# 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.

# Methodology

Once the geocoordinate of each neighbourhood is available, the Foursquare API calls are started. The search radius is set to 1 km. The maximum number of venues is set to 100. Categories are extracted from Foursquare's response on venues. Only the aforementioned amenity related information are transferred to a data frame for further processing. There are two major steps in the data analysis: ranking and clustering. Their workflows are as follows.

## Amenity Ranking of Neighbourhoods

It is a preparatory step to the clustering one. An initial visualization using Folium is done to check the geographic distribution of Ottawa's neighbourhoods. The catergories of interests are reduced to 26 final features. It is done by combining similar features and dropping unintended ones. This can be considered as an example of a *dimension reduction*, which is also a machine learning technique. The *dimension reduction* is performed on data set before applying machine learning algorithms. It is a way of avoiding the *curse of dimensionality*. The final data frame consists of the frequency of each feature for all neighbourhoods.

It is observed that the range of data is different for different features. This is because some amenities are much more frequent than others. This implies the necessity of applying a data normalization technique. The *min-max* scaling is applied to the amenity data set. The *min-max*

scaling is the simplest normalization technique. It is also a widely used pre-processing method that brings all data in the range of [0, 1].

To build a ranking of neighbourhoods, it is necessary to define or adopt a metric. A metric named 'Amenity Score' is defined to obtain the ranking. It is defined as the sum of all normalized frequencies of features in a neighbourhood.

## Amenity Clustering of Neighbourhoods

The final step is to cluster Ottawa's neighbourhoods based on their amenities. The k-means clustering method is selected. It is considered to be an efficient machine learning technique. It has a wide range of applications including vector quantization, cluster analysis, and feature learning. The current project applies the k-means for cluster analysis to identify similarities among neighbourhood amenities.

In the k-means clustering, the number of clusters 'k' is an input parameter. The optimal value of 'k' is detemined by the Elbow method. Figure 1 shows the plot obtained using the final data set in this project. As there is no sharp elbow point, the best value of 'k' is not very clear from the plot. It could be 5, 6, 7, 8, or 9. By considering the number of neighbourhoods in the data set, k = 5 is set. The number of clusters should not approach the number of neighbourhoods.

A final visualization of clustered neighbourhoods is performed to determine their geographic proximity. Finally, average amenity score of each cluster is calculated.

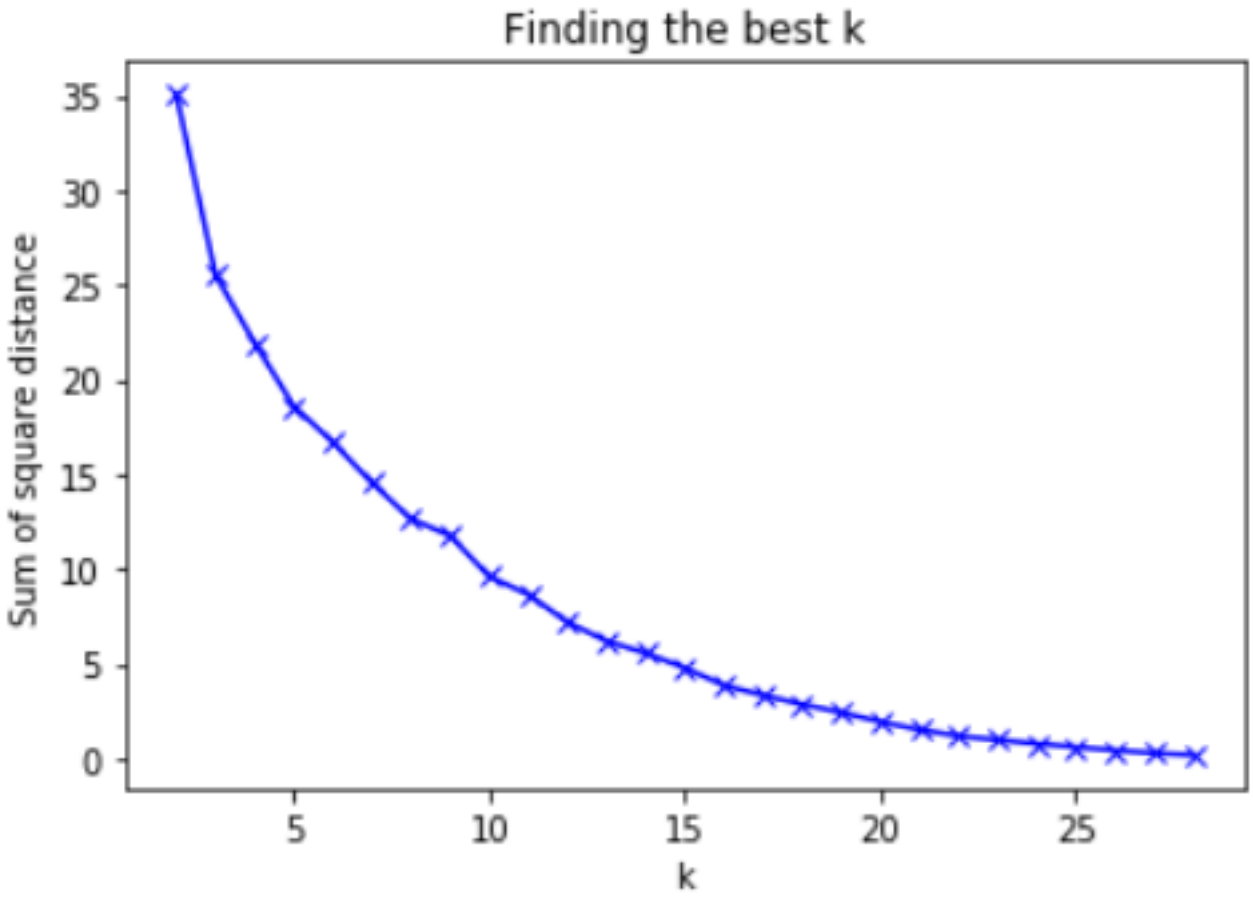


Figure 1: The Elbow method to find 'k' for the final data set.

# Results

Figure 2 shows the initial visualization of Ottawa's neighbourhoods. The list of Ottawa's top 20 neighbourhoods is provided by Table 1. It also shows amenity scores. Figure 3 depicts the clustered neighbourhoods with different marker colors. Table 2 summarizes the clusters. Finally, Table 3 provides the overall summary of Ottawa's neighbourhoods based on amenities.

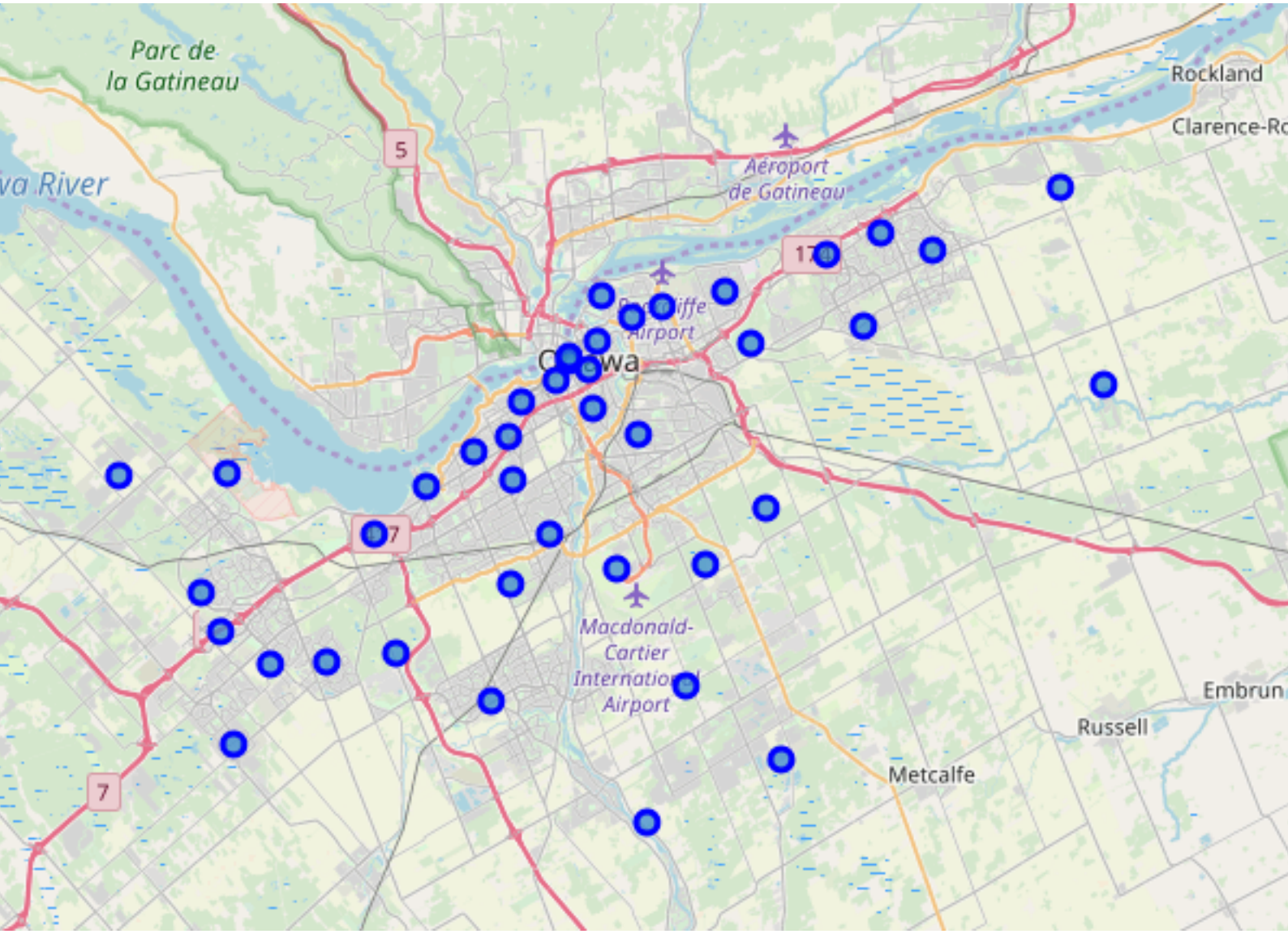


Figure 2: A visualization of Ottawa's neighbourhoods.

| Rank# | Neighbourhood | Amenity Score |
|---|---|---|
| 1 | Centretown | 16.658950 |
| 2 | Government of C. office ... | 15.947554 |
| 3 | Lower Town / Byward Market ... | 14.791857 |
| 4 | Downtown | 14.346140 |
| 5 | Dalhousie Ward | 10.434267 |
| 6 | Civic Hospital / Island Park ... | 9.044369 |
| 7 | The Glebe / Old Ottawa S. ... | 8.921706 |
| 8 | Westboro / Carlington | 7.982408 |
| 9 | Vanier, McKay Lake area | 6.494166 |
| 10 | Queenswood | 5.695661 |
| 11 | Britannia /Whitehaven ... | 5.055445 |
| 12 | Highland Park / McKellar Park ... | 5.045258 |
| 13 | Overbrook, Forbes, ... | 4.878796 |
| 14 | Orleans | 4.394452 |
| 15 | Barrhaven | 4.340723 |
| 16 | Queensway / Copeland Park ... | 3.932067 |
| 17 | Bridlewood | 3.714552 |
| 18 | Beacon Hill / Cyrville / Carson ... | 2.986010 |
| 19 | Alta Vista / Billings Bridge | 2.816997 |
| 20 | Stittsville | 2.656096 |

Table 1: Top 20 neighbourhoods in Ottawa

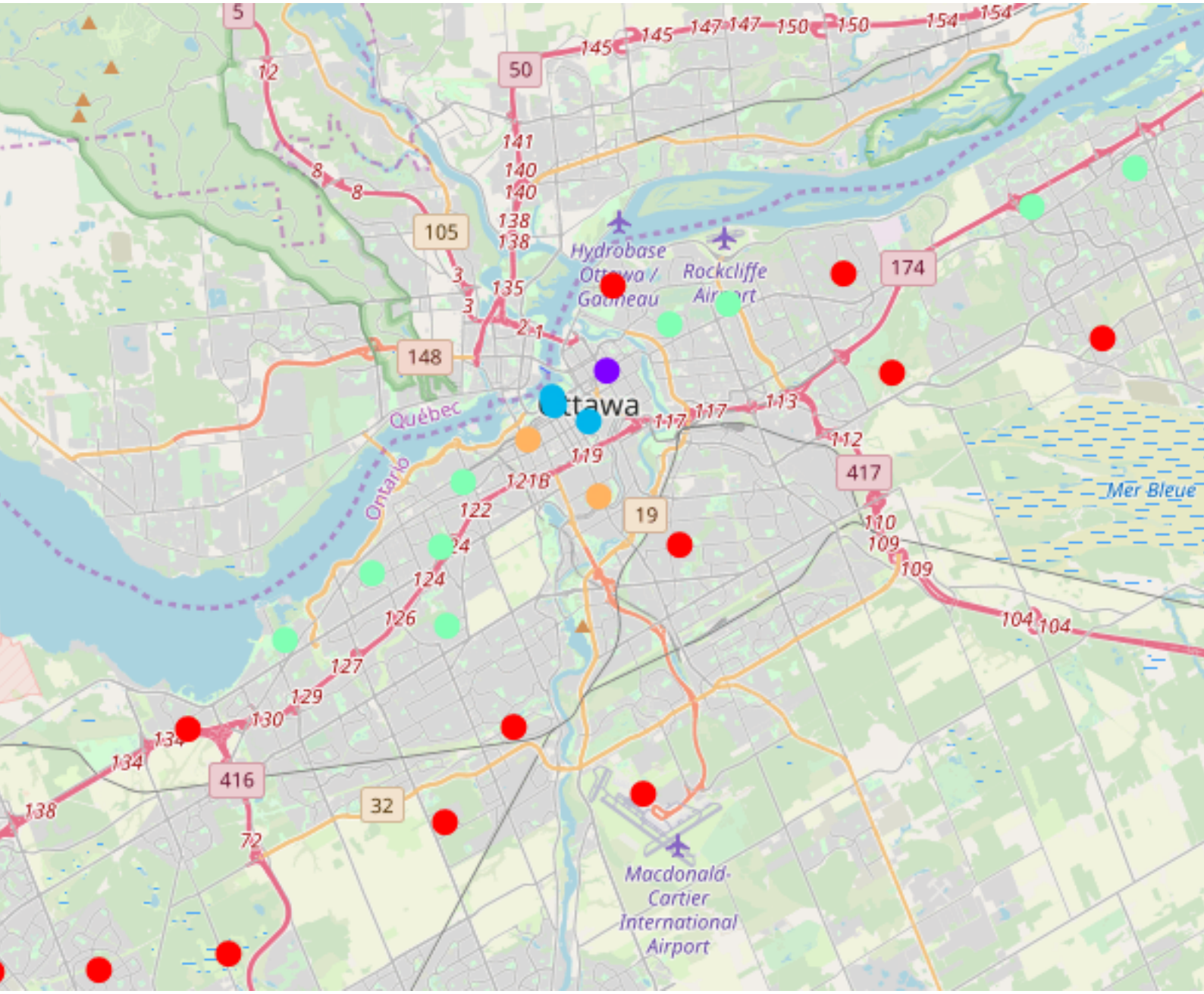


Figure 3: A visualization of Ottawa's clustered neighbourhoods.

| Cluster Label | Size | Average Amenity Score | Marker Color |
|---|---|---|---|
| 0 | 17 | 1.437103 | Red |
| 1 | 1 | 14.791857 | Navy Blue |
| 2 | 3 | 15.650881 | Pale Blue |
| 3 | 10 | 5.686335 | Green |
| 4 | 2 | 9.677986 | Orange |

Table 2: Clusters and amenity scores

| **Group Description** | **Neighbourhoods** |
|---|---|
| Cluster Label **0** | Blackburn Hamlet / Pine View / Sheffield Glen, Alta Vista / Billings Bridge, Beacon Hill / Cyrville / Carson Grove, Rockcliffe Park / New Edinburgh, Heron Gate / Heron Park / Riverside Park / Hun..., Chapel Hill South / Blackburn, South Gloucester, Eastern Nepean: Fisher Heights/ Parkwood Hills..., Centrepointe, Meadowlands, City View, Craig He..., Bells Corners / Arlington Woods/Redwood / Qual..., Katimavik-Hazeldean / Glen Cairn, Bridlewood, Fallowfield Village / Cedarhill Estates / Orch..., Stittsville, Marchwood, Terry Fox / Palladium, and Fallingbrook. |
| Cluster Label **1** | Lower Town / Byward Market / Sandy Hill /University ... |
| Cluster Label **2** | Government of Canada Ottawa and Gatineau office..., Downtown, and Centretown. |
| Cluster Label **3** | Orleans, Queenswood, Overbrook, Forbes, Manor Park, Viscount Alexan..., Vanier, McKay Lake area, Civic Hospital / Island Park / Hintonburg / Me... , Westboro / Carlington, Highland Park / McKellar Park /Westboro /Glaba..., Britannia /Whitehaven / Bayshore / Pinecrest, Queensway / Copeland Park / Central Park / Bel..., and Barrhaven. |
| Cluster Label **4** | Dalhousie Ward and The Glebe / Old Ottawa South / Old Ottawa East... |
| **No data** available at the time stamp of 12:10 AM - 01:00 AM UTC, Sept. 8th, 2019 | Riverview / Hawthorne / Canterbury / Hunt Club..., Blossom Park / Greenboro / Leitrim / Findlay ..., Beaverbrook / South March, North March, Navan, Cumberland, Manotick, and Greely |

Table 3: Summary of Ottawa Neighbourhoods

# Discussion

The results presented here are obtained based on Foursquare's response for the timestamp of 12:10 am - 01:00 am, Sept. 8th, 2019. In that time, amenity data regarding Ottawa's 8 neighbourhoods were unavailable. Moreover, Foursquare's response highly depends on the time instance. Therefore, results could vary if data were collected in another moment. This is the main drawback of this project.

Most of the Foursquare's venues are related to restaurants or foods. For this reason, category-specific search is preferred in this project. However, only few root categories are searched due to the daily limit of usage. The root categories return results that contain a large amount of unintended information. The filtering of necessary features is not always a smooth operation.

The results show Ottawa's amenities are heavily concentrated in the downtown and centretown areas. It is also implied that most of the neighbourhoods are lack of amenities. The city of Ottawa has a long way to go in terms of neighbourhood amenities.

# Conclusion and Future Directions

This project has met its goal of exploring Ottawa's neighbourhood amenities. It summarizes amenity-specific information based on the Foursquare API's response. The summary includes an amenity ranking and amenity-based clusters of neighbourhoods. It identifies the least amenity neighbourhoods as 'cluster label 0' and most amenity neighbourhoods as 'cluster 1' and 'cluster 2'. Stakeholders from real estate industry, homebuyers, development authorities, and entrepreneurs can find useful information in the summary. There is a big difference between Ottawa's most and least amenities neighbourhoods. Interested stakeholders can develop their future projects to reduce the difference.

For future directions, use of multiple sources of amenity data could be considered. As Foursquare search results vary with time stamps, additional sources can improve the reliability of summarized information. In many cases, the frequency of certain features or categories is not a reliable metric. New metrics such the amenities versus population density of a neighbourhood should be considered in the future. This will address the capacity of service issues of amenities.