⇒ When to use :-

- When data contains zero or negative values.
- When you want a Box-Cox-like transformation but without positivity restrictions.
- When your data distribution is heavily skewed in either direction.
- Works for both positive and negative continuous feature.

→ This transformation is somewhat of an adjustment to the Box-Cox transformation, by which we can apply it to negative numbers.

* Binning And Binarization :- (Day-32)

⇒ Encoding Numerical Values :-

"Encoding numerical data means transforming or modifying numerical values to make them more meaningful or usable for machine learning models. It includes techniques that convert raw numbers into structured or scaled forms that improve model performance.

→ Why it is used :-

- To reduce the effect of large numerical ranges.
- To highlight important numerical patterns.
- To prepare numbers for algorithms that can't handle raw numeric scales.
- To convert continuous numbers into categories when needed.

EXPT. NO. | NAME | Page No: | Date: | YOUVA

→ Encoding numerical data ma raw numbers ne modify, categorize, ya scale kari ne models ne vadhu meaningful input apva ma aave che.

→ Etle for example have tamri pase subscriber no column che like 13, 120, 10000, 12454357, ... to ena par model kyarek na sarkhu chale. etle apde ene categorical data ma convert kariye to sari rete chale. to etle apde ene 10+, 100+, 10K+, ... Evi rete convert kariye to saru chale.

→ Types of Encoding:-

(1) Discretization (Binning)
(2) Binarization.

(1) Binning:-

"Binning is the process of transforming continuous variables into discrete variables by creating a set of contiguous intervals that span the range of the variable's value. Binning is also called binning, where bin is an alternative name for interval."

→ Why to use:-

- To handle Outliers.
- To improve the value spread.

Teacher's Signature:

Schach

→ Types of Binning:-

1] Unsupervised Binning:-
   - Equal width / Uniform Binning
   - Equal frequency / Quantile Binning.
   - KMeans Binning.

2] Supervised Binning:-
   - Decision Tree Binning.

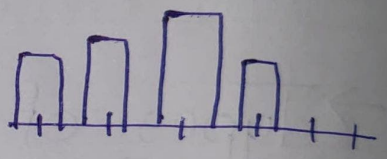
3] Custom Binning.

⇒ Equal Width:-

Age : (27), 32, 84, 56, ... Bins = 10

max = 100 min = 0

$\frac{max - min}{bins} = \frac{100 - 0}{10} = 10.$



(0-10), (10-20), (20-30), ... (90, 100).

Total: 5, 16, 1, 17, 5

→ Equal width ma data range ne same-size na intervals ma divide kari ne values ne bins ma mukva ma aave che.

Pros:-
- Outliers handle thai jai
- Spread of data change na thai.

⇒ Equal Frequency:- (Default) (Mostly Used)

Intervals = 10.

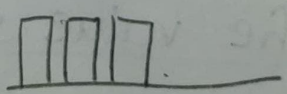
Each interval contains 10% of total observations.

Intervals:-

0-16, 16-20, 20-22, 22-25, ... , 50-74.

10%. 20%. 30%. 40%.

Pros:-
- Outliers handle thai jai.
- Value spread uniform



EXPT. NO.

NAME

M T W T F S S

Page No :

Date:

YOUVA

⇒ K-Means Binning :-

"K-Means binning ma continuous values ne clusters ma group kari ne natural, similarity-based bins banava ma aave che instead of fixed intervals."

Teacher's Signature: