

## Day 10 Assignment

### 10 Hypothesis That could Justify the attrition Patterns in the data.

#### 1) Average PercentSalaryHike rate of attrition=Yes is $\geq 14$

```
In [12]: df_attr_yes.PercentSalaryHike.value_counts()

Out[12]: 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
```

#### 2) Average Distance from home of attrition=Yes is $\leq 10$

```
In [63]: df_attr_yes.DistanceFromHome.value_counts()

Out[63]: 2    114
         1    96
        10    60
         7    48
         9    36
         4    33
         8    33
        15    33
         3    30
         5    30
         6    21
        19    18
        18    15
        16    15
        12    15
        29    12
        17    12
        28    12
        23    12
        24     9
        27     9
        11     9
        22     9
        20     9
        25     6
        21     6
        26     3
        13     3
        14     3
         Name: DistanceFromHome, dtype: int64
```

### 3) Average Age of attrition=Yes is 33

```
In [64]: df_attr_yes.Age.value_counts()

Out[64]: 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 [65]: df_attr_yes.Age.mean()
```

```
Out[65]: 33.607594936708864
```

### 4) The Divorced People are Mostly not Leaving

### 5) Single People Are leaving mostly

```
In [21]: total_attr = df["Attrition"].value_counts()["Yes"]
print("Total Employes 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 Employes 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

## 6) Attrition rate of Female are more than male

```
Out[15]:  
  
In [16]: 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 : 270 37.9746835443038 %  
Number of Female Leaving : 441 62.0253164556962 %
```

**So we Conclude from above result that rate of female attrition is more than Male ¶**

ie. Female = 62%

Male = 37.9%

## 7) People from Research and Development Department have a high Attrition Rate

```
In [40]: df_attr_yes.Department.value_counts()  
  
Out[40]: Research & Development    453  
Sales                            201  
Human Resources                   57  
Name: Department, dtype: int64
```

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

## 8) The people who are at the company >=11 years have a low attrition Rate

**Where as the people <11 Years have High Attrition Rate**

```
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) Sales Executive, Research Scientist and Laboratory Technician have a high Attrition Rate As compared to other JobRole

```
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

## 10.) Employees of Life Science and Medical Field have a high Attrition Rate

### 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