In [114…
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
from warnings import filterwarnings
filterwarnings("ignore")
pd.set_option("display.max_columns",None)
```

In [115…
```
A=pd.read_csv("C:/Users/ASUS/Downloads/WA_Fn-UseC_-HR-Employee-Attrition.csv")
```

In [116…
```
A.head()
```

Out[116]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumbe
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	

In [117…
```
A.isna()
```

Out[117]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNum
0	False	False	False	False	False	False	False	False	False	Fa
1	False	False	False	False	False	False	False	False	False	Fa
2	False	False	False	False	False	False	False	False	False	Fa
3	False	False	False	False	False	False	False	False	False	Fa
4	False	False	False	False	False	False	False	False	False	Fa
...	...	...	...	...	...	...	...	...	...	
1465	False	False	False	False	False	False	False	False	False	Fa
1466	False	False	False	False	False	False	False	False	False	Fa
1467	False	False	False	False	False	False	False	False	False	Fa
1468	False	False	False	False	False	False	False	False	False	Fa
1469	False	False	False	False	False	False	False	False	False	Fa

1470 rows × 35 columns

In [118…
```
A.isna().sum()
```

Out[118]:
```
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
```

replacer

In [119…
```
from PM8 import replacer
replacer(A)
```

In [120…
```
A['Department'].mode()[0]
```

Out[120]:
```
'Research & Development'
```

In [121…
```
A['Age'].mean()
```

Out[121]:
```
36.923809523809524
```

In [122…
```
replacer(A)
```

In [123…
```
A
```

Out[123]:

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

1470 rows × 35 columns

preprocessing

In [124…
```
from PM8 import preprocessing
Anew=preprocessing(A)
```

In [125…
```
Anew.head()
```

Out[125]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvem
0	0.446350	0.742527	-1.010909	-0.891688	0.0	-1.701283	-0.660531	1.383138	0.379
1	1.322365	-1.297775	-0.147150	-1.868426	0.0	-1.699621	0.254625	-0.240677	-1.026
2	0.008343	1.414363	-0.887515	-0.891688	0.0	-1.696298	1.169781	1.284725	-1.026
3	-0.429664	1.461466	-0.764121	1.061787	0.0	-1.694636	1.169781	-0.486709	0.379
4	-1.086676	-0.524295	-0.887515	-1.868426	0.0	-1.691313	-1.575686	-1.274014	0.379

In [126…
```
for i in A.columns:
    print(i, len(A[i].unique()))
```

```
Age 43
Attrition 2
BusinessTravel 3
DailyRate 886
Department 3
DistanceFromHome 29
Education 5
EducationField 6
EmployeeCount 1
EmployeeNumber 1470
EnvironmentSatisfaction 4
Gender 2
HourlyRate 71
JobInvolvement 4
JobLevel 5
JobRole 9
JobSatisfaction 4
MaritalStatus 3
MonthlyIncome 1349
MonthlyRate 1427
NumCompaniesWorked 10
Over18 1
OverTime 2
PercentSalaryHike 15
PerformanceRating 2
RelationshipSatisfaction 4
StandardHours 1
StockOptionLevel 4
TotalWorkingYears 40
TrainingTimesLastYear 7
WorkLifeBalance 4
YearsAtCompany 37
YearsInCurrentRole 19
YearsSinceLastPromotion 16
YearsWithCurrManager 18
```

In [127…
```
A.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
```

Define X & Y

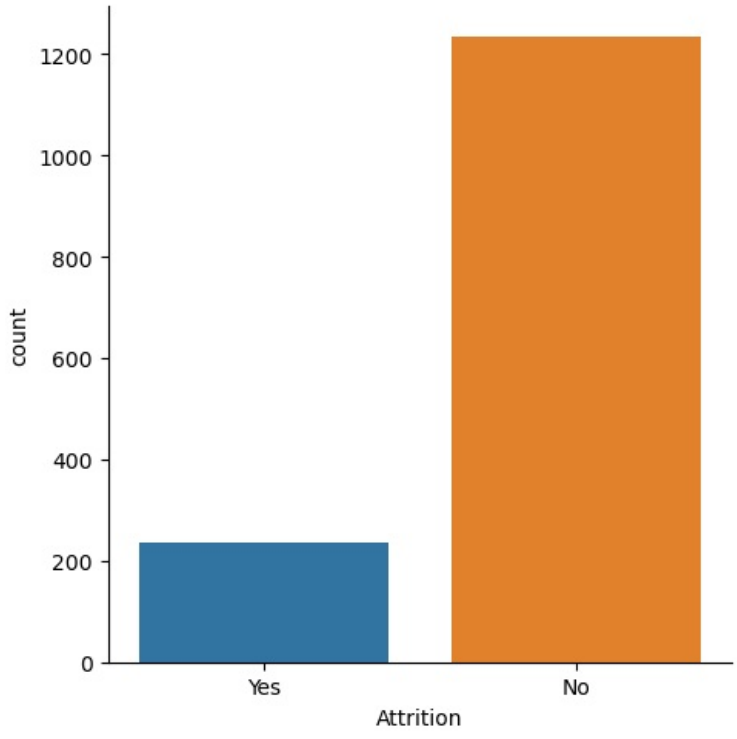
In [128…
```
X = A.drop(['EmployeeNumber',"MonthlyRate"], axis=1)
Y = A[["MonthlyRate"]]
```

Exploratory data analysis Univariate Analysis

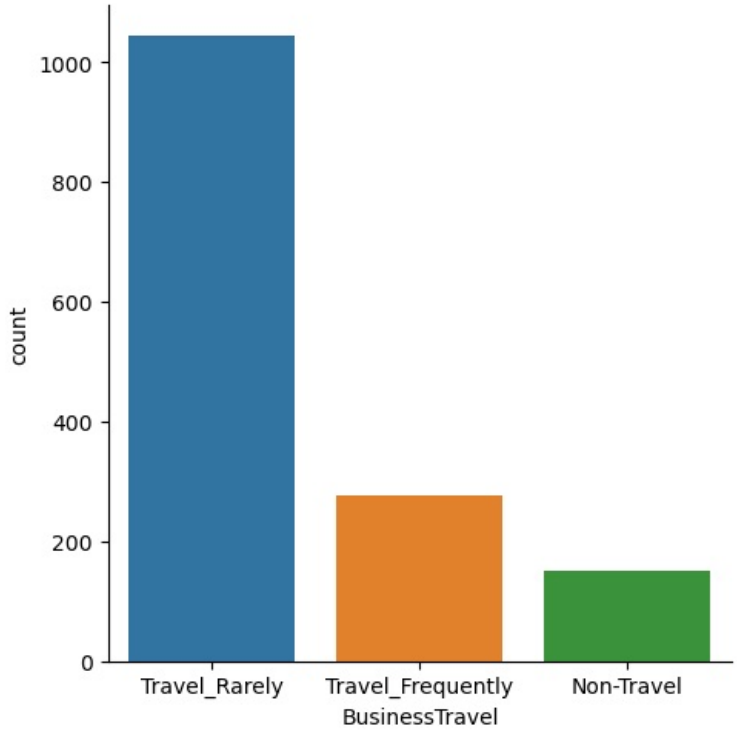
In [137…
```
cat_Data=A.select_dtypes(include='object')

for i in cat_Data:
    plt.figure(figsize=(15, 15))
    sns.catplot(data=A,x=i,kind='count')
```

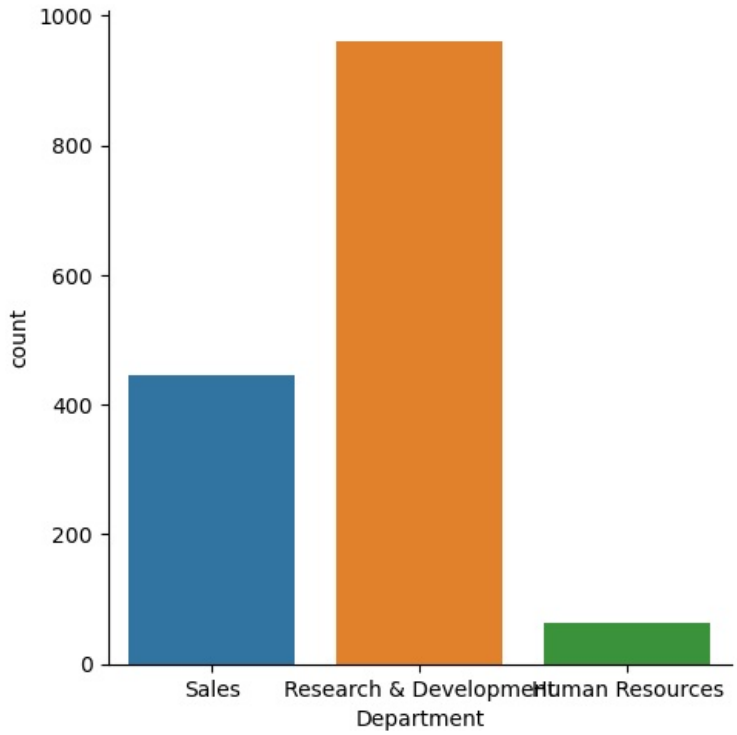
```
<Figure size 1500x1500 with 0 Axes>
```



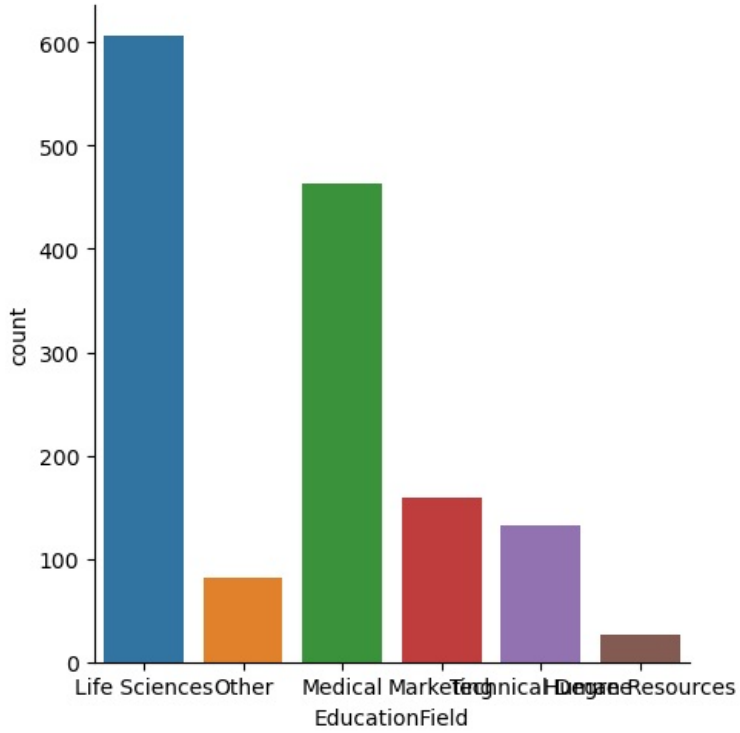
<Figure size 1500x1500 with 0 Axes>



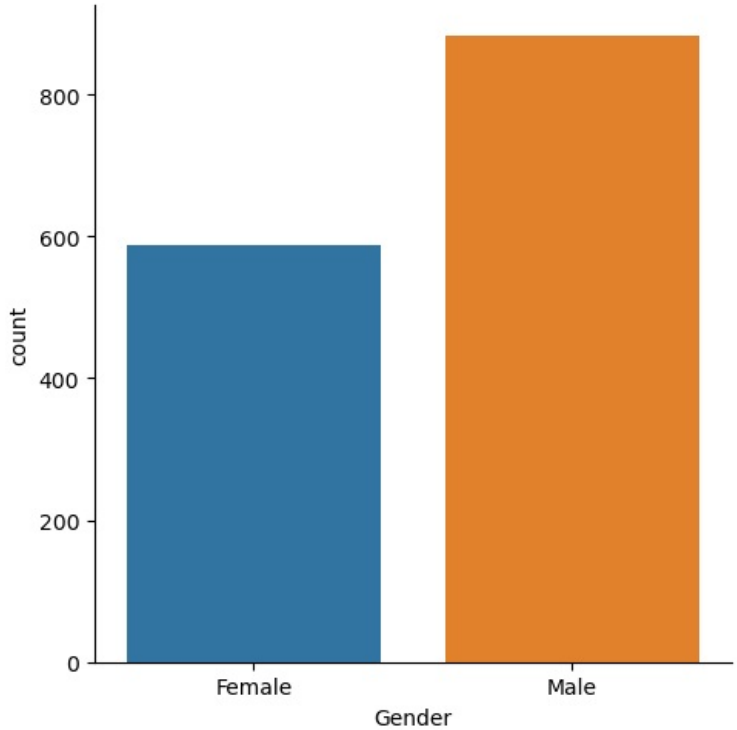
<Figure size 1500x1500 with 0 Axes>



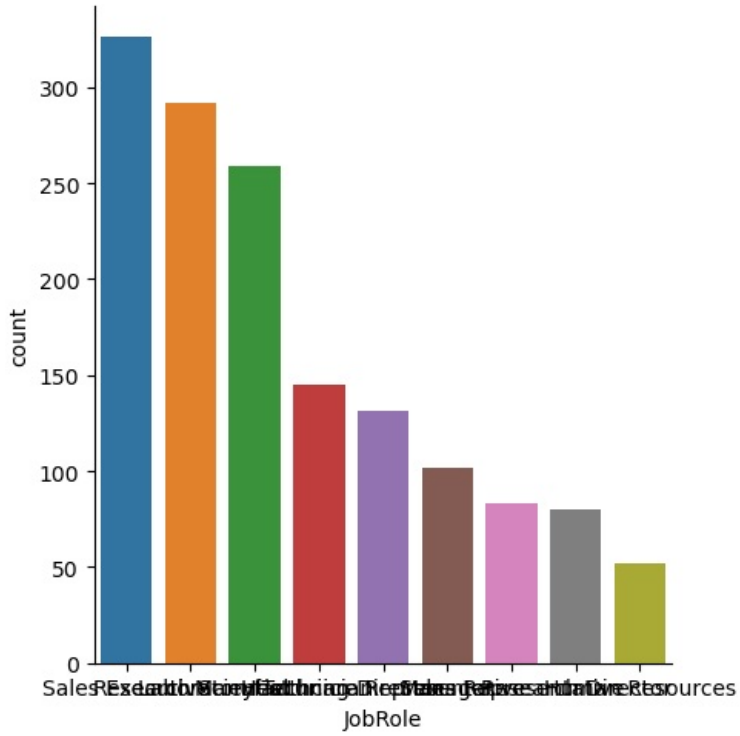
<Figure size 1500x1500 with 0 Axes>



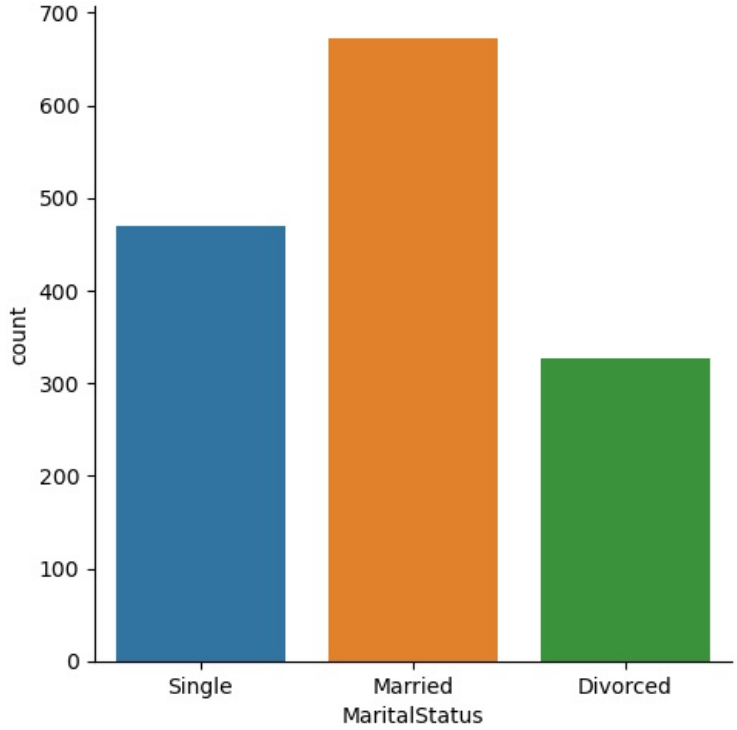
<Figure size 1500x1500 with 0 Axes>



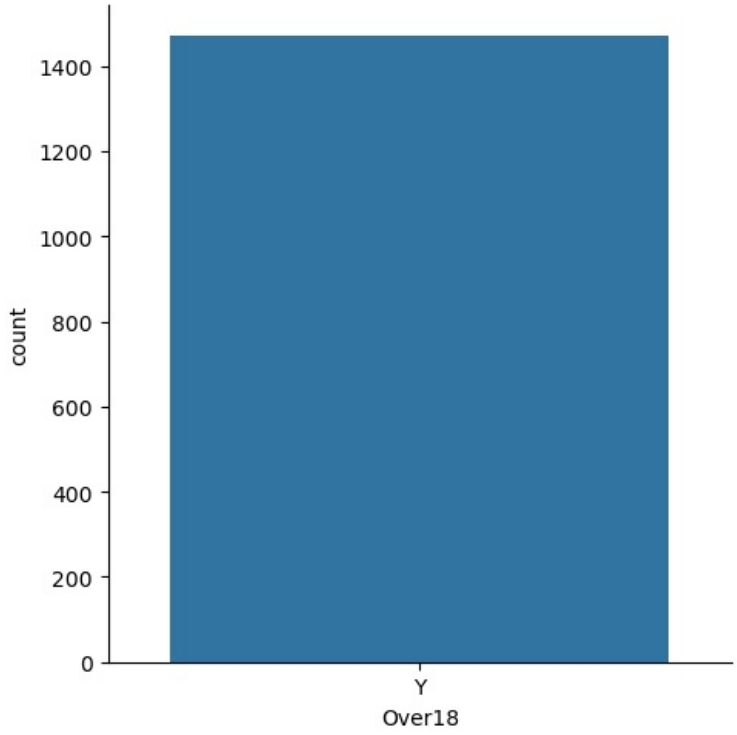
<Figure size 1500x1500 with 0 Axes>



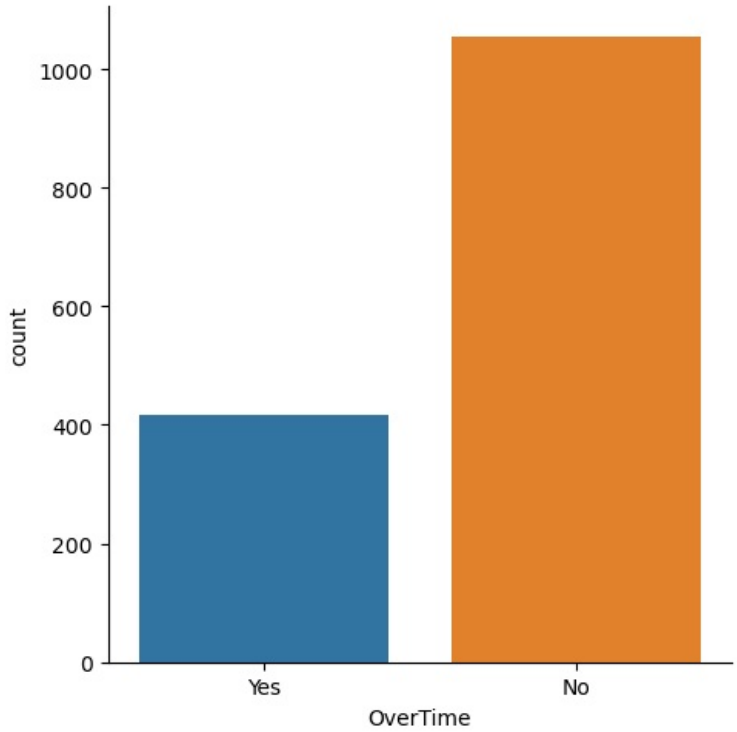
<Figure size 1500x1500 with 0 Axes>



<Figure size 1500x1500 with 0 Axes>




```
<Figure size 1500x1500 with 0 Axes>
```

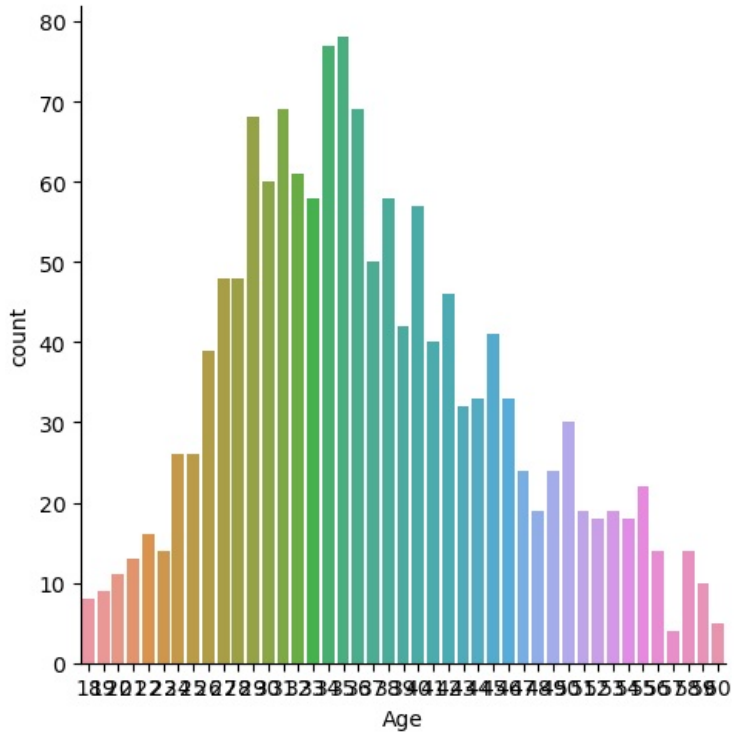


In [141…

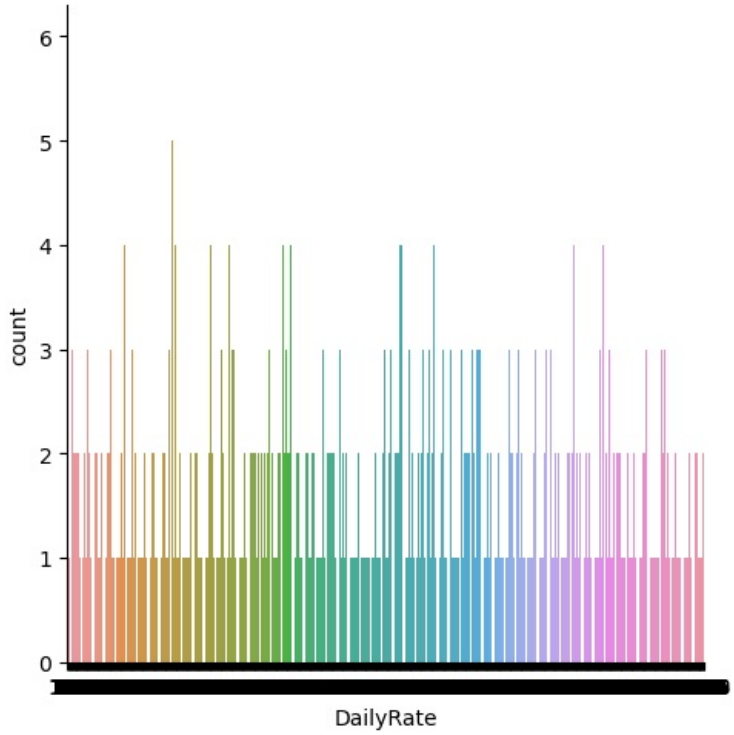
```
con_Data=A.select_dtypes(include='int')
```

```
for i in con_Data:
    plt.figure(figsize=(15, 15))
    sns.catplot(data=A,x=i,kind='count')
```

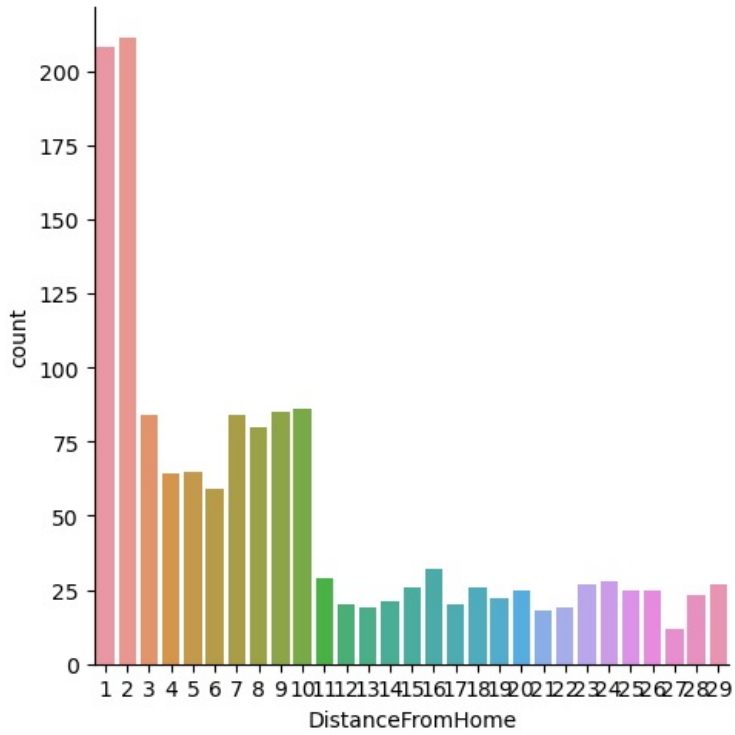
```
<Figure size 1500x1500 with 0 Axes>
```



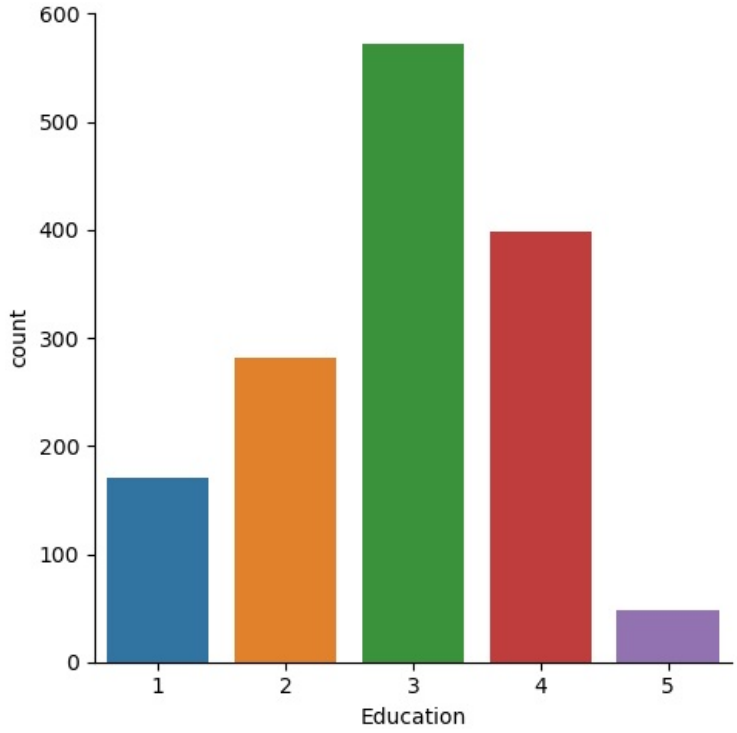

```
<Figure size 1500x1500 with 0 Axes>
```



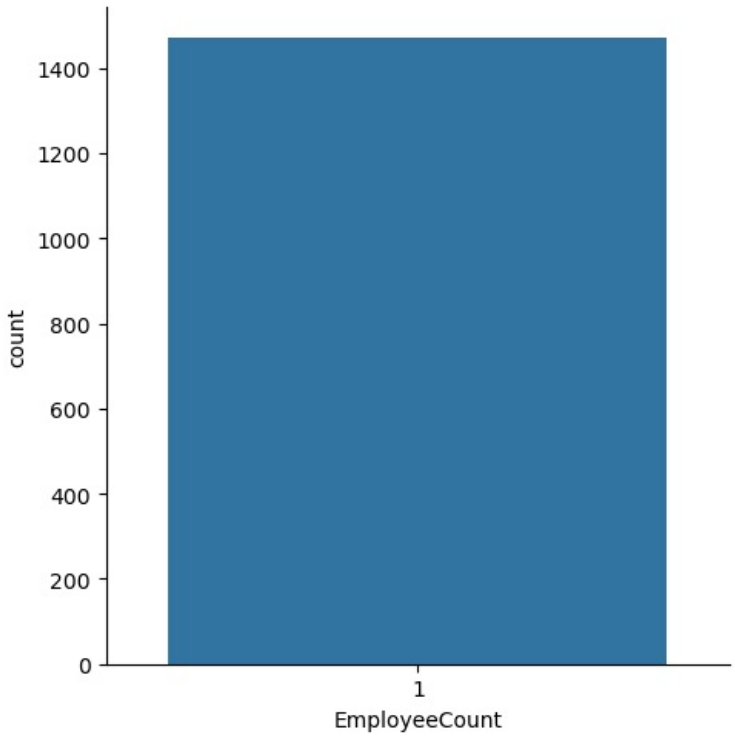

```
<Figure size 1500x1500 with 0 Axes>
```



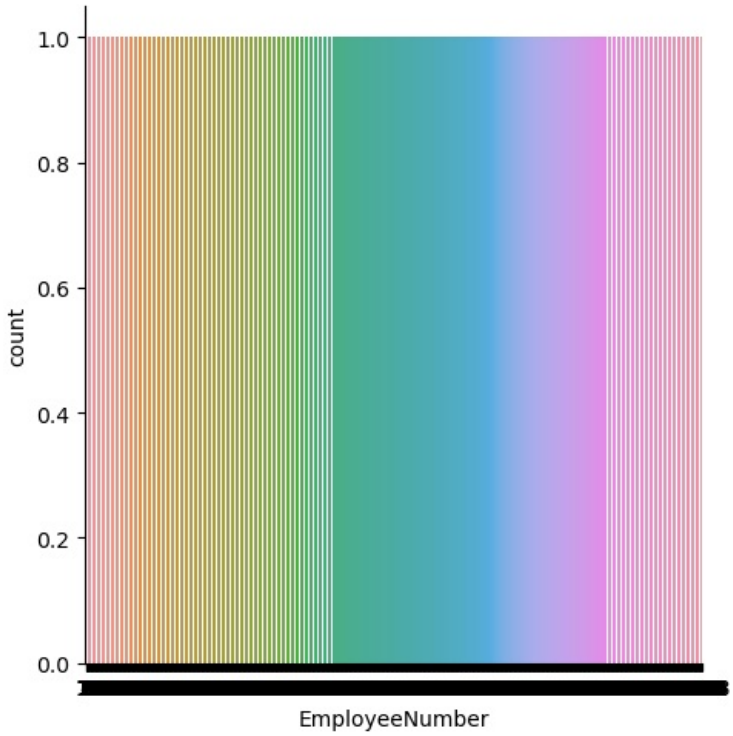
<Figure size 1500x1500 with 0 Axes>



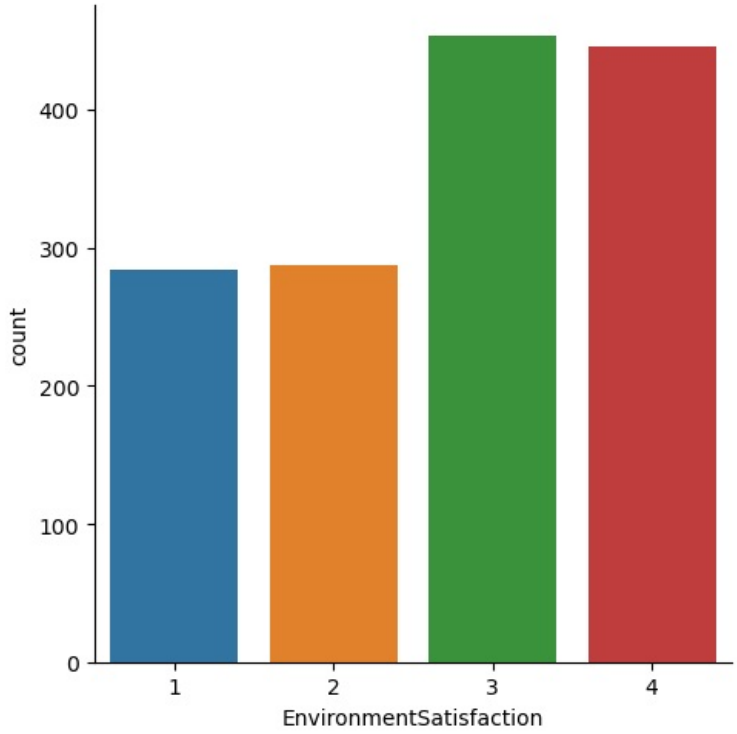
<Figure size 1500x1500 with 0 Axes>



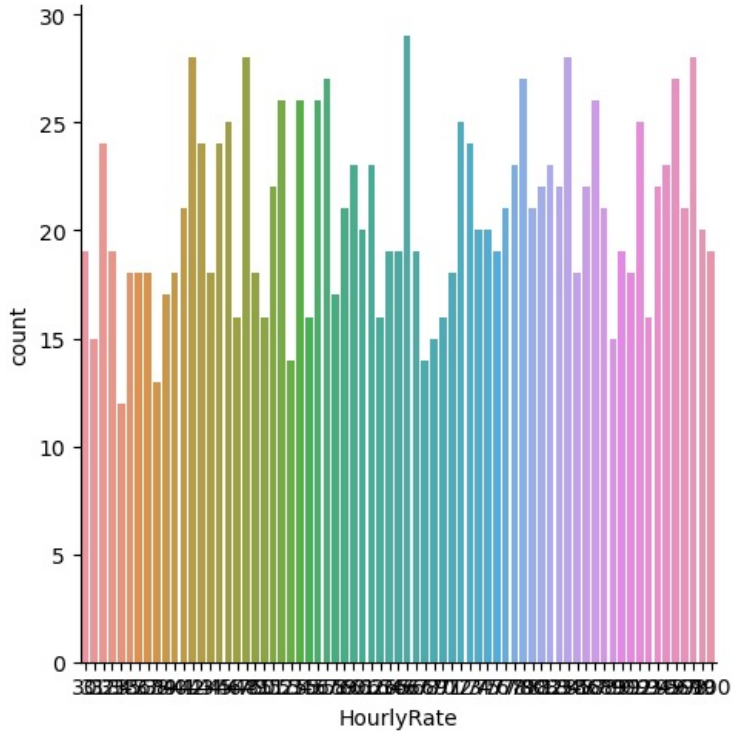

```
<Figure size 1500x1500 with 0 Axes>
```



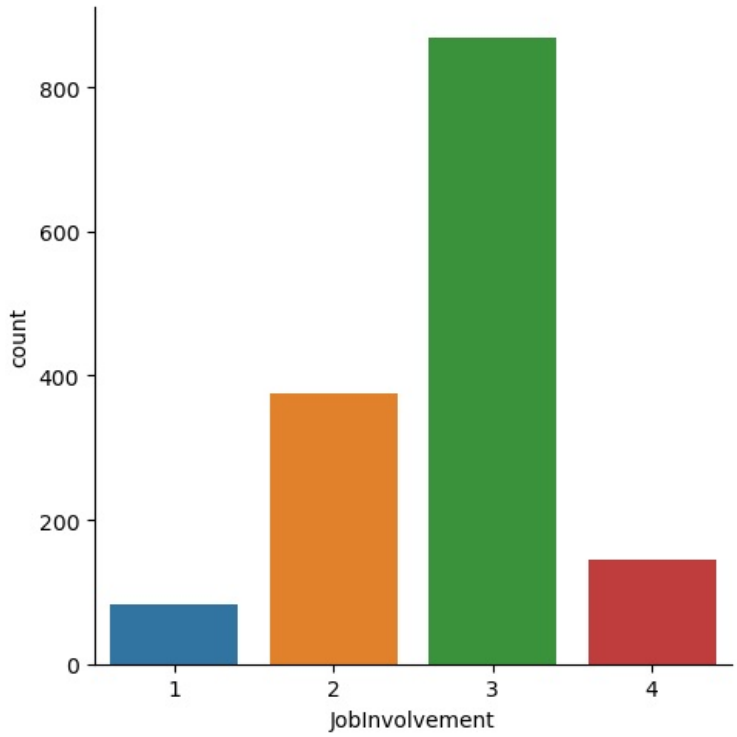

```
<Figure size 1500x1500 with 0 Axes>
```



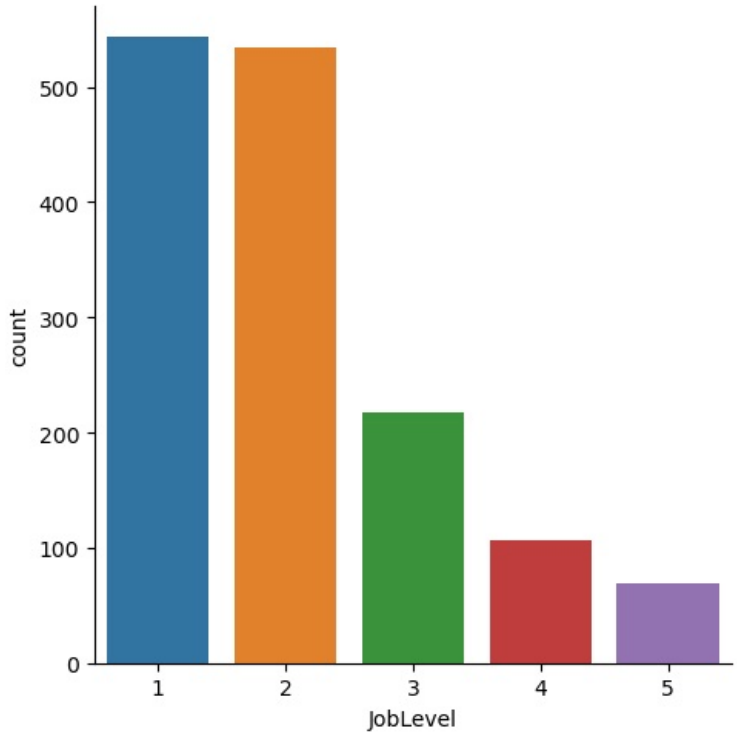

```
<Figure size 1500x1500 with 0 Axes>
```



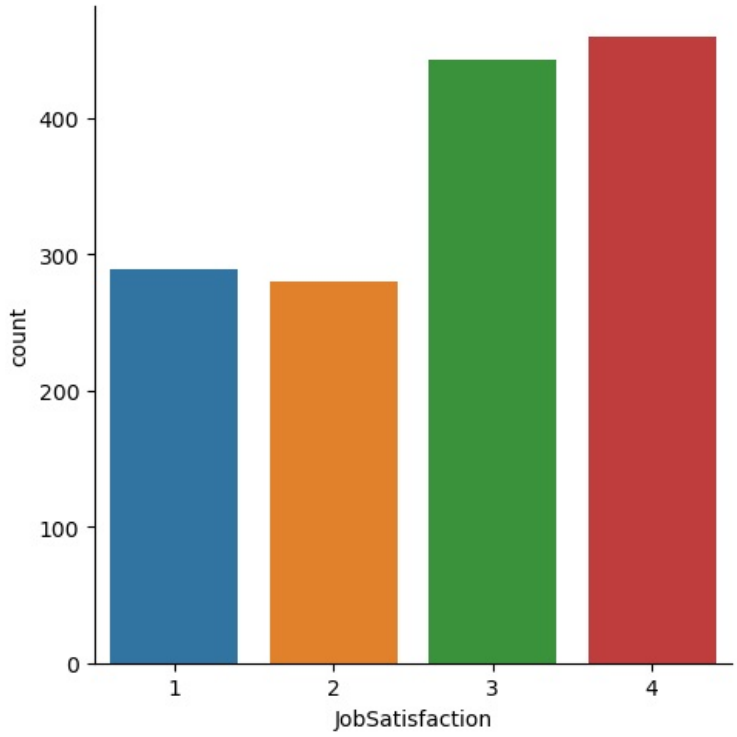

```
<Figure size 1500x1500 with 0 Axes>
```



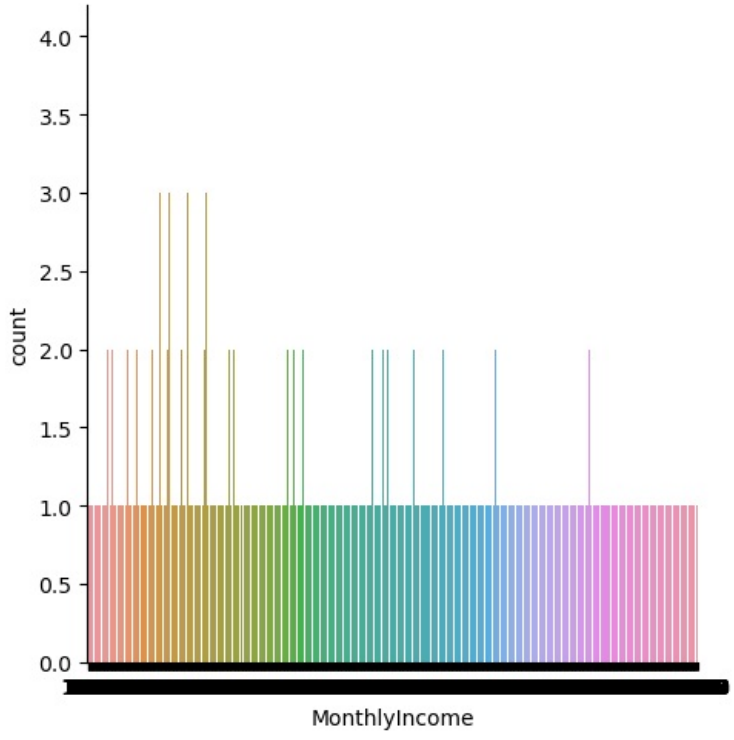
<Figure size 1500x1500 with 0 Axes>



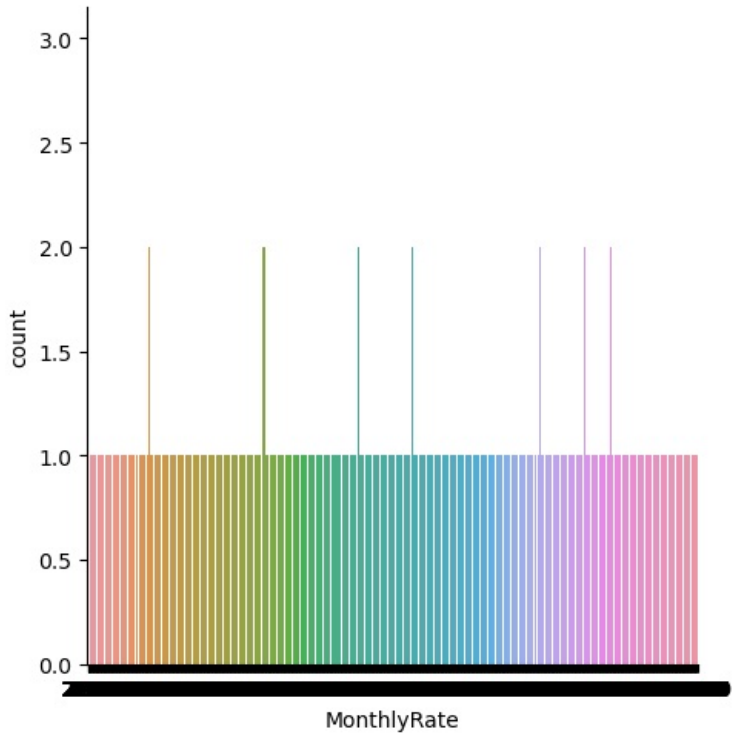
<Figure size 1500x1500 with 0 Axes>



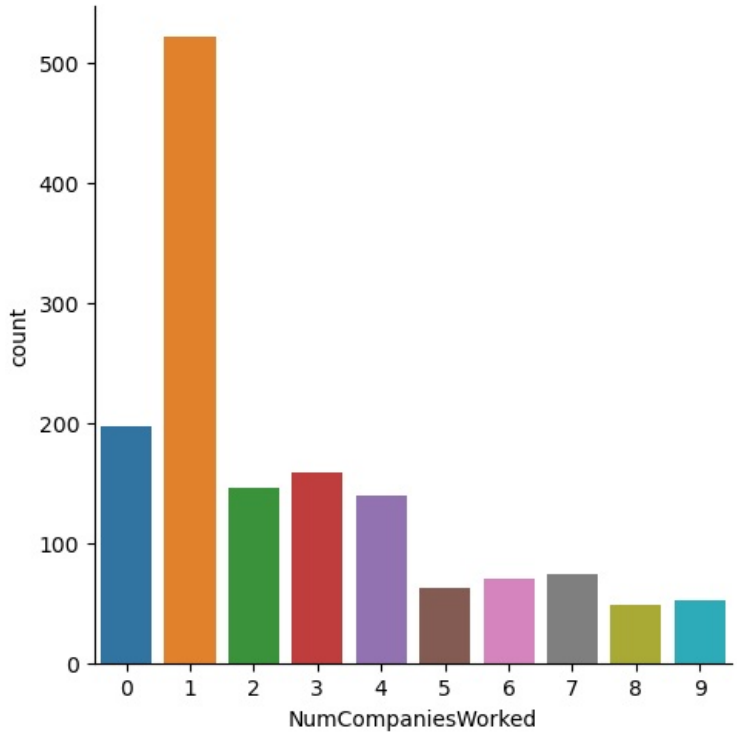

```
<Figure size 1500x1500 with 0 Axes>
```



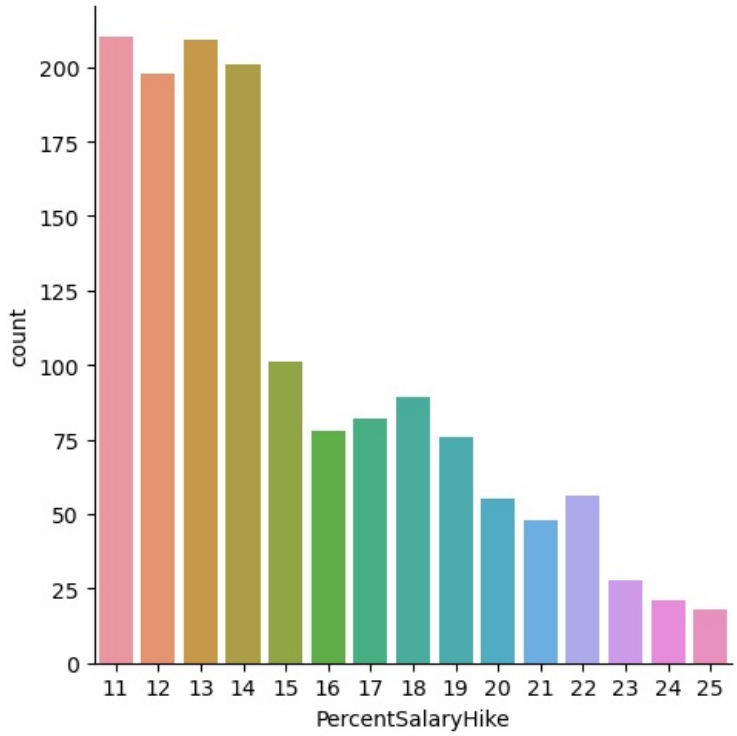

```
<Figure size 1500x1500 with 0 Axes>
```



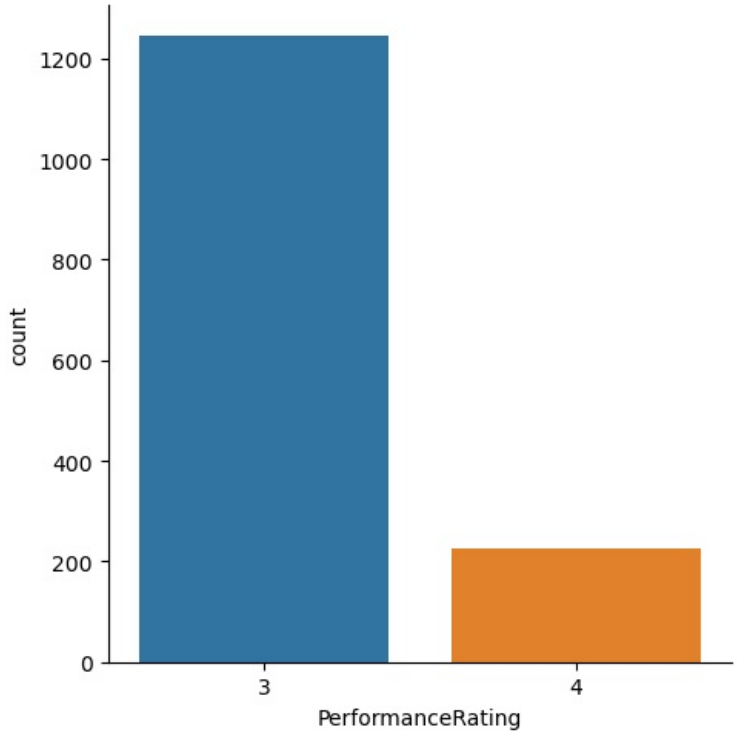
<Figure size 1500x1500 with 0 Axes>



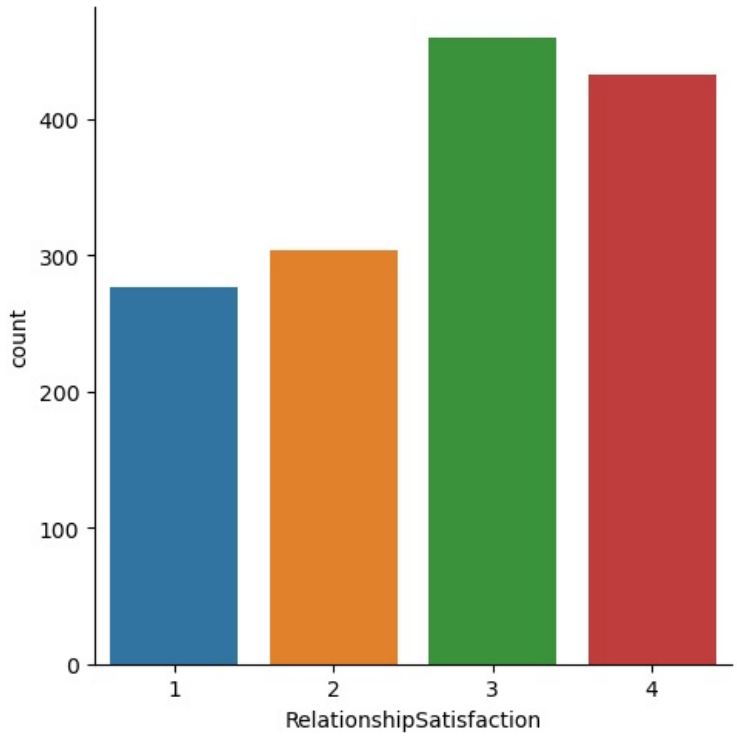
<Figure size 1500x1500 with 0 Axes>



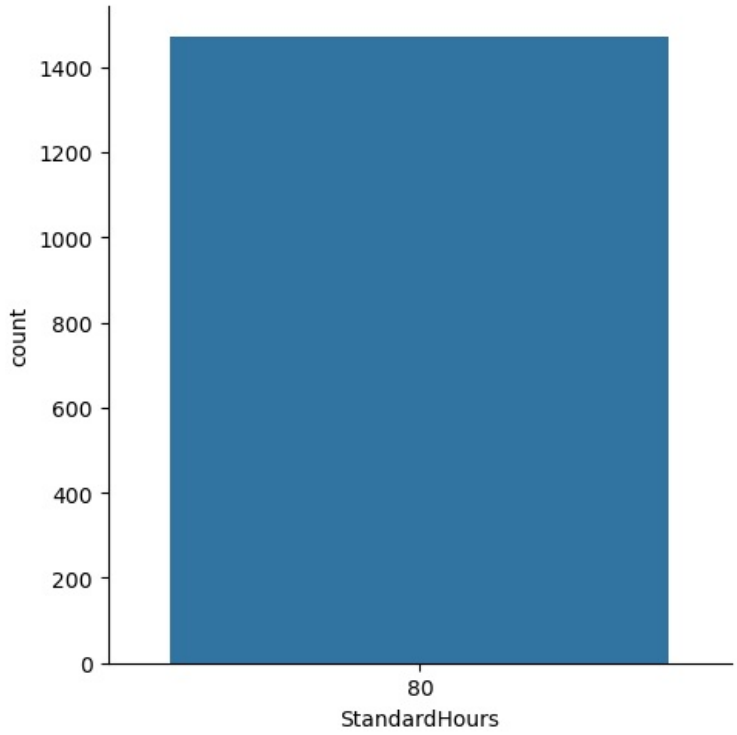

```
<Figure size 1500x1500 with 0 Axes>
```



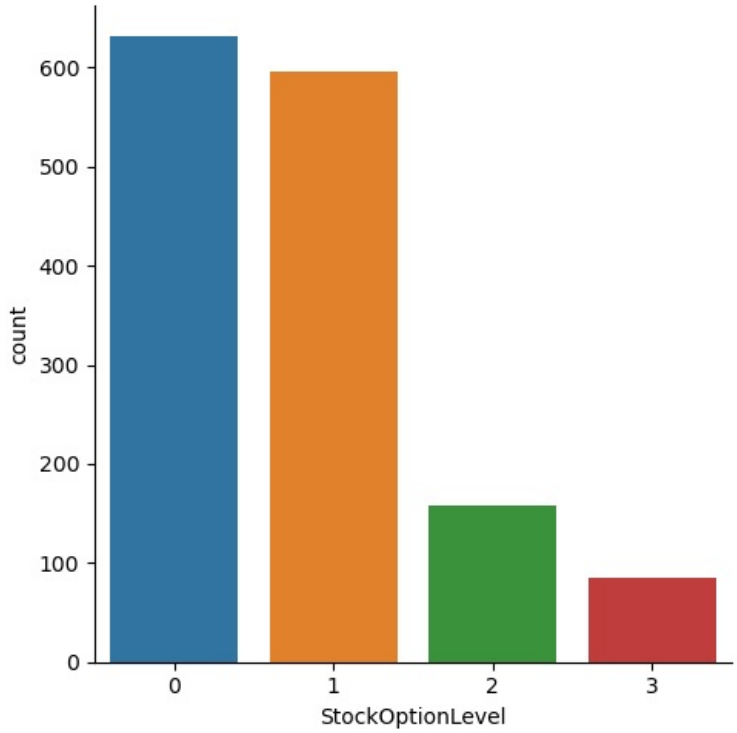

```
<Figure size 1500x1500 with 0 Axes>
```



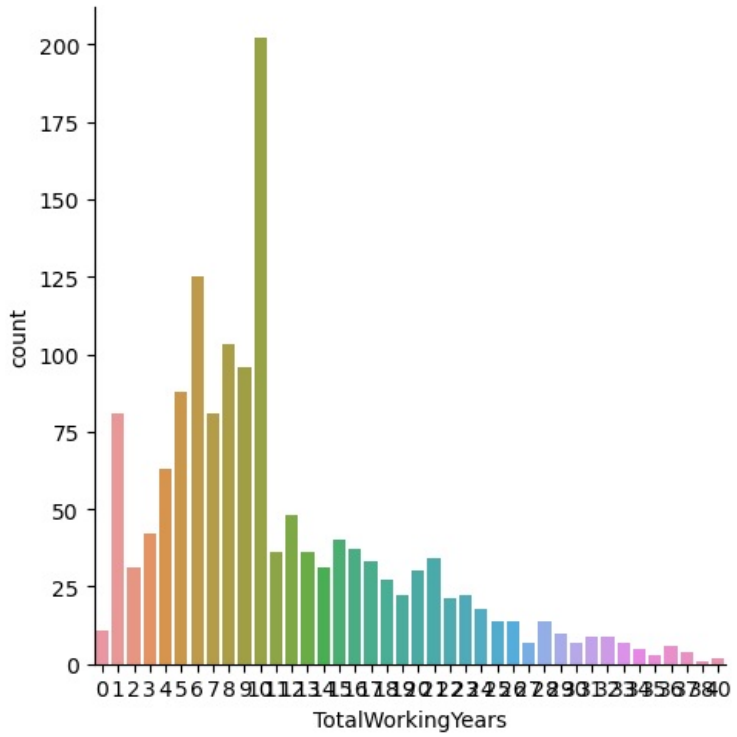

```
<Figure size 1500x1500 with 0 Axes>
```



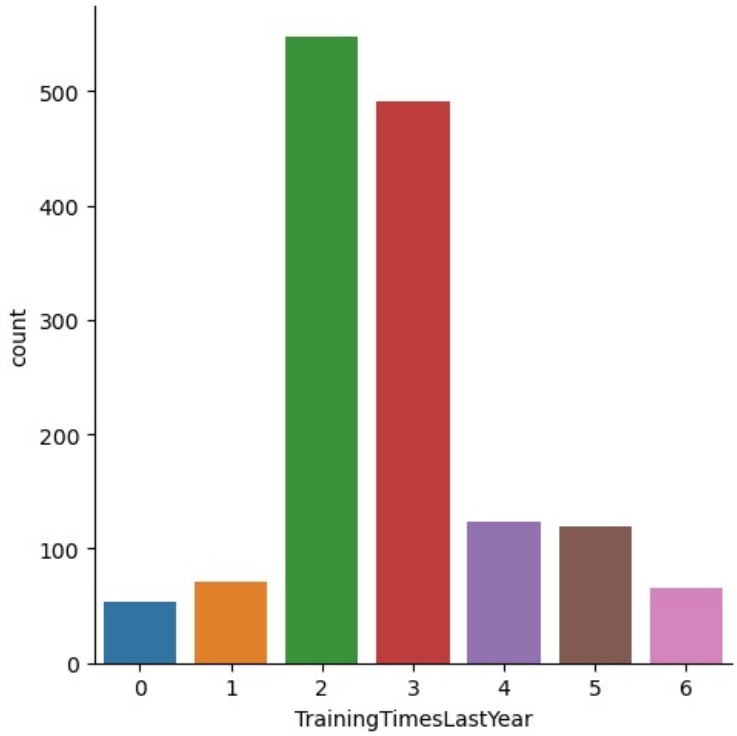

```
<Figure size 1500x1500 with 0 Axes>
```



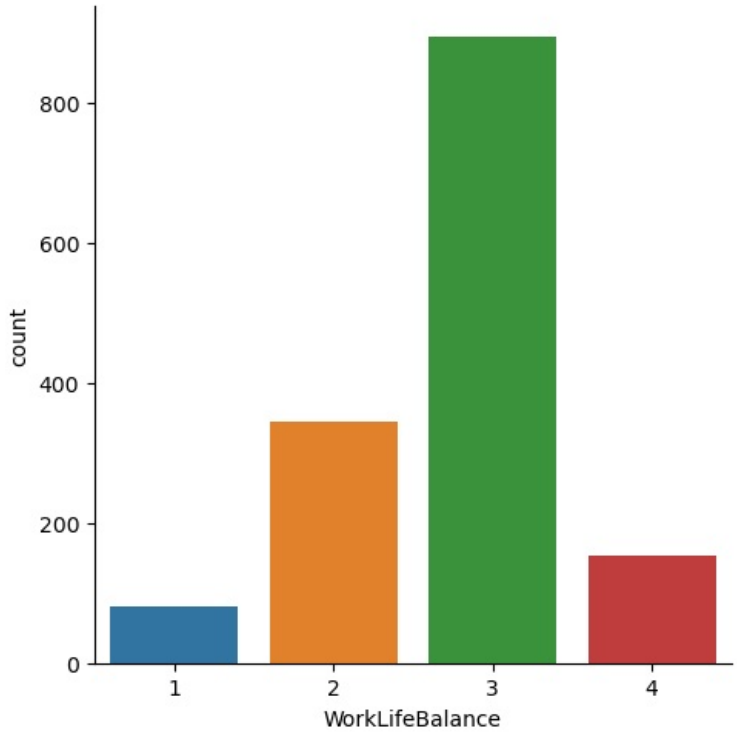

```
<Figure size 1500x1500 with 0 Axes>
```



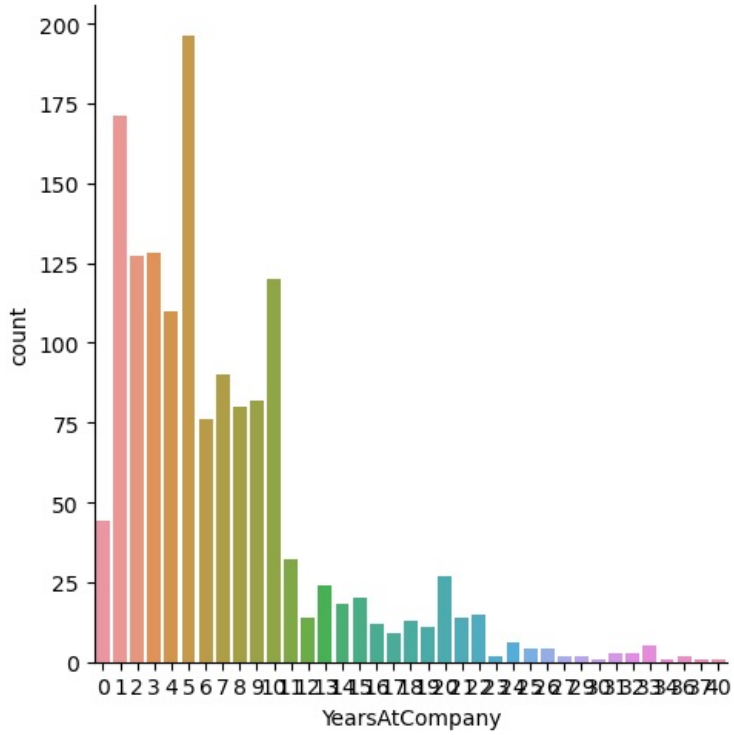

```
<Figure size 1500x1500 with 0 Axes>
```



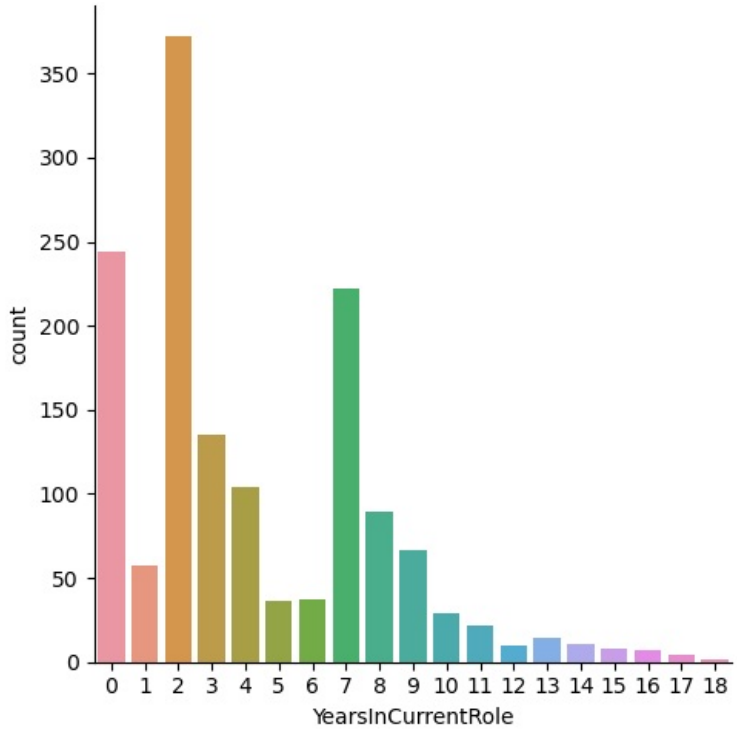

```
<Figure size 1500x1500 with 0 Axes>
```



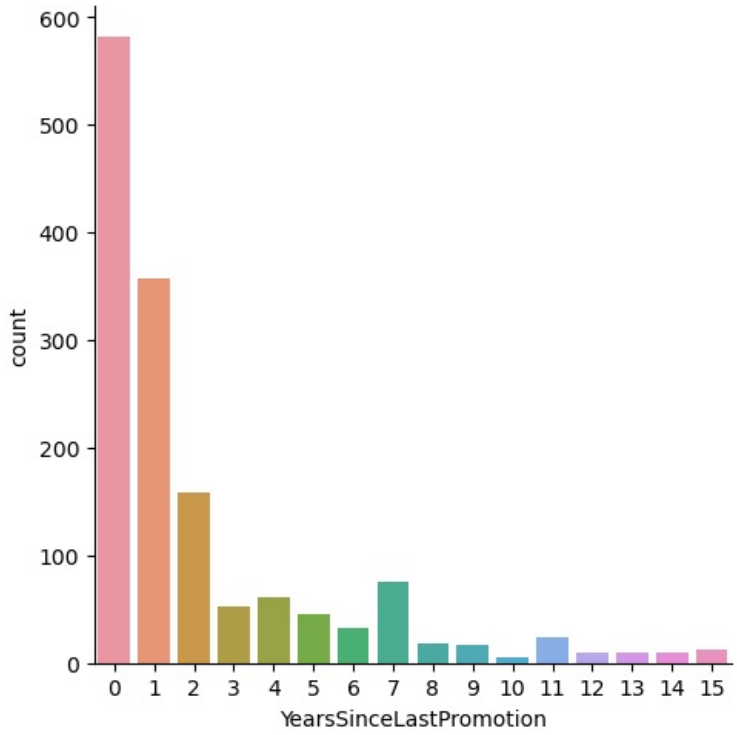

```
<Figure size 1500x1500 with 0 Axes>
```



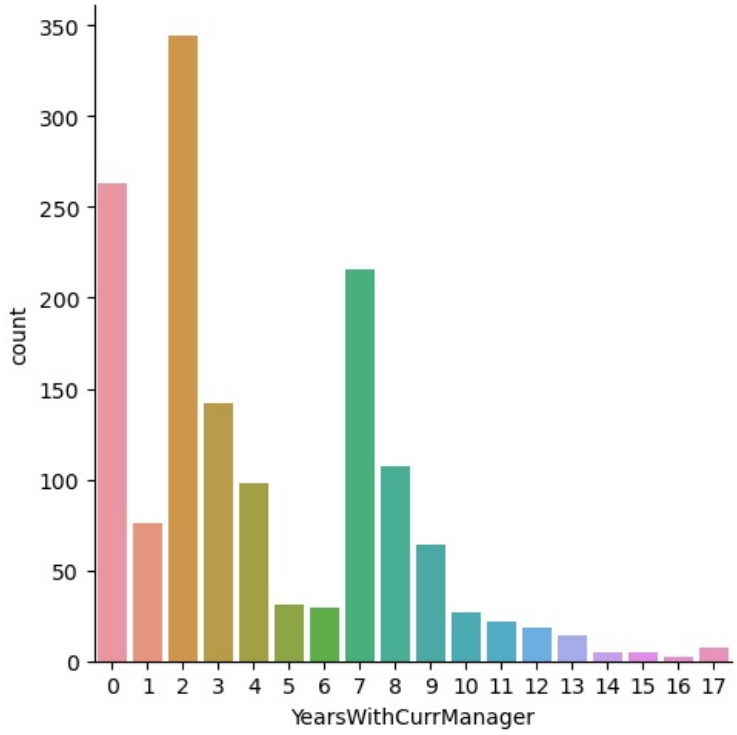
<Figure size 1500x1500 with 0 Axes>



<Figure size 1500x1500 with 0 Axes>




```
<Figure size 1500x1500 with 0 Axes>
```



In [145…
```
Anew.corr()
```

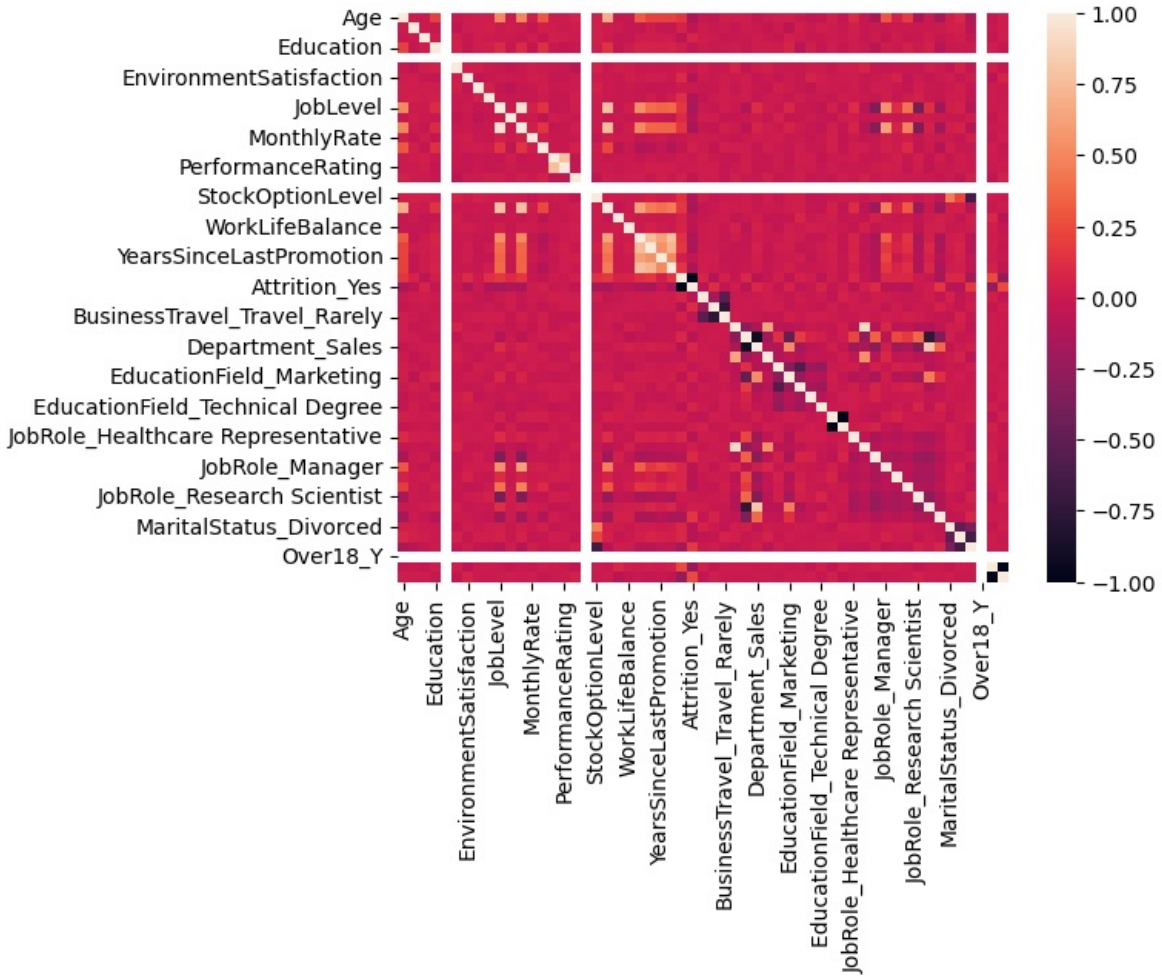
Out[145]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisf
Age	1.000000	0.010661	-0.001686	0.208034	NaN	-0.010145	0.0
DailyRate	0.010661	1.000000	-0.004985	-0.016806	NaN	-0.050990	0.0
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	NaN	0.032916	-0.0
Education	0.208034	-0.016806	0.021042	1.000000	NaN	0.042070	-0.0
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeNumber	-0.010145	-0.050990	0.032916	0.042070	NaN	1.000000	0.0
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	NaN	0.017621	1.0
HourlyRate	0.024287	0.023381	0.031131	0.016775	NaN	0.035179	-0.0
JobInvolvement	0.029820	0.046135	0.008783	0.042438	NaN	-0.006888	-0.0

JobLevel	0.509604	0.002966	0.005303	0.101589	NaN	-0.018519	0.0
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	NaN	-0.046247	-0.0
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	NaN	-0.014829	-0.0
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	NaN	0.012648	0.0
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	NaN	-0.001251	0.0
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	NaN	-0.012944	-0.0
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	NaN	-0.020359	-0.0
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	NaN	-0.069861	0.0
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	NaN	0.062227	0.0
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	NaN	-0.014365	-0.0
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	NaN	0.023603	-0.0
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	NaN	0.010309	0.0
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	NaN	-0.011240	0.0
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	NaN	-0.008416	0.0
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254	NaN	-0.009019	0.0
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	NaN	-0.009197	-0.0
Attrition_No	0.159205	0.056652	-0.077924	0.031373	NaN	0.010577	0.1
Attrition_Yes	-0.159205	-0.056652	0.077924	-0.031373	NaN	-0.010577	-0.1
BusinessTravel_Non-Travel	-0.011215	0.012096	0.023605	0.004524	NaN	0.022272	0.0
BusinessTravel_Travel_Frequently	-0.024743	-0.011776	0.005081	-0.008292	NaN	-0.007980	-0.0
BusinessTravel_Travel_Rarely	0.028791	0.002078	-0.020116	0.004126	NaN	-0.007976	0.0
Department_Human Resources	0.020523	-0.026726	-0.012901	0.011435	NaN	0.063431	-0.0
Department_Research & Development	0.017883	0.014871	-0.008117	-0.018604	NaN	-0.041923	0.0
Department_Sales	-0.027549	-0.003616	0.014085	0.014215	NaN	0.015441	-0.0
EducationField_Human Resources	0.001696	-0.043144	-0.002624	0.026479	NaN	0.035345	-0.0
EducationField_Life Sciences	0.016824	0.004028	-0.024499	0.013184	NaN	-0.000609	-0.0
EducationField_Marketing	0.038162	-0.064449	0.039294	0.072405	NaN	-0.014487	0.0
EducationField_Medical	-0.006354	0.034202	0.013486	-0.072335	NaN	-0.008689	-0.0
EducationField_Other	-0.041466	-0.003893	-0.007969	0.038043	NaN	0.010432	0.0
EducationField_Technical Degree	-0.027604	0.030869	-0.014802	-0.026742	NaN	0.005938	0.0
Gender_Female	0.036311	0.011716	0.001851	0.016547	NaN	-0.022556	-0.0
Gender_Male	-0.036311	-0.011716	-0.001851	-0.016547	NaN	0.022556	0.0
JobRole_Healthcare Representative	0.098825	0.040141	0.022916	0.024270	NaN	0.025945	0.0
JobRole_Human Resources	-0.029856	-0.021156	-0.024089	-0.005295	NaN	0.067287	-0.0
JobRole_Laboratory Technician	-0.143176	-0.006728	0.012369	-0.063566	NaN	-0.019722	-0.0
JobRole_Manager	0.294248	-0.013224	-0.039190	0.028453	NaN	-0.035058	0.0
JobRole_Manufacturing Director	0.049726	-0.005302	0.011848	-0.005290	NaN	-0.014350	0.0
JobRole_Research Director	0.185891	-0.000021	-0.022351	0.049694	NaN	-0.013983	-0.0
JobRole_Research Scientist	-0.146518	-0.002624	-0.010986	0.000709	NaN	-0.017686	0.0
JobRole_Sales Executive	-0.002001	-0.000513	0.030761	0.053398	NaN	0.023263	-0.0
JobRole_Sales Representative	-0.175785	0.005375	-0.015994	-0.091465	NaN	0.006255	0.0
MaritalStatus_Divorced	0.033120	0.037080	-0.005440	-0.002439	NaN	-0.025149	0.0
MaritalStatus_Married	0.083919	0.040035	0.030232	-0.001865	NaN	0.053933	-0.0
MaritalStatus_Single	-0.119185	-0.075835	-0.027445	0.004168	NaN	-0.035189	0.0
Over18_Y	NaN	NaN	NaN	NaN	NaN	NaN	
OverTime_No	-0.028062	-0.009135	-0.025514	0.020322	NaN	0.024037	-0.0
OverTime_Yes	0.028062	0.009135	0.025514	-0.020322	NaN	-0.024037	0.0

In [147…
```
sb.heatmap(Anew.corr())
```

Out[147]:
```
<Axes: >
```



Anova Creation

In [148…
```
cat=[]
con=[]
for i in A.columns:
    if(A[i].dtypes=="object"):
        cat.append(i)
    else:
        con.append(i)
```

In [149…
```
con
```

Out[149]:
```
['Age',
 'DailyRate',
 'DistanceFromHome',
 'Education',
 'EmployeeCount',
 'EmployeeNumber',
 'EnvironmentSatisfaction',
 'HourlyRate',
 'JobInvolvement',
 'JobLevel',
 'JobSatisfaction',
 'MonthlyIncome',
 'MonthlyRate',
 'NumCompaniesWorked',
 'PercentSalaryHike',
 'PerformanceRating',
 'RelationshipSatisfaction',
 'StandardHours',
 'StockOptionLevel',
 'TotalWorkingYears',
 'TrainingTimesLastYear',
 'WorkLifeBalance',
 'YearsAtCompany',
 'YearsInCurrentRole',
 'YearsSinceLastPromotion',
 'YearsWithCurrManager']
```

In [150…
```
cat
```

Out[150]:
```
['Attrition',
 'BusinessTravel',
 'Department',
 'EducationField',
 'Gender',
 'JobRole',
 'MaritalStatus',
 'Over18',
 'OverTime']
```

Anova of predictor

In [156…
```
from statsmodels.api import OLS
from statsmodels.formula.api import ols
model = ols("EmployeeCount ~ MaritalStatus",A).fit()
from statsmodels.stats.anova import anova_lm
anova_result = anova_lm(model)
DF = pd.DataFrame(anova_result)
```

In [157…
```
DF
```

Out[157]:

	df	sum_sq	mean_sq	F	PR(>F)
MaritalStatus	2.0	7.891691e-30	3.945845e-30	4.449087	0.011847
Residual	1467.0	1.301066e-27	8.868888e-31	NaN	NaN