
Employee Attrition

SHREYA GHORPADE, PEEYUSHA GOTETI AND UNNIKRISHNAN SIR →

*True elegance
is defined by
its simplicity*

CONTENT

1. Introduction
2. Motivation
3. Objectives
4. Methodology
5. Result & Discussion
6. Conclusion



Introduction

What is Employee Attrition?

Employee Attrition refers to the percentage of workers who leave an organization and are replaced by new employees. A high rate of attrition in an organization leads to increased recruitment, hiring and training costs. Not only it is costly, but qualified and competent replacements are hard to find. In most industries, the top 20% of people produce about 50% of the output.

How the prediction of the analysis will be useful?

Employee attrition analytics is specifically focused on identifying why employees voluntarily leave, what might have prevented them from leaving, and how we can use data to predict attrition risk. Most importantly, this type of employee predictive analytics can be used to help organizations understand and design the interventions that will be most effective in reducing unwanted attrition.

Motivation

Attrition is commonly used to describe the deliberate downsizing of a company's workforce.
Downsizing happens when employees resign or retire.

Objectives

- The main objective of this research work is to develop a model that can help to predict whether an employee will leave the company or not.
- To identify the factors which makes the employee satisfactory to work in the organization.

Methodology

1. COLLECTING THE DATA FOR BUSINESS DECISION-MAKING, STRATEGIC PLANNING, RESEARCH AND OTHER PURPOSES.

2. CHECKING FOR MISSING AND NULL VALUES

3. EDA USING MATPLOTLIB AND SEABORN LIBRARIES.

4. SPLITTING THE DATA INTO TRAINING TESTING DATASETS.

5. SPECULATING ACCURACY USING RANDOM FOREST CLASSIFIER.

6. PREDICTING THE TARGETED VARIABLE.

DATA PRE-PROCESSING

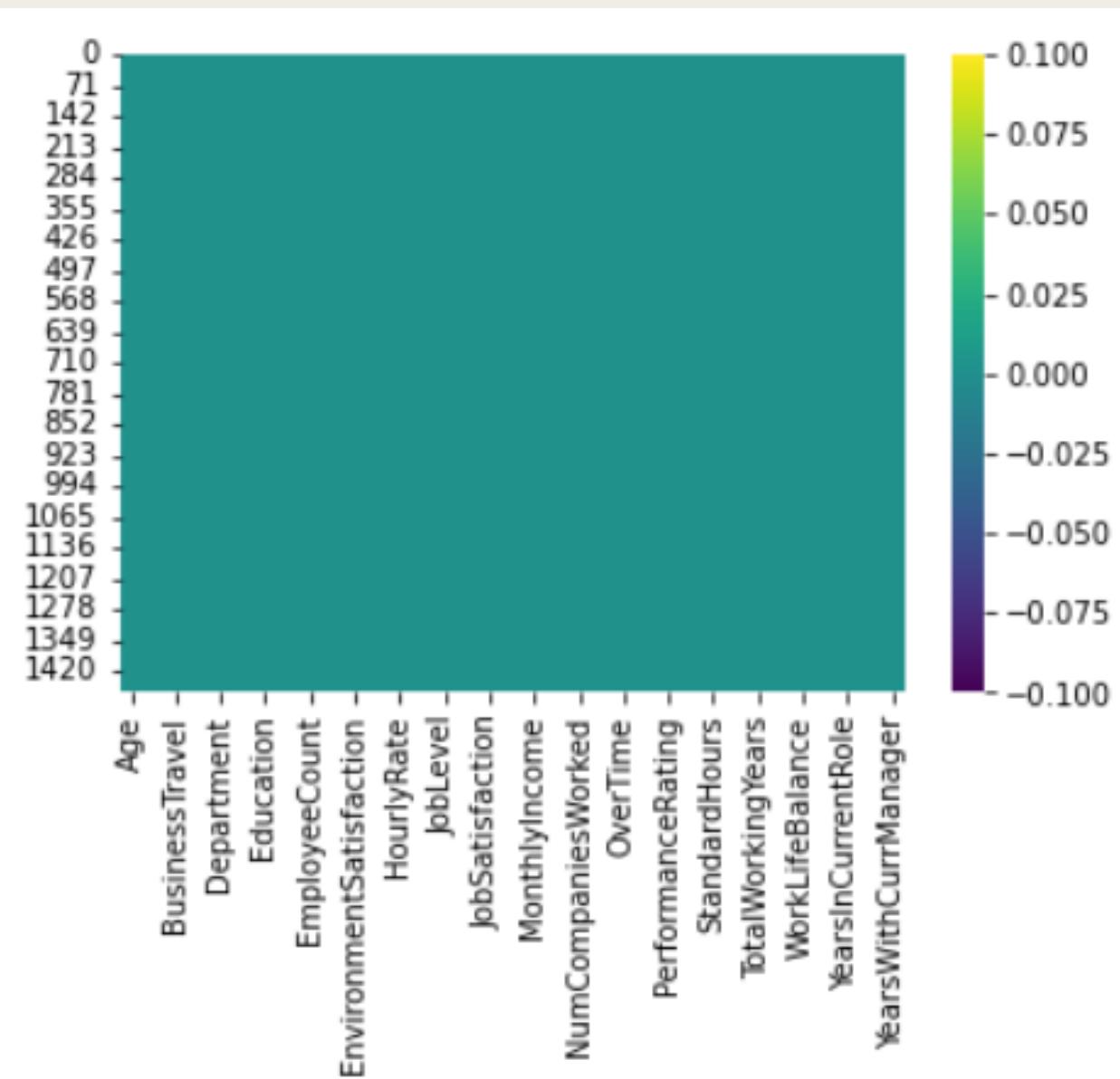
- Data obtained from Kaggle.
- Language used for prediction :- Python
- Tool used for prediction :- Jupyter Notebook
- Integrated Development Environment :- Python IDLE
- Importing libraries such as numpy, pandas, matplotlib and seaborn.

- After importing the dataset in the Python Notebook, we examined for null values and no null values were found.

- By executing this code

```
sns.heatmap(df.isnull(),cmap='viridis')
```

We can observe that there are no null values present.



EXPLORATORY DATA ANALYSIS

Exploratory data analysis (EDA) is a term for certain kinds of initial analysis and findings done with data sets, usually early on in an analytical process.

Why is data exploration necessary?

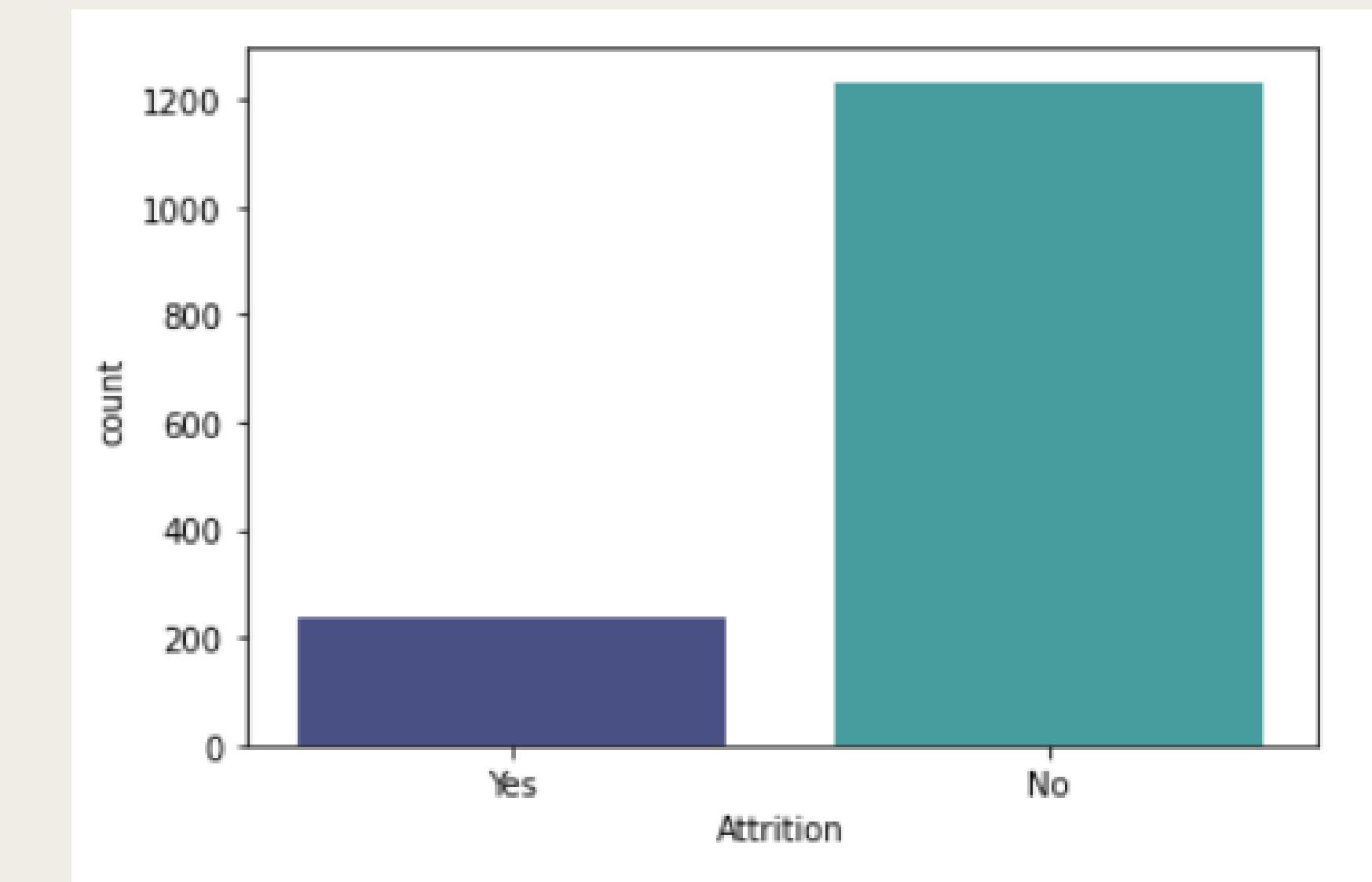
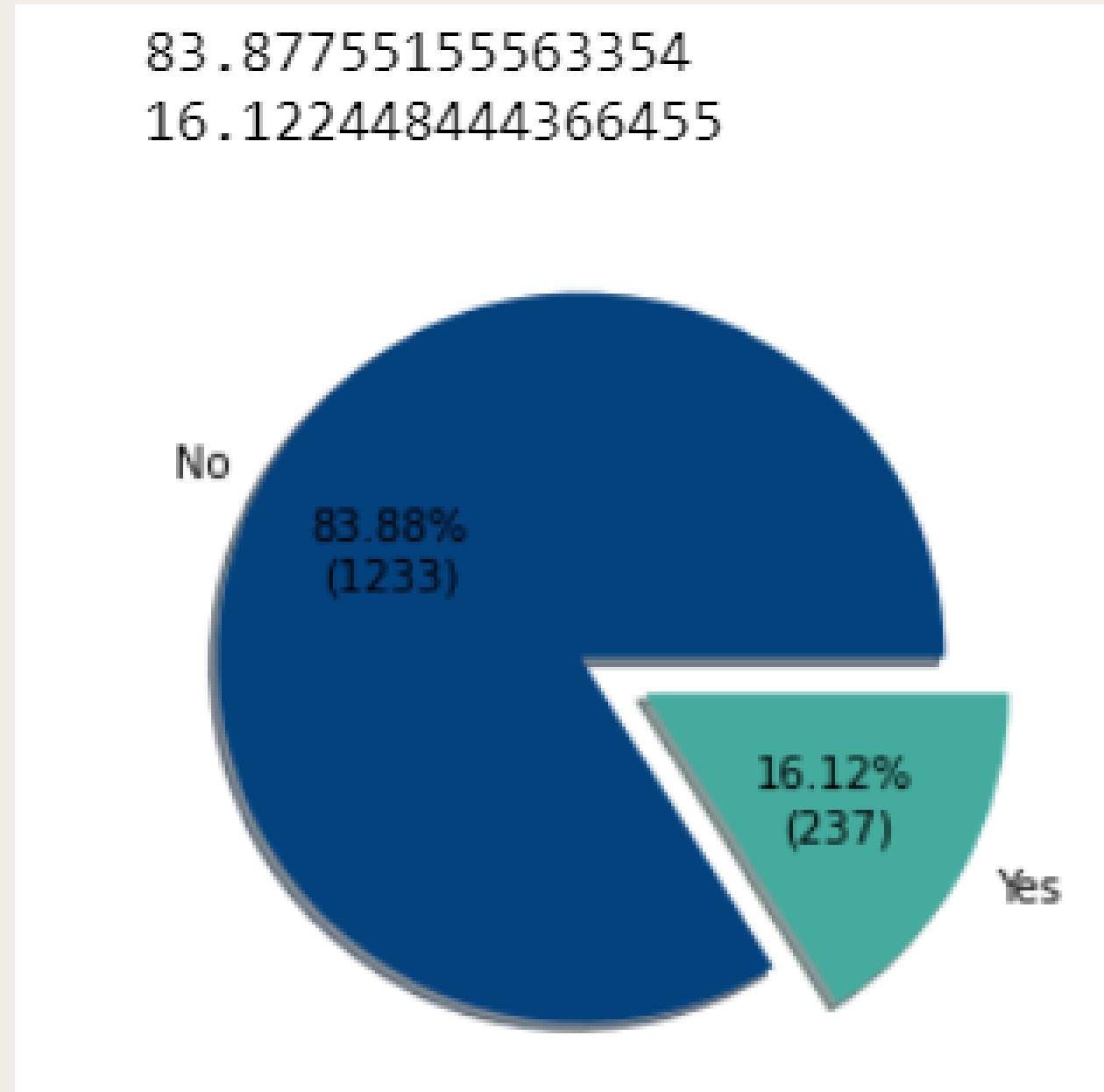
Two key advantages of data exploration are...

- To enable unexpected discoveries in the data.
- To foster a deep understanding of the data as an important fundamental for successful and efficient data science projects.

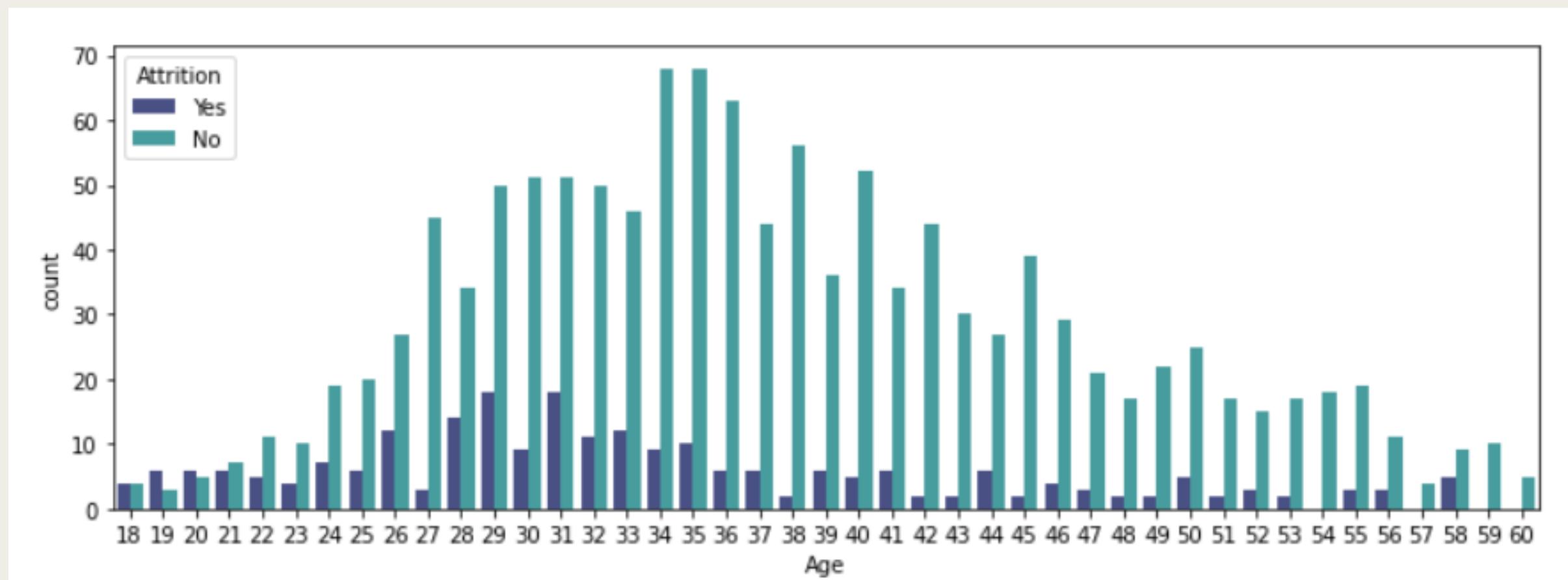
ATTRITION COUNT

The number of employees present in the company :- 1233

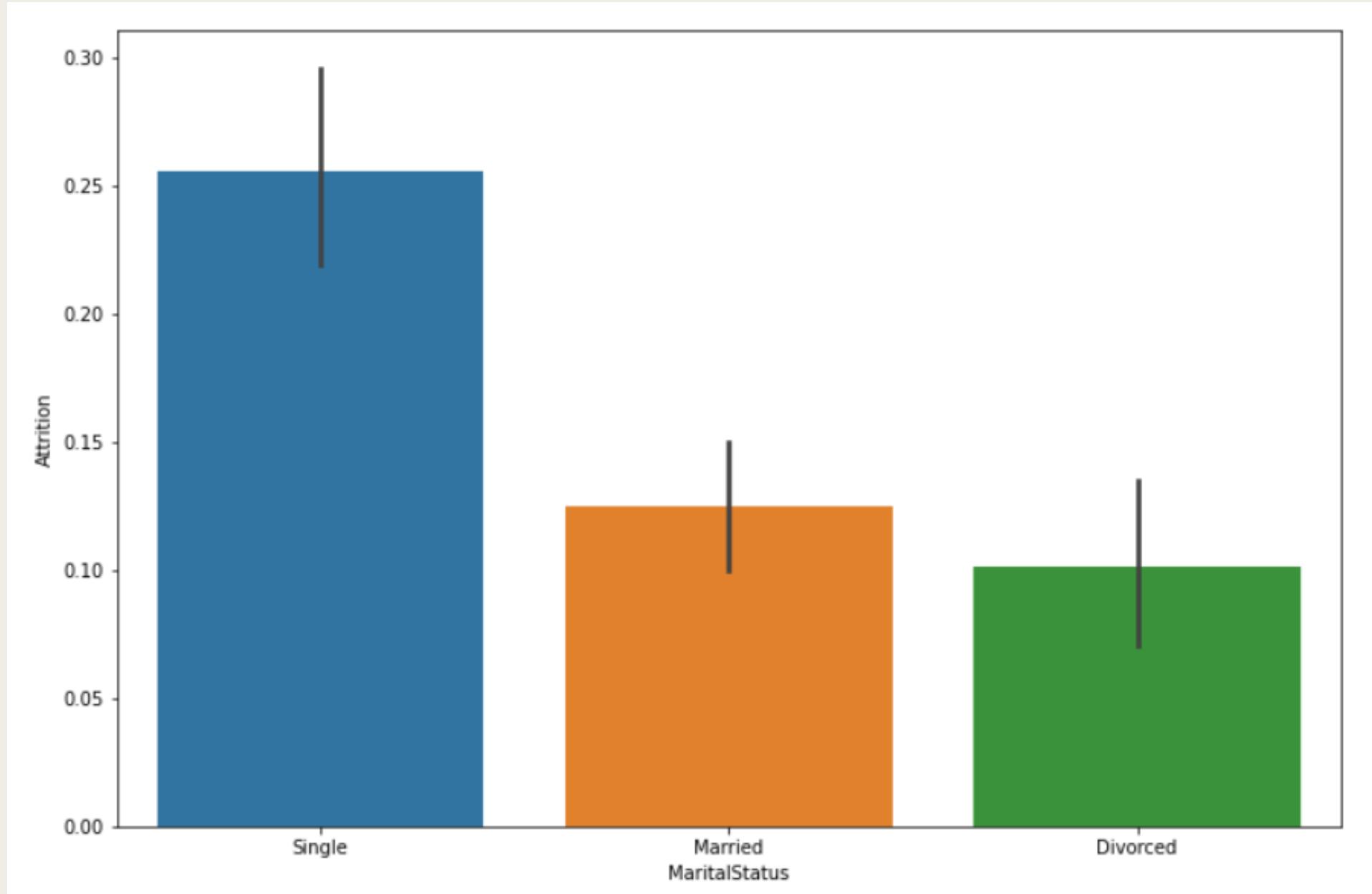
The number of employees left in the company :- 237



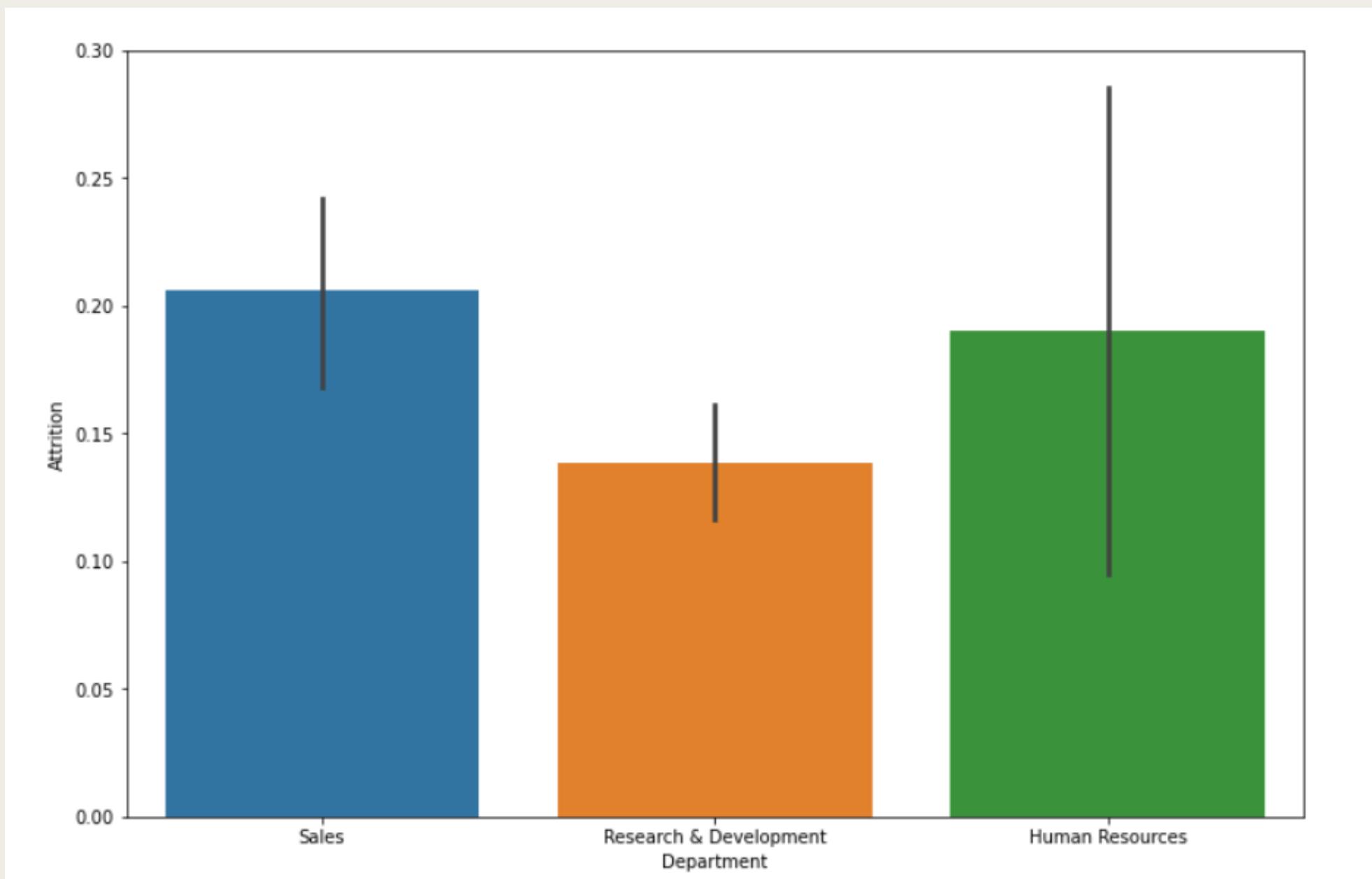
#SHOWS THE NUMBER OF EMPLOYEES THAT LEFT AND STAYED BY AGE



#SHOWS THE RELATION BETWEEN ATTRITION AND MARITAL STATUS



#SHOWS THE RELATION BETWEEN ATTRITION AND DEPARTMENT



Result

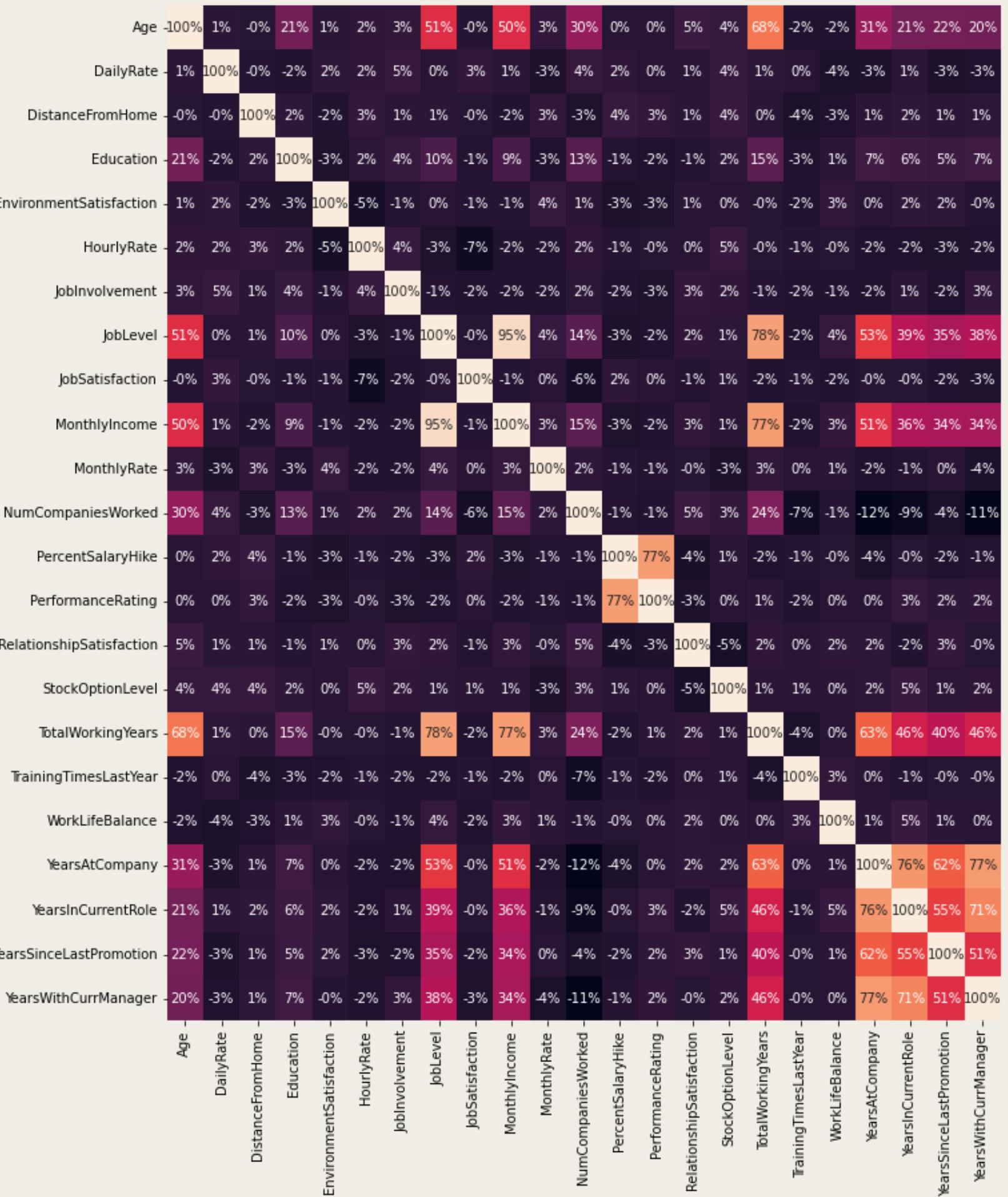
And

Discussion

CORRELATION

CORRELATION IS A STATISTICAL
MEASURE THAT EXPRESSES THE EXTENT
TO WHICH TWO VARIABLES ARE
LINEARLY RELATED.

IT'S A COMMON TOOL FOR
DESCRIBING SIMPLE RELATIONSHIPS
WITHOUT MAKING A STATEMENT
ABOUT CAUSE AND EFFECT.



- Dropping the unwanted columns.
- Converting the categorical columns to numeric form with the help of Label Encoder.

```
from sklearn.preprocessing import LabelEncoder  
  
for column in df.columns:  
    if df[column].dtype == np.number:  
        continue  
    df[column] = LabelEncoder().fit_transform(df[column])
```

- Separating the targeted variable from other columns
- Splitting the data into two parts, training and testing datasets.

Training DataSet is 75%

Testing DataSet is 25%

RANDOM FOREST CLASSIFIER

RANDOM FOREST IS A SUPERVISED MACHINE LEARNING ALGORITHM THAT IS USED WIDELY IN CLASSIFICATION AND REGRESSION PROBLEMS. IT BUILDS DECISION TREES ON DIFFERENT SAMPLES AND TAKES THEIR MAJORITY VOTE FOR CLASSIFICATION AND AVERAGE CASE OF REGRESSION.

HOW DOES THE ALGORITHM WORK?

- Select random samples from a given dataset.
- Construct a decision tree for each sample and get a prediction result from each decision tree.
- Perform a vote for each predicted result
- Select the prediction result with the most votes as the final prediction.

- Importing Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
```

- The Actual Values and The Predicted Values

```
Actual Values :- [0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 1]
```

```
Predicted Values :- [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]
```

- The Model Accuracy is **86%**.

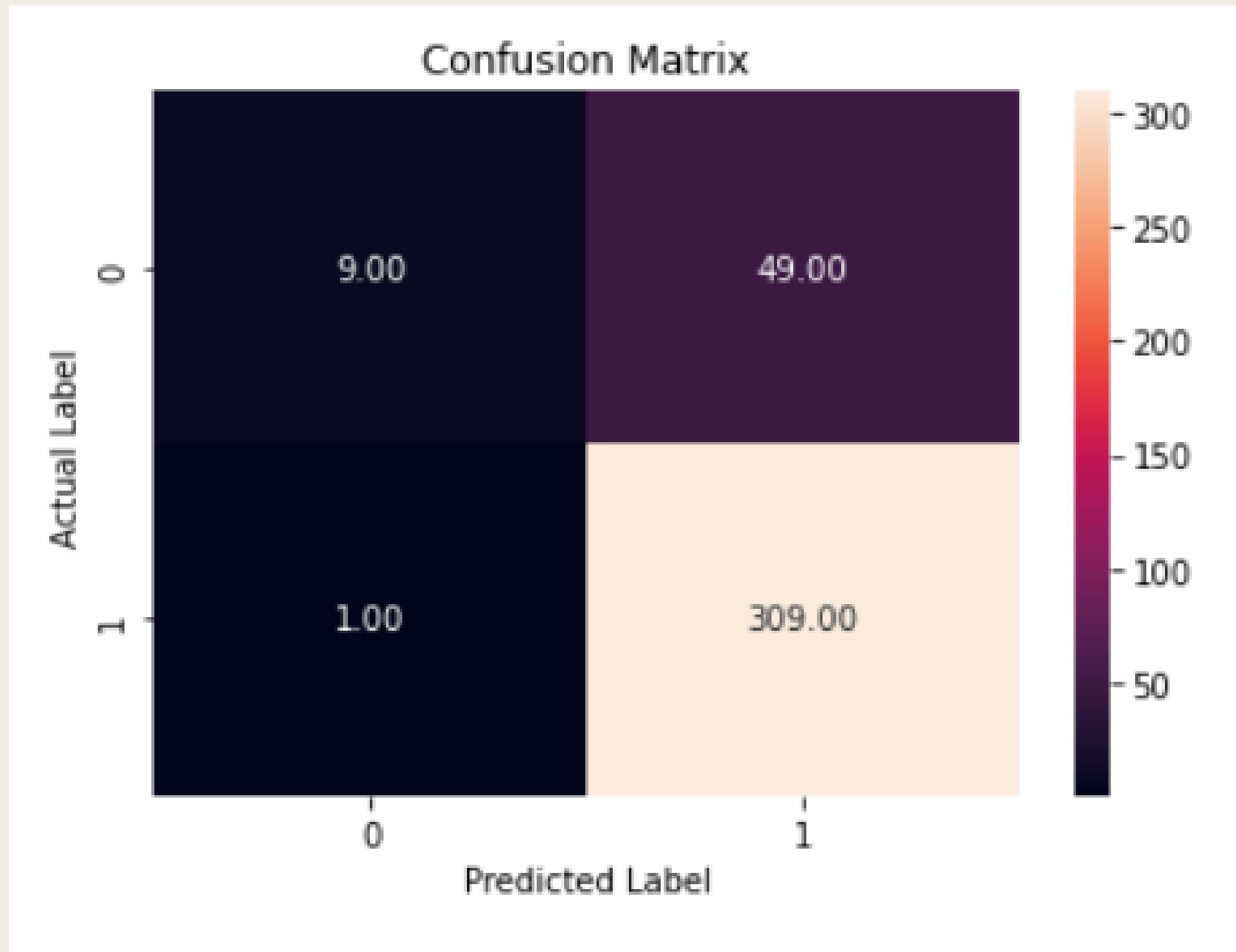
CONFUSION MATRIX

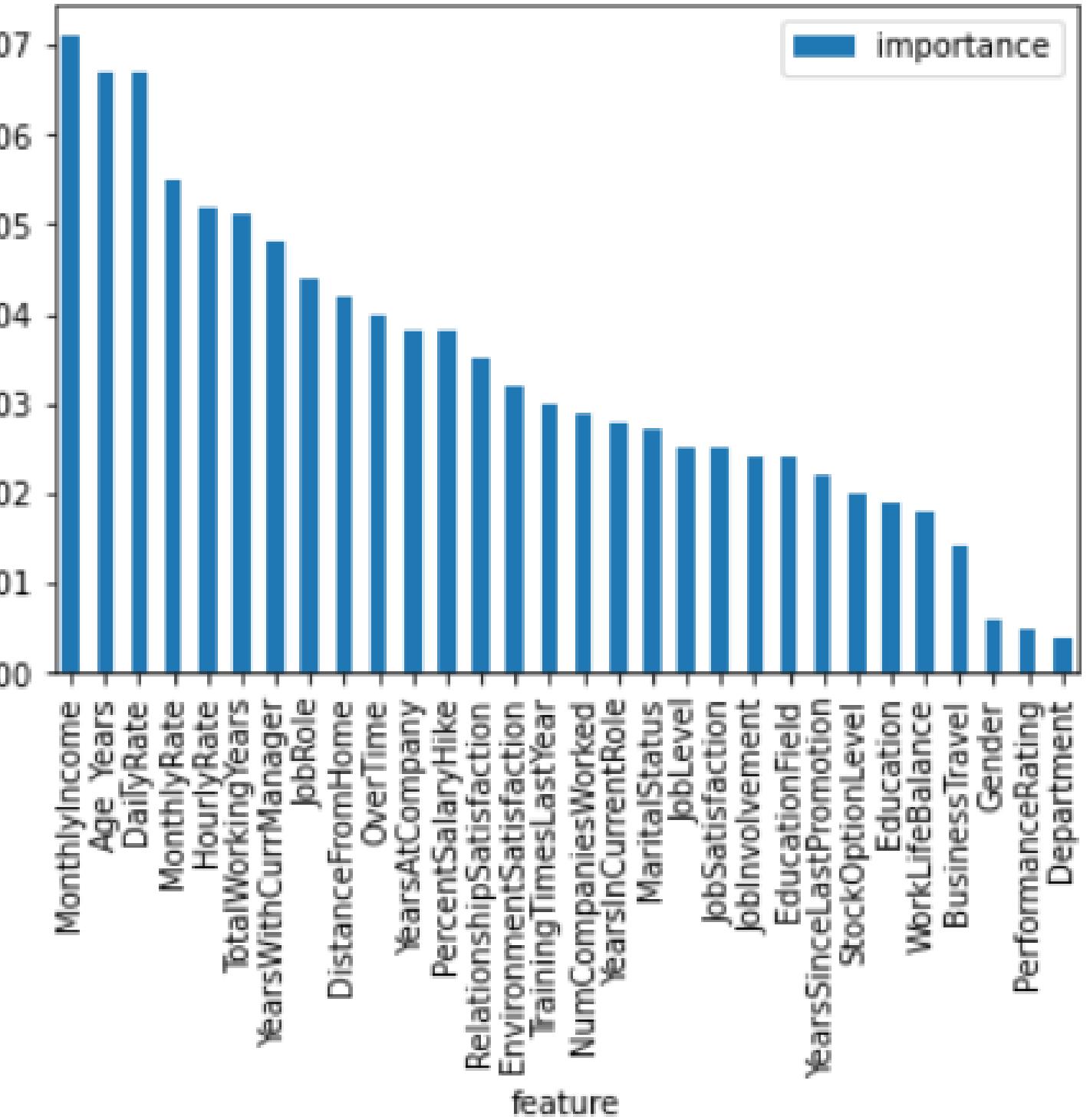
A CONFUSION MATRIX IS A TABLE THAT IS OFTEN USED TO DESCRIBE THE PERFORMANCE OF A CLASSIFICATION MODEL (OR "CLASSIFIER") ON A SET OF TEST DATA FOR WHICH THE TRUE VALUES ARE KNOWN.

		—Actual—	
		Has Heart Disease	Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	False Positives
	Does Not Have Heart Disease	False Negatives	True Negatives

		Actual Values
		Positive (1) Negative (0)
Predicted Values	Positive (1)	TP FP
	Negative (0)	FN TN

#CONFUSION MATRIX USING RANDOM FOREST CLASSIFIER





RANDOM FORESTS CAN BE USED TO RANK THE IMPORTANCE OF VARIABLES IN A CLASSIFICATION. AFTER BEING FIT, THE MODEL PROVIDES A FEATURE IMPORTANCES PROPERTY THAT CAN BE ACCESSED TO RETRIEVE THE RELATIVE IMPORTANCE SCORES FOR EACH INPUT FEATURE. HENCEFORTH, IN THIS CASE MONTHLY INCOME HAS HIGH IMPORTANCE SCORE FOLLOWED BY OTHERS.

UNKNOWN DATA PREDICTION

- Using a Random Data converting it into numpy array.
- Assigning the unknown data and predict if the employee left the company or not.

No = 0

Yes = 1

```
► if(prediction[0]==0):
    print("The employee left the company.")
else:
    print("The employee is currently working in the company.")
```

The employee left the company.

CONCLUSION

- WITH THE HELP OF THE MODEL WE PREDICTED IF THE EMPLOYEE WITH LEAVE THE COMPANY OR NO.
- WE FOUND THE SATISFACTORY FACTOR WHICH MAKES THE EMPLOYEE STAY IN THE COMPANY.

Thank You :)