Finding Optimum Location for opening an Indian Grocery Store in Toronto (using Machine Learning)

Introduction (Background)

For this Capstone project, I'm building a solution for a friend who is an entrepreneur based in Canada. He is willing to explore and open an Indian grocery store in Toronto and expand to Asian Groceries chain. There is healthy population of Asians in Canada and there might be a great demand for Indian cuisines and ingredients, thanks to all the Asian and Indian restaurants. However, there will be heavy competition in the market and there is an absolute need to find the best place and resources to succeed.

Problem Statement –

The objective of this capstone project is to find the most suitable location for my friend to open a new Indian grocery store and later expand to Asian grocery chain in Toronto, Canada.

We need to identify the high demand and low demand neighborhoods as well as population density, income per capita, crime rate and near-by venues/attractions. We will be using data science methods and machine learning methods such as clustering to provide solutions to answer the following business question:

What is the best place to open an Indian grocery store in Toronto?

Data Requirements –

We need to have access to location and geographical data to predict the optimum location for opening an Indian grocery store. The following are some of the data points necessary to perform this analysis and predictions.

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venue data related to Asian restaurants.

The Foursquare API location data is a great place to start the analysis to identify the clusters of neighborhood and optimum location for opening the grocery store.

Data Collection –

The following data needs to be collected to perform the necessary analysis – We will make use of Wikipedia information for scraping Toronto neighborhood related data. We'd also use Geocoordinates (Longitude and Latitude) via Geocoder package

Finally, we'd extract Venue and Nearby location information using Foursquare API.