```
In [2]:
           import pandas as pd
           import numpy as np
In [3]:
           data=pd.read_csv("HR-Employee-Attrition.csv")
In [4]:
           data.head()
                  Attrition
                               BusinessTravel DailyRate
                                                          Department DistanceFromHome Education EducationFie
Out[4]:
             Age
               41
                        Yes
                                 Travel_Rarely
                                                   1102
                                                                 Sales
                                                                                                    2
                                                                                                          Life Scienc
                                                           Research &
          1
               49
                        No
                             Travel_Frequently
                                                    279
                                                                                         8
                                                                                                    1
                                                                                                          Life Scienc
                                                         Development
                                                           Research &
          2
               37
                        Yes
                                 Travel_Rarely
                                                   1373
                                                                                         2
                                                                                                    2
                                                                                                                Oth
                                                         Development
                                                           Research &
                             Travel_Frequently
          3
               33
                        No
                                                   1392
                                                                                         3
                                                                                                    4
                                                                                                          Life Scienc
                                                         Development
                                                           Research &
                                                    591
                                                                                         2
               27
                        No
                                 Travel_Rarely
                                                                                                    1
                                                                                                              Medic
                                                         Development
         5 rows × 35 columns
In [5]:
           data.tail()
Out[5]:
                 Age
                     Attrition
                                  BusinessTravel
                                                 DailyRate
                                                             Department DistanceFromHome Education Education
                                                               Research &
                                Travel_Frequently
                                                                                                        2
          1465
                  36
                            No
                                                       884
                                                                                           23
                                                                                                                  Μ
                                                             Development
                                                               Research &
          1466
                  39
                            No
                                    Travel_Rarely
                                                       613
                                                                                            6
                                                                                                        1
                                                                                                                  Μ
                                                             Development
                                                               Research &
          1467
                  27
                            No
                                    Travel_Rarely
                                                       155
                                                                                            4
                                                                                                        3
                                                                                                             Life Sc
                                                             Development
          1468
                            No
                                Travel_Frequently
                                                       1023
                                                                    Sales
                                                                                                        3
                                                                                                                  Μ
                                                               Research &
          1469
                                                       628
                                                                                            8
                                                                                                        3
                  34
                            No
                                    Travel_Rarely
                                                                                                                  Μ
                                                             Development
         5 rows × 35 columns
In [6]:
           data.isnull()
                      Attrition
                                 BusinessTravel
                                                 DailyRate Department DistanceFromHome
                                                                                             Education
                                                                                                         EducationF
Out[6]:
                 Age
             0 False
                                                                                                   False
                           False
                                          False
                                                     False
                                                                   False
                                                                                       False
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
1	False	False	False	False	False	False	False	ŀ
2	False	False	False	False	False	False	False	ŀ
3	False	False	False	False	False	False	False	ŀ
4	False	False	False	False	False	False	False	ŀ
•••								
1465	False	False	False	False	False	False	False	I
1466	False	False	False	False	False	False	False	I
1467	False	False	False	False	False	False	False	I
1468	False	False	False	False	False	False	False	ŀ
1469	False	False	False	False	False	False	False	ŀ

1470 rows × 35 columns

```
In [8]:
         data.isnull().sum()
                                     0
        Age
Out[8]:
        Attrition
                                     0
        BusinessTravel
                                     0
        DailyRate
                                     0
        Department
                                     0
        DistanceFromHome
                                     0
        Education
                                     0
        EducationField
                                     0
        EmployeeCount
        EmployeeNumber
                                     0
        EnvironmentSatisfaction
                                     0
        Gender
        HourlyRate
                                     0
        JobInvolvement
                                     0
        JobLevel
                                     0
        JobRole
                                     0
        JobSatisfaction
                                     0
        MaritalStatus
                                     0
        MonthlyIncome
                                     0
        MonthlyRate
                                     0
        NumCompaniesWorked
                                     0
        Over18
        OverTime
                                     0
        PercentSalaryHike
        PerformanceRating
        RelationshipSatisfaction
                                     0
        StandardHours
                                     0
        StockOptionLevel
                                     0
        TotalWorkingYears
                                     0
        TrainingTimesLastYear
                                     0
        WorkLifeBalance
                                     0
        YearsAtCompany
                                     0
```

0

YearsInCurrentRole

```
YearsSinceLastPromotion 0
YearsWithCurrManager 0
```

dtype: int64

```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(15,8))
sns.heatmap(data.corr(), annot=True, cmap="coolwarm")
```

Out[7]: <AxesSubplot:>

```
1.0
                                1 0.01-D.001 0.21
                                                               0.01 0.010.0240.03 <mark>0.51</mark>0.004<mark> 0.5 </mark>0.028 0.30.00360019.05
                                                                                                                                               .038<mark>0.68</mark> -0.02-0.0210.31 0.21 0.22 0.2
                         Age
                                                               .05 D.0180.0230.0460.0030.03 D.00740.03 20.0380.028000 47007
.0330.0160.03 D.00808.00508.00340.0170.0270.0290.040.0270.006
                                                                                                                                               .0420.0150.00250.0380.034.00940.0330.02
                 DailvRate
      DistanceFromHome
                                                                                                                                               .049.0040.0370.020.0095.0190.010.01
                                 001-0.005 1 0.02
                 Education -0.21-0.0170.021 1
                                                                0420.0270.0170.042 0.1 -0.0110.0950.0260.13-0.0120.0240.009
                                                                                                                                              0.0180.15-0.025.0098.0690.060.0540.0
          EmployeeCount
                                                                                                                                                                                                               0.8
        EmployeeNumber
                               0.010.0180.0160.02
                                                                    1 -0.050.0080800102.00608.00608.0380.0130.0320.030.007
                                                                                                                                               0034.0027.0190.028.0015.0180.0160.00
EnvironmentSatisfaction
                                                               .035-0.05 1 0.0430.0280.0740.0160.0150.0240.00901.00202001
               HourlyRate -0.0240.0230.0310.01
                                                                                                                                              0.050.002030080500460.020.0240.0270.02
                                                               Jobinvolvement
                   JobLevel -0.51 0.00B.0053 0.1
                                                                                                                                              0.014<mark>0.78-</mark>0.0180.038<mark>0.53 0.39 0.35 0.38</mark>
                                                                                                                                                                                                              - 0.6
                                                                046.0068.07<u>1</u>0.021<mark>91.4</mark>0.000200064.0560.020.00230.01
                                                                                                                                              0.011-0.020.0058.019.00308002-8.0180.02
           JobSatisfaction → .0049.03-D.0030.01
                                                               .015.0068.0160.015<mark>0.950</mark>.007<mark>2 1 0.035 0.15</mark>-0.0270.0170.026
.0130.0380.0150.0160.04.000644035 1 0.018.00644098004
                                                                                                                                              005<mark>40.77-</mark>0.0220.031<mark>0.51 0.36 0.34 0.34</mark>
0.034.026.0015.0080.0240.018.0016.03
           MonthlyIncome
              MonthlyRate -0.0280.0320.0270.02
                                                               0010.0130.0220.015 0.14 0.05 (0.15 0.018 1 -0.010.0140.05
  NumCompaniesWorked - 0.3 0.0380.0290.13
                                                                                                                                              0.03 <mark>0.24</mark> 0.066.00840.120.0910.0370.11
                                                                                                                                                                                                               0.4
       PercentSalaryHike -
                                                                0130.030.0090.0170.0350.02-0.02-0.00640.01 1 0.7
      PerformanceRating -.0009000407.0270.02
                                                               0.02-0.030.00242.0290.020.00240.0140.00948.0140.7
                                                               0.070.0070.001B.0340.0220.0120.0240.0040.053-0.040.031 1
 RelationshipSatisfaction -
                                .0540.00708006/6.00
                                                                                                                                              0.0460.0240.00250.020.0190.0150.0308.00
           StandardHours
                                                                 06D.00340.05 0.0220.0140.01D.00540.0340.030.0075.00345.04
                                                                                                                                                   0.01 0.01 D.004 D.0150.05 10.0140.02
         StockOptionLevel -0.0380.0420.0450.01
                                                                                                                                                                                                               0.2
        TotalWorkingYears -0.68 0.015.004 0.15
                                                               1 -0.0360.001 0.63 0.46 0.4 0.46
                                                                                                                                              0.011<mark>0.036 1 0.028.0036.00507.0021.00</mark>
   TrainingTimesLastYear
                                                               0.010.0280.0046.0150.0380.0150.0310.0080.008040030800260.02
         WorkLifeBalance -0.0210.0380.020.00
                                                                                                                                              0.015<mark>0.630.0036.012 1 0.76 0.62 0</mark>
                                                                .010.00150.02-0.021<mark>0.53</mark>).003 <mark>0.51-</mark>0.0240.12-0.036.0034.019
         YearsAtCompany -0.31 0.034 0095.06
                                                                                                                                                                                                               0.0
       YearsInCurrentRole -0.210.0099.0190.0
                                                                0.390.024.008<mark>0.39</mark>0
                                                                                              .002 <mark>0.36</mark> 0.0130.090.0016.0350.01
                                                                                                                                                   0.460.00570.05 0.7
YearsSinceLastPromotion -0.22-0.0330.010.05
                                                                0090.0160.0270.024<mark>0.35</mark>-0.018<mark>0.34</mark>0.00140.0370.0220.0180.03
                                                                                                                                                   4 0.4 0.00201008 0.62 0.55 1 0.51
                                                                                                                                               .0250.460.004100280.77 0.71 0.51
  YearsWithCurrManager - 0.2 -0.0260.0140.06
                                                                 090.0050.020.026<mark>0.38</mark>0.028<mark>0.34</mark>0.0370.110.010.028.000
                                                                                                                                                     Total Working Years
                                                         EmployeeCount
                                                               EmployeeNumber
                                                                    EnvironmentSatisfaction
                                                                           HourlyRate
                                                                                 oblnvolvement
                                                                                       JobLevel
                                                                                              obSatisfaction
                                                                                                          MonthlyRate
                                                                                                                NumCompaniesWorked
                                                                                                                      PercentSalaryHike
                                                                                                                                   RelationshipSatisfaction
                                                                                                                                         StandardHours
                                                                                                                                              StockOptionLevel
                                                                                                                                                           FainingTimesLastYear
                                                                                                                                                                 WorkLifeBalance
                                                                                                                                                                              YearsInCurrentRole
                                                                                                                                                                                    EarsSinceLastPromotion
                                             DistanceFromHome
                                                  Education
                                                                                                                            PerformanceRating
```

In [8]: data.dtypes

 $\begin{array}{c} \text{Out[8]:} & \text{Age} & \text{int64} \\ \text{Attrition} & \text{object} \\ \text{BusinessTravel} & \text{object} \end{array}$

DailyRate int64
Department object
DistanceFromHome int64
Education int64
EducationField object

EmployeeCount int64
EmployeeNumber int64
EnvironmentSatisfaction int64

Gender object
HourlyRate int64
JobInvolvement int64
JobLevel int64

JobRole object
JobSatisfaction int64
MaritalStatus object
MonthlyIncome int64

MonthlyRate int64 NumCompaniesWorked int64 object Over18 OverTime object PercentSalaryHike int64 PerformanceRating int64 RelationshipSatisfaction int64 StandardHours int64 StockOptionLevel int64 TotalWorkingYears int64 TrainingTimesLastYear int64 WorkLifeBalance int64 YearsAtCompany int64 YearsInCurrentRole int64 YearsSinceLastPromotion int64 YearsWithCurrManager int64 dtype: object

In [9]:

data[data.duplicated()]

 $\verb"Out[9]: \textbf{Age Attrition BusinessTravel DailyRate Department DistanceFromHome Education EducationField}$

0 rows × 35 columns

→

In [10]:

data.dropna()

Out[10]:		Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educatio
	0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sc
	1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sc
	2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
	3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sc
	4	27	No	Travel_Rarely	591	Research & Development	2	1	М
	•••								
	1465	36	No	Travel_Frequently	884	Research & Development	23	2	М
	1466	39	No	Travel_Rarely	613	Research & Development	6	1	М
	1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sc
	1468	49	No	Travel_Frequently	1023	Sales	2	3	М
	1469	34	No	Travel_Rarely	628	Research & Development	8	3	М

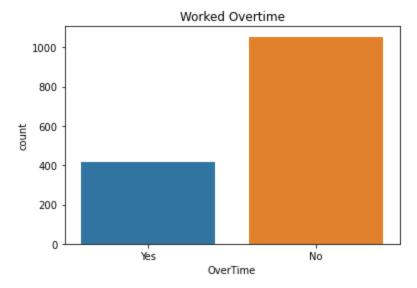
1470 rows × 35 columns

In [11]: data.drop_duplicates()

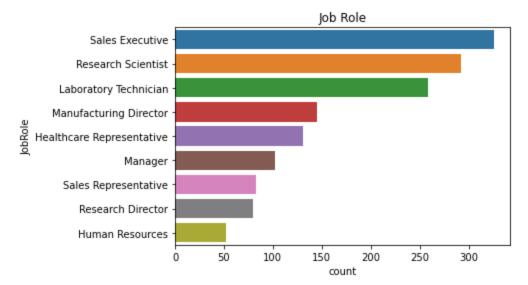
Out[11]:		Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educatio
	0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sc
	1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sc
	2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
	3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sc
	4	27	No	Travel_Rarely	591	Research & Development	2	1	Μ
	•••								
1	465	36	No	Travel_Frequently	884	Research & Development	23	2	M
1	466	39	No	Travel_Rarely	613	Research & Development	6	1	M
1	467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sc
1	468	49	No	Travel_Frequently	1023	Sales	2	3	М
1	469	34	No	Travel_Rarely	628	Research & Development	8	3	М

1470 rows × 35 columns

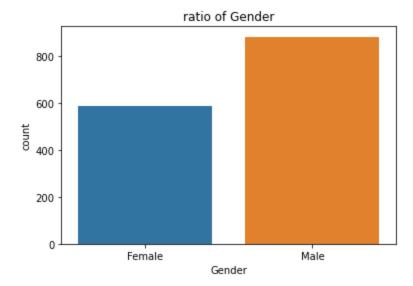
```
In [12]:
    sns.countplot(data=data, x='OverTime')
    plt.title('Worked Overtime')
    plt.show()
```



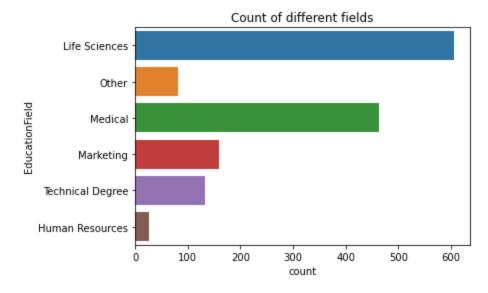
```
In [13]:
    sns.countplot(data=data,y='JobRole')
    plt.title('Job Role')
    plt.show()
```



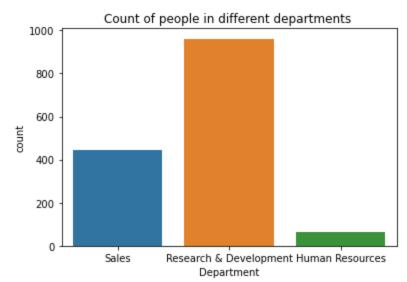
```
In [14]:
    sns.countplot(data=data,x='Gender')
    plt.title('ratio of Gender')
    plt.show()
```



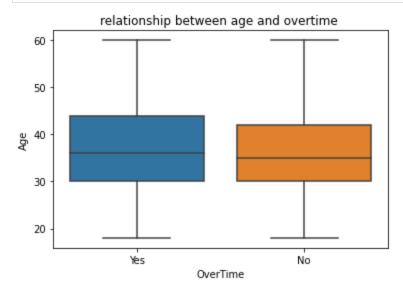
```
In [15]:
    sns.countplot(data=data,y='EducationField')
    plt.title('Count of different fields')
    plt.show()
```



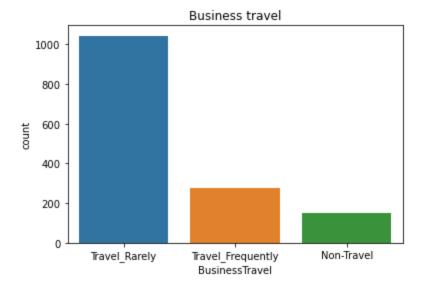
```
sns.countplot(data=data,x='Department')
plt.title('Count of people in different departments ')
plt.show()
```



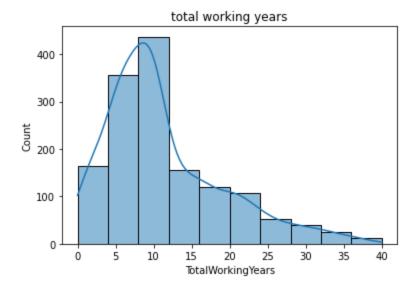
```
sns.boxplot(data=data,x='OverTime',y='Age')
plt.title('relationship between age and overtime ')
plt.show()
```



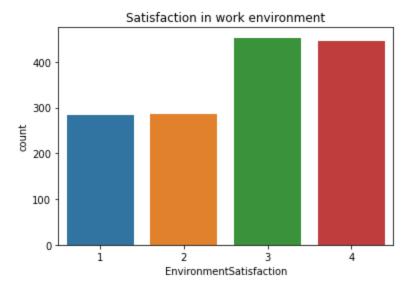
```
sns.countplot(data=data,x='BusinessTravel')
plt.title('Business travel')
plt.show()
```



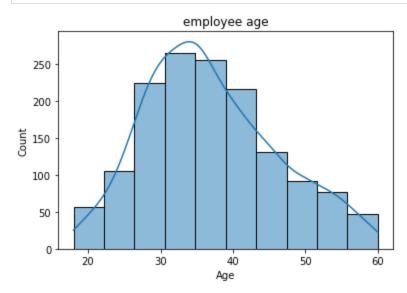
```
sns.histplot(data=data,x='TotalWorkingYears',bins=10,kde=True)
plt.title('total working years')
plt.show()
```



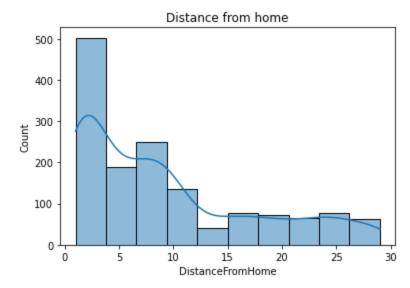
```
In [25]:
    sns.countplot(data=data,x='EnvironmentSatisfaction')
    plt.title('Satisfaction in work environment')
    plt.show()
```



```
In [28]:
    sns.histplot(data=data,x='Age',bins=10,kde=True)
    plt.title('employee age')
    plt.show()
```



```
In [34]:
    sns.histplot(data=data,x='DistanceFromHome',bins=10,kde=True)
    plt.title('Distance from home')
    plt.show()
```



In []: