**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?
    - By considering the 'season' column the 'fall' season have maximum bikes rented.
    - In the season 'spring' lower number of bikes rented.
    - By considering weather condition larger number of bikes rented in 'clear' condition.
    - And lower number of bikes rented in the season 'Light Snow'
2. Why is it important to use drop_first=True during dummy variable creation?
    - While we are creating dummy variables, if there are n variable then with the help of n-1 variable we can able to represent all the data. So we prefer to drop_first column
3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?
    - The 'temperature' variable have the largest correlation with the target variable(0.63).
4. How did you validate the assumptions of Linear Regression after building the model on the training set?
    - With the help of residual analysis, VIF values
    - With the help statsmodel OLS(Ordinary least squares) summary
5. Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?
    - 'temp', 'hum' and 'windspeed' are the 3 features contributing significantly towards explaining the demand of the shared bikes.

**General Subjective Questions**

1. Explain the linear regression algorithm in detail.
    - In Machine learning there are two types of learning techniques supervised learning and non supervised learning.
    - Linear regression is a supervised ML algorithm.
    - In LR model there are two types of variables dependent variable and independent variable.
    - With the help of LR model we can predict the dependent variable by using independent variable.
    - The type of relationship between dependent and independent variableis linear in nature.
2. Explain the Anscombe's quartet in detail.
    - Anscombe's quartet  consist of four data sets have almost same statistical properties.
    - But when we pictorize the data in graph it shows four different characteristic.
    - So this suggest that eventhogh the statistical features shows the same mean, median etc.. we have to plot the graph to understand the data clearly.

3. What is Pearson's R?
    - Pearson's R is also known as Pearson's correlation coefficient
    - It measures the linear correlation between two variables.
    - Its correlation value lies between -1 to 1. -1 means high negative correlation and +1 means high positive correlation
4. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
    - The data which we get for analysis consist of different range of datas and with different measurement units.
    - So we have to bound this in to a particular range.
    - With the help scaling we can able to achieve this.
    - Mainly there are two types of scaling techniques Normalized/Min-Max Scaling and Standardised scaling.
    - In min-max scaling we populate all the datas in the range 0-1.
    - In Standardised scaling all the values the data frame replaces with its z-scores.
5. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
    - The equation to find VIF is  VIF = 1/(1-R2)
    - So VIF is infinity means R2 =1
    - Which means there is a perfect correlation.
6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
    - Q-Q plot means Quantile-Quantile plot.
    - It helps us to find weather two population data sets are with a common distribution
    - If the two sample distributions are linearly related, then the points will be some what in the same line.