Capstone Project - The Battle of Neighborhoods

Chinese Restaurant Opening in New York City

Jasmine Liu

May, 2020

1. Introduction

1.1. Background

As the most famous city in the world, New York City is seen as a business hub with a variety of people. Whilst being a business district and hub, New York is in a major hunt to ensure people have the opportunity to enjoy the worldwide food. In the meantime, Chinese food has been very popular over the western world. Hotpot, as one of the most Chinese cooking method, has already become the major attraction across North America.

1.2. Problem

A traditional Chinese hot pot group is looking to open a flagship restaurant in one of New York's neighborhoods. This will be the first restaurant of that hot pot brand in the US. It does not understand the area and the availability of Chinese restaurant in each neighborhood and requires an investigation to take place in order to determine the best place for the flagship.

1.3. Decision Metrics:

The following decision metrics are requested in order to make an informed decision by the hot pot group to where the restaurant should be located.

1. Density of people for each Borough
2. Number of Neighborhoods in each Borough
3. Number of Chinese restaurants in each Borough
4. Chinese restaurant in the Neighborhood with the best density metric per restaurant
5. Cluster Chinese restaurant in Neighborhood with the best density metric per restaurant

2. Data Requirements and sources

2.1. Data Sources

For the investigation, the following data sources will be used:

1. Wikipedia to obtain density of each Borough in New York city.
    a. Source: https://en.wikipedia.org/wiki/New_York_City

   b. Description: New York Boroughs and the density of each Borough in the New York area.

2. New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
   a. Source: https://cocl.us/new_york_dataset
   b. This contains the data as mentioned above and will be used for investigating the Borough and Neighborhoods using Foursquare API.

3. Chinese restaurant in each neighborhood of New York city.
   a. Source: Foursquare API
   b. The API will return all known Chinese restaurant in each Borough and Neighborhood.

2.2. Data cleaning

During the data extraction process from the various sources, it is important to validate and verify clean data to work with throughout the analysis. Each data source required validations and visual confirmation to whether there are discrepancies in the data we received/sourced.

It was vital that these data feeds had validation to ensure the accuracy of the results obtained. Each feed/data source went through its own unique data extractions and validation based on the source of the data. The following was done to each data source to ensure a robust approach in terms of the analysis and the comparing of data sources.

2.2.1 Wikipedia

The data source to obtain density, populations and other key metrics for New York city came from a Wikipedia page. The page required extracting as well as validation on the extraction of the data as seen in the code. It was of utmost importance to ensure the correct extraction criteria occurred to ensure the validity of the data. In this process it was noted that the naming convention of the Boroughs were not the same as the other feeds and this required adaptation to ensure linking to other data sources was seamless and accurate.

2.2.2. New York City Data

This data source was important to obtaining the data for borough and neighborhoods in the New York city area. We required this data to validate correct geo locations could be sources for the foursquare API. This data source was robust and required minimal changes to align with the other datasets.

2.2.3 Foursquare data for Chinese Restaurant

This data set required a lot of cleaning in order to ensure only Chinese Restaurant information was extracted for the analysis in New York City. The cleaning of the data required all other “noise” to be dropped and that only valid data of restaurants was required. This process of Chinese Restaurant validation required the checks to ensure the Chinese Restaurant we in the correct mapping location based on the New York city Boroughs and ensuring the data received was recent and not old.

3. Methodology and explanation of Data analysis.

In order to answer the various decision metric as stated in 1.3 above, it was vital to ensure each decision metric was fully analyzed and a decision could be made. Thus, it is reporting below is the analysis to obtaining each result as well as the data analysis done for each.

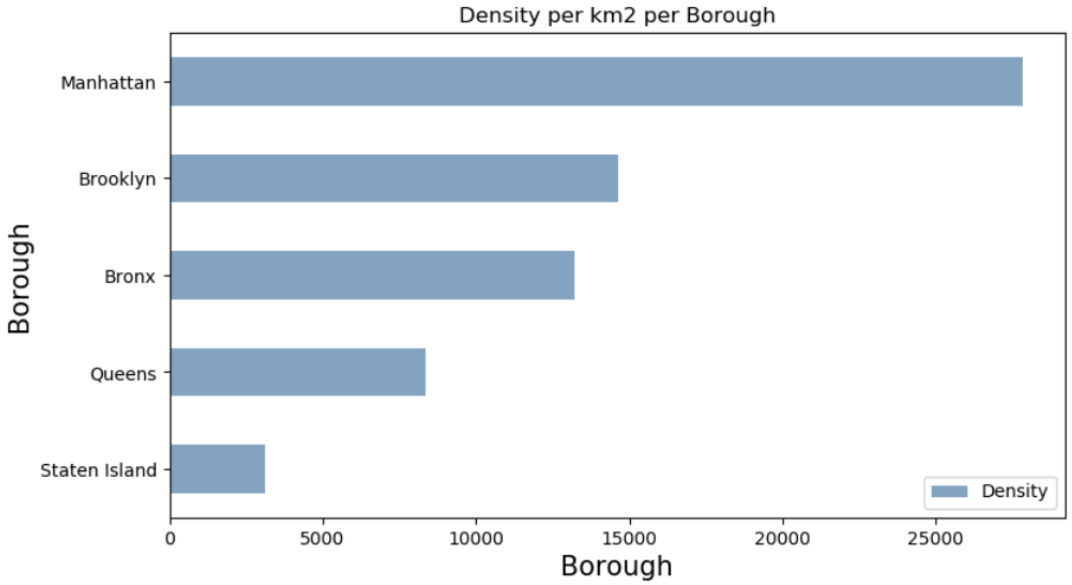
3.1. Density of people for each Borough

The density of each borough in New York city was achieved through the extraction of data from the Wikipedia page as stated in the data source section 2.1 above. With the various scrapping of data from the source the following data could be extracted (Table 1).

	Borough	Density
0	Bronx	13231
1	Brooklyn	14649
2	Manhattan	27826
3	Queens	8354
4	Staten Island	3132

Table 1: Density per Borough in New York City

The density of each borough helps to visualize the density of people per borough and later will be used to ensure the density of people per restaurant that is currently in each area. With targeting the area with the best density will allow for more potential customers. The graph below helps to visually see the major difference in the boroughs based on density.



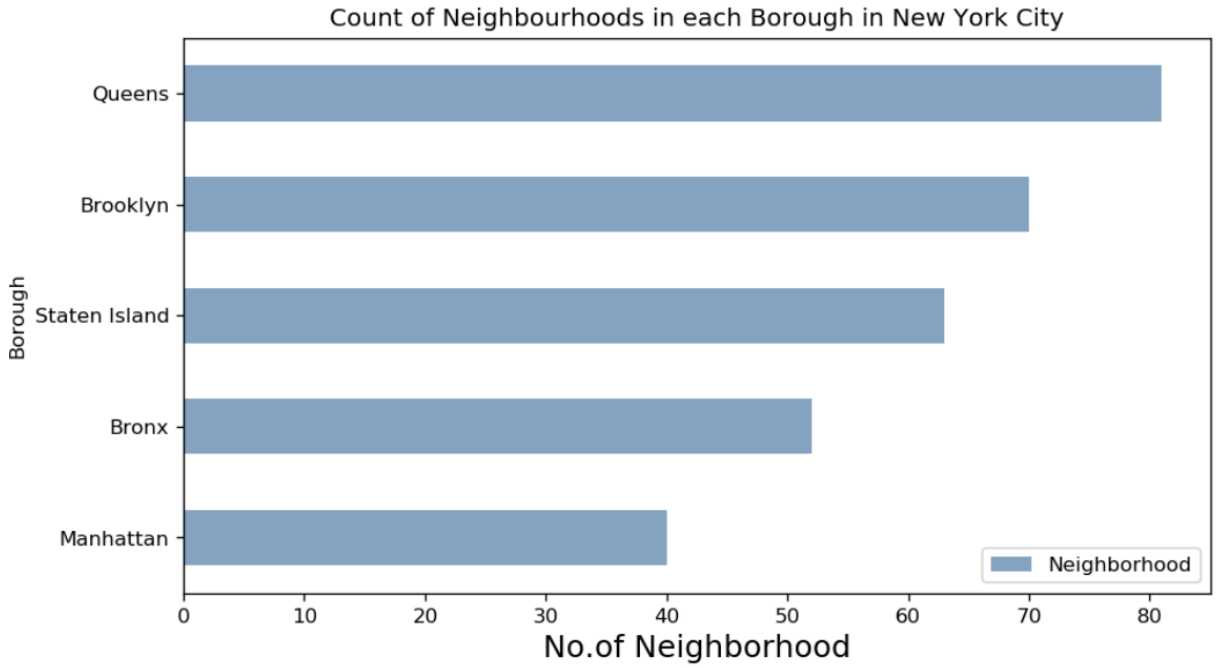
Graph 1: Bar graph of density of people in each borough of New York City.

With graph 1 above, it is easily noticeable that Manhattan has the most people per km2 of any borough. This shows population density that can be used in further analysis of population density per restaurant in each borough.

3.2. Number of Neighborhoods in each Borough

The next important metric was to ensure that number of neighborhoods in each borough. This decision metric aids in the density to determine the spread of the population density. This decision metric aids the ability to open a Chinese restaurant with few neighborhoods to ensure a greater buy in when opening a new Chinese restaurant in an area.

The following was obtained through the New York city data as seen in section 2.2.2 above. This was used to graph the neighborhoods in each borough as seen in table 2 below.

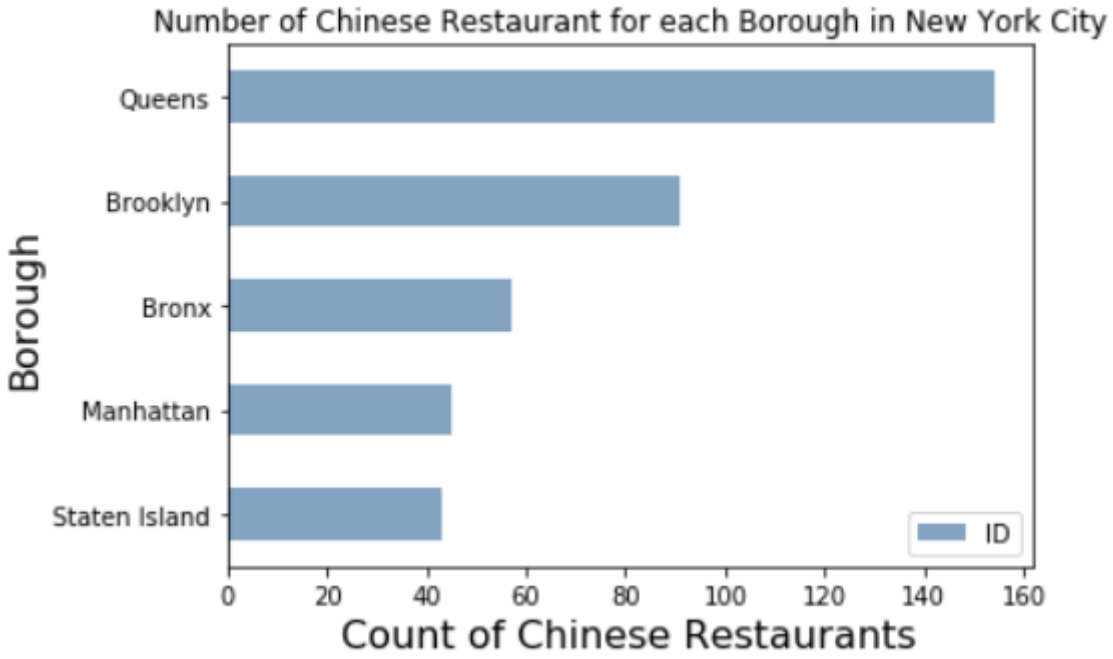


Graph 2: Count of Neighborhoods in each borough.

As seen above, the Manhattan borough has the fewest number of neighborhoods whilst potentially having the highest population density as seen in section 3.1. Both allow for valuable insight in decision making further on.

3.3. Number of Chinese restaurants in each Borough

The next metric was to validate and source the number of Chinese restaurants per borough. This will give valuable insight of competitors and the location of the competitors. When extracting data from Foursquare the follow number of Chinese restaurants were noted in each area.

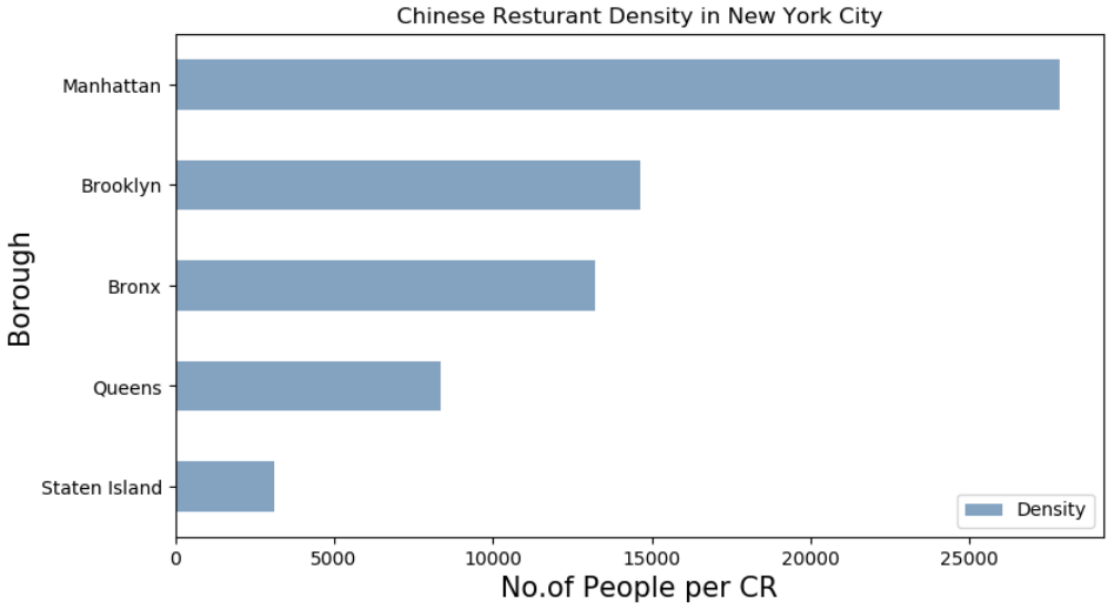


Graph 3: Count of Chinese restaurant per borough

As seen above, there is major difference in the count of Chinese restaurant in each borough. This helps us to see a potential gap in the market of the other boroughs whilst we use this data to determine the borough with the maximum density of people per restaurant. This can also be valuable insight in determining a rating factor for decision making for further analysis on best location.

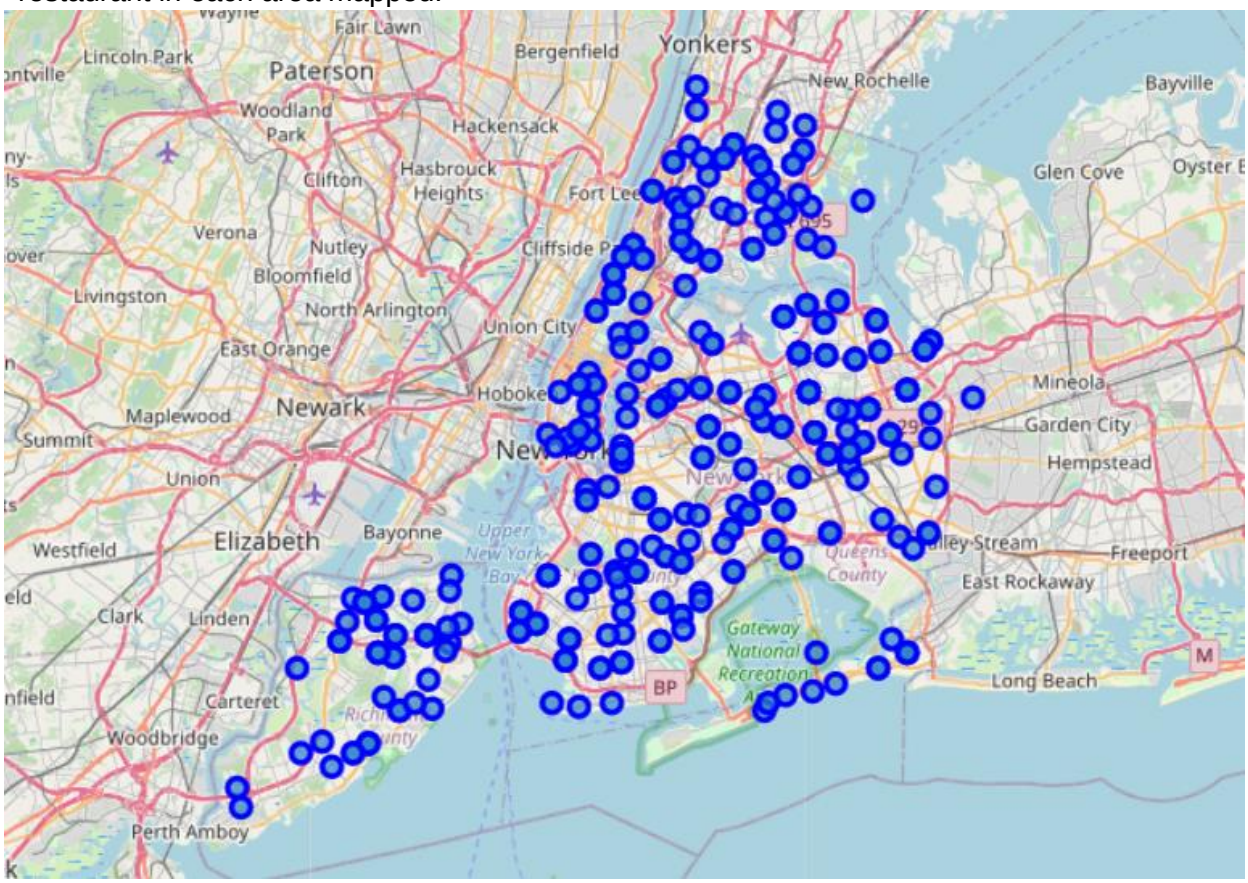
3.4. Chinese restaurant in the Neighborhood with the best density metric per restaurant

With all the above complete, a combination of each result was needed to identify a local best location. With merging data and evaluation, the boroughs per density and restaurant count. It was discovered that Manhattan still produced the highest density of people per restaurant. This statistic goes against the data represented above if seen in isolation. Thus, was of utmost importance that the combination as seen below in graph 4 was designed to evaluate each borough in the same manner. This allows for opportunity not to be weighed heavier toward one borough creating a bias in the data analysis.



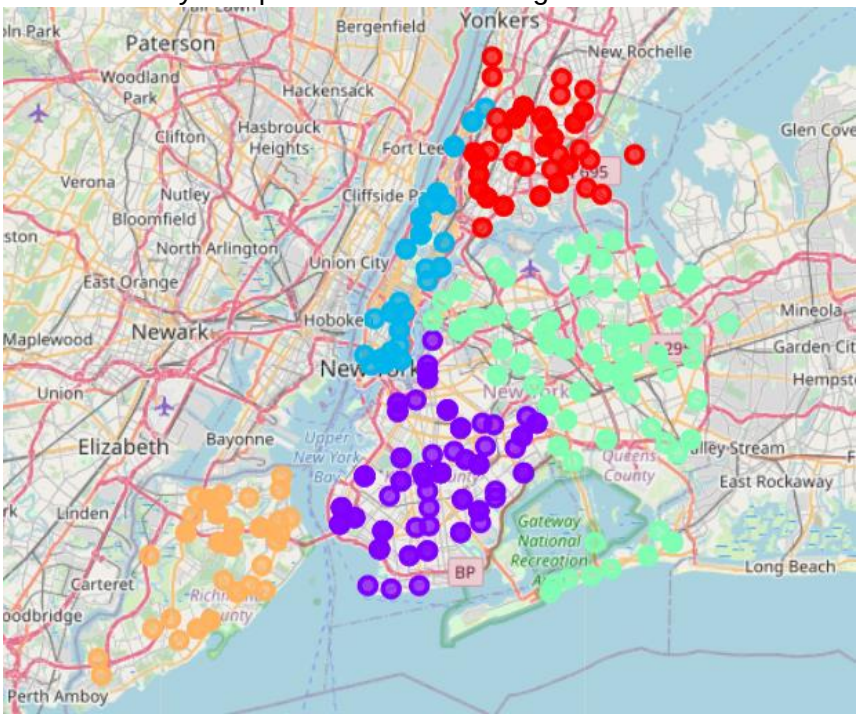
Graph 4: Number of people per Chinese restaurants in each borough

It was also important to see the spread of Chinese restaurant across the entire network of Chinese restaurant to check fair distribution and not over population of the Chinese restaurant in a localized area. The following map displays the density of Chinese restaurant in each area mapped.



Map 1: Density of Chinese restaurant across New York to evaluate the overcrowding of a localized area.

With the above seen, it was more important to visualize the borough density individually by clustering them based on their borough. This allowed for further visual investigation of the borough's restaurant density and possible overcrowding of Chinese restaurant.



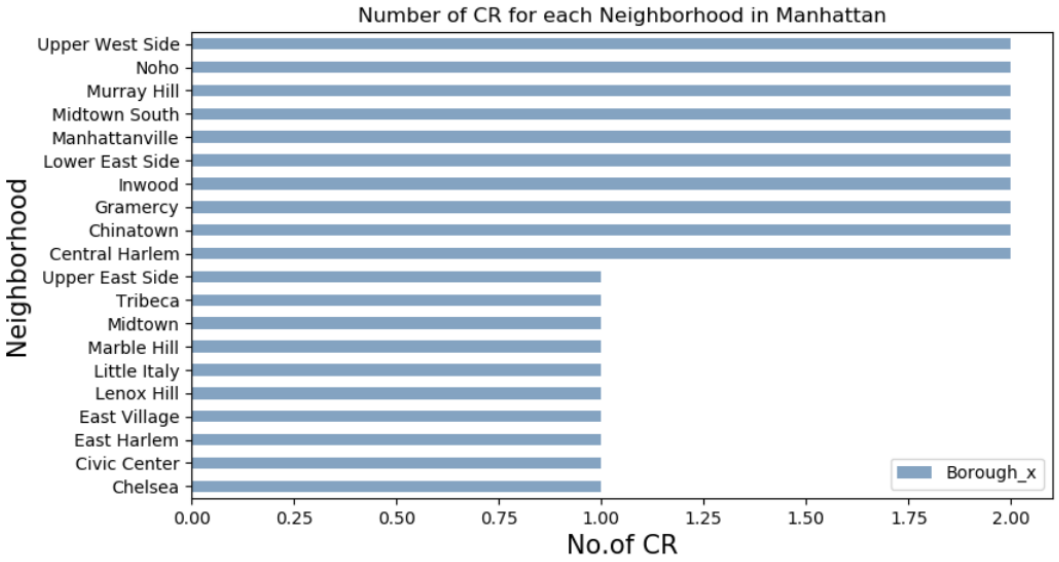
Map 2: Density of Chinese restaurant across New York to evaluate the overcrowding of a localized area clustered into boroughs

As seen above, the special density of the Chinese restaurant illustrated in map 2 above, show that all the Chinese restaurant are evenly spread. This allows for a normalization of the data to determine the best borough and thus looking into neighborhood potential further.

3.5. Cluster Chinese restaurant in Neighborhood with the best density metric per restaurant

With the above taken into consideration, it is noted that the best potential for the restaurant franchise was to evaluate the neighborhoods in Manhattan based on population density per restaurant as well as restaurant spread in the Manhattan area.

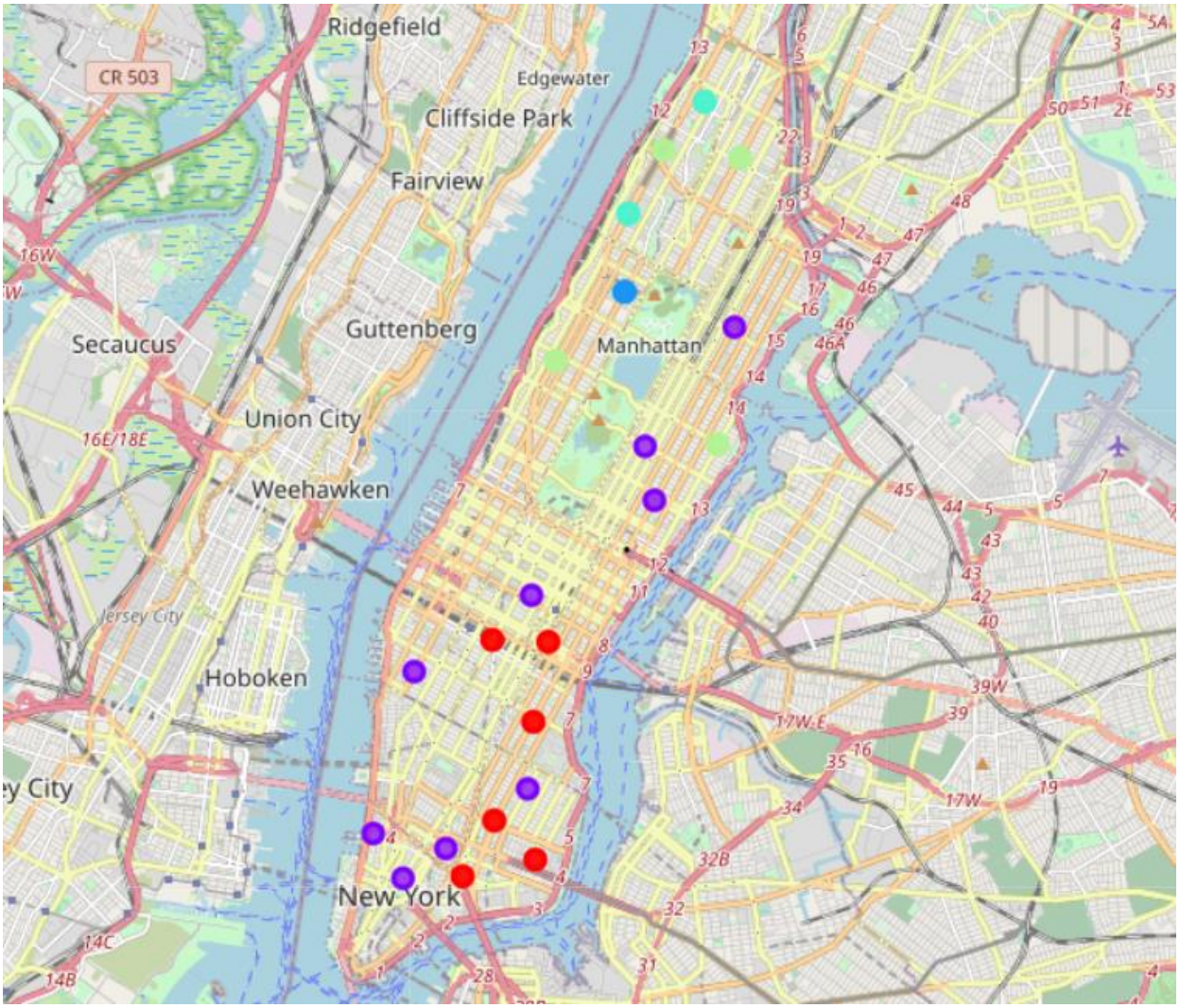
The count of Chinese restaurant in each neighborhood is required to ensure a full picture of the neighborhood competitors is evaluated. With the location data of each restaurant, it is possible to normalize the data to view the count of Chinese restaurant per neighborhood.



Graph 5: Chinese restaurant per neighborhood in Manhattan

With the above information (Graph 5), the data was used to determine restaurant spread and clustering of the neighborhoods to evaluate each neighborhood on the merit of restaurant capacity and population density.

The top restaurant ratings were used to evaluate a visual representation of the Chinese restaurant in the Manhattan area. The following was observer in map 3 below.



Map 3: Best Chinese Restaurants ratings in Manhattan clustered for density spread of the Chinese restaurant.

With confirmation of spread across the neighborhoods being equal, it can be determined that evaluating the number of Chinese restaurants in each neighborhood would be a fair analysis as this is a normalized data set.

4. Results Discussion

With the above decision metrics being consider the following was observed in the data. The population density of each borough is majorly skewed with some boroughs having a much higher population density that other. The Chinese restaurant per a borough was vital for to determine the possibility of the customers joining the restaurant. The ratings of each restaurant ensured a robust approach to determining the impact of opening a restaurant when competing with some of the best Chinese restaurant in the area.

Finally, the spread of the Chinese restaurant aided in deciding if there is a bias in the exact location of the Chinese restaurant in each borough to ensure the robustness of the analysis driving the decision is not skewed with overpopulated areas of restaurant with the best rating.

The neighborhoods in Manhattan also allowed for the full understanding of Chinese restaurant in each neighborhood whilst ensure again the spread of the best-known Chinese restaurant across the neighborhood was not skewed to overcrowding.

5. Conclusion

With all the above taken into consideration, it can be sent that Manhattan has the best chance of making money when opening a restaurant. This is confirmed with the density of people per restaurant in the Manhattan area. The spread of Chinese restaurant also ensured that there is not a skew/ bias in the data and that the Chinese restaurant with the best rating are well spread and determination of the

neighborhood could be based on the neighborhood restaurant count. With this it is determine that a restaurant in any neighborhood with Chinese restaurant less than 3 can be opportunity to open a restaurant. There are multiple neighborhoods that fall within this criteria, and final decision can be made on the merit of costs of location and logistics in these areas. Refer to graph 5 above, for further understanding of the possible neighborhoods.

6. Future Decision

Whilst the analysis is robust, factors specific to the franchise such as cost of location, logistics and other will need to be considered for the exact neighborhood to choose in Manhattan with restaurant count below 1. The availability of space to rent or buy will also need to be incorporated to ensure the decision has the lowest impact on the franchises CAPEX whilst allow for the best opportunity to make money from obtain/retaining customers.