Assignment Day 10

In [1]: import pandas as pd

In [2]: dataset=pd.read\_excel("Attrition\_Rate\_analysis.xls",sheet\_name=1)

In [3]: dataset.head()

Out[3]:

DistanceFromHome\_YES ... YearsWithCurrManager\_NO

0 10.0 ... 0

1 11.0 ... 3

2 1.0 ... 5

3 1.0 ... 4

4 4.0 ... 7

[5 rows x 14 columns]

In [4]: dataset.columns

Out[4]:

Index(['DistanceFromHome\_YES', 'DistanceFromHome\_NO', 'MonthlyIncome\_YES',

'MonthlyIncome\_NO', 'NumCompaniesWorked\_YES', 'NumCompaniesWorked\_NO',

'PercentSalaryHike\_YES', 'PercentSalaryHike\_NO',

'TotalWorkingYears\_YES', 'TotalWorkingYears\_NO', 'YearsAtCompany\_YES',

'YearsAtCompany\_NO', 'YearsWithCurrManager\_YES',

'YearsWithCurrManager\_NO'],

dtype='object')

In [6]: from scipy.stats import mannwhitneyu

...: a1=dataset.DistanceFromHome\_YES

...: a2=dataset.DistanceFromHome\_NO

...: stat, p=mannwhitneyu(a1,a2)

...: print(stat, p)

1317879.0 0.0

In [7]: #As the P value of 0.0 is < 0.05, the H0 is rejected and Ha is accepted

...: Ha: There is significant differences in the Distance From Home between attrition (Y) and attirition(N)

In [8]: #Attrition vs income

In [9]: a1=dataset.MonthlyIncome\_YES

...: a2=dataset.MonthlyIncome\_NO

...: stat, p=mannwhitneyu(a1,a2)

...: print(stat, p)

1365088.5 0.0

In [10]: #As P value of 0.0 is < 0.05 ,th H0 is rejected and Ha is accepted

...: # Ha: There is significant difference in the income between attrition(Y) and attrition(N)

In [11]: a1=dataset.TotalWorkingYears\_YES

...: a2=dataset.TotalWorkingYears\_NO

...: stat, p=mannwhitneyu(a1,a2)

...: print(stat, p)

1726497.0 0.0

In [12]: #As P value of 0.0 is < 0.05 ,th H0 is rejected and Ha is accepted

...: # Ha: There is significant difference in the Total Working hours between attrition(Y) and attrition(N)/

In [13]: ### Attrition vs YearsAtCompany

In [14]: a1=dataset.YearsAtCompany\_YES

...: a2=dataset.YearsAtCompany\_NO

...: stat, p=mannwhitneyu(a1,a2)

...: print(stat, p)

1706751.0 0.0

In [15]: # As P value of 0.0 is < 0.05 ,th H0 is rejected and Ha is accepted

...: # Ha: There is significant difference in the YearsAtCompany between attrition(Y) and attrition(N)

In [16]: ## Attrition vs YearswithCurrentManager

In [17]: a1=dataset.YearsWithCurrManager\_YES

...: a2=dataset.YearsWithCurrManager\_NO

...: stat, p=mannwhitneyu(a1,a2)

...: print(stat, p)

1672735.5 0.0

In [18]: # As P value of 0.0 is < 0.05 ,th H0 is rejected and Ha is accepted

...: # Ha: There is significant difference in the YearsWithCurrManager between attrition(Y) and attrition(N)