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?

They play a huge role In effecting the dependent variable, it should be ensured that we do hot encoding for categorial values . pd.get_dummies(0 can be used to achieve this

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

As indicated in adjusted r square formula, it is better for us to reduce the number of features as we can . hence we should drop redundant columns when we get a chance

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

Registered

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

Found linear relationship with one or more variable

Found normal distribution of residyals

Error terms are independent of each other

Have constant variance

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

Holiday

Bad weather (extracted from weathersit)

summer