

Handling Outliers – IQR Method

➤ What is the IQR Method?

The **Interquartile Range (IQR)** method identifies outliers based on the **spread of the middle 50% of the data**.

- **IQR = Q3 – Q1**
 - Q1: 25th percentile (lower quartile)
 - Q3: 75th percentile (upper quartile)
 - **Outlier Criteria:**
 - Any data point **below** $Q1 - 1.5 \times IQR$
 - Or **above** $Q3 + 1.5 \times IQR$
-

➤ Why Use the IQR Method?

- It is **non-parametric** – doesn't assume a normal distribution.
 - **Robust** to skewed data and resistant to the influence of extreme values.
-

➤ Steps to Apply IQR Method:

1. Calculate Q1 and Q3.
 2. Compute $IQR = Q3 - Q1$.
 3. Determine lower and upper bounds:
 - **Lower bound** = $Q1 - 1.5 \times IQR$
 - **Upper bound** = $Q3 + 1.5 \times IQR$
 4. **Filter or cap** values outside this range.
-

➤ Alternate Approach: Capping Outliers

Instead of removing, you can **cap** outliers:

```
df[col] = df[col].clip(lower=lower_bound, upper=upper_bound)
```

➤ Key Takeaway:

The IQR method is a **simple, powerful, and widely used** technique to detect outliers — especially effective when the data is **not normally distributed**.
