**Attrition Assignment Solution**

**Step1 - Launching**

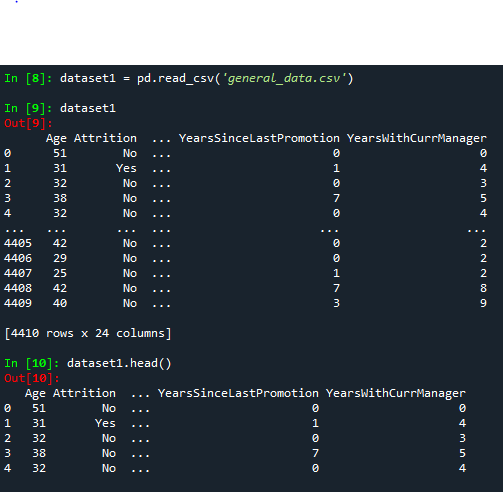
import pandas as pd

import numpy as np

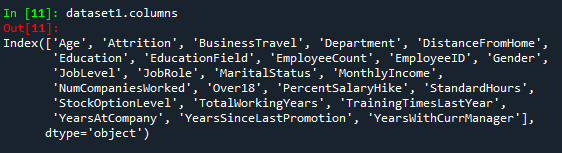
import matplotlib.pyplot as plt

dataset1=pd.read\_excel(‘general\_data.xlsx', sheet\_name=0)

dataset1.head()

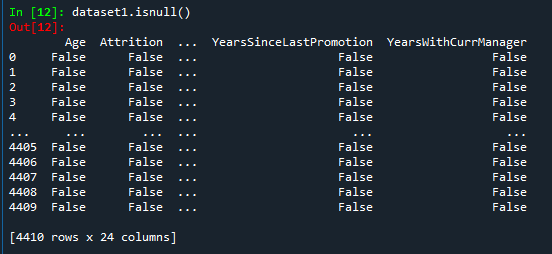


dataset1.columns

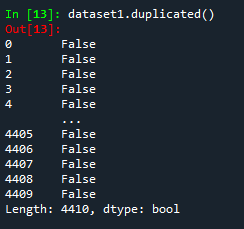


**Step 2 - Data Treatment:**

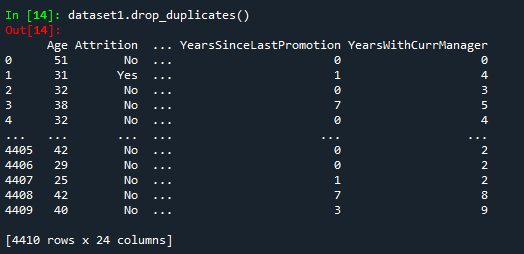
dataset1.isnull()



dataset1.duplicated()



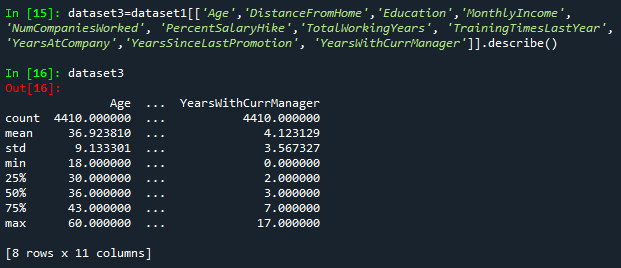
dataset1.drop\_duplicates()



**Step 3 – Univariate Analysis:**

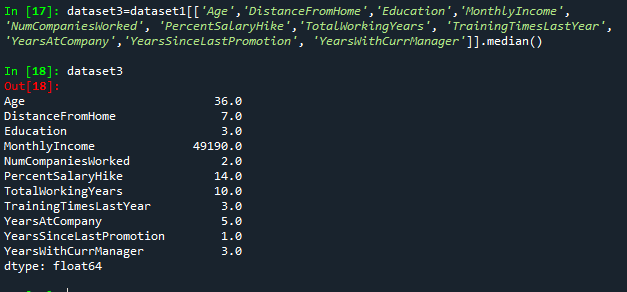
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()

dataset3



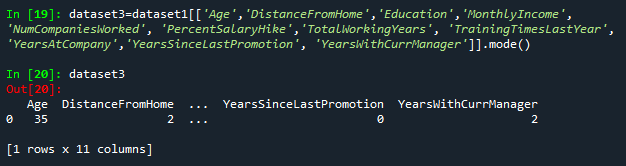
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].median()

dataset3



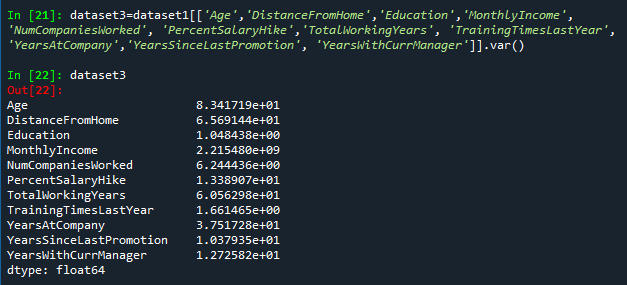
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()

dataset3



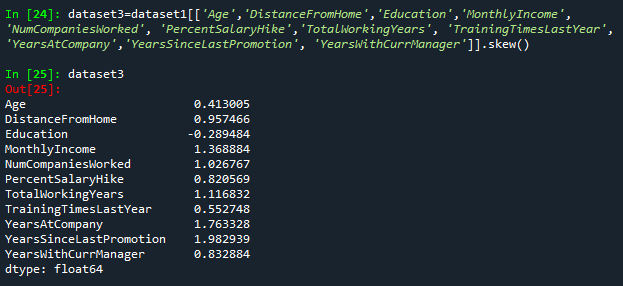
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].var()

dataset3



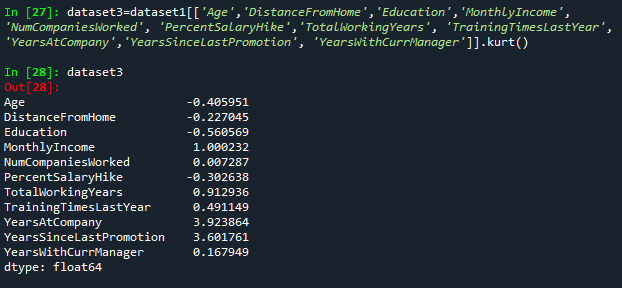
dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].skew()

dataset3



dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].kurt()

dataset3



**Inference from the analysis:**

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

• The Mean\_Monthly\_Income’s IQR is at 54K suggesting company wide attrition across all income bands

• Mean age forms a near normal distribution with 13 years of IQR

**Outliers:**

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

box\_plot=dataset1.Age

plt.boxplot(box\_plot)

