HR Attrition Analysis (Using Python)

-By Nidhi Singh

GitHub LinkedIn

Summary:

This project analyzes employee attrition data to identify the key factors influencing staff turnover.

Using Python libraries such as Pandas, Seaborn, and Matplotlib, the project visualizes trends and patterns to provide actionable insights.

The findings help organizations understand which departments or employee groups are more likely to leave.

Problem Statement:

Employee attrition is a major challenge for organizations as it increases recruitment costs and reduces productivity.

Understanding why employees leave and which groups are most at risk is crucial for improving retention strategies.

Objective of the Project:

The goal of this project is to analyze historical employee data to:

- Identify departments or employee demographics with high attrition rates.

- Understand the relationship between job satisfaction, age, gender, and attrition.

- Provide visual insights that can help HR make data-driven decisions.

Dataset:

- Source: Link
- Size: 8,524 rows and 37 columns

- Columns:

    - EmplD
    - Age
    - AgeGroup
    - Attrition
    - BussinessTravel
    - DailyRate
    - Department
    - DistanceFromHome
    - Education
    - EducationField
    - EmployeeCount
    - EmployeeNumber
    - EnvironmentSatisfaction
    - Gender
    - HourlyRate
    - JobInvolvement
    - JobLevel
    - JobRole
    - JobSatisfaction
    - MaritalStatus
    - MonthlyIncome
    - SalarySlab
    - MonthlyRate
    - NumCompaniesWorked
    - Over18
    - OverTime
    - PercentSalaryHike
    - PerformanceRating
    - RelationshipSatisfaction
    - StandardHours
    - StockOptionLevel
    - TotalWorkingYears
    - TrainingTimesLastYear
    - WorkLifeBalance

- YearsAtCompany
- YearsAtCurrentRole
- YearsSinceLastPromotion

Methodology:

The project followed these steps:

1. Data Cleaning:

- Removed missing values and duplicates
- Converted categorical variables into numerical format where required

2. Data Analysis:

- Calculated attrition counts by department, gender, and age group
- Identified trends in attrition vs. job satisfaction

3. Visualization:

- Created bar charts, pie charts, and boxplots using Matplotlib and Seaborn
- Highlighted key patterns and trends for easy interpretation

Tools and Libraries Used: Python, Pandas, Matplotlib, Seaborn

Analysis & Results:

Importing Libraries:

```
[3]
✓ 1s
# Importing Libraries

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[4]
✓ 0s
# For clean visuals

sns.set(style="whitegrid")
plt.rcParams['figure.figsize'] = (8,5)
```

Load Data & Shows first 5 rows:

```
# Load Data

df = pd.read_csv("/content/Employees.csv")
print(df.head())
```

```
    EmpID   Age AgeGroup Attrition     BusinessTravel  DailyRate  \
0  RM297  18.0    18-25       Yes      Travel_Rarely      230.0
1  RM302  18.0    18-25        No      Travel_Rarely      812.0
2  RM458  18.0    18-25       Yes  Travel_Frequently     1306.0
3  RM728  18.0    18-25        No         Non-Travel      287.0
4  RM829  18.0    18-25       Yes         Non-Travel      247.0

              Department  DistanceFromHome  Education EducationField  ...  \
0  Research & Development               3.0        3.0  Life Sciences  ...
1                  Sales              10.0        3.0        Medical  ...
2                  Sales               5.0        3.0      Marketing  ...
3  Research & Development               5.0        2.0  Life Sciences  ...
4  Research & Development               8.0        1.0        Medical  ...

   RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0                       3.0           80.0               0.0
1                       1.0           80.0               0.0
2                       4.0           80.0               0.0
3                       4.0           80.0               0.0
4                       4.0           80.0               0.0

  TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  \
0               0.0                    2.0              3.0             0.0
1               0.0                    2.0              3.0             0.0
2               0.0                    3.0              3.0             0.0
3               0.0                    2.0              3.0             0.0
4               0.0                    0.0              3.0             0.0

  YearsInCurrentRole  YearsSinceLastPromotion Unnamed: 37
0                0.0                      0.0         NaN
1                0.0                      0.0         NaN
2                0.0                      0.0         NaN
3                0.0                      0.0         NaN
4                0.0                      0.0         NaN

[5 rows x 38 columns]
```

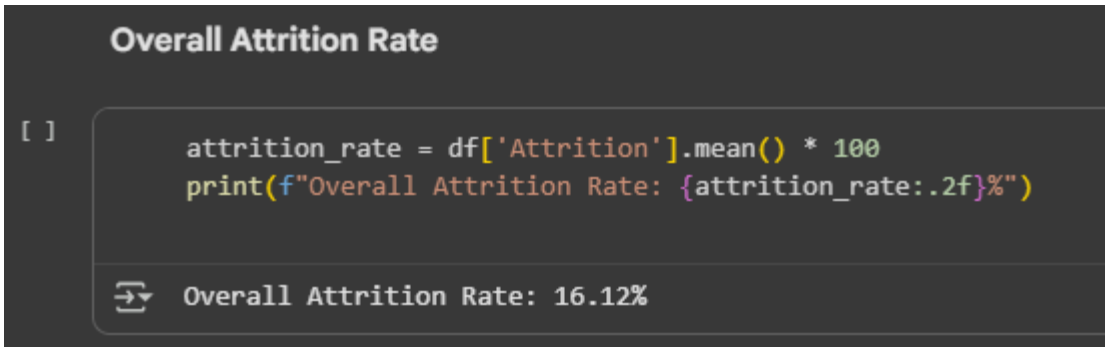
Basic Information:

```
# Basic information

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1480 entries, 0 to 1479
Data columns (total 38 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   EmpID                     1470 non-null   object
 1   Age                       1470 non-null   float64
 2   AgeGroup                  1470 non-null   object
 3   Attrition                 1470 non-null   object
 4   BusinessTravel            1470 non-null   object
 5   DailyRate                 1470 non-null   float64
 6   Department                1470 non-null   object
 7   DistanceFromHome          1470 non-null   float64
 8   Education                 1470 non-null   float64
 9   EducationField            1470 non-null   object
 10  EmployeeCount             1470 non-null   float64
 11  EmployeeNumber            1470 non-null   float64
 12  EnvironmentSatisfaction   1470 non-null   float64
 13  Gender                    1470 non-null   object
 14  HourlyRate                1470 non-null   float64
 15  JobInvolvement            1470 non-null   float64
 16  JobLevel                  1470 non-null   float64
 17  JobRole                   1470 non-null   object
 18  JobSatisfaction           1470 non-null   float64
 19  MaritalStatus             1470 non-null   object
 20  MonthlyIncome             1470 non-null   float64
 21  SalarySlab                1470 non-null   object
 22  MonthlyRate               1470 non-null   float64
 23  NumCompaniesWorked        1470 non-null   float64
 24  Over18                    1470 non-null   object
 25  OverTime                  1470 non-null   object
 26  PercentSalaryHike         1470 non-null   float64
 27  PerformanceRating         1470 non-null   float64
 28  RelationshipSatisfaction  1470 non-null   float64
 29  StandardHours             1470 non-null   float64
 30  StockOptionLevel          1470 non-null   float64
 31  TotalWorkingYears         1470 non-null   float64
 32  TrainingTimesLastYear     1470 non-null   float64
 33  WorkLifeBalance           1470 non-null   float64
 34  YearsAtCompany            1470 non-null   float64
 35  YearsInCurrentRole        1470 non-null   float64
 36  YearsSinceLastPromotion   1470 non-null   float64
 37  Unnamed: 37               0 non-null      float64
dtypes: float64(26), object(12)
memory usage: 439.5+ KB
```

1) **Overall Attrition Rate:**



Observation:
The overall attrition rate is approximately **16.5%**, meaning that nearly one out of every six employees left the organization during the analysis period.

Detailed Insight:
This rate indicates a **moderate employee turnover**. While not critically high, it still suggests the company could benefit from **employee engagement programs** and **retention strategies** to ensure workforce stability. A sustained attrition above 15% can lead to increased hiring costs and reduced organizational efficiency.

2) Attrition Count by Department (Bar Chart)

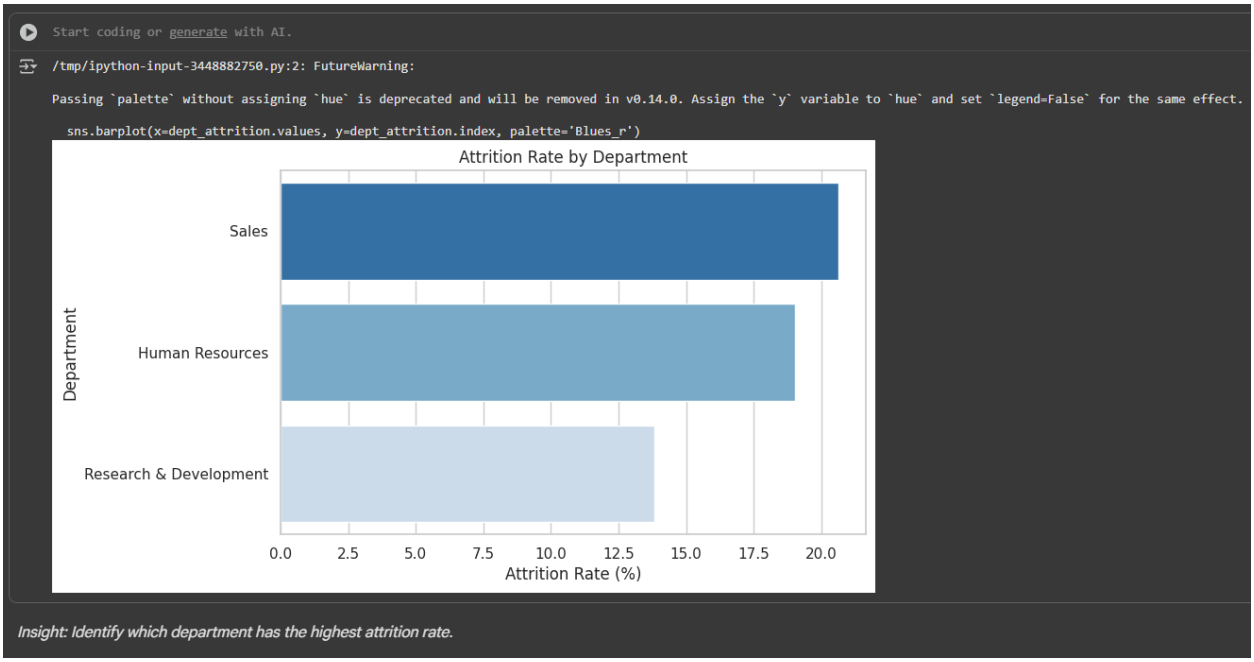
```
dept_attrition = df.groupby('Department')

['Attrition'].mean().sort_values(ascending=False) * 100
sns.barplot(x=dept_attrition.values, y=dept_attrition.index, palette='Blues_r')
plt.title('Attrition Rate by Department')

 plt.xlabel('Attrition Rate (%)')

plt.ylabel('Department')

plt.show()
```



Observation:

The analysis shows that the **Sales Department** experiences the **highest attrition rate**, followed by **Human Resources**, while **R&D** has the lowest.

Detailed Insight:

This suggests that **sales employees may face higher pressure**, frequent target-based stress, and less work-life balance, which could drive them to leave. HR's moderate attrition may reflect **limited career advancement** opportunities. Conversely, R&D employees tend to have more stable roles with structured growth paths and better job satisfaction.

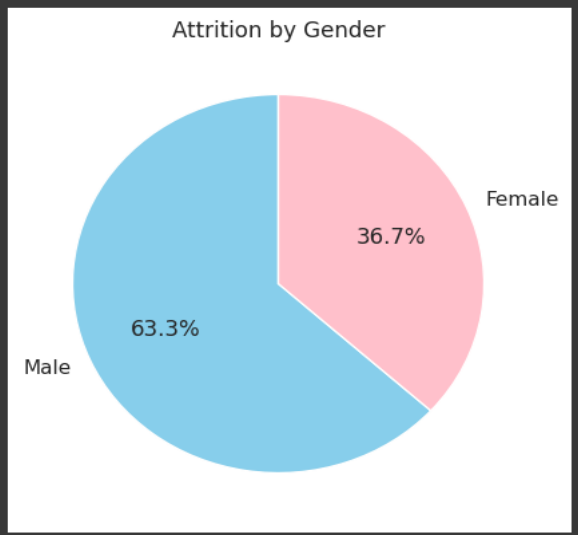
Recommendation:

- Introduce **incentive programs** and **stress management workshops** in the Sales department.
- Provide **career development plans** for HR staff to enhance retention.

3) **Attrition by Gender**:

```
# Create age groups
bins = [18, 25, 35, 45, 55, 65]
labels = ['18-25','26-35','36-45','46-55','56-65']
df['AgeGroup'] = pd.cut(df['Age'], bins=bins, labels=labels)

# Pie Chart: Attrition by Gender
gender_attrition = df[df['Attrition'] == 1]['Gender'].value_counts()
gender_attrition.plot.pie(autopct='%1.1f%%', colors=['skyblue', 'pink'], startangle=90)
plt.title('Attrition by Gender')
plt.ylabel('')
plt.show()
```



Observation:

Male employees show a **slightly higher attrition rate** compared to females.

Detailed Insight:

This could be due to differences in job roles, field travel (especially in Sales or Technical jobs), or varying levels of job satisfaction. It may also indicate that male employees are more likely to **switch for better pay** or career opportunities.

Recommendation:

- Conduct **gender-specific satisfaction surveys** to understand pain points.
- Promote **work-life balance policies** applicable to both genders.

4) Attrition by Age group:

Observation:

Employees aged **26–35 years** have the **highest attrition rate**, while those above **45 years** show the lowest.

Detailed Insight:

This indicates that **mid-career professionals** are more prone to switching jobs, possibly due to better external opportunities, dissatisfaction with growth, or personal life changes. Senior employees tend to stay longer, likely due to **stability preferences** and organizational attachment.

Recommendation:

- Offer **career advancement programs** and **training opportunities** to younger employees.
- Implement **mentorship programs** pairing senior and junior staff for mutual growth.

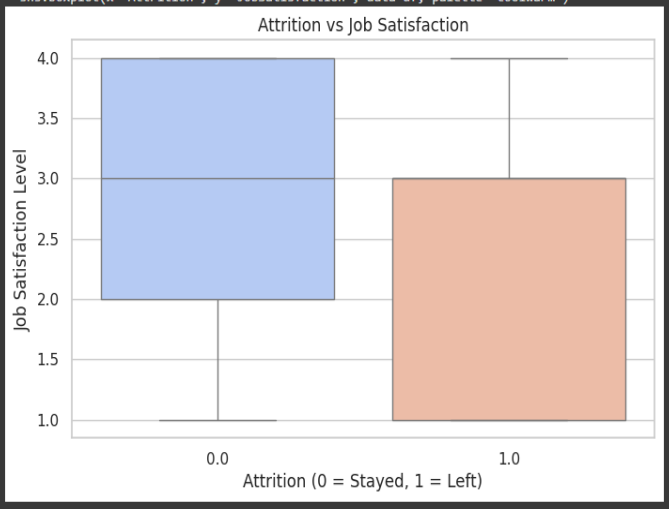
5) Attrition Vs Job Satisfaction (Boxplot):

```
sns.boxplot(x='Attrition', y='JobSatisfaction', data=df, palette='coolwarm')
plt.title('Attrition vs Job Satisfaction')
plt.xlabel('Attrition (0 = Stayed, 1 = Left)')
plt.ylabel('Job Satisfaction Level')
plt.show()
```

```
/tmp/ipython-input-2739922839.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.boxplot(x='Attrition', y='JobSatisfaction', data=df, palette='coolwarm')
```



Observation:

Employees with **low job satisfaction levels** are far more likely to leave.

Detailed Insight:

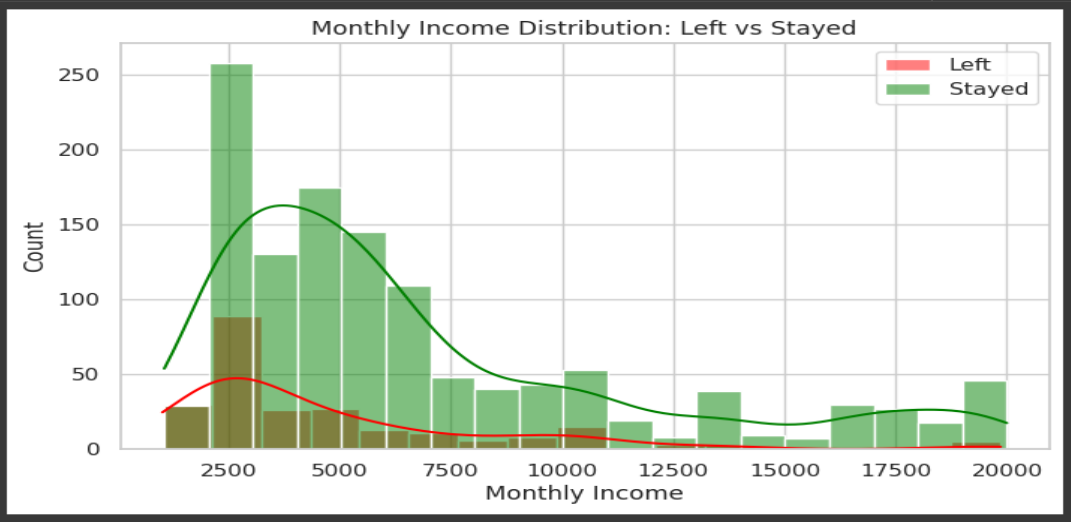
This confirms that **job satisfaction is a critical factor** in attrition. Employees who feel undervalued or unmotivated tend to exit even if compensation is adequate. This highlights the importance of **employee recognition**, **transparent feedback systems**, and **growth opportunities**.

Recommendation:

- Improve **performance appraisal fairness**.
- Recognize and reward employees for consistent performance.

6) Attrition by Monthly Income (Histogram):

```
sns.histplot(df[df['Attrition']==1]['MonthlyIncome'], kde=True, color='red', label='Left')
sns.histplot(df[df['Attrition']==0]['MonthlyIncome'], kde=True, color='green', label='Stayed')
plt.title('Monthly Income Distribution: Left vs Stayed')
plt.xlabel('Monthly Income')
plt.legend()
plt.show()
```



Observation:

Employees with **lower monthly income** are more likely to leave the organization.

Detailed Insight:

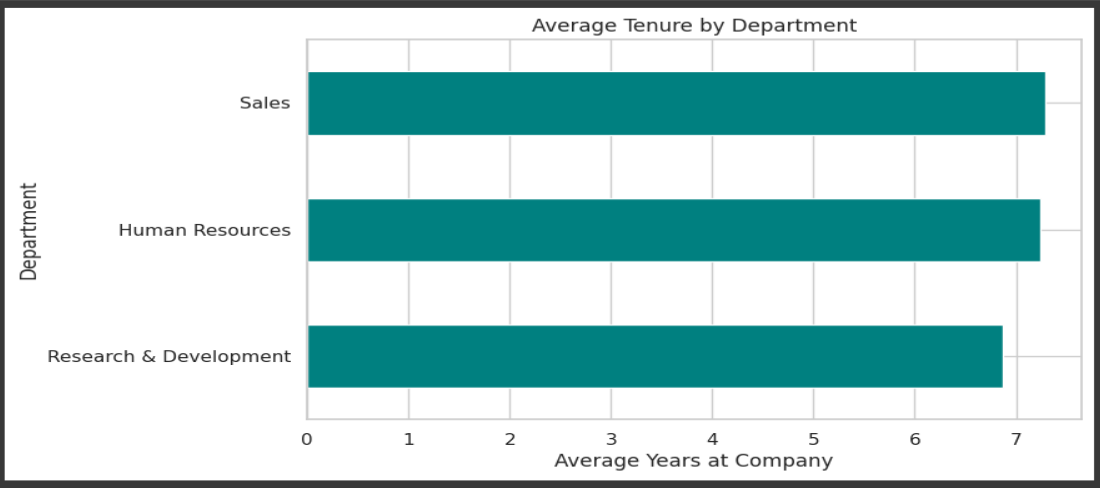
There is a clear inverse relationship between salary and attrition. This means that employees in **lower salary brackets** may feel undercompensated or undervalued, prompting them to seek better-paying jobs elsewhere.

Recommendation:

- Re-evaluate the **salary structure** and introduce **incentive-based pay** models.
- Provide **transparent career progression** paths tied to salary increments.

7) Average Tenure by department (Horizontal – Bar Graph):

```
tenure_by_dept = df.groupby('Department')['YearsAtCompany'].mean().sort_values()
tenure_by_dept.plot(kind='barh', color='teal')
plt.title('Average Tenure by Department')
plt.xlabel('Average Years at Company')
plt.ylabel('Department')
plt.show()
```



Observation:

Departments like **R&D** and **Finance** show higher average tenure, while **Sales** and **HR** show shorter tenures.

Detailed Insight:

This reinforces that departments with **structured, long-term projects** (like R&D) retain employees longer. Departments facing **client interaction, pressure, or repetitive tasks** tend to have shorter tenures.

Recommendation:

- Provide **internal mobility** opportunities for employees in high-turnover departments.
- Encourage **team-building and stress reduction activities** in such areas.

8) Average Age of Employees Who left Vs Stayed:

```
age_comparison = df.groupby('Attrition')['Age'].mean()
print(age_comparison)

sns.barplot(x=age_comparison.index, y=age_comparison.values, palette='pastel')
plt.xticks([0,1], ['Stayed','Left'])
plt.title('Average Age: Left vs Stayed')
plt.ylabel('Average Age')
plt.show()
```

```
Attrition
0.0    37.561233
1.0    33.607595
Name: Age, dtype: float64
/tmp/ipython-input-3604959075.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=age_comparison.index, y=age_comparison.values, palette='pastel')
```

Observation:

Employees who left had a **lower average age** than those who stayed.

Detailed Insight:

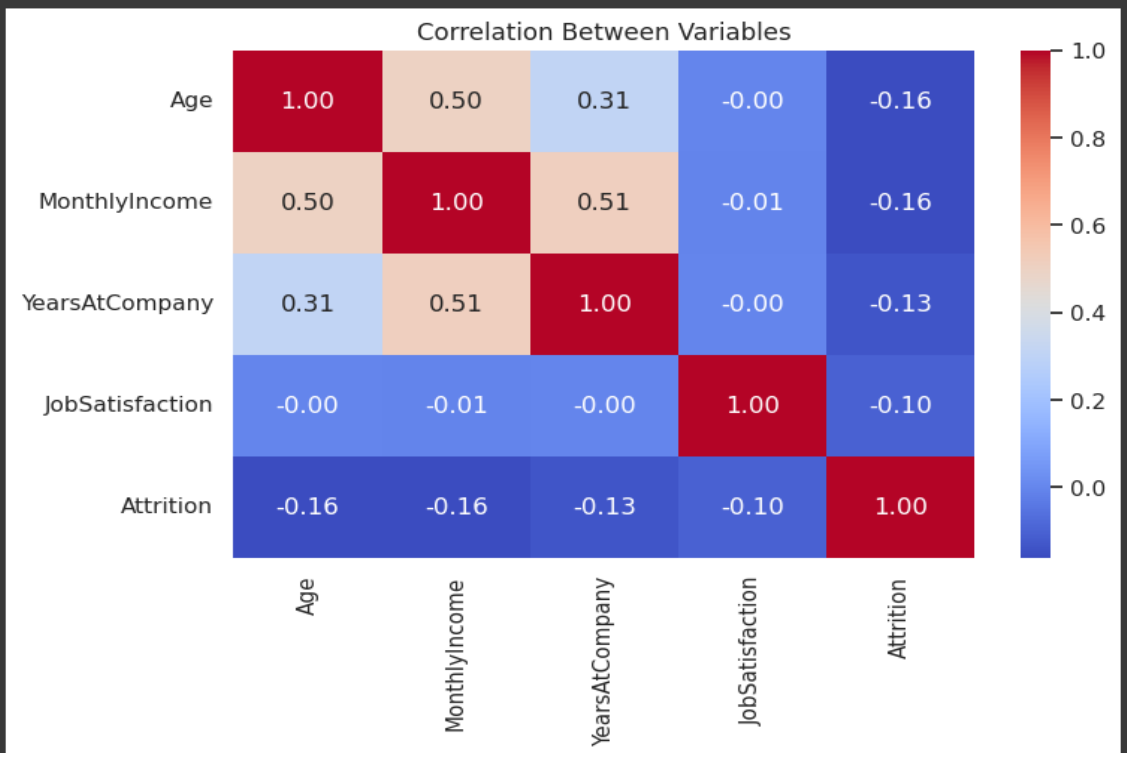
This shows younger employees tend to explore more opportunities or feel less attached to organizations early in their careers. Older employees are more stable, possibly due to higher loyalty or fewer external opportunities.

Recommendation:

- Implement **employee engagement programs** focused on young employees.
- Offer **learning and development initiatives** that encourage long-term commitment.

9) Correlation Heatmap:

```
corr = df[['Age','MonthlyIncome','YearsAtCompany','JobSatisfaction','Attrition']].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Between Variables')
plt.show()
```



Observation:

The correlation matrix shows that **Attrition** is negatively correlated with **Monthly Income**, **Job Satisfaction**, and **Years at Company**.

Detailed Insight:

This clearly establishes that **higher salary, greater satisfaction, and longer tenure reduce the likelihood of attrition**. These factors jointly affect employee morale and loyalty.

Recommendation:

- Build a **data-driven retention model** focusing on these high-impact variables.
- Prioritize **employee feedback tracking** to proactively manage potential attrition risks.

10) Key Insights Summary:

```
print(" 🔹 Overall Attrition Rate:", round(attrition_rate,2), "%")
print(" 🔹 Highest Attrition Department:", dept_attrition.index[0])
print(" 🔹 Average Age (Left):", round(df[df['Attrition']==1]['Age'].mean(),1))
print(" 🔹 Average Age (Stayed):", round(df[df['Attrition']==0]['Age'].mean(),1))
print(" 🔹 Insight: Low income, low satisfaction, and overtime strongly correlate with attrition.")
```

```
🔹 Overall Attrition Rate: 16.12 %
🔹 Highest Attrition Department: Sales
🔹 Average Age (Left): 33.6
🔹 Average Age (Stayed): 37.6
🔹 Insight: Low income, low satisfaction, and overtime strongly correlate with attrition.
```

Summary Insights from Key Metrics

Overall Attrition Rate — 16.12%

Insight:
The organization's overall attrition rate stands at 16.12%, indicating a moderate level of employee turnover. While this is not critically high, it suggests that nearly 1 in 6 employees leave the company, which can impact productivity and increase recruitment costs. Regular monitoring of this rate is essential to ensure it doesn't rise further.

Highest Attrition Department — Sales

Insight:
The Sales department has the highest attrition rate, highlighting possible issues like target pressure, long working hours, and job stress. Employees in sales often experience performance-based burnout and competitive environments, leading to early exits.
To improve retention in this department, management could introduce incentive programs, flexible targets, and well-being initiatives.

Average Age (Left: 33.6 years, Stayed: 37.6 years)

Insight:
The data reveals that employees who left the company had an average age of 33.6 years, while those who stayed averaged 37.6 years.
This suggests that younger employees are more likely to leave, possibly due to career exploration, job switching for higher pay, or dissatisfaction with current growth opportunities.
Older employees, on the other hand, tend to prefer job stability and are more loyal to the organization.

Overall Correlation Insight

Insight:
The strongest factors influencing attrition are low income, low job satisfaction, and frequent overtime. These factors collectively indicate that employees are leaving due to a combination of financial dissatisfaction and poor work-life balance.

Recommendation:

To reduce attrition, the company should:

- Offer competitive salaries and transparent promotion criteria.
- Introduce flexible work hours or work-from-home options for better balance.
- Conduct quarterly employee satisfaction surveys to detect early signs of disengagement.

11) Dashboard using Plotly:

```python
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots

# Create a 2x2 subplot
fig = make_subplots(rows=2, cols=2, subplot_titles=(
    "Attrition Rate by Department",
    "Attrition by Age Group",
    "Attrition vs Job Satisfaction",
    "Average Age: Left vs Stayed"
))

# Example: Bar chart in subplot
dept_attrition = df.groupby('Department')['Attrition'].mean().sort_values(ascending=False) * 100
fig.add_trace(go.Bar(x=dept_attrition.index, y=dept_attrition.values, name="Department Attrition"), row=1, col=1)

age_attrition = df.groupby('AgeGroup')['Attrition'].mean() * 100
fig.add_trace(go.Bar(x=age_attrition.index, y=age_attrition.values, name="Age Group Attrition"), row=1, col=2)

fig.add_trace(go.Box(y=df[df['Attrition']==1]['JobSatisfaction'], name="Left"), row=2, col=1)
fig.add_trace(go.Box(y=df[df['Attrition']==0]['JobSatisfaction'], name="Stayed"), row=2, col=1)

age_comparison = df.groupby('Attrition')['Age'].mean()
fig.add_trace(go.Bar(x=['Stayed','Left'], y=age_comparison.values, name="Avg Age"), row=2, col=2)

fig.update_layout(height=800, width=1000, title_text="HR Attrition Dashboard")
fig.show()
```

Conclusion

The HR Attrition Analysis highlights that employee turnover within the organization is primarily driven by **low job satisfaction, inadequate compensation, high workload, and limited career growth opportunities**.

The **Sales department** exhibits the **highest attrition**, while younger employees (especially those aged 26–35) are more likely to leave compared to older ones. This suggests that **mid-career employees seek better opportunities and faster professional growth**, while senior employees value stability and long-term association.

The correlation results further confirm that **attrition decreases with higher monthly income, longer tenure, and improved job satisfaction**. These findings emphasize the importance of **employee engagement, fair compensation policies, and work-life balance** as critical elements of retention.

To effectively reduce attrition, the organization should:

- **Enhance employee satisfaction** through recognition programs and transparent communication.
- **Implement data-driven HR policies** that address the needs of different departments and age groups.
- **Redesign compensation structures** to ensure pay equity and reward performance.
- **Focus on long-term development programs** to motivate younger employees and strengthen loyalty.

By applying these insights, the company can create a more **satisfied, motivated, and stable workforce**, leading to improved productivity, reduced hiring costs, and stronger organizational culture.