

# DATA ANALYSIS PORTFOLIO

PREPARED BY

Aaliyah Motala



# Professional Background

I am a BSc Biological Science honors graduate. Research and data analysis were the core competencies of my degree. I received a dean's commendation for my Communication for Biologist module, where I was second in my year. My honors project was funded by the South African Medical Research Council (SAMRC).

I started my professional career off with the South African National Biodiversity Institute (SANBI) where I was tasked with database management, social media management, and the City Nature Challenge eThekweni 2020.

I later became a research assistant for the University of KwaZulu-Natal (UKZN) and then a clinical trial recruiter at Subjectwell, where I received an award for the most quality assurance.

I recently completed a data analyst course with Entry Level



# Portfolio Outline

Professional Background	1
Table of Contents	2
Udemy Project Description	3
The Problem	4
Design	5
Findings	6
Analysis	9
Conclusion	10
Student Mental Health Project Description	11
Data Design	12
Findings	13
Data Analysis	14
Conclusion	15
Appendix (optional)	16



# Udemy Project Description

This project was on four different Udemy courses (business finance, graphic design, musical instruments, and web development). The objective of this project was to analyze the datasets and compare and identify any trends amongst the courses. The aim was to discover if the number of subscribers and revenue increased over the years.

I was tasked with cleaning the datasets; I merged the datasets into one workbook, removed any duplicates and blank cells, provided clear and concise headers, and ensured the data in the cells were consistent and formatted correctly.

The data was then analyzed using Google Sheets and Tableau. It was found that even though business finance and web development had the same amount of courses, web development subscribers' dwarfed those of business finance.

# The Problem

The aim was to find out the trend in the different courses and if the number of subscribers to the courses increased over the years. This would in turn increase the amount of revenue received.

I had 1 week to work on this project. The four courses' data (business finance, graphic design, musical instruments, and web development) were extracted from the Udemy database and analyzed. Any other trends found would be presented in graphs.

A few business problem questions: What is your business strategy? Who is your target customer? What actions must we take based on the insights learned?

# Data Design

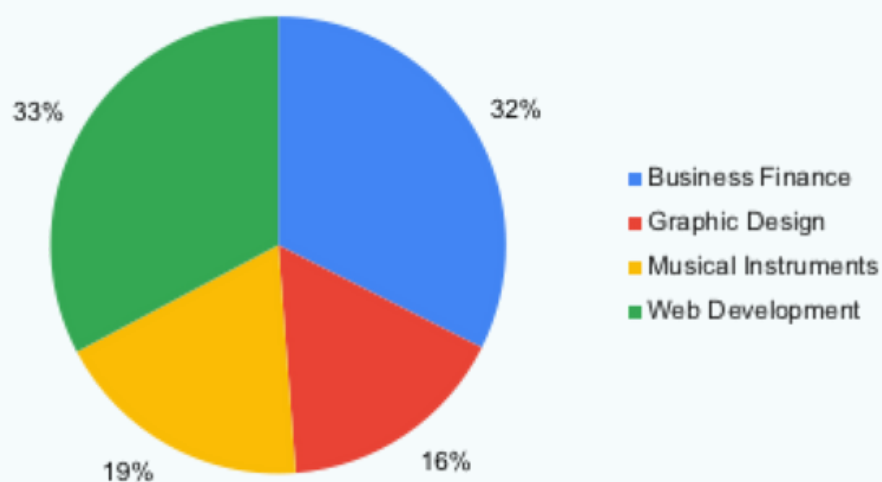
The four datasets were merged into a singular workbook. The data was cleaned by removing duplicates, removing blank cells, providing clear and concise headers, and lastly ensuring the data was consistent and formatted correctly.

Google Sheets and Tableau were used to analyze the data. Various tables and charts were created. Google Sheets was used to cleaning, manipulate, and create pivot tables. Tableau was used to analyze the data even further as it provided different insights.

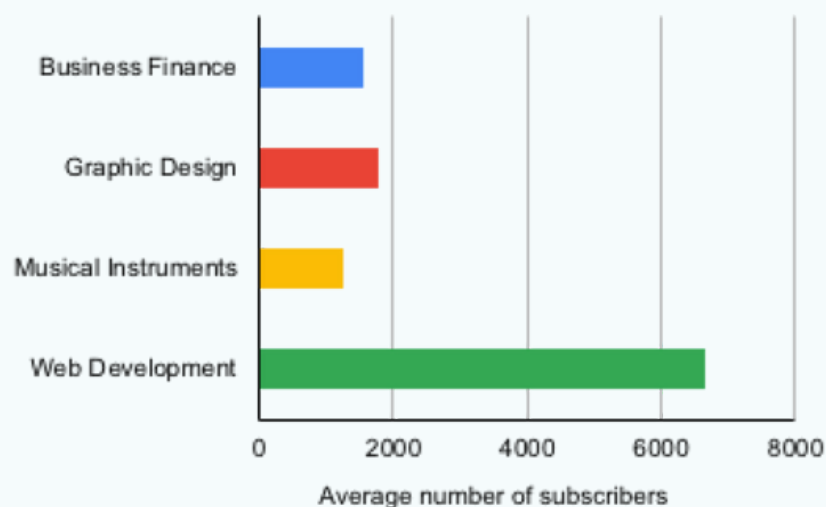
# Findings

Please see below the findings of the project separated into subsections

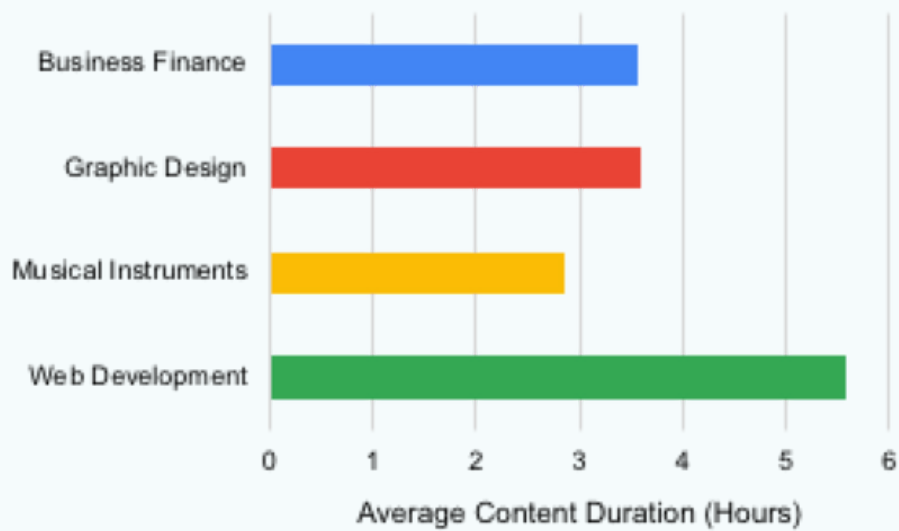
## Finding 1: Getting to know the data



**Figure 1.1: Total percentage of courses available by subject**

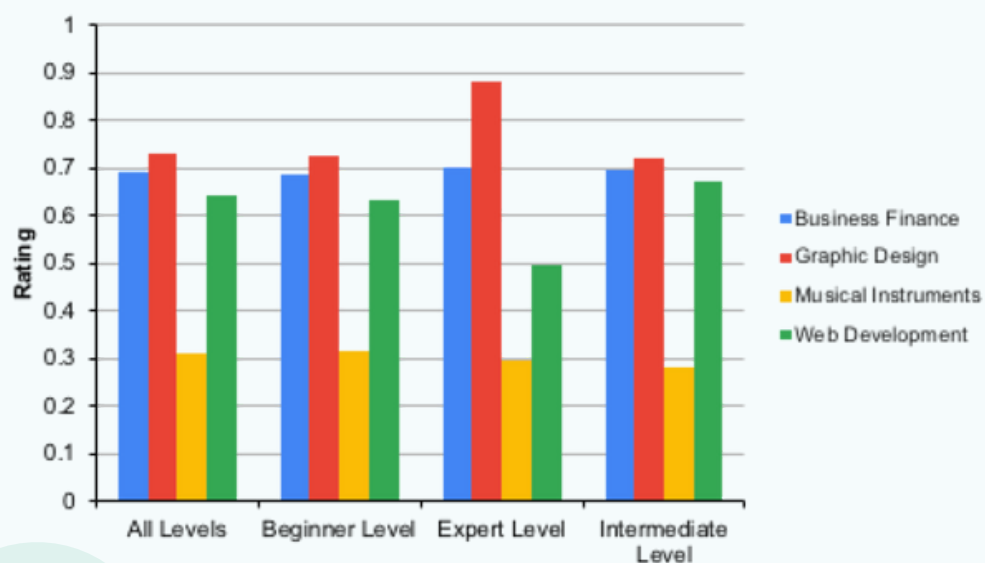


**Figure 1.2: Average number of subscribers per subject**



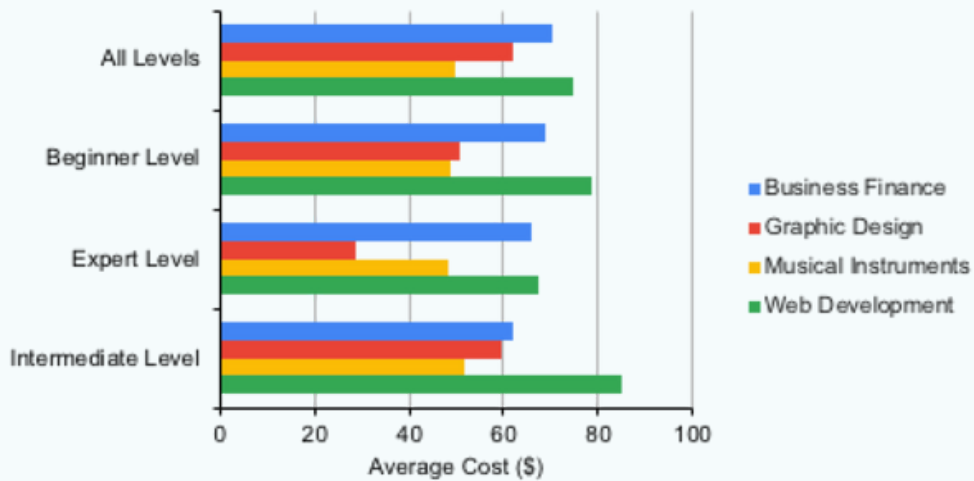
**Figure 1.3 Average content duration (hours) per subject**

## Finding 2: Skill levels



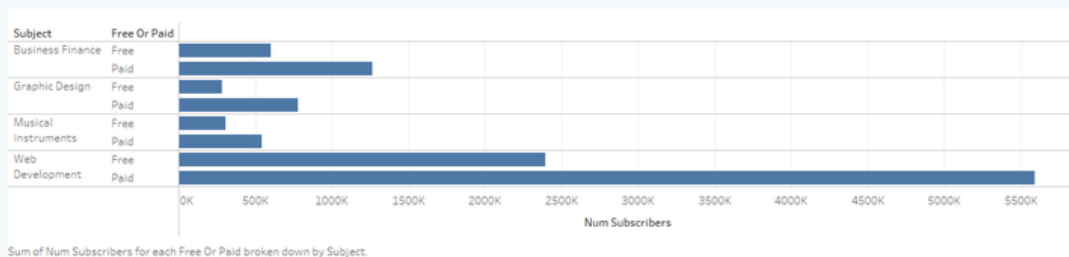
**Figure 2.1: Average rating of courses per level for each subject**



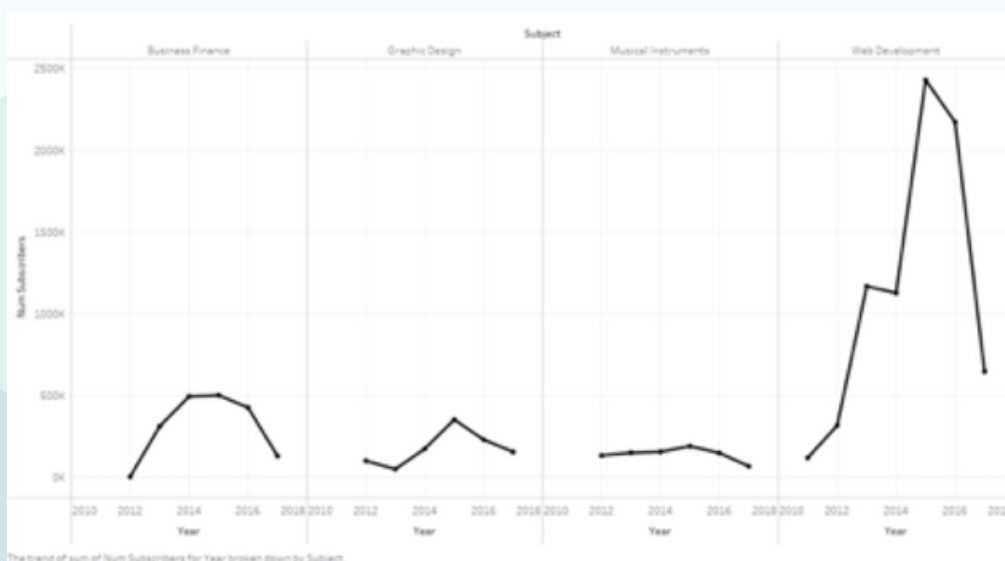


**Figure 2.2: Average cost (\$) of courses per level for each subject**

## Finding 3: Free vs. Paid



**Figure 3.1: Number of subscribers for the free and paid courses per each subject**



**Figure 3.2: Trend of subscribers for each subject over the years 2011 to 2017**

# Analysis



In our analysis of the data, we see that web development has the most courses available at 33%, followed by business finance at 32% (figure 1.1). However, looking at the number of subscribers and content duration, web development is the highest in both regards. Having a higher content duration could incite people to subscribe and have a higher value for money.

One other important factor is rating, when looking at courses, ratings from other students are important, figure 2.1 shows that graphic design at all course levels, has the highest rating, followed by business finance and then web development. Graphic design has the highest rating and has the second lowest cost for the paid courses. This show that cost can play a role in ratings.

Expanding more on the costs of the courses, we compared the free and paid courses. Figure 3.1 shows us that subscribers for paid web development courses dwarf any other subject. This is interesting since web development is the highest cost course. This could be due to the high content duration making it appear as a course worthwhile. Lastly, figure 3.2 shows a decrease in subscribers in all courses. With web development experiencing the largest decline, this could be due to the rating, the high course price, unfriendly user interface, not enough or good enough feedback from staff etc.

# Conclusion

There was a decrease in subscribers in the year 2017 and therefore a decrease in revenue. The decrease can be due to many factors that need to be looked into, starting with the teachers/ course leaders. Questions such as: How was the interaction with the students? Could the students ask questions and be answered timeously? Were there enough tests/assignments? Was there feedback? and so forth. Thereafter a review of course prices compared to other online academies needs to be analyzed. Create a new database of the ratings and comments, consider the negatives and positives, analyze the data, and as a team decided what needs to be changed.

There is a drastic drop and difference from the previous years. Feedback from students should be taken into consideration.



# Student Mental Health Project Description

People are becoming increasingly aware of mental health and more people are speaking out about it. It is stated by Kessler et al. (2007) that the majority of mental health issues, manifest and develop in the early stages of life as almost 75% of chronic mental disorders arise by the 25 years old. Social media, financial difficulties, sexual abuse, social barriers, and other external factors all play a role in students' mental health (Brown, J.S.L, 2018).

This data analysis focuses on the results of a survey done on 101 students on their mental health. Data was collected via Google forms from a University student. The aim of this project is to see if there is any relation between the students' CGPA and mental health. It is hypothesized that the students with a higher CGPA would have more students with depression, anxiety, and panic attacks than students with a low CGPA.

# Data Design

The student mental health dataset was downloaded from Kaggle. Data was collected via Google forms from a University student. There were a total of 101 students who participated.

The data was cleaned as some answers from students were in various letter cases. Data was also standardized, and abbreviations of courses were spelled out in the full course. The various Islamic courses were changed into a broad label of "Religious studies" to make the data more uniform.

The data was analyzed with Google Sheets and Tableau. Google sheets was used to manipulate and clean the data. Tableau provides good insights on different charts and diagrams of the data.

# Findings

## Finding 1: Getting to know the participants

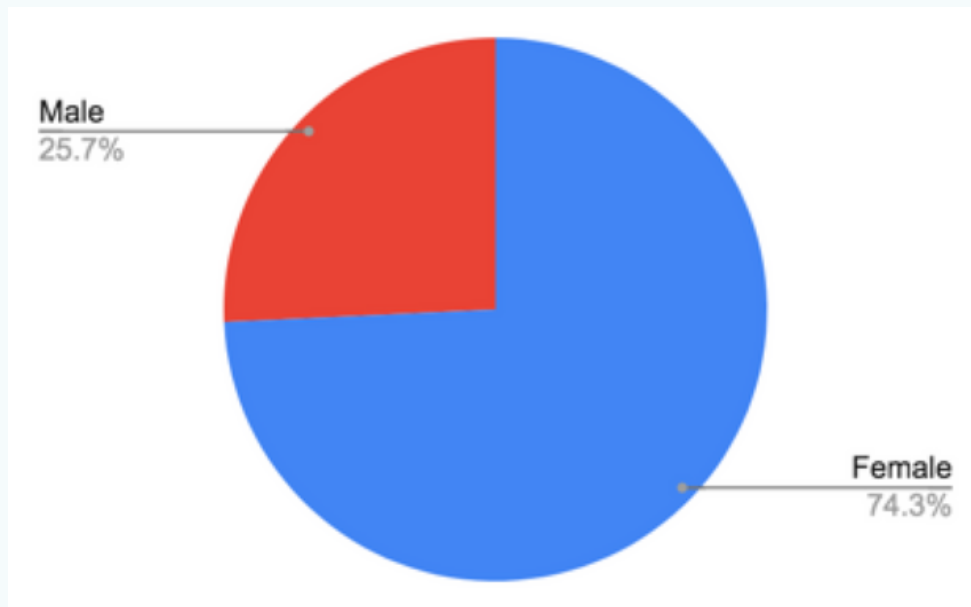


Figure 1.1: The percentage of male and females participating in the survey

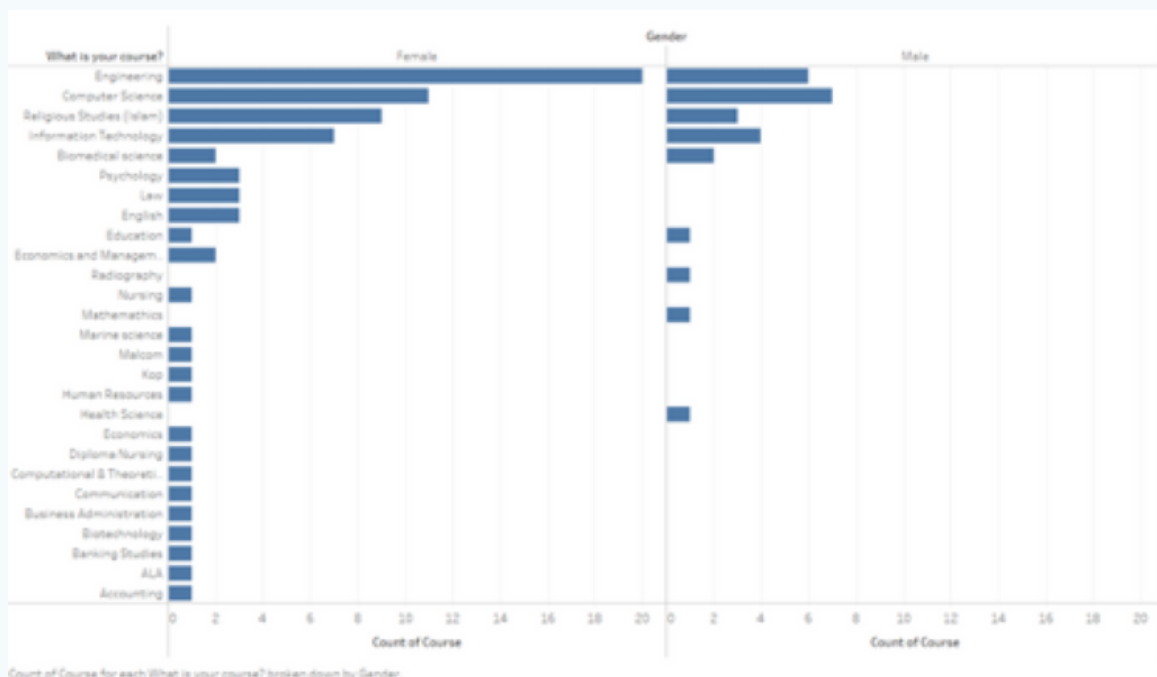
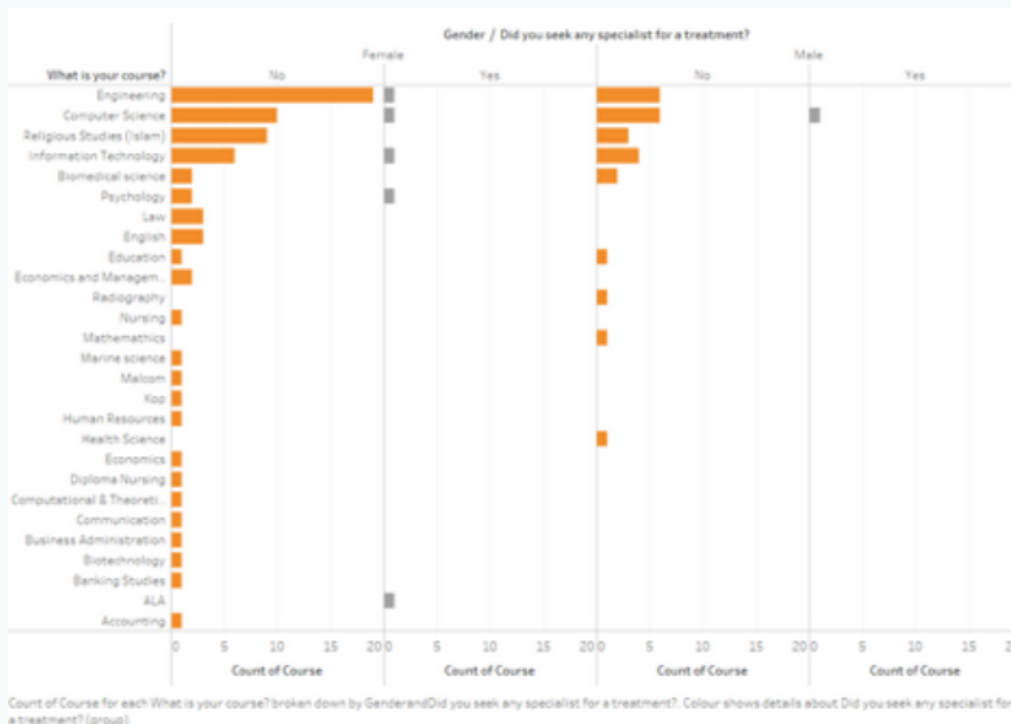


Figure 1.2: The male and female participates and their different courses

## Finding 2: Differences between male and female students' mental health

**Table 2.1: The male and female responses to the mental health questions**

Do you have Depression?			Do you have Anxiety?		Do you have Panic attacks?	
Gender	No	Yes	No	Yes	No	Yes
Female	46	29	51	24	50	25
Male	20	6	16	10	18	8
Total	66	35	67	34	68	33



**Figure 2.1: The male and female participants who have sought any treatment**

# Finding 3: Does marriage play a role in students mental health?

Table 3.1: The married participants mental health

Marital Status	Do you have Depression?		Do you have Anxiety?		Do you have Panic attack?	
	No	Yes	No	Yes	No	Yes
Yes	0	16	9	7	6	10

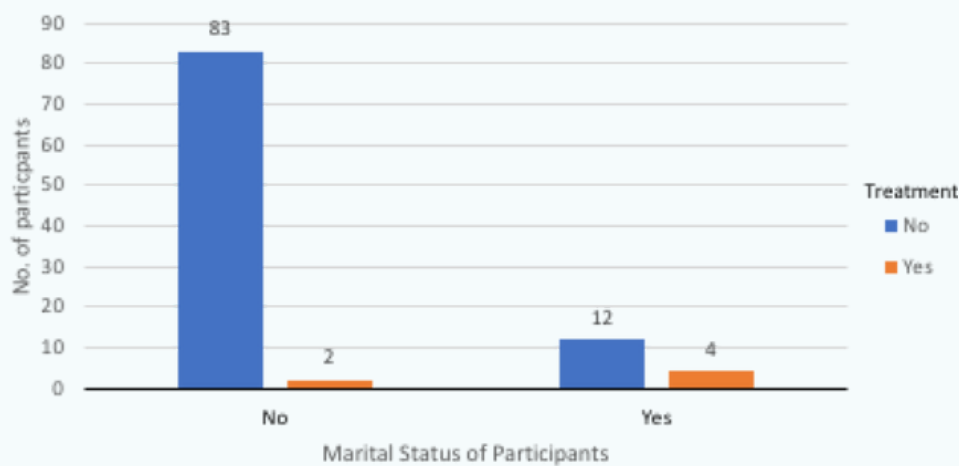


Figure 3.1: The married participants who have sought treatment



## Finding 4: CGPA and mental health

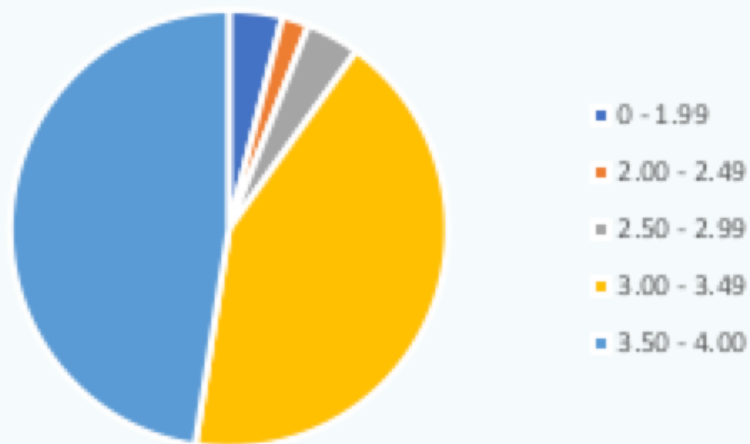
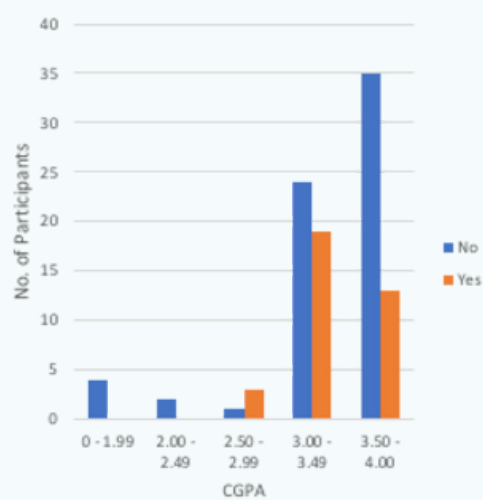
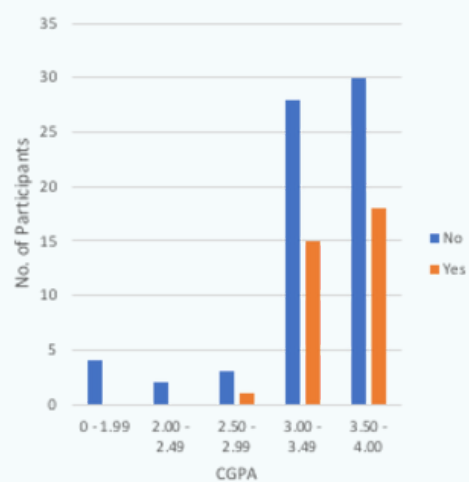


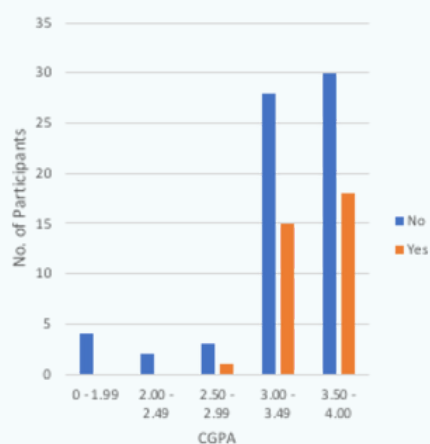
Figure 4.1: The students CGPAS distribution



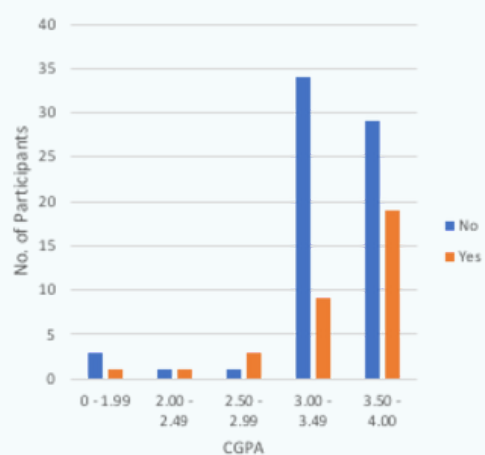
Depression



Anxiety



Panic attacks



Treatment

Figure 4.2: The participants mental health compared to their CGPA

# Analysis

The survey participants are comprised of 74% female and 26% male participants and with the majority of participants from the engineering course. Table 2.1 shows a total of participants who have depression, anxiety, and panic attacks which averages out to 34% of participants having 1 of the 3 mental health issues. Figure 2.1 shows us that only 6% of the 34% with mental health issues sought out treatment for the issues. With only 1 male seeking out any treatment.

Out of 101 survey participants, 16 were married. Interestingly, all 16 married participants said they were depressed as seen in table 3.1. Seven out of the 16 participants had anxiety and 10 out the 16 had panic attacks. In figure 3.1, we can see that out of the 6 participants that sought out treatment, 4 of them were married. This shows that marriage or a supportive partner can play a role in dealing with mental health issues, however, marriage itself can put more strain on students mentally since all 16 of the married had depression.

It is shown that the 3.50 to 4.00 CGPA was the highest with 48% of the participants. Figure 4,2 shows a trend, the higher the CGPA gets, the more the occurrence of depression, anxiety, and panic attacks. However, it is seen that anxiety starts at the lowest CGPA. This could be due to many factors such as financial strain, personal life issues, anxiety about their grades etc.

# Conclusion

Overall, the hypothesis stating that the students with the higher CGPA have been supported. A larger sample size is needed as the difference between males' and females' mental health issues could be because of the small number of participants. The reason for that could be because inadequate circulation of the survey. This could be because students didn't feel the need to participate in a survey. The questions were inadequate and broad, a more comprehensive set of questions such as: if there were any clinics at universities? Do you have supportive family or friends? have you spoken to any people you're close to about your mental health? etc. The marital status is interesting and needs to be investigated more. Students from different backgrounds to increase the datasets' cultural diversity and people with different backgrounds and upbringings as it can play a huge role in the ideas or stigmas attached to mental health.

# Appendix

## Google Sheets Data Set for Udemy Project

[https://docs.google.com/spreadsheets/d/1crO8tTwHmdsH\\_D7O7DF4MCYJhSIXzog7uCgE4RpSFNk/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1crO8tTwHmdsH_D7O7DF4MCYJhSIXzog7uCgE4RpSFNk/edit?usp=sharing)

## Google Sheets Data Set for Student Mental Health Project

[https://docs.google.com/spreadsheets/d/1WOtIKZUNrHcTVUoPssBd1lBluUJcB\\_l6sVS\\_g\\_fhUYg/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1WOtIKZUNrHcTVUoPssBd1lBluUJcB_l6sVS_g_fhUYg/edit?usp=sharing)

## References:

Brown, J.S., 2018. Student mental health: some answers and more questions. *Journal of Mental Health*, 27(3), pp.193-196.

Kessler, R.C., Angermeyer, M., Anthony, J.C., De Graaf, R.O.N., Demyttenaere, K., Gasquet, I., De Girolamo, G., Gluzman, S., Gureje, O.Y.E., Haro, J.M. and Kawakami, N., 2007. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World psychiatry*, 6(3), p.168.