The Battle of Neighborhoods

- What type of restaurant to open and where -

By Hupfstrudel

Assignment:

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

1. Introduction / Problem:

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

With an estimated population of 8.39 million in an area of 302 square miles, New York City is the most densely populated city in the United States. The City is split into five boroughs (Brooklyn, Queens, Manhattan, Bronx) with a total of 306 neighborhoods. Furthermore, with an overall GMP of 2 trillion US dollars provides endless avenues for potential businesses. However, most of the markets are fairly saturated and as such “bloody”.

Let’s assume a large corporation wants to open a new restaurant in New York City. What types of restaurants are not already oversaturated (e.g. Chinese vs. Indian vs. Italian)? What would be the optimal location for a specific type of restaurant given the demographic of each borough?

2. Data:

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

1. **Datasource:** https://opendata.cityofnewyork.us/
   From „NYC OpenData“ we are able to obtain various data regarding, population, crime rates, health inspections and zip code boundaries

2. **Datasource:** https://geo.nyu.edu/catalog/nyu_2451_34572
   Locational data about the individual boroughs and neighborhoods in New York City.

3. **Datasource:** Geopy
   Just as in the previous assignments we can utilize the geopy package to obtain the latitude and longitude values of New York City and the 5 boroughs.

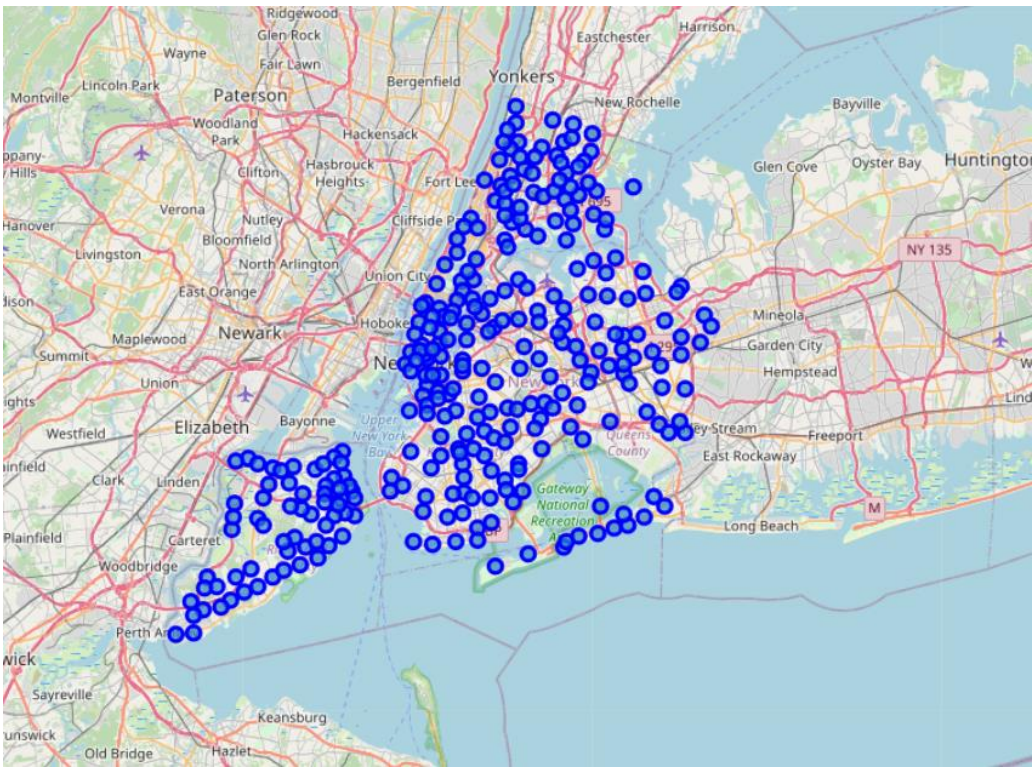
4. **Datasource:** Foursquare API
   The Foursquare API is utilized to obtain the locations and types of all restaurants in New York City

With these 4 datasources at hand we can first put together a descriptive analysis of the individual boroughs and the restaurant market in general. In the following step we will identify our best option of type of restaurant and corresponding borough. Finally, we can utilize a cluster analysis to group the neighborhoods in the chosen borough to identify the best possible location for the new restaurant.

3. Methodology:

Borough Data and Neighborhood Data

I utilized the data from the NY Spatial Data Repository. The link to the free data is: https://geo.nyu.edu/catalog/nyu_2451_34572. For convinience I downloaded the data and am opening is as json before converting it into a pandas dataframe. With the converted dataframe we can utilize the geopy package to find the geographical ccordinates of New York City and use folium to create a map of our first dataframe.



Income and Population Data

Using BeautifulSoup we can cook ourselves a nice meal and scrape the income and population demographics from Wikipedia (https:// en.wikipedia.org/wiki/ Demographics of New York City) and again transfer this information into a pandas dataframe:

	Jurisdiction		Population	Gross Domestic Product		Land area		Density	
	Borough	County	Estimate (2019)[12]	billions(US$)[13]	per capita(US$)	square miles	squarekm	persons / sq. mi	persons /km2
0	The Bronx	Bronx	1418207	42.695	30100	42.10	109.04	33867	13006
1	Brooklyn	Kings	2559903	91.559	35800	70.82	183.42	36147	13957
2	Manhattan	New York	1628706	600.244	368500	22.83	59.13	71341	27544
3	Queens	Queens	2253858	93.310	41400	108.53	281.09	20767	8018
4	Staten Island	Richmond	476143	14.514	30500	58.37	151.18	8157	3150

The city of New York is split into 5 boroughs: Bronx, Brooklyn, Manhatten, Queens and Staten Island. Overall Manhatten has the highest population density and the highest per capita income. Now we can obtain GeoJSON data to visualize the demographics with a choropleth map.

Per capita income:

Population density:

Since Manhatten is already oversaturated with people and the markets are very competitive, we will focus our attention on the borough BROOKLYN which has the second highest population density and the third highest per capita income. Since the market for reastaurants is very likely not over saturated and rent fees will be lower compared to Manhatten, Brooklyn offers a suitable location to open a new restaurant.

Foursquare Data about Restaurants

With the Foursquare API we can obtain data about certain venues in New York City, especially in Brooklyn. First we have to identify our area of interest. For us the “FOOD” category is most interesting, so lets list all the different venue types in the “FOOD” category. We can easily see that this list is quite crowded, thus we will decide to focus on the different types of restaurants. However, we will exclude generic types such as “Fast Food Restaurant etc”.

Venue Category	
Fast Food Restaurant	643
Chinese Restaurant	638
Italian Restaurant	473
Mexican Restaurant	374
Caribbean Restaurant	351
American Restaurant	315
Restaurant	207
Sushi Restaurant	163
Seafood Restaurant	147
Latin American Restaurant	144
Japanese Restaurant	143
Asian Restaurant	142
Spanish Restaurant	130
Thai Restaurant	106
Indian Restaurant	104
Southern / Soul Food Restaurant	69
Korean Restaurant	62
Greek Restaurant	58
Middle Eastern Restaurant	56
Vegetarian / Vegan Restaurant	54
New American Restaurant	52
Ramen Restaurant	39
French Restaurant	39
Mediterranean Restaurant	38
Vietnamese Restaurant	38
Peruvian Restaurant	37
Turkish Restaurant	29
Halal Restaurant	29
South American Restaurant	28
Filipino Restaurant	26
African Restaurant	26
Cuban Restaurant	23
Tapas Restaurant	23
Dim Sum Restaurant	22
Eastern European Restaurant	20
Empanada Restaurant	19
Cantonese Restaurant	16
Dumpling Restaurant	16
Comfort Food Restaurant	15
Falafel Restaurant	13

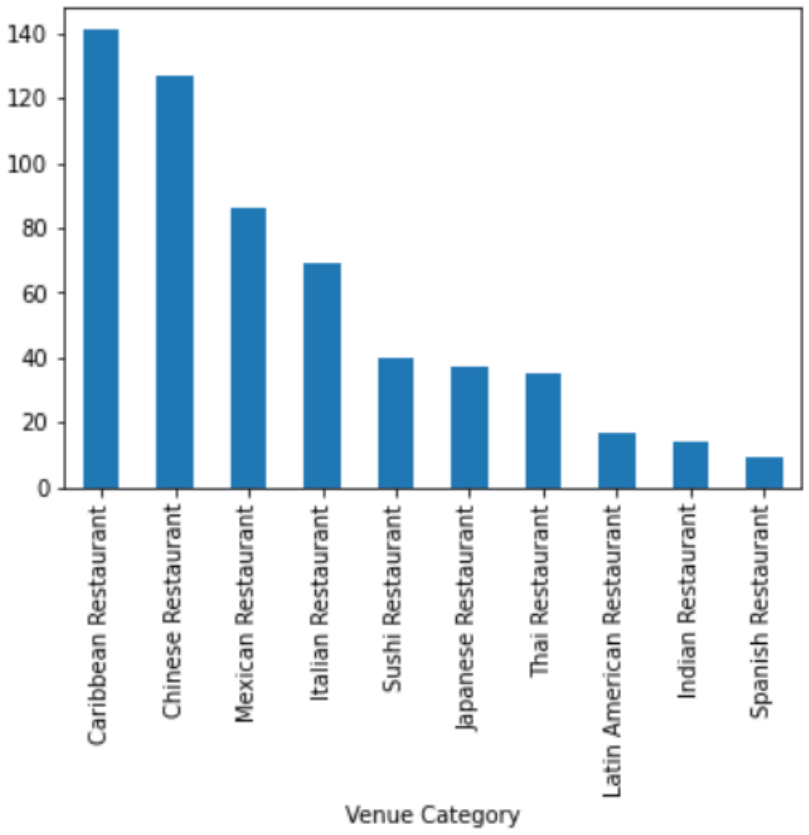
Name: Venue Category, dtype: int64

```
Venue Category
Chinese Restaurant            638
Italian Restaurant            473
Mexican Restaurant            374
Caribbean Restaurant          351
Sushi Restaurant              163
Latin American Restaurant     144
Japanese Restaurant           143
Spanish Restaurant            130
Thai Restaurant               106
Indian Restaurant             104
Name: Venue Category, dtype: int64
```

With the reduced list we can notice that country specific types of restaurants are very popular. At the top are Chinese restaurants with 638 unique restaurants in New York city. Given the high popularity of Chinese and Italian restaurants we choose to focus on opening a Mexican restaurant in Brooklyn. Hispanics provide the second highest demographic in New York City's population, thus this seems to be a good choice and potential market.

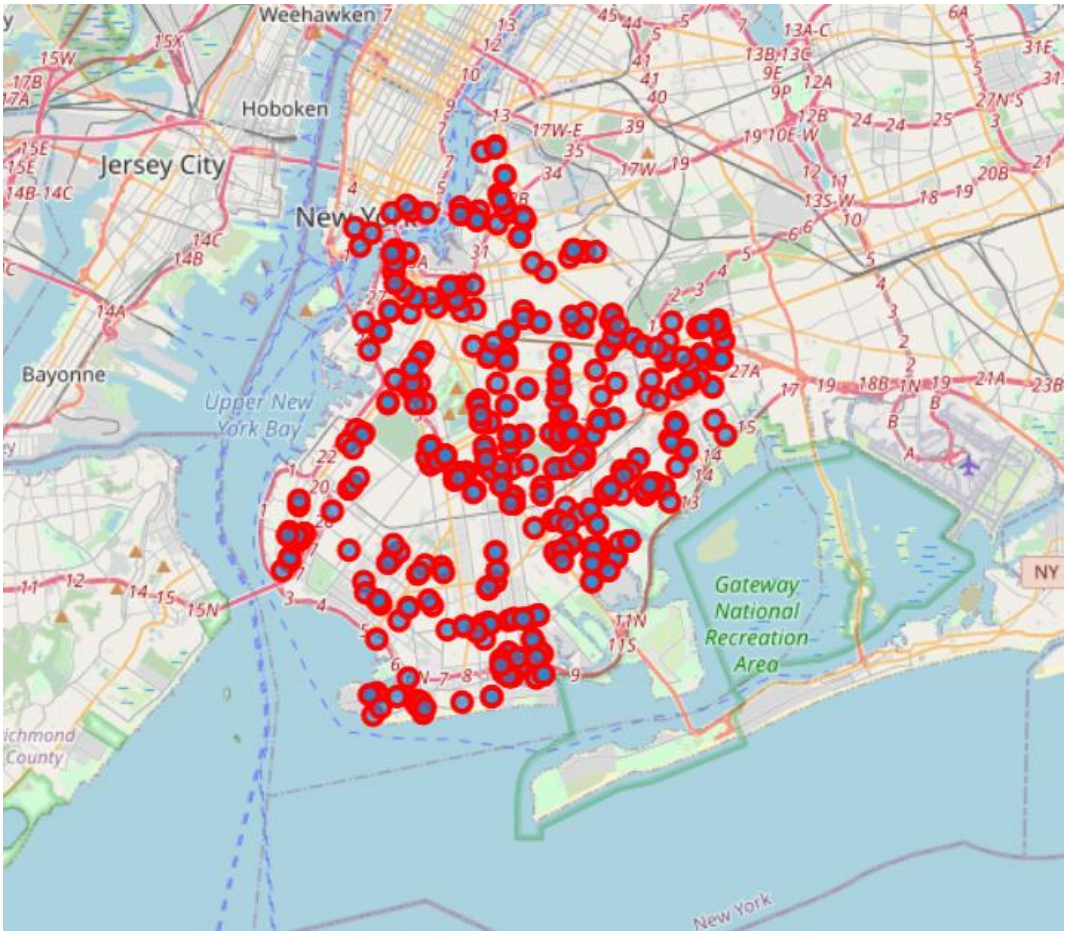
Analyzing venues in Brooklyn

Since we already decided to focus on Brooklyn let's analyze the top10 types of restaurants in Brooklyn:



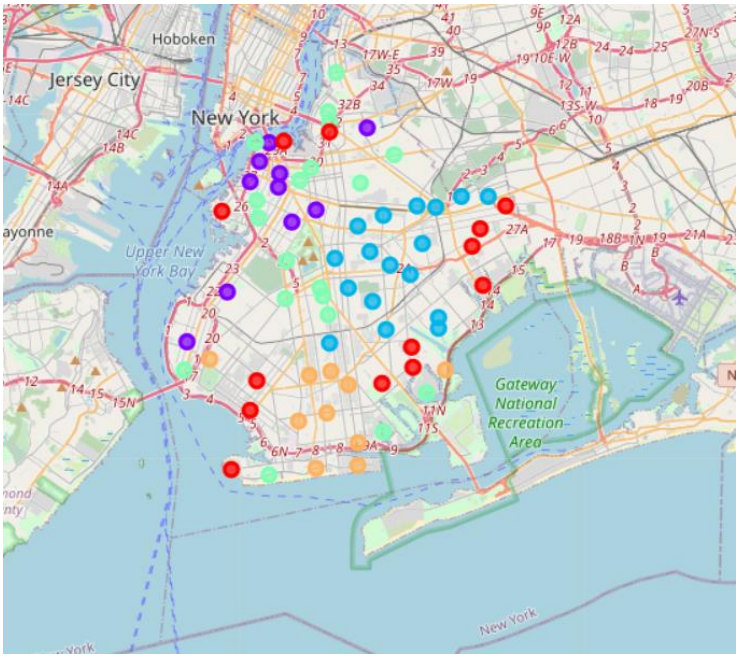
Surprisingly Caribbean restaurants are more popular in Brooklyn compared to the entire New York City averages. Again we see that Mexican restaurants are third famous, thus the market will potentially not be as competitive. And our choice is still robust.

To get a better feeling of the perfect location for our restaurant let us visualize all restaurants in the top 10 categories in Brooklyn:



Clustering

Finally, we will utilize k-means clustering to find the perfect neighborhood for our Mexican restaurant. With the elbow method we identify the optimal amount of clusters k=5. And merge our analysis with the geographical data:

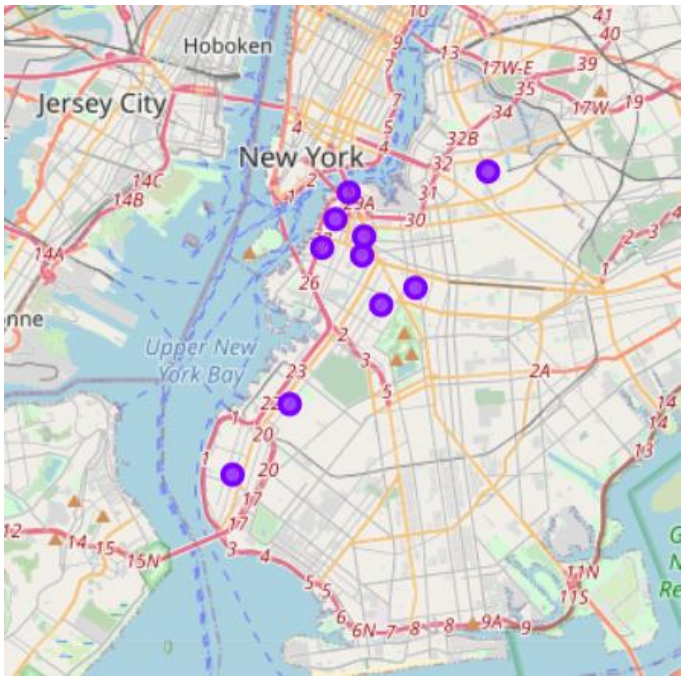


4. Results:

Given the data for the 5 clusters. We can identify the first cluster as our perfect neighborhood:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Borough	Latitude	Longitude
1	1	Bay Ridge	Mexican Restaurant	Japanese Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Brooklyn	40.625801	-74.030621
5	1	Boerum Hill	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Latin American Restaurant	Brooklyn	40.685683	-73.983748
8	1	Brooklyn Heights	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Latin American Restaurant	Brooklyn	40.695864	-73.993782
15	1	Cobble Hill	Mexican Restaurant	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Brooklyn	40.687920	-73.998561
20	1	Downtown	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Latin American Restaurant	Brooklyn	40.690844	-73.983463
21	1	Dumbo	Mexican Restaurant	Italian Restaurant	Japanese Restaurant	Thai Restaurant	Sushi Restaurant	Brooklyn	40.703176	-73.988753
25	1	East Williamsburg	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Spanish Restaurant	Brooklyn	40.708492	-73.938858
52	1	Park Slope	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Latin American Restaurant	Brooklyn	40.672321	-73.977050
53	1	Prospect Heights	Mexican Restaurant	Thai Restaurant	Japanese Restaurant	Sushi Restaurant	Spanish Restaurant	Brooklyn	40.676822	-73.964859
63	1	Sunset Park	Mexican Restaurant	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Brooklyn	40.645103	-74.010316

Since in the neighborhoods in this cluster the type Mexican restaurant is the number one most popular venue. Thus the ideal spot for our restaurant ist one of the following locations in New York City:



5. Conclusion:

All in all, we find that our best option is to open a Mexican restaurant in the borough Brooklyn in the first cluster, so in the area between the neighborhoods from Bay Ridge to Sunset Park.