# 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)

# Categorical variables had medium VIF and low p-value so there was no much problem with multicolinearity

# 

**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)

# Dummy variables should be m-1 hence first varibale can be removed as when we have m variables

**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)

Year has highest correlation

**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)

Validated by using r2 error on test dataset.

**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)

# Windspeed ,year, season-spring

# 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>

# It depends on statistical values between dependent and independent variable , if independent variables changes the value , related dependent variable also changes the value. It has simple linear regression where one dependent variable is corelated with another independent variable . Multiple linear regression where one dependent variable corelates with multiple independent variables

**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>

It is a set of four datasets 1.Linear Relationship 2.Non Linear relationship 3.Outliers 4.Vertical line with one outlier**.** Anscombe’s quartet helps in visualization. Statistics can be misleading so always have graphical analysis along with it is the conclusion from Anscombe’s quartet

**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>

Linear correlation between two values I.e -1 to 1 , where -1 indicates one variable increases and other variable decreases. 1 indicates one variable increases and other one as well increases.

**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>

Scaling is transformation. Scaling is performed to get values within one spcified range which makes it easier to predict corrrect values in model. Standard scaling is Xscale=X-mue/sigma where mean is 0 and standdard deviation is 1. Normalized scaling is getting the value in range of 0 to 1

**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>

# VIF can go to infinite value when one predictor variable is perfectly corelated with another predictor variable

**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 is a graphical tool to access the data following the particular theoretical data I.e normal distribution. Interpretation of q-q plot straight line-points fall on the line so values likely to follow theoretical values. Curve line if points do not fall in straight it means values likely not to follow theoretical values