**Interview Question**

**1. What is Normalization & Standardization and how is it helpful?**

**ANS -> Normalization** and **Standardization** are two techniques used to rescale or adjust the values of numerical data, which are often required when you are working with machine learning models or data analysis. Both techniques make the data comparable, improve model performance, and ensure that variables with different units or scales do not disproportionately influence the results.

1. **Normalization**

Normalization refers to rescaling the data to a fixed range, typically **0 to 1** or **-1 to 1**. The goal is to transform the data such that it falls within a specific range, which is helpful when the data has varying scales and units.

Normalization = x – x(min) / x(max)-x(min)

1. **Standardization (Z-Score Scaling)**

Standardization, also known as **Z-score normalization**, transforms data such that it has a **mean of 0** and a **standard deviation of 1**. This method does not bound the values within a specific range, but instead centers the data around the mean and scales according to the standard deviation.

Standardization = X – mean / std. dev

**2.What techniques can be used to address multicollinearity in multiple linear regression?**

Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other. This means that these variables carry overlapping information, making it difficult to determine their individual effects on the dependent variable.

Let suppose we have horse power and Engine size are two independent factor to decide the car price.

But these two factor are quite show relation in between then as the high engine size leads to high horsepower.

So while implementing linear model. The model get overlap in predicting car price.

To address this issue we can use Lasso and ridge regresson model. As the lasso reduced the same metric factors and Ridge turn down the weightage of independent factors.

We can find the correlation between the independent variables and correlation > .8 needs to study.

Then we can remove one of the column/independent variable from dataset.