Student Mental Health

Libararies

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Load Dataset

In [2]:
```
SMH = pd.read_csv('Student Mental health.csv')
SMH
```

Out[2]:

	Timestamp	Choose your gender	Age	What is your course?	Your current year of Study	What is your CGPA?	Marital status	Do you have Depression?	Do you have Anxiety?	Do you have Panic attack?	Did you seek any specialist for a treatment?
0	08-07-2020 12:02	Female	18.0	Engineering	year 1	3.00 - 3.49	No	Yes	No	Yes	No
1	08-07-2020 12:04	Male	21.0	Islamic education	year 2	3.00 - 3.49	No	No	Yes	No	No
2	08-07-2020 12:05	Male	19.0	BIT	Year 1	3.00 - 3.49	No	Yes	Yes	Yes	No
3	08-07-2020 12:06	Female	22.0	Laws	year 3	3.00 - 3.49	Yes	Yes	No	No	No
4	08-07-2020 12:13	Male	23.0	Mathemathics	year 4	3.00 - 3.49	No	No	No	No	No
...	...	...	...	...	...	...	...	...	...	...	...
96	13-07-2020 19:56	Female	21.0	BCS	year 1	3.50 - 4.00	No	No	Yes	No	No
97	13-07-2020 21:21	Male	18.0	Engineering	Year 2	3.00 - 3.49	No	Yes	Yes	No	No
98	13-07-2020 21:22	Female	19.0	Nursing	Year 3	3.50 - 4.00	Yes	Yes	No	Yes	No
99	13-07-2020 21:23	Female	23.0	Pendidikan Islam	year 4	3.50 - 4.00	No	No	No	No	No
100	18-07-2020 20:16	Male	20.0	Biomedical science	Year 2	3.00 - 3.49	No	No	No	No	No

101 rows × 11 columns

Basic operations and Data cleaning

In [3]:
```
SMH.head()
```

Out[3]:

	Timestamp	Choose your gender	Age	What is your course?	Your current year of Study	What is your CGPA?	Marital status	Do you have Depression?	Do you have Anxiety?	Do you have Panic attack?	Did you seek any specialist for a treatment?
0	08-07-2020 12:02	Female	18.0	Engineering	year 1	3.00 - 3.49	No	Yes	No	Yes	No
1	08-07-2020 12:04	Male	21.0	Islamic education	year 2	3.00 - 3.49	No	No	Yes	No	No
2	08-07-2020 12:05	Male	19.0	BIT	Year 1	3.00 - 3.49	No	Yes	Yes	Yes	No
3	08-07-2020 12:06	Female	22.0	Laws	year 3	3.00 - 3.49	Yes	Yes	No	No	No
4	08-07-2020 12:13	Male	23.0	Mathemathics	year 4	3.00 - 3.49	No	No	No	No	No

In [4]:
```
SMH.tail()
```

Out[4]:

	Timestamp	Choose your gender	Age	What is your course?	Your current year of Study	What is your CGPA?	Marital status	Do you have Depression?	Do you have Anxiety?	Do you have Panic attack?	Did you seek any specialist for a treatment?
96	13-07-2020 19:56	Female	21.0	BCS	year 1	3.50 - 4.00	No	No	Yes	No	No
97	13-07-2020 21:21	Male	18.0	Engineering	Year 2	3.00 - 3.49	No	Yes	Yes	No	No
98	13-07-2020 21:22	Female	19.0	Nursing	Year 3	3.50 - 4.00	Yes	Yes	No	Yes	No
99	13-07-2020 21:23	Female	23.0	Pendidikan Islam	year 4	3.50 - 4.00	No	No	No	No	No
100	18-07-2020 20:16	Male	20.0	Biomedical science	Year 2	3.00 - 3.49	No	No	No	No	No

In [5]:
```
SMH.shape
```

Out[5]: (101, 11)

In [6]:
```
SMH.sample()
```

Out[6]:

	Timestamp	Choose your gender	Age	What is your course?	Your current year of Study	What is your CGPA?	Marital status	Do you have Depression?	Do you have Anxiety?	Do you have Panic attack?	Did you seek any specialist for a treatment?
72	08-07-2020 22:35	Female	19.0	BIT	Year 3	3.00 - 3.49	Yes	Yes	No	No	No

In [7]:
```
SMH.describe()
```

Out[7]:

	Age
count	100.00000
mean	20.53000
std	2.49628
min	18.00000
25%	18.00000
50%	19.00000
75%	23.00000
max	24.00000

In [8]:
```
SMH.describe(include = 'object')
```

Out[8]:

	Timestamp	Choose your gender	What is your course?	Your current year of Study	What is your CGPA?	Marital status	Do you have Depression?	Do you have Anxiety?	Do you have Panic attack?	Did you seek any specialist for a treatment?
count	101	101	101	101	101	101	101	101	101	101
unique	90	2	49	7	6	2	2	2	2	2
top	13-07-2020 10:12	Female	BCS	year 1	3.50 - 4.00	No	No	No	No	No
freq	3	75	18	41	47	85	66	67	68	95

In [9]:
```
SMH.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101 entries, 0 to 100
Data columns (total 11 columns):
 #   Column                                        Non-Null Count  Dtype  
---  ------                                        --------------  -----  
 0   Timestamp                                     101 non-null    object 
 1   Choose your gender                            101 non-null    object 
 2   Age                                           100 non-null    float64
 3   What is your course?                          101 non-null    object 
 4   Your current year of Study                    101 non-null    object 
 5   What is your CGPA?                            101 non-null    object 
 6   Marital status                                101 non-null    object 
 7   Do you have Depression?                       101 non-null    object 
 8   Do you have Anxiety?                          101 non-null    object 
 9   Do you have Panic attack?                     101 non-null    object 
 10  Did you seek any specialist for a treatment?  101 non-null    object 
dtypes: float64(1), object(10)
memory usage: 8.8+ KB
```

In [10]:
```
SMH.columns
```

Out[10]:
```
Index(['Timestamp', 'Choose your gender', 'Age', 'What is your course?',
       'Your current year of Study', 'What is your CGPA?', 'Marital status',
       'Do you have Depression?', 'Do you have Anxiety?',
       'Do you have Panic attack?',
       'Did you seek any specialist for a treatment?'],
      dtype='object')
```

In [11]:
```
SMH.isnull().sum()
```

Out[11]:
```
Timestamp                                      0
Choose your gender                             0
Age                                            1
What is your course?                           0
Your current year of Study                     0
What is your CGPA?                             0
Marital status                                 0
Do you have Depression?                        0
Do you have Anxiety?                           0
Do you have Panic attack?                      0
Did you seek any specialist for a treatment?   0
dtype: int64
```

In [12]:
```
SMH.dropna(inplace=True)
SMH.isna().sum()
```

Out[12]:
```
Timestamp                                      0
Choose your gender                             0
Age                                            0
What is your course?                           0
Your current year of Study                     0
What is your CGPA?                             0
Marital status                                 0
Do you have Depression?                        0
Do you have Anxiety?                           0
Do you have Panic attack?                      0
Did you seek any specialist for a treatment?   0
dtype: int64
```

In [13]:
```
SMH.isnull().sum()
```

Out[13]:
```
Timestamp                                      0
Choose your gender                             0
Age                                            0
What is your course?                           0
Your current year of Study                     0
What is your CGPA?                             0
Marital status                                 0
Do you have Depression?                        0
Do you have Anxiety?                           0
Do you have Panic attack?                      0
Did you seek any specialist for a treatment?   0
dtype: int64
```

In [14]:
```
columns_names = ['Timestamp', 'Gender', 'Age', 'Course', 'Year of Study', 'CGPA', 'Is Married?', 'Has Depression?','Ha
SMH.columns = columns_names
SMH.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 100 entries, 0 to 100
Data columns (total 11 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Timestamp                 100 non-null    object
 1   Gender                    100 non-null    object
 2   Age                       100 non-null    float64
 3   Course                    100 non-null    object
 4   Year of Study             100 non-null    object
 5   CGPA                      100 non-null    object
 6   Is Married?               100 non-null    object
 7   Has Depression?           100 non-null    object
 8   Has Anxiety?              100 non-null    object
 9   Has Panic Attacks?        100 non-null    object
 10  Is Undergoing Treatment?  100 non-null    object
dtypes: float64(1), object(10)
memory usage: 9.4+ KB
```

In [15]:
```
SMH['Year of Study'].unique()
```

Out[15]:
```
array(['year 1', 'year 2', 'Year 1', 'year 3', 'year 4', 'Year 2',
       'Year 3'], dtype=object)
```

In [16]:
```
SMH['Year of Study'] = SMH['Year of Study'].str.split().str[1].astype(int)
SMH['Year of Study'].unique()
```

Out[16]:
```
array([1, 2, 3, 4])
```

In [17]:
```
SMH
```

Out[17]:

	Timestamp	Gender	Age	Course	Year of Study	CGPA	Is Married?	Has Depression?	Has Anxiety?	Has Panic Attacks?	Is Undergoing Treatment?
0	08-07-2020 12:02	Female	18.0	Engineering	1	3.00 - 3.49	No	Yes	No	Yes	No
1	08-07-2020 12:04	Male	21.0	Islamic education	2	3.00 - 3.49	No	No	Yes	No	No
2	08-07-2020 12:05	Male	19.0	BIT	1	3.00 - 3.49	No	Yes	Yes	Yes	No
3	08-07-2020 12:06	Female	22.0	Laws	3	3.00 - 3.49	Yes	Yes	No	No	No
4	08-07-2020 12:13	Male	23.0	Mathemathics	4	3.00 - 3.49	No	No	No	No	No
...	...	...	...	...	...	...	...	...	...	...	...
96	13-07-2020 19:56	Female	21.0	BCS	1	3.50 - 4.00	No	No	Yes	No	No
97	13-07-2020 21:21	Male	18.0	Engineering	2	3.00 - 3.49	No	Yes	Yes	No	No
98	13-07-2020 21:22	Female	19.0	Nursing	3	3.50 - 4.00	Yes	Yes	No	Yes	No
99	13-07-2020 21:23	Female	23.0	Pendidikan Islam	4	3.50 - 4.00	No	No	No	No	No
100	18-07-2020 20:16	Male	20.0	Biomedical science	2	3.00 - 3.49	No	No	No	No	No

100 rows × 11 columns

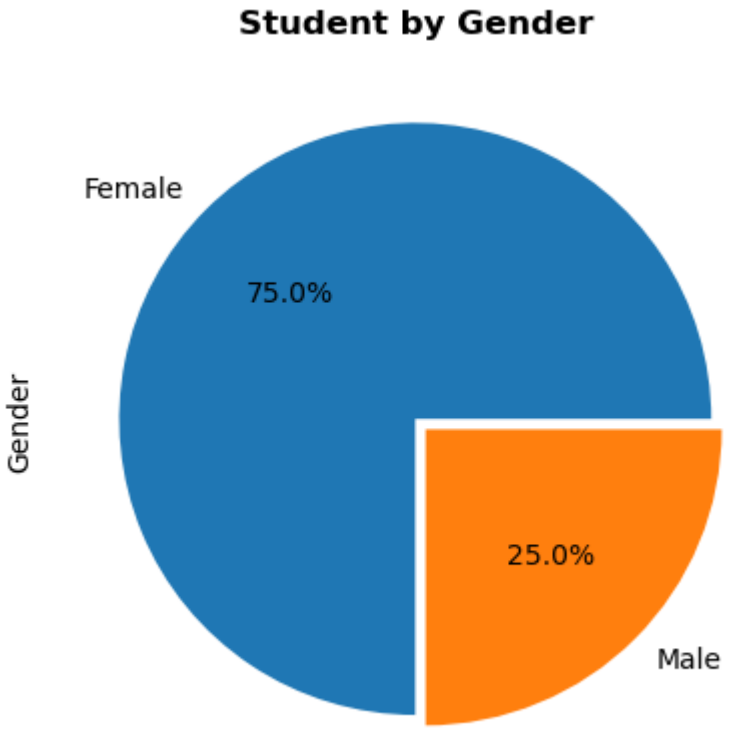
Data Visualization

Students by Gender

In [18]:
```
SMH['Gender'].value_counts()
```

Out[18]:
```
Female    75
Male      25
Name: Gender, dtype: int64
```

In [19]:
```
SMH['Gender'].value_counts().plot(kind='pie',autopct = '%1.1f%%',explode = [0,0.05])
plt.title('Student by Gender',fontweight = 'bold')
plt.show()
```

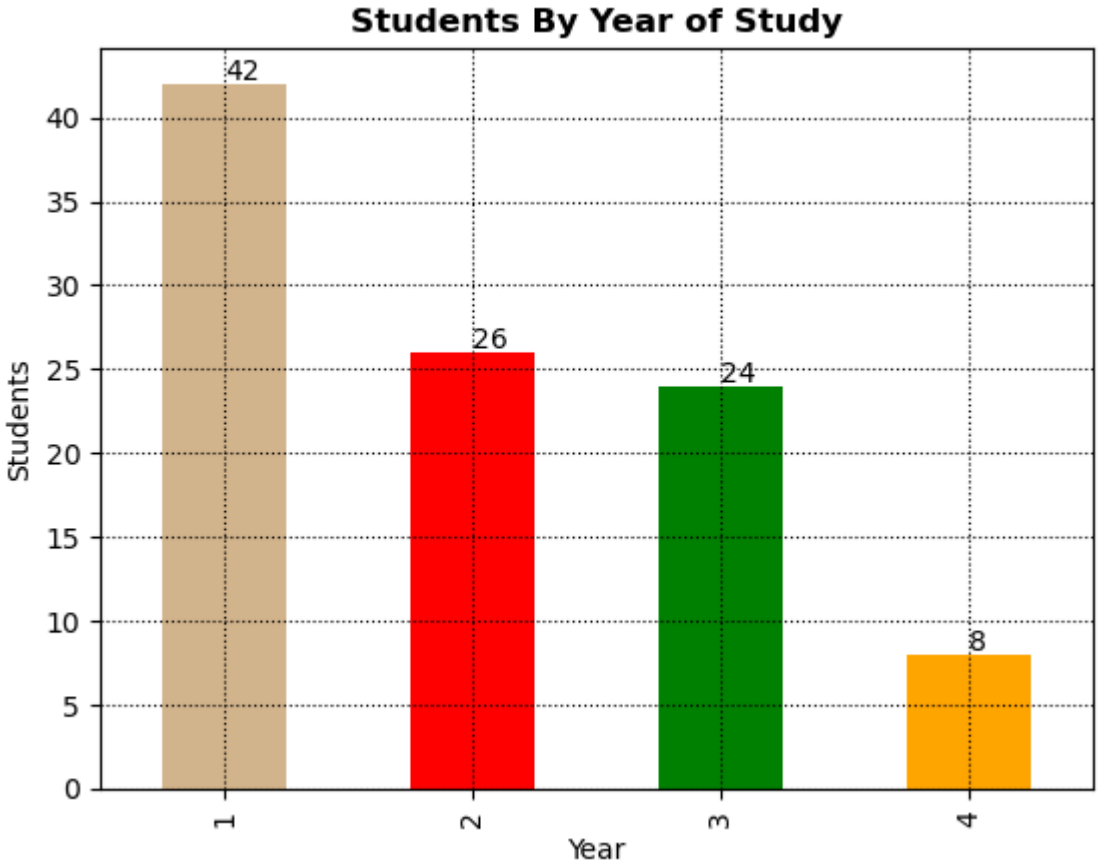


Students year of study

In [20]:
```
SMH['Year of Study'].value_counts()
```

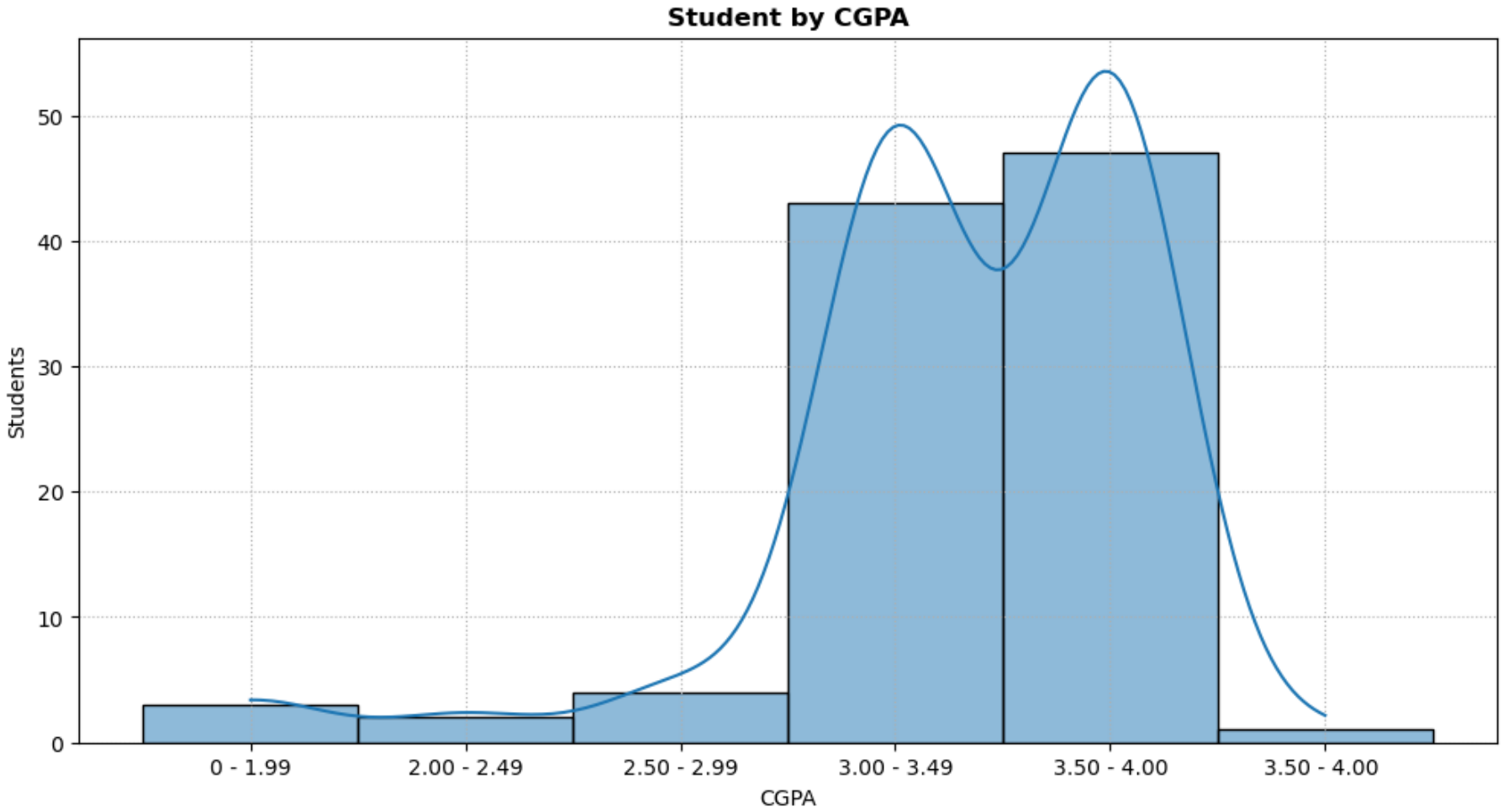
Out[20]:
```
1    42
2    26
3    24
4     8
Name: Year of Study, dtype: int64
```

In [21]:
```
SMH['Year of Study'].value_counts().plot(kind='bar',color=['tan','red','green','orange'])
plt.grid(c='k',ls=':')
plt.title ('Students By Year of Study',fontweight = 'bold')
plt.xlabel('Year')
plt.ylabel('Students')
plt.text(0,42.2,'42',c='k')
plt.text(1,26.2,'26',c='k')
plt.text(2,24.2,'24',c='k')
plt.text(3,8.2,'8',c='k')
plt.show()
```



Students CGPA

In [22]:
```
plt.figure(figsize=(12,6))
sns.histplot(SMH['CGPA'].sort_values(),kde=True)
plt.title('Student by CGPA',fontweight='bold')
plt.ylabel('Students')
plt.grid(ls=':')
plt.show()
```

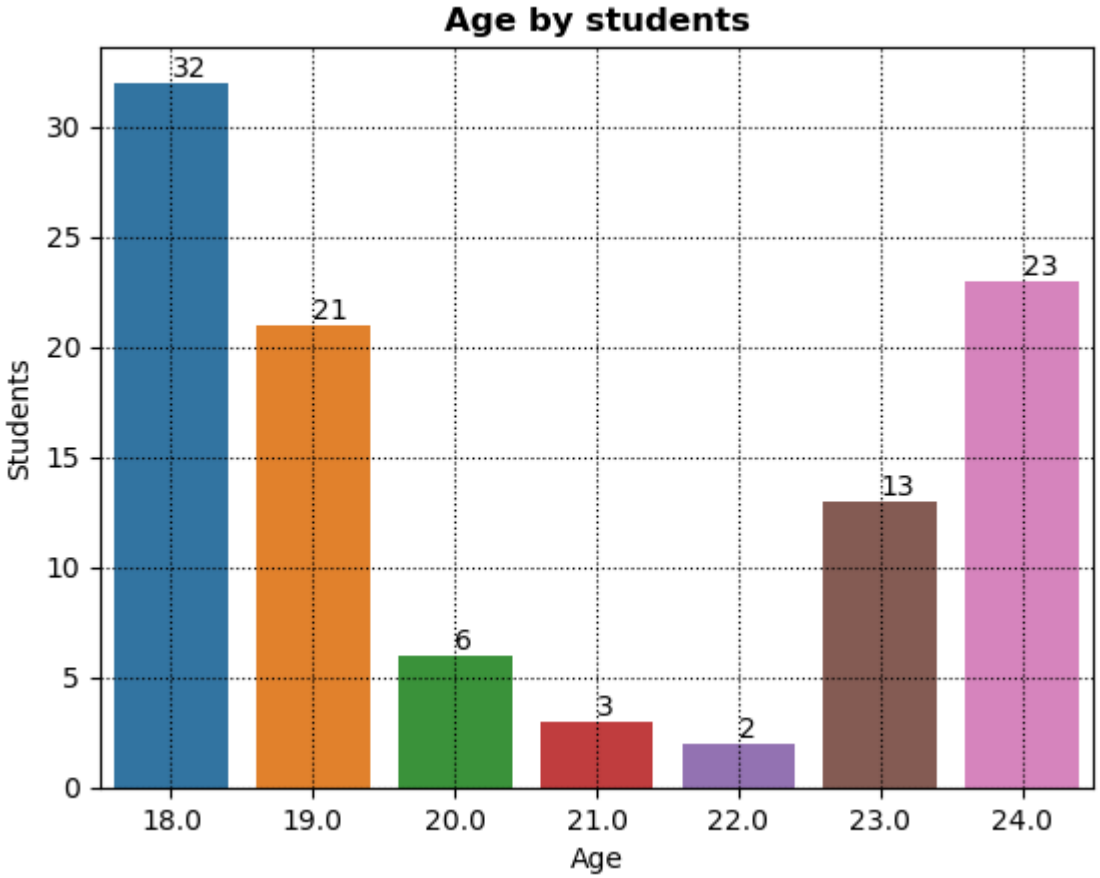


Age by student

In [23]:
```
SMH['Age'].value_counts()
```

Out[23]:
```
18.0    32
24.0    23
19.0    21
23.0    13
20.0     6
21.0     3
22.0     2
Name: Age, dtype: int64
```

In [24]:
```
sns.countplot(data=SMH, x='Age')
plt.xlabel('Age')
plt.ylabel('Students')
plt.title('Age by students', fontsize=12, fontweight='bold')
plt.grid(ls=':',c='k')
plt.text(0,32.3,'32')
plt.text(1,21.3,'21')
plt.text(2,6.3,'6')
plt.text(3,3.3,'3')
plt.text(4,2.3,'2')
plt.text(5,13.3,'13')
plt.text(6,23.3,'23')
plt.show()
```



Exploratory Data Analysis

Students Conditions

In [25]:
```
depressed_filter = SMH['Has Depression?'] == 'Yes'
anxiety_filter = SMH['Has Anxiety?'] == 'Yes'
panicking_filter = SMH['Has Panic Attacks?'] == 'Yes'
conditions_filter = anxiety_filter | depressed_filter | panicking_filter
no_conditions_filter = depressed_filter & anxiety_filter & panicking_filter
```

In [26]:
```
SMH['Has Condition?'] = np.where(conditions_filter, 'Yes',"No")
SMH
```

Out[26]:

	Timestamp	Gender	Age	Course	Year of Study	CGPA	Is Married?	Has Depression?	Has Anxiety?	Has Panic Attacks?	Is Undergoing Treatment?	Has Condition?
0	08-07-2020 12:02	Female	18.0	Engineering	1	3.00 - 3.49	No	Yes	No	Yes	No	Yes
1	08-07-2020 12:04	Male	21.0	Islamic education	2	3.00 - 3.49	No	No	Yes	No	No	Yes
2	08-07-2020 12:05	Male	19.0	BIT	1	3.00 - 3.49	No	Yes	Yes	Yes	No	Yes
3	08-07-2020 12:06	Female	22.0	Laws	3	3.00 - 3.49	Yes	Yes	No	No	No	Yes
4	08-07-2020 12:13	Male	23.0	Mathemathics	4	3.00 - 3.49	No	No	No	No	No	No
...	...	...	...	...	...	...	...	...	...	...	...	...
96	13-07-2020 19:56	Female	21.0	BCS	1	3.50 - 4.00	No	No	Yes	No	No	Yes
97	13-07-2020 21:21	Male	18.0	Engineering	2	3.00 - 3.49	No	Yes	Yes	No	No	Yes
98	13-07-2020 21:22	Female	19.0	Nursing	3	3.50 - 4.00	Yes	Yes	No	Yes	No	Yes
99	13-07-2020 21:23	Female	23.0	Pendidikan Islam	4	3.50 - 4.00	No	No	No	No	No	No
100	18-07-2020 20:16	Male	20.0	Biomedical science	2	3.00 - 3.49	No	No	No	No	No	No

100 rows × 12 columns

In [27]:
```
SMH['Condition'] = ''
SMH.loc[anxiety_filter, 'Condition'] += 'A'
SMH.loc[depressed_filter, 'Condition'] += 'D'
SMH.loc[panicking_filter, 'Condition'] += 'P'
SMH.loc[no_conditions_filter, 'Condition'] += 'None'
SMH
```

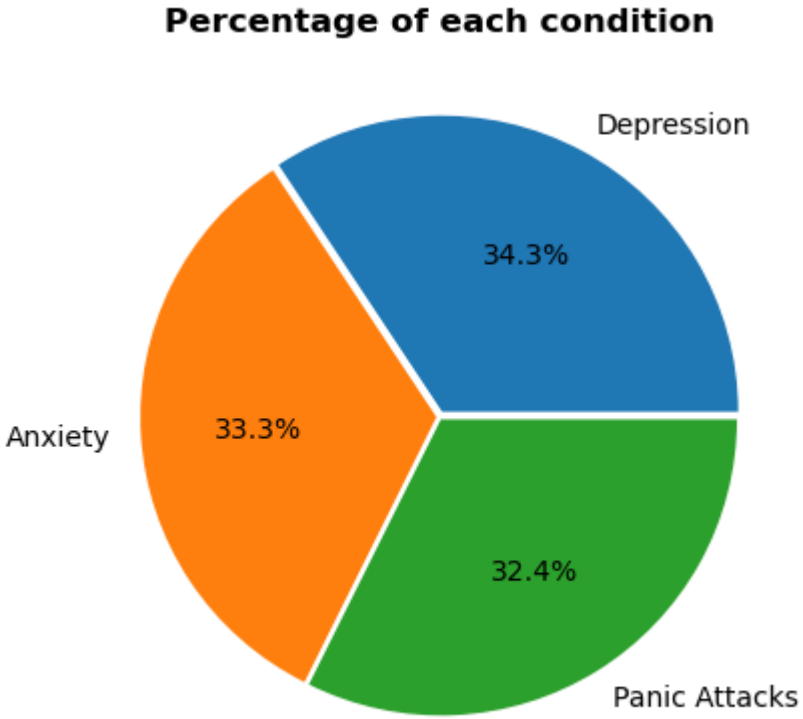
Out[27]:

	Timestamp	Gender	Age	Course	Year of Study	CGPA	Is Married?	Has Depression?	Has Anxiety?	Has Panic Attacks?	Is Undergoing Treatment?	Has Condition?	Condition
0	08-07-2020 12:02	Female	18.0	Engineering	1	3.00 - 3.49	No	Yes	No	Yes	No	Yes	DP
1	08-07-2020 12:04	Male	21.0	Islamic education	2	3.00 - 3.49	No	No	Yes	No	No	Yes	A
2	08-07-2020 12:05	Male	19.0	BIT	1	3.00 - 3.49	No	Yes	Yes	Yes	No	Yes	ADPNone
3	08-07-2020 12:06	Female	22.0	Laws	3	3.00 - 3.49	Yes	Yes	No	No	No	Yes	D
4	08-07-2020 12:13	Male	23.0	Mathemathics	4	3.00 - 3.49	No	No	No	No	No	No	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
96	13-07-2020 19:56	Female	21.0	BCS	1	3.50 - 4.00	No	No	Yes	No	No	Yes	A
97	13-07-2020 21:21	Male	18.0	Engineering	2	3.00 - 3.49	No	Yes	Yes	No	No	Yes	AD
98	13-07-2020 21:22	Female	19.0	Nursing	3	3.50 - 4.00	Yes	Yes	No	Yes	No	Yes	DP
99	13-07-2020 21:23	Female	23.0	Pendidikan Islam	4	3.50 - 4.00	No	No	No	No	No	No	
100	18-07-2020 20:16	Male	20.0	Biomedical science	2	3.00 - 3.49	No	No	No	No	No	No	

100 rows × 13 columns

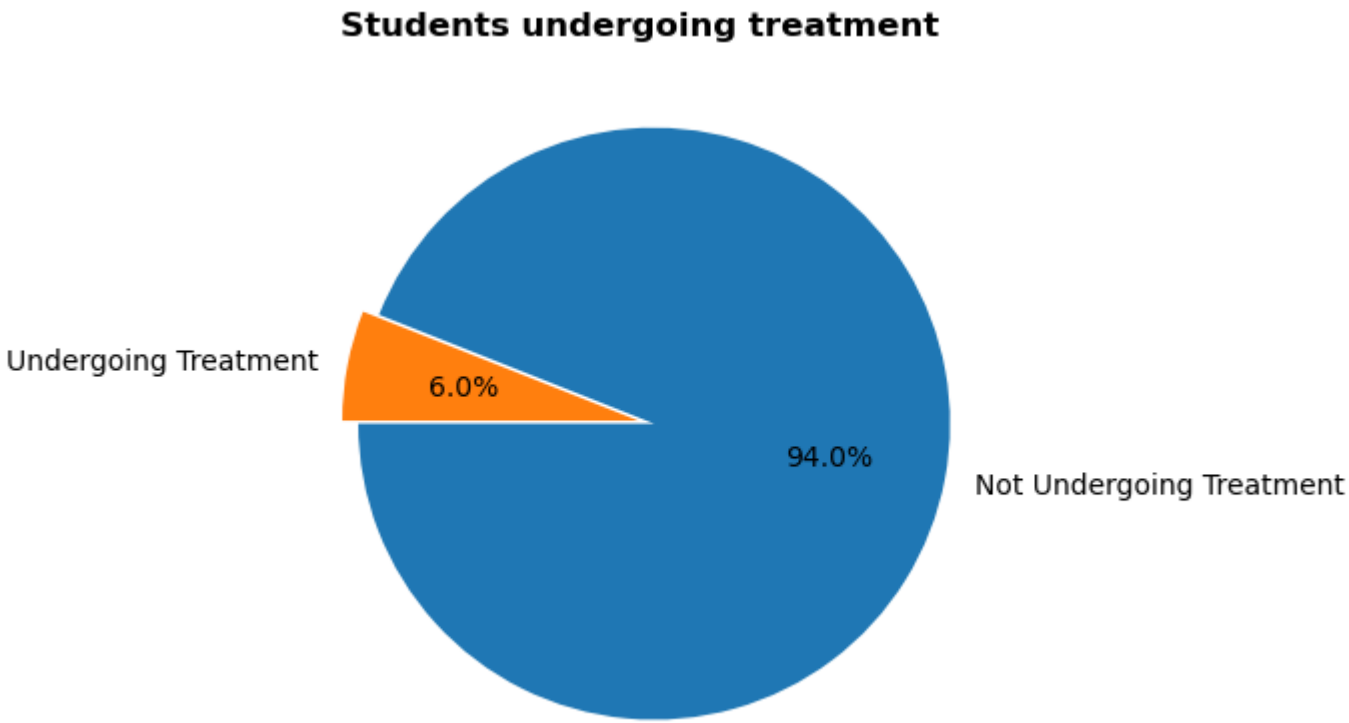
In [28]:
```
SHM = SMH[SMH['Has Condition?'] == 'Yes']
```

In [29]:
```
total_entries = len(SHM)
percentage_depression = (SHM['Has Depression?'].value_counts()['Yes'] / total_entries) * 100
percentage_anxiety = (SHM['Has Anxiety?'].value_counts()['Yes'] / total_entries) * 100
percentage_panic_attacks = (SHM['Has Panic Attacks?'].value_counts()['Yes'] / total_entries) * 100

sizes = [percentage_depression, percentage_anxiety, percentage_panic_attacks]

plt.pie(sizes, explode=(0.02, 0.01, 0.01), labels=['Depression', 'Anxiety', 'Panic Attacks'], autopct='%1.1f%%')

plt.title('Percentage of each condition',fontsize=12, fontweight='bold')

plt.show()
```



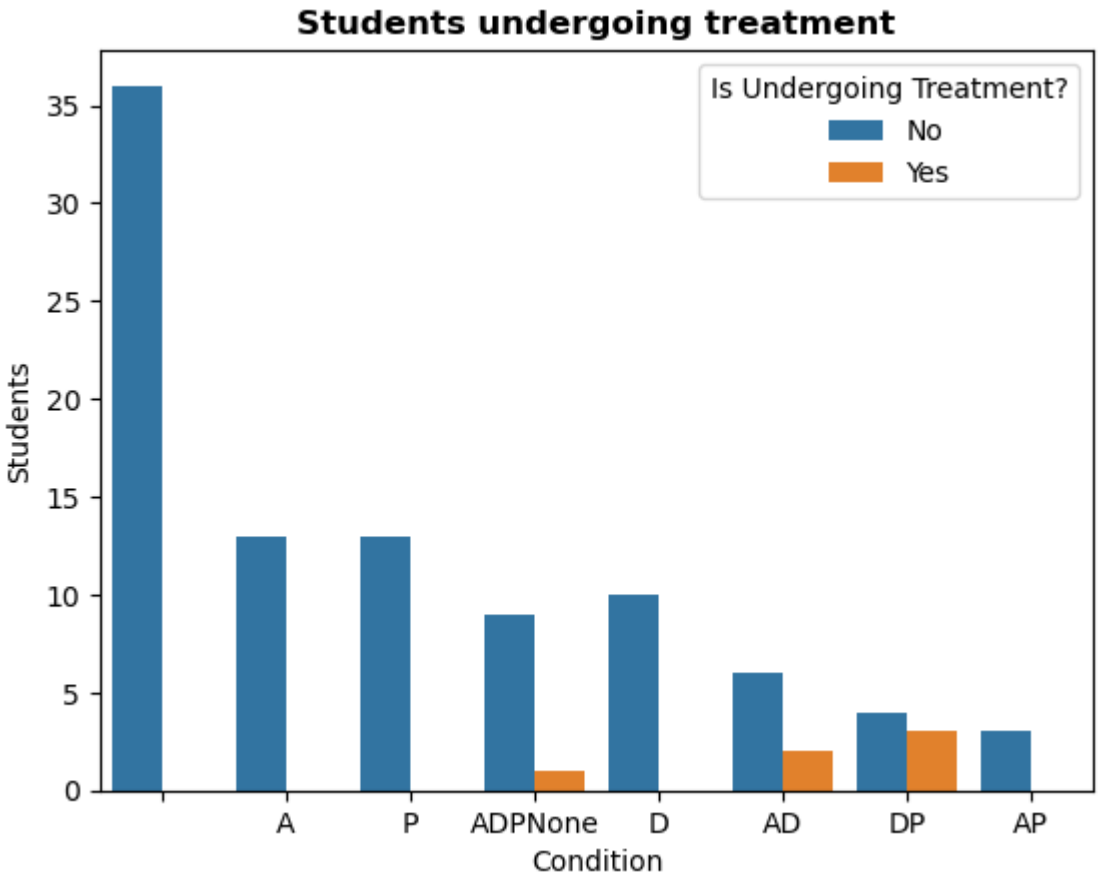
Under going treatment

In [30]:
```
labels = ['Not Undergoing Treatment', 'Undergoing Treatment']

plt.pie(SMH['Is Undergoing Treatment?'].value_counts(),labels=labels, autopct='%1.1f%%', explode=(0.05,0.005),startang
plt.title('Students undergoing treatment', fontsize=12, fontweight='bold')
plt.show()
```



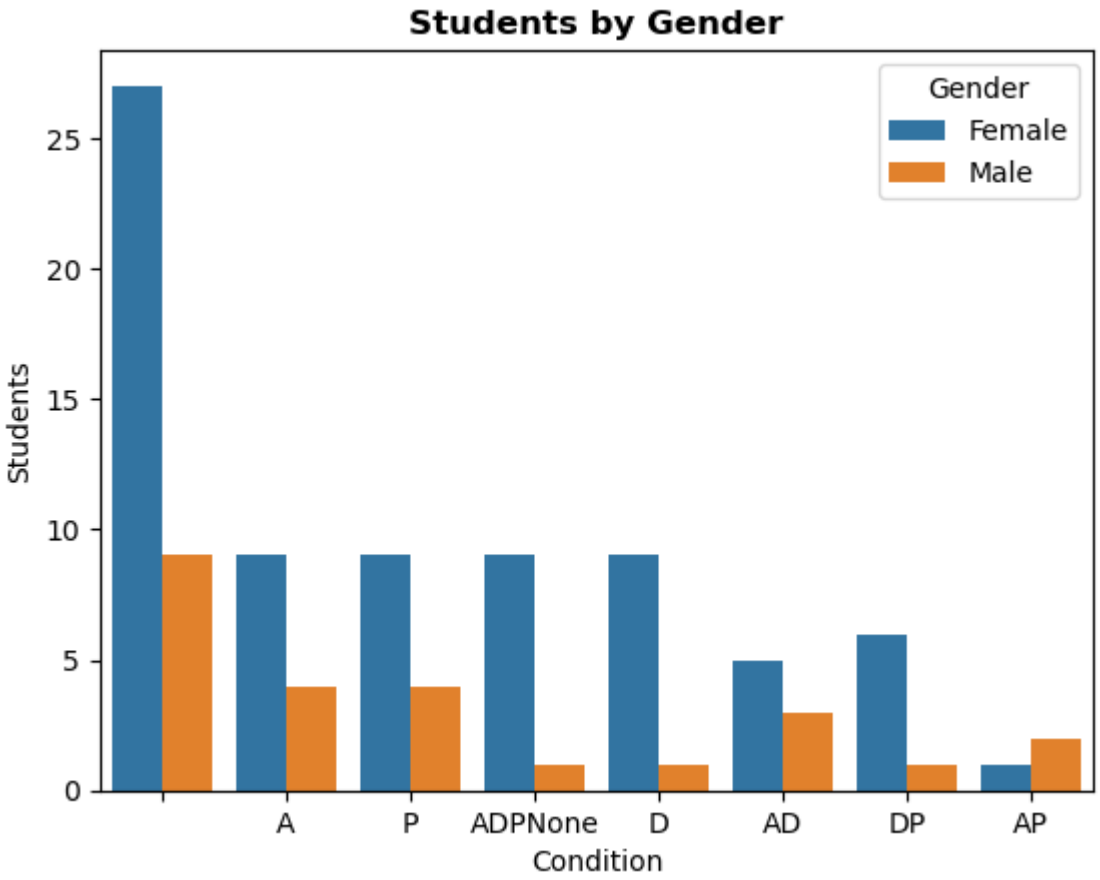
In [31]:
```
sort = SMH['Condition'].value_counts().index.tolist()
```

In [32]:
```
sns.countplot(data=SMH, x='Condition',hue='Is Undergoing Treatment?', order=sort)
plt.xlabel('Condition')
plt.ylabel('Students')
plt.title('Students undergoing treatment', fontsize=12, fontweight='bold')
plt.show()
```



Students by Gender

In [33]:
```
sns.countplot(data=SMH, x='Condition',hue='Gender', order=sort)
plt.xlabel('Condition')
plt.ylabel('Students')
plt.title('Students by Gender', fontsize=12, fontweight='bold')
plt.show()
```



----------End----------

In [ ]: