

1.0 Introduction

1.1 Business Problem & Analysis Objective

In this project, we will hypothesize that an online shopping fashion company – Boozt, decides to expand their business to the city of Toronto. To ensure that every dollar is strategically spent, this fashion company must selectively choose the most optimal regions for advertising to grab the potential customers.

Problem Statement: This project seeks to determine the most strategic region in Toronto for a fashion company to spread of advertising.

1.2 Data Requirements

In this project, we will take into account of the supply of advertising location and demand from customers to assess the potential of each region. Furthermore, bus stop has decide as main advertising location, the number of nearby bus stop within a 3km radius to a neighborhood is used to determine the advertising density. On the other hand, the fashion demand from a customer is assumed to be driven by purchasing power of each neighborhood's population. The purchasing power is both determined by population size and average household income.

Data Sources:

1. Nearby bus stop- [Foursquare](#)
2. Location data - [Wikipedia](#)
3. Demographic/ socioeconomic data (population size by neighbourhood, average after tax household income by neighbourhood) - [Wellbeing Toronto](#)

1.3 Methodology

1. Various data sets were imported and subsequently combined into a single data frame.
2. Foursquare API is used to determine the number of nearby bus stop within a 3km radius to a neighborhood, this is returned as bus stop count and appended to the combined dataframe.
3. To determine the most strategic location, neighbourhoods will be segregated into different regions via k-means clustering.
4. The statistical data on each region (or cluster) is then generated and explored.