

## **Data Science Project - By : Lau Yi Ting, Glendys**

**Topic :** Cyberbullying

**Dataset :** [Kaggle](#)

### **Introduction**

As social media continues to gain prominence and widespread usage in contemporary times, its influence on our society becomes increasingly significant. While social media and the internet offer avenues for expression, they also provide an easy platform for anonymous attacks, leading to the pervasive issue of cyberbullying. Cyberbullying encompasses the use of digital platforms to intimidate, harass, or harm individuals through offensive messages, threats, or online behaviour. As the digital divide continues to narrow through ongoing efforts, the widespread accessibility of the internet is anticipated to lead to an inevitable increase in social media usage. Consequently, this rise in online engagement is likely to correlate with a higher incidence of cyberbullying.

According to the National Center for Education Statistics (NCES) in the United States, during the 2019 school year, approximately 20% of students aged 12 to 18 reported being bullied, with 15% experiencing online or cyberbullying. It is also proven that 7 in 10 young people experience cyberbullying before they hit the age of 18.

There has been a substantial effort in countering cyberbullying through various strategies and initiatives, such as education in gaining awareness of cyberbullying, anti-bullying policies and legislation and collaboration with technology companies.

### **Motivation**

Despite the efforts mentioned above in the introduction, countering cyberbullying is still facing ongoing challenges as technology is evolving every minute through the introduction of more digital platforms (tiktok) and advanced technology (deep fakes) thus, there is a need to continue refining strategies and evaluate the effectiveness of the strategies.

In order to do so, we first need to understand the effects of cyberbullying and what people in our current society think of cyberbullying. A Google survey was conducted to collect public opinions and personal experiences regarding this issue. The dataset will undergo exploratory data analysis to visualize the impact of cyberbullying and draw meaningful conclusions, guiding us towards appropriate measures to combat this problem.

Some questions I aim to answer with this research includes but not limited to:

- a) What age group are more vulnerable to cyberbullying and why?
- b) Do the type of platforms affect the chances of cyberbullying?
- c) Do people's opinions on cyberbullying differ based on their age group and why?

## Dataset

Dataset is taken from [Kaggle](#), which was collected by user, **PRIYANSHU SETHI**, for a school analytical report and includes responses from people with different geographical backgrounds. Google forms were filled out by people on Discord, and have responses that will help to deep dive into the public opinion and impact of cyberbullying on their lives. The dataset has 68 responses with 18 columns. Columns include answers to survey questions such as “What do you consider as Cyber bullying?”, and whether they have been a victim of cyber bullying themselves.

	Timestamp	Email Address	What's Your Age?	What do you consider as Cyber bullying?	What Age group of people are most Vulnerable to Cyber bullying	Have you ever been cyber bullied?	If Yes , What was the way you dealt with the same
1	1/22/2022 11:39:31	armadhav2007@gmail.com	15 - 18	Hacking / Phishing your private details, Spamming , Unlawf...	19-30	No	NA
2	1/22/2022 11:41:15	guptasayyam7@gmail.com	15 - 18	Spamming , Unlawfully morphing your images, Messages an...	15 - 18	Yes	Blocked / Reported The Bully
3	1/22/2022 11:41:57	sonaissuhani@gmail.com	15 - 18	Mean/Inappropriate Comments, Pranks Calls by strangers, S...	15 - 18	Yes	I was scared to do anything, Blocked / Reported The Bully, F...
4	1/22/2022 11:42:02	madmaxhu69@gmail.com	15 - 18	Mean/Inappropriate Comments, Spamming , Unlawfully mo...	Below 15	No	Aaj mere paas bangle 8904 hai, gadi hai 898—, bank-balanc...
5	1/22/2022 11:49:28	aryanjhbcsgaming@gmail.com	15 - 18	Hacking / Phishing your private details, Mean/Inappropriate...	19-30	No	NA
6	1/22/2022 11:50:23	samonwaychakroborty2@gmail.com	15 - 18	Hacking / Phishing your private details, Mean/Inappropriate...	Below 15	Yes	Fought Back the Bully
7	1/22/2022 11:51:15	sehajgun24@gmail.com	15 - 18	Hacking / Phishing your private details	equally vulnerable	Yes	Stopped using Social Media
8	1/22/2022 11:51:18	scsjangid1976@gmail.com	15 - 18	Mean/Inappropriate Comments, Spamming , Unlawfully mo...	15 - 18	No	NA
9	1/22/2022 11:54:58	ekamwaraich5@gmail.com	15 - 18	Mean/Inappropriate Comments, Spamming , Unlawfully mo...	equally vulnerable	No	NA
10	1/22/2022 11:55:09	aroramann50@gmail.com	15 - 18	Messages and Calls containing threats	equally vulnerable	Yes	I was scared to do anything
11	1/22/2022 11:56:51	kashish2004p@gmail.com	15 - 18	Hacking / Phishing your private details, Mean/Inappropriate...	equally vulnerable	Yes	Blocked / Reported The Bully, Talked to a Near one for emot...

Variable	Data Type	Variable Content
Timestamp	character	Date and time of survey response
Email Address	character	Email of respondent
What's Your Age?	character	Age of respondent: < 15, 15-18, above 18
What do you consider as Cyber bullying?	character	Mulitple choice - Hacking / Phishing your private details, Spamming , Unlawfully morphing your images, Messages and Calls containing threats, Leaking private conversations, Pranks Calls by strangers, Stalking
What Age group of people are most Vulnerable to Cyber bullying?	character	Either - equally vulnerable, 15 – 18, Below 15, 19-30, above 30
Have you ever been cyber bullied?	character	Either – Yes or No
If Yes , What was the way you dealt with the same?	character	Multiple choice – Blocked / Reported The Bully, Fought Back the Bully, Talked to a Near one for emotional support, Stopped using Social Media or free text
What do you think made you the victim of cyber bullying?	character	Unstructured text
What Platform according to you leads in cyber bullying?	character	Either – unstructured text, Discord, Facebook, Instagram, All of the above

What type of people are usually the bullies on the internet?	character	Unstructured text
Have You ever knowingly / Unknowingly bullied someone on the internet?	character	Either – Yes or No
What lead you to bully someone ? ( if yes )	character	Unstructured text
What kind of people are usually the victims of cyber-bullying?	character	Either - All of above, Influencers / Famous People, People with different sexual orientation, People with Distinctive Physical appearance
What are the causes of cyber bullying according to you?	character	Either - Lack of awareness, Mental Health Problems, Casteism / Racism, Sadism, Gender Bias
What has more impact on an individual , Cyber Bullying or Bullying in real life ?	character	Either – Cyber Bullying , Bullying in real life, Equal Impact
Who are bullies usually?	character	Either – A stranger, Someone you know, Can be Both
Youtube Roasting is a form of bullying. Do you agree ?	character	Either – Yes or No
Your opinion on the dank communities on the internet , usually involved in bullying?	character	Unstructured Text

## Data Processing / Cleaning

### Cleaning columns with multiple answers

There are some columns with multiple answers separated by “,” such as “What do you consider as Cyber bullying?” and “If Yes , What was the way you dealt with the same?”. We will not be able to use the data effectively if we were to use the columns’ data as it is.

Thus, there is a need to clean the columns by splitting the columns into their individual separate values, removing duplicates and transforming the data into a more structured format.

```
> cleaned_causes_of_bullying <- data %>% separate_rows(`What do you consider as Cyber bullying?`, sep = ",")

> cleaned_causes_of_bullying <- cleaned_causes_of_bullying %>% count(`What do you consider as Cyber bullying?`, sort = TRUE)
```

	What do you consider as Cyber bullying?	n		If Yes , What was the way you dealt with the same	n
1	Unlawfully morphing your images	49	1	NA	39
2	Spamming	47	2	Blocked / Reported The Bully	14
3	Leaking private conversations	45	3	Fought Back the Bully	6
4	Messages and Calls containing threats	42	4	Fought Back the Bully	4
5	Hacking / Phishing your private details	40	5	Talked to a Near one for emotional support	3
6	Mean/Inappropriate Comments	28	6	I was scared to do anything	3
7	Stalking	23	7	Filed an official complaint against the Bully	2
8	Pranks Calls by strangers	16	8	Stopped using Social Media	2

## Text Preprocessing

There are some columns that contains unstructured text as those are responses to open-ended questions in the survey. It will also be inaccurate and inefficient to analyse the text as it is and thus, I will be performing tokenization, stemming and removing stopwords on the open-ended columns such as “Your opinion on the dank communities on the internet , usually involved in bullying?”, “What type of people are usually the bullies on the internet?”.

```
> tidy_cyberbullies <- data %>% select(`What type of people are usually the
bullies on the internet?`)

> tidy_cyberbullies <- tidy_cyberbullies %>% unnest_tokens(word,`What type
of people are usually the bullies on the internet?`)

> tidy_cyberbullies <- tidy_cyberbullies %>% anti_join(stop_words)

> stemmed_tidy_cyberbullies <- tidy_cyberbullies %>% mutate(stemmed_text =
wordStem(word))

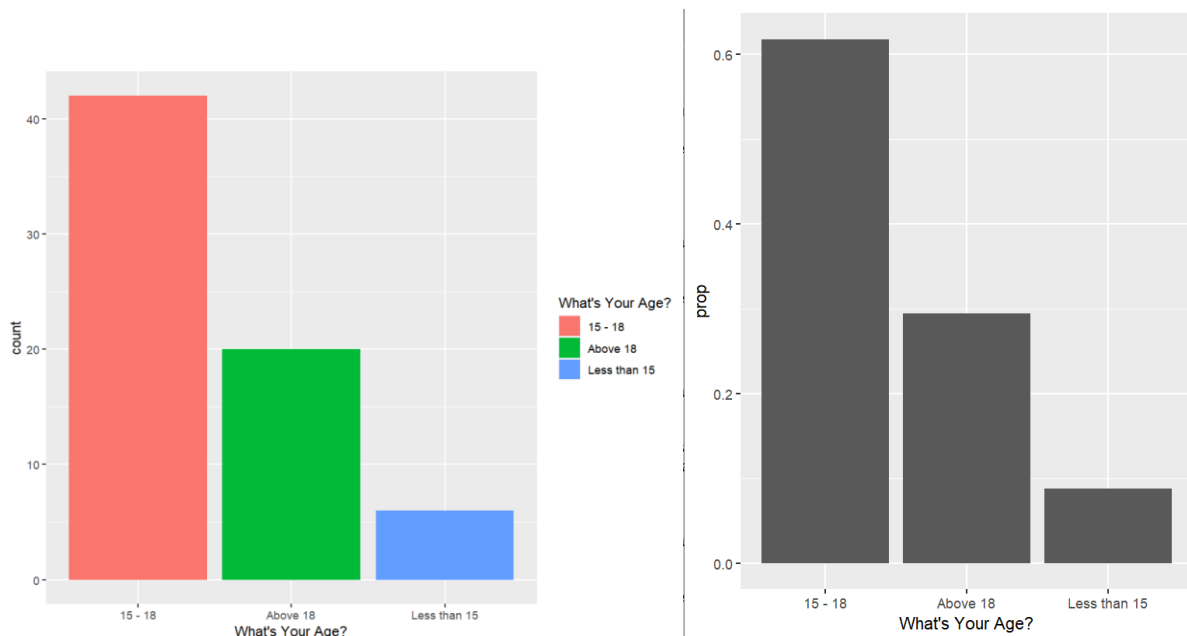
> stemmed_tidy_cyberbullies <- stemmed_tidy_cyberbullies %>%
select("stemmed_text")

> stemmed_tidy_cyberbullies%>% count(stemmed_text, sort=TRUE)
```

	stemmed_text	n
1	peopl	23
2	NA	14
3	bulli	10
4	insecur	7
5	hacker	4
6	life	3
7	person	3
8	ppl	3
9	real	3

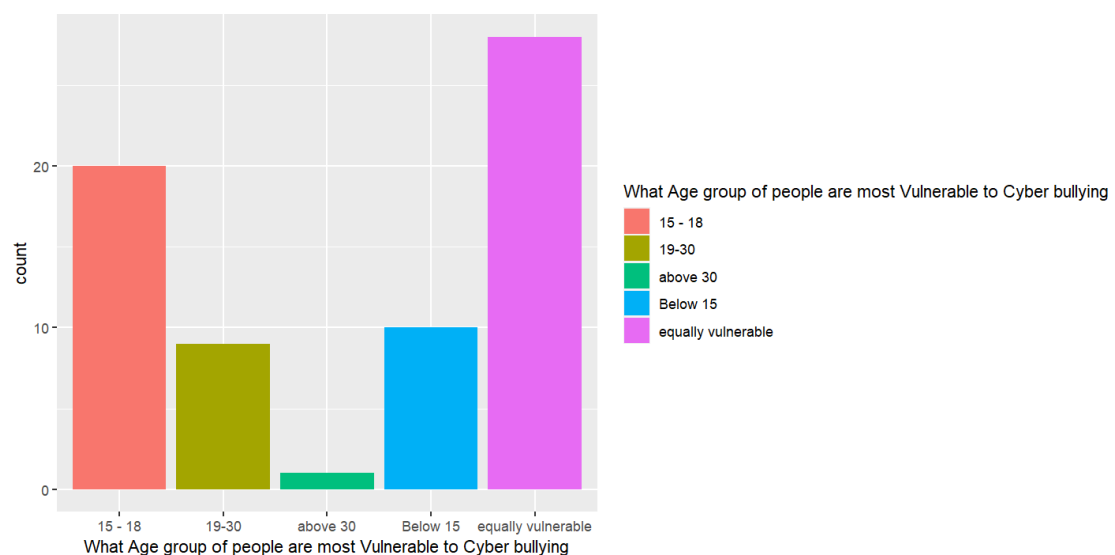
# Exploratory Data Analysis and Visualisation

## EDA on Respondents' Profile (Age Group)

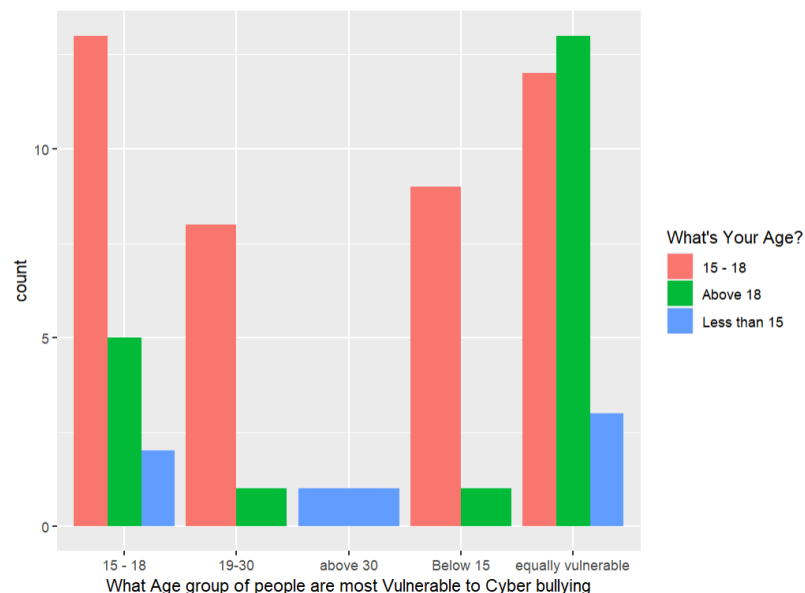


From the above barcharts, most of the respondents (at least 60%) are of the age group of 15-18. Age groups 15-18 are considered teenagers; thus, we can safely assume that the results of this data analysis on the effects and view on cyberbullying will mainly apply to teenagers.

## EDA on "What Age Group of people are most vulnerable to Cyberbullying"

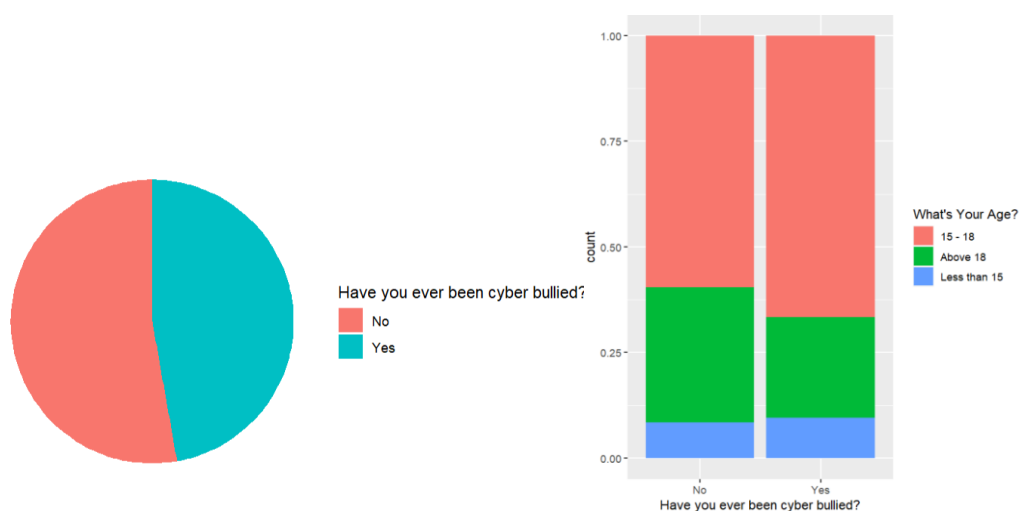


Base on the barcharts above, most respondents feel that all ages are equally vulnerable to cyberbullying, this is most likely because of the well known fact that people of all ages have access to the internet and are actively using social media as well. However, it is also interesting to take note that significant number of respondents feel that people of the age group 15-18 are the most vulnerable to cyberbullying, to further analyse why, I factored in the age group of the respondents for the next visualisation.



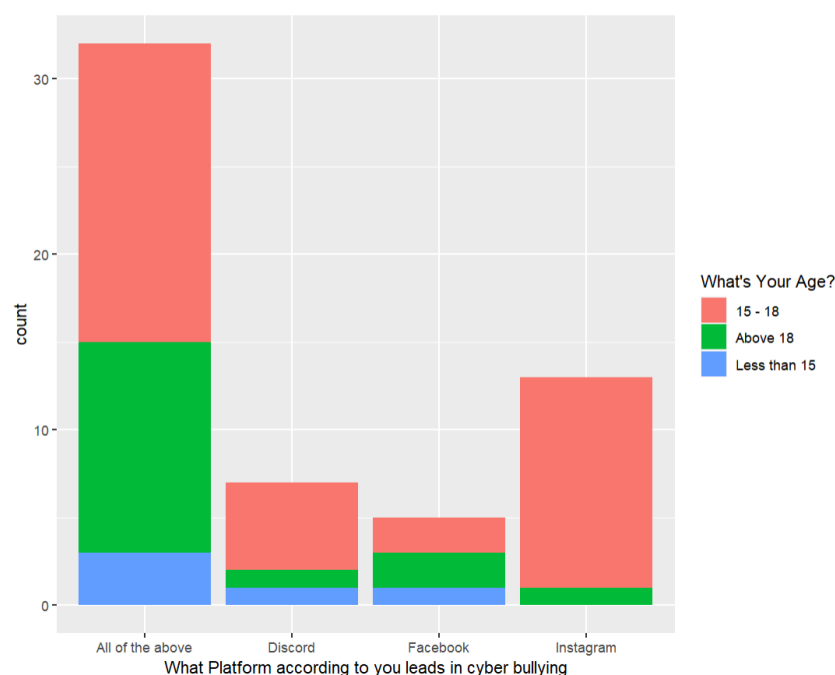
Base on the above bar chart, only respondents of the age group < 15 years old felt that people who are aged above 30 years old are most vulnerable to cyberbullying. Respondents of the age group 15-18 seem to be slightly equally divided among the different options compared to respondents above 18, who mostly agreed on the “equally vulnerable” option.

### EDA on “Have you ever been cyber bullied?”

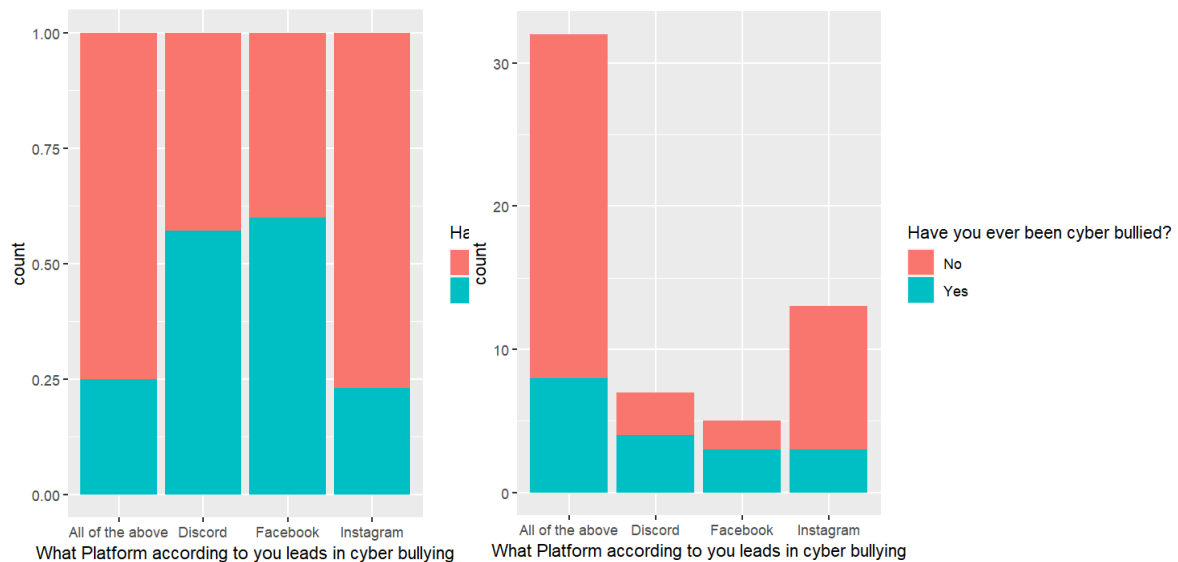


Out of 68 respondents, almost half of them have experienced being cyberbullied before which again strongly implies that cyberbullying is a common phenomenon among teenagers. After deep diving into the answers given by each age group, it seems like respondents of the age group 15-18 are more likely to experience being cyberbullied as compared to the other age groups. Base on research, it is suggested that young people are more likely cyber bully someone online rather than in real life as they feel less accountable for their actions due to the nature of the online world. And thus, people of age group 15-18 are more likely to interact with people of the same age group online which could make them more vulnerable to cyber bullying actions.

### EDA on “What Platform according to you leads to cyber bullying?”



This bar chart shows that around half of the respondents felt that all the different platforms (discord, facebook, instagram) are platforms that leads to cyberbullying, followed by Instagram, Discord and lastly Facebook. In order to discover if the platform makes a difference in the chances of getting cyberbullied, there is a need to do further analysis.



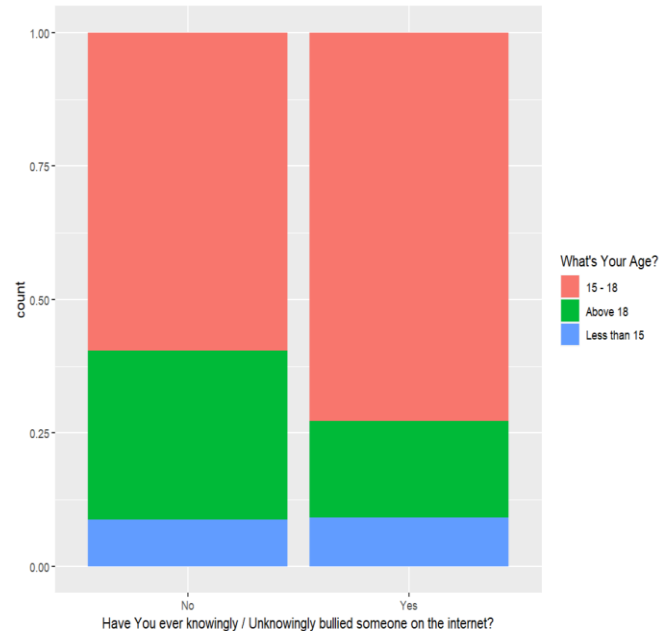
According to the barchart above, most people who answered “Yes” to being cyberbullied felt that platforms - Discord and Facebook are the most likely to lead to cyber bullying. This can mean that respondents who are victims to cyberbullying most likely have been cyberbullied on these 2 platforms as compared to Instagram.

*Possible reasons that support the theory are :*

1. Platform demographics – Facebook and Discord have different user demographics, Facebook has a broader range of users of all age groups (especially older individuals). Whereas Instagram is more popular among younger users, particularly teenagers. This can imply that different age groups have varying levels of awareness and sensitivity regarding cyberbullying and thus, impacting the prevalence n these platforms.
2. Communication formats – Facebook and Discord offer more extensive communication options such as longer-form posts, group discussions, private messaging and longer comments as compared to Instagram.
3. Anonymity and privacy settings- Discord offers higher anonymity than Instagram, allowing users to create pseudonymous accounts and engage in various servers without revealing their identities. This anonymity encourages cyberbullying without immediate consequences. Additionally, Facebook's privacy settings and customization options make it easier for bullies to target and harass specific individuals.
4. Different purposes and user behavior- Facebook is primarily a social networking site where people connect with friends, family, and acquaintances. Discord, on the other hand, is known for its focus on communities and interest-based groups. The specific purposes and dynamics of these platforms can shape the interactions and, unfortunately, facilitate cyberbullying in certain contexts.

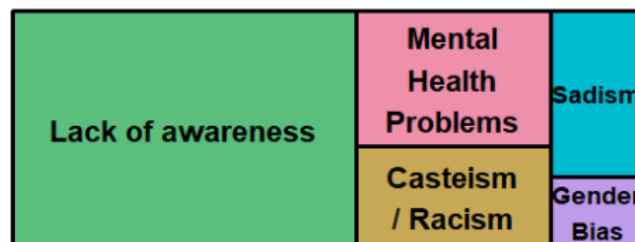


EDA on “Have You ever knowingly / unknowingly bullied someone in the internet?”

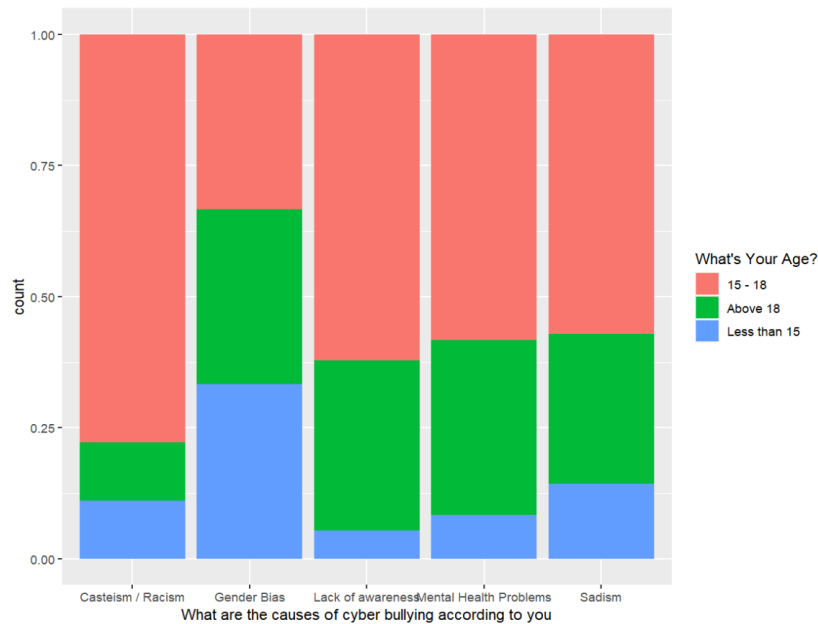


We can tell from the bar chart that respondents of the age group 15-18 are more likely to knowingly or unknowingly cyberbully someone on the internet. This insight is parallel to the finding from the previous section that people of the same age group are also more likely to be cyberbullied. More research is needed to understand this phenomenon.

EDA on “What are the causes of cyber bullying according to you?”



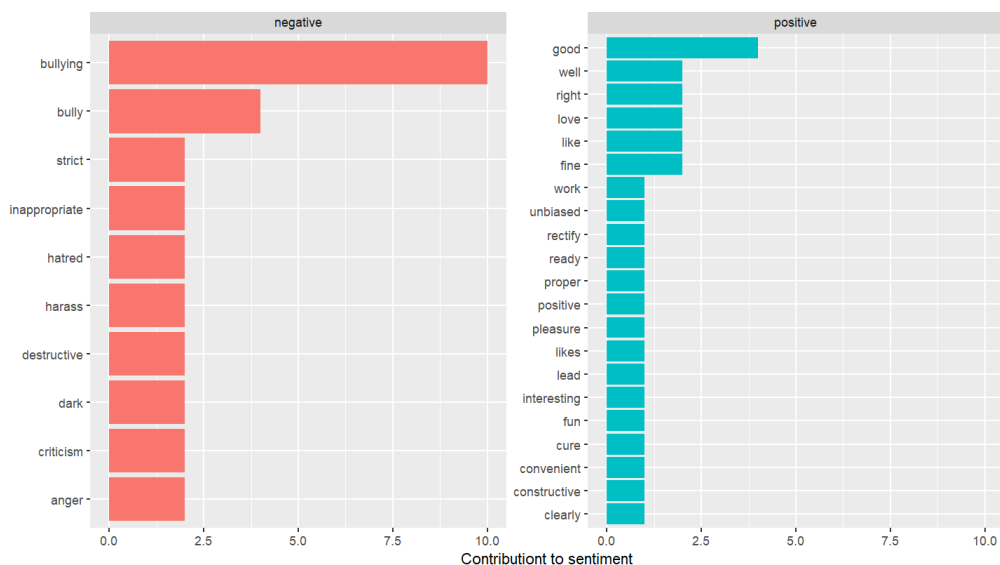
From the treemap above, we can safely assume that almost half of the respondents picked “Lack of awareness” as a cause of cyberbullying.



Majority of the respondents who are of an age group < 15 picked gender bias at the main cause which is completely opposite to those of age group 15-18, while respondents of age group above 18 equally picked all the options except for Casteism / Racism.

## EDA on “Your opinion on the dank communities on the internet , usually involved in bullying?”

For context, dank communities are social groups, primarily found online, that revolve around the appreciation and sharing of edgy, high-quality, and often humorous content known as "dank" content. These communities embrace internet culture, memes, and inside jokes, fostering a sense of shared identity among their members. They engage in activities such as meme creation, remixing content, and participating in challenges, ultimately creating a space where like-minded individuals can connect, laugh, and discuss dank content.



The above visualisation was created after performing sentiment analysis. As we can tell, the answers provided by respondents have a higher negative sentiment as compared to positive sentiment and a lot of the words answered had “bullying” sentiments which could imply that respondents feel that dank communities contribute to cyberbullying.

After research, possible reasons as to why dank communities contribute to cyberbullying other than anonymity,

1. Boundary pushing humour - the culture within dank communities often revolves around humor and pushing boundaries. While this can be entertaining for many, it can also create an environment where offensive or demeaning content is normalized or even celebrated. This normalization can contribute to the spread of cyberbullying, as members may not always consider the impact their words or actions have on others.
2. Group vs Stranger mentality - Moreover, within dank communities, there can be a sense of group dynamics or tribalism. This can lead to the targeting of individuals who are perceived as outsiders or different from the majority. Such targeting can manifest as cyberbullying, with members of the community ganging up on or harassing those they deem as different or unwelcome.

## Insights gained

Going back to the initial research questions :

### *a) What age group are more vulnerable to cyberbullying and why?*

Age group 15-18 are more vulnerable to cyberbullying and are more likely to be cyberbullies as compared to the other age group. Scientific research has always stated that teenagers belonging to this age group have developmental factors that may contribute to this theory, such as: heightened emotions, increased impulsivity, desire for peer acceptance and thus, more likely to be peer pressure. These factors can contribute to risky behaviour including cyberbullying.

Teenagers also have stronger digital literacy and online presence as technology / social media have already been prominent during their early stages of childhood and thus, they may spend significant time online and interact with their peers. This increased online presence provides more opportunities to be both victims or perpetrators to cyberbullying.

### *b) Do the type of platforms affect the chances of cyberbullying?*

It seems like the type of platforms significantly affect the chances of cyberbullying. According to cyberbullying statistics, among all the social media platforms, Facebook was the source of cyberbullying for 75% of online harassment victims in the United States in 2021 ([Beatrice M. 2023](#)).

Thus, further research is needed to understand how some social media platforms decrease the likelihood of cyberbullying such as Instagram's privacy settings etc. Insights learned from the research can be used to apply on platforms such as Facebook and Discord.

*c) Do people's opinions on cyberbullying differ based on their age group and why?*

Based on the barcharts, people's opinions on cyberbullying is largely affected by their age group, for example, how majority of the people of age group < 15 felt that gender is the biggest reason for cyberbullying. However, due to the limitations of the dataset, it is hard to deduce why it differs.