In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [2]:

```
df=pd.read_csv("HR Data.csv")
df
```

Out[2]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical
...	...	...	...	...	...	...	...	...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical

1470 rows × 35 columns

In [3]:

```
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
```

In [4]:

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

Out[4]:

```
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
```

In [5]:

```
df["Department"].value_counts()
```

Out[5]:

```
Department
Research & Development    961
Sales                     446
Human Resources            63
Name: count, dtype: int64
```

In [6]:

```
df["BusinessTravel"].value_counts()
```

Out[6]:

```
BusinessTravel
Travel_Rarely        1043
Travel_Frequently     277
Non-Travel            150
Name: count, dtype: int64
```

In [7]:

```
df["EducationField"].value_counts()
```

Out[7]:

```
EducationField
Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: count, dtype: int64
```

In [8]:

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

Out[8]:

```
Attrition
No     1233
Yes     237
Name: count, dtype: int64
```

In [9]:

```
df["Attrition"]=df["Attrition"].map({"Yes":1,"No":0})
```

In [10]:
```
df
```

Out[10]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
0	41	1	Travel_Rarely	1102	Sales	1	2	Life Sciences
1	49	0	Travel_Frequently	279	Research & Development	8	1	Life Sciences
2	37	1	Travel_Rarely	1373	Research & Development	2	2	Other
3	33	0	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
4	27	0	Travel_Rarely	591	Research & Development	2	1	Medical
...	...	...	...	...	...	...	...	...
1465	36	0	Travel_Frequently	884	Research & Development	23	2	Medical
1466	39	0	Travel_Rarely	613	Research & Development	6	1	Medical
1467	27	0	Travel_Rarely	155	Research & Development	4	3	Life Sciences
1468	49	0	Travel_Frequently	1023	Sales	2	3	Medical
1469	34	0	Travel_Rarely	628	Research & Development	8	3	Medical

1470 rows × 35 columns

In [11]:
```
df.drop(columns=["BusinessTravel","DailyRate","YearsWithCurrManager","StandardHours","Distance
```

In [12]:
```
df["Gender"]=df["Gender"].map({"Male":1,"Female":0})
```

In [13]:
```
df
```

Out[13]:

	Age	Attrition	Department	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentS
0	41	1	Sales	2	Life Sciences	1	1	
1	49	0	Research & Development	1	Life Sciences	1	2	
2	37	1	Research & Development	2	Other	1	4	
3	33	0	Research & Development	4	Life Sciences	1	5	
4	27	0	Research & Development	1	Medical	1	7	
...	...	...	...	...	...	...	...	
1465	36	0	Research & Development	2	Medical	1	2061	
1466	39	0	Research & Development	1	Medical	1	2062	
1467	27	0	Research & Development	3	Life Sciences	1	2064	
1468	49	0	Sales	3	Medical	1	2065	
1469	34	0	Research & Development	3	Medical	1	2068	

1470 rows × 30 columns

In [14]:

```
df1=df[["Department","Gender"]]
df1
```

Out[14]:

	Department	Gender
0	Sales	0
1	Research & Development	1
2	Research & Development	1
3	Research & Development	0
4	Research & Development	1
...	...	...
1465	Research & Development	1
1466	Research & Development	1
1467	Research & Development	1
1468	Sales	1
1469	Research & Development	1

1470 rows × 2 columns

In [15]:

```
x=df1.groupby("Department").sum()
x
```

Out[15]:

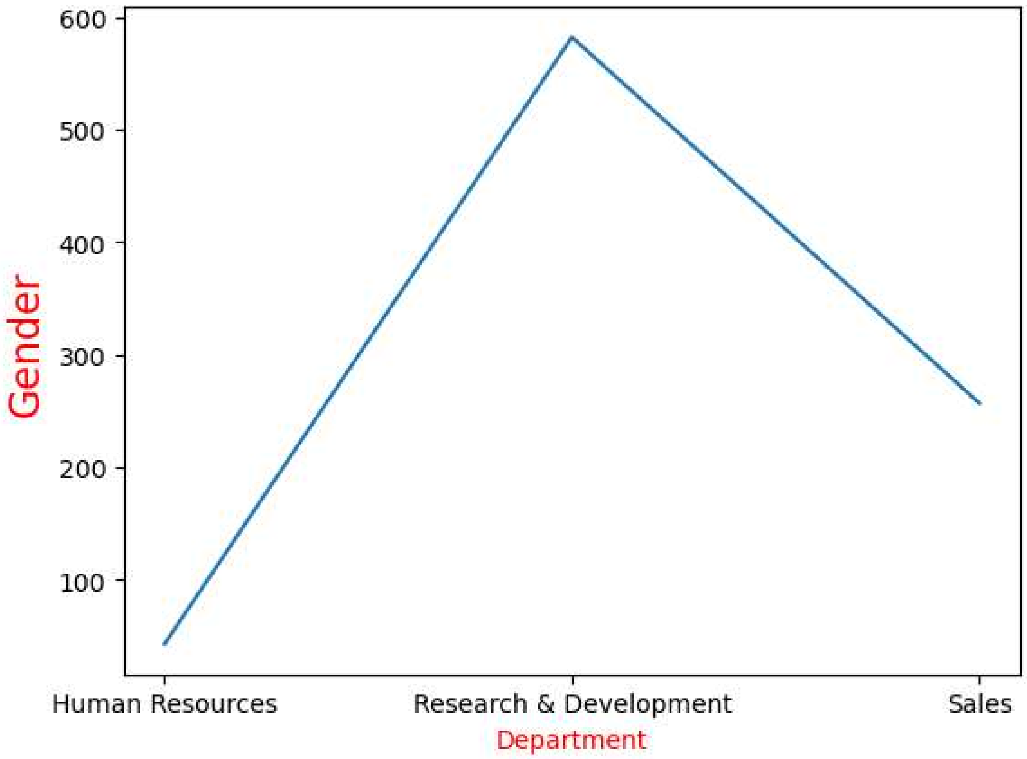
	Gender
Department	
Human Resources	43
Research & Development	582
Sales	257

In [16]:

```
plt.plot(x.index,x["Gender"])
plt.ylabel("Gender",fontsize=16, color="r")
plt.xlabel("Department",fontsize=10, color="r")
plt.plot()
```

Out[16]:

```
[]
```



In [17]:

```
df2=df[["Department","TotalWorkingYears"]]
df2
```

Out[17]:

	Department	TotalWorkingYears
0	Sales	8
1	Research & Development	10
2	Research & Development	7
3	Research & Development	8
4	Research & Development	6
...	...	...
1465	Research & Development	17
1466	Research & Development	9
1467	Research & Development	6
1468	Sales	17
1469	Research & Development	6

1470 rows × 2 columns

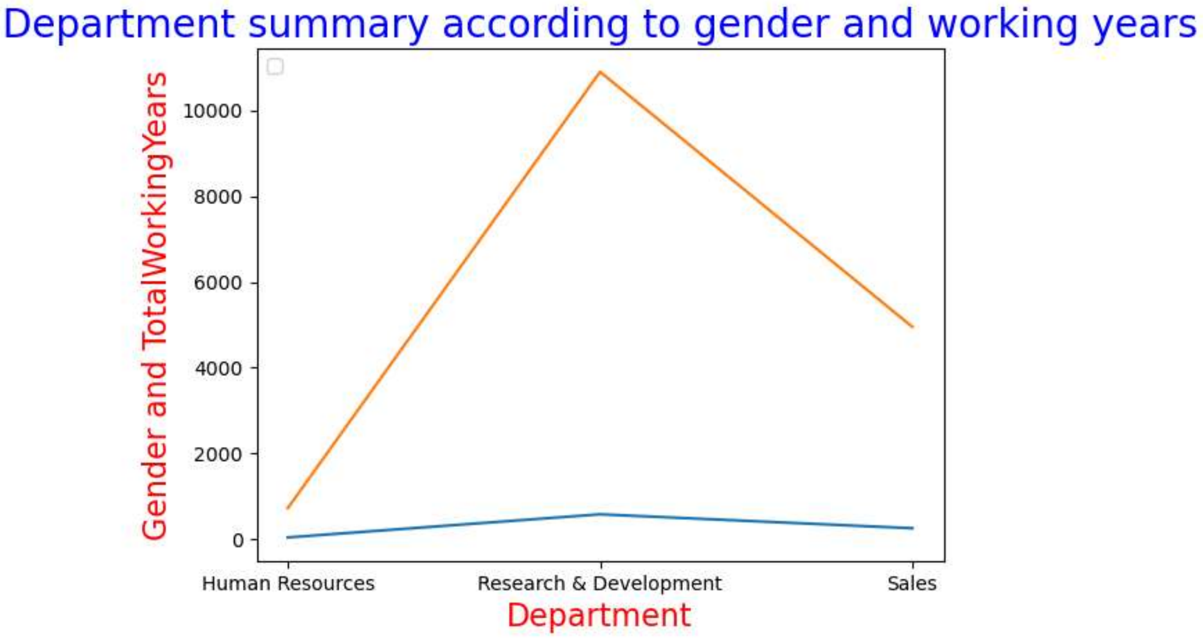
In [18]:
```
y=df2.groupby("Department").sum()
y
```

Out[18]:

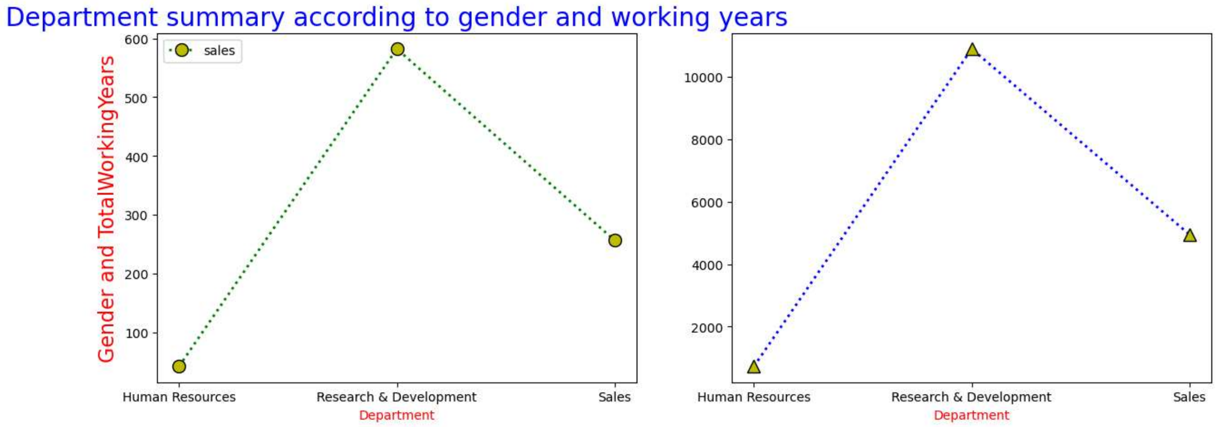
	TotalWorkingYears
Department	
Human Resources	728
Research & Development	10900
Sales	4953

In [19]:
```
from matplotlib import style
plt.plot(x.index,x["Gender"])
plt.plot(y.index,y["TotalWorkingYears"])
plt.xlabel("Department",fontsize=16, color="r")
plt.ylabel("Gender and TotalWorkingYears",fontsize=16, color="r")
plt.title("Department summary according to gender and working years",fontsize=20, color="b")
plt.legend(loc=2)
plt.show()
```

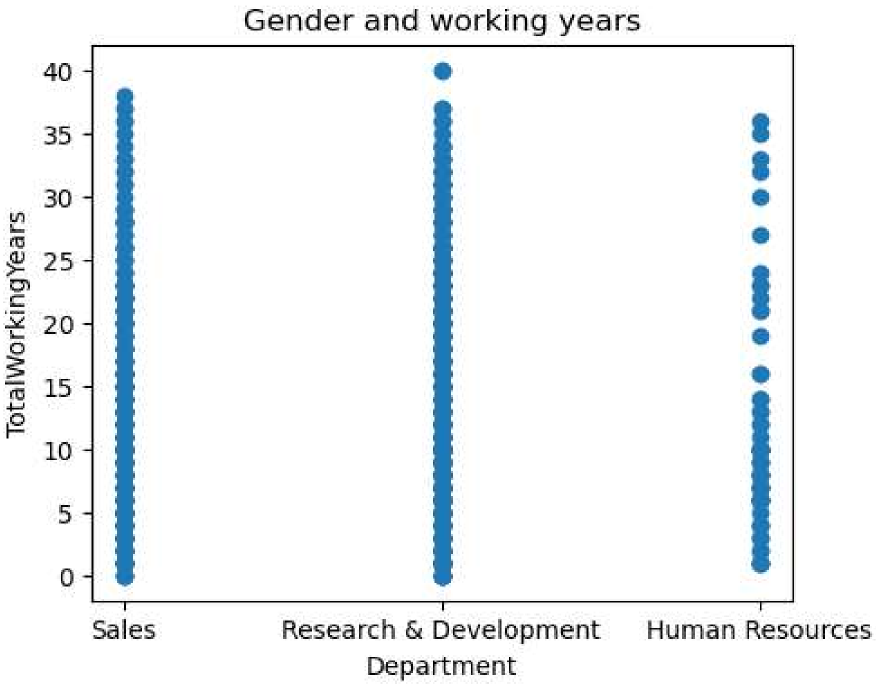
```
No artists with labels found to put in legend.  Note that artists whose label start with an un
derscore are ignored when legend() is called with no argument.
```


Department summary according to gender and working years

In [20]:
```
plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
plt.plot(x.index,x["Gender"],color="g",linewidth=2,marker="o",ms=10,mec="k",mfc="y",
         label="sales",linestyle=":")
plt.ylabel("Gender and TotalWorkingYears",fontsize=16, color="r")
plt.xlabel("Department",fontsize=10, color="r")
plt.title("Department summary according to gender and working years",fontsize=20, color="b")
plt.legend(loc=2)
plt.subplot(1,2,2)
plt.plot(y.index,y["TotalWorkingYears"],color="b",linewidth=2,marker="^",ms=10,mec="k",mfc="y"
         label="profit",linestyle=":")
plt.xlabel("Department",fontsize=10, color="r")
plt.show()
```



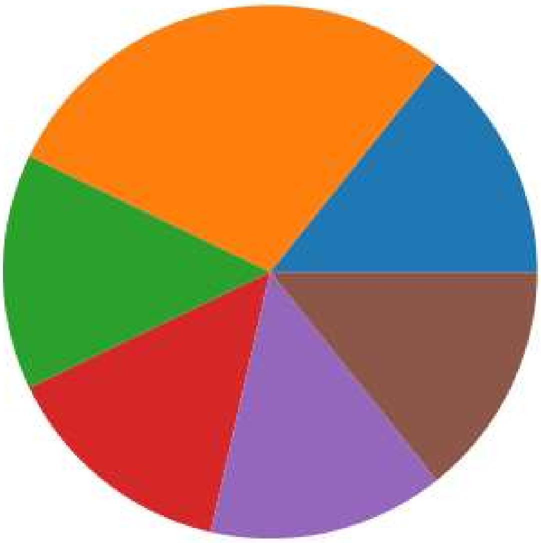
```
In [21]: plt.figure(figsize=(5,4))
         plt.scatter(df2.Department,df2.TotalWorkingYears)
         plt.xlabel("Department")
         plt.ylabel("TotalWorkingYears")
         plt.title("Gender and working years")
         plt.show()
```



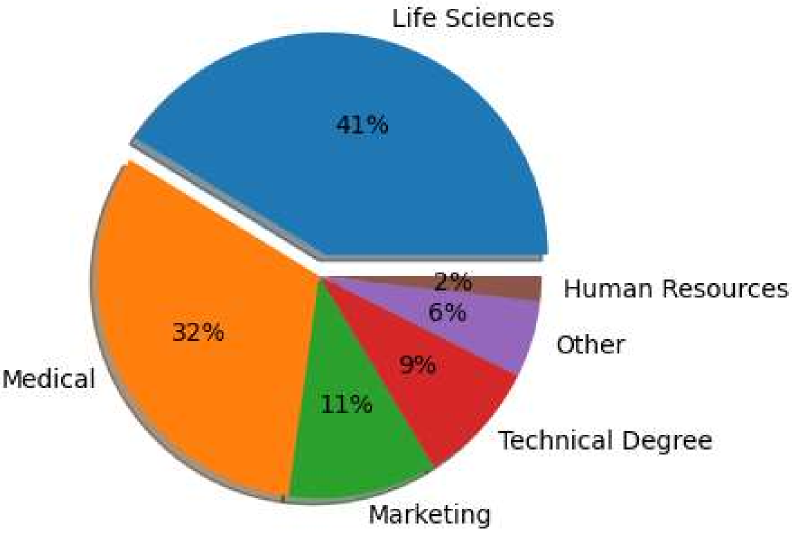
```
In [22]: EducationField=["Life Sciences","Medical","Marketing","Technical Degree","Other","Human Resour
         Values=[606,464,159,132,82,27]

         # Life Sciences       606
         # Medical             464
         # Marketing           159
         # Technical Degree    132
         # Other                82
         # Human Resources      27
```

```
In [23]: plt.pie([1,2,1,1,1,1])
         plt.show()
```

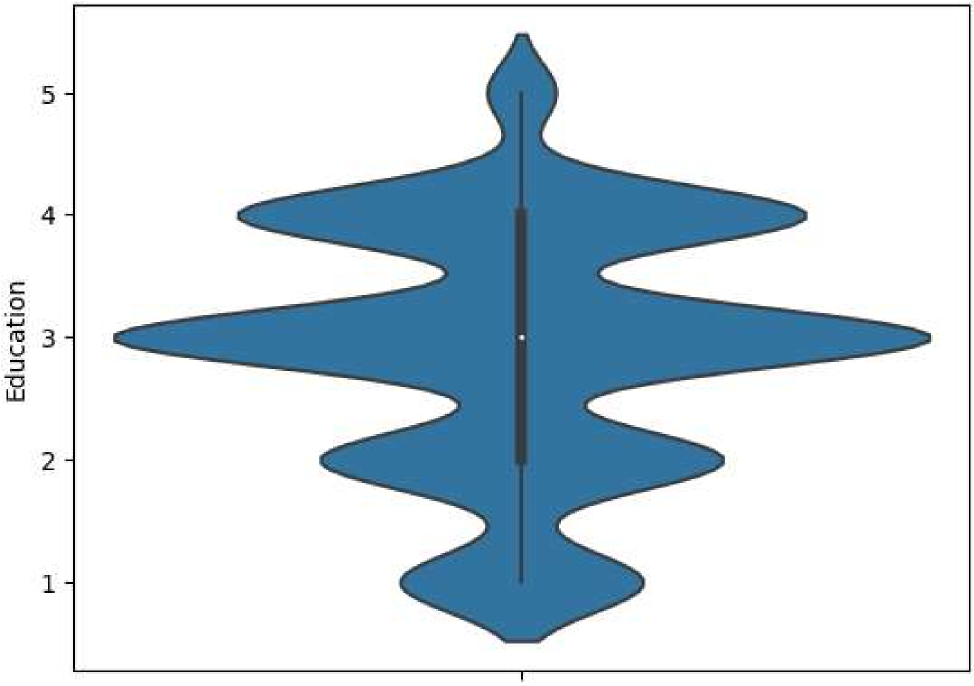


In [24]:
```
plt.figure(figsize=(5,4))
explode=[0.1,0,0,0,0,0]
plt.pie(Values,labels=EducationField,autopct="%0.0f%%",explode=explode,shadow=True)
plt.show()
```

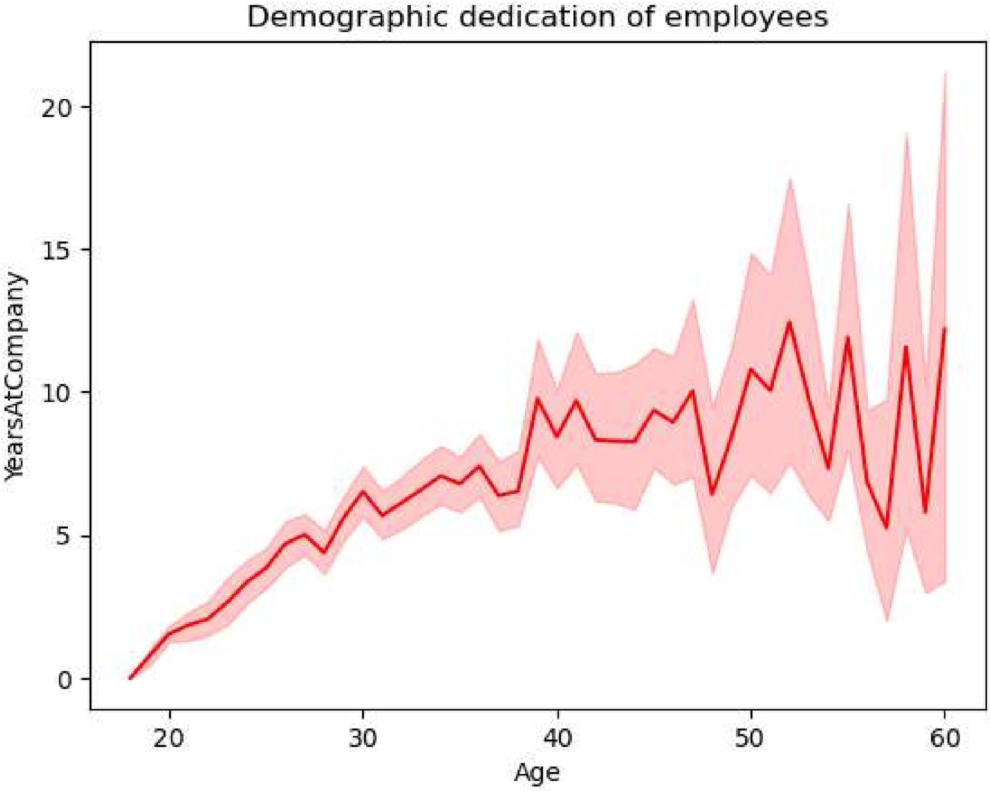


In [25]:
```
sns.violinplot(y="Education",data=df)
plt.plot()
```

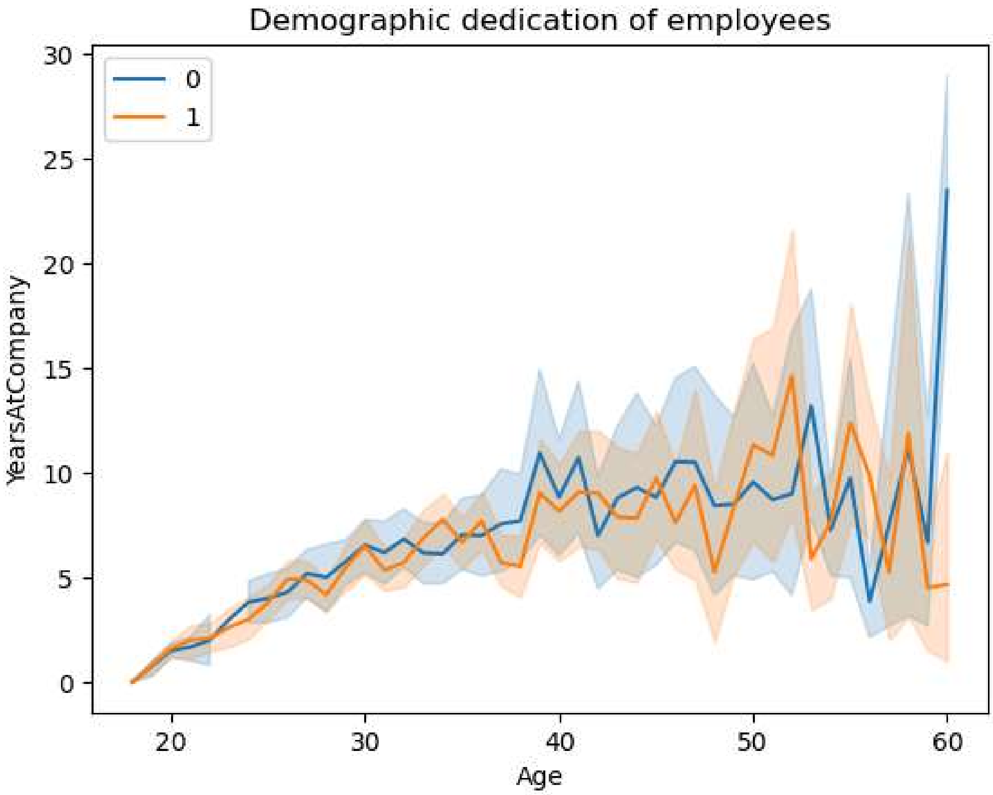
Out[25]:
```
[]
```



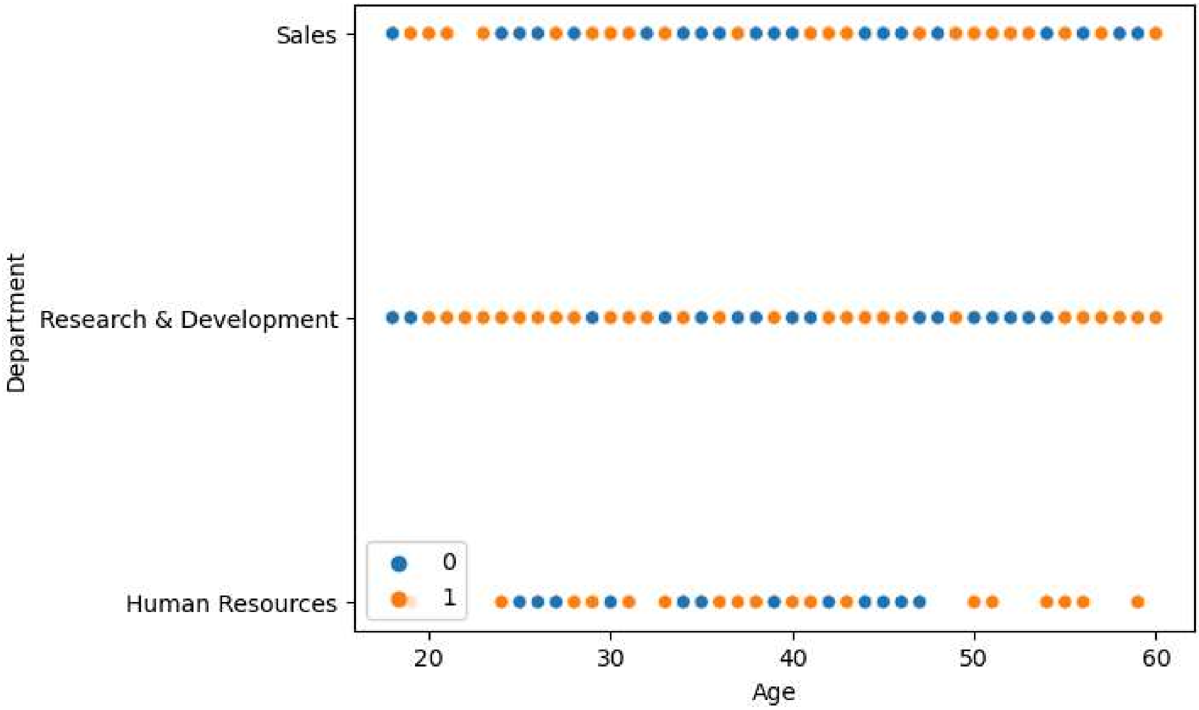
In [26]:
```
sns.lineplot(x="Age",y="YearsAtCompany",data=df,color="r")
plt.title("Demographic dedication of employees")
plt.show()
```



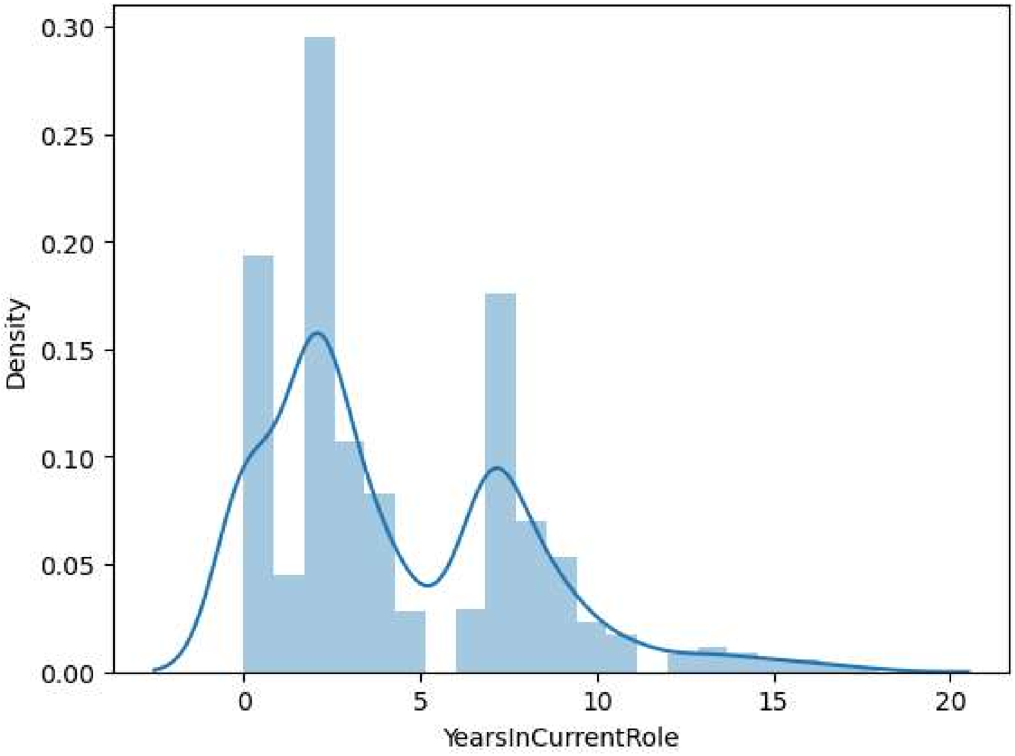
In [27]:
```
sns.lineplot(x="Age",y="YearsAtCompany",data=df,color="r",hue="Gender")
plt.title("Demographic dedication of employees")
plt.legend(loc=2)
plt.show()
```



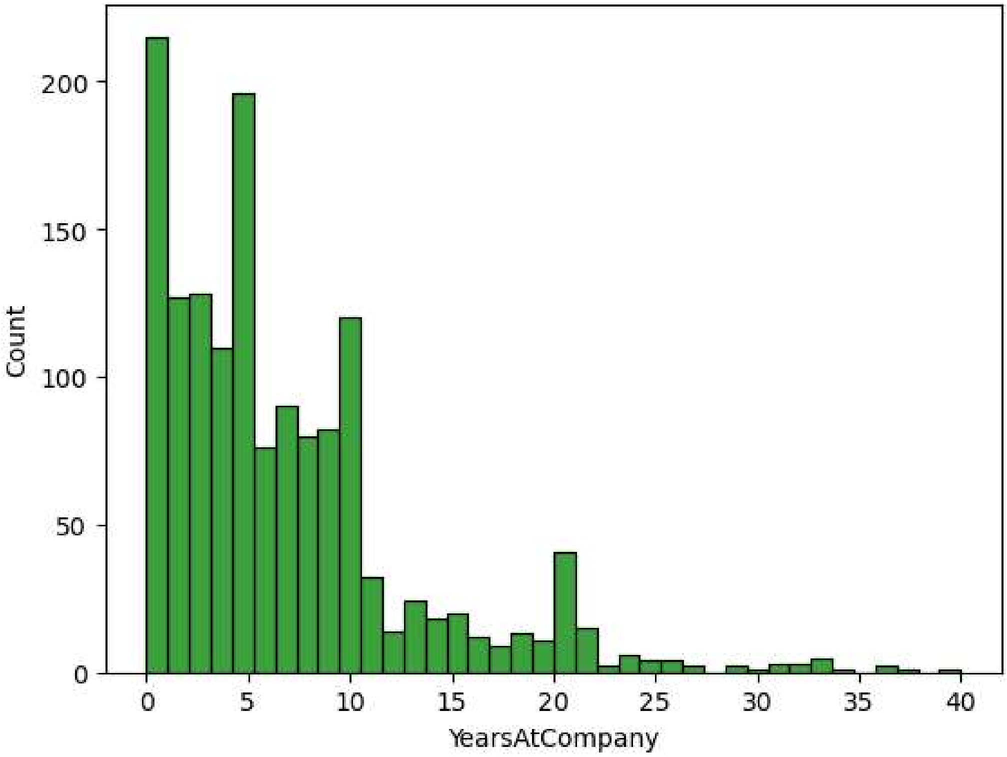
In [28]:
```
sns.scatterplot(x="Age",y="Department",data=df,hue="Gender")
plt.legend(loc=3)
plt.show()
```



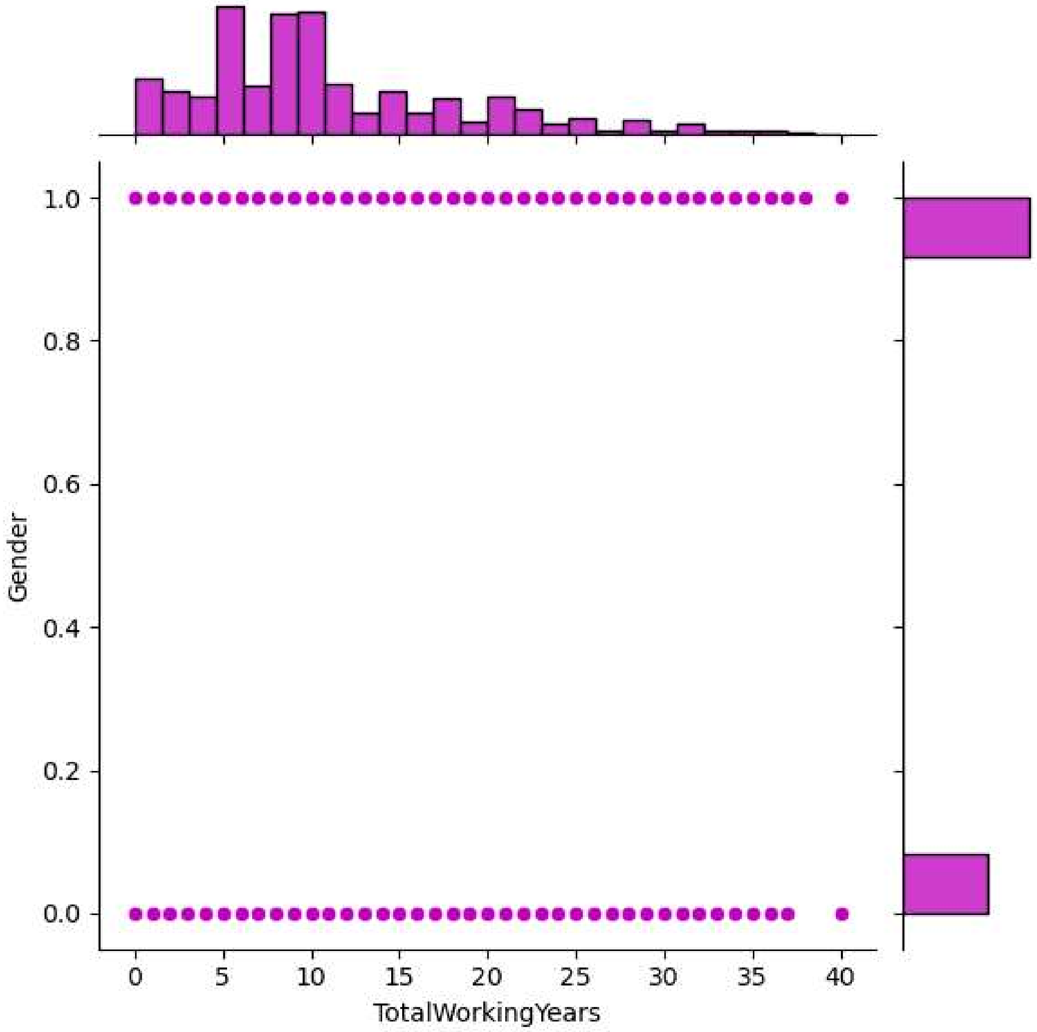
In [29]:
```
sns.distplot(df["YearsInCurrentRole"])
plt.show()
```



In [30]:
```
sns.histplot(df["YearsAtCompany"],color="g")
plt.show()
```



In [31]:
```
sns.jointplot(x="TotalWorkingYears",y="Gender",data=df,color="m")
plt.show()
```



In [32]:
```
df.describe()
```

Out[32]:

	Age	Attrition	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	
count	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470
mean	36.923810	0.161224	2.912925	1.0	1024.865306	2.721769	0
std	9.135373	0.367863	1.024165	0.0	602.024335	1.093082	0
min	18.000000	0.000000	1.000000	1.0	1.000000	1.000000	0
25%	30.000000	0.000000	2.000000	1.0	491.250000	2.000000	0
50%	36.000000	0.000000	3.000000	1.0	1020.500000	3.000000	1
75%	43.000000	0.000000	4.000000	1.0	1555.750000	4.000000	1
max	60.000000	1.000000	5.000000	1.0	2068.000000	4.000000	1

8 rows × 24 columns

In [37]:
```
df.drop(columns=["Department","EducationField"],axis=1,inplace=True)
```

In [38]:
```
df
```

Out[38]:

	Age	Attrition	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRa
0	41	1	2	1	1	2	0	
1	49	0	1	1	2	3	1	
2	37	1	2	1	4	4	1	
3	33	0	4	1	5	4	0	
4	27	0	1	1	7	1	1	
...	...	...	...	...	...	...	...	
1465	36	0	2	1	2061	3	1	
1466	39	0	1	1	2062	4	1	
1467	27	0	3	1	2064	2	1	
1468	49	0	3	1	2065	4	1	
1469	34	0	3	1	2068	2	1	

1470 rows × 28 columns

In [40]:

```
df.head()
```

Out[40]:

	Age	Attrition	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate
0	41	1	2	1	1	2	0	94
1	49	0	1	1	2	3	1	61
2	37	1	2	1	4	4	1	92
3	33	0	4	1	5	4	0	56
4	27	0	1	1	7	1	1	40

5 rows × 28 columns

In [ ]: