**ATTRION PROJECT – CORRELATION**

**Find correlation between Attrition with other variables (10 different combinations)**

import pandas as pd

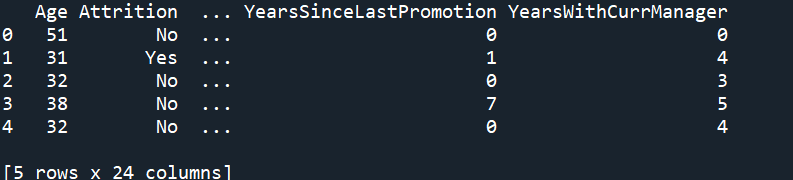
import numpy as np

import matplotlib.pyplot as plt

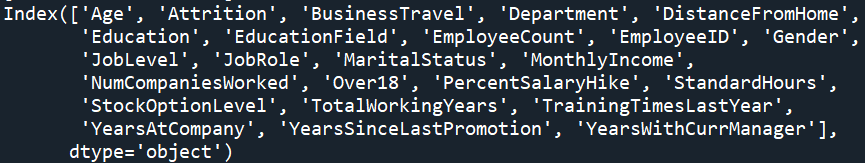
from scipy.stats import pearsonr

dataset1=pd.read\_excel("my\_data.xlsx",sheet\_name=0)

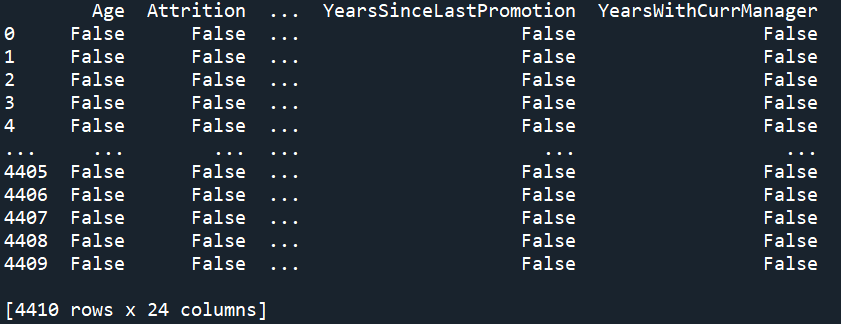
print(dataset1.head())



print(dataset1.columns)



print(dataset1.isnull())



print(dataset1.duplicated())

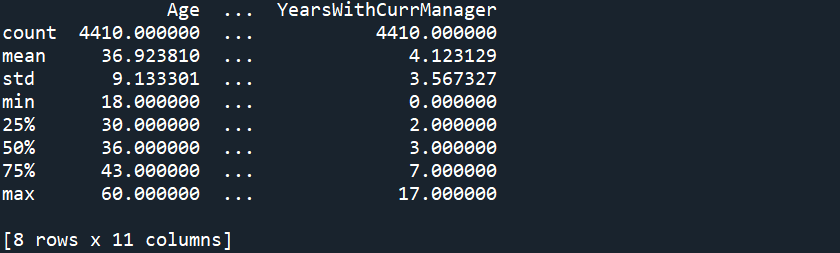


dataset2=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()

print(dataset2)



1. Correlation between Attrition and Age:

H0: There is no significance Correlation between Attrition and Age

H1: There is significance Correlation between Attrition and Age

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

print(stats,p)



**P > 0.05 => Accept H0 & Reject H1**

1. Correlation between Attrition and DistanceFromHome:

H0: There is no significance Correlation between Attrition and DistanceFromHome

H1: There is significance Correlation between Attrition and DistanceFromHome

stats,p1=pearsonr(dataset1.Attrition,dataset1.DistanceFromHome)

print(stats,p1)



**P < 0.05 => Reject H0 & Accept H1**

1. Correlation between Attrition and Education:

H0: There is no significance Correlation between Attrition and Education

H1: There is significance Correlation between Attrition and Education

stats,p2=pearsonr(dataset1.Attrition,dataset1.Education)

print(stats,p2)



**P > 0.05 => Accept H0 & Reject H1**

1. Correlation between Attrition and JobLevel:

H0: There is no significance Correlation between Attrition and JobLevel

H1: There is significance Correlation between Attrition and JobLevel

stats,p3=pearsonr(dataset1.Attrition,dataset1.JobLevel)

print(stats,p3)



**P > 0.05 => Accept H0 & Reject H1**

1. Correlation between Attrition and MonthlyIncome:

H0: There is no significance Correlation between Attrition and MonthlyIncome

H1: There is significance Correlation between Attrition and MonthlyIncome

stats,p4=pearsonr(dataset1.Attrition,dataset1.MonthlyIncome)

print(stats,p4)



**P < 0.05 => Reject H0 & Accept H1**

1. Correlation between Attrition and StockOptionLevel:

H0: There is no significance Correlation between Attrition and StockOptionLevel

H1: There is significance Correlation between Attrition and StockOptionLevel

stats,p5=pearsonr(dataset1.Attrition,dataset1.StockOptionLevel)

print(stats,p5)



**P > 0.05 => Accept H0 & Reject H1**

1. Correlation between Attrition and NumCompaniesWorked:

H0: There is no significance Correlation between Attrition and NumCompaniesWorked

H1: There is significance Correlation between Attrition and NumCompaniesWorked

stats,p9=pearsonr(dataset1.Attrition,dataset1.NumCompaniesWorked)

print(stats,p9)



**P < 0.05 => Reject H0 & Accept H1**

1. Correlation between Attrition and YearsWithCurrManager:

H0: There is no significance Correlation between Attrition and YearsWithCurrManager

H1: There is significance Correlation between Attrition and YearsWithCurrManager

stats,p5=pearsonr(dataset1.Attrition,dataset1.YearsWithCurrManager)

print(stats,p5)



**P < 0.05 => Reject H0 & Accept H1**

1. Correlation between Attrition and YearsAtCompany:

H0: There is no significance Correlation between Attrition and YearsAtCompany

H1: There is significance Correlation between Attrition and YearsAtCompany

stats,p7=pearsonr(dataset1.Attrition,dataset1.YearsAtCompany)

print(stats,p7)



**P < 0.05 => Reject H0 & Accept H1**

1. Correlation between Attrition and PercentSalaryHike:

H0: There is no significance Correlation between Attrition and PercentSalaryHike

H1: There is significance Correlation between Attrition and PercentSalaryHike

stats,p8=pearsonr(dataset1.Attrition,dataset1.PercentSalaryHike)

print(stats,p8)



**P < 0.05 => Reject H0 & Accept H1**