Title: Employee Attrition & Performance Analysis using Python

---

Objective:

The objective of this project is to analyze employee attrition trends and identify key factors influencing employees to leave the organization. The goal is to provide actionable insights to the HR department to reduce attrition and improve employee retention.

Tools & Technologies Used:

- Python
- Pandas (Data Handling)
- Seaborn & Matplotlib (Data Visualization)
- Jupyter Notebook
- CSV File (Dataset)

Dataset Overview:

The dataset contains employee records with demographic, job role, compensation, and satisfaction-related information.

Total Records: 1470 employees

Key Columns:

Age, Gender, Department, EducationField, JobRole, MonthlyIncome, TotalWorkingYears, WorkLifeBalance, JobSatisfaction, MaritalStatus, Attrition (Yes/No)

Data Exploration & Visual Analysis:

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

sns.set_style('whitegrid')

plt.rcParams['figure.figsize'] = (8,5)
```

 # Load Data

```
df = pd.read_csv(r"E:/HR_Analytics_Python_Project/data/HR_Attrition.csv")

print(df.head())

print(df.info())
```

```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1           80                 0
1  ...                         4           80                 1
2  ...                         2           80                 0
3  ...                         3           80                 0
4  ...                         4           80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
3                  8                      3               3               8
4                  6                      3               3               2

   YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                   4                        0                     5
1                   7                        1                     7
2                   0                        0                     0
3                   7                        3                     0
4                   2                        2                     2

[5 rows x 35 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
None
```

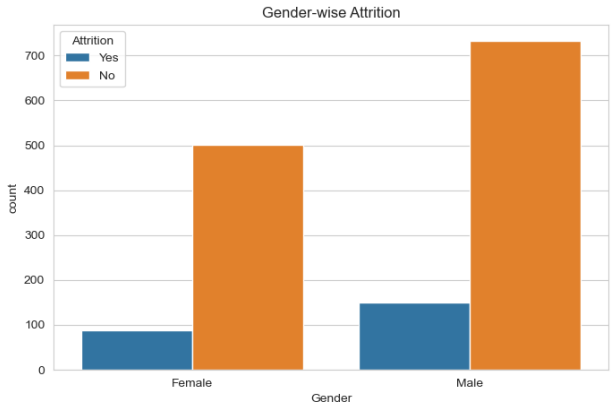
 # Basic Stats

```
print("Total Employees:", len(df))

print("Attrition Rate:\n", df['Attrition'].value_counts(normalize=True) * 100)
```

```
Total Employees: 1470
Attrition Rate:
 Attrition
No     83.877551
Yes    16.122449
Name: proportion, dtype: float64
```

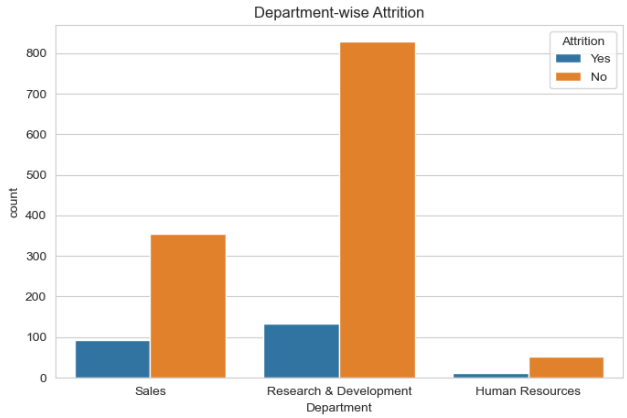
 # Gender-wise Attrition

```
sns.countplot(x='Gender', hue='Attrition', data=df)
plt.title("Gender-wise Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/gender_attrition.png')
plt.show()
plt.clf()
```



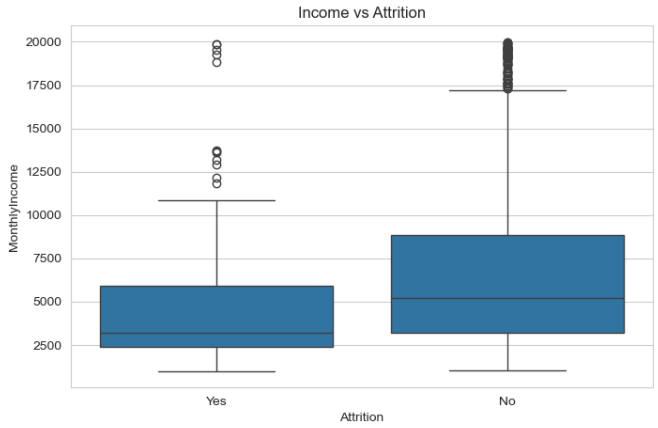
 # Department-wise Attrition

```
sns.countplot(x='Department', hue='Attrition', data=df)
plt.title("Department-wise Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/department_attrition.png')
plt.show()
plt.clf()
```



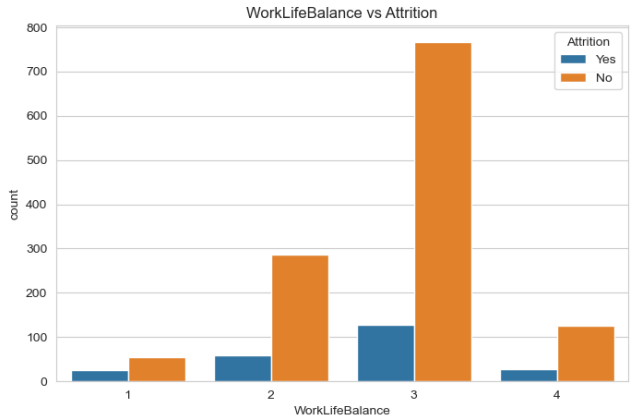
 # Income vs Attrition

```
sns.boxplot(x='Attrition', y='MonthlyIncome', data=df)
plt.title("Income vs Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/income_attrition.png')
plt.show()
plt.clf()
```



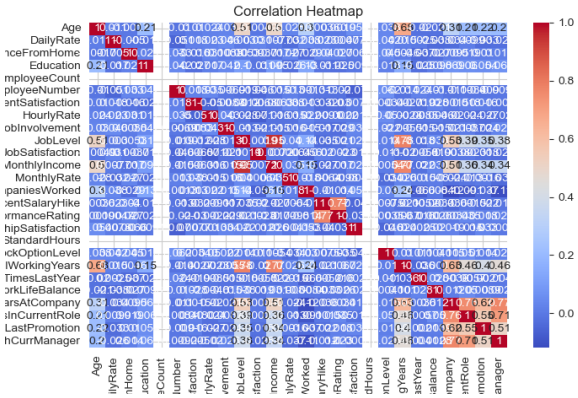
 # WorkLifeBalance impact

```
sns.countplot(x='WorkLifeBalance', hue='Attrition', data=df)
plt.title("WorkLifeBalance vs Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/wlb_attrition.png')
plt.show()
plt.clf()
```



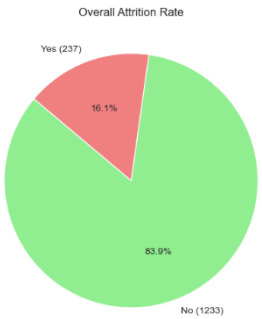
 # Correlation Heatmap

```
corr = df.corr(numeric_only=True)

sns.heatmap(corr, annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/correlation.png')

plt.show()

plt.clf()
```



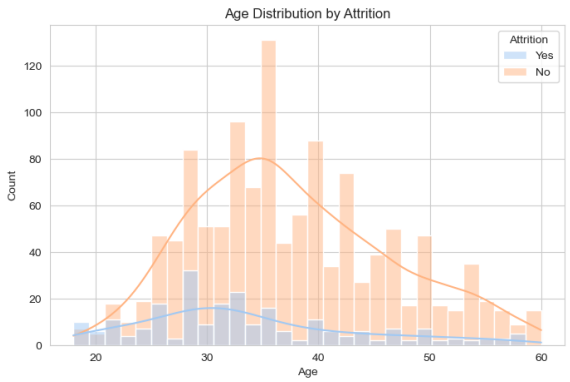
 # Overall Attrition Pie Chart

```
attrition_counts = df['Attrition'].value_counts()
plt.figure(figsize=(6,6))
plt.pie(attrition_counts, labels=[f"{label} ({count})" for label, count in zip(attrition_counts.index, attrition_counts.values)], autopct='%1.1f%%', colors=['lightgreen','lightcoral'], startangle=140)
plt.title('Overall Attrition Rate')
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/overall_attrition_pie.png')
plt.show()
plt.clf()
```



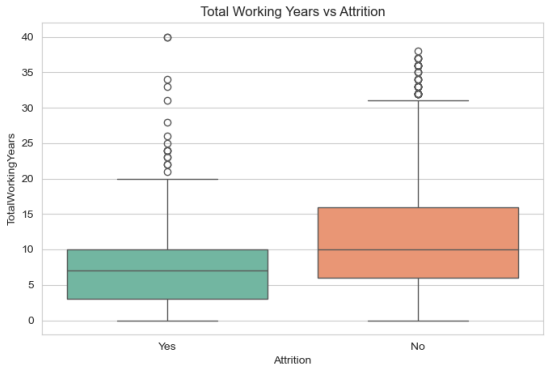
 # Age Distribution with Attrition

```
plt.figure(figsize=(8,5))
sns.histplot(data=df, x='Age', hue='Attrition', kde=True, palette='pastel', bins=30)
plt.title('Age Distribution by Attrition')
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/age_distribution.png')
plt.show()
plt.clf()
```



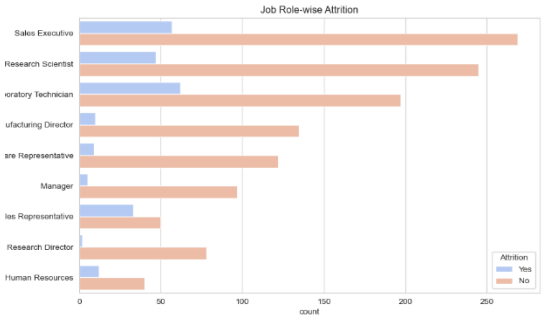
 # Total Working Years vs Attrition

```
plt.figure(figsize=(8,5))
sns.boxplot(x='Attrition', y='TotalWorkingYears', data=df, palette='Set2')
plt.title('Total Working Years vs Attrition')
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/working_years_vs_attrition.png')
plt.show()
plt.clf()
```



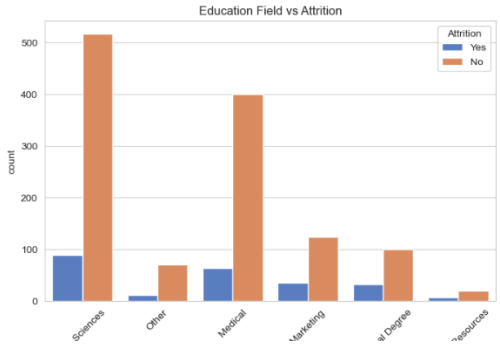
 # Job Role-wise Attrition

```
plt.figure(figsize=(10,6))
sns.countplot(y='JobRole', hue='Attrition', data=df, palette='coolwarm')
plt.title('Job Role-wise Attrition')
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/jobrole_attrition.png')
plt.show()
plt.clf()
```



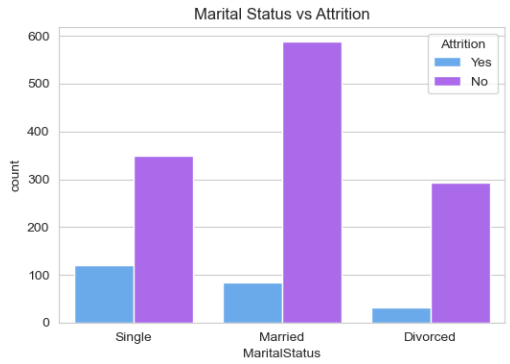
 # Education Field ka Impact Analysis

```
plt.figure(figsize=(8,5))
sns.countplot(x='EducationField', hue='Attrition', data=df, palette='muted')
plt.title("Education Field vs Attrition")
plt.xticks(rotation=45)
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/education_field_attrition.png')
plt.show()
plt.clf()
```



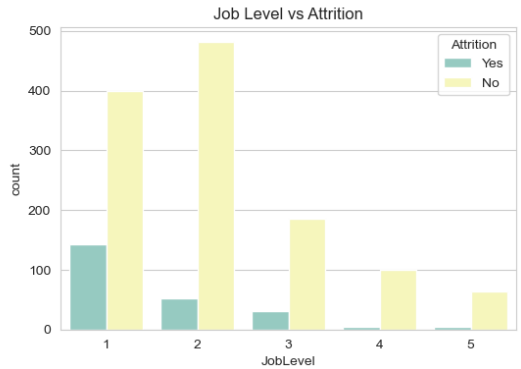
 # Marital Status vs Attrition

```
plt.figure(figsize=(6,4))
sns.countplot(x='MaritalStatus', hue='Attrition', data=df, palette='cool')
plt.title("Marital Status vs Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/marital_status_attrition.png')
plt.show()
plt.clf()
```



 # Attrition by Job Level

```
plt.figure(figsize=(6,4))
sns.countplot(x='JobLevel', hue='Attrition', data=df, palette='Set3')
plt.title("Job Level vs Attrition")
plt.savefig(r'E:/HR_Analytics_Python_Project/output/graphs/job_level_attrition.png')
plt.show()
plt.clf()
```



Summary

Higher attrition observed among employees in Sales department.
Female employees show slightly higher attrition trends..
Employees with lower Work-Life Balance tend to leave more
Low-income and low Job Satisfaction employees are at greater risk.
Education field and job level also play a role in attrition patterns.

Conclusion

- Improve work-life balance programs, especially in Sales department
- Review compensation for employees in lower income brackets
- Conduct regular employee engagement and satisfaction surveys
- Focus retention efforts on departments with historically high attrition
- Provide career growth paths to employees with low job levels