**CODE:**

**The first code to import the given file and check for the duplicates and for the drop of the duplicates.**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import statistics as sat

dataset = pd.read\_excel("C:\letsupgrage assignment\general data.xlsx")

print(dataset.head())

print(dataset.columns)

print(dataset.isnull())

print(dataset.duplicated())

print(dataset.drop\_duplicates())

**Output:**

*Python 3.7.6 (default, Jan 8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)]*

*Type "copyright", "credits" or "license" for more information.*

*IPython 7.16.1 -- An enhanced Interactive Python.*

*runcell(0, 'C:/letsupgrage assignment/new one spyder.py')*

*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]*

*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')*

*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]*

*0 False*

*1 False*

*2 False*

*3 False*

*4 False*

*4405 False*

*4406 False*

*4407 False*

*4408 False*

*4409 False*

*Length: 4410, dtype: bool*

*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*

*... ... ... ... ...*

*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*

**NEXT CODE:**

**To find the mean for the given data:**

dataset1 = dataset[["Age", "Attrition", "DistanceFromHome", "MonthlyIncome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

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

print(dataset1)

**Output:**

*Age 36.923810*

*DistanceFromHome 9.192517*

*MonthlyIncome 65029.312925*

*NumCompaniesWorked 2.694830*

*PercentSalaryHike 15.209524*

*StandardHours 8.000000*

*TotalWorkingYears 11.279936*

*TrainingTimesLastYear 2.799320*

*YearsAtCompany 7.008163*

*YearsSinceLastPromotion 2.187755*

*YearsWithCurrManager 4.123129*

*dtype: float64*

**NEXT CODE:**

**This code is to find the median for the given data:**

dataset2 = dataset[["Age", "Attrition", "DistanceFromHome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

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

print(dataset2)

**Output:**

*Age 36.0*

*DistanceFromHome 7.0*

*NumCompaniesWorked 2.0*

*PercentSalaryHike 14.0*

*StandardHours 8.0*

*TotalWorkingYears 10.0*

*TrainingTimesLastYear 3.0*

*YearsAtCompany 5.0*

*YearsSinceLastPromotion 1.0*

*YearsWithCurrManager 3.0*

*dtype: float64*

**NEXT CODE:**

**This code tells about the mode of the given data:**

dataset3 = dataset[["Age", "Attrition", "DistanceFromHome", "MonthlyIncome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

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

print(dataset3)

**Output:**

*Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager*

*0 35 No ... 0 2*

*[1 rows x 12 columns]*

**NEXT CODE:**

**This code shows the variance of the given data:**

dataset4 = dataset[["Age", "Attrition", "DistanceFromHome", "MonthlyIncome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

"YearsAtCompany", "YearsSinceLastPromotion", "YearsWithCurrManager"]].var()

print(dataset4)

**Output:**

*Age 8.341719e+01*

*DistanceFromHome 6.569144e+01*

*MonthlyIncome 2.215480e+09*

*NumCompaniesWorked 6.244436e+00*

*PercentSalaryHike 1.338907e+01*

*StandardHours 0.000000e+00*

*TotalWorkingYears 6.056298e+01*

*TrainingTimesLastYear 1.661465e+00*

*YearsAtCompany 3.751728e+01*

*YearsSinceLastPromotion 1.037935e+01*

*YearsWithCurrManager 1.272582e+01*

*dtype: float64*

**NEXT CODE:**

**This code is used to show the skewness of the given data:**

dataset5 = dataset[["Age", "Attrition", "DistanceFromHome", "MonthlyIncome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

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

print(dataset5)

**Output:**

*Age 0.413005*

*DistanceFromHome 0.957466*

*MonthlyIncome 1.368884*

*NumCompaniesWorked 1.026767*

*PercentSalaryHike 0.820569*

*StandardHours 0.000000*

*TotalWorkingYears 1.116832*

*TrainingTimesLastYear 0.552748*

*YearsAtCompany 1.763328*

*YearsSinceLastPromotion 1.982939*

*YearsWithCurrManager 0.832884*

*dtype: float64*

**NEXT CODE:**

**This code tells about the kurtosis for the given data:**

dataset6 = dataset[["Age", "Attrition", "DistanceFromHome", "MonthlyIncome",

"NumCompaniesWorked", "PercentSalaryHike", "StandardHours",

"TotalWorkingYears", "TrainingTimesLastYear",

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

print(dataset6)

**Output:**

*Age -0.405951*

*DistanceFromHome -0.227045*

*MonthlyIncome 1.000232*

*NumCompaniesWorked 0.007287*

*PercentSalaryHike -0.302638*

*StandardHours 0.000000*

*TotalWorkingYears 0.912936*

*TrainingTimesLastYear 0.491149*

*YearsAtCompany 3.923864*

*YearsSinceLastPromotion 3.601761*

*YearsWithCurrManager 0.167949*

*dtype: float64*

**NEXT CODE:**

**This code is helps to show the total description of the given data in the one shot:**

dataset7 = dataset.describe()

print(dataset7)

**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 16 columns]*

**NEXT CODE:**

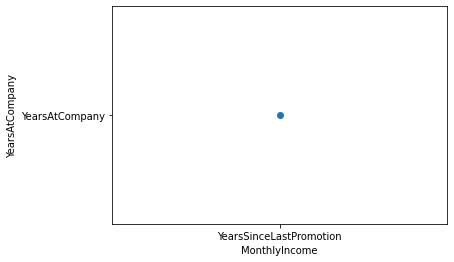
**This code shows the output in the form of plotting. The command used to make this is scatterplot.**

plt.scatter("YearsSinceLastPromotion", "YearsAtCompany")

plt.xlabel("MonthlyIncome")

plt.ylabel("YearsAtCompany")

**Output:**



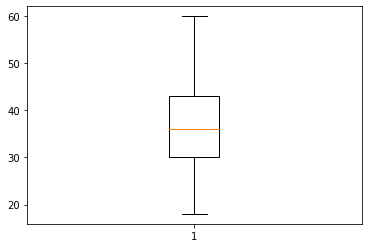
**NEXT CODE:**

**These codes show the output in the form of plotting in the graph. This type is known as the box plot method.**

box\_plot = dataset1.Age

plt.boxplot(box\_plot)

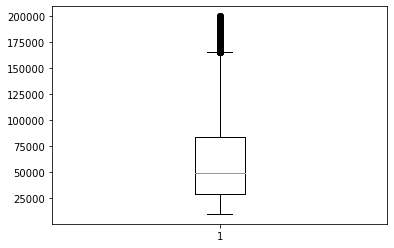
**Output:**



**This code is to tell the boxplot of monthly income:**

box\_plot = dataset1.MonthlyIncome

plt.boxplot(box\_plot)



**LIKE THIS WE CAN ABLE TO DO FOR ALL THE COLUMNS IN THE GIVEN DATA:**

**MY ANALYSIS:**

In this company the age of the workers are working in between the age of 36 to 37. The most frequently ages are 35 years of old workers are working in this company.

The age above 35 years old workers are not going out from the company. So the aged workers below the 30 are looking for the other jobs in Some other Company. This people are getting the salary of approximately in 50,000 per month.

The monthly income for the workers has a lot of out-layers these out-layers are based on their position in the company. And their experience at the company.

To reduce the value of the attrition value in the company, the company should give some gifts and some expenses to the workers to make them to feel good to work in the company. By the way the company should concentrate to take workers with minimum experience and a good salary to maintain the worker to stay in the company. If the worker is working in the company when all the steps are given, the company has the responsibility to give concealing to make the worker to change his or her mode. At last the worker is leaving the company and taking another job from another company, ask him to give some feedback and reviews from the worker to maintain company attrition level in low.