Battle of the Neighborhoods
Week 1
By Zeyar Oo

1.1 Business Problem

For investors, it is very helpful to know what is popular in which neighborhood and demographics such as age, education levels, marital status, income, ethnic background, and other factors that explain and influence consumer behavior. There are some restaurants, for example, that open in a neighborhood with considerable sunk costs and only after a few months they close down because they fail to enough customers. If they understood their customers better in the first place, such failures could be avoided.

1.2 Background

With a population of close to 3 million, Toronto is an important financial center for Canada. Being a cosmopolitan city with diverse demographics, it is only natural that it offers everything you need in terms of restaurants, parks, spas, pubs, gyms, business services and so on. Like in every major city, certain venues such as corporate offices can be concentrated in one area such as in a financial district. Households with higher income and education levels can also be concentrated in certain neighborhoods. Using data on neighborhoods, I will attempt to explain why certain venues are more prevalent in a particular neighborhood. I will also make recommendations for investment opportunities in case there are places with little competition but potential for attracting significant number of customers.

1.3 Data

The data used in this research comes from Foursquare and Toronto open data portal (https://open.toronto.ca/catalogue/). Foursquare API gives information on each neighborhood with venues and their location coordinates and category. The 2016 census for Toronto contains aforementioned important demographics for each neighborhood. By combining the two datasets, I plan to come up with recommendations for type of restaurants that are suitable for a particular neighborhood.

With the use of clustering algorithms, it is possible to get an overview of the similarities or differences between the neighborhoods. Similar neighborhoods can be further analyzed to find out the concentration of certain venues, income levels, education, age, marital status, ethno-linguistic backgrounds and other demographic factors.