```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
```

```
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Emplo
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	

5 rows × 35 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Age                       1470 non-null   int64 
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64 
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64 
 6   Education                 1470 non-null   int64 
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64 
 9   EmployeeNumber            1470 non-null   int64 
 10  EnvironmentSatisfaction   1470 non-null   int64 
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64 
 13  JobInvolvement            1470 non-null   int64 
 14  JobLevel                  1470 non-null   int64 
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64 
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64 
 19  MonthlyRate               1470 non-null   int64 
 20  NumCompaniesWorked        1470 non-null   int64 
 21  Over18                    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 
 30  WorkLifeBalance           1470 non-null   int64 
 31  YearsAtCompany            1470 non-null   int64 
 32  YearsInCurrentRole        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
```

```
df.shape
```

```
(1470, 35)
```

```
df.Attrition.value_counts()
```

```
No     1233
Yes     237
Name: Attrition, dtype: int64
```

```
df.corr()
```

```
<ipython-input-7-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a
  df.corr()
```

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfac
Age	1.000000	0.010661	-0.001686	0.208034	NaN	-0.010145	0.01
DailyRate	0.010661	1.000000	-0.004985	-0.016806	NaN	-0.050990	0.01
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	NaN	0.032916	-0.01
Education	0.208034	-0.016806	0.021042	1.000000	NaN	0.042070	-0.02
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeNumber	-0.010145	-0.050990	0.032916	0.042070	NaN	1.000000	0.01
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	NaN	0.017621	1.00
HourlyRate	0.024287	0.023381	0.031131	0.016775	NaN	0.035179	-0.04
JobInvolvement	0.029820	0.046135	0.008783	0.042438	NaN	-0.006888	-0.00
JobLevel	0.509604	0.002966	0.005303	0.101589	NaN	-0.018519	0.00
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	NaN	-0.046247	-0.00
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	NaN	-0.014829	-0.00
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	NaN	0.012648	0.03
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	NaN	-0.001251	0.01
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	NaN	-0.012944	-0.03
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	NaN	-0.020359	-0.02
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	NaN	-0.069861	0.00
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	NaN	0.062227	0.00
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	NaN	-0.014365	-0.00
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	NaN	0.023603	-0.01
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	NaN	0.010309	0.02
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	NaN	-0.011240	0.00
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	NaN	-0.008416	0.01
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254	NaN	-0.009019	0.01
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	NaN	-0.009197	-0.00

26 rows × 26 columns

```
df.isnull().any()
```

```
Age                       False
Attrition                 False
BusinessTravel            False
DailyRate                 False
Department                False
DistanceFromHome          False
Education                 False
EducationField            False
EmployeeCount             False
EmployeeNumber            False
EnvironmentSatisfaction   False
Gender                    False
HourlyRate                False
JobInvolvement            False
JobLevel                  False
JobRole                   False
JobSatisfaction           False
MaritalStatus             False
MonthlyIncome             False
MonthlyRate               False
NumCompaniesWorked        False
Over18                    False
OverTime                  False
PercentSalaryHike         False
PerformanceRating         False
RelationshipSatisfaction  False
StandardHours             False
StockOptionLevel          False
TotalWorkingYears         False
TrainingTimesLastYear     False
```

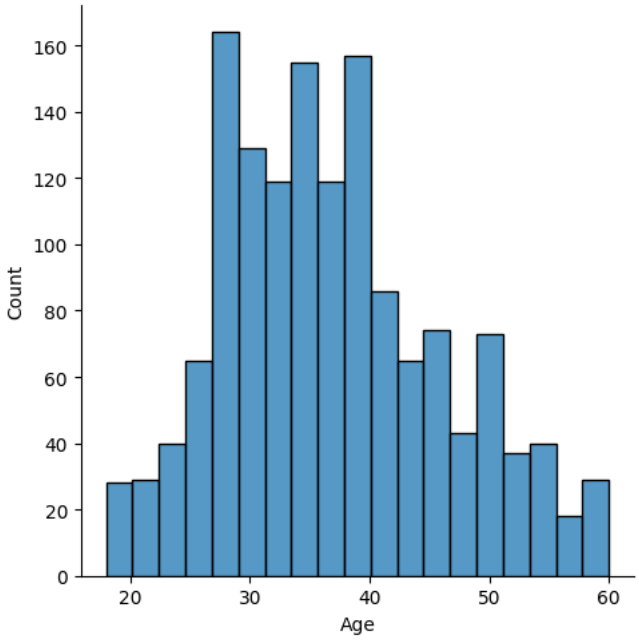
```
WorkLifeBalance             False
YearsAtCompany              False
YearsInCurrentRole          False
YearsSinceLastPromotion     False
YearsWithCurrManager        False
dtype: bool
```

```
df.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
Over18                      0
OverTime                    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
```

```
sns.displot(df["Age"])
```

```
<seaborn.axisgrid.FacetGrid at 0x7f005861f5b0>
```




```
plt.scatter(df['Attrition'],df['BusinessTravel'])
```

```
<matplotlib.collections.PathCollection at 0x7f0018a06320>
```


```
plt.scatter(df['Attrition'],df['DistanceFromHome'])
```

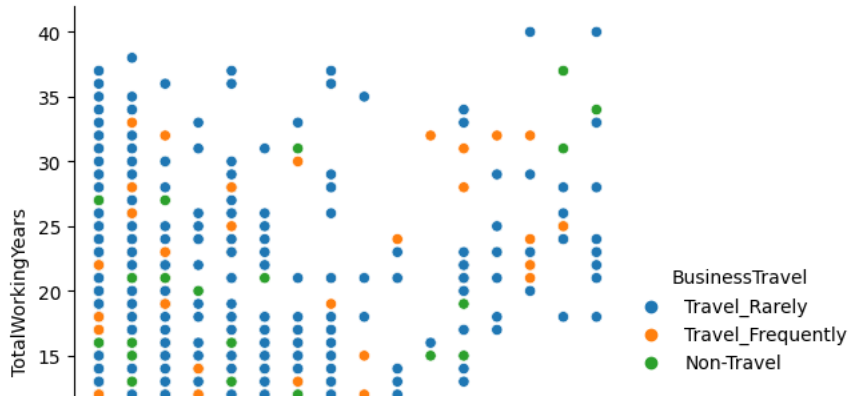
```
<matplotlib.collections.PathCollection at 0x7f001887cdf0>
```


```
plt.scatter(df['Attrition'],df['StandardHours'])
```

```
<matplotlib.collections.PathCollection at 0x7f00188d7310>
```

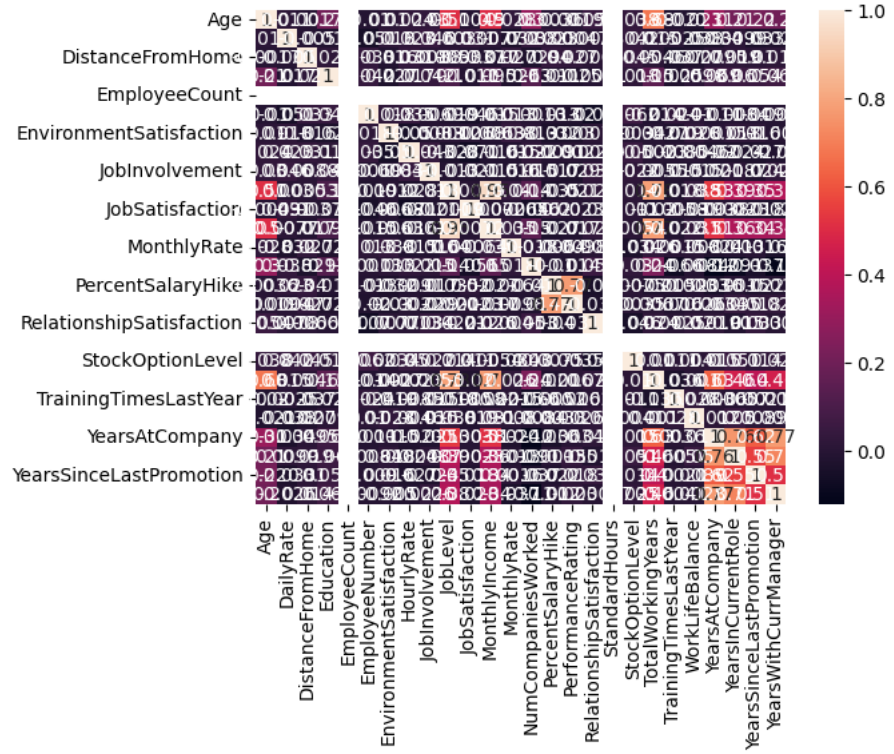

```
sns.relplot(x="YearsSinceLastPromotion",y="TotalWorkingYears",data=df,hue="BusinessTravel")
```

```
<seaborn.axisgrid.FacetGrid at 0x7f0018a47700>
```



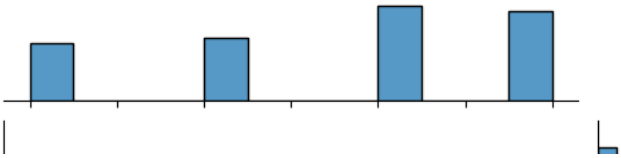

```
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-15-8df7bcac526d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a
  sns.heatmap(df.corr(),annot=True)
<Axes: >
```



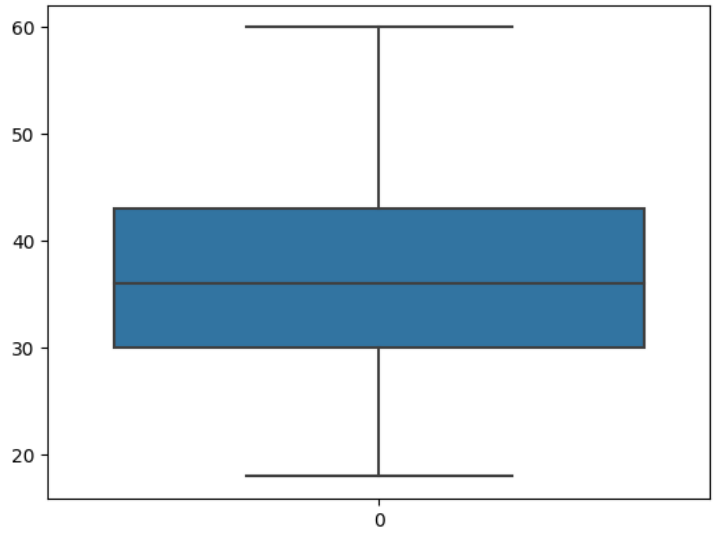

```
sns.jointplot(x="RelationshipSatisfaction",y="Attrition",data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7f00187e6d40>
```



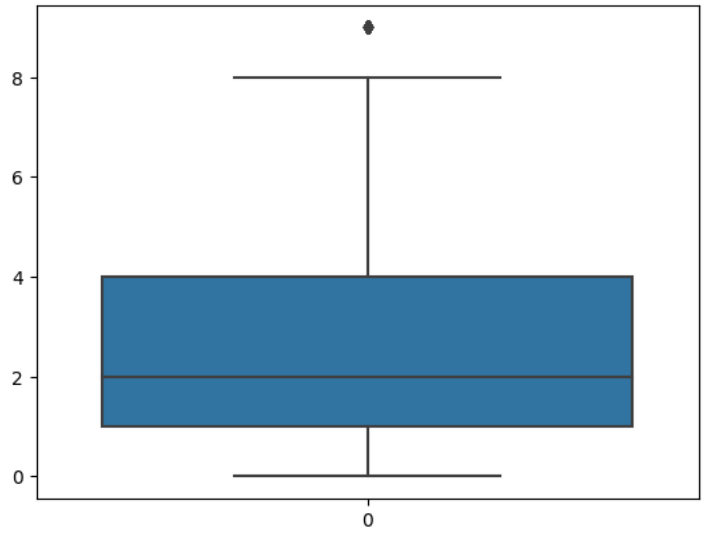
```
sns.boxplot(df.Age)
```

```
<Axes: >
```




```
sns.boxplot(df.NumCompaniesWorked)
```

```
<Axes: >
```




```
q1=df.NumCompaniesWorked.quantile(0.25)
q3=df.NumCompaniesWorked.quantile(0.75)
```

```
print(q1)
print(q3)
```

```
1.0
4.0
```

```
IQR=q3-q1
```

```
IQR
```

```
3.0
```

```
upper_limit=q3+1.5*IQR
upper_limit
```

```
8.5
```

```
df=df[df.NumCompaniesWorked<upper_limit]
```

```
sns.boxplot(df.NumCompaniesWorked)
```

```
<Axes: >
```