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

**1.      Problem Definition:**

Attrition is a very familiar word often used in Human Resources (HR). But what is attrition and how important is it with respect to an organization? To understand the subject better, first we need to take a step back and understand about the hiring process that companies follow.

Organizations hire a lot of employees every year to meet their work requirements. In order to make these employees job ready, these companies invest a lot of time and money for their training. Even trainings are scheduled within the organization for existing employees as well. These training programs aim to increase the productivity and effectiveness of the employees. The job role of HR professionals is to assume a leadership role in designing company compensation programs, work culture, and motivation systems that assists the organization retain top employees. Still, a lot of employees tend to leave their job and shift to other companies.

Attrition in human resources refers to this gradual loss of employees over a period of time. Generally, a higher attrition rate is problematic for companies. The major problem that companies face is the cost attached to each employee. Some of the common expenses of losing employees and replacing them include job postings, hiring processes, documentation, and new hire training**.** Another problem for the companies would be an increase in errors if they have new workers constantly. To clearly understand and bring out a solution to this problem statement, the provided dataset is to be analysed. Here comes the role of 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.

**2.      Data Analysis:**

Now, comes the role of Data Analysis, which is a process of systematically applying statistical or logical techniques to describe, evaluate and illustrate data. It is also used to analyse patterns in observations through the entire dataset.

In the said project,HR-Employee-Attrition dataset is considered for analysis. As per the observation in this dataset, there are no missing values and the data types of the variables are integer and object type comprise 35 variables which might negatively or positively impact the attrition rate.

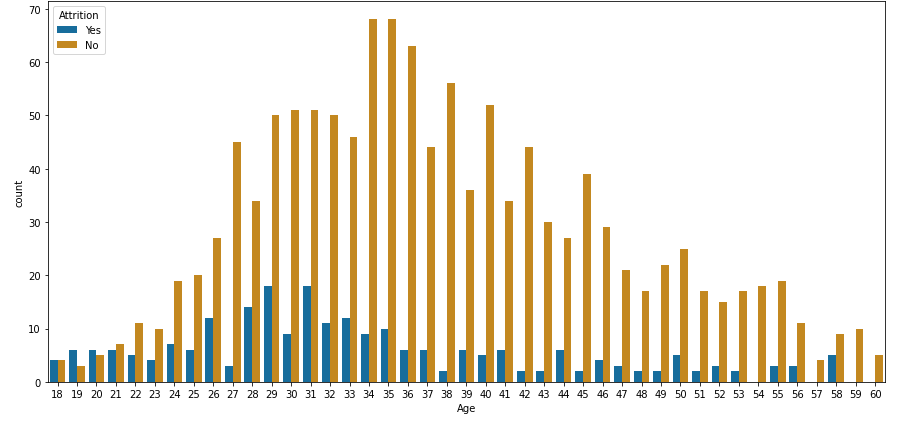
Below mentioned are the variables that are considered to impact the attrition rate:

* **Age**
* **Business Travel**
* **Daily Rate**
* **Department**
* **Distance from Home**
* **Education**
* **Education Field**
* **Employee Count**
* **Employee Number**
* **Environment Satisfaction**
* **Gender**
* **Hourly Rate**
* **Job Involvement**
* **Job Level**
* **Job Role**
* **Job Satisfaction**
* **Marital Status**
* **Monthly Income**
* **Monthly Rate**
* **Number of Companies Worked**
* **Over 18**
* **Overtime**
* **Percent Salary Hike**
* **Performance Rating**
* **Relationship Satisfaction**
* **Standard Hours**
* **Stock Option Level**
* **Total Working Years**
* **Training Times Last Year**
* **Work Life Balance**
* **Years at Company**
* **Years in Current Role**
* **Years since last promotion**
* **Years with Current Manager**

**3.    EDA Concluding Remarks:**

The observation and graphical representations of the dataset as follows:

**Impact of Age on Attrition rate**



Findings from the above observation are mentioned below:

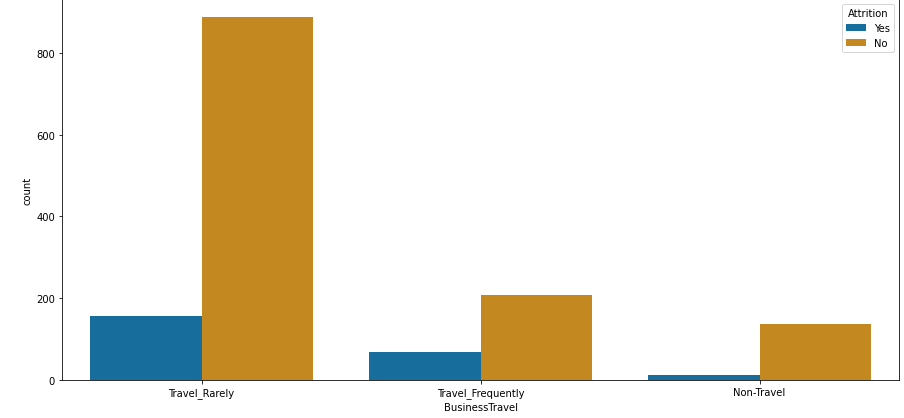
1) Age group from 18 to 21 has high no. of Attrition, but we can also observe that from age 22 it started decreasing.

2) Age group from 30 to 56 has very less Attrition.

3) There is no Attrition in 57 years however 58 years has good amount of Attrition.

4) 59 to 60 has no Attrition at all.

**Impact of Business Travel on Attrition rate**



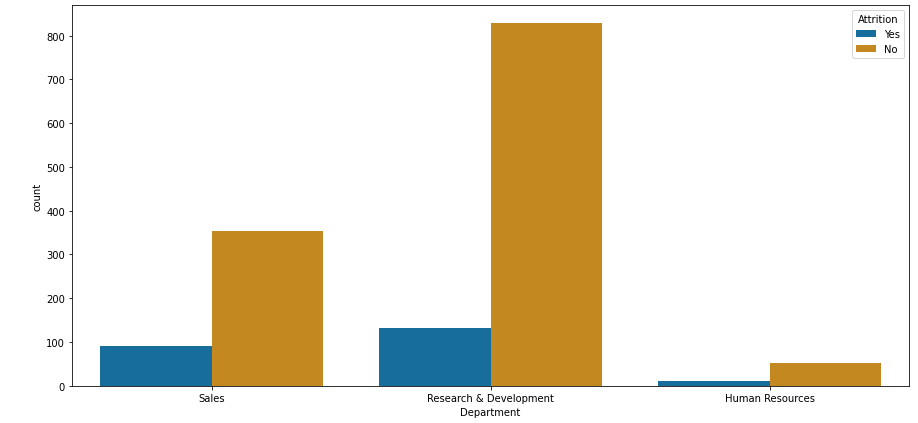
Findings from the above observation are mentioned below:

1) Out of 1044 total no. of employees who Travel Rarely, 156 has left the organization and 887 are still working in that company.

2) Out of 277 total no. of employees who Travel Frequently, 69 has left the organization and 208 are still working in that company.

3) Out of 150 total no. of employees who are Non-Traveller, 12 has left the organization and 138 are still working in that company.

**Impact of Department on Attrition rate**



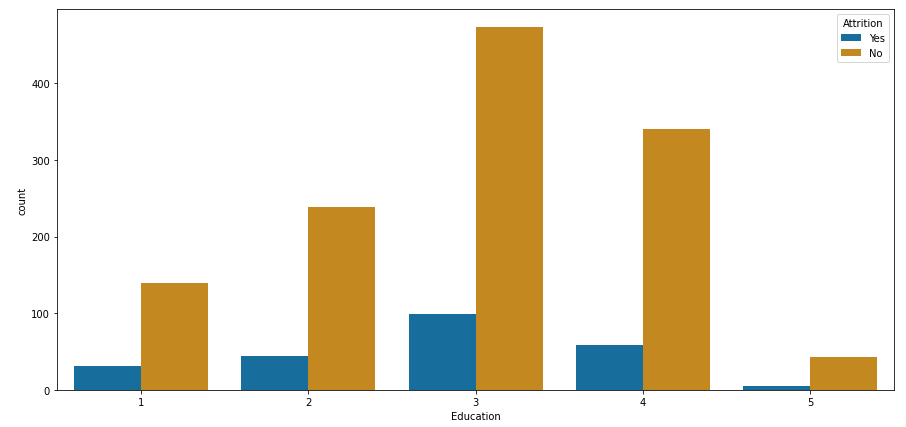
Findings from the above observation are mentioned below:

1) Out of 446 total no. of employees who are working in Sales department, 92 has left the organization and 354 are still working in that company.

2) Out of 961 total no. of employees who are working in Research & Development department, 133 has left the organization and 828 are still working in that company.

3) Out of 63 total no. of employees who are working in Human Resources department, 12 has left the organization and 51 are still working in that company.

**Impact of Education on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 170 total no. of employees who have Education 1, 31 has left the organization and 139 are still working in that company.

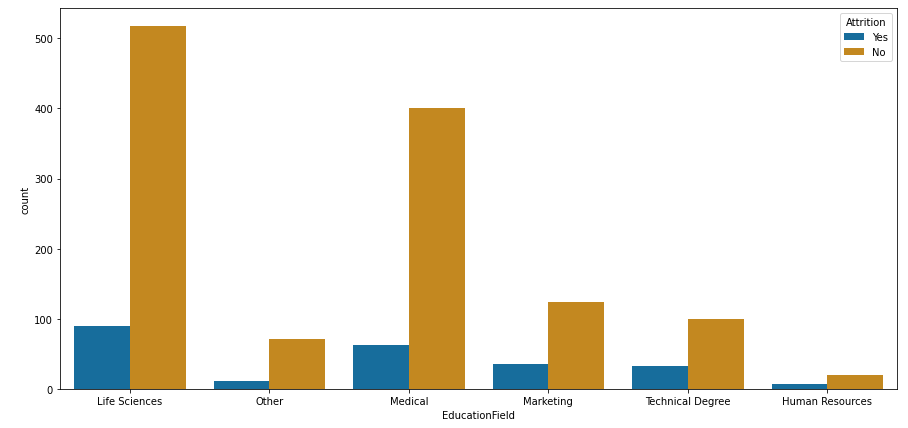
2) Out of 282 total no. of employees who have Education 2, 44 has left the organization and 238 are still working in that company.

3) Out of 572 total no. of employees who have Education 3, 99 has left the organization and 473 are still working in that company.

4) Out of 398 total no. of employees who have Education 4, 58 has left the organization and 340 are still working in that company.

5) Out of 48 total no. of employees who have Education 5, 5 has left the organization and 43 are still working in that company.

**Impact of Education Field** **on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 606 total no. of employees who have Life science as Education Field, 89 has left the organization and 517 are still working in that company.

2) Out of 464 total no. of employees who have Medical as Education Field, 63 has left the organization and 401 are still working in that company.

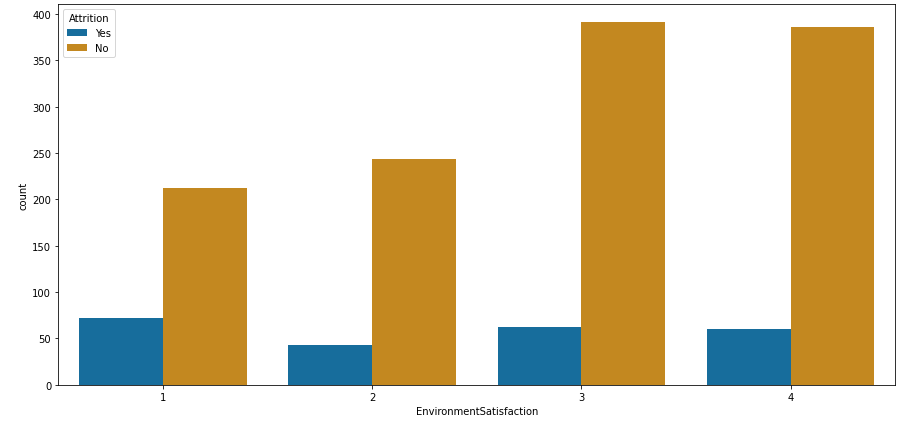
3) Out of 159 total no. of employees who have Marketing as Education Field, 35 has left the organization and 124 are still working in that company.

4) Out of 27 total no. of employees who have Human resource as Education Field, 7 has left the organization and 24 are still working in that company.

5) Out of 132 total no. of employees who have Technical Degree as Education Field, 32 has left the organization and 100 are still working in that company.

6) Out of 82 total no. of employees who have Other Education Field, 11 has left the organization and 71 are still working in that company.

**Impact of Environment Satisfaction** **on Attrition rate**



Findings from the above observation are mentioned below:

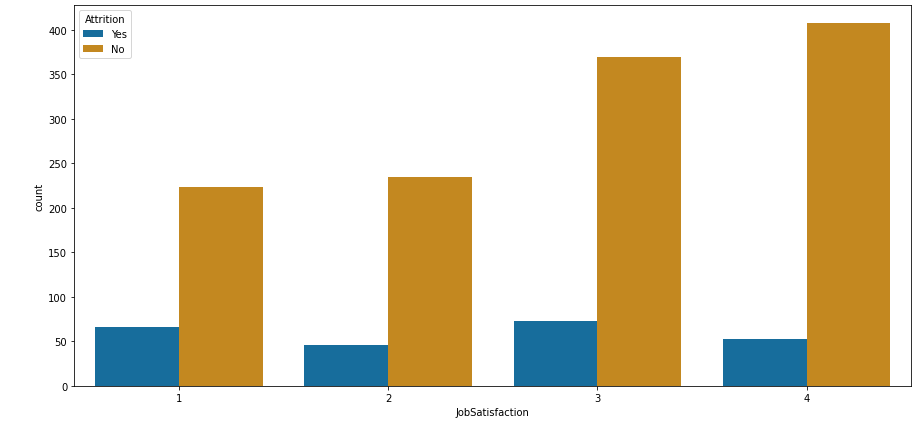
1) Out of 284 total no. of employees who have Environment Satisfaction 1, 72 has left the organization and 212 are still working in that company.

2) Out of 287 total no. of employees who have Environment Satisfaction 2, 43 has left the organization and 244 are still working in that company.

3) Out of 453 total no. of employees who have Environment Satisfaction 3, 62 has left the organization and 391 are still working in that company.

4) Out of 446 total no. of employees who have Environment Satisfaction 4, 60 has left the organization and 386 are still working in that company.

**Impact of Job Satisfaction** **on Attrition rate**



Findings from the above observation are mentioned below:

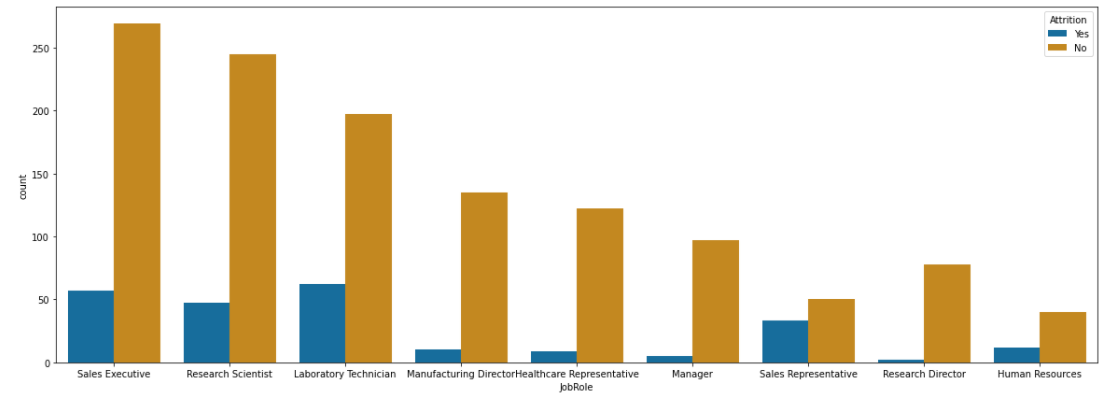
1) Out of 289 total no. of employees who have Job Satisfaction 1, 66 has left the organization and 223 are still working in that company.

2) Out of 280 total no. of employees who have Job Satisfaction 2, 46 has left the organization and 234 are still working in that company.

3) Out of 442 total no. of employees who have Job Satisfaction 3, 73 has left the organization and 369 are still working in that company.

4) Out of 459 total no. of employees who have Job Satisfaction 4, 52 has left the organization and 407 are still working in that company.

**Impact of Job Role** **on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 326 total no. of employees whose Job Role is Sales Executive, 57 has left the organization and 269 are still working in that company.

2) Out of 292 total no. of employees whose Job Role is Research Scientist, 47 has left the organization and 245 are still working in that company.

3) Out of 259 total no. of employees whose Job Role is Laboratory Technician, 62 has left the organization and 197 are still working in that company.

4) Out of 145 total no. of employees whose Job Role is Manufacturing Director, 10 has left the organization and 135 are still working in that company.

5) Out of 131 total no. of employees whose Job Role is Healthcare Representative, 9 has left the organization and 122 are still working in that company.

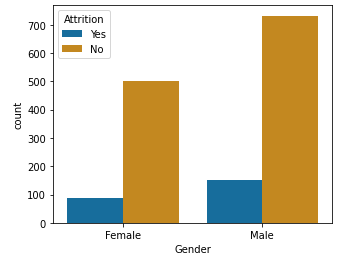
6) Out of 102 total no. of employees whose Job Role is Manager, 5 has left the organization and 97 are still working in that company.

7) Out of 83 total no. of employees whose Job Role is Sales Representative, 33 has left the organization and 50 are still working in that company.

8) Out of 80 total no. of employees whose Job Role is Research Director, 2 has left the organization and 78 are still working in that company.

9) Out of 52 total no. of employees whose Job Role is Human Resources, 12 has left the organization and 40 are still working in that company.

**Impact of Gender** **on Attrition rate**

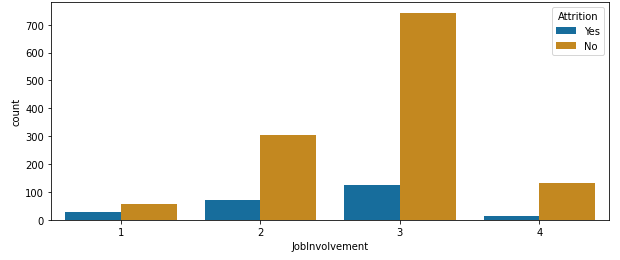


Findings from the above observation are mentioned below:

1) Out of 882 total no. of employees whose Gender is Male, 150 has left the organization and 732 are still working in that company.

2) Out of 588 total no. of employees whose Gender is Female, 87 has left the organization and 501 are still working in that company.

**Impact of Job Involvement** **on Attrition rate**



Findings from the above observation are mentioned below:

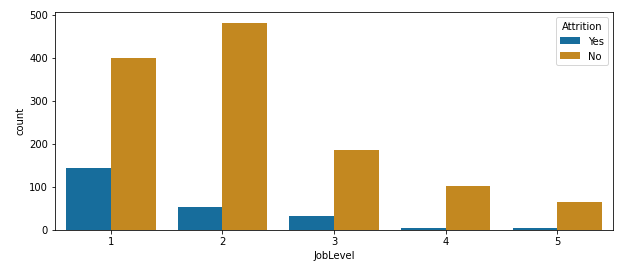
1) Out of 83 total no. of employees under Job Involvement 1, 28 has left the organization and 55 are still working in that company.

2) Out of 375 total no. of employees under Job Involvement 2, 71 has left the organization and 304 are still working in that company.

3) Out of 868 total no. of employees under Job Involvement 3, 125 has left the organization and 743 are still working in that company.

4) Out of 144 total no. of employees under Job Involvement 4, 13 has left the organization and 131 are still working in that company.

**Impact of Job Level** **on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 543 total no. of employees whose Job Level 1, 143 has left the organization and 400 are still working in that company.

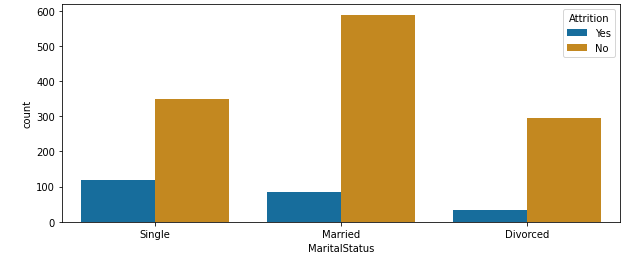
2) Out of 534 total no. of employees whose Job Level 2, 52 has left the organization and 482 are still working in that company.

3) Out of 218 total no. of employees whose Job Level 3, 32 has left the organization and 186 are still working in that company.

4) Out of 106 total no. of employees whose Job Level 4, 5 has left the organization and 101 are still working in that company.

5) Out of 69 total no. of employees whose Job Level 5, 5 has left the organization and 64 are still working in that company.

**Impact of Marital Status** **on Attrition rate**



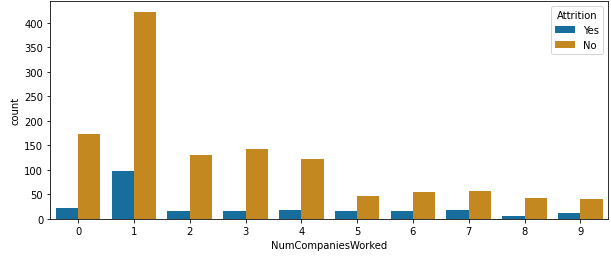
Findings from the above observation are mentioned below:

1) Out of 470 total no. of employees whose Marital Status is Single, 120 has left the organization and 350 are still working in that company.

2) Out of 673 total no. of employees whose Marital Status is Married, 84 has left the organization and 589 are still working in that company.

3) Out of 327 total no. of employees whose Marital Status is Divorced, 33 has left the organization and 294 are still working in that company.

**Impact of Number of companies worked** **on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 197 total no. of employees whose Number of Companies Worked 0, 23 has left the organization and 174 are still working in that company.

2) Out of 521 total no. of employees whose Number of Companies Worked 1, 98 has left the organization and 423 are still working in that company.

3) Out of 146 total no. of employees whose Number of Companies Worked 2, 16 has left the organization and 130 are still working in that company.

4) Out of 159 total no. of employees whose Number of Companies Worked 3, 16 has left the organization and 143 are still working in that company.

5) Out of 139 total no. of employees whose Number of Companies Worked 4, 17 has left the organization and 122 are still working in that company.

6) Out of 63 total no. of employees whose Number of Companies Worked 5, 16 has left the organization and 47 are still working in that company.

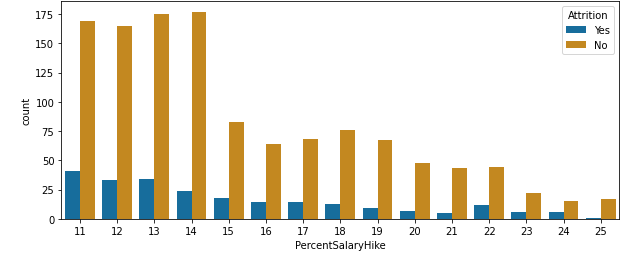
7) Out of 70 total no. of employees whose Number of Companies Worked 6, 16 has left the organization and 54 are still working in that company.

8) Out of 74 total no. of employees whose Number of Companies Worked 7, 17 has left the organization and 57 are still working in that company.

9) Out of 49 total no. of employees whose Number of Companies Worked 8, 6 has left the organization and 43 are still working in that company.

10) Out of 52 total no. of employees whose Number of Companies Worked 9, 12 has left the organization and 40 are still working in that company.

**Impact of Percent Salary Hike on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 210 total no. of employees whose Percent Salary Hike is 11, 41 has left the organization and 169 are still working in that company.

2) Out of 198 total no. of employees whose Percent Salary Hike is 12, 33 has left the organization and 165 are still working in that company.

3) Out of 189 total no. of employees whose Percent Salary Hike is 13, 34 has left the organization and 175 are still working in that company.

4) Out of 201 total no. of employees whose Percent Salary Hike is 14, 24 has left the organization and 177 are still working in that company.

5) Out of 101 total no. of employees whose Percent Salary Hike is 15, 18 has left the organization and 83 are still working in that company.

6) Out of 78 total no. of employees whose Percent Salary Hike is 16, 14 has left the organization and 64 are still working in that company.

7) Out of 82 total no. of employees whose Percent Salary Hike is 17, 14 has left the organization and 68 are still working in that company.

8) Out of 89 total no. of employees whose Percent Salary Hike is 18, 13 has left the organization and 76 are still working in that company.

9) Out of 76 total no. of employees whose Percent Salary Hike is 19, 9 has left the organization and 67 are still working in that company.

10) Out of 55 total no. of employees whose Percent Salary Hike is 20, 7 has left the organization and 48 are still working in that company.

11) Out of 48 total no. of employees whose Percent Salary Hike is 21, 5 has left the organization and 43 are still working in that company.

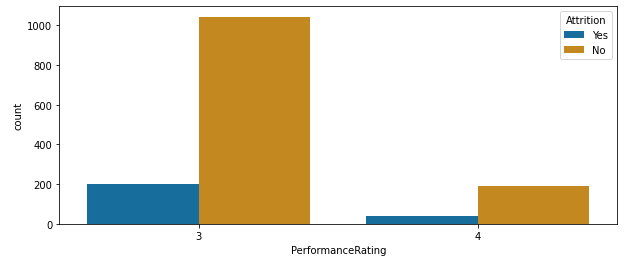
12) Out of 56 total no. of employees whose Percent Salary Hike is 22, 12 has left the organization and 44 are still working in that company.

13) Out of 28 total no. of employees whose Percent Salary Hike is 23, 6 has left the organization and 22 are still working in that company.

14) Out of 21 total no. of employees whose Percent Salary Hike is 24, 6 has left the organization and 15 are still working in that company.

15) Out of 18 total no. of employees whose Percent Salary Hike is 25, 1 has left the organization and 17 are still working in that company.

**Impact of Performance Rating on Attrition rate**



Findings from the above observation are mentioned below:

1) Out of 1244 total no. of employees whose Performance Rating is 3, 200 has left the organization and 1044 are still working in that company.

2) Out of 226 total no. of employees whose Performance Rating is 4, 37 has left the organization and 189 are still working in that company.

# Impact of Years in Current Role on Attrition rate

# 

Findings from the above observation are mentioned below:

1) Out of 244 total no. of employees whose Years in Current Role is 0, 73 has left the organization and 171 are still working in that company.

2) Out of 57 total no. of employees whose Years in Current Role is 1, 11 has left the organization and 46 are still working in that company.

3) Out of 372 total no. of employees whose Years in Current Role is 2, 68 has left the organization and 304 are still working in that company.

4) Out of 135 total no. of employees whose Years in Current Role is 3, 16 has left the organization and 119 are still working in that company.

5) Out of 104 total no. of employees whose Years in Current Role is 4, 15 has left the organization and 89 are still working in that company.

6) Out of 36 total no. of employees whose Years in Current Role is 5, 1 has left the organization and 35 are still working in that company.

7) Out of 37 total no. of employees whose Years in Current Role is 6, 2 has left the organization and 35 are still working in that company.

8) Out of 222 total no. of employees whose Years in Current Role is 7, 31 has left the organization and 191 are still working in that company.

9) Out of 89 total no. of employees whose Years in Current Role is 8, 7 has left the organization and 82 are still working in that company.

10) Out of 67 total no. of employees whose Years in Current Role is 9, 6 has left the organization and 61 are still working in that company.

11) Out of 29 total no. of employees whose Years in Current Role is 10, 2 has left the organization and 27 are still working in that company.

12) Out of 22 total no. of employees whose Years in Current Role is 11, 0 has left the organization and 22 are still working in that company.

13) Out of 10 total no. of employees whose Years in Current Role is 12, 1 has left the organization and 9 are still working in that company.

14) Out of 14 total no. of employees whose Years in Current Role is 13, 1 has left the organization and 13 are still working in that company.

15) Out of 11 total no. of employees whose Years in Current Role is 14, 1 has left the organization and 10 are still working in that company.

16) Out of 8 total no. of employees whose Years in Current Role is 15, 2 has left the organization and 6 are still working in that company.

17) Out of 7 total no. of employees whose Years in Current Role is 16, 0 has left the organization and 7 are still working in that company.

18) Out of 4 total no. of employees whose Years in Current Role is 17, 0 has left the organization and 4 are still working in that company.

19) Out of 2 total no. of employees whose Years in Current Role is 18, 0 has left the organization and 2 are still working in that company

# Impact of Work Life Balance on Attrition rate

# 

Findings from the above observation are mentioned below:

1) Out of 80 total no. of employees whose Work Life Balance is 1, 25 has left the organization and 55 are still working in that company.

2) Out of 344 total no. of employees whose Work Life Balance is 2, 58 has left the organization and 286 are still working in that company.

3) Out of 893 total no. of employees whose Work Life Balance is 3, 127 has left the organization and 766 are still working in that company.

4) Out of 153 total no. of employees whose Work Life Balance is 4, 27 has left the organization and 126 are still working in that company.

# Impact of Relationship Satisfaction on Attrition rate

# 

Findings from the above observation are mentioned below:

1) Out of 276 total no. of employees whose Relationship Satisfaction is 1, 57 has left the organization and 219 are still working in that company.

2) Out of 304 total no. of employees whose Relationship Satisfaction is 2, 45 has left the organization and 258 are still working in that company.

3) Out of 459 total no. of employees whose Relationship Satisfaction is 3, 71 has left the organization and 388 are still working in that company.

4) Out of 432 total no. of employees whose Relationship Satisfaction is 4, 64 has left the organization and 368 are still working in that company.

# Impact of Years Since Last Promotion on Attrition rate

# 

Findings from the above observation are mentioned below:

1) Out of 581 total no. of employees whose Years Since Last Promotion is 0, 110 has left the organization and 471 are still working in that company.

2) Out of 357 total no. of employees whose Years Since Last Promotion is 1, 49 has left the organization and 308 are still working in that company.

3) Out of 154 total no. of employees whose Years Since Last Promotion is 2, 27 has left the organization and 127 are still working in that company.

4) Out of 52 total no. of employees whose Years Since Last Promotion is 3, 9 has left the organization and 43 are still working in that company.

5) Out of 61 total no. of employees whose Years Since Last Promotion is 4, 5 has left the organization and 56 are still working in that company.

6) Out of 45 total no. of employees whose Years Since Last Promotion is 5, 2 has left the organization and 43 are still working in that company.

7) Out of 32 total no. of employees whose Years Since Last Promotion is 6, 6 has left the organization and 26 are still working in that company.

8) Out of 76 total no. of employees whose Years Since Last Promotion is 7, 16 has left the organization and 60 are still working in that company.

9) Out of 18 total no. of employees whose Years Since Last Promotion is 8, 0 has left the organization and 18 are still working in that company.

10) Out of 17 total no. of employees whose Years Since Last Promotion is 9, 4 has left the organization and 13 are still working in that company.

11) Out of 6 total no. of employees whose Years Since Last Promotion is 10, 1 has left the organization and 5 are still working in that company.

12) Out of 24 total no. of employees whose Years Since Last Promotion is 11, 2 has left the organization and 22 are still working in that company.

13) Out of 10 total no. of employees whose Years Since Last Promotion is 12, 0 has left the organization and 10 are still working in that company.

14) Out of 10 total no. of employees whose Years Since Last Promotion is 13, 2 has left the organization and 8 are still working in that company.

15) Out of 9 total no. of employees whose Years Since Last Promotion is 14, 1 has left the organization and 8 are still working in that company.

16) Out of 13 total no. of employees whose Years Since Last Promotion is 15, 3 has left the organization and 10 are still working in that company.

# Impact of Over Time on Attrition rate

# 

Findings from the above observation are mentioned below:

1) Out of 416 total no. of employees getting Over Time, 127 has left the organization and 289 are still working in that company.

2) Out of 1054 total no. of employees not getting Over Time, 110 has left the organization and 944 are still working in that company

**4.     Pre-processing Pipeline:**

# Checking Missing Values:

# While analysing missing values in the dataset, no missing values were found. Values, as calculated from dataset through coding, are presented below for ready reference.

# 

# Checking Data types of the columns:

# While analysing data types in the dataset, two types of data were found – Integer and Object. Values, as calculated from the dataset through coding, are presented below for ready reference.

# 

# Using map function:

# As data type of target column (Attrition) is object, hence using map function to convert categorical data into Numerical data type.

# Checking Correlation of variables with target column:

# While checking correlation of variables with target column, the following are the observations:

1) Age: 15.92% negative correlation with target column.

2) Daily Rate: 5.67% negative correlation with target column.

3) Distance from Home: 7.79% positive correlation with target column.

4) Education: 3.14% negative correlation with target column.

5) Environment Satisfaction: 10.34% negative correlation with target column.

6) Hourly Rate: 0.68% negative correlation with target column.

7) Job Involvement: 13% negative correlation with target column.

8) Job Level: 16.91% negative correlation with target column.

9) Job Satisfaction: 10.35% negative correlation with target column.

10) Monthly Income: 15.98% negative correlation with target column.

11) Monthly Rate: 1.52% positive correlation with target column.

12) Number of Companies Worked: 4.35% positive correlation with target column.

13) Percent Salary Hike: 1.35% negative correlation with target column.

14) Performance Rating: 0.29% positive correlation with target column.

15) Relationship Satisfaction: 4.59% negative correlation with target column.

16) Stock Option Level: 13.71% negative correlation with target column.

17) Total Working Years: 17.11% negative correlation with target column.

18) Training Times Last Year: 5.95% negative correlation with target column.

19) Work Life Balance: 6.39% negative correlation with target column.

20) Years at Company: 13.44% negative correlation with target column.

21) Years in Current Role: 16.05% negative correlation with target column.

22) Years Since Last Promotion: 3.30% negative correlation with target column.

23) Years with Current Manager: 15.62% negative correlation with target column.

**Checking Outlier:**

Following are the columns in which outliers have been observed.

* Monthly Income
* Number of Companies Worked
* Performance Rating
* Stock Option Level
* Total Working Years
* Training Times Last Year
* Work Life Balance
* Years At Company
* Years in Current Role
* Years Since Last Promotion
* Years With Current Manager

# Checking Skewness:

Following are the columns in which skewness are observed.

Distance From Home -------- -- 0.958118

Job Level --------- 1.025401

Monthly Income --------- 1.369817

Number of Companies Worked ---------- 1.026471

Percent Salary Hike ---------- 0.821128

Performance Rating---------- 1.921883

Stock Option Level ---------- 0.968980

Total Working Years---------- 1.117172

Training Times Last Year---------- 0.553124

Work Life Balance---------- -0.552480

Years At Company ---------- 1.764529

Years In Current Role---------- 0.917363

Years Since Last Promotion---------- 1.984290

Years With Current Manager---------- 0.833451

# Checking Normal Distribution:

# As per observation, the values of the columns of the analysed dataset are not normally distributed.

# Using encoding technique:

# Label encoder is used for converting Categorical data into Numerical data.

# Removing Outliers:

# Using np.where function so that it will return the index where data points are outlier

# Extracting the data points those are in the range of (-3,3)

# Checking Total data loss:

# 

# Losing 5.64% of total data after removing outliers.

# Dividing the Dataset into features and vectors where x is feature and y is vector:

# 

# Transforming data to remove skewness with the help of power transform technique:

# 

# Treating Imbalance dataset:

# 

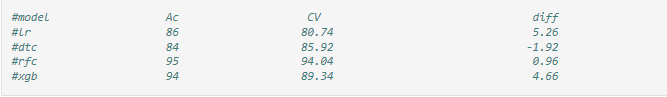
# As per above observation, total number of records for 1 and 0 are equal.

# To get best accuracy score on best Random state:

# 

**5.     Building Machine Learning Models:**

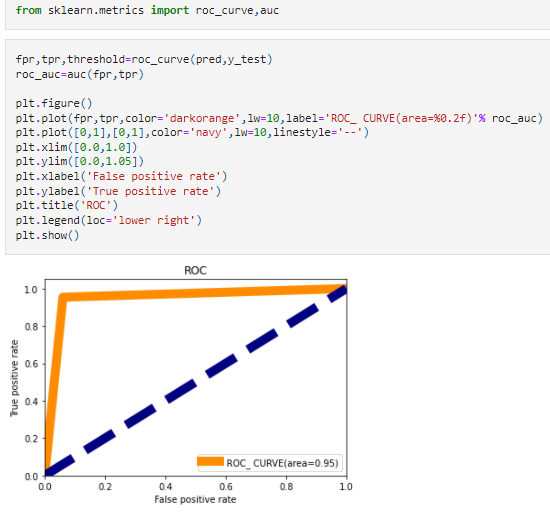
Lastly, in the pre-processing steps, worked on to get the best accuracy score on best random state and then moved on to build different machine learning models based on the random state value. For this dataset analysis, four machine learning models were built, namely, Logistic Regression(lr), Decision Tree Classifier(dtc), Random Forest Classifier (rfc) and XGB Classifier(xgb). After building these four models, the values were compared with the cross-validation score and the difference in the individual values were taken into account. The lesser the difference value, the more viable is the model. In this case, the difference value of Random Forest Classifier is observed to be the lesser amongst the four, hence this model is considered for this particular dataset.



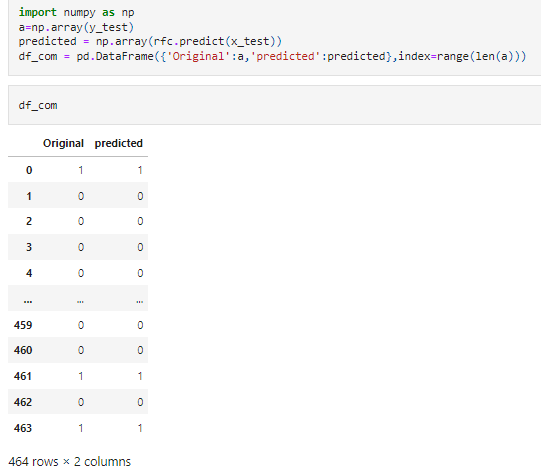
After finalising the model, Hyperparameter Tuning technique is applied to further tune the model and increase the accuracy score. The Hyperparameter Tuning helped in selecting the best parameters out of the ones provided and further assisted in building the final model based on the selected parameters.

**6.     Concluding Remarks:**

After building the model, checked AUC (Area under curve) and ROC (Receiver operating characteristic). It was observed that 95% of area is getting covered which signifies that 95% of times the outcome will be correct.



Compared **Original values** with **Predicted values** to check how the model is performing, observed that 95% of area is getting covered which signifies that 95% of times the outcome will be correct.



Hence, after checking all the necessary steps it can be concluded that the selected model is performing well as it is predicting correct values 95% of times.