```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('/content/Attrition data.csv')
df
```

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...
...	...	...	...	...	...	...	...	...	...	...	...
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...
4409	4410	40	No	Travel_Rarely	Research & Development	28	3	Medical	1	Male	...

4410 rows × 29 columns

```
df.head()
```

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...	Tot
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...	
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...	
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...	
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...	
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...	

5 rows × 29 columns

```
df.dtypes
```

```
EmployeeID                int64
Age                       int64
Attrition                object
BusinessTravel           object
Department               object
DistanceFromHome          int64
Education                 int64
EducationField           object
EmployeeCount             int64
Gender                   object
JobLevel                  int64
JobRole                  object
MaritalStatus            object
MonthlyIncome             int64
NumCompaniesWorked      float64
Over18                   object
PercentSalaryHike         int64
```

```
StandardHours               int64
StockOptionLevel            int64
TotalWorkingYears         float64
TrainingTimesLastYear       int64
YearsAtCompany              int64
YearsSinceLastPromotion     int64
YearsWithCurrManager        int64
EnvironmentSatisfaction   float64
JobSatisfaction           float64
WorkLifeBalance           float64
JobInvolvement              int64
PerformanceRating           int64
dtype: object
```

```
len(df)
```

```
4410
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 29 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   EmployeeID                4410 non-null   int64  
 1   Age                       4410 non-null   int64  
 2   Attrition                 4410 non-null   object 
 3   BusinessTravel            4410 non-null   object 
 4   Department                4410 non-null   object 
 5   DistanceFromHome          4410 non-null   int64  
 6   Education                 4410 non-null   int64  
 7   EducationField            4410 non-null   object 
 8   EmployeeCount             4410 non-null   int64  
 9   Gender                    4410 non-null   object 
 10  JobLevel                  4410 non-null   int64  
 11  JobRole                   4410 non-null   object 
 12  MaritalStatus             4410 non-null   object 
 13  MonthlyIncome             4410 non-null   int64  
 14  NumCompaniesWorked        4391 non-null   float64
 15  Over18                    4410 non-null   object 
 16  PercentSalaryHike         4410 non-null   int64  
 17  StandardHours             4410 non-null   int64  
 18  StockOptionLevel          4410 non-null   int64  
 19  TotalWorkingYears         4401 non-null   float64
 20  TrainingTimesLastYear     4410 non-null   int64  
 21  YearsAtCompany            4410 non-null   int64  
 22  YearsSinceLastPromotion   4410 non-null   int64  
 23  YearsWithCurrManager      4410 non-null   int64  
 24  EnvironmentSatisfaction   4385 non-null   float64
 25  JobSatisfaction           4390 non-null   float64
 26  WorkLifeBalance           4372 non-null   float64
 27  JobInvolvement            4410 non-null   int64  
 28  PerformanceRating         4410 non-null   int64  
dtypes: float64(5), int64(16), object(8)
memory usage: 999.3+ KB
```

```
df.isnull().sum()
```

```
EmployeeID                  0
Age                         0
Attrition                   0
BusinessTravel              0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
Gender                      0
JobLevel                    0
JobRole                     0
MaritalStatus               0
MonthlyIncome               0
NumCompaniesWorked         19
Over18                      0
PercentSalaryHike           0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           9
TrainingTimesLastYear       0
YearsAtCompany              0
```

```
YearsSinceLastPromotion     0
YearsWithCurrManager        0
EnvironmentSatisfaction    25
JobSatisfaction            20
WorkLifeBalance            38
JobInvolvement              0
PerformanceRating           0
dtype: int64
```

```
df = df.dropna()
df
```

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...
...	...	...	...	...	...	...	...	...	...	...	...
4404	4405	29	No	Travel_Rarely	Sales	4	3	Other	1	Female	...
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...

4300 rows × 29 columns

```
df.isnull().sum()
```

```
EmployeeID                0
Age                       0
Attrition                 0
BusinessTravel            0
Department                0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeCount             0
Gender                    0
JobLevel                  0
JobRole                   0
MaritalStatus             0
MonthlyIncome             0
NumCompaniesWorked        0
Over18                    0
PercentSalaryHike         0
StandardHours             0
StockOptionLevel          0
TotalWorkingYears         0
TrainingTimesLastYear     0
YearsAtCompany            0
YearsSinceLastPromotion   0
YearsWithCurrManager      0
EnvironmentSatisfaction   0
JobSatisfaction           0
WorkLifeBalance           0
JobInvolvement            0
PerformanceRating         0
dtype: int64
```

```
df = df.drop_duplicates()
df
```

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...
...	...	...	...	...	...	...	...	...	...	...	...
4404	4405	29	No	Travel_Rarely	Sales	4	3	Other	1	Female	...
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...

4300 rows × 29 columns

```
df.to_csv('Attrition Cleaned Data.csv', index=False)
```