

HR Employee

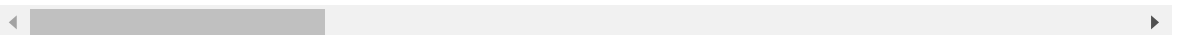
```
In [1]: 1 import pandas as pd
        2 import numpy as np
        3 import matplotlib.pyplot as plt
        4 import seaborn as sns
        5 import plotly.express as px
```

```
In [19]: 1 import pandas as pd
        2
        3 file_path = r'C:\Users\nicho\Downloads\HR Analytics Project\HR-Employee
        4
        5 # Reading the dataset
        6 dataset = pd.read_csv(file_path)
        7 dataset
        8
        9
```

Out[19]:

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

1470 rows × 35 columns

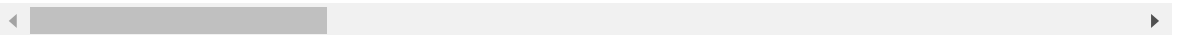


In [20]: 1 dataset.head()

Out[20]:

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

5 rows × 35 columns

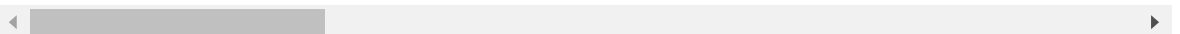


In [21]: 1 dataset.tail()

Out[21]:

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

5 rows × 35 columns



In [22]: 1 dataset.shape

Out[22]: (1470, 35)

In [23]:

1	dataset.info
---	--------------

```

Out[23]: <bound method DataFrame.info of          Age Attrition      BusinessTravel  Da
ilyRate          Department \
0      41      Yes      Travel_Rarely      1102      Sales
1      49      No      Travel_Frequently      279      Research & Development
2      37      Yes      Travel_Rarely      1373      Research & Development
3      33      No      Travel_Frequently      1392      Research & Development
4      27      No      Travel_Rarely      591      Research & Development
...      ...      ...      ...      ...      ...
1465    36      No      Travel_Frequently      884      Research & Development
1466    39      No      Travel_Rarely      613      Research & Development
1467    27      No      Travel_Rarely      155      Research & Development
1468    49      No      Travel_Frequently      1023      Sales
1469    34      No      Travel_Rarely      628      Research & Development

      DistanceFromHome  Education  EducationField  EmployeeCount \
0              1          2  Life Sciences          1
1              8          1  Life Sciences          1
2              2          2      Other          1
3              3          4  Life Sciences          1
4              2          1      Medical          1
...      ...      ...      ...      ...
1465          23          2      Medical          1
1466           6          1      Medical          1
1467           4          3  Life Sciences          1
1468           2          3      Medical          1
1469           8          3      Medical          1

      EmployeeNumber  ...  RelationshipSatisfaction  StandardHours \
0              1  ...              1          80
1              2  ...              4          80
2              4  ...              2          80
3              5  ...              3          80
4              7  ...              4          80
...      ...  ...      ...      ...
1465         2061  ...              3          80
1466         2062  ...              1          80
1467         2064  ...              2          80
1468         2065  ...              4          80
1469         2068  ...              1          80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear \
0              0          8          0
1              1         10          3
2              0          7          3
3              0          8          3
4              1          6          3
...      ...      ...      ...
1465           1         17          3
1466           1          9          5
1467           1          6          0
1468           0         17          3
1469           0          6          3

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole \
0              1          6          4
1              3         10          7
2              3          0          0
3              3          8          7
4              3          2          2
...      ...      ...      ...
1465           3          5          2

```

1466	3	7	7
1467	3	6	2
1468	2	9	6
1469	4	4	3

	YearsSinceLastPromotion	YearsWithCurrManager
0	0	5
1	1	7
2	0	0
3	3	0
4	2	2
...
1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

[1470 rows x 35 columns]>

In [24]:

1	dataset.columns
---	-----------------

Out[24]:

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
      'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
      'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
      'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
      'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
      'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
      'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
      'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
      'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
      'YearsWithCurrManager'],
      dtype='object')
```

```
In [25]: 1 dataset.isnull().sum()
```

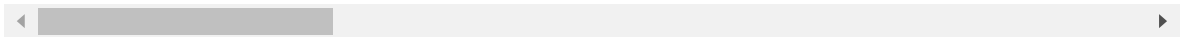
```
Out[25]: Age                                0
Attrition                                0
BusinessTravel                          0
DailyRate                              0
Department                             0
DistanceFromHome                       0
Education                              0
EducationField                          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 [26]: 1 dataset = dataset.drop_duplicates()
2 dataset = dataset.dropna()
3 dataset
```

Out[26]:

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

1470 rows × 35 columns

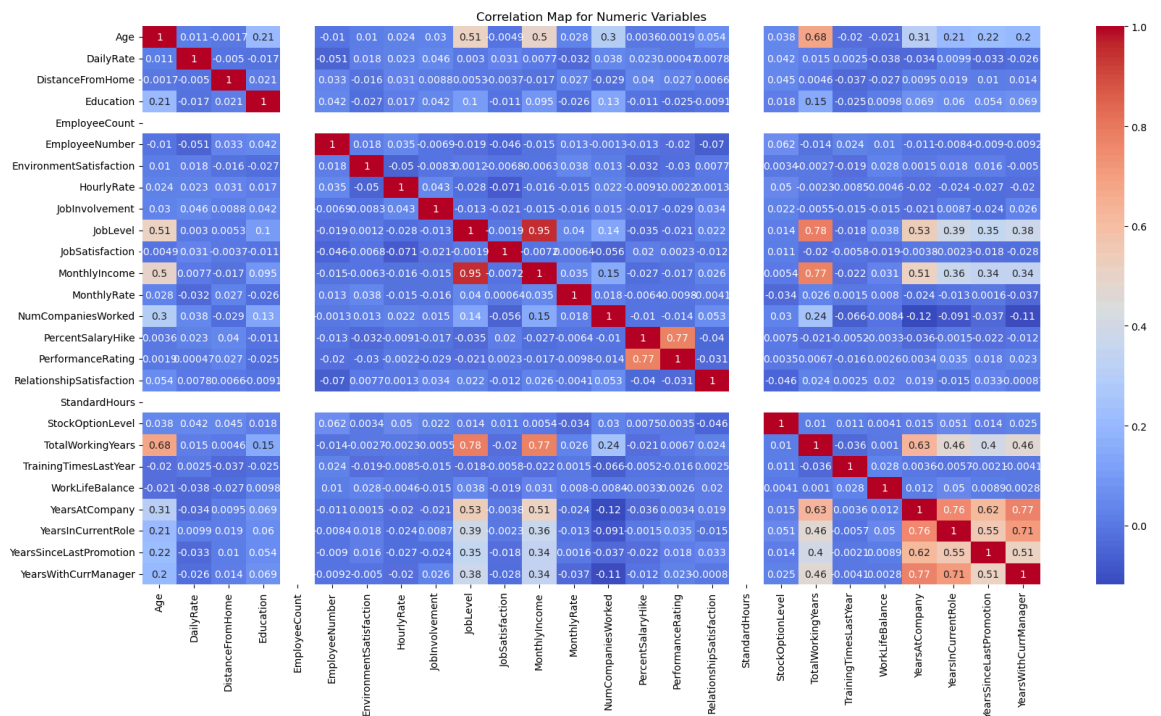


In [27]:

```

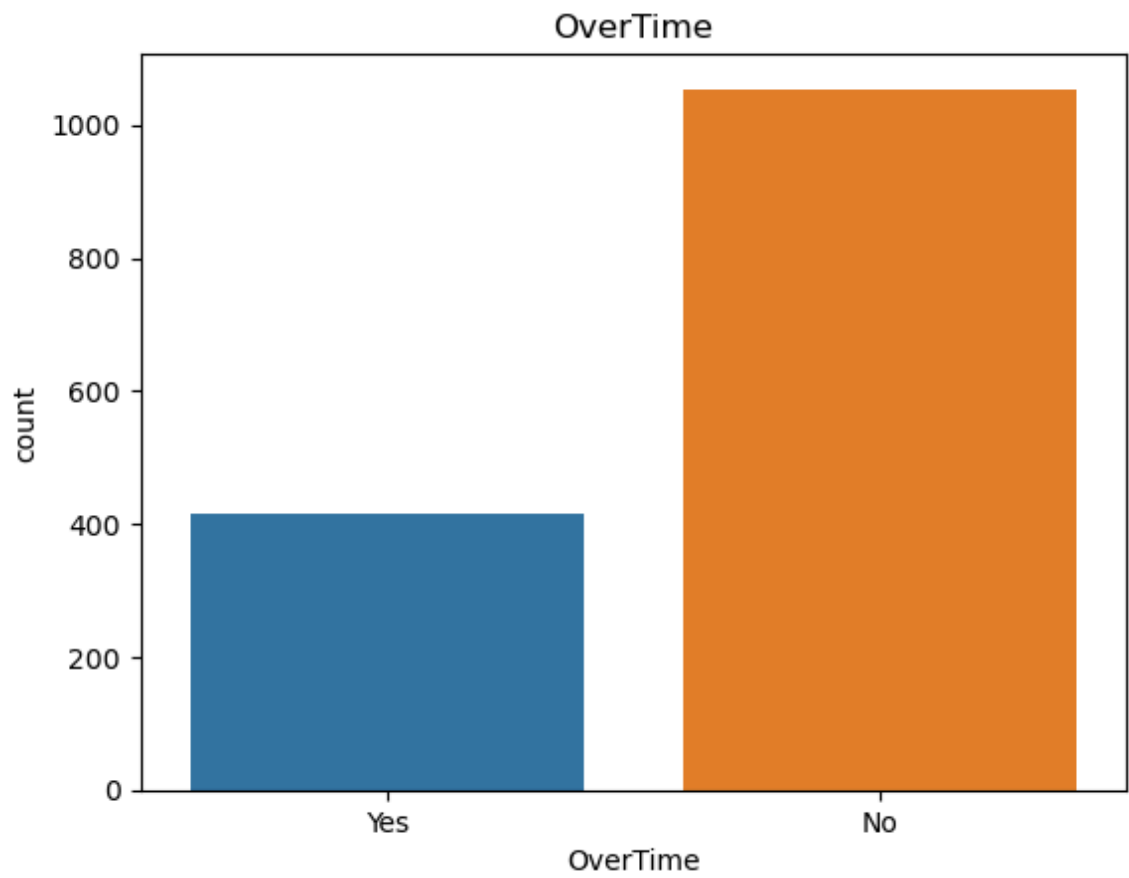
1
2 # Correlation map for all numeric variables
3
4 numeric_df = dataset.select_dtypes(include=['number'])
5
6 corr_matrix = numeric_df.corr()
7
8
9 plt.figure(figsize=(20, 10))
10 sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
11 plt.title('Correlation Map for Numeric Variables')
12 plt.show()
13

```



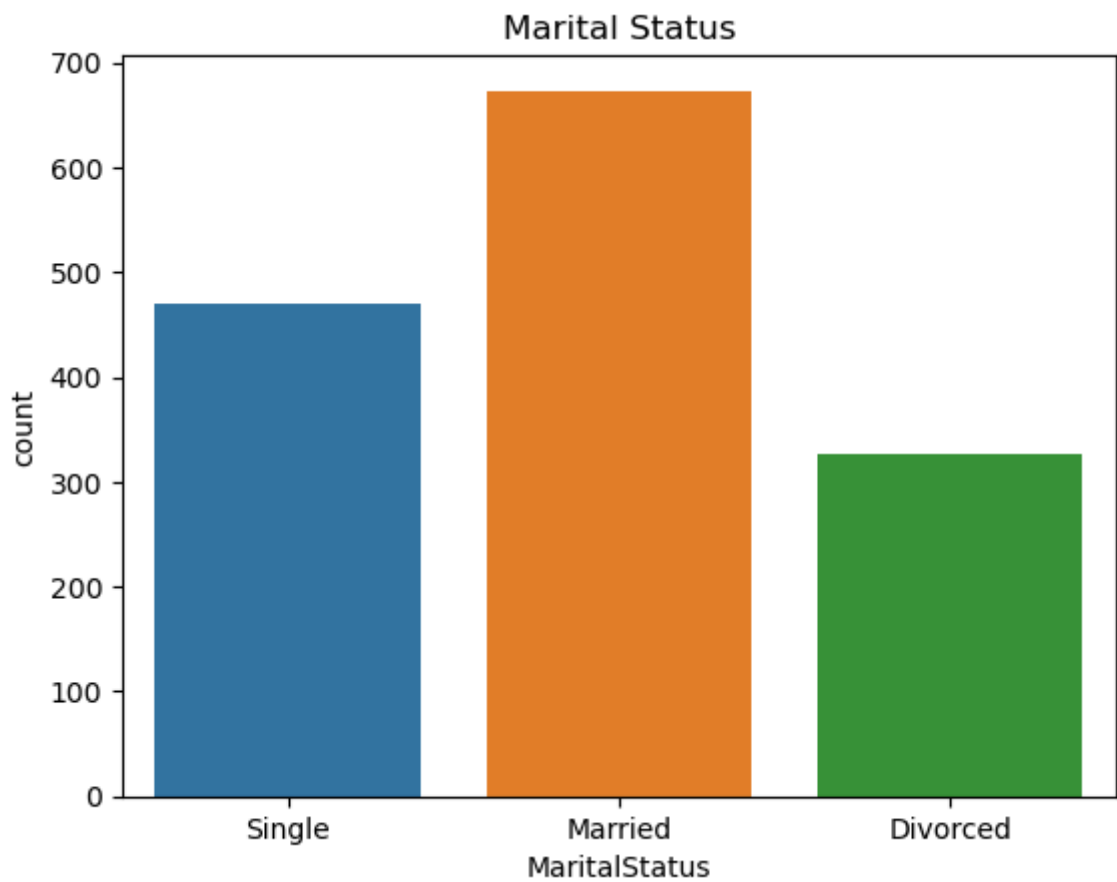
Visualize the categorical variables


```
In [28]: 1 data_1 = dataset
2 # Overtime
3 sns.countplot(data=data_1, x='OverTime')
4 plt.title('OverTime')
5 plt.show()
```



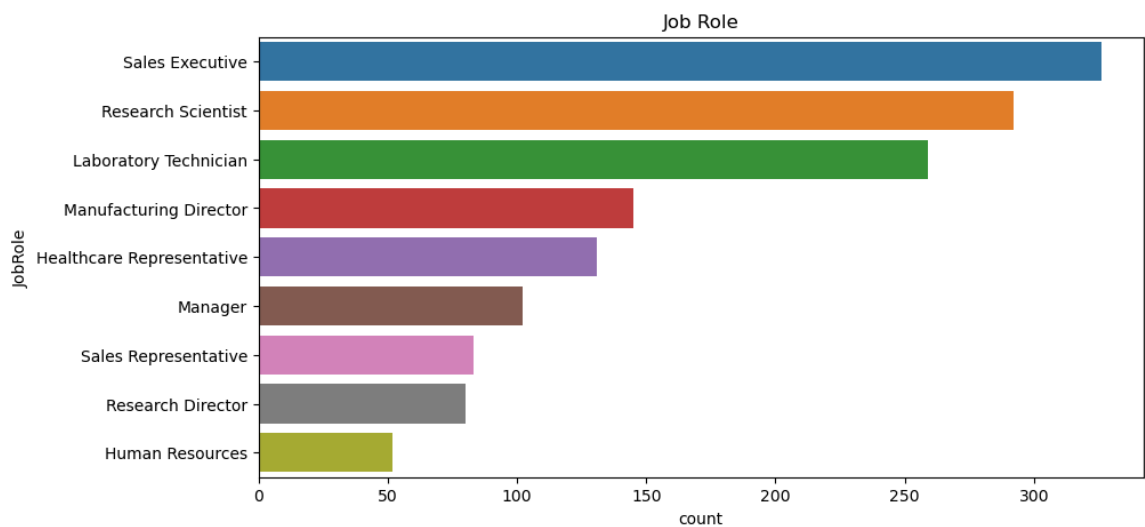
In [29]:

```
1 # Marital Status
2 sns.countplot(data=data_1, x='MaritalStatus')
3 plt.title('Marital Status')
4 plt.show()
```



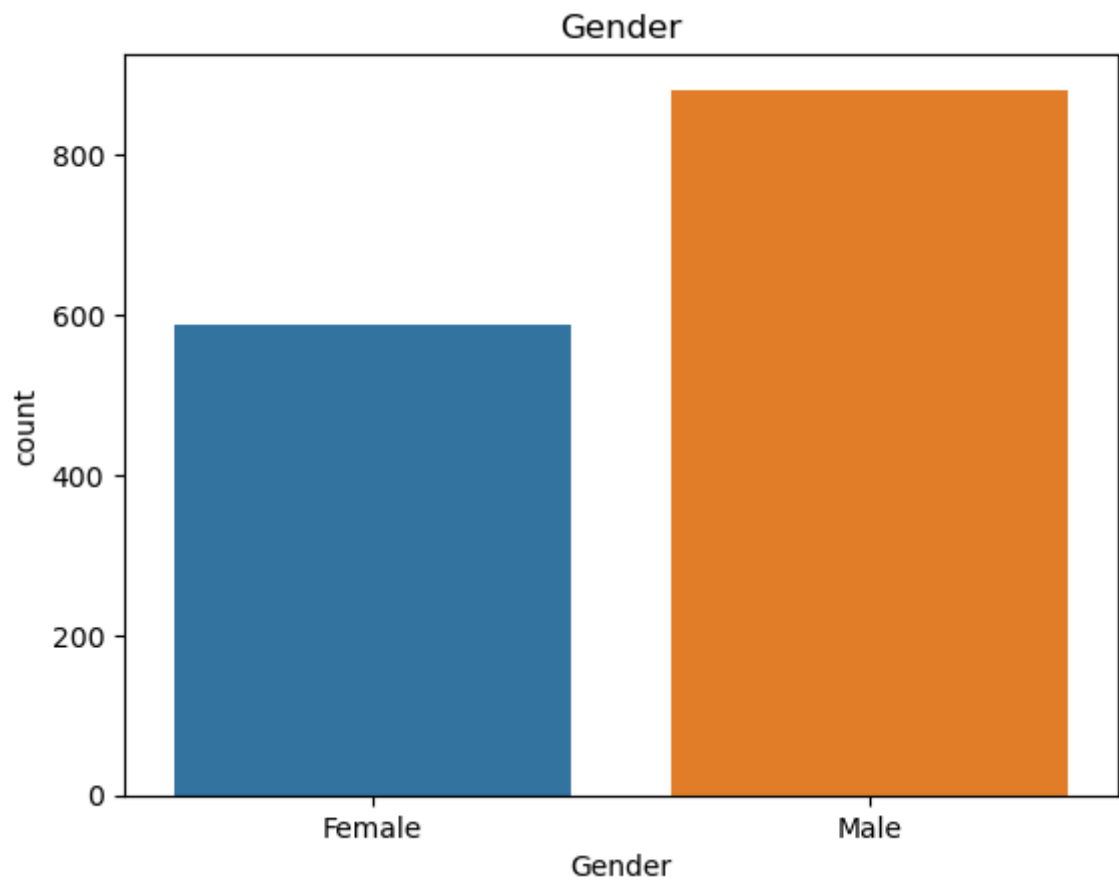
In [30]:

```
1 # Job Role
2 plt.figure(figsize=(10, 5))
3 sns.countplot(data=data_1, y='JobRole')
4 plt.title('Job Role')
5 plt.show()
6
```



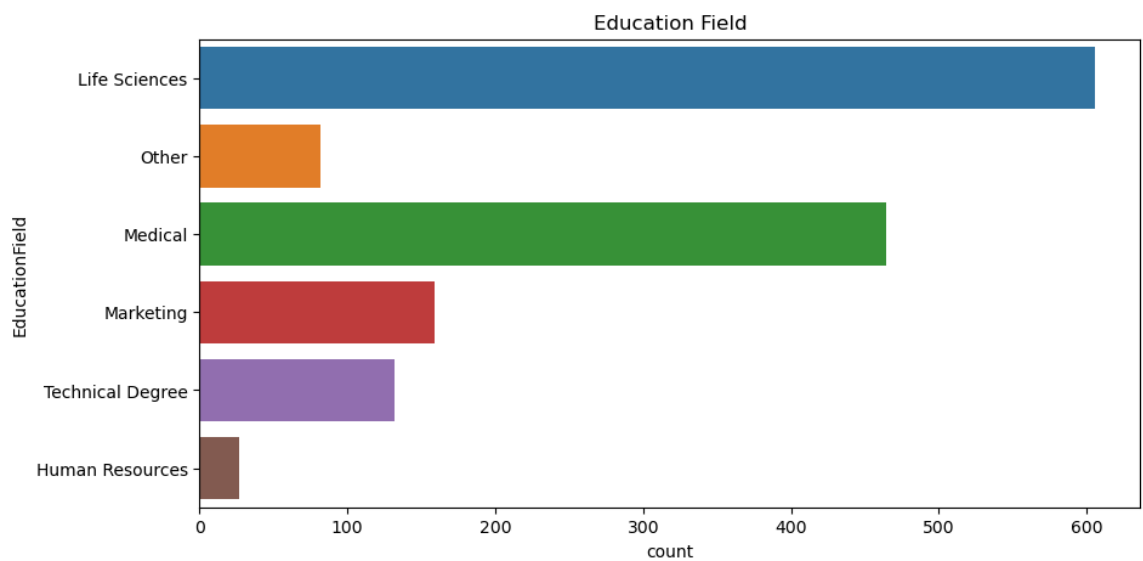
In [31]:

```
1 # Gender
2 sns.countplot(data=data_1, x='Gender')
3 plt.title('Gender')
4 plt.show()
5
```

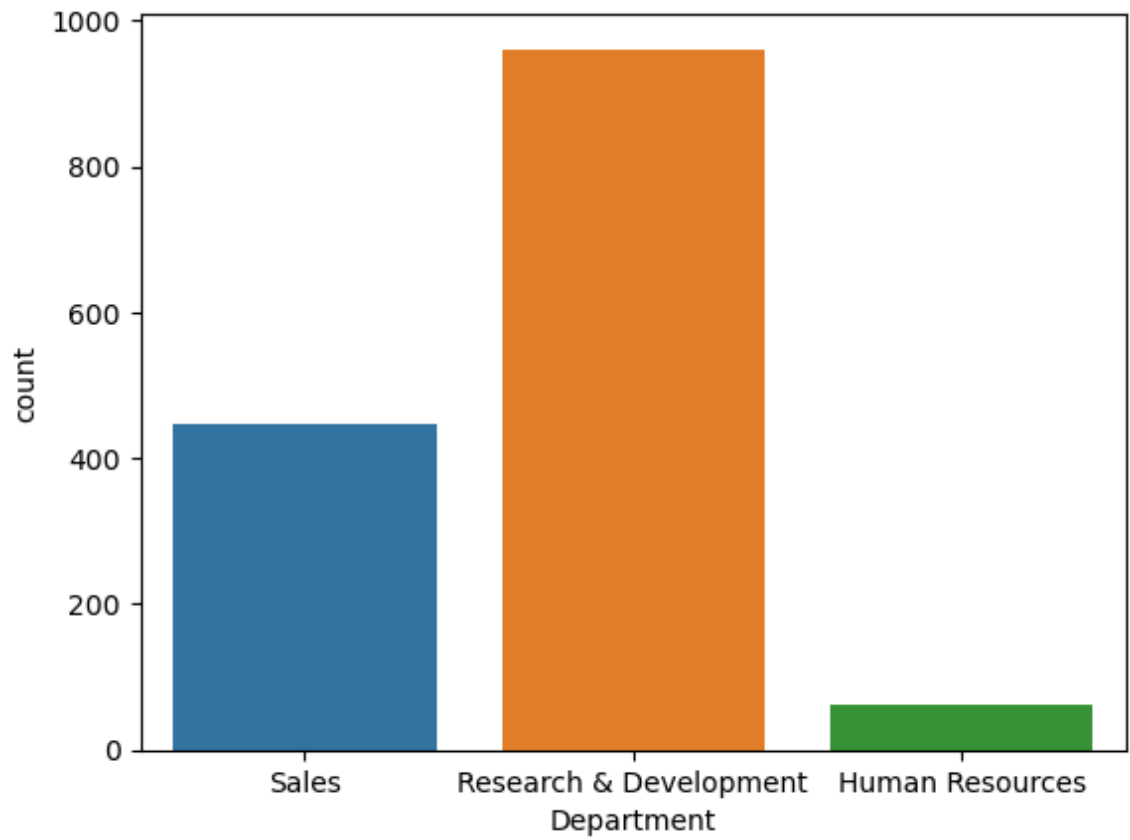


In [32]:

```
1 # Education Field
2 plt.figure(figsize=(10,5))
3 sns.countplot(data=data_1, y = 'EducationField')
4 plt.title('Education Field')
5 plt.show()
```

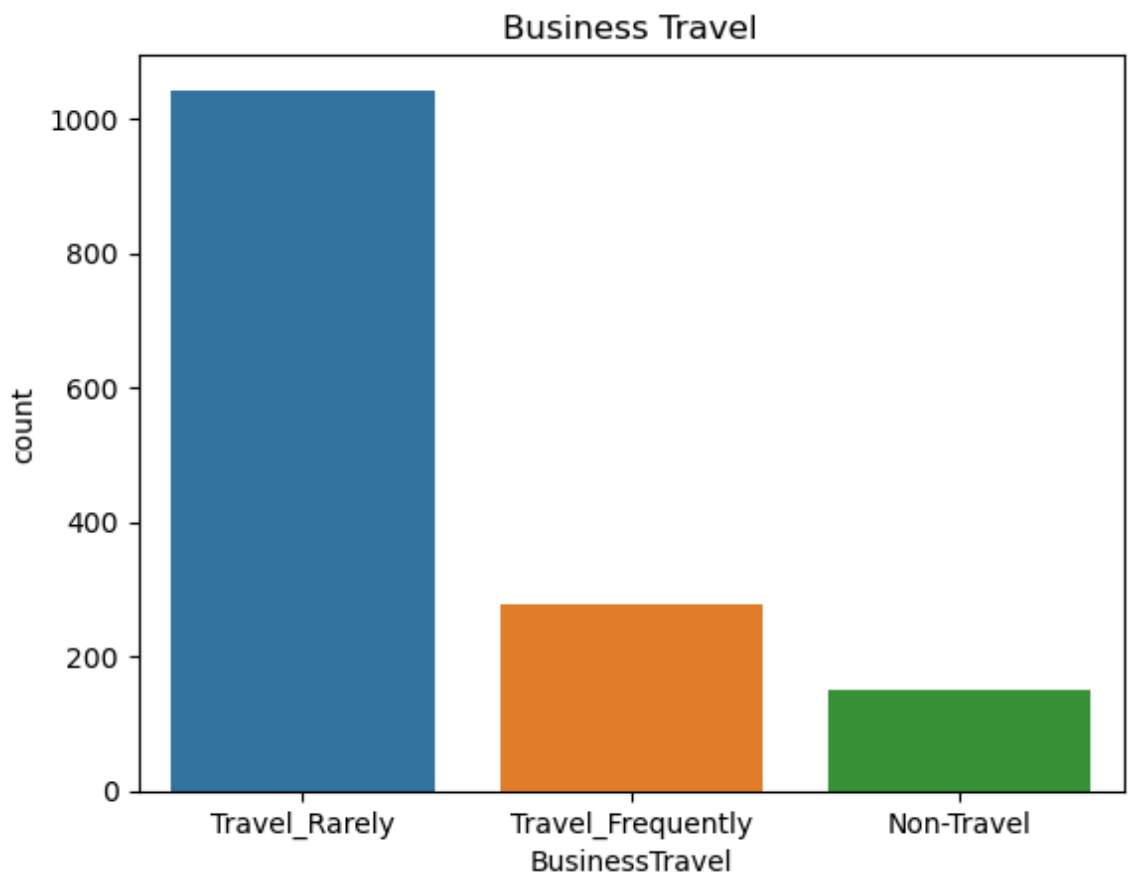


```
In [33]: 1 # Department
2 sns.countplot(data =data_1, x='Department')
3 plt.show()
```



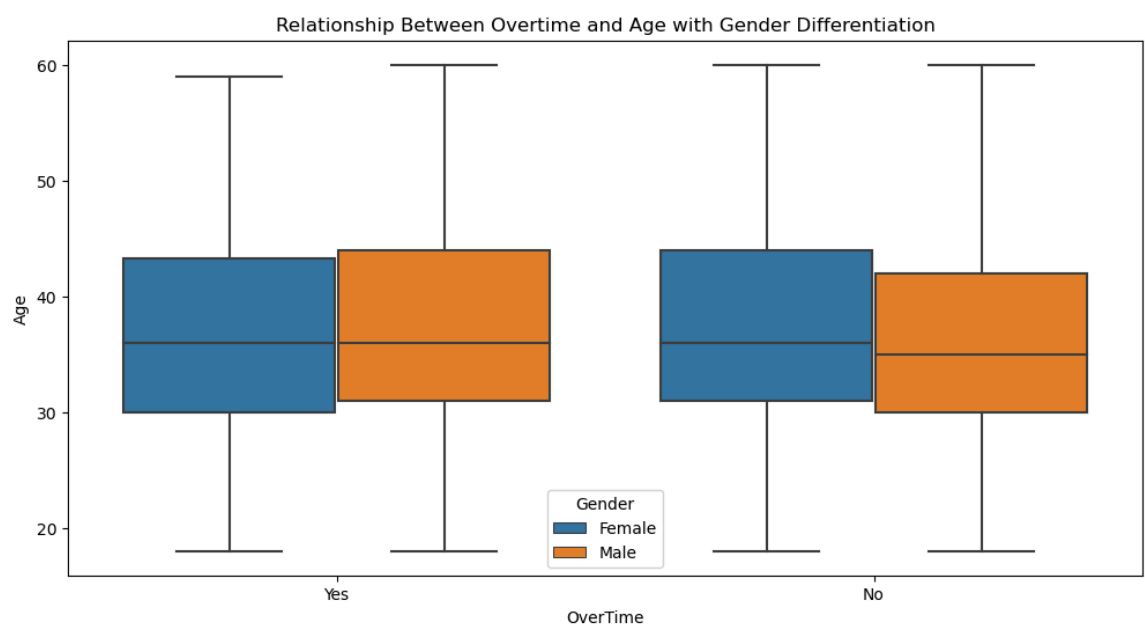
In [34]:

```
1 # Business Travel
2 sns.countplot(data= data_1, x='BusinessTravel')
3 plt.title('Business Travel')
4 plt.show()
```



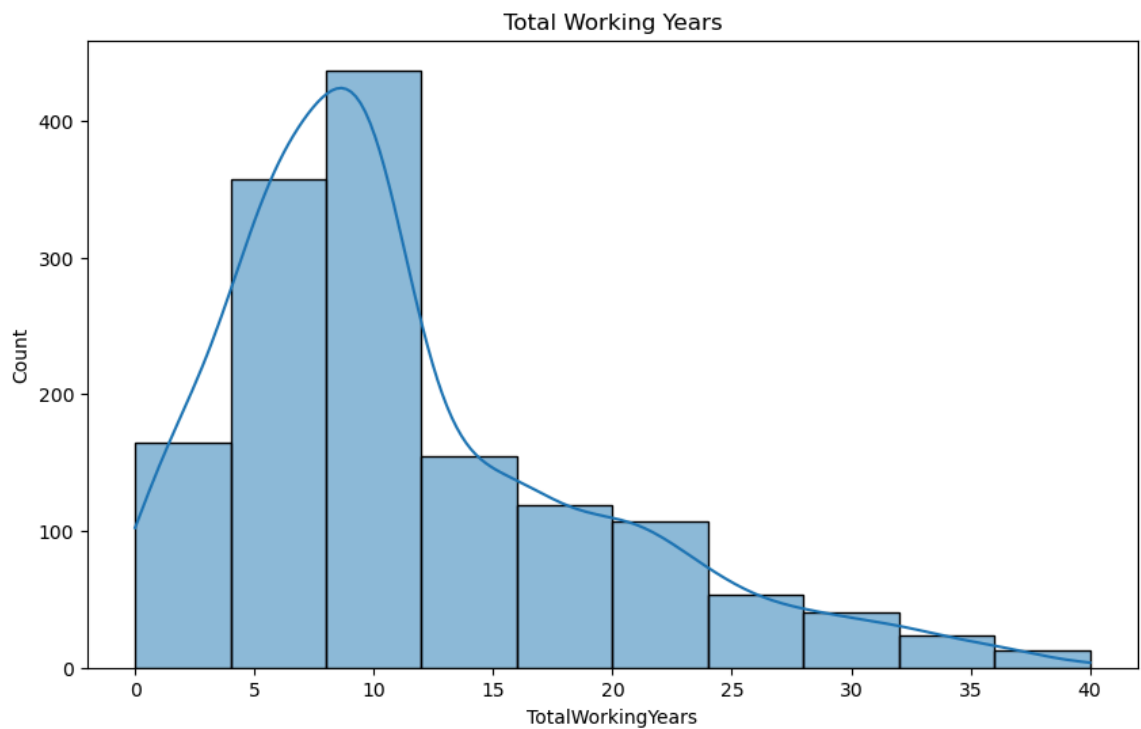
In [35]:

```
1 # Relationship Between Overtime and Age with Gender differentiation
2 plt.figure(figsize=(12, 6))
3 sns.boxplot(data=data_1, x='OverTime', y='Age', hue='Gender')
4 plt.title('Relationship Between Overtime and Age with Gender Differenti
5 plt.show()
```

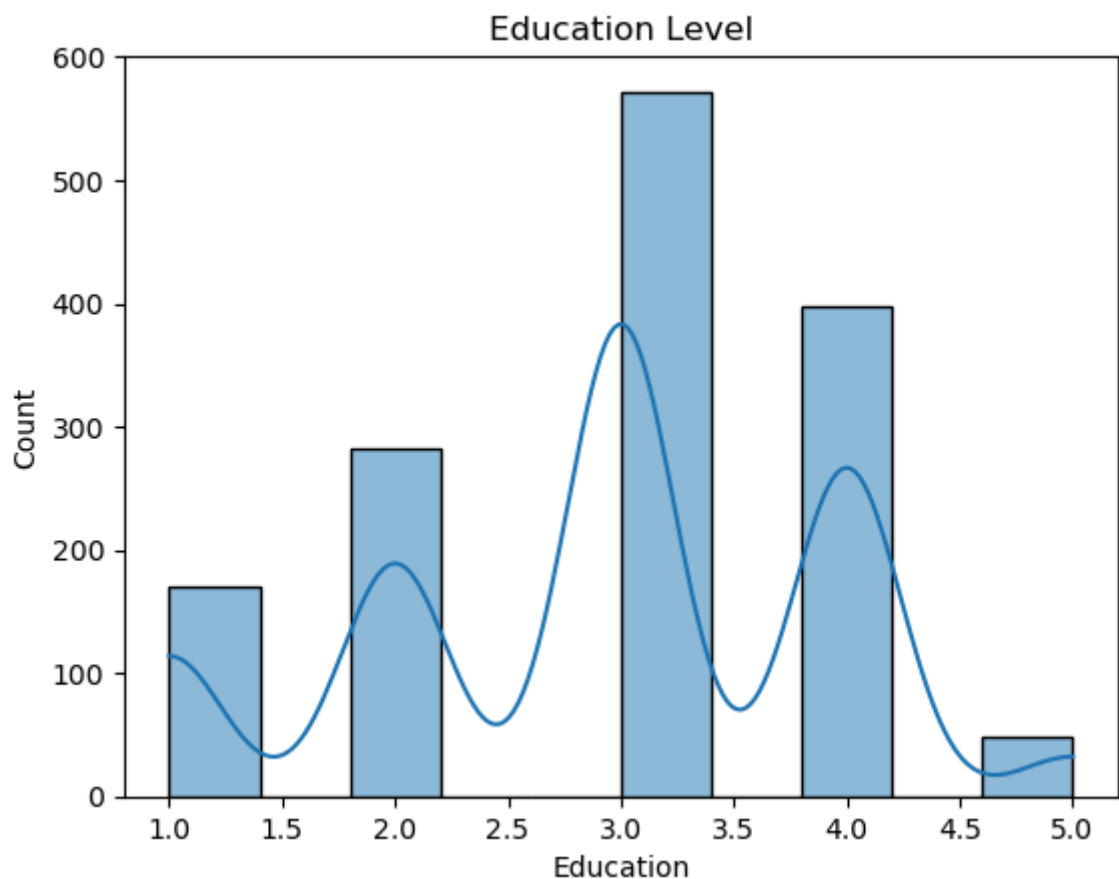


Plotting numerical variables

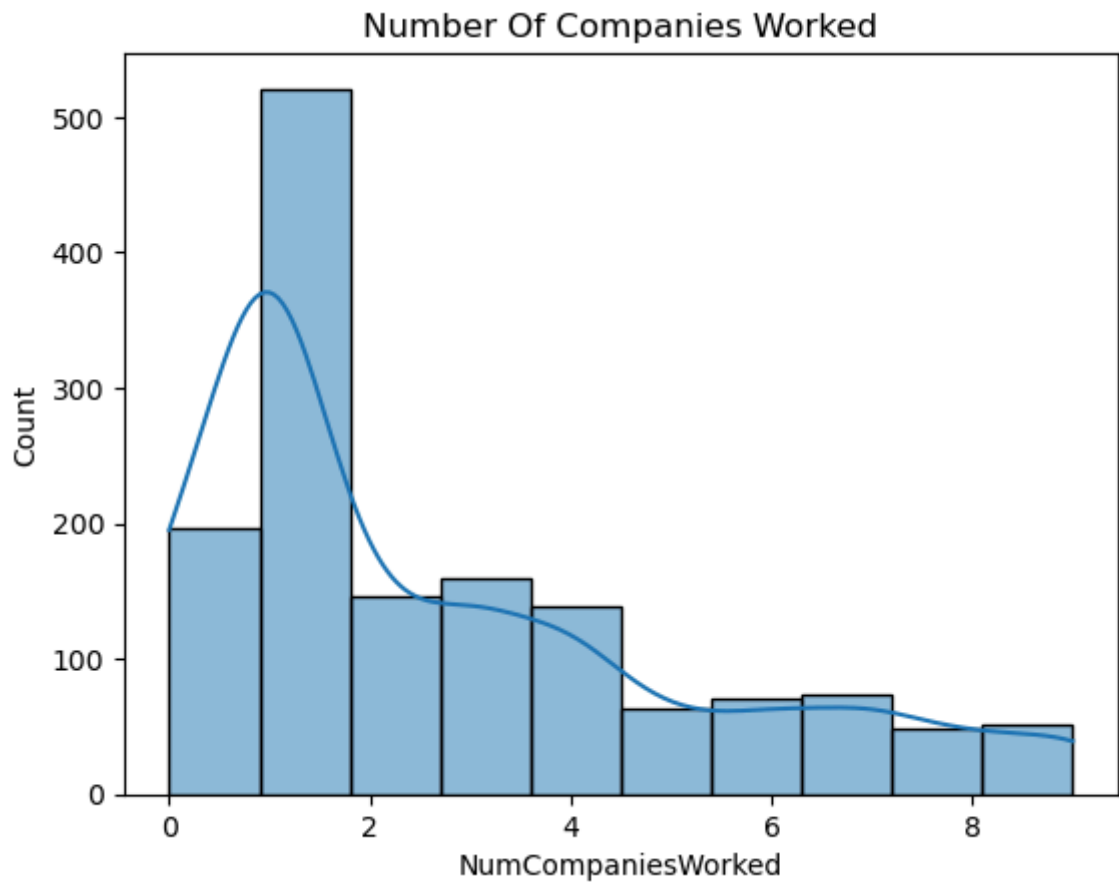
```
In [36]: 1 #working years total
2 plt.figure(figsize=(10, 6))
3 sns.histplot(data=data_1, x='TotalWorkingYears', bins=10, kde=True)
4 plt.title('Total Working Years')
5 plt.show()
```



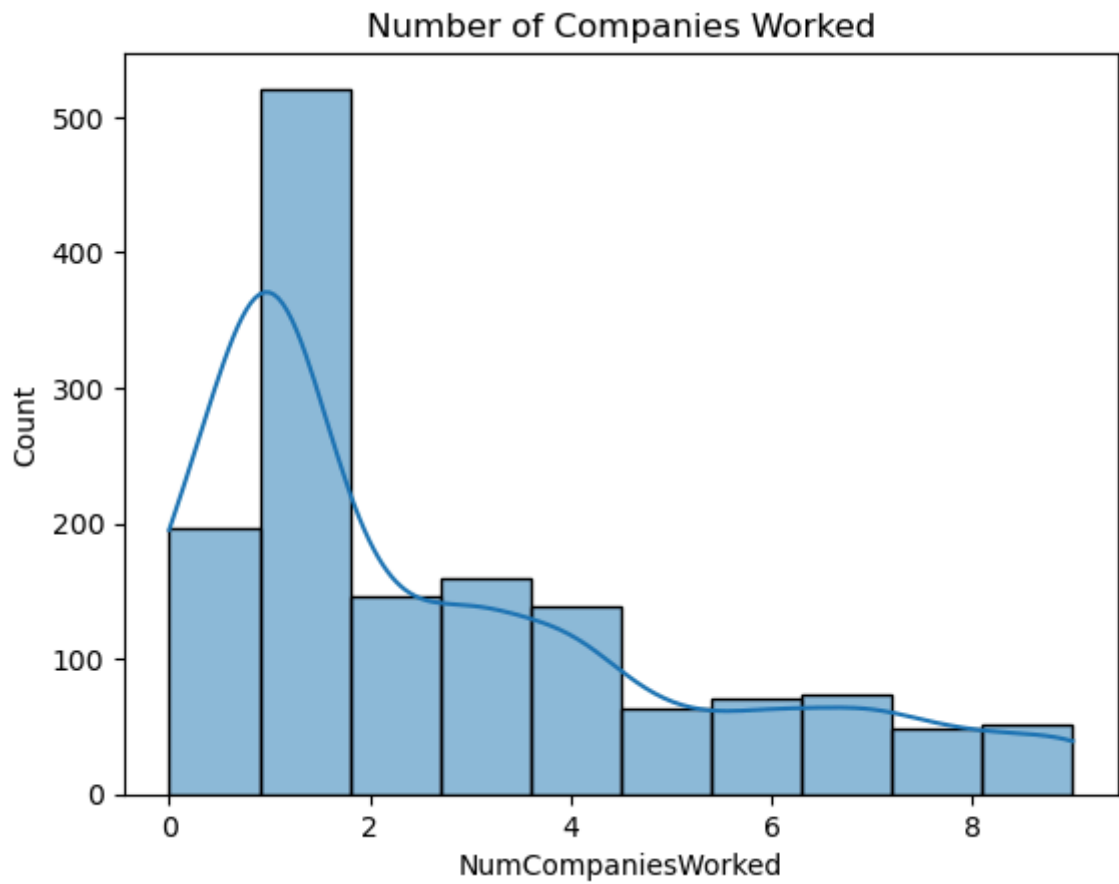
```
In [37]: 1 # Education Level
2 plt.title('Education Level')
3 sns.histplot(data=data_1, x='Education', bins=10, kde=True)
4 plt.show()
```



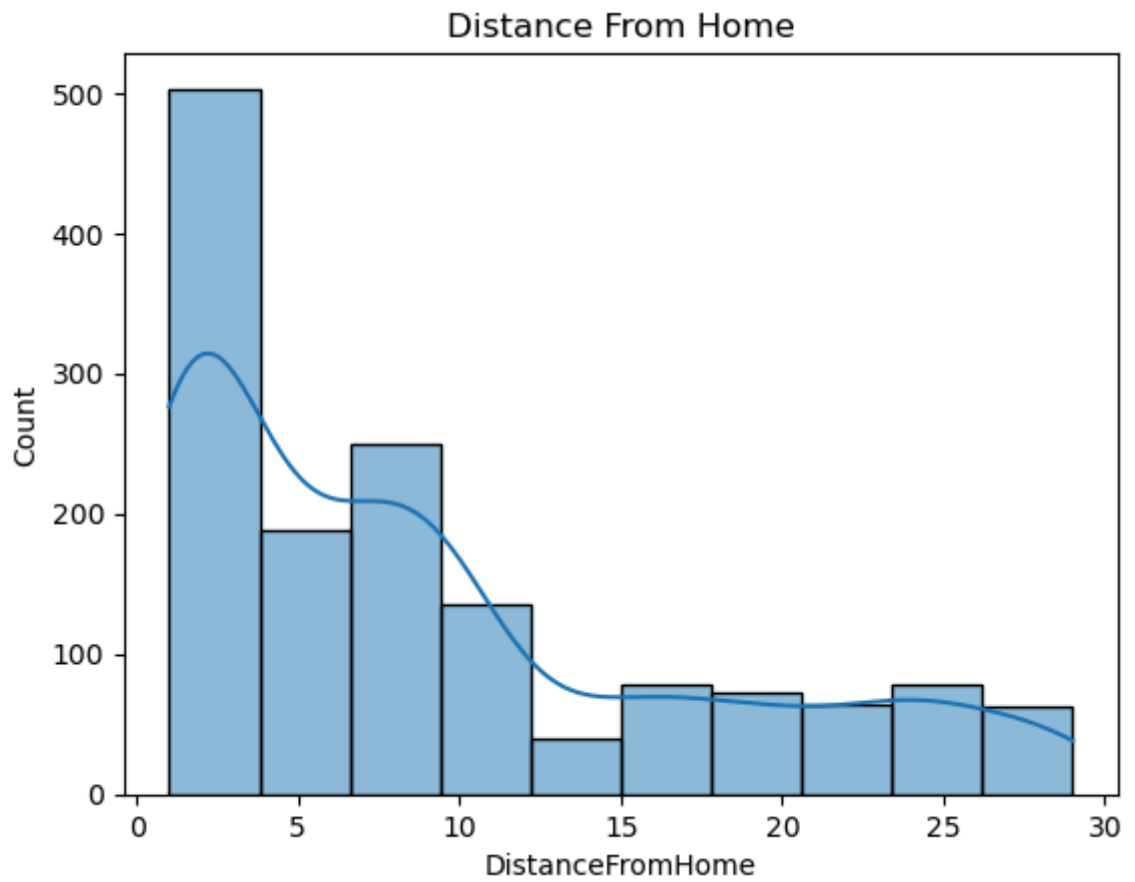
```
In [38]: 1 # Number of Companies Worked
2 sns.histplot(data= data_1, x='NumCompaniesWorked', bins=10, kde=True)
3 plt.title('Number Of Companies Worked')
4 plt.show()
```



```
In [39]: 1 # Number of Companies Worked  
2 sns.histplot(data= data_1, x='NumCompaniesWorked', bins=10, kde=True)  
3 plt.title('Number of Companies Worked')  
4 plt.show()
```

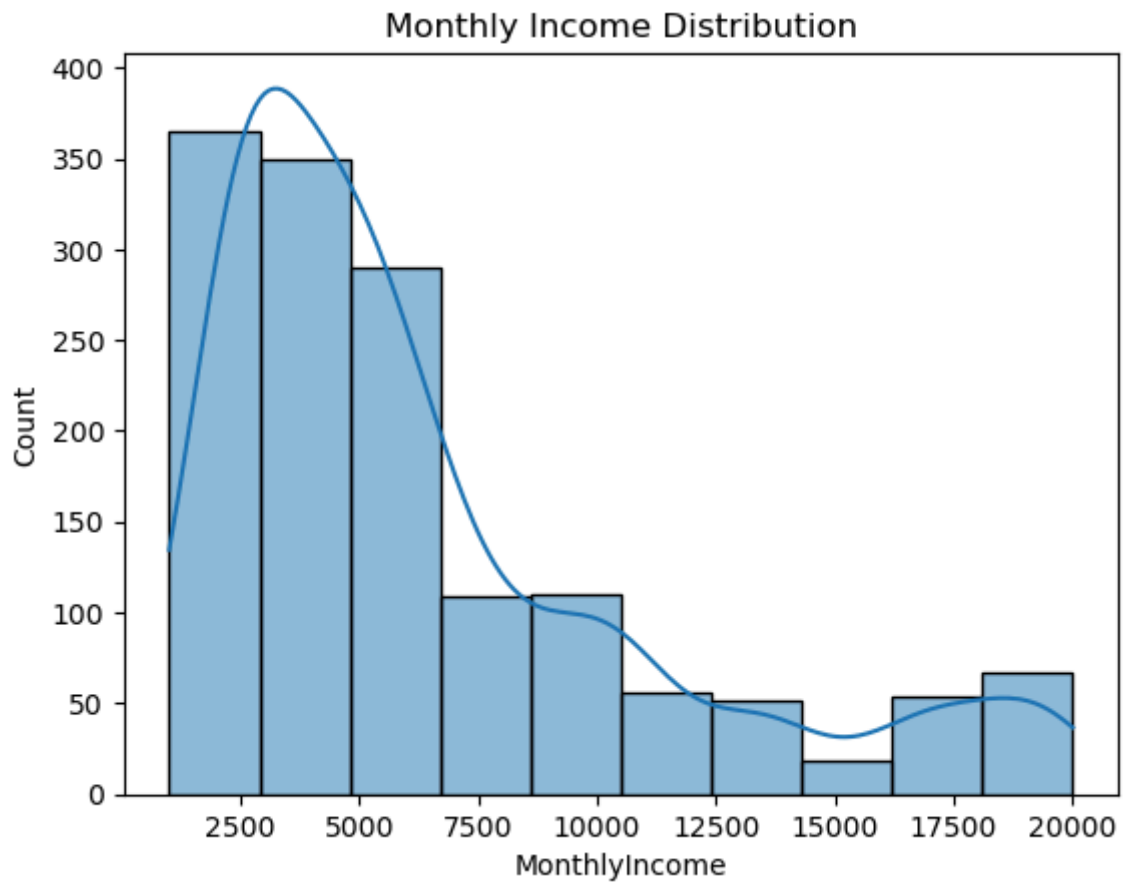



```
In [40]: 1 # Distance From Home
2 plt.title('Distance From Home')
3 sns.histplot(data=data_1, x='DistanceFromHome', bins= 10, kde = True)
4 plt.show()
```



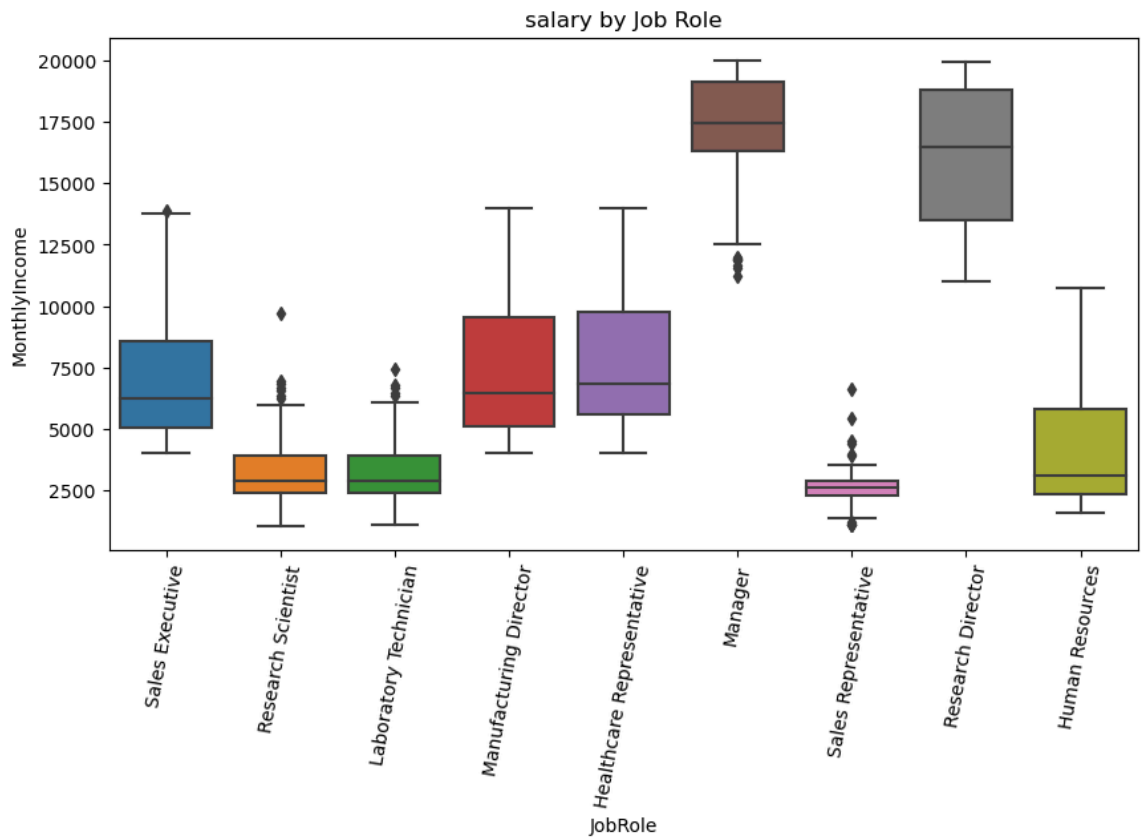
Salary Analysis

```
In [41]: 1 # Monthly Income Distribution
2 sns.histplot(data=data_1, x='MonthlyIncome', bins = 10, kde=True)
3 plt.title('Monthly Income Distribution')
4 plt.show()
```



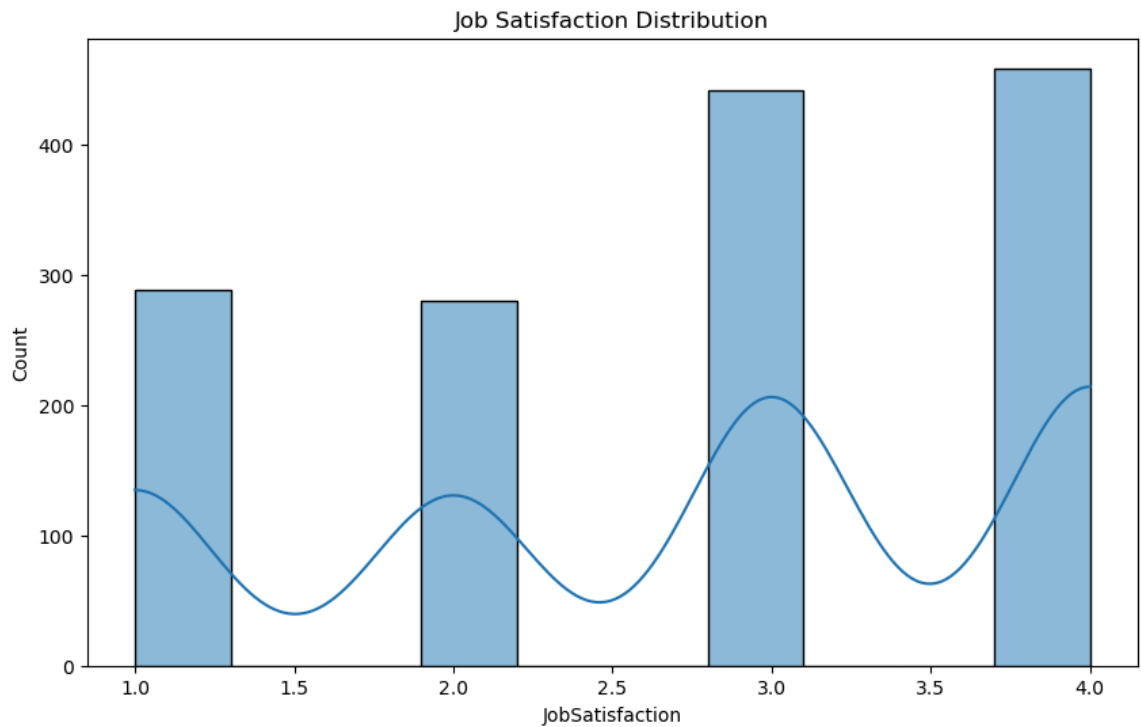
In [42]:

```
1 # Salary by Job Role
2 plt.figure(figsize=(10,5))
3 sns.boxplot(data=data_1, x='JobRole', y='MonthlyIncome')
4 plt.title('salary by Job Role')
5 plt.xticks(rotation = 80)
6 plt.show()
```

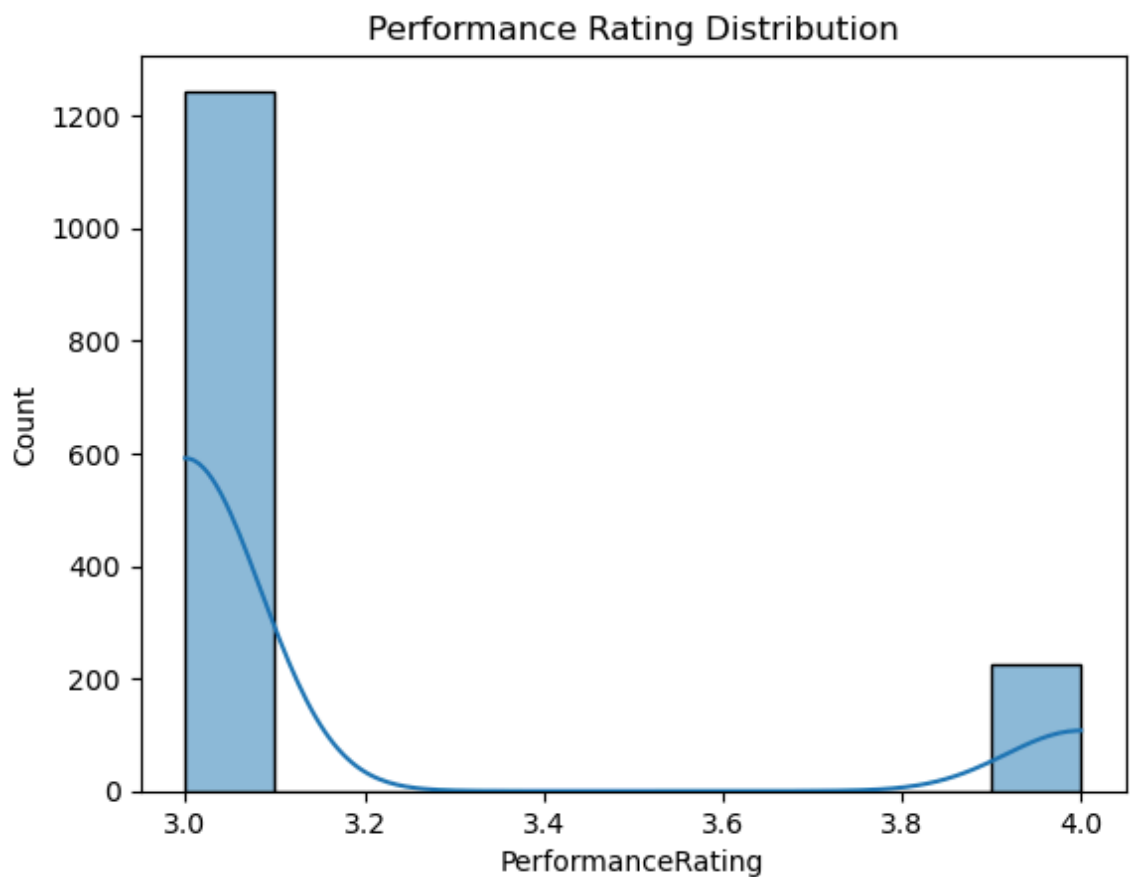


Employee Satisfaction and Performance

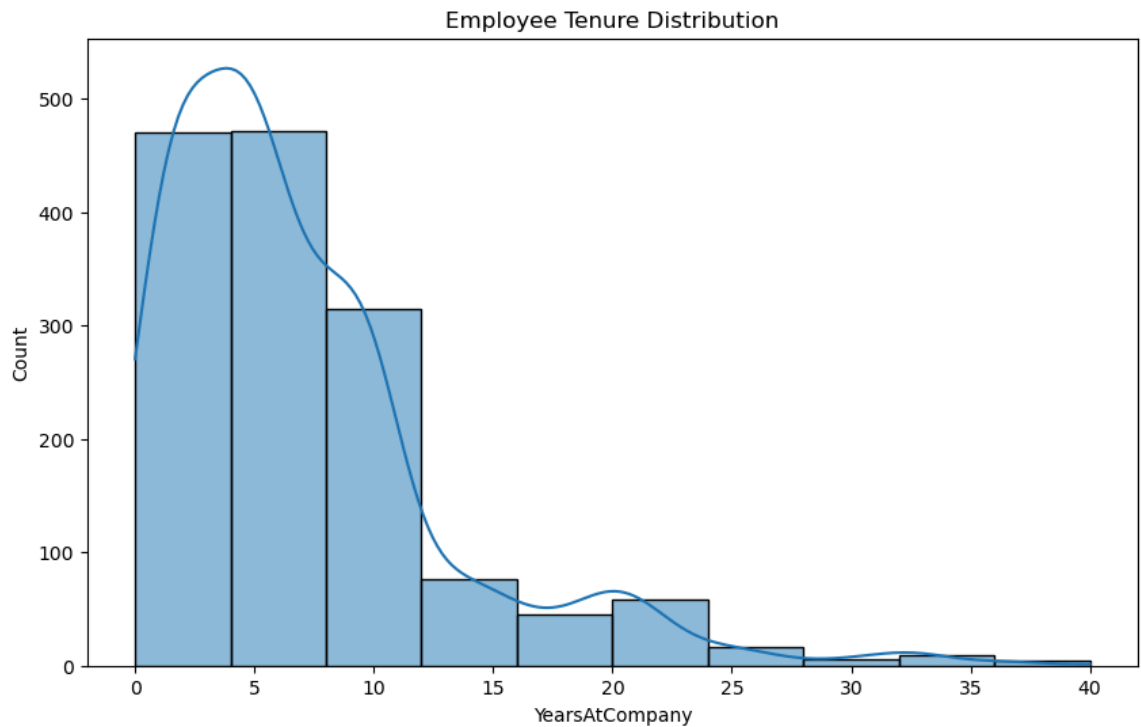
```
In [43]: 1 # Job Satisfaction Distribution
2 plt.figure(figsize=(10, 6))
3 sns.histplot(data=data_1, x='JobSatisfaction', bins=10, kde=True)
4 plt.title('Job Satisfaction Distribution')
5 plt.show()
```



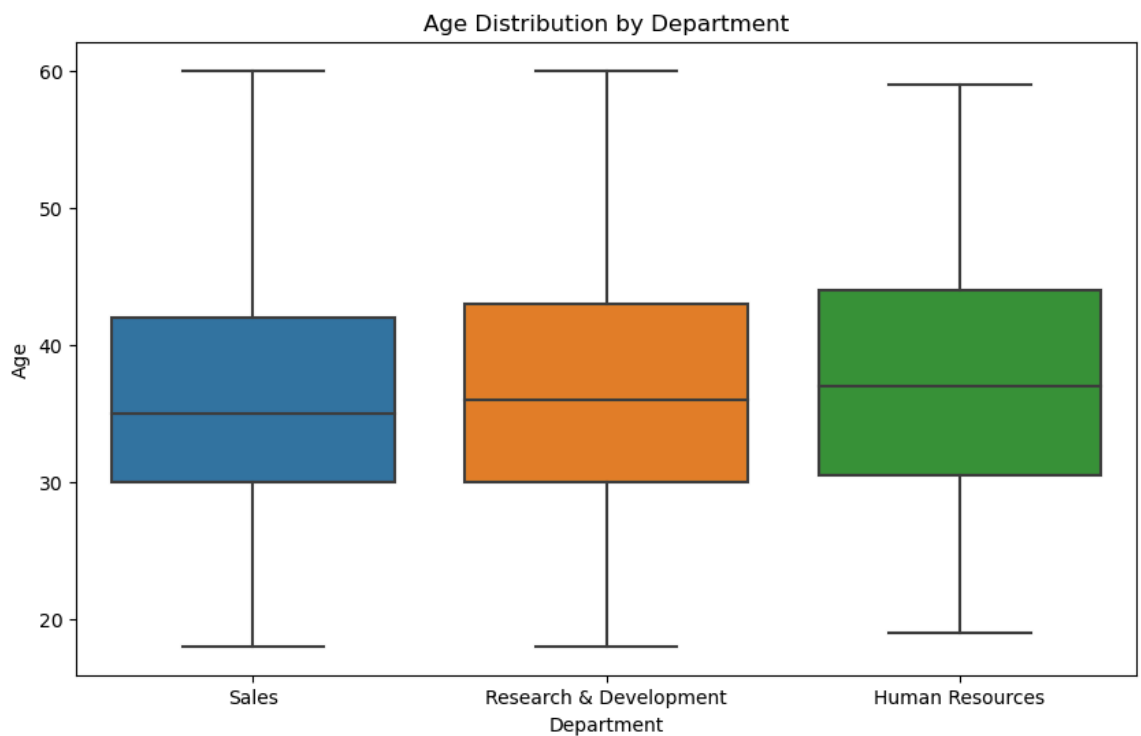
```
In [44]: 1 # Performance Rating distribution
2 sns.histplot(data =dataset, x='PerformanceRating', bins=10, kde=True)
3 plt.title('Performance Rating Distribution')
4 plt.show()
```



```
In [45]: 1 # Employee Tenure
2 plt.figure(figsize=(10, 6))
3 sns.histplot(data=data_1, x='YearsAtCompany', bins=10, kde=True)
4 plt.title('Employee Tenure Distribution')
5 plt.show()
```



```
In [46]: 1 # Age Distribution by Department
2 plt.figure(figsize=(10,6))
3 sns.boxplot(data= data_1, x='Department', y='Age')
4 plt.title('Age Distribution by Department')
5 plt.show()
```



In [47]:

```
1 # Attrition by Job Role
2 plt.figure(figsize=(10,5))
3 sns.countplot(data= data_1, x='JobRole', hue='Attrition')
4 plt.xticks(rotation=90)
5 plt.title('Attrition by Job Role')
6 plt.show()
```

