Federico Piriz
IBM Data Science Professional Certificate
Final course: Coursera Capstone

The Battle of the Neighborhoods – Final Report

1- Introduction and Business Problem:

Background:

New York City is one of the most popular cities in the world. It is a leader city in multiple industries, such as finance, banking, services of all kinds, art, fashion, etcetera. This means that this city has one of the greatest costs to start a company or business, thanks to the extremely competitive markets.

In this project, I will be analyzing the opportunities and threats of starting a new business in NYC. By doing this, a company can be able to determine if it is profitable to start business or not, and where and where to start it.

Problem Description:

New York City is famous for its great food, and the different types of restaurants, from all over the world. It includes cuisine from Eastern Europe, Italy, Ireland, China, Middle East, etc. Restaurants go from top of the class, fine restaurants, to informal food trucks.

Such concentration and competitivity in restaurants in NYC make it difficult for a new company, Chivito's Inc., to start a new restaurant. That is why, it is important for the company to choose a strategic place to start, with not much competitive restaurants nearby, but in a populated area, to increase sales.

Many factors are considered when deciding the new place, such as:

1. NYC Population
2. The existence of Farmers Markets nearby, to buy the ingredients at a low cost and fresh.
3. The existence of venues, such as parks, malls, cinemas, gyms, etc., where population is highly concentrated.
4. Competitors analysis
5. Segmentation of the different boroughs
6. Saturation of the markets
7. Etc.

All these aspects will be taken into account by Chivito's Inc. when deciding where to start up the new restaurant.

Target Audience:

Chivito's Inc. has hired me to analyze which neighborhood would be the best to start the new restaurant. This analysis would be in the best interests of anybody who is thinking of starting a new restaurant in New York City.

Success Criteria:

This project would be successful if a good recommendation of the neighborhood to start the new restaurant was made to Chivito's Inc. This recommendation would be based on the absence of competitors and the presence of Farmers Markets in the area.

2- Data :

New York City is going to be analyzed in this project.

Data source 1: It is a data frame that consists of a total of 306 neighborhoods and 5 boroughs. Also, it includes the latitude and longitude of each neighborhood. This data source is free and it can be acceded in https://geo.nyu.edu/catalog/nyu_2451_34572.

Data source 2: This data frame includes the DOCHMH (Department of Health and Mental Hygiene) Farmers Markets. Farmers Markets are public spaces where producers gather to sell primary goods, more specifically, raw food ingredients, to consumers. This data frame is available in https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2.

Data source 3: Information from Wikipedia will be used to analyze New York City population. The data frame is available in https://en.wikipedia.org/wiki/Demographics_of_New_York_City.

Data source 4: Foursquare API will be used to extract venues from different neighborhoods in New York City. The geographical coordinates of NYC will be the input for the Foursquare API. An example of the result is given below.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
2	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Estrellita Poblana V	40.879687	-73.906257	Mexican Restaurant

3- Methodology:

The objective is to find for Chivito's Inc., the correct location for a new restaurant to start.

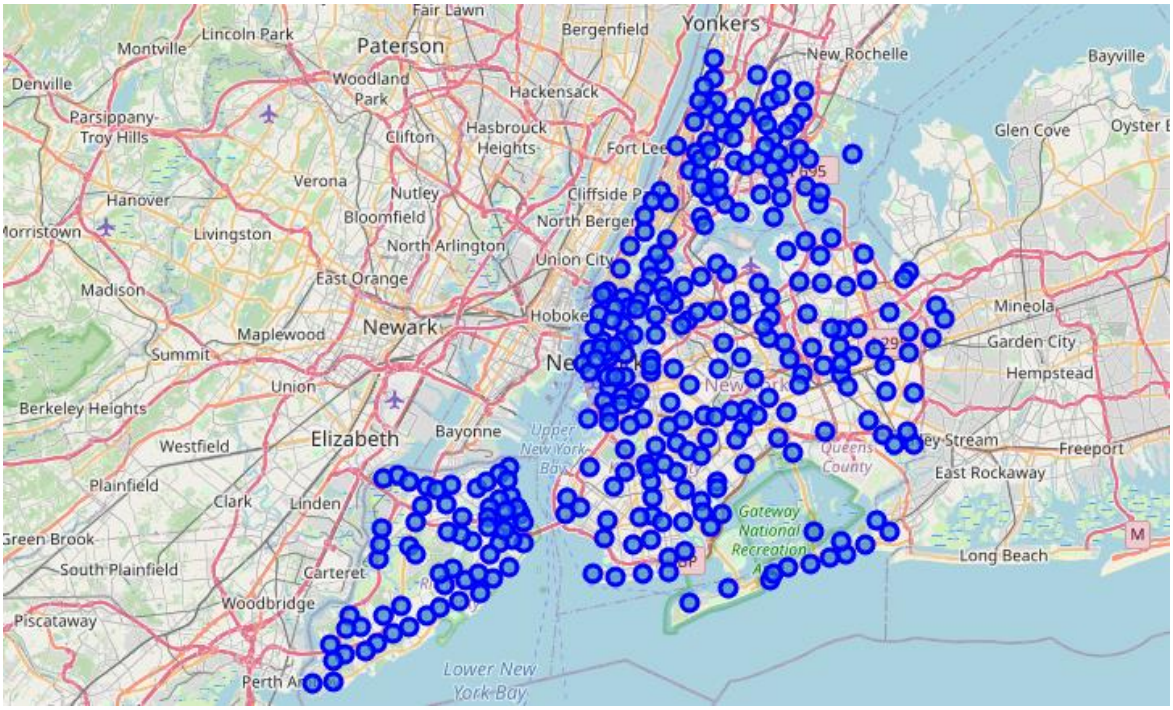
To accomplish this objective, I will analyze the data with the Foursquare API.

The Jupyter Notebooks in my Github Repository have the code needed to analyze this data.

The link to the repository is: https://github.com/federicopiriz/Coursera_Capstone.

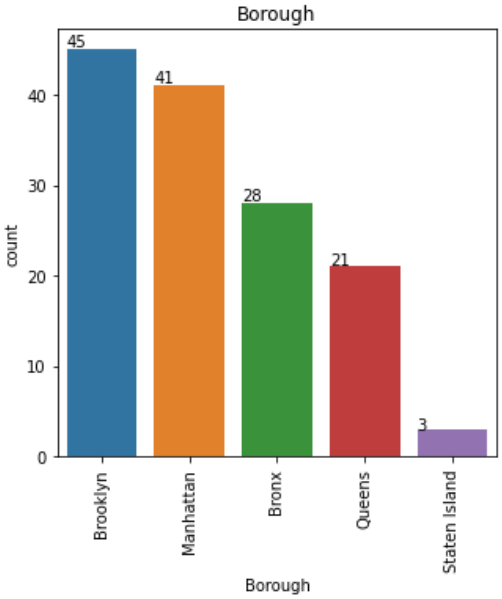
Data source 1: in this part, I use the Foursquare API and the Folium and Geopy libraries to get the locations of NYC neighborhoods. The result is showed below:

Neighborhoods in NYC



Data source 2: here I use Seaborn, Geopy, Matplotlib and Folium to render the map and graph of NYC Farmers Markets by neighborhood.

Brooklyn and Manhattan are the two boroughs with most farmers markets. This is convenient for a new restaurant to start. The results are:



Farmers Markets in NYC:

Data source 3: for this part, I use Beautiful Soup library, to scrape data from NYC Wikipedia website.

The results are:

1- Manhattan is NYC most dense in population and smallest borough.
2- Brooklyn is the most populous borough.
3- Queens is the largest borough.

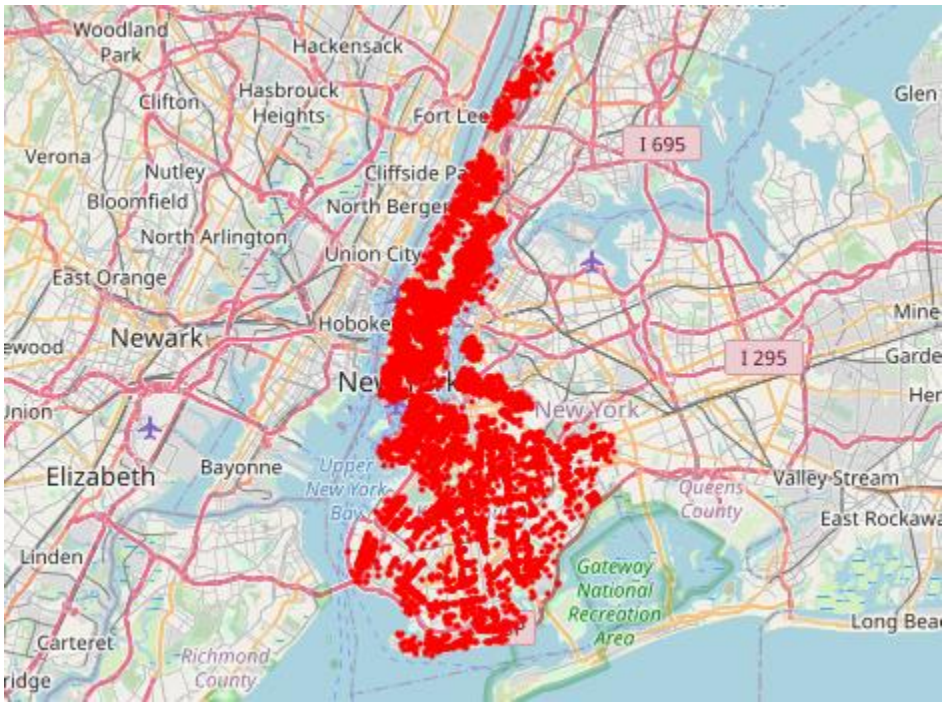
	Borough	County	Estimate_2017	GDP\n	square_miles	square_km	persons_sq_mi	persons/mi2
0	The Bronx	Bronx	1,418,207	42.695\n	42.10	109.04	33,867\n	
1	Brooklyn	Kings	2,559,903	91.559\n	70.82	183.42	36,147\n	
2	Manhattan	New York	1,628,706	600.244\n	22.83	59.13	71,341\n	
3	Queens	Queens	2,253,858	93.310\n	108.53	281.09	20,767\n	
4	Staten Island	Richmond	476,143	14.514\n	58.37	151.18	8,157\n	
5		City of New York	8,336,817	302.64	842.343	783.83	10,636\n	27,547
6		State of New York	19,453,561	47,126.40	1,731.910	122,056.82	159\n	412

Data source 4: I use the Foursquare API to extract venues of the different boroughs in NYC.

The results are:

1- Brooklyn and Manhattan venues

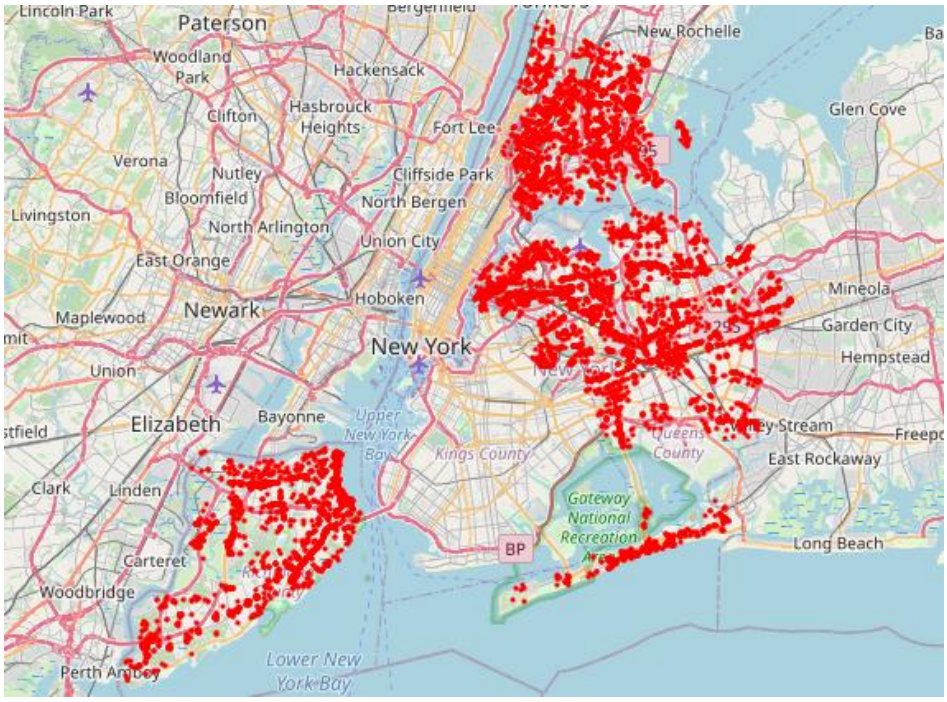
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
2	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Estrellita Poblana V	40.879687	-73.906257	Mexican Restaurant



2- Bronx, Queens and Staten Island venues

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy

Federico Piriz
IBM Data Science Professional Certificate
Final course: Coursera Capstone

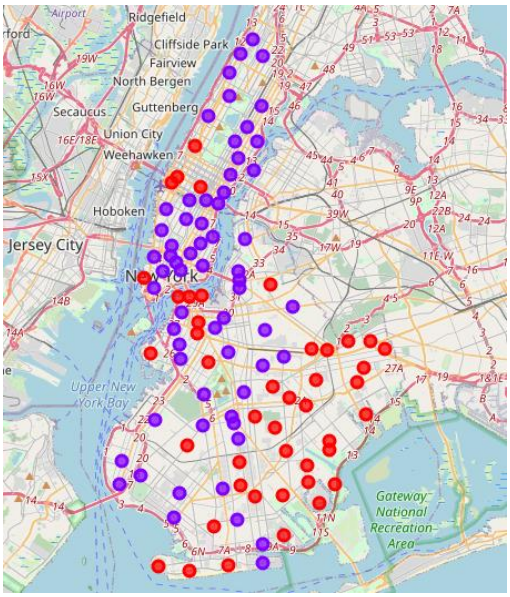


4- Results:

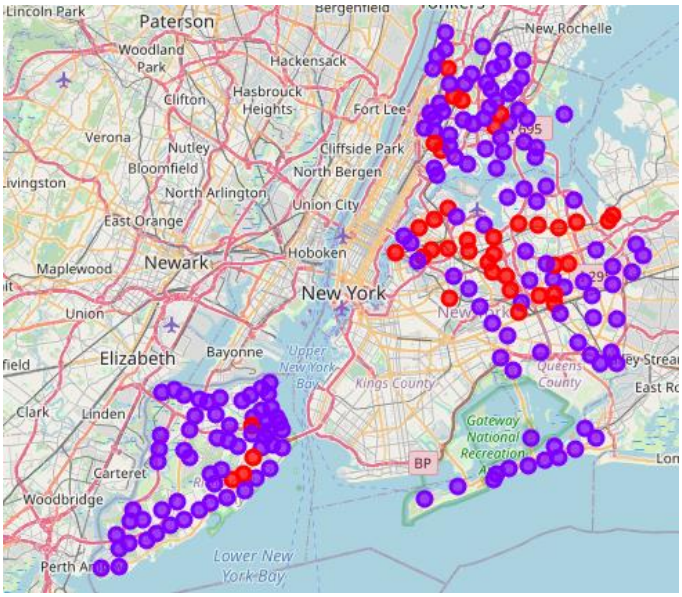
K-Means clustering was used for segmenting neighborhoods in NYC. Only restaurants venues were taken into consideration.

Neighborhoods were divided in two clusters, one with high saturation of restaurants and the other one with low saturation.

Manhattan and Brooklyn clustering



Bronx, Queens and Staten Island clustering



As it is shown, Neighborhoods in NYC are divided in two clusters. **Cluster 0** (Purple) includes the neighborhoods that are not saturated in number of restaurants. On the other hand, **Cluster 1** (Red) is composed by neighborhoods that have a market that is very saturated, number of restaurants is high.

5- Conclusion:

In general, New York City is highly saturated in restaurant venues. Anyway, there are options to consider when deciding to open a new restaurant. I would recommend Chivito's Inc. to start their first restaurant in Manhattan, as there are some neighborhoods that are not saturated (Purple), and the population there is enormous.

Another option would be to start in southern Queens, but this decision would imply less traffic of people nearby the new restaurant.

Further analysis is recommended to take a better decision, as the data source used for this project is really limited. Either way, it is enough to do a respectable recommendation.