Clustering Neighbourhoods to make Business Decisions

Clustering of Toronto Neighbourhoods
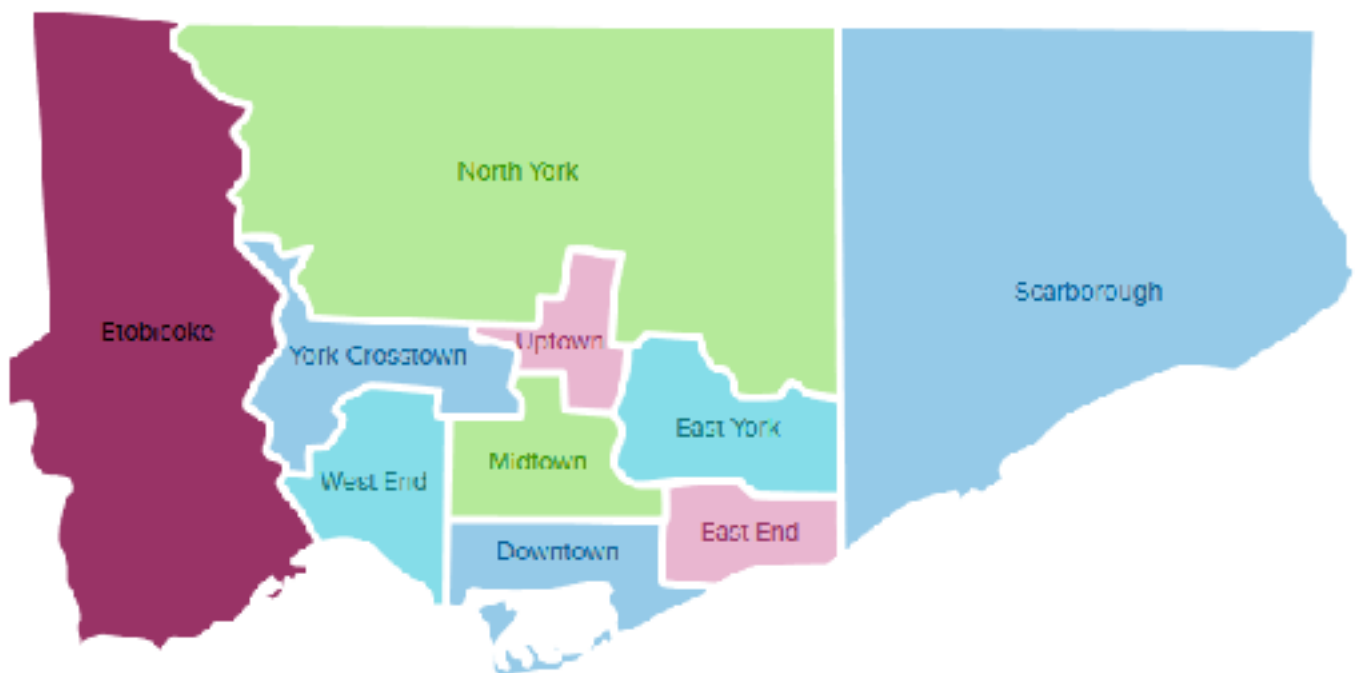
Shardul Kavale - June 11, 2020

Introduction

Different neighbourhoods and different cities are a house to unique venues that define the culture of those places. Some cities and neighbourhoods might be similar in terms of the kind of venues they house. What if we could group these neighbourhoods based on the venues? How would it help investors/ Businesses to make decisions?

Problem

Clustering and Grouping similar neighbourhoods, which will provide insights into what kind of venues they have, thereby aiding potential business owners determine the most feasible location to open their business in.

Eg: *North York has fewer cafes, thus opening a cafe here would prove to be a wise investment due to minimal competition*

Data sources and cleaning

The project works with the data of Torontos's different Boroughs and their postal codes. This table can be found on wikipedia.

Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This data table was web scraped using pandas library. After dropping not assigned values a data frame like below was obtained:

	PostalCode	Borough	Neighborhood
0	M5G	Downtown Toronto	Central Bay Street
1	M2H	North York	Hillcrest Village
2	M4B	East York	Parkview Hill, Woodbine Gardens
3	M1J	Scarborough	Scarborough Village
4	M4G	East York	Leaside
5	M4M	East Toronto	Studio District
6	M1R	Scarborough	Wexford, Maryvale
7	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
8	M9L	North York	Humber Summit
9	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...
10	M1B	Scarborough	Malvern, Rouge
11	M5A	Downtown Toronto	Regent Park, Harbourfront