PROJECT NAME : DATA SCIENCE PROJECT ON H.R DATA SET OF I.B.M COMPANY (USA)

DATA ANALYSIS :

- Data analysis is a process of cleaning, inspecting,transforming and discovering the information.
- In other words we can say that data analysis is used to get insights by summarizing the past data.
- It is simplest and most common technique used in business process.
- It allow us to make accurate decision and to stop guessing.

LANGUAGE USED ......> "PYTHON"

INTRODUCTION OF PROJECT :

TARGET COLUMN=="ATTRITION"

- ATTRITION........> Attrition is the departure of employee from the organization for any reason.
- In other words we want to say the employee leave their company.
- ATTRITION "YES"...........>Employee wants to leave the company.
- ATTRITION "NO"............>Employee don't want to leave.

WHAT WE ARE GOING TO DO IN THIS:

- IT will uncover the factors that lead to employee 'attrirtion'.
- In this project we are oing to analyse that how other fatures of ata set is effecting "ATTRITION".

ADVANTAGE OF DATA ANALYSIS :

- 1. Make decisions at the speed of your business.
- 2.Quickly detect and address operation analysis.
- 3.Increase business and address operational issues.
- 4.Improve customer service with up to data information.

PROJECT BY:

SHIVAM KOTALWAR

DOMAIN ANALYSIS:

IMPORTING THE PYTHON LIBRARIES

[ ]:

```
#importing librarys
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import warnings
warnings.filterwarnings('ignore')
```

[2]:

```
df_=pd.read_csv(r"D:\#python dataset\general_data.csv")
```

[3]:

```
df1=pd.read_csv(r"D:\#python dataset\employee_survey_data.
↪csv",encoding="cp1252")
```

Merging the dataset

[4]:

```
df=df_.merge(df1,how='left')
df
```

[4]:

```
      Age Attrition     BusinessTravel              Department  \
0      51        No      Travel_Rarely                   Sales
1      31       Yes  Travel_Frequently  Research & Development
2      32        No  Travel_Frequently  Research & Development
3      38        No         Non-Travel  Research & Development
4      32        No      Travel_Rarely  Research & Development
…     …        …                …                      …
4405   42        No      Travel_Rarely  Research & Development
4406   29        No      Travel_Rarely  Research & Development
4407   25        No      Travel_Rarely  Research & Development
4408   42        No      Travel_Rarely                   Sales
4409   40        No      Travel_Rarely  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount  EmployeeID  \
0                    6          2  Life Sciences              1           1
1                   10          1  Life Sciences              1           2
2                   17          4          Other              1           3
3                    2          5  Life Sciences              1           4
```

```
4                   10          1        Medical              1           5
…                …        …            …            …         …
4405                 5          4        Medical              1        4406
4406                 2          4        Medical              1        4407
4407                25          2  Life Sciences              1        4408
4408                18          2        Medical              1        4409
4409                28          3        Medical              1        4410

      Gender  …  StandardHours StockOptionLevel TotalWorkingYears  \
0     Female  …              8                0               1.0
1     Female  …              8                1               6.0
2       Male  …              8                3               5.0
3       Male  …              8                3              13.0
4       Male  …              8                2               9.0
…      …  …            …              …               …
4405  Female  …              8                1              10.0
4406    Male  …              8                0              10.0
4407    Male  …              8                0               5.0
4408    Male  …              8                1              10.0
4409    Male  …              8                0               NaN

      TrainingTimesLastYear  YearsAtCompany YearsSinceLastPromotion  \
0                         6               1                       0
1                         3               5                       1
2                         2               5                       0
3                         5               8                       7
4                         2               6                       0
…                     …             …                     …
4405                      5               3                       0
4406                      2               3                       0
4407                      4               4                       1
4408                      2               9                       7
4409                      6              21                       3

      YearsWithCurrManager  EnvironmentSatisfaction  JobSatisfaction  \
0                        0                      3.0              4.0
1                        4                      3.0              2.0
2                        3                      2.0              2.0
3                        5                      4.0              4.0
4                        4                      4.0              1.0
…                    …                      …              …
4405                     2                      4.0              1.0
4406                     2                      4.0              4.0
4407                     2                      1.0              3.0
4408                     8                      4.0              1.0
4409                     9                      1.0              3.0
```

```
      WorkLifeBalance
0                 2.0
1                 4.0
2                 1.0
3                 3.0
4                 3.0
…               …
4405              3.0
4406              3.0
4407              3.0
4408              3.0
4409              NaN

[4410 rows x 27 columns]
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 27 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Age                      4410 non-null   int64
 1   Attrition                4410 non-null   object
 2   BusinessTravel           4410 non-null   object
 3   Department               4410 non-null   object
 4   DistanceFromHome         4410 non-null   int64
 5   Education                4410 non-null   int64
 6   EducationField           4410 non-null   object
 7   EmployeeCount            4410 non-null   int64
 8   EmployeeID               4410 non-null   int64
 9   Gender                   4410 non-null   object
 10  JobLevel                 4410 non-null   int64
 11  JobRole                  4410 non-null   object
 12  MaritalStatus            4410 non-null   object
 13  MonthlyIncome            4410 non-null   int64
 14  NumCompaniesWorked       4391 non-null   float64
 15  Over18                   4410 non-null   object
 16  PercentSalaryHike        4410 non-null   int64
 17  StandardHours            4410 non-null   int64
 18  StockOptionLevel         4410 non-null   int64
 19  TotalWorkingYears        4401 non-null   float64
 20  TrainingTimesLastYear    4410 non-null   int64
 21  YearsAtCompany           4410 non-null   int64
 22  YearsSinceLastPromotion  4410 non-null   int64
 23  YearsWithCurrManager     4410 non-null   int64
 24  EnvironmentSatisfaction  4385 non-null   float64
 25  JobSatisfaction          4390 non-null   float64
 26  WorkLifeBalance          4372 non-null   float64
dtypes: float64(5), int64(14), object(8)
memory usage: 930.4+ KB
```

[6]:

```
df.isnull().sum()
df.dropna(inplace=True)
```

#Dropping the not required columns basis of unique data

[7]:

```
df[["Over18","EmployeeCount"]].nunique()#not need to perform analysis on unique␣
 ↪value coloumns
df.drop(["Over18","EmployeeCount"],inplace=True,axis=1)# droping the columns␣
 ↪from the dataset
```

[8]:

```
percentage_of_attrition=df["Attrition"].value_counts(normalize=True)
percentage_of_attrition
```

[8]:

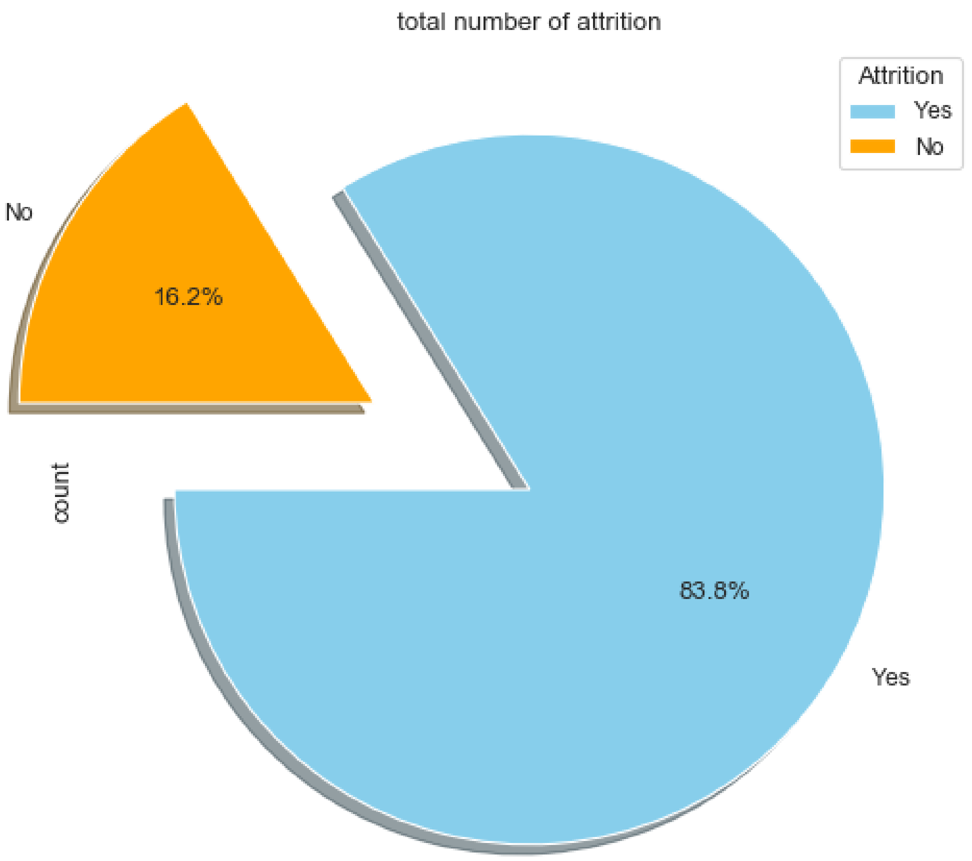
```
Attrition
No     0.838372
Yes    0.161628
Name: proportion, dtype: float64
```

Attritionsaying yes=83% ,No=16% something.

[129]:

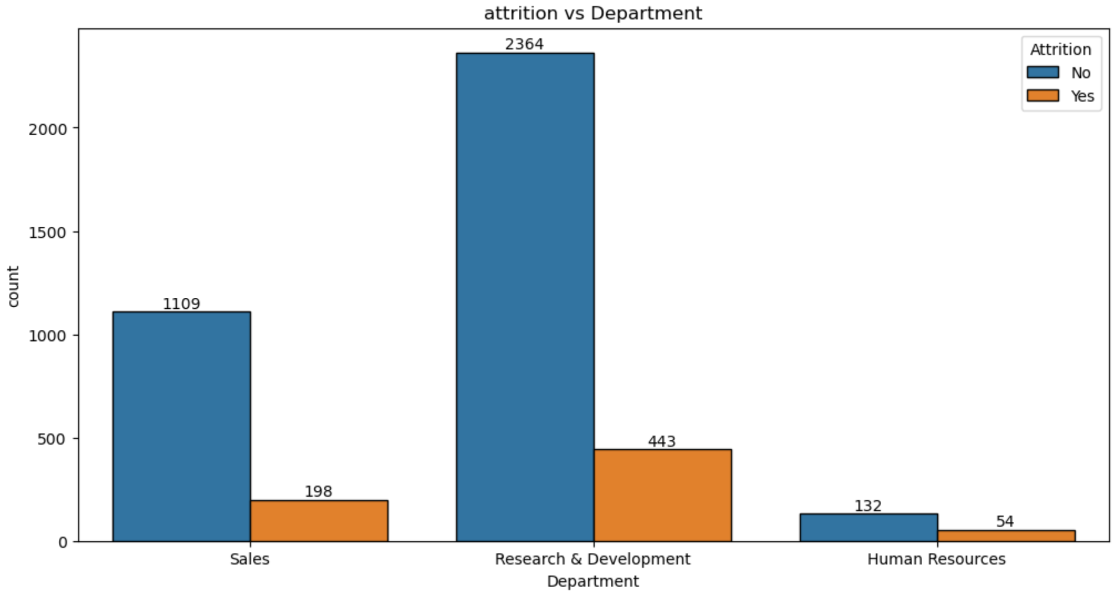
```
plt.figure(figsize=(12,6))
df["Attrition"].value_counts().
 ↪plot(kind='pie',labels=['Yes','No'],shadow=True,autopct='%1.1f%%',
                                    explode=[0,0.
 ↪5],colors=["skyblue","orange"],startangle=180,)
plt.legend(title="Attrition")
plt.title("total number of attrition")
plt.show()
```

PERFORMING EDA



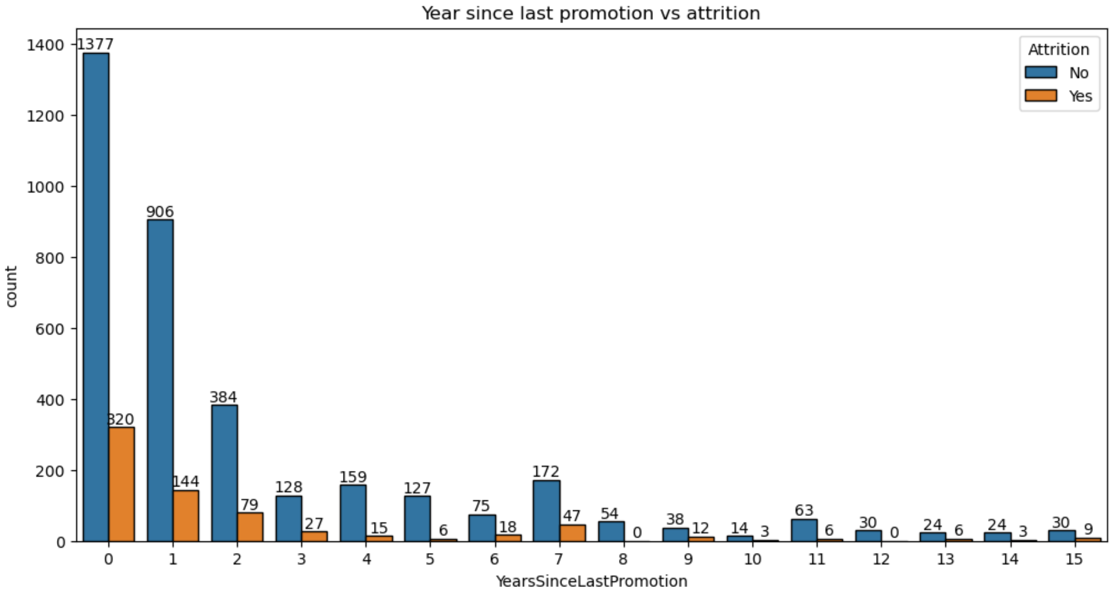
```
[9]: #attrition vs department
     plt.figure(figsize=(12,6))
     att1=sns.countplot(df,x="Department",hue="Attrition",edgecolor="black")
     plt.title("attrition vs Department")
     for bars in att1.containers:
        att1.bar_label(bars)

     plt.show()
     #in this we see the attrition on different department
```



The impact of attrition in various Department. * in human resources department are says about half of employee are gradually decreased.

[11]:
```
# attrition vs promotion
plt.figure(figsize=(12,6))
promotion_=df["YearsSinceLastPromotion"].value_counts(normalize=True)
promotion_=df["Attrition"].value_counts()
promotion_
promotion_=sns.
 ↪countplot(data=df,x="YearsSinceLastPromotion",hue="Attrition",edgecolor="black",)
plt.title("Year since last promotion vs attrition")
for bars in promotion_.containers:
   promotion_.bar_label(bars)
plt.show()
```

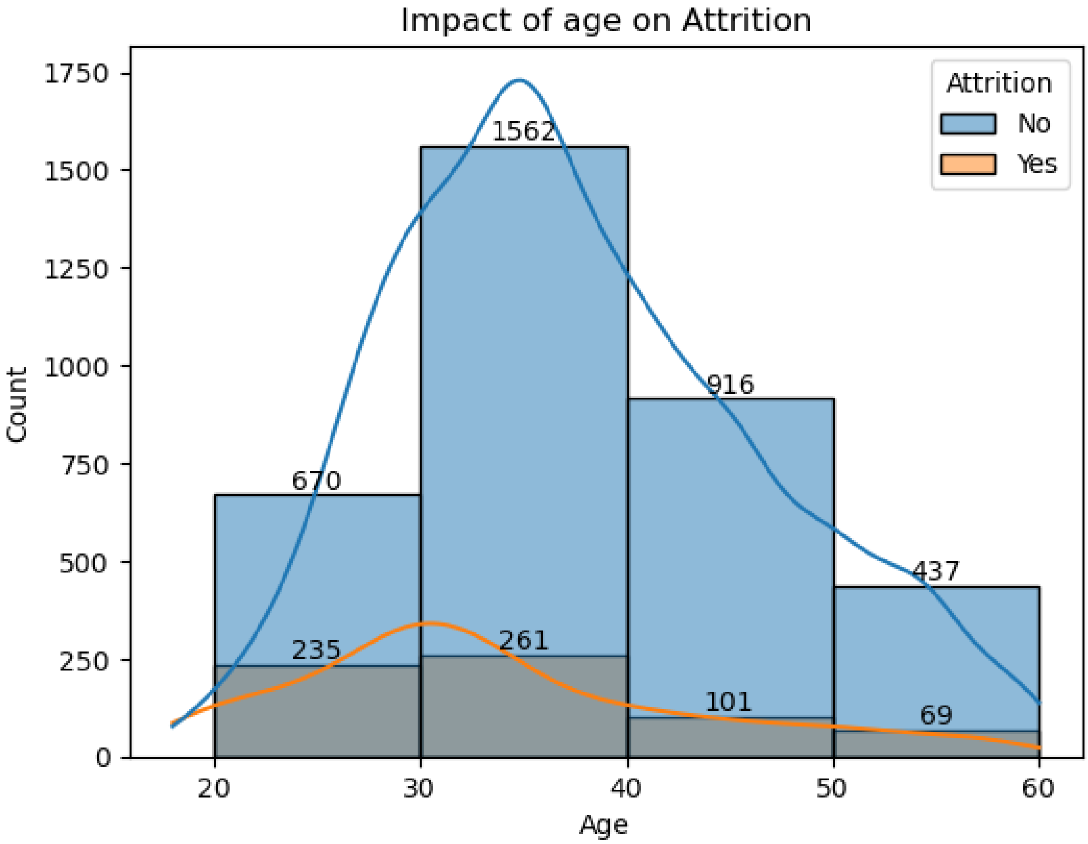


The chart describe how the Attrition occur with Promotion * The impact of attrition lesser the promotion higher the impact. * More affect of Attrition are seen to be from 0 to 1 year promotion employee

- Impact of Age on attrition of company

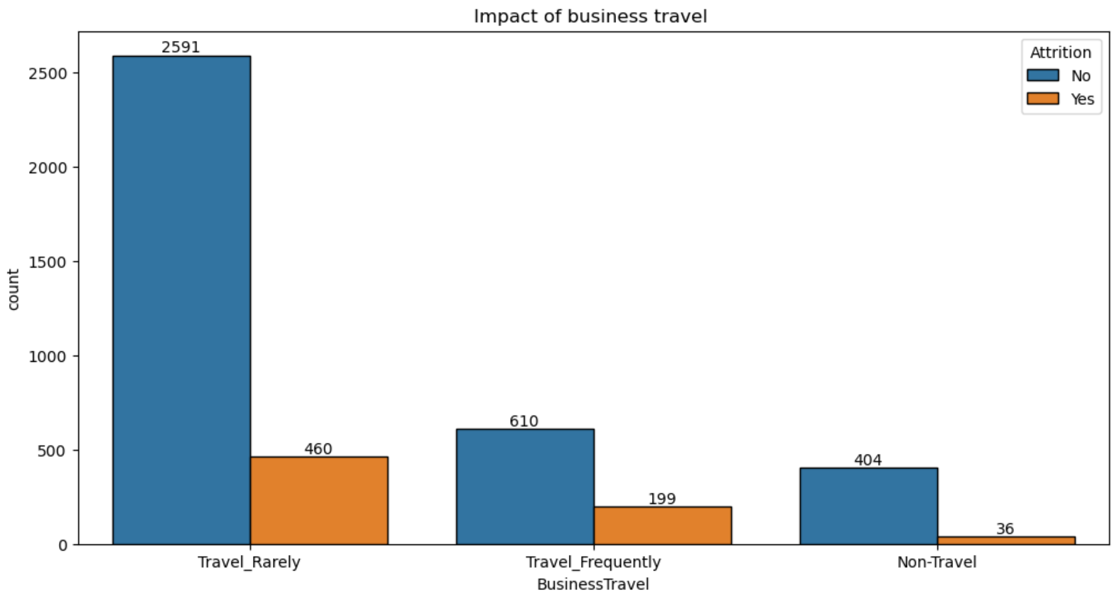
[12]:

```
 Age_Impact=sns.histplot(hue=df.Attrition,x=df.
 ↪Age,bins=[20,30,40,50,60],kde=True)
plt.title("Impact of age on Attrition")
for bars in Age_Impact.containers:
   Age_Impact.bar_label(bars)
plt.show()
```



The current scenario are Lookng the , * The data says 20 to 35 age employee are drop/leave the job. * The 40 to 60 data said higher the age lesser the Attrition

[13]:
```
plt.figure(figsize=(12,6))
travel=sns.countplot(x=df.BusinessTravel,hue=df.Attrition,edgecolor="Black")
plt.title("Impact of business travel ")
for bars in travel.containers:
    travel.bar_label(bars)
plt.show()
```

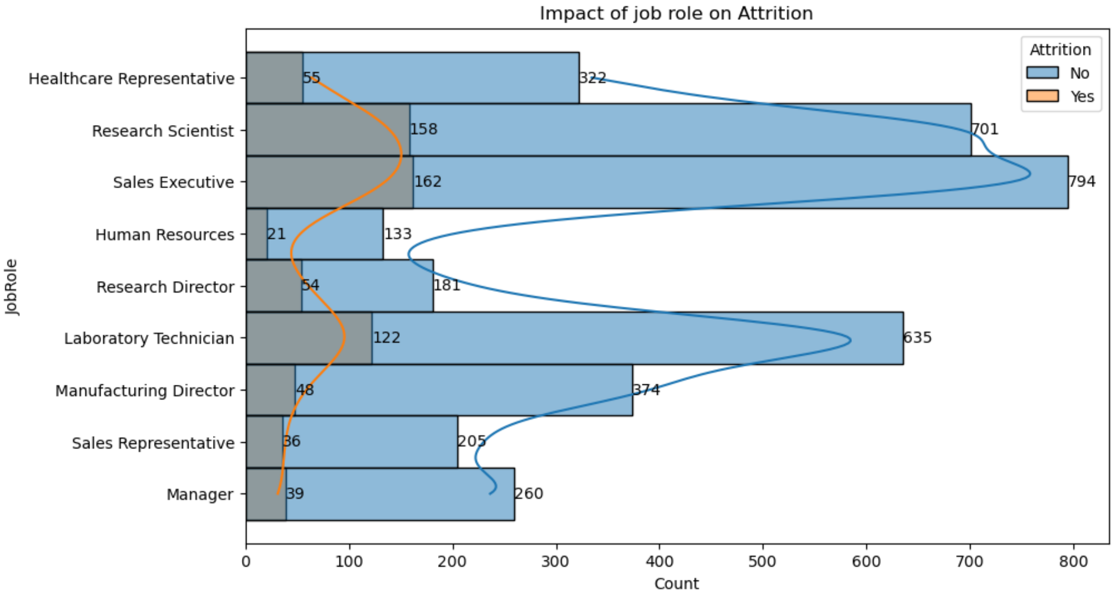


After Visulization of data impact of Attrition on Travel : * Those who are travel frequently are leave the job.

```
* Impact of job role on Attrition
```

[15]:

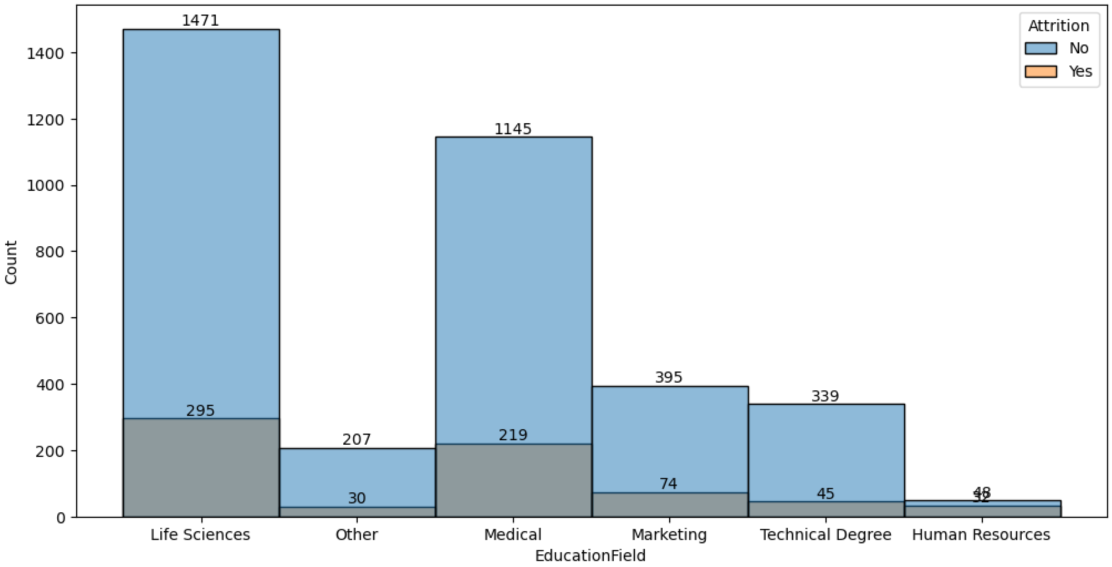
```
plt.figure(figsize=(10,6))
job_role=sns.histplot(hue=df.Attrition,y=df.JobRole ,edgecolor='black',kde=True)
plt.title("Impact of job role on Attrition")
for bars in job_role.containers:
    job_role.bar_label(bars)
plt.show()
```



The most Impacted Job role are : * Reserch scientist,sales executive,Laboratry Tecnician

```
Impact of Education field on Attrition
```

```
[16]: plt.figure(figsize=(12,6))
      edu=sns.histplot(x=df.EducationField,hue=df.Attrition,color="")
      for bars in edu.containers:
          edu.bar_label(bars)
      plt.show()
```



- Attrition most occur in human resource,Marketing,or more in medical

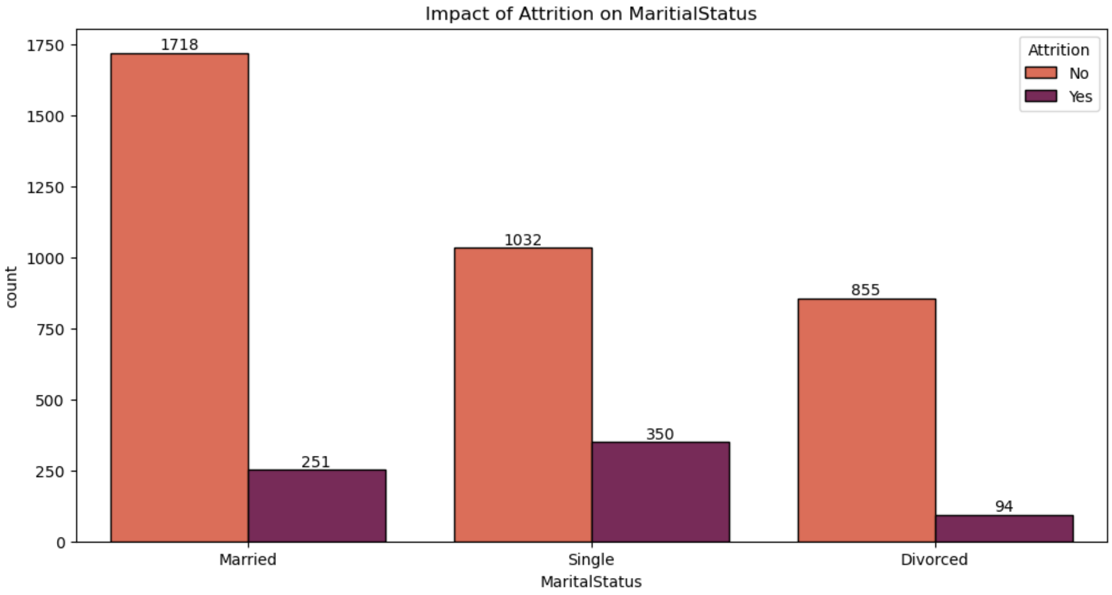
```
[17]: df.columns
```

```
[17]: Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
             'Education', 'EducationField', 'EmployeeID', 'Gender', 'JobLevel',
             'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',
             'PercentSalaryHike', 'StandardHours', 'StockOptionLevel',
             'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany',
             'YearsSinceLastPromotion', 'YearsWithCurrManager',
             'EnvironmentSatisfaction', 'JobSatisfaction', 'WorkLifeBalance'],
            dtype='object')
```

```
Impact of Attrition on maritial status
```

```
[18]: plt.figure(figsize=(12,6))
```

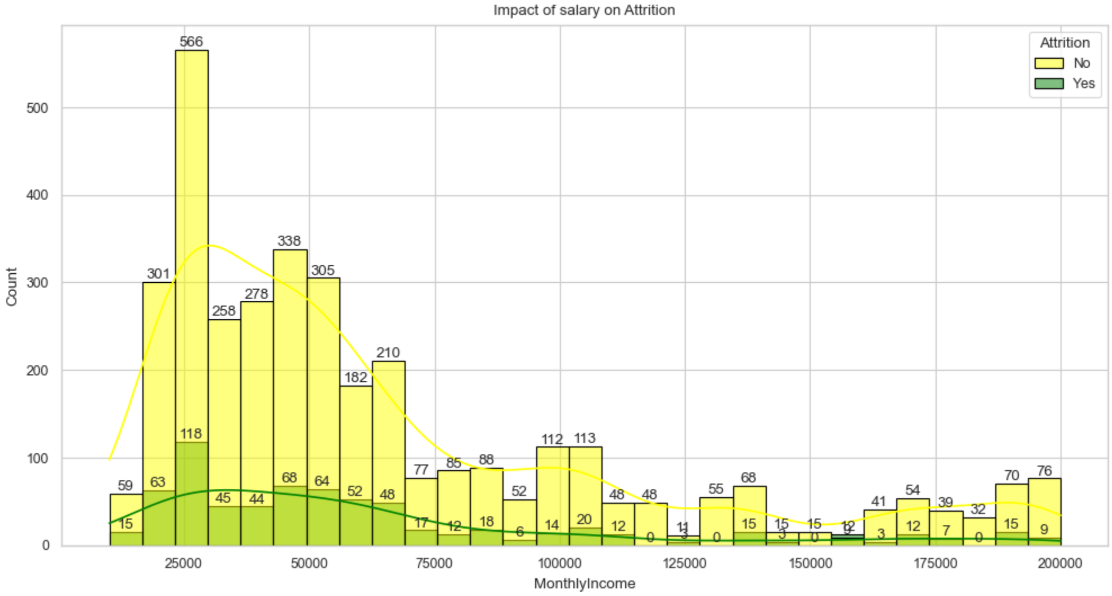
```
maritial_status= sns.countplot(x=df.MaritalStatus,hue=df.
 ↪Attrition,palette='rocket_r',edgecolor="Black")
plt.title("Impact of Attrition on MaritialStatus")
sns.color_palette("rocket_r", as_cmap=True)
for bars in maritial_status.containers:
    maritial_status.bar_label(bars)
```



- Single Maritial status employee are mostly resign their job.

`Monthly income vs Attrition of company`

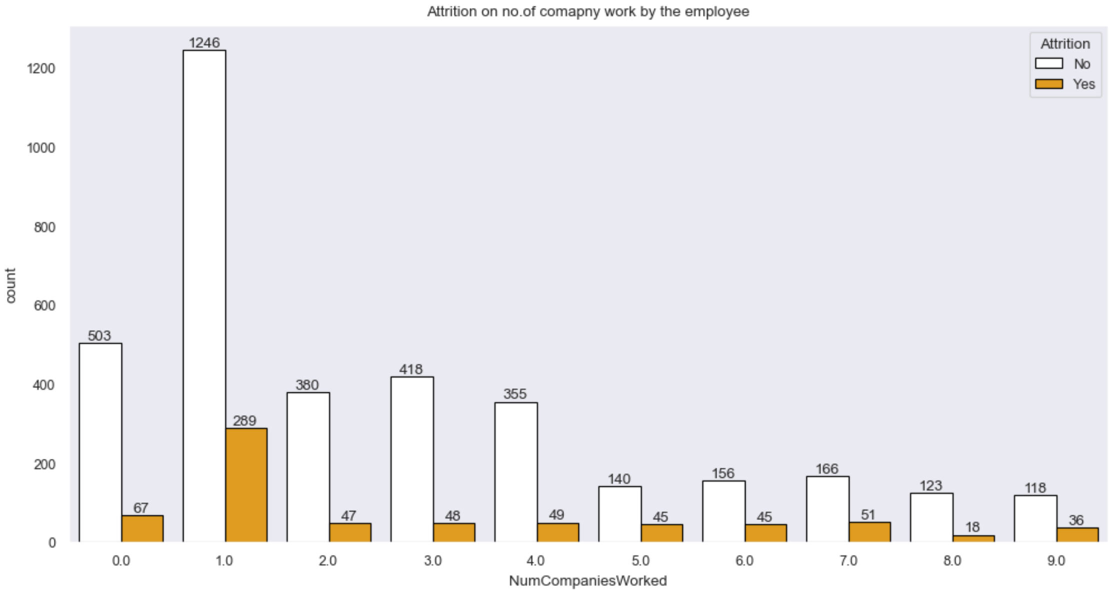
```
[19]: sns.set_style('whitegrid')
sns.set_context('paper')
plt.figure(figsize=(12,6))
monthaly_income = sns.histplot(x=df.MonthlyIncome,hue=df.
 ↪Attrition,palette=["yellow",'green'],edgecolor='black',kde=True,)
for bars in monthaly_income.containers:
    monthaly_income.bar_label(bars)
    plt.title("Impact of salary on Attrition")
```



- lesser the salary higher the Attrition.25000 to 70000
- higher the salary lesser the attrition.>75000

No.of comapny work by the employee impact on Attrition

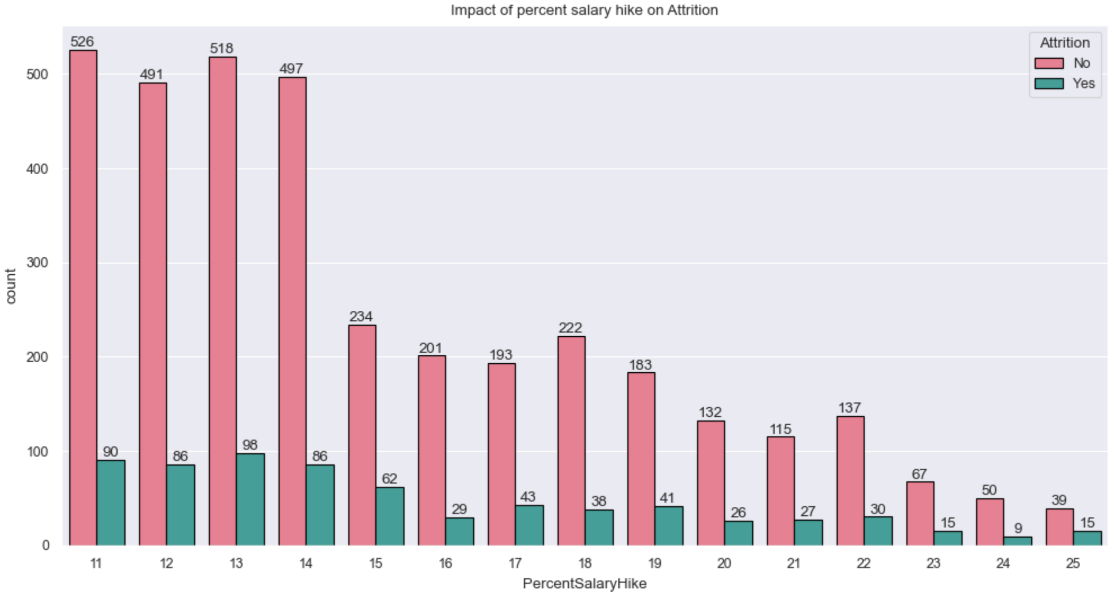
[31]:
```
plt.figure(figsize=(12,6))
sns.set_context("paper")
company_worked=sns.countplot(x=df.NumCompaniesWorked,hue=df.Attrition,␣
 ↪palette=['white',"orange"],edgecolor='black')
plt.title("Attrition on no.of comapny work by the employee")
for bars in company_worked.containers:
    company_worked.bar_label(bars)
plt.show()
print()
print("*The chart shows the the more attrition come from,\n the employee work␣
 ↪more than 5 companys the have more Attrition.")
```



*The chart shows the the more attrition come from,
 the employee work more than 5 companys the have more Attrition.

Impact of salary hike on Attrition

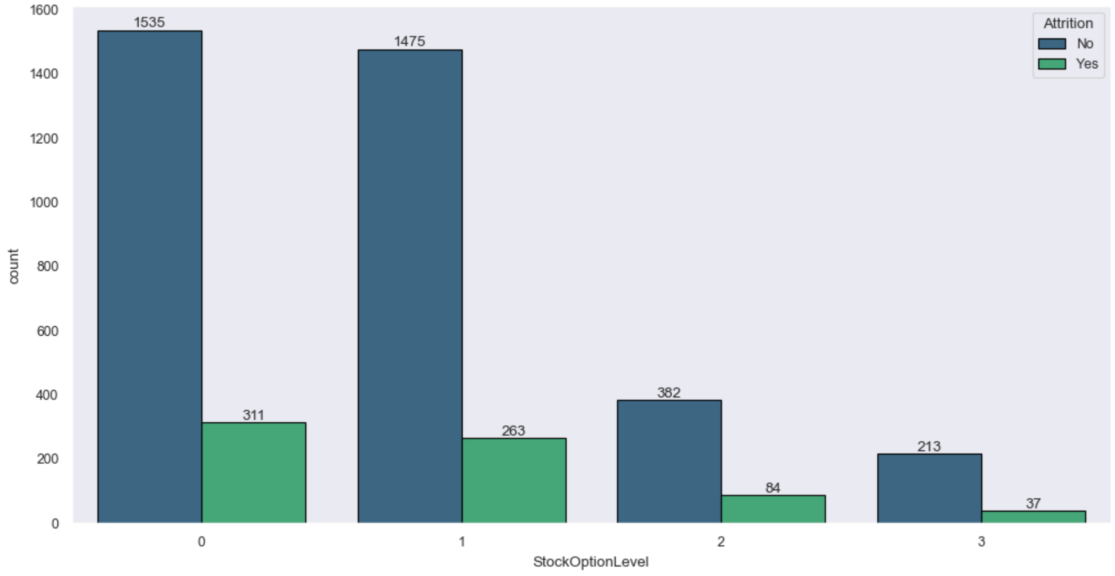
[25]:
```
sns.set_context('paper')
sns.set_style("darkgrid")
plt.figure(figsize=(12,6))
salary_hike=sns.countplot(x=df.PercentSalaryHike,hue=df.Attrition,palette="husl",edgecolor='black')
plt.title("Impact of percent salary hike on Attrition")
for bars in salary_hike.containers:
    salary_hike.bar_label(bars)
plt.savefig("salary.jpg")
plt.show()
print("All numbers are percentage")
print("chart clearlly show more number of emplyee more the Attrition.\n less number of employee same the Attrtion ")
print("The attrition come more when salary increased 15% to 25%")
```



```
All numbers are percentage
chart clearlly show more number of emplyee more the Attrition.
 less number of employee same the Attrtion
The attrition come more when salary increased 15% to 25%
```

[30]:

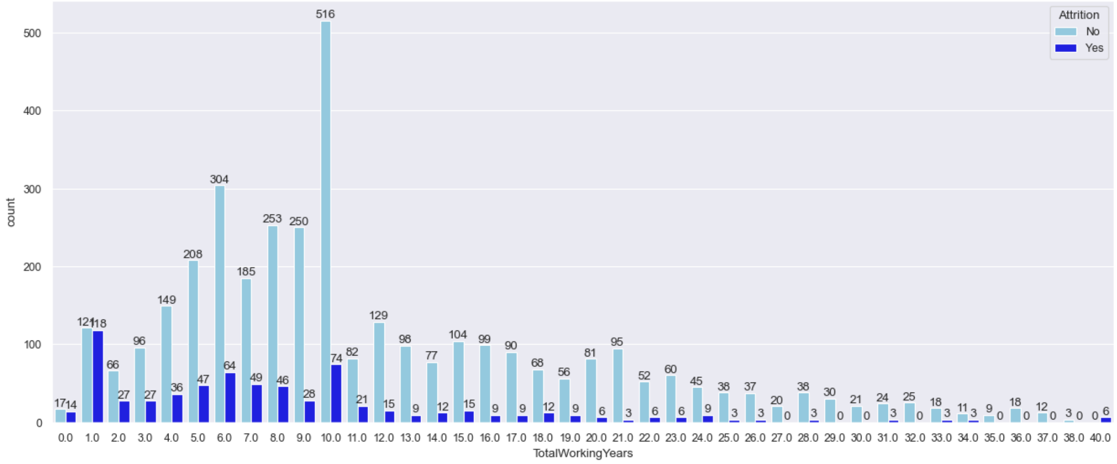
```
plt.figure(figsize=(12,6))
sns.set_context("paper")
sns.set_style("dark")
stockoption=sns.countplot(x=df.StockOptionLevel,hue=df.
  ↪Attrition,palette="viridis",edgecolor='black')
for bars in stockoption.containers:
    stockoption.bar_label(bars)
plt.show()
```



- The employee have zero stock option they leaved their company.

IMPACT OF ATTRITION ON TOTAL WORKING YEARS

[45]:
```
plt.figure(figsize=(15,6))
sns.set_context('paper')
sns.set_style("darkgrid")
workyears=sns.countplot(hue=df.Attrition,x=df.
 ↪TotalWorkingYears,palette=['skyblue','blue'])
for bars in workyears.containers:
    workyears.bar_label(bars)
```

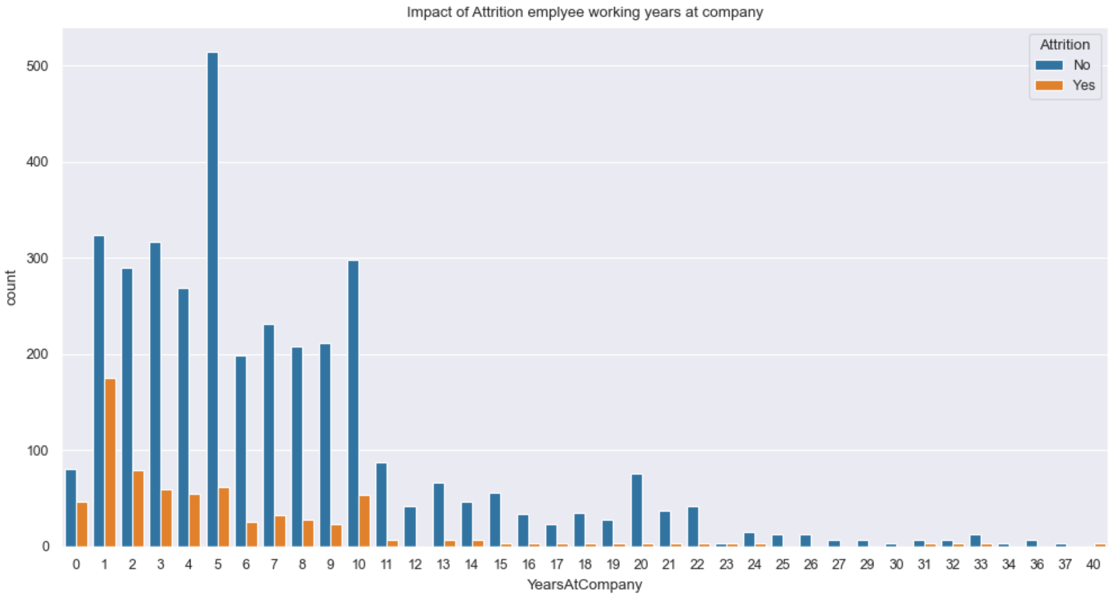


- Attrition come from those employee who have 0 to 5 years of experince.

- Attrition more in 0 or 1 year of experince emplyee

[70]:

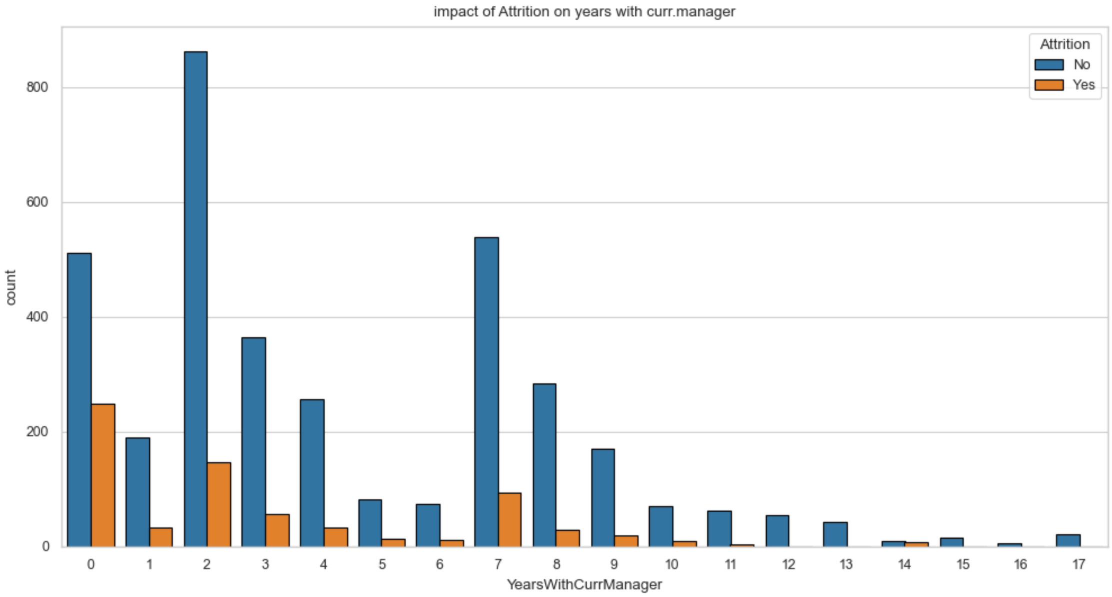
```
plt.figure(figsize=(12,6))
sns.set_context("paper")
fig=sns.countplot(x=df.YearsAtCompany,hue=df.Attrition)
plt.title("Impact of Attrition emplyee working years at company")
plt.show()
```



- The graph describe thos employee who hace worked 0 to 5 years they leave the company.
- Attrition more in 0 or 1 year.
- the employee who work with company 1 year they leave .

[79]:

```
sns.set_context("paper")
sns.set_style('whitegrid')
plt.figure(figsize=(12,6))
manager=sns.countplot(x=df.YearsWithCurrManager,hue=df.
  ↪Attrition,edgecolor='black')
plt.title("impact of Attrition on years with curr.manager")
plt.show()
```

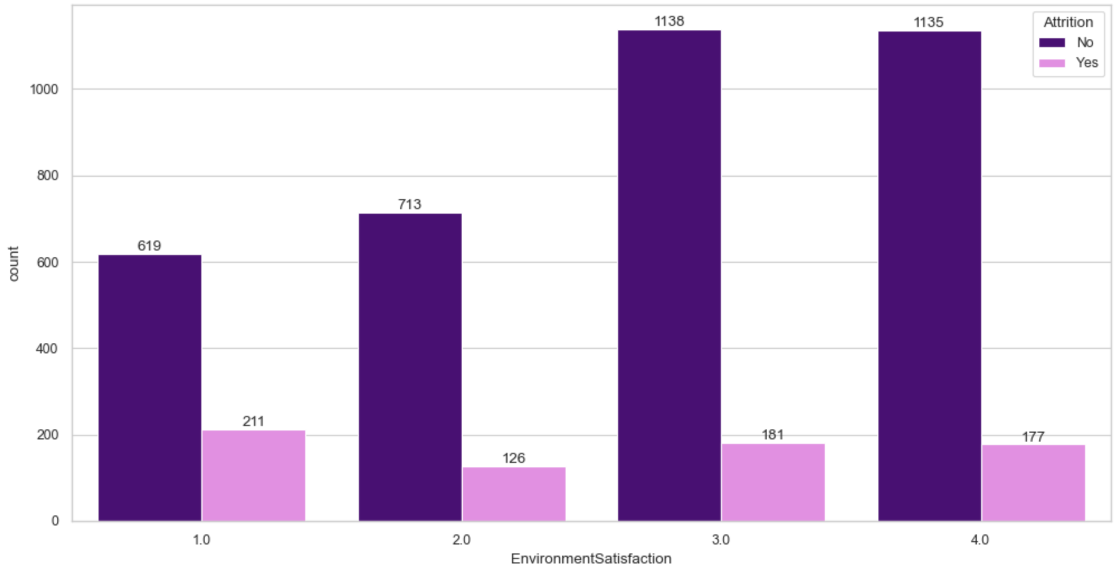


- the employee have 0 year with current manager they fired most.

[97]:
```
df.columns
```

[97]:
```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
       'Education', 'EducationField', 'EmployeeID', 'Gender', 'JobLevel',
       'JobRole', 'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',
       'PercentSalaryHike', 'StandardHours', 'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany',
       'YearsSinceLastPromotion', 'YearsWithCurrManager',
       'EnvironmentSatisfaction', 'JobSatisfaction', 'WorkLifeBalance'],
      dtype='object')
```

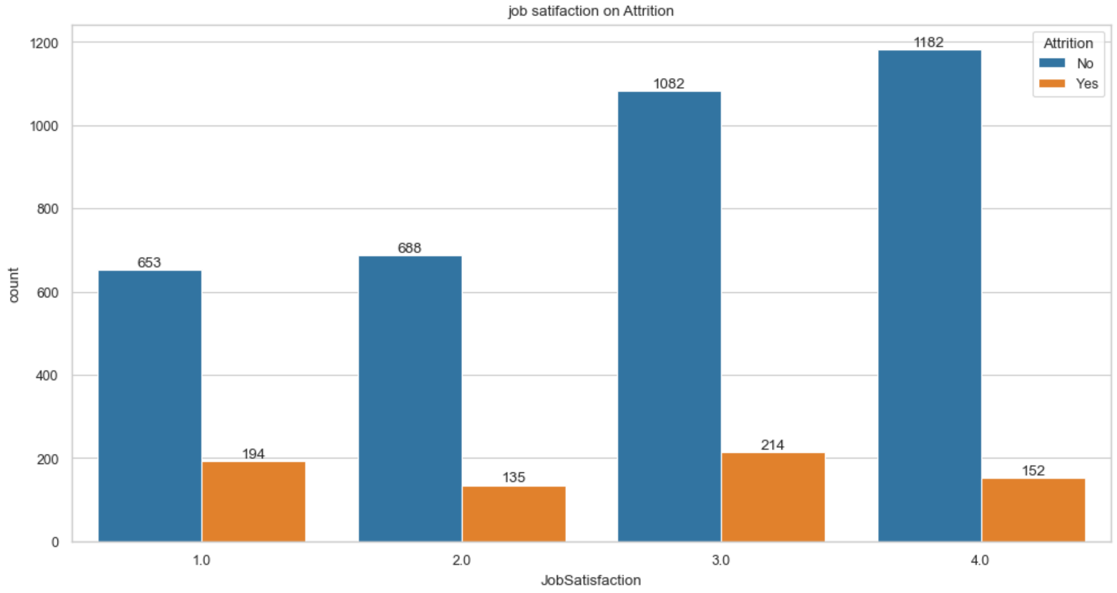
[107]:
```
plt.figure(figsize=(12,6))
enviorment=sns.countplot(x=df.EnvironmentSatisfaction,hue=df.
 ↪Attrition,palette=['Indigo','Violet'])
for bars in enviorment.containers:
    enviorment.bar_label(bars)
```



```
[ ]:
```

The chart describe employee have 1 rating they won't satisfied to the compony enviorment.

```
[116]: plt.figure(figsize=(12,6))
jobsatisfaction=sns.countplot(hue=df.Attrition,x=df.JobSatisfaction)
for bars in jobsatisfaction.containers:
    jobsatisfaction.bar_label(bars)
    plt.title('job satifaction on Attrition')
plt.show()
```

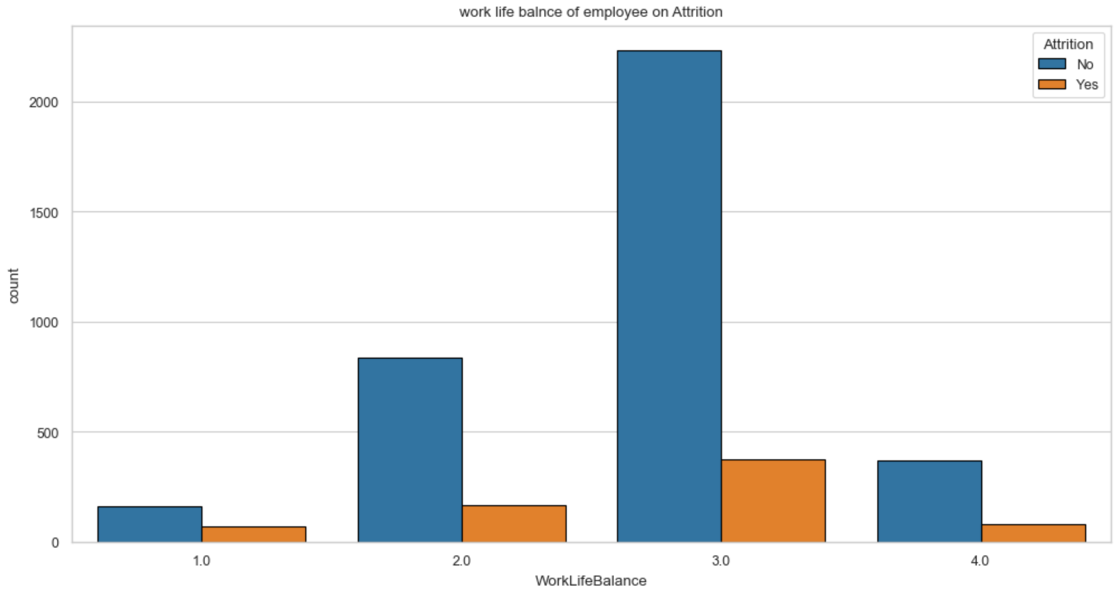


The chart describe employee have 1 or 2 rating they won’t satisfied to the job.

*Work life balance of employee on Attrition

[132]:

```
sns.set_context("paper")
sns.set_style('whitegrid')
plt.figure(figsize=(12,6))
sns.countplot(x=df.WorkLifeBalance,hue=df.Attrition,edgecolor='black')
plt.title("work life balnce of employee on Attrition")
plt.show()
```



the most of the employee not maintain their work life balance.

[ ]:

```
```