Coursera Capstone Course: IBM Data Science Professional

Introduction and Business Problem

The goal of the Coursera Capstone for the IBM Data Science Professional Certification is to utilize Data Science to solve a real-life issue. For this final Coursera Capstone project, we are solving the problem of a restauranteur in New York City. This restauranteur is looking for the best location to open their new fine-dining Italian restaurant. The restauranteur needs to know how many other Italian restaurants are in the are which would lead to competition and where he would find success with an Italian restaurant.

The results of this analysis will assist the restauranteur when he pitches his restaurant idea and location proposal to potential investors.
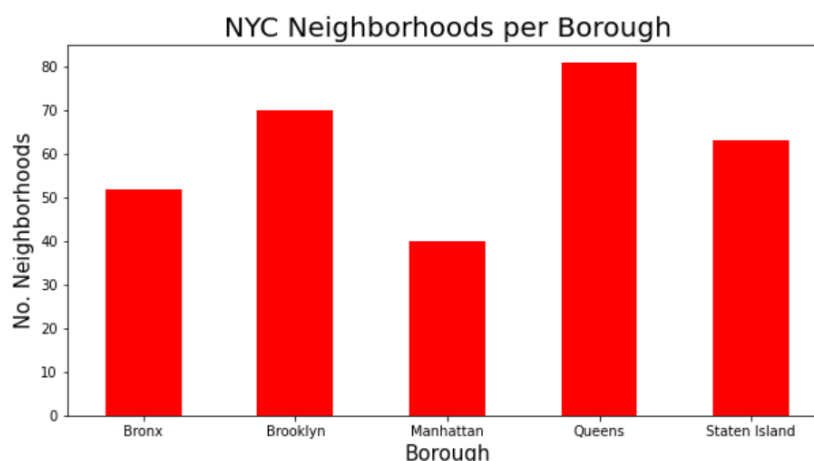
Data

For this problem, we will use data pulled from the Foursquare API to search the neighborhoods of New York City to see which area a fine-dining Italian restaurant would thrive. To find this information we will utilize the following data:

- New York City Neighborhoods
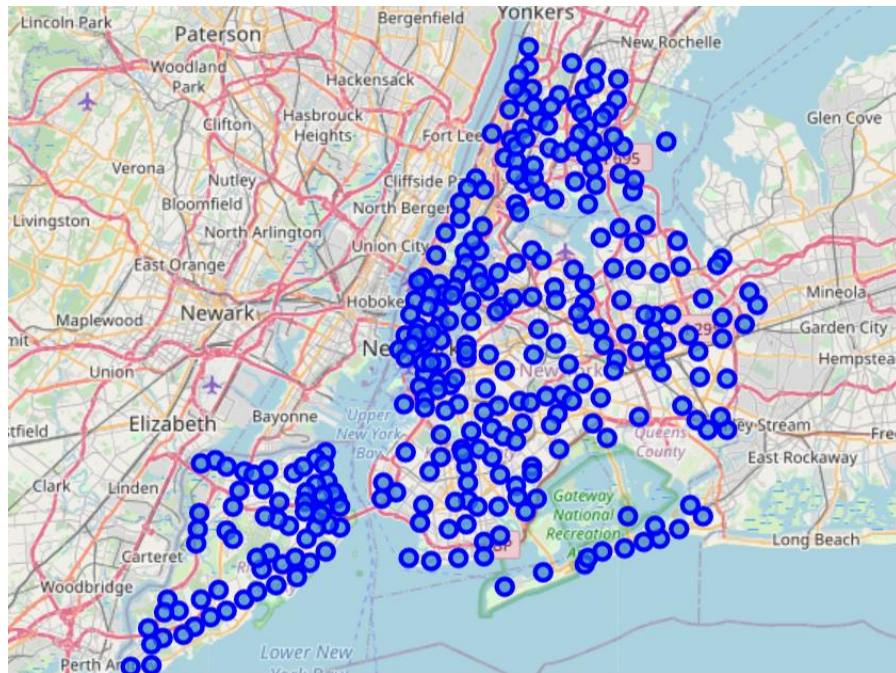- Neighborhood geolocations
- Restaurants in each neighborhood

We will cluster the data by neighborhoods and pull the information regarding other restaurants in the area to analyze for the best fit for our restauranteur.
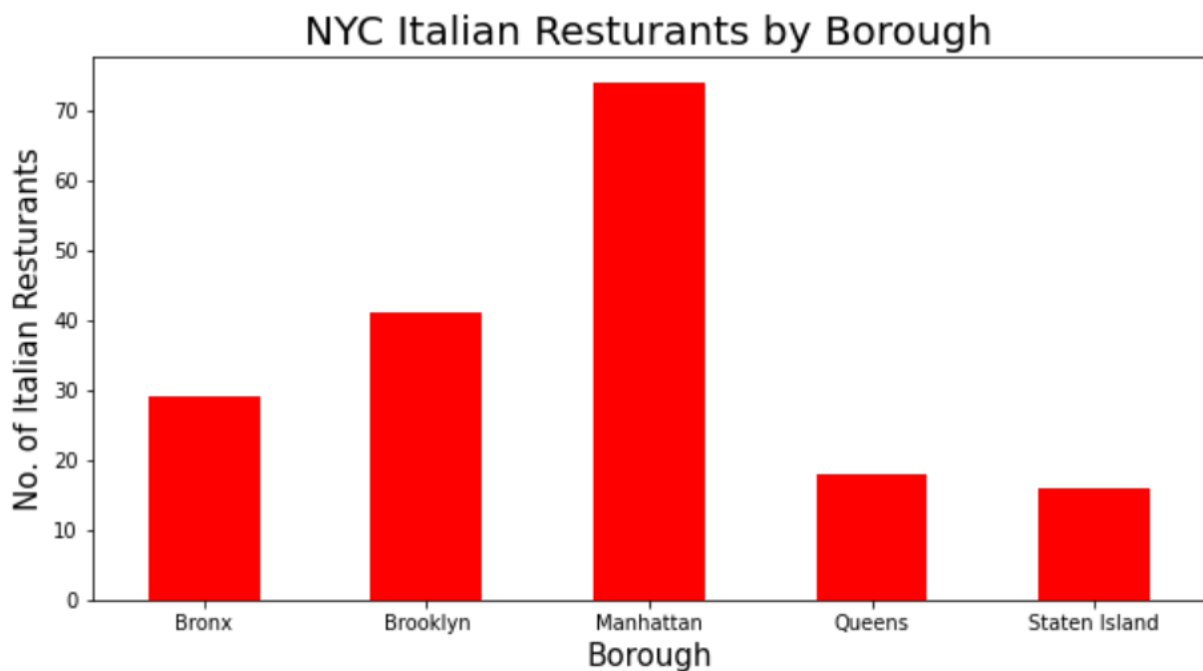
Methodology and Discussion

To start, I pulled all of the neighborhoods per borough for New York City to visualize the various neighborhoods to select from for our Italian restaurant. To visualize this, I utilized plt to organize the number of neighborhoods by borough (see chart below).
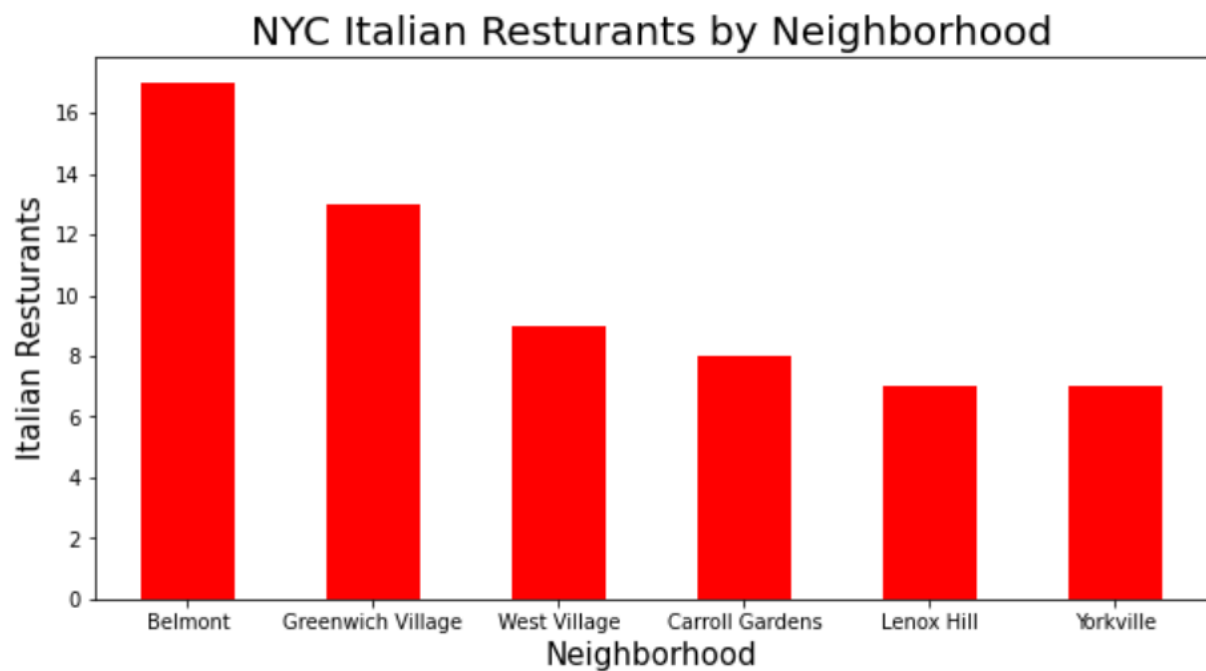
In addition to this visualization, I also used folium to showcase the neighborhoods on a map of NYC (see map below).
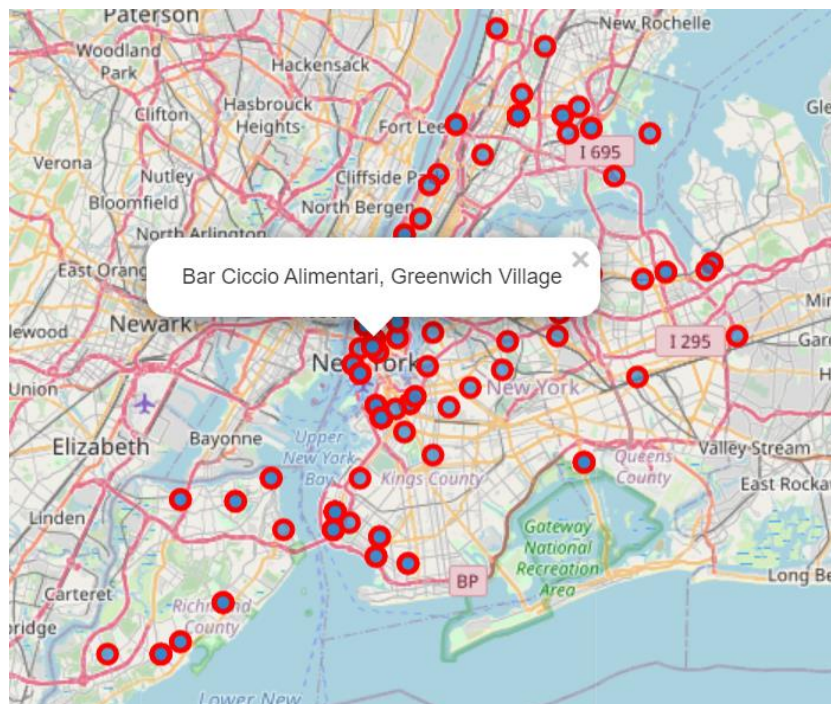

Now that I had the neighborhoods outlined for a potential restaurant location, it was time to utilize the Foursquare API to pull the information on the various Italian restaurants within the city that show both areas where there is interest in Italian cuisine as well as competition. For starters, I showcased the number of Italian restaurants by borough (see chart below).

Once that information was captured, I narrowed it down by the neighborhoods with the highest quantity of Italian restaurants. The result was the Belmont neighborhood in the Bronx.

I wanted to be able to visualize where the restaurants were located within the city and utilized another folium map to do so (see map below).

Results

As a result of the data analysis, our restauranteur is going to open his Italian restaurant in the neighborhood of Greenwich Village. It is a more affluent area of New York City. It also has an appetite (pun intended!) for Italian cuisine due to the number of other Italian restaurants in the neighborhood. Yet it isn't overly saturated as the Belmont area is.

Conclusion

In conclusion, as a result of the data analysis, our restauranteur is going to make a pitch presentation to investors to open an Italian restaurant in Greenwich Village, NYC.