HR Analytics - Project 3

Kaashi Srinivauslu

MeriSKILL!

In [13]:
```
# importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df = pd.read_csv('HR-Employee-Attrition.csv')
```

In [3]:
```
df.head()
```

Out[3]:

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

5 rows × 35 columns

In [4]:
```
df.shape
```

Out[4]:
```
(1470, 35)
```

In [5]:
```
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
```

In [6]:
```
df.columns
```

Out[6]:
```
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 [7]:
```
df.isnull().sum()
```

Out[7]:
```
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
```

Tasks to perform

Data Cleaning

- Deleting Redundent Columns
- Renaming the columns
- Dropping duplicates
- Cleaning Individual columns
- Removing NaN values from dataset

In [8]:
```
# dropping the duplicates from dataset
df= df.drop_duplicates()
```

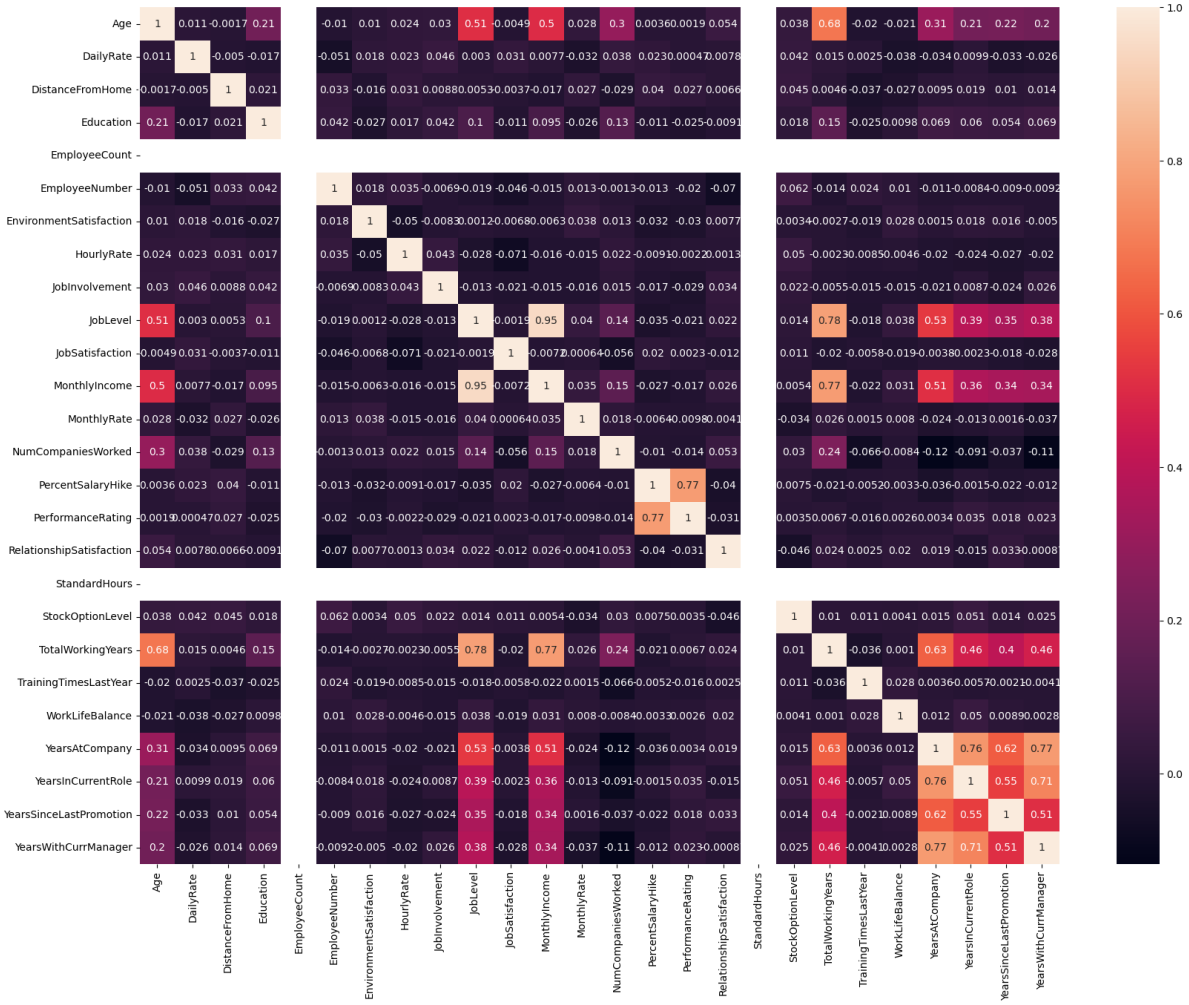
In [9]:
```
# removing the NaN values
df = df.dropna()
```

Data Vusialization

- Plot of correlation map for all numeric values

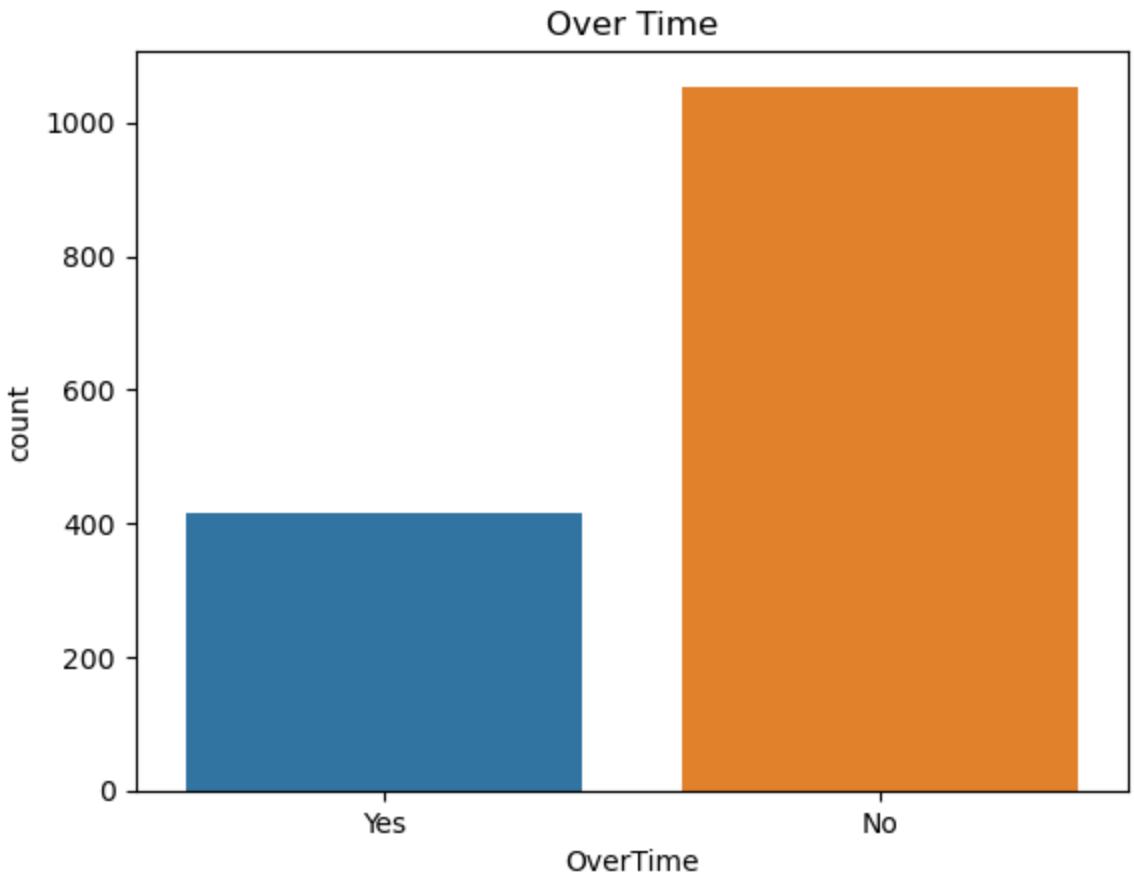
In [15]:
```
plt.figure(figsize=(20,15))
sns.heatmap(df.corr(),annot=True)
```

Out[15]:
```
<AxesSubplot:>
```



Overtime

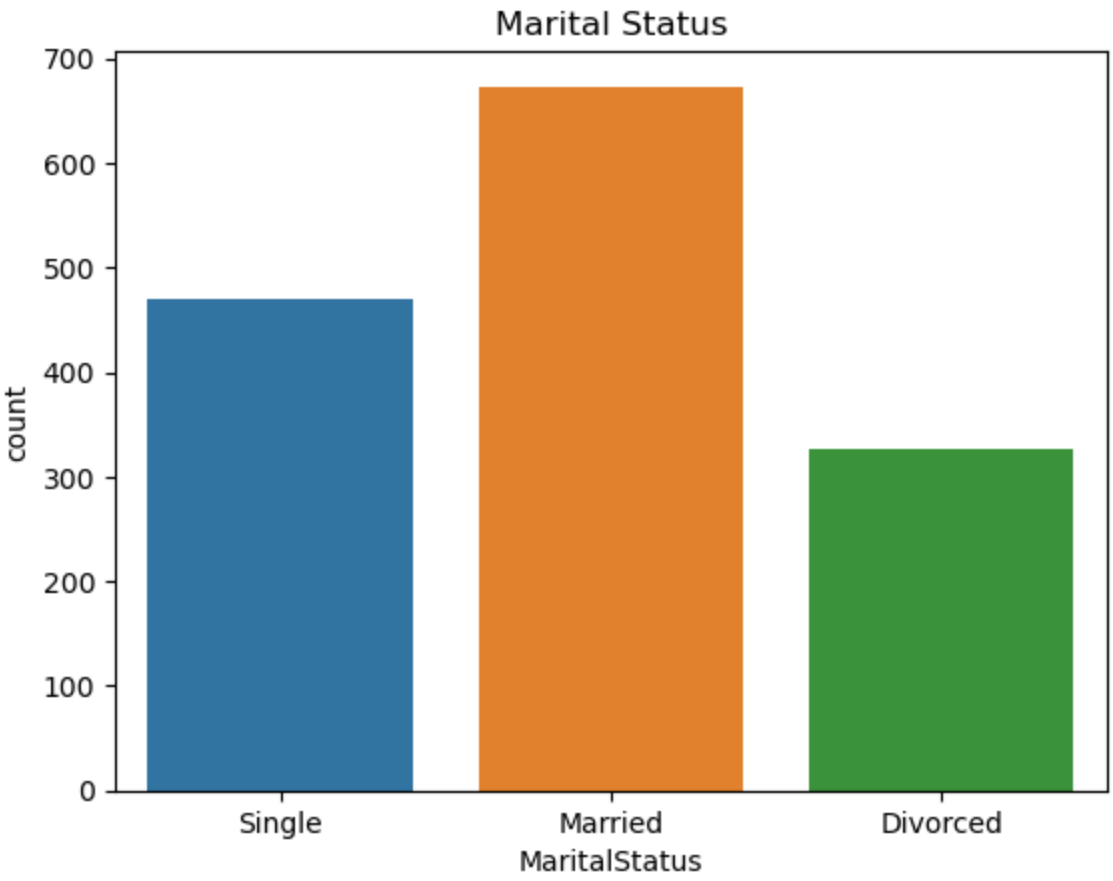
In [18]:
```
sns.countplot(df['OverTime'])
plt.title('Over Time')
plt.show()
```



- Most of the employees prefer not to work overtime

Marital Status

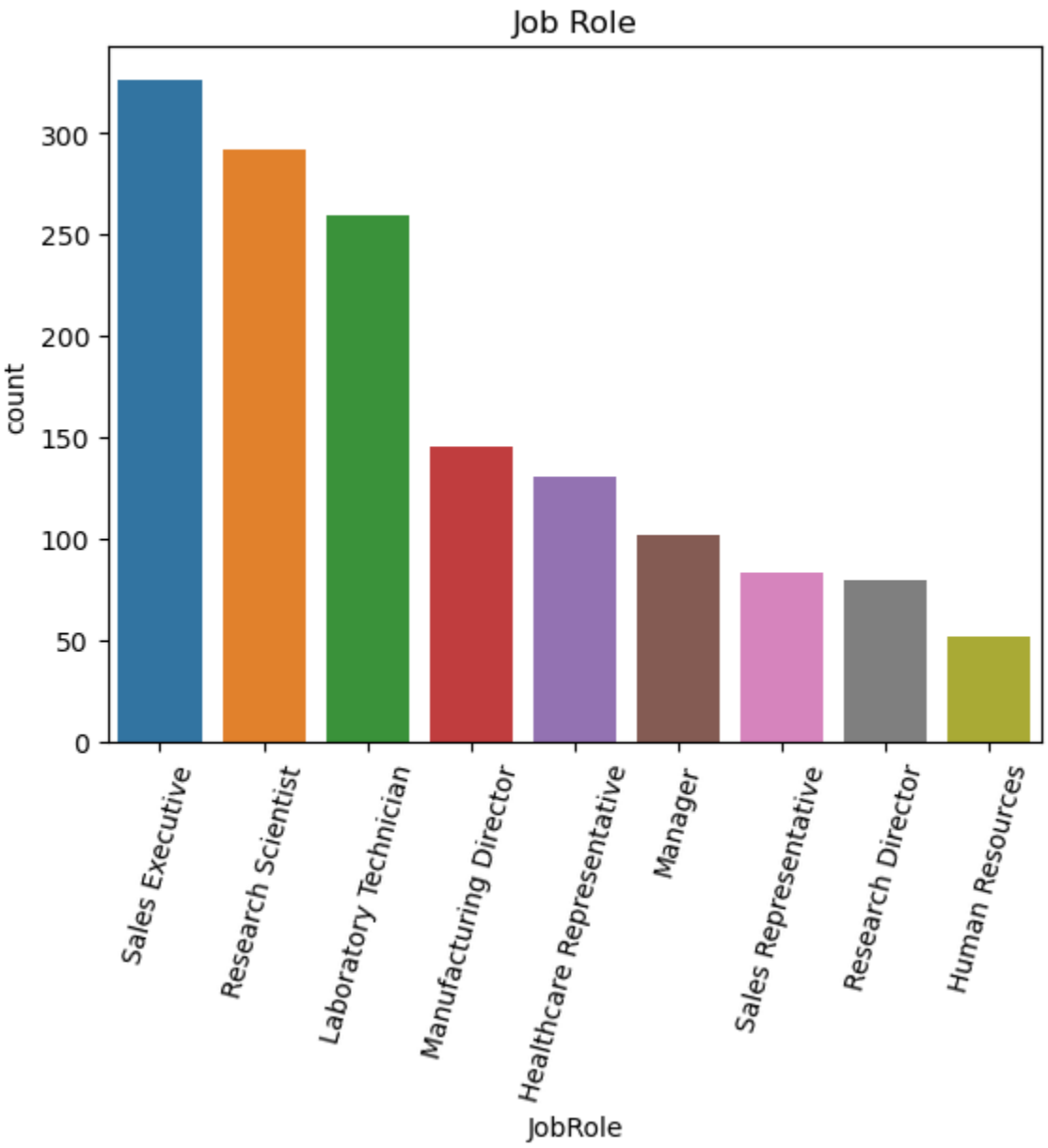
In [19]:
```
sns.countplot(df['MaritalStatus'])
plt.title('Marital Status')
plt.show()
```



- Employees are mostly married or single

Job Role

In [22]:
```
sns.countplot(df['JobRole'])
plt.title('Job Role')
plt.xticks(rotation=75)
plt.show()
```

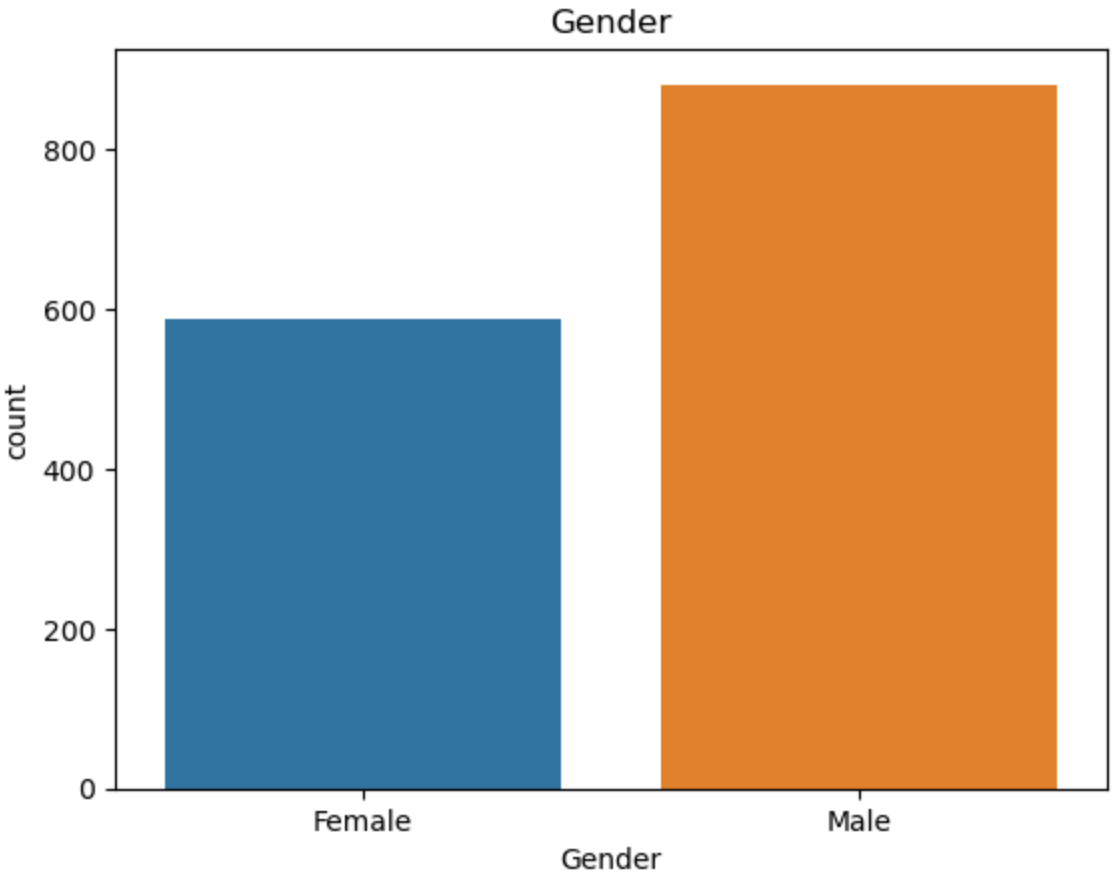


- Sales team has highest number of employees and HR team with least

---

Gender

In [23]:
```
sns.countplot(df['Gender'])
plt.title('Gender')
plt.show()
```

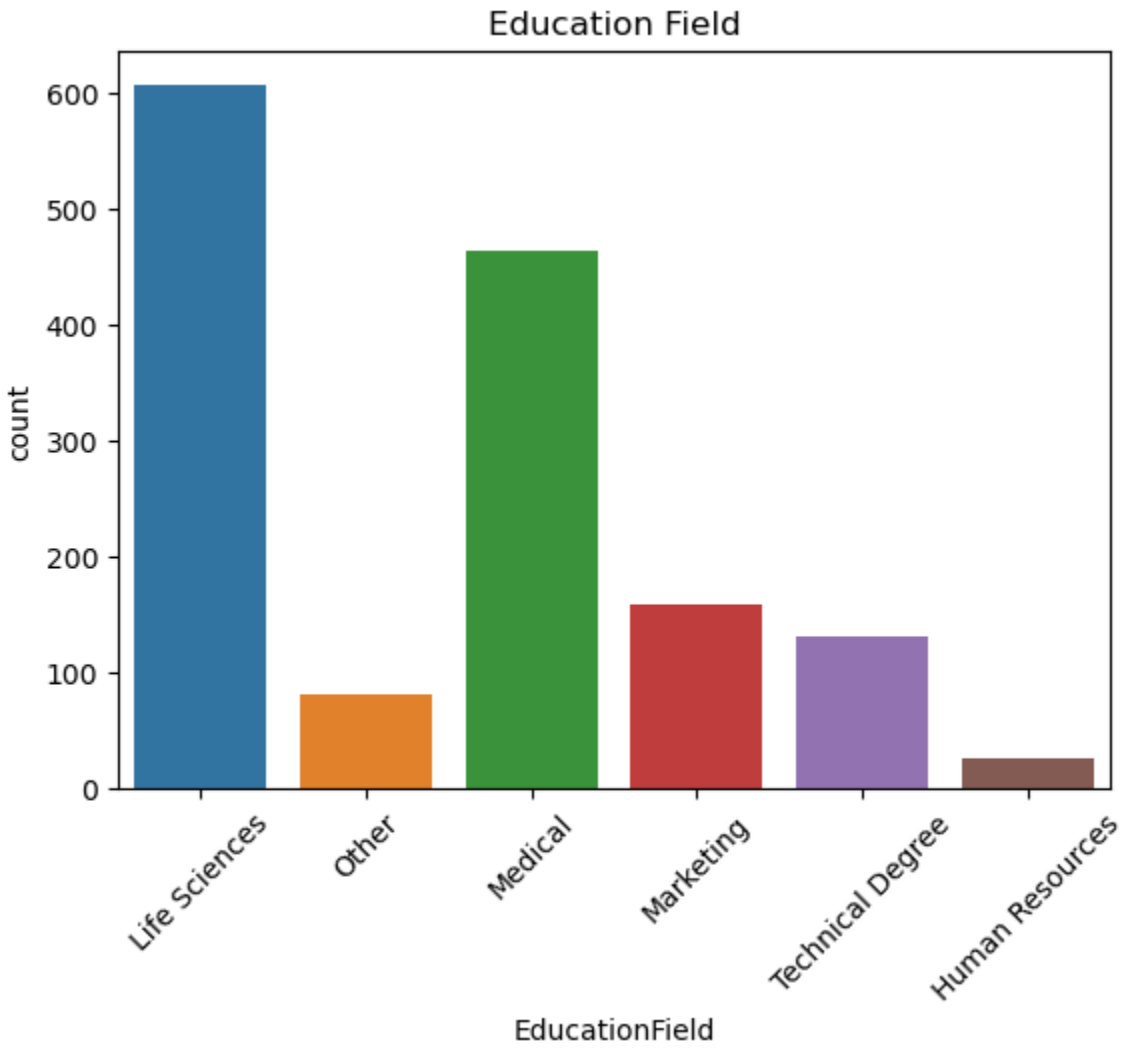


- Number of Male employees are relatively more when comapred to female employees

---

Education Field

In [25]:
```
sns.countplot(df['EducationField'])
plt.title('Education Field')
plt.xticks(rotation=45)
plt.show()
```

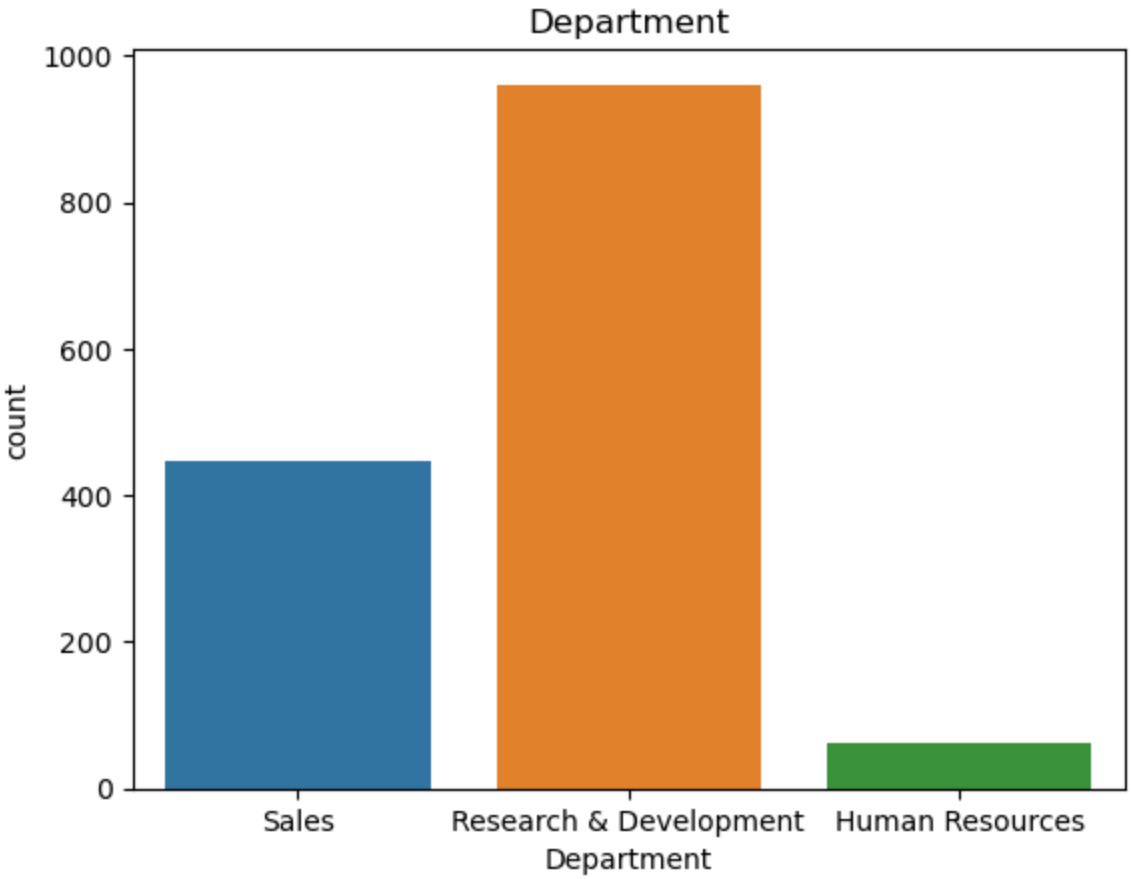


- Employees are moslty from Life sciences or medical as their educaiton background

---

Department

In [26]:
```
sns.countplot(df['Department'])
plt.title('Department')
plt.show()
```

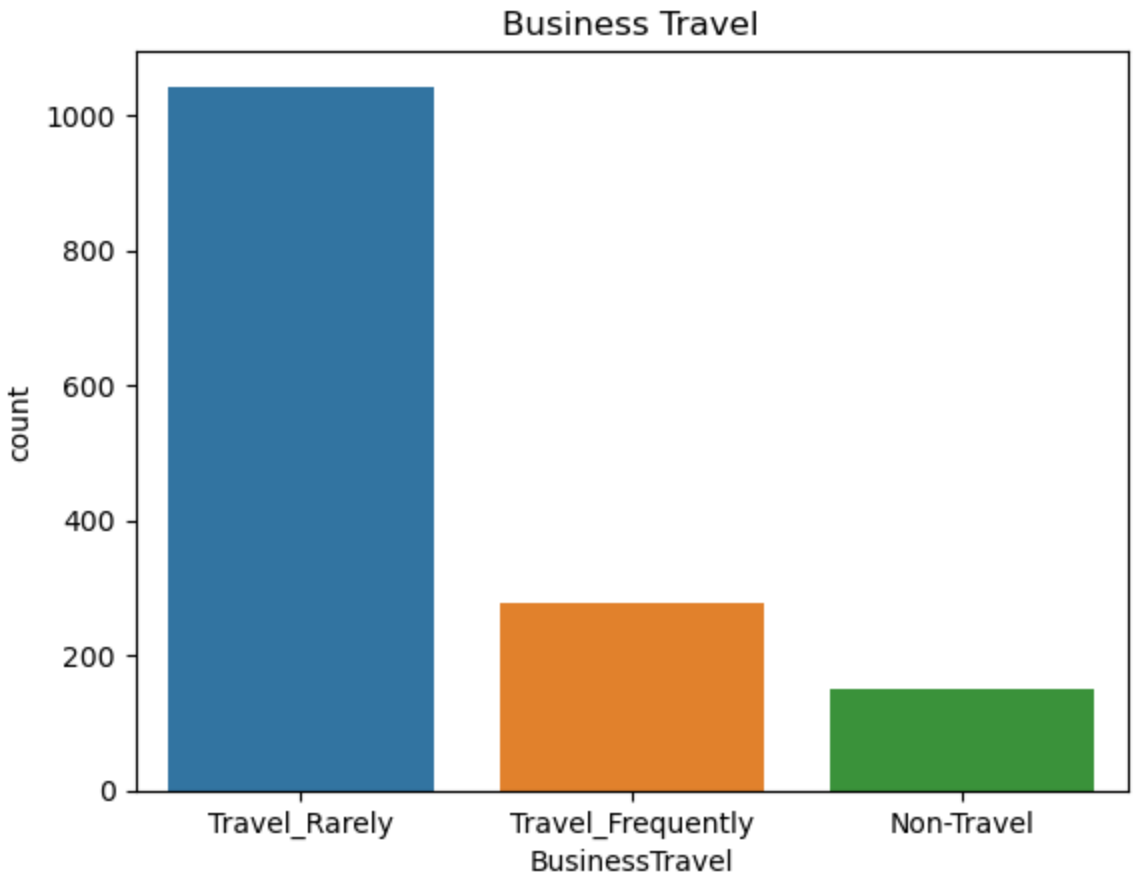


- R&D Department has highest number of employee count followed by sales department

---

Business Travel

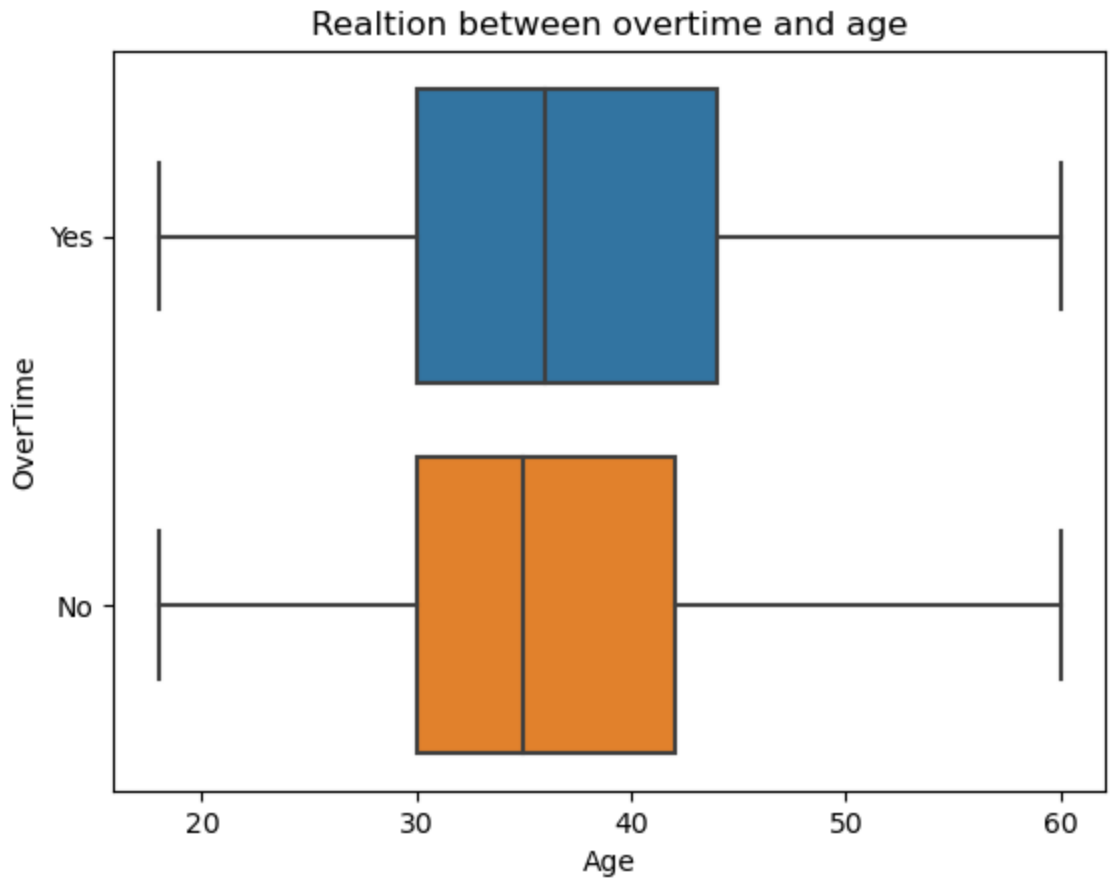
In [27]:
```
sns.countplot(df['BusinessTravel'])
plt.title('Business Travel')
plt.show()
```



- There are business travel very rarely or else no travel

Relation Between overtime and age

In [33]:
```
sns.boxplot(df['Age'],df['OverTime'])
plt.title('Realtion between overtime and age')
plt.show()
```

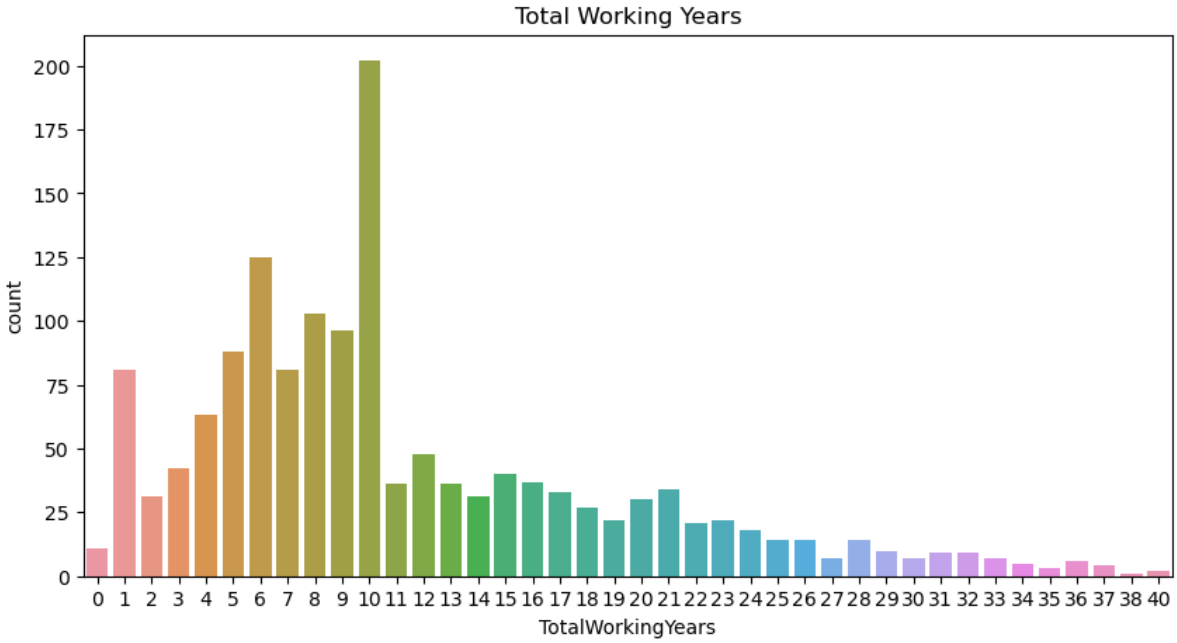


- Thought he age range of overtime and non-overtime employees is same the median age of overtime employees is slightly higher

---

Total Woking years

In [38]:
```
plt.figure(figsize=(10,5))
sns.countplot(df['TotalWorkingYears'])
plt.title('Total Working Years')
plt.show()
```

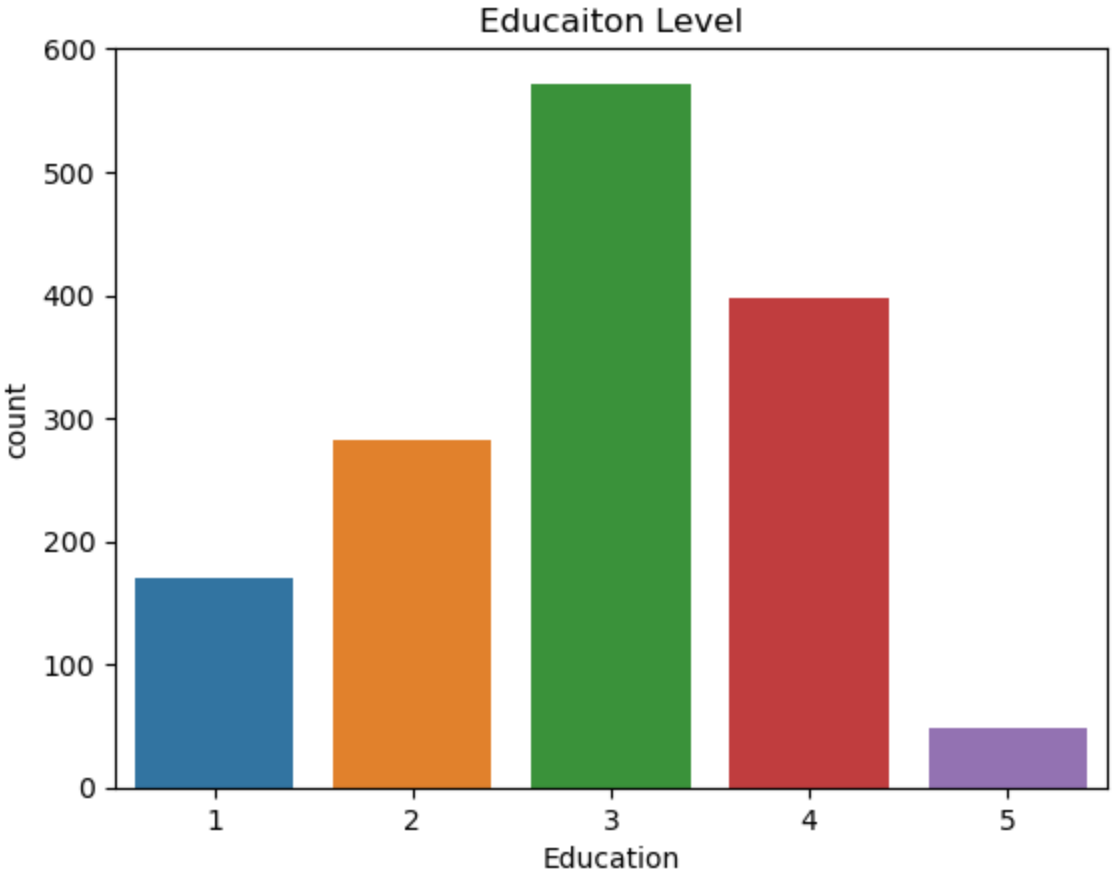


- The work experince of employees with 10 years are most in the company followed by 6 and 8 years of experience

---

Education Level

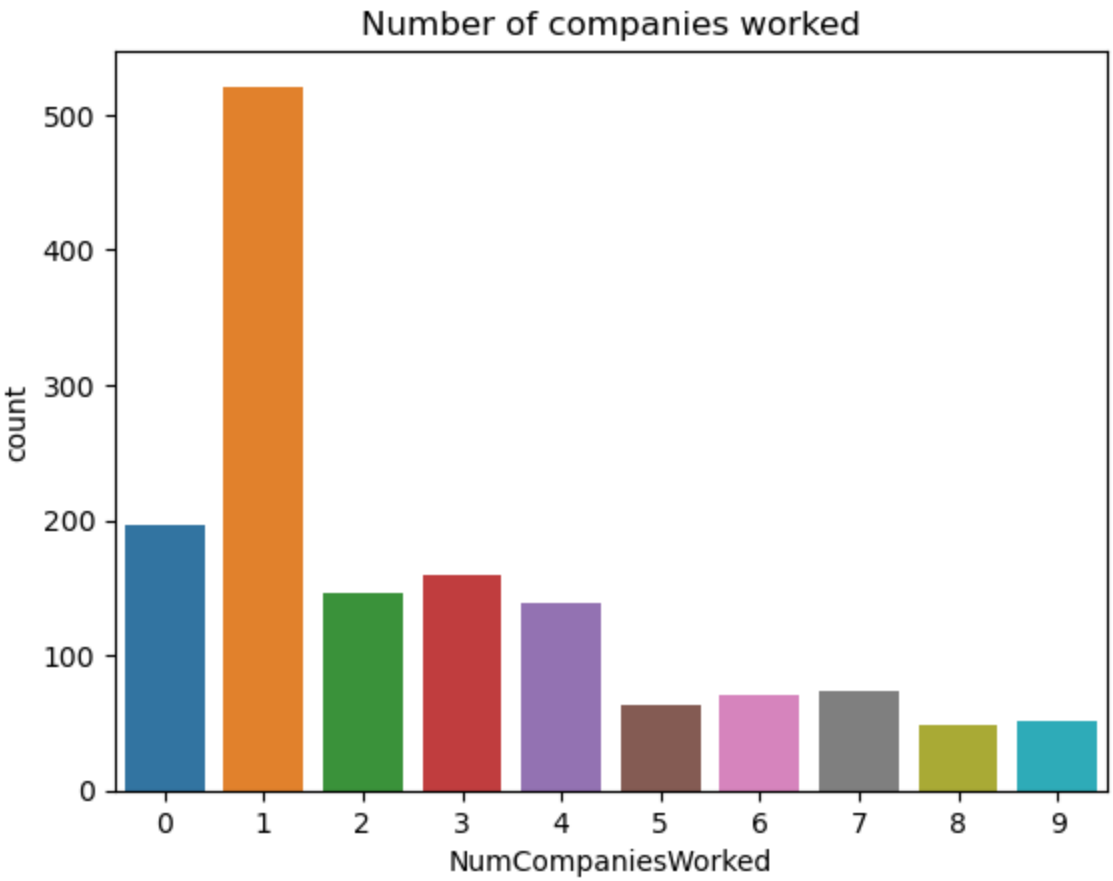
In [39]:
```
sns.countplot(df['Education'])
plt.title('Educaiton Level')
plt.show()
```



- Mostly employees are of level 3 or 4 on education

Number of Companies worked

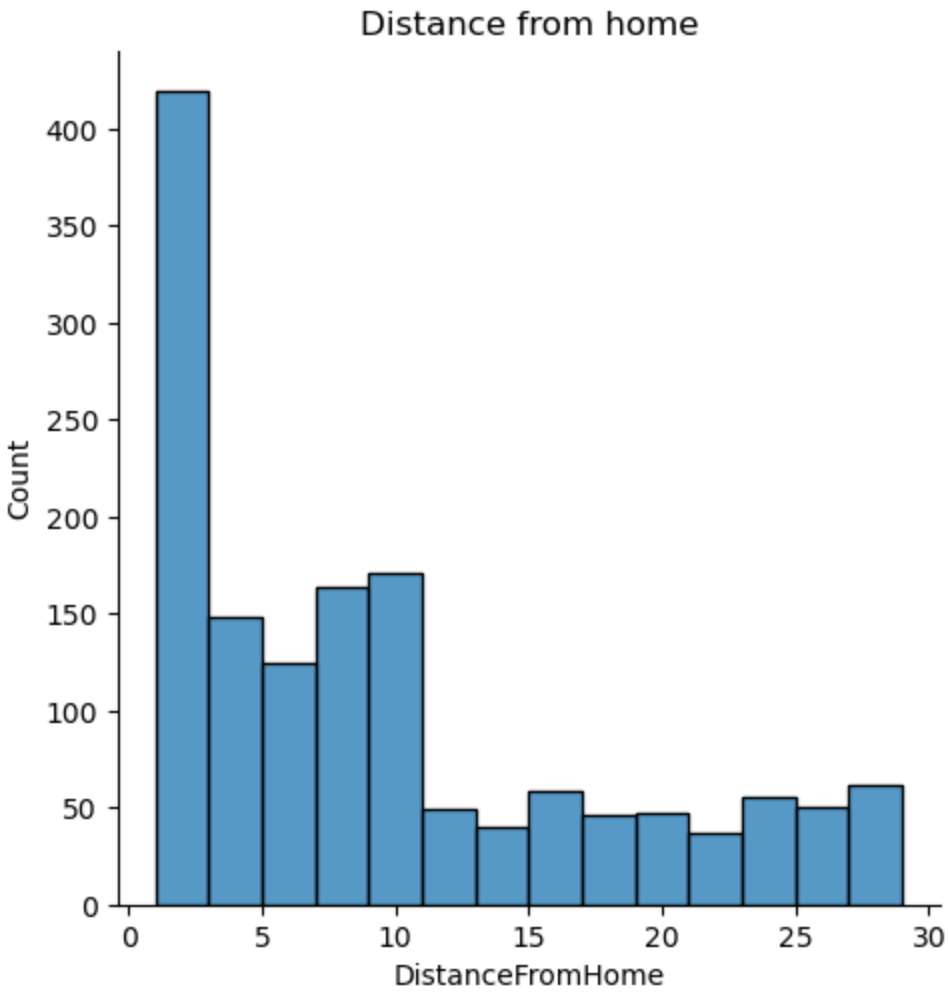
In [40]:
```
sns.countplot(df['NumCompaniesWorked'])
plt.title('Number of companies worked')
plt.show()
```



- There more number number of employees with number of companies they have worked as 1

Distance from home

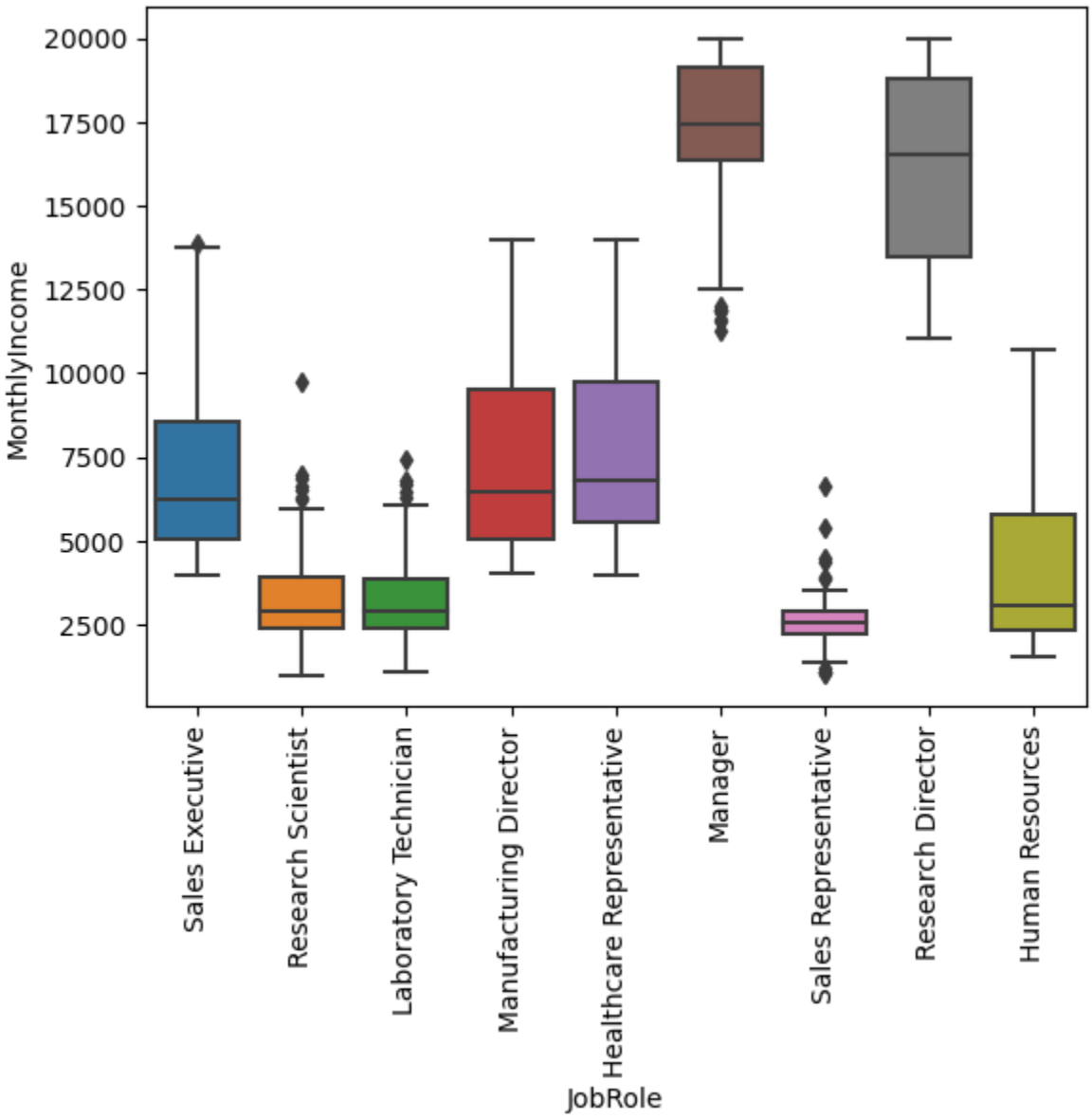
In [42]:
```
sns.displot(df['DistanceFromHome'])
plt.title('Distance from home')
plt.show()
```



- The distance from home to office is very near for most of the emoloyees

---

In [48]:
```
# Income by job role
sns.boxplot(df['JobRole'],df['MonthlyIncome'])
plt.xticks(rotation=90)
plt.show()
```



- The monthly income of Managers ans Research Directors are highest and Sales representatives are lowest

In [ ]:

HR Analytics Dashboard

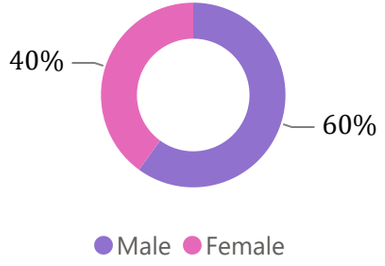
Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

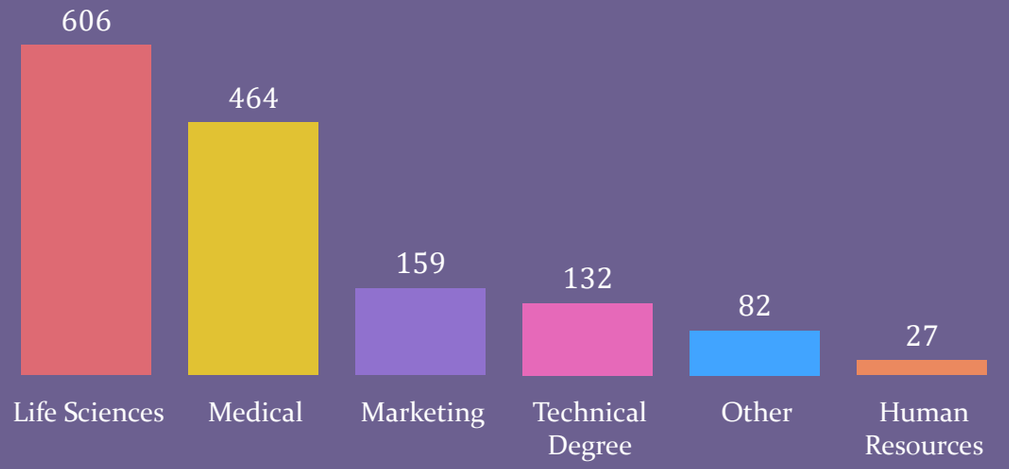
Gender
- [ ] Female
- [ ] Male

1470
Head Count

Headcount by Gender



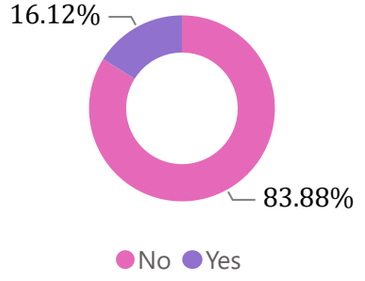
Employees by Education Field



Employees by Department

237
Total attrition

Attrition vs Retention



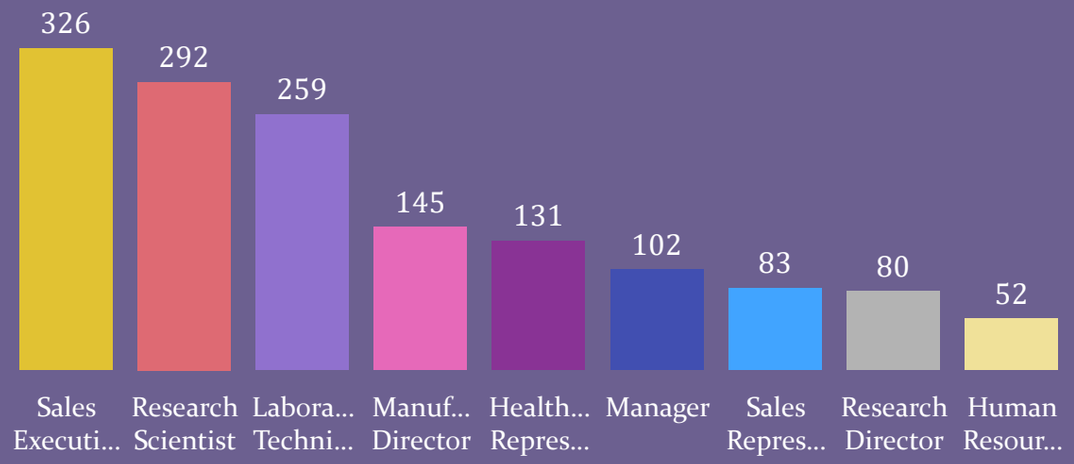
HR Analytics Dashboard

Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

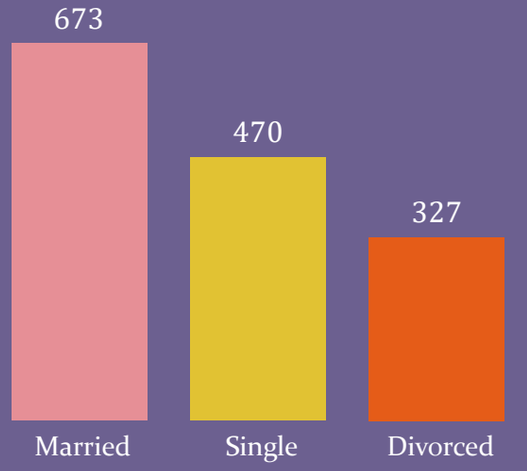
EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

Gender
- [ ] Female
- [ ] Male

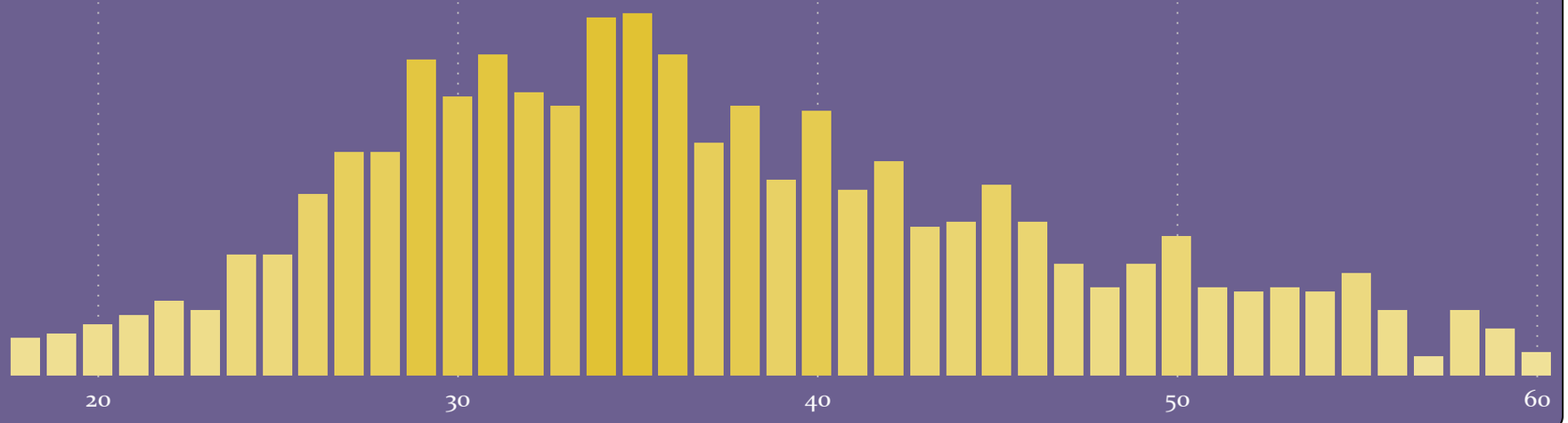
Employees by Job Role



Employees by Marital Status



Employees by Age



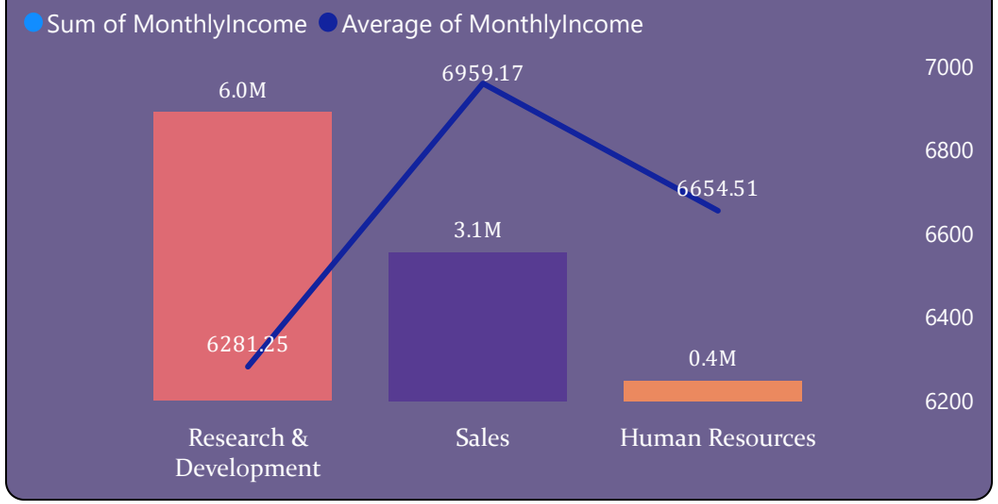
HR Analytics Dashboard

Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

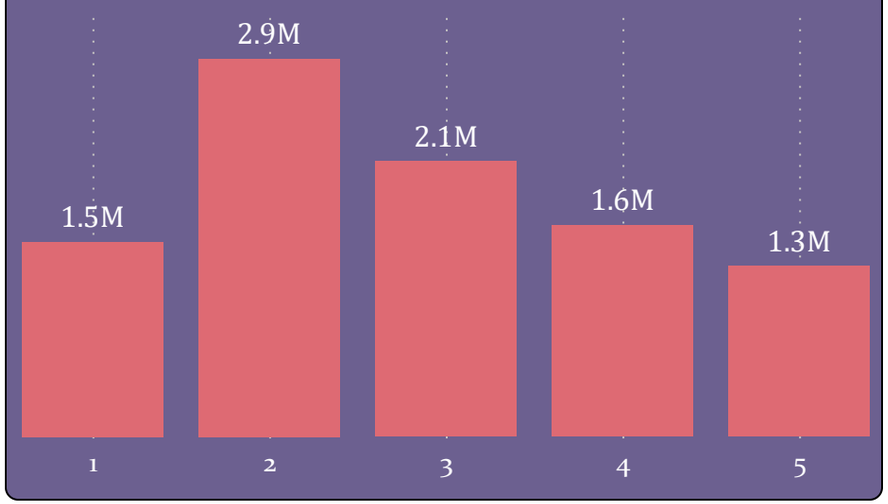
EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

Gender
- [ ] Female
- [ ] Male

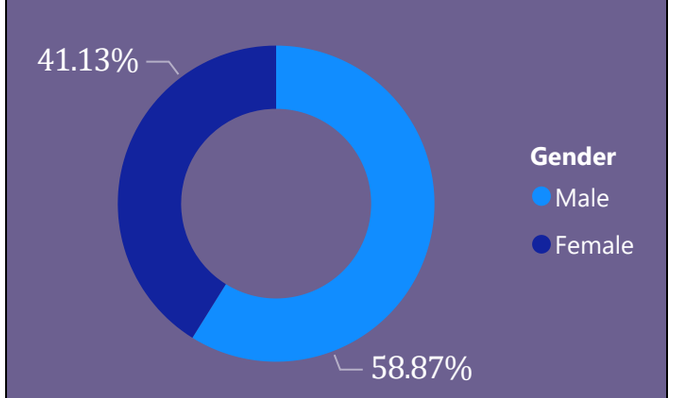
Monthly and Average Inocme by Department



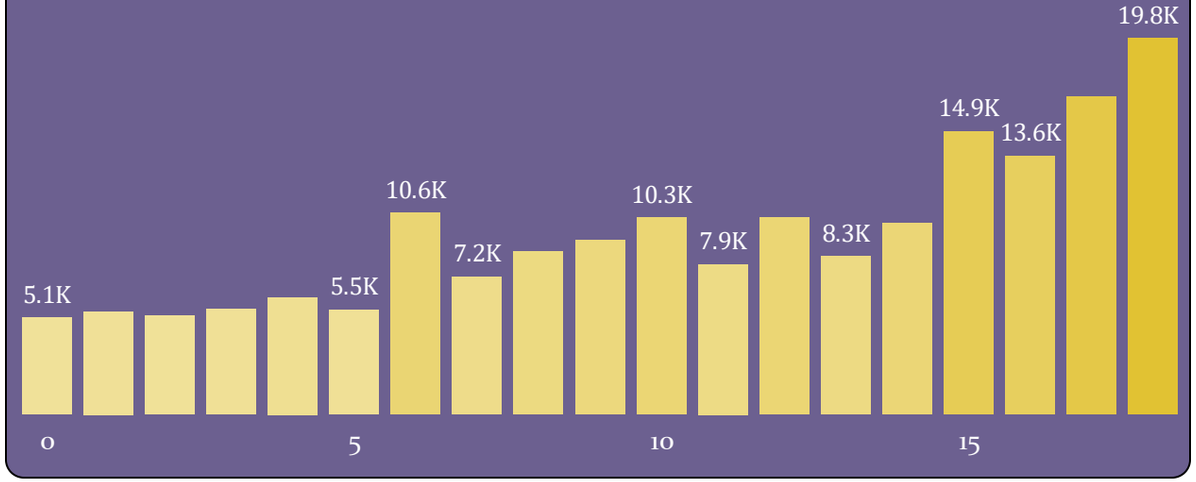
Sum of MonthlyIncome by JobLevel



Monthly Income by Gender



Average Monthly Income by YearsInCurrentRole



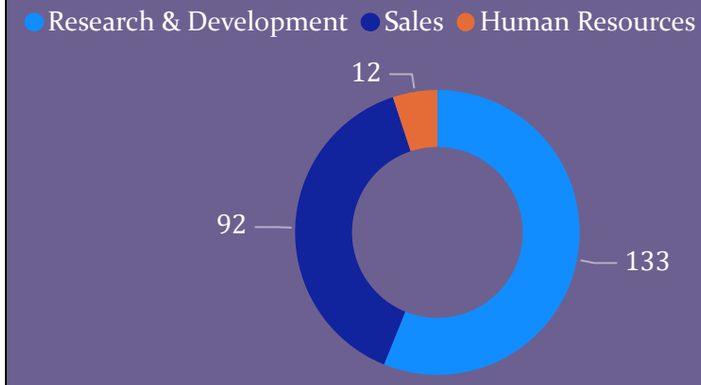
HR Analytics Dashboard

Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

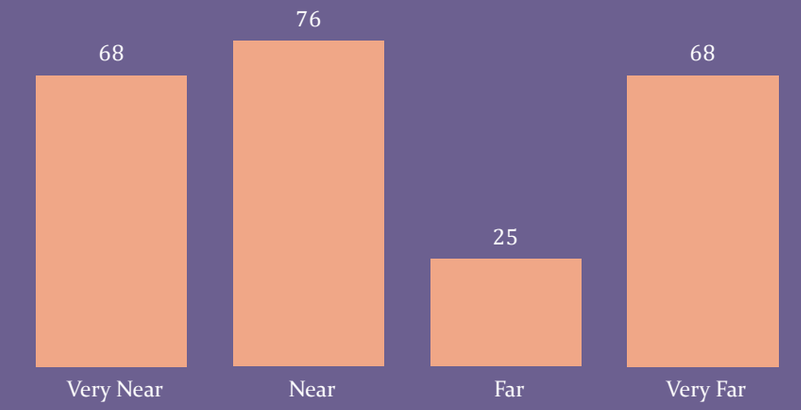
EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

Gender
- [ ] Female
- [ ] Male

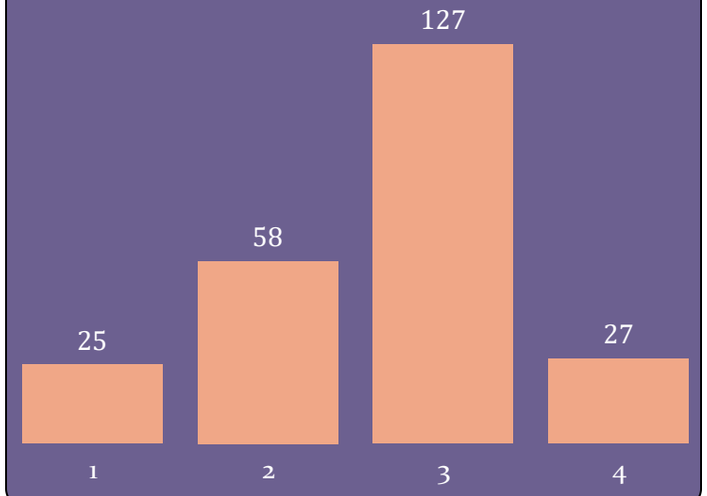
Attrition by Department and gender



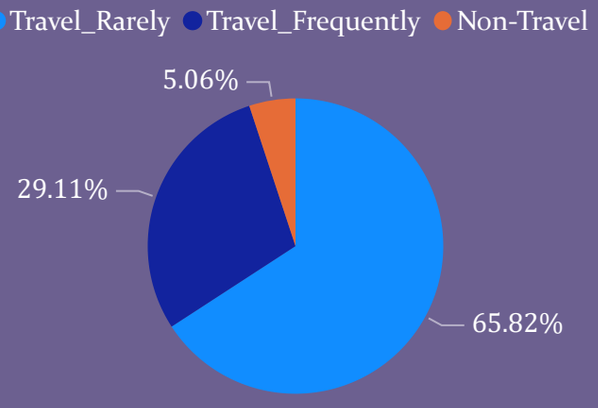
Attrition by Distance Category



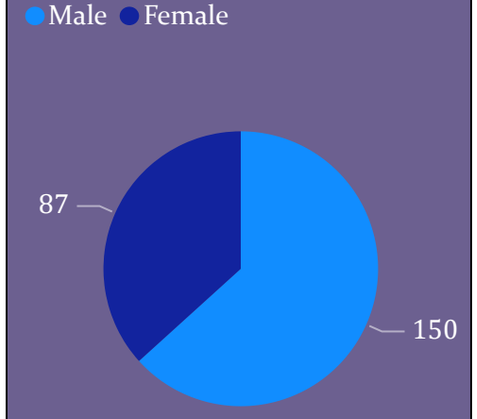
Attrition by Work Life Balance



Total attrition by BusinessTravel



Attrition by Gender



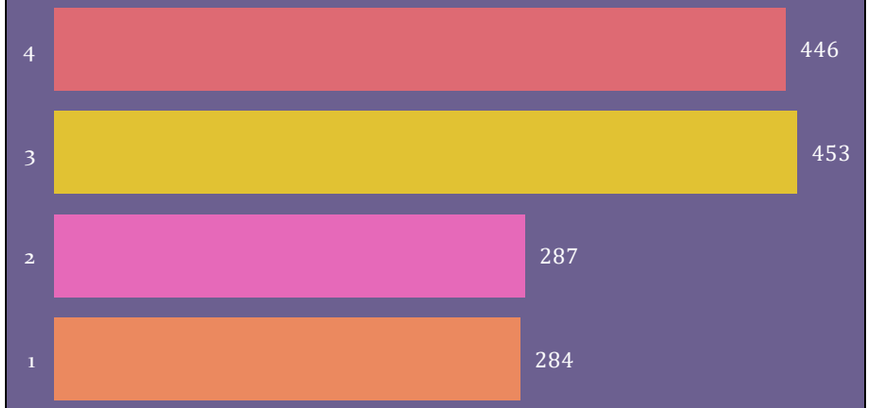
HR Analytics Dashboard

Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

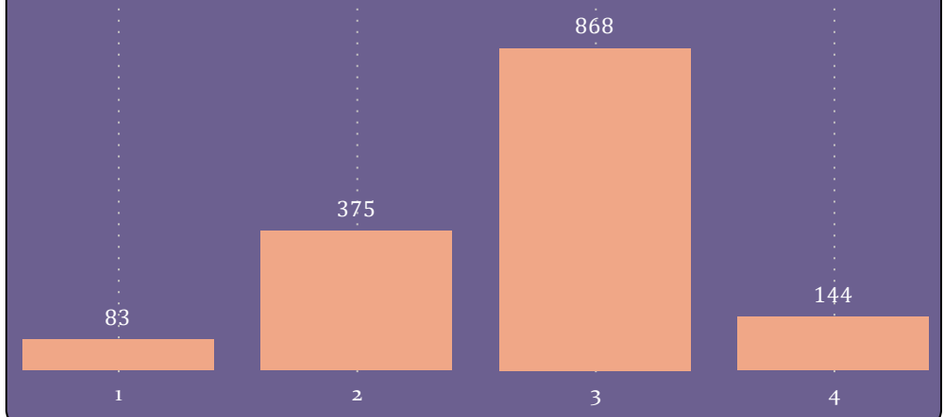
EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

Gender
- [ ] Female
- [ ] Male

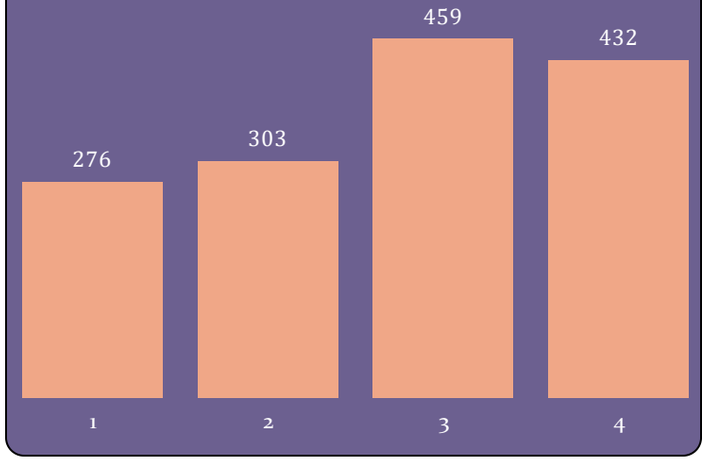
Employees by Job Satisfaction



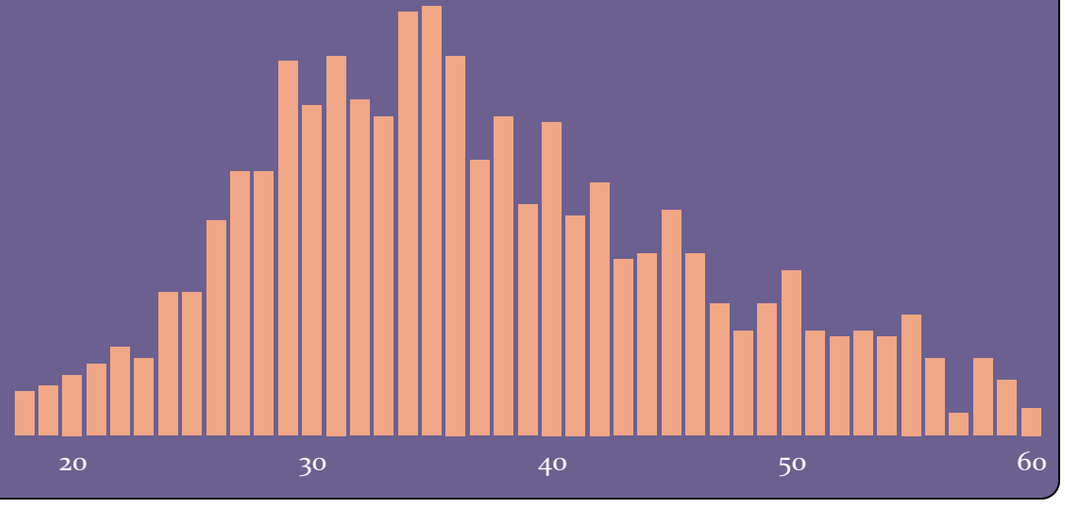
Employees by Job Involment



Employees by Relation Satisfaction



Over Time by Age



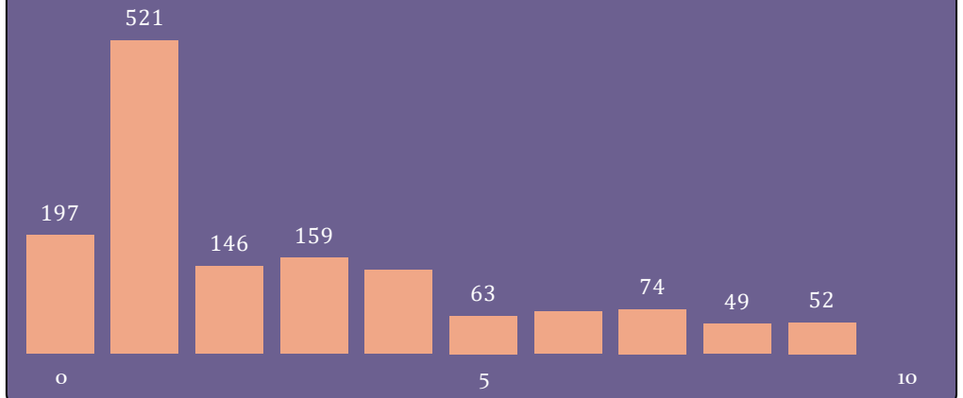
HR Analytics Dashboard

Department
- [ ] Human Resources
- [ ] Research & Developm...
- [ ] Sales

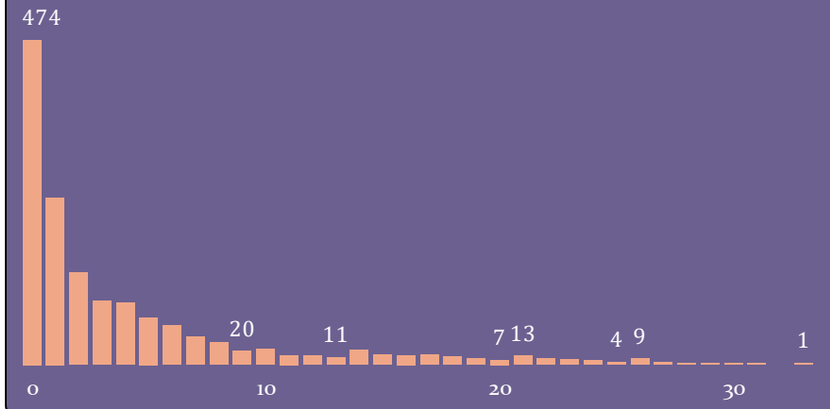
EducationField
- [ ] Human Resources
- [ ] Life Sciences
- [ ] Marketing
- [ ] Medical
- [ ] Other
- [ ] Technical Degree

Gender
- [ ] Female
- [ ] Male

Number of Companies previous worked



Employees with Previous Experince



Total Working Years