Part -1 Assignment based questions

1. From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?
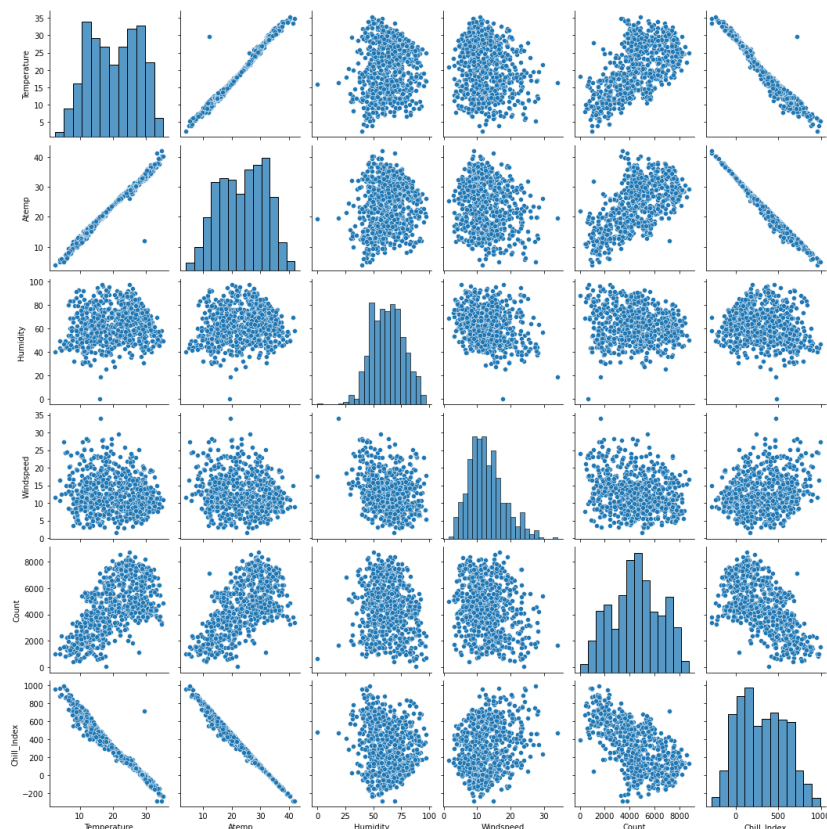
Ans. i. We can see that Fall Season has highest Counts

ii. In Year 2019, bike riding increased significantly.

iii. People don't prefer much - **Cold months from Dec- Feb**, They most prefer **period Jun-Oct**

iv. Sunday have low counts.

v. Cold weather has least counts.

2. Why is it important to use drop_first=True during dummy variable creation?

Ans. We use **drop_first=True** during the dummy variable creation because it helps in reducing the extra column created during dummy variable creation. It reduces the correlations among dummy variables.
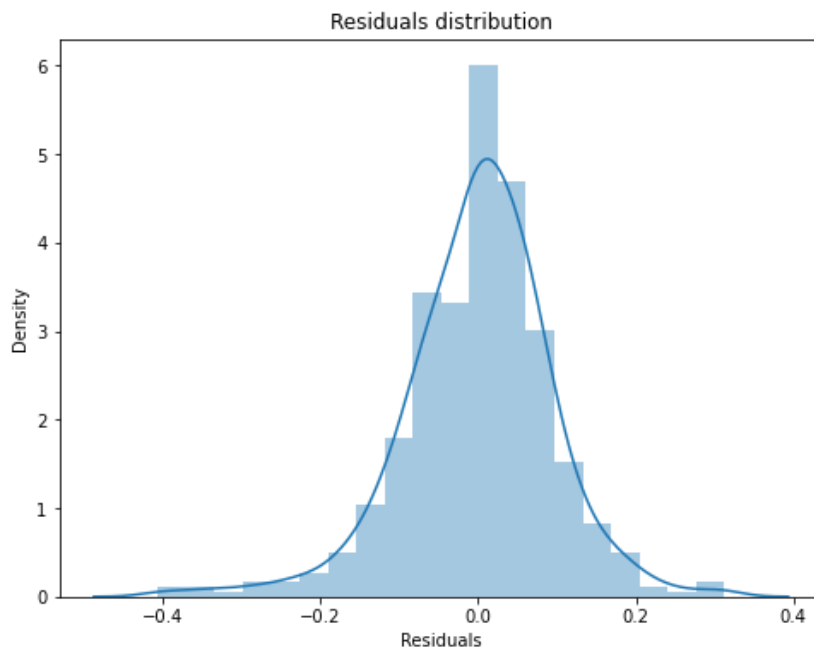
3. Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

Ans. Temp – Atemp has highest correlation with target variable

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

Ans. Looking at residuals

Residuals distribution

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

Ans. 1. Demand is increasing with year.

2. Jun to October demand is high

3. Weather conditions affect the demand.