In [2]:
```
import pandas as pd
import numpy as np
```

In [3]:
```
data=pd.read_csv("HR-Employee-Attrition.csv")
```

In [4]:
```
data.head()
```

Out[4]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationFie
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Scienc
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Scienc
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Oth
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Scienc
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medic

5 rows × 35 columns

In [5]:
```
data.tail()
```

Out[5]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educatio
1465	36	No	Travel_Frequently	884	Research & Development	23	2	M
1466	39	No	Travel_Rarely	613	Research & Development	6	1	M
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sc
1468	49	No	Travel_Frequently	1023	Sales	2	3	M
1469	34	No	Travel_Rarely	628	Research & Development	8	3	M

5 rows × 35 columns

In [6]:
```
data.isnull()
```

Out[6]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationF
0	False	False	False	False	False	False	False	F

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationF
1	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	
1465	False	False	False	False	False	False	False	F
1466	False	False	False	False	False	False	False	F
1467	False	False	False	False	False	False	False	F
1468	False	False	False	False	False	False	False	F
1469	False	False	False	False	False	False	False	F

1470 rows × 35 columns

In [8]:
```
data.isnull().sum()
```

Out[8]:
```
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
```

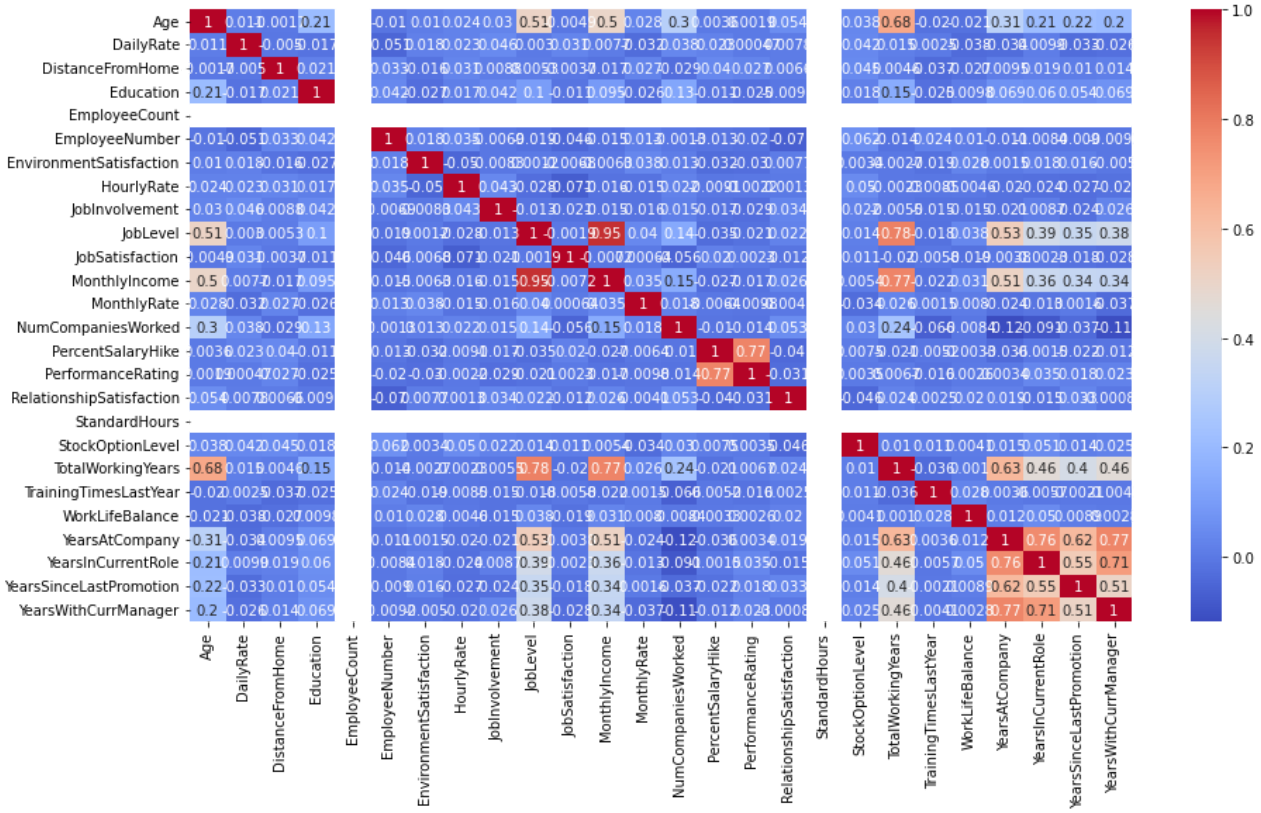
```
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

In [7]:

```
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:>
```



In [8]:

```
data.dtypes
```

Out[8]:

```
Age                          int64
Attrition                   object
BusinessTravel              object
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
Over18                    object
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()]
```

Out[9]:

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	M
...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	M
1466	39	No	Travel_Rarely	613	Research & Development	6	1	M
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sc
1468	49	No	Travel_Frequently	1023	Sales	2	3	M
1469	34	No	Travel_Rarely	628	Research & Development	8	3	M

1470 rows × 35 columns

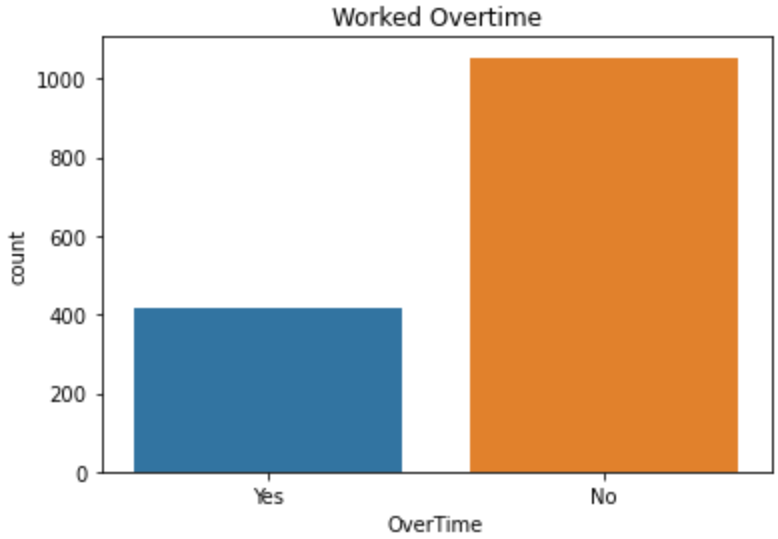
In [11]:
```
data.drop_duplicates()
```

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	M
...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	M
1466	39	No	Travel_Rarely	613	Research & Development	6	1	M
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sc
1468	49	No	Travel_Frequently	1023	Sales	2	3	M
1469	34	No	Travel_Rarely	628	Research & Development	8	3	M

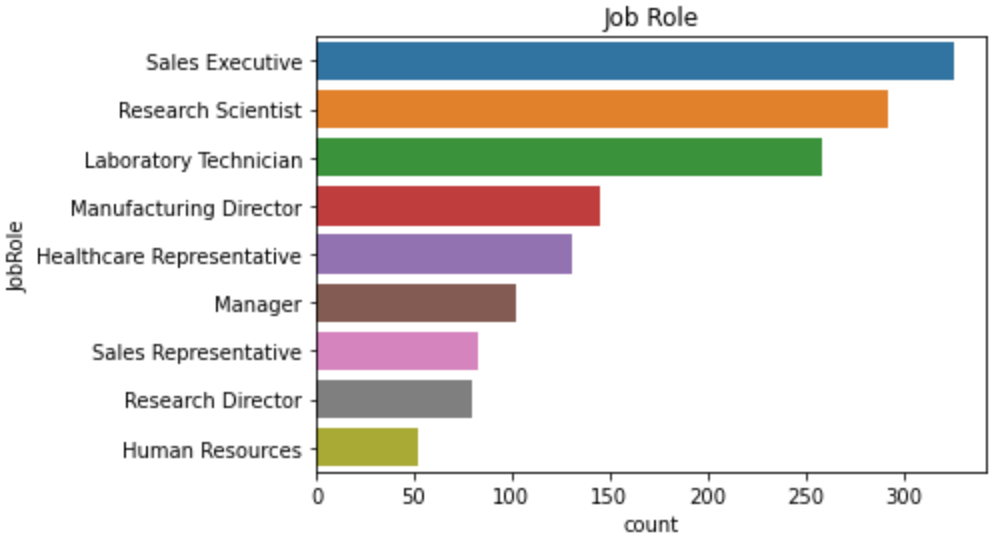
1470 rows × 35 columns

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



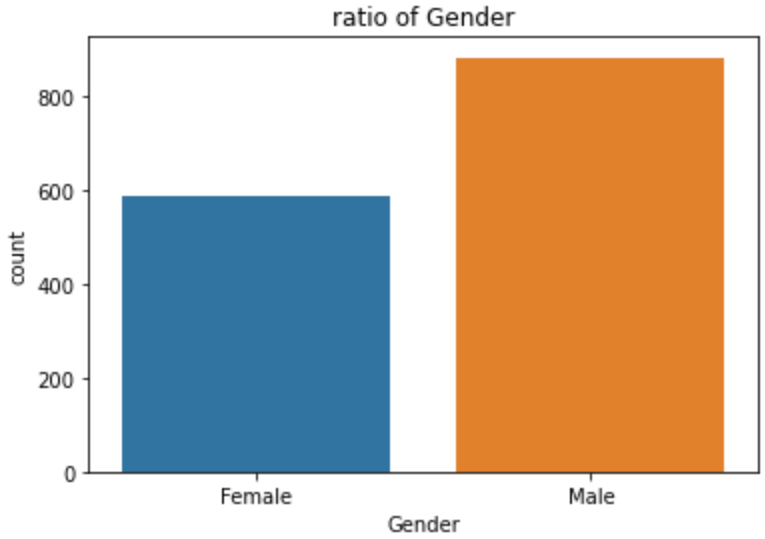
In [13]:

```
sns.countplot(data=data,y='JobRole')
plt.title('Job Role')
plt.show()
```



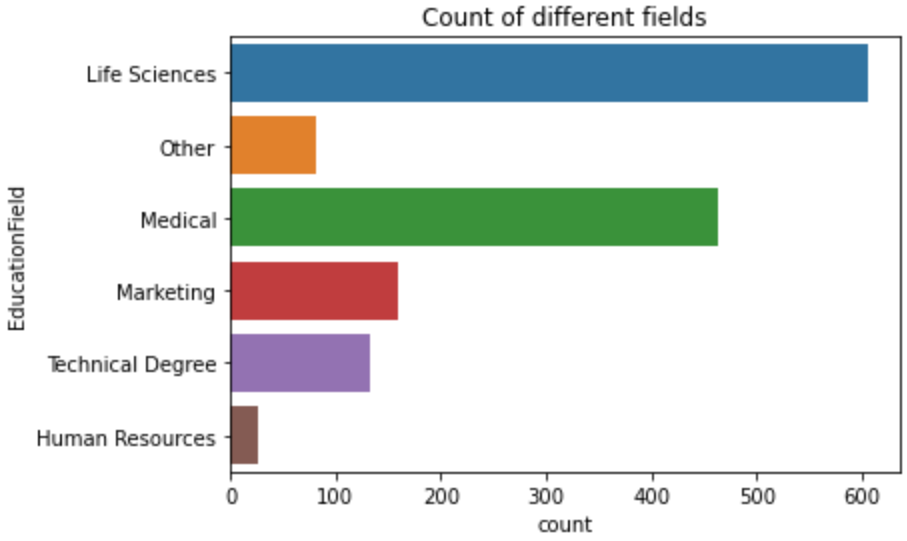
In [14]:

```
sns.countplot(data=data,x='Gender')
plt.title('ratio of Gender')
plt.show()
```



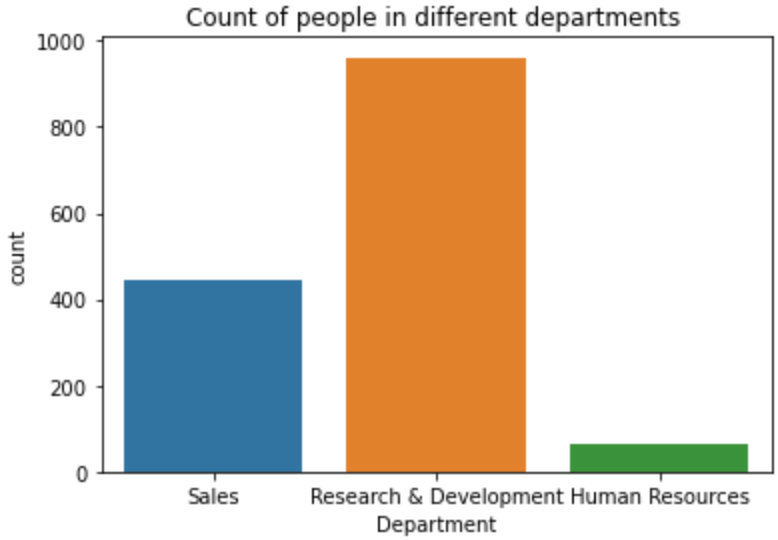
In [15]:

```
sns.countplot(data=data,y='EducationField')
plt.title('Count of different fields')
plt.show()
```



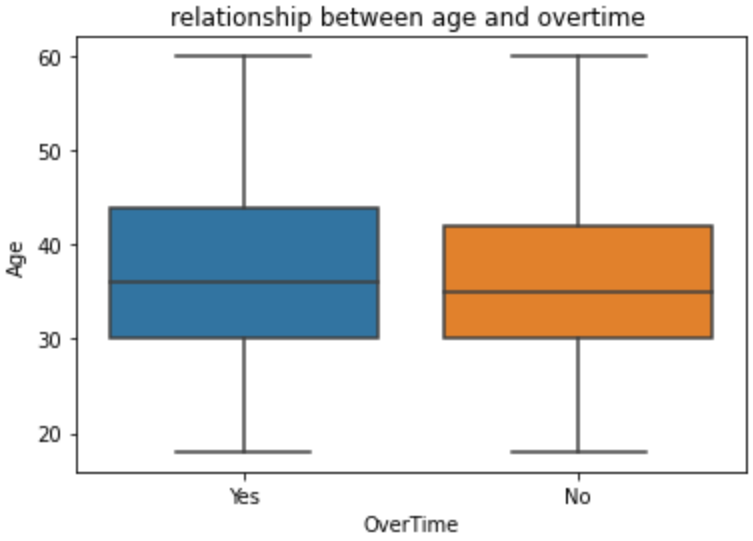
In [17]:

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



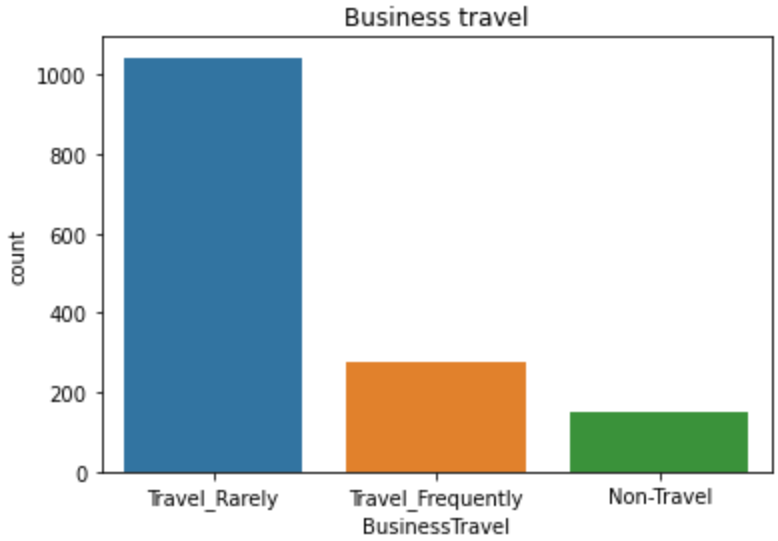
In [19]:

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



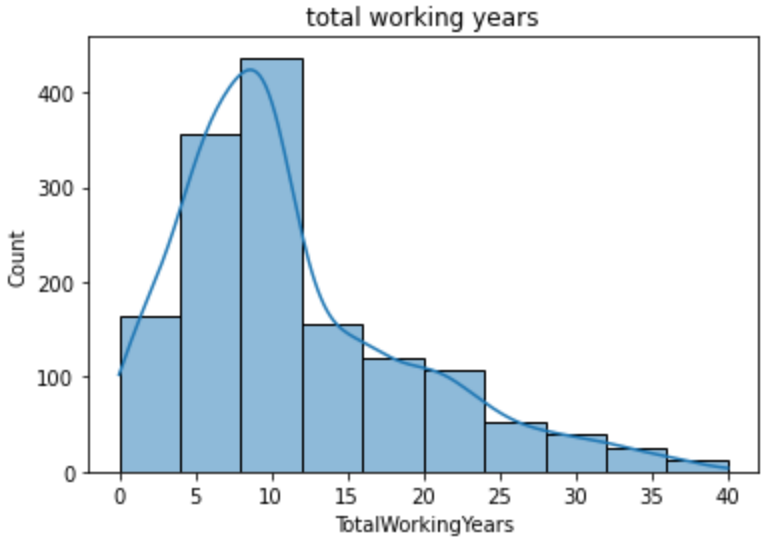
In [20]:

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



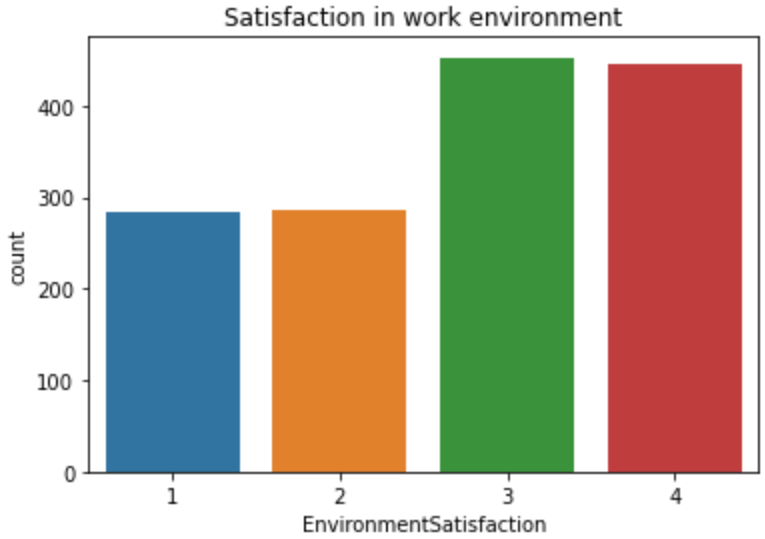
In [23]:

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



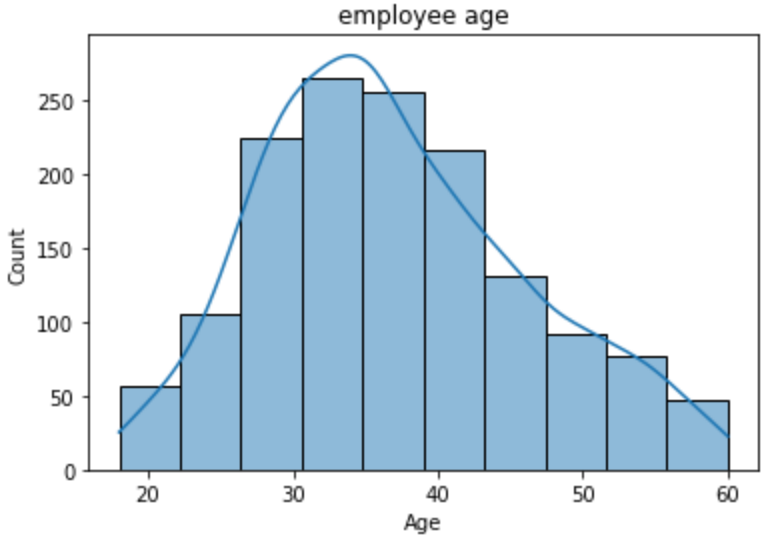
In [25]:

```
sns.countplot(data=data,x='EnvironmentSatisfaction')
plt.title('Satisfaction in work environment')
plt.show()
```



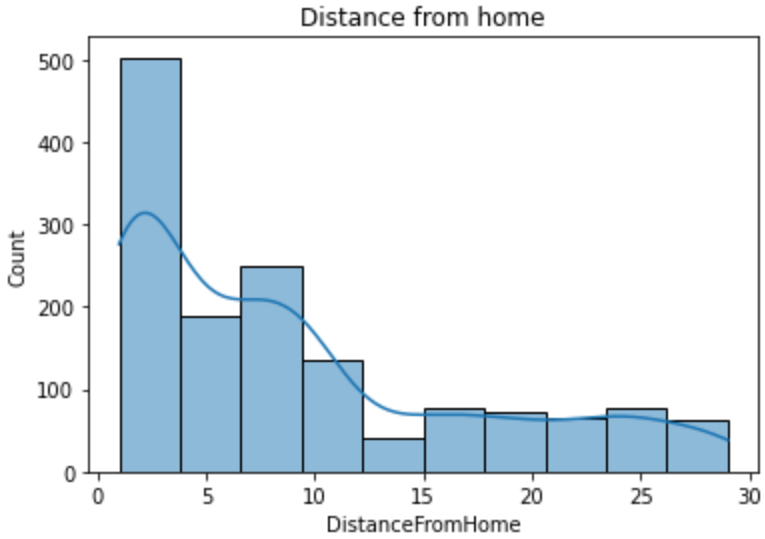
In [28]:

```
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()
```


Distance from home
Count
DistanceFromHome
0
100
200
300
400
500
0
5
10
15
20
25
30

In [ ]: