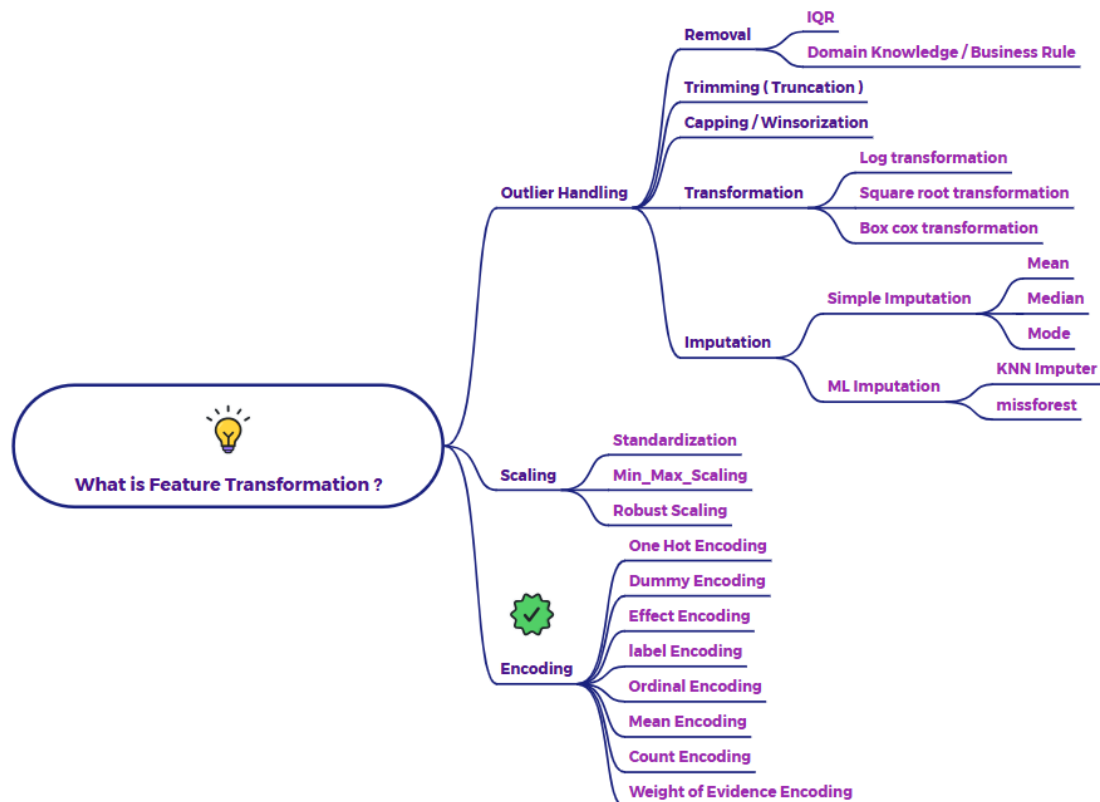


## What is Encoding of Categorical Variable?



## Encoding Categorical Variables

Categorical variables are those that represent qualities or characteristics. Examples include:

- Colors (e.g., Red, Blue, Green)
- Product types (e.g., T-shirt, Shorts)
- City names (e.g., New York, London, Tokyo)

Encoding, in the context of machine learning, is the process of converting these categorical variables into a numerical format. This conversion is necessary because most machine learning algorithms can only process numerical input.

## Why Do We Need Encoding?

Machine learning models, at their core, perform mathematical operations. They work with numbers to find patterns in data. Therefore, to use categorical data in these models, we need to transform it into a numerical representation.

### Here's why encoding is crucial:

- **Compatibility with Algorithms:** Algorithms like linear regression, logistic regression, support vector machines, and neural networks require numerical input. Encoding bridges the gap between categorical data and these algorithms.
- **Meaningful Representation:** Encoding methods aim to represent the relationships between categories in a way that the model can understand. For example, one-hot encoding creates binary variables that indicate the presence or absence of a specific category.
- **Improved Model Performance:** Proper encoding can significantly impact a model's ability to learn from the data. If categorical variables are not encoded correctly, the model may produce inaccurate or meaningless results.