In [ ]:
```
import pandas as pd
```

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

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

Out[ ]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	8
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	10
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	7
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	8
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	6

5 rows × 35 columns

In [ ]:
```
df = df.drop(columns=["EmployeeCount","JobLevel","JobInvolvement","JobSatisfaction","EnvironmentSatisfaction","RelationshipSatisfaction","StockOptionLevel","WorkLifeBalance"])
```

In [ ]:
```
df.duplicated().sum()
```

Out[ ]:
```
0
```

In [ ]:
```
df.isna().sum()
```

Out[ ]:
```
Age                       0
Attrition                 0
BusinessTravel            0
DailyRate                 0
Department                0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeNumber            0
Gender                    0
HourlyRate                0
JobRole                   0
MaritalStatus             0
MonthlyIncome             0
MonthlyRate               0
NumCompaniesWorked        0
OverTime                  0
PercentSalaryHike         0
PerformanceRating         0
StandardHours             0
TotalWorkingYears         0
TrainingTimesLastYear     0
YearsAtCompany            0
YearsInCurrentRole        0
YearsSinceLastPromotion   0
YearsWithCurrManager      0
Age Category              0
Income Category           0
dtype: int64
```

In [ ]:
```
df.isnull().sum()
```

Out[ ]:
```
Age                       0
Attrition                 0
BusinessTravel            0
DailyRate                 0
Department                0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeNumber            0
Gender                    0
HourlyRate                0
JobRole                   0
MaritalStatus             0
MonthlyIncome             0
MonthlyRate               0
NumCompaniesWorked        0
Over18                    0
OverTime                  0
PercentSalaryHike         0
PerformanceRating         0
StandardHours             0
TotalWorkingYears         0
TrainingTimesLastYear     0
YearsAtCompany            0
YearsInCurrentRole        0
YearsSinceLastPromotion   0
YearsWithCurrManager      0
dtype: int64
```

In [ ]:
```
categ_data = df.select_dtypes(include=['object'])

for colname in categ_data.columns:
    print (categ_data[colname].value_counts(), '\n')
```

```
Attrition
No     1233
Yes     237
Name: count, dtype: int64

BusinessTravel
Travel_Rarely        1043
Travel_Frequently     277
Non-Travel            150
Name: count, dtype: int64

Department
Research & Development    961
Sales                     446
Human Resources            63
Name: count, dtype: int64

EducationField
Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: count, dtype: int64

Gender
Male      882
Female    588
Name: count, dtype: int64

JobRole
Sales Executive              326
Research Scientist           292
Laboratory Technician        259
Manufacturing Director       145
Healthcare Representative    131
Manager                      102
Sales Representative          83
Research Director             80
Human Resources               52
Name: count, dtype: int64

MaritalStatus
Married     673
Single      470
Divorced    327
Name: count, dtype: int64

Over18
Y    1470
Name: count, dtype: int64

OverTime
No     1054
Yes     416
Name: count, dtype: int64
```

In [ ]:
```
df = df.drop(columns='Over18')
```

In [ ]:
```
for i, age in enumerate(df['Age']):
    if age >= 18 and age <= 30:
        df.loc[i, 'Age Category'] = '18 - 30'
    elif age > 30 and age <= 40:
        df.loc[i, 'Age Category'] = '31 - 40'
    elif age > 40 and age <= 50:
        df.loc[i, 'Age Category'] = '41 - 50'
    elif age > 50 and age <= 64:
        df.loc[i, 'Age Category'] = '51 - 64'
    else:
        df.loc[i, 'Age Category'] = '65+'
```

In [ ]:
```
for i, income in enumerate(df['MonthlyIncome']):
    if income >= 0 and income < 5000:
        df.loc[i, 'Income Category'] = '<5k'
    elif income >= 5000 and income < 10000:
        df.loc[i, 'Income Category'] = '5k - 10k'
    elif income >= 10000 and income < 15000:
        df.loc[i, 'Income Category'] = '10k - 15k'
    else:
        df.loc[i, 'Income Category'] = '15k+'
```

In [ ]:
```
df.to_csv('Hr Analytics.csv', index=False)
```