More than walls: Choosing a place to call home

APPLIED DATA SCIENCE CAPSTONE

By

Francis Frimpong

Introduction

- The world has seen major growth and development. However, the basic needs of humans have remained the same.
- Food, clothing and shelter (housing).
- Individuals and real estate investors look out for different factors when deciding on places to buy homes or invest in real estate.
- Examples of factors considered are public safety (absence of crimes) and proximity to social and economic amenities.

Data

- Using data from the City of Edmonton, Alberta, Canada, we explore which neighborhoods are most preferable.
- Data sources
  - Census Data
  - Location data
  - Crime occurrence data
  - Location/amenities data from Foursquare
- Python (version 3.8) is used for the data analysis procedures and executing methodology.

Methodology

- Data wrangling is used to combine the relevant datasets and attributes, including combining demographic data and crime occurrence data.
- Exploratory analysis is applied to identify and visualize the neighborhoods of interest, i.e. must have a population density higher than the average within the city.

Neighbourhood Description (Occurrence)	Neighbourhood Name_y	# Occurrences
ABBOTTSFIELD	Abbottsfield	976
ALBANY	Albany	221
ALBERTA AVENUE	Alberta Avenue	5048
ALBERTA PARK INDUSTRIAL	Alberta Park Industrial	260
ALDERGROVE	Aldergrove	937

Table 3 – Neighbourhood and Crime occurrences

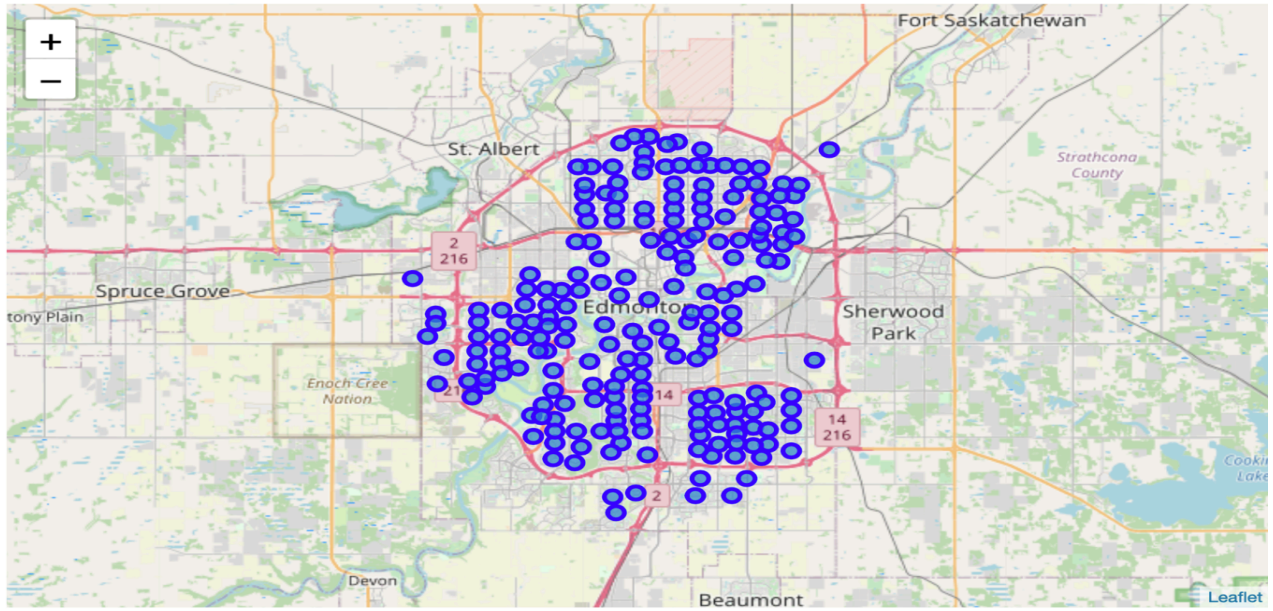

Fig 1 - Map of the City of Edmonton

Methodology

- The K-Nearest Model (KNN) machine learning model is applied.
- The k-means algorithm is used to find the optimal number if clusters – 3 clusters
- The 3 clusters are visualized on the map of the city of Edmonton using Python packages and libraries

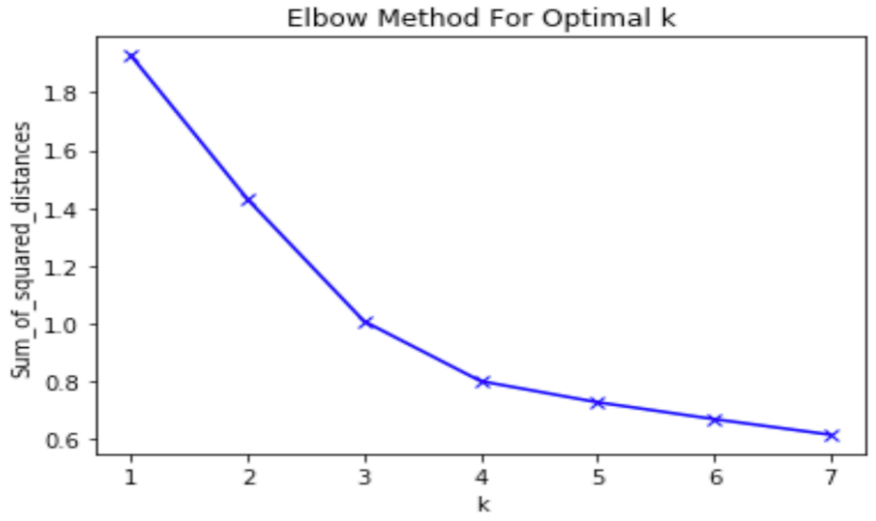

Figure 2 – The elbow method for selecting best k

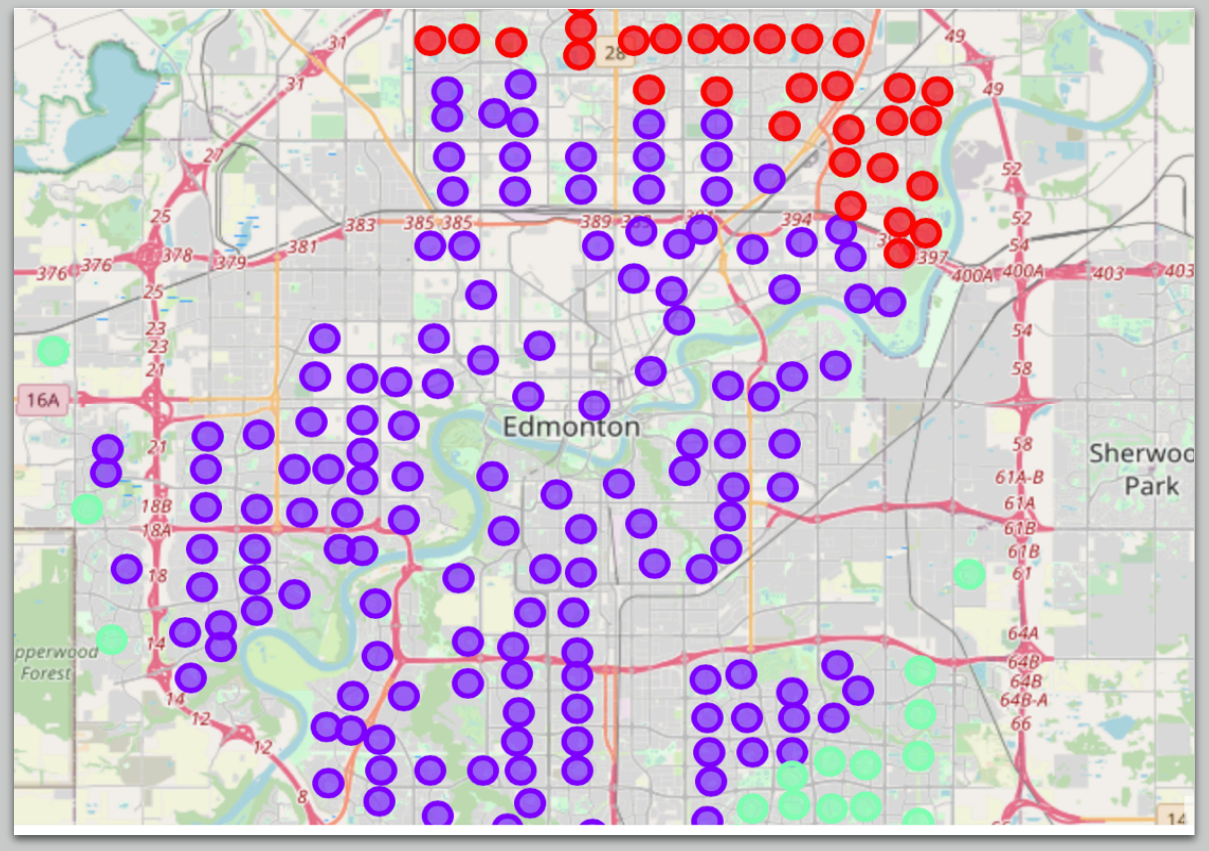

Figure 3 – Neighbourhood clusters

Results

- Neighbourhoods in cluster 0 were ranked the highest among all three clusters.
- These neighbourhoods featured predominantly amenities such as grocery shops and supermarkets, and are closer to the city center, i.e. downtown.
- The neighbourhoods in cluster 1 were the least ranked and are relatively farther away from downtown.

Neighbourhood	Cluster Labels	Crime Frequency	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
Klarvatten	0	0.085616	4	3.6	4.0	2.8	1.2	1.5	0.8	0.3	0.2	0.1	18.585616
Kirkness	0	0.069061	4	3.6	3.2	3.5	1.2	0.5	1.2	0.6	0.2	0.1	18.169061
Bannerman	0	0.062972	4	3.6	3.2	3.5	1.2	0.5	1.2	0.6	0.2	0.1	18.162972
Lago Lindo	0	0.093985	4	3.6	1.6	3.5	2.4	1.5	0.8	0.3	0.2	0.1	18.093985
Miller	0	0.087873	4	3.6	4.0	2.8	1.2	1.0	0.4	0.3	0.6	0.1	18.087873
Casselman	0	0.060241	4	3.6	4.0	2.8	1.2	1.0	0.4	0.3	0.2	0.3	17.860241
Ozerna	0	0.138122	4	4.5	3.2	2.8	1.2	1.0	0.4	0.3	0.2	0.1	17.838122
Hollick-Kenyon	0	0.076687	4	3.6	3.2	1.4	3.0	0.5	1.2	0.3	0.4	0.1	17.776687
Matt Berry	0	0.130548	4	3.6	4.0	2.8	1.2	1.0	0.4	0.3	0.2	0.1	17.730548
Mayliewan	0	0.126582	4	3.6	4.0	1.4	2.4	1.0	0.4	0.3	0.2	0.3	17.726582
Kilkenny	0	0.031606	4	4.5	3.2	2.8	1.2	0.5	0.4	0.3	0.6	0.1	17.631606
Clareview Town Centre	0	0.029343	4	3.6	4.0	2.8	1.2	1.0	0.4	0.3	0.2	0.1	17.629343
Kernohan	0	0.091912	4	3.6	4.0	2.8	0.6	1.0	0.4	0.6	0.2	0.3	17.591912
Beaumaris	0	0.056561	5	3.6	3.2	2.8	1.2	0.5	0.4	0.3	0.2	0.3	17.556561
Sifton Park	0	0.082781	4	3.6	4.0	2.8	0.6	1.0	0.8	0.3	0.2	0.1	17.482781
York	0	0.044287	4	3.6	4.0	2.8	0.6	1.0	0.8	0.3	0.2	0.1	17.444287
Belle Rive	0	0.113895	4	4.5	3.2	1.4	2.4	0.5	0.4	0.6	0.2	0.1	17.413895
Hairsine	0	0.076336	4	4.5	3.2	2.8	1.2	0.5	0.4	0.3	0.2	0.1	17.276336
Lorelei	0	0.064851	5	3.6	3.2	1.4	2.4	0.5	0.4	0.3	0.2	0.2	17.264851
Schonsee	0	0.187970	4	3.6	3.2	1.4	0.6	1.5	0.8	1.5	0.2	0.1	17.087970

Table 5 – Scores: ratings of neighbourhood preference – ranked in descending order

Discussion and Conclusion

- Neighbourhoods closest to the city center were highly ranked and were in near proximity to key amenities such as colleges, school, hospitals, and principal recreational facilities, as compared to distant neighbourhoods in other clusters.
- Safety (lower crime rates) and proximity to key facilities are relevant in choosing places of dwelling for individuals and from a business point, communities real estate developers choose to invest in housing developments