**Correlation and P-values**

**Hypothesis testing**

**import packages:**

import pandas as pd

from scipy.stats import pearsonr

from matplotlib import pyplot as plt

**import data:**

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

**checking null values and duplicated value are existing:**

dataset.isnull()

da=dataset.dropna()

da.duplicated().sum()

da.info()

<class 'pandas.core.frame.DataFrame'>

Int64Index: 4382 entries, 0 to 4408

Data columns (total 24 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Age 4382 non-null int64

1 Attrition 4382 non-null object

2 BusinessTravel 4382 non-null object

3 Department 4382 non-null object

4 DistanceFromHome 4382 non-null int64

5 Education 4382 non-null int64

6 EducationField 4382 non-null object

7 EmployeeCount 4382 non-null int64

8 EmployeeID 4382 non-null int64

9 Gender 4382 non-null object

10 JobLevel 4382 non-null int64

11 JobRole 4382 non-null object

12 MaritalStatus 4382 non-null object

13 MonthlyIncome 4382 non-null int64

14 NumCompaniesWorked 4382 non-null float64

15 Over18 4382 non-null object

16 PercentSalaryHike 4382 non-null int64

17 StandardHours 4382 non-null int64

18 StockOptionLevel 4382 non-null int64

19 TotalWorkingYears 4382 non-null float64

20 TrainingTimesLastYear 4382 non-null int64

21 YearsAtCompany 4382 non-null int64

22 YearsSinceLastPromotion 4382 non-null int64

23 YearsWithCurrManager 4382 non-null int64

dtypes: float64(2), int64(14), object(8)

memory usage: 855.9+ KB

**Convert the attrition column values as 1 and 0**

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

da.head()

Out[4]:

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

**1.correlation and hypothesis of Attrition and distance from home.**

p= pearsonr(da.Attrition,da.DistanceFromHome)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

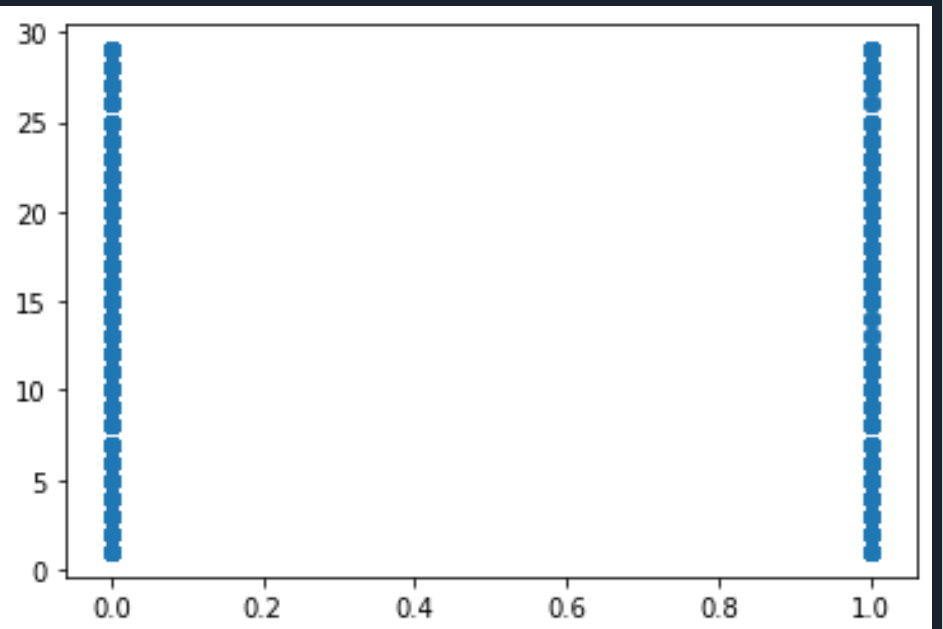
(-0.009448638515156243, 0.5317715668019558)

Accept the hypothesis

negative correlation

**scatter plot:**

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



**2.correlation and hypothesis of Attrition and Education**

p= pearsonr(da.Attrition,da.Education)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

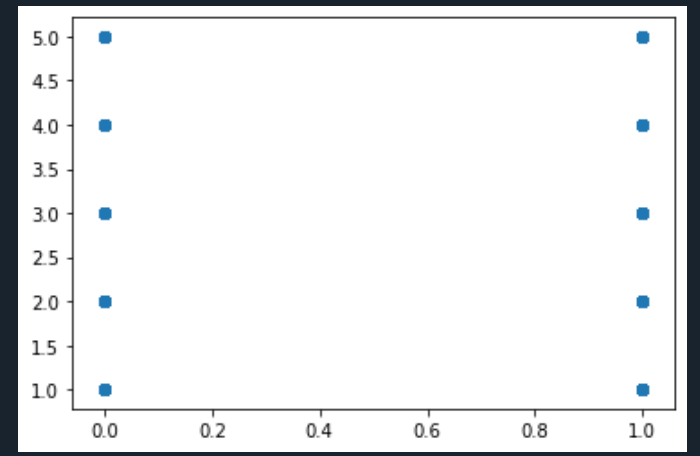
**scatter plot:**

plt.scatter(da.Attrition,da.Education)

(-0.017106307050278727, 0.25757539308157945)

Accept the hypothesis

negative correlation



**3.correlation and hypothesis of Attrition and JobLevel:**

p= pearsonr(da.Attrition,da.JobLevel)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

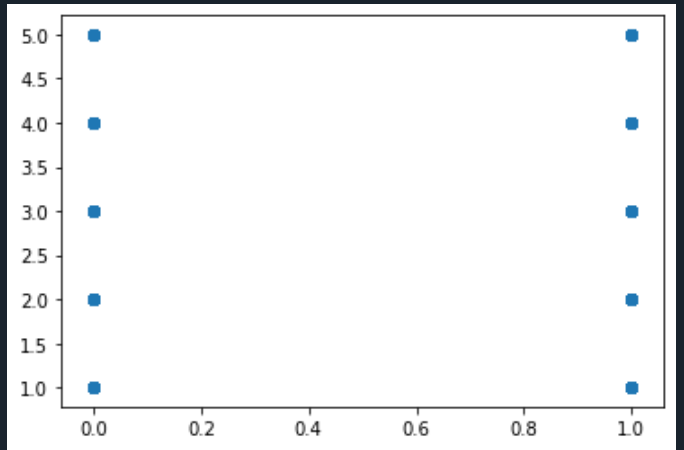
**scatter plot:**

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

(-0.012381569720790865, 0.4125489150380087)

Accept the hypothesis

negative correlation



**4.correlation and hypothesis of Attrition and MonthlyIncome:**

p= pearsonr(da.Attrition,da.MonthlyIncome)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

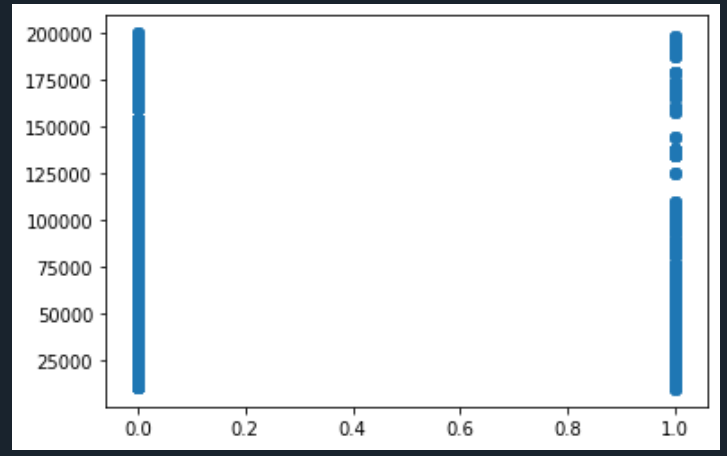
**scatter plot:**

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

(-0.030160293808460668, 0.045890862744719166)

Reject the hypothesis

negative correlation



**5.correlation and hypothesis of Attrition and NumCompaniesWorked:**

p= pearsonr(da.Attrition,da.NumCompaniesWorked)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

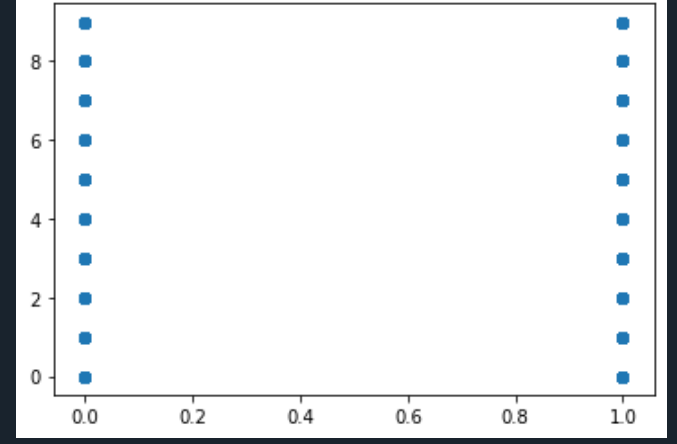
**scatter plot:**

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

(0.042830567244720875, 0.004572057121620842)

Reject the hypothesis

positive correalation



**6.correlation and hypothesis of Attrition and precentsalaryhike:**

p= pearsonr(da.Attrition,da.PercentSalaryHike)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

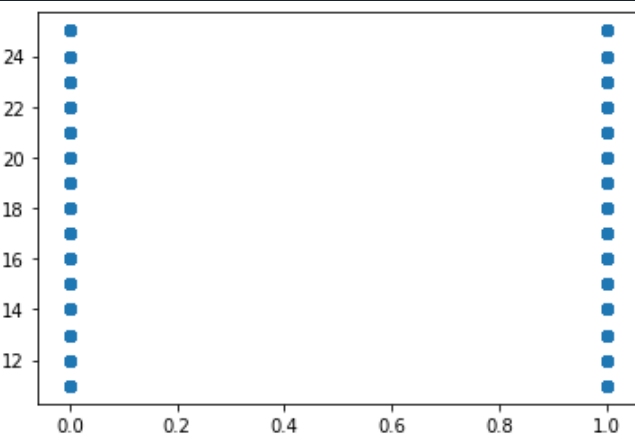
**scatter plot:**

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

(0.03315303713546665, 0.028192446935106235)

Reject the hypothesis

positive correalation



**7.correlation and hypothesis of Attrition and StockOptionLevel:**

p= pearsonr(da.Attrition,da.StockOptionLevel)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

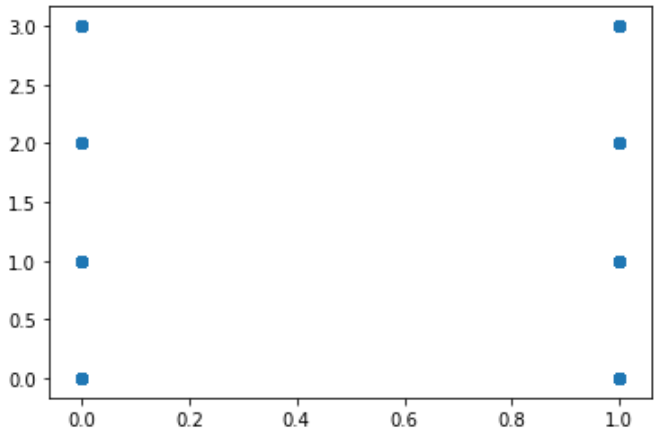
**scatter plot:**

plt.scatter(da.Attrition,da.StockOptionLevel)

(-0.008164026684984324, 0.588999635831226)

Accept the hypothesis

negative correlation



**8.correlation and hypothesis of Attrition and TotalWorkingYears**

*p= pearsonr(da.Attrition,da.TotalWorkingYears)*

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

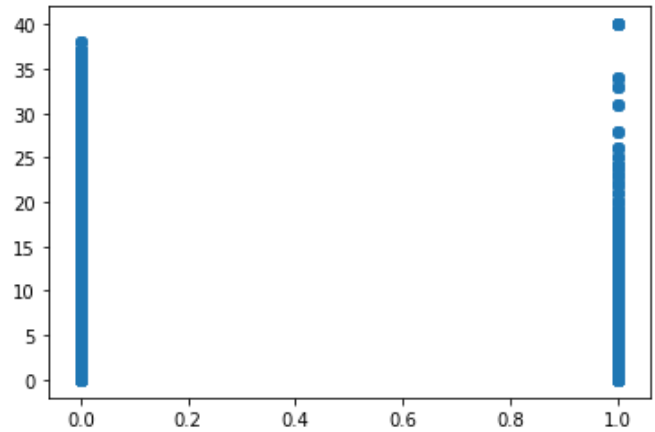
**scatter plot:**

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

(-0.16966991684723917, 1.1645434967091854e-29)

Reject the hypothesis

negative correlation



**9.correlation and hypothesis of Attrition and TrainingTimesLastYear**

p= pearsonr(da.Attrition,da.TrainingTimesLastYear)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

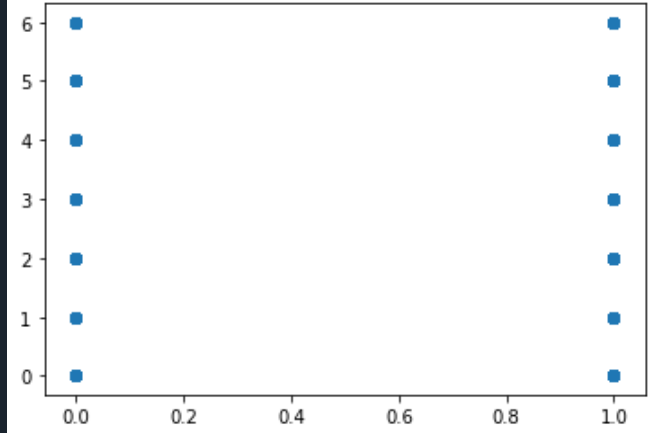
**scatter plot:**

plt.scatter(da.Attrition,da.TrainingTimesLastYear)

(-0.04758573693081737, 0.0016276603635477602)

Reject the hypothesis

negative correlation



**10.correlation and hypothesis of Attrition and YearsAtCompany**

p= pearsonr(da.Attrition,da.YearsAtCompany)

print(p)

if p[1]>=0.05:

print("Accept the hypothesis")

if p[1]<0.05:

print("Reject the hypothesis")

if p[0]==0:

print("there is no correlation")

if p[0]>0:

print("positive correalation")

if p[0]<0:

print("negative correlation")

**scatter plot:**

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

(-0.13300261842521538, 9.476118084840815e-19)

Reject the hypothesis

negative correlation

