MDA 620

Capstone 2

“Student Mental Health Analysis”

Sergio Fernandez

Contents

Introduction:

The importance of mental health in college students cannot be stressed enough and is often underestimated. As an international student, I had to leave all my family in my country and start, in a way, a new life in a foreign country and without knowing the language very well. This added to all the subjects that I had and being on a tennis team caused on me emotional and mental stress. Students of different grades, ages, and years can experience a mental breakdown at some point in their lives. Some aspects are above all: depression, anxiety, and even panic attacks.

Problem Analysis:

In many cases the mental health of students is not treated with the importance it should be and, in this project, we will see how age and gender can affect mental health problems. In this analysis, we are going to look at different university students of different ages and different majors.

Objective of the project:

- Filter, understand, and explore the dataset
- Create a prediction model
- Evaluate the different models and choose the best one for prediction

Data Manipulation and Data Cleaning:

	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	8/7/2020 12:02	Female	18.0	Engineering	year 1	3.00 - 3.49	No	Yes	No	Yes	No
1	8/7/2020 12:04	Male	21.0	Islamic education	year 2	3.00 - 3.49	No	No	Yes	No	No
2	8/7/2020 12:05	Male	19.0	BIT	Year 1	3.00 - 3.49	No	Yes	Yes	Yes	No

First, we are going to import all the necessary libraries for the project and read the dataset to be able to analyze it. Thanks to the info() command we see that there are some values missing.

New Column Names

We can see that some columns contain more information than necessary so I'm going to rename some of them.

	Datetime	Gender	Age	Major	Year	GPA	Marital_Status	Depression	Anxiety	Panic_Attack	Treatment
0	8/7/2020 12:02	Female	18.0	Engineering	year 1	3.00 - 3.49	No	Yes	No	Yes	No
1	8/7/2020 12:04	Male	21.0	Islamic education	year 2	3.00 - 3.49	No	No	Yes	No	No

Fix of Major Names

We can observe that there are 49 majors in total. This is a false number since some courses are interpreted differently even though they are the same. So, I'm going to organize them and the total of majors at the end is 37.

```
data['Major'].replace({'engin': 'Engineering' , 'Engine':'Engineering' , 'Islamic education':'Islamic Education' ,
'Pendidikan islam':'Pendidikan Islam' , 'BIT':'IT', 'psychology':'Psychology',
'koe': 'Koe', 'Kirkhs': 'Irkhs', 'KIRKHS': 'Irkhs', 'Benl': 'BENL', 'Fiqh fatwa ': 'Fiqh', 'Laws': 'Law'} , inplace = True)
len(data['Major'].unique())
```

```
37
```

Fix of “Year” Column

We see that there are 4 years, I'm going to remove the word "year" since we don't need it since the name of the column is "Year".

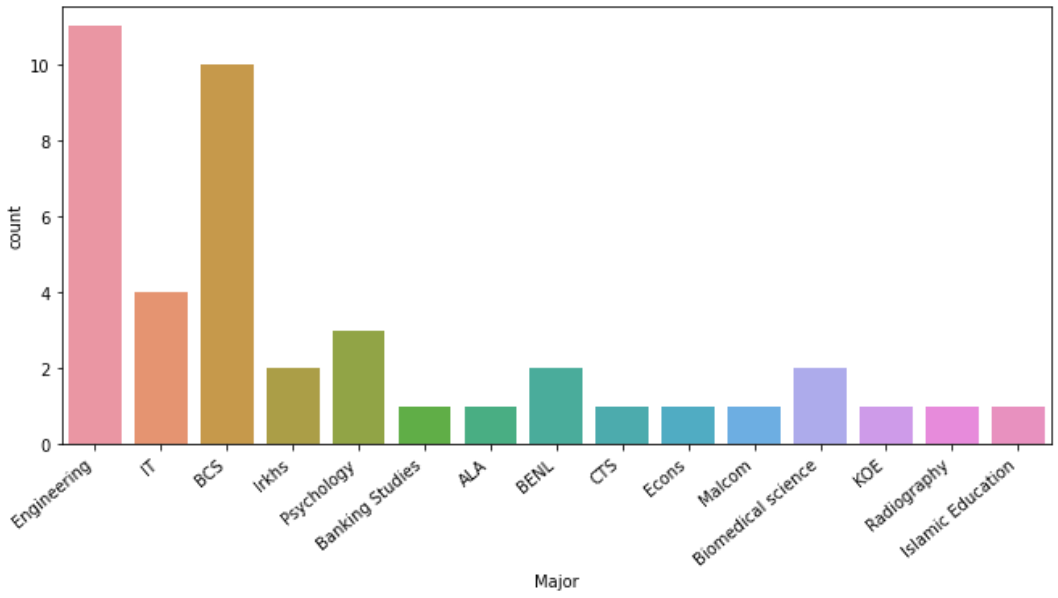
	Datetime	Gender	Age	Major	Year	GPA	Marital_Status	Depression	Anxiety	Panic_Attack	Treatment
0	8/7/2020 12:02	Female	18.0	Engineering	1	3.00 - 3.49	No	Yes	No	Yes	No
1	8/7/2020 12:04	Male	21.0	Islamic education	2	3.00 - 3.49	No	No	Yes	No	No
2	8/7/2020 12:05	Male	19.0	BIT	1	3.00 - 3.49	No	Yes	Yes	Yes	No
3	8/7/2020 12:06	Female	22.0	Laws	3	3.00 - 3.49	Yes	Yes	No	No	No
4	8/7/2020 12:13	Male	23.0	Mathemathics	4	3.00 - 3.49	No	No	No	No	No

Data Exploration and Data Visualization:

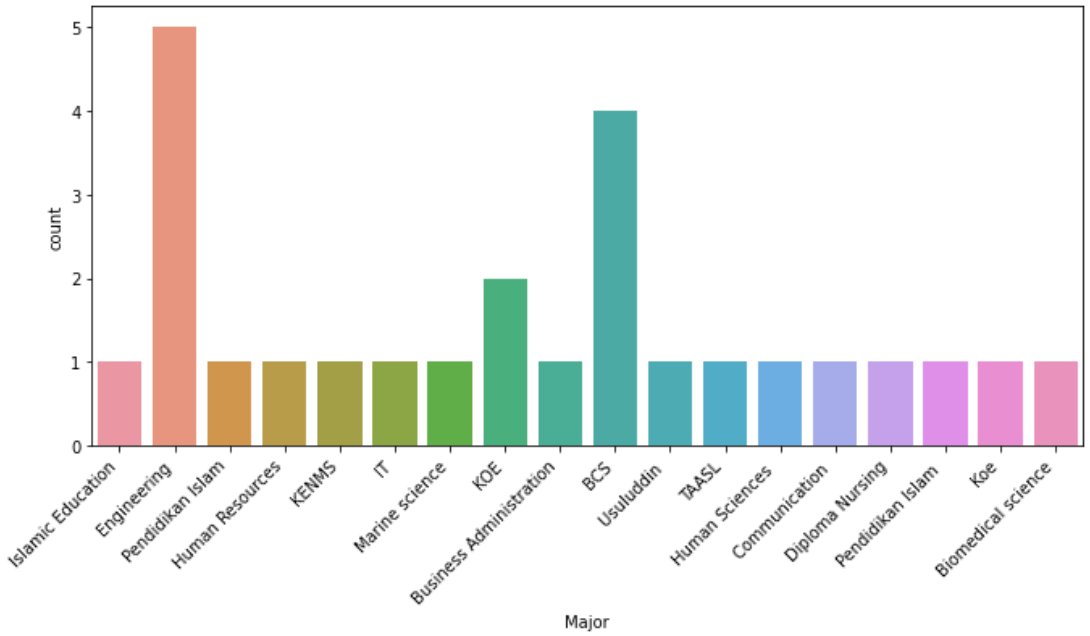
Count of Students by Major

I will be counting how many students are in the different majors during every year.

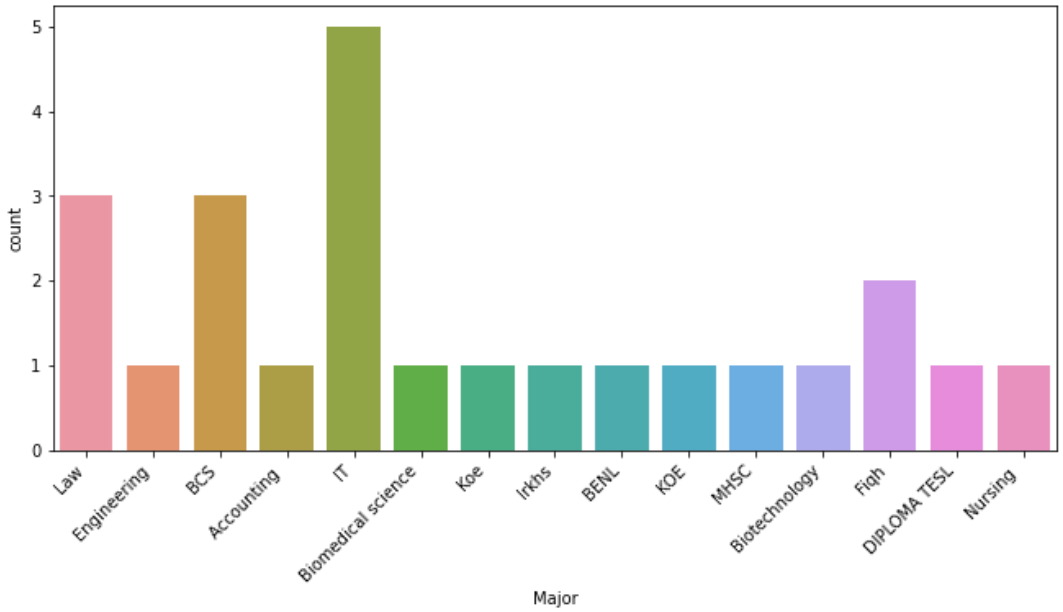
YEAR 1



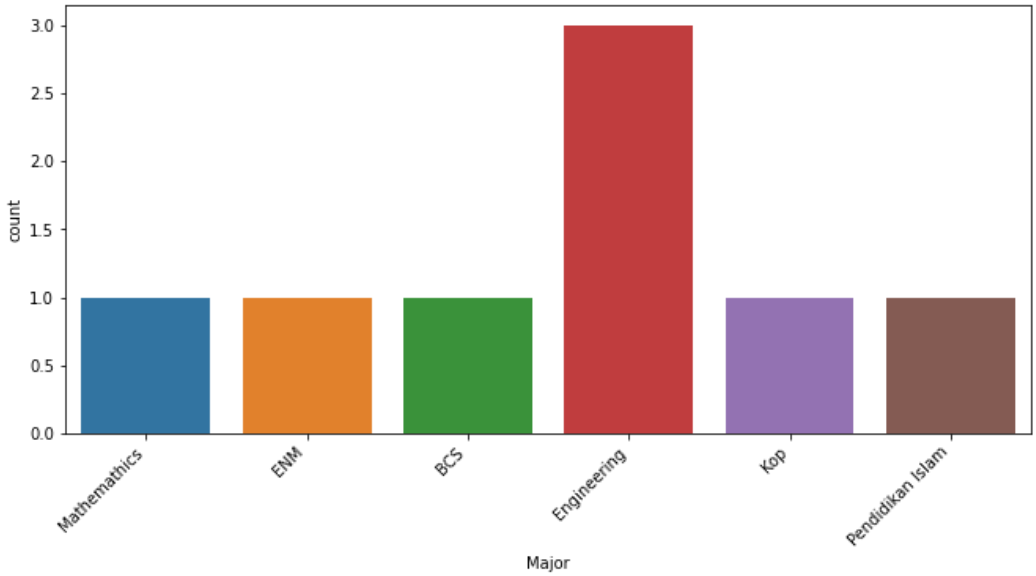
YEAR 2


count
0
1
2
3
4
5
Islamic Education
Engineering
Pendidikan Islam
Human Resources
KENMS
IT
Marine science
KOE
Business Administration
BCS
Usuluddin
TAASL
Human Sciences
Communication
Diploma Nursing
Pendidikan Islam
Koe
Biomedical science
Major

YEAR 3


count
0
1
2
3
4
5
Law
Engineering
BCS
Accounting
IT
Biomedical science
Koe
Irkhs
BENL
KOE
MHSC
Biotechnology
Fiqh
DIPLOMA TESL
Nursing
Major

YEAR 4



We can observe that almost in every year 'Engineering' course is the most active followed by 'BCS' and 'IT'

We can also observe that there are some courses missing in year 4

Percentage of Depression, Anxiety, and Panic Attack

1. Depression: 35%
2. Anxiety: 34%
3. Panic Attack: 33%

```
#Percentage of depression
facing_deppression = data['Depression'][data['Depression']=='Yes'].count()

dep_percent= (facing_deppression/data.shape[0])*100
dep_percent
```

```
35.0
```

```
#percentage of anxiety
facing_anxiety = data['Anxiety'][data['Anxiety']=='Yes'].count()

anx_percent= (facing_anxiety/data.shape[0])*100
anx_percent
```

```
34.0
```

```
#percentage of panic attack
facing_panic_attack = data['Panic_Attack'][data['Panic_Attack']=='Yes'].count()

pa_percent= (facing_panic_attack/data.shape[0])*100
pa_percent
```

```
33.0
```

People with the three mental problems.

Now I will create a data frame with the people that have the three problems.

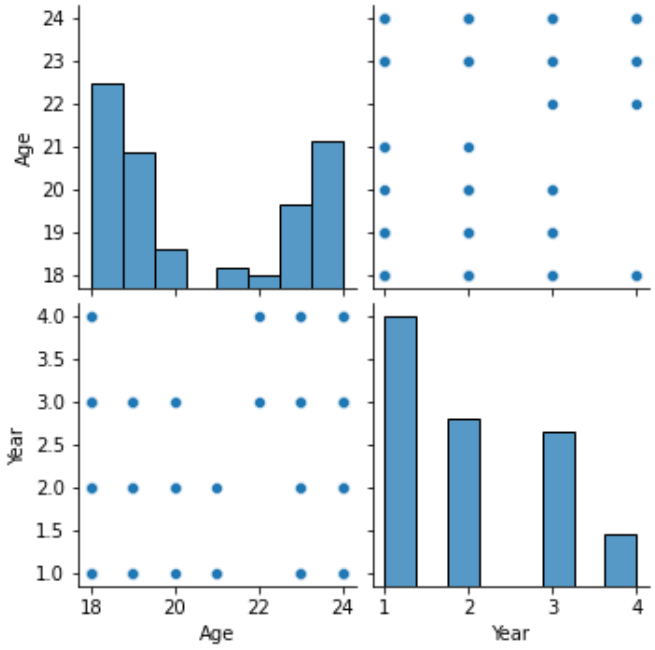
	Datetime	Gender	Age	Major	Year	GPA	Marital_Status	Depression	Anxiety	Panic_Attack	Treatment
2	8/7/2020 12:05	Male	19.0	IT	1	3.00 - 3.49	No	Yes	Yes	Yes	No
17	8/7/2020 12:52	Female	24.0	ENM	4	3.00 - 3.49	Yes	Yes	Yes	Yes	No
19	8/7/2020 13:07	Female	18.0	Marine science	2	3.50 - 4.00	Yes	Yes	Yes	Yes	No
24	8/7/2020 13:17	Female	23.0	BCS	3	3.50 - 4.00	No	Yes	Yes	Yes	No
28	8/7/2020 13:58	Female	24.0	IT	3	3.50 - 4.00	Yes	Yes	Yes	Yes	Yes
34	8/7/2020 14:41	Female	19.0	IT	1	3.00 - 3.49	No	Yes	Yes	Yes	No
37	8/7/2020 14:45	Female	19.0	IT	1	2.50 - 2.99	No	Yes	Yes	Yes	No
53	8/7/2020 15:48	Female	20.0	Koe	3	3.00 - 3.49	Yes	Yes	Yes	Yes	No
80	13/07/2020 10:11:26	Female	24.0	Communication	2	3.50 - 4.00	Yes	Yes	Yes	Yes	No
87	13/07/2020 11:46:13	Female	18.0	Psychology	1	3.50 - 4.00	No	Yes	Yes	Yes	No

We can observe that there are 10 people that have the three problems. 9 of them are women and 4/10 have IT as major. And only one of them is receiving treatment.

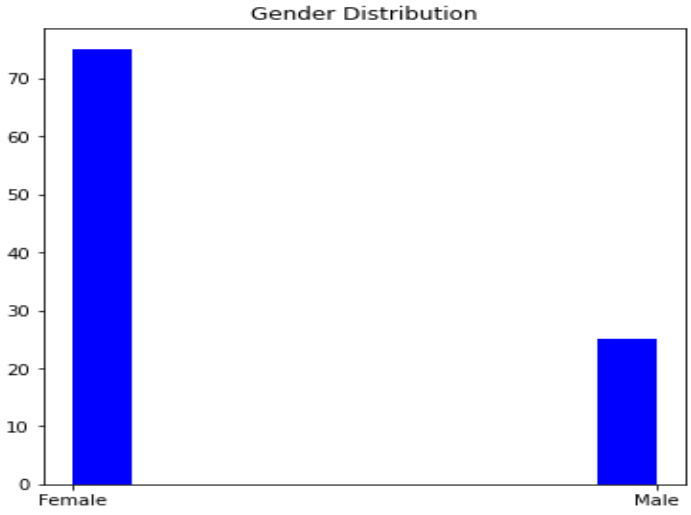
And we can also see that 94/100 people do not receive treatment.

Outliers.

We can see how there are not outliers.

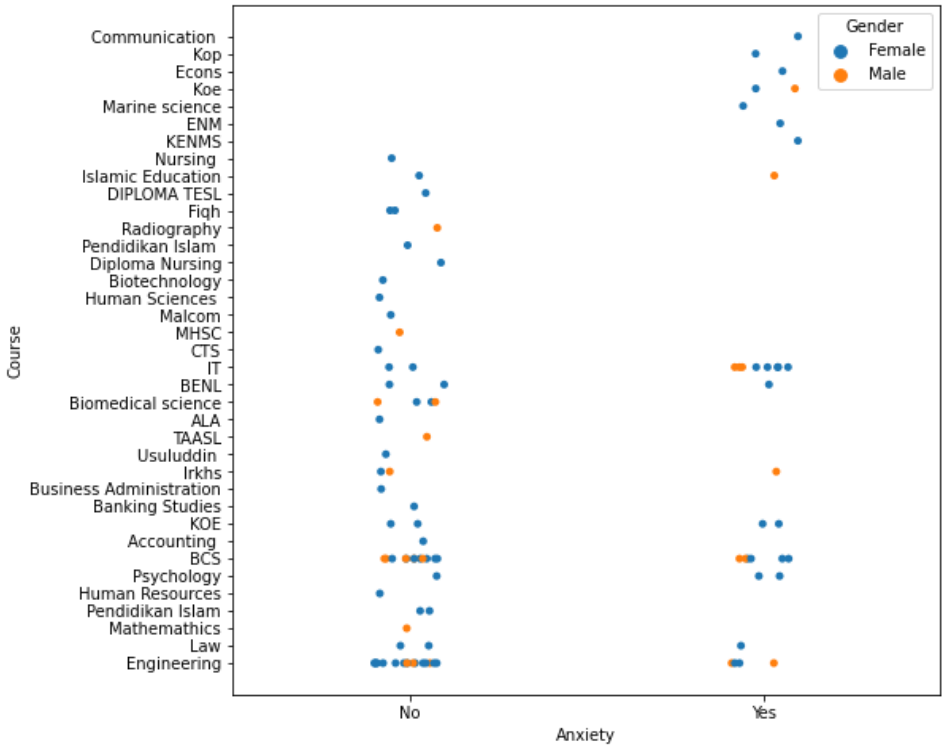


Gender Distribution



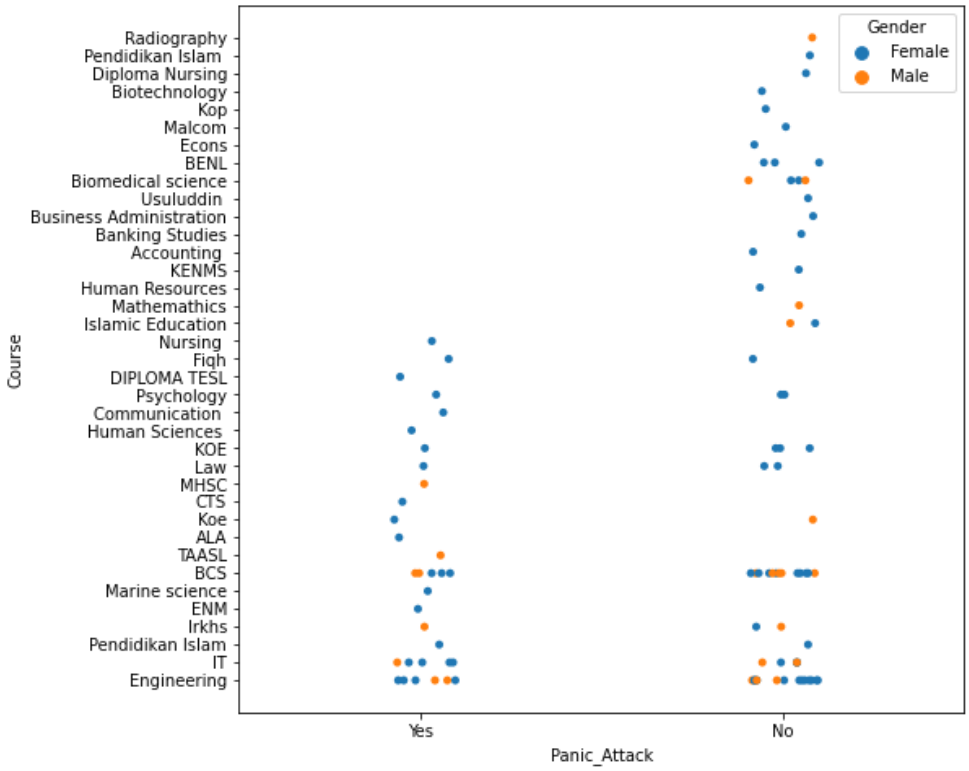
We can observe thank the majority are females.

Anxiety vs Gender and Major



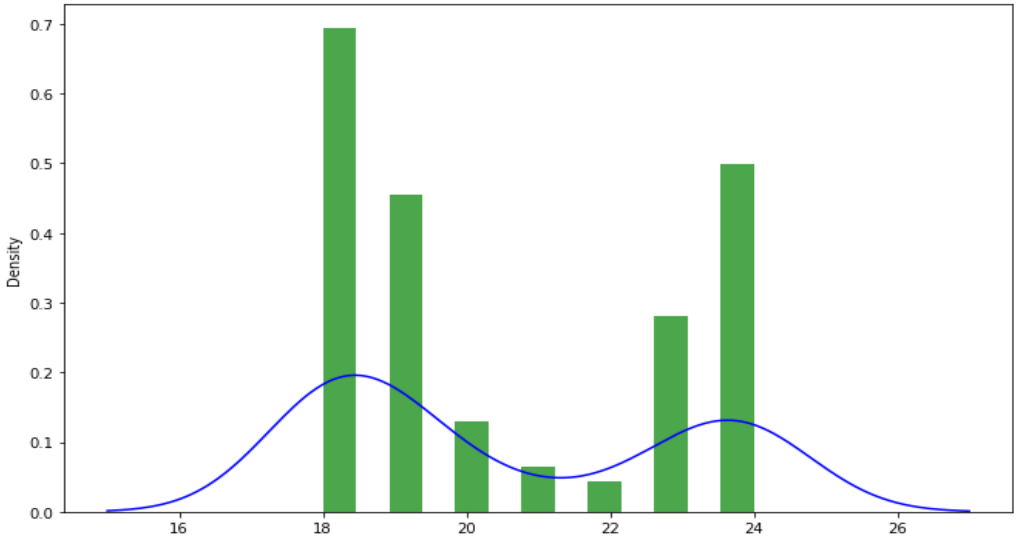
- Males are less prone to experience depression as compared to females.
- 2/3 females in Psychology experience depression.
- Around 50% of the students in IT experience depression.

Panic Attack vs Gender and Major

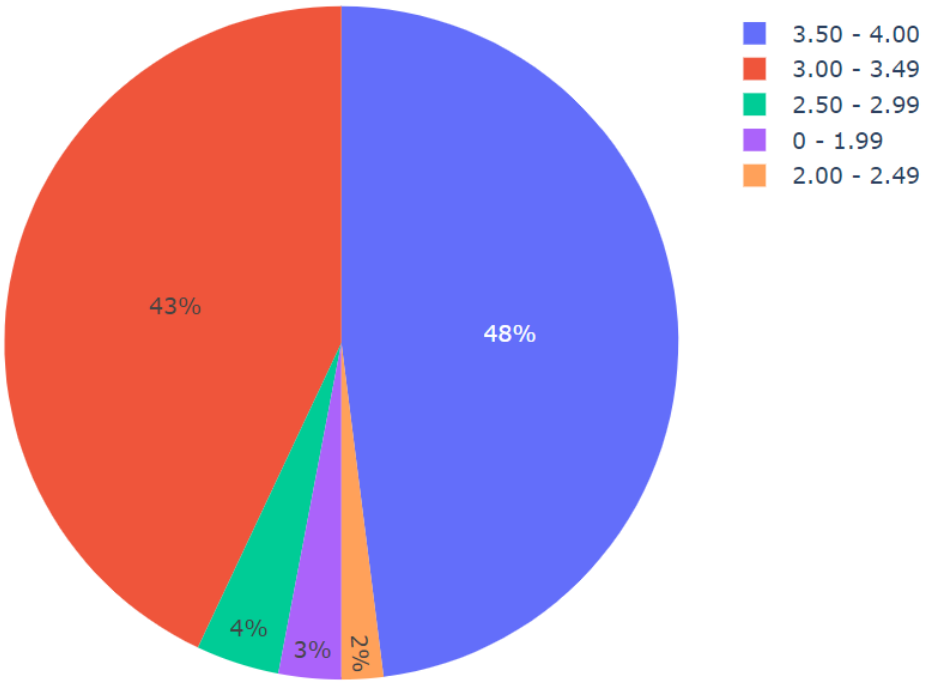


- Males are less prone to experience panic attacks as compared to females.
- Approximately 37.5% of Engineering students experience panic attacks.
- About 62.5% of IT students experience panic attacks.
- About 18% of BCS students experience panic attacks.

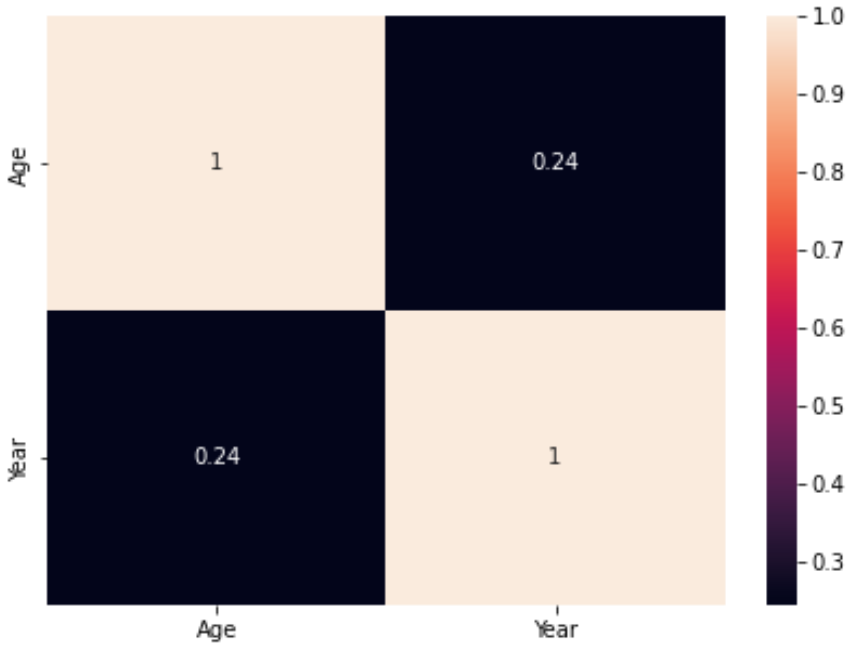
Age of Students



GPA Distribution.



Correlation Matrix

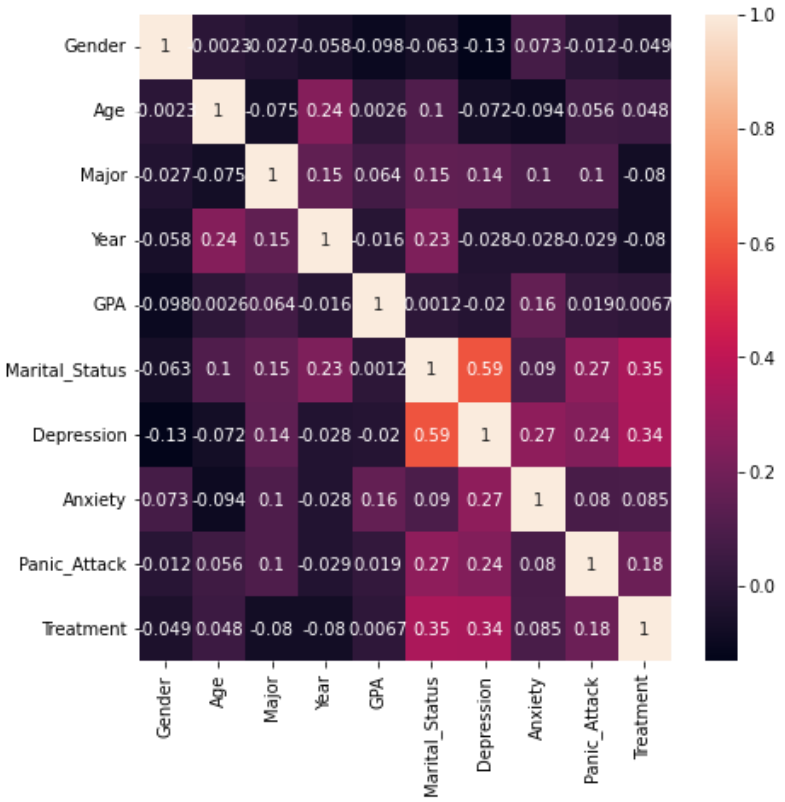


Data Processing:

On this section I will be unique numerical values to each attribute.

Since we do not need Datetime column, we can delete it for this.

And after this we can correlate a matrix figure.



- Marital status shows a close association with Depression.
- Depression, anxiety, panic attack shows a significant correlation.
- Treatment shows a slight correlation with marital status.

Model Selection:

For the model selection, I will be using pipelines to run three different model predictions:

1. Linear regression model
2. Random Forest
3. Decision Tree

```
pipeline_lr=Pipeline([('lr_classifier',LogisticRegression(random_state=42))])
pipeline_dt=Pipeline([ ('dt_classifier',DecisionTreeClassifier(random_state=42))])
pipeline_rf=Pipeline([('rf_classifier',RandomForestClassifier(random_state=42))])
```

After running these models, we can observe that the most accurate among all is the random forest one.

```
cv_results_accuracy = []
for i, model in enumerate(pipelines):
    cv_score = cross_val_score(model, X_train,y_train, cv=10 )
    cv_results_accuracy.append(cv_score)
    print("%s: %f " % (pipe_dict[i], cv_score.mean()))
```

```
Logistic Regression: 0.557143
Decision Tree: 0.471429
Random Forest: 0.585714
```

And now I will evaluate this model.

Model Evaluation:

After building the Random Forest model, I will create a classification report in order to get more information and this is the result.

```
print(classification_report(y_test, predictions))
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         2
           1       0.00      0.00      0.00         1
           2       0.00      0.00      0.00         1
           3       0.50      0.83      0.62        12
           4       0.50      0.36      0.42        14

    accuracy                           0.50        30
   macro avg       0.20      0.24      0.21        30
weighted avg       0.43      0.50      0.44        30
```

Conclusions and Recommendations:

In conclusion, we can observe females are more sensitive than males, and students have more depression in the first year. To solve this problem, the field of study must be clearly defined for the student to have confidence in his or her choice and provide psychological support to students more in the first year and those of young ages and give support to whose study Engineering and IT.

The mental health of students is greatly underestimated and as we have seen in the analysis there are many students who have mental health problems and do not receive any help. We should worry more about mental health on some occasions than getting a good grade on an exam or a project because in the end, it doesn't matter how good you are in a subject if you don't have a well-shaped head, it will not end well, and the pressure will be too much.

My recommendation is that in the future and somehow universities have mandatory psychologists for students, because many times students with problems do not resort to any help because they believe they can bear it, that this pain will not go away. Furthermore, a psychologist is not going to help... That is why I would like there to be at the end of each university year a person in charge of asking students what their mental health is like and thus preventing students from suffering these mental problems.

Work Cited:

https://www.kaggle.com/search?q=student+mental+health