

Identifying the best neighborhood cluster for new immigrants in Toronto

Swapnil Deshbhratar

04/08/2020

Introduction

1.1 Background

Canada is a popular destination for immigration around the globe. In 2016, Canada had 1,212,075 new immigrants who had permanently settled in Canada from 2011 to 2016. These recent immigrants represented 3.5 % of Canada's total population in 2016 according to Statistics Canada which is a federal government body responsible for government statistics. The introduction of Skilled workers Express Entry program by the Canadian government and its commitment to the program has attracted an increasing number of immigrants to the Canadian shores. Toronto being the financial centre with a reputation for multiculturalism, is favoured by most immigrants as their new home. However, moving to a new country is not an easy task. Apart from completely rebooting their lives and starting afresh, most of the skilled workforce that lands in Toronto has to actively pursue various job opportunities and master relevant skills set to start their integration with the Canadian workspace. Moving around in a new city, either to look for new job opportunities or for attending any educational institute can be troublesome, especially in the first couple of months after arrival. Thus, in the early days it is imperative that a new immigrant family chooses an appropriate neighbourhood for their stay.

1.2 Problem

The objective of this IBM Applied Data Science Capstone project is to identify a suitable cluster of neighbourhoods for a new immigrant in Toronto. The choice of a perfect neighbourhood may differ for individuals as per their likes and preferences. However, the purpose of this project is not to find the perfect neighbourhood for every immigrant for the long term but to find for new immigrants the best cluster of neighbourhoods to stay in the early days after their arrival to Canada .

1.3 Interests

Identification of suitable neighborhoods to stay should be of interest to new immigrants who are planning to start their lives in Toronto.

Data Acquisition and Cleaning

2.1 Data Sources

The City of Toronto is divided into 140 neighbourhoods as per the Municipality of Metropolitan Toronto open data set . The data on these 140 neighbourhoods with latitudes and longitudes is available on the City of Toronto open [data portal](#). Apart from this, the location of various Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes is available in 3 different datasets on the City of Toronto open data portal. The Foursquare location data for the 93 neighbourhoods along with location data of Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes was used in this capstone.

2.2 Data Cleaning

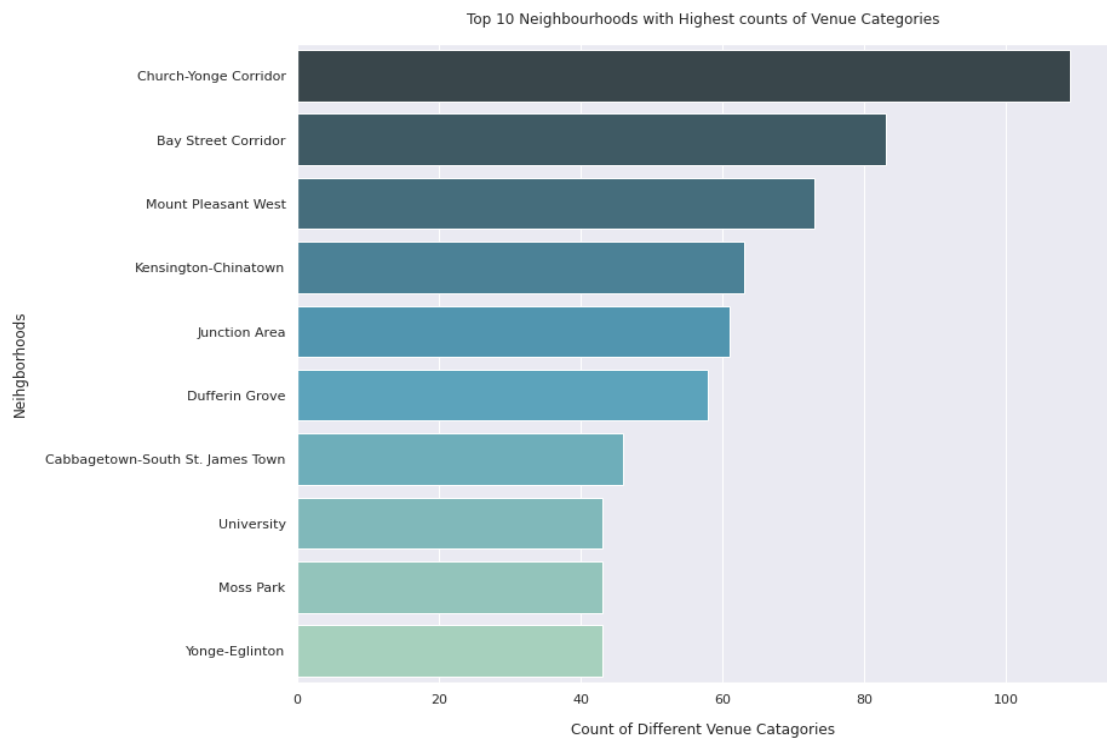
The open data available on the portal of Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes located in the city of Toronto does not have the latitudes and longitudes of these places. But these 3 datasets contain the neighborhoods in which these institutes are located. However, there is some dissimilarity in the names of the neighbourhoods in the neighbourhoods data obtained from the portal and other 3 datasets. The names of the neighborhoods were formatted to be similar in all the 4 datasets obtained from the portal (Neighborhoods, Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes). The cleaning was done on the basis of the numeric codes given to each neighborhood which is available in all the 4 datasets. For example, Roncesyalles is coded as Roncesvalles (86).

Methodology

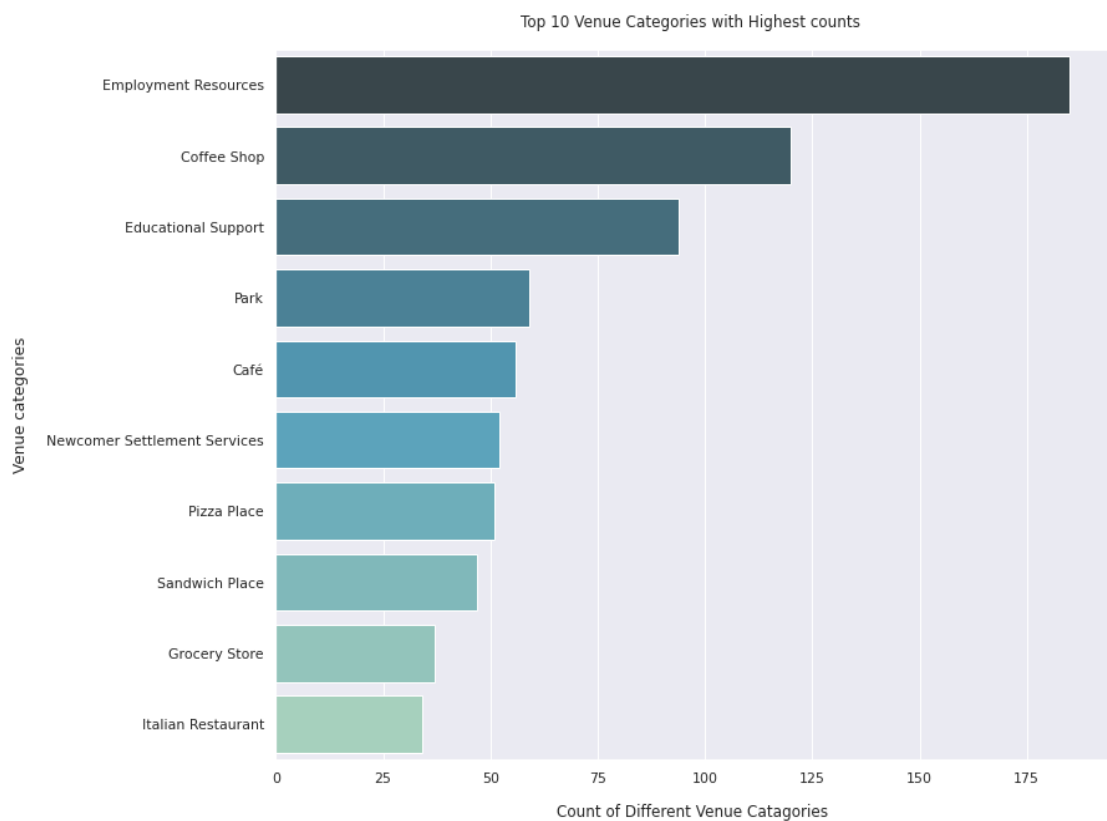
3.1 Data Analysis

For a new immigrant it is favourable to have Newcomer settlement services institutes, Educational Support institutes or Employment resources institutes in their chosen neighborhood. Thus, the project focused on neighborhoods that have either of the above institutes present. By using the Foursquare location data for the 93 neighbourhoods containing either of the above institutes, the locations and category of different types of venues within 500 metres of a point in the neighborhoods was obtained in a pandas dataframe. This dataframe was merged with 3 data frames containing the Newcomer settlement services institutes, Educational Support institutes and

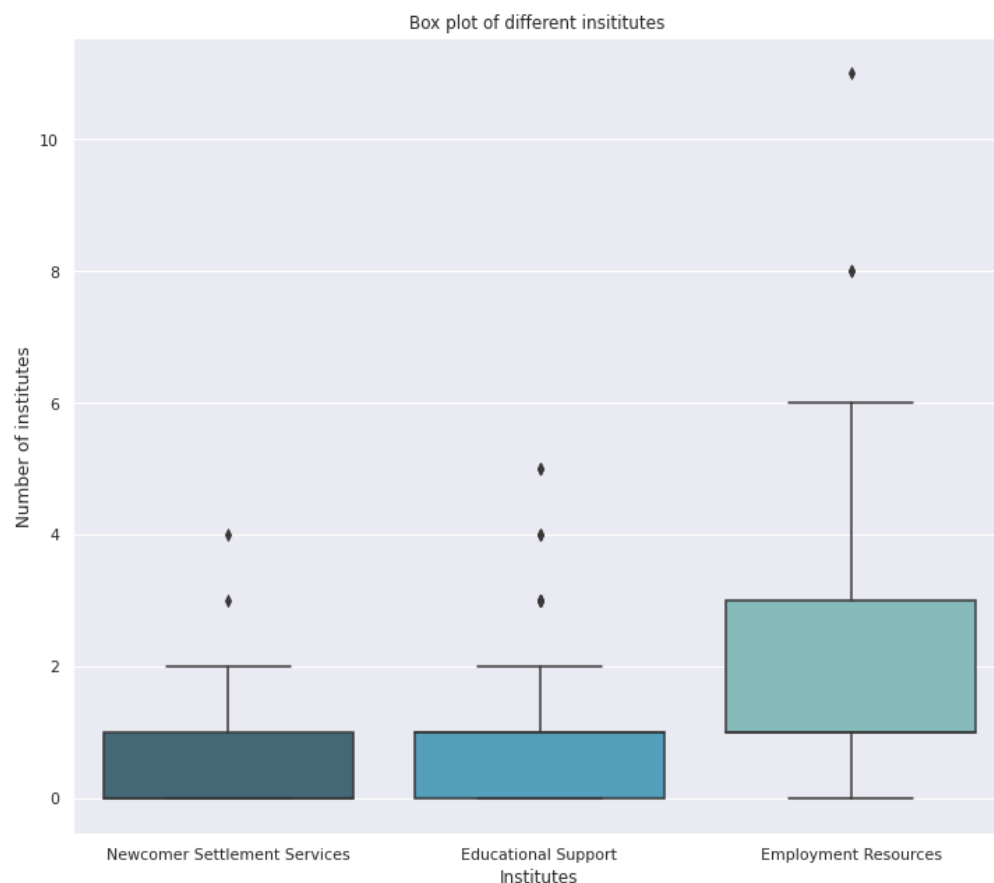
Employment resources institutes data to obtain the final dataframe. The top 10 neighborhoods with the maximum number of venues locations are shown in figure 1.



Also, the value counts of top 10 different venues category is shown in figure 2.

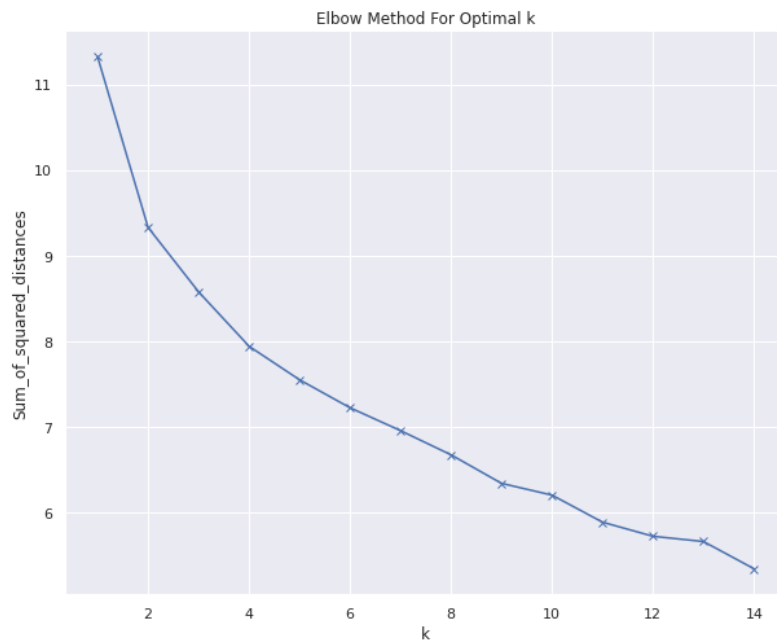


It can be seen from the above figure that Newcomer settlement services institutes, Educational Support institutes or Employment resources institutes are among the top venue categories and will largely influence the characteristics of neighborhoods. From the Foursquare API and additional data, 261 unique categories of venues were found in Toronto. The box plots showing Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes is shown in figure 3.

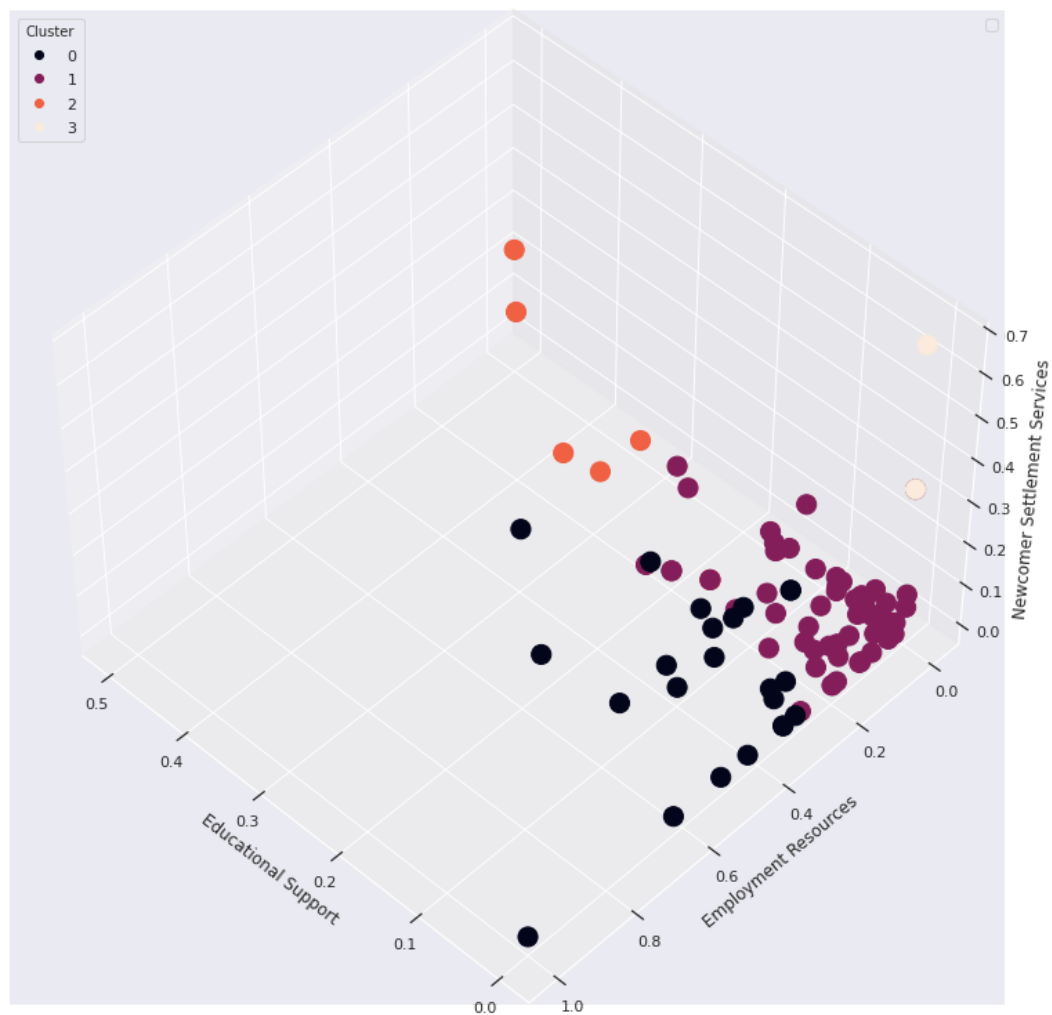


3.2 Clustering

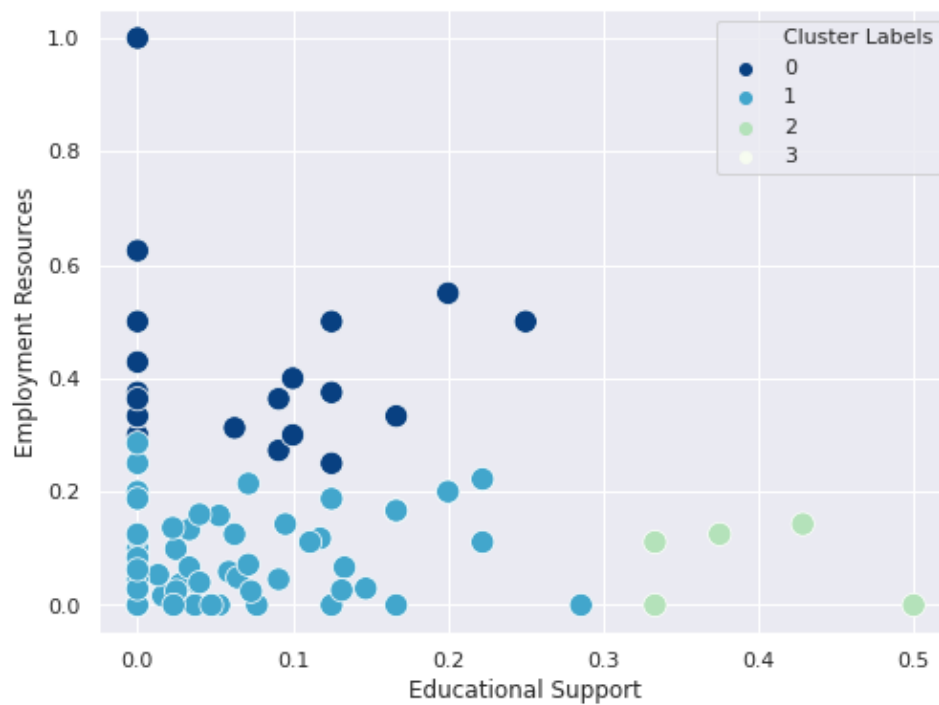
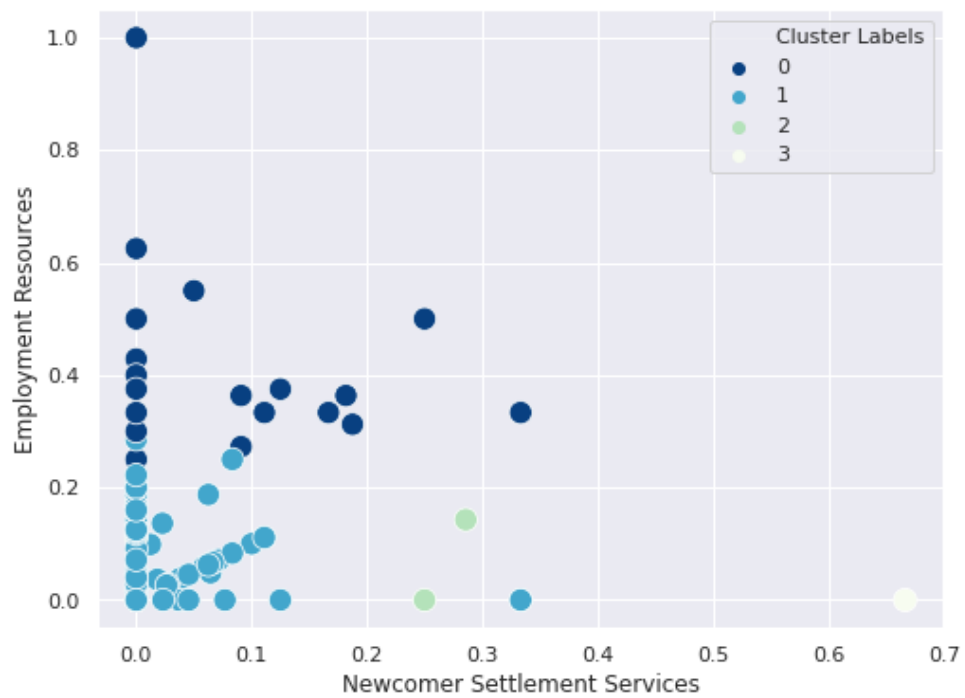
The pandas dataframe containing mean number of venues in each neighborhood for all the features(venue categories) is transformed using StandardScaler() function from sklearn library. By using K-means clustering method, neighbourhood clusters were found out. The optimum value of k was found using elbow method and is 4. The plot for elbow method is shown in figure 4

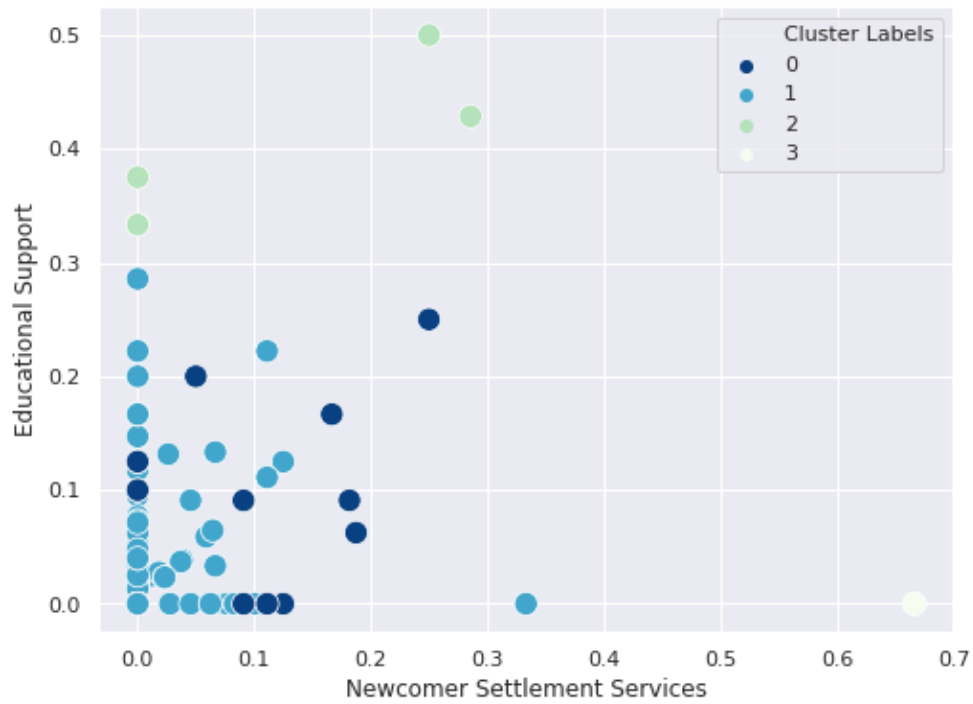


The distribution of the clusters with respect to Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes is shown in figure 5.



The Distribution of clusters with respect to two features is shown in Figure 6,7 and 8.





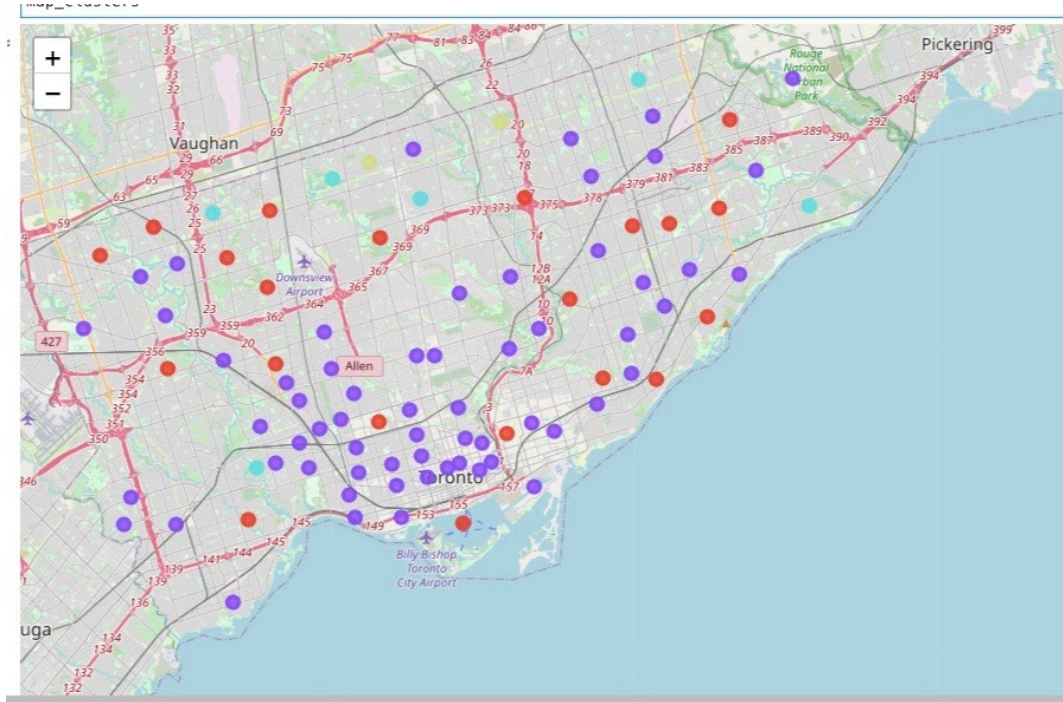
The mean top 5 venue categories for each cluster is shown in table1.

Cluster Labels	Number of Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	23	Employment Resources	Park	Newcomer Settlement Services	Educational Support	Coffee Shop
1	62	Employment Resources	Coffee Shop	Educational Support	Pizza Place	Café
2	6	Educational Support	Park	Employment Resources	Newcomer Settlement Services	River
3	2	Newcomer Settlement Services	Park	Residential Building (Apartment / Condo)	Photography Studio	Egyptian Restaurant

Results

4.1 Cluster Map

The map of Toronto showing various neighborhood clusters is shown in figure 6.



4.2 Cluster Characteristics

Cluster '0' or High Employment Resource Cluster : From the figure 5 and table 1 it can be seen that this cluster has sufficient number of Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes with highest values for employment resources.

Cluster '1' or Mean Cluster : From the figure 5 and table 1 it can be seen that this cluster has a maximum number of neighborhoods. It has sufficient number of Educational Support institutes and Employment resources institutes but less number of Newcomer settlement services institutes. The values of Employment resources institutes is less than that of Cluster '0'.

Cluster '2' or High Educational Support Cluster : From the figure 5 and table 1 it can be seen that this cluster has sufficient number of Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes with highest values for educational support institutes.

Cluster'3' or Single Feature Cluster: This cluster has only 2 neighborhoods and only has Newcomer settlement services institutes.

Discussion

For a new immigrant **High Employment Resource Cluster** or **High Educational Support Cluster** is more suitable as it has appropriate concentration of the features. The final choice however will depend upon the presence of other venues in the neighborhood and preferences of the immigrant. Also, in order to find the ideal neighborhood cluster for a new immigrant in Toronto, only neighborhoods with at least one of the 3 (Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes) institutes were considered for K means clustering. Thus, out of 140 neighborhoods only 93 were considered for clustering. Out of 93 neighborhoods High Employment Resource Cluster with 23 neighborhoods and High Educational Support Cluster with 6 neighborhoods were found to have good concentration of the 3 institutes.

Conclusion

New immigrants who are planning to start their lives in Toronto can choose either **High Employment Resource Cluster** or **High Educational Support Cluster** depending upon their preference for education or job search. If the immigrant is looking for job opportunities then Brookhaven-Amesbury, Cliffcrest, Danforth Village, Kingsview Village, The Westway, Taylor-Massey, Waterfront Communities-The island, Woburn from **High Employment Resource Cluster** would be ideal choices and if immigrant is looking for stronger educational support then Westminster-Branson, Willowdale East from **High Educational Support Cluster** would be ideal choice.