

# The Battle of Neighborhoods

## Report

### 1. Introduction

A Chinese couple have earned a lot of money during their lives in China, they want to go to Toronto relax a few years. In these years, they also want to open a new mall so that they will not feel so boring if nothing to do. They ask me if I have a good idea where to open the mall it the best.

The business problem is obviously: In the city of Toronto, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

We Know there are many ways to help people choose the location that they want to open the mall, but I don't have that so much time and money to focus on many things. So just choosing one factors as a key to solve the problem is the better way. This factor is the clusters of the Toronto with shopping mall.

### 2. Data

This project is mainly about investing and open a shopping mall in the Toronto, which go first is to search useful data to fulfill the goal.

- ✓ The neighborhoods in Toronto. Finding and list the neighborhoods so we can divide it easily.
- ✓ Latitude and longitude coordinates of those Neighborhoods. Using latitude and longitude help us to visualize and plot the neighborhoods.
- ✓ Using Foursquare API to get Venue data, particularly data related to shopping malls. This means we can Perform cluster by python.

### 3. Methodology

Firstly, I use the borough data from wiki and clean、 arrange it as a list. Then adding latitude and longitude to this list. Through this way the postcode、 neighborhoods and location in the Toronto is clearly presented on the plot in python.

Secondly, using the Foursquare API and create the account. By the free package, we can get the venue of these neighborhood. Finding out the unique categories, like the "Chinese restaurant"、 "Discount house"、 "Video store" and so on.

Thirdly, defining the clusters of these neighborhoods, using the factor of shopping mall by K-means way.  $k$ -means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining.  $k$ -means clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

#### **4. Results and Conclusion**

The results that I find the cluster 1 is good place to open the shopping mall. Because there are very few shopping malls in these place, but there are so many consumers which means it is a good way to choose these place.