HOUSING IN EDMONTON

Reginald Acquah

AUGUST 11, 2020

Reginald Acquah

Battle of Neighborhoods

Table of Contents

1. Introduction

Real estate is one of the most profitable businesses in the 21st century. An investor wants to build an apartment in Edmonton area purposely to rent out. Now, it is assumed that rent rates in Edmonton are affected by specific Neighborhood features. These are listed in order of importance as follows:

1. Security perception
2. Proximity to transit stations
3. Supermarkets and Shopping malls
4. Recreational parks
5. Hospitals
6. Schools

Due to how sparsely these factors are located it has become an issue for the investor to determine which neighborhood to build the apartment. The investor therefore is seeking for a way to model the city in order to rank which residential area has on average a greater proportion of these features more attached to it.

The investor is not certain of which neighborhood to build the apartment. We have been tasked to make a recommendation about which neighborhood will be most profitable.

2. Data

Data for neighborhoods in Edmonton will extracted from City of Edmonton Our focus will be on neighborhoods with a population density greater than the mean population density in Edmonton. This is to consider the effect population plays in determining rent rate. We will further export neighborhood centroid points from the City of Edmonton

Also, we will export data from City of Edmonton on occurrence of crime in each neighborhood to account for the security perception. This data contains crime in each neighborhood from 2009 to 2019.

In order to explore the various venues in Edmonton, we will use Foursquare location data to explore venues to aid us in clustering the residential neighborhoods with potential of high rate of investment. We will focus on venues related to transit stations, supermarkets and shopping malls, recreational parks, hospitals and schools.

3. Methodology

3.1. Data Wrangling

We import demographic data of Edmonton and sum across to get total population for each neighbourhood. Table A shows a subset of the data.

Table A: Demographic Data of Edmonton

Neighbourhood Number	Neighbourhood Name	Population
3140	CRESTWOOD	1129
3330	PARKVIEW	1764
6110	CPR IRVINE	12
5350	RHATIGAN RIDGE	1399
4140	ELMWOOD	1506

We then merge with Edmonton neighborhood centroid data from City of Edmonton which consists of geographical coordinates and land size for each neighborhood as the subset shown in Table B.

Table B: Edmonton Geographical Data

Neighbourhood Number	Neighbourhood Name_x	Population	Neighbourhood Name_y	Area Sq Km	Latitude	Longitude
3140	CRESTWOOD	1129	Crestwood	1.168158	53.535434	-113.569038
3330	PARKVIEW	1764	Parkview	1.546448	53.524060	-113.567914
6110	CPR IRVINE	12	CPR Irvine	0.663610	53.507527	-113.490549
5350	RHATIGAN RIDGE	1399	Rhatigan Ridge	1.344078	53.474506	-113.587569
4140	ELMWOOD	1506	Elmwood	1.025925	53.515738	-113.605993

We now import neighborhood crime data from City of Edmonton and sum occurrence of crime by neighborhood to get the crime score for each neighborhood. Table C shows a subset of the data.

Table C: Edmonton Crime Data

Neighbourhood Description (Occurrence)	# Occurrences
ABBOTTSFIELD	976
ALBANY	221
ALBERTA AVENUE	5048
ALBERTA PARK INDUSTRIAL	260
ALDERGROVE	937

We merge Table 2 and Table 3 to get a data for Edmonton neighborhoods consisting of geographical coordinates, population density and crimes. We focused on neighborhoods with a population density greater than the mean population density in Edmonton and replaced missing crime data with the mean of crime in Edmonton neighborhoods. This is shown in Table D.

Table D: Geographical coordinates with Crime and Population Density Data

Neighborhood	Latitude	Longitude	Crimes	Population density
Crestwood	53.535434	-113.569038	566.0	966.478990
Parkview	53.524060	-113.567914	564.0	1140.678328
CPR Irvine	53.507527	-113.490549	1073.0	18.082917
Rhatigan Ridge	53.474506	-113.587569	194.0	1040.862586
Elmwood	53.515738	-113.605993	426.0	1467.943293

3.2. Exploratory Data Analysis

In this section we explore the various neighborhoods in Edmonton with folium library. We use geopy library to generate the geographical coordinates of Edmonton and created a map of Edmonton neighborhoods with a population density greater than the mean population density. This is shown in Figure A below.

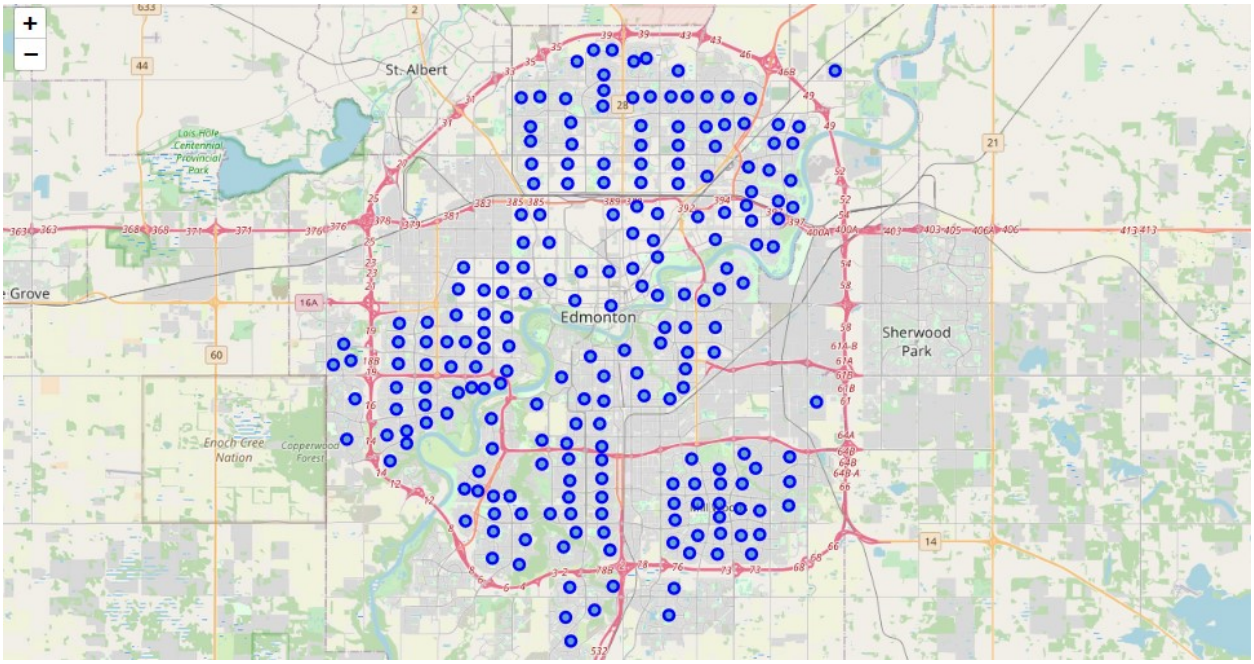

Figure A:Map of Edmonton Showing Neighborhoods with Higher Population Density

We then used the Foursquare API to generate venues specifying category Id of supermarkets, grocery shops, light train stations, schools and hospitals within a radius of 10km of each neighborhood. Table E shows a subset of this data.

Table E: Foursquare Location Data

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Crestwood	53.535434	-113.569038	IGA Andy's Valleyview	53.525746	-113.566761	Grocery Store
Crestwood	53.535434	-113.569038	Save-On-Foods	53.542695	-113.508737	Supermarket
Crestwood	53.535434	-113.569038	T&T Supermarket	53.523360	-113.623934	Supermarket
Crestwood	53.535434	-113.569038	Safeway Oliver	53.547432	-113.518189	Grocery Store
Crestwood	53.535434	-113.569038	Safeway Canada	53.559390	-113.553442	Supermarket

Table F shows a subset of the 10 most common venues in each neighborhood

Table F: Top 10 Venues in Each Neighborhood

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
Abbottsfield	Grocery Store	Hospital	Light Rail Station	Supermarket	Community College	High School	Big Box Store	College Academic Building	Pharmacy	Medical School
Alberta Avenue	Grocery Store	Light Rail Station	Hospital	College Academic Building	Supermarket	Big Box Store	High School	Community College	University	Pharmacy
Aldergrove	Grocery Store	College Academic Building	Hospital	Light Rail Station	High School	Supermarket	College Administrative Building	Medical School	University	College Gym
Allendale	Grocery Store	Hospital	Light Rail Station	College Academic Building	High School	Supermarket	Big Box Store	Medical School	College Administrative Building	General College & University
Argyll	Grocery Store	Hospital	Light Rail Station	Big Box Store	College Academic Building	High School	Supermarket	Medical School	College Administrative Building	General College & University

3.3. Model Development

K-means algorithm will be employed in our study to cluster neighborhoods. This will give as translucent idea of which cluster our prefer neighborhood will be located. Firstly, we find the optimal number of clusters using the elbow method. The elbow was at 3 so we used 3 clusters in the model. The Figure B below illustrates the elbow graph.

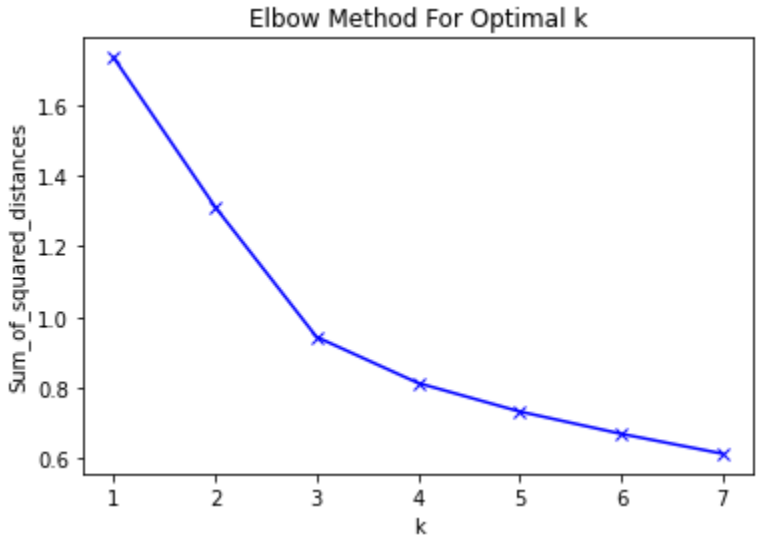

Figure B: Elbow Optimal K

Based on the conclusion from the elbow method we train the k-means algorithm to cluster neighborhoods into 3. Figure C below visualizes how Edmonton neighborhoods are clustered.

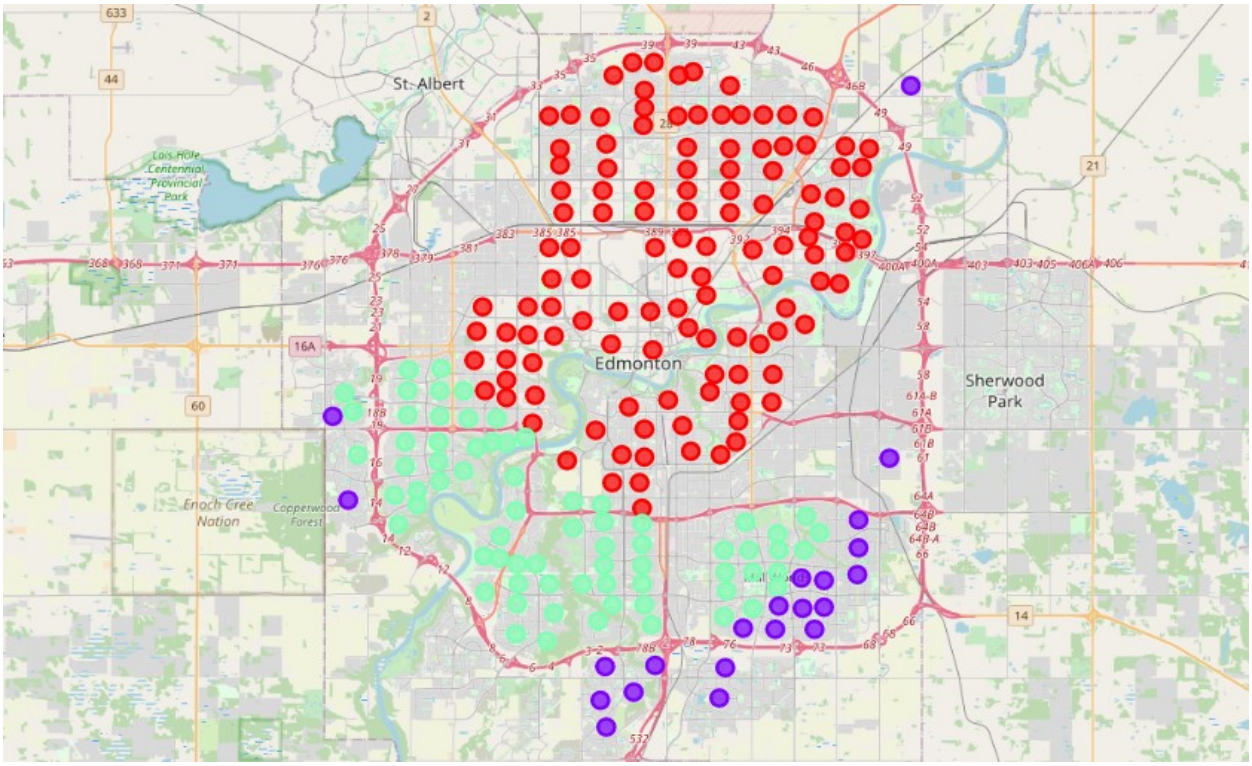

Figure C: Visualizing Clustered Neighborhoods in Edmonton

In order to make an informed decision, we will now apply scores to the various features as displayed in the table below and assign weights according to how common these venues can be found in the neighborhoods. Weight to be applied will range from 0.1 to 1. We will furthermore find the inverse of the crime score and multiply by 50 for the crime scores to be comparable to the venue scores. Finally, we will sum up the scores and then rank clusters according to the total score we gain. The cluster with the highest rank is what we will advise the investor to go for. Table G shows the score of each neighborhood feature score. Table H shows the ranking of top 5 neighborhoods by scores

Table G: Neighborhood Feature Scores

Neighborhood Feature	Score
Proximity to transit stations	5
Supermarkets and Shopping malls	4
Recreational parks	3
Hospitals	2
Schools	1

Table H: Neighborhood Rankings

Neighborhood	Cluster Labels	Crimes	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	Score
Casselman	0	0.980392	4	3.6	1.6	3.5	0.6	2.0	0.4	0.3	0.4	0.1	17.480392
Matt Berry	0	0.130548	4	4.5	3.2	1.4	0.6	2.0	0.4	0.6	0.2	0.3	17.330548
Ozerna	0	0.138122	4	4.5	3.2	1.4	0.6	2.0	0.4	0.6	0.2	0.1	17.138122
Central McDougall	0	0.222222	4	4.5	1.6	2.8	2.4	0.5	0.4	0.3	0.2	0.2	17.122222
Oliver	0	0.007758	4	4.5	1.6	2.8	2.4	0.5	0.4	0.6	0.2	0.1	17.107758

4. Results

Neighborhoods in Cluster 0 have mostly grocery shops, and supermarkets in the environs. They also have easy access transit and bus stations as well as hospitals. Schools are less relatively less visible in these neighborhoods.

Neighborhoods in Cluster 1 also have lots of grocery shops but not supermarkets. Hospitals are much common in this cluster than transit stations. Like cluster 1, Schools are less relatively accessible in these neighborhoods.

Neighborhoods in Cluster 2 have lots of grocery as well as schools relative to the other clusters. Like cluster 1 Hospitals are much common. Transit stations are mostly the 3rd most common venues in this cluster.

After aggregating the neighborhood scores, we realise that neighborhoods in cluster 0 are place where we can charge higher rent rates in Edmonton. They have a score above 17 points

5. Discussions

Our analysis shows that neighborhoods in cluster 0 are have most of the important neighborhood characteristics that can increase rent price. This cluster is more concentrated around the center of

the city. This can be attributed to the concentration of light rail station in the center of the city and the location of the Alberta hospital and the University of Alberta which are all located around the center of Edmonton.

The best neighborhood to build the apartment will be in cluster 0 and from our rankings, Casselman is the appropriate choice.

6. Conclusion

The aim of the project was to identify which area in Edmonton will best suite an apartment that will yield higher rent rates. By clustering and assigning crime scores to the neighborhoods we conclude that cluster 0 is best suite and more specifically Casselman. We recommend that the investor can choose any neighborhood in cluster 0 since they are similar.