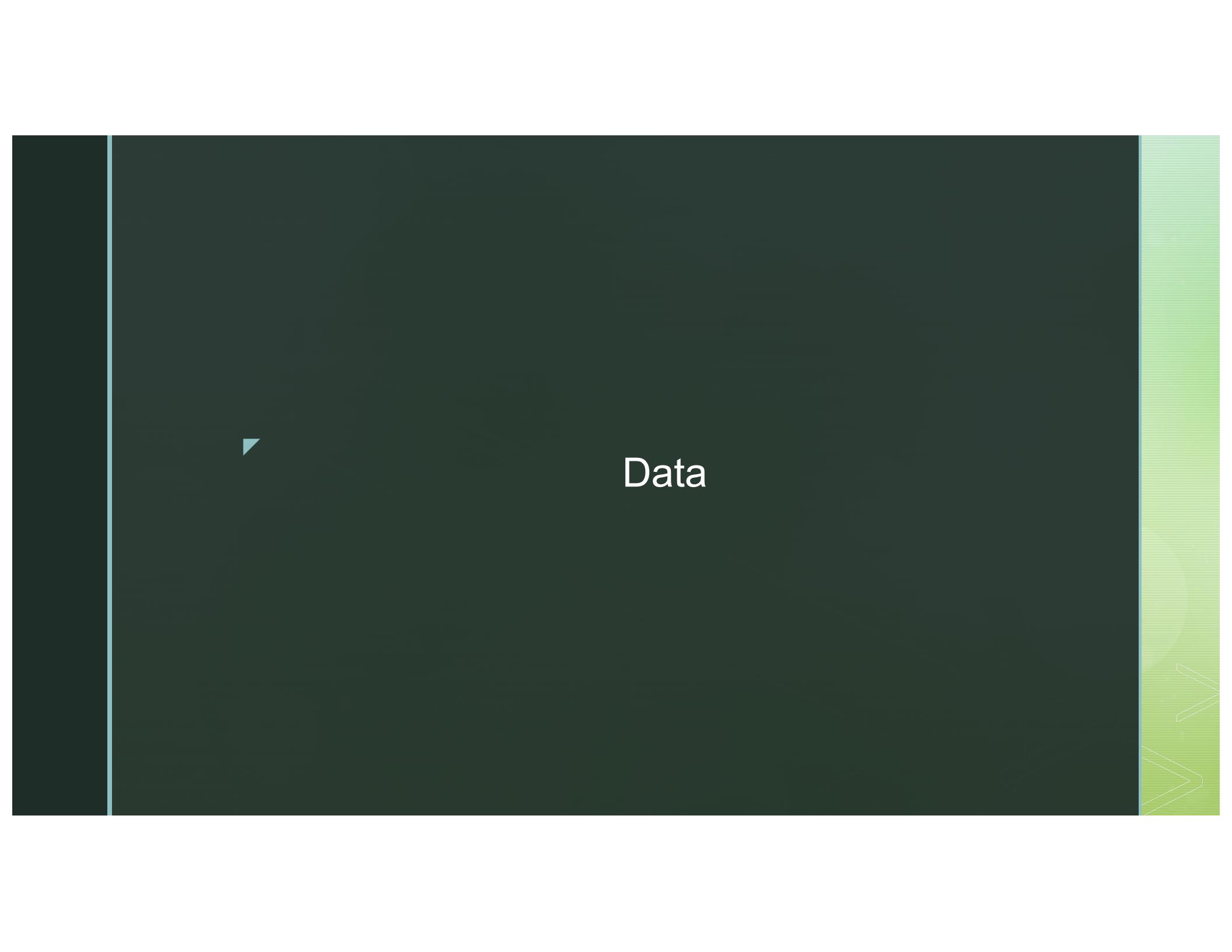


Zheng Zhao

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

Introduction

- City: Halifax, Nova Scotia, Canada.
- Interest: Big city like London, Toronto, New York etc, could have been studied enough for starting up new businesses. As a middle-size and one of the most comfortable cities in Canada is worth for studying.



Data

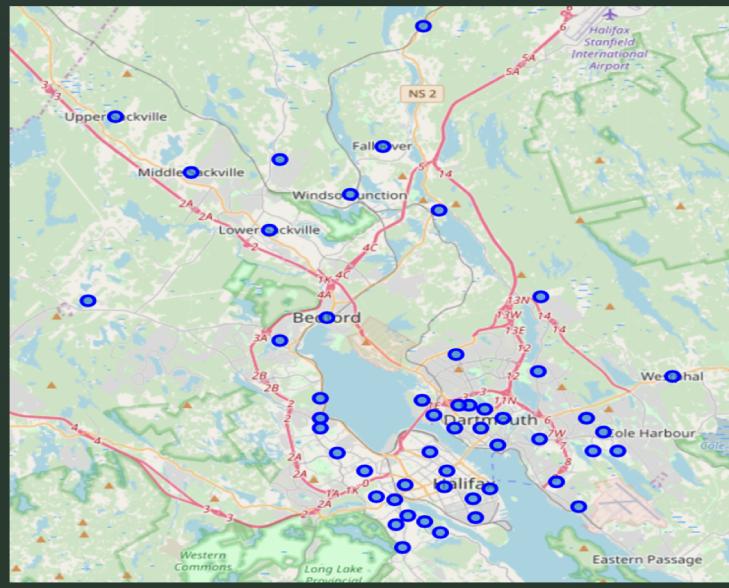
Data

- First: Finding out the communities partition in Halifax

Halifax [edit]	Dartmouth [edit]	Bedford-Sackville Region
Halifax Peninsula [edit] <ul style="list-style-type: none">• Downtown Halifax• The Hydrostone• North End Halifax• West End, Halifax• Quinpool district• Richmond• South End Halifax• Spring Garden	<ul style="list-style-type: none">• Albro Lake• Crystal Heights• Highfield Park• Brightwood• Burnside• Crichton Park• Downtown Dartmouth• Imperoyal• Manor Park• Nantucket• Portland Estates• Portland Hills• Portobello• Port Wallace• Shannon Park• Tuft's Cove• Wallace Heights• Westphal• Woodlawn• Woodside	<ul style="list-style-type: none">• Bedford• Beaver Bank• Birch Cove• Fall River• Glen Moir• Hammonds Plains• Lower Sackville• Middle Sackville• Princes Lodge• Upper Sackville• Waverley• Wellington• Windsor Junction
Mainland Halifax [edit] <ul style="list-style-type: none">• Armdale• Boulderwood• Clayton Park• Cowie Hill• Fairview• Fairmount• Jollimore• Melville Cove• Rockingham• Spryfield		

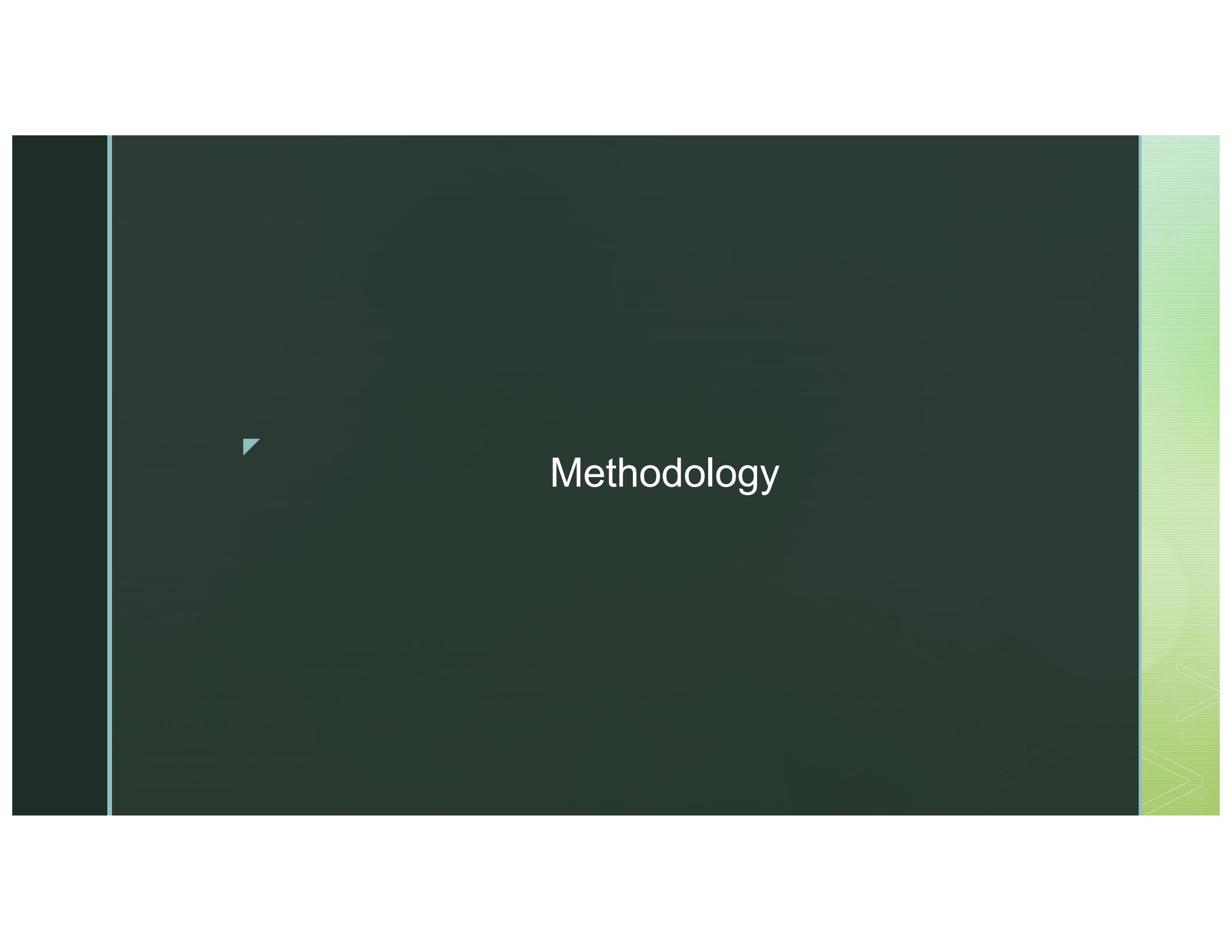
Data

- Second: In Python, BeautifulSoup package was used for scrapping the communities' names and passing these names to Geocoder for obtaining the coordinates of these communities.
- Third: Mapping the communities.



Data

- Fourth: Passing the coordinates to the Foursquare database for obtaining the venues of the communities.
- Fifth: Data pre-processing for getting the highest frequency of the venue for each communities.



Methodology

Methodology

- K – Means Clustering:
Cluster the neighborhoods base on the frequency of venues.
- Euclidean Distance:
Finding out the best K value for clustering

Methodology

- The silhouette score of Euclidean distance:

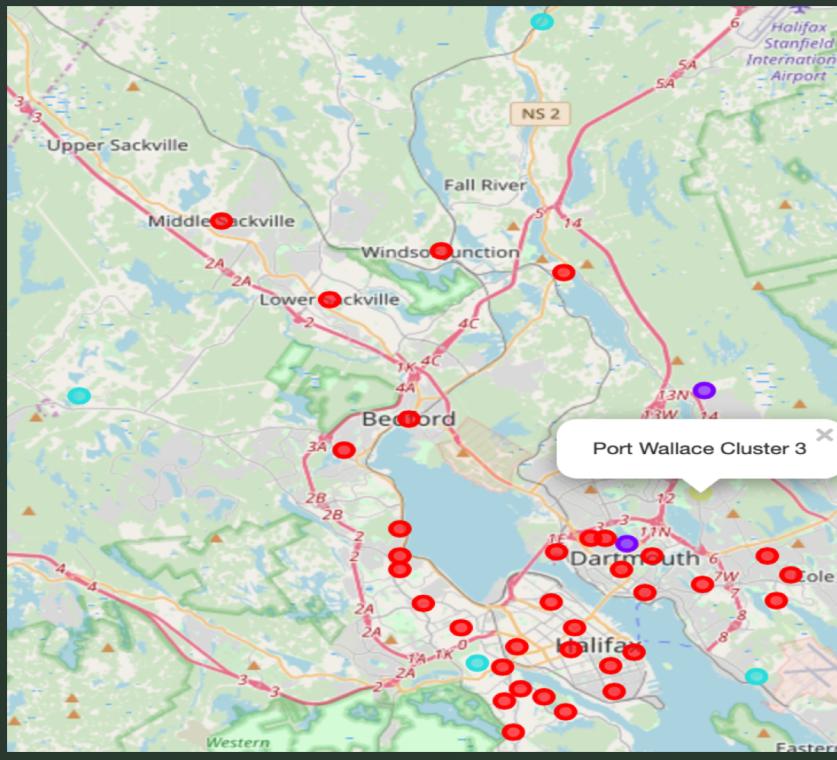
```
When K = 3 , Score = 0.15359302732508787
When K = 4 , Score = 0.25966386244100376
When K = 5 , Score = 0.22871654263920038
When K = 6 , Score = 0.18121332306314558
When K = 7 , Score = 0.10110490528422152
When K = 8 , Score = 0.1955301180819614
When K = 9 , Score = 0.20599854020411382
When K = 10 , Score = 0.2131359741099515
When K = 11 , Score = 0.11202990259915721
When K = 12 , Score = 0.15747207341120173
When K = 13 , Score = 0.13927152164303724
When K = 14 , Score = 0.12105654137797353
When K = 15 , Score = 0.09857041449624598
When K = 16 , Score = 0.11177854135719534
When K = 17 , Score = 0.09815318229958116
When K = 18 , Score = 0.07982371362230808
When K = 19 , Score = 0.09973204296764528
```

The best K value is 4

Discussion

Discussion

- The results of clustering:



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.

Discussion

- The results of cluster 1, cluster 2, and cluster 3:

```
[263]: YHZ_merged.loc[YHZ_merged['Cluster Labels'] == 1, YHZ_merged.columns[[0] + list(range(4, YHZ_merged.shape[1]))]]
```

	communities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	Albro Lake	Beach	Lake	Food Truck	Food Court	Food & Drink Shop
29	Portobello	Lake	Zoo	Food Truck	Food Court	Food & Drink Shop

```
[264]: YHZ_merged.loc[YHZ_merged['Cluster Labels'] == 2, YHZ_merged.columns[[0] + list(range(4, YHZ_merged.shape[1]))]]
```

	communities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
12	Fairmount, Halifax	Trail	Pizza Place	Dumpling Restaurant	Food & Drink Shop	Fishing Spot
24	Imperial	Pizza Place	Dumpling Restaurant	Food & Drink Shop	Fishing Spot	Fast Food Restaurant
42	Hammonds Plains	Pizza Place	Athletics & Sports	Zoo	Dumpling Restaurant	Food & Drink Shop
48	Wellington, Halifax	Athletics & Sports	Zoo	Event Space	Food Stand	Food Court

```
[265]: YHZ_merged.loc[YHZ_merged['Cluster Labels'] == 3, YHZ_merged.columns[[0] + list(range(4, YHZ_merged.shape[1]))]]
```

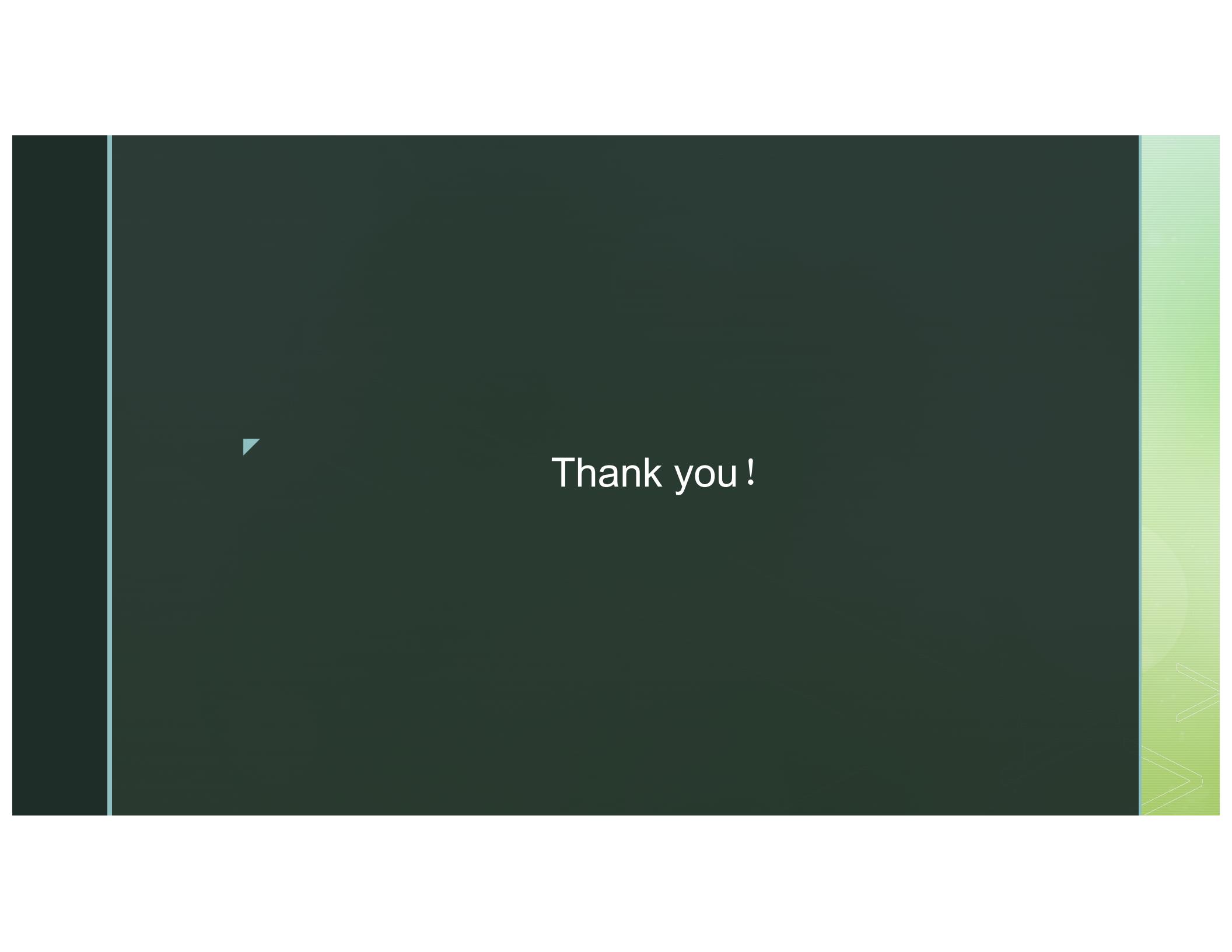
	communities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
30	Port Wallace	Soccer Field	Electronics Store	Food Court	Food & Drink Shop	Fishing Spot

Conclusion



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.



Thank you !