The Battle of the Neighborhoods

1 - Introduction & Business Problem

1.1 - Problem Background

Among many cities in the United States of America, little do we doubt that New York City, is certainly one of the most important ones, either by its financial center, touristic places, sports teams, as well as the food and landscapes it has. It is one of the most important multicultural cities in the entire world, fully packed with job opportunities for everyone. In relation to this, the city is also well known as one of the major centers for financial activity, banks, world trading, shipments, advertising, fashion, arts and music.

Taking this characteristics into consideration, however, also means that it is highly competitive for everyone in this financial or comercial world, specially when trying to start a new businness, with all the competition that is already in the market. Consequently, each business or startup idea needs to be carefully analyzed. The resulting insights from this analysis would defenitely give a good understanding of the business environment, helping to target the market in a proper way, and reducing risks as far as possible.

1.2 - Problem Description¶

In this city, we have several markets and places to catch people's attention. Indeed, one of the most visited places, are restaurants, which are business that essentially prepare and serve food for different kind of customers. There is a wide variety of cuisine and fodd traditions in USA, and New York City is no Exception, fully crowded with people from all over the world, bringing their food for people to taste and order. Italian, Chinese, Indian and Japanese are just a few examples of it, providing the well known Chinese Food, Sushi, Pizzas, cafés, and spicy meals. Due to this variety of food, there are many places to choose from, when thinking about what to eat. Related to this, we could say that the there is really strong and demanding competitive market there, which leads to a very important aspect that anyone who want to start a business or restaurant should be taking care of: The Location. Regarding to the location of a market of restaurant, these are just a few factors that we should consider:

- **City Population**: The amount of People that live in New York City.
- **City Demographics**: The statistical data of the population.
- **Farmer Markets and Food sources**: Food markets nearby to buy fresh ingredients.
- **Cinemas, Cafés, Gyms and Entertaiment**: Places that are likely to attract a high amount of people.
- **Competition**: The local compatitors in that place.
- **Predominant Type of food**: The kind of food that people usually consume in that neighborhood.
- **Segmentation and Distribution of the Borough**: How the city and neighborhoods are distributed.

It is quite clear that, when planning to start up a market for the first time, location should be taking into account in a serious way.

1.3 - Target Audience¶

Recommending the best location in the city for a food place would definitely be suitable for everyone who is planning to take their first move in this kind of market, and who are also planning what they should do before starting their own restaurant. By finding the best location for them, we will ensure that they take as much benefit as possible from this location: a considerable ammount of clients, as well as a known or recognized place, with other attractions near them, to catch people's attention.

2 – Data

2.1 - Data 1

New York City has 5 boroughs and 304 neighborhoods, so we need a dataset with that information, giving a name for each one of them, as well as their location, with latitude and longitude. We can get that information from the following link: https://geo.nyu.edu/catalog/nyu-2451-34572 , which shows the neighborhoods names in 2014, with the information that we need, as it is shown in Figure 1.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 1: New York neighborhoods information

2.2 - Data 2

As we need to known places where to buy fresh food for the restaurant, we need information on farmer markets or grocery stores. We can get updated information about this from the following link : https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2/data While a farmer market and a grocery store may sound similar, a farmer market is a place where you can find several shops and people selling things fresh vegetables and fruits directly from the farm, which may result better in this case. Figure 2 shows an example of the data contained in the above link.

	Borough	Market Name	Street Address	Latitude	Longitude	Days of Operation	Hours of Operations	Season Dates	Accepts EBT	Open Year-Round	Stellar Cooking Demonstrations	Food Activities for Kids	Location Point
0	Brooklyn	Woodhull Hospital Youthmarket	Broadway & Flushing Ave	40.700726	-73.941932	Wednesday	9 a.m. - 2 p.m.	07/10/2019-11/27/2019	Yes	No	No	No	(40.700726, -73.941932)
1	Manhattan	Mount Sinai Hospital Greenmarket	E 99th St bet Madison & Park Aves	40.789169	-73.952743	Wednesday	8 a.m. - 5 p.m.	06/12/19-11/27/19	Yes	No	No	No	(40.789169, -73.952743)
2	Bronx	170 Farm Stand	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30 - 6:30 p.m.	07/10/2019-11/27/2019	Yes	No	No	Yes	(40.839882, -73.916783)
3	Manhattan	Greenmarket at Oculus Plaza	Church & Fulton Sts, on Oculus Plaza	40.711535	-74.010464	Tuesday	7 a.m. - 7 p.m.	07/09/2019-11/30/19	Yes	Yes	No	No	(40.711535, -74.010464)
4	Queens	Ditmars Park Youthmarket	Steinway St bet Ditmars Blvd & 23rd Ave, at Di...	40.772854	-73.906061	Saturday	8 a.m. - 3 p.m.	07/13/2019-11/23/2019	Yes	No	No	No	(40.772854, -73.906061)

Figure 2: Farm Markets information

2.3 - Data 3

We would also need, as mentioned before, information about the city's population, as well as their demographics and cuisine styles, which can easily be obtained from wikipedia, with the following links:

- https://en.wikipedia.org/wiki/New_York_City
- https://en.wikipedia.org/wiki/Economy_of_New_York_City
- https://en.wikipedia.org/wiki/Portal:New_York_City
- https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

2.4 - Data 4: Foursquare API

We would use the Foursquare API to gather information about venues for each neighborhood, giving coordinates as input. The information obtained from this API can be seen in Figure 3.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Figure 3: Example of Foursquare API information result.

3 - Methodology

As we mentioned before, our goal in this project is to find an optimum location for a new restaurant business in New York, for a certain company or person planning to set it up.

3.1 - Analytic Approach

From the data and explanation in the **data section** , we can appreciate that New York City has 5 boroughs and 306 neighborhoods. In relation with this boroughs, we will make two clusters. The first one, would be to cluster Manhattan and Brooklyn. As well as this, a second cluster will be made for Bronx, Queens and Staten Island. We will explain why we would make clusters in this way.

3.2 - Exploratory Data Analysis :

3.2.1 - Data 1 - New York Geographical Coordinates Data

In this part of the report, we will take a closer look at the data from the section **Data 1**. First of all, we transform that data, from a json file, into a pandas dataframe, so as to work with it in python. This dataframe contains geographical coordinates of New York City neighborhoods, and will also be used to get Venues from Foursquare. Using **geopy** and **folium** libraries, we create a map of New York with color dots, that represents each neighborhood, as it can be seen in figure 4.

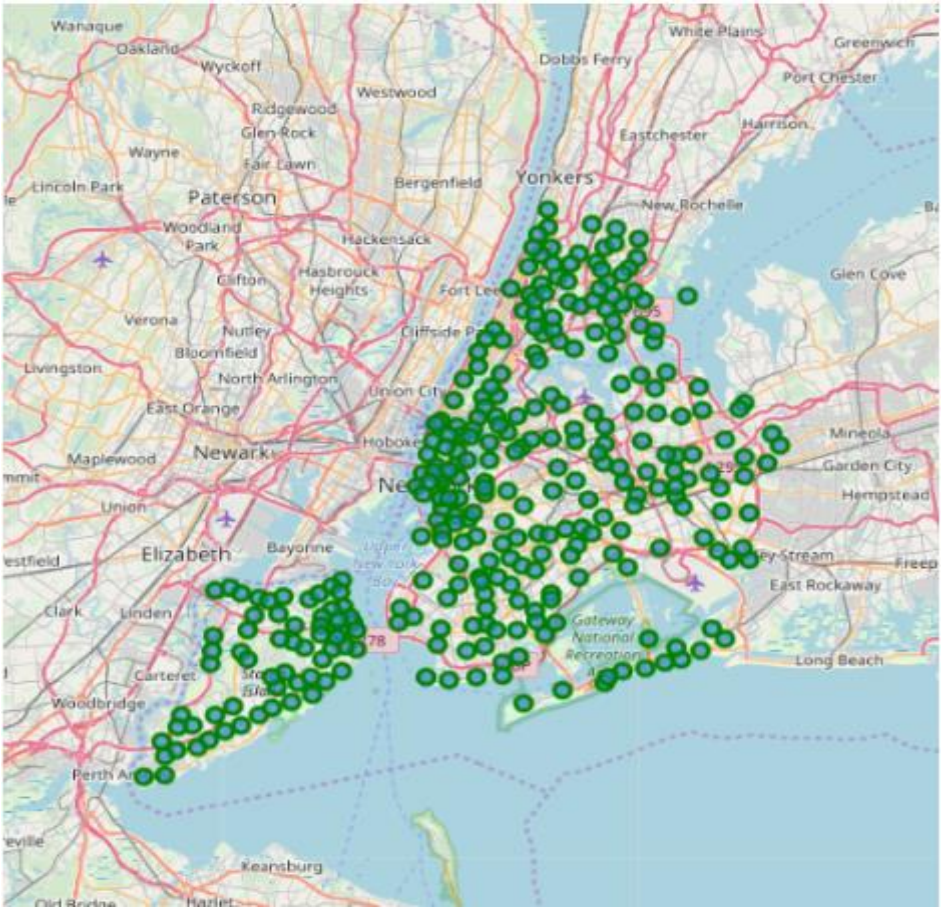

Figure 4: Map of New York City with Neighborhoods

3.2.2 - Data 2 - Farmer Markets

The information gathered from the files in **Data 2** section, shows us that there are 144 Farmer Markets in New York City. Nevertheless, the highest number of markets can be found in Manhattan and Brooklyn. On the other hand, a lower quantity can also be found in Queens,Bronx and Staten Island. If we make a barplot like the one on Figure 5, we could see the difference in markets between each of this Boroughs.

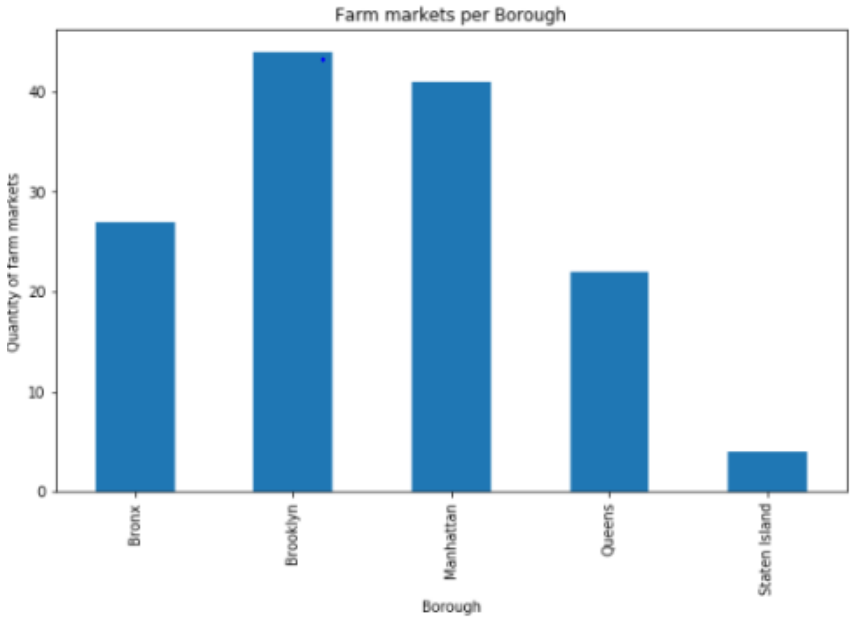

Figure 5: Barplot of farm markets per Borough

We can also create a new map of New York City with folium, showing the farm markets, like the ones in Figure 6.

Figure 6: Map of New York City with Farm Markets

3.2.3 - Data 3 - Population and Demographics

In order to analyze New York's Population, Demographics and styles of Cuisine, we use the data available from the links in the the **Data 3** section (2.2) When analyzing the city's population, we are able to see that Manhattan is the most geographically small and densely populated borough, whereas Staten Island is one of the boroughs with a smaller density of people. As well as this, we can also see that Brooklyn is the most populous borough among people, and also Queens is geographically the largest borough. We can arribe to this conclusions by looking at Figure 7.

	Borough	County	Estimate_2017	billions	per capita	square_miles	suare_km	persons_sq_mi,	persons_km2
0	The Bronx	Bronx	1471160	42.695	29200	42.10	109.04	34653	13231
1	Brooklyn	Kings	2648771	91.559	34600	70.82	183.42	37137	14649
2	Manhattan	New York	1664727	600.244	360600	22.83	59.13	72033	27826
3	Queens	Queens	2358582	93.310	39600	108.53	281.09	21460	8354
4	Staten Island	Richmond	479458	14.514	30300	58.37	151.18	8112	3132
5	City of New York	City of New York	8622698	842.343	97700	302.64	783.83	28188	10947
6	State of New York	State of New York	19849399	1701.399	85700	47214	122284	416.4	159

Figure 7: Population Table

3.2.4 - Data 4 - New York's Cuisine

The data for this section was manually gathered from the Cuisine Wikipedia link provided in the section **data 3**.

Bronx Cuisine

As we see in the word cloud of Figure 8, the most popular cuisines were puerto-rican and Italian.

Figure 8: Bronx Cuisine Word Cloud

Queens Cuisine

Figure 9 shows that the most popular cuisines in Queens are Indian,Pakistani and Italian.

Figure 9: Queens Cuisine Word Cloud

Brooklyn Cuisine

On the other hand, Figure 10 shows us that the most popular cuisines in Brooklyn are Italian,Mexican and Puerto-Rican.

Figure 10: Brooklyn Cuisine Word Cloud

Staten Island Cuisine

Staten Island has far less cuisine alternatives. Still, the word cloud on Figure 11 shows that the most popular ones are Indian and Italian.

Figure 11: Staten Island Cuisine Word Cloud

Manhattan Cuisine

From Figure 12 we can see that Manhattan has a wide variety of cuisines: Chinese,Indian,American,Italian and Puerto Rican.

Figure 12: Manhattan Cuisine Word Cloud

3.2.5 - Data 5 - Foursquare API

In order to explore and get information about New York City neighborhoods, we will use geographical coordinates as input for the Foursquare API, which gives as an output, venues information for each neighborhood. As we are going to make two clusters, we will create two new dataframes: one for the cluster that contains Brooklyn and Manhattan information, and the other one for the remaining boroughs, as it is shown in Figures 13 and 14.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
1	Marble Hill	40.876551	-73.910660	Boston Market	40.877430	-73.905412	American Restaurant
2	Bay Ridge	40.625801	-74.030621	Georgian Dream Cafe and Bakery	40.625586	-74.030196	Caucasian Restaurant
3	Bay Ridge	40.625801	-74.030621	Karam	40.622931	-74.028316	Middle Eastern Restaurant
4	Bay Ridge	40.625801	-74.030621	Elia Restaurant	40.623090	-74.031156	Greek Restaurant

Figure 13: Dataframe for Manhattan and Brooklyn

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898083	-73.850259	Caribbean Restaurant
1	Co-op City	40.874294	-73.829939	Arby's	40.870280	-73.828611	Fast Food Restaurant
2	Co-op City	40.874294	-73.829939	Townhouse Restaurant	40.876086	-73.828868	Restaurant
3	Co-op City	40.874294	-73.829939	Guang Hui Chinese Restaurant	40.876651	-73.829092	Chinese Restaurant
4	Eastchester	40.887556	-73.827806	Fish & Ting	40.885656	-73.829197	Caribbean Restaurant

Figure 14: Dataframe for the remaining Boroughs

3.2 Clustering and Results

From the venues data obtained from foursquare API, we filtered only the restaurant data for Brooklyn and Manhattan clustering, as well as the other remaining Boroughs (Queens,Bronx and Staten Island)

In order to cluster the neighborhoods into two clusters, we used the K-Means Algorithm, which aims to partitionate n observations into k clusters in which each observation belongs to the cluster with the nearest mean, using iterative refinement approach.

3.2.1 Brooklyn and Manhattan

By applying this algorithm to the correspondent data for brooklyn and Manhattan, we obtained the map showed in Figure 15.

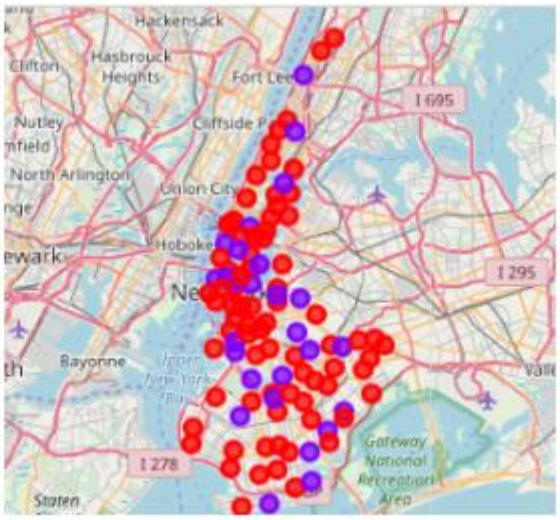

Figure 15: Clustering result map for Brooklyn and Manhattan

Cluster0(red): The total Sum of cluster 0 has a greater value, which also shows that the market is saturated, and the number of restaurants is very high.

Cluster1(purple): On the other hand, the total sum of cluster 1 has a smaller value, which shows that the market is not saturated.

Above all, we can say that there are no untapped neighborhoods in Brooklyn and Manhattan

3.2.2 Queens,Bronx and Staten Island

Regarding to this boroughs, we obtained the following map

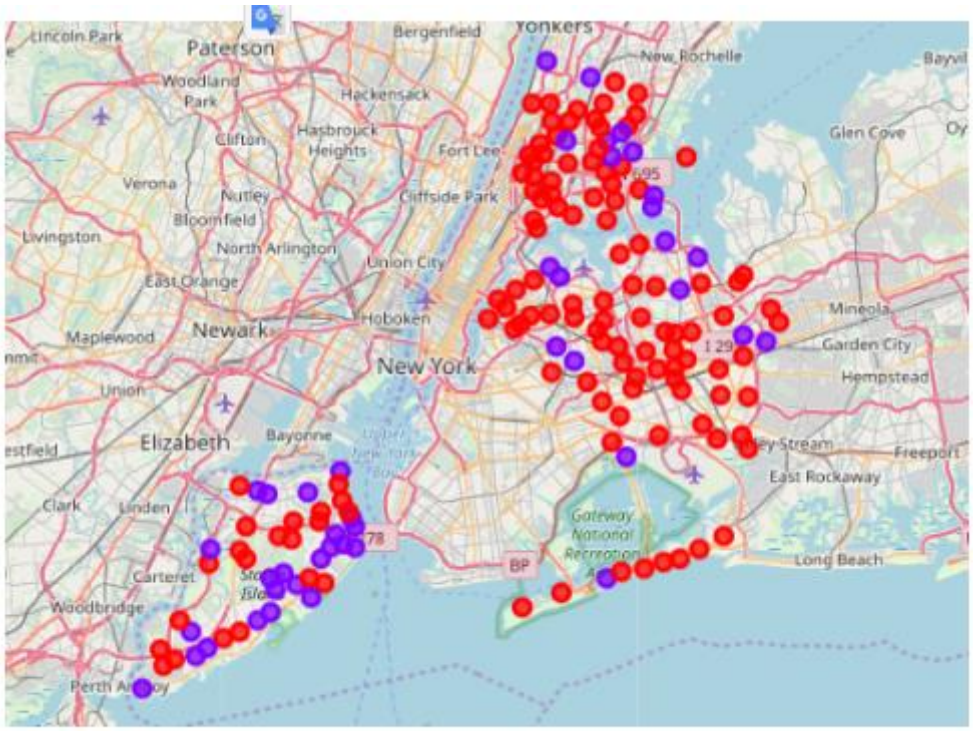

Figure 16: Clustering result map for Queens,Bronx and Staten Island

Cluster0(red): The total of Cluster 0 shows in general a greater value in this case, indicating that the market is saturated, with a high quantity of restaurants. In contrast with Brooklyn and Manhattan, however, we can also see that there are untapped neighborhoods.

Cluster1(purple): The cluster 1 indicates a smaller value, which means that the markets are not saturated, and there are not so many restaurants.

4 - Discussion

Taking into consideration the results metioned before, one of the facts that we can get from those results, is the possibility to increase farmer markets in Bronx, Queens and Staten Island, as there are not so many of them. Furthermore, we are also able to explore and promove different kind of cuisines, due to the fact that compared to Manhattan and Brooklyn, there is not a wide variety of them. As well as this, we can also discuss about what was previously mentioned about Manhattan and Brooklyn. Based on the results that we were able to obtain, the word clouds shows us that this two Boroughs have a really wide variety of cuisines. As ae result of this, there is also a great competition and revalry between them, reason for which would be more risky to start up a restaurant in this places. However, it could result in a succesful move if it is planned well, with a great menu and attendance.

5 - Conclusion

While the analysis in this report was performed with limited data, from public web pages, it was extremely important to apply a wide amount of knowledge, learnt from all the courses that were taken. From analyzing the data, to making graphics and machine learning algorithms, this courses helped to understand the whole process a Data Scientist must make, in order to arrive to conclusions and results. Despite being a large process, finishing it and getting results really makes you feel capable of working in this area of programming and technology.

In relation with the results that we got, we can say that if there are a lot of restaurants, there might also be a lot of demand for them. Brooklyn and Manhattan are just one example of this high demand for the restaurant business. On the other hand, Bronx, Queens and Staten Island, also have a good and reasonable number of restaurants, but no as many as they might be required.