

## **Assignment-based Subjective Questions**

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

Categorical variables such as months, season, weather situation had strong influence to the dependent variable (cnt). For instance, bike rentals were high during holidays, clear weather, days like wednesday, thursday & saturday, months like aug, sep & oct and during summer and fall. These are few inferences around categorical variables

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

Drop first will lead to optimization and overcoming redundancy of independent variables that are created while preparing dummy variables for categorical variables.

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

temp and atemp has strong correlation

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

1. through distribution of error terms with mean as 0
2. scatter plot between error terms and x variables showing no patterns
3. homoscedasticity via uniform variance inferred from scatter plot between x and y\_pred

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

1. Temperature, 2. weather (spring, summer), 3. year