Recommendation for The Location and Style of a New Chinese Restaurant in New York City Neighborhoods

Xiaozhou (John) Zhang

December 30th, 2020

Contents

1. The business problem

For this project, we are going to utilize the Foursquare API for referencing location data in the New York City (NYC), NY area. This data will allow us to report on and provide feedback to interested parties looking at locations and styles to open a new Chinese restaurant in a neighborhood.

The NYC area is home to the largest and most prominent ethnic Chinese population outside of Asia, hosting Chinese populations representing all 34 provincial-level administrative units of China. Chinese cuisines are an important part of Chinese culture, which includes cuisines originating from the diverse regions of China, as well as from Overseas Chinese who have settled in other parts of the world. There are a variety of styles of cooking in China and these styles are distinctive from one another due to factors such as availability of resources, climate, geography, history, cooking techniques and lifestyle. Through the location data analysis, we are going to examine the taste preference of different neighborhoods and identify the hot spots for Chinese restaurants.

With the use of Python and its vast amount of Data Science libraries, we can accurately predict where to open a new Chinese restaurant and decide its style to fit the taste of the favor of the neighborhoods well.

2. The Data

Following the Business Problem, the following data will be used to tailored to retrieve the data for the systematic analysis:

- Neighborhood data of the NYC which has a total of 5 boroughs and 306 neighborhoods
- The list of Chinese restaurants category IDs that were retrieved from Foursquare

We are going to use the NYC neighborhood data to get the geographical coordinates of all its neighborhoods. The geographical coordinates will be used as the center locations to retrieve the information for recommended Chinese restaurants around it with Foursquare's Endpoint "explore". The preferred Chinese restaurants around certain neighborhoods will be determined with the clustering analysis and visualized on the map.

3. Methodology

3.1 Libraries used in this study

- pandas library for data analysis
- numpy library for handling data in a vectorized manner
- json library for handling JSON files
- matplotlib library and associated plotting modules for data visualization
- KMeans module from sklearn.cluster library for clustering analysis
- folium library for map rendering

3.2 Data retrieve and processing

The coordinates for the 306 neighborhoods of NYC were retrieved from the JSON file and casted into a data frame (Fig. 1, check Notebook for details). To narrow down the targeted search results to only Chinese restaurants, 33 category IDs of Chinese restaurants were retrieved from the Foursquare website and casted into a data frame (Fig. 2, check Notebook for details). The comma separated category IDs including all 33 category IDs were created for generating the Foursquare API explore Endpoint request. The function "getNearbyCNRestaurants" was defined with a loop for retrieving nearby recommended Chinese restaurants within the radius of 500 meters of all NYC neighborhoods and cast the results into a data frame. The retrieved data was analyzed with KMeans clustering with the determined optimal k value and the clusters were visualized by map rendering with folium library.

Figure 1 Coordinates of neighborhoods of NYC

	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

Figure 2 List of category IDs of Chinese restaurants

```
CNRestaurants_IDs.head(10)
```

```
(33, 2)
```

	Restaurant Category	Category ID
0	Anhui Restaurant	52af3a5e3cf9994f4e043bea
1	Beijing Restaurant	52af3a723cf9994f4e043bec
2	Cantonese Restaurant	52af3a7c3cf9994f4e043bed
3	Cha Chaan Teng	58daa1558bbb0b01f18ec1d3
4	Chinese Aristocrat Restaurant	52af3a673cf9994f4e043beb
5	Chinese Breakfast Place	52af3a903cf9994f4e043bee
6	Dim Sum Restaurant	4bf58dd8d48988d1f5931735
7	Dongbei Restaurant	52af3a9f3cf9994f4e043bef
8	Fujian Restaurant	52af3aaa3cf9994f4e043bf0
9	Guizhou Restaurant	52af3ab53cf9994f4e043bf1

4. Results

With the category IDs were restricted to only Chinese restaurants and set the radius for the search at 100 meters, in total 241 restaurants were retrieved (Fig. 3, check Notebook for details).

Figure 3 List of recommended Chinese restaurants in NYC neighborhoods

```
NYC_CNRestaurants.head()
```

```
(241, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Restaurant	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	University Heights	40.855727	-73.910416	Merry Land Buffet	40.853287	-73.907745	Dim Sum Restaurant
1	Bay Ridge	40.625801	-74.030621	Kathy's Dumplings	40.627894	-74.029275	Dim Sum Restaurant
2	Bensonhurst	40.611009	-73.995180	大華小炒皇	40.614777	-73.994774	Cantonese Restaurant
3	Bensonhurst	40.611009	-73.995180	Double Ben Cafe 大家樂	40.612389	-74.000717	Cha Chaan Teng
4	Greenpoint	40.730201	-73.954241	Chiko	40.731871	-73.954369	Szechuan Restaurant

Due to the overlap of the search radius in some crowded neighborhoods, there were some duplicated records in the resulted data frame. Duplicated records were checked and dropped if the records have the same restaurant names and coordinates. In total there were 48 duplicated records were dropped, and this leave us 193 restaurants in our final data frame for the further analysis (Fig. 4).

Figure 4 Check and drop duplicated Chinese restaurants retrieved from NYC neighborhoods

```
# Data cleaning to remove duplicated records
NYC_CNRestaurants.drop_duplicates(subset = ['Restaurant', 'Restaurant Latitude', 'Restaurant Longitude'],
                                  keep = "first", inplace = True)
NYC_CNRestaurants.shape
```

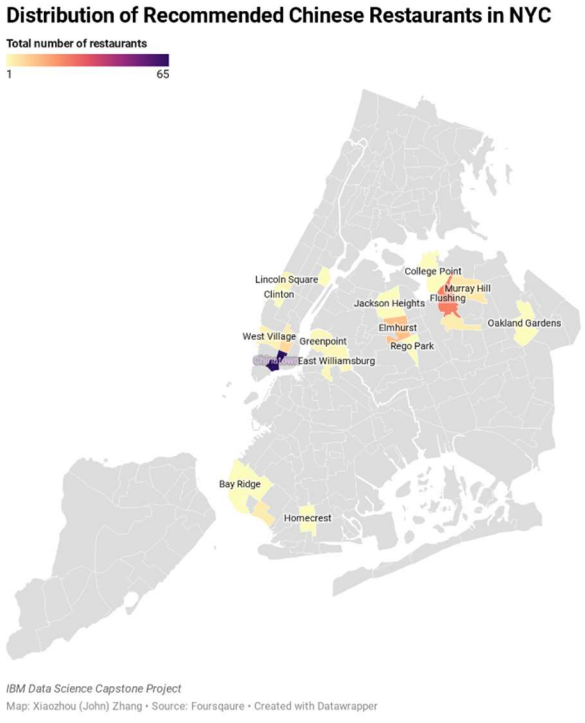
```
(193, 7)
```

The total counts of the recommended Chinese restaurants from different neighborhoods are shown in Fig. 5. To visualize the hot spots that have the high numbers of recommended Chinese restaurants, the choropleth map was generated based on the above counting results with the Datawrapper online web UI (https://www.datawrapper.de). The map clearly shows the hot spots on the map where have the Chinese restaurants got recommended. Note that since some of the neighborhood names in the retrieved data didn't match with the neighborhood names in Datawrapper's database, some data is missing from the map. However, the map still generally reflects the hot spots we summarized.

Figure 5 Counts of the recommended Chinese restaurants from different neighborhoods

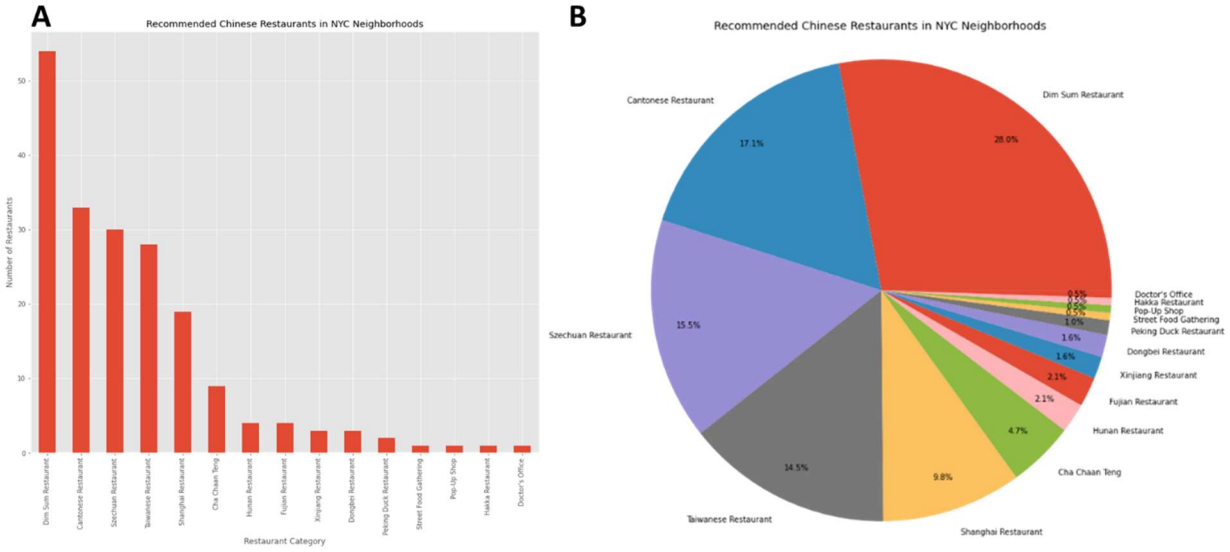
Neighborhood			
Bath Beach	4	Jackson Heights	1
Battery Park City	2	Lenox Hill	1
Bay Ridge	1	Lincoln Square	3
Bayside	1	Little Italy	5
Bensonhurst	2	Little Neck	4
Chelsea	1	Manhattan Valley	4
Chinatown	65	Midtown	4
Clinton	1	Midtown South	2
College Point	1	Murray Hill	5
Downtown	2	New Dorp	1
Dyker Heights	1	Noho	2
East Village	8	North Side	2
East Williamsburg	2	Oakland Gardens	1
Elmhurst	13	Queensboro Hill	4
Financial District	1	Rego Park	1
Flushing	25	Sutton Place	1
Gramercy	4	Tudor City	1
Greenpoint	2	Turtle Bay	2
Greenwich Village	1	University Heights	1
Hillcrest	1	Upper East Side	2
Homecrest	1	Vinegar Hill	1
Hunters Point	2	West Village	3
		Yorkville	1

Figure 6 Visualize the hot spots that have the high numbers of recommended Chinese restaurants



The total counts of all different Chinese restaurants were summarized and plotted in Fig. 7. It is showed that the Dim Sum is the mostly recommended restaurant (54 counts, 28% of the total recommended restaurants). The following four highly recommended restaurants after Dim Sum are Cantonese (17.1%), Szechuan (15.5%), Taiwanese (14.5%), and Shanghai (9.8%) restaurants (Fig. 7B).

Figure 7 The most recommend Chinese restaurants in NYC neighborhoods



To prepare the data for the clustering analysis, the data was transformed with the pandas get_dummies() function and the rows in the transformed data was grouped by neighborhoods and by taking the mean of the frequency of occurrence of each restaurant category (Fig. 8). The top 5 most commonly recommended Chinese restaurants from different neighborhoods were analyzed and summarized in Fig. 9 (check Notebook for details).

Figure 8 The transformed data for clustering analysis

```
# group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
NYC_grouped = NYC_onehot.groupby('Neighborhood').mean().reset_index()
NYC_grouped.head()
```

	Neighborhood	Cantonese Restaurant	Cha Chaan Teng	Dim Sum Restaurant	Doctor's Office	Dongbei Restaurant	Fujian Restaurant	Hakka Restaurant	Hunan Restaurant	Peking Duck Restaurant	Pop-Up Shop	Shanghai Restaurant	Street Food Gathering	Szec Resta
0	Bath Beach	0.75	0.0	0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	Battery Park City	0.00	0.0	0.50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	
2	Bay Ridge	0.00	0.0	1.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	Bayside	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	
4	Bensonhurst	0.50	0.5	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Figure 9 The top 5 most commonly recommended Chinese restaurants from different neighborhoods

----Chinatown----

	restaurant	freq
0	Dim Sum Restaurant	0.28
1	Cantonese Restaurant	0.25
2	Taiwanese Restaurant	0.11
3	Cha Chaan Teng	0.09
4	Shanghai Restaurant	0.08

----Flushing----

	restaurant	freq
0	Dim Sum Restaurant	0.28
1	Szechuan Restaurant	0.24
2	Taiwanese Restaurant	0.16
3	Shanghai Restaurant	0.12
4	Cantonese Restaurant	0.08

----Clinton----

	restaurant	freq
0	Dim Sum Restaurant	1.0
1	Cantonese Restaurant	0.0
2	Cha Chaan Teng	0.0
3	Doctor's Office	0.0
4	Dongbei Restaurant	0.0

----Gramercy----

	restaurant	freq
0	Dim Sum Restaurant	0.50
1	Szechuan Restaurant	0.25
2	Taiwanese Restaurant	0.25
3	Cantonese Restaurant	0.00
4	Cha Chaan Teng	0.00

----College Point----

	restaurant	freq
0	Taiwanese Restaurant	1.0
1	Cantonese Restaurant	0.0
2	Cha Chaan Teng	0.0
3	Dim Sum Restaurant	0.0
4	Doctor's Office	0.0

----Greenpoint----

	restaurant	freq
0	Shanghai Restaurant	0.5
1	Szechuan Restaurant	0.5
2	Cantonese Restaurant	0.0
3	Cha Chaan Teng	0.0
4	Dim Sum Restaurant	0.0

----Downtown----

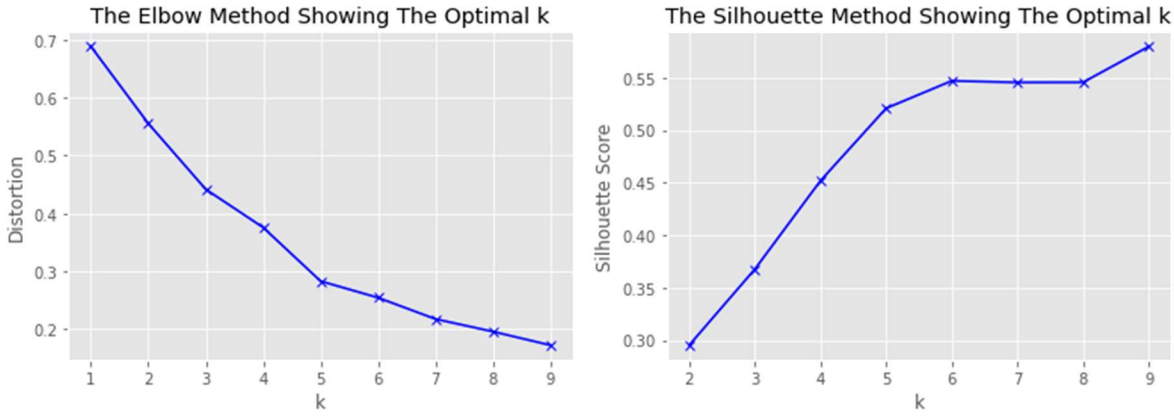
	restaurant	freq
0	Shanghai Restaurant	1.0
1	Cantonese Restaurant	0.0
2	Cha Chaan Teng	0.0
3	Dim Sum Restaurant	0.0
4	Doctor's Office	0.0

----Greenwich Village----

	restaurant	freq
0	Szechuan Restaurant	1.0
1	Cantonese Restaurant	0.0
2	Cha Chaan Teng	0.0
3	Dim Sum Restaurant	0.0
4	Doctor's Office	0.0

To achieve the optimized clustering results for the neighborhoods, the optimal k value was determined by using two different approaches: the elbow method and the silhouette method. Both methods showed the turning point at k=5 and then 5 was selected for the further clustering analysis (Fig. 10).

Figure 10 Determine the optimal k for clustering analysis



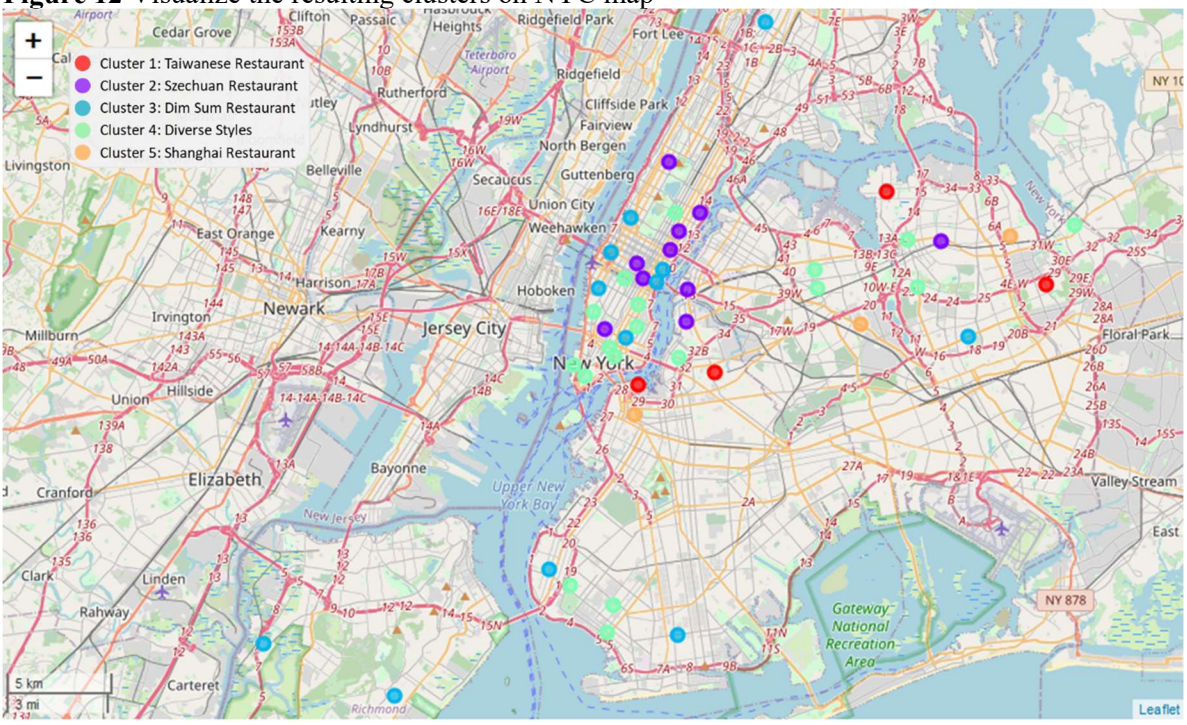
The clustering results were shown in the Fig. 11 (check Notebook for details). The resulting clusters were visualized on the NYC neighborhoods map (Fig. 12).

Figure 11 Neighborhoods with cluster labels and their 10 most commonly recommended restaurants

Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th M Comm Restau
Bronx	University Heights	40.855727	-73.910416	2	Dim Sum Restaurant	Xinjiang Restaurant	Taiwanese Restaurant	Szechuan Restaurant	Street Food Gathering	Shanghai Restaurant	Pop-Up Shop	Peking Duck Restaurant	Hunan Restaurant	Ha Restau
Brooklyn	Bay Ridge	40.625801	-74.030621	2	Dim Sum Restaurant	Xinjiang Restaurant	Taiwanese Restaurant	Szechuan Restaurant	Street Food Gathering	Shanghai Restaurant	Pop-Up Shop	Peking Duck Restaurant	Hunan Restaurant	Ha Restau
Brooklyn	Bensonhurst	40.611009	-73.995180	3	Cha Chaan Teng	Cantonese Restaurant	Xinjiang Restaurant	Taiwanese Restaurant	Szechuan Restaurant	Street Food Gathering	Shanghai Restaurant	Pop-Up Shop	Peking Duck Restaurant	Hu Restau
Brooklyn	Greenpoint	40.730201	-73.954241	1	Szechuan Restaurant	Shanghai Restaurant	Xinjiang Restaurant	Taiwanese Restaurant	Street Food Gathering	Pop-Up Shop	Peking Duck Restaurant	Hunan Restaurant	Hakka Restaurant	Fu Restau

Different clusters were examined (check Notebook for details) and were highlighted with their unique feature based on their first most commonly recommended restaurants. As we can see, neighborhoods in cluster 4 is quite different which have the very diverse choice on their first most commonly recommended restaurants (including styles of Cantonese, Dim Sum, Szechuan, Cha Chaan, Street Food Gathering, and Shanghai). All the other four clusters of neighborhoods have vey high common selection for their first most commonly recommended restaurants (Cluster 1: Taiwanese, Cluster 2: Szechuan, Cluster 3: Dim Sum, and Cluster 5: Shanghai).

Figure 12 Visualize the resulting clusters on NYC map



5. Discussion

As we can see the results, the Chinatown in the Manhattan area definitely is the highly preferred place to try out some highly recommended Chinese restaurants. The second place will be the Flushing area at Queens borough (Fig.6). Dim Sum restaurants is the top 1 recommend restaurant for both Chinatown and Flushing, and followed by Cantonese and Szechuan restaurants. These three types of restaurants are also the generally highly recommended restaurants for the whole NYC neighborhoods (Fig. 7B). As showed in the Fig. 12, neighborhoods of NYC can be clustered into five clusters with at least the obvious distinction in their top 1 recommendation of the Chinese restaurants. The recommended Chinese restaurants are mostly located in the boroughs of Manhattan and Northern Queen which centralizing at the Chinatown and Flushing neighborhoods, which is not surprising since these two neighborhoods have the largest Chinese American population and they are famous and attractive places for international tourists who want to experience Chinese culture and have a taste of Chinese dishes.

One interesting thing I tried is to change the radius for the Foursquare API search request from 500 meters to 5000 meters gradually to see whether I can retrieve more recommended Chinese restaurants from the NYC neighborhoods. It was found that the number of the retrieved restaurants did increase significantly (with the duplicated records dropped due to the overlap of the radius in the crowded neighborhoods). However, I also found that with the current defined search request function the names of the neighborhood were frequently mislabeled. I also retrieved some sample data from Foursquare and realized that the name of the neighborhood of a venue is not always added in the database which make it impossible to retrieve and cast the right name of the neighborhood of a restaurant directly into my data frame. It will be ideal that if one can define a function that can just retrieve interested venues within the boarder of a given neighborhood with the corresponding GEOJASON data.

6. Conclusion

Based on the analysis of the retrieved data from Foursquare and the neighborhoods clustering results, I would recommend one to open their new Chinese restaurant in the area of Chinatown or Flushing neighborhoods from where there are highest population of Chinese Americans and with high traffic for tourists who want to experience Chinese culture and food. To brace the competition, I will recommend the style of the restaurant to be Dim Sum, Cantonese, and Szechuan restaurants.

Keep it in mind that some neighborhoods in Manhattan and Northern Queen (Fig. 12, cluster 4: Diverse Styles) are quite different which have the very diverse choice on their first most commonly recommended restaurants (including styles of Cantonese, Dim Sum, Szechuan, Cha Chaan, Street Food Gathering, and Shanghai). I guess this opens the door for the restaurant owner to try some new Chinese cuisines.

7. References

- Notebook on Github repository
- Lab Notebook for Applied Data Science Capstone - Segmenting and Clustering Neighborhoods in New York City
- New York City - Wikipedia
- Neighborhoods in New York City - Wikipedia
- 'Places API' documentation - Foursquare