

COURSERA PROJECT

Where is the best place to open a new restaurant in New York City?

Ana Martínez García



Ana Martínez García

1 CONTENIDO

2	Business Description	2
2.1	Introduction.....	2
2.2	Problem approach.....	2
2.3	Target	2
3	Data acquisition and cleaning	2
3.1	Data Sources.....	2
4	Methodology	3
5	Analysis.....	3
6	Result.....	6
7	Conclusions	6
8	Recommendations	7

2 BUSINESS DESCRIPTION

2.1 INTRODUCTION

This Project will access to the different restaurants and its categories in the borough and neighborhoods of New York City, EEUU, with the main objective to study the distribution of the Chinese restaurants around the city.

New York has a long history with Chinese food so it's a wise idea make a study of how spread the different Chinese restaurants are in New York and try to stablsh where is the best place for open a new restaurant.

2.2 PROBLEM APPROACH

The information available will be treated using Foursquare API that will allow us to access the venues in each borough and neighborhoods and the use of Data Science will provide the clustered data based.

The visualization of the different locations that we will study will be provided by the Folium visualization library that it will help to visualize the suitable locations for the new restaurant displaying the surrounded restaurants that can be used as reference.

2.3 TARGET

The main stakeholder will be any small or big chain business owners that want to open a new restaurant, that need the answer for the following questions:

- Correlation between population and borough.
- Correlation between population and number of restaurants.
- Study of the rating and valuated Chinese restaurants.

3 DATA ACQUISITION AND CLEANING

3.1 DATA SOURCES

New York is one of the largest cities in the EEUU with several borough and plenty of neighborhoods. The dataset must contain the borough, neighborhood and the coordinates of each one, latitude and longitude (figure 1). The New York dataset contains a total 306 neighborhoods.

	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

```
new_york_data.shape
```

```
(306, 4)
```

Figure 1.- New York Dataset

4 METHODOLOGY

The first steps will require the study for cleaning the data and adapt it into the borough and neighborhood with its coordinates to continue with a calculation and exploration of venues across the different borough of New York City. This new dataset will be visualized using Folium python library.

The workflow allows us to study and provide the stakeholder an answer of where will be the most suitable place to open the new restaurant by identifying relationships between the different features of the datasets.

The update of the final reports and code will be done in GitHub

5 ANALYSIS

The study of the general New York dataset shows that the borough with more of the neighborhood in New York is Queens, this is translated into that the most populated borough in the studied city result to be Queens (figure 2).

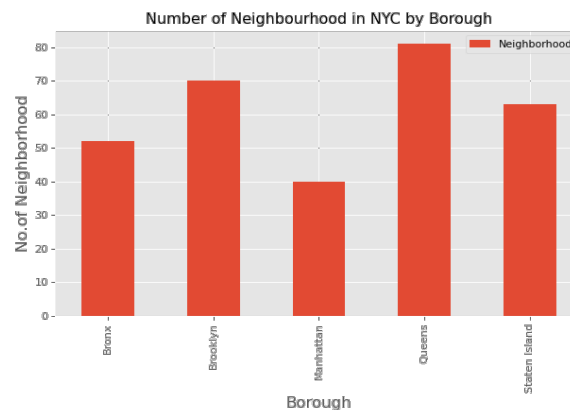


Figure 2- Number of neighborhood by borough in New York city.

Using this base, we are going to see how is correlate this information with where are most of the Chinese restaurants in the city, studying by one side by borough and by neighborhood. This study shows that the correlation between the number of restaurants is also higher in Queens than in any other borough in New York (figure 3 a).

This same trend is observing when we try to obtain where are higher concentration of Chinese restaurant by neighborhood, resulting in the Flushing neighborhood, which is actually located in Queens borough (figure 3 b).

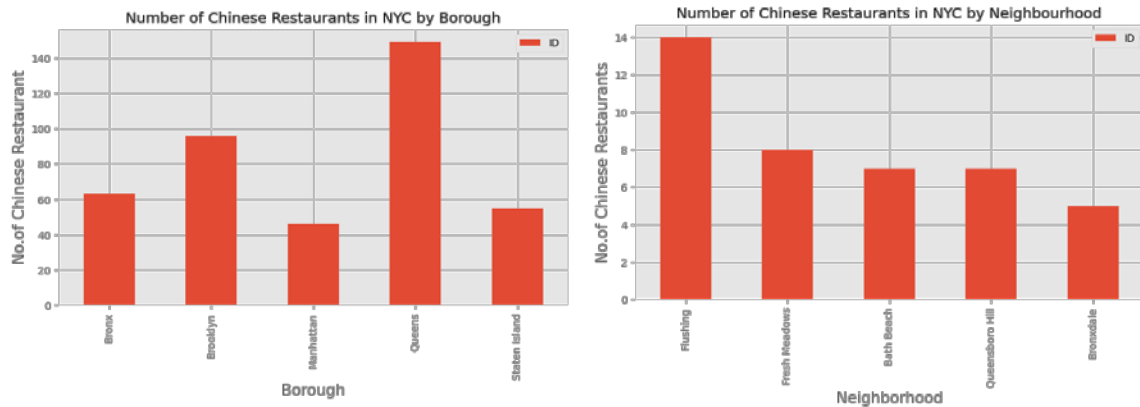


Figure 3 – a) Number of Chinese restaurants in New York City by borough b) by neighborhood.

With the dataset selected we compute the name, likes, rating and tips from each restaurant that its register in the dataset and displaying it and checking what restaurants are in the top of the following categories; likes, tips and rating (figure 4).

```
( 2 / 409 ) processed
Empty DataFrame
Columns: [ID, Name, Likes, Rating, Tips]
Index: []
No data available for id= 4c9d5f2654c8a1cd2e71834b
( 3 / 409 ) processed
```

ID	Name	Likes	Rating	Tips
0 4c66dcfaaebea593955a74d0	Chinese Buffet	7	5.2	10

```
( 4 / 409 ) processed
```

ID	Name	Likes	Rating	Tips
0 4dabc3dc93a04642f09ccabd	Xing Lung Chinese Restaurant	6	6.7	0

```
( 5 / 409 ) processed
```

ID	Name	Likes	Rating	Tips
0 4d965fd061a3a1cde4989a42	Lucky Star Chinese Restaurant	4	7.8	

```
Tips
0 5
( 6 / 409 ) processed
Empty DataFrame
Columns: [ID, Name, Likes, Rating, Tips]
Index: []
No data available for id= 4b97e39ef964a520f21c35e3
( 7 / 409 ) processed
```

ID	Name	Likes	Rating	Tips
0 4e4d0387bd413c4cc66dfd72	Hung Hing Chinese Restaurant	5	6.6	2

```
( 8 / 409 ) processed
```

```
Borough      Bronx
Neighborhood  Fordham
ID            566f33e7498e44c2501bda81
Name          Panda Express
Likes         12
Rating        7.5
Tips          2
Name: 22, dtype: object

Borough      Bronx
Neighborhood  Co-op City
ID            4c66dcfaaebea593955a74d0
Name          Chinese Buffet
Likes         7
Rating        5.2
Tips         10
Name: 3, dtype: object

Borough      Manhattan
Neighborhood  Marble Hill
ID            4d965fd061a3a1cde4989a42
Name          Lucky Star Chinese Restaurant
Likes         4
Rating        7.8
Tips          5
Name: 5, dtype: object
```

Figure 4 – List of all the valuated restaurants in New York. Top restaurants in the categories of likes, rating and tips.

The Chinese restaurant with most of the likes is the one named 'Panda Express', the Chinese restaurant with the biggest rating value is named 'Lucky Star Chinese Restaurant', finally the highest Chinese restaurant with most of the tips is named 'Chinese Buffet'.

In figure 5 we are listing the 10 top rating Chinese restaurant, being the top one valuated with 7.9 value. We want also to display an average rating values by borough to see some trends or changes in the distribution of the valuated restaurants (figure 6).

	Neighborhood	Average Rating
5	Bay Ridge	7.90
110	Marble Hill	7.80
65	Fordham	7.50
139	Parkchester	7.10
176	Unionport	7.10
52	Eastchester	6.70
133	Norwood	6.45
10	Bedford Park	6.45
31	City Island	6.30
131	North Riverdale	5.95

Figure 5 – Top ten rating Chinese restaurants in New York City by neighborhood.

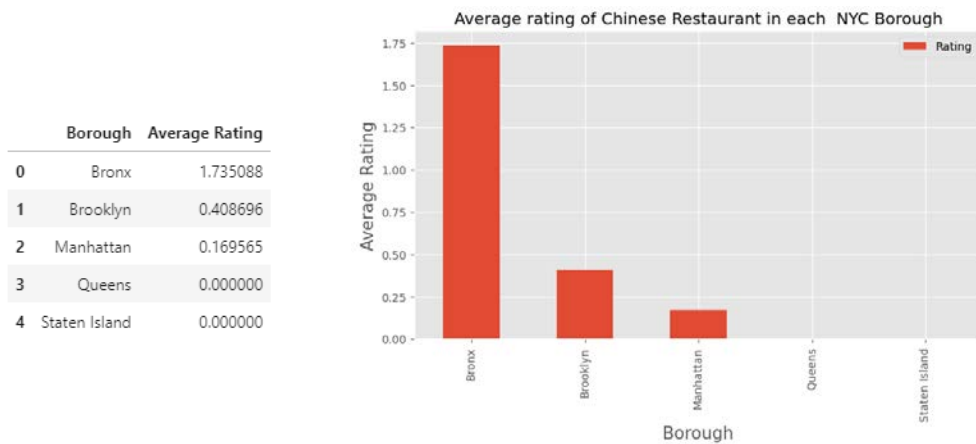


Figure 6- Average rating Chinese restaurant in each New York borough.

To visualize where are the Chinese restaurants we need to combine both dataset information, the first one, with the borough, neighborhoods and its coordinates with the dataset with the rates with the restaurants (figure 7).

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	Borough_y	Latitude	Longitude
0	Manhattan	Marble Hill	4d965fd061a3a1cde4989a42	Lucky Star Chinese Restaurant	4	7.8	5	Manhattan	40.876551	-73.910660
1	Bronx	Pelham Parkway	4b9d6b45f964a52078ab36e3	Mr. Q's Chinese Restaurant	10	7.7	10	Bronx	40.857413	-73.854756
2	Bronx	Fordham	566f33e7498e44c2501bda81	Panda Express	12	7.5	2	Bronx	40.860997	-73.896427
3	Bronx	Parkchester	4b89b62df964a520ff4c32e3	Sabor Latino Seafood Restaurant	5	7.1	8	Bronx	40.837938	-73.856003
4	Bronx	Concourse	4d9a695ee5fd6ea8e3096df5	Lucky House Chinese Restaurant	5	7.0	2	Bronx	40.834284	-73.915589
5	Bronx	Unionport	4b89b62df964a520ff4c32e3	Sabor Latino Seafood Restaurant	5	7.1	8	Bronx	40.829774	-73.850535
6	Brooklyn	Bay Ridge	5553702c498ecb85f99d8f37	XIN	6	7.5	4	Brooklyn	40.625801	-74.030621
7	Brooklyn	Bay Ridge	4b1c3b4cf964a520bb0424e3	Grand Sichuan House	37	8.3	24	Brooklyn	40.625801	-74.030621
8	Brooklyn	Bensonhurst	574a1a53498e6a1b7cb336cb	Orchid Dynasty	6	7.6	1	Brooklyn	40.611009	-73.995180
9	Brooklyn	Bensonhurst	5b5fa5476dcf04003993a767	Mr. Bun 包大人	6	7.2	0	Brooklyn	40.611009	-73.995180
10	Brooklyn	Bensonhurst	4bdcc1043904a593ba9e4f9e	New Ruans	9	7.0	2	Brooklyn	40.611009	-73.995180

Figure 7- new dataset combination between the valuated restaurants and its coordinates.

Finally, we will display these top ten rating Chinese restaurants in the New York city to see the distribution of the most valued restaurants (figure 8).

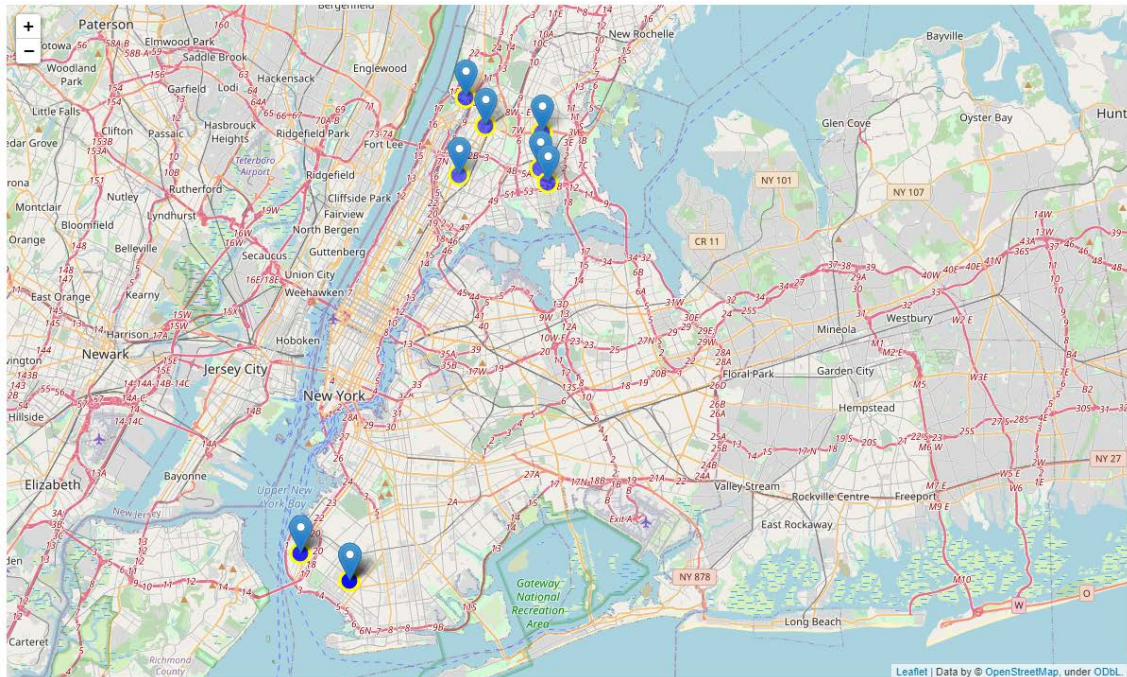


Figure 8 - New York map with the best Chinese restaurants locations.

6 RESULT

Studying the dataset, we found that there are a lot of missing valued Chinese restaurants, so the solution will be quite limited by the data info that we are able to use. Most of the evaluated restaurants are located in the Bronx borough and a most of the restaurants located in the most populated borough, Queens, are without evaluation so it was not possible to study them limiting our study.

The highest valued restaurant is located in Bay Ridge (Brooklyn) with a total value of 8.3 which is also the most liked restaurant with the highest tip values. But it's an exceptional restaurant.

7 CONCLUSIONS

- The most populated borough is Queens which also contains more of the Chinese restaurants in New York but these restaurants are not valued.
- This neighborhood is closed to Brooklyn, which has the best rating Chinese restaurants.
- May a good place to start a new restaurant could be between both boroughs due to its close to the area with the best valued Chinese restaurants and closer to the big populated area of New York.

8 RECOMMENDATIONS

- Required more evaluated info about the Chinese restaurant in the New York city.
- Specified from the stakeholder preferable some important aspect that determine can be aspect that limited the preferable area to open new restaurants such as; customers, neighborhood, budget, quality of restaurant...