**Assignment-based Subjective Questions**

1) Categorical variables like weekday (Saturday), weathersit(clear), month (September) has positively increase count values whereas few Categorical variables like holiday, season (spring) had reduced the count value.
2) If there are N values for a Categorical Variable, then those N values can be represented using N-1 dummy variables. If all those N-1 dummy variables are equal to Zero, then it will calculate the remaining $N^{th}$ value of the Categorical Variable.
3) I observed both temp and atemp have the highest correlation with target variable count. This happened because both temp and atemp are highly correlated.
4) I validated the model for Linearity, Independence, Homoscedasticity, Normality, no multicollinearity, no endogeneity using R2, VIF, standard scaler, and correlation values.
5) The top three features that are defining the demand for shared bikes are: atemp, weathersit_clear, mnth_sept

**General Subjective Questions**

1) A Linear Regression model assumes that dependent variables have a linear relationship with independent variables. By defining cost function, linear regression model calculates the parameters of independent variables that reduces the overall error by minimizing the cost function. R2 value is generally used to know if a model is good predictor of dependent variable.

2) Anscombe's quartet: Sometimes the Statistics summary (Mean, variance, R2, Correlation) of different data set might be same but the distribution of the dataset will follow different patterns. That is why it is essential to plot the data and see if the relationship holds despite different datasets having similar statistic summary.

3) Pearson R: It is used to define correlations between two entities. It ranges between [-1,1]. Value less than Zero means negatively Correlated and value greater than 0 is positively correlated. If magnitude is greater than 0.5 then we can say two values are strongly correlated.

4) Scaling is a way of transforming a value from one dimension scale to another dimension scale using mathematical formula.
   Normalized scaling converts all values in the range of [0,1] by using min-max Scaling.
   xn=(x-xmin)/(xmax-xmin)
   Standardized Scaling converts all values by using mean and standard deviation.
   xs=(x-sigma)/ (standard deviation)

5) I observed VIF of few variables will be infinite if there is strong correlation between any one of the variables. Example if speed in miles per her and speed in km per hour are used in the same dataset, then the VIF of both variables will become infinity as they are strongly correlated with mathematical formula.

6) Q-Q Plots will help us to access if two data sets are following the same theoretical distribution such as Normal, exponential or Uniform distribution using quantiles. We can use Q-Q plot to verify if our regression model is correct by plotting actual vs predicted and or by selecting input samples.