Capstone Project - The Battle of Neighbourhoods

Table of Contents:

1.Introduction

1.1 Background

Edmonton is the capital city of the Canadian province of Alberta. It is the second largest city and Canada's fifth largest municipality. Edmonton being the major economic centre for northern and central Alberta, is a favourable city to start a new in demand business.

1.2 Business Problem

As Edmonton is one of the highly populous cities there might be gas stations located across its neighbourhoods. This project aims to predict the neighbourhoods suitable to open a gas station in Edmonton.

2. Data acquisition and cleaning

2.1 Data Sources

Since the requirement is to find the neighbourhoods, it can be found in

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

FOURSQUARE API can be used to get the information about the available venues present in the neighbourhoods.

2.2 Data Cleaning

Data downloaded from the above sources and stored in a table.

There are some neighbourhoods without latitude and longitude positions in the dataset. So, removed those neighbourhoods from the table.

3. Methodology

In order to extract information about the venues present in different neighbourhoods, assumption is to have them clustered. I chose to use "K-Means Clustering Algorithm". K-means is a type of unsupervised learning, which is generally used for unlabelled data (data without defined categories or groups).

K-means groups each data point present in the data, given by the number K. It iteratively assigns each data point to one of K groups based on the features of that data point.
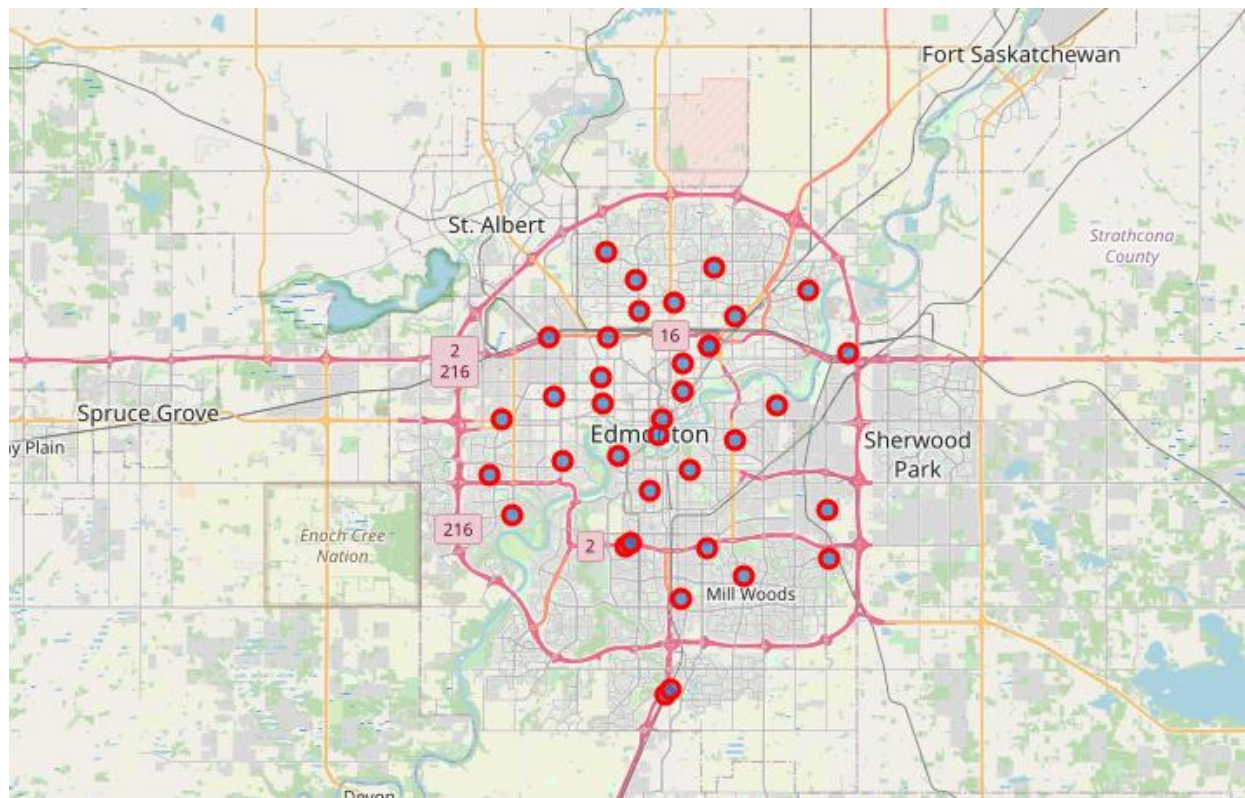
3.1 Analysis

Steps involved during analysis:
- Download and explore Dataset:

Information about the neighbourhood location were downloaded, cleansed.
- Create a map of Edmonton with neighbourhoods superimposed on top:

This is to visualize the neighbourhood location present in Edmonton.

- Explore neighbourhoods in Edmonton to get details of the venues using FOURSQUARE API:

This is to get the details about the venues present in every neighbourhood of Edmonton

- Analyse every neighbourhood:

Group the neighbourhoods depending on the frequency of each category venues.

- Cluster neighbourhoods using K-Means:
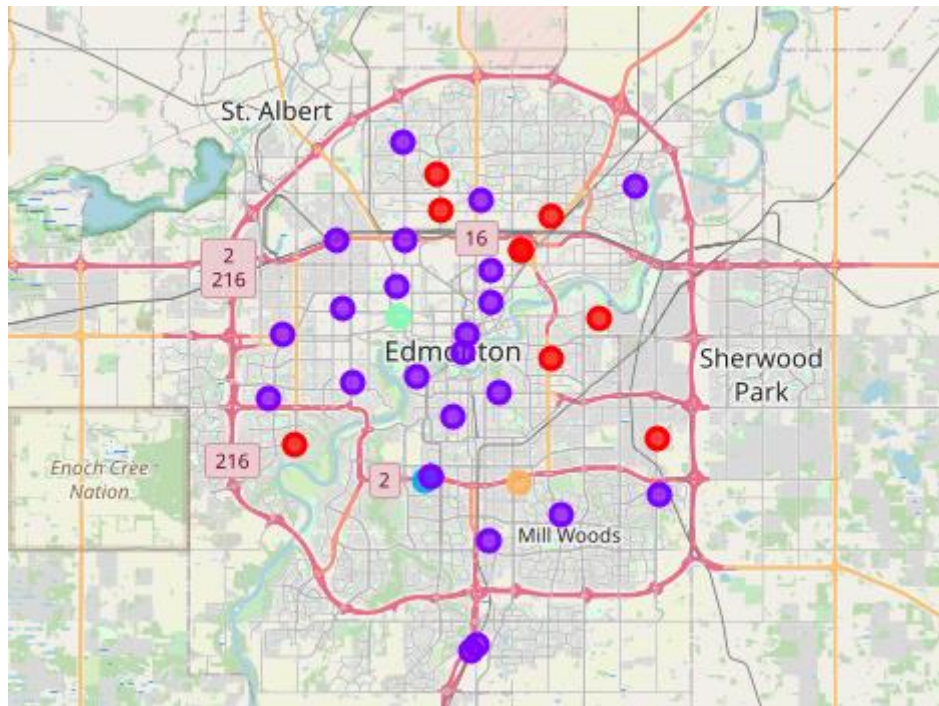
Clustered neighbourhoods into 5 categories.

- Examine Clusters:

This is to get the list of neighbourhoods with top 10 venues for each neighbourhood.

4. Results

Gas stations are more common in the neighbourhoods segmented to cluster 2. However, frequency of Gas Station in other clusters 1,3,4,5 is insignificant.

5. Discussion

In this project, I tried to predict the suitable neighbourhood for a new gas station in Edmonton. I took the data about the neighbourhoods along with its locations present in Edmonton. I used K-Means clustering algorithm to segment the neighbourhoods and used FOURSQUARE API to get the venue details present in those neighbourhoods. Finally, able to cluster the neighbourhoods with most frequent top 10 venues.

6.Conclusion

Model in this study focused on finding the neighbourhoods with less number of gas stations. However, there would be other factors that might also contribute to open a new gas station such as land availability, prices, proximity to public places, population etc.