# Coffee Crunch

Assessing Edmonton Neighbourhoods' Coffee Options

## Business Problem

Coffee is a $6.2 billion dollar industry with 160,000+ café/coffee shop jobs in Canada[1]. Access to a nearby coffee shop is a must for many caffeine crazed Canadians, which means business opportunities for entrepreneurs. Knowing which neighbourhoods are over or undersaturated with coffee options would provide immense value when deciding on the location of a new café. Furthermore, it would be useful to know whether those existing cafés are capturing the premium or inexpensive market segment, as an inexpensive Tim Hortons franchise and an upscale option would compete in different market segments.

Visualizing and clustering neighbourhoods based on coffee shop quantity and quality provides entrepreneurs a simple way to identify market opportunities, for example opening a coffee shop where there is a lack of existing ones or opening a premium café in a neighbourhood with only low-quality existing options.

## Data

The City of Edmonton's Open Data Portal provides geo coordinates of each neighborhood. The centroid of the neighbourhood is approximated by taking the average of the boundary point coordinates. [2]

Coffee shops within a 1km radius of each neighbourhood centroid will be used as a proxy for the cafés accessible from that neighbourhood. The number of coffee shops as well as the top rating of the neighbourhood's cafes is gathered from Foursquare's location API for each neighbourhood in the City of Edmonton.

## Methodology

The neighborhood latitude and longitude coordinates are imported as a pandas dataframe and mapped to check the accuracy. The number of local coffee shops within 1km as well as the Foursquare ID for the top result is then gathered using the 'venues/explore' Basic endpoint.

The ratings for each top neighbourhood coffee shop are obtained with the Premium 'venues/' endpoint. Blank/NaN values for the coffee shop ratings are set to '0' where there are no existing coffee shops while the others are set to the citywide average rating.

Exploratory data analysis reveals that there are 401 neighbourhoods, with 145 having no coffee shop ratings. 98 out of these 145 have no coffee shops while 47 have at least one coffee shop, but no ratings:

```
dfEdmonton.loc[dfEdmonton['Top Coffee Shop Rating'] == 0].describe()
```

```
topCoffeeShopRatings.count(np.nan)
```

| | Cluster Labels | Latitude | Longitude | # of Coffee Shops | Top Coffee Shop Rating |
|---|---|---|---|---|---|
| count | 98.0 | 98.000000 | 98.000000 | 98.0 | 98.0 |

```
145
```

[1] https://www.coffeeassoc.com/coffee-facts/
[2] https://data.edmonton.ca/Geospatial-Boundaries/City-of-Edmonton-Neighbourhood-Boundaries/jfvj-x253

The mean and median number of coffee shops per neighbourhood is 2.9 and 2 respectively, with a minimum of 0 and a maximum of 20. The 75th percentile neighbourhood has 4 coffee shops and only the top 7.5% (30/401) have greater than six coffee shops. This indicates a right-skewed distribution.

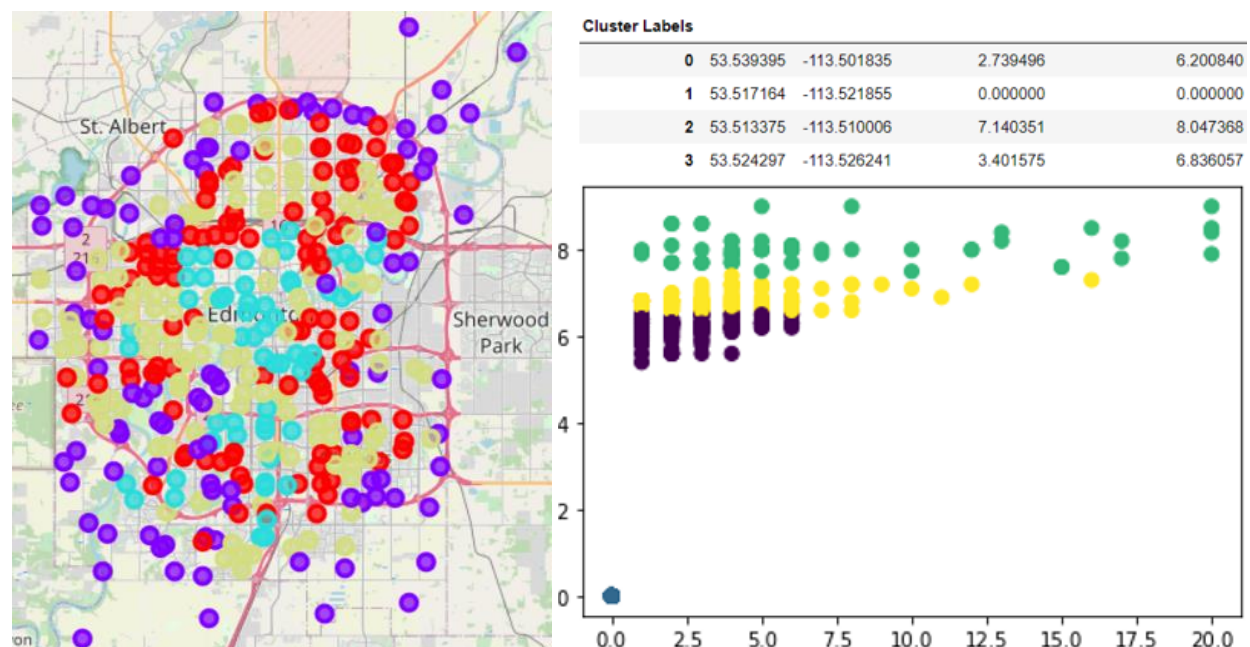|  | Cluster Labels | Latitude | Longitude | # of Coffee Shops | Top Coffee Shop Rating |
|---|---|---|---|---|---|
| count | 401.000000 | 401.000000 | 401.000000 | 401.000000 | 401.000000 |
| mean | 1.478803 | 53.525482 | -113.515619 | 2.905237 | 5.149076 |
| std | 1.216624 | 0.066111 | 0.088071 | 3.395732 | 2.994968 |
| min | 0.000000 | 53.365261 | -113.703122 | 0.000000 | 0.000000 |
| 25% | 0.000000 | 53.472776 | -113.584271 | 1.000000 | 5.600000 |
| 50% | 1.000000 | 53.526059 | -113.513574 | 2.000000 | 6.400000 |
| 75% | 3.000000 | 53.577381 | -113.447312 | 4.000000 | 6.814453 |
| max | 3.000000 | 53.684593 | -113.283408 | 20.000000 | 9.000000 |

```
dfEdmonton.loc[dfEdmonton['# of Coffee Shops'] > 6].describe()
```

|  | Cluster Labels | Latitude | Longitude | # of Coffee Shops | Top Coffee Shop Rating |
|---|---|---|---|---|---|
| count | 30.000000 | 30.000000 | 30.000000 | 30.000000 | 30.000000 |

K-means clustering is the machine learning application applied to find meaningful groups of neighbourhoods. Both the number and rating of the neighbourhood's coffee shops was tested but the clustering on only Top Coffee Shop Rating proved better for making useful insights. A k value of 4 was chosen as this level of clustering makes it easy to identify neighborhoods of opportunity and neighborhoods of oversaturation.

Once the clusters are determined the neighbourhood centroids are mapped, colour-coded with their respective clusters. A scatter plot of the clusters helps visualize the difference between the clusters.

## Results



| Cluster Labels |  |  |  |  |
|---|---|---|---|---|
| 0 | 53.539395 | -113.501835 | 2.739496 | 6.200840 |
| 1 | 53.517164 | -113.521855 | 0.000000 | 0.000000 |
| 2 | 53.513375 | -113.510006 | 7.140351 | 8.047368 |
| 3 | 53.524297 | -113.526241 | 3.401575 | 6.836057 |

Four clusters of neighbourhoods were determined, following a relatively well-defined pattern:

Cluster 1 (Blue/Purple) highlights neighbourhoods with no coffee shops at all. Most of these neighbourhoods lie on the suburban and rural low-density periphery of the city, however there are many central and higher density neighbourhoods represented here as well.

Cluster 0 (Red/Indigo) shows the neighbourhoods with existing coffee shop options, albeit of low quality. These neighbourhoods are dispersed across the city.

Cluster 3 (Yellow-Green/Yellow) shows neighbourhoods that have existing coffee shop options of middling quality. These are mostly located in the interior of the city but outside of the core.

Cluster 2 (Cyan/Green) shows neighbourhoods with an abundance of high-quality coffee shops. These neighbourhoods are mostly located in high density downtown/shopping areas of the city.

## Discussion

Cluster 1 (Blue/Purple) neighbourhoods represent good opportunities for any coffee shop to locate since they would face no existing competition. Care must be taken however to ensure there is high enough local density to support one.

Cluster 0 (Red/Indigo) neighbourhoods represent opportunities for up-scale and high-quality cafes to locate, as their competition is clearly not providing a quality coffee option.

Cluster 3 (Yellow-Green/Yellow) neighbourhoods offer some opportunity for new entrants of high-quality, however the competition faced in these neighbourhoods would be noticeably higher than in Cluster 0. Therefore Cluster 3 represents areas of lower opportunity compared to Clusters 1 and 0.

Cluster 2 (Cyan/Green) neighbourhoods are best avoided for coffee shop entrepreneurs as they are already saturated with good quality coffee options and so would face stiff local competition. Alternatively, they may represent areas of high density and coffee drinking, however that analysis is outside the scope of this research.

## Conclusion

Coffee shops are a staple of the urban economy. By leveraging location data from the City of Edmonton and Foursquare, café entrepreneurs can use this machine learning model is used to identify which neighbourhoods to open a new shop in. Using k-means clustering, neighbourhoods lacking existing quality options as well as over-saturated neighbourhoods are determined, revealing the most opportune neighbourhoods. Out of 4 neighbourhood clusters, Cluster 1 and Cluster 0 would make the best locations for new coffee shops as they offer either no existing coffee shops (Cluster 1), or coffee shops of low-quality (Cluster0).