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?
   The effect on dependent variable is surprising in many events,
   1. such as humidity has less role to play in actual demand.
   2. Casual users are more in number so can be potential buyers by looking in to data about who are quite frequent travellers. (3 marks)
2. Why is it important to use **drop first=True** during dummy variable creation? (2 mark)
3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable? Registered users has highest co relation
   (1 mark)
4. How did you validate the assumptions of Linear Regression after building the model on the training set? By checking on actual data and the r squared value
   (3 marks)
5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes? (2 marks)

General Subjective Questions

1. Explain the linear regression algorithm in detail.

Answer: Linear regression states the relation between one or more variables in statistics. There are two types of variables dependent and independent. The value you want to predict is dependent where- as value you are using for predicting is independent.
(4 marks)

2. Explain the Anscombe's quartet in detail.
   Answer: Group of data sets which have same mean, standard deviation, and regression line but they are different qualitatively. That is why it's important to look data graphically as well.
   (3 marks)
3. What is Pearson's R?
   Answer: It is the strength of the curve of two variables. It varies between -1 and 1. -1 being negative and 1 is opposite of that. (3 marks)
   4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
   Answer: It is a step of data Pre-Processing which is applied to independent variables to normalize the data within a particular range. It also helps in speeding up the calculations in an algorithm (3 marks)

5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
(3 marks) 6.

What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
(3 marks)