**Assignment-based Subjective Questions**

1. **From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?**

Three categorical variables such as ‘seasons’ ‘holiday’ and ‘working day’ can infer their effect on dependent variables

1. **Why is it important to use drop\_first=True during dummy variable creation?**

It helps in reducing the extra column created during dummy variable creation. Hence it reduces the correlations created among dummy variables. It helps in reducing dimensionality

1. **Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?**

By looking at the pair-plot analysis the ‘Temperature’ has the highest correlation with the target variable

1. **How did you validate the assumptions of Linear Regression after building the model on the training set?**

After building the model on training set we should import

1. Linear relationship between X and Y
2. Error terms- normally distrubted with mean 0
3. Error terms are independent to each other
4. Error terms-Constant variance (homoscedasticity)

Steps

1. train and split from sklearn “**from** sklearn.model\_selection **import** train\_test\_split”
2. Train the model by using random forest by model.fit
3. Then plot the graph for Y test and Y Predicited values
4. From sklearn metrics import mean.square error to check the error value
5. **Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?**

Based on the final model, the top 3 features contributing significantly towards explaining the demand of the shared bikes are

1. Year
2. Holiday
3. Temperature