The Battle of the Neighborhoods

By: Lu Yang
Jan 3, 2020

1. Introduction

With the globalization of the world, it is getting more and more common that people move to a new place or country to study, work, and travel. The United States is a country welcomes people from all over the world. There are many new immigrants coming to the US to look for a new and better life. Among the immigrants, Chinese have been increasing greatly in the past decades. In some neighborhoods, the density of Chinese is quite high, and this triggered a lot of Chinese related businesses to arise. As new immigrants from China settle down in the US, they would think about how they would be making their livings. Also, there are government funded multicultural associations in each state of the US to help people settle down and start their new life. Staff in these associations are often asked the questions about opening new businesses.

We know that a lot of people like Chinese foods. So, why not open a Chinese restaurant? However, the first question would be "where to open it?". This project would help people who would like to open a Chinese restaurant select where to open the restaurant in Queens borough, New York City, where Asian people are heavily populated.

1.1. Problem

Queens borough has a lot of neighborhoods in which a lot Asian, especially Chinese live. There are many Chinese restaurants and other countries' restaurants. In this project, I would aim to offer recommendations where to open a new Chinese restaurant in the neighborhoods of Queens, New York City.

1.2. Target Audience

The new immigrants from China would be interested in where to open a new Chinese restaurant in Queens's neighborhoods. The government funded multicultural associations helping new immigrants would be interested in this as well.

2. Data Acquisition and Cleaning

2.1 Data Source

The venues data of Queens neighborhoods can be obtained from Foursquare by calling its free APIs. The venues categories can be divided into 274 columns. They are restaurants, bars, museums, hotels, banks, and so on. The venues categories from Foursquare is further processed by cleaning, filtering, and modifying for the purposes of solving the proposed problem.

In order to solve the problem, I also retrieve the population of Queens's neighborhoods from the official website of New York City's Department of Health: https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoodpop.htm. This data is combined with the processed Foursquare venues data of Queens neighborhoods for data exploration, visualization, and analysis.

2.2 Data Cleaning
2.2.1 Venues data obtained from Foursquare

The original venues data obtained from Foursquare contains 274 columns in which many do not help on solving our problem. The following snapshot, Table 1, is part of the dataset returned from Foursquare.

Table 1. The original venues categories of Queens neighborhoods returned by Foursquare.

	Neighborhood	Accessories Store	Afghan Restaurant	Airport Terminal	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop
0	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The categories of the venues are divided into 274 columns. It is obvious that we would not need this many columns. What we need is the data of east Asian restaurants, especially Chinese restaurants. Therefore, I removed all columns that the column labels do not contain the string 'Restaurant'. After the removal of those columns, I obtained the dataset that is shown in Table 2. The number of columns decreased to 60.

Table 2. The venues of all restaurants.

	Neighborhood	Afghan Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cantonese Restaurant	Caribbean Restaurant	Chinese Restaurant	Colombian Restaurant	Comfort Food Restaurant	Cuban Restaurant
0	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Astoria	0	0	0	0	0	1	0	0	0	0	0	0	0
2	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Astoria	0	0	0	0	0	0	0	0	0	0	0	0	0

Since we aim to find a neighborhood to open a new Chinese restaurant, we are more interested in the existing Chinese restaurants and east Asian restaurants because they would be competitors to the new restaurant. Restaurants like Cuban restaurants, Italian restaurants are very different from Chinese and east Asian restaurants. Therefore, we do not consider them for choosing the location. The process picking out existing Chinese restaurants and east Asian restaurants is manual based on the column names. Table 3 shows the venues only containing Chinese and east Asian restaurants.

Table 3. The venues of Chinese and east Asian restaurants.

Neighborhood	Cantonese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Filipino Restaurant	Hotpot Restaurant	Indonesian Restaurant	Japanese Restaurant	Korean Restaurant	Malay Restaurant	Shanghai Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Vietnamese Restaurant
Arverne	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Astoria	0	1	0	0	0	0	0	2	1	0	0	0	0	0
Astoria Heights	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Auburndale	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Bay Terrace	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bayside	0	1	0	0	0	0	0	0	1	0	1	0	0	1
Bayswater	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beechhurst	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Bellaire	0	1	0	0	0	0	0	0	0	0	0	0	0	0

After the cleaning and filtering, we have 14 columns representing the Chinese and east Asian restaurants in Queens.

2.2.2 Population in Queens Neighborhoods

In order to analyze the distributions of the restaurants, population is key to help the analysis. However, Foursquare does not provide the information about population. The official website of New York City's Department of Health https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoodpop.htm provides the annual average population. We are only interested in the section for Queens borough. Table 4 shows the population of each neighborhood in Queens.

Table 4. Annual average population in Queens's neighborhoods.

Queens				
	Astoria & Long Island City	88,021	92,150	180,171
	Jackson Heights & North Corona	98,331	90,820	189,150
	Flushing, Murray Hill & Whitestone	123,550	134,764	258,315
	Bayside, Douglaston & Little Neck	58,155	63,468	121,623
	Queens Village, Cambria Heights & Rosedale	90,730	104,801	195,531
	Briarwood, Fresh Meadows & Hillcrest	76,454	83,163	159,617
	Elmhurst & South Corona	76,755	74,131	150,886
	Forest Hills & Rego Park	54,531	62,495	117,027
	Sunnyside & Woodside	70,244	68,429	138,673
	Ridgewood, Glendale & Middle Village	86,201	90,134	176,336
	Richmond Hill & Woodhaven	73,306	74,235	147,541
	Jamaica, Hollis & St. Albans	105,917	122,616	228,533
	Howard Beach & Ozone Park	61,637	65,928	127,565
	Far Rockaway, Breezy Point & Broad Channel	56,053	62,989	119,043

There are totally 14 neighborhoods in the above table. Since this is the officially announced neighborhoods, we make our neighborhoods returned from Foursquare conformed to the above neighborhoods. In order to do this, I combine the rows and sum up the number of restaurants for the same neighborhoods in Table 3 to make the neighborhoods the same as the ones in Table 4. Also, the population of neighborhoods is added as a new column that is shown in Table 5.

Table 5. Refined neighborhoods and combined population.

Neighborhood	Cantonese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Filipino Restaurant	Hotpot Restaurant	Indonesian Restaurant	Japanese Restaurant	Korean Restaurant	Malay Restaurant	Shanghai Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Vietnamese Restaurant	Population
Astoria, Astoria Height, Long Island City	0	4	0	0	0	0	0	2	1	0	0	0	0	0	180171
Jackson Heights, Corona	0	0	0	0	0	0	0	0	0	0	0	0	0	0	189150
Flushing, Murray Hill, Whitestone	1	3	1	1	0	6	0	1	26	1	1	1	0	0	258315
Bayside, Douglaston, Little Neck	1	6	0	0	0	0	0	1	4	0	1	0	0	2	121623
Queens Village, Cambria Heights, Rosedale	0	2	0	0	0	0	0	0	0	0	0	0	0	0	195531
Briarwood, Fresh Meadows & Hillcrest	0	2	1	0	0	0	0	0	0	0	0	0	0	0	159617
Elmhurst, East Elmhurst	0	4	0	0	0	1	1	0	0	1	0	0	0	2	150886
Forest Hills, Rego Park	0	2	0	0	0	0	0	1	0	0	1	0	0	0	117027
Sunnyside, Woodside	0	4	0	0	4	0	0	1	1	0	0	0	0	0	138673
Ridgewood, Glendale, Middle Village	0	3	0	0	0	0	0	0	1	0	0	0	0	0	176336
Richmond Hill, Woodhaven	0	0	0	0	0	0	0	1	0	0	0	0	0	0	147541
Jamaica, Hollis, St. Albans	0	3	0	0	0	0	0	1	0	0	0	0	0	0	228533
Howard Beach, Ozone Park	0	2	0	0	0	0	0	1	0	0	0	0	0	0	127565
Far Rockaway, Breezy Point, Broad Channel	0	3	0	0	0	0	0	0	0	0	0	0	0	0	119043

Table 5 corresponds to the dataframe 'queens_14neighborhoods_asian_restaurants' in my notebook code. The exploration of the data is mainly conducted based on this dataframe.

3. Methodology - Exploratory Data Analysis

Before looking at the data, we can think about some general criteria that would be positive to opening a new Chinese restaurant in some neighborhood of Queens:

(1) Big population
(2) Heavily populated by Chinese/east Asian people
(3) Small number of existing Chinese restaurants in neighborhoods satisfying (2)

When we explore the cleaned dataset, we would try to observe if the data would satisfy the general criteria above.

In order to explore and find the insights of the data, we first should observe the demographics of Queens because it would play a big role of the Chinese and east Asian restaurant distributions. According to the wikipedia of demographics of Queens, we learn that (https://en.wikipedia.org/wiki/Demographics_of_Queens#2013):

Asians are numerous throughout the borough but most concentrated in Northeastern and Central Queens in areas such as

- Flushing
- Little Neck
- Bayside
- Fresh Meadows
- Jamaica Estates
- Elmhurst
- Woodside
- Richmond Hill
- Ozone Park

Please note that it lists the neighborhoods that numerous Asians live, but not necessarily Chinese or east Asian people.

We can assume that the number of Chinese restaurants in the above neighborhoods should be relatively higher than other neighborhoods due to a higher density of Asian people in them. Observing Table 5, we can find that our assumption was right, but with some exceptions. For example, in neighborhood 'Richmond Hill, Woodhaven', there is no Chinese restaurant, but the neighborhood has a lot of Asian residents. We assume this is a neighborhood with very small Chinese population, but a huge Asian population.

Now that we only have information about neighborhoods with Asian heavily populated, we should explore more about the population and the number of east Asian restaurants. We can easily sum up the number of all restaurants on each row and create a new column "Total Restaurants" in the dataframe. This is shown in Table 6.

Table 6. "Total Restaurant" added at the end of the dataframe.

Neighborhood	Cantonese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Filipino Restaurant	Hotpot Restaurant	Indonesian Restaurant	Japanese Restaurant	Korean Restaurant	Malay Restaurant	Shanghai Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Vietnamese Restaurant	Population	Total Restaurant
Astoria, Astoria Height, Long Island City	0	4	0	0	0	0	0	2	1	0	0	0	0	0	180171	7
Jackson Heights, Corona	0	0	0	0	0	0	0	0	0	0	0	0	0	0	189150	0
Flushing, Murray Hill, Whitestone	1	3	1	1	0	6	0	1	26	1	1	1	0	0	258315	42
Bayside, Douglaston, Little Neck	1	6	0	0	0	0	0	1	4	0	1	0	0	2	121623	15
Queens Village, Cambria Heights, Rosedale	0	2	0	0	0	0	0	0	0	0	0	0	0	0	195531	2
Briarwood, Fresh Meadows & Hillcrest	0	2	1	0	0	0	0	0	0	0	0	0	0	0	159617	3
Elmhurst, East Elmhurst	0	4	0	0	0	1	1	0	0	1	0	0	0	2	150886	9
Forest Hills, Rego Park	0	2	0	0	0	0	0	1	0	0	1	0	0	0	117027	4
Sunnyside, Woodside	0	4	0	0	4	0	0	1	1	0	0	0	0	0	138673	10
Ridgewood, Glendale, Middle Village	0	3	0	0	0	0	0	0	1	0	0	0	0	0	176336	4
Richmond Hill, Woodhaven	0	0	0	0	0	0	0	1	0	0	0	0	0	0	147541	1
Jamaica, Hollis, St. Albans	0	3	0	0	0	0	0	1	0	0	0	0	0	0	228533	4
Howard Beach, Ozone Park	0	2	0	0	0	0	0	1	0	0	0	0	0	0	127565	3
Far Rockaway, Breezy Point, Broad Channel	0	3	0	0	0	0	0	0	0	0	0	0	0	0	119043	3

In Table 6, we can find that the neighborhoods with higher number of east Asian restaurants (column "Total Restaurant") conform to what the demographics of Queens illustrations about Asian populated neighborhoods.

So, generally, we can conclude that neighborhoods with bigger population and numerous Asian people would be more promising places to open a new Chinese restaurant. Furthermore, if there are not many Chinese restaurant, but many east Asian restaurants, it would be perfect for opening a new Chinese restaurant.

This implies that the density of Chinese restaurant among east Asian restaurants would be very helpful information for the decision. We would add the ratio of the number of Chinese restaurant to the number of east Asian restaurant in each neighborhood to a new column "Chinese/Total". We call it the density of Chinese restaurants among east Asian restaurants. Table 7 shows the dataset with the new column.

Table 7. The density of Chinese restaurants among east Asian restaurants.

Neighborhood	Cantonese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Filipino Restaurant	Hotpot Restaurant	Indonesian Restaurant	Japanese Restaurant	Korean Restaurant	Malay Restaurant	Shanghai Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Vietnamese Restaurant	Population	Total Restaurant	Chinese/Total
Astoria, Astoria Height, Long Island City	0	4	0	0	0	0	0	2	1	0	0	0	0	0	180171	7	0.57
Jackson Heights, Corona	0	0	0	0	0	0	0	0	0	0	0	0	0	0	189150	0	0.00
Flushing, Murray Hill, Whitestone	1	3	1	1	0	6	0	1	26	1	1	1	0	0	258315	42	0.07
Bayside, Douglaston, Little Neck	1	6	0	0	0	0	0	1	4	0	1	0	0	2	121623	15	0.40
Queens Village, Cambria Heights, Rosedale	0	2	0	0	0	0	0	0	0	0	0	0	0	0	195531	2	1.00
Briarwood, Fresh Meadows & Hillcrest	0	2	1	0	0	0	0	0	0	0	0	0	0	0	159617	3	0.67
Elmhurst, East Elmhurst	0	4	0	0	0	1	1	0	0	1	0	0	0	2	150886	9	0.44
Forest Hills, Rego Park	0	2	0	0	0	0	0	1	0	0	1	0	0	0	117027	4	0.50

A perfect neighborhood for opening a new Chinese restaurant would be the ones with low density of Chinese restaurants, but big number of east Asian restaurants and big Asian population.

With these explorations, we can analyze the data by visualizing the dataset in the next section.

4. Data Visualization and Result Analysis

In this section, we visualize the dataset explored in the last section by scattered plots. Then, we analyze the data in the plot to verify our explorations.

First, we plot the numbers of existing Chinese restaurants and the population in each neighborhood. This is shown in Fig 1.

Fig 1. The number of Chinese restaurants and population.
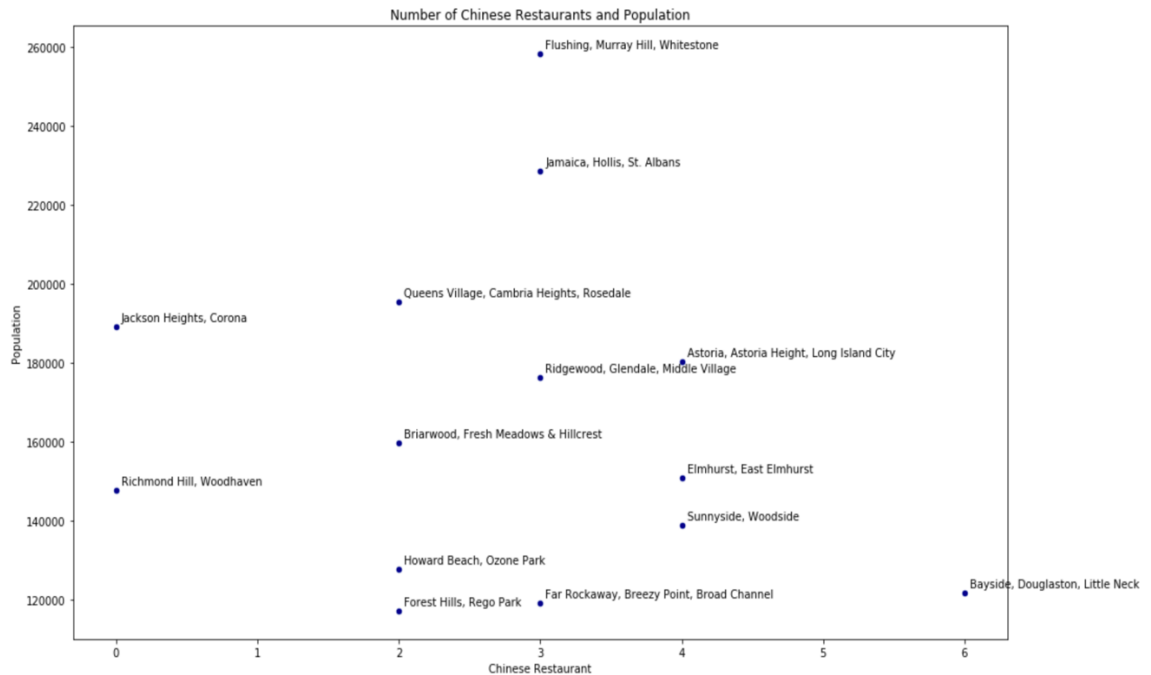
In the plot, there are 6 (the maximum number of Chinese restaurants in the plot) in 'Bayside, Douglaston, Little Neck'. This makes sense because Bayside and Little Neck is a neighborhood with many Asian people. However, the population of Bayside, Douglaston and Little Neck is quite low compared to other neighborhoods. This indicates that there already could be enough Chinese restaurants. So, we would not recommend open a new Chinese restaurant here. The neighborhoods, 'Astoria, Astoria Height, Long Island City', 'Elmhurst, East Elmhurst', and 'Sunnyside, Woodside' are all areas that many Asian people live and it is reasonable that they have relatively higher number of Chinese restaurants. However, with relatively low population in these areas, I would not recommend open a new Chinese restaurant in these three areas.

The next group of neighborhoods, 'Flushing, Murrey Hill, Whitestone', 'Jamaica, Hillis, St. Albans', 'Ridgewood, Glendale, Middle Village', and 'Far Rockaway, Breezy Point, Broad Channel', all have 3 Chinese restaurants, but with very big difference of population. It is easy to observe that the bigger the population, the more restaurants needed. So, it is reasonable that I recommend open a new Chinese restaurant in the neighborhood 'Flushing, Murrey Hill, Whitestone' which has a huge population, 258,315. 'Jamaica, Hillis, St. Albans' is also a good area to open a new Chinese restaurant because of its relatively smalls number of Chinese restaurants and huge population.

Among all the neighborhoods that have either 2 or 0 Chinese restaurants, only one neighborhood 'Richmond Hill, Woodhaven' was reported as a neighborhood with many Asians. But the overall population, 147,541, is quite low compared to other neighborhoods. Therefore, this neighborhood would not be recommended.

Overall, according the analysis of the above plot, the top neighborhood to open a new Chinese restaurant is 'Flushing, Murrey Hill, Whitestone'.

Next, we visualize the number of east Asian restaurants and population in Fig 2.

Fig 2. The number of east Asian restaurants and population.
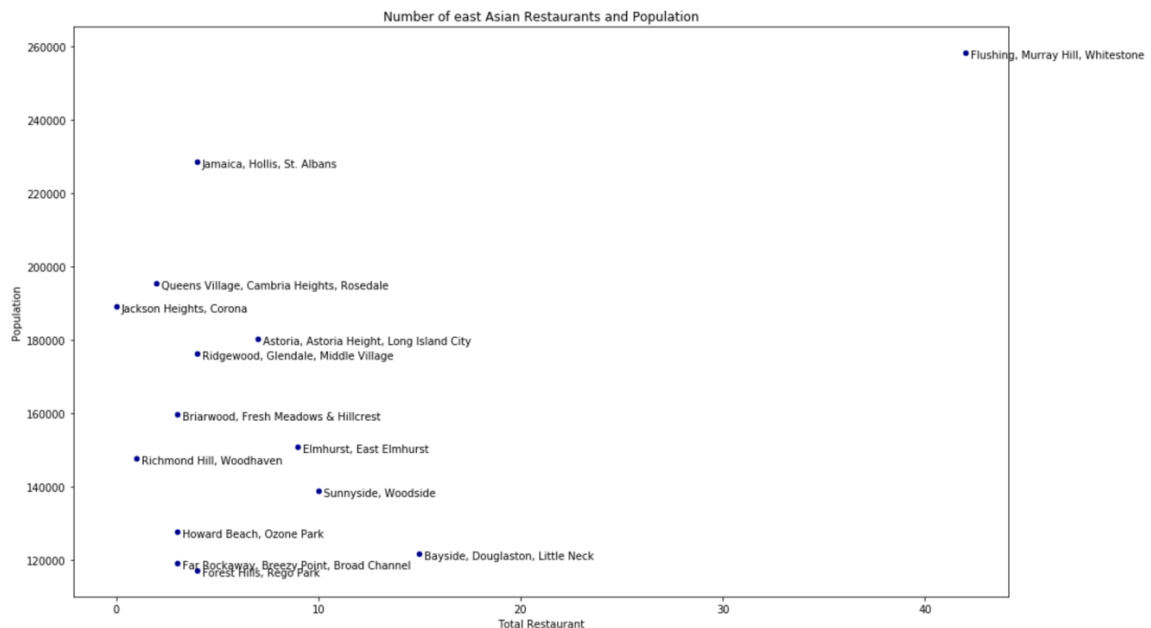
The plot of east Asian restaurants and the population shows some commonalities to the plot in Fig 1. However, there are also some interesting insights and differences.

Not surprisingly, the neighborhood 'Flushing, Murray Hill, and Whitestone' with the biggest population has the biggest number which is 42 of east Asian restaurants. This number is far bigger than the number of east Asian restaurants in other neighborhoods. However, in our previous plot, the number of Chinese restaurants in this neighborhood was not the biggest one and it was close to other neighborhoods. Observing the number of restaurants in the dataset, we see that there are totally 26 Korean restaurants in 'Flushing, Murray Hill, and Whitestone'. This contributes more than half of the east Asian restaurants in the neighborhood. This is an interesting observation because it indicates that the number of Chinese restaurant is not that big compared to other east Asian restaurants while this neighborhood is heavily populated with Asian people.

Looking at other neighborhoods, I see that 'Bayside, Douglaston, Little Neck' has a relatively bigger number of east Asian restaurants with a very low population which again shows this is not a good neighborhood to open a new Chinese restaurant. Almost all other neighborhoods have small number of east Asian restaurants and relatively low population except 'Jamaica, Hillis, St. Albans'. This neighborhood has 4 east Asian restaurants and a big population, 228,533. So, this neighborhood is also a good choice to open a new Chinese restaurant. Overall, from the observation of the above plot, I recommend open a new Chinese restaurant in 'Flushing, Murray Hill, and Whitestone' or 'Jamaica, Hillis, St. Albans'.

We already see that the density of the Chinese restaurants among the east Asian restaurants is very helpful to our analysis. Therefore, we plot the density of Chinese restaurants in Fig 3.

Fig 3. The density of Chinese restaurants and population.

Let's focus more on the neighborhoods that the ratio is lower than 0.5. A ratio above 0.5 indicates that more than half of the restaurants in the neighborhood are Chinese restaurants.

So, the competition would be much tougher than other neighborhoods. And the residents may not need more.

Among the neighborhoods with the ratio values lower than 0.5, all of them have much smaller population than 'Flushing, Murray Hill, and Whitestone'. Furthermore, 'Flushing, Murray Hill, and Whitestone' has a quite small ratio of 0.07. This implies that even though there are very many east Asian restaurants in this neighborhood, Chinese restaurants take a very small portion of them. Considering that 'Flushing, Murray Hill, and Whitestone' is a neighborhood concentrated with Asians and the population is the biggest one among all neighborhoods, I do recommend open a new Chinese restaurant in this neighborhood.

5. Conclusion

This project solves the problem which neighborhood to choose to open a new Chinese restaurant in the Queens borough, New York City. The project uses the venues data of neighborhoods of Queens borough of New York City from Foursquare and the population of the neighborhoods from the official website of New York's Department of Health. The dataset from the two sources was cleaned, filtered, modified, and combined for the exploration, visualization and analysis. Using the scattered plots generated, I am able to analyze the relations of the Chinese restaurants, the east Asian restaurants, and the population in Queens's neighborhoods. Finally, the neighborhood 'Flushing, Murray Hill, and Whitestone' is recommended as the best choice to open a new Chinese restaurant.

6. Future Work

Due to the limitations of the data that I am able to obtain from Foursquare and other sources, I am not able to do more data understanding, data cleaning, model creation, data visualization, and analysis. For example, if we know the revenue of both the Chinese and east Asian restaurants, the numbers of Chinese population in Queens's neighborhoods, the Demographics of visitors to Queens borough, and so on, we would be able to analyze if the number of Chinese population has a linear regression relation with the population although we assume this should be true, which neighborhood's Chinese restaurants have the highest revenue, and which neighborhoods have the biggest number of visitors, especially Chinese visitors.