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',

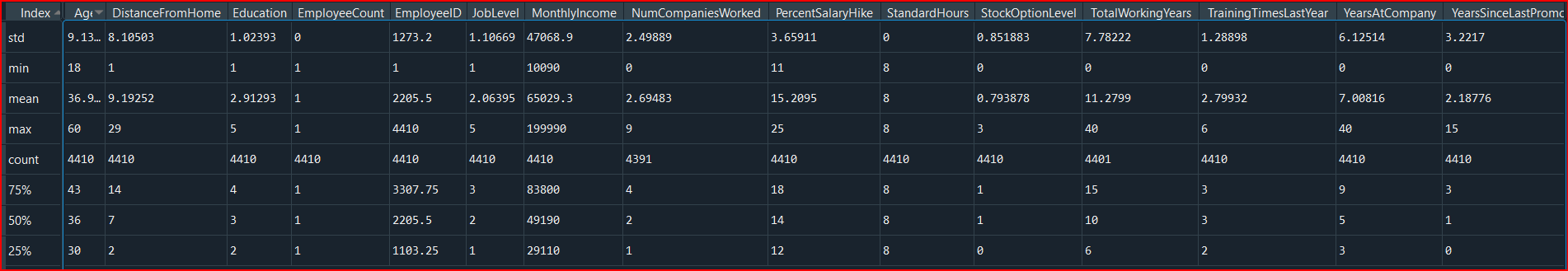
'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()

print(des)



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

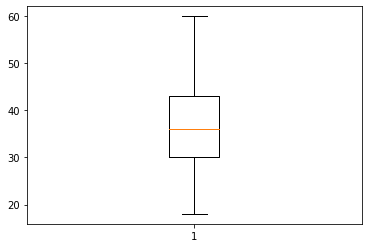
2.the rest all are leptokurtic

Plots

1.Age

plot1=var.Age

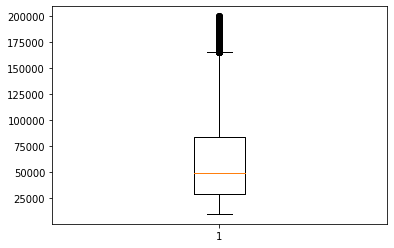
m.boxplot(plot1)



Inference:

age is normally distrubuted

2.Monthly income



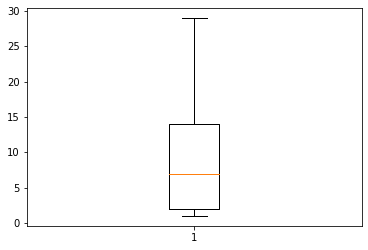
Inference:

Positively skewed with outliers

3.Distane form home

plot3=var.DistanceFromHome

m.boxplot(plot3)



Inference:

Positively skewed without any outliers