Finding the Optimal Neighborhood to Expand Restaurant Business

Malachi Jones

March 14, 2021

1. Introduction

A problem facing many successful restaurants is where they should put their new locations when they decide to expand their business. One way a restaurant can attempt to find the best area for their new locations is to search for the areas that are most like wOuhere their current restaurant is located. It makes sense to try and find similar areas because given that the initial restaurant was successful, a restaurant in an area with similar characteristics should have a higher probability of success.

In this report, we will be analyzing a specific scenario under these guidelines. It is assumed the original restaurant is in the Rosedale neighborhood of Downtown Toronto. The restaurant owner wants to open a new location either in Toronto or another Canadian city. The question is what city and neighborhood has characteristics most like Rosedale. The map above shows where each area in Toronto is located within the city. The blue dot represents Rosedale. While this report uses a specific example, the methodology and general framework can be applied for different neighborhoods and cities.

2. Discussion of Data

2.1 Data Sources

For the analysis, neighborhood location data and data on the types of venues in the area will be needed. Two data sets are needed to get the necessary information. First, a data set consisting of a list of all the postal codes in Canada along with the name associated with that postal code and its latitude and longitude coordinates. This data set came from AggData (https://www.aggdata.com/free/canada-postal-codes). To get data on the types of venues in each area, the Foursquare API is used.

2.2 Data Cleaning

First, the postal code dataset was downloaded and cleaned. After importing the data set, the postal codes from the desired cities were selected. For the analysis, only the top seven cities in Canada by population were used. It is likely that the areas most closely resembling the area with the original restaurant will be in an urban area. These cities are Toronto, Montreal, Vancouver, Calgary, Edmonton, Ottawa, and Winnipeg.

Next, venue data was imported using the Foursquare API for each of the postal codes. The venue type was extracted for each venue and grouped by postal code. For each postal code, the proportion of each venue type compared to the total venues in that area was calculated. For example, Calgary (Bridgeland / Greenview / Zoo / YYC) has a proportion of 0.111 for water parks. This means that 11.1% of the venues in the area are classified as water parks. The proportions along with the names of each area were then put into a table. In total, there are 201 different areas in our data set. Also, there are 388 different categories of venues.

3. Methodology

3.1 Exploratory Analysis

Using the proportions that were derived as described above, the top venues for each area can be calculated. Below is a table showing the top ten most common venues for Downtown Toronto (Rosedale).

Table 1: Most Common Venues for Rosedale		
Rank	Type	Proportion
1	Park	0.294

|		
2	Trail	0.118
3	Playground	0.059
4	Grocery Store	0.059
5	Candy Store	0.059
6	Café	0.059
7	Historic Site	0.059
8	Scenic Lookout	0.059
9	Farmers Market	0.059
10	Gym \ Fitness Center	0.059

3.2 K-Means Clustering

The objective our analysis will be to find other areas that have a similar breakdown of venue types. The method that is used for this analysis is k-means clustering. K-means clustering divides the data set into k-number of clusters based on how similar their characteristics are. The model is run for many different k's to find the optimal number of clusters for the data. The goal is to find the cluster of which Downtown Toronto (Rosedale) is a member. These are the areas that are most similar to the one where the original restaurant is, so they will be good spots to possibly expand.

4. Results

First, the k-means clustering model is computed for each k from 2 to 15. An elbow plot was made to see if the optimal k can be determined. Below is the elbow plot created by the models:
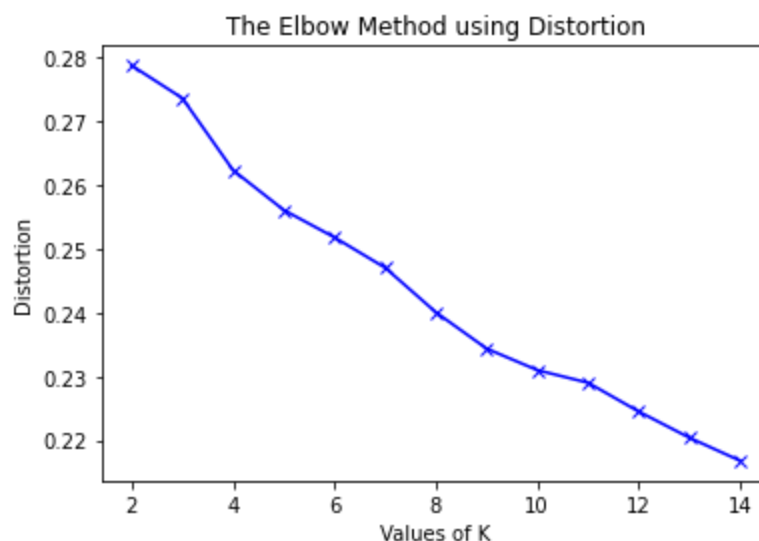
There is no distinct elbow in the plot where the slope flattens quickly, so it is unclear how many clusters is appropriate. Another method for determining the optimal k is by calculating the silhouette score for each k. Below this plot is displayed:

Silhouette Scores for a Variety of Ks
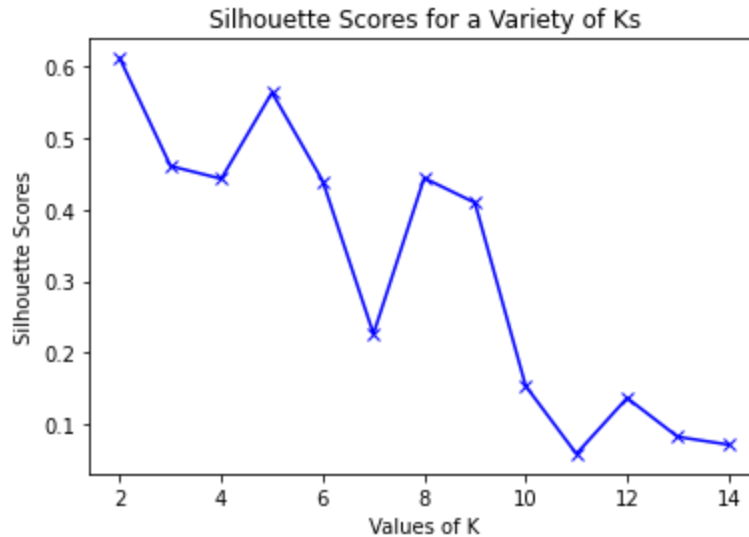
Based on these two tests, the number of clusters our analysis uses is 5. It has a high silhouette score and looks like it has close to an elbow in the elbow plot.

Now, the k-means clustering algorithm is conducted on the data set with 5 clusters. The following is the list of areas who belong to the same cluster as Downtown Toronto (Rosedale).

1. Vancouver (Chaldecutt / South University Endowment)
2. North Vancouver East Central
3. North Vancouver North Central
4. North Vancouver Northwest Central
5. North Vancouver Northwest
6. West Vancouver North
7. West Vancouver West
8. Calgary (Brentwood / Collingwood / Nose Hill)
9. Ottawa (Rockcliffe Park / New Edinburgh)
10. Winnipeg (Seven Oaks West)
11. Winnipeg (Inkster West)
12. Winnipeg (St. James-Assiniboia SE)
13. Winnipeg (Grassie / Pequis)

The cities that have areas that are close characteristically to Downtown Toronto (Rosedale) are Vancouver, Calgary, Ottawa, and Winnipeg. Below are maps of each of these cities showing

where the areas (represented by the blue dots) that are clustered with Rosedale are located within the city:
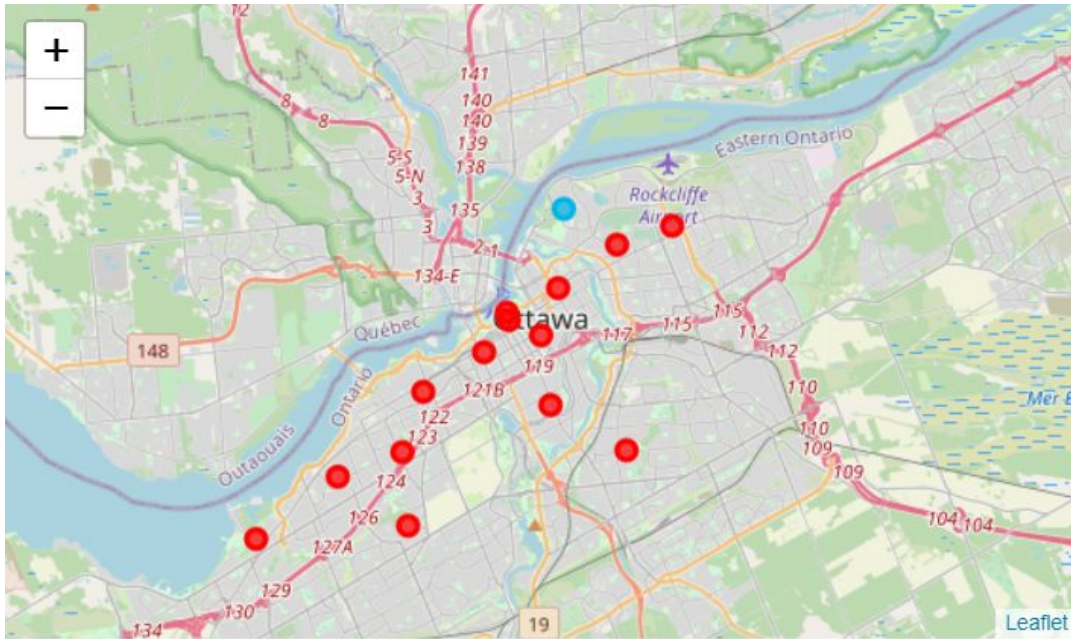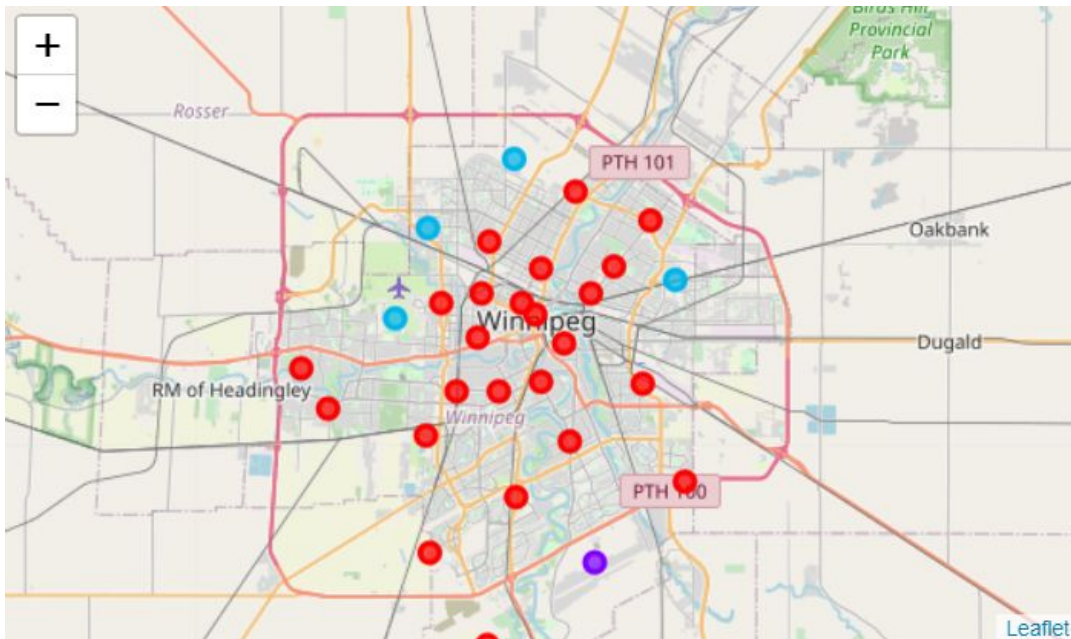
Vancouver
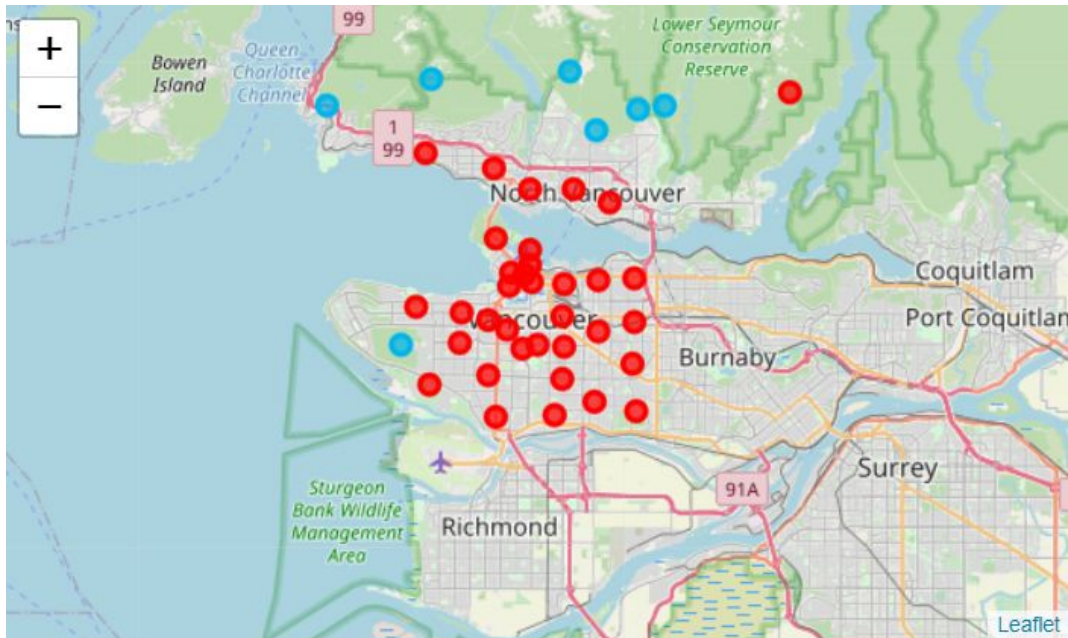
Calgary
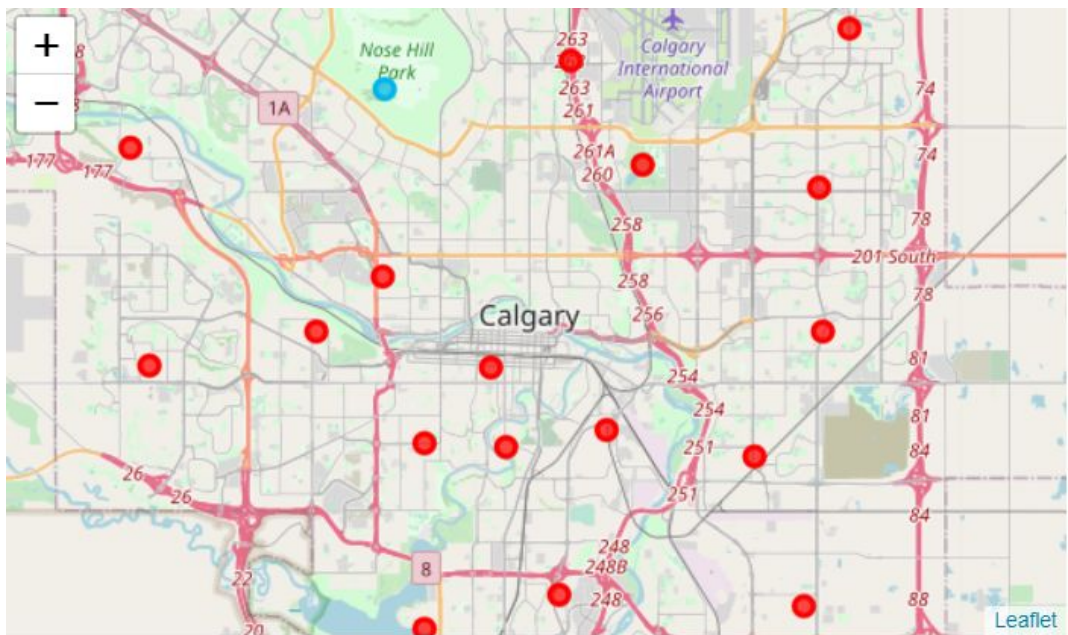
Ottawa

Winnipeg

From the list and maps, it is clear that both Vancouver and Winnipeg have the most areas similar to Rosedale.

5. Discussion

From our analysis the following recommendations could be made to the owner of the restaurant in Rosedale about possible areas to open new locations. Any of the areas that are in the same cluster which are listed in the results section are good areas to begin investigating further. Because of the quantity of locations in both Vancouver and Winnipeg, it is recommended to look into locations in those areas first. Ideally, the location would be put in an area where even the surrounding area has some of the similar characteristics.

6. Conclusion

In conclusion, using k-means clustering, the areas that are most like Downtown Toronto (Rosedale) were able to be determined. Using the same methodology with different datasets, the same analyses and determinations could be done for other cities or businesses. To make the results even stronger, demographic data of the areas could be added. For instance, the socioeconomic, racial, and age demographics of each area would help further distinguish each area and tell a fuller picture of the make-up of the area. This would allow even better recommendations to be made. Overall, this is a good beginning analysis for a restaurant business looking to open new locations in a different region.