Best Location for New Italian Restaurant in Toronto

Hyeon Gu Kim

June 30, 2020

1. Introduction

1.1 Background

An international renowned, multi-Michelin starred chef Gordon Ramsey has opened a string of award-winning restaurants across the globe and now he is ready to open his new Italian restaurant in Toronto. Looking for a perfect location for his new restaurant in Toronto, Gordon Ramsey asked me to use my data science knowledge to provide valuable insights and recommendation for his new restaurant.
He specifically wants to know:

- Floating population in Toronto
- Restaurants that could be competitors to his new restaurant
- Whether there are markets or farms near the location for affordable, quick and fresh ingredients
- Characteristics of each neighborhood

1.2 Business Problem

The challenge is finding a best location for Ramsey's new Italian restaurant in Toronto. Such location should have high demand for Ramsey's luxurious dishes. In addition, it should be close enough to markets/farms for fresh and quick ingredients distribution to his restaurant.

1.3 Interest

Beside my client Gordon Ramsey, the target audience would be the stakeholders and employees of Ramsey's food company. Anyone who are interested in food business also would be the target audience.

2.1 Data Description

I will be using the data below in order to solve the business problem.

- Neighborhoods in Toronto including its latitude and longitude
- Floating population data in each borough
- Location of markets/farms
- Average income of residents in each neighborhood

I will collect data as follows:

- Foursquare and Geopy will be used to retrieve geodata (latitude and longitude)

- Address and geodata for markets/farms will be achieved by Geopy and Nominatim
- Beautiful Soup will be utilized for retrieving data from URL (for floating population and avg income data)
- Use Folium for drawing map for Toronto and neighborhood

Successful collection of the data will allow us to answer the following questions:

- Which location has high floating population?
- What would be the possible candidates in terms of neighborhood?
- What is the characteristics of residents in each neighborhood?
- What is the distance between our possible candidates and markets/farms for each neighborhood?
- Who is the possible competitor?

3. Methodology

3.1 Neighborhoods with high floating population

From https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods, I could achieve overall demographic data of each neighborhoods with help of Beautiful Soup library. I filtered the data set and convert it to Pandas data frame so that the data frame has columns such as population, population density, population change, and average income.

I utilized Geopy Nominatim in order to get latitude and longitude of each neighborhoods, making another Pandas data frame. I ruled out the neighborhoods that is not available in Geopy search. Here is the result of merging these two data frames into one.

	Neighborhood	Latitude	Longitude	Population	Density (people/km2)	PopulationChange %	AvgIncome
0	Agincourt	43.785353	-79.278549	44,577	3580	4.6	25,750
1	Alderwood	43.601717	-79.545232	11,656	2360	-4.0	35,239
2	Alexandra Park	43.650787	-79.404318	4,355	13,609	0.0	19,687
3	Allenby	43.711351	-79.553424	2,513	4333	-1.0	245,592
4	Amesbury	43.706162	-79.483492	17,318	4,934	1.1	27,546
...	...	...	...	...	...	...	...
159	Woburn	43.759824	-79.225291	48,507	3636	-1.5	26,190
160	Wychwood	43.682122	-79.423839	4,182	6150	-2.0	53,613
161	York Mills	43.744039	-79.406657	17,564	2409	2.0	92,099
162	York University Heights	43.758781	-79.519434	26,140	1979	-1.2	24,432
163	Yorkville	43.671386	-79.390168	6,045	10,795	62.6	105,239

Then I sorted the data frame to find the neighborhoods with high floating population.

	Neighborhood	Population	Density (people/km2)
0	St. James Town	14666	63765
1	Bay Street Corridor	4787	43518
2	North York City Centre	10427	37239
3	Church and Wellesley	13397	24358
4	Crescent Town	8157	20393
5	Garden District	8240	15846
6	Parkdale	28367	13974
7	Alexandra Park	4355	13609
8	Niagara	6524	11862
9	Wallace Emerson	10338	11748

Because high floating population indicates continuous exposure of Ramsey's restaurant, we have to pay more attention to the density of each neighborhoods rather than its population. From the Pandas data frame above, we can observe top 10 neighborhoods with high population density.

3.2 Characteristics of each neighborhoods

Here are the characteristics of the neighborhoods with high population density that I found:

- **St. James Town**
  - St. James Town is the largest high-rise community in Canada. It has been identified as one of 13 economically deprived neighborhoods within the city. Due to its cultural and minority demographics, St. James Town is often thought as "the world within a block".
- **Bay Street Corridor**
  - "Bay Street" is frequently used as a metonym to refer to Toronto's Financial District and the Canadian financial sector as a whole, similar to Wall Street in the United States.
- **North York City Centre**
  - North York City Centre became the largest of four central business districts in the new city outside Downtown Toronto
- **Church and Wellesley**
  - It is an LGBT-oriented enclave with the core commercial strip located along Church Street from Wellesley south to Alexander.
- **Crescent Town**
  - It mainly consists of high-rise apartment complexes, built originally to take advantage of the opening of the adjacent Victoria Park subway station, which connects to the central quadrangle via a partially covered walkway.

We can easily notice that Church and Wellesley and Crescent Town are not ideal for Ramsey's restaurant. Since it is LGBT-oriented neighborhood, Church and Wellesley is not efficient location for success of Ramsey's business. In addition, High-rise apartment complexes in Crescent Town would most likely limit the customers for Ramsey's restaurant.

Therefore, we can claim that the possible candidates for Ramsey's new restaurant are St. James Town, Bay Street Corridor, and North York City Centre.

3.3 Distance between the neighborhoods and food distributors

After ruling out most of the neighborhoods that do not qualify for Ramsey's new Italian restaurant, we should pay attention of the distance between the neighborhoods and food distributors. Since all three candidates are in Ontario, Toronto, we will focus on food distributors in Ontario.

Again, utilizing Beautiful Soup library, we could read data of the food distributors in Ontario from Better Food Concepts (http://betterfoodconcepts.ca/distributors/). I also used Geopy Nominatim for the latitude and longitude.

	Food Distributors	Address	Latitude	Longitude
0	Macgregors Meat & Seafood	265 Garyray Dr	43.768908	-79.539571
1	Atlantic Grocery Distributors	Bay Roberts	43.735988	-79.258963
2	Centennial	Richmond Hill	43.812589	-79.263370
3	Chicago58	Woodbridge	43.759586	-79.598073
4	FKK Wholesale Cash & Carey	North York	43.754326	-79.449117
5	Flanagan Foodservice	Kitchener	43.731949	-79.253564
6	GFS	Milton	43.695302	-79.332080
7	The Butcher Shoppe	Etobicoke	43.643556	-79.565633
8	Burton Meats	Mississauga	43.666855	-79.587956
9	Nossack Fine Meats	Red Deer	43.691766	-79.274040
10	Independent Fish	Winnipeg	43.669121	-79.371860
11	Mercury Wholesale	Hamilton	43.670618	-79.374682
12	Morton Wholesale	Windsor	43.664636	-79.378061
13	Northern Meat Services	Winnipeg	43.669121	-79.371860
14	Stewarts	Barrie	43.681135	-79.432282
15	Sysco	Peterborough	43.674714	-79.441812
16	Vending Products of Canada	Etobicoke	43.643556	-79.565633
17	Worldwide Specialty Foods	Calgary	43.669491	-79.374354

In order to find the distance between the neighborhoods and the food distributors, I used Haversine Formula, passing lat1 and lng1 as the latitude and longitude of the neighborhoods, and lat2 and lng2 as the latitude and longitude of the food distributors.

```python
# lat and lng distance calculator (Haversine Formula)
import math
def dist_calculator(lat1, lng1, lat2, lng2):
    radius = 6371
    toRadian = math.pi / 180
    delta_Lat = abs(lat1 - lat2) * toRadian
    delta_Lng = abs(lng1 - lng2) * toRadian

    sin_Lat = math.sin(delta_Lat / 2)
    sin_Lng = math.sin(delta_Lng / 2)

    squareRoot = math.sqrt(
        sin_Lat * delta_Lat +
        math.cos(lat1 * toRadian) * math.cos(lat2 * toRadian) * sin_Lng * sin_Lng)

    distance = 2 * radius * math.asin(squareRoot)

    return distance
```

For St. James Town, the head of the distances between St. James Town and the food distributors is:

	Food Distributors	Distance to St.James (km)
0	Macgregors Meat & Seafood	20.61
1	Atlantic Grocery Distributors	13.90
2	Centennial	24.17
3	Chicago58	23.00
4	FKK Wholesale Cash & Carey	14.70

Distance to St.James (km)	
count	18.000000
mean	10.308889
std	8.397219
min	0.080000
25%	1.927500
50%	11.205000
75%	16.040000
max	24.170000

For Bay Street Corridor, the head of the distances between Bay Street Corridor and the food distributors is:

	Food Distributors	Distance to Bay Street (km)
0	Macgregors Meat & Seafood	20.06
1	Atlantic Grocery Distributors	15.00
2	Centennial	24.95
3	Chicago58	22.23
4	FKK Wholesale Cash & Carey	14.52

Distance to Bay Street (km)	
count	18.000000
mean	10.449444
std	7.993305
min	0.940000
25%	2.055000
50%	12.245000
75%	14.965000
max	24.950000

For North York, the head of the distances between North York and the food distributors is:

	Food Distributors	Distance to North York (km)
0	Macgregors Meat & Seafood	7.62
1	Atlantic Grocery Distributors	15.54
2	Centennial	17.50
3	Chicago58	11.99
4	FKK Wholesale Cash & Carey	0.00

Distance to North York (km)	
count	18.000000
mean	14.131667
std	4.614362
min	0.000000
25%	12.700000
50%	14.770000
75%	16.902500
max	19.780000

Comparison of these three data gives insights that St. James Town and Bay Street Corridor are located closer to food distributors in Ontario than North York Centre. Therefore, I ruled out North York Centre.

3.4 Venues near the neighborhoods

Now that we have only two neighborhoods in our hands, we have to see the overall distribution of venues near the neighborhoods. I used Foursquare API with radius of 500 in order to retrieve data of venues near the neighborhoods.

Here is the data frame of the top 10 most common venues near St. James Town:

	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
125	1	St. James Town	Coffee Shop	Pizza Place	Gym / Fitness Center	Grocery Store	Café	Hotel	Bar	Beer Store	Bistro	Breakfast Spot

We can clearly notice that the overall distribution of the most common venues is focused on café and bar. As I expected from the neighborhood's characteristics, this indicates that St. James Town is not ideal for Ramsey's restaurant because of its high density of residents.

Let's check Bay Street Corridor:

	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
8	1	Bay Street Corridor	Clothing Store	Coffee Shop	Sushi Restaurant	French Restaurant	Boutique	Italian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Café	Smoke Shop

The overall distribution of the most common venues near Bay Street Corridor suggests that Bay Street Corridor is the best location for Ramsey to open his new Italian restaurant in Toronto. Wide distribution of various restaurants and other venues indicates high floating population of various people, not just its residents.

3.5 Potential Competitors

Since the sixth most common venue in Bay Street Corridor is Italian restaurant, I focused on finding Ramsey's potential competitor restaurant. Utilizing Foursquare API, I explored Italian restaurants in Bay Street Corridor. Here is the data frame of the restaurants:

	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Trattoria Nervosa	43.671019	-79.391081	Italian Restaurant
2	Dal Moro's Fresh Pasta To Go	43.666641	-79.385400	Italian Restaurant
3	7 West Cafe	43.668665	-79.386830	Italian Restaurant
4	Blu Ristorante and Lounge	43.671685	-79.388614	Italian Restaurant
5	Dimmi	43.670062	-79.392680	Italian Restaurant
7	Vaticano Trattoria	43.670738	-79.391043	Italian Restaurant
8	Scaccia Restaurant	43.668794	-79.388296	Italian Restaurant
9	Sofia	43.670592	-79.391849	Italian Restaurant
10	Focaccia	43.669932	-79.386110	Italian Restaurant
11	Zaza paninoteca	43.670739	-79.391159	Italian Restaurant
12	Bacaro Italian Eatery	43.670824	-79.384675	Italian Restaurant

Considering Ramsay's restaurants are luxurious (probably between 3 to 5-dollar signs), the possible competitors against his restaurant would be Blu Ristorante and Lounge, and Sofia.

4. Result

Under the assumption that Gordon Ramsey is looking for a location for his new Italian restaurant in Toronto, Bay Street Corridor would be the best choice for his new Italian restaurant. Bay Street Corridor neighborhood has the highest floating population in Toronto except St. James Town where its high population is highly concentrated on residents only. In addition, Ramsey's restaurant would have the upper hand compared to other restaurants because Bay Street Corridor is located near variety of food distributors, giving Ramsey's restaurant the ability to get fresh ingredients. When we think about his restaurant's possible competitors, there are Blu Ristorante and Lounge and Sofia.

5. Discussion

Toronto is a big city with diverse culture and population density. My approach to the business problem was simple: I tried to focus on the floating population of each neighborhoods in Toronto. With K-cluster of 5, I utilized K-Means Clustering method in order to divide the data into separate clusters. I then used Folium to map the result out.
I then found the characteristics of each of the neighborhoods which have high floating population. I could exclude the neighborhoods that does not fit to Ramsey's Italian restaurant. The final 3 neighborhoods were St. James Town, Bay Street Corridor, and North York City Centre.
Since all three neighborhoods are located in Ontario, Toronto, I then focused on the distance between the neighborhoods and food distributors in Ontario. I utilized Beautiful Soup in order to read data of food distributors in Better Food Concepts (http://betterfoodconcepts.ca/distributors/). I also utilized geocoder in order to find the latitude and longitude for each of the food distributors. I then used Haversine Formula for distance between each of the neighborhoods and the food distributors. From its results, I could rule out North City Centre which has longest mean distance of 14.13 km.
Through Foursquare API, I explored venues near St. James Town and Bay Street Corridor in radius of 500. As I expected through the neighborhoods' characteristics, the most common venues near St. James Town were cafe, gyms, and beer shops as St. James Town is the resident-centered neighborhood. On the other hand, the most common venues near Bay Street Corridor were restaurants and clothing shops. From this data analysis, I could derive insights that Bay Street Corridor is the best neighborhood for Ramsey's new Italian restaurant.
Lastly, I found possible competitors against Ramsey's restaurant. Also using Foursquare API, I could gather data on Italian restaurants in Bay Street Corridor. Among 11 Italian restaurants in the neighborhood, I ruled out the restaurants that are as not luxurious as most of Ramsey's restaurant. The possible competitors against Ramsey's restaurant were Blu Ristorante and Lounge and Sofia.

6. Conclusion

In conclusion, Bay Street Corridor is the best neighborhood for Ramsey's new Italian restaurant when we take account of floating population, characteristics of the neighborhood, and distance between the neighborhood and food distributors.

Opening an Italian restaurant in Bay Street Corridor will not only expose the restaurant to many people but also maintain Ramsey's reputation for Michelin-star chef with quick distribution of fresh ingredients.