Socio Economic survey of impact of COVID-19 on College students

Gokhale Education Society's

H.P.T ARTS & R.Y.K SCIENCE COLLEGE NASHIK – 422005

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the project work entitled

"SOCIO-ECONOMIC SURVEY OF IMPACT OF COVID-19 ON COLLEGE STUDENTS"

is a bonafied work carried out by,

Roll no	Name of the student
08	Gaikwad Pratiksha Sharad
15	Jagtap Subodh Prajokt
22	Kulthe Yash Pramod
25	Mogal Anisha Bhanudas
29	Sayed Areeb Nisar
35	Shukla Ekta Shyamdhar

With partial fulfilment for the statistics project of the Savitribai Phule Pune University during the year 2021-22. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Signature of the guide,

Prof. Mrs. V. S. Joshi.

Examiner

Head of Department,

Prof. Mrs. V. S. Joshi.

Acknowledgement

A project usually falls short of its expectations unless guided by the right person at the right time. Success of a project is an outcome of sincere efforts, channeled in the right direction, efficient supervision and the most valuable professional guidance. This project would not have been completed without the direct and indirect help and guidance of such luminaries. They provided us with the necessary resources and atmosphere conductive for healthy learning and training. We would like to thank Savitribai Phule Pune University for giving us an opportunity to perform the project because of which could apply the theoretical knowledge in Statistics at an undergraduate level we express our gratitude to Principal **Prof.V.N.Suryavanshi**, H.P.T ARTS AND R.Y.K SCIENCE COLLEGE, NASHIK for allowing us to present this project. At the outset we would like to guidance that we have received from **Prof**. **Mrs. V. S. Joshi**. We would also like to thank the teaching staff **Mr. Umesh. R. Yeole**, **Mrs. P. D. Dugaje**, **Ms. H. B. Jadhav**, **Mrs. J. D. Vetal**, **Ms. S. M. Pawar** and Non-teaching staff **Mr. Patil**. Without their critical evaluation and suggestion at every stage of the project, this project could not have reached its present form. Faculty has critically evaluated each step in developing this project.

We would like to extend the special thanks to our respondents who gave us fruitful information to analyze the Socio-economic survey on impact of covid-19 on college students.

And finally, the students of our college, friends and family for their support without which the project could not have been a successful one.

Heartfelt gratitude to all of you.

INDEX

INTRODUCTION.

COVID-19 has overwhelmed the entire world, and India also has borne the brunt of the same.The spread was so colossal that the World Health Organization (WHO) had to declare it as a pandemic. The only way to control and defeat this mammoth pandemic was to make people follow social distancing and also to restrain them from moving out to avoid social connect. The impact of COVID-19 was seen in every sphere of life. The emergence of COVID-19 led the world to an unprecedented public health crisis. It had created an unstable environment for individuals, loss of business activities, loss of employment, education system, usage of digital platforms, effect on food and health habits of individual, etc. This had halted a large number of economic activities and routine activities because of infectious nature.

Although COVID-19 has mostly affected the elder generation's health, it has also disproportionately affected youth (aged 30 and below) in almost every other significant aspect of their life creating a drastic change in their education, health, work, spirit, finances, emotions and overall environment resulting in us selecting of this topic.

Education is the biggest sector that has been adversely affected by this pandemic. It forced many great nations to enforce lockdown thereby bringing everything to an abrupt halt for a certain period of time. Right from the big businesses to educational institutions this pandemic has literally altered and devastated the traditional ways of carrying out these things. The educational sector has been fighting to survive the crises. The teaching learning and evaluation methodologies have been completely altered by this pandemic. The digitalization of education became a necessity in order to provide seamless education. All the educational institutions in India had to be shut down right from the primary schools to the universities. The outbreak of COVID-19 had forced many states to enforce lockdowns that brought everything to a standstill including the teaching and learning process. According to the UNESCO report more than 90% of total student population in the India was affected due to the pandemic during the initial phase of its outbreak. It has caused a serious and very deep-rooted impact on the social, economic and also psychological life of people in different parts of the world. With the closure of educational institutions, the need for a rapid transition from physical learning to the digital sphere of learning emerged.

Another effect was seen on the employment of the individual People lost their jobs, decrease in their daily wages, etc. were some factors that were observed during pandemic. Covid-19 halted a large number of economic activities because of infectious nature. Even in areas with relatively low disease risks, people were exposed to extensive communication about the risks of COVID-19, which was likely to have caused some of them stress. The measures to be adopted were the lockdowns, self-isolation, social distancing, which have a wide impact on the digital media consumption, to keep themselves occupied, people entertained, socially connected and be in touch with the friends and family. Digital platforms were the best option for people to be connected. Social networking has been evidenced to have both positive and negative effects on our youths.COVID-19 has silently and rapidly spread across India.

A 38-set questionnaire was developed, which included a variety of multiple-choice questions, Likert scale and for a few questions, the respondents were allowed to enter free texts. The survey was administered using the Google Forms platform, which requires subjects to be logged in to an e-mail account to participate in the survey, it restricted multiple entries from an individual account. The distribution of the questionnaire was conducted through the outreach of social media platforms and e-mails. Clear instructions with the google form were provided to ensure the respondent must be a student. A web-based survey was conducted to students through the medium of Google online platforms from May 9 ,2022 to May 15, 2022. The online survey questionnaire contained four subgroups:

(a) Participants were asked to describe their general demographics, such as age, the region of residence.

(b) Information about the daily online learning routine following the transition from offline learning in educational institutions in India: average time spent for online study (hours) /day;

(c) Assessment of the experience of online learning to evaluate the levels of satisfaction among students.

(d) Assessment of health due to the change in lifestyle: average time spent on sleep (hours)/day; average time spent of internet; frequency of exercise (hours)/day.

The aim of this survey study is to investigate the impact of the COVID-19 pandemic on the education, health, and lifestyle of college students from different age-groups.

OBJECTIVES.

The novel coronavirus disease, named COVID-19 on 11 February 2020, is caused by SARS-CoV-2 virus. Since a pandemic like COVID-19 has happened in the world after a century, this pandemic has led to a drastic change in people's life by resulting in a complete lockdown. First lockdown was implemented from 23th March 2019 in India leading to more lockdowns over the span of two years. This time period caused a significant and drastic change in food related and health habits. A variety of COVID-19 related psychological changes were also seen. Further impacts were seen on usage of digital platforms and every other sphere of life.

Thus resulting in our objectives –

1. To understand students' perception of COVID-19's impact on their future.
2. To analyze the impact of the ongoing health crisis on youth and their attitude or outlook towards the future prospects.
3. In this project we would be discussing the problems faced by the students who are pursuing higher education during this pandemic.
4. To find out the positive & negative impact COVID-19 had on the teaching, learning and evaluation methodologies at undergraduate level.
5. Due to this pandemic the mode of education has changed, this project analyzes the student's response towards online or offline mode of education.
6. To determine the changes in the preferred mode (online/offline) of ordering food, shopping and mode of payment pre/post pandemic.
7. To analyze whether the students tested positive or not by the means of 3 variables with the help of a decision tree.

DATA COLLECTION

- Way of collection of data and method of data collection is very important part of the project.
- We select the survey method for our project's data collection from sample survey, either primary or secondary data can be collected.

  i. Primary data:

  Primary data means original data obtained by an investigator himself/herself. Primary data is also called as raw data.

  ii. Secondary data:

  Secondary data means data taken from sources like reports or office record etc., which is already collected by some other influx.

- To investigate our project objectives we decided to collect primary data by preparing proper questionnaire. We have prepared the questionnaire considering all the rules and regulations that we have studied in topic.

RESEARCH METHODOLOGY

We discuss the methodology adopted to fulfil the objectives of the present study. The study was conducted in the Nashik district of Maharashtra. Present chapter is divided into following sub-head.

- Sampling Technique: Simple random sampling
- Sample size: 244
- Sampling unit: All college students below age 30
- Collection of data and method of enquiry: Present study was based on the primary data. The data was collected from the respondents through Google Forms.
- Period of enquiry: 1 week
- Data collection tools: Questionnaire consisted of multiple-choice question.

QUESTIONNAIRE

Email:

____

1. Name :

____

2. Gender :

a) Male.   b) Female.

3. Age Group :

a) Below 18   b) 18-24   c) 24-27   d) 27-30

4. Educational Qualification (Pursuing) :

a) HSC   b) Under Graduate   c) Post Graduate   d) Other

5. Stream :

a) Science   b) Commerce   c) Arts

6. Annual family income?

a) Below 1 lakh   b) 1 lakh – 2.5 lakh   c) 2.5 lakh – 5 lakh

d) 5 lakh – 7lakh   e) Above 7 lakh

7. Average monthly family expenditure?

a) Below 10k   b) 10k – 30k   c) 30k – 50k

d) 50k – 70k   e) Above 70k

8. Where do you stay?

    a) P.G.   b) Hostel   c) With Parents/ Guardian

9. During the lockdown where did you live?

    a) P.G.   b) Hostel   c) With Parents/ Guardian

10. Have you tested positive for Covid-19?

    a) Yes   b) No

11. If yes, how many times?

    a) Once   b) Twice   c) More than twice

12. Have you suffered a loss in your immediate family due to Covid-19?

    a) Yes   b) No

13. Did you start working after Covid-19?

    a) Yes   b) No

14. If yes, Do you work _________

    a) Part time   b) Full time

15. Which mode of education do you prefer?

    a) Online   b) Offline

16. If online, Reasons for preference ______

    a) Accessibility of time and place.

b) Convenience and flexibility.
c) Affordability.
d) Efficiency.
e) Other.

17.If offline, Reason for preference ______

a) Less distraction.
b) Better learning environment.
c) Less gadget addiction
d) Better student - teacher interaction
e) Other

18.Did you attend online lectures regularly?

a) Yes b) No

19.Are MCQ pattern exams beneficial to students from academic point of view?

a) Yes b) No

20.How many hours per day did you study?

(i) Before COVID:

a) Less than 1 hour b) 1 hour - 2 hour c) 2 hour - 4 hour
d) 4 hour - 6 hour e) more than 6 hour

(ii) During COVID:

a) Less than 1 hour b) 1 hour - 2 hour c) 2 hour - 4 hour
d) 4 hour - 6 hour e) more than 6 hour

(iii) After COVID:

a) Less than 1 hour b) 1 hour - 2 hour c) 2 hour - 4 hour
d) 4 hour - 6 hour e) more than 6 hour

21. How would you rate the following :
(1 being the lowest rating and 5 being the highest rating)

(i) Understanding in online mode –

a) 1 b) 2 c) 3 c) 4 d) 5

(ii) Understanding in offline mode –

a) 1 b) 2 c) 3 c) 4 d) 5

(iii) Online exams –

a) 1 b) 2 c) 3 c) 4 d) 5

(iv) Offline exams –

a) 1 b) 2 c) 3 c) 4 d) 5

22.Did Covid-19 have a negative impact on your physical health?

a) Yes b) No

23. How did Covid-19 impact your mental health?

a) Positively b) No impact c) Negatively

24.Thinking of the situation since Introduction of Covid-19 restrictions, How has the
following things changed for you?
1] Increased 2] No change 3] Decreased

(i) Consumption of junk food and sweet –

a) 1 b) 2 c) 3

(ii) Consumption of fresh vegetables and fruits –

a) 1 b) 2 c) 3

(iii) Frequency of exercising –

a) 1 b) 2 c) 3

(iv) Stress level –

a) 1 b) 2 c) 3

(v) CGPA / Percentage –

a) 1 b) 2 c) 3

(vi) Time spent on internet –

a) 1 b) 2 c) 3

(vii) Average sleep time –

a) 1 b) 2 c) 3

25.How stressed are you about your future after Covid-19?

a) Not at all b) Somewhat c) Very d) Extremely

26.Preferred mode for shopping :

(i) Before pandemic:

a) Mostly online b) Mostly offline c) Both

(ii) After pandemic:

a) Mostly online b) Mostly offline c) Both

27.Preferred mode for ordering food :

(i) Before pandemic:

a) Online b) Dine-in c) Take out d) Home made

(ii) After pandemic:

a) Online b) Dine-in c) Take out d) Home made

28.Preferred mode of payment :

(i) Before pandemic:

a) Cash b) BHIM / PayTm / Google Pay c) Net banking
d) Debit / Credit Card

(ii) During pandemic:

a) Cash b) BHIM / PayTm / Google Pay c) Net banking
d) Debit / Credit Card

(iii) After pandemic:

a) Cash b) BHIM / PayTm / Google Pay c) Net banking
d) Debit / Credit Card

29. Did Covid-19 affect your mode of transportation?

a) Yes b) No

30.After Covid-19 which mode of transportation do you prefer?

a) Cycle b) Auto – Rickshaw c) Bus d) Private vehicle

31. Which platform do you prefer for watching movies?

a) T.V b) Downloaded c) OTT d) Theatre

32.If time spent on internet increased, then for what purpose?

a) Gaming b) Social media c) Work d) Educational purpose e) Other

33.Are you aware of currents affairs?

a) Yes b) No

34.If yes, then which of the following source of information do you use stay informed?

a) Newspaper b) Radio c) Television
d) Social media e) Mainstream media f) Family, friends and colleagues

35.Did you learn any new skills during Covid-19?

a) Yes b) No

If yes, then please mention –

____

36.How concerned are you about each of the following impacts of Covid-19?
1] Not at all 2] Somewhat 3] Very 4] Extremely

(i) Maintaining social ties

a) 1 b) 2 c) 3 d) 4

(ii) Impact of interrupted education

a) 1 b) 2 c) 3 d) 4

(iii) How satisfied are you with your life these days?

a) 1 b) 2 c) 3 d) 4

37. In your opinion, which aspect of your life did Covid-19 have a great impact on?

____

38. As the restrictions have been lifted, what precautions do you think should be implemented to prevent another pandemic?

____

CODING

1. Gender	Coding
Male	1
Female	0

2.Age Group	Coding
Below 18	1
18 – 21	2
21 – 24	3
24 – 27	4
27 – 30	5

3. Stream	Coding
Science	1
Commerce	2
Arts	3

4.Education Qualification (Pursuing):	Coding
HSC	1
Under Graduate	2
Post Graduate	3
Other	4

5.Annual family income ?	Coding
Below 1 lakh	1
1 lakh - 2.5 lakh	2
2.5 lakh - 5 lakh	3
5 lakh - 7 lakh	4
Above 7 lakh	5

6.Average monthly family expenditure?	Coding
Below 10k	1
10k - 30k	2
30k - 50k	3
50k - 70k	4
Above 70k	5

7.Where do you stay ?	**Coding**
P.G.	1
Hostel	2
With Parents/ Guardian	3

8.During the lockdown where did you live ?	**Coding**
P.G.	1
Hostel	2
With Parents/ Guardian	3

9. Have you tested positive for Covid-19?	**Coding**
Yes	1
No	0

10. If yes, How many times?	**Coding**
Once	1
Twice	2
More than twice	3

11. Have you suffered a loss in your immediate family due to Covid-19?	**Coding**
No	0
Yes	1

12. Did you start working after Covid-19?	**Coding**
No	0
Yes	1

13.If yes, Do you work __________	**Coding**
Part time	0
Full time	1

14.Which mode of education do you prefer?	**Coding**
Online	0
Offline	1

15.If online, Reasons for preference ______	**Coding**
Accessibility of time and place	1
Convenience and flexibility	2
Affordability	3
Efficiency	4
Other	5

16.If offline, Reason for preference ____	**Coding**
Less distraction	1
Better learning environment	2
Less gadget addiction	3
Better student - teacher interaction	4
Other	5

17.Did you attend online lectures regularly ?	**Coding**
No	0
Yes	1

18.Are MCQ pattern exams beneficial to students from academic point of view ?	**Coding**
No	0
Yes	1

20. How many hours per day did you study?	
Before COVID	**Coding**
Less than 1 hour	1
1 hour - 2 hour	2
2 hour - 4 hour	3
4 hour - 6 hour	4
more than 6 hour	5
During COVID	**Coding**
Less than 1 hour	1
1 hour - 2 hour	2
2 hour - 4 hour	3
4 hour - 6 hour	4
more than 6 hour	5
After COVID	**Coding**

Less than 1 hour	1
1 hour - 2 hour	2
2 hour - 4 hour	3
4 hour - 6 hour	4
more than 6 hour	5

21.How would you rate the following:	
Understanding in online mode -	**Coding**
Lowest	1
Highest	5
Understanding in offline mode -	**Coding**
Lowest	1
Highest	5
Online exams -	**Coding**
Lowest	1
Highest	5
Offline exams -	**Coding**
Lowest	1
Highest	5
Online exams -	**Coding**
Lowest	1
Highest	5

22.Did Covid-19 have a negative impact on your physical health ?	**Coding**
No	0
Yes	1

23.How did Covid-19 impact your mental health?	**Coding**
Positively	1
No impact	2
Negatively	3

24. Thinking of the situation of COVID -19 restrictions , how has the following things changed for you:	
Consumption of junk food and sweet -	**Coding**
Increased	1
No change	2
Decreased	3
Consumption of fresh vegetables and fruits-	**Coding**
Increased	1
No change	2
Decreased	3
Frequency of exercising -	**Coding**
Increased	1
No change	2
Decreased	3

Stress level -	**Coding**
Increased	1
No change	2
Decreased	3
CGPA / Percentage -	**Coding**
Increased	1
No change	2
Decreased	3
Time spent on internet -	**Coding**
Increased	1
No change	2
Decreased	3
Average sleep time -	**Coding**
Increased	1
No change	2
Decreased	3

25.Preferred mode for Shopping:	
Before Pandemic -	**Coding**
Mostly online	1
Mostly offline	2
Both	3
After Pandemic -	**Coding**
Mostly online	1
Mostly offline	2
Both	3

26. Preferred mode of payment -	
Before pandemic -	**Coding**
Cash	1
BHIM / PayTm / Google Pay	2
Net banking	3
Debit / Credit Card	4
During pandemic -	**Coding**
Cash	1
BHIM / PayTm / Google Pay	2
Net banking	3
Debit / Credit Card	4
After pandemic -	**Coding**
Cash	1
BHIM / PayTm / Google Pay	2
Net banking	3
Debit / Credit Card	4

27.Did Covid-19 affect your mode of transportation ?	**Coding**
No	0
Yes	1

28.After Covid-19 which mode of transportation do you prefer ?	**Coding**
Cycle	1

Auto – Rickshaw	2
Bus	3
Private vehicle	4

29. Which platform do you prefer for watching movies ?	**Coding**
T.V.	1
Downloaded	2
OTT	3
Theatre	4

30.If time spent on internet increased, then for what purpose?	**Coding**
Gaming	1
Social media	2
Work	3
Educational purpose	4
Other	5

31.Are you aware of currents affairs ?	**Coding**
No	0
Yes	1

32.If yes, then which of the following source of information do you use stay informed?	**Coding**
Newspaper	1
Radio	2
Television	3
Social media	4
Mainstream media	5
Family, Friends, Colleagues	6

33. Did you learn any new skills during Covid-19?	**Coding**
No	0
Yes	1

34. How concerned are you about each of the following impacts of Covid-19 ?	
Maintaining social ties -	**Coding**
Not at all	1
Somewhat	2
Very	3
Extremely	4
Impact of interrupted education -	**Coding**
Not at all	1
Somewhat	2
Very	3
Extremely	4
How satisfied are you with your life these days?	**Coding**
	1
Somewhat	2
Very	3
Extremely	4

35.How stressed are you about your future after Covid-19?	**Coding**
Not at all	1
Somewhat	2
Very	3
Extremely	4

STATISTICAL TOOLS USED:

(i) Graphical representation

   (a) Simple bar diagram.
   (b) Multiple bar diagram.
   (c) Sub-divided bar diagram.
   (d) Histogram.
   (e) Pie chart.

(ii) Proportion test.

(iii) Chi-square test.

(iv) F-test.

(v) t-Test and Mann Whitney U-test.

(vi) Logistic regression.

(vii) Decision tree.

(viii) Anova and Kruskal-Wallis test.

THEORY OF STATISTICAL TOOLS USED:

(i) Graphical representation

Graphical Representation is a visual display of data and statistical results. It is often more effective than presenting the data in tabular form. There are many different types of graphical representations which is used depending upon the nature of data and type of the statistical results. It is very effective way to serve the purpose of comparison at a glance and revealing the patterns in the data. Graphs and diagrams are easy to understand and create an effect. Graphs and charts are often used to ease understanding of large quantities of data and relationships between parts of the data. Graphs can usually be read more quickly than the raw data that they are produced form. They are used in wide variety of fields and can be created by hands often on graph papers or by Computer using a chart application. Therefore, Graphs and Charts are believed to be powerful tools to convey information. The different types of graphical representation used in this project are:

- **Simple bar diagram:**
  This is the simplest way of presenting the statistical data classified according to single characteristics. It can be used to present the data like population of different cities, exports of different cities, exports of different countries, etc. In general, it can be used for representing any single series but generally it is used to show the categorical series.

- **Multiple Bar Diagram:**
  A multiple bar diagram is used for two or three-dimensional comparison. For comparison of magnitudes of one variable in two or three aspects or comparison of magnitudes of two or three variables, rectangles in a group are placed side by side.

- **Sub-divided diagram:**
  A subdivided bar diagram is a way of representing data in which the total magnitude is divided into different segments. In this diagram, first of all, we draw the simple bars for each class taking the total magnitude in that class then we divide that bar into segments of its various components. Sub-divided bar diagrams are those diagrams which simultaneously present, total values as

well as part values of a set of data. Different parts of a bar must be shown in the same order for all bars of a diagram.

- **Histogram:**
  A histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size. In the most common form of histogram, the independent variable is plotted along the horizontal axis and the dependent variable is plotted along the vertical axis.

- **Pie Chart:**
  It is a special type of diagram used to represent the whole quantity by a circle and the sub-division of the whole quantity is shown by the sectors of that circle. This diagram is a two-dimensional diagram.

(ii) Proportion Test (Testing of equality of two proportions (P_1=P_2)):

Suppose we draw two samples. Suppose these samples give proportions of specific items as P1 and P2 respectively. One may be interested in knowing that the population proportions from which these samples are chosen are same. In other words, we want to whether difference between two sample proportions is negligible and it has arisen merely due to sampling variations.

Let, P_1=proportion of specific items in first population

P_2=proportion of specific items in second population

n_1=size of sample drawn from first sample

n_2=size of sample drawn from second sample

X_1-Number of items of specific type in first sample

X_2-Number of items of specific type in second sample

P_1=X1/n1-proportion of specific items in first sample

P_2=X2/n2= proportion of specific items in second sample

The hypothesis for such problems will be:

Ho: P1=P2 versus H1:$P_1 \neq P_2$

R commands for null hypothesis $H_o:P_1=P_2$

(a) Consider the alternative hypothesis : $H_1:P_1=P_2$

prop.test (x,n,conf.level=c)

(b) Consider the alternative hypothesis $H_1:P_1>P_2$.

prop.test (x,n,conf.level=c alternative="greater")

(c) Consider the alternative hypothesis $H_1:P_1<P_2$.

prop.test(x,n,conf.level=c alternative="less")

Decision: Reject H_0 at α % l.o.s if p-value is less than l.o.s, otherwise accept it.

Has χ2 distribution with 1 d.f.

(iii) Chi-square test for independence of two attributes.

Suppose that the given data is classified into r-levels of attributes A denoted as A_1,…..,A_r and s levels of attribute B represented by B_1,…..,$B_{s.}$

Then different class frequencies can be represented in the following tabular form:

B A	B1	B2	…	Bi	…	Bs	Total
A1	O11	O12	…	O1i	…	O1s	(A1)
A2	O21	O22	…	O2j	…	O2s	(A2)
…	…	…	…	…	…	…	…
Ai	Oi1	Oi2	…	Oij	…	Ois	(Ai)
…	…	…	…	…	…	…	…
Ar	Or1	Or2	…	Orj	…	Orj	(Ar)
Total	(B2)	(B2)	…	(Bj)	…	(Bs)	N

This table is known as (r x s) contingency table.

N=$\sum\sum O_{ij}$= Total observed frequency

(Ai)= $\sum O_{ij}$= Total of observed frequencies in i^{th} row;   i=1,2,…,r

(Bj)= $\sum O_{ij}$= Total of observed frequencies in j^{th} row;   j=1,2,…,s

Here, Hypothesis under consideration is,

H_0: Two attributes A and B are independent

v/s

H_1: Two attributes A and B are not independent.

$E_{ij=}(A_i)(B_j)$/N ;   i=1,2,….,r ;   j=1,2,…..,s

The test statistics under H_o is,

$\chi 2 = \sum\sum (O_{ij}-e_{ij})^{2/eij}=\sum\sum (O_{ij}^{2}-e_{ij})-N$

Criteria: (1) Reject H_o at α% l.o.s if $\chi 2_{r-s-1} \geq \chi 2_{(r-s-1),\alpha}$, otherwise accept it.

(2) Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

Yates Correction: If in 2x2 contingency table any cell frequency is less than any cell frequency is less than 5 then the statistics correctly in specific way. This correction is due to Yate's and hence is known as Yates Correction Modified Formula obtained by him in this case it is as follows

$\chi_1^2 = N[(ad-bc) -(N/2)]^2 / [(a+b) (c+d) (a+c) (b+d)]$

Hence, whenever a cell frequency is less than 5 Yate's corrected formula given above should be used accept test procedure rejects H_o if $\chi_1^2 \geq \chi^2_{1, \alpha}$ at 100 α % l.o.s. otherwise accept it.

In R software following command is used for performing the test:

chisq.test(y, conf. level =, correct =)

(iv) F-test

We consider here F-test for equality of two variances. In this test we test the null hypothesis,

H_0: $\sigma_1^2 = \sigma_2^2$, against one of the alternative hypothesis

v/s

H_1: $\sigma_1^2 \neq \sigma_2^2$

H_1: $\sigma_1^2 < \sigma_2^2$

H_1: $\sigma_1^2 > \sigma_2^2$

a) Consider, H_1: $\sigma_1^2 \neq \sigma_2^2$ then the R command will be,

```
var.test(x,y,conf.level=c)
```

Where x & y are vectors of observations and c is the confidence coefficient (1-α) for l.o.s α.

b) If H_1: $\sigma_1^2 < \sigma_2^2$ then R command is,

```
var.test(x,y,conf.level=c,alternative="less")
```

c) If H_1: $\sigma_1^2 > \sigma_2^2$ then R command is ,

```
var.test(x,y,conf.level=c,alternative="greater")
```

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

(v) t-Test and Mann Whitney U-test

i. t-Test

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.The null hypothesis and alternative hypothesis for t-test are as follows:-

$H_0: \mu_1 = \mu_2$ v/s $H_1: \mu_1 < \mu_2$

$H_1: \mu_1 > \mu_2$

$H_1: \mu_1 \neq \mu_2$

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

R command for this test is: t.test(x,y,var.equal=,conf.level=c,alternative=””)

The main assumption for application of t-test is assumption of normality if this is violated we cannot use t-test . Hence we use the analogous nonparametric test known as Mann Whitney test.

ii. Mann Whitney U-test

The Mann-Whitney U-test is based on the magnitude of Y observations in relation to X observations.

The null hypothesis and alternative hypothesis for t-test are as follows:-

$H_0: M_x = M_y$ v/s $H_1: M_x < M_y$

$H_1: M_x > M_y$

$H_1: M_x \neq M_y$

Where, M_x and M_y are the medians of X and Y respectively.

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

In R-software the command for Mann-Whitney test is:
wilcox.test(x,y,conf.level=c,alternative=””)

The Mann-Whitney U-test will give very similar results to performing an ordinary parametric two sample t-test on rankings of the data.

(vi) Logistic regression.

Logistic regression analysis is used to examine the association of (categorical or continuous) independent variables with one dichotomous dependent variable.

Consider a situation which involves a dichotomous variable Y and a single regressor X. As conditional distribution of response is Bernoulli with probability given by π(X). We write regression model as

$$Y= E(Y|X) + \varepsilon$$

$$= \pi (X) + \varepsilon$$

Where the error term e is Bernoulli random variable with E(ε)=0 and var= (ε) = π(X) (1-π(X)). The model obtained by using logistic distribution function is called logistic regression model and is given by,

$$\pi (x)= \ln [\pi(x)/1-\pi(x)]$$

Or,

$$Y= [e^{(\beta o+\beta 1x)}/1+e^{(\beta o+\beta 1x)}] +\varepsilon$$

Where, β_o and β_1 are regression coefficients.

A transformation of π (x) that is useful in our study or logistic regression is the logit transformation. It is defined as follows

$$h(x) = \ln [\pi(x)/1 - \pi (x)]$$

$$= \ln [(e\hat{}(Bo+ B1x)/(1+e\hat{}(Bo + B1x)) * (1+ e\hat{} (B1_0 + B1_x))/1]$$

$$= \beta o + \beta 1x$$

The method of maximum likelihood can be used to estimate parameters in logistic regression. This method gives the values of unknown parameters which maximizes the likelihood or probability of obtaining the observed set of data. Logistic regression with a dichotomous regress or coded as 0 and 1, relationship between the odds ratio and the regression coefficient is,

$\Psi = e\hat{}\ \beta 1$

Logistic regression is a powerful analytical research tool due to this relationship. The odds ratio is a measurement of association between Y and X.

Note that $\beta 1 = 0 = \psi = 1$ (No association),

$\beta 1 > 0 = \psi > 1$(Positive association) and

$\beta 1 < 0 = \psi < 1$ (Negative association).

Testing significance of the model:

After estimating the vector parameter β, we wish to test significance of the regressors in the model that is we want to test significance of k coefficients of regression in the model. The likelihood ratio test for all k coefficients of the regressors in the model is performed in the same way as in single regressor case. Here we want to test,

$H_0 = \beta_1 = \beta_2 = \ldots.. = \beta_k = 0$

v/s

H_1= At least one β_i is non zero ; i=1,2,……k

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

(vii) Decision tree

Most common Machine-Learning methods, such as classic linear regression, classification, k-nearest neighbors, use a metric cost function to evaluate performance. As an example, we use the Euclidean distance in a kNN algorithm to find the closest k data points to our unseen one and use the classes of these observed points to assign a value.

There are some datasets or problem questions that do not align perfectly with this method of analysis. Classifying fruit based on its description does not benefit from a calculation of Euclidean distances at any point in the process, so we need an alternative method that can attack these specific problems.

What we might notice about problems that do not lend themselves to distance calculations is that they might benefit from a model that aims to “categorize” them better. Essentially, a model that is less mathematically intense (on the surface) and aims to find common data points by splitting a dataset into progressively smaller and smaller groups at every iteration. The idea is that when we reach the end of our model, we will have a series of groups that can be uniquely identifiable as belonging to a specific class.

This is essentially the process of a decision tree. Decision trees apply a sequence of decisions or rules that often depend on a single variable at a time. These trees partition our input in regions, refining the level of detail at each iteration/level until we reach the end of our tree, also called leaf node, which provides the final predicted label.

- **Components of a Tree**

A decision tree has the following components:

1. Node — a point in the tree between two branches, in which a rule is declared.
2. Root Node — the first node in the tree.
3. Branches — arrow connecting one node to another, the direction to travel depending on how the data point relates to the rule in the original node.
4. Leaf node — a final node in the tree, a point at which a label is assigned to the data point.

5. Level — a number assigned to each set of nodes, starting from 0 that denote how many levels those nodes are from the root.

6. Branching factor — the branching factor B at level l is equal to the number of branches is has to nodes at level L+1.

So how does an instance pass through a decision tree? We begin at the root node. We evaluate our instance with respect to this function and decide which branch we must now go down to reach the next level. It's important to note that the branch decisions must be mutually exclusive, meaning that there cannot be any uncertainty as to which direction the input will move to (it must be purely objective). The node that we have moved to is now considered the root node of this new subtree and we start the process again. This continues iteratively until we reach a leaf node and assign the associated class to our instance. Our trees will also calculate, through training, what the most influential feature variables are, so we may not even need to assess all feature variables in order to make a decision on the class.

Tree classification is considered to be extremely interpretable and very quick to execute, since simple decisions are being made at every stage and an entire tree does not need to be traversed in order to find the correct class.

(viii) Anova and Kruskal-Wallis test

(i) Anova

When a numerical data is collected, we are bound to get some variation in the data. While analyzing the data, there are two factors which govern the variation in the data – assignable causes of variation and non-assignable causes of variation. The technique of partitioning the total variation (assignable and non-assignable causes of variation) present in the set of numerical data into number of different components, estimating the variation due to these components and performing certain tests to draw conclusion regarding the components is known as ANOVA. The procedure of ANOVA was introduced by R.A.Fisher in 1920.

In one-way classification data are classified according to the factor suppose there are k classes of sizes n_1, n_2, …, k_r respectively. Let x_{ij} be the j^{th} observation corresponding to i^{th} class.

(i=1,2,…,k) and (j=1,2,…,ni)

Let, N=∑ni= Total number of observations

Mathematical models; $x_{ij} = \mu+\alpha i+\varepsilon_{ij}$; i=1,2,…,k ; j=1,2,…,ni

Where,

μ = General mean effect

αi= Effect of i^{th} class of the factor

ε_{ij}= Random Error

Here we test the null Hypothesis ,

H_0 : $\mu_1 = \mu_2 =. = \ldots = \mu_k = \mu$ or

H_o : $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$

v/s

H_1: At least one of the α_i’s is not equal to zero.

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

The assumptions of ANOVA are -

i) Observations should be independent and homoscedastic.

ii) Parent population from which the data is collected should be normally distributed.

iii) Various effects of the factors should be additive in nature.

Therefore before applying ANOVA we have to check the normality of data. For this we use Shapiro-Wilk test.

(ii) **Shapiro-Wilk Test:**

The Shapiro-Wilk Test is a test of normality in frequentist statistics.The Shapiro-Wilk test is one of the general normality tests designed to detect all departures from normality.

The hypothesis to be tested under this test is:

Ho: A sample X_1..., X_n came from a normally distributed population.

H1: A sample X_1..., X_n is not from a normally distributed population.

The Test Statistic is : -

$$W= (\sum a_i x_{(i)})^2/\sum(x_i- \overline{x})^2$$

Where, x(i)=i^th^ order statistic i.e. the i^{th} smallest number in the sample.

$\overline{x}$ = the sample mean.

a_i= (a_i, ..., a_n) = constants.

Test Criterion:

If p-value is less than chosen alpha level, then the null hypothesis is rejected at 100α∑% l.o.s and there is evidence that the data tested are not from a normally distributed population i.e. the data are not normal. On the contrary, if p-value is greater than the chosen alpha level then the null hypothesis is accepted i.e. the data came from a normally distributed population.

Command for Shapiro Test in R-software is:

```
shapiro.test()
```

If the assumption of normality for ANOVA gets violated we opt for the analogous nonparametric test known as Kruskal-Wallis test.

(iii) Kruskal-Wallis test.

The Kruskal-Wallis test (sometimes also called the "one-way ANOVA on ranks") is a rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.

The Kruskal–Wallis test by ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes.

The Kruskal–Wallis Non-Parametric Hypothesis Test is to compare medians among k groups (k > 2).The null and alternative hypotheses for the Kruskal-Wallis test are as follows:

H_0: Population medians are equal. (M_D=M_0)

v/s

H_1: Population medians are not all equal.

i.e. H_1: (M_D<M_0) or H_1: (M_D>M_0) or H_1: ($M_D \neq M_0$)

Test criteria: Reject H_o at α% l.o.s if p- value less than l.o.s, otherwise accept it.

In R software Kruskal-Wallis rank sum test is performed by command:

```
kruskal.test()
```

ANALYSIS

Graphical representation of data:

1. Multiple bar plot for preferred mode of ordering food :

Before Pandemic -

Mode of ordering food	Frequency
Dine-in	57
Home made	114
Online	43
Take out	30
Grand Total	**244**

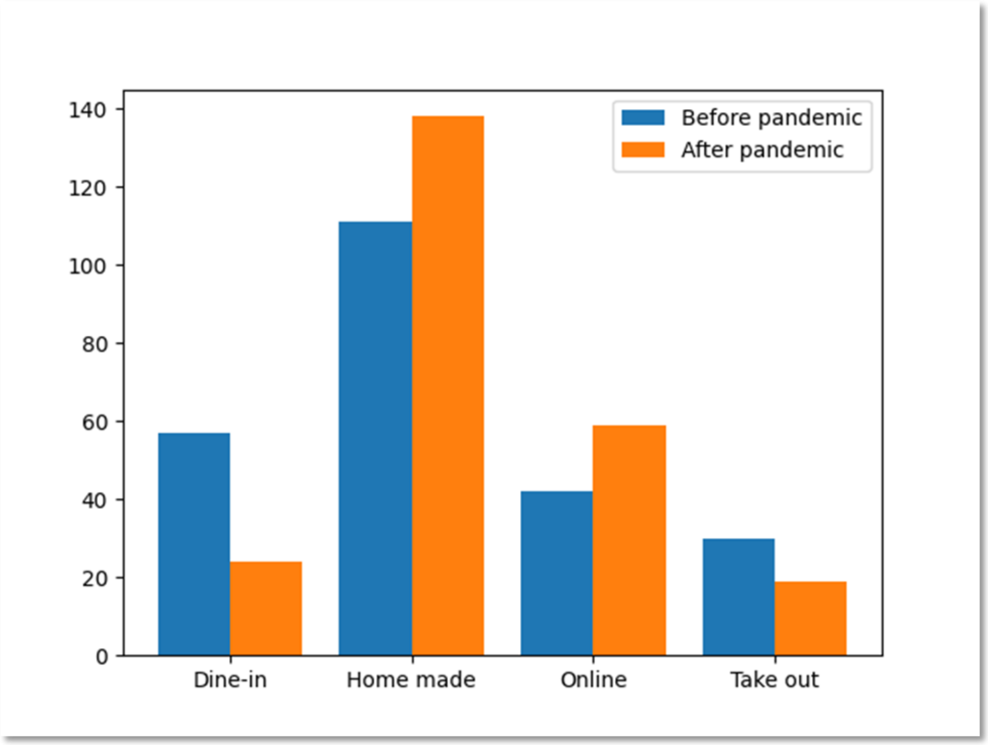
After Pandemic –

Mode of ordering food	Frequency
Dine-in	24
Home made	142
Online	59
Take out	19
Grand Total	**244**

Python commands –

```
import matplotlib.pyplot as p

import numpy as np

Mode_of_ordering_food=['Dine-in','Home made','Online','Take out']

B=[57,111,42,30]

A=[24,138,59,19]

w=0.4

bar1=np.arange(len(Mode_of_ordering_food))

bar1

bar2=[i+w for i in bar1]
```

```
bar2
p.bar(bar1,B,w,label="Before pandemic")
p.bar(bar2,A,w,label="After pandemic")
p.xticks(bar1+w/2,Mode_of_ordering_food)
p.legend()
p.show()
```



Interpretation: It can be observed that most of the people preferred homemade food before as well as after pandemic, also there is slight increase in its preference after pandemic. While there is decrease in preference of dine-in and take out after pandemic. Preference of ordering food online after pandemic has increased.

2. Multiple Bar Plot for Preferred mode of Shopping

Before Pandemic –

Mode of shopping	Frequency
Both	102
Mostly offline	94
Mostly online	48
Grand Total	**244**

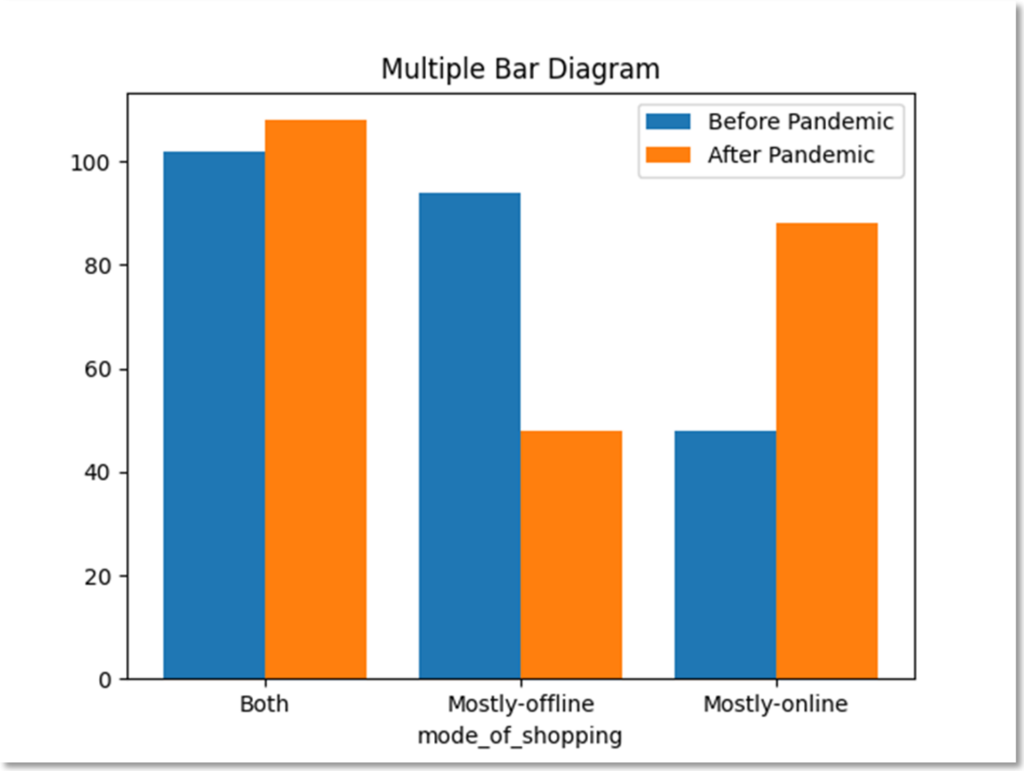
After Pandemic –

Mode of shopping	Frequency
Both	108
Mostly offline	48
Mostly online	88
Grand Total	**244**

Python Commands –

```
import matplotlib.pyplot as p

import numpy as np

mode_of_shopping=["Both","Mostly-offline","Mostly-online"]

B=[102,94,48]

A=[108,48,88]

w=0.4

bar1=np.arange(len(mode_of_shopping))

bar1

bar2=[i+w for i in bar1]

bar2
```

```
p.bar(bar1,B,w,label="Before Pandemic")
p.bar(bar2,A,w,label="After Pandemic")
p.xticks(bar1+w/2,mode_of_shopping)
p.title("Multiple Bar Diagram")
p.xlabel("mode_of_shopping")
p.legend()
p.show()
```

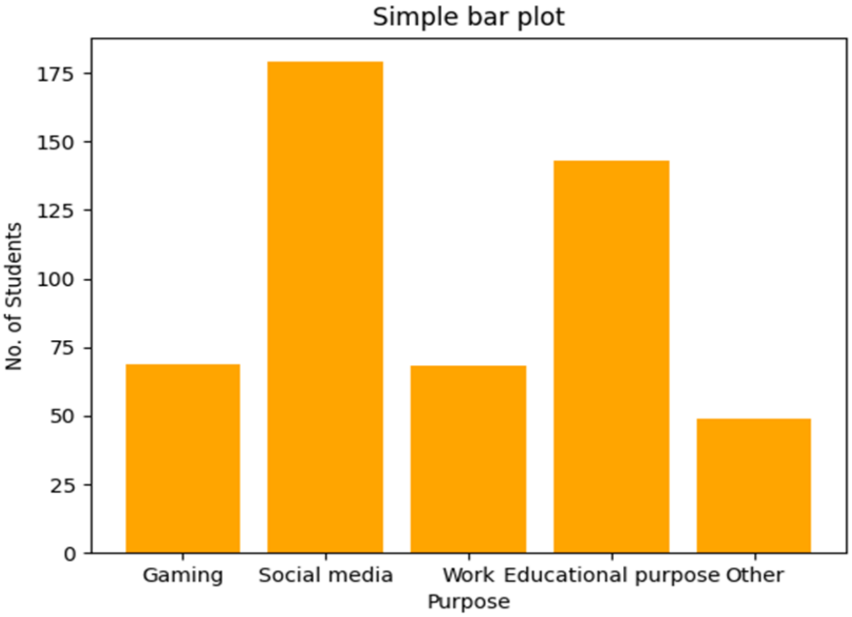


Interpretation: It can be observed that most of the people preferred both modes of shopping. While preference for offline mode of shopping shows a significant decrease after pandemic, it increases for online mode shopping, which is better from social distancing point of view.

3. Simple bar plot about usage of time on internet by students.

Python Commands –

```
import matplotlib.pyplot as p
P=['Gaming','Social media','Work','Educational purpose','Other']
O=[69,179,68,143,49]
p.bar(P,O,color="orange")
p.xlabel("Purpose")
p.ylabel("No. of Students")
p.title("Simple bar plot")
p.show()
```



Interpretation: It can be observed that maximum time spent on internet by students was on social media. Thereafter some of students utilized it for educational purpose while only a few students used it for gaming and other work purpose.

4. Sub-divided bar graph for preferred mode of payment.

Before Pandemic –

Mode of payment	Frequency
BHIM / PayTm / Google Pay	60
Cash	160
Debit / Credit Card	13
Net banking	11
Grand Total	**244**

During Pandemic –

Mode of payment	Frequency
BHIM / PayTm / Google Pay	166
Cash	45
Debit / Credit Card	9
Net banking	24
Grand Total	**244**

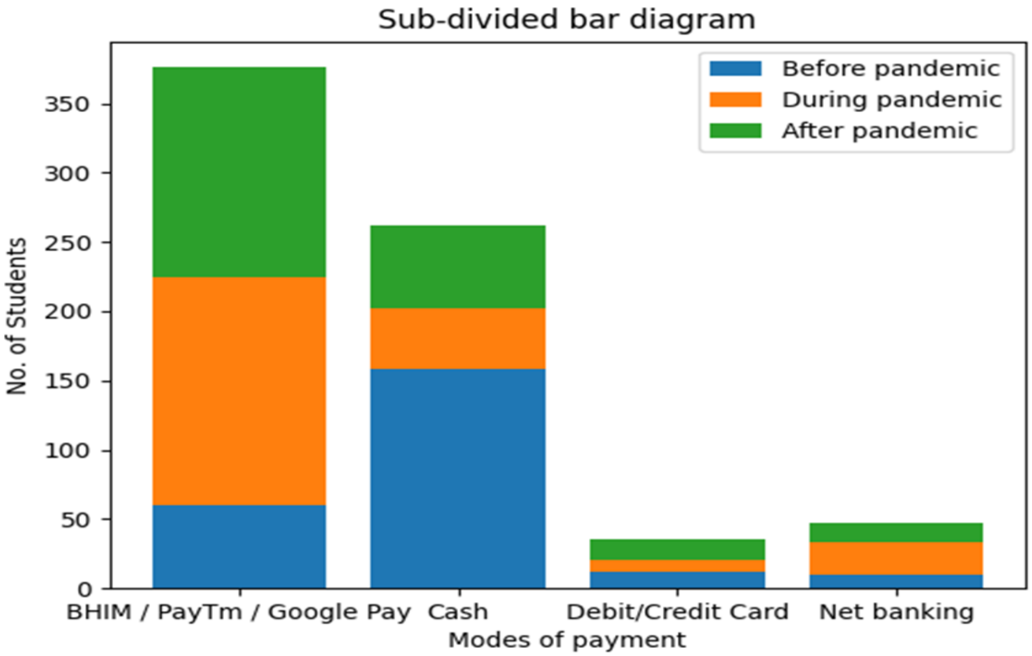
After Pandemic –

Mode of payment	Frequency
BHIM / PayTm / Google Pay	154
Cash	62
Debit / Credit Card	14
Net banking	14
Grand Total	**244**

Python commands –

```
import matplotlib.pyplot as p

import numpy as np

X=['BHIM / PayTm / Google Pay','Cash','Debit/Credit Card','Net banking']

Before_pandemic=[60,158,12,10]

During_pandemic=[164,44,9,23]

After_pandemic=[152,60,14,14]

A_After_pandemic=list(np.add(Before_pandemic,During_pandemic))
```

```
A_After_pandemic
p.bar(X,Before_pandemic,label="Before pandemic")
p.bar(X,During_pandemic,bottom=Before_pandemic,label="During pandemic")
p.bar(X,After_pandemic,bottom=A_After_pandemic,label="After pandemic")
p.title("Sub-divided bar diagram")
p.xlabel("Modes of payment")
p.ylabel("No. of Students")
p.legend()
p.show()
```



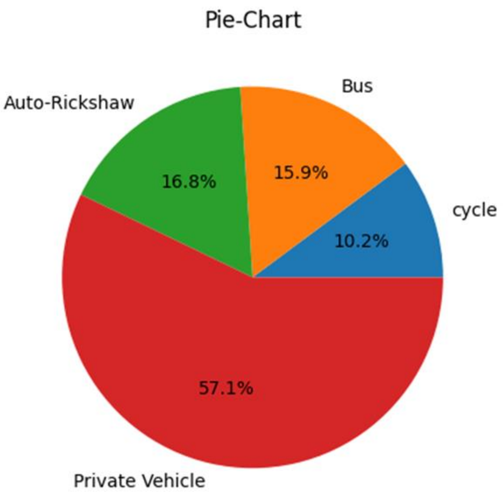
Interpretation: It can be observed that there is notable difference in using cash for payments before, during and after pandemic. It was observed that before pandemic only a few number of students preferred UPI (BHIM/PayTm/Google Pay) as a mode of payment and it increased during pandemic. And there are very few using cards and net banking. Due to secure payments and it being convenient to use, most of the students use UPI for their regular payments.

5. Pie-chart for Preferred Mode of Transportation.

Mode of transportation	Central angle
Cycle	13.1147541
Bus	20.4918033
Auto-Rickshaw	21.7213115
Private vehicle	73.7704918

Python commands –

```
import matplotlib.pyplot as p
responses=[13.1147541,20.4918033,21.7213115,73.7704918]
mode_of_transportation=["cycle","Bus","Auto-Rickshaw","Private Vehicle"]
p.pie(responses,labels=mode_of_transportation,autopct="%2.1f%%")
p.title("Pie-Chart")
p.show()
```



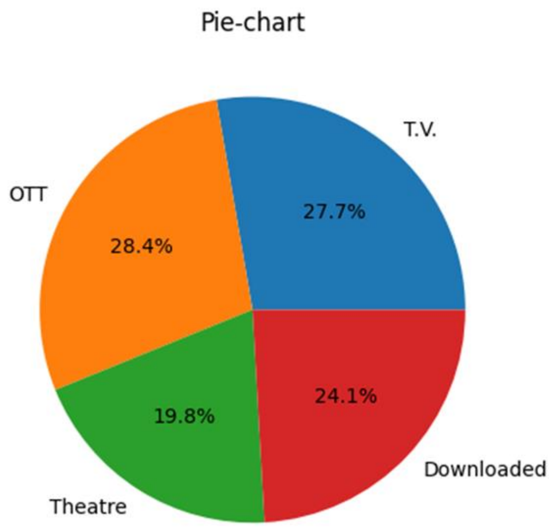
Interpretation: It can be observed that most of the students prefer private vehicles after pandemic. The number of students who prefer bus or auto rickshaw to other mode of transport is almost the same. Keeping social distancing norms in mind the preference for private vehicles has increased.

6. Pie-Chart for preferred mode of watching movies after pandemic.

T.V.	115
OTT	118
Theatre	82
Downloaded	100

Python commands –

```
import matplotlib.pyplot as p

x=[115,118,82,100]

platforms=["T.V.","OTT","Theatre","Downloaded"]

p.pie(x,labels=platforms,autopct="%2.1f%%")

p.title("Pie-chart")

p.show()
```



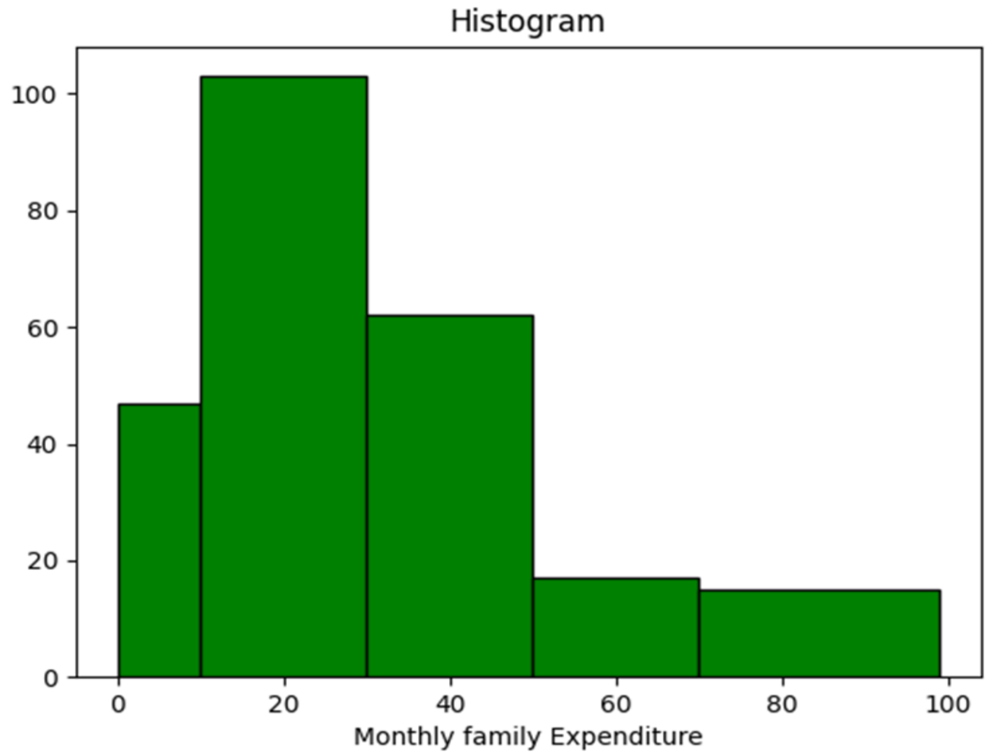
Interpretation: It can be observed that very few students prefer theatre to watch movie after pandemic which is better from social distancing point of view. Students preferring other platforms (OTT, T.V. and Downloaded) is almost the same.

7. Histogram for Monthly Family Expenditure.

Monthly family expenditure	**Frequency**
10k - 30k	103
30k - 50k	62
50k - 70k	17
Above 70k	15
Below 10k	47
Grand Total	**244**

Python Commands –

```
import numpy as np

lb=[0,10,30,50,70]

ub=[10,30,50,70,99]

l=np.array(lb)

u=np.array(ub)

mid=(l+u)/2

f=[47,103,62,17,15]

f1=np.array(f)

y=np.repeat(mid,f1)

interval=[0,10,30,50,70,99]

p.hist(y,interval,edgecolor="Black",color="Green")

p.xlabel("Monthly family Expenditure")

p.title("Histogram")

p.show()
```



Interpretation: It can be observed that monthly family expenditure of students is positively skewed.

TESTS

1. Proportion Test

a) **To check whether proportion of people opting for offline mode of shopping before and after COVID is same or not.**

x1= Number of people who preferred offline mode of shopping before COVID.

x2=Number of people who prefer offline mode of shopping after COVID.

n=Total number of responses.

Here we want to test,

$H_0 : P_1 = P_2$ (Proportion of people opting for offline mode of shopping before and after COVID is same.)

v/s

$H_1 : P_1 \neq P_2$ (Proportion of people opting for offline mode of shopping before and after COVID is not the same.)

```
>x1=94

>x2=48

> x=c(94,48)

> n=c(244,244)

> prop.test(x,n)


        2-sample test for equality of proportions with continuity correction


data:  x out of n

X-squared = 20.113, df = 1, p-value = 7.299e-06
```

```
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1055818 0.2714674
sample estimates:
   prop 1    prop 2
0.3852459 0.1967213
```

Decision : Here p-value = 7.299e-06 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: Proportion of people opting for offline mode of shopping before and after COVID is not the same.

b) To check whether proportion of people attending online lectures regularly of two age groups is same or not.

x1= Number of people attending online lectures regularly of (18 to 24) age group.

x2= Number of people attending online lectures regularly of (24 t0 27) age group.

n= Total number of responses.

Here we want to test,

H_0 : P_1 = P_2 (Proportion of people attending online lectures regularly of two age groups is same.)

v/s

$H_1 : P_1 \neq P_2$ (Proportion of people attending online lectures regularly of two age groups is not the same.)

```
> x=c(106,39)

> n=c(154,58)

> prop.test(x,n)

        2-sample test for equality of proportions with continuity correction


data:  x out of n

X-squared = 0.0031663, df = 1, p-value = 0.9551

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.1371812  0.1689770

sample estimates:

   prop 1    prop 2

0.6883117 0.6724138
```

Decision : Here p-value = 0.9551 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: Proportion of people attending online lectures regularly of two age groups is same.

c) To check whether proportion of males and females regularly attending online lectures is same or not.

x1= Number of males attending online lectures regularly.

x2= Number of females attending online lectures regularly.

n= Total number of responses.

Here we want to test,

$H_0 : P_1 = P_2$ (Proportion of males and females regularly attending online lectures is same.)

v/s

$H_1 : P_1 \neq P_2$ (Proportion of males and females regularly attending online lectures is not the same.)

```
>x=c(79,85)

>n=c(164,164)

>prop.test(x,n)
```

```
2-sample test for equality of proportions with continuity correction


data:  x out of n

X-squared = 0.30488, df = 1, p-value = 0.5808

alternative hypothesis: two.sided

95 percent confidence interval:

-0.15083138  0.07766065

sample estimates:

prop 1    prop 2
```

```
0.4817073 0.5182927
```

Decision : Here p-value = 0.5808 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: Proportion of males and females regularly attending online lectures is same.

2. Chi-square tests for independence of attributes.

a) To test independence of Stream and preferred mode of education.

		Stream		
Mode of education	**Arts**	**Commerce**	**Science**	**Grand Total**
Offline	24	37	110	171
Online	8	20	45	73
Grand Total	**32**	**57**	**155**	**244**

Here we want to test,

H_0 : Preferred mode of education is independent of stream of education.

v/s

H_1 : Preferred mode of education is dependent of stream of education.

```
> x=c(25,37,110,8,20,45)
> x
[1]  25  37 110   8  20  45
> y=matrix(x,nrow=2,byrow=T)
> rownames(y)<-c("Offline","Online")
> rownames(y)
[1] "Offline" "Online"
> colnames(y)<-c("Arts","Commerce","Science")
> colnames(y)
[1] "Arts"     "Commerce" "Science"



> y
```

```
        Arts Commerce Science
Offline   25      37     110
Online     8      20      45
> chisq.test(y)

          Pearson's Chi-squared test

data:  y
X-squared = 1.2928, df = 2, p-value = 0.5239
```

Decision : Here p-value = 0.5239 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: Preferred mode of education is independent of stream of education.

b) To test independence of time spent on internet during COVID and age group.

	Age Groups						
Time spent on internet	**Below 18**	**18 - 21**	**21 – 24**	**24 - 27**	**27 - 30**	**Above 30**	**Grand Total**
Decreased	4	44	22		2		72
Increased	9	81	19	6	3		118
No change	4	29	17	2	1	1	54
Grand Total	**17**	**154**	**58**	**8**	**6**	**1**	**244**

Here we want to test,

H_0 : Time spent on internet during COVID is independent of age group.

v/s

H_1 : Time spent on internet during COVID is dependent on age group.

```
> x=c(4,44,22,0,2,0,9,81,19,6,3,0,4,29,17,2,1,1)
> x
 [1]  4 44 22  0  2  0  9 81 19  6  3  0  4 29 17  2  1  1
> y=matrix(x,nrow=3,byrow=T)
> rownames(y)<-c("Increased","No Change","Decreased")
> rownames(y)
[1] "Increased" "No Change" "Decreased"
> colnames(y)<-c("Below 18","18 - 21","21 - 24","24 - 27","27 - 30","Above 30")
> colnames(y)
[1] "Below 18" "18 - 21"  "21 - 24"  "24 - 27"  "27 - 30"  "Above 30"
> y
          Below 18 18 - 21 21 - 24 24 - 27 27 - 30 Above 30
Increased        4      44      22       0       2        0
No Change        9      81      19       6       3        0
Decreased        4      29      17       2       1        1
> chisq.test(y)

        Pearson's Chi-squared test

data:  y
X-squared = 14.531, df = 10, p-value = 0.1501
```

Decision : Here p-value = 0.1501 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: Time spent on internet during COVID is independent of age group.

c) To test independence of impact of COVID on mental health and interrupted education due to COVID.

		Interrupted education			
Impact on mental health	**Not at all**	**Somewhat**	**Very**	**Extremely**	**Grand Total**
Negatively	5	18	52	33	108
No impact	10	30	41	17	98
Positively	3	5	18	12	38
Grand Total	**18**	**53**	**111**	**62**	**244**

Here we want to test,

H_0 : Impact of COVID on mental health is independent of interrupted education due to COVID.

v/s

H_1 : Impact of COVID on mental health is dependent on interrupted education due to COVID.

```
> x=c(5,18,52,33,10,30,41,17,3,5,18,12)

> x

 [1]  5 18 52 33 10 30 41 17  3  5 18 12

> y=matrix(x,nrow=3,byrow=T)

> rownames(y)<-c("Positive","No impact","Negative")

> rownames(y)

[1] "Positive"  "No impact" "Negative"

> colnames(y)<-c("Not at all","Somewhat","Very","Extremely")
> colnames(y)
[1] "Not at all" "Somewhat"   "Very"      "Extremely"

> y
          Not at all    Somewhat    Very  Extremely
Positive          5          18           52       33
No impact       10        30           41       17
Negative          3          5            18       12
> chisq.test(y,correct = T)


          Pearson's Chi-squared test


data:  y
X-squared = 12.987, df = 6, p-value = 0.04325
```

Decision : Here p-value = 0.04325 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: Impact of COVID on mental health is dependent on interrupted education due to COVID.

3. t-Test

a) To test whether average understanding in offline mode of education is greater than average understanding in online mode of education.

X= Understanding in online mode of education.

Y= Understanding in offline mode of education.

H_0 : μ_1 = μ_2 (Average understanding in online mode of education and average understanding in offline mode of education is same.)

v/s

H_1 : μ_1 < μ_2 (Average understanding in online mode of education is less than average understanding in offline mode of education.)

```
> x=scan("clipboard")

Read 244 items

> x

  [1] 3 1 2 3 3 3 3 2 3 3 2 3 4 2 3 2 1 3 5 2 3 2 3 4 3 3 3 4 2 3 3 4 3 1 1 1 3 4 5 2 3 5 2
4 4 4 5 2 4 3 1 2 2

 [54] 1 5 2 1 2 4 3 1 3 3 2 5 3 3 3 1 2 5 3 2 4 3 3 4 1 1 3 5 2 2 3 3 3 4 2 3 3 3 2 3 2 3 2
3 5 4 3 2 2 3 2 3 3

[107] 4 4 3 2 3 3 3 3 2 2 2 3 3 1 2 2 3 1 3 3 2 3 1 5 2 3 3 2 3 5 2 2 3 1 2 4 5 1 2 3 3 3 2
2 3 3 2 2 3 3 5 1 4

[160] 1 2 3 2 1 3 3 3 3 2 3 2 5 3 3 4 3 2 3 3 5 2 3 5 1 4 3 3 5 2 5 2 3 4 3 2 4 1 2 1 5 4 2
2 3 2 2 2 3 5 3 3 3

[213] 5 3 3 2 1 2 5 3 3 5 5 1 1 3 2 2 3 2 3 4 3 2 3 3 2 5 1 2 4 3 3 4

> y=scan("clipboard")

Read 244 items
```

```
> y
  [1] 5 3 4 5 4 4 5 4 3 5 4 4 4 4 5 5 4 5 5 5 4 3 4 5 5 5 4 5 4 3 4 4 5 5 5 3 5 3 3 5 5 4 4
5 4 4 5 4 5 4 4 4 5
 [54] 5 2 3 1 4 4 5 5 4 4 2 5 4 4 4 3 4 5 4 5 3 4 4 4 5 4 3 1 4 5 4 4 4 3 4 3 4 4 5 5 4 5 4
3 1 4 3 4 4 4 4 4 4
[107] 4 5 4 5 5 4 3 5 2 4 4 3 5 5 4 4 4 3 5 5 5 4 5 2 4 5 4 4 5 1 3 5 2 4 5 4 1 4 4 4 4 4 4
4 5 5 4 4 5 4 1 4 5
[160] 5 4 4 5 5 4 5 4 4 3 4 4 1 3 4 5 4 4 4 5 5 2 3 1 5 3 3 3 3 2 4 5 4 3 3 4 4 5 4 3 4 3 4
4 4 4 4 5 5 4 4 5 5
[213] 2 5 4 4 4 5 3 5 4 1 2 4 4 5 5 4 3 4 4 5 3 4 5 4 5 5 4 5 5 4 4 4
```

To apply t-Test to the given data we first check normality of the data.

Here we want to test,

H_0 : The data is normal.

v/s

H_1 : The data is not normal.

(i) For x,

```
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.8964, p-value = 6.581e-12
```

Decision: Here p-value = 6.581e-12 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: The data is not normal.

(ii) For y,

```
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.80105, p-value < 2.2e-16
```

Decision: Here p-value = 2.2e-16 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: The data is not normal.

Conclusion: Both x and y are not normal.

This violates the assumption of normality required for t-Test. Hence we opt for Mann Whitney U test.

```
> wilcox.test(x,y,alternative = "l")

        Wilcoxon rank sum test with continuity correction

data:  x and y
W = 12413, p-value < 2.2e-16
alternative hypothesis: true location shift is less than 0
```

Decision : Here p-value = 2.2e-16 < l.o.s (5%)

Hence we may reject H_0 at 5 l.o.s.

Conclusion: Average understanding in online mode of education is less than average understanding in offline mode of education.

b) To test whether average monthly family expenditure of PG and hostel students is same.

X= Monthly family expenditure of PG students.

Y= Monthly family expenditure of Hostel students.

H_0 : $\mu_1 = \mu_2$ (Average monthly family expenditure of PG and hostel students is same.)

v/s

H_1 : $\mu_1 \neq \mu_2$ (Average monthly family expenditure of PG and hostel students is not the same.)

```
> x=scan("clipboard")
Read 11 items
> x
 [1] 2 1 0 1 1 1 2 3 1 1 1
> y=scan("clipboard")
Read 19 items
> y
```

```
 [1] 4 0 1 1 0 1 3 1 2 0 2 1 0 2 2 1 3 3 2
```

To apply t-Test to the given data we first check normality of the data.

Here we want to test,

H_0 : The data is normal.

v/s

H_1 : The data is not normal.

(i) For x,

```
> shapiro.test(x)

        Shapiro-Wilk normality test


data:  x
W = 0.80992, p-value = 0.1271
```

Decision: Here p-value = 0.1271 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: The data is normal.

(ii) For y

```
> shapiro.test(y)

        Shapiro-Wilk normality test


data:  y
```

```
W = 0.9136, p-value = 0.0862
```

Decision: Here p-value = 0.0862 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: The data is normal.

```
> var.test(x,y)

        F test to compare two variances

data:  x and y
F = 0.44983, num df = 10, denom df = 18, p-value = 0.198
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1569319 1.5534196
sample estimates:
ratio of variances
         0.4498259
```

Decision: Here p-value = 0.198 > l.o.s (5%)

Conclusion: Variances are equal.

```
> t.test(x,y,var.equal = T)

        Two Sample t-test
```

```
data:  x and y
t = -0.63696, df = 28, p-value = 0.5293
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.0691094  0.5619324
sample estimates:
mean of x mean of y
 1.272727  1.526316
```

Decision : Here p-value = 0.5293 > l.o.s (5%)

Hence we may accept H_0 at 5% l.o.s.

Conclusion: Average monthly family expenditure of PG and hostel students is the same.

4. Logistic Regression.

Y= Responses about whether Covid-19 had a negative impact on physical health.

x1= Responses about change in consumption of junk food and sweets.

x2= Responses about testing positive for COVID.

x3= Responses about change in stress level due to COVID.

```
> y=scan("clipboard")
Read 244 items
> y
  [1] 1 0 1 0 0 1 0 0 0 1 0 1 0 0 1 0 1 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
0 1 0 0 0 1 0 1 0 1
 [54] 1 0 1 1 1 0 1 0 0 1 1 1 1 1 0 1 0 0 0 1 0 0 0 1 1 1 0 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1
0 1 0 1 0 0 0 0 1 1
[107] 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 1 1 1 0 0 1 0 0 0 1 0 1 1 0 0 0 0 0 0 1 1 1 1 0 1 0
1 0 0 1 1 1 1 0 0 0
[160] 1 1 0 1 1 1 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 1 1 1 1 0 0
0 0 1 1 0 1 1 0 1 0
[213] 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1 1 1 1 0 1 0 1 1 0 1 0 0 1 1 0
> x1=scan("clipboard")
Read 244 items
> x1
  [1] 3 1 1 2 1 3 3 2 1 2 2 2 1 2 3 2 2 2 2 1 2 1 1 2 2 2 3 1 2 1 3 3 2 1 1 2 3 3 1 3 3 2 1
3 3 2 2 3 3 2 1 2 1
 [54] 1 3 3 2 3 3 2 1 3 3 2 3 2 3 2 2 3 3 1 3 2 2 1 1 2 2 3 2 1 3 3 2 2 2 3 1 1 2 2 2 2 3 3
3 1 2 2 2 2 1 1 1 2
[107] 2 1 1 2 2 2 1 2 2 3 2 2 3 3 1 2 2 3 2 3 2 3 1 2 1 3 3 2 1 2 2 2 1 3 3 2 3 1 3 3 2 2 3
1 3 2 3 2 3 3 3 3 2
```

```
[160] 1 3 1 1 3 3 2 2 3 1 2 2 2 2 2 2 3 3 3 2 2 2 1 3 3 2 2 3 2 1 3 3 3 1 2 2 3 1 2 2 3 2 2
1 2 3 3 2 1 2 1 2 1

[213] 2 1 3 2 1 2 2 1 2 1 1 1 2 1 1 3 1 2 3 2 2 2 3 2 2 2 2 1 2 2 3 2

> x2=scan("clipboard")

Read 244 items

> x2

  [1] 1 0 0 0 0 1 0 0 1 1 1 1 0 0 0 0 1 1 0 1 0 0 1 0 1 1 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 1 0 0 1

 [54] 0 0 1 0 0 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 0 1 1 0 1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1 0

[107] 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0
0 0 0 0 0 0 0 0 0 0

[160] 1 0 0 1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 1 1 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0

[213] 0 1 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 1 0

> x3=scan("clipboard")

Read 244 items

> x3

  [1] 2 1 2 2 2 2 1 1 1 3 3 2 2 1 1 2 2 1 1 1 2 1 1 3 2 3 1 3 1 1 1 1 2 1 1 3 3 3 1 1 3 2 3
3 1 1 3 3 1 2 3 2 2

 [54] 2 1 1 2 1 1 1 3 2 2 2 1 1 1 2 2 2 3 3 1 2 2 1 2 2 2 2 3 1 2 2 1 1 3 3 1 1 3 2 3 3 1 1
2 2 2 2 2 1 1 1 1 1

[107] 1 1 1 2 3 1 1 1 2 1 1 1 3 1 1 3 2 2 2 1 2 1 1 2 1 2 2 2 1 3 2 2 2 3 1 2 3 1 1 2 2 1 2
3 3 2 1 2 1 1 2 1 2

[160] 3 3 3 1 2 1 2 2 1 3 2 2 1 2 2 1 1 3 1 1 2 1 3 1 1 2 1 2 2 2 3 1 2 1 2 3 2 1 1 2 1 2 2
1 1 1 1 1 3 3 2 2 1

[213] 2 3 2 3 3 1 2 3 1 1 3 3 2 1 1 1 1 3 1 2 2 1 3 1 2 1 1 3 3 3 1 3


> lfit=glm(y~(x1+x2+x3),family=binomial)
> summary(lfit)

Call:
glm(formula = y ~ (x1 + x2 + x3), family = binomial)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-1.7910  -1.0697   0.6702   1.1690   1.5004

Coefficients:
                 Estimate      Std. Error    z value         Pr(>|z|)
(Intercept)   -0.1808           0.5201       -0.348           0.728
x1              0.1594           0.1828        0.872           0.383
x2              1.3193           0.3244        4.067          4.77e-05 ***
x3             -0.2373           0.1725       -1.376          0.169
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

   Null deviance: 338.11  on 243  degrees of freedom
Residual deviance: 317.66  on 240  degrees of freedom
AIC: 325.66

Number of Fisher Scoring iterations: 4
```

Testing significance of regression:

Here we want to test,

H_0 : Regression coefficient is insignificant.

v/s

H_1 : Regression coefficient is significant.

For each Regression coefficient: β_1, β_2, β_3.

(i) For regression coefficient $\boldsymbol{\beta_1}$,

p-value for $\boldsymbol{\beta_1}$ = 0.383
Decision: Here p-value > l.o.s (5%)
Hence we may accept H_0 at 5% l.o.s.
Conclusion: Regression coefficient $\boldsymbol{\beta_1}$ is insignificant.

(ii) For regression coefficient $\boldsymbol{\beta_2}$,

p-value for $\boldsymbol{\beta_2}$ = 4.77e-05 ***
Decision: Here p-value < l.o.s (5%)
Hence we may reject H_0 at 5% l.o.s.

Conclusion: Regression coefficient $\boldsymbol{\beta_2}$ is significant.

(iii) For regression coefficient $\boldsymbol{\beta_3}$,

p-value for $\boldsymbol{\beta_3}$ = 0.169
Decision: Here p-value > l.o.s (5%)
Hence we may accept H_0 at 5% l.o.s.

Conclusion: Regression coefficient $\boldsymbol{\beta_3}$ is insignificant.

Conclusion: It can be concluded that regression coefficients $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_3}$ are insignificant and regression coefficient $\boldsymbol{\beta_2}$ is significant.

That is testing positive for COVID has a significant effect on COVID causing negative impact on physical health.

5. Analysis of variance (ANOVA) and Kruskal-Wallis test.

a) To test whether the average study time is same before, during and after COVID.

Here,

$\mu_1 =$ Average study time before COVID.

μ_2 = Average study time during COVID.

μ_3 = Average study time after COVID.

Here we want to test,

H_0 : $\mu_1 = \mu_2 = \mu_3$ (Average study time before, during and after COVID is same.)

v/s

H_1 : $\mu_1 \neq \mu_2 \neq \mu_3$ (Average study time before, during and after COVID is not the same.)

```
>x1=scan("clipboard")

Read 244 items

> x1

  [1] 3 1 3 3 3 1 2 2 2 4 2 2 1 1 3 2 3 4 3 3 4 3 2 3 3 4 2 2 3 1 3 3 4 5 2 3 5 2 1 4 3 4 1
4 3 2 1 1 4 3 3 2 2

 [54] 3 3 3 1 4 3 4 5 2 3 2 3 2 3 4 3 3 2 1 3 2 3 2 3 3 3 3 3 3 2 4 4 2 1 3 1 3 5 4 2 4 3
1 4 3 1 3 2 3 5 3 1

[107] 2 2 3 3 3 3 5 3 2 1 2 3 2 3 3 2 2 3 2 5 3 2 3 2 3 4 2 2 4 2 2 2 2 3 4 2 3 4 5 4 2 3
3 3 5 2 3 2 2 3 3 3 3

[160] 2 4 1 5 3 1 4 3 3 3 3 1 3 3 3 4 2 2 2 3 4 1 2 1 1 3 1 1 3 3 3 3 3 3 2 3 3 3 2 2 5 2
3 3 2 3 3 3 3 2 3 3 3

[213] 1 2 2 2 1 3 4 3 2 5 5 3 1 2 3 2 2 3 2 3 3 3 2 3 3 2 3 3 2 3 1 4

> x2=scan("clipboard")

Read 244 items
```

```
> x2

  [1] 1 1 3 3 4 1 3 1 2 2 1 1 1 1 3 2 1 2 2 1 5 1 1 2 3 2 2 1 2 1 2 1 2 5 3 2 5 4 1 2 1 4 2
3 3 2 1 1 3 1 2 4 1

 [54] 1 4 1 1 4 2 3 2 3 1 2 2 1 1 2 2 2 1 1 2 1 3 1 2 1 4 2 3 2 2 2 3 3 1 2 1 1 1 1 2 2 2 1
1 3 2 1 1 1 2 3 1 1

[107] 2 1 1 3 2 2 2 2 1 1 1 2 3 4 1 1 1 2 2 5 1 2 4 1 2 3 1 1 2 1 1 1 1 4 2 2 2 1 2 3 1 2
2 2 3 1 1 2 1 2 2 1 3

[160] 1 5 2 5 2 1 1 2 1 2 1 1 5 3 1 2 3 1 2 3 4 1 2 2 1 2 1 1 1 1 2 2 1 4 3 1 3 1 1 1 5 2
2 2 2 1 2 1 2 3 1 3 3

[213] 2 1 1 1 1 2 3 2 1 5 5 2 1 1 3 1 2 3 1 2 4 2 4 2 2 2 1 3 1 3 1 3

> x3=scan("clipboard")

Read 244 items

> x3

  [1] 2 2 3 3 3 2 2 3 2 4 2 3 2 2 3 2 2 2 2 2 4 2 3 3 4 2 2 4 4 3 3 2 5 2 3 5 4 2 2 2 4 2
2 3 2 2 2 3 2 2 4 2

 [54] 2 3 2 2 4 3 4 2 2 2 2 3 2 2 4 3 2 2 2 2 2 3 2 3 2 3 2 2 3 2 3 4 2 2 2 2 2 2 5 4 2 3 4
2 4 2 2 2 2 3 3 2 5

[107] 2 2 4 3 3 2 4 2 3 2 2 2 3 3 2 2 2 3 2 5 2 2 3 2 2 3 2 2 3 3 2 2 3 2 5 2 3 4 3 3 2 3
3 3 3 2 2 2 3 3 3 2 4

[160] 3 5 2 5 2 2 2 3 2 2 3 2 2 3 3 3 3 2 2 3 4 2 2 2 2 3 2 2 3 2 2 2 3 4 2 3 3 2 2 2 5 2
3 2 2 2 2 2 3 2 2 3 3

[213] 2 2 3 2 2 3 3 3 2 5 5 2 2 2 3 2 2 3 2 2 3 3 2 3 3 2 2 3 2 3 2 2
```

To apply ANOVA to the given data we first check normality of the data.

Here we want to test,

H_0 : The data is normal.

v/s

H_1 : The data is not normal

(i) For x1:

```
> shapiro.test(x1)
        Shapiro-Wilk normality test
data:  x1
W = 0.89494, p-value = 5.258e-12
```

Decision: Here p-value = 5.258e-12 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: The data is not normal.

(ii) For x2:

```
> shapiro.test(x2)
        Shapiro-Wilk normality test
data:  x2
W = 0.80759, p-value < 2.2e-16
```

Decision : Here p-value = 2.2e-16 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: The data is not normal.

(iii) For x3:

```
> shapiro.test(x3)
        Shapiro-Wilk normality test
data:  x3
W = 0.72503, p-value < 2.2e-16
```

Decision: Here p-value = 2.2e-16 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: The data is not normal.

Conclusion: All x1 , x2 and x3 are not normal.

This violates the assumption of normality required for ANOVA. Hence we opt for Kruskal-Wallis test.

```
> x=list(x1,x2,x3)
> kruskal.test(x)


        Kruskal-Wallis rank sum test


data:  x
Kruskal-Wallis chi-squared = 91.271, df = 2, p-value < 2.2e-16
```

Decision : Here p-value = 2.2e-16 < l.o.s (5%)

Hence we may reject H_0 at 5% l.o.s.

Conclusion: Average study time before, during and after COVID is not the same.

6. Decision tree

Here, Target variable is whether a student tested positive for COVID or not.

We make a decision regarding the target variable on the basis of following variables: (i) Gender. (ii) Age group. (iii) Residential condition.

Testing positive for COVID	Frequency
No	178
Yes	66
Grand Total	**244**

Entropy(S)	0.836640742

	Testing positive for COVID		
Gender	**No**	**Yes**	**Grand Total**
Female	83	31	114
Male	95	35	130
Grand Total	**178**	**66**	**244**

Entropy(S,female)	0.828643675
Entropy(S,male)	0.843349289
Gain(S,Gender)	9.27993E-05

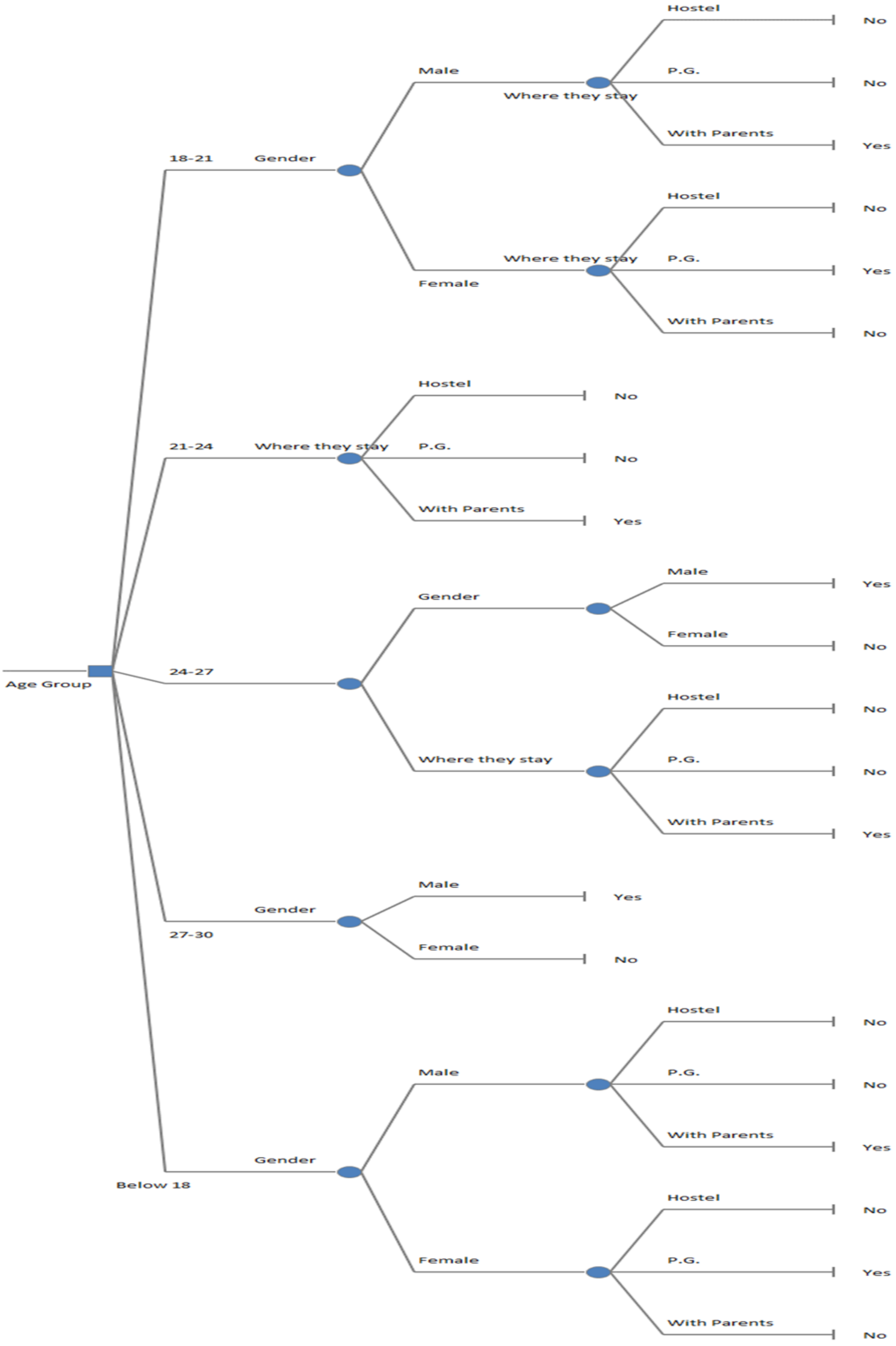
	Testing positive for COVID		
Age Group	**No**	**Yes**	**Grand Total**
18 – 21	116	37	153
21 – 24	44	14	58
24 – 27	3	5	8
27 – 30	6	2	8
Below 18	9	8	17
Grand Total	**178**	**66**	**244**

Entropy (S,18-21)	0.792387151
Entropy (S,21-24)	0.79732651
Entropy (S,24-27)	0.954434003
Entropy (S,27-30)	0.650022422
Entropy (S,below 18)	0.997502546
Gain (S,Age group)	0.026688462

	Testing positive for COVID		
Residential condition	**No**	**Yes**	**Grand Total**
Hostel	9	5	14
P.G.	18	8	26
With Parents/ Guardian	151	53	204
Grand Total	**178**	**66**	**244**

Entropy(S,hostel)	0.940285959
Entropy(S,P.G.)	0.870864469
Entropy(S,With parents/Guardian)	0.822836884
Gain(S,Where do you stay)	0.002149903

Here as the gain for "Age group" is maximum, it is the decision node.


Age Group
18-21
Gender
Male
Where they stay
Hostel
No
P.G.
No
With Parents
Yes
Female
Where they stay
Hostel
No
P.G.
Yes
With Parents
No
21-24
Where they stay
Hostel
No
P.G.
No
With Parents
Yes
24-27
Gender
Male
Yes
Female
No
Where they stay
Hostel
No
P.G.
No
With Parents
Yes
27-30
Gender
Male
Yes
Female
No
Below 18
Gender
Male
Hostel
No
P.G.
No
With Parents
Yes
Female
Hostel
No
P.G.
Yes
With Parents
No

Interpretation:

Here the decision node is Age group with 5 decision branches

(i) Below 18

For this decision branch it can be observed that for the collected data the male students living with their parents tested positive for COVID while for female students staying at a PG resulted in testing positive for COVID.

(ii) 18 – 21

For this decision branch it can be observed that for the collected data the male students living with their parents tested positive for COVID while for female students staying at a PG resulted in testing positive for COVID.

(iii) 21 – 24

For this decision branch it can be observed that for the collected data students living with their parents tested positive for COVID irrespective of their gender.

(iv) 24 – 27

For this decision branch it can be observed that for the collected data only the male students living with their parents tested positive for COVID.

(v) 27 – 30

For this decision branch it can be observed that for the collected data only the male students tested positive for COVID irrespective of their residential condition.

CONCLUSIONS

1. It can be concluded that proportion of people opting for offline mode of shopping before and after COVID is not the same.
2. It can also be concluded that proportion of people attending online lectures regularly of two age groups is same and the proportion of males and females attending online lectures regularly is also the same.
3. We can conclude that preferred mode of education and stream of education are not associated with each other.
4. Change in time spent on internet during COVID is not dependent on age group.
5. It can be concluded that impact of COVID on mental health of students is dependent on interruption in education caused by COVID.
6. Average understanding in online mode of education is less than average understanding in offline mode of education for students, which shows that students prefer traditional way of learning over digitalization of education .
7. The average monthly family expenditure of PG and hostel students is the same.
8. It can be observed that testing positive for COVID has a significant effect on COVID causing negative impact on physical health.
9. The average study time for students before, during and after COVID differs.
10. From the decision tree it can be concluded that for all age groups, almost all male students tested positive with respect to their residential condition and for female students aged 16 to 24 staying at a PG proved to be a factor resulting in testing positive for COVID.

MAJOR FINDINGS

1. From the collected data 26.7% students tested positive for COVID. Out of those students 85.9% tested positive once for COVID and the remaining students tested positive more than once for COVID.
2. 28.7% students suffered a loss in their immediate family due to COVID-19.
3. Most of the students chose offline mode of education because of better learning environment and better interaction between students and teachers. Some of the students chose online mode of education because it is convenient and flexible as well as it has accessibility of time and place.
4. From the collected data 52.9% students prefer that the MCQ pattern exam is beneficial for the students from their academic point of view.
5. COVID-19 had an effect on students' studying habits, as we found that most students studied 2 hours to 4 hours before the pandemic, which reduced to less than 1 hour during the pandemic.
6. An increase in time spent on the internet for 48.3% of students was observed during COVID.
7. 82.5% of students are aware about current affairs and most of the students refer to social media for information thus making a better use of it.
8. 64.6% of students learned new skills during the pandemic.
9. 33.6% of students have started working after COVID-19 pandemic out of which 53.9% of students are working full time jobs and remaining 33.% of students works part-time.
10. 51.7% of students had negative impact on their Physical health during pandemic and 44.6% of students had negative impact on their mental health during pandemic.
11. It has been observed that 45% of people had no change in their consumption of junk food and sweets. On the Other hand, it has also been observed that 37.9% of students have increased their consumption of fresh Vegetables and fruits.

12. An increase in frequency of exercising was observed in 35.8% of students.

13. According to the survey, the stress level of 42.9% of students have increased.

14. Also, there was no change observed in the average sleep time of 45.4% students.

15. It was observed that preferred mode of shopping was both online and offline for 42.1% of students before pandemic and was slightly increased to 43.8% after pandemic.

16. 46.3% of students preferred homemade food before pandemic which later on increased to 57.5 % after pandemic. This shows that students preferred healthy diet over outside food.

17. It was also observed that 25% of students preferred Digital payment options before pandemic which drastically increased to 63.3% during pandemic and remained constant after pandemic showing that students find digitalization of payment more convenient over other payment options.

18. Most of the students i.e. 73.8% prefer private vehicles for transportation after COVID which shows that they are now more health conscious.

19. 35.4% of students are very much satisfied with their life these days.

OPINION OF RESPONDENTS.

1. **Did you learn any new skills during COVID-19? If yes, then please mention-**

- Cooking, new language etc.
- I became nutritionist.
- Power BI, Qlik Sense, Knime.
- Foreign language, cooking.
- Drawing.
- Gardening.
- Python, Digital marketing, Photoshop, CorelDraw, Website development.
- Online courses.
- Video editing, technical skills.
- I already having a hobby of write poems in my own words but during pandemic my writing skills was improved and I write some perfect poems in suitable words.
- Reading books.
- Stock market, making tasty food, Tech knowledge.
- Control of emotions.
- Piano.
- Patience, critical thinking.
- Cooking.
- Speaking English.
- Worli painting.
- Books reading, drawings.
- Lot of programming languages.
- Playing flute, Cycling, Yoga, trekking.

- Sketching.
- Stitching.
- I have learnt to play Harmonica (Mouth Organ) during Covid-19.
- Origami, Solid works.
- Photography.
- Swimming, Car driving, cooking.
- Making food, and playing chess.
- Business Marketing.
- Mandala art.
- Dancing.
- Cloth painting.

2. **In your opinion, which aspect of your life did Covid-19 have a great impact on?**

- Personal.
- Mental health, physical health and overall health.
- I lost my closed ones so I feel depressed about losing them due to which mental health got affected.
- My mental health.
- On health and education.
- Staying alone.
- Way of job.
- Health and education.
- Educational purposes and family relations.
- My writing hobby.
- Relationships.
- Personal hygiene.
- Digital.

- Not proper education.
- Social, Mental, Physical.
- College.
- Adverse impact on education quality but also positive impact like participation in various extra-curricular activities.
- My health and focus.
- Academic.
- It tore many families into pieces.
- Earning, Lifestyle, health.
- Overall perspective towards life.

3. **As the restrictions have been lifted, what precautions do you think should be implemented to prevent another pandemic?**
    - People should maintain hygiene and continue to wear mask.
    - Eat n drink healthy, exercise regularly, when feeling ill concern your doctor, stay calm, pray to Almighty for no other pandemic.
    - Keep clean and help everyone who needs help.
    - The precautions which we are taking before should be taken now also but in a limited way.
    - One should firstly follow every rules and regulations given by government.
    - Wearing Mask must be mandatory at social gatherings, avoid spitting on the road, personal hygiene.
    - Vaccination drive should be made more efficient and fast.
    - Eat fresh foods, live life with good habits, stay safe.
    - Wearing mask and maintenance of social distance.
    - Wearing a mask and maintaining the social distance taking precautions against viral infection.

- Self-care, precautions while going out and after returning home, a regular medical examination.
- Hygiene, healthy diet, exercise, medical awareness and a rational mind.
- We should wear a mask while going in public places, Use of sanitizer to protect ourselves.
- Social Distancing and Sanitization.
- Following Guidelines.
- Masks and vaccination
- People should lift themselves from the fact that the 2nd dose of the vaccine cannot prevent or cannot be used as an absolute means for prevention of another pandemic.
- People should understand the fact that the fight is not over yet. Going out every weekend would somehow lead to pandemic.
- Sanitizing, wear mask, regular exercise and healthy diet which leads to good immunity.
- Updating and assisting the affected people to stay quarantine.
- Everyone should be vaccinated first, then everyone should wear a mask, use a sanitizer and make sure that everyone we live with should be vaccinated. This kind of instructions follow than we all are living peace fully.
- Boosting immune system with healthy food, exercise & good mental health.

Limitations

1. This data is limited to 244 respondents. Therefore, whatever results obtained are limited to only this sample. Hence, these results are not suitable for worldwide purpose.

2. Data of the project was mostly collected from the local college students of Nasik. So, the conclusion of this project's result may differ for other places.

3. Since the data is collected for the age group 13-30, the conclusion will not be same if other age groups were to be included.

4. The conclusions may vary from sample to sample.

Bibliography

Books:

1. Statistical computing using R software

2. Common Statistical Test - M. B. Kulkarni

Software Used:

1. Python.

2. R-software.

3. MS Excel.

4. Advanced MS Excel.

5. MS Word.

Websites

1. https://www.frontiersin.org/articles/10.3389/fnut.2021.635859/full

2.https://www.clinicaladvisor.com/home/topics/diet-and-nutrition-information-center/views-fast-food-changed-during-pandemic/

3.https://ieeexplore.ieee.org/document/9243379

4.https://www.sciencedirect.com/science/article/pii/S019074092032288X