Capstone Project

The Battle of Neighbourhoods

Problem statement and background information

Toronto is a bustling city full of opportunity for new businesses. While there is opportunity to be had, this has attracted many to start a business already and to unlock opportunity a new business will need to compete with the existing market.

Joe, a 35 old from Toronto has saved up working in a law firm and now would like to start his own business. It is crucial to understand where to locate his business to ensure he is not marginalised by the competition. Joe has a wide array of skills so happy to consider a range of different industries to enter.

The key questions Joe needs to answer are:

1. Which industry should he consider?
2. Where should Joe locate his first venue?
3. Are there any further expansion opportunities for the future?

Data sources and how these will solve the problem

Data set	Source	Key features	How this will be used
Toronto post code	https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M	Postcode district and area information	To segment Toronto geographically into areas (postcodes) and return basic information of the area (postcode)
Geo-coordinates	Geocoder API	Location data	To enhance Toronto postcode data with geo co-ordinates
Foursquare venue API	https://api.foursquare.com/v2/venues/explore?<API call>	Venue types and locations	Identify existing business in Toronto, their location and type to help answer Joe's questions

Methodology

Data methodology

- Import postcode data from Wikipedia and enhance with geo-coordinates from geocoder
- Use the Foursquare API to gather the venue data for all postcodes
- Group the data into postcode level dataframe and another at venue level for the analysis

Analysis key steps

- Analyse which type of business is the most popular
- Understand which districts have high / low penetration of the industry selected and how the venue types correlate with each other
- Cluster to understand which is best location to start
- Use K-means algorithm to understand the distinct groups of neighbourhoods.