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

data=pd.read_csv('/content/Attrition data.csv')
data
```

```
{"type":"dataframe","variable_name":"data"}
```

```
data.info()
data.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 29 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   EmployeeID                4410 non-null   int64  
 1   Age                       4410 non-null   int64  
 2   Attrition                 4410 non-null   object 
 3   BusinessTravel            4410 non-null   object 
 4   Department                4410 non-null   object 
 5   DistanceFromHome          4410 non-null   int64  
 6   Education                 4410 non-null   int64  
 7   EducationField            4410 non-null   object 
 8   EmployeeCount             4410 non-null   int64  
 9   Gender                    4410 non-null   object 
 10  JobLevel                  4410 non-null   int64  
 11  JobRole                   4410 non-null   object 
 12  MaritalStatus             4410 non-null   object 
 13  MonthlyIncome             4410 non-null   int64  
 14  NumCompaniesWorked        4391 non-null   float64
 15  Over18                    4410 non-null   object 
 16  PercentSalaryHike         4410 non-null   int64  
 17  StandardHours             4410 non-null   int64  
 18  StockOptionLevel          4410 non-null   int64  
 19  TotalWorkingYears         4401 non-null   float64
 20  TrainingTimesLastYear     4410 non-null   int64  
 21  YearsAtCompany            4410 non-null   int64  
 22  YearsSinceLastPromotion   4410 non-null   int64  
 23  YearsWithCurrManager      4410 non-null   int64  
 24  EnvironmentSatisfaction   4385 non-null   float64
 25  JobSatisfaction           4390 non-null   float64
 26  WorkLifeBalance           4372 non-null   float64
 27  JobInvolvement            4410 non-null   int64  
 28  PerformanceRating         4410 non-null   int64  
dtypes: float64(5), int64(16), object(8)
memory usage: 999.3+ KB
```

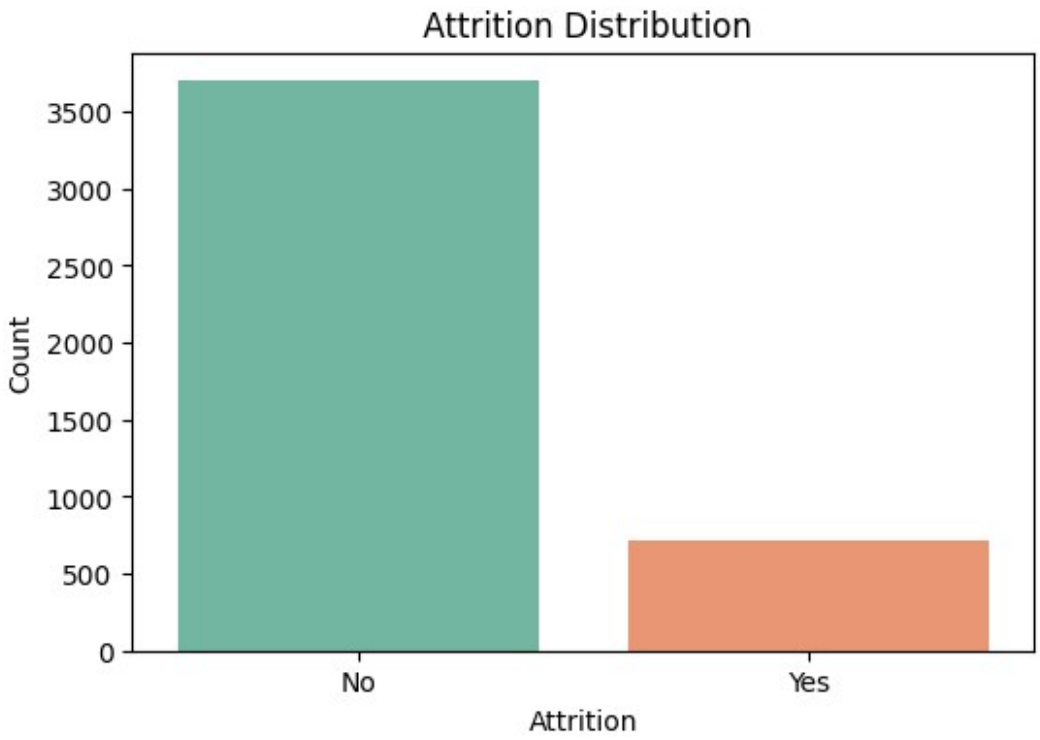
```
{"type":"dataframe"}
```

```
missed_value=data.isnull().sum()
missed_value
```

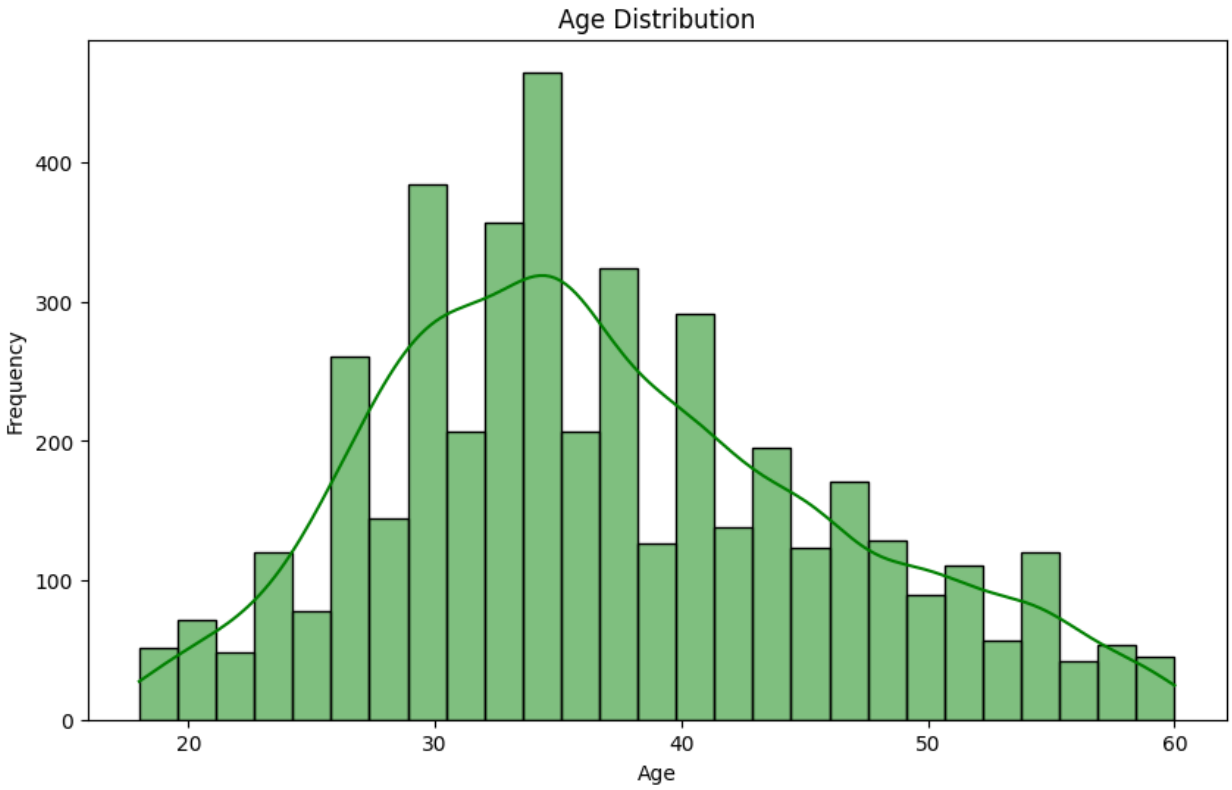
```
EmployeeID                 0
Age                        0
Attrition                  0
BusinessTravel             0
Department                 0
DistanceFromHome           0
Education                  0
EducationField             0
EmployeeCount              0
Gender                     0
JobLevel                   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
```

```
#Attrition Distribution
plt.figure(figsize=(6,4))
sns.countplot(x='Attrition', data=data, palette='Set2')
plt.title(' Attrition Distribution')
plt.xlabel('Attrition')
plt.ylabel('Count')
plt.show()
```

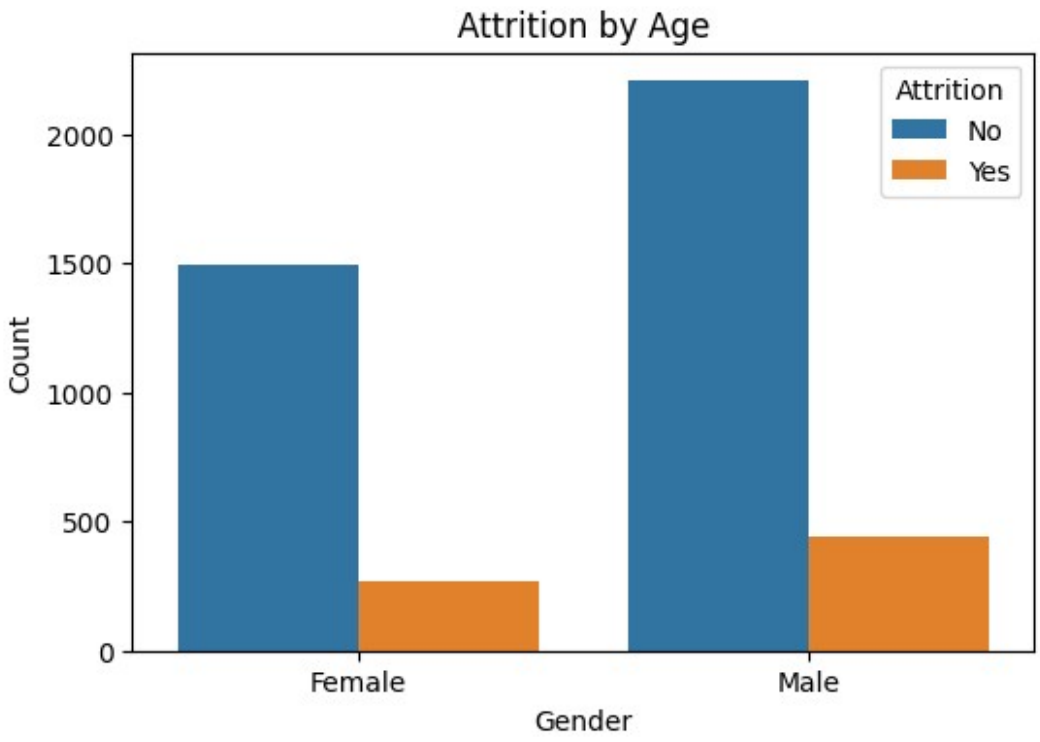
```
<ipython-input-5-7445962d7610>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.countplot(x='Attrition', data=data, palette='Set2')
```



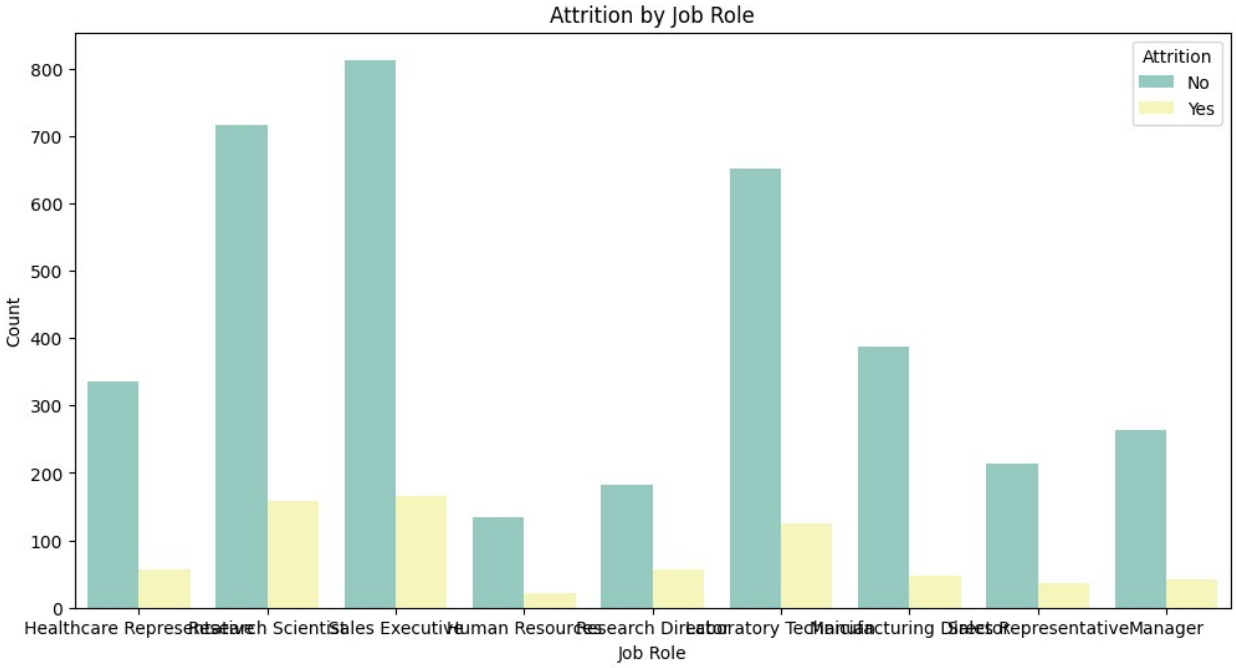

```
#Age Distribution
plt.figure(figsize=(10,6))
sns.histplot(data['Age'],kde=True,color='green')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```




```
#Gender_attrition
plt.figure(figsize=(6,4))
sns.countplot(x='Gender',hue='Attrition',data=data)
plt.title('Attrition by Age')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```




```
# Job role_Attrition
plt.figure(figsize=(12, 6))
sns.countplot(x='JobRole', hue='Attrition', data=data, palette='Set3')
plt.title('Attrition by Job Role')
plt.xlabel('Job Role')
plt.ylabel('Count')
plt.show()
```


Attrition by Job Role
Attrition
No
Yes
Count
0
100
200
300
400
500
600
700
800
Healthcare Representative
Research Scientist
Sales Executive
Human Resources
Research Director
Laboratory Technician
Manufacturing Director
Sales Representative
Manager
Job Role