## **Time Based Anomaly Detection**

### **Objective:**

Analyze sales trends over time.

Detect seasonal variations and holiday effects on sales.

Use time-series analysis for understanding store and department performance over time.

Detects anomalies in time series Sales data using specialized methods.

💡 Approach:

📊 Create Rolling Statistics: Calculate rolling averages, moving sums, and standard deviations to track data variations over time.

🔄 Apply Exponential Smoothing: Use exponential moving averages (EMA) and anomaly detection techniques to highlight unusual patterns in time series data.

🚨 Highlight Anomalies: Set thresholds and visual cues to identify significant deviations from smoothed values, aiding in anomaly detection and trend analysis.

**Time Based Anomaly**

Time-based anomaly detection refers to the process of identifying unusual patterns or deviations from expected behavior in time-series data. This technique is commonly used in various fields such as cybersecurity, finance, manufacturing, and healthcare to detect abnormal events or anomalies that may indicate potential problems or threats.

Some common techniques used in time-based anomaly detection include:

* Statistical Methods: Such as mean, median, standard deviation, z-score, or percentiles to detect deviations from the expected statistical properties of the data.
* Machine Learning Algorithms: Including supervised learning methods like Support Vector Machines (SVM), Random Forests, or neural networks, as well as unsupervised learning techniques like k-means clustering, isolation forests, or autoencoders.
* Time-Series Analysis: Techniques like seasonality decomposition, trend analysis, autocorrelation, or spectral analysis to identify abnormal patterns in the time-series data.

Overall, time-based anomaly detection plays a crucial role in monitoring and maintaining the integrity, security, and reliability of systems and processes by identifying and mitigating potential threats or abnormalities in time-series data.

## **Challenges in detecting anomalies in time-based data:**

Temporal shifts refer to changes or variations in patterns, trends, or distributions over time in time-based data. These shifts can manifest in various ways and can complicate anomaly detection, as they often make it challenging to distinguish between genuine anomalies and shifts that are part of the normal evolution of the data. Here's a more detailed look at temporal shifts in time-based data, how they complicate anomaly detection, and strategies to account for them:

### **Definition and Examples of Temporal Shifts:**

* **Seasonal Shifts:**These are regular patterns that occur over specific time intervals, such as daily, weekly, or annually. For example, retail sales may see a surge in December due to the holiday season.
* **Trend Shifts:** A change in the overall direction of data points over time. For instance, a tech company's stock price might have a long-term upward trend, but this trend could shift to a downward direction due to market changes.
* **Sudden Events:** Unpredictable events like natural disasters, political upheavals, or unexpected product failures can introduce sudden shifts in time-based data. For example, a sudden drop in website traffic due to a server outage.
* **Gradual Changes:** These shifts occur slowly over time and may be caused by factors like changing consumer preferences, technological advancements, or demographic shifts. An example is the gradual decline in the use of landline phones.

### **How Temporal Shifts Complicate Anomaly Detection:**

Temporal shifts complicate anomaly detection because they can lead to variations in the data that are not necessarily anomalous but are due to the underlying dynamics. Anomaly detection models trained on historical data may struggle to adapt to these shifts, resulting in increased false positives or false negatives. For example, if you have a model trained on retail sales data from non-holiday periods, it may flag higher sales during holiday seasons as anomalies if it doesn't account for temporal shifts.

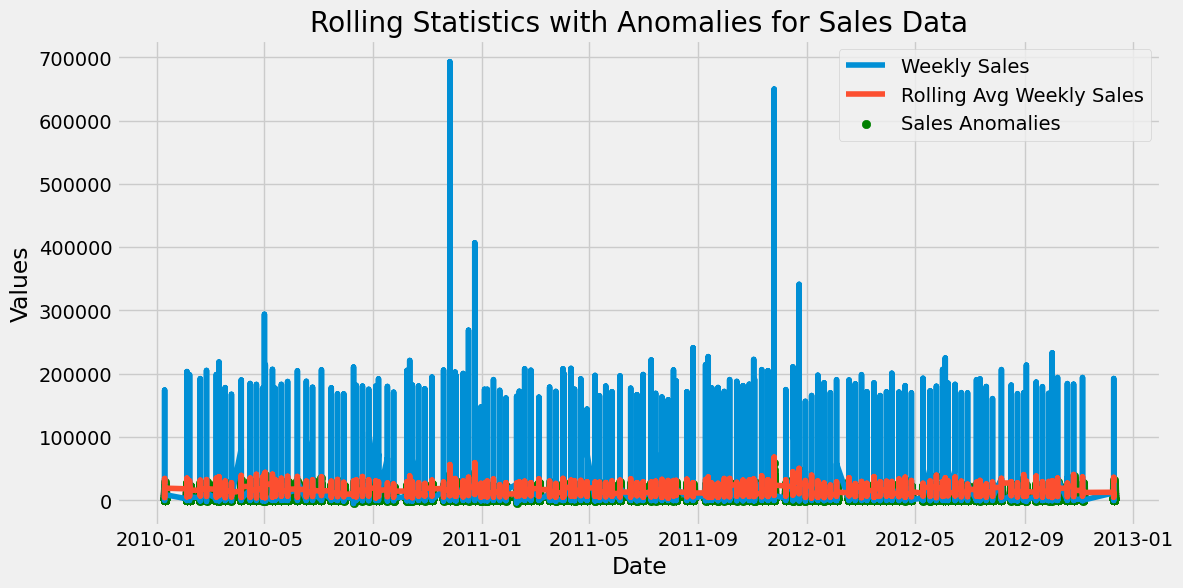
### **Strategies to Account for Temporal Shifts in Anomaly Detection:**

* **Rolling Statistics:** Calculate rolling statistics (e.g., moving averages or moving standard deviations) over a fixed window of historical data. Use these statistics as a reference to detect anomalies in new data. This approach can help capture gradual shifts in data.
* **Seasonal Decomposition:** Decompose time series data into its seasonal, trend, and residual components using techniques like seasonal decomposition of time series (STL). Anomalies can then be detected in the residual component, which should be free from seasonality and trend.
* **Prophet Model (Facebook Prophet):** Prophet is a forecasting model that handles time series data with seasonality and trend components. It can be used for anomaly detection by comparing predicted values with observed values. Significant deviations can be flagged as anomalies.
* **Machine Learning Models:** Train machine learning models, such as Random Forest, LSTM (Long Short-Term Memory), or ARIMA (AutoRegressive Integrated Moving Average), with features that capture temporal information (e.g., lag values, seasonal indicators). These models can learn to adapt to temporal shifts.

We have used Rolling statistics to detect the anomalies

Rolling statistics, also known as rolling or moving averages, are statistical measures calculated over a rolling window or moving time frame within a time-series dataset. Instead of calculating the statistic for the entire dataset at once, rolling statistics are computed iteratively for each subset of the data defined by the rolling window.

The rolling window is a fixed-size time interval or number of observations that moves sequentially through the dataset. At each step, the statistic is calculated using the data points within the window, and then the window advances by one observation, including the next data point and excluding the oldest one.



**Observations:**

The graph shows the rolling statistics with anomalies for sales data.

The rolling average of sales does not show steadily increasing or decreasing over time, indicating that overall sales are trending constant.

There are a few anomalies in the sales data, as indicated by the green dots on the graph. These anomalies could be caused by a variety of factors, such as markdown events, seasonal fluctuations, or holidays.

**Exponential Weighted Moving Average (EWMA)**

Exponential Weighted Moving Average (EWMA) is a statistical technique used for smoothing time-series data by assigning exponentially decreasing weights to observations over time. Unlike simple moving averages, where all observations in the window are weighted equally, EWMA gives more weight to recent data points while gradually decreasing the influence of older observations.

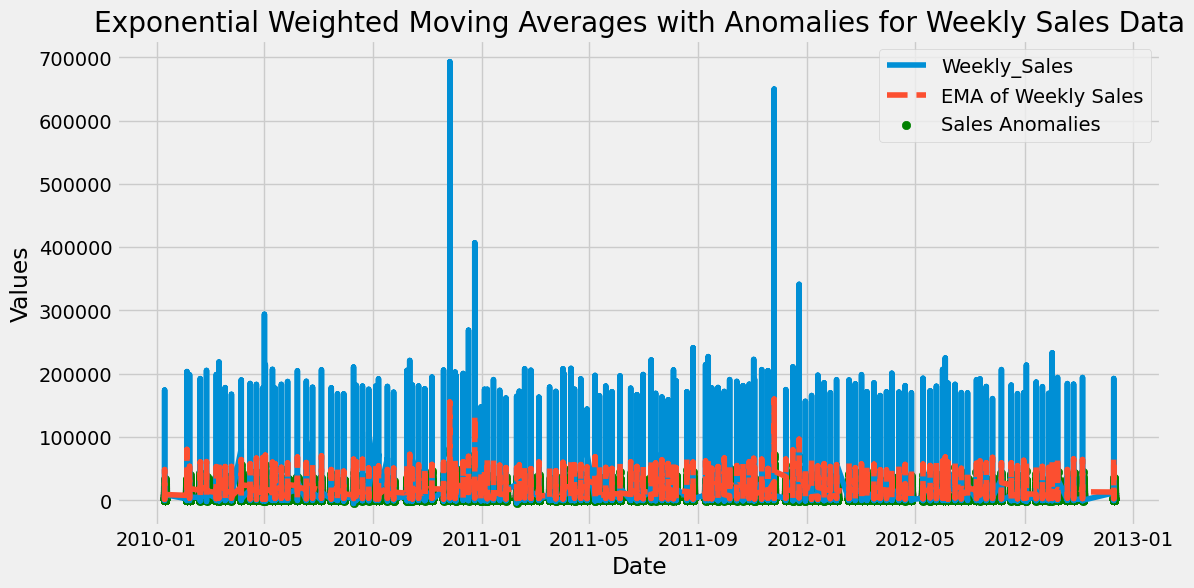
Key features and advantages of EWMA include:

1. Adaptive Smoothing: EWMA adapts to changes in the underlying data pattern by giving more weight to recent observations. This makes it more responsive to short-term fluctuations and trends compared to simple moving averages.
2. Efficient Computation: Since EWMA only requires the previous moving average and the current observation, it can be efficiently computed in real-time or on streaming data without storing the entire history of observations.
3. Flexibility: The smoothing factor
4. 𝛼
5. *α* allows users to control the degree of smoothing and the trade-off between responsiveness and stability. Higher values of
6. 𝛼
7. *α* result in smoother curves but may lag behind abrupt changes in the data.
8. Noise Reduction: EWMA helps reduce the impact of random noise or outliers in the time series, resulting in a smoother and more interpretable signal.

Applications of EWMA include:

* Financial Analysis: EWMA is commonly used in finance for risk management, portfolio optimization, and technical analysis of stock prices and financial indicators.
* Demand Forecasting: EWMA can be applied to smooth historical demand data and generate forecasts for future demand, helping businesses optimize inventory levels and production schedules.
* Quality Control: EWMA is used in manufacturing and process industries for detecting deviations from target values in quality control metrics such as temperature, pressure, or chemical concentrations.

Overall, Exponential Weighted Moving Average is a versatile and widely used technique for smoothing time-series data, capturing underlying trends, and making data-driven decisions in various domains.



**Observations:**

The exponential Weighted moving average (EWMA) of Amazon sales data has been nearly steady with a few unexpected spikes over time, indicating that overall sales are trending constant with a few seasonal fluctuations.

There are a few anomalies in the sales data, as indicated by the green dots on the graph. These anomalies could be caused by a variety of factors, such as special holidays, seasonal fluctuations, or unexpected events.

#### **Conclusion:**

In this project, we employed time-based anomaly detection techniques to analyze sales trends and detect irregular patterns in time-series data. Our approach involved the creation of rolling statistics and the application of exponential smoothing methods to identify anomalies and understand store and department performance over time.

The analysis of rolling statistics revealed insights into sales trends over time. We observed that the rolling averages of sales did not exhibit a steady increase or decrease, indicating that overall sales were relatively constant. However, the presence of anomalies, as indicated by the green dots on the graph, suggested deviations from expected patterns. These anomalies could be attributed to various factors such as markdown events, seasonal fluctuations, or holidays.

Similarly, the application of exponential weighted moving average (EWMA) allowed us to smooth the sales data and identify unexpected spikes or dips. While the EWMA of Amazon sales data remained relatively steady, occasional anomalies were detected, signaling potential disruptions in the sales pattern. These anomalies could be attributed to special holidays, seasonal fluctuations, or unexpected events impacting sales performance.

Overall, our analysis provided valuable insights into sales trends and anomalies in the time-series data. By understanding the underlying patterns and identifying irregularities, businesses can make informed decisions regarding inventory management, marketing strategies, and resource allocation. Moving forward, continued monitoring and analysis of time-series data will be essential for detecting anomalies and optimizing business operations in a dynamic and evolving market environment.