HR Analytics Project- Understanding the Attrition in HR

**Introduction**

Attrition has become a critical concern in terms of a company's competitive advantage in recent years. Finding, hiring, and training new employees is quite costly. It is more cost-effective for a corporation to keep its current staff. To keep their staff for a longer amount of time, a corporation must maintain a nice working environment. It was done manually a few years ago, but now it is done by the emerging trend using artificial intelligence and data analysis.

**Logistic Regression in Analysis**

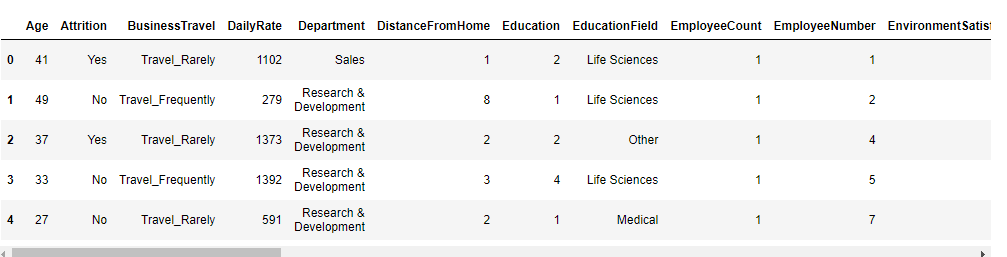
Whether an employee will stay or quit a company, his or her response is binary, i.e. it can be either "YES" or "NO." As a result, we can see what our dependent variable is. Employee attrition is merely a categorical variable. We can't utilise linear regression with a dependent categorical variable; instead, we have to utilise "LOGISTIC REGRESSION."

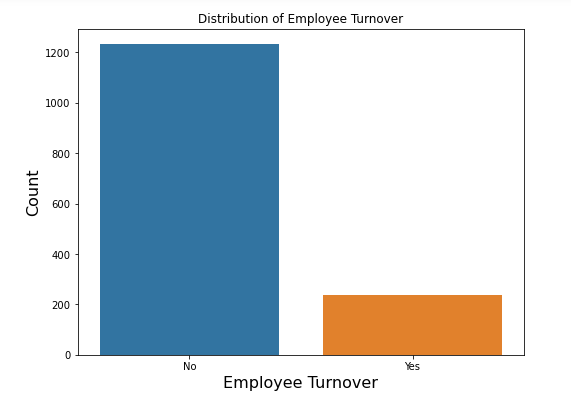
**Steps in this method**

* Data collection
* Data preprocessing
* Dividing the data in to testing and training
* Building the model using training data
* Testing the accuracy on th testing data

**Exploratory Data Analysis**

The dataset contains 1470 observations and 35 variables. Within 35 variables “Attrition” is the dependent variable.

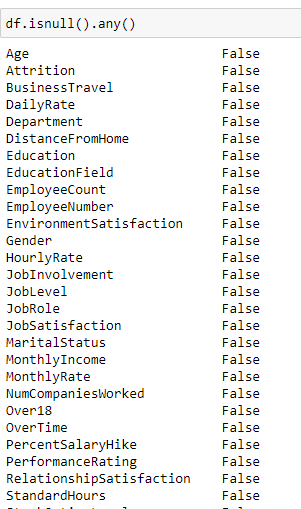




Take a look on the attrition of employees ,number of counts are imbalanced

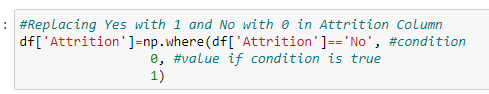
**Data Preparation**

Checking the null are present in the datasets



There are no null values present in the dataset so it shows ‘false’

**Changing the datatypes**



the data type of the dependent variable “Attrition”. It is given as “Yes” and “No” form i.e. it is a categorical variable. To make a proper model we have to convert it into numeric form. To do so, we will assign value 1 to “Yes” and value 0 to “No” and convert it into numeric.

Splitting the datasets in to ‘Train’ and ‘Test’ data

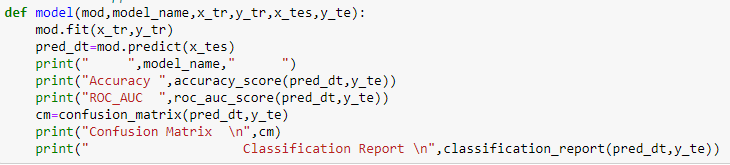
With the help of the Training data set we will build up our model and test its accuracy using the Testing Data set.

**Building the model**

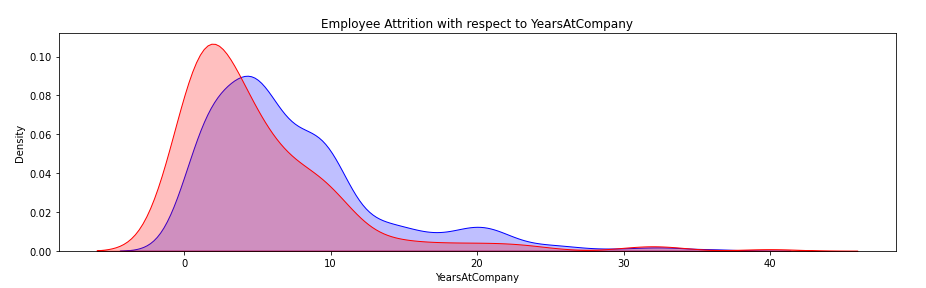
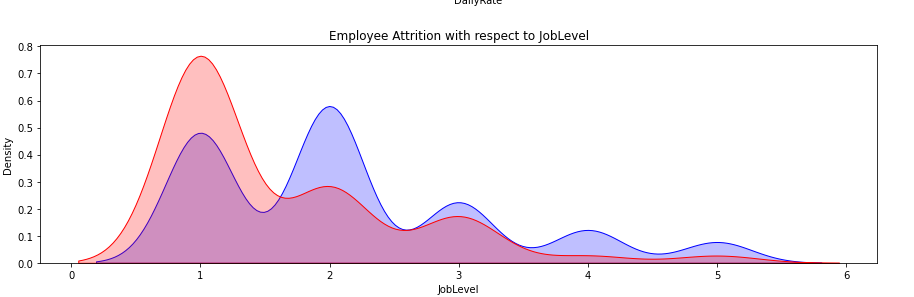
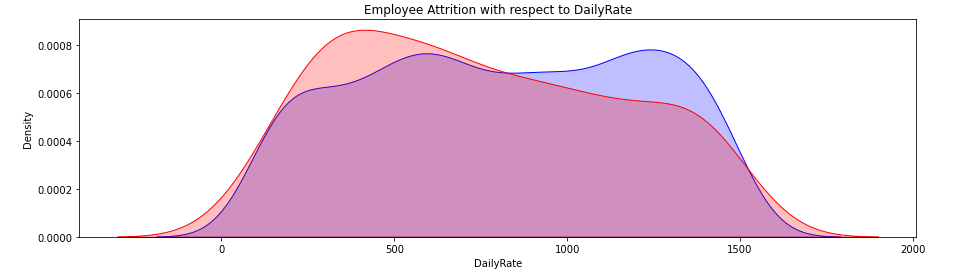
Identifying the independent variables,incorporate the dependent variable ‘Attrition’ in the model .Transform the datatypes in to formula and then split the train and test data







**Visualization of how attrition is affected**



From our above result we can see, Business travel, Distance from home, Environment satisfaction, Job involvement, Job satisfaction, Marital status, Number of companies worked, Over time, Relationship satisfaction, Total working years, Years at the company, years since last promotion, years in the current roleall these are most significant variables in determining employee attrition.

**Conclusion**

This model provides the best fit with accuracy 85% better than the logistic regression which obtained 84% .