

**PROJECT TITLE:- HR ANALYTICS PROJECT**

****

**SUBMITTED BY :-**

**VISHNU RATHORE**

**ACKNOWLEDGMENT**

I would like to express my special thanks of gratitude to “**Data trained**” that made this project possible. I would like thank to all **Data trained technical team** for his guidance in building this project. I would also like to thank Data Trained Institution to making me capable of making proper decisions in the field of damascene and Machine learning.

**INTRODUCTION**

**What HR analytics?**

**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.

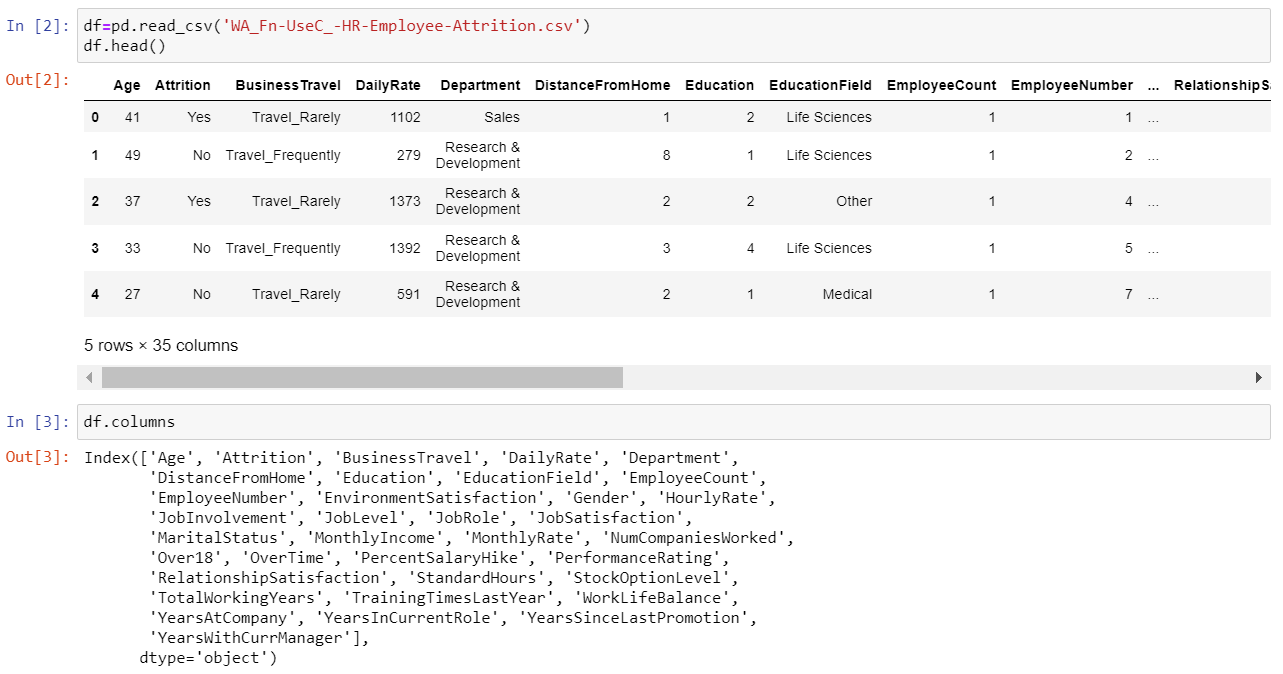
### **Data Description**

**1. Age:** Age of the employee in numeric.  
**2. Attrition:** Attrition of employee categorical : Yes/ No.  
**3. Business Travel:** It is a categorical column about employee business travel : Travel Frequently/ Rarely Travel/ Non- Travel.  
**4. Daily Rate:** It is a numerical column signify daily working rate of employee.  
**5. Department:** It is a categorical column signify employee's working department.  
**6. Distance From Home:** It is a numerical column signify total distance of employee's home from office.  
**7. Education:** It is a categorical column signify level of employee's education : (1 'Below College', 2 'College', 3 'Bachelor', 4 'Master', 5 'Doctor')  
**8. Education Field:** It is a categorical column signify education field of employees.  
**9. Employee Count:** Number of employees in numerical. It has only one unique value i.e., 1  
**10.Employee Number:** It is a unique numerical Id assigned to every employee.  
**11. Environment Satisfaction:** It is a categorical column having following values : (1 'Low', 2 'Medium', 3 'High', 4 'Very High')  
**12. Gender:** It is a categorical column having only 2 values Male and female.  
**13. Hourly Rate:** It is a numerical column signify hourly working rate of employee.  
**14. Job Involvement:** It is a categorical column having following values : (1 'Low',2 'Medium',3 'High',4 'Very High')  
**15. Job Level:** It is a categorical column signify level of job i.e., 1,2,3,4 and 5  
**16. Job Role:** It is a categorical column signifying role of employee i.e., sales executive, research scientists etc.  
**17. Job Satisfaction:** It signifies how satisfied an employee is with the job : (1 'Low',2 'Medium',3 'High',4 'Very High')  
**18. Marital Status:** It is a categorical column signifies marital status of employee i.e., Married, Single, Divorced etc.  
**19. Monthly Income:** It is a numerical column signify monthly income of the employee.  
**20. Monthly Rate:** It is a numerical column signify monthly rate of working by employee.  
**21. No. Companies Worked:** It is a numerical column signify number of companies an employee previously worked for.  
**22. Over18:** It is a boolean field signify whether employee age is over 18 or not.  
**23. Over Time:** It is a boolean field signify whether employee work overtime or not.  
**24. Percent Salary Hike:** It is a numerical column signify percentage of hike received recently by an employee.

**25. PerformanceRating:** It is a categorical column signify performance rating of employee (1 'Low',2 'Good', 3 'Excellent', 4 'Outstanding')  
**26. Relationship Satisfaction:** It is a categorical column signify relationship bonding status with co workers (1 'Low',2 'Medium',3 'High',4 'Very High')  
**27. Standard Hours:** It is a numerical column signify standard working hours of employees.  
**28. Stock Option Level:** It is a categorical column. It has four categories 0,1,2 and 3  
**29. Total Working Years:** It is a numerical column signify total working years of employees.  
**30. Training Times Last Year:** It is a numerical column signify total number of times employee got training.  
**31. Work Life Balance:** It is a categorical column signify work life balance (1 'Bad',2 'Good',3 'Better',4 'Best')  
**32. Years At Company:** It is a numerical column signify total number of years spent by employee in current company.  
**33. Years In Current Role:** It is a numerical column signify total number of years spent by employee in current job role in current company.  
**34. Years Since Last Promotion:** It is a numerical column signify total number of years since last promotion of employee in current company.  
**35. Years With Current Manager:** It is a numerical column signify total number of years employee work under current manager in current company.

**1.Data Used**

* Data was used from Kaggle which is a freely available platform for data scientists and machine learning enthusiasts.
* Source: <https://github.com/dsrscientist/IBM_HR_Attrition_Rate_Analytics>
* We are using Jupyter-notebook to run **HR ANALYTICS PROJECT**.



## 2. Data Analysis

The procedure of extracting information from given raw data is called data analysis. Here we will use **EDA** module of **DATA PREPRATION**  library to do this step.

**What is Exploratory Data Analysis ?**

Exploratory Data Analysis or (EDA) is understanding the data sets by summarizing their main characteristics often plotting them visually. This step is very important especially when we arrive at modeling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plot, Scatter plot and many more. It often takes much time to explore the data. Through the process of EDA, we can ask to define the problem statement or definition on our data set which is very important.

## Importing the required libraries for EDA

## 

### **Features and Targets**

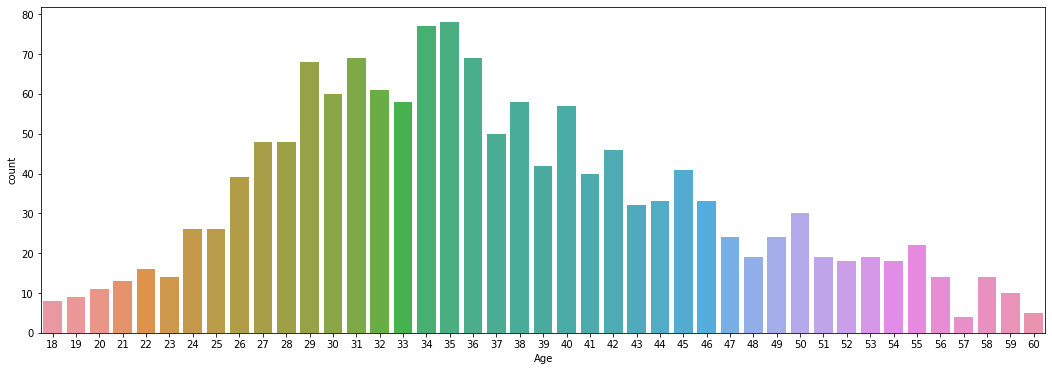
****

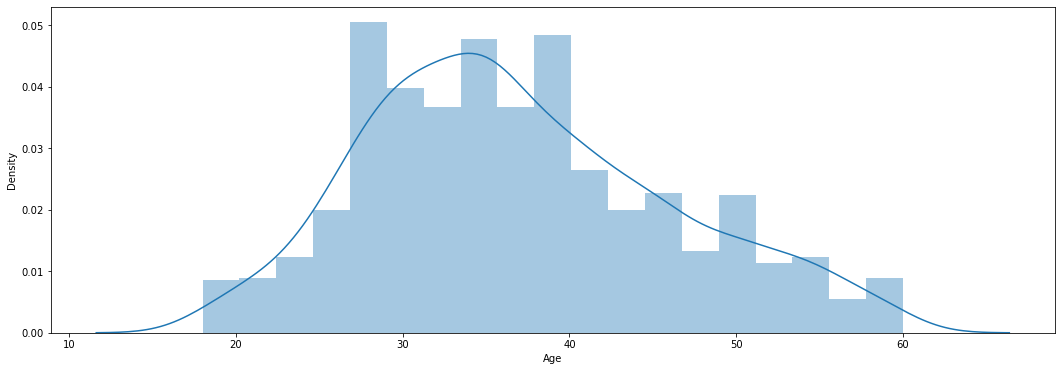
**Goal:-**

* We have to find the if the employee is likely to leave the company based on the historical data of the employees of a company
* **Attrition** in human resources refers to the gradual loss of employees over time.

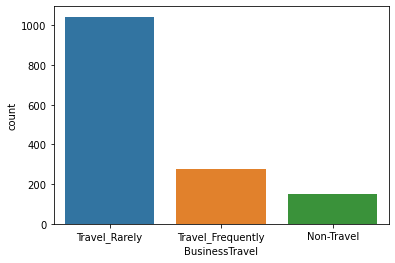
### **Variables**

* **Age:-**

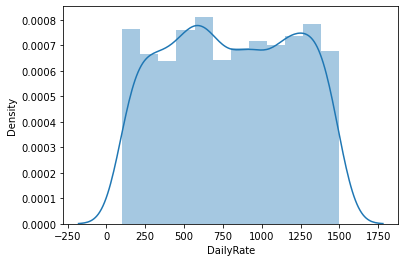
****

****

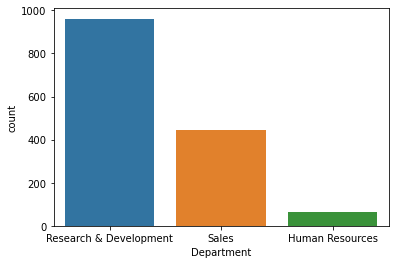
* Most of the employee in the company are from the age group of 26 to 38
* The minimum age required to work in the company is 18yrs.
* No workers are from the over 60 years of age that is from the senior citizen category.
* **Business Travel**

****

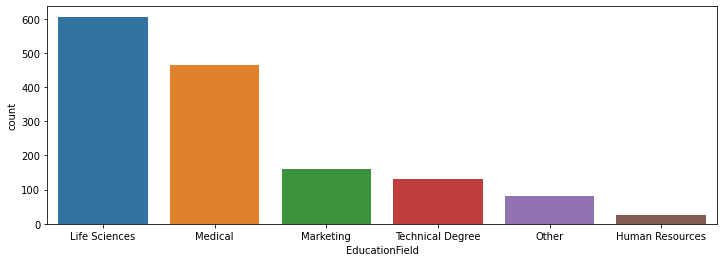
* Most of the people in the company travel rarely but there are very few non travellers.
* The frequent travellers only make up about 18.85 percent of the total employees.
* **Daily rate**

****

* Daily rate has a pretty uniform distribution.
* A day rate is the cost of an individual's work for a single day.
* **Department**

****

* The Company has more than 65 percent of employees in the R&D department.
* Every company invests the maximum in Research and Development Departments.
* Sales and HR makes the other 35 percent of the company employees.
* **Education field**

****

* Most of the employees belong to life sciences , medical.
* Very few employees belong to human resources.

## Model Building

## 1.Split dataset into train and test set in order to prediction w.r.t X\_test

## 2.If needed do scaling of data

## 3.Import model

## 4.Fit the data

## 5.Predict w.r.t X test

## 6. Classification report

## 7.Hyper Parameter Tuning for better score

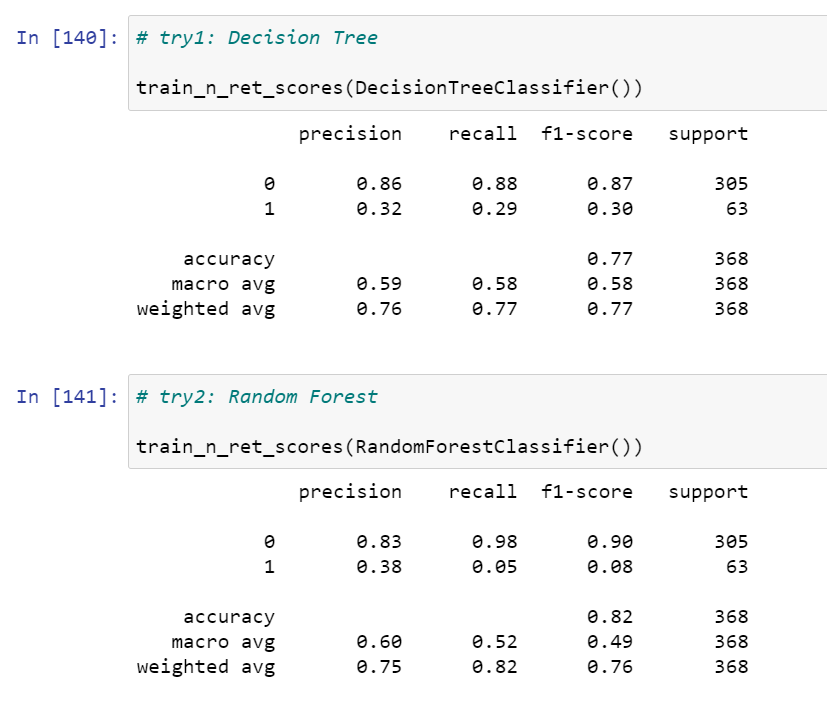
## SCALE

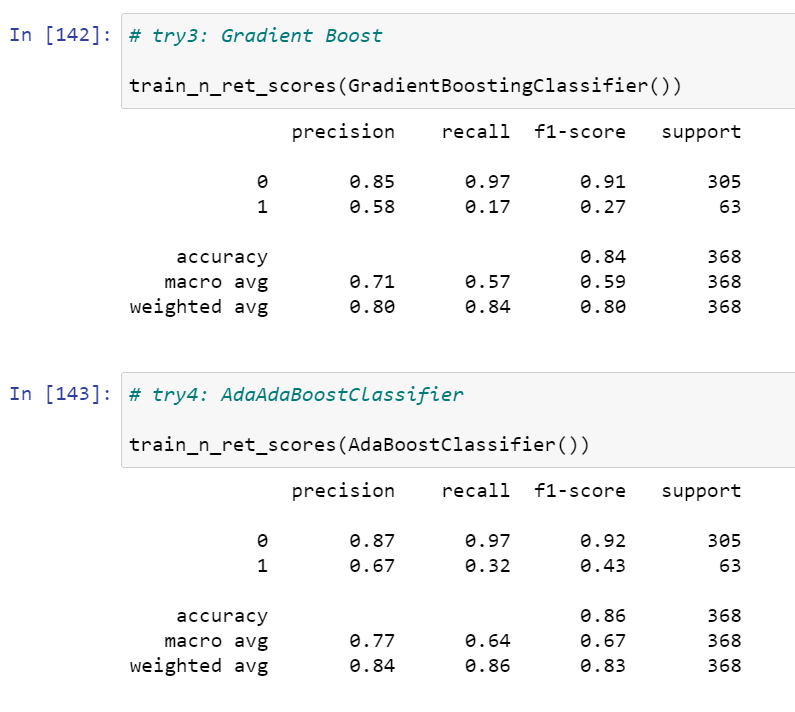
## 

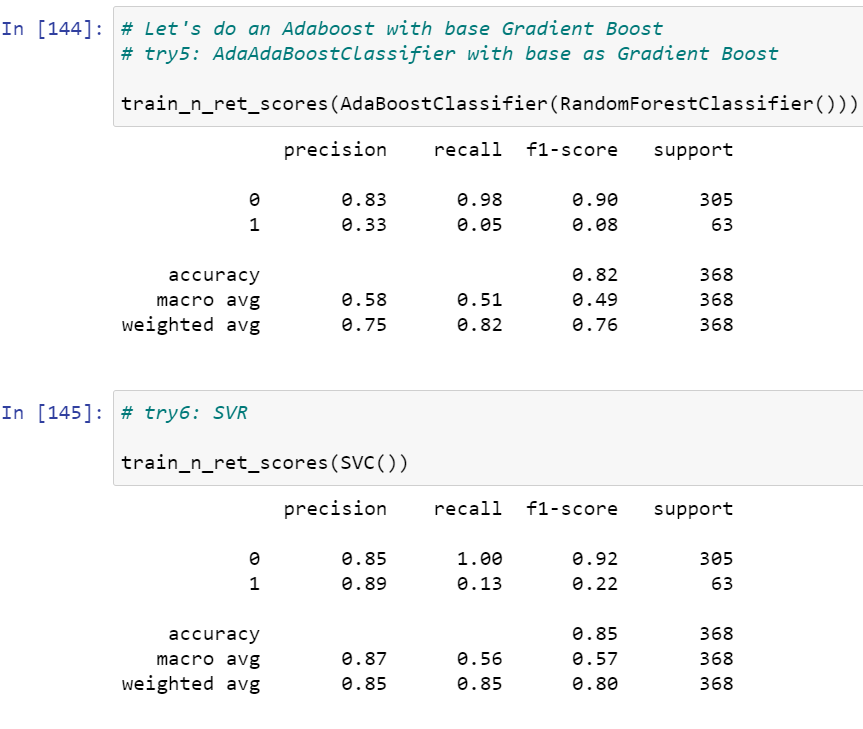
* **Model Building**

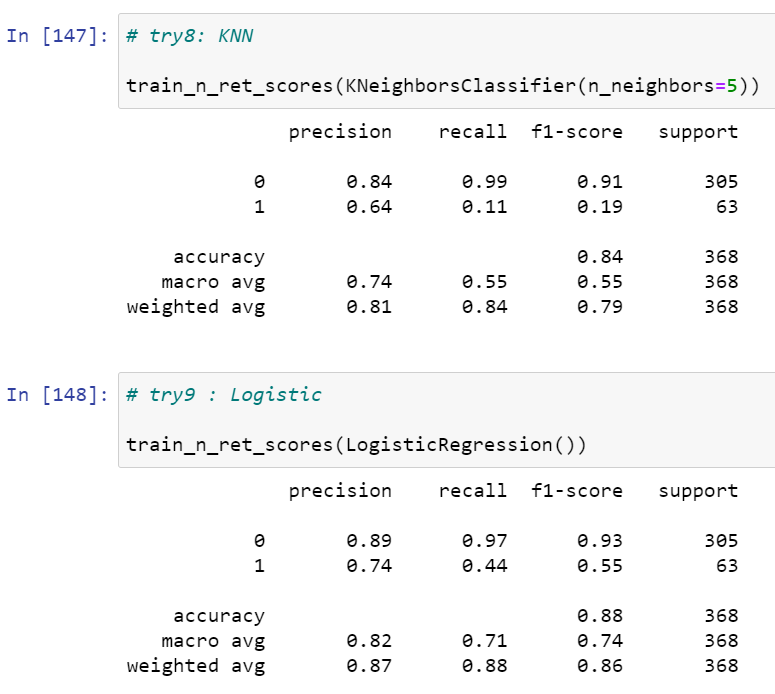
****

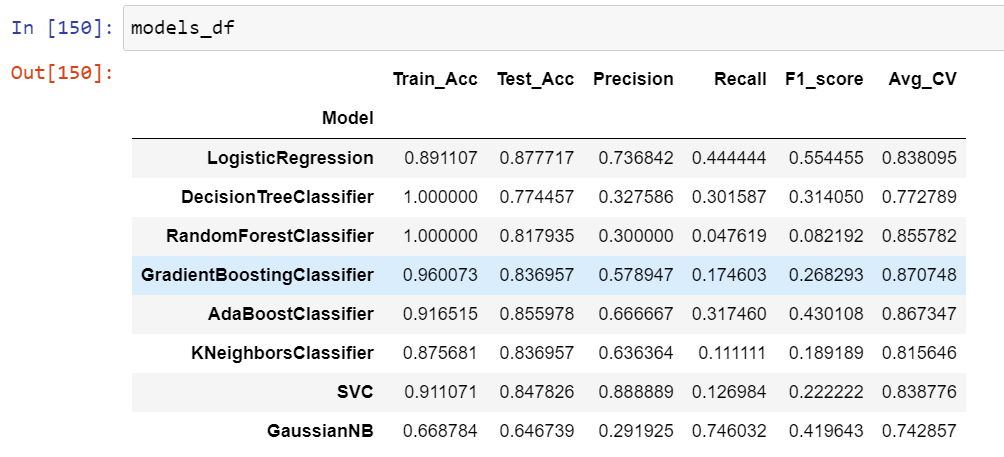
* **Fit & Predict the Data**

****

****

****





* **Gradient boosting classifier** gives best result and score .
* **Conclusion**

Employee satisfaction, years At Company, and evaluation were the three biggest factors in determining turnover.

1. Employees with low to medium salaries are more likely to leave the company. It is observed that the employees with high salaries are least likely to leave the company.
2. The employee turnover is high in the departments Sales, technical, Support. The Management department has the least employee turnover.
3. The employee turnover is high for employees with 2,6 and 7 projects. Majority of the employees with 3,4,5 projects did not leave the company. All the employees with 7 projects left the company.
4. Employees with low and high performance tend to leave the company more. The sweet spot for employees that stayed is within 0.6-0.8 evaluation.
5. Employees who had less hours of work (150hours or less) left the company more. Employees who had too many hours of work (250 or more) left the company Employees who left generally were underworked or overworked.
6. Employees who had really low satisfaction levels (0.2 or less) left the company more Employees who had low satisfaction levels (0.3~0.5) left the company more Employees who had really high satisfaction levels (0.7 or more) left the company more.
7. More than half of the employees with 4 and 5 years left the company.
8. Looks like employees who did not leave the company had an average evaluation of around 70% even with different project Counts. There is a huge skew in employees who had a turnover though. It drastically changes after 3 project Counts. Employees that had two projects and a horrible evaluation left. Employees with more than 3 projects and super high evaluations left.

* **Reference**
* Data collection from [www.kaggle.com](http://www.kaggle.com)
* <https://www.analyticsvidhya.com>
* Data trained LMS video.