ASSIGNMENT DAY-11

#IMPORTING PACKAGES, CSV FILE TO ANALYSE AND REMOVING NAN –DATA

```
import os

os.chdir(r"C:\Users\User\Downloads")

import pandas as pd

import matplotlib.pyplot as plt

import scipy.stats as sci

datasheet=pd.read_csv("general_data (1).csv")

datasheet=datasheet.dropna()
```

#REPLACING ATTRITION "YES"-1 AND "NO"-0 IT IS REQUIRED IN ORDER TO CALCULATE CORRELATION AND PROBABILITY , GENDER :MALE-1, FEMALE-0

datasheet.Attrition.replace(('Yes', 'No'), (1, 0), inplace=True)

datasheet.Gender.replace(('Male', 'Female'), (1, 0), inplace=True)

In [9]: datasheet

Out[9]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	...	NumCompanie
0	51	0	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	...	
1	31	1	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	...	
2	32	0	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	...	
3	38	0	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	...	
4	32	0	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	...	
...	...	...	...	...	...	...	...	...	...	...	...	
4404	29	0	Travel_Rarely	Sales	4	3	Other	1	4405	Female	...	
4405	42	0	Travel_Rarely	Research & Development	5	4	Medical	1	4406	Female	...	

In [11]: datasheet

Out[11]:

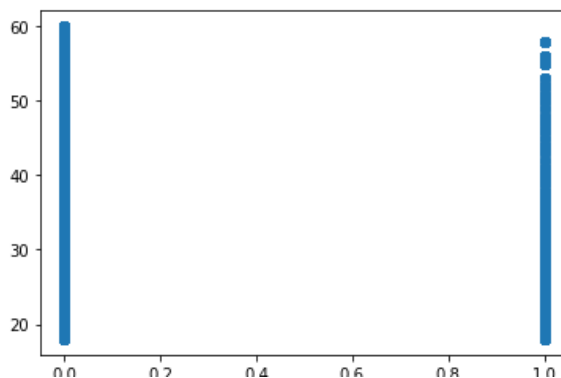| cation	EducationField	EmployeeCount	EmployeeID	Gender	...	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWor
2	Life Sciences	1	1	0	...	1.0	Y	11	8	0	
1	Life Sciences	1	2	0	...	0.0	Y	23	8	1	
4	Other	1	3	1	...	1.0	Y	15	8	3	
5	Life Sciences	1	4	1	...	3.0	Y	11	8	3	
1	Medical	1	5	1	...	4.0	Y	12	8	2	
...	...	...	...	...	...	...	...	...	...	...	
3	Other	1	4405	0	...	1.0	Y	18	8	0	
4	Medical	1	4406	0	...	3.0	Y	17	8	1	
4	Medical	1	4407	1	...	2.0	Y	15	8	0	
2	Life Sciences	1	4408	1	...	0.0	Y	20	8	0	
2	Medical	1	4409	1	...	0.0	Y	14	8	1	

#CALCULATING PROBABILITY AND CORRELATION BTWN ATTRITION AND VARIOUS OTHER FACTORS

--AGE VS ATTRITION

```
In [19]: stats,p=pearsonr(datasheet.Attrition,datasheet.Age)
         print(stats,p)
         plt.scatter(datasheet.Attrition,datasheet.Age)

         -0.15839867954096146 5.126598219398078e-26

Out[19]: <matplotlib.collections.PathCollection at 0x5918848>
```
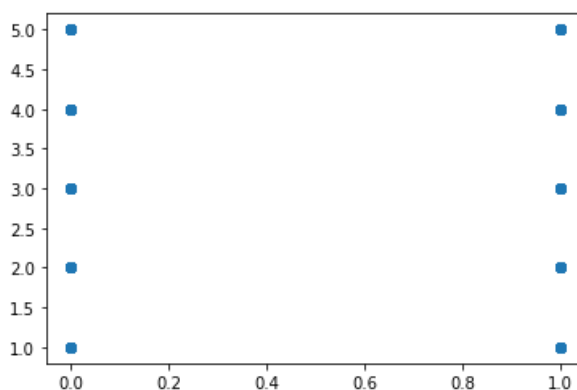

--EDUCATION VS ATTRITION

```
In [20]: stats1,p1=pearsonr(datasheet.Attrition,datasheet.Education)
         print(stats1,p1)
         plt.scatter(datasheet.Attrition,datasheet.Education)

         -0.017106307050278116 0.2575753930816995

Out[20]: <matplotlib.collections.PathCollection at 0x59183c8>
```
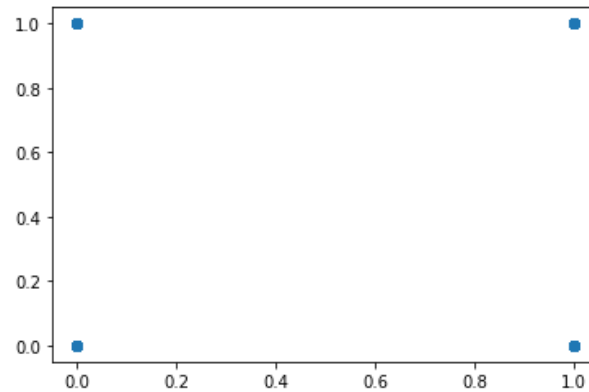
--GENDER VS ATTRITION

```
In [21]: stats2,p2=pearsonr(datasheet.Attrition,datasheet.Gender)
         print(stats2,p2)
         plt.scatter(datasheet.Attrition,datasheet.Gender)

         0.018396425194440548 0.22339966086836377

Out[21]: <matplotlib.collections.PathCollection at 0x6d8ae88>
```
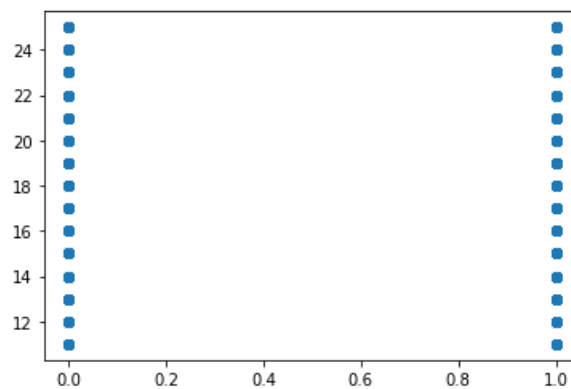

--PERCENTAGE_SAL_HIKE VS ATTRITION

```
In [22]: stats3,p3=pearsonr(datasheet.Attrition,datasheet.PercentSalaryHike)
         print(stats3,p3)
         plt.scatter(datasheet.Attrition,datasheet.PercentSalaryHike)

         0.03315303713546523 0.028192446935107012

Out[22]: <matplotlib.collections.PathCollection at 0x6f40a88>
```
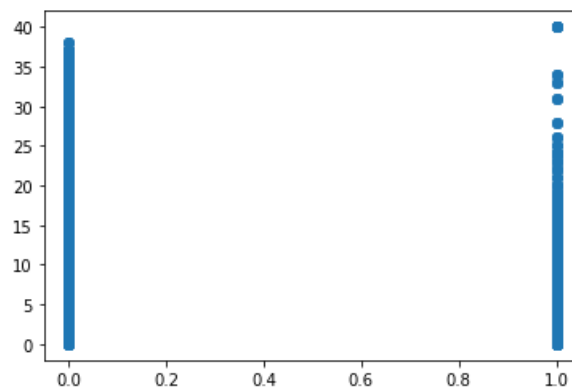
--TOTAL_WORKING_YEARS VS ATTRITION

```
In [23]: stats4,p4=pearsonr(datasheet.Attrition,datasheet.TotalWorkingYears)
         print(stats4,p4)
         plt.scatter(datasheet.Attrition,datasheet.TotalWorkingYears)

         -0.1696699168472327 1.1645434967153252e-29

Out[23]: <matplotlib.collections.PathCollection at 0x6f40988>
```
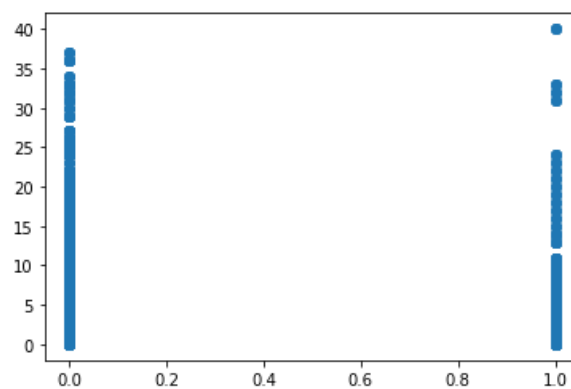

--YEARS_AT_COMPANY VS ATTRITION

```
In [24]: stats5,p5=pearsonr(datasheet.Attrition,datasheet.YearsAtCompany)
         print(stats5,p5)
         plt.scatter(datasheet.Attrition,datasheet.YearsAtCompany)

         -0.13300261842521083 9.476118084864852e-19

Out[24]: <matplotlib.collections.PathCollection at 0xb9b5288>
```
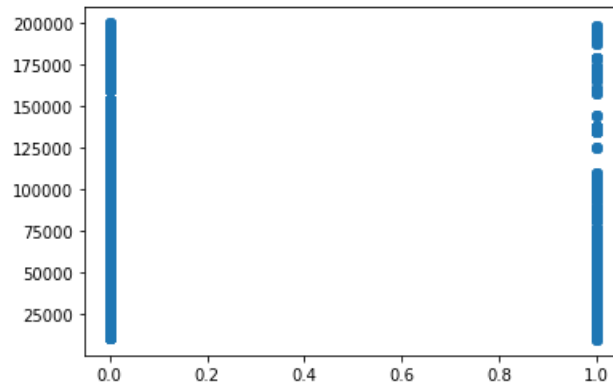
--MONTHLY_INCOME VS ATTRITION

```
In [25]: stats6,p6=pearsonr(datasheet.Attrition,datasheet.MonthlyIncome)
         print(stats6,p6)
         plt.scatter(datasheet.Attrition,datasheet.MonthlyIncome)
```

```
-0.03016029380845958 0.04589086274474114
```

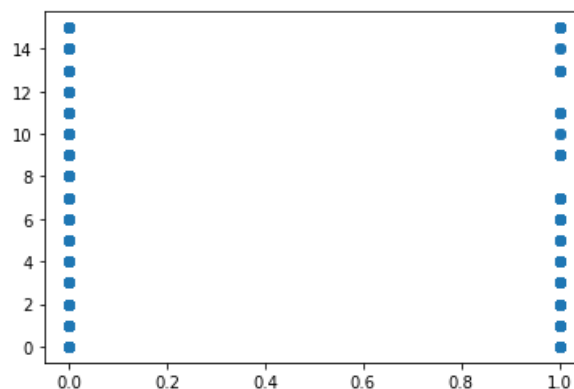Out[25]: <matplotlib.collections.PathCollection at 0xba25848>

--YEAR_SINCE_LAST+PROMOTION VS ATTRITION

```
In [27]: stats7,p7=pearsonr(datasheet.Attrition,datasheet.YearsSinceLastPromotion)
         print(stats7,p7)
         plt.scatter(datasheet.Attrition,datasheet.YearsSinceLastPromotion)
```

```
-0.03142315056330794 0.03752293607393637
```

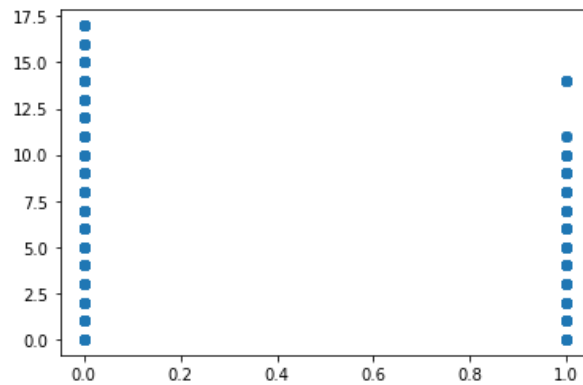Out[27]: <matplotlib.collections.PathCollection at 0xbf2e2c8>

--YEARS_WITH_CURR_MANAGER VS ATTRITON

```
In [28]: stats8,p8=pearsonr(datasheet.Attrition,datasheet.YearsWithCurrManager)
         print(stats8,p8)
         plt.scatter(datasheet.Attrition,datasheet.YearsWithCurrManager)
```

         -0.154691536902868 7.1053696467956645e-25

Out[28]: <matplotlib.collections.PathCollection at 0xb972488>

#CORRELATION MATRIX

```
In [29]: corr_matrix=datasheet.corr()
```

```
In [30]: corr_matrix
```

Out[30]:

	Age	Attrition	DistanceFromHome	Education	EmployeeCount	EmployeeID	Gender	JobLevel	MonthlyIncome	NumCompar
Age	1.000000	-0.158399	0.007376	-0.033900	NaN	0.008105	-0.040625	-0.001137	-0.045163	
Attrition	-0.158399	1.000000	-0.009449	-0.017106	NaN	-0.004621	0.018396	-0.012382	-0.030160	
DistanceFromHome	0.007376	-0.009449	1.000000	-0.007491	NaN	-0.000326	-0.042610	-0.039990	-0.022757	
Education	-0.033900	-0.017106	-0.007491	1.000000	NaN	-0.009389	-0.017671	0.045822	0.007289	
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeID	0.008105	-0.004621	-0.000326	-0.009389	NaN	1.000000	-0.012812	-0.003090	0.007865	
Gender	-0.040625	0.018396	-0.042610	-0.017671	NaN	-0.012812	1.000000	-0.026019	0.006805	
JobLevel	-0.001137	-0.012382	-0.039990	0.045822	NaN	-0.003090	-0.026019	1.000000	0.046688	
MonthlyIncome	-0.045163	-0.030160	-0.022757	0.007289	NaN	0.007865	0.006805	0.046688	1.000000	
NumCompaniesWorked	0.299527	0.042831	-0.014449	-0.016210	NaN	0.000719	-0.066860	-0.009759	-0.021446	
PercentSalaryHike	-0.032561	0.033153	0.037720	-0.041054	NaN	-0.004877	0.013207	0.010874	0.004607	
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	-0.031504	-0.008164	0.009353	0.002386	NaN	-0.013488	0.022043	0.000365	0.027242	
TotalWorkingYears	0.680037	-0.169670	0.009574	-0.009228	NaN	-0.001688	-0.028251	-0.036293	-0.034398	
TrainingTimesLastYear	-0.028962	-0.047586	-0.008957	0.009939	NaN	-0.012102	-0.032597	-0.031931	0.049780	
YearsAtCompany	0.311281	-0.133003	0.030746	0.005997	NaN	0.004117	-0.018659	-0.063360	0.000088	
YearsSinceLastPromotion	0.215650	-0.031423	0.002243	0.023457	NaN	0.000814	-0.023018	-0.059680	0.064470	
YearsWithCurrManager	0.201580	-0.154692	0.021773	0.005645	NaN	0.009079	0.002688	-0.053898	0.023095	

CONCLUSION:

SR.NO	QUANTITY VS ATTRITION	CORRELATION	PROBABILITY	H0 STATUS
1.	Age	-0.1583	5.12e-26	Rejected
2.	Education	-0.0170	0.2575	Accepted
3.	Gender	0.0183	0.2233	Accepted
4.	PercentSalaryHike	0.0331	0.0281	Rejected
5.	TotalWorkingHours	-0.1696	1.1645e-29	Rejected
6.	YearsAtCompany	-0.1330	9.4761e-19	Rejected
7.	MonthlyIncome	-0.0301	0.0458	Rejected
8.	YearsSinceLastPromotion	-0.0314	0.03752	Rejected
9.	YearsWithCurrentManager	-0.1546	7.1053e-25	Rejected