**TESTS ON DATASET- DAY 12 STAT TEST**

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

d1 = pd.read\_csv("general\_data.csv")

d1=d1.dropna()

print(d1.columns)

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')

print(d1[['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())

Age -0.409517

DistanceFromHome -0.230691

Education -0.565008

EmployeeCount 0.000000

EmployeeID -1.198607

JobLevel 0.388189

MonthlyIncome 0.990836

NumCompaniesWorked 0.014307

PercentSalaryHike -0.306951

StandardHours 0.000000

StockOptionLevel 0.356755

TotalWorkingYears 0.909316

TrainingTimesLastYear 0.494215

YearsAtCompany 3.930726

YearsSinceLastPromotion 3.592162

YearsWithCurrManager 0.170703

dtype: float64

**CHI SQUARE TEST between Gender and Attrition**

H0 >> Attrition behaviour is not same for both male and female

Ha >> Attrition behaviour is same for both male and female

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

c=pd.crosstab(d1.Attrition,d1.Gender)

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

print(stats,p)

1.3825823839528295 0.23966176275638887

**Conclusion:**

As p-value > 0.05, Hence H0 will be accepted

**CHI SQUARE TEST between MaritalStatus and Attrition**

H0 >> The attrition rate is not the same among employees of different marital status.

Ha >> The attrition rate is not the same among employees of different marital status.

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

c=pd.crosstab(d1.Attrition,d1.MaritalStatus)

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

print(stats,p)

133.85785802925156 8.573051828219379e-30

**Conclusion:**

As we can see that, p < 0.05 so Ha will be accepted and H0 will be rejected

**MANN-WHITNEY TEST between Attrition and MonthlyIncome**

H0 >> Monthly income has no significant effect on attrition

Ha >> Monthly income has significant effect on attrition

d1['Attrition'] = pd.factorize(d1.Attrition)[0]

**from** **scipy.stats** **import** mannwhitneyu

stats,p = mannwhitneyu(d1.Attrition,d1.MonthlyIncome)

print(stats,p)

0.0 0.0

**Conclusion:**

Here p value is 0.0, so Ha will be accepted

**MANN-WHITNEY TEST between Attrition and PercentSalaryHike**

H0 >> Salary Hike has no impact on Attrition

Ha >> Salary Hike leads to less Attrition

**from** **scipy.stats** **import** mannwhitneyu

stats,p = mannwhitneyu(d1.Attrition,d1.PercentSalaryHike)

print(stats,p)

0.0 0.0

**Conclusion:**

p-value < 0.05, therefore Ha accepted