**The Battle of Neighborhood: The Best Community for Starting a Restaurant**

**Zheng Zhao**

Table of Contents

[1. Introduction 1](#_Toc28855267)

[2. Data 1](#_Toc28855268)

[3. Methodology 2](#_Toc28855269)

[4. Discussion 3](#_Toc28855270)

[5. Conclusion 4](#_Toc28855271)

**Table of Captions**

[*Figure 1 2*](#_Toc28855320)

[*Figure 2 3*](#_Toc28855321)

[*Figure 3 3*](#_Toc28855322)

[*Figure 4 4*](#_Toc28855323)

# Introduction

The main idea of the capstone project is to find out the best area for running a restaurant. That would base on the exploration of Halifax city which in an eastern province, Nova Scotia, Canada. It is not one of the biggest cities in the world, however, it is considered as one of the best cities for living in in Canada. Hence, the assumption is basically made base on this city. The reason why not choose the big city like New York, Toronto or London was that restaurants are satiate already. There is less meaningful to study into that sort of cities. Besides, Halifax is identified as one of the fastest developing cities in Canada. Consequently, studying into the areas of Halifax is the best for running a restaurant is meaningful for those who intend to running their own catering business in Halifax.

# Data

Data was implemented of this project is mainly the coordinates of Halifax.

1. The first step is to find out the communities, area or neighborhood classification of Halifax. The first idea is to search in Wikipedia. Hence, the communities are found in

<https://en.wikipedia.org/wiki/Communities_in_the_Halifax_Regional_Municipality>.

There are four urban communities was captured into the study: Halifax Peninsula, Mainland Halifax, Dartmouth and Bedford. The sub-communities were scraped by using the BeautifulSoup package. Also, the coordinates of each communities were captured from Wikipedia link.

1. The second step would be passing the coordinates to Foursquare database and explore the venues of the communities and transform into map and find the areas with low supply of restaurants.
2. However, during the programming section, the coordinates from Wikipedia is not completable for folium package. Therefore, geocoder package was used for obtain the coordinates for Halifax communities. This step is to pass the communities' names to geocoder and generate coordinates for mapping.
3. Mapping the communities:

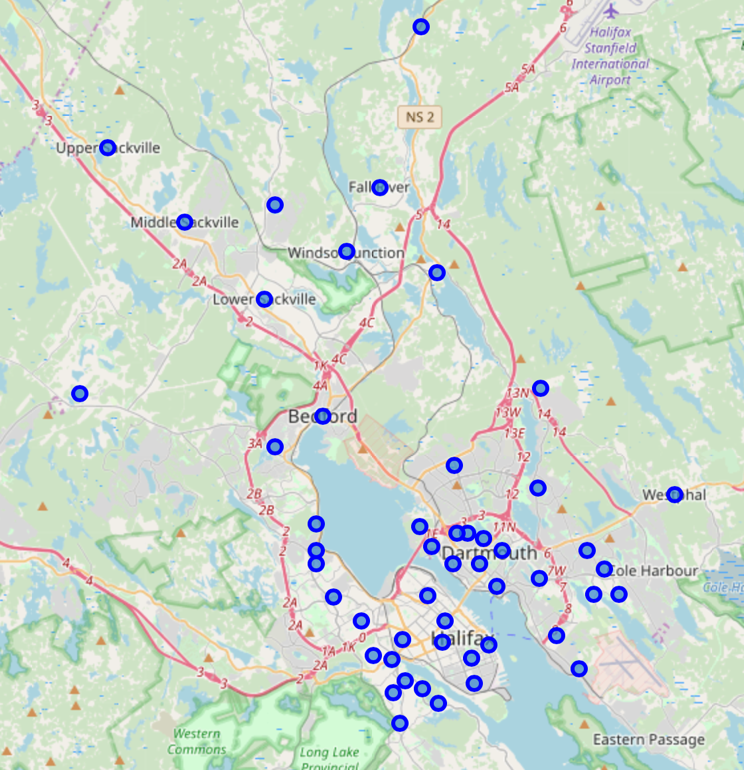


Figure 1

# Methodology

The main methodology to find out the area for opening a restaurant is K-means clustering, which is a unsupervised machine learning approach that grouping data base on their similarity. In the instance of this research purpose, K-means clustering was applied to group neighborhoods based on venues were searched from Foursquare database.

For identifying the best K value for clustering, Euclidean distance was applied for calculating silhouette score on K value from 3 to 20. Due to there are over 40 neighborhoods needed to be analysis, Hence, 3 clusters were determined as the minimum. The result was snapshotted into follow:

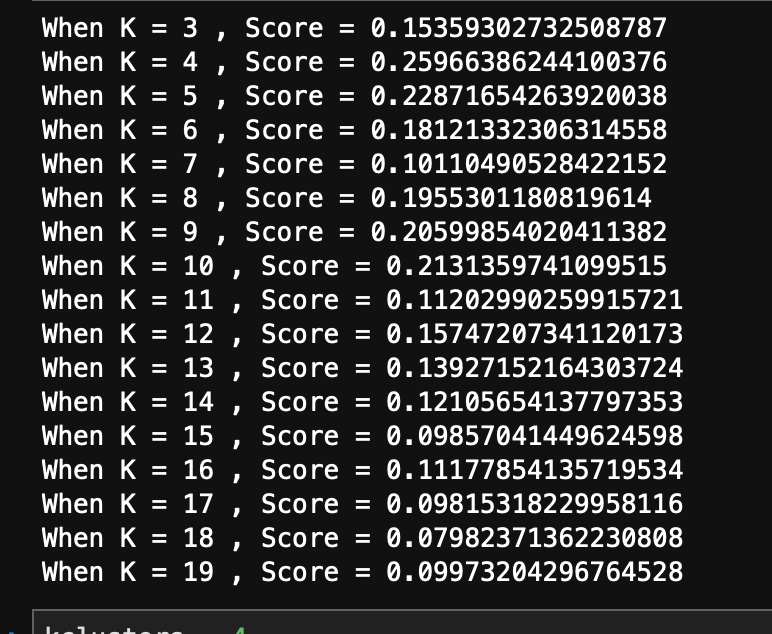


Figure 2

Consequently, when K equals to 4, the silhouette score is maximum which is around 0.26. After applied K-means clustering, and set different colors for clusters. The map had been conducted into follow:

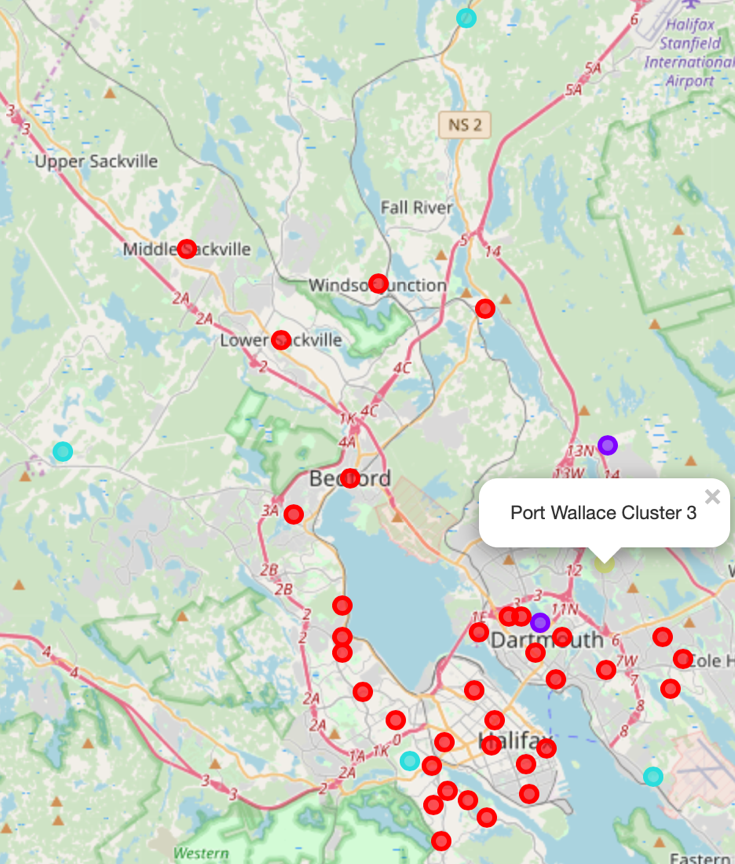


Figure 3

Due to the color of cluster 3 is yellow, which is not cleared in the map, the label was snapshotted. As we can see, there are two purple dots, which is cluster 1 and four blue dot is cluster 2.

# Discussion

According to Figure 3, most of the flourishing area was clustered into label 0, which are in red color. Hence, if starting a new restaurant in one of the cluster 0 area, there are high competition between others. Thus, the neighborhood that we will finally recommend should be chosen from cluster 1, cluster 2 and cluster3.

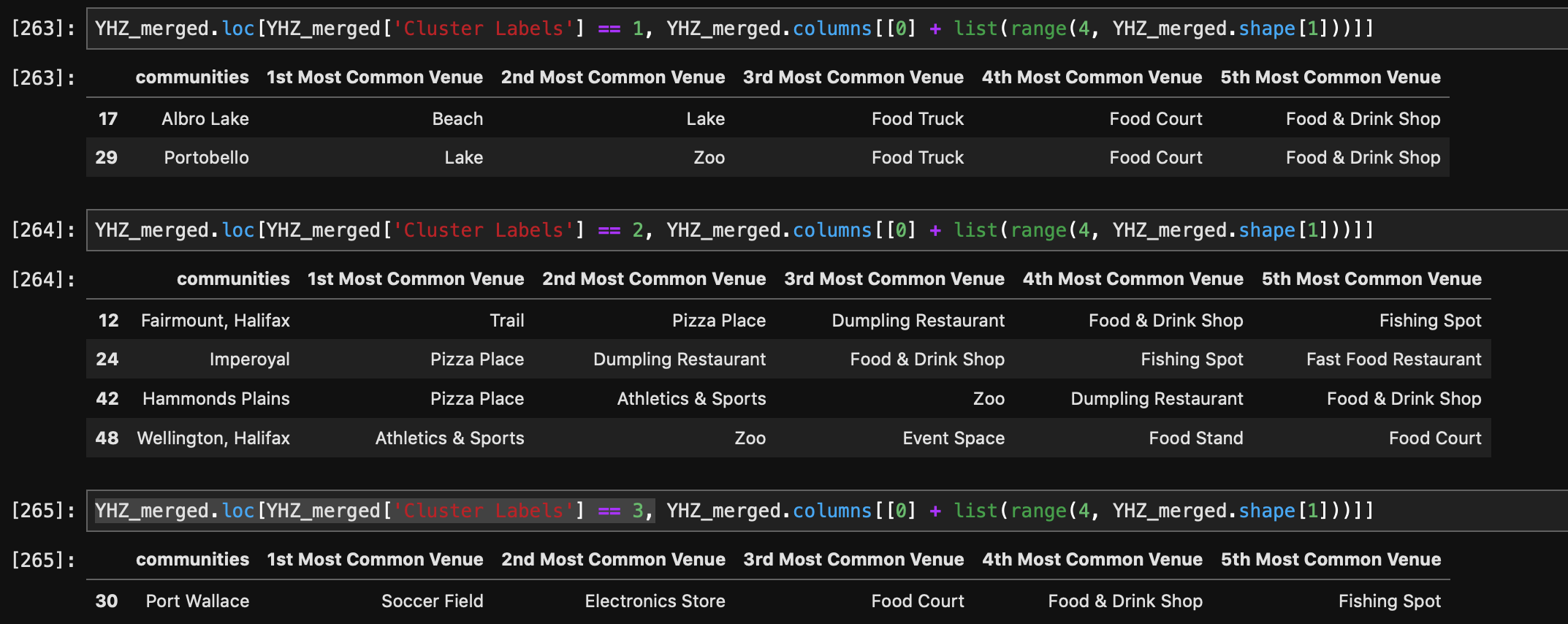


Figure 4

From Figure 4 above, there is only a few communities was included into these three clusters. For cluster 1, Albro Lake and Portobello, the first two most common venue are infrastructure of entertainment, which are Beach, Lake and Zoo. But, the third and after most common venues for them are all with food supply. Consequently, these communities should be carefully considered. For cluster 2, three of four communities are having most common venues with restaurants. These three communities should not be considered as candidates for starting new restaurant. However, Wellington communities is having infrastructures of common used for top three the most common venues. For cluster 3, there is one community, Port Wallace. The most common venue is soccer field, and the second is stores, while the third is food court.

# Conclusion

For the best choice for starting a new restaurant, the first recommendation should be Wellington, Halifax in cluster 2, which with three the most common venues with sports spaces, zoo and event space. Apparently, less competition between other restaurants and demand of food supply could be high enough. The second choice could be filtered could be Albro Lake and Portobello community from cluster 1, and Por Wallace from cluster 3.

However, in this research, the population of communities and travel rate of locals or visitors have not being considered. Even though the recommended communities are appropriate for restaurant positioning, the population might be low as there are low supply of restaurants. This might generate a bias that rural areas could be always considered as the communities for recommendation. Hence, the limitation of this research is obvious, which due to lack of data of population and data travel from any authorized organisations. Never the less, future studies of Halifax communities on business starting should consider the demand of respective domain.