**Attrition Assignment Solution**

**Step1: Launching**

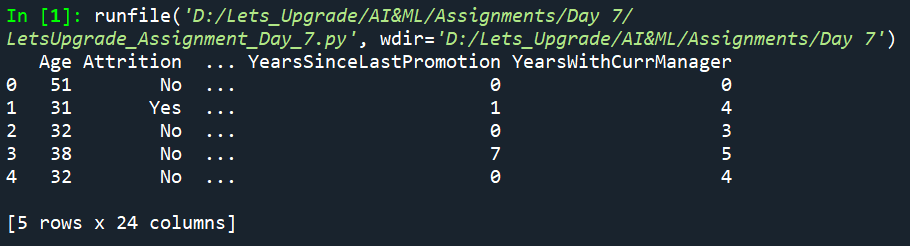
import pandas as pd

import numpy as np

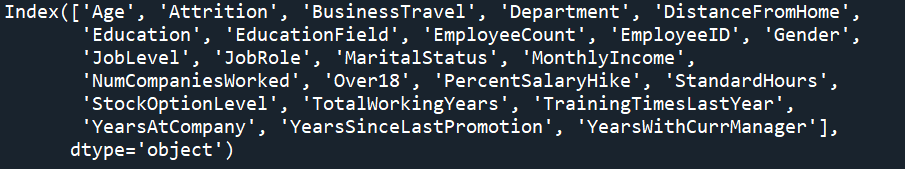
import matplotlib.pyplot as plt

dataset1=pd.read\_excel("my\_data.xlsx",sheet\_name=0)

print(dataset1.head())

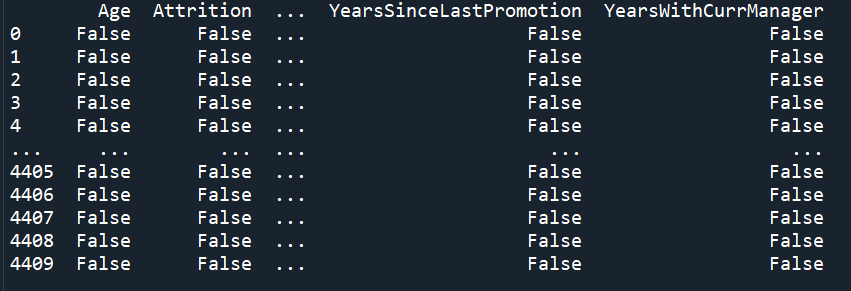


print(dataset1.columns)

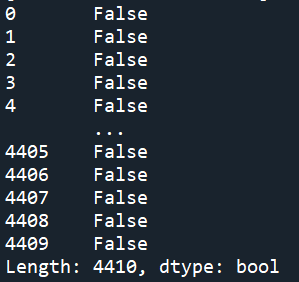


**Step2: Data Treatment**

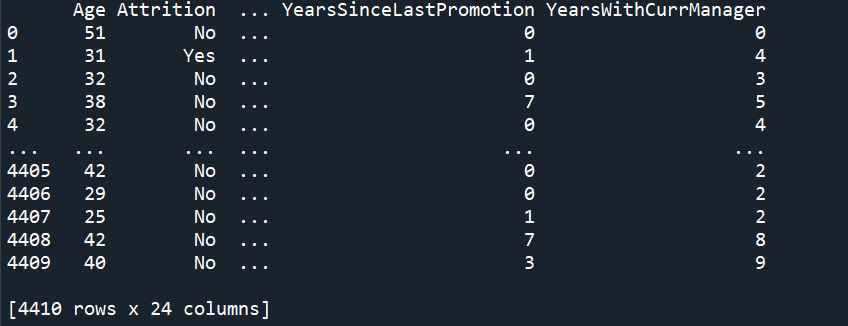
print(dataset1.isnull())



print(dataset1.duplicated())



print(dataset1.drop\_duplicates())



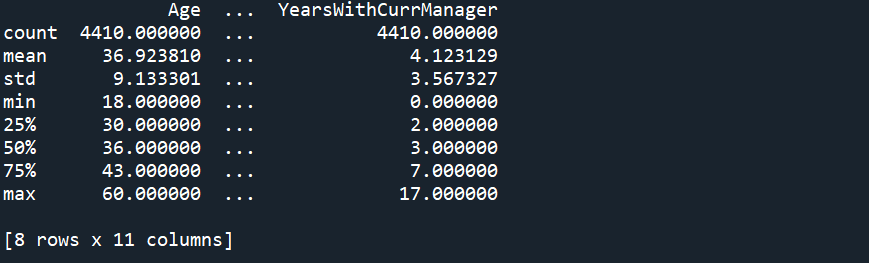
**Step 3 – Univariate Analysis:**

dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

print(dataset3)

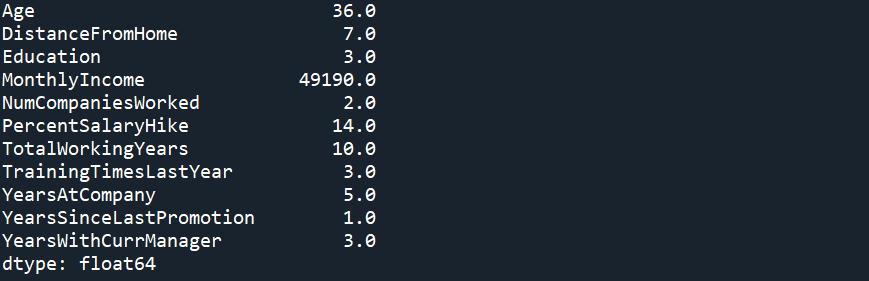


dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

print(dataset3)

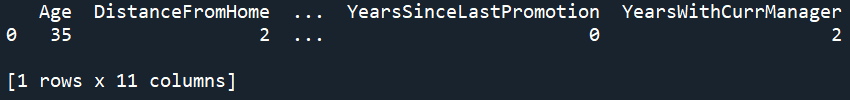


dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

print(dataset3)

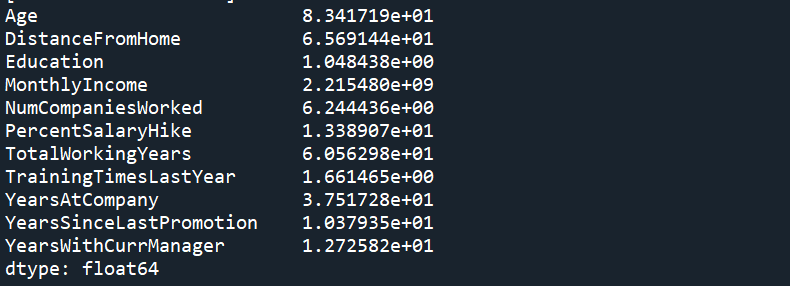


dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

print(dataset3)

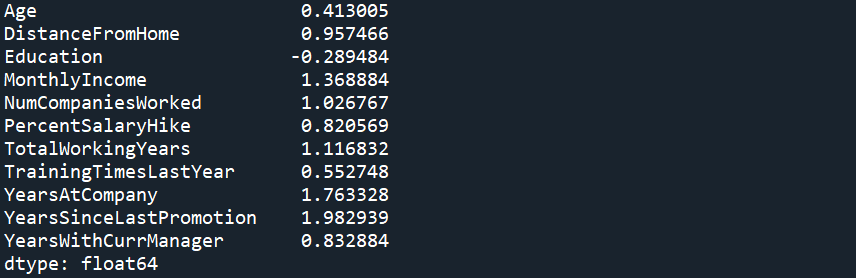


dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

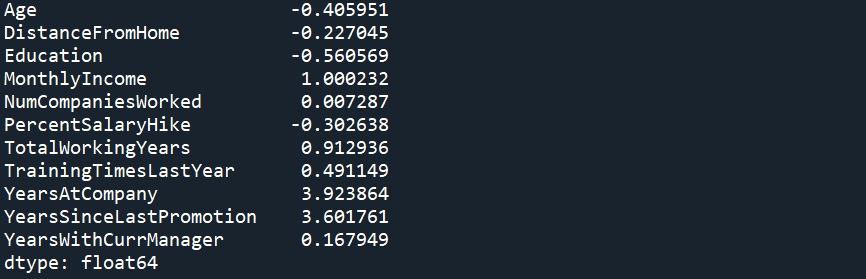
print(dataset3)



dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

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

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

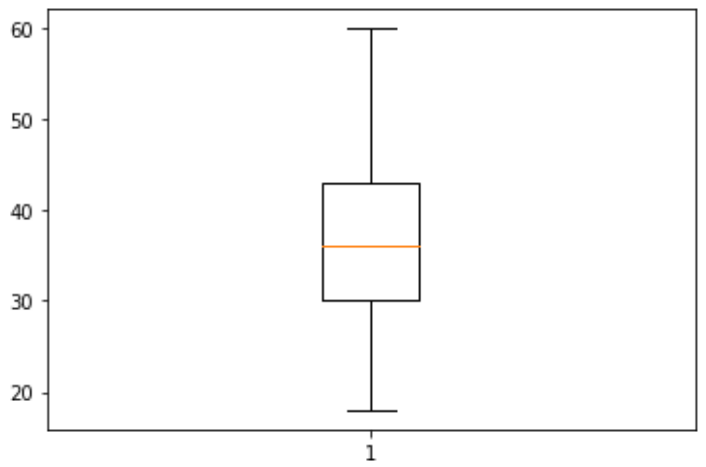
print(dataset3) 

**Inference from the analysis:**

1. Education has a negative skewness: Most of the people need to improve their knowledge.
2. Age, Distance from home, education and percent salary hike has negative kurtosis (platyokurtic)

box\_plot=dataset1.Age

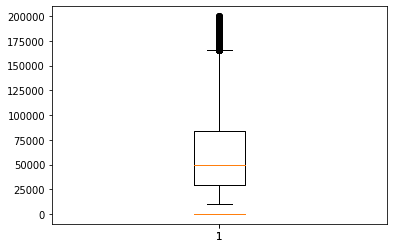
plt.boxplot(box\_plot)



Age is normally distributed with no outliers

box\_plot=dataset1.MonthlyIncome

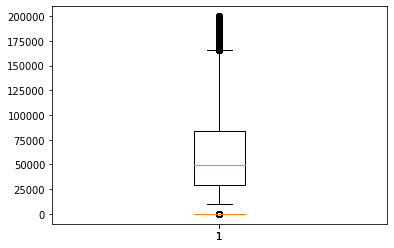
plt.boxplot(box\_plot)



Monthly Income is Right skewed with several outliers

box\_plot=dataset1.YearsAtCompany

plt.boxplot(box\_plot)



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