Assignment-based Subjective Questions

Answers:

1. **The demad of bike is less in the month of spring when compared with other seasons**. **The demand bike increased in the year 2019 when compared with year 2018**.

2. drop_first=True is important to use, as it **helps in reducing the extra column created during dummy variable creation**. Hence it reduces the correlations created among dummy variables

3. **The numerical variable 'registered'** has the highest correlation with the target variable 'cnt' , if we consider all the features.

4. Pair-wise scatterplots may be helpful in validating the linearity assumption as it is easy to visualize a linear relationship on a plot.

General Subjective Questions

Answers:

1. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models **a target prediction value based on independent variables**. ... Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x)

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. The **Pearson correlation coefficient** is also known as **Pearson's *r***, the **Pearson product-moment correlation coefficient** (**PPMCC**), the **bivariate correlation** or colloquially simply as **the correlation coefficient** is a measure of linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations; thus it is essentially a normalized measurement of the covariance, such that the result always has a value between −1 and 1.

4. The two most discussed scaling methods are Normalization and Standardization. Normalization typically means rescales the values into a range of [0,1]. Standardization typically means rescales data to have a mean of 0 and a standard **deviation** of 1 (unit variance).

5. If all the independent variables are orthogonal to each other, then VIF = 1.0. **If there is perfect correlation**, then VIF = infinity. A large value of VIF indicates that there is a correlation between the variables.

6. The purpose of Q Q plots is **to find out if two sets of data come from the same distribution**. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line. ... This particular type of Q Q plot is called a normal quantile-quantile (QQ) plot.