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? (3 marks)

They change values of r significantly after encoding them, They are major factors to be taken in consideration while making prediction.

2. Why is it important to use drop_first=True during dummy variable creation? (2 mark)

Because an additional factor is not required as it already there when no case applies also it creates a dummy variable trap which leads to multi collinearity

3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable? (1 mark)

Atemp - Feels temp was the one

4. How did you validate the assumptions of Linear Regression after building the model on the training set? (3 marks)

Residual Analysis

- Error value sould be normal distibuted
- Error value should not have a pattern

5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes? (2 marks)

- Atemp
- Season
- Weather

General Subjective Questions

1. Explain the linear regression algorithm in detail. (4 marks)

Linear regression is a ML technique used to make prediction on a continous scale using a straight curve in our distibution.

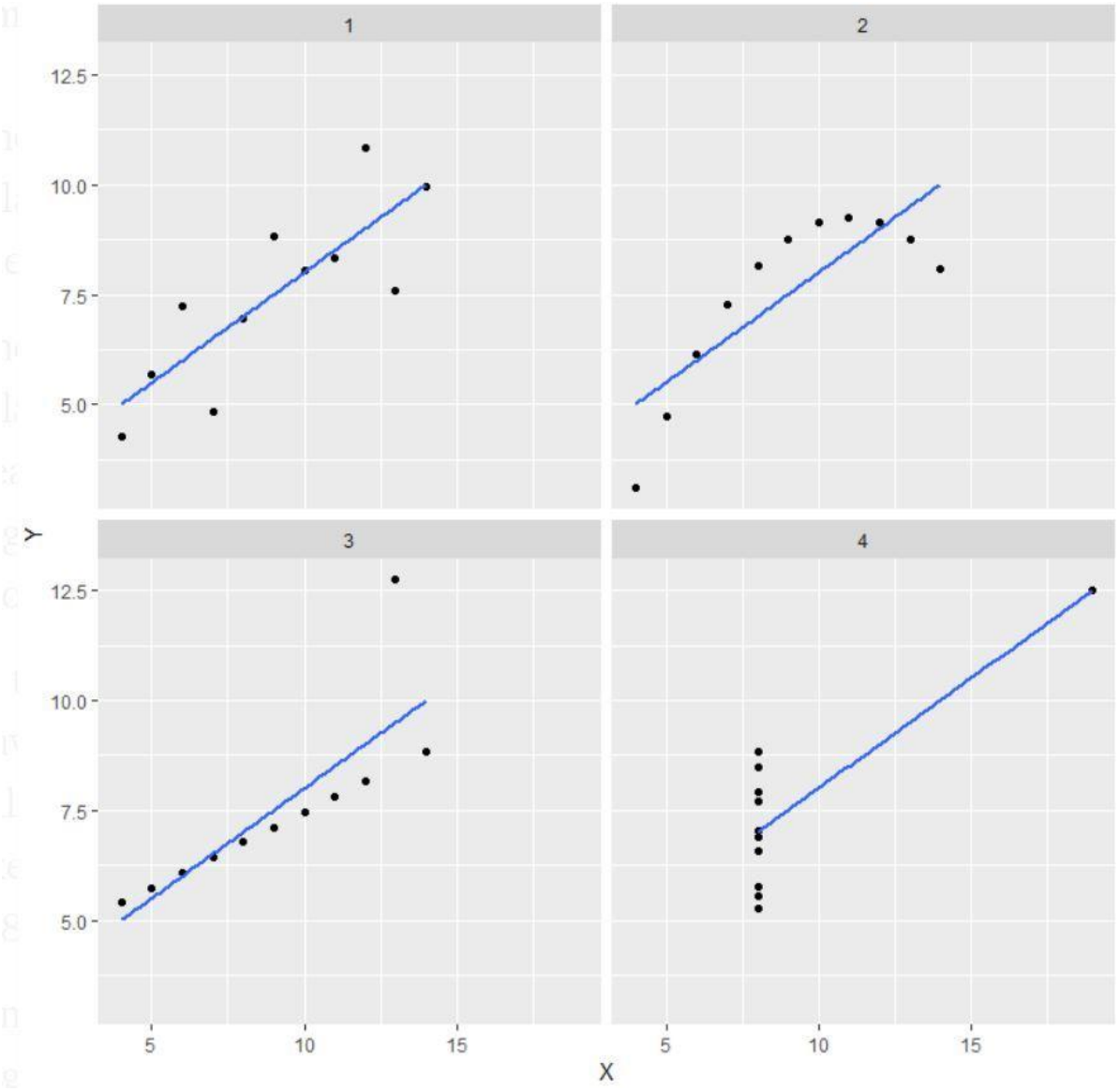
y = M x + c

M = Slope c=constant

Multiple model : y = mx1+c + mx2+c +...+mxn+c

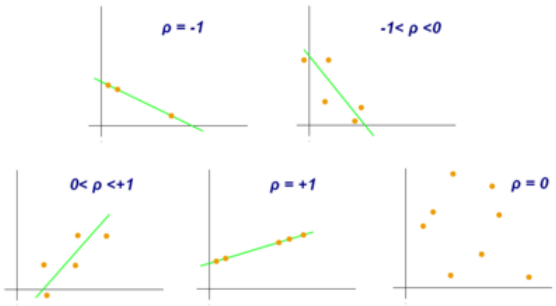
2. Explain the Anscombe's quartet in detail. (3 marks)

It consist of 4 division of number which graphical picture gives a different picture than the numbers values



3. What is Pearson's R? (3 marks)

It is the measure of linear correlation between 2 sets of data.



4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling? (3 marks)

In order to get all values on one scale we do we can get comparable

Normalized - divide it in 0 and 1

standardized scaling – values are centered around mean with unit standard deviation

5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?

When correlation is perfect we found value as inf, it happens as they are orthogonal towards each other in this case

6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

The q-q plot is the plot of the quantiles of the first data set against the quantiles of the second data set, quantile means the fraction of points below the given values with this we can check with it if two are in same distribution or not.