Assignment-Based Subjective Questions

1. **Effect of categorical variables**
   Categorical variables can significantly influence the dependent variable by introducing distinct group effects. For example, "yr","weekday","weathersit" ,"casual", "registered" can affect bike demand by determining suitability for biking.

2. **Importance of `drop_first=True`**
   It avoids the **dummy variable trap** by dropping one category to prevent perfect multicollinearity, ensuring the regression model remains interpretable.

3. **Highest correlation from pair-plot**
   Identify the numerical variable with the strongest linear relationship (highest absolute correlation coefficient) with the target variable, as seen in the pair-plot.

4. **Validation of Linear Regression assumptions**
   - **Normality**: Use a dist plot of residuals.
   - **Multicollinearity**: Check VIF scores.

5. **Top 3 features for bike demand**
   Based on model coefficients or p-values, features like **registered**, **yr**, and **casual** typically emerge as significant.

General Subjective Questions

1. **Linear regression algorithm**
   A predictive model to determine the linear relationship between independent (X) and dependent (Y) variables by minimizing residual sum of squares using the equation $Y = \beta_0 + \beta_1 X + \epsilon$.

2. **Anscombe's quartet**
   Four datasets with identical summary statistics (mean, variance, correlation, regression line) but distinct visual patterns, emphasizing the importance of visualizing data.

3. **Pearson's R**
   A measure of the linear relationship between two variables, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). $r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$.

4. **Scaling and its types**
   - **Scaling** adjusts feature ranges for model stability and convergence.
   - **Normalized Scaling**: Rescales to [0, 1] or [-1, 1].
   - **Standardized Scaling**: Centers data (mean=0, SD=1).

5. **Infinite VIF**
   Occurs due to **perfect multicollinearity**, i.e., when one predictor is a perfect linear combination of others.

6. **Q-Q plot**
   Plots residual quantiles against theoretical quantiles to check if residuals follow a normal distribution. Essential for validating the normality assumption in linear regression.