Predicting a suitable location for a Chinese Restaurant in Queens, New York, USA

Adeola Ayandeyi
May, 2020

1. Introduction

1.1 Background

The world has become a global village and with this comes diversification of cultures. Different cultures from different regions of the world have and still mixing up, helping people understand each other and their perspective. This has birthed the exploration of different cultures by different people and this include work ethics, food, fashion, entertainment, music and general way of life. Food is critical to human health and this diversification has made people from one culture try out food of other cultures. United States of America is one of the countries of the world that is diversified. Different people from different cultures lives here and this has made exploration of cultures in that countries very significant. For instance, the craving for Chinese food is on the rise and this is as a result of culture diversification. Chinese foods are not being taken only by the Chinese in the US but also other people from different cultural background. This is so because of the unique blend of their meals and of course healthy. Although there a lot of restaurants in New York, the project focus on the suitable location of a new Chinese restaurant in Queens, New York. Some important characteristics of this location is that it has fewer restaurants within 250m and no Chinese restaurants within 400m. Suitable locations will be extracted and based on other factors, the stakeholders will determine the best location to plant a new Chinese restaurant in Queens, New York.

1.2 Problem

This project aims to predict locations where to set up a new Chinese restaurant in Queens neighborhood where there is none within 400m. Other factors will be used to determine which exact location to focus on.

1.3 Interest

The stakeholders will be interested in the accurate prediction of the location for a new Chinese restaurant for competitive advantage and business values. Others that will be interested are people of the neighborhood having a Chinese restaurant in their locality.

2. Data

2.1 Data sources

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

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Python library geopy**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of New York center will be obtained using **Google Maps API geocoding**

2.2 Data Collection

We created the latitude & longitude coordinates for centroids of our candidate neighborhoods. We also created a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around New York city center. We also created the grid of area candidates, which will be spaced equally and within 4km from Queens. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.
To accurately calculate distances we created our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we projected those coordinates back to latitude/longitude degrees to be shown on Folium map. We created a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all its neighbors**. 301 neighborhoods centers within ~4km from Queens were retrieved. These locations are visualized below:

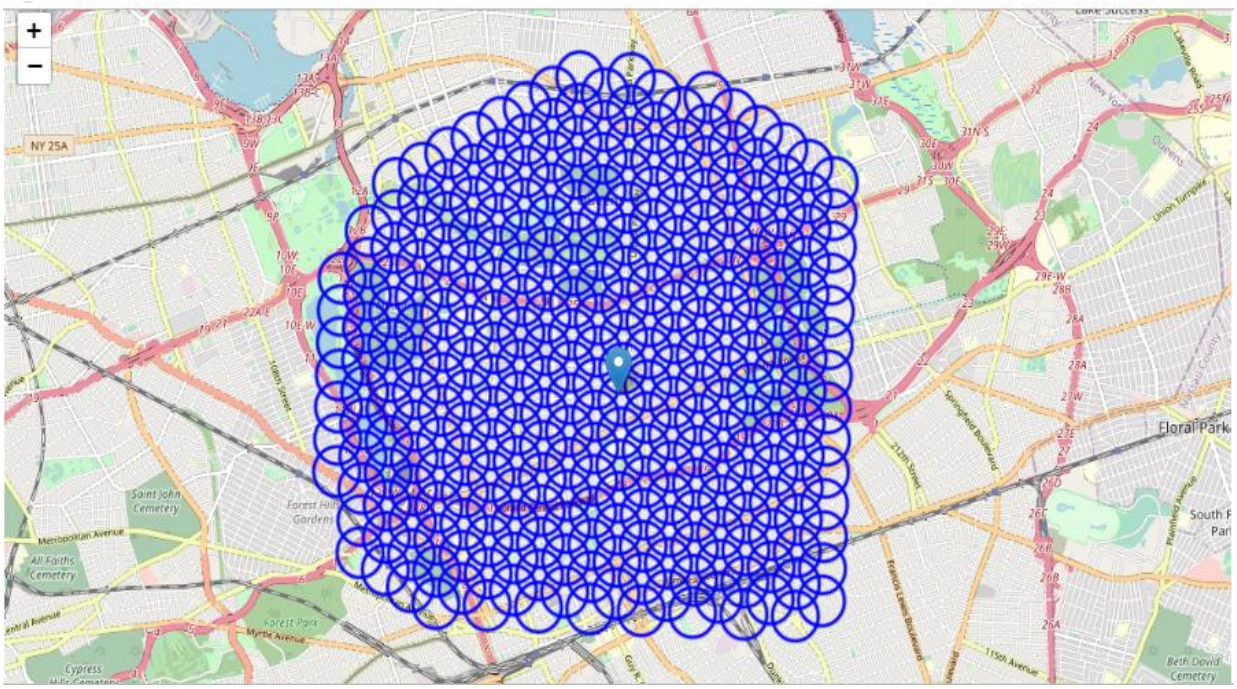

Figure 1. 301 Neighborhood centers ~4km from Queens

With Google API, the addresses of these locations were retrieved and a dataframe was created with the following features: Addresses, Longitudes, Latitudes, X and Y coordinates and the distance from the center.

After generating the location candidates, we used Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants. We included in out list only venues that have 'restaurant' in category name, and we made sure to detect and include all the subcategories of specific 'Chinese restaurant' category, as we need info on Chinese restaurants in the neighborhood.

- The following information were collected:
- Total number of restaurants collected is 485.
- Total number of Chinese restaurants is 17.
- Percentage of Chinese restaurants is 3.51%.
- Average number of restaurants in neighborhood is 3

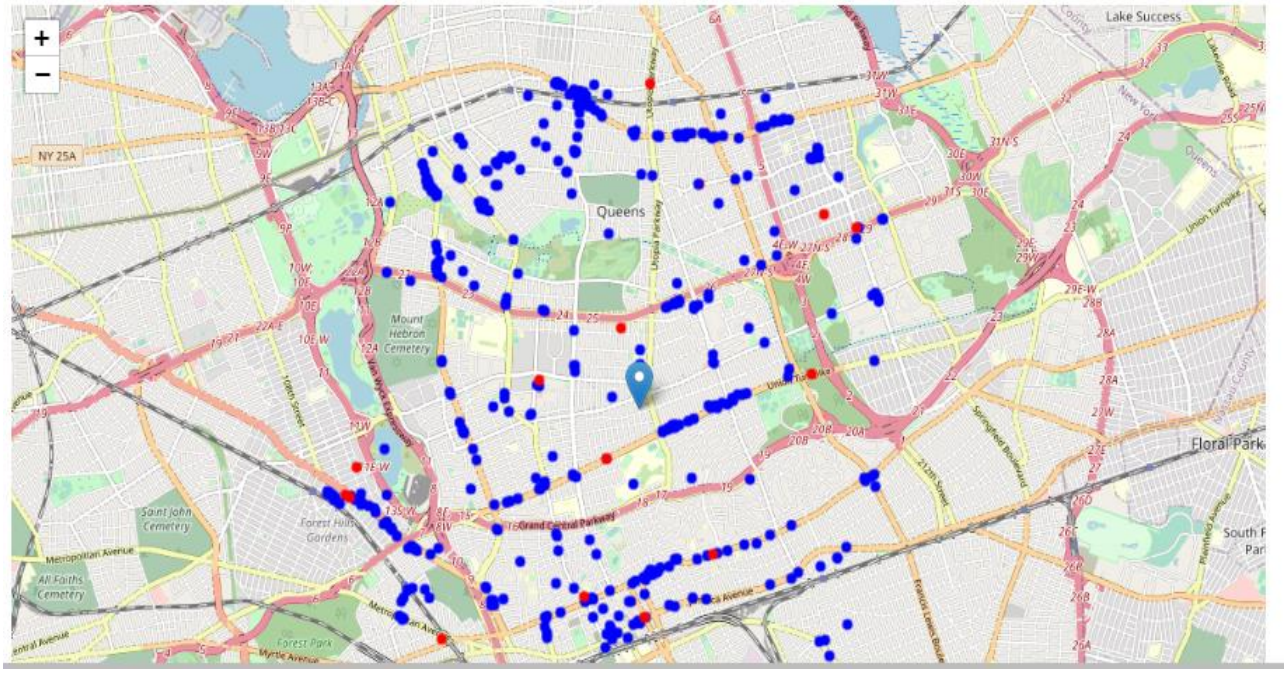

Figure 2: Collected Restaurants in the neighborhood (Chinese restaurants marked red, others, blue)

3. Methodology

In this project we will directed our efforts on detecting areas of Queens, New York that have low restaurant density, particularly those with low number of Chinese restaurants. We limited our analysis to area ~4km around city center.

In first step we collected the required data: location and type (category) of every restaurant within 4km from Queens, New York center. We also identified Chinese restaurants

In the second step of the analysis,we calculated and explored 'restaurant density' across different areas of Queens - we used heatmaps to identify a few promising areas close to center with low number of restaurants in general (*and* no Chinese restaurants in vicinity) and focused our attention on those areas.

In the third step we focus on most promising areas, created clusters of locations that meet some basic requirements established in discussion with stakeholders. We also considered locations with no more than two restaurants in radius of 250 meters, and we wanted locations without Chinese restaurants in radius of minimum of 400 meters. We presented the map of all such locations but also created 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.

3.1 Exploratory Analysis

After performing some basic explanatory data analysis and derive some additional info from our raw data, first, we counted the number of restaurants in every area candidate and we noticed that the average number of restaurants in every area with radius of 300m is approximately 3. We also calculated the distance to nearest Chinese restaurant from every area candidate center and **the** average distance closest Chinese restaurant from each area center is approximately 1386. This means that Chinese restaurant can be found within approximately 1386m from every area center candidate. The visuals below show the density of restaurants around the neighborhood.

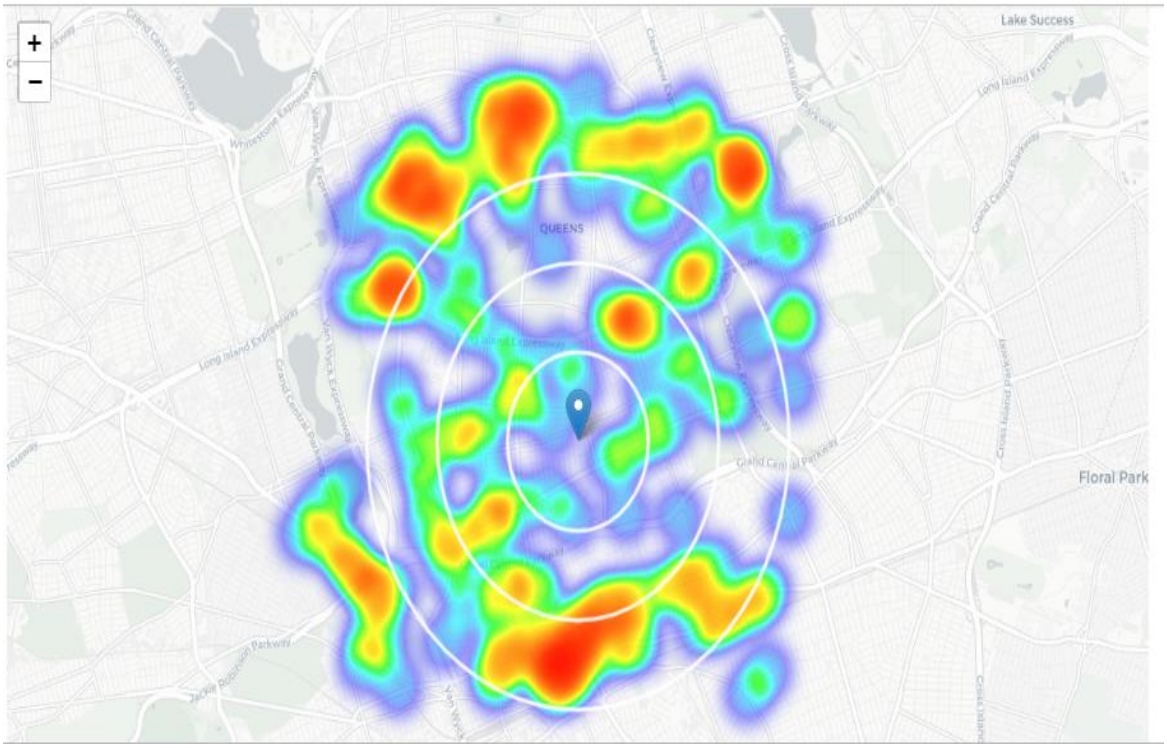

Figure 3: Density of all the restaurants around Queens

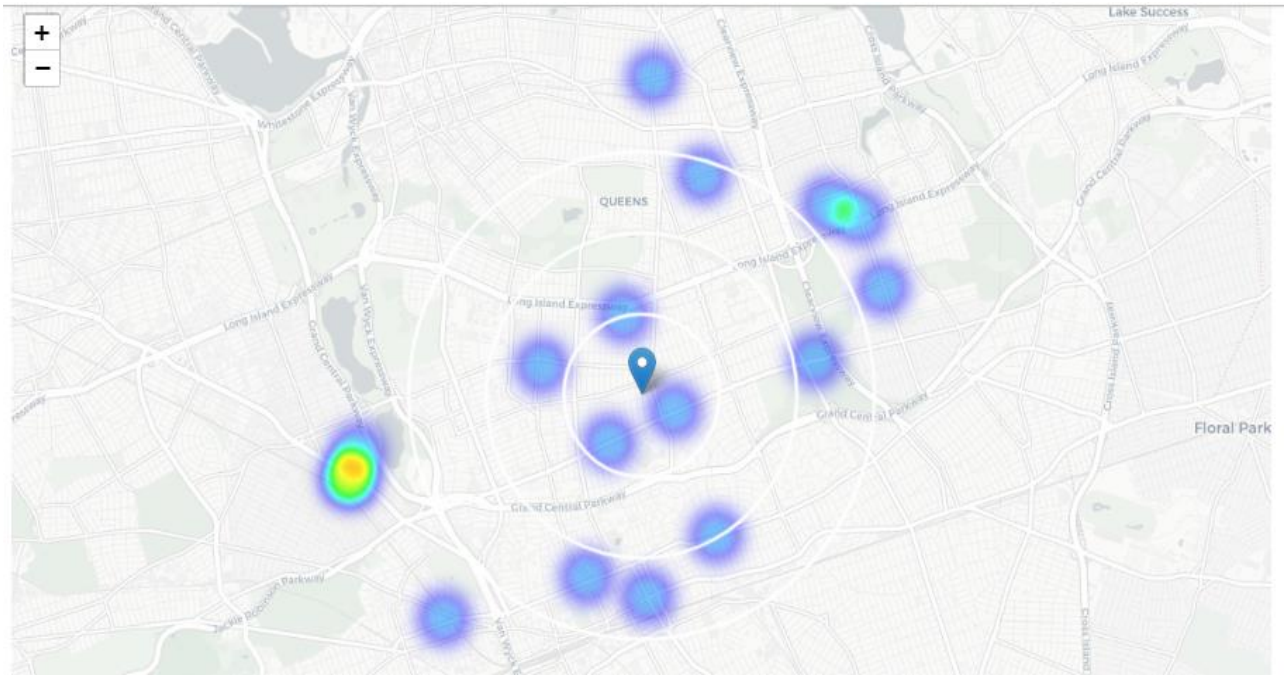

Figure 4: The density of Chinese restaurants only around Queens

Detecting the number of restaurants that are nearby with minimum distance of 3km, we were able to extract the number of nearby restaurants per location and also the distance to a Chinese restaurant. Based on this information, we filtered those locations: we're interested only in which are locations with no more than two restaurants in radius of 250 meters**,** and no Chinese restaurants in radius of 400 meters. The following information were extracted:

- Locations with no more than two restaurants nearby is 269
- Locations with no Chinese restaurants within 400m is 281
- Locations with both conditions met: 255

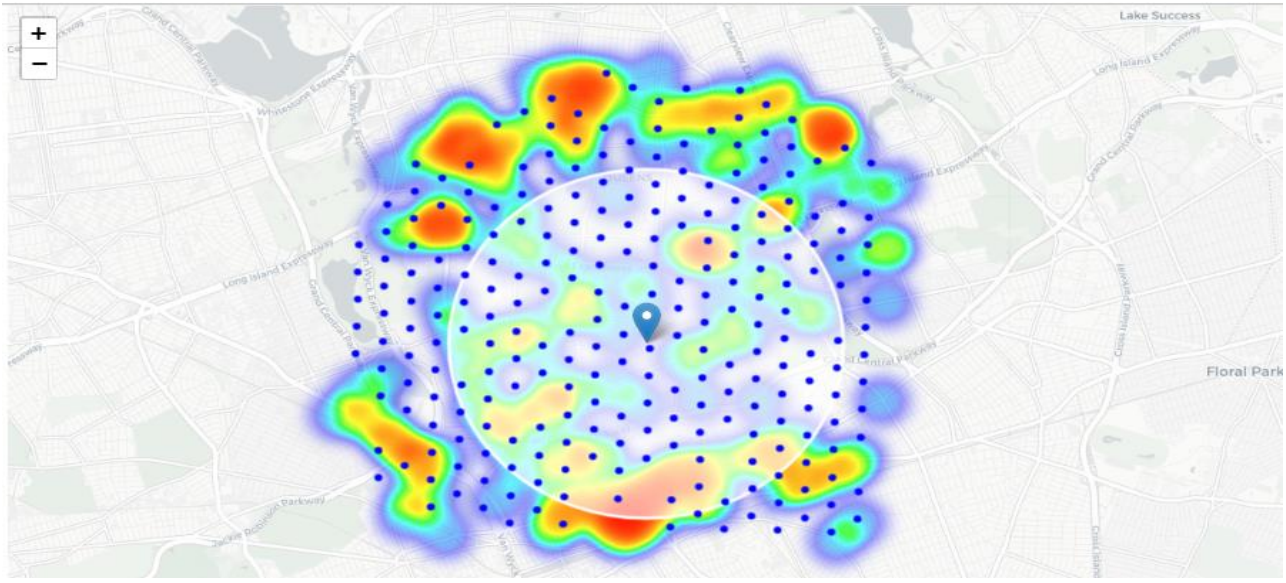

Figure 5: Locations with no more than 2 restaurants in radius of 250m and no Chinese restaurants closer than 400m

Locations on the map above has

- no more than two restaurants in radius of 250m,
- no Chinese restaurant closer than 400m.

Any of those locations is a potential candidate for a new Chinese restaurant

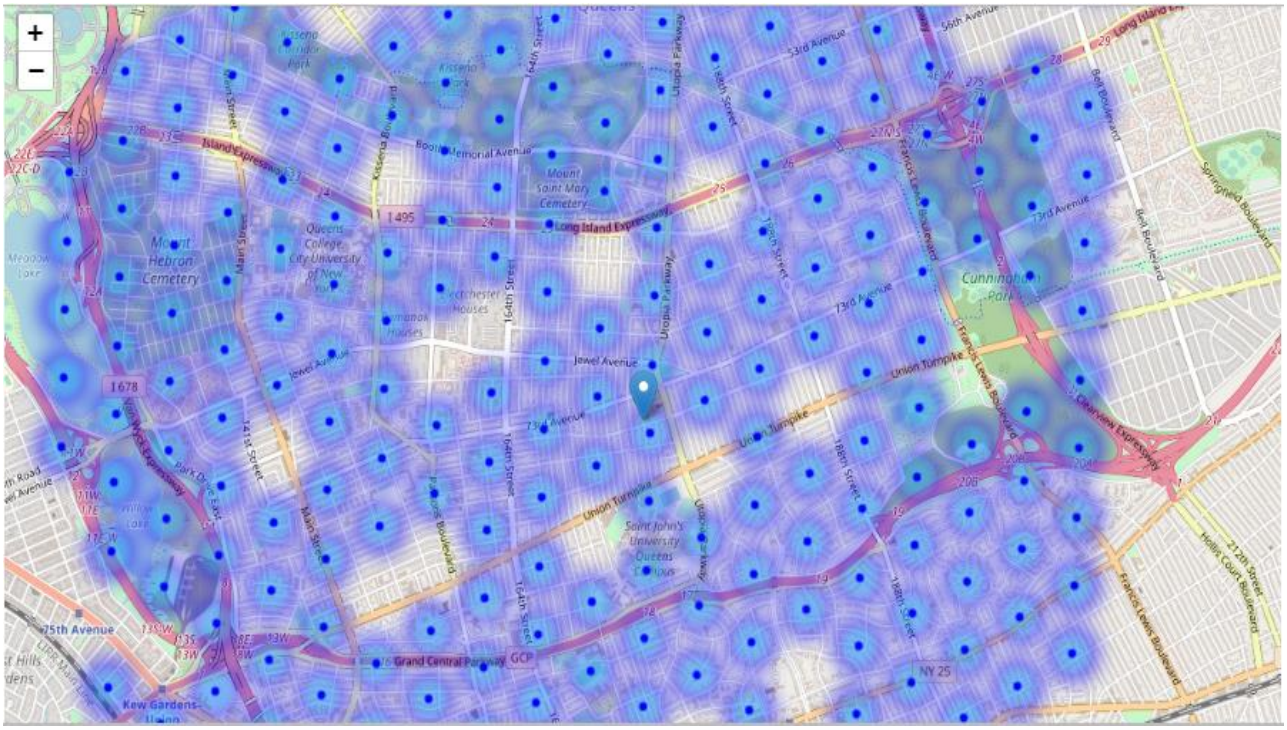

Figure 6: Heatmap visualization of suitable locations

Visualizing these locations with a heatmap, we have a clear indication of zones with low number of restaurants in vicinity, and no Chinese restaurants at all nearby

4. Predictive Modeling (KMeans Clustering)

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms which identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

The model, KMeans clustering was used to cluster those locations and created centers of zones containing good locations. Restricting the number of clusters to 15, the map below shows the clusters which represent the groupings of most of the candidate locations and the cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

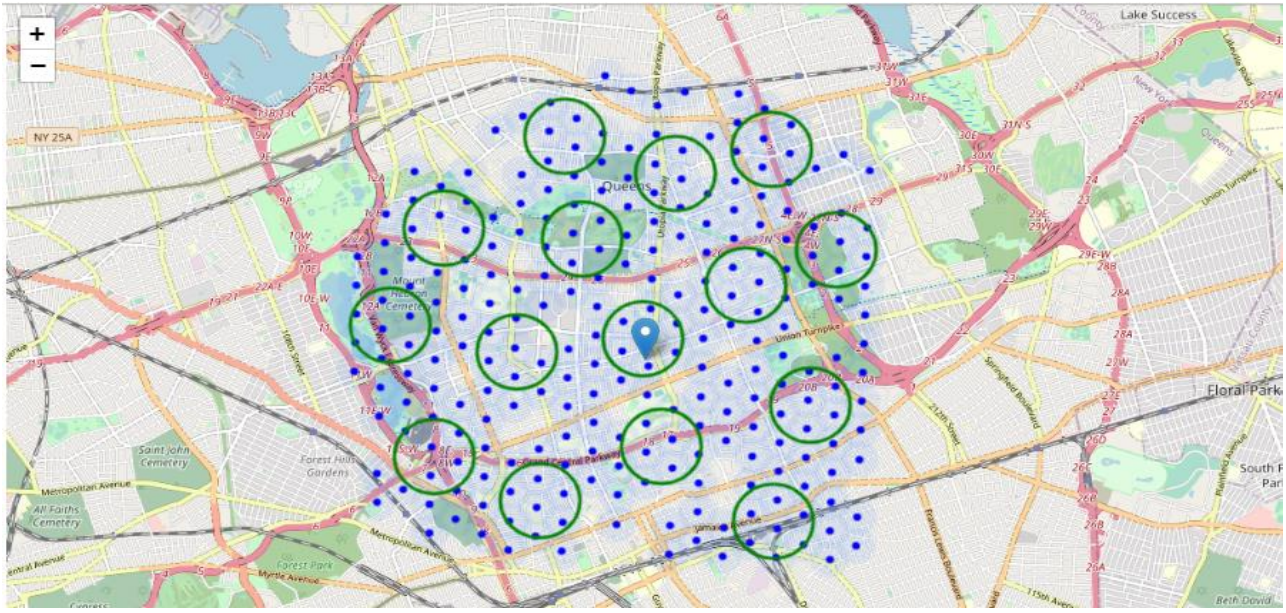

Figure 7: Clusters of centers of zones containing good locations

A good place to start exploring the neighborhood is the addresses of these clusters so as to know which is best suited for a new Chinese Restaurant.

With **reverse geocode, we retrieved the addresses** of **those candidate area centers** which can be presented to stakeholders. The following are the addresses:

154-44 71st Ave, Flushing, NY 11367 => 2.5km from Queens
86-31 Sancho St, Jamaica, NY 11423 => 3.3km from Queens
Grand Central Pkwy, Jamaica, NY 11435 => 4.5km from Queens
46-61 189th St, Flushing, NY 11358 => 3.9km from Queens
172-28 83rd Ave, Jamaica, NY 11432 => 1.8km from Queens
18402 Jamaica Ave, Queens, NY 11423 => 4.2km from Queens
67-1 210th St, Oakland Gardens, NY 11364 => 4.4km from Queens
164 St/Booth Memorial Av, Queens, NY 11355 => 2.8km from Queens
14453 85th Ave, Briarwood, NY 11435 => 3.5km from Queens
45-28 160th St, Flushing, NY 11358 => 4.9km from Queens
174-20 Jewel Ave, Flushing, NY 11365 => 0.5km from Queens
140-15 58th Rd, Flushing, NY 11355 => 4.8km from Queens
Unnamed Road, Flushing, NY 11367 => 5.0km from Queens
204-27 47th Rd, Flushing, NY 11361 => 5.0km from Queens
67-21G 193rd Ln, Fresh Meadows, NY 11365 => 2.5km from Queens

The 15 addresses shown above represent the centers of zones containing locations with low number of restaurants and no Chinese restaurants nearby, all zones being fairly close to city center. The centers/addresses of these zones (the green circles on the map below) should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.

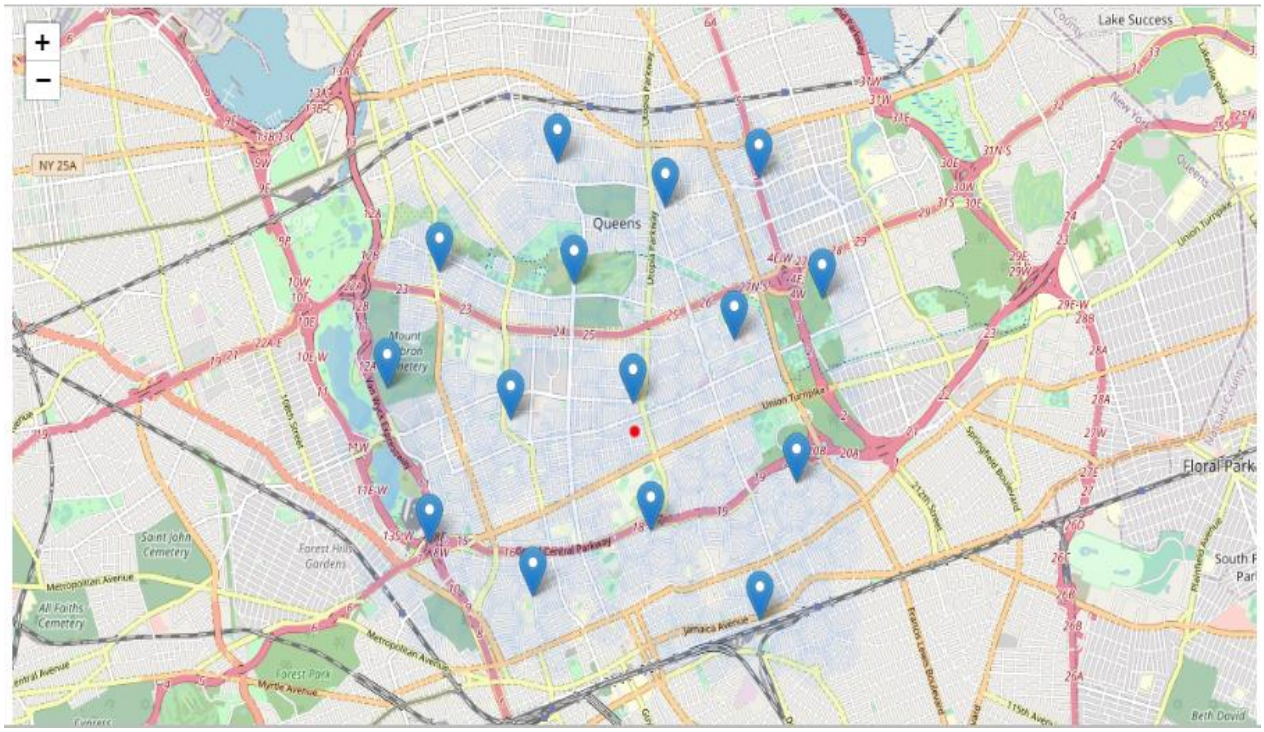

Figure 8: Location of the 15 addresses presented to stakeholders

5. Results and Discussion

Our analysis shows that although there is a great number of restaurants in New York, there are pockets of low restaurant density fairly close to the city centre.

After directing our attention to areas with low restaurant in Queens, we first created a dense grid of location candidates; those locations were then filtered so that those with more than two restaurants in radius of 250m and those with any Chinese restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as starting points for more detailed local analysis based on other factors.

Result of all these are 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Chinese restaurants particularly. 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 New York center but not crowded with existing restaurants (particularly Chinese) - 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 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.

6. Conclusion

Purpose of this project was to identify New York areas close to center with low number of restaurants (particularly Chinese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Chinese restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified Queens borough that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created 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 (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.