

:CORRELATION BETWEEN ATTRITION AND OTHER VARIABLES:

1) Correlation between Attrition and Distance From Home:

- Correlation value ' r ' = -0.00963, which means the variables are weak negatively correlated.
- Probability ' p ' value = 0.522, so we can accept the Null Hypothesis as it affects 52% of data.

```
In [34]: dataset["Attrition1"].isna().values.any()
Out[34]: True

In [35]: dataset.dropna(subset=["Attrition1"], inplace=True)

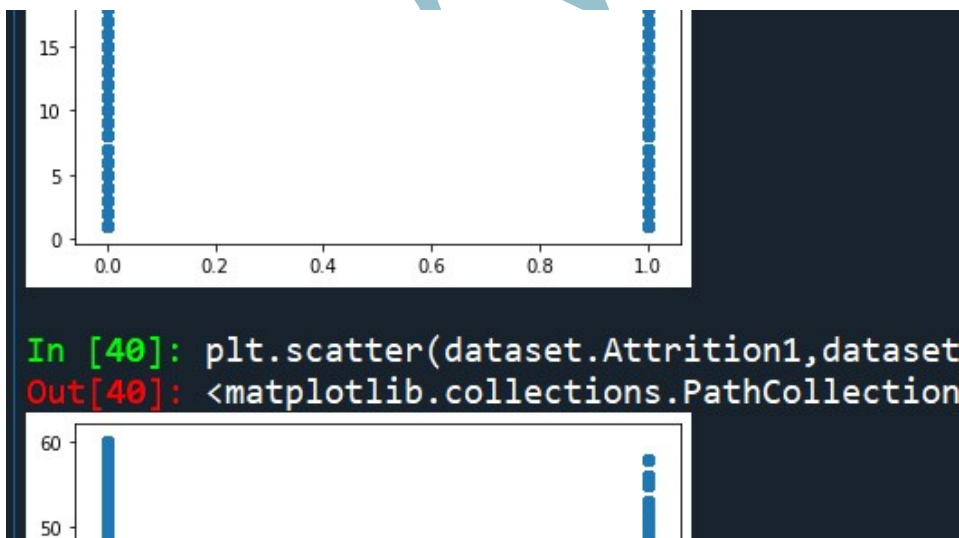
In [36]: dataset["Attrition1"].isna().values.any()
Out[36]: False

In [37]: stats, p = pearsonr(dataset.Attrition1, dataset.DistanceFromHome)

In [38]: print(stats, p)
-0.009638784678344565 0.5223162968450248
```

2) Correlation between Attrition and Age:


- Correlation value ' r ' = -0.159, which means the variables are negatively correlated.
- Probability ' p ' value = 2.09e-26, so we can reject the Null Hypothesis.



3) Correlation between Attrition and Education Level:

- Correlation value ' r ' = -0.014, which means the variables are weak negatively correlated.

- Probability 'p' value = 0.319, so we can accept the Null Hypothesis as it affects around 32% of data.



```

In [41]: stats,p=pearsonr(dataset.Attrition1,da

In [42]: print(stats,p)
-0.15917551489227316 2.0935226759299425e-26

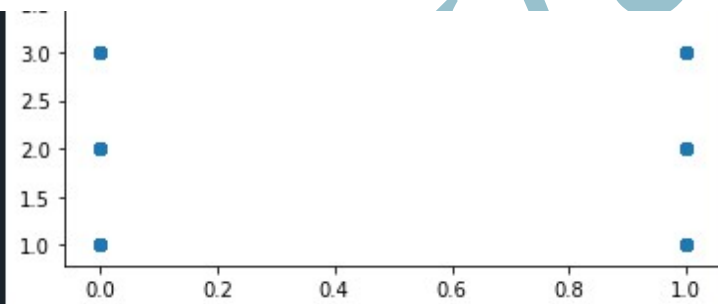
In [43]: stats,p=pearsonr(dataset.Attrition1,da

In [44]: print(stats,p)
-0.0149996869479451 0.31942298183450585

```

4) Correlation between Attrition and Job Level:

- Correlation value 'r' = -0.0102, which means the variables are weakly negatively correlated.
- Probability 'p' value = 0.497, so we can reject the Null Hypothesis.



```

In [46]: stats,p=pearsonr(dataset.Attrition1,

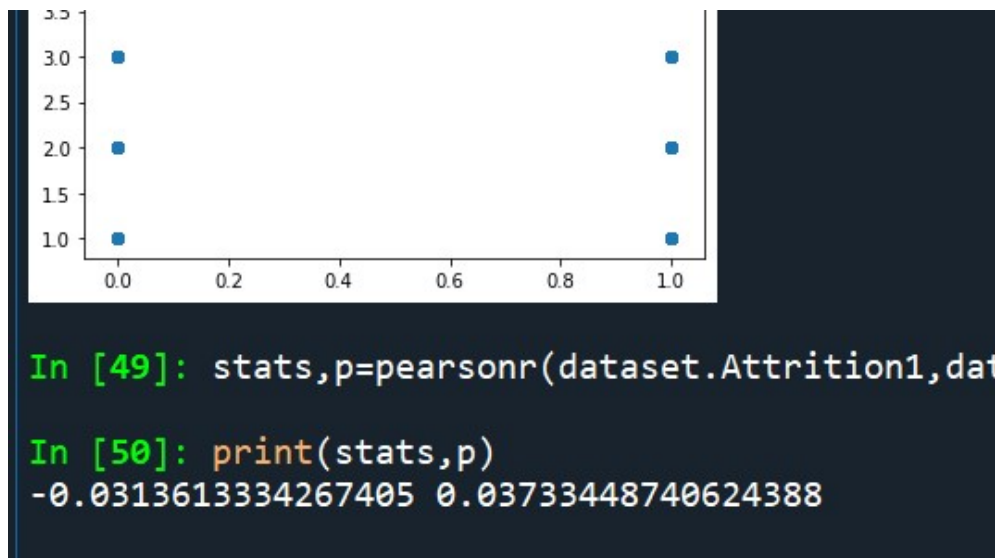
In [47]: print(stats,p)
-0.010216389132645505 0.497695798535647

```

5) Correlation between Attrition and Monthly Income:

- Correlation value 'r' = -0.0313, which means the variables are weakly negatively correlated.

- Probability 'p' value = 0.037, so we can reject the Null Hypothesis as only 3% of the data is affected.



6) Correlation between Attrition and Salary Hike Percentage:

- Correlation value 'r' = 0.0324, which means the variables are positively correlated.
- Probability 'p' value = 0.031, so we can reject the Null Hypothesis as only 3% of data is affected.

```

a = np.asarray_chkfinite(a)

File "C:\Users\dell\anaconda3\lib\site-packages\numpy\core\numeric.py", line 100, in _asarray_chkfinite
    "array must not contain infs or NaNs")

ValueError: array must not contain infs or NaNs

In [53]: stats,p=pearsonr(dataset.Attrition1,dataset.SalaryHike)

In [54]: print(stats,p)
0.032469604641723576 0.03110678713888575

```

7) Correlation between Attrition and Number of trainings conducted:

- Correlation value 'r' = -0.049, which means the variables are negatively correlated.
- Probability 'p' value = 0.00106, so we can reject the Null Hypothesis.

```

a = np.asarray_chkfinite(a)

File "C:\Users\dell\anaconda3\lib\site-packages\numpy\
"array must not contain infs or NaNs")

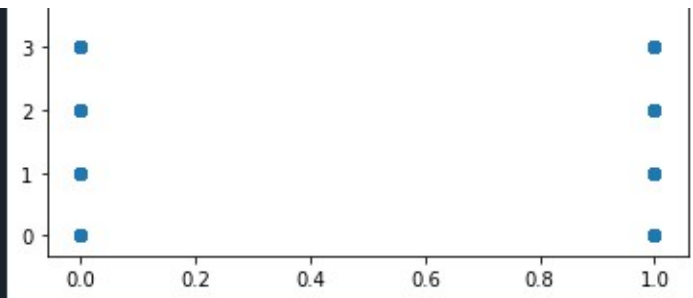
ValueError: array must not contain infs or NaNs

In [57]: stats,p=pearsonr(dataset.Attrition1,dataset.
In [58]: print(stats,p)
-0.04928403650627486 0.001063504750108934

```

8) Correlation between Attrition and Number of Years at company:

- Correlation value ' r ' = -0.134, which means the variables are negatively correlated.
- Probability ' p ' value = 3.33, so we can accept the Null Hypothesis as 33% of data is affected.



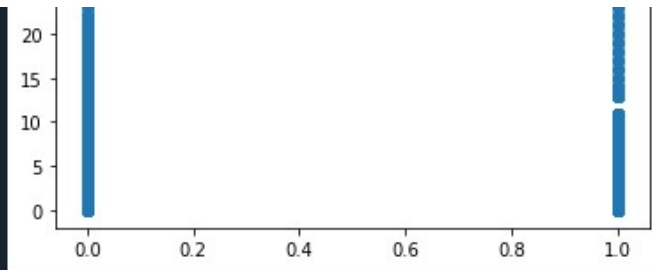
```

In [60]: stats,p=pearsonr(dataset.Attrition1,dat
In [61]: print(stats,p)
-0.13433741288732426 3.33237622244819e-19

```

9) Correlation between Attrition and Number of years worked since last Promotion:

- Correlation value ' r ' = -0.0331, which means the variables are weakly negatively correlated.
- Probability ' p ' value = 0.027, so we can reject the Null Hypothesis as only 2% data is affected.

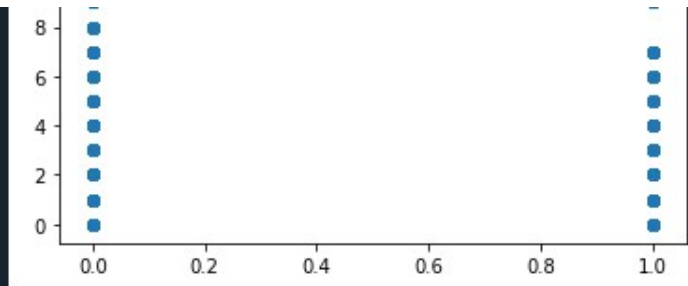


```
In [63]: stats,p=pearsonr(dataset.Attrition1,dataset.YearsWorked)

In [64]: print(stats,p)
-0.033126684437498256 0.027852890171090966
```

10) Correlation between Attrition and Number of years worked with Current Manager:

- Correlation value ' r ' = -0.156, which means the variables are negatively correlated.
- Probability ' p ' value = $1.96e-25$, so we can reject the Null Hypothesis.



```
In [66]: stats,p=pearsonr(dataset.Attrition1,dataset.YearsWorked)

In [67]: print(stats,p)
-0.15605022524858336 1.976589930519286e-25
```