Launching and Data Treatment –

```
import pandas as pd
from scipy.stats import pearsonr
import matplotlib.pyplot as plt

dataset = pd.read_csv("general_data_Correlation.csv")

dataset.drop_duplicates()
Out[5]:
      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]
```

To find correlation of dataset –

dataset.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]

1. Correlation of Attrition with Age -

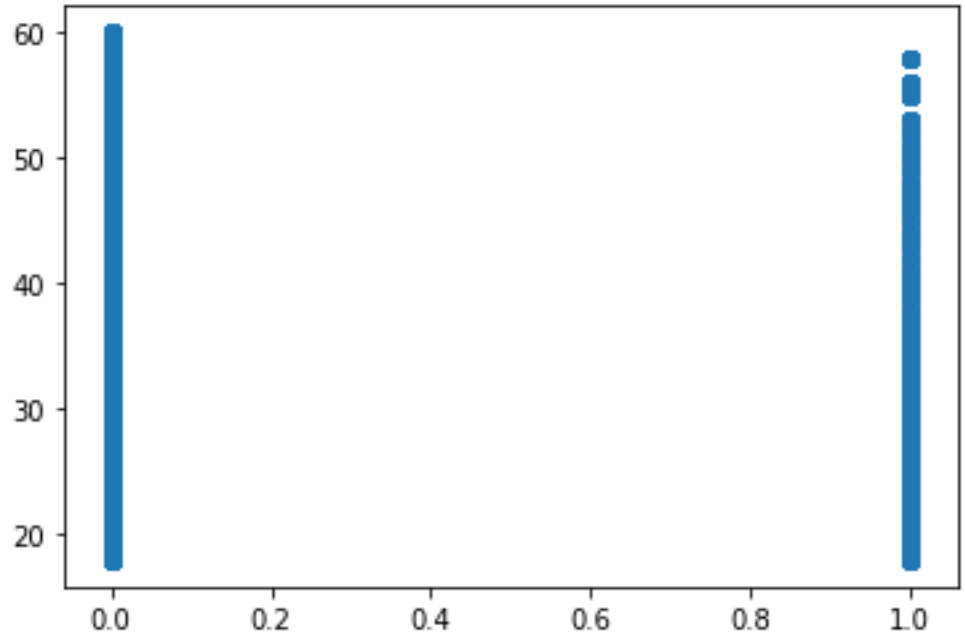
stats,p = pearsonr(dataset.Attrition,dataset.Age)

print(stats,p)

-0.15920500686577965 1.996801615886744e-26

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

Out[9]: <matplotlib.collections.PathCollection at 0x1cec7f92b88>



2. Correlation of Attrition with MonthlyIncome -

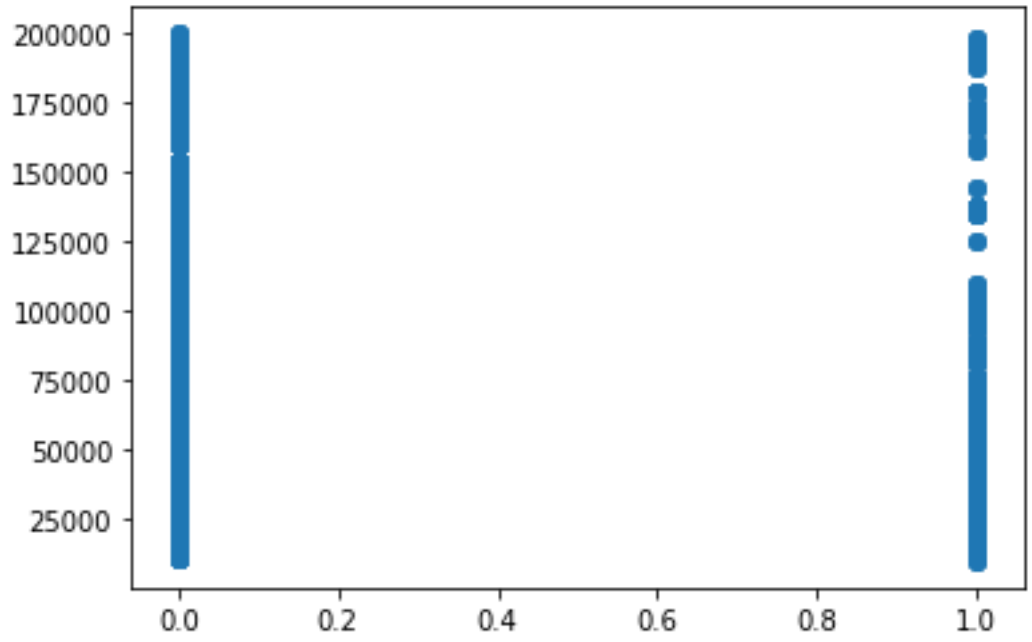
stats,p = pearsonr(dataset.Attrition,dataset.MonthlyIncome)

print(stats,p)

-0.031176281698115007 0.03842748490600132

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

Out[12]: <matplotlib.collections.PathCollection at 0x1cec7833208>



3. Correlation of Attrition with DistanceFromHome -

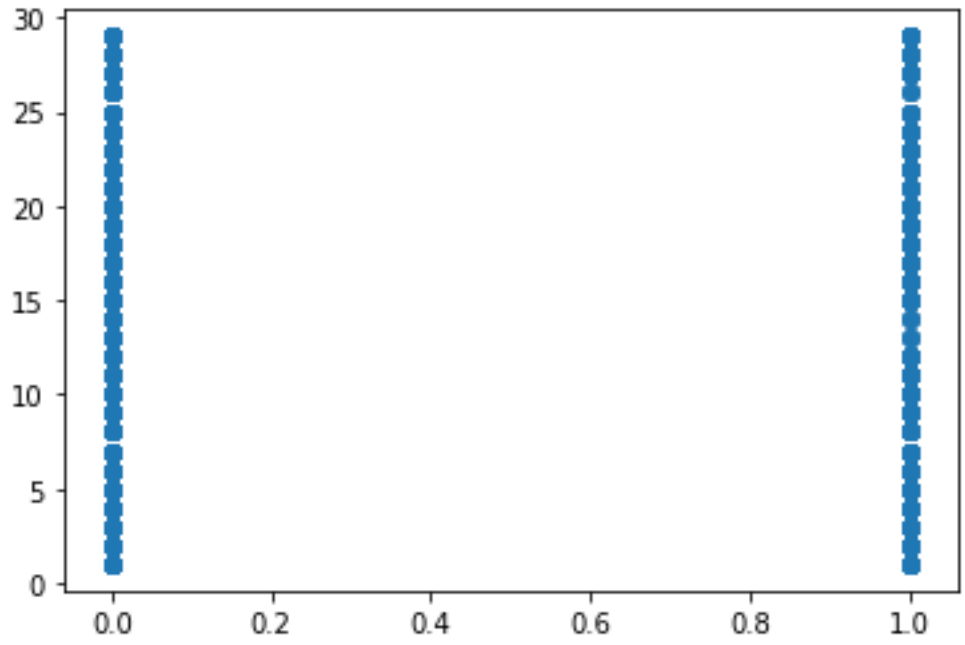
stats,p = pearsonr(dataset.Attrition,dataset.DistanceFromHome)

print(stats,p)

-0.009730141010179674 0.5182860428050771

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

Out[15]: <matplotlib.collections.PathCollection at 0x1cec804e388>



4. Correlation of Attrition with JobLevel -

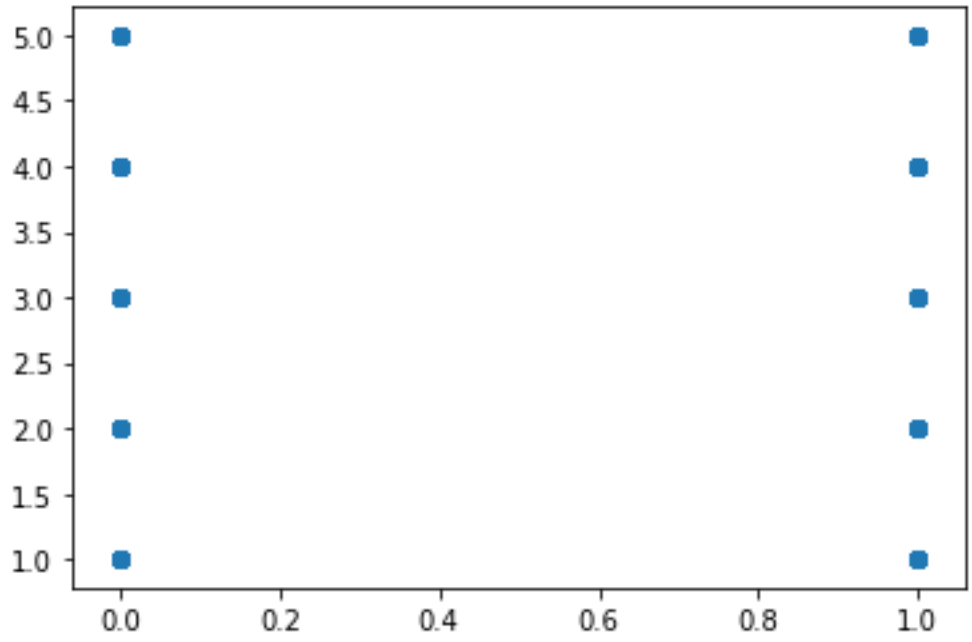
stats,p = pearsonr(dataset.Attrition,dataset.JobLevel)

print(stats,p)

-0.010289713287495042 0.49451717271828405

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

Out[18]: <matplotlib.collections.PathCollection at 0x1cec80bc488>



4. Correlation of Attrition with YearsAtCompany -

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

print(stats,p)

-0.1343922139899772 3.1638831224877484e-19

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

Out[21]: <matplotlib.collections.PathCollection at 0x1cec8118dc8>