IBM HR Analytics Employee Attrition & Performance

HR Attrition Analysis:

In the business world, companies often face the challenge of retaining talented employees. One of the most pressing issues is the increasing rate of employee turnover, commonly known as HR attrition. Turnover can have a significant impact on a company's productivity, stability, and long-term sustainability. High attrition rates can lead to increased recruitment and training costs, disrupt team dynamics, and result in the loss of valuable institutional knowledge. Therefore, understanding the factors contributing to attrition and implementing effective retention strategies is crucial for maintaining a competitive edge and ensuring

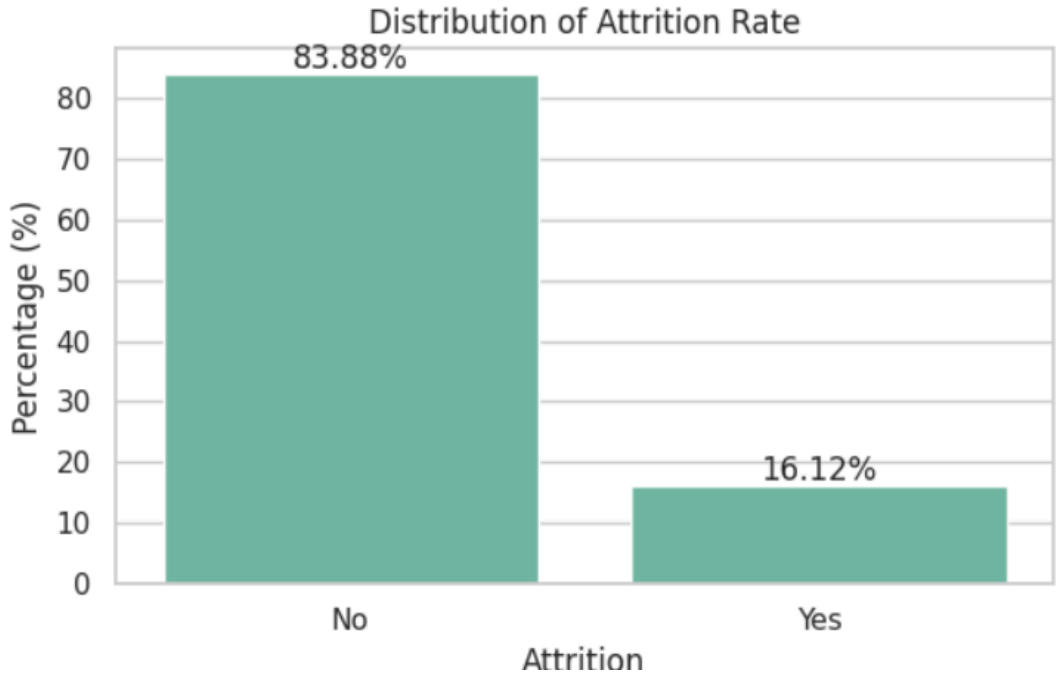
Objectives of the Analysis:

1. Understand Current Turnover Rates: Gain a comprehensive understanding of the current employee turnover rate and analyze the demographic distribution of attrition by age, gender, education, department, and job role.

2. Identify Key Factors Influencing Turnover: Examine the main factors contributing to employee attrition, including job satisfaction indicators (job involvement and work-life balance), salary factors (monthly income and salary hikes), and benefit factors (stock option levels), to uncover patterns and correlations that drive higher attrition rates.

Employee Attrition Rate Distribution:

```
Attrition Distribution:
 Attrition
No     0.838776
Yes    0.161224
Name: proportion, dtype: float64
```



The image displays a **bar chart** titled **"Distribution of Attrition Rate"**, which visualizes how many employees have left the organization ("Yes") versus those who have stayed ("No").

Key Details from the Output:

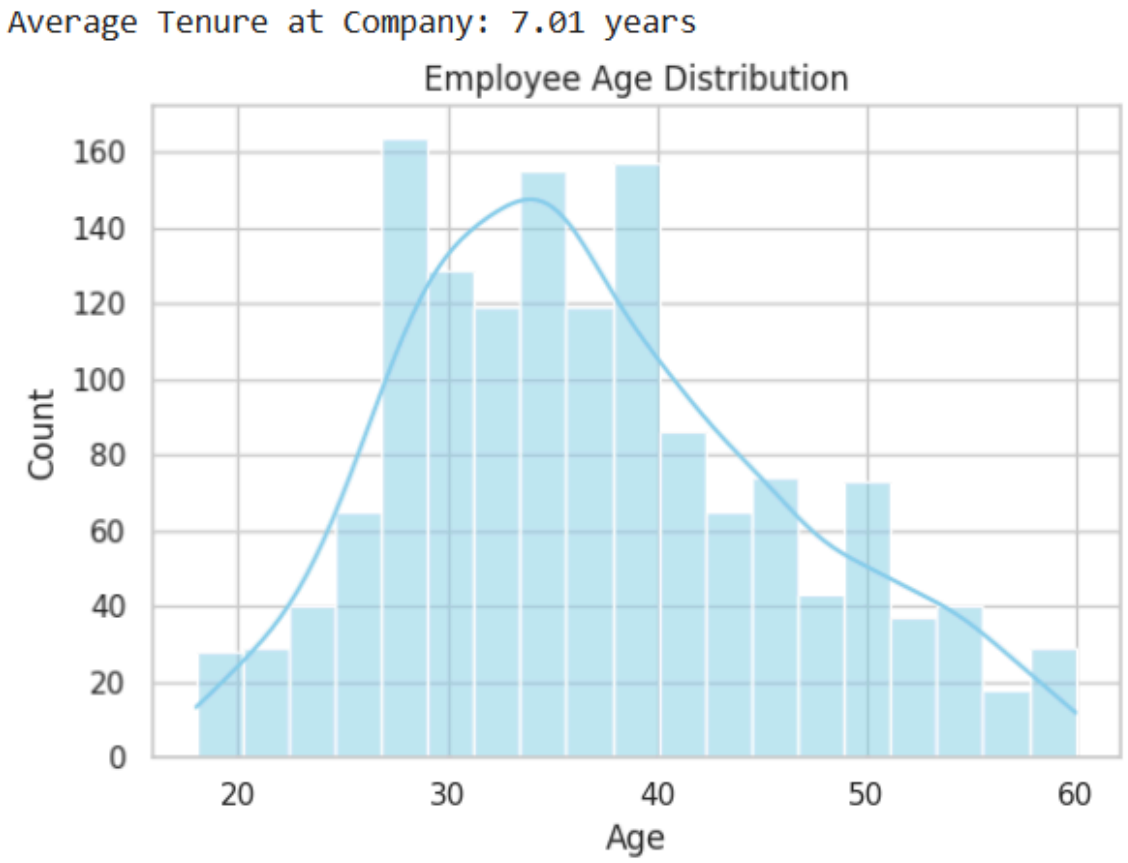
- **Attrition Rate (Yes):** 16.12% of employees left the company.
- **Retention Rate (No):** 83.88% of employees remained in the company.

This insight is derived from a **value count normalization** (`value_counts(normalize=True)`) on the `Attrition` column of the dataset. The chart helps to quickly identify that attrition is not widespread in this organization, but still notable (more than 1 in 10 employees).

Demographic Overview of Employees (Age, Gender, Department)

Average Tenure:

- The average number of years an employee stays at the company is **7.01 years**, indicating relatively **stable employment** within the organization.



Employee Age Distribution (Left Plot):

- A **histogram with KDE (smoothed line)** shows that:
    - Most employees fall between **30 and 40 years of age**.
    - There is a symmetric, bell-like curve, suggesting a **normal distribution** centered around the early 30s.
    - Very few employees are younger than 25 or older than 55.

Gender Distribution (Middle Plot):

A **bar chart** reveals:

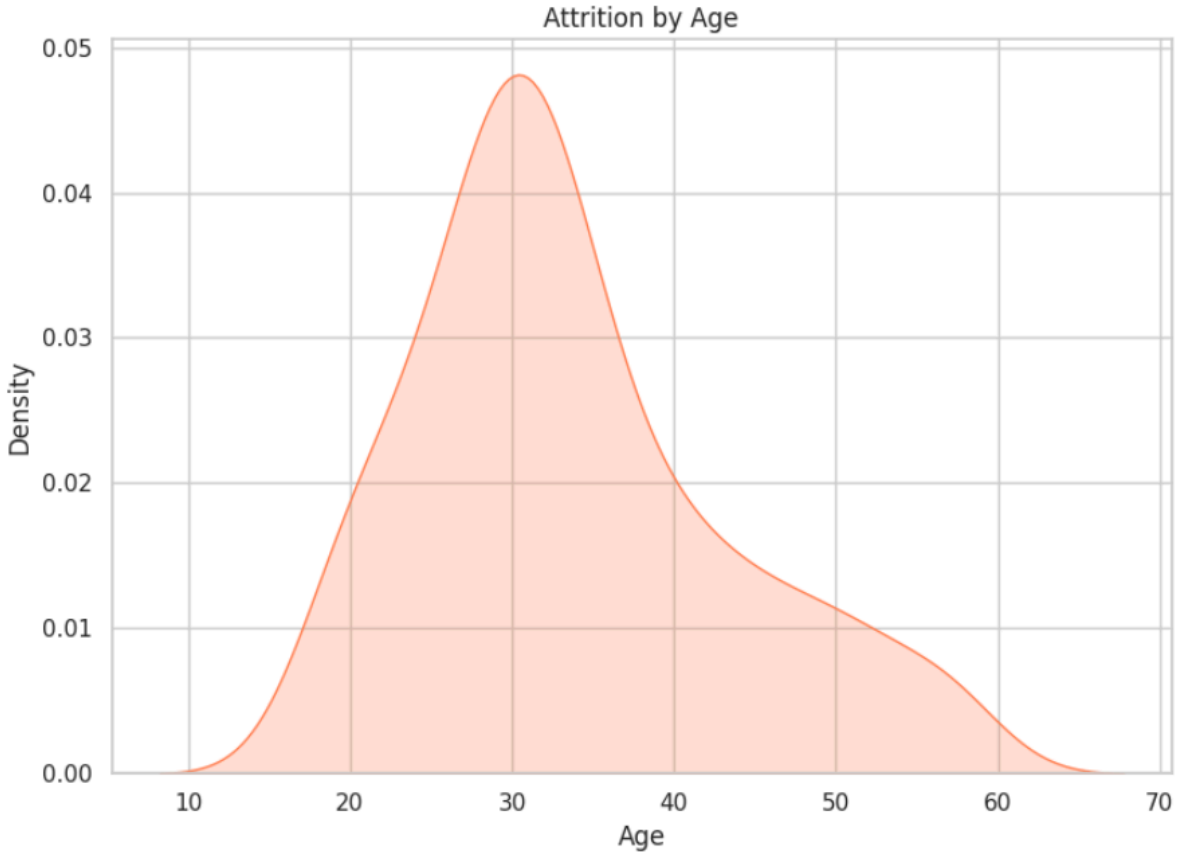
- A higher proportion of **male employees (880)** than **female employees (580)**.
- This indicates a **gender imbalance**, which may warrant further inclusion efforts depending on organizational goals.

Department Distribution (Right Plot):

This chart highlights:

- The majority of employees work in **Research & Development (950)**.
- **Sales** is the next largest department (450 employees).
- **Human Resources** has the fewest employees (70).

Attrition by Age:



This **KDE plot** (Kernel Density Estimate) visualizes the distribution of employees who left the company (i.e., attrition) with respect to age. The curve peaks around the **early 30s**, indicating that **younger employees (around 30 years old)** are more likely to leave the organization. The density then tapers off, meaning **older employees are less likely to leave**. This suggests that **age is a significant factor in attrition**, possibly due to younger professionals seeking better growth, pay, or career transitions.

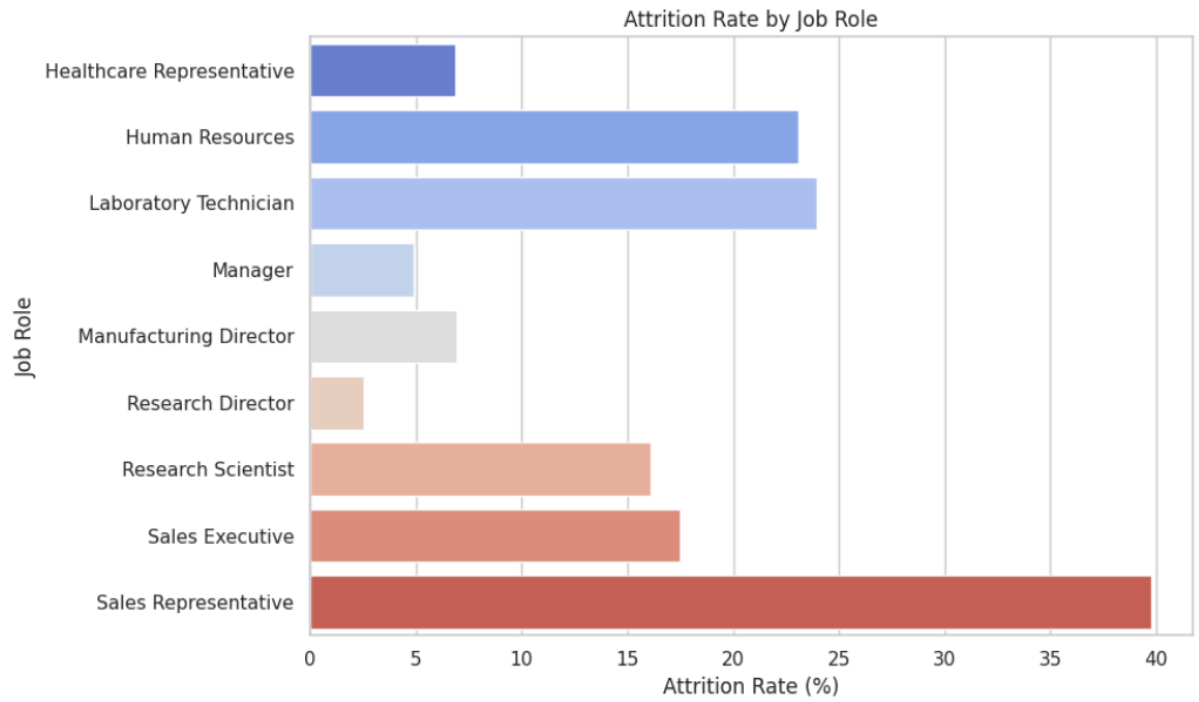
Attrition Rate by Gender:

This **bar chart** shows the attrition rate split by gender. The attrition rate for **male employees is slightly higher** than that of female employees.

- **Male Attrition Rate = 17%**
- **Female Attrition Rate = 15%**

This implies that while the difference isn't drastic, **male employees may be slightly more prone to leaving**, possibly due to external opportunities, job dissatisfaction, or other role-based expectations. It also emphasizes the need for **gender-inclusive retention strategies**.

Job Role and Employee Attrition:

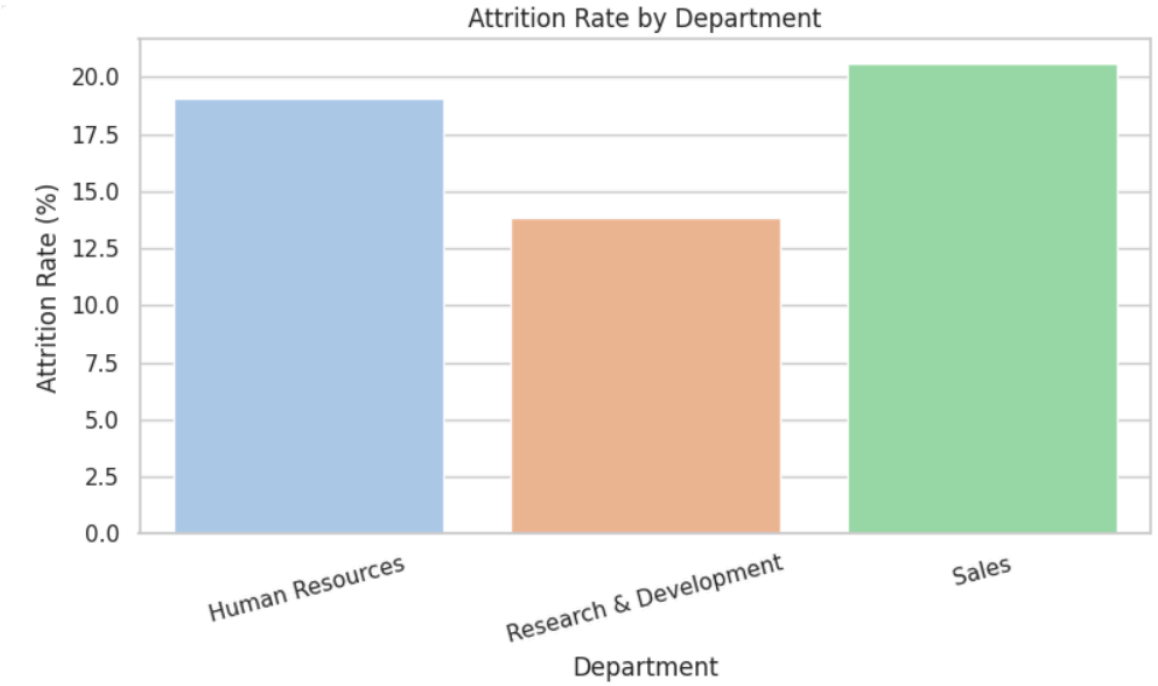


This horizontal bar chart illustrates the attrition rate across various job roles. It reveals that:

- **Sales Representatives** have the highest attrition rate, nearing 40%.
- **Human Resources** and **Laboratory Technicians** also show notably high attrition.
- **Research Directors**, **Managers**, and **Healthcare Representatives** exhibit the lowest attrition rates.

This suggests certain job roles are more prone to turnover, likely due to stress, career growth stagnation, or external opportunities.

Department-wise Employee Attrition:

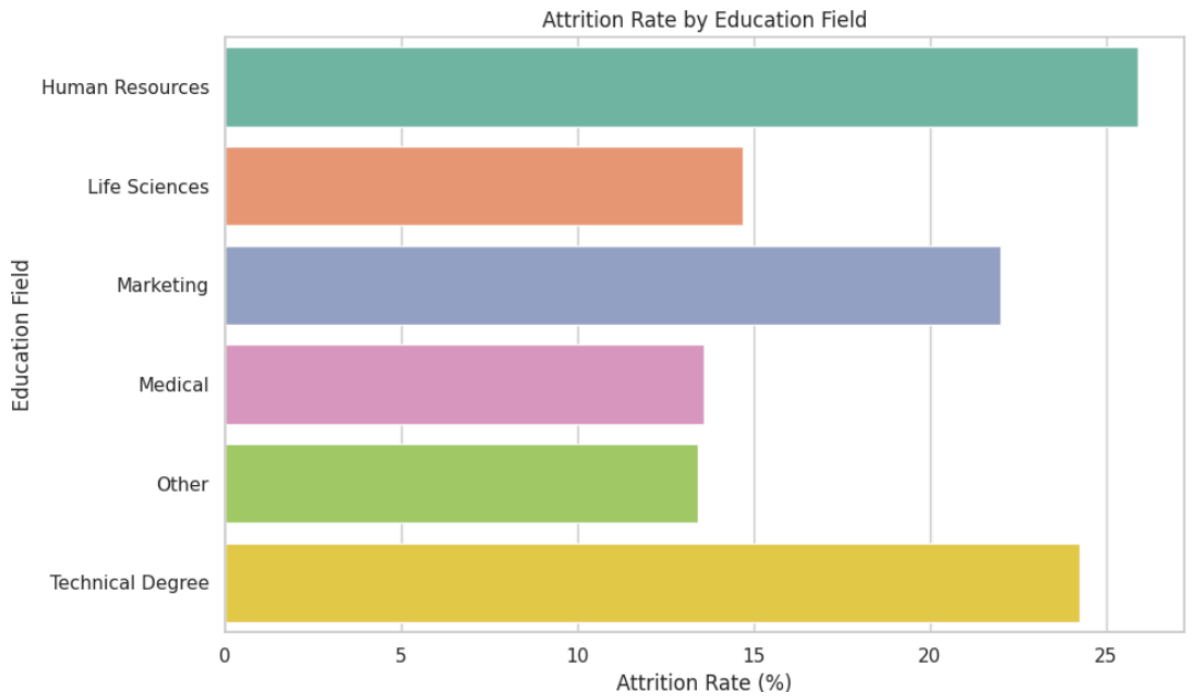


This vertical bar plot compares attrition rates among departments:

- The **Sales** department has the highest attrition (above 20%).
- **Human Resources** follows closely.
- **Research & Development** has the lowest attrition, suggesting more job satisfaction or retention strategies at play.

Understanding departmental attrition helps target interventions more effectively.

Attrition Rate by Education Field:

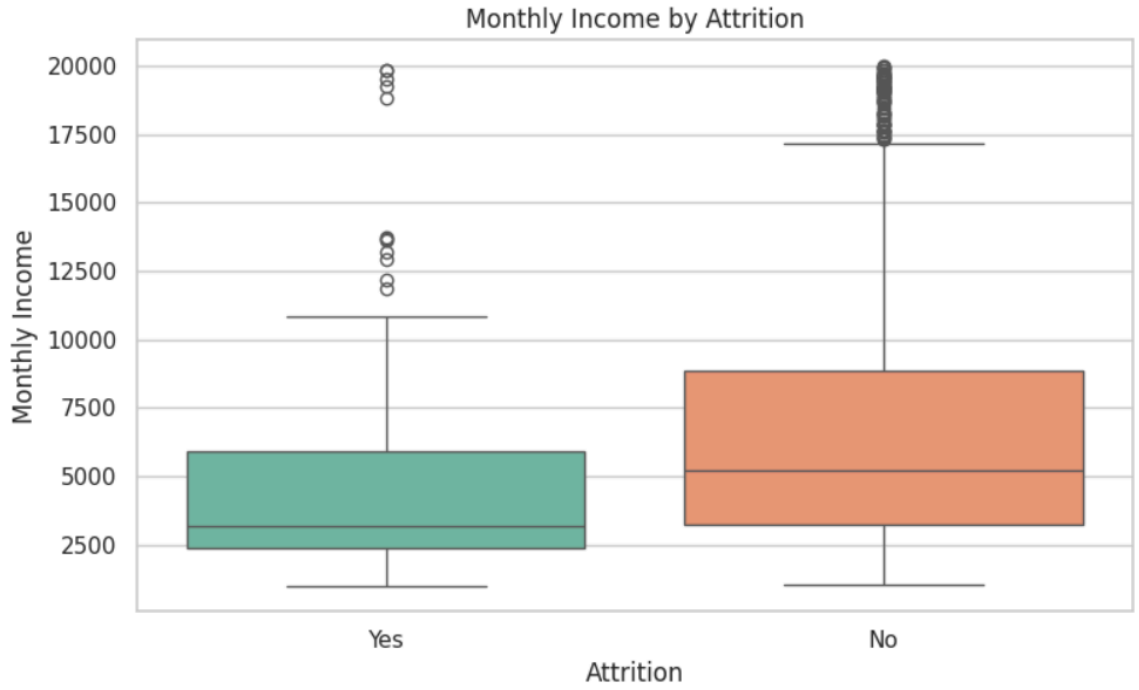


This horizontal bar chart explores how attrition varies across employees' education fields:

- **Human Resources**, **Technical Degrees**, and **Marketing** show higher attrition rates (22–25%).
- Fields like **Medical**, **Other**, and **Life Sciences** see comparatively lower turnover.

This may reflect job-market dynamics or alignment between job expectations and academic preparation.

Employee Attrition vs. Monthly Income:



This boxplot compares the **monthly income** of employees who **left the company ("Yes")** versus those who **stayed ("No")**.

Observations:

- Employees who **stayed** generally have **higher median incomes** than those who left.
- The **income range** is wider for employees who stayed, with more high-income outliers.
- Employees with **lower income are more prone to attrition**, as shown by the lower median in the "Yes" group.

Higher income is associated with lower attrition, suggesting that compensation might be a key factor in employee retention.

Employee Attrition vs. Work-Life Balance:

This bar chart shows the **attrition rate (%)** based on employees' **work-life balance ratings**, where:

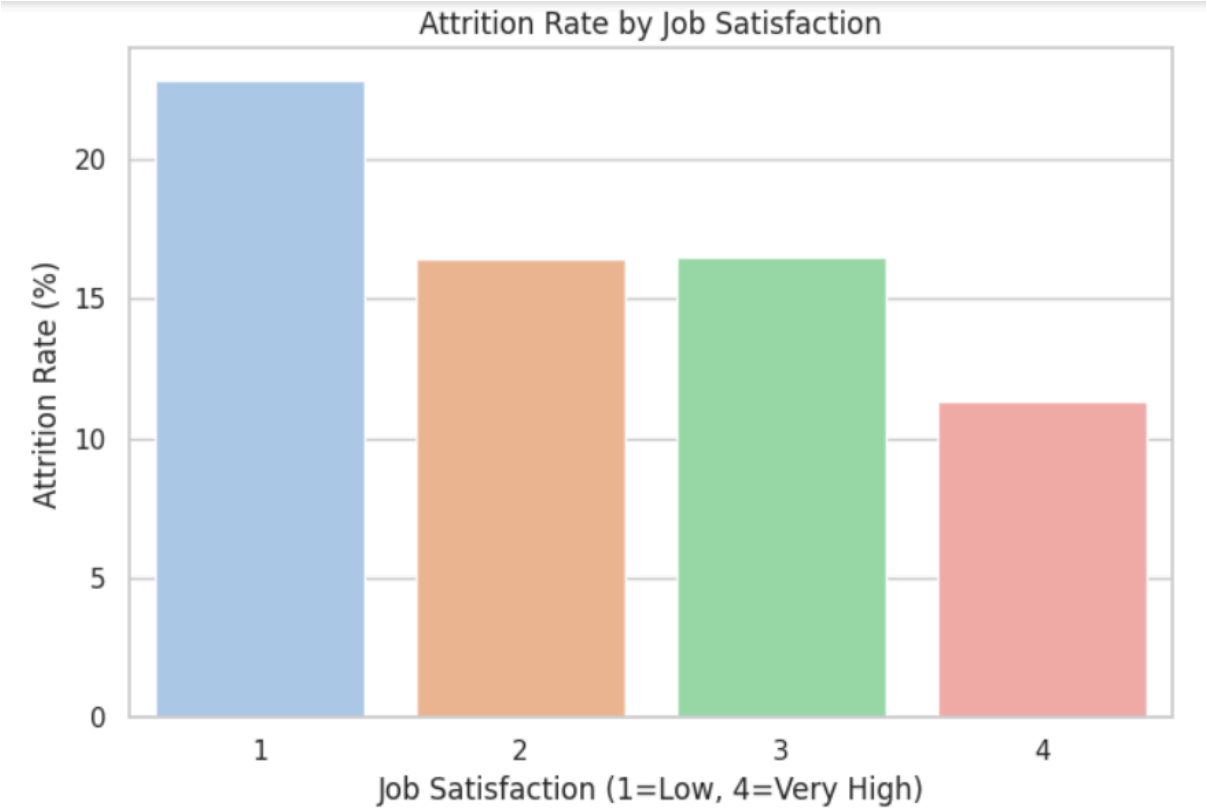
1 = Bad
4 = Best

Observations:

- Employees rating their work-life balance as **"1 (Bad)"** have the **highest attrition rate** (over 30%).
- Those who rated it **better (2–4)** have **significantly lower attrition rates**.
- Interestingly, attrition slightly rises again from **rating 3 to 4**, though it's still much lower than rating 1.

Poor work-life balance is a strong predictor of attrition, suggesting companies should improve work-life policies to retain staff.

Job Satisfaction and Employee Attrition:

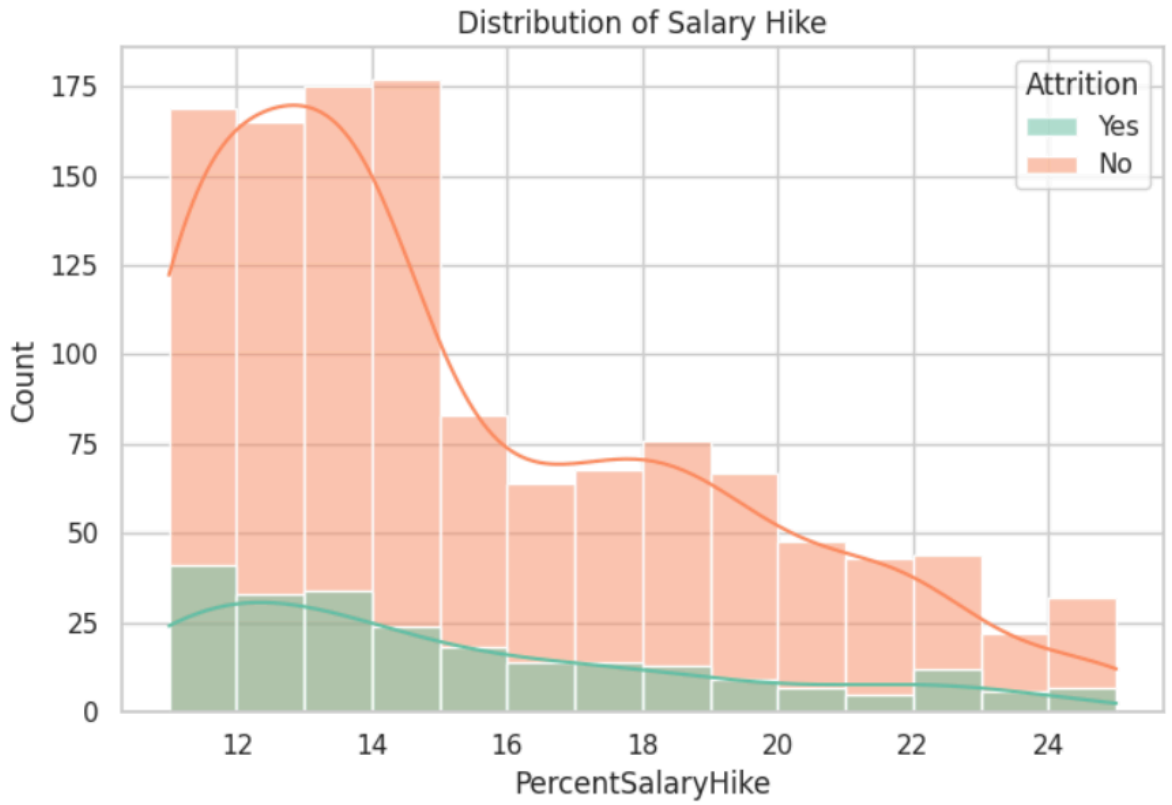


This vertical bar chart illustrates how employee attrition varies by job satisfaction levels (1 = Low, 4 = Very High):

- Employees with **lowest job satisfaction (1)** have the **highest attrition rate** at over **22%**, suggesting dissatisfaction is a strong driver of turnover.
- Attrition decreases steadily with increasing satisfaction, dropping to around **11%** for those with **very high job satisfaction (4)**.
- Interestingly, levels 2 and 3 (moderate satisfaction) show similar attrition rates (16%).

There's a clear inverse relationship between job satisfaction and attrition: the more satisfied the employees, the less likely they are to leave. This insight can help HR teams focus on improving workplace satisfaction to reduce turnover.

Salary Hike Distribution vs. Employee Attrition:



This histogram (with KDE overlay) displays the distribution of **Percent Salary Hike** for employees who **left the company (Yes)** and those who **did not (No)**.

Observations:

- Most employees, regardless of attrition, received salary hikes between **11% to 15%**.
- However, a **larger proportion of employees who stayed** (No) had higher frequency across nearly all salary hike ranges.
- The **density of those who left** decreases more sharply after about **13% hike**, suggesting higher hikes might be linked with better retention.

Insight:

Higher salary hikes are loosely associated with lower attrition, though the difference is not very strong or dramatic.

Comparative Analysis of Salary Hike and Attrition:

This boxplot compares the **percent salary hikes** between employees who **left ("Yes")** and who **stayed ("No")**.

Observations:

- The **median salary hike** appears almost the **same (14%)** for both groups.
- The **range and interquartile spread** are similar, but slightly more **upper range values** are present among employees who stayed.
- There's **no significant difference** in salary hike distribution between the two groups.

Insight:

Salary hike **alone may not be a strong predictor** of attrition, as the distributions are very similar for both groups.

Employee Attrition vs. Performance Rating:

This **stacked bar chart** compares the **number of employees** who left (`Yes`) and stayed (`No`) based on their **performance ratings** (either 3 or 4).

Observations:

- The **majority of employees** fall under **Performance Rating 3**.
- A **significant portion of attrition** occurs among those rated **3**.
- Very **few employees** are rated as **4**, and even fewer among them have left.
- **Attrition is proportionally lower** among employees with a **performance rating of 4**.

Insight:

- Employees with **higher performance ratings** tend to **stay longer**, indicating that **high-performing employees are more likely to be retained**.
- This could also suggest that **underappreciation or limited recognition** among average performers might contribute to higher attrition.

Model Performance for Predicting Employee Attrition:

1. **Classification Report Analysis:**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.95      0.90       370
           1       0.34      0.14      0.20        71

    accuracy                           0.82       441
   macro avg       0.60      0.54      0.55       441
weighted avg       0.77      0.82      0.79       441
```

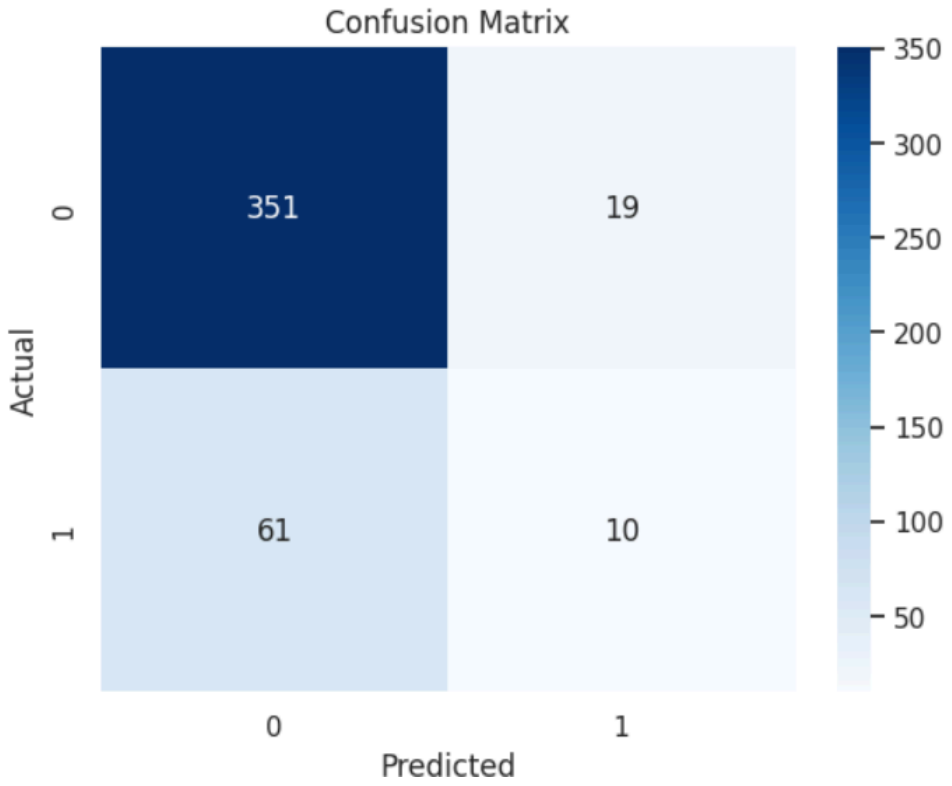
Class	Precision	Recall	F1-Score	Support
0 (No Attrition)	0.85	0.95	0.90	370
1 (Yes Attrition)	0.34	0.14	0.20	71

Accuracy: 82% - This means the model predicts correctly for 82% of cases overall.

Macro Avg (Equal weight for both classes): F1-score: 0.55 - low due to poor prediction of class 1.

Weighted Avg (Weighted by support): F1-score: 0.79 - skewed high due to class imbalance (more "No Attrition").

2. Confusion Matrix Breakdown:



Actual \ Predicted	Predicted 0	Predicted 1
Actual 0 (No Attrition)	**351** (True Negative)	**19** (FalsePositive)
Actual 1 (Yes Attrition)	**61** (False Negative)	**10** (True Positive)

- **True Positives (10)**: Correctly predicted attrition.
- **False Negatives (61)**: Actual attrition but model failed to detect.
- **False Positives (19)**: Incorrectly flagged as attrition.
- **True Negatives (351)**: Correctly predicted non-attrition.

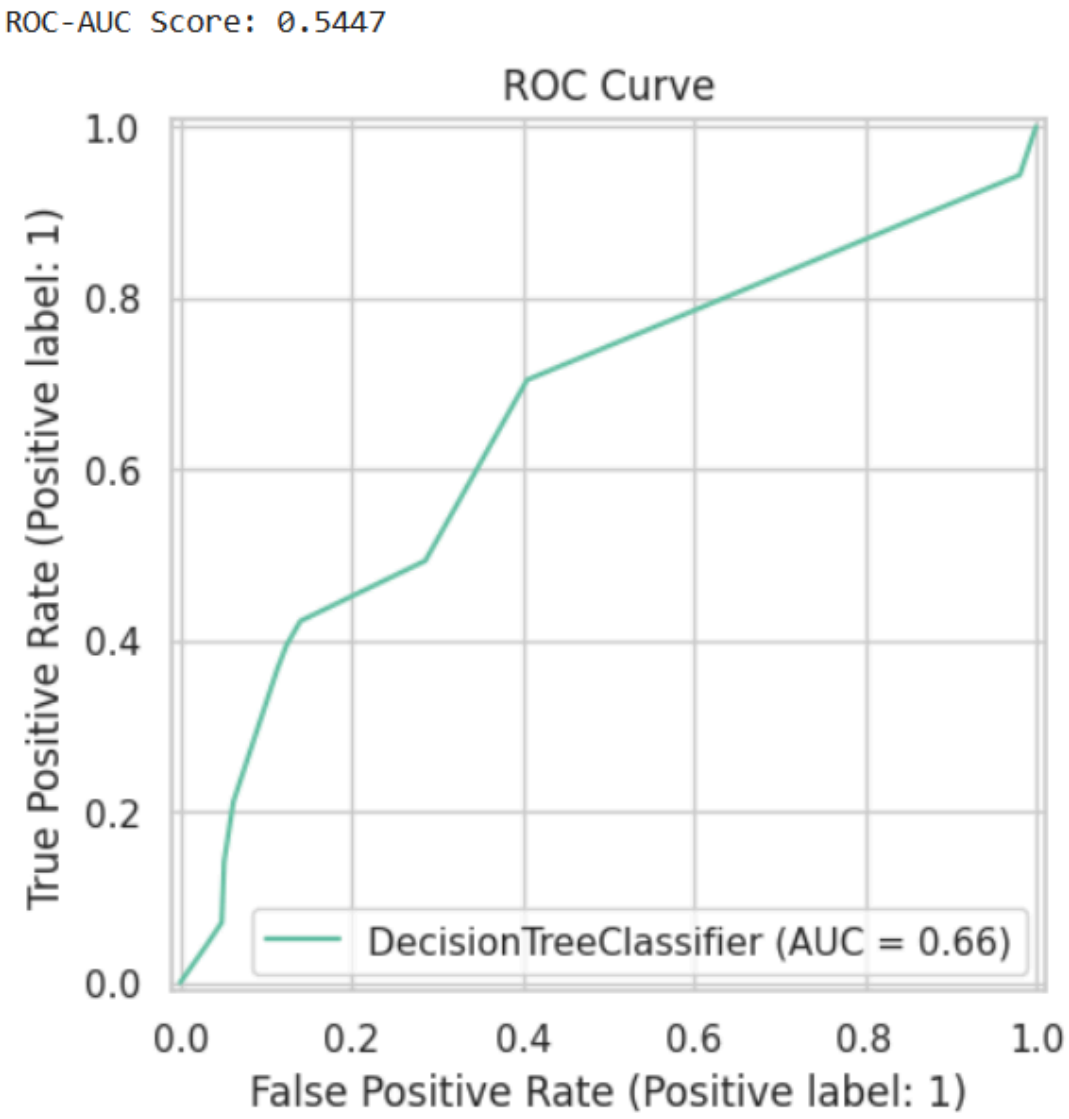
📉 Key Insight:

- The model is **biased toward predicting non-attrition (class 0)**.

- It performs **very poorly on detecting actual attrition (class 1)**:
  - Only **14% recall** for attrition class (it catches just 10 out of 71).
  - **High false negative rate (61 out of 71 attrition cases missed)**.

- This could be due to:
  - **Class imbalance** (much more “No” than “Yes” attrition).
  - **Model not learning attrition-specific patterns well.**

ROC Curve & AUC Score (Evaluating the Discriminative Power of the Classifier):



This plot displays the **ROC Curve** (Receiver Operating Characteristic) for a **Decision Tree Classifier**.

The curve illustrates the trade-off between:

- **True Positive Rate (TPR)** = Sensitivity
- **False Positive Rate (FPR)** = 1 - Specificity

Key Metrics:

- **AUC (Area Under Curve):** 0.66 (from the legend)
- **ROC-AUC Score (top left):** 0.5447

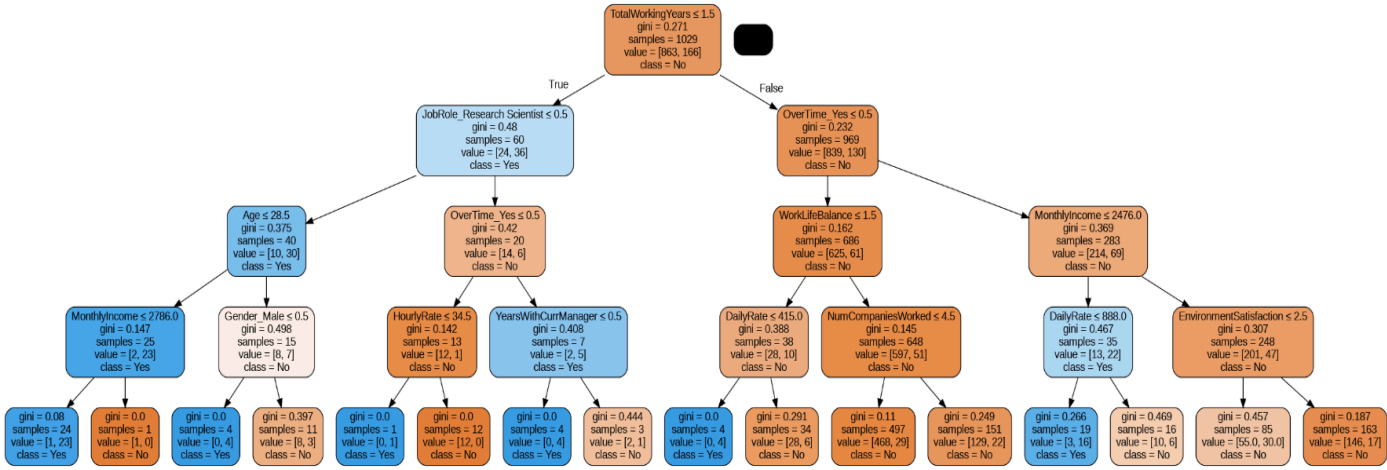
There is a discrepancy between the printed ROC-AUC score (0.5447) and the legend (AUC = 0.66), possibly due to different test splits or calculation mismatches.

Interpretation:

- An AUC of **0.5** indicates random guessing; **1.0** indicates perfect prediction.
- **0.5447 to 0.66** suggests the model is **slightly better than random**, but **not reliably distinguishing attrition from non-attrition**.
- This supports what we saw in the confusion matrix: poor identification of the "Yes" class.

Decision Tree Visualization (Decision Rules for Predicting Attrition):

```
Requirement already satisfied: pydotplus in /usr/local/lib/python3.11/dist-packages (2.0.2)
Requirement already satisfied: graphviz in /usr/local/lib/python3.11/dist-packages (0.21)
Requirement already satisfied: pyparsing>=2.0.1 in /usr/local/lib/python3.11/dist-packages (from pydotplus) (3.2.3)
```



This is the full structure of a trained Decision Tree model used to predict employee attrition.

Each node shows:

- Feature & split condition (e.g., `TotalWorkingYears ≤ 1.5`)
- Gini index (purity of the node)
- Samples at that node
- Value = [count of class 0 (No), class 1 (Yes)]
- Predicted class

Observations:

- The root node splits on `TotalWorkingYears ≤ 1.5`. Employees with low tenure tend to leave.

- Key features involved in predicting attrition:
    - `OverTime`
    - `WorkLifeBalance`
    - `MonthlyIncome`
    - `YearsWithCurrManager`
    - `EnvironmentSatisfaction`
    - `Age`, `Gender`, `JobRole`

Insights:

- Employees with:
    - Low work experience
    - Low income
    - Frequent overtime
    - Low work-life balance
    - Are more likely to leave.
- This tree helps interpret the decision logic, but also indicates overfitting risk (many deep branches).

Combined Summary:

The Decision Tree model is interpretable and captures some attrition patterns (e.g., low tenure, poor work-life balance), but:

- Performs poorly at identifying actual attrition cases.
- ROC-AUC and classification report confirm that it's struggling with class imbalance and may overfit.

Summary Table of Actions:

Risk Factor	Action
Low total working years	Strong onboarding, mentorship, and early-career support
Overtime	Limit excessive hours, offer flexibility, and monitor burnout
Poor work-life balance	Promote wellness, flexible policies, and train managers on workload
Low monthly income	Competitive pay, performance-based incentives, and clear promotion paths
Dissatisfaction with the environment	Improve culture, office conditions, and feedback handling
Weak manager relationships	Train managers, support early relationships, and mentoring programs
Job-specific risks	Deeper analysis and tailored retention strategies for critical roles

```
######################## THIS IS A WORK OF 'SURENDRAN L'
```