School Clustering in Calgary and Edmonton

Introduction

Edmonton and Calgary are the two large cities in Alberta, Canada. Every year, an Alberta school rank is released to the public according to grade 6 and/or grade 3 students' test performance.
(http://alberta.compareschoolrankings.org/elementary/Fraser_School/Edmonton/Report_Card.aspx)

Each participating school is rated from 0 to 10 based on the test performance, with 10 being the highest and 0 being the lowest. The parents who are interested in choosing a good school for their children's education would pay attention to the rank and consider purchasing a property close by. However, it is my interest to add the geoinformation to the data set and observe school clustering in Edmonton and Calgary. In specific, which area or school zone has better education quality than other areas. To analyze the school scores and their geo-codes and determine if there are clusters formed, we created CSV file that contains all elementary school rankings in Alberta. We will also add the geocode to the database to help us visualize the clusters. The project is helpful to those young parents who want to buy a property with good schools nearby.

Data

Data is collected from the website which includes the school's name, 2017-2018 ratings, rating in the most recent five years, and we added the geocode, i.e. the latitude and longitude to the CSV file for elementary schools in Edmonton and Calgary.

Calgary school information:

https://github.com/clairvoan/Coursera_Capstone_NY/blob/master/Calgaryschool.csv

Edmonton school information:

https://github.com/clairvoan/Coursera_Capstone_NY/blob/master/Edmontonschool.csv

Methodology

I use the KMeans to find if there are school clusters based on their 2017-2018 rating, most recent five years average, latitude, and longitude. To differentiate school zones, I decide to use five categories in the KMeans clustering method.

Next, I use the folium map to visualize the school cluster on the map. For the five clusters, I use color-coding to illustrate their clusters. Green is the cluster with the highest rate and Red is the cluster with the lowest rate.

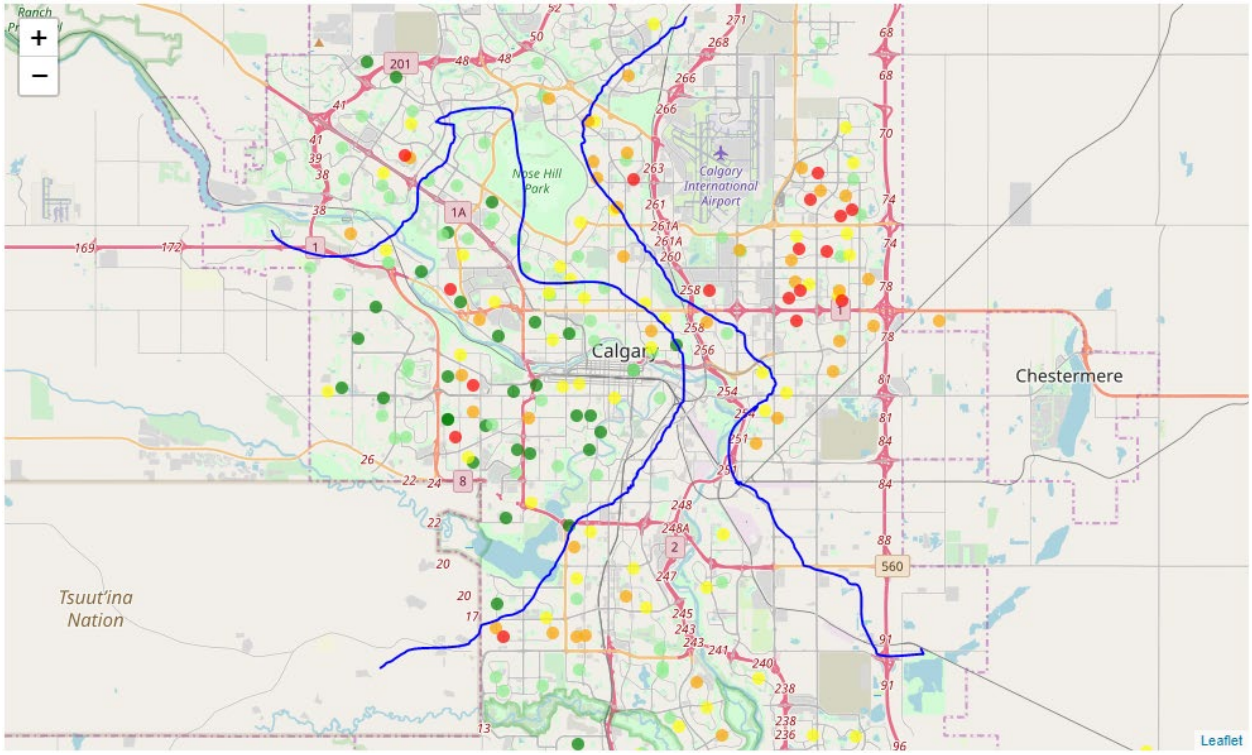
Results

The results tell us in both cities, we can find a prominent difference among five clusters.

The table below is the Calgary school clusters with the average 2017-18 rating, average five years rating, and geo center of each cluster.

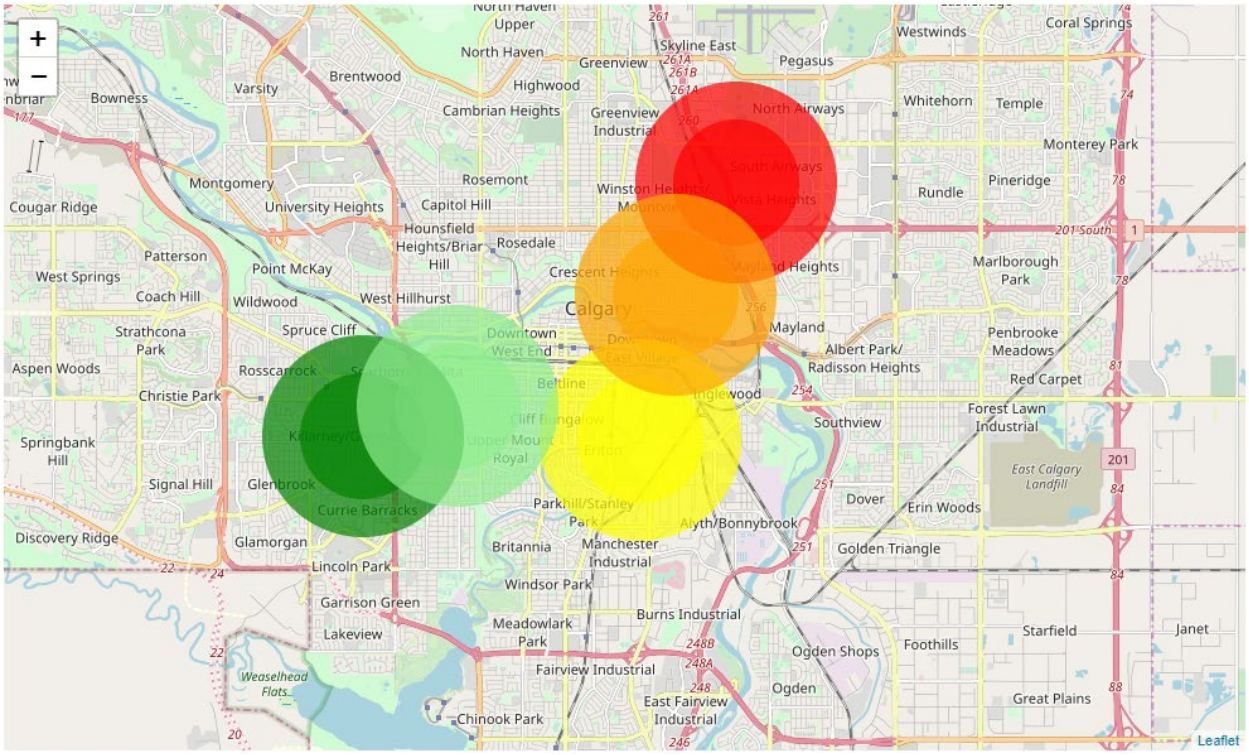
	2017-2018Rating	Adj5yrs	latitude	Longitude
Clus_km				
0	6.530189	6.194340	51.030711	-114.051286
1	8.788571	8.577143	51.031393	-114.126749
2	3.282353	4.047059	51.074880	-114.024902
3	7.447826	7.366667	51.036599	-114.100932
4	5.060526	5.323684	51.055639	-114.041350

To mark each school on the map with color-coding:



* I manually added the two blue curves as we can visually find the school clusters distributions.

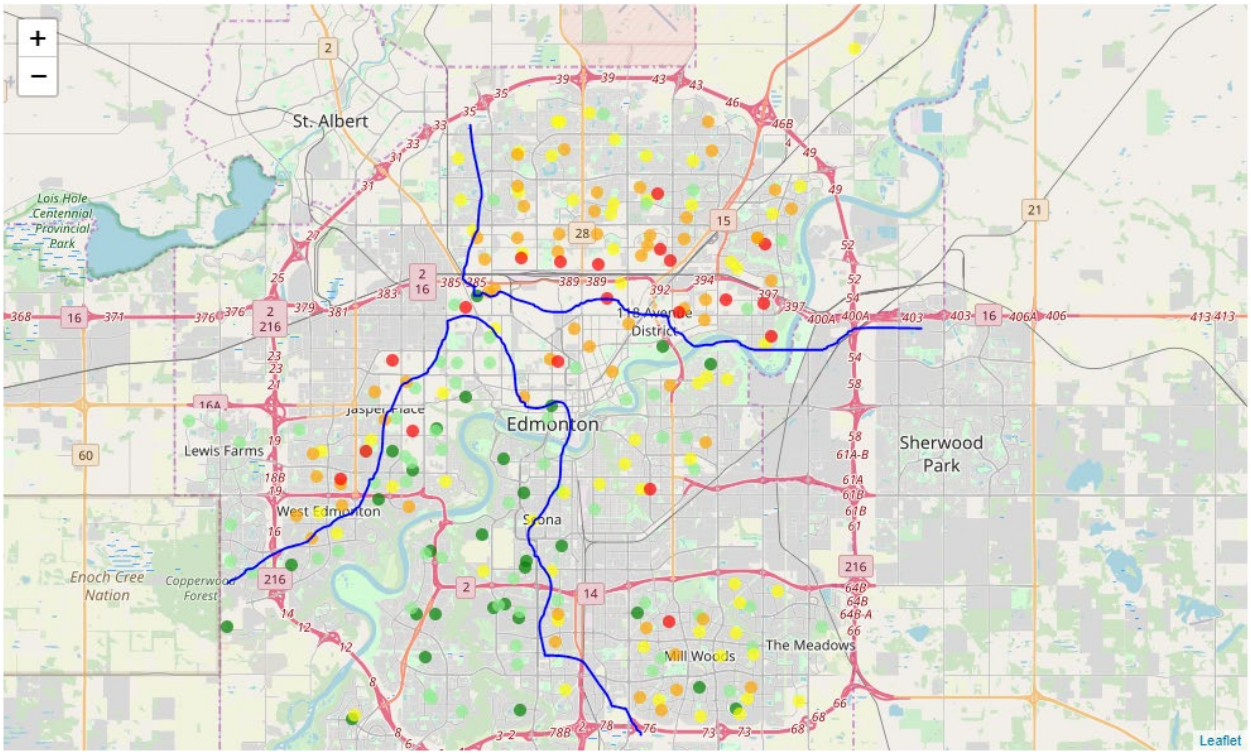
The map-based on grouped school clusters with their average values:

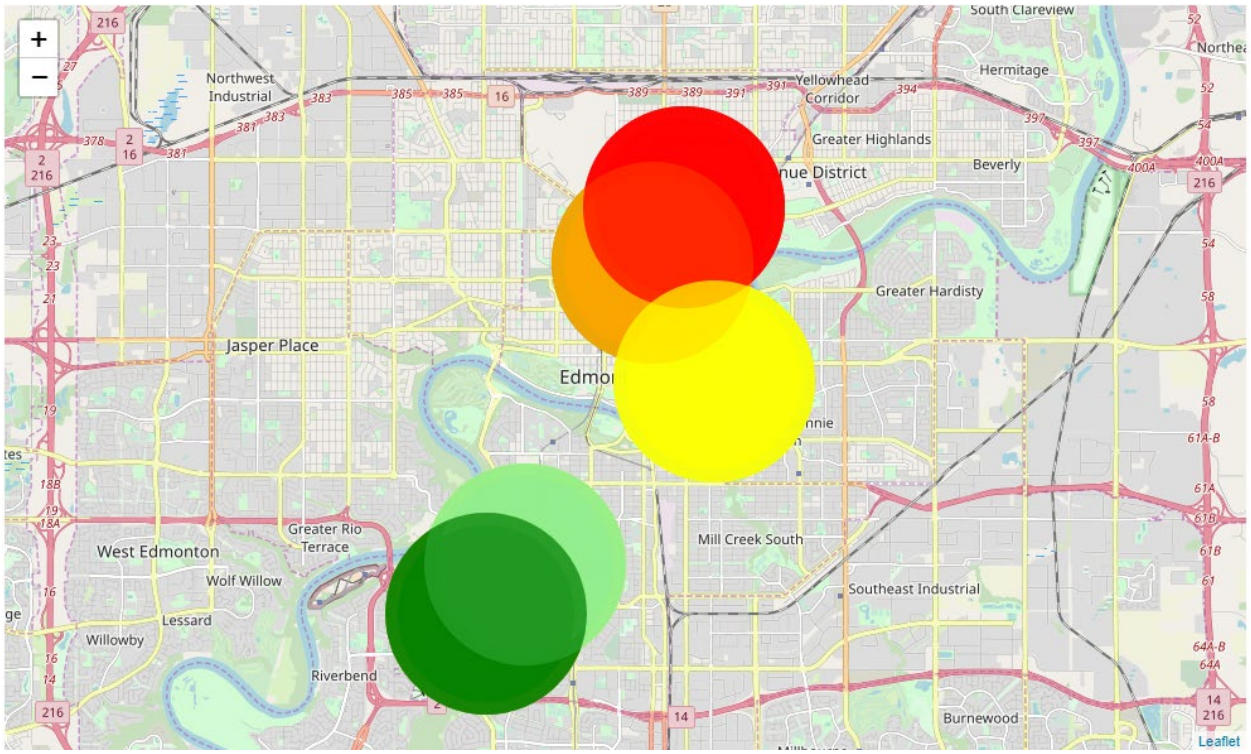


Edmonton school clusters table is presented as follow:

	2017-2018Rating	Adj5yrs	latitude	Longitude
Clus_km1				
0	7.075362	6.997101	53.505077	-113.529910
1	4.329310	4.670690	53.553680	-113.495149
2	2.335000	2.665000	53.562676	-113.486807
3	5.795455	5.843939	53.534498	-113.478428
4	8.313793	8.258621	53.497126	-113.540522

And the two respective maps:


St. Albert
Lois Hole Centennial Provincial Park
Edmonton
Sherwood Park
Lewis Farms
West Edmonton
Enoch Cree Nation
Mill Woods
The Meadows
Leaflet


Northwest Industrial
Yellowhead Corridor
Hermitage
Greater Highlands
Beverly
Greater Hardisty
Jasper Place
West Edmonton
Greater Rio Terrace
Wolf Willow
Lessard
Willowby
Riverbend
Mill Creek South
Southeast Industrial
Burnewood
Leaflet

Discussion

	2017-2018Rating	Adj5yrs	latitude	Longitude
Clus_km				
0	6.530189	6.194340	51.030711	-114.051286
1	8.788571	8.577143	51.031393	-114.126749
2	3.282353	4.047059	51.074880	-114.024902
3	7.447826	7.366667	51.036599	-114.100932
4	5.060526	5.323684	51.055639	-114.041350

According to the cluster table of Calgary, we find that the diverging trend. In specific, zone 1 is showing the highest rating, it is improving in recent 5 years, as the most recent one-year rating 2017-2018, is 8.79, higher than the five-year average, 8.58. Meanwhile, zone 2 is showing the lowest rating, and the most recent one-year rating is lower than the five-year average (3.28, dropping from 4.05).

The two maps associated with the data table tell us the west side of Calgary is much better than the east side.

Similarly, in Edmonton, we found the same trend, the school zone with the highest rating is improving. It is also noticeable that the average rating of each cluster in Edmonton is lower than in Calgary. So averagely, the rating of Calgary schools is higher than Edmonton. And the two maps tell us that the education quality at Southwest Edmonton is much better than Northeast Edmonton.

	2017-2018Rating	Adj5yrs	latitude	Longitude
Clus_km1				
0	7.075362	6.997101	53.505077	-113.529910
1	4.329310	4.670690	53.553680	-113.495149
2	2.335000	2.665000	53.562676	-113.486807
3	5.795455	5.843939	53.534498	-113.478428
4	8.313793	8.258621	53.497126	-113.540522

Conclusion

The school clustering analysis presents the decisive result that the West side of Calgary and Southwest area of Edmonton has better education quality.