Employee Attrition Analysis

A case study on analysis attrition of employee and building a machine learning model to classify the employee who likely to resign.

Content

1. Project INTRODUCTION.
2. ANALYTICS process flow.
3. What does historical DATA tell us ?
4. PREDICTION model
5. Model prediction PERFORMANCE evaluation
6. PRESCRIPTIVE analytics

Attrition of employee is departure of staff from organization over time. This can lead to a loss of valuable talent and increased recruitment costs.

PROJECT INTRODUCTION

Prediction if an organization's employee will leave or stay

Analytics Process

Understanding Requirement

Determine the target responds

Exploratory Data Analysis

Data Cleaning
Univariate Analysis
Multivariate Analysis

Model Evaluation

Confusion Matrix

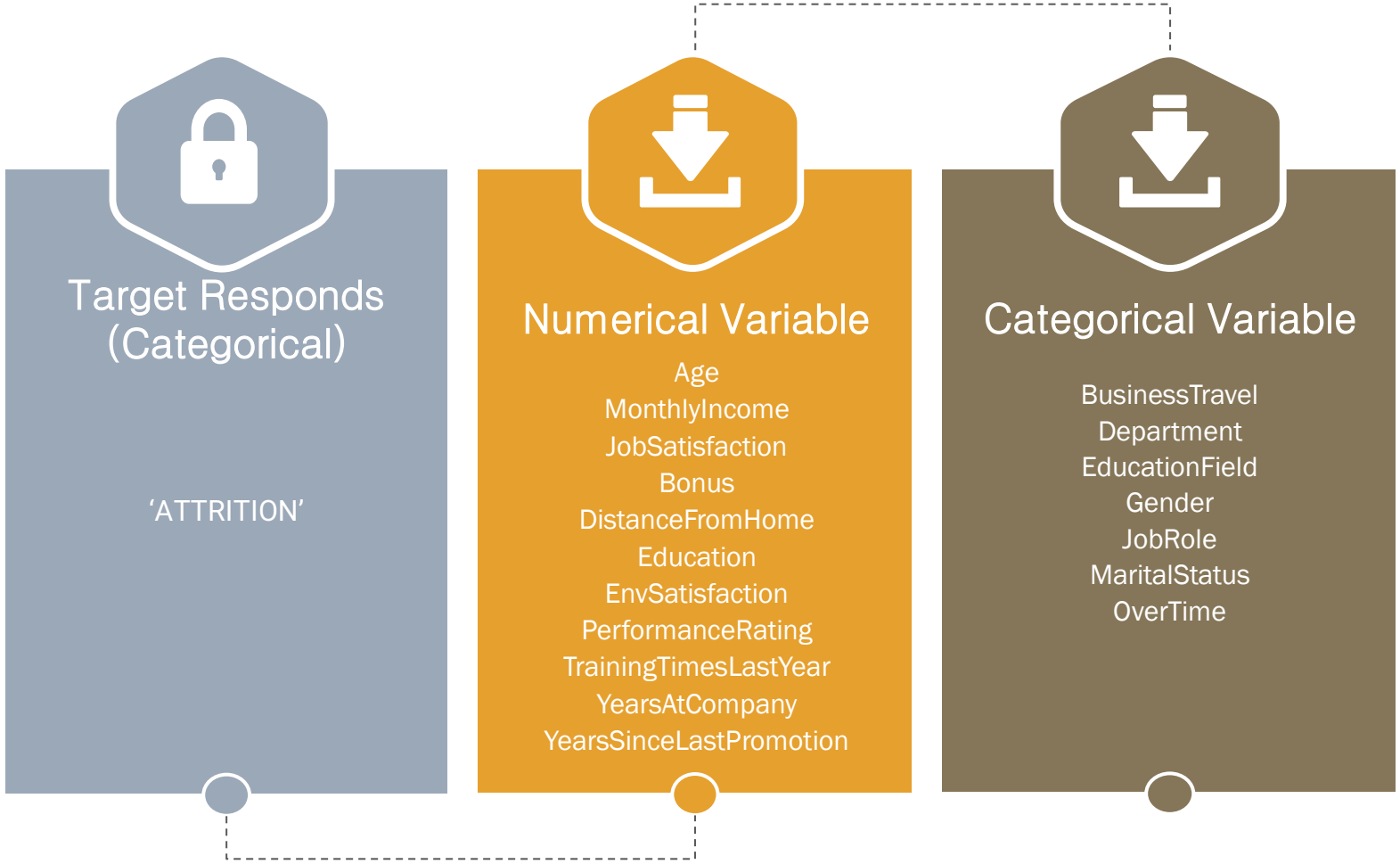
Data Preparation

Find / collect data source

Modeling

Feature Selection
Partitioning
Scaling
Training
Model Testing

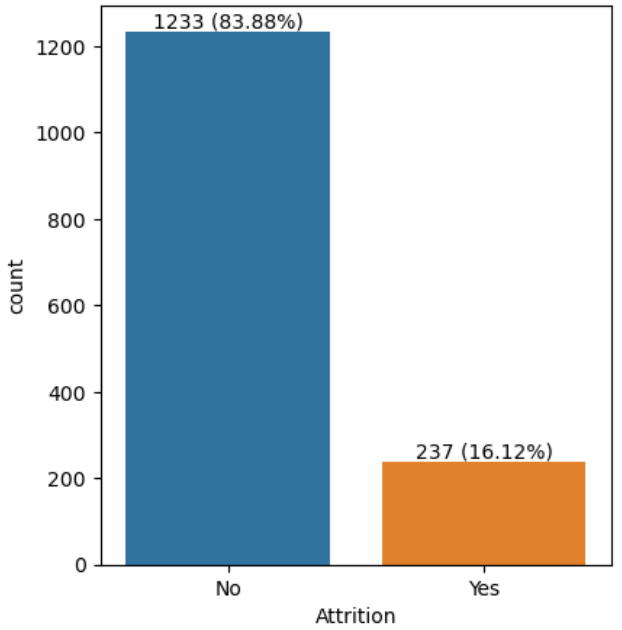
EXPLORATORY DATA ANALYSIS

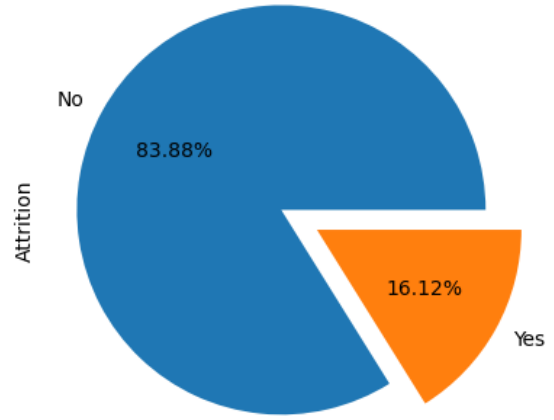


Target Responds

Attrition variable

- We need to predict the attrition variable based on input variables given.
- Note that the target responds is categorical variable





- In the company, there are **1470 employees** based on the database received. **237** employees who compose 16.12% of the total number of employees **left** the company for some reasons. Besides that, **1233** employee is currently continuing to work in the same company.

Numerical Vs Categorical

Input	Analytics Result	Relationship
Age	T-Test Result : P-value < 0.05	✓
Monthly Income	T-Test Result : P-value < 0.05	✓
Job Satisfaction	T-Test Result : P-value < 0.05	✓
Bonus	T-Test Result : P-value < 0.05	✓
Distance from Home	T-Test Result : P-value < 0.05	✓
Education Level	T-Test Result : P-value > 0.05	✗
Env Satisfaction	T-Test Result : P-value < 0.05	✓

Numerical Vs Categorical

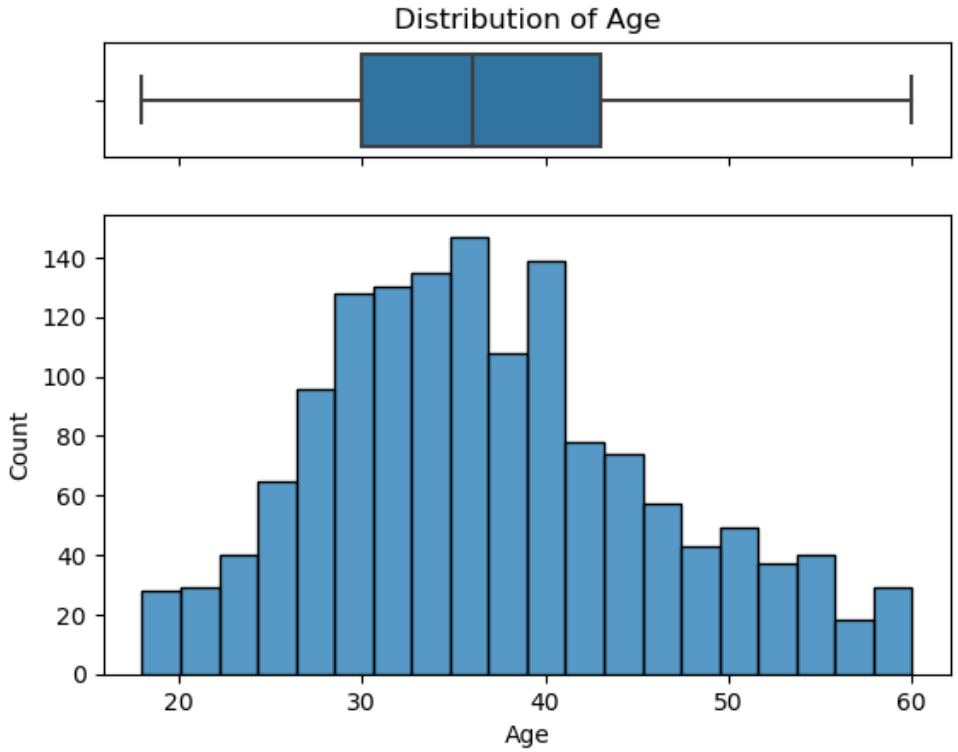
Input	Analytics Result	Relationship
Performance Rating	T-Test Result : P-value > 0.05	
Training Times LastYear	T-Test Result : P-value < 0.05	
Years At Company	T-Test Result : P-value < 0.05	
Years Since Last Promotion	T-Test Result : P-value < 0.05	

cont

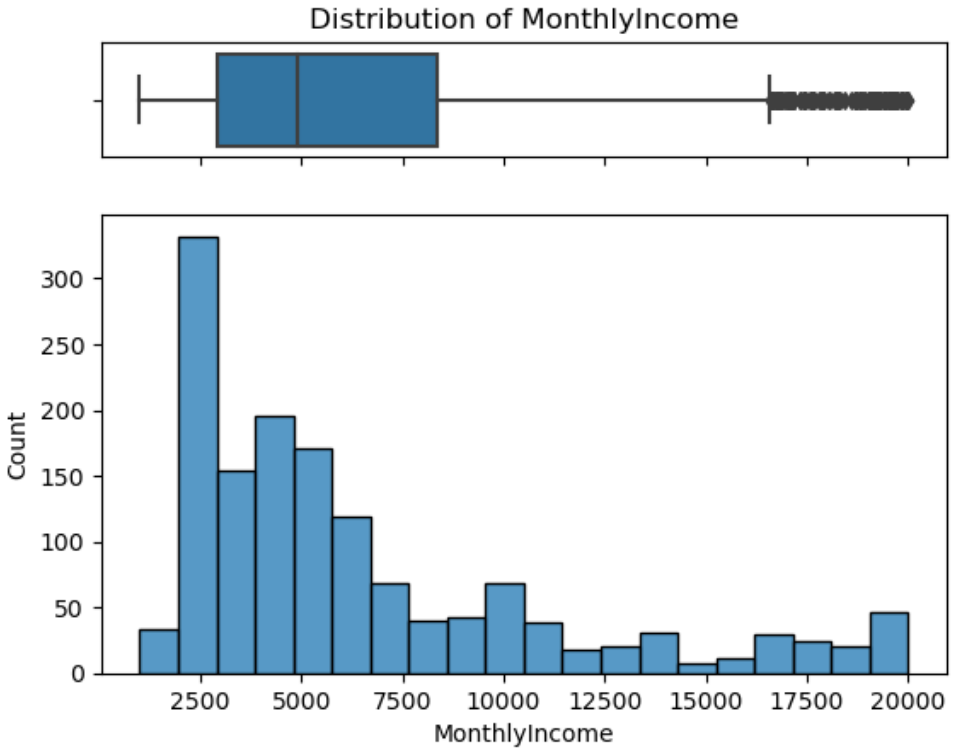
Categorical Vs Categorical

Input	Analytics Result	Relationship
Business Travel	Chi2 Test Result : P-value < 0.05	✓
Department	Chi2 Test Result : P-value < 0.05	✓
Education Field	Chi2 Test Result : P-value < 0.05	✓
Gender	Chi2 Test Result : P-value > 0.05	✗
Job Role	Chi2 Test Result : P-value < 0.05	✓
Marital Status	Chi2 Test Result : P-value < 0.05	✓
Overtime	Chi2 Test Result : P-value < 0.05	✓

Univariate Analysis

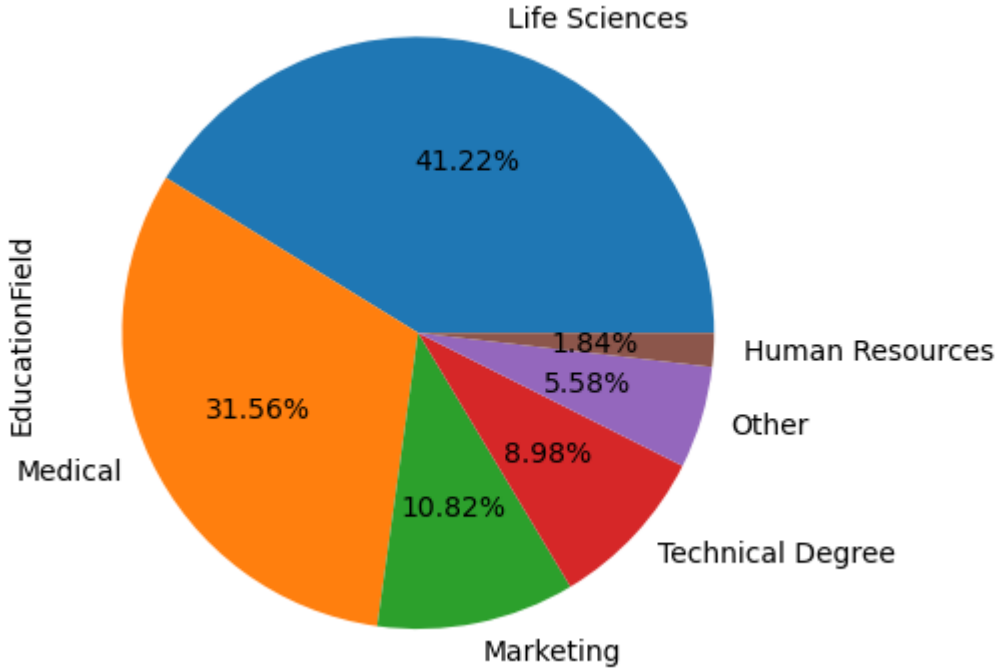


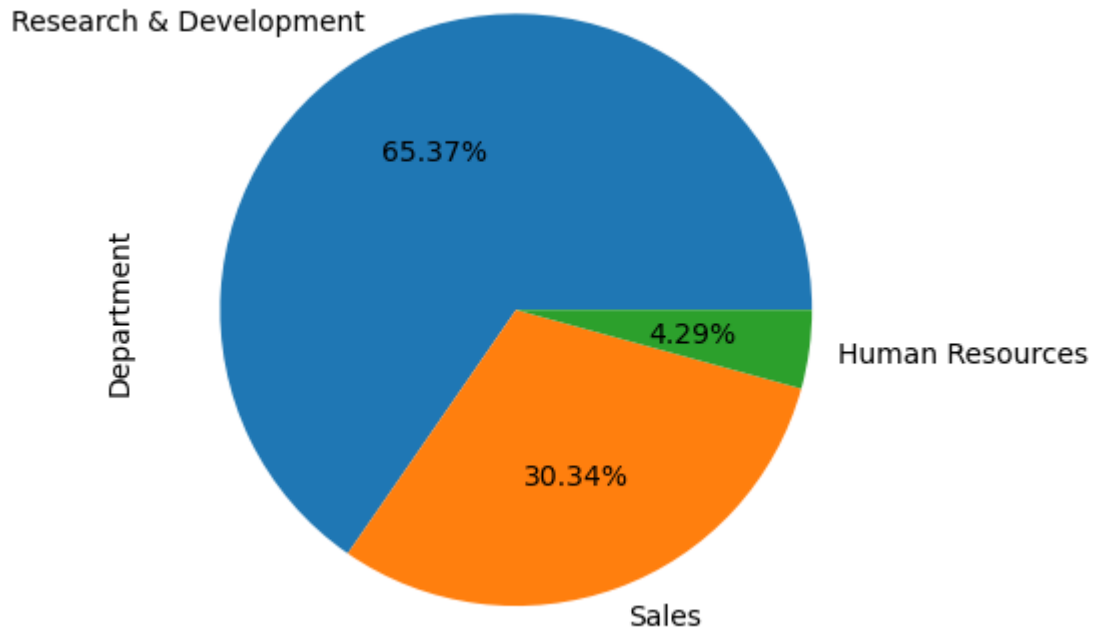
- 25% OF THE EMPLOYEE WITH AGE LESS THAN 30 YEARS OLD
- 25% OF THE EMPLOYEE WITH AGE MORE THAN 43 YEARS OLD
- 50% OF THE EMPLOYEE WITH AGE BETWEEN 30 TO 40 YEARS OLD



- MEAN VALUE OF THE MONTHLY INCOME IS 6502.93
- 50% OF THE EMPLOYEE RECEIVED MONTHLY INCOME IN RANGE OF 2911 TO 8379

Univariate Analysis



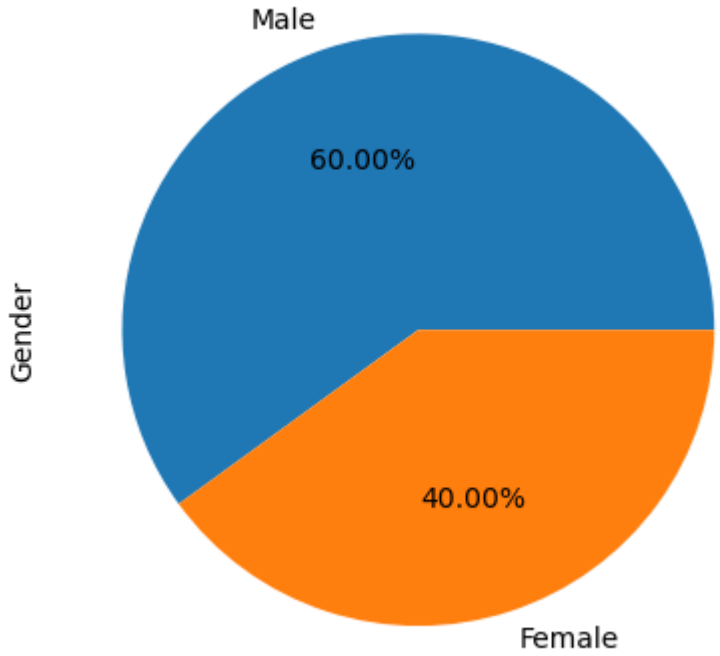


TOP 3 OF EDUCATION FIELD IN THE ORGANIZATION IS

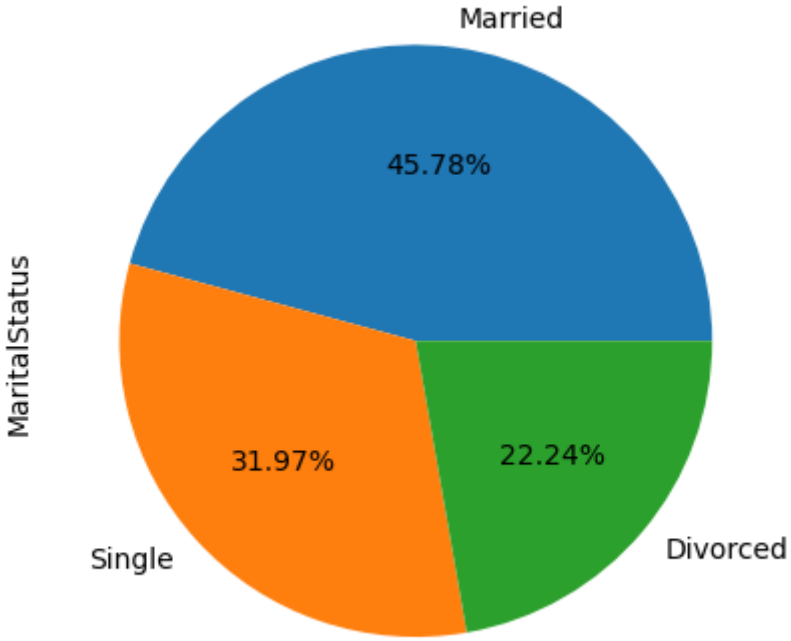
1. LIFE SCIENCE
2. MEDICAL
3. MARKETING

- THERE ARE 3 DEPARTMENT IN THE ORGANIZATION, BUT THE MAJOR DEPARTMENT IS **R&D** WHICH IS COVER **65.37%** FOR THE WHOLE ORGANIZATION. **SALES** DEPARTMENT COMPOSE THE **30.34%** AND THE **REST** IS IN **HR DEPARTMENT**.

Univariate Analysis

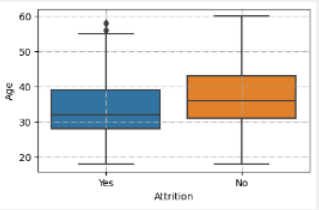
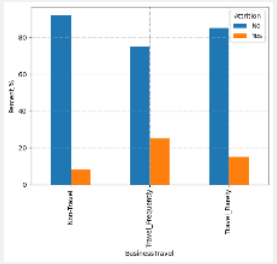
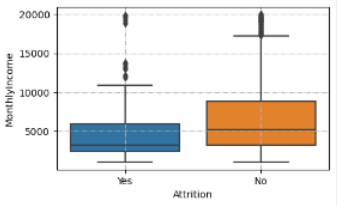
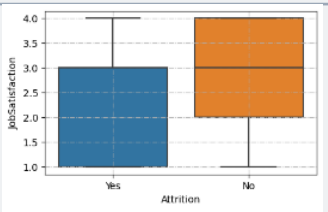


- Male employees are more in number, contributing to 60%

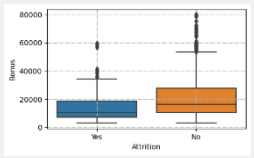
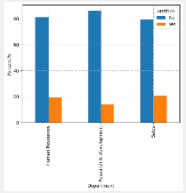
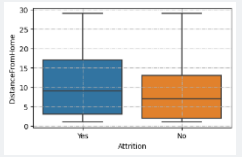
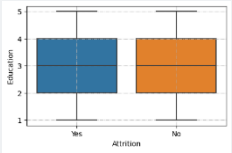
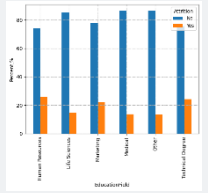


- Married employees are more in the company with 45.78% total contribution.
- Single employee contribute 31.97%
- 22.24% of the organization employee is Divorced

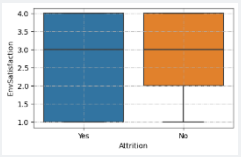
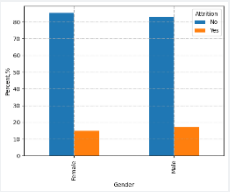
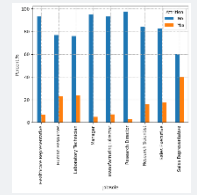
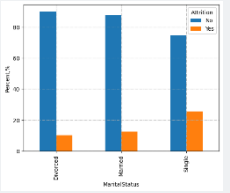
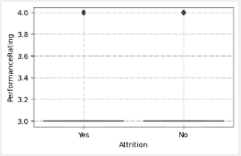
Multivariate Analysis

No	Input Variable	Business Question	Analytic Query	Analytics Result	Business Insight	
1	Age	Is Attrition influence by the Age of the employee ?	Is there any correlation between attrition and age ?	P-value: 8.356e-10 There is significant different between Age and Attrition	The attrition is affected by the Age. 50% of the attrition employee was from 28 to 39 years old.	
2	BusinessTravel	Is Attrition influence by the Business Travel of the employee ?	Is there any significant different of Attrition between Business Travel class ?	P-value: 5.6086E-06 There is significant different between Age and Attrition	Travel_frequently give the highest percentage of Attrition which is 25% and Non_travel only 8% who Attrition.	
3	MonthlyIncome	Is Attrition influence by the Monthly Income of the employee ?	Is there any correlation between attrition and montly income ?	P-value: 4.4336e-13 There is significant different between Age and Attrition	There are higher chances that monthly income below 6k attrition	
4	JobSatisfaction	Is Attrition influence by the Job Satisfaction of the employee ?	Is there any correlation between attrition and Job satisfaction ?	P-value: 4.4336e-13 There is significant different between Age and Attrition	Obviously, lower job satisfaction the higher attrition rate.	

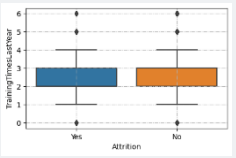
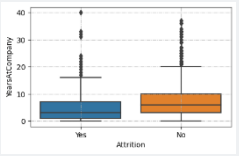
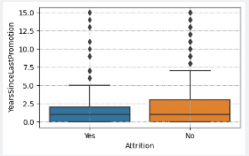
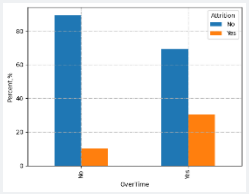
Multivariate Analysis

5	Bonus	Is Attrition influence by the Bonus that the employee received?	Is there any correlation between attrition and bonus received ?	P-value: 3.692e-14 There is significant different between Bonus and Attrition	There is higher chances of Attrition to employee that received below 20k bonus	
6	Department	Is Attrition influence by the Department of the employee work in ?	Is there any significant different of Attrition between each department ?	P-value: 0.004526 There is slightly significant different between Department and Attrition	Sales department is the highest percentage of Attrition which is higher that 20% of their staff attrition	
7	DistanceFromHome	Is Attrition influence by the distance from home of the employee ?	Is there any significant different of Attrition between distance from home ?	P-value: 0.00414 There is slightly significant different between Distance from home and Attrition	There are chances that long distance to work could attrition rather than short distance.	
8	Education	Is Attrition influence by the Education level of the employee ?	Is there any correlation between attrition and education level ?	P-value: 0.23 Accepting Hnull which is there is no relationship between education level with Attriton status.	The attrition is not affected by education level.	
9	EducationField	Is Attrition influence by the Education field of the employee ?	Is there any significant different of Attrition between each education field ?	P-value: 0.00677 There is slightly significant different between Education Field and Attrition	There is higher chances of employee Attrition from HR, Marketing and Technical Degree education field.	

Multivariate Analysis

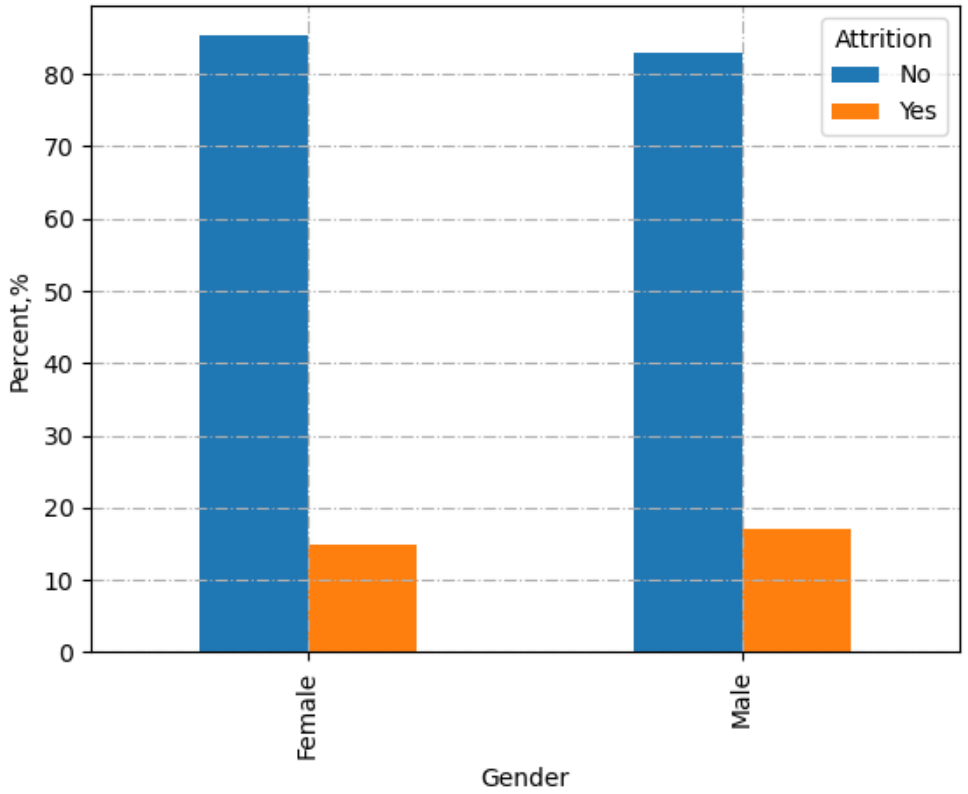
10	EnvSatisfaction	Is Attrition influence by the work enviroment satisfaction of the employee ?	Is there any correlation between attrition and enviroment satisfaction level ?	P-value: 0.00029 There is slightly significant different between Env Satisfaction and Attrition	Obviously, lower enviroment satisfaction the higher attrition rate.	
11	Gender	Is Attrition influence by the gender of the employee ?	Is there any significant different of Attrition between gender ?	P-value: 0.3 Accepting Hnull which is there is no relationship between performance rating with Attriton status.	The attrition is not affected by gender.	
12	JobRole	Is Attrition influence by the job role of the employee ?	Is there any significant different of Attrition between JobRole ?	P-value: 2.7525e-15 There is slightly significant different between JobRole and Attrition	Job role of Sales Representative contribute the highest percentage of Attrition which is almost 40% of the Sales Representative JobRole.	
13	MaritalStatus	Is Attrition influence by the maritial status of the employee ?	Is there any significant different of Attrition between Maritial Status ?	P-value: 9.4555e-11 There is slightly significant different between JobRole and Attrition	Single employee have higher chance to Attrition other than employee who were married and divorced.	
14	PerformanceRating	Is Attrition influence by the performance rating of the employee ?	Is there any correlation between attrition and performance rating ?	P-value: 0.91 Accepting Hnull which is there is no relationship between performance rating with Attriton status.	The attrition is not affected by performance rating.	

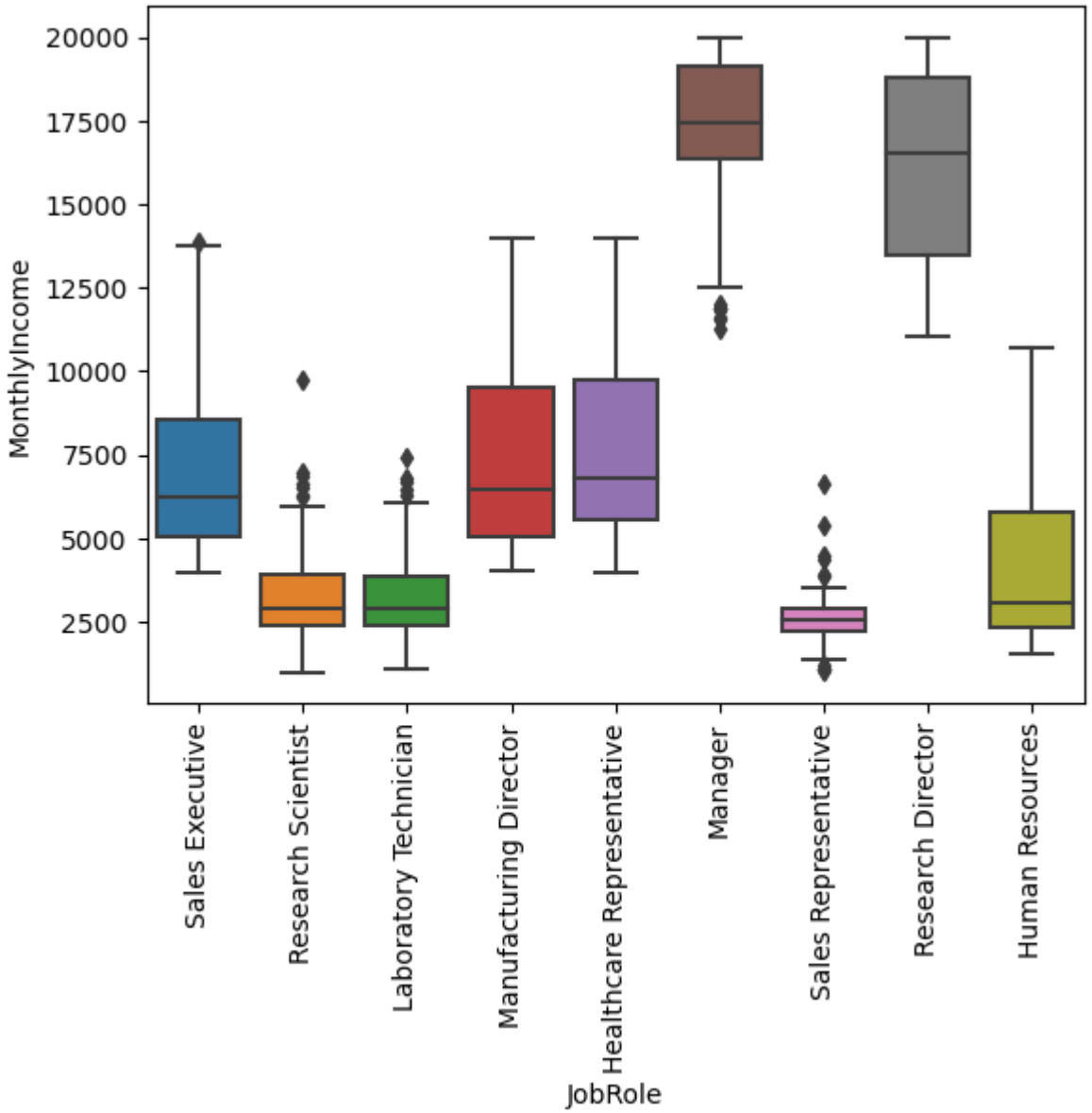
Multivariate Analysis

15	TrainingTimesLastYear	Is Attrition influence by the training time last year of the employee ?	Is there any correlation between attrition and TrainingTimesLastYear ?	P-value: 0.023 There is slightly significant different between TrainingTimesLastYear and Attrition	Training Times Last Year somehow effecting the Attrition rate	
16	YearsAtCompany	Is Attrition influence by the year at company of the employee ?	Is there any correlation between attrition and YearsAtCompany ?	P-value: 2.3189e-07 There is significant different between YearsAtCompany and Attrition	Low years at company is higher tend to Attrition	
17	YearsSinceLastPromotion	Is Attrition influence by the years since last promotion of the employee ?	Is there any correlation between attrition and performance rating ?	P-value: 0.205 Accepting Hnull which is there is no relationship between YearsSinceLastPromotion with Attriton status.	The attrition is not affected by yearssincelastpromotion.	
18	OverTime	Is Attrition influence by the overtime of the employee ?	Is there any significant different of Attrition between overtime ?	P-value: 8.156e-21 There is significant different between Overtime and Attrition	Employee that performing overtime higher chances to Attrition than employee that no overtime.	

Multivariate Analysis

- There is no relationship between both gender with Attrition
- **Monthly income vs job role**





Feature Selection

AGE

MONTHLY INCOME

JOB SATISFACTION

BONUS

DISTANCE FROM HOME

ENVIROMENT SATISFACTION

TRAINING TIMES LAST YEAR

YEARS AT COMPANY

BUSINESS TRAVEL

DEPARTMENT

EDUCATION FIELD

JOB ROLE

MARITAL STATUS

OVERTIME

~~GENDER~~

~~EDUCATION~~

~~PERFORMANCE RATING~~

~~YEARS SINCE LAST PROMOTION~~

Model Algorithm Selection

Machine learning algorithm

Logistic Regression

Attrition	Yes	No
row_0		
Yes	12	6
No	59	364

	precision	recall	f1-score	support
No	0.86	0.98	0.92	370
Yes	0.67	0.17	0.27	71
accuracy			0.85	441
macro avg	0.76	0.58	0.59	441
weighted avg	0.83	0.85	0.81	441

Naïve Bayes

Attrition	Yes	No
row_0		
Yes	42	104
No	29	266

	precision	recall	f1-score	support
No	0.90	0.72	0.80	370
Yes	0.29	0.59	0.39	71
accuracy			0.70	441
macro avg	0.59	0.66	0.59	441
weighted avg	0.80	0.70	0.73	441

Prescriptive Analytics

Targeted Retention Strategies

•**Action:** Focus on implementing targeted retention strategies for employees who are predicted to leave (attrition: yes).
•**Insight:** The model's recall of 0.59 suggests that it captures a significant portion of actual attrition cases. Prioritize resources to retain these individuals, as they are at a higher risk of leaving

Exit Interviews and Feedback

•**Action:** Conduct exit interviews and gather feedback from employees who actually leave (attrition: yes).
•**Insight:** The model's precision of 0.29 indicates room for improvement in identifying employees who will leave. Exit interviews can help identify the reasons behind attrition and guide improvements.

Workload balancing

•**Action:** Conduct workload balancing study.
•**Insight:** Based on EDA there is high significant different of Attrition between employee who don’t work overtime and employee that work overtime.

Hybrid Workplace

•**Action:** Deployed hybrid workplace for the employee.
•**Insight:** Higher distance from home contribute higher chance of Attrition. With hybrid workplace, employee able work at home and office when required.

Prescriptive Analytics

New data feature

•**Action:** Addon new feature data
•**Insight:** The model can be improved by adding new feature data for example sick leave , total working hours per month and etc. This will increase the precision ,sensitivity and accuracy of the model.