Assignment No 11

#correlation Analysis

In [2]: import pandas as pd

...: dataset=pd.read\_excel("Attrition\_Rate\_analysis.xls",sheet\_name=0)

In [3]: dataset.head()

Out[3]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 51 No ... 0 0

1 31 Yes ... 1 4

2 32 No ... 0 3

3 38 No ... 7 5

4 32 No ... 0 4

[5 rows x 24 columns]

In [4]: from scipy.stats import pearsonr

...: dataset['TotalWorkingYears']=dataset['TotalWorkingYears'].fillna(11.28)

...: dataset.columns

Out[4]:

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')

In [11]: from sklearn import preprocessing

In [12]: Label\_encoder=preprocessing.LabelBinarizer()

In [13]: dataset["Attrition"]=Label\_encoder.fit\_transform(dataset["Attrition"])

In [14]: stats,p=pearsonr(dataset.Attrition,dataset.DistanceFromHome)

...: print(stats, p)

-0.009730141010179674 0.5182860428050771

In [15]: stats, p=pearsonr(dataset.Attrition, dataset.MonthlyIncome)

...: print(stats, p)

-0.031176281698115007 0.03842748490600132

In [16]: stats, p=pearsonr(dataset.Attrition, dataset.TotalWorkingYears)

...: print(stats, p)

-0.17011136355965176 5.473159751791005e-30

In [17]: stats, p=pearsonr(dataset.Attrition, dataset.YearsAtCompany)

...: print(stats, p)

-0.1343922139899772 3.1638831224877484e-19

In [18]: stats, p=pearsonr(dataset.Attrition, dataset.YearsWithCurrManager)

...: print(stats, p)

-0.15619931590162847 1.7339322652896276e-25

#####################################################################

Attrition & DistanceFromHome:

As r = -0.009, there’s low negative correlation between Attrition and DistanceFromHome

As the P value of 0.518 is > 0.05, we are accepting H0 and hence there’s no significant correlation

between Attrition & DistanceFromHome

Attrition & MonthlyIncome:

As r = -0.031, there’s low negative correlation between Attrition and MonthlyIncome

As the P value of 0.038 is < 0.05, we are accepting Ha and hence there’s significant correlation

between Attrition & MonthlyIncome

Attrition & TotalWorkingYears:

As r = -0.17, there’s low negative correlation between Attrition and TotalWorkingYears

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between

Attrition & TotalWorkingYears

Attrition & YearsAtCompany:

As r = -0.1343, there’s low negative correlation between Attrition and YearsAtCompany

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between

Attrition & YearsAtCompany

Attrition & YearsWithCurrManager:

As r = -0.1561, there’s low negative correlation between Attrition and YearsWithCurrManager

As the P value is < 0.05, we are accepting Ha and hence there’s significant correlation between

Attrition & YearsWithCurrManager