**ADJUSTED R SQAURE**

* We use R2 as discussed, as a goodness of fit parameter. The higher the value of R2, the better the model is – the close R2 is to 1, the better the model.
* This is all good and works perfectly. But the problem starts to occur when you add more variables to your model.
* When you add a new variable to your current regression model, R2 will never decrease. There are only two possibilities for R2 when you add new variables to the current regression model – 1. It will stay the same.

2. R2 will increase.

* The newly added variable will never be equal to 0, it will always be somehow slightly correlated to the independent variable.
* So, due to those slight random correlations, there would be minor changes in R2 – it could increase or decrease slightly.
* That is why there is a problem with R2, you can add variables, and you will not know if those variables are helping your model or if they’re not. Because your R2 is biased and is always increasing regardless of actual improvement or degradation of fit.
* And this is where Adjusted R-square comes in.
* p – number of regressors (independent variables)

n – sample size

* It is said that Adjusted R square has a penalization factor. It penalizes you for adding variables that don’t help the model.