```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")

from google.colab import files
uploaded = files.upload()
```

```
Choose Files WA_Fn-Us...-Attrition.csv
• WA_Fn-UseC_-HR-Employee-Attrition.csv(text/csv) - 227977 bytes, last modified: 7/11/2025 - 100% done
Saving WA_Fn-UseC_-HR-Employee-Attrition.csv to WA_Fn-UseC_-HR-Employee-Attrition.csv
```

```
df=pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv')
print(df.head)
```

```
<bound method NDFrame.head of       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
...   ...       ...                ...        ...                     ...
1465   36        No  Travel_Frequently        884  Research & Development
1466   39        No      Travel_Rarely        613  Research & Development
1467   27        No      Travel_Rarely        155  Research & Development
1468   49        No  Travel_Frequently       1023                   Sales
1469   34        No      Travel_Rarely        628  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount  \
0                    1          2  Life Sciences              1
1                    8          1  Life Sciences              1
2                    2          2          Other              1
3                    3          4  Life Sciences              1
4                    2          1        Medical              1
...                ...        ...            ...            ...
1465                23          2        Medical              1
1466                 6          1        Medical              1
1467                 4          3  Life Sciences              1
1468                 2          3        Medical              1
1469                 8          3        Medical              1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
0                  1  ...                         1            80
1                  2  ...                         4            80
2                  4  ...                         2            80
3                  5  ...                         3            80
4                  7  ...                         4            80
...              ...  ...                       ...           ...
1465            2061  ...                         3            80
1466            2062  ...                         1            80
1467            2064  ...                         2            80
1468            2065  ...                         4            80
1469            2068  ...                         1            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                    0                  8                      0
1                    1                 10                      3
2                    0                  7                      3
3                    0                  8                      3
4                    1                  6                      3
...                ...                ...                    ...
1465                 1                 17                      3
1466                 1                  9                      5
1467                 1                  6                      0
1468                 0                 17                      3
1469                 0                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0                  1               6                  4
1                  3              10                  7
2                  3               0                  0
3                  3               8                  7
4                  3               2                  2
```

```
# Check for nulls
df.isnull().sum()
```

```
# Convert Attrition to 0/1
df['AttritionFlag'] = df['Attrition'].map({'Yes': 1, 'No': 0})
```

```
# Add Age Groups
```

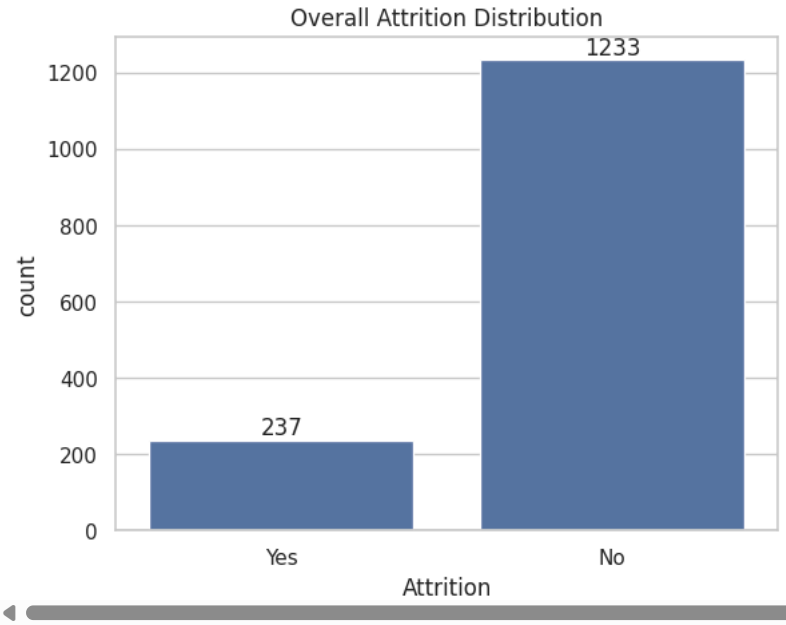
```
df['AgeGroup'] = pd.cut(df['Age'], bins=[17, 30, 40, 50, 60],
                        labels=['18-30', '31-40', '41-50', '51-60'])

# Add Income Levels
df['IncomeLevel'] = pd.cut(df['MonthlyIncome'],
                           bins=[0, 3000, 6000, 9000, 15000, df['MonthlyIncome'].max()],
                           labels=['Very Low', 'Low', 'Medium', 'High', 'Very High'])
```

```
df.head(10)
```

ationField	EmployeeCount	EmployeeNumbe
.ife Sciences	1	
.ife Sciences	1	
Other	1	
.ife Sciences	1	
Medical	1	
.ife Sciences	1	
Medical	1	
.ife Sciences	1	
.ife Sciences	1	
Medical	1	

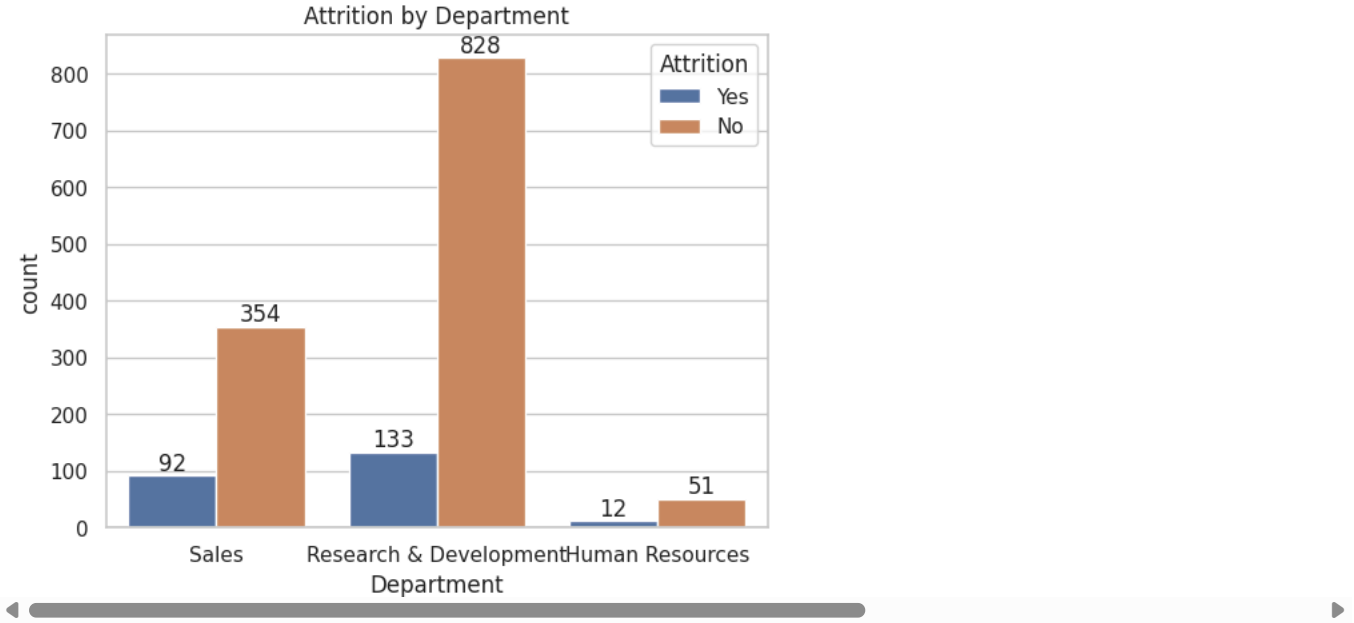
```
ax=sns.countplot(x='Attrition', data=df)
plt.title("Overall Attrition Distribution")
ax.bar_label(ax.containers[0])
```

```
[Text(0, 0, '237'), Text(0, 0, '1233')]
```



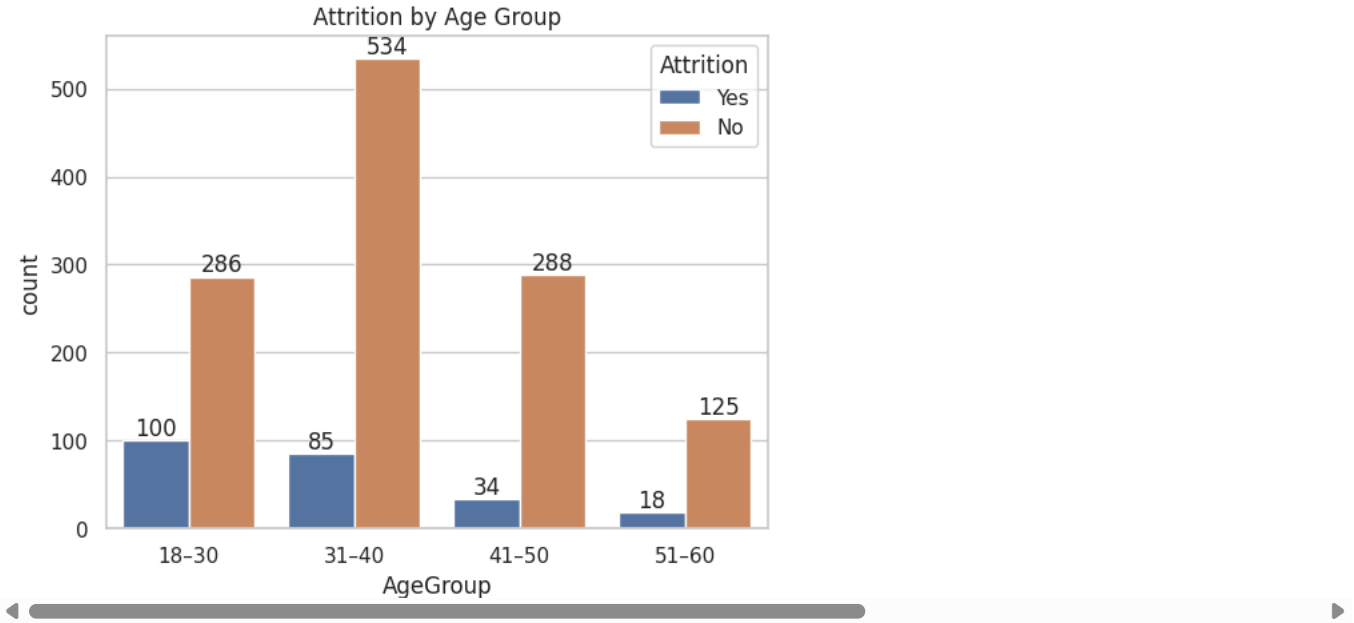

```
bx=sns.countplot(x='Department', hue='Attrition', data=df)
plt.title("Attrition by Department")
bx.bar_label(bx.containers[0])
bx.bar_label(bx.containers[1])
```

```
[Text(0, 0, '354'), Text(0, 0, '828'), Text(0, 0, '51')]
```



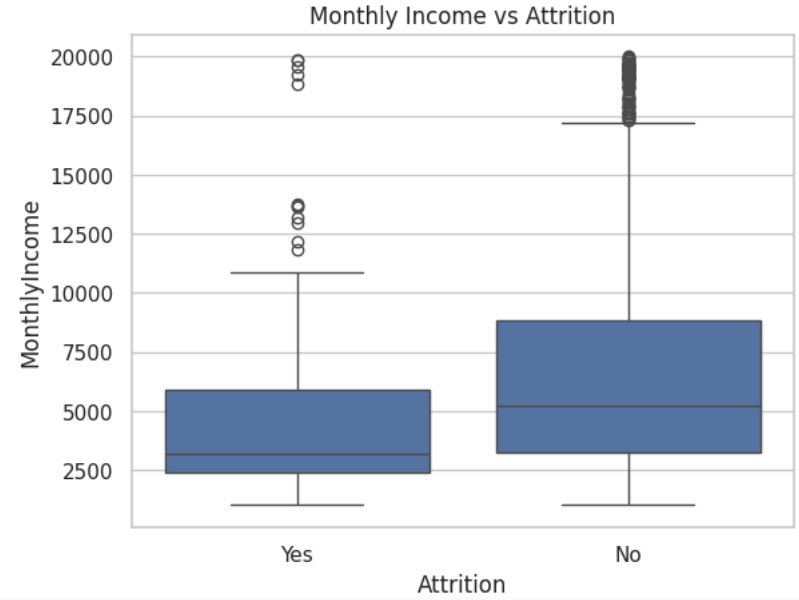

```
cx=sns.countplot(x='AgeGroup', hue='Attrition', data=df)
plt.title("Attrition by Age Group")
cx.bar_label(cx.containers[0])
cx.bar_label(cx.containers[1])
```

```
[Text(0, 0, '286'), Text(0, 0, '534'), Text(0, 0, '288'), Text(0, 0, '125')]
```




```
dx=sns.boxplot(x='Attrition', y='MonthlyIncome', data=df)
plt.title("Monthly Income vs Attrition")
```

```
Text(0.5, 1.0, 'Monthly Income vs Attrition')
```




```
df.to_csv("Cleaned_Attrition_Data.csv", index=False)
files.download("Cleaned_Attrition_Data.csv")
```