

Day-7 Assignment

Analysing the Data and finding reason of Attrition and ways to minimize Attrition

1) Analysing Total number of People who left from whole data

```
In [8]: df["Attrition"].value_counts()
```

```
Out[8]: No      3699  
       Yes       711  
       Name: Attrition, dtype: int64
```

From the Given Data 711 People Have left

```
In [9]: df["Attrition"].value_counts(normalize=True)
```

```
Out[9]: No      0.838776  
       Yes      0.161224  
       Name: Attrition, dtype: float64
```

Which means 16% People Left

2) There is 62% Male Attrition and 37% Female Attrition

Analysis of Attrition on gender Basis

```
In [5]: # Getting Number of Attrition == Yes and Gender == Male  
attr_yes_male_count = (df[df.Attrition=="Yes"].Gender == "Male").value_counts()[1]  
attr_yes_male_count
```

```
Out[5]: 441
```

```
In [6]: # Getting Number of Attrition == Yes and Gender == Female  
attr_yes_female_count = (df[df.Attrition=="Yes"].Gender == "Female").value_counts()[1]  
attr_yes_female_count
```

```
Out[6]: 270
```

```
In [7]: total_attr = df["Attrition"].value_counts()["Yes"]  
print("Total Employees Who leave :",total_attr)  
print("Number of Male leaving :",attr_yes_male_count,(attr_yes_male_count/total_attr)*100,"%")  
print("Number of Female Leaving :",attr_yes_female_count,(attr_yes_female_count/total_attr)*100,"%")
```

```
Total Employees Who leave : 711  
Number of Male leaving : 441 62.0253164556962 %  
Number of Female Leaving : 270 37.9746835443038 %
```

So we Conclude from above result that rate of male attrition is more than Female

ie. male = 62%

Female = 37.9%

3) Analysis Based on Marital Status of Employees

```
In [21]: total_attr = df["Attrition"].value_counts()["Yes"]
print("Total Employees Who leave :",total_attr)
print("Number of Married People leaving :",attr_yes_married_count,(attr_yes_married_count/total_attr)*100,"%")
print("Number of Single People Leaving :",attr_yes_single_count,(attr_yes_single_count/total_attr)*100,"%")
print("Number of Divorced People Leaving :",attr_yes_divorced_count,(attr_yes_divorced_count/total_attr)*100,"%")

Total Employees Who leave : 711
Number of Married People leaving : 252 35.44303797468354 %
Number of Single People Leaving : 360 50.63291139240506 %
Number of Divorced People Leaving : 99 13.924050632911392 %
```

Conclusion From Above analysis :

Single People Are leaving mostly. Rate: 50.63%
Married Rate : 35.44%
Divorced Rate: 13.92%

The Divorced People are Mostly not Leaving

4) Analysis based on PercentSalaryHike

Analysis Based on Percent Salary Hikes

```
In [43]: df_attr_yes.PercentSalaryHike.value_counts(ascending=False)
```

```
Out[43]: 13    102
         12     90
         11     90
         14     87
         15     63
         17     45
         19     42
         18     39
         22     30
         16     30
         21     27
         20     27
         25     15
         23     15
         24      9
         Name: PercentSalaryHike, dtype: int64
```

Person Having Salary Hike more than 19% are Leaving Less

Lets See The Analysis of PercentSalary Increase by Grouping in range

```
In [37]: # Groups 13-15 , 16-18 , 19 - 21 , 22-24 + 25
```

```
In [23]: def group_PercentSalaryHike(n,m):
          sum=0
          for i in range(n,m+1):
              sum += int(df_attr_yes.PercentSalaryHike.value_counts()[i])
          return sum
```

```
In [24]: print("%Hike | Count")
print("13-15 :",group_PercentSalaryHike(13,15))
print("16-18 :",group_PercentSalaryHike(16,18))
print("19-21 :",group_PercentSalaryHike(19,21))
print("22-25 :",group_PercentSalaryHike(22,25))

%Hike | Count
13-15 : 252
16-18 : 114
19-21 : 96
22-25 : 69
```

Person with 13 - 15% and 16 - 18% are Having high attrition Rate

5) Analysis based on Age

Attrition Analysis Based On Age

```
In [27]: df_attr_yes["Age"].value_counts()
```

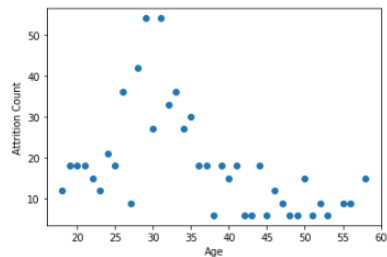
```
Out[27]: 29    54
          31    54
          28    42
          26    36
          33    36
          32    33
          35    30
          30    27
          34    27
          24    21
          25    18
          36    18
          44    18
          21    18
          20    18
          41    18
          19    18
          39    18
          37    18
          22    15
          58    15
          40    15
          50    15
          18    12
          23    12
          46    12
          27     9
          55     9
          52     9
          56     9
          47     9
          38     6
          49     6
          45     6
          51     6
          43     6
          53     6
          42     6
          48     6
          Name: Age, dtype: int64
```

```
In [28]: ages = df_attr_yes["Age"].value_counts()
         ages.index
```

```
Out[28]: Int64Index([29, 31, 28, 26, 33, 32, 35, 30, 34, 24, 25, 36, 44, 21, 20, 41, 19,
                    39, 37, 22, 58, 40, 50, 18, 23, 46, 27, 55, 52, 56, 47, 38, 49, 45,
                    51, 43, 53, 42, 48],
                    dtype='int64')
```

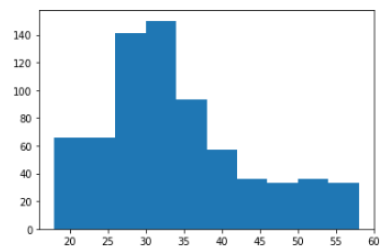
```
In [29]: plt.scatter(ages.index, ages)
         plt.xlabel("Age")
```

```
Out[29]: Text(0, 0.5, 'Attrition Count')
```



```
In [22]: plt.hist(df_attr_yes.Age)
```

```
Out[22]: (array([ 66.,  66., 141., 150.,  93.,  57.,  36.,  33.,  36.,  33.]),
         array([18., 22., 26., 30., 34., 38., 42., 46., 50., 54., 58.]),
         <a list of 10 Patch objects>)
```



We Observe that Age Range 27-37 Have a High Attrition Rate

6) Analysis Department wise

Analysis Department Wise

```
In [40]: df_attr_yes.Department.value_counts()
```

```
Out[40]: Research & Development    453  
Sales                               201  
Human Resources                    57  
Name: Department, dtype: int64
```

```
In [24]: attr_yes_rnd_count = df_attr_yes.Department.value_counts()["Research & Development"]  
attr_yes_rnd_count
```

```
Out[24]: 453
```

```
In [25]: attr_yes_sales_count = df_attr_yes.Department.value_counts()["Sales"]  
attr_yes_sales_count
```

```
Out[25]: 201
```

```
In [26]: attr_yes_humRes_count = df_attr_yes.Department.value_counts()["Human Resources"]  
attr_yes_humRes_count
```

```
Out[26]: 57
```

```
In [27]: total_attr = df["Attrition"].value_counts()["Yes"]  
print("Total Employees Who leave :",total_attr)  
print("Number of People leaving of R & D:",attr_yes_rnd_count,(attr_yes_rnd_count/total_attr)*100,"%")  
print("Number of People Leaving of Sales:",attr_yes_sales_count,(attr_yes_sales_count/total_attr)*100,"%")  
print("Number of People Leaving of Human Resources:",attr_yes_humRes_count,(attr_yes_humRes_count/total_attr)*100,"%")  
  
Total Employees Who leave : 711  
Number of People leaving of R & D: 453 63.71308016877637 %  
Number of People Leaving of Sales: 201 28.270042194092827 %  
Number of People Leaving of Human Resources: 57 8.016877637130802 %
```

Therefore we can analyse from above data that People from Research and Development Department have a high Attrition Rate

7) Analysis Based On Job-Role

Analysis Based On Job Role of Employee

```
In [67]: df_attr_yes.JobRole.value_counts()
```

```
Out[67]: Sales Executive           165  
Research Scientist                159  
Laboratory Technician            126  
Research Director                 57  
Healthcare Representative        57  
Manufacturing Director           48  
Manager                          42  
Sales Representative              36  
Human Resources                  21  
Name: JobRole, dtype: int64
```

We Observe from above data that Sales Executive, Research Scientist and Laboratory Technician have a high Attrition Rate As compared to other JobRole

8) Analysis Based on Years at Company

Analysis Based on Years At Company

```
In [66]: df_attr_yes.YearsAtCompany.value_counts()
```

```
Out[66]: 1    177
         2     81
         5     63
         3     60
         4     57
        10     54
         0     48
         7     33
         6     27
         8     27
         9     24
        13      6
        11      6
        14      6
        33      3
        15      3
        16      3
        17      3
        18      3
        19      3
        20      3
        21      3
        22      3
        23      3
        24      3
        31      3
        32      3
        40      3
        Name: YearsAtCompany, dtype: int64
```

We Observe from above data that The people who are at the company ≥ 11 years have a low attrition Rate

Where as the people < 11 Years have High Attrition Rate

9) Analysis Based On Education Field of the Employee

Analysis based on Education Field of the Employee

```
In [17]: df_attr_yes.EducationField.value_counts()
```

```
Out[17]: Life Sciences    303
         Medical          225
         Marketing        75
         Technical Degree  45
         Human Resources   33
         Other            30
         Name: EducationField, dtype: int64
```

From above we observe that Employees of Life Science and Medical Field have a high Attrition Rate

10.) Analysis Based on total work Experience of Employees

Analysis Based on Total Work Experience on employees

ie.TotalWorkingYears

```
In [23]: df_attr_yes.TotalWorkingYears.value_counts()
```

```
Out[23]: 1.0    119
        10.0   75
         6.0   66
         7.0   54
         5.0   48
         8.0   47
         4.0   36
         9.0   30
         3.0   27
         2.0   27
        11.0   21
         0.0   15
        15.0   15
        12.0   15
        18.0   12
        14.0   12
        19.0    9
        24.0    9
        16.0    9
        17.0    9
        13.0    9
        40.0    6
        22.0    6
        20.0    6
        23.0    6
        25.0    3
        33.0    3
        21.0    3
        26.0    3
        31.0    3
        28.0    3
        34.0    3
        Name: TotalWorkingYears, dtype: int64
```

```
In [36]: # Grouping The data
def getExpTotal(n,m):
    sum=0
    for i in range(n,m+1):
        try:
            sum += df_attr_yes.TotalWorkingYears.value_counts()[float(i)]
        except:
            pass
    return sum
```

```
In [33]: workExp_less_than_16 = getExpTotal(1,15)
workExp_greater_than_15 = getExpTotal(16,40)
```

```
In [35]: total_attr = df["Attrition"].value_counts()["Yes"]
print("Total Employes Who leave :",total_attr)
print("Number of People leaving Having Experience between 1-15:",workExp_less_than_16 ,(workExp_less_than_16/total_attr)*100,"%")
print("Number of People leaving Having Experience between 16-40:",workExp_greater_than_15 ,(workExp_greater_than_15/total_attr)*100,"%")

Total Employes Who leave : 711
Number of People leaving Having Experience between 1-15: 601 84.52883263009845 %
Number of People leaving Having Experience between 16-40: 93 13.080168776371309 %
```

From Above analysis People having Experience less than 16 years have a attrition Attrition Rate

Where As people having Experience Above 15 Years Have less Attrition Rate

Summarizing the Analysis

- 1) Male Attrition is more than the Females,** So a survey regarding the issues faced by Males in the organization could help the organization to reduce Male Attrition
- 2) Less PercentSalaryHike** given to the employees may be one of the reasons too so Salary hikes can be given.
- 3) People of age 27-37 have a high attrition rate,** so the Salary hikes can be granted to this Age-range people.
- 4) R & D Department People have a high Attrition Rate,** a survey regarding the issues faced by the R&D Dept. People can help to know problem in that Department.
- 5) The Sales Executives, Research Scientists and Laboratory Technician Have a high Attrition Rate,** So the survey for the people of this job- Roles can help to know about the issues.
- 6) The People who are with the company less than 11 Years have a high attrition rate.**
- 7) The People having total Experience less than 16 years have a high Attrition rate,** so for these kind of people some interesting activities can be conducted.

Mostly the young Employees who have less experience are leaving, so a reason might be their like they might be not getting the chance to work on what technology they like, so a survey for the young employees can be conducted and the company can put employees on the technologies which they like.

This Solution might help reduce the Attrition rate of young Employees.