Edmonton Neighbourhood Selection

Report Prepared By:

Adam Asquini, Coursera / IBM Applied Data Science Professional Certificate Capstone Project

Part 1: March 18, 2020

Introduction

Background Context

Edmonton is a major Canadian city with have a population of over 1M people (including surrounding suburban communities). Edmonton is located in the northern half of the Canadian province of Alberta, Canada and is Canada's most northerly major city[1]. Often referred to as the gateway to the north, Edmonton serves as a major economic centre for the oil and gas industry and provides a logistical and transportation hub for communities further north in the Alberta oil sands such as Fort McMurray. Edmonton also has a large public sector workforce and is growing its technology sector. As a large urban centre, Edmonton has thousands of real estate transactions across its hundreds of neighbourhoods each year. Having knowledge of these neighbourhoods would be a benefit to buyers, sellers and real estate professionals

Business Problem

According to the Alberta Real Estate Association, approximately 19,000 real estate transactions occur in the Edmonton region each year[2]. Given that many real estate transactions see hundreds of thousands (or millions) of dollars exchange hands, real estate purchases require careful consideration and can have profound impacts on both buyers and sellers. With thousands of real estate transactions occurring annually, both customers and real estate professionals would benefit from data-driven analysis to aide in these important investments.

This report examines a broad set of publicly available, online sources of data to aide customers and real estate professionals in making decisions around real estate transactions. Without this type of data-driven analysis, stakeholders are often left to more traditional and limited data analysis techniques such as looking at various sources of data independently or are reliant on the advice of people such as friends, advisors and workplace companions, which can be rife with bias.

The analysis completed in this report will attempt to segment and cluster neighbourhoods based on their similarities to other neighbourhoods (not including location). Data analyzed about neighbourhoods to conduct this grouping will include attributes such as amenities, crime statistics, demographics and property assessment values. The output of this analysis will be tool that can be used to better understand the general characteristics of Edmonton's neighbourhoods, which could then be mapped to a client's unique needs and preferences. For example, a client with a budget of $500K and a desire to live in an area with low crime rates could use the tool to find a cluster of neighbourhoods across the city that best meet this criteria. From here, other personal preferences such as the client's commute to work and the location of family and friends could be applied to further refine the areas the client may look for a new home.

While some tools already exist to complete similar analysis, such as the Multiple Listing Service (MLS), this tool will integrate other non-traditional sources of data from diverse sources to provide a more thorough analysis of neighbourhoods in Edmonton.

---

[1] https://globalnews.ca/news/5861019/census-edmonton-population-2019/
[2] https://albertarealtor.ca/page/data-explorer

Audience for this Report

The primary intended audience of this report are the clients (both buyers and sellers) that are interested in participating in a real estate transaction. The buyer's stake in this report is having more knowledge of neighbourhoods they may move to so that they can make a more informed choice about their potential home purchase. The seller's stake in this report is that they may be able to better understand the characteristics of the neighbourhood in which their property is located, which may help them with pricing and promoting positive attributes of their neighbourhood (i.e. location to parks or amenities).

A secondary audience of this report is real estate professionals. While Realtors and other real estate professionals may come to similar conclusions about neighbourhoods through their experience, this report may be useful when discussing options with clients as the report provides a less biased view of certain neighbourhoods in Edmonton.

Data Sources and Acquisition

The following data sources and their respective description have been utilized in this report.

Date Source (hyperlinked)	Description
City of Edmonton Open Data Portal – Edmonton Police Service Neighbourhood Criminal Occurrences	Statistics on crime reported to the Edmonton Police Service by neighbourhood in Edmonton in the year 2009 – 2019. Aggregate crime rates by neighbourhood will be calculated as an indicator of the level of crime in the respective neighbourhood
City of Edmonton Open Data Portal – Property Assessment Data (Current Calendar Year)	Property Assessment values for all homes in the city of Edmonton. The average assessed values of homes in each respective neighbourhood will be used as a comparator.
City of Edmonton Open Data Portal – City of Edmonton Neighbourhoods	List of all neighbourhoods and their respective neighbourhood id in the city of Edmonton. This list forms the basis for comparison across neighbourhoods.
City of Edmonton Open Data Portal – Neighbourhood (Centroid Point)	List of the latitude and longitude of the centre point (centroids) of each Edmonton neighbourhood. Used as a coordinate point to determine nearby amenities.
City of Edmonton Open Data Portal – 2016 Census Population by Household Income (Neighbourhood / Ward)	List of the number of people belonging to each income bracket in each neighbourhood in the city of Edmonton. Used to determine the average household income in each neighbourhood.
City of Edmonton Open Data Portal – 2016 Census - Population by Age (Neighbourhood / Ward)	List of the number of people belonging to each age bracket in each neighbourhood in the city of Edmonton. Used to determine the average age in each neighbourhood.
City of Edmonton Open Data Portal – 2016 Census –	List of the number of each home type (single-family dwelling, apartment etc..) in each neighbourhood in the city of Edmonton.

| [Population by Structure Type (Neighbourhood/Ward)](#) | Used to determine the rates of each structure type in each neighbourhood. |
| Foursquare location data (No hyperlink available. Data pulled through Foursquare API) | List of nearby (within 500 metres) venues from each neighbourhood centroid. |

*Notes
1) As I continue with my analysis for the week 2 – full report requirements, I may add more data sources to this list or remove them as required.
2) The headings for the remainder of the report (week 2 requirements) are shown below, but have yet to be completed.

Methodology – To be completed in week 2

Exploratory Data Analysis

Filtering out data (missing items etc..)

Normalization

Analysis of each cluster

Modeling Techniques

K-Means Neighbours for predictive modeling

Calculation of ideal K

Results

Discussion

Conclusion