**Statistical Tests**

**Step 1: Launching and Data Treatment:**

import pandas as pd

dataset = pd.read\_excel("Attrition Analysis Data.xlsx", sheet\_name = 0)

dataset.head()

Out[3]:

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[4]:

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')

dataset1 =dataset.drop\_duplicates()

dataset1 =dataset1.dropna()

dataset\_yes = pd.read\_excel("Attrition Analysis Data.xlsx", sheet\_name = 'Attrition\_yes')

dataset\_yes.head()

Out[8]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 31 Yes ... 1 4

1 28 Yes ... 0 0

2 47 Yes ... 9 9

3 44 Yes ... 0 0

4 26 Yes ... 0 2

[5 rows x 24 columns]

dataset\_yes = dataset\_yes.drop\_duplicates()

dataset\_yes = dataset\_yes.dropna()

dataset\_no = pd.read\_excel("Attrition Analysis Data.xlsx", sheet\_name = 'Attrition\_no')

dataset\_no.head()

Out[12]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 51 No ... 0 0

1 32 No ... 0 3

2 38 No ... 7 5

3 32 No ... 0 4

4 46 No ... 7 7

[5 rows x 24 columns]

dataset\_no = dataset\_no.drop\_duplicates()

dataset\_no = dataset\_no.dropna()

**Non-Parametric Tests:**

**Mann Whitney Test**

**Attrition vs DistanceFromHome**

from scipy.stats import mannwhitneyu

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

print(stats,p)

1295261.0 0.488538986087403

H0: There is no significant differences in the Distance From Home between attrition\_yes and attirition\_no

Ha: There is a significant difference in the Distance From Home between attrition\_yes and attirition\_no

Since p value > 0.05, H0 is accepted.

**Attrition vs MonthlyIncome**

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

print(stats,p)

1249573.5 0.06508807631576838

H0: There is no significant differences in the MonthlyIncome between attrition\_yes and attirition\_no

Ha: There is a significant difference in the MonthlyIncome between attrition\_yes and attirition\_no

Since p value > 0.05, H0 is accepted.

**Attrition vs** **NumCompaniesWorked**

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

print(stats,p)

1238814.5 0.02793197853866981

H0: There is no significant differences in the NumCompaniesWorked between attrition\_yes and attirition\_no

Ha: There is a significant difference in the NumCompaniesWorked between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs** **PercentSalaryHike**

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

print(stats,p)

1231873.5 0.017810794960084964

H0: There is no significant differences in the PercentSalaryHike between attrition\_yes and attirition\_no

Ha: There is a significant difference in the PercentSalaryHike between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs TotalWorkingYears**

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

print(stats,p)

895173.5 2.741211827689903e-39

H0: There is no significant differences in the TotalWorkingYears between attrition\_yes and attirition\_no

Ha: There is a significant difference in the TotalWorkingYears between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs YearsAtCompany**

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

print(stats,p)

912579.0 3.3433144809752036e-36

H0: There is no significant differences in the YearsAtCompany between attrition\_yes and attirition\_no

Ha: There is a significant difference in the YearsAtCompany between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs YearsWithCurrManager**

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

print(stats,p)

945958.5 5.420302388722274e-31

H0: There is no significant differences in the YearsWithCurrManager between attrition\_yes and attirition\_no

Ha: There is a significant difference in the YearsWithCurrManager between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs YearsSinceLastPromotion**

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

print(stats,p)

1196606.0 0.00037904698157957496

H0: There is no significant differences in the YearsSinceLastPromotion between attrition\_yes and attirition\_no

Ha: There is a significant difference in the YearsSinceLastPromotion between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Chi -Square Test**

**Attrition vs BusinessTravel**

from scipy.stats import chi2\_contingency

chitable = pd.crosstab(dataset1.Attrition,dataset1.BusinessTravel)

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

print(stats,p)

70.07594084831366 6.07019702736392e-16

chitable

Out[36]:

BusinessTravel Non-Travel Travel\_Frequently Travel\_Rarely

Attrition

No 412 621 2644

Yes 36 204 465

H0 - There is no dependency between Attrition and BusinessTravel

Ha - There is dependency between Attrition and BusinessTravel

Since p value is < 0.05, Ha is accepted.

**Attrition vs Gender**

chitable = pd.crosstab(dataset1.Attrition,dataset1.Gender)

chitable

Out[38]:

Gender Female Male

Attrition

No 1488 2189

Yes 268 437

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

print(stats,p)

1.3825823839528295 0.23966176275638887

H0 - There is no dependency between Attrition and Gender

Ha - There is dependency between Attrition and Gender

Since p value is > 0.05, H0 is accepted.

**Attrition vs JobRole**

chitable = pd.crosstab(dataset1.Attrition,dataset1.JobRole)

chitable

Out[42]:

JobRole Healthcare Representative ... Sales Representative

Attrition ...

No 333 ... 210

Yes 56 ... 36

[2 rows x 9 columns]

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

print(stats,p)

21.893724958847 0.005116592717526599

H0 - There is no dependency between Attrition and JobRole

Ha - There is dependency between Attrition and JobRole

Since p value is < 0.05, Ha is accepted.

**Attrition vs MaritalStatus**

chitable = pd.crosstab(dataset1.Attrition,dataset1.MaritalStatus)

chitable

Out[46]:

MaritalStatus Divorced Married Single

Attrition

No 872 1756 1049

Yes 98 251 356

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

print(stats,p)

133.85785802925156 8.573051828219379e-30

H0 - There is no dependency between Attrition and MaritalStatus

Ha - There is dependency between Attrition and MaritalStatus

Since p value is < 0.05, Ha is accepted.

**Two Sample Separate t -Test**

**Attrition vs DistanceFromHome**

from scipy.stats import ttest\_ind

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

print(stats,p)

-0.6253536318706914 0.5317715668047676

H0: There is no significant differences in the Distance From Home between attrition\_yes and attirition\_no

Ha: There is a significant difference in the Distance From Home between attrition\_yes and attirition\_no

Since p value > 0.05, H0 is accepted.

**Attrition vs MonthlyIncome**

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

print(stats,p)

-1.9969640177214658 0.045890862744972095

H0: There is no significant differences in the MonthlyIncome between attrition\_yes and attirition\_no

Ha: There is a significant difference in the MonthlyIncome between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.

**Attrition vs YearsAtCompany**

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

print(stats,p)

-8.881225486705604 9.476118084889976e-19

H0: There is no significant differences in the YearsAtCompany between attrition\_yes and attirition\_no

Ha: There is a significant difference in the YearsAtCompany between attrition\_yes and attirition\_no

Since p value < 0.05, Ha is accepted.