The Battle of Neighborhoods | Business Proposal | Introduction

Introduction:

Since, lots of people are migrating to various states of Canada due to various reasons and they needs lot of research for accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, good housing prices and schools for their children. People are looking for better neighborhoods also.

This Project aim to create an analysis of neighborhood to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toranto.

Problem:

1. Determine the cluster label for each capital
2. Visualize the cluster label in map
3. Determine the most common venue

Data Section

Datalink: https://en.wikipedia.org/wiki/List_of_cities_in_Canada Will use Scarborough dataset which we scrapped from wikipedia on Week 3.

we have csv dataset consisting of latitude and longitude, zip codes in capstone.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within

an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

3. Methodology Section

Clustering Approach:

To compare the similarities of two cities, we decided to explore venues, segment them, and group them into clusters to find similar capitals. To be able to do that, we need to cluster data this is a form of unsupervised machine learning: k-means clustering algorithm.

Work Flow:

Using credentials of Foursquare API features of near-by places of the capital city would be mined. Due to http request limitations the number

of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

The canada capital city was segmented into five clusters and was plotted in map to show the similar cluster group with the same color.

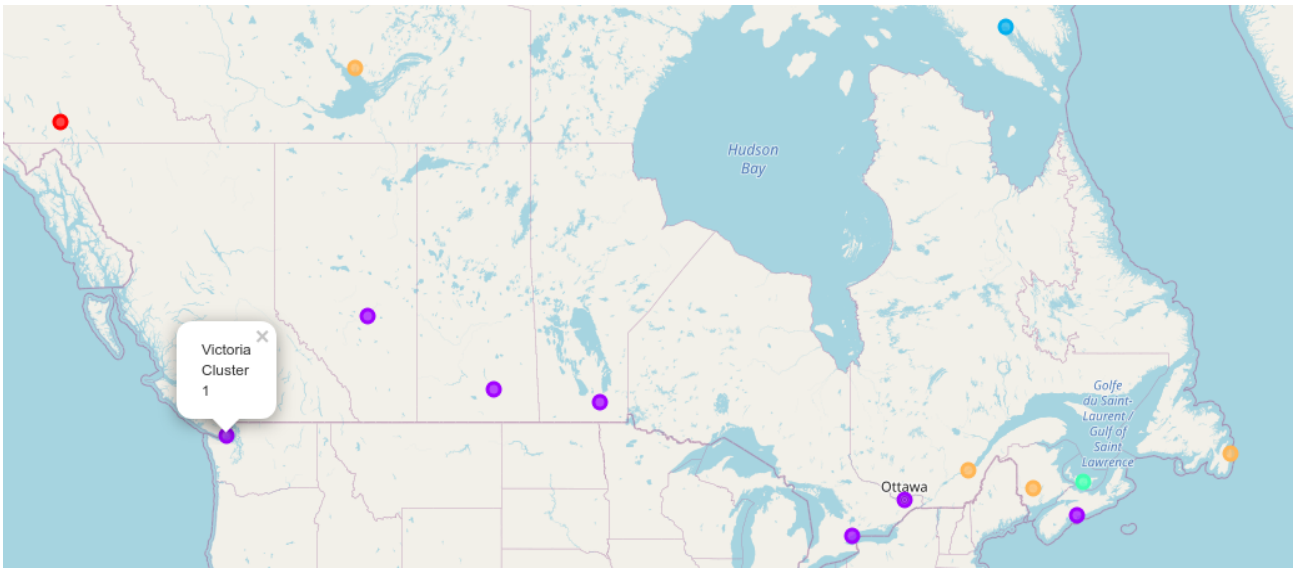

figure: segmented map of canada

Number of capital city in cluster 0, cluster 1, cluster 2, cluster 3 and cluster 4 was found to be 1 , 7, 1, 1, and 4 respectively.

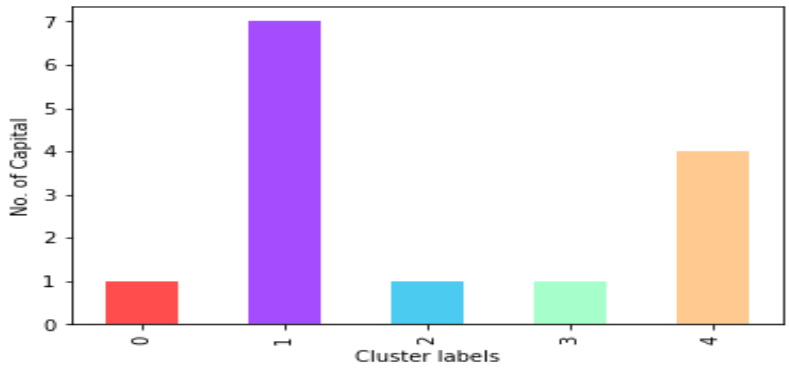

figure: bar-plot of cluster vs no. of capital

The most common venue found in cluster 0, cluster 1, cluster 2, cluster 3 and cluster 4 was coffee shop, coffee shop, coffee shop, Fast Food Restaurant, Restaurant.

5. Discussion Section

The major purpose of this project is to suggest visitor in Canada about the Capital city of each state and showing them which capital city are similar to other in terms of the venues located nearby.

Problem which i Tried to Solve:

1) Determine the cluster label for each capital.

2) Visualize the cluster label in the map.

3) Determine most common venue in each cluster.

Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 5 different clusters. And by locking the most common place in each cluster I came to know that pizza shop, coffee shop are very common and a nice startup business to anyone who wants to start their own business in canada.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.