PreProcessingPart1

April 24, 2025

```python
[5]: import pandas as pd
```

```python
[6]: df = pd.read_csv("student_depression_dataset.csv")
     df.head(10)
```

```
[6]:    id  Gender   Age           City Profession  Academic Pressure  \
     0   2    Male  33.0  Visakhapatnam    Student                5.0
     1   8  Female  24.0      Bangalore    Student                2.0
     2  26    Male  31.0       Srinagar    Student                3.0
     3  30  Female  28.0       Varanasi    Student                3.0
     4  32  Female  25.0         Jaipur    Student                4.0
     5  33    Male  29.0           Pune    Student                2.0
     6  52    Male  30.0          Thane    Student                3.0
     7  56  Female  30.0        Chennai    Student                2.0
     8  59    Male  28.0         Nagpur    Student                3.0
     9  62    Male  31.0         Nashik    Student                2.0

        Work Pressure  CGPA  Study Satisfaction  Job Satisfaction  \
     0            0.0  8.97                 2.0               0.0
     1            0.0  5.90                 5.0               0.0
     2            0.0  7.03                 5.0               0.0
     3            0.0  5.59                 2.0               0.0
     4            0.0  8.13                 3.0               0.0
     5            0.0  5.70                 3.0               0.0
     6            0.0  9.54                 4.0               0.0
     7            0.0  8.04                 4.0               0.0
     8            0.0  9.79                 1.0               0.0
     9            0.0  8.38                 3.0               0.0

             Sleep Duration Dietary Habits      Degree  \
     0           '5-6 hours'        Healthy    B.Pharm
     1           '5-6 hours'       Moderate        BSc
     2  'Less than 5 hours'        Healthy         BA
     3           '7-8 hours'       Moderate        BCA
     4           '5-6 hours'       Moderate     M.Tech
     5  'Less than 5 hours'        Healthy        PhD
     6           '7-8 hours'        Healthy        BSc
     7  'Less than 5 hours'      Unhealthy  'Class 12'
```

```
8          '7-8 hours'         Moderate         B.Ed
9  'Less than 5 hours'         Moderate          LLB
```

```
   Have you ever had suicidal thoughts ?  Work/Study Hours  Financial Stress  \
0                                    Yes               3.0               1.0
1                                     No               3.0               2.0
2                                     No               9.0               1.0
3                                    Yes               4.0               5.0
4                                    Yes               1.0               1.0
5                                     No               4.0               1.0
6                                     No               1.0               2.0
7                                     No               0.0               1.0
8                                    Yes              12.0               3.0
9                                    Yes               2.0               5.0
```

```
   Family History of Mental Illness  Depression
0                                No           1
1                               Yes           0
2                               Yes           0
3                               Yes           1
4                                No           0
5                                No           0
6                                No           0
7                               Yes           0
8                                No           1
9                                No           1
```

[7]: `df.describe()`

[7]:
```
                    id           Age  Academic Pressure  Work Pressure  \
count  27901.000000  27901.000000       27901.000000   27901.000000
mean   70442.149421     25.822300           3.141214       0.000430
std    40641.175216      4.905687           1.381465       0.043992
min        2.000000     18.000000           0.000000       0.000000
25%    35039.000000     21.000000           2.000000       0.000000
50%    70684.000000     25.000000           3.000000       0.000000
75%   105818.000000     30.000000           4.000000       0.000000
max   140699.000000     59.000000           5.000000       5.000000
```

```
               CGPA  Study Satisfaction  Job Satisfaction  Work/Study Hours  \
count  27901.000000        27901.000000      27901.000000      27901.000000
mean       7.656104            2.943837          0.000681          7.156984
std        1.470707            1.361148          0.044394          3.707642
min        0.000000            0.000000          0.000000          0.000000
25%        6.290000            2.000000          0.000000          4.000000
50%        7.770000            3.000000          0.000000          8.000000
75%        8.920000            4.000000          0.000000         10.000000
```

```
max        10.000000              5.000000           4.000000          12.000000


           Depression
count   27901.000000
mean        0.585499
std         0.492645
min         0.000000
25%         0.000000
50%         1.000000
75%         1.000000
max         1.000000
```

[8]: `print(df.dtypes)`

```
id                                    int64
Gender                               object
Age                                 float64
City                                 object
Profession                           object
Academic Pressure                   float64
Work Pressure                       float64
CGPA                                float64
Study Satisfaction                  float64
Job Satisfaction                    float64
Sleep Duration                       object
Dietary Habits                       object
Degree                               object
Have you ever had suicidal thoughts ?   object
Work/Study Hours                    float64
Financial Stress                     object
Family History of Mental Illness     object
Depression                            int64
dtype: object
```

0.0.1 Notes:

Looking at the head of the data and the data type of Financial Stress we can say that Financial Stress shouldn't be an Object type. We also need to convert float columns to int64 except CGPA column.