In [80]:
```
import pandas as pd
import numpy as np
import matplotlib as plt
import matplotlib.pyplot as pl
import seaborn as sns
import math
```

In [81]:
```
data=pd.read_csv("C:/Users/Admin/Downloads/Attrition.csv")
```

In [82]:
```
data
```

Out[82]:

	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	Othe
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medica
5	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences
6	59	No	Travel_Rarely	1324	Research & Development	3	3	Medica

In [83]:
```
data.shape
```

Out[83]:
```
(1470, 35)
```

In [84]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
Age                         1470 non-null int64
Attrition                   1470 non-null object
BusinessTravel              1470 non-null object
DailyRate                   1470 non-null int64
Department                  1470 non-null object
DistanceFromHome            1470 non-null int64
Education                   1470 non-null int64
EducationField              1470 non-null object
EmployeeCount               1470 non-null int64
EmployeeNumber              1470 non-null int64
EnvironmentSatisfaction     1470 non-null int64
Gender                      1470 non-null object
HourlyRate                  1470 non-null int64
JobInvolvement              1470 non-null int64
JobLevel                    1470 non-null int64
JobRole                     1470 non-null object
JobSatisfaction             1470 non-null int64
MaritalStatus               1470 non-null object
MonthlyIncome               1470 non-null int64
MonthlyRate                 1470 non-null int64
NumCompaniesWorked          1470 non-null int64
Over18                      1470 non-null object
OverTime                    1470 non-null object
PercentSalaryHike           1470 non-null int64
PerformanceRating           1470 non-null int64
RelationshipSatisfaction    1470 non-null int64
StandardHours               1470 non-null int64
StockOptionLevel            1470 non-null int64
TotalWorkingYears           1470 non-null int64
TrainingTimesLastYear       1470 non-null int64
WorkLifeBalance             1470 non-null int64
YearsAtCompany              1470 non-null int64
YearsInCurrentRole          1470 non-null int64
YearsSinceLastPromotion     1470 non-null int64
YearsWithCurrManager        1470 non-null int64
dtypes: int64(26), object(9)
memory usage: 402.0+ KB
```

In [85]:
```
data.describe()
```

Out[85]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	Environn
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306	
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335	
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000	
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000	
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000	
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000	

8 rows × 26 columns

In [86]:
```
data.head()
```

Out[86]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Emp
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	

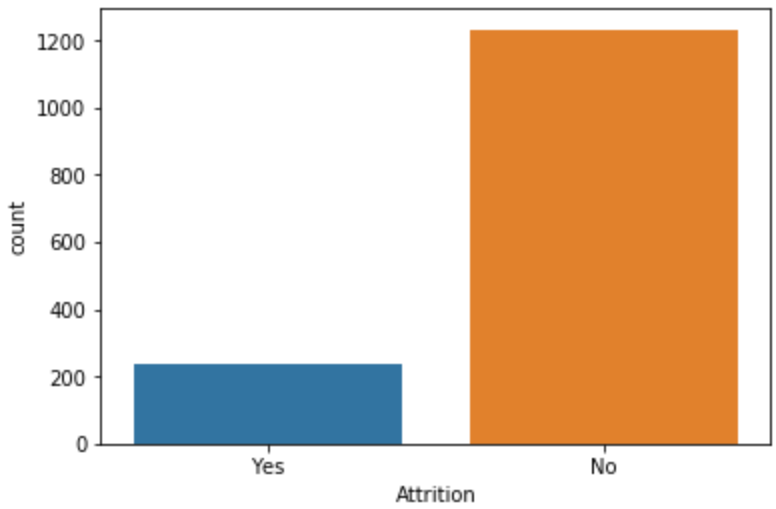
5 rows × 35 columns

In [87]:
```
data.isnull().sum()
```

Out[87]:
```
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
```

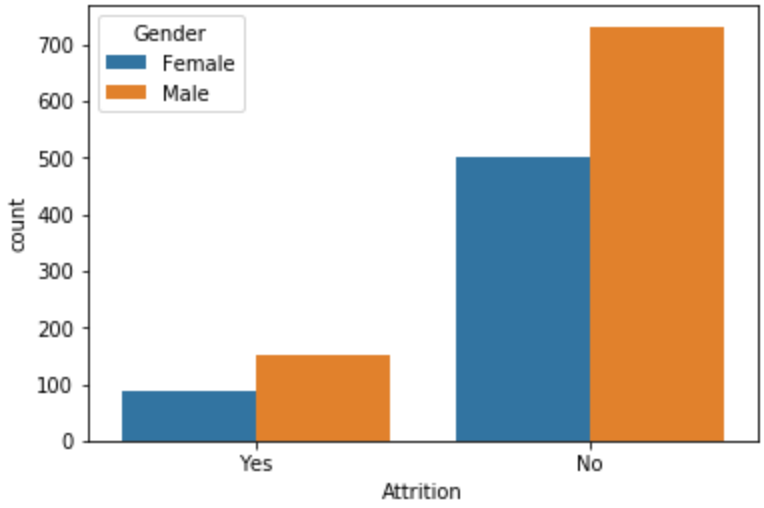
In [88]:
```
sns.countplot(x="Attrition",data=data)
```

Out[88]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xf3d34e0>
```



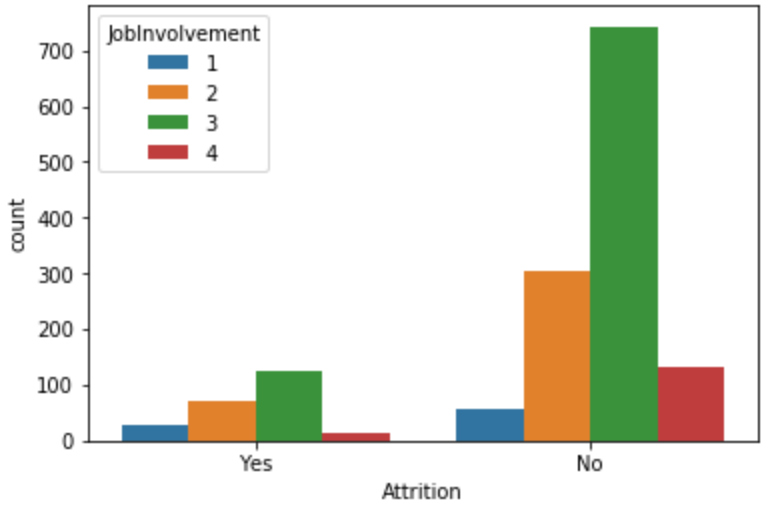
In [89]:
```
sns.countplot(x="Attrition",hue="Gender",data=data)
```

Out[89]: <matplotlib.axes._subplots.AxesSubplot at 0xf42ae10>



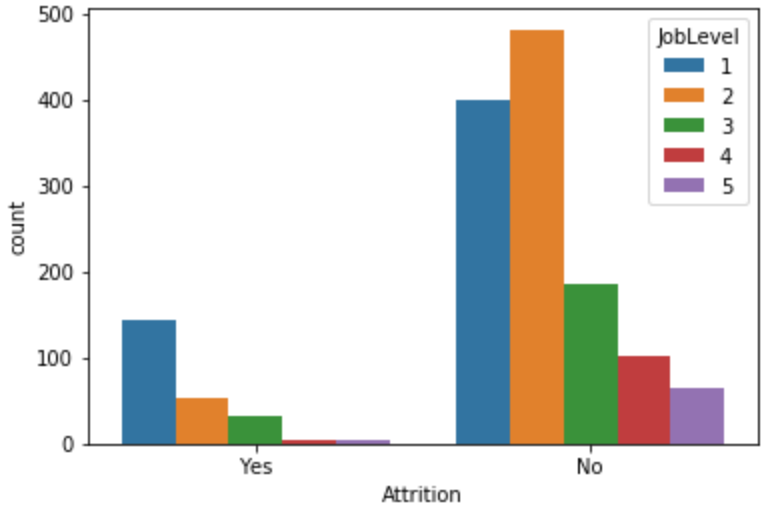
In [90]:
```
sns.countplot(x="Attrition",hue="JobInvolvement",data=data)
```

Out[90]: <matplotlib.axes._subplots.AxesSubplot at 0xf20e860>



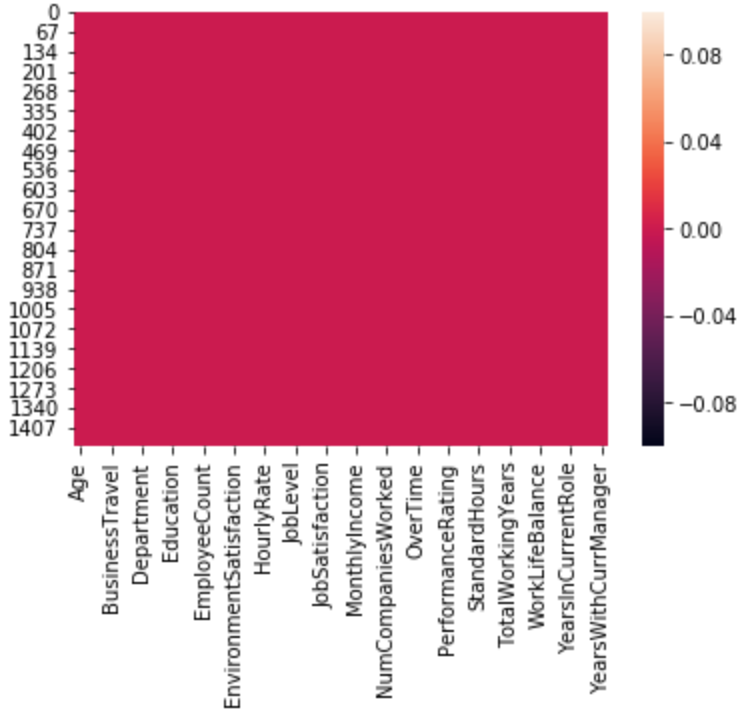
In [91]:
```
sns.countplot(x="Attrition",hue="JobLevel",data=data)
```

Out[91]: <matplotlib.axes._subplots.AxesSubplot at 0xf1e63c8>



In [92]:
```
sns.heatmap(data.isnull())
```

Out[92]: <matplotlib.axes._subplots.AxesSubplot at 0xf4dfb38>

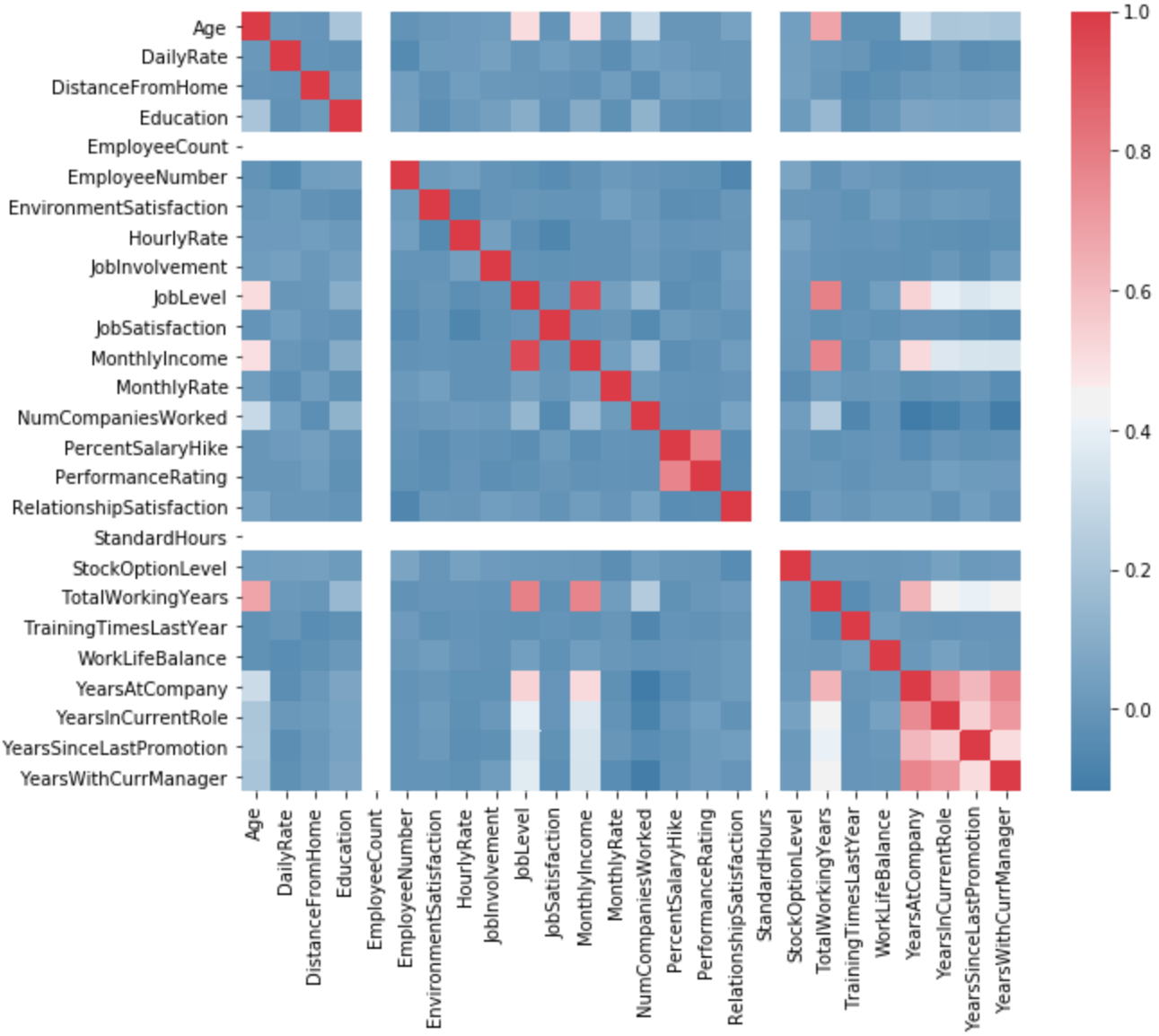


In [93]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
Age                         1470 non-null int64
Attrition                   1470 non-null object
BusinessTravel              1470 non-null object
DailyRate                   1470 non-null int64
Department                  1470 non-null object
DistanceFromHome            1470 non-null int64
Education                   1470 non-null int64
EducationField              1470 non-null object
EmployeeCount               1470 non-null int64
EmployeeNumber              1470 non-null int64
EnvironmentSatisfaction     1470 non-null int64
Gender                      1470 non-null object
HourlyRate                  1470 non-null int64
JobInvolvement              1470 non-null int64
JobLevel                    1470 non-null int64
JobRole                     1470 non-null object
JobSatisfaction             1470 non-null int64
MaritalStatus               1470 non-null object
MonthlyIncome               1470 non-null int64
MonthlyRate                 1470 non-null int64
NumCompaniesWorked          1470 non-null int64
Over18                      1470 non-null object
OverTime                    1470 non-null object
PercentSalaryHike           1470 non-null int64
PerformanceRating           1470 non-null int64
RelationshipSatisfaction    1470 non-null int64
StandardHours               1470 non-null int64
StockOptionLevel            1470 non-null int64
TotalWorkingYears           1470 non-null int64
TrainingTimesLastYear       1470 non-null int64
WorkLifeBalance             1470 non-null int64
YearsAtCompany              1470 non-null int64
YearsInCurrentRole          1470 non-null int64
YearsSinceLastPromotion     1470 non-null int64
YearsWithCurrManager        1470 non-null int64
dtypes: int64(26), object(9)
memory usage: 402.0+ KB
```

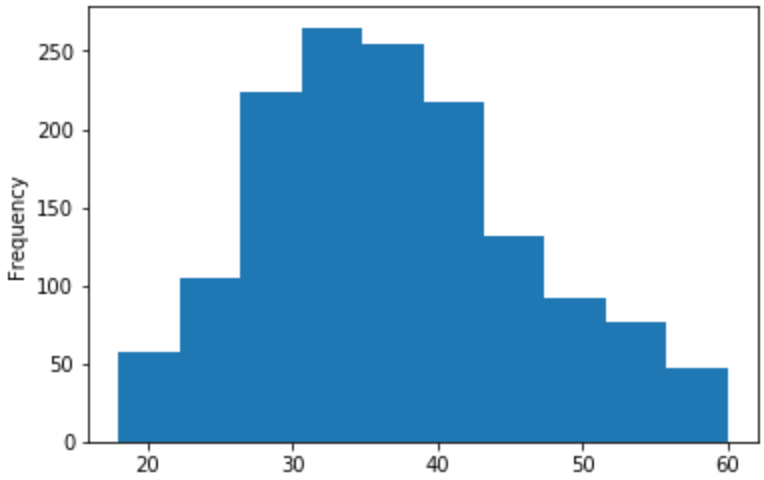
In [94]:
```
f, ax=pl.subplots(figsize=(10,8))
corr=data.corr()
sns.heatmap(corr,mask=np.zeros_like(corr,dtype=np.bool),cmap=sns.diverging_palette(240,10,
```

Out[94]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xf770978>
```



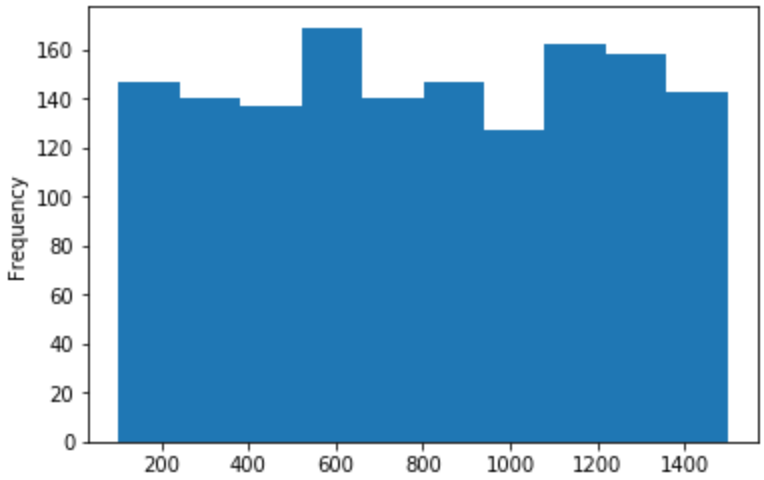
In [95]:
```
data['Age'].plot.hist()
```

Out[95]: <matplotlib.axes._subplots.AxesSubplot at 0xfa18cf8>



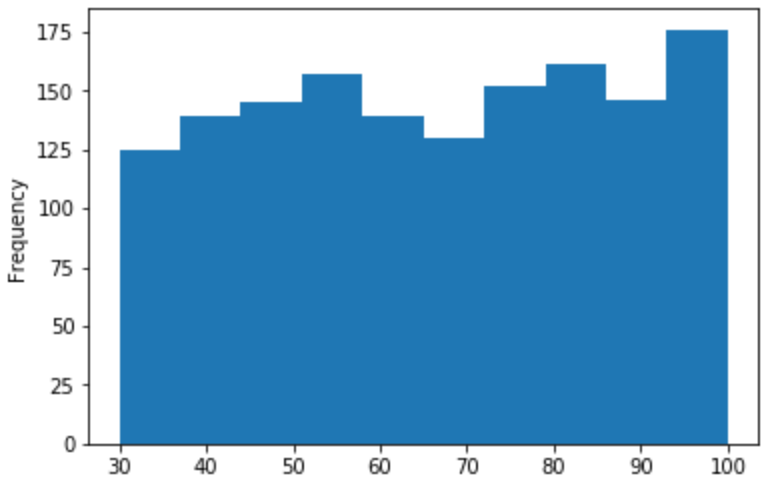
In [96]:
```
data['DailyRate'].plot.hist()
```

Out[96]: <matplotlib.axes._subplots.AxesSubplot at 0xf847c18>



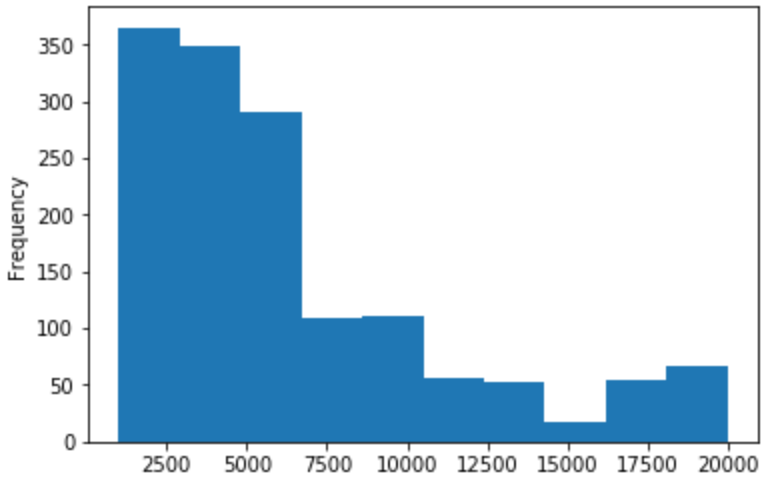
In [97]:
```
data['HourlyRate'].plot.hist()
```

Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0xf8f8898>



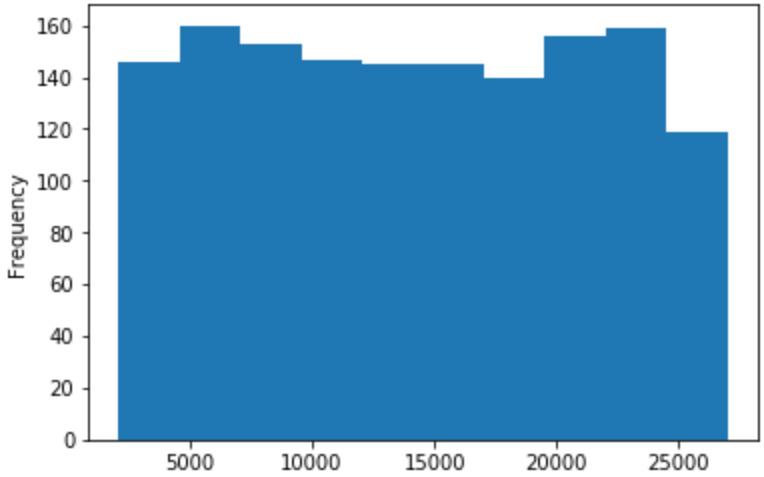
In [98]:
```
data['MonthlyIncome'].plot.hist()
```

Out[98]: <matplotlib.axes._subplots.AxesSubplot at 0xf962ef0>



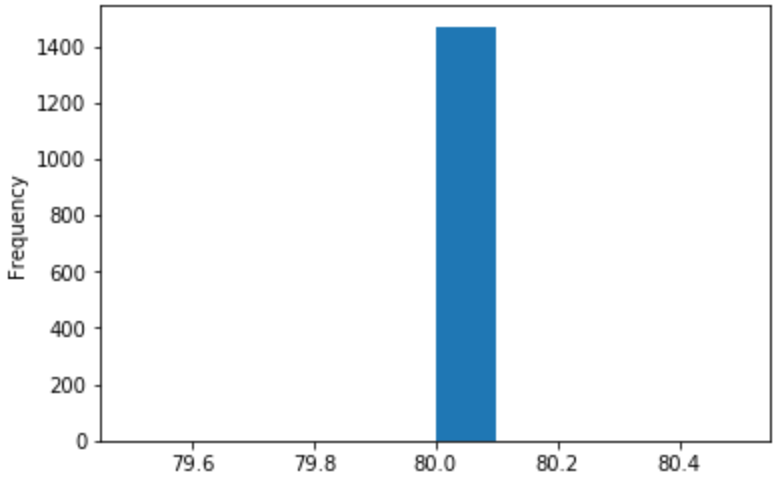
In [99]:
```
data['MonthlyRate'].plot.hist()
```

Out[99]: <matplotlib.axes._subplots.AxesSubplot at 0xfc0f780>



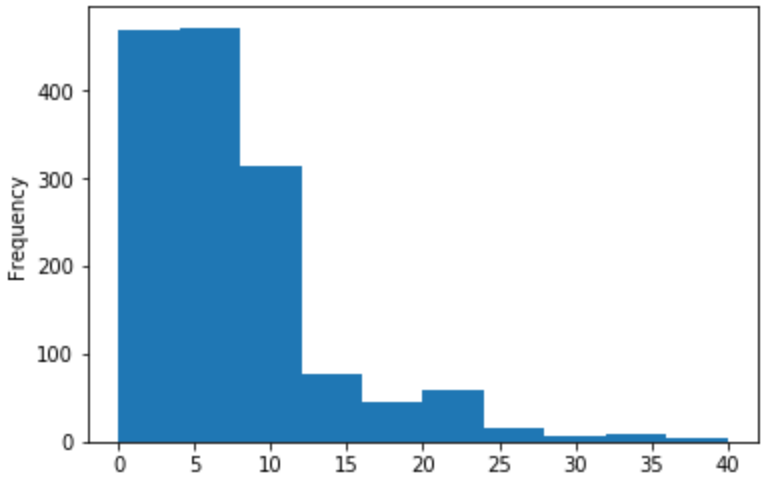
In [100]:
```
data['StandardHours'].plot.hist()
```

Out[100]: <matplotlib.axes._subplots.AxesSubplot at 0xfc76828>



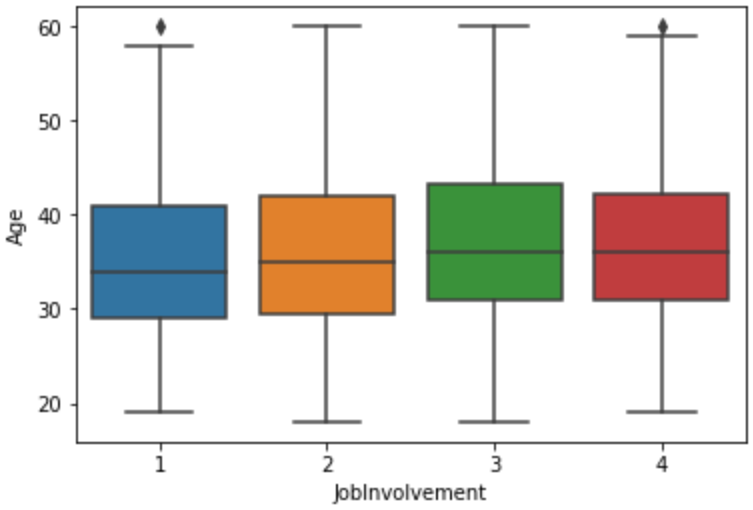
In [101]:
```
data['YearsAtCompany'].plot.hist()
```

Out[101]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xfcd4400>
```



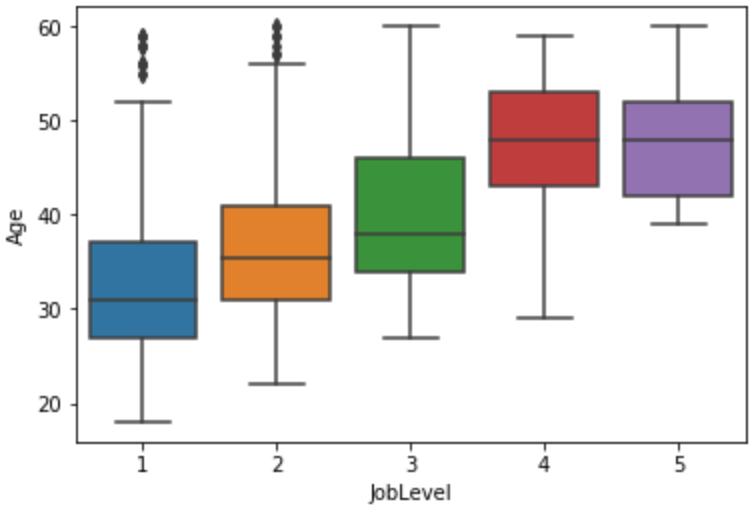
In [102]:
```
sns.boxplot(x="JobInvolvement",y="Age",data=data)
```

Out[102]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xfd4cda0>
```



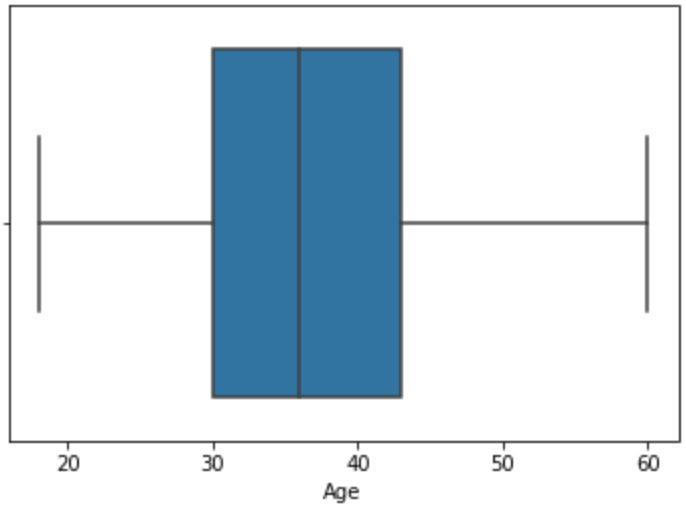
In [103]:
```
sns.boxplot(x="JobLevel",y="Age",data=data)
```

Out[103]: <matplotlib.axes._subplots.AxesSubplot at 0x10db70b8>



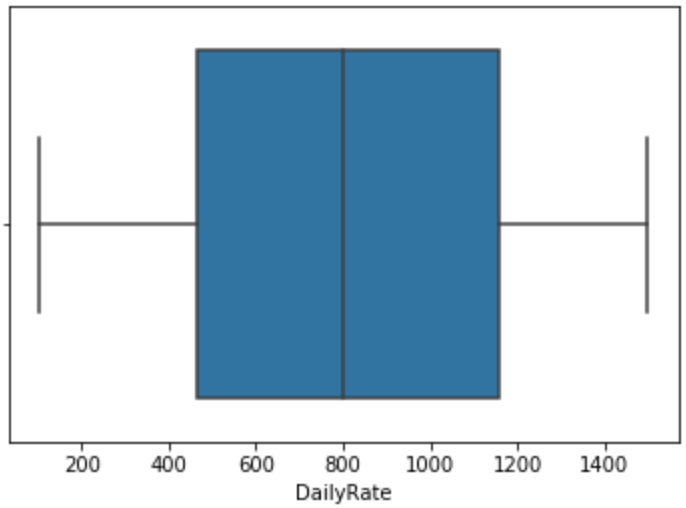
In [104]:
```
sns.boxplot(x="Age",data=data)
```

Out[104]: <matplotlib.axes._subplots.AxesSubplot at 0x10e492e8>



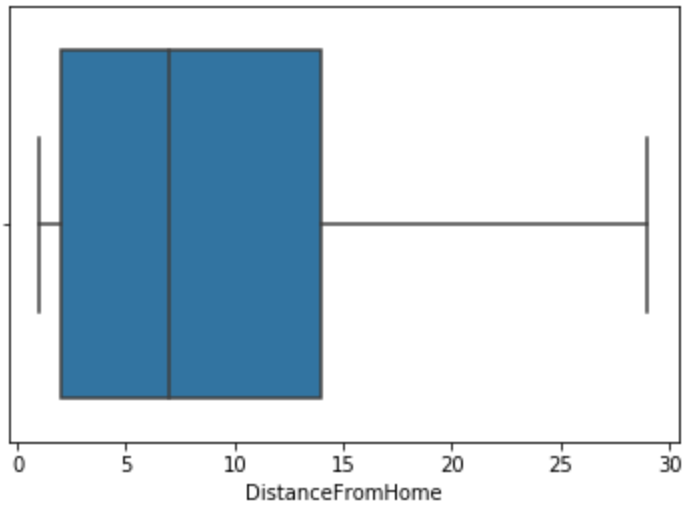
In [105]:
```
sns.boxplot(x="DailyRate",data=data)
```

Out[105]: <matplotlib.axes._subplots.AxesSubplot at 0x10e75128>



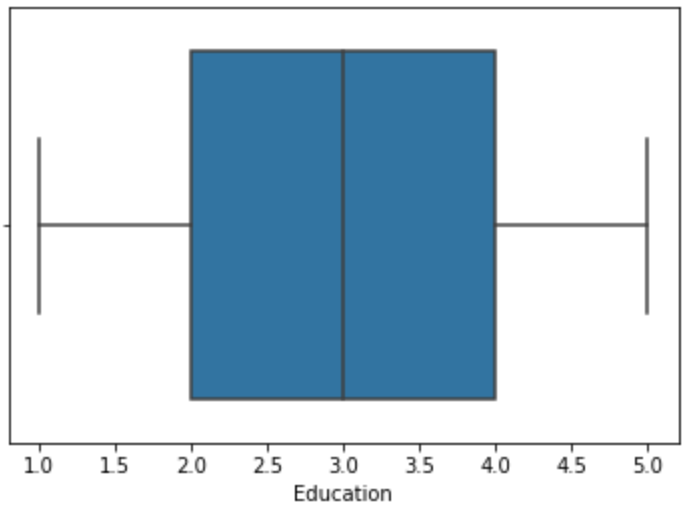
In [106]:
```
sns.boxplot(x="DistanceFromHome",data=data)
```

Out[106]: <matplotlib.axes._subplots.AxesSubplot at 0x10ec4908>



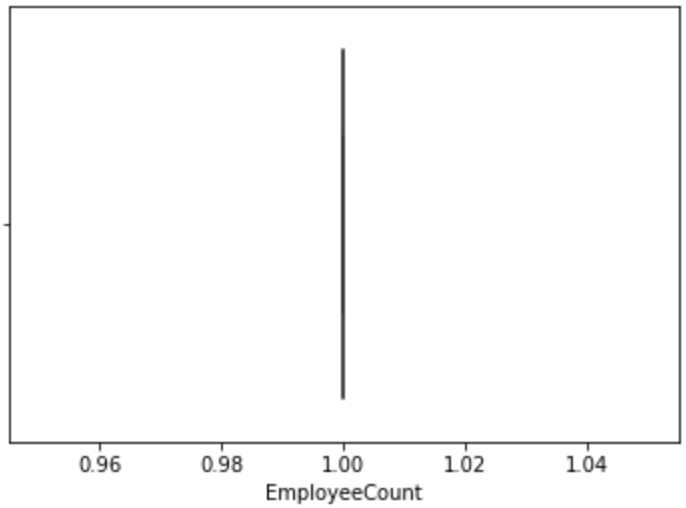
In [107]:
```
sns.boxplot(x="Education",data=data)
```

Out[107]: <matplotlib.axes._subplots.AxesSubplot at 0x10f1df98>



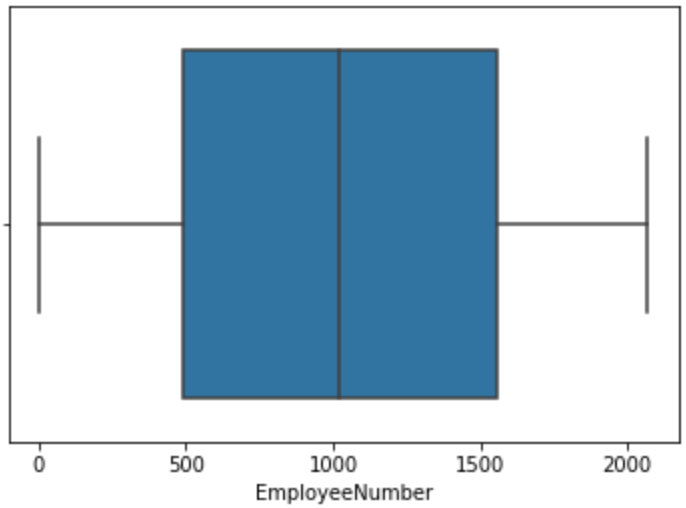
In [108]:
```
sns.boxplot(x="EmployeeCount",data=data)
```

Out[108]: <matplotlib.axes._subplots.AxesSubplot at 0x10f69160>



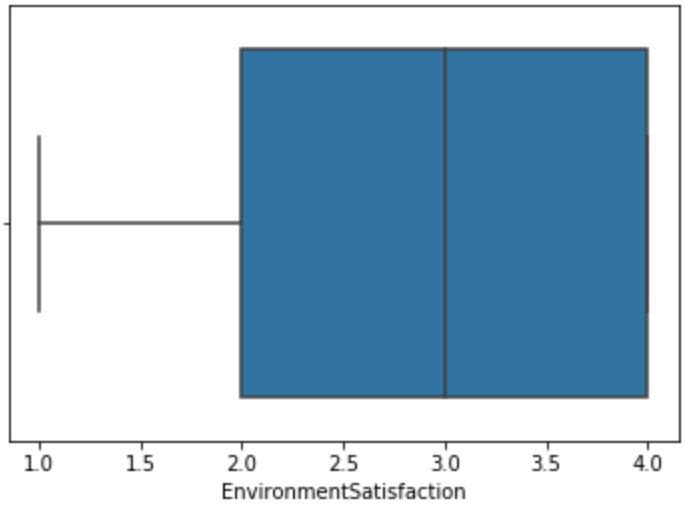
In [109]:
```
sns.boxplot(x="EmployeeNumber",data=data)
```

Out[109]: <matplotlib.axes._subplots.AxesSubplot at 0x10fd8ac8>



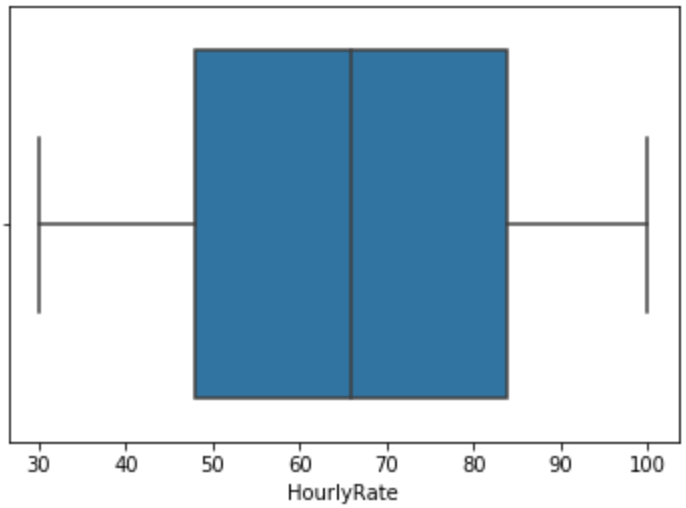
In [110]:
```
sns.boxplot(x="EnvironmentSatisfaction",data=data)
```

Out[110]: <matplotlib.axes._subplots.AxesSubplot at 0x110299b0>



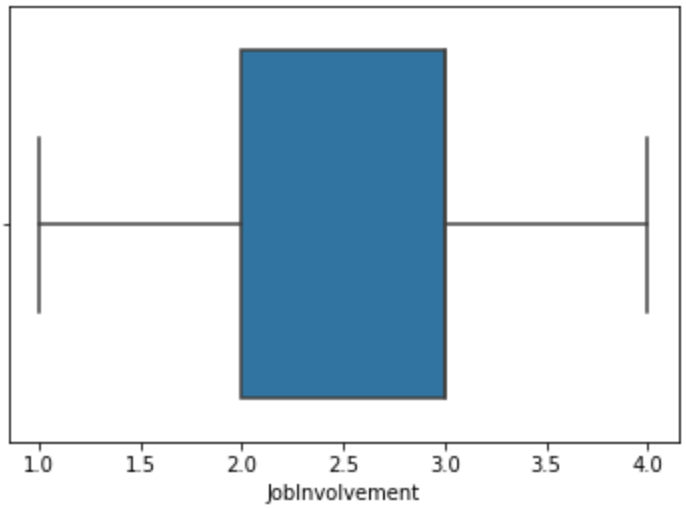
In [111]:
```
sns.boxplot(x="HourlyRate",data=data)
```

Out[111]: <matplotlib.axes._subplots.AxesSubplot at 0x1107a7b8>



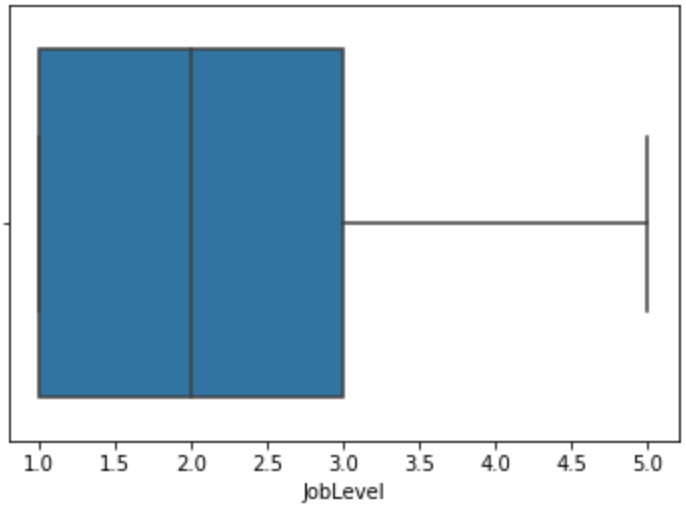
In [112]:
```
sns.boxplot(x="JobInvolvement",data=data)
```

Out[112]: <matplotlib.axes._subplots.AxesSubplot at 0x110d57b8>



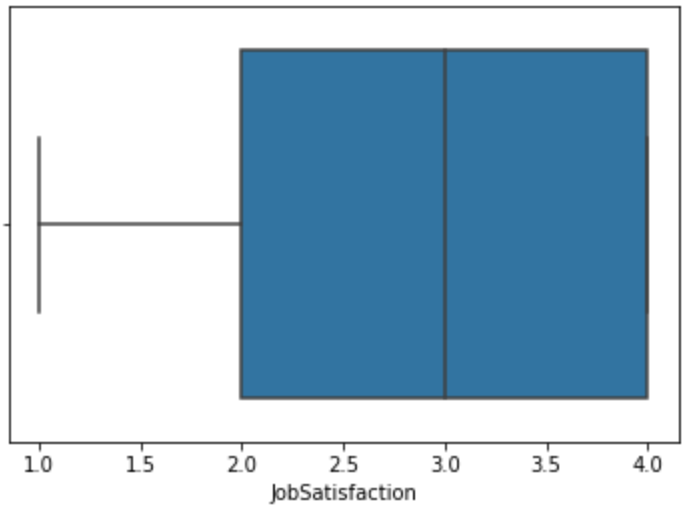
In [113]:
```
sns.boxplot(x="JobLevel",data=data)
```

Out[113]: <matplotlib.axes._subplots.AxesSubplot at 0x11132160>



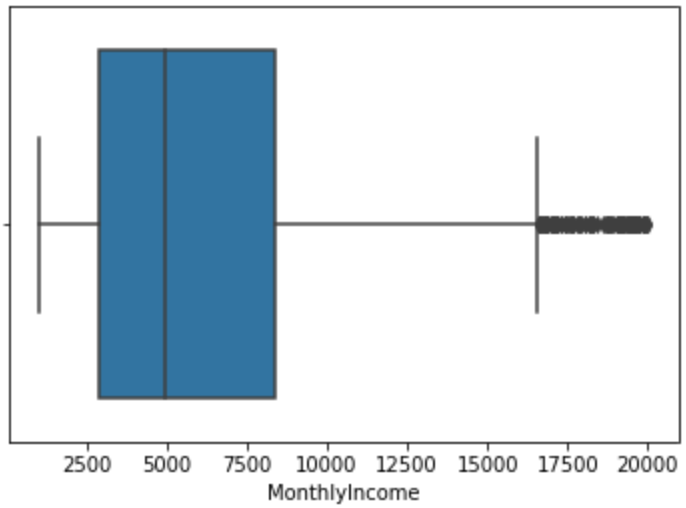
In [114]:
```
sns.boxplot(x="JobSatisfaction",data=data)
```

Out[114]: <matplotlib.axes._subplots.AxesSubplot at 0x1107a1d0>



In [115]:
```
sns.boxplot(x="MonthlyIncome",data=data)
```

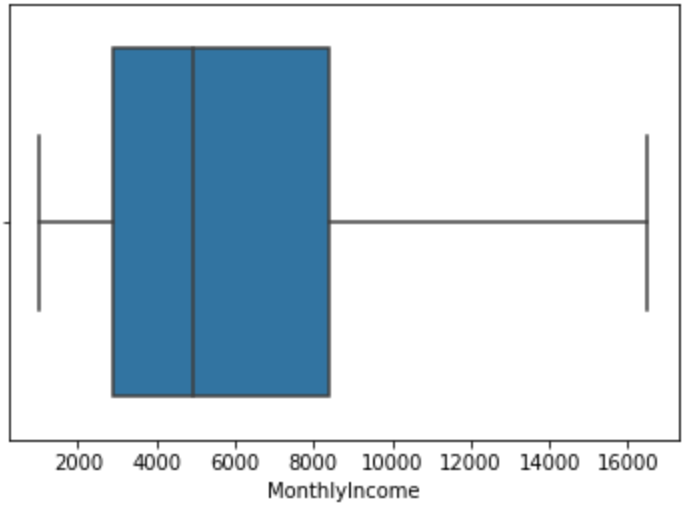
Out[115]: <matplotlib.axes._subplots.AxesSubplot at 0x111e8be0>



In [116]:
```
data['MonthlyIncome']=np.where(data['MonthlyIncome']>16500,16500,data['MonthlyIncome'])
```

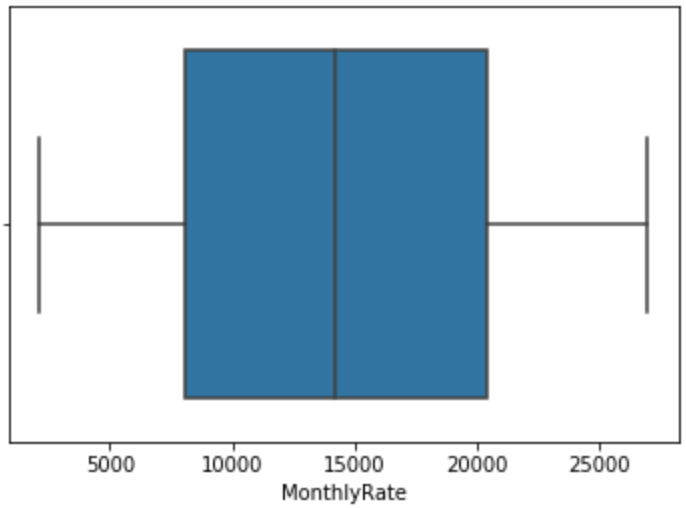
In [117]:
```
sns.boxplot(data['MonthlyIncome'])
```

Out[117]: <matplotlib.axes._subplots.AxesSubplot at 0xf3d3d68>



In [118]:
```
sns.boxplot(x="MonthlyRate",data=data)
```

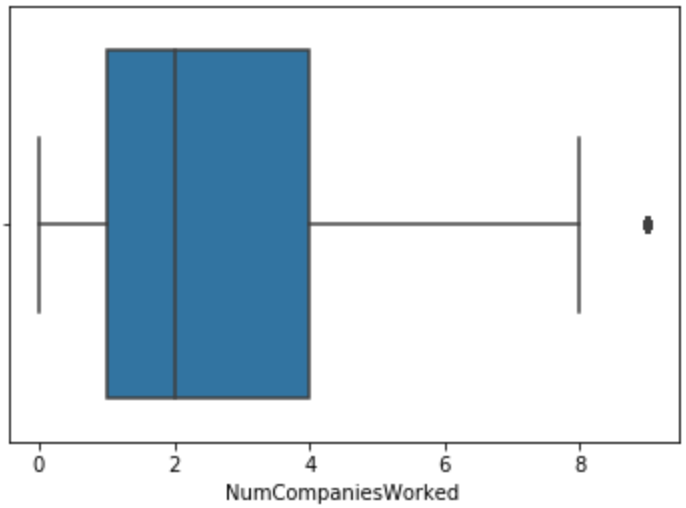
Out[118]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1129f6a0>
```



Type *Markdown* and LaTeX: α^2

In [119]:
```
sns.boxplot(x="NumCompaniesWorked",data=data)
```

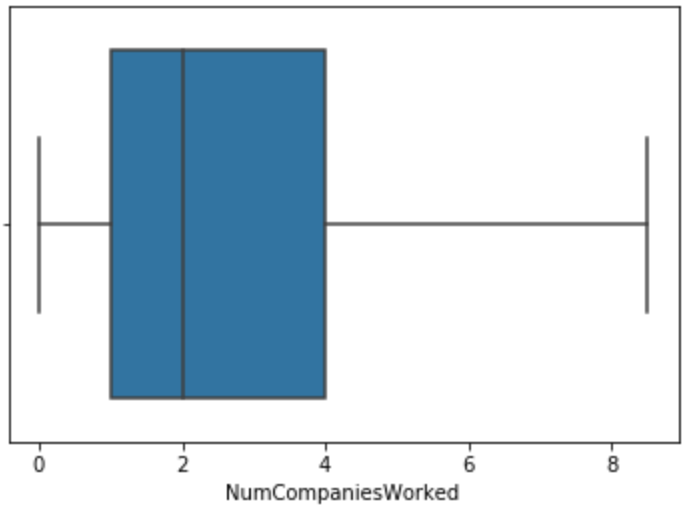
Out[119]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x112f2630>
```



In [120]:
```
data['NumCompaniesWorked']=np.where(data['NumCompaniesWorked']>8.5,8.5,data['NumCompaniesW
```

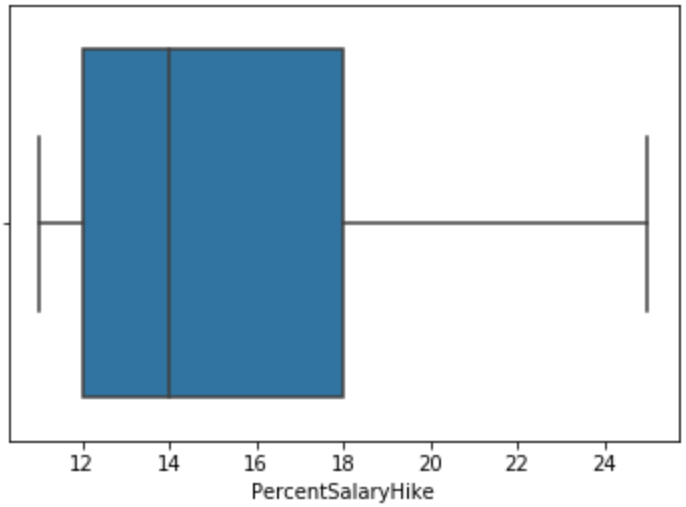
In [121]:
```
sns.boxplot(data['NumCompaniesWorked'])
```

Out[121]: <matplotlib.axes._subplots.AxesSubplot at 0x11345a58>



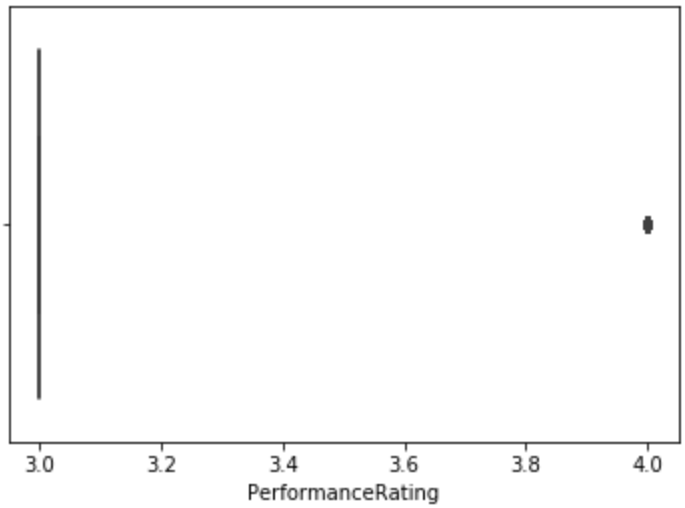
In [122]:
```
sns.boxplot(x="PercentSalaryHike",data=data)
```

Out[122]: <matplotlib.axes._subplots.AxesSubplot at 0x11399748>



In [123]:
```
sns.boxplot(x="PerformanceRating",data=data)
```

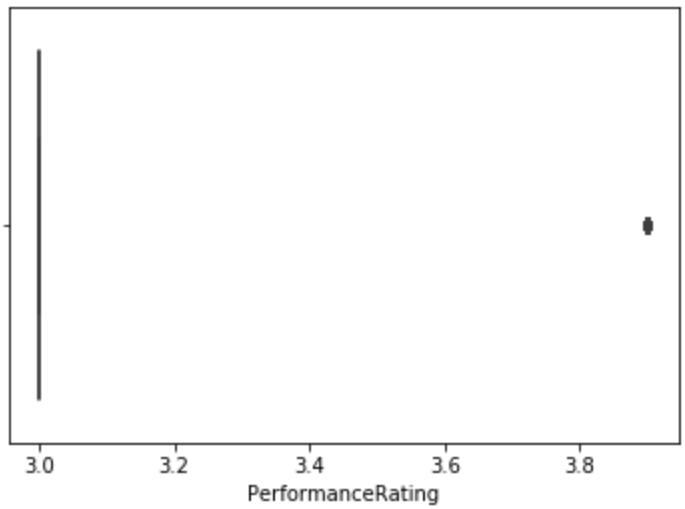
Out[123]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x113f5a58>
```



In [124]:
```
data['PerformanceRating']=np.where(data['PerformanceRating']>3.9,3.9,data['PerformanceRati
```

In [125]:
```
sns.boxplot(data['PerformanceRating'])
```

Out[125]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1144a400>
```



In [126]:
```
data=data.drop(["EmployeeNumber"],axis=1)
```

In [127]:
```
data
```

Out[127]:

	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	Othe
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medica
5	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences
6	59	No	Travel_Rarely	1324	Research & Development	3	3	Medica

In [128]:
```
data=data.drop(["Department"],axis=1)
```

In [129]:
```
data
```

Out[129]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EducationField	EmployeeCour
0	41	Yes	Travel_Rarely	1102	1	2	Life Sciences	
1	49	No	Travel_Frequently	279	8	1	Life Sciences	
2	37	Yes	Travel_Rarely	1373	2	2	Other	
3	33	No	Travel_Frequently	1392	3	4	Life Sciences	
4	27	No	Travel_Rarely	591	2	1	Medical	
5	32	No	Travel_Frequently	1005	2	2	Life Sciences	
6	59	No	Travel_Rarely	1324	3	3	Medical	
7	30	No	Travel_Rarely	1358	24	1	Life Sciences	
8	38	No	Travel_Frequently	216	23	3	Life Sciences	
9	36	No	Travel_Rarely	1299	27	3	Medical	
10	35	No	Travel_Rarely	809	16	3	Medical	
11	29	No	Travel_Rarely	153	15	2	Life Sciences	
12	31	No	Travel_Rarely	670	26	1	Life Sciences	
13	34	No	Travel_Rarely	1346	19	2	Medical	
14	28	Yes	Travel_Rarely	103	24	3	Life Sciences	
15	29	No	Travel_Rarely	1389	21	4	Life Sciences	
16	32	No	Travel_Rarely	334	5	2	Life Sciences	
17	22	No	Non-Travel	1123	16	2	Medical	
18	53	No	Travel_Rarely	1219	2	4	Life Sciences	
19	38	No	Travel_Rarely	371	2	3	Life Sciences	
20	24	No	Non-Travel	673	11	2	Other	
21	36	Yes	Travel_Rarely	1218	9	4	Life Sciences	
22	34	No	Travel_Rarely	419	7	4	Life Sciences	
23	21	No	Travel_Rarely	391	15	2	Life Sciences	
24	34	Yes	Travel_Rarely	699	6	1	Medical	
25	53	No	Travel_Rarely	1282	5	3	Other	
26	32	Yes	Travel_Frequently	1125	16	1	Life Sciences	
27	42	No	Travel_Rarely	691	8	4	Marketing	
28	44	No	Travel_Rarely	477	7	4	Medical	
29	46	No	Travel_Rarely	705	2	4	Marketing	
...	...	...	...	...	...	...	...	.
1440	36	No	Travel_Frequently	688	4	2	Life Sciences	
1441	56	No	Non-Travel	667	1	4	Life Sciences	
1442	29	Yes	Travel_Rarely	1092	1	4	Medical	
1443	42	No	Travel_Rarely	300	2	3	Life Sciences	
1444	56	Yes	Travel_Rarely	310	7	2	Technical Degree	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EducationField	EmployeeCour
1445	41	No	Travel_Rarely	582	28	4	Life Sciences	
1446	34	No	Travel_Rarely	704	28	3	Marketing	
1447	36	No	Non-Travel	301	15	4	Marketing	
1448	41	No	Travel_Rarely	930	3	3	Life Sciences	
1449	32	No	Travel_Rarely	529	2	3	Technical Degree	
1450	35	No	Travel_Rarely	1146	26	4	Life Sciences	
1451	38	No	Travel_Rarely	345	10	2	Life Sciences	
1452	50	Yes	Travel_Frequently	878	1	4	Life Sciences	
1453	36	No	Travel_Rarely	1120	11	4	Marketing	
1454	45	No	Travel_Rarely	374	20	3	Life Sciences	
1455	40	No	Travel_Rarely	1322	2	4	Life Sciences	
1456	35	No	Travel_Frequently	1199	18	4	Life Sciences	
1457	40	No	Travel_Rarely	1194	2	4	Medical	
1458	35	No	Travel_Rarely	287	1	4	Life Sciences	
1459	29	No	Travel_Rarely	1378	13	2	Other	
1460	29	No	Travel_Rarely	468	28	4	Medical	
1461	50	Yes	Travel_Rarely	410	28	3	Marketing	
1462	39	No	Travel_Rarely	722	24	1	Marketing	
1463	31	No	Non-Travel	325	5	3	Medical	
1464	26	No	Travel_Rarely	1167	5	3	Other	
1465	36	No	Travel_Frequently	884	23	2	Medical	
1466	39	No	Travel_Rarely	613	6	1	Medical	
1467	27	No	Travel_Rarely	155	4	3	Life Sciences	
1468	49	No	Travel_Frequently	1023	2	3	Medical	
1469	34	No	Travel_Rarely	628	8	3	Medical	

1470 rows × 33 columns

In [130]:
```
data=data.drop(["Gender"],axis=1)
```

In [131]:
```
data
```

Out[131]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EducationField	EmployeeCour
0	41	Yes	Travel_Rarely	1102	1	2	Life Sciences	
1	49	No	Travel_Frequently	279	8	1	Life Sciences	
2	37	Yes	Travel_Rarely	1373	2	2	Other	
3	33	No	Travel_Frequently	1392	3	4	Life Sciences	
4	27	No	Travel_Rarely	591	2	1	Medical	
5	32	No	Travel_Frequently	1005	2	2	Life Sciences	
6	59	No	Travel_Rarely	1324	3	3	Medical	
7	30	No	Travel_Rarely	1358	24	1	Life Sciences	
8	38	No	Travel_Frequently	216	23	3	Life Sciences	
9	36	No	Travel_Rarely	1299	27	3	Medical	
10	35	No	Travel_Rarely	809	16	3	Medical	
11	29	No	Travel_Rarely	153	15	2	Life Sciences	
12	31	No	Travel_Rarely	670	26	1	Life Sciences	
13	34	No	Travel_Rarely	1346	19	2	Medical	
14	28	Yes	Travel_Rarely	103	24	3	Life Sciences	
15	29	No	Travel_Rarely	1389	21	4	Life Sciences	
16	32	No	Travel_Rarely	334	5	2	Life Sciences	
17	22	No	Non-Travel	1123	16	2	Medical	
18	53	No	Travel_Rarely	1219	2	4	Life Sciences	
19	38	No	Travel_Rarely	371	2	3	Life Sciences	
20	24	No	Non-Travel	673	11	2	Other	
21	36	Yes	Travel_Rarely	1218	9	4	Life Sciences	
22	34	No	Travel_Rarely	419	7	4	Life Sciences	
23	21	No	Travel_Rarely	391	15	2	Life Sciences	
24	34	Yes	Travel_Rarely	699	6	1	Medical	
25	53	No	Travel_Rarely	1282	5	3	Other	
26	32	Yes	Travel_Frequently	1125	16	1	Life Sciences	
27	42	No	Travel_Rarely	691	8	4	Marketing	
28	44	No	Travel_Rarely	477	7	4	Medical	
29	46	No	Travel_Rarely	705	2	4	Marketing	
...	...	...	...	...	...	...	...	.
1440	36	No	Travel_Frequently	688	4	2	Life Sciences	
1441	56	No	Non-Travel	667	1	4	Life Sciences	
1442	29	Yes	Travel_Rarely	1092	1	4	Medical	
1443	42	No	Travel_Rarely	300	2	3	Life Sciences	
1444	56	Yes	Travel_Rarely	310	7	2	Technical Degree	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EducationField	EmployeeCour
1445	41	No	Travel_Rarely	582	28	4	Life Sciences	
1446	34	No	Travel_Rarely	704	28	3	Marketing	
1447	36	No	Non-Travel	301	15	4	Marketing	
1448	41	No	Travel_Rarely	930	3	3	Life Sciences	
1449	32	No	Travel_Rarely	529	2	3	Technical Degree	
1450	35	No	Travel_Rarely	1146	26	4	Life Sciences	
1451	38	No	Travel_Rarely	345	10	2	Life Sciences	
1452	50	Yes	Travel_Frequently	878	1	4	Life Sciences	
1453	36	No	Travel_Rarely	1120	11	4	Marketing	
1454	45	No	Travel_Rarely	374	20	3	Life Sciences	
1455	40	No	Travel_Rarely	1322	2	4	Life Sciences	
1456	35	No	Travel_Frequently	1199	18	4	Life Sciences	
1457	40	No	Travel_Rarely	1194	2	4	Medical	
1458	35	No	Travel_Rarely	287	1	4	Life Sciences	
1459	29	No	Travel_Rarely	1378	13	2	Other	
1460	29	No	Travel_Rarely	468	28	4	Medical	
1461	50	Yes	Travel_Rarely	410	28	3	Marketing	
1462	39	No	Travel_Rarely	722	24	1	Marketing	
1463	31	No	Non-Travel	325	5	3	Medical	
1464	26	No	Travel_Rarely	1167	5	3	Other	
1465	36	No	Travel_Frequently	884	23	2	Medical	
1466	39	No	Travel_Rarely	613	6	1	Medical	
1467	27	No	Travel_Rarely	155	4	3	Life Sciences	
1468	49	No	Travel_Frequently	1023	2	3	Medical	
1469	34	No	Travel_Rarely	628	8	3	Medical	

1470 rows × 32 columns

In [132]:
```
data=data.drop(["EducationField"],axis=1)
```

In [133]:
```
data
```

Out[133]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EmployeeCount	EnvironmentS
0	41	Yes	Travel_Rarely	1102	1	2	1	
1	49	No	Travel_Frequently	279	8	1	1	
2	37	Yes	Travel_Rarely	1373	2	2	1	
3	33	No	Travel_Frequently	1392	3	4	1	
4	27	No	Travel_Rarely	591	2	1	1	
5	32	No	Travel_Frequently	1005	2	2	1	
6	59	No	Travel_Rarely	1324	3	3	1	
7	30	No	Travel_Rarely	1358	24	1	1	
8	38	No	Travel_Frequently	216	23	3	1	
9	36	No	Travel_Rarely	1299	27	3	1	
10	35	No	Travel_Rarely	809	16	3	1	
11	29	No	Travel_Rarely	153	15	2	1	
12	31	No	Travel_Rarely	670	26	1	1	
13	34	No	Travel_Rarely	1346	19	2	1	
14	28	Yes	Travel_Rarely	103	24	3	1	
15	29	No	Travel_Rarely	1389	21	4	1	
16	32	No	Travel_Rarely	334	5	2	1	
17	22	No	Non-Travel	1123	16	2	1	
18	53	No	Travel_Rarely	1219	2	4	1	
19	38	No	Travel_Rarely	371	2	3	1	
20	24	No	Non-Travel	673	11	2	1	
21	36	Yes	Travel_Rarely	1218	9	4	1	
22	34	No	Travel_Rarely	419	7	4	1	
23	21	No	Travel_Rarely	391	15	2	1	
24	34	Yes	Travel_Rarely	699	6	1	1	
25	53	No	Travel_Rarely	1282	5	3	1	
26	32	Yes	Travel_Frequently	1125	16	1	1	
27	42	No	Travel_Rarely	691	8	4	1	
28	44	No	Travel_Rarely	477	7	4	1	
29	46	No	Travel_Rarely	705	2	4	1	
...	...	...	...	...	...	...	...	
1440	36	No	Travel_Frequently	688	4	2	1	
1441	56	No	Non-Travel	667	1	4	1	
1442	29	Yes	Travel_Rarely	1092	1	4	1	
1443	42	No	Travel_Rarely	300	2	3	1	
1444	56	Yes	Travel_Rarely	310	7	2	1	
1445	41	No	Travel_Rarely	582	28	4	1	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EmployeeCount	EnvironmentS
1446	34	No	Travel_Rarely	704	28	3	1	
1447	36	No	Non-Travel	301	15	4	1	
1448	41	No	Travel_Rarely	930	3	3	1	
1449	32	No	Travel_Rarely	529	2	3	1	
1450	35	No	Travel_Rarely	1146	26	4	1	
1451	38	No	Travel_Rarely	345	10	2	1	
1452	50	Yes	Travel_Frequently	878	1	4	1	
1453	36	No	Travel_Rarely	1120	11	4	1	
1454	45	No	Travel_Rarely	374	20	3	1	
1455	40	No	Travel_Rarely	1322	2	4	1	
1456	35	No	Travel_Frequently	1199	18	4	1	
1457	40	No	Travel_Rarely	1194	2	4	1	
1458	35	No	Travel_Rarely	287	1	4	1	
1459	29	No	Travel_Rarely	1378	13	2	1	
1460	29	No	Travel_Rarely	468	28	4	1	
1461	50	Yes	Travel_Rarely	410	28	3	1	
1462	39	No	Travel_Rarely	722	24	1	1	
1463	31	No	Non-Travel	325	5	3	1	
1464	26	No	Travel_Rarely	1167	5	3	1	
1465	36	No	Travel_Frequently	884	23	2	1	
1466	39	No	Travel_Rarely	613	6	1	1	
1467	27	No	Travel_Rarely	155	4	3	1	
1468	49	No	Travel_Frequently	1023	2	3	1	
1469	34	No	Travel_Rarely	628	8	3	1	

1470 rows × 31 columns

In [134]:
```
data=data.drop(["EmployeeCount"],axis=1)
```

In [135]:
```
data
```

Out[135]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
0	41	Yes	Travel_Rarely	1102	1	2	2	
1	49	No	Travel_Frequently	279	8	1	3	
2	37	Yes	Travel_Rarely	1373	2	2	4	
3	33	No	Travel_Frequently	1392	3	4	4	
4	27	No	Travel_Rarely	591	2	1	1	
5	32	No	Travel_Frequently	1005	2	2	4	
6	59	No	Travel_Rarely	1324	3	3	3	
7	30	No	Travel_Rarely	1358	24	1	4	
8	38	No	Travel_Frequently	216	23	3	4	
9	36	No	Travel_Rarely	1299	27	3	3	
10	35	No	Travel_Rarely	809	16	3	1	
11	29	No	Travel_Rarely	153	15	2	4	
12	31	No	Travel_Rarely	670	26	1	1	
13	34	No	Travel_Rarely	1346	19	2	2	
14	28	Yes	Travel_Rarely	103	24	3	3	
15	29	No	Travel_Rarely	1389	21	4	2	
16	32	No	Travel_Rarely	334	5	2	1	
17	22	No	Non-Travel	1123	16	2	4	
18	53	No	Travel_Rarely	1219	2	4	1	
19	38	No	Travel_Rarely	371	2	3	4	
20	24	No	Non-Travel	673	11	2	1	
21	36	Yes	Travel_Rarely	1218	9	4	3	
22	34	No	Travel_Rarely	419	7	4	1	
23	21	No	Travel_Rarely	391	15	2	3	
24	34	Yes	Travel_Rarely	699	6	1	2	
25	53	No	Travel_Rarely	1282	5	3	3	
26	32	Yes	Travel_Frequently	1125	16	1	2	
27	42	No	Travel_Rarely	691	8	4	3	
28	44	No	Travel_Rarely	477	7	4	1	
29	46	No	Travel_Rarely	705	2	4	2	
...	...	...	...	...	...	...	...	
1440	36	No	Travel_Frequently	688	4	2	4	
1441	56	No	Non-Travel	667	1	4	3	
1442	29	Yes	Travel_Rarely	1092	1	4	1	
1443	42	No	Travel_Rarely	300	2	3	1	
1444	56	Yes	Travel_Rarely	310	7	2	4	
1445	41	No	Travel_Rarely	582	28	4	1	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
1446	34	No	Travel_Rarely	704	28	3	4	
1447	36	No	Non-Travel	301	15	4	4	
1448	41	No	Travel_Rarely	930	3	3	3	
1449	32	No	Travel_Rarely	529	2	3	4	
1450	35	No	Travel_Rarely	1146	26	4	3	
1451	38	No	Travel_Rarely	345	10	2	1	
1452	50	Yes	Travel_Frequently	878	1	4	2	
1453	36	No	Travel_Rarely	1120	11	4	2	
1454	45	No	Travel_Rarely	374	20	3	4	
1455	40	No	Travel_Rarely	1322	2	4	3	
1456	35	No	Travel_Frequently	1199	18	4	3	
1457	40	No	Travel_Rarely	1194	2	4	3	
1458	35	No	Travel_Rarely	287	1	4	3	
1459	29	No	Travel_Rarely	1378	13	2	4	
1460	29	No	Travel_Rarely	468	28	4	4	
1461	50	Yes	Travel_Rarely	410	28	3	4	
1462	39	No	Travel_Rarely	722	24	1	2	
1463	31	No	Non-Travel	325	5	3	2	
1464	26	No	Travel_Rarely	1167	5	3	4	
1465	36	No	Travel_Frequently	884	23	2	3	
1466	39	No	Travel_Rarely	613	6	1	4	
1467	27	No	Travel_Rarely	155	4	3	2	
1468	49	No	Travel_Frequently	1023	2	3	4	
1469	34	No	Travel_Rarely	628	8	3	2	

1470 rows × 30 columns

In [136]:
```
data=data.drop(["Over18"],axis=1)
```

In [137]:
```
data
```

Out[137]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
0	41	Yes	Travel_Rarely	1102	1	2	2	
1	49	No	Travel_Frequently	279	8	1	3	
2	37	Yes	Travel_Rarely	1373	2	2	4	
3	33	No	Travel_Frequently	1392	3	4	4	
4	27	No	Travel_Rarely	591	2	1	1	
5	32	No	Travel_Frequently	1005	2	2	4	
6	59	No	Travel_Rarely	1324	3	3	3	
7	30	No	Travel_Rarely	1358	24	1	4	
8	38	No	Travel_Frequently	216	23	3	4	
9	36	No	Travel_Rarely	1299	27	3	3	
10	35	No	Travel_Rarely	809	16	3	1	
11	29	No	Travel_Rarely	153	15	2	4	
12	31	No	Travel_Rarely	670	26	1	1	
13	34	No	Travel_Rarely	1346	19	2	2	
14	28	Yes	Travel_Rarely	103	24	3	3	
15	29	No	Travel_Rarely	1389	21	4	2	
16	32	No	Travel_Rarely	334	5	2	1	
17	22	No	Non-Travel	1123	16	2	4	
18	53	No	Travel_Rarely	1219	2	4	1	
19	38	No	Travel_Rarely	371	2	3	4	
20	24	No	Non-Travel	673	11	2	1	
21	36	Yes	Travel_Rarely	1218	9	4	3	
22	34	No	Travel_Rarely	419	7	4	1	
23	21	No	Travel_Rarely	391	15	2	3	
24	34	Yes	Travel_Rarely	699	6	1	2	
25	53	No	Travel_Rarely	1282	5	3	3	
26	32	Yes	Travel_Frequently	1125	16	1	2	
27	42	No	Travel_Rarely	691	8	4	3	
28	44	No	Travel_Rarely	477	7	4	1	
29	46	No	Travel_Rarely	705	2	4	2	
...	...	...	...	...	...	...	...	
1440	36	No	Travel_Frequently	688	4	2	4	
1441	56	No	Non-Travel	667	1	4	3	
1442	29	Yes	Travel_Rarely	1092	1	4	1	
1443	42	No	Travel_Rarely	300	2	3	1	
1444	56	Yes	Travel_Rarely	310	7	2	4	
1445	41	No	Travel_Rarely	582	28	4	1	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
1446	34	No	Travel_Rarely	704	28	3	4	
1447	36	No	Non-Travel	301	15	4	4	
1448	41	No	Travel_Rarely	930	3	3	3	
1449	32	No	Travel_Rarely	529	2	3	4	
1450	35	No	Travel_Rarely	1146	26	4	3	
1451	38	No	Travel_Rarely	345	10	2	1	
1452	50	Yes	Travel_Frequently	878	1	4	2	
1453	36	No	Travel_Rarely	1120	11	4	2	
1454	45	No	Travel_Rarely	374	20	3	4	
1455	40	No	Travel_Rarely	1322	2	4	3	
1456	35	No	Travel_Frequently	1199	18	4	3	
1457	40	No	Travel_Rarely	1194	2	4	3	
1458	35	No	Travel_Rarely	287	1	4	3	
1459	29	No	Travel_Rarely	1378	13	2	4	
1460	29	No	Travel_Rarely	468	28	4	4	
1461	50	Yes	Travel_Rarely	410	28	3	4	
1462	39	No	Travel_Rarely	722	24	1	2	
1463	31	No	Non-Travel	325	5	3	2	
1464	26	No	Travel_Rarely	1167	5	3	4	
1465	36	No	Travel_Frequently	884	23	2	3	
1466	39	No	Travel_Rarely	613	6	1	4	
1467	27	No	Travel_Rarely	155	4	3	2	
1468	49	No	Travel_Frequently	1023	2	3	4	
1469	34	No	Travel_Rarely	628	8	3	2	

1470 rows × 29 columns

In [138]:
```
data=data.drop(["JobRole"],axis=1)
```

In [139]:
```
data
```

Out[139]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
0	41	Yes	Travel_Rarely	1102	1	2	2	
1	49	No	Travel_Frequently	279	8	1	3	
2	37	Yes	Travel_Rarely	1373	2	2	4	
3	33	No	Travel_Frequently	1392	3	4	4	
4	27	No	Travel_Rarely	591	2	1	1	
5	32	No	Travel_Frequently	1005	2	2	4	
6	59	No	Travel_Rarely	1324	3	3	3	
7	30	No	Travel_Rarely	1358	24	1	4	
8	38	No	Travel_Frequently	216	23	3	4	
9	36	No	Travel_Rarely	1299	27	3	3	
10	35	No	Travel_Rarely	809	16	3	1	
11	29	No	Travel_Rarely	153	15	2	4	
12	31	No	Travel_Rarely	670	26	1	1	
13	34	No	Travel_Rarely	1346	19	2	2	
14	28	Yes	Travel_Rarely	103	24	3	3	
15	29	No	Travel_Rarely	1389	21	4	2	
16	32	No	Travel_Rarely	334	5	2	1	
17	22	No	Non-Travel	1123	16	2	4	
18	53	No	Travel_Rarely	1219	2	4	1	
19	38	No	Travel_Rarely	371	2	3	4	
20	24	No	Non-Travel	673	11	2	1	
21	36	Yes	Travel_Rarely	1218	9	4	3	
22	34	No	Travel_Rarely	419	7	4	1	
23	21	No	Travel_Rarely	391	15	2	3	
24	34	Yes	Travel_Rarely	699	6	1	2	
25	53	No	Travel_Rarely	1282	5	3	3	
26	32	Yes	Travel_Frequently	1125	16	1	2	
27	42	No	Travel_Rarely	691	8	4	3	
28	44	No	Travel_Rarely	477	7	4	1	
29	46	No	Travel_Rarely	705	2	4	2	
...	...	...	...	...	...	...	...	
1440	36	No	Travel_Frequently	688	4	2	4	
1441	56	No	Non-Travel	667	1	4	3	
1442	29	Yes	Travel_Rarely	1092	1	4	1	
1443	42	No	Travel_Rarely	300	2	3	1	
1444	56	Yes	Travel_Rarely	310	7	2	4	
1445	41	No	Travel_Rarely	582	28	4	1	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hour
1446	34	No	Travel_Rarely	704	28	3	4	
1447	36	No	Non-Travel	301	15	4	4	
1448	41	No	Travel_Rarely	930	3	3	3	
1449	32	No	Travel_Rarely	529	2	3	4	
1450	35	No	Travel_Rarely	1146	26	4	3	
1451	38	No	Travel_Rarely	345	10	2	1	
1452	50	Yes	Travel_Frequently	878	1	4	2	
1453	36	No	Travel_Rarely	1120	11	4	2	
1454	45	No	Travel_Rarely	374	20	3	4	
1455	40	No	Travel_Rarely	1322	2	4	3	
1456	35	No	Travel_Frequently	1199	18	4	3	
1457	40	No	Travel_Rarely	1194	2	4	3	
1458	35	No	Travel_Rarely	287	1	4	3	
1459	29	No	Travel_Rarely	1378	13	2	4	
1460	29	No	Travel_Rarely	468	28	4	4	
1461	50	Yes	Travel_Rarely	410	28	3	4	
1462	39	No	Travel_Rarely	722	24	1	2	
1463	31	No	Non-Travel	325	5	3	2	
1464	26	No	Travel_Rarely	1167	5	3	4	
1465	36	No	Travel_Frequently	884	23	2	3	
1466	39	No	Travel_Rarely	613	6	1	4	
1467	27	No	Travel_Rarely	155	4	3	2	
1468	49	No	Travel_Frequently	1023	2	3	4	
1469	34	No	Travel_Rarely	628	8	3	2	

1470 rows × 28 columns

In [140]:
```
BusinessTravel=data['BusinessTravel']=pd.get_dummies(data['BusinessTravel'],drop_first=Tru
```

In [141]:
```
BusinessTravel
```

Out[141]:

	Travel_Frequently	Travel_Rarely
0	0	1
1	1	0
2	0	1
3	1	0
4	0	1
5	1	0
6	0	1
7	0	1
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1
13	0	1
14	0	1
15	0	1
16	0	1
17	0	0
18	0	1
19	0	1
20	0	0
21	0	1
22	0	1
23	0	1
24	0	1
25	0	1
26	1	0
27	0	1
28	0	1
29	0	1
...	...	...
1440	1	0
1441	0	0
1442	0	1
1443	0	1
1444	0	1
1445	0	1

	Travel_Frequently	Travel_Rarely
1446	0	1
1447	0	0
1448	0	1
1449	0	1
1450	0	1
1451	0	1
1452	1	0
1453	0	1
1454	0	1
1455	0	1
1456	1	0
1457	0	1
1458	0	1
1459	0	1
1460	0	1
1461	0	1
1462	0	1
1463	0	0
1464	0	1
1465	1	0
1466	0	1
1467	0	1
1468	1	0
1469	0	1

1470 rows × 2 columns

In [142]:
```
MaritalStatus=data['MaritalStatus']=pd.get_dummies(data['MaritalStatus'],drop_first=True)
```

In [ ]:
```
OverTime=data['OverTime']=pd.get_dummies(data['OverTime'],drop_first=True)
```

In [143]:
```
MaritalStatus
```

Out[143]:

	Married	Single
0	0	1
1	1	0
2	0	1
3	1	0
4	1	0
5	0	1
6	1	0
7	0	0
8	0	1
9	1	0
10	1	0
11	0	1
12	0	0
13	0	0
14	0	1
15	0	0
16	0	0
17	0	0
18	1	0
19	0	1
20	0	0
21	0	1
22	0	1
23	0	1
24	0	1
25	0	0
26	0	1
27	1	0
28	1	0
29	0	1
...	...	...
1440	0	0
1441	0	0
1442	1	0
1443	1	0
1444	1	0
1445	1	0

	Married	Single
1446	1	0
1447	0	0
1448	0	0
1449	0	1
1450	0	1
1451	1	0
1452	0	0
1453	1	0
1454	0	1
1455	0	1
1456	1	0
1457	1	0
1458	1	0
1459	1	0
1460	0	1
1461	0	0
1462	1	0
1463	0	1
1464	0	1
1465	1	0
1466	1	0
1467	1	0
1468	1	0
1469	1	0

1470 rows × 2 columns

In [144]:
```
Attrition=data['Attrition']=pd.get_dummies(data['Attrition'],drop_first=True)
```

In [145]:
```
Attrition
```

Out[145]:

	Yes
0	1
1	0
2	1
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	1
15	0
16	0
17	0
18	0
19	0
20	0
21	1
22	0
23	0
24	1
25	0
26	1
27	0
28	0
29	0
...	...
1440	0
1441	0
1442	1
1443	0
1444	1
1445	0

	Yes
1446	0
1447	0
1448	0
1449	0
1450	0
1451	0
1452	1
1453	0
1454	0
1455	0
1456	0
1457	0
1458	0
1459	0
1460	0
1461	1
1462	0
1463	0
1464	0
1465	0
1466	0
1467	0
1468	0
1469	0

1470 rows × 1 columns

In [146]:
```
data=pd.concat([data,BusinessTravel,Attrition,MaritalStatus],axis=1)
```

In [147]:
```
data
```

Out[147]:

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hourly
0	41	1	0	1102	1	2	2	
1	49	0	1	279	8	1	3	
2	37	1	0	1373	2	2	4	
3	33	0	1	1392	3	4	4	
4	27	0	0	591	2	1	1	
5	32	0	1	1005	2	2	4	
6	59	0	0	1324	3	3	3	
7	30	0	0	1358	24	1	4	
8	38	0	1	216	23	3	4	
9	36	0	0	1299	27	3	3	
10	35	0	0	809	16	3	1	
11	29	0	0	153	15	2	4	
12	31	0	0	670	26	1	1	
13	34	0	0	1346	19	2	2	
14	28	1	0	103	24	3	3	
15	29	0	0	1389	21	4	2	
16	32	0	0	334	5	2	1	
17	22	0	0	1123	16	2	4	
18	53	0	0	1219	2	4	1	
19	38	0	0	371	2	3	4	
20	24	0	0	673	11	2	1	
21	36	1	0	1218	9	4	3	
22	34	0	0	419	7	4	1	
23	21	0	0	391	15	2	3	
24	34	1	0	699	6	1	2	
25	53	0	0	1282	5	3	3	
26	32	1	1	1125	16	1	2	
27	42	0	0	691	8	4	3	
28	44	0	0	477	7	4	1	
29	46	0	0	705	2	4	2	
...	...	...	...	...	...	...	...	
1440	36	0	1	688	4	2	4	
1441	56	0	0	667	1	4	3	
1442	29	1	0	1092	1	4	1	
1443	42	0	0	300	2	3	1	
1444	56	1	0	310	7	2	4	
1445	41	0	0	582	28	4	1	

	Age	Attrition	BusinessTravel	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Hourly
1446	34	0	0	704	28	3	4	
1447	36	0	0	301	15	4	4	
1448	41	0	0	930	3	3	3	
1449	32	0	0	529	2	3	4	
1450	35	0	0	1146	26	4	3	
1451	38	0	0	345	10	2	1	
1452	50	1	1	878	1	4	2	
1453	36	0	0	1120	11	4	2	
1454	45	0	0	374	20	3	4	
1455	40	0	0	1322	2	4	3	
1456	35	0	1	1199	18	4	3	
1457	40	0	0	1194	2	4	3	
1458	35	0	0	287	1	4	3	
1459	29	0	0	1378	13	2	4	
1460	29	0	0	468	28	4	4	
1461	50	1	0	410	28	3	4	
1462	39	0	0	722	24	1	2	
1463	31	0	0	325	5	3	2	
1464	26	0	0	1167	5	3	4	
1465	36	0	1	884	23	2	3	
1466	39	0	0	613	6	1	4	
1467	27	0	0	155	4	3	2	
1468	49	0	1	1023	2	3	4	
1469	34	0	0	628	8	3	2	

1470 rows × 33 columns

In [148]:
```
data=data.drop(['BusinessTravel','Attrition','MaritalStatus'],axis=1)
```

In [149]:
```
data
```

Out[149]:

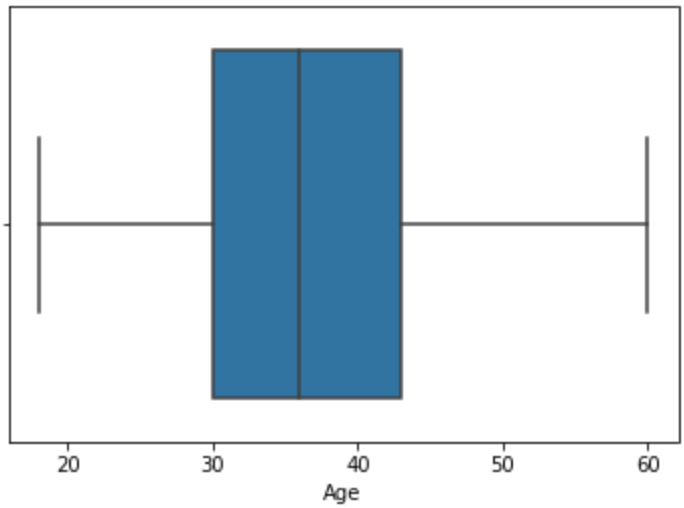
	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	HourlyRate	JobInvolvement	Jo
0	41	1102	1	2	2	94	3	
1	49	279	8	1	3	61	2	
2	37	1373	2	2	4	92	2	
3	33	1392	3	4	4	56	3	
4	27	591	2	1	1	40	3	
5	32	1005	2	2	4	79	3	
6	59	1324	3	3	3	81	4	
7	30	1358	24	1	4	67	3	
8	38	216	23	3	4	44	2	
9	36	1299	27	3	3	94	3	
10	35	809	16	3	1	84	4	
11	29	153	15	2	4	49	2	
12	31	670	26	1	1	31	3	
13	34	1346	19	2	2	93	3	
14	28	103	24	3	3	50	2	
15	29	1389	21	4	2	51	4	
16	32	334	5	2	1	80	4	
17	22	1123	16	2	4	96	4	
18	53	1219	2	4	1	78	2	
19	38	371	2	3	4	45	3	
20	24	673	11	2	1	96	4	
21	36	1218	9	4	3	82	2	
22	34	419	7	4	1	53	3	
23	21	391	15	2	3	96	3	
24	34	699	6	1	2	83	3	
25	53	1282	5	3	3	58	3	
26	32	1125	16	1	2	72	1	
27	42	691	8	4	3	48	3	
28	44	477	7	4	1	42	2	
29	46	705	2	4	2	83	3	
...	...	...	...	...	...	...	...	
1440	36	688	4	2	4	97	3	
1441	56	667	1	4	3	57	3	
1442	29	1092	1	4	1	36	3	
1443	42	300	2	3	1	56	3	
1444	56	310	7	2	4	72	3	
1445	41	582	28	4	1	60	2	

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	HourlyRate	JobInvolvement	Jol
1446	34	704	28	3	4	95	2	
1447	36	301	15	4	4	88	1	
1448	41	930	3	3	3	57	2	
1449	32	529	2	3	4	78	3	
1450	35	1146	26	4	3	31	3	
1451	38	345	10	2	1	100	3	
1452	50	878	1	4	2	94	3	
1453	36	1120	11	4	2	100	2	
1454	45	374	20	3	4	50	3	
1455	40	1322	2	4	3	52	2	
1456	35	1199	18	4	3	80	3	
1457	40	1194	2	4	3	98	3	
1458	35	287	1	4	3	62	1	
1459	29	1378	13	2	4	46	2	
1460	29	468	28	4	4	73	2	
1461	50	410	28	3	4	39	2	
1462	39	722	24	1	2	60	2	
1463	31	325	5	3	2	74	3	
1464	26	1167	5	3	4	30	2	
1465	36	884	23	2	3	41	4	
1466	39	613	6	1	4	42	2	
1467	27	155	4	3	2	87	4	
1468	49	1023	2	3	4	63	2	
1469	34	628	8	3	2	82	4	

1470 rows × 30 columns

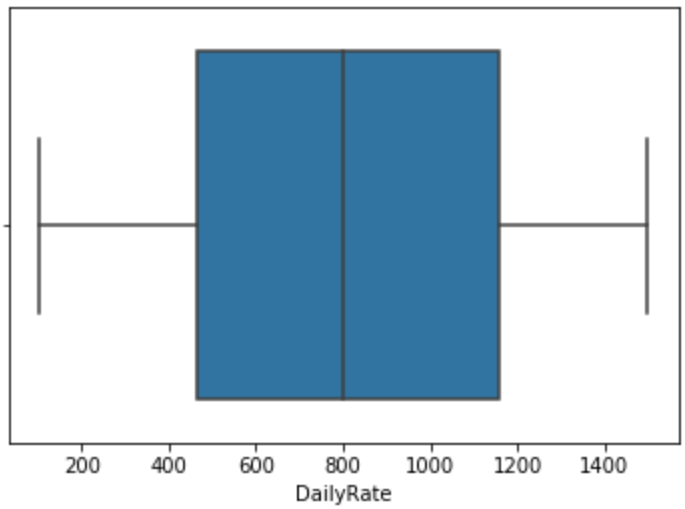
In [150]:
```
sns.boxplot(x="Age",data=data)
```

Out[150]: <matplotlib.axes._subplots.AxesSubplot at 0x114ceac8>



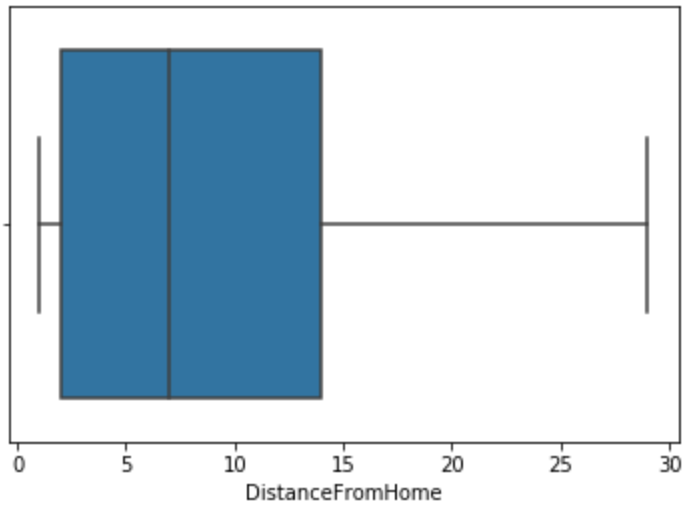
In [151]:
```
sns.boxplot(x="DailyRate",data=data)
```

Out[151]: <matplotlib.axes._subplots.AxesSubplot at 0x114ce4e0>



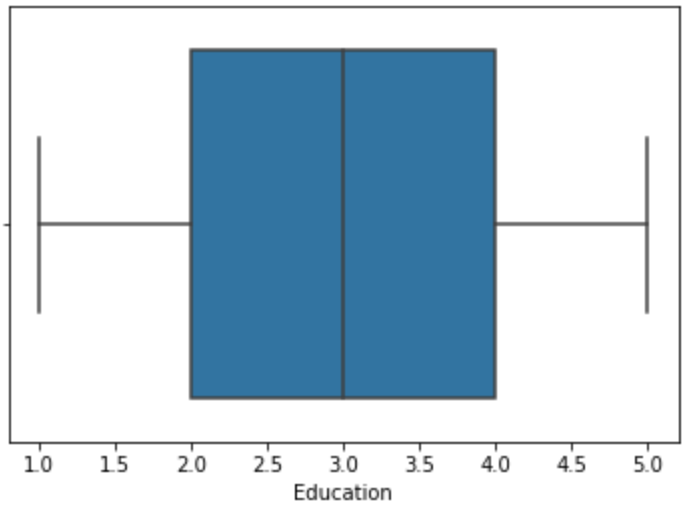
In [152]:
```
sns.boxplot(x="DistanceFromHome",data=data)
```

Out[152]: <matplotlib.axes._subplots.AxesSubplot at 0x12539860>



In [153]:
```
sns.boxplot(x="Education",data=data)
```

Out[153]: <matplotlib.axes._subplots.AxesSubplot at 0x114a0630>



In [154]:
```
sns.boxplot(x="EmployeeCount",data=data)
```

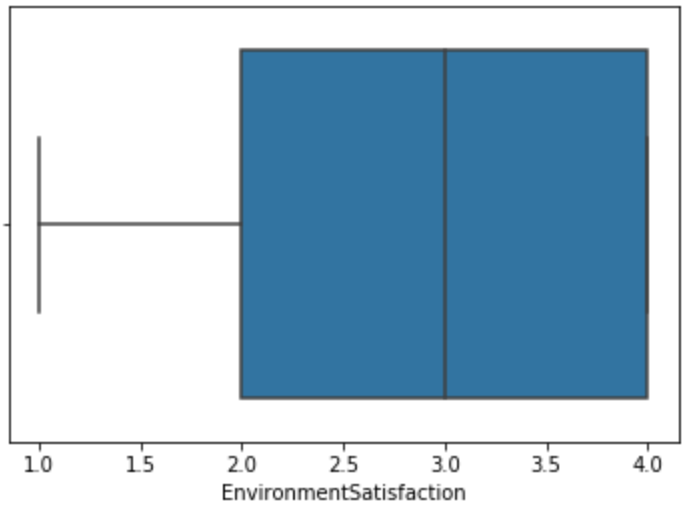
```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-154-8ce17f43ec59> in <module>
----> 1 sns.boxplot(x="EmployeeCount",data=data)

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in boxplot(x, y, hue, d
ata, order, hue_order, orient, color, palette, saturation, width, dodge, fliersize, linew
idth, whis, notch, ax, **kwargs)
   2229     plotter = _BoxPlotter(x, y, hue, data, order, hue_order,
   2230                           orient, color, palette, saturation,
-> 2231                           width, dodge, fliersize, linewidth)
   2232
   2233     if ax is None:

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in __init__(self, x, y,
hue, data, order, hue_order, orient, color, palette, saturation, width, dodge, fliersize,
linewidth)
    444                  width, dodge, fliersize, linewidth):
    445
--> 446         self.establish_variables(x, y, hue, data, orient, order, hue_order)
    447         self.establish_colors(color, palette, saturation)
    448

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in establish_variables
(self, x, y, hue, data, orient, order, hue_order, units)
    153                 if isinstance(input, string_types):
    154                     err = "Could not interpret input '{}'".format(input)
--> 155                     raise ValueError(err)
    156
    157             # Figure out the plotting orientation

ValueError: Could not interpret input 'EmployeeCount'
```

In [155]:
```
sns.boxplot(x="EmployeeNumber",data=data)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-155-f85b1900c61a> in <module>
----> 1 sns.boxplot(x="EmployeeNumber",data=data)

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in boxplot(x, y, hue, d
ata, order, hue_order, orient, color, palette, saturation, width, dodge, fliersize, linew
idth, whis, notch, ax, **kwargs)
   2229     plotter = _BoxPlotter(x, y, hue, data, order, hue_order,
   2230                           orient, color, palette, saturation,
-> 2231                           width, dodge, fliersize, linewidth)
   2232
   2233     if ax is None:

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in __init__(self, x, y,
hue, data, order, hue_order, orient, color, palette, saturation, width, dodge, fliersize,
linewidth)
    444                  width, dodge, fliersize, linewidth):
    445
--> 446         self.establish_variables(x, y, hue, data, orient, order, hue_order)
    447         self.establish_colors(color, palette, saturation)
    448

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py in establish_variables
(self, x, y, hue, data, orient, order, hue_order, units)
    153                 if isinstance(input, string_types):
    154                     err = "Could not interpret input '{}'".format(input)
--> 155                     raise ValueError(err)
    156
    157             # Figure out the plotting orientation

ValueError: Could not interpret input 'EmployeeNumber'
```

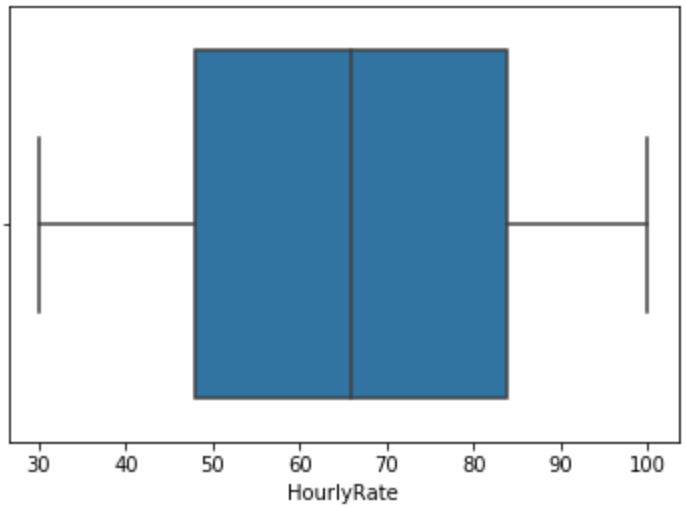
In [156]:
```
sns.boxplot(x="EnvironmentSatisfaction",data=data)
```

Out[156]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x125fc860>
```



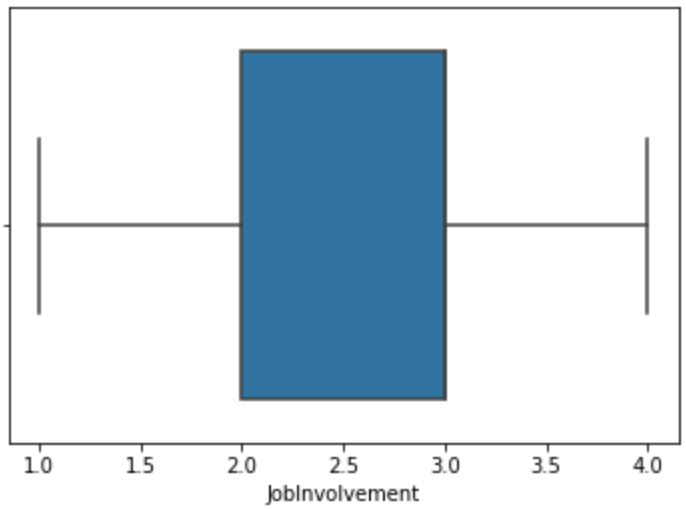
In [157]:
```
sns.boxplot(x="HourlyRate",data=data)
```

Out[157]: <matplotlib.axes._subplots.AxesSubplot at 0x12665048>



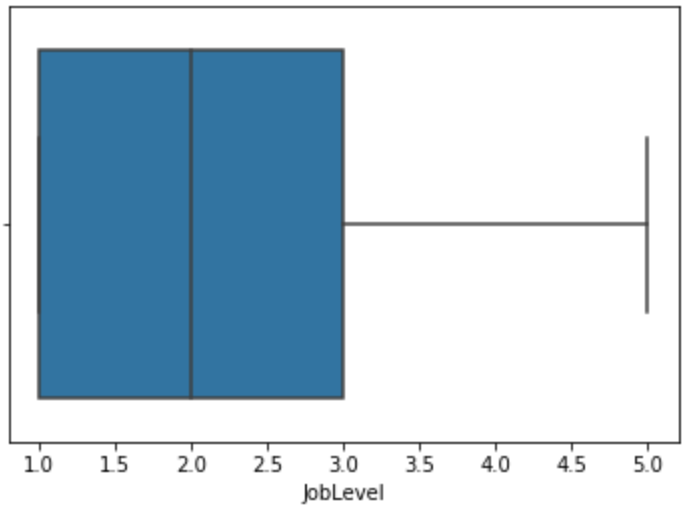
In [158]:
```
sns.boxplot(x="JobInvolvement",data=data)
```

Out[158]: <matplotlib.axes._subplots.AxesSubplot at 0x126bf2b0>



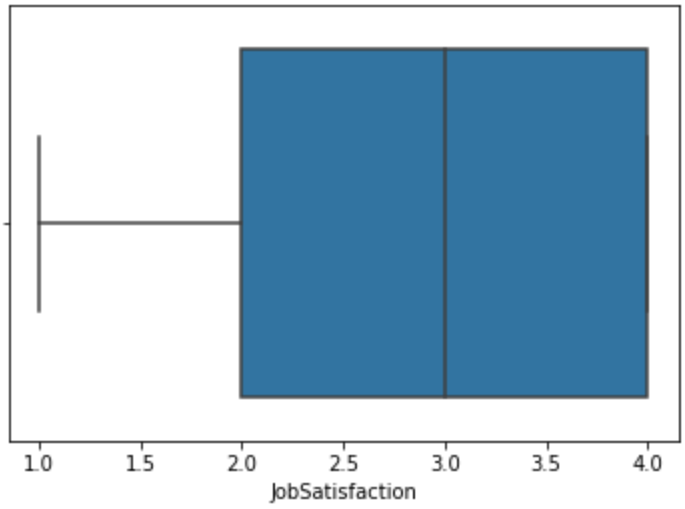
In [159]:
```
sns.boxplot(x="JobLevel",data=data)
```

Out[159]: <matplotlib.axes._subplots.AxesSubplot at 0x125fc3c8>



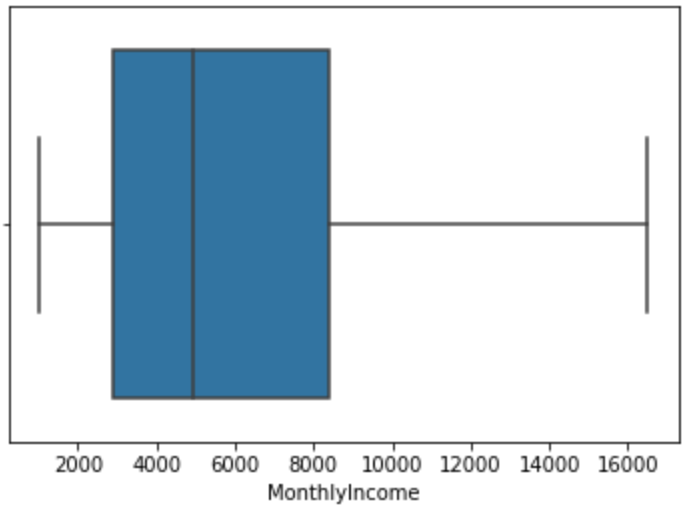
In [160]:
```
sns.boxplot(x="JobSatisfaction",data=data)
```

Out[160]: <matplotlib.axes._subplots.AxesSubplot at 0x127780b8>



In [161]:
```
sns.boxplot(x="MonthlyIncome",data=data)
```

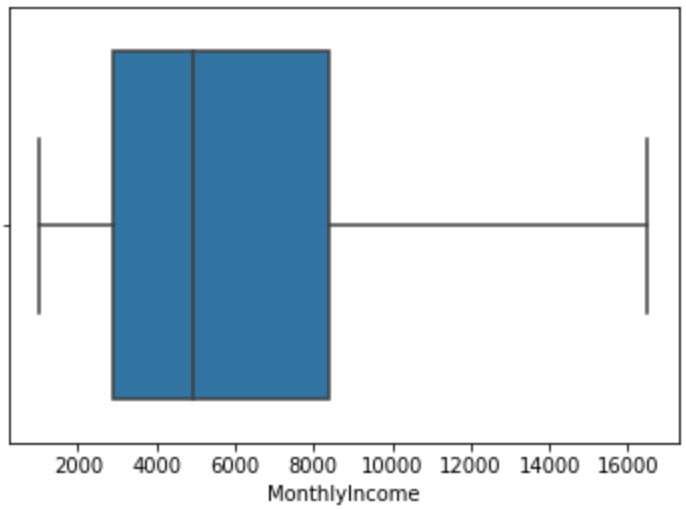
Out[161]: <matplotlib.axes._subplots.AxesSubplot at 0x127d6668>



In [162]:
```
data['MonthlyIncome']=np.where(data['MonthlyIncome']>16500,16500,data['MonthlyIncome'])
```

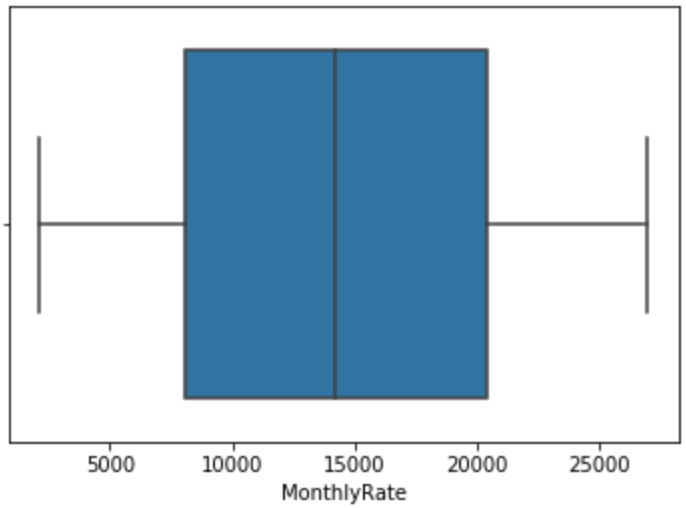
In [163]:
```
sns.boxplot(data['MonthlyIncome'])
```

Out[163]: <matplotlib.axes._subplots.AxesSubplot at 0x12822cf8>



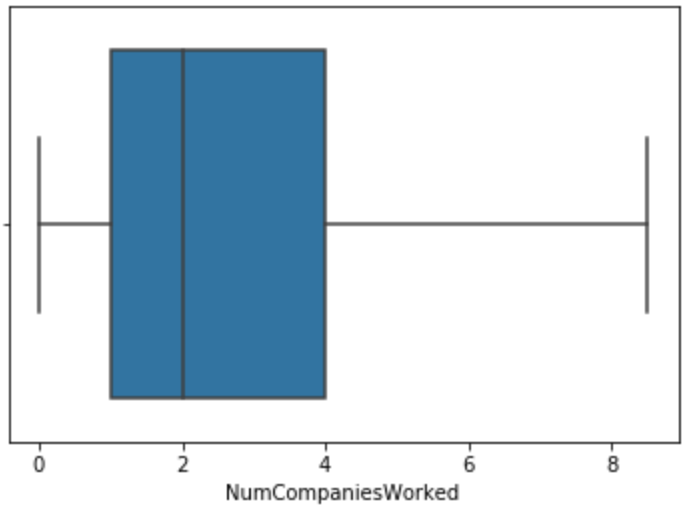
In [164]:
```
sns.boxplot(x="MonthlyRate",data=data)
```

Out[164]: <matplotlib.axes._subplots.AxesSubplot at 0x1286e908>



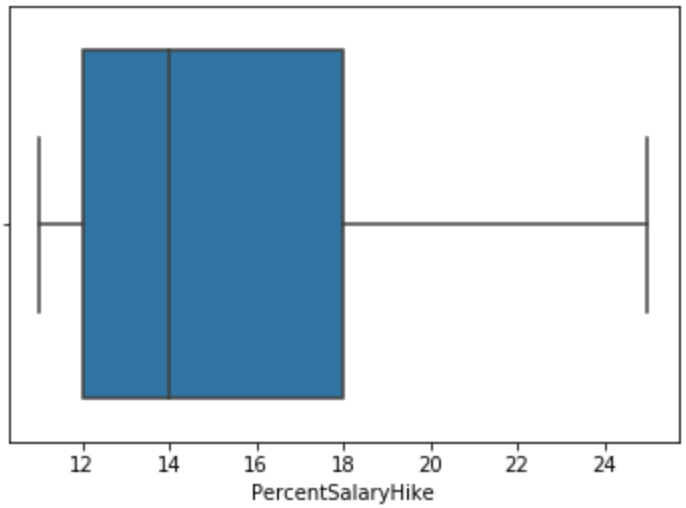
In [165]:
```
sns.boxplot(x="NumCompaniesWorked",data=data)
```

Out[165]: <matplotlib.axes._subplots.AxesSubplot at 0x128ca4a8>



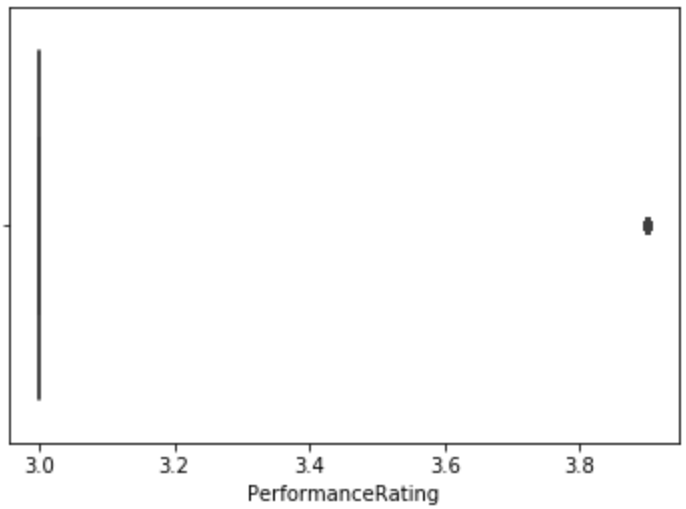
In [166]:
```
sns.boxplot(x="PercentSalaryHike",data=data)
```

Out[166]: <matplotlib.axes._subplots.AxesSubplot at 0x1291d2e8>



In [167]:
```
sns.boxplot(x="PerformanceRating",data=data)
```

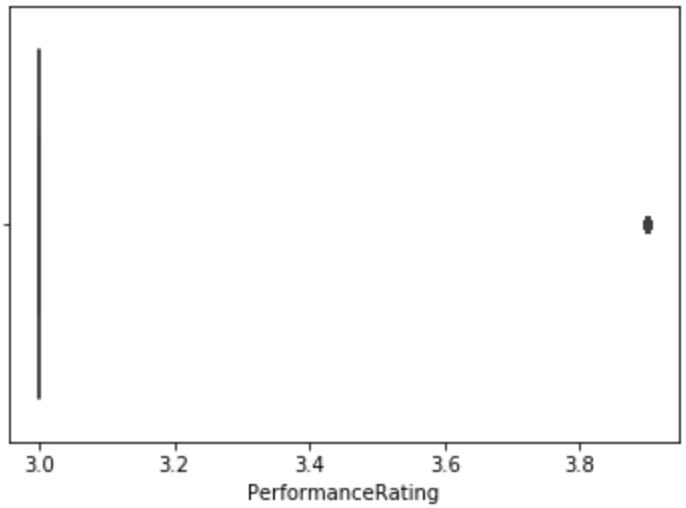
Out[167]: <matplotlib.axes._subplots.AxesSubplot at 0x129708d0>



In [168]:
```
data['PerformanceRating']=np.where(data['PerformanceRating']>3.9,3.9,data['PerformanceRati
```

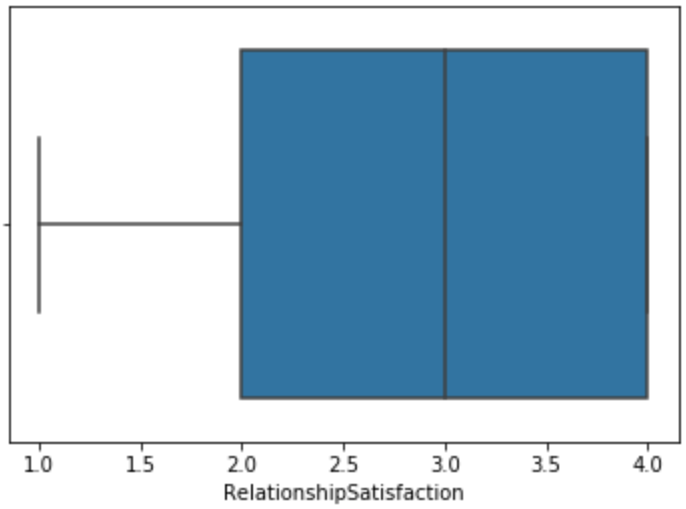
In [169]:
```
sns.boxplot(data['PerformanceRating'])
```

Out[169]: <matplotlib.axes._subplots.AxesSubplot at 0x5c3d588>



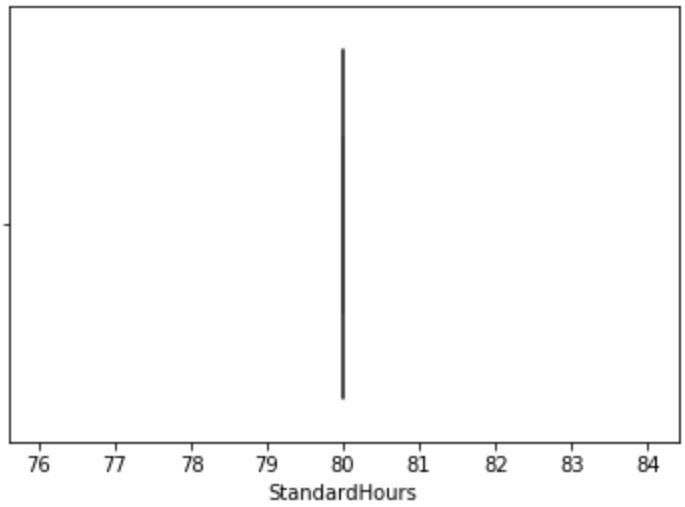
In [170]:
```
sns.boxplot(data['RelationshipSatisfaction'],data=data)
```

Out[170]: <matplotlib.axes._subplots.AxesSubplot at 0x5c8c4e0>



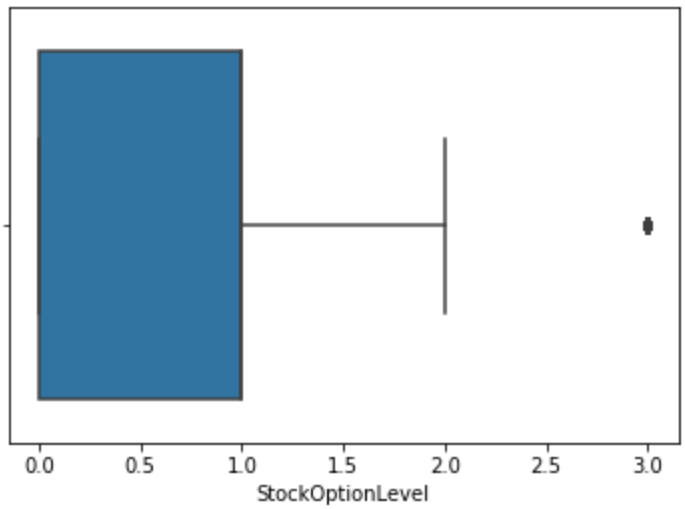
In [171]:
```
sns.boxplot(data['StandardHours'],data=data)
```

Out[171]: <matplotlib.axes._subplots.AxesSubplot at 0x5cd7390>



In [172]:
```
sns.boxplot(data['StockOptionLevel'],data=data)
```

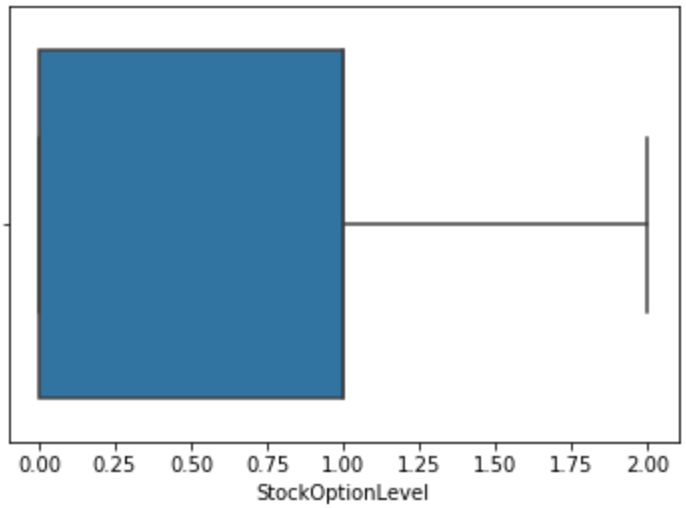
Out[172]: <matplotlib.axes._subplots.AxesSubplot at 0x5d325f8>



In [173]:
```
data['StockOptionLevel']=np.where(data['StockOptionLevel']>2,2,data['StockOptionLevel'])
```

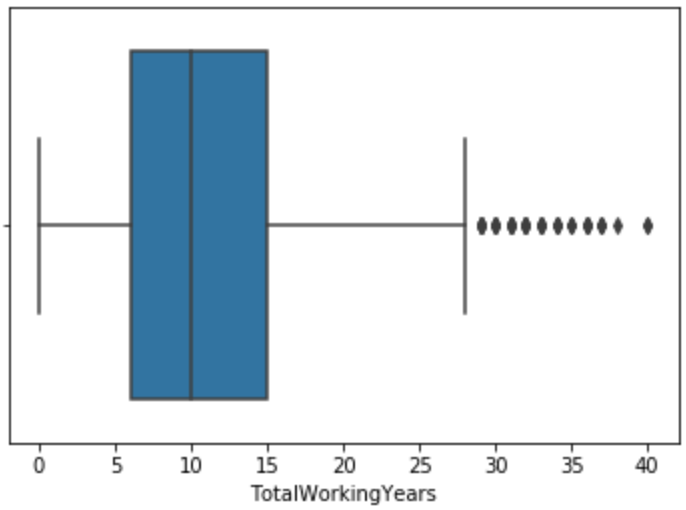
In [174]:
```
sns.boxplot(data['StockOptionLevel'])
```

Out[174]: <matplotlib.axes._subplots.AxesSubplot at 0x5da3a90>



In [175]:
```
sns.boxplot(data['TotalWorkingYears'],data=data)
```

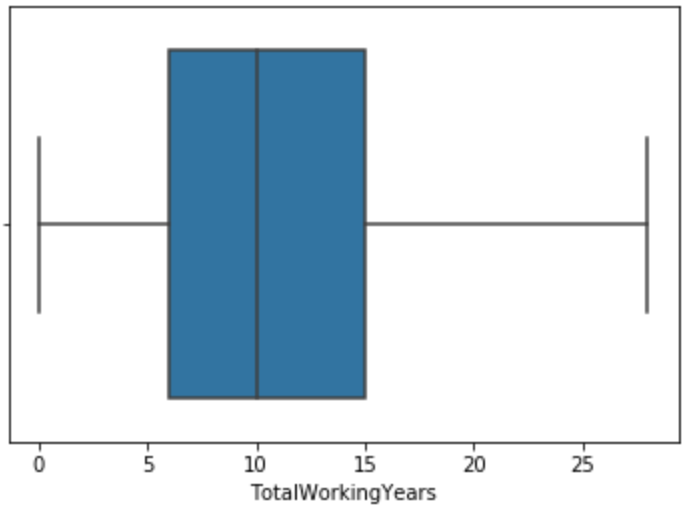
Out[175]: <matplotlib.axes._subplots.AxesSubplot at 0x129d9eb8>



In [176]:
```
data['TotalWorkingYears']=np.where(data['TotalWorkingYears']>28,28,data['TotalWorkingYears
```

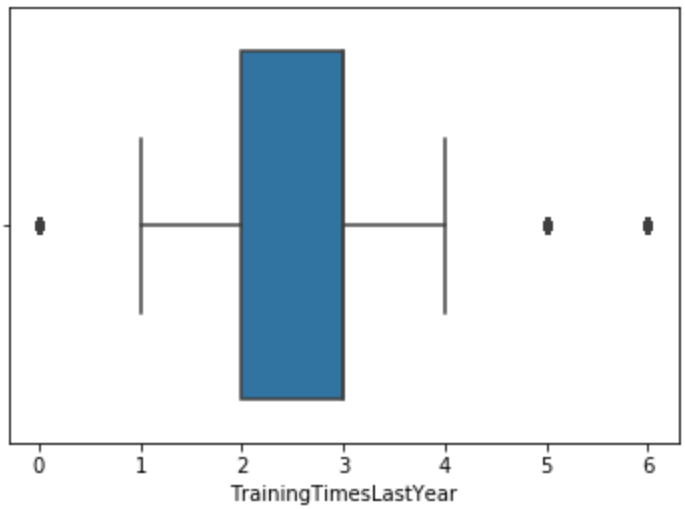
In [177]:
```
sns.boxplot(data['TotalWorkingYears'])
```

Out[177]: <matplotlib.axes._subplots.AxesSubplot at 0x12a33588>



In [178]:
```
sns.boxplot(data['TrainingTimesLastYear'],data=data)
```

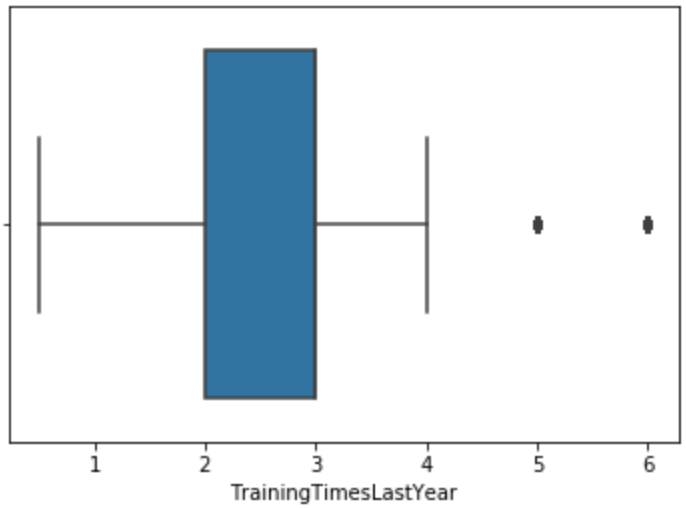
Out[178]: <matplotlib.axes._subplots.AxesSubplot at 0x12a7b588>



In [179]:
```
data['TrainingTimesLastYear']=np.where(data['TrainingTimesLastYear']<0.5,0.5,data['Trainin
```

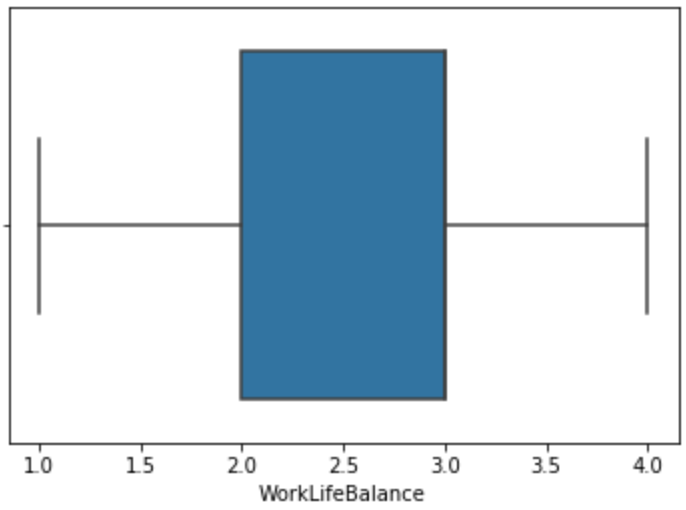
In [180]:
```
sns.boxplot(data['TrainingTimesLastYear'])
```

Out[180]: <matplotlib.axes._subplots.AxesSubplot at 0x13aa8d68>



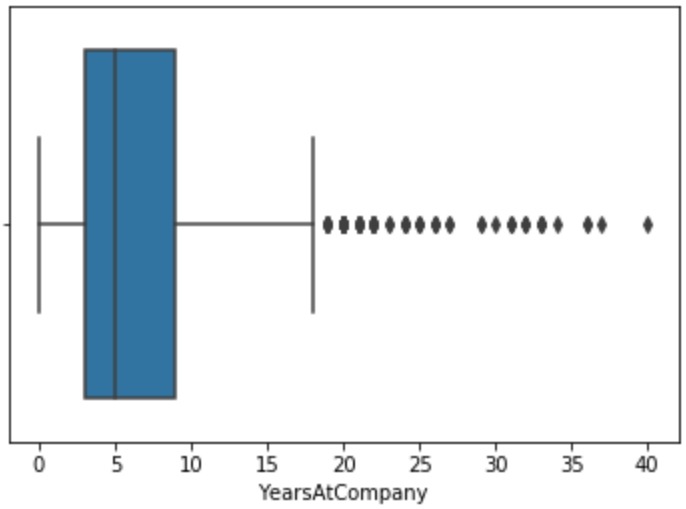
In [181]:
```
sns.boxplot(data['WorkLifeBalance'],data=data)
```

Out[181]: <matplotlib.axes._subplots.AxesSubplot at 0x13afc518>



In [182]:
```
sns.boxplot(data['YearsAtCompany'],data=data)
```

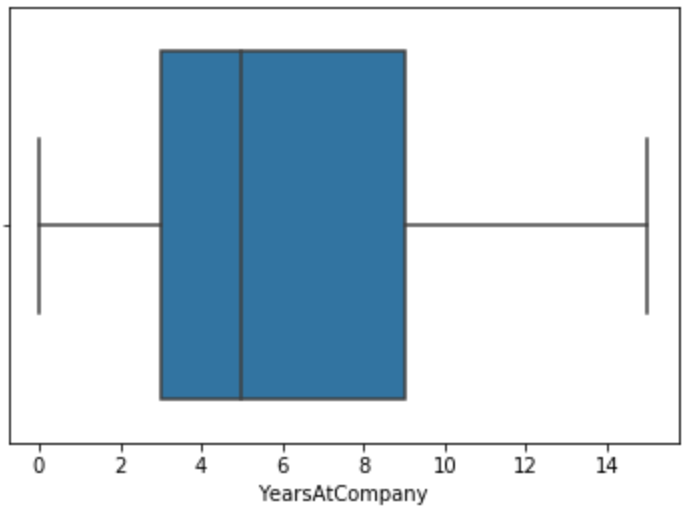
Out[182]: <matplotlib.axes._subplots.AxesSubplot at 0x13b60358>



In [183]:
```
data['YearsAtCompany']=np.where(data['YearsAtCompany']>15,15,data['YearsAtCompany'])
```

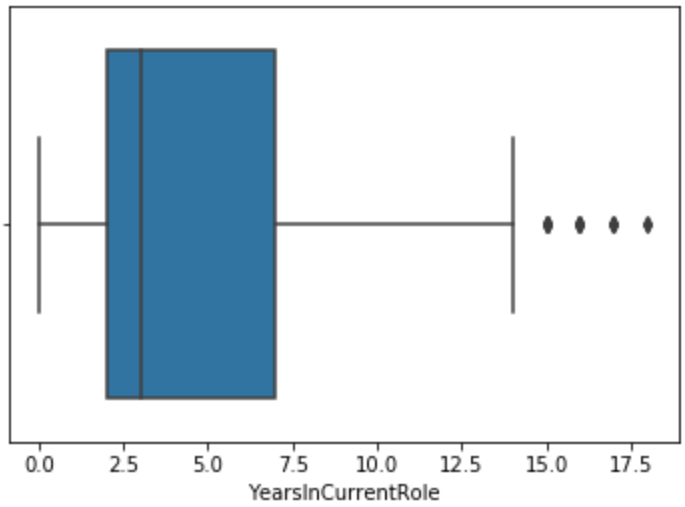
In [184]:
```
sns.boxplot(data['YearsAtCompany'])
```

Out[184]: <matplotlib.axes._subplots.AxesSubplot at 0x13bbca58>



In [185]:
```
sns.boxplot(data['YearsInCurrentRole'],data=data)
```

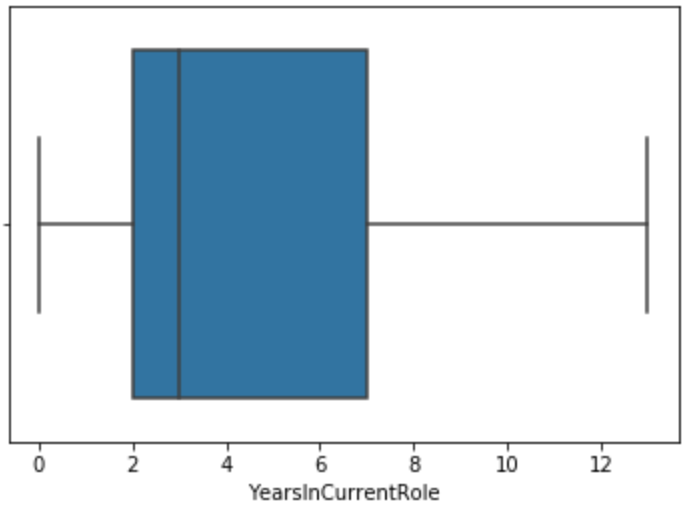
Out[185]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x13c18128>
```



In [186]:
```
data['YearsInCurrentRole']=np.where(data['YearsInCurrentRole']>13.0,13.0,data['YearsInCurr
```

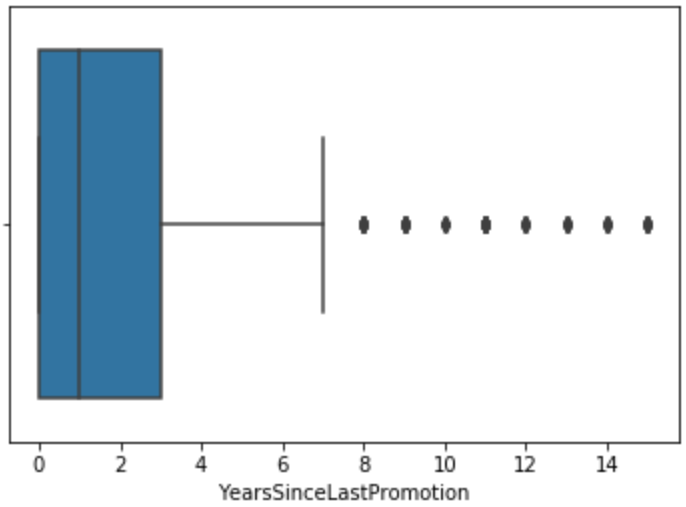
In [187]:
```
sns.boxplot(data['YearsInCurrentRole'])
```

Out[187]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x13c72d68>
```



In [188]:
```
sns.boxplot(data['YearsSinceLastPromotion'],data=data)
```

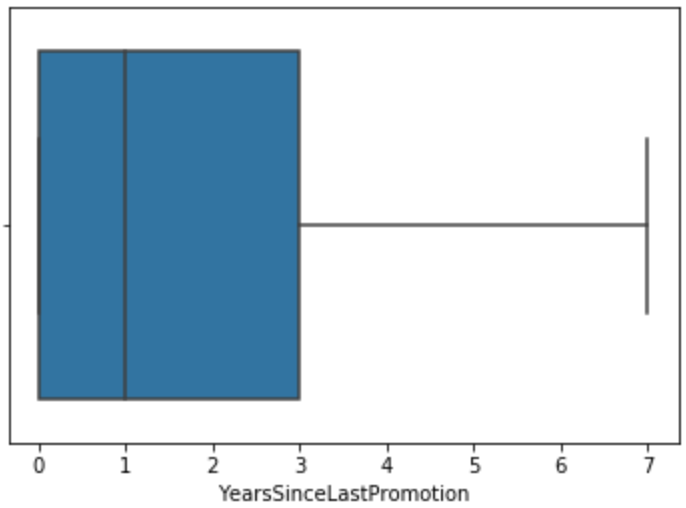
Out[188]: <matplotlib.axes._subplots.AxesSubplot at 0x13cc9240>



In [189]:
```
data['YearsSinceLastPromotion']=np.where(data['YearsSinceLastPromotion']>7,7,data['YearsSi
```

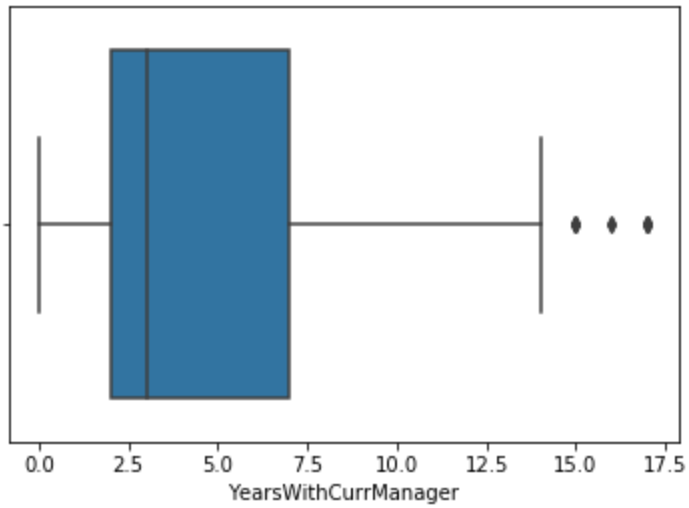
In [190]:
```
sns.boxplot(data['YearsSinceLastPromotion'])
```

Out[190]: <matplotlib.axes._subplots.AxesSubplot at 0x13d2cc50>



In [191]:
```
sns.boxplot(data['YearsWithCurrManager'],data=data)
```

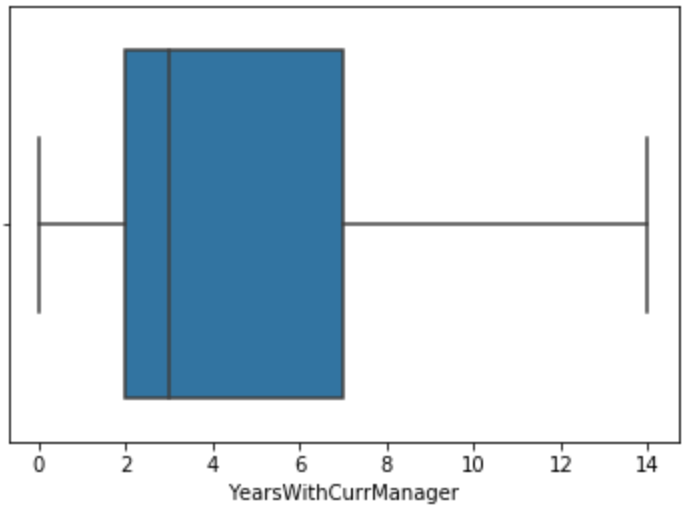
Out[191]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x13d89e80>
```



In [192]:
```
data['YearsWithCurrManager']=np.where(data['YearsWithCurrManager']>14,14,data['YearsWithCu
```

In [193]:
```
sns.boxplot(data['YearsWithCurrManager'])
```

Out[193]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x13d17898>
```



In [194]:
```
x=data.drop(['Yes'],axis=1)
```

In [195]:
```
y=data.Yes
```

In [ ]:
```
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import cohen_kappa_score as kappa
from sklearn.metrics import confusion_matrix
from sklearn import metrics
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
```

In [ ]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=10)
```

In [ ]:
```
from sklearn.linear_model import LinearRegression
```

In [ ]:
```
 classifier=(LogisticRegression())
```

In [ ]:
```
classifier.fit(x_train,y_train)
```

In [ ]:
```
y_pred=classifier.predict(x_test)
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: