Load Data

```
import pandas as pd
dataset=pd.read csv("employee-attrition-dataset.csv")
dataset.head()
dataset.tail()
      Age Attrition
                             BusinessTravel DailyRate
Department \
1465
          36
                     No
                         Travel Frequently
                                                     884 Research &
Development
1466
          39
                     No
                              Travel Rarely
                                                     613
                                                          Research &
Development
1467
                     No
                              Travel Rarely
                                                     155
                                                          Research &
          27
Development
                     No
1468
          49
                          Travel Frequently
                                                    1023
Sales
1469
          34
                     No
                              Travel Rarely
                                                     628
                                                          Research &
Development
      DistanceFromHome
                          Education EducationField
                                                      EmployeeCount
1465
                     23
                                  2
                                            Medical
                                                                   1
                                  1
                                                                   1
1466
                      6
                                            Medical
                      4
                                  3
                                                                   1
1467
                                     Life Sciences
                      2
                                  3
1468
                                            Medical
                                                                   1
                      8
                                  3
1469
                                            Medical
                                                                   1
      EmployeeNumber
                             RelationshipSatisfaction StandardHours
                        . . .
1465
                 2061
                                                      3
                                                                    80
                                                      1
1466
                 2062
                                                                    80
                                                      2
1467
                 2064
                                                                    80
                        . . .
1468
                 2065
                                                      4
                                                                    80
                                                      1
1469
                 2068
                                                                    80
      StockOptionLevel
                         TotalWorkingYears
                                              TrainingTimesLastYear
1465
                      1
                                          17
                                                                    3
                                                                    5
                      1
1466
                                           9
                      1
                                           6
                                                                    0
1467
                      0
                                          17
                                                                    3
1468
                      0
                                                                    3
1469
                                           6
                       YearsAtCompany YearsInCurrentRole \
     WorkLifeBalance
1465
                    3
                    3
                                                          7
                                      7
1466
                                                          2
1467
                    3
                                      6
                    2
                                     9
                                                          6
1468
                                                          3
1469
                    4
```

1465 1466	YearsSinceLastPromotion 0 1	YearsWithCurrManager 3 7
1467 1468	0 0	3
1469	1	2
[5 ro	ws x 35 columns]	

import liberaries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import chi2 contingency, f oneway, ttest ind
from sklearn.preprocessing import StandardScaler
import plotly.express as px
import plotly graph objects as go
import statsmodels.api as sm
# Load dataset
dataset=pd.read csv("employee-attrition-dataset.csv")
dataset
                            BusinessTravel DailyRate
      i»¿Age Attrition
Department \
          41
                   Yes
                             Travel Rarely
0
                                                  1102
Sales
          49
                    No
                        Travel Frequently
                                                   279
                                                        Research &
Development
          37
                             Travel Rarely
                                                  1373
                   Yes
                                                        Research &
Development
          33
                    No
                         Travel Frequently
                                                  1392
                                                        Research &
Development
          27
                    No
                             Travel Rarely
                                                   591
                                                        Research &
Development
                                                        Research &
                         Travel Frequently
                                                   884
1465
          36
                    No
Development
                    No
                                                        Research &
1466
          39
                             Travel Rarely
                                                   613
Development
                                                        Research &
1467
          27
                    No
                             Travel Rarely
                                                   155
Development
1468
          49
                    No
                         Travel_Frequently
                                                  1023
Sales
1469
                    No
                             Travel Rarely
                                                   628
                                                        Research &
          34
```

Devel	lopment		
0 1 2 3 4 1465 1466	DistanceFromHome 1 8 2 3 2 23 6	Education EducationField 2 Life Sciences 1 Life Sciences 2 Other 4 Life Sciences 1 Medical 2 Medical 1 Medical	1 1 1 1 1
1467 1468 1469	4 2 8	3 Life Sciences 3 Medical 3 Medical	1 1
0 1 2 3 4	EmployeeNumber 1 2 4 5 7	<pre>. RelationshipSatisfact</pre>	ion StandardHours \ 1
1465 1466 1467 1468 1469	2061 2062 2064 2065 2068		3 80 1 80 2 80 4 80 1 80
0 1 2 3 4	StockOptionLevel 0 1 0 0 1	TotalWorkingYears Train 8 10 7 8 6	ingTimesLastYear \ 0 3 3 3 3 3
1465 1466 1467 1468 1469	1 1 1 0 0	17 9 6 17 6	3 5 0 3 3
0 1 2 3 4	1 3 3 3	earsAtCompany YearsInCurr 6 10 0 8	rentRole \ 4 7 0 7
4 1465 1466	3 3 3	2 5 7	2 2 7

```
1467
                    3
                                    6
                                                        2
                    2
                                                        6
1468
                                    9
1469
                    4
                                    4
                                                        3
      YearsSinceLastPromotion YearsWithCurrManager
0
                                                    7
1
                             1
2
                             0
                                                    0
3
                             3
                                                    0
4
                             2
                                                    2
. . .
                             0
                                                    3
1465
1466
                             1
                                                    7
                                                    3
1467
                             0
                                                    8
                             0
1468
1469
[1470 rows x 35 columns]
dataset.size
51450
dataset.shape
(1470, 35)
dataset.index
RangeIndex(start=0, stop=1470, step=1)
dataset.columns
Index(['i»¿Age', 'Attrition', 'BusinessTravel', 'DailyRate',
'Department',
       'DistanceFromHome', 'Education', 'EducationField',
'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender',
'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours',
'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear',
'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole',
'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

```
dataset.columns.to list()
['Age',
 'Attrition',
 'BusinessTravel',
 'DailyRate',
 'Department',
 'DistanceFromHome',
 'Education',
 'EducationField',
 'EmployeeCount',
 'EmployeeNumber',
 'EnvironmentSatisfaction',
 'Gender',
 'HourlyRate',
 'JobInvolvement',
 'JobLevel',
 'JobRole',
 'JobSatisfaction',
 'MaritalStatus',
 'MonthlyIncome',
 'MonthlyRate',
 'NumCompaniesWorked',
 '0ver18',
 'OverTime',
 'PercentSalaryHike',
 'PerformanceRating',
 'RelationshipSatisfaction',
 'StandardHours',
 'StockOptionLevel',
 'TotalWorkingYears',
 'TrainingTimesLastYear',
 'WorkLifeBalance',
 'YearsAtCompany',
 'YearsInCurrentRole',
 'YearsSinceLastPromotion',
 'YearsWithCurrManager']
dataset.head()
                         BusinessTravel
   Age Attrition غ«ت
                                          DailyRate
Department \
                          Travel Rarely
0
       41
                Yes
                                               1102
Sales
                     Travel Frequently
                                                279
                                                     Research &
       49
                  No
Development
       37
                Yes
                          Travel Rarely
                                               1373
                                                     Research &
Development
                 No Travel Frequently
       33
                                               1392 Research &
Development
```

4 27 Development	No	Travel_Rare	ly	591 Resear	rch &
DistanceF EmployeeNumb		ation Educa	tionField	EmployeeCo	ount
0	1	2 Life	Sciences		1
1	8	1 Life	Sciences		1
2 2	2	2	0ther		1
4					
3 5	3	4 Life	Sciences		1
4 7	2	1	Medical		1
	+ionchinCo+io	faction Cta	nda rdUourc	C+ockOn+i	anloval \
0 1	tionshipSatis	1 4	матаноитs 80 80	•	lonLevel \ 0 1
2		2 3	80 80		0 0
4		4	80		1
TotalWork		iningTimesL	astYear Wo	rkLifeBalar	nce
YearsAtCompa 0 6	ny \ 8		0		1
1	10		3		3
10 2	7		3		3
0	8		3		3
8					
4 2	6		3		3
YearsInCur 0 1 2 3	rentRole Yea 4 7 0 7 2	rsSinceLast	Promotion 0 1 0 3 2	YearsWithO	CurrManager 5 7 0 0 2
[5 rows x 35	columns]				
dataset.tail	()				
	Attrition \ No T	BusinessT ravel_Frequ		lyRate 884 Res	search &

Develo	•						
1466	39	No	Travel_	Rarely	613	Research &	צ
	opment	N.a	Tuescal	Danaly	155	Dagaanah C	
1467	27	No	Travel_	Rarety	155	Research &	X
1468	opment 49	No Tr	avel_Freq	uently	1023		
Sales				_			
1469	34	No	Travel_	Rarely	628	Research &	X
pevero	opment						
	DistanceF	romHome Ed	ucation E	ducation	Field Fmnl	LoyeeCount	\
1465	DIStancer	23	2		dical	1	`
1466		6	1		dical	ī	
1467		4		Life Sci		1	
1468		2	3	Me	dical	1	
1469		8	3	Me	dical	1	
	- 1 N		D 1 . '	1 . 6	· · · · · · · · · · · · · · · · · · ·		
1465	EmployeeN		Relation	shipSati	sfaction St		
1465		2061			3		30
1466 1467		2062 2064			1 2		30 30
1468		2065			4		30
1469		2068			1		30
1103		2000 111			-		,,
	StockOpti	onLevel To	talWorkin	gYears	TrainingTi	nesLastYear	_ \
1465		1		17		3	3
1466		1		9		5	
1467		1		6		9)
1468		0		17		3	3
1469		Θ		6		J	3
V	√orkLifeBa	lance Year	sAtCompan	v YearsT	nCurrentRo ¹	le \	
1465				5		2	
1466		3 3		7		7	
1467		3		6		2	
1468		2		9		6	
1469		4		4		3	
	V	-1 t D t	'	-11' + h C			
1465	rearssinc	eLastPromot	_	sWithCur			
1465 1466			0 1		3 7		
1467			0		3		
1468			0		8		
1469			1		2		
			_		_		
[5 rov	ws x 35 co	lumns]					
datase	et.info(ve	rbose= <mark>False</mark>	,show_cou	nts=True)		

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Columns: 35 entries, i»¿Age to YearsWithCurrManager
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
dataset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
#
     Column
                                Non-Null Count
                                                 Dtype
     _ _ _ _ _ _
 0
                                1470 non-null
                                                 int64
     ï≫;Age
 1
     Attrition
                                1470 non-null
                                                 object
 2
     BusinessTravel
                                1470 non-null
                                                 object
 3
     DailyRate
                                1470 non-null
                                                 int64
 4
                                1470 non-null
     Department
                                                 object
 5
     DistanceFromHome
                                1470 non-null
                                                 int64
                                1470 non-null
 6
     Education
                                                 int64
 7
     EducationField
                                1470 non-null
                                                 object
 8
     EmployeeCount
                                1470 non-null
                                                 int64
 9
     EmployeeNumber
                                1470 non-null
                                                 int64
                                1470 non-null
 10
    EnvironmentSatisfaction
                                                 int64
 11
     Gender
                                1470 non-null
                                                 object
 12
     HourlvRate
                                1470 non-null
                                                 int64
 13
     JobInvolvement
                                1470 non-null
                                                 int64
     JobLevel
 14
                                1470 non-null
                                                 int64
 15
                                1470 non-null
    JobRole
                                                 obiect
 16
     JobSatisfaction
                                1470 non-null
                                                 int64
 17
    MaritalStatus
                                1470 non-null
                                                 object
    MonthlyIncome
 18
                                1470 non-null
                                                 int64
 19
     MonthlyRate
                                1470 non-null
                                                 int64
 20
     NumCompaniesWorked
                                1470 non-null
                                                 int64
 21
     0ver18
                                1470 non-null
                                                 object
 22
     OverTime
                                1470 non-null
                                                 object
 23
    PercentSalaryHike
                                1470 non-null
                                                 int64
 24 PerformanceRating
                                1470 non-null
                                                 int64
 25
     RelationshipSatisfaction
                                1470 non-null
                                                 int64
 26
    StandardHours
                                1470 non-null
                                                 int64
 27
    StockOptionLevel
                                1470 non-null
                                                 int64
 28
    TotalWorkingYears
                                1470 non-null
                                                 int64
 29
    TrainingTimesLastYear
                                1470 non-null
                                                 int64
    WorkLifeBalance
 30
                                1470 non-null
                                                 int64
 31
    YearsAtCompany
                                1470 non-null
                                                 int64
    YearsInCurrentRole
32
                                1470 non-null
                                                 int64
 33
    YearsSinceLastPromotion
                                1470 non-null
                                                 int64
 34
    YearsWithCurrManager
                                1470 non-null
                                                 int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

dataset.d	describe()			
	Age	DailyRate	DistanceFromHome	Education
Employee(count 14	Count \ 170.000000	1470.000000	1470.000000	1470.000000
1470.0				
mean 1.0	36.923810	802.485714	9.192517	2.912925
std	9.135373	403.509100	8.106864	1.024165
0.0				
min 1.0	18.000000	102.000000	1.000000	1.000000
25%	30.000000	465.000000	2.000000	2.000000
1.0	26 000000	002 00000	7 000000	2 000000
50% 1.0	36.000000	802.000000	7.000000	3.000000
75%	43.000000	1157.000000	14.000000	4.000000
1.0	60.000000	1400 000000	29.000000	F 000000
max 1.0	00.00000	1499.000000	29.00000	5.000000
En JobInvolv	nployeeNumb /ement \	er Environme	ntSatisfaction H	HourlyRate
count	1470.0000	00	1470.000000 14	170.000000
1470.0000				
mean 2.729932	1024.8653	06	2.721769	65.891156
std	602.0243	35	1.093082	20.329428
0.711561				
min 1.000000	1.0000	90	1.000000	30.000000
25%	491.2500	00	2.000000	48.000000
2.000000	1020 5000	00	2 000000	66 000000
50% 3.000000	1020.5000	90	3.000000	66.000000
75%	1555.7500	00	4.000000	83.750000
3.000000	2069 0000	00	4 000000	100.000000
max 4.000000	2068.0000	00	4.000000	100.00000
				5
count 14	JobLevel 170.000000	Relatio	nshipSatisfaction 1470.000000	StandardHours \ 1470.0
mean	2.063946		2.712245	80.0
std	1.106940		1.081209	0.0
min 25%	1.000000 1.000000		1.000000 2.000000	80.0 80.0
50%	2.000000		3.000000	80.0
75%	3.000000		4.000000	80.0
max	5.000000		4.000000	80.0

```
TotalWorkingYears
       StockOptionLevel
                                             TrainingTimesLastYear
            1470.000000
                                1470.000000
                                                        1470.000000
count
mean
               0.793878
                                  11.279592
                                                           2.799320
               0.852077
                                   7.780782
                                                           1.289271
std
min
               0.000000
                                   0.000000
                                                           0.000000
25%
               0.000000
                                   6.000000
                                                           2.000000
50%
               1.000000
                                  10.000000
                                                           3.000000
75%
               1.000000
                                  15,000000
                                                           3.000000
               3.000000
                                  40.000000
                                                           6.000000
max
       WorkLifeBalance
                        YearsAtCompany
                                         YearsInCurrentRole \
           1470.000000
                            1470.000000
                                                 1470.000000
count
mean
              2.761224
                               7.008163
                                                    4.229252
              0.706476
                               6.126525
                                                    3,623137
std
min
              1.000000
                               0.000000
                                                    0.000000
25%
              2,000000
                               3,000000
                                                    2.000000
50%
              3.000000
                               5.000000
                                                    3.000000
75%
              3.000000
                               9.000000
                                                    7.000000
              4.000000
                              40.000000
                                                   18.000000
max
       YearsSinceLastPromotion YearsWithCurrManager
                   1470.000000
                                          1470.000000
count
                       2.187755
                                              4.123129
mean
                       3,222430
                                              3.568136
std
min
                       0.000000
                                             0.000000
25%
                       0.000000
                                             2.000000
50%
                       1.000000
                                             3.000000
75%
                       3.000000
                                             7.000000
                      15.000000
                                            17.000000
max
[8 rows x 26 columns]
dataset.describe().columns
Index(['i»¿Age', 'DailyRate', 'DistanceFromHome', 'Education',
'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobSatisfaction',
'MonthlyIncome',
       'MonthlyRate', 'NumCompaniesWorked', 'PercentSalaryHike',
       'PerformanceRating', 'RelationshipSatisfaction',
'StandardHours',
       'StockOptionLevel', 'TotalWorkingYears',
'TrainingTimesLastYear',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager'],
      dtype='object')
list(set(dataset.columns)-set(dataset.describe().columns))
```

```
['JobRole',
 '0ver18',
 'EducationField',
 'Attrition',
 'Department',
 'MaritalStatus'
 'BusinessTravel',
 'OverTime',
 'Gender']
dataset["Gender"]
0
        Female
1
          Male
2
          Male
3
        Female
4
          Male
          . . .
1465
          Male
          Male
1466
1467
          Male
1468
          Male
1469
          Male
Name: Gender, Length: 1470, dtype: object
dataset.Gender
0
        Female
1
          Male
2
          Male
3
        Female
4
          Male
          . . .
1465
          Male
1466
          Male
1467
          Male
1468
          Male
          Male
1469
Name: Gender, Length: 1470, dtype: object
dataset[["Gender","MonthlyIncome"]]
              MonthlyIncome
      Gender
0
      Female
                        5993
1
        Male
                        5130
2
        Male
                        2090
3
      Female
                        2909
4
        Male
                        3468
                         . . .
1465
        Male
                        2571
                        9991
1466
        Male
```

```
1467
        Male
                        6142
1468
                        5390
        Male
1469
        Male
                        4404
[1470 rows x 2 columns]
dataset.nunique()
ï»;Aqe
                                43
Attrition
                                 2
                                 3
BusinessTravel
DailyRate
                               886
Department
                                 3
                                29
DistanceFromHome
                                 5
Education
EducationField
                                 6
EmployeeCount
                                 1
EmployeeNumber
                             1470
EnvironmentSatisfaction
                                 4
                                 2
Gender
HourlyRate
                                71
JobInvolvement
                                 4
JobLevel
                                 5
                                 9
JobRole
JobSatisfaction
                                 4
MaritalStatus
                                 3
MonthlyIncome
                             1349
MonthlyRate
                             1427
NumCompaniesWorked
                               10
0ver18
                                 1
                                 2
OverTime
PercentSalaryHike
                                15
                                 2
PerformanceRating
RelationshipSatisfaction
                                 4
                                 1
StandardHours
                                4
StockOptionLevel
TotalWorkingYears
                                40
TrainingTimesLastYear
                                7
WorkLifeBalance
                                4
                                37
YearsAtCompany
YearsInCurrentRole
                                19
YearsSinceLastPromotion
                                16
YearsWithCurrManager
                               18
dtype: int64
dataset["Gender"].unique()
array(['Female', 'Male'], dtype=object)
dataset.isnull()
```

0 1 2 3 4 1465 1466 1467 1468 1469	Age False False False False False False False False	Attrition False	Business	ravel False	DailyRate False	Department False	
0 1 2 3 4 1465 1466 1467 1468 1469			Education False		tionField False	EmployeeCount False	
0 1 2 3 4 1465 1466 1467 1468 1469	Employe	eNumber False		onshipSa	atisfaction False False False False False False False False False	Fale Fale Fale Fale Fale Fale Fale Fale	se se se se se se se
0 1 2 3 4 1465 1466 1467 1468	Stock0p	tionLevel T False False False False False False False	otalWorki	False False False False False False False False	e e e e	TimesLastYear False False False False False False False	

1469		False		Fals	se	I	False
0 1 2 3 4	WorkLifeBa	False False False False	Fa Fa Fa	cany Nalse alse alse alse alse	⁄earsInCurre	ralse False False False False False	\
1465 1466 1467 1468 1469		False False False False False	Fa Fa	alse alse alse alse alse		False False False False False	
0 1 2 3 4	YearsSince	eLastPro	motion Yea False False False False False	arsWith	nCurrManager False False False False False		
1465 1466 1467 1468 1469			False False False False False		False False False False False		
_	rows x 35 na(dataset)		1				
0 1 2 3 4 	Age A-False False False False False False False	ttrition False False False False False		Travel False False False False False False	False False False False	Fa Fa Fa Fa Fa	ent \ lse lse lse lse
1466 1467 1468 1469	False False False False	False False False False		False False False False	False False False	Fa [°] Fa [°] Fa [°]	lse lse lse lse
0 1 2 3	DistanceF	romHome False False False False	Education False False False False	Educa	ationField False False False False	 	Count \ False False False False

4	False	False	False	False
1465 1466 1467 1468 1469	False False False False False	False False False False False	False False False False False	False False False False False
0 1 2 3 4 1465 1466 1467 1468 1469	False .	Relationship	Satisfaction St False False False False False False False False False	randardHours False
0 1 2 3 4	StockOptionLevel False False False False False	TotalWorkingYea Fal Fal Fal Fal	se se se se	esLastYear \ False False False False False False
1465 1466 1467 1468 1469	False False False False False	Fal Fal Fal Fal	se se se	False False False False False
0 1 2 3 4 1465 1466 1467 1468 1469	WorkLifeBalance False	False	Fal Fal Fal Fal hCurrManager	se se se se se se se
0		False False	False False	

2	False	False
3	False	False
4	False	False
1465	False	False
1466	False	False
1467	False	False
1468	False	False
1469	False	False

[1470 rows x 35 columns]

dataset.isnull().sum()

Age	0
Attrition	0
BusinessTravel	0
DailyRate	0
Department	0
DistanceFromHome	0
Education	0
EducationField	0
EmployeeCount	0
EmployeeNumber	0
EnvironmentSatisfaction	0
Gender	0
HourlyRate	0
JobInvolvement	0
JobLevel	0
JobRole	0
JobSatisfaction	0
MaritalStatus	0
MonthlyIncome	0
MonthlyRate	0
NumCompaniesWorked	0
0ver18	0
0verTime	0
PercentSalaryHike	0
PerformanceRating	0
RelationshipSatisfaction	0
StandardHours	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	0
YearsInCurrentRole	0
YearsSinceLastPromotion	0
YearsWithCurrManager	0
dtype: int64	

```
dataset.isnull().mean()
                             0.0
ï≫¿Age
Attrition
                             0.0
BusinessTravel
                             0.0
DailyRate
                             0.0
Department
                             0.0
DistanceFromHome
                             0.0
Education
                             0.0
EducationField
                             0.0
EmployeeCount
                             0.0
EmployeeNumber
                             0.0
EnvironmentSatisfaction
                             0.0
Gender
                             0.0
HourlyRate
                             0.0
JobInvolvement
                             0.0
JobLevel
                             0.0
JobRole
                             0.0
JobSatisfaction
                             0.0
MaritalStatus
                             0.0
MonthlyIncome
                             0.0
MonthlyRate
                             0.0
NumCompaniesWorked
                             0.0
0ver18
                             0.0
OverTime
                             0.0
PercentSalaryHike
                             0.0
PerformanceRating
                             0.0
RelationshipSatisfaction
                             0.0
StandardHours
                             0.0
StockOptionLevel
                             0.0
TotalWorkingYears
                             0.0
TrainingTimesLastYear
                             0.0
WorkLifeBalance
                             0.0
YearsAtCompany
                             0.0
YearsInCurrentRole
                             0.0
YearsSinceLastPromotion
                             0.0
YearsWithCurrManager
                             0.0
dtype: float64
dataset.isnull().mean()*100
                             0.0
ï»;Aqe
Attrition
                             0.0
                             0.0
BusinessTravel
DailyRate
                             0.0
                             0.0
Department
DistanceFromHome
                             0.0
Education
                             0.0
EducationField
                             0.0
EmployeeCount
                             0.0
```

```
EmployeeNumber
                             0.0
EnvironmentSatisfaction
                             0.0
Gender
                             0.0
HourlyRate
                             0.0
JobInvolvement
                             0.0
JobLevel
                             0.0
JobRole
                             0.0
JobSatisfaction
                             0.0
MaritalStatus
                             0.0
MonthlyIncome
                             0.0
MonthlyRate
                             0.0
NumCompaniesWorked
                             0.0
0ver18
                             0.0
OverTime
                             0.0
PercentSalaryHike
                             0.0
PerformanceRating
                             0.0
RelationshipSatisfaction
                             0.0
StandardHours
                             0.0
StockOptionLevel
                             0.0
TotalWorkingYears
                             0.0
TrainingTimesLastYear
                             0.0
WorkLifeBalance
                             0.0
YearsAtCompany
                             0.0
YearsInCurrentRole
                             0.0
YearsSinceLastPromotion
                             0.0
YearsWithCurrManager
                             0.0
dtype: float64
def missing value(dataframe):
round(dataframe.isnull().mean()*100,2).sort_values(ascending=False)
missing value(dataset)
ï≫¿Age
                             0.0
StandardHours
                             0.0
NumCompaniesWorked
                             0.0
0ver18
                             0.0
0verTime
                             0.0
PercentSalaryHike
                             0.0
PerformanceRating
                             0.0
RelationshipSatisfaction
                             0.0
StockOptionLevel
                             0.0
MonthlyIncome
                             0.0
TotalWorkingYears
                             0.0
TrainingTimesLastYear
                             0.0
WorkLifeBalance
                             0.0
YearsAtCompany
                             0.0
YearsInCurrentRole
                             0.0
YearsSinceLastPromotion
                             0.0
```

```
MonthlyRate
                             0.0
MaritalStatus
                             0.0
Attrition
                             0.0
EmployeeCount
                             0.0
BusinessTravel
                             0.0
DailyRate
                             0.0
Department
                             0.0
DistanceFromHome
                             0.0
                             0.0
Education
EducationField
                             0.0
EmployeeNumber
                             0.0
JobSatisfaction
                             0.0
EnvironmentSatisfaction
                             0.0
                             0.0
Gender
HourlyRate
                             0.0
JobInvolvement
                             0.0
                             0.0
JobLevel
JobRole
                             0.0
YearsWithCurrManager
                             0.0
dtype: float64
null=missing value(dataset)[missing value(dataset)>50]
null
Series([], dtype: float64)
null.index
Index([], dtype='object')
dataset.drop(columns=null.index,inplace=False)
                           BusinessTravel DailyRate
      i»¿Age Attrition
Department \
          41
0
                   Yes
                             Travel Rarely
                                                  1102
Sales
                                                        Research &
          49
                    No
                        Travel Frequently
                                                   279
Development
                   Yes
                             Travel Rarely
                                                  1373
                                                        Research &
          37
Development
          33
                    No
                         Travel Frequently
                                                  1392 Research &
Development
          27
                    No
                             Travel Rarely
                                                   591 Research &
Development
                                                   . . .
1465
                        Travel Frequently
                                                        Research &
          36
                    No
                                                   884
Development
                    No
                             Travel Rarely
          39
                                                   613
                                                        Research &
1466
Development
```

1467	27 No	Travel_R	Rarely	155 Res	earch &
Developmer 1468 Sales	49 No	Travel_Frequ	iently :	1023	
1469 Developmer	34 No	Travel_R	Rarely	628 Res	earch &
	tanceFromHome	Education Ed	lucationField	Employe	eCount \
0	1	2 L	ife Sciences.ife Sciences	Linptoye	1
2	2	2	Other ife Sciences		1
4	2	1	Medical		1
1465 1466	23 6	2 1	Medical Medical		1 1
1467 1468	4		ife Sciences. Medical		1
1469	8	3	Medical		1
Empl	loyeeNumber 1	Relations	shipSatisfact:	ion Standa 1	ardHours \ 80
1 2	2 4			4 2	80 80
3 4	5 7			3 4	80 80
1465	2061			 3	80
1466 1467	2062 2064			1 2	80 80
1468 1469	2065 2068			4 1	80 80
	ckOptionLevel	TotalWorking		ingTimesLa	_
0	0		8 10 7		0 3
2 3 4	0		7 8		3 3 3 3
	1		6		
1465 1466 1467	1 1 1		17 9 6		3 5
1467 1468 1469	9		17 6		0 3 3
	-	YearsAtCompany		entRole `	_
0 1	1 3	6 10	j	4 7	,
2	3	0		0	

3 4	3 3	8 2	7 2
1465	3	5	2
1466	3	7	7
1467	3	6	2
1468	2	9	6
1469	4	4	3

	YearsSinceLastPromotion	YearsWithCurrManager
0	0	5
1	1	7
2	0	0
3	3	Θ
4	2	2
1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

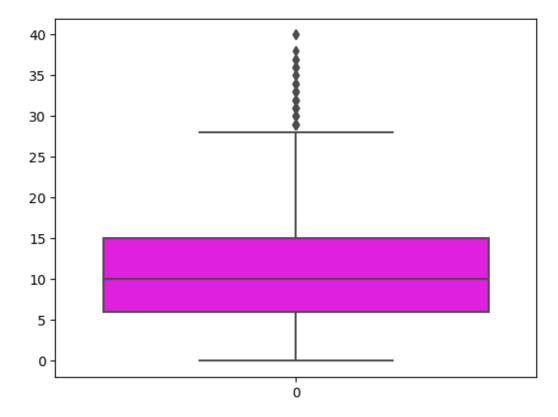
[1470 rows x 35 columns]

dataset[["i»¿Age", "Attrition", "BusinessTravel", "DailyRate", "Department
", "DistanceFromHome", "Education", "EducationField", "EmployeeCount", "Emp
loyeeNumber", "RelationshipSatisfaction", "StandardHours", "StockOptionLe
vel", "TotalWorkingYears", "TrainingTimesLastYear", "WorkLifeBalance", "Ye
arsAtCompany",

"YearsInCurrentRole", "YearsSinceLastPromotion", "YearsWithCurrManager"]
].isnull().sum()

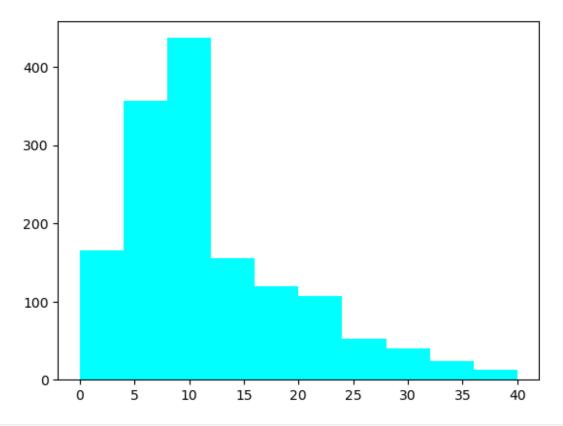
Age	0
Attrition	0
BusinessTravel	0
DailyRate	0
Department	0
DistanceFromHome	0
Education	0
EducationField	0
EmployeeCount	0
EmployeeNumber	0
RelationshipSatisfaction	0
StandardHours	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	0

```
YearsInCurrentRole 0
YearsSinceLastPromotion 0
YearsWithCurrManager 0
dtype: int64
sns.boxplot(dataset["TotalWorkingYears"],color="magenta")
<Axes: >
```



```
plt.hist(dataset["TotalWorkingYears"],color="cyan")

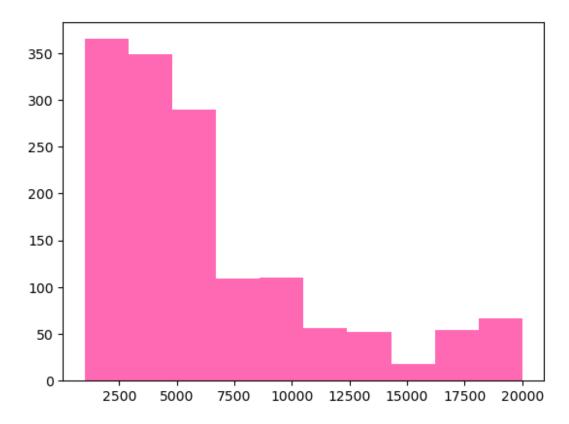
(array([165., 357., 437., 155., 119., 107., 53., 40., 24., 13.]),
  array([ 0., 4., 8., 12., 16., 20., 24., 28., 32., 36., 40.]),
  <BarContainer object of 10 artists>)
```



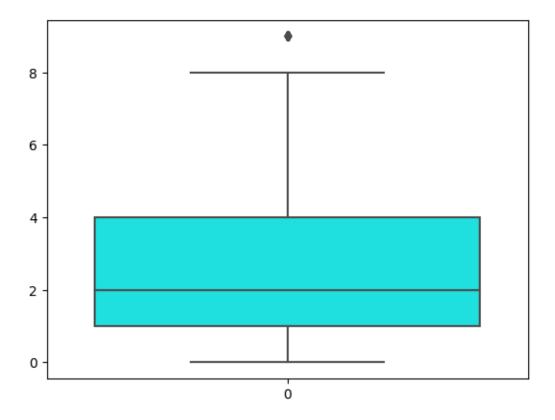
```
dataset["TotalWorkingYears"].describe()
         1470.000000
count
mean
           11.279592
            7.780782
std
            0.000000
min
25%
            6.000000
50%
           10.000000
           15.000000
75%
           40.000000
max
Name: TotalWorkingYears, dtype: float64
q1=np.quantile(dataset["TotalWorkingYears"].dropna(),0.25)
q1
6.0
q2=np.quantile(dataset["TotalWorkingYears"].dropna(),0.50)
q2
10.0
q3=np.quantile(dataset["TotalWorkingYears"].dropna(),0.75)
q3
15.0
```

```
IQR=q3-q1
IQR
9.0
upper_fence=q3+1.5*(IQR)
lower fence=1-1.5*(IQR)
1=[]
for i in dataset["TotalWorkingYears"]:
    if i>upper fence or i<lower fence:</pre>
         l.append(i)
print(l)
[31, 29, 37, 38, 30, 40, 36, 34, 32, 33, 37, 30, 36, 31, 33, 32, 37, 31, 32, 32, 30, 34, 30, 40, 29, 35, 31, 33, 31, 29, 32, 30, 33, 30,
29, 31, 32, 33, 36, 34, 31, 36, 33, 31, 29, 33, 29, 32, 31, 35, 29,
32, 34, 36, 32, 30, 36, 29, 34, 37, 29, 29, 35]
missing values = dataset.isnull().sum()
print("\nMissing Values Per Column:")
print(missing values)
Missing Values Per Column:
ï≫¿Age
                               0
                               0
Attrition
                               0
BusinessTravel
                               0
DailyRate
                               0
Department
                               0
DistanceFromHome
                               0
Education
                               0
EducationField
                               0
EmployeeCount
                               0
EmployeeNumber
                               0
EnvironmentSatisfaction
                               0
Gender
                               0
HourlyRate
                               0
JobInvolvement
                               0
JobLevel
                               0
JobRole
JobSatisfaction
                               0
                               0
MaritalStatus
MonthlyIncome
                               0
                               0
MonthlyRate
                               0
NumCompaniesWorked
0ver18
                               0
                               0
OverTime
PercentSalaryHike
                               0
PerformanceRating
                               0
RelationshipSatisfaction
                               0
```

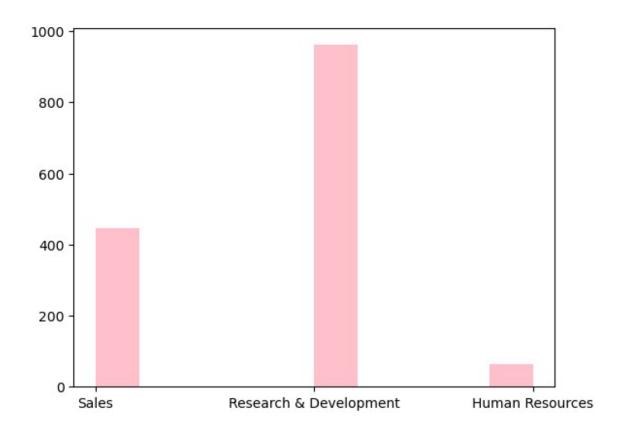
```
StandardHours
                            0
StockOptionLevel
                            0
TotalWorkingYears
                            0
                            0
TrainingTimesLastYear
WorkLifeBalance
                            0
                            0
YearsAtCompany
YearsInCurrentRole
                            0
YearsSinceLastPromotion
                            0
YearsWithCurrManager
                            0
dtype: int64
dataset['MonthlyIncome'].fillna(dataset['MonthlyIncome'].median(),
inplace=True)
dataset['NumCompaniesWorked'].fillna(dataset['NumCompaniesWorked'].mod
e(), inplace=True)
dataset['MonthlyIncome']
0
        5993
1
        5130
2
        2090
3
        2909
        3468
        . . .
1465
        2571
1466
        9991
1467
        6142
1468
        5390
1469
        4404
Name: MonthlyIncome, Length: 1470, dtype: int64
plt.hist(dataset["MonthlyIncome"],color="hotpink")
(array([365., 349., 290., 109., 110., 56., 52., 18., 54., 67.]),
array([ 1009., 2908., 4807., 6706., 8605., 10504., 12403.,
14302.,
        16201., 18100., 19999.]),
 <BarContainer object of 10 artists>)
```



```
dataset['NumCompaniesWorked']
0
1
2
        8
        1
        6
3
        1
        9
1465
        4
1466
        4
        1
1467
        2
1468
1469
Name: NumCompaniesWorked, Length: 1470, dtype: int64
sns.boxplot(dataset["NumCompaniesWorked"],color="cyan")
<Axes: >
```



```
dataset.dropna(subset=['Department'], inplace=True)
dataset['Department']
0
                      Sales
1
      Research & Development
2
       Research & Development
3
       Research & Development
      Research & Development
1465
      Research & Development
      Research & Development
1466
1467
      Research & Development
1468
                      Sales
      Research & Development
1469
Name: Department, Length: 1470, dtype: object
plt.hist(dataset["Department"],color="pink")
<BarContainer object of 10 artists>)
```



Drop columns that are irrelevant

```
# List of columns to drop
columns_to_drop = ['EmployeeCount', 'Over18', 'StandardHours',
'EmployeeID']
# Drop columns if they exist in the dataset
dataset.drop(columns=[col for col in columns to drop if col in
dataset.columns], axis=1, inplace=True)
# Display remaining columns
print("\nColumns after dropping irrelevant ones:")
print(dataset.columns)
Columns after dropping irrelevant ones:
Index(['i»¿Age', 'Attrition', 'BusinessTravel', 'DailyRate',
'Department',
       'DistanceFromHome', 'Education', 'EducationField',
'EmployeeNumber',
       'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
'JobInvolvement',
       'JobLevel', 'JobRole', 'JobSatisfaction', 'MaritalStatus',
       'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
```

```
'OverTime',
       PercentSalaryHike', 'PerformanceRating',
'RelationshipSatisfaction',
       'StockOptionLevel', 'TotalWorkingYears',
'TrainingTimesLastYear',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager'],
      dtype='object')
# Map categorical variables to numeric
# dataset['Attrition'] = dataset['Attrition'].map({'Yes': 1, 'No': 0})
# dataset['OverTime'] = dataset['OverTime'].map({'Yes': 1, 'No': 0})
# dataset['Department'] = dataset['Department'].map({ 'Sales': 1,
'Research & Development': 2, 'Human Resources': 3})
print(dataset.head())
   i»¿Age Attrition
                        BusinessTravel DailyRate
Department \
       41
                Yes
                       Travel Rarely
                                             1102
Sales
                 No Travel Frequently
                                              279
                                                   Research &
1
       49
Development
                         Travel Rarely
                                             1373
                                                   Research &
       37
                Yes
Development
       33
                 No Travel Frequently
                                             1392
                                                   Research &
Development
                         Travel Rarely
                                              591 Research &
       27
                 No
Development
   DistanceFromHome
                     Education EducationField EmployeeNumber \
0
                             2 Life Sciences
                  1
                                                             1
1
                  8
                             1 Life Sciences
                                                             2
2
                  2
                                                             4
                                        0ther
3
                  3
                                                             5
                                Life Sciences
                  2
                             1
                                      Medical
   EnvironmentSatisfaction ... PerformanceRating
RelationshipSatisfaction \
                                                3
0
1
1
                         3
                                                4
4
2
                                                3
2
3
                                                3
3
4
                                                3
                         1
4
```

```
StockOptionLevel TotalWorkingYears TrainingTimesLastYear
WorkLifeBalance \
                                   8
                                                        0
1
1
                                  10
                                                        3
3
2
                                                        3
3
3
                                                        3
3
4
                                                        3
                                   6
3
                YearsInCurrentRole YearsSinceLastPromotion \
  YearsAtCompany
0
1
             10
                                 7
                                                        1
2
              0
                                 0
                                                        0
              8
                                 7
                                                        3
3
4
              2
                                                        2
  YearsWithCurrManager
0
                    5
                    7
1
2
                    0
3
                    0
4
[5 rows x 32 columns]
dataset['YearsAtCompanyBins'] = pd.cut(dataset['YearsAtCompany'],
bins=[0, 1, 5, 10, 20, 40],
                                   labels=['<1', '1-5', '5-10', '10-
20', '>20'1)
print("Data cleaning and preprocessing complete.")
Data cleaning and preprocessing complete.
dataset['AgeGroup']
       40-50
1
       40-50
2
       30-40
3
       30-40
       18-30
       . . .
       30-40
1465
1466
       30-40
       18-30
1467
```

```
1468
        40-50
1469
        30-40
Name: AgeGroup, Length: 1470, dtype: category
Categories (4, object): ['18-30' < '30-40' < '40-50' < '50-60']
dataset['YearsAtCompanyBins']
        5-10
1
        5 - 10
2
        NaN
3
        5 - 10
        1-5
1465
        1-5
1466
        5 - 10
1467
        5-10
1468
        5 - 10
1469
        1-5
Name: YearsAtCompanyBins, Length: 1470, dtype: category
Categories (5, object): ['<1' < '1-5' < '5-10' < '10-20' < '>20']
```

Re-importing all liberaries for perform analysis

```
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
import warnings
warnings.filterwarnings("ignore")
```

introduction of project

ATTRITION --> Attrition of employees from organization for any reason or in simple words we can say that employees who wants to leave the company

ATTRITION "YES"-->Employee wants to leave the company.

ATTRITON "NO"-->Eemployee don't want to leave the company

domain analysis

1.AGE

Age of an employee

2.BUSSINESS TRAVEL

That employee is travelling or not for bussines purpose

3.DEPARTMENT

In which department of company employee is working

4.DISTANCE FROM HOME

that how much distance is there of an employee from his/her house to office

5.HOURLY/DAILY/MONTHLY RATING

Rating given to employee on daily, hourly and montly basis

6.EDUCATION FIELD

from which education background employee belongs to

7.JOB AND ENVIRONMENT SATISFACTION

It is that employee is satisfiedd with their job as well as (given in 1-5 rating)

8.JOB INVOLVEMENT

How much employee is involvment in his work(given in rating 1-5)

9.JOB LEVEL

Level of job, higher rating==higher job level

10.PERFORMANCE RATING

How employee is performing according to his job(given rating 1-5)

11.MONTHLY INCOME

Monthly salary of an employee in doller

12.PERCEMTAGE SALARY HIKE

How many percent salary in incresing anually

<pre>dataset.describe(include='0')</pre>					
Gender	Attrition \	BusinessTravel	Department	EducationField	
count	1470	1470	1470	1470	
1470					
unique	2	3	3	6	
2					
top	No	Travel_Rarely	Research & Development	Life Sciences	
Male					
freq	1233	1043	961	606	
882					

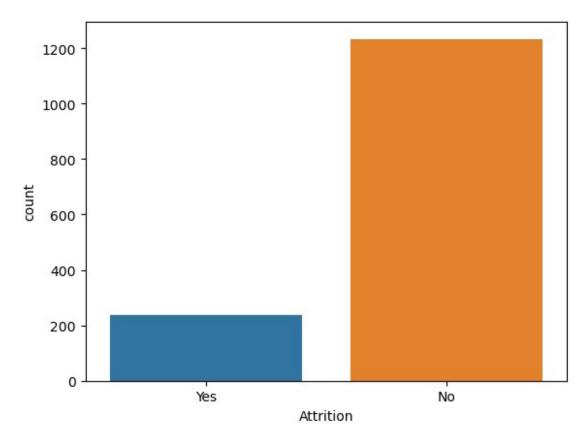
freq	Sales Execu	Role Marita 1470 9 Itive N 326	lStatus 1470 3 Married 673	OverTime 1470 2 No 1054		
dataset.	describe()					
count 1 mean std min 25% 50% 75% max	">¿Age 470.000000 36.923810 9.135373 18.000000 30.000000 43.000000 60.000000	DailyRate 1470.000000 802.48571 403.509100 102.000000 465.0000000 802.000000 1157.0000000	9 4 9 9 9 9	nceFromHor 1470.00000 9.1925 8.10680 1.00000 2.00000 7.00000 14.00000	30 1470.000000 17 2.912925 64 1.024165 30 1.000000 30 2.000000 30 3.000000 4.000000	
	mployeeNumb	er Environ	mentSati	isfaction	HourlyRate	
JobInvol count	vement \ 1470.0000	000	147	70.000000	1470.000000	
1470.000 mean	000 1024.8653	k06		2.721769	65.891156	
2.729932						
std 0.711561	602.0243	335		1.093082	20.329428	
min	1.0000	000		1.000000	30.000000	
1.000000 25%	491.2500	000		2.000000	48.000000	
2.000000 50%	1020.5000	000		3.000000	66.000000	
3.000000						
75% 3.000000	1555.7500	טטט		4.000000	83.750000	
max 4 000000	2068.0000	000		4.000000	100.000000	
4.000000	JobLevel	JobSatisfa	ction .	Perfo	rmanceRating \	
count 1 mean	.470.000000 2.063946	1470.00	00571		1470.000000 3.153741	
std	1.106940				0.360824	
min 25%	1.000000 1.000000		2000		3.000000 3.000000	
50%	2.000000	3.00	90000 .		3.000000	
75% max	3.000000 5.000000		. 00000 90000 .		3.000000 4.000000	
				Ontion		
RelationshipSatisfaction StockOptionLevel TotalWorkingYears \						
count		1470.000000	9	1470.00000	90 1470.0	000000

mean	2.7122	45 0.79	93878	11.279592
std	1.0812	09 0.85	52077	7.780782
min	1.0000	00 0.00	00000	0.000000
25%	2.0000	00 0.00	00000	6.000000
50%	3.0000	00 1.00	00000	10.000000
75%	4.0000	00 1.00	00000	15.000000
max	4.0000	00 3.00	00000	40.000000
count mean std min 25% 50% 75% max	lanager 1470.000000 4.229252 3.623137 0.000000 2.000000 3.000000 7.000000 18.000000	2.1 3.2 0.6 0.6 1.6	1470.00 4 7.00 5 6.12 9 9.00 9.00 9.00 9.00 9.00	

Analysis on categorical data

TARGET COLUMN==ATTRITION

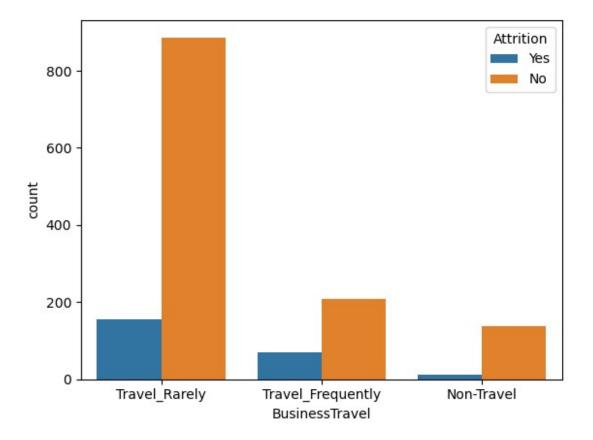
sns.countplot(x=dataset.Attrition)
plt.show()



- -->Data of attrition says that it has more number of "No" values and less number of "Yes"
- -->It can be seen that there is big diffrence in counts of the values so we can says that it is IMBALANCE DATA

1 IMPACT OF BUSSINES TRAVEL ON ATTRITION

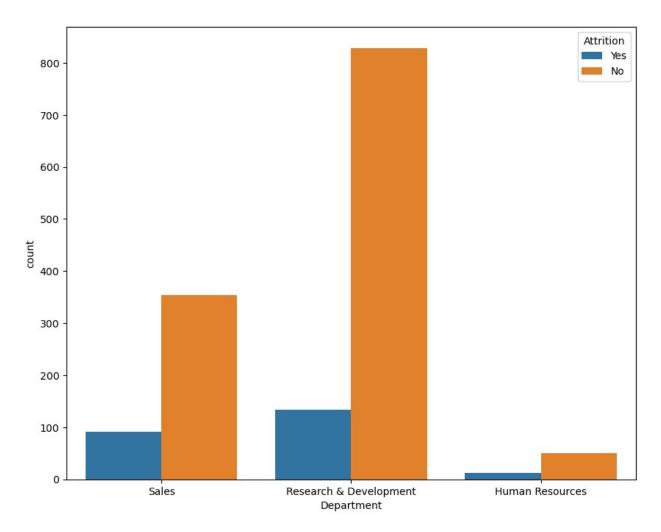
sns.countplot(hue=dataset.Attrition,x=dataset.BusinessTravel)
plt.show()



- -->The graph tell us that company has more count or more no of employees travel rearly
- -->There are more employees travel rearly and not satisfied with their jobs
- -->Non-traveller have least count as well as least attrition

2 IMPACT OF DEPARTMENT ON ATTRITION

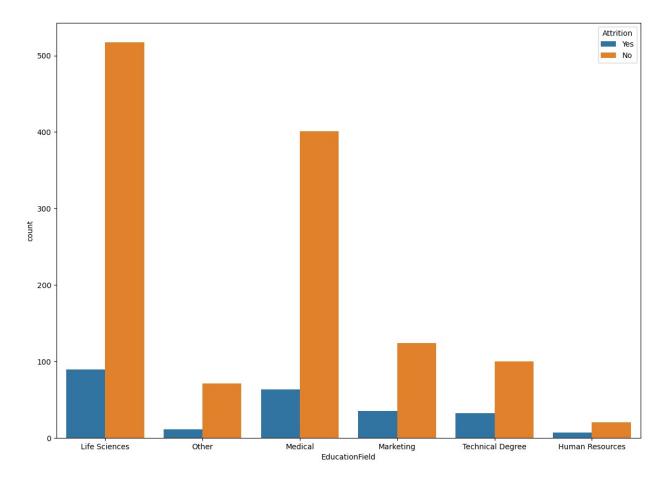
```
plt.figure(figsize=(10,8))
sns.countplot(hue=dataset.Attrition,x=dataset.Department)
plt.show()
```



- -->There are 3 no. of department are there = 1.Sales,2.Research & Development,3.Human Resources
- -->"Research & Development" department have more number of department of Attrition(150 employees) as compare to other two department
- -->"HR Department" have least Attrition with 5 to 10 employees

3 IMPACT OF EDUCATION ON ATTRITION

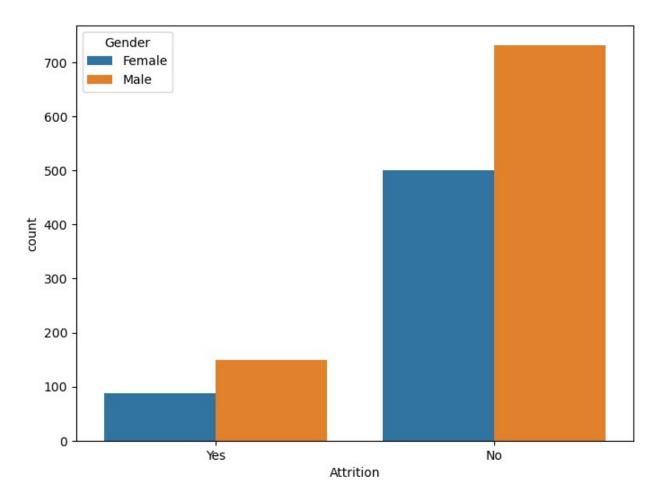
```
plt.figure(figsize=(14,10))
sns.countplot(hue=dataset.Attrition,x=dataset.EducationField)
plt.show()
```



- -->First and foremost things is that Employees are from "life science" and "Medical" background are more as comapre to other education field
- --> Nearly 100 number of employee are there who are from Lifes Sciences education background will leave the company and follow the Medical education employees
- -->As we conclude from analysis of Department and Attrition .here also HR education background employees have least Attrition

4. IMPACT GENDER ON ATTRITION

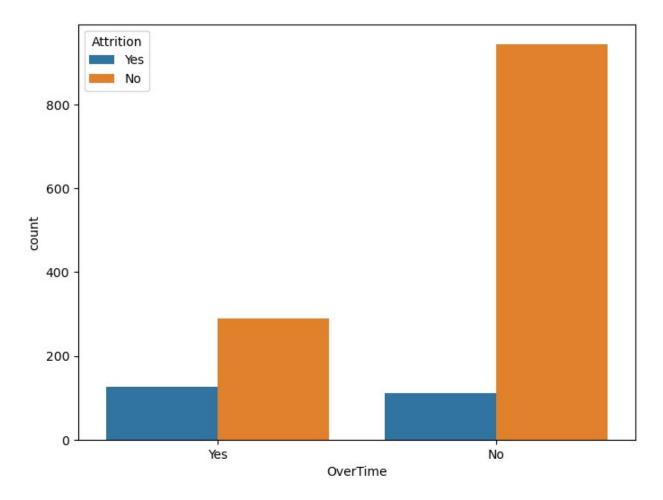
```
plt.figure(figsize=(8,6))
sns.countplot(x=dataset.Attrition, hue=dataset.Gender)
plt.show()
```



- --> Male employees are more as compared to Female employees
- --> Males are more likely to quit rather than Females

5 OVERTIME AND ATTRITION

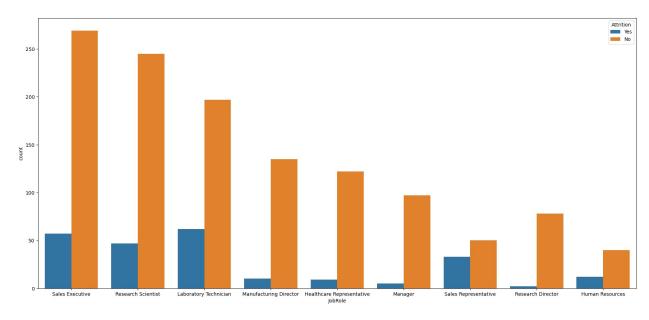
```
plt.figure(figsize=(8,6))
sns.countplot(hue=dataset.Attrition,x=dataset.OverTime)
plt.show()
```



- --> As for "Attrition yes" there is the minor diffrents between Employees who are doing OverTime and who are not doing OverTime
- --> So we can say overtime feature is not mutch effecting Attrition
- --> But we can conclude that most of employees are not doing overtime

6 IMPACT OF JOBOLE ON ATTRITION

```
plt.figure(figsize=(22,10),facecolor='white')
sns.countplot(x=dataset.JobRole,hue=dataset.Attrition)
plt.show()
```



- --> There is less no of Research director who leaves the company
- -->Laboratory technician, sales executive and reasearch scientist are the top 3job roles in which employees have their "Attrition yes"
- -->It can also seen that more number of employees in sales executive job role

ANALYSIS ON CONTINUOS DATA WITH RESPECT TO TARGET COLUMN

```
numerical_col=[]
for column in dataset.columns:
    if dataset[column].dtype=="int64" and
len(dataset[column].unique())>=10:
        numerical_col.append(column)
numerical col
[ˈi»¿Age',
 'DailyRate',
 'DistanceFromHome',
 'EmployeeNumber',
 'HourlyRate',
 'MonthlyIncome',
 'MonthlyRate',
 'NumCompaniesWorked',
 'PercentSalaryHike',
 'TotalWorkingYears',
 'YearsAtCompany',
 'YearsInCurrentRole',
```

```
'YearsSinceLastPromotion',
'YearsWithCurrManager']
```

graphical repesentation of continuos data

```
dataset2=dataset[
    ['Age',
    'DailyRate',
    'DistanceFromHome',
    'EmployeeNumber',
    'HourlyRate',
    'MonthlyIncome',
    'MonthlyRate',
    'NumCompaniesWorked',
    'PercentSalaryHike',
    'TotalWorkingYears',
    'YearsAtCompany',
    'YearsInCurrentRole',
    'YearsSinceLastPromotion',
    'YearsWithCurrManager']]
```

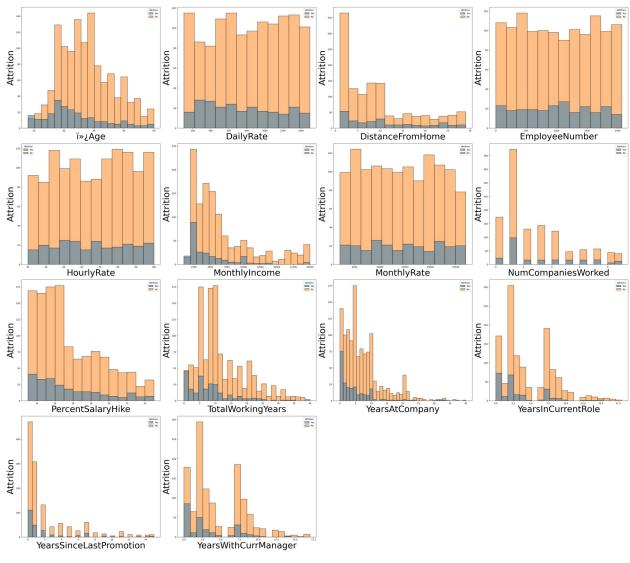
Another method of visualization

```
plt.figure(figsize=(40,35),facecolor='white')
plotnumber = 1

for column in dataset2:
    if plotnumber<=16:
        ax=plt.subplot(4,4,plotnumber)

sns.histplot(x=dataset2[column].dropna(axis=0),hue=dataset.Attrition)

    plt.xlabel(column,fontsize=40)
    plt.ylabel('Attrition',fontsize=40)
    plotnumber+=1
plt.tight_layout()</pre>
```



```
import matplotlib.pyplot as plt
import seaborn as sns

def plot_in_pairs(dataset2, hue_column):
    columns = dataset2.columns
    num_columns = len(columns)
    pair_count = 2

for i in range(0, num_columns, pair_count):
    plt.figure(figsize=(20, 10), facecolor='white')

    for j in range(pair_count):
        if i + j < num_columns:
            ax = plt.subplot(1, pair_count, j + 1)

        sns.histplot(
            x=dataset2[columns[i + j]].dropna(),</pre>
```

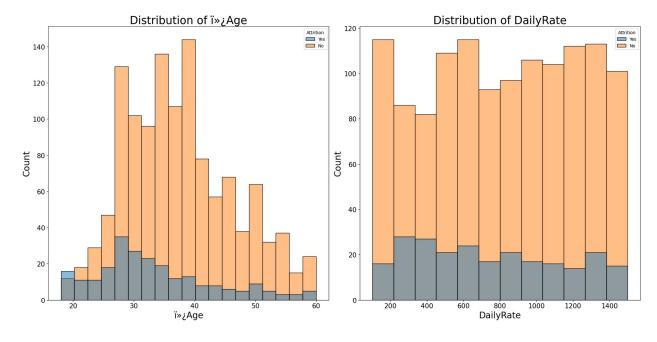
```
hue=hue_column,
kde=False,
ax=ax
)

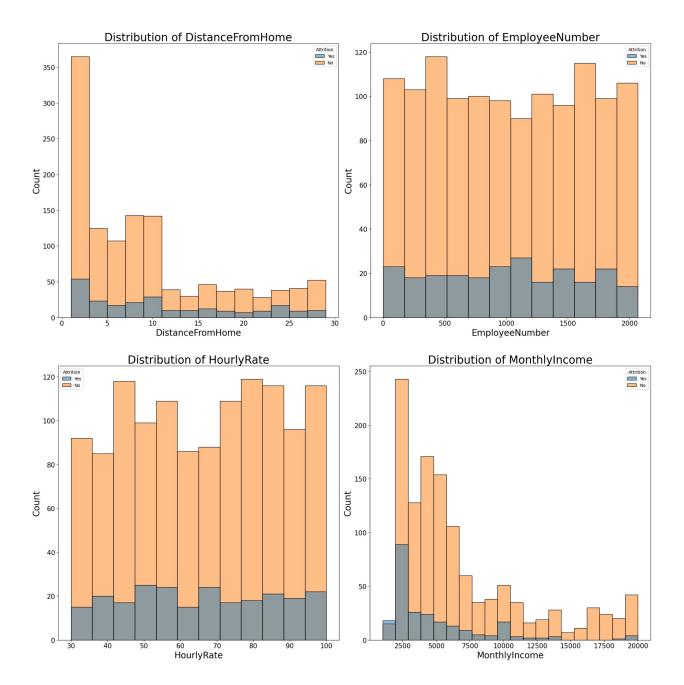
plt.xlabel(columns[i + j], fontsize=20)
plt.ylabel('Count', fontsize=20)
ax.set_title(f'Distribution of {columns[i + j]}',
fontsize=25)

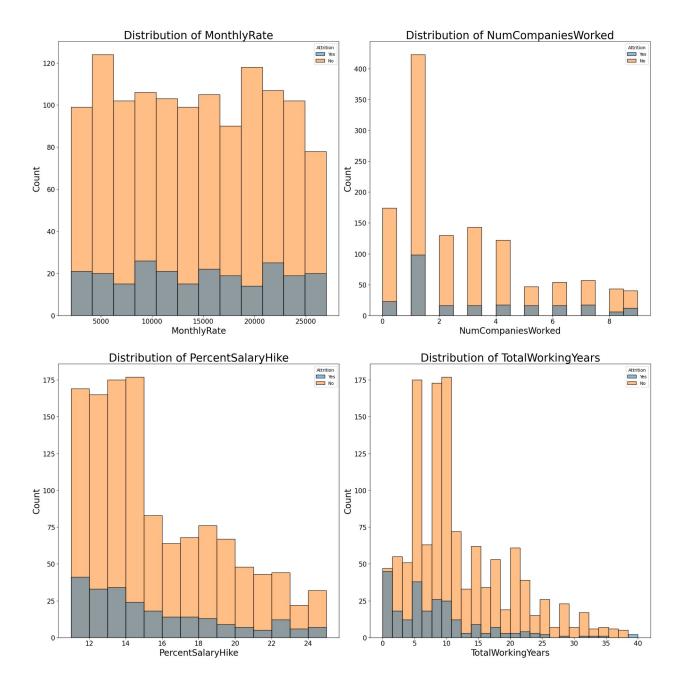
ax.tick_params(axis='both', which='major',
labelsize=15)

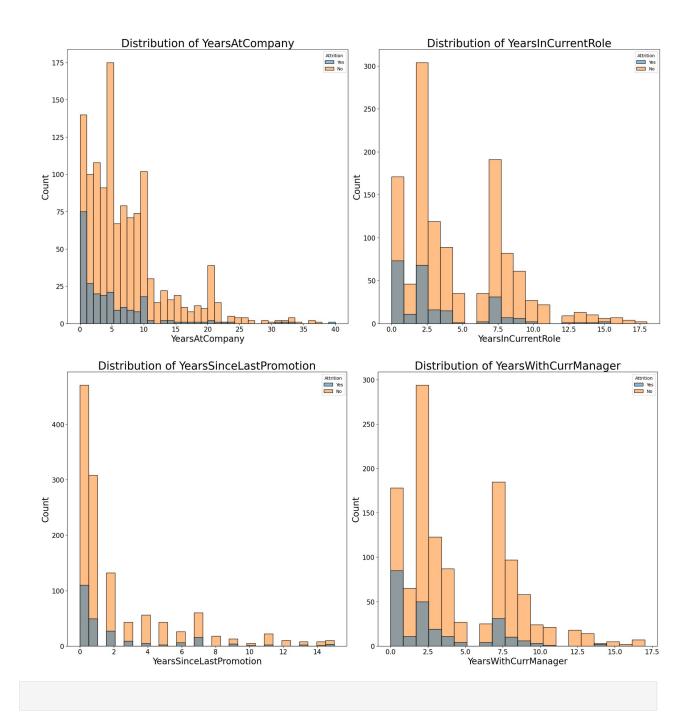
plt.tight_layout()
plt.show()

plot_in_pairs(dataset2, hue_column=dataset.Attrition)
```



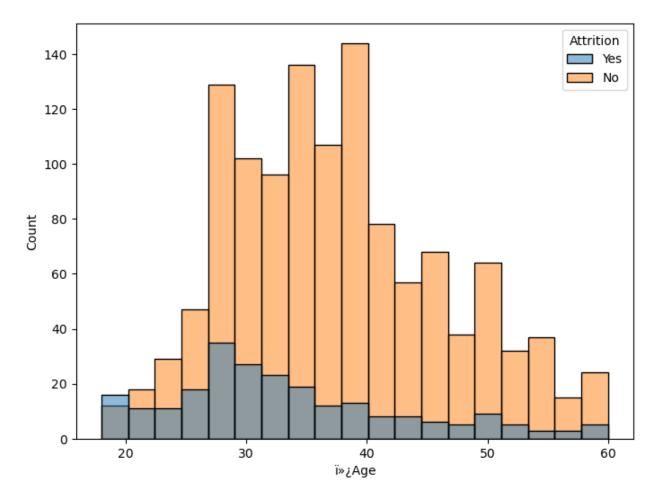






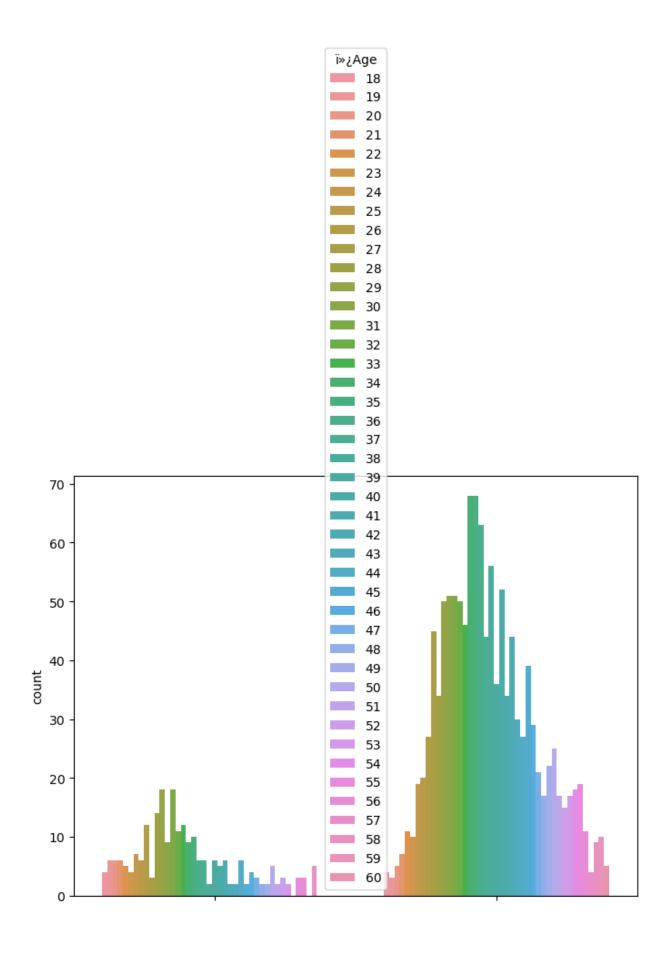
1 IMPACT OF AGE ON ATTRITION

```
plt.figure(figsize=(8,6),facecolor='white')
sns.histplot(x=dataset['Age'],hue=dataset.Attrition)
plt.show()
```



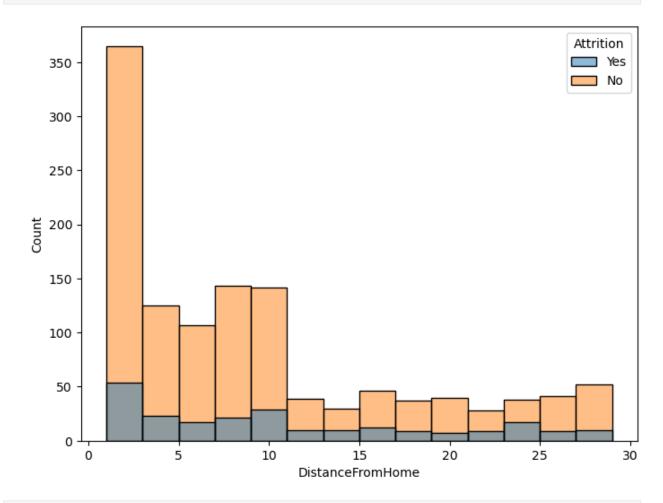
- -->Employees in age 25 to 35 are more likely to leave their job
- -->After the age 40. the distribution tell us that "Higher the age lesser will we Attrition"

```
plt.figure(figsize=(8,6),facecolor='white')
sns.countplot(hue=dataset['Age'],x=dataset.Attrition)
plt.show()
```

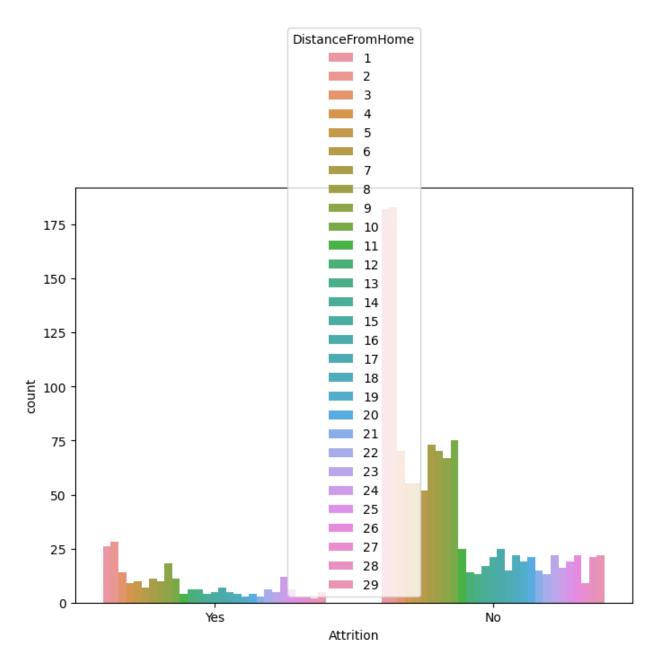


2 DISTANCE FROM HOME AND ATTRITION

```
plt.figure(figsize=(8,6), facecolor='white')
sns.histplot(x=dataset.DistanceFromHome, hue=dataset.Attrition)
plt.show()
```



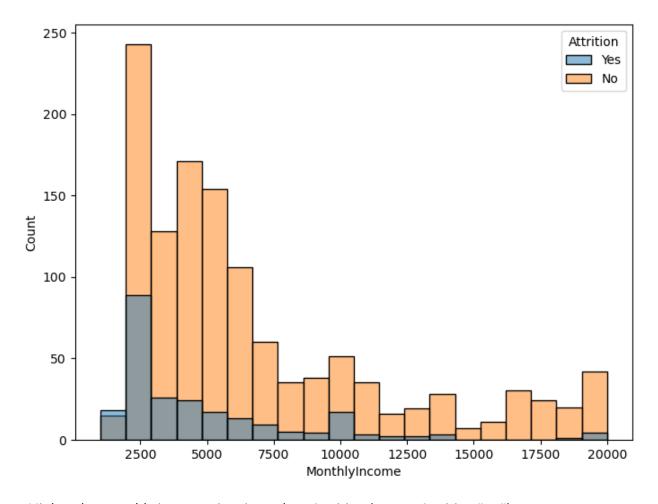
```
plt.figure(figsize=(8,6),facecolor='white')
sns.countplot(hue=dataset.DistanceFromHome,x=dataset.Attrition)
plt.show()
```



- --> Employees who has distance range "0-100" km are more likely to leave their job.
- --> we can also conclude that lesser the distance more number of employees working.

3 how monthly income give trends with respect to attrition

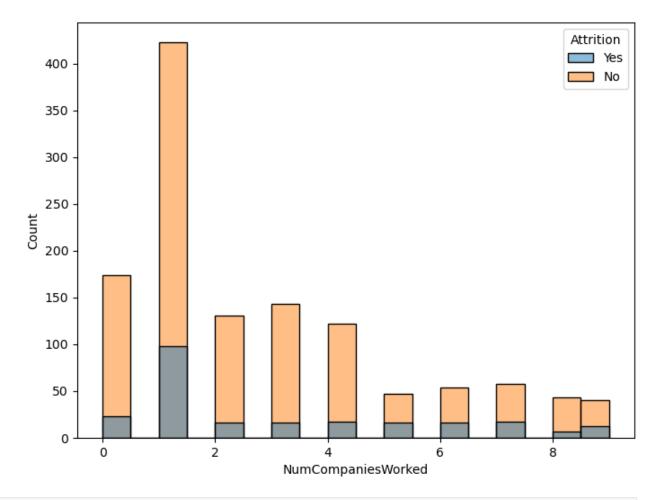
```
plt.figure(figsize=(8,6), facecolor='white')
sns.histplot(x=dataset.MonthlyIncome, hue=dataset.Attrition)
plt.show()
```



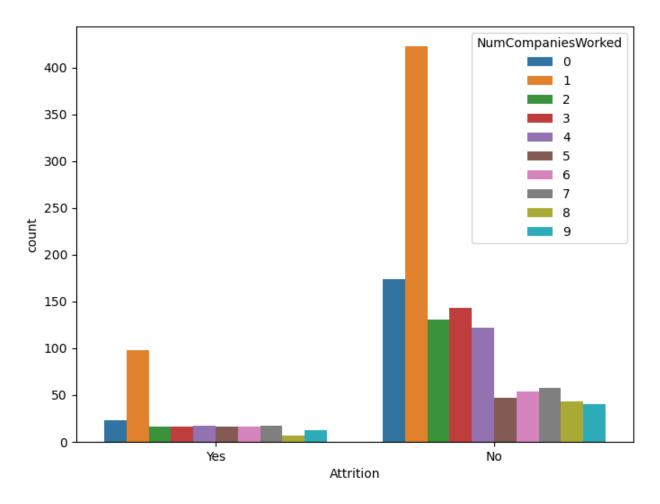
- --> Higher the monthly income give rise to less Attrition (means Attrition "no")
- --> Employees who have their income 2500 aprox are more likely to quit their job because 2500 is the least range of income

4.IMPACT OF NO. OF COMPANIES WORKED

```
plt.figure(figsize=(8,6),facecolor='white')
sns.histplot(x=dataset.NumCompaniesWorked,hue=dataset.Attrition)
plt.show()
```



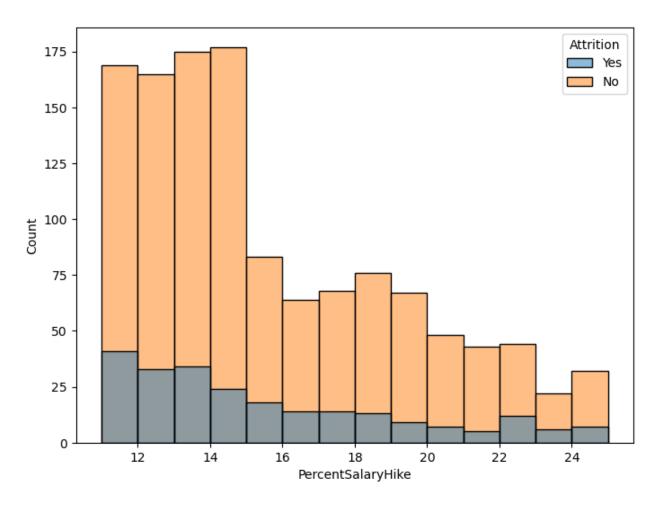
plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(hue=dataset.NumCompaniesWorked, x=dataset.Attrition)
plt.show()



--> Olny the employees(no. of employees=100) who work with one company before have more (Attrition yes) rest have similar data

5 HOW SALARY HIKE IMPACT THE ATTRITION

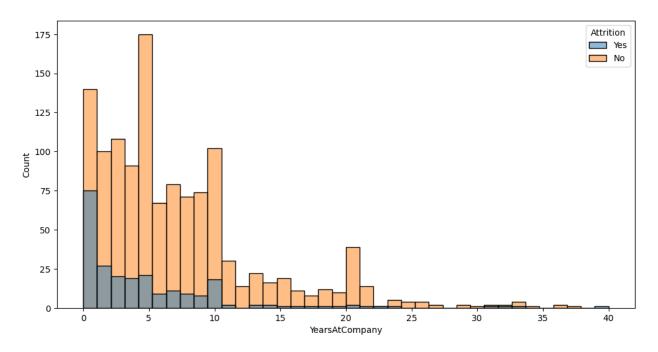
```
plt.figure(figsize=(8,6),facecolor='white')
sns.histplot(x=dataset.PercentSalaryHike,hue=dataset.Attrition)
plt.show()
```



--> Higher the salary percentage hike lesser the Attrition ("no")

6 YEAR AT THE COMPANY

```
plt.figure(figsize=(12,6),facecolor='white')
sns.histplot(x=dataset. YearsAtCompany,hue=dataset.Attrition)
plt.show()
```



- --> Fresher have higher Attrition "yes" that is of 75 no. of workers are more than half of feshers
- --> Apart from employees who ranges fron 1 to 10 years working on company are less likely to quit their job.

ANALYSIS OF DISCRETE DATA WITH RESPWCT TO TARGET COLUMN

```
discrete col=[]
for column in dataset.columns:
    if dataset[column].dtype=="int64" and
len(dataset[column].unique())>=10:
        discrete col.append(column)
discrete_col
,'Age',
 'DailyRate',
 'DistanceFromHome',
 'EmployeeNumber',
 'HourlyRate',
 'MonthlyIncome',
 'MonthlyRate',
 'NumCompaniesWorked',
 'PercentSalaryHike',
 'TotalWorkingYears',
 'YearsAtCompany',
 'YearsInCurrentRole',
```

```
'YearsSinceLastPromotion',
 'YearsWithCurrManager']
dataset.describe()
             Age
                       DailyRate
                                   DistanceFromHome
                                                         Education
       1470.000000
                     1470.000000
                                        1470.000000
                                                      1470.000000
count
         36.923810
                      802.485714
                                            9.192517
                                                          2.912925
mean
std
          9.135373
                      403.509100
                                            8.106864
                                                          1.024165
min
         18.000000
                      102.000000
                                            1.000000
                                                          1.000000
                                                          2.000000
25%
         30,000000
                      465.000000
                                            2.000000
         36,000000
                      802.000000
50%
                                            7.000000
                                                          3.000000
75%
         43.000000
                     1157.000000
                                           14.000000
                                                          4.000000
         60,000000
                     1499.000000
                                           29,000000
                                                          5.000000
max
       EmployeeNumber
                        EnvironmentSatisfaction
                                                    HourlyRate
JobInvolvement \
          1470.000000
                                                   1470.000000
count
                                     1470.000000
1470.000000
          1024.865306
                                        2.721769
                                                     65.891156
mean
2.729932
std
           602.024335
                                        1.093082
                                                     20.329428
0.711561
              1.000000
                                        1.000000
                                                     30.000000
min
1.000000
25%
           491.250000
                                        2.000000
                                                     48.000000
2.000000
50%
          1020.500000
                                        3.000000
                                                     66.000000
3.000000
75%
          1555.750000
                                        4.000000
                                                     83.750000
3.000000
          2068.000000
                                        4.000000
                                                    100.000000
max
4.000000
          JobLevel
                     JobSatisfaction
                                             PerformanceRating
       1470.000000
                         1470.000000
                                                   1470.000000
count
          2.063946
                             2.728571
                                                      3.153741
mean
                             1.102846
std
          1.106940
                                                      0.360824
min
          1.000000
                             1.000000
                                                      3.000000
25%
          1.000000
                             2.000000
                                                      3.000000
50%
          2.000000
                             3.000000
                                                      3.000000
                                        . . .
75%
          3,000000
                             4,000000
                                                      3,000000
          5.000000
                             4,000000
                                                      4.000000
max
       RelationshipSatisfaction StockOptionLevel
TotalWorkingYears
count
                     1470.000000
                                        1470.000000
                                                             1470.000000
                        2.712245
                                            0.793878
                                                               11.279592
mean
```

### 1.000000									
25% 2.000000 0.000000 6.000000 50% 3.000000 1.000000 10.000000 75% 4.000000 1.000000 15.000000 max 4.000000 3.000000 40.000000 TrainingTimesLastYear WorkLifeBalance YearsAtCompany \ count 1470.000000 1470.000000 1470.000000 mean 2.799320 2.761224 7.008163 std 1.289271 0.706476 6.126525 min 0.000000 1.000000 0.000000 25% 2.000000 2.000000 3.000000 5.000000 50% 3.000000 3.000000 5.000000 75% 3.000000 3.000000 9.000000 max 6.000000 1470.000000 40.000000 YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager count 1470.000000 mean 4.229252 2.187755 4.123129 std 3.623137 3.222430 3.568136 min 0.000000 25% 2.000000 0.000000 25% 3.000000 1.000000 25% 3.000000 1.000000 25% 3.000000 1.000000 75% 3.000000 1.000000 8000000 50% 3.000000 1.000000 17.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement', 'EnvironmentSatisfaction', 'JobInvolvement',	std	1.0812	09	0.852	077	7.780782			
TrainingTimesLastYear	min	1.0000	00	0.000	000	0.000000			
75%	25%	2.0000	00	0.000	000	6.000000			
TrainingTimesLastYear WorkLifeBalance YearsAtCompany \ count	50%	3.0000	00	1.000	000	10.000000			
TrainingTimesLastYear WorkLifeBalance YearsAtCompany \ Count	75%	4.0000	00	1.000	000	15.000000			
count 1470.000000 1470.000000 1470.000000 mean 2.799320 2.761224 7.008163 std 1.289271 0.706476 6.126525 min 0.000000 1.000000 0.000000 25% 2.000000 2.000000 3.000000 50% 3.000000 3.000000 5.000000 75% 3.000000 3.000000 9.000000 max 6.000000 4.000000 40.00000 YearsIncurrentRole YearsSinceLastPromotion YearsWithCurrManager Count 1470.000000 1470.000000 WearsWithCurrManager Count 1470.000000 1470.000000 WearsWithCurrManager 2.187755 4.123129 \$1.000000 0.000000 2.000000 0.000000 2.000000 0.000000 2.000000 0.000000 2.000000 2.000000 0.000000 3.000000	max	4.0000	00	3.000	000	40.000000			
count 1470.000000 1470.000000 1470.000000 mean 2.799320 2.761224 7.008163 std 1.289271 0.706476 6.126525 min 0.000000 1.000000 0.000000 25% 2.000000 2.000000 3.000000 50% 3.000000 3.000000 5.000000 75% 3.000000 3.000000 40.00000 YearsIncurrentRole YearsSinceLastPromotion YearsWithCurrManager count 1470.000000 1470.000000 WearsWithCurrManager count 1470.000000 1470.000000 Hears SinceLastPromotion YearsWithCurrManager count 1470.000000 Hears SinceLastPromotion YearsWithCurrManager Curr Manager A.129252 2.187755 4.129129 Store SinceLastPromotion Value 3.000000 0.000000 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td></t<>									
YearsWithCurrManager count 1470.000000 1470.000000 1470.000000 mean 4.229252 2.187755 4.123129 std 3.623137 3.222430 3.568136 min 0.000000 25% 2.000000 25% 2.000000 50% 3.000000 75% 7.000000 75% 7.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',	count mean std min 25% 50% 75%	1470.000000 2.799320 1.289271 0.000000 2.000000 3.000000 3.000000	1470 2 0 1 2 3 3	000000 761224 706476 000000 000000 000000	1470.00 7.00 6.12 0.00 3.00 5.00 9.00	90000 98163 26525 90000 90000 90000			
count 1470.000000 1470.000000 mean 4.229252 2.187755 4.123129 std 3.623137 3.222430 3.568136 min 0.000000 0.000000 0.000000 25% 2.000000 50% 3.000000 1.000000 3.000000 75% 7.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',			arsSinceL	.astPromo	tion				
mean 4.229252 2.187755 4.123129 std 3.623137 3.222430 3.568136 min 0.000000 0.000000 25% 2.000000 0.000000 2.000000 50% 3.000000 1.000000 75% 7.000000 3.000000 7.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',	count			1470.00	0000				
<pre>std</pre>	mean	4.229252		2.18	7755				
min 0.000000 0.000000 0.000000 0.000000 0.000000	std	3.623137		3.22	2430				
25% 2.000000 2.000000 50% 3.000000 75% 7.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',	min	0.00000		0.00	0000				
<pre>50%</pre>	25%	2.000000		0.00	0000				
75% 7.000000 3.000000 7.000000 max 18.000000 15.000000 17.000000 [8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',	50%	3.000000		1.00	0000				
<pre>max</pre>	75%	7.000000		3.00	0000				
<pre>[8 rows x 24 columns] dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',</pre>	max	18.000000		15.00	0000				
<pre>dataset3=dataset[['Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',</pre>	17.000000								
<pre>'Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',</pre>	[8 rows x 24	columns]							
	<pre>'Education', 'EmployeeNumber', 'EnvironmentSatisfaction', 'JobInvolvement',</pre>								

```
'JobSatisfaction',
     'NumCompaniesWorked',
    'PerformanceRating',
    'RelationshipSatisfaction',
    'StockOptionLevel',
    'TrainingTimesLastYear',
    'WorkLifeBalance'
]]
dataset3
      Education EmployeeNumber EnvironmentSatisfaction
JobInvolvement \
                                                          2
3
1
               1
                                                          3
2
2
               2
                                                          4
2
3
                               5
                                                          4
3
4
                                                          1
3
. . .
. .
                                                          3
1465
                            2061
1466
               1
                            2062
                                                          4
2
               3
                                                          2
1467
                            2064
               3
                                                          4
1468
                            2065
2
               3
                                                          2
1469
                            2068
4
      JobLevel JobSatisfaction NumCompaniesWorked PerformanceRating
\
0
                                                                         3
1
                                                                         4
2
                                                                         3
3
                                                                         3
                                                                         3
```

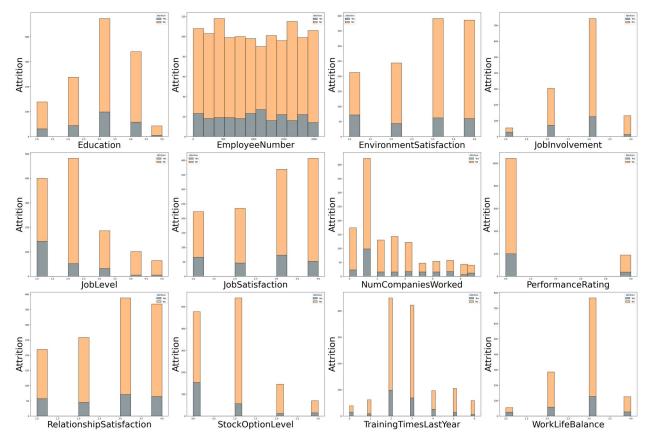
1465	2	4		4	3
1466	3	1		4	3
1467	2	2		1	4
1468	2	2		2	3
1469	2	3		2	3
D 1	1		10.11.1		
TrainingTime	onshipSatisfact sLastYear \				
0		1	0		
0 1 3 2 3 3 3 4 3		4	1		
3					
2		2	0		
3		3	0		
3		J	U		
4		4	1		
3					
1465		3	1		
3					
1466		1	1		
5 1467		2	1		
0		۷	1		
1468		4	0		
3		1	0		
1469 3		1	0		
5					
	feBalance				
0	1				
1	3				
2 3 4	3 3 3 3				
4	3				
1465					
1466	3				
1467	3				
1468	3 3 3 2 4				
1469	4				
[1470 rows x	12 columns]				

```
plt.figure(figsize=(40,35), facecolor='white')
plotnumber = 1

for column in dataset3:
    if plotnumber<=16:
        ax=plt.subplot(4,4,plotnumber)

sns.histplot(x=dataset3[column].dropna(axis=0), hue=dataset.Attrition)

    plt.xlabel(column, fontsize=40)
    plt.ylabel('Attrition', fontsize=40)
    plotnumber+=1
plt.tight_layout()</pre>
```

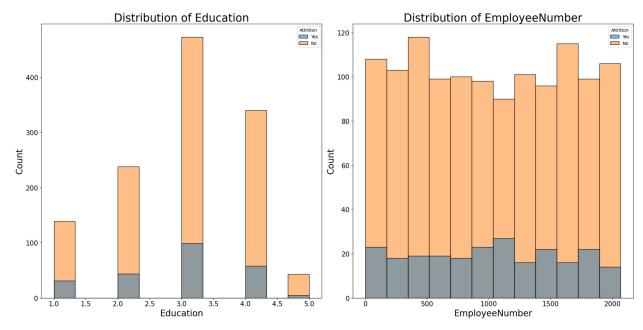


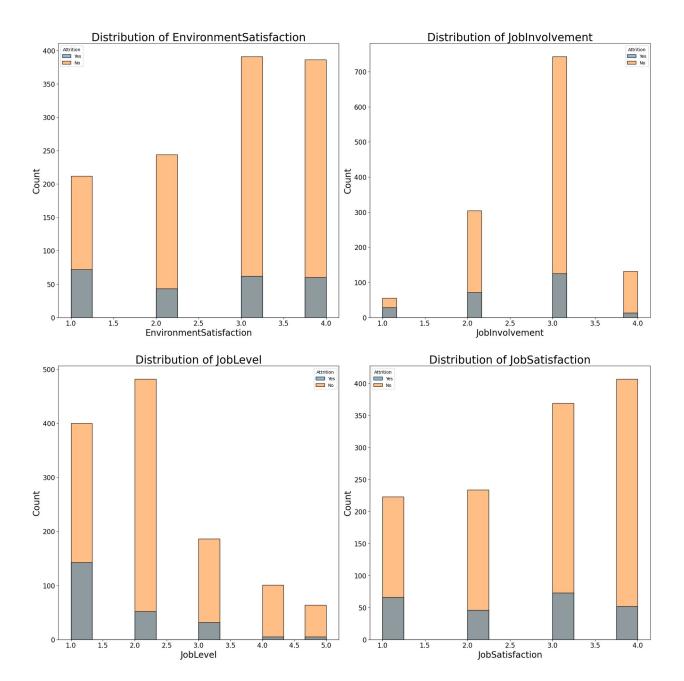
GRAPHICAL REPSENTATION

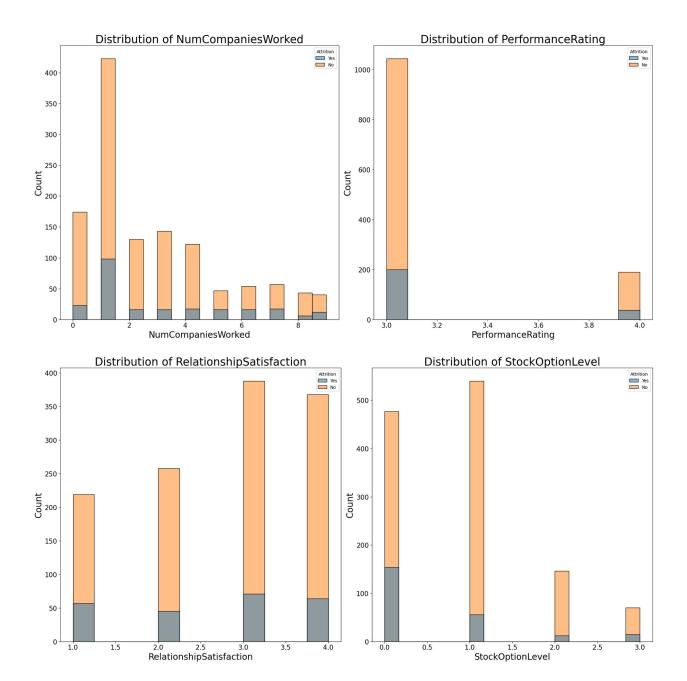
```
import matplotlib.pyplot as plt
import seaborn as sns

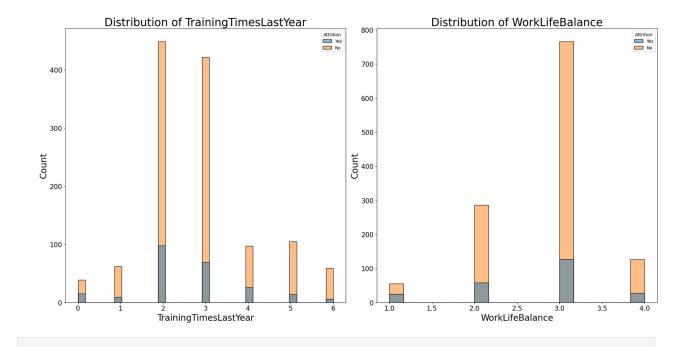
def plot_in_pairs(dataset3, hue_column):
    columns = dataset3.columns
    num_columns = len(columns)
```

```
pair_count = 2
    for i in range(0, num_columns, pair_count):
        plt.figure(figsize=(20, 10), facecolor='white')
        for j in range(pair count):
            if i + j < num_columns:</pre>
                ax = plt.subplot(1, pair_count, j + 1)
                sns.histplot(
                    x=dataset3[columns[i + j]].dropna(),
                    hue=hue column,
                    kde=False,
                    ax=ax
                )
                plt.xlabel(columns[i + j], fontsize=20)
                plt.ylabel('Count', fontsize=20)
                ax.set_title(f'Distribution of {columns[i + j]}',
fontsize=25)
                ax.tick params(axis='both', which='major',
labelsize=15)
        plt.tight layout()
        plt.show()
plot in pairs(dataset3, hue column=dataset.Attrition)
```



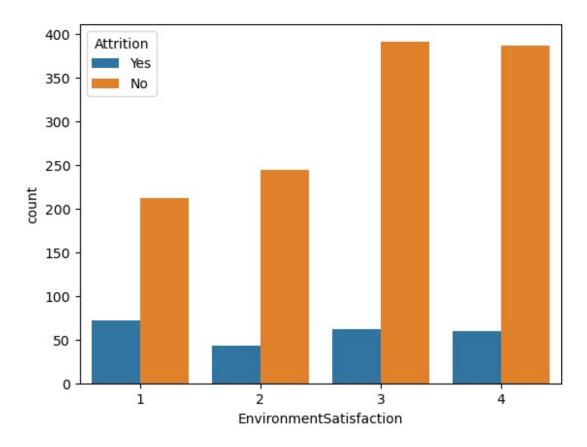






1 IMPACT OF ENVIRNMENT AND JOB SATISFACTION ON ATTRITION

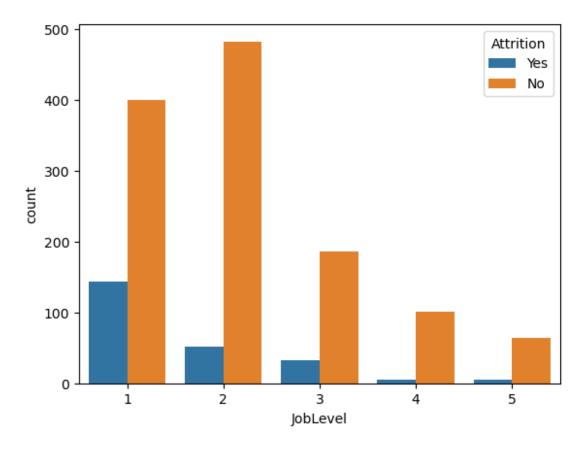
#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.EnvironmentSatisfaction, hue=dataset.Attrition)
plt.show()



--> Increase in rate of environment as well as job satisfaction give rise to increase in Attrtion "no" (means not willing to quit)

2 HOW JOB LEVEL IMPACTING ON ATTRITION

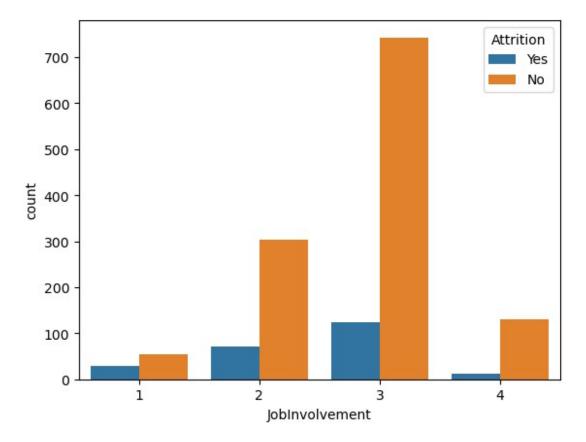
```
#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.JobLevel, hue=dataset.Attrition)
plt.show()
```



--> Increase in job level .Decrease in chances of leaving the company for Employees

3 JOBINVOLVEMENT IMPACT ON ATTRITION

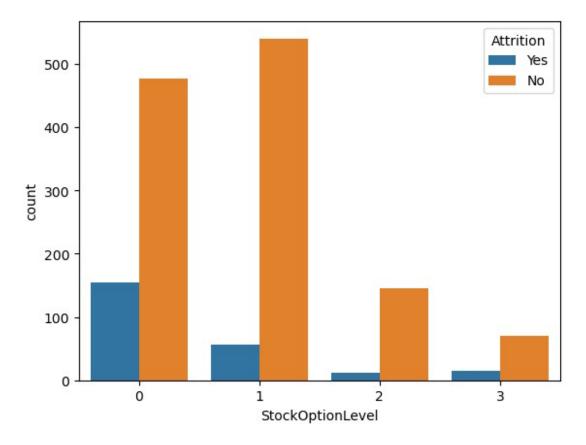
#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.JobInvolvement, hue=dataset.Attrition)
plt.show()



- --> The employees who involved in job more than sufficient are more likely to quit or we can say that they have mmore pressure of work
- --> Somehow. there are some emp. who not involvd fully in there job but still they are likely to quit

4 IMPACT OF STOCK OPTION LEVEL ON ATTRITION

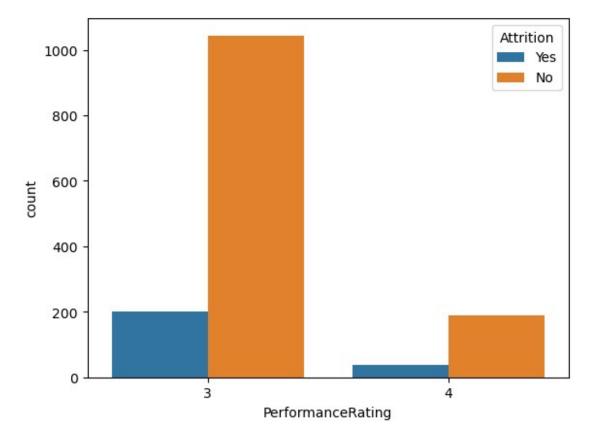
```
#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.StockOptionLevel, hue=dataset.Attrition)
plt.show()
```



-->For the employees who're not having stock option are likely to quit

5 PERFORMANCE RATING AND ATTRITION

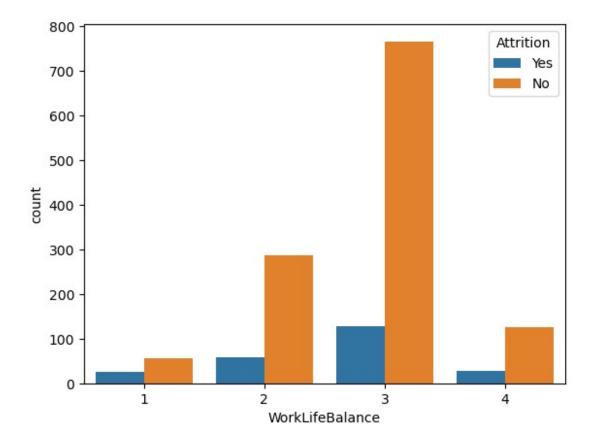
#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.PerformanceRating, hue=dataset.Attrition)
plt.show()



- --> On an average, most of the employees are moderately performed (because performance rating lies in 3-4).
- --> However employees having less performance rating are more likely to quit we can say that company wants to fire that employees

6 WORK LIFE BALANCE IMPACT ON ATTRITION

#plt.figure(figsize=(8,6), facecolor='white')
sns.countplot(x=dataset.WorkLifeBalance, hue=dataset.Attrition)
plt.show()



 $\mbox{--->}$ More the employees life is balance , lesser the Attrition.