In [1]:
```
import pandas as pd
dataset = pd.read_csv("general_data.csv")

dataset.head()
```

Out[1]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID
0	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1
1	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2
2	32	No	Travel_Frequently	Research & Development	17	4	Other	1	3
3	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4
4	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5

5 rows × 24 columns

In [2]:
```
#drop the row with null values
dataset1=dataset.dropna()

from scipy.stats import pearsonr

#convert Attrition values yes and no as 0 and 1 to be able to correlate and calculate the r
 value
dataset1.Attrition = dataset1.Attrition.replace('No', 0)
dataset1.Attrition = dataset1.Attrition.replace('Yes', 1)

dataset1.head()
```

```
C:\Users\Joy\anaconda3\lib\site-packages\pandas\core\generic.py:5303: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guid
e/indexing.html#returning-a-view-versus-a-copy
  self[name] = value
```

Out[2]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID
0	51	0	Travel_Rarely	Sales	6	2	Life Sciences	1	1
1	31	1	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2
2	32	0	Travel_Frequently	Research & Development	17	4	Other	1	3
3	38	0	Non-Travel	Research & Development	2	5	Life Sciences	1	4
4	32	0	Travel_Rarely	Research & Development	10	1	Medical	1	5

5 rows × 24 columns

1. Correlation between Attrition and Age

In [3]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.Age)
print(stats,p)
```

```
-0.15839867954096146 5.126598219398078e-26
```

Here in the above test we found the correlation between the age and attrition and we got the p-value is less than 0.005 so (h0) is rejected and (ha) is accepted and we can conclude that there is significant correlation between the age and attrition

2. Next to find the correlation between the Attrition and DistanceFromHome

In [4]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.DistanceFromHome)
print(stats,p)
```

```
-0.009448638515156003 0.5317715668019634
```

The p-value is greater than 0.005 so (h0) is accepted and we can conclude that there is no correlation between the DistanceFromHome and Attrition

3. Correlation btween Attrition and Education

In [5]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.Education)
print(stats,p)
```

```
-0.017106307050278116 0.25757539308616995
```

Here in the above test we found the correlation between Education and Attrition, the p-value is greater than 0.005 so (h0) is accepted and we can conclude that there is no correlation between the Education and Attrition

4. Correlation between the Attrition and NumCompaniesWorked

In [6]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.NumCompaniesWorked)
print(stats,p)
```

```
0.04283056724471901 0.004572057121625154
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is correlation between the NumCompaniesWorked and Attrition.

5. Correlation between the Attrition and PercentSalaryHike

In [7]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.PercentSalaryHike)
print(stats,p)
```

```
0.03315303713546523 0.028192446935107012
```

The p-value is greater than 0.005 so (h0) is accepted and we can conclude that there is no correlation between the PercentSalaryHike and Attrition

6. Correlation between the Attrition and StockOptionLevel

In [8]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.StockOptionLevel)
print(stats,p)
```

```
-0.008164026684984027 0.5889996358312328
```

The p-value is greater than 0.005 so (h0) is accepted and we can conclude that there is no correlation between the StockOptionLevel and Attrition

7. Correlation between the Attrition and TotalWorkingYears

In [9]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.TotalWorkingYears)
print(stats,p)
```

```
-0.1696699168472327 1.1645434967153252e-29
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is significant correlation between the TotalWorkingYears and Attrition

8. Correlation between the Attrition and TrainingTimesLastYear

In [10]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.TrainingTimesLastYear)
print(stats,p)
```

```
-0.047585736930815525 0.0016276603635481809
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is correlation between the TrainingTimesLastYear and Attrition

8. Correlation between the Attrition and YearsAtCompany

In [11]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.YearsAtCompany)
print(stats,p)
```

```
-0.13300261842521083 9.476118084864852e-19
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is correlation between the YearsAtCompany and Attrition

9. Correlation between the Attrition and YearsSinceLastPromotion

In [12]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.YearsSinceLastPromotion)
print(stats,p)
```

```
-0.03142315056330794 0.03752293607393637
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is correlation between the YearsSinceLastPromotion and Attrition

10. Correlation between the Attrition and YearsWithCurrManager

In [13]:
```
stats,p=pearsonr(dataset1.Attrition, dataset1.YearsWithCurrManager)
print(stats,p)
```

```
-0.1546915369028688 7.10536964679566 45e-25
```

The p-value is less than 0.005 so (h0) is rejected and we can conclude that there is correlation between the YearsWithCurrManager and Attrition

Correlation Matrix

In [17]:
```
dataset1.corr()
```

Out[17]:

	Age	Attrition	DistanceFromHome	Education	EmployeeCount	EmployeeID	JobLevel	Monthl
Age	1.000000	-0.158399	0.007376	-0.033900	NaN	0.008105	-0.001137	-
Attrition	-0.158399	1.000000	-0.009449	-0.017106	NaN	-0.004621	-0.012382	-
DistanceFromHome	0.007376	-0.009449	1.000000	-0.007491	NaN	-0.000326	-0.039990	-
Education	-0.033900	-0.017106	-0.007491	1.000000	NaN	-0.009389	0.045822	
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeID	0.008105	-0.004621	-0.000326	-0.009389	NaN	1.000000	-0.003090	
JobLevel	-0.001137	-0.012382	-0.039990	0.045822	NaN	-0.003090	1.000000	
MonthlyIncome	-0.045163	-0.030160	-0.022757	0.007289	NaN	0.007865	0.046688	
NumCompaniesWorked	0.299527	0.042831	-0.014449	-0.016210	NaN	0.000719	-0.009759	-
PercentSalaryHike	-0.032561	0.033153	0.037720	-0.041054	NaN	-0.004877	0.010874	
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	-0.031504	-0.008164	0.009353	0.002386	NaN	-0.013488	0.000365	
TotalWorkingYears	0.680037	-0.169670	0.009574	-0.009228	NaN	-0.001688	-0.036293	-
TrainingTimesLastYear	-0.028962	-0.047586	-0.008957	0.009939	NaN	-0.012102	-0.031931	
YearsAtCompany	0.311281	-0.133003	0.030746	0.005997	NaN	0.004117	-0.063360	
YearsSinceLastPromotion	0.215650	-0.031423	0.002243	0.023457	NaN	0.000814	-0.059680	
YearsWithCurrManager	0.201580	-0.154692	0.021773	0.005645	NaN	0.009079	-0.053898	