**Correlation:**

**Step1: Launching**

import pandas as pd

from scipy.stats import pearsonr

import matplotlib.pyplot as plt

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

**dataset1.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

**dataset1.columns**

Out[5]:

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[3]:

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.dropna(inplace=True)

Out[4]:

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

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

4404 29 No ... 1 5

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

[4382 rows x 24 columns]

dataset.duplicated()

Out[5]:

0 False

1 False

2 False

3 False

4 False

4405 False

4406 False

4407 False

4408 False

4409 False

Length: 4410, dtype: bool

dataset.drop\_duplicates

Out[6]:

<bound method DataFrame.drop\_duplicates of 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

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

4405 42 No ... 0 2

4406 29 No ... 0 2

4407 25 No ... 1 2

4408 42 No ... 7 8

4409 40 No ... 3 9

[4410 rows x 24 columns]>

dataset["Attrition"]=dataset["Attrition"].map({"Yes":1,"No":0})

dataset["Gender"]=dataset["Gender"].map({"Male":1,"Female":0})

dataset.head(5)

Out[10]:

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]

**Correlation between Attrition and Age:**

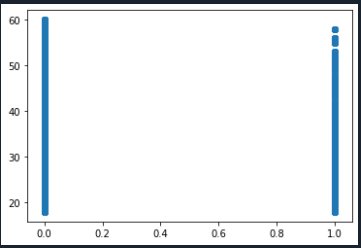
**stats,p=pearsonr(dataset.Attrition,dataset.Age)**

print(stats,p)

-0.159205006865775 1.996801615893625e-26

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

Out[16]: <matplotlib.collections.PathCollection at 0x2368c9fb808>



H0- Age doesn’t have significant impact on Attrition.

H1- Age have significant impact on Attrition.

**From r values we can conclude the age is negatively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and Gender:**

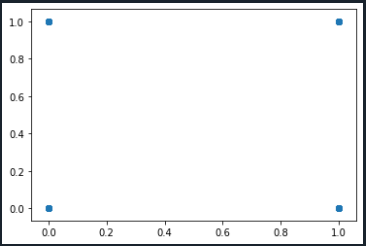
**stats,p=pearsonr(dataset.Attrition,dataset.Gender)**

print(stats,p)

0.018125078877010366 0.22881970951790567

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

Out[19]: <matplotlib.collections.PathCollection at 0x2368caacdc8>



H0- Gender doesn’t have significant impact on Attrition.

H1- Gender have significant impact on Attrition.

**From r values we can conclude the gender is positively correlated with attrition.**

**As p >0.05, H0 i.e. null hypothesis is accepted.**

**Correlation between Attrition and MonthlyIncome:**

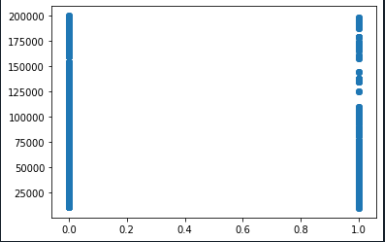
**stats,p=pearsonr(dataset.Attrition,dataset.MonthlyIncome)**

print(stats,p)

-0.031176281698114025 0.0384274849060192

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

Out[22]: <matplotlib.collections.PathCollection at 0x2368caacc48>



H0- Monthly Income doesn’t have significant impact on Attrition.

H1- Monthly Income have significant impact on Attrition.

**From r values we can conclude the Monthly Income is negatively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and PercentSalaryHike:**

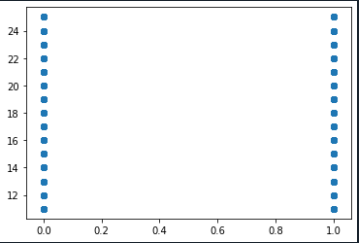
**stats,p=pearsonr(dataset.Attrition,dataset.PercentSalaryHike)**

print(stats,p)

0.03253259489105223 0.030743386433369824

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

Out[25]: <matplotlib.collections.PathCollection at 0x2368cb89c88>



H0- Percent salary hike doesn’t have significant impact on Attrition.

H1- Percent salary hike has significant impact on Attrition.

**From r values we can conclude the percent salary hike is positively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and DistanceFromHome:**

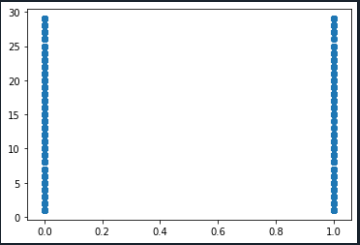
**stats,p=pearsonr(dataset.Attrition,dataset.DistanceFromHome)**

print(stats,p)

-0.009730141010179435 0.5182860428049617

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

Out[28]: <matplotlib.collections.PathCollection at 0x2368cbecd88>



H0- Distance from Home doesn’t have significant impact on Attrition.

H1- Distance from Home have significant impact on Attrition.

**From r values we can conclude the percent distance from home is negatively correlated with attrition.**

**As p >0.05, H0 i.e. null hypothesis is accepted.**

**Correlation between Attrition and JobLevel:**

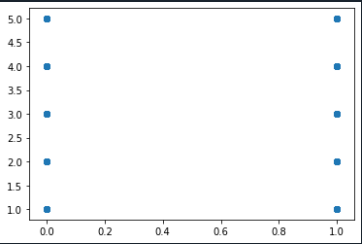
**stats,p=pearsonr(dataset.Attrition,dataset.JobLevel)**

print(stats,p)

-0.010289713287495079 0.49451717271828405

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

Out[31]: <matplotlib.collections.PathCollection at 0x2368cbec148>



H0- Job Level doesn’t have significant impact on Attrition.

H1- Job Level have significant impact on Attrition.

**From r values we can conclude the job level is negatively correlated with attrition.**

**As p >0.05, H0 i.e. null hypothesis is accepted.**

**Correlation between Attrition and YearsSinceLastPromotion:**

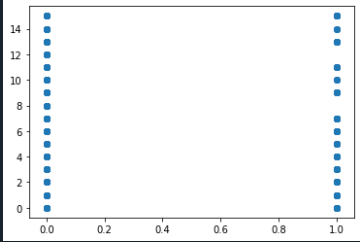
**stats,p=pearsonr(dataset.Attrition,dataset.YearsSinceLastPromotion)**

print(stats,p)

-0.03301877514258329 0.02833033618939086

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

Out[43]: <matplotlib.collections.PathCollection at 0x2368cc55248>



H0- Years Since Last Promotion doesn’t have significant impact on Attrition.

H1- Years Since Last Promotion have significant impact on Attrition.

**From r values we can conclude the job level is negatively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and TotalWorkingYears:**

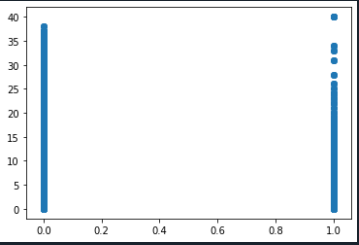
**stats,p=pearsonr(dataset.Attrition,dataset.TotalWorkingYears)**

print(stats,p)

-0.16966991684723265 1.1645434967153252e-29

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

Out[10]: <matplotlib.collections.PathCollection at 0x201d8473e88>



H0- Total Working Years doesn’t have significant impact on Attrition.

H1- Total Working Years have significant impact on Attrition.

**From r values we can conclude the Total Working Years** **is negatively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and NumCompaniesWorked:**

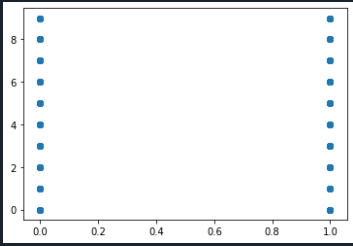
**stats,p=pearsonr(dataset.Attrition,dataset.NumCompaniesWorked)**

print(stats,p)

0.04283056724471892 0.004572057121624155

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

Out[14]: <matplotlib.collections.PathCollection at 0x201d85a7848>



H0- Number of companies worked doesn’t have significant impact on Attrition.

H1- Number of companies worked have significant impact on Attrition.

**From r values we can conclude the Number of companies worked** **is Positively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**

**Correlation between Attrition and YearsAtCompany:**

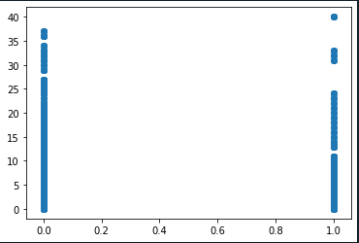
**stats,p=pearsonr(dataset.Attrition,dataset.YearsAtCompany)**

print(stats,p)

-0.1330026184252109 9.476118084864852e-19

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

Out[17]: <matplotlib.collections.PathCollection at 0x201d860a348>



H0- Years at Company worked doesn’t have significant impact on Attrition.

H1- Years at Company worked have significant impact on Attrition.

**From r values we can conclude the Number of companies worked** **is negatively correlated with attrition.**

**As p <0.05, H0 i.e. null hypothesis is rejected.**