

IBM Capstone Project

Identifying the best neighborhood cluster for new immigrants in Toronto

**Swapnil
Deshbhratar**

Introduction

1

Background

Toronto is a popular destination for immigration. Choosing an appropriate neighborhood to begin a new life in Toronto is essential.

2

The Problem

Identify suitable cluster of neighborhoods for a new immigrant in Toronto

3

Interests

New immigrants who are planning to start their lives in Toronto.

Data Sources

The location of neighborhoods, newcomer settlement services institutes, educational support institutes and employment resources institutes in Toronto obtained from City of Toronto open data portal

Data Acquisition

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. The Foursquare location data for the 93 neighbourhoods containing either of the above institutes and the Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes data was obtained in pandas dataframe

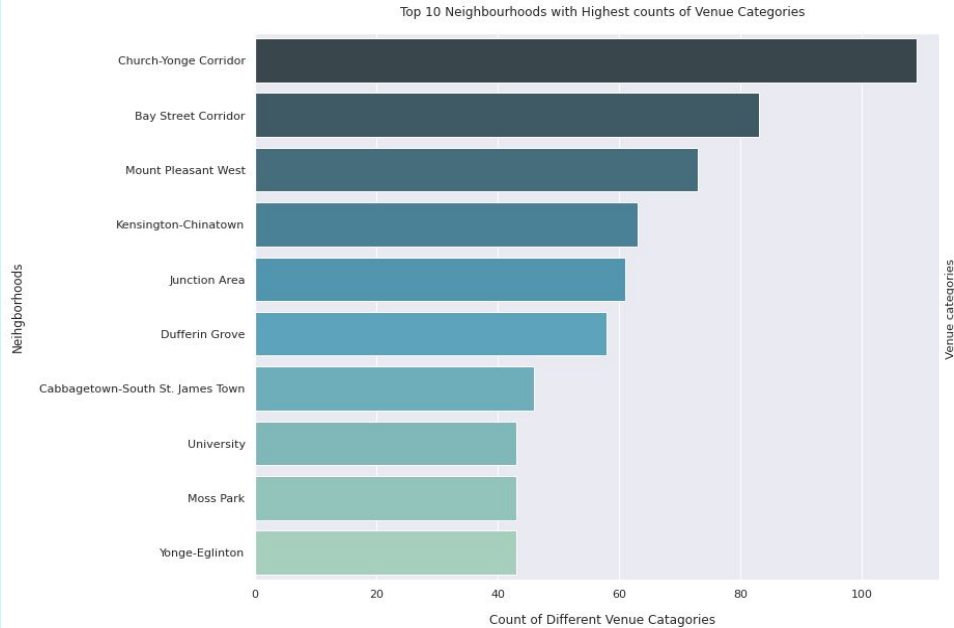


Data Wrangling

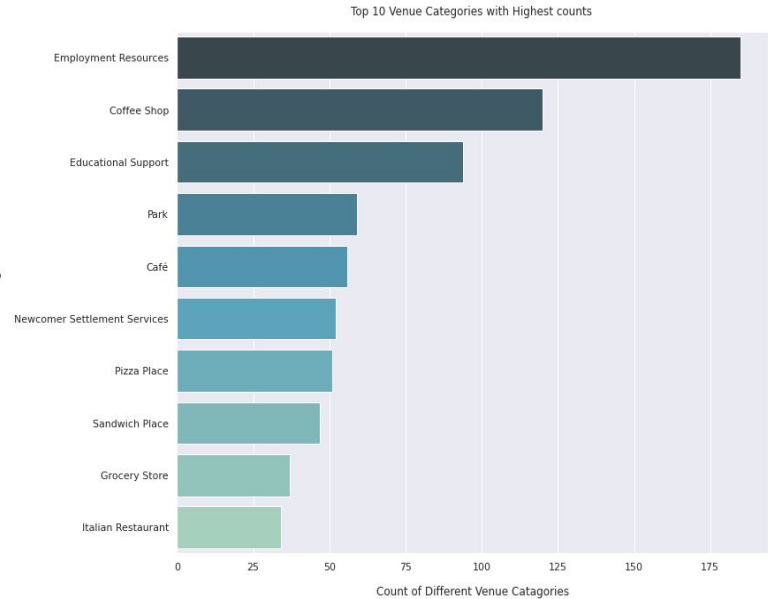
The names of the neighborhoods were formatted to be similar in all the 4 datasets on the basis of the numeric codes given to each neighborhood available in all the 4 datasets

Exploratory Data Analysis

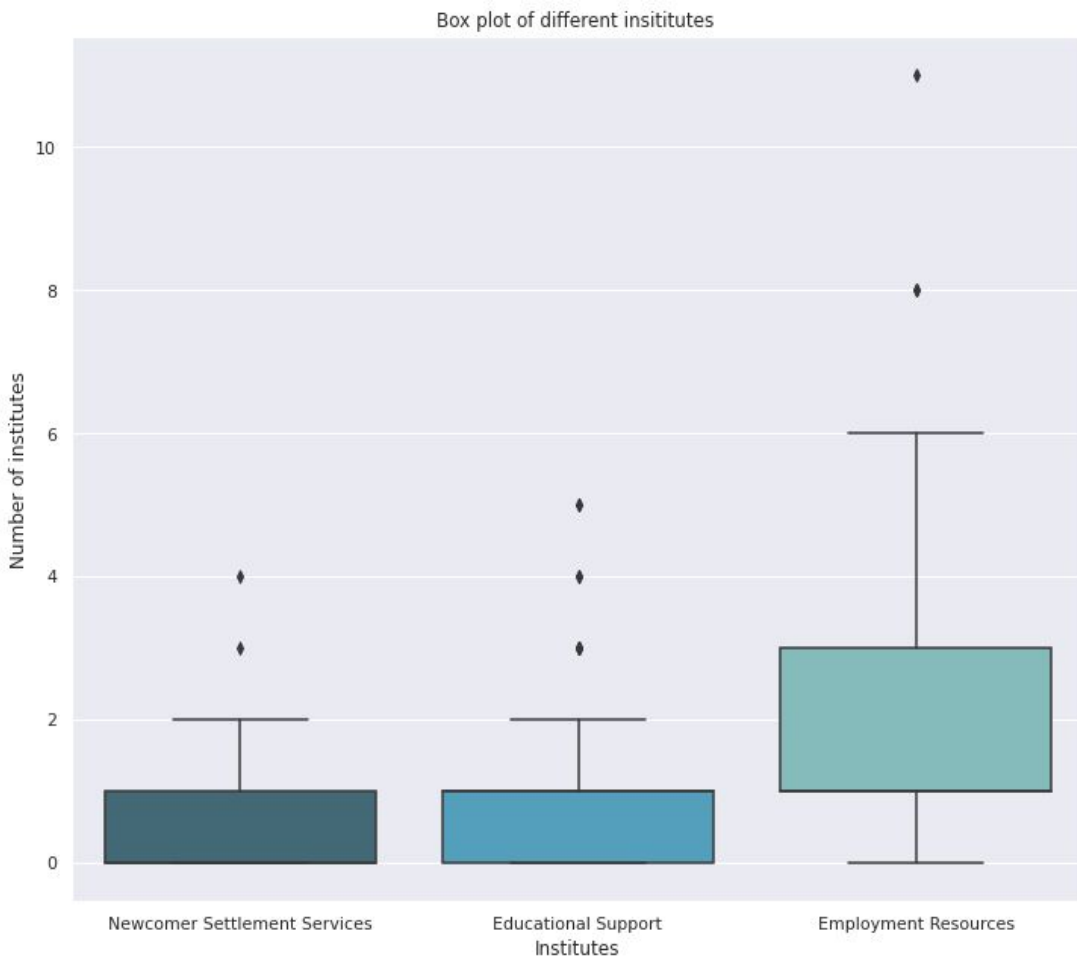
The top 10 neighborhoods with the maximum venues locations



The value count of top 10 venues category

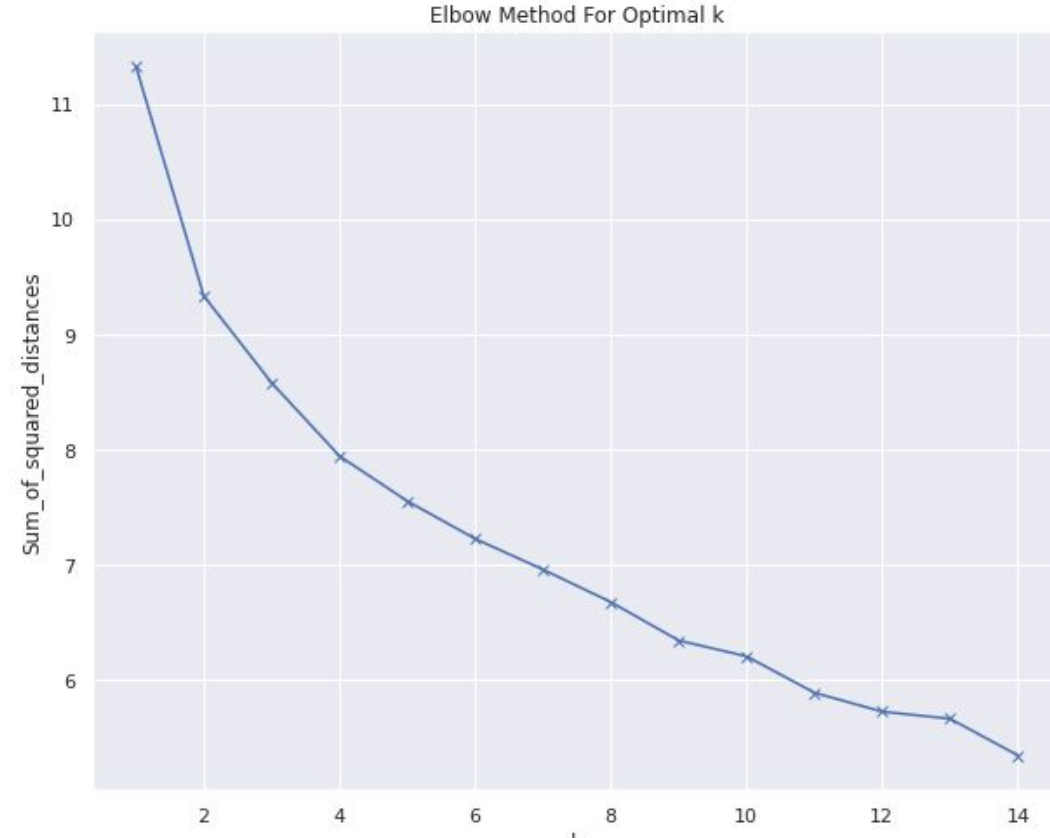


Data Analysis



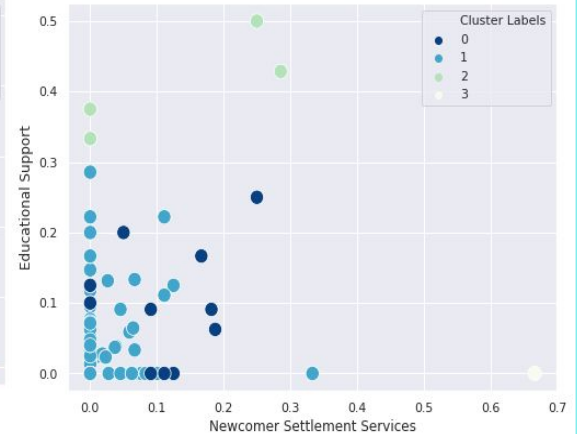
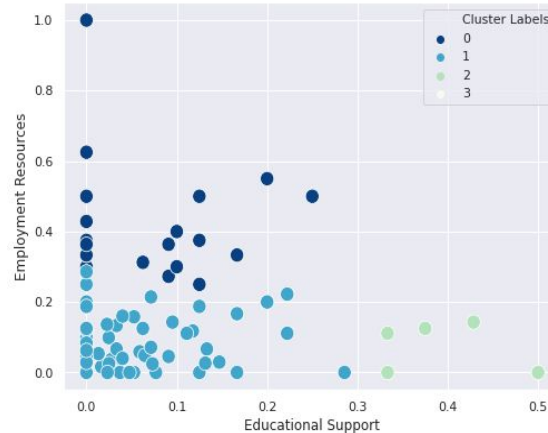
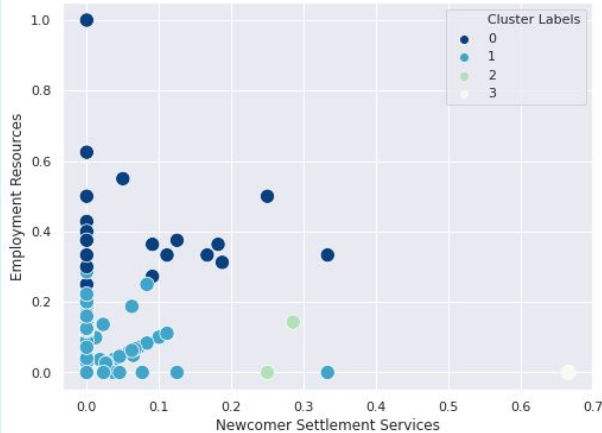
Newcomer settlement services institutes, Educational Support institutes or Employment resources institutes are among the top venue categories and will largely influence the characteristics of neighborhoods. From the Foursquare API and additional data, 261 unique categories of venues were found in Toronto

K Means Clustering



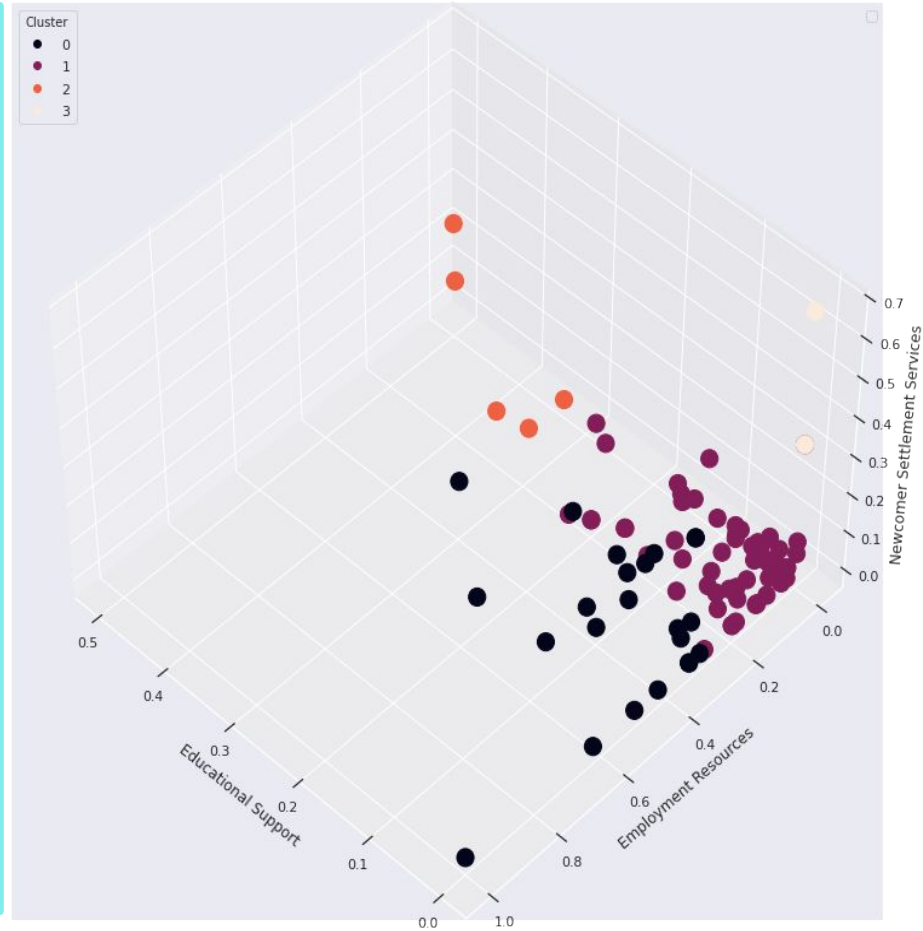
The pandas dataframe containing mean number of venues in each neighborhood for all the features(venue categories) is transformed using StandardScaler() function from sklearn library. By using K-means clustering method, neighbourhood clusters were found out. The optimum value of k was found using elbow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

Cluster Distribution with two features



Clusters

4 neighborhood clusters of Toronto with respect to Newcomer settlement services institutes, Educational Support institutes and Employment resources institutes



Cluster Characteristics

High Employment Resource Cluster(0)

Contains all 3 institutes with
high number of Employment
Resources Institutes

Mean Cluster(1)

Contains all 3 institutes with
lesser number of
Employment Resources
Institutes than Cluster 0

High Educational Support Cluster(2)

Contains all 3 institutes
with high number of
Educational Support
Institutes

Single Feature Cluster(3)

Contains only Newcomer
Settlement Services
Institutes

Cluster Labels	Count	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

Conclusion



High Employment Resources Cluster

New immigrants looking for employment resources institutes can choose the neighborhoods in this cluster



High Educational Support Cluster

New Immigrants looking for educational support can choose from this cluster

Thanks!