


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

 Choose Files


No file chosen

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
import pandas as pd

# Paste the correct file name below
df = pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")

# Now continue the same
print(df.head())
```




	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	8	0	1	6	4	0	5
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	10	3	3	10	7	1	7
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	7	3	3	0	0	0	0
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	8	3	3	8	7	3	0
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	6	3	3	2	2	2	2

[5 rows x 35 columns]

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

# Load the dataset (update filename if different)
df = pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
df.head()
```



	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	8	0	1	6	4	0	5
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	10	3	3	10	7	1	7
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	7	3	3	0	0	0	0
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	8	3	3	8	7	3	0
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	6	3	3	2	2	2	2

5 rows x 35 columns

```
df.shape

(1470, 35)

df.info()
```



```
<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
```

```
df.describe()
```



	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470.000000	1470.000000
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306	2.721769	65.891156	2.729932
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335	1.093082	20.329428	0.711561
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	1.000000	30.000000	1.000000
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000	2.000000	48.000000	2.000000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000	3.000000	66.000000	3.000000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000	4.000000	83.750000	3.000000
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000	4.000000	100.000000	4.000000

8 rows × 26 columns

```
df.columns
```



```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
      'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
      'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
      'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
      'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
      'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
      'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
      'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
      'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
      'YearsWithCurrManager'],
      dtype='object')
```

```
df['Attrition'].value_counts()
```



count

Attrition

No	1233
Yes	237

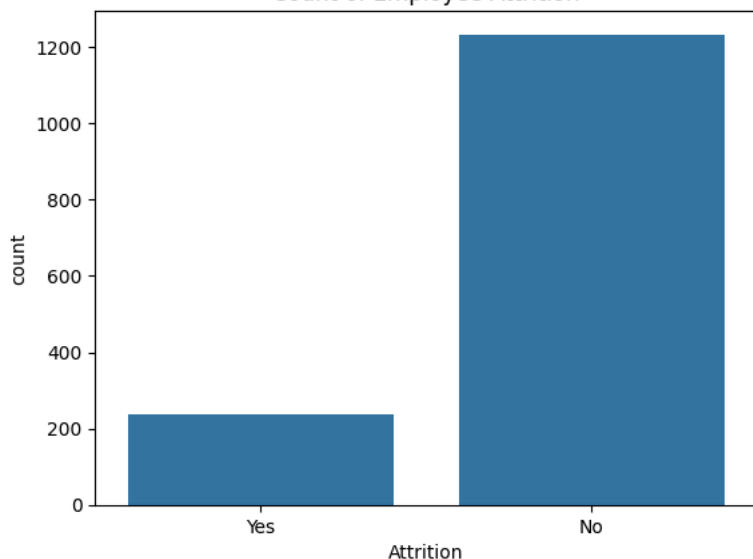
```
df.groupby('Attrition').count()
```

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
sns.countplot(x='Attrition', data=df)
plt.title('Count of Employee Attrition')
plt.savefig('attrition_overall.png')
plt.show()
```



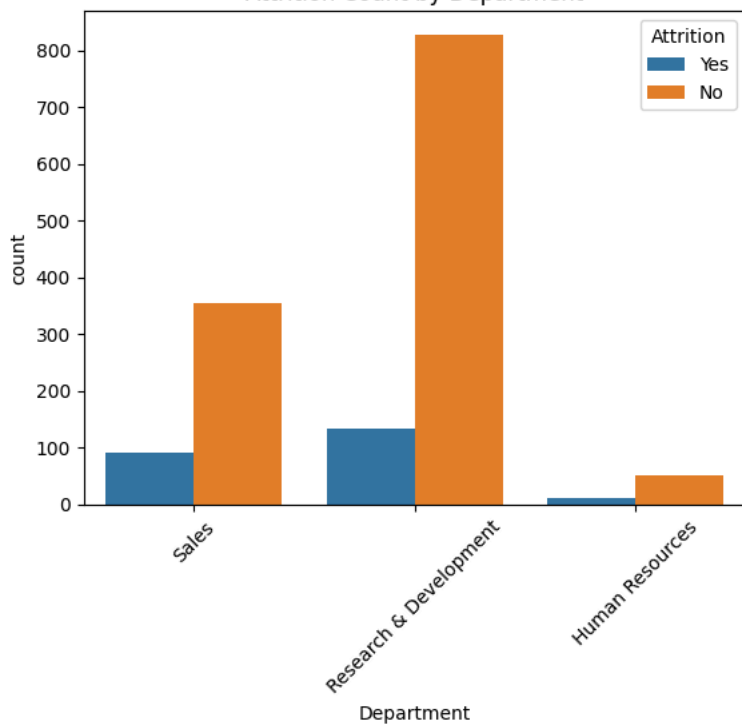
Count of Employee Attrition



```
sns.countplot(x='Department', hue='Attrition', data=df)
plt.title('Attrition Count by Department')
plt.xticks(rotation=45)
plt.savefig('attrition_by_department.png')
plt.show()
```

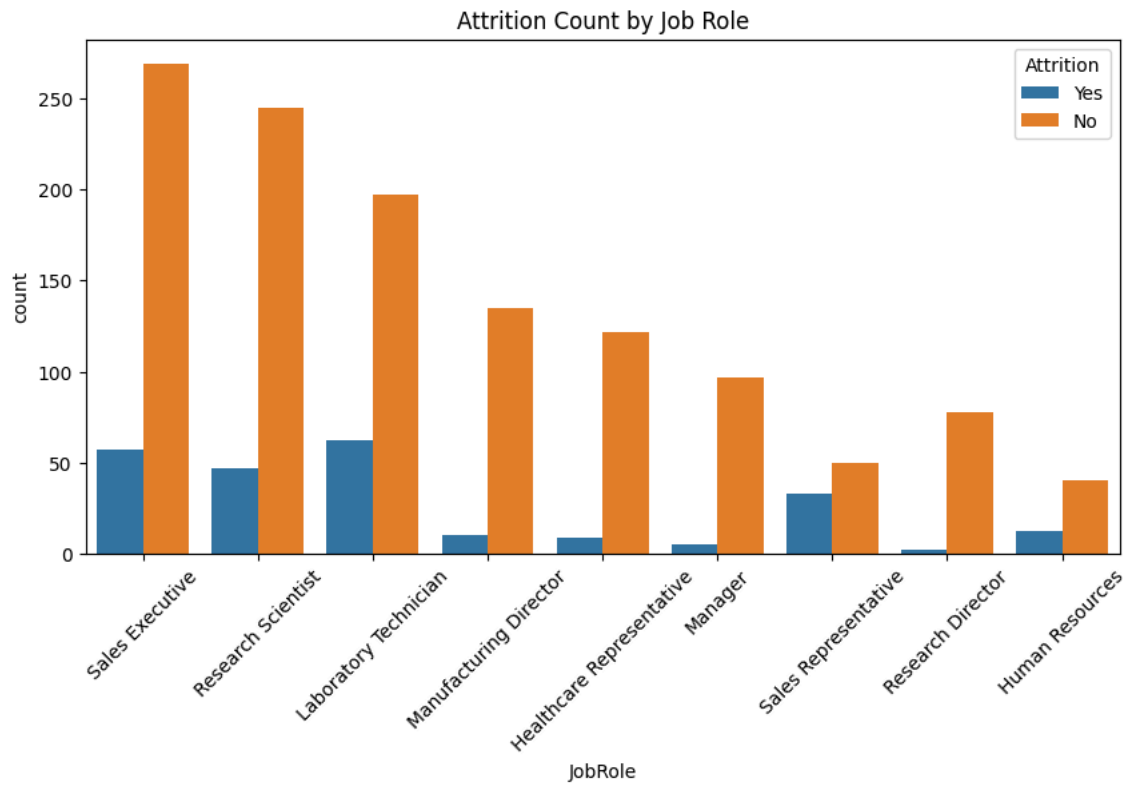


Attrition Count by Department



```
plt.figure(figsize=(10,5))
sns.countplot(x='JobRole', hue='Attrition', data=df)
plt.title('Attrition Count by Job Role')
plt.xticks(rotation=45)
```

```
plt.savefig('attrition_by_jobrole.png')
plt.show()
```



- Most attrition is observed in the Sales department.
- Job roles like Laboratory Technician and Sales Executive show higher attrition.
- The company has significantly more employees staying than leaving.