Assignment-based Subjective Questions

Question 1. From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?
Total Marks: 3 marks
Answer: The bike rental rates are more in the months of September & October, specially on Sat, Wed & Thurs in 2019.

Question 2. Why is it important to use **drop_first=True** during dummy variable creation?
Total Marks: 2 marks
Answer: It helps to reduce the additional column created during the dummy variables to avoid redundancy.

Question 3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?
Total Marks: 1 mark
Answer: temp.

Question 4. How did you validate the assumptions of Linear Regression after building the model on the training set?
Total Marks: 3 marks
Answer: Will check VIF, linear relationship between dependent variable & a feature variable.

Question 5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?
Total Marks: 2 marks
Answer: temp, yr, holiday

General Subjective Questions

Question 6. Explain the linear regression algorithm in detail.
Total Marks:
Answer: It is used to model the relationship between a target variable (in this case study Cnt is target variable), and one or more independent variables (temp, yr, holiday). It is mainly used to find a line that best fits the data, minimizing the error between predicted and actual values.
There are two types of linear regression viz,
1) Simple Linear Regression (One independent and one dependent variable)
2) Multiple Linear Regression (Two or more independent variables).

To execute linear regression algorithm there are some assumptions are as follows:-
1) The relationship between the dependent & independent variable is linear.
2) Observations are independent of each other.

3) The variance of residuals is constant across all levels of the independent variables which is called Homoscedasticity.
4) Residuals are normally distributed.
5) Independent variables are not highly correlated.

A linear line showing the relationship between the variables is called a regression line. A positive linear relationship is when the dependent variable on Y axis along with the independent variable in the X- axis. However, if dependent variable value decreases with increase in independent variable value increases in X-axis, it is called a negative linear relationship.

Question 7. Explain the Anscombe's quartet in detail. (Do not edit)
Total Marks: 3 marks (Do not edit)
Answer: Please write your answer below this line. (Do not edit)

<Your answer for Question 7 goes here>

Question 8. What is Pearson's R? (Do not edit)
Total Marks: 3 marks (Do not edit)
Answer: Please write your answer below this line. (Do not edit)

<Your answer for Question 8 goes here>

Question 9. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
Total Marks: 3 marks
Answer: Scaling is the process of transforming the features of data so that they all fall within a similar range of distribution. In ML, variables in ds often have different units & ranges, which can affect the performance and interpretation of MLA. Scaling ensures that each feature contributes equally to the model, preventing bias due to difference in the magnitude of the features. Scaling is performed to improve the model performance.

Question 10. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
Total Marks: 3 marks
Answer: The value of VIF is infinite, when there is a perfect correlation between the two independent variables. The R-Squared value is 1 in this case. This leads to VIF infinity as VIF = 1/(1-R2). This suggest that there is a problem of multicollinearity & one of these variables need to be dropped in order to define a working model for regression.

Question 11. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression. (Do not edit)

Total Marks: 3 marks (Do not edit)

Answer: Please write your answer below this line. (Do not edit)

<Your answer for Question 11 goes here>
