Data Science Capstone

AN ANALYSIS OF THE LOCATIONS OF ITALIAN RESTURAUNTS FOR VARIOUS NEIGHBORHOODS IN THE BOUROUGH OF BROOKLYN, NEW YORK CITY

ZACHARY SHEEHAN

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Data Science Capstone

Introduction

New York City, being the most populous city in the United States, naturally has significant competition for any business looking to get its start. This is only compounded when considering the restaurant industry, specifically Italian food and pizza, of which the city is world renowned.

In this scenario, famous Italian brothers Mario and Luigi are looking to relocate their beloved restaurant, Mario & Luigi's Italian Ristorante & Pizzeria, to the borough of Brooklyn in New York City but are weary of the city's reputation for having so many established Italian Restaurants and Pizzerias competing with their business. They contacted IBM to help them utilize the power of data science to solve their business problem.

Data

In order to gain an understanding of the surrounding neighborhoods, we need data on their locations, specifically latitude and longitude coordinates for more precise geolocating. NYU has a public JSON dataset which contains all of this information for the 5 boroughs and 306 neighborhoods of New York City (https://geo.nyu.edu/catalog/nyu_2451_34572).

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

For restaurant data, we shall utilize the Foursquare API that allows for searching of businesses within specific latitude and longitude constraints. We will tie this to the previous step with the coordinates obtained through the neighborhood data in order to generate a dataset of businesses by neighborhood.

Methodology

Using the data from the NYU neighborhoods dataset, we defined a method to call 100 business listings of proximity to the neighborhood coordinates provided. From here, we add businesses by category including location coordinates to our neighborhoods dataset and analyze for unique business categories. An additional dataset is then formed in order to analyze the frequency of the captured categories within each neighborhood. Finally, we use this information to form a dataset of the top 10 most common categories of businesses for each neighborhood of Brooklyn.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Georgian Dream Cafe and Bakery	40.625586	-74.030196	Caucasian Restaurant
4	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot

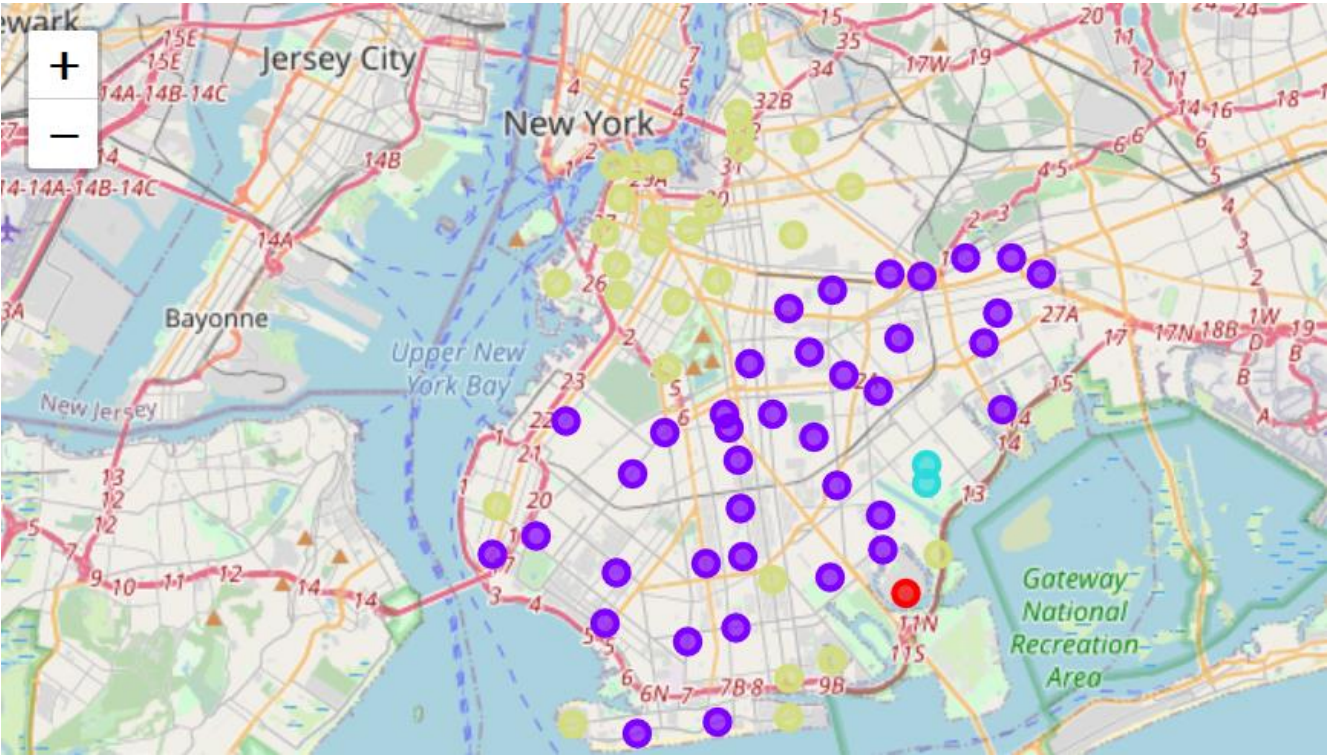|	Neighborhood	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store
0	Bath Beach	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.00
1	Bay Ridge	0.012346	0.0	0.000000	0.037037	0.000000	0.0	0.0	0.000000	0.00
2	Bedford Stuyvesant	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.00
3	Bensonhurst	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.00
4	Bergen Beach	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.00

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bath Beach	Pharmacy	Fast Food Restaurant	Gas Station	Bubble Tea Shop	Cantonese Restaurant	Italian Restaurant	Chinese Restaurant	Pizza Place	Donut Shop	Dessert Shop
1	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Bar	American Restaurant	Greek Restaurant	Playground	Hookah Bar	Ice Cream Shop	Pharmacy
2	Bedford Stuyvesant	Coffee Shop	Pizza Place	Café	Bar	Deli / Bodega	BBQ Joint	Park	Cocktail Bar	New American Restaurant	Bus Stop
3	Bensonhurst	Chinese Restaurant	Italian Restaurant	Park	Sushi Restaurant	Pizza Place	Ice Cream Shop	Donut Shop	Pet Store	Record Shop	Cosmetics Shop
4	Bergen Beach	Harbor / Marina	Baseball Field	Playground	Athletics & Sports	Hockey Field	Park	Food & Drink Shop	Food	Food Court	Food Stand

Now that we have the information for the most common categories of businesses for each neighborhood, we can begin the process of K-means clustering in order to create a more generalized grouping of specific neighborhoods that may hold a different makeup of categories of businesses. We decided upon 4 clusters as that would give more of a direct answer on business composition differences.

After clustering our neighborhoods, we use mapping plug-in Folium to visualize the location and size of our clusters to determine relevancy to our business problem. Finally, we will examine the exact neighborhoods that make up each cluster to understand the makeup of the cluster and if the cluster would be a good place for our client to establish their business.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	3	Italian Restaurant	Spa	Pizza Place	Bar	American Restaurant	Greek Restaurant	Playground	Hookah Bar	Ice Cream Shop	Pharmacy
1	Brooklyn	Bensonhurst	40.611009	-73.995180	1	Chinese Restaurant	Italian Restaurant	Park	Sushi Restaurant	Pizza Place	Ice Cream Shop	Donut Shop	Pet Store	Record Shop	Cosmetics Shop
2	Brooklyn	Sunset Park	40.645103	-74.010316	1	Pizza Place	Latin American Restaurant	Bank	Bakery	Mexican Restaurant	Mobile Phone Shop	Fried Chicken Joint	Deli / Bodega	Ice Cream Shop	Gym
3	Brooklyn	Greenpoint	40.730201	-73.954241	3	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Record Shop	Yoga Studio	Mexican Restaurant	French Restaurant	Nail Salon	Deli / Bodega
4	Brooklyn	Gravesend	40.595260	-73.973471	1	Lounge	Pizza Place	Bakery	Chinese Restaurant	Italian Restaurant	Gym	Diner	Cosmetics Shop	Furniture / Home Store	Men's Store

Results

The resulting map of our clustering process is shown above. As we can see from the purple and yellow clusters, there is a significant overlap between the composition of categories of businesses in Brooklyn, which is expected as the borough is well-known for its Italian food and pizza. Shown below is the makeup of Cluster 1 (marked in purple), which as you can see holds many different types of restaurants, most of which having Italian or Pizza in the top 5, in some cases both.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bensonhurst	Chinese Restaurant	Italian Restaurant	Park	Sushi Restaurant	Pizza Place	Ice Cream Shop	Donut Shop	Pet Store	Record Shop	Cosmetics Shop
2	Sunset Park	Pizza Place	Latin American Restaurant	Bank	Bakery	Mexican Restaurant	Mobile Phone Shop	Fried Chicken Joint	Deli / Bodega	Ice Cream Shop	Gym
4	Gravesend	Lounge	Pizza Place	Bakery	Chinese Restaurant	Italian Restaurant	Gym	Diner	Cosmetics Shop	Furniture / Home Store	Men's Store
5	Brighton Beach	Eastern European Restaurant	Restaurant	Russian Restaurant	Mobile Phone Shop	Beach	Sushi Restaurant	Pharmacy	Gourmet Shop	Diner	Taco Place
7	Manhattan Terrace	Ice Cream Shop	Pizza Place	Donut Shop	Chinese Restaurant	Bagel Shop	Burger Joint	Mobile Phone Shop	Organic Grocery	Steakhouse	Grocery Store
8	Flatbush	Coffee Shop	Deli / Bodega	Mexican Restaurant	Caribbean Restaurant	Bank	Pharmacy	Juice Bar	Bagel Shop	Hardware Store	Lounge
9	Crown Heights	Pizza Place	Museum	Café	Bakery	Bagel Shop	Burger Joint	Candy Store	Supermarket	Sushi Restaurant	Salon / Barbershop
10	East Flatbush	Department Store	Liquor Store	Pharmacy	Park	Chinese Restaurant	Caribbean Restaurant	Fast Food Restaurant	Food & Drink Shop	Print Shop	Supermarket

Cluster 3 (marked in yellow) is largely the same as Cluster 1, the main difference being the presence of bars in addition to restaurants. From this information, we can rule out both of these clusters as viable locations for opening an Italian restaurant. Interestingly, Cluster 0 (marked in red) has only Fast Food in their top 5, though upon further review the Mill Island neighborhood appears to hold significant industrial land areas so we will rule out this cluster as well. Finally, Cluster 3 (marked in teal, shown below) seems to be the perfect solution to our business problem as it holds commercial districts as well as other restaurants, however Italian and Pizza restaurants are not in the top 10 most common for either neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
28	Canarsie	Thai Restaurant	Asian Restaurant	Gym	Grocery Store	Caribbean Restaurant	Food	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop
59	Paerdegat Basin	Asian Restaurant	Child Care Service	Auto Garage	Food	Women's Store	Field	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market

Discussion

Some limitations that were faced include that there is little demographic information considered in our initial NYU dataset and upon further research the Mill Island neighborhood may not be the best location for a restaurant as it is not as residential as other neighborhoods and the population is relatively low.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
30	Mill Island	Pool	Women's Store	Flower Shop	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market	Food

Conclusion

In this report, we have utilized geolocation and business category data in order to analyze the composition of businesses in various neighborhoods of Brooklyn in order to help a fledgling restaurant find the location with the least amount of localized competition. To conclude, we have decided to recommend the Mario brothers to open their restaurant within one of the Cluster 3 neighborhoods.