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

data=pd.read_csv('generaldata.csv')

data.head()
Out[4]:
   Age Attrition  ... YearsSinceLastPromotion YearsWithCurrManager
0   51       No  ...                     0                    0
1   31      Yes  ...                     1                    4
2   32       No  ...                     0                    3
3   38       No  ...                     7                    5
4   32       No  ...                     0                    4

[5 rows x 24 columns]
data.columns
Out[6]:
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')

data.isnull()
Out[7]:
       Age  Attrition  ...  YearsSinceLastPromotion  YearsWithCurrManager
```

0 False False ... False False

1 False False ... False False

2 False False ... False False

3 False False ... False False

4 False False ... False False

... ...

4405 False False ... False False

4406 False False ... False False

4407 False False ... False False

4408 False False ... False False

4409 False False ... False False

[4410 rows x 24 columns]

data.duplicated()

Out[8]:

0 False

1 False

2 False

3 False

4 False

4405 False

4406 False

4407 False

4408 False

4409 False

Length: 4410, dtype: bool

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()
```

```
data1
```

Out[10]:

```
              Age  ...  YearsWithCurrManager
count  4410.000000  ...         4410.000000
mean     36.923810  ...            4.123129
std       9.133301  ...            3.567327
min      18.000000  ...            0.000000
25%      30.000000  ...            2.000000
50%      36.000000  ...            3.000000
75%      43.000000  ...            7.000000
max      60.000000  ...           17.000000

[8 rows x 11 columns]
```

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].median()
```

```
data1
```

Out[12]:

```
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

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mean()
```

```
data1
```

Out[14]:

Age 36.923810

DistanceFromHome 9.192517

Education 2.912925

MonthlyIncome 65029.312925

NumCompaniesWorked 2.694830

PercentSalaryHike 15.209524

TotalWorkingYears 11.279936

TrainingTimesLastYear 2.799320

YearsAtCompany 7.008163

YearsSinceLastPromotion 2.187755

YearsWithCurrManager 4.123129

dtype: float64

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()
```

```
data1
```

Out[16]:

```
   Age  DistanceFromHome  ...  YearsSinceLastPromotion  YearsWithCurrManager
0   35                 2  ...                        0                     2

[1 rows x 11 columns]
```

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].var()
```

```
data1
```

Out[18]:

```
Age                       8.341719e+01
DistanceFromHome          6.569144e+01
Education                 1.048438e+00
MonthlyIncome             2.215480e+09
NumCompaniesWorked        6.244436e+00
PercentSalaryHike         1.338907e+01
TotalWorkingYears         6.056298e+01
TrainingTimesLastYear     1.661465e+00
YearsAtCompany            3.751728e+01
YearsSinceLastPromotion   1.037935e+01
YearsWithCurrManager      1.272582e+01
dtype: float64
```

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].skew()
```

```
data1
```

```
Out[20]:
```

```
Age                    0.413005
DistanceFromHome       0.957466
Education             -0.289484
MonthlyIncome          1.368884
NumCompaniesWorked     1.026767
PercentSalaryHike      0.820569
TotalWorkingYears      1.116832
TrainingTimesLastYear  0.552748
YearsAtCompany         1.763328
YearsSinceLastPromotion  1.982939
YearsWithCurrManager   0.832884
dtype: float64
```

```
data1=data[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked',
'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].kurt()
```

```
data1
```

```
Out[22]:
```

```
Age                   -0.405951
DistanceFromHome      -0.227045
Education             -0.560569
MonthlyIncome          1.000232
NumCompaniesWorked     0.007287
PercentSalaryHike     -0.302638
TotalWorkingYears      0.912936
TrainingTimesLastYear  0.491149
```

YearsAtCompany 3.923864

YearsSinceLastPromotion 3.601761

YearsWithCurrManager 0.167949

dtype: float64

	Mean	Median	Mode	var	skew	kurt
Age	36.9	36.0	35	8.3	0.41	-0.4
Distance from home	9.1	7.0	2	6.5	0.95	-0.22
Education	2.9	3.0	3	1.0	-0.28	-0.5
Monthly Income	65029	49190	23420	2.21	1.36	1.00
NumCompaniesWorked	26	2.0	1	6.2	1.02	0.0
PercentSalaryHike	15.2	14.0	11	1.3	0.8	-0.3
TotalWorkingYears	11.2	10.0	10	6.0	1.11	0.91
TrainingTimesLastYear	2.7	3.0	2	1.6	0.55	0.49
YearsAtCompany	7.0	5.0	5	3.7	1.76	3.92
YearsSinceLastPromotion	2.18	1.0	0	1.03	1.98	3.601
YearsWithCurrManager	4.13	3.0	2	1.2	0.83	0.16

All the above variables show positive skewness

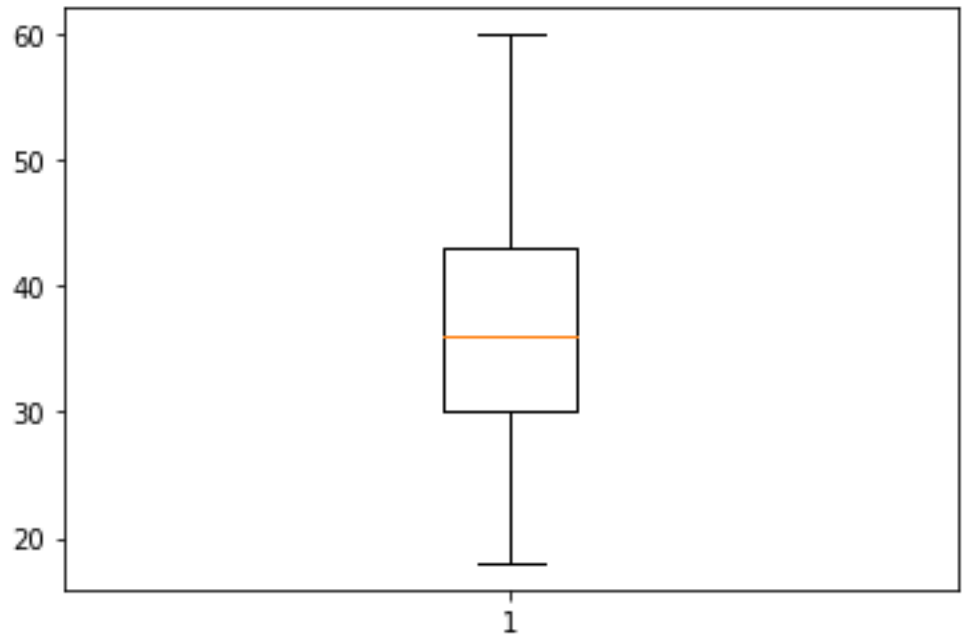
Age,DistanceFromHome,Education,PercentSalaryHike are Platykurtic and remaining all are Leptokurtic

Boxplot

```
box_plot=data.Age
plt.boxplot(box_plot)
```

Out[23]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1bbf15349c8>,
  <matplotlib.lines.Line2D at 0x1bbf1534788>],
 'caps': [<matplotlib.lines.Line2D at 0x1bbf154ac88>,
  <matplotlib.lines.Line2D at 0x1bbf154ad88>],
 'boxes': [<matplotlib.lines.Line2D at 0x1bbf1531dc8>],
 'medians': [<matplotlib.lines.Line2D at 0x1bbf154ae48>],
 'fliers': [<matplotlib.lines.Line2D at 0x1bbf1534ec8>],
 'means': []}
```



Age is normally distributed.It has no outliers

box_plot=data.MonthlyIncome

plt.boxplot(box_plot)

Out[24]:

{'whiskers': [<matplotlib.lines.Line2D at 0x1bbf142b448>,

<matplotlib.lines.Line2D at 0x1bbf142b248>],

'caps': [<matplotlib.lines.Line2D at 0x1bbf1434a48>,

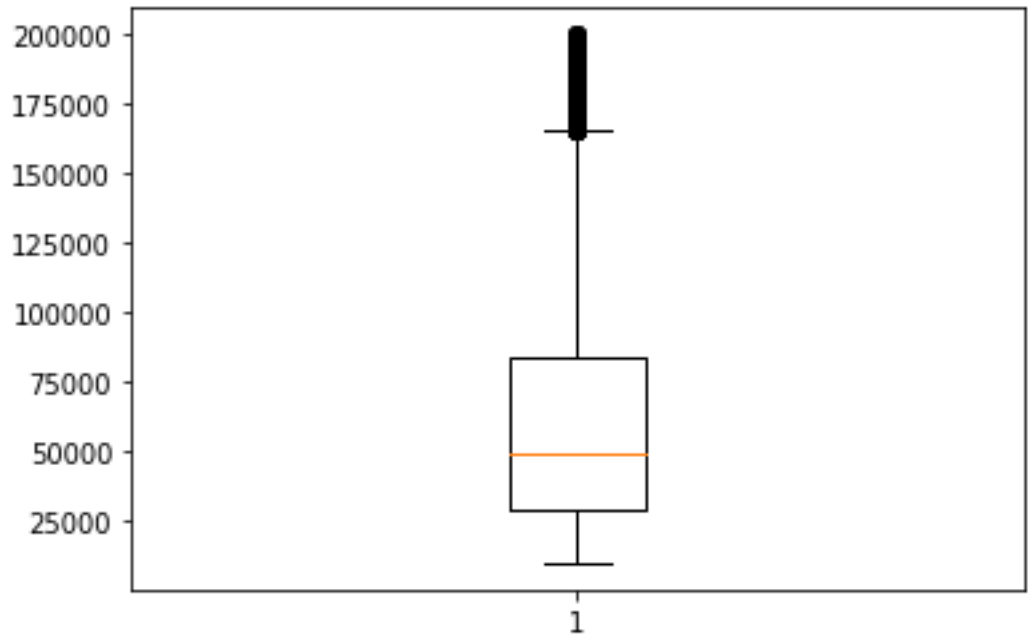
<matplotlib.lines.Line2D at 0x1bbf1434488>],

'boxes': [<matplotlib.lines.Line2D at 0x1bbf14230c8>],

'medians': [<matplotlib.lines.Line2D at 0x1bbf142ff08>],

'fliers': [<matplotlib.lines.Line2D at 0x1bbf12a9948>],

'means': []}



MonthlyIncome is positively skewed and it has outliers

box_plot=data.YearsAtCompany

plt.boxplot(box_plot)

Out[25]:

{'whiskers': [<matplotlib.lines.Line2D at 0x1bbf2d23ec8>,

<matplotlib.lines.Line2D at 0x1bbf2d28e88>],

'caps': [<matplotlib.lines.Line2D at 0x1bbf2d28f88>,

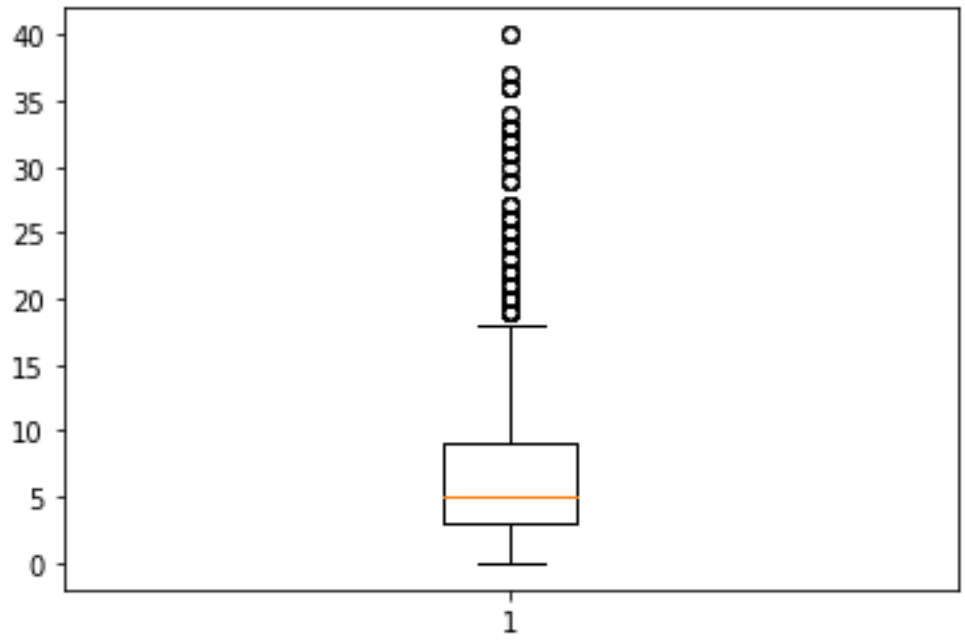
<matplotlib.lines.Line2D at 0x1bbf2d2ce08>],

'boxes': [<matplotlib.lines.Line2D at 0x1bbf2d23d48>],

'medians': [<matplotlib.lines.Line2D at 0x1bbf2d2cf08>],

'fliers': [<matplotlib.lines.Line2D at 0x1bbf2d30d88>],

'means': []}



.YearsAtCompany is positively skewed and it has several outliers

box_plot=data.YearsSinceLastPromotion

plt.boxplot(box_plot)

Out[27]:

{'whiskers': [<matplotlib.lines.Line2D at 0x1bbf2d9be08>,

 <matplotlib.lines.Line2D at 0x1bbf2d9bf08>],

 'caps': [<matplotlib.lines.Line2D at 0x1bbf2d9fdc8>,

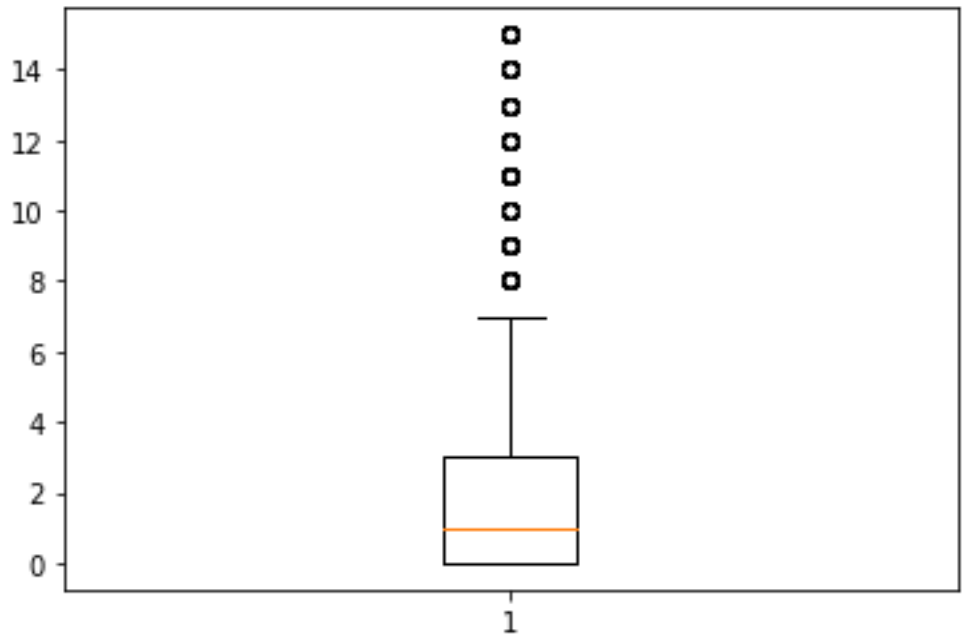
 <matplotlib.lines.Line2D at 0x1bbf2d9fec8>],

 'boxes': [<matplotlib.lines.Line2D at 0x1bbf2d9b4c8>],

 'medians': [<matplotlib.lines.Line2D at 0x1bbf2da3ec8>],

 'fliers': [<matplotlib.lines.Line2D at 0x1bbf2da3fc8>],

 'means': []}



YearsSinceLastPromotion is also positively skewed and has several outliers.