PROJECT – EMPLOYEE ATTRITION

Import Important Labraries

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

Importing Dataset

```python
data=pd.read_csv(r"C:\Users\tinu9\Downloads\Attrition data.csv")
```

```python
data.head()
```

```
   EmployeeID  Age Attrition     BusinessTravel              Department  \
0           1   51        No      Travel_Rarely                   Sales
1           2   31       Yes  Travel_Frequently  Research & Development
2           3   32        No  Travel_Frequently  Research & Development
3           4   38        No         Non-Travel  Research & Development
4           5   32        No      Travel_Rarely  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  Gender  ...  \
0                 6          2  Life Sciences              1  Female  ...
1                10          1  Life Sciences              1  Female  ...
2                17          4          Other              1    Male  ...
3                 2          5  Life Sciences              1    Male  ...
4                10          1        Medical              1    Male  ...

   TotalWorkingYears  TrainingTimesLastYear  YearsAtCompany  \
0                1.0                      6               1
1                6.0                      3               5
2                5.0                      2               5
3               13.0                      5               8
4                9.0                      2               6

   YearsSinceLastPromotion  YearsWithCurrManager  EnvironmentSatisfaction  \
0                        0                     0                      3.0
1                        1                     4                      3.0
2                        0                     3                      2.0
3                        7                     5                      4.0
4                        0                     4                      4.0
```

```
     JobSatisfaction  WorkLifeBalance  JobInvolvement  PerformanceRating
0              4.0              2.0               3                  3
1              2.0              4.0               2                  4
2              2.0              1.0               3                  3
3              4.0              3.0               2                  3
4              1.0              3.0               3                  3

[5 rows x 29 columns]
```

Basic checks

data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 29 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   EmployeeID               4410 non-null   int64
 1   Age                      4410 non-null   int64
 2   Attrition                4410 non-null   object
 3   BusinessTravel           4410 non-null   object
 4   Department               4410 non-null   object
 5   DistanceFromHome         4410 non-null   int64
 6   Education                4410 non-null   int64
 7   EducationField           4410 non-null   object
 8   EmployeeCount            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
 27  JobInvolvement           4410 non-null   int64
 28  PerformanceRating        4410 non-null   int64
```

```
dtypes: float64(5), int64(16), object(8)
memory usage: 999.3+ KB
```

. Here information tell us that there is no null values are present in data,So need not to handle null values.

data.describe

```
<bound method NDFrame.describe of        EmployeeID  Age Attrition
BusinessTravel              Department   \
0                1   51        No      Travel_Rarely                   Sales
1                2   31       Yes  Travel_Frequently  Research & Development
2                3   32        No  Travel_Frequently  Research & Development
3                4   38        No         Non-Travel  Research & Development
4                5   32        No      Travel_Rarely  Research & Development
...            ...  ...       ...                ...                     ...
4405          4406   42        No      Travel_Rarely  Research & Development
4406          4407   29        No      Travel_Rarely  Research & Development
4407          4408   25        No      Travel_Rarely  Research & Development
4408          4409   42        No      Travel_Rarely                   Sales
4409          4410   40        No      Travel_Rarely  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount  Gender  ...
\
0                    6          2  Life Sciences              1  Female  ...
1                   10          1  Life Sciences              1  Female  ...
2                   17          4          Other              1    Male  ...
3                    2          5  Life Sciences              1    Male  ...
4                   10          1        Medical              1    Male  ...
...                ...        ...            ...            ...     ...  ...
4405                 5          4        Medical              1  Female  ...
4406                 2          4        Medical              1    Male  ...
4407                25          2  Life Sciences              1    Male  ...
4408                18          2        Medical              1    Male  ...
4409                28          3        Medical              1    Male  ...

      TotalWorkingYears TrainingTimesLastYear YearsAtCompany  \
0                   1.0                     6              1
1                   6.0                     3              5
2                   5.0                     2              5
3                  13.0                     5              8
4                   9.0                     2              6
...                 ...                   ...            ...
4405               10.0                     5              3
```

```
4406                    10.0                    2                3
4407                     5.0                    4                4
4408                    10.0                    2                9
4409                     NaN                    6               21
```

```
        YearsSinceLastPromotion  YearsWithCurrManager  EnvironmentSatisfaction
\
0                            0                     0                      3.0
1                            1                     4                      3.0
2                            0                     3                      2.0
3                            7                     5                      4.0
4                            0                     4                      4.0
...                        ...                   ...                      ...
4405                         0                     2                      4.0
4406                         0                     2                      4.0
4407                         1                     2                      1.0
4408                         7                     8                      4.0
4409                         3                     9                      1.0
```

```
        JobSatisfaction  WorkLifeBalance  JobInvolvement  PerformanceRating
0                   4.0              2.0               3                  3
1                   2.0              4.0               2                  4
2                   2.0              1.0               3                  3
3                   4.0              3.0               2                  3
4                   1.0              3.0               3                  3
...                 ...              ...             ...                ...
4405                1.0              3.0               3                  3
4406                4.0              3.0               2                  3
4407                3.0              3.0               3                  4
4408                1.0              3.0               2                  3
4409                3.0              NaN               4                  3
```

[4410 rows x 29 columns]>

. Here two columns(employ_count , standard_Hours)have std==0.so,it means they have unique values it means they are not able to use in analysis.

data.isnull().sum()

```
EmployeeID              0
Age                     0
Attrition               0
BusinessTravel          0
Department              0
```

```
DistanceFromHome              0
Education                     0
EducationField                0
EmployeeCount                 0
Gender                        0
JobLevel                      0
JobRole                       0
MaritalStatus                 0
MonthlyIncome                 0
NumCompaniesWorked           19
Over18                        0
PercentSalaryHike             0
StandardHours                 0
StockOptionLevel              0
TotalWorkingYears             9
TrainingTimesLastYear         0
YearsAtCompany                0
YearsSinceLastPromotion       0
YearsWithCurrManager          0
EnvironmentSatisfaction      25
JobSatisfaction              20
WorkLifeBalance              38
JobInvolvement                0
PerformanceRating             0
dtype: int64
```

EDA(EXPLORATORY DATA ANALYSIS)

ANALYSIS ON CATAGORICAL COLUMNS W.R.T TARGET COLUMN(ATTRITION)

TARGET COLUMN==ATTRITION

```
sns.countplot(x=data.Attrition)
plt.show()
```
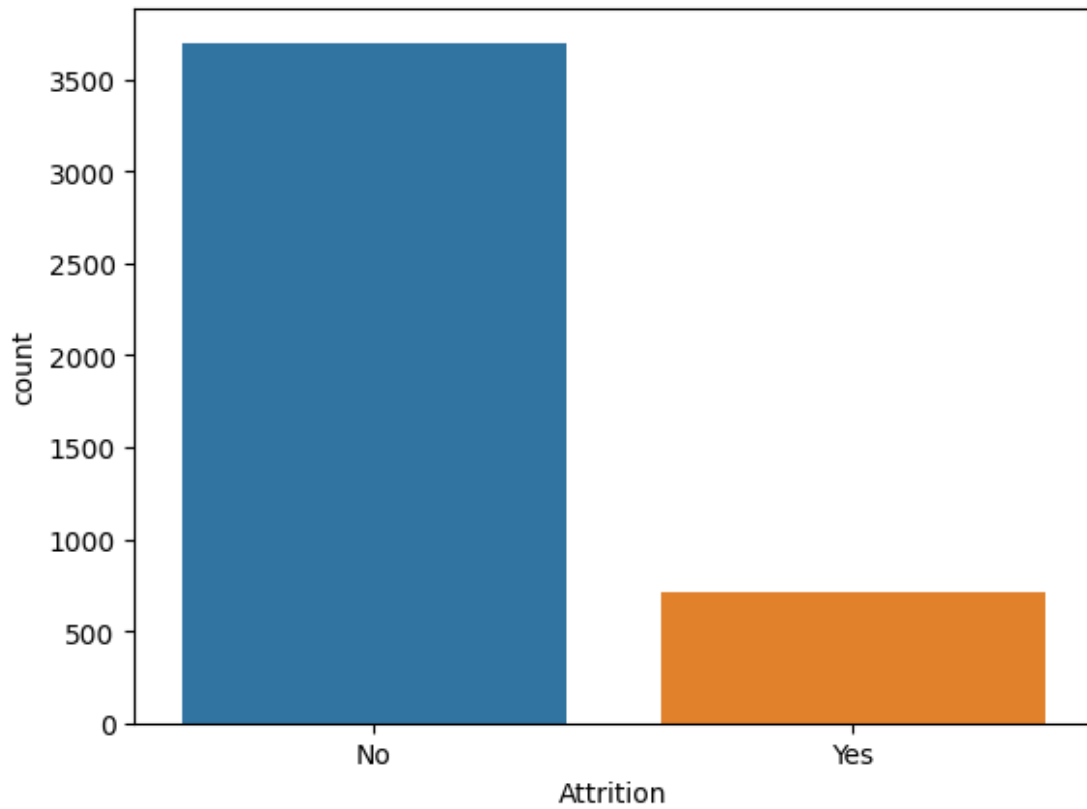
IMPACT OF BUSINESS TRAVEL ON ATTRITION

```
sns.countplot(hue=data.Attrition,x=data.BusinessTravel)
plt.show()
```
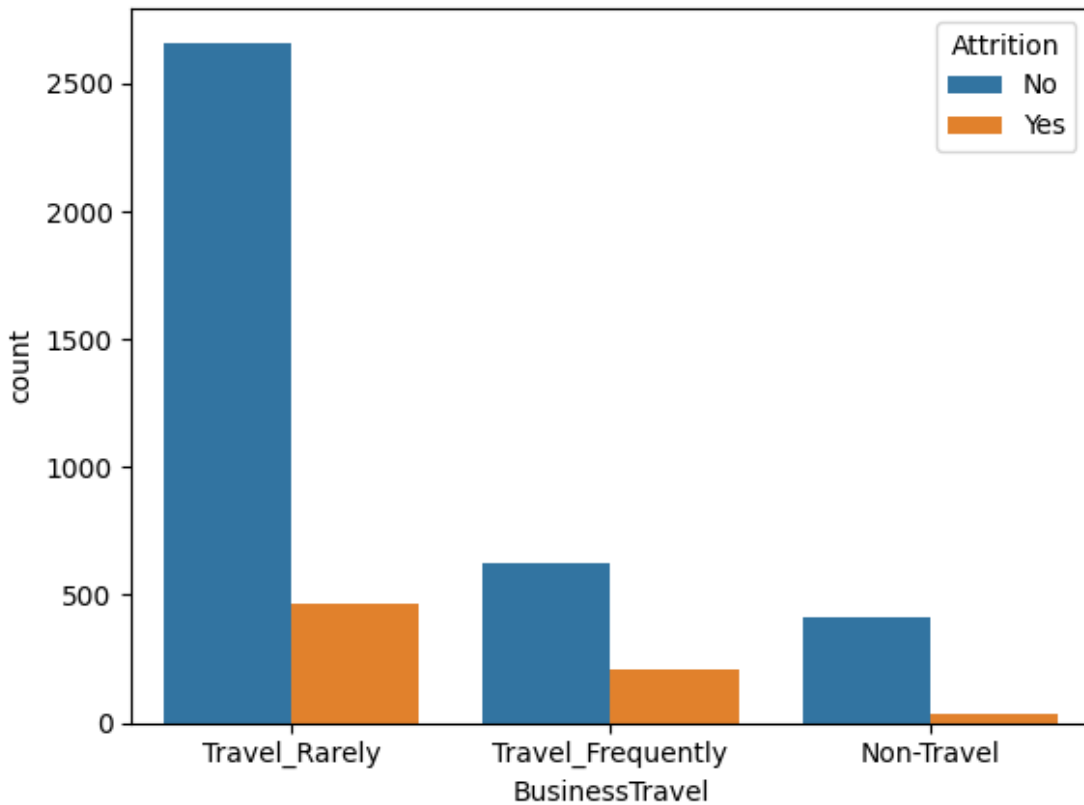
.Graph tell us that company has more count or more no.of employes who travels rarely.it means travels rate of company is less.

.There are more employes which travel rarely and are not satisfied with their job.

IMPACT OF DEPARTMENT ON ATTRITION

```
sns.countplot(hue=data.Attrition,x=data.Department)
plt.show()
```
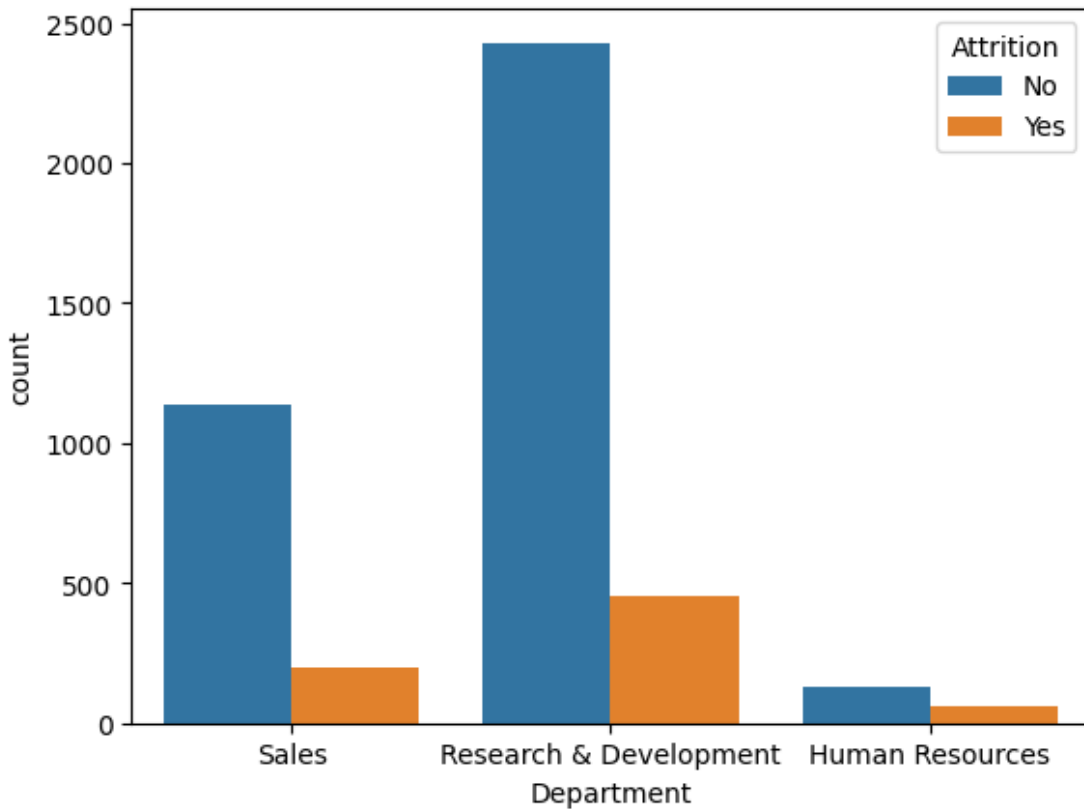
.Research & Development Department have more number of attrition (150 employees) as compared to other two department.

IMPACT OF EDUCATION FIELD ON ATTRITION

```
sns.countplot(x=data.Attrition,hue=data.EducationField)
plt.show()
```
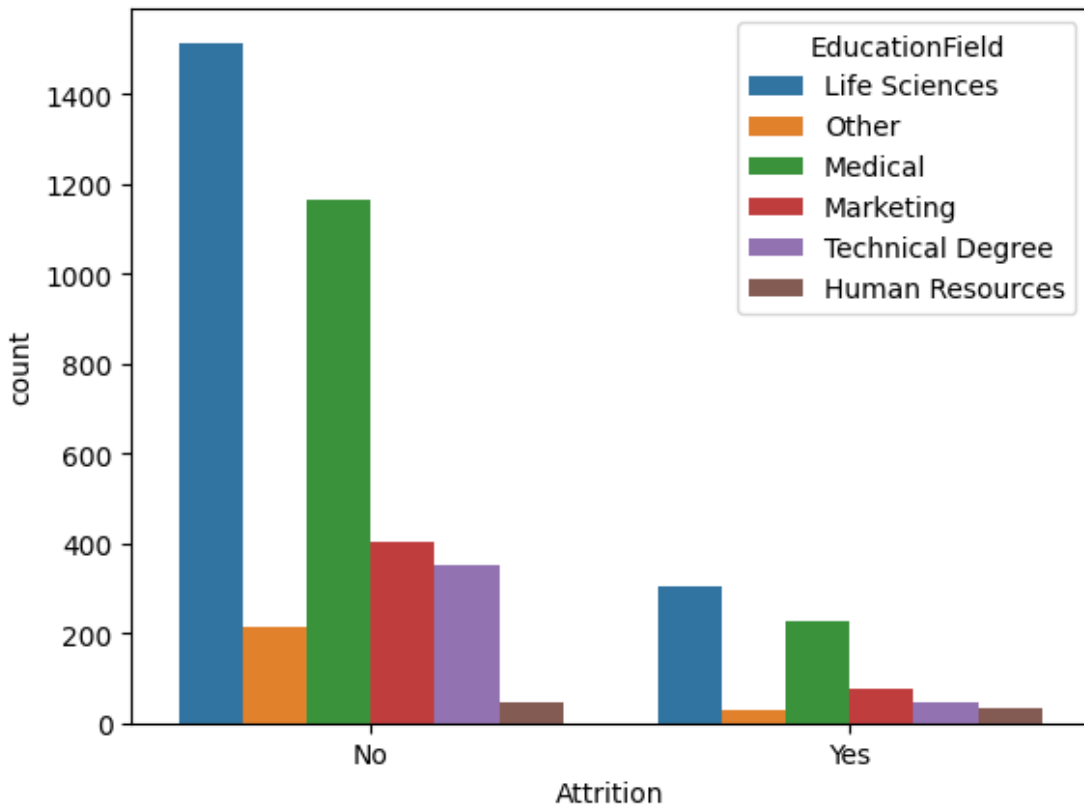
.First and foremost thing is that employee who are from "Life science & Medical" backgrounds are more as compared to other education field.

.Nearly 100 number of employees are there who are from Life science education background will leave the company and followed by Medical education Employees.

GENDER AND ATTRITION

```
sns.countplot(x=data.Attrition,hue=data.Gender)
plt.show()
```
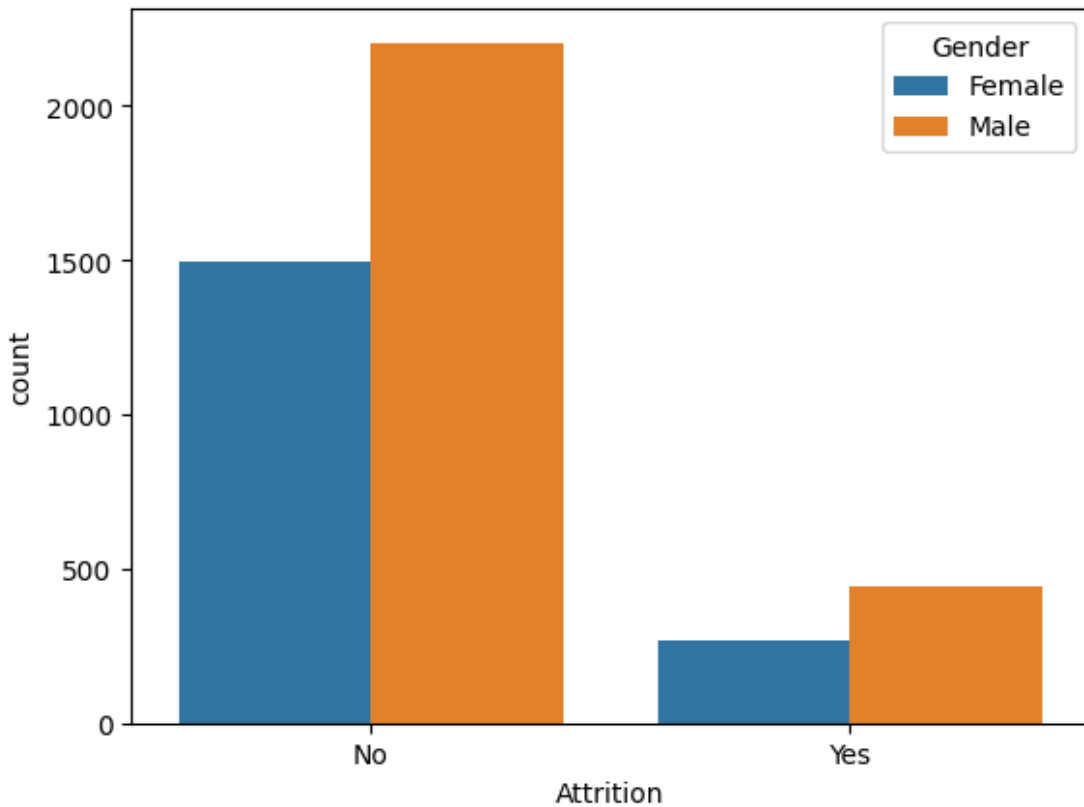
.Male employee are more as compared to females

.Males are more likely to quit the job rather then females.

DISTANCE FROM HOME AND ATTRITION

```
sns.histplot(hue=data.Attrition,x=data.DistanceFromHome)
plt.show()
```
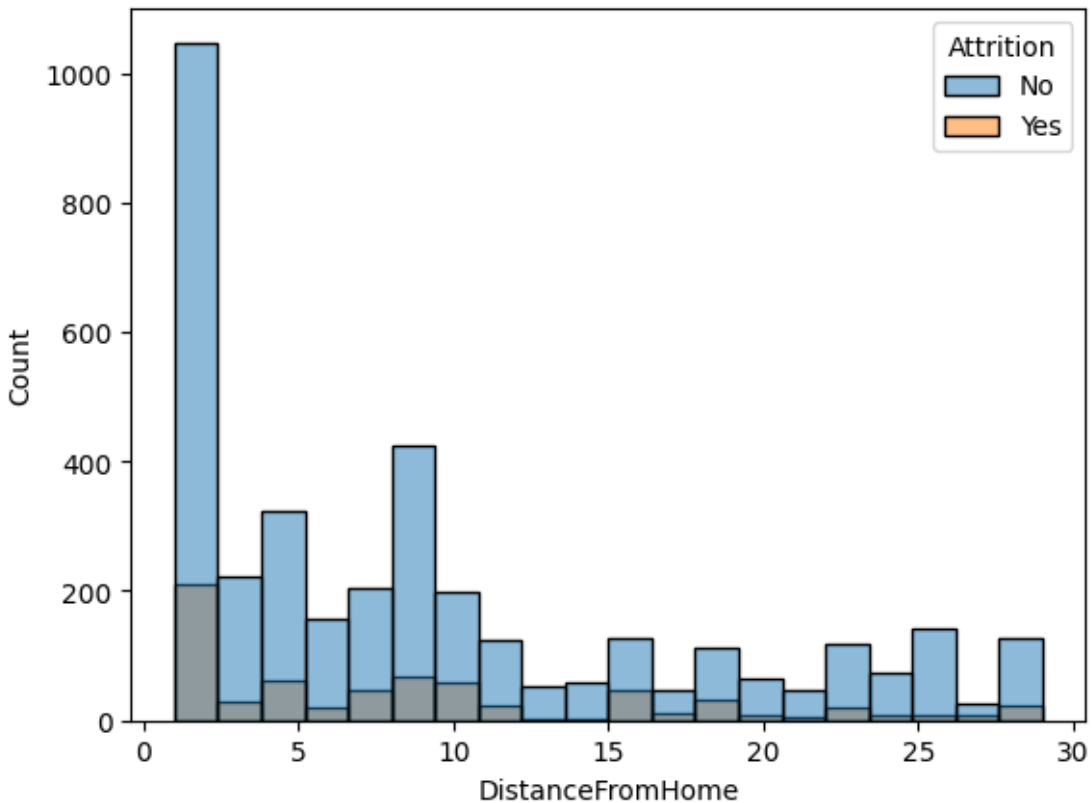
.Employees who has distance range of 0-10 km , are more likely to leave the job.

.We can also conclude that lesser the distance more number of employees are working.

IMPACT OF JOB ROLE ON ATTRITION

```
plt.figure(figsize=(20,10),facecolor="white")
sns.countplot(x="JobRole",hue='Attrition',data=data)
plt.xlabel('JobRole',fontsize=10)

Text(0.5, 0, 'JobRole')
```
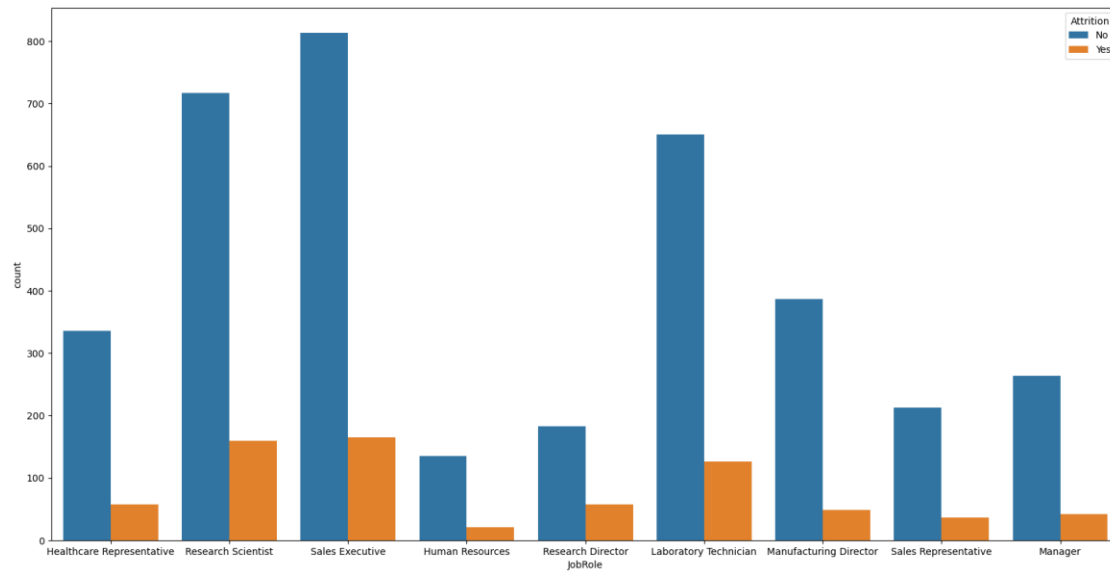
.There are less no.of Research Direactor who leaves the company.

.Laboratory technician,sales Exceutive and Research scientist are the top 3 job roles in which employees have their Attrition "yes".

.Apart from these it can also seen that there are more no.of employees in Sales Exceutive job role.

IMPACT OF NO. OF COMPANIES WORKED

```
sns.histplot(hue=data.Attrition,x=data.NumCompaniesWorked)
plt.show()
```
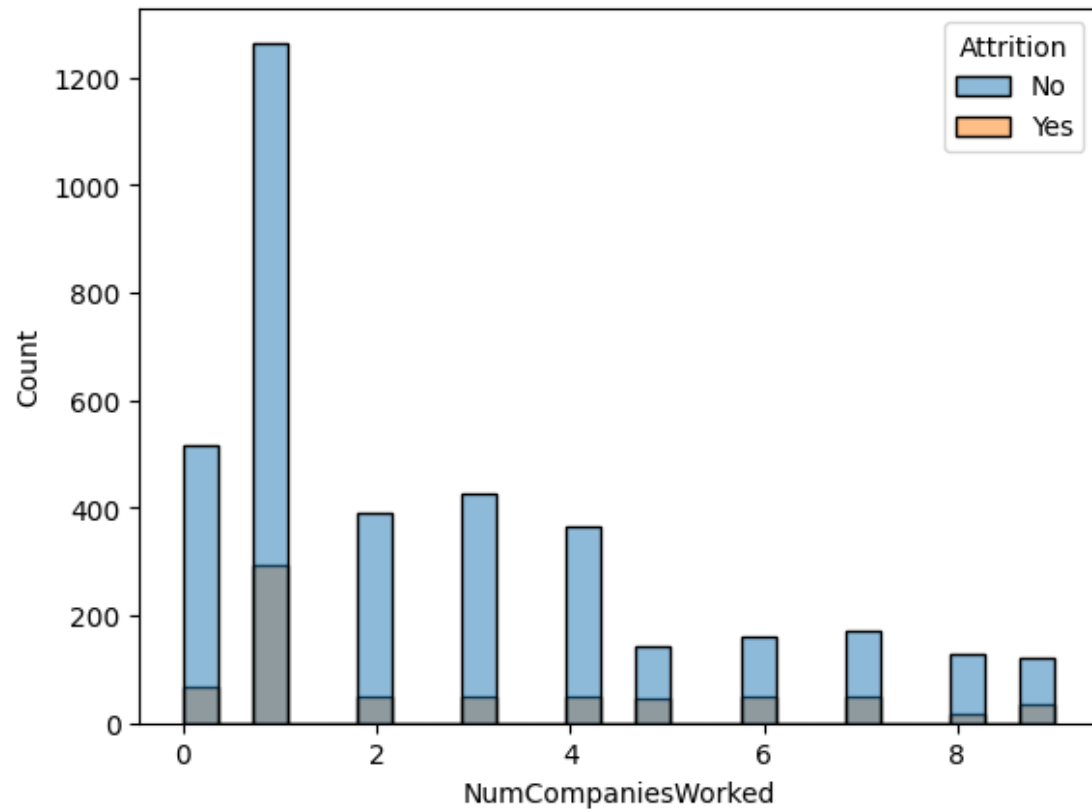
.Only that employees(no.of emp=100)who worked with 1 company before have most Attrition(Attrition "YES"),rest have similar data.

IMPACT OF MARITAL STATUS

```
sns.histplot(hue=data.Attrition,x=data.MaritalStatus)
plt.show()
```
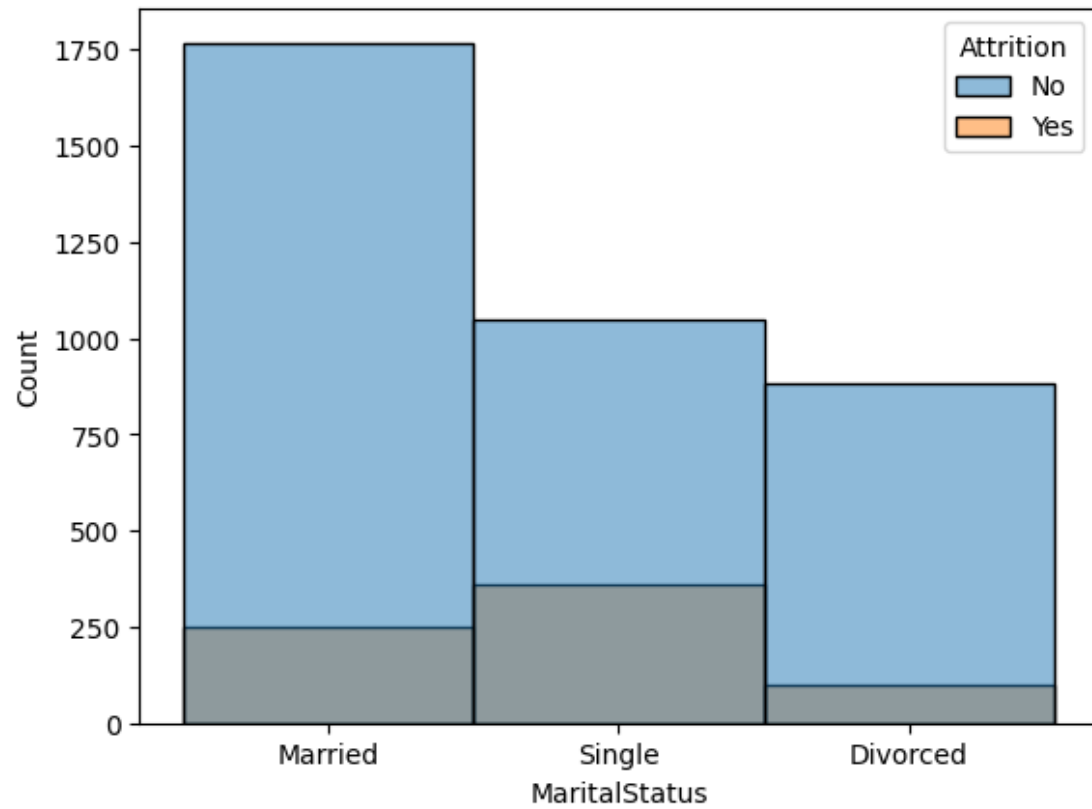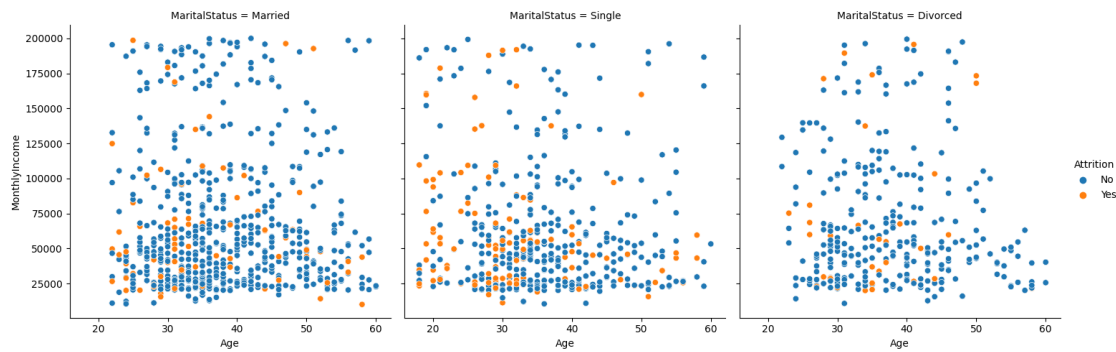
Can low salary be one of the reason?

```python
sns.violinplot(x='EducationField',y='MonthlyIncome',hue='Attrition',
             data=data,split=True)

plt.show()
```
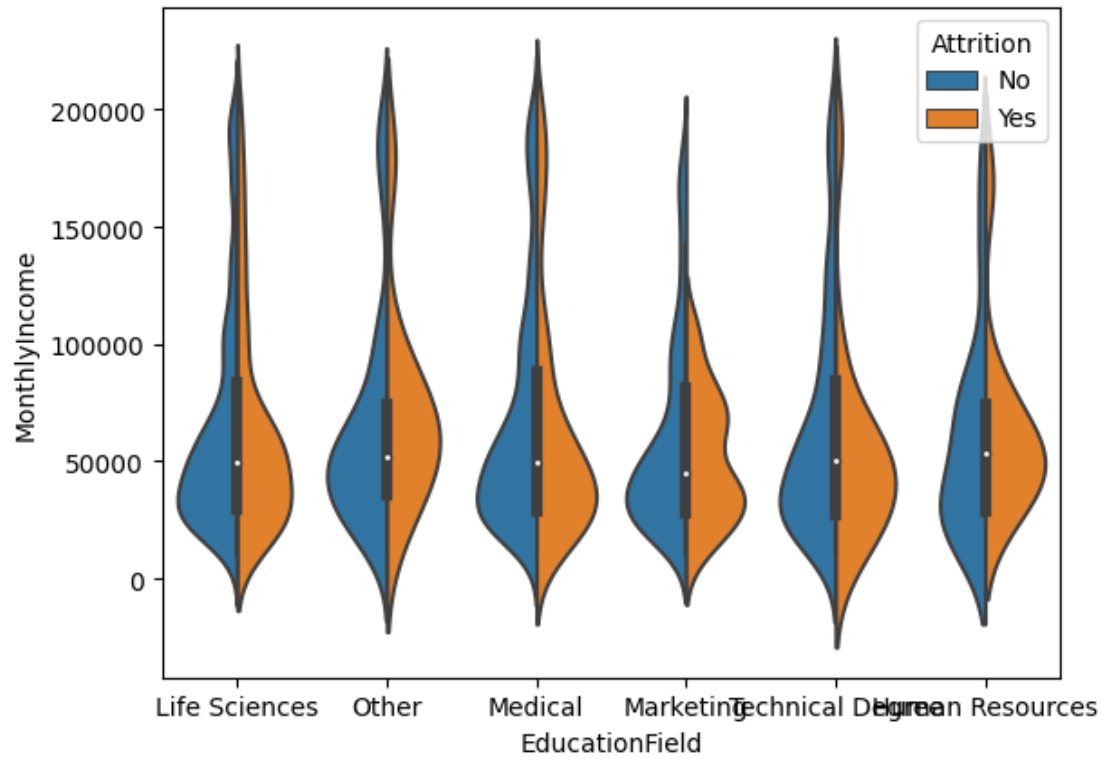
Subplot

```
sns.relplot(x='Age',y='MonthlyIncome',hue='Attrition',col='MaritalStatus',dat
a=data)
plt.show()
```

Who is leaving more? Married ,un-married or divorced employees?

```python
labels=['Married','Single','Divorced']
plt.pie(data.MaritalStatus.value_counts(),labels=labels,autopct='%.1f%%')
plt.show()
```