Battle of Neighborhoods in New York City

Matteo Tortoli

Introduction

New York is a famous city, which attracts many tourists and workers from all over the world.

Opening a business activity in this city can be very profitable but also very risky, as it is necessary to know well the neighborhoods and the commercial businesses. Not all neighborhoods in New York City are the same, and for each of them there can be advantages and disadvantages, for example, a less famous neighborhood can attract less tourists and workers, but there should be less competition from competing businesses.

Business Problem

In this project the main goal is to find the best neighborhood/area to open an italian restaurant or pizzeria. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant/pizzeria** in **New York**, USA.

It's important to understand which is the right neighborhood to open this kind of activity, because in NYC there are already many italian restaurant/pizzeria activities, and therefore opening it in the right neighborhood can make the difference.

We will try to detect **locations that are not already crowded with restaurants**, especially we are particularly interested in **areas with no Italian restaurants/pizzeria in vicinity**. We would also prefer locations **as close to the biggest borough as possible**, assuming that first two conditions are met.

Data

Based on the business problem defined above, fist of all we must find the biggest borough in New York, or the most densely populated borough. After that, the factors on which we will base our choice are the following:

- number of restaurants in each neighborhood
- number of Italian restaurants or pizzeria in the neighborhood, if any
- Population density in each neighborhood
- Latitude and longitude of each neighborhood

Following data sources will be needed to extract/generate the required information:

- New York City boroughs are taken from wikipedia page: https://en.wikipedia.org/wiki/New_York_City#Boroughs. We will use this table to determine which is the "target borough".

- Neighborhoods, latitudes and longitudes are taken here: https://geo.nyu.edu/catalog/nyu_2451_34572 .
- Neighborhoods population density are taken here: https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp
- Foursquare API for find restaurants in each neighbohood, especially italian restaurant and pizzeria.

Boroughs Candidate

As mentioned before, first of all we need to understand which is our "target borough". To do so, we need to escrape data from the wikipedia page mentioned before.

The result is:

	Borough	County	Estimate (2018) [151]	billions(US$) [152]	per capita(US$)	square miles	squarekm	persons / sq. mi	persons /km2
0	The Bronx	Bronx	1,432,132	42.695	29,200	42.10	109.04	34,653	13,231
1	Brooklyn	Kings	2,582,830	91.559	34,600	70.82	183.42	37,137	14,649
2	Manhattan	New York	1,628,701	600.244	360,900	22.83	59.13	72,033	27,826
3	Queens	Queens	2,278,906	93.310	39,600	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	476,179	14.514	30,300	58.37	151.18	8,112	3,132

As we can see in the dataframe above, New York has 5 boroughs. Although Manhattan has the highest density per square meter (72,033), the estimated population is higher in Brooklyn (2,582,830), so is the city's most populous borough. Moreover, Brooklyn is known for its cultural, social, and ethnic diversity, therefore we can choose **Brooklyn** as our target borough.

Now that we have find our target borough, the next step is to analyze NYC map, more precisely Brooklyn borough.

Exploration of the target borough

After downloading the data and processing it, the result is the follow:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471
5	Brooklyn	Brighton Beach	40.576825	-73.965094
6	Brooklyn	Sheepshead Bay	40.586890	-73.943186
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
8	Brooklyn	Flatbush	40.636326	-73.958401
9	Brooklyn	Crown Heights	40.670829	-73.943291
10	Brooklyn	East Flatbush	40.641718	-73.936103

(to see the operations performed and the complete table, please consult the notebook file).

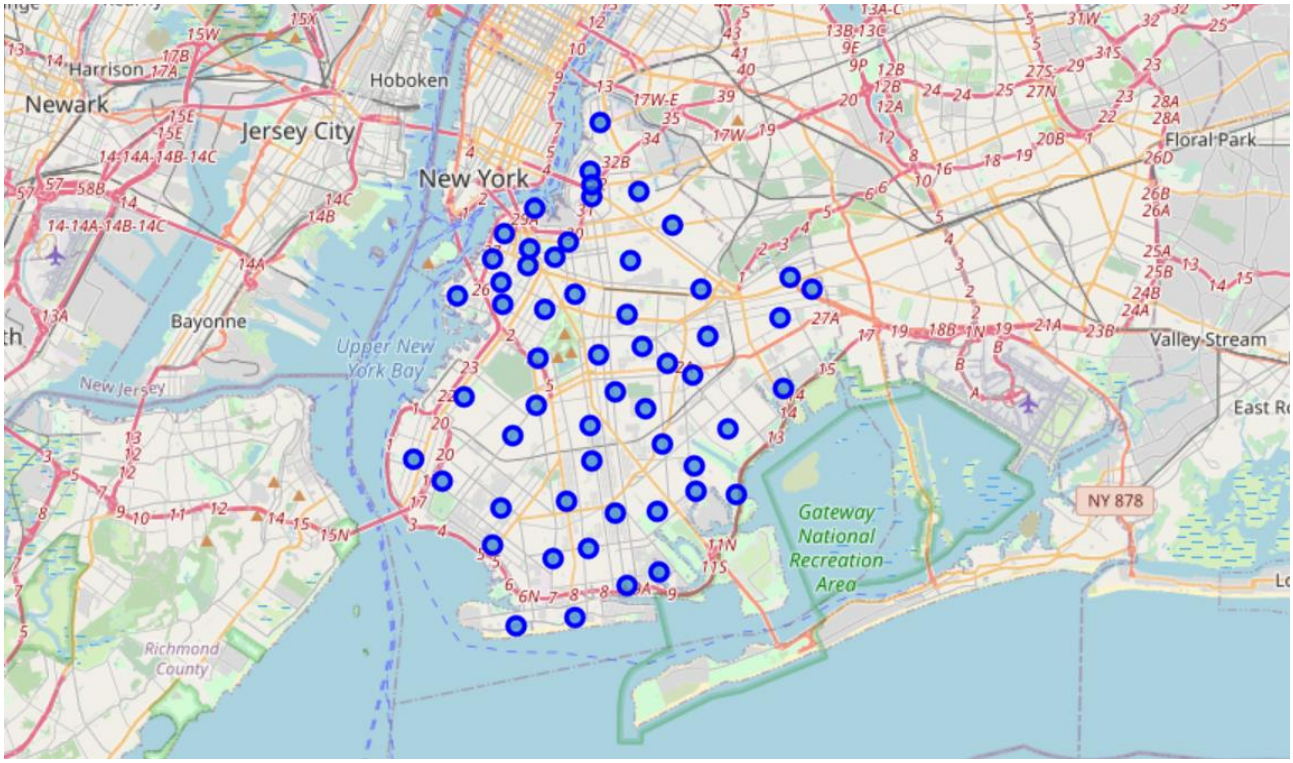
The next step is to find the population density for each Brooklyn neighborhood. We can download a dataset containing the population density from the following site: https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp

The result is:

	Borough	Neighborhood	Latitude	Longitude	Population Density
0	Brooklyn	Bay Ridge	40.625801	-74.030621	79371
1	Brooklyn	Bensonhurst	40.611009	-73.995180	62978
2	Brooklyn	Sunset Park	40.645103	-74.010316	72340
3	Brooklyn	Greenpoint	40.730201	-73.954241	34719
4	Brooklyn	Gravesend	40.595260	-73.973471	29436

Brooklyn map

Now we have all the geographical data for visualize map of Brooklyn:



Methodology

In this project we will direct our efforts on detecting areas of NYC that have low italian restaurant/pizzeria density and the highest possible population density. The target borough is Brooklyn, because is the borough with the highest population density.

In first step we have found the target borough, and find all the **geographical informations about it**: neighbohoords name, latitude, longitude, population density, and finally visualize a map of it.

Second step in our analysis will be exploration of **restaurants** across different neighborhoods of Brooklyn. To do so, we will use **Foursquare API**.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements**: the lowest density of italian restaurants/pizzeria and the highest population density. We will create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / to search the optimal venue location by stakeholders.

Analysis

As mentioned before, we analyze venues for each neighborhood of Brooklyn.

The result of this analys is a table containing the top ten most common venues in each neighborhood of Brooklyn:

	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	Bath Beach	Chinese Restaurant	Pharmacy	Bubble Tea Shop	Gas Station	Fast Food Restaurant	Italian Restaurant	Sushi Restaurant	Pakistani Restaurant	Sandwich Place	Surf Spot
1	Bay Ridge	Italian Restaurant	Spa	Pizza Place	American Restaurant	Greek Restaurant	Pharmacy	Bar	Chinese Restaurant	Playground	Bagel Shop
2	Bedford Stuyvesant	Pizza Place	Coffee Shop	Bar	Café	Discount Store	Juice Bar	Gift Shop	Cocktail Bar	Gourmet Shop	New American Restaurant
3	Bensonhurst	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Chinese Restaurant	Grocery Store	Donut Shop	Bridal Shop	Noodle House	Factory	Liquor Store
4	Bergen Beach	Harbor / Marina	Baseball Field	Athletics & Sports	Playground	Donut Shop	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant

(To see the full table, please consult the notebook file).

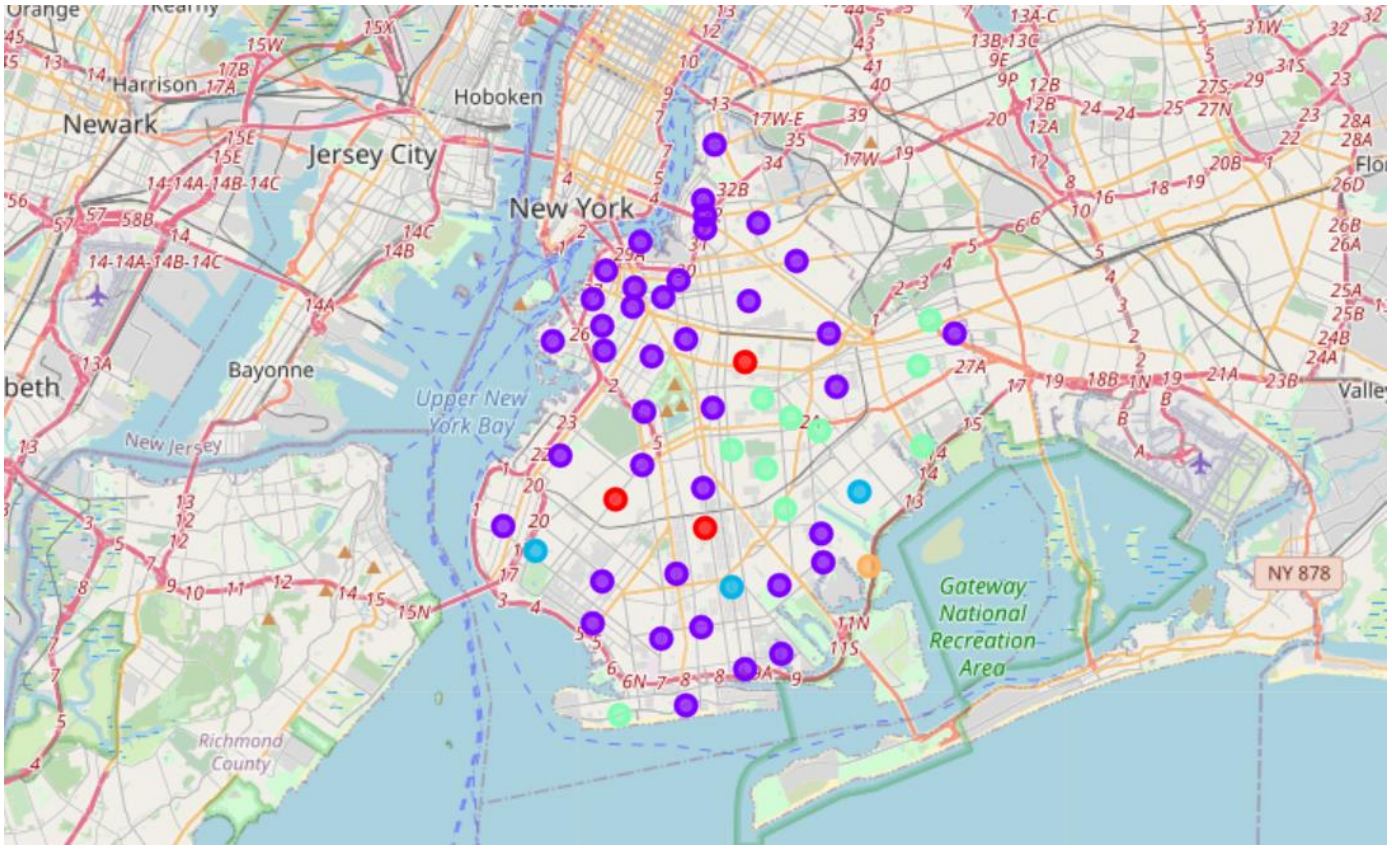
Cluster Neighborhoods

The third and final step is to cluster the neighborhood using **K-means clustering algorithm**. The table output after executed the algorithm is the following:

	Borough	Neighborhood	Latitude	Longitude	Population Density	Cluster Labels	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
0	Brooklyn	Bay Ridge	40.625801	-74.030621	79371	1	Italian Restaurant	Spa	Pizza Place	American Restaurant	Greek Restaurant	Pharmacy	Bar
1	Brooklyn	Bensonhurst	40.611009	-73.995180	62978	1	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Chinese Restaurant	Grocery Store	Donut Shop	Bridal Shop
2	Brooklyn	Sunset Park	40.645103	-74.010316	72340	1	Latin American Restaurant	Pizza Place	Mexican Restaurant	Bank	Bakery	Gym	Mobile Phone Shop
3	Brooklyn	Greenpoint	40.730201	-73.954241	34719	1	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Yoga Studio	Thrift / Vintage Store	Record Shop
4	Brooklyn	Gravesend	40.595260	-73.973471	29436	1	Chinese Restaurant	Pizza Place	Lounge	Bakery	Diner	Gym	Music Store

(To see the full table, please consult the notebook file).

The dataframe above contains the borough, the neighborhood, population density, the latitude and longitude of each zone, the cluster of which the neighborhood is part of, and the 10 most common venues. Now we have all the informations for visualize the clusters, where each dot color represents a different cluster.



Results and discussion

Finally we have all the informations for analyze the results and discuss which is the best neighborhood to open a an italian restaurant/pizzeria. Let's start analyzing each cluster:

Cluster 0

	Neighborhood	Population Density	Cluster Labels	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
9	Crown Heights	39670	0	Pizza Place	Café	Museum	Burger Joint	Coffee Shop	Convenience Store	Bank	Bakery	Bagel Shop	Candy Store
34	Borough Park	106357	0	Bank	Pizza Place	Deli / Bodega	Pharmacy	Grocery Store	Café	American Restaurant	Hotel	Farmers Market	Fast Food Restaurant
46	Midwood	52835	0	Pizza Place	Ice Cream Shop	Pharmacy	Candy Store	Bakery	Bagel Shop	Electronics Store	Video Game Store	Convenience Store	Flower Shop

Although there is a high population density (especially at Borough Park), the most common venue in this cluster is Pizza Place, a therefore it is not recommended to open a pizzeria. However, considering the low presence of ethnic restaurants and the high population density in **Borough Park**, this neighborhood could be a **good option** to open an Italian restaurant.

Cluster 1

	Neighborhood	Population Density	Cluster Labels	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
0	Bay Ridge	79371	1	Italian Restaurant	Spa	Pizza Place	American Restaurant	Greek Restaurant	Pharmacy	Bar	Chinese Restaurant	Playground
1	Bensonhurst	62978	1	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Chinese Restaurant	Grocery Store	Donut Shop	Bridal Shop	Noodle House	Factory
2	Sunset Park	72340	1	Latin American Restaurant	Pizza Place	Mexican Restaurant	Bank	Bakery	Gym	Mobile Phone Shop	Fried Chicken Joint	Pharmacy
3	Greenpoint	34719	1	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Yoga Studio	Thrift / Vintage Store	Record Shop	New American Restaurant	Tea Room
4	Gravesend	29436	1	Chinese Restaurant	Pizza Place	Lounge	Bakery	Diner	Gym	Music Store	Spa	Music Venue
5	Brighton Beach	35547	1	Restaurant	Eastern European Restaurant	Russian Restaurant	Beach	Mobile Phone Shop	Bank	Sushi Restaurant	Gourmet Shop	Supplement Shop
6	Sheepshead Bay	64518	1	Turkish Restaurant	Dessert Shop	Sandwich Place	Yoga Studio	Buffet	Diner	Miscellaneous Shop	Café	Chinese Restaurant

(to see the full table, please consult the notebook file)

this is the largest cluster, in fact, it contains most of the neighborhoods of Brooklyn. This cluster is characterized by a massive presence of ethnic restaurants of all kinds (Indian, Japanese, Italian, Chinese, etc.), especially in larger neighborhoods. Open an Italian restaurant/pizzeria in one of these neighborhoods could be a risky choice.

Cluster 2

	Neighborhood	Population Density	Cluster Labels	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
28	Canarsie	83693	2	Deli / Bodega	Asian Restaurant	Food	Caribbean Restaurant	Thai Restaurant	Grocery Store	Chinese Restaurant	Gym	Dry Cleaner	Discount Store
35	Dyker Heights	42419	2	Burger Joint	Dance Studio	Golf Course	Bagel Shop	Food	Yoga Studio	Fish & Chips Shop	Farmers Market	Fast Food Restaurant	Field
68	Madison	38917	2	Bagel Shop	Dessert Shop	Italian Restaurant	Deli / Bodega	Candy Store	Pizza Place	Pilates Studio	Filipino Restaurant	Farm	Farmers Market

in this cluster the italian restaurants and pizzeria represent the most popular venues, especially in Madison neighborhood, where opening an italian restaurant or pizzeria could lead to high competition due to the other similar commercial businesses already present. However, **Canarsie** neighborhood could be a valid choice to open italian restaurant or pizzeria, because it has a high population density and there aren't competing activities. However, the presence of other ethnic restaurants, although not italian, should not be underestimated.

Cluster 3

	Neighborhood	Population Density	Cluster Labels	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
10	East Flatbush	50355	3	Food & Drink Shop	Pharmacy	Chinese Restaurant	Fast Food Restaurant	Park	Supermarket	Caribbean Restaurant	Hardware Store	Print Shop
25	Cypress Hills	49223	3	Fried Chicken Joint	Donut Shop	Latin American Restaurant	Fast Food Restaurant	Ice Cream Shop	Food	Women's Store	Coffee Shop	Baseball Field
26	East New York	29343	3	Deli / Bodega	Spanish Restaurant	Gym	Pizza Place	Convenience Store	Salon / Barbershop	Event Service	Caribbean Restaurant	Fast Food Restaurant
27	Starrett City	13354	3	Moving Target	Pizza Place	Convenience Store	Cosmetics Shop	Donut Shop	Pharmacy	American Restaurant	Caribbean Restaurant	Food & Drink Shop
29	Flatlands	64762	3	Pharmacy	Fried Chicken Joint	Fast Food Restaurant	Caribbean Restaurant	Lounge	Paper / Office Supplies Store	Park	Electronics Store	Nightclub
32	Coney Island	31965	3	Caribbean Restaurant	Baseball Stadium	Brewery	Athletics & Sports	Pharmacy	Music Venue	Skating Rink	Food Court	Monument / Landmark

(to see the full table, please consult the notebook file)

also in this cluster there are many neighborhoods that have a high presence of ethnic restaurants (especially caribbean restaurant), some of them can be valid candidates to open an Italian restaurant/pizzeria, because there is a high population density and low competition by similar activities.

Cluster 4

	Neighborhood	Population Density	Cluster Labels	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
45	Bergen Beach	45231	4	Harbor / Marina	Baseball Field	Athletics & Sports	Playground	Donut Shop	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant

in this cluster there is only one neighborhood. Although in this neighborhood there are no Italian restaurants or pizzeria, compared to the previous candidates the population density is rather low, therefore other neighborhoods could be better candidates than this.

Conclusion

The purpose of this project was to identify neighborhoods in New York with a low presence of restaurants, especially Italian restaurants or pizzeria, in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant/pizzeria.

First of all, we analyzed which could be the best borough to open the restaurant. The analysis led us to choose Brooklyn, thanks to its higher population density than other boroughs. Then, by finding restaurants presence in each Brooklyn'S neighborhood from Foursquare data we have clustered them. Clustering those locations was performed in order to create major zones of interest (containing greatest number of potential locations) which satisfy some basic requirements regarding existing nearby italian restaurants/pizzeria.

The final decision about the best italian restaurant/pizzeria location will be made by stakeholders based on the specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.