Correlation Analysis and Model Testing on HR Dataset

February 8, 2024

0.0.1 About Dataset:

0.0.2 Context:

The data is details of business employees from a company.

0.0.3 Contents:

Employees Personal Details (Age, Maritial Status, Distance from home, Educational Background).

Employees Official Details (Hourly Wages, Salary Hike, Overtime, Department, Jobrole, Years in current role, Department, Jobrole, Years in current role, Total Working Hours, Training times,etc.).

0.0.4 Problem Statement:

Analyse and visualize the given data.

0.0.5 Import Libraries and Dataset

[1]:
```
# Import Libraries:
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
# Import Dataset:
employee_df = pd.read_csv("C:/Users/amitm/Desktop/New folder/Task Impetus/Class/
↪Python/Case Study/Human_Resources.csv")
```

[3]:
```
employee_df.head()
```

[3]:
```
   Age Attrition     BusinessTravel  DailyRate              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
```

```
   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
2  …                         2            80                 0
3  …                         3            80                 0
4  …                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

[4]:
```
employee_df.tail()
```

[4]:
```
      Age Attrition     BusinessTravel  DailyRate              Department  \
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  \
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  \
```

```
1465           2061  ...                       3            80
1466           2062  ...                       1            80
1467           2064  ...                       2            80
1468           2065  ...                       4            80
1469           2068  ...                       1            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
1465                 1                 17                      3
1466                 1                  9                      5
1467                 1                  6                      0
1468                 0                 17                      3
1469                 0                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
1465               3               5                  2
1466               3               7                  7
1467               3               6                  2
1468               2               9                  6
1469               4               4                  3

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

[5 rows x 35 columns]
```

[5]:
```
employee_df.shape
```

[5]:
```
(1470, 35)
```

[6]:
```
employee_df.columns
```

[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')
```

[7]:
```python
# Print Columns:
print(employee_df.columns)
```

```
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')
```

Describing Data:

[8]:
```python
employee_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
```

[9]:
```
employee_df[['Age','Over18','Attrition']]
```

[9]:

	Age	Over18	Attrition
0	41	Y	Yes
1	49	Y	No
2	37	Y	Yes
3	33	Y	No
4	27	Y	No
…	…	…	…
1465	36	Y	No
1466	39	Y	No
1467	27	Y	No
1468	49	Y	No
1469	34	Y	No

[1470 rows x 3 columns]

[10]:
```
employee_df['Over18'].unique()
```

[10]:
```
array(['Y'], dtype=object)
```

[11]:
```
employee_df['Attrition'].unique()
```

[11]:
```
array(['Yes', 'No'], dtype=object)
```

[12]:
```
# Employees leaving the company:
employee_df[(employee_df['Attrition']=='Yes')&(employee_df['Over18']=='Y')].
 ↪shape[0]
print("No of Employees leaving company: ",employee_df[np.logical_and␣
 ↪(employee_df['Attrition']=='Yes',
                                                                                ␣
 ↪   employee_df['Over18']=='Y')].shape[0])
# There are no employees below 18 of age
```

```
No of Employees leaving company:  237
```

[13]:
```
# Checking for null  values:
employee_df.isnull().sum()
```

[13]:
```
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
```

[14]:
```
employee_df[['Attrition','OverTime','Over18']]
```

[14]:
```
     Attrition OverTime Over18
0          Yes      Yes      Y
1           No       No      Y
2          Yes      Yes      Y
3           No      Yes      Y
```

```
4          No       No      Y
...      ...      ...     ...
1465       No       No      Y
1466       No       No      Y
1467       No      Yes      Y
1468       No       No      Y
1469       No       No      Y

[1470 rows x 3 columns]
```

[15]:
```python
employee_df.describe().T
```

[15]:

	count	mean	std	min	25%
Age	1470.0	36.923810	9.135373	18.0	30.00
DailyRate	1470.0	802.485714	403.509100	102.0	465.00
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00
Education	1470.0	2.912925	1.024165	1.0	2.00
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00
JobLevel	1470.0	2.063946	1.106940	1.0	1.00
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00
StandardHours	1470.0	80.000000	0.000000	80.0	80.00
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00

\

	50%	75%	max
Age	36.0	43.00	60.0
DailyRate	802.0	1157.00	1499.0
DistanceFromHome	7.0	14.00	29.0
Education	3.0	4.00	5.0
EmployeeCount	1.0	1.00	1.0
EmployeeNumber	1020.5	1555.75	2068.0

```
EnvironmentSatisfaction        3.0      4.00      4.0
HourlyRate                    66.0     83.75    100.0
JobInvolvement                 3.0      3.00      4.0
JobLevel                       2.0      3.00      5.0
JobSatisfaction                3.0      4.00      4.0
MonthlyIncome               4919.0   8379.00  19999.0
MonthlyRate                14235.5  20461.50  26999.0
NumCompaniesWorked             2.0      4.00      9.0
PercentSalaryHike             14.0     18.00     25.0
PerformanceRating              3.0      3.00      4.0
RelationshipSatisfaction       3.0      4.00      4.0
StandardHours                 80.0     80.00     80.0
StockOptionLevel               1.0      1.00      3.0
TotalWorkingYears             10.0     15.00     40.0
TrainingTimesLastYear          3.0      3.00      6.0
WorkLifeBalance                3.0      3.00      4.0
YearsAtCompany                 5.0      9.00     40.0
YearsInCurrentRole             3.0      7.00     18.0
YearsSinceLastPromotion        1.0      3.00     15.0
YearsWithCurrManager           3.0      7.00     17.0
```

[16]:
```python
def convertToZeroorOne(x):
    if x=='Yes':
        return 1
    else:
        return 0
```

[17]:
```python
convertToZeroorOne('Yes')
```

[17]:
```
1
```

[18]:
```python
employee_df_new=employee_df['Attrition'].apply(convertToZeroorOne)
```

[19]:
```python
employee_df_new
```

[19]:
```
0       1
1       0
2       1
3       0
4       0
       ..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Attrition, Length: 1470, dtype: int64
```

[20]:
```python
employee_df['Attrition']=employee_df['Attrition'].apply(lambda x:1 if x=='Yes'␣
 ↪else 0)
employee_df['OverTime']=employee_df['OverTime'].apply(lambda x:1 if x=='Yes'␣
 ↪else 0)
employee_df['Over18']=employee_df['Over18'].apply(lambda x:1 if x=='Y' else 0)
```

[21]:
```python
employee_df.head()
```

[21]:
```
   Age  Attrition     BusinessTravel  DailyRate              Department  \
0   41          1      Travel_Rarely       1102                   Sales
1   49          0  Travel_Frequently        279  Research & Development
2   37          1      Travel_Rarely       1373  Research & Development
3   33          0  Travel_Frequently       1392  Research & Development
4   27          0      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
2  …                         2            80                 0
3  …                         3            80                 0
4  …                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

[22]:
```python
employee_df[['Attrition','OverTime','Over18']]
```

[22]:

```
      Attrition  OverTime  Over18
0             1         1       1
1             0         0       1
2             1         1       1
3             0         0       1
4             0         0       1
…           …         …       …
1465          0         0       1
1466          0         0       1
1467          0         1       1
1468          0         0       1
1469          0         0       1

[1470 rows x 3 columns]
```

[23]:

```python
g=employee_df.groupby(['Age'])
```

[24]:

```python
type(g)
```

[24]:

```
pandas.core.groupby.generic.DataFrameGroupBy
```

[25]:

```python
g.get_group(30)
```

[25]:

```
      Age  Attrition     BusinessTravel  DailyRate              Department  \
7      30          0      Travel_Rarely       1358  Research & Development
32     30          0      Travel_Rarely        125  Research & Development
44     30          0  Travel_Frequently        721  Research & Development
80     30          0      Travel_Rarely        852  Research & Development
88     30          0      Travel_Rarely        288  Research & Development
92     30          0      Travel_Rarely       1334                   Sales
120    30          0  Travel_Frequently       1312  Research & Development
139    30          0      Travel_Rarely       1240         Human Resources
143    30          0      Travel_Rarely        438  Research & Development
145    30          0      Travel_Rarely        201  Research & Development
146    30          0      Travel_Rarely       1427  Research & Development
167    30          0      Travel_Rarely       1339                   Sales
173    30          0         Non-Travel        111  Research & Development
211    30          0         Non-Travel        829  Research & Development
214    30          1      Travel_Rarely       1005  Research & Development
216    30          1  Travel_Frequently        334                   Sales
324    30          0      Travel_Rarely       1275  Research & Development
338    30          0      Travel_Rarely        570                   Sales
354    30          0         Non-Travel        641                   Sales
381    30          0      Travel_Rarely        202                   Sales
385    30          1  Travel_Frequently        464  Research & Development
402    30          0      Travel_Rarely       1082                   Sales
410    30          0      Travel_Rarely        317  Research & Development
```

419	30	0	Non-Travel	1400	Research & Development
423	30	0	Non-Travel	1398	Sales
426	30	0	Non-Travel	1116	Research & Development
437	30	0	Travel_Rarely	413	Sales
480	30	1	Travel_Frequently	448	Sales
501	30	0	Travel_Frequently	160	Research & Development
545	30	0	Travel_Rarely	501	Sales
581	30	0	Travel_Rarely	921	Research & Development
602	30	0	Travel_Rarely	946	Research & Development
623	30	0	Travel_Frequently	1012	Research & Development
702	30	0	Travel_Rarely	231	Sales
720	30	1	Travel_Rarely	138	Research & Development
730	30	0	Travel_Rarely	153	Research & Development
732	30	1	Travel_Frequently	109	Research & Development
782	30	0	Travel_Rarely	1176	Research & Development
844	30	0	Travel_Rarely	852	Sales
865	30	0	Travel_Rarely	1329	Sales
874	30	0	Travel_Rarely	853	Research & Development
886	30	0	Travel_Rarely	1465	Research & Development
931	30	0	Non-Travel	879	Research & Development
941	30	0	Travel_Rarely	1138	Research & Development
948	30	0	Travel_Rarely	634	Research & Development
1013	30	0	Travel_Rarely	855	Sales
1049	30	0	Travel_Rarely	1358	Sales
1052	30	0	Non-Travel	990	Research & Development
1064	30	0	Travel_Rarely	330	Human Resources
1106	30	1	Travel_Rarely	740	Sales
1109	30	0	Travel_Rarely	1288	Sales
1141	30	0	Travel_Rarely	241	Research & Development
1233	30	0	Travel_Rarely	793	Research & Development
1244	30	0	Travel_Frequently	1312	Research & Development
1246	30	1	Travel_Frequently	600	Human Resources
1251	30	0	Travel_Rarely	979	Sales
1259	30	0	Travel_Rarely	305	Research & Development
1296	30	0	Travel_Rarely	1092	Research & Development
1338	30	1	Travel_Rarely	945	Sales
1412	30	0	Travel_Rarely	911	Research & Development

	DistanceFromHome	Education	EducationField	EmployeeCount	\
7	24	1	Life Sciences	1	
32	9	2	Medical	1	
44	1	2	Medical	1	
80	1	1	Life Sciences	1	
88	2	3	Life Sciences	1	
92	4	2	Medical	1	
120	23	3	Life Sciences	1	
139	9	3	Human Resources	1	

143	18	3	Life Sciences	1
145	5	3	Technical Degree	1
146	2	1	Medical	1
167	5	3	Life Sciences	1
173	9	3	Medical	1
211	1	1	Life Sciences	1
214	3	3	Technical Degree	1
216	26	4	Marketing	1
324	28	2	Medical	1
338	5	3	Marketing	1
354	25	2	Technical Degree	1
381	2	1	Technical Degree	1
385	4	3	Technical Degree	1
402	12	3	Technical Degree	1
410	2	3	Life Sciences	1
419	3	3	Life Sciences	1
423	22	4	Other	1
426	2	3	Medical	1
437	7	1	Marketing	1
480	12	4	Life Sciences	1
501	3	3	Medical	1
545	27	5	Marketing	1
581	1	3	Life Sciences	1
602	2	3	Medical	1
623	5	4	Life Sciences	1
702	8	2	Other	1
720	22	3	Life Sciences	1
730	8	2	Life Sciences	1
732	5	3	Medical	1
782	20	3	Other	1
844	10	3	Marketing	1
865	29	4	Life Sciences	1
874	7	4	Life Sciences	1
886	1	3	Medical	1
931	9	2	Medical	1
941	6	3	Technical Degree	1
948	17	4	Medical	1
1013	7	4	Marketing	1
1049	16	1	Life Sciences	1
1052	7	3	Technical Degree	1
1064	1	3	Life Sciences	1
1106	1	3	Life Sciences	1
1109	29	4	Technical Degree	1
1141	7	3	Medical	1
1233	16	1	Life Sciences	1
1244	2	4	Technical Degree	1
1246	8	3	Human Resources	1

1251	15	2	Marketing	1
1259	16	3	Life Sciences	1
1296	10	3	Medical	1
1338	9	3	Medical	1
1412	1	2	Medical	1

	EmployeeNumber	…	RelationshipSatisfaction	StandardHours	\
7	11	…	2	80	
32	41	…	1	80	
44	57	…	4	80	
80	104	…	3	80	
88	117	…	1	80	
92	121	…	2	80	
120	159	…	3	80	
139	184	…	4	80	
143	194	…	3	80	
145	197	…	4	80	
146	198	…	4	80	
167	228	…	3	80	
173	239	…	3	80	
211	292	…	3	80	
214	297	…	3	80	
216	299	…	3	80	
324	441	…	4	80	
338	456	…	3	80	
354	475	…	2	80	
381	508	…	1	80	
385	514	…	3	80	
402	533	…	2	80	
410	548	…	3	80	
419	562	…	3	80	
423	567	…	3	80	
426	571	…	3	80	
437	585	…	1	80	
480	648	…	3	80	
501	680	…	3	80	
545	747	…	4	80	
581	806	…	3	80	
602	833	…	2	80	
623	861	…	2	80	
702	982	…	1	80	
720	1004	…	2	80	
730	1015	…	3	80	
732	1017	…	1	80	
782	1084	…	3	80	
844	1179	…	1	80	
865	1211	…	3	80	

874	1224	…	1	80
886	1241	…	1	80
931	1298	…	3	80
941	1311	…	1	80
948	1321	…	4	80
1013	1428	…	2	80
1049	1479	…	3	80
1052	1482	…	2	80
1064	1499	…	1	80
1106	1562	…	4	80
1109	1568	…	2	80
1141	1609	…	2	80
1233	1729	…	2	80
1244	1745	…	4	80
1246	1747	…	3	80
1251	1754	…	1	80
1259	1763	…	3	80
1296	1816	…	2	80
1338	1876	…	3	80
1412	1989	…	3	80

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
7	1	1	2	
32	0	10	5	
44	0	12	2	
80	2	10	1	
88	3	11	3	
92	3	11	4	
120	3	10	2	
139	0	12	2	
143	0	5	4	
145	1	8	3	
146	0	6	3	
167	1	12	2	
173	2	12	4	
211	0	12	2	
214	0	8	5	
216	0	9	5	
324	2	11	2	
338	3	10	2	
354	1	4	2	
381	1	1	3	
385	0	3	4	
402	0	6	6	
410	0	11	2	
419	1	9	3	
423	0	10	3	

426	0	12	2
437	0	4	3
480	1	1	2
501	1	1	2
545	1	10	2
581	2	7	2
602	0	12	4
623	1	10	3
702	1	10	2
720	0	7	2
730	3	9	4
732	0	4	3
782	1	7	1
844	1	10	3
865	3	8	3
874	3	10	4
886	1	12	2
931	0	10	3
941	1	10	6
948	2	9	2
1013	2	8	3
1049	2	4	2
1052	2	1	2
1064	1	6	3
1106	1	10	4
1109	1	9	3
1141	1	6	3
1233	1	10	2
1244	0	10	2
1246	1	6	0
1251	1	12	2
1259	1	10	3
1296	0	9	3
1338	0	1	3
1412	0	12	6

	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole \
7	3	1	0
32	3	10	0
44	3	12	8
80	2	10	8
88	3	11	10
92	2	11	8
120	2	10	7
139	1	11	9
143	2	5	4
145	3	3	2

146	3	5	3
167	3	10	9
173	3	12	9
211	3	11	8
214	3	5	2
216	2	6	3
324	3	10	8
338	3	10	9
354	4	2	2
381	3	1	0
385	3	1	0
402	3	5	4
410	3	5	4
419	1	5	3
423	3	9	8
426	2	11	7
437	3	3	2
480	4	1	0
501	3	1	0
545	2	8	7
581	3	2	2
602	2	0	0
623	2	5	4
702	4	8	4
720	3	5	2
730	2	8	7
732	3	3	2
782	2	6	2
844	3	10	3
865	3	4	3
874	2	10	7
886	3	11	9
931	3	8	4
941	3	9	2
948	3	9	1
1013	3	3	2
1049	2	2	1
1052	2	1	0
1064	4	5	3
1106	3	10	8
1109	3	4	2
1141	2	6	4
1233	2	10	0
1244	3	9	7
1246	2	4	2
1251	3	7	7
1259	3	7	0

1296	3	7	7
1338	2	1	0
1412	2	12	8

	YearsSinceLastPromotion	YearsWithCurrManager
7	0	0
32	1	8
44	3	7
80	3	0
88	10	8
92	2	7
120	0	9
139	4	7
143	0	4
145	2	2
146	1	2
167	7	4
173	6	10
211	5	8
214	0	4
216	0	1
324	1	9
338	1	2
354	2	2
381	0	0
385	0	0
402	4	4
410	0	2
419	1	4
423	7	8
426	6	7
437	1	2
480	0	0
501	0	0
545	7	7
581	0	2
602	0	0
623	0	3
702	7	7
720	0	1
730	1	7
732	1	2
782	0	2
844	1	4
865	0	3
874	8	9
886	5	7

```
931                          1                      7
941                          6                      7
948                          0                      8
1013                         0                      2
1049                         2                      2
1052                         0                      0
1064                         1                      3
1106                         6                      7
1109                         1                      3
1141                         1                      1
1233                         0                      8
1244                         0                      7
1246                         1                      2
1251                         1                      7
1259                         1                      7
1296                         0                      2
1338                         0                      0
1412                         1                      7

[60 rows x 35 columns]
```

[26]:
```
employee_df.groupby(['Age','Attrition']).agg('count')
```

[26]:
```
               BusinessTravel  DailyRate  Department  DistanceFromHome  \
Age Attrition
18  0                       4          4           4                 4
    1                       4          4           4                 4
19  0                       3          3           3                 3
    1                       6          6           6                 6
20  0                       5          5           5                 5
…                         …        …         …                …
57  0                       4          4           4                 4
58  0                       9          9           9                 9
    1                       5          5           5                 5
59  0                      10         10          10                10
60  0                       5          5           5                 5

               Education  EducationField  EmployeeCount  EmployeeNumber  \
Age Attrition
18  0                  4               4              4               4
    1                  4               4              4               4
19  0                  3               3              3               3
    1                  6               6              6               6
20  0                  5               5              5               5
…                   …             …            …             …
57  0                  4               4              4               4
58  0                  9               9              9               9
```

```
    1               5               5               5               5
59  0              10              10              10              10
60  0               5               5               5               5

               EnvironmentSatisfaction  Gender  …  RelationshipSatisfaction  \
Age Attrition                                   …
18  0                                4       4  …                         4
    1                                4       4  …                         4
19  0                                3       3  …                         3
    1                                6       6  …                         6
20  0                                5       5  …                         5
…                                  …      …  …                        …
57  0                                4       4  …                         4
58  0                                9       9  …                         9
    1                                5       5  …                         5
59  0                               10      10  …                        10
60  0                                5       5  …                         5

               StandardHours  StockOptionLevel  TotalWorkingYears  \
Age Attrition
18  0                      4                 4                  4
    1                      4                 4                  4
19  0                      3                 3                  3
    1                      6                 6                  6
20  0                      5                 5                  5
…                        …                 …                  …
57  0                      4                 4                  4
58  0                      9                 9                  9
    1                      5                 5                  5
59  0                     10                10                 10
60  0                      5                 5                  5

               TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  \
Age Attrition
18  0                              4                4               4
    1                              4                4               4
19  0                              3                3               3
    1                              6                6               6
20  0                              5                5               5
…                                …                …               …
57  0                              4                4               4
58  0                              9                9               9
    1                              5                5               5
59  0                             10               10              10
60  0                              5                5               5

               YearsInCurrentRole  YearsSinceLastPromotion  \
```

```
Age Attrition
18  0                    4                    4
    1                    4                    4
19  0                    3                    3
    1                    6                    6
20  0                    5                    5
…                        …                    …
57  0                    4                    4
58  0                    9                    9
    1                    5                    5
59  0                   10                   10
60  0                    5                    5

               YearsWithCurrManager
Age Attrition
18  0                             4
    1                             4
19  0                             3
    1                             6
20  0                             5
…                                 …
57  0                             4
58  0                             9
    1                             5
59  0                            10
60  0                             5

[82 rows x 33 columns]
```

[27]:
```
g.head()
```

[27]:
```
      Age  Attrition     BusinessTravel  DailyRate              Department  \
0      41          1      Travel_Rarely       1102                   Sales
1      49          0  Travel_Frequently        279  Research & Development
2      37          1      Travel_Rarely       1373  Research & Development
3      33          0  Travel_Frequently       1392  Research & Development
4      27          0      Travel_Rarely        591  Research & Development
…     …          …                  …          …                       …
727    18          0         Non-Travel        287  Research & Development
828    18          1         Non-Travel        247  Research & Development
879    60          0      Travel_Rarely        696                   Sales
1053   57          0      Travel_Rarely        405  Research & Development
1209   60          0      Travel_Rarely        370  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
…                …        …            …            …
727                  5          2  Life Sciences             1
828                  8          1       Medical              1
879                  7          4     Marketing              1
1053                 1          2  Life Sciences             1
1209                 1          4       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
…                …  …                       …           …
727             1012  …                         4            80
828             1156  …                         4            80
879             1233  …                         2            80
1053            1483  …                         1            80
1209            1697  …                         3            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                    0                  8                      0
1                    1                 10                      3
2                    0                  7                      3
3                    0                  8                      3
4                    1                  6                      3
…                …                …                     …
727                  0                  0                      2
828                  0                  0                      0
879                  1                 12                      3
1053                 1                 13                      2
1209                 1                 19                      2

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0                  1               6                  4
1                  3              10                  7
2                  3               0                  0
3                  3               8                  7
4                  3               2                  2
…              …              …                …
727                3               0                  0
828                3               0                  0
879                3              11                  7
1053               2              12                  9
```

```
1209                4               1                  0

      YearsSinceLastPromotion  YearsWithCurrManager
0                           0                     5
1                           1                     7
2                           0                     0
3                           3                     0
4                           2                     2
…                         …                   …
727                         0                     0
828                         0                     0
879                         1                     9
1053                        2                     8
1209                        0                     0

[214 rows x 35 columns]
```

0.0.6 Visualizing Dataset:

[28]:
```
# Let's replace the 'Attritition' and 'overtime' column with integers before␣
↪performing any visualizations
#employee_df['Attrition'] = employee_df['Attrition'].apply(lambda x: 1 if x ==␣
↪'Yes' else 0)
#employee_df['OverTime'] = employee_df['OverTime'].apply(lambda x: 1 if x ==␣
↪'Yes' else 0)
#employee_df['Over18'] = employee_df['Over18'].apply(lambda x: 1 if x == 'Y'␣
↪else 0)
```

[29]:
```
employee_df.head()
```

[29]:
```
   Age  Attrition     BusinessTravel  DailyRate              Department  \
0   41          1      Travel_Rarely       1102                   Sales
1   49          0  Travel_Frequently        279  Research & Development
2   37          1      Travel_Rarely       1373  Research & Development
3   33          0  Travel_Frequently       1392  Research & Development
4   27          0      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
```

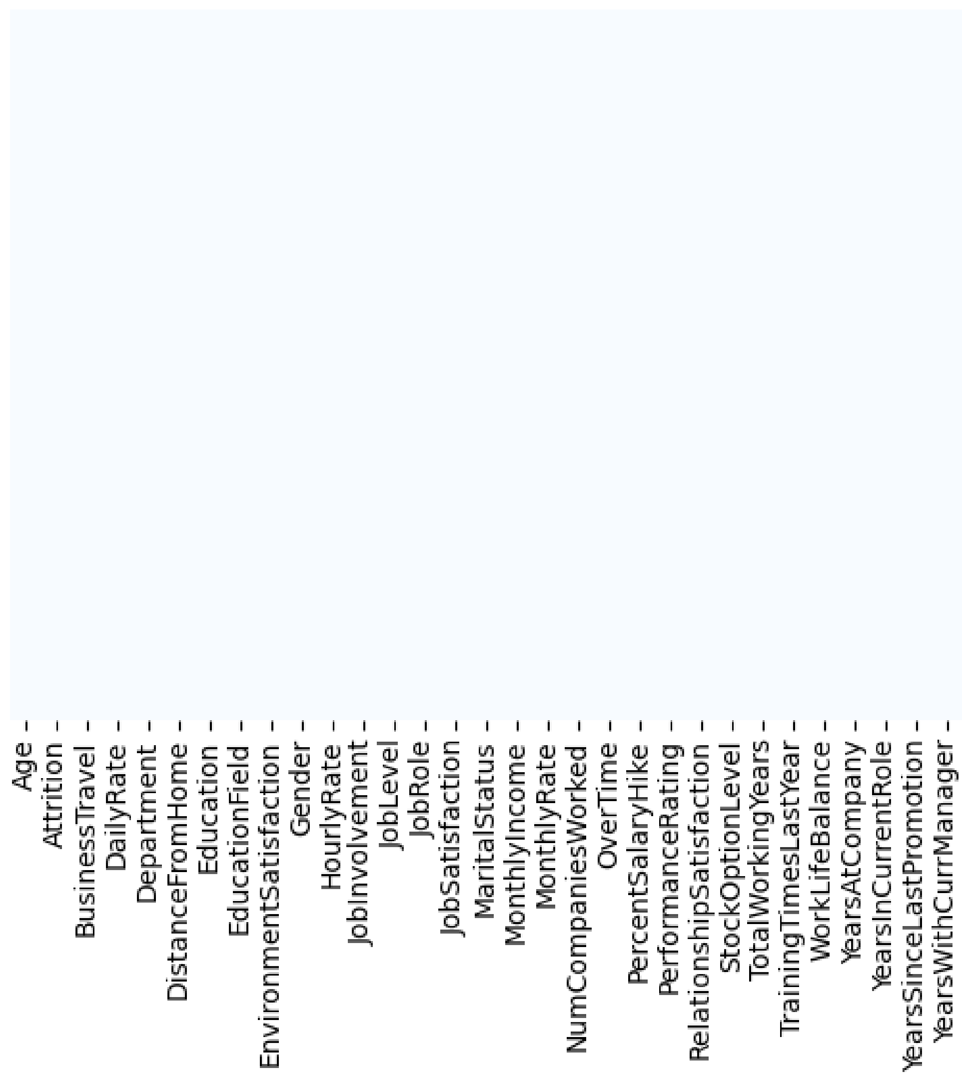
```
2  …                     2           80                 0
3  …                     3           80                 0
4  …                     4           80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

[30]:

```
# Drop Employee Count,Standard Hours,Over 18, Employee Number
employee_df.drop(['EmployeeCount', 'StandardHours', 'Over18',␣
 ↪'EmployeeNumber'],axis=1, inplace=True)
```

[31]:

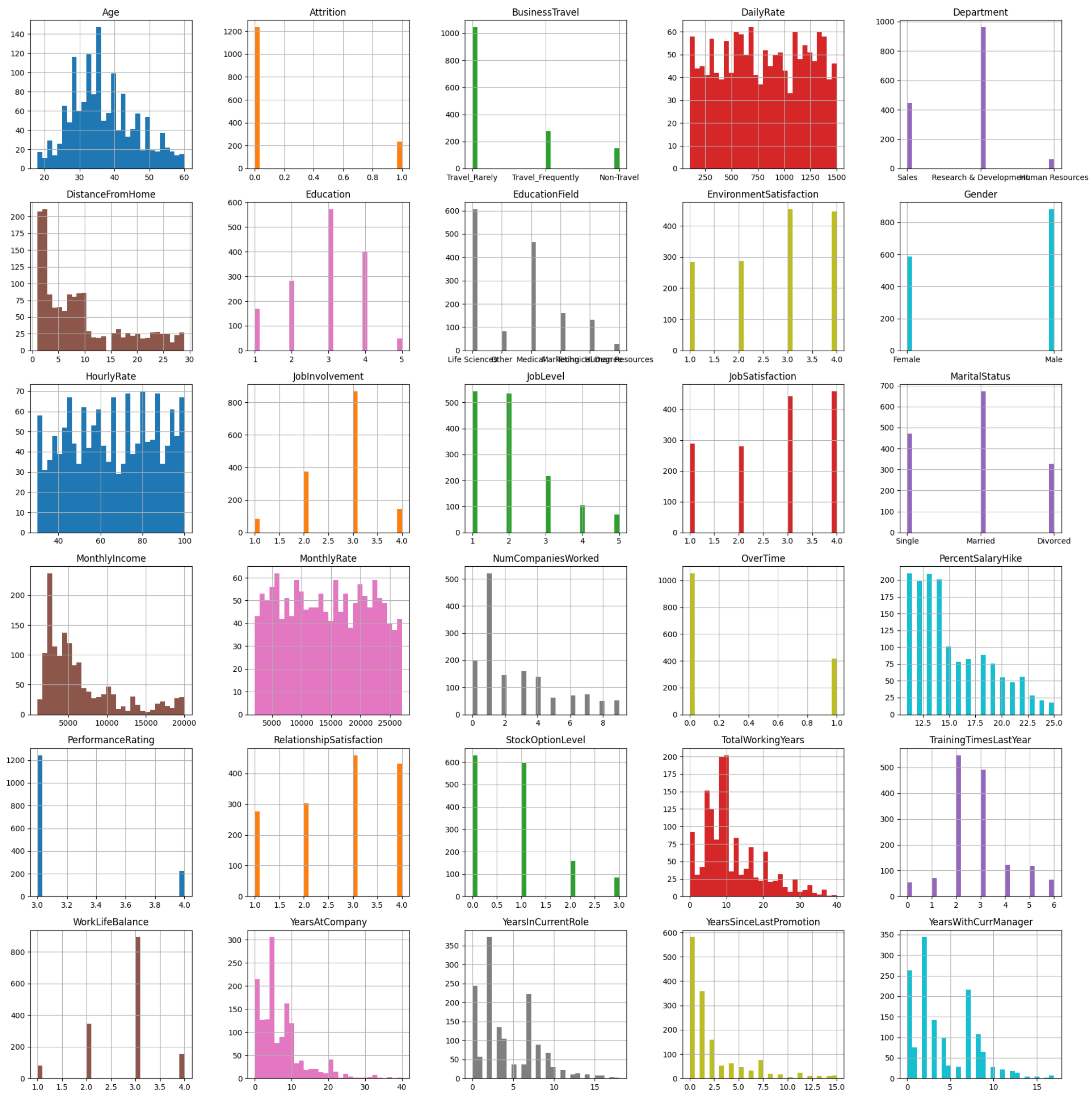
```
# Check For Missing Values:
sns.heatmap(employee_df.isnull(), yticklabels=False, cbar=False, cmap="Blues")
plt.show()
```



[32]:

```
# Get the list of columns in the DataFrame
columns = employee_df.columns
columns = [col for col in columns if col != 'JobRole']

# Calculate the number of rows and columns based on the number of columns in the DataFrame
num_cols = len(columns)
num_rows = (num_cols + 4) // 5
# Create a figure and axes objects
fig, axes = plt.subplots(nrows=num_rows, ncols=5, figsize=(20, 20))
# Flatten axes if necessary
axes = axes.flatten()
# Plot histograms in a grid layout
```

```
for i, col in enumerate(columns):
    employee_df[col].hist(ax=axes[i], bins=30, color='C{}'.format(i))
    axes[i].set_title(col)
# Remove any empty subplots at the end if the number of columns is not a multiple of 5
if num_cols % 5 != 0:
    for j in range(num_cols % 5, 5):
        fig.delaxes(axes[-j])
plt.tight_layout()
plt.show()
```



[33]:
```
employee_df.head(5)
```

```
[33]:    Age  Attrition     BusinessTravel  DailyRate              Department  \
      0   41          1      Travel_Rarely       1102                   Sales
      1   49          0  Travel_Frequently        279  Research & Development
      2   37          1      Travel_Rarely       1373  Research & Development
      3   33          0  Travel_Frequently       1392  Research & Development
      4   27          0      Travel_Rarely        591  Research & Development

         DistanceFromHome  Education EducationField  EnvironmentSatisfaction  \
      0                 1          2  Life Sciences                        2
      1                 8          1  Life Sciences                        3
      2                 2          2          Other                        4
      3                 3          4  Life Sciences                        4
      4                 2          1        Medical                        1

         Gender  …  PerformanceRating  RelationshipSatisfaction  StockOptionLevel  \
      0  Female  …                  3                         1                 0
      1    Male  …                  4                         4                 1
      2    Male  …                  3                         2                 0
      3  Female  …                  3                         3                 0
      4    Male  …                  3                         4                 1

        TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
      0                 8                      0               1               6
      1                10                      3               3              10
      2                 7                      3               3               0
      3                 8                      3               3               8
      4                 6                      3               3               2

         YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
      0                   4                        0                     5
      1                   7                        1                     7
      2                   0                        0                     0
      3                   7                        3                     0
      4                   2                        2                     2

      [5 rows x 31 columns]
```

[34]:
```
# Count the number of employees who stayed and left:
left_df= employee_df[employee_df['Attrition'] == 1]
stayed_df= employee_df[employee_df['Attrition'] == 0]
```

[35]:
```
print("Total =", len(employee_df))

print("Number of employees who left the company =", len(left_df))
print("Percentage of employees who left the company =", 1.*len(left_df)/
 ↪len(employee_df)*100.0, "%")
```

```
print("Number of employees who did not leave the company (stayed) =",␣
 ↪len(stayed_df))
print("Percentage of employees who did not leave the company (stayed) =", 1.
 ↪*len(stayed_df)/len(employee_df)*100.0, "%")
```

```
Total = 1470
Number of employees who left the company = 237
Percentage of employees who left the company = 16.122448979591837 %
Number of employees who did not leave the company (stayed) = 1233
Percentage of employees who did not leave the company (stayed) =
83.87755102040816 %
```

[36]:
```
# Describing Employees who left:
left_df.describe().T
```

[36]:

	count	mean	std	min	25%	\
Age	237.0	33.607595	9.689350	18.0	28.0	
Attrition	237.0	1.000000	0.000000	1.0	1.0	
DailyRate	237.0	750.362869	401.899519	103.0	408.0	
DistanceFromHome	237.0	10.632911	8.452525	1.0	3.0	
Education	237.0	2.839662	1.008244	1.0	2.0	
EnvironmentSatisfaction	237.0	2.464135	1.169791	1.0	1.0	
HourlyRate	237.0	65.573840	20.099958	31.0	50.0	
JobInvolvement	237.0	2.518987	0.773405	1.0	2.0	
JobLevel	237.0	1.637131	0.940594	1.0	1.0	
JobSatisfaction	237.0	2.468354	1.118058	1.0	1.0	
MonthlyIncome	237.0	4787.092827	3640.210367	1009.0	2373.0	
MonthlyRate	237.0	14559.308017	7208.153264	2326.0	8870.0	
NumCompaniesWorked	237.0	2.940928	2.678519	0.0	1.0	
OverTime	237.0	0.535865	0.499768	0.0	0.0	
PercentSalaryHike	237.0	15.097046	3.770294	11.0	12.0	
PerformanceRating	237.0	3.156118	0.363735	3.0	3.0	
RelationshipSatisfaction	237.0	2.599156	1.125437	1.0	2.0	
StockOptionLevel	237.0	0.527426	0.856361	0.0	0.0	
TotalWorkingYears	237.0	8.244726	7.169204	0.0	3.0	
TrainingTimesLastYear	237.0	2.624473	1.254784	0.0	2.0	
WorkLifeBalance	237.0	2.658228	0.816453	1.0	2.0	
YearsAtCompany	237.0	5.130802	5.949984	0.0	1.0	
YearsInCurrentRole	237.0	2.902954	3.174827	0.0	0.0	
YearsSinceLastPromotion	237.0	1.945148	3.153077	0.0	0.0	
YearsWithCurrManager	237.0	2.852321	3.143349	0.0	0.0	

	50%	75%	max
Age	32.0	39.0	58.0
Attrition	1.0	1.0	1.0
DailyRate	699.0	1092.0	1496.0
DistanceFromHome	9.0	17.0	29.0

```
Education                    3.0      4.0      5.0
EnvironmentSatisfaction      3.0      4.0      4.0
HourlyRate                  66.0     84.0    100.0
JobInvolvement               3.0      3.0      4.0
JobLevel                     1.0      2.0      5.0
JobSatisfaction              3.0      3.0      4.0
MonthlyIncome             3202.0   5916.0  19859.0
MonthlyRate              14618.0  21081.0  26999.0
NumCompaniesWorked           1.0      5.0      9.0
OverTime                     1.0      1.0      1.0
PercentSalaryHike           14.0     17.0     25.0
PerformanceRating            3.0      3.0      4.0
RelationshipSatisfaction     3.0      4.0      4.0
StockOptionLevel             0.0      1.0      3.0
TotalWorkingYears            7.0     10.0     40.0
TrainingTimesLastYear        2.0      3.0      6.0
WorkLifeBalance              3.0      3.0      4.0
YearsAtCompany               3.0      7.0     40.0
YearsInCurrentRole           2.0      4.0     15.0
YearsSinceLastPromotion      1.0      2.0     15.0
YearsWithCurrManager         2.0      5.0     14.0
```

[37]:

```python
# Describing Employees who stayed:
stayed_df.describe().T
```

[37]:

```
                          count          mean          std     min     25%  \
Age                      1233.0     37.561233     8.888360    18.0    31.0
Attrition                1233.0      0.000000     0.000000     0.0     0.0
DailyRate                1233.0    812.504461   403.208379   102.0   477.0
DistanceFromHome         1233.0      8.915653     8.012633     1.0     2.0
Education                1233.0      2.927007     1.027002     1.0     2.0
EnvironmentSatisfaction  1233.0      2.771290     1.071132     1.0     2.0
HourlyRate               1233.0     65.952149    20.380754    30.0    48.0
JobInvolvement           1233.0      2.770479     0.692050     1.0     2.0
JobLevel                 1233.0      2.145985     1.117933     1.0     1.0
JobSatisfaction          1233.0      2.778589     1.093277     1.0     2.0
MonthlyIncome            1233.0   6832.739659  4818.208001  1051.0  3211.0
MonthlyRate              1233.0  14265.779400  7102.260749  2094.0  7973.0
NumCompaniesWorked       1233.0      2.645580     2.460090     0.0     1.0
OverTime                 1233.0      0.234388     0.423787     0.0     0.0
PercentSalaryHike        1233.0     15.231144     3.639511    11.0    12.0
PerformanceRating        1233.0      3.153285     0.360408     3.0     3.0
RelationshipSatisfaction 1233.0      2.733982     1.071603     1.0     2.0
StockOptionLevel         1233.0      0.845093     0.841985     0.0     0.0
TotalWorkingYears        1233.0     11.862936     7.760719     0.0     6.0
TrainingTimesLastYear    1233.0      2.832928     1.293585     0.0     2.0
WorkLifeBalance          1233.0      2.781022     0.681907     1.0     2.0
```

```
YearsAtCompany           1233.0    7.369019   6.096298   0.0    3.0
YearsInCurrentRole       1233.0    4.484185   3.649402   0.0    2.0
YearsSinceLastPromotion  1233.0    2.234388   3.234762   0.0    0.0
YearsWithCurrManager     1233.0    4.367397   3.594116   0.0    2.0

                             50%      75%      max
Age                         36.0     43.0     60.0
Attrition                    0.0      0.0      0.0
DailyRate                  817.0   1176.0   1499.0
DistanceFromHome             7.0     13.0     29.0
Education                    3.0      4.0      5.0
EnvironmentSatisfaction      3.0      4.0      4.0
HourlyRate                  66.0     83.0    100.0
JobInvolvement               3.0      3.0      4.0
JobLevel                     2.0      3.0      5.0
JobSatisfaction              3.0      4.0      4.0
MonthlyIncome             5204.0   8834.0  19999.0
MonthlyRate              14120.0  20364.0  26997.0
NumCompaniesWorked           2.0      4.0      9.0
OverTime                     0.0      0.0      1.0
PercentSalaryHike           14.0     18.0     25.0
PerformanceRating            3.0      3.0      4.0
RelationshipSatisfaction     3.0      4.0      4.0
StockOptionLevel             1.0      1.0      3.0
TotalWorkingYears           10.0     16.0     38.0
TrainingTimesLastYear        3.0      3.0      6.0
WorkLifeBalance              3.0      3.0      4.0
YearsAtCompany               6.0     10.0     37.0
YearsInCurrentRole           3.0      7.0     18.0
YearsSinceLastPromotion      1.0      3.0     15.0
YearsWithCurrManager         3.0      7.0     17.0
```

[38]:
```python
# Correlating Dataset:
numeric_df = employee_df.select_dtypes(include='number')
correlations = numeric_df.corr()
correlations
```

[38]:
```
                              Age  Attrition  DailyRate  DistanceFromHome  \
Age                      1.000000  -0.159205   0.010661         -0.001686
Attrition               -0.159205   1.000000  -0.056652          0.077924
DailyRate                0.010661  -0.056652   1.000000         -0.004985
DistanceFromHome        -0.001686   0.077924  -0.004985          1.000000
Education                0.208034  -0.031373  -0.016806          0.021042
EnvironmentSatisfaction  0.010146  -0.103369   0.018355         -0.016075
HourlyRate               0.024287  -0.006846   0.023381          0.031131
JobInvolvement           0.029820  -0.130016   0.046135          0.008783
JobLevel                 0.509604  -0.169105   0.002966          0.005303
```

JobSatisfaction	-0.004892	-0.103481	0.030571	-0.003669
MonthlyIncome	0.497855	-0.159840	0.007707	-0.017014
MonthlyRate	0.028051	0.015170	-0.032182	0.027473
NumCompaniesWorked	0.299635	0.043494	0.038153	-0.029251
OverTime	0.028062	0.246118	0.009135	0.025514
PercentSalaryHike	0.003634	-0.013478	0.022704	0.040235
PerformanceRating	0.001904	0.002889	0.000473	0.027110
RelationshipSatisfaction	0.053535	-0.045872	0.007846	0.006557
StockOptionLevel	0.037510	-0.137145	0.042143	0.044872
TotalWorkingYears	0.680381	-0.171063	0.014515	0.004628
TrainingTimesLastYear	-0.019621	-0.059478	0.002453	-0.036942
WorkLifeBalance	-0.021490	-0.063939	-0.037848	-0.026556
YearsAtCompany	0.311309	-0.134392	-0.034055	0.009508
YearsInCurrentRole	0.212901	-0.160545	0.009932	0.018845
YearsSinceLastPromotion	0.216513	-0.033019	-0.033229	0.010029
YearsWithCurrManager	0.202089	-0.156199	-0.026363	0.014406

	Education	EnvironmentSatisfaction	HourlyRate	\
Age	0.208034	0.010146	0.024287	
Attrition	-0.031373	-0.103369	-0.006846	
DailyRate	-0.016806	0.018355	0.023381	
DistanceFromHome	0.021042	-0.016075	0.031131	
Education	1.000000	-0.027128	0.016775	
EnvironmentSatisfaction	-0.027128	1.000000	-0.049857	
HourlyRate	0.016775	-0.049857	1.000000	
JobInvolvement	0.042438	-0.008278	0.042861	
JobLevel	0.101589	0.001212	-0.027853	
JobSatisfaction	-0.011296	-0.006784	-0.071335	
MonthlyIncome	0.094961	-0.006259	-0.015794	
MonthlyRate	-0.026084	0.037600	-0.015297	
NumCompaniesWorked	0.126317	0.012594	0.022157	
OverTime	-0.020322	0.070132	-0.007782	
PercentSalaryHike	-0.011111	-0.031701	-0.009062	
PerformanceRating	-0.024539	-0.029548	-0.002172	
RelationshipSatisfaction	-0.009118	0.007665	0.001330	
StockOptionLevel	0.018422	0.003432	0.050263	
TotalWorkingYears	0.148280	-0.002693	-0.002334	
TrainingTimesLastYear	-0.025100	-0.019359	-0.008548	
WorkLifeBalance	0.009819	0.027627	-0.004607	
YearsAtCompany	0.069114	0.001458	-0.019582	
YearsInCurrentRole	0.060236	0.018007	-0.024106	
YearsSinceLastPromotion	0.054254	0.016194	-0.026716	
YearsWithCurrManager	0.069065	-0.004999	-0.020123	

	JobInvolvement	JobLevel	JobSatisfaction	…	\
Age	0.029820	0.509604	-0.004892	…	
Attrition	-0.130016	-0.169105	-0.103481	…	

```
DailyRate                   0.046135  0.002966          0.030571  …
DistanceFromHome            0.008783  0.005303         -0.003669  …
Education                   0.042438  0.101589         -0.011296  …
EnvironmentSatisfaction    -0.008278  0.001212         -0.006784  …
HourlyRate                  0.042861 -0.027853         -0.071335  …
JobInvolvement              1.000000 -0.012630         -0.021476  …
JobLevel                   -0.012630  1.000000         -0.001944  …
JobSatisfaction            -0.021476 -0.001944          1.000000  …
MonthlyIncome              -0.015271  0.950300         -0.007157  …
MonthlyRate                -0.016322  0.039563          0.000644  …
NumCompaniesWorked          0.015012  0.142501         -0.055699  …
OverTime                   -0.003507  0.000544          0.024539  …
PercentSalaryHike          -0.017205 -0.034730          0.020002  …
PerformanceRating          -0.029071 -0.021222          0.002297  …
RelationshipSatisfaction    0.034297  0.021642         -0.012454  …
StockOptionLevel            0.021523  0.013984          0.010690  …
TotalWorkingYears          -0.005533  0.782208         -0.020185  …
TrainingTimesLastYear      -0.015338 -0.018191         -0.005779  …
WorkLifeBalance            -0.014617  0.037818         -0.019459  …
YearsAtCompany             -0.021355  0.534739         -0.003803  …
YearsInCurrentRole          0.008717  0.389447         -0.002305  …
YearsSinceLastPromotion    -0.024184  0.353885         -0.018214  …
YearsWithCurrManager        0.025976  0.375281         -0.027656  …

                          PerformanceRating  RelationshipSatisfaction  \
Age                                0.001904                  0.053535
Attrition                          0.002889                 -0.045872
DailyRate                          0.000473                  0.007846
DistanceFromHome                   0.027110                  0.006557
Education                         -0.024539                 -0.009118
EnvironmentSatisfaction           -0.029548                  0.007665
HourlyRate                        -0.002172                  0.001330
JobInvolvement                    -0.029071                  0.034297
JobLevel                          -0.021222                  0.021642
JobSatisfaction                    0.002297                 -0.012454
MonthlyIncome                     -0.017120                  0.025873
MonthlyRate                       -0.009811                 -0.004085
NumCompaniesWorked                -0.014095                  0.052733
OverTime                           0.004369                  0.048493
PercentSalaryHike                  0.773550                 -0.040490
PerformanceRating                  1.000000                 -0.031351
RelationshipSatisfaction          -0.031351                  1.000000
StockOptionLevel                   0.003506                 -0.045952
TotalWorkingYears                  0.006744                  0.024054
TrainingTimesLastYear             -0.015579                  0.002497
WorkLifeBalance                    0.002572                  0.019604
YearsAtCompany                     0.003435                  0.019367
```

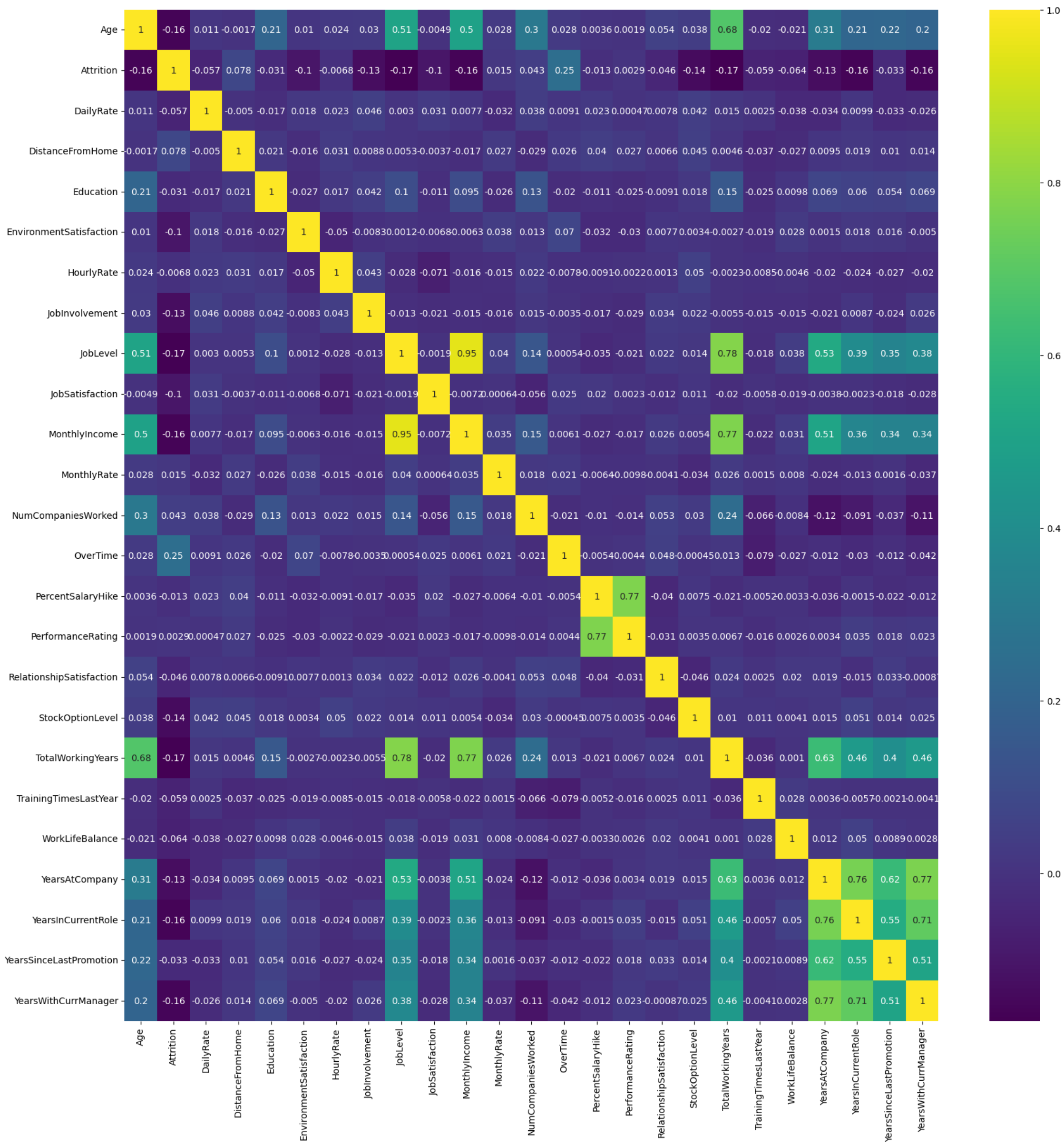
```
YearsInCurrentRole                  0.034986                 -0.015123
YearsSinceLastPromotion             0.017896                  0.033493
YearsWithCurrManager                0.022827                 -0.000867

                          StockOptionLevel  TotalWorkingYears  \
Age                               0.037510           0.680381
Attrition                        -0.137145          -0.171063
DailyRate                         0.042143           0.014515
DistanceFromHome                  0.044872           0.004628
Education                         0.018422           0.148280
EnvironmentSatisfaction           0.003432          -0.002693
HourlyRate                        0.050263          -0.002334
JobInvolvement                    0.021523          -0.005533
JobLevel                          0.013984           0.782208
JobSatisfaction                   0.010690          -0.020185
MonthlyIncome                     0.005408           0.772893
MonthlyRate                      -0.034323           0.026442
NumCompaniesWorked                0.030075           0.237639
OverTime                         -0.000449           0.012754
PercentSalaryHike                 0.007528          -0.020608
PerformanceRating                 0.003506           0.006744
RelationshipSatisfaction         -0.045952           0.024054
StockOptionLevel                  1.000000           0.010136
TotalWorkingYears                 0.010136           1.000000
TrainingTimesLastYear             0.011274          -0.035662
WorkLifeBalance                   0.004129           0.001008
YearsAtCompany                    0.015058           0.628133
YearsInCurrentRole                0.050818           0.460365
YearsSinceLastPromotion           0.014352           0.404858
YearsWithCurrManager              0.024698           0.459188

                          TrainingTimesLastYear  WorkLifeBalance  \
Age                                   -0.019621        -0.021490
Attrition                             -0.059478        -0.063939
DailyRate                              0.002453        -0.037848
DistanceFromHome                      -0.036942        -0.026556
Education                             -0.025100         0.009819
EnvironmentSatisfaction               -0.019359         0.027627
HourlyRate                            -0.008548        -0.004607
JobInvolvement                        -0.015338        -0.014617
JobLevel                              -0.018191         0.037818
JobSatisfaction                       -0.005779        -0.019459
MonthlyIncome                         -0.021736         0.030683
MonthlyRate                            0.001467         0.007963
NumCompaniesWorked                    -0.066054        -0.008366
OverTime                              -0.079113        -0.027092
PercentSalaryHike                     -0.005221        -0.003280
```

```
PerformanceRating                   -0.015579         0.002572
RelationshipSatisfaction             0.002497         0.019604
StockOptionLevel                     0.011274         0.004129
TotalWorkingYears                   -0.035662         0.001008
TrainingTimesLastYear                1.000000         0.028072
WorkLifeBalance                      0.028072         1.000000
YearsAtCompany                       0.003569         0.012089
YearsInCurrentRole                  -0.005738         0.049856
YearsSinceLastPromotion             -0.002067         0.008941
YearsWithCurrManager                -0.004096         0.002759

                          YearsAtCompany  YearsInCurrentRole  \
Age                             0.311309            0.212901
Attrition                      -0.134392           -0.160545
DailyRate                      -0.034055            0.009932
DistanceFromHome                0.009508            0.018845
Education                       0.069114            0.060236
EnvironmentSatisfaction         0.001458            0.018007
HourlyRate                     -0.019582           -0.024106
JobInvolvement                 -0.021355            0.008717
JobLevel                        0.534739            0.389447
JobSatisfaction                -0.003803           -0.002305
MonthlyIncome                   0.514285            0.363818
MonthlyRate                    -0.023655           -0.012815
NumCompaniesWorked             -0.118421           -0.090754
OverTime                       -0.011687           -0.029758
PercentSalaryHike              -0.035991           -0.001520
PerformanceRating               0.003435            0.034986
RelationshipSatisfaction        0.019367           -0.015123
StockOptionLevel                0.015058            0.050818
TotalWorkingYears               0.628133            0.460365
TrainingTimesLastYear           0.003569           -0.005738
WorkLifeBalance                 0.012089            0.049856
YearsAtCompany                  1.000000            0.758754
YearsInCurrentRole              0.758754            1.000000
YearsSinceLastPromotion         0.618409            0.548056
YearsWithCurrManager            0.769212            0.714365

                          YearsSinceLastPromotion  YearsWithCurrManager
Age                                      0.216513              0.202089
Attrition                               -0.033019             -0.156199
DailyRate                               -0.033229             -0.026363
DistanceFromHome                         0.010029              0.014406
Education                                0.054254              0.069065
EnvironmentSatisfaction                  0.016194             -0.004999
HourlyRate                              -0.026716             -0.020123
JobInvolvement                          -0.024184              0.025976
```

```
JobLevel                                      0.353885                 0.375281
JobSatisfaction                              -0.018214                -0.027656
MonthlyIncome                                 0.344978                 0.344079
MonthlyRate                                   0.001567                -0.036746
NumCompaniesWorked                           -0.036814                -0.110319
OverTime                                     -0.012239                -0.041586
PercentSalaryHike                            -0.022154                -0.011985
PerformanceRating                             0.017896                 0.022827
RelationshipSatisfaction                      0.033493                -0.000867
StockOptionLevel                              0.014352                 0.024698
TotalWorkingYears                             0.404858                 0.459188
TrainingTimesLastYear                        -0.002067                -0.004096
WorkLifeBalance                               0.008941                 0.002759
YearsAtCompany                                0.618409                 0.769212
YearsInCurrentRole                            0.548056                 0.714365
YearsSinceLastPromotion                       1.000000                 0.510224
YearsWithCurrManager                          0.510224                 1.000000

[25 rows x 25 columns]
```

[39]:

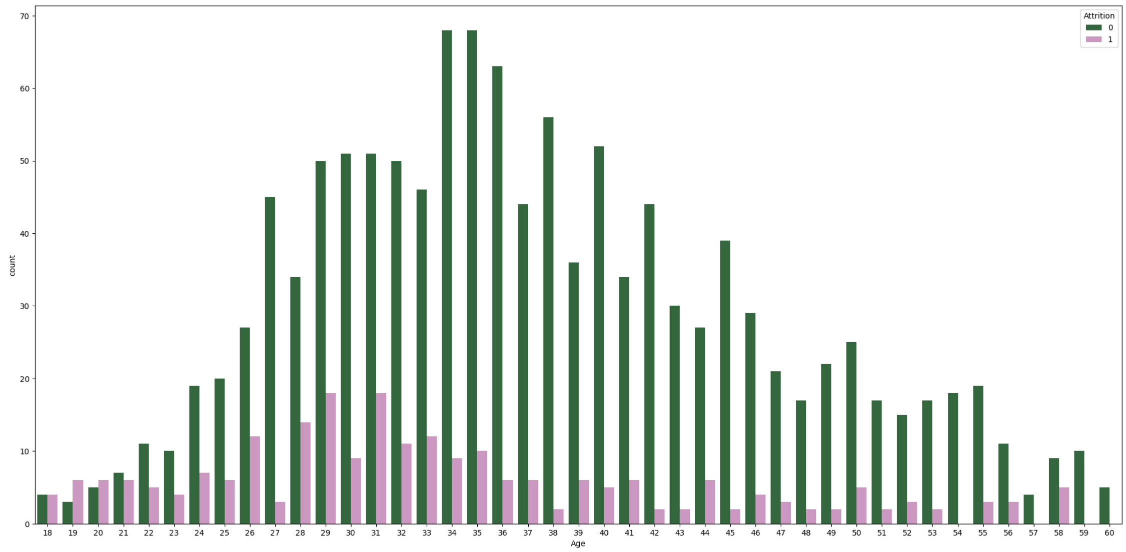
```
custom_cmap = sns.color_palette("viridis", as_cmap=True)
f, ax = plt.subplots(figsize = (20, 20))
sns.heatmap(correlations, annot = True,cmap=custom_cmap)
plt.show()
```



- Job level is strongly correlated with total working Years
- Monthly income is strongly correlated with Job level
- Monthly income is strongly correlated with total working Years
- Age is stongly correlated with monthly income

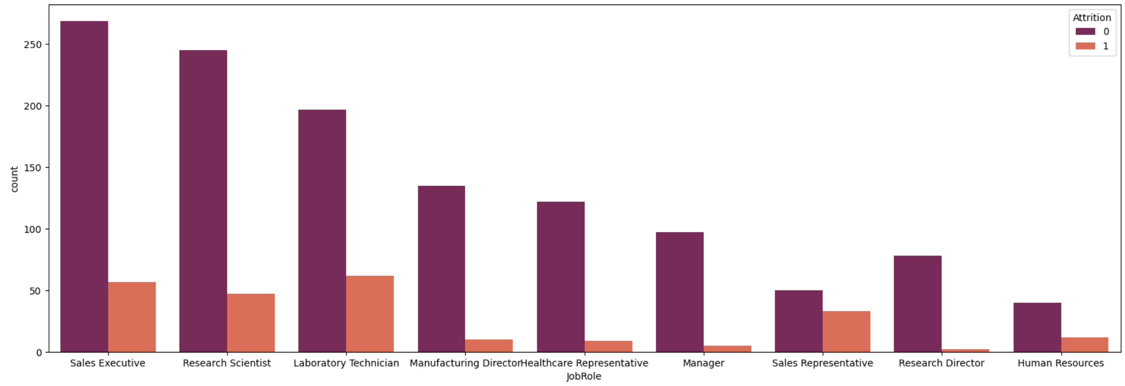
[40]:
```
plt.figure(figsize=[25, 12])
sns.countplot(x = 'Age', hue = 'Attrition', data = employee_df,␣
↪palette="cubehelix")
```

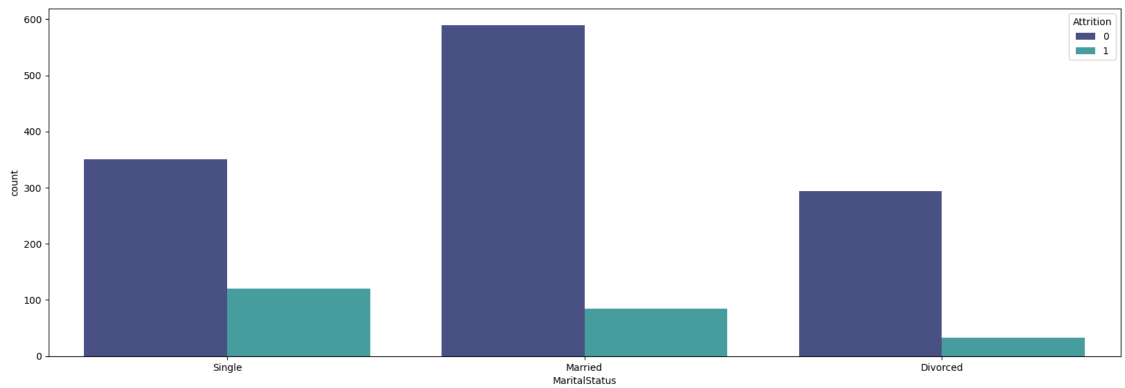
[40]: <Axes: xlabel='Age', ylabel='count'>

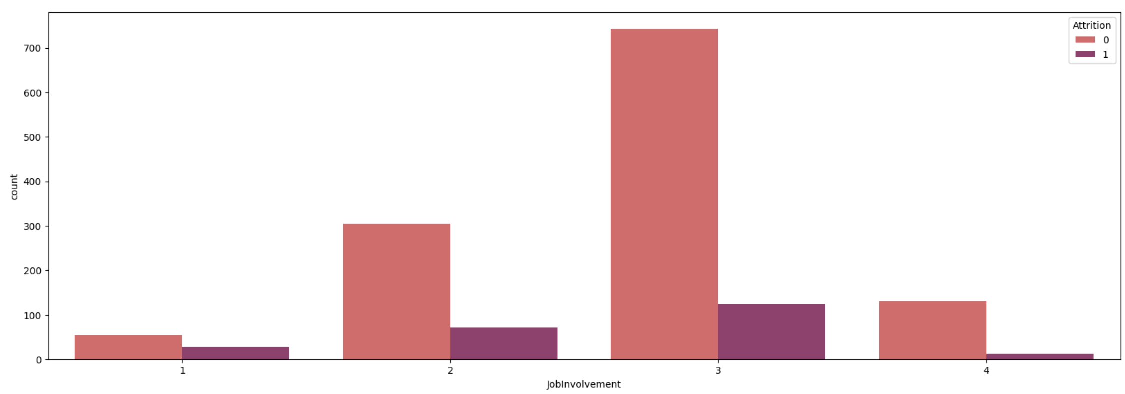


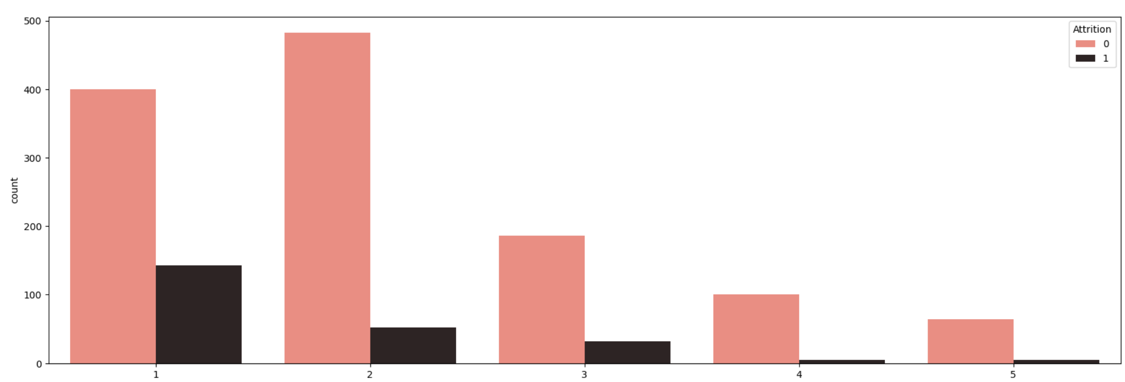
[41]:
```
plt.figure(figsize=[20,30])
plt.subplot(411)
sns.countplot(x = 'JobRole', hue = 'Attrition', data = employee_df,␣
↪palette="rocket")
plt.subplot(412)
sns.countplot(x = 'MaritalStatus', hue = 'Attrition', data = employee_df,␣
↪palette="mako")
plt.subplot(413)
sns.countplot(x = 'JobInvolvement', hue = 'Attrition', data = employee_df,␣
↪palette="flare")
plt.subplot(414)
sns.countplot(x = 'JobLevel', hue = 'Attrition', data = employee_df,␣
↪palette="dark:salmon_r")
```

[41]: <Axes: xlabel='JobLevel', ylabel='count'>


Attrition
0
1
count
250
200
150
100
50
0
Sales Executive
Research Scientist
Laboratory Technician
Manufacturing Director
Healthcare Representative
Manager
Sales Representative
Research Director
Human Resources
JobRole


Attrition
0
1
count
600
500
400
300
200
100
0
Single
Married
Divorced
MaritalStatus

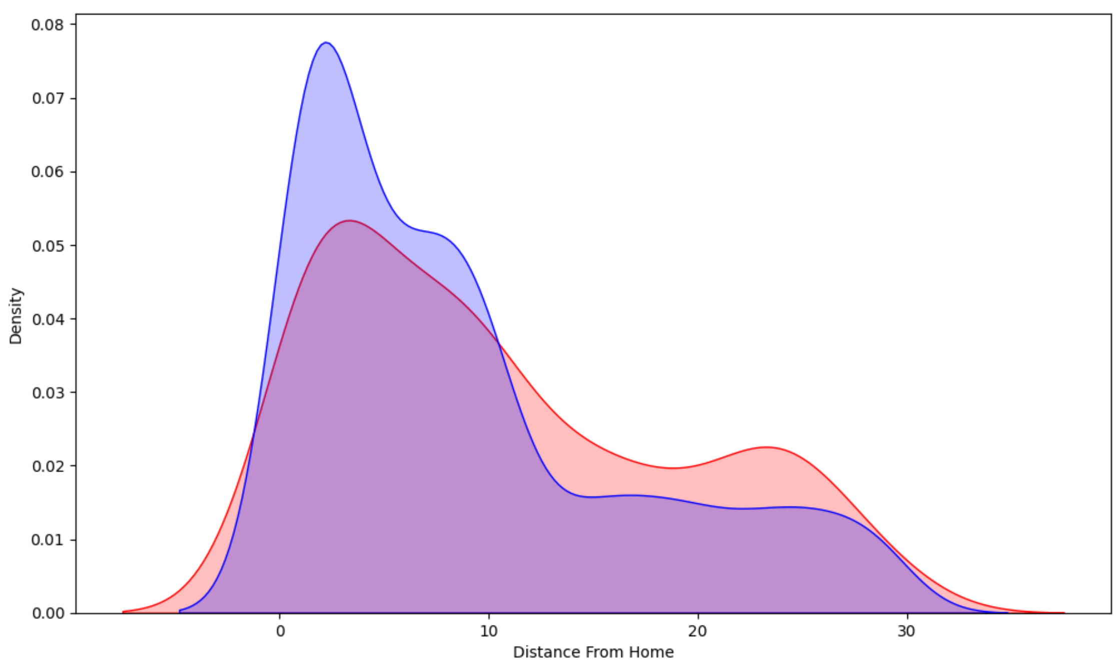

Attrition
0
1
count
700
600
500
400
300
200
100
0
1
2
3
4
JobInvolvement


Attrition
0
1
count
500
400
300
200
100
0
1
2
3
4
5
JobLevel

- Single employees tend to leave compared to married and divorced
- Sales Representitives tend to leave compared to any other job
- Less involved employees tend to leave the company
- Less experienced (low job level) tend to leave the company

[42]:
```
# KDE (Kernel Density Estimate) is used for visualizing the Probability Density␣
↪of a continuous variable.
# KDE describes the probability density at different values in a continuous␣
↪variable.
plt.figure(figsize=(12,7))

sns.kdeplot(left_df['DistanceFromHome'], label = 'Employees who left', fill =␣
↪True, color = 'red')
sns.kdeplot(stayed_df['DistanceFromHome'], label = 'Employees who Stayed', fill␣
↪= True, color = 'blue')

plt.xlabel('Distance From Home')
```

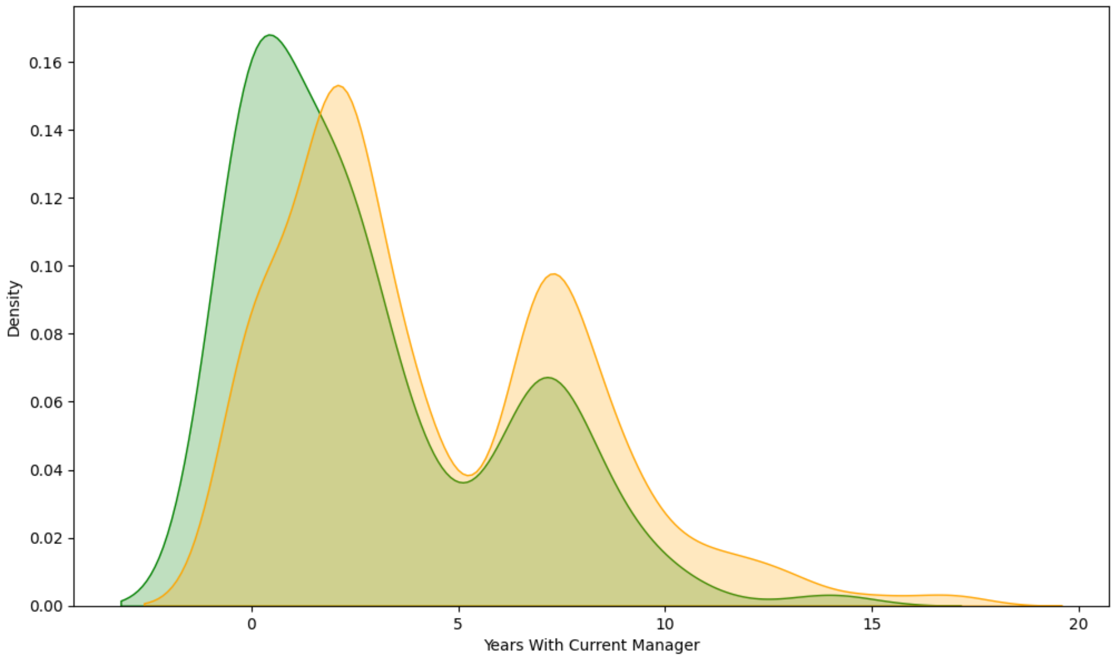
[42]: Text(0.5, 0, 'Distance From Home')



[43]:
```
plt.figure(figsize=(12,7))

sns.kdeplot(left_df['YearsWithCurrManager'], label = 'Employees who left', fill␣
↪= True, color = 'green')
```

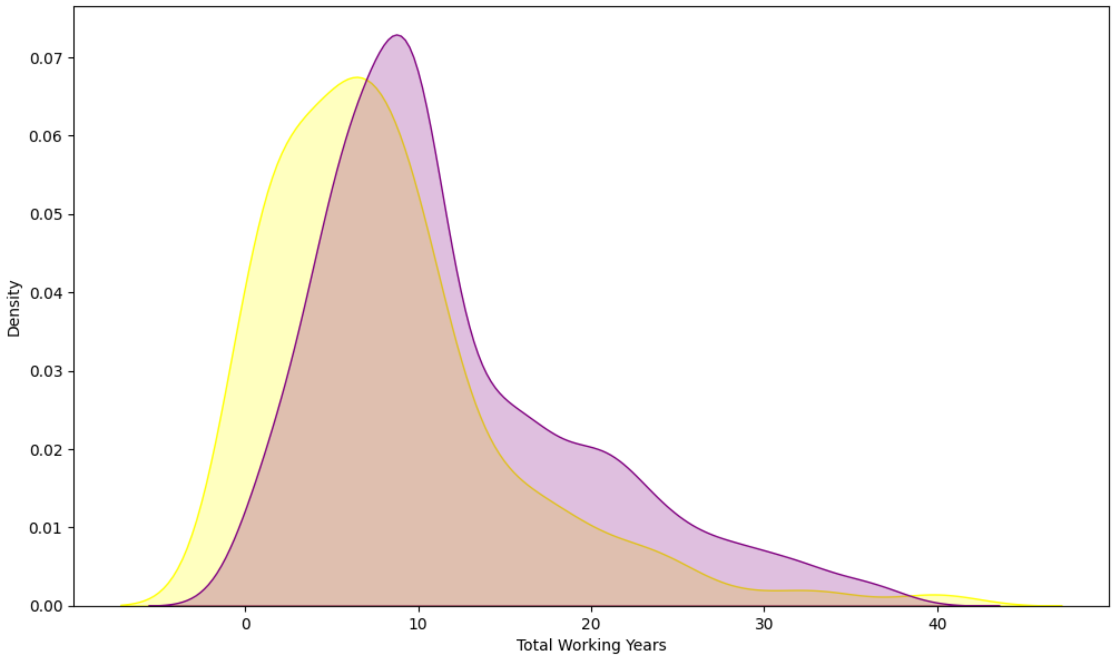
```
sns.kdeplot(stayed_df['YearsWithCurrManager'], label = 'Employees who Stayed',␣
↪fill = True, color = 'orange')

plt.xlabel('Years With Current Manager')
```

[43]: Text(0.5, 0, 'Years With Current Manager')



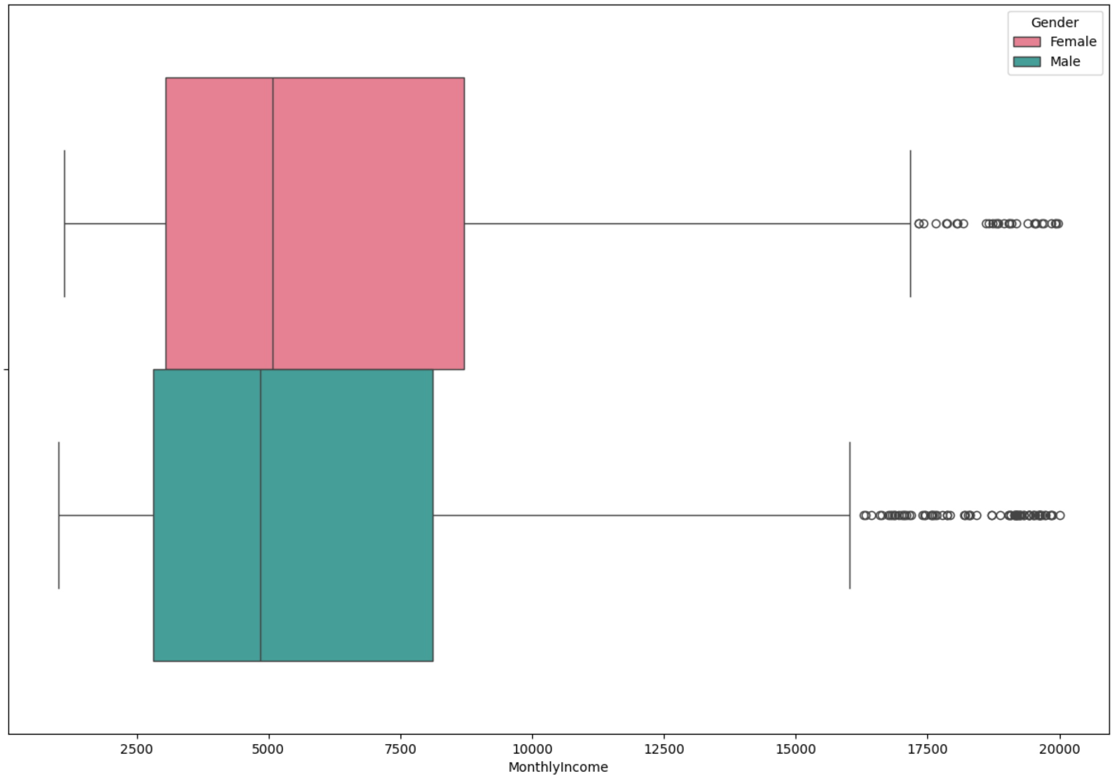
[44]:
```
plt.figure(figsize=(12,7))

sns.kdeplot(left_df['TotalWorkingYears'], fill = True, label = 'Employees who␣
↪left', color = 'yellow')
sns.kdeplot(stayed_df['TotalWorkingYears'], fill = True, label = 'Employees who␣
↪Stayed', color = 'purple')

plt.xlabel('Total Working Years')
```

[44]: Text(0.5, 0, 'Total Working Years')



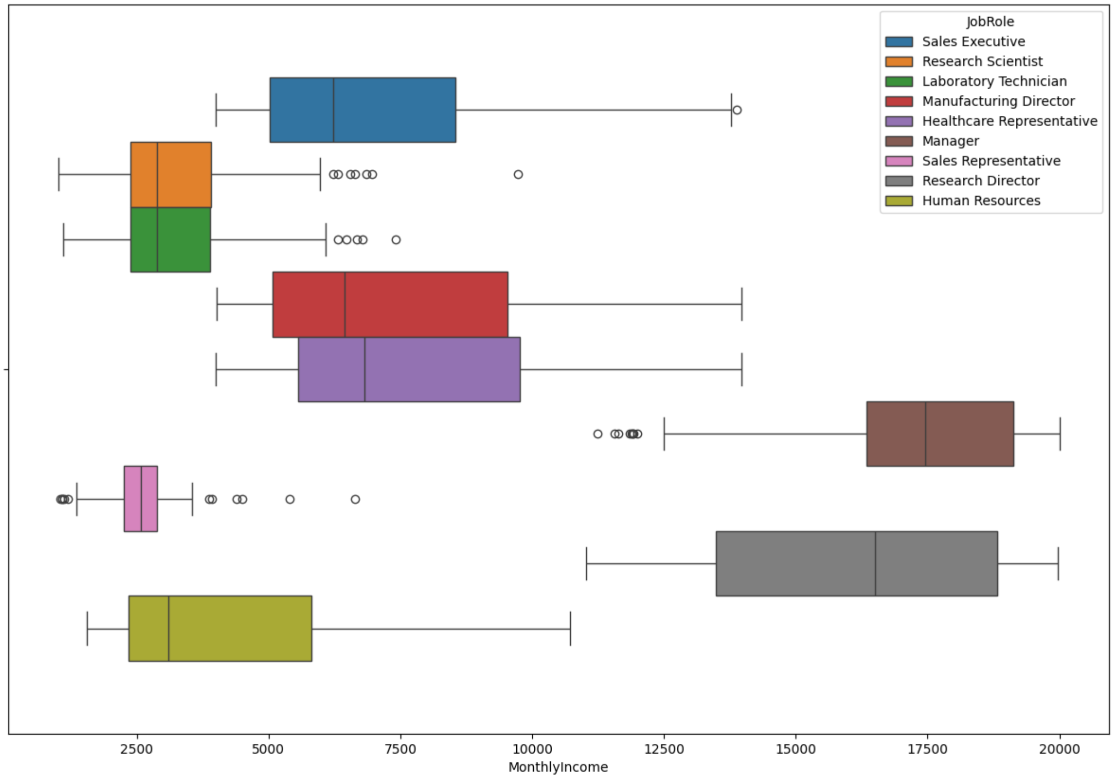
[45]:
```
# Let's see the Gender vs. Monthly Income
plt.figure(figsize=(15, 10))
sns.boxplot(x = 'MonthlyIncome', hue= 'Gender', data = employee_df,␣
 ↪palette='husl')
```

[45]:
```
<Axes: xlabel='MonthlyIncome'>
```



[46]:
```
# Let's see the monthly income vs. job role
plt.figure(figsize=(15, 10))
sns.boxplot(x= 'MonthlyIncome', hue = 'JobRole', data = employee_df,␣
 ↪palette="tab10", legend=True)
```

[46]: <Axes: xlabel='MonthlyIncome'>



0.0.7 Create Testing And Training Dataset:

[47]:
```
employee_df.describe(include='object')
```

[47]:
```
       BusinessTravel              Department EducationField Gender  \
count            1470                    1470           1470   1470
unique              3                       3              6      2
top     Travel_Rarely  Research & Development  Life Sciences   Male
freq             1043                     961            606    882

                JobRole MaritalStatus
count              1470          1470
unique                9             3
top     Sales Executive       Married
freq                326           673
```

[48]:
```
employee_df.head(5)
```

[48]:
```
   Age  Attrition     BusinessTravel  DailyRate              Department  \
0   41          1      Travel_Rarely       1102                   Sales
1   49          0  Travel_Frequently        279  Research & Development
2   37          1      Travel_Rarely       1373  Research & Development
```

```
3   33           0  Travel_Frequently       1392  Research & Development
4   27           0      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EnvironmentSatisfaction  \
0                 1          2  Life Sciences                        2
1                 8          1  Life Sciences                        3
2                 2          2          Other                        4
3                 3          4  Life Sciences                        4
4                 2          1        Medical                        1

   Gender  …  PerformanceRating  RelationshipSatisfaction  StockOptionLevel  \
0  Female  …                  3                         1                 0
1    Male  …                  4                         4                 1
2    Male  …                  3                         2                 0
3  Female  …                  3                         3                 0
4    Male  …                  3                         4                 1

  TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                 8                      0               1               6
1                10                      3               3              10
2                 7                      3               3               0
3                 8                      3               3               8
4                 6                      3               3               2

   YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                   4                        0                     5
1                   7                        1                     7
2                   0                        0                     0
3                   7                        3                     0
4                   2                        2                     2

[5 rows x 31 columns]
```

[49]:
```
employee_df.shape
```

[49]:
```
(1470, 31)
```

[50]:
```
employee_df.describe(include='object')
```

[50]:
```
       BusinessTravel              Department EducationField Gender  \
count            1470                    1470           1470   1470
unique              3                       3              6      2
top     Travel_Rarely  Research & Development  Life Sciences   Male
freq             1043                     961            606    882

                JobRole MaritalStatus
count              1470          1470
```

```
unique                 9           3
top      Sales Executive     Married
freq                 326         673
```

[51]:
```
employee_df['BusinessTravel'].unique()
```

[51]:
```
array(['Travel_Rarely', 'Travel_Frequently', 'Non-Travel'], dtype=object)
```

0.0.8 Categorical Column Encoding

[52]:
```
# Creating dataset with catgeorial columns:
X_cat = employee_df[['BusinessTravel', 'Department', 'EducationField',␣
 ↪'Gender', 'JobRole', 'MaritalStatus']]
X_cat
```

[52]:
```
         BusinessTravel              Department EducationField  Gender  \
0         Travel_Rarely                   Sales  Life Sciences  Female
1     Travel_Frequently  Research & Development  Life Sciences    Male
2         Travel_Rarely  Research & Development          Other    Male
3     Travel_Frequently  Research & Development  Life Sciences  Female
4         Travel_Rarely  Research & Development        Medical    Male
…                   …                       …            …      …
1465  Travel_Frequently  Research & Development        Medical    Male
1466      Travel_Rarely  Research & Development        Medical    Male
1467      Travel_Rarely  Research & Development  Life Sciences    Male
1468  Travel_Frequently                   Sales        Medical    Male
1469      Travel_Rarely  Research & Development        Medical    Male

                        JobRole MaritalStatus
0               Sales Executive        Single
1            Research Scientist       Married
2         Laboratory Technician        Single
3            Research Scientist       Married
4         Laboratory Technician       Married
…                           …           …
1465      Laboratory Technician       Married
1466  Healthcare Representative       Married
1467     Manufacturing Director       Married
1468            Sales Executive       Married
1469      Laboratory Technician       Married

[1470 rows x 6 columns]
```

[53]:
```
from sklearn.preprocessing import OneHotEncoder

onehotencoder = OneHotEncoder()
```

```
X_cat = onehotencoder.fit_transform(X_cat).toarray() #return the valyes in␣
 ↪series
type(X_cat)
```

[53]: numpy.ndarray

[54]:
```
X_cat.shape
```

[54]: (1470, 26)

[55]:
```
X_cat = pd.DataFrame(X_cat)
```

[56]:
```
X_cat
```

[56]:
```
        0    1    2    3    4    5    6    7    8    9  …   16   17   18  \
0     0.0  0.0  1.0  0.0  0.0  1.0  0.0  1.0  0.0  0.0  …  0.0  0.0  0.0
1     0.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …  0.0  0.0  0.0
2     0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  0.0  …  1.0  0.0  0.0
3     0.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …  0.0  0.0  0.0
4     0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …  1.0  0.0  0.0
…     …    …    …    …    …    …    …    …    …    …  …    …    …    …
1465  0.0  1.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …  1.0  0.0  0.0
1466  0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …  0.0  0.0  0.0
1467  0.0  0.0  1.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …  0.0  0.0  1.0
1468  0.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0  1.0  …  0.0  0.0  0.0
1469  0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …  1.0  0.0  0.0

       19   20   21   22   23   24   25
0     0.0  0.0  1.0  0.0  0.0  0.0  1.0
1     0.0  1.0  0.0  0.0  0.0  1.0  0.0
2     0.0  0.0  0.0  0.0  0.0  0.0  1.0
3     0.0  1.0  0.0  0.0  0.0  1.0  0.0
4     0.0  0.0  0.0  0.0  0.0  1.0  0.0
…     …    …    …    …    …    …    …
1465  0.0  0.0  0.0  0.0  0.0  1.0  0.0
1466  0.0  0.0  0.0  0.0  0.0  1.0  0.0
1467  0.0  0.0  0.0  0.0  0.0  1.0  0.0
1468  0.0  0.0  1.0  0.0  0.0  1.0  0.0
1469  0.0  0.0  0.0  0.0  0.0  1.0  0.0

[1470 rows x 26 columns]
```

[57]:
```
X_numerical = employee_df[['Age', 'DailyRate', 'DistanceFromHome', 'Education',␣
 ↪'EnvironmentSatisfaction', 'HourlyRate',
                              'JobInvolvement', 'JobLevel',␣
 ↪'JobSatisfaction',        'MonthlyIncome', 'MonthlyRate',
```

```
                              'NumCompaniesWorked', 'OverTime',␣
↪'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction',
                              'StockOptionLevel',␣
↪'TotalWorkingYears',        'TrainingTimesLastYear',␣
↪'WorkLifeBalance',
                              'YearsAtCompany'        ,'YearsInCurrentRole',␣
↪'YearsSinceLastPromotion', 'YearsWithCurrManager']]
X_numerical
```

[57]:

```
      Age  DailyRate  DistanceFromHome  Education  EnvironmentSatisfaction  \
0      41       1102                 1          2                        2
1      49        279                 8          1                        3
2      37       1373                 2          2                        4
3      33       1392                 3          4                        4
4      27        591                 2          1                        1
…     …         …                 …        …                      …
1465   36        884                23          2                        3
1466   39        613                 6          1                        4
1467   27        155                 4          3                        2
1468   49       1023                 2          3                        4
1469   34        628                 8          3                        2

      HourlyRate  JobInvolvement  JobLevel  JobSatisfaction  MonthlyIncome  \
0             94               3         2                4           5993
1             61               2         2                2           5130
2             92               2         1                3           2090
3             56               3         1                3           2909
4             40               3         1                2           3468
…            …              …        …               …            …
1465          41               4         2                4           2571
1466          42               2         3                1           9991
1467          87               4         2                2           6142
1468          63               2         2                2           5390
1469          82               4         2                3           4404

      …  PerformanceRating  RelationshipSatisfaction  StockOptionLevel  \
0     …                  3                         1                 0
1     …                  4                         4                 1
2     …                  3                         2                 0
3     …                  3                         3                 0
4     …                  3                         4                 1
…    …                …                        …                …
1465  …                  3                         3                 1
1466  …                  3                         1                 1
1467  …                  4                         2                 1
1468  …                  3                         4                 0
1469  …                  3                         1                 0
```

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

      YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
0                  6                   4                        0
1                 10                   7                        1
2                  0                   0                        0
3                  8                   7                        3
4                  2                   2                        2
…                …                   …                      …
1465               5                   2                        0
1466               7                   7                        1
1467               6                   2                        0
1468               9                   6                        0
1469               4                   3                        1

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

[1470 rows x 24 columns]
```

[58]:
```
X_all = pd.concat([X_cat, X_numerical], axis = 1)
X_all
```

[58]:
```
        0    1    2    3    4    5    6    7    8    9  …  \
0     0.0  0.0  1.0  0.0  0.0  1.0  0.0  1.0  0.0  0.0  …
```

```
1     0.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …
2     0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  0.0  …
3     0.0  1.0  0.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …
4     0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …
…    …   …   …   …   …   …   …   …   …   …   …   …
1465  0.0  1.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …
1466  0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …
1467  0.0  0.0  1.0  0.0  1.0  0.0  0.0  1.0  0.0  0.0  …
1468  0.0  1.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0  1.0  …
1469  0.0  0.0  1.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  …

      PerformanceRating  RelationshipSatisfaction  StockOptionLevel  \
0                     3                         1                 0
1                     4                         4                 1
2                     3                         2                 0
3                     3                         3                 0
4                     3                         4                 1
…                   …                       …                 …
1465                  3                         3                 1
1466                  3                         1                 1
1467                  4                         2                 1
1468                  3                         4                 0
1469                  3                         1                 0

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

      YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
0                  6                   4                        0
1                 10                   7                        1
2                  0                   0                        0
3                  8                   7                        3
4                  2                   2                        2
…                …                  …                      …
1465               5                   2                        0
1466               7                   7                        1
1467               6                   2                        0
```

```
1468              9                  6                        0
1469              4                  3                        1

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

[1470 rows x 50 columns]
```

0.0.9 Feature Scaling

[59]:
```python
from sklearn.preprocessing import MinMaxScaler

X_all.columns = X_all.columns.astype(str)
scaler = MinMaxScaler()
X = scaler.fit_transform(X_all)
```

[60]:
```python
type(X)
```

[60]:
```
numpy.ndarray
```

[61]:
```python
X
```

[61]:
```
array([[0.        , 0.        , 1.        , …, 0.22222222, 0.        ,
        0.29411765],
       [0.        , 1.        , 0.        , …, 0.38888889, 0.06666667,
        0.41176471],
       [0.        , 0.        , 1.        , …, 0.        , 0.        ,
        0.        ],
       …,
       [0.        , 0.        , 1.        , …, 0.11111111, 0.        ,
        0.17647059],
       [0.        , 1.        , 0.        , …, 0.33333333, 0.        ,
        0.47058824],
       [0.        , 0.        , 1.        , …, 0.16666667, 0.06666667,
        0.11764706]])
```

[62]:
```python
y = employee_df['Attrition']
y
```

[62]:
```
0       1
1       0
2       1
3       0
4       0
       ..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Attrition, Length: 1470, dtype: int64
```

0.0.10 Model Testing:

[63]:
```python
# Understanding Logistic Regression, Decision Tree Classifier and Random forest␣
↪Classifier :

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25)
```

[64]:
```python
X_train.shape
```

[64]:
```
(1102, 50)
```

[65]:
```python
X_test.shape
```

[65]:
```
(368, 50)
```

[66]:
```python
# Testing For Logistic Regression:
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

model = LogisticRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
```

[67]:
```python
y_pred
```

[67]:
```
array([1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
```

```
       0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)
```

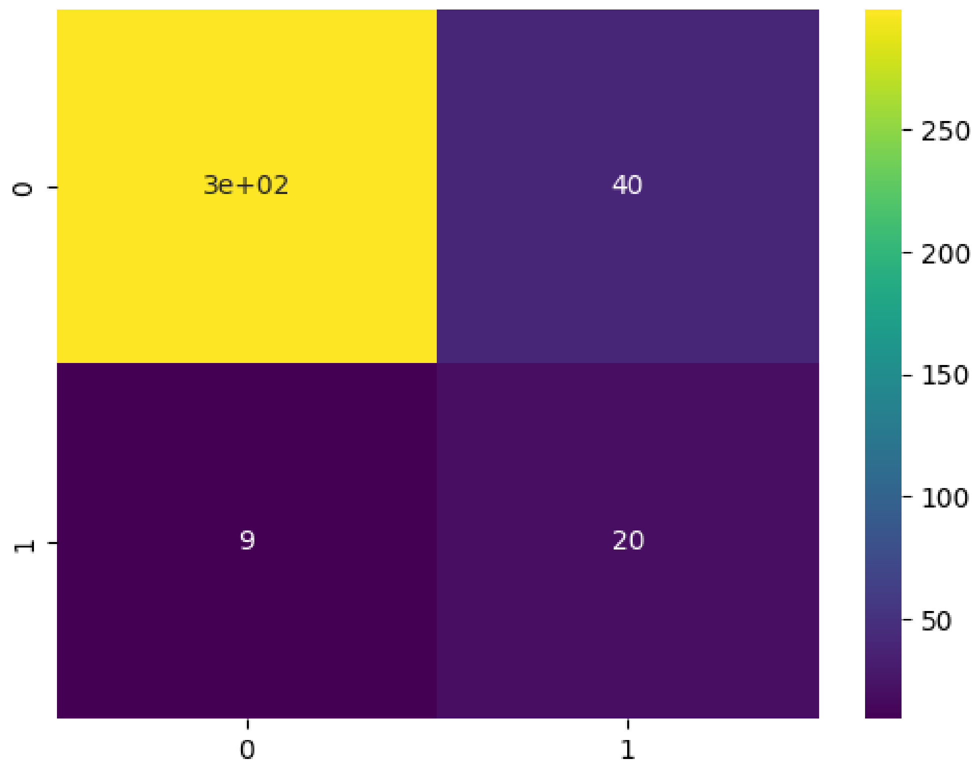
- Prediction of 0 suggests that the model estimates the probability of the instance belonging to the negative class is greater than 0.5.
- Prediction of 1 suggests that the model estimates the probability of the instance belonging to the positive class is greater than 0.5.

[68]:
```
# Testing for Confusion matrix:
from sklearn.metrics import confusion_matrix, classification_report

print("Accuracy of prediction teast: {} %".format( 100 * accuracy_score(y_pred,␣
 ↪y_test)))
```

```
Accuracy of prediction teast: 86.68478260869566 %
```

[69]:
```
# Testing Set Performance
cus_cmap = sns.color_palette("viridis", as_cmap=True)
cm = confusion_matrix(y_pred, y_test)
sns.heatmap(cm, annot=True, cmap=cus_cmap)
```

[69]: `<Axes: >`



[70]:
```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.88      0.97      0.92       308
           1       0.69      0.33      0.45        60

    accuracy                           0.87       368
   macro avg       0.79      0.65      0.69       368
weighted avg       0.85      0.87      0.85       368
```

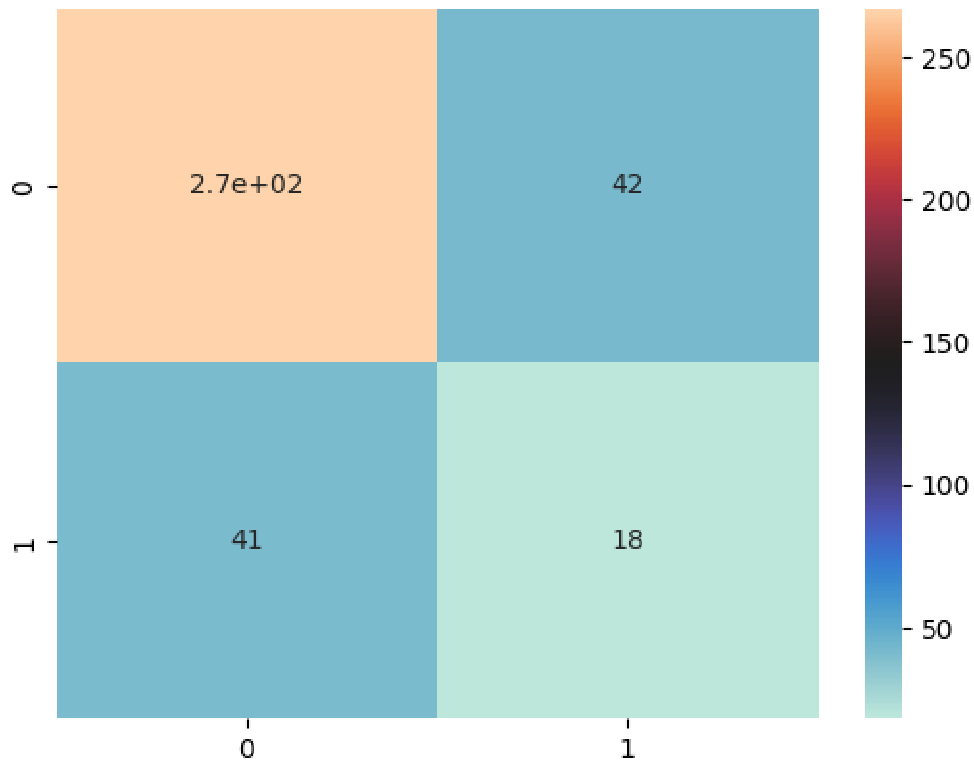
- For class 0, F1-score for class 0 is 0.94, reflecting a good balance between precision and recall. With a high support of 311, indicating a large number of instances for class 0, the model's performance on this class seems robust.
- For class 1, F1-score for class 1 is 0.53, indicating a moderate balance between precision and recall. With a support of 57, indicating a smaller number of instances for class 1, the model's performance on this class is less reliable compared to class 0.
- The overall accuracy of the model is 0.89, indicating that it correctly classified approximately 89% .

[71]:
```
print("Total Record of 1 ",employee_df[employee_df['Attrition']==1].shape[0])
print("Total Record of 0 ",employee_df[employee_df['Attrition']==0].shape[0])
```

```
Total Record of 1  237
Total Record of 0  1233
```

[72]:
```python
# Testing for Decision Tree Classifier:
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
# Testing Set Performance
cus_cmap = sns.color_palette("icefire", as_cmap=True)
cm = confusion_matrix(y_pred, y_test)
sns.heatmap(cm, annot=True,cmap=cus_cmap)
```

[72]: <Axes: >



[73]:
```python
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.86      0.87      0.87       308
           1       0.31      0.30      0.30        60
```

```
    accuracy                           0.77       368
   macro avg       0.58      0.58      0.58       368
weighted avg       0.77      0.77      0.77       368
```

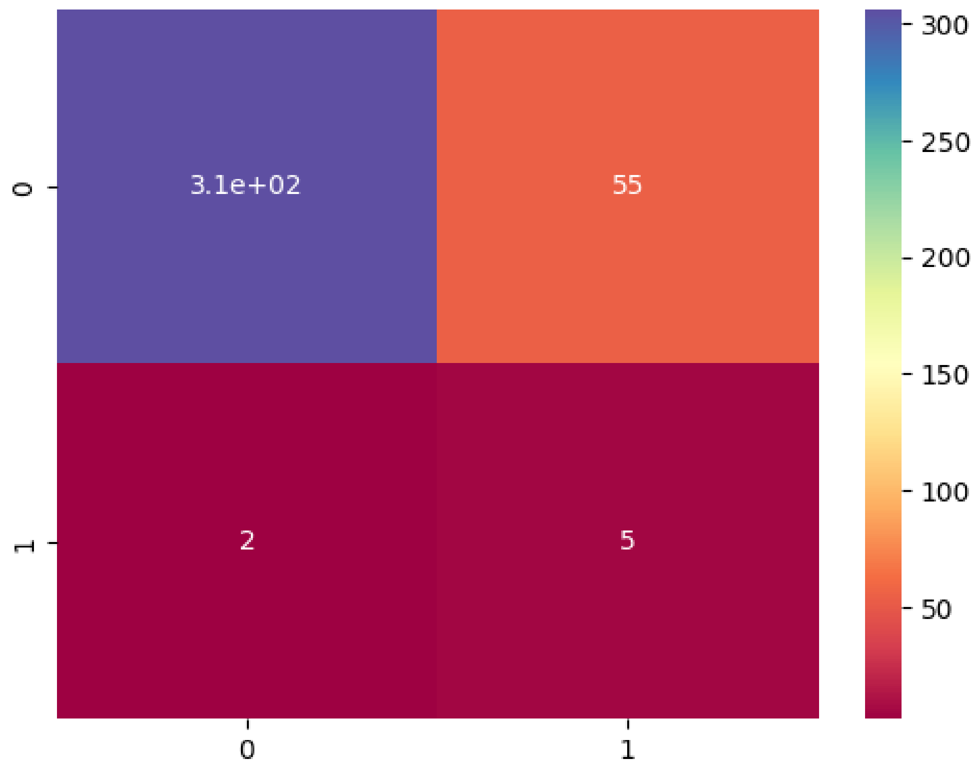
- For class 0, F1-score for class 0 is 0.87, reflecting a good balance between precision and recall. With a support of 311, indicating a large number of instances for class 0, the model's performance on this class seems robust.
- For class 1, F1-score for class 1 is 0.36, indicating a relatively low balance between precision and recall. With a support of 57, indicating a smaller number of instances for class 1, the model's performance on this class is less reliable compared to class 0.
- The overall accuracy of the model is 0.79, indicating that it correctly classified approximately 79%.

[74]:
```
# Testing for Random Forest Classifier
from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier()
model.fit(X_train, y_train)
```

[74]:
```
RandomForestClassifier()
```

[75]:
```
# Testing Set Performance
model=RandomForestClassifier(n_estimators=150,criterion='entropy',random_state=100)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

cus_cmap = sns.color_palette("Spectral", as_cmap=True)
cm = confusion_matrix(y_pred, y_test)
sns.heatmap(cm, annot=True, cmap=cus_cmap)
```

[75]:
```
<Axes: >
```



[76]:
```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.85      0.99      0.91       308
           1       0.71      0.08      0.15        60

    accuracy                           0.85       368
   macro avg       0.78      0.54      0.53       368
weighted avg       0.83      0.85      0.79       368
```

- For class 0, F1-score for class 0 is 0.93, reflecting a high balance between precision and recall. With a support of 311, indicating a large number of instances for class 0, the model's performance on this class seems robust.
- For class 1, F1-score for class 1 is 0.35, indicating a relatively low balance between precision and recall. With a support of 57, indicating a smaller number of instances for class 1, the model's performance on this class is less reliable compared to class 0.
- The overall accuracy of the model is 0.88, indicating that it correctly classified approximately 88%

Conclusion For Model Testing:

Logistic Regression:

- Achieves an accuracy of 0.89 with relatively balanced precision and recall for class 0.
- Struggles with recall for class 1, indicating difficulty in correctly identifying instances of that class.

Decision Tree Classifier:

- Shows decent accuracy at 0.79 with better precision and recall for class 0 compared to class 1.
- Struggles with recall for class 1, similar to logistic regression.

Random Forest Classifier:

- Demonstrates the highest accuracy of 0.88 among the three classifiers.
- Shows excellent precision and recall for class 0, but struggles with recall for class 1 similar to the other models.

Overall, while all three classifiers perform well in classifying instances of class 0, they exhibit challenges in correctly identifying instances of class 1, especially in recall. This suggests potential issues related to class imbalance or difficulty in capturing the characteristics of class 1. Further investigation and possibly model tuning, such as adjusting class weights, collecting more data for class 1, or using different feature engineering techniques, may be necessary to improve the performance, particularly for class 1 predictions.