IBM DATA SCIENCE CAPSTONE PROJECT

Find the Best Location for Living and Opening a Vietnamese Restaurant in City of Toronto

Hai Nguyen
nhhaidee@gmail.com

1. Introduction

1. 1 Problem Description

For this capstone project, assuming that immigrant family is planning to migrate to Canada and choose Toronto as their destination. One of the biggest concerns for immigrant families is to choose area which has very crime rate but large population to open a business which is a Vietnamese Restaurant.

1. 2 Project Requirement

Propose the best area for the immigrant family with following requirements:

- Area must have low crime rate.
- Area must not have the common venues are Asian Restaurants so low number of competitors.
- Population must be greater than 25000.

2. Data

For analysis, we use the following data sources :

- Toronto Crime by Neighbourhood: this source of data is used to find neighborhoods which have low crime rate.

  https://opendata.arcgis.com/datasets/af500b5abb7240399853b35a2362d0c0_0.geojson

- Toronto Neighbourhoods Geo: this source of data is used to plot Choropleth diagram to crime rate distribution among Toronto Neighborhoods. The data contains geographical coordinate (latitude, longitude) of 140 neighborhoods of Toronto.

  https://open.toronto.ca/dataset/neighbourhoods/

- Foursquare Developers Access to venue data: this source of data is used to find venues for each Toronto Neighborhood which is then used for clustering method.

  https://foursquare.com/

3. Data Analysis Tools

Python libraries such as: Pandas, Scikitlearn, Numpy, matplotlib, Folium, Foursquare API will be used for data analysis

4. Methodology

The analysis will include three main steps:

Step1: Find neighborhoods which have low crime rate based on Toronto Crime by Neighbourhood. This step will answer the first requirement of project. With the Choropleth map, we can have big picture about crime rate distribution among neighborhoods. This is important when making decision to choose location for living.

Step 2: After having neighborhoods with low crime rate, then will explore venues of these neighborhood by using Foursquare API. We will have information about the most common venues for each neighborhood. This is important when making decision to choose location for Vietnamese restaurant

Step 3: We will use clustering method called K-Means Cluster to cluster neighborhood. The step 2 and 3 will answer the second and third requirement of project.

5. Result and Discussion

Below is the distribution map of crime rate among Neighborhood of Toronto (**Figure 4.1**) in which 60 Neighborhoods have a low crime rate (**Figure 4.2**)

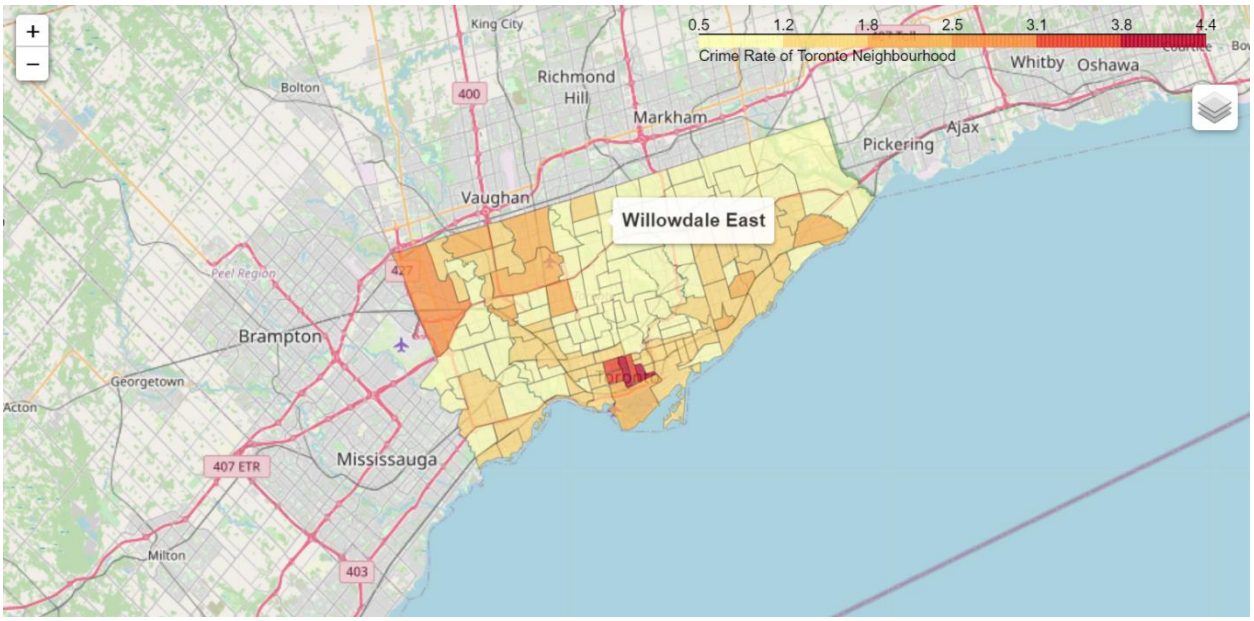

Figure 4.1 Distribution map of crime rate among Neighborhood of Toronto

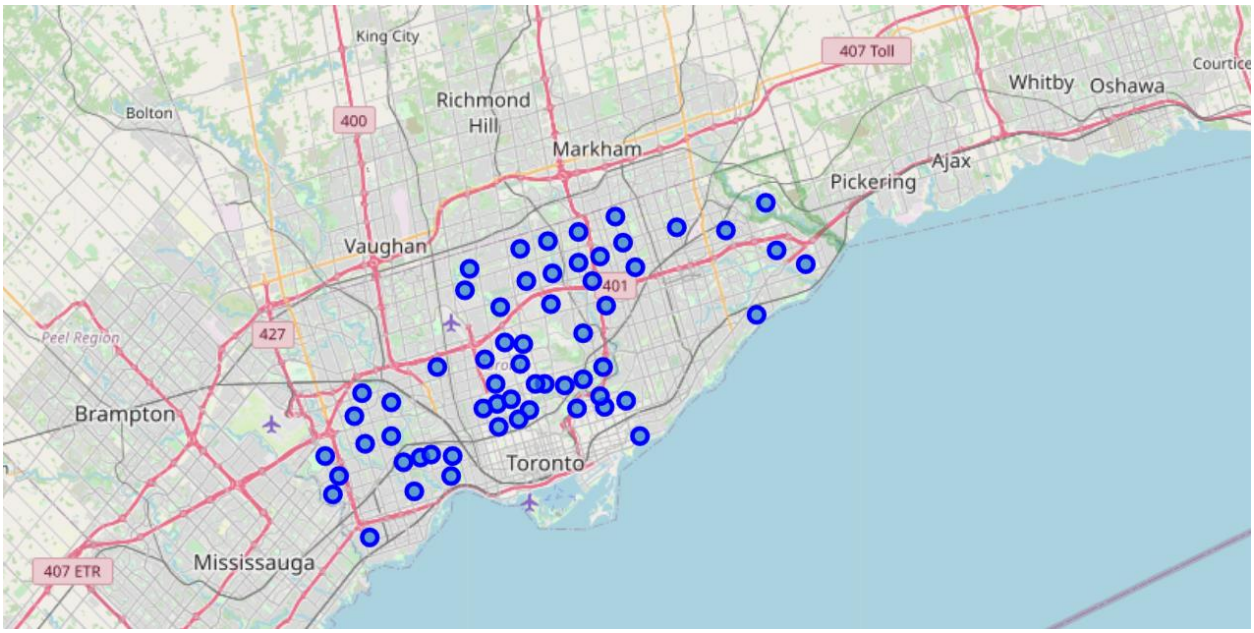

Figure 4.2 60 Neighborhoods have a low crime rate

K-Means Clustering Neighborhood without filtering Population as Figure 4.3

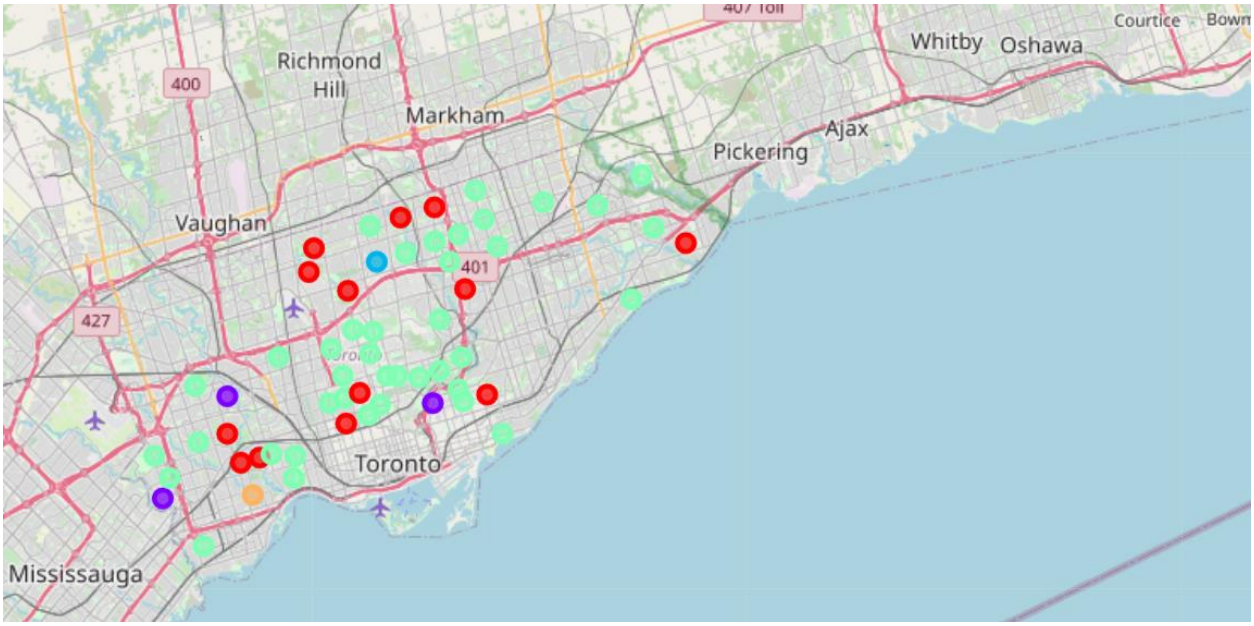

Figure 4.3 K-Means Clustering Neighborhood

After filtering cluster in which we choose neighborhoods have large population (> 25000).

1. Cluster 0 (Red): has 2 neighborhoods which are Westminster-Branson and Parkwoods-Donald. These neighborhoods have good population density and most common venues in these areas are Western Restaurant styles (Fast Food, French restaurants).

2. Cluster 1 (Purple): there is on neighborhood found.

3. Cluster 2 (Blue): 1 neighborhood found which is Willowdale East, this neighborhood has large population (50000). Hotel and other entertainment facilities are the most common venues. This is the best neighborhood to open Vietnamese restaurant with very low competitor. Furthermore, it has homicide rate is 0 so immigrant likely to choose area for living and open a Vietnamese restaurant

4. Cluster 3 (Green): 8 neighborhoods found in which Asian Restaurant is the most common venues (Chinese Restaurant, Vietnamese Restaurant). We should not choose this cluster because of many competitors

5. Cluster 4 (Orange): 1 neighborhood found which is Stonegate-Queensway. This is also a good location to open restaurant, but it has low population (25051)

Based on the observation, the Willowdale East area of Cluster 2 is the best location.

6. Conclusion

There are 140 neighborhoods in the city of Toronto and based on the analytic methods, we can find the best location the meet the requirement for the immigrant family. This method discussed in this report can be applied for any city in the world.