Time Series Data Characteristics

Time series datasets are defined not only by the individual data values but also by their temporal orders which are

- Temporal Dependency: Past values influence future values.
- Seasonality: Regular repeating patterns over specific intervals (like daily or monthly).
- Trends: Long-term upward or downward movements in the data.

Types of Anomalies in Time Series

Anomalies in time series are in 3 main types: point, contextual, collective. For this project, we will mainly work with contextual and collective for this project due to their effectiveness in time series.

Point Anomalies:

- A single data point that deviates significantly from the rest of the data.
- Detection Techniques:
- Statistical thresholds using mean, median, and standard deviation, Z-score
- Clustering-based algorithms and methods like DBSCAN to flag isolated points.
- Isolation Forest (works for time series)
- Note: Noise can lead to false positives for point anomalies

Contextual Anomalies:

- Data points that are anomalous in a specific context. In time series, context is defined by the trend, seasonality, or local temporal behavior.

Challenges & Addressing Them

- Breaking down the time series into different parts of trend, seasonal, and residuals will make it easier to determine and understand the context
- The context of a time series is dynamic and changes with time, making it challenging to capture with static thresholds but we can address this by focusing in on different parts of the time series after having broken it up

Time Series–Specific Detection Approaches and Implementations

- Prediction-based Methods: Train models (such as LSTMs, RNNs, or even statistical methods like SARIMA) to forecast the next data point. Should also work effectively in time series
- Large prediction errors that are relative to a like a set dynamic threshold might indicate the presence of a context anomaly.
- Break down the time series into trend, seasonal, and residual parts (using STL or wavelet methods). Then, analyze the leftovers/residuals perhaps using a rolling window-based threshold.
- Use dynamic thresholds that adjust based on recent seasonal or trend behavior, which is especially important when seasonal patterns vary over time.

Collective Anomalies:

- A sequence or a group of data points that, when taken together, deviate from the expected pattern, even though individual points might not be anomalous on their own.

Challenges:

- Depending on the relationships between multiple sequential data points rather than on isolated values, hard to comprehend this relation

Time Series–Specific Detection Approaches and Implimentations:

- Pattern Recognition with Sliding Windows: Apply sliding window techniques to segment the time series into different sub-sequences. Then compare each sub-sequence to a model of normal behavior and compare that in order to detect unusual sequences and behavior.
- Change Point Detection: Identify abrupt changes or shifts in the statistical properties (mean, variance) of the time series. Sudden changes can indicate the start or end of a collective anomaly. Algorithms like Bayesian change point detection or cumulative sum can often be effective in these situations
- Utilize LSTMs or GRUs which can be used to learn the underlying temporal patterns over longer sequences. A high reconstruction error or significant deviation in prediction over a window can indicate the presence of a collective anomaly.