**Article**

**On**

**HR Analytics Project- Understanding the Attrition in HR**

**Author : Sidharth Das**

# Problem Definition :

# 

**Every year a lot of companies hire a number of employees. The companies invest time and money in training those employees, not just this but there are training programs within the companies for their existing employees as well. The aim of these programs is to increase the effectiveness of their employees. But where HR Analytics fit in this? and is it just about improving the performance of employees?**

**HR Analytics**

**Human resource analytics (HR analytics) is an area in the field of analytics that refers to applying analytic processes to the human resource department of an organization in the hope of improving employee performance and therefore getting a better return on investment. HR analytics does not just deal with gathering data on employee efficiency. Instead, it aims to provide insight into each process by gathering data and then using it to make relevant decisions about how to improve these processes.**

**Attrition in HR**

**Attrition in human resources refers to the gradual loss of employees overtime. In general, relatively high attrition is problematic for companies. HR professionals often assume a leadership role in designing company compensation programs, work culture, and motivation systems that help the organization retain top employees.**

**How does Attrition affect companies? and how does HR Analytics help in analyzing attrition? We will discuss the first question here and for the second question, we will write the code and try to understand the process step by step.**

**Attrition affecting Companies**

**A major problem in high employee attrition is its cost to an organization. Job postings, hiring processes, paperwork, and new hire training are some of the common expenses of losing employees and replacing them. Additionally, regular employee turnover prohibits your organization from increasing its collective knowledge base and experience over time. This is especially concerning if your business is customer-facing, as customers often prefer to interact with familiar people. Errors and issues are more likely if you constantly have new workers.**

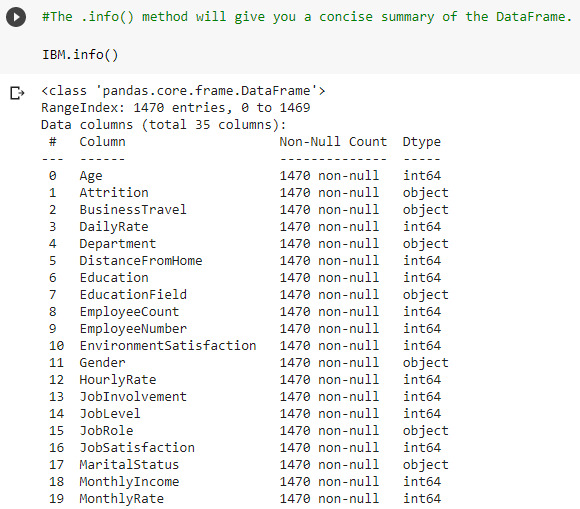
# Data Analysis :

**For this undertaking, we've got used ‘ibm hr analytics worker attrition overall performance’ dataset, which is downloaded from github, and incorporates the statistics for 1470 personnel. The capabilities within the dataset consist of ‘age’, ‘attrition’, ‘department’, ‘education’, ‘employeecount’, ‘gender’, ‘jobrole’, ‘jobsatisfcation’, ‘monthlyincome’, ‘percentsalaryhike’, ‘performancerating’ and ‘yearsatcompany’ among others. We can study the features of this dataset to are expecting whether or not a particular worker is going to depart the corporation or now not.**

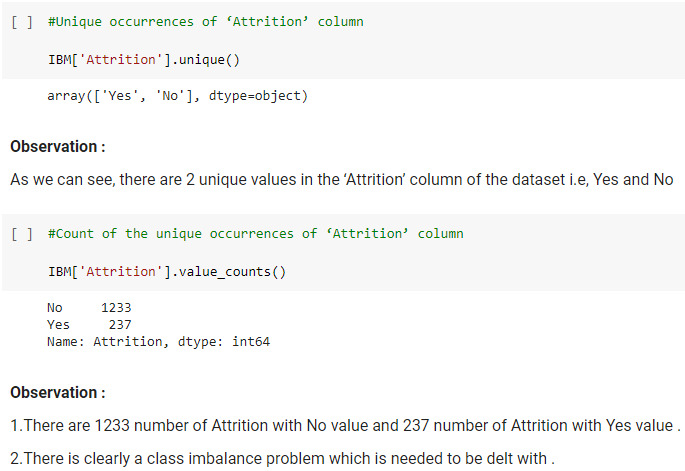
**First we import the csv file. The dataset has 1470 rows and 35 columns.**

****

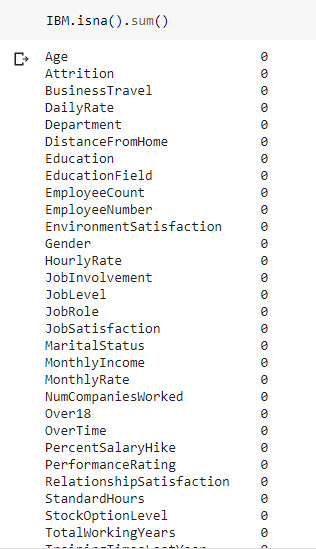
**The dataset contains several numerical and categorical features that give the details of the employee. Fortunately, none of the columns of the dataset contain any missing values.**

****

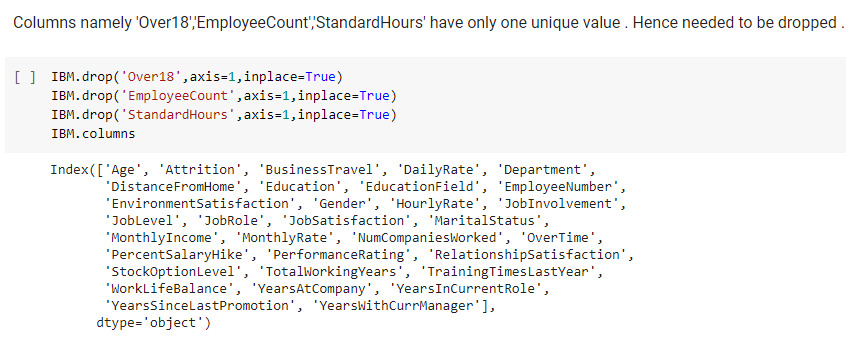
**Checking uniqueness of Target Variable ‘Attrition’**



**Checking for Null Values**

****

**Dropping Irrelevant Columns :**

****

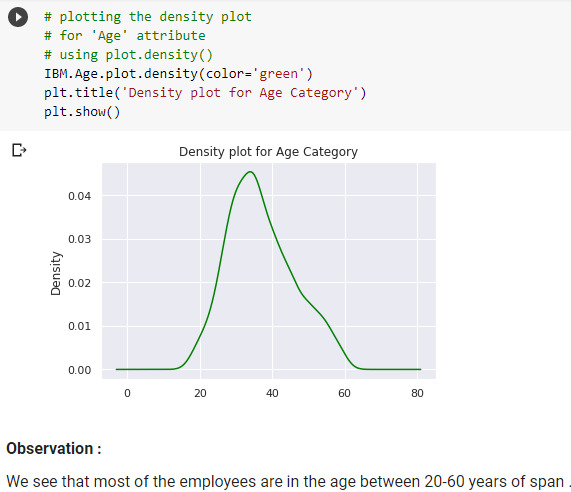
# EDA Concluding Remarks :

# Target Variable (Attrition)

# C:\Users\lenovo\Pictures\Blogs_Pic\41.jpg

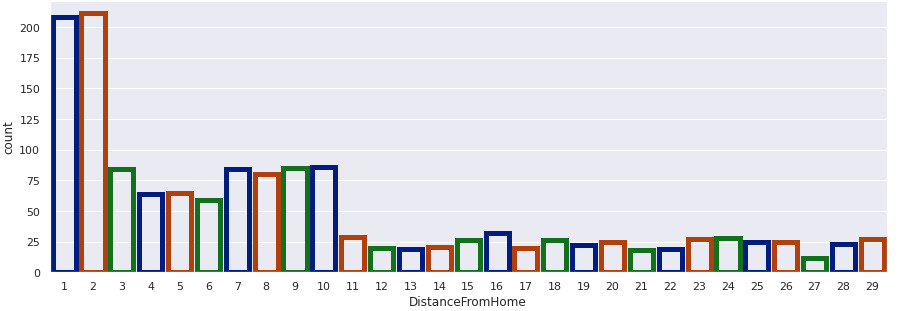
**Remarks : In this dataset, 1233 of the 1470 employees are still working in the company and 237 employees have left. This plot tells us that the dataset is imbalanced. Hence, we will have to make sure we split the dataset properly.**

**Age :**

****

**Remarks : The Age column has a normal distribution. Most of the employees in the company are in their 30’s. The mean age of the employees that leave the company is 33, and the mean age of the employees that stay with the company is 37.**

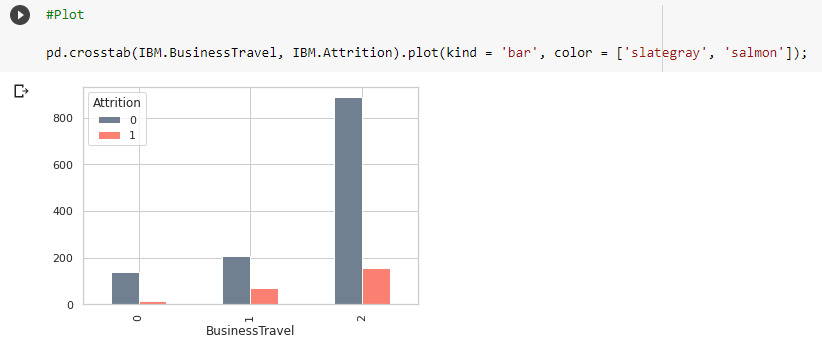
**Distance from Home :**

****

**Remarks : DistanceFromHome is heavily right skewed. The mean of distance from home to the company for the employees is 9.2. The average distance from home to the company for employees who are currently working in the company is 7 and for the employees who have left the company is 9. We will have to treat the column for skewness before building the model.**

**Buisness Travel :**

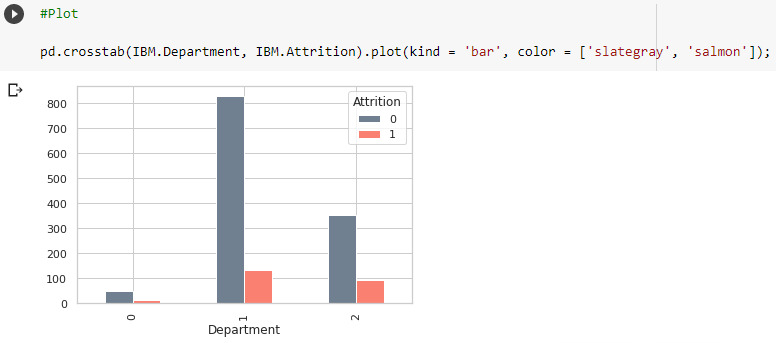
****

****

**Remarks : From 1043 employees that rarely travel for business, 887 employees are still working in the company and 156 have left the company. Out of 227 that travel frequently, 69 have left and 208 are still with the company. And among the employees who do not travel at all 138 are still in the company and 12 have left.**

**Department :**

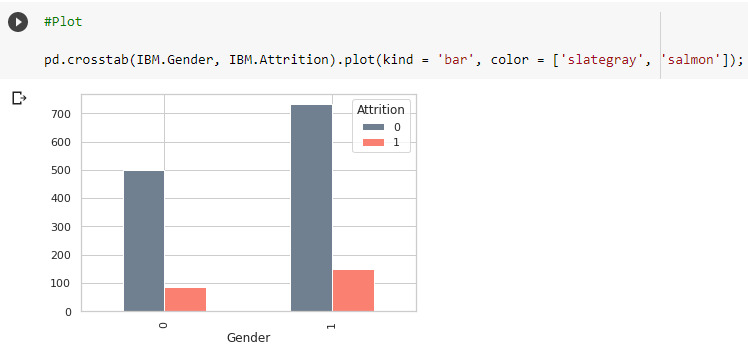
****

****

**Remarks :** **13.84% of the employees working in the Research & Development department have left the company while 20. 62% of the employees from Sales department have also left. From Human Resources, 19.05% of the employees have left the company.**

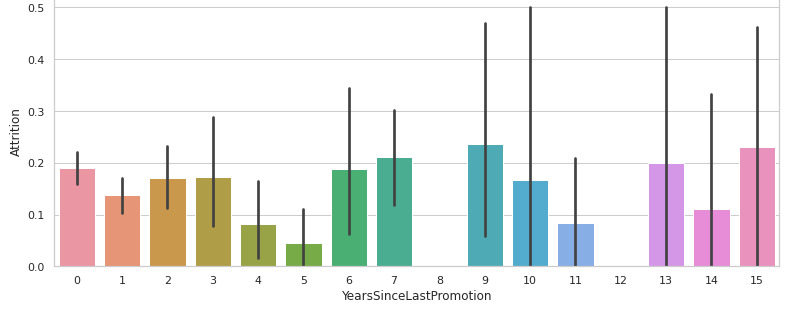
**Gender :**

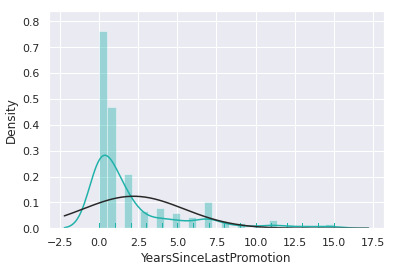
****

****

**Remarks : We can see that the company has a higher number of male employees than female. We can also see that the number of male employees leaving the company is higher than the female employees leaving the company.**

**YearsSinceLastPromotion :**

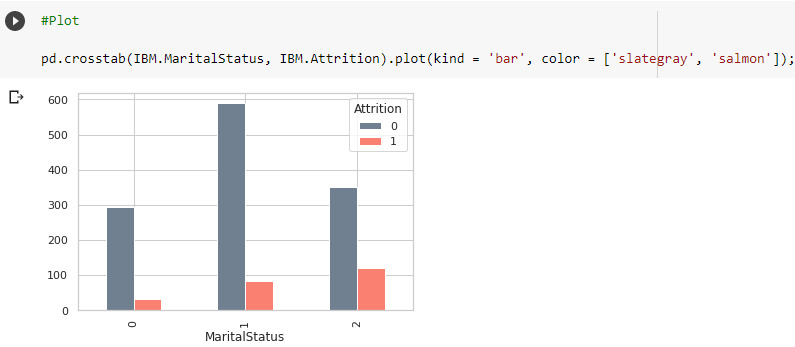
****

****

**Remarks : YearsSinceLastPromotion column is highly right skewed and it has a lot of outliers. The average years before an employee gets a promotion is roughly 2. Both the employees who are still working in the company and those who have left have an average of 1 year since their last promotion. We will have to treat the data for skewness and outliers before training the model.**

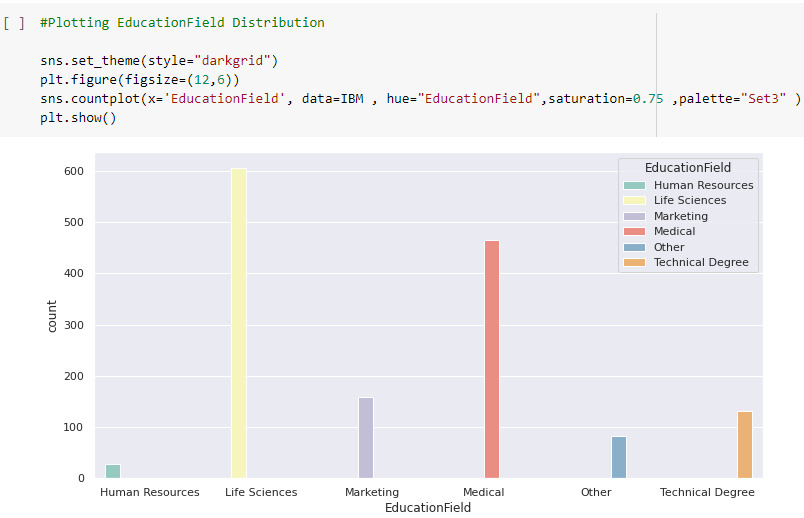
**MaritalStatus :**

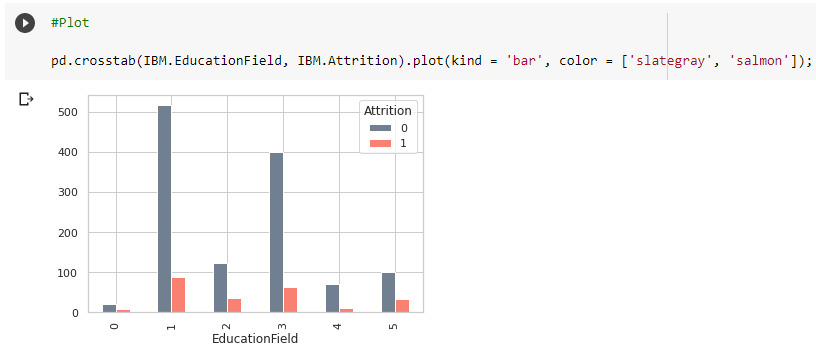
****

****

**Remarks : More than 45% of the employees are married, of those only 12.5% employees leave the company. On the other hand, more than 25% of the employees that are not in any committed relationship and over 10% of the ones who are divorced leave the company. From this we can conclude that employees who are single are more likely to leave the company.**

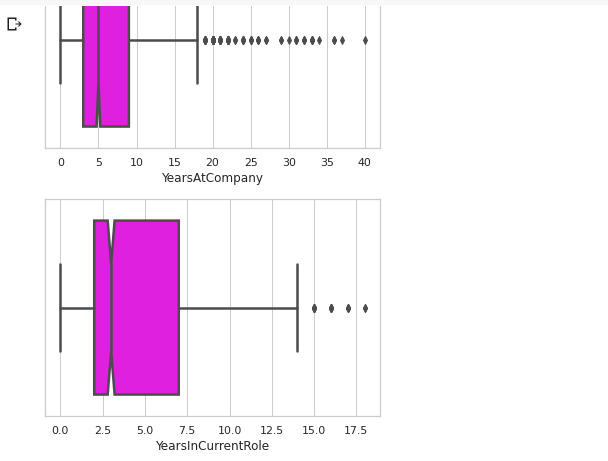
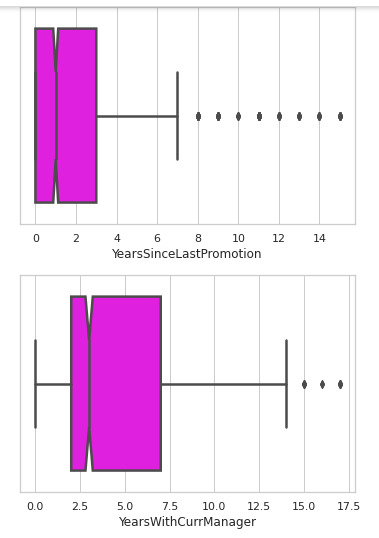
**Education\_Field :**

****

****

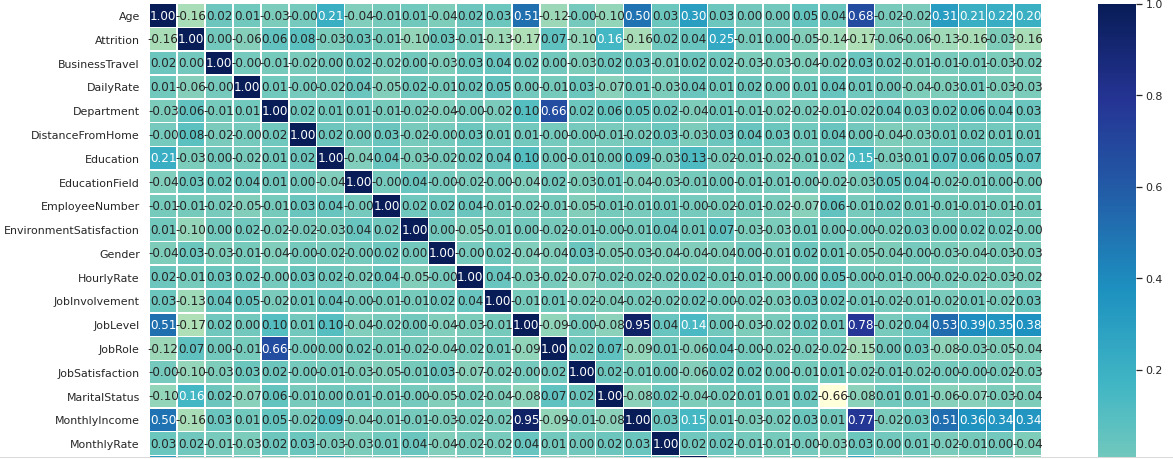
**Remarks : The employees listed in the dataset have various education backgrounds. Majority of the employees have studied Life Sciences. From the above graphs we can see that employees of Life Science education background are more likely to stay with the company and employees with a technical degree are more likely to leave the company.**

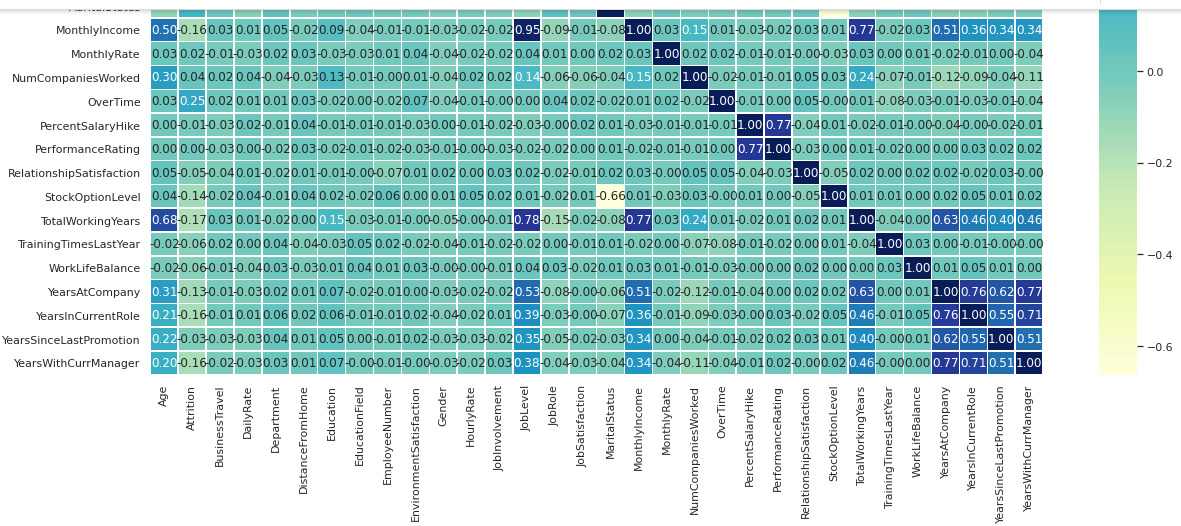
**Outliers :**

**** ****

**Remarks : We use boxplots to plot the outliers in the numerical columns of the dataset. From the below plots we can see that ‘MonthlyIncome’, ‘YearsSinceLastPromotion’, ‘YearsAtCompany’, ‘TotalWorkingYears’, and ‘YearsWithCurrManager’ columns have outliers. We will have to treat these outliers before building the predictive model.**

**Correlation\_Matrix :**

****

****

**Remarks : Features which have strong correlations:**

**Percent Salary Hike and Performance Rating, Total Working Years, Monthly Income and Job Level, Years at Company, Years with Current Manager, and Years in Current Role,**

**Features which have moderate correlations:**

**Age has moderate correlation with Total Working Years, Monthly Income, and Job Level, Job Level has moderate correlation with Years at Company and Age, Total Working Years has moderate correlation with Years with Current Manager, Years Since Last Promotion, Years in Current Role, Years at Company, and Age, Years at Company has moderate correlation with Years Since Last Promotion, Total Working Years, Monthly Income, Job Level, Years in Current Role has moderate correlation with Years Since Last Promotion, Total Working Years, Years Since Last Promotion has moderate correlation with Years with Current Manager, Years in Current Role, Years at Company, Total Working Years, Years with Current Manager has moderate correlation with Years Since Last Promotion, Total Working Years.**

# Preprocessing Pipeline :

# LabelEncoder:

# C:\Users\lenovo\Pictures\Blogs_Pic\58.jpg

# Checking for Outliers :

# C:\Users\lenovo\Pictures\Blogs_Pic\61.jpg

# Removing Skewness

# C:\Users\lenovo\Pictures\Blogs_Pic\62.jpg

# Split Data and Target Variable

# C:\Users\lenovo\Pictures\Blogs_Pic\63.jpg

# Feature Scalling

# C:\Users\lenovo\Pictures\Blogs_Pic\64.jpg

# Treating Class Imbalance Problem

# C:\Users\lenovo\Pictures\Blogs_Pic\65.jpg

# Finding Best Random State

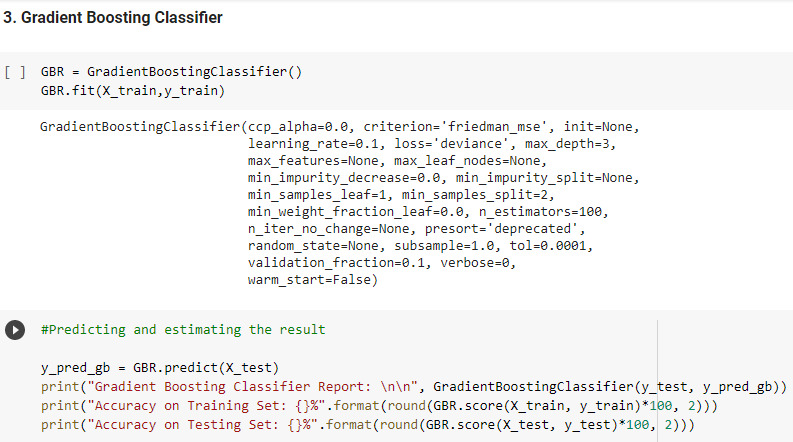
# C:\Users\lenovo\Pictures\Blogs_Pic\66.jpg

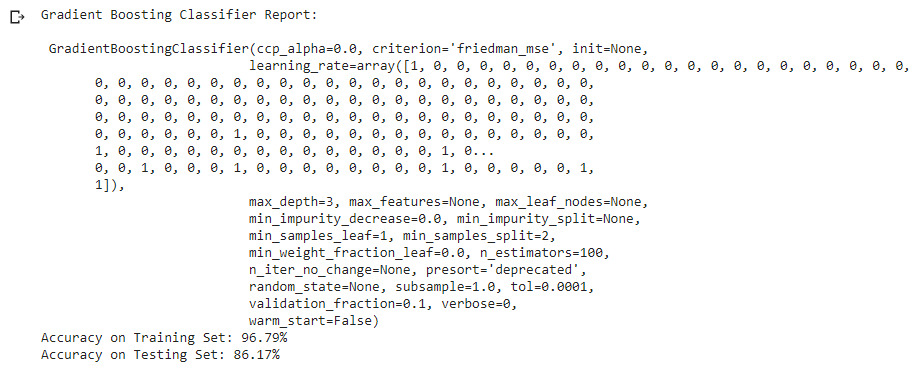
# Splitting the Dataset with train and test

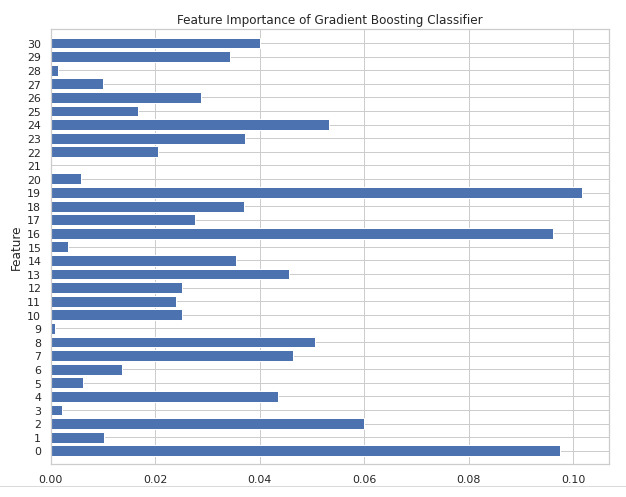
# C:\Users\lenovo\Pictures\Blogs_Pic\67.jpg

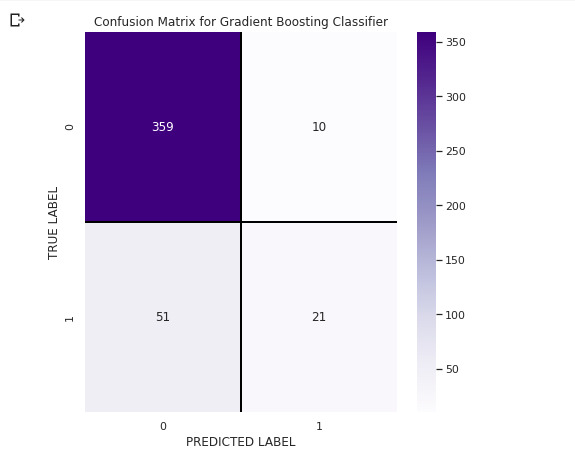
# Building Machine Learning Models

**We applied many Machine learning Models such as DecisionTree,LogisticRegression,SVM,AdaaBoost,GradientBoosting,GaussianNB etc where in I got the least difference between accuracy\_score(86.17%) and CVS(86.12%) i.e, (0.05) in Gradient Boosting Classifier . Hence it is our Best Fitted Model for ML .**





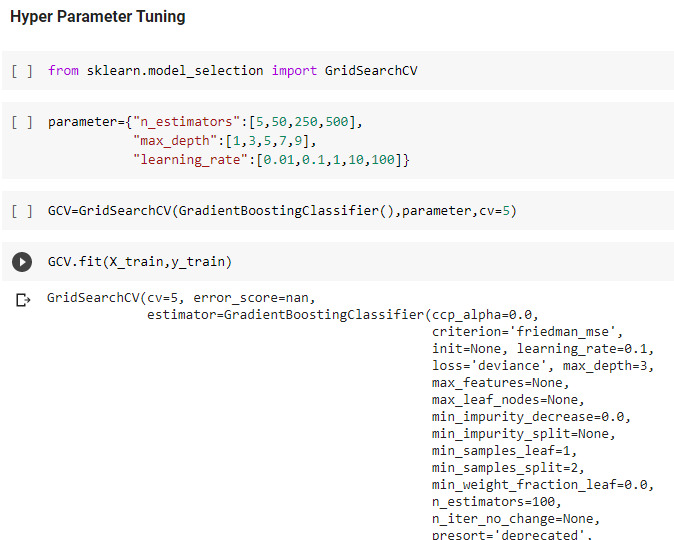
****

****

# C:\Users\lenovo\Pictures\Blogs_Pic\72.jpg

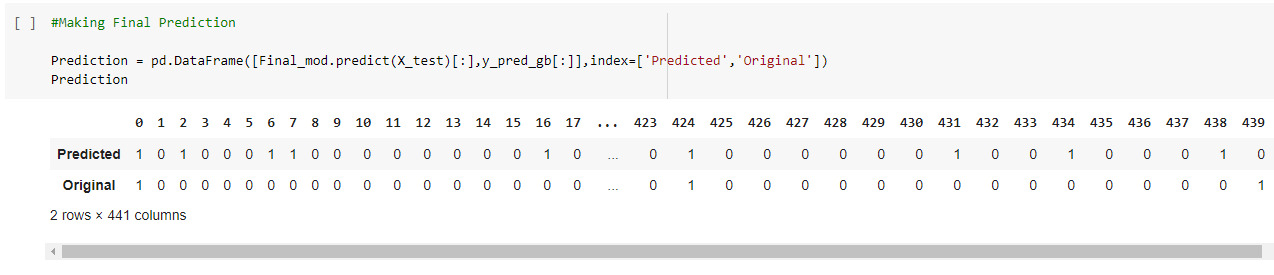
**Hyperparameter Tuning**

**After getting the scores for baseline algorithms, we perform hyperparameter tuning on model which give the best scores to improve the accuracy and reduce overfitting or underfitting. We use RandomizedSearchCV for hyperparameter tuning for the models. RandomizedSearchCV is used as we consider more than one parameter for tuning and it gives results faster than GridSearchCV.**

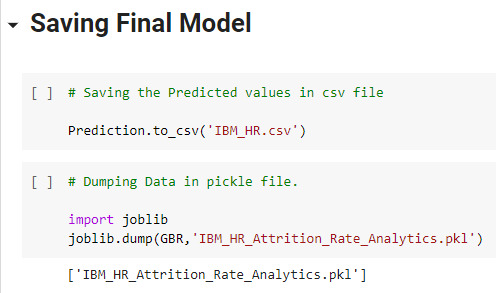
****

****

**Making Final Prediction**

****

**Saving Final\_Model**

****

# Conclusion

# As new entries are delivered to the dataset and the dataset grows, we are able to retrain the model with the brand new records to get greater correct predictions. On getting higher predictions we will flow directly to in addition classify the personnel who're maximum probable to leave the agency, people who can also or won't depart the organization and those who're least in all likelihood to go away the corporation. This facts also can be used while hiring new personnel to reduce the attrition rate.

# The most applicable capabilities that indicate humans leaving are:

# I. Monthlyincome: employees with lower wages are more likely to go away the corporation. Consequently, the corporation can provide regular bonuses and incentives for those employees.

# Ii. Distancefromhome: employees who stay in the direction of the place of business are much less in all likelihood to leave the enterprise compared to people who stay farther away. So, the company ought to provide proper communal transport for the employees who live far from the administrative center to make touring easier and more convenient.

# Iii. Age: personnel underneath 30 years of age are more likely to leave the organization. Screening the personnel of their age might be considered as discrimination, so the business enterprise can paintings in the direction of making the workplace friendlier toward that age group as opposed to turning them away.

# Iv. Additional time: employees who work overtime are much more likely to depart the enterprise. Therefore, the agency can paintings in the direction of having right working hours for all the employees and sufficient workforce to lessen overtime as a whole lot as feasible.

# V. Yearswithcurrmanager: personnel which have labored below the modern manager for much less than 2 years are most possibly to go away the employer. To keep away from that, the corporation need to strive now not to change managers very frequently.

# Similarly to the above listed tips, a through interview with the employee before hiring and know-how of preceding jobs will assist in reducing the attrition price.