

In [1]:

import pandas as pd

In [2]:

import numpy as np
import matplotlib.pyplot as plt

In [3]:

import seaborn as sns

In [4]:

from plotly import __version__
print(__version__)
import cufflinks as cf
from plotly.offline import download_plotlyjs,init_notebook_mode,plot,iplot
init_notebook_mode(connected=True)
cf.go_offline()

4.8.2

In [5]:

import scipy.stats as stats

In [7]:

df=pd.read_csv(r"C:\Users\siyad\AppData\Local\Temp\Temp3_Day-7-20200715T141046Z-001.zip\Day-7\Assignment\general_data.csv")

In [8]:

df.head()

Out[8]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	...	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel	TotalV
0	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	...	1.0	Y	11	8	0	
1	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	...	0.0	Y	23	8	1	
2	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	...	1.0	Y	15	8	3	
3	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	...	3.0	Y	11	8	3	
4	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	...	4.0	Y	12	8	2	

5 rows × 24 columns

In [9]:

from scipy.stats import pearsonr as pr

In [10]:

stats,p=pr(df.YearsAtCompany,df.YearsSinceLastPromotion)

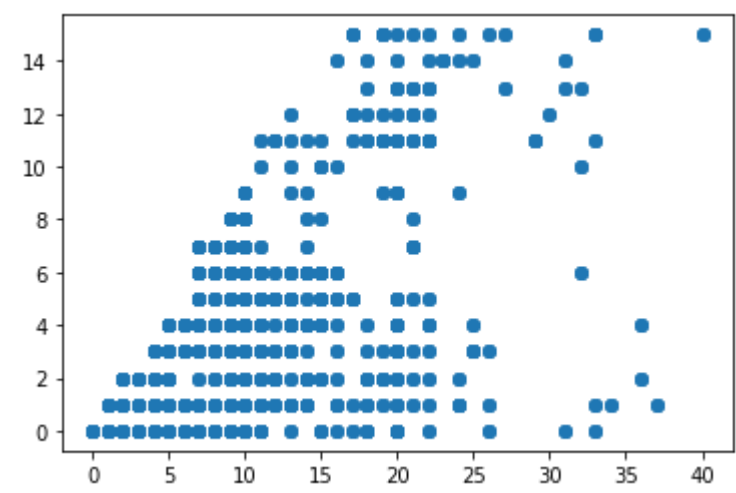
In [11]:

print(stats,p)

0.6184088652176049 0.0

```
In [12]: plt.scatter(df.YearsAtCompany,df.YearsSinceLastPromotion)
```

Out[12]: <matplotlib.collections.PathCollection at 0x167d3a82b00>



```
In [18]: df.columns
```

Out[18]: Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender', 'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'], dtype='object')

```
In [19]: df.isnull()
```

Out[19]:

Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	...	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWork
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
...	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	
False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	

rows × 24 columns



```
In [20]: df.isnull().sum()
```

```
Out[20]: Age                0
Attrition                 0
BusinessTravel            0
Department               0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeCount             0
EmployeeID                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
dtype: int64
```

```
In [21]: df1=df.dropna()
```

```
In [22]: df1.isnull().sum()
```

```
Out[22]: Age                0
Attrition                0
BusinessTravel           0
Department              0
DistanceFromHome        0
Education               0
EducationField           0
EmployeeCount           0
EmployeeID              0
Gender                  0
JobLevel                0
JobRole                 0
MaritalStatus           0
MonthlyIncome           0
NumCompaniesWorked      0
Over18                  0
PercentSalaryHike       0
StandardHours           0
StockOptionLevel        0
TotalWorkingYears       0
TrainingTimesLastYear   0
YearsAtCompany          0
YearsSinceLastPromotion 0
YearsWithCurrManager    0
dtype: int64
```

```
In [23]: df1.duplicated()
```

```
Out[23]: 0      False
1      False
2      False
3      False
4      False
...
4404   False
4405   False
4406   False
4407   False
4408   False
Length: 4382, dtype: bool
```

```
In [24]: df1.duplicated().sum()
```

Out[24]: 0

```
In [29]: df3=df1[['Age', 'DistanceFromHome', 'Education', 'MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike', 'TotalWorkingYears',
                'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']]
```

In [143]:

df3.describe()

Out[143]:

	Age	DistanceFromHome	Education	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager
count	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000	4382.000000
mean	36.933364	9.198996	2.912369	65061.702419	2.693291	15.210634	11.290278	2.798266	7.010497	2.191693	4.126198
std	9.137272	8.105396	1.024728	47142.310175	2.497832	3.663007	7.785717	1.289402	6.129351	3.224994	3.569674
min	18.000000	1.000000	1.000000	10090.000000	0.000000	11.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	30.000000	2.000000	2.000000	29110.000000	1.000000	12.000000	6.000000	2.000000	3.000000	0.000000	2.000000
50%	36.000000	7.000000	3.000000	49190.000000	2.000000	14.000000	10.000000	3.000000	5.000000	1.000000	3.000000
75%	43.000000	14.000000	4.000000	83790.000000	4.000000	18.000000	15.000000	3.000000	9.000000	3.000000	7.000000
max	60.000000	29.000000	5.000000	199990.000000	9.000000	25.000000	40.000000	6.000000	40.000000	15.000000	17.000000

In [33]:

df3.median()

Out[33]:

Age	36.0
DistanceFromHome	7.0
Education	3.0
MonthlyIncome	49190.0
NumCompaniesWorked	2.0
PercentSalaryHike	14.0
TotalWorkingYears	10.0
TrainingTimesLastYear	3.0
YearsAtCompany	5.0
YearsSinceLastPromotion	1.0
YearsWithCurrManager	3.0
dtype:	float64

In [34]:

df3.mode()

Out[34]:

	Age	DistanceFromHome	Education	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager
0	35	2	3	23420	1.0	11	10.0	2	5	0	2

In [35]: df3.var()

Out[35]: Age 8.348974e+01
DistanceFromHome 6.569744e+01
Education 1.050068e+00
MonthlyIncome 2.222397e+09
NumCompaniesWorked 6.239165e+00
PercentSalaryHike 1.341762e+01
TotalWorkingYears 6.061739e+01
TrainingTimesLastYear 1.662558e+00
YearsAtCompany 3.756894e+01
YearsSinceLastPromotion 1.040059e+01
YearsWithCurrManager 1.274257e+01
dtype: float64

In [36]: df3.skew()

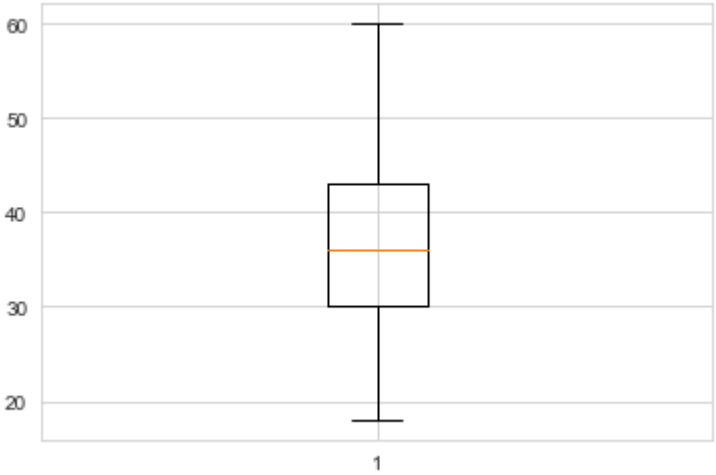
Out[36]: Age 0.413048
DistanceFromHome 0.955517
Education -0.288977
MonthlyIncome 1.367457
NumCompaniesWorked 1.029174
PercentSalaryHike 0.819510
TotalWorkingYears 1.115419
TrainingTimesLastYear 0.551818
YearsAtCompany 1.764619
YearsSinceLastPromotion 1.980992
YearsWithCurrManager 0.834277
dtype: float64

In [37]: df3.kurt()

Out[37]: Age -0.409517
DistanceFromHome -0.230691
Education -0.565008
MonthlyIncome 0.990836
NumCompaniesWorked 0.014307
PercentSalaryHike -0.306951
TotalWorkingYears 0.909316
TrainingTimesLastYear 0.494215
YearsAtCompany 3.930726
YearsSinceLastPromotion 3.592162
YearsWithCurrManager 0.170703
dtype: float64

```
In [97]: plt.boxplot(df1['Age'])
```

Out[97]: {'whiskers': [<matplotlib.lines.Line2D at 0x2a4625e1ac8>, <matplotlib.lines.Line2D at 0x2a4625e1e80>], 'caps': [<matplotlib.lines.Line2D at 0x2a462602278>, <matplotlib.lines.Line2D at 0x2a462602630>], 'boxes': [<matplotlib.lines.Line2D at 0x2a4625e1780>], 'medians': [<matplotlib.lines.Line2D at 0x2a4626029e8>], 'fliers': [<matplotlib.lines.Line2D at 0x2a462602da0>], 'means': []}

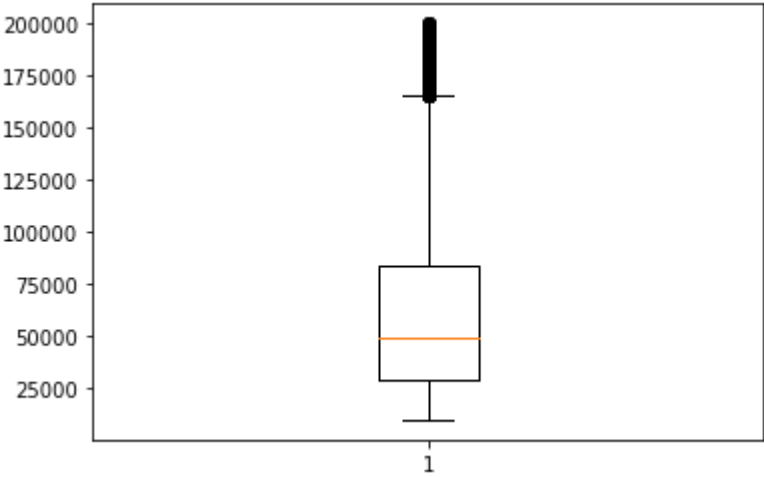


```
In [98]: stats.iqr(df1['Age'])
```

Out[98]: 13.0

```
In [39]: plt.boxplot(df1['MonthlyIncome'])
```

Out[39]: {'whiskers': [<matplotlib.lines.Line2D at 0x2a45da94358>, <matplotlib.lines.Line2D at 0x2a45b9ee0b8>], 'caps': [<matplotlib.lines.Line2D at 0x2a45da9f400>, <matplotlib.lines.Line2D at 0x2a45da9f7b8>], 'boxes': [<matplotlib.lines.Line2D at 0x2a45da94160>], 'medians': [<matplotlib.lines.Line2D at 0x2a45da87940>], 'fliers': [<matplotlib.lines.Line2D at 0x2a45da802e8>], 'means': []}

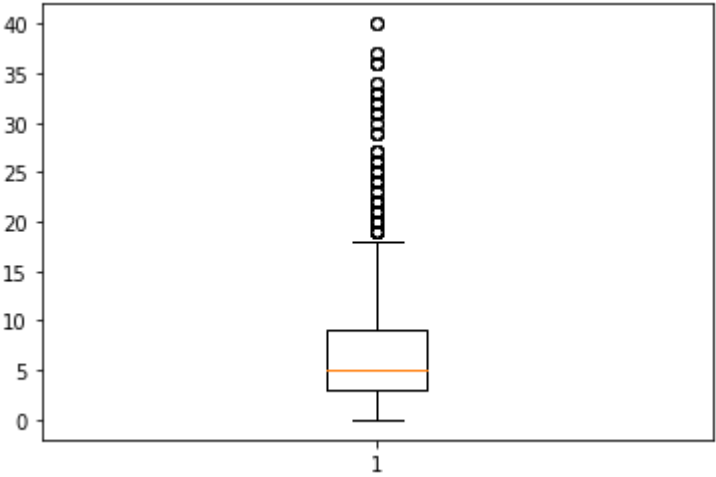


```
In [99]: stats.iqr(df1['MonthlyIncome'])
```

Out[99]: 54680.0


```
In [40]: plt.boxplot(df1['YearsAtCompany'])
```

Out[40]: {'whiskers': [<matplotlib.lines.Line2D at 0x2a45db144e0>,
<matplotlib.lines.Line2D at 0x2a45db14860>],
'caps': [<matplotlib.lines.Line2D at 0x2a45db14be0>,
<matplotlib.lines.Line2D at 0x2a45db14f60>],
'boxes': [<matplotlib.lines.Line2D at 0x2a45db141d0>],
'medians': [<matplotlib.lines.Line2D at 0x2a45db20320>],
'fliers': [<matplotlib.lines.Line2D at 0x2a45db206a0>],
'means': []}

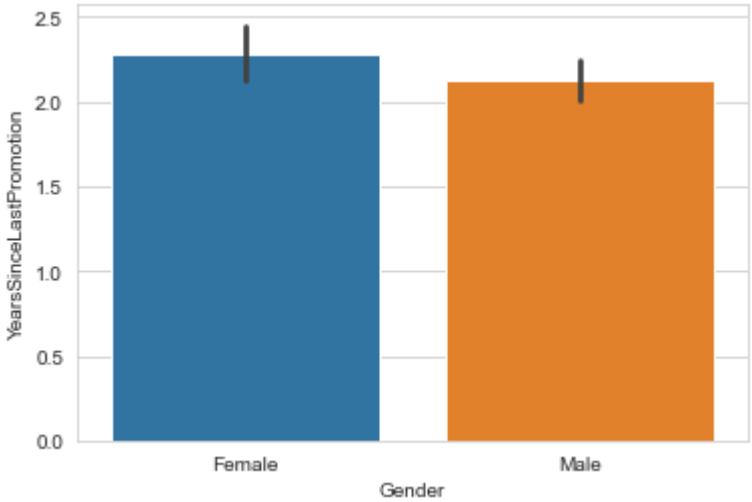


```
In [100]: stats.iqr(df1['YearsAtCompany'])
```

Out[100]: 6.0

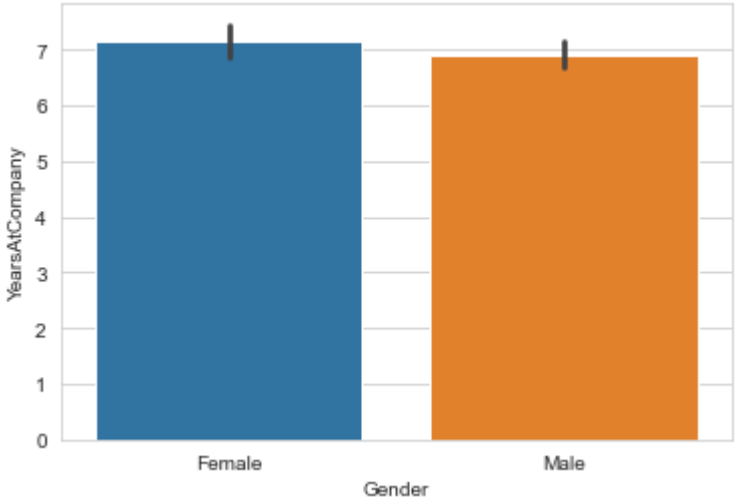
```
In [173]: sns.barplot(x='Gender',y='YearsSinceLastPromotion',data=df1,estimator=np.mean)
```

Out[173]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4652a4358>



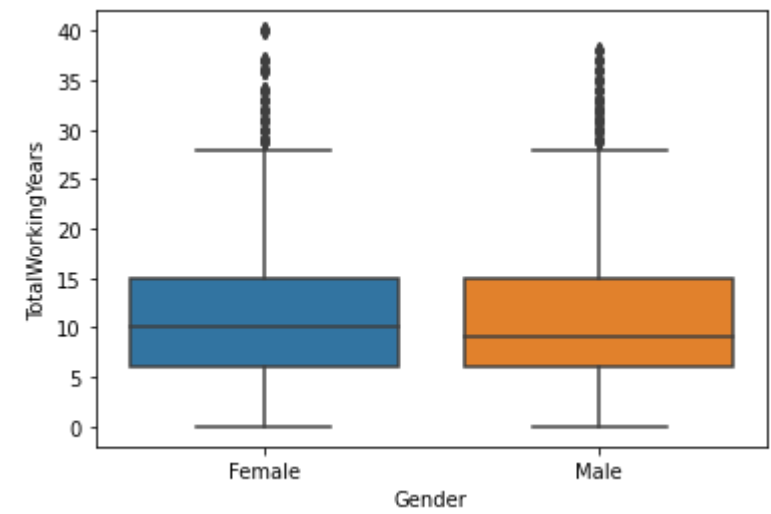
```
In [174]: sns.barplot(x='Gender',y='YearsAtCompany',data=df1,estimator=np.mean)
```

Out[174]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4634b8128>



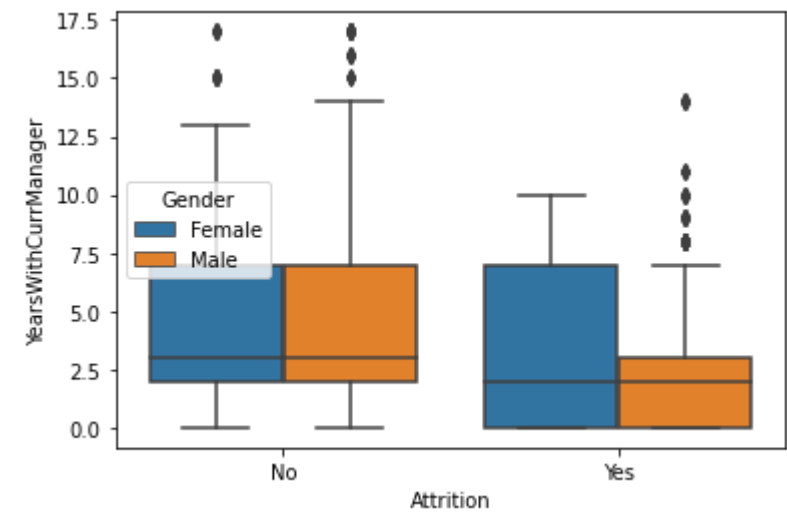
```
In [43]: sns.boxplot(x='Gender',y='TotalWorkingYears',data=df1)
```

Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45dbfc9e8>



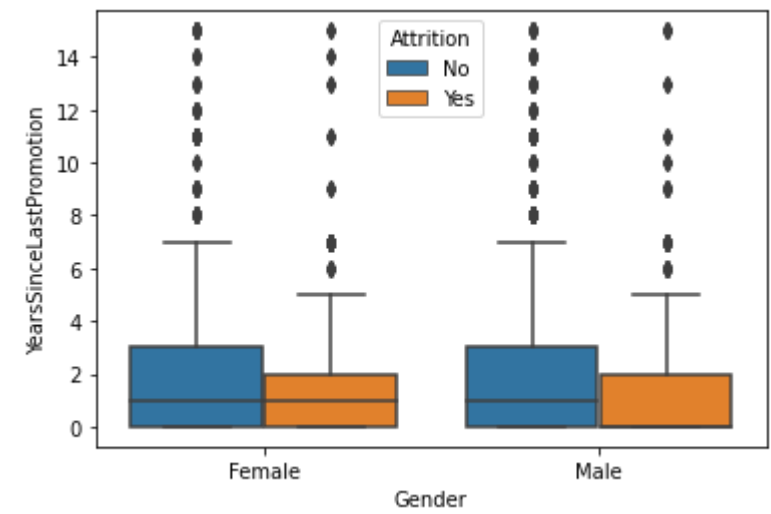
```
In [44]: sns.boxplot(x='Attrition',y='YearsWithCurrManager',data=df1,hue='Gender')
```

Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45dc7b7b8>



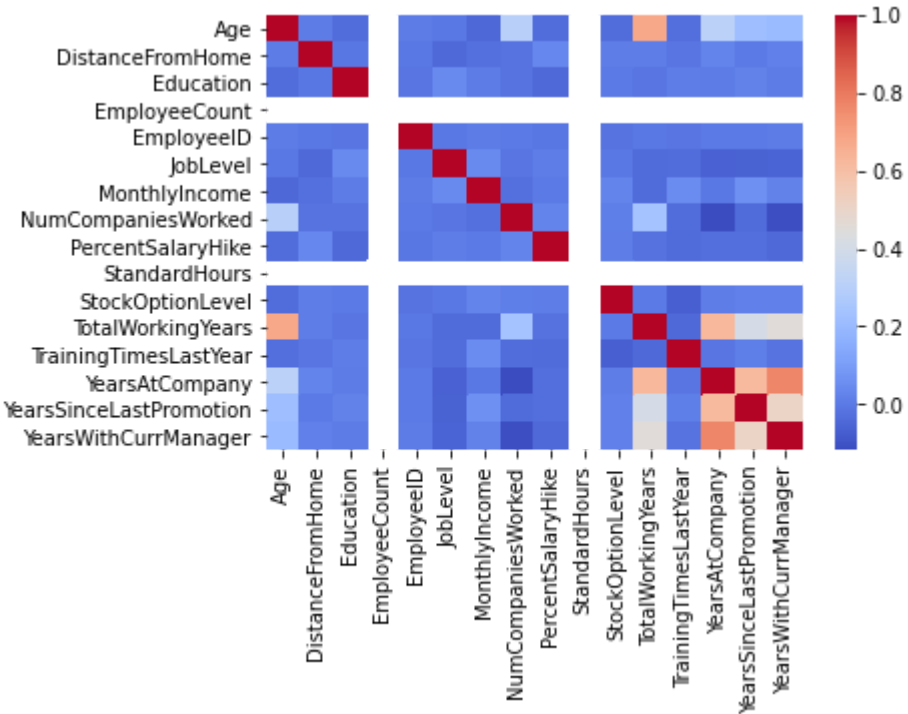
```
In [45]: sns.boxplot(x='Gender',y='YearsSinceLastPromotion',data=df1,hue='Attrition')
```

Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45dd533c8>



```
In [46]: dc=df1.corr()  
sns.heatmap(dc,cmap='coolwarm')
```

Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x2a435f0c710>



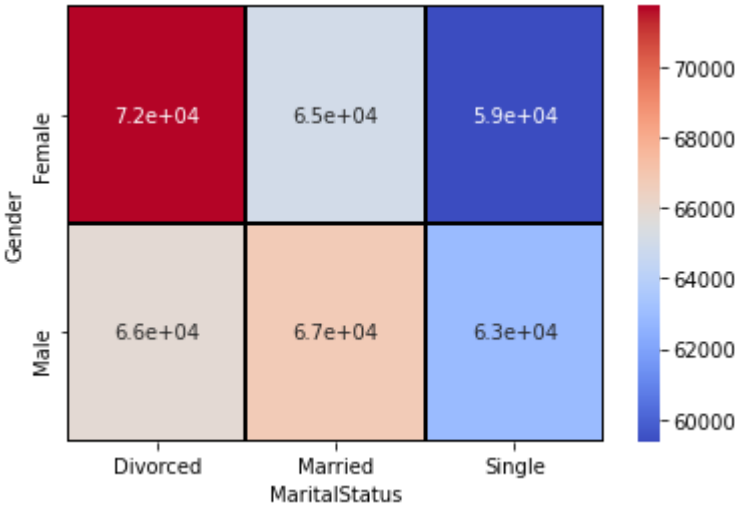
```
In [47]: df1.corr()
```

Out[47]:

	Age	DistanceFromHome	Education	EmployeeCount	EmployeeID	JobLevel	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	Training1
Age	1.000000	0.007376	-0.033900	NaN	0.008105	-0.001137	-0.045163	0.299527	-0.032561	NaN	-0.031504	0.680037	
DistanceFromHome	0.007376	1.000000	-0.007491	NaN	-0.000326	-0.039990	-0.022757	-0.014449	0.037720	NaN	0.009353	0.009574	
Education	-0.033900	-0.007491	1.000000	NaN	-0.009389	0.045822	0.007289	-0.016210	-0.041054	NaN	0.002386	-0.009228	
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeID	0.008105	-0.000326	-0.009389	NaN	1.000000	-0.003090	0.007865	0.000719	-0.004877	NaN	-0.013488	-0.001688	
JobLevel	-0.001137	-0.039990	0.045822	NaN	-0.003090	1.000000	0.046688	-0.009759	0.010874	NaN	0.000365	-0.036293	
MonthlyIncome	-0.045163	-0.022757	0.007289	NaN	0.007865	0.046688	1.000000	-0.021446	0.004607	NaN	0.027242	-0.034398	
NumCompaniesWorked	0.299527	-0.014449	-0.016210	NaN	0.000719	-0.009759	-0.021446	1.000000	0.030064	NaN	0.016291	0.238807	
PercentSalaryHike	-0.032561	0.037720	-0.041054	NaN	-0.004877	0.010874	0.004607	0.030064	1.000000	NaN	0.012104	-0.018480	
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	-0.031504	0.009353	0.002386	NaN	-0.013488	0.000365	0.027242	0.016291	0.012104	NaN	1.000000	0.003138	
TotalWorkingYears	0.680037	0.009574	-0.009228	NaN	-0.001688	-0.036293	-0.034398	0.238807	-0.018480	NaN	0.003138	1.000000	
TrainingTimesLastYear	-0.028962	-0.008957	0.009939	NaN	-0.012102	-0.031931	0.049780	-0.031335	-0.036720	NaN	-0.068028	-0.041459	
YearsAtCompany	0.311281	0.030746	0.005997	NaN	0.004117	-0.063360	0.000088	-0.117213	-0.029022	NaN	0.009369	0.627738	
YearsSinceLastPromotion	0.215650	0.002243	0.023457	NaN	0.000814	-0.059680	0.064470	-0.035855	-0.028654	NaN	0.019889	0.404058	
YearsWithCurrManager	0.201580	0.021773	0.005645	NaN	0.009079	-0.053898	0.023095	-0.109372	-0.039687	NaN	0.019398	0.458640	

```
In [48]: fp=df1.pivot_table(index='Gender',columns='MaritalStatus',values='MonthlyIncome')
sns.heatmap(fp,annot=True,cmap='coolwarm',linecolor='black',linewidth=1)
```

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45debbc50>



In [49]:

data=df1[['YearsWithCurrManager','YearsAtCompany']]
data.corr()

Out[49]:

	YearsWithCurrManager	YearsAtCompany
YearsWithCurrManager	1.000000	0.769161
YearsAtCompany	0.769161	1.000000

In [51]:

df1.columns

Out[51]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
 'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
 'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],
 dtype='object')

In [52]:

crosstab1 = pd.crosstab(df1["Gender"], df1["JobRole"])
crosstab1

Out[52]:

JobRole	Healthcare Representative	Human Resources	Laboratory Technician	Manager	Manufacturing Director	Research Director	Research Scientist	Sales Executive	Sales Representative
Gender									
Female	158	60	299	146	175	96	322	396	104
Male	231	96	474	159	254	141	550	579	142

In [53]:

stats.chi2_contingency(crosstab1)

Out[53]:

(12.882211411612944,
 0.11596711816900672,
 8,
 array([[155.8840712 , 62.51392058, 309.7644911 , 122.22272935,
 171.91328161, 94.97307166, 349.43678686, 390.71200365,
 98.579644],
 [233.1159288 , 93.48607942, 463.2355089 , 182.77727065,
 257.08671839, 142.02692834, 522.56321314, 584.28799635,
 147.420356]]))

In [54]:

df1['Attrition'].value_counts()

Out[54]:

No 3677
Yes 705
Name: Attrition, dtype: int64

In [55]:

atr_yes=df1[df1['Attrition']=='Yes']

```
In [56]: fem_yes=atr_yes[atr_yes['Gender']=='Female']

In [57]: male_yes=atr_yes[atr_yes['Gender']=='Male']

In [58]: atr_no=df1[df1['Attrition']=='No']

In [59]: fem_no=atr_no[atr_no['Gender']=='Female']

In [60]: male_no=atr_no[atr_no['Gender']=='Male']

In [61]: atr_yes['Gender'].value_counts()
Out[61]: Male      437
         Female    268
         Name: Gender, dtype: int64

In [62]: atr_yes['JobRole'].value_counts()
Out[62]: Sales Executive      165
         Research Scientist   158
         Laboratory Technician 125
         Healthcare Representative 56
         Research Director    54
         Manufacturing Director 48
         Manager              42
         Sales Representative  36
         Human Resources       21
         Name: JobRole, dtype: int64

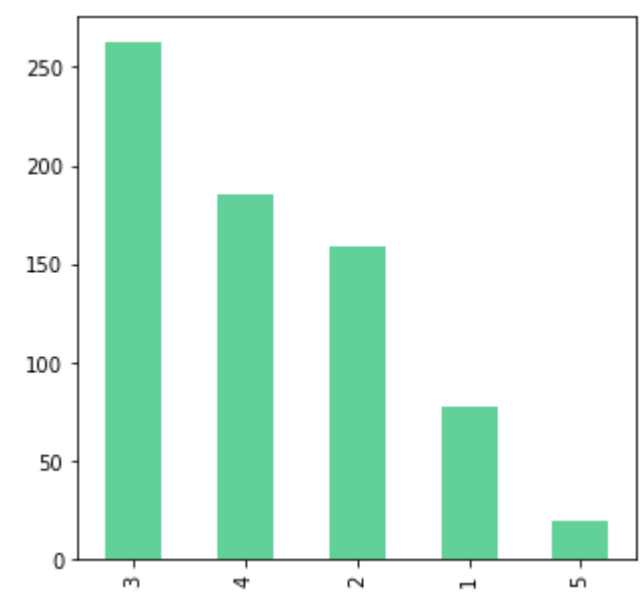
In [63]: atr_yes['JobLevel'].value_counts()
Out[63]: 2      283
         1      250
         3       96
         4       51
         5       25
         Name: JobLevel, dtype: int64

In [64]: atr_yes['EmployeeCount'].value_counts()
Out[64]: 1      705
         Name: EmployeeCount, dtype: int64
```



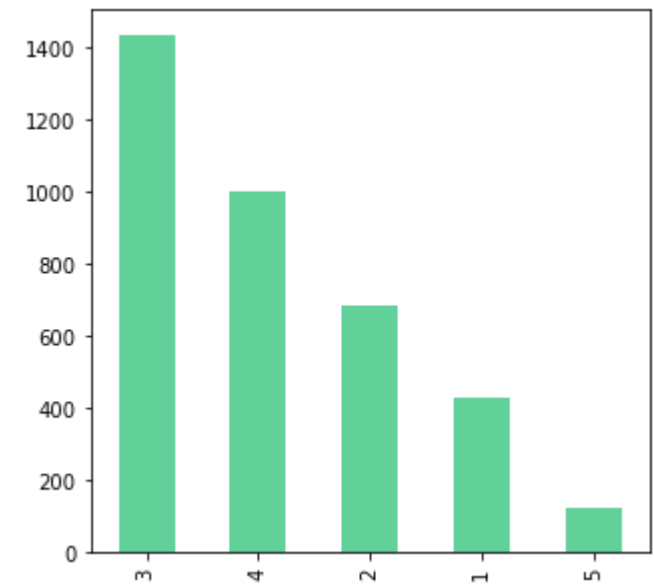
```
In [65]: atr_yes['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[65]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fbe9358>



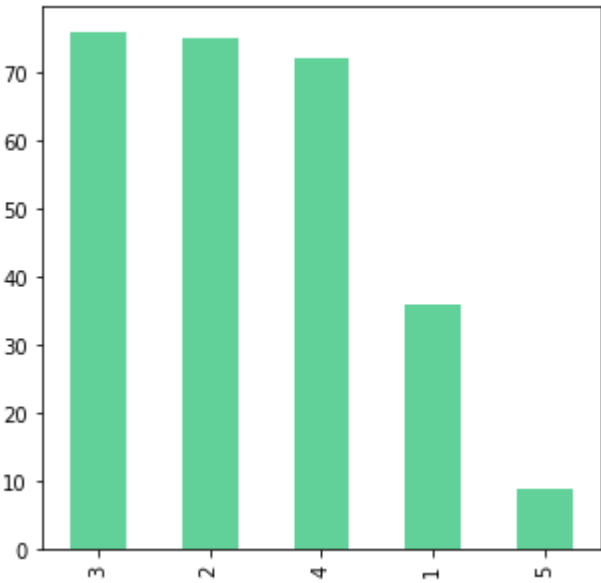
```
In [66]: atr_no['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[66]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fc3e390>



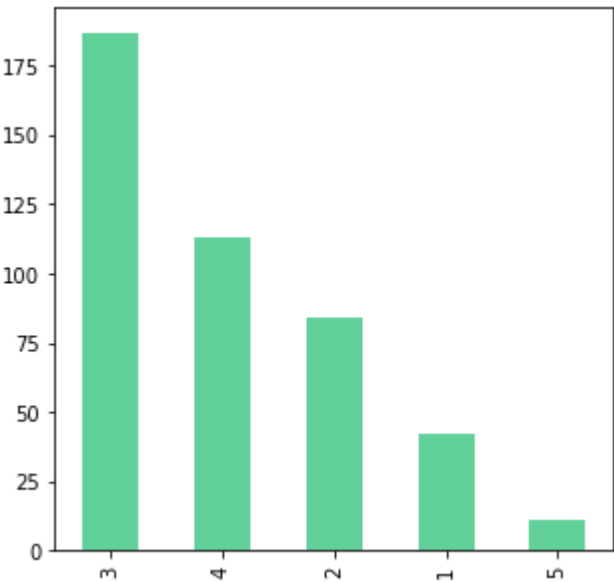
```
In [67]: fem_yes['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fc8ac18>



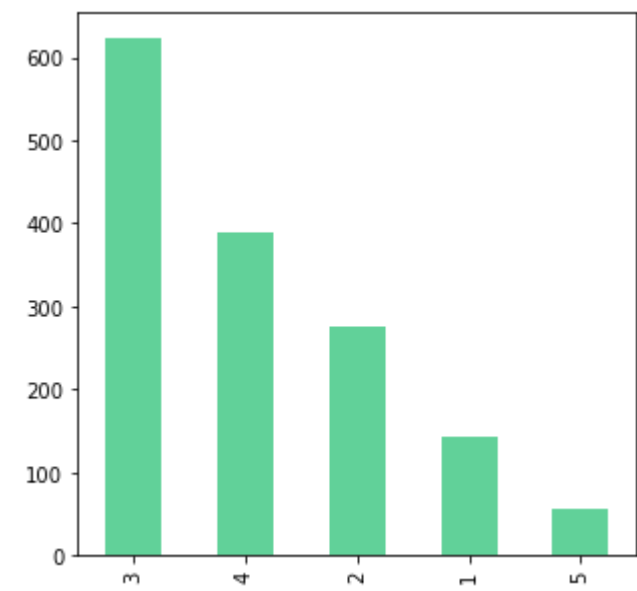
```
In [68]: male_yes['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[68]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fcf1cc0>



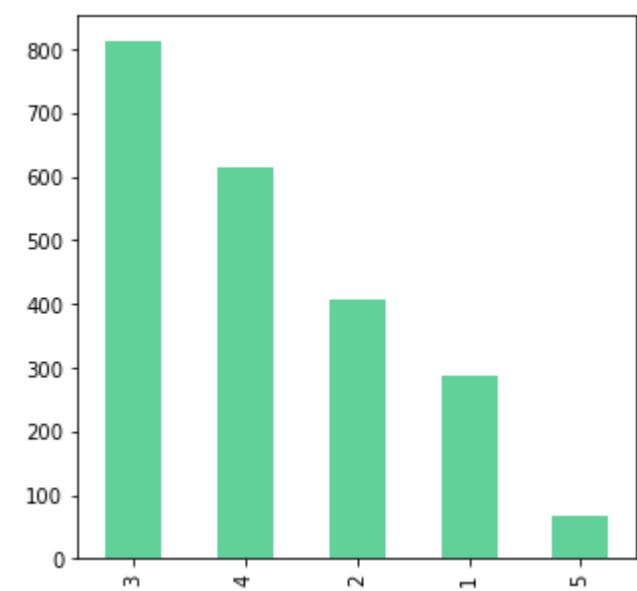
```
In [69]: fem_no['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[69]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fd60358>



```
In [70]: male_no['Education'].value_counts().plot(kind="bar", figsize=(5,5), color="#61d199")
```

Out[70]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fdc89e8>

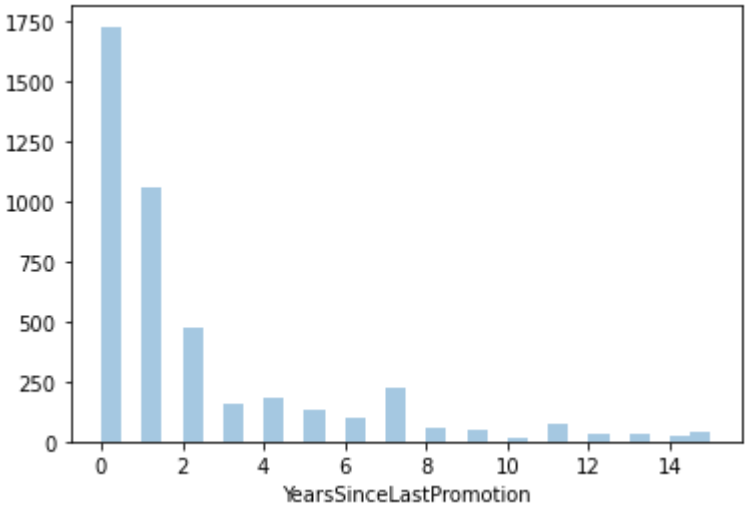


```
In [71]: df1.columns
```

Out[71]: Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender', 'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'], dtype='object')

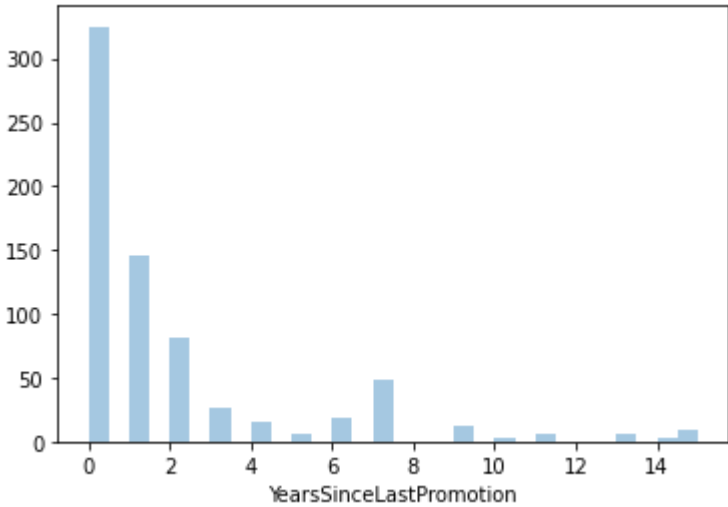
```
In [72]: sns.distplot(df1['YearsSinceLastPromotion'],kde=False,bins=30)
```

Out[72]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fe38a58>



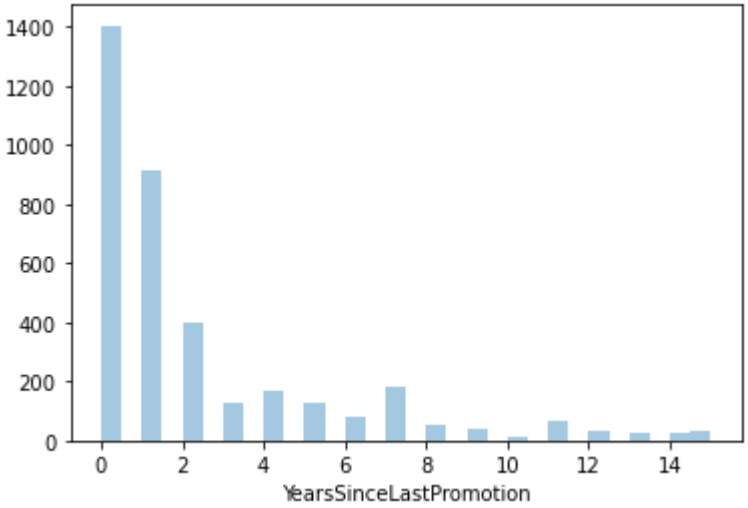
```
In [73]: sns.distplot(atr_yes['YearsSinceLastPromotion'],kde=False,bins=30)
```

Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fed4780>



```
In [74]: sns.distplot(atr_no['YearsSinceLastPromotion'],kde=False,bins=30)
```

Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45ffbb4e0>

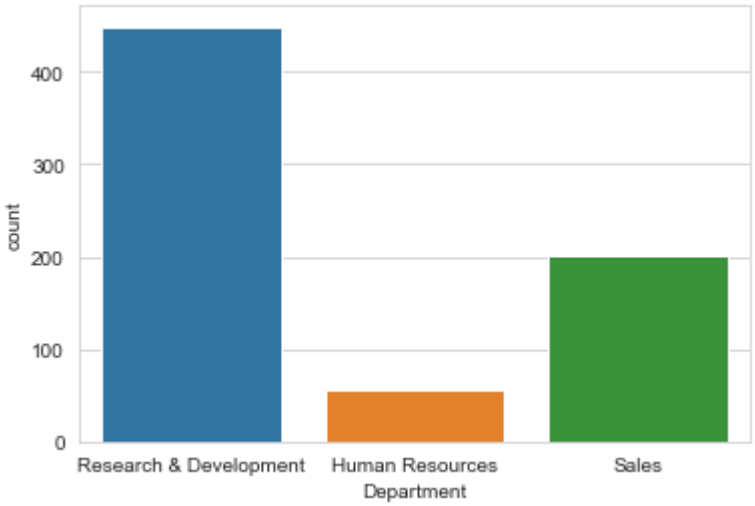


```
In [75]: atr_yes['BusinessTravel'].value_counts()
```

Out[75]: Travel_Rarely 465
Travel_Frequently 204
Non-Travel 36
Name: BusinessTravel, dtype: int64

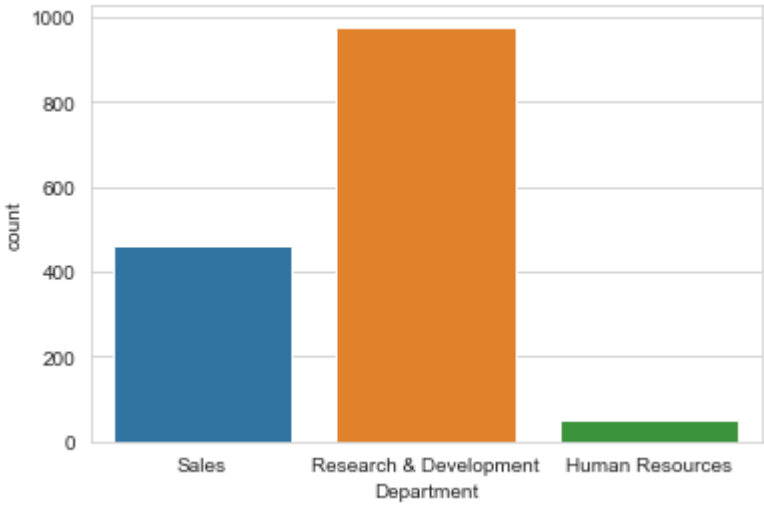
```
In [76]: sns.set_style('whitegrid')  
sns.countplot(x='Department',data=atr_yes)
```

Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0x2a460067a90>



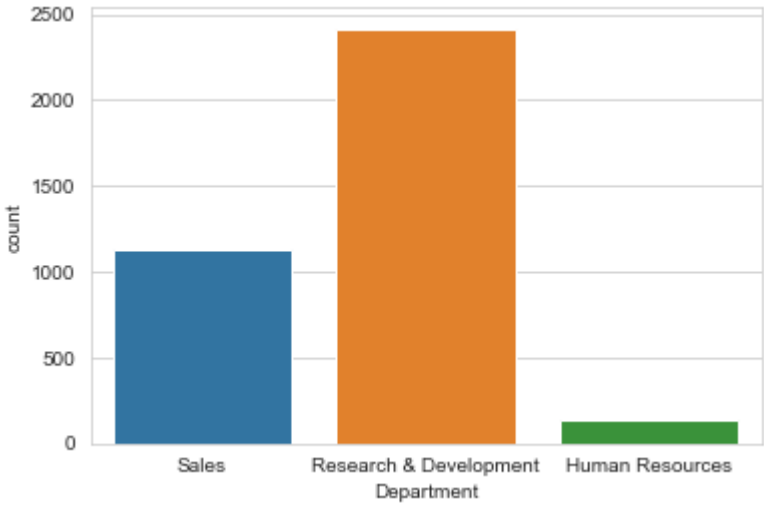
```
In [77]: sns.set_style('whitegrid')
sns.countplot(x='Department',data=fem_no)
```

Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fd68e10>



```
In [78]: sns.set_style('whitegrid')
sns.countplot(x='Department',data=atr_no)
```

Out[78]: <matplotlib.axes._subplots.AxesSubplot at 0x2a45fdc1668>



```
In [79]: atr_yes['Department'].value_counts()
```

Out[79]: Research & Development 449
Sales 201
Human Resources 55
Name: Department, dtype: int64

```
In [96]: stats.iqr(df1['YearsAtCompany'])
```

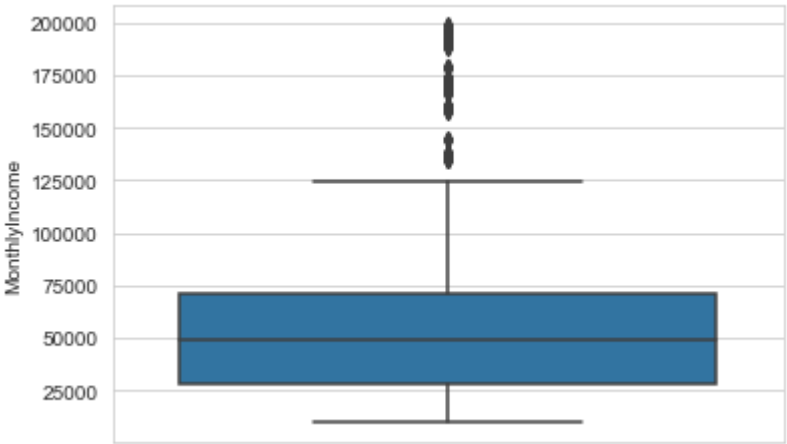
Out[96]: 6.0

```
In [101]: df1.columns
```

Out[101]: Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender', 'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'], dtype='object')

```
In [105]: sns.boxplot(y='MonthlyIncome',data=atr_yes)
```

Out[105]: <matplotlib.axes._subplots.AxesSubplot at 0x2a462c57240>

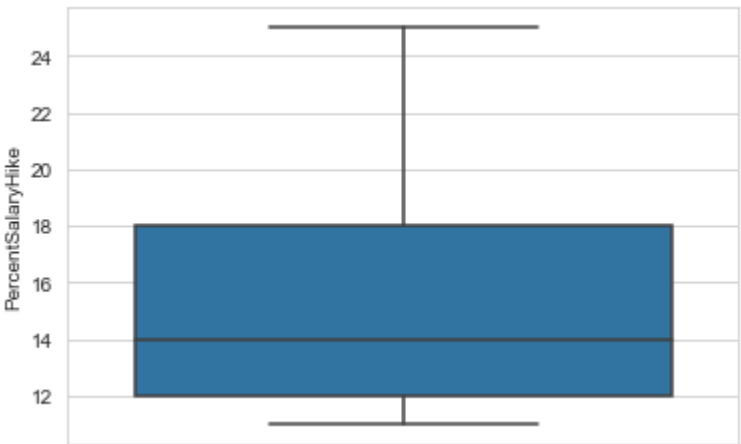


```
In [106]: stats.iqr(atr_yes['MonthlyIncome'])
```

Out[106]: 42600.0

```
In [108]: sns.boxplot(y='PercentSalaryHike',data=atr_yes)
```

Out[108]: <matplotlib.axes._subplots.AxesSubplot at 0x2a462c41828>

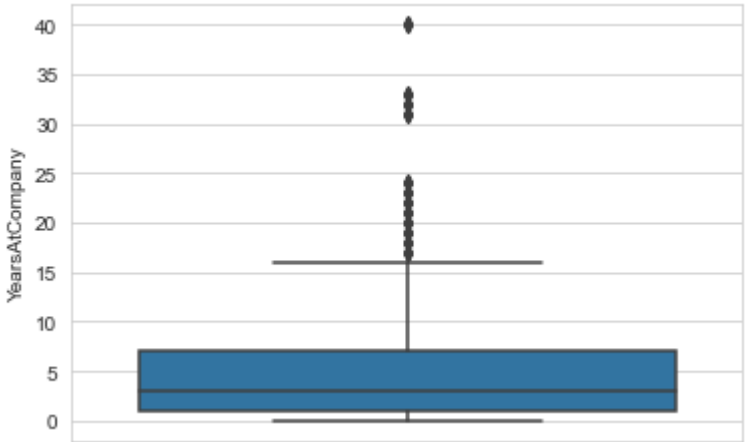


```
In [109]: stats.iqr(atr_yes['PercentSalaryHike'])
```

Out[109]: 6.0

```
In [110]: sns.boxplot(y='YearsAtCompany',data=atr_yes)
```

Out[110]: <matplotlib.axes._subplots.AxesSubplot at 0x2a46225a7b8>

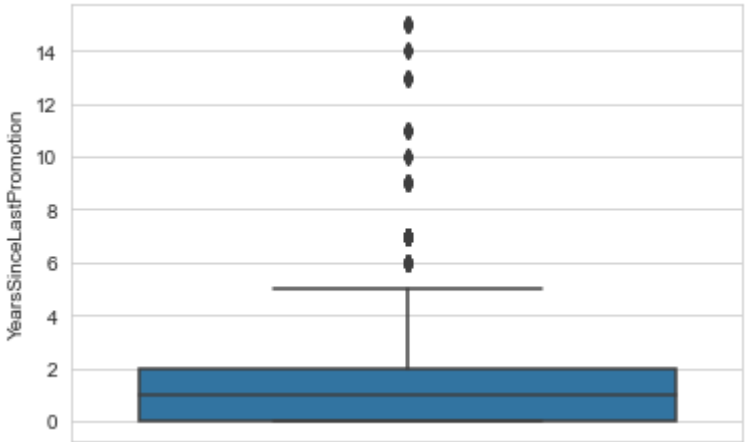


```
In [111]: stats.iqr(atr_yes['YearsAtCompany'])
```

Out[111]: 6.0

```
In [121]: sns.boxplot(y='YearsSinceLastPromotion',data=atr_yes)
```

Out[121]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4633129b0>



```
In [122]: stats.iqr(atr_yes['YearsSinceLastPromotion'])
```

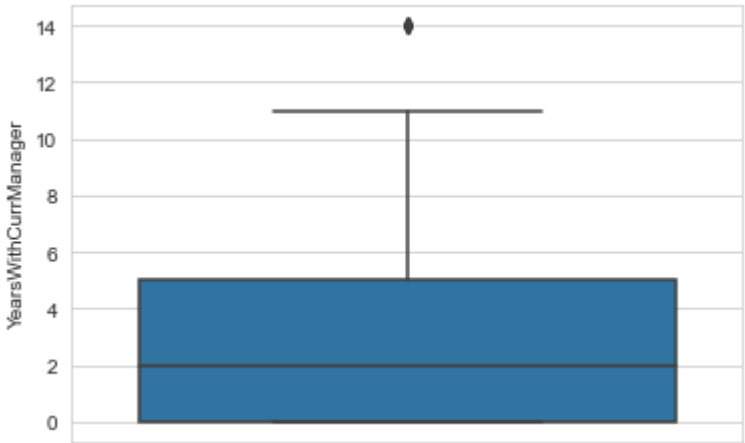
Out[122]: 2.0

```
In [129]: atr_yes['YearsSinceLastPromotion'].mean()
```

Out[129]: 1.9602836879432624


```
In [135]: sns.boxplot(y='YearsWithCurrManager',data=atr_yes)
```

Out[135]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4637410f0>



```
In [138]: stats.iqr(atr_yes['YearsWithCurrManager'])
```

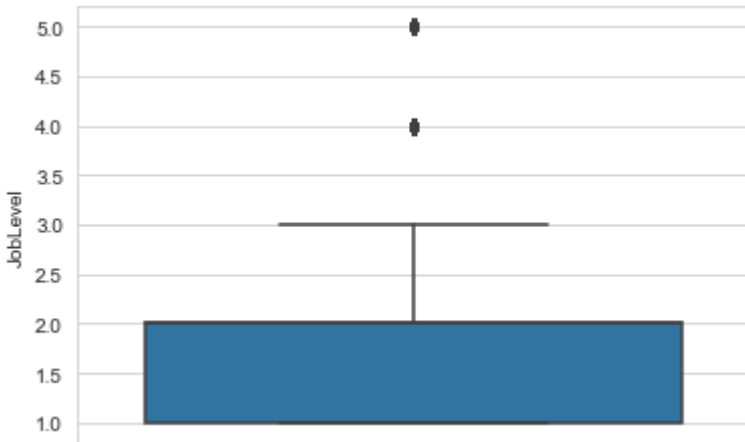
Out[138]: 5.0

```
In [139]: atr_yes['YearsWithCurrManager'].mean()
```

Out[139]: 2.8652482269503547

```
In [151]: sns.boxplot(y='JobLevel',data=atr_yes)
```

Out[151]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4632c3208>



```
In [152]: stats.iqr(atr_yes['JobLevel'])
```

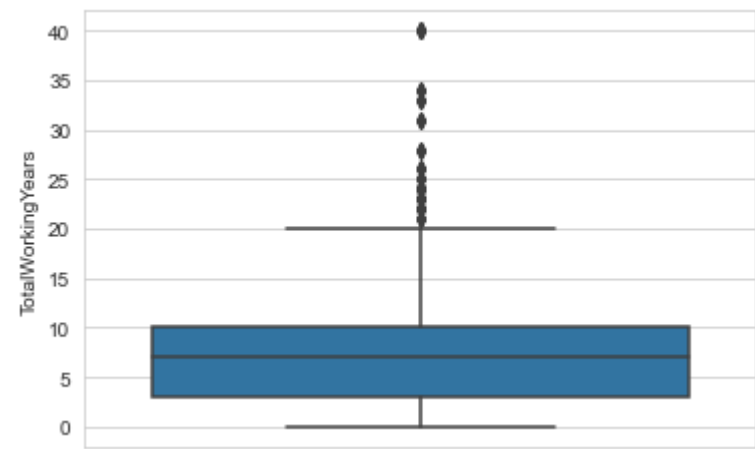
Out[152]: 1.0

```
In [153]: atr_yes['JobLevel'].mean()
```

Out[153]: 2.0326241134751775

```
In [165]: sns.boxplot(y='TotalWorkingYears',data=atr_yes)
```

Out[165]: <matplotlib.axes._subplots.AxesSubplot at 0x2a463d6c1d0>



```
In [166]: stats.iqr(atr_yes['TotalWorkingYears'])
```

Out[166]: 7.0

```
In [167]: atr_yes['TotalWorkingYears'].mean()
```

Out[167]: 8.273758865248228

```
In [140]: atr_yes.describe()
```

Out[140]:

	Age	DistanceFromHome	Education	EmployeeCount	EmployeeeID	JobLevel	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear
count	705.000000	705.000000	705.000000	705.0	705.000000	705.000000	705.000000	705.000000	705.000000	705.0	705.000000	705.000000	705.000000
mean	33.628369	9.024113	2.872340	1.0	2194.387234	2.032624	61814.950355	2.937589	15.487943	8.0	0.778723	8.273759	2.658156
std	9.678836	7.755184	1.014463	0.0	1270.408987	1.048707	44890.457122	2.681283	3.785842	0.0	0.857355	7.176762	1.155899
min	18.000000	1.000000	1.000000	1.0	2.000000	1.000000	10090.000000	0.000000	11.000000	8.0	0.000000	0.000000	0.000000
25%	28.000000	2.000000	2.000000	1.0	1075.000000	1.000000	28440.000000	1.000000	12.000000	8.0	0.000000	3.000000	2.000000
50%	32.000000	7.000000	3.000000	1.0	2165.000000	2.000000	49080.000000	1.000000	14.000000	8.0	1.000000	7.000000	3.000000
75%	39.000000	15.000000	4.000000	1.0	3343.000000	2.000000	71040.000000	5.000000	18.000000	8.0	1.000000	10.000000	3.000000
max	58.000000	29.000000	5.000000	1.0	4403.000000	5.000000	198590.000000	9.000000	25.000000	8.0	3.000000	40.000000	6.000000

```
In [144]: atr_yes.median()
```

```
Out[144]: Age                32.0
DistanceFromHome           7.0
Education                   3.0
EmployeeCount               1.0
EmployeeID                 2165.0
JobLevel                    2.0
MonthlyIncome              49080.0
NumCompaniesWorked          1.0
PercentSalaryHike           14.0
StandardHours               8.0
StockOptionLevel            1.0
TotalWorkingYears           7.0
TrainingTimesLastYear        3.0
YearsAtCompany              3.0
YearsSinceLastPromotion      1.0
YearsWithCurrManager         2.0
dtype: float64
```

```
In [145]: atr_yes.mean()
```

```
Out[145]: Age                33.628369
DistanceFromHome           9.024113
Education                   2.872340
EmployeeCount               1.000000
EmployeeID                 2194.387234
JobLevel                    2.032624
MonthlyIncome              61814.950355
NumCompaniesWorked          2.937589
PercentSalaryHike           15.487943
StandardHours               8.000000
StockOptionLevel            0.778723
TotalWorkingYears           8.273759
TrainingTimesLastYear        2.658156
YearsAtCompany              5.148936
YearsSinceLastPromotion      1.960284
YearsWithCurrManager         2.865248
dtype: float64
```

```
In [147]: atr_yes.skew()
```

```
Out[147]: Age                0.714479
DistanceFromHome          0.957983
Education                -0.217029
EmployeeCount             0.000000
EmployeeID                0.001186
JobLevel                  1.053776
MonthlyIncome             1.536183
NumCompaniesWorked        0.864889
PercentSalaryHike         0.760011
StandardHours             0.000000
StockOptionLevel          0.956491
TotalWorkingYears         1.674396
TrainingTimesLastYear     0.416718
YearsAtCompany            2.659829
YearsSinceLastPromotion   2.195541
YearsWithCurrManager      1.017582
dtype: float64
```

```
In [148]: atr_yes.kurt()
```

```
Out[148]: Age                -0.073388
DistanceFromHome          -0.050913
Education                -0.632511
EmployeeCount             0.000000
EmployeeID               -1.211662
JobLevel                  0.624243
MonthlyIncome             1.636083
NumCompaniesWorked       -0.552946
PercentSalaryHike        -0.432517
StandardHours             0.000000
StockOptionLevel          0.245353
TotalWorkingYears         3.673837
TrainingTimesLastYear     0.964115
YearsAtCompany            9.366724
YearsSinceLastPromotion   4.711054
YearsWithCurrManager      0.219179
dtype: float64
```

```
In [149]: atr_yes.var()
```

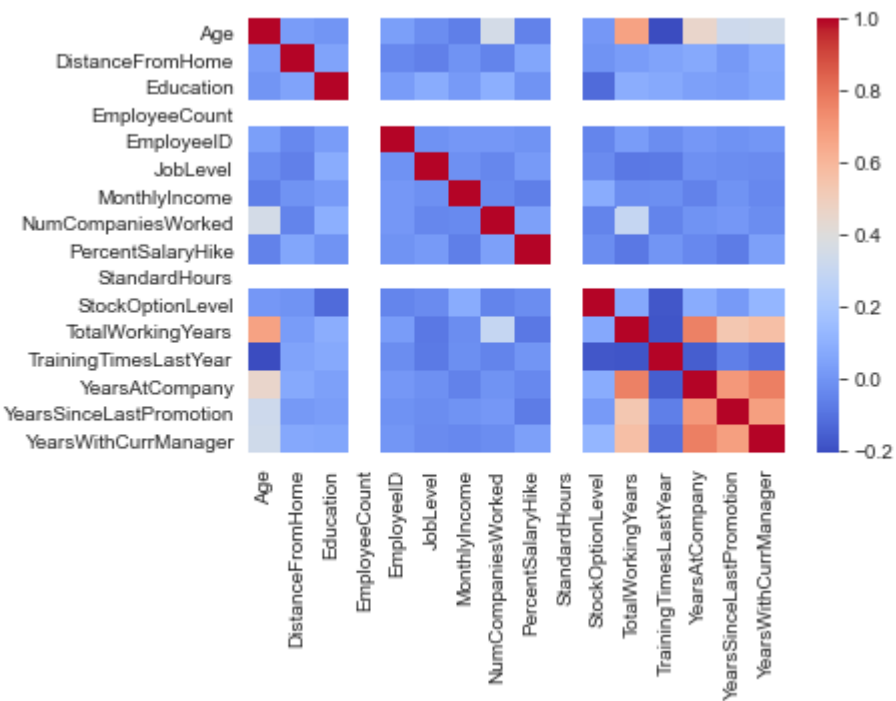
```
Out[149]: Age                9.367988e+01
DistanceFromHome          6.014288e+01
Education                 1.029134e+00
EmployeeCount             0.000000e+00
EmployeeID               1.613939e+06
JobLevel                 1.099786e+00
MonthlyIncome            2.015153e+09
NumCompaniesWorked        7.189281e+00
PercentSalaryHike         1.433260e+01
StandardHours             0.000000e+00
StockOptionLevel          7.350580e-01
TotalWorkingYears         5.150592e+01
TrainingTimesLastYear     1.336102e+00
YearsAtCompany            3.553318e+01
YearsSinceLastPromotion   9.970011e+00
YearsWithCurrManager      9.900850e+00
dtype: float64
```

```
In [150]: atr_yes.std()
```

```
Out[150]: Age                9.678836
DistanceFromHome           7.755184
Education                 1.014463
EmployeeCount             0.000000
EmployeeID              1270.408987
JobLevel                 1.048707
MonthlyIncome          44890.457122
NumCompaniesWorked        2.681283
PercentSalaryHike         3.785842
StandardHours             0.000000
StockOptionLevel          0.857355
TotalWorkingYears         7.176762
TrainingTimesLastYear     1.155899
YearsAtCompany            5.960972
YearsSinceLastPromotion   3.157532
YearsWithCurrManager      3.146562
dtype: float64
```

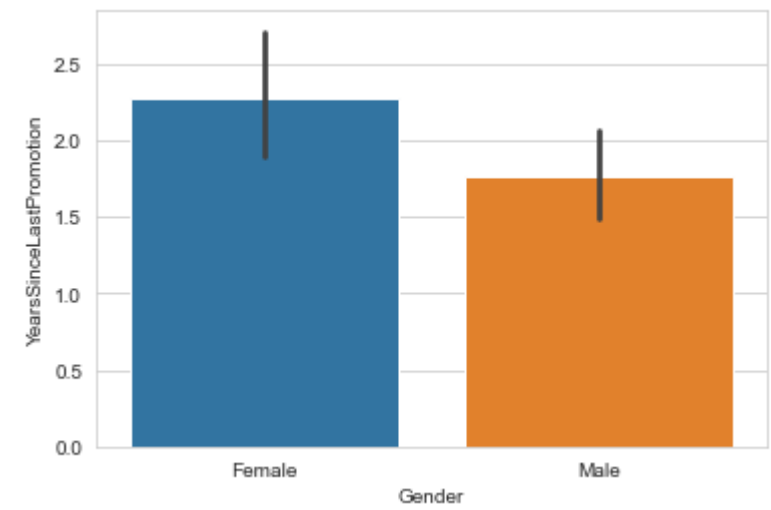
```
In [169]: sns.heatmap(atr_yes.corr(),cmap='coolwarm')
```

Out[169]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4652a46d8>



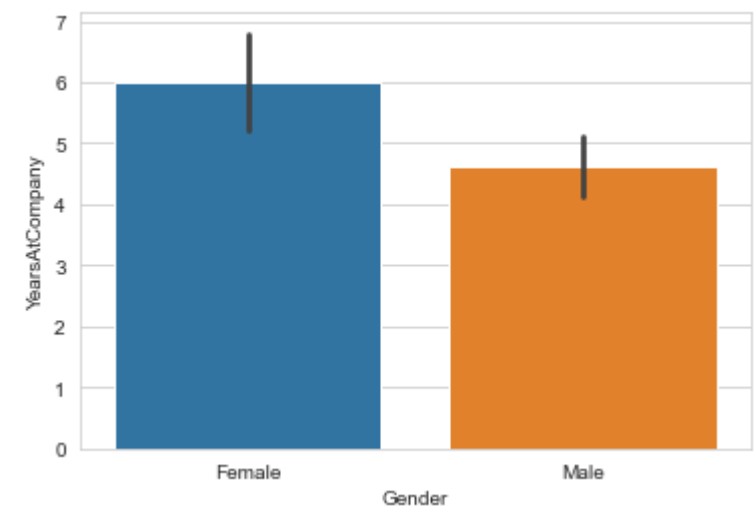
```
In [175]: sns.barplot(x='Gender',y='YearsSinceLastPromotion',data=atr_yes,estimator=np.mean)
```

Out[175]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4632a97b8>



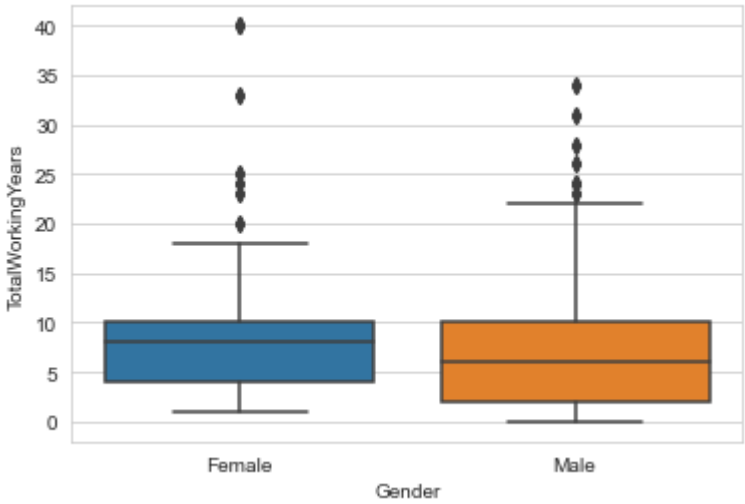
```
In [176]: sns.barplot(x='Gender',y='YearsAtCompany',data=atr_yes,estimator=np.mean)
```

Out[176]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4630d8f98>



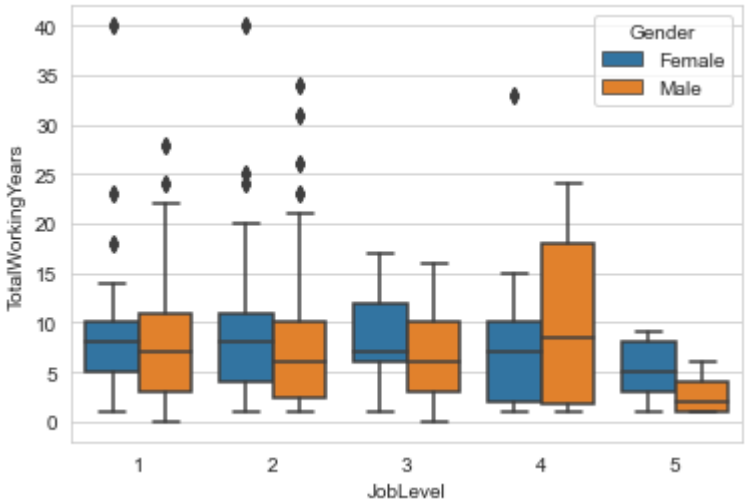
```
In [177]: sns.boxplot(x='Gender',y='TotalWorkingYears',data=atr_yes)
```

Out[177]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4635e6b70>



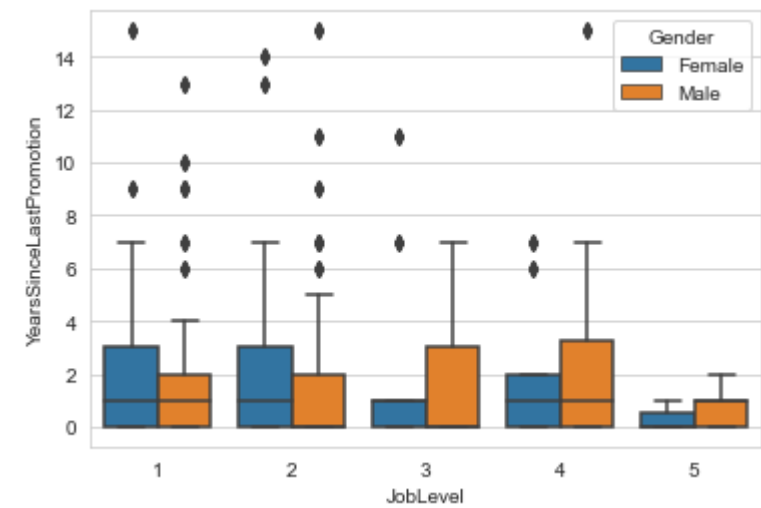
```
In [178]: sns.boxplot(x='JobLevel',y='TotalWorkingYears',data=atr_yes,hue='Gender')
```

Out[178]: <matplotlib.axes._subplots.AxesSubplot at 0x2a46329cb00>



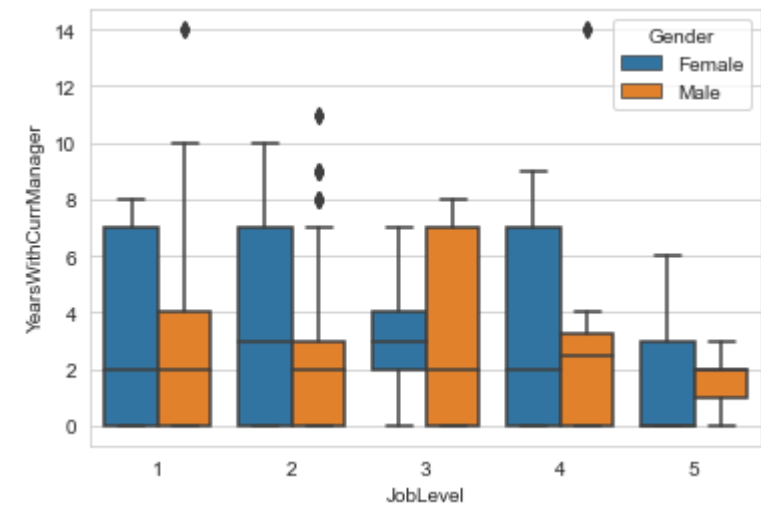

```
In [179]: sns.boxplot(x='JobLevel',y='YearsSinceLastPromotion',data=atr_yes,hue='Gender')
```

Out[179]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4631ed780>



```
In [180]: sns.boxplot(x='JobLevel',y='YearsWithCurrManager',data=atr_yes,hue='Gender')
```

Out[180]: <matplotlib.axes._subplots.AxesSubplot at 0x2a4656a7898>



In []: