Handling Outliers – Percentile Method & Winsorization

➢ What is the Percentile Method?

- This method **detects outliers** by identifying values below or above specific **percentile cutoffs**.
- Common thresholds:
  - **Lower bound**: 1st or 5th percentile
  - **Upper bound**: 95th or 99th percentile
- Unlike IQR or Z-score, this method is **distribution-agnostic** and simple to apply.

➢ What is Winsorization?

- **Winsorization** is the technique of **limiting extreme values** by **replacing outliers** with values at a certain percentile.
- Instead of removing data points, you **cap** them to reduce the impact of outliers.

➢ Example Winsorization Process:

If using 5th and 95th percentiles:

- Any value below the 5th percentile is **replaced with the 5th percentile value**.
- Any value above the 95th percentile is **replaced with the 95th percentile value**.

➢ Why Use Winsorization?

- Keeps **all data points**, making it **useful for small datasets**.
- Reduces the effect of outliers **without deletion**.
- Improves **model stability** while preserving dataset size.

➢ Key Takeaway:

The **Percentile Method + Winsorization** is a **gentle and effective way** to handle outliers — especially when you want to **retain data** and reduce the influence of extreme values without distorting the overall distribution.