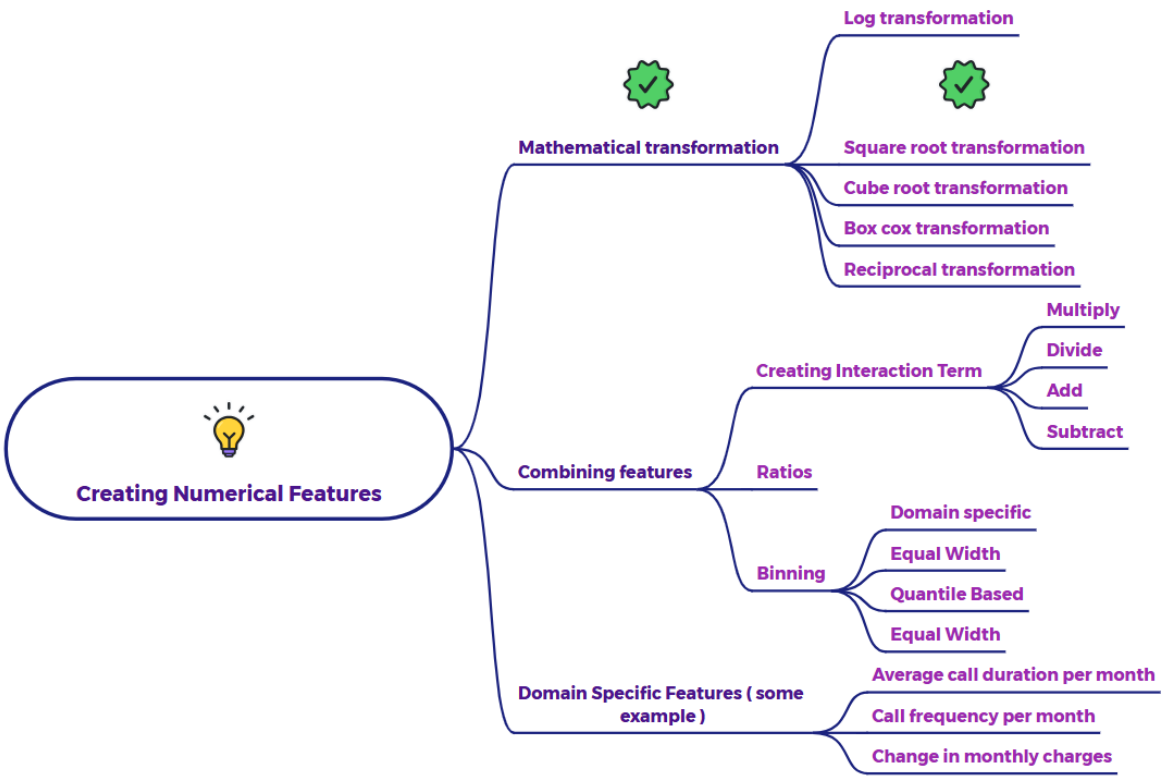
Explain Square root transformation



Square root transformation is a common data transformation technique used to make data more suitable for statistical modeling. It involves applying the square root function to each data point.

Why use it?

- **Reduces Skewness:** Often used for right-skewed data, where large values are more spread out, to make the distribution more symmetrical.
- **Stabilizes Variance:** Can help when the variance of the data is not constant across different levels of a variable (heteroscedasticity).

Example

Let's say we have a dataset of income values (in thousands of dollars):

$20, $25, $30, $35, $40, $45, $50, $60, $70, $80, $100, $150, $200

Here's how to apply the square root transformation to these values:

Square root Transformation
√20 ≈ 4.472
√25 = 5
√30 ≈ 5.477
√35 ≈ 5.916
√40 ≈ 6.325
√45 ≈ 6.708
√50 ≈ 7.071
√60 ≈ 7.746
√70 ≈ 8.367
√80 ≈ 8.944
√100 = 10
√150 ≈ 12.247
√200 ≈ 14.142

If we were to plot the distribution of these square root-transformed values, we would likely see a distribution that is less skewed and more closely resembles a normal distribution compared to the original income data. Check the plot below: