Outliers in Data Analysis

Definition

- An outlier is a data point that significantly deviates from the majority of the other data points in a dataset.

Types

- Outliers can be classified as either statistical or contextual outliers.

- Statistical outliers are data points that lie far from the mean or median,

- While contextual outliers are observations that defy the expected pattern within a specific context.

Causes of Outliers

- Causes of outliers in data can be attributed to factors such as measurement errors, data entry mistakes, extreme or rare events, or inherent variability in the data.

Detection Techniques

- Various statistical techniques, such as the Z-score method, box plots, or Mahalanobis distance, can help identify outliers.

- However, domain knowledge and context-specific analysis are equally important for accurate identification and interpretation.

Handling Outliers

- Depending on the analysis objectives and domain requirements, outliers can be treated differently.

- Options include removing outliers, transforming data, or applying robust statistical techniques to mitigate their influence.