Data Description

Since we will need to explore, segment, and cluster the neighborhoods in the city of Toronto, the Toronto neighborhood data is key for this project.

Unfortunately, the data for the Toronto neighborhoods is not widely available on the internet in a structured format. Therefore, we will need to scrape it through an existing Wikipedia page that has all the information we need to explore and cluster the neighborhoods. The data should contain the coordinates for each of the neighborhoods in Toronto that will help us to obtain more crucial information for the project.

We will then need to leverage on features in a reliable location information provider such as Foursquare to explore the various types of venues and categories available in each neighborhood. We will also need to understand the trends for these venues in the respective neighborhoods.

We will also use clustering techniques such as the K-Clustering to segment and cluster these neighborhoods so that we can group them together to understand their similarities and what best we can do for these types of neighborhoods.

With all these features, and data, we will then be able to come up with the best recommendations AnyWords shareholders for them to settle in a new location.

The success of this project will be based on two key factors:

1. The lack of bookshops available, meaning a lower competition rate.
2. A higher number of residences for a higher demand.