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?

Ans: From the categorical variables like season, for summer and fall when temperatures and atemp are high, more users opt for bike sharing. So categorical variables does have impact on dependent variables.

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

Ans: Drop\_first allows n-1 variables to be and delete extra column while creating dummy variable.

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

Ans: temp & atemp has higher correlation with cnt variable.

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

Ans: After model development, we did residual analysis and did a prediction to see if it linearly shows the relationship between variables.

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

Ans: Temperature, Season, Month

General Subjective Questions

1. Explain the linear regression algorithm in detail?

Ans: Linear regression is type of machine learning model that is supervised in nature and works with labelled datasets. It works in classification and regression. This algorithm predicts based on linear relationships between dependent variables and one or more independent variables. As part of testing we do best fit line along with residual analysis for model testing and validation.

1. Explain the Anscombe’s quartet in detail.

Ans: Anscombe’s quartet comprises of four datasets that have identical summary statistics. It will have same means, variance, R-squared, correlations and linear regression lines. It helps in knowing the purpose of exploratory data analysis.

1. What is Pearson’s R?

Ans: Pearson’s R is Pearson Correlation Coefficient. It is most common way of measuring linear correlation.

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

Ans: Scaling helps in changing the range of data and it helps with keep all values compatible. We do it so that all variables have compatible values to compare. Normalized scaling is min max scaling which shows graph in bell curve. Standardized scaling is another scaling method centered around mean with standard deviation.

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

Ans: This happens when there is perfect correlation between independent variables in regression model.

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

Ans: Q-Q Plots is used to compare dataset to a theoretical model. This helps in linear regression when we have training and test data set received separately.