

Capstone Project - The Battle of Neighborhoods

12.12.2019

Houssam AlRachid

Introduction

Choosing the right location for your restaurant is critical to your success. It is easy to be impulsive and end the search altogether. However, establishing roots for your restaurant is not something that should be taken lightly. There are many factors to consider that can make all the difference in your ultimate success. Your restaurant's location affects the success of your business and the speed at which your investments are paid back. An astutely chosen location provides you with a stable traffic of customers and high profits.

An investor is looking to open a new **restaurant in New York City (NYC)**, but he is not sure about the best location for his new venue. NYC is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles, New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities. A global power city, New York City has been described as the cultural, financial and media capital of the world.

Therefore it is very natural that the city contains a large number of restaurants, making the market highly competitive. Thus, opening a new restaurant needs to be analysed carefully.

In this project, we will do data analysis to find an optimal location to open a new restaurant. In this project, we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a Japanese restaurant in NYC. This project would interest investors who want to open a new Japanese restaurant in NYC.

Data

Based on our problem comprehension, several factors will influence our decision:

1. Number of existing restaurants in each neighborhood of NYC;
2. Number of existing Japanese restaurants in each neighborhood of NYC;
3. Distance between Japanese restaurants in each neighborhood of NYC;
4. Distance of neighborhood from **NYC Center**.

The necessary data will be needed are:

1. NYC has a total of 5 boroughs and 306 neighborhoods. We will essentially need a dataset that contains all the boroughs and neighborhoods that exist in each borough along with their latitude and longitude.

Source: https://geo.nyu.edu/catalog/nyu_2451_34572

- Information on venues in the neighborhoods of NYC.

Source: Fousquare API;

- Japanese restaurants in each neighborhood of NYC.

Source: Fousquare API.

- Category IDs corresponding to Japanese restaurants were taken from Foursquare.
Source : <https://developer.foursquare.com/docs/resources/categories>

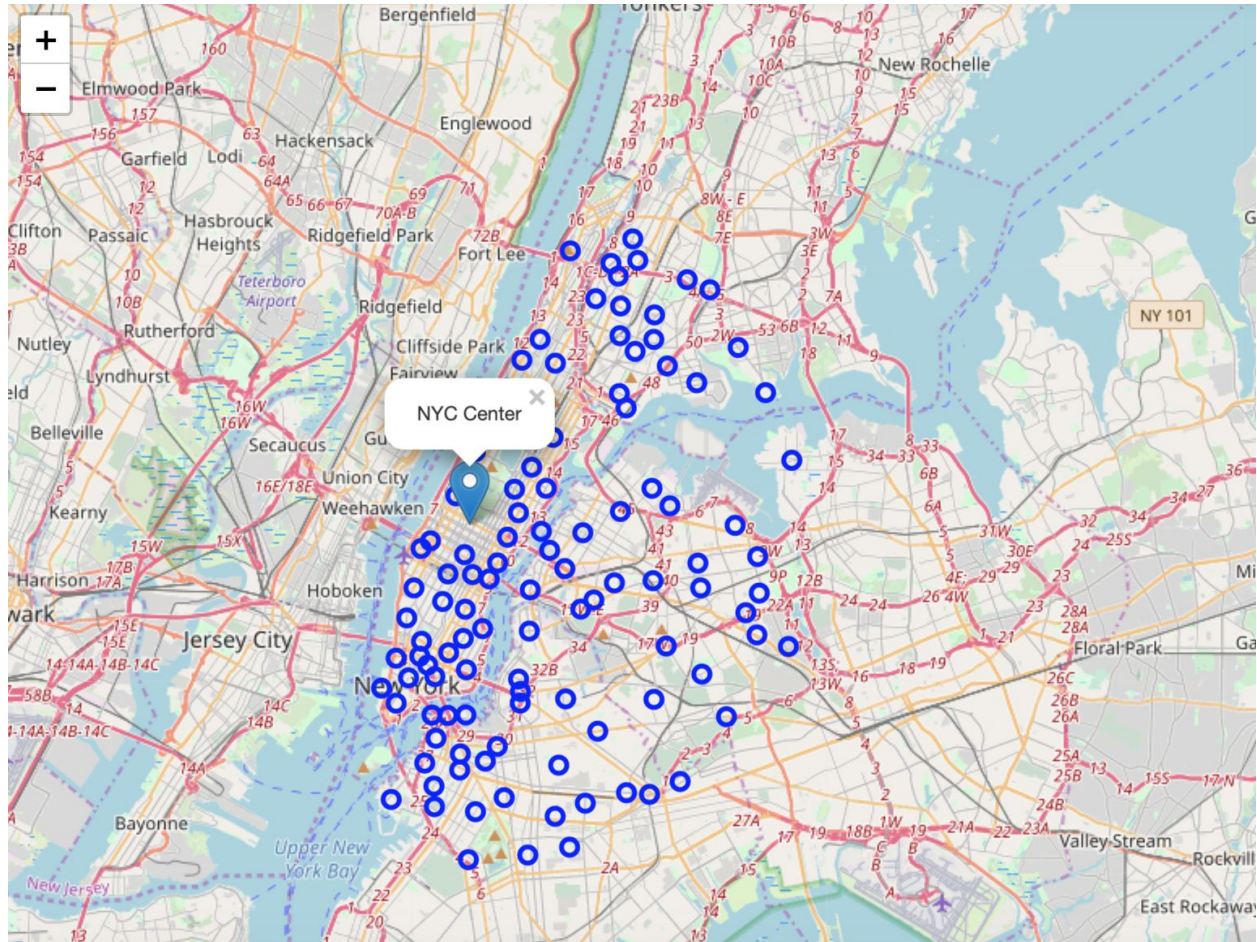
We start by downloading data in order to segment the **neighborhoods in NYC** and explore them. We will essentially need a dataset that contains all the 5 boroughs and the 306 neighborhoods associated with their latitude and longitude coordinates. All the relevant data is in the features key, which is basically a list of the neighborhoods. So, we define a new variable that includes this data. The next task is essentially transforming this data of nested Python dictionaries into a pandas dataframe. We then obtain a dataframe table containing 306 neighborhoods of this form:

	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

Next, we compute the distances of each Neighbourhood from NYC center and we keep only the neighborhoods **within ~15 km** from NYC Center.

	Borough	Neighborhood	Latitude	Longitude	Distance from center	X	Y
0	Bronx	University Heights	40.855727	-73.910416	14337.285190	4.649657e+06	5.862515e+06
1	Bronx	Morris Heights	40.847898	-73.919672	12948.680744	4.649537e+06	5.861096e+06
2	Bronx	East Tremont	40.842696	-73.887356	14290.222060	4.652610e+06	5.862538e+06
3	Bronx	West Farms	40.839475	-73.877745	14657.713936	4.653661e+06	5.862787e+06
4	Bronx	High Bridge	40.836623	-73.926102	11293.397863	4.649930e+06	5.859480e+06

Let's visualize the data we have so far: NYC Center location and candidate neighborhood centers **within ~15km** :

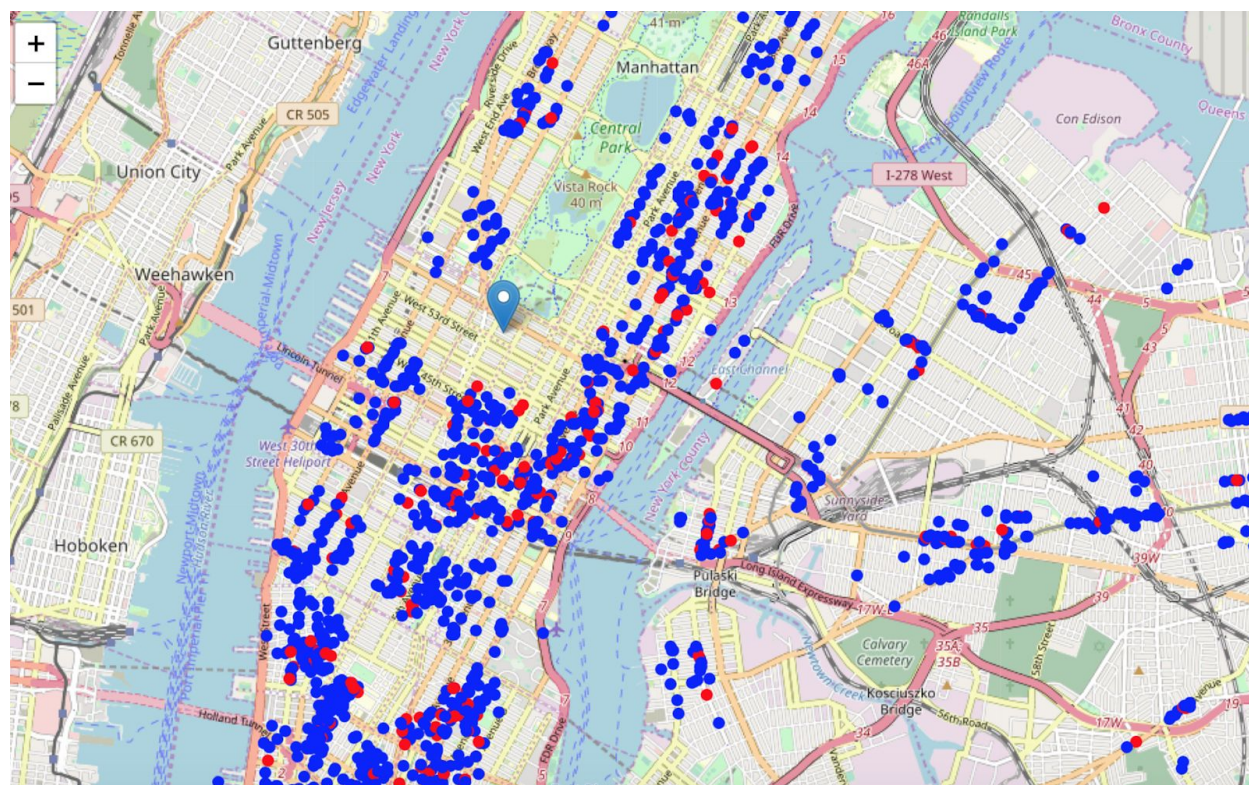


Next, we are going to start utilizing the **Foursquare API** to explore the restaurants in each neighborhood. We will include in our list only venues that have '**restaurant**' in category name, and we'll make sure to detect and include all the subcategories of specific '**Japanese restaurant**' category, as we need info on Japanese restaurants in the neighborhood.

After this, we search for all restaurants locations and get nearby ones in a **radius of 500 m** from each neighborhood. To do this, we start by getting relevant part of JSON and transform it into a pandas dataframe. We obtain the following results:

1. Total number of restaurants: 2734
2. Total number of Japanese restaurants: 289
3. Percentage of Japanese restaurants: 10.57%
4. Average number of restaurants in neighborhood: 26.85

Let's now see all the collected restaurants in our area of interest on the map, and let's also show Japanese restaurants in a different color.



The map illustrates all the restaurants in an area within a few kilometers from New York City Center, and we know which ones are Japanese restaurants (red circles)! We also know which restaurants are in the vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Japanese restaurant!

Methodology

In this present project, we will focus on detecting areas of New York City that have **low restaurant density**, particularly those with **low number of Japanese restaurants**. We will limit our analysis to area ~15km around NYC Center.

The methodology can be resumes in three big steps as follows:

1. We have collected the required data: **location and type (category) of every restaurant within 15km from NYC Center**. We have also **identified Japanese restaurants** (according to Foursquare categorization);

2. We will compute and explore '**restaurant density**' across different areas of NYC - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Japanese restaurants in vicinity) and focus our attention on those areas;
3. We focus on the most promising areas and within those **create clusters of locations that meet some basic requirements established in discussion with stakeholders**: we will take into consideration locations with **no more than 5 restaurants in a radius of 500 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Analysis

We start by performing some basic exploratory data analysis and derive some additional info from our raw data. First let's count the number of restaurants in every area candidate. We obtain that the average number of restaurants in every area with a radius of 500m is equal 26.85. The following figure gives us details on each location:

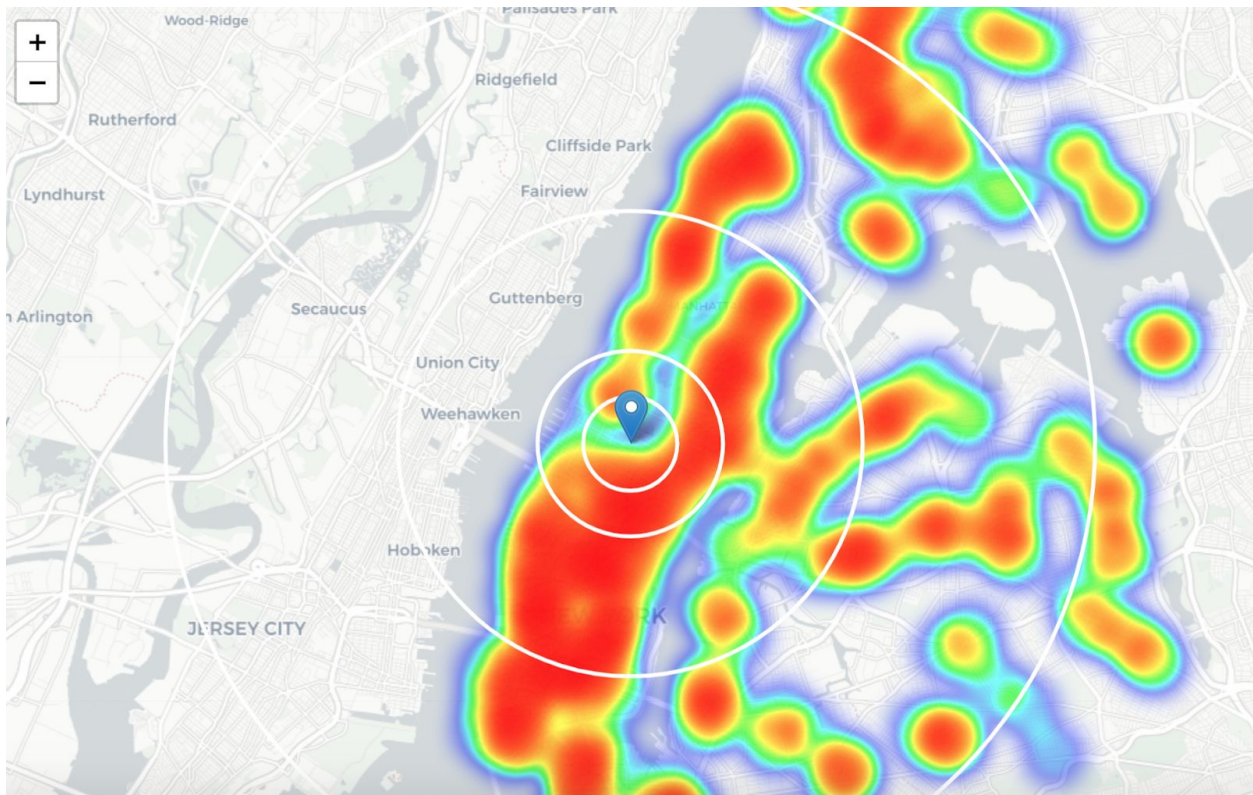
	Borough	Neighborhood	Latitude	Longitude	Distance from center	X	Y	Restaurants in area
0	Bronx	University Heights	40.855727	-73.910416	14337.285190	4.649657e+06	5.862515e+06	8
1	Bronx	Morris Heights	40.847898	-73.919672	12948.680744	4.649537e+06	5.861096e+06	5
2	Bronx	East Tremont	40.842696	-73.887356	14290.222060	4.652610e+06	5.862538e+06	7
3	Bronx	West Farms	40.839475	-73.877745	14657.713936	4.653661e+06	5.862787e+06	3
4	Bronx	High Bridge	40.836623	-73.926102	11293.397863	4.649930e+06	5.859480e+06	14
5	Bronx	Melrose	40.819754	-73.909422	10468.798950	4.652674e+06	5.858693e+06	5
6	Bronx	Mott Haven	40.806239	-73.916100	8704.857021	4.653229e+06	5.856819e+06	7
7	Bronx	Port Morris	40.801664	-73.913221	8544.128814	4.653839e+06	5.856503e+06	6
8	Bronx	Longwood	40.815099	-73.895788	11076.340286	4.654173e+06	5.859036e+06	4
9	Bronx	Hunts Point	40.809730	-73.883315	11701.258801	4.655634e+06	5.859230e+06	3

We calculate the **distance to the nearest Japanese** restaurant from every area candidate center (not only those within 500m - we want distance to closest one, regardless of how distant it is), we obtain:

	Borough	Neighborhood	Latitude	Longitude	Distance from center	X	Y	Restaurants in area	Distance to Japanese restaurant
0	Bronx	University Heights	40.855727	-73.910416	14337.285190	4.649657e+06	5.862515e+06	8	2771.935757
1	Bronx	Morris Heights	40.847898	-73.919672	12948.680744	4.649537e+06	5.861096e+06	5	1776.801415
2	Bronx	East Tremont	40.842696	-73.887356	14290.222060	4.652610e+06	5.862538e+06	7	4183.623152
3	Bronx	West Farms	40.839475	-73.877745	14657.713936	4.653661e+06	5.862787e+06	3	4841.442198
4	Bronx	High Bridge	40.836623	-73.926102	11293.397863	4.649930e+06	5.859480e+06	14	1626.799685

Thus, on average, Japanese restaurant can be found **within ~1km** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

To do this, we start by creating a **Heatmap** showing the density of restaurants and try to extract some meaningful info from that. Also, we added a few circles indicating distance of 1km, 2km, 5km and 10km from NYC Center, as illustrates the following figure:



It looks like a few pockets of low restaurant density closest to the city center can be found **East and North from NYC Center** compared to south. The West is not counted since it does not belong to NYC! Let's illustrate another Heatmap showing the density of Japanese restaurants only.



This map is not so 'hot' (Japanese restaurants represent a subset of ~10% of all restaurants in NYC) but it also indicates **higher density of existing Japanese restaurants directly South from NYC Center, with closest pockets of low Japanese restaurant density positioned North and East from city center.**

Based on this we will now focus our analysis on **areas North and East from NYC Center** - we will move the center of our area of interest and reduce its size to have a **radius of 10km**. This places our location candidates mostly in boroughs **Brooklyn, Queens and the Bronx**.

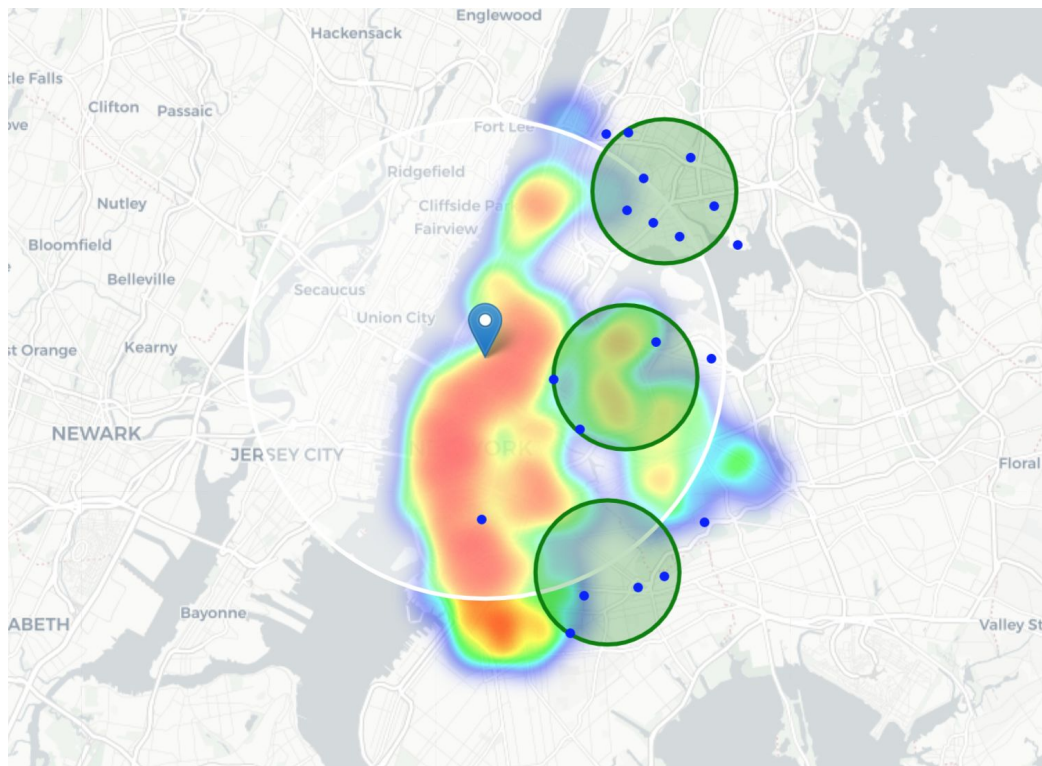
To do this, we reduce our neighborhoods dataframe to contain only parts of Brooklyn, Queens and The Bronx closest to NYC Center. We are interested only in **locations with no more than 5 restaurants in a radius of 500 meters, and no Japanese restaurants in a radius of 500 meters**. We obtain :

1. Locations with no more than 4 restaurants nearby: 21
2. Locations with no Japanese restaurants within 500 m: 41
3. Locations with both conditions met: 19

Let's see those good locations on map....



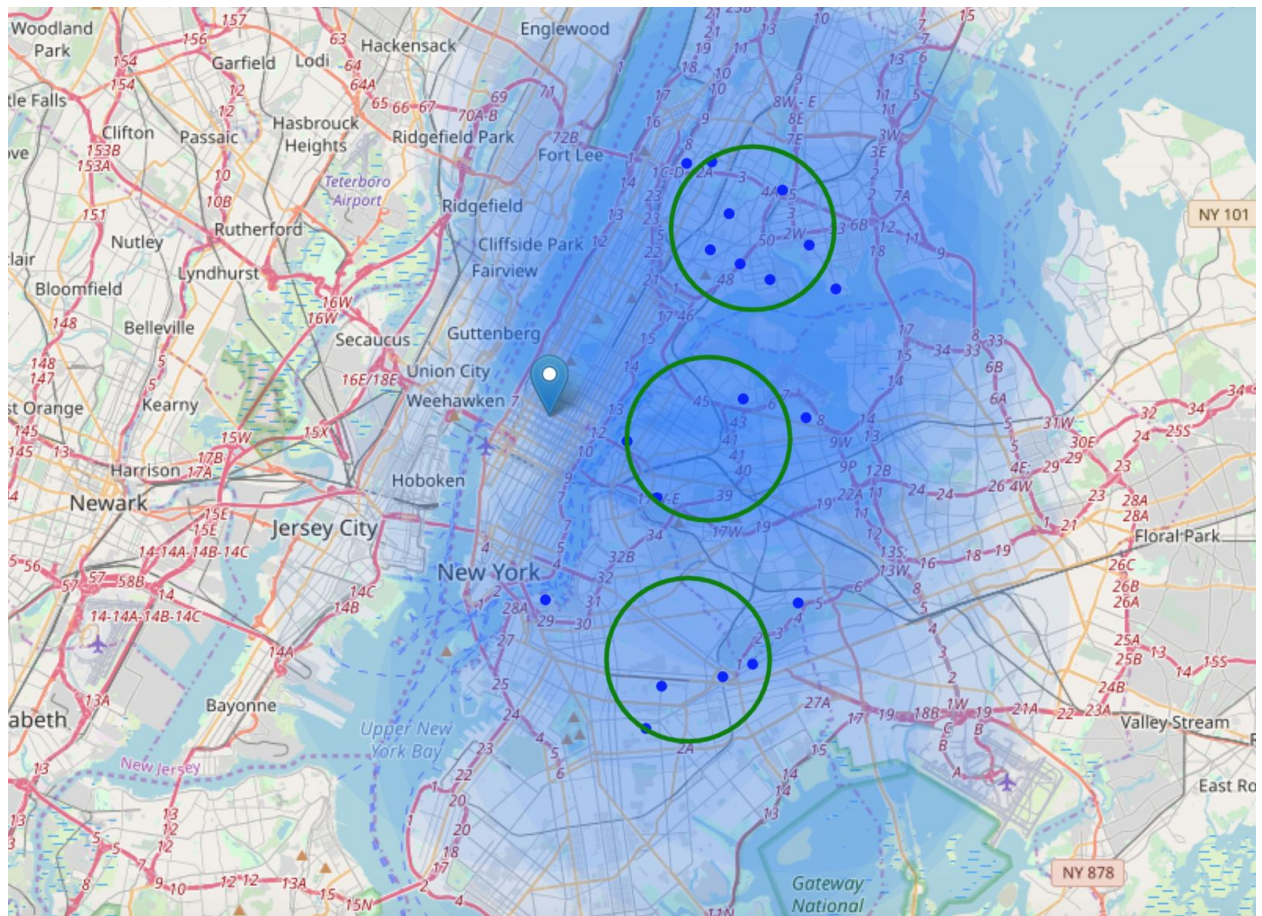
What we have now is a clear indication of zones with low number of restaurants in the vicinity, and no Japanese restaurants at all nearby. Let us now **cluster those locations** to create centers of zones **containing good locations**. Those zones, their centers and addresses will be the final result of our analysis. To do this We chose to use the **KMeans Clustering algorithm**. The following map illustrates the 3 clustering-zones in green color:



Our clusters **represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.**

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:



Finally, we compute the addresses which can be presented to stakeholders:

	Borough	Neighborhood	Latitude	Longitude	Distance from center	X	Y	Restaurants in area	Distance to Japanese restaurant
0	Bronx	Morris Heights	40.847898	-73.919672	12948.680744	4.649537e+06	5.861096e+06	5	1776.801415
1	Bronx	West Farms	40.839475	-73.877745	14657.713936	4.653661e+06	5.862787e+06	3	4841.442198
2	Bronx	Melrose	40.819754	-73.909422	10468.798950	4.652674e+06	5.858693e+06	5	1400.285479
3	Bronx	Longwood	40.815099	-73.895788	11076.340286	4.654173e+06	5.859036e+06	4	2930.439232
4	Bronx	Hunts Point	40.809730	-73.883315	11701.258801	4.655634e+06	5.859230e+06	3	4404.696854
5	Bronx	Soundview	40.821012	-73.865746	14047.736309	4.656154e+06	5.861538e+06	4	5706.396957
6	Bronx	Clason Point	40.806551	-73.854144	14152.526982	4.658287e+06	5.860697e+06	3	6486.965201
7	Queens	East Elmhurst	40.764073	-73.867041	11590.158567	4.660700e+06	5.855308e+06	5	2317.290994
8	Queens	Glendale	40.702762	-73.870742	13929.986305	4.665403e+06	5.848451e+06	1	1580.155182
9	Brooklyn	Wingate	40.660947	-73.937187	14592.365437	4.663341e+06	5.839801e+06	4	1096.176738
10	Queens	Astoria Heights	40.770317	-73.894680	8788.267505	4.657919e+06	5.854267e+06	4	1132.379507
11	Bronx	Claremont Village	40.831428	-73.901199	12204.943544	4.652396e+06	5.860463e+06	5	2196.262824
12	Bronx	Mount Hope	40.848842	-73.908299	13652.583465	4.650392e+06	5.861903e+06	5	2915.432846
13	Queens	Blissville	40.737251	-73.932442	6039.158156	4.657513e+06	5.848350e+06	2	1161.416044
14	Brooklyn	Vinegar Hill	40.703321	-73.981116	8182.565208	4.656271e+06	5.841664e+06	5	784.489451
15	Brooklyn	Weeksville	40.675040	-73.930531	13034.401620	4.662741e+06	5.841739e+06	4	1564.329462
16	Brooklyn	Broadway Junction	40.677861	-73.903317	14039.915848	4.664753e+06	5.843734e+06	5	4065.614978
17	Brooklyn	Highland Park	40.681999	-73.890346	14394.657482	4.665484e+06	5.844987e+06	4	3536.257961
18	Queens	Queensbridge	40.756091	-73.945631	3649.427146	4.654893e+06	5.849570e+06	2	801.863190


This concludes our analysis. We have created **19 addresses** representing centers of zones containing **locations with low number of restaurants and no Japanese restaurants nearby**, all zones being fairly **close to NYC Center** (all less than 10 km from NYC Center). Those zones are identified in **Queens, The Bronx and Brooklyn** boroughs, which we have identified as interesting due to being popular with tourists, fairly close to the city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a great number of restaurants in NYC (~1400 in our initial area of interest which was 15x15km around NYC Center), there are pockets of low restaurant density fairly close to the city center. Highest concentration of restaurants was detected south from NYC Center, so we focused our attention to areas East and North, corresponding to boroughs Queens, The Bronx and Brooklyn which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest, we first filter locations inside those boroughs so that those with more than five restaurants in radius of 500m and those with an Japanese restaurant closer than 500m were removed.

Result of all this is 19 zones containing the largest number of potential new restaurant locations based on number of and distance to existing venues. This, of course, does not



imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to NYC Center but not crowded with existing restaurants (particularly Japanese) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for a more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

The goal of this project is to identify NYC areas close to center with low number of restaurants (particularly Japanese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Queens, Brooklyn and The Bronx), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.