NAME: GOKUL NAIR

EMPLOYMENT ATTRITION PREDICTION AND ANALYSIS

AIM:

This project aims to provide insights into the factors influencing employee attrition and predict which employees are likely to leave the company.

PROBLEM STATEMENT:

Acme Corporation, a leading tech company, is facing a significant challenge with employee turnover. The HR department is concerned about the increasing rate of attrition, as it negatively impacts team dynamics, project continuity, and overall company morale. To address this issue, Acme Corporation wants to leverage data analytics and machine learning to understand the factors influencing employee turnover and predict which employees are likely to leave in the near future.

DATASET:

The dataset provided is related to employee attrition in a company. It contains various features describing the employees and their job roles, demographic information, and performance metrics. This dataset spans the last five years and includes information on employees who have left the company and those who are still currently employed.

Overview

Rows: 1470 (each row represents an individual employee)
Columns: 35 (each column represents a feature or variable related to the employee or their job)

Columns and Their Descriptions

Age: Age of the employees
Attrition: Whether the employee has left the company (Yes/No)
BusinessTravel: Frequency of travel for business purposes (Non-Travel, Travel_Rarely, Travel_Frequently)
DailyRate: Daily salary rate of the employee
Department: Department in which the employee works (Sales, Research & Development, Human Resources)
DistanceFromHome: Distance of the employee's residence from the workplace
Education: Education level of the employee (1-5)
EducationField: Field of education (Life Sciences, Other, Medical, Marketing, Technical Degree, Human Resources)
EmployeeCount: Number of employees (always 1, hence irrelevant)
EmployeeNumber: Unique identifier for each employee
EnvironmentSatisfaction: Satisfaction with the work environment (1-4)
Gender: Gender of the employee (Male, Female)
HourlyRate: Hourly wage rate of the employee
JobInvolvement: Level of job involvement (1-4)

JobLevel: Job level within the company (1-5)
JobRole: Job role in the company (e.g., Sales Executive, Research Scientist)
JobSatisfaction: Satisfaction with the job (1-4)
MaritalStatus: Marital status of the employee (Single, Married, Divorced)
MonthlyIncome: Monthly salary of the employee
MonthlyRate: Monthly rate of the employee
NumCompaniesWorked: Number of companies the employee has worked for
Over18: Whether the employee is over 18 (always 'Yes', hence irrelevant)
OverTime: Whether the employee works overtime (Yes/No)
PercentSalaryHike: Percent increase in salary
PerformanceRating: Performance rating (1-4)
RelationshipSatisfaction: Satisfaction with relationships at work (1-4)
StandardHours: Standard hours worked (always 80, hence irrelevant)
StockOptionLevel: Stock option level (0-3)
TotalWorkingYears: Total years of work experience
TrainingTimesLastYear: Number of training sessions attended last year
WorkLifeBalance: Work-life balance satisfaction (1-4)
YearsAtCompany: Number of years the employee has been with the company
YearsInCurrentRole: Number of years in the current role
YearsSinceLastPromotion: Number of years since the last promotion
YearsWithCurrManager: Number of years with the current manager.

DATA PREPROCESSING:

DATA CLEANING-

Check for Missing values:
There are no missing values in any columns.

Check for Outliers:
Since there are no missing values, the next step is to check for inconsistent or outlier values and handle them accordingly. However, it seems the data is relatively clean.

Dropping Irrelevant Columns:
Columns like EmployeeCount, EmployeeNumber, Over18, and StandardHours are dropped as they do not provide useful information for predictive modeling.

DATA TRANSFORMATION-

<u>Encoding Categorical Variables:</u>

Categorical Variables-
Attrition,BusinessTravel,Department,EducationField,Gender,JobRole,MaritalStatus,OverTime

Convert categorical variables into numerical form.
Attrition is encoded as 1 for 'Yes' and 0 for 'No'.
Gender is encoded as 1 for 'Male' and 0 for 'Female'.
OverTime is encoded as 1 for 'Yes' and 0 for 'No'.

Other categorical columns are one-hot encoded.

DATA LABELLING-
Ensuring that the target variable (Attrition) is in the correct format for modeling.

DATA SUMMARY:

Descriptive Statistics:

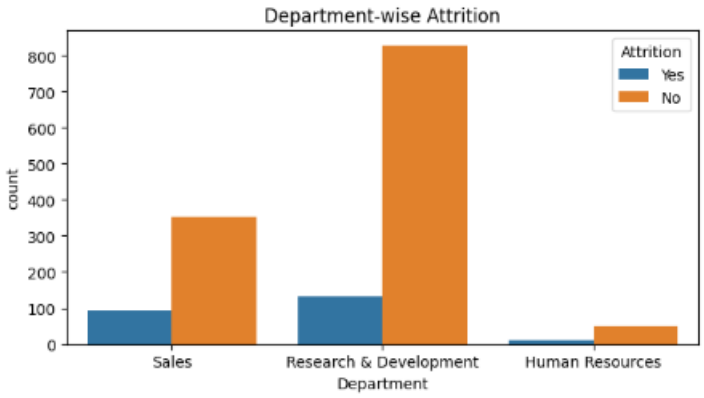
	count	mean	std	min	25%	\
Age	1470.0	36.923810	9.135373	18.0	30.0	
Attrition	1470.0	0.161224	0.367863	0.0	0.0	
DailyRate	1470.0	802.485714	403.509100	102.0	465.0	
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.0	
Education	1470.0	2.912925	1.024165	1.0	2.0	
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.0	
Gender	1470.0	0.600000	0.490065	0.0	0.0	
HourlyRate	1470.0	65.891156	20.329428	30.0	48.0	
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.0	
JobLevel	1470.0	2.063946	1.106940	1.0	1.0	
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.0	
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.0	
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.0	
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.0	
OverTime	1470.0	0.282993	0.450606	0.0	0.0	
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.0	
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.0	
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.0	
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.0	
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.0	
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.0	
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.0	
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.0	
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.0	
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.0	
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.0	

	50%	75%	max
Age	36.0	43.00	60.0
Attrition	0.0	0.00	1.0
DailyRate	802.0	1157.00	1499.0
DistanceFromHome	7.0	14.00	29.0
Education	3.0	4.00	5.0
EnvironmentSatisfaction	3.0	4.00	4.0
Gender	1.0	1.00	1.0
HourlyRate	66.0	83.75	100.0
JobInvolvement	3.0	3.00	4.0
JobLevel	2.0	3.00	5.0
JobSatisfaction	3.0	4.00	4.0
MonthlyIncome	4919.0	8379.00	19999.0
MonthlyRate	14235.5	20461.50	26999.0
NumCompaniesWorked	2.0	4.00	9.0
OverTime	0.0	1.00	1.0
PercentSalaryHike	14.0	18.00	25.0
PerformanceRating	3.0	3.00	4.0
RelationshipSatisfaction	3.0	4.00	4.0
StockOptionLevel	1.0	1.00	3.0
TotalWorkingYears	10.0	15.00	40.0
TrainingTimesLastYear	3.0	3.00	6.0
WorkLifeBalance	3.0	3.00	4.0
YearsAtCompany	5.0	9.00	40.0
YearsInCurrentRole	3.0	7.00	18.0
YearsSinceLastPromotion	1.0	3.00	15.0
YearsWithCurrManager	3.0	7.00	17.0

The dataset includes information on 1470 employees across various demographics and job-related factors. On average, employees are approximately 37 years old. The attrition rate, indicating the proportion of employees who have left the company, stands at around 16%. The average monthly income is approximately $6500, and employees have spent an average of 7 years with the company. Job satisfaction among employees is moderately high, with an average score of 2.73 out of 4. These statistics provide a comprehensive overview of the workforce composition, tenure, and satisfaction levels within the company.

DATA ANALYSIS:

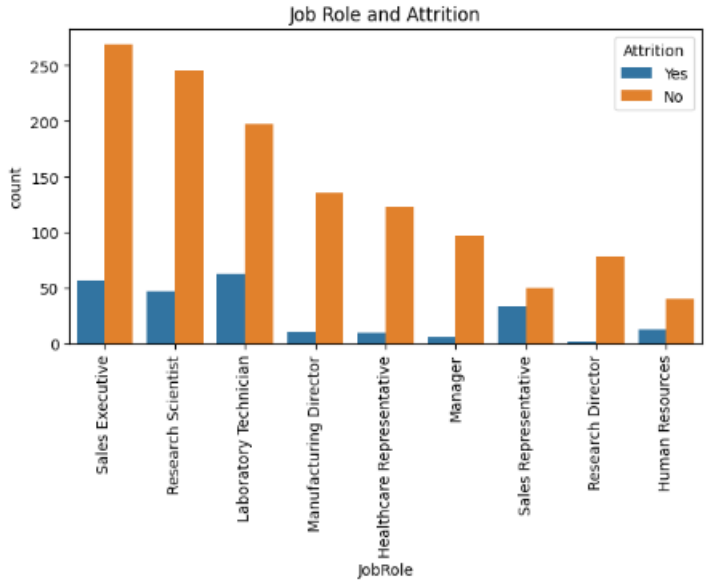
Department-wise Attrition:



Observation: The Research & Development department has the highest number of employees, Sales comes next, followed by Human Resources.

Insight: Attrition proportion seems relatively higher in the Sales and Human Resources departments compared to the overall department size, indicating that these departments might have higher turnover rates.

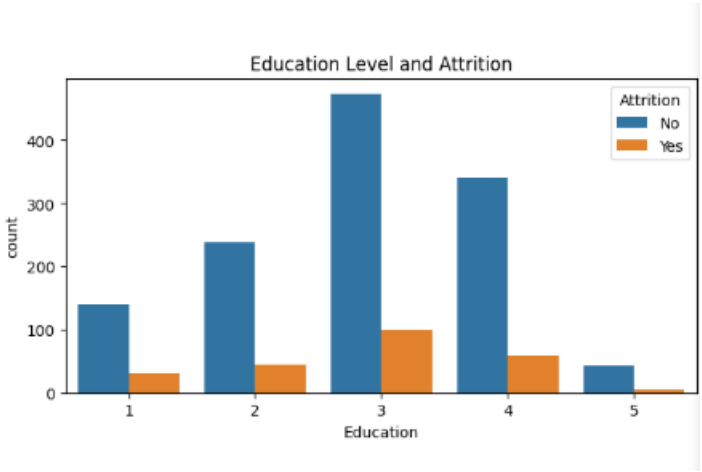
Job Role and Attrition:



Observation: Sales Executives,Research Scientists and Labrotary Technician form the largest groups in terms of total numbers, with a significant portion experiencing attrition. Job roles like Healthcare Representative and Manager have fewer employees, but a noticeable number still leave.

Insight: Certain job roles, particularly Sales Represtative,Sales Executive,Research Scientist and Labrotary Technician show higher rate of attrition. This could suggest job dissatisfaction, high stress, or better opportunities elsewhere in these roles.

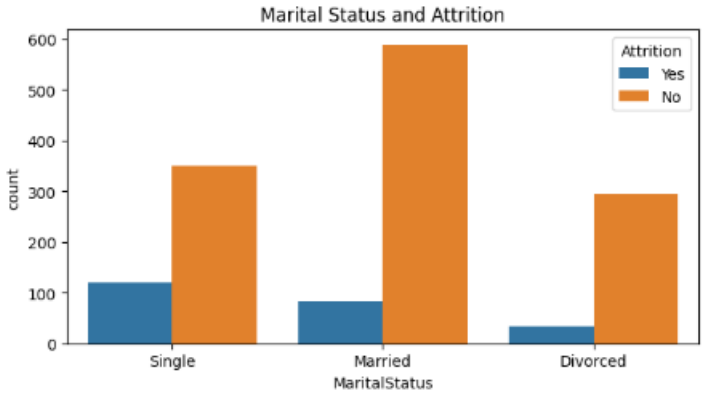
Education Level and Attrition:



Observation: Employees with an education level of 3 (most likely Bachelor's degree) form the largest group. Attrition is noticeable across all education levels, but more frequent at levels 1 and 3.

Insight: There doesn't seem to be a direct correlation between education level and attrition, suggesting other factors may play a more significant role in attrition rates.

Marital Status and Attrition:



Observation: Married employees constitute the largest group, followed by Single and then Divorced employees. Single employees have the highest attrition rate

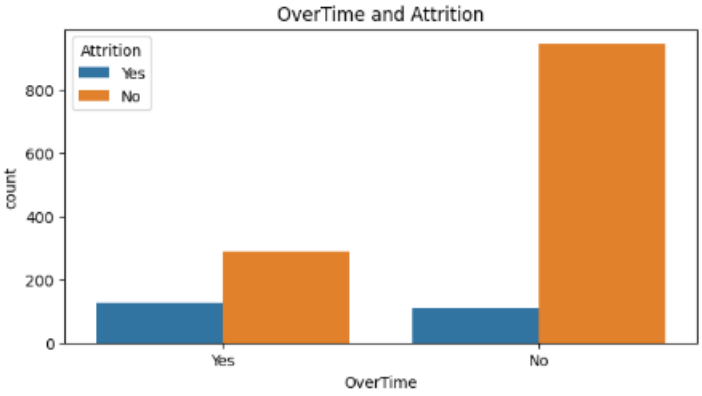
Insight: Single employees seem more likely to leave compared to their married counterparts, possibly due to fewer personal commitments or a greater willingness to relocate for better opportunities.

Work-Life Balance and Attrition:

Observation: Most employees rate their work-life balance as 3, followed by 2. Attrition rate is higher for employees with a work-life balance rating of 1 and 2.

Insight: Poor work-life balance appears to correlate with higher attrition, indicating that improving work-life balance could potentially reduce turnover.

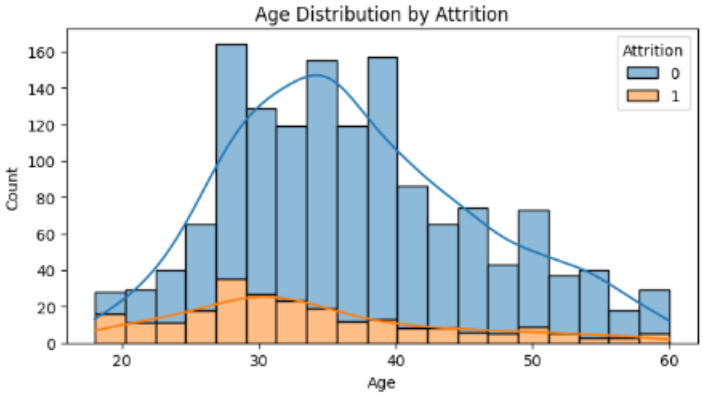
OverTime and Attrition:



Observation: Employees who work overtime have a higher attrition rate compared to those who do not.

Insight: Overtime work is a significant factor in employee attrition. This suggests that excessive overtime might lead to burnout or dissatisfaction, prompting employees to leave.

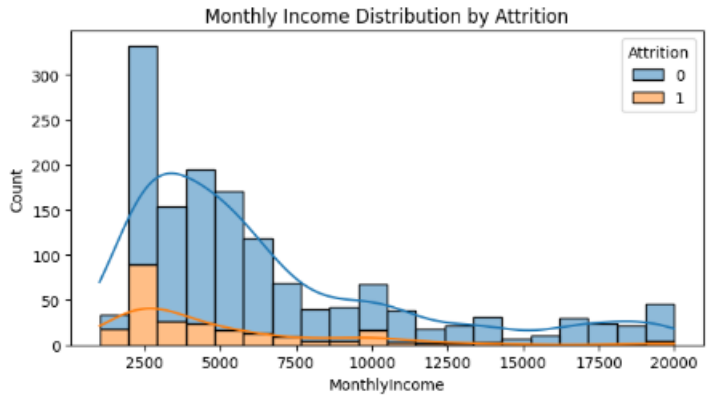
Age Distribution by Attrition:



Observation: The majority of employees fall within the age range of 20 to 40. Within this age range, there is a noticeable number of employees who have left (attrition = 1). As age increases beyond 40, both the number of employees and attrition rates decrease.

Insight: Younger employees, especially those in their late 20s to early 30s, have higher attrition rates. This could be due to career mobility, opportunities for advancement elsewhere, or dissatisfaction with current roles.

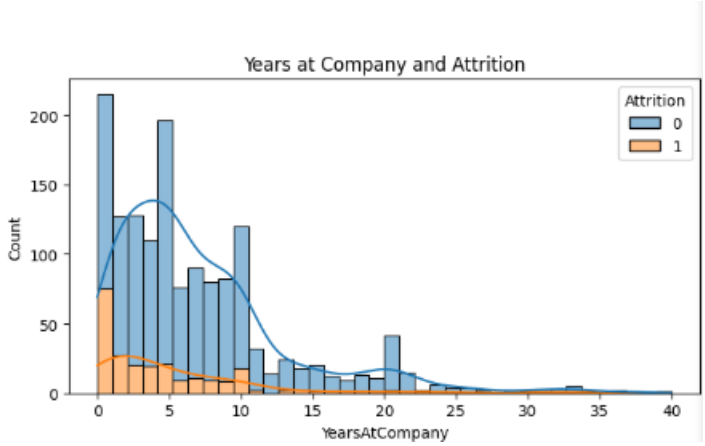
Monthly Income Distribution by Attrition:



Observation: The majority of employees have a monthly income in the range of 2,500 to 7,500. Employees with lower income levels show higher attrition rates compared to those with higher incomes.

Insight: Lower-income employees are more likely to leave the organization, possibly due to financial dissatisfaction or better salary offers from other companies. Higher-income employees tend to stay longer, indicating satisfaction or fewer comparable opportunities.

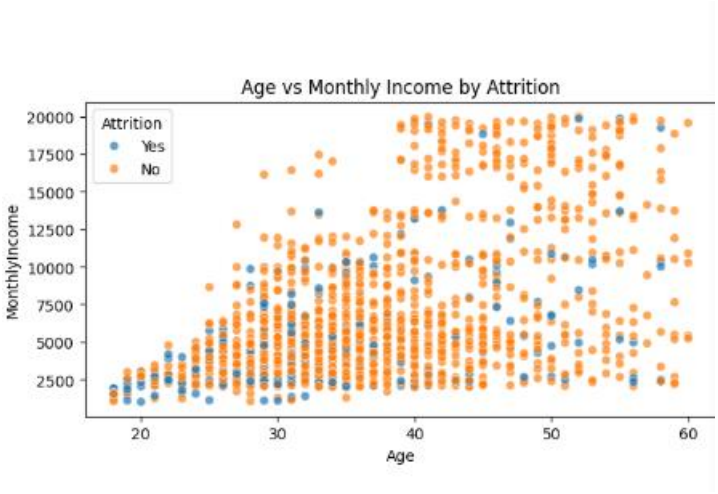
Years at Company and Attrition:



Observation: Most employees have been with the company for 0 to 10 years. Attrition is highest among employees with less than 5 years at the company.

Insight: Newer employees are more prone to leaving. This suggests that improving early employee experiences and retention strategies within the first few years of employment could reduce attrition rates.

Age vs. Monthly Income by Attrition:



Observation: The scatter plot shows a wide distribution of monthly incomes across different ages. However, employees who have left the company (Attrition = Yes) are spread across the income spectrum, with noticeable clusters in the lower-income ranges.

Insight: While there are employees who leave at various income levels and ages, the trend indicates that younger and lower-income employees are more likely to leave. This further reinforces the need to address financial and career development opportunities for younger employees.

CONCLUSION:

The majority of employees tend to remain with the company, indicating overall good retention rates.
Key factors influencing employee attrition-

Department and Job Role Impact: Sales and Human Resources see high turnover rates, suggesting issues like job dissatisfaction or lack of growth. Specific roles like Sales Executive and Research Scientist also show higher attrition, indicating inherent challenges or attractive offers from competitors.

Demographic Factors: Younger employees, especially in their late 20s to early 30s, are more likely to leave. Single employees have higher attrition rates than married ones, possibly due to fewer personal commitments. Gender doesn't significantly affect attrition.

Work-Life Balance and Job Satisfaction: Poor work-life balance and lower job satisfaction are linked to higher attrition rates, emphasizing the importance of maintaining both for retention.

Financial Factors: Lower-income employees and those working overtime are more prone to leaving, suggesting financial dissatisfaction or burnout as key drivers.

Tenure: Newer employees, particularly those with less than 5 years at the company, have higher attrition rates, highlighting the importance of effective onboarding and early career development.

Addressing these factors through targeted HR strategies can reduce attrition rates, improve satisfaction, and foster a stable workforce.

MODEL FITTING:

In this model we have used logistic regression.

Logistic regression is a widely used statistical method for binary classification problems, such as predicting employee attrition. In the context of employee attrition, the target variable indicates whether an employee has left (Attrition = 1) or stayed (Attrition = 0). The model can show how changes in variables like salary, age, or job role influence the probability of an employee leaving.

Model Building

Model Initialization:
A Logistic Regression model was initialized with a maximum iteration of 1000 to ensure convergence.

Model Training:
The model was trained using the standardized training set.

Model Evaluation:
Predictions were made on the test set, and the model's performance was evaluated using accuracy, confusion matrix, and classification report.

CODE-

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

# Split data into features and target variable
X = data_encoded.drop('Attrition', axis=1)
y = data_encoded['Attrition']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Standardize the features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Logistic Regression
log_reg = LogisticRegression(max_iter=1000)
log_reg.fit(X_train, y_train)
y_pred_log_reg = log_reg.predict(X_test)
```

```
print("Logistic Regression:")
print("Confusion Matrix:\n",confusion_matrix(y_test, y_pred_log_reg))
print("Classification Report:\n",classification_report(y_test, y_pred_log_reg))
print("Accuracy:", accuracy_score(y_test, y_pred_log_reg))
```

OUTPUT:

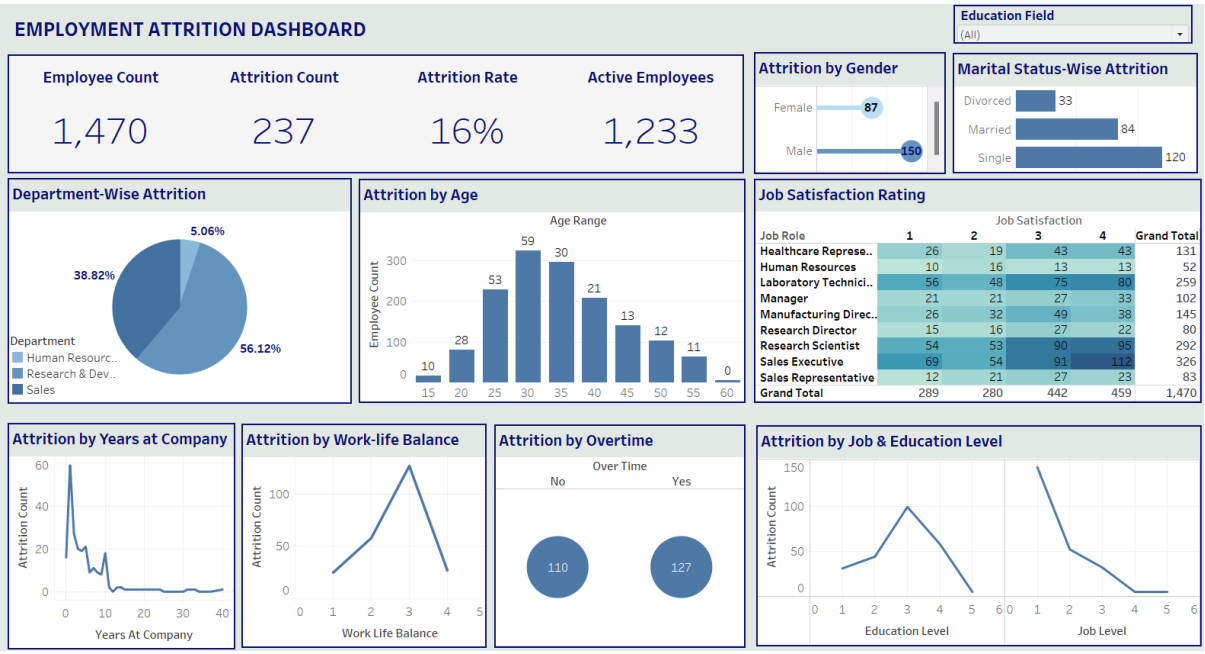
```
Logistic Regression:
Confusion Matrix:
 [[359  21]
 [ 37  24]]
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.94      0.93       380
           1       0.53      0.39      0.45        61

    accuracy                           0.87       441
   macro avg       0.72      0.67      0.69       441
weighted avg       0.85      0.87      0.86       441

Accuracy: 0.8684807256235828
```

CONCLUSION:

The model correctly identified 359 non-attrition cases and 24 attrition cases.
Accuracy: The model achieved an accuracy of 87%.
The model showed high precision and recall for non-attrition cases but moderate performance for predicting attrition cases. This discrepancy is common in imbalanced datasets.

The analysis provided valuable insights into factors influencing employee attrition. The logistic regression model achieved a good accuracy of 87%, indicating its effectiveness in predicting attrition. However, the model's performance for attrition cases suggests room for improvement, possibly through more sophisticated modeling techniques or addressing class imbalance.

DASHBOARD USING TABLEAU:



In dataset we have dropped columns like EmployeeCount, Over18, and StandardHours are dropped as they do not provide useful information.

In next step we converted attrition column values(Yes/No) into binary values(1/0) by creating calculated value and then calculated attrition count for plotting attrition against other variables.

Overall Attrition Metrics

Employee Count: 1,470 employees.

Attrition Count: 237 employees have left the company.

Attrition Rate: 16% of the workforce has experienced attrition (**Employee Count/ Attrition Count)**

Active Employees: 1,233 employees are currently active.

The attrition rate of 16% indicates a moderate level of employee turnover. To further contextualize, industries generally consider a turnover rate of around 10-15% as normal. Therefore, the company's rate is slightly above average, which might warrant further investigation.

All the below metrices gives us the Attrition Count we can navigate and click on Attrition Rate to get the propotion(rate) of attrition of respective metrices.

Attrition by Gender

Female: 87 employees
Male: 150 employees

Observation:
Men have a higher attrition count compared to women.

Insight:
The proportion of attrition is greater in males as compared to females

Marital Status-Wise Attrition

Observation:
Single: 120 employees
Married: 84 employees
Divorced: 33 employees

Insight:
Single employees experience higher attrition rates, which may reflect their flexibility to relocate or pursue career changes. Married and divorced employees might be more stable but could also face different challenges affecting their job retention.

Department-Wise Attrition

Observation: The Research & Development department has the highest number of employees, Sales comes next, followed by Human Resources.

Insight: Attrition proportion seems relatively higher in the Sales and Human Resources departments compared to the overall department size, indicating that these departments might have higher turnover rates.

Attrition by Age

Observation: The majority of employees fall within the age range of 20 to 40. Within this age range, there is a noticeable number of employees who have left (attrition = 1). As age increases beyond 40, both the number of employees and attrition rates decrease.

Insight: Younger employees, especially those in their late 20s to early 30s, have higher attrition rates. This could be due to career mobility, opportunities for advancement elsewhere, or dissatisfaction with current roles.

Job Satisfaction Rating

The heat map provides attrition data by job role and satisfaction level (1 being low, 4 being high).

Observation: Sales Executives,Research Scientists and Labrotary Technician form the largest groups in terms of total numbers, with a significant portion experiencing attrition. Job roles like Healthcare Representative and Manager have fewer employees, but a noticeable number still leave.

Insight: Certain job roles, particularly Sales Represtative,Sales Executive,Research Scientist and Labrotary Technician show higher rate of attrition. This could suggest job dissatisfaction, high stress, or better opportunities elsewhere in these roles.

Attrition by Years at Company

Observation:
0-1 year: High attrition.
2-5 years: Decreases but still significant.
5+ years: Drops sharply.

Insight:
Early-stage employees have higher attrition rates, highlighting the importance of onboarding, training, and early engagement strategies to improve retention. Attrition decreases significantly after the 5-year mark, indicating increased stability among long-term employees.

Attrition by Work-Life Balance:

Observation: Most employees rate their work-life balance as 3, followed by 2. Attrition rate is higher for employees with a work-life balance rating of 1 and 2.

Insight: Poor work-life balance appears to correlate with higher attrition, indicating that improving work-life balance could potentially reduce turnover.

Attrition by Overtime:

Observation: Employees who work overtime have a higher attrition rate compared to those who do not.

Insight: Overtime work is a significant factor in employee attrition. This suggests that excessive overtime might lead to burnout or dissatisfaction, prompting employees to leave.

Attrition by Education Level:

Observation: Employees with an education level of 3 (most likely Bachelor's degree) form the largest group. Attrition is noticeable across all education levels, but more frequent at levels 1 and 3.

Insight: There doesn't seem to be a direct correlation between education level and attrition, suggesting other factors may play a more significant role in attrition rates.

Attrition by Job Level:

Insight: Attrition is higher among employees with lower education levels, which may point to roles that are less fulfilling or offer fewer career advancement opportunities. Higher attrition at specific education levels suggests targeted retention strategies for employees in these categories.

CONCLUSION

The dashboard provides a clear and comprehensive view of the factors contributing to employee attrition. By focusing on key areas such as department-specific challenges, age-specific retention strategies, and work-life balance improvements, the company can take proactive steps to reduce turnover and improve employee satisfaction and retention.