BATTLE OF NEIGHBORHOOD

Toronto

INTRODUCTION

Background
Toronto is the capital city of Canada and is the most populous city with more than 2.7 million people in 2016. Toronto is divided into 6 major districts and more than 100 neighbours with their own characteristics.

Problem
Data around Toronto neighbourhood is widely available from multiple sources. The aim of this study is to find the best neighbourhood to open a restaurant in Toronto with consideration of the geolocation data and demographics data of each neighbourhood.

Target Audience

- New immigrants to Toronto
- People who plan to move to Toronto

DATA SOURCES AND DATA CLEANING

Data Sources

The first data source is neighbourhood list in Toronto scrapped from Wikipedia using beautiful soup package.

The second one is geolocational data from 'foursquare.com' using API calls.

The last one is neighbourhood demographic data is downloaded from 'open.toronto.ca'.

Data Cleaning

(a) Download data from Wikipedia as first data source

(b) Request for geolocational data as 2nd data source

(c) Download demographic data from Census of Population

DATA VISUALIZATION AND CLUSTERING

With a pandas data frame of neighbourhood with coordinates information, we shall be able to visualise the data using folium package. With K-means method, we are able to cluster neighbourhood based on their nearby places. A transition can be found as below-
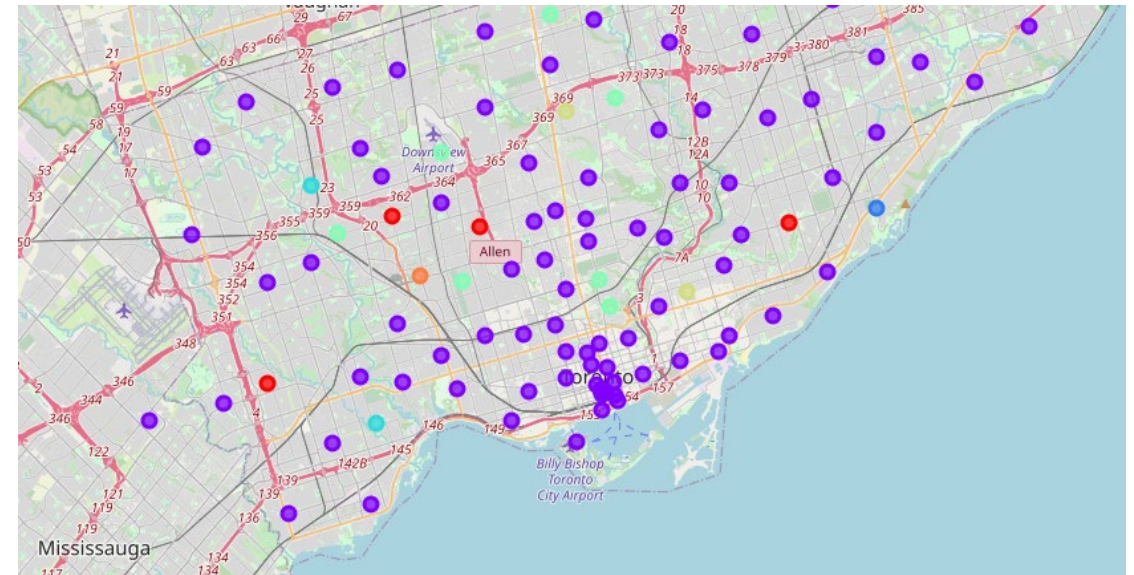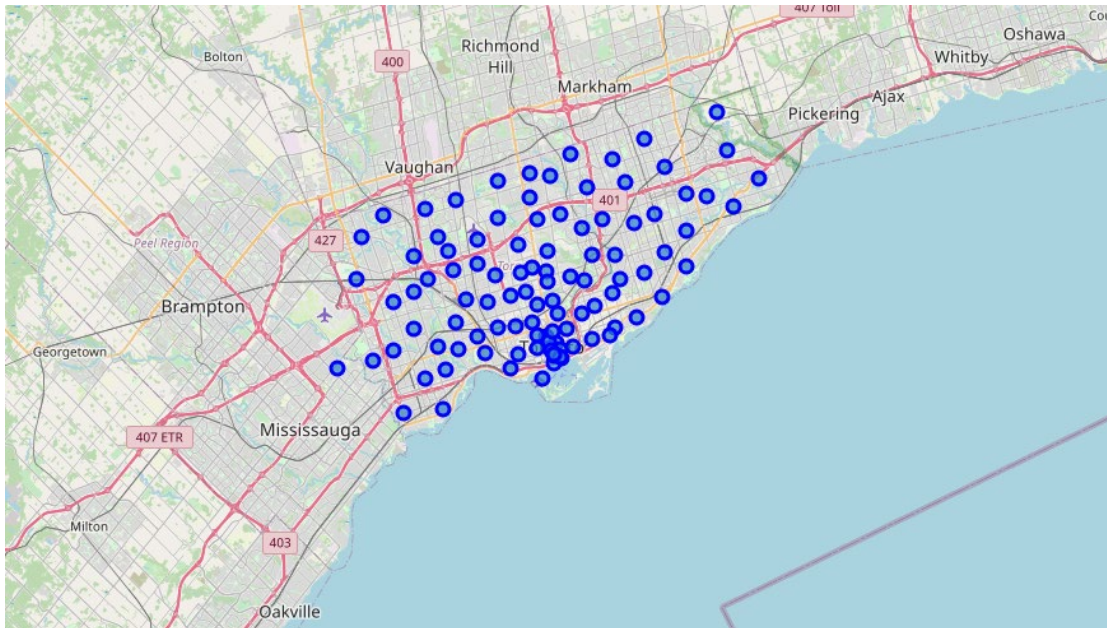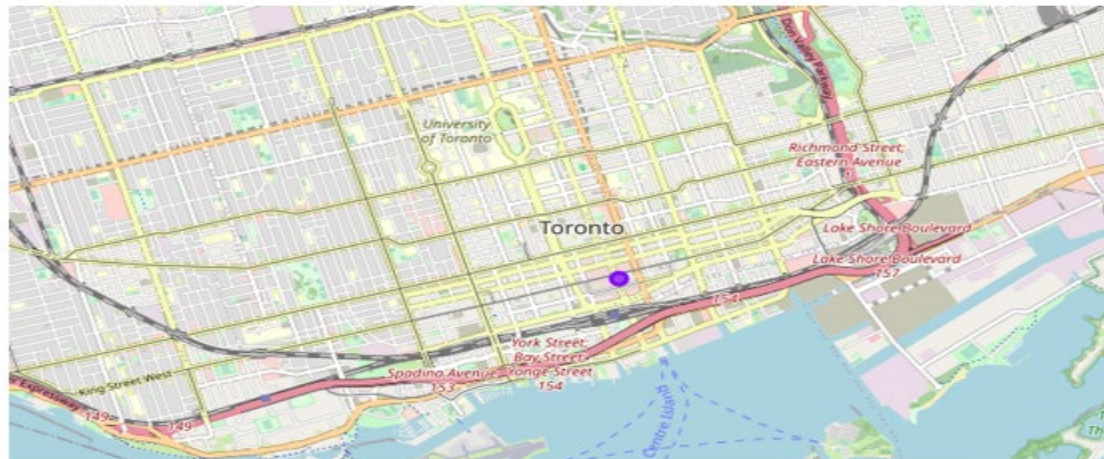
RESULTS

With clustering information as below, we will be able to tell the places with 'restaurants' in the top 3 popular venues. As we have 3rd data source with populate density, we are able to filter further and select the neighbourhood that linked to 'restaurant' keyword with highest population density.

Below is the result of the analysis. The best place selected is Commerce Court with population density of 3700.

	Neighborhood	density_int
48	Commerce Court, Victoria Hotel	3710

It can also be visualised as below –

CONCLUSIONS AND FUTURE DIRECTIONS

In this study, I gathered publicly available information from website and construct a mechanism to cluster neighbourhood information based on popular nearby places. The result set are further filtered based on demographic information of neighbourhood. The final recommendation is to open a restaurant in CBD area of Toronto, specifically area around Commerce Court, Victoria Hotel.

The place turned out to be the CBD area of Toronto with higher population density which is not uncommon. This analysis can be further enhanced using more demographic data points. This will also give better clustering power of K-means to produce a more accurate results.