The Effects of Neighbourhood Characteristics on Crime Incidence

Steven Letourneau*, Nathan Ell†, Peter Cheung‡, Jordan McCaskill§ and Mohamad El-Hajj¶
Department of Computer Science
MacEwan University, Edmonton, Canada
{* letourneaus5, † elln2, ‡ cheungp23, § mccaskillj2}@mymacewan.ca, ¶ elhajjm@macewan.ca

***Abstract*—Using data from the City of Edmonton, Canada Open Data Portal, an exploration process is undergone using data mining techniques to help detect unseen relationships between tangible spatial characteristics and non-tangible crime incidences. These findings will help law enforcement and city planners make empirically based decisions and avoid the misappropriation of public resources. Using frequent pattern analysis to examine neighbourhood attributes that occur alongside crime provides insight into why crime occurs. These techniques include clustering, classification algorithms, and association algorithms. Results of the analysis on neighbourhood spatial characteristics indicate that dwelling structure type and tree density relate to incidence of neighbourhood crime, while other neighbourhood spatial characteristics bear no relationship. Results also show that intangible neighbourhood characteristics indicate that the distribution of yearly household income and employment and school enrollment levels relate to incidence of neighbourhood crime. The distribution of yearly household income bears a relationship to crime type, specifically violent vs non-violent types.**

Keywords*—data mining; crime incidence; neighbourhood clustering; frequent pattern analysis.

I. Introduction

Modern city administrations generate and keep vast disparate data. This is often a result of requirements set in transparency policies, and the practicality of todays modern data-centric technology. Thanks to these factors, we are in an age of unprecedented digital data availability. Wide-ranging data sets include information from precipitation levels and geographic features, to crime statistics and traffic data. The use of such data is imperative for anybody looking to make responsible and informed decisions, like municipal administration. With regard to public safety, much of this data can be mined to analyze trends and causes of crime. The City of Edmonton, Canada regularly releases a wealth of crime statistics, and maintains data sets detailing city owned and maintained assets and their locations. This data can be examined at the neighbourhood level, creating profiles of crime and other characteristics. These neighbourhood profiles may contain unconventional spatial neighbourhood attributes such as public services, transit access, and property type. These spatial attributes have received little attention as potential predictors of non-tangible attributes such as crime incidence and type. Thirty-one different disparate data sets, consisting of 1,081,026 records are joined and analyzed using SQL Server Analytic Services [11] and SQL Business Intelligence Tools for Visual Studio [12]. The following questions are answered: Do spatial characteristics of neighbourhoods predict occurrences of crime? Do intangible characteristics of neighbourhoods predict occurrences of crime? Are there characteristics of neighbourhoods which predict the type of crime incidences?

II. Related Work

The analysis of crime statistics in conjunction with other data is becoming increasingly popular, due to an increase in data availability. Researchers from the University of California developed a method for programmatically identifying and validating predictive relationships between the visual attributes of images taken in a city and the non-visual attributes of those images locations [14]. The types of non-visual attributes which were predicted were crime statistics, housing prices, and population density. In short, using images of locations in a city as input, the researchers applied a trained predictor and a robust regression technique to the images to identify particular visual elements and generate an attribute value as output based on the visual elements identified. The results showed that there exists a relationship between the key visual elements of the images and the non-visual attributes such as crime rate and crime type.

Researchers from the University of Massachusetts developed a crime forecasting methodology with the support of a local Massachusetts police department. First, the study paired crime data with additional spatial data. Second, they divided the city into a large grid of equal sized squares. Third, they built a profile for each square from the grid using the paired data. Finally, classification methods were applied to each profiled square to forecast where, why, and how crime hot spots appear [2]. Several algorithms were employed and tested for correctness. Among these algorithms were a one nearest neighbour (1NN) algorithm, decision tree, and neural network model. It was found that in the short run the neural net strategy outperformed most methods, with the most robust and accurate predictions of crime incidence based on spatial and past events. Most algorithms performed better with a lower resolution grid (squares of larger area) because it offered better spatial variability as opposed to the higher resolution grid. The most accurate model overall was a 1NN classifier modified with set constraints; it was able to predict current hot spots with nearly 80% accuracy.

In the final research project examined, focus was placed on finding predictors of both spatial and temporal incidences. Data sets containing information of past crime incidences were run through an Apriori algorithm to identify frequent occurrences in the data and find frequent criminal patterns [7]. Naive Bayesian and decision tree classifiers were then used to predict the crime type which was most likely to occur at

978-1-5386-4649-6/18/$31.00

a given time and location, based on probability distributions generated by the classifiers. The results were compelling, as the Naive Bayes classification was able to predict crimes with approximately 50% accuracy. The technique was applied in other major US cities (San Francisco and Denver) and the results were consistent. The researchers were cognizant of the demographic data of the locations being studied, and normalized for population density as necessary.

III. Data Mining Overview

All data used is taken from the City of Edmonton Open Data Portal [1]. All of the data being used has been collected within the last 5 years, starting in 2012. Due to the nature of the research questions, it is necessary to aggregate a variety of tables containing both tangible and intangible traits:

Criminal Incidence Data by:

- **Type** - Assault, Sexual Assault, Robbery, Theft
- **Neighbourhood** - neighbourhood name
- **Year** - Year of the crime incident

Demographic Data for Neighbourhood by:

- **Household earnings** - Percent of households in each income bracket
- **Population** - Number of people per neighbourhood
- **Employment and unemployment** - Percent of people employed or unemployed
- **Age demographics** - Percent of people in each age bracket
- **Education enrollment** - Number of people enrolled in some level of school

Tangible Data for Neighbourhood by (per neighbourhood):

- **Tree density** - Number of city maintained trees
- **Libraries** - Number of Libraries
- **Schools** - Number of Catholic/Public
- **Transit access** - Number of Bus Stops
- **Parks** - Number of Parks
- **Sports facilities** - Number of ball parks, tennis courts,soccer fields, swimming pools, rec centers
- **Dwelling types** - Percentage of single detached houses, row houses, apartments, duplexes, institutional housing
- **Youth services** - Number of addiction, homeless, and women's and youth shelters
- **Police stations** - Number of police stations
- **Community halls** - Number of community halls

IV. Data Cleaning and Transformation

Examination of the relationship between crime and spatial traits is performed by analyzing the City of Edmonton using predefined neighbourhood boundaries. This decision simplified analysis by giving a common neighbourhood name dimension that could link multiple data sets together at a meaningful resolution. However, most spatial data sets contained latitude and longitude dimensions, which are transformed into the neighbourhood boundaries using a provided data set. The Shapely Python library [15] is used to perform this transformation. Irrelevant dimensions, such as neighbourhood identifying numbers, are removed. Crime occurrence data sets included the quarter of the year, type of crime, and neighbourhood that the crime incident report occurred in. This data is aggregated to produce a count of incidents per incident type per year for each neighbourhood.

Spatial data sets collected include playgrounds, ball diamonds, community halls, public schools, Catholic schools, youth services centers, soccer fields, parkland, bus stops, and city managed trees. These data sets come in a comma separated value format, and, are parsed to determine the neighbourhood dimension from longitude and latitude dimensions mentioned previously.

Non-spatial data sets collected include property values, household earning brackets, demographics, employment data, and dwelling types. Each of these data sets come in comma separated value format.

To identify trends within the data, the counts for spatial features by neighbourhood are discretized. These counts are used instead of raw counts for each feature in the analysis to simplify identification of outlier cases and interpretation of results. Discretization is accomplished by transformation of continuous values into discrete bins [6] appropriately according to the minimum, average, and maximum value for each field. For example, dwelling types within a neighbourhood are transformed from a number of dwellings per dwelling type to a percentage of the overall dwellings. This allowed for direct comparisons to be made.

V. Crime Data Analysis by Neighbourhood Using Clustering Methods Overview

To better understand the attributes which cause crime occurrences in a neighbourhood, clustering is used. Clustering is the act of partitioning a set of data objects into subsets based on similarities between their dimensions. [4]. This allows partitioning into different sets based on a similar number of occurrences of crime and crime type.

The number of occurrences for each crime type are queried and ordered. From this result, the top ten and bottom ten neighbourhoods for each crime type are selected. Afterwards, neighbourhoods are scored by the number of times they are found in the top or bottom ten for any crime type. The neighbourhoods which occurred with the highest frequency within the top ten formed the "worst crime" neighbourhood cluster and the groups that appeared with the highest frequency within the bottom ten formed our "best crime" neighbourhood cluster. Both clusters are then joined with the demographic and neighbourhood features data.

A heat-map visualization is applied to each dimension individually over the entire range of values in both clusters assisted in deciding which dimensions to investigate. The heat-map also highlights potential differences between the clusters. This gave a good starting point for further exploration of the data [9].

VI. Crime Data Analysis by Type Using Clustering Methods

A. Do spatial characteristics of neighbourhoods predict occurrences of crime?

Examination of spatial characteristics and their relation to crime incidence level is performed. Differences are observed between mean values of several spatial features, As demonstrated in Table 1, various neighbourhood characteristics showed pronounced differences between the high and low crime neighbourhoods.

Characteristic	Mean Value Best Crime	Mean Value Worst Crime	Std Dev. Best Crime	Std Dev. Worst Crime
City maintained trees	1614.95	1007.9	1217.05	880.59
Structure % under 5 stories	9.93	25.07	10.67	28.71
Public parks	3.82	3.30	3.05	2.54
Libraries	0.05	0.22	0.39	0.48
Recreation centres	0.11	0.32	0.35	0.71

TABLE 1: Spacial characteristics of neighbourhoods. These characteristics showed stark differences between high and low crime neighbourhoods.

To compare the mean values for different spatial characteristics in the two clusters, a *t-test* is used - a statistical test between two sample mean values between two groups [13]. In the approach of mean comparison a decision is made with respect to the null hypothesis. The p value obtained from *t-tests* is the probability, given the null hypothesis, of obtaining a difference in mean values that is at least as large as the difference in means observed. The null hypothesis states that that there is no significant difference between population means observed and that any observed difference is due to error or as a result of sampling [3]. *T-tests* performed on each spatial dimension's mean values revealed that the slight variances between nearly all of the attributes are statistically insignificant ($p > 0.05$). The only statistically significant differences in mean values between spatial features is the number of trees per neighbourhood and the dwelling type percentages per neighbourhood ($p < 0.05$). With respect to trees, neighbourhoods that experience lower overall rates of crime have more trees than those that experience higher rates of crime. Crime incidence is found to be more higher in neighbourhoods that have a higher percentage of apartments under five stories, as shown in Figure 1. Although crime incidence varies noticeably between neighbourhoods that have higher percentages of single detached houses as well, the difference is much more pronounced for apartments under 5 stories, differing by over 50%.

Apart from trees and structure type, no other relationship between crime occurrences and spatial characteristics in a particular neighbourhood exists. While differences are observed between mean values of other spatial attributes, performing *t-tests* on these dimensions found them to be statistically insignificant between the best and worst crime neighbourhood clusters. This leads to the conclusion that most typical spatial features in a neighbourhood do not have an effect on the overall crime incidence of a neighbourhood.

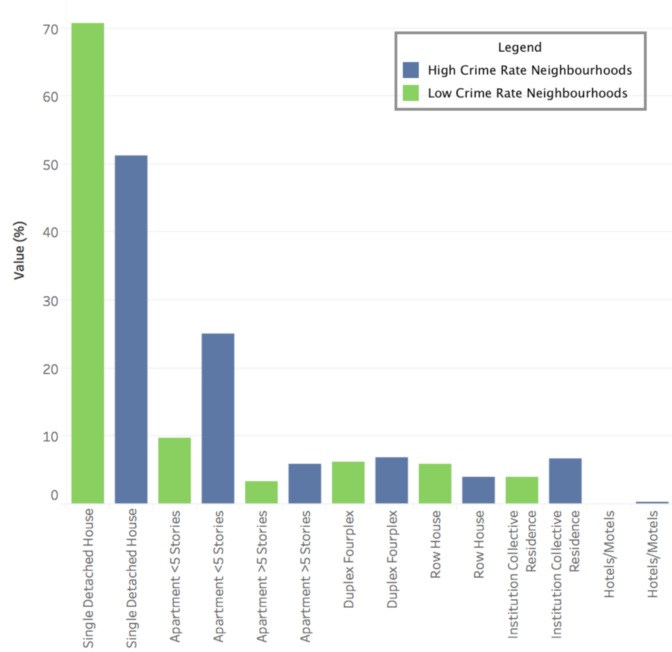

Fig. 1: Comparison of dwelling type occupation between high and low crime neighbourhoods.

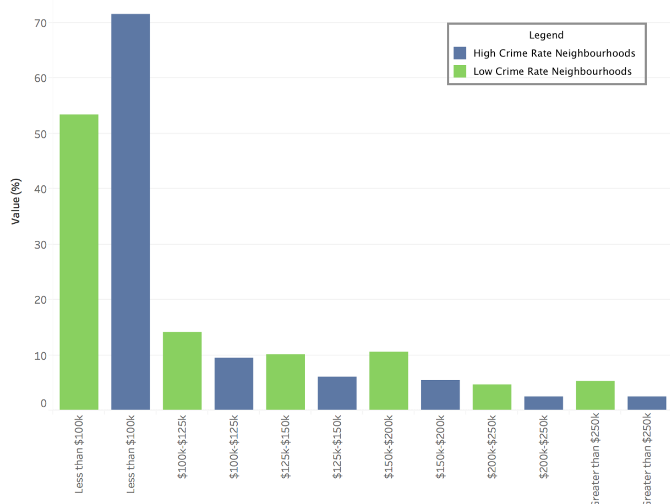

Fig. 2: Comparison of household income between high and low crime neighbourhoods.

B. Do intangible characteristics of neighbourhoods predict occurrences of crime?

Unable to draw conclusions from spatial features of neighbourhoods, census data is used to examine intangible features of neighbourhoods. Heat-map visualization is employed to generate a two-dimensional view of numerical data, where the data are represented by a range of colours. This allowed grouping similar values by colour, to identify trends in data [9]. Heat-maps that are generated and can be seen in Figures 5, and 6.

Over a range of five possible income brackets, neighbourhoods in the high crime cluster have a significantly higher frequency of household incomes in the lower three income brackets. To verify this, a derived bracket containing the sum of all household incomes less than \$100,000 per year is created. As shown in Figure 2, there is a pronounced difference between percentage of households making under \$100,000 per year between the high and low crime clusters. The percentage of households earning under \$100,000 per year have mean values of 51.97 and 71.55 with standard deviation of 18.26 and 22.15 in the best and worst crime neighbourhoods, respectively. Comparing these cluster's mean values for this attribute confirmed that the differences are statistically significant ($p < 0.001$), indicating a strong difference between the mean values of the high crime and low crime clusters.

The heat-map visualization reveals an inverse trend between neighbourhood crime incidence, and employment rate and/or school enrollment rate. A derived attribute is generated to reduce the dimensions needed to explore this trend. The number of dimensions is reduced, to represent the percentage of individuals not employed, in school, and home-makers. Comparing the mean values of this attribute for high and low crime clusters showed significant differences between the percentage of unemployed and not in school population per neighbourhood, as shown in Figure 3. The percentage of neighbourhood population not working and not in school have a mean of 18.05 and 26.43 with standard deviations of 8.03 and 11.30 in the best and worst crime neighbourhoods, respectively.

A comparison of the cluster's means values for employment/education percentage, using a *t-test*, confirms that these differences are statistically significant ($p < 0.001$) [3]. This indicates a pronounced difference between the employment and education levels of the best and worst crime clusters.

Further investigation by taking the mean values of the percentages for education related dimensions (separate from employment) for the adult population is revealing. Aggregating the high school and post secondary dimensions and performing a t-test on these mean values revealed a statistically significant difference ($p < 0.05$). This indicates that higher education enrollment in a neighbourhood bears a relationship to lower crime levels. In summary, the propensity for crime is lower in neighbourhoods containing high employment and high levels of school enrollment.

VII. Crime Data Analysis Frequent Pattern Analysis

Association rules are formed by searching for frequent if/then associations and using two criteria to identify the most meaningful associations: confidence and support [8].

Confidence identifies the number of times that a potential if/then association is, while support is the frequency with which items appear in the data set being examined. Lift is the ratio of observed support for a set of items to the expected support of those items, if those items were independent. An association rule mining is used to find associations, frequent patterns, or causal relationships from data sets [7].

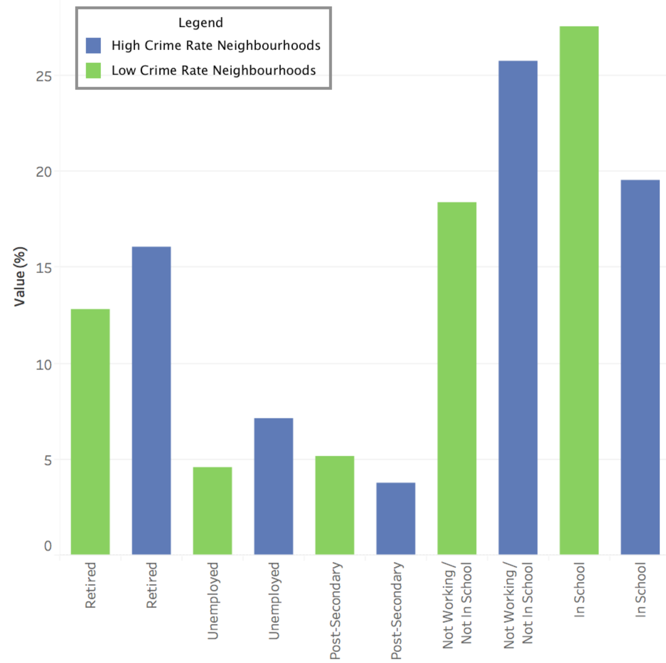

Fig. 3: Comparison of employment status categories between high and low crime neighbourhoods.

A. *Are there any characteristics of neighbourhoods that predict the type of crime incidences?*

For a given neighbourhood, spatial characteristics are used as inputs and crime types are used as a predictable. Exploring the set of rules generated by the algorithm gave little to no insight into any potential relationships. Generated rules have low probabilities and low lift, indicating low confidence. Examining the item sets generated by the algorithm does not provide any more insight than the rule set; these frequent item sets have low support and consist of dimensions irrelevant to the question. Considering the results from the clustering algorithm also show no meaningful relationship between spatial characteristics and crime type, this provides a cross-validation for both algorithms.

#	Probability	Lift	Rule
1	1.000	1.29	30k To 60k $= 20 - 29$, 60k To 100k $< 16 - >$ Assault $= 34 - 66$

TABLE 2: Association rules for low income brackets. The association between low (\$30,000 to \$100,000 per year) household incomes and high assault is high, confirmed with high probability and lift values. Within rules, income bracket items denote percent of total neighbourhood population and crime type items denote an integer range of crime incidence.

B. *Violent vs Non-violent Crime*

Running the same algorithm on each neighbourhood with average property value and average household income as inputs and crime types as the predictable provides interesting results. Based on the rules and frequent item sets generated, neighbourhoods with lower average household income have

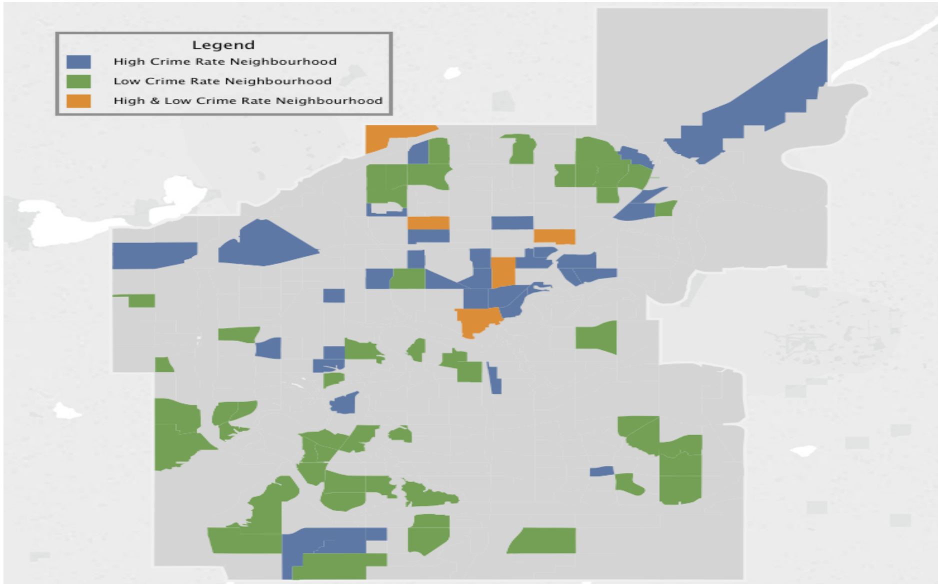

Fig. 4: Visualization of the spatial distribution of neighbourhoods in the best crime and worst cluster.

#	Support	Size	Item Set
1	8	2	Assault = 34 − 66, 60k To100k < 16
2	7	3	Assault = 34 − 66, Less Than30k = 28 − 40, 125k To150k < 3
3	7	3	Assault = 34 − 66, Less Than30k = 28 − 40, 60k To100k < 16
4	10	3	Robbery = 4+, 30k To60k = 20 − 29, 60k To100k = 25 − 32

TABLE 3: Frequent item sets occurring in low income neighbourhoods. In general, neighbourhoods characterized to have low household income frequently have high assault and robbery incidence, relative to neighbourhoods with higher household income. Within rules, income bracket items denote a percentage of total neighbourhood population while crime type items denote an integer range of crime incidence.

more violent crime than neighbourhoods with higher average household income as shown in Tables 2 and 3.

The contrary is observed in high crime neighbourhoods - a lower rate of violent crimes, shown in Tables 4 and 5. This is indicated by the lower probability and lift for rules containing this attribute in higher income neighbourhoods as compared to lower income neighbourhoods.

C. Variations in Types of Crime

All neighbourhoods have at least one occurrence of every type of crime (with the exception of homicide) so analysis of frequent item sets is done for those that have in the top 10% support. Comparison between the frequent item sets shows that the variation in types of crimes is the largest for the neighbourhoods that have a larger percentage of households in the $100,000 to $150,000 income bracket.

This shows that neighbourhoods having a higher percentage of their population earning a household income between $100,000 and $150,000 will experience a higher variance in crime type. This is shown by comparing the frequent item sets in Tables 6 and 7. This was an interesting and surprising result, due the results of the clustering methods where the overall crime occurrence was higher in very low income

#	Probability	Lift	Rule
1	1.000	1.16	200k To250k >= 15, 250k Or More = 7 − 16 − > Theft Of Vehicle = 31+
2	1.000	0.71	250k Or More = 16 − 23, 200k To250k = 6 − 9 − > Break And Enter = 31+

TABLE 4: Association rules for high income brackets. The association between high ($200,000+ per year) household incomes and non-violent crime is high, confirmed by the high probability and lift values. Within rules, income bracket items denote the percentage of neighbourhood population in that bracket, and crime type items denote an integer range of crime incidence.

#	Support	Size	Item Set
1	12	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Theft Of Vehicle = 16 − 30
2	10	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Break And Enter = 16 − 30
3	41	3	250k Or More = 2 − 7, 200k To250k = 2 − 6, Theft From Vehicle = 26 − 50

TABLE 5: Frequent item sets occurring in high ($150,000+ per year) income neighbourhoods. In general, neighbourhoods having high household incomes frequently contain high break and enter and theft involving vehicle incidence, relative to neighbourhoods with fewer high income households. Within rules, income bracket items denote the percentage of neighbourhood population in that bracket, and crime type items denote an integer range of crime incidence.

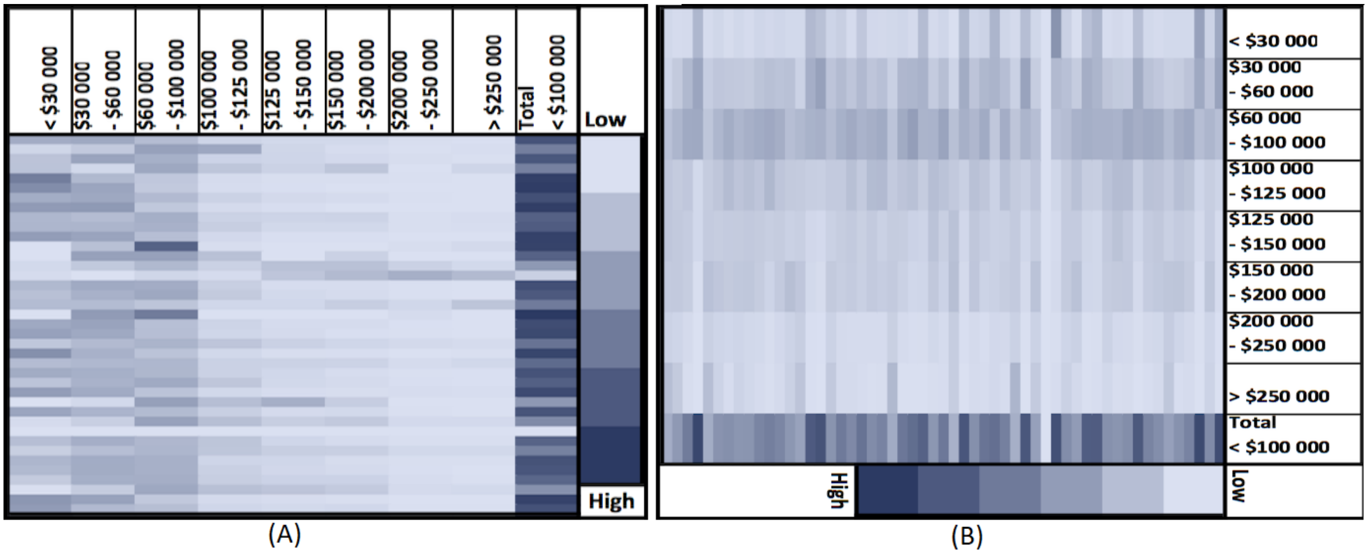

Fig. 5: Heat-map of yearly household income percentage for (A). the low crime neighbourhood cluster. (B) high crime neighbourhood cluster.

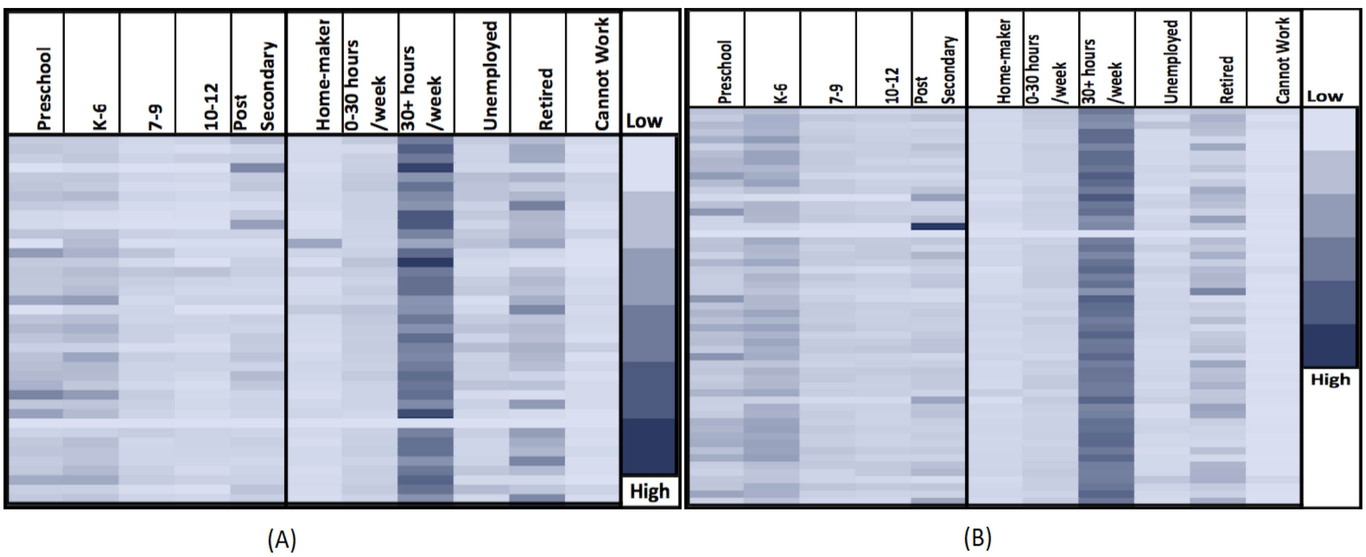

Fig. 6: Heat-map of employment and education percentage for (A) the low crime neighbourhood cluster. (B) the high crime neighbourhood cluster.

neighbourhoods, where attention was paid to the variance of crime type.

VIII. Additional Discussion

During the creation of the high and low crime neighbourhood clusters, the neighbourhoods are superimposed onto a map of the City of Edmonton. This shows the spatial distribution of the best and worst neighbourhoods in relation to each other. In Figure 4 it is apparent that the higher crime neighbourhoods are situated closer to the center of the city when compared to Figure 5, where the lower crime neighbourhoods are situated closer to the city limits. The result raises additional questions about the location of the high and low crime neighbourhoods. Future research includes exploration of how location within a city, population density, neighbourhood age, or other factors may impact crime incidence.

IX. Conclusion

Thanks to the City of Edmonton, Canada Open Data Portal, an opportunity was provided to analyze and interpret crime and neighbourhood data. Most spatial characteristics have no correlation to occurrences of neighbourhood crime. However, certain dwelling types in neighbourhoods have a statistically significant correlation with crime incidence. In neighbourhoods with a higher percentage of apartments under five stories in height, crime incidence was increased. Additionally, tree density have a similar statistic significance, as neighbourhoods

#	Support	Size	Item Set
1	48	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Assault = 1 − 33
2	46	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Theft From Vehicle = 26 − 50
3	41	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Sexual Assault = 1 − 8
4	36	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Break And Enter = 16 − 30
5	28	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Theft Of Vehicle = 16 − 30
6	23	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Robbery = 1 − 2
7	22	3	100k To125k = 11 − 16, 125k To150k = 7 − 12, Theft Over Five = 1 − 8

TABLE 6: Frequent item sets occurring in middle ($100,000 to $150,000 per year) household income neighbourhoods. In general, neighbourhoods characterized as having middle household incomes frequently contain a high incidence of all varieties of crimes, relative to the number of these crimes in neighbourhoods with both high and low income households. Within rules income bracket items denote percentage of total neighbourhood population and crime type items denote an integer range of crime incidence.

#	Support	Size	Item Set
1	15	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Sexual Assault = 1−8
2	15	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Assault = 1−33
3	15	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Theft From Vehicle = 26−50
4	12	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Theft Of Vehicle = 16−30
5	10	3	150k To200k = 3 − 7, 200k To250k = 2 − 6, Break And Enter = 16−30

TABLE 7: Frequent item sets containing high ($200,000+ per year) household income neighbourhoods. In general, neighbourhoods characterized as having high household incomes frequently contain a lower incidence of all crime types, relative to the number of these crimes in neighbourhoods with both high and low income households.

with more trees experienced reduced crime incidence compared to those with less trees.

While spatial characteristics have little bearing on crime incidence, intangible neighbourhood characteristics have a greater effect. Several of these intangible characteristics that are identified to be a strong predictor of crime incidence are employment status, school enrollment, and household income levels. Neighbourhoods with higher employment, school enrollment, and/or household income correlate strongly with lower overall levels of crime.

Some of these neighbourhood characteristics are also predictors of crime type, in addition to incidence rate. neighbourhood household income level is a predictor of crime type (violent or non-violent) as well as crime type variance. Neighbourhoods having a lower average household income experienced more potentially violent crimes such as robbery, assault, and break and enter compared to neighbourhoods with higher average household income. Outliers are identified, with certain household income brackets, such as the $100,000 - $150,000 range, experienced a marked variance in crime types compared with other brackets.

ACKNOWLEDGMENT

The authors would like to thank the City of Edmonton, Alberta, Canada for providing the data used in this research.

REFERENCES

[1] "City of Edmonton Open Data Portal, City of Edmonton, 2017" https://data.edmonton.ca, last accessed: Nov 2017.

[2] C.H. Yu, M. Ward, M. Morabito, and W. Ding, *Crime Forecasting Using Data Mining Techniques*. Volume 3, Number 2, Issue 10, 2013.

[3] G. Rochefort-Maranda, *On the correct interpretation of p values and the importance of random variables*. Synthese, Volume 193, Issue 6, 2016.

[4] H. Gupta and V. Shrivastava, *Review of various Techniques in Clustering*. Volume 3, Number 2, Issue 10, 2013.

[5] H. He, H. Zhang, H. Luo, *Design of Multi-level Teaching System Based on Association Rule Mining*. International Journal of Emerging Technologies in Learning, Volume 11, 2016.

[6] H. Nguyen, E. Mller, J. Vreeken, K. Bhm, *Unsupervised interaction-preserving discretization of multivariate data*. Data Min Knowl Disc, 2014.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Morgan Kaufmann, 3rd ed., 2012.

[8] J. Hipp, U. Gntzer, G. Nakhaeizadeh, *Algorithms for Association Rule Mining A General Survey and Comparison*. SIGKDDExplorations, ACM SIGKDD, 2000.

[9] J. Gu, M. Pitz, S. Breitner, W. Birmili, S. von Klot, A. Schneider, J. Soentgen, A. Reller, A. Peters, J. Cyrys, *Selection of key ambient particulate variables for epidemiological studies Applying cluster and heatmap analyses as tools for data reduction*. Science of The Total Environment, Supplement C, 2012.

[10] Microsoft, *Microsoft Association Algorithm Technical Teference*. https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-association-algorithm-technical-reference

[11] Microsoft, *SQL Server Analysis Services*. https://docs.microsoft.com/en-us/sql/analysis-services/instances/install-windows/install-analysis-service

[12] Microsoft *Visual Studio 2017*. https://imagine.microsoft.com/en-US/Catalog/Product/530 last accessed: Nov 2017.

[13] R. Wall Emerson, *Journal of Visual Impairment & Blindness*. Volume 3, Number 2, Issue 10, 2013.

[14] S. Arietta, A. Efros, R. Ramamoorthi, and M. Agrawala, *City Forensics: Using Visual Elements to Predict Non-Visual City Attributes*. Volume 3, Number 2, Issue 10, 2013.

[15] *Python Shapely*. https://pypi.python.org/pypi/Shapely, last accessed: Nov 2017.