cleaning_process

May 19, 2025

[1]:
```
import pandas as pd

#reads csv file and initializes first row as headers and first column as index␣
 ↪column
df = pd.read_csv(r"uncleaned_student_depression_dataset.csv",index_col=␣
 ↪0,header = 0 )
```

[2]:
```
print(df.shape)
```

```
(27902, 17)
```

[3]:
```
#creates a new index column and keeps the old 'id' column too
df.reset_index(inplace=True)

#changes the name of new index column to 'serial number'
df.index.name = 'Serial Number'

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27902 entries, 0 to 27901
Data columns (total 18 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   id                                     27902 non-null  int64
 1   Gender                                 27902 non-null  object
 2   Age                                    27902 non-null  int64
 3   City                                   27902 non-null  object
 4   Profession                             27902 non-null  object
 5   Academic Pressure                      27902 non-null  int64
 6   Work Pressure                          27902 non-null  int64
 7   CGPA                                   27902 non-null  float64
 8   Study Satisfaction                     27902 non-null  int64
 9   Job Satisfaction                       27902 non-null  int64
 10  Sleep Duration                         27902 non-null  object
 11  Dietary Habits                         27902 non-null  object
 12  Degree                                 27902 non-null  object
 13  Have you ever had suicidal thoughts ?  27902 non-null  object
 14  Work/Study Hours                       27902 non-null  int64
```

```
 15  Financial Stress                        27902 non-null  object
 16  Family History of Mental Illness       27902 non-null  object
 17  Depression                              27902 non-null  int64
dtypes: float64(1), int64(8), object(9)
memory usage: 3.8+ MB
```

[4]:
```
#first five rows of the dataframe
df.head()
```

[4]:

Serial Number	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	1	Male	19	Delhi	Student	4	0	6.00	3	0	'6-7 hours'	Moderate	B.Com	Yes	8	4	No	1
1	2	Male	33	Visakhapatnam	Student	5	0	8.97	2	0	'5-6 hours'	Healthy	B.Pharm	Yes	3	1	No	1
2	8	Female	24	Bangalore	Student	2	0	5.90	5	0	'5-6 hours'	Moderate	BSc	No	3	2	Yes	0
3	26	Male	31	Srinagar	Student	3	0	7.03	5	0	'Less than 5 hours'	Healthy	BA	No	9	1	Yes	0
4	30	Female	28	Varanasi	Student	3	0	5.59	2	0	'7-8 hours'	Moderate	BCA	Yes	4			

```
4                              5                                 Yes           1
```

[5]:
```
df.describe()
```

[5]:
```
                 id           Age  Academic Pressure  Work Pressure  \
count  27902.000000  27902.000000       27902.000000   27902.000000
mean   70439.624830     25.822056           3.141244       0.000430
std    40642.634749      4.905770           1.381450       0.043991
min        1.000000     18.000000           0.000000       0.000000
25%    35035.250000     21.000000           2.000000       0.000000
50%    70673.000000     25.000000           3.000000       0.000000
75%   105817.000000     30.000000           4.000000       0.000000
max   140699.000000     59.000000           5.000000       5.000000

              CGPA  Study Satisfaction  Job Satisfaction  Work/Study Hours  \
count  27902.000000        27902.000000      27902.000000      27902.000000
mean       7.656045            2.943839          0.000681          7.157014
std        1.470714            1.361124          0.044394          3.707579
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  27902.000000
mean       0.585514
std        0.492642
min        0.000000
25%        0.000000
50%        1.000000
75%        1.000000
max        1.000000
```

[6]:
```
df['Gender'].nunique()
```

[6]:
```
2
```

[7]:
```
df['Gender'].unique()
```

[7]:
```
array(['Male', 'Female'], dtype=object)
```

[8]:
```
df['Gender'].value_counts()
```

[8]:
```
Gender
Male      15548
Female    12354
Name: count, dtype: int64
```

[9]:
```
df['Age'].nunique()
```

[9]:
```
34
```

[10]:
```
df['Age'].unique()
```

[10]:
```
array([19, 33, 24, 31, 28, 25, 29, 30, 27, 20, 23, 18, 21, 22, 34, 32, 26,
       39, 35, 42, 36, 58, 49, 38, 51, 44, 43, 46, 59, 54, 48, 56, 37, 41])
```

[11]:
```
df['Age'].value_counts()
```

[11]:
```
Age
24    2258
20    2237
28    2133
29    1950
33    1893
25    1784
21    1726
23    1645
18    1587
19    1561
34    1468
27    1462
31    1427
32    1262
22    1160
26    1155
30    1145
35      10
38       8
36       7
42       4
48       3
39       3
43       2
46       2
37       2
49       1
51       1
44       1
59       1
54       1
58       1
56       1
41       1
Name: count, dtype: int64
```

[12]:
```
df['City'].nunique()
```

[12]:
```
51
```

[13]:
```
df['City'].unique()
```

[13]:
```
array(['Delhi', 'Visakhapatnam', 'Bangalore', 'Srinagar', 'Varanasi',
       'Jaipur', 'Pune', 'Thane', 'Chennai', 'Nagpur', 'Nashik',
       'Vadodara', 'Kalyan', 'Rajkot', 'Ahmedabad', 'Kolkata', 'Mumbai',
       'Lucknow', 'Indore', 'Surat', 'Ludhiana', 'Bhopal', 'Meerut',
       'Agra', 'Ghaziabad', 'Hyderabad', 'Vasai-Virar', 'Kanpur', 'Patna',
       'Faridabad', 'Saanvi', 'M.Tech', 'Bhavna', 'City', '3',
       "'Less than 5 Kalyan'", 'Mira', 'Harsha', 'Vaanya', 'Gaurav',
       'Harsh', 'Reyansh', 'Kibara', 'Rashi', 'ME', 'M.Com', 'Nalyan',
       'Mihir', 'Nalini', 'Nandini', 'Khaziabad'], dtype=object)
```

[14]:
```
df['City'].value_counts()
```

[14]:
```
City
Kalyan                1570
Srinagar              1372
Hyderabad             1340
Vasai-Virar           1290
Lucknow               1155
Thane                 1139
Ludhiana              1111
Agra                  1094
Surat                 1078
Kolkata               1066
Jaipur                1036
Patna                 1007
Visakhapatnam          969
Pune                   968
Ahmedabad              951
Bhopal                 934
Chennai                885
Meerut                 825
Rajkot                 816
Delhi                  770
Bangalore              767
Ghaziabad              745
Mumbai                 699
Vadodara               694
Varanasi               685
Nagpur                 651
Indore                 643
Kanpur                 609
```

```
Nashik                    547
Faridabad                 461
Harsha                      2
Saanvi                      2
Bhavna                      2
City                        2
ME                          1
M.Com                       1
Nalyan                      1
Nandini                     1
Mihir                       1
Nalini                      1
Kibara                      1
Rashi                       1
'Less than 5 Kalyan'        1
Reyansh                     1
Harsh                       1
Gaurav                      1
Vaanya                      1
Mira                        1
3                           1
M.Tech                      1
Khaziabad                   1
Name: count, dtype: int64
```

[15]:
```
#incorrect values in the 'City' column which cannot be fixed
unwanted = ['Saanvi', 'M.Tech', 'Bhavna', 'City', '3',
        "'Less than 5 Kalyan'", 'Mira', 'Harsha', 'Vaanya', 'Gaurav',
        'Harsh', 'Reyansh', 'Kibara', 'Rashi', 'ME', 'M.Com', 'Nalyan',
        'Mihir', 'Nalini', 'Nandini']

#drops all the incorrect values which cannot be fixed
df.drop(df[df['City'].isin(unwanted)].index, inplace = True)

#fixes the fixabel incorrect values in 'City' column
df['City'] = df['City'].apply(lambda x : 'Ghaziabad' if x == 'Khaziabad' else x)
```

[16]:
```
df['Profession'].nunique()
```

[16]:
```
14
```

[17]:
```
df['Profession'].unique()
```

[17]:
```
array(['Student', "'Civil Engineer'", 'Architect', "'UX/UI Designer'",
       "'Digital Marketer'", "'Content Writer'",
       "'Educational Consultant'", 'Teacher', 'Manager', 'Chef', 'Doctor',
       'Lawyer', 'Entrepreneur', 'Pharmacist'], dtype=object)
```

[18]:
```
df['Profession'].value_counts()
```

[18]:
```
Profession
Student                     27847
Architect                       8
Teacher                         6
'Digital Marketer'              3
'Content Writer'                2
Chef                            2
Doctor                          2
Pharmacist                      2
'Civil Engineer'                1
'UX/UI Designer'                1
'Educational Consultant'        1
Manager                         1
Lawyer                          1
Entrepreneur                    1
Name: count, dtype: int64
```

[19]:
```
df['Academic Pressure'].nunique()
```

[19]:
```
6
```

[20]:
```
df['Academic Pressure'].unique()
```

[20]:
```
array([4, 5, 2, 3, 1, 0])
```

[21]:
```
df['Academic Pressure'].value_counts()
```

[21]:
```
Academic Pressure
3    7454
5    6293
4    5152
1    4797
2    4173
0       9
Name: count, dtype: int64
```

[22]:
```
df['Work Pressure'].nunique()
```

[22]:
```
3
```

[23]:
```
df['Work Pressure'].unique()
```

[23]:
```
array([0, 5, 2])
```

[24]:
```
df['Work Pressure'].value_counts()
```

[24]:
```
Work Pressure
0    27875
5        2
2        1
Name: count, dtype: int64
```

[25]:
```
df['CGPA'].nunique()
```

[25]:
```
332
```

[26]:
```
df['CGPA'].unique()
```

[26]:
```
array([ 6.  ,  8.97 ,  5.9  ,  7.03 ,  5.59 ,  8.13 ,  5.7  ,
        9.54 ,  8.04 ,  9.79 ,  8.38 ,  6.1  ,  7.04 ,  8.52 ,
        5.64 ,  8.58 ,  6.51 ,  7.25 ,  7.83 ,  9.93 ,  8.74 ,
        6.73 ,  5.57 ,  8.59 ,  7.1  ,  6.08 ,  5.74 ,  9.86 ,
        6.7  ,  6.21 ,  5.87 ,  6.37 ,  9.72 ,  5.88 ,  9.56 ,
        6.99 ,  5.24 ,  9.21 ,  7.85 ,  6.95 ,  5.86 ,  7.92 ,
        9.66 ,  8.94 ,  9.71 ,  7.87 ,  5.6  ,  7.9  ,  5.46 ,
        6.79 ,  8.7  ,  7.38 ,  8.5  ,  7.09 ,  9.82 ,  8.89 ,
        7.94 ,  9.11 ,  6.75 ,  7.53 ,  9.49 ,  9.01 ,  7.64 ,
        5.27 ,  9.44 ,  5.75 ,  7.51 ,  9.05 ,  6.38 ,  8.95 ,
        9.88 ,  5.32 ,  6.27 ,  7.7  ,  8.1  ,  9.59 ,  8.96 ,
        5.51 ,  7.43 ,  8.79 ,  9.95 ,  5.37 ,  6.86 ,  8.32 ,
        9.74 ,  5.66 ,  7.48 ,  8.23 ,  8.81 ,  6.03 ,  5.56 ,
        5.68 ,  5.14 ,  7.61 ,  6.17 ,  8.17 ,  9.87 ,  8.75 ,
        6.16 ,  9.5  ,  7.99 ,  5.67 ,  8.92 ,  6.19 ,  5.76 ,
        6.25 ,  5.11 ,  5.58 ,  5.65 ,  9.89 ,  8.03 ,  6.61 ,
        9.41 ,  8.64 ,  7.21 ,  8.28 ,  6.04 ,  9.13 ,  8.08 ,
        9.96 ,  5.12 ,  8.35 ,  7.07 ,  9.6  ,  9.24 ,  8.54 ,
        8.78 ,  8.93 ,  8.91 ,  9.04 ,  6.83 ,  5.85 ,  7.74 ,
        6.41 ,  8.9  ,  7.75 ,  7.88 ,  5.42 ,  7.52 ,  7.68 ,
        8.4  ,  9.39 ,  6.84 ,  5.99 ,  8.62 ,  8.53 ,  7.47 ,
        6.78 ,  6.42 ,  9.92 ,  8.39 ,  5.89 ,  7.22 ,  6.81 ,
        9.02 ,  9.97 ,  9.63 ,  9.67 ,  5.41 ,  7.27 ,  6.05 ,
        6.85 ,  9.33 ,  5.81 ,  6.53 ,  5.98 ,  6.02 ,  6.74 ,
        5.26 ,  7.72 ,  7.39 ,  8.43 ,  9.34 ,  5.44 ,  5.82 ,
        5.72 ,  8.19 ,  8.44 ,  8.98 ,  9.37 ,  5.8  ,  7.28 ,
        7.6  ,  7.91 ,  9.17 ,  7.46 ,  9.43 ,  9.91 ,  9.36 ,
        5.16 ,  7.08 ,  9.26 ,  8.83 , 10.   ,  7.8  ,  9.46 ,
        6.63 ,  7.24 ,  6.47 ,  7.77 ,  5.06 ,  7.17 ,  8.24 ,
        6.88 ,  9.03 ,  5.08 ,  5.45 ,  8.46 ,  9.19 ,  6.36 ,
        8.73 ,  7.11 ,  9.12 ,  9.4  ,  8.11 ,  9.98 ,  5.55 ,
        8.61 ,  8.14 ,  6.89 ,  9.84 ,  5.48 ,  8.21 ,  7.82 ,
        8.55 ,  5.79 ,  8.77 ,  8.29 ,  6.92 ,  7.37 ,  9.7  ,
        6.26 ,  7.26 ,  7.5  ,  6.82 ,  7.15 ,  5.77 ,  5.91 ,
        5.1  ,  7.71 ,  9.06 ,  5.71 ,  5.84 ,  9.42 ,  6.23 ,
```

```
       6.29  ,  5.25  ,  9.69  ,  9.9   ,  6.39  ,  8.09  ,  5.83  ,
       5.47  ,  6.56  ,  8.71  ,  9.94  ,  6.69  ,  5.52  ,  7.3   ,
       7.02  ,  6.33  ,  8.07  ,  8.37  ,  8.    ,  7.79  ,  8.65  ,
       6.28  ,  7.35  ,  8.69  ,  7.12  ,  7.32  ,  7.13  ,  5.97  ,
       5.09  ,  6.91  ,  6.76  ,  6.52  ,  7.45  ,  8.56  ,  6.5   ,
       8.63  ,  8.27  ,  8.49  ,  6.59  ,  9.29  ,  5.3   ,  7.06  ,
       5.38  ,  6.65  ,  9.16  ,  8.01  ,  8.25  ,  8.02  ,  8.47  ,
       7.34  ,  8.88  ,  7.14  ,  8.42  ,  5.17  ,  9.1   ,  7.49  ,
       9.85  ,  7.42  ,  9.31  ,  6.35  ,  7.    ,  5.39  ,  5.61  ,
       9.78  ,  9.25  ,  5.69  ,  9.47  ,  8.16  ,  7.23  ,  6.46  ,
       0.    ,  8.26  ,  6.32  ,  6.77  ,  8.85  ,  5.03  ,  7.65  ,
       5.78  ,  6.24  ,  5.35  ,  6.06  ,  7.78  ,  6.64  ,  7.0625,
       6.98  ,  6.44  ,  6.09  ])
```

[27]:
```
df['CGPA'].value_counts()
```

[27]:
```
CGPA
8.04    821
9.96    425
5.74    410
8.95    370
9.21    342
       ...
7.65      1
6.77      1
8.26      1
7.23      1
6.09      1
Name: count, Length: 332, dtype: int64
```

[28]:
```
df['Study Satisfaction'].nunique()
```

[28]:
```
6
```

[29]:
```
df['Study Satisfaction'].unique()
```

[29]:
```
array([3, 2, 5, 4, 1, 0])
```

[30]:
```
df['Study Satisfaction'].value_counts()
```

[30]:
```
Study Satisfaction
4    6356
2    5831
3    5820
1    5442
5    4419
0      10
Name: count, dtype: int64
```

[31]:
```
df['Job Satisfaction'].nunique()
```

[31]:
```
5
```

[32]:
```
df['Job Satisfaction'].unique()
```

[32]:
```
array([0, 3, 4, 2, 1])
```

[33]:
```
df['Job Satisfaction'].value_counts()
```

[33]:
```
Job Satisfaction
0    27870
2        3
4        2
1        2
3        1
Name: count, dtype: int64
```

[34]:
```
df['Sleep Duration'].nunique()
```

[34]:
```
6
```

[35]:
```
df['Sleep Duration'].unique()
```

[35]:
```
array(["'6-7 hours'", "'5-6 hours'", "'Less than 5 hours'", "'7-8 hours'",
       "'More than 8 hours'", 'Others'], dtype=object)
```

[36]:
```
#drops all the incorrect and unfixable values in 'Sleep Duration' column
df.drop( df[df['Sleep Duration'] == 'Others'].index, inplace=True)

df['Sleep Duration'].value_counts()
```

[36]:
```
Sleep Duration
'Less than 5 hours'    8304
'7-8 hours'            7337
'5-6 hours'            6177
'More than 8 hours'    6041
'6-7 hours'               1
Name: count, dtype: int64
```

[37]:
```
df['Dietary Habits'].nunique()
```

[37]:
```
4
```

[38]:
```
df['Dietary Habits'].unique()
```

[38]:
```
array(['Moderate', 'Healthy', 'Unhealthy', 'Others'], dtype=object)
```

[39]:
```
df['Dietary Habits'].value_counts()
```

[39]:
```
Dietary Habits
Unhealthy    10297
Moderate      9909
Healthy       7642
Others          12
Name: count, dtype: int64
```

[40]:
```
df['Degree'].nunique()
```

[40]:
```
28
```

[41]:
```
df['Degree'].unique()
```

[41]:
```
array(['B.Com', 'B.Pharm', 'BSc', 'BA', 'BCA', 'M.Tech', 'PhD',
       "'Class 12'", 'B.Ed', 'LLB', 'BE', 'M.Ed', 'MSc', 'BHM', 'M.Pharm',
       'MCA', 'MA', 'MD', 'MBA', 'MBBS', 'M.Com', 'B.Arch', 'LLM',
       'B.Tech', 'BBA', 'ME', 'MHM', 'Others'], dtype=object)
```

[42]:
```
df['Degree'].value_counts()
```

[42]:
```
Degree
'Class 12'    6075
B.Ed          1863
B.Com         1505
B.Arch        1476
BCA           1431
MSc           1187
B.Tech        1152
MCA           1043
M.Tech        1019
BHM            924
BSc            887
M.Ed           818
B.Pharm        810
M.Com          734
MBBS           696
BBA            696
LLB            670
BE             610
BA             595
M.Pharm        582
MD             571
MBA            560
MA             544
PhD            520
LLM            481
```

```
MHM            191
ME             185
Others          35
Name: count, dtype: int64
```

[43]:
```
df['Have you ever had suicidal thoughts ?'].nunique()
```

[43]:
```
2
```

[44]:
```
df['Have you ever had suicidal thoughts ?'].unique()
```

[44]:
```
array(['Yes', 'No'], dtype=object)
```

[45]:
```
df['Have you ever had suicidal thoughts ?'].value_counts()
```

[45]:
```
Have you ever had suicidal thoughts ?
Yes    17631
No     10229
Name: count, dtype: int64
```

[46]:
```
df['Work/Study Hours'].nunique()
```

[46]:
```
13
```

[47]:
```
df['Work/Study Hours'].unique()
```

[47]:
```
array([ 8,  3,  9,  4,  1,  0, 12,  2, 11, 10,  6,  5,  7])
```

[48]:
```
df['Work/Study Hours'].value_counts()
```

[48]:
```
Work/Study Hours
10    4227
12    3166
11    2889
8     2506
6     2243
9     2025
7     1999
0     1698
4     1612
2     1585
3     1467
5     1296
1     1147
Name: count, dtype: int64
```

[49]:
```
df['Financial Stress'].nunique()
```

[49]:
```
6
```

[50]:
```
df['Financial Stress'].unique()
```

[50]:
```
array(['4', '1', '2', '5', '3', '?'], dtype=object)
```

[51]:
```
df['Financial Stress'].value_counts()
```

[51]:
```
Financial Stress
5    6705
4    5773
3    5219
1    5110
2    5050
?       3
Name: count, dtype: int64
```

[52]:
```
#drops all the incorrect and unfixable values in 'Financial Stress' column
df.drop(df[df['Financial Stress'] == '?'].index, inplace = True)

#changes the data type of 'Financial Stress' column from 'object' to 'int'
df['Financial Stress'] = df['Financial Stress'].astype(int)
```

[53]:
```
df.dtypes['Financial Stress']
```

[53]:
```
dtype('int64')
```

[54]:
```
df['Family History of Mental Illness'].nunique()
```

[54]:
```
2
```

[55]:
```
df['Family History of Mental Illness'].unique()
```

[55]:
```
array(['No', 'Yes'], dtype=object)
```

[56]:
```
df['Family History of Mental Illness'].value_counts()
```

[56]:
```
Family History of Mental Illness
No     14374
Yes    13483
Name: count, dtype: int64
```

[57]:
```
df['Depression'].nunique()
```

[57]:
```
2
```

[58]:
```
df['Depression'].unique()
```

[58]:
```
array([1, 0])
```

[59]:
```
df['Depression'].value_counts()
```

[59]:
```
Depression
1    16313
0    11544
Name: count, dtype: int64
```

[60]:
```
#changes the data type of 'Depression' column from 'int to 'bool'
df['Depression'] = df['Depression'].astype(bool)
```

[61]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 27857 entries, 0 to 27901
Data columns (total 18 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   id                                     27857 non-null  int64
 1   Gender                                 27857 non-null  object
 2   Age                                    27857 non-null  int64
 3   City                                   27857 non-null  object
 4   Profession                             27857 non-null  object
 5   Academic Pressure                      27857 non-null  int64
 6   Work Pressure                          27857 non-null  int64
 7   CGPA                                   27857 non-null  float64
 8   Study Satisfaction                     27857 non-null  int64
 9   Job Satisfaction                       27857 non-null  int64
 10  Sleep Duration                         27857 non-null  object
 11  Dietary Habits                         27857 non-null  object
 12  Degree                                 27857 non-null  object
 13  Have you ever had suicidal thoughts ?  27857 non-null  object
 14  Work/Study Hours                       27857 non-null  int64
 15  Financial Stress                       27857 non-null  int64
 16  Family History of Mental Illness       27857 non-null  object
 17  Depression                             27857 non-null  bool
dtypes: bool(1), float64(1), int64(8), object(8)
memory usage: 3.9+ MB
```

[62]:
```
df.head()
```

[62]:

Serial Number	id	Gender	Age	City	Profession	Academic Pressure	\
0	1	Male	19	Delhi	Student	4	
1	2	Male	33	Visakhapatnam	Student	5	
2	8	Female	24	Bangalore	Student	2	
3	26	Male	31	Srinagar	Student	3	

```
4              30  Female   28       Varanasi    Student                  3

               Work Pressure  CGPA  Study Satisfaction  Job Satisfaction  \
Serial Number
0                          0  6.00                   3                 0
1                          0  8.97                   2                 0
2                          0  5.90                   5                 0
3                          0  7.03                   5                 0
4                          0  5.59                   2                 0

                    Sleep Duration Dietary Habits   Degree  \
Serial Number
0                       '6-7 hours'       Moderate    B.Com
1                       '5-6 hours'        Healthy  B.Pharm
2                       '5-6 hours'       Moderate      BSc
3                 'Less than 5 hours'        Healthy       BA
4                       '7-8 hours'       Moderate      BCA

              Have you ever had suicidal thoughts ?  Work/Study Hours  \
Serial Number
0                                               Yes                8
1                                               Yes                3
2                                                No                3
3                                                No                9
4                                               Yes                4

               Financial Stress Family History of Mental Illness  Depression
Serial Number
0                             4                              No         True
1                             1                              No         True
2                             2                             Yes        False
3                             1                             Yes        False
4                             5                             Yes         True
```

[ ]:
```
#exports the cleaned database as a csv file with headers and indexes
df.to_csv('cleaned_student_depression_dataset.csv')
```