**Correlation Assignment**

**Step 1: Loading Data:**

import pandas as pd

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

dataset.head()

Out[35]:

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]

dataset.columns

Out[36]:

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')

**Step 2 – Data Treatment:**

dataset.isnull()

Out[37]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 False False ... False False

1 False False ... False False

2 False False ... False False

3 False False ... False False

4 False False ... False False

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

4405 False False ... False False

4406 False False ... False False

4407 False False ... False False

4408 False False ... False False

4409 False False ... False False

[4410 rows x 24 columns]

dataset.duplicated()

Out[39]:

0 False

1 False

2 False

3 False

4 False

4405 False

4406 False

4407 False

4408 False

4409 False

Length: 4410, dtype: bool

dataset1 = dataset.drop\_duplicates()

dataset2 = dataset1.dropna()

**Step 3 -Correlation Analysis:**

1. **Correlation between Attrition and Age:**

from scipy.stats import pearsonr

stats, p =pearsonr(dataset2.Attrition, dataset2.Age)

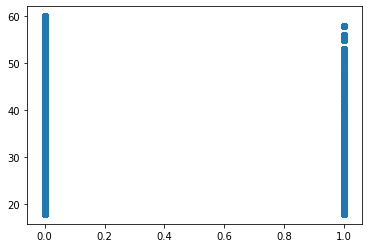
print(stats,p)

-0.15839867954096706 5.1265982193780794e-26

import matplotlib.pyplot as plt

plt.scatter(dataset2.Attrition,dataset2.Age)

Out[46]: <matplotlib.collections.PathCollection at 0x1bb686e59c8>



H0 -> There is no significant correlation between Attrition and Age.

Ha -> There is significant correlation between Attrition and Age.

**Inference:**

The p value is less than 0.05, hence there is significant correlation between Attrition and Age.

The negative value of r represents that the Attrition and Age are negatively correlated and are inversely proportional.

1. **Correlation between Attrition and MonthlyIncome:**

dataset2.corr()

Out[47]:

Age ... YearsWithCurrManager

Age 1.000000 ... 0.201580

Attrition -0.158399 ... -0.154692

DistanceFromHome 0.007376 ... 0.021773

Education -0.033900 ... 0.005645

EmployeeCount NaN ... NaN

EmployeeID 0.008105 ... 0.009079

JobLevel -0.001137 ... -0.053898

MonthlyIncome -0.045163 ... 0.023095

NumCompaniesWorked 0.299527 ... -0.109372

PercentSalaryHike -0.032561 ... -0.039687

StandardHours NaN ... NaN

StockOptionLevel -0.031504 ... 0.019398

TotalWorkingYears 0.680037 ... 0.458640

TrainingTimesLastYear -0.028962 ... -0.015792

YearsAtCompany 0.311281 ... 0.769161

YearsSinceLastPromotion 0.215650 ... 0.510341

YearsWithCurrManager 0.201580 ... 1.000000

[17 rows x 17 columns]

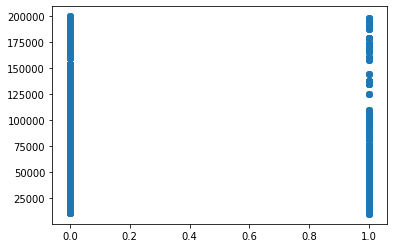
stats, p =pearsonr(dataset2.Attrition, dataset2.MonthlyIncome)

print(stats,p)

-0.030160293808460668 0.045890862744719166

plt.scatter(dataset2.Attrition,dataset2.MonthlyIncome)

Out[50]: <matplotlib.collections.PathCollection at 0x1bb67623c48>



H0 -> There is no significant correlation between Attrition and MonthlyIncome.

Ha -> There is significant correlation between Attrition and MonthlyIncome.

**Inference:**

The p value is less than 0.05, hence there is significant correlation between Attrition and MonthlyIncome.

The negative value of r represents that the Attrition and MonthlyIncome are negatively correlated and are inversely proportional.

1. **Correlation between Attrition and PercentSalaryHike:**

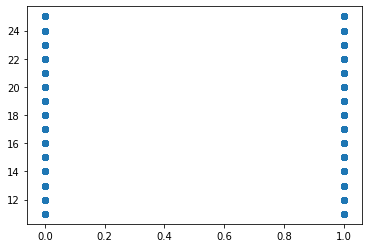
stats, p =pearsonr(dataset2.Attrition, dataset2.PercentSalaryHike)

print(stats,p)

0.03315303713546665 0.028192446935106235

plt.scatter(dataset2.Attrition, dataset2.PercentSalaryHike)

Out[56]: <matplotlib.collections.PathCollection at 0x1bb678fad48>



**Inference:**

There is a positive correlation between Attrition and PercentSalaryHike