The battle of Neighborhoods

Title: Business Venue Recommender System in Toronto

Problem Background :

Toronto is the capital of the province of Ontario, Canada. It is a major Candian city along Lake Ontrario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers. It is also the fastest growing city in North America. Toronto is an international centre of business, finance, arts, and culture, and is recognized as of the most multicultural and cosmopoliton cities of the world.

Problem Description :

A businessman already has a restaurant being operated successfully in one neighborhood of Ontario. Suppose, he/she wants to increase their revenue by opening another branch of the restaurant in other part of the city. In such situation, the type of neighborhood plays an important in choosing an optimum location for the new branch. Factors like the kinds of venues in the neighborhood, population of the neighborhood, income of the neighborhood and so on have a significant effect on the location chosen for the new branch.

Our aim would be to find a location similar to the location of the original branch to minimize the risks.

Data Requirements :

For a Recommender system, we need data and lots of data. Data can answer question which are unimaginable and non answerable by humans because humans do not have the tendency to analyse such large dataset and produce analytics to find solutions.

1. We will need information about all the neighborhoods and the boroughs of the city of Toronto. We would also need each neighborhood's latitude and longitude information. We would also need other information like income, and population of each neighborhood. I found the neighborhood, income, and population information from here: https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv
   I mapped each neighborhood to its postal code and borough from the Wikipedia page of Toronto City.
   I found the latitude and longitude from here mapped to each postal code: http://cocl.us/Geospatial_data
2. We would also need information of venues, their longitude and location of each venue. This is where FourSquare API comes into play. Use of foursquare is focused to fetch nearest venue

locations so that we can use them to form a cluster. Foursquare api leverages the power of finding nearest venues in a radius(in my case : 500mts) and also corresponding coordinates,venue location and names.