Coursera Capstone

IBM Applied Data Science Capstone

Finding Best Neighbourhoods for a New Home in Toronto, Canada

By: Suhail Ur Rahman

February 2020

Data

To solve the problem, we need to find the following data:

- List of Neighbourhoods and their properties.
- Latitude and Longitude of the neighbourhoods.
- Venue data of the neighbourhoods.
- Average Prices of homes in each neighbourhood

Sources of data and methods to extract the data.
To find the list of neighbourhoods and their crime rates in order to determine the safety of the neighbourhood, I accessed the website of Toronto Police. They have open data access to various crimes' ratings from 2014 to 2018. They had a geojson file(https://opendata.arcgis.com/datasets/af500b5abb7240399853b35a2362d0c0_0.geojson) which helped in determining the crime according to the neighbourhoods. Also, they included Latitude and Longitude coordinates for the boundaries of each neighbourhood.

They were 140 neighbourhoods in the geojson file. Since the coordinates were of boundaries, the centre coordinates of each neighbourhood have to be calculated according to the average of all latitudes and longitudes. This would be needed in order to plot the neighbourhood clusters.

To find the venue data, Foursquare API was used. It would show the most popular venues in each neighbourhood of Toronto while using the central coordinates of the respective neighbourhood.

After Segmenting the data, it would be combined with the average home prices of each neighbourhood. The data will be obtained from the blog post from the following link https://www.moneysense.ca/spend/real-estate/where-to-buy-2019-toronto/.
There are some neighbourhoods which are missing from the dataset, and the missing prices will be replaced with the mean price of homes across all neighbourhoods.