CAPSTONE FINAL ASSIGNMENT

EXPLORING VENUES FOR VISITORS IN OTTAWA, ON, CANADA

Table of Content

1. Introduction

This project focuses on exploring venues that best suit for visitors in Ottawa (ON, Canada). Ottawa is the capital of Canada with the area about 2,790 km^2. There are many venues for visitors such as restaurants, hotels and coffee shops.

When visitors plan to travel to a city, they will be looking for places to stay, to visit, to eat and drink... They primarily aim to places having good ratings across all venues and the average prices that fit their budget. Therefore, we will especially focus on the ratings and average prices of the venues.

We need the support of two API providers: Foursquare API and Zomato API. The former is to determine the locations of interest while the latter provides their rating and price information. This work helps to identify places that are fit for different visitors. Once we finish exploring these venues, any company can launch an application and provide suggestions for users.

2. Data Collection

We will fetch the data using the two different APIs as follow:

- Foursquare API fetches venues in Ottawa in the radius of 4 kilometers from the city center.
- Zomato API provides information related to the venues within this area, e.g. the complete address, user ratings, price range, and so on.

The data will be represented as tabular form.

2.1. Foursquare API

We determine venues in Ottawa up to a radius of 4 kilometers from the city center using the Foursquare API. We consider only 4 kilometers because of the quota of number of calling given by Foursquare. We use the *explore* API to find all the venues within the area. We iteratively call the API until obtaining all the venues within the given radius. Each time we call, the API will provide up to 100 venues.

The obtained data is as follow:

	name	categories	lat	lng
0	National Arts Centre - Centre National des Arts	Concert Hall	45.422922	-75.692996
1	Sansotei Ramen 三草亭	Noodle House	45.418920	-75.699328
2	National War Memorial	Memorial Site	45.423909	-75.695221
3	Rideau Canal	Other Great Outdoors	45.424781	-75.695129
4	Alt Hotel Ottawa	Hotel	45.419973	-75.698948

Figure 1. Data provided by Foursquare API

2.2. Zomato API

The Zomato API can search for a venue given certain search filters such as query, latitude, or longitude. The search API will be performed according to the name, latitude, and longitude values of the fetched venues using an URI.

The obtained data is as follow:

	venue	latitude	longitude	price_for_two	price_range	rating	address
0	Le Café	45.4232850000	-75.6946990000	115.0	4.0	3.4	National Arts Centre, 53 Elgin Street, Ottawa ...
1	Sansotei Ramen	45.4189760000	-75.6991850000	0.0	1.0	3.8	153 Bank Street, Ottawa K1P 5N7
2	Le Café	45.4232850000	-75.6946990000	115.0	4.0	3.4	National Arts Centre, 53 Elgin Street, Ottawa ...
3	Wilfrid's Restaurant - Fairmont Château Laurier	45.4251800000	-75.6944100000	120.0	4.0	3.5	Fairmont Château Laurier, 1 Rideau Street, Ott...
4	Juice Monkey	45.4198833333	-75.6996166667	25.0	2.0	3.5	209 Slater Street, Ottawa K1P 5N5

Figure 2. Data provided by Zomato API

3. Data Cleaning

Since the current data is collected from different sources, we need to combine them and then perform data cleansing.

3.1. Merging data from the two data frames

For each venue, we have two data points where each one belongs to the results provided by one of the two APIs. We thus need to find a way combining them properly.

It is typical to guess that there may be a small coordinate variation (i.e. distance) between the two points (of the two data frames) indicating the same venue.

First, such variation may come from the problem of number representation on computer. Therefore, we should try to round the coordinate values with a smaller number of decimal places. Concretely, the latitude and longitude will be rounded to have 4 decimal places. Venues of small location variation are now matched between the two data frames.

There is also a possible risk that the two obtained coordinates for a venue has a significant distance, i.e. they are still not the same after rounding. In this case, we can use a distance threshold to determine whether they indicate the same venue or not. We will set the threshold as 0.0004 for both latitude and longitude displacements.

The current data frame is structured as follow:

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address
0	Sansotei Ramen 三草亭	Noodle House	45.4189	-75.6993	Sansotei Ramen	45.4190	-75.6992	0.0	1.0	3.8	153 Bank Street, Ottawa K1P 5N7
1	Riviera	Modern European Restaurant	45.4232	-75.6962	Riviera	45.4234	-75.6963	25.0	2.0	3.8	62 Sparks Street, Ottawa K1P 5A5
2	Confederation Park	Park	45.4221	-75.6925	Royi Fruta Bar	45.4222	-75.6924	10.0	1.0	0	Confederation Park, Ottawa K1P 5J2
3	North & Navy	Italian Restaurant	45.4167	-75.6990	North & Navy	45.4168	-75.6991	90.0	4.0	4.1	226 Nepean Street, Ottawa K2P 0B8
4	La Bottega	Deli / Bodega	45.4269	-75.6919	La Bottega Nicastro	45.4270	-75.6921	20.0	2.0	4.6	64 George Street, Ottawa K1N 5V9

Figure 3. Combined data frame

It is noticeable that some venues have the mismatched names in the two sources, e.g. Confederation Park and Royi Fruta Bar at index 2.

There are different possible reasons such as: **(1)** they are 2 different venues located in the same building, or **(2)** they are two close and their coordinates are thus quite similar, or **(3)** an old venue was replaced by the other one (due to the different updated dates of the two APIs). In the example of index 2, it seems the third reason since the corresponding rating is 0 (i.e. it is a new venue).

They may be fixed in various ways such as waiting for the updated API version, manual tuning the information, or simply skip them (since users normally focus more on the venue category than its name).

3.2. Normalizing and removing unnecessary information

We first normalize the price by dividing the values in column *price_for_two* by 2 in order to obtain the average price for a single visitor.

We then remove unnecessary columns, in which we decide to work on the venue name from Zomato API.

For the venues that do not have rating, we should remove them since the visitors will not expect going there.

This is a heading view of the final dataset for the further exploration:

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Noodle House	Sansotei Ramen	45.4190	-75.6992	1.0	3.8	153 Bank Street, Ottawa K1P 5N7	0.0
1	Modern European Restaurant	Riviera	45.4234	-75.6963	2.0	3.8	62 Sparks Street, Ottawa K1P 5A5	12.5
2	Italian Restaurant	North & Navy	45.4168	-75.6991	4.0	4.1	226 Nepean Street, Ottawa K2P 0B8	45.0
3	Deli / Bodega	La Bottega Nicastro	45.4270	-75.6921	2.0	4.6	64 George Street, Ottawa K1N 5V9	10.0
4	Market	Maple Valley Tea World	45.4276	-75.6925	1.0	3.5	55 ByWard Market Square, Ottawa K1N 9C3	5.0

Figure 4. Pre-processed data frame

4. Methodology

We perform analysis on the data based on the ratings and the average price of each venue. We first determine venue categories that are majority in the area of interest. According to that information, the visitors can expect these activities when going there. We then explore areas of high ratings and those of low rating using visualization. The results will be useful for users since they tend to visit high rated places and avoid low rated ones. Finally, we discuss and conclude which venues the visitors should explore based on their requirement of rating and cost.

4.1. Categories

We have various types of venues in the final dataset. We will take a look at the venues and check which are the majority venue categories in the list.

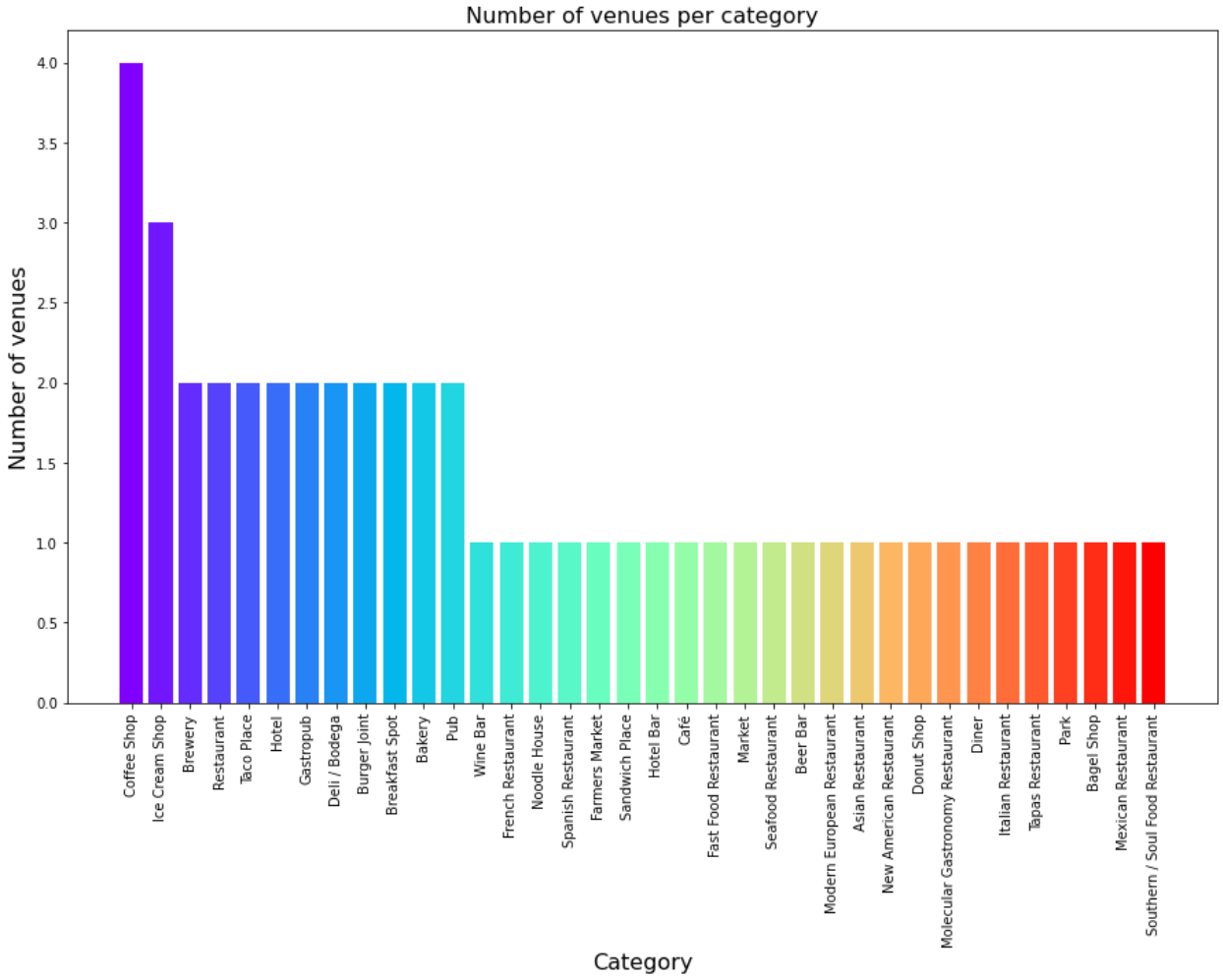

Figure 5. Statistic of venue categories

It is obvious that the two majority venues are coffee shop and ice cream shop. Therefore, there is the good place for visitors who want to relax with a cup of coffee or enjoy ice cream.

4.2. Rating

Venue's rating is an important factor for a visitor to decide whether it is worth it to visit the place. We take a look on the overall statistic of rating.

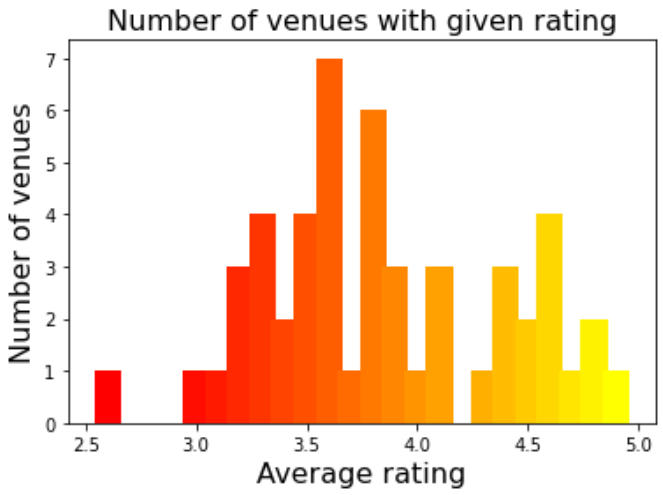

Figure 6. Statistic of venue rating

The plot above shows that most venues have the rating of 3.6, and the distribution tends to skew to the higher rating direction. There are also many venues of high ratings from 4.3 to 4.9.

For a better qualification, we create bins for ratings and plot them in different colors on the map. The ratings will be divided between 4 bins corresponding to the qualities Very good, Good, Average, and Low:

- 1 to 2: Low (highlighted by red color)
- 2 to 3: Average (highlighted by orange color)
- 3 to 4: Good (highlighted by dark green color)
- 4 to 5: Very good (highlighted by green color)

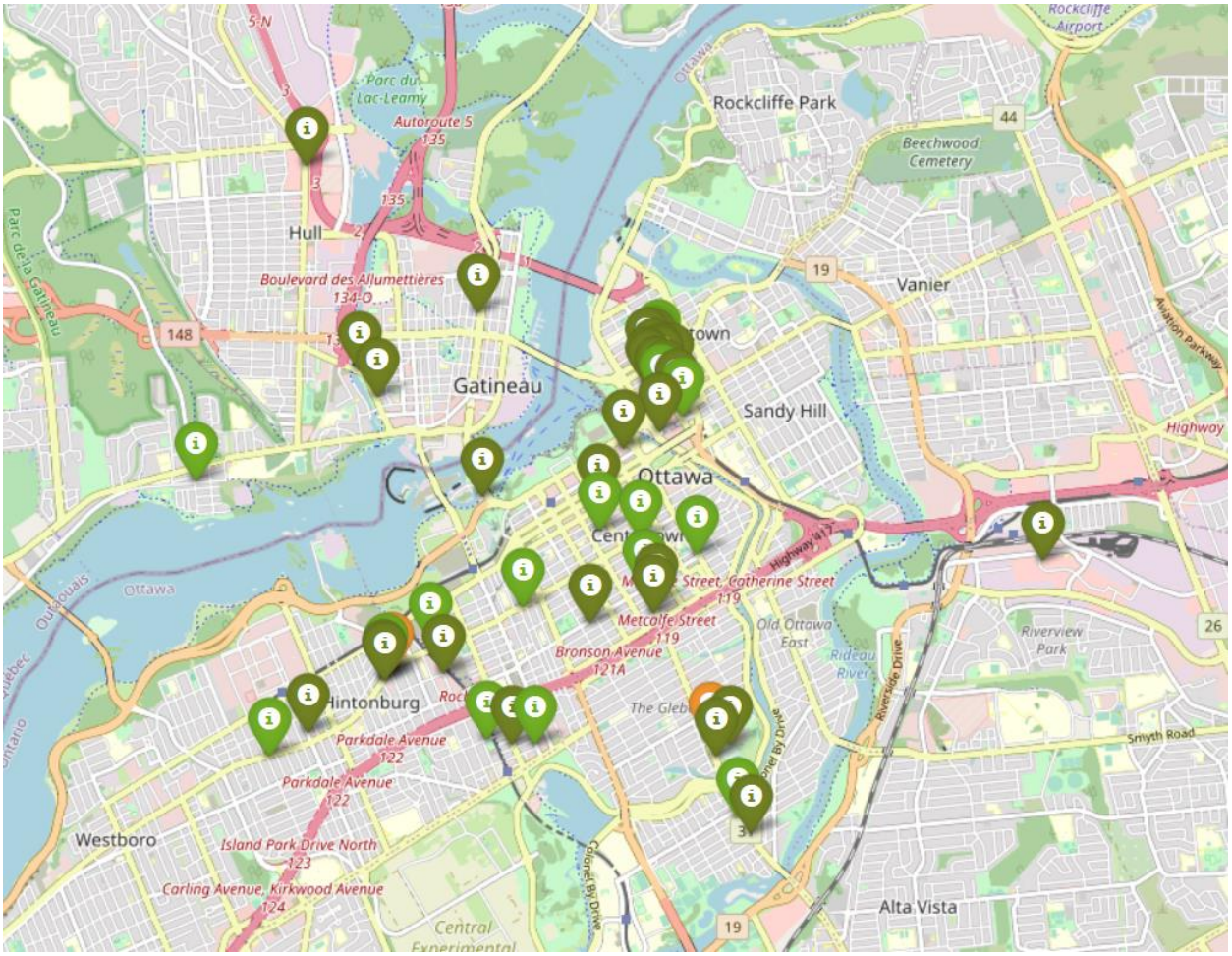

Figure 7. Visualization of venue quality

According to the markers shown on the map, the venues located near an end of the **Pont Alexandra bridge** tend to have good and very good ratings (3.0 to 5.0) and they are close to each other. It is a good idea for visitors to check out that area. The areas **Centertown** and **Hintonburg** are also worth to visit though they have less venues compared to the first area.

4.3. Price

We can check the average price per person. There are two price features including *average_price* which indicates the average cost for one person and *price_range* which determines the price range as defined by Zomato.

We first explore the *average_price* using a scatter plot between the average price and the number of venues. The point size indicates the corresponding price.

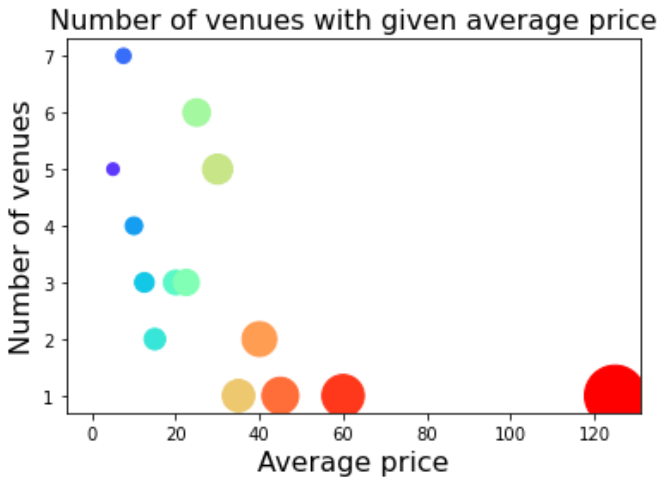

Figure 8. Statistic of venue's average price

The figure shows that most venues have low price (less than 40). Therefore, this area of Ottawa is suitable for visitors who want to travel without spending much for their budget.

In some cases, users focus on the overall price of an area where they plan to spend their money for various activities that fit their budget. We will use the *price_range* information to plot the venues on a map. The color changes from green to red as the increasing of price.

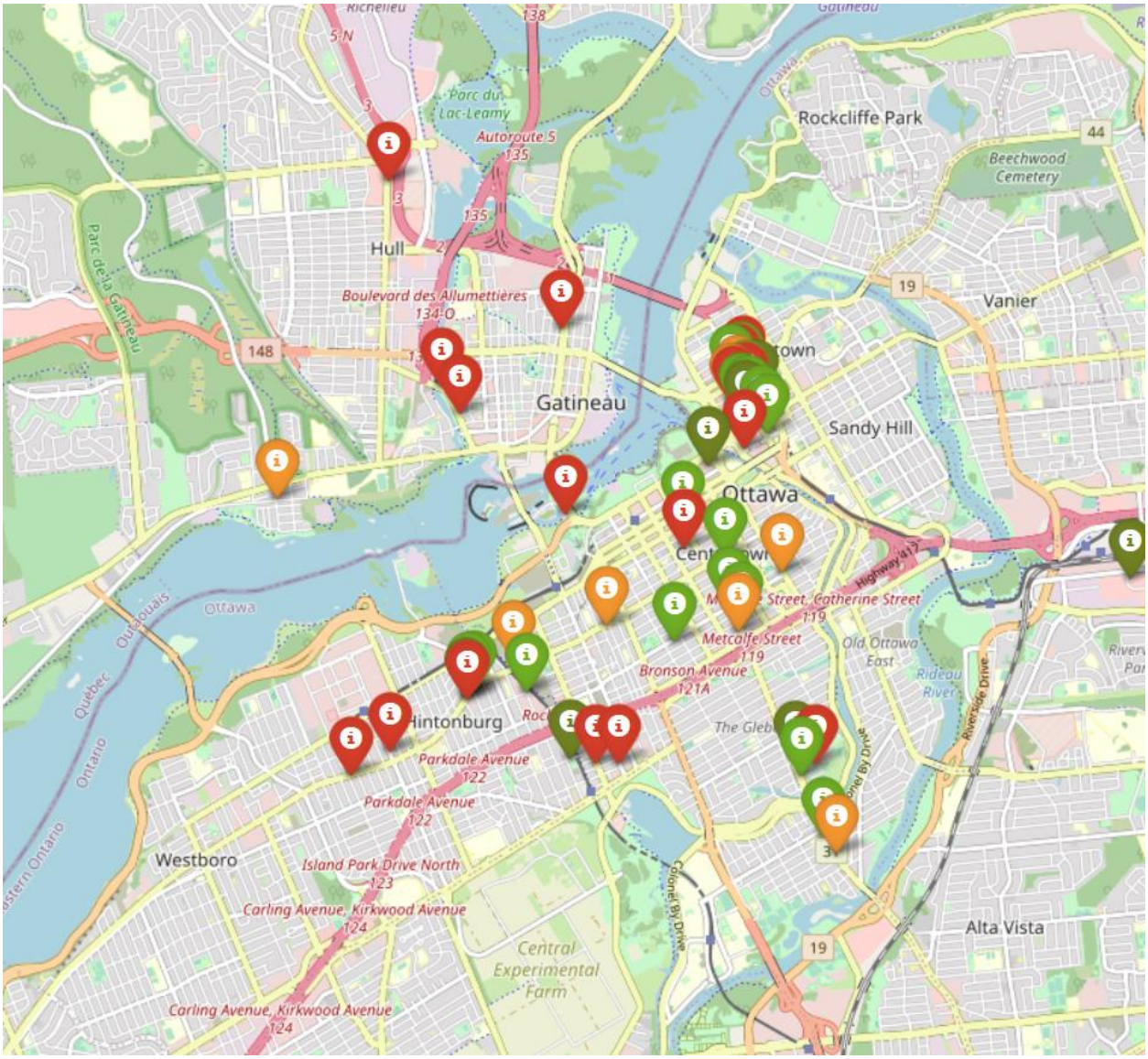

Figure 9. Visualization of venue price range

Once again, we see that the venues located near an end of the **Pont Alexandra bridge** and the **Centertown** areas are suitable for visitors who want to save their pocket. The visitors who want to spend their money a lot can go to nearby areas, including the other end of the bridge.

5. Results and Discussion

According to the analysis above, we obtain a number of conclusions that will be useful to aid visitors at Ottawa, ON, Canada.

First, some mismatches of venue names in the data obtained from Foursquare and Zomato indicate that some venues may be recently replaced by new ones. Therefore, the visitors should check carefully their objective places (via website, other search engines...) before coming there.

The two majority venues in the studied area of Ottawa are coffee shop and ice cream shop. Therefore, that is the good place for visitors who want to relax with a cup of coffee or enjoy ice cream.

Most venues have the rating of 3.6 and there are also many venues of high ratings from 4.3 to 4.9. The visitors thus can satisfy with the services they will use there.

In more details, the venues located near an end of the Pont Alexandra bridge tend to have good and very good ratings (3.0 to 5.0) and they are close to each other. It is a good idea for visitors to check out that area. The areas Centertown and Hintonburg are also worth to visit though they have less venues compared to the first area.

Besides, most venues have low price (less than 40). Therefore, this area of Ottawa is suitable for visitors who want to travel without spending much for their budget. In addition, the venues located near an end of the Pont Alexandra bridge and the Centertown areas are suitable for visitors who want to save their pocket. The visitors who want to spend their money a lot can go to nearby areas, including the other end of the bridge.

A company can use the above information to build up an online website/application, to provide users with up to date information about various venues in Ottawa based on the search criteria (name, rating and price).

6. Conclusion

The objective of this project is to explore the venues that visitors at Ottawa, ON, Canada should visit. The primary criteria are venue rating and price related information. The venues have been identified using Foursquare and Zomato APIs and have been plotted on the map. The exploration can be adapted to other cities and for other criteria.