Connor Penny

Opening an Italian Restaurant in Toronto

Introduction/Business Problem

There are numerous factors that can contribute to a restaurant owner's location choice. Often, potential competition is important to consider when deciding where to locate. If an area is already saturated with restaurants that serve your cuisine, you would likely lose business to these restaurants if you opened a restaurant there. Ideally you would like to be the only one in an area serving your specific cuisine type.

However, once you find areas without restaurants similar to yours, a new question arises. *Why* hasn't this kind of restaurant opened in this area? It is possible that the barriers to entry have just been too high, or the area recently developed into a viable location for restaurants. In this case, you would want to take advantage of this opportunity and monopolize the market for your cuisine in this area. On the other hand, it's possible that a restaurant similar to yours hasn't opened up in this area because this area cannot support it. This area may simply not have a high enough demand for this type of restaurant.

These are important considerations that any potential restaurant owner should think about before investing huge amounts of money into their new business. Thus, if a data driven answer can be found, it could mitigate a large amount of risk for these owners and increase potential profits.

Data

Similar to the lab from earlier in this course, we will be using FourSquare data on venues in different Toronto neighborhoods. It was easiest to do this for Toronto because we had already scraped the postal code of each neighborhood from a Wikipedia page, allowing us to find coordinates for each one using geocoder. While we will use data on venues of all types in our analysis, we will pay special attention to data on Italian restaurants, as this is our restaurant of interest.

Methodology

In order to find Toronto neighborhoods with no Italian restaurants and then determine if these neighborhoods would be viable locations to open an Italian restaurant, I sought to find neighborhoods that had no Italian restaurants but had similar characteristics to those that did have Italian restaurants.

To do this, I started by scraping all the necessary data from the FourSquare API and putting it into a dataframe. After transforming the data, the dataframes we will use for data analysis is the set of all Toronto neighborhoods with frequency of each venue type in that neighborhood.

Connor Penny

Neighborhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports
Berczy Park	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.000000	0.000000	0.00	0.020000	0.00	0.000000	0.000000	0.0000
Brockton, Parkdale Village, Exhibition Place	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
Business reply mail Processing Centre, South C...	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
CN Tower, King and Spadina, Railway Lands, Har...	0.0625	0.0625	0.0625	0.125	0.125	0.0625	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
Central Bay Street	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.000000	0.000000	0.00	0.000000	0.02	0.000000	0.000000	0.0000
Christie	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0625

I also create a data frame that is each neighborhood with just the frequency of Italian Restaurants in that neighborhood to show more readily neighborhoods that did not have Italian restaurants.

	Neighborhood	Italian Restaurant
0	Berczy Park	0.020000
1	Brockton, Parkdale Village, Exhibition Place	0.043478
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.040000
5	Christie	0.062500
6	Church and Wellesley	0.000000
7	Commerce Court, Victoria Hotel	0.000000
8	Davisville	0.060606
9	Davisville North	0.000000
10	Dufferin, Dovercourt Village	0.000000

Now that I knew which neighborhoods did and did not have Italian restaurants, I needed to do determine which of those without could possibly support one.

The way I decided to do this was to cluster the neighborhoods based on the frequency of all their venues. I use a K means clustering. I explored many different numbers of clusters not shown in the notebook, but I found that 7 clusters led to the most differentiation of neighborhoods into different clusters. Other numbers of clusters often left clusters unpopulated by neighborhoods. In this clustering,

we are grouping neighborhoods based on similarities of the kinds of business that exist in that neighborhood. This gives us insight into the different neighborhoods that might have similar consumer demand as well as similar activities such as nightlife.

	Postal Code	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
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Theater
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Yoga Studio	Sushi Restaurant	Sandwich Place	Restaurant	Italian Restaurant	Diner
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	6	Coffee Shop	Clothing Store	Café	Fast Food Restaurant	Middle Eastern Restaurant	Bookstore	Theater
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	6	Café	Coffee Shop	Gastropub	Cocktail Bar	Japanese Restaurant	Restaurant	Farmers Market
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	6	Trail	Pub	Health Food Store	Yoga Studio	Cupcake Shop	Donut Shop	Doner Restaurant

Using this information, I could then examine each of the clusters to see the frequency of Italian restaurants within each cluster. I eliminated any clusters with only neighborhoods with no Italian restaurants from further investigation. If none of the neighborhoods in a cluster have Italian restaurants in them, I interpreted this as meaning the cluster did not have the traits necessary to support an Italian restaurant.

I then examined the clusters that did have neighborhoods with Italian restaurants. I looked in these clusters to find the neighborhoods that did not have Italian restaurants and identified these as neighborhoods of interest for potentially opening a new Italian restaurant. A neighborhood like this (in a cluster with neighborhoods that have Italian restaurants but has no Italian restaurants itself) has the characteristics of neighborhoods where Italian restaurants are successful. However, it does not have an Italian restaurant in its area. Clusters with proportionally more neighborhoods with Italian restaurants than without are assumed to be the more "successful" clusters for identifying neighborhoods of interest.

Once I identified the clusters that did have neighborhoods with Italian restaurants, I decided to take it a step further and perform K means clustering within these clusters. This served the same purpose as the original clustering, further identifying clusters of interest and allowing me to eliminate any clusters with only neighborhoods that did not have Italian restaurants.

Results

Of the 38 total Toronto neighborhoods, I found that 19 did not have an Italian restaurant. Any neighborhood that did have an Italian restaurant was eliminated from consideration as a location for a new Italian restaurant. However, these neighborhoods would still be used in our data analysis, as described in our methodology.

After performing a K means clustering, I found that 5 of the 7 clusters had only one neighborhood in them, and none of these neighborhoods had an Italian restaurant in them. This narrowed my search down to two clusters of interest.

Connor Penny

	Cluster Labels	Italian Restaurant
0	0	0.023197
1	1	0.000000
2	2	0.000000
3	3	0.000000
4	4	0.000000
5	5	0.000000
6	6	0.022496

This narrowed my search down to cluster 0 and cluster 6, which contained a combined 15 neighborhoods that were in these two clusters but did not contain Italian restaurants.

	Neighborhood	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	Italian Restaurant
0	Regent Park, Harbourfront	0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Theater	Yoga Studio	Shoe Store	Brewery	0.000000
1	Queen's Park, Ontario Provincial Government	0	Coffee Shop	Yoga Studio	Sushi Restaurant	Sandwich Place	Restaurant	Italian Restaurant	Diner	Café	Smoothie Shop	Beer Bar	0.029412
6	Central Bay Street	0	Coffee Shop	Bubble Tea Shop	Café	Italian Restaurant	Yoga Studio	Comic Shop	Sandwich Place	Burger Joint	Ramen Restaurant	Portuguese Restaurant	0.040000
12	The Danforth West, Riverdale	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Restaurant	Ice Cream Shop	Bookstore	Furniture / Home Store	Grocery Store	Pizza Place	Brewery	0.069767
23	North Toronto West, Lawrence Park	0	Coffee Shop	Clothing Store	Bagel Shop	Furniture / Home Store	Ice Cream Shop	Fast Food Restaurant	Diner	Mexican Restaurant	Park	Chinese Restaurant	0.000000
24	The Annex, North Midtown, Yorkville	0	Café	Sandwich Place	Coffee Shop	History Museum	Park	Pharmacy	Pizza Place	Pub	Middle Eastern Restaurant	Burger Joint	0.000000

| Neighborhood	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	Italian Restaurant
Garden District, Ryerson	6	Coffee Shop	Clothing Store	Café	Middle Eastern Restaurant	Bookstore	Ramen Restaurant	Theater	Fast Food Restaurant	Pizza Place	Diner	0.020000
St. James Town	6	Café	Gastropub	Coffee Shop	Cocktail Bar	Cosmetics Shop	Restaurant	Farmers Market	Japanese Restaurant	Beer Bar	Bookstore	0.020000
The Beaches	6	Trail	Pub	Health Food Store	Yoga Studio	Cupcake Shop	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Discount Store	0.000000
Berczy Park	6	Coffee Shop	Farmers Market	Cocktail Bar	Restaurant	Cheese Shop	Beer Bar	Bakery	Seafood Restaurant	Hotel	Park	0.020000
Christie	6	Grocery Store	Café	Park	Candy Store	Athletics & Sports	Restaurant	Italian Restaurant	Coffee Shop	Nightclub	Baby Store	0.062500
Richmond, Adelaide, King	6	Coffee Shop	Café	Concert Hall	American Restaurant	Steakhouse	Restaurant	Noodle House	Department Store	Smoke Shop	Seafood Restaurant	0.000000
Dufferin, Dovercourt Village	6	Pharmacy	Bakery	Bank	Café	Park	Brewery	Music Venue	Supermarket	Bar	Grocery Store	0.000000

Connor Penny

After performing a k means clustering within these clusters, I was only able to narrow our search slightly more. I eliminated two more neighborhoods from this clustering, and I was left with 13 neighborhoods with no Italian restaurants but similar characteristics to neighborhoods that do have Italian restaurants. The following table summarizes those neighborhoods and their venue characteristics:

Cluster Labels	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	Italian Restaurant
1	Regent Park, Harbourfront	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Café	Theater	Yoga Studio	Shoe Store	Brewery	0.0
1	The Annex, North Midtown, Yorkville	Café	Sandwich Place	Coffee Shop	History Museum	Park	Pharmacy	Pizza Place	Pub	Middle Eastern Restaurant	Burger Joint	0.0
1	Davisville North	Park	Hotel	Dog Run	Breakfast Spot	Gym / Fitness Center	Sandwich Place	Department Store	Dance Studio	Food & Drink Shop	Doner Restaurant	0.0
1	Commerce Court, Victoria Hotel	Coffee Shop	Café	Hotel	Restaurant	Gym	Seafood Restaurant	Deli / Bodega	Beer Bar	Japanese Restaurant	Gastropub	0.0
1	Church and Wellesley	Coffee Shop	Gay Bar	Restaurant	Yoga Studio	Men's Store	Japanese Restaurant	Burger Joint	Hobby Shop	Hotel	Distribution Center	0.0
1	Summerhill West, Rathnelly, South Hill, Forest...	Coffee Shop	Supermarket	Fried Chicken Joint	Sports Bar	Sushi Restaurant	Bank	Restaurant	Pizza Place	Bagel Shop	Liquor Store	0.0
1	Business reply mail Processing Centre, South C...	Brewery	Pizza Place	Recording Studio	Restaurant	Light Rail Station	Auto Workshop	Butcher	Burrito Place	Farmers Market	Gym / Fitness Center	0.0
4	Dufferin, Dovercourt Village	Pharmacy	Bakery	Bank	Café	Park	Brewery	Music Venue	Supermarket	Bar	Grocery Store	0.0
4	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Airport	Coffee Shop	Rental Car Location	Sculpture Garden	Harbor / Marina	Plane	Boat or Ferry	Boutique	0.0
4	Kensington Market, Chinatown, Grange Park	Mexican Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Bakery	Burger Joint	Dumpling Restaurant	Dessert Shop	Vietnamese Restaurant	Bar	0.0
4	The Beaches	Trail	Pub	Health Food Store	Yoga Studio	Cupcake Shop	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Discount Store	0.0
4	Richmond, Adelaide, King	Coffee Shop	Café	Concert Hall	American Restaurant	Steakhouse	Restaurant	Noodle House	Department Store	Smoke Shop	Seafood Restaurant	0.0
4	First Canadian Place, Underground city	Café	Coffee Shop	Restaurant	Seafood Restaurant	Japanese Restaurant	Hotel	Deli / Bodega	Bar	Gym	American Restaurant	0.0

Discussion

If we investigate the characteristics of these 13 neighborhoods, we see that the most common venues among them are overwhelmingly restaurants, bars, and coffee shops. For many of these neighborhoods, the characteristics are suggestive of a typical town center, an area with a large amount of people looking for entertainment and supporting business that may not be considered necessarily "essential". Thus, it makes sense that these neighborhoods would be able to support the opening of a new Italian restaurant; it is fairly obvious that they are capable of supporting a wide range of other restaurant types.

Connor Penny
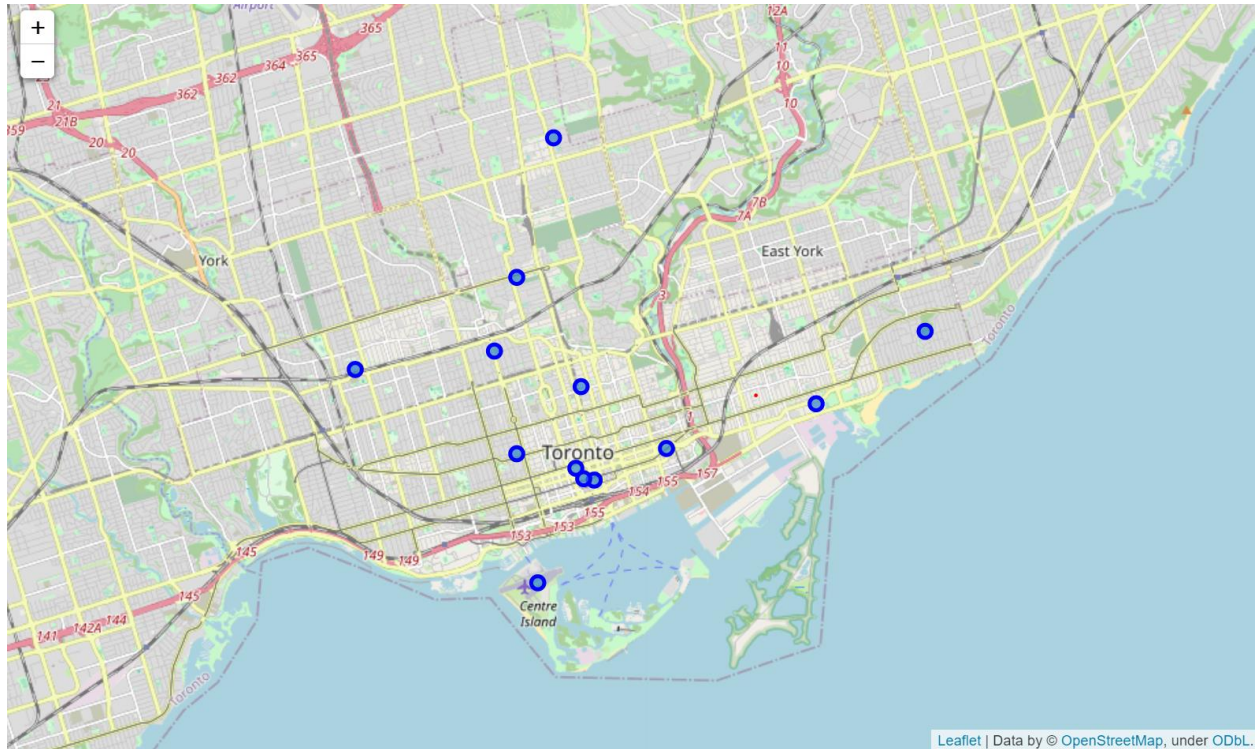
When we look at the distribution of these neighborhoods geospatially, we see that they do tend to cluster around the center of Toronto. However, we see several viable neighborhoods on the outskirts of the city. And although we did not factor rent costs into our data analysis, I would assume that these locations around the outskirts of the city would be cheaper than locations right on the water and/or at the city's center.

Conclusion

The purpose of this data analysis was to find neighborhoods in Toronto that would be viable locations for a new Italian restaurant. In order to do this, I searched for neighborhoods that did not have an Italian restaurant but were similar in characteristics to those that did have one. To do this, we used k means clustering and inferential reasoning and were able to eliminate 25 neighborhoods, leaving 13 as viable candidates for a new Italian restaurant.

Obviously, this does not take into account all factors that are important in determining where to locate a new restaurant. Another factor to consider, as mentioned earlier, is the rent cost in each neighborhood. Any potential restaurant owner needs to balance rent costs with consumer demand to maintain a profitable business. This would be an interesting consideration for further research.