# Subjective Questions Solution

## Assignment based subjective questions

**Q1)** From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?

**Ans)** In our dataset there are following categorical variables

* season [1: Sprint, 2: Summer, 3: Fall, 4: Winter] has 4 levels.
* yr [0: 2018, 1: 2019] has 2 levels.
* mnth [1 to 12] has 12 levels.
* holiday [0: yes, 1: no] has 2 levels.
* weekday [0: yes, 1: no] has 2 levels.
* weatherlist has 4 levels.
  + 1: Clear, Few clouds, Partly cloudy, Partly cloudy
  + 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
  + 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
  + 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

There is total 12 dependent variables earlier & from above categorical variables we have 26 levels, which will increase the dependent variables count to 38.

**Q2)** Why is it important to use **drop\_first=True** during dummy variable creation?

**Ans)** Since categorical variables can have multiple levels which could increate the number of dependent variables in the dataset, hence, to reduce the number of dependent variables in the dataset we use the drop\_first = True. As we can also interpret a categorical variable with N levels into N-1 variables. E.g. Suppose our categorical variable has 2 levels(‘yes’ or ‘no’), then we can map ‘yes’ to 1 and ‘no’ to 0, then with only 1 variable we can interpret the categorical variable value, if it is 1 then yes otherwise no.