HOUSING IN EDMONTON

Battle of Neighborhoods

Introduction

- Real estate is one of the most profitable businesses in the 21st century
- It is assumed that rent rates in Edmonton are affected by specific Neighborhood features
  - Security perception
  - Proximity to transit stations
  - Supermarkets and Shopping malls
  - Recreational parks
  - Hospitals
  - Schools
- We have been tasked to make a recommendation about which neighborhood will be most profitable

Data

Data for neighborhoods in Edmonton was extracted from City of Edmonton website

We will further export neighborhood centroid points from the City of Edmonton website

Crime data for neighborhoods from 2009 to 2019 was also be exported from the City of Edmonton's website

We used Foursquare location data to explore specific venues to aid us in clustering the residential neighborhoods

Our interest is venues related to transit stations, supermarkets and shopping malls, recreational parks, hospitals and schools

Methodology

Our methodology is in three parts

- Data Wrangling
- Exploratory Data Analysis
- Model Development

Data Wrangling

We import demographic data of Edmonton and sum across to get total population for each neighborhood

We then merge with Edmonton neighborhood centroid data from City of Edmonton which consists of geographical coordinates and land size for each neighborhood

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

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.

Data Wrangling

▶ We then merge all the data sets to get a data looking partly like show below

| 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

Exploratory Data Analysis
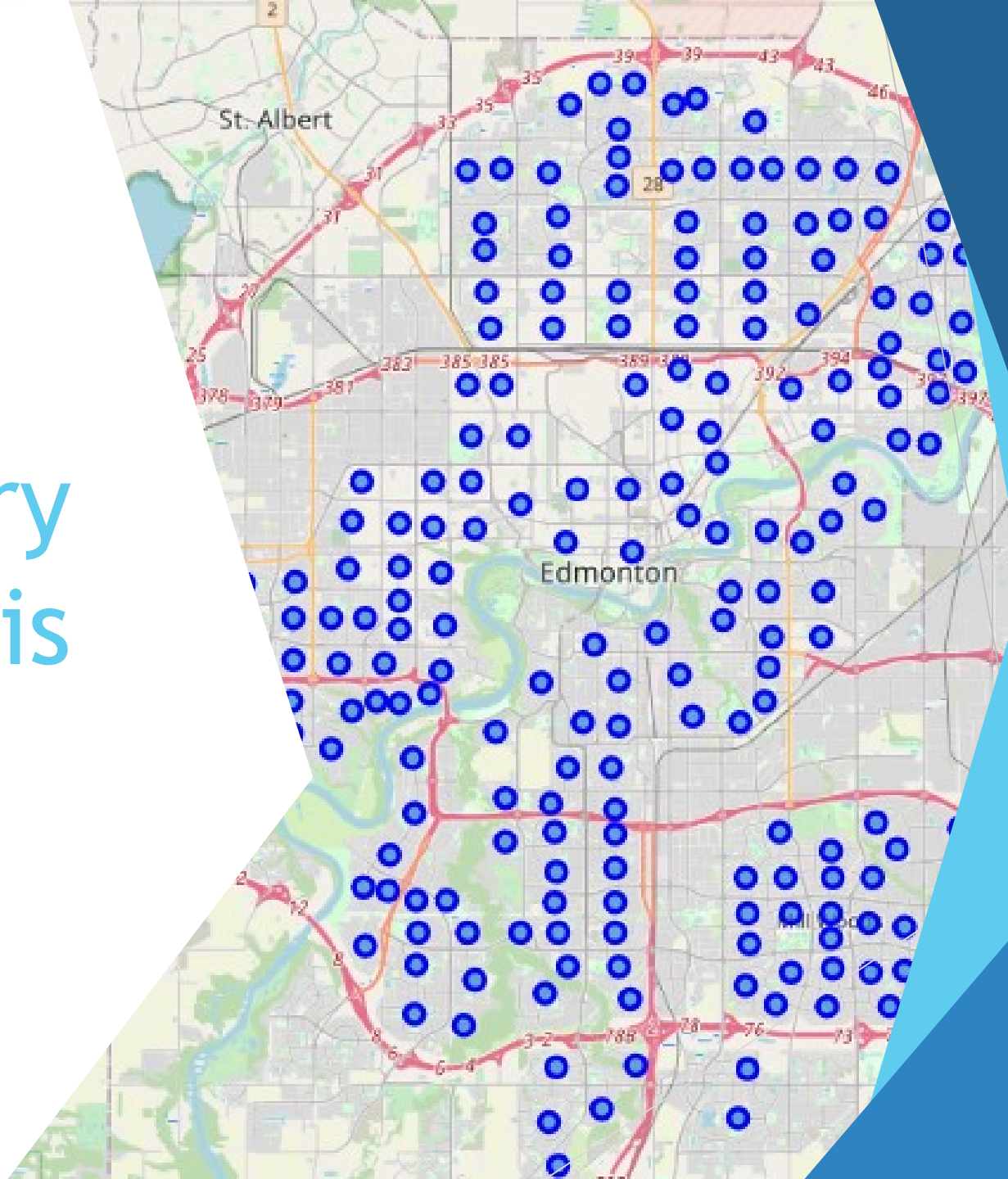
Foursquare Location Data

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

Model Development

▶ We employ the K-means algorithm

▶ 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.

Elbow Method For Optimal k
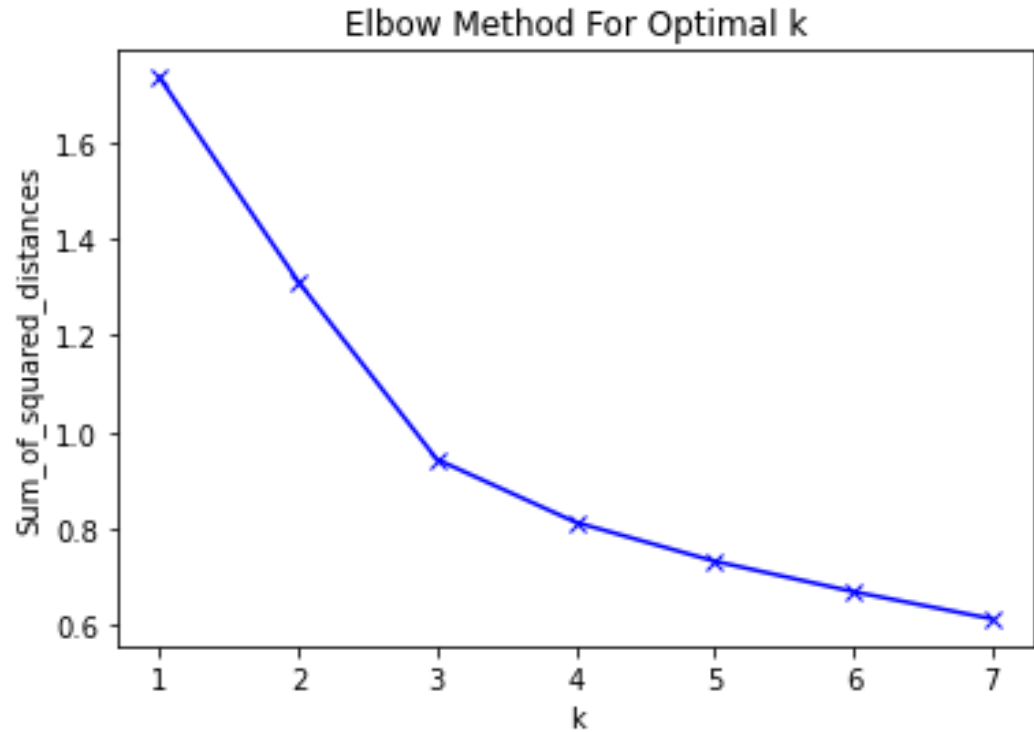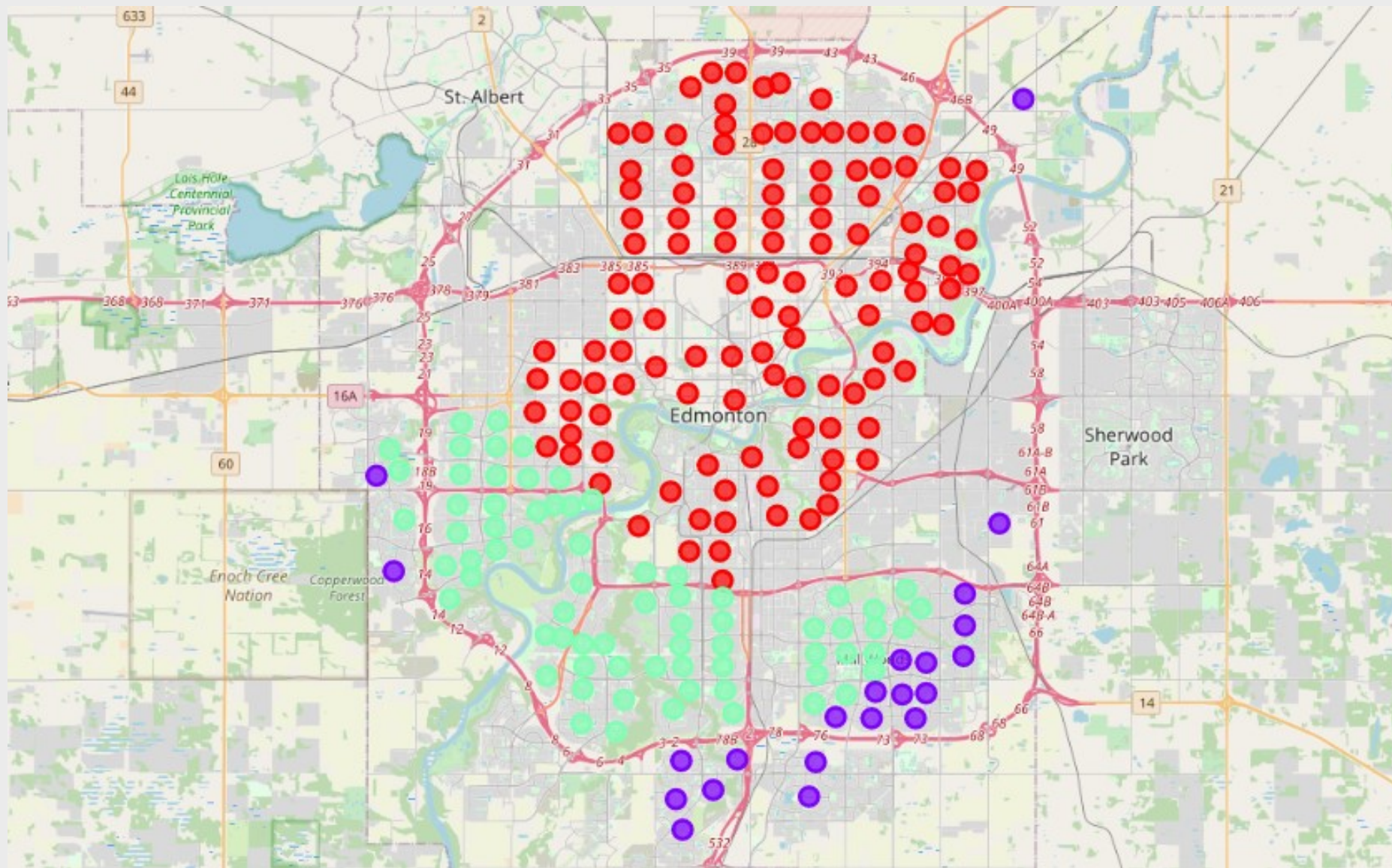
Model Development

We apply scores to the various features 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

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

Model Development

The table shows the score for each neighbourhood feature

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

Neighborhood Rankings

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

Discussion

Our analysis shows that neighborhoods in cluster 0 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.

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.