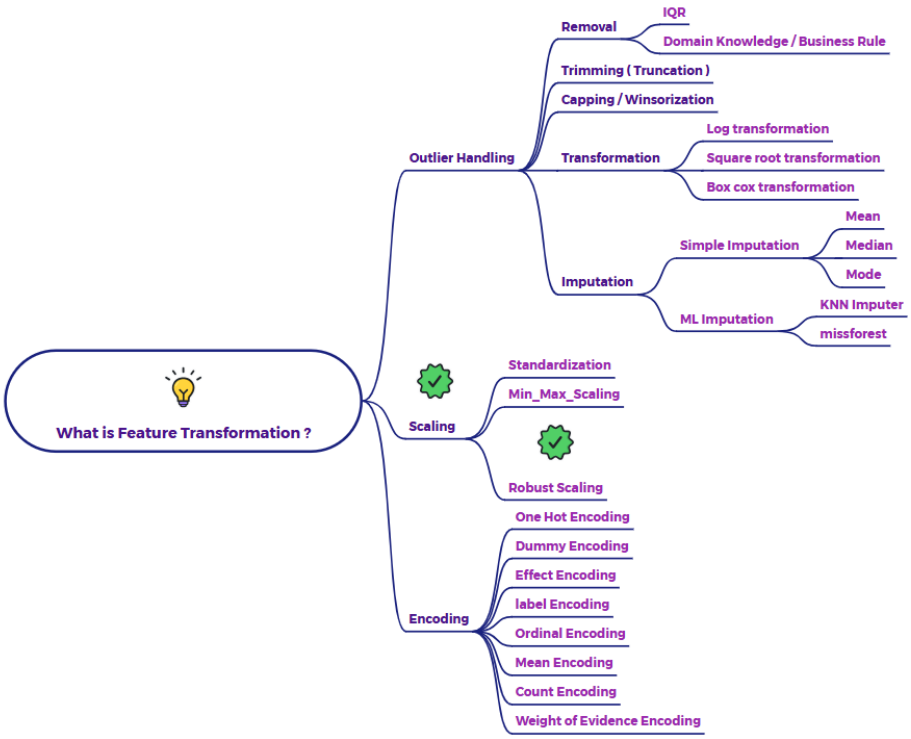
Explain Robust scaling



Robust Scaling

1. **Explanation of Robust Scaling**
    - Robust scaling is a scaling technique that is less sensitive to outliers than Min-Max scaling. It scales the data using the interquartile range (IQR).
2. **How to Calculate Robust Scaling**

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$

The formula for Robust scaling is:

$x' = (x - Q1) / (Q3 - Q1)$

Where:

- (x): Original value
- (Q1): 25th percentile (1st quartile)
- (Q3): 75th percentile (3rd quartile)
- (Q3 - Q1): Interquartile Range (IQR)

Example:

Let's say we have the following data for a variable "Salary" (in thousands of dollars): 40, 50, 60, 70, 80, 90, 100, 110, 120, 300 (Here, 300 is an outlier)

Calculate the first quartile ((Q1)): 50

Calculate the third quartile ((Q3)): 100

Calculate the IQR: (Q3 - Q1) = 100 - 50 = 50

Scale each value:

- For 40: (40 - 50) / 50 = -0.2
- For 50: (50 - 50) / 50 = 0
- For 60: (60 - 50) / 50 = 0.2
- For 70: (70 - 50) / 50 = 0.4
- For 80: (80 - 50) / 50 = 0.6
- For 90: (90 - 50) / 50 = 0.8
- For 100: (100 - 50) / 50 = 1.0
- For 110: (110 - 50) / 50 = 1.2
- For 120: (120 - 50) / 50 = 1.4
- For 300: (300 - 50) / 50 = 5.0

So, the Robust scaled "Salary" values are: -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 5.0

3. **When to Use Robust Scaling**
    - When your data contains outliers.
    - When you want to reduce the effect of outliers on the scaling of your features.
4. **Strengths and Weaknesses of Robust Scaling**
    - **Strengths:**
        - Less sensitive to outliers compared to Min-Max scaling and Standardization.
        - Provides a more stable scaling for data with extreme values.
    - **Weaknesses:**
        - Does not transform data to a specific range like Min-Max scaling.
        - Can compress the majority of the data into a smaller range if outliers are present.