Coursera Capstone Final Project

Open a Chinese restaurant in Manhattan

1. Business Problem

To open a Chinese restaurant

A client had been migrated to America and lived in Manhattan for several years, and now she has a lot of spare time of her own since her youngest child started to go to school this year. She enjoys Chinese food a lot and so do her friends. She wants to open an authentic Chinese restaurant near where she lives in Manhattan. It's known to all that in Chinatown the Asian people centralized.

So she wants to know:

- **Where else does Asian people concentrate in Manhattan like Chinatown?**
- **How about the demographics, especially Asian population density?**
- **In that area how is the Chinese restaurant proportion and density?**
- **How to score these neighborhoods to give a valuable advice?**

Data science will be used to weigh between the location and restaurants, using data gathered from the website foursquare.com. And demographics from WIKI (census) in Manhattan will also be used to weigh the target neighborhoods.

2. Data Section

Data science will be used to weigh between the location and restaurants, using data gathered from the website foursquare.com. And demographics data from WIKI is based on data from the 2010 United States Census, it will also be used to weigh the target neighborhoods.

Description of the Data:

The following data is required to answer the issues of the problem:

- List of Manhattan neighborhoods with their geodata
- List of restaurants in Manhattan in those neighborhood
- Demographics of Manhattan neighborhoods

  (from wiki: https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods)

Assumption:

- We can reasonably assume that all Asian people likes Chinese cuisines and will be our target customer.
- Demographics data is based on 2010 United States Census, we assume demographics in 2019 stay the same as in 2010.
- The number of premium calls for the *'details'* endpoints is very limited. In this project the data may not precisely reflect the truth. But again, we assume it's the truth.

3. Methodology

- Using clustering to get a glance of Manhattan neighborhoods, and check which neighborhoods are similar to Chinatown.
- Getting the demographics of Manhattan neighborhoods to find where Asian population concentrates, so as to aim to target customers(Asian instead of Chinese, as all most all of Asian people loves Chinese food)
- Find the Chinese restaurant proportion and restaurants number in those target neighborhoods
- Candidate evaluation formula design:

  Evaluation Formula:

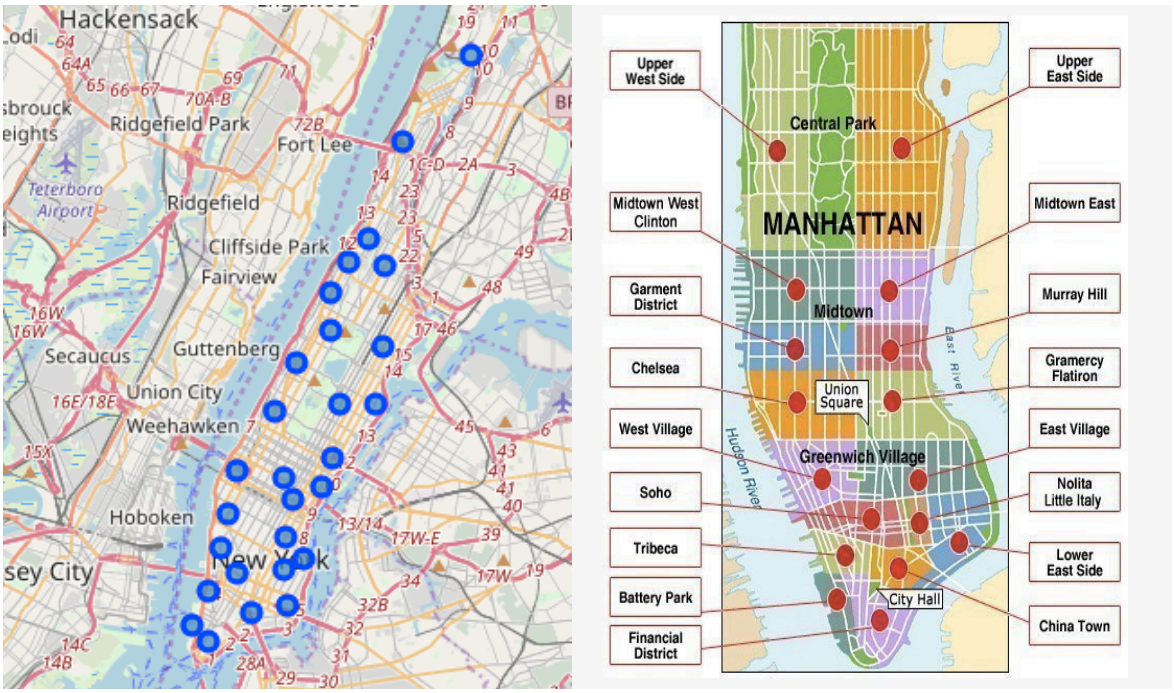
Final_score=Asian_percentage*(1-Chinese_Restaurant_percentage)+population_density/1000

Explanation:

- We assume that Asian people all likes Chinese food, and only consider competition of the same kind.
- Asian people are the key target customers so the higher the better.
- (1-Chinese_Restaurant_percentage) represents the prospect as competition to other Chinese restaurant had been considered and subtracted.
- Population density had also been considered as an adjustment, because the denser the easier to attract more clients.

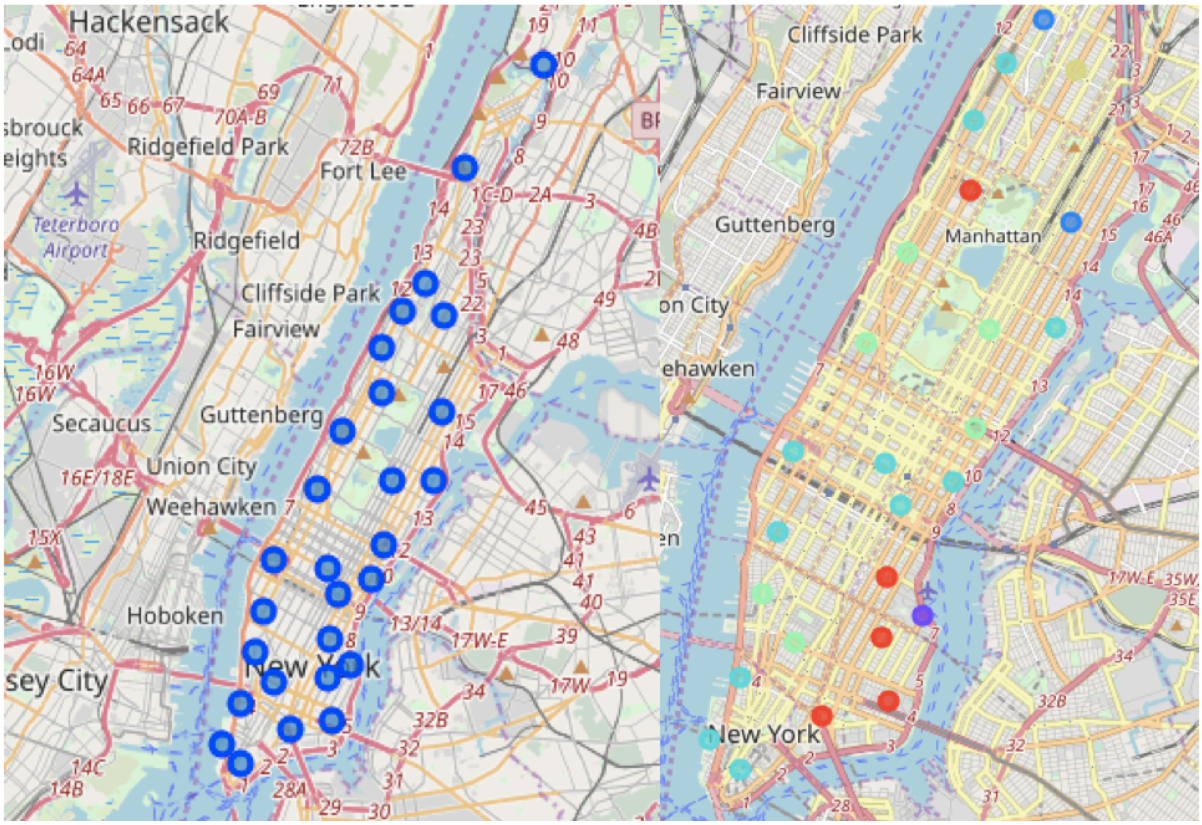
3.1 Finding neighborhoods similar to Chinatown by clustering

There are 30 + neighborhood in Manhattan, but lack of demographic data for some of the neighborhood, those neighborhoods were dropped. Blow is map of the 27 neighborhoods left after data cleaned



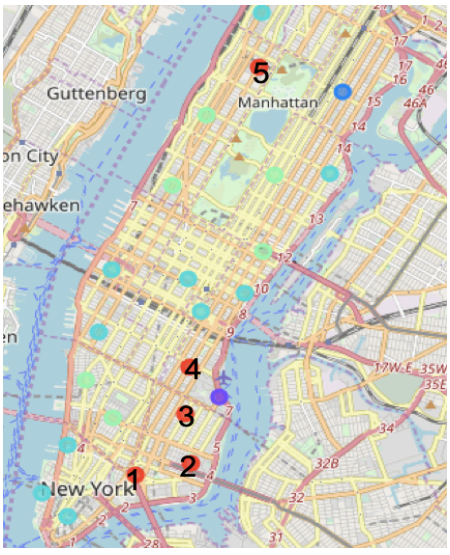
Seven Clustering to get first glance of Manhattan

After iterative clustering, I chose the number of clusters as 7, and below is the final Seven Clustering to get first glance of Manhattan



Five Neighborhoods similar to Chinatown (In red)

It's known to all that the density of Chinese population and restaurants in Chinatown is the highest. So I used clustering to try to find neighborhoods similar to Chinatown and get five similar neighborhoods.

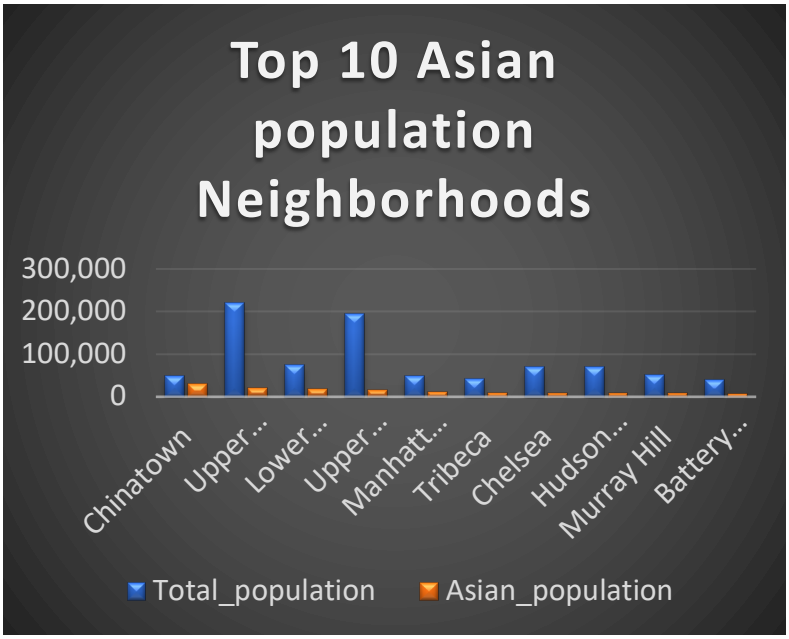


Results:

1, Chinatown
2, Lower East Side
4, Gramercy
3, East Village
5, Manhattan Valley

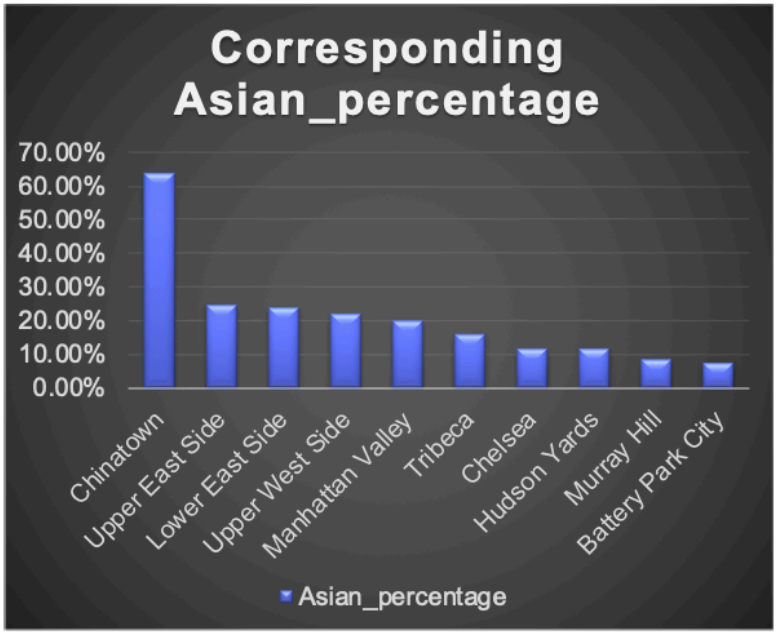
3.2 Check total and Asian population

As we can see that the top Asian population neighborhoods are to some extent similar to the clustering results. We got top 10 Asian population Neighborhoods as Asian people are the most important target customers.



It can be concluded that clustering result of “Gramercy and East Village” will be removed from the candidate list as there are few Asian population.

Asian population proportion:

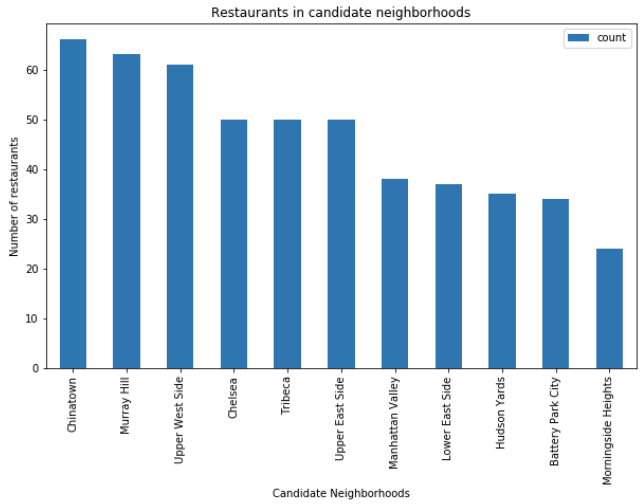


Demographics of the 10 candidate neighborhoods

Neighborhood	Total_population	Asian_percentage	Asian_population	population_density
Chinatown	47,844	63.90%	30,572	144
Lower East Side	72,957	24.90%	18,166	136.1
Manhattan Valley	48,983	24%	11,755	100
Tribeca	42,742	22.20%	9,478	73.5
Battery Park City	39,699	20.20%	8,016	82.7
Murray Hill	50,742	16.20%	8,233	151.5
Chelsea	70,150	11.80%	8,267	82.4
Hudson Yards	70,150	11.80%	8,267	82.4
Upper East Side	219,920	8.60%	18,847	170.3
Upper West Side	193,867	7.60%	14,804	166.8

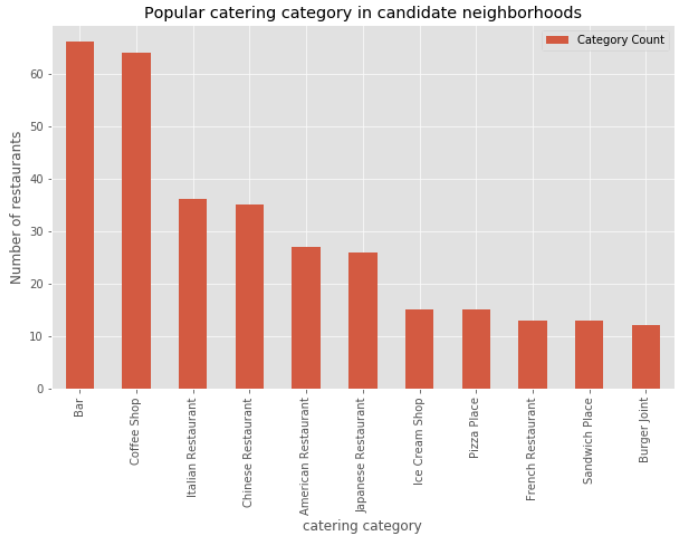
3.3 Explore Chinese restaurants (numbers and proportion)

Below shows the total restaurant numbers in previous candidate neighborhoods.



There are only 35 Chinese Restaurants comparing to the high proportion of Asian immigrants, which means there is a big market left. Chinese food still had a growing prospect in this area.

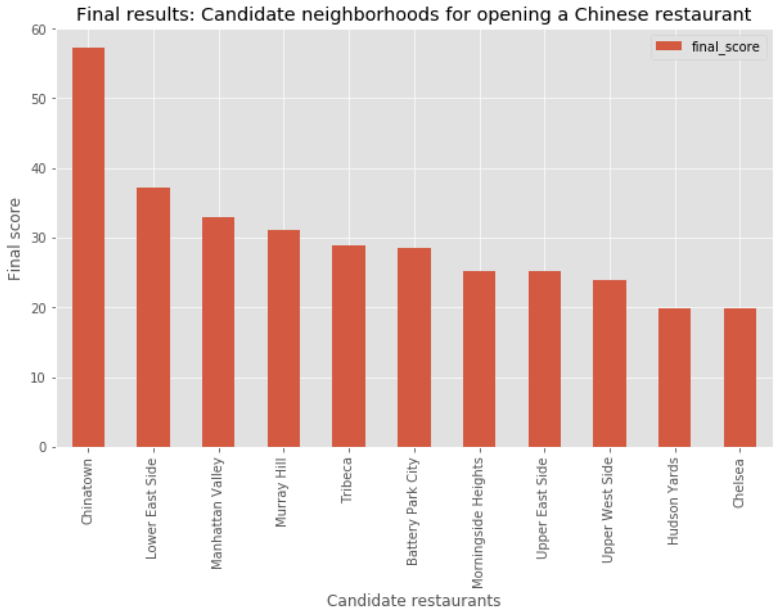
Popular catering category in candidate neighborhoods



Neighborhood	Total_population	Asian_percentage	population_density	Chinese_Restaurant_percentage
Chinatown	47,844	63.90%	144	33.00%
Lower East Side	72,957	24.90%	136.1	5.00%
Manhattan Valley	48,983	24%	100	4.00%
Tribeca	42,742	22.20%	73.5	3.00%
Battery Park City	39,699	20.20%	82.7	0.10%
Murray Hill	50,742	16.20%	151.5	2.00%
Morningside Heights	55,929	13.30%	120.2	0.10%
Hudson Yards	70,150	11.80%	82.4	2.00%
Chelsea	70,150	11.80%	82.4	2.00%
Upper East Side	219,920	8.60%	170.3	5.00%
Upper West Side	193,867	7.60%	166.8	5.00%

4. Results

The Utility Modle considering candidate Asian population density, Chinese restaurant proportion and also use population density as a slight adjustment. The final score sequence will be a good and first indicator and reference to choose a location for opening a new Chinese restaurant even though it's not perfect and can be refined.



5. Discussion

- Future evaluation can also include other factors such as cuisine favorites, transportation and rental price of real state.
- As data science is a highly iterative process, it is likely that the machine learning and predicting approaches could be refined with a smaller set of preselected restaurant types and more data points.

6. Conclusion

- For this assignment location data from the Foursquare API was gathered and was analyzed using multiple data science related packages in Python

- The number of data points for this analysis was limited. Despite that, it possible to get a good overview of the business metrics for specific restaurant types in the area of Manhattan.
- Machine learning was a helpful asset to transfer the insight of the data into a model.