

Selection of a city based on Individual's interest for relocation purpose

Scope

This document is a detailed report on a business case statement. It mentions the different aspects of a business case.

Target Audience

For this problem statement, the prospective target audience are individuals who based on their preferences are relocating to a new city, a city where they are unaware of which location would be best suited according to their choices and current lifestyle. Though this case study deals with the city of Toronto, it can be used for any city based on the postcodes and foursquare data.

Problem Statement

Ravi Iyer, an Indian is on his first ever trip overseas to the city of Toronto for a year. He belongs to an orthodox family and his family had been skeptical if he will be able to keep up with his cultural and food habits in Toronto.

Ravi has been exploring different neighbourhoods of Toronto. His preferences are

- He would not like to spend too much on commuting. Hence he would prefer a location with local transport.
- He is a vegetarian and he would prefer Asian food as much as possible.
- He is an avid reader and fitness conscious and would prefer having a book store and a fitness center close by
- Keen on not spending too much either on food, he would prefer to cook on his own and would hence need a grocery store near his accommodation
- For recreation and relaxation he prefers living close to parks and movie theatres.

With such specific needs, Ravi would like to explore neighbourhoods of Toronto and the possibility of living a lifestyle he expects.

Data Section

The data used for this analysis will be from FourSquare.com and Wikipedia. For all the zip codes available on Wikipedia for Canada, since our problem statement deals with the city of Toronto, we extract data for Toronto, clean the data and analyze it.

After cleaning of data for all neighbourhoods of Toronto, use foursquare.com to fetch venue information. Based on the data extracted and analyzed come up with the most favourable neighbourhood. We also apply statistical methods to cluster the data based on neighbourhoods.

To explain the data processing in detail, we first extract all zip code information from Wikipedia from the link:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The city of Toronto is mainly divided into four major areas viz. East, West, Central and DownTown Toronto, we get a list of venues from foursquare.com, the venues which meet our problem statement criteria. Based on the grouped data figures and the availability of services in each area, we decide on DownTown Toronto as the most favourable place. The selection is based purely on the counts of desired venues in each area.

Once the selection of the area is done, we then go on to get more information about the venues in DownTown Toronto. The venues are categorized, and are ranked based on their popularity. Using statistical method of kmeans clustering we find out the most optimum number of clusters that can be applied and then perform k means clustering on the data based on the neighbourhoods in DownTown Toronto.