

In [ ]:

*#Importing the libraries which are mostly used for data analysis and data visualization.*  
**import** numpy **as** np  
**import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
**import** seaborn **as** sb  
  
*#Importing warning to ignore them.*  
**import** warnings  
warnings**.**filterwarnings('ignore')

In [ ]:

*#Importing CSV file*  
data**=**pd**.**read\_csv('Student Mental health.csv')

Basic and Advanced EDA commands[¶](#gjdgxs)

In [22]:

*#Dipslay top two records*  
print("The top two records:\n",data**.**head(2))  
*#Display bottom two records*  
print("The bottom two records:\n",data**.**tail(2))

The top two records:  
 Choose your gender Age What is your course? Your current year of Study \  
0 0.0 18.0 Engineering year 1   
1 1.0 21.0 Islamic education year 2   
  
 What is your CGPA? Marital status Do you have Depression? \  
0 3.00 - 3.49 No 1.0   
1 3.00 - 3.49 No 0.0   
  
 Do you have Anxiety? Do you have Panic attack? \  
0 0.0 1.0   
1 1.0 0.0   
  
 Did you seek any specialist for a treatment?   
0 No   
1 No   
The bottom two records:  
 Choose your gender Age What is your course? Your current year of Study \  
101 NaN NaN NaN NaN   
102 NaN NaN NaN NaN   
  
 What is your CGPA? Marital status Do you have Depression? \  
101 NaN NaN NaN   
102 NaN NaN NaN   
  
 Do you have Anxiety? Do you have Panic attack? \  
101 NaN NaN   
102 NaN NaN   
  
 Did you seek any specialist for a treatment?   
101 NaN   
102 NaN

In [21]:

*#To list down columns*  
print("The columns of datasets are:\n\n",data**.**columns,"\n\n")  
*#tolist also list down the column*  
print("The columns of datasets are:\n\n",data**.**columns**.**values**.**tolist(),"\n\n")  
*#iloc() and loc() are used for sclicing purpose*  
iloc**=**data**.**iloc[:,[0,2]]  
print("Accesing data:\n\n",iloc,"\n\n")  
loc**=**data**.**loc[0:2,['Choose your gender','Age']]  
print("Accesing data:\n\n",loc)

The columns of datasets are:  
  
 Index(['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')   
  
  
The columns of datasets are:  
  
 ['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?']   
  
  
Accesing data:  
  
 Choose your gender What is your course?  
0 0.0 Engineering  
1 1.0 Islamic education  
2 1.0 BIT  
3 0.0 Laws  
4 1.0 Mathemathics  
.. ... ...  
98 0.0 Nursing   
99 0.0 Pendidikan Islam  
100 1.0 Biomedical science  
101 NaN NaN  
102 NaN NaN  
  
[103 rows x 2 columns]   
  
  
Accesing data:  
  
 Choose your gender Age  
0 0.0 18.0  
1 1.0 21.0  
2 1.0 19.0

In [20]:

*#Checking null values and print their total along woth columns*  
print("Null values with the sum:\n\n",data**.**isnull()**.**sum())

Null values with the sum:  
  
 Choose your gender 2  
Age 3  
What is your course? 2  
Your current year of Study 2  
What is your CGPA? 2  
Marital status 2  
Do you have Depression? 2  
Do you have Anxiety? 2  
Do you have Panic attack? 2  
Did you seek any specialist for a treatment? 2  
dtype: int64

In [7]:

*#Info() is used to et know about number of enteries.number of columns,data type and memory space it acquired*  
print(data**.**info())

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 103 entries, 0 to 102  
Data columns (total 10 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Choose your gender 101 non-null object   
 1 Age 100 non-null float64  
 2 What is your course? 101 non-null object   
 3 Your current year of Study 101 non-null object   
 4 What is your CGPA? 101 non-null object   
 5 Marital status 101 non-null object   
 6 Do you have Depression? 101 non-null object   
 7 Do you have Anxiety? 101 non-null object   
 8 Do you have Panic attack? 101 non-null object   
 9 Did you seek any specialist for a treatment? 101 non-null object   
dtypes: float64(1), object(9)  
memory usage: 8.2+ KB  
None

In [8]:

*#Describe() is used to get statistical analysis and it only deal with numeric data*  
print("Statistical analysis:\n\n",data**.**describe())

Statistical analysis:  
  
 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 [9]:

*#Returns the count of unqiue values in column*  
print("The counts of unqiue value:\n\n",data**.**value\_counts())

The counts of unqiue value:  
  
 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?  
Female 18.0 BCS year 1 3.00 - 3.49 No Yes No No No 1  
 24.0 Biomedical science year 1 3.00 - 3.49 No No No No No 1  
 Kop year 4 3.00 - 3.49 No No Yes No No 1  
 KOE year 1 3.50 - 4.00 No No Yes Yes No 1  
 Fiqh Year 3 0 - 1.99 No No No Yes No 1  
 ..  
 19.0 Biotechnology Year 3 0 - 1.99 No No No No No 1  
 Biomedical science year 3 3.00 - 3.49 No No No No No 1  
 BIT year 1 3.00 - 3.49 No Yes Yes Yes No 1  
 2.50 - 2.99 No Yes Yes Yes No 1  
Male 24.0 Engineering Year 2 2.00 - 2.49 No No No Yes No 1  
Length: 100, dtype: int64

In [10]:

*#Sample()-returns the random sample of your data*  
print("Random row:\n",data**.**sample())

Random row:  
 Choose your gender Age What is your course? Your current year of Study \  
77 Female 18.0 Human Sciences Year 2   
  
 What is your CGPA? Marital status Do you have Depression? \  
77 3.00 - 3.49 No No   
  
 Do you have Anxiety? Do you have Panic attack? \  
77 No Yes   
  
 Did you seek any specialist for a treatment?   
77 No

In [11]:

*#For coverting catagorical data into numeric data*   
print(data['Do you have Depression?']**.**replace(['No','Yes'],[0,1],inplace**=True**))  
print(data['Do you have Anxiety?']**.**replace(['No','Yes'],[0,1],inplace**=True**))  
print(data['Do you have Panic attack?']**.**replace(['No','Yes'],[0,1],inplace**=True**))  
print(data['Choose your gender']**.**replace(['Female','Male'],[0,1],inplace**=True**))

None  
None  
None  
None

In [12]:

print(data**.**info())

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 103 entries, 0 to 102  
Data columns (total 10 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Choose your gender 101 non-null float64  
 1 Age 100 non-null float64  
 2 What is your course? 101 non-null object   
 3 Your current year of Study 101 non-null object   
 4 What is your CGPA? 101 non-null object   
 5 Marital status 101 non-null object   
 6 Do you have Depression? 101 non-null float64  
 7 Do you have Anxiety? 101 non-null float64  
 8 Do you have Panic attack? 101 non-null float64  
 9 Did you seek any specialist for a treatment? 101 non-null object   
dtypes: float64(5), object(5)  
memory usage: 8.2+ KB  
None

In [13]:

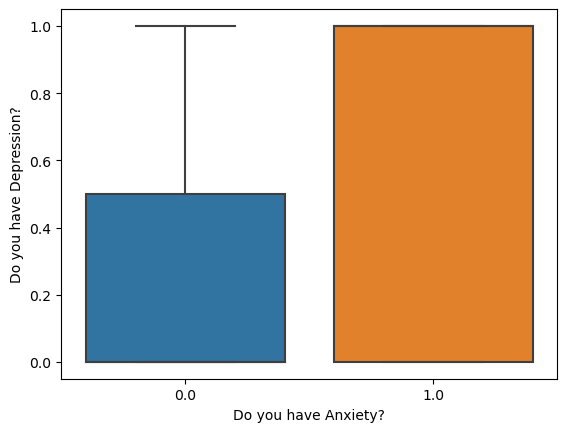
*#It retunrs the duplicated value*  
print("The duplicated values are:\n",data[data**.**duplicated()])

The duplicated values are:  
 Choose your gender Age What is your course? Your current year of Study \  
102 NaN NaN NaN NaN   
  
 What is your CGPA? Marital status Do you have Depression? \  
102 NaN NaN NaN   
  
 Do you have Anxiety? Do you have Panic attack? \  
102 NaN NaN   
  
 Did you seek any specialist for a treatment?   
102 NaN

Visualization of data[¶](#30j0zll)

In [14]:

*#Box plot:Used to visualize the distribution of a set of data values.*  
sb**.**boxplot(x**=**'Do you have Anxiety?',y**=**'Do you have Depression?',data**=**data)  
plt**.**show()



Obersvations:[¶](#1fob9te)

* Students who have depression they also have anxiety.

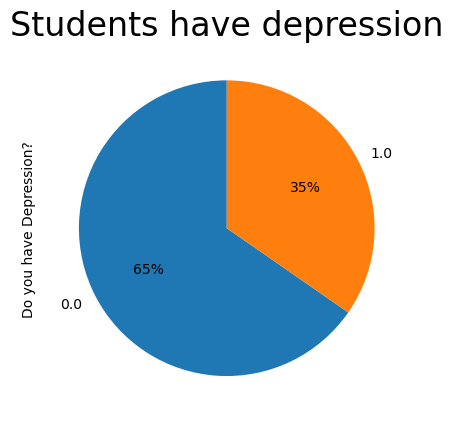
Univariate Analysis[¶](#3znysh7)

* This type of data consists of only one variable.

In [15]:

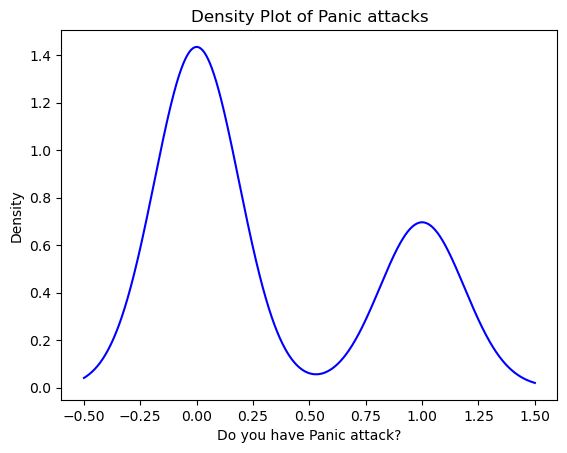
*#Pie charts-It is used to visualize the distribution of categorical data.*  
counts0**=**data['Do you have Depression?']**.**value\_counts()  
plt**.**title('Students have depression',size**=**24)  
print(counts0**.**plot(kind**=**'pie',autopct**=lambda** x:f'{x:.0f}%',startangle**=**90))  
plt**.**show()

Axes(0.22375,0.11;0.5775x0.77)



In [16]:

*# Create a density plot (KDE) to visualize the distribution of the 'skewed\_variable'*  
data['Do you have Panic attack?']**.**plot(kind**=**'density', color**=**'b')  
plt**.**xlabel('Do you have Panic attack?')  
plt**.**ylabel('Density')  
plt**.**title('Density Plot of Panic attacks')  
plt**.**show()



Obersvations:[¶](#1fob9te)

* Students who have depression they also have anxiety.
* 35% students have anxiety and depression.

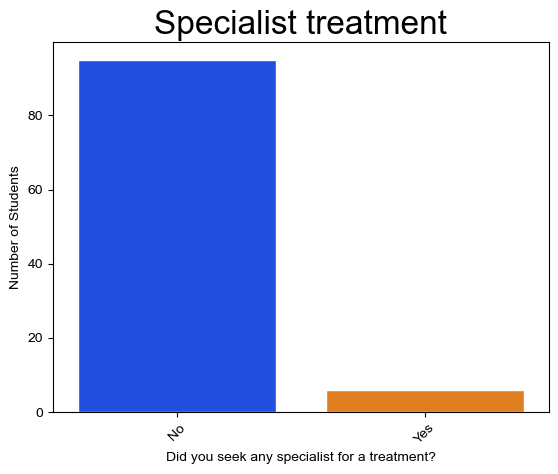
Bivariate analysis:[¶](#2et92p0)

* It is a statistical method examining how two different things are related.

In [17]:

*#Bar plot:It is uesd for comparin the frequency of different categories*  
counts**=**data['Did you seek any specialist for a treatment?']**.**value\_counts()  
print(counts**.**plot(kind**=**'bar'))  
sb**.**set\_style('whitegrid')  
ax **=** sb**.**barplot(x**=**counts**.**index, y**=**counts**.**values, palette**=**'bright')  
ax**.**set\_xlabel('Did you seek any specialist for a treatment?')  
plt**.**xticks(rotation**=**45)  
ax**.**set\_ylabel('Number of Students')  
plt**.**title('Specialist treatment',size**=**24)  
plt**.**show()

Axes(0.125,0.11;0.775x0.77)

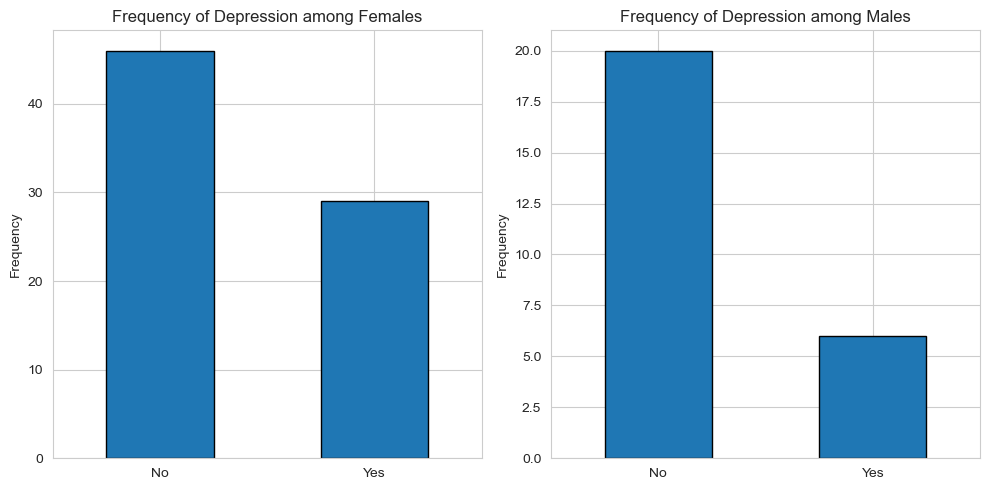


Observation[¶](#tyjcwt)

* It is observe that students who is taking specialist treatment is less compared to students suffering for depression.

In [18]:

Female **=** data[data['Choose your gender'] **==** 0]  
Male **=** data[data['Choose your gender'] **==** 1]  
  
*# Calculate the frequency of depression (1 for Yes, 0 for No) for each gender*  
c1 **=** Female['Do you have Depression?']**.**value\_counts()  
c2 **=** Male['Do you have Depression?']**.**value\_counts()  
  
*# Create the bar plot*  
fig, ax **=** plt**.**subplots(nrows**=**1, ncols**=**2, figsize**=**(10, 5))  
c1**.**plot(kind**=**'bar', ax**=**ax[0],edgecolor**=**'Black')  
c2**.**plot(kind**=**'bar', ax**=**ax[1],edgecolor**=**'Black')  
  
*# Add labels and title*  
ax[0]**.**set\_title('Frequency of Depression among Females')  
ax[0]**.**set\_xticklabels(['No', 'Yes'], rotation**=**0)  
ax[0]**.**set\_ylabel('Frequency')  
ax[1]**.**set\_title('Frequency of Depression among Males')  
ax[1]**.**set\_xticklabels(['No', 'Yes'], rotation**=**0)  
ax[1]**.**set\_ylabel('Frequency')  
  
plt**.**tight\_layout()  
plt**.**show()

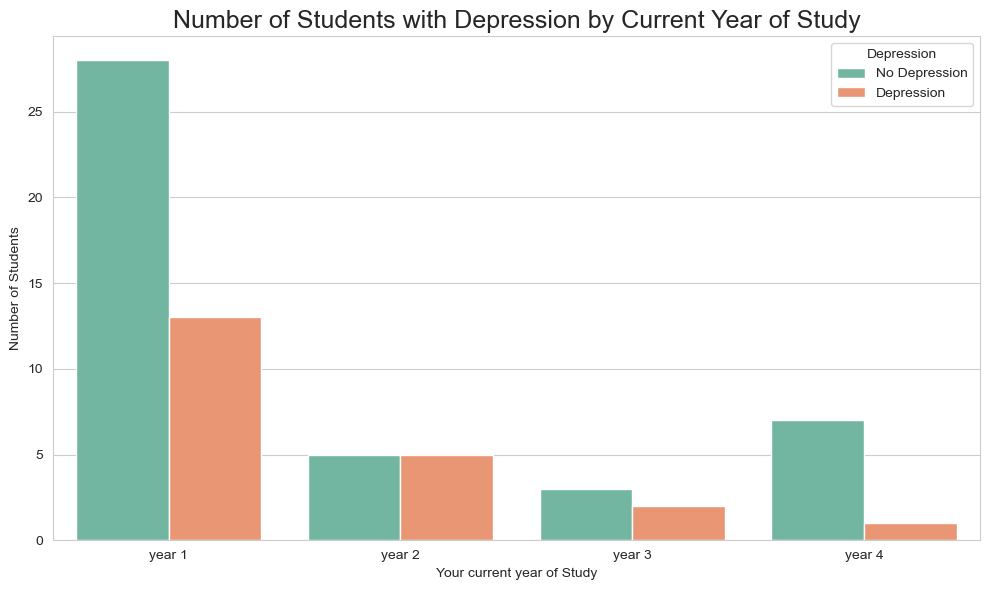


Observation[¶](#tyjcwt)

* Female count compared to male is more.

In [25]:

*# Assuming you have a pandas DataFrame called 'data' with columns 'Do you have Depression?' and 'Your current year of Study'*  
*# Set the order of the x-axis categories (optional, adjust as needed)*  
year\_order **=** ['year 1', 'year 2', 'year 3', 'year 4']  
plt**.**figure(figsize**=**(10, 6))  
  
*# Create the count plot with specified order and rotated x-axis labels*  
sb**.**countplot(data**=**data, x**=**'Your current year of Study', hue**=**'Do you have Depression?', palette**=**'Set2', order**=**year\_order)  
plt**.**xlabel('Your current year of Study')  
plt**.**ylabel('Number of Students')  
plt**.**title('Number of Students with Depression by Current Year of Study', size**=**18)  
  
*# Calculate and display percentages for each category*  
total\_students\_per\_year **=** data['Your current year of Study']**.**value\_counts()  
**for** i, year **in** enumerate(year\_order):  
 count\_no\_depression **=** data[(data['Your current year of Study'] **==** year) **&** (data['Do you have Depression?'] **==** 'No')]**.**shape[0]  
 count\_depression **=** data[(data['Your current year of Study'] **==** year) **&** (data['Do you have Depression?'] **==** 'Yes')]**.**shape[0]  
 total\_students **=** total\_students\_per\_year[year]  
plt**.**legend(title**=**'Depression', labels**=**['No Depression', 'Depression'])  
plt**.**tight\_layout() *# To prevent label cutoff*  
plt**.**show()



Multivariate analysis[¶](#3dy6vkm)

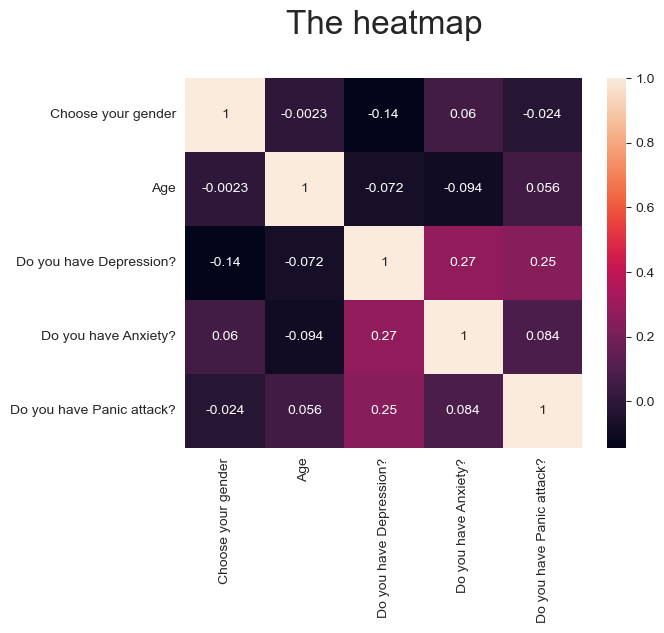
It involves evaluating multiple variables (more than two) to identify any possible association among them

In [26]:

*#Heatmap:It uses for visual representation of data with different colors.It is uesd to visualize correlation between variables.*  
*#It lies between -1 to +1*  
sb**.**heatmap(data**.**corr(),annot**=True**)  
plt**.**title('The heatmap\n',size**=**24)

Out[26]:

Text(0.5, 1.0, 'The heatmap\n')



Conclusion:[¶](#1t3h5sf)

* We observe that student of first year is suffering more it may be because of new environment,adaptibility,etc.
* We observe that female suffers more than male.
* We observe that students who have depression they also have anxiety and panic attacks.
* We observe that students consulting doctor is less the reason ma be that they are not aware about it.