

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

Categorical features like season, month, weathersit have a good amount of impact on the model which is proved through their coefficient factors and p-values.

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

It is important to use drop_First = True because ifor any dummy variable creation with n levels only n-1 features are required.

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

Column 'registered' has the highest correlation with the target variable 'cnt'

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

Validated the above by doing residual analysis on error terms which should be a normal distribution with mean zero according to the assumption and this was achieved in the final analysis.

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

Year , temperature, winter season.