Final Assignement IBM

**Report Capstone Project**

BY

MOLINA Jérémy

**2019**

Table des matières

[I. Introduction 2](#_Toc12414248)

[1. Description of the problem 2](#_Toc12414249)

[2. Background Discussion 2](#_Toc12414250)

[3. Audience 2](#_Toc12414251)

[4. Data 3](#_Toc12414252)

[II. Methodology 3](#_Toc12414253)

[a. Geopy and Map of New york 3](#_Toc12414254)

[b. Geopy and Manhattan map 4](#_Toc12414255)

[c. Explore and analyse neighborhood 4](#_Toc12414256)

[d. Explore data with K-mean and cluster analyse 6](#_Toc12414257)

[III. Conclusion 9](#_Toc12414258)

# Introduction

## 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.

## 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.

## 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.

## 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.

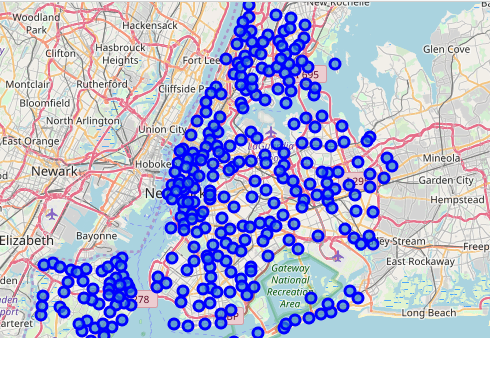
# Methodology

#### Data Exploration

### Geopy and Map of New york

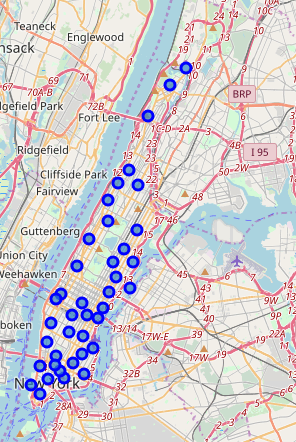
So after transform the panda into dataframe, I used geopy to get the latitude and longitude of new york, thereby allowing me to get geospatial data. Thanks to that, I was able to generate a map of New York showing all Borough. This allowed me to have an overview before focusing on Manhattan. This overview 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.



### Geopy and Manhattan map

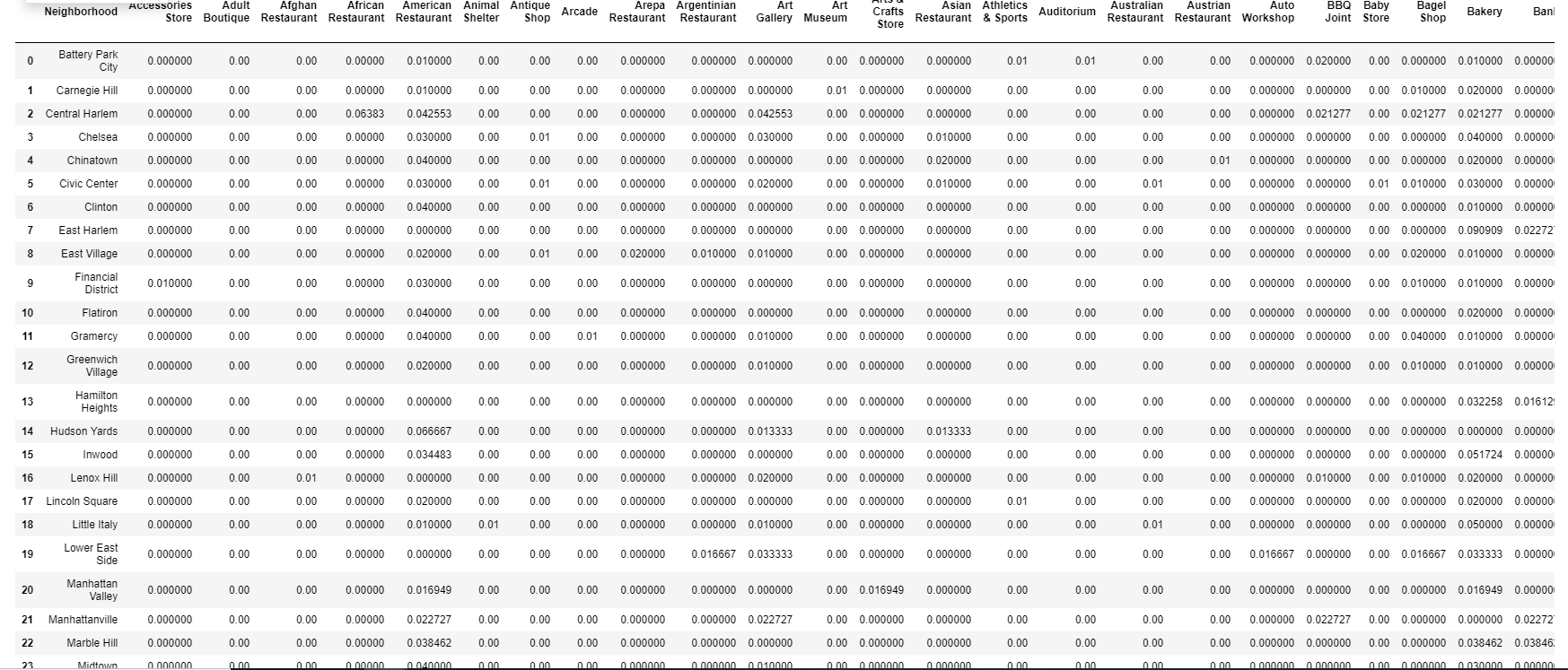
After this step, I decided to get geospatial data from Manhattan, allowing me to directly view all the neighborhoods of Manhattan, again for a better understanding of the study I generate a map so that it is more readable and easier located the different places of Manhattan.



### 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.



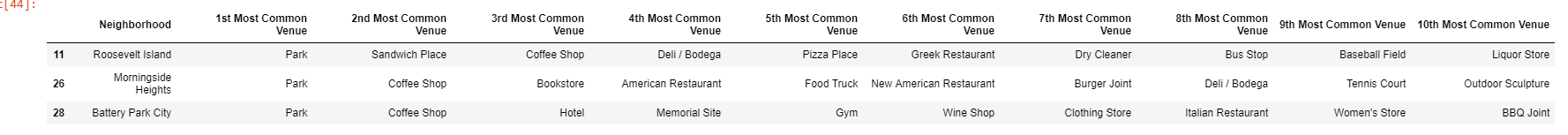
After having generated this statistical table showing all the shops for each neighborhood, it was necessary to obtain data that were both more precise, but also simpler to process, which would make it possible to add the attractiveness of a type of trade to the neighborhood. within a neighborhood. That's 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.



### 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 :



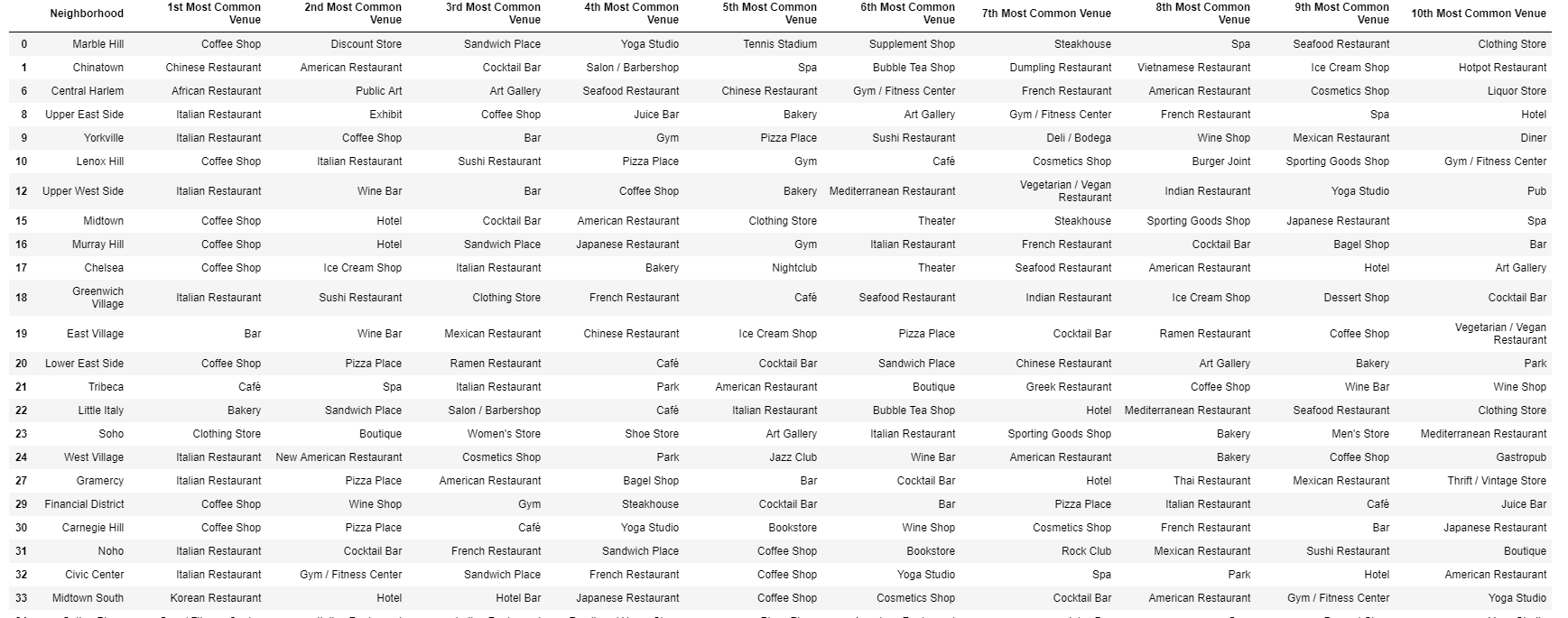
we realize that there is no French restaurant which could be a good news, but we can notice that it is mainly fast food that have a great success more than restaurants.

For the second cluster :



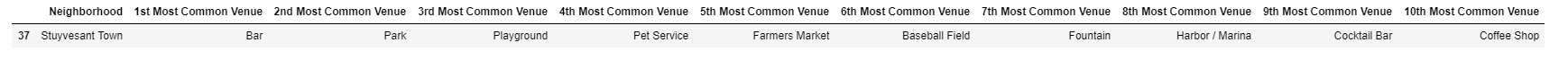
However, for this one, on the contrary, there are too many restaurants, two Italians, a Frenchman and American restaurants. However one can see a point of interest, indeed, we can see that the French restaurant is in 7th position. We can make two assumptions, either the restaurant is not good enough 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 :

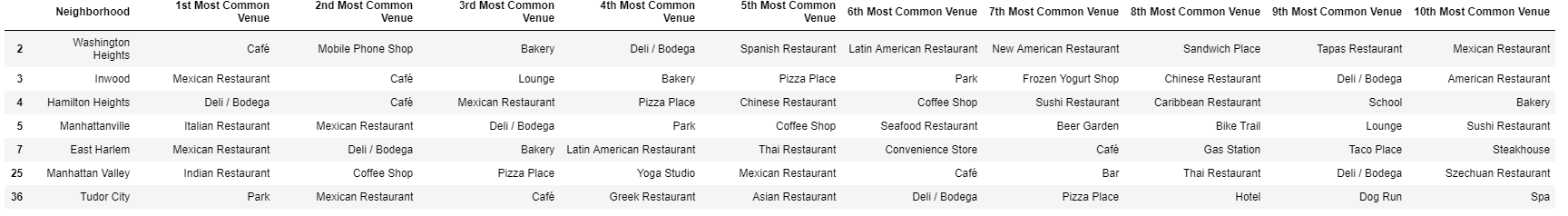


You can see that there are four French restaurants in all, all placed between 7th and 8th position. Thus even if we can see that the populations in these Neighborhoods have the area interested in French food, we can still assume that there are no positions in favour of opening our client’s restaurant,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

Cluster four :



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

Cluster five :

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.

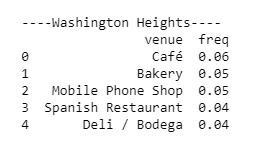
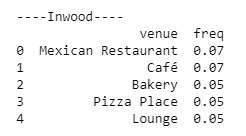
We can therefore offer the customer to create restaurants in Tudor City; Manhattan Valley; East Harlem; Manhattanville; Hamilton Heights; Inwood or Washington Heights

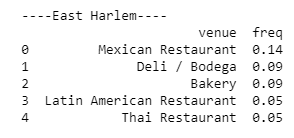
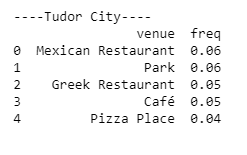
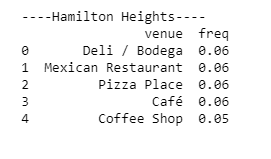
# Conclusion

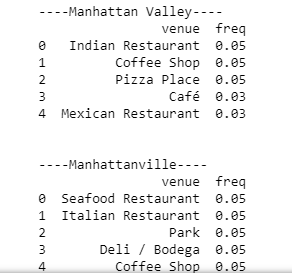
If we did a cluster analysis, we can see that the most favorable locations for the restaurant to be the most profitable are: in Tudor City; Manhattan Valley; East Harlem; Manhattanville; Hamilton Heights; Inwood or Washington Heights.

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 :

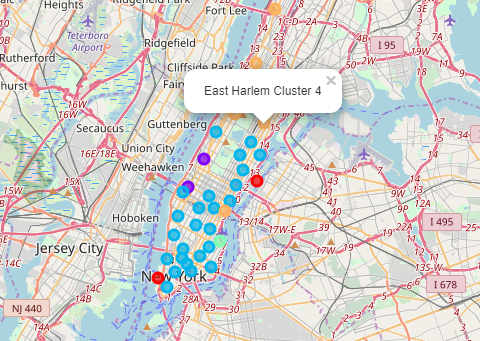
****



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**

****