NAME: SAKSHI SUDHAKAR SHIRORE

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-

Encoding Categorical Variables:

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:

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 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 has information on 1470 employees, including their age, job details, and other factors. Here are some key points:

- The average age of employees is about 37 years.
- Around 16% of employees have left the company.
- The average monthly income is about $6500.
- Employees have worked at the company for an average of 7 years.
- Job satisfaction is fairly high, with an average score of 2.73 out of 4.

These details give a clear picture of the employees' age, income, time at the company, and job satisfaction.

DATA ANALYSIS:

Department-wise Attrition:

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

Insights: The proportion of employees leaving the company appears to be higher in the Sales and Human Resources departments compared to their overall size. This suggests that these departments may have higher turnover rates.

Job Role and Attrition:

Job Role and Attrition
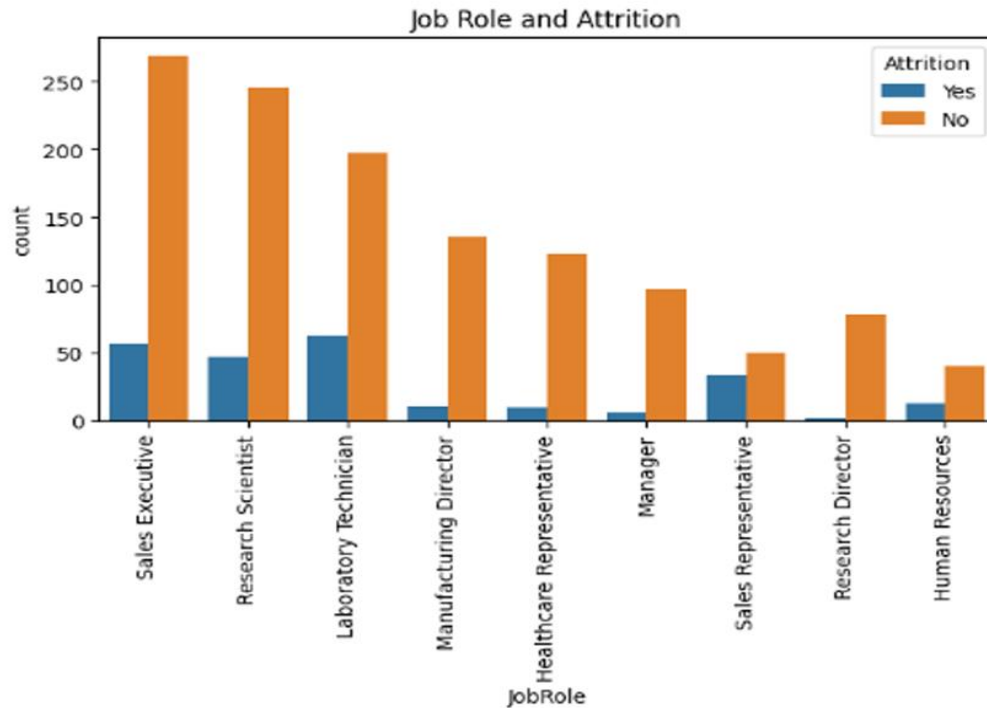
Observation: Sales Executives, Research Scientists, and Laboratory Technicians are the largest groups of employees, with many of them leaving the company. Job roles like Healthcare Representatives and Managers have fewer employees, but a noticeable number still leave.

Insight: Certain job roles, like Sales Representatives, Sales Executives, Research Scientists, and Laboratory Technicians, have higher rates of employees leaving. This might mean they are unhappy, stressed, or finding better jobs elsewhere.

Education Level and Attrition:

Education Level and Attrition
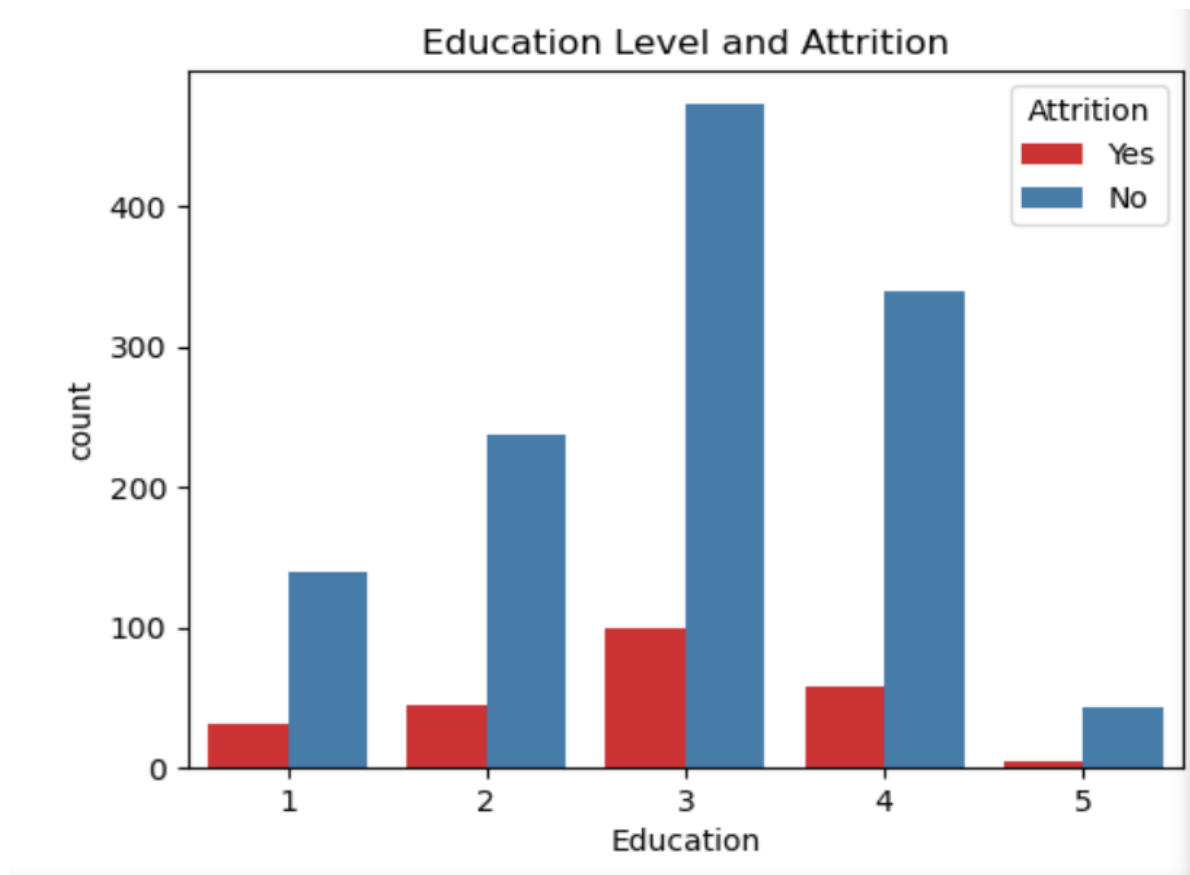
Observation: Employees with an education level of 3 (probably a Bachelor's degree) are the largest group. Attrition happens at all education levels, but it's more common at levels 1 and 3.

Insight: There doesn't seem to be a direct link between education level and attrition, suggesting other factors might be more important in why employees leave.

Marital Status and Attrition:

Marital Status and Attrition
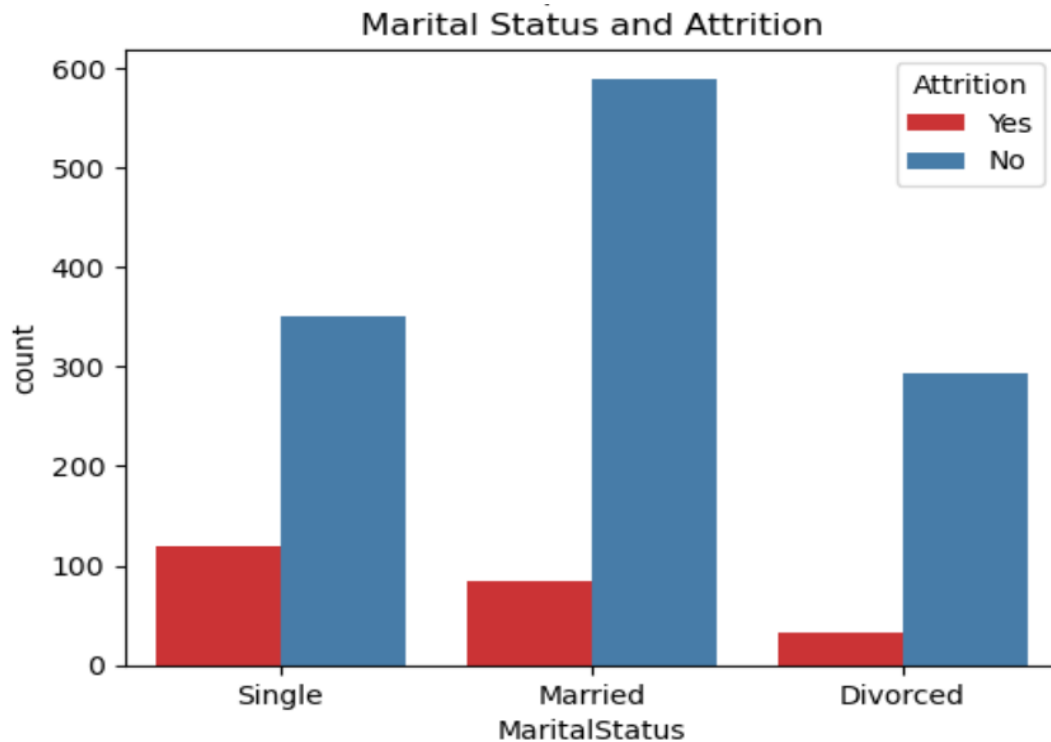
Observation: Married employees are the largest group, followed by single and then divorced employees. Single employees have the highest attrition rate.

Insight: Single employees are more likely to leave, possibly because they have fewer personal commitments or are more willing to move for better opportunities.

Work-Life Balance and Attrition:

Work-Life Balance and Attrition
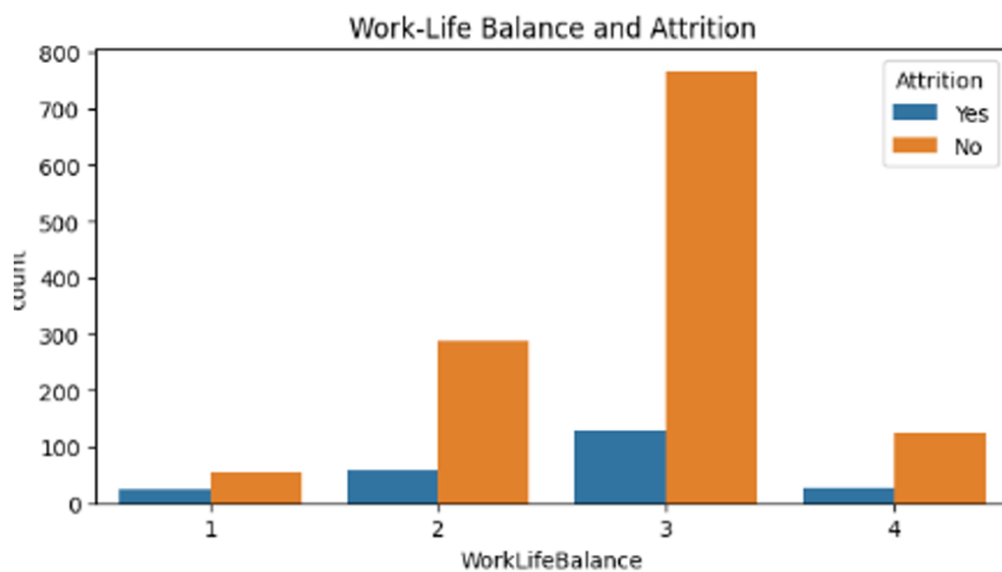
Observation: Most employees rate their work-life balance as 3, followed by 2. Employees with a work-life balance rating of 1 and 2 have higher attrition rates.

Insight: Poor work-life balance seems to be linked to higher attrition, suggesting that improving work-life balance could help reduce turnover.
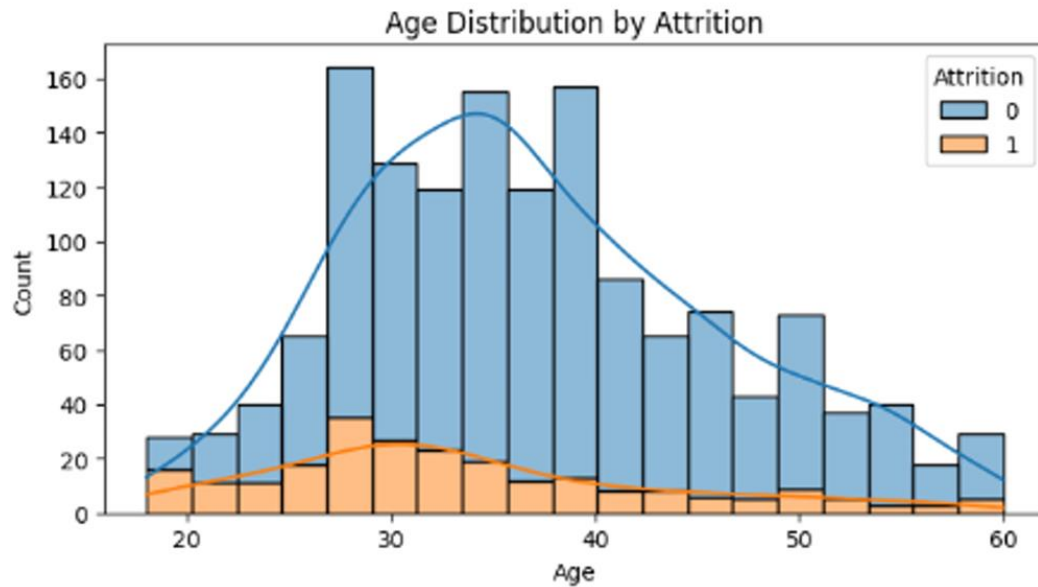
Over Time and Attrition:
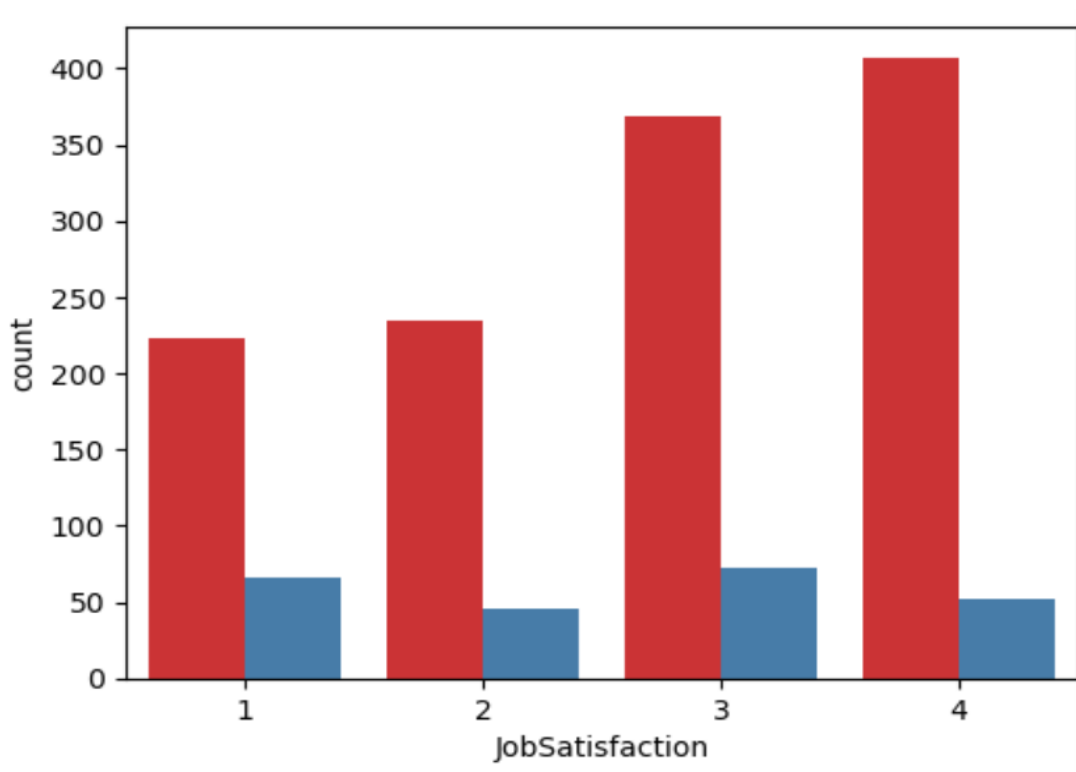
Observation: Employees who work overtime are more likely to leave than those who don't.

Insight: Working overtime is a big reason why employees leave. It seems that too much overtime could make employees feel burned out or unhappy, so they decide to quit.

Age Distribution by Attrition:

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:

Monthly Income Distribution by Attrition
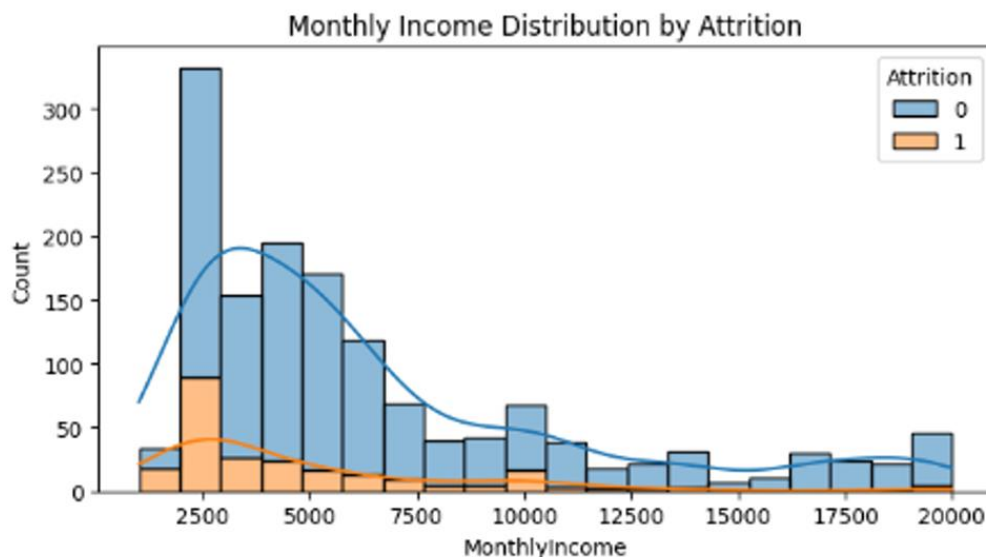
Observation: Most employees earn between $2,500 and $7,500 per month. Those with lower incomes are more likely to leave than those with higher incomes.

Insight: Employees who earn less are more likely to quit, maybe because they're not happy with their pay or because they find better-paying jobs elsewhere. Employees who earn more tend to stay longer, suggesting they're satisfied or have fewer options for higher-paying jobs elsewhere.

Years at Company and Attrition:
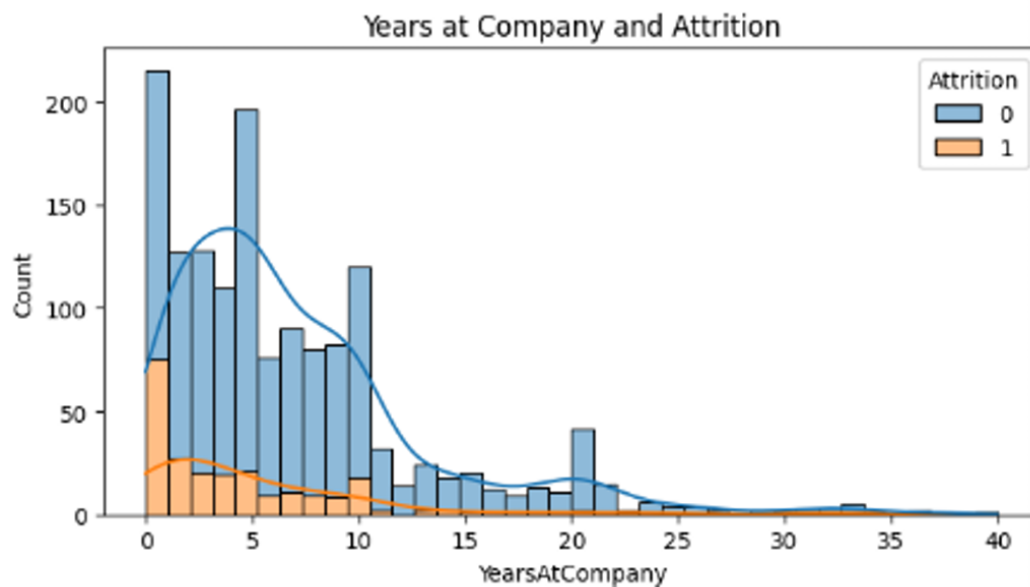
Observation: Most employees have worked at the company for 0 to 10 years. Attrition is highest among those who have been with the company for less than 5 years.

Insight: Newer employees are more likely to leave. This indicates that focusing on improving the experience and retention strategies for new hires in their first few years could help reduce turnover.

Age vs. Monthly Income by Attrition:

Age vs Monthly Income by Attrition
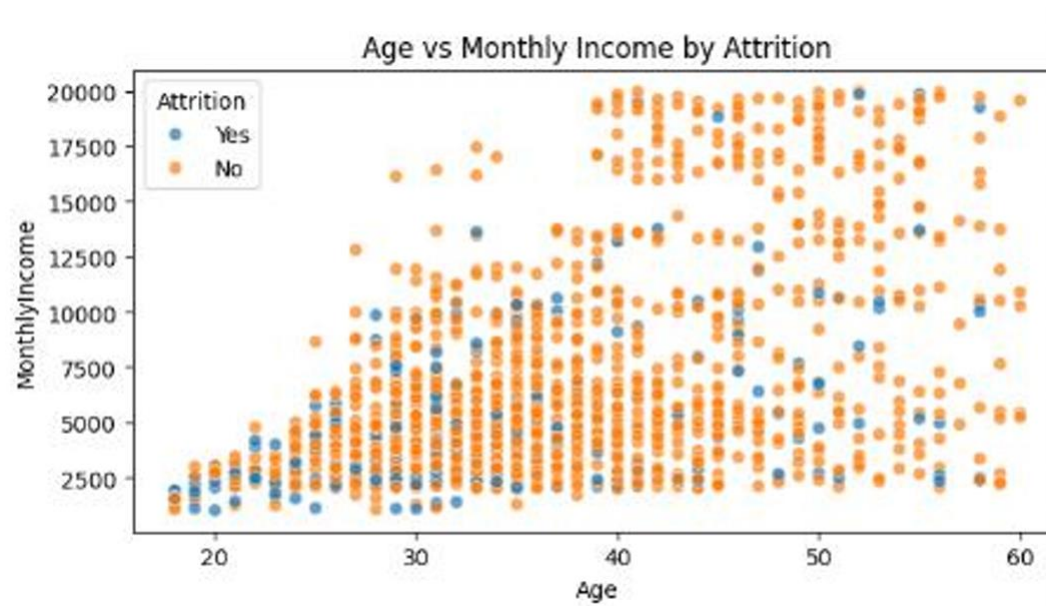
Observation: The scatter plot displays various monthly incomes across different ages. However, employees who have left the company (Attrition = Yes) are mostly found in the lower-income ranges.

Insight: Although employees leave at different income levels and ages, younger and lower-income employees are more prone to leaving. This highlights the importance of addressing financial and career development needs for younger employees.

CONCLUSION: Most employees stay with the company, indicating good retention overall. Here are the main factors affecting attrition:

1. Department and Job Role: Sales and Human Resources have high turnover, as do roles like Sales Executive and Research Scientist.
2. Demographics: Younger employees, especially in their late 20s to early 30s, are more likely to leave. Single employees leave more than married ones.
3. Work-Life Balance and Job Satisfaction: Poor balance and low satisfaction lead to higher attrition.
4. Financial Factors: Lower-income employees and those working overtime tend to leave more, possibly due to financial concerns or burnout.
5. Tenure: Newer employees, especially those with less than 5 years at the company, are more likely to leave, showing the need for better onboarding and early career development.

To reduce attrition, HR strategies should focus on these areas.

MODEL FITTING:

We used logistic regression, a method for binary classification like predicting employee attrition. The model helps understand how factors like salary, age, or job role affect the likelihood of an employee leaving.

Model Building: We initialized a Logistic Regression model with a maximum iteration of 1000 for convergence.

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

Model Evaluation: Predictions were made on the test set, and we assessed the model's performance 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 cases where employees stayed and 24 cases where employees left. It achieved an accuracy of 87%.

While the model performed well in predicting employees who stayed, its performance in identifying those who left was moderate. This is common in datasets where one class (attrition) is much smaller than the other (non-attrition).

Overall, the analysis gave us useful insights into what influences employee attrition. Although the logistic regression model showed good accuracy, it could be improved, especially in predicting attrition cases. This might require using more advanced modelling techniques or addressing the imbalance between the two classes.

DASHBOARD:
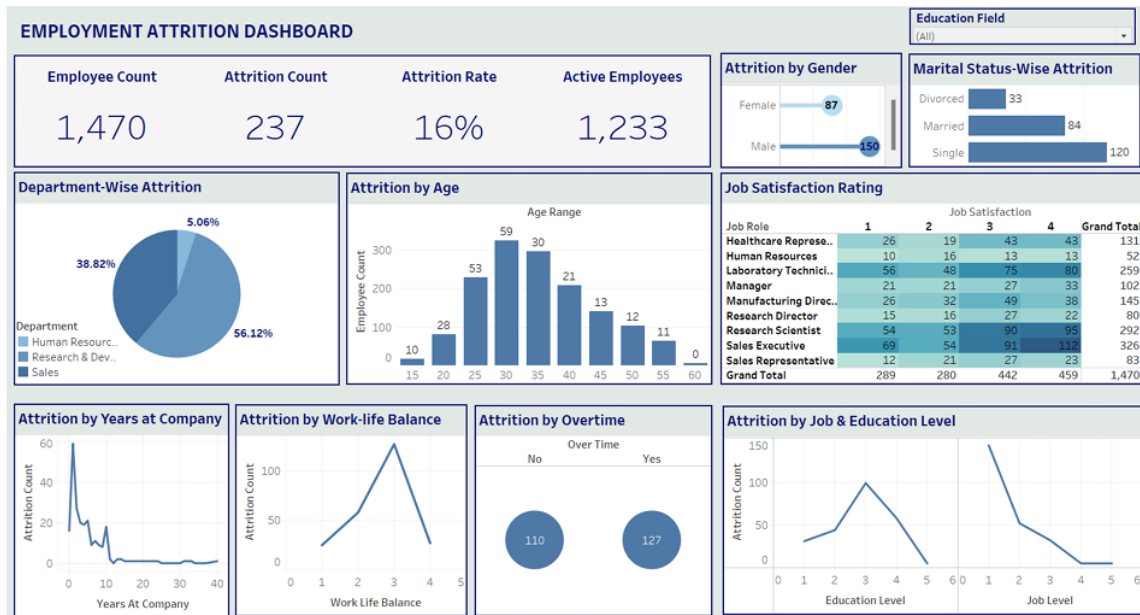
We removed the columns Employee Count, Over18, and Standard Hours from the dataset because they didn't provide useful information. Then, we converted the attrition column from Yes/No values to binary values (1 for Yes, 0 for No). After that, we calculated the number of employees who left the company to use this information 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 proportion (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 Laboratory 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.