

By Jens Greve
January 23

Table of content

Cover

Introduction

Business problem description

Target audience

Data

Data sources

Methodology

Results

Discussion

Conclusion

Appendix – link to notebook

Introduction

New York City is the largest city in US with 8.4 million people according to Wikipedia and with a wide and diverse range of ethnicities. It also has a large history of immigration and 48.6% of NY citizens are speakers of a non-English language according to <https://datausa.io/>. And New York Times has claimed that New York is the most diverse city in the world with 800 languages spoken in New York. So it has a long tradition of different ethnical restaurants.

The final project aims to recommend the best location to open a Mexican restaurant in New York City. Obviously, much data is needed to really make a solid analysis, e.g. including availability and price of rental space for restaurant purpose, robberies, buying power etc. But for simplicity for this project, we will limit to analyze the availability of Mexican restaurants in New York and their attributes such as ratings and likes, combine it with demographic data and make some discussions of how the data could be used to find the best location for a Mexican restaurant in the area of New York city.

The Mexican cuisine has been raising in popularity in USA for many years, while it has always been popular with people of Hispanic origin. So the thesis is to find a borough or neighborhood with many Hispanic people, few or poor rating Mexican restaurants. Alternatively, an area with many Mexican restaurants which could be taken as an interest for Mexican food in this area.

So, this project aims to analyze the ethnical demographics of the New York boroughs and the availability and ranking of Mexican restaurants in boroughs and neighborhoods to understand the competition and decide the best location to open a Mexican restaurant in New York.

Business problem description

The objective of this Capstone project is to retrieve location-based data from New York City related to demographics and restaurants and analyze it in order to select the best locations in New York city to open a Mexican restaurant. The project will use data science retrieving, analyzing and visualization methodologies taught in the Coursera course.

Questions aimed to be answered in the report:

- In which boroughs of New York do people of Hispanic origin live?
- Which neighborhoods or boroughs have the largest number of Mexican restaurants and/or the highest ratings / most likes?
- Which kind of Mexican restaurants exist and how do they rate? How does the big chains compare to the local, authentic shop?

- What is / are the best location(s) to open a Mexican restaurant in New York City?

Target audience of this report

The target audience for this report is potential business owners that consider opening or invest in a food business in New York City. New York city may already have a ton of restaurants, but it is also a vibrant city with people from many ethnical backgrounds. So New Yorkers have an open mind towards ethnical cuisines and new culinary experiences. With its diverse culture comes diversity in restaurants. New York have many restaurants in different categories like Italian, Chinese, Japanese etc.

The idea for a Mexican restaurant came as the Mexican cuisine has been one of the fastest growing cuisines in the world, and besides being very tasty, much Mexican food is healthy. So, an authentic Mexican restaurant is to likely appeal to the Hispanic population while also offering a healthy alternative to the common American eating habits.

Data

To answer the above questions, we need to following New York City data:

- List of neighborhoods and boroughs of New York city
- Latitude and longitude of the neighborhoods and boroughs
- Venue data of the Mexican restaurants with attributes like rating and likes
- Demographic data of ethnical origin by borough

Data Sources

- New York City data ethnical orgin of New York will be obtained at this web site: https://en.wikipedia.org/wiki/Demographics_of_New_York_City
- New York City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source: https://cocl.us/new_york_dataset
- All data related to locations and quality of Mexican restaurants will be obtained via the FourSquare API utilized via the Request library in Python.

Methodology

The project required using many data analysis and science skills taught in the Coursera course, e.g. web scraping, retrieve JSON data from an API, data cleaning and data wrangling to visualize data as bar charts or as maps.

- Data will be scraped from a table from https://en.wikipedia.org/wiki/Demographics_of_New_York_City and cleaned, sorted and put into a Panda dataframe.
- The demographics data will be visualized using bar charts and displaying Hispanic population by borough and by total Hispanic population.
- Data will be collected from https://cocl.us/new_york_dataset and cleaned, sorted and processed into a Panda dataframe.
- The data will be enriched with GPS coordinates using the Python geocoder package
- FourSquare APIs will be used to locate all venues and then filtered by Mexican restaurants, incl. their ratings, likes and tips and added to the Panda dataframe.
- Venue data will be sorted based on sum, ranking, borough and neighborhood.
- Finally, the data will be visually assessed using graphing from Python libraries such as bar charts and mapping the venue on a map using the Folium library.
- No machine learning is used in this project

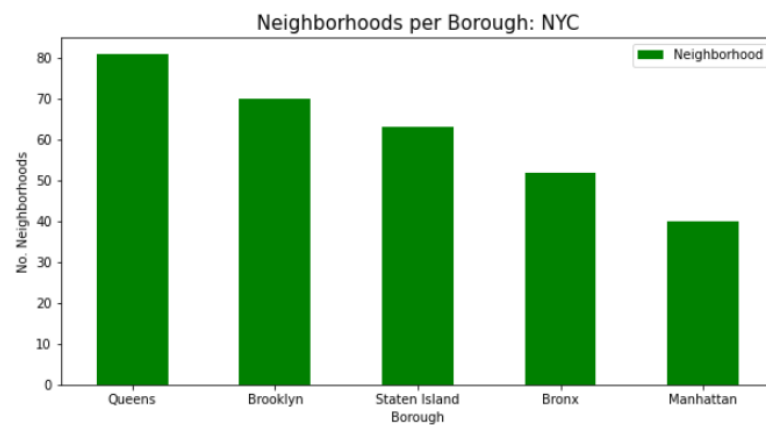
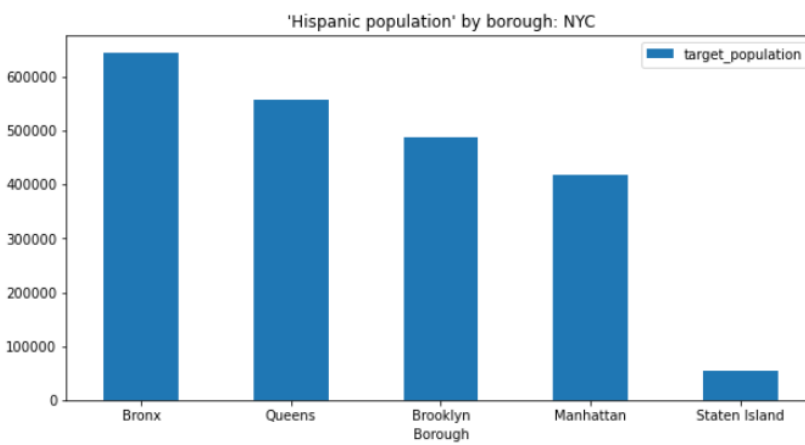
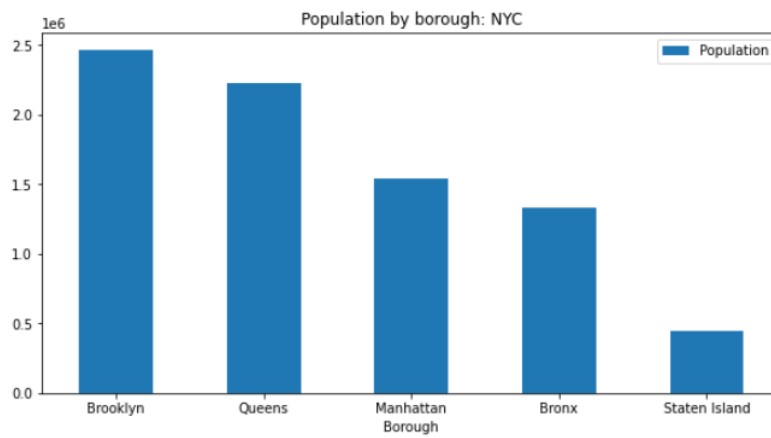
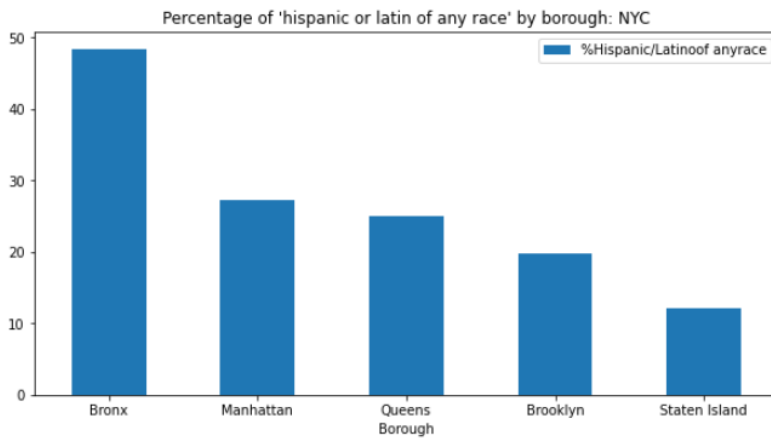
Results

The results of our analysis below.

Demographics by borough

We can see that Bronx has the highest percentage of Hispanic people – almost twice as large as the second largest, but Bronx has the second smallest population. Factoring in the population size and percentage, Bronx still has the largest Hispanic population, but the other boroughs have a very Hispanic population as well.

Queens and Brooklyn have the highest number of neighborhoods.



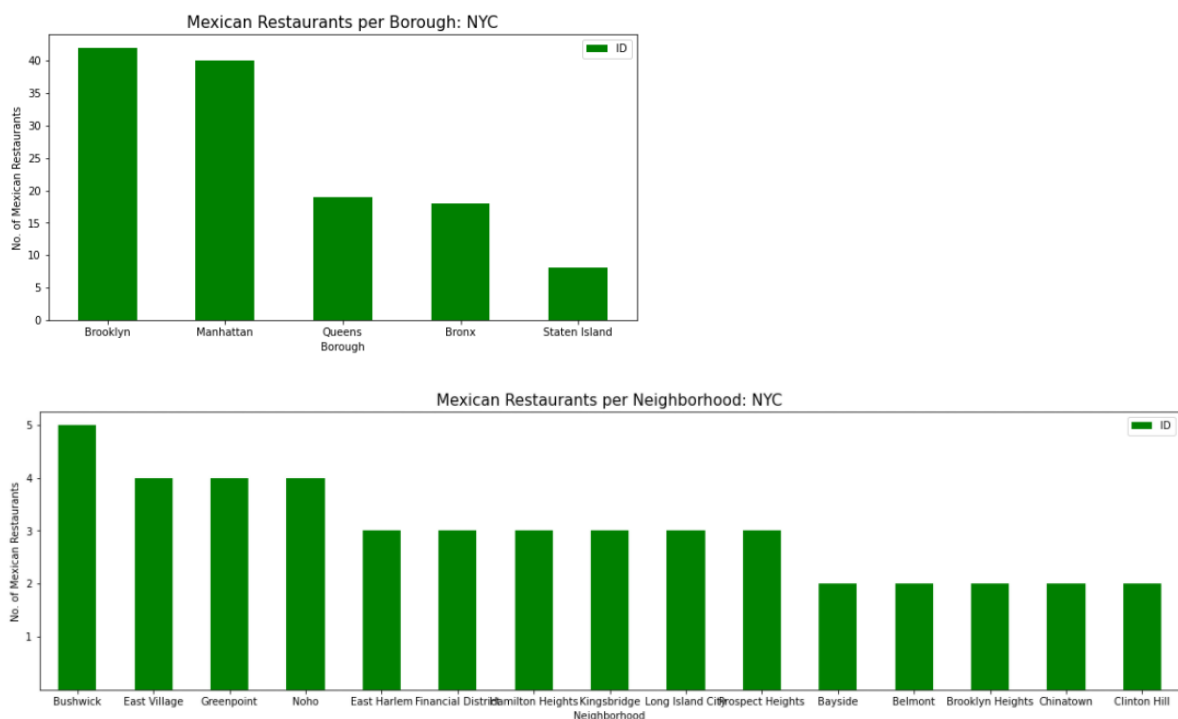
Mexican restaurants by borough and neighborhood

We can see that Brooklyn has the highest number of Mexican restaurants, closely followed by Manhattan.

Bronx with the highest number of Hispanic population only have 50% of the number of Mexican restaurants in both Brooklyn and Manhattan.

The 2 neighborhoods in Brooklyn with the highest number of Mexican restaurants are Bushwick and Greenpoint.

The 2 neighborhoods in Manhattan with the highest number of Mexican restaurants are East Village and Noho.



We can also notice that Bushwick seems to have small authentic Mexican restaurants and Bronx have many from the big chain Chipotle.

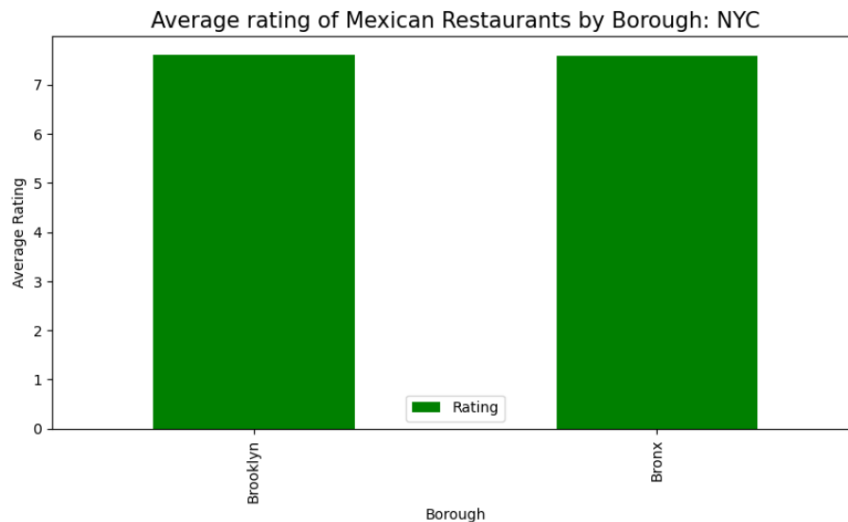
98	Queens	Jamaica Center	5a74e9b8a4c39c7583d2c6aa	Chipotle Mexican Grill
1	Bronx	Kingsbridge	553d5376498e322eb4d37a1b	Chipotle Mexican Grill
8	Bronx	Westchester Square	5e7dc30676db98000832f0d6	Chipotle Mexican Grill
10	Bronx	Belmont	5529d964498ec65bfc27ad91	Chipotle Mexican Grill
114	Bronx	Concourse Village	58d3f7f79435a92450b97b0d	Chipotle Mexican Grill
86	Queens	Corona	5982f0d5f5e9d71716cdc9d5	Cienega Las Tlayudas de Oaxaca Mexican Cuisine
95	Queens	Bayside	4babe1e3f964a520e6d13ae3	Cinco De Mayo

Unnamed: 0	Borough	Neighborhood	ID	Name	
29	29	Brooklyn	Bushwick	54a5a765498e6b7e2ce128c8	El Kucho
30	30	Brooklyn	Bushwick	4a79c17ef964a52093e71fe3	El Sol de Cholula
28	28	Brooklyn	Bushwick	507c9d44e4b07390823b8d98	Taqueria Sofia
32	32	Brooklyn	Bushwick	49d7cb08f964a520625d1fe3	Zefe's
31	31	Brooklyn	Bushwick	4dd6d110814d85e931ff48e4	Zefe's Mexican Restaurant

Mexican restaurants by ratings, likes and tips

NOTE: Fourthsquare has a limit of 50 premium API calls per day in sandbox, so we were unable to retrieve all data

I was unable to retrieve data for all 127 Mexican restaurants due the constraint, so I continue my analysis using the retrieved data which fortunately includes Brooklyn data. We can see the average rating between Bronx and Brooklyn is about the same.



We can see that Greenpoint in Brooklyn has almost the same average rating as Bushwick, but significantly more Likes and Tips.

Average rating		Average tips		Average Likes	
Neighborhood		Neighborhood		Neighborhood	
Parkchester	9.100000	Parkchester	73.000000	South Side	194.000000
North Side	8.400000	South Side	71.000000	Parkchester	131.000000
Westchester Square	8.050000	Sunset Park	49.500000	Greenpoint	97.000000
Prospect Heights	8.033333	Downtown	39.000000	Downtown	88.000000
South Side	8.000000	Clinton Hill	38.000000	Sunset Park	84.500000
Brooklyn Heights	7.950000	Windsor Terrace	36.000000	Prospect Heights	81.666667
Bushwick	7.900000	Greenpoint	29.750000	Clinton Hill	46.000000
Sunset Park	7.850000	Prospect Heights	29.000000	Belmont	45.000000
Kingsbridge	7.833333	East Williamsburg	22.500000	Windsor Terrace	40.000000
Kensington	7.800000	Belmont	21.000000	Brooklyn Heights	37.500000
Greenpoint	7.700000	Bushwick	15.500000	Bushwick	36.750000
Park Slope	7.650000	Westchester Square	14.500000	East Williamsburg	27.500000
Georgetown	7.600000	Flatbush	11.000000	North Side	26.000000
Prospect Park South	7.500000	North Riverdale	8.000000	Kingsbridge	24.666667
Flatbush	7.400000	Kingsbridge	7.333333	Westchester Square	23.000000
Belmont	7.300000	North Side	6.000000	Bay Ridge	19.000000
Clinton Hill	7.200000	Bedford Park	4.000000	Flatbush	17.000000
Schuylerville	7.200000	Bay Ridge	4.000000	Park Slope	15.000000
Bedford Park	7.100000	Kensington	3.000000	Prospect Park South	12.000000
Downtown	7.100000	Melrose	3.000000	North Riverdale	9.000000
Melrose	6.900000	Brooklyn Heights	3.000000	Melrose	8.000000
East Williamsburg	6.750000	Schuylerville	3.000000	Bedford Park	6.000000
Windsor Terrace	6.700000	Prospect Park South	2.000000	Kensington	5.000000
Bay Ridge	6.700000	Park Slope	1.500000	Schuylerville	5.000000
North Riverdale	6.500000	Georgetown	0.000000	Georgetown	2.000000
Name: Rating, dtype: float64		Name: Tips, dtype: float64		Name: Likes, dtype: float64	

Finally, we mapped the Brooklyn restaurants on a map using the Folium library.

You can click on each restaurant on the map and the borough, neighborhood and average rating.



Discussion

From the results and my analysis, I would say that Manhattan and Brooklyn are the best places to eat Mexican food in New York City. To our surprise Bronx does not have many Mexican restaurants, but Bronx has the same average rating as Brooklyn, but lower tips and likes. Both Brooklyn and Manhattan have more x2 the number of restaurants compared with Bronx. Manhattan and Brooklyn have the highest number of Mexican restaurants and Manhattan has the highest rating by a wide margin. Bronx has the lowest rating almost 20% lower rating than Manhattan.

So, this could suggest that a Hispanic population does not equal a taste for Mexican food. Perhaps these people cook at home or prefer or other types of Latin cuisine, e.g. food from Puerto Rico or Dominican Republic.

We could also notice that the chain Chipotle has many restaurants, but that they have an average rating. This would suggest there could be an opening for a real authentic Mexican cuisine to out-compete the chain's restaurants.

One could argue that Manhattan with almost as many Mexican restaurants as Brooklyn and the influx of people working daily in Manhattan, would be the best choice for a Mexican restaurant, but I would rate Brooklyn as the best location to open a Mexican restaurant. Compared to Manhattan, it has a larger total population

and a larger Hispanic population and the lower average rating than Manhattan makes it easier competition. We should also consider that Brooklyn real estate prices are much cheaper than Manhattan.

I would argue that both Bushwick and Greenpoint are good locations and both have restaurants with top ratings and with lower ratings so there is room to out compete the poorly rated restaurants. Greenpoint rates better than Bushwick for Likes and Tips, so that put **Greenpoint as the best location to open a Mexican restaurant.**

However, I would also recommend to open a genuine, real Mexican restaurant.

Finally, I would go to Oxomoco in Brooklyn for the best Mexican cuisine with 9 (event though the best scores 9.1), it has 250 likes.

Limitations and ideas for further research

First, our data primarily came from Fourthsquare and it has a limitation of 50 premium API calls per day in sandbox, so we were unable to retrieve all data and can question the accuracy of the acquired data. To get better results, we would need more data, e.g. real estate prices, restaurant size, price, profit and citizens buying power and use a Fourthsquare paid account to bypass any limitations.

Conclusions

In this Capstone project, I have followed a data science methodology as taught in the Coursera course, incl. identifying a business problem, analyzing which data source are required, retrieving the data, cleaning it, and performing analysis along with visualizations. And as the last step I have provided a discussion and recommendation for the business problem.

So, our conclusion is that Brooklyn is the best borough and Greenpoint the best neighborhood to open a Mexican restaurant.

Appendix – link to notebook of Capstone project

https://github.com/mrgreve/Data_Science_Coursera_Capstone/blob/master/Capstone%20Project%20-%20Opening%20a%20Mexican%20restaurant%20in%20NYC.ipynb