ATTRITION ASSIGNMENT

STEP 1 = LAUNCHING :

import pandas as pd

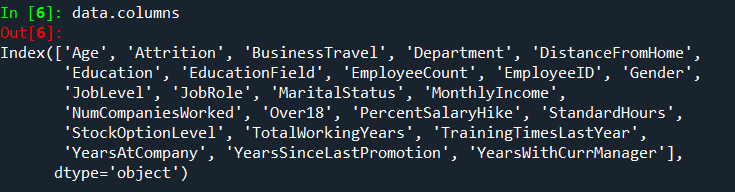
import numpy as n

import matpolib.pyplot as pl

data=pd.read\_csv(“general\_data.csv”)

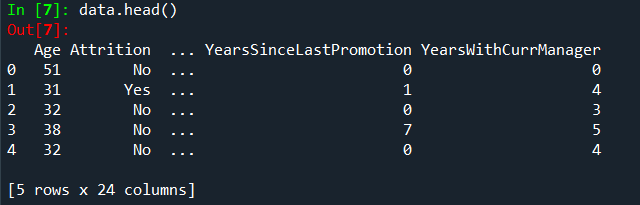
->To find column names .

data.columns



->To find the data of first 5 rows.

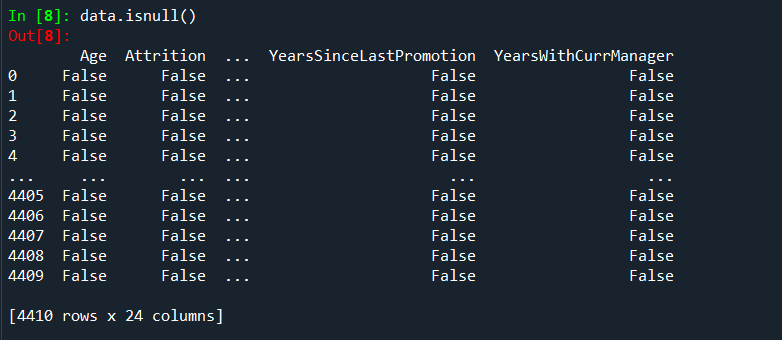
data.head( )



STEP 2 = DATA TREATMENT :

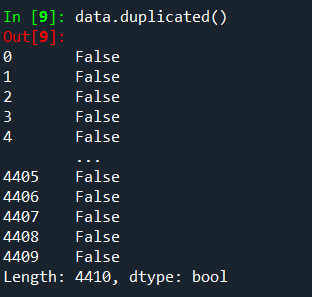
->To find out null values in the table.

data.isnull( )



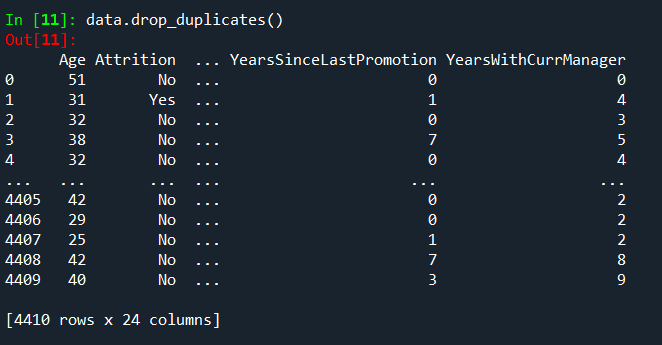
->To find out duplicated values of table.

data.duplicated()



->To drop all duplicated values of the table.

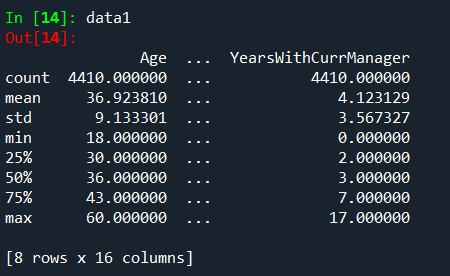
data.drop\_duplicates()



STEP 3 = UNIVARIATE ANALYSIS :

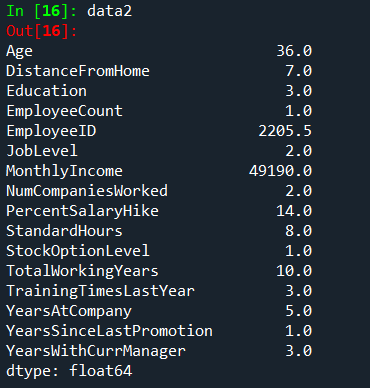
->To describe the whole table.

data1=data[['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()



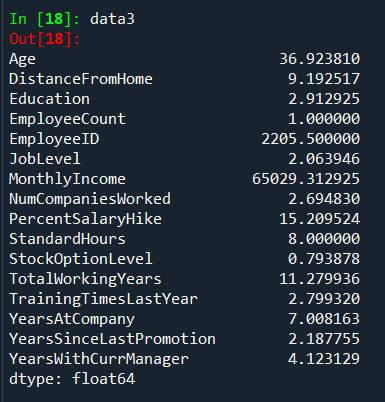
->To find out median of each column .

data2=data[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome','Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender','JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours','StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].median()



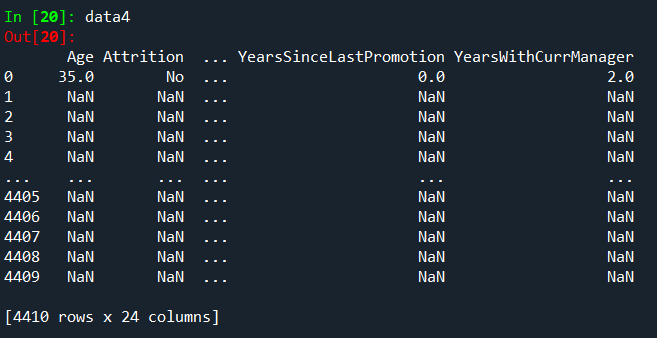
->To find out mean of each column.

data3=data[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome','Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender','JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours','StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].mean()



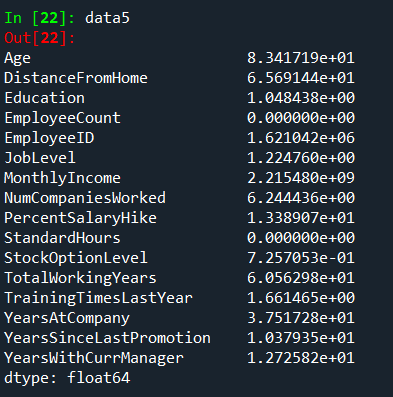
->To find out mode.

data4=data[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome','Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender','JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours','StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()



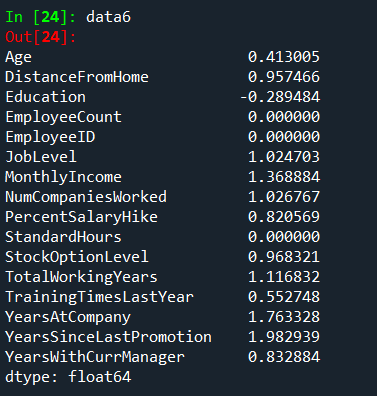
->To find variance of each column.

data5=data[['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()



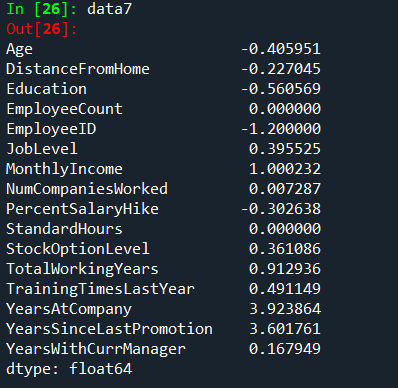
->To find skewness.

data6=data[['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()



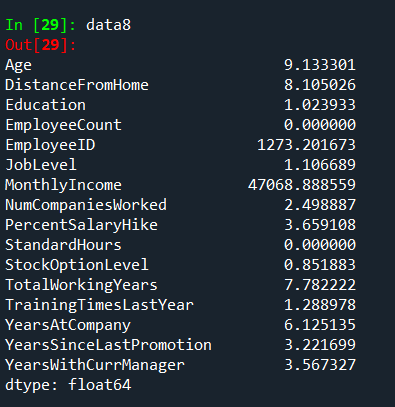
->To find out kurtosis.

data7=data[['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()

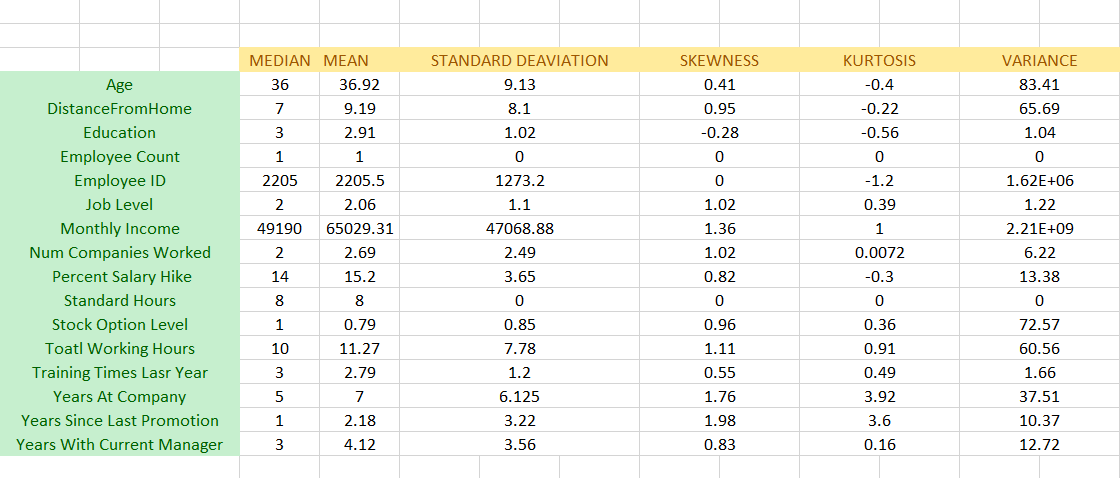


->To find standard deviation .

data8=data[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome','Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender','JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours','StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear','YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].std()



INFERENCE :



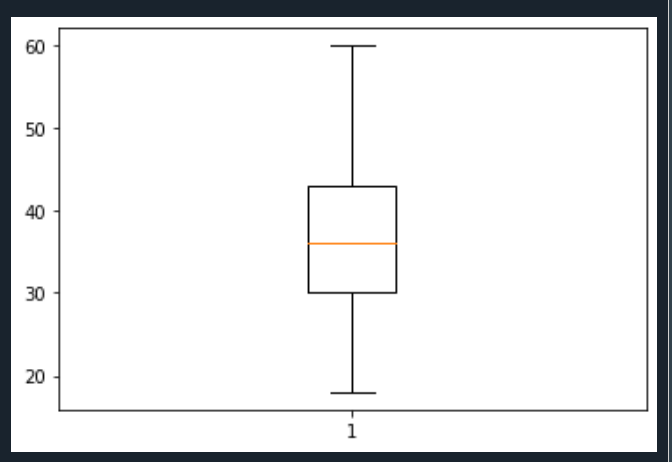
->All the above variables show positive skewness; while Age &Mean\_distance\_from\_home are leptokurtic and all other variables are platykurtic.

OUTLIERS :

There’s no regression found while plotting Age, MonthlyIncome, TotalWorkingYears , YearsAtCompany, etc., on a scatter plot.

box\_plot=data.Age

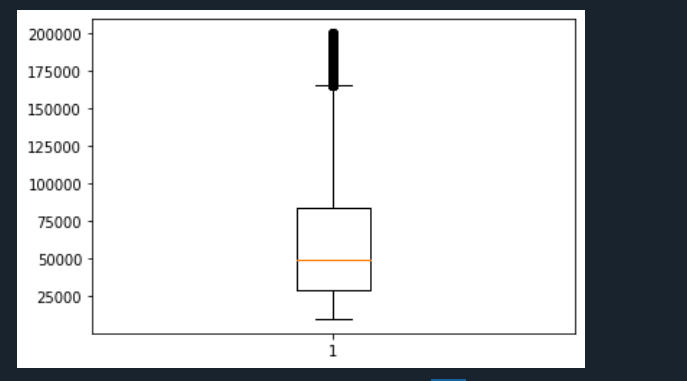
pl.boxplot(box\_plot)



Age is normally distributed without any outliers

box\_plot=data.MonthlyIncome

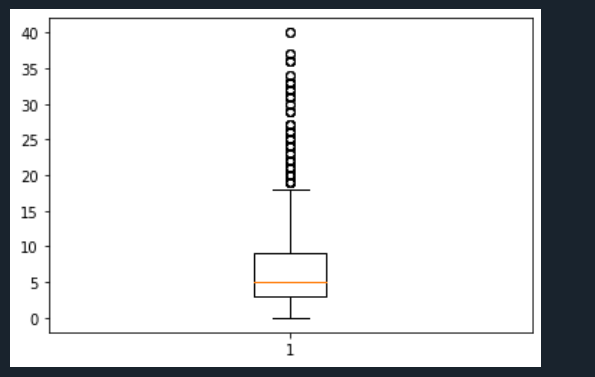
pl.boxplot(box\_plot)



Monthly Income is Right skewed with several outliers

box\_plot=data.YearsAtCompany

pl.boxplot(box\_plot)



Years at company is also Right Skewed with several outliers observed.

# Hypothesis 1

The Employees who all are got Attrited with a Population age of 33, and the sts is 8.8. To determine the sample mean of 20 attrited Employees with a sample mean of 31.

**Hypothesis 2**

The Employees who all are got Attrited with a Population

Of Montly Salary 61682, and sts is 44792. To determine the sample mean of 50 Employees with a sample mean of 57562.

**Hypothesis 3**

The Employees who all are got Attrited with a Population

Of Montly Salary 198590, and sts is 44792. To determine the sample mean of 50 Employees with a sample mean of 57562 ( Find the value of Alpha mostly to satisfy the above scenario)

**Hypothesis 4**

The Employees who all are got Attrited with a Population of Number of Companies Worked 2.96, with a std of

2.67. . To determine the sample mean of 70 Employees with a sample mean of 3.02

**Hypothesis 5**

The Employees who all are got Attrited with a Population of Percentage of Hike 15.48, with a std of

3.77. To determine the sample mean of 100 Employees with a sample mean of 15.5

**Hypothesis 6**

The Employees who all are working with a Population age of 37, and the sts is 8.8. To determine the sample mean of 100 attrited Employees with a sample mean of 32.51.

**Hypothesis 7**

The Employees who all are woking with a Population

Of Montly Salary 65672, and sts is 47472. To determine the sample mean of 250 Employees with a sample mean of 62137.

**Hypothesis 8**

The Employees who all are Working with a Population of Number of Companies Worked 2.64, with a std of

2.46. To determine the sample mean of 170 Employees with a sample mean of 2.88.

**Hypothesis 9**

The Employees who all are Working with a Population of Percentage of Hike 15.15, with a std of

3.63. To determine the sample mean of 150 Employees with a sample mean of 15.54.

**Hypothesis 10**

The Employees who all are Working wit ha Population of TotalWorkingYears 11.86, and STS is 7.76. To determine the sample mean of 200 Employees with a sample mean of 8.04.