Analyzing Neighborhoods in **Alberta** for starting a restaurant

IBM Data Science Professional Certificates

Introduction

- Alberta is considered as one of the most diversified cities in the canada. It is also the center of the nation's film and television industry. Alberta also has a diverse economy and hosts businesses in a broad range of professional and cultural fields, it has become one of the world's premier travel destinations, attracting more than 50 million visitors in 2018. Thus, alberta will be the best place to open a restaurant to welcome people from all over the world. The aim of this project is to study the neighborhoods in alberta in order to decided best locations for opening a restaurant by collecting and analyzing appropriate data.

Data Collection

- **Venue Data**
  - Foursquare API location data will be used in this project to help solving the problems. Foursquare API provids venue recommendations for all neighborhoods in Alberta.

- **Neighborhoods Data**
  - The data of the neighborhoods in Los Angeles County was scraped from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

- **Geographical Coordinates**
  - The geographical coordinates for Alberta data has been obtained from the GeoPy library in python. This data is relevant for plotting the map of Alberta using the Folium library in python. The geocoder library in python has been used to obtain latitude and longitude data for various neighborhoods in Alberta.

Methodology

Number of Neighborhoods Grouped by Location in Albe
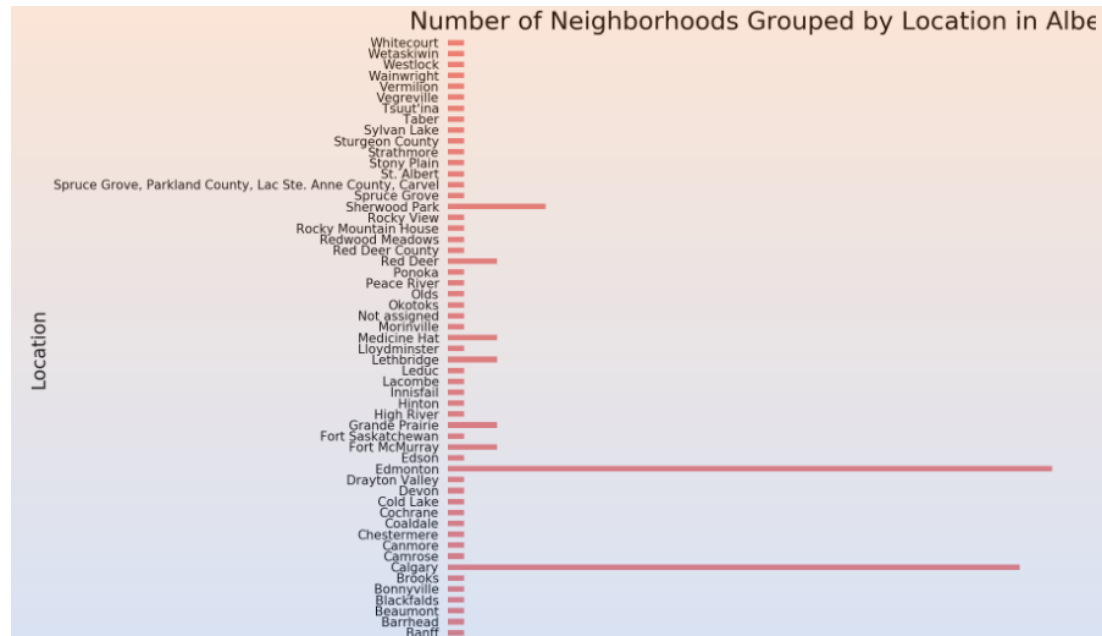
- First, we have collect the list of postal codes in Alberta, although the data we gained included latitude and longitude information, but we still reconfirmed these coordinates using Geocoder in order to get final dataframe.

- Second, in order to understand data better, we can plot this data and realized Sherwood Park and, Edmonton and Calgary have the greatest number of neighborhoods. We can also visualized the neighborhoods on a map using **Folium**, we will gained the geographical coordinates of Alberta by using **GeoPy**

Methodology

- Third, we will use Foursquare API to gain venue recommendations. We will first create credentials on the Foursquare API website and start by exploring the first neighborhood dataframe using Foursquare API get the neighborhood is Central Medicine Hat. Then we will extract the top 200 venues near Central Medicine Hat with a radius of 1000m. We will use **GET method** to get our results.

Methodology

Most Common Venue Types in 6 Prioritized Neighborhoods
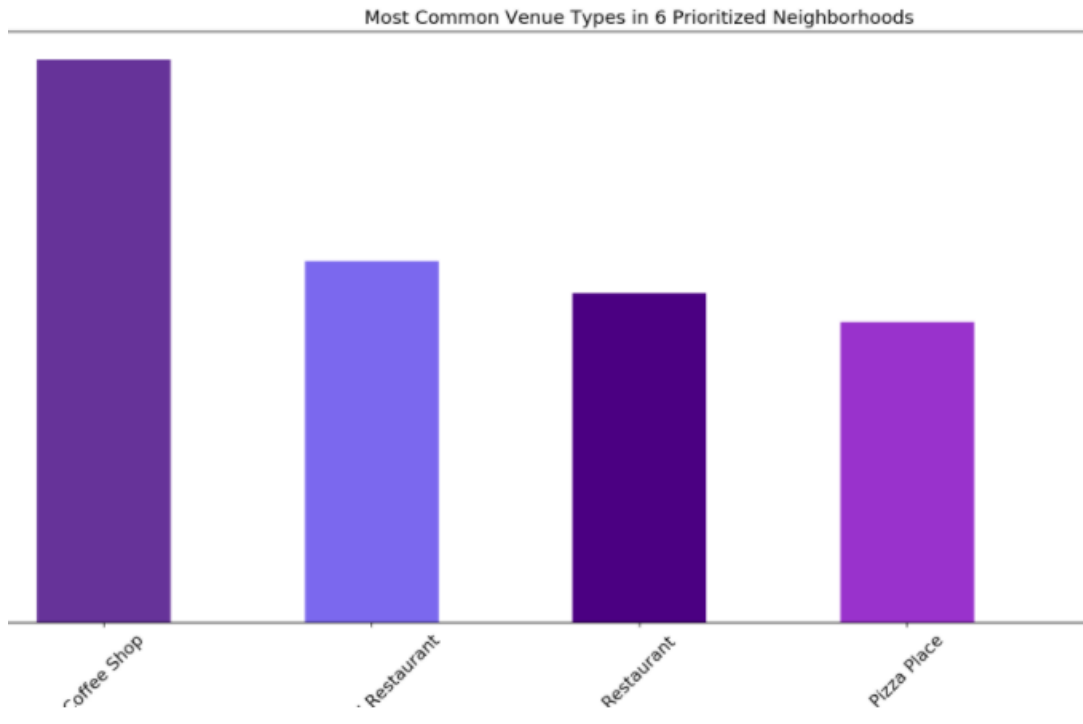
- Forth, we can start to generalize **foursquare API** to get nearby venues and how many venues were returned for each neighborhood. Then will check unique categories and display the most common venue categories in 6 prioritized neighborhoods which are Coffee shop, Restaurant, Pizza place and Grocery store.

Methodology

- Fifth, we can start to analyze each neighborhood by **One-hot Encoding** to see which categories belong in which neighborhoods and we able to create a dataframe with the top 10 common venues for each neighborhood as our final dataframe.

- Sixth, we will use **K-Means clustering method** to cluster the neighborhoods. We will first display how many clusters to use by using the **Silhouette Score.** We can also display the scores for different number of clusters and plot the data.

- Last, we will create a dataframe that includes cluster labels along with top 10 venues and visualized the clustering by creating a map using **Folium**.

Results and Discussion

- We have already pulled data on every neighborhood in Alberta and used the data to narrow down our neighborhood in each cluster and their top 10 most common venues.

- Now we have the results from five clusters gained. Cluster 1 and Cluster 3 are more suited for restaurants while cluster 2, cluster 4 and cluster 5 are less suited. Cluster 2 include too many other venues like schools and factories. Cluster 1 and cluster 3 include higher degree of coffee shop, restaurant, hotels, supermarket and plaza complex. Thus, the neighborhoods in cluster 1 and cluster 3 will be the best area for opening a restaurant.

- Compare with cluster 1 and cluster 3, cluster 3 is a better place to open a restaurant. Cluster 1 include a great number of schools, farmers market, electronics store and hotels, while cluster 3 contains less store required smaller space. As a result, Cluster 1 would be the best place to opening a restaurant.

Conclusion

- As a conclusion, we have completed this project by using common python and machine learning libraries and packages to manipulated data sets we gained from Wikipedia, Foursquare API to explore the neighborhoods of Alberta, Folium and K-Means to cluster and segment neighborhoods.

Thank you

- After the project, I realized theses tools like K-Means will helping people making effective decisions across different industries such as financial, energy, and healthcare industry. I am looking forward to exploring more and more models and tools from python and machine leaning to assist improved my decision making.