# 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: <Your answer for Question 1 goes below this line> (Do not edit)

# Some of the categorical variable have significant impact on the dependent variable while others don’t. such variable are,

# Season: Summer and Fall have positive impact on bike rental count. Spring has the least amount of bike rental bookings.

# Month: Similar trends were observed in months as well, warmer months (April to October) have higher rental count.

# Weather: There is a positive correlation between how good the weather is and bike rental. Days with rain have least number of rentals.

# Holiday: Though this does not have major significance but there is a slight drop in number of rentals on holiday which is an interesting insight.

# 

**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: <Your answer for Question 2 goes below this line> (Do not edit)

# Setting drop\_first to true drops the first dummy column. As we only need n-1 columns for a given categorical field which supports n different types of values.

# For example, for a categorical column supporting three categorical values (A, B, C), drop\_first = True will only create two columns, B and C and drop A.

# Interpretation:

# B = 0, C = 0 => A

# B = 1, C = 0 => B

# B = 0, C = 1 => C

# The default value for drop\_first is False.

**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: <Your answer for Question 3 goes below this line> (Do not edit)

# By looking at the pair-plot both temp and atemp seems to have highest correlation with the target variable.

**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: <Your answer for Question 4 goes below this line> (Do not edit)

The assumption of linear regression model has been validated be performing following steps.

* Validation of Linear relationship between X and Y
* Normal distribution of error terms: Using residual analysis
* Independence of Error terms
* Constant Variance (homoscedasticity)

**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: <Your answer for Question 5 goes below this line> (Do not edit)

# Temp (coff: 0.449)

# Light\_Rain (coff: -0.285)

# Yr (coff: 0.234)

# 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>

**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? (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>

**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>

**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>