**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date-20/8/2020**

**Content flow:-**

**1- Introduction:-**

**2- About the dataset**

**3- Advantage of the model**

**4- Problem Statement**

**5- Importing the required libraries**

**6- Importing the dataset and checking the dataset shape**

**7- Checking the data type**

**8- Checking for missing values**

**9- EDA(Exploratory data analysis)**

**Univariate Analysis:-**

1🡪Age

2🡪EnvironmentSatisfaction

3🡪Jobinvolvement

4🡪 JobLevel

5🡪 TotalWorkingYears

6🡪 YearSinceLastPromotion

7🡪 YearsinCurrentrole

8🡪 MonthlyIncome

**Multivariate Analysis:-**

1🡪Age, Attrition & Gender

2🡪MonthlyIncome,Attrition,OverTime

**10- Feature Engineering and data preprocessing**

**11-Divideing training and testing set as we don’t want to scale the target variable.**

**12-skewness removal**

**13-Data Scaling**

**14-Model Building**

**15-Choosing and saving the final model**

**16-Testing the test data on the final saved model**

**17-Saving the test results in a csv format in local drive**

**18-Conclusion.**

**An Article on Employee attrition in a company(IBM dataset)**

**Introduction:-**

Employee leaving (attrition) is a major problem in any industry. The resource spent by the company on the employee become a complete loss when an employee leaves a company. There are a many reasons which strike our mind directly when we think why the employee left the company.

But there are many other features also which impact the employee attrition which we don’t take into consideration while thinking why the employee left.

In this Project we will analyze many such features which affect the employee attrition and checkout the major reason of employee attrition.

We will also do a predictive analysis on the dataset and try to predict which employee is likely to leave the company in the near future so that we can take necessary steps to prevent the employee from leaving the company.

**About the dataset:-**

The dataset contains 35 features & 1470 rows.

**Advantage of the model:-**

We can take preventive measures to stop the employee from leaving the company.

**Problem Statement:-**

**My goal is to predict the employee Attrition-Yes (1) or No (0).**

**Importing the required libraries:-**

This is a binary class classification problem.

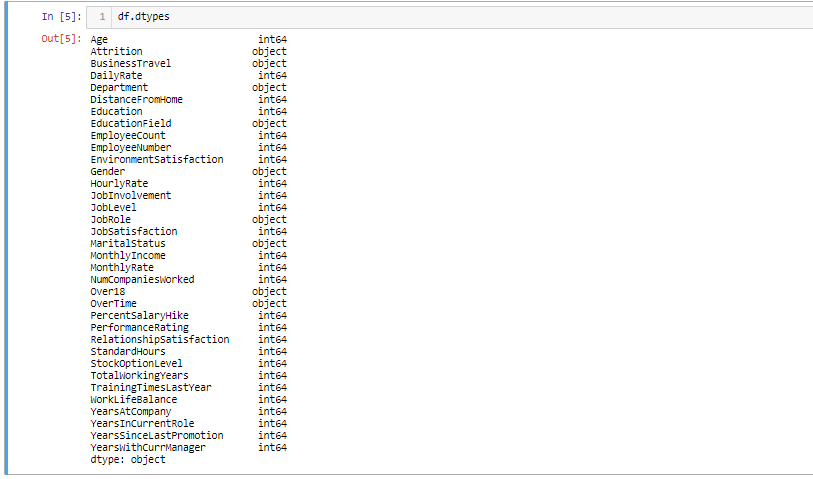
We will use accuracy\_score as our evaluation metrics to predict the employee Attrition-Yes(1) or No (0).



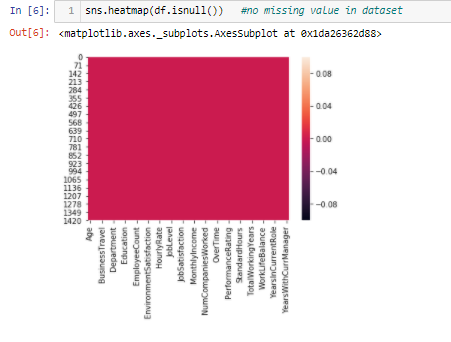
**Importing the dataset and checking the dataset shape :-**

****

**Checking the datatype:-**

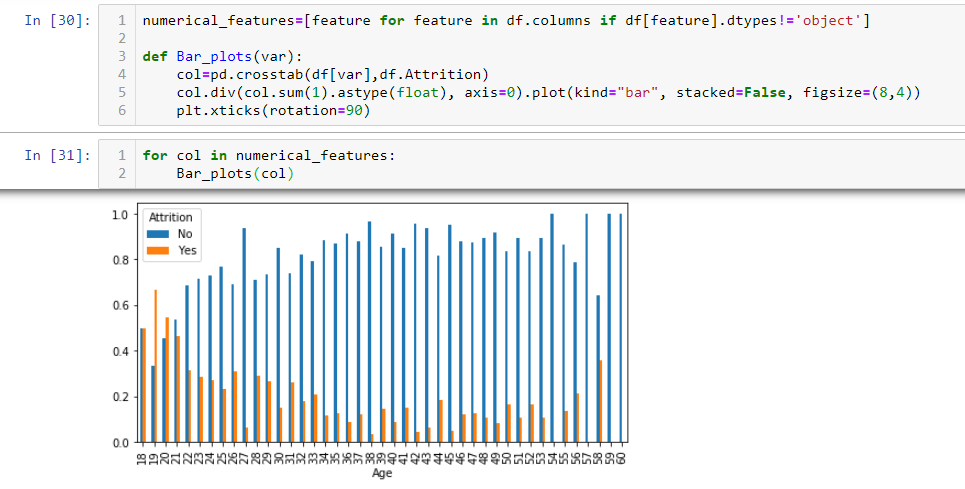
****

**Checking for missing values:-**

****

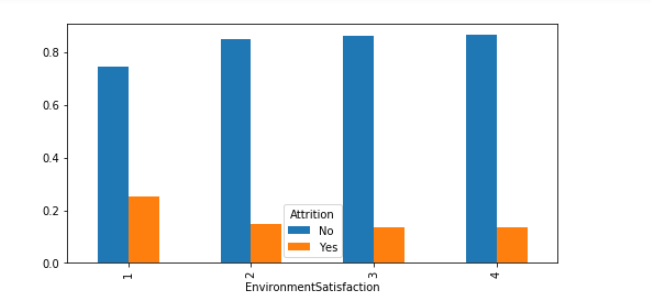
**EDA(Exploratory data analysis)**

**1-Age**

****

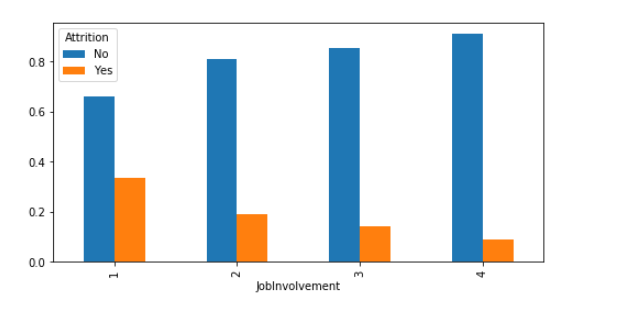
**From graph above We can analyze that Age is almost inversely proportional to attrition.Employee below age 22 have more attrition.**

**2-EnvironmentSatisfaction**

****

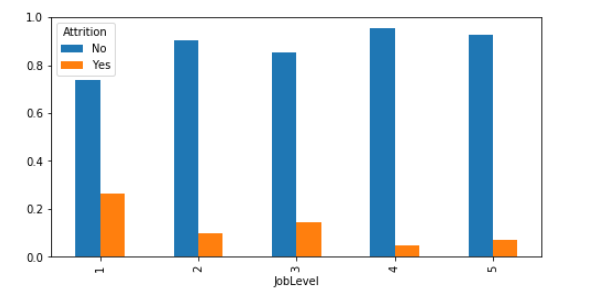
**Level-1 EnvironmentSatisfaction have higher attrition in comparison to level-2,3,4 .**

**3-Jobinvolvement**

****

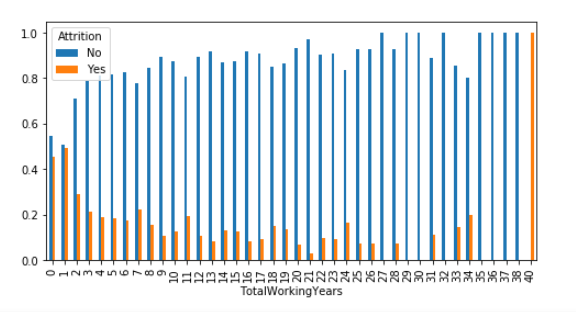
**Level-1 Jobinvolvement have higher attrition in comparison to level-2,3,4 .**

**4-JobLevel**

****

**Level-1 JobLevel have higher attrition in comparison to level-2,3,4 .**

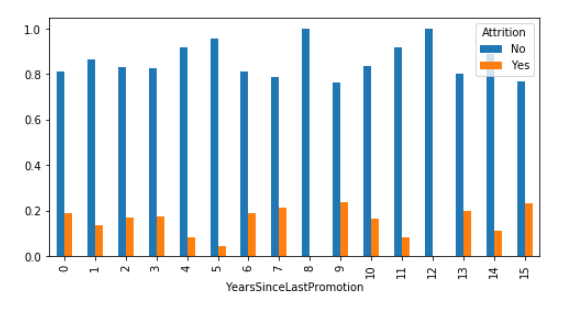
**5-TotalWorkingYears**

****

**We can see that for TotalWorkingYears=40 the attrition is 100%.**

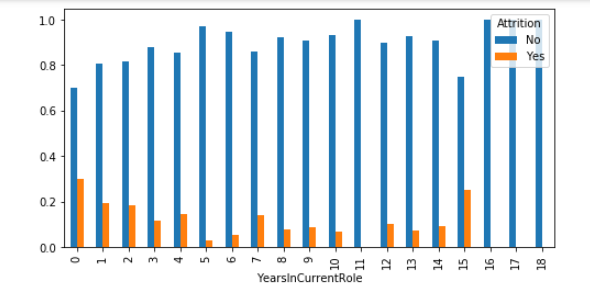
**It might be that the employee retires when TotalWorkingYears=40.**

**6-YearSinceLastPromotion**

****

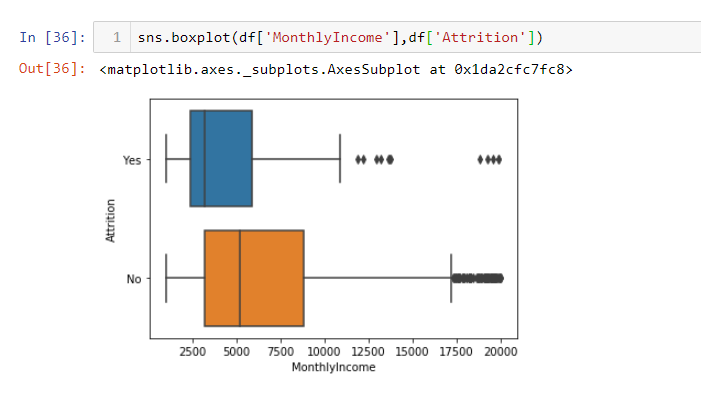
**There is a decrease in attrition when an employee is not promoted in 5th and 6th years.Thats strange.**

**7-YearsinCurrentrole**

****

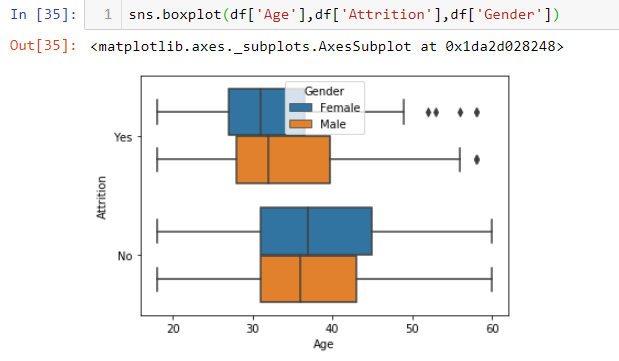
**Employees who are in current role for around 0 to 4 years have high attrition.**

**8-MonthlyIncome**

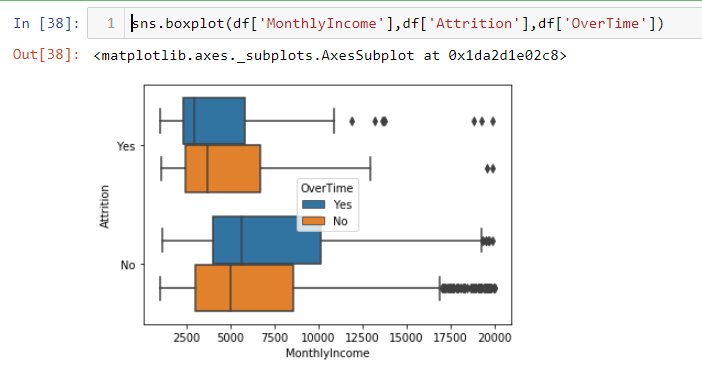
****

**Employees whose monthly income is below 5000 Units have high attrition.**

**Multivariate analysis**

****

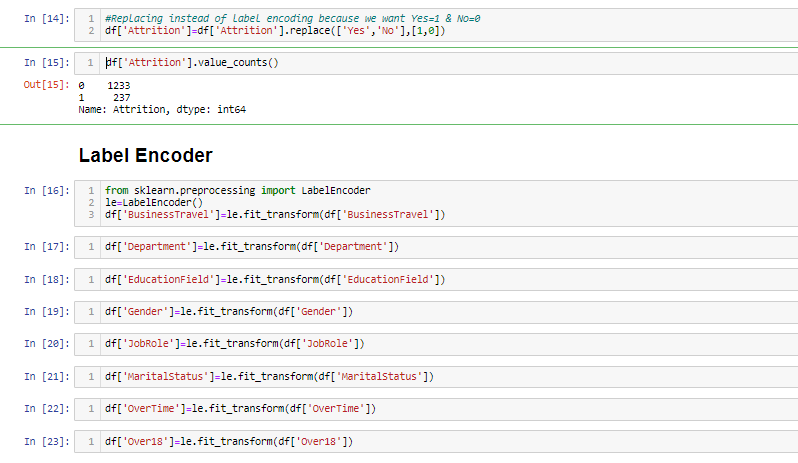
**Attrition is more common in the younger age groups and it is more likely with females As Expected it is more common amongst single Employees who leave for getting married may be(just a hypothesis)**.



**Employees who work over time have high attrition than employees who did not do over time.**

**Feature Engineering and data preprocessing:-**

**Replacing the target variable with desired integers**

****

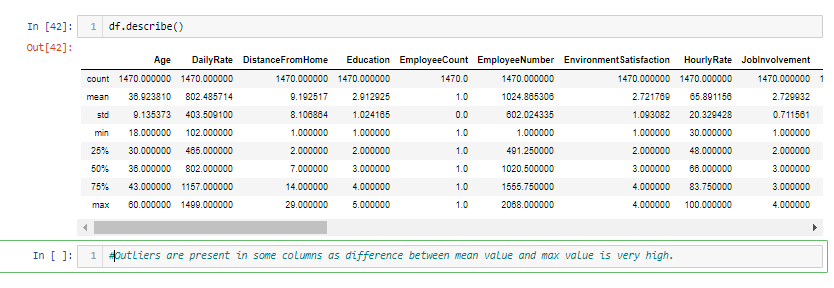
****

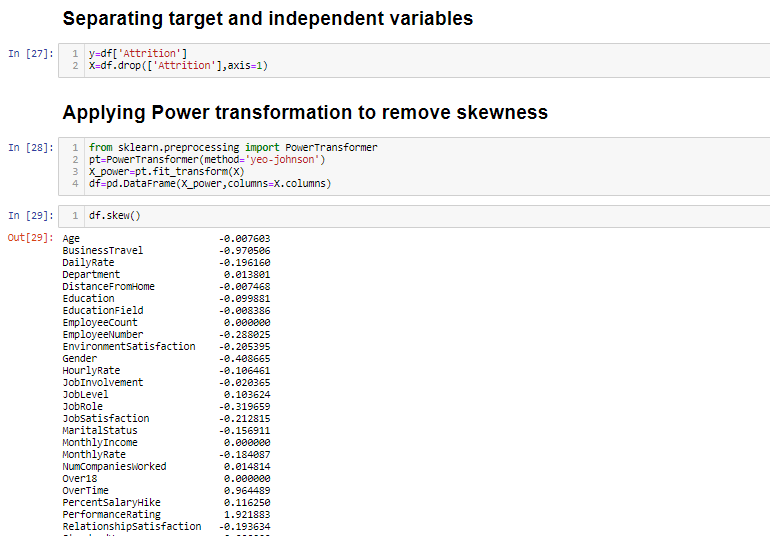
**All categorical data label encoded and converted to integer for feeding into the model building process.**

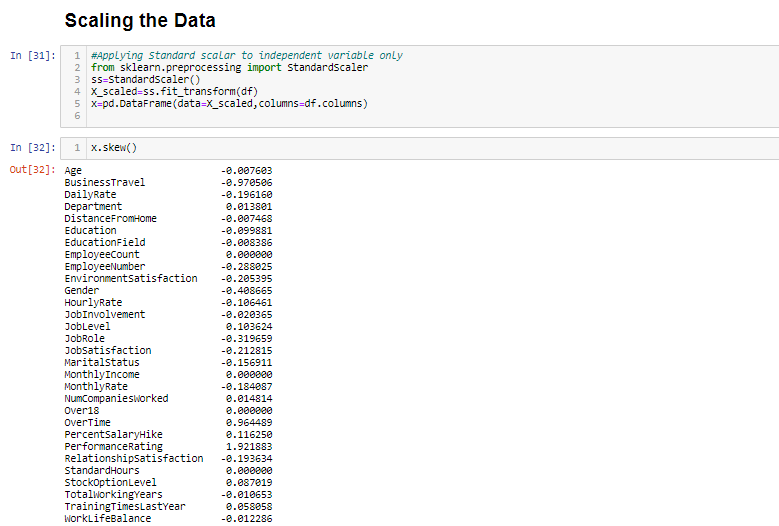
**HeatMap:-**

****

**Lighter the color shade higher is the correlation between the features as per the scale on the right.**

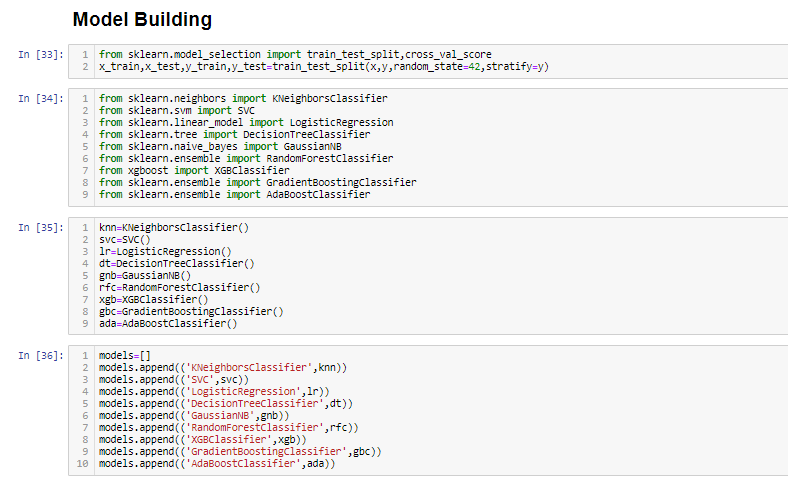
****

****

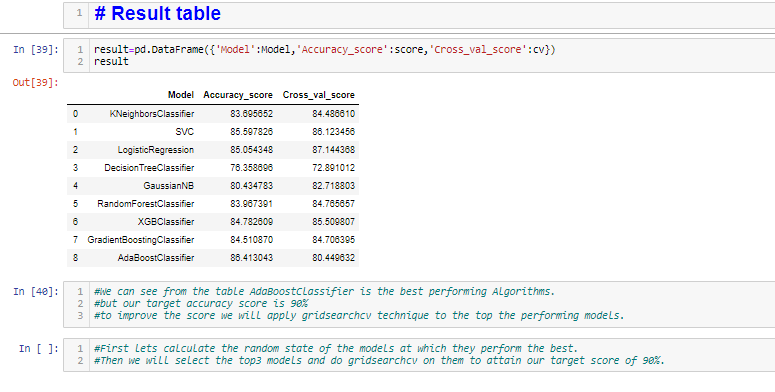
****

**Checking the data we can see that all the data have been scaled and in numerical form.**

**Hence the data is ready to feed to the model building process as we know our machine learning algorithms only can process numerical data.**

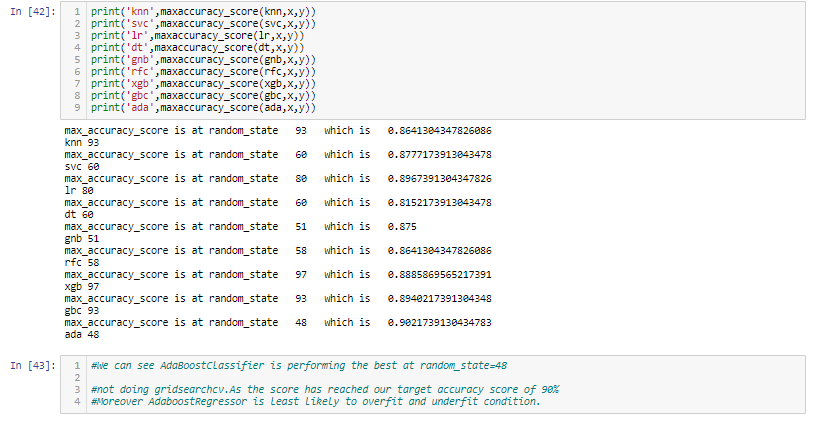
****

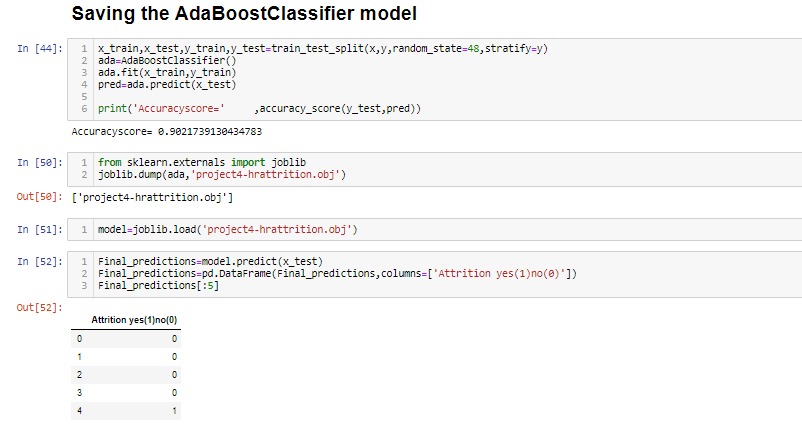
****

****

****

**Output of beat random\_state for each model**

****

****

****

**Conclusion:-**

**1-Wether the employee will leave the company(attrition) yes(1) or not(0) can be predicted using the model above.**

**2-By using sklearn(AdaBoostClassifier)we have build a machine learning model that will predict whether the employee leave the company or not.**

**3-By predicting which employee will leave We can take preventive measures to stop that employee from leaving the company.**