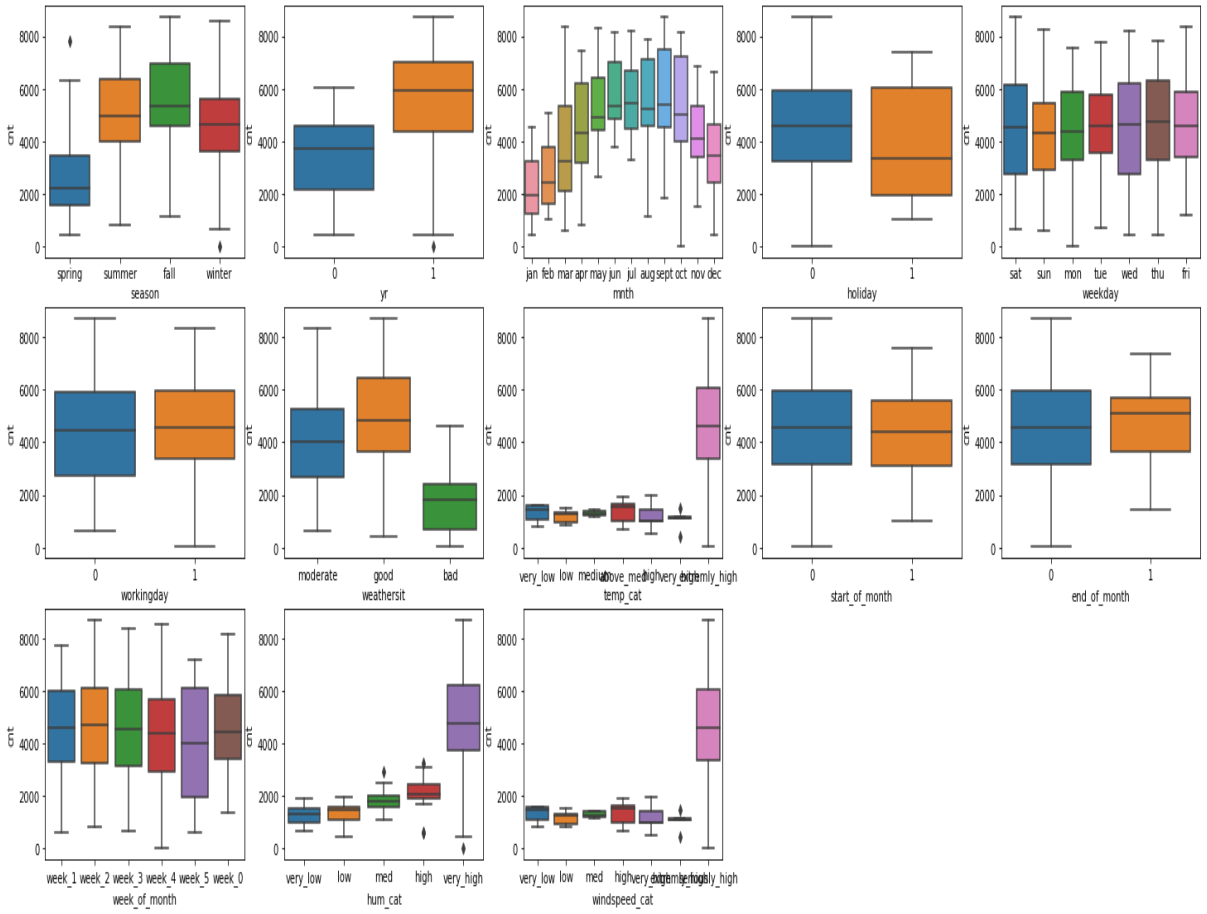
Assignment-based Subjective

Questions 1. From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable? (3 marks)
Ans:

Categorical values that are considered in the dataset, and few also derived are:
'season','yr','mnth','holiday','weekday','workingday','weathersit', 'temp_cat',
'start_of_month','end_of_month','week_of_month', 'hum_cat', 'windspeed_cat'



Note: Graph can be checked in notebook for better visualization

1. Fall has the Highest Demand, whereas spring has the lowest demand
2. Demand in bike sharing business has increased in 2019 compared to 2018
3. Demand keeps on increasing till September, with September Month being month of Highest demand. And there is decreased demand from Oct to Jan.
4. Higher demand when it is not a Holiday.
5. Weekends have slightly higher demand, but the mean is more or less the same. We can derive that there is not much impact of day of week in the data. And similar is with working day
6. Bad weather has the lowest demand, whereas when the weather is good, there is a high demand.

7. Start of the Month has less deviation in demand compared to the rest of the days.
8. End of the Month also has less deviation in demand compared to the rest of the days.
9. Week of the Month does not have much to say

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

Ans:
drop_first=True parameter in the dummy variable creation is important in order to maintain the N-1 Columns,
Example:
If we have 12 Columns of the Months, then the final number of columns for dummy variable needs to be N-1 which is achieved by drop_first=True