Clustering Food Venues in Canadian Cities

Author: Lawrence Lee

Date: September, 2019

For IBM Data Science Professional Certificate

Background

- Food businesses are one of the most numerous and diverse small businesses in Canada.

- How do geographic areas within a city and between cities compare, with regard to concentrations of different categories of food businesses?

Data Sources

Canadian Postal Codes

- 261 Forward Sortation Areas (FSA) (e.g. H1M)

- Scraped from Wikipedia html tables

FourSquare Venues

- 5,798 Venues
- Food Categories
- Geo coordinates (lat, lng)

- Search Endpoint Called and data extracted from JSON

Venue Data Sample

Place	Place Latitude	Place Longitude	Name	Latitude	Longitude	Category
H0M	45.69	-73.50	Tenuta	45.69	-73.50	Italian Restaurant
H0M	45.69	-73.50	Dairy Queen Store	45.69	-73.49	Fast Food Restaurant
H0M	45.69	-73.50	Tim Hortons	45.69	-73.49	Coffee Shop

Method: Jupyter Notebook – Python 3.7.3

Core Modules
numpy
pandas
collections
requests
geopy.geocoders.Nominatim
matplotlib
sklearn.cluster.KMeans
folium

Debugging Modules
time
dill

Method: Exclusions

FSA

- Not Assigned
- Province of Alberta outside Calgary
- No corresponding venues

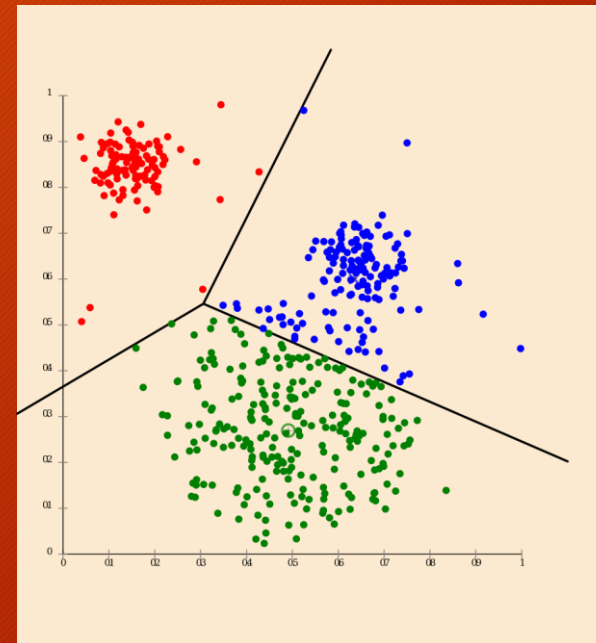## Venue

- Primary Category not Food
- Coffee Shops and Cafes

Method: KMeans Cluster

What is Clustering?

- Machine learning

- Segment data points into mutually exclusive clusters.

- Unsupervised: pre-existing labels are not provided. Inferences are made based on data similarities.
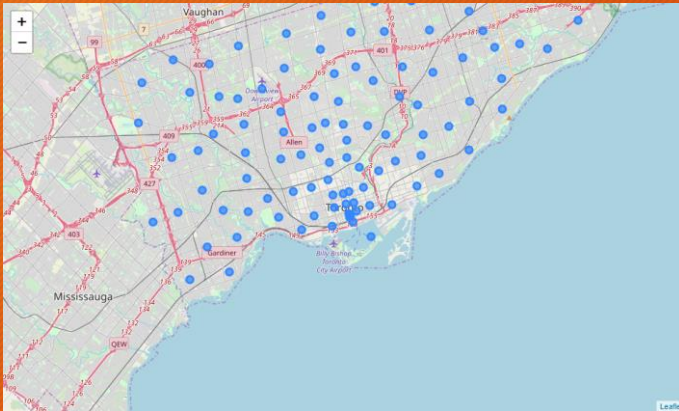
What is K-Means?

- Clustering algorithm

- Partition based

- Sphere-like clusters

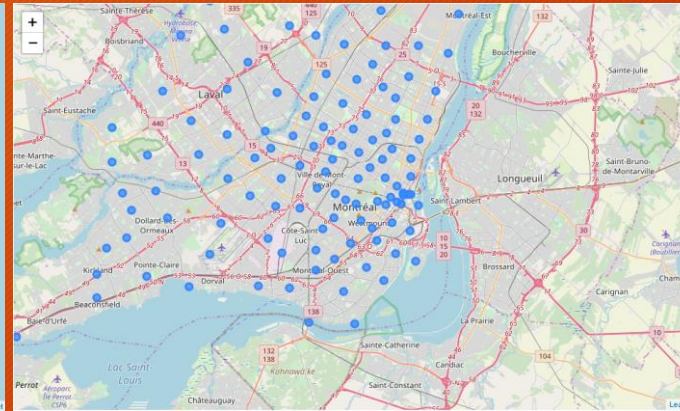- Relatively efficient for medium and large datasets.
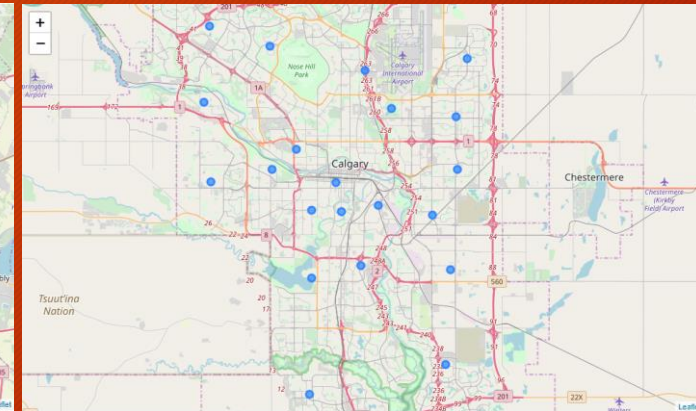
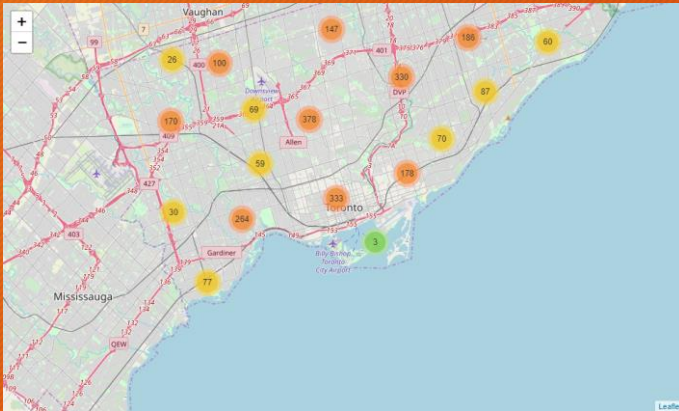# Results: FSA Maps

Toronto, Ontario

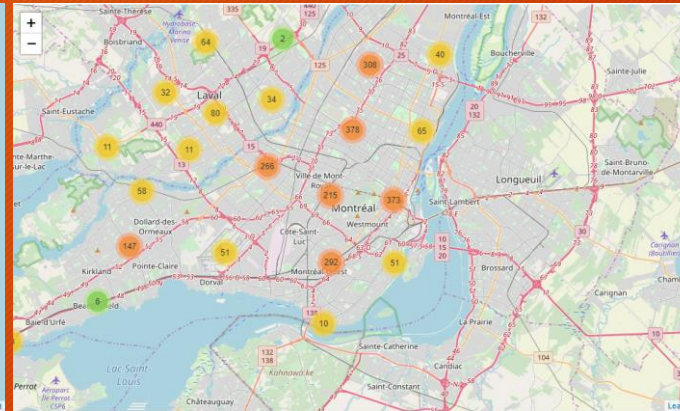### Montreal, Quebec

Calgary, Alberta

Results: Venue Maps
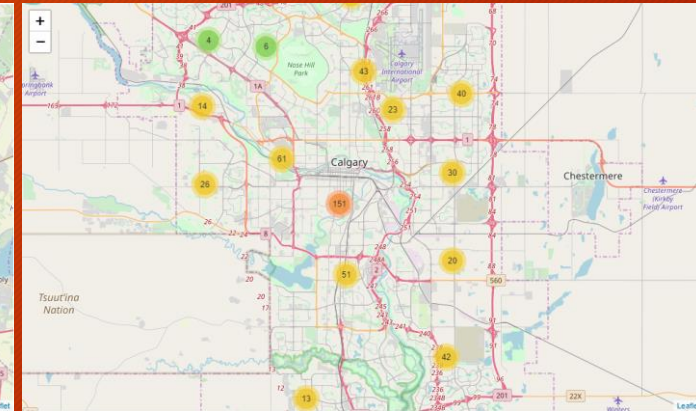
Toronto, Ontario

Montreal, Quebec

Calgary, Alberta
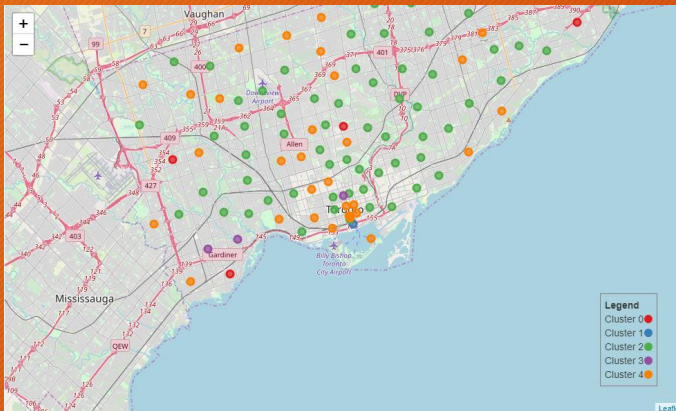
Results: Clusters

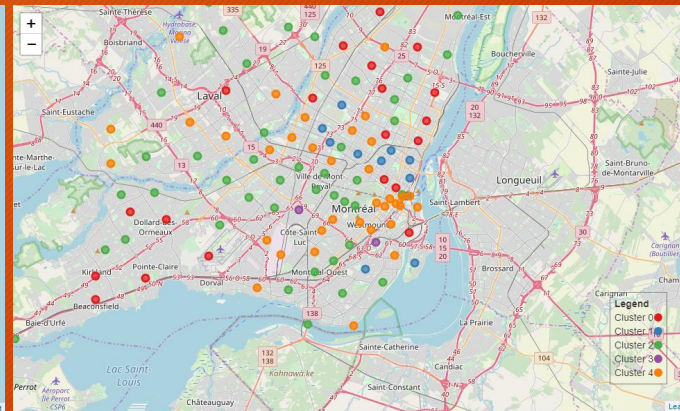| Cluster	Calgary, Alberta	Montreal, Quebec	Toronto, Ontario	1st Most Frequent	2nd Most Frequent	3rd Most Frequent	4th Most Frequent	5th Most Frequent
0	7	21	4	Asian Restaurant	Pizza Place	Fast Food Restaurant	Restaurant	Dessert Shop
1	2	11	1	Asian Restaurant	Fast Food Restaurant	Pizza Place	Dessert Shop	Restaurant
2	7	45	61	Asian Restaurant	Pizza Place	Fast Food Restaurant	Restaurant	Bakery
3	2	2	4	Asian Restaurant	Fast Food Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant
4	7	36	31	Asian Restaurant	Fast Food Restaurant	Restaurant	Pizza Place	Japanese Restaurant

Results: Cities and Clusters

City	0	1	2	3	4	1st Most Frequent	2nd Most Frequent	3rd Most Frequent	4th Most Frequent	5th Most Frequent
Calgary, Alberta	7	2	7	2	7	Asian Restaurant	Pizza Place	Fast Food Restaurant	Sandwich Place	Vietnamese Restaurant
Montreal, Quebec	21	11	45	2	36	Asian Restaurant	Fast Food Restaurant	Restaurant	Pizza Place	Dessert Shop
Toronto, Ontario	4	1	61	4	31	Asian Restaurant	Pizza Place	Fast Food Restaurant	Restaurant	Japanese Restaurant

Results: Venue Maps

Toronto, Ontario

Montreal, Quebec

Calgary, Alberta

Legend
Cluster 0 ●
Cluster 1 ●
Cluster 2 ●
Cluster 3 ●
Cluster 4 ●

Legend

Cluster 0 🔴

Cluster 1 🔵

Cluster 2 🟢

Cluster 3 🟣

Cluster 4 🟠

Legend

Cluster 0 🔴
Cluster 1 🔵
Cluster 2 🟢
Cluster 3 🟣
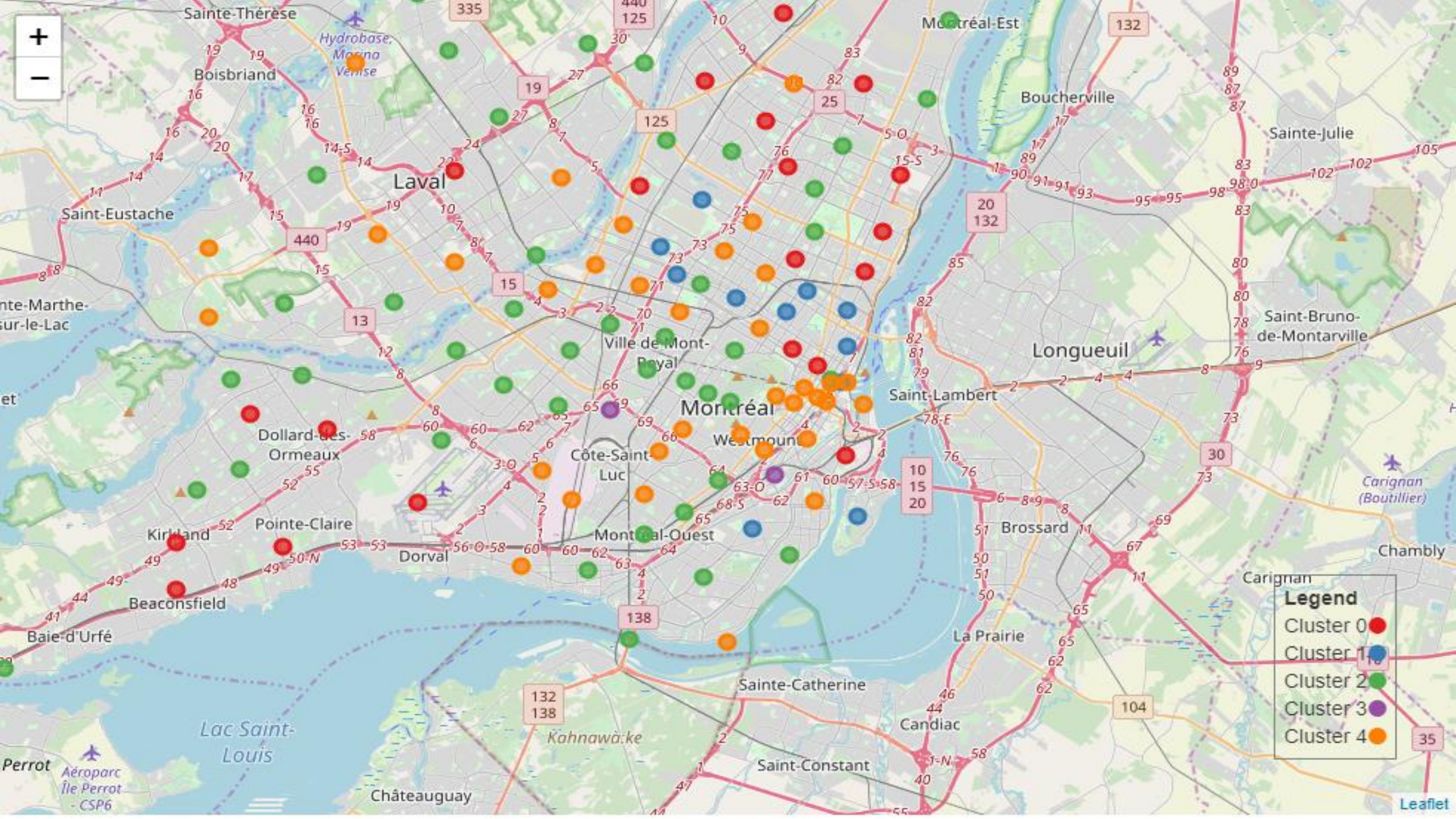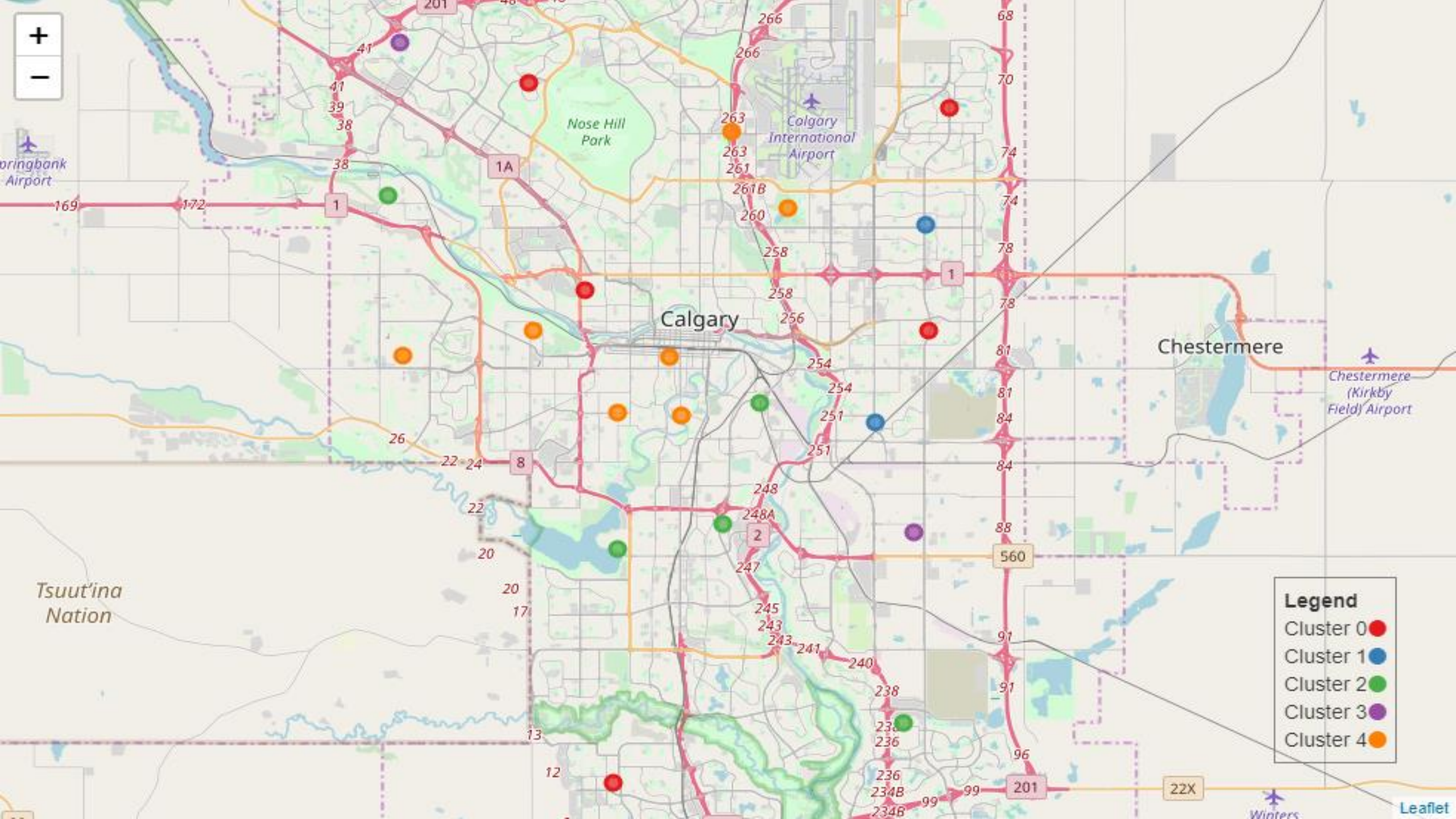Cluster 4 🟠
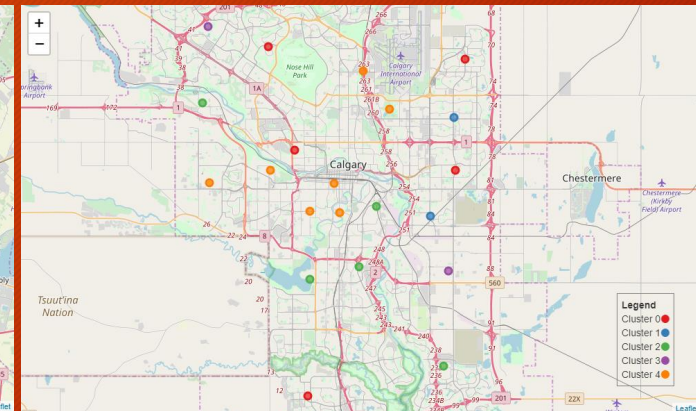
Discussion and Conclusions

- Low density of FSAs in Calgary
  - Lower density population, larger area
  - Some FSAs had no applicable venues
  - Less accurate representation
- High frequency of Coffee Shops, Cafes, Asian Restaurants, Fast Food and Pizza Places
- Calgary: Vietnamese
- Montreal: Dessert Shop
- Toronto: Japanese