2009-2016 Analysis

To develop a problem statement regarding the consistently high Violent Crime Severity Index in the City of Edmonton, we analyzed data from Stats Canada's Canada Crime Index for the years 1998-2021. This dataset provided statistics on various crime severity indexes, including overall crime severity, violent crime severity, non-violent crime severity, youth crime severity, and youth violent crime severity. Initially, we focused on the overall crime severity index for the province of Alberta and its two major cities, Edmonton, and Calgary.

Our analysis revealed that between 2000 and 2004, the crime severity index for Edmonton was significantly higher than both the Alberta and Calgary indexes, as per figure 1, below. However, from 2004 onwards, all indexes began to decline, with the Edmonton index eventually aligning with the Alberta index in 2012 and remaining on a similar trajectory thereafter. It is worth noting that throughout this period, the Calgary index consistently remained much lower than both the Edmonton and Alberta indexes.

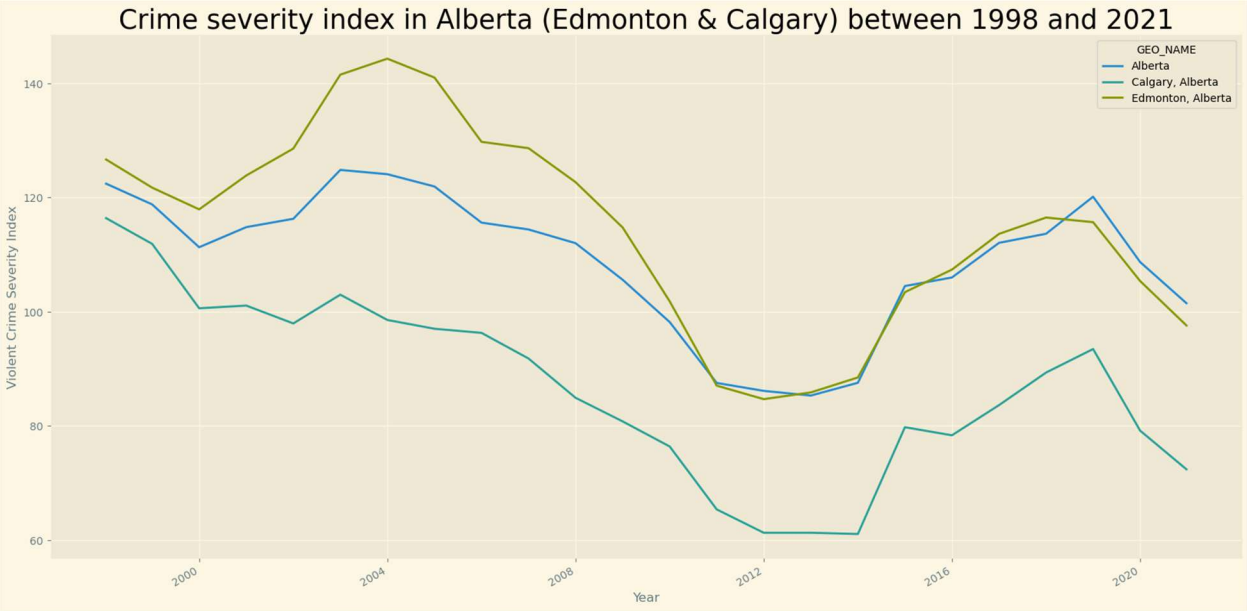

Figure 1 Overall Crime Severity Index for Alberta, Calgary, and Edmonton.

After analyzing the overall crime severity index data, we focused specifically on the Violent Crime Severity Index for Alberta, Edmonton, and Calgary. To our surprise, we found that the Edmonton index was consistently higher than both the Alberta and Calgary indexes. Even during the period between 2008 and 2013, when all three indexes showed a decline, the Edmonton index remained higher than the overall provincial index and significantly higher than the Calgary index.

As showing in Figure 2, below, although there was a sharp drop in all three indexes in 2016, the Edmonton index continued to be consistently higher than the Alberta and Calgary indexes, indicating that the city was facing a significant challenge in addressing violent crime.

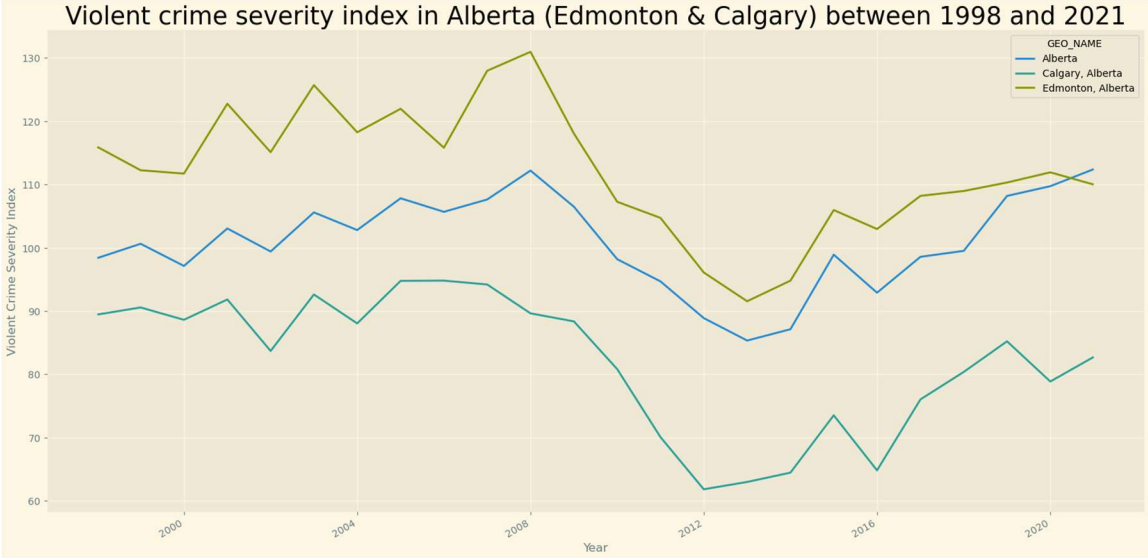

Figure 2 Violent Crime Severity Index for Alberta, Calgary, and Edmonton.

To start our analysis, we gathered crime data statistics from the Edmonton Police Department data portal. The dataset contained several columns, including 'NGH_Name', 'NGH_Number', 'Latitude', 'Longitude', 'Violation_Type', 'Sum_Occurences', 'DT_Year', and 'DT_Month'. The 'NGH_Name' column represented the neighborhood, while the 'Violation_Type' column included categories such as Theft from Vehicle, Break and Enter, Assault, Theft of Vehicle, Robbery, Sexual Assaults, Theft Over $5000, and Homicide.

For our preliminary analysis, we focused primarily on severe crimes such as Assault, Robbery, Sexual Assaults, and Homicide, which constitute the Severe Crime Index as per Statistics Canada. By analyzing the occurrences of these crimes in different neighborhoods, we aimed to identify patterns and trends in crime across the city of Edmonton.

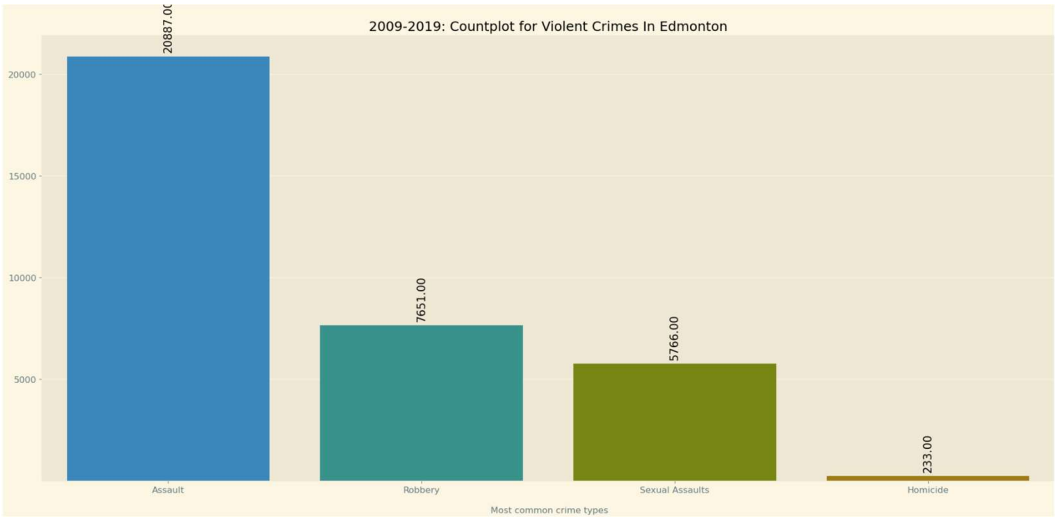

Figure 3: 2009-2019 Countplot for Violent Crimes in Edmonton.

To gain an overview of severe crime trends in Edmonton between 2009 and 2019, we plotted a Countplot of the total occurrences of severe crimes, which included Assault, Sexual Assaults, Robbery,

and Homicide. Figure 3 depicts the results, showing that Assault was the most prevalent severe crime with a count of 20,887, followed by Robbery with 7,651, Sexual Assaults with 5,766, and Homicide with 233. This initial analysis provided a baseline understanding of the types of severe crimes occurring in Edmonton and their relative frequencies. We also see that the crime of Assault capture 60% of the proportion all severe crimes in Edmonton.

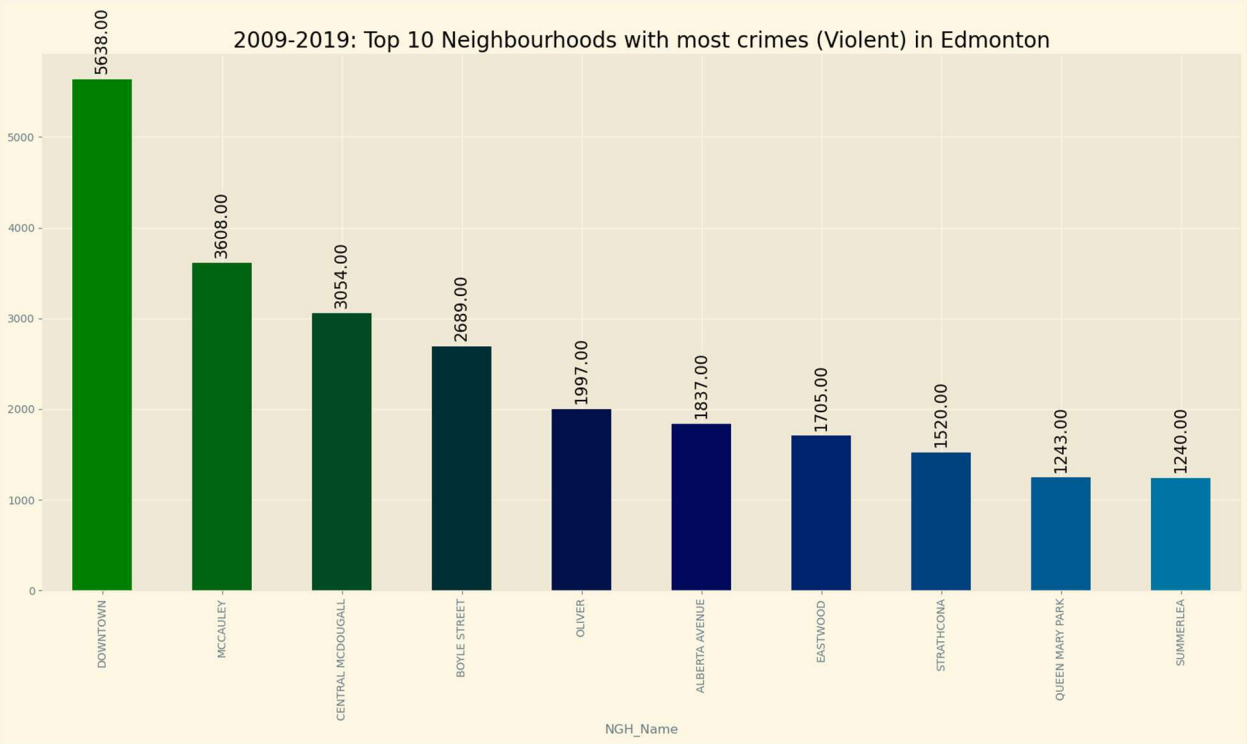

Figure 4: 2009-2019 Top 10 Neighborhoods with highest Violent Crimes in Edmonton.

To further analyze severe crime trends in Edmonton, we investigated which neighborhoods had the highest occurrences of severe crimes. Based on our analysis, we plotted a graph of the top 10 neighborhoods based on their total counts of severe crimes. The results from figure 4, indicated that Downtown had the highest count at 5,638, followed by McCauley with 3,608, Central McDougall with 3,054, Boyle Street with 2,689, Oliver with 1,997, Alberta Avenue with 1,837, Eastwood with 1,705, Strathcona with 1,520, Queen Mary Park with 1,243, and rounding off the top 10 was Summerlea with 1,240. This analysis revealed that severe crimes were most prevalent in certain neighborhoods in Edmonton and may provide insights for targeted interventions and crime prevention strategies.

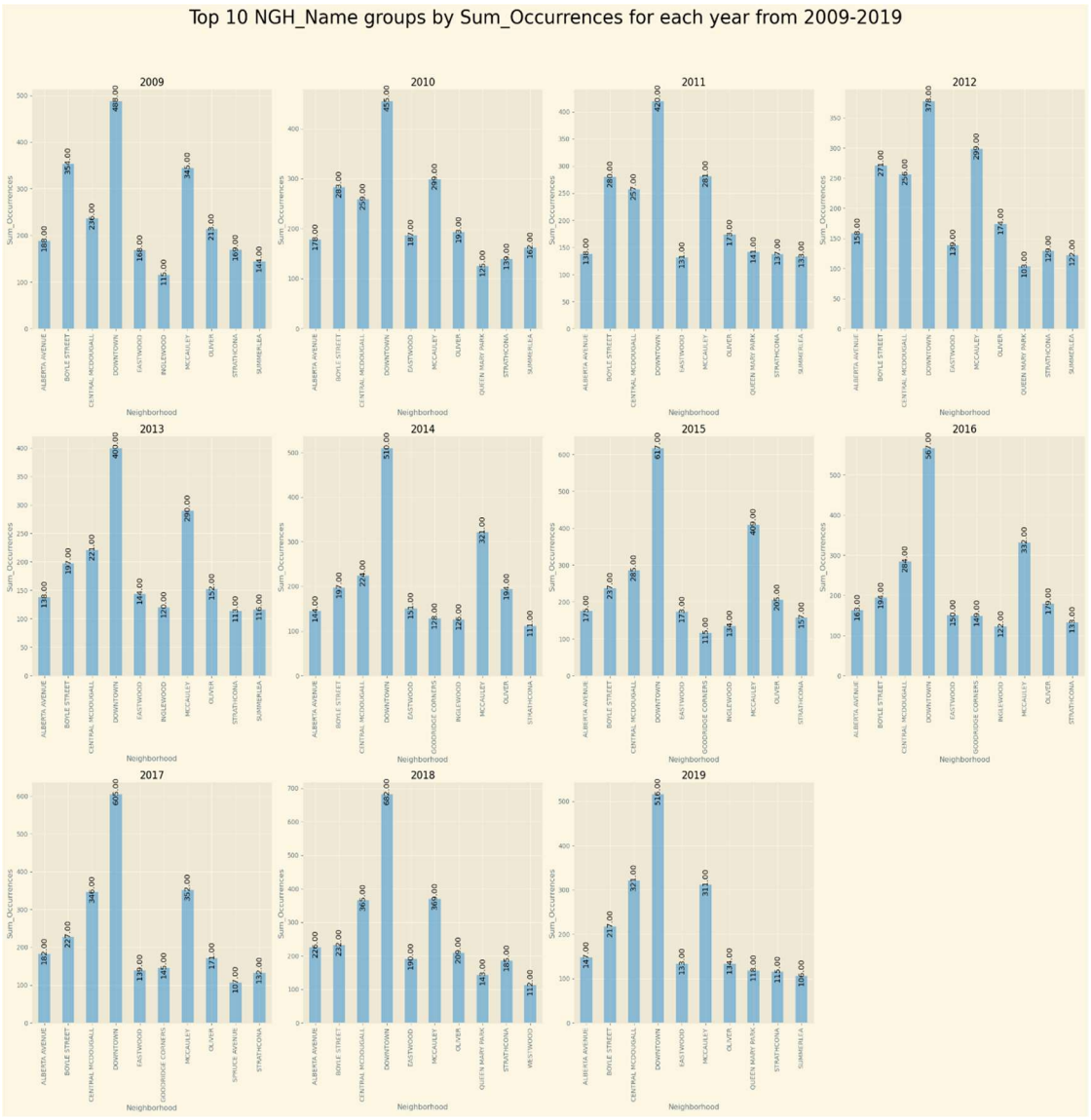

Figure 5: Top 10 Neighborhoods for all the years between 2009 and 2019.

To gain a deeper understanding of severe crime trends in Edmonton neighborhoods, we plotted charts for the top 10 neighborhoods each year between 2009 and 2019. The goal was to identify any consistent patterns or neighborhoods that consistently appeared in the top 10. By doing so, we could provide insights into which neighborhoods experienced high occurrences of severe crimes over time and could potentially benefit from targeted interventions and crime prevention strategies.

The results of this analysis indicated that certain neighborhoods consistently appeared in the top 10 each year, while others showed more fluctuation. This information provided a basis for further investigation into the factors that may contribute to severe crime in these neighborhoods and potential strategies for addressing them.

Based on our analysis of severe crime trends in Edmonton neighborhoods between 2009 and 2019, certain neighborhoods consistently appeared in the top 10 each year. From figure 5, we see that these neighborhoods include Downtown, Alberta Avenue, Boyle Street, Central McDougall, McCauley, and Oliver. This information can provide valuable insights into areas that may require targeted interventions and crime prevention strategies to improve public safety and reduce severe crime occurrences.

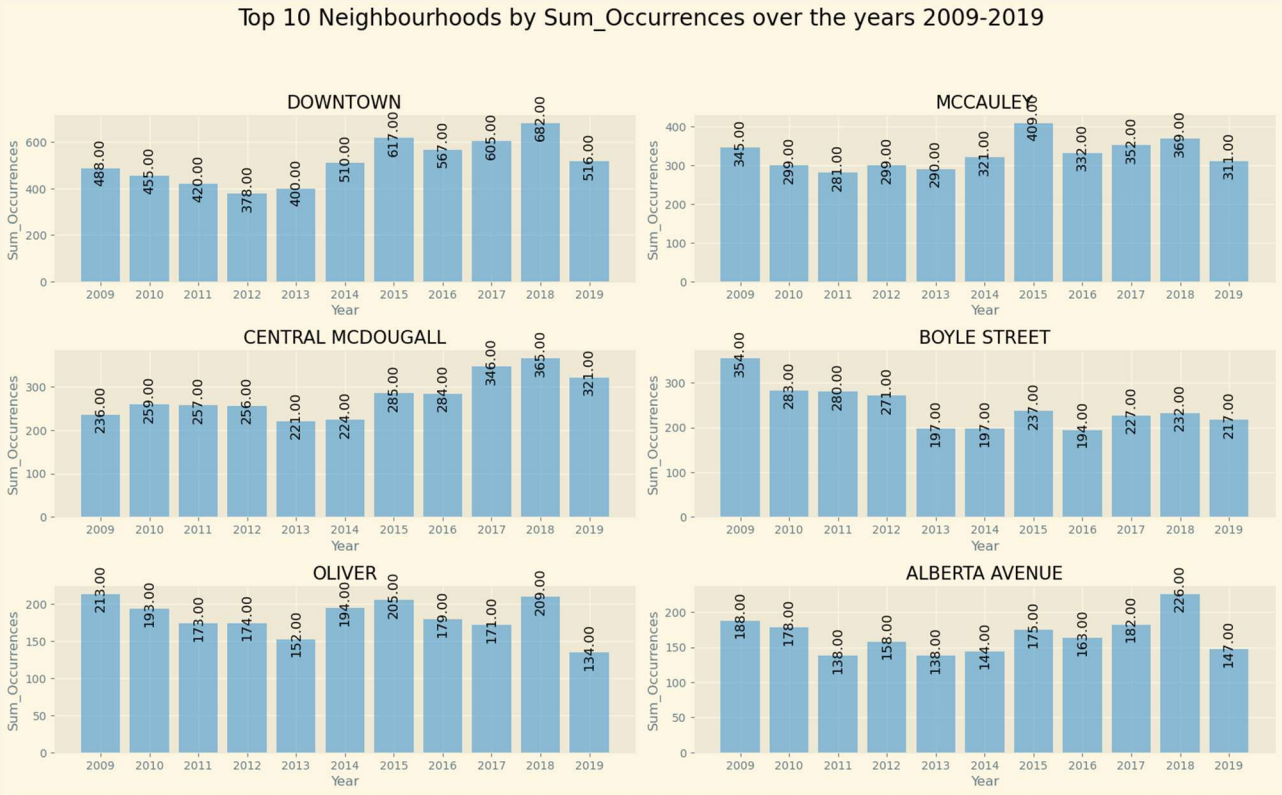

Figure 6: Top 6 Neighborhoods between 2009-2019.

To gain a more detailed understanding of severe crime trends in Edmonton neighborhoods, we decided to focus on the six neighborhoods that consistently appeared in the top 10 between 2009 and 2019. These neighborhoods included Downtown Alberta Avenue, Boyle Street, Central McDougall, McCauley, and Oliver. As shown in figure 6, we then plotted the severe crime rates for each of these neighborhoods individually between 2009 and 2019 to identify any patterns or trends in severe crime occurrences. This analysis allowed us to better understand the severity and frequency of crimes in these neighborhoods over time and provided a foundation for targeted interventions and crime prevention strategies.

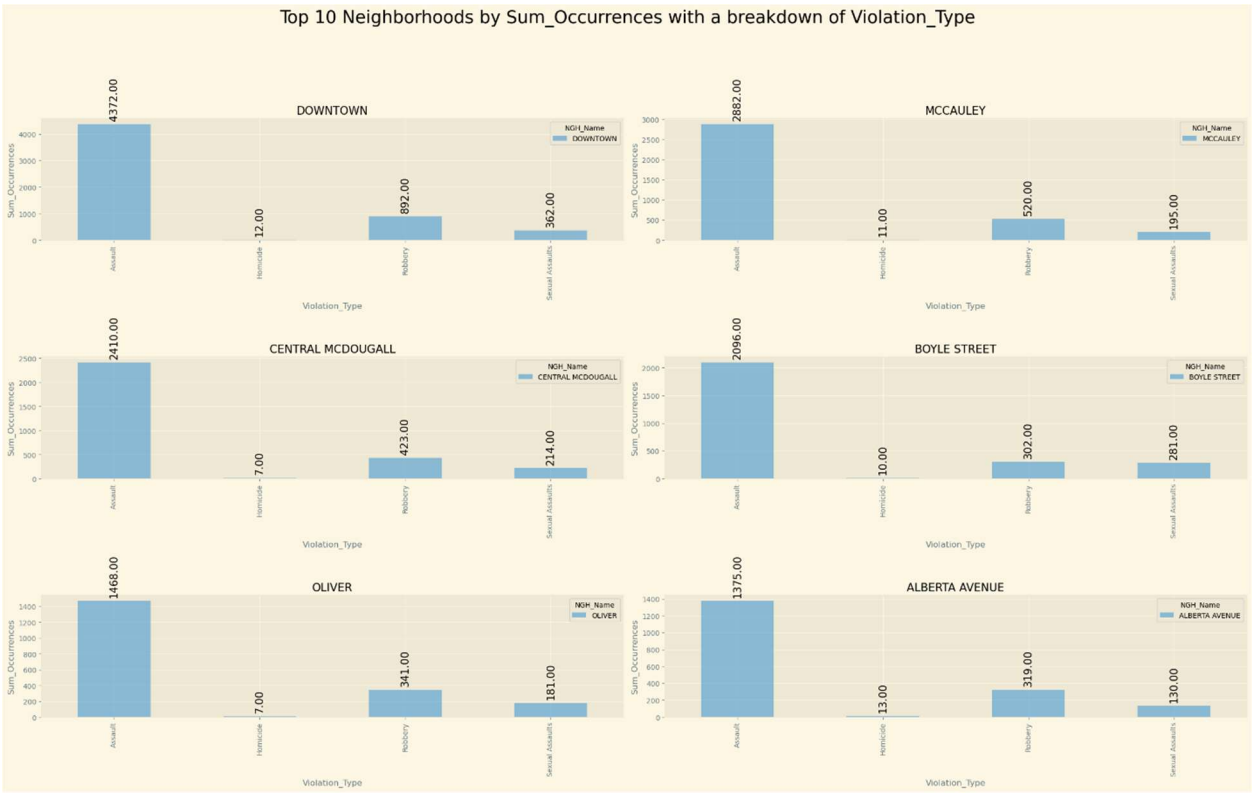

Figure 7: Breakdown of Violent Crimes for the Top 6 Neighborhoods.

To gain a more comprehensive understanding of severe crime trends in the six neighborhoods that consistently appeared in the top 10 between 2009 and 2019, we analyzed the breakdown of each Violent crime category. This included Assault, Sexual Assaults, Robbery, and Homicide, and our analysis aimed to identify the counts of each crime in these neighborhoods between 2009 and 2019. By conducting this analysis, we could better understand the types of severe crimes occurring in these neighborhoods and their relative frequencies, allowing us to identify patterns and trends and inform potential interventions and crime prevention strategies. As shown in Figure 7 above, Assault captured much of the proportion of Violent crimes for all six neighborhoods. In fact, if we recall from figure 3, The crime of Assault captured much of the overall severe crime proportion.

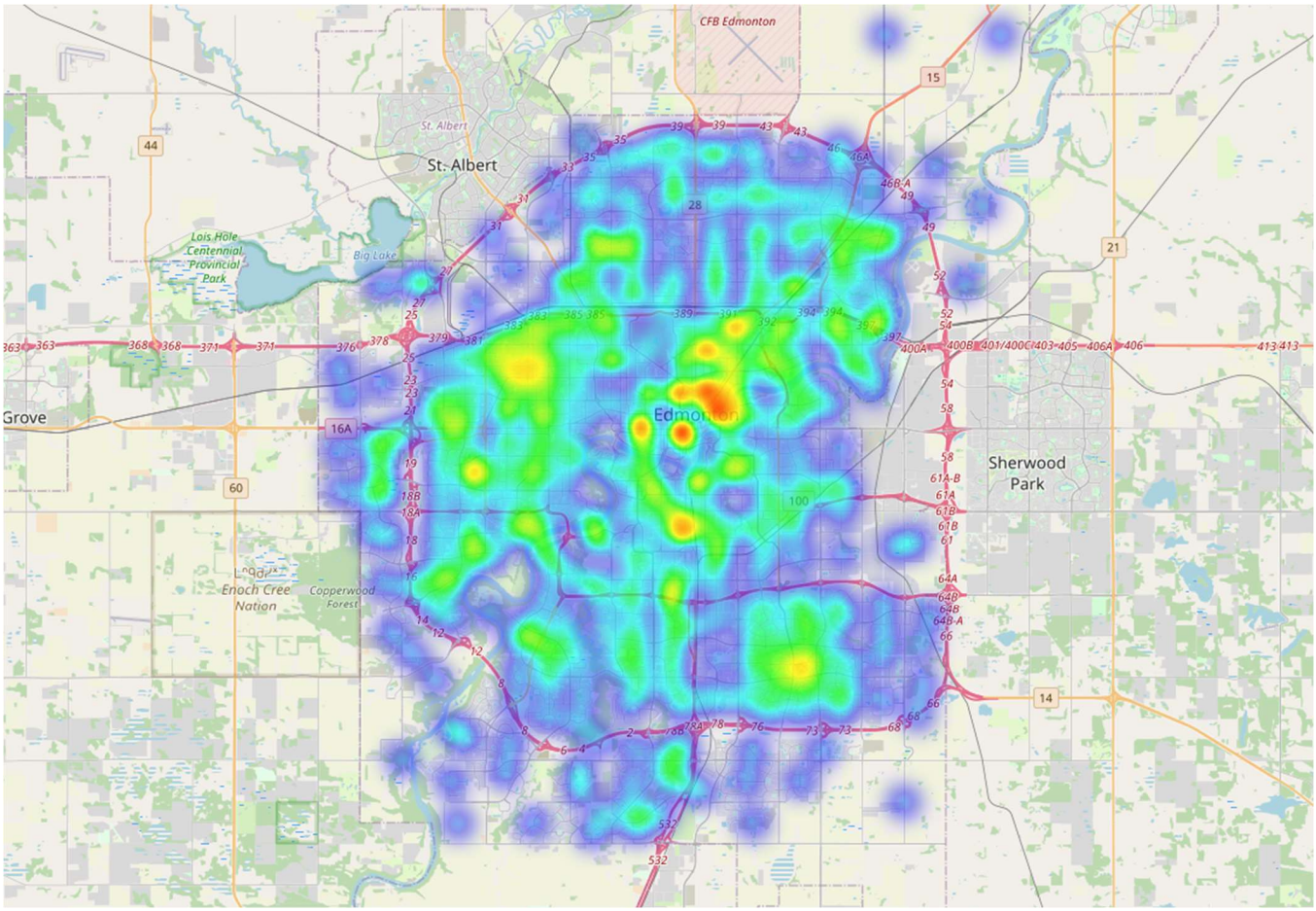

Figure 8:2009-2019, Folium Heatmap showing the distribution of Violent Crime Counts by Neighborhood in Edmonton.

To gain a visual understanding of severe crime trends in Edmonton neighborhoods and how they correlate with the top six neighborhoods that consistently appeared in the top 10 between 2009 and 2019, we used Folium Heat Map, as shown in figure 8 above, and Geopandas to map the overall occurrence of severe crimes in Edmonton, figure 9. By visualizing the distribution of severe crimes across neighborhoods, we were able to identify the dark areas on the heatmap that correlated with the top six neighborhoods. This analysis provided valuable insights into the areas in Edmonton that experienced high occurrences of severe crimes and may help guide targeted interventions and crime prevention strategies.

Before proceeding with our analysis of severe crime trends in Edmonton neighborhoods, we first used Pandas to pivot our filtered dataset. This involved creating a new dataframe, 'pivoted_df_filtered,' using the Pandas pivot_table function. The pivot_table function allowed us to aggregate the crime data by neighborhood, location, average temperature, year, and month, while also breaking it down by Violation Type. Specifically, we set the index to include 'NGH_Name,' 'NGH_Number,' 'Latitude,' 'Longitude,' 'AVG_Temp,' 'DT_Year,' and 'DT_Month,' and set the columns to 'Violation_Type.' The values were set to 'Sum_Occurences,' and we used 'aggfunc=sum' to aggregate the values by summing them. This process allowed us to analyze the crime data and identify patterns and trends in severe crime occurrences across neighborhoods in Edmonton more easily.

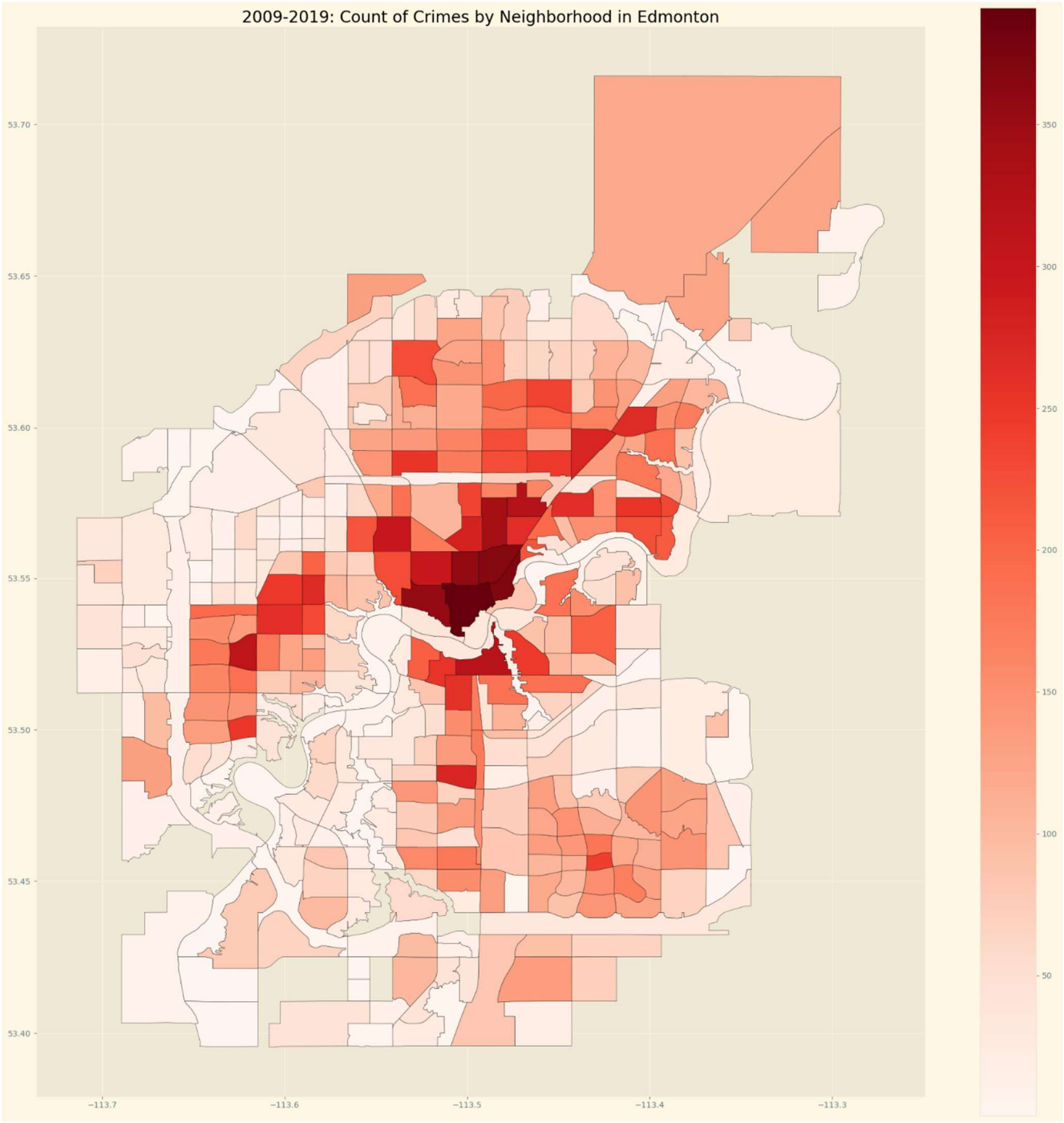

Figure 9: 2009-2019 Geopandas Map showing count of Violent Crimes in Edmonton.

2016

Although our initial dataset spans a decade from 2009 to 2019 and includes neighborhood crime data, we have chosen to concentrate on 2016 due to the availability of Income data by neighborhood for that year. By examining the demographics of each neighborhood through income data, we aim to gain better understanding of the relationship between socioeconomic status and crime rates. We also recognize that the year 2020 and beyond were impacted by the COVID-19 pandemic and may not be representative of typical crime rates, so we have chosen to limit our analysis to 2019. While we acknowledge that looking at a single year may limit our analysis, we believe that focusing on 2016 will provide valuable insights into the factors that contribute to crime within each neighborhood.

As we move forward with our analysis of severe crime trends in Edmonton, it is important to keep in mind the top six neighborhoods that consistently appeared in the top 10 between 2009 and 2019. These neighborhoods include Downtown, Alberta Avenue, Boyle Street, Central McDougall, McCauley, and Oliver, and represent areas that experienced high occurrences of severe crimes. By focusing on these neighborhoods in our analysis, we can better understand the factors that contribute to severe crime and inform targeted interventions and crime prevention strategies.

For our 2016 analysis, we used a curated dataset that included several columns, such as 'NGH_Name,' 'Latitude,' 'Longitude,' 'Violation_Type,' 'Sum_Occurrences,' and various demographic and income variables for each neighborhood. The demographic variables included Preschool, Kindergarten, Gr7_Gr9, Gr10_Gr12, Post_Secondary, Homemaker, Employedage0-30, Employed 30+, Unemployed, Retired, Permanently U, and Employment_No Response. The income variables included Income_Less than $30,000, Income_$30,000 to less than $60,000, Income_$60,000 to less than $100,000, Income_$100,000 to less than $125,000, Income_$125,000 to less than $150,000, Income_$150,000 to less than $200,000, Income_$200,000 to less than $250,000, Income_$250,000 or more, and Income_No Response.

During our analysis of the curated dataset for the year 2016, we observed a considerable number of null values in the income columns. Additionally, we also noticed many zero values in the demographic and income columns. Since census data is typically only collected in residential areas, we suspected that these zero values corresponded to industrial areas in Edmonton.

To ensure data quality and improve the accuracy of our analysis, we filtered these columns to only include values greater than zero (>0). This process automatically removed any rows that contained null values in the income columns, which were likely associated with industrial areas. By focusing on residential neighborhoods with meaningful demographic and income data, we were able to better understand the relationship between socioeconomic factors and severe crime in Edmonton.

Also, during our analysis, we encountered a substantial number of 'Employment_No Response' and 'Income_No Response' values in the curated dataset. To improve the accuracy of our analysis, we attempted to redistribute these no response values proportionately to their respective employment and income columns. However, after conducting this analysis, we determined that the redistribution of these values did not have a significant impact on the data.

As a result, we ultimately decided not to include the no response columns in any of our subsequent analyses. Despite this limitation, we still believe that our analysis provides valuable insights into the relationship between socioeconomic factors and severe crime in Edmonton.

Unlike the previous analysis, for this analysis we decided to keep the property crimes with the violent crimes to help us further our analysis of relationships.

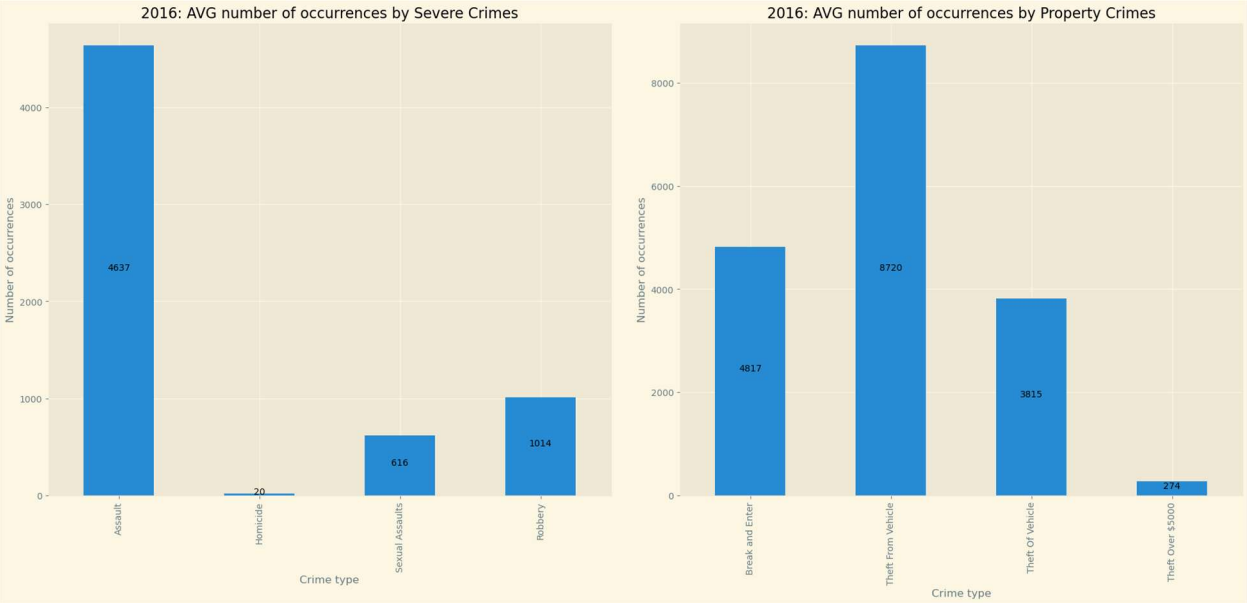

Figure 10: 2016 Count of Violent Crimes and Property Crimes for Edmonton.

To begin our analysis of crime in Edmonton in 2016, we plotted the counts of Violent Crimes and Property Crimes for the entire city. To accomplish this, we created a grid with two subplots using matplotlib. The first subplot displays the total number of severe crimes, including Assault, Homicide, Sexual Assaults, and Robbery. The second subplot displays the total number of property crimes, including Break and Enter, Theft from Vehicle, Theft of Vehicle, and Theft Over $5000.

As we can see for 2016, Violent Crimes of Assault were at 4637. On the other hand, Theft from Vehicle from Property crimes was at 8720 counts. Figure 10 visual representation provides an overview of the distribution of crime types in Edmonton in 2016, serving as a starting point for further analysis.

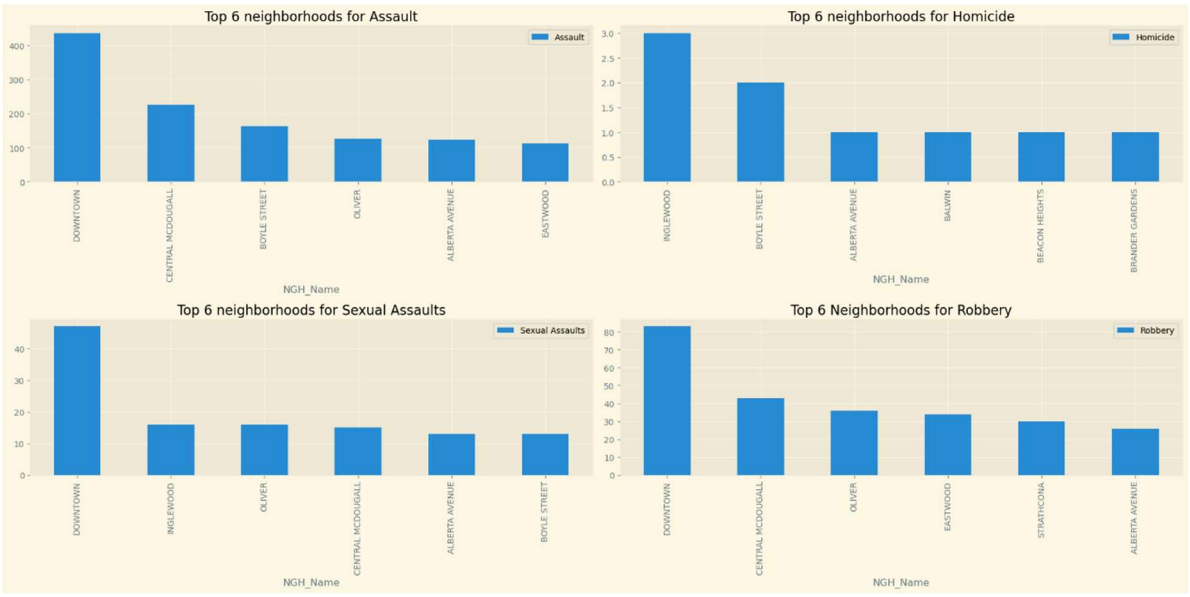

Figure 11: 2016 Top 6 Neighborhoods with highest Violent Crimes

To further our analysis, we plotted the top 6 neighborhoods for both Violent Crime (figure 11) and Property Crime (figure 12) in 2016. The results showed that the same neighborhoods that consistently showed up in our previous analysis for 2009-2019, also appeared as top neighborhoods for 2016 in terms of both Violent and Property Crimes. Downtown, Central McDougall, McCauley, Boyle Street, Alberta Avenue, and Oliver were the top 6 neighborhoods for both Violent and Property Crimes. This consistency across multiple years suggests that these neighborhoods may have underlying socio-economic factors that contribute to their higher crime rates.

Figure 12: 2016 Top 6 Neighborhoods with highest Property Crimes in Edmonton.

After plotting the top 6 neighborhoods for violent crimes and property crimes in 2016, we merged the analysis data with property assessment values for the same year. First, we filtered the property assessment values for 2016 and then grouped them by neighborhood while aggregating the values to a mean. We then merged the assessment data with our analysis data, resulting in two null values for the assessment values. The two corresponding neighborhoods were dropped as they represented new areas with very low crime rates, and therefore, were not relevant to our analysis.

The figure 13 below shows the distribution of the Assessment Values for 2016 over all neighborhoods.

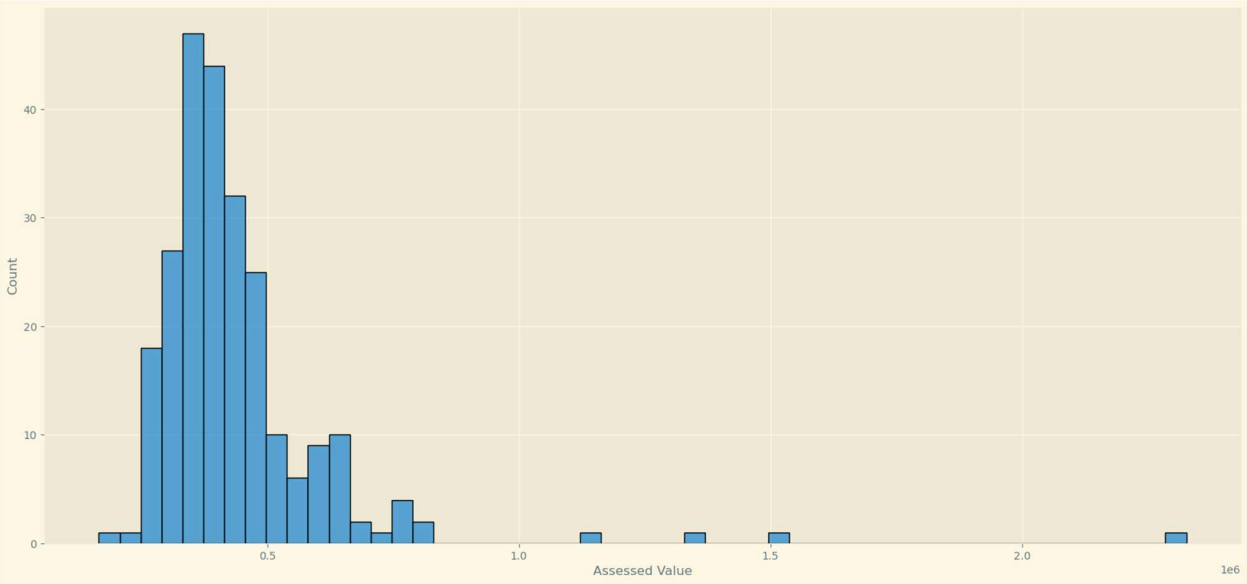

Figure 13: 2016 Distribution of Property Assessment Values in Edmonton.

We then proceeded with the new dataframe by plotting a correlation heatmap to further investigate the violent crimes.

The correlation heatmap, as shown in figure 14, is a useful tool to visually examine the relationships between variables in the dataset. To create the heatmap, we used the Seaborn library, which provides a heatmap function that generates a color-coded matrix of the correlation coefficients between pairs of variables.

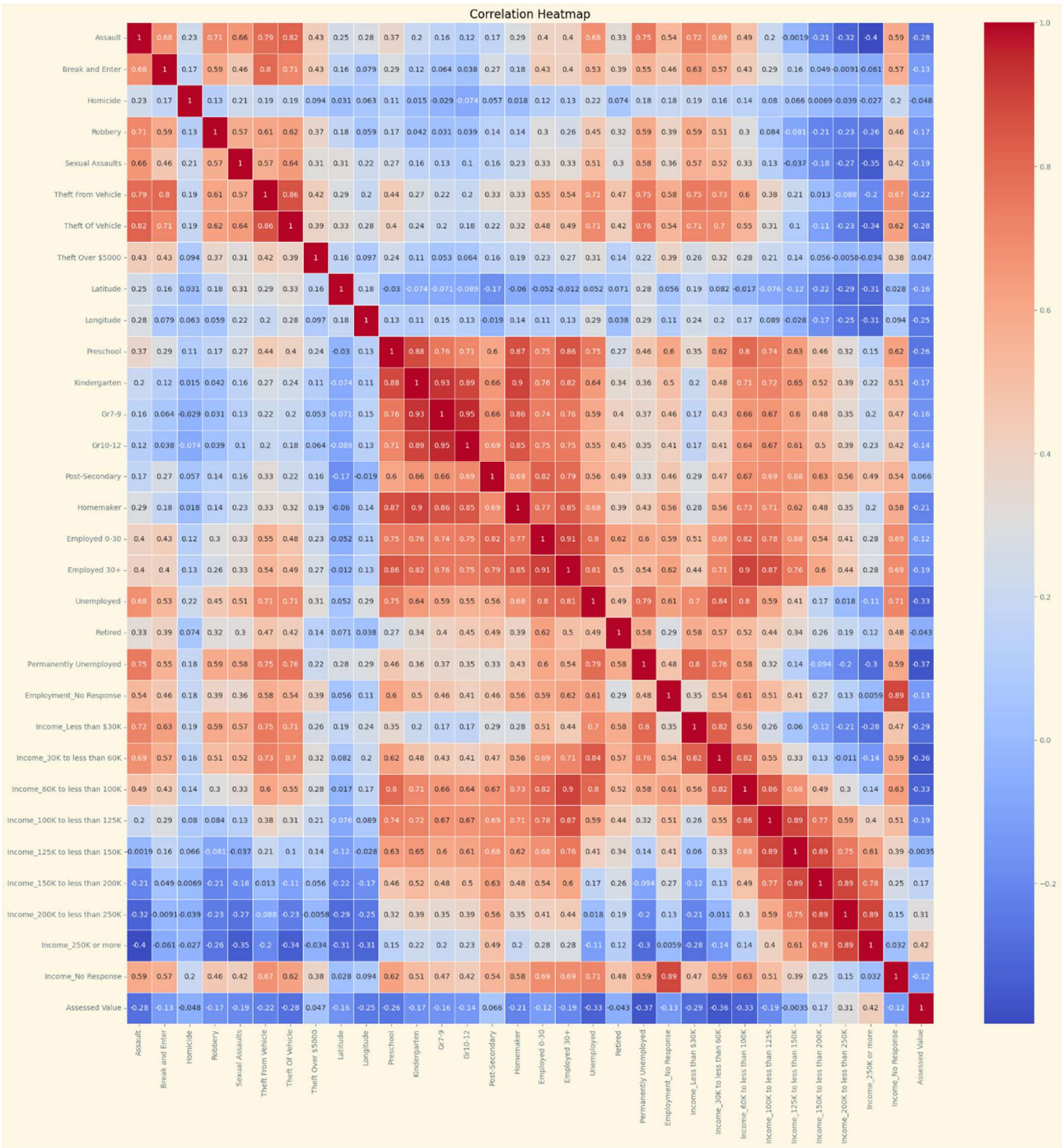

Figure 14: 2016 Correlation Heatmap

2016 Correlation Analysis:

It is important to note that, however, that correlation does not necessarily imply causation, and further analysis would be needed to fully understand the relationships between these variables.

Top 4 Correlations with 'Assessed Value':

Income_250K or more: positive correlation (0.42)

Permanently Unemployed: negative correlation (0.37)

Income_30K to less than 60K: negative correlation (0.36)

Unemployed: negative correlation (0.33)

These results suggest that there are some interesting relationships between assessed property values and other variables in the final DataFrame. Here are a few possible interpretations:

Higher-income neighborhoods tend to have higher assessed property values. This is indicated by the positive correlation between Assessed Value and Income_250K or more. It's possible that this correlation reflects the fact that more affluent neighborhoods tend to have more expensive homes and properties, which are valued higher for property tax purposes.

Unemployment and low income are negatively associated with assessed property values. This is indicated by the negative correlations between Assessed Value and Permanently Unemployed, Income_30K to less than 60K, and Unemployed. It's possible that these correlations reflect the fact that neighborhoods with higher rates of unemployment and lower median incomes tend to have less valuable homes and properties, which are valued lower for property tax purposes

Top 4 Correlations with 'Assault':

Theft Of Vehicle: positive correlation (0.82)

Theft From Vehicle: positive correlation (0.79)

Permanently Unemployed: positive correlation (0.75)

Income_Less than $30K: positive correlation (0.72)

Areas with higher rates of theft from and theft of vehicles tend to also have higher rates of assault. This is indicated by the positive correlations between Assault and Theft Of Vehicle (0.82) and Theft From Vehicle (0.79). It's possible that this correlation reflects the fact that areas with higher rates of property crime may also have higher rates of violent crime.

Areas with higher rates of permanent unemployment and lower median incomes tend to have higher rates of assault. This is indicated by the positive correlations between Assault and Permanently Unemployed (0.75) and Income_Less than $30K (0.72). It's possible that this correlation reflects the fact that areas with higher rates of poverty and unemployment may also have higher rates of violent crime.

Top 4 Correlations with ‘Theft from Vehicle:

Theft Of Vehicle: positive correlation (0.86)

Break and Enter: positive correlation (0.80)

Assault: positive correlation (0.79)

Permanently Unemployed: positive correlation (0.75)

Areas with higher rates of theft from vehicles tend to also have higher rates of theft of vehicles and break and enter. This is indicated by the positive correlations between Theft from Vehicle and Theft Of Vehicle (0.86) and Break and Enter (0.80). It's possible that this correlation reflects the fact that areas with higher rates of property crime may have more than one type of property crime.

Areas with higher rates of theft from vehicles also tend to have higher rates of assault. This is indicated by the positive correlation between Theft from Vehicle and Assault (0.79). It's possible that this correlation reflects the fact that areas with higher rates of property crime may also have higher rates of violent crime.

Areas with higher rates of permanent unemployment tend to have higher rates of theft from vehicles. This is indicated by the positive correlation between Theft from Vehicle and Permanently Unemployed (0.75). It's possible that this correlation reflects the fact that areas with higher rates of poverty and unemployment may have higher rates of property crime.

Top 4 Correlations with ‘Robbery’:

Assault: positive correlation (0.71)

Theft Of Vehicle: positive correlation (0.62)

Theft From Vehicle: positive correlation (0.61)

Income_Less than $30K: positive correlation (0.59)

For 'Robbery', we found that higher counts were positively correlated with counts of assault, theft of vehicle, theft from vehicle, and having an income less than 30K.

To aid in our analysis, we created a custom function named **plot_crime_correlations ()** that allows us to focus on specific subsets of our data. This function takes in a dataframe and two crime types: severe crime and property crime. The function then filters the dataframe to include only the specified crime types and generates a correlation heatmap. This allows us to easily identify correlations between different variables within a particular subset of the data. By using this function, we were able to gain insights into how different demographic factors and assessed property values relate to both severe and property crimes. As per figure 15 we can see that Assault is highly correlated to Theft of Vehicle , Theft from Vehicle and Robbery.

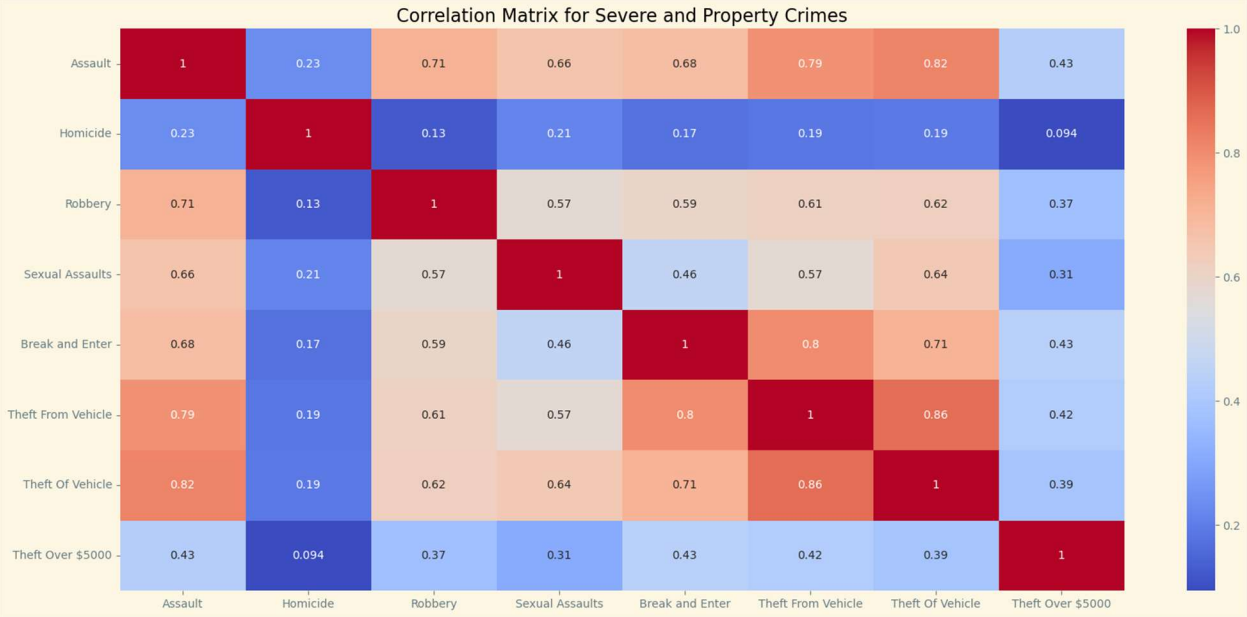

Figure 15: 2016 Correlation Heatmap between Violent Crimes and Property Crimes in Edmonton.

2016: Clustering

To further our analysis, we performed KMeans clustering on our dataset. We used the dendrogram, figure 16, to determine the optimal number of clusters and set it to 4. We then plotted a scatter plot of Latitude vs. Longitude, colored by cluster to visualize the distribution of our data, figure 17.

Overall, the KMeans clustering allowed us to identify distinct patterns in crime rates across the city, which could help inform targeted crime prevention strategies in different areas.

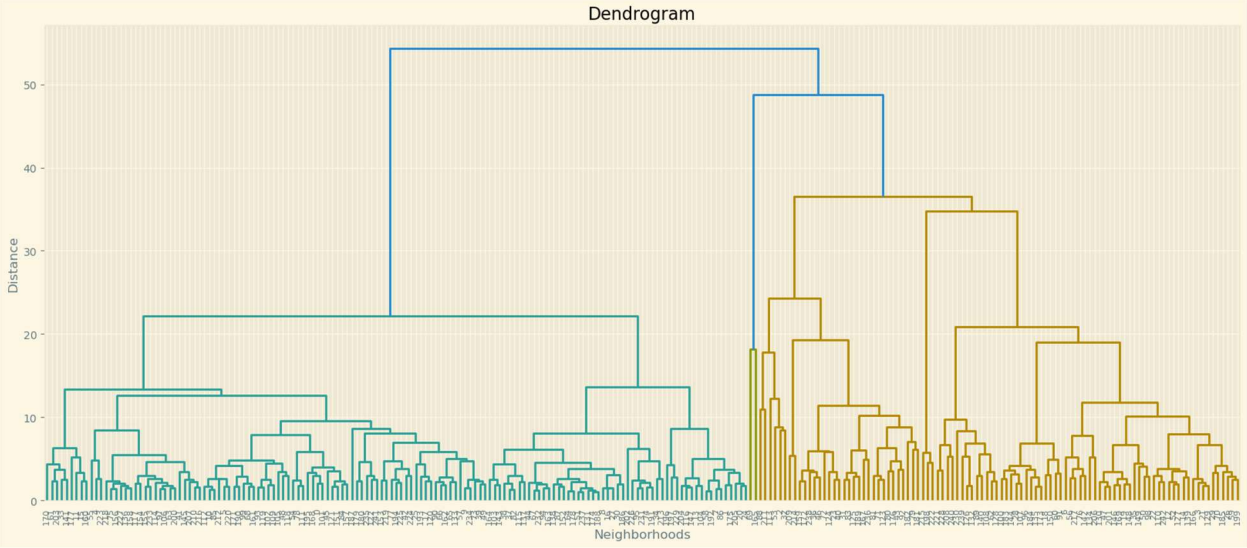

Figure 16: Dendrogram for Number of Clusters to use.

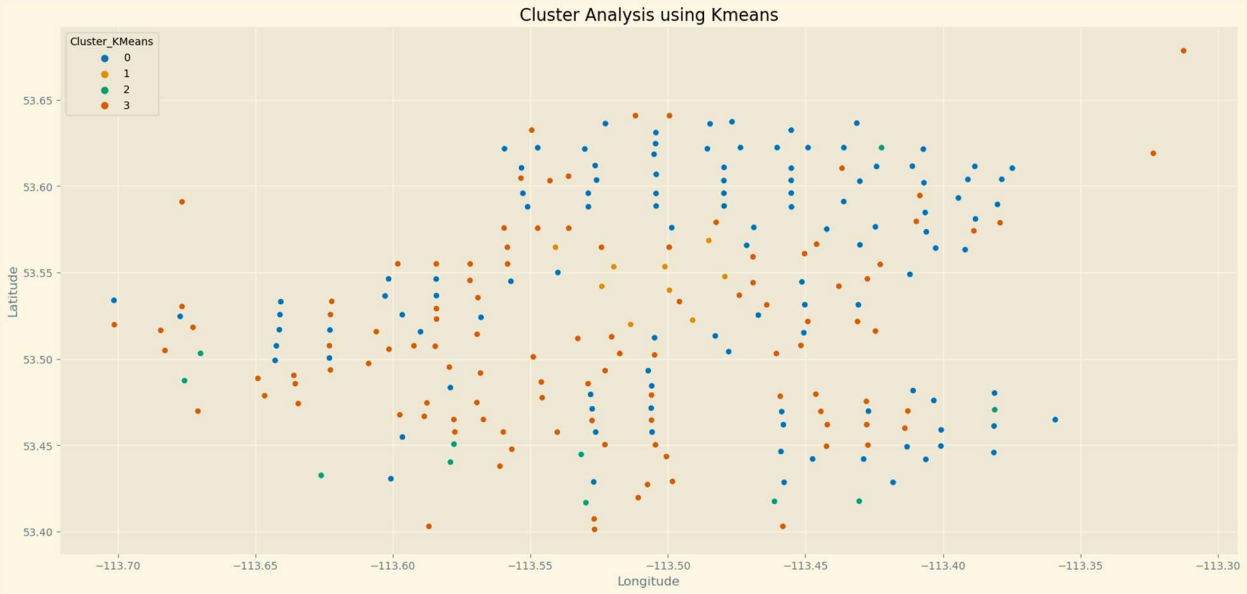

Figure 17: 2016 KMeans Cluster Analysis with Latitude vs Longitude

To visualize the KMeans clustering results, we used the Geopandas library to plot each cluster on a map of Edmonton. The map shows the spatial distribution of each cluster, with each block representing a neighborhood. From the map, we can see that the four clusters are well-defined and have distinct spatial patterns.

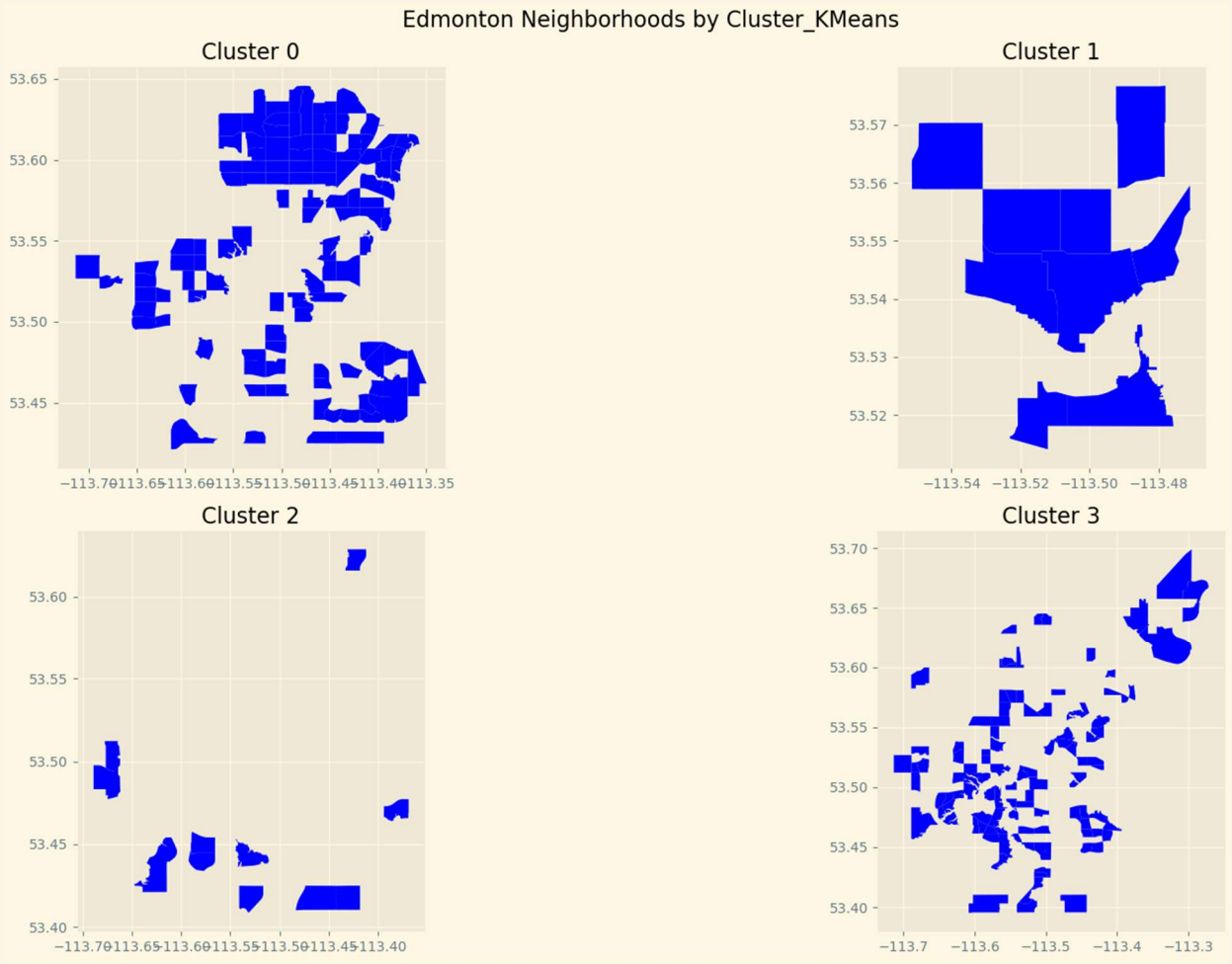

Figure 18: 2016 Geopandas visual of the clusters produced by KMeans.

Our analysis using KMeans clustering identified one cluster (Cluster 1) that contained neighborhoods with the highest counts of both property and violent crimes. Interestingly, this cluster contained the same neighborhoods that were identified as having high crime rates in our previous analysis (Alberta Avenue, Boyle Street, Central McDougall, Oliver, Downtown, and Queen Mary Park), but also included two additional neighborhoods (Garneau and Strathcona). The use of KMeans clustering provided additional validation of our previous findings and allowed us to identify additional neighborhoods with high crime rates.

Figure 19: 2016 Pie Chart showing distribution of Violent and Property Crimes by Cluster

To gain further insights into the characteristics of each cluster, we plotted pie charts, figure 19 above, to visualize the proportions of Violent Crimes and Property Crimes within each cluster.

As expected, Cluster 1 had the highest proportions of both Violent Crimes and Property Crimes, with Assault being the most common type of Violent Crime and Theft from Vehicle being the most common type of Property Crime, as shown in figure 20.

We also noticed that Neighborhoods in cluster 0 and cluster 2 had similarly reported crimes too. However, cluster 3 was the lowest in both Violent and Property crimes.

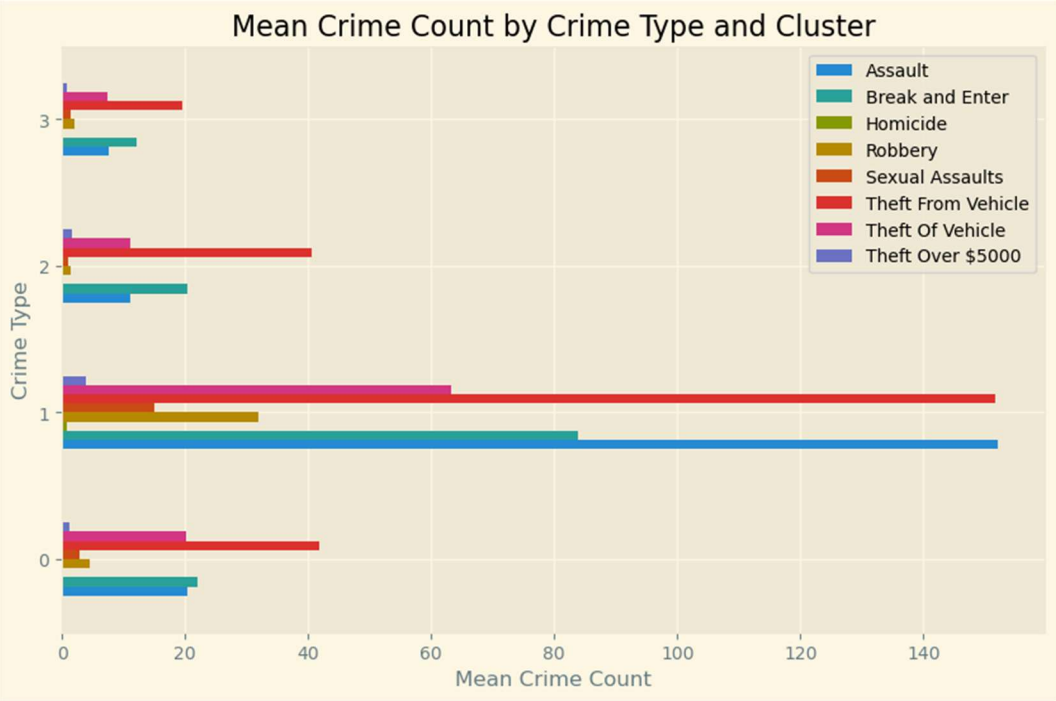

Figure 20: 2016 Horizontal Bar chart showing mean of Violent and Property crimes by Cluster.

The table1 below, shows the mean value of each type of crime (Assault, Break and Enter, Homicide, Robbery, Sexual Assaults, Theft From Vehicle, Theft Of Vehicle, Theft Over $5000) by KMeans cluster.

Cluster 0 has a relatively low mean value for Assault and Robbery, but a high mean value for Theft from Vehicle.

Cluster 1 has a high mean value for Assault, Theft From Vehicle, and Theft Of Vehicle.

Table 1: Mean Values of Violent and Property Crimes by Cluster

Cluster_KMeans	Assault	Break and Enter	Homicide	Robbery	Sexual Assaults	Theft From Vehicle	Theft Of Vehicle	Theft Over $5000
0	20.469027	22.044248	0.070796	4.415929	2.858407	41.840708	20.238938	1.203540
1	152.111111	83.888889	0.777778	31.888889	15.111111	151.555556	63.333333	3.888889
2	11.090909	20.272727	0.000000	1.454545	0.909091	40.545455	11.090909	1.636364
3	7.536364	12.154545	0.036364	1.927273	1.318182	19.554545	7.481818	0.763636

Cluster 2 has a relatively low mean value for most types of crimes, with the exception of Break and Enter.

Cluster 3 has the lowest mean value for all types of crimes, indicating that neighborhoods in this cluster have the lowest crime rates as also visually shown in figure 20.

```
cluster_1['NGH_Name']
```

```
2          ALBERTA AVENUE
32           BOYLE STREET
53      CENTRAL MCDOUGALL
69              DOWNTOWN
88               GARNEAU
113            INGLEWOOD
165              OLIVER
182       QUEEN MARY PARK
211           STRATHCONA
```

Figure 21: 2016 Neighborhoods in Cluster 1

```
['ABBOTTSFIELD', 'ALBANY', 'ALLARD', 'ALLENDALE', 'ARGYLL', 'ASPEN GARDENS',
 'AVONMORE', 'BARANOW', 'BEARSPAW', 'BELGRAVIA', 'BELLEVUE', 'BERGMAN', 'BLACKBURNE',
 'BLACKMUD CREEK', 'BLUE QUILL ESTATES', 'BRANDER GARDENS', 'BRECKENRIDGE GREENS', 'BROOKSIDE',
 'BULYEA HEIGHTS', 'CALLAGHAN', 'CALLINGWOOD NORTH', 'CAMERON HEIGHTS', 'CANON RIDGE', 'CAPILANO',
 'CARTER CREST', 'CHAMBERY', 'CHAPPELLE AREA', 'CLOVERDALE', 'CRESTWOOD', 'CROMDALE', 'DECHENE', 'DELTON',
 'DONSDALE', 'DOVERCOURT', 'EDGEMONT', 'EKOTA', 'ELMWOOD', 'ELSINORE', 'FALCONER HEIGHTS',
 'FULTON PLACE', 'GAINER INDUSTRIAL', 'GARIEPY', 'GRANDVIEW HEIGHTS', 'GRANVILLE', 'GREENVIEW',
 'GROVENOR', 'HENDERSON ESTATES', 'HIGH PARK', 'HODGSON', 'HUDSON', 'IDYLWYLDE', 'JAMIESON PLACE',
 'JASPER PARK', 'KAMEYOSEK', 'KEHEEWIN', 'KENILWORTH', 'LANSDOWNE', 'LAURIER HEIGHTS', 'LEE RIDGE',
 'LEGER', 'LENDRUM PLACE', 'MACTAGGART', 'MAGRATH HEIGHTS', 'MALMO PLAINS', 'MAYFIELD', 'MCKERNAN', 'MCLEOD',
 'MCQUEEN', 'MEYOKUMIN', 'MICHAELS PARK', 'MINCHAU', 'NORTH GLENORA', 'OGILVIE RIDGE', 'OLESKIW', 'PARKALLEN',
 'PATRICIA HEIGHTS', 'PEMBINA', 'POTTER GREENS', 'PRINCE CHARLES', 'PRINCE RUPERT', 'RHATIGAN RIDGE', 'RICHFORD',
 'RIDEAU PARK', 'RIO TERRACE', 'RIVERDALE', 'ROSENTHAL', 'ROSSDALE', 'RURAL NORTH EAST HORSE HILL',
 'RURAL NORTH EAST SOUTH STURGEON', 'SHERBROOKE', 'SHERWOOD', 'SIFTON PARK', 'SKYRATTLER',
 'SPRUCE AVENUE', 'STEINHAUER', 'STRATHEARN', 'SUMMERLEA', 'SWEET GRASS', 'TAWA', 'TERRA LOSA',
 'TERRACE HEIGHTS', 'THE ORCHARDS AT ELLERSLIE', 'TRUMPETER AREA', 'TWEDDLE PLACE', 'VIRGINIA PARK',
 'WEBBER GREENS', 'WEDGEWOOD HEIGHTS', 'WEINLOS', 'WESTRIDGE', 'WOODCROFT']
```

Figure 22: 2016 Neighborhoods in Cluster 3

Looking at the Pairplot of cluster 1, figure 23 below, the cluster with the highest crime rates, which include our initial Top 6 neighborhoods, we see that Assault and Robbery is highly correlated almost all the other crimes, which confirms our correlation analysis earlier over all the neighbourhoods. Here we are just looking at cluster 1, that includes the neighborhood as per figure 21, on previous page.

Figure 23: Pairplot of Cluster 1: with Highest Crimes Count.

To investigate further, we utilized the geopy package to calculate the distances between each neighborhood in cluster 1 and Downtown Edmonton, which we used as the center point. The resulting distances were added as a column to the cluster 1 dataframe, and a GeoDataFrame was created for mapping purposes.

We then created a map of the neighborhoods in cluster 1 using the center point as Downtown Edmonton, and the distances from the center were labeled for each neighborhood. This allowed us to visualize the spatial distribution of these high-crime neighborhoods, and how far they were from the center of the city.

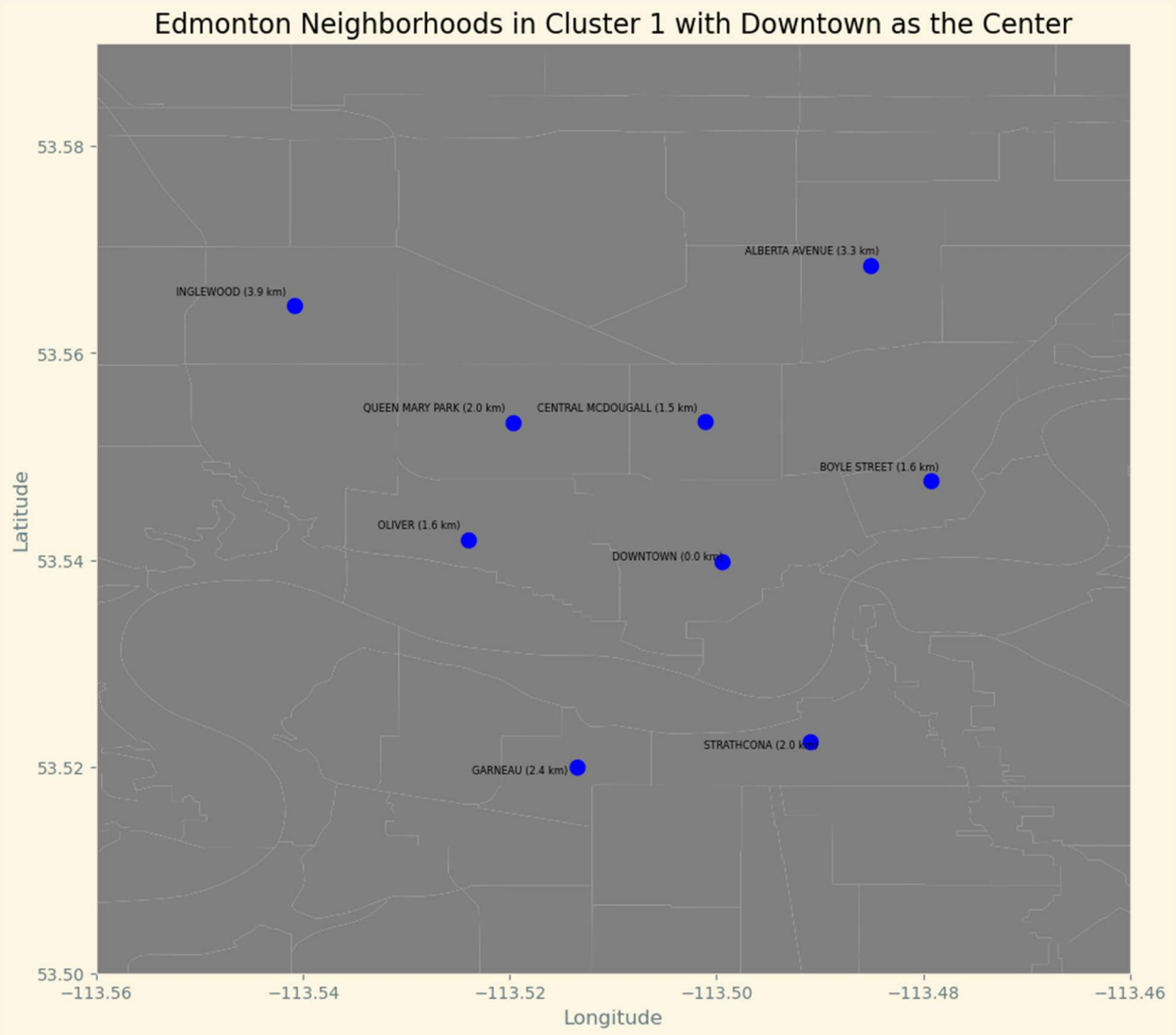

Figure 24: 2016 Distance Between Neighborhoods within Cluster 1.

On average, these neighborhoods were approximately 2.5 km away from Downtown Edmonton, as shown in Figure 24 above.

Figure 25: 2016 Proportion of Violent Crimes by Cluster

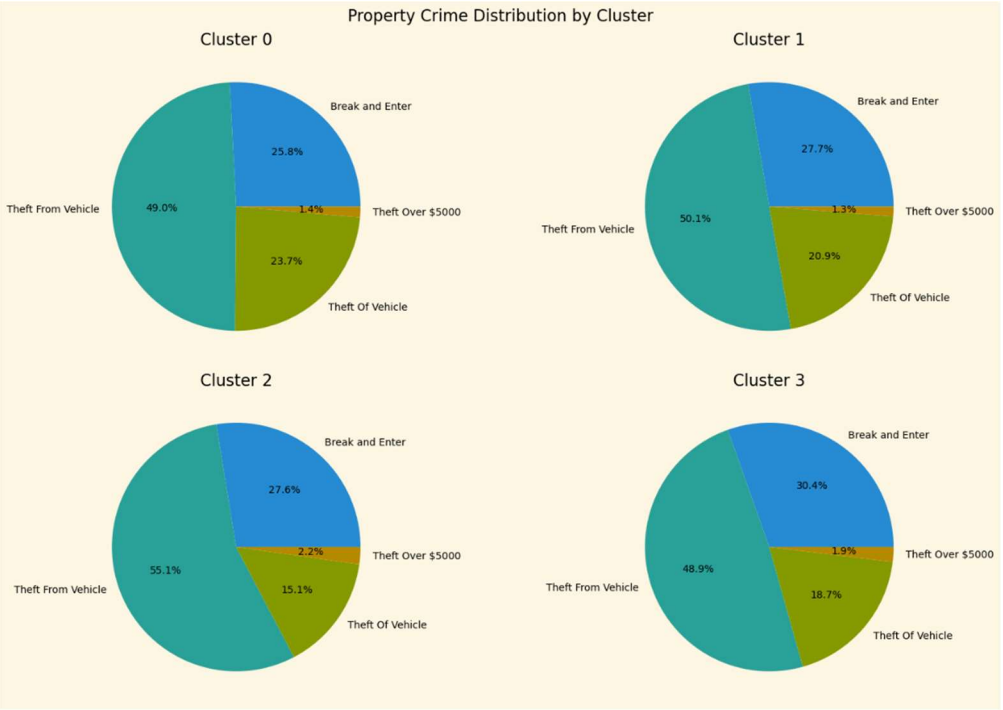

Figure 26: 2016 Proportion of Property Crimes by Cluster.

We created the pie chart in figures 25 and 26 to complement Table 1, allowing us to visually compare the proportions of severe crimes by cluster. The percentages for assault seem high across all clusters, but it is important to note that these calculations are based on the number of assaults within each cluster, rather than as an overall assault rate. While Cluster 0 had an average assault rate of only 20.47 and Cluster 1 had an average of 152.11, the proportion percentages appear similar. However, it is crucial to be cautious of statistics, as unclear charts like these can be manipulated to convey misinformation.

Comparison between Cluster 1 (High Crime) and Cluster 3 (low crime rates)

From the comparison of the income distribution between cluster 1 (the cluster with the highest crime rates) and cluster 3 (the cluster with the lowest crime rates), figure 27 and 28, we can observe that in cluster 3, the majority of the population belongs to the income categories of $30K to less than $60K, $60K to less than $100K, and $100K to less than $125K, whereas in cluster 1, the majority of the population belongs to the income category of less than $30K. Additionally, the income distribution in cluster 1 is more skewed towards the lower income categories compared to cluster 3. This suggests that income inequality might be a contributing factor to the higher crime rates in cluster 1.

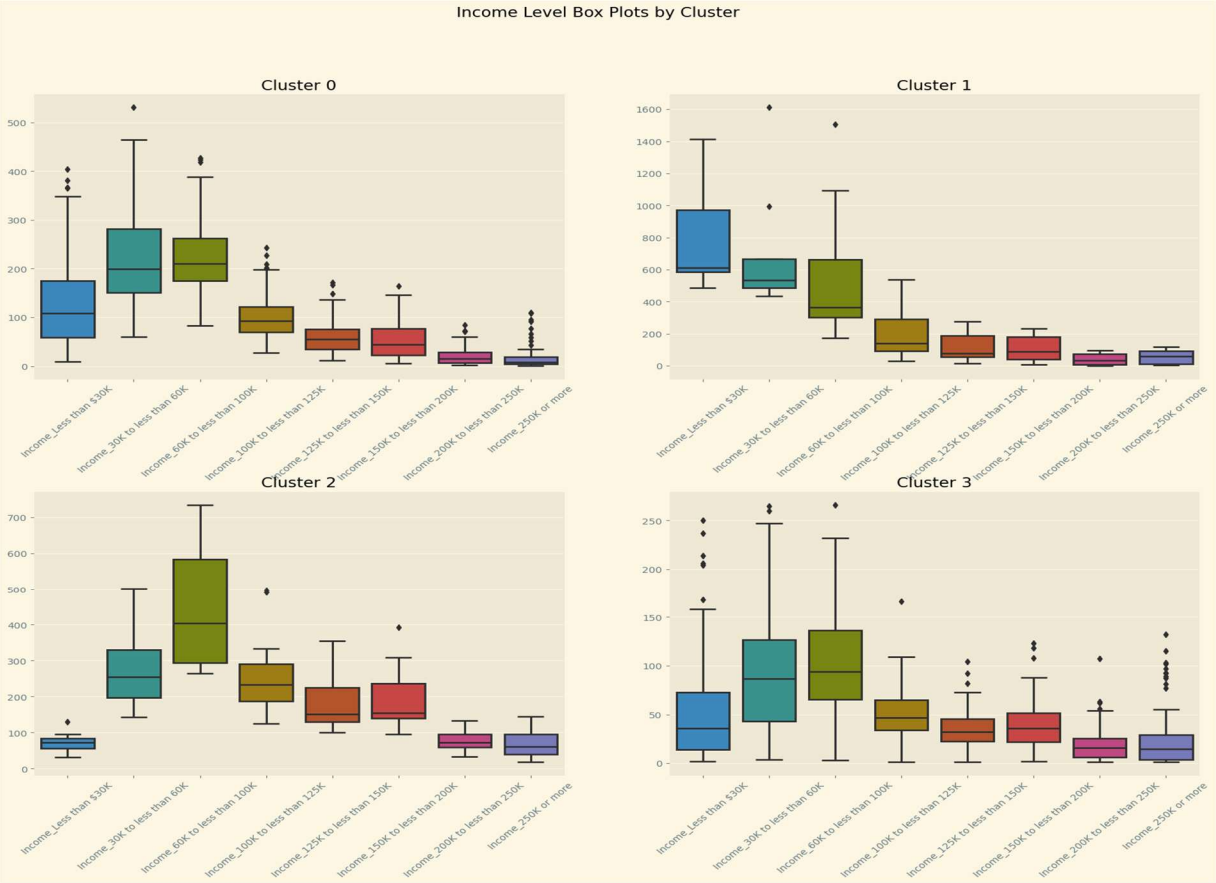

Figure 27: 2016 Boxplots to identify Central Tendency for Income Levels within Clusters.

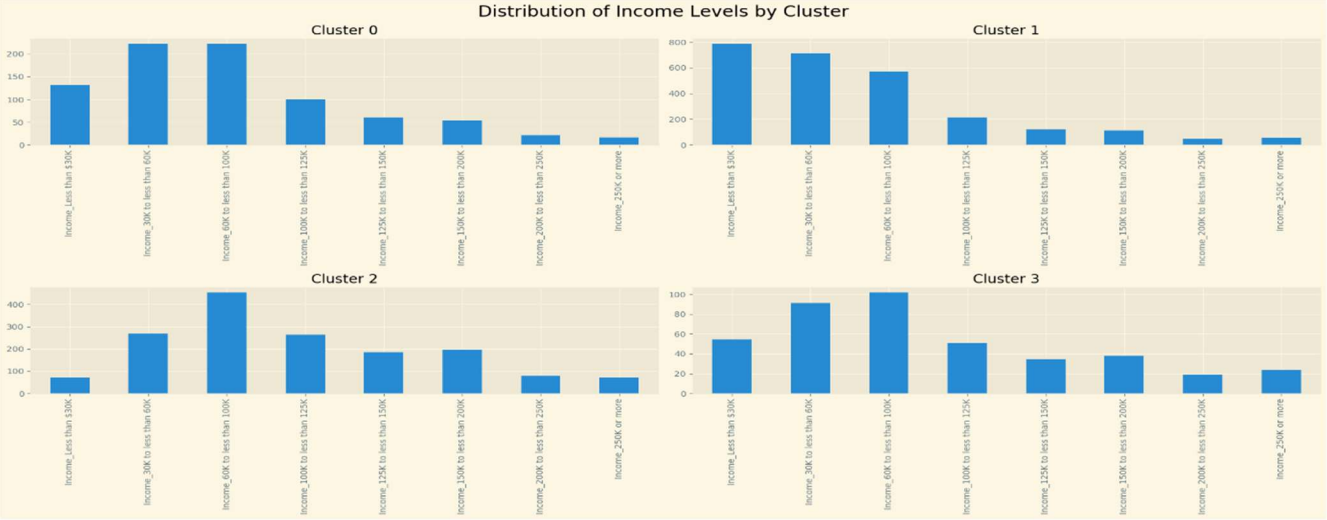

Figure 28: 2016 Bar plot of Income Levels by Cluster.

Comparing Cluster 1 and 3 side by side gives a better visualization.

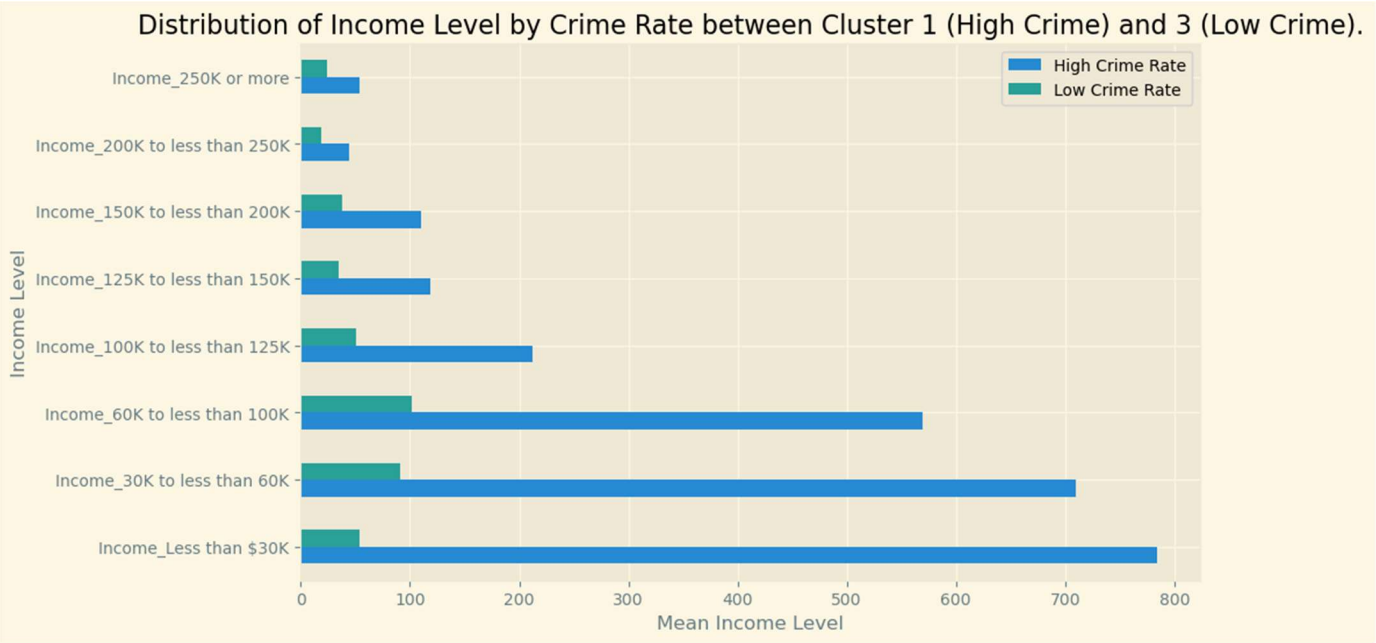

Figure 29: 2016 Distribution of Income Levels by Crime Rate between Cluster 1 and Cluster 3.

Comparing the clusters on educational levels from figure 30 and 31, from the data, it appears that Cluster 1 (the cluster with high crime rates) has a higher count of individuals with post-Secondary education levels compared to Cluster 3 (the cluster with low crime rates). However, Cluster 3 has higher counts of individuals in the Preschool and Kindergarten education levels compared to Cluster 1. Additionally, both clusters have similar counts of individuals in the Gr7-9 and Gr10-12 education levels.

However, looking at the neighborhoods in Cluster 3, we see that cluster 3 includes some very new Edmonton neighborhoods.

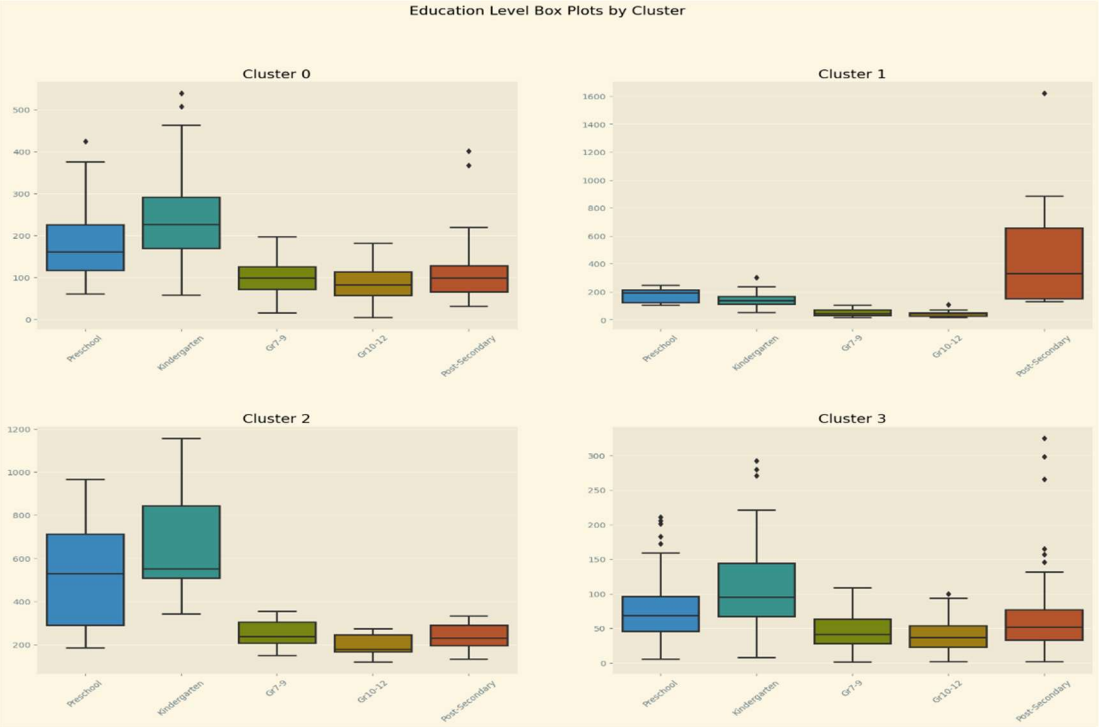

Figure 30: 2016 Boxplots for identifying Central Tendency for Education Levels within Clusters.

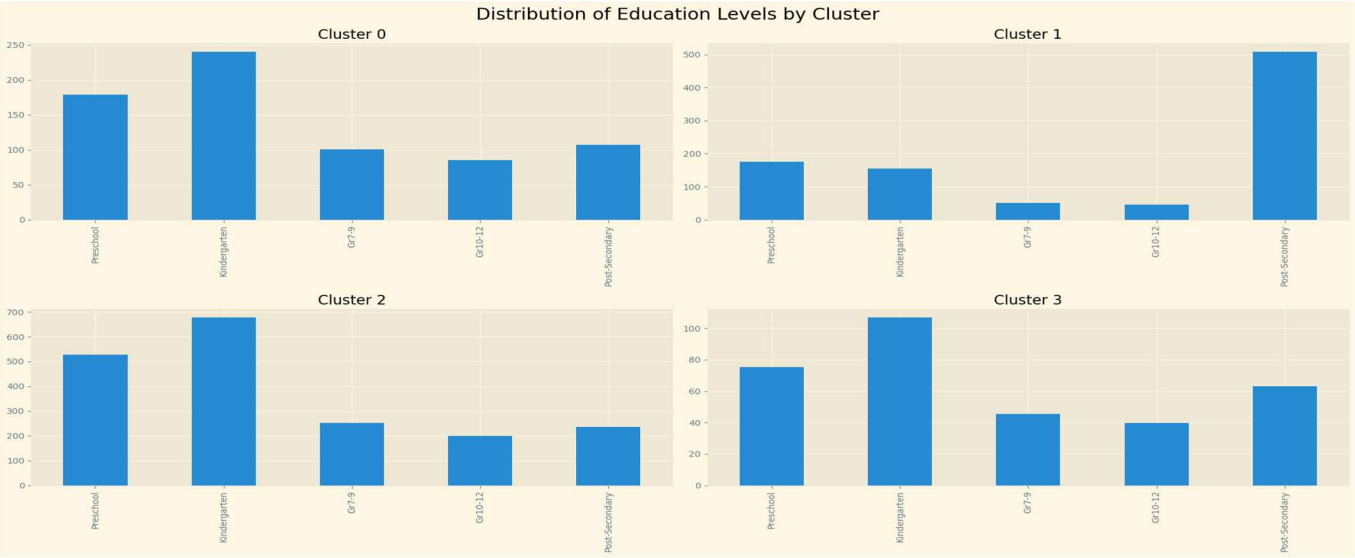

Figure 31: 2016 Education Levels by Cluster.

The below figure 32 shows a better visualization, showing our high crime area having the highest post-secondary education compared to cluster 3, that is the cluster with the lowest crime rate.

Cluster 1 has a very high count of Post Secondary educated residents. And since people with higher education are more understanding of reporting crimes promptly, hence leading to the high crimes rates being reported.

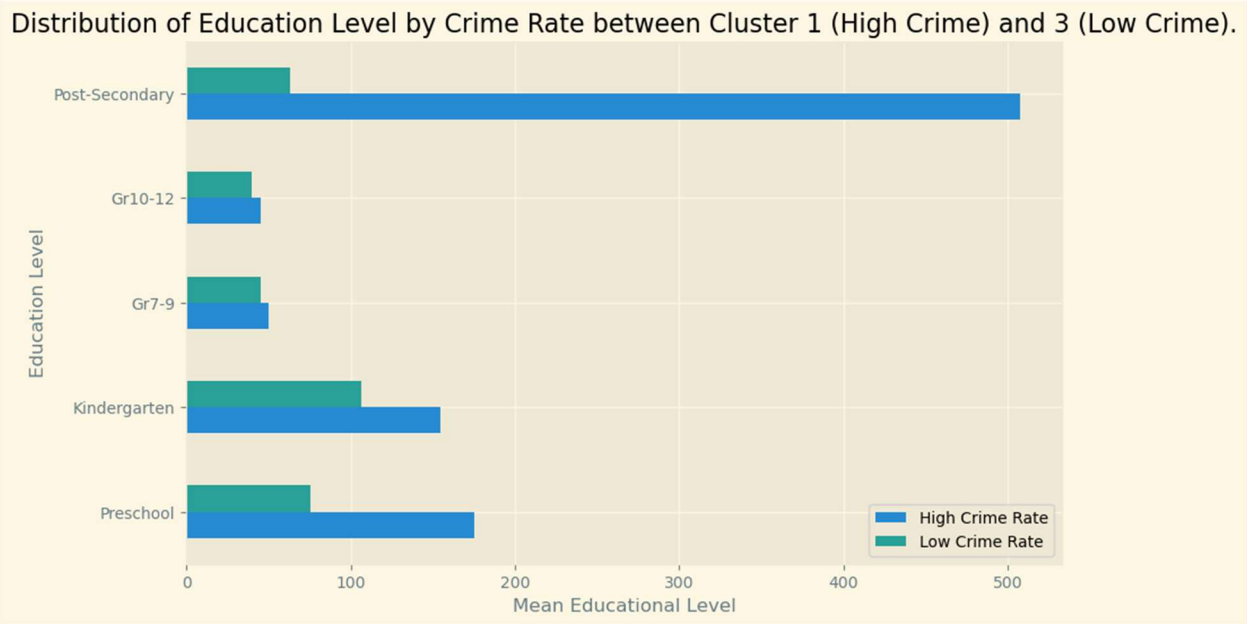

Figure 32: 2016 Comparing Clusters between Highest and Lowest Crime by Education Level.

Conclusion of Crime Analysis for the City of Edmonton.

Based on the analysis so far, we can conclude that there are certain neighborhoods in Edmonton that have higher crime rates than others. Specifically, neighborhoods in Cluster 1 have the highest rates of severe crimes such as Assault, Robbery, and theft. On the other hand, neighborhoods in Cluster 3 have the lowest crime rates.

There also seems to be a correlation between the income levels of neighborhood and its crime rate, with higher crime rates tending to occur in neighborhoods with lower incomes. However, there are some exceptions to this, such as Cluster 1 where there are high levels of post-secondary education despite the high crime rates.

Overall, the analysis highlights the importance of understanding the unique characteristics of each neighborhood when trying to address and reduce crime in Edmonton.

If given more time, some potential avenues for further analysis could include:

Examining the relationship between crime rates and other environmental factors such as Air Quality, proximity to public transit, access to public spaces and density of liquor stores or other businesses that may attract crime.

Exploring whether there are any patterns in the types of crimes that are committed in different neighborhoods, whether certain types of crimes are more likely to occur in clusters.

Conducting a deeper dive into the demographic factors associated with crime by examining more granular data such as age, gender and ethnicity as well as exploring the population characteristics of high crime neighborhoods and analyzing the exact time and location of crimes being committed would provide a more comprehensive understanding of factors driving crime in Edmonton.

Additionally, examining the characteristics of the perpetrators of crimes, such as their age, gender, and previous criminal history, could provide insights into how to prevent and reduce crime in the city of Edmonton.

Exploring the potential impact of economic factors, such as inflation, cost of living, and employment rates, on crime rates in Edmonton and analyzing any correlations between them could provide valuable insights into how economic conditions affect crime in the city. This could involve analysing and examining the economic characteristics of specific neighbourhoods or clusters with different levels of crime.

2016 Comparison Study Between Edmonton and Calgary.

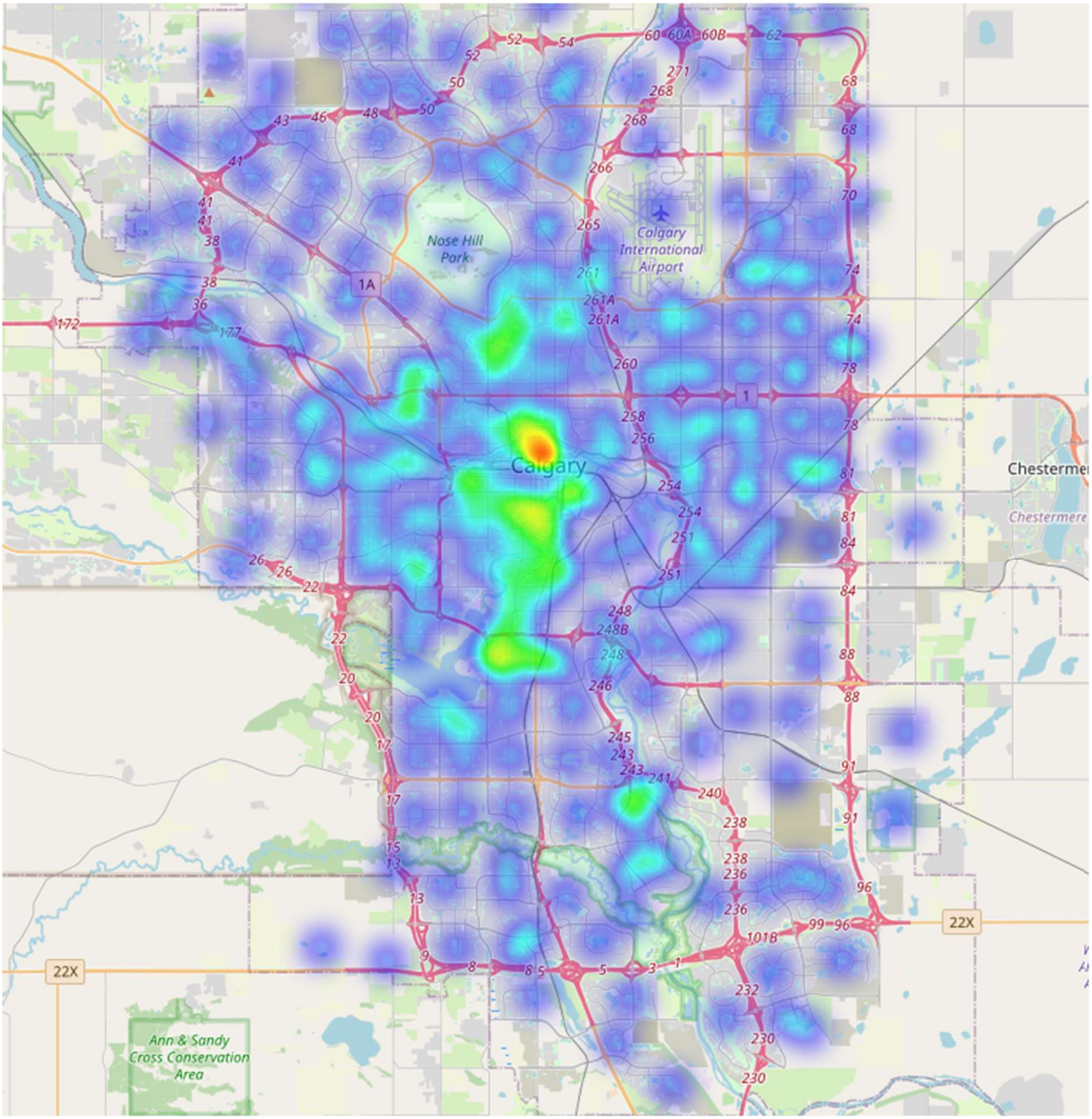

Figure 33: 2016 Folium Heatmap to show count of crimes by neighborhood in Calgary.

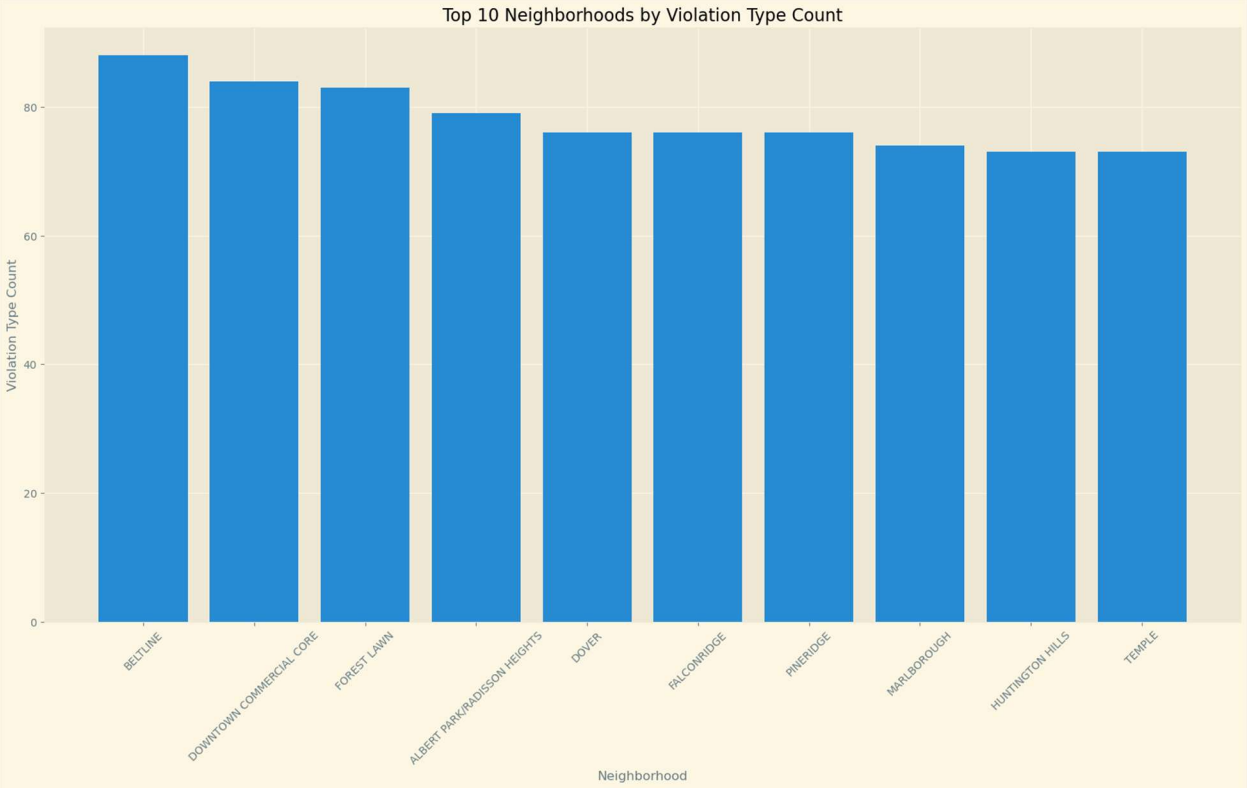

Figure 34: 2016 Top 10 Neighborhoods with the highest count of Crimes in Calgary.

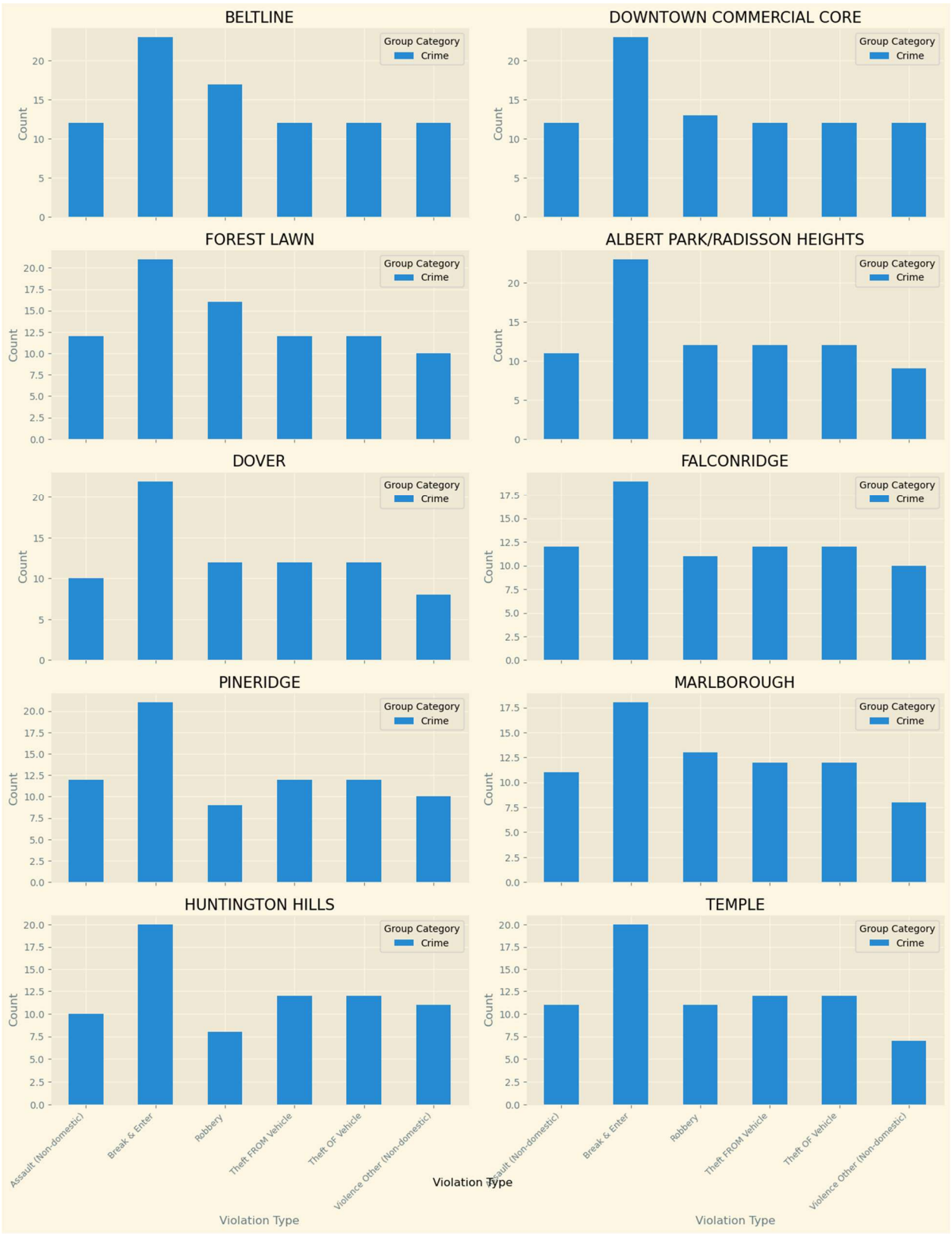

Figure 35: 2016 Top 10 Neighborhoods with breakdown of Crimes Occurred in Calgary.

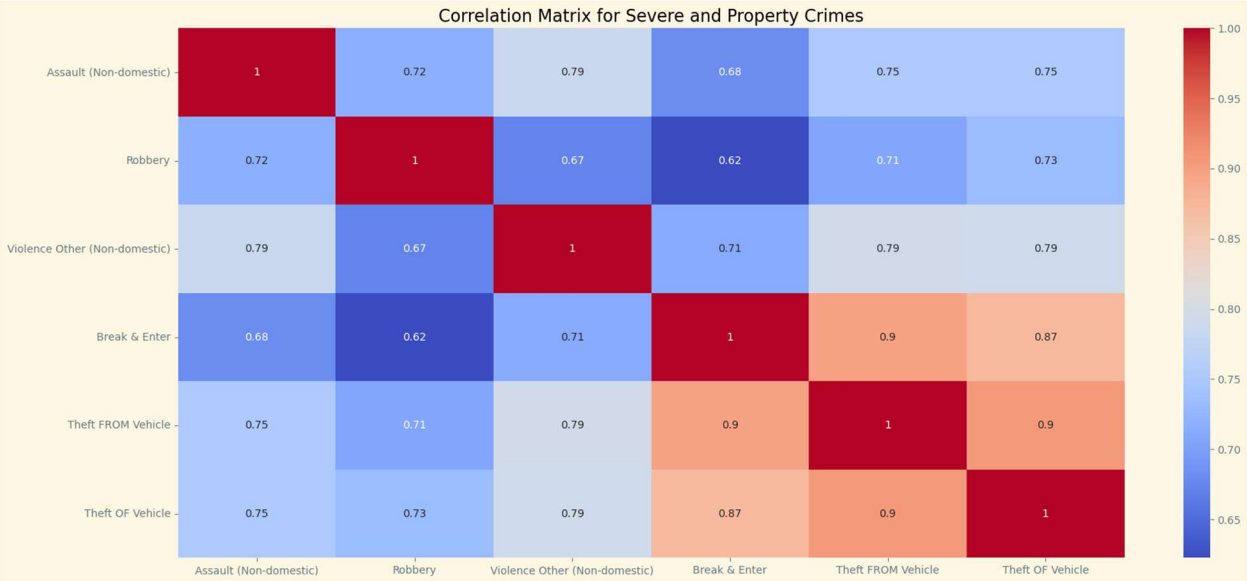

Figure 36: 2016 Correlation Heatmap between Violent and Property Crimes in Calgary.

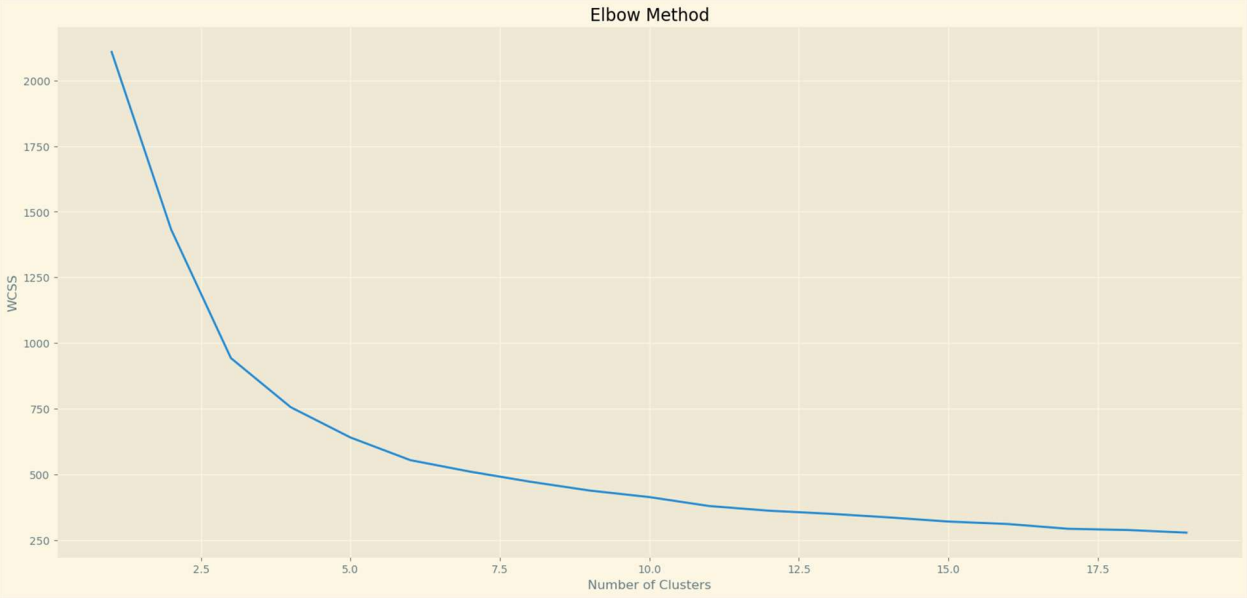

Figure 37: 2016 Elbow Method to Identify Number of Clusters

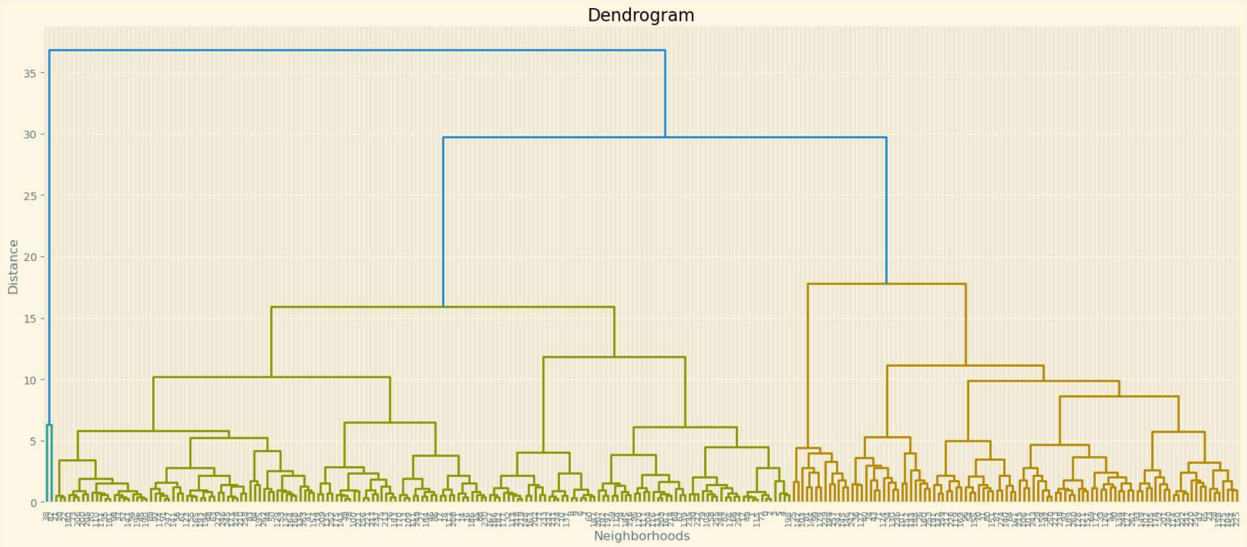

Figure 38: 2016 Dendrogram to identify number of Clusters.

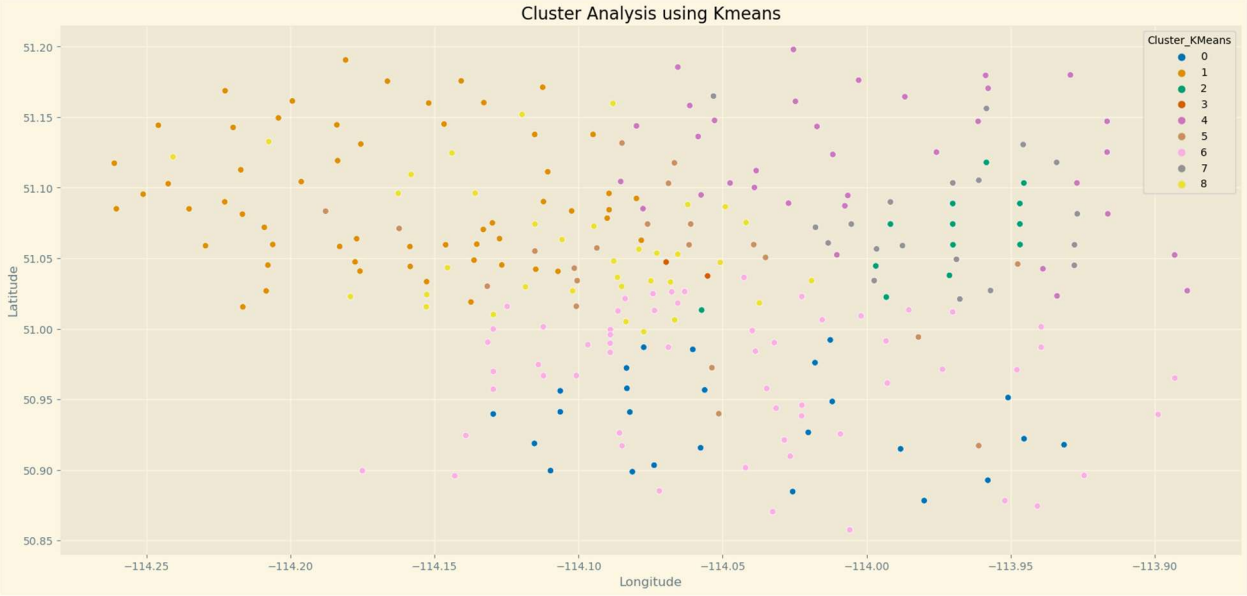

Figure 39: 2016 Result of KMeans Cluster Analysis on Calgary data.

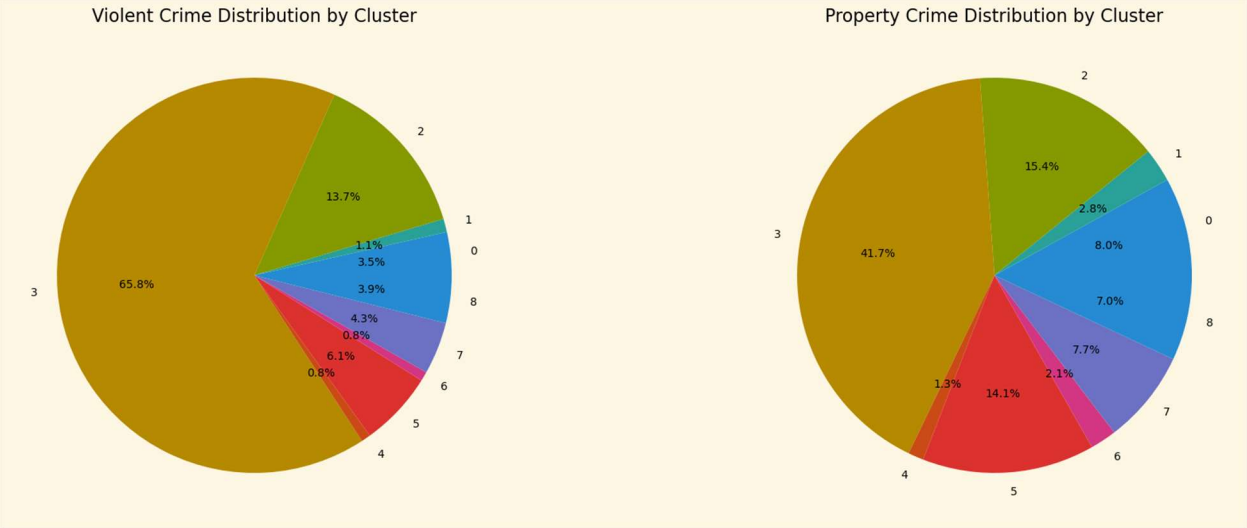

Figure 40: 2016 Proportion of Violent Crimes and Property Crimes by Cluster for Calgary.

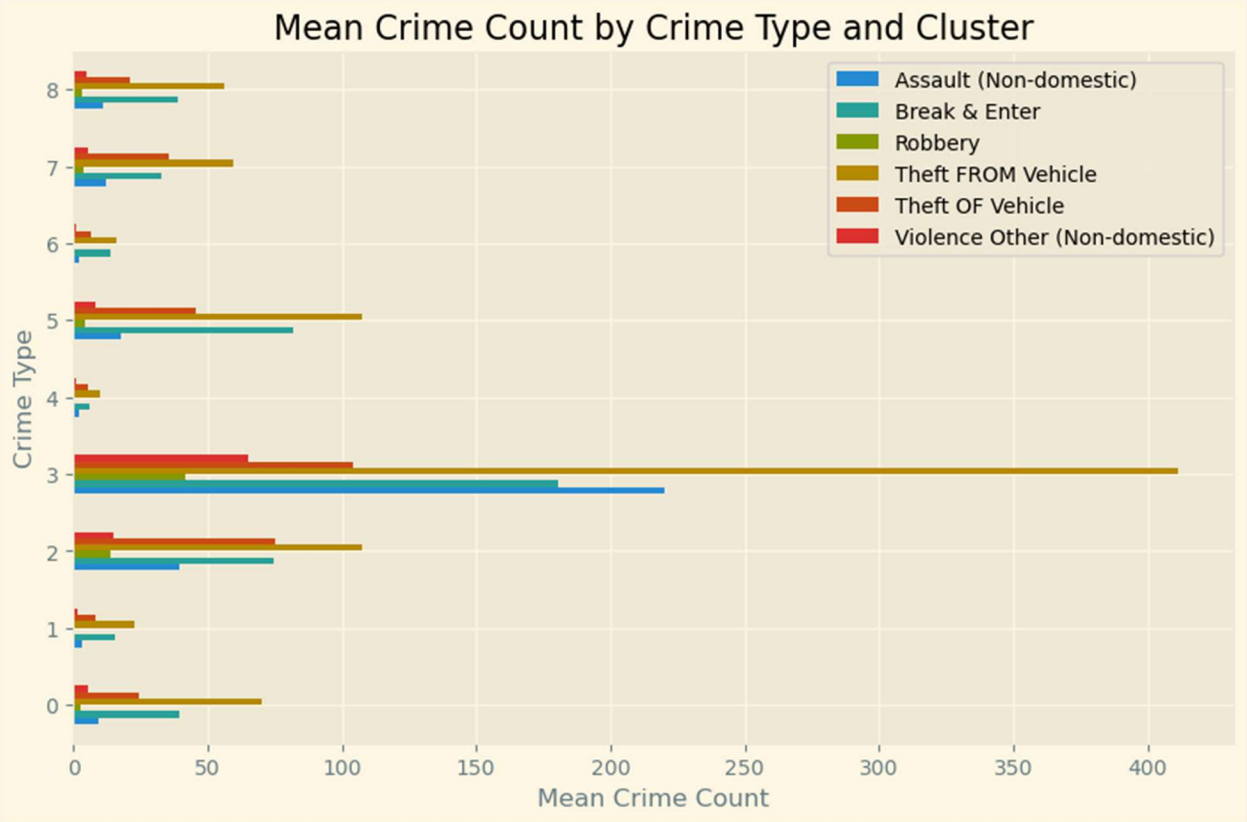

Figure 41: 2016 Distribution of Average Property and Violent Crimes by Clusters for Calgary.

Table 2: 2016 Table showing total number of crimes by Cluster in Calgary.

Cluster_KMeans	Assault (Non-domestic)	Break & Enter	Robbery	Theft FROM Vehicle	Theft OF Vehicle	Violence Other (Non-domestic)
0	9.320000	39.520000	2.640000	70.400000	24.240000	5.520000
1	3.403509	15.421053	0.438596	22.526316	8.175439	1.596491
2	39.692308	74.538462	13.615385	107.307692	75.153846	15.000000
3	220.000000	180.500000	42.000000	411.500000	104.000000	65.000000
4	2.264706	6.000000	0.500000	9.882353	5.382353	1.058824
5	17.952381	82.095238	4.428571	107.714286	45.904762	8.047619
6	2.327586	13.603448	0.293103	15.810345	6.465517	1.155172
7	12.105263	32.842105	3.736842	59.368421	35.684211	5.421053
8	11.285714	39.142857	3.314286	56.485714	20.857143	4.714286

```
Ccluster_2['NGH_Name']
```

```
25     ALBERT PARK/RADISSON HEIGHTS
81                           DOVER
99                      FALCONRIDGE
103                     FOREST LAWN
142           MANCHESTER INDUSTRIAL
144                     MARLBOROUGH
145                MARLBOROUGH PARK
146                     MARTINDALE
179                      PINERIDGE
199                         RUNDLE
239                       SUNRIDGE
241                         TEMPLE
257                       WHITEHORN
```

```
Ccluster_3['NGH_Name']
```

```
38                      BELTLINE
82     DOWNTOWN COMMERCIAL CORE
```

Figure 42: 2016 Code Extract showing Cluster 1& 2 for Calgary.

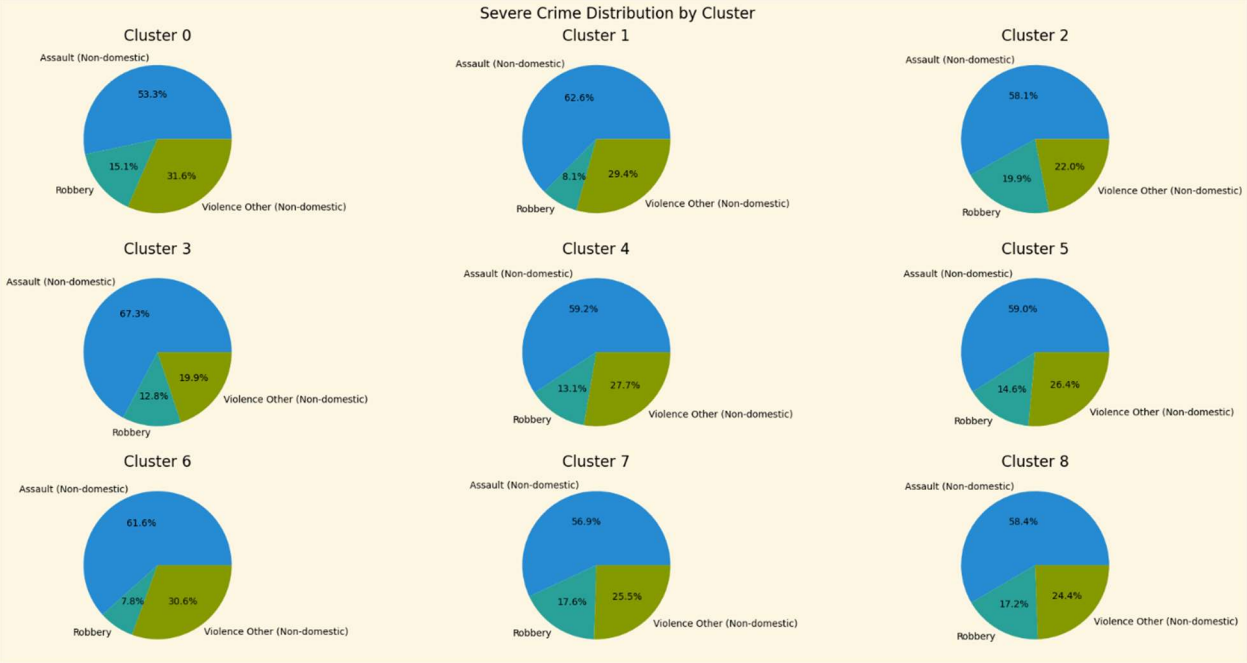

Figure 43: 2016 Proportions of Violent Crimes by Cluster in Calgary.

Figure 44: 2016 Proportions of Property Crimes by Clusters in Calgary.

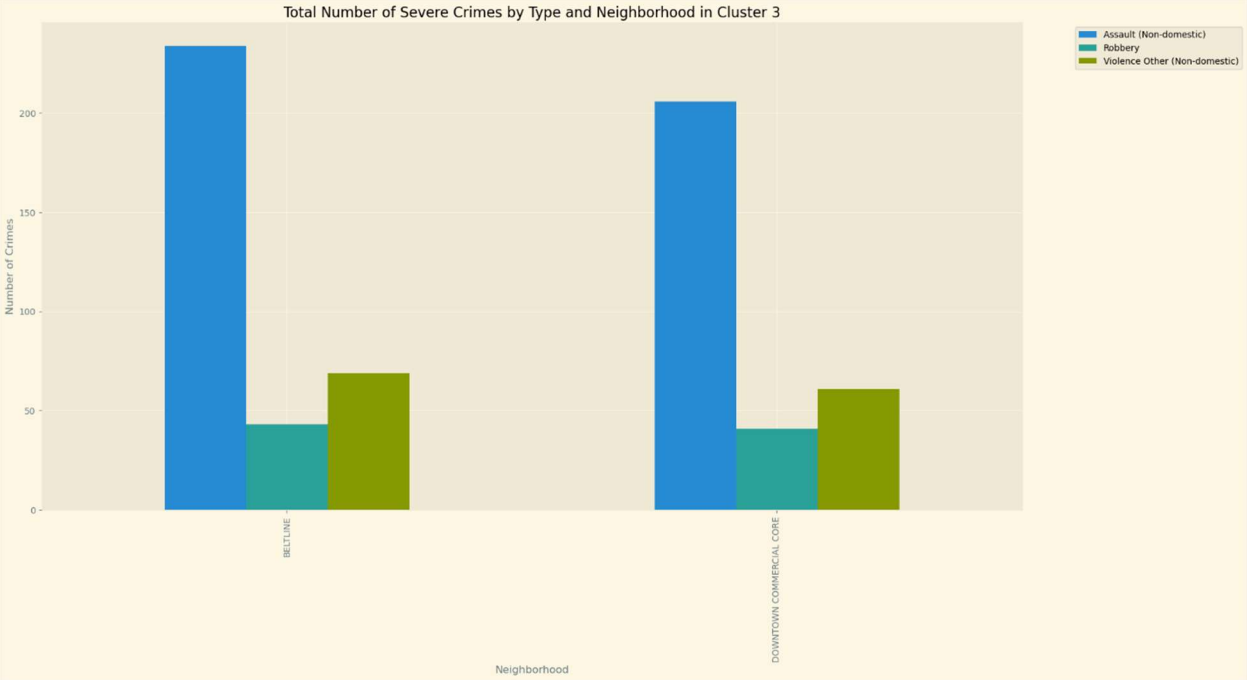

Figure 45: 2016 Cluster 3 showing breakdown of Violent Crimes.

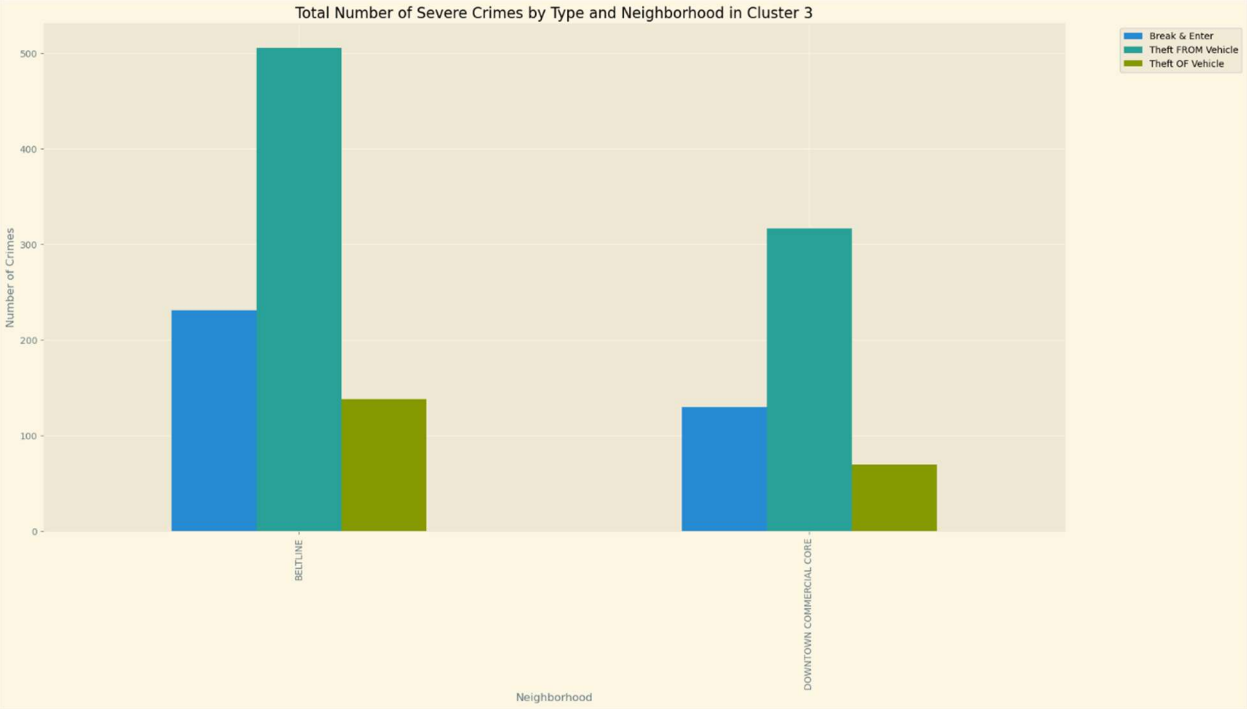

Figure 46: 2016 Cluster 3 showing breakdown of Property Crimes.

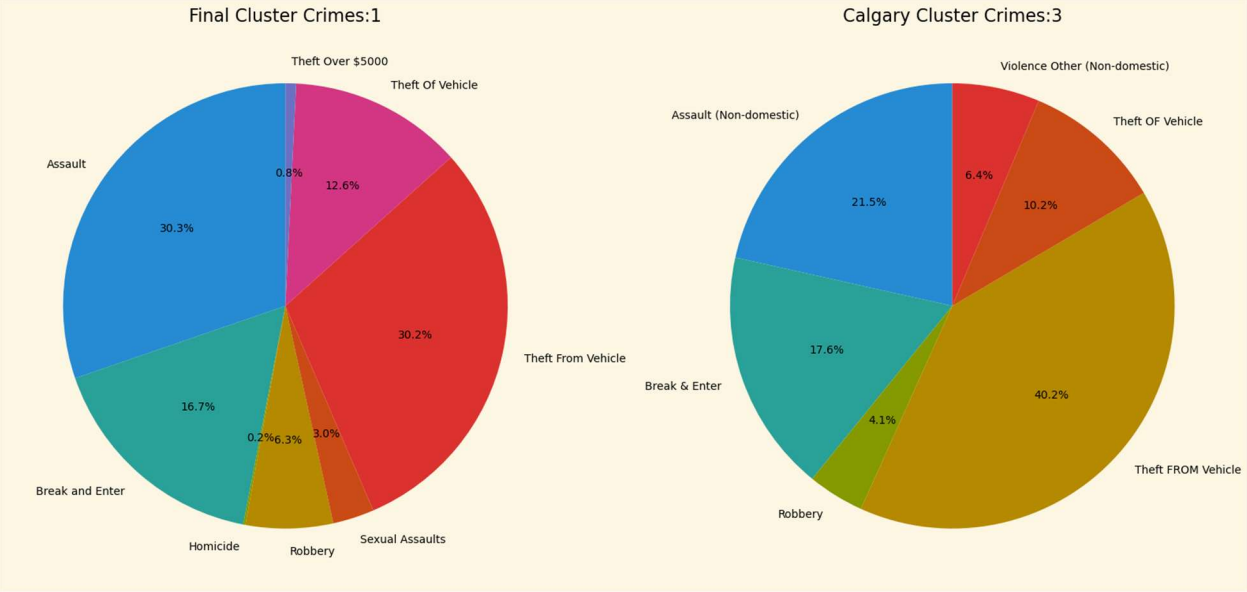

Figure 47: 2016 Comparison between Cluster 1 Edmonton and Cluster 3 Calgary

Table 3: 2016 Table showing the Neighborhoods within Edmonton Cluster 1

NGH_Name	Assault	Break and Enter	Homicide	Robbery	Sexual Assaults	Theft From Vehicle	Theft Of Vehicle	Theft Over $5000
ALBERTA AVENUE	123.0	119.0	1.0	26.0	13.0	156.0	99.0	2.0
BOYLE STREET	163.0	50.0	2.0	16.0	13.0	128.0	44.0	4.0
CENTRAL MCDOUGALL	226.0	50.0	0.0	43.0	15.0	122.0	70.0	5.0
DOWNTOWN	436.0	123.0	1.0	83.0	47.0	316.0	87.0	10.0
GARNEAU	33.0	78.0	0.0	17.0	2.0	83.0	33.0	1.0
INGLEWOOD	80.0	76.0	3.0	23.0	16.0	101.0	47.0	3.0
OLIVER	127.0	118.0	0.0	36.0	16.0	196.0	75.0	5.0
QUEEN MARY PARK	84.0	46.0	0.0	13.0	8.0	138.0	61.0	3.0
STRATHCONA	97.0	95.0	0.0	30.0	6.0	124.0	54.0	2.0

Table 4: 2016 Table showing the Neighborhoods within Calgary Cluster 3

NGH_Name	Assault (Non-domestic)	Break & Enter	Robbery	Theft FROM Vehicle	Theft OF Vehicle	Violence Other (Non-domestic)
BELTLINE	234.0	231.0	43.0	506.0	138.0	69.0
DOWNTOWN COMMERCIAL CORE	206.0	130.0	41.0	317.0	70.0	61.0

Summary & Conclusion

In 2016, crimes were reported in all neighborhoods in the City of Edmonton. However, certain neighborhoods such as Downtown, Central McDougall, Boyle Street, Oliver, and Alberta Avenue consistently had the highest number of Violent and Property Crimes. Although we do not have precise information on the exact locations within these neighborhoods where the crimes occurred, a comparison with the City of Calgary showed a similar pattern, with the concentration of crimes occurring within the city center, specifically in Downtown Calgary Commercial Core and the Calgary Beltline. Our KMeans cluster analysis confirmed this for both cities.

However, we must note that 2016 was not the most representative year for our analysis, as there was a sharp decrease in violent crimes for both cities between 2015 and 2016. Additionally, in May 2016, a wildfire began southwest of Fort McMurray, Alberta, which forced upwards of 88,000 people from their homes, making it the largest wildfire evacuation in Alberta's history. On a more positive note, the Rogers Place, a multi-use indoor arena in Edmonton and the new home of the Edmonton Oilers hockey team, was officially opened on September 8, 2016.

Despite the decrease in crime rates, Edmonton still had the highest Crime Severity Index. However, due to time constraints, we were unable to perform a more in-depth analysis of the data. Given additional time, some of the areas we would have liked to explore include the ratio of crimes committed by neighborhood to the total population of that neighborhood, the precise locations and times of crimes, and the role of the municipal government in combating crime. We are particularly interested in examining the impact of diverse types of municipal government, given that Edmonton has had a more liberal municipal government for decades, while Calgary has historically had a more conservative municipal government up until Mayor Nenshi, who was a centrist. This would allow us to answer the question of whether conservative municipal governments are more effective than liberal ones in fighting crime.

Hence, in addition, given the patterns observed in Edmonton and Calgary, we would have liked to extend our analysis to other Canadian cities to determine if similar trends exist. Specifically, we would have explored the concentration of crimes in the downtown cores of other major Canadian cities, such as Toronto, Vancouver, and Montreal, and compared their crime rates to those of Edmonton and Calgary. However, due to time constraints, we were unable to pursue this line of inquiry.