**Attrition Assignments**

**Step1: Launcing :**

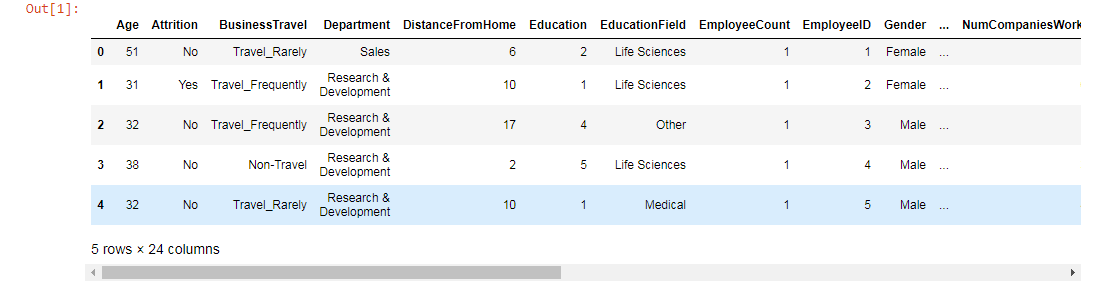
**import pandas as pd**

**import matplotlib.pyplot as plt**

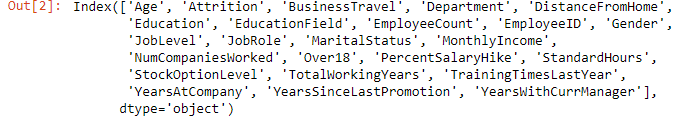
**import numpy as np**

**dataset=pd.read\_csv("general\_data.csv")**

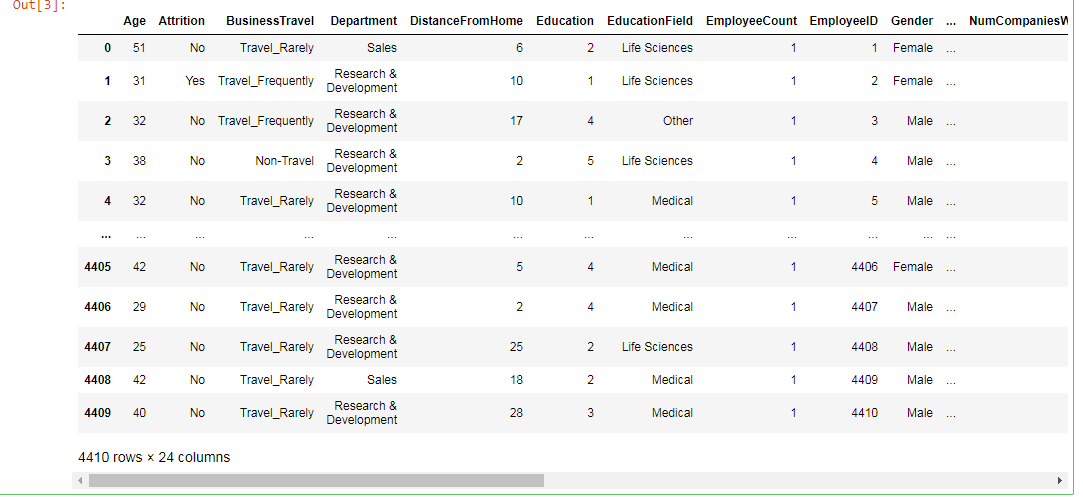
**dataset.head()**

****

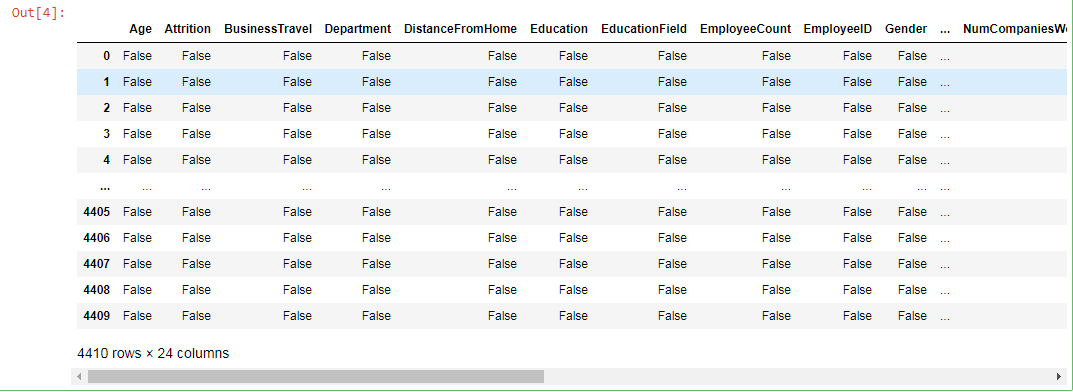
**dataset.columns**

****

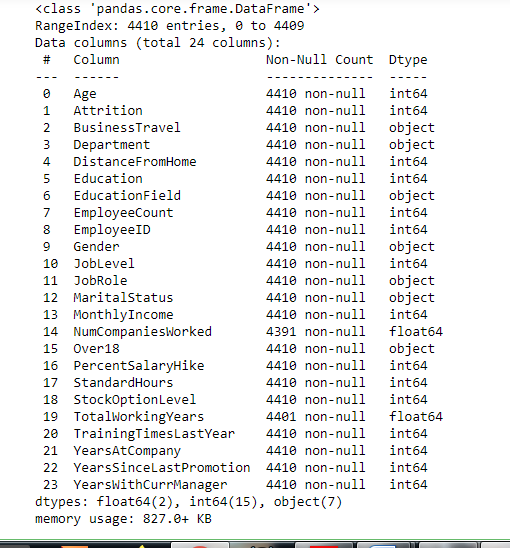
**Dataset**

****

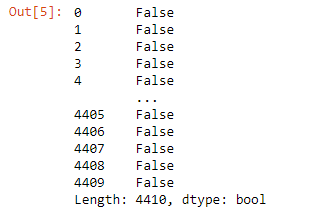
**dataset.isnull()**

****

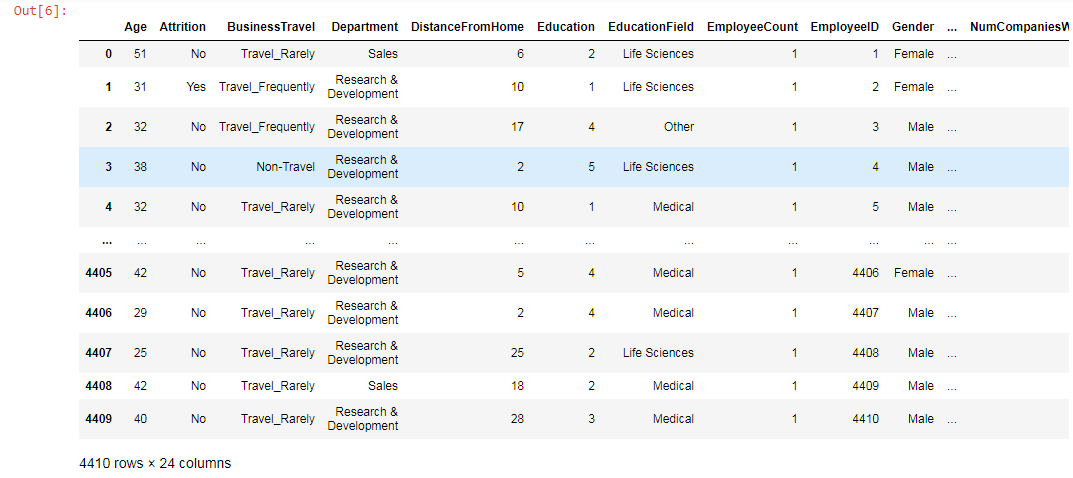
**dataset.info()**

****

**dataset.duplicated()**

****

**dataset.drop\_duplicates()**

****

**dataset2=dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

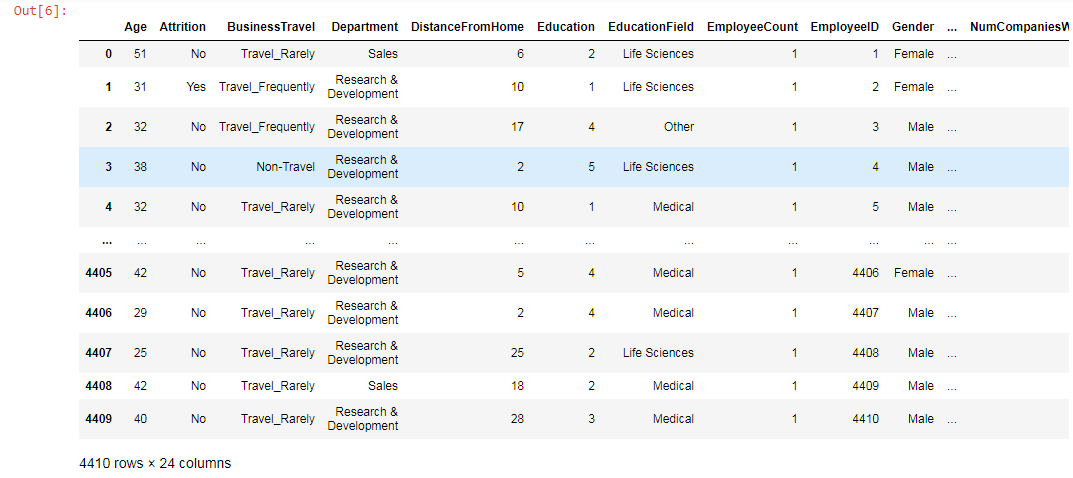
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset3=dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

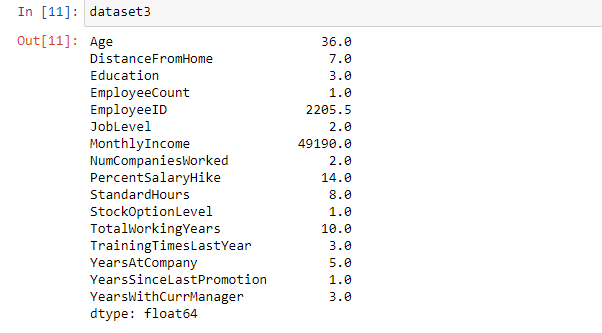
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset4=dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

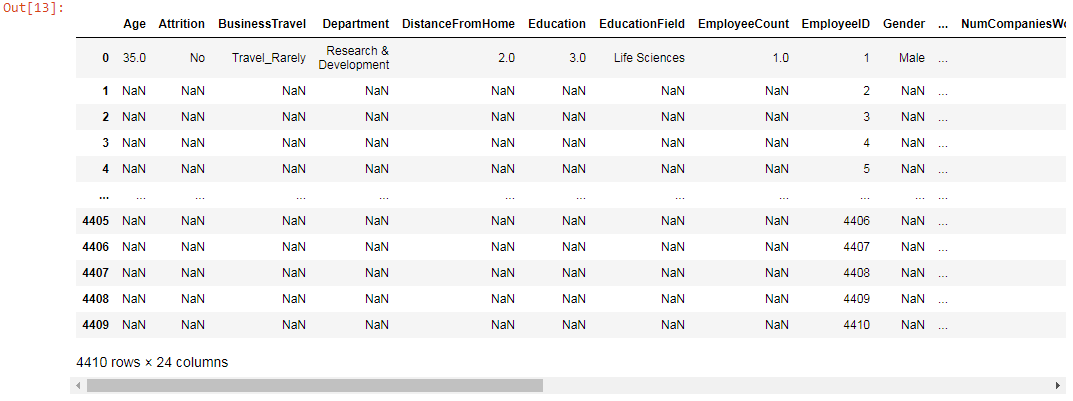
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

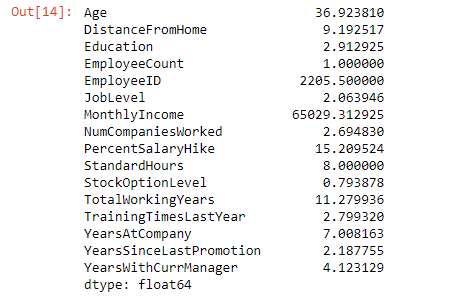
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

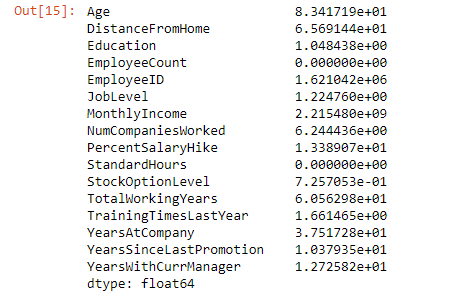
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

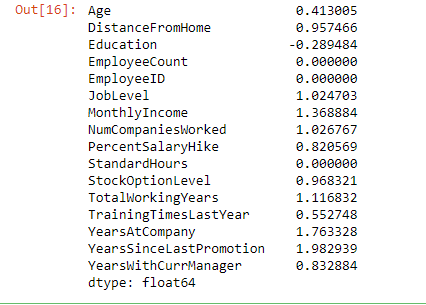
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

****

**dataset[['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',**

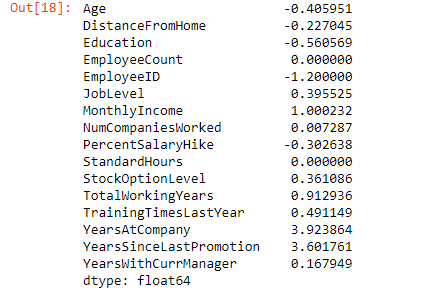
**'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',**

**'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',**

**'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

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

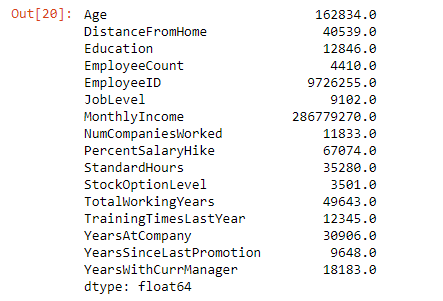
****

**dataset[['Age', 'DistanceFromHome','Education', 'EmployeeCount', 'EmployeeID',**

**'JobLevel', 'MonthlyIncome','NumCompaniesWorked', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

**'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].sum()**

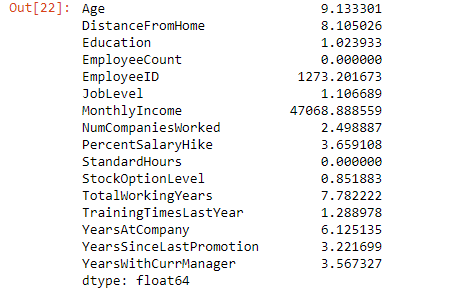
****

**dataset[['Age', 'DistanceFromHome','Education', 'EmployeeCount', 'EmployeeID',**

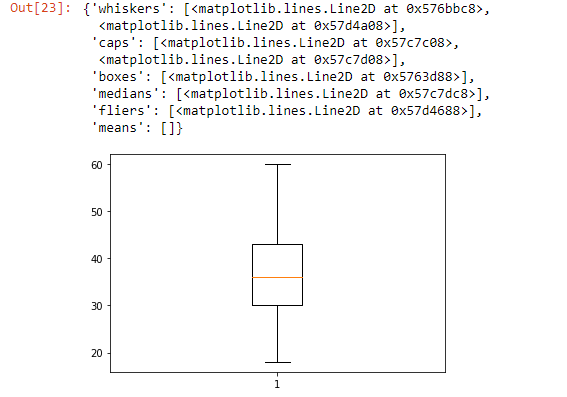
**'JobLevel', 'MonthlyIncome','NumCompaniesWorked', 'PercentSalaryHike', 'StandardHours',**

**'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',**

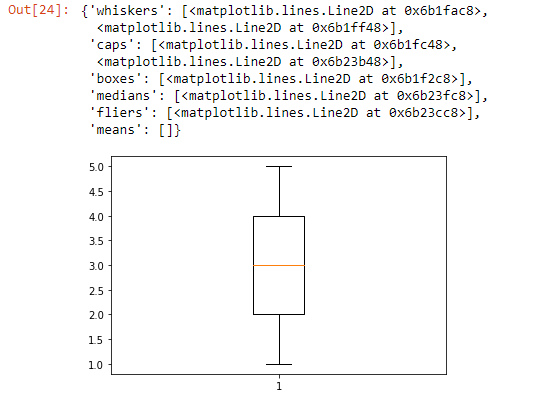
**'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].std()**

****

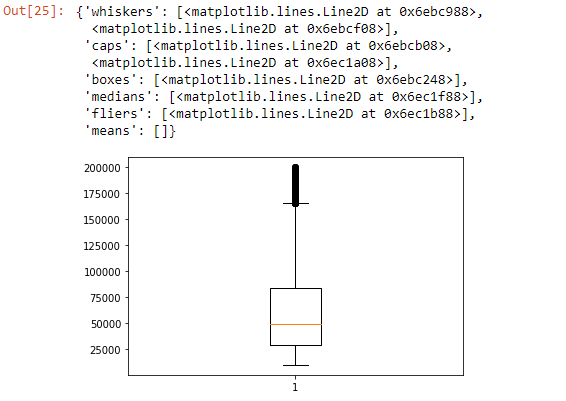
**plt.boxplot(dataset.Age)**

****

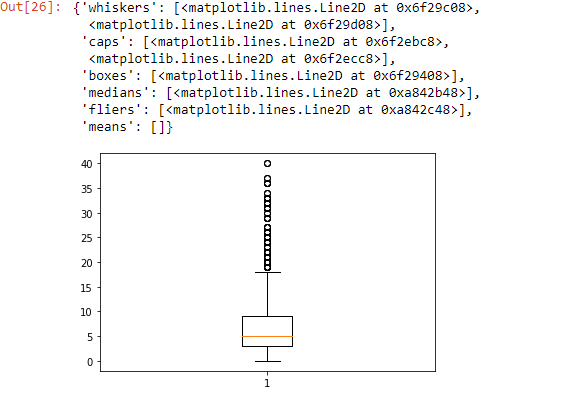
**plt.boxplot(dataset.Education)**

****

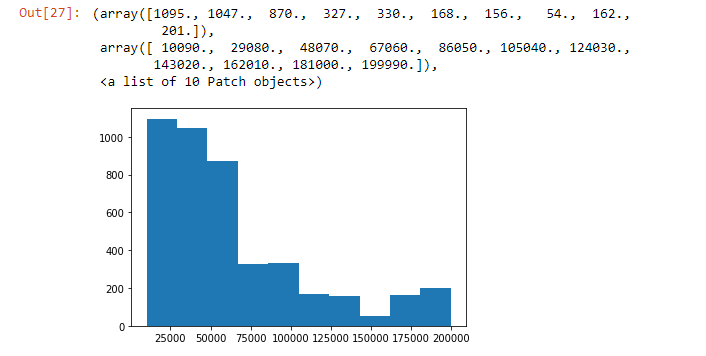
**plt.boxplot(dataset.MonthlyIncome)**

****

**plt.boxplot(dataset.YearsAtCompany)**

****

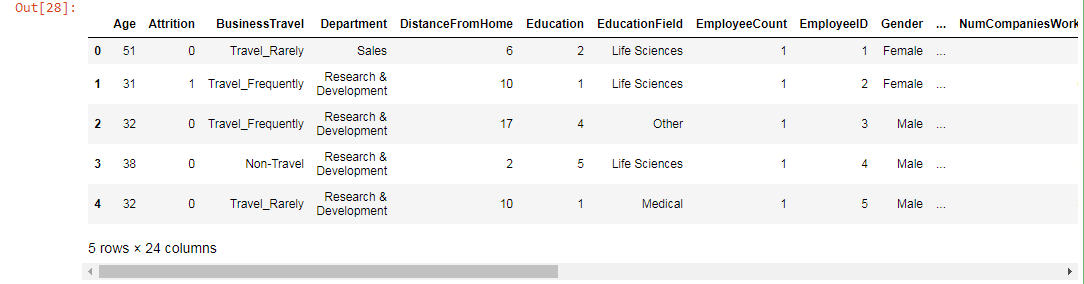
**plt.hist(dataset.MonthlyIncome)**

****

**# Attrition with some other fields:**

**dataset["Attrition"]=dataset["Attrition"].map({"Yes":1,"No":0})**

**dataset.head()**

****

**Check for Hypothesis:**

**import pandas as pd**

**dataset\_yes= pd.read\_excel('general\_data.xlsx',sheet\_name="Attrition\_Yes")**

**dataset\_no= pd.read\_excel('general\_data.xlsx',sheet\_name="Attrition\_No")**

**from scipy.stats import mannwhitneyu**

**stats,p = mannwhitneyu(dataset\_yes.Age, dataset\_no.Age)**

**print("Stats=",stats,"And P=",p)**

**if p<0.05:**

**print("Hypothesis is rejectected.")**

**else:**

**print("Alternate hypothesis accepted")**

Stats= 252760.5 And P= 8.533914665285945e-273

Hypothesis is rejectected.

**# Attrition with income**

**stats,p = mannwhitneyu(dataset\_yes.PercentSalaryHike, dataset\_no.PercentSalaryHike)**

**print("Stats=",stats,"And P=",p)**

**if p<0.05:**

**print("Hypothesis is rejectected.")**

**else:**

**print("Alternate hypothesis accepted")**

Stats= 273787.5 And P= 5.0375111124363596e-263

Hypothesis is rejectected.

**# Attrition with Distance fromHome**

**stats,p = mannwhitneyu(dataset\_yes.DistanceFromHome, dataset\_no.DistanceFromHome)**

**print("Stats=",stats,"And P=",p)**

**if p<0.05:**

**print("Hypothesis is rejectected.")**

**else:**

**print("Alternate hypothesis accepted")**

Stats= 273787.5 And P= 5.0375111124363596e-263

Hypothesis is rejectected.

**stats,p = mannwhitneyu(dataset\_yes.YearsWithCurrManager, dataset\_no.YearsWithCurrManager)**

**print("Stats=",stats,"And P=",p)**

**if p<0.05:**

**print("Hypothesis is rejectected.")**

**else:**

**print("Alternate hypothesis accepted")**

Stats= 273052.5 And P= 2.3104478144262257e-263

Hypothesis is rejectected.

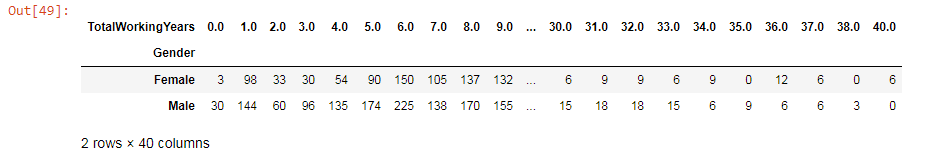
**# Chi square test:**

**dataset1=pd.read\_csv("general\_data.csv")**

**from scipy.stats import chi2\_contingency**

**chitable=pd.crosstab(dataset1.Gender,dataset.TotalWorkingYears)**

**chitable**

****

**Statistical test:**

**from scipy.stats import ttest\_ind**

**stats,p=ttest\_ind(dataset\_yes.YearsWithCurrManager, dataset\_no.YearsWithCurrManager)**

**print(stats,p)**

-37.694238548131764 6.582076911148772e-271

**stats,p=ttest\_ind(dataset\_yes.MonthlyIncome, dataset\_no.MonthlyIncome)**

**print(stats,p)**

-1623.8364867084538 0.0

**stats,p=ttest\_ind(dataset\_yes.YearsAtCompany, dataset\_no.YearsAtCompany)**

**print(stats,p)**

-37.694238548131764 6.582076911148772e-271

**# Correlation,Unsupervised Learning:**

**from scipy.stats import pearsonr**

**stats,p=pearsonr(dataset.Attrition,dataset.Age)**

**print(stats,p)**

-0.15920500686577965 1.996801615886744e-26

**stats,p=pearsonr(dataset.Attrition,dataset.DistanceFromHome)**

**print(stats,p)**

-0.009730141010179674 0.5182860428050771

**stats,p=pearsonr(dataset.Attrition,dataset.JobLevel)**

**print(stats,p)**

-0.010289713287495042 0.49451717271828405

**stats,p=pearsonr(dataset.Attrition,dataset.PercentSalaryHike)**

**print(stats,p)**

0.03253259489105349 0.030743386433355353

**stats,p=pearsonr(dataset.YearsAtCompany,dataset.YearsAtCompany)**

**print(stats,p)**

1.0 0.0