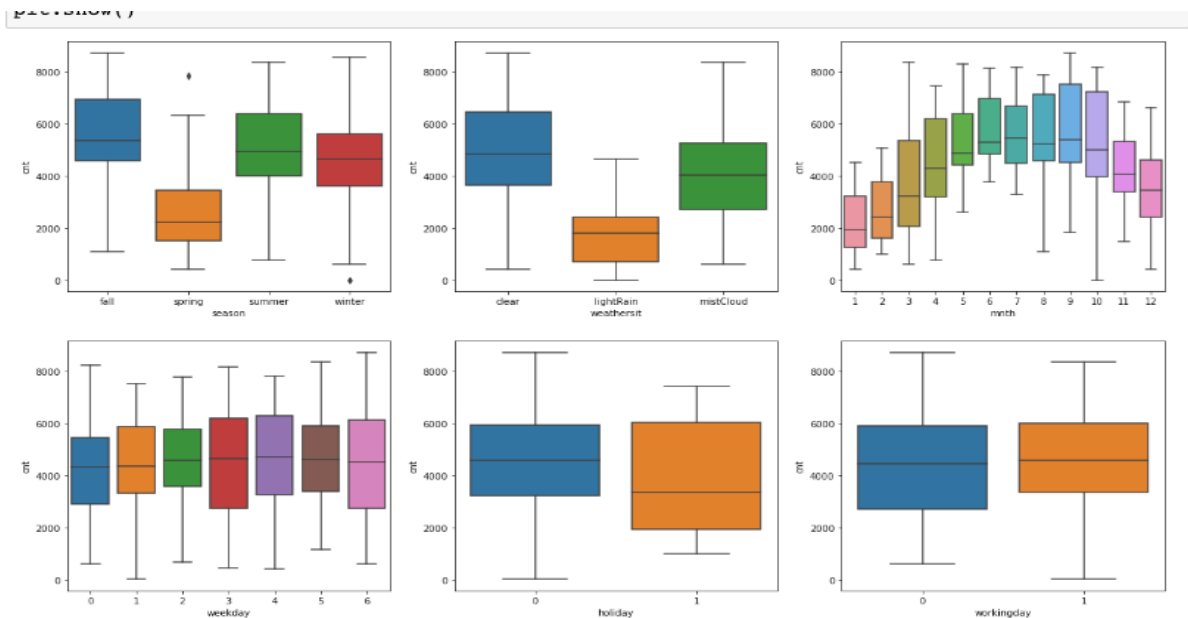


Assignment-based Subjective Questions

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

Ans:- Season, month ,weekday ,weathersit , holiday , working days are categorical variables:-
These categorical influence the dependent variable in following ways :-



1:Season:-Spring season had lowest value of cnt value and fall season is highest value of cnt value. Summer and winter have median almost to 5k-6k.

2: weather situations:- cnt is more when weather clear and mild cloud

3:month: September is having more rental while January is less .

4:weekday :Across cnt median is same . weekday is very less influence towards dependent variable

5:holiday:- Rental found lower during holiday

8:working day: Rental more on working day

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

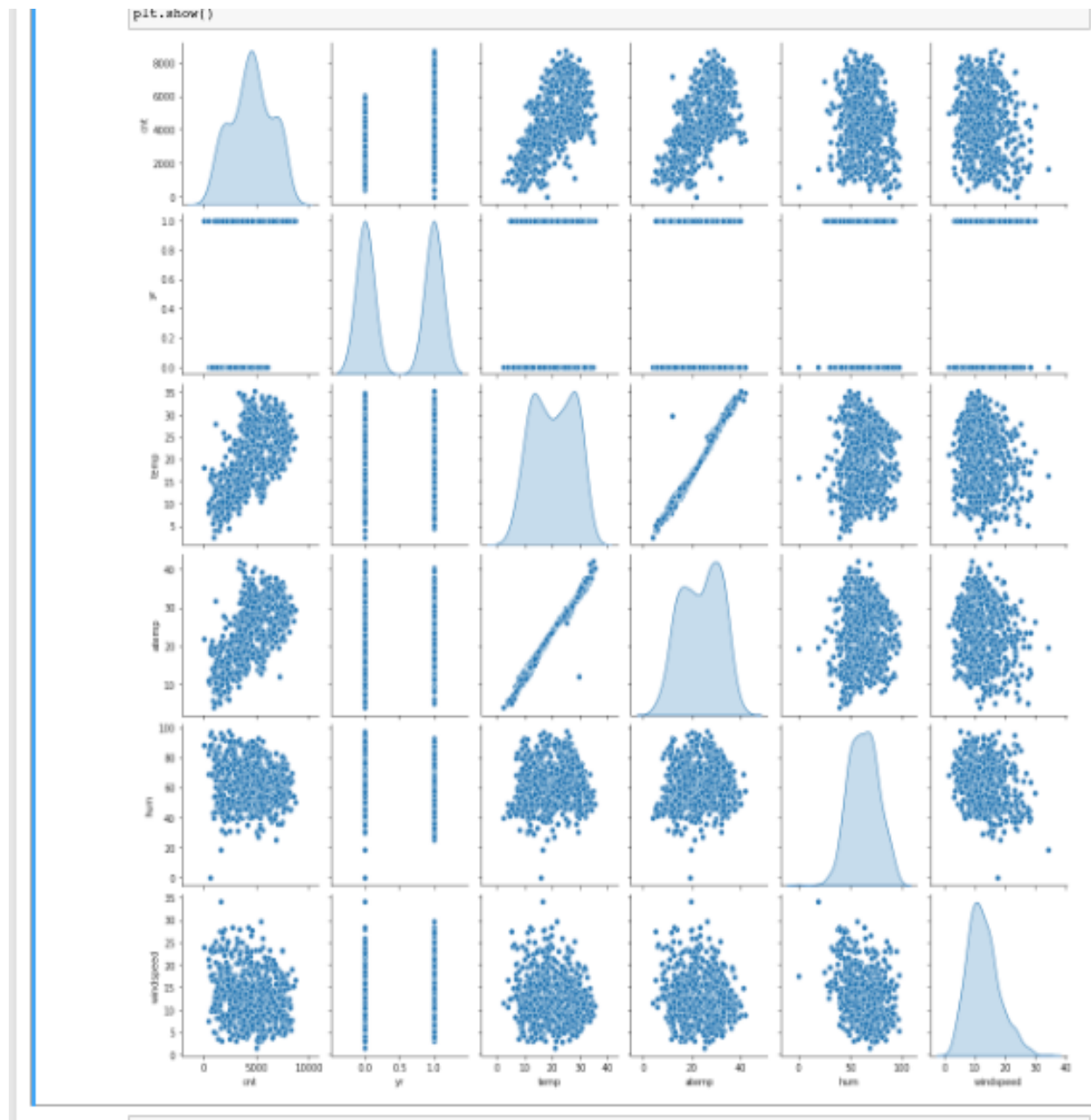
creating dummy variables from categorical data, the parameter `drop_first=True` is used to prevent multicollinearity in regression analysis and to improve the interpretability of the model. Let's break down why this parameter is important and how it addresses these issues:

1: Multicollinearity: Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated.

2: Interpretability:

Including all dummy variables in a regression model can make it difficult to interpret the effects of each category. By dropping the first category, you establish a reference point against which the effects of the other categories can be compared. This makes it easier to understand the impact of each category on the response variable.

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



temp and atemp are two parameter which is highly correlated with cnt.

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

The dis