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. Season and year has good impact on dependent variable and other variables has low impact on it.
2. Why is it important to use drop_first=True during dummy variable creation?
   Ans. If we doesn't apply drop_first=True then in dataframe it contain redundant variable unnecessary.
3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?
   Ans. Registered has high correlation apart from them casual has good impact and then temp and atemp.
4. How did you validate the assumptions of Linear Regression after building the model on the training set?
   Ans.
   A. Relationship between X and Y should be linear
   B. Error should be normally distributed
   C. Error should not be depend on each other.
   D. Homodescity.
5. Based on the final model which are the top 3 features contributing significantly towards explaining the demand of the share bikes?
   Ans. Registered,casual,temp.

General Subjective Question

1. Explain the linear regression algorithm in details.
   Ans. Linear regression is all about to modelling the relationship b/w a dependent and oneor more independent variable by fitting a straight line equation of data.
2. Explain the Anscombe's quartet in detail.
   Ans. It's a set of small four small dataset.
3. What is Pearson's R?
   Ans. It is a statistical measurement to check the relationship between 2 variables. Range -1 to 1
4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
   Ans. Scaling is a technique to transform a data in same range of data. Scaling is important when data is not distributed normally.
   Normalized scaling :- data range between 0 to 1 .
   Standard scaling :- Its done by the help of mean and SD (where mean=0 and SD=1)
5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
   Ans. When data contains duplicate variables.
6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
   It represent the distribution of data.