CORRELATION ANALYSIS

1. Attrition vs. Years at company

import numpy as n

import pandas as pd

import matplotlib.pyplot as m

from scipy.stats import pearsonr as pear, mannwhitneyu as man,ttest\_ind as ttest

from sklearn.preprocessing import LabelEncoder

count=0

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

ds.dropna()

ds.drop\_duplicates()

le=LabelEncoder()

ds["Attrition"]=le.fit\_transform(ds["Attrition"])

stats,p=pear(ds.Attrition,ds.YearsAtCompany)

print("the value pf p is: ",p)

**the value pf p is: 3.1638831224877484e-19**

since the value of p is <0.05 Ha is true and there is significant relation between attrition and years at company

1. Attrition vs. Monthly income

stats,p=pear(ds.Attrition,ds.MonthlyIncome)

print("the value pf p is: ",p)

**the value pf p is: 0.03842748490600132**

since the value of p is <0.05 Ha is true and there is significant relation between attrition and Monthly income

1. Attrition vs. Training Times Last Year

stats,p=pear(ds.Attrition,ds.TrainingTimesLastYear)

print("the value pf p is: ",p)

**the value pf p is: 0.0010247061915362814**

since the value of p is <0.05 Ha is true and there is significant relation between attrition and Monthly income

1. Attrition vs years with curret manager

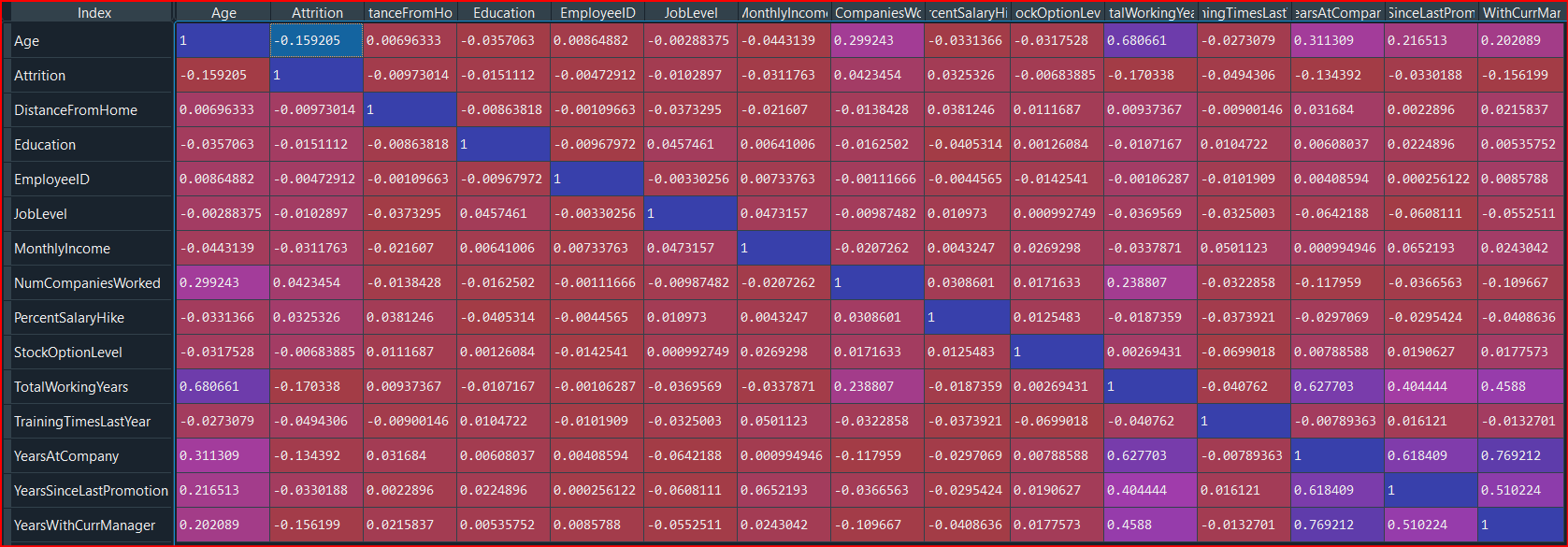
stats,p=pear(ds.Attrition,ds.YearsWithCurrManager)

print("the value pf p is: ",p)

**the value pf p is: 1.7339322652900218e-25**

since the value of p is <0.05 Ha is true and there is significant relation between attrition and years with current manager

Correlation matrix



**INFERENCE:**

**1.theres low and negative correlation between attrition and age,total working years ,with current manager,distance from home,job level,monthly income ,education,stock option level,**

**2. theres low and positive correlation between attrition and number of companies worked,%salary hike**