**Step1: Launching**

import pandas as pd

dataset=pd.read\_excel("general\_data1.xlsx",sheet\_name=0)

**dataset.head()**

Out[4]:

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]

**dataset.columns**

Out[5]:

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')

**Step 2 - Data Treatment:**

**dataset.isnull()**

Out[6]:

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]

**dataset.dropna(inplace=True)**

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

Attr\_no= dataset[dataset['Attrition']== 0]

Attr\_no

Out[13]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 51 0 ... 0 0

2 32 0 ... 0 3

3 38 0 ... 7 5

4 32 0 ... 0 4

5 46 0 ... 7 7

... ... ... ... ...

4404 29 0 ... 1 5

4405 42 0 ... 0 2

4406 29 0 ... 0 2

4407 25 0 ... 1 2

4408 42 0 ... 7 8

[3677 rows x 24 columns]

Attr\_yes= dataset[dataset['Attrition']== 1]

Attr\_yes

Out[15]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

1 31 1 ... 1 4

6 28 1 ... 0 0

13 47 1 ... 9 9

28 44 1 ... 0 0

30 26 1 ... 0 2

... ... ... ... ...

4381 29 1 ... 0 1

4386 33 1 ... 0 4

4388 33 1 ... 1 7

4391 32 1 ... 1 2

4402 37 1 ... 0 0

[705 rows x 24 columns]

**Mannwhitney test:**

**Attirition and Distance From Home:**

from scipy.stats import mannwhitneyu

stats,p=mannwhitneyu(Attr\_yes.DistanceFromHome,

Attr\_no.DistanceFromHome)

print(stats,p)

1295261.0 0.488538986087403

H0: There is no significant differences in the Distance from Home between those who left and those who stayed in the company.

Ha: There is significant differences in the Distance from Home between those who left and those who stayed in the company.

**Since p > 0.05, we reject the Alternate Hypothesis and null hypothesis is accepted.**

**Attirition and Monthly Income:**

stats, p = mannwhitneyu(Attr\_yes.MonthlyIncome,Attr\_no.MonthlyIncome)

print(stats,p)

1249573.5 0.06508807631576838

H0: There is no significant differences in the Monthly Income between those who left and those who stayed in the company.

Ha: There is significant differences in the Monthly Income between those who left and those who stayed in the company.

**Since p > 0.05, we reject the Alternate Hypothesis and null hypothesis is accepted.**

**Attirition and PercentSalaryHike:**

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

print(stats,p)

1231873.5 0.017810794960084964

H0: There is no significant differences in the percentage of salary hike between those who left and those who stayed in the company.

Ha: There is significant differences in the percentage of salary hike between those who left and those who stayed in the company.

**Since p < 0.05, alternative hypothesis is accepted and null hypothesis is rejected.**

**Chi Square Test:**

from scipy.stats import chi2\_contingency

MaritalStatus = {'Married': 0,'Single' : 1,'Divorced' : 2}

**Attrition and MaritalStatus:**

chitable = pd.crosstab (dataset.Attrition,dataset.MaritalStatus)

print (chitable)

MaritalStatus Divorced Married Single

Attrition

0 872 1756 1049

1 98 251 356

stats, p, dof, expeted = chi2\_contingency (chitable)

print(stats,p)

133.85785802925156 8.573051828219379e-30

H0: There is no dependency between Attrition and Marital Status.

Ha: There is dependency between Attrition and Marital Status.

**Since p < 0.05, alternative hypothesis is accepted and null hypothesis is rejected.**

**Attrition and Department:**

chitable = pd.crosstab(dataset.Attrition, dataset.Department)

stats, p, dof, expeted = chi2\_contingency (chitable)

print (chitable)

Department Human Resources Research & Development Sales

Attrition

0 132 2416 1129

1 55 449 201

print(stats,p)

25.89432541916022 2.382970570769315e-06

H0: There is no dependency between Attrition and Department.

Ha: There is dependency between Attrition and Department.

**Since p < 0.05, alternative hypothesis is accepted and null hypothesis is rejected.**