Name: Bishakha
SJSU ID: 019178354

Project Proposal

Fa25: ISE-201 Math Foundation for Decision and Data Sciences

List of Databases selected

Out of curiosity to understand current sociological conditions , I chose databases which help me understand new patterns in our society that shapes how societies now function.Below are the three databases :

1. Student Success Factors and Insights (Source: Kaggle)
2. Impact of Screen Time on Mental Health (Source: Kaggle)
3. Student Stress Factors (Source : kaggle)

Student Success Factors and Insights

This database explores factors contributing to student success. The dataset has 20 attributes and 6607 rows. It explores several aspects like parental education level, parental involvement , motivation level, Family income , extracurricular activities , Access to resources, physical activity , sleep hours , hours studied and many more.
Most of the attributes are of string type, belonging to different categories. The data looks authentic , because most columns hold some string type categorical values.
Therefore, a lot of encoding will be required to be able to use all the columns for analysis as well as modelling.

Impact of Screen Time on Mental Health

Most of the time while awake , we spend more time on screen than with people around. Be it phone or watch or TV or laptop, we are constantly looking at some screen. In recent years , particularly after COVID , mental health conditions have deteriorated. Mental health crises can be felt and seen all around globally.
This dataset helps understand the impact of screen time on mental health. It has 25 columns and 2000 rows. It explores several factors like phone , tablet , tv usage hours, work related screen time, eating habits , location type , gender , social media hours, caffeine intake, anxiety and depression scores etc.

However , the dataset does not look realistic. It is not clearly shown how the anxiety and depression score is calculated and scaled between 0-20. Also it does not reflect on how the mental health score has been calculated. Also the total screen time columns is not equal to the sum of all the screen time columns for a user_id. So, the dataset does not seem authentic.

Student Stress Factors

With people interacting more with technology and less with peers, stress levels have increased in society. This dataset explores several factors causing stress in students. For example , factors like living conditions, study load, social support , peer pressure, basic needs, bullying , anxiety level , depression , mental health history etc.

However , it is not mentioned how anxiety level , depression , self esteem etc has been calculated . Most of the columns have already been encoded and scaled on varied ranges. The dataset has not explained the range used for different columns and also does not explain encoding values used. So the conclusions will be hard to explain with this dataset .

Student Success Factors and Insights, Why ?

After initial look at the dataset and performing descriptive statistics , I chose 'Student success Factors and Insight Datasets.

I chose this dataset because it feels more authentic, realistic as well as sensible. Most columns hold some string value and need to be encoded before they can be used for EDA.

Initial Set of Question

1. How many attributes are in the dataset ?
2. How many students are in the dataset ?
3. What is average attendance ?
4. How many average hours of physical activities do students do every day ?
5. What is the average sleep hours for students ?
6. How many students go to public school and how many go to private school ?
7. How many average hours of tutoring sessions do students take ?
8. How many students have high parental involvement ?
9. What is the average , max and min previous scores for students ?
10. How many parents have a postgraduate education ?

To answer these questions , I performed the initial descriptive statistical analysis .we will see the answer in the next section

Initial Descriptive Statistics and Conclusions

To perform statistical analysis , I used python and its libraries like pandas , numpy , matplotlib etc , and derived the following conclusions, some of which are answers to the *initial set of questions*.

1. There are a total of 20 attributes in the dataset.
2. A total of 6607 rows of student data are in the dataset.
3. The average attendance is around 80%.
4. An average 2.9 hours of physical activities do students do everyday.
5. An average 7 hours of sleep students get everyday.
6. Based on charts plotted for categorical data , almost 4600 students go to public school , whereas almost 2000 goes to private school.
7. Each day an average of 1.5 hours of tutoring sessions students take.
8. Almost 1900 students reported high parental involvement.
9. Average previous score is 75 , while max is 100 and min is 50.
10. Almost 1300 parents have postgraduate education.

A few more points about the dataset .

1. A lot of encoding needs to be done before using the dataset for training.
2. A few rows have blank/empty columns for a few attributes , like teacher_quality , parental_education_level. That needs to be fixed too.
3. The distribution of data for categorical columns looks normal, realistic and convincing as it resembles very much what I observe in society around me.

Appendix

October 21, 2025

0.1 Python code

FA25: ISE-201 Sec 33 - Math Dec. and Data Science

Project Proposal Python code

Datasets :
Source : Kaggle.com

1.Student Success: Factors & Insights

https://www.kaggle.com/datasets/anassarfraz13/student-success-factors-and-insights

2.Impact of Screen Time on Mental Health

https://www.kaggle.com/datasets/khushikyad001/impact-of-screen-time-on-mental-health/data

3.Student Stress Factors: A Comprehensive Analysis

https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis

[1]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

[2]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

1 *Exploring Student Success Dataset*

[3]:
```
dataset1 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/
↪Fall25_ISE_Mathematics/Project_Datasets/StudentPerformanceFactors.csv')
```

[4]:
```
dataset1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6607 entries, 0 to 6606
Data columns (total 20 columns):
 #   Column                      Non-Null Count  Dtype
```

```
---  ------                       --------------  -----
 0   Hours_Studied                6607 non-null   int64
 1   Attendance                   6607 non-null   int64
 2   Parental_Involvement         6607 non-null   object
 3   Access_to_Resources          6607 non-null   object
 4   Extracurricular_Activities   6607 non-null   object
 5   Sleep_Hours                  6607 non-null   int64
 6   Previous_Scores              6607 non-null   int64
 7   Motivation_Level             6607 non-null   object
 8   Internet_Access              6607 non-null   object
 9   Tutoring_Sessions            6607 non-null   int64
 10  Family_Income                6607 non-null   object
 11  Teacher_Quality              6529 non-null   object
 12  School_Type                  6607 non-null   object
 13  Peer_Influence               6607 non-null   object
 14  Physical_Activity            6607 non-null   int64
 15  Learning_Disabilities        6607 non-null   object
 16  Parental_Education_Level     6517 non-null   object
 17  Distance_from_Home           6540 non-null   object
 18  Gender                       6607 non-null   object
 19  Exam_Score                   6607 non-null   int64
dtypes: int64(7), object(13)
memory usage: 1.0+ MB
```

[5]:
```
dataset1.describe()
```

[5]:

	Hours_Studied	Attendance	Sleep_Hours	Previous_Scores	\
count	6607.000000	6607.000000	6607.00000	6607.000000	
mean	19.975329	79.977448	7.02906	75.070531	
std	5.990594	11.547475	1.46812	14.399784	
min	1.000000	60.000000	4.00000	50.000000	
25%	16.000000	70.000000	6.00000	63.000000	
50%	20.000000	80.000000	7.00000	75.000000	
75%	24.000000	90.000000	8.00000	88.000000	
max	44.000000	100.000000	10.00000	100.000000	

	Tutoring_Sessions	Physical_Activity	Exam_Score
count	6607.000000	6607.000000	6607.000000
mean	1.493719	2.967610	67.235659
std	1.230570	1.031231	3.890456
min	0.000000	0.000000	55.000000
25%	1.000000	2.000000	65.000000
50%	1.000000	3.000000	67.000000
75%	2.000000	4.000000	69.000000
max	8.000000	6.000000	101.000000

[6]:
```
dataset1.shape
```

[6]: (6607, 20)

[7]:
```
dataset1.head(10)
```

[7]:

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources
0	23	84	Low	High
1	19	64	Low	Medium
2	24	98	Medium	Medium
3	29	89	Low	Medium
4	19	92	Medium	Medium
5	19	88	Medium	Medium
6	29	84	Medium	Low
7	25	78	Low	High
8	17	94	Medium	High
9	23	98	Medium	Medium

	Extracurricular_Activities	Sleep_Hours	Previous_Scores	Motivation_Level
0	No	7	73	Low
1	No	8	59	Low
2	Yes	7	91	Medium
3	Yes	8	98	Medium
4	Yes	6	65	Medium
5	Yes	8	89	Medium
6	Yes	7	68	Low
7	Yes	6	50	Medium
8	No	6	80	High
9	Yes	8	71	Medium

	Internet_Access	Tutoring_Sessions	Family_Income	Teacher_Quality
0	Yes	0	Low	Medium
1	Yes	2	Medium	Medium
2	Yes	2	Medium	Medium
3	Yes	1	Medium	Medium
4	Yes	3	Medium	High
5	Yes	3	Medium	Medium
6	Yes	1	Low	Medium
7	Yes	1	High	High
8	Yes	0	Medium	Low
9	Yes	0	High	High

	School_Type	Peer_Influence	Physical_Activity	Learning_Disabilities
0	Public	Positive	3	No
1	Public	Negative	4	No
2	Public	Neutral	4	No
3	Public	Negative	4	No
4	Public	Neutral	4	No
5	Public	Positive	3	No

```
6     Private        Neutral                  2                    No
7      Public       Negative                  2                    No
8     Private        Neutral                  1                    No
9      Public       Positive                  5                    No

  Parental_Education_Level Distance_from_Home  Gender  Exam_Score
0              High School               Near    Male          67
1                  College           Moderate  Female          61
2              Postgraduate               Near    Male          74
3              High School           Moderate    Male          71
4                  College               Near  Female          70
5              Postgraduate               Near    Male          71
6              High School           Moderate    Male          67
7              High School                Far    Male          66
8                  College               Near    Male          69
9              High School           Moderate    Male          72
```

[8]:
```
dataset1.isnull().sum()
```

[8]:
```
Hours_Studied                 0
Attendance                    0
Parental_Involvement          0
Access_to_Resources           0
Extracurricular_Activities    0
Sleep_Hours                   0
Previous_Scores               0
Motivation_Level              0
Internet_Access               0
Tutoring_Sessions             0
Family_Income                 0
Teacher_Quality              78
School_Type                   0
Peer_Influence                0
Physical_Activity             0
Learning_Disabilities         0
Parental_Education_Level     90
Distance_from_Home           67
Gender                        0
Exam_Score                    0
dtype: int64
```

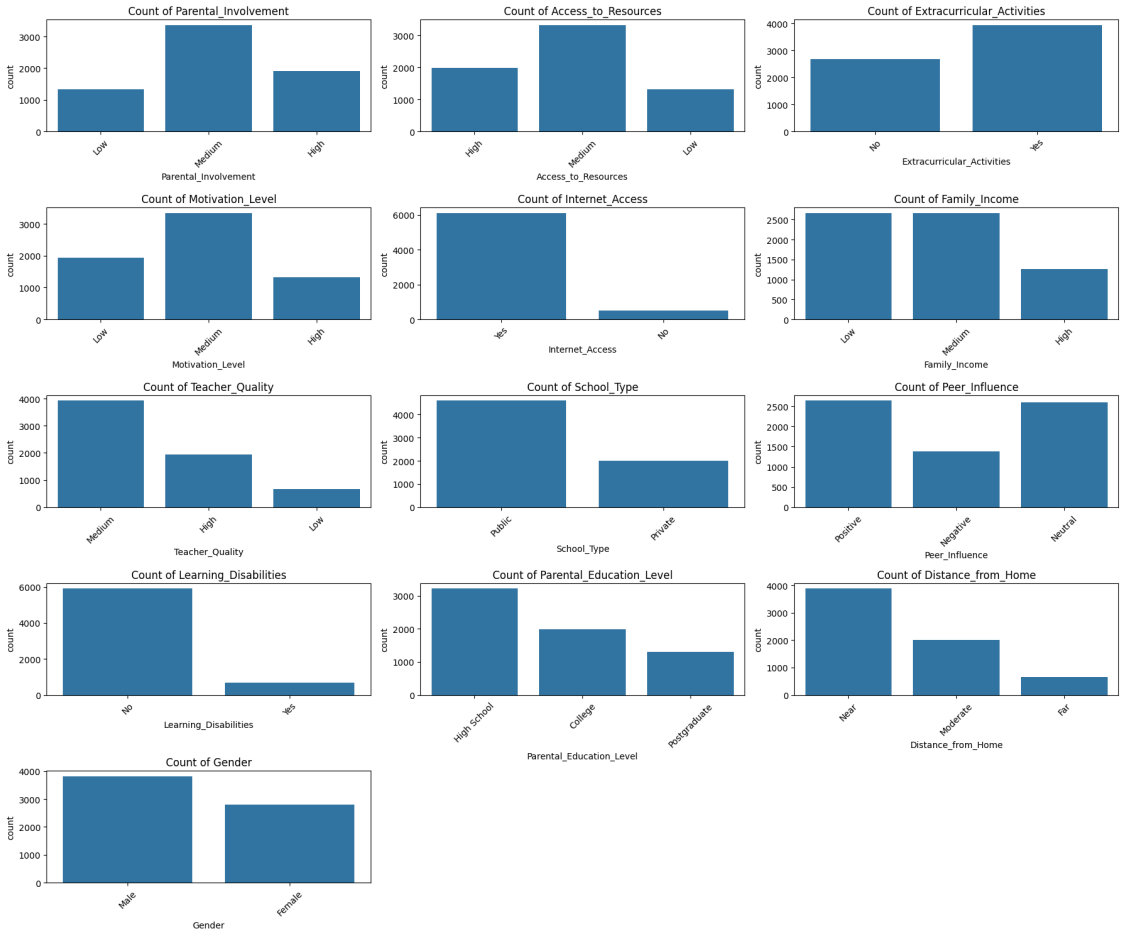
[10]:
```
#### Distribution chart for categorical data
cat_cols =␣
 ↪["Parental_Involvement","Access_to_Resources","Extracurricular_Activities",
          ␣
 ↪"Motivation_Level","Internet_Access","Family_Income","Teacher_Quality",
           "School_Type","Peer_Influence","Learning_Disabilities",
```

```
            "Parental_Education_Level","Distance_from_Home","Gender"]

plt.figure(figsize=(18,15))

for i, col in enumerate(cat_cols, 1):
    plt.subplot(5, 3, i)   # 5 rows, 3 columns grid (total 15 slots for 13␣
 ↪plots)
    sns.countplot(x=col, data=dataset1)
    plt.title(f"Count of {col}")
    plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```




```
[ ]:
```

2 *Exploring Mental heath dataset*

[13]:
```
dataset2 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/
↪Fall25_ISE_Mathematics/Project_Datasets/digital_diet_mental_health.csv')
```

[37]:
```
dataset2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 25 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   user_id                           2000 non-null   object
 1   age                               2000 non-null   int64
 2   gender                            2000 non-null   object
 3   daily_screen_time_hours           2000 non-null   float64
 4   phone_usage_hours                 2000 non-null   float64
 5   laptop_usage_hours                2000 non-null   float64
 6   tablet_usage_hours                2000 non-null   float64
 7   tv_usage_hours                    2000 non-null   float64
 8   social_media_hours                2000 non-null   float64
 9   work_related_hours                2000 non-null   float64
 10  entertainment_hours               2000 non-null   float64
 11  gaming_hours                      2000 non-null   float64
 12  sleep_duration_hours              2000 non-null   float64
 13  sleep_quality                     2000 non-null   int64
 14  mood_rating                       2000 non-null   int64
 15  stress_level                      2000 non-null   int64
 16  physical_activity_hours_per_week  2000 non-null   float64
 17  location_type                     2000 non-null   object
 18  mental_health_score               2000 non-null   int64
 19  uses_wellness_apps                2000 non-null   int64
 20  eats_healthy                      2000 non-null   int64
 21  caffeine_intake_mg_per_day        2000 non-null   float64
 22  weekly_anxiety_score              2000 non-null   int64
 23  weekly_depression_score           2000 non-null   int64
 24  mindfulness_minutes_per_day       2000 non-null   float64
dtypes: float64(13), int64(9), object(3)
memory usage: 390.8+ KB
```

[38]:
```
dataset2.describe()
```

[38]:

	age	daily_screen_time_hours	phone_usage_hours	\
count	2000.000000	2000.000000	2000.000000	
mean	38.805500	6.025600	3.023700	
std	14.929203	1.974123	1.449399	
min	13.000000	0.000000	0.000000	

```
25%      26.000000                4.700000           2.000000
50%      39.000000                6.000000           3.000000
75%      51.000000                7.325000           4.000000
max      64.000000               13.300000           8.400000

       laptop_usage_hours  tablet_usage_hours  tv_usage_hours  \
count         2000.000000         2000.000000     2000.000000
mean             1.999950            0.995650        1.503700
std              0.997949            0.492714        0.959003
min              0.000000            0.000000        0.000000
25%              1.300000            0.600000        0.800000
50%              2.000000            1.000000        1.500000
75%              2.700000            1.300000        2.200000
max              5.600000            2.500000        4.700000

       social_media_hours  work_related_hours  entertainment_hours  \
count         2000.000000         2000.000000           2000.00000
mean             2.039200            2.010250              2.46735
std              1.133435            1.116111              1.23686
min              0.000000            0.000000              0.00000
25%              1.200000            1.200000              1.60000
50%              2.000000            2.000000              2.40000
75%              2.800000            2.800000              3.30000
max              5.800000            5.900000              6.80000

       gaming_hours  …  mood_rating  stress_level  \
count     2000.0000  …  2000.000000   2000.000000
mean         1.2795  …     5.591000      5.541500
std          0.8945  …     2.899814      2.885731
min          0.0000  …     1.000000      1.000000
25%          0.6000  …     3.000000      3.000000
50%          1.2000  …     6.000000      6.000000
75%          1.9000  …     8.000000      8.000000
max          4.0000  …    10.000000     10.000000

       physical_activity_hours_per_week  mental_health_score  \
count                       2000.000000          2000.000000
mean                           3.087150            49.650500
std                            1.885258            17.546717
min                            0.000000            20.000000
25%                            1.600000            35.000000
50%                            3.000000            49.000000
75%                            4.400000            64.250000
max                            9.700000            80.000000

       uses_wellness_apps  eats_healthy  caffeine_intake_mg_per_day  \
count         2000.000000   2000.000000                 2000.00000
```

```
mean              0.387500      0.507500                  148.07970
std               0.487301      0.500069                   48.86066
min               0.000000      0.000000                    0.80000
25%               0.000000      0.000000                  113.90000
50%               0.000000      1.000000                  147.45000
75%               1.000000      1.000000                  180.70000
max               1.000000      1.000000                  364.90000

       weekly_anxiety_score  weekly_depression_score  \
count           2000.000000                2000.00000
mean               9.887500                  10.04900
std                6.027853                   6.05334
min                0.000000                   0.00000
25%                5.000000                   5.00000
50%               10.000000                  10.00000
75%               15.000000                  15.00000
max               20.000000                  20.00000

       mindfulness_minutes_per_day
count                  2000.000000
mean                     10.753750
std                       7.340269
min                       0.000000
25%                       4.900000
50%                      10.400000
75%                      15.800000
max                      36.400000

[8 rows x 22 columns]
```

[26]:
```
dataset2.shape
```

[26]:
```
(2000, 25)
```

[29]:
```
dataset2.head(10)
```

[29]:
```
    user_id  age  gender  daily_screen_time_hours  phone_usage_hours  \
0    user_1   51  Female                      4.8                3.4
1    user_2   64    Male                      3.9                3.5
2    user_3   41   Other                     10.5                2.1
3    user_4   27   Other                      8.8                0.0
4    user_5   55    Male                      5.9                1.7
5    user_6   20  Female                      9.9                3.2
6    user_7   33    Male                      5.8                4.0
7    user_8   51  Female                      7.4                2.9
8    user_9   31  Female                      6.0                2.3
9   user_10   35    Male                      6.8                6.4
```

	laptop_usage_hours	tablet_usage_hours	tv_usage_hours	social_media_hours	\
0	1.3	1.6	1.6	4.1	
1	1.8	0.9	2.0	2.7	
2	2.6	0.7	2.2	3.0	
3	0.0	0.7	2.5	3.3	
4	1.1	1.5	1.6	1.1	
5	2.7	1.2	3.3	1.6	
6	3.2	1.9	1.3	0.9	
7	3.0	0.9	3.4	1.2	
8	2.9	1.4	0.4	1.5	
9	2.1	1.4	0.5	1.2	

	work_related_hours	…	stress_level	physical_activity_hours_per_week	\
0	2.0	…	10	0.7	
1	3.1	…	6	4.3	
2	2.8	…	5	3.1	
3	1.6	…	5	0.0	
4	3.6	…	7	3.0	
5	2.5	…	2	3.3	
6	1.7	…	9	6.8	
7	2.2	…	2	3.5	
8	2.1	…	9	0.0	
9	2.2	…	4	3.5	

	location_type	mental_health_score	uses_wellness_apps	eats_healthy	\
0	Urban	32	1	1	
1	Suburban	75	0	1	
2	Suburban	22	0	0	
3	Rural	22	0	1	
4	Urban	64	1	1	
5	Suburban	72	1	0	
6	Urban	45	0	0	
7	Urban	23	0	1	
8	Suburban	35	1	0	
9	Urban	47	1	0	

	caffeine_intake_mg_per_day	weekly_anxiety_score	weekly_depression_score	\
0	125.2	13	15	
1	150.4	19	18	
2	187.9	7	3	
3	73.6	7	2	
4	217.5	8	10	
5	102.8	17	16	
6	162.8	6	4	
7	147.1	9	8	
8	109.7	0	0	

```
9                      206.8                    3                       8

   mindfulness_minutes_per_day
0                          4.0
1                          6.5
2                          6.9
3                          4.8
4                          0.0
5                         11.5
6                          7.9
7                          4.4
8                         21.2
9                          8.1

[10 rows x 25 columns]
```

[15]:
```
dataset2.isnull().sum()
```

[15]:
```
user_id                          0
age                              0
gender                           0
daily_screen_time_hours          0
phone_usage_hours                0
laptop_usage_hours               0
tablet_usage_hours               0
tv_usage_hours                   0
social_media_hours               0
work_related_hours               0
entertainment_hours              0
gaming_hours                     0
sleep_duration_hours             0
sleep_quality                    0
mood_rating                      0
stress_level                     0
physical_activity_hours_per_week 0
location_type                    0
mental_health_score              0
uses_wellness_apps               0
eats_healthy                     0
caffeine_intake_mg_per_day       0
weekly_anxiety_score             0
weekly_depression_score          0
mindfulness_minutes_per_day      0
dtype: int64
```

3 *Exporing Student stress level dataset*

[16]:
```
dataset3 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/
↪Fall25_ISE_Mathematics/Project_Datasets/StressLevelDataset.csv')
```

[39]:
```
dataset3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1100 entries, 0 to 1099
Data columns (total 21 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   anxiety_level                 1100 non-null   int64
 1   self_esteem                   1100 non-null   int64
 2   mental_health_history         1100 non-null   int64
 3   depression                    1100 non-null   int64
 4   headache                      1100 non-null   int64
 5   blood_pressure                1100 non-null   int64
 6   sleep_quality                 1100 non-null   int64
 7   breathing_problem             1100 non-null   int64
 8   noise_level                   1100 non-null   int64
 9   living_conditions             1100 non-null   int64
 10  safety                        1100 non-null   int64
 11  basic_needs                   1100 non-null   int64
 12  academic_performance          1100 non-null   int64
 13  study_load                    1100 non-null   int64
 14  teacher_student_relationship  1100 non-null   int64
 15  future_career_concerns        1100 non-null   int64
 16  social_support                1100 non-null   int64
 17  peer_pressure                 1100 non-null   int64
 18  extracurricular_activities    1100 non-null   int64
 19  bullying                      1100 non-null   int64
 20  stress_level                  1100 non-null   int64
dtypes: int64(21)
memory usage: 180.6 KB
```

[40]:
```
dataset3.describe()
```

[40]:

	anxiety_level	self_esteem	mental_health_history	depression	\
count	1100.000000	1100.000000	1100.000000	1100.000000	
mean	11.063636	17.777273	0.492727	12.555455	
std	6.117558	8.944599	0.500175	7.727008	
min	0.000000	0.000000	0.000000	0.000000	
25%	6.000000	11.000000	0.000000	6.000000	
50%	11.000000	19.000000	0.000000	12.000000	
75%	16.000000	26.000000	1.000000	19.000000	
max	21.000000	30.000000	1.000000	27.000000	

```
          headache  blood_pressure  sleep_quality  breathing_problem  \
count  1100.000000     1100.000000    1100.000000        1100.000000
mean      2.508182        2.181818       2.660000           2.753636
std       1.409356        0.833575       1.548383           1.400713
min       0.000000        1.000000       0.000000           0.000000
25%       1.000000        1.000000       1.000000           2.000000
50%       3.000000        2.000000       2.500000           3.000000
75%       3.000000        3.000000       4.000000           4.000000
max       5.000000        3.000000       5.000000           5.000000

       noise_level  living_conditions  …  basic_needs  academic_performance  \
count  1100.000000        1100.000000  …  1100.000000           1100.000000
mean      2.649091           2.518182  …     2.772727              2.772727
std       1.328127           1.119208  …     1.433761              1.414594
min       0.000000           0.000000  …     0.000000              0.000000
25%       2.000000           2.000000  …     2.000000              2.000000
50%       3.000000           2.000000  …     3.000000              2.000000
75%       3.000000           3.000000  …     4.000000              4.000000
max       5.000000           5.000000  …     5.000000              5.000000

        study_load  teacher_student_relationship  future_career_concerns  \
count  1100.000000                   1100.000000             1100.000000
mean      2.621818                      2.648182                2.649091
std       1.315781                      1.384579                1.529375
min       0.000000                      0.000000                0.000000
25%       2.000000                      2.000000                1.000000
50%       2.000000                      2.000000                2.000000
75%       3.000000                      4.000000                4.000000
max       5.000000                      5.000000                5.000000

       social_support  peer_pressure  extracurricular_activities     bullying  \
count     1100.000000    1100.000000                 1100.000000  1100.000000
mean         1.881818       2.734545                    2.767273     2.617273
std          1.047826       1.425265                    1.417562     1.530958
min          0.000000       0.000000                    0.000000     0.000000
25%          1.000000       2.000000                    2.000000     1.000000
50%          2.000000       2.000000                    2.500000     3.000000
75%          3.000000       4.000000                    4.000000     4.000000
max          3.000000       5.000000                    5.000000     5.000000

       stress_level
count   1100.000000
mean       0.996364
std        0.821673
min        0.000000
25%        0.000000
```

```
50%         1.000000
75%         2.000000
max         2.000000

[8 rows x 21 columns]
```

[30]:
```
dataset3.shape
```

[30]:
```
(1100, 21)
```

[31]:
```
dataset3.head(10)
```

[31]:

	anxiety_level	self_esteem	mental_health_history	depression	headache	\
0	14	20	0	11	2	
1	15	8	1	15	5	
2	12	18	1	14	2	
3	16	12	1	15	4	
4	16	28	0	7	2	
5	20	13	1	21	3	
6	4	26	0	6	1	
7	17	3	1	22	4	
8	13	22	1	12	3	
9	6	8	0	27	4	

	blood_pressure	sleep_quality	breathing_problem	noise_level	\
0	1	2	4	2	
1	3	1	4	3	
2	1	2	2	2	
3	3	1	3	4	
4	3	5	1	3	
5	3	1	4	3	
6	2	4	1	1	
7	3	1	5	3	
8	1	2	4	3	
9	3	1	2	0	

	living_conditions	…	basic_needs	academic_performance	study_load	\
0	3	…	2	3	2	
1	1	…	2	1	4	
2	2	…	2	2	3	
3	2	…	2	2	4	
4	2	…	3	4	3	
5	2	…	1	2	5	
6	4	…	4	5	1	
7	1	…	1	1	3	
8	3	…	3	3	3	
9	5	…	2	2	2	

```
   teacher_student_relationship  future_career_concerns  social_support  \
0                             3                       3               2
1                             1                       5               1
2                             3                       2               2
3                             1                       4               1
4                             1                       2               1
5                             2                       5               1
6                             4                       1               3
7                             2                       4               1
8                             2                       3               3
9                             1                       5               1

   peer_pressure  extracurricular_activities  bullying  stress_level
0              3                           3         2             1
1              4                           5         5             2
2              3                           2         2             1
3              4                           4         5             2
4              5                           0         5             1
5              4                           4         5             2
6              2                           2         1             0
7              4                           4         5             2
8              3                           2         2             1
9              5                           3         4             1

[10 rows x 21 columns]
```

[18]:
```
dataset3.isnull().sum()
```

[18]:
```
anxiety_level                 0
self_esteem                   0
mental_health_history         0
depression                    0
headache                      0
blood_pressure                0
sleep_quality                 0
breathing_problem             0
noise_level                   0
living_conditions             0
safety                        0
basic_needs                   0
academic_performance          0
study_load                    0
teacher_student_relationship  0
future_career_concerns        0
social_support                0
peer_pressure                 0
```

```
extracurricular_activities      0
bullying                        0
stress_level                    0
dtype: int64
```