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?
    a. Based on the analysis we saw that the temp variable is very much associated with the count variable.
    b. Summer and fall season are good for bike business
    c. Same can be seen in months time as well.
    d. Saturday(s) has more broad spectrum but overall mean is same.
    e. It is more on holidays
    f. It is increasing yoy
    g. It is affected by weather.
2. Why is it important to use drop_first=True during dummy variable creation?
    a. So that while creating dummy column for categorical column instead of n , we will only keep n-1. Where n is distinct values in column.
3. Looking at the pair plot among the numerical variables, which one has the highest correlation with the target variable?
    a. Temp and atemp variable as highest variable with the target variable.
4. How did you validate the assumptions of Linear Regression after building the model on the training set?
    a. Based on adjusted R2, P values, VIF (variation inflation factor) and Validating it on test data.
5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?
    a. Temp
    b. Weather
    c. year

General Subjective Questions:

1. Explain the linear regression algorithm in detail.
    a. Linear regression is a way to calculate the resultant variable based on the equation (y=m(i)x(i)+c). and where I is from 1 to N. N is equal to the number of the features we have used.
2. Explain Anscombe's quartet in detail.
    a. Anscombe's quartet talks about the importance of plotting the graph or visualization of the data. As the mean and standard deviation could be sometimes deceiving.
3. What is Pearson's R?
    a. It is a way for measuring the relationship between 2 variables. It varies b/w 0 to 1.
4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
    a. Scaling is done to reduce the coefficient values for the bring into normal range. There are multiple kind of scaling done.
        i. MinMixScaler
        ii. StandardScaler

iii. Robust Scaler

5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
    a. When two features are 100% matched to each other. Then it would be infinite.
6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
    a. It is a graph representation of 2 probability distribution against each other for comparison.