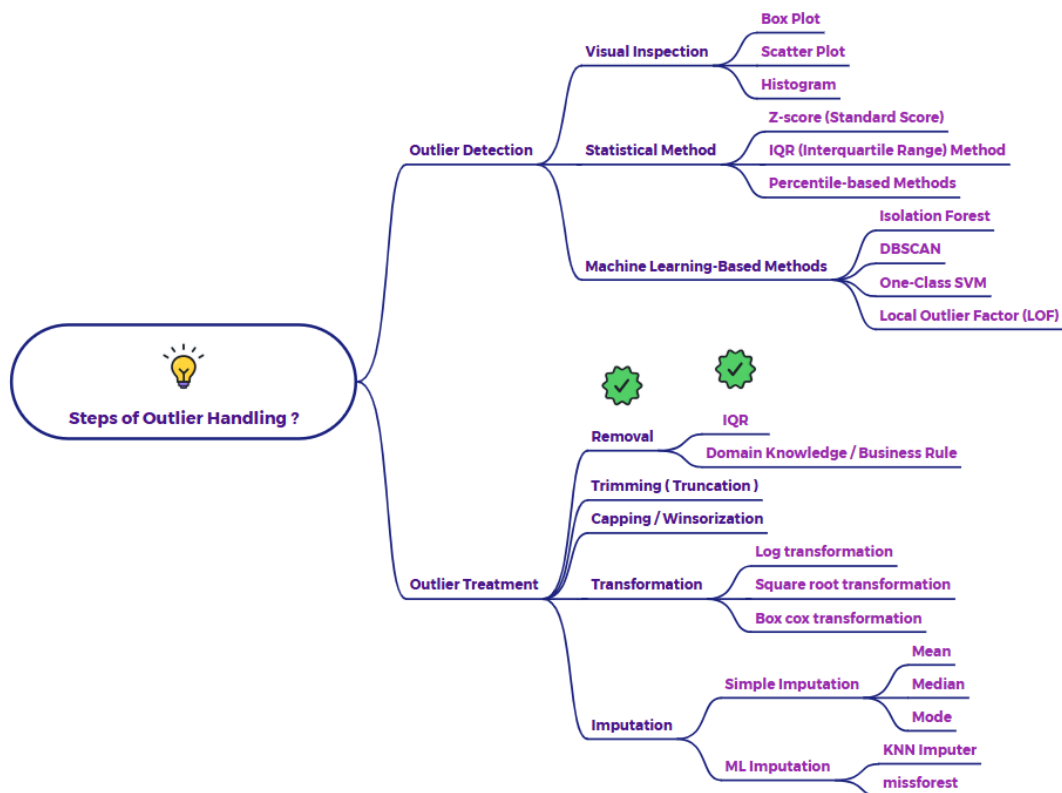


Explain Outlier treatment through removal (IQR) with an example



Outlier Treatment: Removal (IQR Method)

As the name suggests, this method involves identifying outliers using the Interquartile Range (IQR) and then removing them from the dataset.

Steps:

1. **Calculate Q1 and Q3:** Determine the first quartile (Q1, the 25th percentile) and the third quartile (Q3, the 75th percentile) of the data.
2. **Calculate IQR:** Compute the IQR as $IQR = Q3 - Q1$.
3. **Define Outlier Boundaries:**
 - Lower Bound = $Q1 - 1.5 * IQR$
 - Upper Bound = $Q3 + 1.5 * IQR$
4. **Remove Outliers:** Any data points falling below the Lower Bound or above the Upper Bound are considered outliers and are removed from the dataset.

Example:

Let's say we have a dataset of daily sales for a small online store over a month:

[120, 150, 130, 160, 140, 170, 125, 155, 135, 165, 145, 180, 110, 1000, 125, 155, 135, 165, 145, 180, 120, 150, 130, 160, 140, 170, 125, 155, 135, 165]

Here, 1000 is a significant outlier. Let's apply the IQR method to remove it.

1. Calculate Q1 and Q3:

- $Q1 = 130$
- $Q3 = 165$

2. Calculate IQR:

- $IQR = 165 - 130 = 35$

3. Define Outlier Boundaries:

- Lower Bound = $130 - 1.5 * 35 = 77.5$
- Upper Bound = $165 + 1.5 * 35 = 217.5$

4. Remove Outliers: The value 1000 is far above 217.5, so it's removed. The cleaned dataset becomes:

[120, 150, 130, 160, 140, 170, 125, 155, 135, 165, 145, 180, 110, 125, 155, 135, 165, 145, 180, 120, 150, 130, 160, 140, 170, 125, 155, 135, 165]

When to Use Removal (IQR):

- When you are confident that the outliers are errors or are not representative of the population you are studying.
- When you want to clean your data to make it suitable for certain statistical analyses or machine learning models that are sensitive to outliers.

Caution:

- Removing outliers can lead to loss of information, so it should be done judiciously.
- If outliers represent genuine extreme values that are important to your analysis, removing them might not be the best approach.

