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

```
df=pd.read_csv("C:\Users\prajd\Downloads\HR_Fix-Good_HR-Employee-Attrition.csv")
df
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	Tra
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	8	0
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	10	3
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	7	3
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	8	3
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	6	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	3	80	1	17	3
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	1	80	1	9	3
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	2	80	1	6	3
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	4	80	0	17	3
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	1	80	0	6	3

1470 rows × 35 columns

```
df.shape
```

```
(1470, 35)
```

```
df.isnull().sum()
```

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

```
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
```

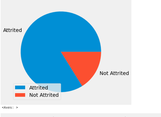
```
df.columns
```

```
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')
```

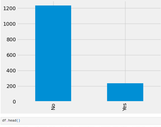
```
df.Attrition
```

```
0       Yes
1        No
2       Yes
3        No
4        No
       ...
1465     No
1466     No
1467     No
1468     No
1469     No
Name: Attrition, Length: 1470, dtype: object
```

```
Target=df.iloc[:,1].value_counts()
plt.pie(Target,labels=Target.index[:2],autopct='%1.0f%%')
plt.title('Attrition')
plt.legend(loc = 'lower left')
plt.show()
df['Attrition'].value_counts().plot.bar()
```




```
<Axes: >
```

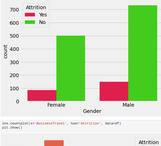



```
df.head()
```

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

5 rows × 35 columns

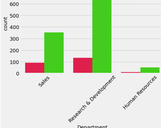
```
sns.countplot(x='Gender', hue='Attrition', data=df, palette='prism_r')
plt.show()
```



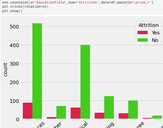

```
sns.countplot(x='BusinessTravel', hue='Attrition', data=df)
plt.show()
```

```
sns.countplot(x='MaritalStatus', hue='Attrition', data=df,palette='prism_r')
plt.show()
```

```
sns.countplot(x='Department', hue='Attrition',data=df,palette='prism_r')
plt.xticks(rotation=45)
plt.show()
```




```
sns.countplot(x='EducationField', hue='Attrition',data=df,palette='prism_r')
plt.xticks(rotation=45)
plt.show()
```




```
plt.figure(figsize=(8,4))
sns.countplot(x='DistanceFromHome', hue='Attrition', data=df,palette='prism_r')
```

```
<Axes: xlabel='DistanceFromHome', ylabel='count'>
```

```
sns.countplot(x='RelationshipSatisfaction', hue='Attrition',data=df,palette='prism_r')
```

```
<Axes: xlabel='RelationshipSatisfaction', ylabel='count'>
```

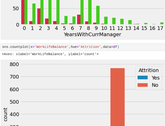
```
plt.figure(figsize=(8,4))
sns.countplot(x='YearsAtCompany', hue='Attrition',data=df,palette='prism_r')
plt.show()
```

```
sns.countplot(x='YearsSinceLastPromotion', hue='Attrition',data=df,palette='prism_r')
```

```
<Axes: xlabel='YearsSinceLastPromotion', ylabel='count'>
```

```
sns.countplot(x='YearsWithCurrManager', hue='Attrition',data=df,palette='prism_r')
```

```
<Axes: xlabel='YearsWithCurrManager', ylabel='count'>
```




```
sns.countplot(x='WorkLifeBalance', hue='Attrition',data=df)
```

```
<Axes: xlabel='WorkLifeBalance', ylabel='count'>
```