Assignment-based Subjective Questions:

1.) All of the categorical variables are equally significant in terms of looking at the p-value and in terms of VIF (Variance inflation factor) the 'holiday' column makes it the most reliable and the other categorical columns are also well below the level 5.
2.) Suppose if a column has 3 unique values, with 2 dummy variables we can actually identify the third variable as n variables require n-1 dummy variables. Hence, we use the drop_first=true in order to drop the first dummy variable.
3.) temp and atemp are almost equally correlated.
4.) The validation of the assumption that the error terms are normally distributed has been proved by taking the difference of the actual point subtracted from the predicted point and the difference is plotted as a histogram.
5.) Holiday, yr and the windspeed contributed more significantly in terms of the demand of the shared bikes.

General Subjective Questions:

1.) Linear Regression algorithm is used to explain the relationship between two variables. The simple linear regression is given by the equation y=mx+c where m is the coefficient(slope of the line) and the c is the intercept. The best fit line is given by minimizing the error between the predicted and the true point. Simple Linear Regression is used when a variable is predicted with only one variable and it is termed as multiple linear regression in case of multiple independent columns
2.) In 1973 , Francis Anscombe has created for datasets with x,y fields containing 11 points. They all seem to be similar in terms of numbers and also the best fit line ends up to be the same but when visualized they appear quite different and this proves the importance of visualization as the numbers can be deceptive
3.) The Pearsons correlation coefficient measures the linear correlation between two variables. It is given by dividing the covariance of the two variables by the product of their standard deviations. It is not applicable for non-linear relationships
4.) Scaling means bring all the numerical values into a common range of values. It is done to maintain a standard range of values across all the columns as it may effect the coefficients. Normal scaling brings the values between 0 and 1 where as the standard scaling makes the mean of the data to be 0 and the standard deviation will be 1
5.) The VIF is infinite if the variables are perfectly correlated i.e has high correlation. The presence of multicollinearity contributes to high VIF
6.) QQ plots are quantile-quantile plots i.e plots of two quantiles against each other. The main goal of a qq plot is to determine if two sample sets of data is coming from a common distribution.