

```
In [3]: import pandas as pd
```

```
In [5]: df=pd.read_csv("/home/ubuntu/Downloads/Aman_Singh_DSABDL/archive(2)/c
```

```
In [6]: df
```

Out[6]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	q
0	1	8	5	2	1	1	1	
1	1	6	1	11	1	1	1	
2	1	1	5	5	1	1	1	
3	1	8	2	15	1	1	1	
4	2	12	1	3	0	1	1	
...
4419	1	1	6	15	1	1	1	
4420	1	1	2	15	1	1	19	
4421	1	1	1	12	1	1	1	
4422	1	1	1	9	1	1	1	
4423	1	5	1	15	1	1	9	

4424 rows × 35 columns

In [7]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 columns):
 #   Column                                          Non-Null Count
Dtype  -----
0   Marital status                                4424 non-null
int64
1   Application mode                              4424 non-null
int64
2   Application order                            4424 non-null
int64
3   Course                                         4424 non-null
int64
4   Daytime/evening attendance                  4424 non-null
int64
5   Previous qualification                       4424 non-null
int64
6   Nacionality                                  4424 non-null
int64
7   Mother's qualification                      4424 non-null
int64
8   Father's qualification                      4424 non-null
int64
9   Mother's occupation                         4424 non-null
int64
10  Father's occupation                         4424 non-null
int64
11  Displaced                                   4424 non-null
int64
12  Educational special needs                   4424 non-null
int64
13  Debtor                                       4424 non-null
int64
14  Tuition fees up to date                    4424 non-null
int64
15  Gender                                       4424 non-null
int64
16  Scholarship holder                         4424 non-null
int64
17  Age at enrollment                          4424 non-null
int64
18  International                               4424 non-null
int64
19  Curricular units 1st sem (credited)         4424 non-null
int64
20  Curricular units 1st sem (enrolled)         4424 non-null
int64
21  Curricular units 1st sem (evaluations)      4424 non-null
int64
22  Curricular units 1st sem (approved)         4424 non-null
int64
23  Curricular units 1st sem (grade)            4424 non-null
float64
24  Curricular units 1st sem (without evaluations) 4424 non-null
int64

```

```

    25 Curricular units 2nd sem (credited)          4424 non-null
int64
    26 Curricular units 2nd sem (enrolled)          4424 non-null
int64
    27 Curricular units 2nd sem (evaluations)       4424 non-null
int64
    28 Curricular units 2nd sem (approved)          4424 non-null
int64
    29 Curricular units 2nd sem (grade)             4424 non-null
float64
    30 Curricular units 2nd sem (without evaluations) 4424 non-null
int64
    31 Unemployment rate                           4424 non-null
float64
    32 Inflation rate                              4424 non-null
float64
    33 GDP                                           4424 non-null
float64
    34 Target                                       4424 non-null
object
dtypes: float64(5), int64(29), object(1)
memory usage: 1.2+ MB

```

In [8]: `df.describe()`

Out[8]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	
count	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424
mean	1.178571	6.886980	1.727848	9.899186	0.890823	2.531420	
std	0.605747	5.298964	1.313793	4.331792	0.311897	3.963707	
min	1.000000	1.000000	0.000000	1.000000	0.000000	1.000000	
25%	1.000000	1.000000	1.000000	6.000000	1.000000	1.000000	
50%	1.000000	8.000000	1.000000	10.000000	1.000000	1.000000	
75%	1.000000	12.000000	2.000000	13.000000	1.000000	1.000000	
max	6.000000	18.000000	9.000000	17.000000	1.000000	17.000000	

8 rows × 34 columns

```
In [9]: df.head(10)
```

```
Out[9]:
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	N quali
0	1	8	5	2	1	1	1	
1	1	6	1	11	1	1	1	
2	1	1	5	5	1	1	1	
3	1	8	2	15	1	1	1	
4	2	12	1	3	0	1	1	
5	2	12	1	17	0	12	1	
6	1	1	1	12	1	1	1	
7	1	9	4	11	1	1	1	
8	1	1	3	10	1	1	15	
9	1	1	1	10	1	1	1	

10 rows × 35 columns

```
In [10]: df.tail(10)
```

```
Out[10]:
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	q
4414	1	1	1	8	1	1	1	
4415	4	12	1	12	1	12	1	
4416	1	14	2	12	1	1	1	
4417	1	1	1	5	1	1	1	
4418	1	15	1	5	1	14	1	
4419	1	1	6	15	1	1	1	
4420	1	1	2	15	1	1	19	
4421	1	1	1	12	1	1	1	
4422	1	1	1	9	1	1	1	
4423	1	5	1	15	1	1	9	

10 rows × 35 columns

```
In [11]: df=df.dropna()
```

In [12]: df

Out[12]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	q
0	1	8	5	2	1	1	1	
1	1	6	1	11	1	1	1	
2	1	1	5	5	1	1	1	
3	1	8	2	15	1	1	1	
4	2	12	1	3	0	1	1	
...
4419	1	1	6	15	1	1	1	
4420	1	1	2	15	1	1	19	
4421	1	1	1	12	1	1	1	
4422	1	1	1	9	1	1	1	
4423	1	5	1	15	1	1	9	

4424 rows × 35 columns

In [13]: df.describe()

Out[13]:

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	l
count	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000
mean	1.178571	6.886980	1.727848	9.899186	0.890823	2.531420	
std	0.605747	5.298964	1.313793	4.331792	0.311897	3.963707	
min	1.000000	1.000000	0.000000	1.000000	0.000000	1.000000	
25%	1.000000	1.000000	1.000000	6.000000	1.000000	1.000000	
50%	1.000000	8.000000	1.000000	10.000000	1.000000	1.000000	
75%	1.000000	12.000000	2.000000	13.000000	1.000000	1.000000	
max	6.000000	18.000000	9.000000	17.000000	1.000000	17.000000	

8 rows × 34 columns

In [14]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 columns):
 #   Column                                          Non-Null Count
Dtype  -----
----  -
0    Marital status                                4424 non-null
int64
1    Application mode                             4424 non-null
int64
2    Application order                           4424 non-null
int64
3    Course                                         4424 non-null
int64
4    Daytime/evening attendance                 4424 non-null
int64
5    Previous qualification                     4424 non-null
int64
6    Nacionality                                 4424 non-null
int64
7    Mother's qualification                     4424 non-null
int64
8    Father's qualification                     4424 non-null
int64
9    Mother's occupation                         4424 non-null
int64
10   Father's occupation                         4424 non-null
int64
11   Displaced                                  4424 non-null
int64
12   Educational special needs                 4424 non-null
int64
13   Debtor                                    4424 non-null
int64
14   Tuition fees up to date                   4424 non-null
int64
15   Gender                                    4424 non-null
int64
16   Scholarship holder                        4424 non-null
int64
17   Age at enrollment                         4424 non-null
int64
18   International                             4424 non-null
int64
19   Curricular units 1st sem (credited)       4424 non-null
int64
20   Curricular units 1st sem (enrolled)       4424 non-null
int64
21   Curricular units 1st sem (evaluations)    4424 non-null
int64
22   Curricular units 1st sem (approved)       4424 non-null
int64
23   Curricular units 1st sem (grade)          4424 non-null
float64
24   Curricular units 1st sem (without evaluations) 4424 non-null
int64

```

```

25 Curricular units 2nd sem (credited)          4424 non-null
int64
26 Curricular units 2nd sem (enrolled)          4424 non-null
int64
27 Curricular units 2nd sem (evaluations)        4424 non-null
int64
28 Curricular units 2nd sem (approved)          4424 non-null
int64
29 Curricular units 2nd sem (grade)             4424 non-null
float64
30 Curricular units 2nd sem (without evaluations) 4424 non-null
int64
31 Unemployment rate                            4424 non-null
float64
32 Inflation rate                              4424 non-null
float64
33 GDP                                          4424 non-null
float64
34 Target                                      4424 non-null
object
dtypes: float64(5), int64(29), object(1)
memory usage: 1.2+ MB

```

```
In [18]: df['Target']=df['Target'].astype(str)
```

```
In [19]: df
```

```
Out[19]:
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	q
0	1	8	5	2	1	1	1	
1	1	6	1	11	1	1	1	
2	1	1	5	5	1	1	1	
3	1	8	2	15	1	1	1	
4	2	12	1	3	0	1	1	
...	
4419	1	1	6	15	1	1	1	
4420	1	1	2	15	1	1	19	
4421	1	1	1	12	1	1	1	
4422	1	1	1	9	1	1	1	
4423	1	5	1	15	1	1	9	

4424 rows × 35 columns

```
In [20]: df.dtypes
```

```
Out[20]: Marital status          int64
Application mode                int64
Application order               int64
Course                         int64
Daytime/evening attendance     int64
Previous qualification          int64
Nacionality                    int64
Mother's qualification         int64
Father's qualification         int64
Mother's occupation            int64
Father's occupation            int64
Displaced                      int64
Educational special needs      int64
Debtor                         int64
Tuition fees up to date        int64
Gender                         int64
Scholarship holder             int64
Age at enrollment              int64
International                   int64
Curricular units 1st sem (credited)  int64
Curricular units 1st sem (enrolled)  int64
Curricular units 1st sem (evaluations) int64
Curricular units 1st sem (approved)  int64
Curricular units 1st sem (grade)      float64
Curricular units 1st sem (without evaluations) int64
Curricular units 2nd sem (credited)   int64
Curricular units 2nd sem (enrolled)   int64
Curricular units 2nd sem (evaluations) int64
Curricular units 2nd sem (approved)   int64
Curricular units 2nd sem (grade)      float64
Curricular units 2nd sem (without evaluations) int64
Unemployment rate                float64
Inflation rate                   float64
GDP                              float64
Target                           object
dtype: object
```

```
In [22]: df=df.astype({'Target':'string'})
```


In [23]: df.dtypes

```
Out[23]: Marital status          int64
Application mode          int64
Application order         int64
Course                   int64
Daytime/evening attendance int64
Previous qualification     int64
Nacionality               int64
Mother's qualification     int64
Father's qualification     int64
Mother's occupation       int64
Father's occupation       int64
Displaced                 int64
Educational special needs int64
Debtor                    int64
Tuition fees up to date   int64
Gender                    int64
Scholarship holder        int64
Age at enrollment         int64
International              int64
Curricular units 1st sem (credited) int64
Curricular units 1st sem (enrolled) int64
Curricular units 1st sem (evaluations) int64
Curricular units 1st sem (approved) int64
Curricular units 1st sem (grade) float64
Curricular units 1st sem (without evaluations) int64
Curricular units 2nd sem (credited) int64
Curricular units 2nd sem (enrolled) int64
Curricular units 2nd sem (evaluations) int64
Curricular units 2nd sem (approved) int64
Curricular units 2nd sem (grade) float64
Curricular units 2nd sem (without evaluations) int64
Unemployment rate        float64
Inflation rate            float64
GDP                       float64
Target                    string
dtype: object
```

In [24]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 columns):
 #   Column                                          Non-Null Count
Dtype  -----
----  -
0   Marital status                                4424 non-null
int64
1   Application mode                              4424 non-null
int64
2   Application order                            4424 non-null
int64
3   Course                                         4424 non-null
int64
4   Daytime/evening attendance                   4424 non-null
int64
5   Previous qualification                       4424 non-null
int64
6   Nacionality                                  4424 non-null
int64
7   Mother's qualification                      4424 non-null
int64
8   Father's qualification                      4424 non-null
int64
9   Mother's occupation                          4424 non-null
int64
10  Father's occupation                         4424 non-null
int64
11  Displaced                                   4424 non-null
int64
12  Educational special needs                   4424 non-null
int64
13  Debtor                                       4424 non-null
int64
14  Tuition fees up to date                     4424 non-null
int64
15  Gender                                       4424 non-null
int64
16  Scholarship holder                          4424 non-null
int64
17  Age at enrollment                           4424 non-null
int64
18  International                               4424 non-null
int64
19  Curricular units 1st sem (credited)         4424 non-null
int64
20  Curricular units 1st sem (enrolled)         4424 non-null
int64
21  Curricular units 1st sem (evaluations)      4424 non-null
int64
22  Curricular units 1st sem (approved)         4424 non-null
int64
23  Curricular units 1st sem (grade)            4424 non-null
float64
24  Curricular units 1st sem (without evaluations) 4424 non-null
int64

```

```

25 Curricular units 2nd sem (credited)      4424 non-null
int64
26 Curricular units 2nd sem (enrolled)      4424 non-null
int64
27 Curricular units 2nd sem (evaluations)   4424 non-null
int64
28 Curricular units 2nd sem (approved)      4424 non-null
int64
29 Curricular units 2nd sem (grade)         4424 non-null
float64
30 Curricular units 2nd sem (without evaluations) 4424 non-null
int64
31 Unemployment rate                       4424 non-null
float64
32 Inflation rate                         4424 non-null
float64
33 GDP                                    4424 non-null
float64
34 Target                                4424 non-null
string
dtypes: float64(5), int64(29), string(1)
memory usage: 1.2 MB

```

In [25]: `pd.get_dummies(df['Target'])`

Out[25]:

	Dropout	Enrolled	Graduate
0	1	0	0
1	0	0	1
2	1	0	0
3	0	0	1
4	0	0	1
...
4419	0	0	1
4420	1	0	0
4421	1	0	0
4422	0	0	1
4423	0	0	1

4424 rows × 3 columns

In []:

In [26]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 columns):
 #   Column                                          Non-Null Count
Dtype  -----
----  -
0    Marital status                               4424 non-null
int64
1    Application mode                             4424 non-null
int64
2    Application order                           4424 non-null
int64
3    Course                                       4424 non-null
int64
4    Daytime/evening attendance                 4424 non-null
int64
5    Previous qualification                     4424 non-null
int64
6    Nacionality                                4424 non-null
int64
7    Mother's qualification                     4424 non-null
int64
8    Father's qualification                     4424 non-null
int64
9    Mother's occupation                         4424 non-null
int64
10   Father's occupation                        4424 non-null
int64
11   Displaced                                 4424 non-null
int64
12   Educational special needs                 4424 non-null
int64
13   Debtor                                    4424 non-null
int64
14   Tuition fees up to date                   4424 non-null
int64
15   Gender                                    4424 non-null
int64
16   Scholarship holder                        4424 non-null
int64
17   Age at enrollment                         4424 non-null
int64
18   International                             4424 non-null
int64
19   Curricular units 1st sem (credited)       4424 non-null
int64
20   Curricular units 1st sem (enrolled)       4424 non-null
int64
21   Curricular units 1st sem (evaluations)    4424 non-null
int64
22   Curricular units 1st sem (approved)       4424 non-null
int64
23   Curricular units 1st sem (grade)          4424 non-null
float64
24   Curricular units 1st sem (without evaluations) 4424 non-null
int64

```

```
25 Curricular units 2nd sem (credited)          4424 non-null
int64
26 Curricular units 2nd sem (enrolled)          4424 non-null
int64
27 Curricular units 2nd sem (evaluations)        4424 non-null
int64
28 Curricular units 2nd sem (approved)          4424 non-null
int64
29 Curricular units 2nd sem (grade)             4424 non-null
float64
30 Curricular units 2nd sem (without evaluations) 4424 non-null
int64
31 Unemployment rate                            4424 non-null
float64
32 Inflation rate                              4424 non-null
float64
33 GDP                                          4424 non-null
float64
34 Target                                      4424 non-null
string
dtypes: float64(5), int64(29), string(1)
memory usage: 1.2 MB
```