**Project: HR Attrition**

**Problem Statement:** 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.

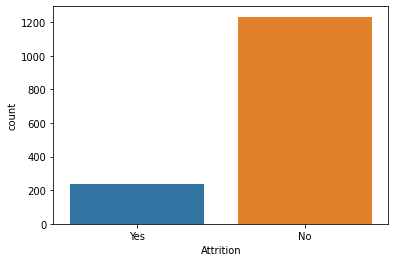
**Attrition affecting Companies**

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.

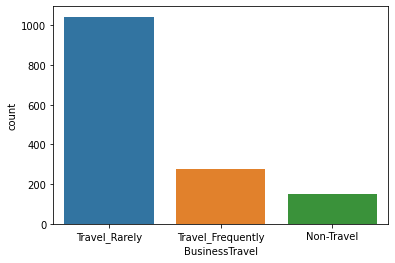
.

**Data Analysis**

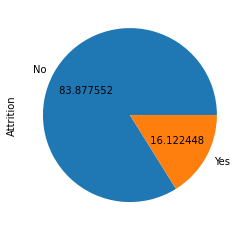
1. The Shape of the data set is 1470 rows and 35 Columns
2. When the null values is checked using isnull( ) then It was found that there was no null values
3. There are no duplicated values also
4. In below graph we can easily see that Attrition (yes) is less when compared to (No) this means less number of employees are leaving the organization.



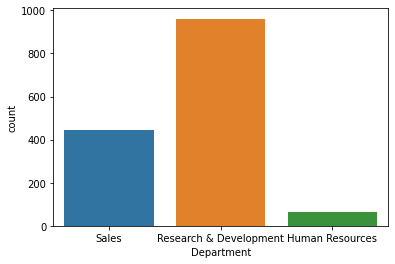
1. In the below graph we can see that the number of employees who travel rarely is more



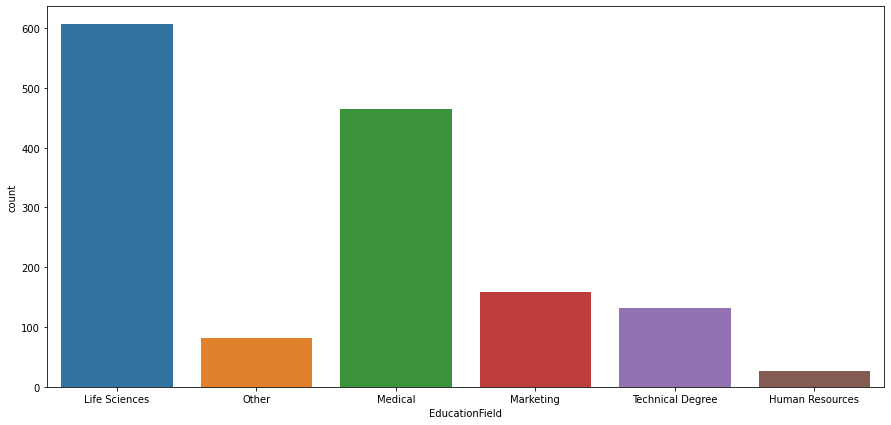
1. In the below pie chart is clearly show that the employees who are not leaving the organization is almost 84% when compared to employees who are leaving the organization i.e. almost 16 %



1. In the below graph we can note that most of the employees are from research and development department



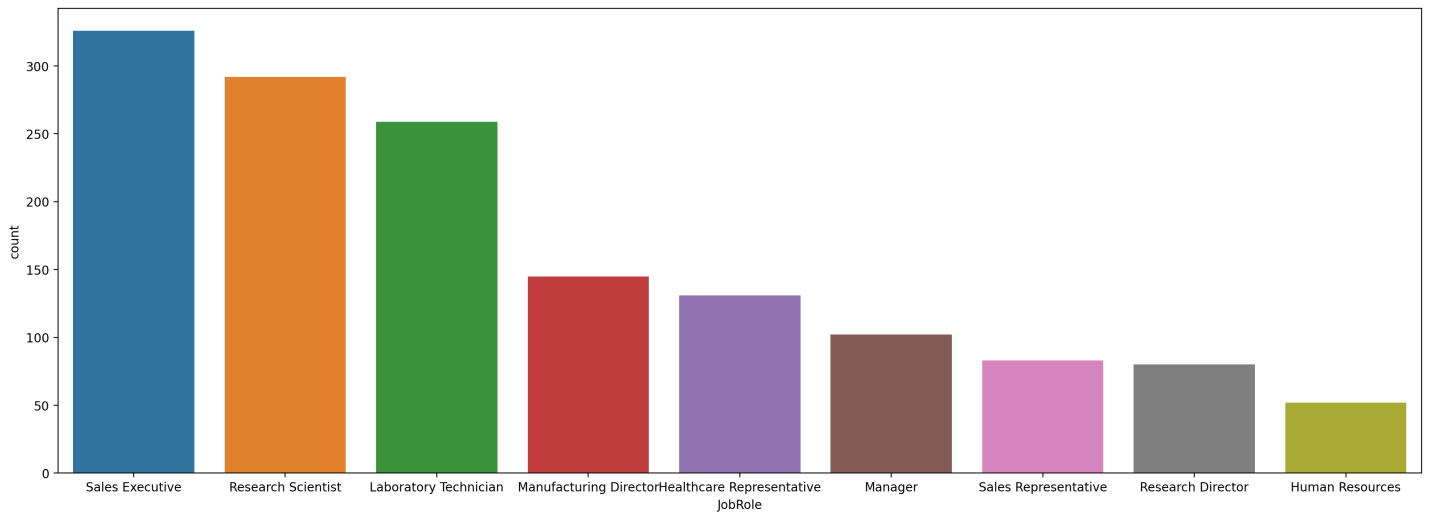
1. With the below graph we can know that most of the employees are from Life Sciences and Medical field



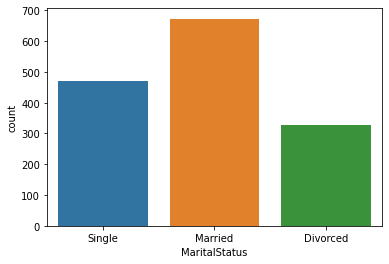
1. With the below graph we can note that most of employees are Male

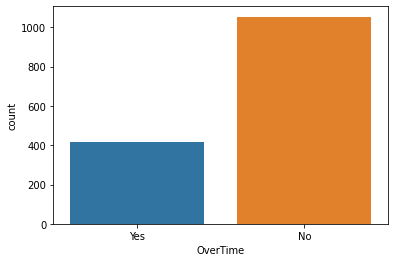


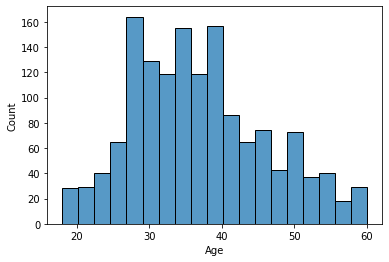
1. With the above graph we can note that most of the employees are Sales Executive and Research Scientist



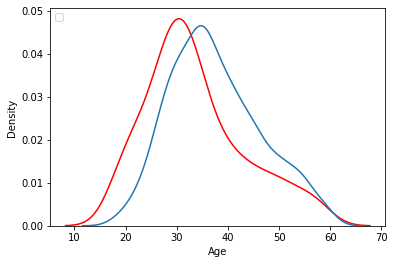
1. With the blow graph we can easily know that most of the employees are married



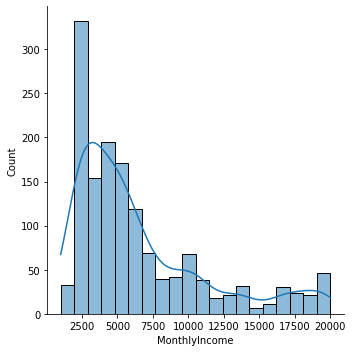
1. With the below graph we can note that most of the employees does have overtime
2. We can note that in below The age is between 18 to 60 of the employees and most number of employees are between 25 to 40



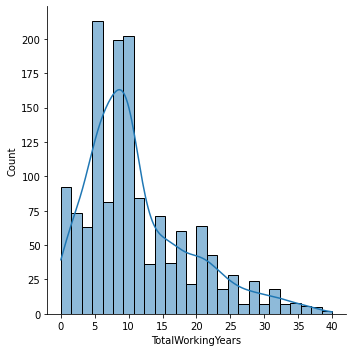
1. With the below graph we can note that employees with age around 30 is leaving the organization when compared to employees with age around 50



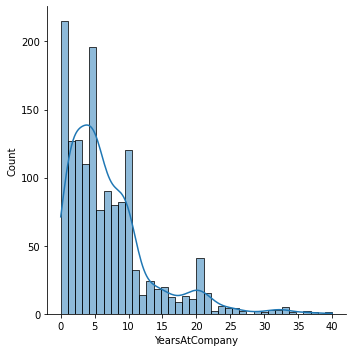
Red= Attrition: yes, Blue = Attrition=N0

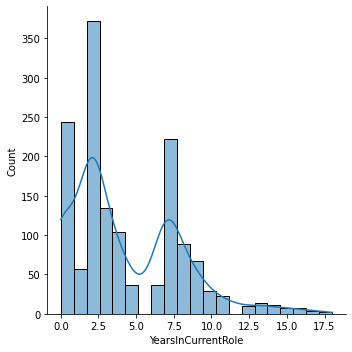
1. We can note in the below graph that most of the employees monthly income in the range 2500 to 3000

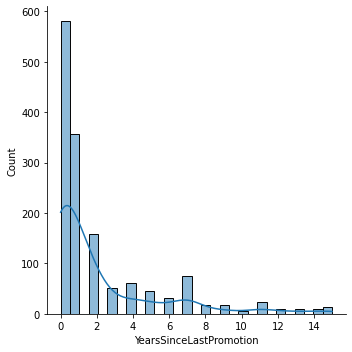
1. Here we can note that most of the employees are 5 years of experience



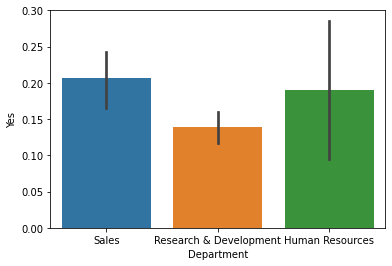
1. With the below graph can note that most of the employees are new to the company

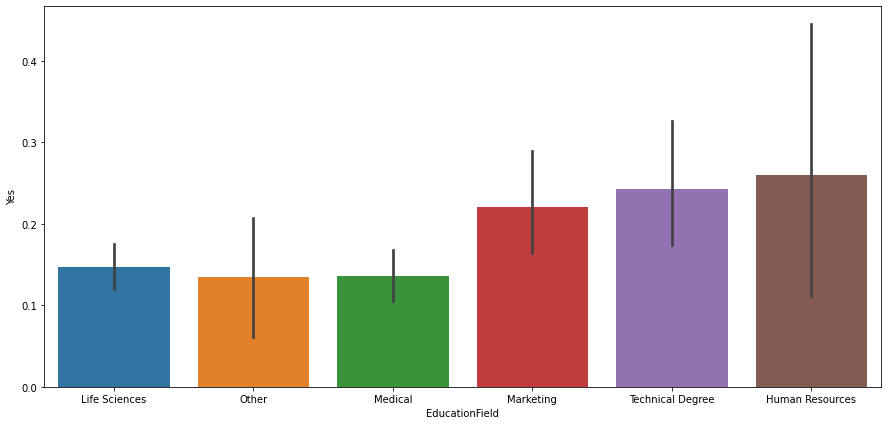


1. In the below graph we can note that most of the employees around 400 are in the same role since 2.5 years
2. Here we can note that most of the employees are getting not getting promotion

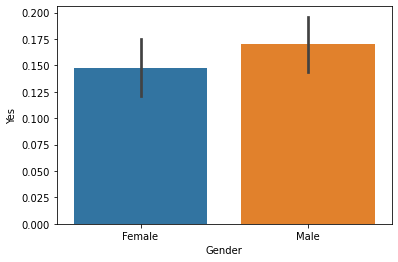


1. With the below Graphs we can understand that Human Resources is been attired more when compared to others

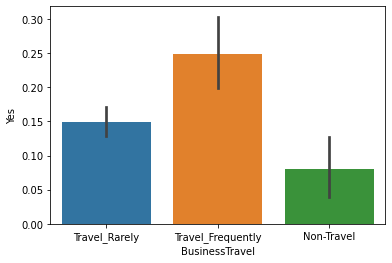




1. With the below we and understand that Male is been attired more when compared to female



1. In the similar Manner Employees that travel frequently tend to leave more the company.



**EDA Conclusion:**

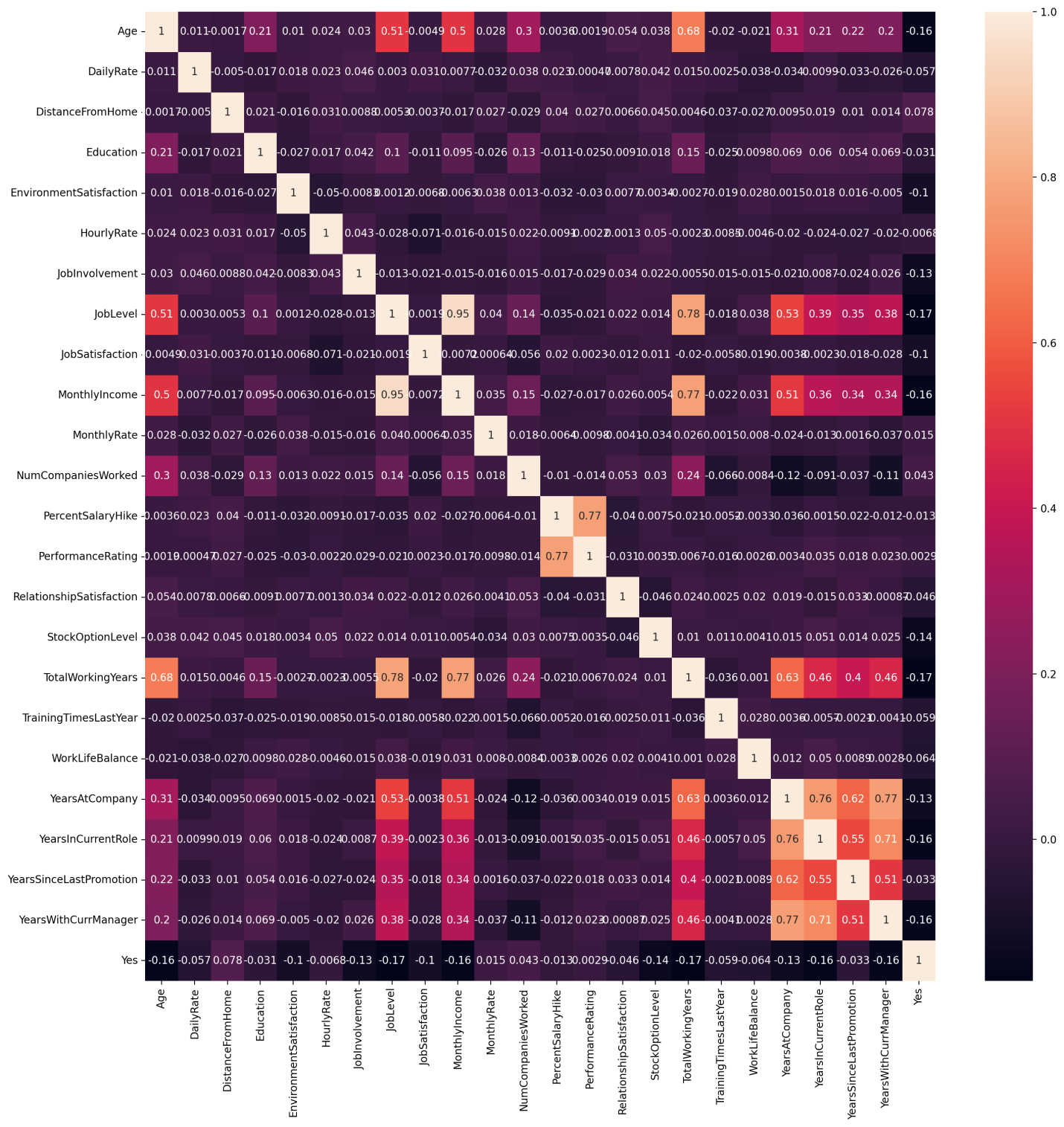
With EDA we had the insight of the data and we found that Human resource has department has more attrition when compared to others and as per the analysis we also found out that employees we not getting promotion and in the same role since 2.5 years and monthly income of most of the employees are less i.e. 2500 to 3000

**Pre processing**

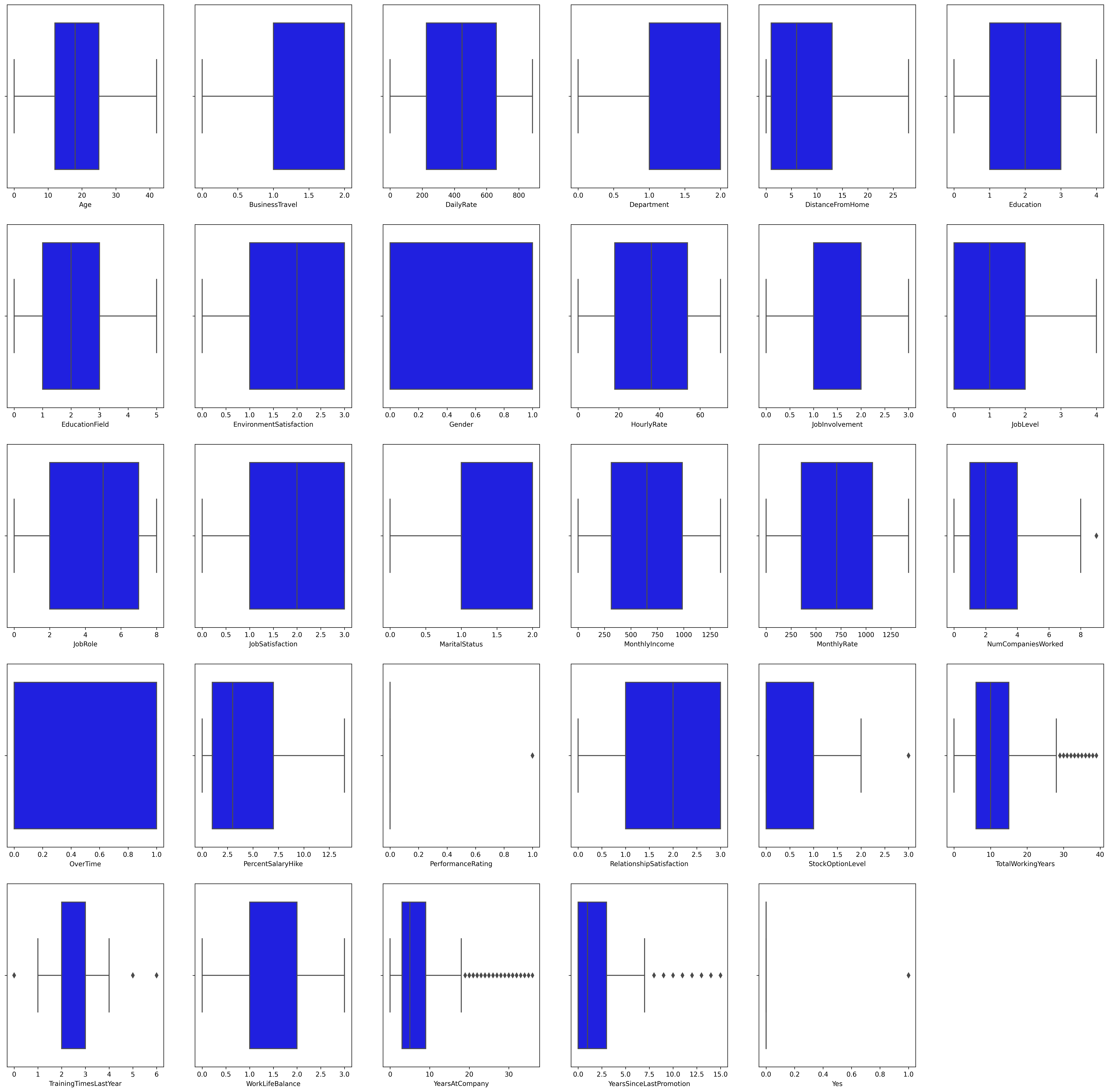
1. In pre processing step we have dropped 1) EmployeeNumber 2)Over18 3) StandardHours 4)EmployeeNumber since these columns are note that useful
2. **Checking the Correlation:** After checking the correlation, multicollinearity

We have dropped the following columns

1. **YearsAtCompany**
2. **YearsWithCurrManager**

****

1. **Handling Categorical Values:** I have used Label Encoder to handle Categorical Values
2. **Checking the Outliers :** We the above graph we can see that we have outliers so in this problem Statement I used IQR method to remove outliers



1. **Feature Scaling :** After removing outliers I have done feature scaling to bring the values on one scale for that I have used StandardScaler() from skelearn.
2. **Train test split:**  After Feature scaling I have used train\_test\_split from sklearn.model\_selection the split the data in training and testing sets.

**Building Machine Learning Models**

1. **Logistic Regression:** Algorithm is implemented using the linear\_model library of python-sklearn that has Logistic Regression model. After training the model we got accuracy of 86%
2. **Cross Validation of Logistic Regression** : After doing the cross validation of Logistic Regression got the score of 86.55%
3. **Hyperparmeter tuning using Grid Search CV:** After tuning the parameters using Hyperparameter tuning checked accuracy and it was found to be 85.75 %
4. **Random forest classifier** : Algorithm is implemented using the ensemble library of python-Sklearn,that has RandomForestClassifier.After training the model we got the accuracy of 86.01%
5. **Cross Validation of Random forest classifier :** After doing the cross validation of Random forest classifiergot the score of 84%
6. **Hyperparmeter tuning using Grid Search CV:** After tuning the parameters using Hyperparameter tuning checked accuracy and it was found to be 84.97%
7. **DecisionTreeClassifier:** Algorithm is implemented using the tree library of python-sklearn that has DecisionTreeClassifier model. After training the model we got the accuracy of 75.90%
8. **Cross Validation of DecisionTreeClassifier :** After doing the cross validation of Random forest classifiergot the score of 75.09%
9. **Hyperparmeter tuning using Grid Search CV:** After tuning the parameters using Hyperparameter tuning checked accuracy and it was found to be 83.41%
10. **Auc and Roc Curves to select the best model :**

|  |  |
| --- | --- |
| **Model Name** | **ROC** |
| **Logistic Regression** |  |
| **Logistic Regression with Hyper pameter tuning** | **logisticHyperparamter.png** |
| **Random Forest with Hyper Parameter tuning** | **RandomForestHyper.png** |
| **Random Forest Classifier** | **randomforest.png** |
| **Decision Tree Classifier** | **Descitiontree.png** |
| **Decision Tree Classifier with Hyper Parameter Tuning** | **DescitiontreeHyper.png** |

**Conclusion:**

In this Project we have detailed the various steps when implementing an analytics use case in HR, employee attrition.

We used various model like Logistic Regression, Random Forest Classifier, DecisonTree Classifier for training and obtained ROC curves. So the model Random Forest with Hyper parameter tuning is the best model which we have saved using pickle library.

**So concluding the project, male employees who travel frequently, work in Human Resource department, have a low salary hike, and live far from workplace have a high probability of leaving the company.**