ATTRITION PROJECT

```
import numpy as n
import pandas as pd
import matplotlib.pyplot as m
count=0
var=pd.read_csv("general_data.csv")
var.dropna()
var.drop_duplicates()
```

1.to find mean/min/max/1st,2nd,3rd quartile/number of entries

```
des=var[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
      'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
      'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
      'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
      'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
      'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()
print(des)
```

Index	Age	DistanceFromHome	Education	EmployeeCount	EmployeeID	JobLevel	MonthlyIncome	NumCompaniesWorked	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromo
std	9.13...	8.10503	1.02393	0	1273.2	1.10669	47068.9	2.49889	3.65911	0	0.851883	7.78222	1.28898	6.12514	3.2217
min	18	1	1	1	1	1	10090	0	11	8	0	0	0	0	0
mean	36.9...	9.19252	2.91293	1	2205.5	2.06395	65029.3	2.69483	15.2095	8	0.793878	11.2799	2.79932	7.00816	2.18776
max	60	29	5	1	4410	5	199990	9	25	8	3	40	6	40	15
count	4410	4410	4410	4410	4410	4410	4410	4391	4410	4410	4410	4401	4410	4410	4410
75%	43	14	4	1	3307.75	3	83800	4	18	8	1	15	3	9	3
50%	36	7	3	1	2205.5	2	49190	2	14	8	1	10	3	5	1
25%	30	2	2	1	1103.25	1	29110	1	12	8	0	6	2	3	0

2.to find varience

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

```
print(varience)
```

```
Age                    8.341719e+01
DistanceFromHome       6.569144e+01
Education              1.048438e+00
EmployeeCount          0.000000e+00
EmployeeID             1.621042e+06
JobLevel               1.224760e+00
MonthlyIncome          2.215480e+09
NumCompaniesWorked     6.244436e+00
PercentSalaryHike      1.338907e+01
StandardHours          0.000000e+00
StockOptionLevel       7.257053e-01
TotalWorkingYears      6.056298e+01
TrainingTimesLastYear  1.661465e+00
YearsAtCompany         3.751728e+01
YearsSinceLastPromotion  1.037935e+01
YearsWithCurrManager   1.272582e+01
```

3.to find skewness

```
skew=var[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
      'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
      'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
      'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
      'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
      'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].skew()
print(skew)
```

```
Age                    0.413005
DistanceFromHome       0.957466
Education             -0.289484
EmployeeCount          0.000000
EmployeeID             0.000000
JobLevel               1.024703
```

MonthlyIncome 1.368884

NumCompaniesWorked 1.026767

PercentSalaryHike 0.820569

StandardHours 0.000000

StockOptionLevel 0.968321

TotalWorkingYears 1.116832

TrainingTimesLastYear 0.552748

YearsAtCompany 1.763328

YearsSinceLastPromotion 1.982939

YearsWithCurrManager 0.832884

Inference:

1.Education is negatively skewed and its mean<median

2.Rest all are positively skewed

4.to find kurtosis

```
kurt=var[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
      'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
      'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
      'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
      'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
      'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].kurt()
print(kurt)
```

DistanceFromHome -0.227045

Education -0.560569

EmployeeCount 0.000000

EmployeeID -1.200000

JobLevel 0.395525

MonthlyIncome 1.000232

NumCompaniesWorked 0.007287

PercentSalaryHike -0.302638

StandardHours 0.000000

StockOptionLevel 0.361086

TotalWorkingYears 0.912936

TrainingTimesLastYear 0.491149

YearsAtCompany 3.923864

YearsSinceLastPromotion 3.601761

YearsWithCurrManager 0.167949

Inference:

1.DistanceFromHome, Education, EmployeeID, PercentSalaryHike are all are mesokurtic

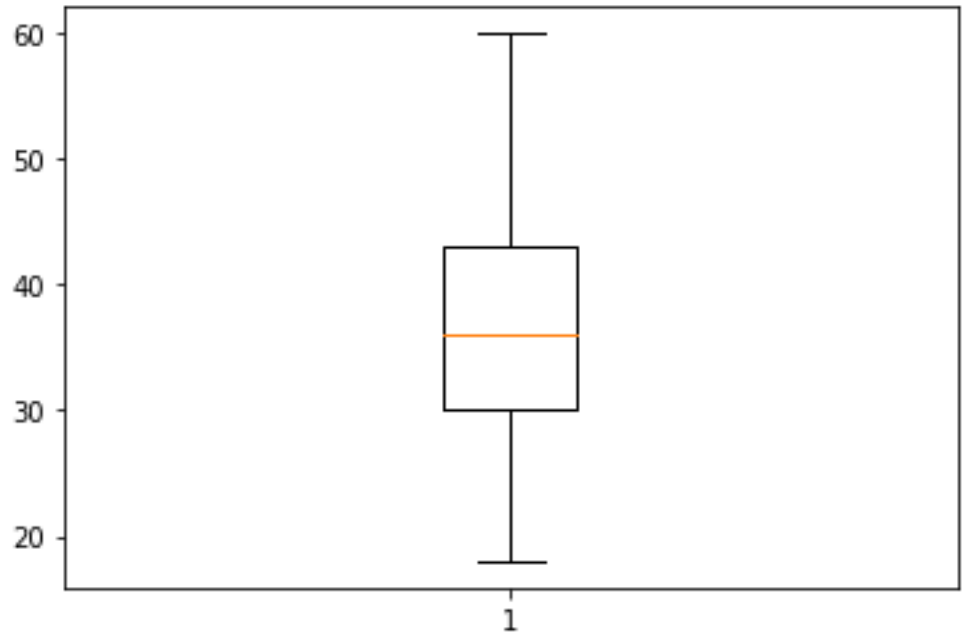
2.the rest all are leptokurtic

Plots

1.Age

plot1=var.Age

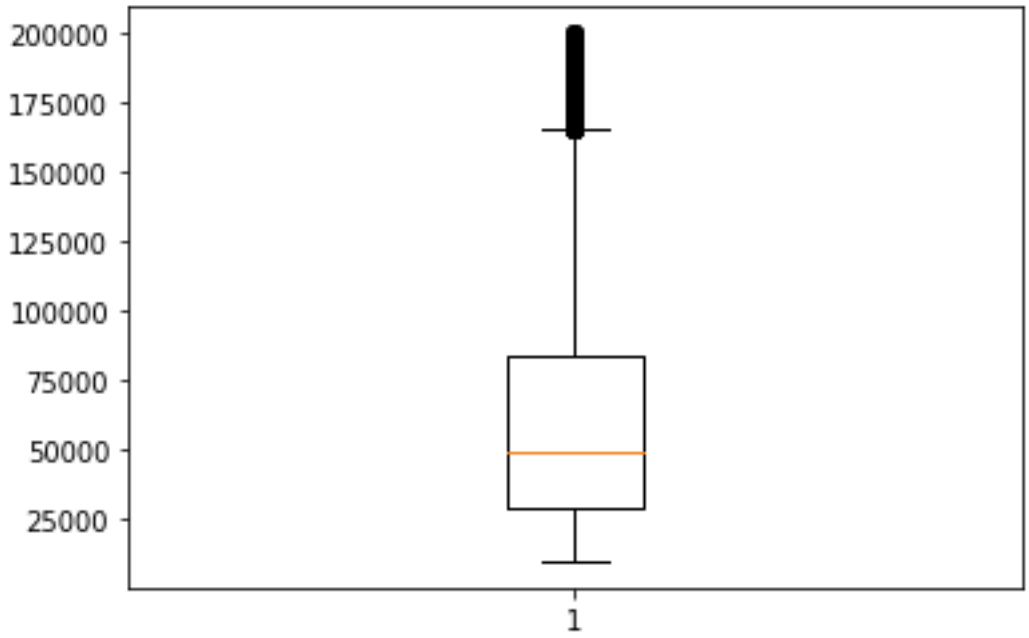
m.boxplot(plot1)



Inference:

age is normally distrubuted

2.Monthly income



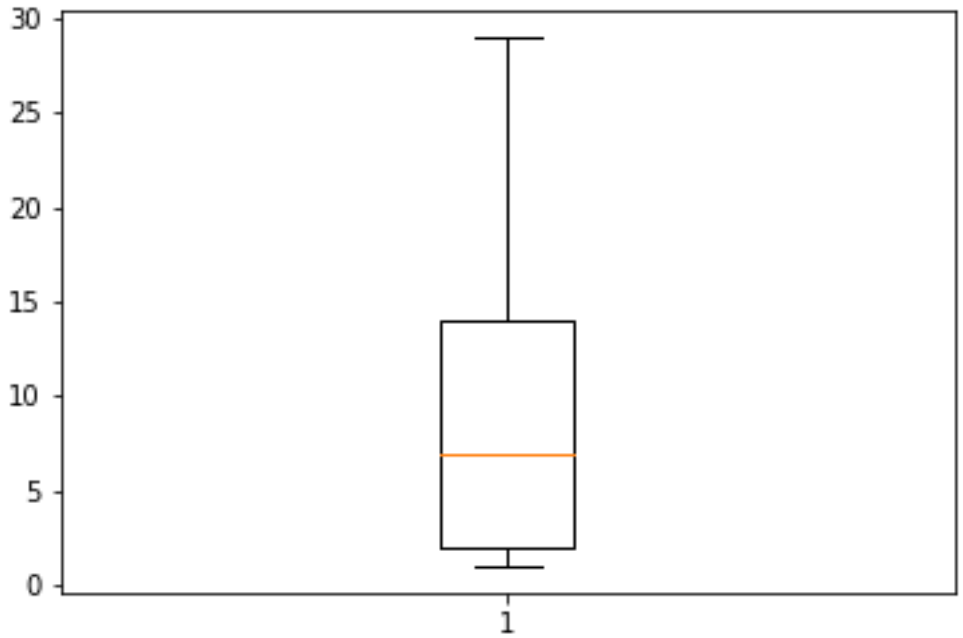
Inference:

Positively skewed with outliers

3.Distane form home

plot3=var.DistanceFromHome

m.boxplot(plot3)



Inference:

Positively skewed without any outliers