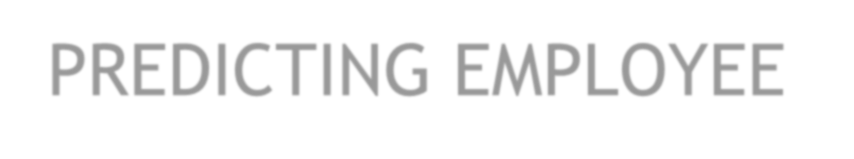
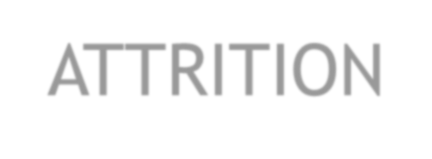
PREDICTING EMPLOYEE ATTRITION



PREDICTING EMPLOYEE



ATTRITION

Objective and Scope of the study

* The objective of this project is to predict the attrition rate for each employee, to find out who’s more likely to leave the organization.
* It will help organizations to find ways to prevent attrition or to plan in advance the hiring of new candidate.
* Attrition proves to be a costly and time consuming problem for organization and it also leads to loss of productivity.
* The scope of the project extends to companies in all industries

# ANALYTICS APPROACH

* Check for missing values in the data, and if any, will process the data accordingly.
* Understand how the features are related with our target variable-attrition.
* Convert target variable into numeric form.
* Apply the feature selection and feature engineering to make it model ready.
* Apply various algorithms to check which one is the most suitable.
* Draw out recommendations based our analysis.
* The model is online host Streamlit

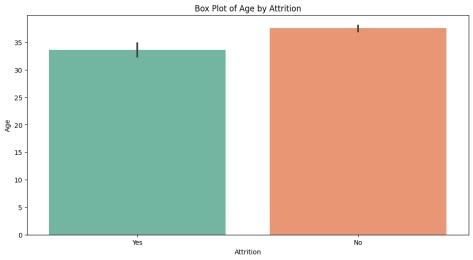
# Tools and Techniques

* We have selected Python as our analytics tools.
* Python includes many packages such as Numpy,Pandas,Seaborn,Matplotlib,Streamlit.
* Algorithms such as Logistic Regression,Randomforest,SVM,XGBoost have been used for prediction.

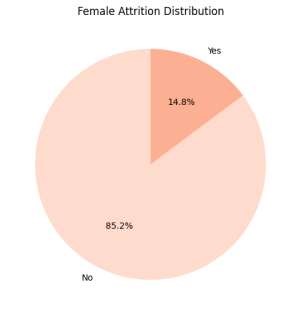
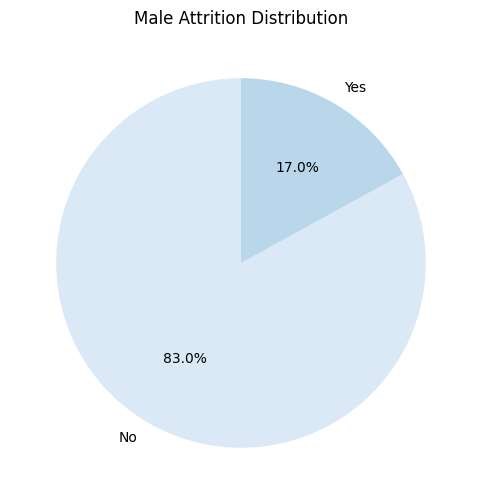
# Visualization(EDA)

* Age vs Attrition
* Gender vs Attrition
* Department vs Attrition
* Distancefromhome vs Attrition
* Business Travel vs Attrition
* Environment Satisfaction vs Attrition
* Education field vs Attrition
* Job Role vs Attrition
* Marital Status vs Attrition
* Monthly Income vs Attrition

AGE VS ATTRITION

* Most of the emloyees are between age 30 to 40.
* We can clearly observe a trend that as the age is increasing the attrition is decreasing.
* The medain age of employee who left the organization is less than the employees who are working.
* Employees with young age leaves the company more compared to elder employees.

## GENDER VS ATTRITION



 Male employees accounts for a higher proportion than female employees by more than 20%.1



Attrition in male employees is higher compared to female employees.

### DEPARTMENT VS ATTRITION

* Most of the employees are from Research & Development Department (65.4%).
* Highest Attrition is in the Sales Department.
* Human Resources Department Attrition rate is also very high.
* Attrition in Research & Development Department is least compared to other departments.

## BUSSINESS TRAVEL VS ATTRITION

* Most of the employees in the organization Travel Rarely.
* Highest employee attrition can be observed by those employees who Travels Frequently.
* Lowest employee attrition can be observed by those employees who are Non-Travel.

### ENVIROMENT SATISFACTION VS ATTRITION

* Most of the employees have rated the organization environment satisfaction High & Very High.
* Attrition Rate is high among the employee with high level of environment satisfication.

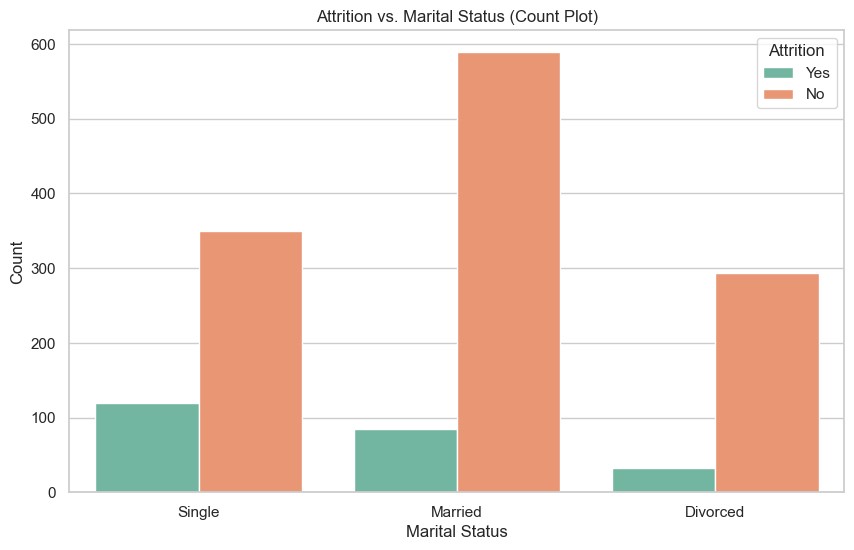
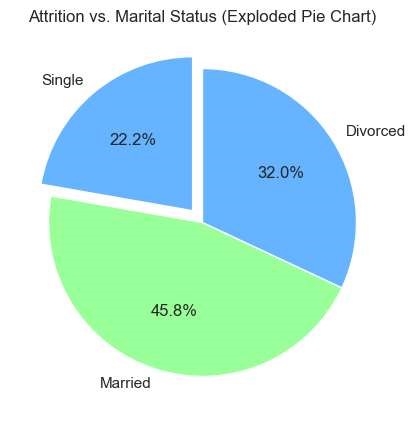
## EDUCATION FIELD VS ATTRITION

* Most of the employees are either from Life Science or Medical Education Field.
* Very few employees are from Human Resources Education Field.
* Education Fields like Human Resources, Technical, Marketing is having very high attrition rate.
* This may be because of work load becuase there are very few employees in these education fields compared to education field with less attrition rate.

## JOB ROLE VS ATTRITON

* Most employees is working as Sales executive, Research Scientist or Laboratory Technician.
* Highest attrition rates are in role of Sale Representative.

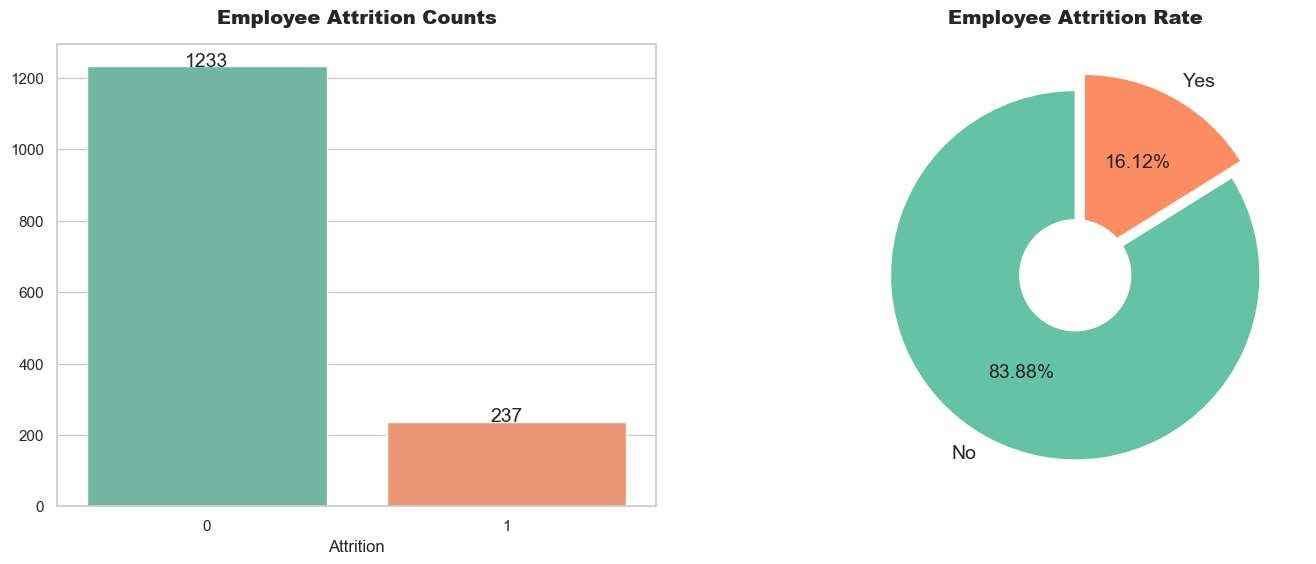
sMARITAL STATUS VS ATTRITION



### PERFORMANCERATING VS ATTRITION

* Most of the employees are having excellent performance rating.
* Both the categories in this field is having same attriton rate.
* No meaningful insight from this feature.

EMPLOYEE ATTRITON



MACHINE LEARINIG PACKAGES

* from sklearn.tree import DecisionTreeClassifier
* from sklearn.ensemble import RandomForestClassifier
* from sklearn.naive\_bayes import GaussianNB
* from sklearn.neighbors import KNeighborsClassifier
* from sklearn.svm import SVC
* from sklearn.ensemble import AdaBoostClassifier
* from sklearn.ensemble import

GradientBoostingClassifier

* from sklearn.linear\_model import LogisticRegression
* from sklearn.model\_selection import train\_test\_split
* from sklearn.metrics import accuracy\_score
* from xgboost import XGBClassifier
* from sklearn import metrics
* from sklearn.metrics import roc\_curve
* from sklearn.metrics import recall\_score
* from sklearn.metrics import confusion\_matrix
* from sklearn.metrics import precision\_score
* from sklearn.metrics import f1\_score
* from sklearn.metrics import classification\_report
* from sklearn.metrics import roc\_auc\_score
* from sklearn.model\_selection import cross\_val\_score
* from sklearn.model\_selection import GridSearchCV
* from sklearn.metrics import accuracy\_score
* from sklearn.model\_selection import RepeatedStratifiedKFold
* from sklearn.model\_selection import

RandomizedSearchCV

* from scipy.stats import randint
* from scipy.stats import uniform
* from sklearn import metrics
* from sklearn.metrics import \*

SCORE OF ALL ML MODELS

**Algorithms**

**Training Score**

**Testing Score**

**Precision**

**Recall**

**AUC Score**

**0**

Logistic

Regression

81.896552

78.744939

0.781377

0.790984

0.875721

**1**

KNN

90.263692

83.198381

0.776632

0.926230

0.936943

**2**

Naive Bayes

73.073022

71.862348

0.685512

0.795082

0.821443

**3**

Decision Tree

100.000000

80.971660

0.815126

0.795082

0.809541

**4**

Random Forest

100.000000

88.056680

0.911111

0.840164

0.955361

**5**

Gradient Boost

94.168357

89.068826

0.913043

0.860656

0.941148

**6**

XGBoost

100.000000

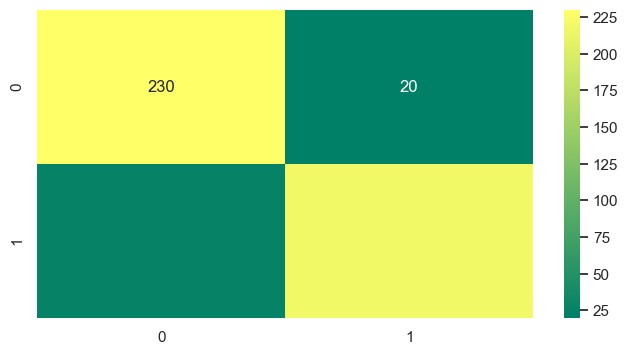
90.890688

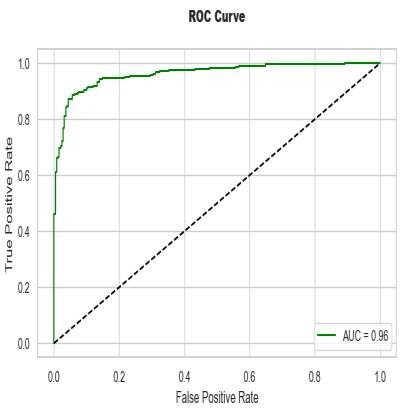
0.916318

0.897541

0.962230

WE USE XGBOOST BECAUSE THERE TESTING SCORE AND AUC SCORE IS GOOD :





STREAMLIT



# THANK YOU