**DAY – 11 Assignments**

**Unsupervised Learning – correlation analysis**

dataset1=pd.read\_excel("general\_data.xls",sheet\_name=0)

dataset1.head()

Out[27]:

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]

stats,p=pearsonr(dataset1.Attrition,dataset1.DistanceFromHome)

print(stats,p)

-0.00973014101017966 0.5182860428050771

stats,p=pearsonr(dataset1.Attrition,dataset1.MonthlyIncome)

print(stats,p)

-0.031176281698115007 0.03842748490600132

stats,p=pearsonr(dataset1.Attrition,dataset1.PercentSalaryHike)

print(stats,p)

0.03253259489105351 0.030743386433355353

stats,p=pearsonr(dataset1.Attrition,dataset1.YearsAtCompany)

print(stats,p)

-0.1343922139899772 3.1638831224877484e-19

stats,p=pearsonr(dataset1.Attrition,dataset1.YearsWithCurrManager)

print(stats,p)

-0.15619931590162842 1.7339322652900218e-25

The inference from above analysis are as follows :

**Attrition vs DistanceFromHome**

r = -0.009 it measns there is low negative correlation between attrition and DistanceFromHome

Also value of p=0.518 is > 0.05 ,so Null Hypothesis is accepted and alternate hypothesis is rejected hence there is no significant difference between Attrition and DistanceFromHome.

**Attrition vs MonthlyIncome**

r = -0.031it measns there is low negative correlation between attrition and MonthlyIncome

Also value of p = 0.038 < 0.05 ,so null hypothesis is rejected and Alternate hypothesis is accepted and there is significant difference between Attrition and MonthlyIncome.

**Attrition vs PercentSalaryHike**

r = 0.032 it measns there is positive correlation between Attrition and PercentSalaryHike

Also value of p = 0.030 < 0.05, so Null Hypothesis is rejected and alternate hypothesis is accepted hence there is significant difference between Attrition and PercentSalaryHike

**Attrition vs YearsAtCompany**

r = -0.134 it measns there is low negative correlation between attrition and DistanceFromHome

Also value of p < 0.05, so null hypothesis is rejected and Alternate hypothesis is accepted and there is significant difference between Attrition and YearsAtCompany.

**Attrition vs YearsWithCurrManager**

r = -0.156 it measns there is low negative correlation between attrition and DistanceFromHome.

Also value of p < 0.05, so null hypothesis is rejected and Alternate hypothesis is accepted and there is significant difference between Attrition and YearsWithCurrManager.