
2. DATA

2.1 Sources

We will be using data from a variety of sources to get the information required;

- Wikipedia(https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary) : Using the data we will be able to get the neighbourhoods, population and population density.
- Foursquare API: Using the information from foursquare we will get the venues, restaurants for the neighbourhoods.
- Income Data: For all intensive purposes we will use the data from Great News (<https://great-news.ca/demographics/>) and assume the information is representative of median household income in Calgary.

2.2 Cleaning

Data will need to be scraped from the sources mentions in section 2.1 and combined. Missing data or incomplete data was present for some of the income figures related to neighbourhoods. We will make a few assumptions;

- Only neighbourhoods with assigned data will be processed.
 - Neighbourhoods missing data will be dropped.
 - Income values from given year will be used to characterize neighbourhood spending.
-