**CORRELATION PRACTICE**

import pandas as p

from scipy.stats import pearsonr

dataset=p.read\_csv('general\_data.csv')

a=dataset.head()

print(a)

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0}) // Changing attrition values to 0 & 1

**Output:**



stats, s=pearsonr(dataset.Attrition,dataset.Education) // Attrition and education

print('P value for Attrition and education',s)

print('Covariance value',stats)

Output:

P value for Attrition and education : 0.3157293177118575

Covariance value: -0.01511116771096876

**Analysis :** R value is negative so negative correlated. P value is >0.05 , so null hypothesis is accepted.

***Checking Attrition & Age co related or not***

stats, t=pearsonr(dataset.Attrition,dataset.Age) // Age

print('P value for Attrition and education',t)

print('Covariance value',stats)

**Output:**

P value for Attrition and education 1.996801615887666e-26

Covariance value -0.15920500686577943

**Analysis**: Variables are negative co-related

***Checking Attrition & Distance From home co-related or not***

stats, t=pearsonr(dataset.Attrition,dataset.DistanceFromHome)

print('P value for Attrition and education',t)

print('Covariance value',stats)

**Output :**

P value for Attrition and education 0.5182860428050771

Covariance value -0.00973014101017969

**Analysis**: Attrition is not happening due to DFH and Variables are negative co-related

***Checking Attrition & Job level co-related or not***

stats,J=pearsonr(dataset.Attrition, dataset.JobLevel)

print('P value for Attrition and education',J)

print('covariance', stats)

***Output:***

P value for Attrition and education 0.4945171727183978

covariance -0.010289713287494997

**Analysis**: Attrition is happening not due to Job level.Variables are negative co-related

***Checking Attrition & Monthlyincome co-related or not***

stats,M=pearsonr(dataset.Attrition, dataset.MonthlyIncome)

print('P value for Attrition and education',M)

print('covariance', stats)

**Output:**

P value for Attrition and education 0.03842748490600132

covariance -0.031176281698115017

**Analysis:**

Due to monthly income attrition is happening and variables are negative co-related.

***Checkinh Attrition & PercentSalaryHike***

stats,K=pearsonr(dataset.Attrition, dataset.PercentSalaryHike)

print('P value for Attrition and education',K)

print('covariance', stats)

**Output:**

P value for Attrition and education 0.030743386433355353

covariance 0.032532594891053514

**Analysis**:

Due to Percantage of Hike Attrition is happening and varaibles are positive co- related

***Checking Attrition & EmployeeCount***

stats,H=pearsonr(dataset.Attrition, dataset.EmployeeCount)

print('P value for Attrition',H)

print('covariance', stats)

**Output:**

covariance nan

C:\ProgramData\Anaconda3\lib\site-packages\scipy\stats\stats.py:3845: PearsonRConstantInputWarning: An input array is constant; the correlation coefficent is not defined.

warnings.warn(PearsonRConstantInputWarning())

***Checking Attrition & TrianingTimesLastYear***

stats,H=pearsonr(dataset.Attrition, dataset.TrainingTimesLastYear)

print('P value for Attrition',H)

print('covariance', stats)

**Output:**

P value for Attrition 0.0010247061915365072

covariance -0.04943057624425494

**Analysis:**

Attrition is not happening due to Training Times and varaibles are negatively co-related.