

Assignment-based Subjective Questions

1_Ans -: Bike Demand is higher in the year 2019 to 2018. There is very less demand for the bikes on Spring. January has the least demand for bikes and also the demand for bikes are less on holidays.

2_Ans-: It is important to use `drop_first=True` during the dummy variable creation because it removes the extra column and also benefits from less correlations between columns.

3_Ans-: temp has 0.63 correlation with target variable which is the highest.

4_Ans -: The categories has correlation with the numerical values.

5_Ans – temp, atemp, weathersit

General Subjective Questions

1_Ans – Linear Regression Algorithm is a form of predictive modelling technique which tells us the relationship between the dependent and independent variables.

We have 2 types of Linear Regression:

- i) Simple Linear Regression
- ii) Multiple linear regression

Simple Linear Regression The most basic type of regression model is the simple linear regression which explains the relationship between a dependent variable and one independent variable using a straightline. The straight line is plotted on the scatter plot of these two points.

Multiple linear regression is a statistical technique to understand the relationship between one dependent variable and several independent variables. The objective of multiple regression is to find a linear equation that can best determine the value of dependent variable Y for different values independent variables in X

2_Ans – There are these four data sets in Anscombe's quartet plots which have nearly same statistical observations, which provides same statistical information that involves variance, and mean of all x,y points in all four datasets.

This tells us about the importance of visualising the data before applying various algorithms out there to build models out of them which suggests that the data features must be plotted in order to see the distribution of the samples that can help you identify the various anomalies present in the data like outliers, diversity of the data, linear separability of the data, etc.

3_Ans -: Pearson's R is a number between -1 and 1 that measures the strength between two variables

4_Ans -: It is a step of data Pre-Processing 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.

Most of the times, collected data set 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.

It is important to note that scaling just affects the coefficients and none of the other parameters like t-statistic, F-statistic, p-values, R-squared, etc.

The Difference between Scaling and Normalization is 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_Ans-: VIF will be infinity when there is perfect correlation This shows a perfect correlation between two independent variables. In the case of perfect correlation, we get $R^2 = 1$ which gives us $1/(1-R^2)$ infinity. To overcome this issue we need to drop one of the variables from the dataset which is causing this perfect multicollinearity.

If the Value of VIF is infinity then the corresponding variable may be expressed exactly by a linear combination of other variables

6_Ans-: Q-Q Plots which are Quantile-Quantile plots are plots of two quantiles which are against each other. A quantile is a fraction where certain values fall below that quantile. For example, the median is a quantile where 50% of the data fall below that point and 50% lie above it. 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.

