Multiple linear regression Interview Questions:

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

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

**1. Normalization & Standardization and their Benefits**

**Normalization** and **Standardization** are techniques used to scale numerical features in a dataset. These techniques are crucial for many machine learning algorithms, especially those that rely on distance metrics or gradient descent optimization.

**Normalization:**

* Scales features to a specific range, typically between 0 and 1.
* Useful when the features have different scales but a specific range is desired.
* **Formula:**
* Normalized value = (value - min\_value) / (max\_value - min\_value)

**Standardization:**

* Scales features to have a mean of 0 and a standard deviation of 1.
* Useful when features have different scales and the distribution is not known.
* **Formula:**
* Standardized value = (value - mean) / standard\_deviation

**Benefits of Normalization and Standardization:**

* **Improved Model Performance:** Many algorithms, like gradient descent, converge faster and more reliably when features are on a similar scale.
* **Fair Feature Comparison:** Features with different scales can have a disproportionate impact on the model. Normalization and standardization ensure that all features are treated equally.
* **Better Distance Calculations:** Techniques like K-Nearest Neighbors and clustering algorithms rely on distance metrics. Normalization and standardization can improve the accuracy of these calculations.

**2. Techniques to Address Multicollinearity in Multiple Linear Regression**

Multicollinearity occurs when two or more independent variables in a regression model are highly 1 correlated. This can lead to unstable and unreliable model coefficients. Here are some techniques to address this issue:

**1. Feature Removal:**

* Identify the highly correlated features and remove one of them.
* Use techniques like correlation matrix or Variance Inflation Factor (VIF) to identify the culprits.

**2. Feature Engineering:**

* Combine highly correlated features into a single feature.
* For example, if "height" and "weight" are highly correlated, create a new feature "body\_mass\_index".

**3. Principal Component Analysis (PCA):**

* Reduces the dimensionality of the dataset by creating new uncorrelated features called principal components.
* However, it can be difficult to interpret the resulting model.

**4. Ridge Regression:**

* Adds a penalty term to the loss function, which discourages large coefficient values.
* This can help to reduce the impact of multicollinearity.

**5. Lasso Regression:**

* Similar to Ridge Regression, but it can also perform feature selection by setting some coefficients to zero.
* This can help to identify the most important features and reduce the impact of multicollinearity.

By effectively addressing multicollinearity, you can improve the accuracy, stability, and interpretability of your multiple linear regression models.