Capstone Project – The Battle of Neighborhood

The Optimal Location for a Planned Mexican Restaurant in New York City

Xin Zhang

12-3-2020

1. Introduction

1.1 Project Description

This project is to find an optimal location to open a Mexican restaurant in New York City. To find this location, several factors should be considered, including: 1) the number of the same type restaurants nearby; 2) the population density near the new restaurant.

Less Mexican restaurants nearby means less competition, while higher population density means there are more potential customers of the planned restaurant. They're both essential parts of the restaurant business and will be the very first thing that should be considered.

In this project, we should first determine the candidate borough, and then find out the best neighborhood to open the restaurant.

New York City is composed of five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. From the table below, we can get the population density of each borough.

New York City's five boroughs								
Jurisdiction		**Population**	**Gross Domestic Product**		**Land area**		**Density**	
Borough	*County*	*Estimate (2019)*	*billions (2012 US$)*	*per capita (US$)*	*square miles*	*square km*	*persons / mi^2*	*persons / km^2*
The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04	33,867	13,006
Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42	36,147	13,957
Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13	71,341	27,544
Queens	Queens	2,253,858	93.310	41,400	108.53	281.09	20,767	8,018
Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18	8,157	3,150

Obviously, the population density of Manhattan is much higher than the other boroughs. So, I'll focus on Manhattan and try to find optimal restaurant location in

this area.

After that, all neighborhoods in Manhattan will be considered. By comparing the number of exist Mexican restaurants in each neighborhood and the population density of each neighborhood, we'll obtain 5 best location. This result will be used for the restaurant owner and stakeholders to decide the final location.

1.2 Background

Manhattan is the most densely populated of the five boroughs of New York City. Manhattan contains more than 40 neighborhoods, the population is 1.628 million. There are over 25,000 restaurants and bars in Manhattan, of which over 500 are Mexican restaurants. But these Mexican restaurants distribute unevenly. Some neighborhoods barely have such restaurants. Normally they are either very small neighborhoods or dominated by one ethnicity. In some neighborhoods the situation is quite the opposite. There could be more than two Mexican restaurants within a few miles around. So, our goal is to find the optical location for the new Mexican restaurant, where neither isolated nor crowded.

2. Data Acquisition and Cleaning

2.1 Data Sources

- New York City population density by borough can be downloaded here.
- Manhattan neighborhoods data can be download from the here.
- Mexican restaurants data can be found through Foursquare API.
- Manhattan population by neighborhood data is downloaded here.
- New York City geojson file used to generate map can be downloaded here.

2.2 Data Cleaning

The Manhattan neighborhoods dataset can be generated from newyork_data.json file. Create a query through Foursquare, using "Mexican" as key word to get the venues

data. Download Manhattan population dataset.

There are several problems with the venues data.

First, the venues category contains some non-conformance terms, such as “Mediterranean Restaurant”, “Dessert Shop”, “Cheese Shop”. I checked all the venues category and dropped the non-conformance terms.

The second problem is that some neighborhood names in “Neighborhood-Population” dataset is different from the ones in “newyork_data”. For example, some neighborhoods have more than one name, some neighborhoods with large area are divided into several child-neighborhoods. I unified the names to make sure that the datasets can be merged based on neighborhood name.

After merging all the datasets together, the dataframe contains 5 columns and 43 rows.

3. Methodology

Now we have a dataframe which is 5 columns * 43 rows. The columns are “Neighborhood”, “Number of Restaurants”, “Neighborhood Latitude”, “Neighborhood Longitude” and “Population”.

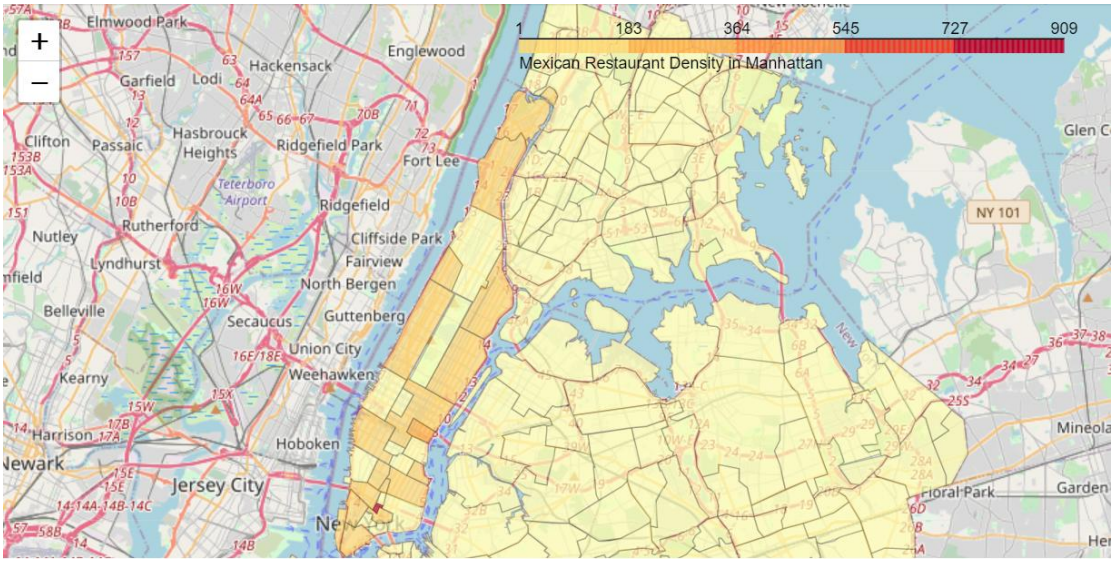
	Neighborhood	Number of Restaurants	Neighborhood Latitude	Neighborhood Longitude	Population
0	noho	34.0	40.723259	-73.988434	24846.000000
1	midtown	26.0	40.754691	-73.981669	391371.000000
2	flatiron	25.0	40.739673	-73.990947	NaN
3	financial district	24.0	40.707107	-74.010665	60976.000000
4	east harlem	20.0	40.792249	-73.944182	115921.000000

To better analyze the potential capacity for consumption in each neighborhood, we need to add a column name “Restaurant Density”, which can be calculate by “Number of Restaurants”/“Population”.

	Neighborhood	Number of Restaurants	Neighborhood Latitude	Neighborhood Longitude	Population	Restaurant Density
0	upper east side	4.0	40.775639	-73.960508	229688.000000	0.000017
1	upper west side	5.0	40.787658	-73.977059	209084.000000	0.000024
2	stuyvesant town	1.0	40.731000	-73.974052	21049.000000	0.000048
3	midtown	26.0	40.754691	-73.981669	391371.000000	0.000066
4	washington heights	11.0	40.851903	-73.936900	158318.000000	0.000069

Then, use python folium library to visualize geographic details of Manhattan and its neighborhoods.

I created a Choropleth map to show the restaurant density of each neighborhood. In order to create Choropleth map, we need a GeoJSON file that defines the areas/boundaries of Manhattan. Since I only find the GeoJSON file of New York City, I used this file in the project. That's why the map below also shows the other boroughs of New York City.



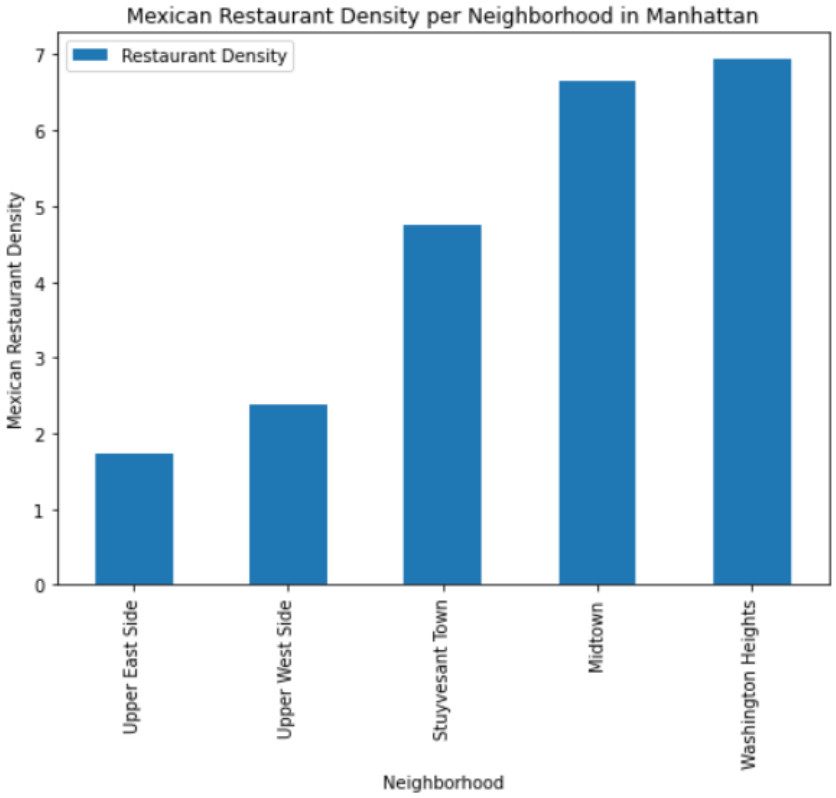
The restaurant density in the table above is between $1.7*10^{-5}$~$908*10^{-5}$, which is too small to be used for legend scale. So I multiplied the value in the column of "Restaurant Density" by 10^5. The revised legend scale is shown in the top right corner.

In the Choropleth map above, the red colored area means there is more Mexican restaurants per person. So, the optimal location should be at the light yellow colored areas.

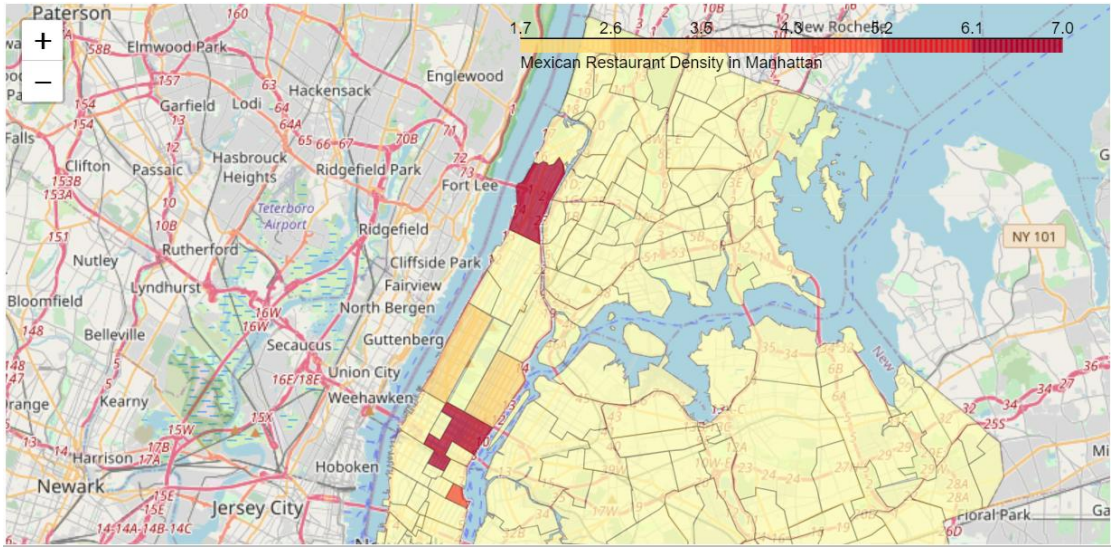
4. Result

As I mention in the first section, the goal of this project is to provide 5 best location to the restaurant owner and stakeholders to make the final decision. The final 5 candidate neighborhoods is shown below.

	Neighborhood	Number of Restaurants	Neighborhood Latitude	Neighborhood Longitude	Population	Restaurant Density
0	Upper East Side	4.0	40.775639	-73.960508	229688.0	1.741493
1	Upper West Side	5.0	40.787658	-73.977059	209084.0	2.391383
2	Stuyvesant Town	1.0	40.731000	-73.974052	21049.0	4.750820
3	Midtown	26.0	40.754691	-73.981669	391371.0	6.643313
4	Washington Heights	11.0	40.851903	-73.936900	158318.0	6.948041



Their locations are also shown in the map.



Among the five neighborhoods, four of the them are relatively large neighborhoods, which has more Mexican restaurants and higher population. Only Stuyvesant Town is a small area and also has less population. Restaurant owner can make a final decision based on these circumstances.

5. Discussion

Because of the limitation of the data sources, some data is missing, for example, the Manhattan population by neighborhoods is incomplete, most of which are small neighborhood. So, I ignored those neighborhoods in the analysis section.

There are many other factors that should be considered when chooses location for a new restaurant. In this project, we only focus on the customers density. In the future, I hope we can gather even comprehensive data and deliver a project that includes more conditions.

6. Conclusion

In this project, I gathered New York City boroughs and neighborhoods data, population data, venues information. I built a method to evaluate the potential capacity for consumption in each neighborhood. I created a map to visualize the restaurant density of each neighborhood. In the end, I gave 5 optimal locations as candidates for restaurant owner and stakeholders to make the final decision.