Restaurant Diversity and Demand Analysis of Major Canadian Cities

1. Introduction

Canada is considered one of the most culturally diverse countries on the planet and is home to a wide variety of different cuisines, which is especially clear in the nation's largest and most populous cities. For this data science project, I hope to analyze the restaurant scenes in Toronto, Montreal, and Vancouver to see how they compare to one another concerning cultural diversity, and the demands of Canadians generally. This topic is quite pertinent to those interested in opening restaurants in these competitive cities, particularly during the COVID-19 pandemic (which has drastically altered the food & beverage industry worldwide).

2. Data

To analyze these restaurant scenes, great amounts of accurate data are necessary to discern meaningful results. The data used in this project include:

- Restaurant venue lists near given coordinates of neighbourhoods within each city, which was procured using the **Foursquare API**. This data contains the coordinates of each restaurant, category/type of restaurant, and name of each venue within a specific radius of a point [1].
- The demand of each type of restaurant among Canadians, which was represented using keyword search volume from **SEMRush** keyword data. This keyword data is a rough heuristic of what types of restaurants Canadians most often search for [2].
- The neighbourhood information of each city, and the corresponding postal codes to each grouping provided by **Wikipedia** [3].
- Specific location data (longitude and latitude coordinates) of neighbourhoods based on postal code information provided by **Service Objects** [4].

3. Methodology

To store the data of the project and code, I used a **GitHub** repository to host the Python Jupyter Notebook, alongside several CSV files [5]. The first part of the project entailed web-scraping the location data of the neighbourhoods in each city, which was accomplished by using the BeautifulSoup Python Library. Using the data from Wikipedia and CSV files procured from Service Object (and Coursera), I was able to create data frames for each city, like the one below:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

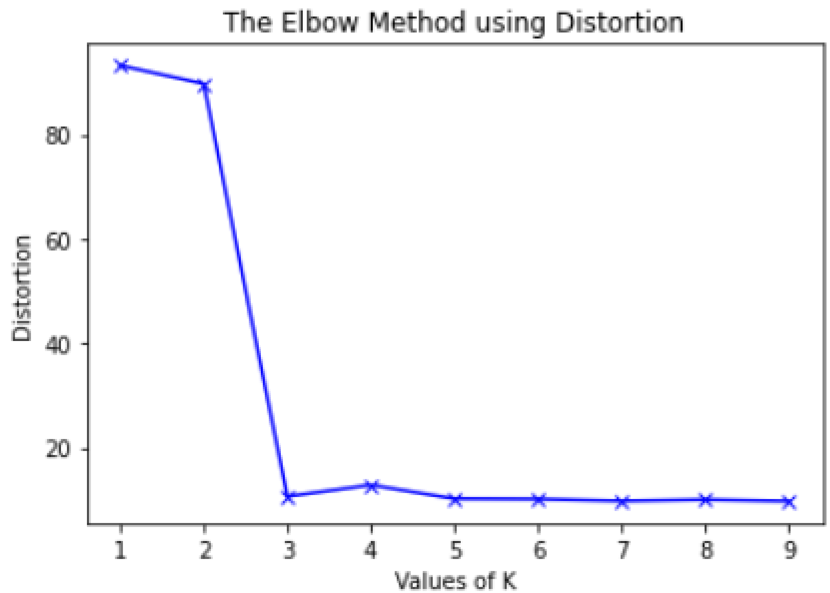
Next, using the Foursquare API to locate restaurants nearby each neighbourhood, I created data frames that contained the venue data:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Pointe-aux-Trembles	45.651381	-73.50238	Cora	45.650538	-73.511365	Breakfast Spot
1	Pointe-aux-Trembles	45.651381	-73.50238	Tomate Basilic	45.650852	-73.511610	Italian Restaurant
2	Pointe-aux-Trembles	45.651381	-73.50238	Rôtisserie Scores	45.657079	-73.509816	Restaurant
3	Pointe-aux-Trembles	45.651381	-73.50238	Thai Express	45.654907	-73.508971	Thai Restaurant
4	Pointe-aux-Trembles	45.651381	-73.50238	Rôtisserie St-Hubert	45.650479	-73.512087	Restaurant

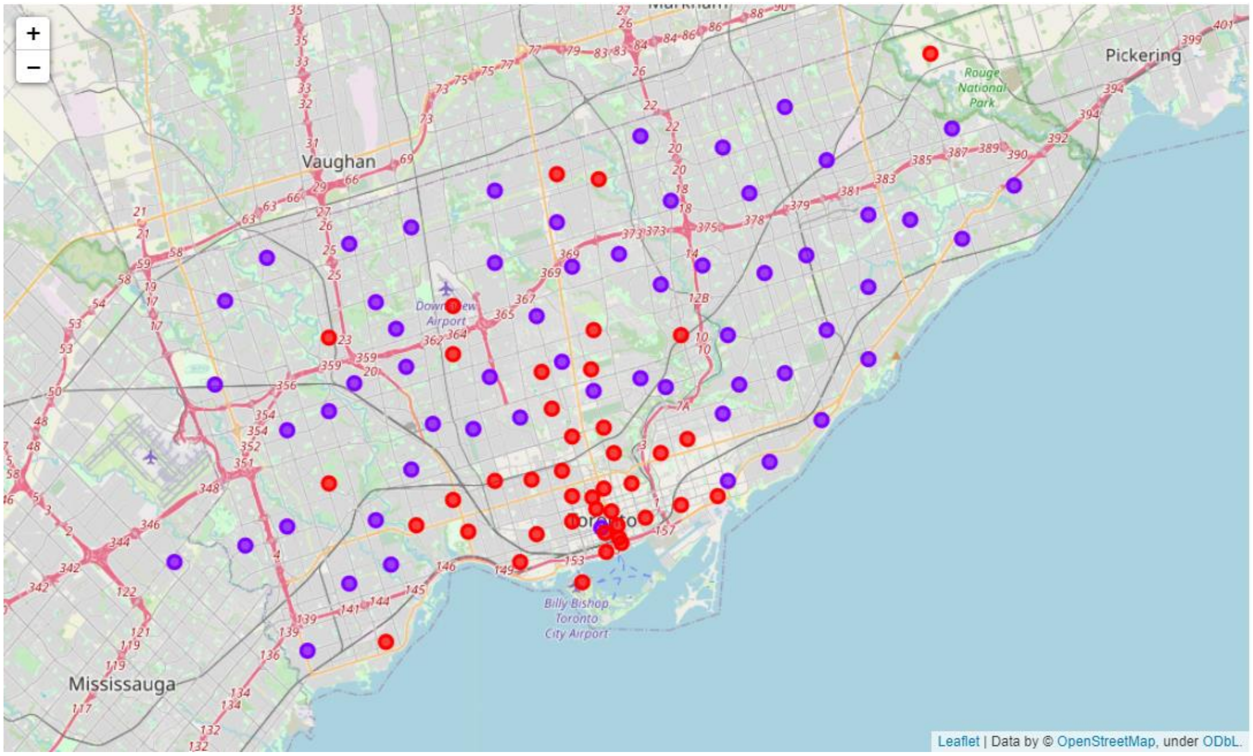
After that, I used one-hot encoding to reformat the restaurant data so it could be analyzed quantitatively even though the data is qualitative. The data were reformatted in terms of the venue categories so that each neighbourhood could be analyzed based on the types of restaurants nearby. With some formatting, and reshaping of the data, the top restaurant categories for each neighbourhood was determined as below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ahuntsic Central	Restaurant	Café	Fast Food Restaurant	Italian Restaurant	Asian Restaurant
1	Ahuntsic East	Café	Sushi Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Breakfast Spot
2	Ahuntsic North	Italian Restaurant	Bakery	Sandwich Place	Fast Food Restaurant	Sushi Restaurant
3	Ahuntsic Southeast	Café	Sandwich Place	Breakfast Spot	Middle Eastern Restaurant	Fast Food Restaurant
4	Ahuntsic Southwest	Restaurant	Pizza Place	Bakery	Middle Eastern Restaurant	Café

From this data, I then began to analyze the neighbourhoods by using the K-Means unsupervised learning algorithm. Before doing so, I performed the Elbow-method analysis in each data set to determine the number of clusters for each city. This helps to show the number of groupings that are useful to describe a dataset:



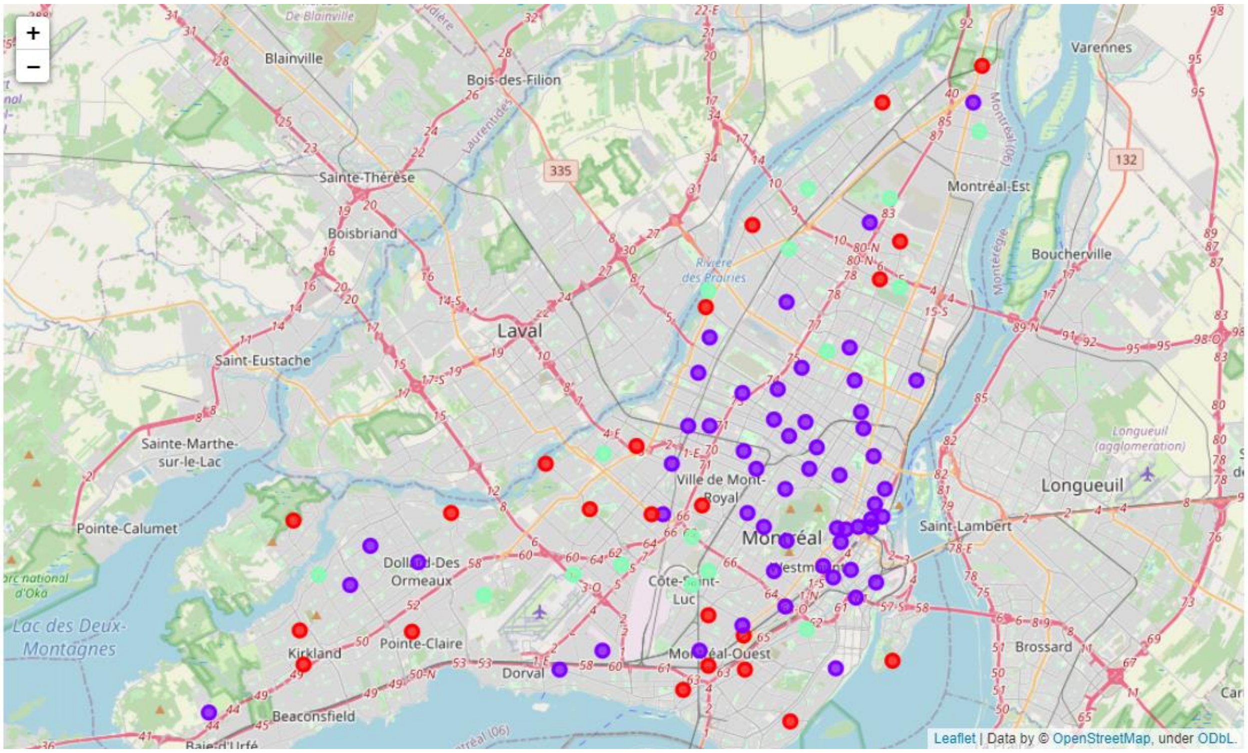
This graph shows that for the analyzed dataset (Montreal Neighborhoods), 3 is the optimal cluster size. Then, using the Python Folium Library, I was able to plot the neighbourhood clusters of each city with each point coloured corresponding to its grouping.



As you can see, Toronto is clustered into two groups which represent two different types of neighbourhoods.

- Red: Cafés, Italian Restaurants, and traditional Dine-In restaurants (Japanese, Mexican, Greek).
- Purple: Pizza Places, Fast Food, Sandwiches, North American Food (Burgers, Diners), and Takeout favourites (Chinese, Indian, General Asian).

A majority of the red clustered neighbourhoods lie at the heart of Toronto, indicating that lunch food and quick eats dominate the city centre.



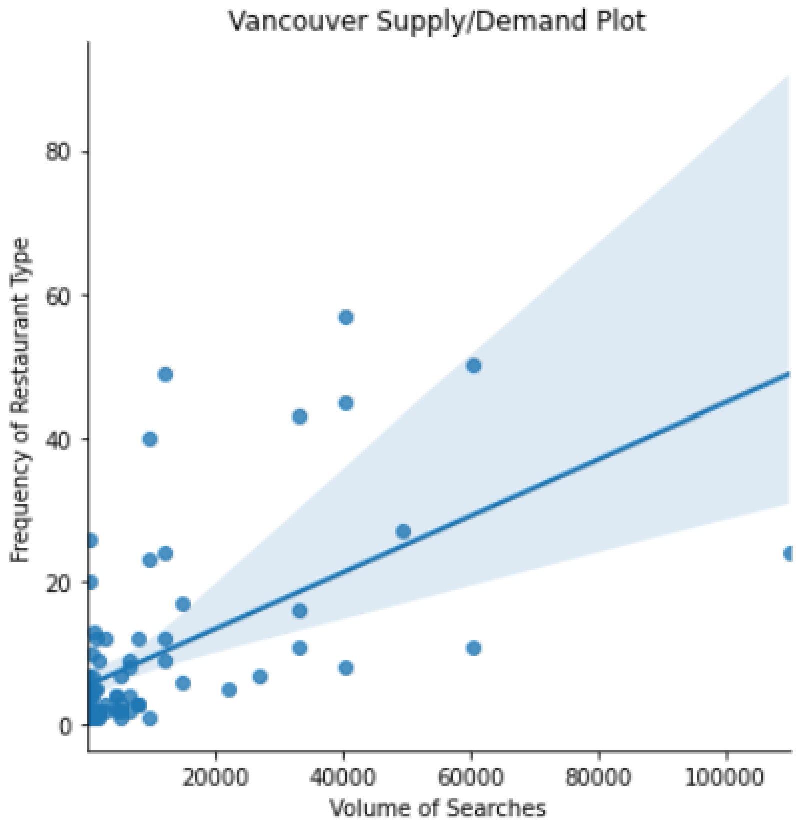
Montreal on the other hand is clustered into three groups rather than two.

- Red: Pizza, Fast Food, and Italian Restaurants.
- Purple: Cafés, Bakeries, and Asian Restaurants.
- Teal: General Restaurants, Sandwiches, and Diners.

The purple clusters dominate most of Montreal, which should be unsurprising given the popularity of Cafés and Bakeries among French-Canadians.

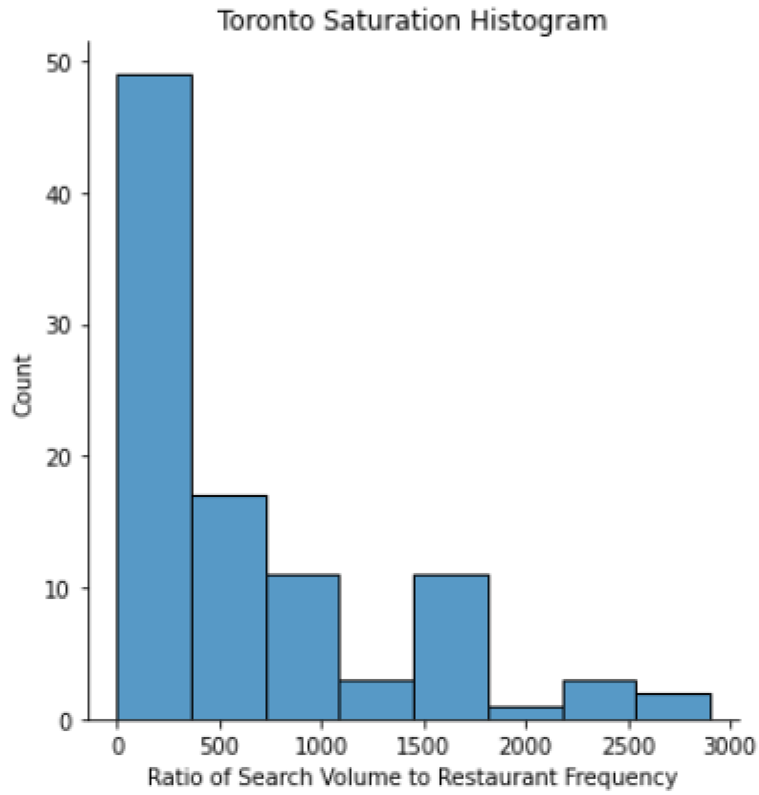
Vancouver was a very interesting case, as each neighbourhood was quite diverse, and the K-means algorithm was unable to detect any meaningful clusters of the data.

From there, using SEMRush data for each Restaurant category in the datasets, I created a spreadsheet corresponding to the frequency of each restaurant type (in each city) as compared to the search volume for that category among Canadians. Plotting the data for Vancouver of search volume as compared to the number of each kind of restaurant can be seen below:



This graph may be difficult to interpret due to the shortcomings of using search volume as a demand heuristic. Even so, there is a positive relationship between search volume and restaurant type frequency.

To assess the saturation of each restaurant scene, we can plot histograms based on the ratio of search volume to the frequency of restaurant type.



We can tell from this chart that most restaurant types in Toronto have a low ratio of search volume to restaurant frequency (less than 500 searches/venue). This indicates that there are far too many restaurants of certain types that customers are not actively searching for. One example would be Pizza Places: there is a relatively high search volume for these places, but there is also a surplus of Pizza Places!

4. Results

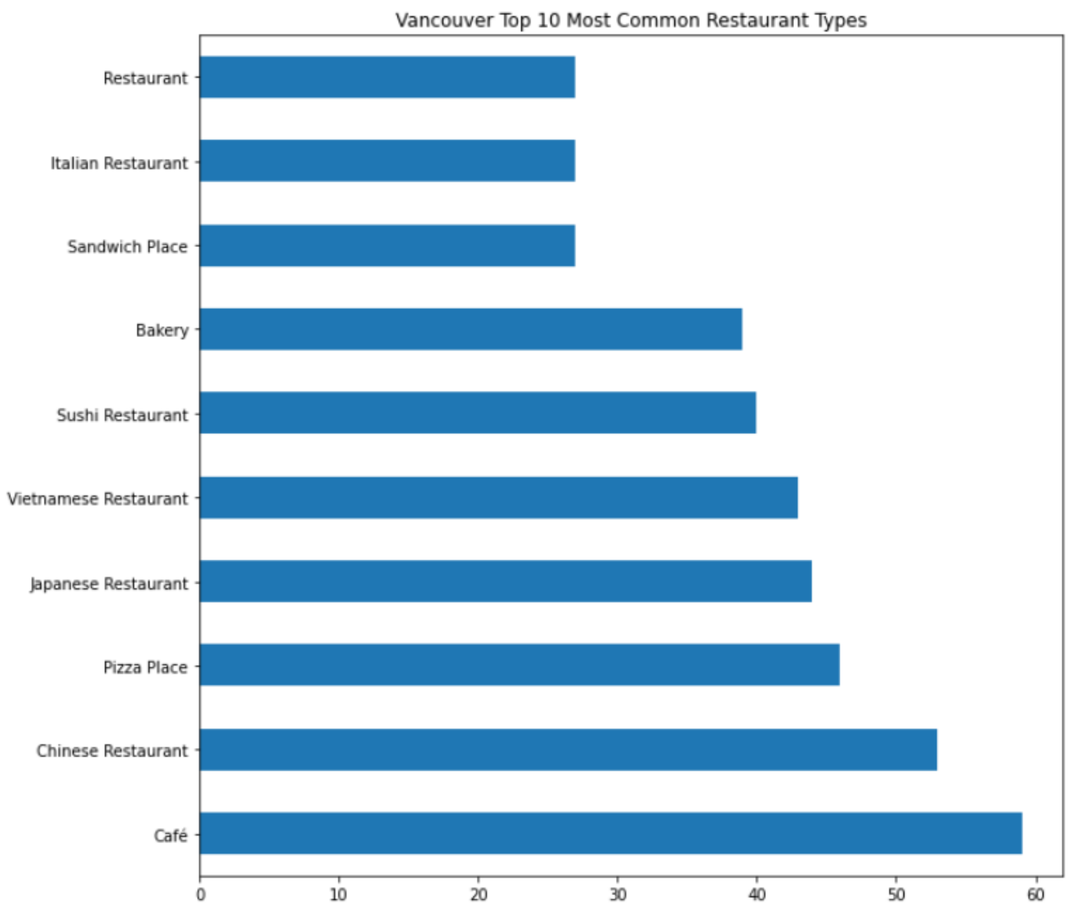
4.1 Restaurant Diversity

In terms of restaurant diversity, each city has different strong points.

When it comes to the overall number of different restaurant categories, Toronto is the winner with 104 different restaurant types, followed by Montreal with 98 unique venue types, and Vancouver in last with 81 types. This metric is interesting, though Toronto is also the greatest in land size, so if you were to assess the data as different categories/area, the results would be the opposite, as Vancouver has the greatest diversity in such a small land size.

Regarding neighbourhood diversity, as mentioned previously, Vancouver has the most individualistic neighbourhoods of the bunch. This is followed by Montreal, with Toronto having the least neighbourhood diversity.

As for the distribution of the most frequently occurring restaurant types, the city with the most even distribution is Vancouver, followed by Montreal, and with Toronto in last.



4.2 Meeting the Demand of The Public

When assessing each city in terms of how well it satisfies its residents' demands, we can look at the keyword data based on search volume.

For the Supply/Demand plots generated (like the one for Vancouver in the Methodology section), Vancouver's plot had the highest Pearson correlation of 0.56, followed by 0.51 for Toronto, and 0.41 for Montreal. These results are pertinent to the discussion as the Pearson correlation is a measure of how strongly two variables are correlated. In this case, that means how the frequency of restaurant types reflect the demand for them.

Something also to note for this information: Vancouver has fewer total categories, so even though its restaurant frequency most strongly reflects Canadian search demand, it also lacks many restaurant types entirely.

5. Discussion

From the results given above, we can see that each city does better in certain metrics, and worse in others. There is no one obvious answer to which is the most diverse, nor to which city meets the demands of Canadians the most.

Looking at the problem from the pure size perspective, Toronto is the clear winner. Even if its restaurant scene does not strongly reflect the demands of Canadians, it has the most restaurants and the most different kinds of restaurants. Further, it is the least saturated scene as indicated by the search ratio histograms.

When viewing the data in terms of diversity/size, Vancouver has the upper hand. It is by far the smallest of the three cities, though it reaches similar metrics in each category, and is even the best in neighbourhood diversity and meeting the search demands. When it comes to value/size, Vancouver beats out the other two cities by far.

Montreal is a little different from the other two, as it almost always landed in between the two other cities in each metric. It does not have as high neighbourhood diversity as Vancouver, but it is better than Toronto in that respect. It loses to Toronto in the sheer number of categories but beats Vancouver at that. Overall, Montreal is the middle ground option between the two other cities.

6. Conclusion

As you can see, each of the three cities is unique in their own way and have strengths that the others do not possess. I am quite confident in saying that each city offers a wide variety of different cuisines and are representative of the multicultural nature of Canada. For those interested in what kind of restaurants are in highest demand, some of the highest in the list include:

- Indian Cuisine
- Diners
- Malay Cuisine
- Bistros
- Steakhouses
- Afghan Cuisine
- Food Trucks
- Turkish Cuisine

The most saturated restaurant types by far are Pizza Places, Sushi Restaurants, and Cafés.

For those interested in more comprehensive results, feel free to message me, or look at my GitHub Repository (in the Resources section).

This project is the culmination of what I learned from the IBM Applied Data Science Specialization, offered by Coursera.

Resources

[1] Foursquare API: https://developer.foursquare.com/docs/

[2] SEMRush Marketing Platform: https://www.semrush.com/

[3] Wikipedia: https://en.wikipedia.org/wiki/Main_Page

[4] Service Objects: https://www.serviceobjects.com/

[5] Github Repository: https://github.com/ReidZaffino/Coursera_Capstone