eda

August 22, 2025

[2]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[8]:
```
df = pd.read_csv("C:\\Users\\HP\\Desktop\\WA_Fn-UseC_-HR-Employee-Attrition.
↪csv")
```

[9]:
```
# Basic info
print(df.head())
print(df['Attrition'].value_counts())
```

```
   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
```

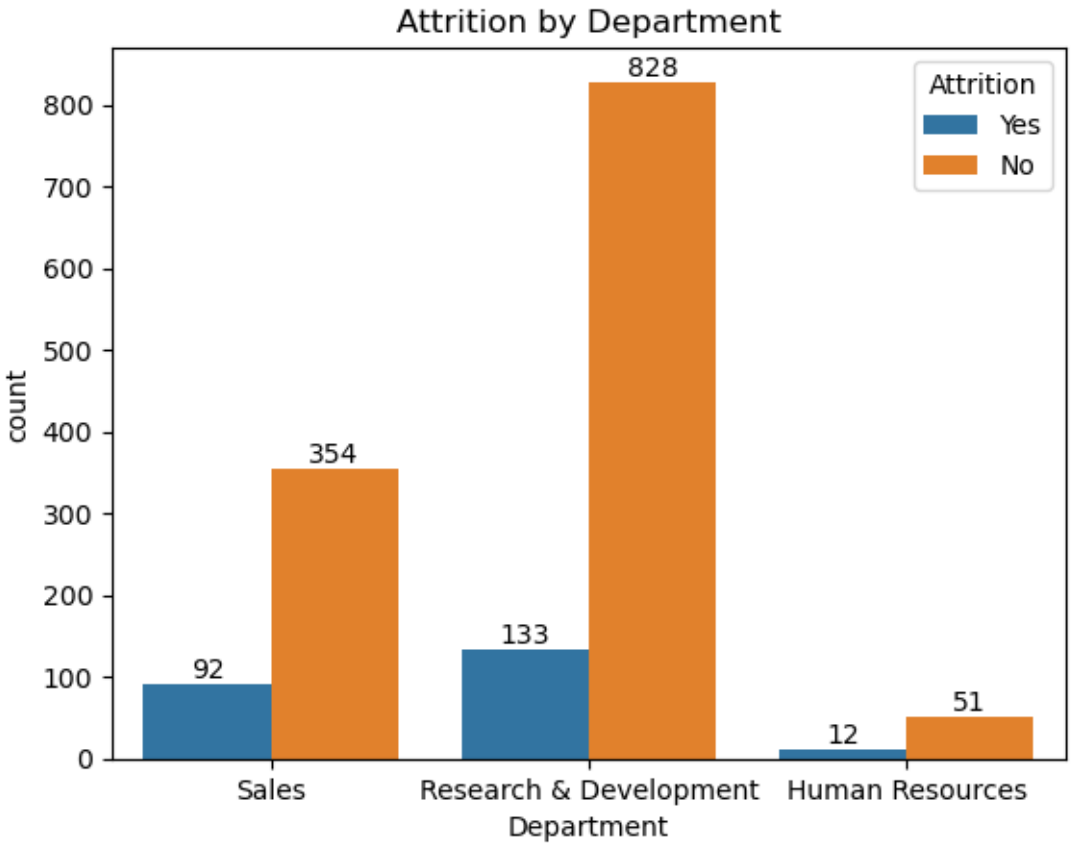
```
  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7    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

[5 rows x 35 columns]
Attrition
No     1233
Yes     237
Name: count, dtype: int64
```

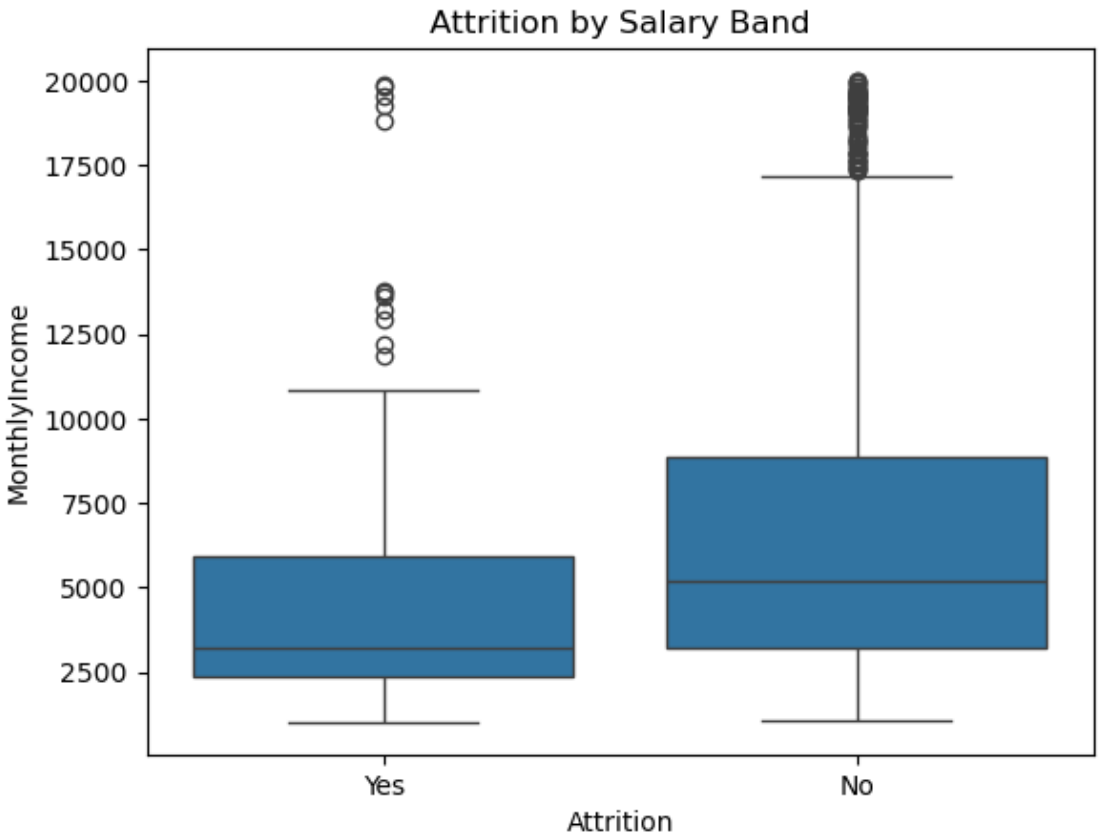
[14]:
```
# Department-wise attrition with value labels
ax = sns.countplot(x='Department', hue='Attrition', data=df)
plt.title("Attrition by Department")

# Add values on each bar
for container in ax.containers:
    ax.bar_label(container, fmt='%d')

plt.show()
```



[15]:
```
# Salary bands vs attrition
sns.boxplot(x='Attrition', y='MonthlyIncome', data=df)
plt.title("Attrition by Salary Band")
plt.show()
```



[ ]: