Association Rules between Crime, Shelters and Social-economic Factors in the City of Toronto

INF 2190: Data Analytics

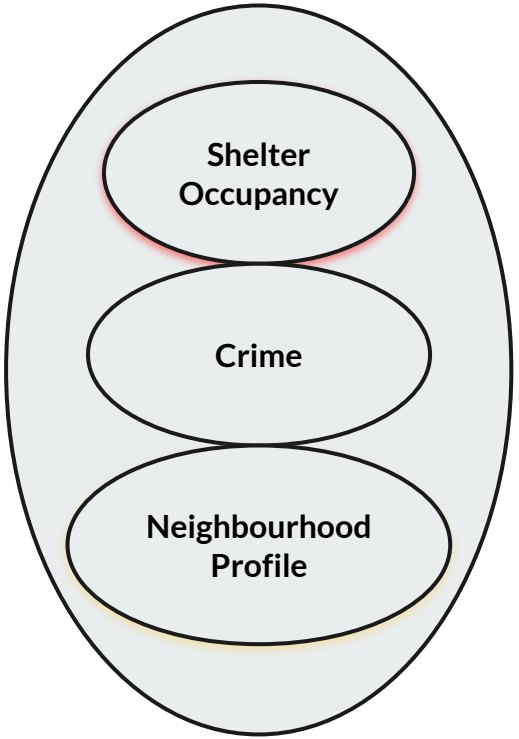
Group 12 - Komal, Lauren (Seim), Mounica, Ngoné, Rutvik

Introduction, Problem Definition & Motivation

- Does a neighbourhood with higher employment rate have less crime?
- Does having **higher shelter capacity** or/and **occupancy** translate to **less crime**?
- How do various **social-economic factors** affect crime rate?
- Can we identify potential dimensions and rules governing the relationship between **crimes** and **shelters** in Toronto?

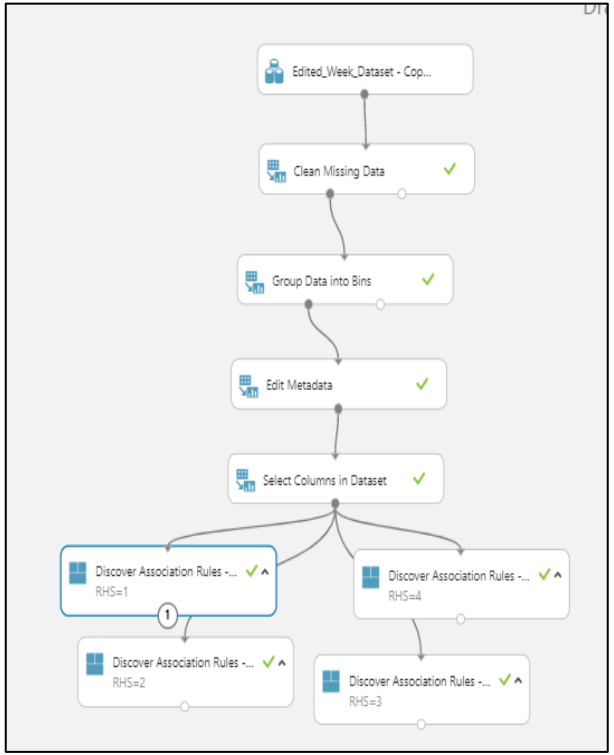
Brief Description of the Data Sets

- **Shelter**: Neighbourhood, Occupancy, and Capacity
- **Crime**: Assault, Auto-Theft, Break-and-Enter, Robbery, and Theft-Over, Overall Crime
- **Neighbourhood Profile**: Population, Household, Income and Education
- **Final Dataset**:
  - ➔ Integration by Neighbourhood and Date
  - ➔ Aggregation by Week and Neighbourhood.
  - ➔ 33 attributes and 1596 tuples (31 Neighbourhoods)
  - ➔ Format: Neighbourhood, Week, Shelter Attributes, Crime Attributes, Socio-Economic Attributes



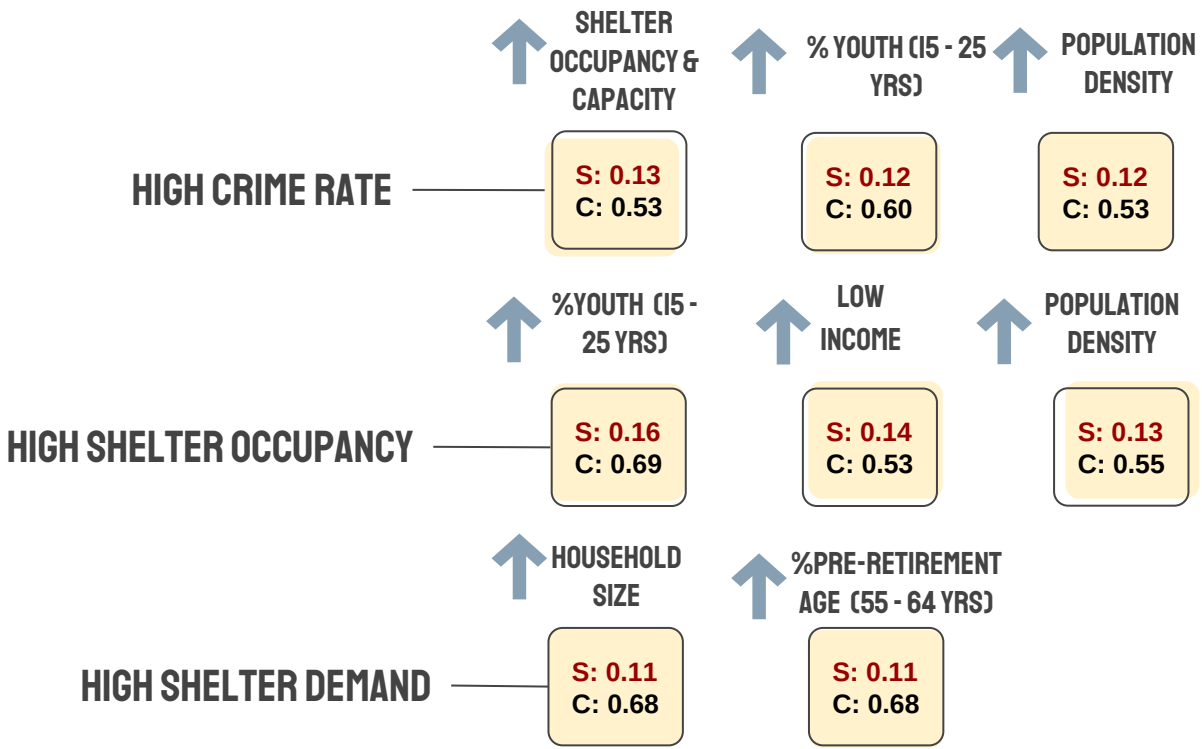
Data Analysis Method/Task

- **Step 1: Data Cleaning**
  - ➔ Removing tuples, filling in empty & zero cells, smoothing outliers: mean value and linear regression methods
- **Step 2: Data Integration & Aggregation**
- **Step 3: Data Reduction**
  - ➔ Dimensionality: Identify variables that are correlated to compress representation of the original data
  - ➔ Missing Data: Probability PCA
  - ➔ Binning: Quantiles with 4 Bins. Skewed Shelter and Crime variables
  - ➔ Transformation: Numeric variables into categorical variables
  - ➔ Data Selection: of columns
- **Step 4: Data Mining**
  - ➔ **Association Rule** - Min Support = 0.1 & Min Confidence = 0.5
  - ➔ RHS = crime & shelter variable



Experimental Results

*Minimum Support → 10 %
*Minimum Confidence → 50%
*Average Running Time → 67s



- **Four major crime variables** (i.e. breaking and entering, auto-theft, robberies, theft over) were associated with one another
- **Assault** was not associated with any of the other crime variables

Discussion

What did we learn?

- High shelter occupancy, high shelter capacity, high population density, and high number of young adults were found to be associated with high crime.
- Low number of shelters were found to be associated with crimes: Break & Enter, Robbery, Auto-theft
- No association rules relating employment and crime were found.

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.699e+00  2.820e+00   2.021  0.04349 *
TOTAL_CAPACITY     -7.535e-03  4.064e-03  -1.854  0.06389 .
TOTAL_OCCUPANCY     1.138e-02  4.130e-03   2.756  0.00592 **
Population_density  9.856e-05  6.314e-05   1.561  0.11875
Youth               3.464e+02  1.787e+01  19.391  < 2e-16 ***
Children           -1.554e+02  1.303e+01 -11.921  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.91 on 1503 degrees of freedom
Multiple R-squared:  0.3863,    Adjusted R-squared:  0.3843
F-statistic: 189.2 on 5 and 1503 DF,  p-value: < 2.2e-16
```

Next Steps:

- Examining frequently occurring itemsets specific (including socio-economic variables) to safe and unsafe neighbourhoods.
- Visualise this data.
- Find discrepancies between safe and unsafe neighbourhood-specific itemsets. Create action-oriented intelligence..

Data Sources

https://open.toronto.ca/dataset/daily-shelter-occupancy/

http://data.torontopolice.on.ca/datasets/mci-2014-to-2018?orderBy=occurrencedate&where=occurrencedate%20%3E%3D%20TIMESTAMP%20%272017-01-01%2000%3A00%3A00%27%20AND%20occurrencedate%20%3C%3D%20TIMESTAMP%20%272017-12-31%2023%3A59%3A59%27

https://open.toronto.ca/dataset/neighbourhood-profiles/