STUDENT MENTAL HEALTH ASSESMENT

Submitted by

RUTHRESHWARAN.M

1P22CS024

In partial fulfilment of the requirements for the award of the Degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from Bharathiar University, Coimbatore.

Under the Internal Supervision of

Dr. Suganya S. M.C.A., Ph.D.

Associate Professor

Estd 1986

RVS CAS

Building Intellectual Capital

SCHOOL OF COMPUTER STUDIES (PG)

RVS COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Sulur, Coimbatore – 641 402.

March 2024.

RVS COLLEGE OF ARTS AND SCIENCE
(AUTONOMOUS)

Sulur, Coimbatore – 641 402.

School of Computer Studies (PG)

Register Number: 1P22CS024

Certified Bonafide Project Work done by RUTHRESHWARAN M

Submitted for the Project Evaluation and Viva - Voce held at the School of Computer Studies (PG), RVS College of Arts and Science, Sulur, Coimbatore on

_____

Supervisor **HOD**

Internal Examiner **External Examiner**

Certificate

Certificate

This is to certify that the project work entitled **STUDENT MENTAL HEALTH ASSESMENT**, submitted to the School of Computer Studies(PG), RVS College of Arts and Science impartial fulfillment of the requirements for the award of the Degree of Master of Science in Computer Science is a record of original project work done by **RUTHRESHWARAN M** during the period December 2023- March 2024 of her study in the **Master of Science in Computer Science, RVS College of Arts and Science**, under my internal supervision and the project work has not formed thebasis for the award of any Degree/Diploma/Associateship/Fellowship or other similar title to any candidate of any University.

Internal Supervisor

Declaration

Declaration

I, **RUTHRESHWARAN M**, hereby declare that the project entitled **STUDENT MENTAL HEALTH ASSESMENT,** submitted to the School of Computer Studies (PG), RVS College of Arts and Science, in partial fulfillment of the requirements for the award of the Degree of Master of Science in Computer Science is a record of original project work done by me during the period December 2023 to March 2024 under the internal supervision of **Dr. Suganya S. M.C.A., Ph.D., Associate Professor, RVS College Of Arts and Science (Autonomous)** From Bharathiar University, Coimbatore.

Signature of the Candidate

Acknowledgement

Acknowledgement

I express my sincere thanks to our Managing Trustee **Dr. K. Senthil Ganesh MBA (USA)., MS (UK)., Ph.D.,** for providing us with adequate faculty and laboratory resources for completing my project successfully.

I take this as a fine opportunity to express my sincere thanks to **Dr. T. Sivakumar M.Sc., M. Phil., Ph.D., Principal,** RVS College of Arts and Science (Autonomous) for giving me the opportunity to undertake this project.

I express my sincere thanks to **Dr. P. Navaneetham M.Sc., M.Phil., Ph.D., Director (Administration), School of Computer Studies** for the help and advice throughout the project.

I express my sincere thanks to **Dr. D. Maheswari, M.Sc.CS., M.Phil., Ph.D., Head and Research Coordinator, School of Computer Studies(PG)** for her support and advice throughout the project.

I express my gratitude to **Dr. Suganya S. M.C.A., Ph.D., Associate Professor, School of Computer Studies (PG)** for his valuable guidance, support, encouragement, and motivation rendered by her throughout this project.

Finally, I express my sincere thanks to all other staff members and my dear friends, dear and near forhelping me to complete this project.

RUTHRESHWARAN M

STUDENT MENTAL HEALTH ASSESMENT

Abstract

Abstract

In recent years, there has been a growing recognition of the importance of mental health in the context of higher education. The transition to university life, academic pressures, social dynamics, and personal challenges all contribute to the complex landscape of student well-being. This abstract presents an overview of a comprehensive analysis aimed at understanding the various facets of student mental health.Drawing upon a rich dataset encompassing diverse variables such as demographics, academic performance indicators, psychological measures, lifestyle factors, and support networks, this analysis seeks to uncover patterns and trends in student mental health. By employing advanced statistical techniques and data mining approaches,

we aim to identify risk factors associated with poor mental health outcomes, as well as protective factors that promote resilience and well-being.Furthermore, this analysis delves into the intersectionality of mental health with other aspects of student life, including academic success, social relationships, and lifestyle choices. By examining the interplay between these factors, we hope to gain a deeper understanding of the nuanced challenges faced by students and the potential pathways to effective intervention.

By analyzing data collected through surveys and interviews, we endeavor to discern patterns and correlations between these variables and their impact on students' mental health and academic performance. This research not only aims to identify risk factors but also to highlight protective factors and coping strategies that contribute to resilience and success in university settings.

CONTENT

TABLE OF CONTENTS

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER I - BUSINESS UNDERSTANDING

1.1 INTRODUCTION

The dataset represents mental health evaluations of students. This dataset seeks to provide valuable insights into the mental health of students by capturing a number of factors that may impact their mental health. Mental health significantly impacts a student's ability to learn, cope with stress, and engage in academic and social activities. Positive mental health contributes to better academic performance, increased motivation, and improved interpersonal relationships.

With today's fierce competition and increasing pressures in life, college students' mental health problems have become more visible, and their mental health conditions are concerning. People with severe mental disorders or mental illnesses are forced to suspend school, drop out of school self-harm commit suicide, and even break the law in an endless stream among college students. It is critical and urgent to improve college students' overall quality, particularly their psychological quality cultivate exceptional social talents, improve mental health education, and predict mental health.

College students are outstanding members of the youth population, representing a high intellectual group, and their mental health is critical. College students are in a critical transition period in their development and maturity. They will face a variety of issues during this time, including emotions and socialization If they are not handled properly, they can lead to depression, anxiety, and other psychological issues. This is extremely harmful to college students' development. It is not uncommon to come across examples of exceptional college students who failed to deal with the final suicide due to emotional issues.

1.2 OBJECTIVE

In analysing student mental health assessment data, researchers often employ a multifaceted approach to gain comprehensive insights into various facets of students' well-being. One approach involves quantitative analysis, where statistical methods are applied to numerical data gathered from standardized mental health assessments. Researchers may use tools like regression analysis or machine learning algorithms to identify patterns,

correlations, and predictive factors associated with mental health outcomes. This approach enables the identification of overarching trends, risk factors, and potential interventions on a broader scale, allowing educational institutions to implement targeted strategies to support student mental health. Complementing quantitative analysis, a qualitative approach is also crucial in gaining a deeper understanding of the subjective experiences and nuances of students' mental health. Qualitative methods involve the analysis of non-numerical data such as open-ended survey responses, interviews, and focus group discussions. By delving into the qualitative aspects, researchers can uncover the unique challenges, perceptions, and coping mechanisms of students.

1.3 TOOLS USED

For analyzing student mental health data, one could use a combination of Tableau for visualization, Python for data preprocessing, R for statistical analysis, and machine learning techniques. Python libraries like pandas, numpy, and scikit-learn could be employed for data manipulation, feature engineering, and building predictive models. R packages such as psych and caret could be useful for statistical analysis and model validation. This integrated approach allows for comprehensive exploration, analysis, and interpretation of student mental health data, aiding in the identification of patterns and factors influencing mental well-being.

CHAPTER II - DATA UNDERSTANDING

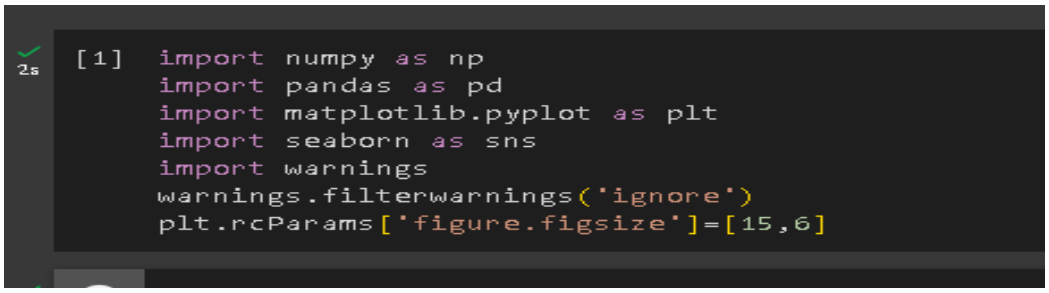
DATA UNDERSTANDING

In today's world, data has become a critical component of our professional and personal daily lives. It helps us decide where to buy a home, what advice we give our children, and what location to visit on vacation. For marketers, data helps us decide which audience to target, what message to send, and what offer to provide. There's ample data out there to help us make these important decisions, but it's more crucial than ever to understand the insights—and even the human connections—that the numbers alone can't tell us.

Importing the libraries

A solid selection of libraries is an essential element of a developer's toolkit for researching and developing complicated applications without having to write a lot of code. In general, a library is a collection of code designed to make common operations go faster.

1. Pandas – for reading the dataset files
2. Numpy – for numerical calculations
3. Matplotlib – for graphic visualization
4. Seaborn – for graphical visualization of the data
5. Sklearn – for scaling the dataset



Reading Dataset

Read the dataset using the pandas library.

```
df=pd.read_csv('/content/students_mental_health_survey.csv')
df.head()
```

	Age	Course	Gender	CGPA	Stress_Level	Depression_Score	Anxiety_Score	Sleep_Quality	Physical_Activity	Diet_
0	25	Others	Male	3.56	3	3	2	Good	Moderate	
1	24	Engineering	Female	2.44	0	3	0	Average	Low	
2	19	Business	Female	3.74	4	0	3	Good	Low	
3	19	Computer Science	Male	3.65	2	1	0	Average	Low	
4	18	Business	Male	3.40	3	3	4	Good	Low	

2.1 DATA COLLECTION

Data collection is the process of gathering and measuring information on targeted variables of interest in an organized system, which then allows you to answer relevant questions and decide future outcomes.

A public dataset available from Kaggle, the actual source of the data is not mention as it confidential. No matter where you obtain the data from, Make sure the data is relevant and Validated. Quality is the key!

2.2 DATA DESCRIPTION

A variable consists of two parts – the label and the data type. Data types can be numeric (integers, real numbers) or strings. The data type can sometimes be tricky; for example, US postal codes are numeric but need to be treated as strings. Once the labels and data types are known, you can group attributes into two kinds for modelling.

Continuous Variables: These are numbers which can range from negative infinity to positive infinity. You would associate with the labels a sense of magnitude, maximum and minimum. You can sort on such variables and filter by ranges.

Categorical Variables: These variables can have a limited set of values, each of which indicate a sub-type. For example, Direction is a categorical variable because it can be either North, South, East, or West. You can filter on or group by a specific value or values of a categorical variable.

Now look into the variables of our dataset. Once you have identified the variables of interest, summary statistics help you understand the nature of each variable. Each attribute's summary statistics such as **count, standard deviation, mean, minimum, maximum and IQR values** are calculated using the **describe()** function. To dive into the dataset, Python Programming is used for further process.

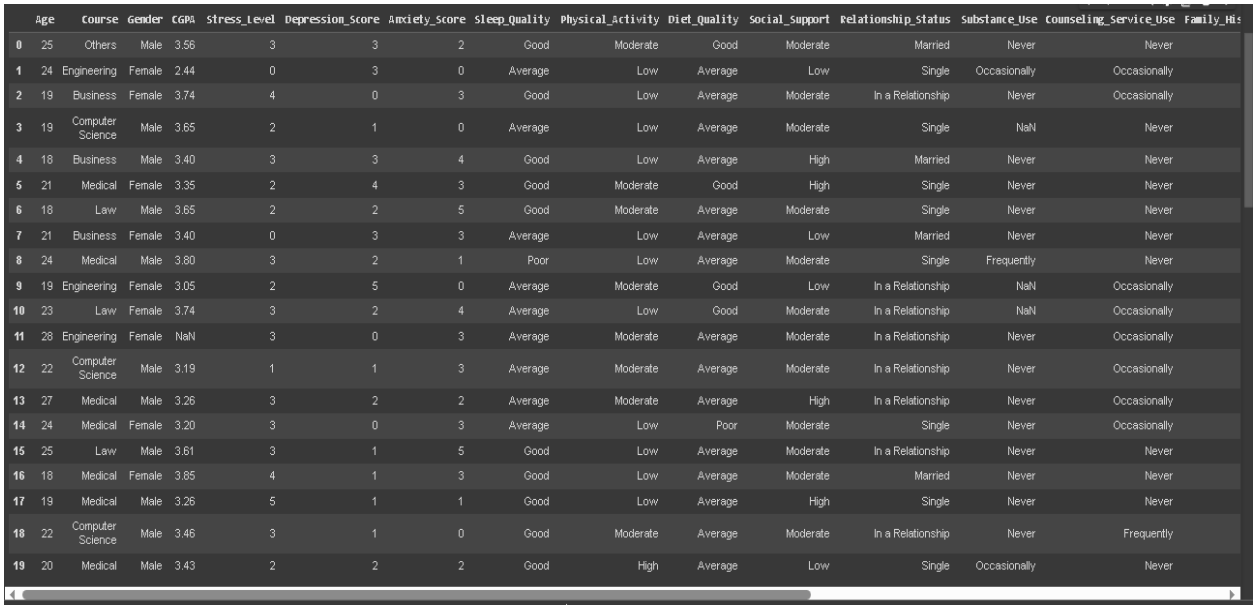
Here **describe()** function is used in python to derive the overall summary of the dataset. But in the dataset most of the values are categories. Therefore **all** datatypes are included in the function as describe() function only takes numeric datatype as default.

```
df.describe(include='all')
```

	Age	Course	Gender	CGPA	Stress_Level	Depression_Score	Anxiety_Score	Sleep_Quality	Physical_
count	7022.000000	7022	7022	7010.00000	7022.000000	7022.000000	7022.000000	7022	
unique	NaN	6	2	NaN	NaN	NaN	NaN	3	
top	NaN	Medical	Male	NaN	NaN	NaN	NaN	Good	
freq	NaN	2105	3547	NaN	NaN	NaN	NaN	3589	
mean	23.003418	NaN	NaN	3.49127	2.427941	2.254486	2.300484	NaN	
std	3.853978	NaN	NaN	0.28742	1.638408	1.625193	1.624305	NaN	
min	18.000000	NaN	NaN	2.44000	0.000000	0.000000	0.000000	NaN	
25%	20.000000	NaN	NaN	3.29000	1.000000	1.000000	1.000000	NaN	
50%	22.000000	NaN	NaN	3.50000	2.000000	2.000000	2.000000	NaN	
75%	25.000000	NaN	NaN	3.70000	4.000000	3.000000	4.000000	NaN	
max	35.000000	NaN	NaN	4.00000	5.000000	5.000000	5.000000	NaN	

Checking for the shape of the dataset

The student mental health Data frame has 7023 rows and 20 columns.



	Age	Course	Gender	CGPA	Stress_Level	Depression_Score	Anxiety_Score	Sleep_Quality	Physical_Activity	Diet_Quality	Social_Support	Relationship_Status	Substance_Use	Counseling_Service_Use	Family_His
0	25	Others	Male	3.56	3	3	2	Good	Moderate	Good	Moderate	Married	Never	Never	
1	24	Engineering	Female	2.44	0	3	0	Average	Low	Average	Low	Single	Occasionally	Occasionally	
2	19	Business	Female	3.74	4	0	3	Good	Low	Average	Moderate	In a Relationship	Never	Occasionally	
3	19	Computer Science	Male	3.65	2	1	0	Average	Low	Average	Moderate	Single	NaN	Never	
4	18	Business	Male	3.40	3	3	4	Good	Low	Average	High	Married	Never	Never	
5	21	Medical	Female	3.35	2	4	3	Good	Moderate	Good	High	Single	Never	Never	
6	18	Law	Male	3.65	2	2	5	Good	Moderate	Average	Moderate	Single	Never	Never	
7	21	Business	Female	3.40	0	3	3	Average	Low	Average	Low	Married	Never	Never	
8	24	Medical	Male	3.80	3	2	1	Poor	Low	Average	Moderate	Single	Frequently	Never	
9	19	Engineering	Female	3.05	2	5	0	Average	Moderate	Good	Low	In a Relationship	NaN	Occasionally	
10	23	Law	Female	3.74	3	2	4	Average	Low	Good	Moderate	In a Relationship	NaN	Occasionally	
11	28	Engineering	Female	NaN	3	0	3	Average	Moderate	Average	Moderate	In a Relationship	Never	Occasionally	
12	22	Computer Science	Male	3.19	1	1	3	Average	Moderate	Average	Moderate	In a Relationship	Never	Occasionally	
13	27	Medical	Male	3.26	3	2	2	Average	Moderate	Average	High	In a Relationship	Never	Occasionally	
14	24	Medical	Female	3.20	3	0	3	Average	Low	Poor	Moderate	Single	Never	Occasionally	
15	25	Law	Male	3.61	3	1	5	Good	Low	Average	Moderate	In a Relationship	Never	Never	
16	18	Medical	Female	3.85	4	1	3	Good	Low	Average	Moderate	Married	Never	Never	
17	19	Medical	Male	3.26	5	1	1	Good	Low	Average	High	Single	Never	Never	
18	22	Computer Science	Male	3.46	3	1	0	Good	Moderate	Average	Moderate	In a Relationship	Never	Frequently	
19	20	Medical	Male	3.43	2	2	2	Good	High	Average	Low	Single	Occasionally	Never	

This data frame contains the following columns:

Stress Level:

The level of stress experienced by the individual.

Depression Score:

The score representing the level of depression experienced by the individual.

Anxiety Score:

The score representing the level of anxiety experienced by the individual.

Sleep Quality:

The quality of sleep experienced by the individual.

Physical Activity:

The level of physical activity.

Diet Quality:

The quality of the individual's diet.

Social Support:

The level of social support received by the individual.

Substance Use:

The frequency of substance use such as alcohol, cigarettes or other drugs.

Family History:

Whether the individual has a family history of mental health issues.

Chronic Illness Financial Stress:

The level of financial stress experienced by the individual.

Semester Credit Load:

The number of credits the individual is taking in the semester.

Summarizing each attributes

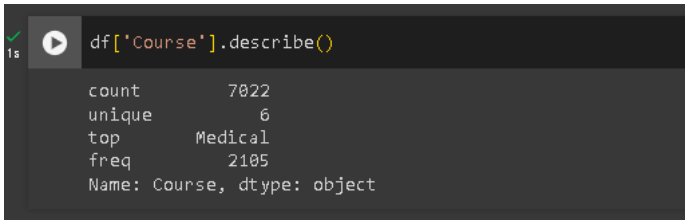
Here know about each and every variables below using **describe()** function individually.

age:

Age is a continuous variable. It denotes the age of the patients. The statistical summary of age variable is given below. As we see the smallest value is in negative, I assume that the data is not accurate. Let's ignore the age as the mental health does not depends on age. It affects people of every age.

Course:

Course is categorical variable The academic program or subject that a student is enrolled in. The specific area of study that a student is pursuing, such as Computer Science, Biology, History, etc.

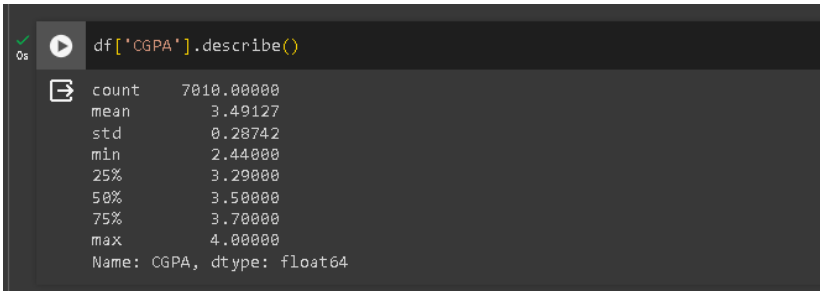


Gender:

Gender is a categorical variable, The classification of individuals as male, female, or another gender identity.

CGPA:

This is a continuous variable, A measure of a student's academic performance, calculated by averaging the grades earned in all completed courses. A numerical representation of a student's overall academic achievement.

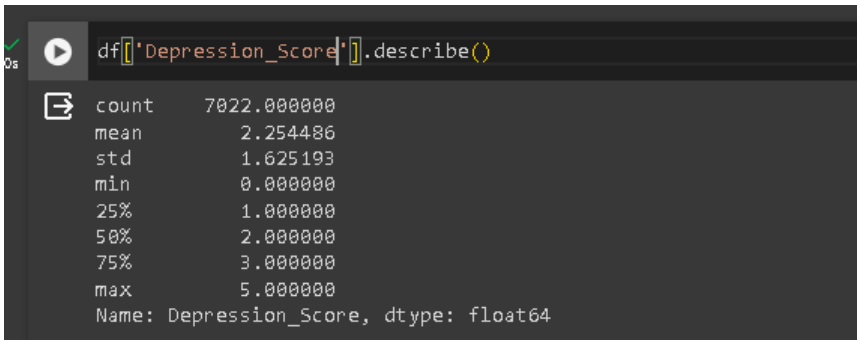


Stress level:

The degree of psychological and emotional pressure experienced by an individual. The subjective feeling of being overwhelmed or strained due to various factors, such as academic demands, personal issues, etc.

Depression score:

Depression score is a continuous variable. A numerical assessment of an individual's level of depression symptoms. A quantification of the severity of depressive symptoms experienced by a person, often measured using standardized assessment tools.



Anxiety score:

Anxiety score is a continuous variable. A quantification of the severity of anxiety symptoms experienced by an individual, typically assessed using standardized instruments.

```
df['Anxiety_Score'].describe()

count    7022.000000
mean        2.300484
std         1.624305
min         0.000000
25%         1.000000
50%         2.000000
75%         4.000000
max         5.000000
Name: Anxiety_Score, dtype: float64
```

Sleep quality:

Sleep quality is a continuous variable, The perceived or measured effectiveness and restfulness of an individual's sleep. The subjective or objective evaluation of how well an individual sleeps, considering factors like duration, depth, and disturbances.

```
df['Sleep_Quality'].describe()

count     7022
unique       3
top       Good
freq      3589
Name: Sleep_Quality, dtype: object
```

Physical activity:

Physical activity is a continuous variable. The level of engagement in bodily movements and exercise. The extent to which an individual participates in physical activities, such as sports, workouts, or any form of exercise.

```
df['Physical_Activity'].describe()

count         7022
unique           3
top       Moderate
freq          3521
Name: Physical_Activity, dtype: object
```

Diet quality:

Diet quality Is a continuous variable. The nutritional value and healthiness of an individual's dietary habits. The overall nutritional content and balance of the foods consumed by a person.

```
df['Diet_Quality'].describe()

count        7022
unique          3
top       Average
freq         4268
Name: Diet_Quality, dtype: object
```

Social support:

Social support is a continuous variable. The presence and effectiveness of relationships and assistance from friends, family, or other social connections.

```
df['Social_Support'].describe()

count         7022
unique           3
top       Moderate
freq          3470
Name: Social_Support, dtype: object
```

CHAPTER III – DATA PREPARATION

3.1 DATA CLEANING

Data cleaning is a fundamental aspect of data science, essential for ensuring the accuracy, reliability, and usability of datasets. It involves identifying and rectifying errors, inconsistencies, and missing values within the data. By removing noise and irrelevant information, data cleaning enhances the quality of the dataset, enabling more accurate analysis and modeling. Moreover, it helps in maintaining consistency in formats, units, and representations, facilitating the integration of data from multiple sources. Additionally, data cleaning plays a vital role in reducing biases that may be present in the data, thereby ensuring fairness and impartiality in analysis and decision-making processes. Ultimately, investing time and effort in data cleaning upfront saves resources and enhances the effectiveness of subsequent data analysis and modeling tasks.

Around 80% of your time will be spent cleaning data. Cleaning your data is a process of ensuring your data is in the correct format; consistent and errors are identified and dealt with appropriately. The actions below lead to a cleaner dataset:

- Remove duplicate values (This is usually the case when combining multiple datasets)
- Remove irrelevant observations (observations need to be specific to the problem you are solving)
- Address missing values (e.g., Imputation techniques, drop features/observations)
- Reformat data types (e.g., Boolean, numeric, Datetime)
- Filter unwanted outliers (if you have a legitimate reason)
- Reformat strings (e.g., remove white spaces, mislabelled/misspelt categories)
- Validate (does the data make sense? does the data adhere to the defined business rules?)
- Cleaning your data will allow for higher-quality information and ultimately lead to more conclusive and accurate decision.

- Before getting into this step, Let's take look of the overall summary of the dataset using **info()** function in python.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7022 entries, 0 to 7021
Data columns (total 20 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Age                           7022 non-null   int64
 1   Course                        7022 non-null   object
 2   Gender                        7022 non-null   object
 3   CGPA                          7010 non-null   float64
 4   Stress_Level                  7022 non-null   int64
 5   Depression_Score              7022 non-null   int64
 6   Anxiety_Score                 7022 non-null   int64
 7   Sleep_Quality                 7022 non-null   object
 8   Physical_Activity             7022 non-null   object
 9   Diet_Quality                  7022 non-null   object
 10  Social_Support                7022 non-null   object
 11  Relationship_Status           7022 non-null   object
 12  Substance_Use                 7007 non-null   object
 13  Counseling_Service_Use        7022 non-null   object
 14  Family_History                7022 non-null   object
 15  Chronic_Illness               7022 non-null   object
 16  Financial_Stress              7022 non-null   int64
 17  Extracurricular_Involvement   7022 non-null   object
 18  Semester_Credit_Load          7022 non-null   int64
 19  Residence_Type                7022 non-null   object
dtypes: float64(1), int64(6), object(13)
memory usage: 1.1+ MB
```

As we know the total records in the dataset is 7022, in the picture above, it is stated that the variable **CGPA** has 7010 and **SUBSTANCE USE** has 7007 non null values, which means the variable has null values.

3.2 HANDLING NA VALUES

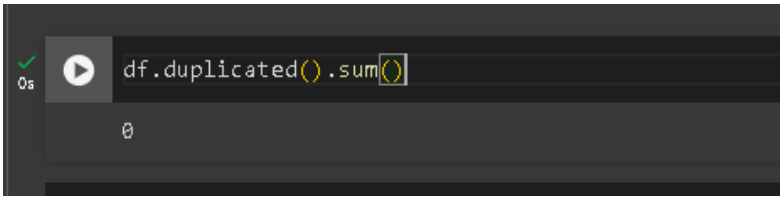
```
df=df.dropna()
df.head()
```

	Age	Course	Gender	CGPA	Stress_Level	Depression_Score	Anxiety_Score	Sleep_Quality	Physical_Activity	Diet_
0	25	Others	Male	3.56	3	3	2	Good	Moderate	
1	24	Engineering	Female	2.44	0	3	0	Average	Low	
2	19	Business	Female	3.74	4	0	3	Good	Low	
4	18	Business	Male	3.40	3	3	4	Good	Low	
5	21	Medical	Female	3.35	2	4	3	Good	Moderate	

In the image above, NA values as we cannot fill correct values so I drop that record.

3.3 CHECKING FOR DUPLICATE VALUES

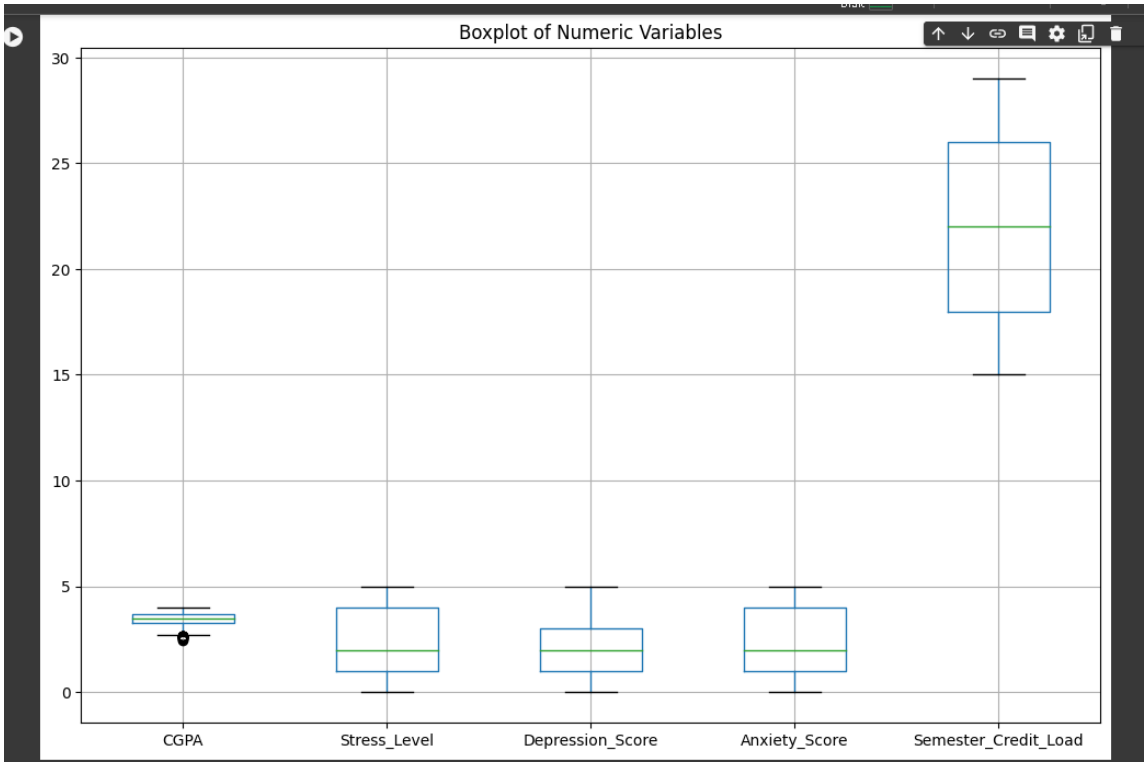
It is very important for you to remove duplicates from the dataset to maintain accuracy and to avoid misleading statistics. To check there is duplicates or not. The Python pandas library has a method for it, that is **duplicated().** Use **sum()** along with it, then it will return the total number of the duplicates in the dataset.



3.4 OUTLIER DETECTION

Outlier is a data point in the dataset that differs significantly from the other data or observations. It can mess up your analysis. There are many ways to deal with outliers. There are some techniques used to deal with outliers.

- Deleting observations
- Transforming values
- Imputation
- Separately treating
- Deleting observations



According to the above graph there are no Outliers.

CHAPTER IV – EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is the crucial process of using summary statistics and graphical representations to perform preliminary investigations on data in order to uncover patterns, detect anomalies, test hypotheses, and verify assumptions.

In simple words, EDA is a data exploration technique to understand the various aspects of the data. EDA is often used to see what data may disclose outside of formal modelling and to learn more about the variables in a data

collection and how they interact. It could also help us figure out if the statistical procedures we are considering for data analysis are appropriate. Before modelling the data, it gives insight into all of the data and the numerous interactions between the data elements.

4.1 DATA VISUALIZATION

Data visualization is extremely useful in understanding the data and obtaining useful insights. It can allow you to get an instant understanding of the data that is just not possible by observing rows of data in a table. That's what makes it so important in Data Science! Let's see some more reasons why data visualization is so important.

Data Visualization discovers the trends in data. It is interactive and provides a perspective on the data. It explains a data process, tells a data story and puts the data into the correct context. Data visualization is educational for users and saves time.

Some charts below to understand the data visually. Before starting the analysis, Three new variables are created and stored some values in order to perform the analysis efficiently.

Col variable holds every column names of the dataset.

```
col=list(df.columns)
col
```

```
['Age',
 'Course',
 'Gender',
 'CGPA',
 'Stress_Level',
 'Depression_Score',
 'Anxiety_Score',
 'Sleep_Quality',
 'Physical_Activity',
 'Diet_Quality',
 'Social_Support',
 'Relationship_Status',
 'Substance_Use',
 'Counseling_Service_Use',
 'Family_History',
 'Chronic_Illness',
 'Financial_Stress',
 'Extracurricular_Involvement',
 'Semester_Credit_Load',
 'Residence_Type']
```

categorical_var stores the categorical variables.

Categorical variables

```
categorical_vars = df.select_dtypes(include='object').columns
print("Categorical Variables:", categorical_vars)
```

```
Categorical Variables: Index(['Course', 'Gender', 'Sleep_Quality', 'Physical_Activity',
       'Diet_Quality', 'Social_Support', 'Relationship_Status',
       'Substance_Use', 'Counseling_Service_Use', 'Family_History',
       'Chronic_Illness', 'Extracurricular_Involvement', 'Residence_Type'],
      dtype='object')
```

```
total_categorical_vars = len(categorical_vars)
```

```
print("Total Count of categorical Variables:", total_categorical_vars)
```

```
Total Count of categorical Variables: 13
```

numeric_var stores the numerical variables.

Continuous variables

```
[ ]  numeric_vars = df.select_dtypes(include=['int64', 'float64']).columns
     print("Continuous Variables:", numeric_vars)
```

```
Continuous Variables: Index(['Age', 'CGPA', 'Stress_Level', 'Depression_Score', 'Anxiety_Score',
       'Financial_Stress', 'Semester_Credit_Load'],
      dtype='object')
```

Count of continuous variables

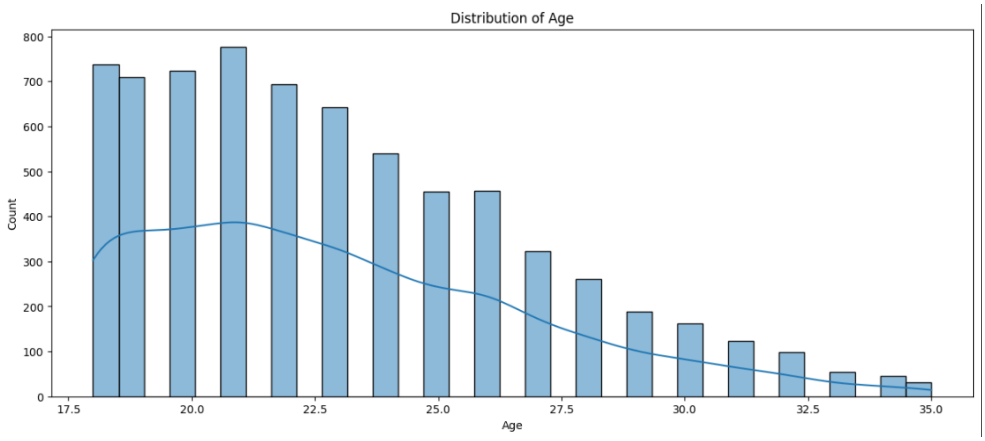
```
[ ]  numeric_vars = df.select_dtypes(include=['int64', 'float64']).columns
```

```
[ ]  total_numeric_vars = len(numeric_vars)
```

```
[ ]  print("Total Count of Numeric (Continuous) Variables:", total_numeric_vars)
```

```
Total Count of Numeric (Continuous) Variables: 7
```

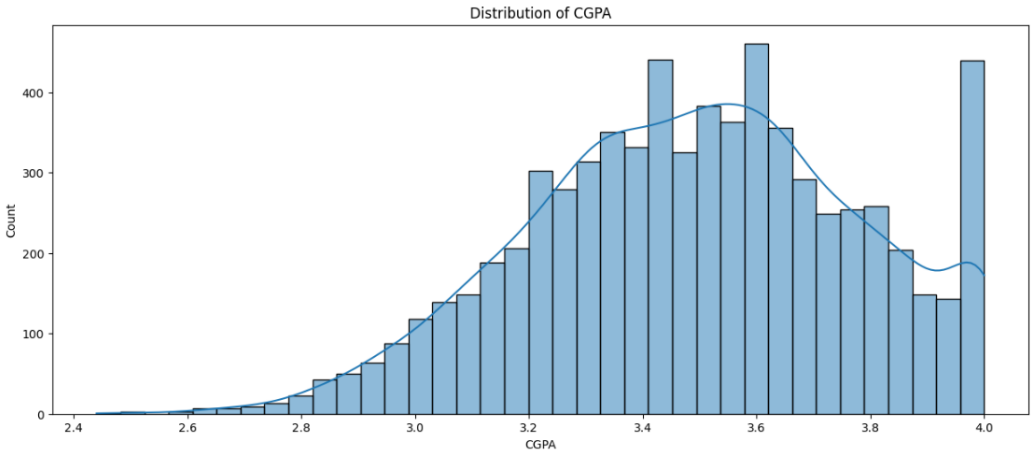
Distribution of age in the dataset



The chart has age ranges on the x-axis and counts on the y-axis. The age ranges are 17.5, 20.0, 22.5, 25.0, 27.5, and 30.0, which could represent the midpoints of the age groups. The counts on the y-axis indicate how many individuals fall within each age range.

If the bar for the age range 20.0 is the highest, it means that the majority of the population is between the ages of 19.5 and 20.5. Similarly, lower bars for other age ranges would indicate fewer individuals in those age groups.

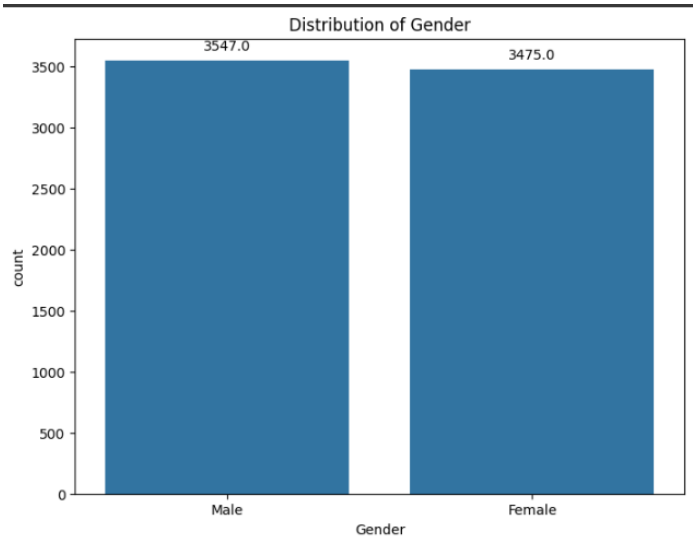
Distribution of CGPA



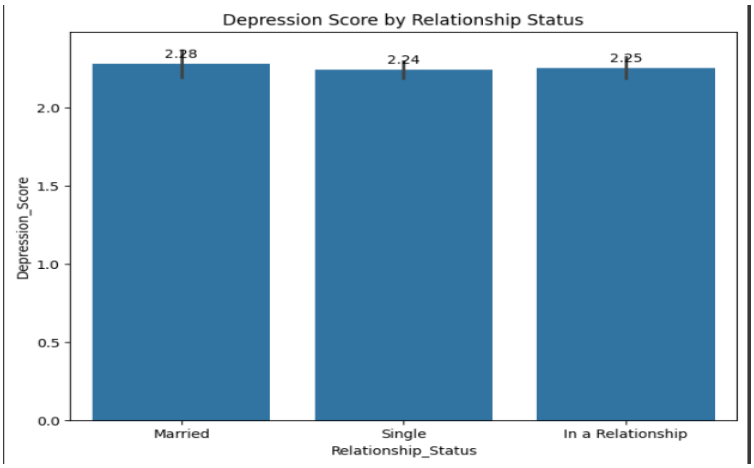
There is a chart showing the distribution of CGPA (Cumulative Grade Point Average). The CGPA values are ranging from 2.4 to 4.0 with intervals of 0.2. The y-axis shows the number of students with a given CGPA range.

It can be observed that the majority of the students fall in the CGPA range of 3.0 to 3.2, followed by a slightly smaller number of students in the CGPA range of 2.8 to 3.0. The number of students with a CGPA below 2.6 or above 3.4 seems to be significantly lower.

Distribution of Gender



Depression score by relationship status

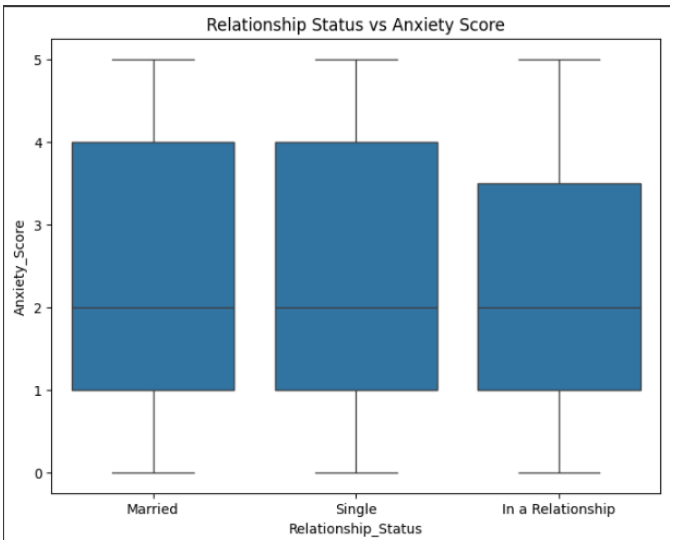


The chart shows the "depression score" across three categories of relationship status: Married, Single, and In a Relationship.The chart appears to be a bar chart or a column chart with the relationship statuses on the x-axis. The y-axis represents the depression scores, with a range of 0.0 to 2.0 in increments of 0.5.

The average depression scores are:

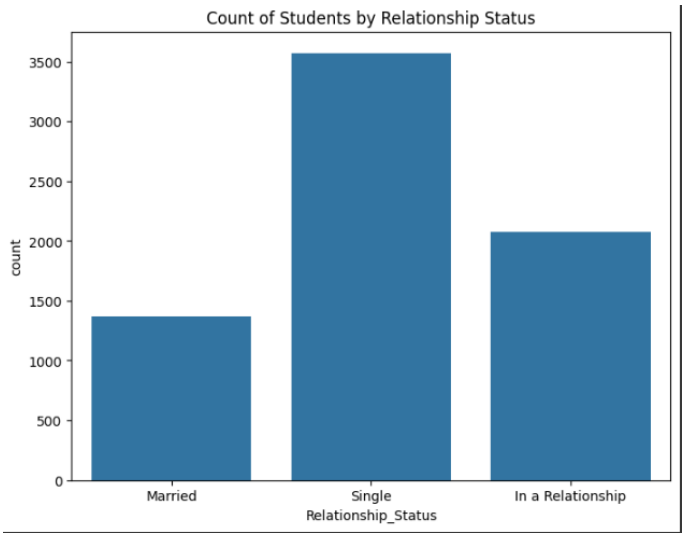
- Approximately 0.5 for Married individuals.
- Around 1.0 for those who are Single.
- Around 1.5 for individuals in a Relationship.

Anxiety score by relationship status



The chart plots Relationship Status against Anxiety Score. The x-axis represents the Relationship Status, with three categories: Married, Single, and In a Relationship. The y-axis corresponds to the Anxiety Score, which ranges from 1 to 5, with a given interval.The average Anxiety Score for Married individuals is approximately 1.8, while for those In a Relationship, it is about 2.2.

Count of students by relationship status

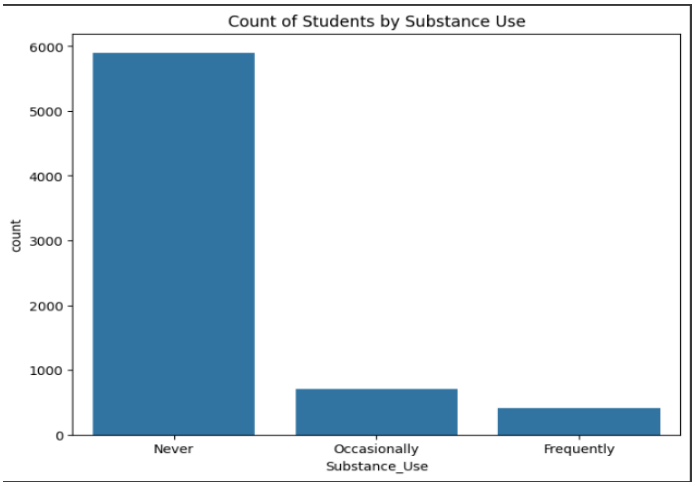


The chart shows a "Count of Students by Relationship Status," with three categories of relationship status: Married, Single, and In a Relationship on the x-axis. The y-axis represents the number of students.

- There are approximately 1,000 married students
- The number of single students is around 3,500
- The count of students in a relationship is about 2,500

Married students make up the smallest group and The single students are the largest group and The number of students who are In a Relationship falls in between the two other groups.

Count of students by substance use

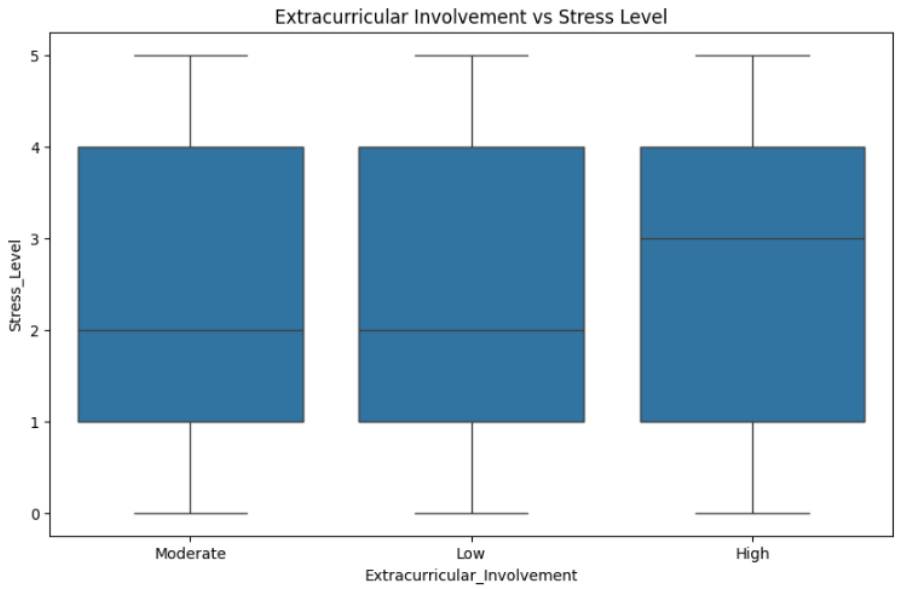


The bar chart displaying the "Count of Students" in relation to their levels of substance use. The categories of substance use are:

1. Never
2. Occasionally
3. Frequently

The chart shows the number of students that fall into each category. It appears that the majority of students selected "Never" as their substance use, as indicated by the tallest bar in the chart. The numbers 6000, 5000, 4000, 3000, 2000, and 1000 are likely labels for each bar, representing different counts.

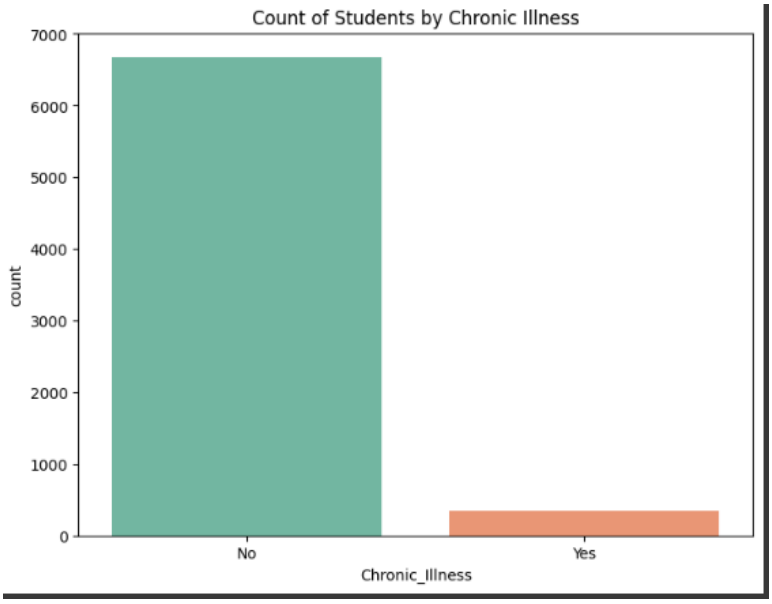
Extracurricular involvement vs stress level



Scatter plot showing the relationship between two variables: "Extracurricular Involvement" and "Stress Level." The x-axis represents "Extracurricular Involvement," which is likely a continuous variable ranging from 0 to 5. The y-axis represents "Stress Level," which is also a continuous variable.

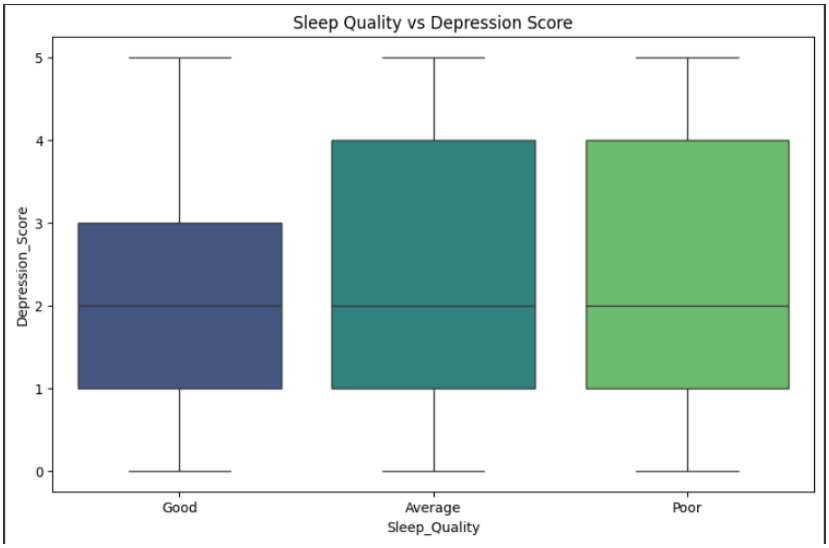
Each point on the chart represents a data point with a specific value for both "Extracurricular Involvement" and "Stress Level. It appears that there might be a general trend of higher stress levels associated with higher extracurricular involvement

Count of students by chronic illness



The bar chart displaying the "Count of Students" based on whether they have a chronic illness or not**.** which has two categories: "No" and "Yes." The y-axis represents the count of students in each category. The chart shows that the majority of students do not have a chronic illness, as indicated by the taller bar on the left side.

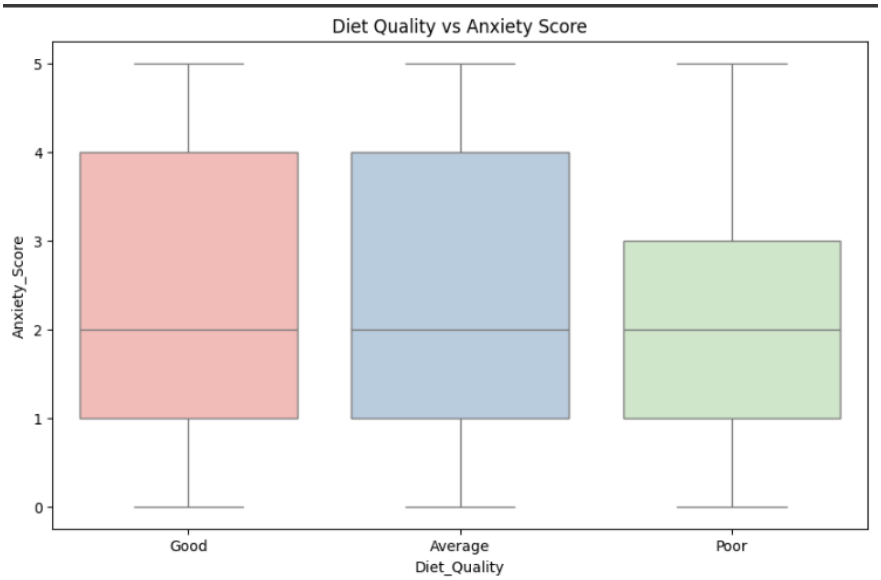
Sleep quality vs depression score



Scatter plot showing the relationship between two variables: "Sleep Quality" and "Depression Score." The x-axis represents the "Sleep Quality" variable, which has three categories: "Good," "Average," and "Poor." The y-axis represents the "Depression Score" variable, which is likely a continuous variable.

Each point on the chart represents a data point with a specific value for both "Sleep Quality" and "Depression Score." The chart shows that higher depression scores are generally associated with poorer sleep quality, as indicated by the trend of points moving from the bottom left to the top right of the chart.

Diet quality vs anxiety score

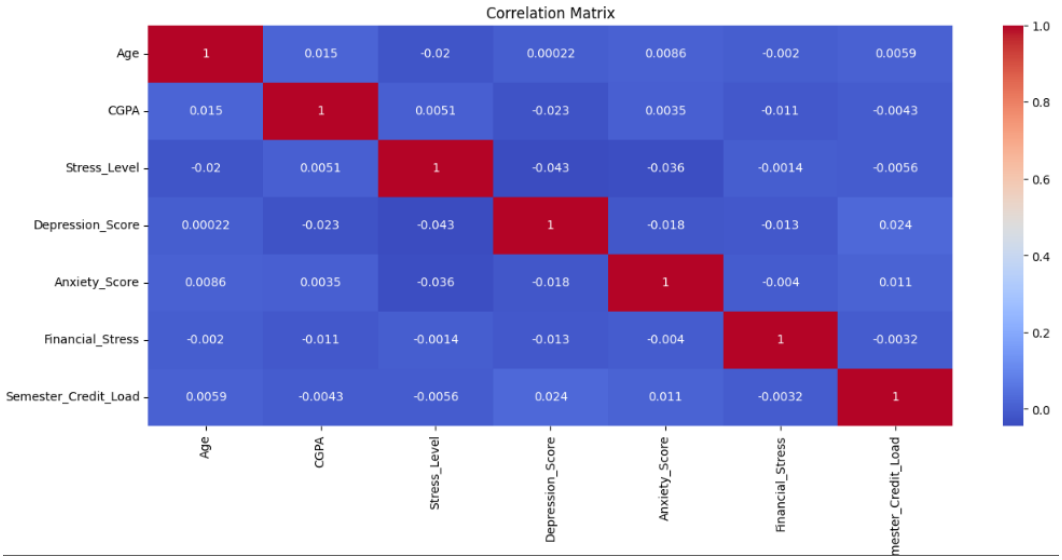


plot showing the relationship between two variables: "Diet Quality" and "Anxiety Score." The x-axis represents the "Diet Quality" variable, which has three categories: "Good," "Average," and "Poor".

Each point on the chart represents a data point with a specific value for both "Diet Quality" and "Anxiety Score." The chart shows that higher anxiety scores are generally associated with poorer diet quality.

4.2 CORRELATION

The correlation coefficient helps in measuring the extent of the relationship between two variables in one figure. It analysis facilitates the understanding of economic behavior and helps in locating the critically important variables on which others depend. When two variables are correlated, the value of one variable can be estimated, given the value of another. This is done with the help of regression equations. Correlation facilitates the decision-making in the business world. It reduces the range of uncertainty as predictions based on correlation are likely to be more reliable and near to reality.



4.3 Overall Insights

- High stress levels can be a crucial factor affecting student mental health. Analyze the stress levels reported by students to understand the extent of pressure they may be experiencing, potentially due to academic demands, personal issues, or other sources.

- Poor sleep quality is often linked to mental health issues. Explore the sleep quality variable to determine if students with lower sleep quality scores also report higher levels of stress, depression, or anxiety. Adequate and quality sleep is essential for overall well-being.

- The availability of social support can significantly impact mental health. Investigate how students' mental health is influenced by the level of support they receive from their social networks. Higher social support may act as a protective factor against mental health challenges.

- Physical activity is known to have positive effects on mental health. Examine the relationship between students' mental health and their reported levels of physical activity. Higher physical activity levels may correlate with lower stress, depression, and anxiety scores.

- Relationship status can impact mental well-being. Investigate whether students in certain relationship statuses (single, in a relationship, etc.) exhibit variations in mental health indicators. Relationship-related stressors or support can influence mental health outcomes.

- Utilization of counselling services can indicate students' awareness of and proactive approach to managing their mental health. Analyze whether students who use counselling services report lower levels of stress, depression, and anxiety, suggesting the effectiveness of mental health support.

- Explore the variable related to financial stress. Financial difficulties can contribute to heightened stress and negatively impact mental health. Investigate whether students facing financial stress report higher levels of stress, depression, or anxiety.

CHAPTER V – CONCLUSION

Conclusion

The dataset provided offers a glimpse into the multifaceted landscape of student mental health, showcasing various factors that can influence well-being during the academic journey. It appears that academic performance, as indicated by CGPA, may significantly impact stress levels, with the pressure to maintain high grades potentially exacerbating feelings of anxiety and depression. Moreover, lifestyle factors such as sleep quality, physical activity, and diet quality emerge as critical determinants of mental well-being, underscoring the importance of holistic health practices.

Social support, both within personal relationships and through counselling services, appears to play a pivotal role in buffering against the negative effects of stress and fostering resilience. However, challenges such as substance use, chronic illness, financial stress, and the balancing act of extracurricular involvement underscore the need for tailored interventions that address the unique needs of each individual. Moving forward, efforts to promote student mental health must encompass a comprehensive approach that integrates academic support, access to mental health resources, and the cultivation of supportive communities within educational institutions.

By recognizing and addressing the complex interplay of factors influencing student mental health, we can strive towards creating environments that nurture not only academic success but also holistic well-being.

CHAPTER VI– BIBLIOGRAPHY

[1] Kaggle: Students Mental Health Assessments (kaggle.com)

[2] Medium: https://towardsdatascience.com/random-forest-regression-5f605132d19d

[3] Java point: https://www.javatpoint.com/linear-regression-in-machine-learning

[4] Data Analysis with Python: Zero to Pandas by Jovian https://jovian.ai/learn/data-analysis-with-python-zero-to-pandas

[5] Pandas user guide: https://pandas.pydata.org/docs/user_guide/index.html

[6] Matplotlib user guide: https://matplotlib.org/3.3.1/users/index.html

[7] Seaborn user guide: https://seaborn.pydata.org/tutorial.html