Exploring towns in Newfoundland using Foursquare API

Joseph Franklin

Table of Contents

Table of Contents.....2

I.Introduction.....3

II. Data Sources 3

III. Methodology 4

IV. Results and Discussions.....8

V. Conclusion 9

I.Introduction

In Newfoundland, vast majority of the dwelling units are spread across many towns and cities. Each town or city is designed such that it can be self-contained with living, education, healthcare, and recreational needs. When someone is looking for a place to stay in Newfoundland, the question will be which town/city is the best to live in?

To determine the ideal town or city to stay, the available amenities within the town or city is one of the common factors to consider.

Therefore, in this project, the objective is to investigate in the amenities within each town and city. The goal is to give an insight of the most common amenities that can be found in each town and city, as well as which town and cities are similar in terms of the available amenities. The findings might be interesting to the target audiences as follows:

1. An individual/family looking for a place to stay in Newfoundland
2. Property agents/potential customers looking to buy/sell their property
3. Property developers doing market study for potential development areas

II.Data Sources

The data that is required for this project is as follows:

1. Newfoundland town and cities names and total area The list of Newfoundland town and cities is retrieved from Wikipedia page: https://en.wikipedia.org/wiki/List_of_municipalities_in_Newfoundland_and_Labrador

2. Geographical location of each town and cities Data location of each town and cities is retrieved using GeoPy's Nominatim API based on the names of the town and city.

3. Amenities in each town and cities The Foursquare API will be used to explore the common amenities that can be found within the radius of each town and city. The explore function will be used to obtain the top 100 venues within the defined radius of each town and city. 4

III.Methodology

1. Web-scraping

The list of Newfoundland towns and cities is obtained through web-scraping from Wikipedia using the BeautifulSoup package. The data field of interest in this project are the names of the towns and cities, as well as the total area, and only these relevant data fields are extracted. The extracted data are then presented in a data frame for exploration. There are a total of 23 towns and 3 cities in Newfoundland and the top 5 rows of the data frame is as shown in Fig. 1. Figure 1:

	Town/City	Area
0	Corner Brook	148.26
1	Mount Pearl	15.76
2	St. John's	445.88
3	Admirals Beach	24.42
4	Anchor Point	2.41

Top 5 rows of data frame

2. Geocoding

With the list of the towns and cities, geocoding is performed to obtain the geographical coordinates of each town and city. The coordinates of the town/city are added to the data frame as shown in Fig. 2. Folium is then used to visualize the markers of each town and city on the OpenStreetMap according to the geographical coordinates. A visual check on the location of the markers indicate that the geographical locations data are acceptable before moving to the next step.

	Town/City	Area	Latitude	Longitude
0	Corner Brook	148.26	48.951	-57.9501
1	Mount Pearl	15.76	47.5189	-52.8063
2	St. John's	445.88	47.5608	-52.7122
3	Admirals Beach	24.42	47.0172	-53.6267
4	Anchor Point	2.41	51.2334	-56.7981

Figure 2: Data frame with geographical coordinates of town

3. RESTful API

Foursquare API explore function is then used to retrieve the information on the venues within each town and city. The key information required for the request is geographical coordinates of each town, radius and the maximum number of venues to retrieve. The geographical coordinates are provided from data obtained through geocoding and the default maximum number of venues that can be retrieved is 100 for the free license from Foursquare.

The area of the town/cities varies and in order to capture data that is most representative of the whole area of the town/cities, the radius for each town/city is individually defined according to the total area of the town/city, assuming that area of each town/city is circular. The calculated radius are added to the data frame as shown in Fig. 4.

	Town/City	Area	Latitude	Longitude	Radius (m)
0	Corner Brook	148.26	48.951	-57.9501	6869.69
1	Mount Pearl	15.76	47.5189	-52.8063	2239.77
2	St. John's	445.88	47.5608	-52.7122	11913.4
3	Admirals Beach	24.42	47.0172	-53.6267	2788.03
4	Anchor Point	2.41	51.2334	-56.7981	875.858

Figure 4: Radius of each town/city

4. Feature Engineering

The venues obtained are then grouped by the towns/city and one-hot encoding is used to convert the categorical variables to numerical values. The mean of frequency of occurrence of each category is calculated to find the top 10 venues in each town/city. An example of the top 10 most common venues of first 5 rows of the towns are shown in Fig. 5.

Town/City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mos Commo Venu
Anchor Point	Hotel	Wings Joint	Fishing Spot	Construction & Landscaping	Convenience Store	Cosmetics Shop	Dance Studio	Department Store	Diner	Discount Store
Appleton	Construction & Landscaping	Hotel	Wings Joint	Fishing Spot	Convenience Store	Cosmetics Shop	Dance Studio	Department Store	Diner	Discount Store
Arnold's Cove	Harbor / Marina	Pharmacy	Diner	Pizza Place	American Restaurant	Fishing Spot	Convenience Store	Cosmetics Shop	Dance Studio	Department Store
Badger	Boat or Ferry	Wings Joint	Fishing Spot	Convenience Store	Cosmetics Shop	Dance Studio	Department Store	Diner	Discount Store	Dive Bar
Baie Verte	Auto Garage	Sandwich Place	Electronics Store	Bistro	Wings Joint	Fish & Chips Shop	Convenience Store	Cosmetics Shop	Dance Studio	Department Store

Figure 5: Top 10 most common venue in each town/city

5. Unsupervised learning Unsupervised learning is implemented to determine the similarities among the towns/cities in terms of the available amenities. K-means clustering is chosen as the model and the optimal number of clusters is determined by using a combination of the elbow method and silhouette score.

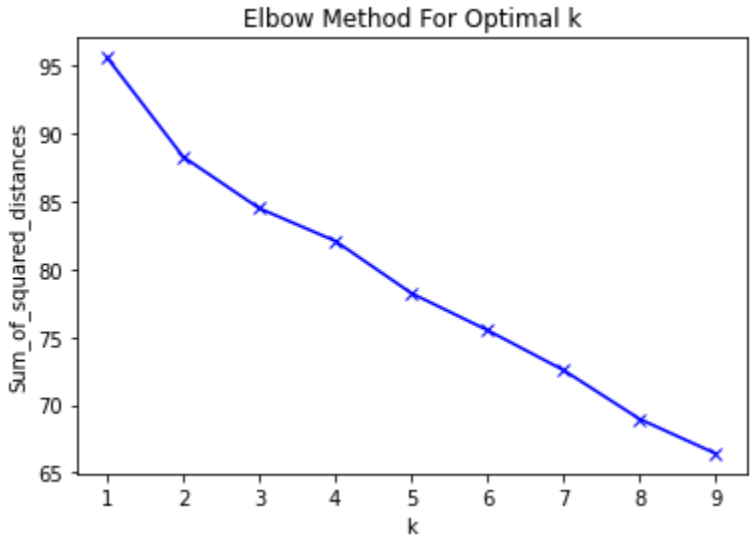

Figure 6: Elbow method

Figure 7: Silhouette score

The elbow method shows the elbow point at k=2, which coincides with the highest silhouette score at k=2. Therefore, the number of clusters chosen is 2. The results and discussions are presented in the next section.

IV.Results and Discussions

The towns/cities in Newfoundland can be grouped in 2 main clusters as visualized by the below map in Fig. 8.

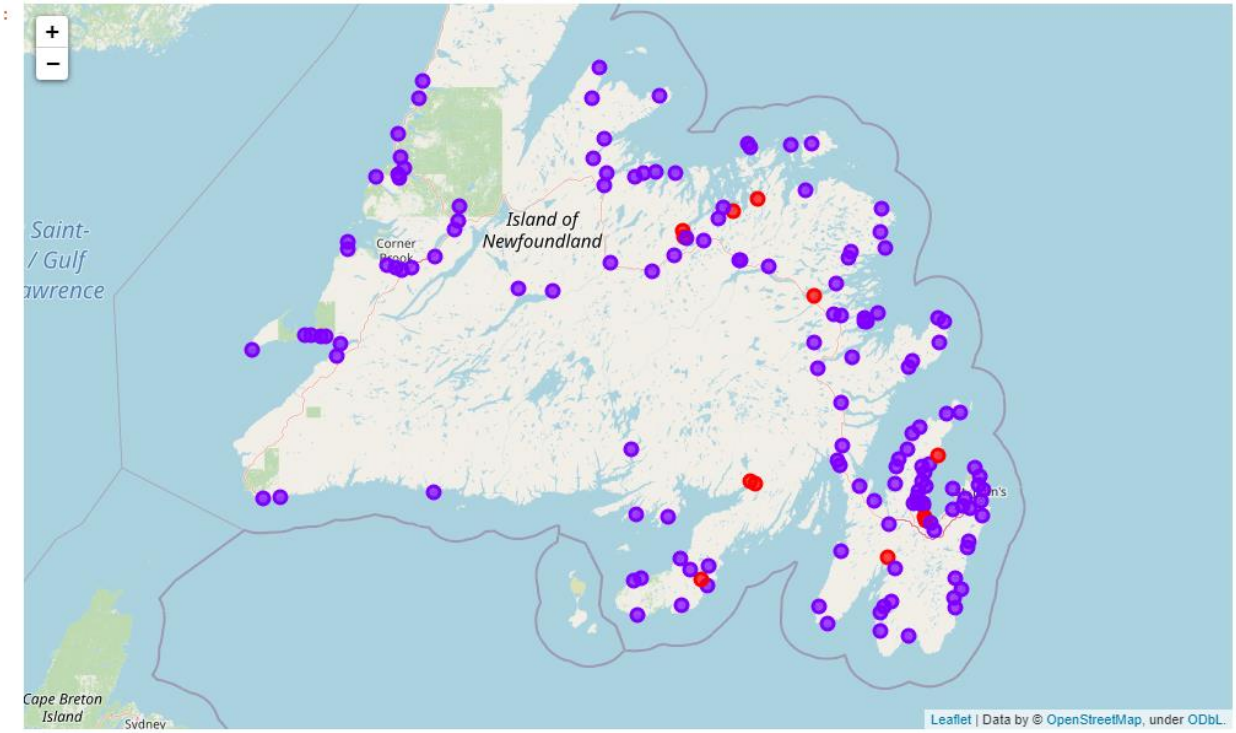

Figure 8: Clusters of town/cities

Based on the results above, it is seen that the clusters are segmented based on certain type of venue that are most common in that each neighborhood. Some observations can be made from each cluster's characteristics.

In Cluster 1, the most common venues are convenience stores, ATMs, and fishing spots. This doesn't leave much in terms of things to do and see in this cluster of towns.

In Cluster 2, the most common venues are coffee shops, Restaurants, and Marinas. This cluster provides many more things to visit, such as Cafes, Bakeries, and diners.

V.Conclusion

In this project, the main goal is to give an insight of the most common amenities that can be found in each town and city in Newfoundland, as well as the similarities between the town/cities in terms of the available amenities.

The data sources were identified to obtain the list of town/cities in Newfoundland, area, and geographical coordinates. These information are used to enquire 100 most common venues in each town/city using Foursquare API. K-means clustering is then used to perform clustering of the towns/city to two main clusters.

An analysis of the results shows that cluster 1 doesn't have much in terms of reasons to visit. The biggest tourist attraction found were fishing spots. Cluster 2 presents more dining places where the most common venues are coffee shops and cafes. These findings can be a good starting point for an individual/family to decide on a suitable town/city to stay in Newfoundland.

Future work can investigate in incorporating other factors such as the availability of transport facilities, schools, population density and housing prices to this study.