IQR Method Refinement

1. Skew-Adjusted IQR Thresholding

The standard IQR was calculated for each sensor variable to determine outliers. To enhance detection accuracy, skew-adjusted bounds were applied, taking into account each feature's skewness, which dynamically adjusted the thresholds.

Skewness was calculated using

```
skew = data[column].skew()
```

which measures the asymmetry in the data distribution for each sensor reading:

- **Positive skew**: Indicates a longer or fatter tail on the right side of the distribution.
- **Negative skew**: Indicates a longer or fatter tail on the left side.
- **Zero or near-zero skew**: Suggests a symmetric distribution.

Parameter Values:

- **Multiplier (k):** 1.5, adjusted based on skewness for each feature.
- **Skew Adjustment:** For each feature, the multiplier was modified based on skewness as follows:

  Lower Multiplier= 1.5 * (1 - skew)

  Upper Multiplier= 1.5 * (1 + skew)

- This adjustment creates skew-sensitive bounds, refining outlier detection.

Outlier Bounds Calculation:

Lower Bound = Q1 - (Lower Multiplier * IQR)

Upper Bound = Q3 + (Upper Multiplier * IQR)

2. Results and Observations

Effect of Skew Adjustment:

- Features with higher skewness, such as [specific feature with skew], had their multipliers adjusted, resulting in tighter or wider bounds based on the direction of skew. This led to a reduction in false positives and more accurately identified true anomalies.

Outlier Detection Patterns:

- The skew-adjusted IQR method effectively detected genuine anomalies, particularly in high-variance accelerometer features like Acc Z.
- Features with lower skew, such as gyro_x, had minimal adjustment, retaining traditional IQR bounds with effective detection results.

Conclusion

The skew-adjusted IQR method, with a base multiplier of 1.5, effectively balanced sensitivity and specificity in outlier detection. This method reduced false positives, particularly for features with inherent skewness, making it suitable for high-variance, sensor-based anomaly detection tasks.