Coursera Capstone Project

Zhonghao Sun

Introduction

Toronto is the capital of Canada. An international supermarket chain company is planning to open its first supermarket in Toronto. The aim of this project is to find a proper neighborhood location for its first store. A number of factors need to be considered to determine a proper location, these include demand, competition, and reputation for future development. The company may have different development strategies. For example, a less competitive area may be not good for the company to improve its reputation for future development, while it may be difficult to be profit in a more competitive area. The task for a data scientist is to collect and analyze Toronto neighborhoods' data and provide statistical information for the board to make decision. Specifically, the following need to be explored:

1) Compare population (demand), number of supermarket and grocery stores (competition), number of other venues such as restaurants, bars, hotels, schools that can increase demand,
2) Cluster neighborhoods and find the current development patterns,
3) Identify proper neighborhoods based on different development strategies of the company.

The target audience is the marketing and development department of the supermarket company.

Data

To solve the problem, we need to know neighborhood name, geographical information, population, and venue data.

1) Neighborhood name and population. The name and population of Toronto neighborhoods can be find in Wikipedia: https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods. It contains a table list. We can use web scraping techniques to extract the data with pandas and beautifulsoup packages. The data is stored in a dataframe and can be further processed.
2) Geographical information of neighborhoods. The geographical information is required to request venue data. The latitude and longitude coordinates of neighborhoods can be

obtained using Geocoder package or the csv file from the class. This will be a table list of neighborhoods name, and their latitude and longitude coordinates.

3) Venue data. We use Foursquare API to get venue data for neighborhoods. Foursquare.com is one of the largest databases of venues. We can obtain a table list of neighborhoods and nearby venues, including venue name and category by using the Foursquare API.

Methodology

Results

Discussion

Conclusion