# Capstone Project - The Battle of Neighborhoods

*Anna Dudareva*

## 1. Introduction

### 1.1. Background

Edmonton is the capital city of the Canadian province of Alberta. The current metro area population of Edmonton in 2020 is 1,461,000, a 2.17% increase from 2019, that makes it one of the most growing cities in the whole Canada. Moreover, Edmonton is a multinational city which arouses increased interest among investors. Having a part number of firms, technologies in various areas and industries located in and around Edmonton, this city attracts many people from different parts of Canada and rest of the world. Due to increasing population in the city, there is a large scope to set up businesses like shopping stores, hotels and hostels, restaurants, coffee shops, beauty and health shops, etc.

### 1.2. Business problem

The objective of this project is to analyze the neighbourhoods in Edmonton and divide them into different clusters based on the popular venues at each neighbourhood, by using data science methodology and machine learning techniques like clustering.

### 1.3 Interest

The target audience of this project is small business owners who want to set up their businesses like restaurants, coffee shops, beauty and health stores, etc. It would help them:

- to find out what kind venues already exist
- to find the optimal location based on the category of their business

## 2. Data requirements

### 2.1. Data source

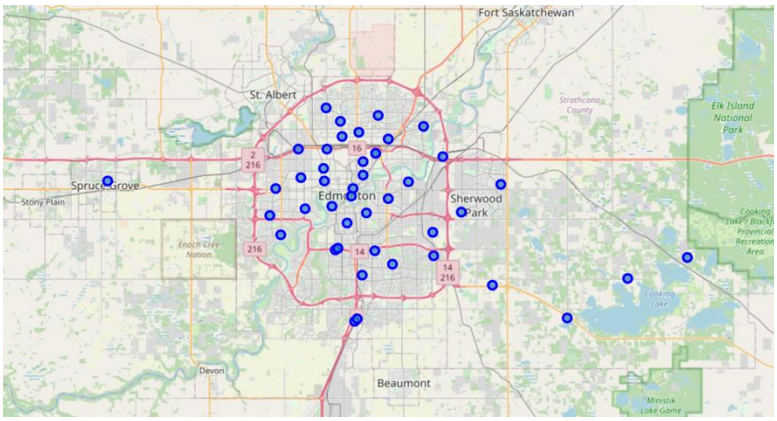
To achieve the project aim I will need the following data sets:

- A ready dataset with Borough name, Neighborhoods, Post Code and GPS coordinates of Alberta, Canada can be found on Wikipedia website: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T. Furthermore, sorting information about neighborhoods only for Edmonton city

| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | T1A | Medicine Hat | Central Medicine Hat | 50.036460 | -110.679250 |
| 1 | T2A | Calgary | Penbrooke Meadows, Marlborough | 51.049680 | -113.964320 |
| 2 | T3A | Calgary | Dalhousie, Edgemont, Hamptons, Hidden Valley | 51.126060 | -114.143158 |
| 3 | T4A | Airdrie | East Airdrie | 51.272450 | -113.986980 |
| 4 | T5A | Edmonton | West Clareview, East Londonderry | 53.5899 | -113.4413 |

Image 1. Alberta neighborhoods dataset

- The Foursquare API was used to access the venues in the neighborhoods: www.foursquare.com
- Then neighborhoods will be clustered based on their venues using Data Science Techniques, namely k-means algorithm, silhouette coefficient and plots
- Folium library can be used to visualize the clusters on the map of Edmonton




Map 1. Edmonton city with neighborhoods

## 2.2. Data cleaning

Data cleaning and pre-processing involve scraping data of Alberta region and then of Edmonton city with all neighborhoods. Further, extracting directly neighborhoods of Edmonton and scraping venue items. All the extracted data was converted in to data frame for further analysis. There were a lot of missing values, so I decided to drop missing data.

In Image 1 you see the data from Foursquare API.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | West Clareview, East Londonderry | 53.5899 | -113.4413 | Café del Sol | 53.592441 | -113.441455 | Mexican Restaurant |
| 1 | West Clareview, East Londonderry | 53.5899 | -113.4413 | Buffet Royale Carvery | 53.587229 | -113.439075 | Buffet |
| 2 | West Clareview, East Londonderry | 53.5899 | -113.4413 | Red Claw Gaming | 53.586937 | -113.439775 | Toy / Game Store |
| 3 | West Clareview, East Londonderry | 53.5899 | -113.4413 | My Grandma's Attic | 53.586033 | -113.441629 | Record Shop |
| 4 | West Clareview, East Londonderry | 53.5899 | -113.4413 | Belvedere Transit Centre | 53.587932 | -113.435254 | Bus Station |

Image 2. Venues and venue categories in each neighborhood.

## 3. Methodology

I have the neighborhoods data of Edmonton, in total 38 neighborhoods. I also have the most popular venues in each neighborhood obtained using Foursquare API. A total of 309 venues have been obtained in the whole city and 123 unique categories.

Due to multiple neighborhoods with less than 5 venues were returned, I considered to use only the neighborhoods with more then 5 venues, in order to make a better analysis.

Using one hot encoding I can find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset and K - Nearest Neighbor clustering technique have been used. To find the optimal number of clusters silhouette coefficient is used.

The clusters can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest business people suitable locations based on the category.

## 4. Analysis

4.1 As I mentioned in Methodology part, due to multiple neighborhoods with less than 5 venues, I decided to use only those which have more than 5 venues. Thereby, I have visualized those neighborhoods.

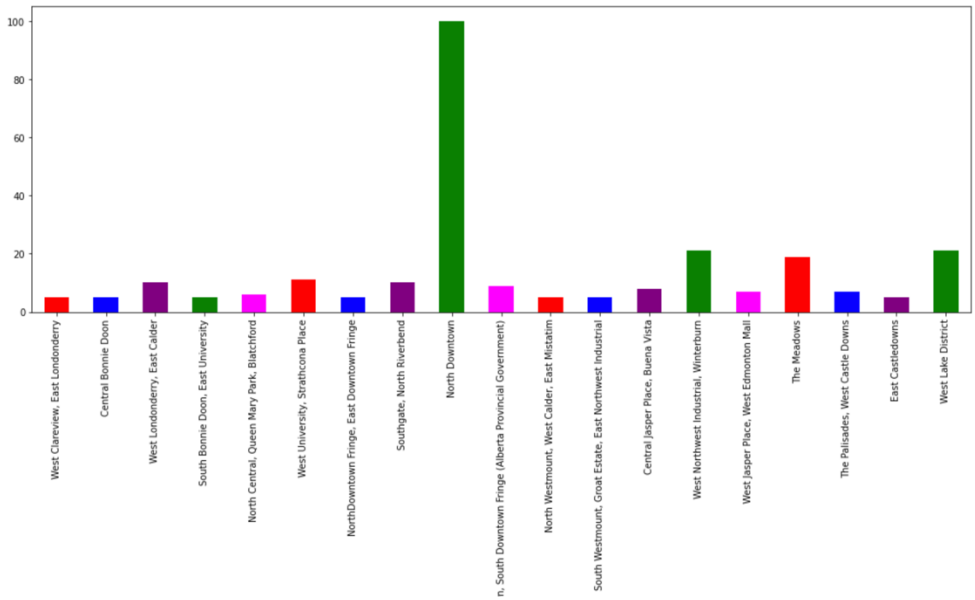



Image 3. Neighborhoods with more than 5 venues

4.2 Next was one hot encoding on the filtered data to obtain the venue categories in each neighborhood. Then group the data by neighborhood and take the mean value of the frequency of occurrence of each category. The outcome is shown below.

| | Neighborhood | American Restaurant | Arts & Crafts Store | Asian Restaurant | Auto Workshop | Bakery | Bank | Bar | Baseball Field | Baseball Stadium | Big Box Store | Bookstore | Breakfast Spot | Brewery | Buffet | Burger Joint | Bus Station | Business Service | But |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Central Beverly | 0.0 | 0.0 | 0.0 | 0.0 | 0.000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 1 | Central Bonnie Doon | 0.2 | 0.0 | 0.0 | 0.0 | 0.000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 2 | Central Jasper Place, Buena Vista | 0.0 | 0.0 | 0.0 | 0.0 | 0.125 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 3 | Central Mistatim | 0.0 | 0.0 | 0.0 | 0.0 | 0.000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 4 | East Castledowns | 0.0 | 0.0 | 0.0 | 0.0 | 0.000 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |

Image 3. Mean of frequency of occurrence of each category

4.3 One hot encoding was performed to obtain the 10 most common venues in each neighborhood, then stored these data in a new dataframe. A sample is shown below.

| | Neighborhood | 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 Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Central Beverly | Smoke Shop | Construction & Landscaping | Grocery Store | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 1 | Central Bonnie Doon | American Restaurant | Water Park | Trail | Liquor Store | Electronics Store | Falafel Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant |
| 2 | Central Jasper Place, Buena Vista | Sandwich Place | Sushi Restaurant | Pizza Place | Convenience Store | Salad Place | Bakery | Café | Fast Food Restaurant | Flower Shop | Falafel Restaurant |
| 3 | Central Mistatim | Warehouse Store | Casino | Liquor Store | Wine Shop | Falafel Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Electronics Store |
| 4 | East Castledowns | Plaza | Recreation Center | Liquor Store | Skating Rink | Food Truck | Department Store | Diner | Discount Store | Distribution Center | Dog Run |

Image 4. Top 10 common venues in each neighborhood

This dataset was used for the clustering algorithm, namely K-Nearest Neighbor (KNN) clustering was used. It is an unsupervised machine learning technique that clusters the given data into K number of clusters. For optimal result I had to select the best value of K. So, the silhouette coefficient was used to find the best value for K. A range of value from 2 to 10 was considered. A range of coefficients is demonstrated in the table below. As we can see, n_clusters=3 has highest Silhouette Coefficient. This means that 3 should be the optimal number of clusters.

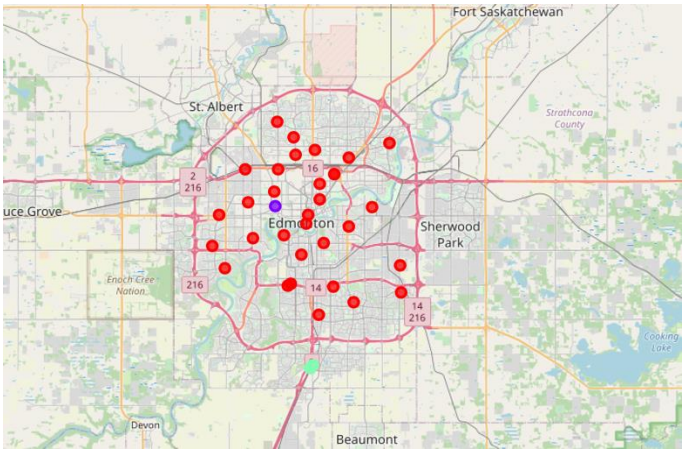
```
print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))
```

```
For n_clusters=2, The Silhouette Coefficient is 0.3720631959243243
For n_clusters=3, The Silhouette Coefficient is 0.37746557374507167
For n_clusters=4, The Silhouette Coefficient is 0.20421151393585438
For n_clusters=5, The Silhouette Coefficient is 0.21701449692631913
For n_clusters=6, The Silhouette Coefficient is 0.1518278010326688
For n_clusters=7, The Silhouette Coefficient is 0.1929511880723329
For n_clusters=8, The Silhouette Coefficient is 0.1808537468980641
For n_clusters=9, The Silhouette Coefficient is 0.1718769853636927
```

Image 5. The Silhouette Coefficient results

## 5. Results

Let's examine the three clusters and find the discriminating venue categories that identify each cluster. For this purpose, let's have a look into 10 most common venue category in each cluster.




Map 2. Edmonton map with clusters: red-cluster 1, violet-cluster 2, green-cluster 3

### 5.1 Cluster 1

The top venue categories in Cluster 1 are Fast Food Restaurant, Discount Store, Eastern European Restaurant, Distribution Center, Dog Run, Flower Shop, Diner, Electronics Store, Food & Drink Shop, Falafel Restaurant.

| | Neighborhood | Clusters | 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 Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | West Clareview, East Londonderry | 0.0 | Bus Station | Buffet | Mexican Restaurant | Record Shop | Toy / Game Store | Fast Food Restaurant | French Restaurant | Food Truck | Food & Drink Shop | Flower Shop |
| 1 | North Capilano | 0.0 | Bus Station | Playground | Ski Trail | Department Store | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 2 | East North Central, West Beverly | 0.0 | Smoke Shop | Construction & Landscaping | Grocery Store | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 3 | SE Capilano, West Southeast Industrial, East B... | 0.0 | Business Service | Home Service | Playground | Bar | Falafel Restaurant | French Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant |
| 5 | Central Bonnie Doon | 0.0 | American Restaurant | Water Park | Trail | Liquor Store | Electronics Store | Falafel Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant |
| 6 | West Londonderry, East Calder | 0.0 | Shopping Mall | Arts & Crafts Store | Hockey Arena | Butcher | Bakery | Dog Run | Baseball Field | Comic Shop | Grocery Store | Recreation Center |
| 7 | South Bonnie Doon, East University | 0.0 | American Restaurant | Pharmacy | Coffee Shop | Mediterranean Restaurant | Flower Shop | Fried Chicken Joint | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant |
| 8 | North Central, Queen Mary Park, Blatchford | 0.0 | Pharmacy | Café | Bakery | Bank | Music Venue | Theater | Fast Food Restaurant | Food Truck | Food & Drink Shop | Flower Shop |
| 9 | West University, Strathcona Place | 0.0 | College Gym | Theater | Diner | College Residence Hall | Fast Food Restaurant | Coffee Shop | Pub | Bank | Sandwich Place | East Europe Restaurant |
| 10 | NorthDowntown Fringe, East Downtown Fringe | 0.0 | Grocery Store | Café | Soccer Stadium | Gift Shop | Gym | Falafel Restaurant | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Wine Shop |
| 11 | Southgate, North Riverbend | 0.0 | Bus Station | Distribution Center | Restaurant | Sandwich Place | Furniture / Home Store | French Restaurant | Diner | Discount Store | Dog Run | Eastern European Restaurant |
| 12 | North Downtown | 0.0 | Coffee Shop | Sandwich Place | Pub | Restaurant | Italian Restaurant | Café | Fast Food Restaurant | Hotel | Nightclub | New American Restaurant |
| 13 | Kaskitayo, Aspen Gardens | 0.0 | Lake | Wine Shop | Furniture / Home Store | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store | Falafel Restaurant |
| 14 | South Downtown, South Downtown Fringe (Alberta... | 0.0 | Park | Thai Restaurant | Baseball Stadium | Sandwich Place | Museum | Gift Shop | French Restaurant | Hotel | Eastern European Restaurant | Electronics Store |
| 15 | West Mill Woods | 0.0 | Soccer Field | Construction & Landscaping | Gym / Fitness Center | Wine Shop | Falafel Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Eastern European Restaurant |
| 16 | North Westmount, West Calder, East Mistatim | 0.0 | Furniture / Home Store | Middle Eastern Restaurant | Auto Workshop | Pub | Breakfast Spot | Fast Food Restaurant | French Restaurant | Food Truck | Food & Drink Shop | Flower Shop |
| 17 | East Mill Woods | 0.0 | Pub | Bakery | Wine Shop | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 18 | South Westmount, Groat Estate, East Northwest ... | 0.0 | Pizza Place | Coffee Shop | Discount Store | Bank | Gym | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Falafel Restaurant |
| 19 | Southwest Edmonton | 0.0 | Home Service | Indian Restaurant | Paintball Field | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 21 | South Industrial | 0.0 | Park | Gym | Golf Driving Range | Wine Shop | French Restaurant | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant |
| 22 | North Jasper Place | 0.0 | Convenience Store | Furniture / Home Store | Rental Car Location | Liquor Store | Wine Shop | Electronics Store | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Falafel Restaurant |
| 23 | East Southeast Industrial, South Clover Bar | 0.0 | Housing Development | Convenience Store | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store | Falafel Restaurant |
| 24 | Central Jasper Place, Buena Vista | 0.0 | Sandwich Place | Sushi Restaurant | Pizza Place | Convenience Store | Salad Place | Bakery | Café | Fast Food Restaurant | Flower Shop | Falafel Restaurant |
| 25 | Southgate, North Riverbend | 0.0 | Bus Station | Distribution Center | Restaurant | Sandwich Place | Furniture / Home Store | French Restaurant | Diner | Discount Store | Dog Run | Eastern European Restaurant |
| 26 | West Northwest Industrial, Winterburn | 0.0 | Fast Food Restaurant | Hotel | Vietnamese Restaurant | Coffee Shop | Grocery Store | Arts & Crafts Store | Chinese Restaurant | Pizza Place | Ice Cream Shop | Indian Restaurant |
| 28 | West Jasper Place, West Edmonton Mall | 0.0 | Pharmacy | Health & Beauty Service | Photography Studio | Pub | Business Service | Asian Restaurant | Hot Dog Joint | Diner | Discount Store | Distribution Center |
| 29 | The Meadows | 0.0 | Restaurant | Coffee Shop | Fast Food Restaurant | Pharmacy | Salad Place | Clothing Store | Fried Chicken Joint | Discount Store | Gas Station | Bus Station |
| 30 | Central Mistatim | 0.0 | Warehouse Store | Casino | Liquor Store | Wine Shop | Falafel Restaurant | Food Truck | Food & Drink Shop | Flower Shop | Fast Food Restaurant | Electronics Store |
| 31 | The Palisades, West Castle Downs | 0.0 | Restaurant | Convenience Store | Ice Cream Shop | Pharmacy | Halal Restaurant | Gas Station | Fast Food Restaurant | Wine Shop | Flower Shop | Falafel Restaurant |
| 32 | Central Beverly | 0.0 | Smoke Shop | Construction & Landscaping | Grocery Store | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 34 | East Castledowns | 0.0 | Plaza | Recreation Center | Liquor Store | Skating Rink | Food Truck | Department Store | Diner | Discount Store | Distribution Center | Dog Run |
| 36 | Horse Hill, East Lake District | 0.0 | Pharmacy | Coffee Shop | Food & Drink Shop | Wine Shop | Fried Chicken Joint | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 37 | West Lake District | 0.0 | Coffee Shop | Fast Food Restaurant | Pharmacy | American Restaurant | Supermarket | Pizza Place | Chinese Restaurant | Sandwich Place | Furniture / Home Store | Liquor Store |

### 5.2 Cluster 2

The top venue categories in Cluster 2 are Portuguese Restaurant, Wine Shop, Fried Chicken Joint, Dinner and Discount Store.

| | Neighborhood | Clusters | 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 Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | Glenora, SW Downtown Fringe | 1.0 | Portuguese Restaurant | Wine Shop | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store | Falafel Restaurant |

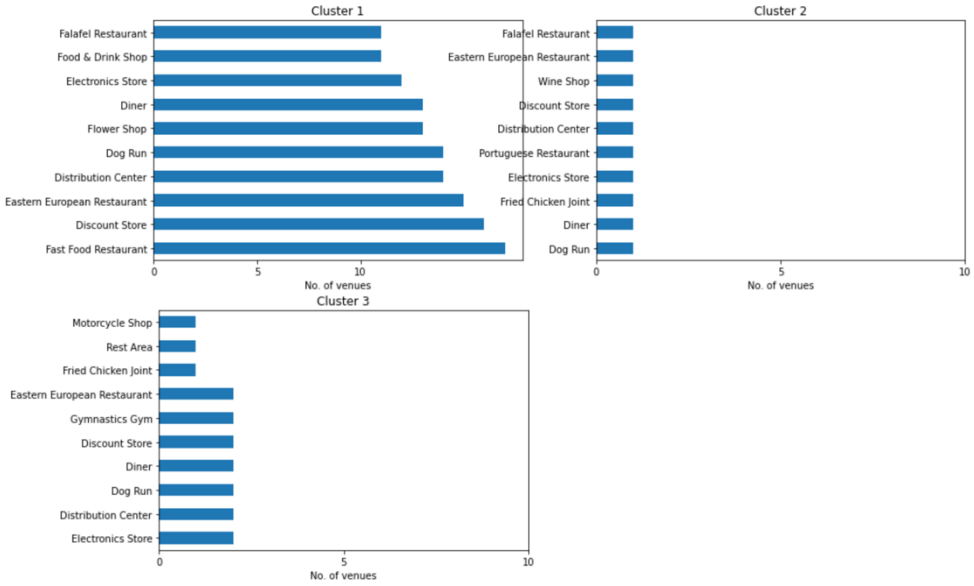
### 5.3 Cluster 3

The top venue categories in Cluster 3 are Gymnastics Gym, Motorcycle Shop, Rest Area, Wine Shop, Home Shop, Fried Chicken Joint, Diner and Discount Store.

| | Neighborhood | Clusters | 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 Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | Heritage Valley | 2.0 | Gymnastics Gym | Rest Area | Wine Shop | Fried Chicken Joint | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store |
| 35 | Ellerslie | 2.0 | Motorcycle Shop | Gymnastics Gym | Furniture / Home Store | Diner | Discount Store | Distribution Center | Dog Run | Eastern European Restaurant | Electronics Store | Falafel Restaurant |

## 6. Discussion and future directions

Now that we have the clusters and the top 10 most common venue categories, let’s visulaize them in each cluster for comparison.




This plot can be used to suggest valuable information to Business persons. Let's discuss a few examples considering they would like to start the following category of business:

- The cluster 1 is situated within the central part of the city, and is saturated with fast food restaurants and some other type restaurants, hence opening one here is not the best choice. I could suggest that this cluster don't need more restaurants, but hotel/hostel, health & beauty shop, or bakery could be a good idea for business.

- On the cotrary, cluster 3 is situated outside the central part of the city, and has one fast food restaurant. Suggestion could be a coffee shop or one more, but another type fast food restaurant. If a deeper analysis

could be done, due to location a major project could be suggested, as large shopping mall, indoor climbing park or amusement park for children etc.

- Cluster 2 is the smallest one and is located in central part of the city. The most popular venues there are food and wine, so may be cluster 2 is a good place for a bar, pub or disco.

In this project, I focus on the most common venues in the neighborhoods as well as the frequency of occurrence of different venues. In order to make an insightful data-driven business decision, there are other factors for consideration such as population in each neighborhood and income of residences, the concentration of commercial buildings that could influence the location decision.

Future research could devise a methodology to estimate mentioned above data to be used in the clustering algorithm to identify the prime locations for a new business. This project uses a free developer account of Foursquare API that came with limitations on how many API calls and results returned. Future research could use a paid account to bypass these limitations for better results.

## 7. Conclusion

This section concludes the predictive analysis for finding suitable location to open a new business in Edmonton. I have used different python packages to extract, scrap the necessary data from Wikipedia and to visualise data, also Foursquare API was used to explore the venues in neighborhoods. Final section of the project was to group venues in clusters and define the characteristics for that particular cluster. I have found three clusters, to of which are situated in the central part of the Edmonton, however the third cluster is located outside. A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. So all this information can be used by stakeholders to decide the location for the particular type of business.