Project 1
Toronto Major Crime Indicators

Group 2:
    Tania Barrera
    Merve Celme
    Fanny Sigouin

October 10th, 2023

Project Goal & Dataset

Goal

Analyze major crime indicators (MCIs) in the city of Toronto from 2014 to 2022 to identify changing crime trends year over year as well as trends by type of offences and by neighbourhoods.

Dataset

Data was retrieved from Open Data Toronto using their API.

Data Limitations ⇨ Data Cleaning

1. **One occurrence may have several rows of data** associated with MCIs as data is provided at the offence and/or victim level. Therefore, we decided to consider rows that are identical (i.e. same event ID and same offence) as a single data point, dropping identical duplicate rows.

2. The dataset includes data about offences going back to 2000, but **the years 2000-2013 have thousands of missing rows of data**. However, the years 2014-2022 have a consistently similar number of rows of data. Therefore, we only looked at data from 2014 onwards.

3. Some offences occurred outside of the city of Toronto or occurred in **unverified locations**. These offences (~ 5,000) were not included in the analysis of neighbourhoods.

4. There are seemingly **no records that occurred at midnight** but there is an abnormal spike at noon. Offences that occurred at noon were excluded from the time of day visualizations.

Questions

1. Year to year, has there been an overall increase or reduction in crime?

2. What are the peak times for crime occurrences?

3. On average, how long after the offence date was it reported?

4. Can we identify trends or patterns in specific neighbourhoods?

5. Is there a correlation between the type of offence and the location? Is there a correlation between certain types of offences?

Question 1:
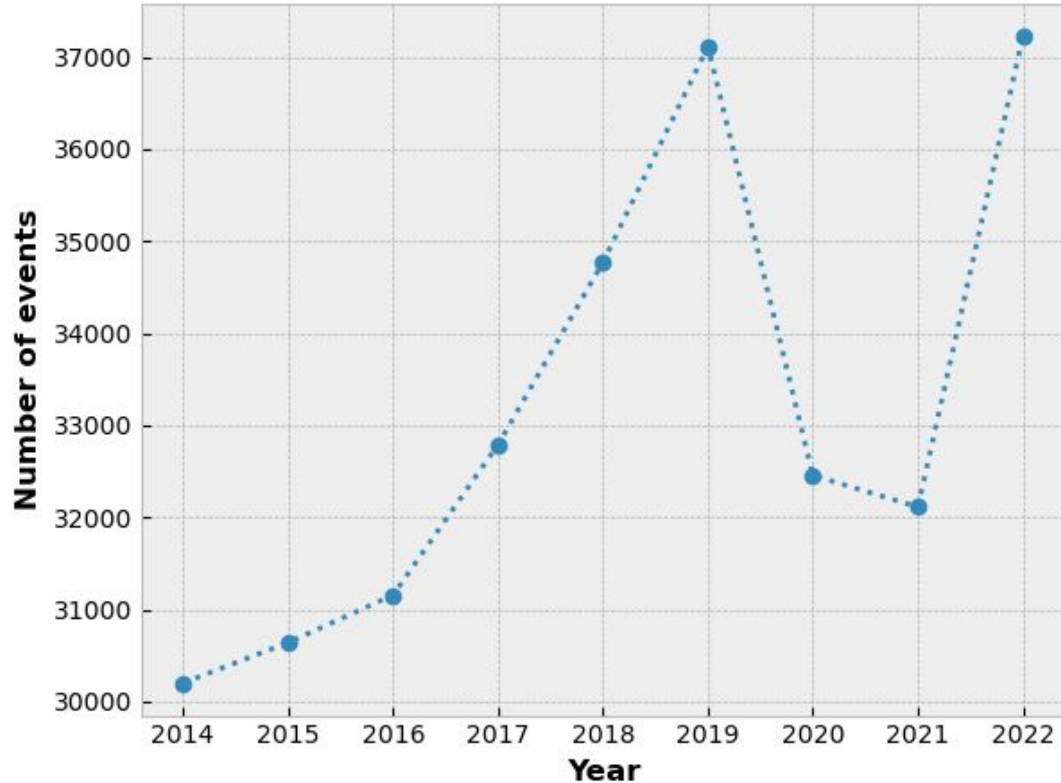Year to year, has there been an overall increase or reduction in crime?

Question 1

Year to year, has there been an overall increase or reduction in crime?

Correlation coefficient: 0.68
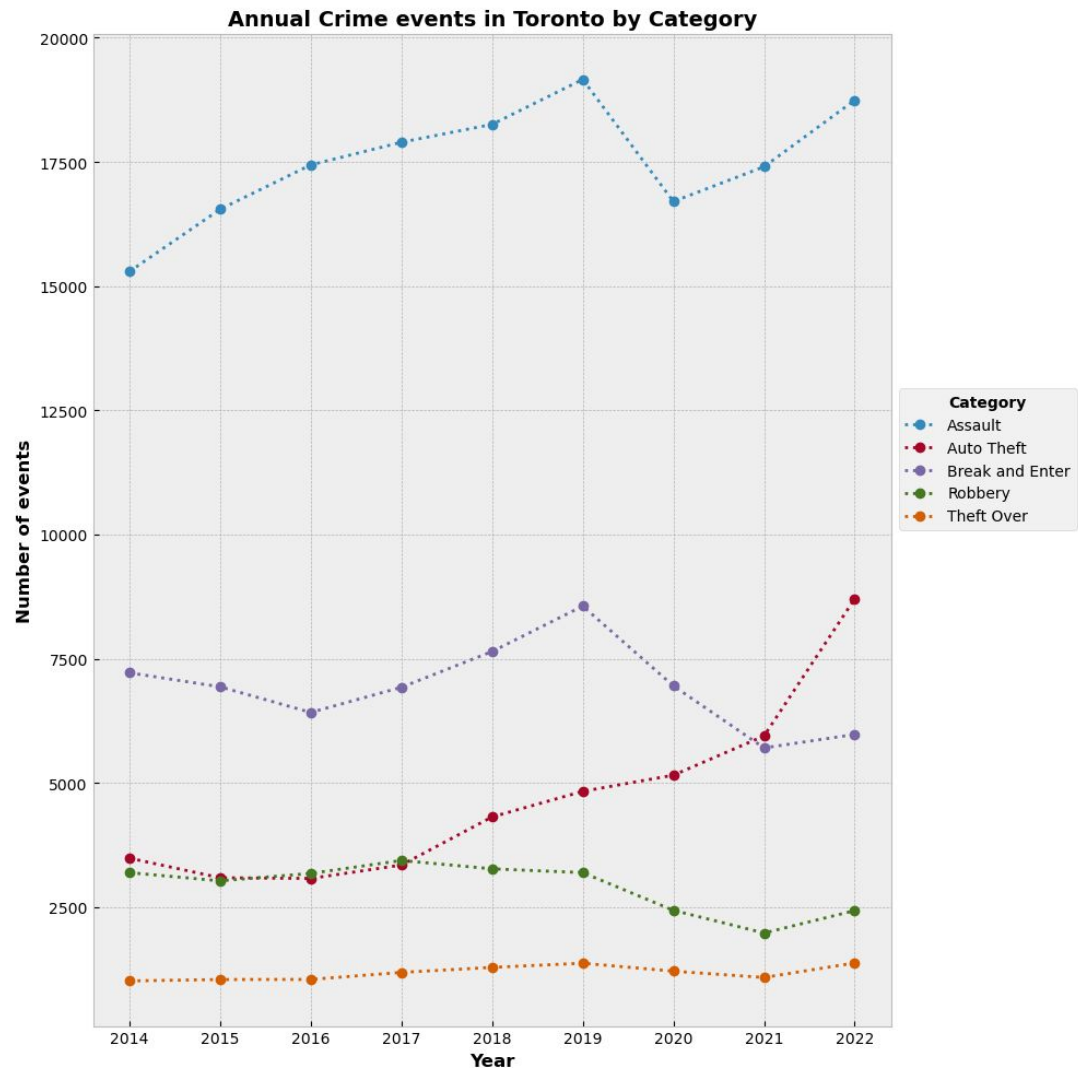

Crime events in Toronto by Year

Question 1

Year to year, has there been an overall increase or reduction in crime?

Correlation coefficients by category:
- Assault: 0.62
- **Auto Theft: 0.88**
- Break and Enter: -032
- **Robbery: -0.73**
- Theft Over: 0.67

Annual Crime events in Toronto by Category

Question 2:
What are the peak times for crime occurrences?
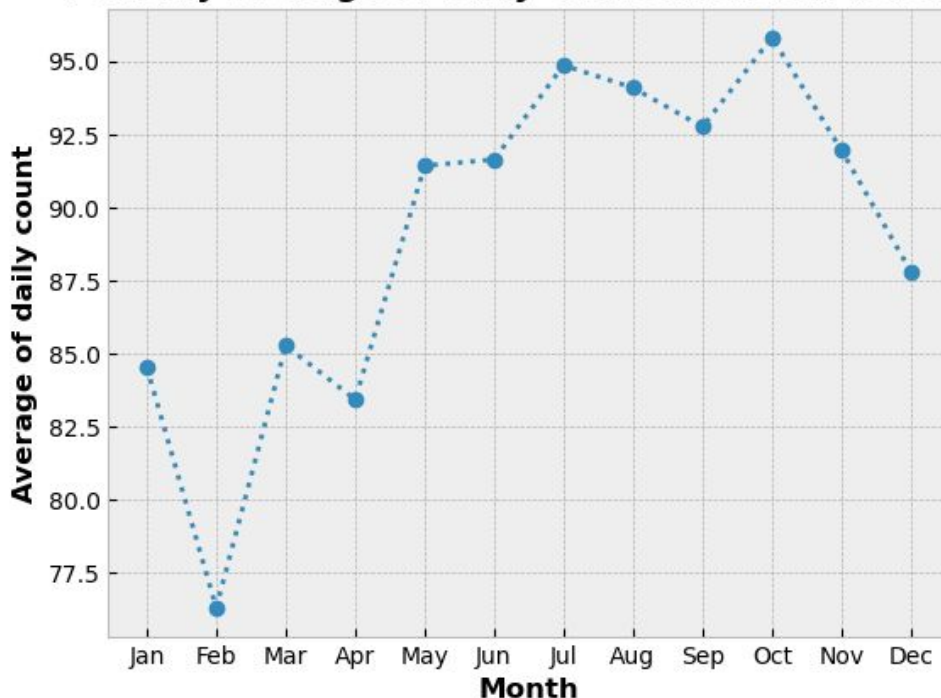
Question 2
What are the peak times for crime occurrences?

- Significant effect of month on the monthly average of daily crime events
  - One-way ANOVA ($p=9.46 \times 10^{-5}$)
- The **winter** months see the lowest daily crime counts, and the **summer** months see the highest.

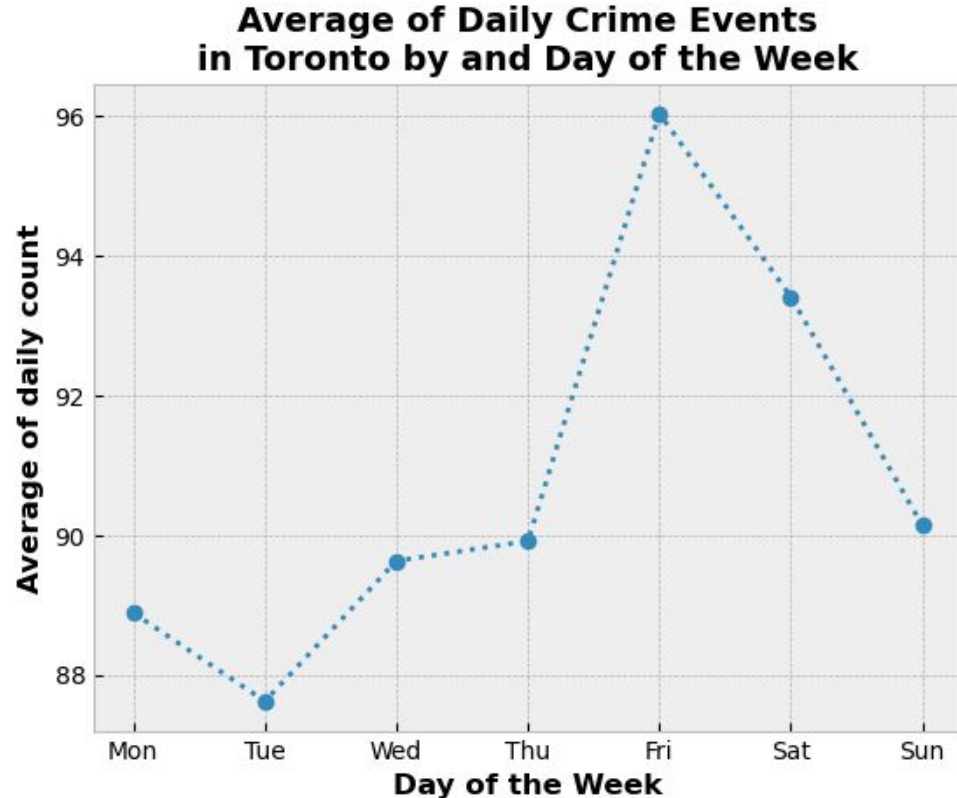**Monthly Average of Daily Crime Events in Toronto**

Question 2
What are the peak times for crime occurrences?

- Significant effect of day of the week on the average of daily crime events
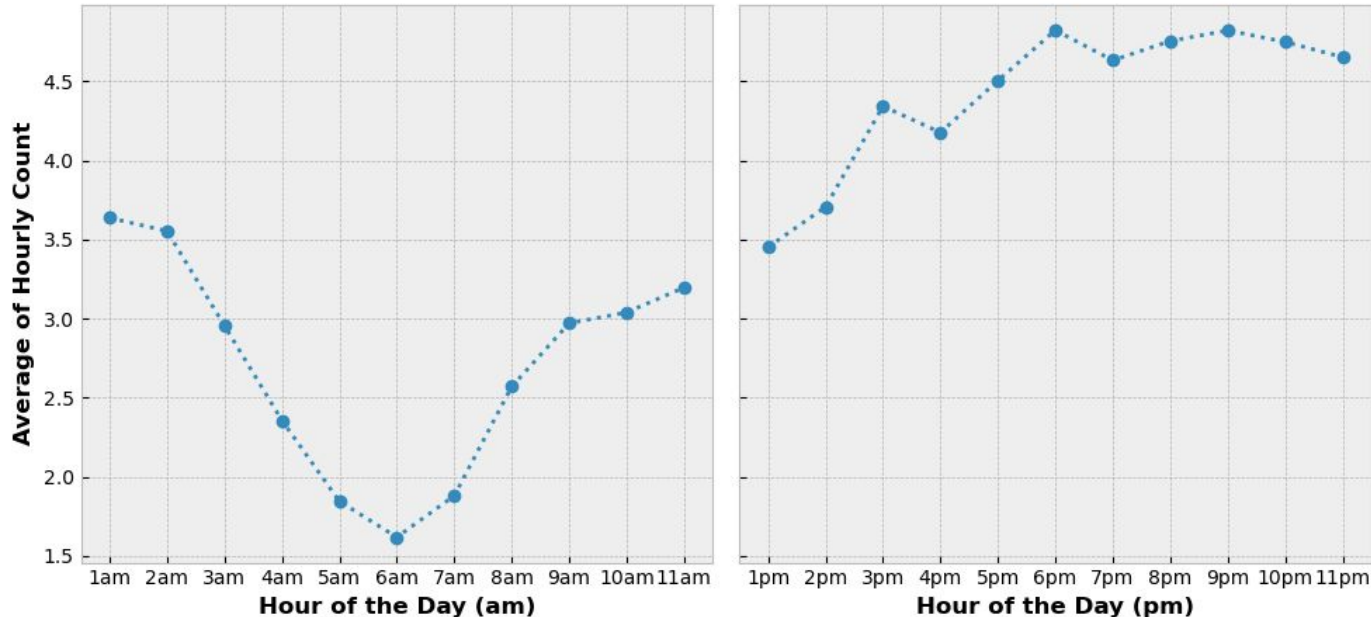   - One-way ANOVA
    - (p=5.17 x10^{-7})
- Friday has the highest crime occurrence, whereas Tuesday has the lowest

Average of Daily Crime Events in Toronto by and Day of the Week

Question 2
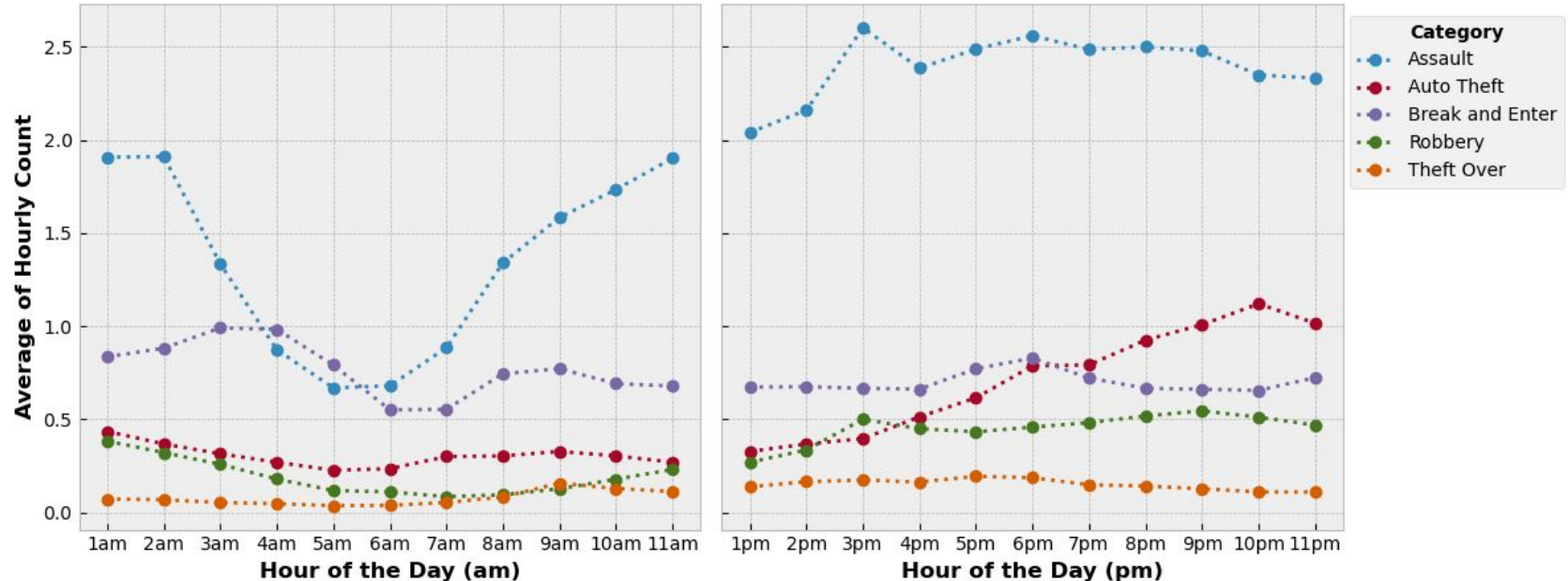What are the peak times for crime occurrences?

Average Hourly Count of Crime Events in Toronto

Question 2
What are the peak times for crime occurrences?

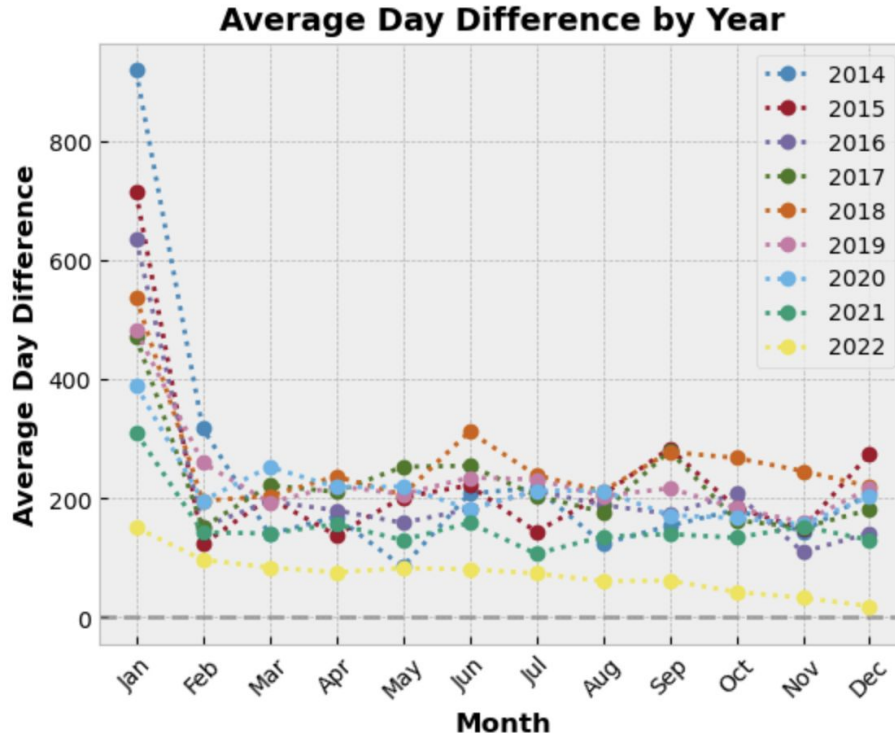**Average Hourly Count of Crime Events in Toronto by Category**

Question 3:

On average, how long after the offence date was it reported?

Question 3

On average, how long after the offence date was it reported?

Average Day Difference by Year

- January consistently records the highest reporting delays.

- Marked reduction in delay days across all months in 2022.

- Months like June and September show fluctuating delay dynamics.

Question 3

On average, how long after the offence date was it reported?

Average Day Difference by Year
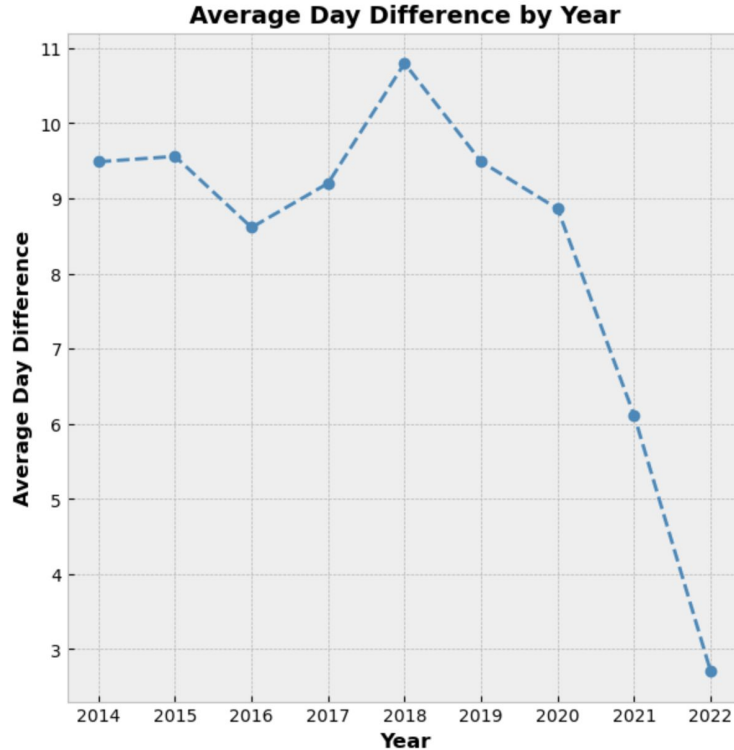
- Studied delay between offence occurrence and reporting from 2014 to 2022.

- 2014-2018: Reporting delay between 8.6 to 10.8 days.

- Peak in 2018: Average 10.8-day delay.

- Decreasing trend from 2019 onwards.

- 2021: Delay reduced to just over 6 days.

- 2022: Sharp decrease to 2.7 days.

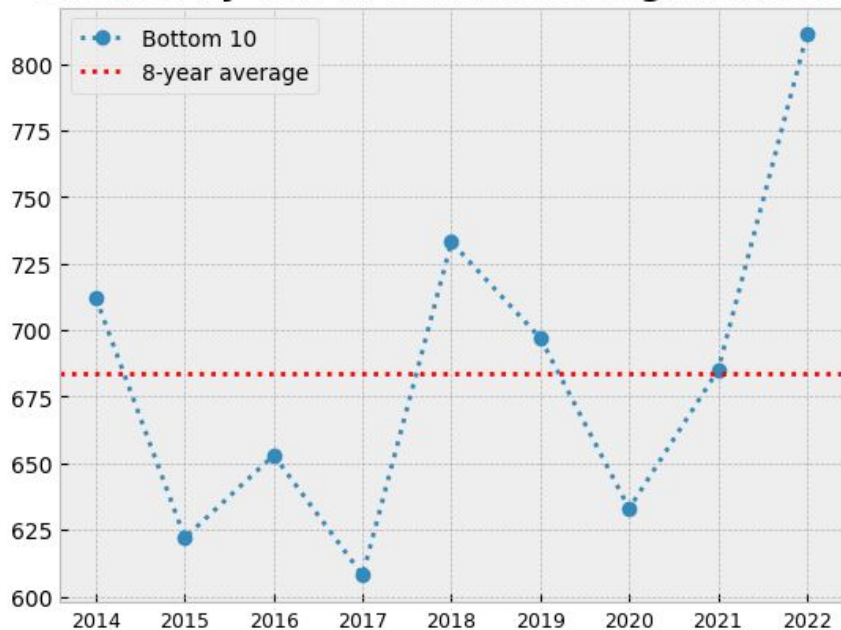- Better public awareness or improved reporting systems.

Question 4:

Can we identify trends or patterns in specific neighbourhoods?

Top and Bottom 10 Neighbourhoods

Offences by Year in Bottom 10 Neighbourhoods

Offences in the bottom neighbourhoods were less consistent in the years prior to the pandemic. They fluctuated from 2014 to 2018, with a 20% increase from 2017 to 2018. They decreased by 13% from 2018 to 2020 but have reached their 8-year peak in 2022 with an 18% increase over 2021.
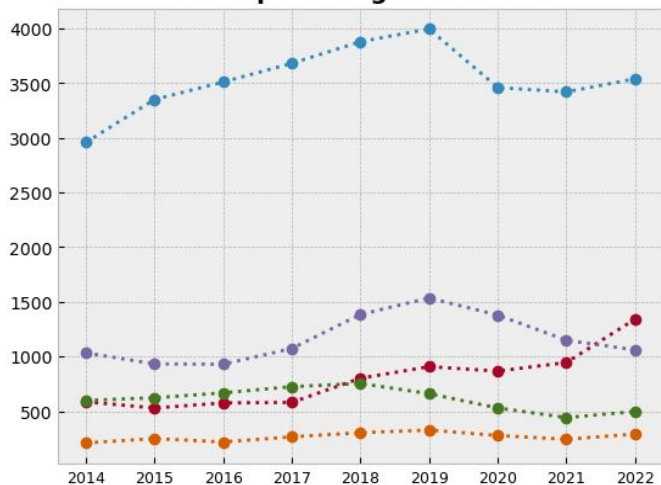
Offence Categories

Auto theft has been on the rise since 2021 in the top 10 neighbourhoods. Assault declined from 2019 to 2021 but remains the top offence category across neighbourhoods.
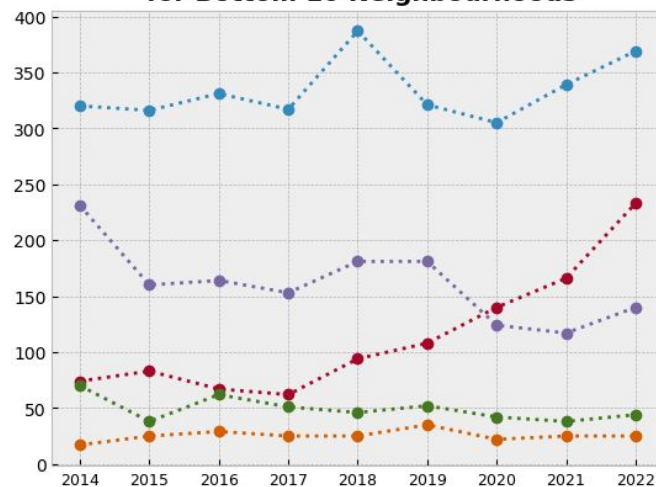
Assault and auto theft are on the rise since 2020 in the bottom 10 neighbourhoods.

Other offences have remained relatively consistent.

Offence Categories by Year for Top 10 Neighbourhoods

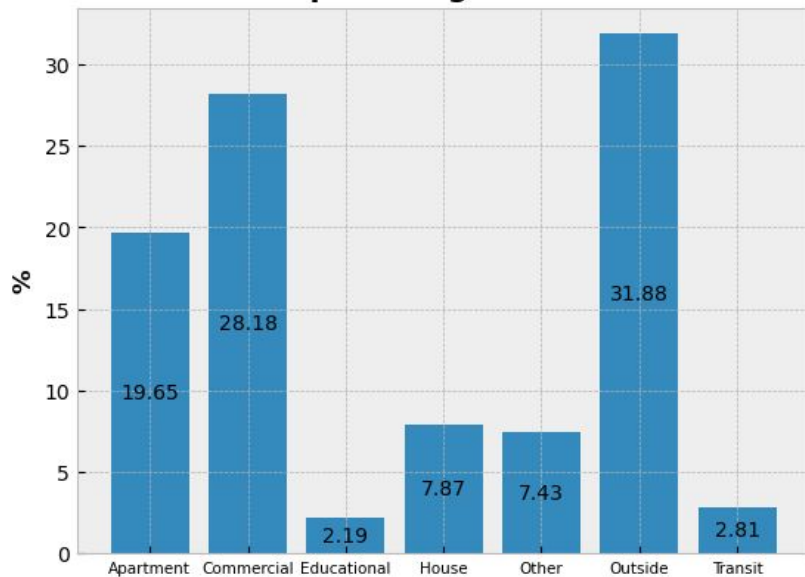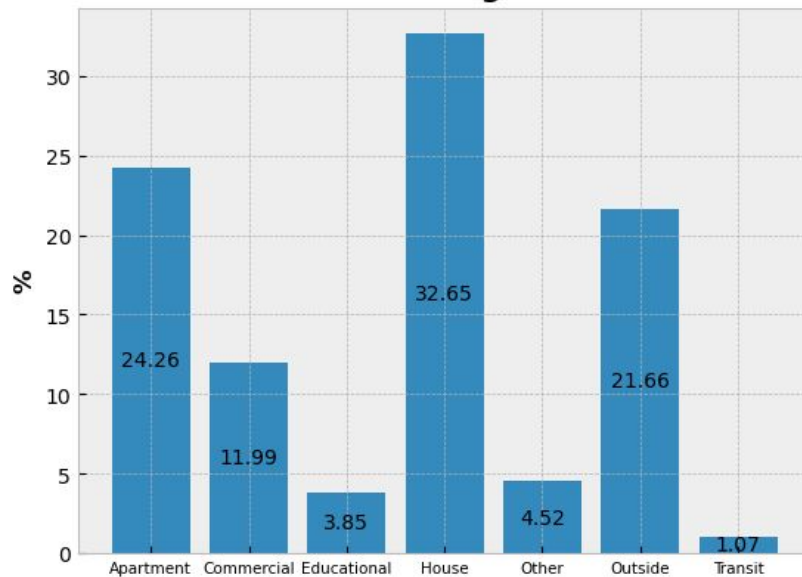**Offence Categories by Year for Bottom 10 Neighbourhoods**

Premises Type

Offences in the top 10 neighbourhoods mostly occur outside (31.8%) and in commercial places (29.2%).

Offences in the bottom 10 neighbourhoods mostly occur in houses (32.6%) and apartments (24.3%).

Premises Type by Percentage of Offences in Top 10 Neighbourhoods

Apartment: 19.65
Commercial: 28.18
Educational: 2.19
House: 7.87
Other: 7.43
Outside: 31.88
Transit: 2.81

Premises Type by Percentage of Offences in Bottom 10 Neighbourhoods

Apartment: 24.26
Commercial: 11.99
Educational: 3.85
House: 32.65
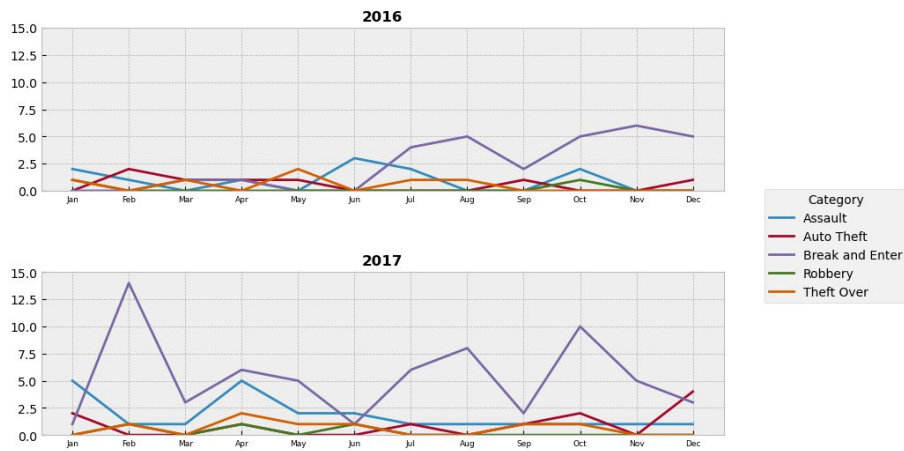Other: 4.52
Outside: 21.66
Transit: 1.07

Neighbourhood Year over Year Changes

Forest Hill South experienced the greatest positive YoY change from 2016 to 2017, with an increase in offences of 94.55%.

This was mostly due to a 120% increase in break & enter and a 100% increase in assault.

Highland Creek experienced the greatest negative YoY change from 2019 to 2020, with a decrease in offences of 52.8%.

There was a 60% reduction in break & enter, a 50% reduction in assault, a 44% reduction in auto theft, and an 84% reduction in robbery.

Offence Categories by Month for Forest Hill South, 2016-2017

Offence Categories by Month for Highland Creek, 2019-2020
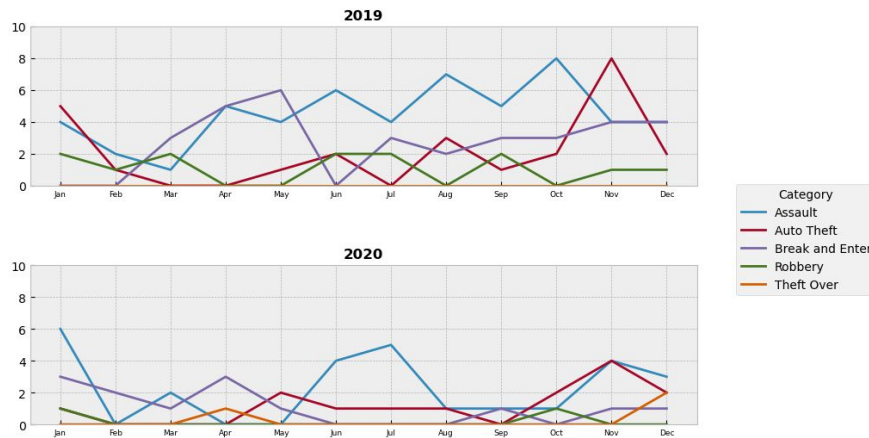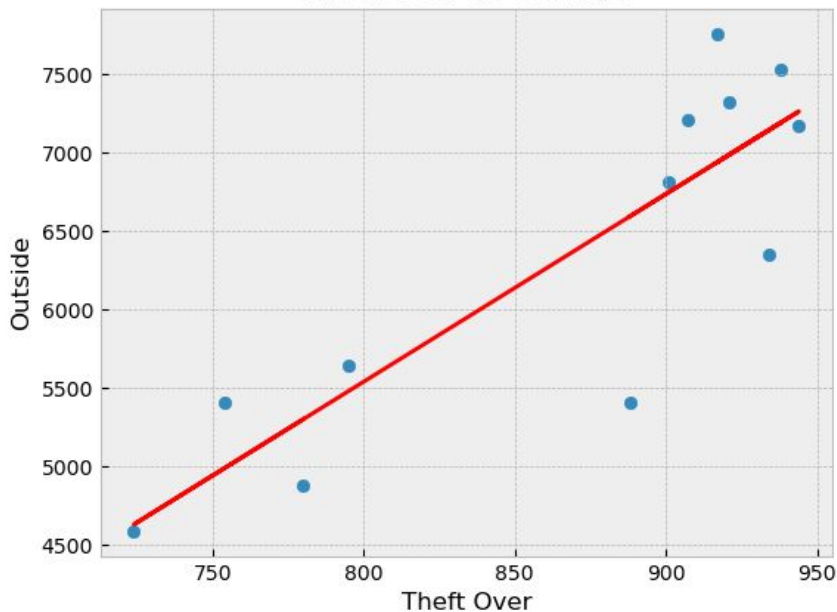
Question 5:
Is there a correlation between the type of offence and the premises?
Is there a correlation between certain types of offences?

Correlation: Premise Type and MCI Category

Theft Over vs Outside

Assault: Extremely strong correlation of 0.95 with outside premises (parking lots, streets/highways, open areas, etc.).

Auto theft: Very strong correlation of 0.9 with houses.

Break & enter: Very strong correlation of 0.77 with commercial premises and 0.87 with houses.

Robbery: Strong correlation of 0.74 with other premises (medical and religious facilities, social services, etc.).

Theft over $5,000: Very strong correlation of 0.89 with houses and 0.86 with outside premises.
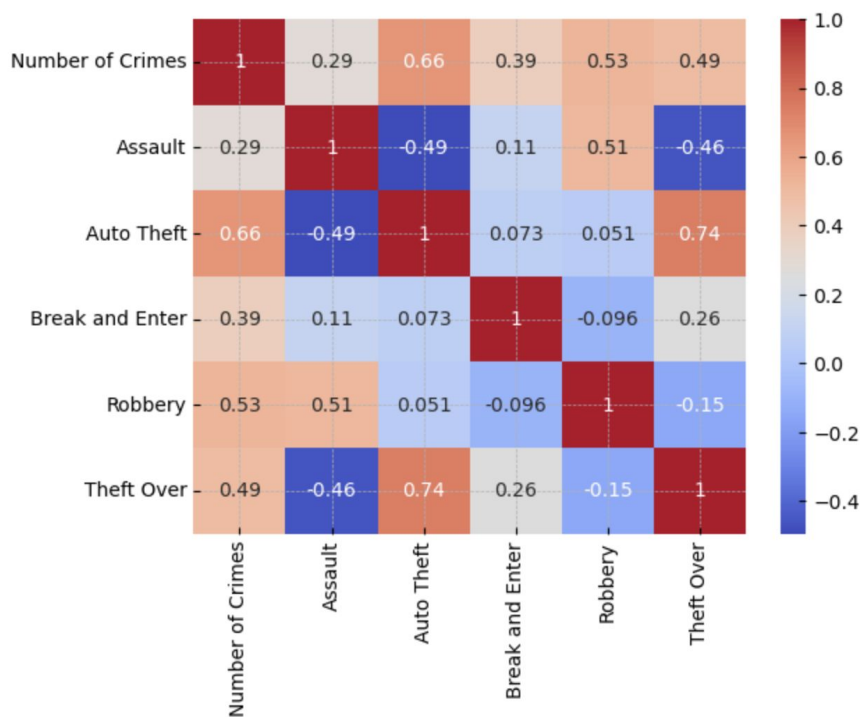
Question 5

Is there a correlation between certain types of offences?
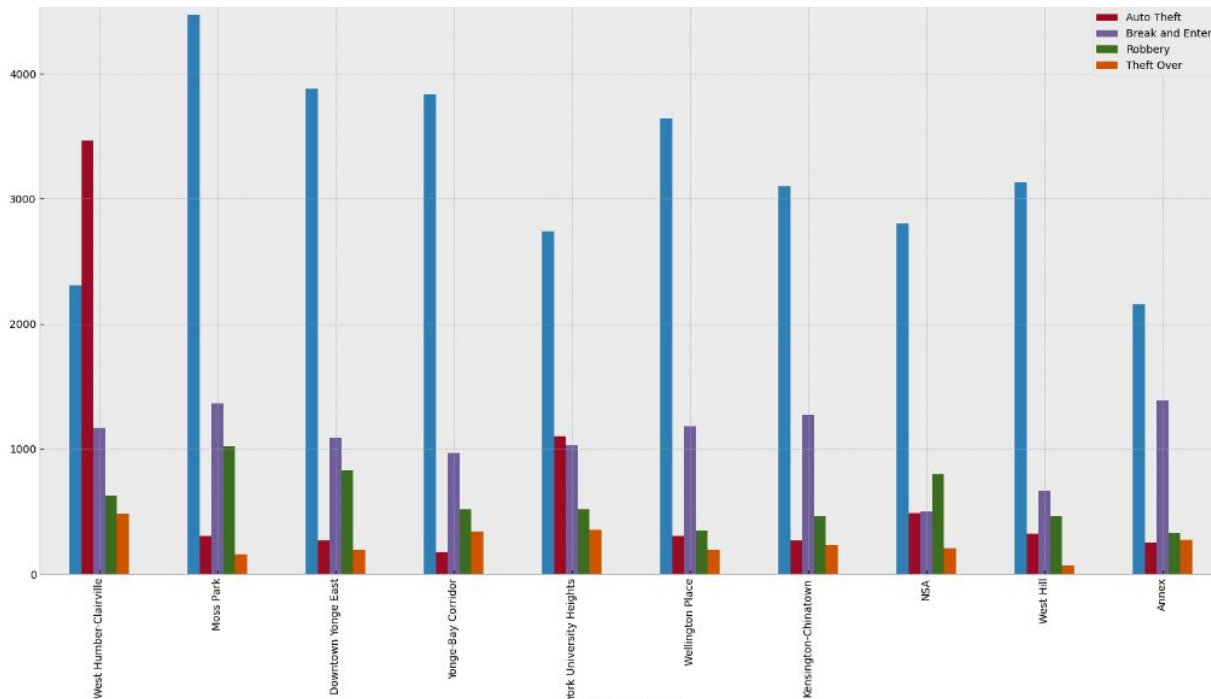
- Top 10 neighbourhoods with most crimes
- Assault is the overall most seen offence category except the first neighbourhood where auto theft is an anomaly
- Break and enter category mostly exist as the second

Question 5

Is there a correlation between certain types of offences?

- Analyzed using a correlation matrix.

- Found a strong link (0.74) between Auto Theft and Theft Over.

- Noticed a negative relationship (-0.49) between Assault and Auto Theft.

- Break and Enter showed weak ties to other crimes, suggesting unique factors at play.

- The data assists in pinpointing areas needing focused interventions.

Conclusions

1. Year to year, has there been an overall increase or reduction in crime?

   There is an upwards trend of crime in Toronto year to year, which was dampened by the pandemic. Auto theft is in very high-rise, while robberies are going down.

2. What are the peak times for crime occurrences?

   The peak times for crime are: the summer months, weekends and the evenings (~6pm to 10pm).

3. On average, how long after the offence date was it reported?

   From 2014 to 2022, the delay in crime reporting significantly reduced, 75% decrease in last four years, it might suggest heightened public awareness or potential improvements in reporting systems.

Conclusions

4. Can we identify trends or patterns in specific neighbourhoods?

    Forest Hill South experienced a 94% increase in offences from 2016 to 2017, mainly due to increases of over 100% in assault and auto theft. Highland Creek experienced a 52% decrease in offences from 2019 to 2020, with reductions across all offence categories.

5. Is there a correlation between the type of offence and the location? Is there a correlation between certain types of offences?

    Assault has a very strong positive correlation (0.95) with outside premises. Auto theft has a very strong positive correlation (0.9) with houses. The correlation matrix reveals strong ties (0.74) between Auto Theft and Theft Over, but Assault and Auto Theft trends diverge (-0.49), with Break and Enter largely independent, guiding targeted interventions.

Implication of Findings

1. Because of the upwards trend of crime in Toronto year to year, mainly in connection with auto theft and related offences, crime-stopping efforts should be focused on this.

2. Since the peak times for crime are the summer months, weekends and the evenings, these are the periods when people in Toronto should be most vigilant and take preventive actions.

3. Residents in high crime areas should try to avoid commercial locations and outdoor areas during peak times, as offences mainly occur in those locations.

Questions?