Statistical Tests (Mann-Whitney)

```
import pandas as pd

dataset=pd.read_excel('general_data_Correlation.xlsx', sheet_name=1)

dataset=pd.read_excel('general_data_Correlation.xlsx', sheet_name=0)

dataset.head()
Out[4]:
   Age  Attrition  ... YearsSinceLastPromotion YearsWithCurrManager
0   51          0  ...                      0                    0
1   31          1  ...                      1                    4
2   32          0  ...                      0                    3
3   38          0  ...                      7                    5
4   32          0  ...                      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')

from scipy.stats import mannwhitneyu

no_data=pd.read_excel('general_data_Correlation.xlsx', sheet_name=1)

yes_data=pd.read_excel('general_data_Correlation.xlsx', sheet_name=2)
```

Attrition Vs Distance from Home

```
stat, p=mannwhitneyu(no_data.DistanceFromHome,yes_data.DistanceFromHome)

print(stat,p)
1312110.0 0.4629185205822659
```

As the P value of 0.4629185205822659 is < 0.05, the H0 is rejected and Ha is accepted.
H0: There is no significant difference in the Distance from Home between attrition (Y) and attrition (N)
Ha: There is a significant difference in the Distance from Home between attrition (Y) and attrition (N)

Attrition Vs Income

```
stat, p=mannwhitneyu(no_data.MonthlyIncome,yes_data.MonthlyIncome)

print(stat,p)
1264900.5 0.053577283839938566
```

As the P value is 0.053577283839938566, which is > than 0.05, the H0 is accepted and ha is rejected.
H0: There is no significant difference in the income between attrition (Y) and attrition (N)
Ha: There is a significant difference in the income between attrition (Y) and attrition (N)

Statistical Tests (CHI-SQUARE TEST)

```
from scipy.stats import chi2_contingency

chi_square = pd.crosstab(dataset.Gender,dataset.BusinessTravel)

chi_square
Out[19]:
BusinessTravel  Non-Travel  Travel_Frequently  Travel_Rarely
Gender
Female             153            330         1281
Male               297            501         1848

stats,p,dof,expected = chi2_contingency(chi_square)

print(stats,p)
7.929887577835395 0.01896910285626416
```

P value is less than 0.05, hence Alternate hypothesis is accepted.
H0 - There is no dependency between gender and businessTravel
H1 - There is dependency between gender and businessTravel

```
chi_square2 = pd.crosstab(dataset.Gender,dataset.Age)

chi_square2
Out[21]:
Age    18  19  20  21  22  23  24  25  26  ...  52  53  54  55  56  57  58  59  60
```

Gender ...
Female 0 3 15 15 18 18 33 30 42 ... 18 12 30 36 15 0 18 9 9
Male 24 24 18 24 30 24 45 48 75 ... 36 45 24 30 27 12 24 21 6

[2 rows x 43 columns]

stats,p,dof,expected = chi2_contingency(chi_square2)

print(stats,p)
144.8889096499983 3.0836271884017946e-13

P value is less than 0.05, hence Alternate hypothesis is accepted.
H0 - There is no dependency between gender and Age
H1 - There is dependency between gender and Age

Statistical Tests (Separate T Test)

from scipy.stats import ttest_ind

Attrition Vs Distance from Home

stat, p=ttest_ind(no_data.DistanceFromHome,yes_data.DistanceFromHome)

print(stats,p)
144.8889096499983 0.518286042805572

As the P value is again 0.518286042805572, which is > than 0.05, the H0 is accepted and ha is rejected.
H0: There is no significant difference in the Distance from Home between attrition (Y) and attrition (N)
Ha: There is a significant difference in the Distance from Home between attrition (Y) and attrition (N)

Attrition Vs MonthlyIncome

stat, p=ttest_ind(no_data.MonthlyIncome,yes_data.MonthlyIncome)

print(stats,p)
144.8889096499983 0.03842748490605113

As the P value is again 0.03842748490605113, which is < than 0.05, the H0 is rejected and ha is accepted.
H0: There is no significant difference in the Monthly Income between attrition (Y) and attrition (N)
Ha: There is a significant difference in the Monthly Income between attrition (Y) and attrition (N)