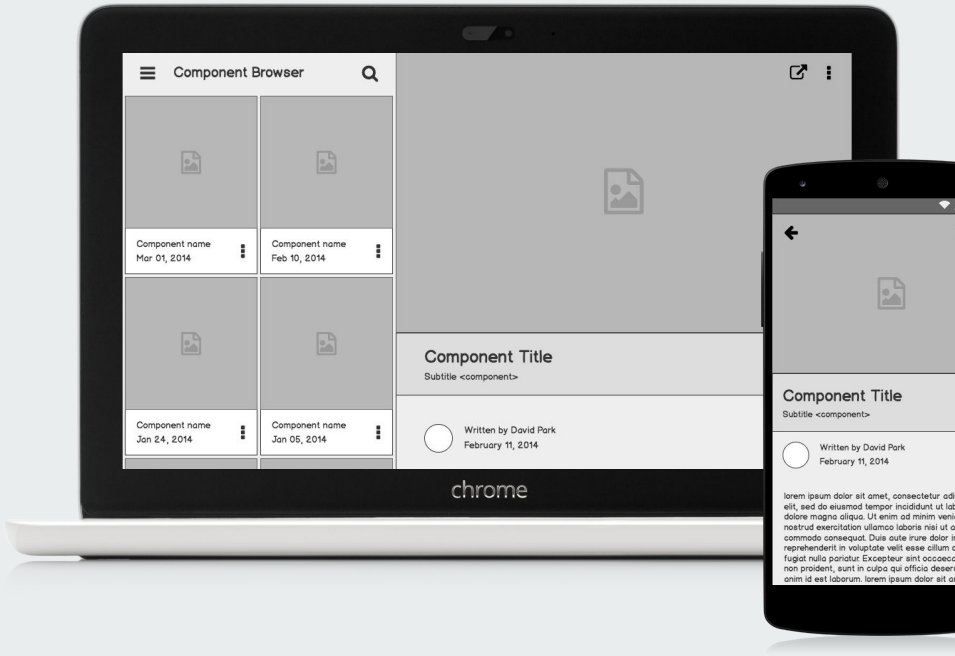
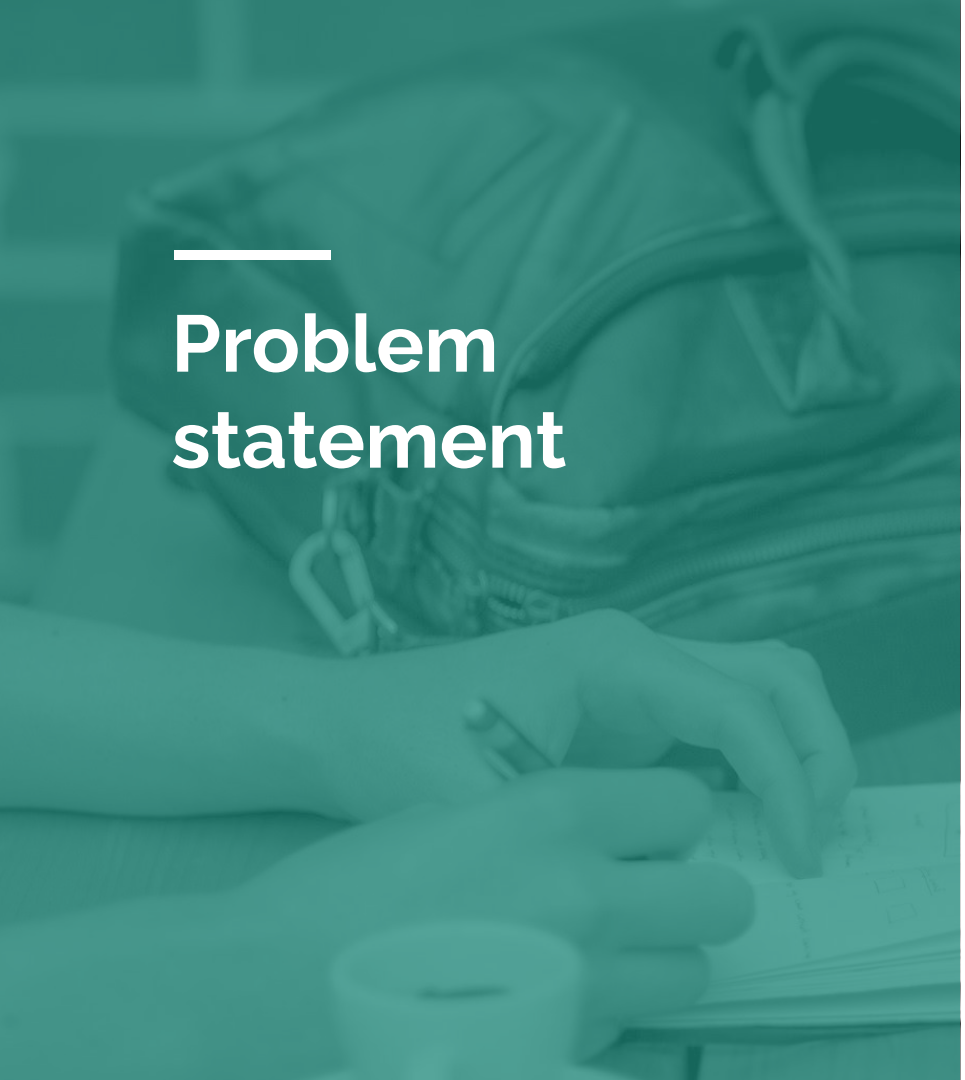
AI Final Project Group 11

Sarah Long
Tasfia Alam
Vidurshan Sribalasuhabiramam



Outline

- Introduction: Background and Business Problem
- Data explanation and data source
- Methodology
- Results



Problem statement

In this project, we decided to create a program to analyze the data patterns between the number of crimes related to theft and income levels in the various neighborhoods in Toronto.

Data Sources

01

Neighborhood Data:

Open-Data Toronto's Neighborhood profiles provides

- Postal codes
- Neighborhood names
- Boroughs

Data Sources

02

Income Statistics for Households:

The household income after taxes for the year of 2015 was acquired from the Open-Data Toronto's Neighborhood profiles

Crime Statistics:

The three crimes we chose to cluster were

- Robberies
- Auto theft
- Theft over $5000

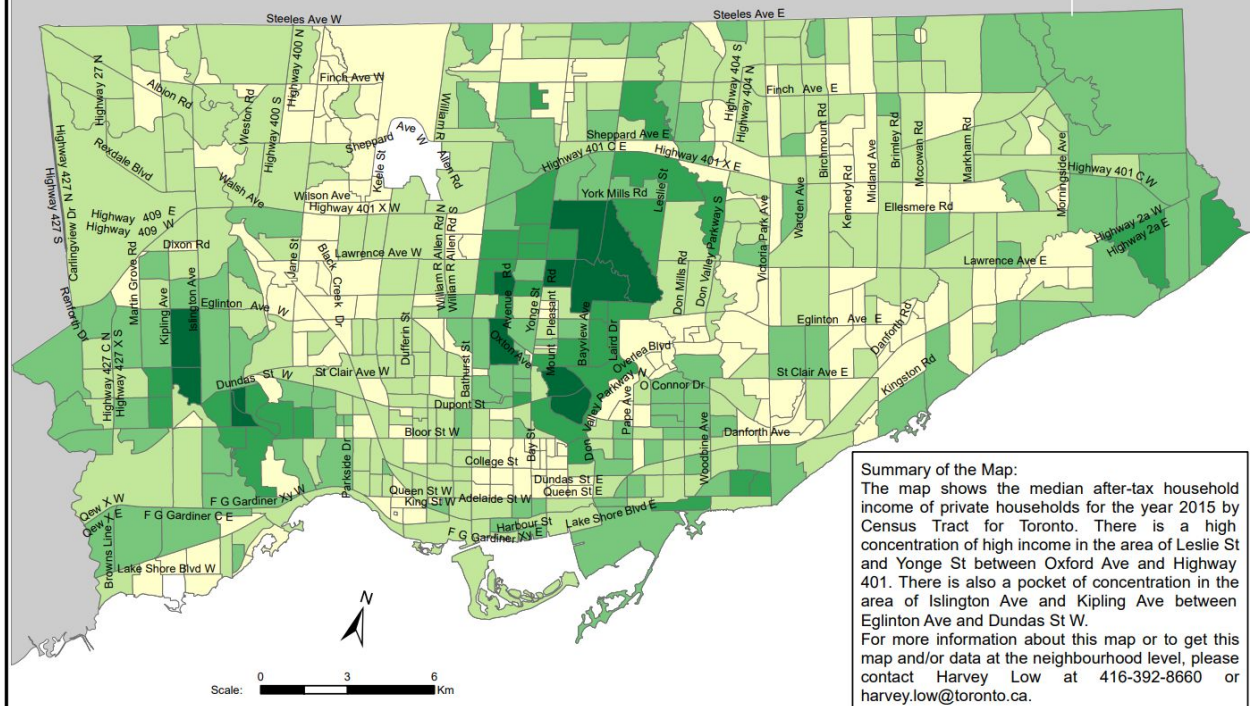
Methodology

Methodology

- K-means clustering would be the best method to cluster neighbourhoods in terms of median after-tax household income levels and crime rates.

- On the next slide you will see a sample heatmap of the median after-tax household income for 2015 that shows what neighborhoods are in which income bracket.

City of Toronto Income - Median After-Tax Household Income 2015



Summary of the Map:
The map shows the median after-tax household income of private households for the year 2015 by Census Tract for Toronto. There is a high concentration of high income in the area of Leslie St and Yonge St between Oxford Ave and Highway 401. There is also a pocket of concentration in the area of Islington Ave and Kipling Ave between Eglinton Ave and Dundas St W.
For more information about this map or to get this map and/or data at the neighbourhood level, please contact Harvey Low at 416-392-8660 or harvey.low@toronto.ca.

Median After Tax income
- $21,941 - $53,018
- $53,019 - $70,059
- $70,060 - $94,185
- $94,186 - $130,816
- $130,817 - $206,336

- Not available/ No data
- Expressway
- Major Arterial

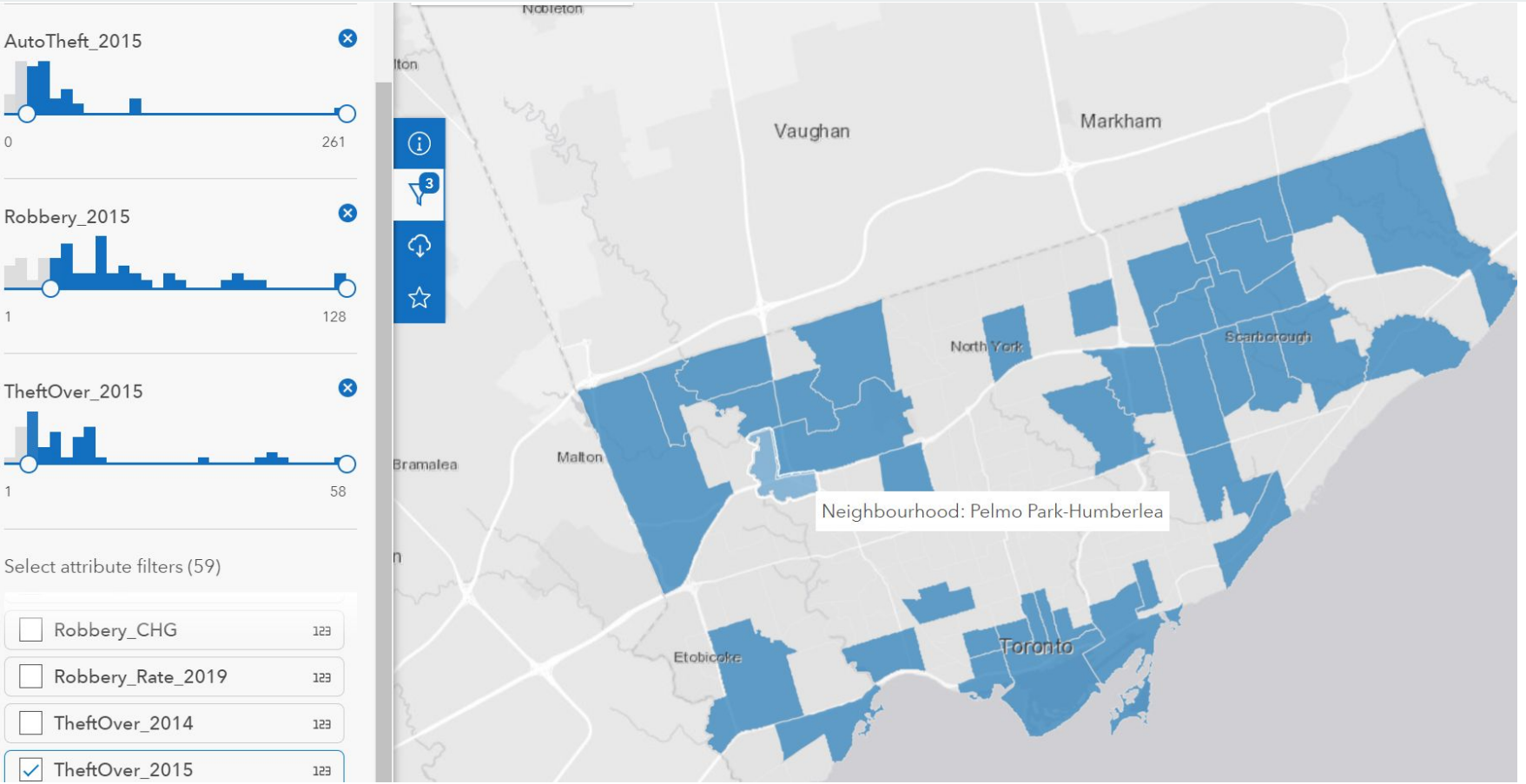
Source: City of Toronto;
Statistics Canada 2016 Census.

Published: May 2018.

Prepared by: Social Research and Analysis Unit.
Contact: spar@toronto.ca.

Methodology

- The crime frequency for auto-theft, robberies, and theft over $5000 are shown below in a sample heatmap. The highlighted neighbourhoods are those with a frequence of any of the three crimes above the city of Toronto’s median crime rate for each category.

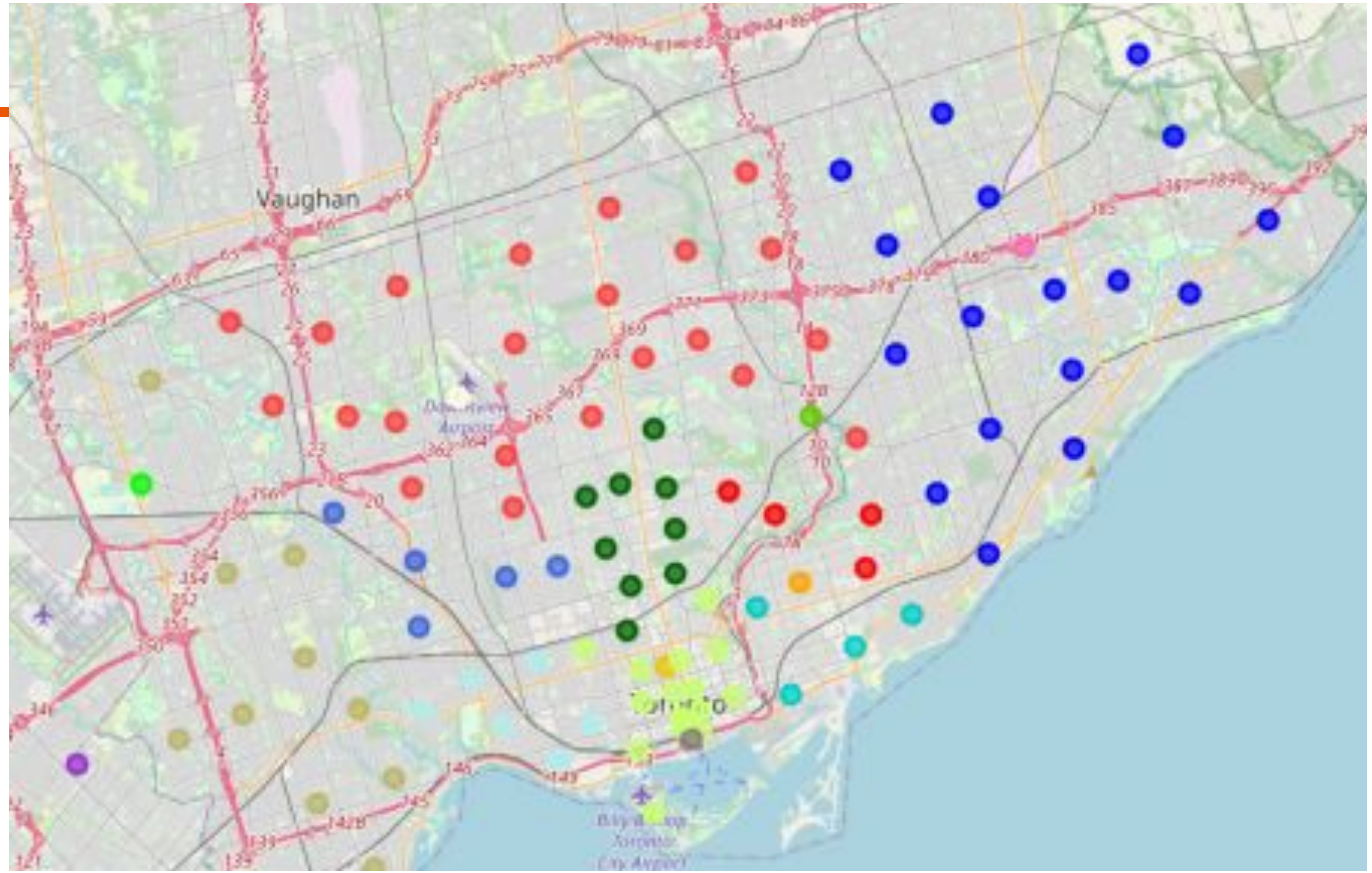

AutoTheft_2015
0
261
Robbery_2015
1
128
TheftOver_2015
1
58
Select attribute filters (59)
Robbery_CHG
Robbery_Rate_2019
TheftOver_2014
TheftOver_2015
Vaughan
Markham
North York
Scarborough
Malton
Bramalea
Etobicoke
Toronto
Neighbourhood: Pelmo Park-Humberlea

K-Means Clustering

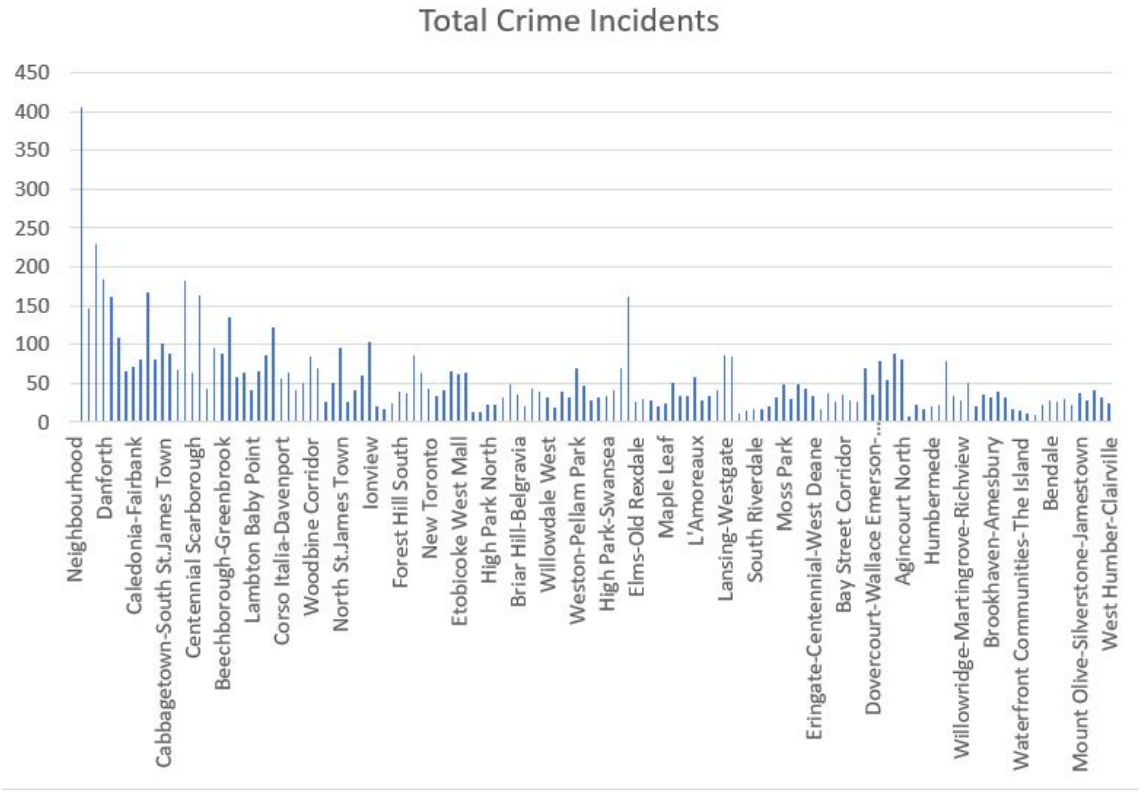
Each neighbourhood's data point in relation to the next would be determined using the Euclidean distance between each other. K-means clustering would cluster the data points based on the minimum variance between points, so as to minimize the within-cluster sum of squares (WCSS).

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

Results


Vaughan

Results


Total Crime Incidents
450
400
350
300
250
200
150
100
50
0
Neighbourhood
Danforth
Caledonia-Fairbank
Cabbagetown-South St.James Town
Centennial Scarborough
Beechborough-Greenbrook
Lambton Baby Point
Corso Italia-Davenport
Woodbine Corridor
North St.James Town
Ionview
Forest Hill South
New Toronto
Etobicoke West Mall
High Park North
Briar Hill-Belgravia
Willowdale West
Weston-Pellam Park
High Park-Swansea
Elms-Old Rexdale
Maple Leaf
L'Amoreaux
Lansing-Westgate
South Riverdale
Moss Park
Eringate-Centennial-West Deane
Bay Street Corridor
Dovercourt-Wallace Emerson-...
Agincourt North
Humbermede
Willowridge-Martingrove-Richview
Brookhaven-Amesbury
Waterfront Communities-The Island
Bendale
Mount Olive-Silverstone-Jamestown
West Humber-Clairville

Discussion and Conclusion

In conclusion, neighbourhoods with lower median after-tax income had higher rates of crime. Even though it would be expected that higher income neighbourhoods would be the victim of theft and robbery crimes, the analysis proved otherwise. K-means clustering provided the means to analyze the data and explore the correlation between income levels and crime rates.

Questions?