Choosing between Toronto and Montreal: A Clustering Approach

I. Introduction

A. Background and Problem

The decision to choose a place to live and settle can be a daunting experience, either with family or alone. Nowadays, many immigrants are directing their attention to Canada, as the country has been very receptive to them.

Canada is a very appealing country with an economy constantly growing and many opportunities. The most popular Canadian cities are truly vibrant and dynamic. Of course, as an immigrant, choosing a city to live in is much more complex than appreciating the beauty of a city from a touristic perspective.

Toronto and Montreal are the two largest economies in Canada. Before COVID-19, both cities had low unemployment rates and offered a high quality of life to their residents. Immigrants have a wide variety of factors to consider when deciding where to live. Among those factors, some are specific for each case (whether the person has a job offer or not, whether French is a barrier or not, etc.). However, this data science project aims to ease the analysis of the non-specific factors to consider when choosing between Toronto and Montreal. I will first use descriptive statistics of economic and demographic data. Then, I will use a clustering approach to analyze how similar or different are both cities.

B. Audience

The audience of this data science are immigrants that face the decision to move either to Toronto or Montreal. The results of this study are not expected to be definitive. They are only going to provide information to these immigrants so that their decision is easier and can be concentrated on the specific factors that matter for them.

II. Data

I will use data from the OECD to make a comparison of the two cities and put them in the context of other Canadian cities. Other governmental or private organizations' data will be part of this first exploratory analysis.

For the clustering section of the analysis, I will use location and zip code data from geonames.org in addition to Toronto data provided by IBM and Coursera for a previous project. I will cluster the zip codes of both cities, as separated datasets, according to the type of venues that exist in each city. This will allow me to understand if the clusters from one city are similar to the clusters from another city. Finally, I will cluster the zip codes of both cities as one dataset to see if the algorithm separates the boroughs in clusters isolated geographically because of being very dissimilar. If otherwise, clusters do not present geographic isolation it would mean that both cities are very similar in terms of the venues located around each zip code.

III. Methodology

The OECD data is from their “Cities” dataset in which they offer data per city of each country. I downloaded the data using the queries system and structured it using python panda’s library. The variables obtained from this dataset were the population of greater areas of Canada (cities including surrounding suburbs and neighborhoods), unemployment rates, real GDP (PPP adjusted), and real GDP PPP per capita. The OECD data was complemented with a comparison of the cost of living in each of the Canadian cities. Data on the cost of living comparison was web-scrapped from the expatistan website.

The “geonames” data was also web-scrapped and formatted using python pandas’ library. Then, with the geolocation data of each zip code, I requested data of the venues around them from the Foursquare API.

I used a k-means machine-learning algorithm to cluster the zip codes. The clustering was performed in two sections. In the first section, I clustered Toronto’s zip codes and then Montreal zip codes as two separate datasets. In the second section, I clustered Toronto and Montreal’s zip codes as one dataset.

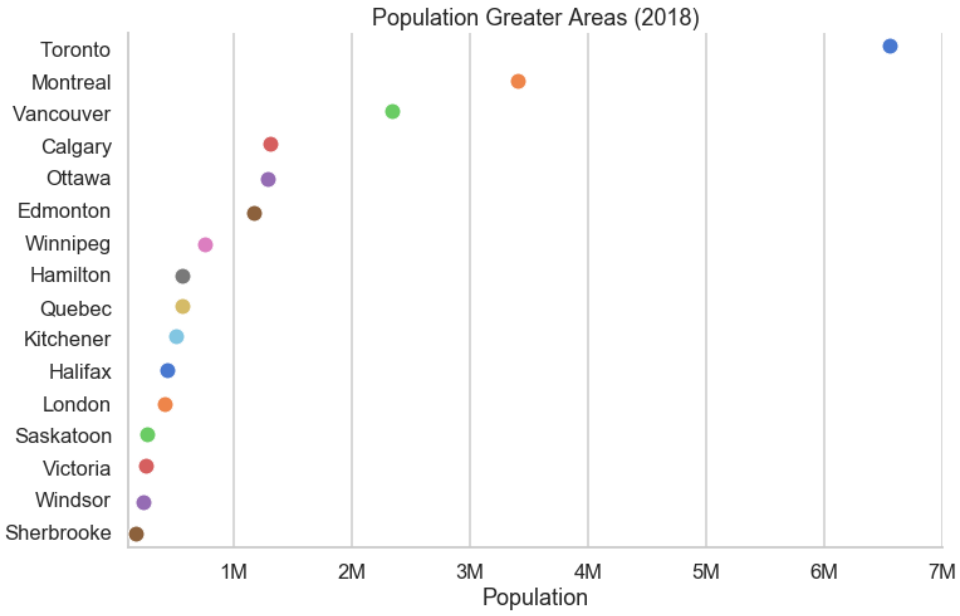
IV. Results

For the sake of objectivity, I decided to perform my Toronto-Montreal comparison using zip codes as a geographical delimitation. In the case of Toronto, we analyzed places with zip codes starting with “M” and, in the case of Montreal the places with zip codes starting with “H” (excluding the North Pole).

The decision of using zip codes as the parameter for delimitation conflicts with the existing data of the two cities for the descriptive statistics analysis. In some cases, it is possible to find data of “greater areas”, which is a broader area than my zip code delimitation. In other cases, data reported is only for the city of Toronto and the city of Montreal.

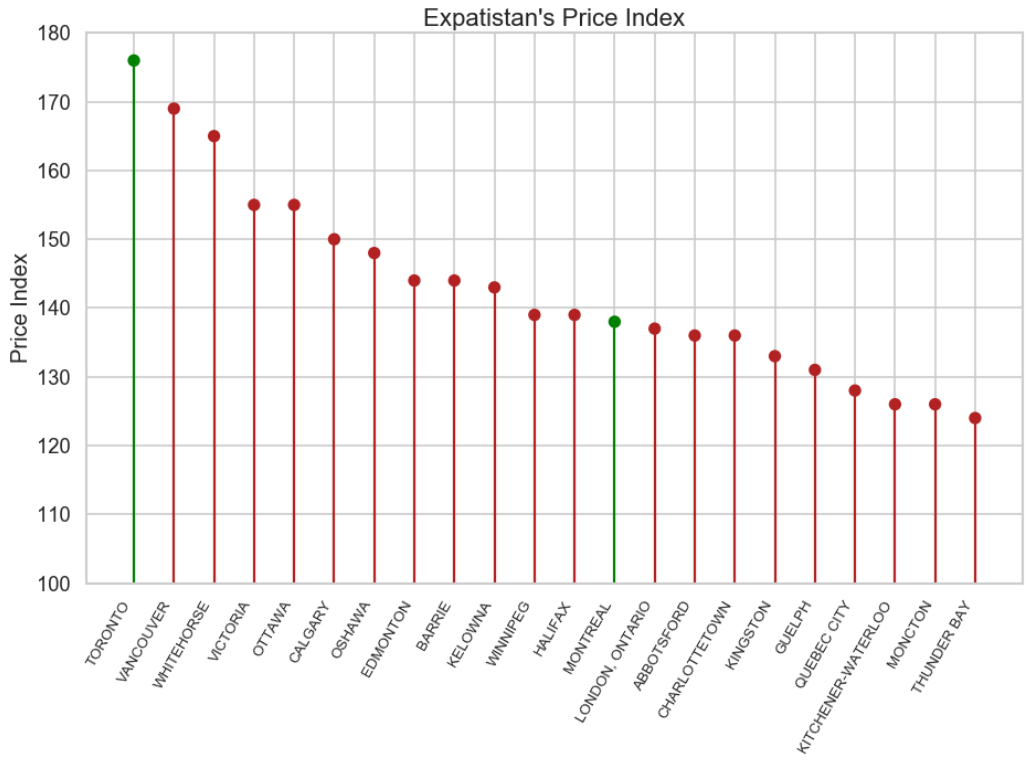
A. Descriptive Statistics

According to the OECD, Toronto Greater Area and Montreal Greater Area were the two most populated regions in Canada in 2018. Toronto Greater Area had a population of approximately 6.5 million people and Montreal Greater Area had a population of approximately 3.4 million people. The following graph puts these numbers into context with other “greater areas” of Canada.

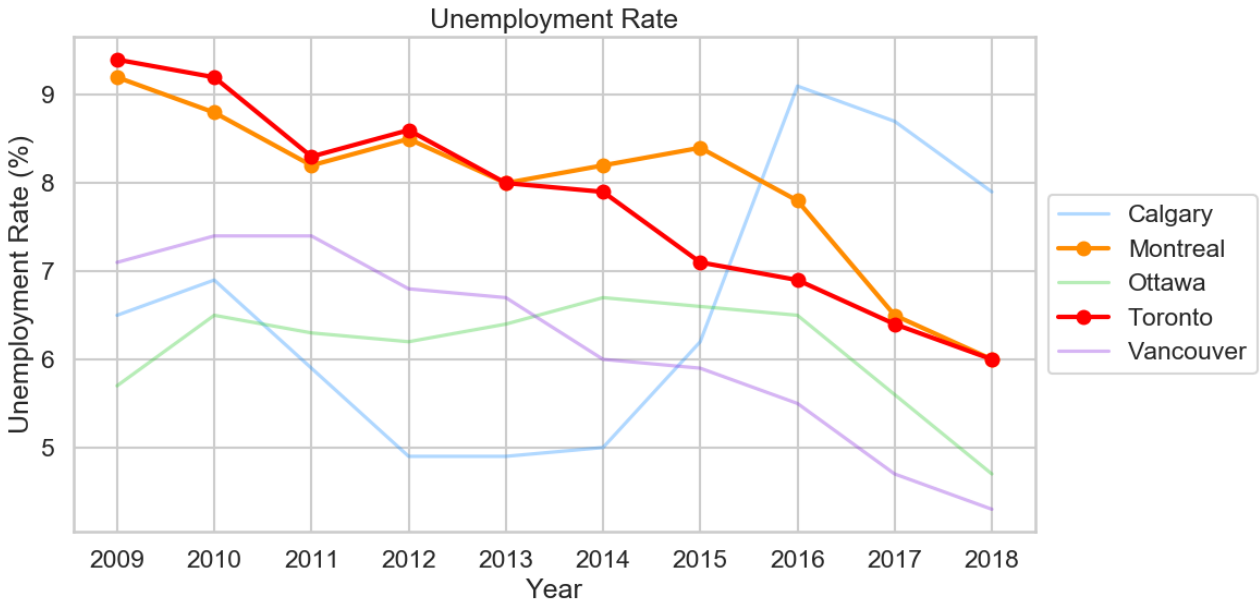


The data extracted from expatistan reveals that Toronto is a more expensive city than Montreal. Expatistan’s methodology compares how much more expensive is a city than a base city (Prague)

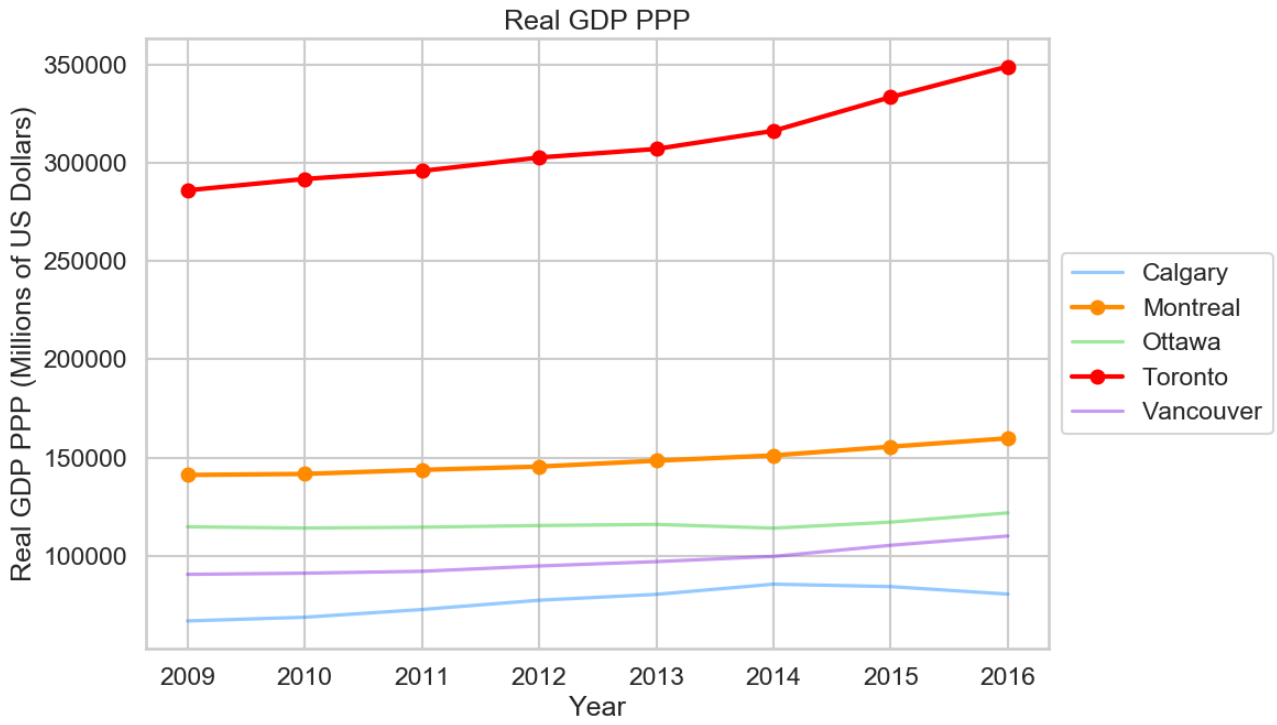
to construct a “Price Index”. Therefore, if a city has a price index of 150, it means that the city is 50% more expensive than Prague. The following graph shows Toronto and Montreal’s price index in comparison with other Canadian cities.



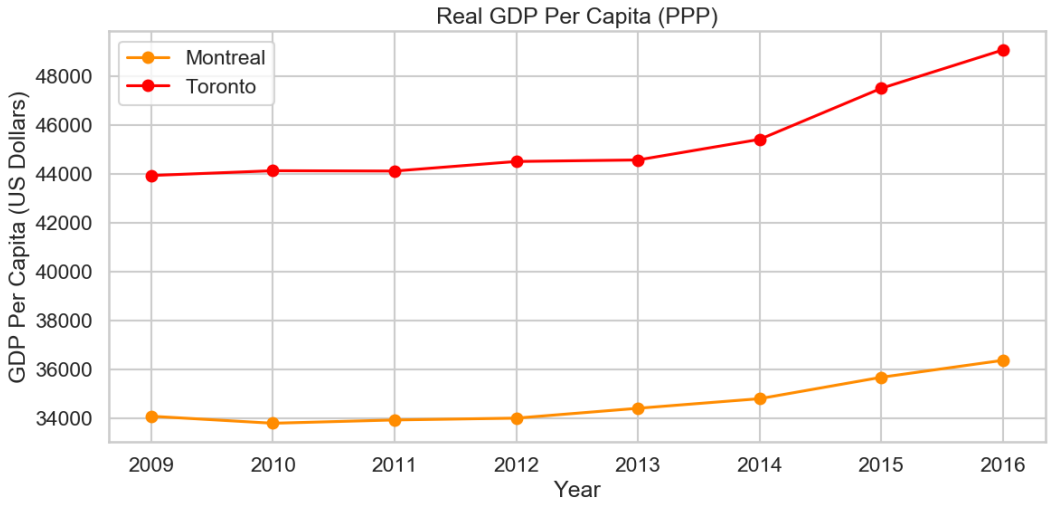
Despite the large population of Toronto and Montreal, before COVID-19, they both presented low unemployment rates. The OECD reported a 6% unemployment rate for Toronto and Montreal, as shown in the following graph.



The OECD reports the real GDP with the purchasing power parity adjustment to make better comparisons across different cities around the world. Unfortunately, their data do not provide observations of Canadian cities after 2016. Despite the missing values, all Canadian cities appear to have a very stable GDP growth trend. The dataset shows that in 2016 Toronto's real GDP PPP was approximately 349 billion US dollars and Montreal's was approximately 160 billion US dollars.

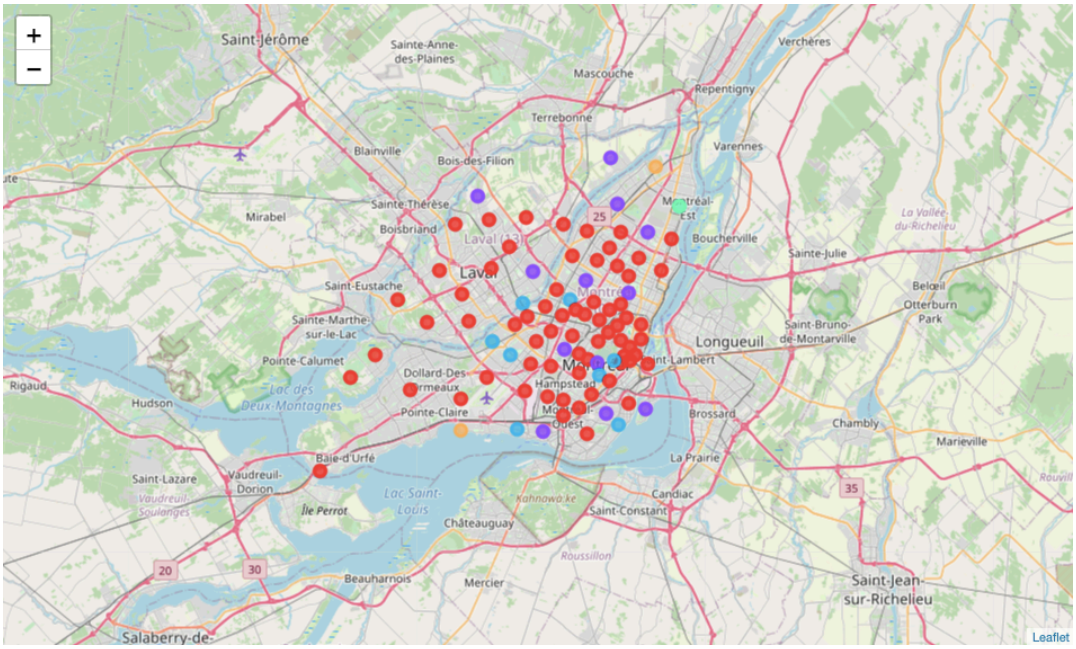


I was also interested in looking at the GDP per capita to understand individual productivity and, more importantly for potential immigrants, use it as a proxy for individual wealth. As expected, both cities present the same trend as the real GDP PPP graph.

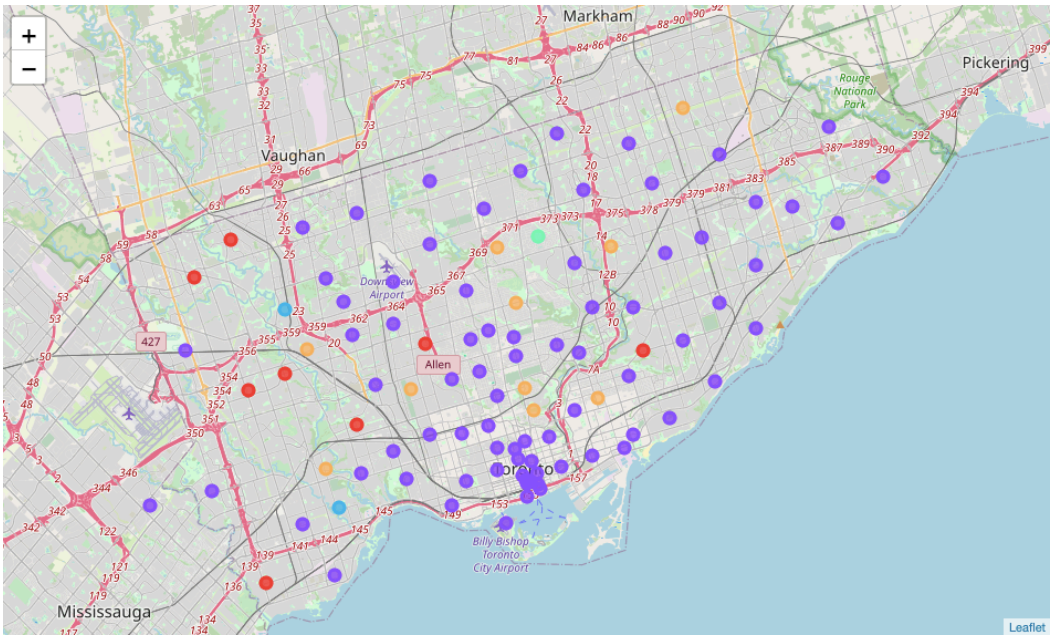


B. Montreal and Toronto Clusters Analysis

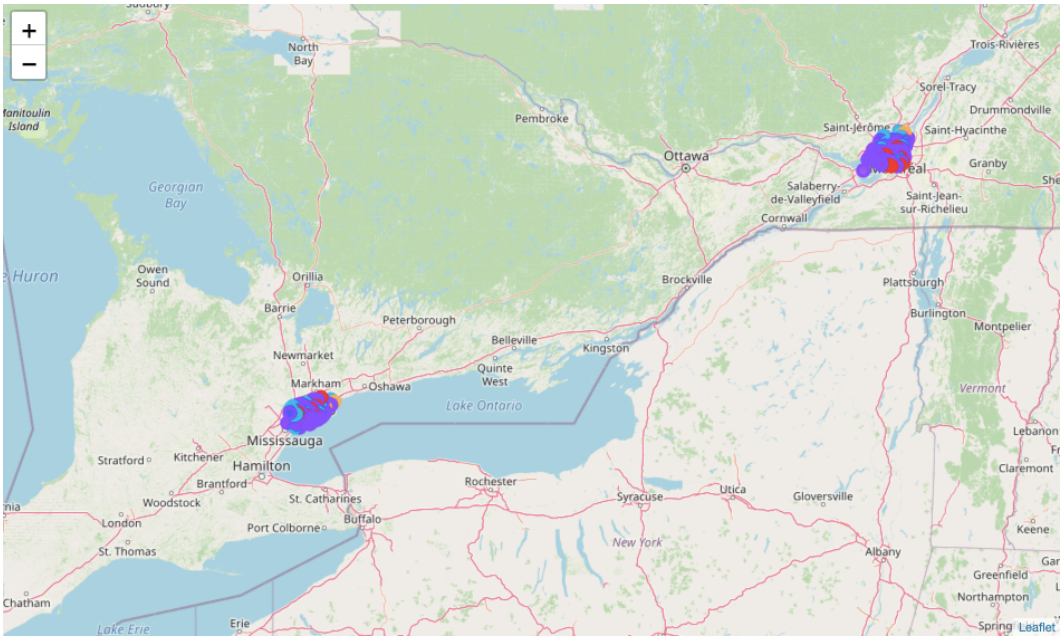
Clusters generated with the k-means algorithm for Montreal's zip codes show the presence of one big cluster, two medium-size clusters, and two very small clusters. The largest cluster contains a lot of food venues like cafes, bars, and international restaurants. The second-largest cluster is one with still a lot of food venues but with a considerable number of parks and other recreation venues. The third cluster shows a predominance of parks and recreation places. The two other clusters show a mixture of venue types with some Asian and Mediterranean food restaurants.



In the case of Toronto, the clustering analysis produces a similar result. One major cluster with mainly food venues, two medium-size clusters similar to Montreal's medium-size clusters. Curiously, one of these two clusters is mainly located in the west of the city around the Etobicoke area. Finally, there are two small clusters with neighborhoods in Etobicoke and North York which also show the presence of Asian and Mediterranean restaurants.



When analyzing the clusters of Toronto and Montreal as one dataset, I confirmed what was observed with the two separate datasets. No geographical isolation of one cluster is present because the two major clusters of each city are indeed very similar. This result was expected given the fact that both cities are both Canadian and are very diverse.



V. Discussion

The results of this comparison between Toronto and Montreal show that both cities present a lot of similarities. Toronto is a more populated city and, thus, it is likely a more diverse city. However, the clustering analysis shows that in terms of venues both cities have a similar degree of diversity, judged by the multicultural offer of food venues. Regarding the population statistics, a potential immigrant might want to analyze the density of the population since Toronto has a significantly larger population.

Before COVID-19, the two cities presented a reasonable degree of opportunities for immigrants looking to join the labor force of either city. Unemployment for both cities in 2018 was around 6%. Toronto is the largest economy within Canada with a substantially higher GDP. Toronto is also ahead of any other Canadian city in the wealth of its population if judged by GDP per capita. However, Toronto appears to be the most expensive city in Canada and it is way ahead of Montreal. Potential immigrants might want to consider GDP per worker if they are aiming to join the labor force of either Toronto or Montreal.

The contrast between wealth and cost of the cities makes it difficult to recommend one city over the other for potential immigrants. It ultimately depends on their profession, individual job opportunities, and the most active sectors of the economy of each city.

The clustering analysis shows that both cities are very similar if judged by the types of venues. Montreal and Toronto present a lot of immigration, which is reflected in the wide variety of international restaurants. Potential immigrants should not be afraid of any type of discrimination against them because both cities are truly multicultural.

Both cities have neighborhoods that are mainly city places. They both also have other neighborhoods that offer access to parks and recreation centers in conjunction with some city life. The clustering analysis does not reveal any major difference between Montreal and Toronto that immigrants should factor in their analysis.

VI. Conclusion

This study reveals that the decision to live in Montreal or Toronto ultimately depends on the specific circumstances of each person. Montreal and Toronto are multicultural cities with similar characteristics and venues.

The descriptive analysis shows that Toronto and Montreal, as the two largest economies within Canada, present a lot of opportunities for potential immigrants. There is a contrast between potentially higher wages in Toronto and a lower cost of life in Montreal.

The clustering analysis strengthens the argument of similarity between the two cities. It does not reveal any major difference between Montreal and Toronto that immigrants should factor in their analysis.