Data Description

The Canadian Anti-Fraud Centre collects fraud and identity crime reports in their Fraud Reporting System database, sourced from public reports, including those submitted online. The accuracy of these reports heavily relies on the information provided by individuals, who have the option to include varying levels of detail. This dataset has already been cleaned and processed. It contains complete information about the fraud report with the following record in each column.

Data Source
Canadian Anti-Fraud Centre Fraud Reporting System Dataset - Canadian Anti-Fraud Centre Reporting Data
https://open.canada.ca/data/en/dataset/6a09c998-cddb-4a22-beff-4dca67ab892f/resource/43c67af5-e598-4a9b-a484-fe1cb5d775b5#additional-info

Meta Data

Data Name: Canadian Anti-Fraud Centre Reporting Data.csv (61.1MB)

Data Numbers: 285855 - Non-Null Data

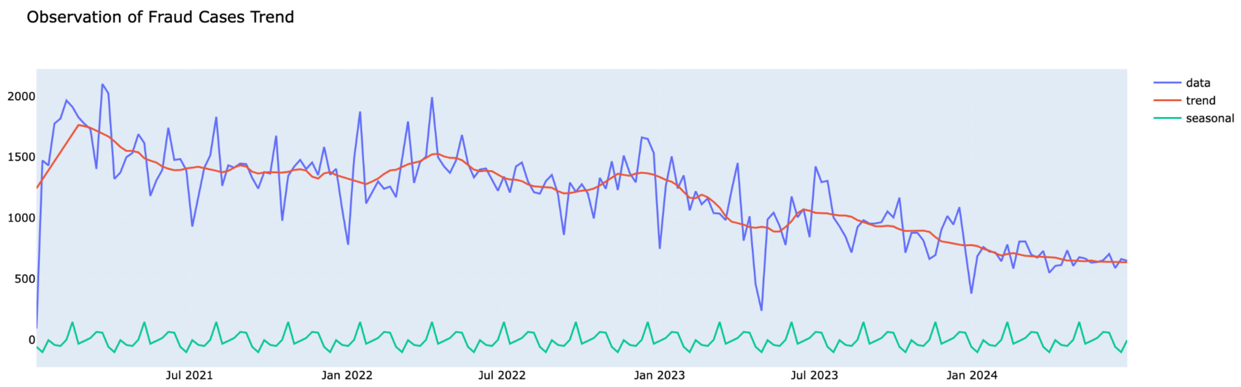
Columns: 15 (after removing duplicate columns with columns in French)

Number ID	Gender
Date Received	Genre
Complaint Received Type	Language of Correspondence
Country	Victim Age Range
Province/State	Complaint Type
Pays	Number of Victims
Fraud and Cybercrime Thematic Categories	Dollar Loss
Solicitation Method	

What are the Trends & Patterns of Cybercrime from 2021 to 2024 in Canada?

Trend Observation

Detect fraudulent activities through data visualization to identify trending patterns in relation to time factors. We apply an *Additive Model* to extract time series data to trend, seasonal and residual. What we can tell from the plot below is that the trend decreases annually..



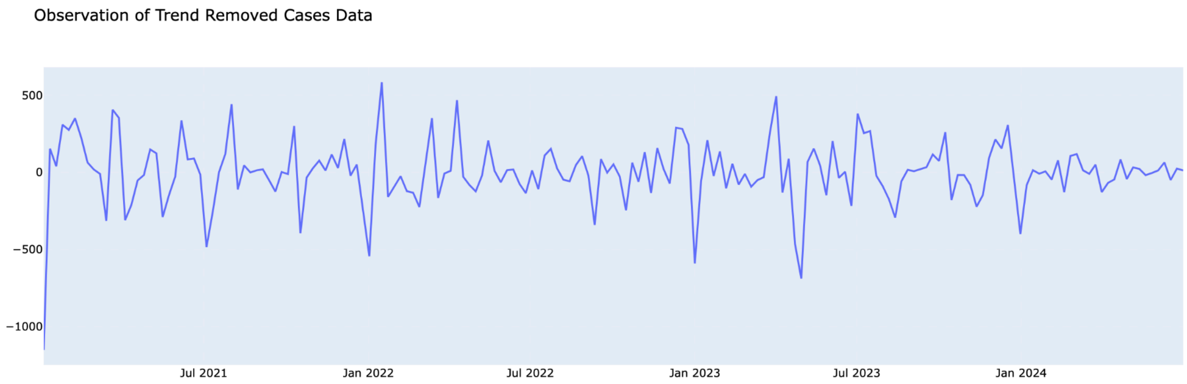
$$*Additive\ Model = Trend + Seasonal + Residual$$

Seasonal Analysis

Trend Removing

To effectively conduct seasonal analysis, it is essential to eliminate the trend component from the observed data. There are two primary methods for achieving this:

1. **Additive Model**: This approach involves calculating the *observed data minus the trend component.* It is straightforward and quick; while it may not completely eliminate all trends, it provides a clearer signal from the observed data.
2. **First-order Difference**: This is the most commonly used method in Time Series Analysis. It helps in making the data more stationary, but it may also result in the loss of important signals from the observed data.



By observed data minus the trend component

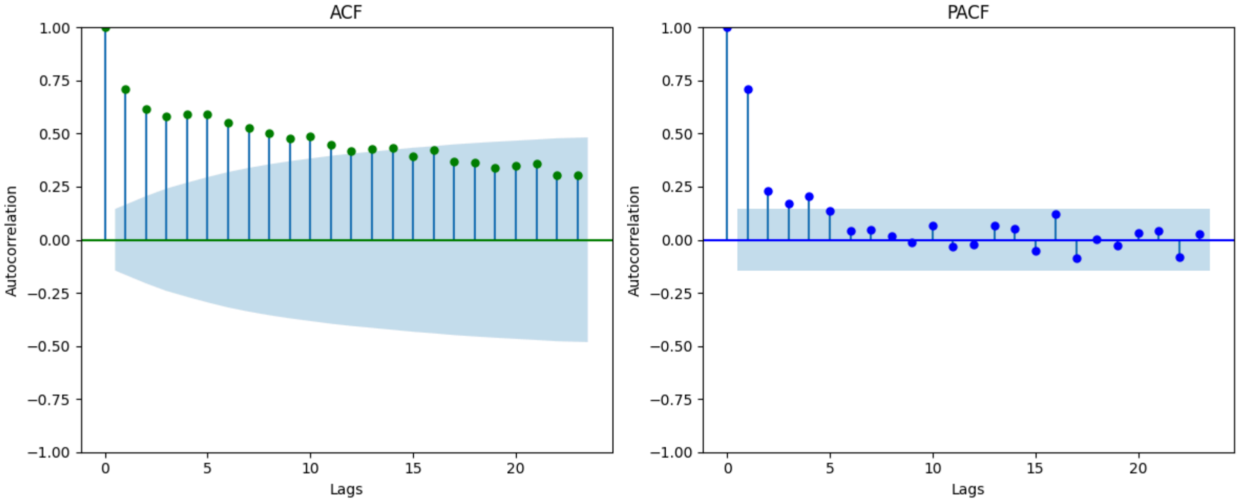
Periodicity Impact Analysis

To determine the routine cycle of fraud data trends, we utilize the ACF and PACF methods to calculate the periodicity.

ACF: This method assesses how today's value correlates with values from several previous time steps.

PACF: This method evaluates how today's value is influenced directly by a specific past value, while disregarding the values in between.

By analyzing the results from the ACF and PACF data, we can effectively identify the cycle period that has the most significant impact on the data points at any given time.

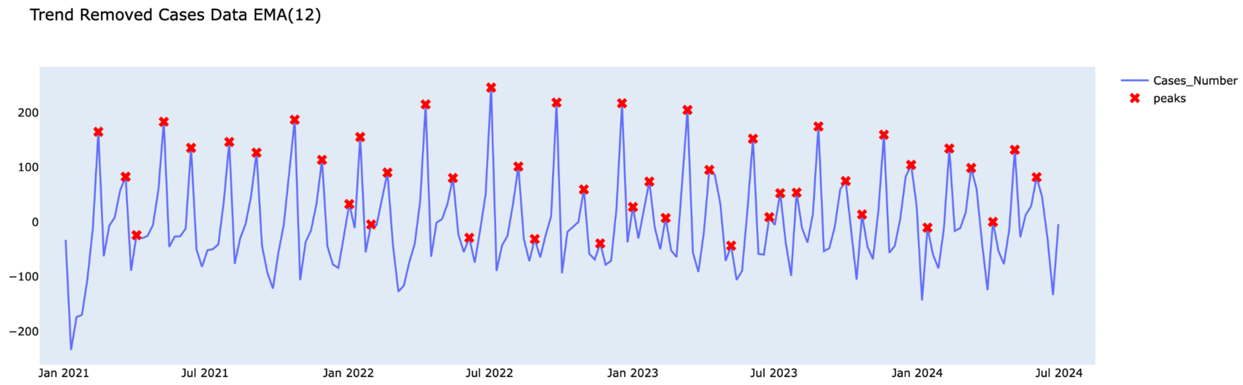


The seasonal period is identified as 12 weeks, or 3 months, based on the ACF. This indicates that the data at any given time is influenced by the data from the preceding 3 months. Additionally, the PACF reveals that the data at any point is also directly impacted by the data from 4 weeks, or 1 month, prior. This suggests that certain routines occur on a monthly basis.

Intensity Frequency Analysis

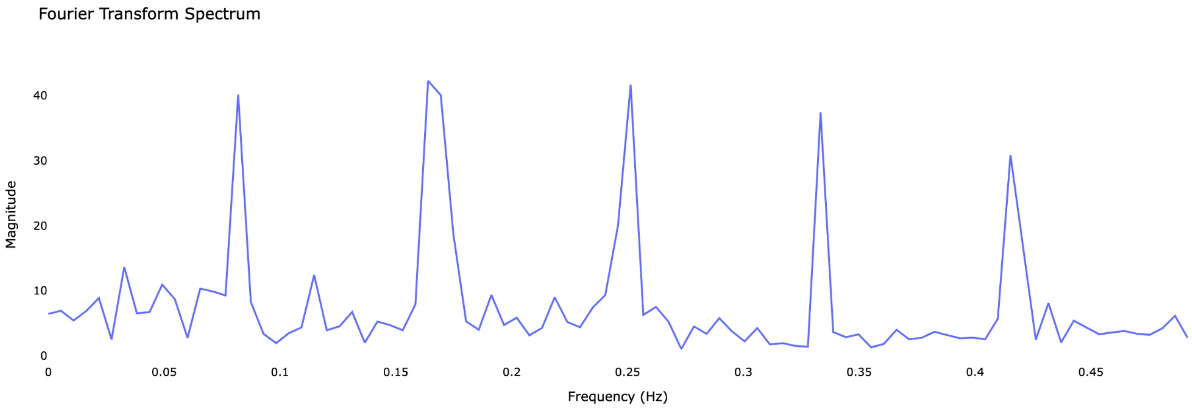
After thoroughly analyzing the regular pattern period, we can utilize the Exponential Moving Average (EMA) to assess and confirm the seasonal intensity of data that does not exhibit a trend. This method allows us to pinpoint specific periods when fraud cases are on the rise and when they are decreasing.

We prefer the Exponential Moving Average (EMA) over the Simple Moving Average (MA) because it minimizes the influence of historical data on future predictions, as illustrated in the Autocorrelation Function (ACF) plot. These factors are essential to our analysis.



The data reveals distinct peaks and valleys, indicating a seasonal pattern in the frequency of cybercrime cases. Consequently, we aim to identify the features and characteristics associated with these peak occurrences.

To begin with, we aim to comprehend the frequency of patterns. While we are aware of the routine period, the intensity frequency remains unclear. To uncover this information, we will utilize the *Fourier Transform.*



These data show that cybercrime occurs with varying frequencies, peaking (top3) at the following intervals:

- 0.08 Hz with a peak of 40.15, occurring approximately every 12.20 weeks.

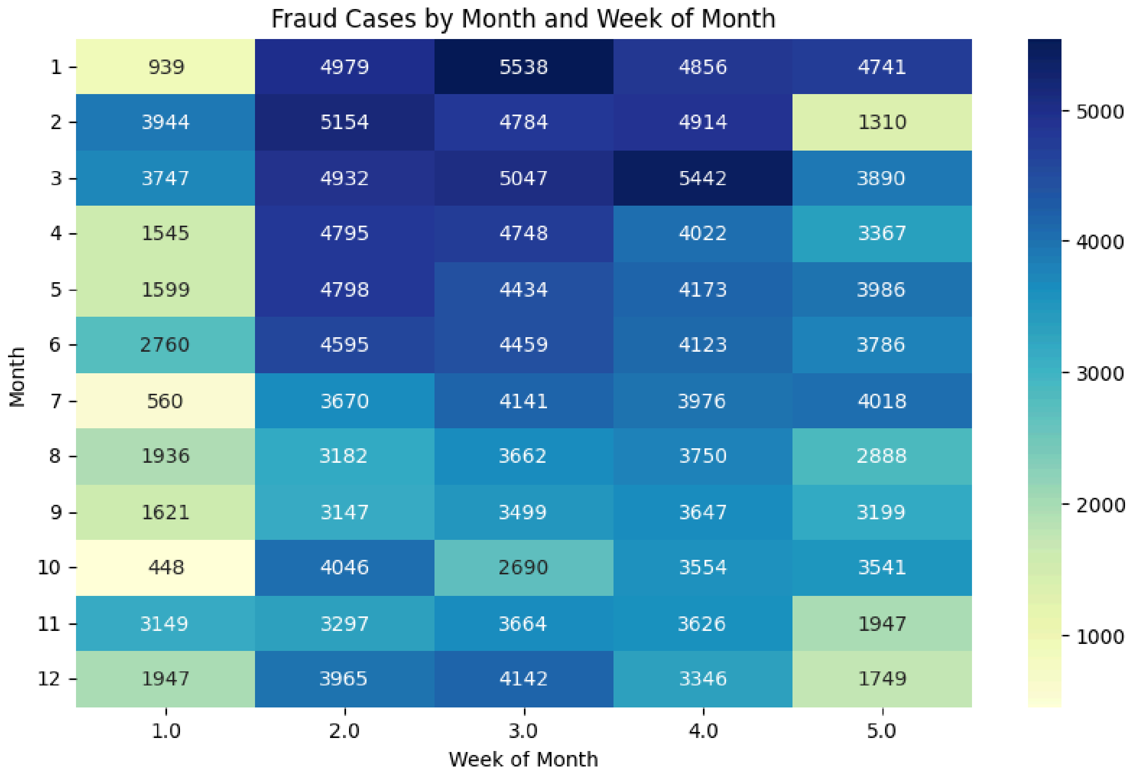
- 0.16 Hz with a peak of 42.26, occurring every 6.10 weeks.
- 0.25 Hz with a peak of 41.66, occurring approximately every 3.98 weeks.

In summary, these frequencies indicate that cybercrime activity exhibits periodic fluctuations ranging between 2.41 and 12.20 weeks, with certain frequencies (such as 0.16 Hz and 0.25 Hz) showing more pronounced peaks. This periodic variation helps to predict the frequency and trends of cybercrime incidents.

Week of Month Pattern Analysis

We performed a *Chi-square* test to assess whether the frequency of fraud cases varied significantly across different weeks of the month and across various months. The objective was to identify any discernible patterns in the timing of fraud incidents, which could help reveal **if particular weeks within a month are more susceptible to fraud, depending on the month**.

Chi-squared: 15472.28, p-value: 0.0, df=55



The findings from the Chi-square test showed a **p-value below 0.05**, suggesting a statistically significant relationship between the month and the specific weeks within each month. This indicates that **fraud cases in certain months are likely to peak during specific weeks**, rather than being uniformly distributed throughout the month.

Conclusion

Based on our thorough analyses, we've gathered some key takeaways that shine a light on the patterns and risks tied to fraud cases during the reviewed timeframe:

1. **Downward Trend in Fraud Cases**: The data shows a clear and significant **downward trend** in the overall number of fraud cases throughout the period. This could mean that our existing preventive measures are hitting the mark, or perhaps there's a natural decline in fraud occurrences. That said, we shouldn't let our guard down, as there may still be occasional spikes that need our ongoing attention.

2. **Periodic Patterns in Fraud Peaks**: Even with the overall decline, we've noticed some periodic patterns, particularly with cycles of **12 weeks and 4 weeks**. This suggests that fraud cases tend to peak at regular intervals. **Looking ahead, we anticipate continued peaks at specific time intervals**, especially around the **4-week, 6-week, and 12-week cycles**. This insight is invaluable for forecasting, as it highlights those predictable periods of heightened risk. Organizations can use this information to **proactively allocate resources** and ramp up monitoring during these peak times to help mitigate potential losses.

3. **Chi-square Test Results**: The findings from the **Chi-square test** further confirm a **strong correlation between the month and the week of the month** concerning fraud cases. This indicates that fraud incidents aren't spread evenly throughout the month but are more concentrated in certain weeks, depending on the time of year. Understanding this correlation can provide organizations with valuable insights for **risk forecasting** and strategic planning. By knowing when they are most vulnerable to fraud, businesses can allocate resources more efficiently, such as boosting fraud detection efforts or adding staff during those high-risk periods.

4. **Implications for Risk Management**: These insights offer solid support for organizations, guiding them to make **data-driven decisions** related to **risk management and resource allocation**. By grasping the predictable nature of fraud patterns—both in terms of periodic peaks and specific weeks within a month—businesses can plan ahead to minimize risks.