Day-8
Encoding Numerical and Mixed variable

Discretization(Binning):

Discretization is the process of converting continuous numerical features into discrete bins or categories. This can be useful for simplifying data and making it easier for some machine learning models to handle.

Types and Examples:

1. **Equal-Width Binning**:

   Description: Divides the range of the feature into a specified number of equal-width bins.

2. **Equal-Frequency Binning**:

   Description: Divides the data into bins such that each bin contains approximately the same number of data points.

3. **Custom Binning**:

   Description: Defines custom bin edges based on domain knowledge or specific criteria.

4. **Discretization with `KBinsDiscretize(k mean binning)`**:

   Description: Provides a more flexible way to discretize data with various binning strategies.Used when data has clusters.

Binarization:

Binarization is the process of converting numerical values into binary format (0s and 1s). This technique is useful for simplifying features and preparing data for algorithms that work with binary data.

How It Works:

- **Threshold-Based Binarization**: Values are converted to 1 if they are above a specified threshold and to 0 if they are below or equal to the threshold.
- Example :

  Used in image processing.

  [0,255] → colour range and threshold(k)=127.5

  Pixel > k → converted to white

  Pixel < k → converted to black

Handling mixed variable:

Example:

	categorical	Number
A75	**A**	**75**
7	**NAN**	**7**
R	**R**	**NAN**