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? (Do not edit)

Total Marks: 3 marks (Do not edit)

Answer: <I have used the cnt as dependent variable here. Usage of rental bikes are high when its NOT **holiday** in the year **2019** year with **clear weather** per my bivariant categorical analysis
> (Do not edit)

Question 2. Why is it important to use **drop_first=True** during dummy variable creation? (Do not edit)

Total Marks: 2 marks (Do not edit)

Answer: <To avoid redundancy ,improve model efficiency and for accuracy.I used for season which has 4 types (n-1) which is 3 columns which internally makes the combinations of
000 -spring,010-fall,001-winter,100-summer and finally removed the first one spring and also redundant season coumn> (Do not edit)

Question 3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable? (Do not edit)

Total Marks: 1 mark (Do not edit)

Answer: <temp has the better linear nature or high correlation with target variable cnt per my pair plot > (Do not edit)

Question 4. How did you validate the assumptions of Linear Regression after building the model on the training set? (Do not edit)

Total Marks: 3 marks (Do not edit)

Answer: < **Rental count = 0.2341*yr - 0.0872 * Holiday + 0.4662*temp - 0.1546 * windspeed - 0.0750 * windspeed -0.2785* Rain/Snow + 0.1209* Summer + 0.0827* fall + 0.1572* winter**
#Overall An R-squared of 0.822 is generally considered a good value, suggesting that the model is a good fit for the data and has a strong predictive power.
And from the graph its adescent model between train and test predictions
> (Do not edit)

Question 5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes? (Do not edit)

Total Marks: 2 marks (Do not edit)

Answer: <The demand of renatal bikes in the year **2019** is significantly high compared to 2018 and **temperature** has significant factor during **warm weathers**, these three terms helped model building steps> (Do not edit)

General Subjective Questions

Question 6. Explain the linear regression algorithm in detail. (Do not edit)
Total Marks: 4 marks (Do not edit)
Answer: Please write your answer below this line. (Do not edit)

<Your answer for Question 6 goes here>

Linear regression is a MAchine learning regression technique
It comprises of Simple linear regressiona nd multiple linear regression
It comprises of data cleaning,EDA,calculations like Residuals,R-squared formulae,RFE,VIF variation inflation factor
Using above methods we can select the features,train the data with 80:20 or 70:30 train and test thee linear regression model and predict the efficiency,accuracy usecases.

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>
I don't know this concept in linear regression

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>
It's a factor to asses good coefficient between two continuous variables
It supports normally distributed data ranges between -1 to 1
0.6-0.8 is treated as a strong correlation factor indicated with 'r'

Question 9. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling? (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 9 goes here>
For numerical values to predict good analysis we make those values between 0 and 1 to normalize or standardize data.
And the formula to be used behind it is
#normalization : (x-xmin)/(xmax-xmin)
#standardization(x-mu)/sigma
This helps in training the model to have the good residuals,vif etc..,

Question 10. You might have observed that sometimes the value of VIF is infinite. Why does this happen? (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 10 goes here>
Yes I have observed the values infinite while preparing the VIF
Later after modifying the season and weathersit columns to dummy then I got the reasonable values w.r.t VIF or P-value to fit the model.

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>
It's a quantile-quantiele plot used in visual checks.
It has the caliber to divide tinto eual quantiles
It compares the theoeretical distribution with quantiles of the datset
It used to draw the insights of the normal distribution
Outlier detection
Can decide or predeict residuals and to improve the fit of the model using QQplot
There are certain