  0.09
3               Café  0.03          3  Latin American Restaurant  0.05
4  Mexican Restaurant  0.03          4         Thai Restaurant  0.05


    ----Manhattanville----
                  venue  freq
0  Seafood Restaurant  0.05
1  Italian Restaurant  0.05
2               Park  0.05
3        Deli / Bodega  0.05
4          Coffee Shop  0.05
```

To conclude, we can see that the economic attractiveness is greater in East Harlemn.

In conclusion the most favorable place, economically speaking East Harlemn, if we take into account the population, as well as their consumption and preference for a type of trade, but also the attractiveness of the geographical area.

Location