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

```
df = pd.read_csv('/content/Gr_4_HR-Employee-/
```

```
df.head()
```

	Age	Attrition	BusinessTravel	Dail
0	41	Yes	Travel_Rarely	
1	49	No	Travel_Frequently	
2	37	Yes	Travel_Rarely	
3	33	No	Travel_Frequently	
4	27	No	Travel_Rarely	

5 rows × 35 columns

```
df.size
```

```
51450
```

```
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
---  ------                    ---------
 0   Age                       1470 non-
 1   Attrition                 1470 non-
 2   BusinessTravel            1470 non-
 3   DailyRate                 1470 non-
 4   Department                1470 non-
 5   DistanceFromHome          1470 non-
```

Gr_4_HR-Employee-Attrition.csv ✕ •••

1 to 10 of 1470 entries Filter

gYears	TrainingTimesLastYear	WorkLifeBalance	Ye
	0	1	6
	3	3	10
	3	3	0
	3	3	8
	3	3	2
	2	2	7
	3	2	1
	2	3	1
	2	3	9
	3	2	7

Show 10 per page

1 2 10 100 140 147

```
  6   Education                 1470 non-
  7   EducationField            1470 non-
  8   EmployeeCount             1470 non-
  9   EmployeeNumber            1470 non-
  10  EnvironmentSatisfaction   1470 non-
  11  Gender                    1470 non-
  12  HourlyRate                1470 non-
  13  JobInvolvement            1470 non-
  14  JobLevel                  1470 non-
  15  JobRole                   1470 non-
  16  JobSatisfaction           1470 non-
  17  MaritalStatus             1470 non-
  18  MonthlyIncome             1470 non-
  19  MonthlyRate               1470 non-
  20  NumCompaniesWorked        1470 non-
  21  Over18                    1470 non-
  22  OverTime                  1470 non-
  23  PercentSalaryHike         1470 non-
  24  PerformanceRating         1470 non-
  25  RelationshipSatisfaction  1470 non-
  26  StandardHours             1470 non-
  27  StockOptionLevel          1470 non-
  28  TotalWorkingYears         1470 non-
  29  TrainingTimesLastYear     1470 non-
  30  WorkLifeBalance           1470 non-
  31  YearsAtCompany            1470 non-
  32  YearsInCurrentRole        1470 non-
  33  YearsSinceLastPromotion   1470 non-
  34  YearsWithCurrManager      1470 non-
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

```
df.describe()
```

	Age	DailyRate	Distanc
count	1470.000000	1470.000000	14
mean	36.923810	802.485714	
std	9.135373	403.509100	
min	18.000000	102.000000	
25%	30.000000	465.000000	
50%	36.000000	802.000000	
75%	43.000000	1157.000000	
max	60.000000	1499.000000	

8 rows × 26 columns

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

	0
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
```

There are no Null value

```
df.duplicated().sum()
```

```
0
```

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

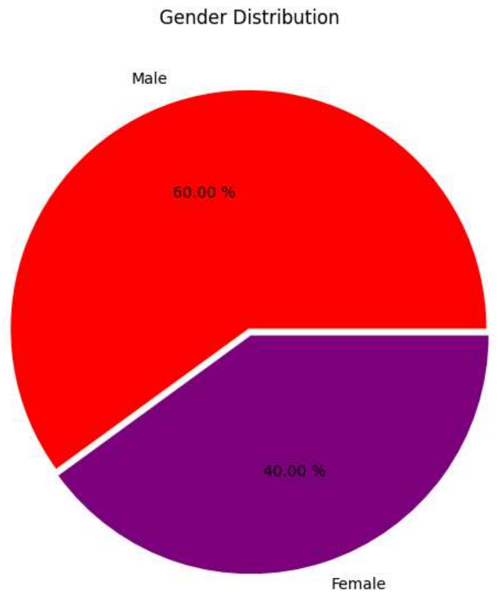
```
           count
Gender
  Male       882
Female       588

dtype: int64
```

```
labels=list(df.Gender.value_counts().to_dict
Male= list(df.Gender.value_counts().to_dict(
Female= list(df.Gender.value_counts().to_dic
```

```
labels
```

```
['Male', 'Female']
```

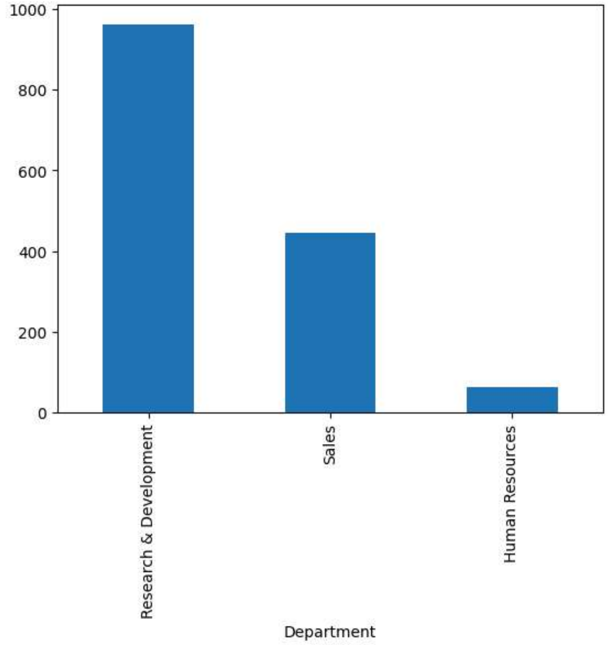
```
plt.figure(figsize=(7,7))
plt.title("Gender Distribution")
plt.pie([Male,Female],
        labels = labels,
        autopct='%.2f %%',
        explode = [0.01, 0.02],
        colors = ['red','purple']);
```



```
df.Department.value_counts()
```

```
                           count
            Department
Research & Development       961
                 Sales       446
       Human Resources        63

dtype: int64
```

```
df.Department.value_counts().plot(kind='bar'
<Axes: xlabel='Department'>
```



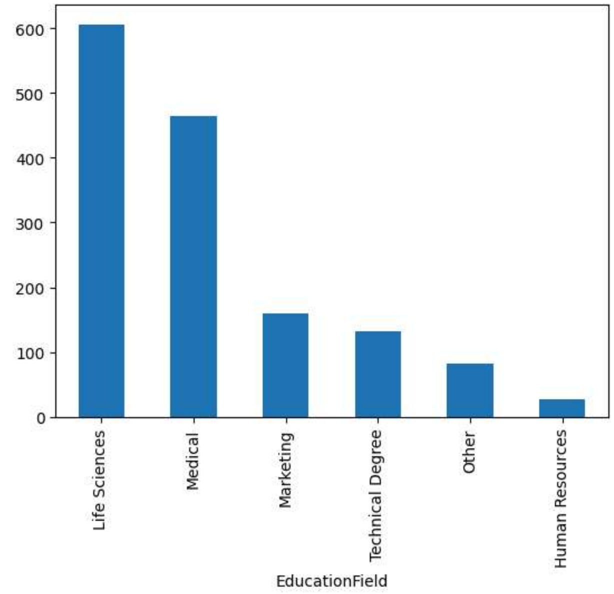
```
df.EducationField.value_counts()
```

EducationField	count
Life Sciences	606
Medical	464
Marketing	159
Technical Degree	132
Other	82
Human Resources	27

dtype: int64

```
df.EducationField.value_counts().plot(kind='
```

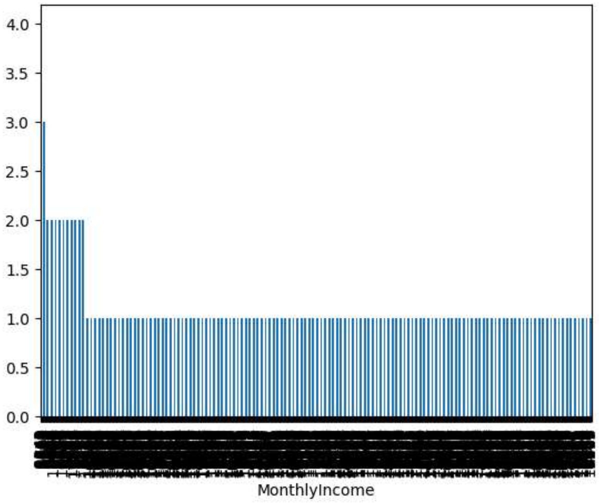
```
<Axes: xlabel='EducationField'>
```



```
df.MonthlyIncome.value_counts()
```

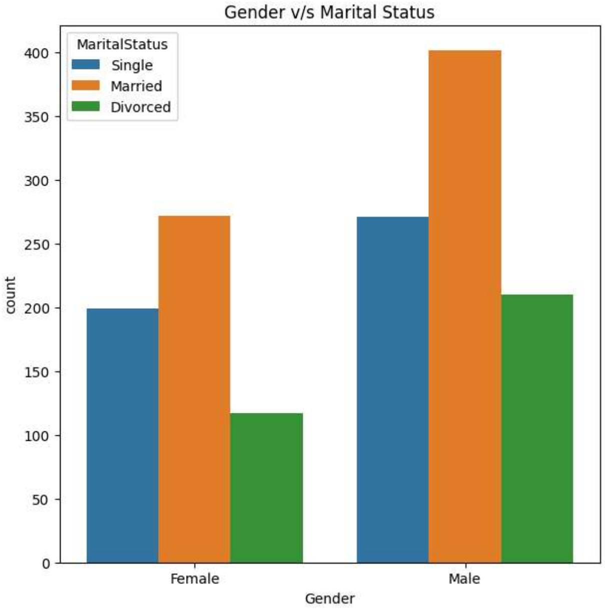
```
                count
MonthlyIncome
     2342           4
     6142           3
     2741           3
     2559           3
     2610           3
      ...         ...
     7104           1
     2773           1
    19513           1
     3447           1
     4404           1

1349 rows × 1 columns

dtype: int64
```

```
df.MonthlyIncome.value_counts().plot(kind='b
<Axes: xlabel='MonthlyIncome'>
```



```
plt.figure(figsize=(7,7))
plt.title("Gender v/s Marital Status")
sns.countplot(x = 'Gender',
              hue = 'MaritalStatus',
              data = df);
```

```
/usr/local/lib/python3.10/dist-packages/
  data_subset = grouped_data.get_group(p
/usr/local/lib/python3.10/dist-packages/
  data_subset = grouped_data.get_group(p
/usr/local/lib/python3.10/dist-packages/
  data_subset = grouped_data.get_group(p
```



```
plt.figure(figsize=(7,7))
plt.title("Job involement v/s Job Role")
sns.countplot(x = 'JobInvolvement',
              hue = 'JobRole',
              data = df);
plt.xlabel('Job involvement')
plt.ylabel('Job Role')
```

```
/usr/local/lib/python3.10/dist-packages/
  data_subset = grouped_data.get_group(p
/usr/local/lib/python3.10/dist-packages/
```