**Day -7 Assignment**

import pandas as p

import matplotlib.pyplot as pt

dataset=p.read\_csv('general\_data.csv')

first\_rows=dataset.head()

print(first\_rows) // first 5 rows

***Output :***

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

runfile('C:/Users/BHAGYA/.spyder-py3/untitled0.py', wdir='C:/Users/BHAGYA/.spyder-py3')

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]

\*\*\*\*\*\*\*\*\*\*\*

last\_rows=dataset.tail()

print(last\_rows) //last 5 rows

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

4409 40 No ... 3 9

[5 rows x 24 columns]

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***Mean function***

m=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(m)

***Output:***

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

\*\*\*\*\*\*\*\*

***Median function***

md=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(md)

***Output:***

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

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Mode function***

mo=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(mo)

***Output:***

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 35 No ... 0 2

[1 rows x 13 columns]

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Using describe function***

des=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(des)

***Output:***

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]

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Skew function***

peak=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(peak)

***Output***

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

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Kurtosis function***

peak=dataset[['Age','Attrition', 'Department','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

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

print(peak)

***Output***:

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

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***Outliers:***

pt.boxplot(dataset.MonthlyIncome)

month=dataset['MonthlyIncome'].median()

print(month)

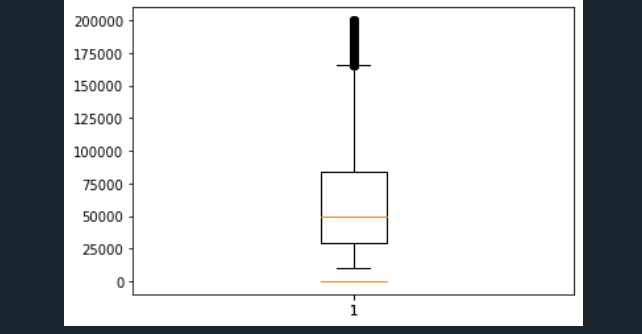
mont=dataset['MonthlyIncome'].mean()

print(mont)

***Output:***

49190.0 // Median

65029.31292517007 // Mean



**Mean is greater than Median so it is Positive Skewness**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ag=dataset['Age'].mean()

print(ag)

a=dataset['Age'].median()

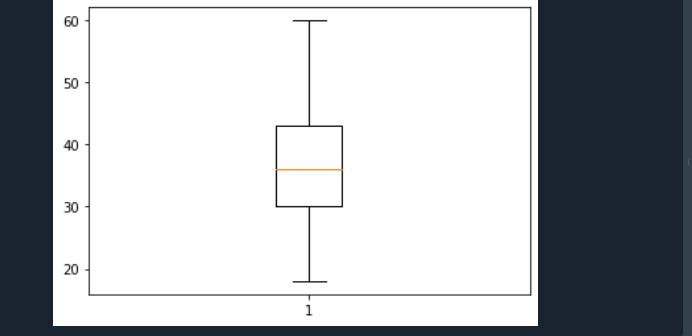
print(a)

pt.boxplot(dataset.Age)

***Output:***

36.923809523809524 // mean

36.0// medain



**Mean is equal to Median so it “Normal Distribution”**

***ScatterPlot***

pt.scatter(dataset.Age,dataset.MonthlyIncome)

****