Multicolinearity

Multicollinearity happens when two or more independent variables in a regression model are highly correlated. This makes it difficult to determine which variable is actually affecting the dependent variable because their effects are overlapping.

<u>Ways to Handle Multicolinearity</u>

When dealing with multicollinearity, the goal is to reduce the overlap among variables

1. **Remove One of the Correlated Variables**
    - *If two variables are highly correlated (e.g., correlation coefficient > 0.8), consider dropping one of them. Select the variable with higher predictive power or business relevance.*
2. **Combine Features into a Single Variable**
    - *Use domain knowledge to merge related features into a composite variable. For example, you can combine "length" and "width" into "area" for geometric datasets.*
3. **Apply PCA (Principal Component Analysis)**
    - *PCA transforms correlated variables into a smaller set of uncorrelated components. This technique retains most of the variance while eliminating multicollinearity.*
4. **Use Regularization Techniques**
    - **Ridge Regression**: *Shrinks coefficients by adding a penalty proportional to the square of their magnitude. This reduces the impact of multicollinearity.*
    - **Lasso Regression**: *Performs feature selection by shrinking some coefficients to zero, effectively removing irrelevant features.*