IMPACT OF COVID-19 ON STUDENTS

GROUP 6

SUBMITTED TO:
MR. MAYUR MORE

PROFESSER IN-CHARGE

ST. XAVIER'S COLLEGE

MUMBAI (AUTONOMOUS)

ST. XAVIER'S COLLEGE (AUTONOUMOUS), MUMBAI

DEPARTMENT OF IT

GROUP PROJECT

(2021 – 2022)

CLASS: **MSc BDA (Part 1)** **GROUP NUMBER:** **6**

TITLE: IMPACT OF COVID-19 ON STUDENTS

GROUP MEMBERS		
UID	**NAME**	**EMAIL ID**
219020	PAYAL SHARMA	payalmsharma1602@gmail.com
219002	REEVE KALE	alvitakale09.as@gmail.com
219035	NOVA ABRAHAM	abrahamnova113@gmail.com
219008	AKHIL ISAAC JOHN	akhilisaacjohn@gmail.com
219030	YASHIKA SACHDEVA	yashikasachdeva900@gmail.com
219043	BHARGAV YERAGI	bhargav.yeragi@gmail.com

PROF. MAYUR MORE
PROFESSOR IN-CHARGE

PROF. ROY THOMAS
HEAD OF IT DEPARTMENT

INDEX

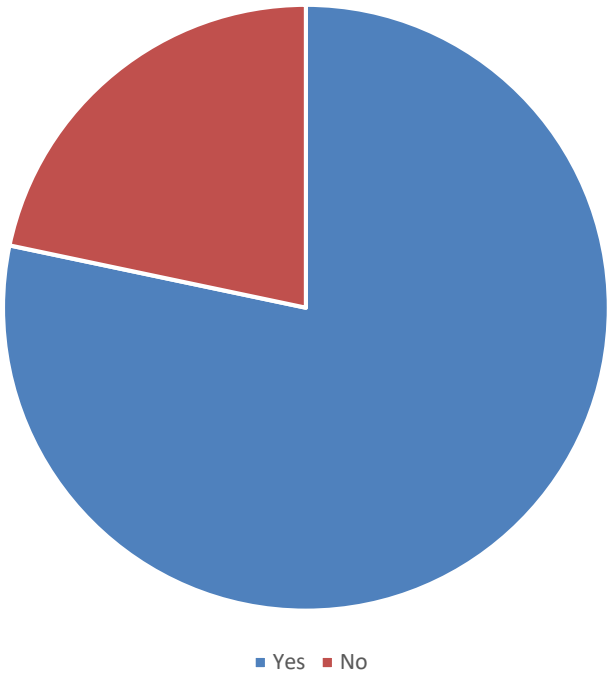
Sr.no	Content	Page no.
1	Introduction	3
2	Objectives	5
3	Assessment of objective 1	6
4	Assessment of objective 2	8
5	Assessment of objective 3	10
6	Assessment of objective 4	12
7	Assessment of objective 5	14
8	Assessment of objective 6	16
9	Assessment of objective 7	18
10	Assessment of objective 8	20
11	Statistical Analysis	22
12	Methodology	33
13	Limitations	35

Introduction

27th January 2020, the date when India witnessed its first ever covid positive patient. Since then, the story of the nation and the world is not a hidden matter to any citizen. The world economy crashed, industries halted, lives being succumbed were just among the few horrific things we all witnessed around us. Each and every individual around us had to go through a lot during these months, be it physically or mentally. As we slowly and gradually transitioned towards taking our work, occupation and even studies online, this new medium came with its own fair share of pros and cons. Through this project we would like to highlight the effect covid-19 had on the lives of students. We surveyed a fair number of students and asked them about their lifestyles, feedback on online education and many other aspects of their life which we felt were affected because of covid-19.

We surveyed 92 students most of them being in the age of 18-23. We asked them certain set of questions which are listed down below in our objectives. Among the questions which we asked, there was one particular question which needed a serious look into, not just by the students from our survey but by each and every individual of our planet.

The question of “Did the COVID-19 pandemic changed your thinking on the importance of climate change?’’ highlighted the importance of how serious of number of issue climate change is. The results we got were as follows.



78.3% of the people have had a changed perspective on climate change.

OBJECTIVES

Through the medium of this project we aim to show the impact COVID- 19 has caused on students life.

We have chosen the following parameters as our objectives:-

1] Mental Health during COVID-19 Lockdown

2] Hours spent working before vs. during the lockdown

3] Physical activity before vs. during the Lockdown

4] Effect of COVID-19 on education platforms

5] Response levels of distance learning

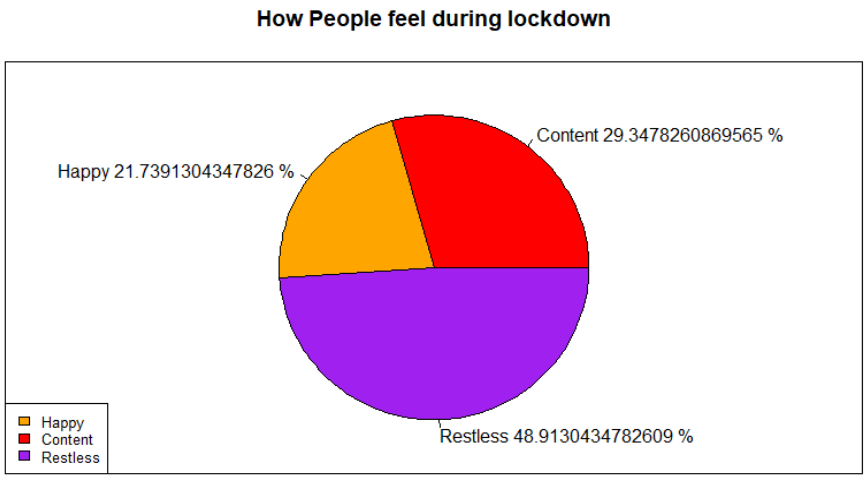
6] Monthly expenses during lockdown

7] Entertainment (OTT platforms)

8] Effect of Social media platforms with increased screen time

ASSESSMENT OF OBJECTIVE 1

The idea behind this was to check on how people have been feeling generally throughout the day.



Insight:

From the total 92 responses that we got we can see that a total of 45 respondents which translates to 48.9% felt restless throughout the day followed by 29.3% feeling content and 21.7% feeling happy. The following results can be observed in the pie chart given above.

R code :

```
data1<-read.csv("dataset.csv",header=TRUE)


data1
data1 %>%select(feel_during)


data2<-data.frame(data1)
detach(data2)
attach(data2)
class(data2)
summary(data2$feel_during)
```

Length Class Mode

92 charactercharacter

```
table(feel_during)
```

Content Happy Restless

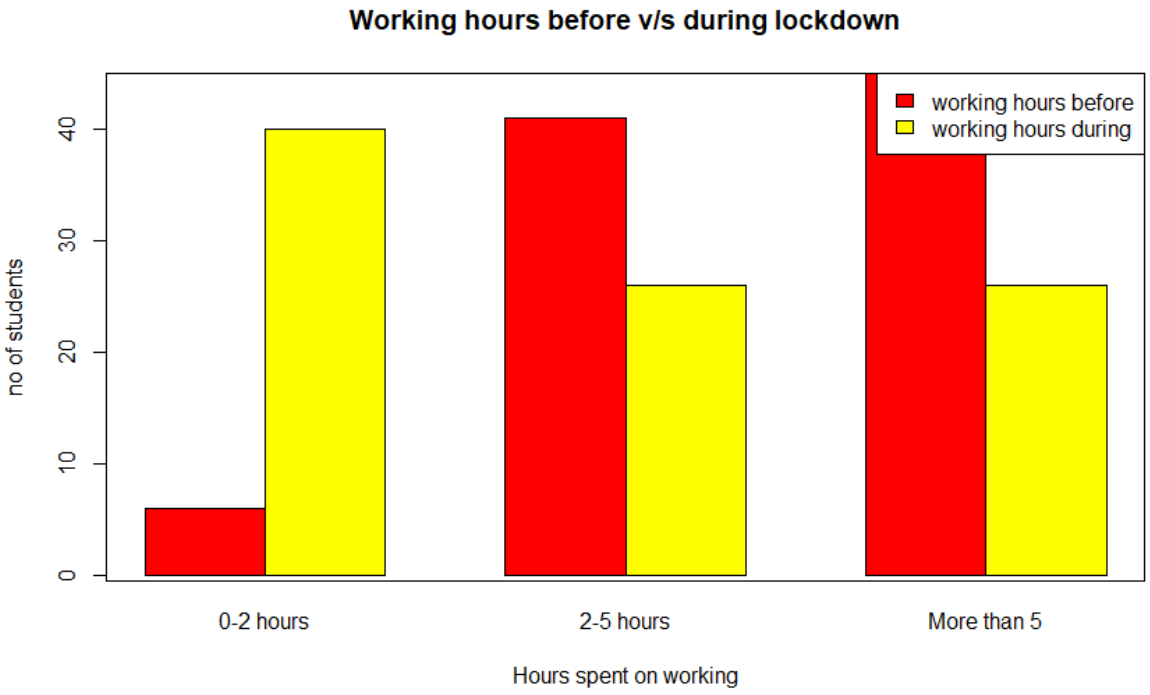
27 20 45

So mode is restless as it has appeared maximum no of times

```
(table(feel_during)/92)*100
lbls<-paste(names(table(feel_during)),(table(feel_during)/92)*100,"%")
pie(table(df$feel_during),labels = lbls,main="How People feel during lockdown",col=c("red","orange","purple"))
legend("bottomleft", c("Happy","Content","Restless")ex = 0.8,fill=c("orange","red","purple"))
box()
```

ASSESSMENT OF OBJECTIVE 2

The thought behind this question was to analyze how much difference has the COVID-19 lockdown made to the working hours of any given individual.



Insights

Here what we can observe is that there is a significant difference between the working hours of students. As we can see that there were 45 responses for people working more than 5 hours before covid, which drops down to 26 responses during covid. Another significant change can be seen in students who used to work for up to 2 hours. There were 6

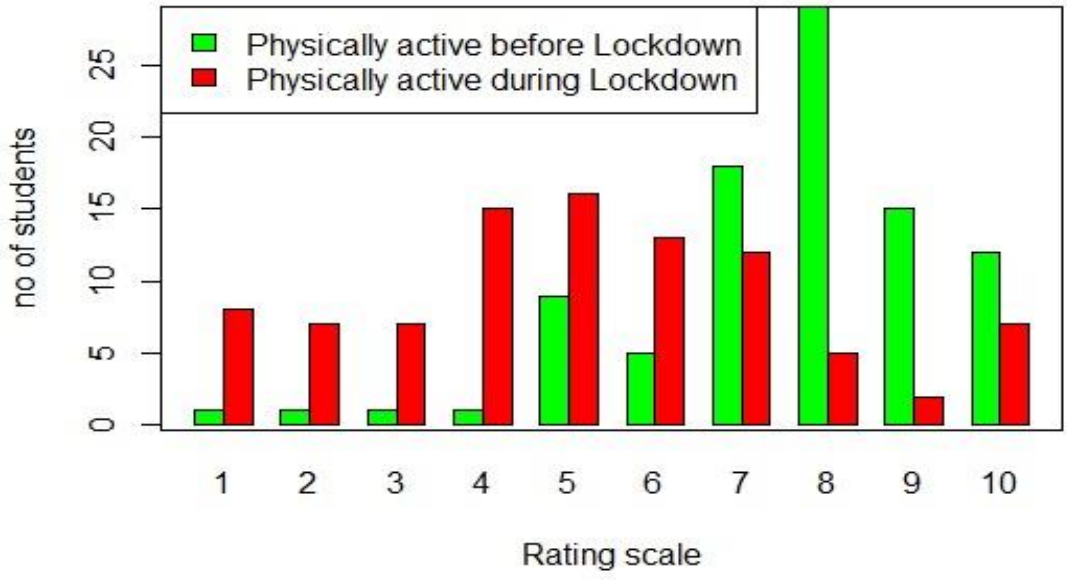
responses for it before covid compared to 40 for working during covid.

R code :

```
working_hours_before<-table(working_hours_before)
working_hours_before
0-2 hours   2-5 hours More than 5
         6         41         45
Mode  ofworking_hours_before  is More than 5
working_hours_during<-table(working_hours_during)
0-2 hours   2-5 hours More than 5
        40         26         26
Mode of working_hours_during is 0-2 hours
ta=rbind(working_hours_before,working_hours_during)
ta
                       0-2 hours 2-5 hours More than 5
working_hours_before         6        41         45
working_hours_during        40        26          2

barplot(height=ta,beside = TRUE,col=c('red','yellow'),xlab='Hours spent on working',ylab='no of students')
legend('topright',legend=c('working hours before','working hours during'),fill=c('red','yellow'))
```

ASSESSMENT OF OBJECTIVE 3

As there was restlessness among students during the lockdown, we also thought of comparing their amount of physical activity during the lockdown. As most of us were restricted to the walls of our home. This parameter was taken into account for the home workouts and other forms of physical activity that could've been done whilst being at home or maintaining social distancing



Insights

As we can see in the bar charts given above us, we have surveyed 92 students. We had asked them to rate on a scale of 1 to 10, how physically active were they before and during

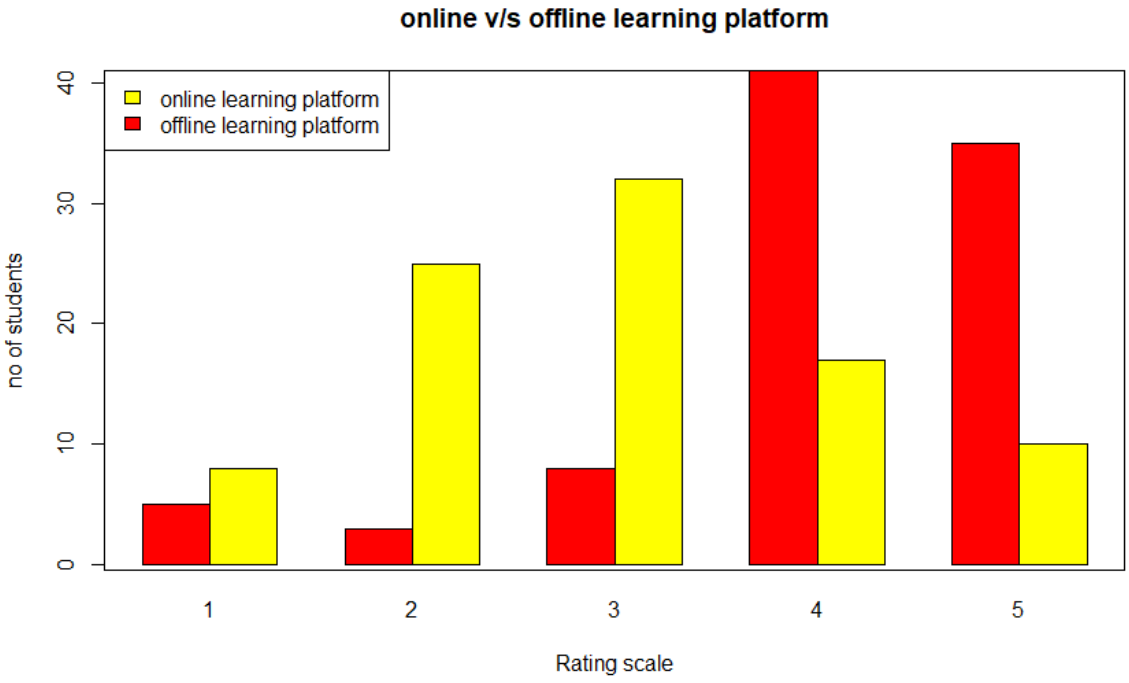
the lockdown. Most students rated 8 as their level of activity before lockdown, 29 respondents to be exact as per the graph. Followed by 18 responses for 7 and 15 students rating 9. Most responses lie in the range of 8 to 10 which is 67.3% of responses. If we see the scale during the lockdown, we see a significant drop in the scale of physical activity. With most responses being rated 5, followed by 4 and 6. The number of respondents for the same is 16, 15 and 13 respectively.

R code :

```
d<-read.csv("C://Users//user//Documents//R PRGS//R prgs//dataset.csv")

physically_active_during<-table(d$physically_active_during)

physically_active_before<-table(d$physically_active_before)

physically_active_during


physically_active_before


t5=rbind(physically_active_before,physically_active_during)

t5


barplot(height=t5,beside = TRUE,col=c('green','red'),xlab='Rating scale',ylab='no of students')

legend('topleft',legend=c('Physically active before Lockdown','Physically active during Lockdown'),fill=c("green","red"))

box()
```

ASSESSMENT OF OBJECTIVE 4

In here we decided to ask the students on how they rate the Education platforms, We compared the online vs offline education platforms and as seen below are the results we acquired.



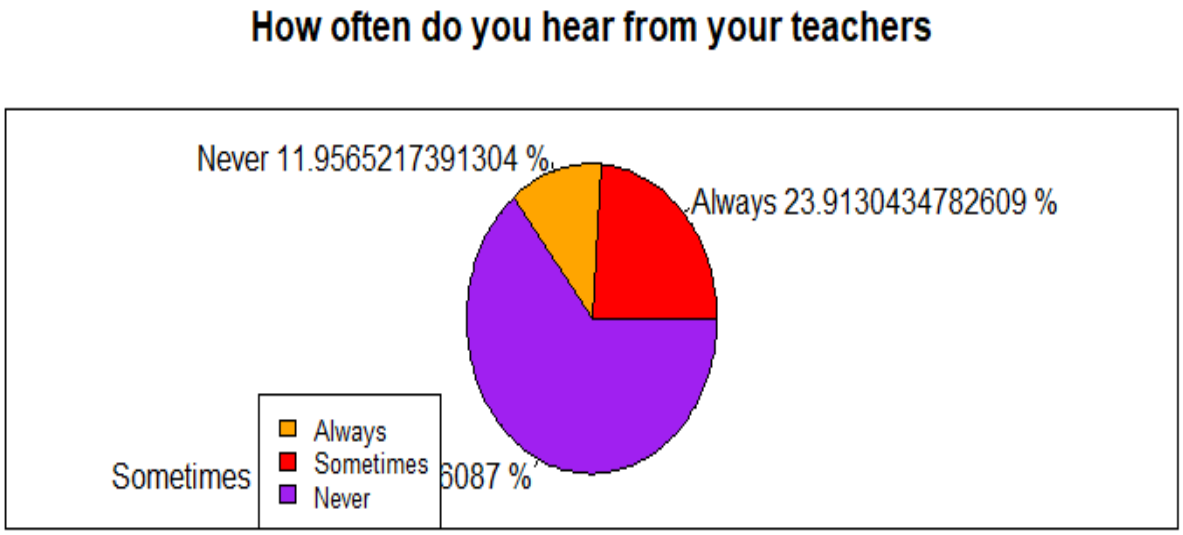
As we can see above we had maximum responses for students rating offline learning platform as more preferable with 41 respondents rating it as 4 and 35 respondents rated it as 5. Comparing that to the online learning platform we see a drop in the scale with majority of the responses being 3 and 2 with the responses being 32 and 25 respectively.

R code :

```
rate_offline<-table(data2$rate_offline_learning_platform)
rate_offline
Ratings :            1  2  3  4  5
No. of students:5  3  8 41 35
rate_online<-table(data2$rate_online_learning_platform)
rate_online
rate_offline<-table(data2$rate_offline_learning_platform)
rate_offline
Ratings :            1  2  3  4  5
No. of students:5  3  8 41 35
rate_online<-table(data2$rate_online_learning_platform)
rate_online
Ratings: 1  2  3  4  5
 No. of students:8 25 32 17 10
gg=rbind(rate_offline,rate_online)
gg
1  2  3  4  5
rate_offline5  3  8 41 35
rate_online  8 25 32 17 10
Maximum people finds offline as better learning platform
barplot(height=gg,beside = TRUE,col=c('red','yellow'),xlab='Rating scale',ylab='no of students')
legend('topleft',legend=c('online learning platform','offline learning platform'),fill=c("yellow","red"))
box()
```

ASSESSMENT OF OBJECTIVE 5

Now that we were aware that how students rate online learning, we further dive into the same aspect by asking them, while they indulge in distance education. How often do they hear form their educators? The results are as follows.



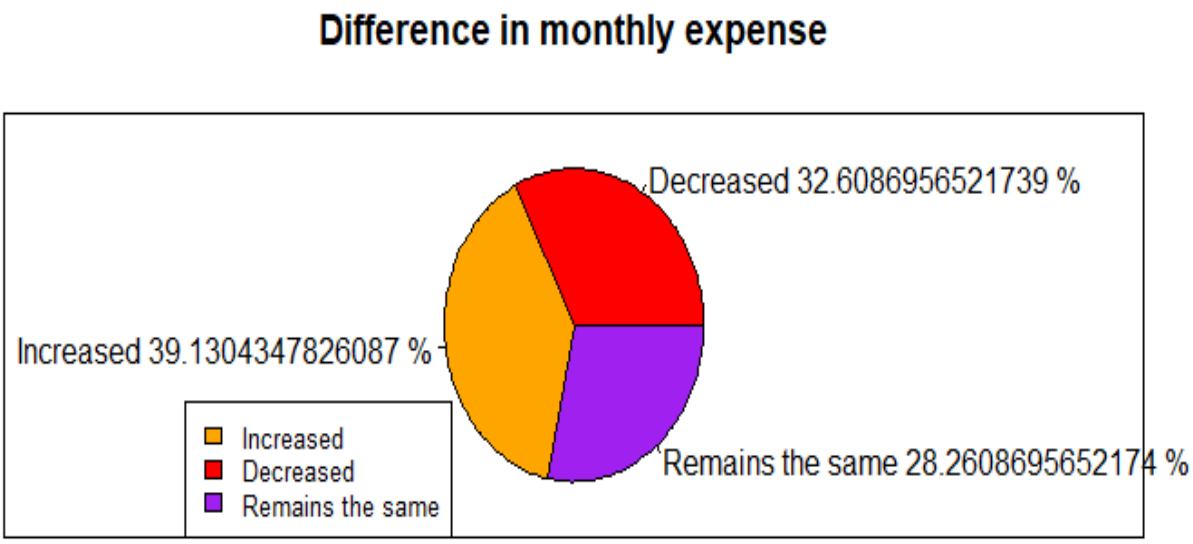
As we can see majority of the students hear from their teachers “sometimes”. 64.1% of all the respondents to be exact. While 11.9% of the students felt they have never heard from their teachers.

R code:

```
table(teacher)
teacher
Always     Never Sometimes
      22        11        59
> names(table(teacher))
[1] "Always"    "Never"     "Sometimes"
> (table(teacher)/92)*100
teacher
   Always     Never Sometimes
23.91304  11.95652  64.13043
> lbls_2<-paste(names(table(teacher)),(table(teacher)/92)*100,"%")
> pie(table(df$teacher),labels = lbls_2,main = "How often do you hear from your teachers",col = c("red","orange","purple"))
> legend("bottomleft",c("Always","Sometimes","Never"),cex = 0.8,fill = c("orange","red","purple"))
>box()
```

ASSESSMENT OF OBJECTIVE 6

Another aspect of life which was affected not just for the students but also for every other individual was their finances. They were disrupted on a large scale for many people. We asked students if they saw any chance in their monthly expenses, be it increase or decrease.



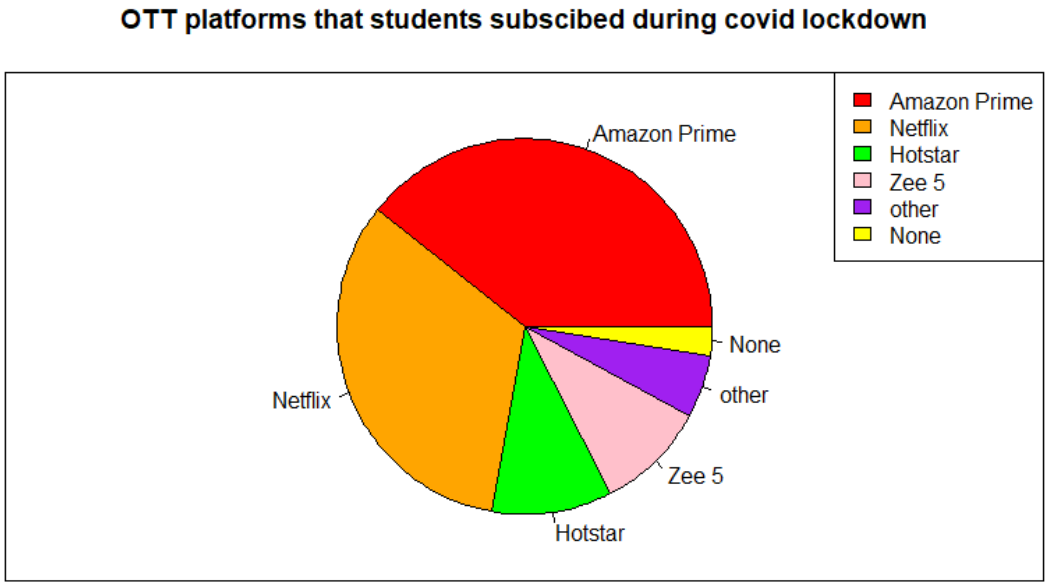
The results for this were a mixed bag with no clear majority for any of the aspect. Although we can see a slight inclination towards an increase in peoples expenses with 39% responses saying the same compared to 32.6% and 28.2% being for decreased and remained the same respectively.

R code :

```
table(monthly.expense)

monthly.expense

      Decreased       Increased Remains the same

             30              36              26

> names(table(monthly.expense))

[1] "Decreased"       "Increased"       "Remains the same"

> (table(monthly.expense)/92)*100

monthly.expense

      Decreased       Increased Remains the same

       32.60870        39.13043        28.26087

> lbls_3<-
paste(names(table(monthly.expense)),(table(monthly.expense)/92)*100,
"%")

> pie(table(df$monthly.expense),labels = lbls_3,main = "Difference in
monthly expense",col = c("red","orange","purple"))

>legend("bottomleft",c("Increased","Decreased","Remains the
same"),cex = 0.8,fill = c("orange","red","purple"))

>box()
```

ASSESSMENT OF OBJECTIVE 7

Our further two objectives are the two things this generation cannot live without, the use of social media and binge watching on OTT platforms. In this part we asked them which OTT platform did they subscribed to. This does not include renewed membership. It is based on which new platform did they subscribe to for the first time. The results for the same are below



As seen above, Netflix and Amazon prime lead the way for the OTT platforms most subscribed to. With major movie releases being shifted from on screen to these famous OTT platforms. The other notable ones being Hotstar, Zee5. Surprisingly we also had students who subscribed to no new streaming platform

R code :

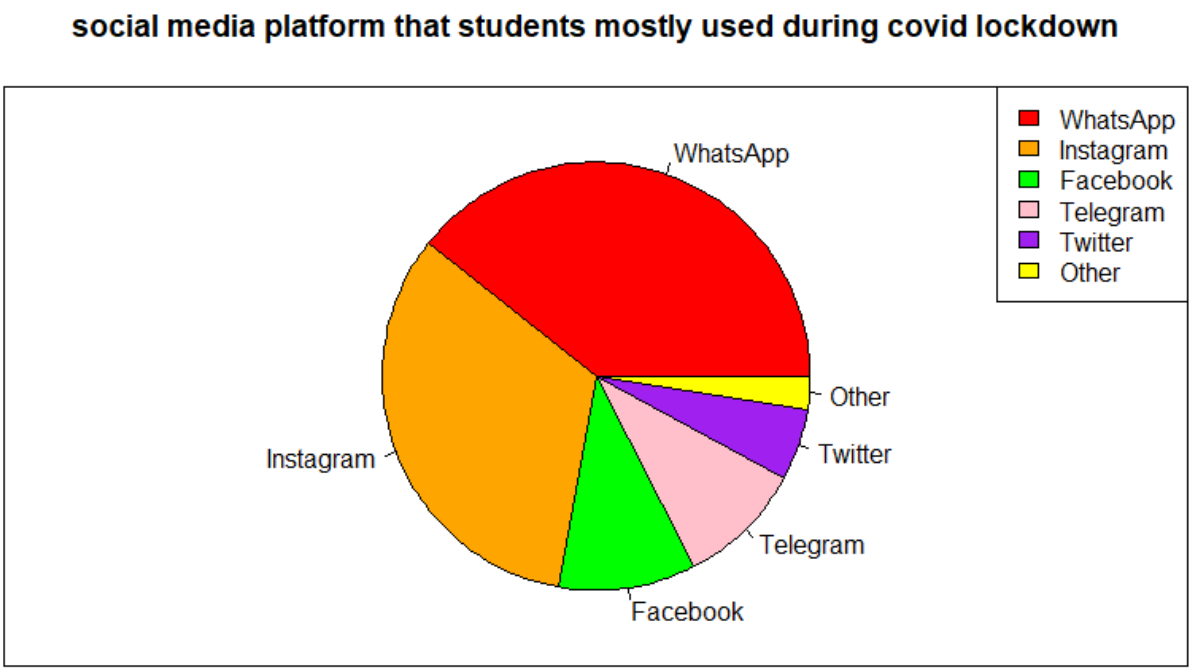
```
data20<-read.csv("dataset3.csv",header=TRUE)

data20
```

```
   subscribed     no_of_subscriptions
1 Amazon Prime          55
2      Netflix          60
3      Hotstar           35
4        Zee 5           12
5        other           2
6         None            7
```

```
indice<-data10[,2]!=0

pie(data10[,2][indice],labels =
data20$subscribed,col=c("red","orange","green","pink","purple","yellow"),
main="OTT platforms that students subscibed during covid lockdown")

legend("topright",legend=data20$subscribed,fill=c("red","orange","green"
,"pink","purple","yellow"))

box()
```

ASSESSMENT OF OBJECTIVE 8

Our final objective being to see which social media platforms were the students most active on ?.



As displayed above, we can see that a huge majority of students were active on WhatsApp took up almost half of the pie chart, followed by Instagram too majorly compared Facebook, Telegram, Twitter and others.

R code :

```
data10<-read.csv("dataset2.csv",header=TRUE)
data10
social_medianno_of_people
1      WhatsApp          79
2     Instagram          66
3      Facebook          21
4      Telegram          19
5       Twitter          11
6         Other           5

indices<-data10[,2]!=0
pie(data10[,2][indices],labels =
data10$social_media,col=c("red","orange","green","pink","purple","yellow
"),main="social media platform that students mostly during covid
lockdown")
legend("topright",legend=data10$social_media,fill=c("red","orange","gree
n","pink","purple","yellow"))
box()
```

STATISTICAL ANALYSIS

Between offline Learning ratings and reopening of School/Colleges thoughts

R Code to find Chis-square & correlation between X(Offline Learning Platform) and Y(Reopening of School/Colleges)

R Code:

```
reopen<- data$Reopening_thoughts
reopen[reopen == 'Yes']<- 1
reopen[reopen == 'No']<- 0
r<- table(data$offline_rating,reopen)
#r


chisq.test(r)
```

```
##
##  Pearson's Chi-squared test
##
## data:  r
## X-squared = 9.6829, df = 4, p-value = 0.04612
```

```
r1<- rbind(r[1,]+r[2,]+r[3,]+r[4,],r[5,])
colnames(r1)<- c("NO","YES")
rownames(r1)<- c("don't prefer","prefer")
r1
```

```
##             NO YES
```

```
## don't prefer 27  30
## prefer       6  29
```

```
Yule(t(r1))
```

```
## [1] 0.6261682
```

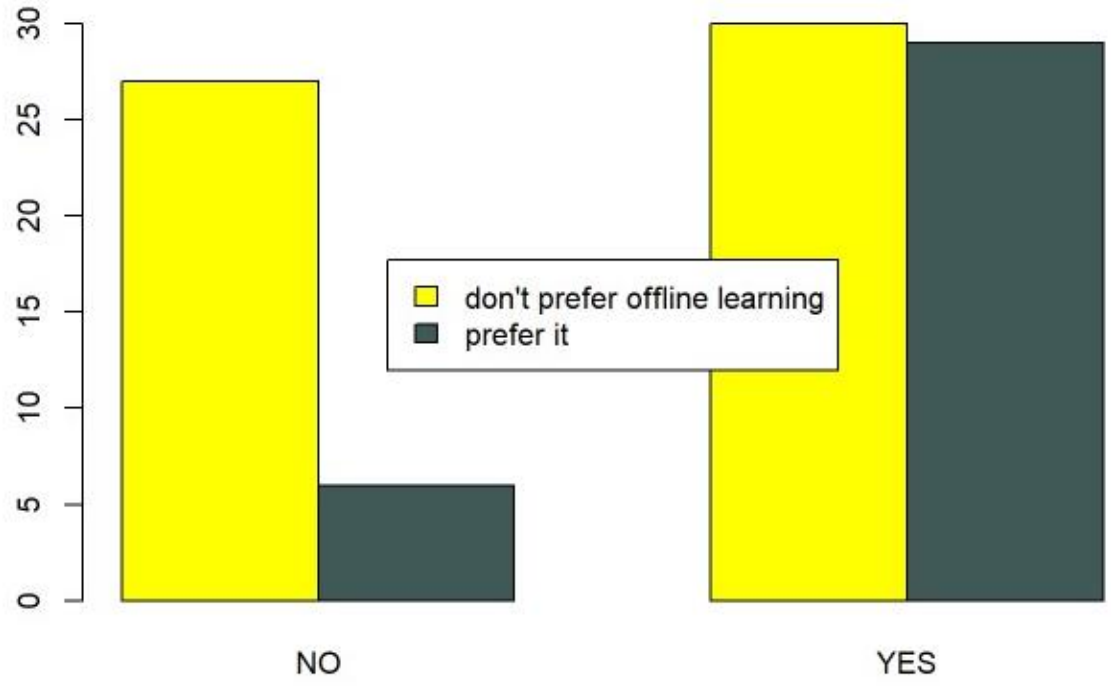
```
epi.2by2(t(r1))
```

```
##             Outcome +    Outcome -      Total       Inc risk *       Odds
## Exposed +          27          6       33            81.8       4.50
## Exposed -          30         29       59            50.8       1.03
## Total              57         35       92            62.0       1.63
##
## Point estimates and 95% CIs:
## -------------------------------------------------------------------
## Inc risk ratio                          1.61 (1.19, 2.17)
## Odds ratio                              4.35 (1.57, 12.08)
## Attrib risk in the exposed *                30.97 (12.64, 49.30)
## Attrib fraction in the exposed (%)          37.85 (16.28, 53.87)
## Attrib risk in the population *             11.11 (-5.05, 27.27)
## Attrib fraction in the population (%)       17.93 (4.70, 29.32)
## -------------------------------------------------------------------
## Uncorrected chi2 test that OR = 1: chi2(1) = 8.612 Pr>chi2 = 0.003
## Fisher exact test that OR = 1: Pr>chi2 = 0.004
##  Wald confidence limits
##  CI: confidence interval
##  * Outcomes per 100 population units
```

```
barplot(r1,beside = TRUE,col=c("#FFFF00","#405856"))
```

```
legend("center",legend = c("don't prefer offline learning","prefer it"),fill= c("#FFFF00","#405856"))
```

```
##Yules Coefficient: 0.6261682
```



Conclusion: On studying the above trend, we conclude that most of the students prefer Offline teaching

Major reasons for this could be,

• Continuous use of electronic gadgets can affect the concentration levels of students drastically.

• Less Physical Activity of students can make student s lazy leading to low concentration levels

Did happier people have less screen time

Here we wish to see if screen time dependent on the happiness of the people. Hence, here we do Chi-square test of significance to find our conclusion.

Hypothesis:

H0:Increase in Screen time is independent of the happy mood of people

H1:Increase in Screen time is dependent on the happy mood of people

Test of significance: Chi square test

Decision Criterion: If the P-value is less than (or equal to) α, then the null hypothesis is rejected in favour of the alternative hypothesis. And, if the P-value is greater than α, then the null hypothesis is not rejected.

Analysis:

Here Level of significance (α) = 0.05

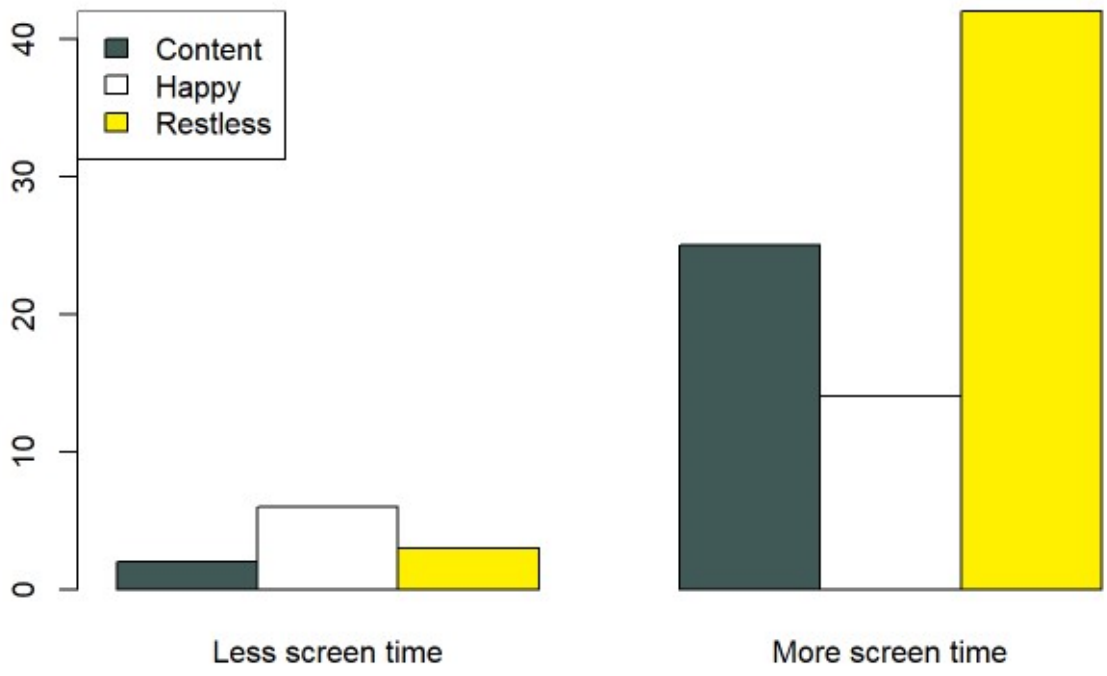
R-Code:

```
z<- table(data$Health,data$Screen_time)
z
##
##          No Yes
##   Content   2  25
##   Happy     6  14
##   Restless  3  42
chisq.test(z)
```

```
barplot(z,beside = TRUE,col=c("#405856","#FFFFFF","#FFF000"))
```

```
legend("topleft",legend = c("Content","Happy","Restless"),fill= c("#405856","#FFFFFF","#FFF000"))
```

```
##
##  Pearson's Chi-squared test
##
## data:  z
## X-squared = 7.9124, df = 2, p-value = 0.01914
```



Conclusion:

Happier people have lesser screen time

Major reasons could be:

- More screen time indicates lesser physical activity which makes student restless.
- Studies found that, too much screen time can interfere with getting enough exercise, doing homework, being with friends, and spending time with family. It also can contribute to obesity, attention and learning problems, and sleep problems.

Monthly expenses on reopening thoughts of Colleges and Schools

R-Code:

```
c1<- table(data$monthly_expense,data$Reopening_thoughts)
c1
##
##                 No Yes
##   Decreased       5  25
##   Increased      21  15
##   Remains the same  7  19
chisq.test(c1)
##
##  Pearson's Chi-squared test
##
## data:  c1
## X-squared = 13.611, df = 2, p-value = 0.001108
c2<- rbind(c1[1,]+c1[3,], c1[2,])
c3<- cbind(c2[,2],c2[1,])
c3[2,2]=21
colnames(c3)<- c("Yes","No")
rownames(c3)<- c("Decreased/the same","Increased")
c3
##                 Yes No
## Decreased/the same  44 12
## Increased         15 21
Yule(c3)
```

```
## [1] 0.673913
```

```
epi.2by2(c3)
```

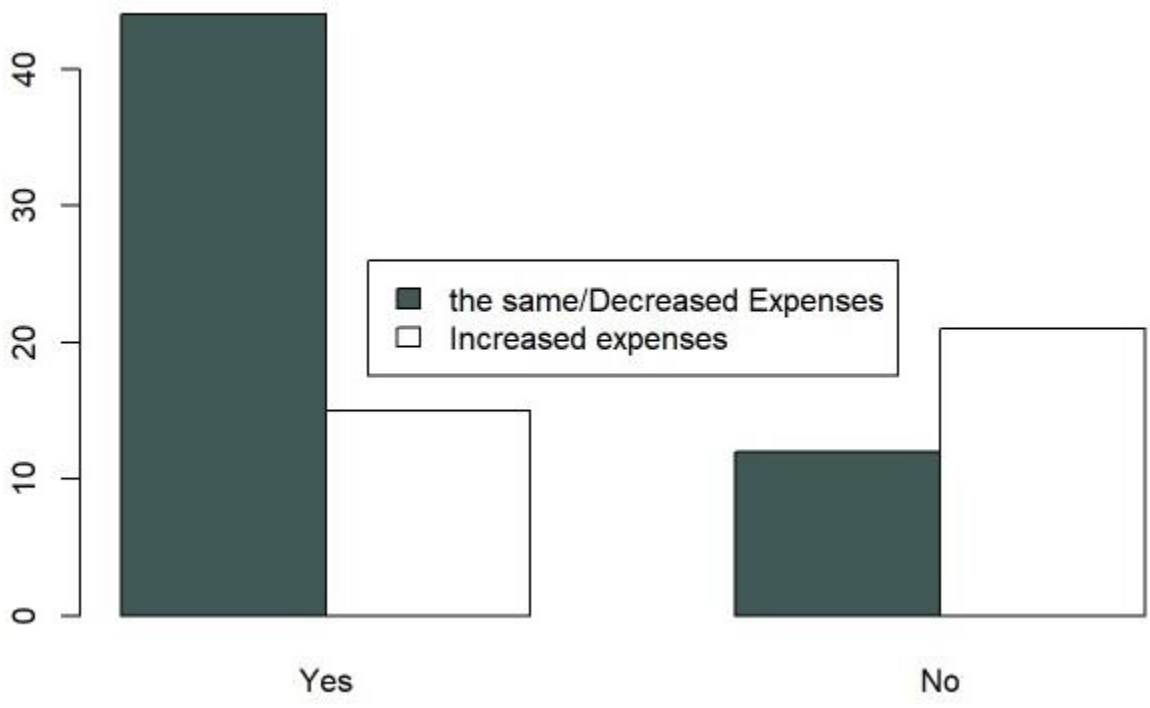
```
##             Outcome +    Outcome -      Total       Inc risk *       Odds
## Exposed +          44          12         56            78.6      3.667
## Exposed -          15          21         36            41.7      0.714
## Total              59          33         92            64.1      1.788
##
## Point estimates and 95% CIs:
## -------------------------------------------------------------------
## Inc risk ratio                          1.89 (1.25, 2.84)
## Odds ratio                              5.13 (2.05, 12.88)
## Attrib risk in the exposed *                36.90 (17.54, 56.27)
## Attrib fraction in the exposed (%)          46.97 (20.09, 64.81)
## Attrib risk in the population *             22.46 (3.61, 41.32)
## Attrib fraction in the population (%)       35.03 (11.32, 52.40)
## -------------------------------------------------------------------
## Uncorrected chi2 test that OR = 1: chi2(1) = 12.974 Pr>chi2 = <0.001
## Fisher exact test that OR = 1: Pr>chi2 = <0.001
##  Wald confidence limits
##  CI: confidence interval
##  * Outcomes per 100 population units
```

```
barplot(c3,beside = TRUE,col=c("#405856","#FFFFFF"))
```

```
legend("center",legend = c("the same/Decreased Expenses","Increased expenses"),fill=
c("#405856","#FFFFFF"))
```

```
##  Pearson's Chi-squared test
##
```

```
## data:  c1
## X-squared = 13.611, df = 2, p-value = 0.001108
##Yules Coefficient: 0.8298677
```



Conclusion:

Since Yule's coefficient = 0.8298677, we say that there is a positive correlation between Monthly expenses and reopening of colleges/ schools.

People who are willing to go to colleges/schools have the same/decreased household expenses.

On the other hand, people who are reluctant to go to school/colleges had increased expenses during lockdown.

Less Significant objectives:

→Online rating depends on engagement with teachers

```
x<- table(data$online_rating,data$engagement_with_teachers)
x
##
##    Always Never Sometimes
##  1     1    3       4
##  2     4    3      18
##  3    11    2      19
##  4     3    1      13
##  5     3    2       5
chisq.test(x)
## Warning in chisq.test(x): Chi-squared approximation may be incorrect
##
##  Pearson's Chi-squared test
##
## data:  x
## X-squared = 10.61, df = 8, p-value = 0.2248
```

Dependency of online rating upon engagement with teachers is less significant as its p-value is 0.2248 i.e p-value>0.05

→Increased screen time caused change in thoughts on social issues

```
d<- table(data$Screen_time[data$Health !="Happy"],          data$effect_on_social_issues[data$Health != "Happy"  ])
d
##
##      No Yes
##   No   0   5
##   Yes 17  50
d1<- cbind(d[,2],d[,1])
d2<- rbind(d1[2,],d1[1,])
d2
##      [,1] [,2]
## [1,]   50   17
## [2,]    5    0
chisq.test(d2)
## Warning in chisq.test(d2): Chi-squared approximation may be incorrect
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  d2
## X-squared = 0.55191, df = 1, p-value = 0.4575
```

Dependency of thoughts on social issues with screen time is less significant as its p-value is 0.4575 i.e p-value>0.05

→Efficiency during lockdown and screen time

```
b<-table(data$efficiency_during,data$Screen_time)
b
##
##             No Yes
##   0-2 hours    6  34
##   2-5 hours    3  23
##   More than 5  2  24
chisq.test(b)
## Warning in chisq.test(b): Chi-squared approximation may be incorrect
##
##  Pearson's Chi-squared test
##
## data:  b
## X-squared = 0.80539, df = 2, p-value = 0.6685
```

Dependency of Efficiency during lockdown on Screen time is less significant as its p-value is 0.6685 i.e p-value>0.05

METHODOLOGY

R studio: helped to plot graphs and calculate chi-square values for Data Analysis and data visualization.

Libraries used:-

1) For importing/retrieving csv files

Library(readr)

Library(dplyr)

Library(tidyr)

2) For Hypothesis testing

library(dplyr)

library(ggplot2)

library(janitor)

library(psych)

library(epiR)

3) For Plotting Graph

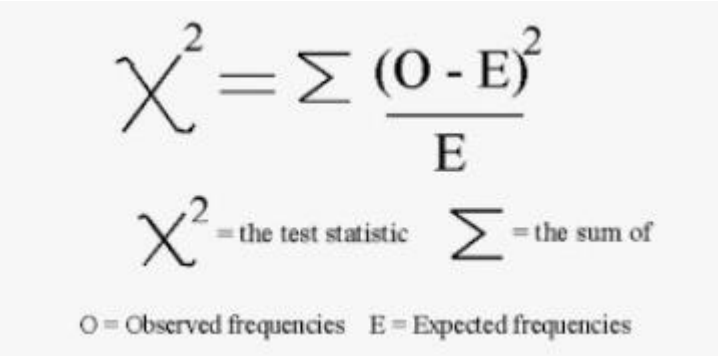
library(ggplot2)

Chi-Square Test- is a hypothesis testing method. Two common Chi-square tests involve checking if observed frequencies in one or more categories match expected frequencies. χ is the Greek symbol for Chi. A **chi-squared test**, also written as **χ^2 test**, is a statistical hypothesis test that is valid to perform when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test.

Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table.

Formulas used:

- For a test of homogeneity, **df = (Rows − 1)×(Cols – 1)**



Google studio-

Google studio helped us to visualize our data accurately with its wide variety of features

REFERENCES:-

- Our Github Repository
  https://github.com/reeve-k-15/DatAnalysis_CovidEffects
- Youtube- Data Analysis references
  epiR function -
  https://www.youtube.com/watch?v=V_YNPQoAyCc
- Books: Lecture notes & data analysis books(pdf)
- Google Form link-
  https://docs.google.com/forms/d/e/1FAIpQLSeuD1aoSfbSk_cvajAG46MmMaW_h_ivNMThSmhBEHiw2SvLVw/viewform

LIMITATIONS

- More variety of questions could have been added to get a better overall idea of how covid affected the various other factors to bring light upon.
- Due to a very limited reach of the survey, we couldn't get a large amount of responses that would have given us better accurate results leading to the optimal solution.
- Human Errors like students not giving true responses or submitting without knowing what it was about also brings a slight negative solution in analysis.