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?

A) It could be seen that on both the year same number of bikes are shared

   Most people rented the bike during non holidays stating it is used for commutation to their work.

2) Why is it important to use drop_first=True during dummy variable creation?

A) It is used to avoid the residual column created during the dummy variable creation

3) Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

A) The temperature has the highest correlation with the target variable

4) How did you validate the assumptions of Linear Regression after building the model on the training set?

A) From the displot plotted we could see that the error is normally distributed with zero mean

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

A) 'yr', 'weathersit', 'atemp' are the top 3 variables contributing towards the demand of the shared bikes

General Subjective Questions

1. Explain the linear regression algorithm in detail?

A) Linear regression is a simple statistical regression method used for predictive analysis and shows the relationship between the continuous variables.

2. Explain the Anscombe’s quartet in detail.?

A) Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed.

3. What is Pearson's R?

A) It is a measure of linear correlation between two sets of data.

4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

A) Scaling is referred to an operation on converting the scale of various variables to one standard scale.

It is required in model building process so that no one variable value is assumed greater than other by the model

A normalized dataset will always have values that range between 0 and 1. A standardized dataset will have a mean of 0 and standard deviation of 1, but there is no specific upper or lower bound for the maximum and minimum values

5 .You might have observed that sometimes the value of VIF is infinite. Why does this happen?

A) An infinite VIF value indicates that the corresponding variable may be expressed exactly by a linear combination of other variables

6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

A) This plot is used to plot the quantiles of a sample distribution against quantiles of a theoretical distribution.