A

PROJECT REPORT ON STATISTICAL ANALISIS

“To Analyze Cybercrime Awareness Among Students.”

By T.Y. Bsc Students.

PRESENTED BY

- **Ms. Sayali Kishor Pise.**
- **Ms. Harshada Kishor Deore.**
- **Ms. Madhuri Tarachand Patil.**
- **Ms. Nikita Pramod Pakhale.**

Submitted to

DEPARTMENT OF STATISTICS

JAI HIND EDUCATIONAL TRUSTS

Z. B. PATIL COLLEGE, DHULE.

Under the guidance of

Prof. P.U. Patil Sir

Kavayatri Bahinabai Chaudhari North Maharashtra University, Jalgoan.

ACADEMIC YEAR 2020-2021

CERTIFICATE

This is to certify that a project report on

"To Analyze Cybercrime Awareness Among Students."

By T.Y. Bsc. Students.

Submitted by,

- **Ms. Sayali Kishor Pise.**
- **Ms. Harshada Kishor Deore.**
- **Ms. Madhuri Tarachand Patil.**
- **Ms. Nikita Pramod Pakhale.**

Under our guidance in partial fulfillment of requirement for award of **"Bachelor of Science (Statistics)" of Kavayatri Bahinabai Chaudhari North Maharashtra University, Jalgoan, in academic year of 2020-2021.**

Project Preceeptor

Prof. P.U. Patil Sir,

Department of Statistics

Z.B. Patil College, Dhule.

Head

Prof. G.T. Pedram Sir,

Department of Statistics

Z.B. Patil College, Dhule.

ACKNOWLEDGEMENT

"WITH GREAT PLEASURE AND EFFORTS, WE ARE ABLE TO PRESENT THIS PROJECT."

WE ARE VERY THANKFUL TO ***PROF. P.U. PATIL SIR***, DEPARTMENT OF STATISTICS FOR HIS GREAT HELP AND ENCOURAGING US DURING THE COURSE OF DEVELOPMENT OF THIS PROJECT.

WE ALSO EXPRESSED OUR PLEASURE TOWARDS,

PROF. G.T. PEDRAM SIR (HEAD OF DEPARMENT OF STATISTICS) AND ***DR. N.S. CHAVAN SIR (DEPT. OF STATISTICS, Z.B. PATIL COLLEGE, DHULE.)*** FOR THEIR VALUABLE GUIDANCE.

WE ARE ALSO THANKFUL TO OUR TECHNICAL LAB ASISTANT ***MR.P.R. JADHAV*** WHO HELPED US DURING THE PROJECT WORK. WE ALSO APPRECIATE ***MR. SACHIN DABHADE*** **SIR** FOR PROVIDING US INFORMATION FROM COLLEGE OFFICE RECORD.

I WHOLE HEARTEDLY THANK MY PARENTS, WITHOUT WHO'S SUPPORT I COULD NEVER HAVE COMPLETED THIS PROJECT.

LASTLY, WE WOULD LIKE TO THANK ALL OUR FRIENDS WHO GAVE US THEIR KIND CO-OPERATION AND VALUABLE SUGGETIONS WHILE MAKING PROJECT.

- **Ms. Sayali Kishor Pise.**
- **Ms. Harshada Kishor Deore.**
- **Ms. Madhuri Tarachand Patil.**
- **Ms. Nikita Pramod Pakhale.**

T.Y. Bsc. (2020-2021)

Dept. Of Statistics.

CONTENT

INTRODUCTION

According to our Hon. Prime Minister **"Cyber related risks are a global threat of bloodless war. India can work towards giving world a shield from the threat of cyber warfare".**

The internet is one of the most sophisticated technologies created. The internet and its related technologies are constantly evolving. The number of internet users is increasing steadily around the globe. Internet offers great benefit to society but also present opportunities for crime using new and highly sophisticated technology tools. Today email and websites have become the preferred means of communication.

Cyber-crime is defined as the **"Illegal activity done using computer system and the internet"** which is punishable by law. With increasing internet penetration, cybercrimes have also increased in the last few years. Between **2015** and **2020**, the number of cyber-crimes registered in the country has gone up to 5 times. **Maharashtra** and **Uttar Pradesh** alone encountered **1/3rd** of these crimes.

There are different types of Cyber-Crimes….

TYPES

CREDIT CARD FRAUDS:

- This **Cyber Crime** involves stealing **credit card** information of a user either through computer malware or through phishing attacks and making heavy purchases using the **credit card** data.
- As per the latest RBI guidelines, a customer is liable to pay to the Bank only upto Rs. 10,000 if he is a **Credit Card Fraud** victim.

CYBER STALKING:

- **Cyberstalking is** the use of the Internet or other electronic **means** to stalk or harass an individual, group, or organization.
- It may include false accusations, defamation, slander and libel. **Cyberstalking is** often accompanied by realtime or offline stalking.

CYBER TERRORISM:

- **Cyberterrorism** is the use of the Internet to conduct violent acts that result in, or threaten, loss of life or significant bodily harm, in order to achieve political or ideological gains through threat or intimidation.

- Cyberterrorism can be also defined as the intentional use of computers, networks, and public internet to cause destruction and harm for personal objectives.

HACKING:

- Hacking refers to activities that seek to compromise digital devices, such as computers, smartphones, tablets, and even entire network.
- Hacking is typically technical in natures. But hackers can also use psychology to trick the user into clicking on a malicious attachment or providing personal data.

ONLINE FRAUDS:

- **Online fraud** is a type of cybercrime fraud or deception which makes use of the Internet and could involve hiding of information or providing incorrect information for the purpose of tricking victims out of money, property, and inheritance.

SPAM MAILS:

- **Email spam**, also referred to as **junk email** or simply **SPAM**, is unsolicited messages sent in bulk by email (spamming) The legal definition and status of spam varies from one jurisdiction to another, but nowhere have laws.
- Spammers collect email addresses from chat rooms, websites, customer lists, newsgroups, and viruses that harvest users' address books.

SURVEY DESIGN

A single survey is made of at least a sample method of data collection (e.g., a questionnaire) and individual questions or items that become data that can be analyzed statistically.

A survey may focus on different types of topics such as preferences, opinions behavior, or factual information, depending on its purpose. Since survey research is almost based on a sample of the population.

The success of the research is dependent on the representativeness of the sample with respect to a target population of interest to the researcher.

- **Selection of Sample Size: -**

For this study sample size of **120** has been taken.

Total strength of students in Jaihind Educational Trust's, Zulal Bhilajirao Patil College.

Total strength of students studying in college (11^{th} to T.Y.) = **3,713**

Source: (From college admission record)

SOURCES OF DATA COLLECTION

Project will be based on two sources:

1. Primary data.
2. Secondary data.

1) Primary Data: -

Questionnaire: - Primary data was collected by preparing online questionnaire and the students were had been requested to fill them.

We send them an online questionnaire link as below: -

https://surveyheart.com/form/6092bef3e819ba72d47e7aa5

2) Secondary Data: -

Secondary data will be consisting of an internet and websites.

https://cybercrime.gov.in/

https://www.metacompliance.com

www.cyberangles.org

STATISTICAL TOOLS

The main statistical tools used for collection and analysis of data in this project are:

- Questionnaire
- Determination of sample size
- Test of randomness
- Chi-Square test of independent attributes
- ANOVA
- Correlation coefficient
- Pie diagram
- Bar diagram
- Column diagram

PROJECT OBJECTIVE: -

- The aim of study is to investigate the attention of cyber-crime among faculty students of Jaihind Educational Trust's, Zulal Bhilajirao Patil College.
- The survey is to analyse the awareness of cyber-crime among youths.

DETERMINATION OF SAMPLE SIZE

Here, we have,

- N = Population size = 3713
- α = Level of significance = 0.05
- p = Population proportion = 0.5
- $Z_{\alpha/2}$ = Confidence level = 95 % = 1.96
- d = Margin error = 0.09

Therefore, To Determine the Sample Size n,

$$= \frac{(N*p(1-p) * Z^2_{\alpha/2})}{(p*(1-p) * Z^2_{\alpha/2}+(N-1) * d^2}$$

$$= \frac{(3713*0.5*(1-0.5) * (1.96)^2)}{(0.5*(1-0.5) * (1.96)^2+((3713-1) * (0.09)^2)}$$

$$= \frac{3565.4}{29.6}$$

n = 120.47 = **120**

TEST OF RANDOMNESS

- H_o= Sample is random.
- H_1 = Sample is not random.

To test above hypothesis first we find the value of r. i.e., number of r.v.s

- m = Student is male.
- f = Student is female.

The sample is,

mmmmfmmffmmffmfmfmmmmmmfmfffffffmfmffmfmmmmfmffmmfmmfm

mmfmffmmmfffmfmmfmffmmfmmmfmmmmffmffmmmffffmmffmfffmfff

mffmmmffmmfffmm

Sample size,

120 > 20

So, we will apply normal approximation

- r = number of runs = 63
- n_1 = number of male students = 62
- n_2 = number of female students = 58

Test statistic =

$$Z_{Cal} = \frac{(r - E(r))}{\sqrt{Var(r)}}$$

Consider,

$$E(r) = \frac{1 + (2*n_1*n_2)}{n1+n2}$$

$$= 1+(2*62*58)/120$$

$E(r) = 59.94$

Now

$$Var(r) = \frac{2*n1*n2\ (2*n1*n2-n1-n2)}{(n1+n2)^2*(n1+n2-1)}$$

$$= 50861824/1713600$$

$Var(r) = 29.68$

$Z = 0.5616$

$|Z| = 0.5616$

At $\alpha = 0.05$ $Z_{tab} = Z_{\alpha/2} = 1.96$

$|Z| < Z_{\alpha/2}$

Therefore, Accept H_0.

Conclusion: - Sample is random.

CHI-SQUARE TEST OF INDEPENDENT ATTRIBUTES

H_0: Choice of social media accounts independent of gender.

H_1: Choice of social media accounts not independent of gender.

- **Observed Frequency**

	Gender		
Accounts	**Female**	**Male**	**Grand Total**
1	14	7	**21**
2	1	-	**1**
3	1	-	**1**
4	-	1	**1**
1,2	7	1	**8**
1,3	5	5	**10**
1,4	-	1	**1**
1,5	3	2	**5**
1,2,3	3	2	**5**
1,2,4	2	5	**7**
1,2,5	2	1	**3**
1,3,5	7	-	**7**
1,4,3	-	3	**3**
1,2,3,4	1	1	**2**
1,2,3,5	7	2	**9**
1,2,4,3	-	10	**10**
1,2,4,5	-	1	**1**
1,2,3,4,5	4	3	**7**
1,2,4,3,5	9	9	**18**
Grand Total	**66**	**54**	**120**

What's app = 1

Instagram = 2

Telegram = 3

Facebook = 4

Snapchat = 5

- **Expected Frequency**

	Gender		
Accounts	**Female**	**Male**	**Grand Total**
1	**11.55**	**9.45**	**21**
2	**0.55**	**0.45**	**1**
3	**0.55**	**0.45**	**1**
4	**0.55**	**0.45**	**1**
1,2	**4.4**	**3.6**	**8**
1,3	**5.5**	**4.5**	**10**
1,4	**0.55**	**0.45**	**1**
1,5	**2.75**	**2.25**	**5**
1,2,3	**2.75**	**2.25**	**5**
1,2,4	**3.85**	**3.15**	**7**
1,2,5	**1.65**	**1.35**	**3**
1,3,5	**3.85**	**3.15**	**7**
1,4,3	**1.65**	**1.35**	**3**
1,2,3,4	**1.1**	**0.9**	**2**
1,2,3,5	**4.95**	**4.05**	**9**
1,2,4,3	**5.5**	**4.5**	**10**
1,2,4,5	**0.55**	**0.45**	**1**
1,2,3,4,5	**3.85**	**3.15**	**7**
1,2,4,3,5	**9.9**	**8.1**	**18**
Grand Total	**66**	**54**	**120**

- **p - value:**

p value	**0.9642**
L.O.S. (α)	**0.05**

P value > α, Here p value greater than level of significance.

Therefore, Accept H_0.

- **Conclusion: -** Choice of social media accounts independent of gender.

ANOVA (TWO WAY CLASSIFICATION)

Age	Response	
	Yes	No
17	15	16
18	17	6
19	12	10
20	25	8
21	27	8

H_{01}: Age do not differ significantly.

$H_{02:}$ Knowledge of cyber-crime do not differ significantly.

Anova: Two-Factor Without Replication.

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
17	2	31	15.5	0.5
18	2	23	11.5	60.5
19	2	22	11	2
20	2	33	16.5	144.5
21	2	35	17.5	180.5
Yes	5	96	19.2	42.2
No	5	48	9.6	14.8

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	70.4	4	17.6	**0.44670051**	**0.772856**	**6.388233**
Columns	230.4	1	230.4	**5.84771574**	**0.072904**	**7.708647**
Error	157.6	4	39.4			
Total	458.4	9				

- **Rows: -**

  Here, $F_{cal} < F_{crit}$

  Then, Accept $H_{01.}$

- **Conclusion: -** There are no significant differences among ages.

- **Columns: -**

  Here, $F_{cal} < F_{crit}$

  Then, Accept H_{02}

- **Conclusion: -** There are no significant differences among knowledge of cyber-crime.

CHI – SQUARE TEST

$H_{0:}$ - Time spend on internet by users is independent of gender.

H_1: - Time spend on internet by users is dependent of gender.

➢ Observed Frequency: -

	Use of Internet					
Gender	**0**	**1**	**2**	**3**	**4**	**Grand Total**
Female	3	30	18	13	2	66
Male	2	19	16	10	7	54
Grand Total	**5**	**49**	**34**	**23**	**9**	**120**

➢ Expected Frequency: -

	Use of Internet					
Gender	**0**	**1**	**2**	**3**	**4**	**Grand Total**
Female	2.75	26.95	18.7	12.65	4.95	66
Male	2.25	22.05	15.3	10.35	4.05	54
Grand Total	**5**	**49**	**34**	**23**	**9**	**120**

➢ p value: -

p value	**0.308**
L.O.S. (α)	**0.05**

p value > α, Here p value greater than level of significance.

Therefore, Accept H_0.

➢ **Conclusion: -** Time spend on internet by users is independent of gender.

CORRELATION COEFFICIENT

- Correlation coefficient between age (%) and time spent on internet (%) preferred by students.

Age (%)	Time spent (%)
19.17	10.00
16.67	15.00
14.17	22.50
21.17	46.67
25.83	5.83

Correlation =	**0.1224**

- **Conclusion: -** Correlation coefficient between age and time spent on internet is positive. Then, we conclude that as age increases percentage of time spend on internet also increases.

Regression:

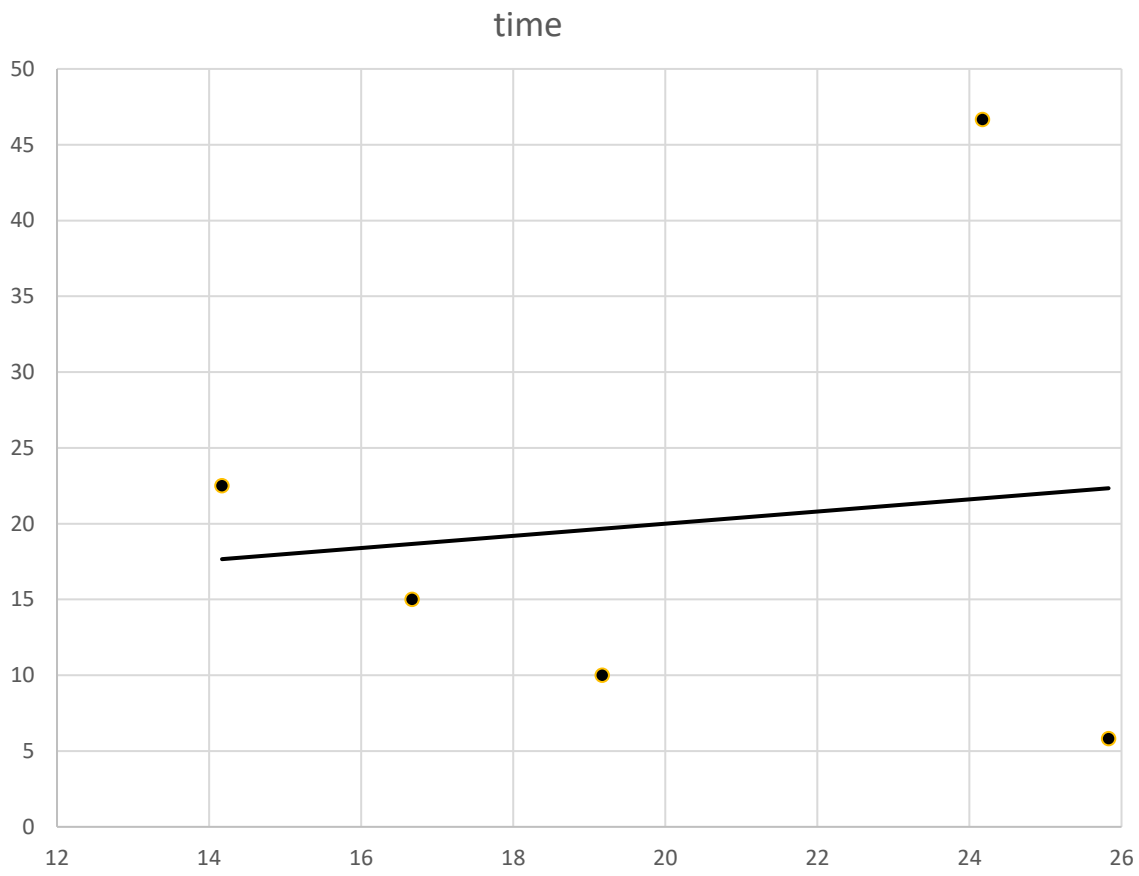
Regression Statistics	
Multiple R	0.122432825
R Square	0.014989797
Square	0.313346938
Standard Error	5.647663352
Observations	5

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.456175991	1.456176	0.04565373	0.844504017
Residual	3	95.68830401	31.8961		
Total	4	97.14448			
	Coefficients	Standard Error	t Stat	p-value	
Intercept	19.25481776	4.31367852	4.463665	0.020941438	
time spent	0.037359112	0.174847091	0.213667	0.844504017	

➢ Conclusion: -

There is significant relationship between the percentage of age and percentage of time spent on internet, r = 0.1224.



QUESTIONNAIRE

1) Name

2) Gender

- Male
- Female

3) Age

- 17
- 18
- 19
- 20
- 21

4) In which faculty you are studying?

- Arts
- Commerce and management
- Science

5) Which of the following social media accounts you use often?

- What's app
- Facebook
- Instagram
- Snapchat
- Telegram
- Other (specify)

6) How much time you spent on internet before COVID - 19?

- 1 hour
- 2 hours
- 3 hours
- 5 and more hours
- Other (specify)

7) How much time you spent on internet after COVID - 19?

- 1 hour
- 2 hours
- 3 hours
- 5 and more hours
- Other (specify)

8) How many times you found someone using your profile, photo, bank details, duplicating your personal information or hacking social media accounts?

- Never.
- 1 time.
- 2 times.
- 3 times.
- 4 times and more than 4 times.

9) How important is to have strong codeword in your social accounts?

- Important
- Not important
- I don't know

10) Are you know about cyber-crime?

- Yes
- No

11) Which of the following cyber-crimes do you know?

- Credit card frauds
- Cyber stalking
- Cyber terrorism
- Hacking
- Online frauds
- All of above
- Other (specify)

12) Do you ever visit unknown websites?

- Yes
- No
- Not sure

13) What you think, is it difficult to identify a fraudulent website?

- Agree
- Disagree
- Neutral

14) Do you have an antivirus programs on your handset?

- Yes, I have
- No
- Other (specify)

15) Have you ever lost money due to cyber-crime?

- Money got deducted from bank account
- Online frauds
- Never

16) Do you agree that cyber-crime laws in country are able to control cyber criminals?

- Agree
- Disagree
- Not sure

17) Is cyber-crime reducing by taking seminar, showing posters of awareness to students or by any other way?

- Yes
- No
- Don’t know

DATA ANALYSIS

1) Gender of the respondents?

Gender	Response (%)
Male	45
Female	55

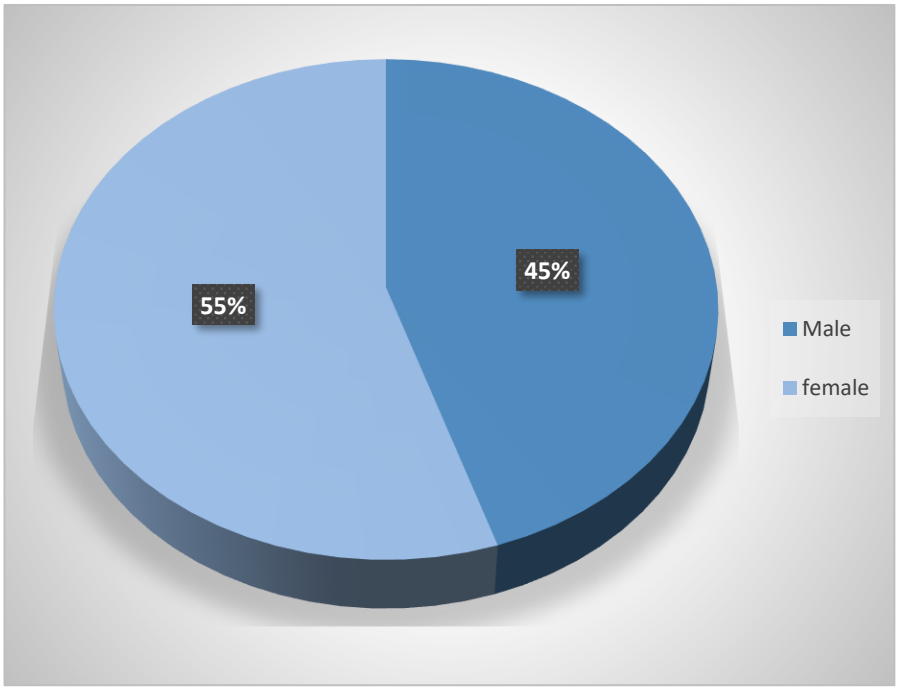


Fig 1: Gender of the respondent

Analysis: fig. 1 shows that among 120 responses whom survey was conducted 45% are male and 55% are female.

2) Social media accounts used by users?

Accounts	%
What's app	31.88
Instagram	19.35
Facebook	13.62
Telegram	19.35
Snapchat	13.90
Others	1.91

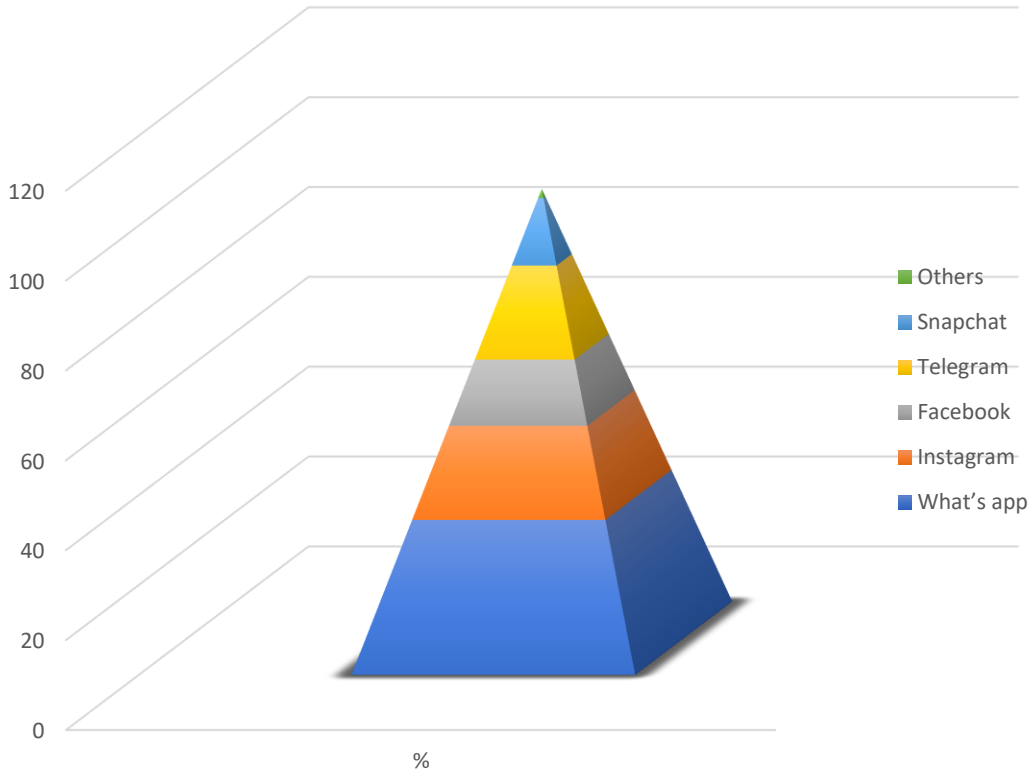


Fig.2. Social media accounts

3)Time spend on internet in COVID-19? (Before and After)

Option	Before (%)	After (%)
1 hour	40.83	10
2 hours	28.33	15
3 hours	19.17	22.50
5 and more hours	7.5	46.67
Others	4.17	5.83

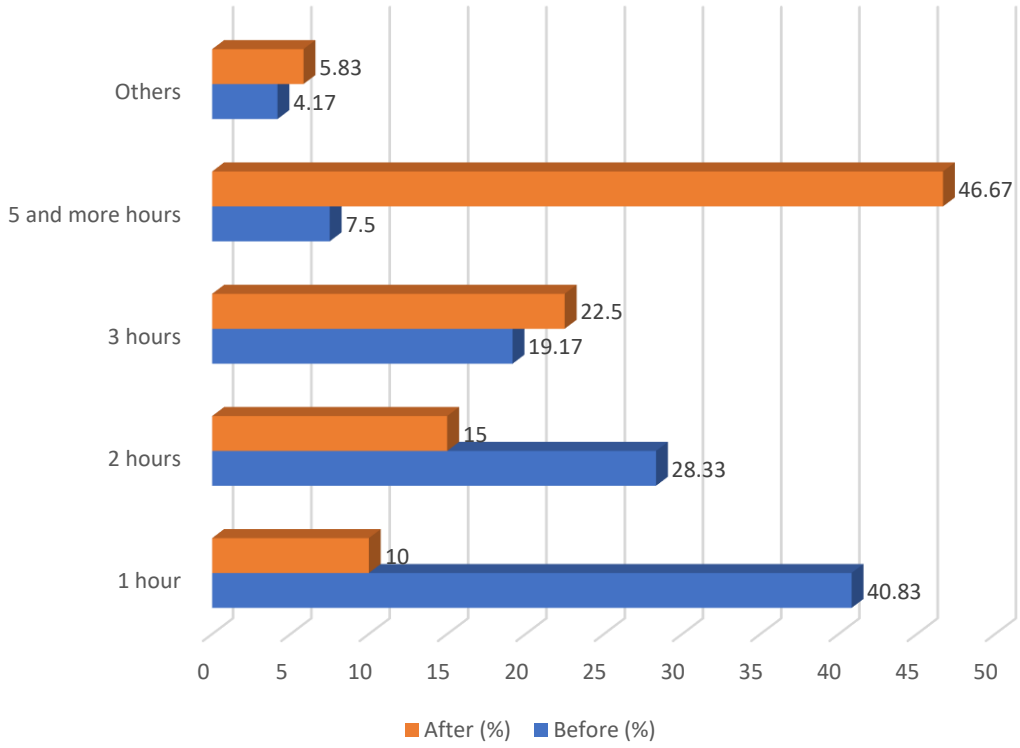


Fig.3 Time spend on internet

Analysis: - From fig.3 it shows that the no hours spend on internet is increases after COVID-19.

4) Cyber-crime rate faced by students?

Options	Counts
Never	71
1 time	17
2 times	14
3 times	7
4 times and more than 4 times	11

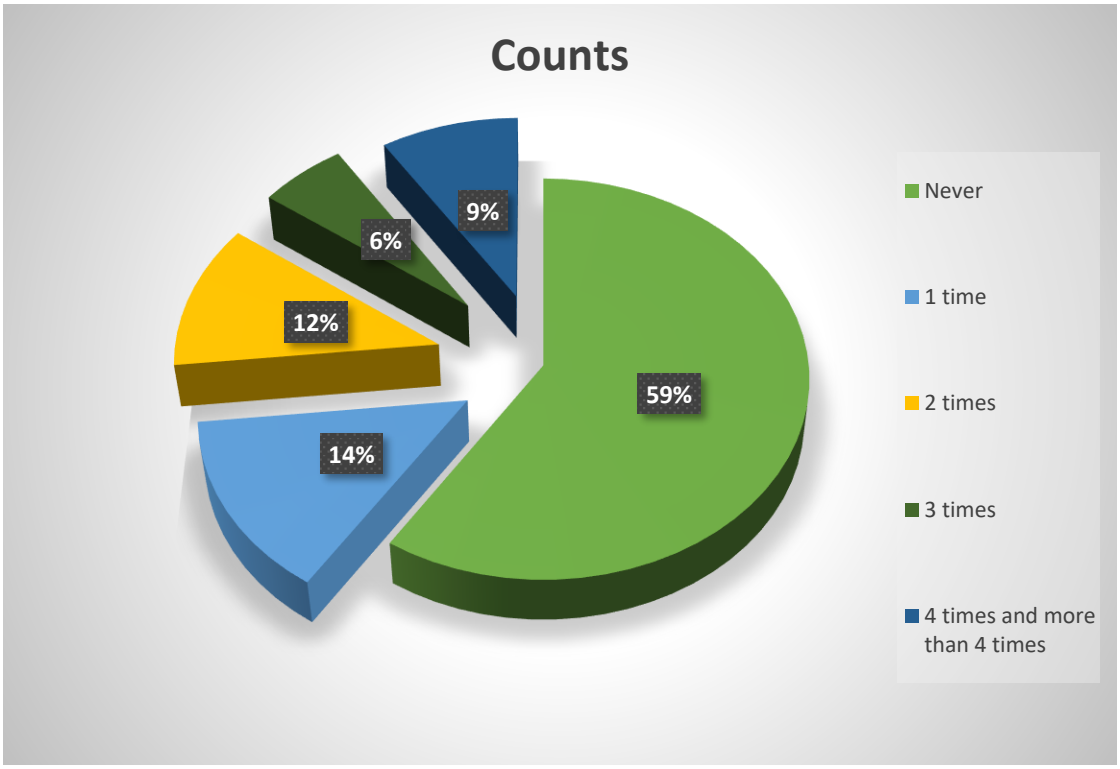


Fig.4 cyber-crime rate

5) How important is to have strong codeword in your social accounts?

Options	%
Important	72.50
Not important	13.33
I don’t know	14.17

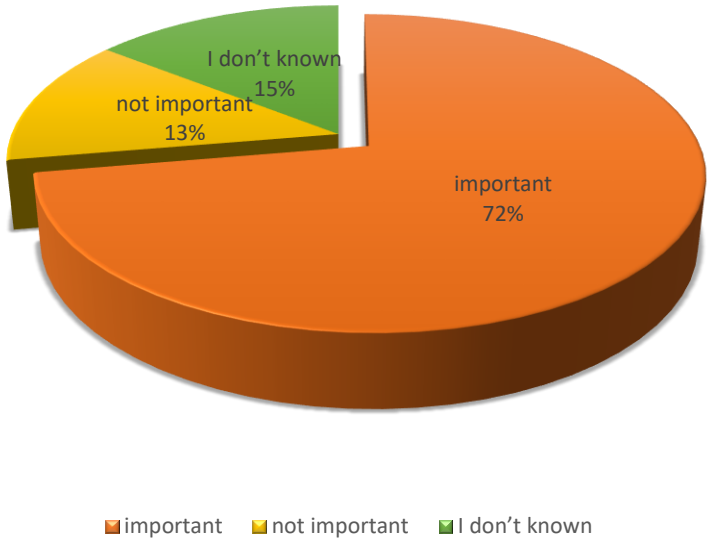


Fig.5. Importance of codeword

Analysis: - fig.5. Shows that 72% respondents agrees that it is important to have a strong codeword in their social media accounts.

6) Are you know about cyber-crime?

Options	%
Yes	80.00
No	20.00

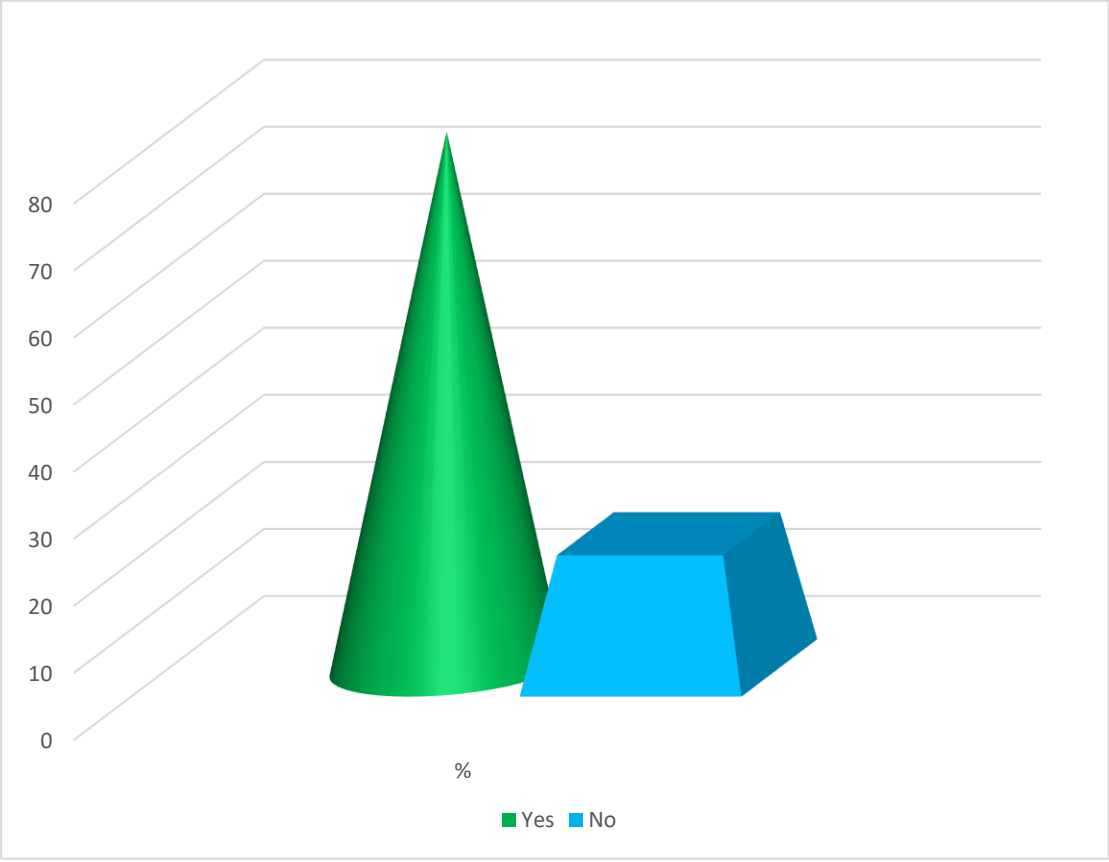
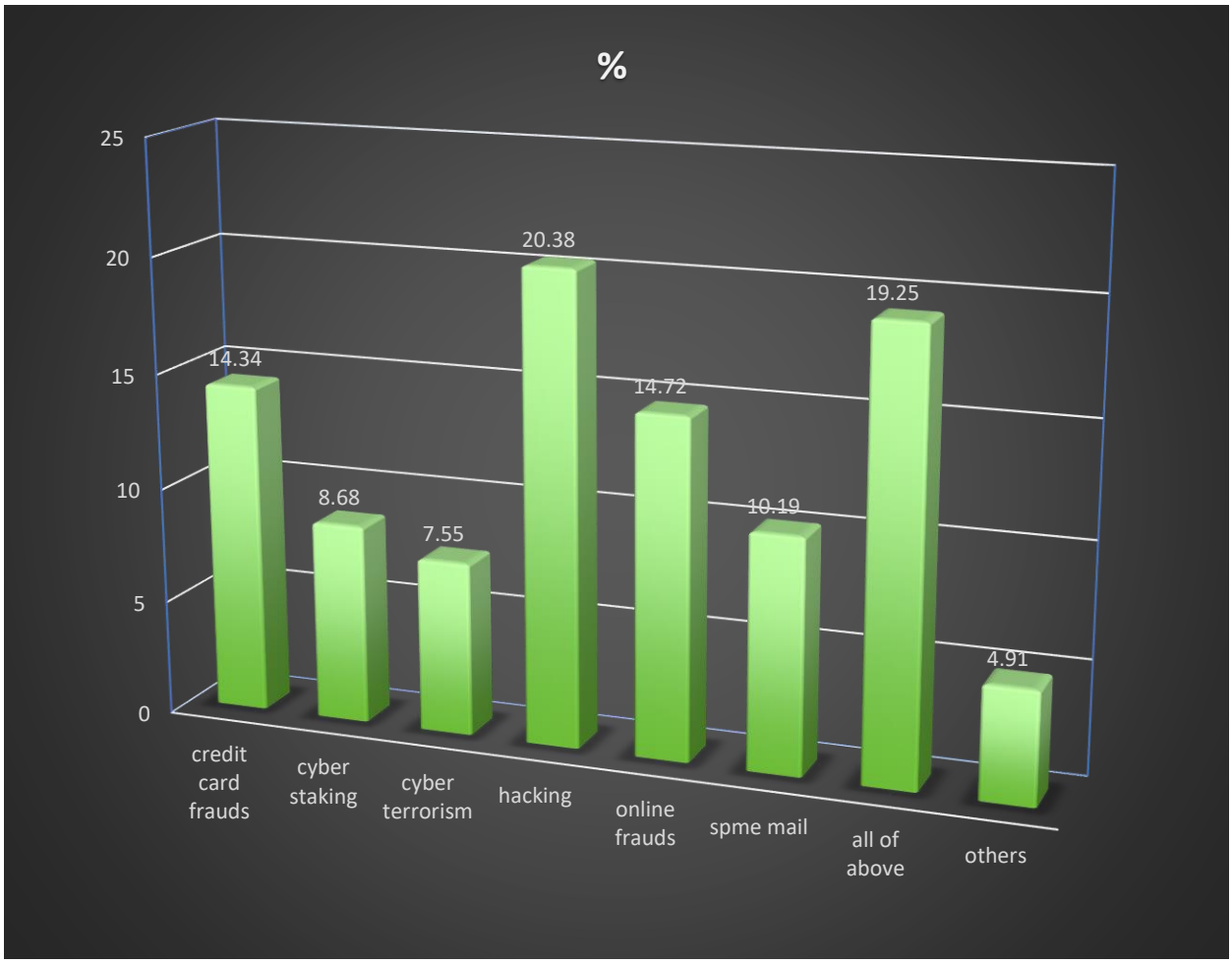


Fig.6. knowing of cyber crime

7) Types of cyber-crime known to students?

Options	%
Credit card frauds	14.34
Cyber Stalking	8.68
Cyber terrorism	7.55
Hacking	20.38
Online Frauds	14.72
Spam mails	10.19
All of the above	19.25
Others	4.91



8) Do you ever visit unknown websites?

Options	Counts
Yes	51
No	41
Not Sure	28

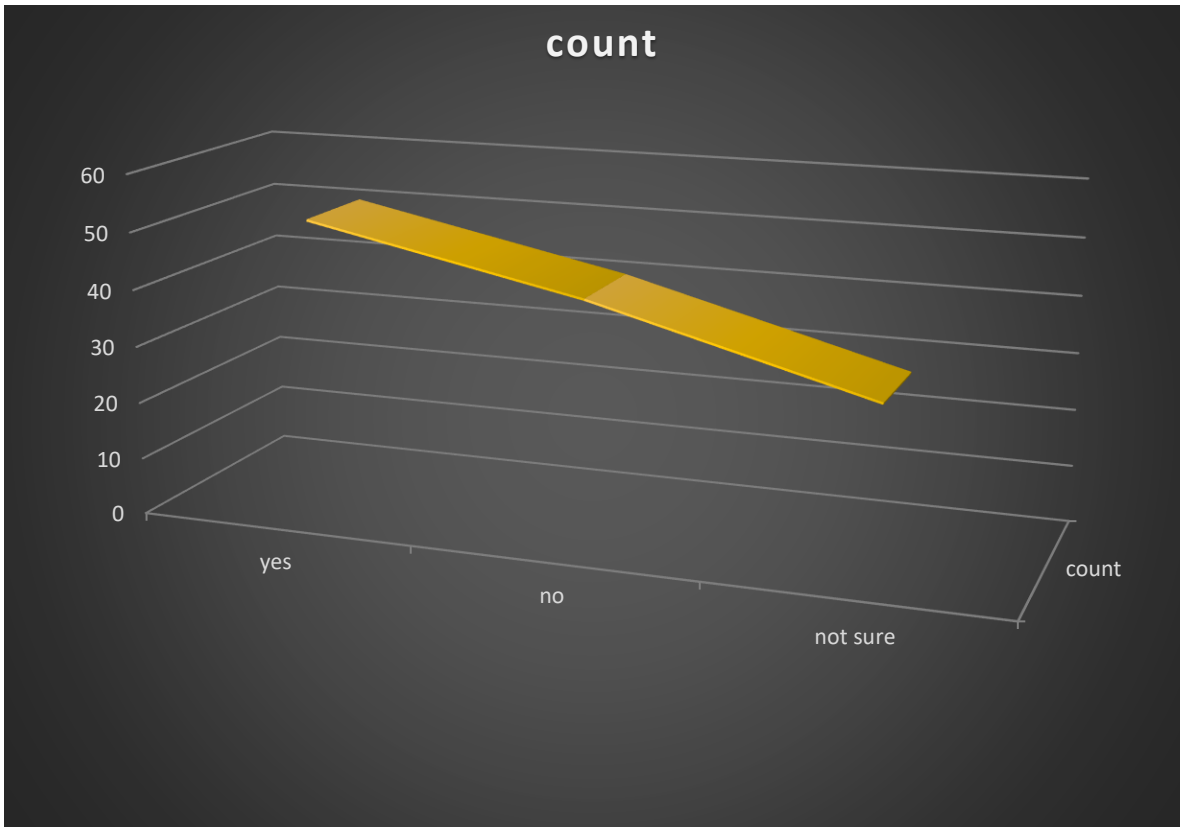


Fig.8. Visiting to unknown websites

Analysis: - fig.8 shows that the average respondents visited unknown websites.

9) Is it difficult to identify fraudulent websites?

Options	%
Agree	49.17
Disagree	21.67
Neutral	29.17

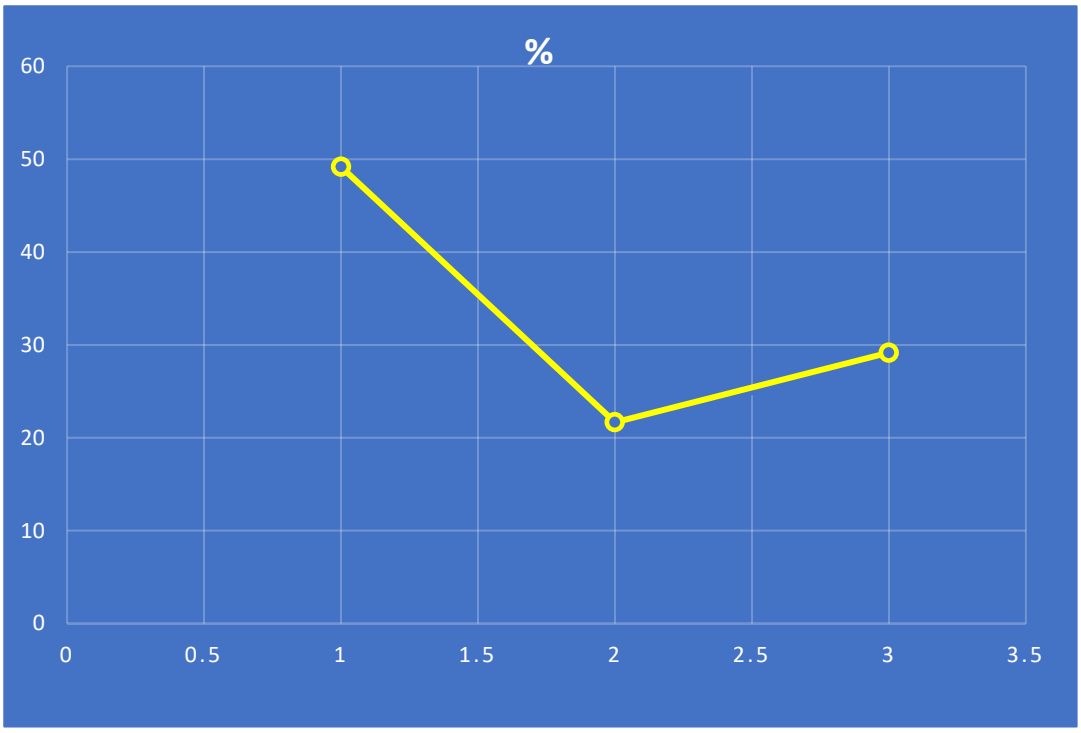


Fig.9 Identification of fraudulent websites

Analysis: - fig.9 shows that the average students agrees that it is difficult to identify fraudulent websites.

10) Do you have an antivirus programme on your handset?

Options	Counts
Yes, I have	43
No	76
others	1

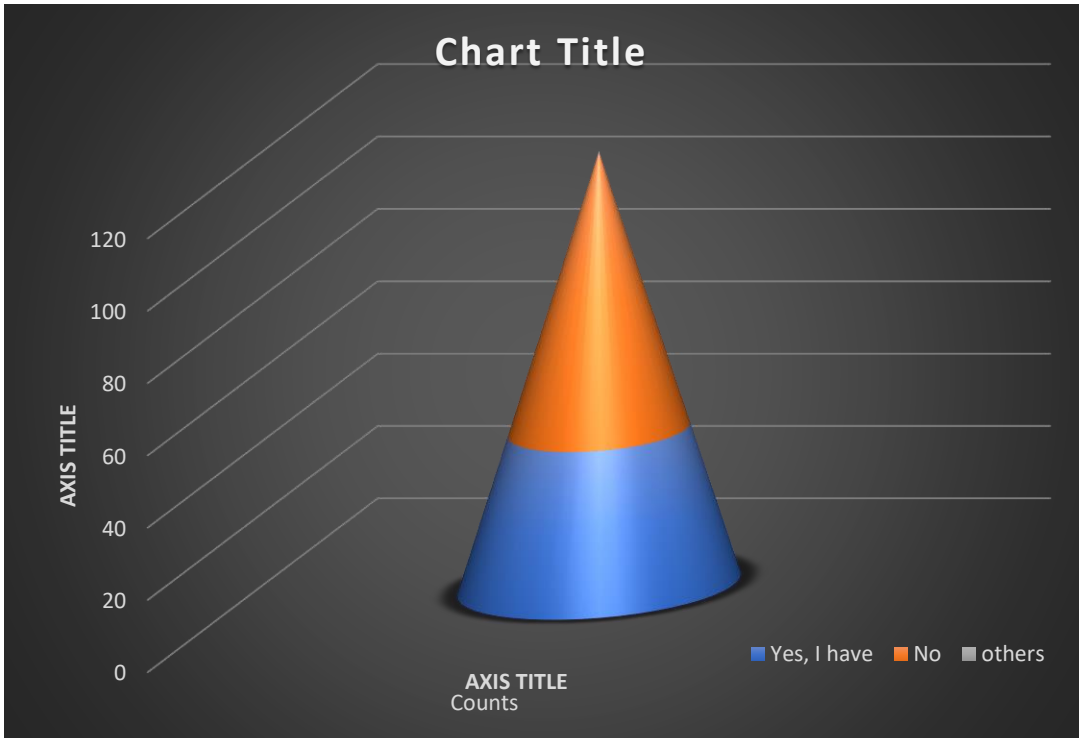


Fig.10 Antivirus programme in handset

Analysis :- fig.10. shows that, most of the students don't have an antivirus programme in there handset.

Quick heal is the best antivirus programme.

11) Have you ever lost money due to cyber-crimes?

Options	%
Money got detected from bank accounts.	16.54
n online frauds.	20.30
Never.	63.16

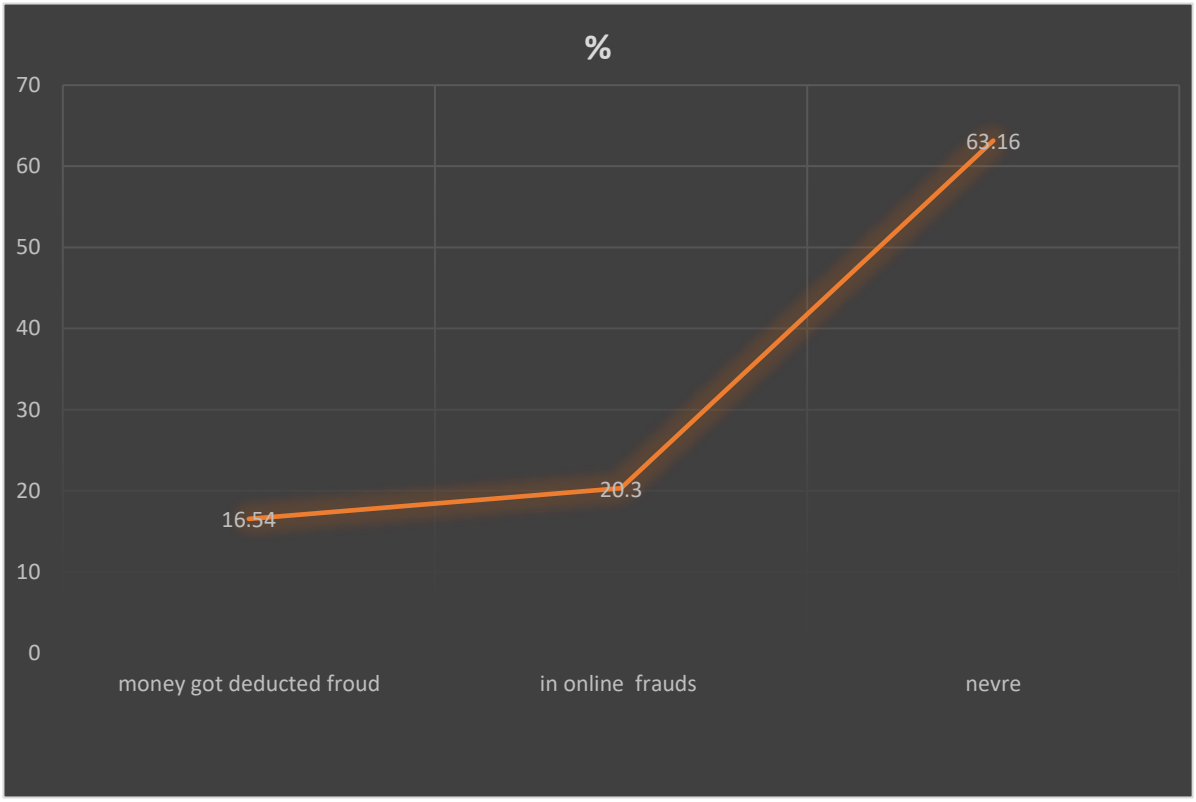


Fig.11. Lost of money

12) Do you agree that cyber-crime laws in our country are able to control cyber criminals?

Options	%
Agree	38.33
Disagree	23.33
Not sure	38.33

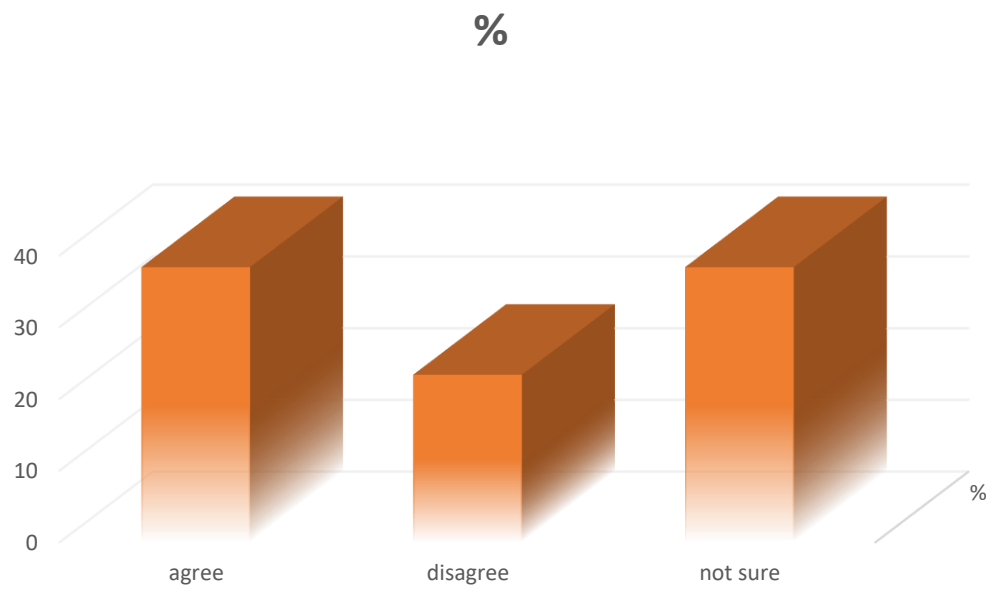
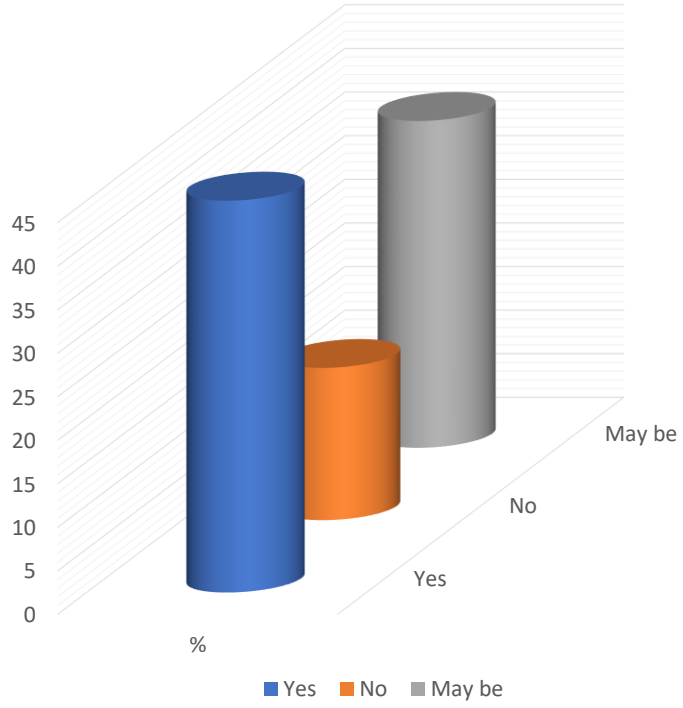


Fig.12. Opinion about Cyber Crime laws

Analysis: - fig.12 Shows that, average percent of respondents agree that cyber-crime laws are not able to control cyber criminals.

13) Is cyber-crime reducing by taking, seminar showing posters of awareness to students or by any other way?

Option	%
Yes	45.00
No	17.50
May be	37.50



Analysis: - fig.13 shows that average percent of respondents agrees that by taking seminar, showing posters of awareness or any other way incidents of cyber-crime reduced.

CONCLUSION

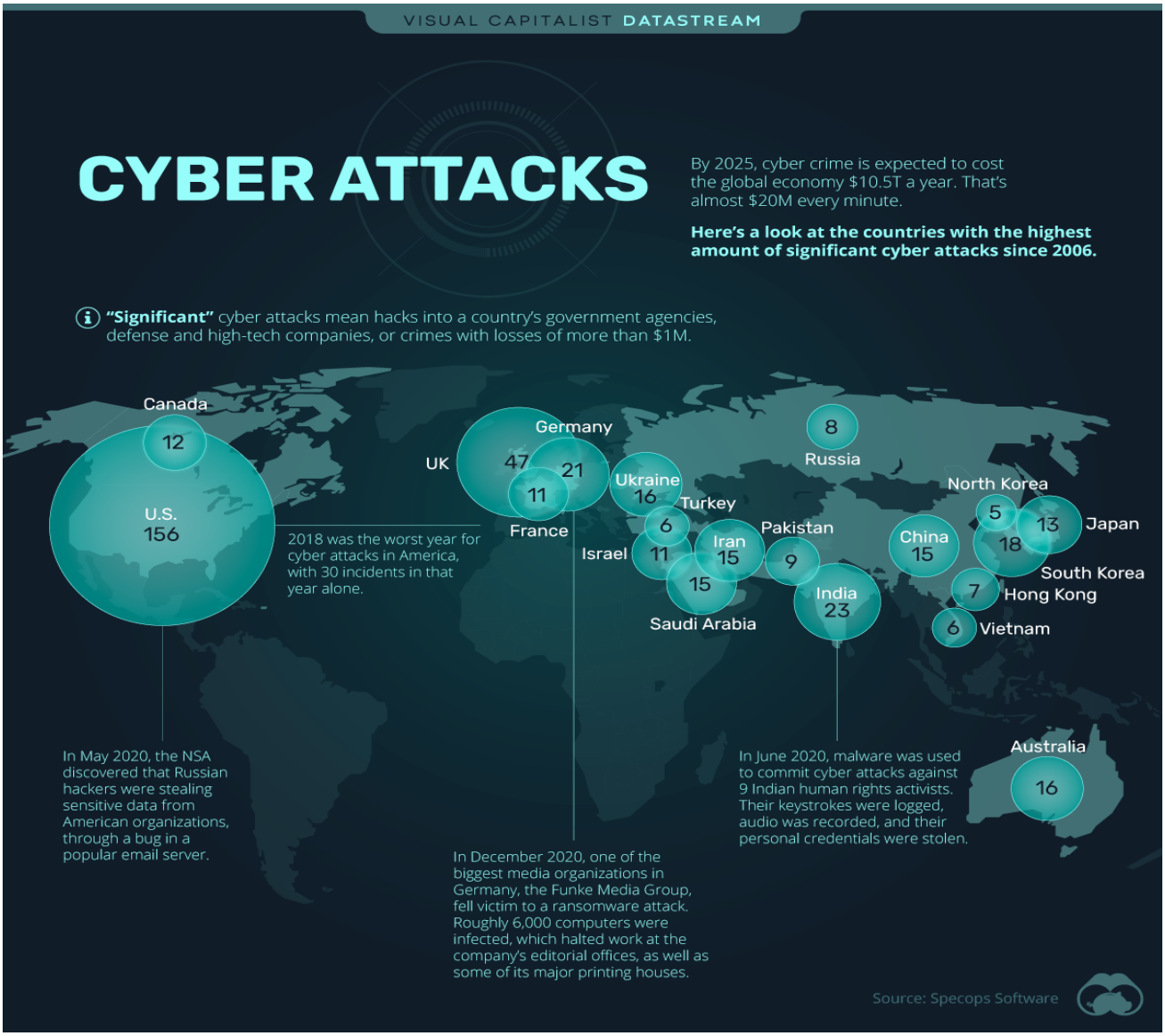
As far as the conclusion is concerned the following points are highlighted.

After online survey we found that most oftenly use social media accounts are What's app, Instagram and Telegram. As age increases percentage of use of internet also increases.

From the statistical analysis we have seen that as use of internet increases rate of cyber-crime also increases. From population of 120 respondents 80 % respondents know about cyber-crime and 20 % are unaware about cyber-crime. Average respondents are found to be sometimes victim of cyber-crime, though they didn't use any antivirus programmes in their handset.

Many respondents are known to following types of cyber-crimes, "Hacking, Credit card frauds, Cyber stalking, Online frauds." Students says that Quick Heal is the best antivirus programme to protect our laptop and pc.

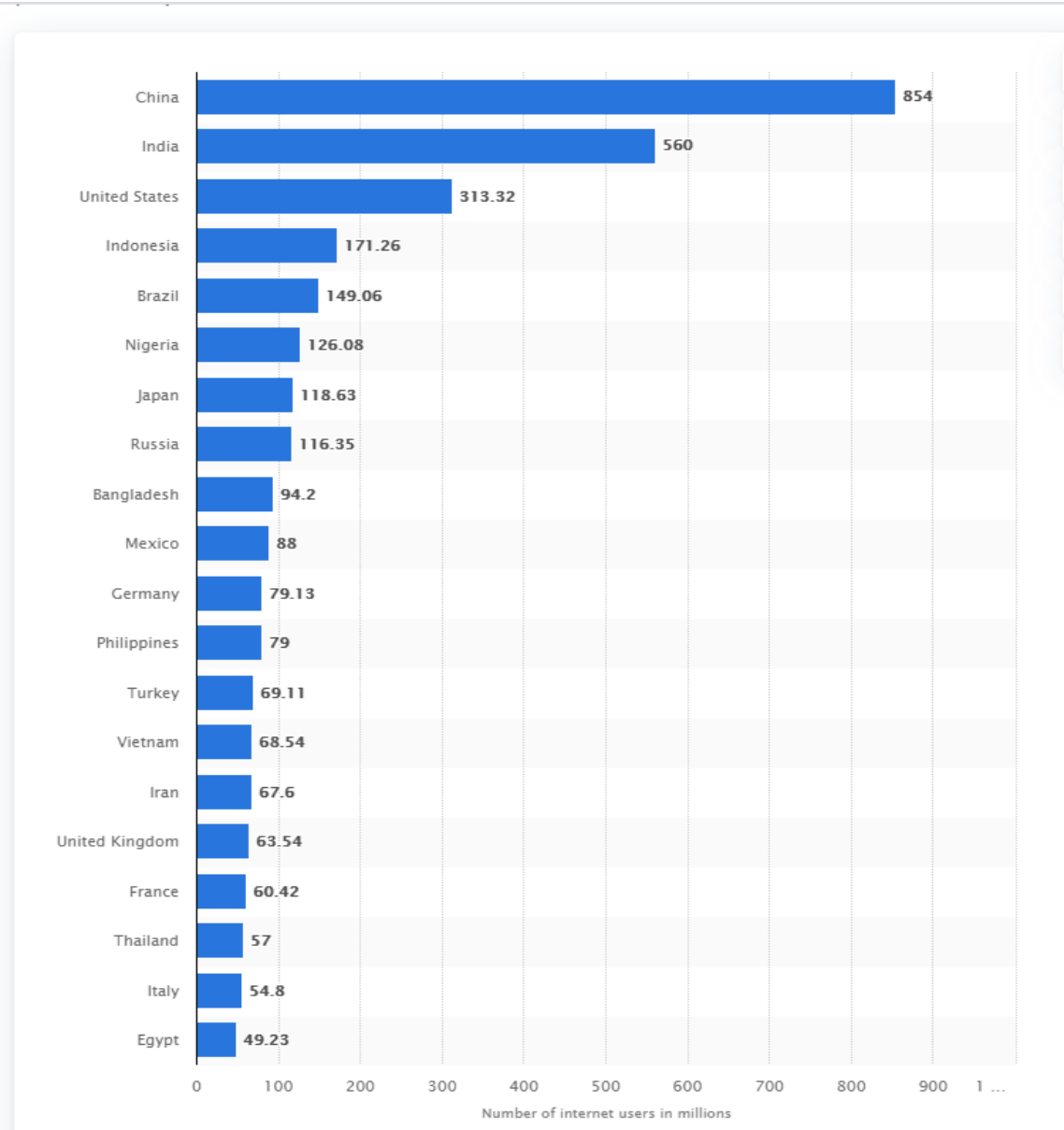
FURTHER SCOPE



❖ INDIA

- As of December 2020, China was ranked first among the countries with the most internet users and **second largest country after China was India with most internet users.**

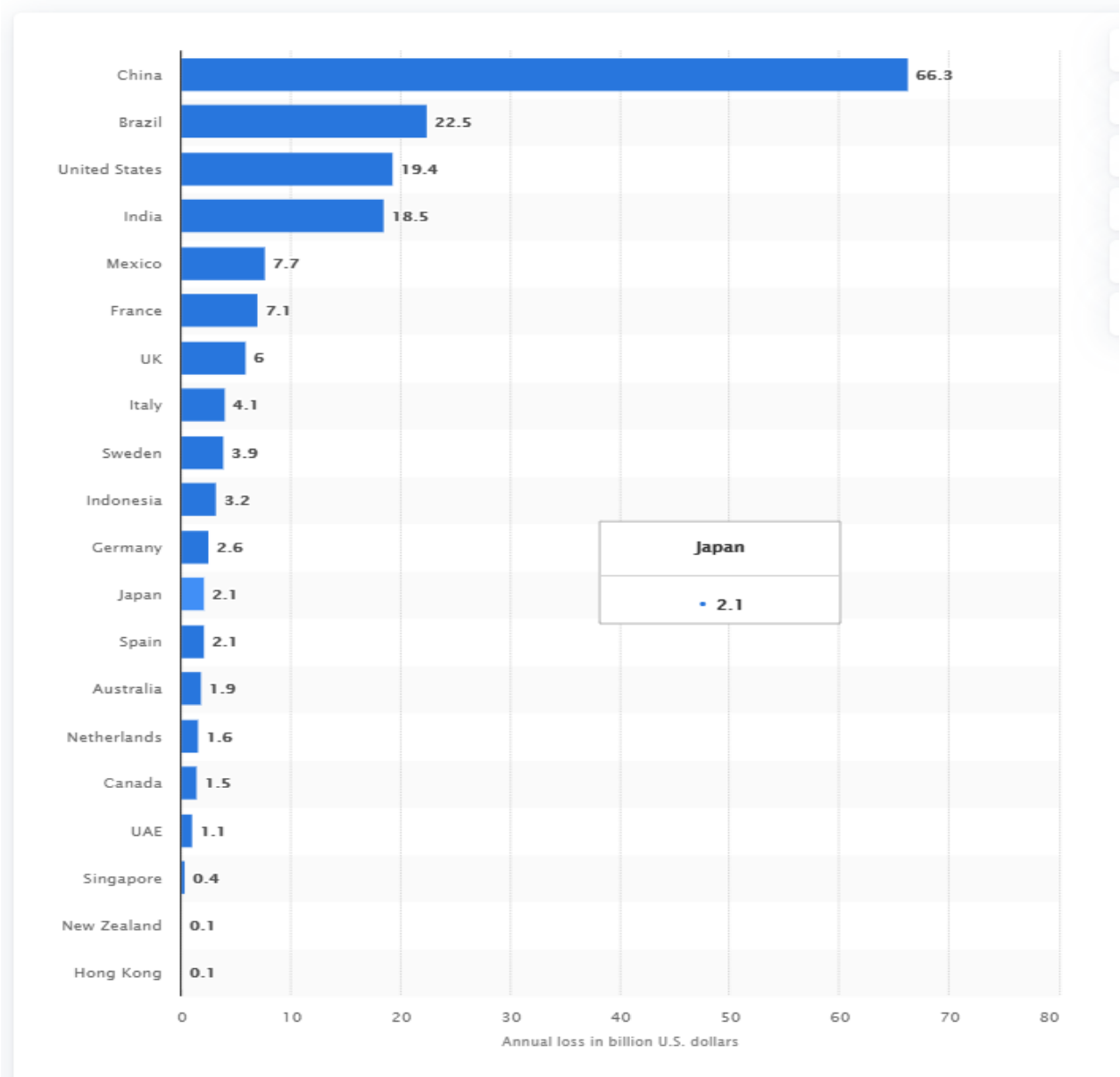
Countries with the most internet users



- India is the world's third largest internet younger users after China and the United State.

❖ CBER-CRIME IN INDIA

- The statistics presents a ranking of the countries most affected by cyber-crime in 2020. India ranked fourth in victims of cyber-crime.
- India was the second most cyber-attacked country in Asia-Pacific in 2020.



Consumer loss through cyber-crime world wide

- 52% Indian adults don't know how to protect themselves from cybercrime survey.

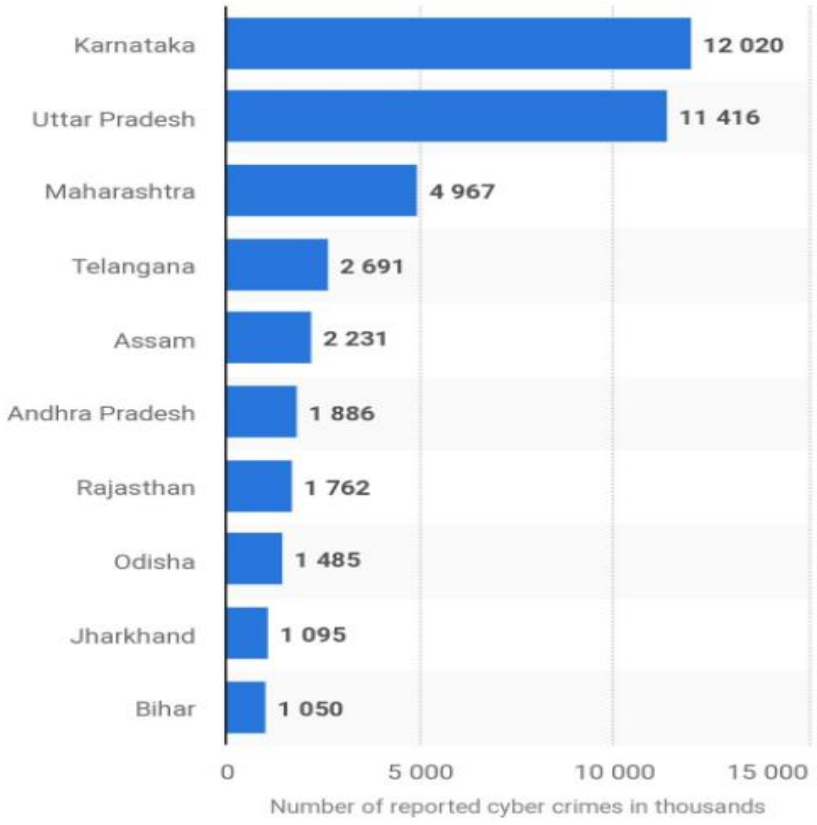
- As per the respondents 7 in 10 Indian adults (70%) believe that remote work has made it much easier for hackers and cybercriminals to take advantage of people.

Analysis: -

- India is the most affected country in the world. It has a high proportion of young people.

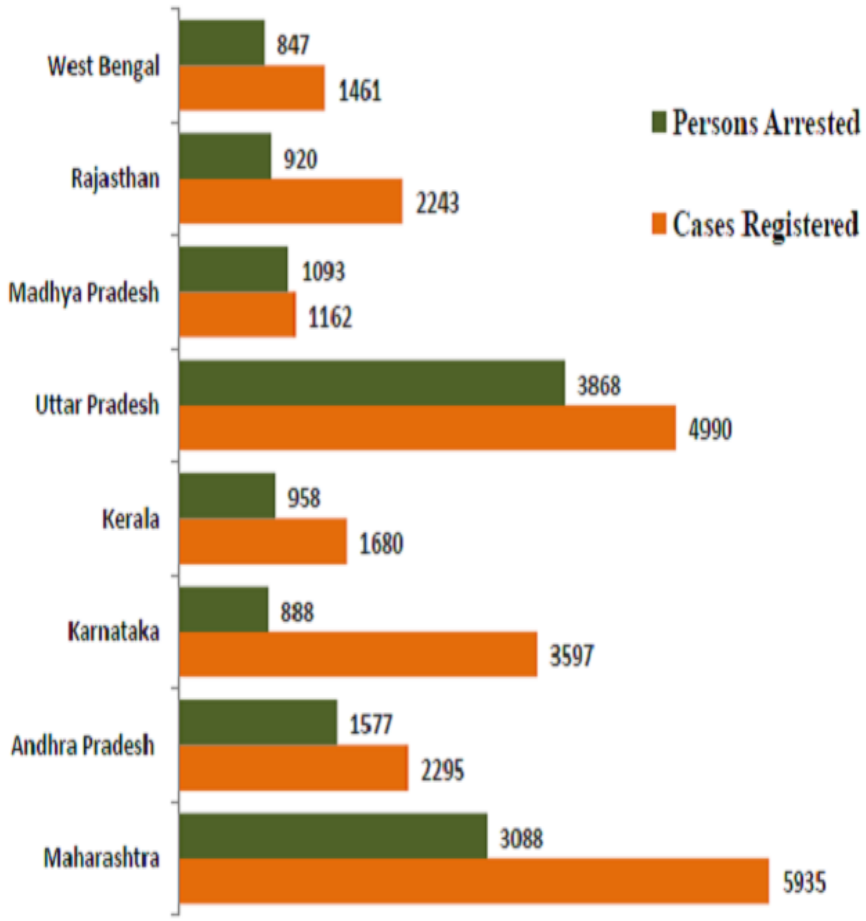
❖ STATES: -

- Karnataka and Uttar Pradesh accounted for the highest share during the measured time period. The northern state of Uttar Pradesh had the highest number of cyber-crimes compared to the rest of the country.



❖ MAHARASHTRA

- Maharashtra is included in top 5 among the states in India in cyber-crime.
- Registered cases of cyber-crime in Maharashtra are more as compared to other states in India.



THANK YOU....