

FINAL REPORT

Xin Zhang

Capstone Project – The Battle of Neighborhoods

Introduction:

New York City (NYC), often called The City or simply New York (NY), is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles, New York is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York is also an important center for international diplomacy.

Situated on one of the world's largest natural harbors, New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.

The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs speak English as alternate language.

With New York City's diverse culture, comes diverse food items. There are many restaurants in New York City, each belong to different categories like Chinese, Indian, French, Italian etc.

Problems should be solved in this project:

1. List and visualize all major parts of New York City that has Chinese restaurants
2. What is the best location in New York City for Chinese Cuisine?
3. Which areas have potential Chinese Restaurant Market?
4. Which all areas lack Chinese Restaurants?
5. Which is the best place to stay if you prefer Cuisine?

Data Section:

New York City (NYC), often called The City or simply New York (NY), is the most populous city in the United States. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce. New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island. While, based on these basic introductions of New York City, we can know that New York City's demographics show that it is a large and ethnically diverse metropolis. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belong to different categories like Chinese, Indian, French etc.

Therefore, in this project, I'd like to know how we can choose a Chinese Restaurant in New York City if we want to eat some Chinese cuisine. So, we might need the following data:

- New York City data that contains Boroughs, Neighborhoods along with their latitude and longitude.
 - Data source: https://geo.nyu.edu/catalog/nyu_2451_34572
 - This dataset contains the required information, such as latitude, longitude. And we will use this dataset to explore various neighborhoods and borough of New York City.
- Chinese Restaurants in each neighborhood of New York City
 - Data source: Foursquare API (only 50 premium calls each day)
 - Description: By using this API we will get all the venues and details of venues in each neighborhood. We can filter these venues to get only Chinese restaurants.
- GeoSpace Data
 - Data source: <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>
 - Description: By using this geo space data we will get the New York Borough boundaries that will help us visualize choropleth map.

Methodology:

1. Firstly, We need to collect the New York city data from https://cocl.us/new_york_dataset.

```
1 def get_new_york_data():
2     url='https://cocl.us/new_york_dataset'
3     resp=requests.get(url).json()
4     # all data is present in features label
5     features=resp['features']
6
7     # define the dataframe columns
8     column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
9     # instantiate the dataframe
10    new_york_data = pd.DataFrame(column_names)
11
12    for data in features:
13        borough = data['properties']['borough']
14        neighborhood_name = data['properties']['name']
15
16        neighborhood_latlon = data['geometry']['coordinates']
17        neighborhood_lat = neighborhood_latlon[1]
18        neighborhood_lon = neighborhood_latlon[0]
19
20        new_york_data = new_york_data.append({'Borough': borough,
21                                              'Neighborhood': neighborhood_name,
22                                              'Latitude': neighborhood_lat,
23                                              'Longitude': neighborhood_lon}, ignore_index=True)
24
25    return new_york_data
```

```
1 new_york_data = get_new_york_data()
2
3 new_york_data.shape
```

(306, 4)

The above result shows that there are 306 different Neighborhoods in New York.

```
1 new_york_data.head()
```

	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

2. We will find all venues for each neighborhood using FourSquare API. And the new Data frame shows just like below, including Venue Id, Venue Name, Venue Category.

```
1 ny_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Id	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	4c537892fd2ea593cb077a28	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	4d375ce799fe8eec99fd2355	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	4c9e50e38afca09379b2ff15	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	4c1bed4eb306c928140763b7	Jimbo's	40.891740	-73.858226	Burger Joint
4	Wakefield	40.894705	-73.847201	4c783cef3badb1f7e4244b54	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop

3. Next we will collect Chinese restaurants for each Neighborhood.

```
1 Chinese_restaurant = ny_venues[ny_venues['Venue Category']=='Chinese Restaurant']
2
3 Chinese_rest_ny = pd.merge(new_york_data, Chinese_restaurant, how='inner', on='Neighborhood').reset_index()
4 Chinese_rest_ny = Chinese_rest_ny.drop(columns=['Neighborhood Latitude', 'Neighborhood Longitude', 'Venue Latitude'])
5 Chinese_rest_ny = Chinese_rest_ny.drop(columns=['index', 'level_0'], axis=1)
6 Chinese_rest_ny.head()
```

Borough	Neighborhood	Latitude	Longitude	Venue Id	Venue Name	Venue Category	
0	Bronx	Baychester	40.866858	-73.835798	4c1a8bac8b3aa593a8c8955f	Dragon City	Chinese Restaurant
1	Bronx	Pelham Parkway	40.857413	-73.854756	4b9d6b45f964a52078ab36e3	Mr. Q's Chinese Restaurant	Chinese Restaurant
2	Bronx	Pelham Parkway	40.857413	-73.854756	4c3f7f2eda3dc928b8f6c5b9	Peking Kitchen	Chinese Restaurant
3	Bronx	City Island	40.847247	-73.786488	4bfc652bf14fa593c1f5c9d4	City Island Chinese Restaurant	Chinese Restaurant
4	Bronx	University Heights	40.855727	-73.910416	566f33e7498e44c2501bda81	Panda Express	Chinese Restaurant

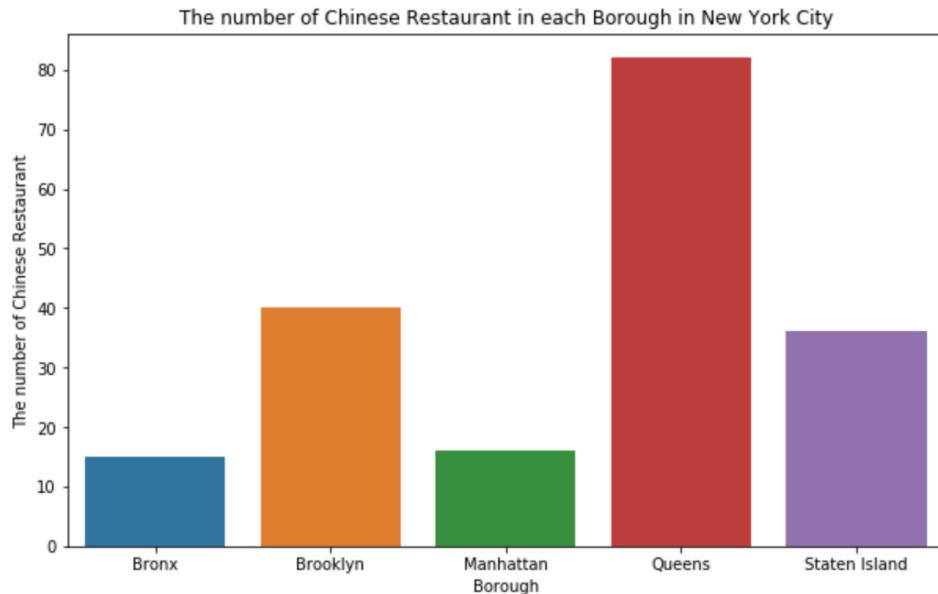
```
1 Chinese_rest_ny.shape
```

(189, 7)

The result clearly shows that there are 189 Chinese restaurants across New York City.

4. Create a Bar Plot to visualize number of Chinese Restaurants for each Borough and each neighborhood in New York City.

```
1 plt.rcParams['figure.figsize'] = [10, 6]
2
3 bor_chinese = Chinese_rest_ny.groupby(['Borough'])['Venue Name'].count().reset_index()
4 sns.barplot(x='Borough', y='Venue Name', data=bor_chinese)
5 plt.title('The number of Chinese Restaurant in each Borough in New York City')
6 plt.ylabel('The number of Chinese Restaurant')
7 plt.show()
```

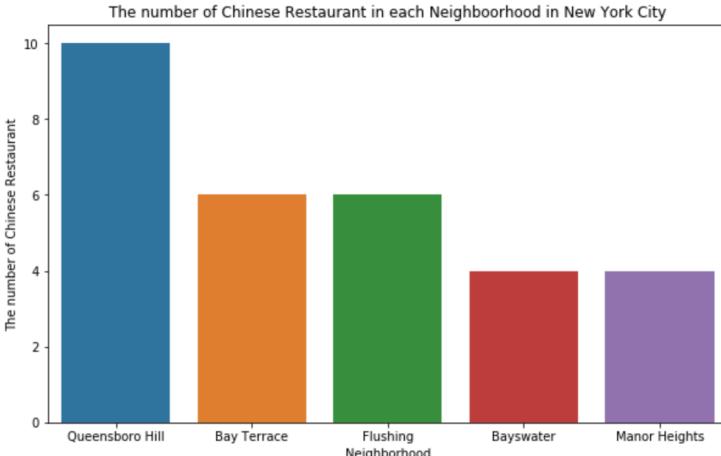


we can see that Queens has the most Chinese restaurants and Bronx and Manhattan have the low number of Chinese restaurants.

```

1 neigh_chinese = Chinese_rest_ny.groupby(['Neighborhood'])['Venue Name'].count().reset_index()
2 sns.barplot(x='Neighborhood', y='Venue Name', data=neigh_chinese.nlargest(5, 'Venue Name'))
3 plt.title('The number of Chinese Restaurant in each Neighborhood in New York City')
4 plt.ylabel('The number of Chinese Restaurant')
5 plt.show()

```



We can see that Queensboro Hill in Queens has the highest number of Chinese Restaurant with a total count of 10.

5. Next using FourSquare API, we will find the Ratings, Tips, and Like count for all the Chinese Restaurants.

```

1 venue_details=[]
2 for item in data:
3     try:
4         venue_data = item['response']['venue']
5         venue_id=venue_data['id']
6         venue_name=venue_data['name']
7         venue_likes=venue_data['likes']['count']
8         venue_rating=venue_data['rating']
9         venue_tips=venue_data['tips']['count']
10        venue_details.append([venue_id, venue_name, venue_likes, venue_rating, venue_tips])
11    except KeyError:
12        pass
13 column_names=['ID','Name','Likes','Rating','Tips']
14 Chinese_rest_details = pd.DataFrame(venue_details,columns=column_names)
1 Chinese_rest_details = pd.concat([Chinese_rest_details_1, Chinese_rest_details_2, Chinese_rest_details_3, Chinese_
2 Chinese_rest_details = Chinese_rest_details.rename(columns={'ID': 'Venue Id'})
3 Chinese_rest_details = Chinese_rest_details.drop('index', axis=1)
4 Chinese_rest_details

```

	Venue Id	Name	Likes	Rating	Tips
0	4c1a8bac8b3aa593a8c8955f	Dragon City	1	6.8	8
1	4b9d6b45f964a52078ab36e3	Mr. Q's Chinese Restaurant	9	7.2	10
2	4c3f7f2eda3dc928b8f6c5b9	Peking Kitchen	9	7.8	2
3	4bfc652b1f14fa593c1f5c9d4	City Island Chinese Resturant	6	6.4	4
4	566f33e7498e44c2501bda81	Panda Express	12	7.6	2
5	4c7e9ddaf13a1cdf5369ea4	El Pabelon De Oro	15	7.7	8
6	4b1c3b4cf964a520bb0424e3	Grand Sichuan House	37	8.3	24
7	4c65da95f7a80f47b6959c7a	Spicy Bampa	28	7.5	10
8	5a7f87920d173f267e27cc36	Golden City Seafood	10	8.0	4
9	4aece020f964a520f7cb21e3	Silver Star Restaurant	7	7.8	7
10	4bfaf805d0382d7f520ac90a	China Delight	9	8.1	7

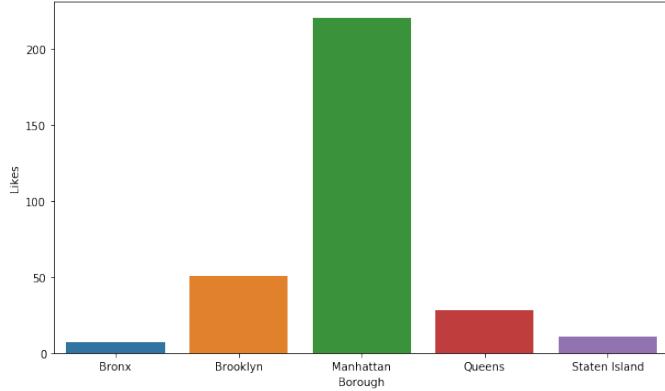
Here I spend four days getting the detail venue information of all Chinese restaurants across the New York City, because I can only have 50 premium calls each day from Foursquare. Therefore, there is a merge step for each day call.

6. Next let's visualize Borough with average Likes, Rating and Tips of restaurants.

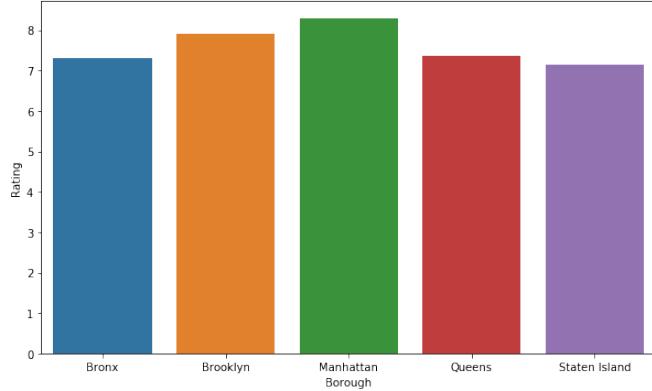
```
1 bor_Chinese = Chinese_rest.groupby('Borough')['Likes', 'Rating', 'Tips'].mean().reset_index()
2 bor_Chinese
```

	Borough	Likes	Rating	Tips
0	Bronx	7.500000	7.312500	5.125000
1	Brooklyn	51.000000	7.894737	17.263158
2	Manhattan	220.571429	8.300000	60.571429
3	Queens	28.312500	7.350000	13.437500
4	Staten Island	11.333333	7.144444	9.444444

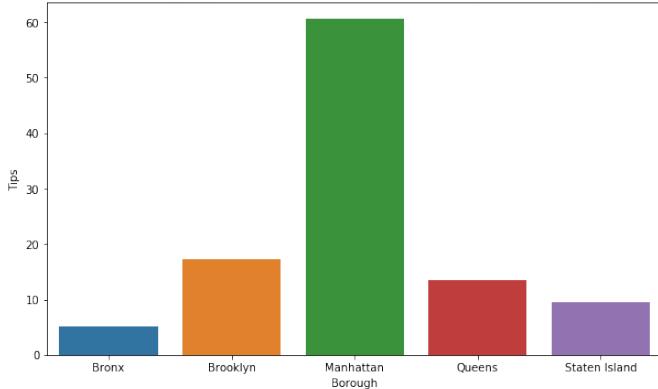
The average Likes of Chinese restaurants in each Borough of New York City



The average Rating of Chinese restaurants in each Borough of New York City



The average Tips of Chinese restaurants in each Borough of New York City



From the above graphs, we can clearly see that the Chinese restaurants in Manhattan will have most likes, tips and best rating.

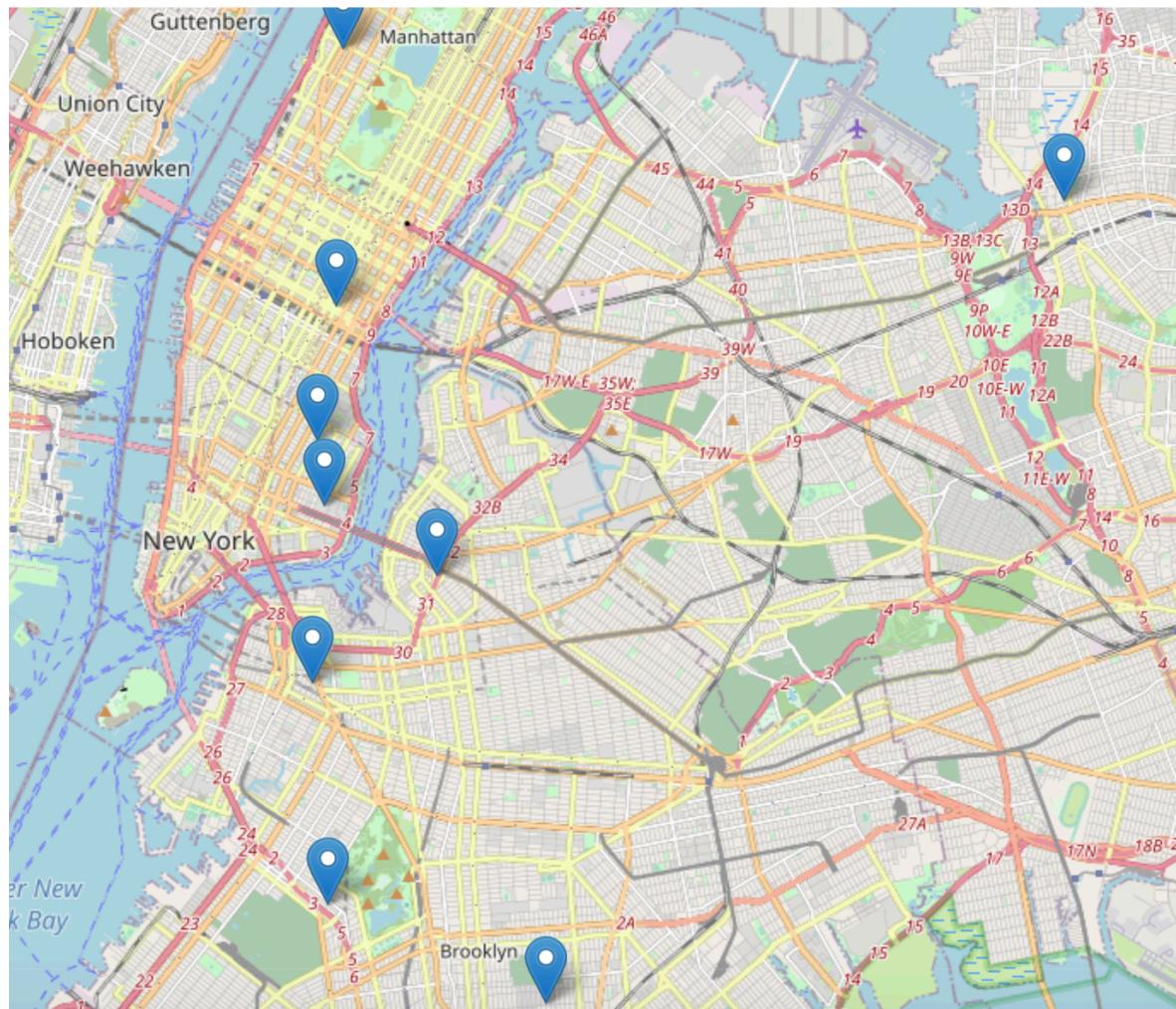
7. Visualize Top10 maximum Rating of Chinese restaurants in New York City and show it on the map using folium package.

```

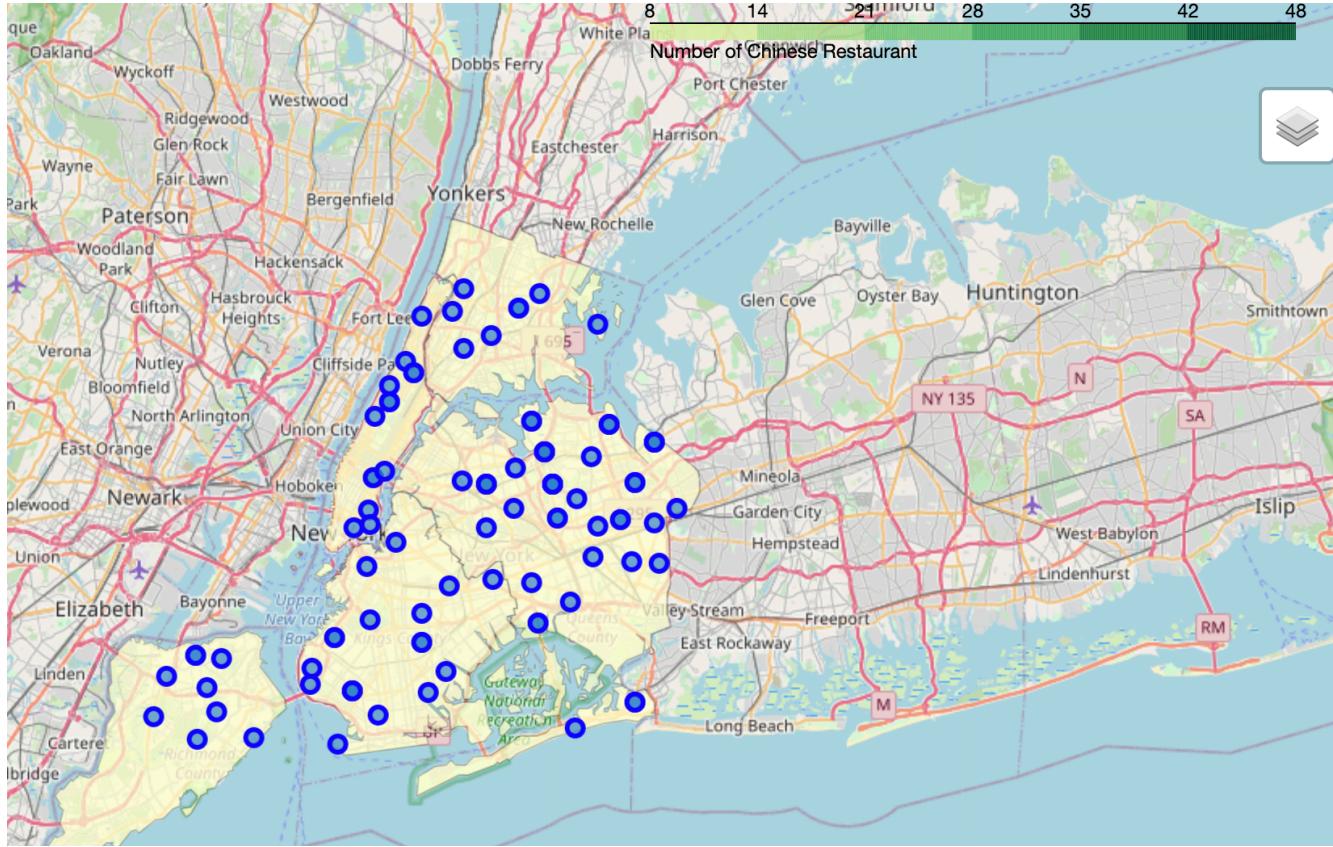
1 ny_rating_top10 = Chinese_rest.sort_values(by='Rating', ascending=False).reset_index().head(10)
2 ny_rating_top10 = ny_rating_top10.drop(['index','level_0'], axis=1)
3 ny_rating_top10

```

	Borough	Neighborhood	Latitude	Longitude	Venue Id	Venue Name	Likes	Rating	Tips
0	Manhattan	Lower East Side	40.717807	-73.980890	5b380f649deb7d00399fdf9d	Kings County Imperial	67	9.2	10
1	Brooklyn	Marine Park	40.609748	-73.931344	4b5a6114f964a52043c128e3	China New Star	58	8.9	19
2	Brooklyn	Williamsburg	40.707144	-73.958115	58df00318cfef546addb99246	Birds of a Feather	341	8.9	71
3	Brooklyn	Windsor Terrace	40.656946	-73.980073	54e39b14498e9c646b42958e	East Wind Snack Shop	138	8.9	50
4	Queens	Flushing	40.764454	-73.831773	4d1bcddc1356224bc6410abe	OK Ryan	48	8.8	19
5	Manhattan	Murray Hill	40.748303	-73.978332	4e3484038877beb5e9a22a0b	Café China	1046	8.8	273
6	Brooklyn	East Flatbush	40.641718	-73.936103	4c1ad571eac020a10a0144c2	China Hao Restaurant	32	8.7	15
7	Manhattan	East Village	40.727847	-73.982226	5647ee82498e8bf0dddef53d	Málà Project	436	8.7	91
8	Manhattan	Upper West Side	40.787658	-73.977059	5215797811d2f1217bb93475	RedFarm	470	8.6	129
9	Brooklyn	Downtown	40.690844	-73.983463	59da9590e1f0aa52976b8f35	Han Dynasty	161	8.6	32



8. Show the distribution of Chinese Restaurants across New York City on the map.



Here, I was using chloropleth map of folium map to visualize the distribution of Chinese restaurants in each borough of New York City. And I tried to figure out the reason why it couldn't show the different levels of color, but I couldn't find an answer for it. So, I am very pleasant to discuss with you if you know how to deal with it. I really appreciate it.

Results:

So, we can give the answer to the question asked in the question above:

1. The best location in New York City for Chinese Cuisine:

Borough	Neighborhood	Latitude	Longitude	Venue Id	Venue Name	Likes	Rating	Tips	
0	Manhattan	Lower East Side	40.717807	-73.980890	5b380f649deb7d00399fdf9d	Kings County Imperial	67	9.2	10
1	Brooklyn	Marine Park	40.609748	-73.931344	4b5a6114f964a52043c128e3	China New Star	58	8.9	19
2	Brooklyn	Williamsburg	40.707144	-73.958115	58df00318cfef46addb99246	Birds of a Feather	341	8.9	71
3	Brooklyn	Windsor Terrace	40.656946	-73.980073	54e39b14498e9c646b42958e	East Wind Snack Shop	138	8.9	50
4	Queens	Flushing	40.764454	-73.831773	4d1bcddc1356224bc6410abe	OK Ryan	48	8.8	19

2. Queens has the most Chinese restaurants and Bronx and Manhattan have the low number of Chinese restaurants. Therefore, Queens borough is the best place to open a Chinese restaurant market.

Borough	Venue Name
0	Bronx
1	Brooklyn
2	Manhattan
3	Queens
4	Staten Island

3. Obviously, Bronx and Manhattan are the two boroughs lacking Chinese restaurant. But both of them have huge potential to open Chinese restaurants to attract more Chinese customers.
4. According to the data shows, I think Manhattan is the best choice for you to enjoy Chinese cuisine. Because the average amount of customers likes and average rating of Chinese restaurant are leading all five boroughs. Therefore, I recommend that you'd better to go to Manhattan if you want to have the best Chinese food experience.

Discussions:

Based on the data that we get from New York University and Foursquare, we can know Queens have the most Chinese restaurants, which may due to there are more Chinese resident living there. So, Queens borough is the potential Chinese market for you if you'd like to open a store or restaurants selling something that Chinese people like.

Generally speaking, Manhattan borough is the best place to enjoy Chinese cuisine. From data analysis, we can clearly see that Chinese restaurant in Manhattan will have the most customers and each customer will give them a high evaluation. On the other hand, Manhattan has the highest consumption level due to the high tips. Therefore, Chinese restaurant must be the best choice for you if you don't worry about the cost.

Conclusion:

In this project, we can conclude that if you want more choice for Chinese restaurant or have a nice Chinese food experience in a not too high price. The Queens borough is a good choice for you. While, you'd better to enjoy Chinese cuisine in Manhattan if you work and live in the center of New York City and don't care about the cost.

All of the results are from my analysis, I know there is always room for improvement and therefore the above solution I provided can also be improved for better results upon the data we have.