EMPLOYEES ATTRITION

TALENT RETENTION TASKFORCE

- Debanjan Nanda
- Ayan Maity

AGENDA

Literature Survey

Data Gathering

Data Preprocessing

Exploratory Data Analysis

Model Selection

WHAT IS EMPLOYEE ATTRITION

Employee attrition refers to the gradual loss of employees over time. This can happen when employees leave the company for various reasons, such as finding a new job, retiring, or personal issues. High attrition rates can be costly for companies because they need to spend time and money to hire and train new workers. Understanding why employees leave can help companies make changes to improve job satisfaction and retain their staff longer.

WHAT ?

- This project aims to predict whether employees will leave their jobs or not.
- We'll gather data on various employee attributes such as age, job role, satisfaction levels, and tenure to train predictive models.
- These models will utilize advanced machine learning techniques of classification to make accurate predictions.
- By analyzing historical data, we'll teach the models to identify patterns associated with turnover and tenure.
- Ultimately, the goal is to develop reliable tools that assist companies in understanding and managing their workforce dynamics effectively.

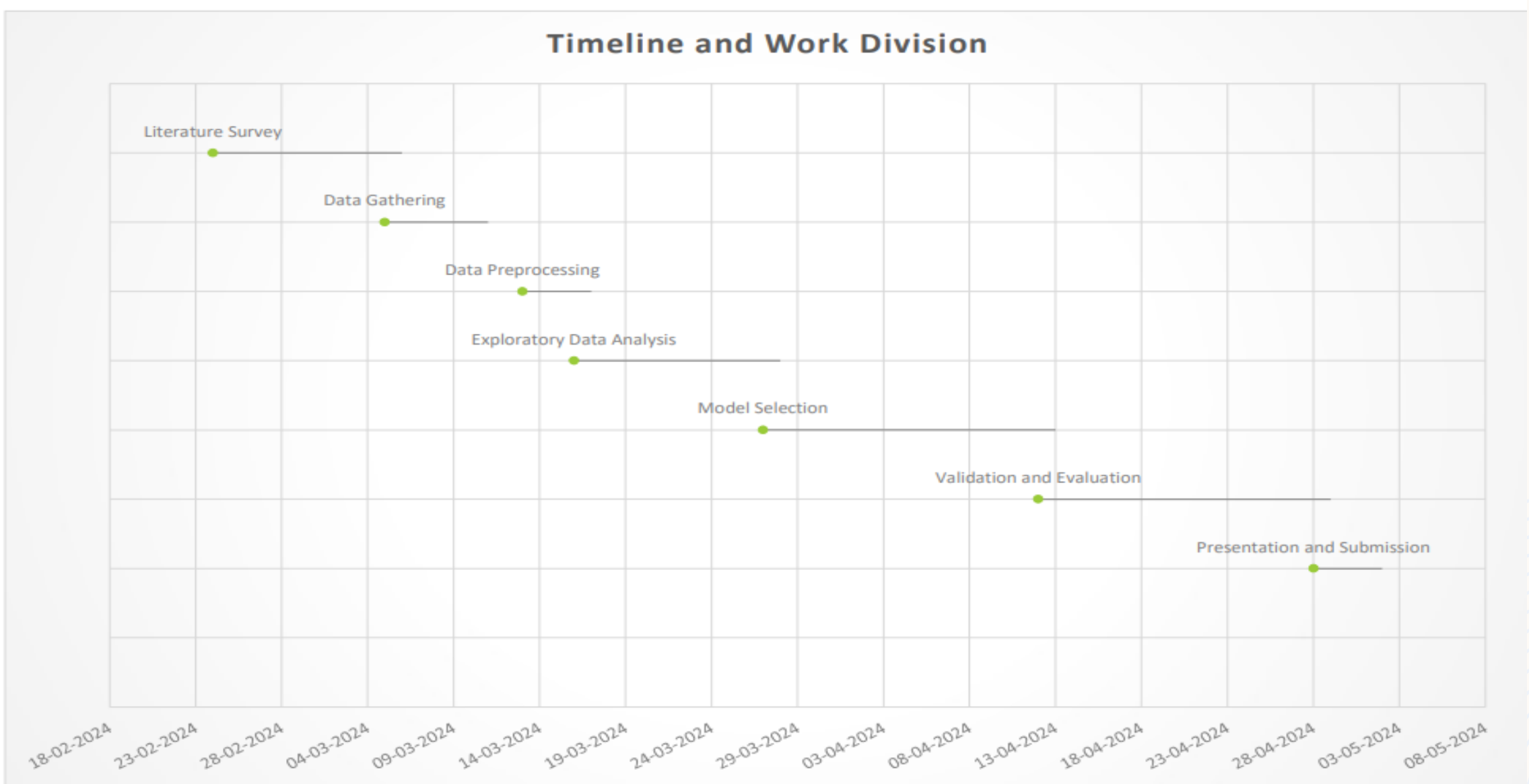
WHY ?

- Employee attrition projects are essential for companies to understand why employees are leaving their jobs.
- Understanding the reasons behind employee departures allows companies to implement changes that encourage employee retention.
- By reducing employee turnover, companies can save money that would otherwise be spent on hiring and training new employees.
- Longer employee tenure leads to increased skill and expertise, benefiting the company's overall productivity and success.
- Ultimately, decreasing turnover contributes to a more positive and stable work environment, improving morale and overall workplace satisfaction.

HOW ?

- The project will employ advanced machine learning techniques to analyze the collected data and develop predictive models.
- classification models like logistic regression, decision trees, random forest can be utilized to predict employee attrition.

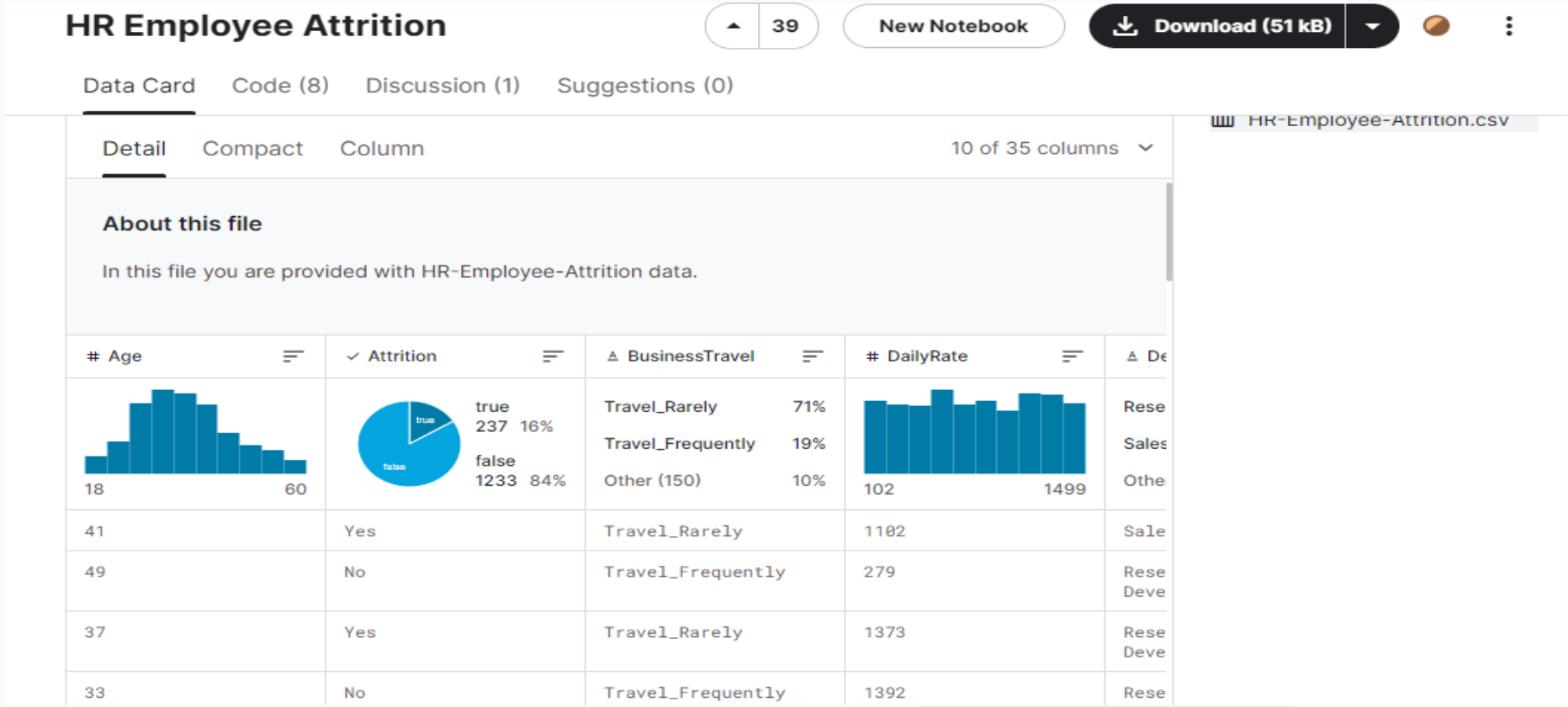
TIMELINE AND WORK DIVISION:



LITERATURE SURVEY

Author	Paper	Data	Method	Result
Ali Raza ,Kashif Munir , Mubarak Almutairi, Faizan Younas and Mian Muhammad Sadiq Fareed	Predicting Employee Attrition Using Machine Learning	IBM HR employee attrition dataset: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset	Four machine learning models—Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree Classifier (DTC), and Extra Trees Classifier (ETC)—were employed for prediction.	The study evaluated the models based on various metrics such as accuracy, precision, recall, F1 score, and ROC curve. The ETC model demonstrated superior performance across these metrics. This model achieved the highest accuracy score of 93%.
By Alao D. & Adeyemo A. B.	ANALYZING EMPLOYEE ATTRITION USING DECISION TREE ALGORITHMS	Employees in a Higher Institution in South-West Nigeria: https://iiste.org/Journals/index.php/CIS/article/view/10148	The authors employed classification techniques, specifically decision tree algorithms, to develop prediction models for employee attrition.	The performance of the developed models was assessed using various performance metrics such as True Positive Rate (TP Rate), False Positive Rate (FP Rate), Precision, F-Measure, and Receiver Operating Characteristic (ROC) curve. The decision tree model achieved the best accuracy level of almost 75%.

DATA GATHERING



https://www.kaggle.com/datasets/itssuru/hr-employee-attrition

Our Dataset contains 35 columns . The description of the columns are given below :

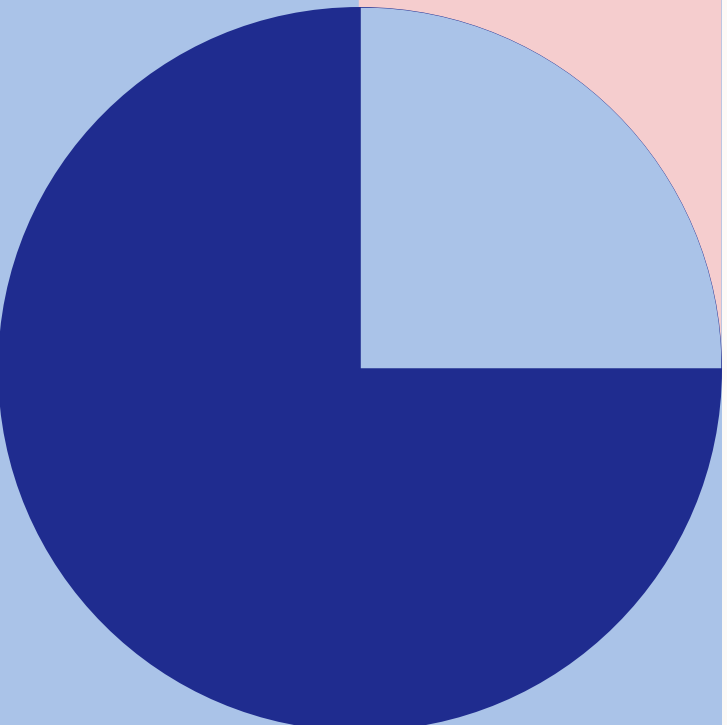
- Age: The age of the employee.
- Attrition: Indicates if the employee has left the company ('Yes' or 'No').
- BusinessTravel: Frequency of business travel ('Travel_Rarely', 'Travel_Frequently', 'Non-Travel').
- DailyRate: The daily rate of the employee.
- Department: The department in which the employee works ('Sales', 'Research & Development', 'Human Resources').
- DistanceFromHome: The distance from the employee's home to the workplace.
- Education: Education level (1 'Below College', 2 'College', 3 'Bachelor', 4 'Master', 5 'Doctor').
- EducationField: Field of education (e.g., 'Life Sciences', 'Medical', 'Marketing').
- EmployeeCount: Always 1 (constant value).
- EmployeeNumber: An Employee Number is a unique number that has been assigned to each current and former State employee and elected official in the Position and Personnel DataBase.
- EnvironmentSatisfaction: Satisfaction with the work environment (1 'Low', 2 'Medium', 3 'High', 4 'Very High').
- Gender: Gender of the employee ('Male' or 'Female').
- HourlyRate: The hourly rate of the employee.
- JobInvolvement: The level of involvement in the job (1 'Low', 2 'Medium', 3 'High', 4 'Very High').
- JobLevel: Job level within the organization.
- JobRole: The role of the employee (e.g., 'Sales Executive', 'Research Scientist').
- JobSatisfaction: Job satisfaction level (1 'Low', 2 'Medium', 3 'High', 4 'Very High').
- MaritalStatus: Marital status of the employee ('Single', 'Married', 'Divorced').
- MonthlyIncome: Monthly income of the employee.
- MonthlyRate: The monthly rate of the employee.

- NumCompaniesWorked: Number of companies the employee has worked for.
- Over18: Whether the employee is over 18 years old (Y).
- OverTime: Whether the employee works overtime ('Yes' or 'No').
- PercentSalaryHike: Percent increase in salary.
- PerformanceRating: Performance rating (1 'Low', 2 'Good', 3 'Excellent', 4 'Outstanding').
- RelationshipSatisfaction: Satisfaction with relationships at work (1 'Low', 2 'Medium', 3 'High', 4 'Very High').
- StandardHours: Standard working hours (constant value of 80).
- StockOptionLevel: Employee stock options, also known as ESOs, are stock options in the company's stock granted by an employer to certain employees. Typically they are granted to those in management or officer-level positions. Stock options give the employee the right to buy a certain amount of stock at a specific price, during a specific period of time. Options typically have expiration dates as well, by which the options must have been exercised, otherwise they will become worthless.
- TotalWorkingYears: Total number of years the employee has worked.
- TrainingTimesLastYear: Number of training sessions attended last year.
- WorkLifeBalance: Work-life balance rating (1 'Bad', 2 'Good', 3 'Better', 4 'Best').
- YearsAtCompany: Number of years the employee has been at the company.
- YearsInCurrentRole: Number of years the employee has been in the current role.
- YearsSinceLastPromotion: Number of years since the employee's last promotion.
- YearsWithCurrManager: Number of years the employee has worked with the current manager.

DATA PREPROCESSING

1. Data Loading and Overview:
    - Loaded dataset and displayed basic information such as shape, first few rows, summary statistics, and data types.
2. Handling Categorical Variables:
    - Converted categorical variables ('Attrition', 'OverTime', 'Over18') into numerical form.
    - 'Attrition': Converted 'Yes' to 1 and 'No' to 0.
    - 'OverTime': Converted 'Yes' to 1 and 'No' to 0.
    - 'Over18': Converted 'Y' to 1 and 'N' to 0.
3. Check for Missing Values:
    - Verified absence of missing data, ensuring data completeness and reliability.
4. Data Quality Assurance:
    - Ensured consistency and reliability of dataset by confirming no missing values.
    - Validated correctness of data transformations for categorical variables.

EXPLORATORY DATA ANALYSIS



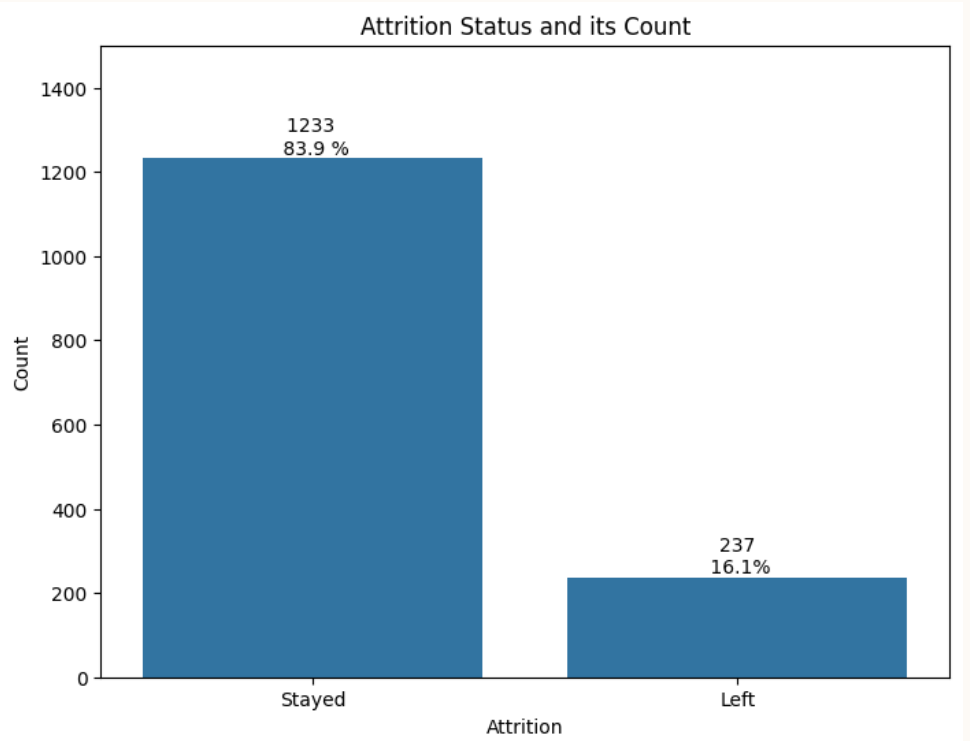
Total Employees: 1470

Number of employees who left: 237

% of employees who left: 16.12%

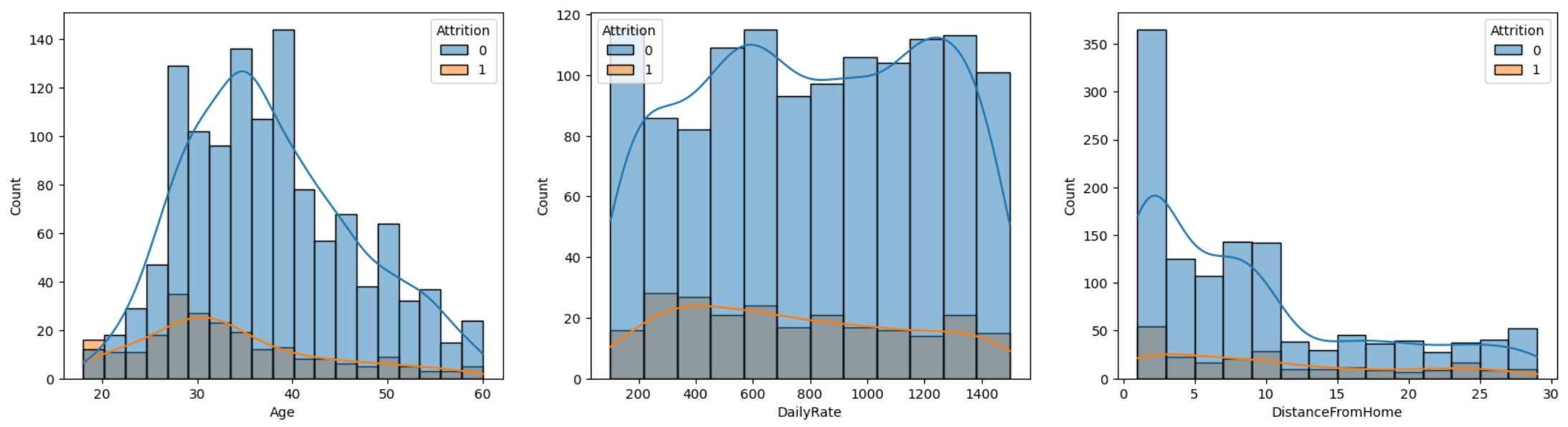
Number of employees who stayed: 1233

% of employees who stayed: 83.88%




Age
Attrition
DailyRate
DistanceFromHome
Education
EmployeeCount
EmployeeNumber
EnvironmentSatisfaction
HourlyRate
JobInvolvement
JobLevel
JobSatisfaction
MonthlyIncome
MonthlyRate
NumCompaniesWorked
Over18
OverTime
PercentSalaryHike
PerformanceRating
RelationshipSatisfaction
StandardHours
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager

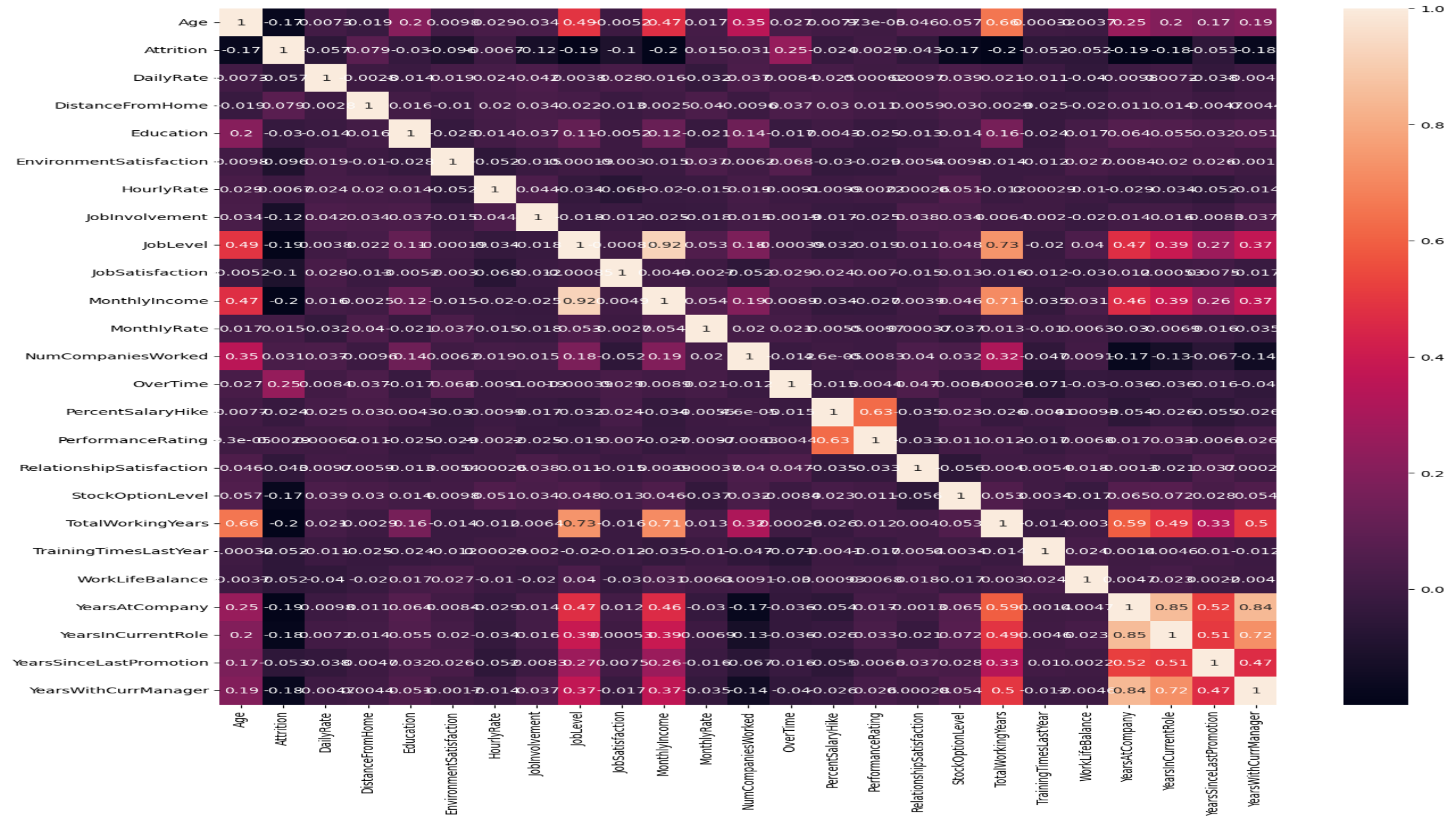
- *plot histogram for each numeric variable/feature of the dataset.*
- *Several features such as 'MonthlyIncome' and 'TotalWorkingYears' are heavy tailed.*
- *It makes sense to drop 'EmployeeCount' and 'Standardhours' since they do not change from one employee to the other.*



- Removed 'EmployeeCount', 'StandardHours', 'Over18', and 'EmployeeNumber' columns as they contain constant values and do not contribute to the analysis.

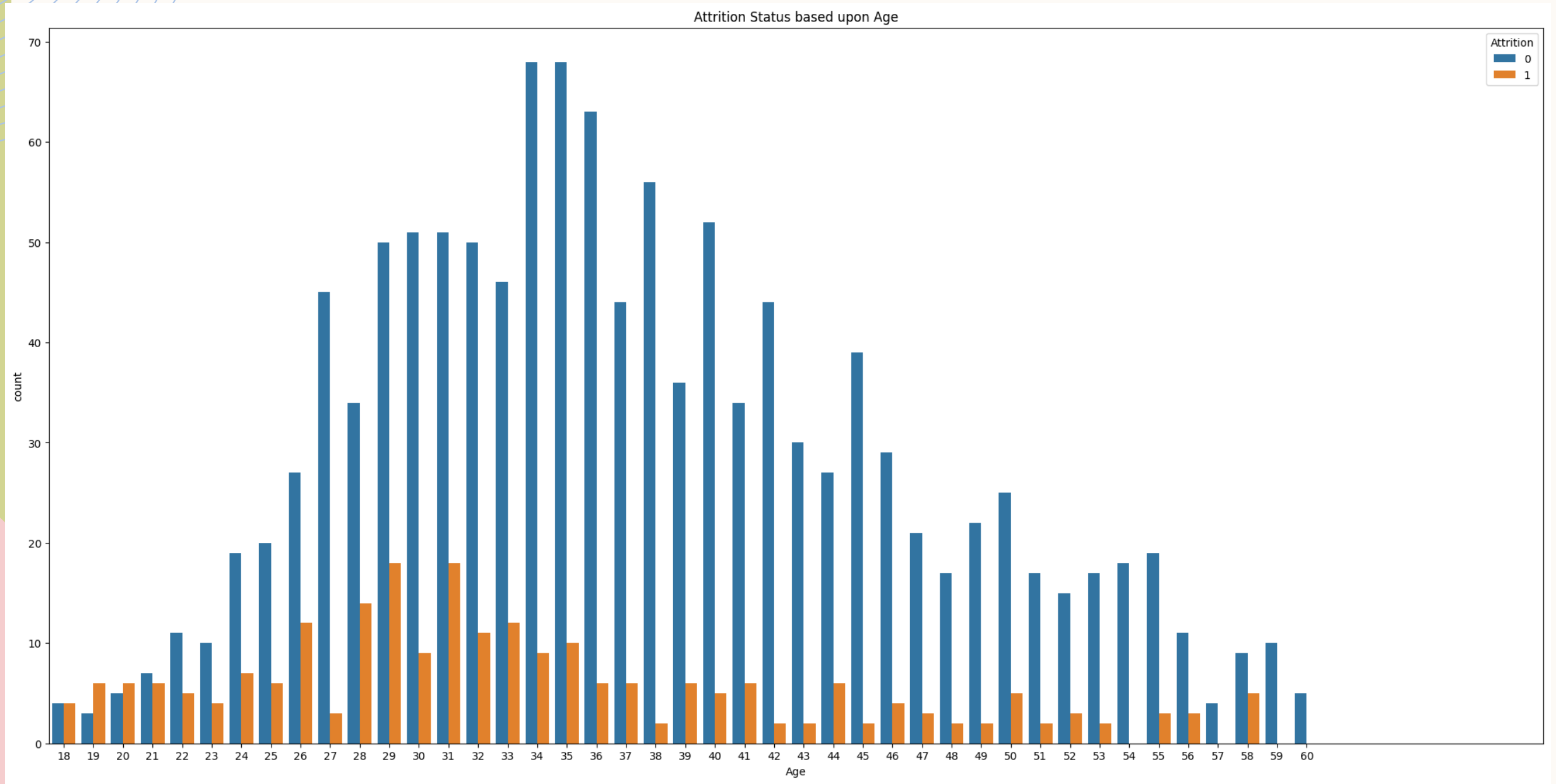
Comparing Mean and Standard Deviation:

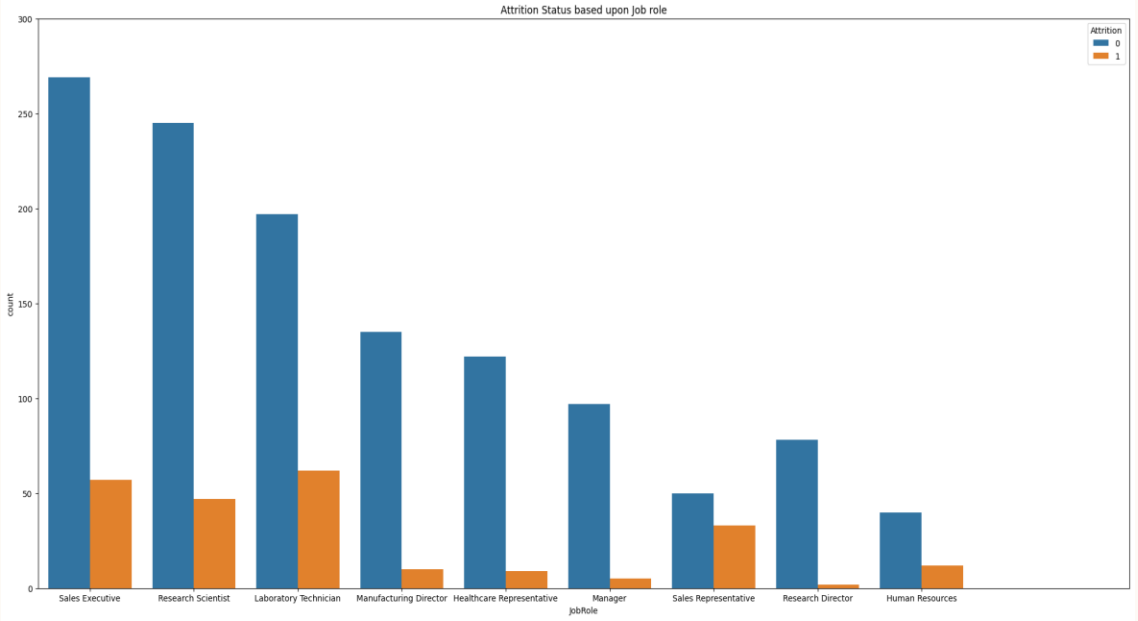
- Age: Employees who remained in the company have a higher mean age compared to those who left.
- DailyRate: The daily rate of employees who stayed is notably higher.
- DistanceFromHome: Employees who stayed tend to reside closer to their workplace.

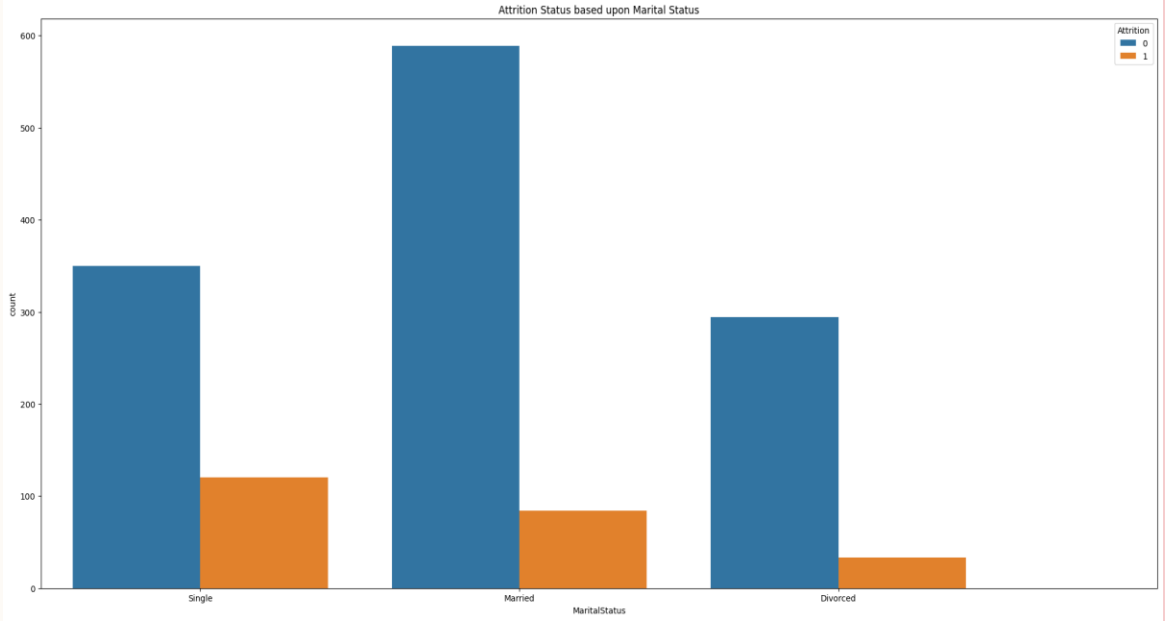

Age
Attrition
DailyRate
DistanceFromHome
Education
EnvironmentSatisfaction
HourlyRate
JobInvolvement
JobLevel
JobSatisfaction
MonthlyIncome
MonthlyRate
NumCompaniesWorked
OverTime
PercentSalaryHike
PerformanceRating
RelationshipSatisfaction
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager
1.0
0.8
0.6
0.4
0.2
0.0

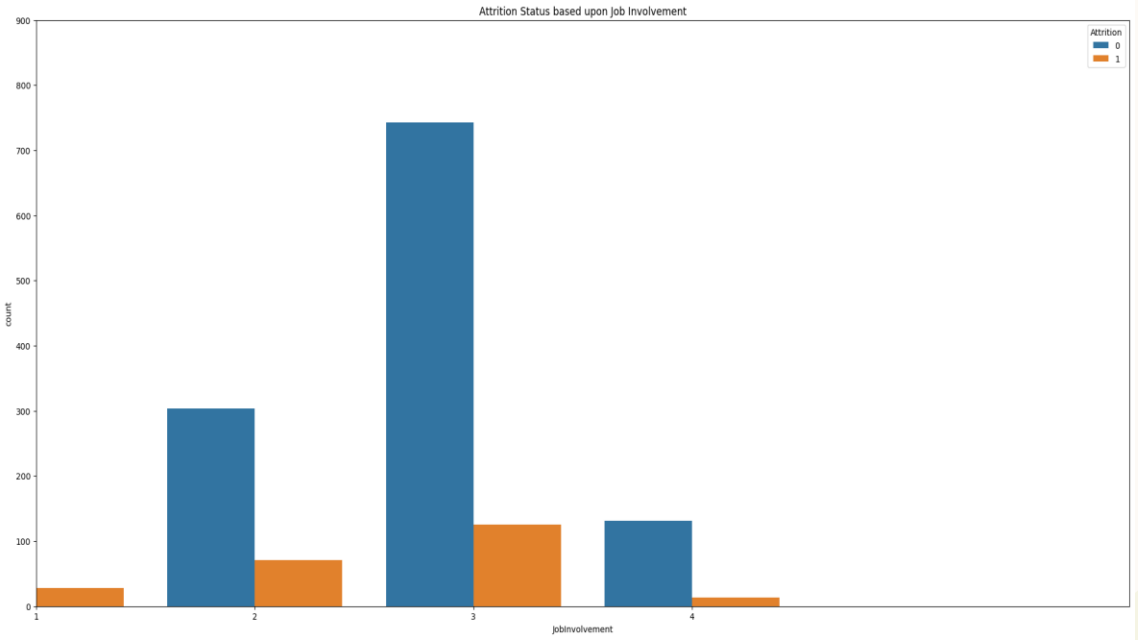
Key Correlations in Employee Data:

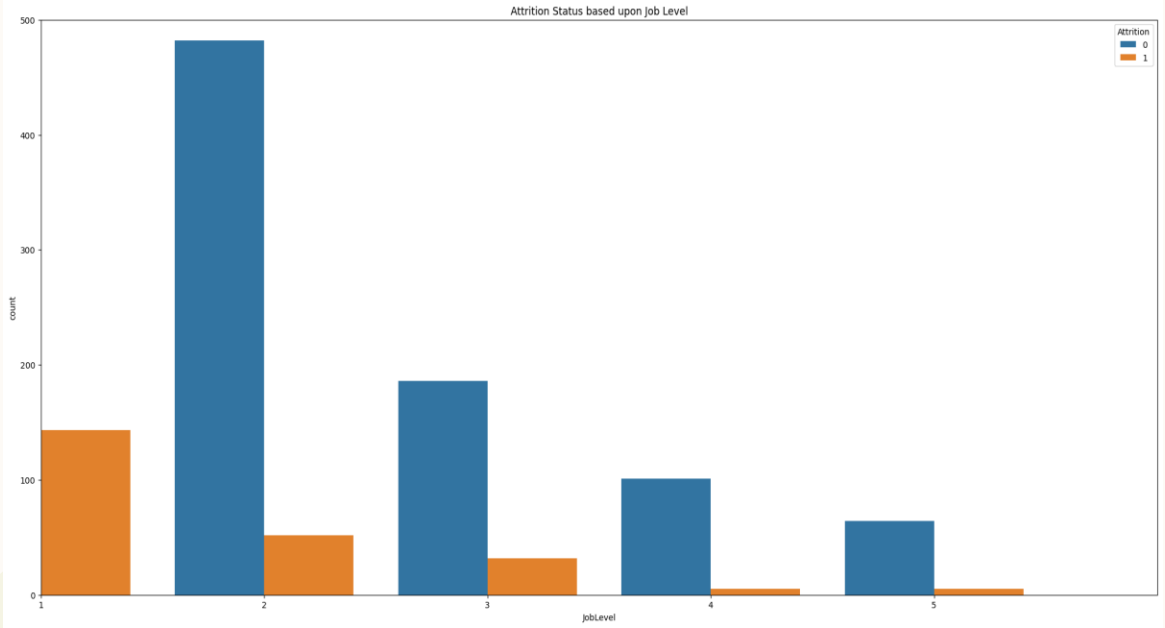
- Job Level and Total Working Hours: Strong positive correlation observed between job level and total working hours.
- Monthly Income and Job Level: Strong positive correlation found between monthly income and job level.
- Monthly Income and Total Working Hours: Strong positive correlation detected between monthly income and total working hours.
- Age and Monthly Income: Strong positive correlation observed between age and monthly income.


Attrition Status based upon Age
Attrition
0
1
count
Age
0
10
20
30
40
50
60
70
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60


Attrition Status based upon Job role
Attrition
0
1
count
JobRole
Sales Executive
Research Scientist
Laboratory Technician
Manufacturing Director
Healthcare Representative
Manager
Sales Representative
Research Director
Human Resources

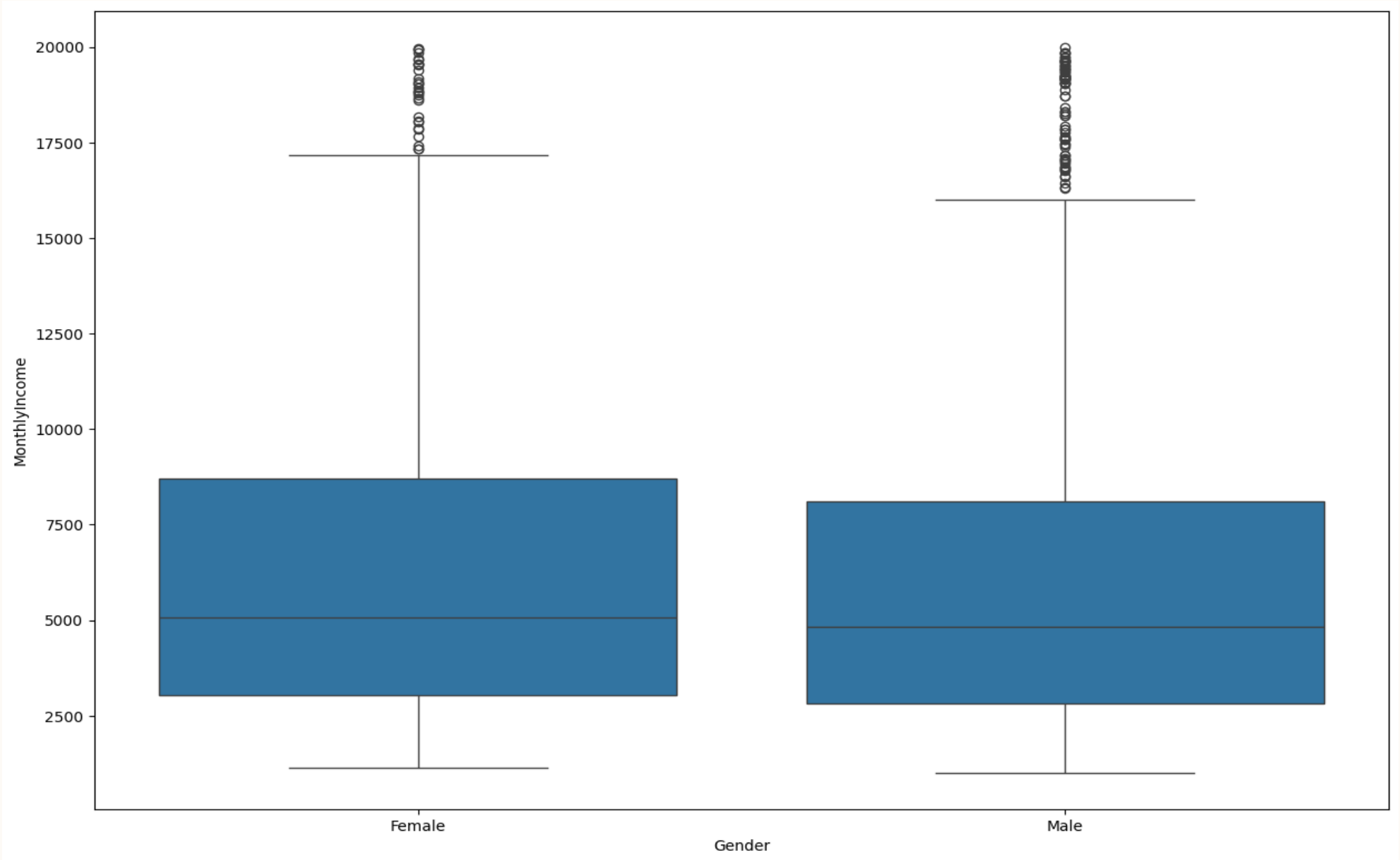

Attrition Status based upon Marital Status
Attrition
0
1
count
MaritalStatus
Single
Married
Divorced

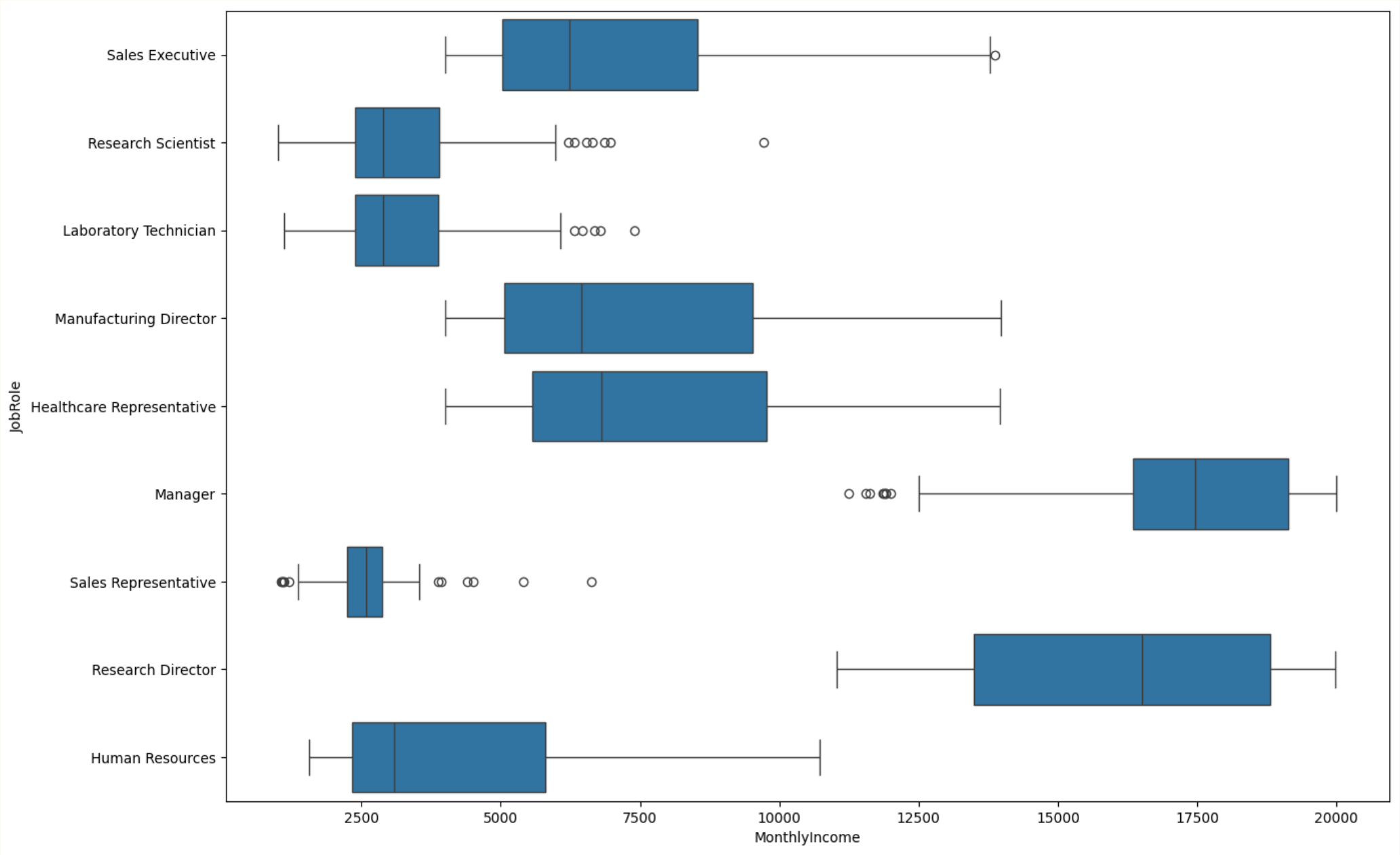

Attrition Status based upon Job Involvement
Attrition
0
1
count
JobInvolvement


Attrition Status based upon Job Level
Attrition
0
1
count
JobLevel

Key Factors Contributing to Employee Attrition:

- Marital Status: Single employees are more likely to leave compared to married and divorced individuals.
- Job Role: Sales Representatives exhibit higher attrition rates compared to employees in other roles.
- Job Involvement: Employees with lower levels of job involvement are more prone to leaving the company.
- Experience Level: Employees with lower job levels (less experienced) tend to leave the company.


20000
17500
15000
12500
10000
7500
5000
2500
MonthlyIncome
Female
Male
Gender


Sales Executive
Research Scientist
Laboratory Technician
Manufacturing Director
Healthcare Representative
Manager
Sales Representative
Research Director
Human Resources
JobRole
2500
5000
7500
10000
12500
15000
17500
20000
MonthlyIncome

MODEL SELECTION

- Total Employees in our dataset is 1470. Number of employees who left: 237(16.12%), Number of employees who stayed: 1233(83.88%)
- From here we can see that, our dataset is imbalanced. For balancing our dataset we have used SMOTE(Synthetic Minority Oversampling Technique)

- We selected some classification models such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Xg Boost, Extra Tree Classifier and Support Vector Machine.
- After balancing the class we have got best accuracy in Gradient Boosting (90%).
- After hyperparameter tuning we have got the best accuracy of 90% in Random Forest and XG Boost Classifier .

BEFORE BALANCING THE CLASS

	Class-0			Class-1			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
Logistic Regression	0.92	0.98	0.95	0.82	0.46	0.59	91
Decision Tree	0.91	0.88	0.89	0.36	0.42	0.39	82
Random Forest	0.89	0.99	0.94	0.82	0.18	0.30	88
Gradient Boosting	0.91	0.97	0.94	0.72	0.42	0.53	90
XGBoost	0.92	0.95	0.93	0.59	0.44	0.51	88
Extra Trees Classifier	0.89	0.99	0.94	0.73	0.22	0.34	88
SVM	0.90	0.99	0.94	0.87	0.26	0.40	89

Table 1: Before Balancing the classes

AFTER BALANCING THE CLASS

	Class-0			Class-1			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
Logistic Regression	0.95	0.76	0.84	0.33	0.76	0.46	76
Decision Tree	0.89	0.81	0.85	0.24	0.38	0.29	75
Random Forest	0.91	0.97	0.94	0.69	0.36	0.47	89
Gradient Boosting	0.92	0.96	0.94	0.66	0.50	0.57	90
XGBoost	0.92	0.94	0.93	0.57	0.48	0.52	88
Extra Trees Classifier	0.91	0.96	0.94	0.64	0.42	0.51	89
SVM	0.92	0.89	0.91	0.43	0.54	0.48	84

Table 2: After Balancing the classes

AFTER HYPERPARAMETER TUNING

	Class-0			Class-1			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
Logistic Regression	0.95	0.76	0.84	0.33	0.76	0.46	76
Decision Tree	0.89	0.93	0.91	0.41	0.30	0.34	85
Random Forest	0.91	0.98	0.94	0.75	0.42	0.54	90
Gradient Boosting	0.94	0.96	0.95	0.70	0.60	0.65	88
XGBoost	0.92	0.96	0.94	0.64	0.50	0.56	90
Extra Trees Classifier	0.91	0.96	0.93	0.61	0.38	0.47	88
SVM	0.87	0.99	0.92	0.43	0.06	0.11	86

Table 3: After Hyperparameter Tuning

CONCLUSION

Summing up, our 'Employee Attrition' project sheds light on the common reasons employees tend to leave their jobs across various industries. Through data analysis and machine learning techniques, we've developed a predictive model to anticipate future attrition. Armed with this tool, companies can take proactive measures to retain talent and enhance workplace contentment.

THANK YOU