# employee-attrition-1

July 24, 2024

***EMPLOYEE_ATTRITION_REPORT***

---

**Problem Statement:**

XYZ company which was established a few years back is facing around a 15% attrition rate for a couple of years. And it's majorly affecting the company in many aspects. In order to understand why employees are leaving the company and reduce the attrition rate XYZ company has approached an HR analytics consultancy for analyzing the data they have. You are playing the HR analyst role in this project and building a dashboard which can help the organization in making data-driven decisions.

[4]:
```
import matplotlib.pyplot as plt
import numpy as np
import warnings
warnings.filterwarnings("ignore")
import pandas as pd
import seaborn as sns
```

[5]:
```
attr = pd.read_csv(r"/content/data.csv")
```

[6]:
```
attr.head(10)
```

[6]:
```
   EmployeeID  Age Attrition     BusinessTravel              Department  \
0           1   51        No      Travel_Rarely                   Sales
1           2   31       Yes  Travel_Frequently  Research & Development
2           3   32        No  Travel_Frequently  Research & Development
3           4   38        No         Non-Travel  Research & Development
4           5   32        No      Travel_Rarely  Research & Development
5           6   46        No      Travel_Rarely  Research & Development
6           7   28       Yes      Travel_Rarely  Research & Development
7           8   29        No      Travel_Rarely  Research & Development
8           9   31        No      Travel_Rarely  Research & Development
9          10   25        No         Non-Travel  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  Gender  …  \
0                 6          2  Life Sciences              1  Female  …
1                10          1  Life Sciences              1  Female  …
2                17          4          Other              1    Male  …
```

```
3                 2          5  Life Sciences              1    Male  …
4                10          1        Medical              1    Male  …
5                 8          3  Life Sciences              1  Female  …
6                11          2        Medical              1    Male  …
7                18          3  Life Sciences              1    Male  …
8                 1          3  Life Sciences              1    Male  …
9                 7          4        Medical              1  Female  …

   TotalWorkingYears TrainingTimesLastYear YearsAtCompany  \
0                1.0                     6              1
1                6.0                     3              5
2                5.0                     2              5
3               13.0                     5              8
4                9.0                     2              6
5               28.0                     5              7
6                5.0                     2              0
7               10.0                     2              0
8               10.0                     2              9
9                6.0                     2              6

   YearsSinceLastPromotion  YearsWithCurrManager EnvironmentSatisfaction  \
0                        0                     0                     3.0
1                        1                     4                     3.0
2                        0                     3                     2.0
3                        7                     5                     4.0
4                        0                     4                     4.0
5                        7                     7                     3.0
6                        0                     0                     1.0
7                        0                     0                     1.0
8                        7                     8                     2.0
9                        1                     5                     2.0

   JobSatisfaction  WorkLifeBalance  JobInvolvement  PerformanceRating
0              4.0              2.0               3                  3
1              2.0              4.0               2                  4
2              2.0              1.0               3                  3
3              4.0              3.0               2                  3
4              1.0              3.0               3                  3
5              2.0              2.0               3                  3
6              3.0              1.0               3                  4
7              2.0              3.0               3                  4
8              4.0              3.0               3                  4
9              1.0              3.0               3                  3

[10 rows x 29 columns]
```

```
[7]: attr.shape
```

[7]:

```
(4410, 29)
```

[8]:

```
attr.describe()
```

[8]:

```
        EmployeeID          Age  DistanceFromHome    Education  EmployeeCount  \
count  4410.000000  4410.000000       4410.000000  4410.000000         4410.0
mean   2205.500000    36.923810          9.192517     2.912925            1.0
std    1273.201673     9.133301          8.105026     1.023933            0.0
min       1.000000    18.000000          1.000000     1.000000            1.0
25%    1103.250000    30.000000          2.000000     2.000000            1.0
50%    2205.500000    36.000000          7.000000     3.000000            1.0
75%    3307.750000    43.000000         14.000000     4.000000            1.0
max    4410.000000    60.000000         29.000000     5.000000            1.0

          JobLevel  MonthlyIncome  NumCompaniesWorked  PercentSalaryHike  \
count  4410.000000    4410.000000         4391.000000        4410.000000
mean      2.063946   65029.312925            2.694830          15.209524
std       1.106689   47068.888559            2.498887           3.659108
min       1.000000   10090.000000            0.000000          11.000000
25%       1.000000   29110.000000            1.000000          12.000000
50%       2.000000   49190.000000            2.000000          14.000000
75%       3.000000   83800.000000            4.000000          18.000000
max       5.000000  199990.000000            9.000000          25.000000

       StandardHours  …  TotalWorkingYears  TrainingTimesLastYear  \
count         4410.0  …        4401.000000            4410.000000
mean             8.0  …          11.279936               2.799320
std              0.0  …           7.782222               1.288978
min              8.0  …           0.000000               0.000000
25%              8.0  …           6.000000               2.000000
50%              8.0  …          10.000000               3.000000
75%              8.0  …          15.000000               3.000000
max              8.0  …          40.000000               6.000000

       YearsAtCompany  YearsSinceLastPromotion  YearsWithCurrManager  \
count     4410.000000              4410.000000           4410.000000
mean         7.008163                 2.187755              4.123129
std          6.125135                 3.221699              3.567327
min          0.000000                 0.000000              0.000000
25%          3.000000                 0.000000              2.000000
50%          5.000000                 1.000000              3.000000
75%          9.000000                 3.000000              7.000000
max         40.000000                15.000000             17.000000

       EnvironmentSatisfaction  JobSatisfaction  WorkLifeBalance  \
count              4385.000000      4390.000000      4372.000000
mean                  2.723603         2.728246         2.761436
```

```
std                   1.092756         1.101253         0.706245
min                   1.000000         1.000000         1.000000
25%                   2.000000         2.000000         2.000000
50%                   3.000000         3.000000         3.000000
75%                   4.000000         4.000000         3.000000
max                   4.000000         4.000000         4.000000

       JobInvolvement  PerformanceRating
count     4410.000000        4410.000000
mean         2.729932           3.153741
std          0.711400           0.360742
min          1.000000           3.000000
25%          2.000000           3.000000
50%          3.000000           3.000000
75%          3.000000           3.000000
max          4.000000           4.000000

[8 rows x 21 columns]
```

[9]:

```
columns = list(attr.columns)
columns
```

[9]:

```
['EmployeeID',
 'Age',
 'Attrition',
 'BusinessTravel',
 'Department',
 'DistanceFromHome',
 'Education',
 'EducationField',
 'EmployeeCount',
 'Gender',
 'JobLevel',
 'JobRole',
 'MaritalStatus',
 'MonthlyIncome',
 'NumCompaniesWorked',
 'Over18',
 'PercentSalaryHike',
 'StandardHours',
 'StockOptionLevel',
 'TotalWorkingYears',
 'TrainingTimesLastYear',
 'YearsAtCompany',
 'YearsSinceLastPromotion',
 'YearsWithCurrManager',
 'EnvironmentSatisfaction',
 'JobSatisfaction',
 'WorkLifeBalance',
 'JobInvolvement',
 'PerformanceRating']
```

[10]:

```
attr.isnull().sum()
```

[10]:

```
EmployeeID                  0
Age                         0
Attrition                   0
BusinessTravel              0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
Gender                      0
JobLevel                    0
JobRole                     0
MaritalStatus               0
MonthlyIncome               0
NumCompaniesWorked         19
Over18                      0
PercentSalaryHike           0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           9
TrainingTimesLastYear       0
YearsAtCompany              0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
EnvironmentSatisfaction    25
JobSatisfaction            20
WorkLifeBalance            38
JobInvolvement              0
PerformanceRating           0
dtype: int64
```

[11]:

```
attr.dropna(inplace=True)
```

[12]:

```
attr.shape
```

[12]:

```
(4300, 29)
```

[13]:

```
attr.describe()
```

```
[13]:         EmployeeID          Age  DistanceFromHome    Education  EmployeeCount  \
      count  4300.000000  4300.000000       4300.000000  4300.000000         4300.0
      mean   2211.695116    36.926977          9.197907     2.913256            1.0
      std    1272.117692     9.146517          8.097059     1.024774            0.0
      min       1.000000    18.000000          1.000000     1.000000            1.0
      25%    1110.750000    30.000000          2.000000     2.000000            1.0
      50%    2215.500000    36.000000          7.000000     3.000000            1.0
      75%    3314.250000    43.000000         14.000000     4.000000            1.0
      max    4409.000000    60.000000         29.000000     5.000000            1.0

                JobLevel  MonthlyIncome  NumCompaniesWorked  PercentSalaryHike  \
      count  4300.000000    4300.000000         4300.000000        4300.000000
      mean      2.066977   65059.844186            2.690000          15.210698
      std       1.106633   47045.398914            2.495764           3.662777
      min       1.000000   10090.000000            0.000000          11.000000
      25%       1.000000   29260.000000            1.000000          12.000000
      50%       2.000000   49360.000000            2.000000          14.000000
      75%       3.000000   83802.500000            4.000000          18.000000
      max       5.000000  199990.000000            9.000000          25.000000

             StandardHours  …  TotalWorkingYears  TrainingTimesLastYear  \
      count         4300.0  …        4300.000000            4300.000000
      mean             8.0  …          11.285116               2.796279
      std              0.0  …           7.790052               1.290142
      min              8.0  …           0.000000               0.000000
      25%              8.0  …           6.000000               2.000000
      50%              8.0  …          10.000000               3.000000
      75%              8.0  …          15.000000               3.000000
      max              8.0  …          40.000000               6.000000

             YearsAtCompany  YearsSinceLastPromotion  YearsWithCurrManager  \
      count     4300.000000              4300.000000           4300.000000
      mean         7.026047                 2.190000              4.132558
      std          6.148036                 3.230818              3.565831
      min          0.000000                 0.000000              0.000000
      25%          3.000000                 0.000000              2.000000
      50%          5.000000                 1.000000              3.000000
      75%          9.250000                 3.000000              7.000000
      max         40.000000                15.000000             17.000000

             EnvironmentSatisfaction  JobSatisfaction  WorkLifeBalance  \
      count              4300.000000      4300.000000      4300.000000
      mean                  2.723953         2.724884         2.761163
      std                   1.093802         1.101875         0.707800
      min                   1.000000         1.000000         1.000000
      25%                   2.000000         2.000000         2.000000
      50%                   3.000000         3.000000         3.000000
```

```
75%                   4.000000          4.000000          3.000000
max                   4.000000          4.000000          4.000000

       JobInvolvement  PerformanceRating
count     4300.000000        4300.000000
mean         2.728837           3.153953
std          0.710769           0.360946
min          1.000000           3.000000
25%          2.000000           3.000000
50%          3.000000           3.000000
75%          3.000000           3.000000
max          4.000000           4.000000

[8 rows x 21 columns]
```

[14]:

```
attr.isnull().sum()
```

[14]:

```
EmployeeID                 0
Age                        0
Attrition                  0
BusinessTravel             0
Department                 0
DistanceFromHome           0
Education                  0
EducationField             0
EmployeeCount              0
Gender                     0
JobLevel                   0
JobRole                    0
MaritalStatus              0
MonthlyIncome              0
NumCompaniesWorked         0
Over18                     0
PercentSalaryHike          0
StandardHours              0
StockOptionLevel           0
TotalWorkingYears          0
TrainingTimesLastYear      0
YearsAtCompany             0
YearsSinceLastPromotion    0
YearsWithCurrManager       0
EnvironmentSatisfaction    0
JobSatisfaction            0
WorkLifeBalance            0
JobInvolvement             0
PerformanceRating          0
dtype: int64
```

[15]:

```
attr.duplicated()
```

[15]:

```
0       False
1       False
2       False
3       False
4       False
        …
4404    False
4405    False
4406    False
4407    False
4408    False
Length: 4300, dtype: bool
```

[16]:

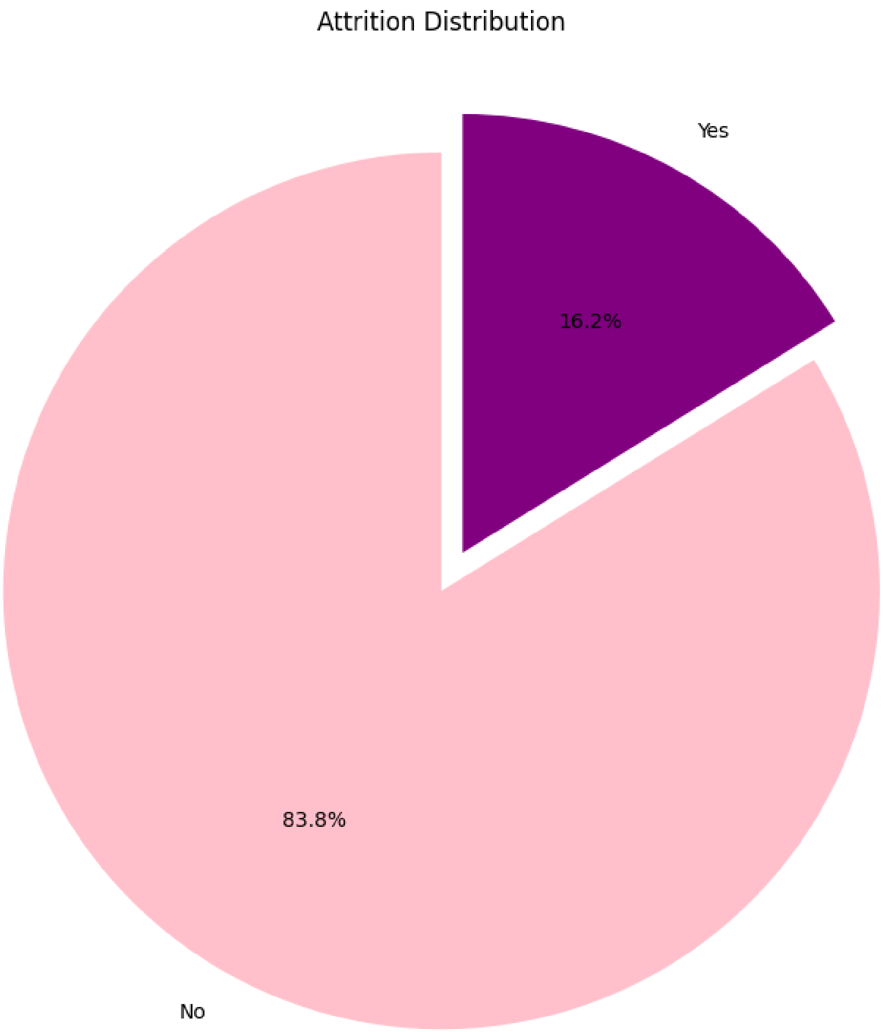
```
attr.nunique()
```

[16]:

```
EmployeeID                 4300
Age                          43
Attrition                     2
BusinessTravel                3
Department                    3
DistanceFromHome             29
Education                     5
EducationField                6
EmployeeCount                 1
Gender                        2
JobLevel                      5
JobRole                       9
MaritalStatus                 3
MonthlyIncome              1349
NumCompaniesWorked           10
Over18                        1
PercentSalaryHike            15
StandardHours                 1
StockOptionLevel              4
TotalWorkingYears            40
TrainingTimesLastYear         7
YearsAtCompany               37
YearsSinceLastPromotion      16
YearsWithCurrManager         18
EnvironmentSatisfaction       4
JobSatisfaction               4
WorkLifeBalance               4
JobInvolvement                4
PerformanceRating             2
dtype: int64
```

[17]:

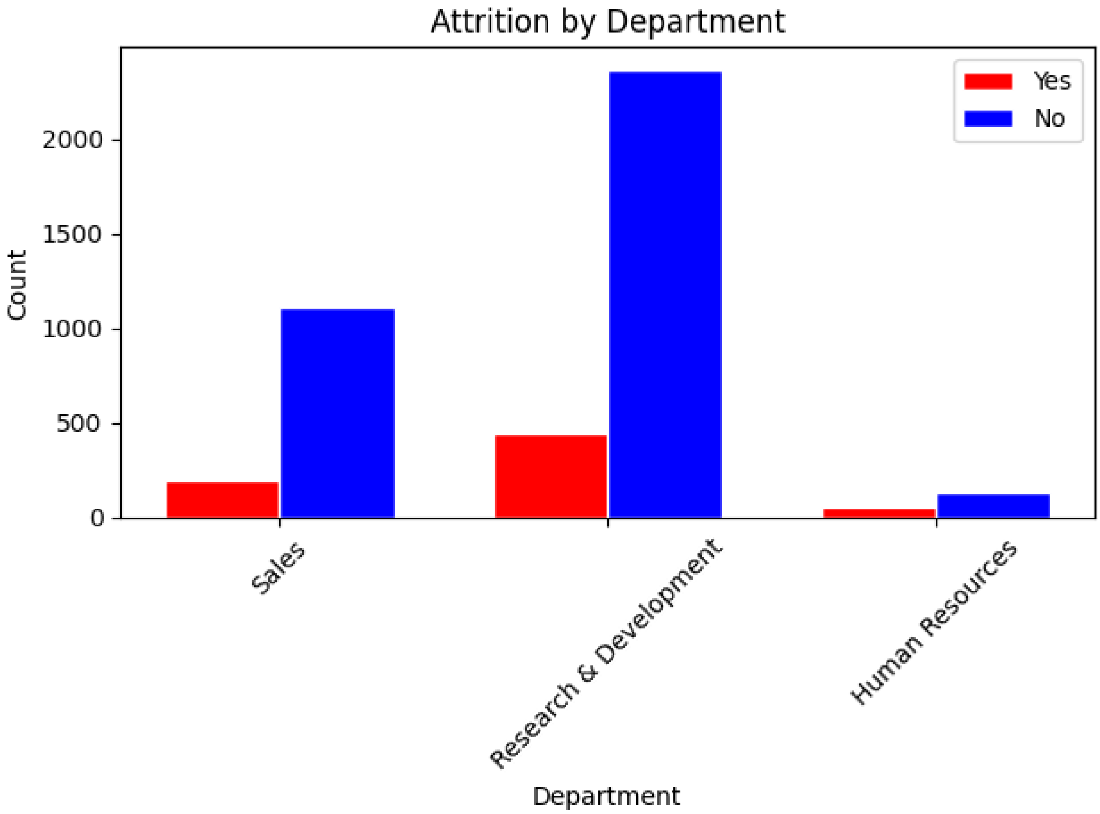
```
attrition_counts = attr['Attrition'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(attrition_counts, labels=attrition_counts.index, autopct='%.1f%%',␣
 ↪explode=(0, 0.1), colors=['pink', 'purple'], startangle=90)
plt.title('Attrition Distribution')
plt.tight_layout()
plt.show()
```



[18]:

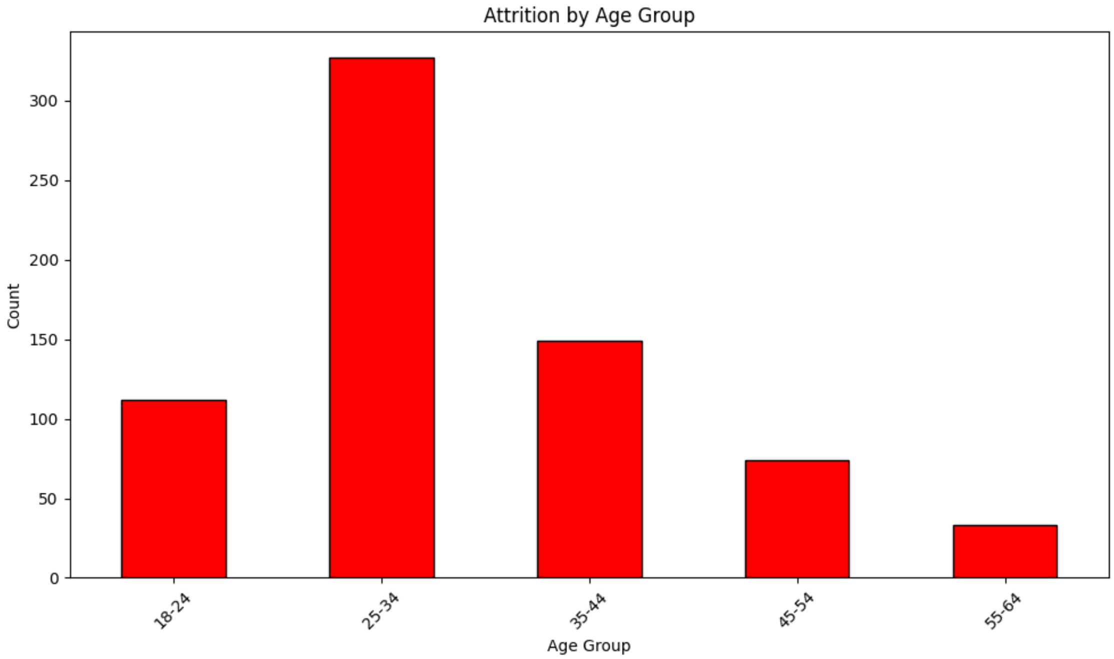
```
departments = attr['Department'].unique()
attrition_yes = attr[attr['Attrition'] == 'Yes'].groupby('Department').size()
attrition_no = attr[attr['Attrition'] == 'No'].groupby('Department').size()
```

```
attrition_yes = attrition_yes.reindex(departments, fill_value=0)
attrition_no = attrition_no.reindex(departments, fill_value=0)

bar_width = 0.35
r1 = range(len(departments))
r2 = [x + bar_width for x in r1]

plt.bar(r1, attrition_yes, color='red', width=bar_width, edgecolor='white',␣
 ↪label='Yes')
plt.bar(r2, attrition_no, color='blue', width=bar_width, edgecolor='white',␣
 ↪label='No')

plt.xlabel('Department')
plt.ylabel('Count')
plt.title('Attrition by Department')
plt.xticks([r + bar_width/2 for r in range(len(departments))], departments,␣
 ↪rotation=45)
plt.legend()
plt.tight_layout()
plt.show()
```



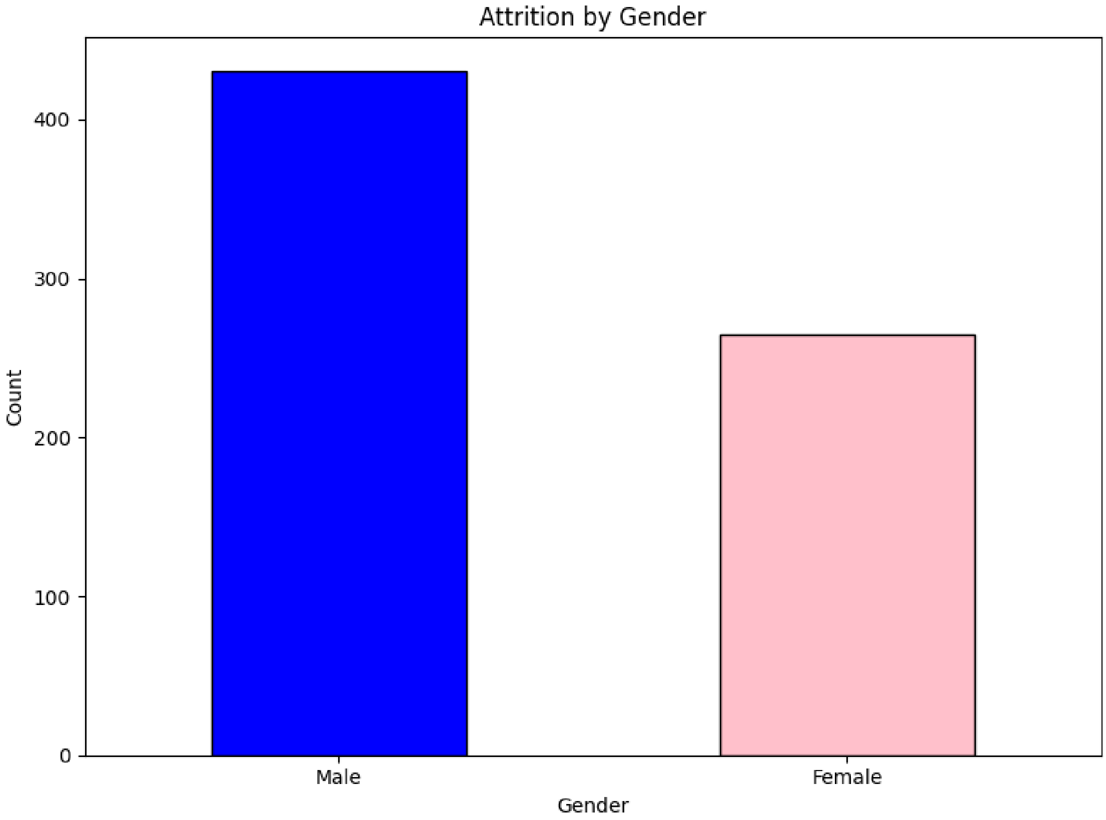
[19]:

```
bins = [18, 25, 35, 45, 55, 65]
labels = ['18-24', '25-34', '35-44', '45-54', '55-64']
attr['AgeGroup'] = pd.cut(attr['Age'], bins=bins, labels=labels, right=False)
attrition_age_group = attr[attr['Attrition'] == 'Yes']['AgeGroup'].
 ↪value_counts().reindex(labels, fill_value=0)

# Plot the histogram
plt.figure(figsize=(10, 6))
attrition_age_group.plot(kind='bar', color='red', edgecolor='black')
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Attrition by Age Group')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

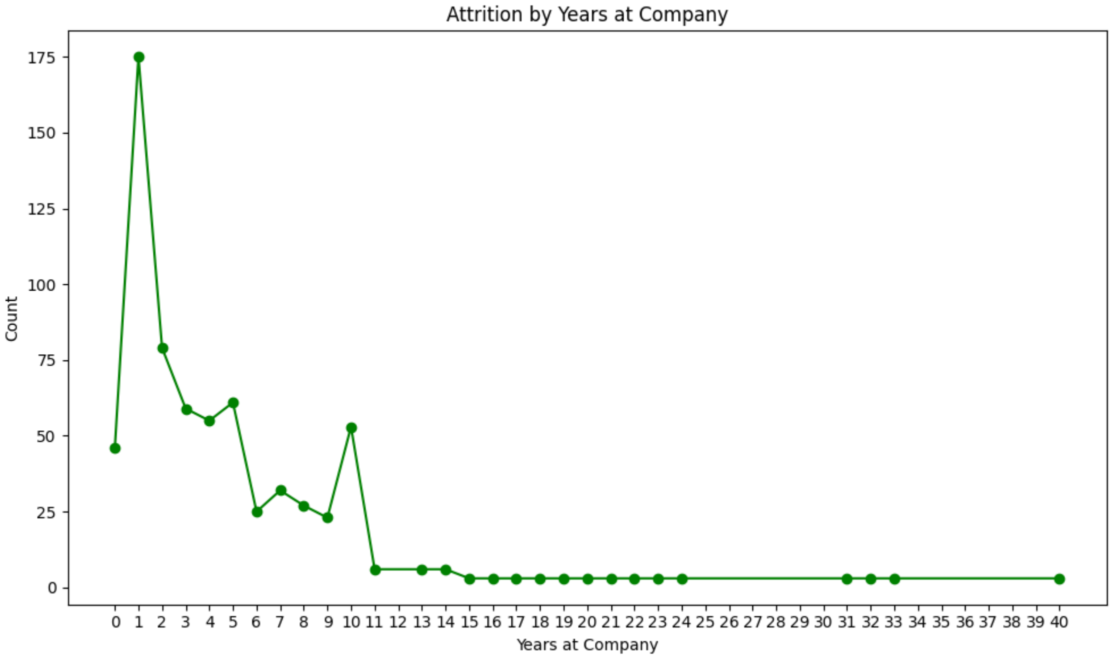


[20]:

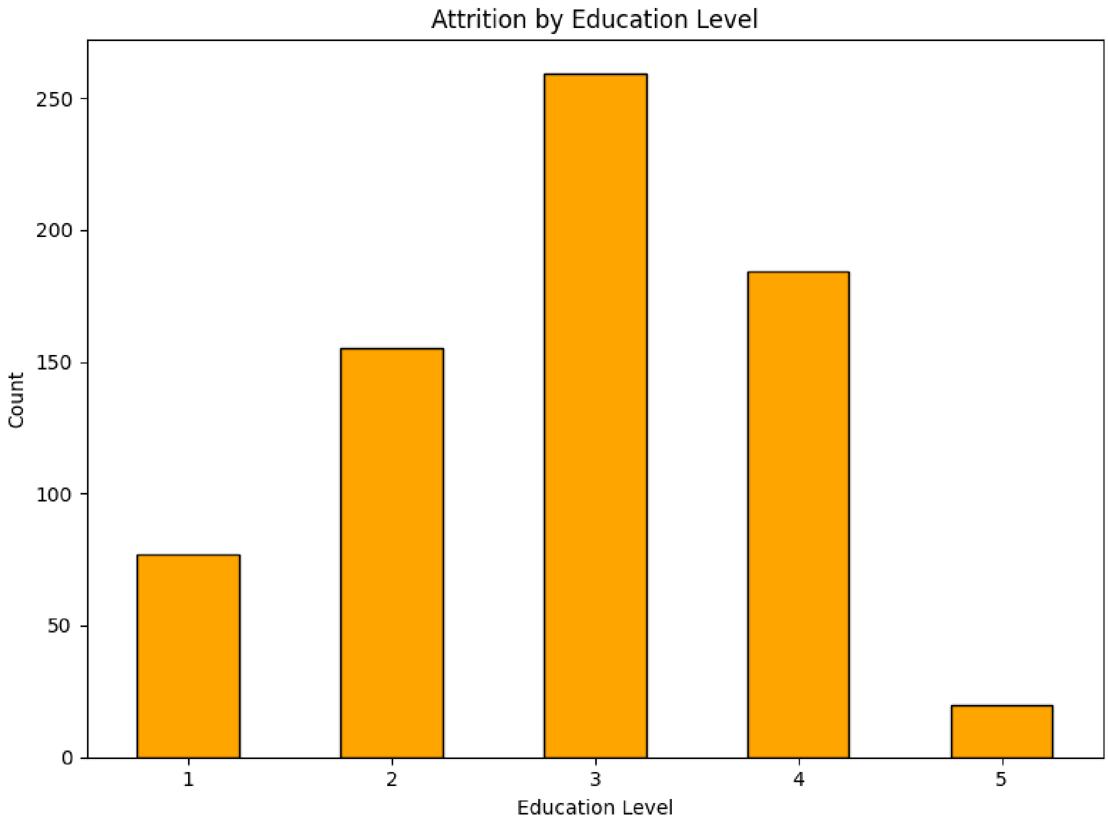
```
attrition_gender = attr[attr['Attrition'] == 'Yes']['Gender'].value_counts()

# Plot the bar chart
plt.figure(figsize=(8, 6))
attrition_gender.plot(kind='bar', color=['blue', 'pink'], edgecolor='black')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Attrition by Gender')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```



[21]:
```
attrition_years_company = attr[attr['Attrition'] == 'Yes'].
  ↪groupby('YearsAtCompany').size()
plt.figure(figsize=(10, 6))
attrition_years_company.plot(kind='line', marker='o', color='green')
plt.xlabel('Years at Company')
plt.ylabel('Count')
plt.title('Attrition by Years at Company')
plt.xticks(range(0, attr['YearsAtCompany'].max()+1, 1))
plt.tight_layout()
plt.show()
```



[22]:
```
attrition_education = attr[attr['Attrition'] == 'Yes']['Education'].
 ↪value_counts().sort_index()
plt.figure(figsize=(8, 6))
attrition_education.plot(kind='bar', color='orange', edgecolor='black')
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.title('Attrition by Education Level')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```



[23]:

```
# Calculate the overall attrition rate
total_employees = attr.shape[0]
total_attrition = attr[attr['Attrition'] == 'Yes'].shape[0]
attrition_rate = (total_attrition / total_employees) * 100

print(f'Overall Attrition Rate: {attrition_rate:.2f}%')
```

```
Overall Attrition Rate: 16.16%
```

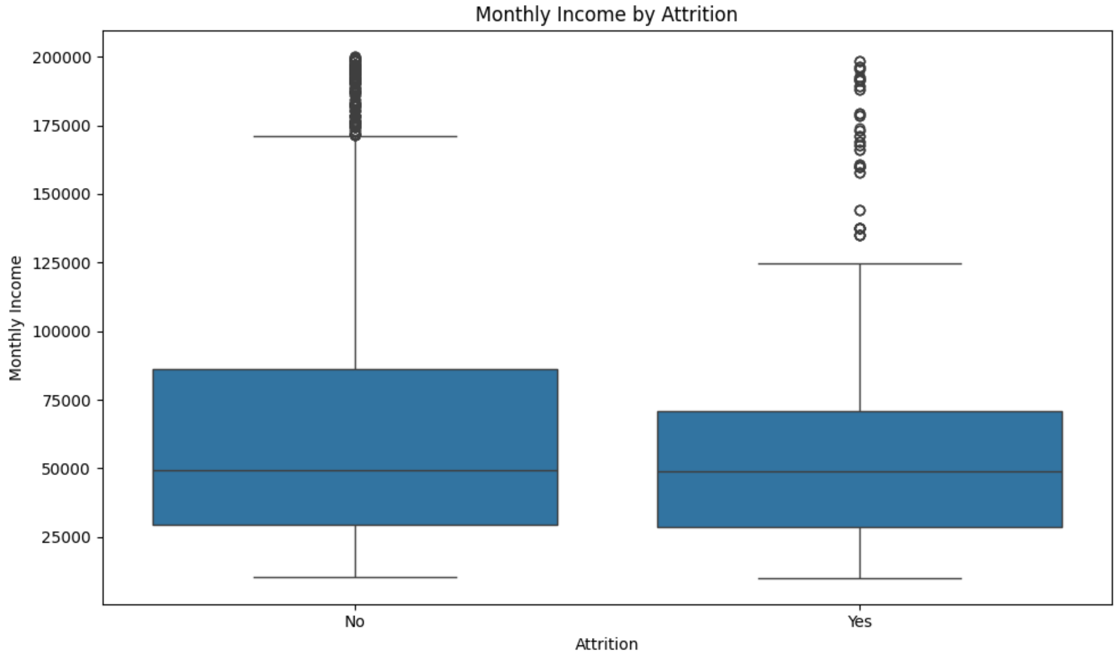
[24]:

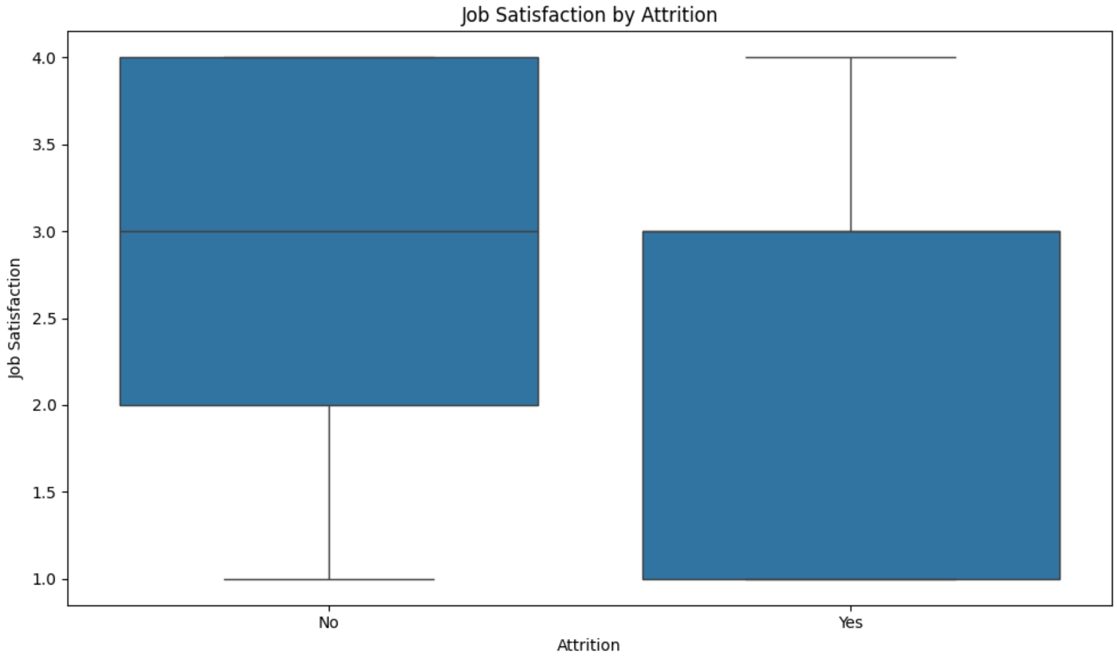
```
# Attrition by Department
attrition_by_department = attr[attr['Attrition'] == 'Yes']['Department'].
 ↪value_counts() / attr['Department'].value_counts() * 100

# Attrition by Job Role
attrition_by_job_role = attr[attr['Attrition'] == 'Yes']['JobRole'].
 ↪value_counts() / attr['JobRole'].value_counts() * 100

print('Attrition Rate by Department:')
print(attrition_by_department)

print('\nAttrition Rate by Job Role:')
print(attrition_by_job_role)
```

```
Attrition Rate by Department:
Department
Research & Development    15.781974
Sales                     15.149197
Human Resources           29.032258
Name: count, dtype: float64

Attrition Rate by Job Role:
JobRole
Healthcare Representative    14.588859
Human Resources              13.636364
Laboratory Technician        16.116248
Manager                      13.043478
Manufacturing Director       11.374408
Research Director            22.978723
Research Scientist           18.393481
Sales Executive              16.945607
Sales Representative         14.937759
Name: count, dtype: float64
```

[25]:

```
# Prepare data for the box plot
plt.figure(figsize=(10, 6))
sns.boxplot(x='Attrition', y='MonthlyIncome', data=attr)
plt.xlabel('Attrition')
plt.ylabel('Monthly Income')
plt.title('Monthly Income by Attrition')
plt.tight_layout()
plt.show()

plt.figure(figsize=(10, 6))
sns.boxplot(x='Attrition', y='JobSatisfaction', data=attr)
plt.xlabel('Attrition')
plt.ylabel('Job Satisfaction')
plt.title('Job Satisfaction by Attrition')
plt.tight_layout()
plt.show()
```

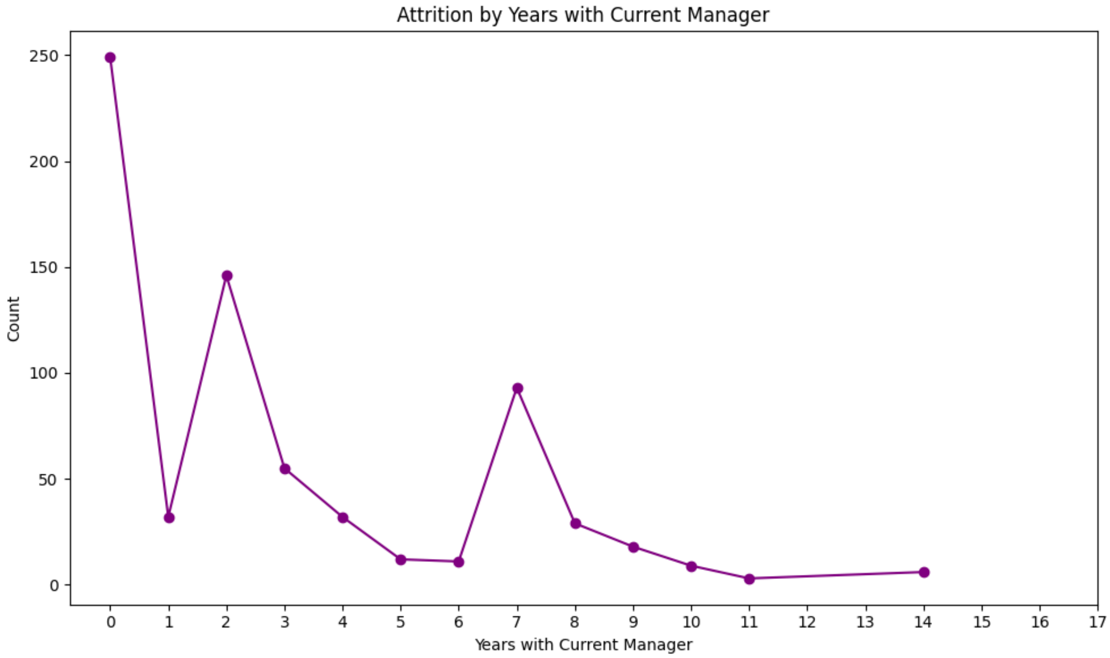




[26]:

```
# Prepare data for the line plot
attrition_years_manager = attr[attr['Attrition'] == 'Yes'].
 ↪groupby('YearsWithCurrManager').size()
```

```
plt.figure(figsize=(10, 6))
attrition_years_manager.plot(kind='line', marker='o', color='purple')
plt.xlabel('Years with Current Manager')
plt.ylabel('Count')
plt.title('Attrition by Years with Current Manager')
plt.xticks(range(0, attr['YearsWithCurrManager'].max()+1, 1))
plt.tight_layout()
plt.show()
```



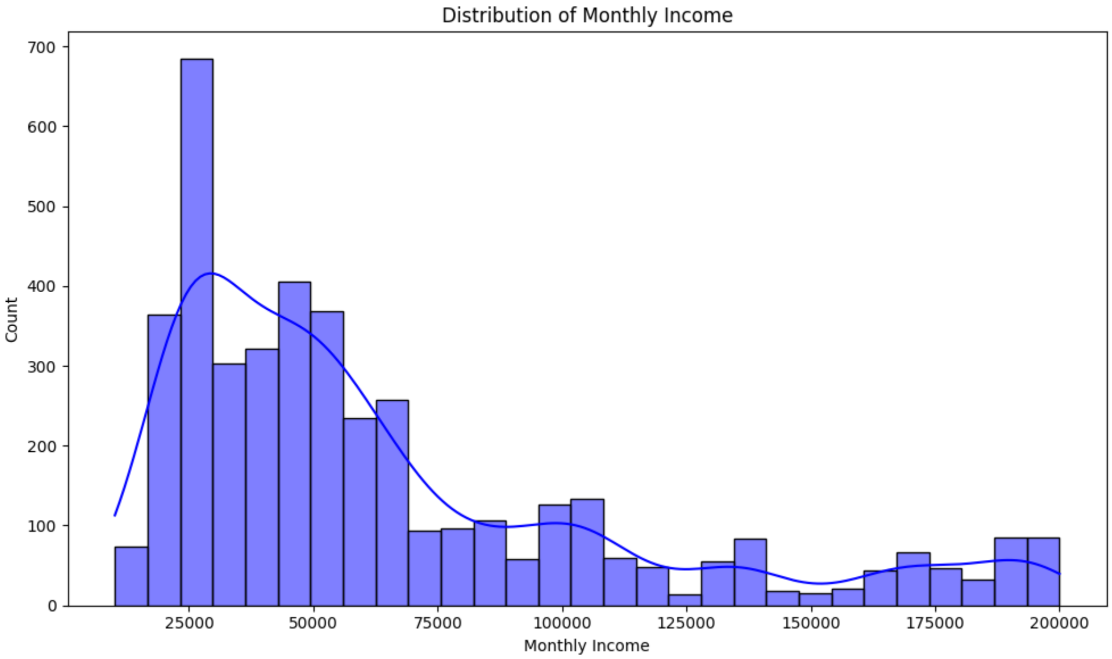
[27]:
```
monthly_income_stats = attr['MonthlyIncome'].describe()
print(monthly_income_stats)
```

```
count      4300.000000
mean      65059.844186
std       47045.398914
min       10090.000000
25%       29260.000000
50%       49360.000000
75%       83802.500000
max      199990.000000
Name: MonthlyIncome, dtype: float64
```

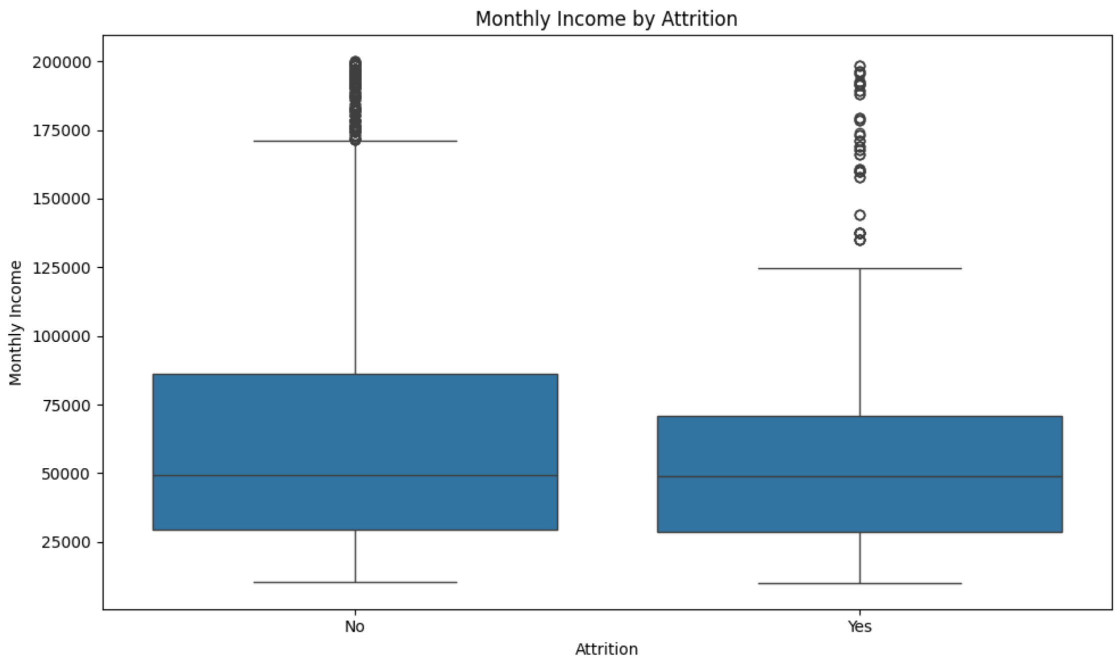
[28]:
```
attr["MonthlyIncome"].nunique()
```

[28]: 1349

[29]:
```
plt.figure(figsize=(10, 6))
sns.histplot(attr['MonthlyIncome'], kde=True, color='blue')
plt.xlabel('Monthly Income')
plt.ylabel('Count')
plt.title('Distribution of Monthly Income')
plt.tight_layout()
plt.show()
```

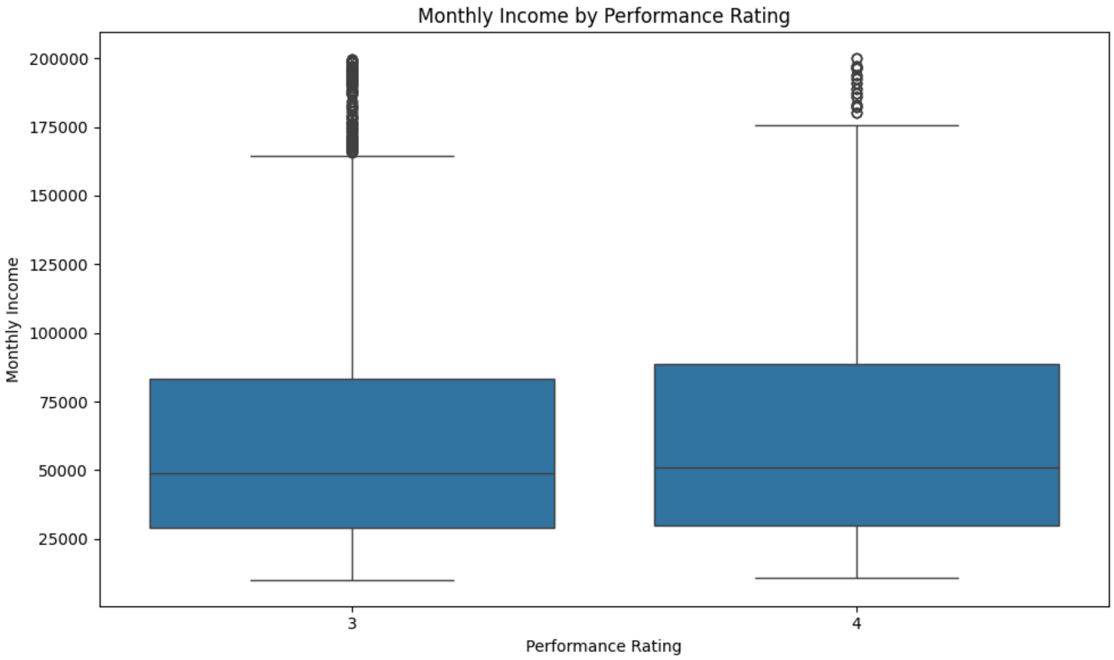


[30]:
```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Attrition', y='MonthlyIncome', data=attr)
plt.xlabel('Attrition')
plt.ylabel('Monthly Income')
plt.title('Monthly Income by Attrition')
plt.tight_layout()
plt.show()
```



[31]:

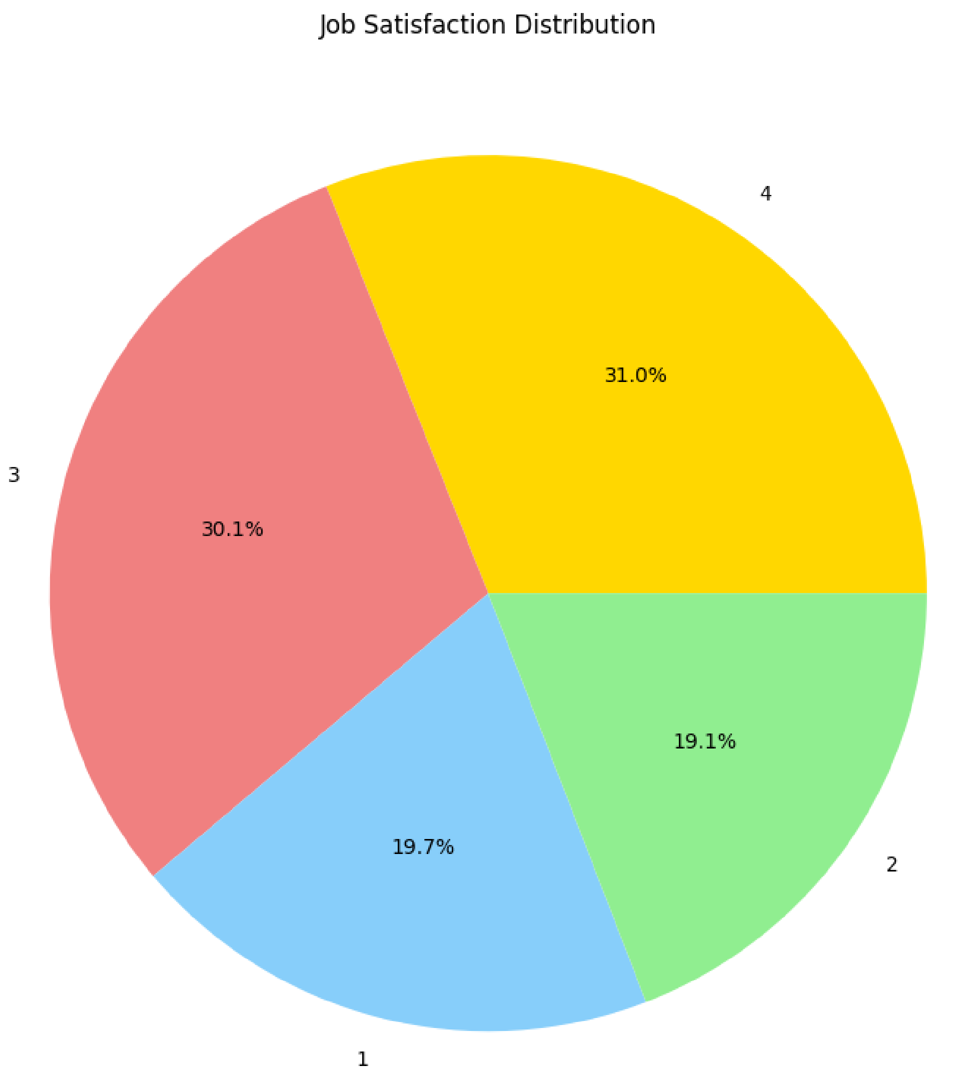
```
plt.figure(figsize=(10, 6))
sns.boxplot(x='PerformanceRating', y='MonthlyIncome', data=attr)
plt.xlabel('Performance Rating')
plt.ylabel('Monthly Income')
plt.title('Monthly Income by Performance Rating')
plt.tight_layout()
plt.show()
```



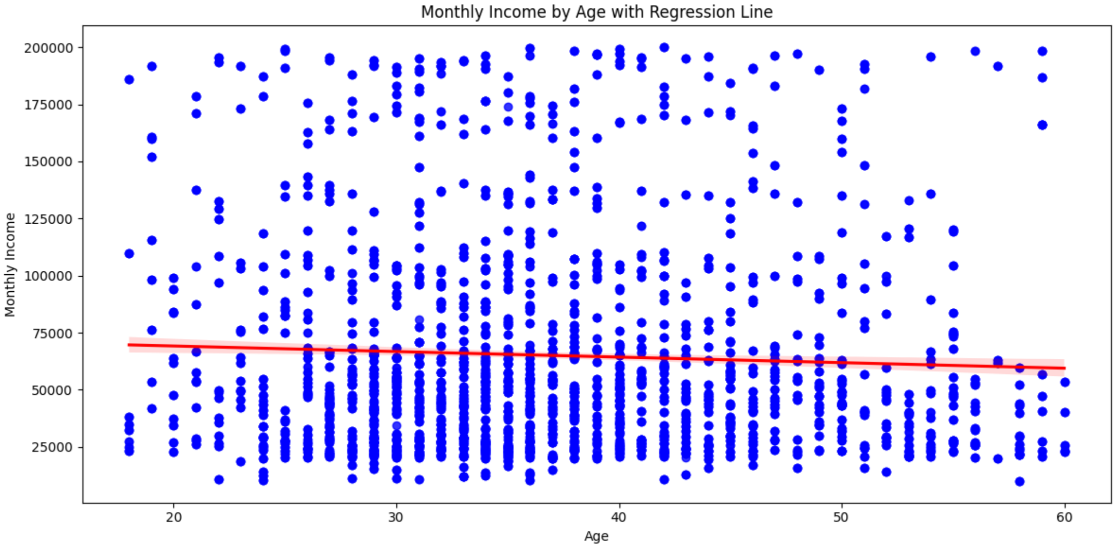
```
[32]: attr["JobSatisfaction"].value_counts()
```

```
[32]: JobSatisfaction
      4.0    1334
      3.0    1296
      1.0     847
      2.0     823
      Name: count, dtype: int64
```

```
[33]: labels = ['4', '3', '1', '2']

      colors = ['gold', 'lightcoral', 'lightskyblue', 'lightgreen']

      plt.figure(figsize=(8, 8))
      plt.pie(attr['JobSatisfaction'].value_counts(), labels=labels, autopct='%.
        ↪1f%%', colors=colors)
      plt.title('Job Satisfaction Distribution')
      plt.tight_layout()
      plt.show()
```



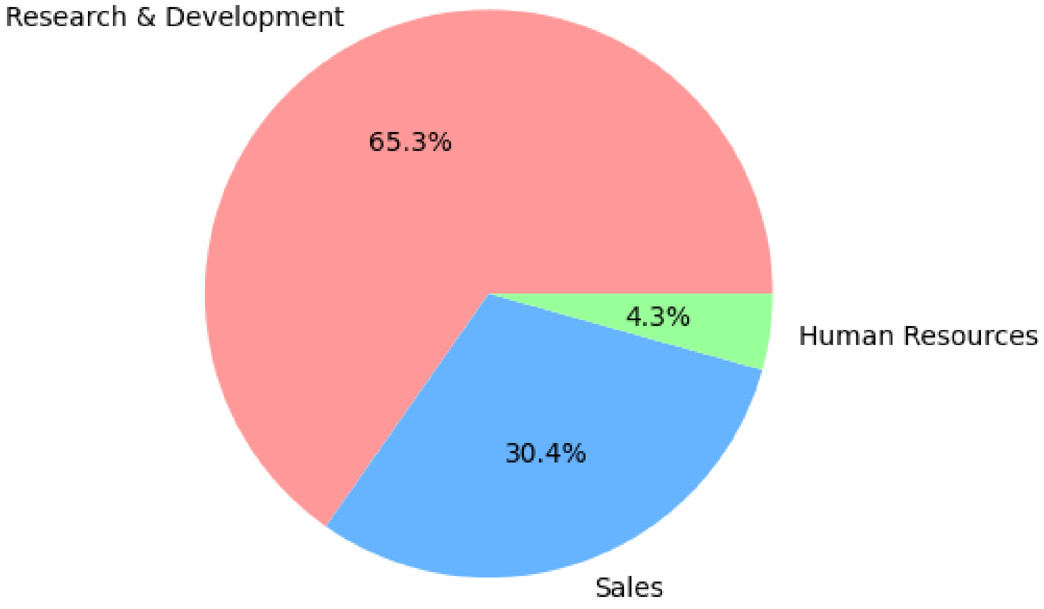
```
[34]: # Regression plot of Monthly Income by Age
      plt.figure(figsize=(12, 6))
      sns.regplot(x='Age', y='MonthlyIncome', data=attr, scatter_kws={'color':
        ↪'blue'}, line_kws={'color':'red'})
      plt.xlabel('Age')
      plt.ylabel('Monthly Income')
      plt.title('Monthly Income by Age with Regression Line')
      plt.tight_layout()
      plt.show()
```



[35]:
```
attr["Department"].value_counts()
```

[35]:
```
Department
Research & Development    2807
Sales                     1307
Human Resources            186
Name: count, dtype: int64
```

[36]:
```
labels = ['Research & Development', 'Sales', 'Human Resources']
colors = ['#ff9999','#66b3ff','#99ff99']
plt.pie(attr.Department.value_counts(), labels=labels, autopct='%.1f%%',␣
  ↪colors=colors)
plt.show()
```



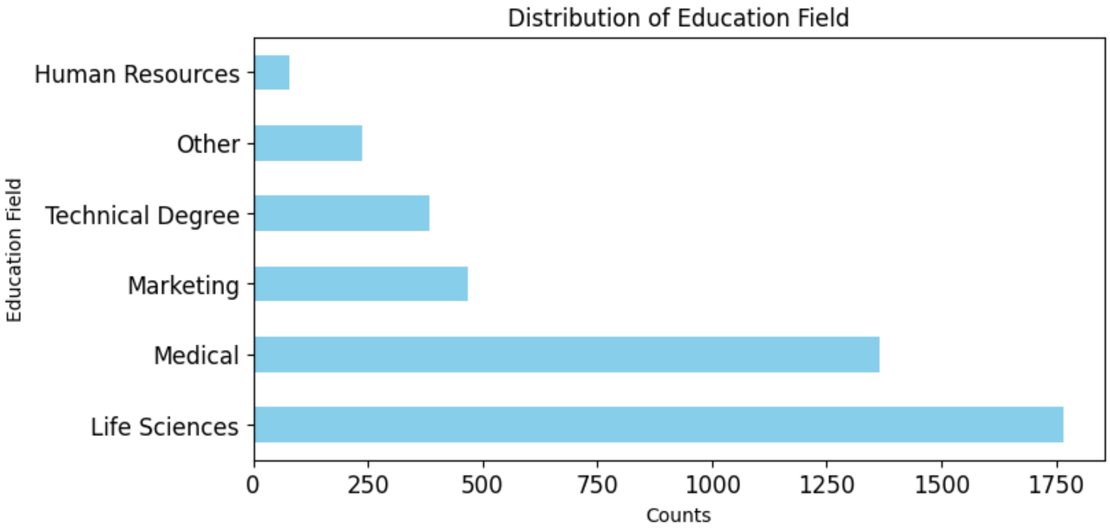
[37]:
```
attr["EducationField"].nunique()
```

[37]:
```
6
```

[38]:
```
attr["EducationField"].value_counts()
```

[38]:
```
EducationField
Life Sciences       1766
Medical             1364
Marketing            469
Technical Degree     384
Other                237
Human Resources       80
Name: count, dtype: int64
```

[39]:
```
attr["EducationField"].value_counts().plot(kind="barh", figsize=(8, 4),␣
  ↪color='skyblue', fontsize=12)

plt.xlabel('Counts')
plt.ylabel('Education Field')
plt.title('Distribution of Education Field')
plt.show()
```



[40]:

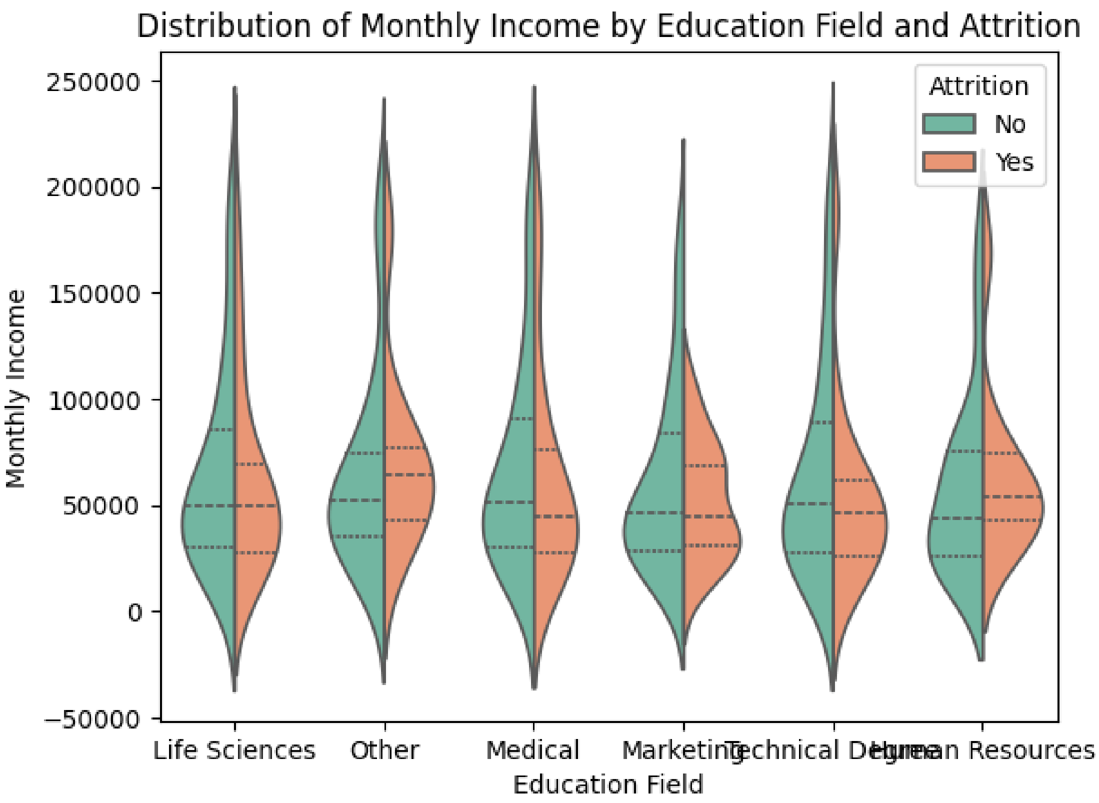
```
attr.groupby('EducationField').MonthlyIncome.mean()
```

[40]:

```
EducationField
Human Resources     60523.125000
Life Sciences       65392.729332
Marketing           59302.857143
Medical             66973.137830
Other               66665.991561
Technical Degree    63717.916667
Name: MonthlyIncome, dtype: float64
```

[41]:

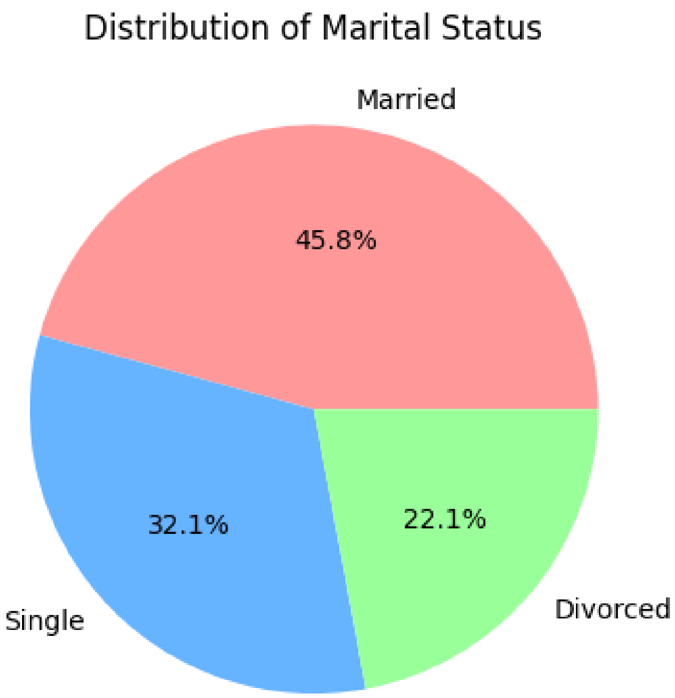
```
sns.violinplot(
    x="EducationField",
    y="MonthlyIncome",
    hue="Attrition",
    data=attr,
    split=True,
    palette="Set2",
    bw=0.5,
    inner="quartile"
)

plt.title('Distribution of Monthly Income by Education Field and Attrition')
plt.xlabel('Education Field')
plt.ylabel('Monthly Income')
plt.show()
```



[42]:
```
attr["MaritalStatus"].value_counts()
```

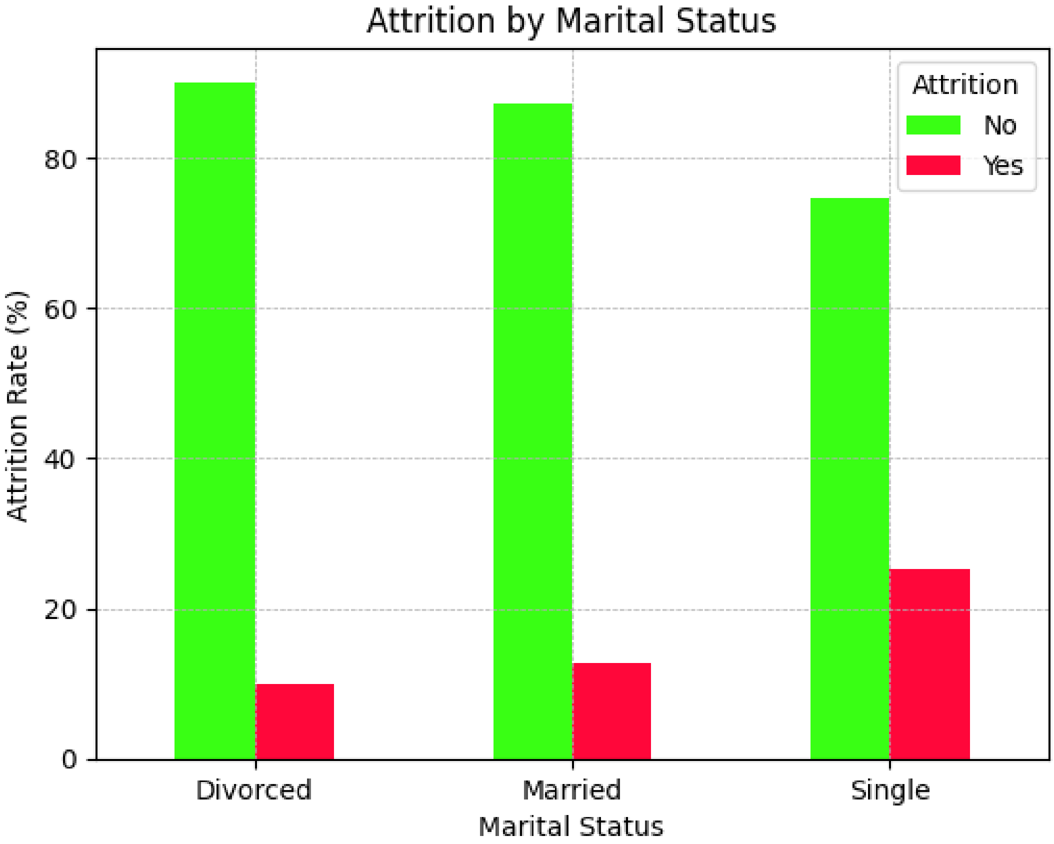
[42]:
```
MaritalStatus
Married     1969
Single      1382
Divorced     949
Name: count, dtype: int64
```

[43]:
```
labels = ['Married', 'Single', 'Divorced']
colors = ['#ff9999', '#66b3ff', '#99ff99']
plt.pie(attr.MaritalStatus.value_counts(), labels=labels, autopct='%.1f%%',␣
 ↪colors=colors)
plt.title('Distribution of Marital Status')
plt.show()
```



[44]:

```
tab = pd.crosstab(attr["MaritalStatus"], attr["Attrition"], normalize='index') * 100
neon_colors = ["#39FF14", "#FF073A", "#FFD700", "#00FFFF", "#FF00FF"]
# Plot the bar chart
tab.plot(kind='bar', color=neon_colors)
plt.title('Attrition by Marital Status')
plt.xlabel('Marital Status')
plt.ylabel('Attrition Rate (%)')
plt.xticks(rotation=0)
plt.legend(title='Attrition')
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.show()
```


Attrition by Marital Status
Attrition
🟩 No
🟥 Yes
Attrition Rate (%)
80
60
40
20
0
Divorced
Married
Single
Marital Status