Conclusions based on analysis:

1) The following points were made from the dependence of target variable on categorical variables:
    i) The analysis has shown that, during summer and fall season, the rental count is more compared to spring and winter.
    ii) The year 2019 has gathered more rentals than the previous year.
    iii) On holidays, the average rental is less compared to non-holiday.
    iv) Almost every weekday has same average rental count.
    v) September and October months has seen better rentals compared to others.
2) When we create dummy variables for a categorical variable, we use n-1 categories.
   So we provide “drop_first = True” parameter to the function, which ignores one category and creates n-1 dummy variables, which leads to less correlation between predictors.
3) “temp” and “atemp” columns have decent correlation with the target variable but so does the “casual” and “registered” columns, but they are to be ignored as they build the target variable itself.
4) After building the model on training set, we can see through the summary that the r-squared value of the model is quite decent (around 0.82), and also, we see the p-values of the predictors which were less that 0.05, hence we validate the assumptions of the Linear Regression.
   Also, the residuals having a normal form, is also a factor to validate the assumptions.
5) Year, temperature and winter columns were the better features contributing towards the demand.

General Subjective Question's:

1. Linear Regression algorithm: It is based on supervised learning, meaning data with labels. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. It performs a task to predict a dependent variable based on a given set of independent variables. The regression technique finds out a linear relation between predictors and the target, and so is called Linear Regression.

2. Anscombe's Quartet can be defined as a group of four data sets which are nearly identical in simple descriptive statistics, but there are some peculiarities in the dataset that fools the regression model if built. They have very different distributions and appear differently when plotted on scatter plots.

3. Correlation coefficients are used to measure how strong a relationship is between two variables. There are several types of correlation coefficient, but the most popular is Pearson's. Pearson's correlation, also called Pearson's R, is a correlation coefficient commonly used in linear regression. Correlation coefficient formulas are used to find how strong a relationship is between data. The formulas return a value between -1 and 1, where: 1 indicates a strong positive relationship. -1 indicates a strong negative relationship. A result of zero indicates no relationship at all.

4. Scaling is a step of data-preprocessing which is applied to independent variables to normalize the data within a particular range. It also helps in speeding up the calculations in an algorithm. Over time, the collected dataset contains features highly varying in magnitudes, units and range. If scaling is not done then algorithm only takes magnitude in account and not units hence incorrect modelling. To solve this issue, we have to do scaling to bring all the variables to the same level of magnitude.
   Normalized Scaling:
   - It brings all the data in the range of 0 to 1.
   - MinMaxScaler from sklearn helps best to normalize the variables.
   - Formula: $x = (x - \min(x)) / (\max(x) - \min(x))$

   Standardized Scaling:
   - Standardization replaces the values by their Z-score. It brings all of the data values into a standard normal distribution which has a mean of 0 and standard deviation of 1.
   - Formula: $x = (x - \text{mean}(x)) / \text{sd}(x)$

   Normalization has a disadvantage that it loses some information in the data like about the outliers, and so mostly Standardization is preferred.

5. An infinite value of VIF indicates that the corresponding variable may be expressed exactly by a linear combination of other variables, basically called as high multicollinearity.

6. Quantile-Quantile (Q-Q) plot, is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal, exponential or Uniform distribution. It helps to determine if two data sets come from populations with a common distribution.
   Advantage:
   - It can be used with sample sizes.
   - Many distributed aspects like shifts in scale, changes in symmetry and the presence of outliers can be all detected from this plot.