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?

I see that demand for next year has grown, Demand is continuously growing every month till June. September is highest demanding month.

After September, demand is decreased, clealy seen that weathershit has highest demand, fall has highest demand , Demand was decreased on holidays ,

More on Demand for bikes is positively correlated to temp as temp increased demand also increases and is opposite for humidity ,demand decreased as humidity increases .

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

drop_first = True, is important to use, as it helps us in reducing the extra column created during creating the dummy variables. Hence it reduces the correlations created among dummy variables

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

Temp and atemp has highest correlation with target variable (cnt)

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

I can make assumptions on the bases of, Linear Relationship between the features and target, no Multicollinearity between the features, Homoscedasticity Assumption, Normal distribution of error terms.

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

In the month of September Company can expand their business , Also in Spring season they should do same ,like to expand their growth , They get low demand in Rainy and Show seasons because of rain fall and show fall people like to stay at home .