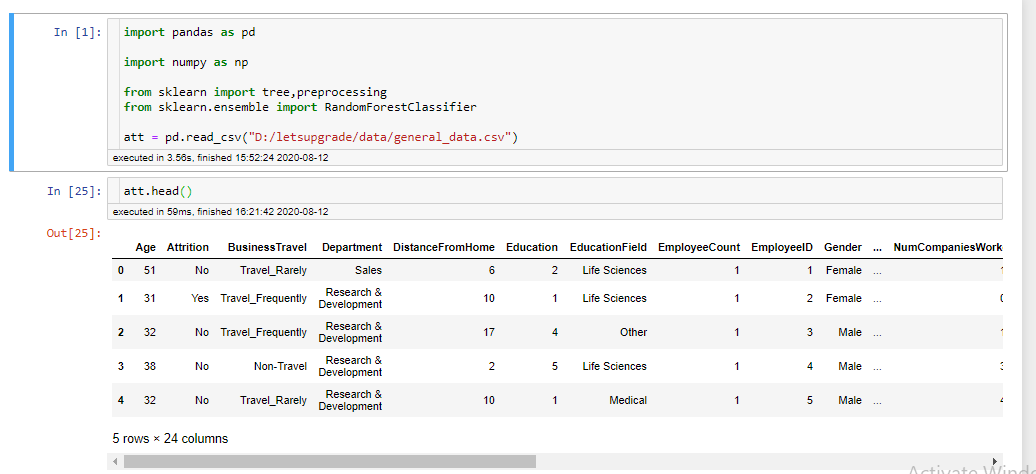
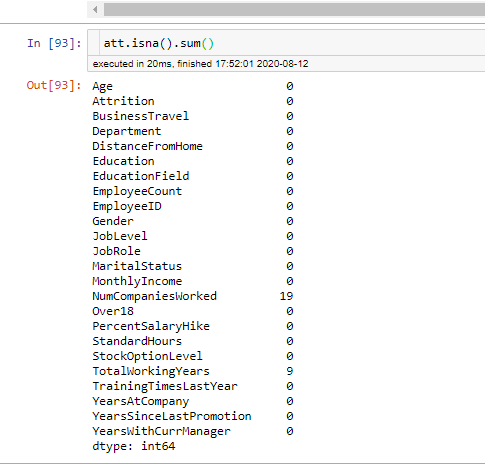
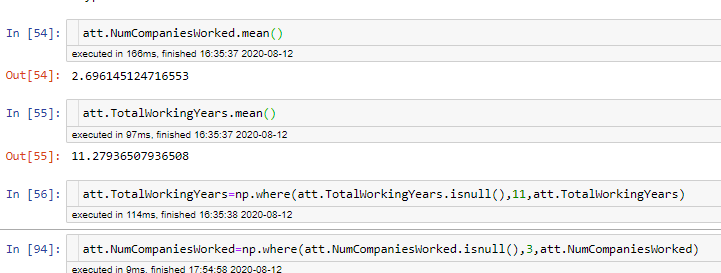
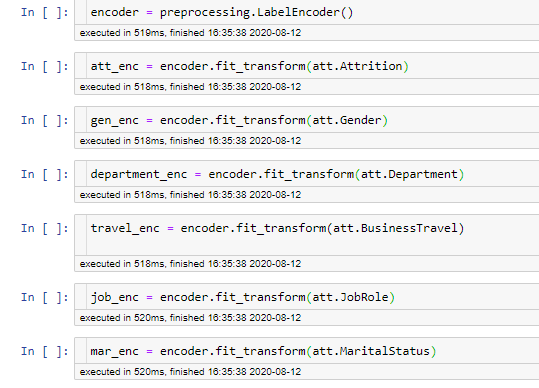
**Attrition data Decision Tree**

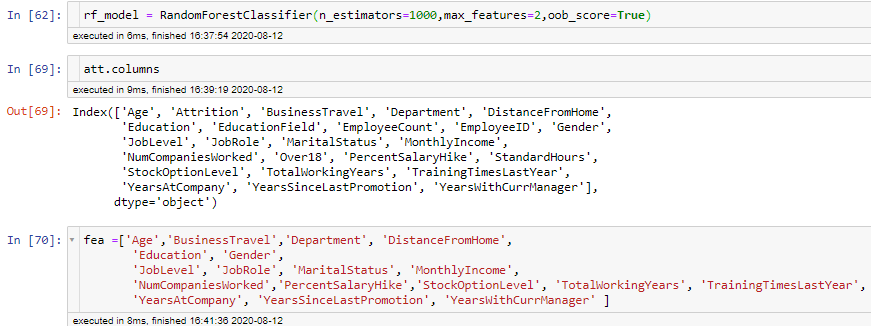
* Import required module or take dataset that we required to perform make Decision tree

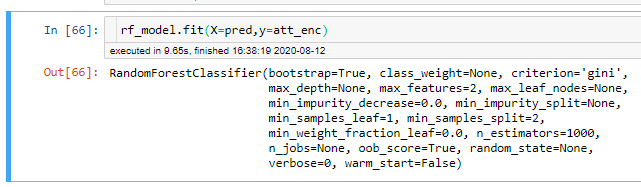
Check how much null data item present in data 

* Fill mean value in continues feature & categorical we fill mode of features



* there are many columns so we required to select only important variable so over tree is not over fit so we apply random forest algo
* encode the data : mean pandas not do operation on string value so we required to do convert into numerical value
* 
* Select feature for random forest algorithm mean that we can ignore id,name like feature for random forest
* Attrition is has two category so we required to do binarysplit so we fix max\_features =2





* Importance of model in sorted order :

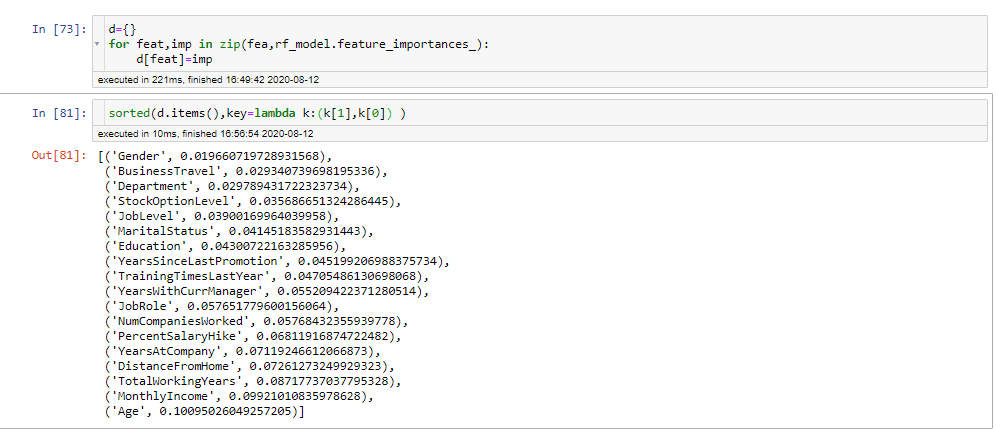
('DistanceFromHome', 0.07261273249929323),

('TotalWorkingYears', 0.08717737037795328),

('MonthlyIncome', 0.09921010835978628),

('Age', 0.10095026049257205)

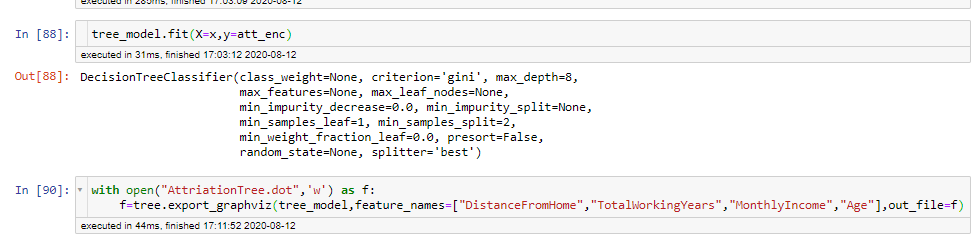
max value in importance

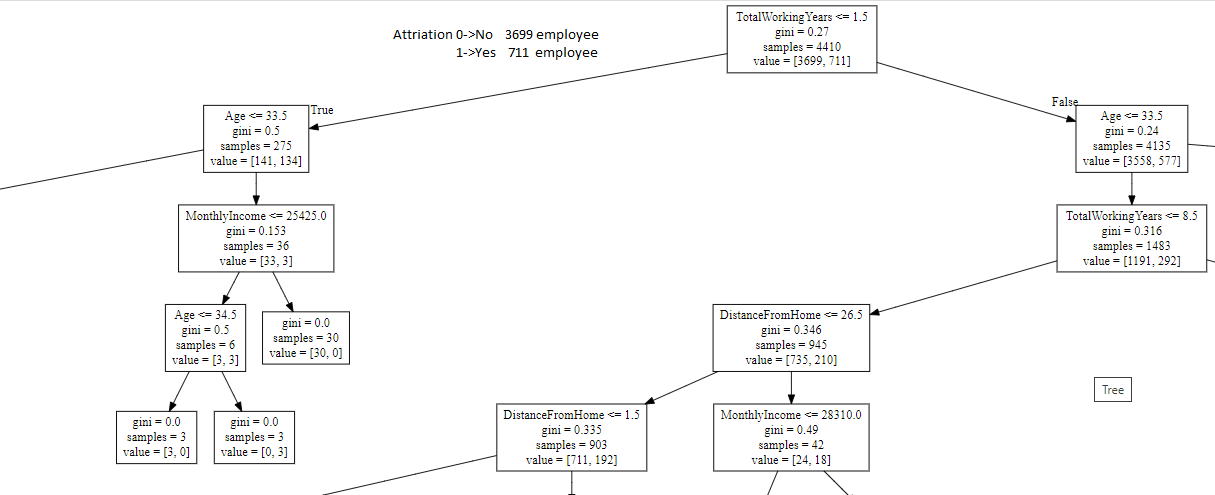


* Give max depth of tree =8 because we like to choose only for variable for decision tree drow so 4\*2 =8



* Draw tree in .dot file



,

1. totalworkingyear<=1.5 & age >33.5 & age <=34.5 & Monthaly income <25425.0 ->no attrirtion
2. totalworkingyear<=1.5 & age >33.5 & Monthaly income >25425.0 ->no attrirtion
3. totalworkingyear<=1.5 & Monthaly income <25425.0 & age >=34.5 ->Yes attrirtion

similarly many rule we can define from decision tree