**Assignment-based Subjective Questions**

Q.1 From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?

**Answer**:

1. Season fall has the highest demand for rental bikes
2. The demand has increased from the last year
3. Demand is continuously growing each month till June. September month has highest demand. After September, demand is decreasing
4. less demand on holidays
5. Weekday is not giving clear picture about demand.
6. Weekday is not giving clear picture about demand.
7. The clear weathershit has highest demand
8. During September, bike sharing is more. During the year end and beginning, it is less, could be due to extreme weather conditions.

Q.2 Why is it important to use drop\_first=True during dummy variable creation?

**Answer:**

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.

Q.3Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

**Answer:**

temp and atemp has the highest correlation with cnt

Q.4 How did you validate the assumptions of Linear Regression after building the model on the training set?

**Answer:**

We can validate the assumptions of linear regression model by looking at the distplot of residuals or error terms. They follow normal distribution and are centered around 0. So, I validated the model by residual analysis.

Q.5 Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?

**Answer:**

**The top three features and there coefficients are as follow:**

1. **Temp -** 0.4784
2. **yr\_2019 -** 1.0285
3. **weathersit\_Light\_Snow - -**1.1228

**General Subjective Questions**

**Q.1** Explain the linear regression algorithm in detail.

**Answer:**

Linear Regression is a type of supervised Machine Learning algorithm that is used for the

prediction of numeric values.Linear Regression is the most basic form of regression

analysis.Regression is the most commonly used predictive analysis model.

Linear regression is based on the popular equation «y = mx + c".It assumes that there is a linear relationship between the dependent variable(y) and the predictor(s)/independent variable(×). In regression, we calculate the best fit line which describes the relationship between the independent and dependent variable. Regression is performed when the dependent Variable is of continuous data type and Predictors or independent variables could be of any data type like continuous, nominal/categorical etc. Regression method tries to find the best fit line which shows the relationship between the dependent variable and predictors with least error.

In regression, the output/dependent variable is the function of an independent variable and the coefficient and the error term.

Regression is broadly divided into simple linear regression and multiple linear regression.

1. Simple Linear Regression : SR is used when the dependent variable is predicted using

only one independent variable.

2. Multiple Linear Regression MLR is used when the dependent variable is predicted

using multiple independent variables.

The equation for MLR will be:

***yi* = 0 + 1*x*i1 + 2*x*i2 + ... p*x*ip + i**

B1 = coefficient for X1 variable

B2 = coefficient for X2 variable

B3 = coefficient for X3 variable and so on..

p0 is the intercept (constant term).