In [3]:

```python
import pandas as pd
df = pd.read_csv('HR Data.csv')
df
```

Out[3]:

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

1470 rows × 35 columns

In [4]:
```
# Display the data types of each column
print("Data Types of Columns:")
print(df.dtypes)
```

```
Data Types of Columns:
Age                          int64
Attrition                   object
BusinessTravel              object
DailyRate                    int64
Department                  object
DistanceFromHome             int64
Education                    int64
EducationField              object
EmployeeCount                int64
EmployeeNumber               int64
EnvironmentSatisfaction      int64
Gender                      object
HourlyRate                   int64
JobInvolvement               int64
JobLevel                     int64
JobRole                     object
JobSatisfaction              int64
MaritalStatus               object
MonthlyIncome                int64
MonthlyRate                  int64
NumCompaniesWorked           int64
Over18                      object
OverTime                    object
PercentSalaryHike            int64
PerformanceRating            int64
RelationshipSatisfaction     int64
StandardHours                int64
StockOptionLevel             int64
TotalWorkingYears            int64
TrainingTimesLastYear        int64
WorkLifeBalance              int64
YearsAtCompany               int64
YearsInCurrentRole           int64
YearsSinceLastPromotion      int64
YearsWithCurrManager         int64
dtype: object
```

In [5]:
```python
# Display common information about the columns
print("\nCommon Information about Columns:")
for column in df.columns:
    unique_values = df[column].nunique()
    data_type = df[column].dtype
    if unique_values <= 10:  # Adjust the threshold as needed
        values = df[column].unique()
        print(f"{column}: {data_type}, {unique_values} unique values, Value
    else:
        print(f"{column}: {data_type}, {unique_values} unique values")
```

```
Common Information about Columns:
Age: int64, 43 unique values
Attrition: object, 2 unique values, Values: ['Yes' 'No']
BusinessTravel: object, 3 unique values, Values: ['Travel_Rarely' 'Travel_
Frequently' 'Non-Travel']
DailyRate: int64, 886 unique values
Department: object, 3 unique values, Values: ['Sales' 'Research & Developm
ent' 'Human Resources']
DistanceFromHome: int64, 29 unique values
Education: int64, 5 unique values, Values: [2 1 4 3 5]
EducationField: object, 6 unique values, Values: ['Life Sciences' 'Other'
'Medical' 'Marketing' 'Technical Degree'
 'Human Resources']
EmployeeCount: int64, 1 unique values, Values: [1]
EmployeeNumber: int64, 1470 unique values
EnvironmentSatisfaction: int64, 4 unique values, Values: [2 3 4 1]
Gender: object, 2 unique values, Values: ['Female' 'Male']
HourlyRate: int64, 71 unique values
JobInvolvement: int64, 4 unique values, Values: [3 2 4 1]
JobLevel: int64, 5 unique values, Values: [2 1 3 4 5]
JobRole: object, 9 unique values, Values: ['Sales Executive' 'Research Sci
entist' 'Laboratory Technician'
 'Manufacturing Director' 'Healthcare Representative' 'Manager'
 'Sales Representative' 'Research Director' 'Human Resources']
JobSatisfaction: int64, 4 unique values, Values: [4 2 3 1]
MaritalStatus: object, 3 unique values, Values: ['Single' 'Married' 'Divor
ced']
MonthlyIncome: int64, 1349 unique values
MonthlyRate: int64, 1427 unique values
NumCompaniesWorked: int64, 10 unique values, Values: [8 1 6 9 0 4 5 2 7 3]
Over18: object, 1 unique values, Values: ['Y']
OverTime: object, 2 unique values, Values: ['Yes' 'No']
PercentSalaryHike: int64, 15 unique values
PerformanceRating: int64, 2 unique values, Values: [3 4]
RelationshipSatisfaction: int64, 4 unique values, Values: [1 4 2 3]
StandardHours: int64, 1 unique values, Values: [80]
StockOptionLevel: int64, 4 unique values, Values: [0 1 3 2]
TotalWorkingYears: int64, 40 unique values
TrainingTimesLastYear: int64, 7 unique values, Values: [0 3 2 5 1 4 6]
WorkLifeBalance: int64, 4 unique values, Values: [1 3 2 4]
YearsAtCompany: int64, 37 unique values
YearsInCurrentRole: int64, 19 unique values
YearsSinceLastPromotion: int64, 16 unique values
YearsWithCurrManager: int64, 18 unique values
```

```
In [6]: # Data Cleansing
        # Remove unnecessary columns
        unnecessary_columns = ['EmployeeCount', 'EmployeeNumber', 'Over18', 'Standa
        df_cleaned = df.drop(columns=unnecessary_columns)
```

```
In [7]: # Rename columns
        new_column_names = {
            'MonthlyIncome': 'Income',
            'YearsAtCompany': 'Tenure'
            # Add more renames as needed
        }
        df_cleaned = df_cleaned.rename(columns=new_column_names)
```

```
In [8]: # Eliminate redundant entries (if any)
        df_cleaned = df_cleaned.drop_duplicates()
```

```
In [9]: # Eliminate NaN values
        df_cleaned = df_cleaned.dropna()
```

In [10]:
```python
# Display the cleaned dataset
print(df_cleaned.head())
```

```
   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  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  StockOptionLe
vel  \
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  Tenure  \
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]
```

In [ ]: