

Capstone Project: Deciding the Next Location of a Popular Chain Restaurant

By Chloe Mandel



Introduction

Background:

A growing Italian restaurant chain is looking to open a new location. So far, restaurants have been located in the Canadian cities of Quebec, Quebec City, Vancouver, and Ottawa. Now, the restaurant's founder is thinking about expanding to the US through a New York location. However, she's wondering if she should hold off on an international expansion for now and instead locate the new restaurant in Toronto. Moreover, once she decides between New York and Toronto, she needs to select the best neighborhood to open the restaurant in.

Business Problem:

In order to answer the question, "Where should the new restaurant in the growing Italian restaurant chain be located?", the founder needs access to some key information about Italian restaurants in the cities of New York and Toronto. This includes the amount of Italian restaurants in each city relative to their size, the individual neighborhoods with the most/least Italian restaurants, and the rankings of said restaurants.

Importance:

It is crucial for my client to access data on existing Italian restaurants in these cities because learning from this information is the best way to come to a decision that will ensure the success of the new restaurant. My client would not want to choose the new location without knowing that the choice she is making will likely lead to higher profits for the franchise. Therefore, she would be interested in viewing data that will indicate which city/neighborhood is the most economically viable choice.

Data

In order to solve my business problem, I needed to use the following data:

- Data containing New York boroughs, neighborhoods, longitudes and latitudes from IBM Developer Skills Network
- Data containing Toronto boroughs, neighborhoods, and postal codes from Wikipedia
- Geographic data from Nominatim to show New York and Toronto neighborhoods on a map through Folium
- Data about venues in New York and Toronto (location, venue category, rating, etc.) from the Foursquare API

As part of my analysis, I scraped data containing New York and Toronto's neighborhoods and boroughs from open data sources. I then found the longitude and latitude of these locations through the Python Geocoder package. In order to obtain information about Italian restaurant venues, I used the Foursquare API, which provides data about venue locations, rankings, ratings, and user comments, among other things.

Methodology

Data Scraping

- I first scraped the Toronto neighborhood/borough/postal code data from Wikipedia and cleaned the data to be in the format I wanted it to be in
- I then obtained the longitude and latitude of each Toronto neighborhood and added these to a dataframe

Exploration and Visualization

- I defined functions to obtain data about Toronto venues from Foursquare
- I made a dataframe of these venues and filtered it to show only Italian restaurants
- I obtained the number of Italian restaurants per square mile in Toronto, then visualized the number of Italian restaurants in each Toronto borough and neighborhood using bar charts

- I used Folium to visualize the restaurants on a map of Toronto
- I then defined a function to get venue details from Foursquare
 - I used this to get details about each Italian restaurant and found the average rating of Italian restaurants in Toronto

New York

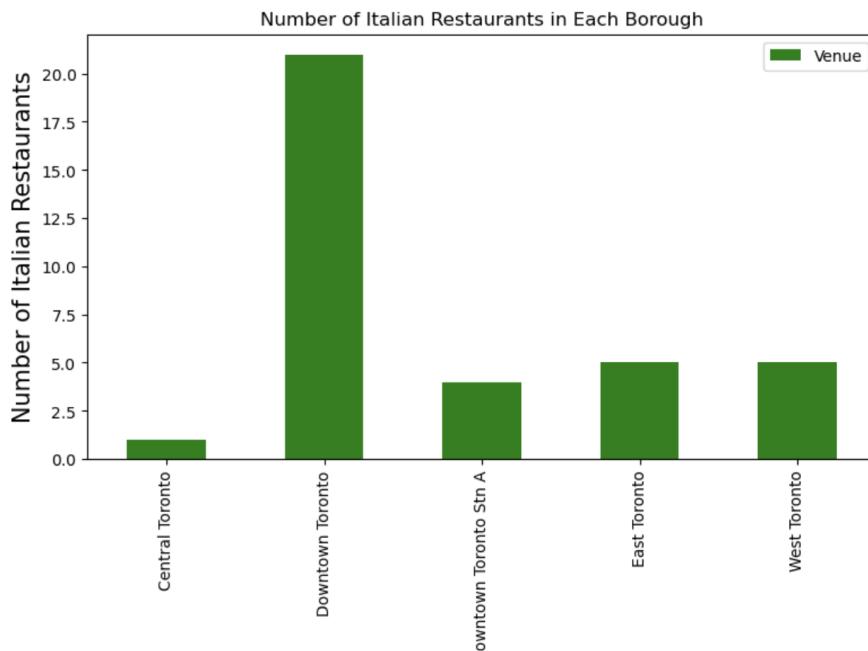
- I repeated the above data scraping, cleaning, visualization and exploration for the city of New York

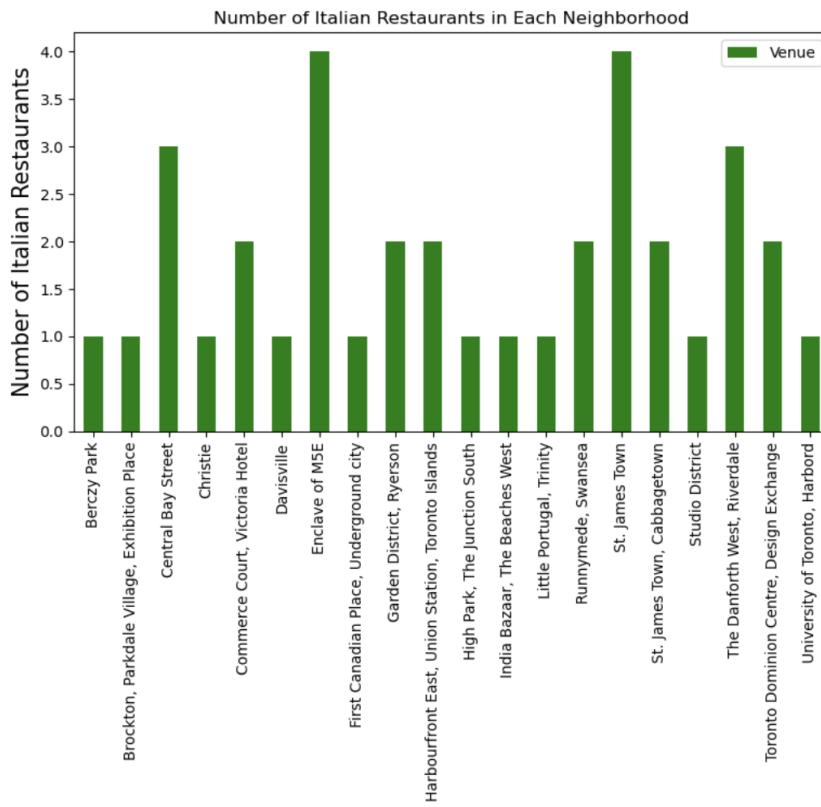
Recommendation

- I used the results from my data exploration to recommend where the restaurant chain owner should locate the new restaurant

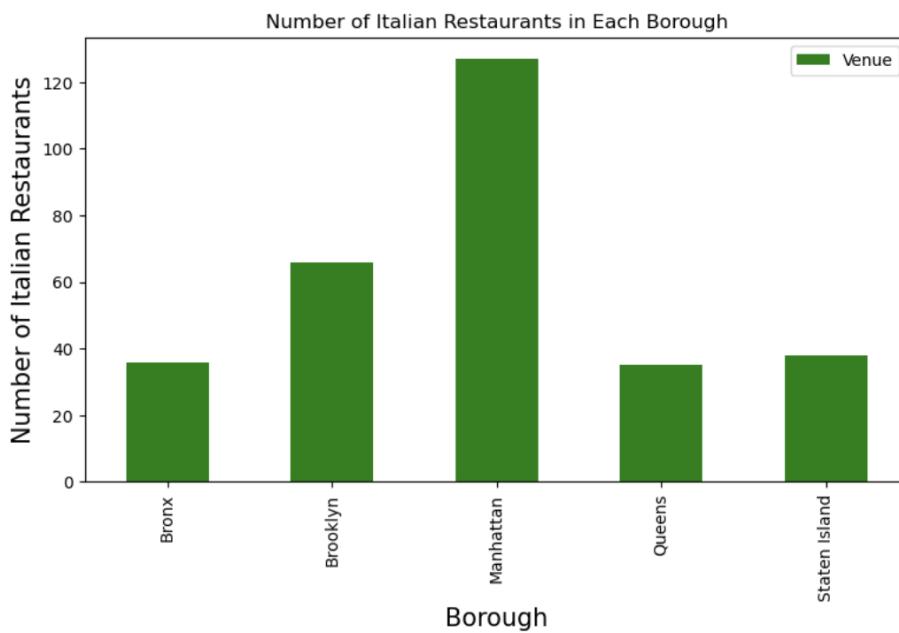
Results

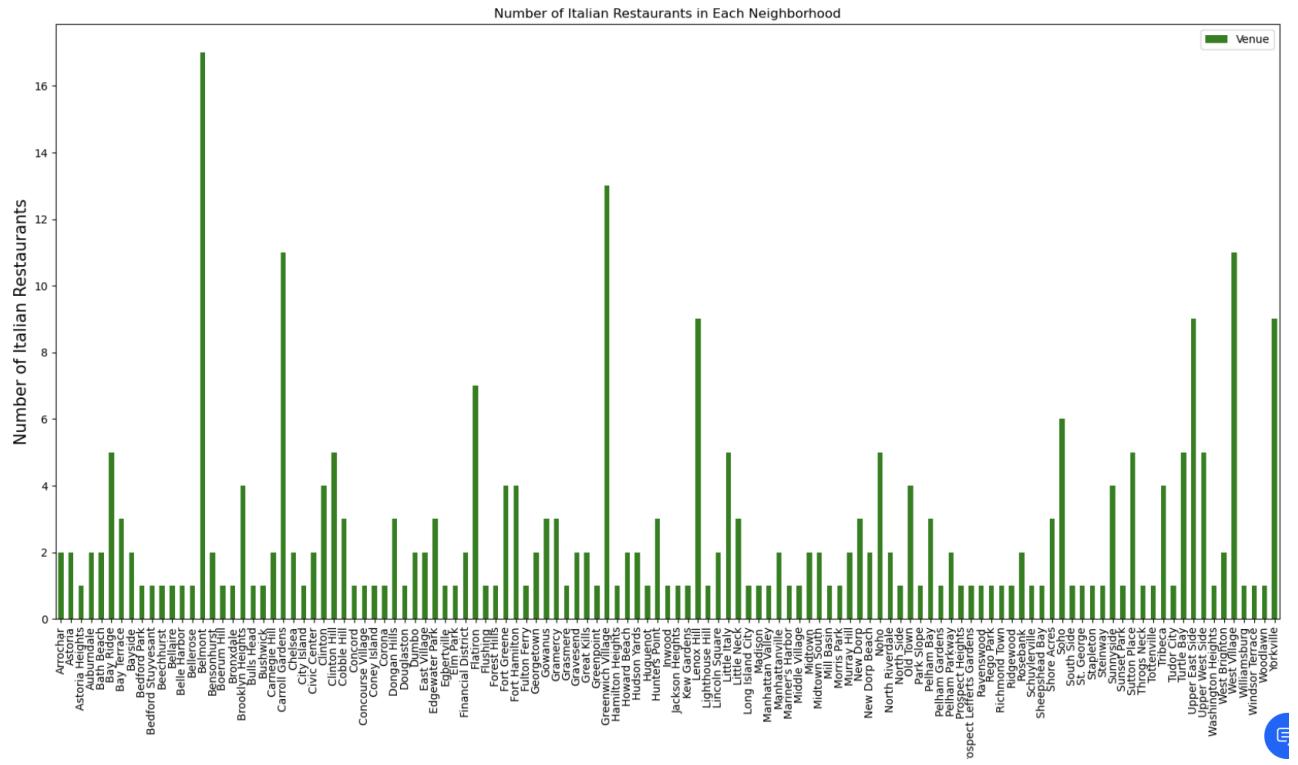
- I found the number of Italian restaurants per borough and neighborhood for Toronto:





- And New York:





- I found the number of Italian restaurants in Toronto to be ~0.15 Italian restaurants per sq. mile with an average rating of ~7.5
 - I found the number of Italian restaurants in New York to be ~0.1 Italian restaurants per sq. mile with an average rating of ~7.8

Discussion

Question #1: Which city should the restaurant be located in?

While Toronto has ~0.15 Italian restaurants per sq. mile and an average rating of ~7.5, New York has ~0.1 Italian restaurants per sq. mile and an average rating of ~7.8. So even though New York has many more Italian restaurants than Toronto, indicating more competition for the Italian restaurant chain owner, the higher average rating suggests that these restaurants are doing well and that there is a demand for Italian food. Based on these stats alone, it seems that the owner should consider locating the restaurant in New York. However, it would also be helpful for the owner to know other information such as the average price to rent a restaurant space in each city, what it might cost to hire workers in each city, whether new restaurants tend to survive in each city, and even how COVID has affected restaurant popularity/success in each city.

Question #2: Which neighborhood should the restaurant be located in?

Should the owner locate the new restaurant in New York, she has some options for specific locations. She could potentially put the restaurant in one of the boroughs that has the least Italian restaurants (Queens, the Bronx, or Staten Island), but in a neighborhood such as Astoria or North Riverdale that has a higher than average number of Italian restaurants for those boroughs. This would allow her to face less overall competition from other restaurants in the borough, but still be in a neighborhood where Italian food is in demand. Alternatively, she could locate the restaurant in Manhattan or Brooklyn (these are the boroughs with the most Italian restaurants) in a neighborhood that has a lower than average number of Italian restaurants, such as Chelsea or Park Slope. This way, she'd face less competition within the neighborhood, but be in a borough where Italian restaurants tend to survive and thrive.

Limitations of my analysis:

Because I have a free account with the Foursquare API, I could only make a certain number of calls per day, which limited the extent to which I could explore and analyze data about the Italian restaurants in each city. For future analysis, getting more data on the likes, tips and ratings of each restaurant would be helpful for solving this business problem. It would also be helpful to use data from other sources in order to reduce dependency on the Foursquare data, which may not be fully up-to-date/accurate.

Conclusion

For this project, I identified a business problem and the data used to solve it, scraped and cleaned the data, visualized and analyzed the data, performed data exploration using the Foursquare API, and made recommendations to a business owner based on my findings. These recommendations should help my client choose which city -- and which neighborhood within that city -- would be the most economically viable location in which to place the new restaurant in her Italian restaurant chain.