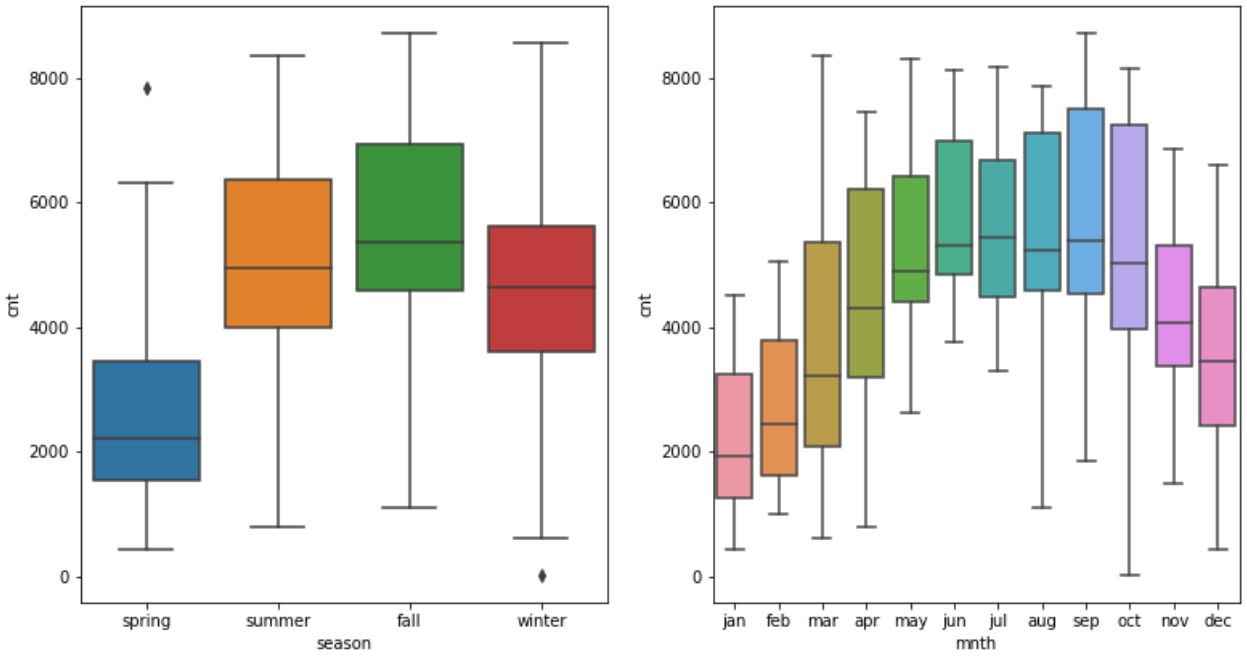
Assignment-based Subjective Questions

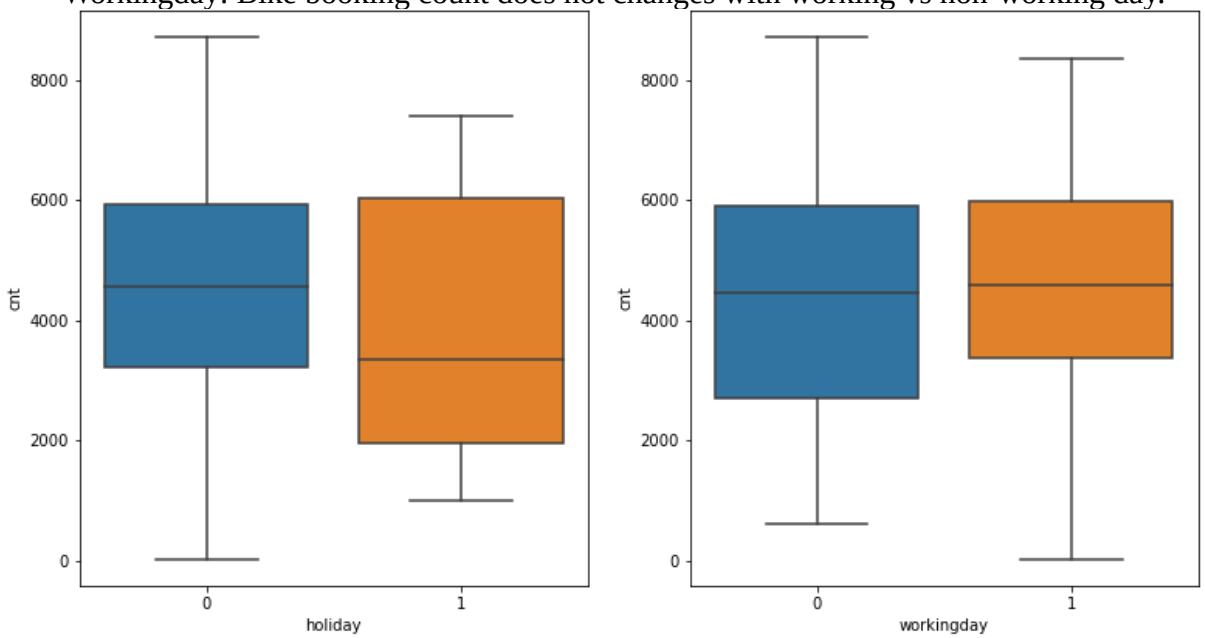
1. **From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable? (3 marks)**

Answer: We are having the following catagorical features in the dataset:

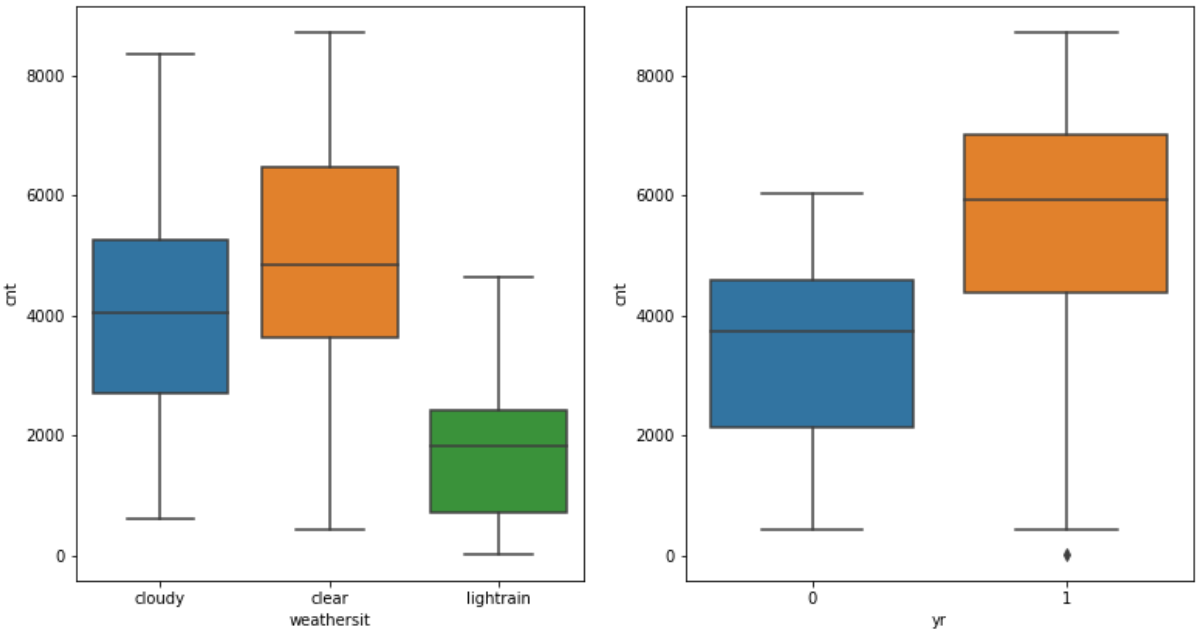
- Season: Counts of rented bikes depends on season, cnt increases in Fall and Summer season and it gets decreases with winter and spring
- mnth: Cnt varies with months as we are having higher booking with May, June, July , August, and September. And booking drcreases with Winters month.



- Holiday: Cnt depends on holiday. On non-holiday we are having higher cnt value.
- Workingday: Bike booking count does not changes with working vs non-working day.



- Yr: Count of booking increases with year. We are having higher no of bike renting in 2019 year compared to 2018.



- Weathersit: Weathersit afftects the bike booking. Clear weather attracts the customers and have higher booking then cloudy and then light rain.

2. **Why is it important to use drop_first=True during dummy variable creation? (2 mark)**
Answer: When we are converting the catagorical features to numeric feature. So that these can be scalarized. We break the single column into the all possible probable columns and fill the values between 0 and 1.
For example: We are having the following catagorical feature:

Season
summer
winter
fall
spring

And it will be converted to following table:

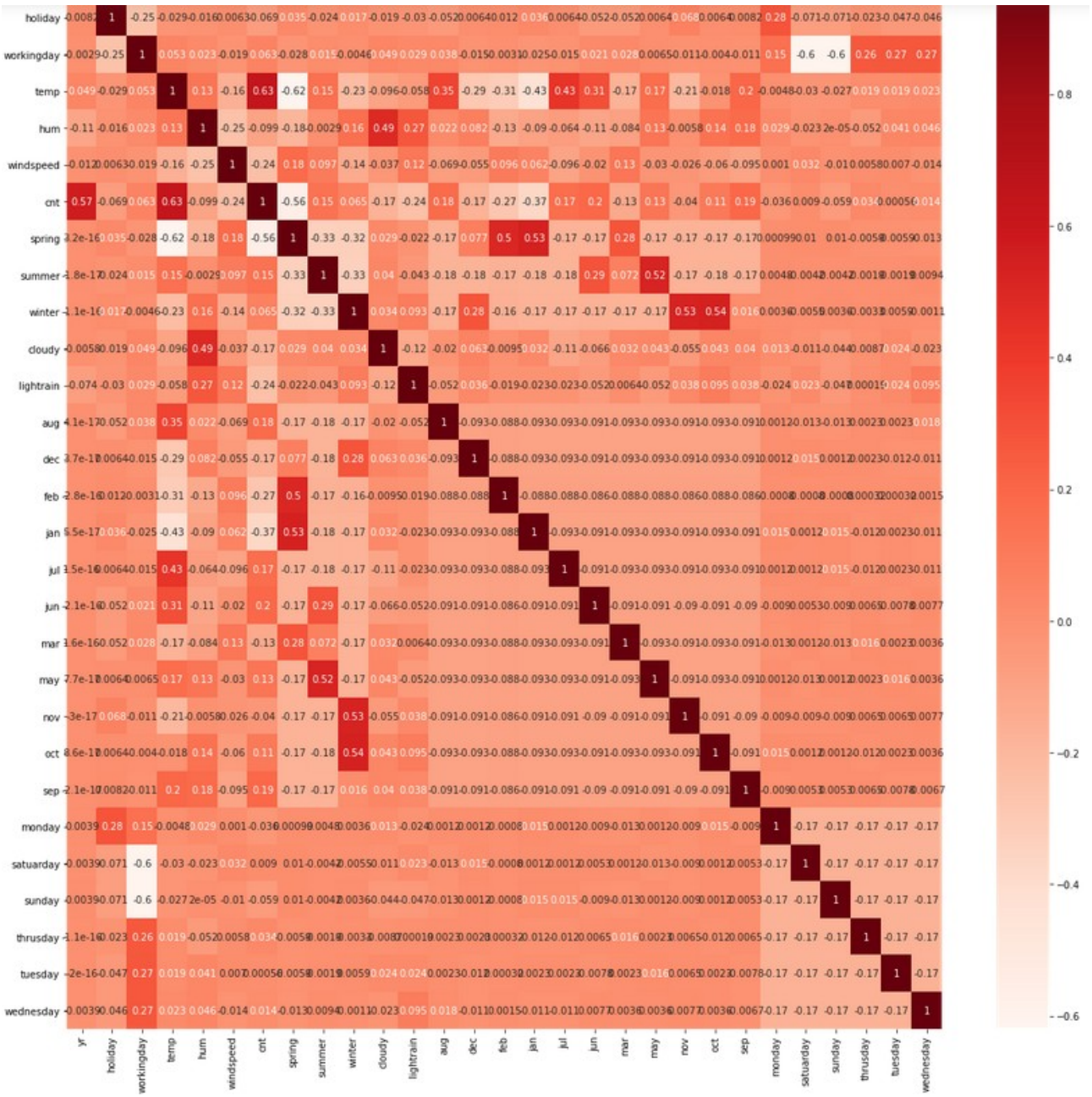
summer	Winter	Fall	spring
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

But we can extract the information with with 1 less column like if all the values are zero the we can assume it would be the final option and we can have the following table:

Winter	Fall	spring
0	0	0
1	0	0
0	1	0
0	0	1

As we can see 4 colums are reduced to 3 columns.

3. **Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable? (1 mark)**
Answer: temp attribute is having the highest correlation with cnt (target variable) with 0.63 value.


yr
holiday
workingday
temp
hum
windspeed
cnt
spring
summer
winter
cloudy
lightrain
aug
dec
feb
jan
jul
jun
mar
may
nov
oct
sep
monday
satuarday
sunday
thrusday
tuesday
wednesday
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6