import pandas as pd

dataset=pd.read\_excel("general\_data.xlsx",sheet\_name=0)

dataset

Out[1]:

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

... ... ... ... ...

4405 42 0 ... 0 2

4406 29 0 ... 0 2

4407 25 0 ... 1 2

4408 42 0 ... 7 8

4409 40 0 ... 3 9

[4410 rows x 24 columns]

from scipy.stats import pearsonr

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

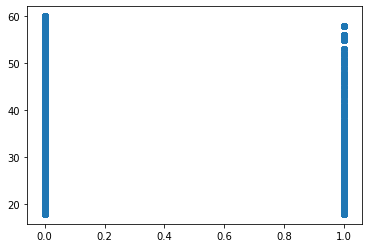
print(stats,p)

-0.15920500686577965 1.996801615886744e-26

import matplotlib.pyplot as plt

plt.scatter(dataset.Attrition,dataset.Age)

Out[3]: <matplotlib.collections.PathCollection at 0x2a9f2decb08>



dataset.corr()

Out[4]:

Age ... YearsWithCurrManager

Age 1.000000 ... 0.202089

Attrition -0.159205 ... -0.156199

DistanceFromHome 0.006963 ... 0.021584

Education -0.035706 ... 0.005358

EmployeeCount NaN ... NaN

EmployeeID 0.008649 ... 0.008579

JobLevel -0.002884 ... -0.055251

MonthlyIncome -0.044314 ... 0.024304

NumCompaniesWorked 0.299243 ... -0.109667

PercentSalaryHike -0.033137 ... -0.040864

StandardHours NaN ... NaN

StockOptionLevel -0.031753 ... 0.017757

TotalWorkingYears 0.680661 ... 0.458800

TrainingTimesLastYear -0.027308 ... -0.013270

YearsAtCompany 0.311309 ... 0.769212

YearsSinceLastPromotion 0.216513 ... 0.510224

YearsWithCurrManager 0.202089 ... 1.000000

[17 rows x 17 columns]

import pandas as pd

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

dataset1.drop\_duplicates()

Out[1]:

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

... ... ... ... ...

4405 42 0 ... 0 2

4406 29 0 ... 0 2

4407 25 0 ... 1 2

4408 42 0 ... 7 8

4409 40 0 ... 3 9

[4410 rows x 24 columns]

from scipy.stats import pearsonr

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

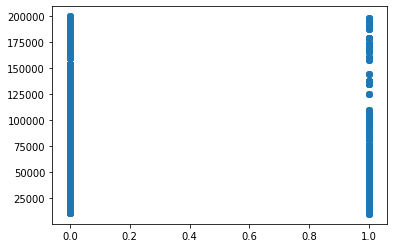
print(stats,p)

-0.031176281698115007 0.03842748490600132

import matplotlib.pyplot as plt

plt.scatter(dataset1.Attrition,dataset1.MonthlyIncome)

Out[5]: <matplotlib.collections.PathCollection at 0x23f7edb9748>



dataset1.corr()

Out[6]:

Age ... YearsWithCurrManager

Age 1.000000 ... 0.202089

Attrition -0.159205 ... -0.156199

DistanceFromHome 0.006963 ... 0.021584

Education -0.035706 ... 0.005358

EmployeeCount NaN ... NaN

EmployeeID 0.008649 ... 0.008579

JobLevel -0.002884 ... -0.055251

MonthlyIncome -0.044314 ... 0.024304

NumCompaniesWorked 0.299243 ... -0.109667

PercentSalaryHike -0.033137 ... -0.040864

StandardHours NaN ... NaN

StockOptionLevel -0.031753 ... 0.017757

TotalWorkingYears 0.680661 ... 0.458800

TrainingTimesLastYear -0.027308 ... -0.013270

YearsAtCompany 0.311309 ... 0.769212

YearsSinceLastPromotion 0.216513 ... 0.510224

YearsWithCurrManager 0.202089 ... 1.000000

[17 rows x 17 columns]