Exploring Ottawa Neighbourhood Amenities

Md M. Hasan

Introduction

Aim and Scope

The City of Ottawa is the capital of Canada. Its population is just over a million. According to news media reports, Ottawa has Canada's one of the hottest real estate market. Selling and buying homes in a neighbourhood are greatly influenced by the nearby amenities. This project will explore and summarize Ottawa's neighbourhoods based on their nearby amenities. It aims to rank the neighbourhoods by groups in terms of amenities.

Business Problem

This project has targeted stakeholders from two different business areas. The primary stakeholders are from the real estate industry and homebuyers. The neighbourhoods with the most amenities are more attractive to homebuyers. Thus, real estate companies can make more profit from knowing about amenities. This project is to produce unbiased results to prevent exaggeration activities of some real estate companies. Thus, homebuyers can also be benefited from the results. On the other hand, the secondary stakeholders are from development authorities and entrepreneurs. Such stakeholders need to identify the neighbourhoods with the least amenities. Such neighbourhoods are the best locations for establishing new amenities. Thus, Ottawa's living standards can be improved by creating new business opportunities.

Project Goal

This project is going to apply data science methodologies and tools to solve a neighbourhood relevant problem. Its goal is to deliver a clear summary of Ottawa's neighbourhoods' amenities that are very useful in citizens' daily life.

Data

Sources

This project requires data sets of Ottawa's neighbourhoods and amenities. The sources of the required data are as follows.

• Neighbourhood data: Initially, the postal codes of Ottawa neighbourhoods are collected from `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_K`. The latitude and longitude for each postal code are obtained from `https://www.geonames.org`.

• Amenities data: Once the geographic coordinates are available, the Foursquare API (`https://developer.foursquare.com`) is used to collect neighbourhood-specific data of amenities.

Cleaning/Filtering

The source formats are refined to construct data frames.

• Acquisition of Postal Codes and Coordinates: It mainly requires text pattern matching and removal of redundant elements from the source structures. Tables are read from *wikipedia* and *geonames* using the "read_html" function.

• Acquisition of Amenities Data from the Foursqure API: The main challenge of using the Foursquare API data is that the most of its urban venues are related to foods or restaurants. Information regarding those venues are very useful to travellers and foodlovers. However, this project is primarily focusing on amenities that attract homebuyers. Thus, restaurants are not the key points of interests. When a Foursquare search result is overflooded with food related venues, it becomes difficult to extract other types of venues. In some cases, many other types of venue may not appear within a specified limit of results. To overcome this problem, this project is to use the *search* endpoint based on individual "*categoryId*" to filter the amenities of interests for each neighbourhood. The search parameter of *intent* will be set to "browse". Here is the list of selected Foursquare categories that corresponds to neighbourhood amenities:

- School
- Library
- Park

- Medical Center
- Pharmacy
- Bus Stop
- Grocery Store
- Community Center
- Shopping Mall
- Market
- Shopping Plaza
- Coffee Shop
- Child Care Service
- Daycare
- Bank
- Post Office
- Outdoors & Recreation
- Convenience Store
- Nursery School
- Preschool
- Fast Food Restaurant
- Fire Station

Feature Extraction/Preparation

To extract the necessary features from Foursquare's search results, the following case-by-case mechanisms will be applied.

• In some cases, features can be directly extracted from the results. For example, *bank* and *fire station* features on a neighbourhood can be directly extracted.

• In some cases, multiple features can be extracted from the same category. For example, *medical center* includes hospitals, pet/animal hospitals, dentist chambers, and many more venues. Another good example of such a category is *school*.

• In some cases, overlapping exists between multiple categories. Only one feature will be extracted from those categories. For example, *shopping mall*, *market*, and *shopping plaza* in the

same neighbourhood can be considered to be a single feature. The same can be done for *nursery school* and *preschool*.

• In many cases, unwanted features will be dropped. For example, the *medical center* category also includes the venues like "alternative healing" facilities. Such features will be dropped from the neighbourhood amenities.

• The venues belong to ethnic restaurants are also dropped since they are very diverse and too difficult to generalized as a single amenity. Fast food restaurants and coffee houses are kept as amenities in a neighbourhood.

After constructing a data frame of the selected features, the k-means clustering approach will be applied to summarize the neighbourhood amenities.