DEPARTMENT OF STATISTICS

"STUDY OF PEOPLE'S OPINION OVER COVID-19"

T.Y.B.Sc.

(2020-2021)

Guided by: Prof. S. M. Pawar

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,

"Study of People's Opinion over Covid-19"

is satisfactorily completed by,

1. **Srushti Bhalerao**
2. **Kshitij Chalke**
3. **Advay Chandratre**
4. **Yukta Chavan**
5. **Priyanka Chinchole**
6. **Devendra Deshmane**
7. **Abhishek Dhumal**

the students of Department of Statistics, H.P.T. Arts & R.Y.K Science College, Nashik.

Project Guide
Prof. S. M. Pawar

Head of Department
Prof. V. S. Joshi

ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance from many people and we are extremely privileged to have got this along the completion of our project. All that we have done is only due to such supervision and we would not forget to thank them.

We thank **Savitribai Phule Pune University** for providing us an opportunity to do the project work in Statistics.

We express our gratitude to **Principal V. N. Suryawanshi**, **HPT Arts & R.Y.K Science College** for giving us all support and allowing us to present the project.

We owe our deep gratitude to our **Head of Department Prof. Mrs. V. S. Joshi** and our **Project guide Prof. Ms. S. M. Pawar**, who took keen interest in our project work and guided us all along, till the completion of our project work by providing all the necessary information for making a good project. Also, we would like to extend our sincere gratitude to all staff in laboratory for their timely support.

Last but not the least we would like to extend our sincere gratitude towards the people who gave the necessary information required for the project.

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

INDEX FOR GRAPHS AND CHARTS

INDEX FOR ANALYSIS OF DATA

INTRODUCTION

COVID-19 (Coronavirus) has affected day to day life and and is slowing down the global economy. This pandemic has affected thousands of peoples, who are either sick or are being killed due to the spread of this disease. The most common symptoms of this viral infection are **fever, cold, cough, bone pain and breathing problems** and ultimately leading to **pneumonia**. The emphasis is on taking extensive precautions such as extensive **hygiene protocol (e.g. regularly washing of hands, avoidance of face to face interaction, etc.), social distancing, and wearing masks**, and so on. This virus is spreading exponentially region wise. Countries are banning gatherings of people to the spread and break the **exponential** curve. Many countries are locking their population and enforcing strict quarantine to control the spread of the havoc of this highly communicable disease.

COVID-19 has rapidly affected our day to day life, businesses, disrupted the world trade and movements. Identification of the disease at early stage is vital to control the spread of the virus because it spreads rapidly from person to person. Most of the countries have slowed down their manufacturing of the products. The various industries and sectors are affected by the cause of this disease; these include the **pharmaceuticals industry, solar power sector, tourism, Information and electronics industry**. This virus creates significant knock-on effects on the daily life of citizens, as well as about the global economy.

Presently the impacts of the COVID-19 in daily life are extensive and have far reaching consequences. These can be divided into various categories:

1. **<u>Healthcare</u>**
    - Challenges in the diagnosis, quarantine and treatment of suspected or confirmed cases.
    - High burden of functioning of the existing medical system.
    - Patients with other disease and health problems are getting neglected.

- Overload on doctors and other healthcare professionals, who are at a very high risk.
- Overloading of medical shops.
- Requirement of high protection.
- Disruption of medical supply chain.

2. **Economic**
    - Slowing of the manufacturing of essential goods.
    - Disrupt the supply chain of products.
    - Losses in national and international business.
    - Poor cash flow in market.
    - Significant slowing down in the revenue growth.
3. **Social**
    - Service sector is not being able to provide their proper service.
    - Cancellation or postponement of large-scale sports or tournaments.
    - Avoiding the national and international travelling and cancellation of services.
    - Disruption of celebration of cultural, religious and festive events.
    - Undue stress among the population.
    - Social distancing with our peers and family members.
    - Closure of the hotels, restaurants and religious places.
    - Closure of places for entertainment such as movie and play theatres, sports clubs, gymnasium, swimming pools, and so on.
    - Postponement of examinations.

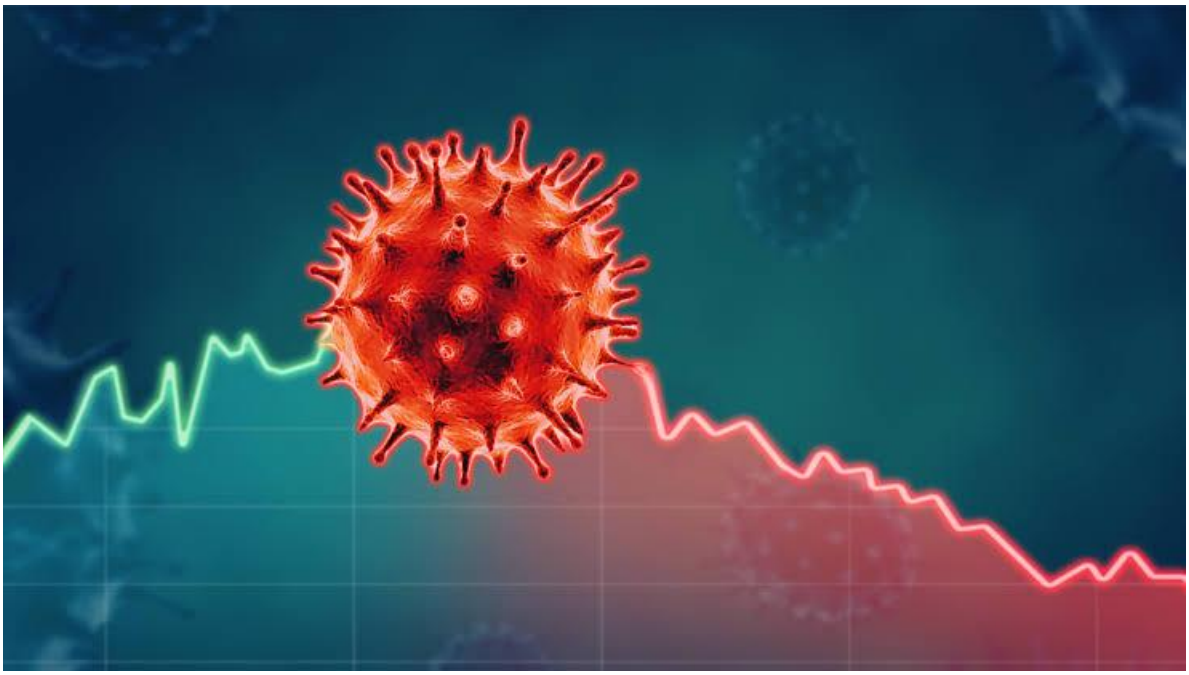
The number of informants taken under consideration for the project are 250. Out of 250 informants, there are 122 males and 128 females. The project includes statistical analysis of data collected through different areas of Nashik city.

OBJECTIVES

The main objectives of our project are listed below:

- To get an idea about people's opinion over COVID-19 pandemic
- To check whether their health conditions depend on their age.
- To get an idea about the treatment people prefer for getting cure from a particular type of disease.
- To study changes in eating and shopping habits of the people.
- To study the effect of pandemic on people' mental and physical health.
- To study effect of COVID-19 on various sectors like education, hospitality, tourism, trade, business, health, etc.
- To analyze people's opinion over lockdown, education pattern, lifestyle after COVID-19, etc.
- To know what option do people think as the best for avoiding spread of COVID-19 pandemic.

FACTS ABOUT COVID-19



Coronavirus disease 2019(COVID-19), is also known as the coronavirus, or COVID, is a contagious disease caused by severe acute respiratory syndrome coronavirus 2(SARS-CoV-2). The first known case was identified in Wuhan, China in December-2019. The disease has since spread worldwide, leading to an ongoing pandemic.

Symptoms of COVID-19 are variable, but often include fever, cough, headache, fatigue, breathing difficulties, and loss of smell and taste. Symptoms may begin one to fourteen days after exposure to the virus. At least a third of people who are infected do not develop noticeable symptoms. Of those people who develop noticeable symptoms enough to be classed as patients, most (81%) develop mild to moderate symptoms (up to mild pneumonia), while 14% develop severe symptoms (dyspnea, hypoxia, or more than 50% lung involvement on imaging), and 5% suffer critical symptoms (respiratory failure shock, or multiorgan dysfunction). Older people are at higher risk of developing severe symptoms. Some people

continue to experience a range of effects for months after recovery, and damge to organs has been observed. Multi-year studies are underway to further investigate the long term effects of the disease.

Transmission of COVID-19 occurs when people are exposed to virus-containing respiratory droplets and airborne particles exhaled by an infected person. Those particles may be inhaled or may reach the mouth, nose, or eyes of a person through touching or direct deposition (i.e. being coughed on). The risk of infection is highest when people are in close proximity for a long time, but particles can be inhaled over longer distances, particularly indoors in poorly ventilated and crowded spaces. In those conditions, small particles can remain suspended in air for minutes to hours. Touching a contaminated surface or object may lead to infection although this not contribute substancially to transmission. People who are infected can transmit the virus to another person up to two days before they themselves show symptoms, as can people who do not experience symptoms. People remain infectious for up to ten days after the onset of symptoms in moderate cases and upto twenty days in severe cases.

Several testing methods have been developed to diagnose the disease. The standard diagnostic method is by detection of virus nucleic acid by real-time reverse transcription polymerase chain reaction (rRT-PCR), transcription-mediated amplification (TMA), or by reverse transcription loop-mediated isothermal amplification (RT-LAMP) from a nasopharyngeal swab.

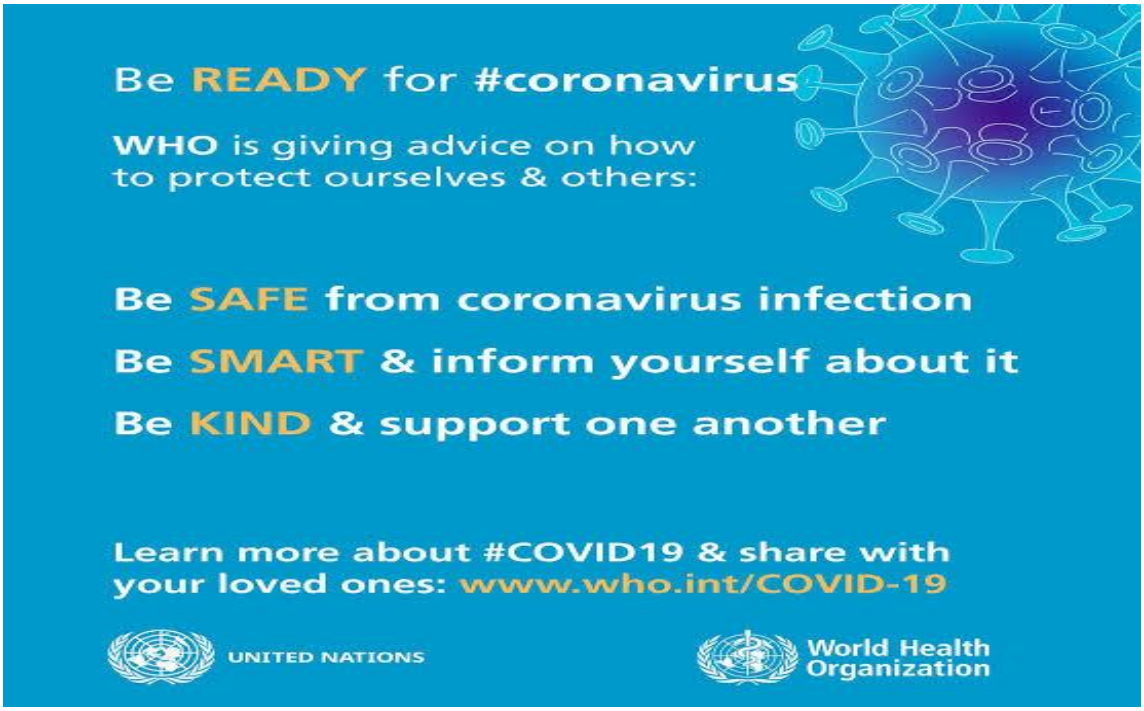
HIGH-RISK GROUPS

1. **Older Adults:-**
   Older Adults are more likely to have long-term health problems that can put them at risk. People's immune systems tend to weaken with age, making it more difficult for older people to flight off infections. Lung tissue becomes more elastic over time, making respiratory diseases like COVID-19 a particular concern for older people. Inflammation in older people can be more intense, causing organ damage.
2. **Heart disease:-**
   Although COVID-19 most often affects the airway and lungs, these organs work together with the heart to drive oxygen to the body tisuues. When the lungs are overtaxed due to illness, the heart has to work harder, which creates challenges for people who are already living with heart disease. The American Heart Association notes that viral illnesses similar to COVID-19 can raise the risk of heart attack in people with a buildup of plaque in their blood vessels. Research shows that viral illness can make it more likely that a piece of plaque lining the vessels could break off and block blood flow to the heart.
3. **Lung disease:-**
   Chronic airway and lung diseases such as chronic obstructive pulmonary disease (COPD), asthma, pulmonary fibrosis and interstitial lung disease can set the stage for a more severe infection with the new coronavirus because of scarring, inflammation, or lung damage. It's very important for people with these conditions to work with their doctors and ensure they have adequate supplies of maintenance and rescue medications on hand.

4. **Diabetes:-**

   People living with diabetes have an increased risk of getting very sick from the new coronavirus. Diabetes type 1 and type 2 both cause an increase in blood sugar. Poorly controlled blood sugar can make viral diseases including COVID-19, more dangerous, possibly because higher blood sugar can create an environment where viruses are likely to thrive. In addition, diabetes increases inflammation and weakens the immune system, making it harder for people living with the condition to fight off disease in general. Those living with diabetes should adhere to their medications regimens and do everything possible to keep their blood sugar under control. Having an adequate supply of medications and staying in close touch with doctors can help.

GUIDELINES BY WHO



To prevent the spread of COVID-19:

- Clean your hands often. Use soap and water, or an alcohol-based hand-rub.
- Maintain a safe distance from anyone who is coughing and sneezing
- Wear a mask when physical distancing is not possible.
- Don’t touch your eyes, nose or mouth.
- Cover your nose and mouth with your bent elbow or tissue when you cough or sneeze.
- Stay home if you feel unwell.
- If you have fever, cough, and difficulty breathing, seek medical attention.

DATA COLLECTION

Collection of data is very important work and it is to be done carefully. The methods of data collection are:

(1)Surveys (2) Laboratory experiments (3) Simulations

For our project we decided to use **survey method** of data collection. With the help of sample surveys primary or secondary data can be collected.

Primary Data: Primary data means original data obtained by an investigator himself. Primary data is also called as raw data.

Secondary Data: Secondary data means data taken from sources like office record, bulletin, reports, etc. which is already collected by some other agencies.

We came to a decision that to investigate the objectives of our project, the best method is sample survey. The primary data required for this survey is collected by doing investigation through questionnaire. So our aim was to prepare a good and appropriate questionnaire. We prepared the questionnaire considering all the rules and regulations that we have studied in topic.

"Sample Survey"

To study the effect of COVID-19 on people, firstly we included their age, gender, marital status, annual family income, number of earning members in family. Further, we enquired whether they are suffering from any disease and what treatment do they prefer for getting cure. Also, it was important to know whether they were having health insurance or not. One of the main objective of our project was to study the effect of the pandemic over sectors like education, tourism, lifestyle, health, etc. So for that we asked them questions related to the changes occurred in their daily habits after COVID-19. We are also interested to know whether they are satisfied with the measures taken by the government. Moreover, what according to them is the best option to avoid spread of virus.

SAMPLING TECHNIQUE

Population, an aggregate of objects or individuals under study**.** In order to study a population, we collect information about each and every element in the population. This method is called as census or complete enumeration. But for infinite population census cannot be used. To overcome the limitations of census, sampling is used.
Sample: Any part of the population under study which represents all the characteristics of population is called as "Sample".
A success of sampling method mainly depends upon proper selection of sampling method. Sampling method are of two types,

(1)**Non-random sampling:** In this method investigator select elements in any manner suitable to him/her.

(2)**Random sampling:** In this method the selection of units in sample is done impartially. Personal or any kind of bias in the selection of unit is avoided.

Types of Random Sampling:

(a)Simple random sampling, (b)Stratified Sampling, (c)Systematic Sampling, (d) Cluster Sampling, etc.

Simple Random Sampling: The technique in which each unit has an equal chance of selection in sample is called as "Simple Random Sampling".

Types of Simple Random Sampling:

Simple Random Sampling with Replacement(SRSWR): Units are replaced back in population.

Simple Random Sampling without Replacement(SRSWOR): Units are not replaced back in the population.

"For our project we used Simple Random Sampling without Replacement method for data collection. We collected the data using following questionnaire prepared using Google forms".

QUESTIONNAIRE

- **Age: _____**
- **Gender:** ☐ Male ☐ Female
- **Marital Status:** ☐ Married ☐ Unmarried
- **Working status,**

☐ Full time ☐ Part-time ☐ Unemployed ☐ Self-employed
☐ Student ☐ Homemaker ☐ Business ☐ Retired ☐ Military
☐ Other:_____

- **Nature of Working (Before COVID-19)**

☐ Office Work ☐ Field Work ☐ Work from Home

- **Nature of Working (After COVID-19)**

☐ Office Work ☐ Field Work ☐ Work from Home

- **Family:** ☐ Joint ☐ Nuclear
- **Annual Family Income(in lakh Rs.)**:

☐ Below 2 ☐ 2-4 ☐ 4-6 ☐ 6-8 ☐ 8-10 ☐ 10 And Above

- **No. of earning members in family:** ☐ Only 1 ☐ More than 1
- **Are you suffering from any disease?** ☐ Yes ☐ No
- **If Yes,**

☐ Diabetes ☐ Cancer ☐ Kidney Disease ☐ High BP
☐ Heart Disease ☐ COPD ☐ Thyroid ☐ Other:_____

- **Treatment Prefered:**

☐ Allopathy ☐ Ayurvedic ☐ Naturopathy ☐ Homeopathy
☐ Other:_____

- Do you have health insurance? ☐ Yes ☐ No
- **If yes,** ☐ Before COVID ☐ After COVID
- **What food do you prefer eating?**

☐ Healthy Food ☐ Junk Food ☐ Diet Food ☐ Homemade Food
☐ Ready to eat Food ☐ Frozen Food

- **Have your eating habits changed after COVID-19?**

☐ Yes ☐ No ☐ Maybe

- **Would You Look For Eating Fast Food If The Effect Of Covid Decreases?**

☐ Yes ☐ No ☐ Maybe

- **Do You Agree That Having Healthy Eating Habits May Help You Stay Away From Covid-19?**

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

- **What Mode Of Shopping Do You Prefer After Covid 19?**

☐ Online ☐ Offline ☐ Both

- **If Online, Reason**

☐ Fear of COVID ☐ Time-Saving ☐ Easy to use ☐ Sale and Cash Back Offers ☐ Other:_____

- **Which Method Of Payment Do You Use?**

☐ Online ☐ Offline ☐ Both

- **Did you travel anywhere since COVID-19?**

☐ Yes ☐ No

- **IF YES, What mode of transport did you prefer?**

☐ Private ☐ Public

- **If Public, what do you prefer traveling from?**

☐ Roadways ☐ Railways ☐ Airways ☐ Other_____

- **Issues faced traveling publicly**

☐ Reservation ☐ Social Distancing not practiced ☐ Delay due to security check ☐ Insufficient Staff ☐ Time for Travel ☐ Other:_____

- **Have COVID-19 affected education?**

☐ Yes ☐ No ☐ Maybe

- **Which Method Of Teaching is more useful?**

☐ Online ☐ Offline ☐ Both

- **Issues faced in Online Teaching**

☐ Communication ☐ Network Issues
☐ Lack of problem solving ☐ Classroom environment
☐ Lack of focus among students ☐ Other:_____

- **Learning is more effective in classroom than in Online**

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree
☐ Strongly Disagree

- **Has COVID affected lives of people?**

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree
☐ Strongly Disagree

- **How has COVID-19 affected you?**

☐ Physically ☐ Mentally ☐ Financially ☐ Other_____

- **Do You Think Lock-down Really Helped Preventing Spread Of Virus?**

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree
☐ Strongly Disagree

- **Are You Satisfied With The Decisions Taken By Government To Avoid Spread Of Virus?**

☐ Yes ☐ No ☐ Maybe

- **Are You OK having another lockdown?**

☐ Yes ☐ No

- **What According To You Is The Best Option To Prevent Covid-19?**

☐ Lockdown ☐ Vaccination
☐ Wearing Masks ☐ Social Distancing
☐ Other:_____

….STAY HOME STAY SAFE….

CODING

Gender	Coding
Male	0
Female	1
Other	2

M.S.	Coding
Married	0
Unmarried	1

W.S.	Coding
Full time	0
Part time	1
Unemployed	2
Self-employed	3
Student	4
Homemaker	5
Business	6
Retired	7
Military	8
Other	9

Family	Coding
Joint	0
Nuclear	1

Nature Of Work	Coding
No response	0
Office Work	1
Field Work	2
Work from Home	3

No. of earning family members	Coding
Only 1	0
More than 1	1

A.I. (in lakh Rs.)	Coding
Below 2	0
2-4	1
4-6	2
6-8	3
8-10	4
10 And Above	5

Disease	Coding
No disease	0
Diabetes	1
Cancer	2
Kidney Disease	3
High BP	4
Heart disease	5
COPD	6
Thyroid	7
Other	8

H.I.	Coding
No insurance	0
Before COVID	1
After COVID	2

Treatment	Coding
No Treatment	0
Allopathy	1
Ayurvedic	2
Naturopathy	3
Homeopathy	4
Other	5

Food	Coding
Healthy food	0
Junk food	1
Diet food	2
Homemade food	3
Ready to eat	4
Frozen Food	5

Mode of Shopping	Coding
Online	0
Offline	1
Both	2

Reason for O.S.	Coding
No Response	0
Fear of COVID	1
Time-Saving	2
Easy to use	3
Sale and Cash Back	4
Other	5

Mode of Payment	Coding
Online	0
Offline	1
Both	2

Mode of Transport	Coding
No Response	0
Private	1
Public	2

Mode Of Public Transport	Coding
No Response	0
Roadways	1
Railways	2
Airways	3
Other	4

Issues faced Traveling Publicly	Coding
No Issues	0
Reservation	1
Social Distancing not practiced	2
Delay due to security check	3
Insufficient Staff	4
Time for Travel	5
Other	6

Useful Method of Teaching	Coding
Online	0
Offline	1
Both	2

Issues faced in Online Teaching	Coding
No issues	0
Communication Issues	1
Network Issues	2
Lack of Problem Solving	3
Classroom Environment	4
Lack of focus among Students	5
Other	6

.

Best Option to Prevent COVID-19	Coding
Lockdown	0
Vaccination	1
Wearing Masks	2
Social Distancing	3
Other	4

ABBREVIATIONS & ACTUAL DATA

Y	Yes
N	No
M.S.	Marital Status
W.S.	Working Status
A.I.	Annual Income
H.I.	Health Insurance
O.S.	Online Shopping
F	Family
E.F.M	Earning Family Members
D	Disease
T	Treatment
On	Online
Off	Offline
F	Food

VARIABLES AND ATTRIBUTES

Variables	**Attributes**
Age	Gender
Annual Income	Marital Status
	Working Status
	Disease
	Treatment
	Reason

STATISTICAL TOOLS USED IN PROJECT

1. Graphs and Charts:

(a) Simple bar diagram

(b) Histogram

(c) Subdivided bar diagram

(d) Pie chart

(f) Rod plot

2. Likert's Scale

3. Fitting of Normal distribution

4. Skewness And Kurtosis

5. Testing of Hypothesis:

(a) Test for proportion for one sample

(b) Test for proportion for two sample

(c)Paired T test

(d) χ^2-Test for independence of two attribute

6. Non- parametric test:

(a) Shapiro-Wilk Test for normality

7. Regression Analysis

Logistic Regression

Packages used in project:

1. R-Software
2. MS-Excel

INFORMATION ABOUT STATISTICAL TOOLS

After collection of data the second stage is the presentation of data.
Classification is the tool of data condensation which makes the analysis easier.
Frequency: The number of observations in a class is called as frequency.
Graphs and Charts: Graphs are easy to understand and create an effect which lasts for longer time. Graph represents data and facts in an attractive and impressive manner.

1. Simple Bar Diagram:

This is the simplest way of presenting the statistical data classified according to single characteristic. It can be used to present the data like population 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.

2. Subdivided Bar Diagram:

Many times we require to represent whole quantity and its subdivisions in the same diagram. In such cases, simple bar diagram is used to represent whole quantities and subdivision can be represented proportionally by dividing each bar into number of parts. This type of diagram is called subdivided bar diagram. Subdivided bar diagram can be used to represent sex wise division of populations of different states, etc.

4. Pie Diagram or Pie Chart:

It is a special type of diagram used to represent the whole quantity by a circle and the subdivisions of the whole quantity are shown by the sectors of that circle. This diagram is a two dimensional diagram. It can be used to represent the subdivision of total budget or total income, etc.

5. Rod or Spike Plot:

It is used to represent graphically the ungrouped frequency distribution of discrete variable. It shows values of variable on x-axis and corresponding frequency on y-axis in form of vertical rod.

6. Multiple Bar Diagram:

A multiple bar graph shows the relationship between different values of data. Each data value is represented by a column in the graph. In multiple bar graph, multiple data points for each category of data are shown in addition of columns.

Skewness And Kurtosis:

Skewness and kurtosis are the characteristics of a frequency distribution for studying shape and pattern. Skewness is the property which describes the lack of symmetry in a distribution. A distribution may be skewed or symmetric. There are two types of skewness namely positive and negative.Here, we considered coefficient of skewness based on moments: If $ß1= \mu3^2/\mu2^3$, then **$\Upsilon 1=sqrt(ß1)$** .

If **$\Upsilon 1<0$**, distribution is negatively skewed.

If **$\Upsilon 1=0$**, distribution is symmetric.

If **$\Upsilon 1>0$**, distribution is positively skewed.

Kurtosis enables to have an idea about the shape and nature of peakness of the curve of frequency distribution. Measure of kurtosis gives the extent to which the distribution is peaked as compare to normal curve. If curve of a frequency distribution is like normal curve then it is known as “mesokurtic”. If the curve is having a high peak than a normal curve then it is known as “leptokurtic”. If the curve is having very low peak than a normal curve then it is known as “platykurtic”.

Measure of kurtosis based on moments is: If $ß2=\mu4/\mu2^2$, then **$\Upsilon 2=ß2-3$.**

If **$\Upsilon 2<0$**, distribution is platykurtic.

If **$\Upsilon 2=0$**, distribution is mesokurtic.

If **$\Upsilon 2>0$**, distribution is leptokurtic.

Testing of hypothesis:

1. Testing population proportion (P) equal to a specified value (Po):
Sometimes observation are taken on qualitative characteristics. In this case number of observations of specific type are counted and its proportion is determined. Here, we develop test in order to test whether proportion of certain type of observations is equal to a specified value.

P=Proportion of certain type of items in population.
N= Sample size.
X=Number of items of certain type in a sample of size n.
P=X/n=Proportion of certain type of item in sample.
Po=Specified value of P.
Ho: P=Po Vs H1: P ǂ Po OR H1: P>Po OR H1: P<Po

Test statistic value under Ho is, **Z= (p-Po)/√PQ/n** follows **N(0,1)** for large n
Decision: We reject Ho at α% l.o.s. if $|Z|>Z_{\alpha/2}$ otherwise accept Ho.
In using these tests, R provides direct command for applying the test.
R commands for null hypothesis **Ho: P=Po**:
a) Consider the alternative hypothesis **H1: P ǂ Po**.

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

b) Consider the alternative hypothesis H1: P>Po.

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

c) Consider the alternative hypothesis H1: P<Po.

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

Decision: We may reject Ho at α% l.o.s. if l.o.s. greater than p-value.

2. Testing equality of two population proportions(P1=P2):
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 know whether difference between two sample proportions is negligible and it has arisen merely due to sampling variations.

Let, P1=proportion of specific items in first population
P2=proportion of specific items in second population

n1 = Size of sample drawn from first sample
n2= Size of sample drawn from second sample
x1= Number of items of specific type in first sample
x2= Number of items of specific type in second sample
p1=x1/n1=proportion of specific item in first sample
p2=x2/n2=proportion of specific item in second sample
The hypothesis for such problems will be:

Ho: P1=P2 versus **H1:P1 ǂ P2**
Under Ho, **$Z=p1-p2/(\sqrt{PQ((1/n1)+(1/n2))} \rightarrow N(0,1)$**
R commands for null hypothesis **Ho: P1=P2**:
(a) consider the alternative hypothesis **H1: P1 ǂ P2.**

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

(b) consider the alternative hypothesis **H1: P1>P2**.

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

(c) consider the alternative hypothesis **H1: P1<P2.**

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

Decision: We may reject Ho at **α**% l.o.s. if l.o.s. greater than p-value.

3. χ^2-Test for independence of attribute:

Suppose that the given data are classified into r levels of attribute A denoted by A1,A1,….,Ar and s levels of attribute B represented by B1,B2,…,Bs. Then different class frequencies can be represented in the following tabular form:

	B1	B2	…..	Bi	…..	Bs	Total
A1	O11	O12	…..	O1j	…..	O1s	(A1)
A2	O21	O22	…..	O2j	…..	O2s	(A2)
…..	…..	……	…..	…..	…..	…..	…..
Ai	Oi1	Oi2	…..	Oij	…..	Ois	(Ai)
…..	…..	…..	…..	…..	…..	…..	…..
Ar	Or1	Or2	…..	Orj	…..	Ors	(As)
Total	(B1)	(B2)	…..	(Bj)	…..	(Bs)	N

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

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

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

$(B_j)=\sum O_{ij}$=Total of observed frequencies in jth column; j=1,2,…,s

Here, Hypothesis under consideration is,

Ho: Two attributes A and B are independent

against H1: Two attributes A and B are not independent.

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

Then test statistic under Ho is,

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

Decision: We reject Ho at α % l.o.s. if $\chi^2_{r-s-1} \geq \chi^2_{r-s-1;\alpha}$.

Otherwise accept Ho.

For (2 x 2) contingency table:

In particular if we have r=2 , s=2 i.e. two attributes A and B are at two levels each, then the contingency table is ,

	B1	B2	TOTAL
A1	a	b	a+b
A2	c	d	c+d
TOTAL	a+c	b+d	TOTAL

Where, N= a+b+c+d

Here, the test statistic used under Ho is ,

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

Decision: We reject Ho at α % l.o.s. if $\chi^2 cal = \chi^2 1,\alpha$. Otherwise accept Ho.

4. χ^2-Test for goodness of fit:

For a given data (frequency distribution) we try to fit some probability distribution and which distribution will fit properly may be a question of interest. Hence we desire to test H0: Fitting of the probability distribution to given data is proper. The test based on chi-square (χ^2) distribution is used. It is called chi-square test of goodness of fit.

In this case we compare the observed and expected frequencies.

Thus we can take

Ho: There is no significant difference between observed and expected frequencies.

$\sum Oi = N = \sum ei$; i=1,2,……,k

p=number of parameters to be estimated for fitting the probability distribution.

The test statistic under Ho is, $\chi^2 = \sum (Oi-ei)^2/ei = \sum (Oi^2/ei)-N$ has χ^2 distribution with (k-p-1) degrees of freedom.

Decision: We reject Ho at α % l.o.s. if χ^2 k-p-1 = χ^2 k-p-1;α. Otherwise accept Ho.

5. Logistic Regression:

In regression analysis there are many situation in which response variable is binary taking only two values say 0 and 1 then we use logistic regression model.

The other way of looking at regression of Y on X is conditional expectation E(Y/X=x). Now here Y is Bernoulli random variable taking the values 0 and 1.

Also E(Y/X=x) = P(Y=1/ X=x) =π(x)

In logistic regression model we assume that,

π (x) = [e^a+bx] /[1+e^a+bx]

where π (x) is called as logistic function in x.

e^a+bx = π (x) / 1- π (x) so that regression model becomes,

ln(π (x) / 1- π (x)) = a+bx

ln(π (x) / 1- π (x)) is called logit transformation and hence the above model is called logistic regression model.

5. Likert Scale:

Various kinds of rating scales have been developed to measure attitudes directly (i.e. the person knows their attitude is being studied). The most widely used is Likert Scale (1932).

In its final form, the Likert Scale is a five (or seven) point scale which is used to allow the individual to express how much they agree or disagree with a particular statement.

For example,

I believe that ecological questions are most important issues facing human beings today.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
(1)	(2)	(3)	(4)	(5)

A Likert Scale assumes that the strength/intensity of an attitude is linear, i.e. on a continuum from strongly agree to strongly disagree, and makes the assumption that attitude can be measured.

For example, each of the five (or seven) responses would have a numerical value which would be used to measure the attitude under investigation.

In addition to measuring statements of agreement, Likert scales can measure other variations such as frequency, quality, importance, and likelihood, etc.

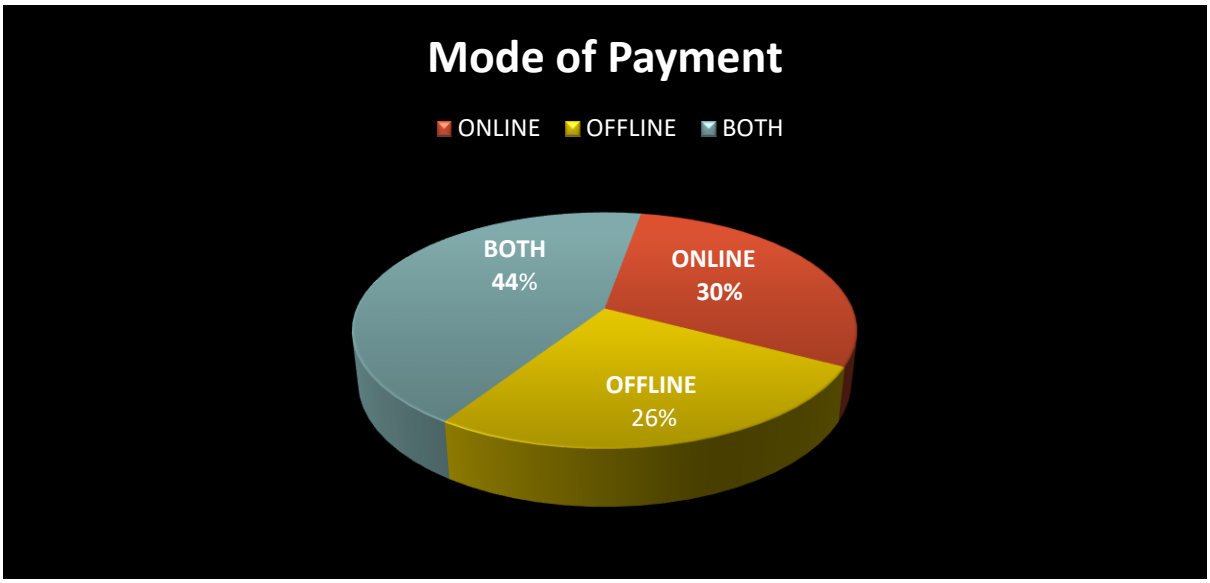
ANALYSIS OF DATA

- **Graphs**

1. Pie Chart:

i. Mode of Payment

MODE OF PAYMENT	No of people
ONLINE	75
OFFLINE	66
BOTH	109



Conclusion:-

From the above diagram, we can observe that most of the people prefer both online & offline method of payment.

ii. Mode of Shopping:

MODE OF SHOPPING	No of people
ONLINE	73
OFFLINE	62
BOTH	115

Conclusion:-

From the above diagram, we can observe that most of the people prefer both online and offline shopping.

iii. Reason for Online Shopping

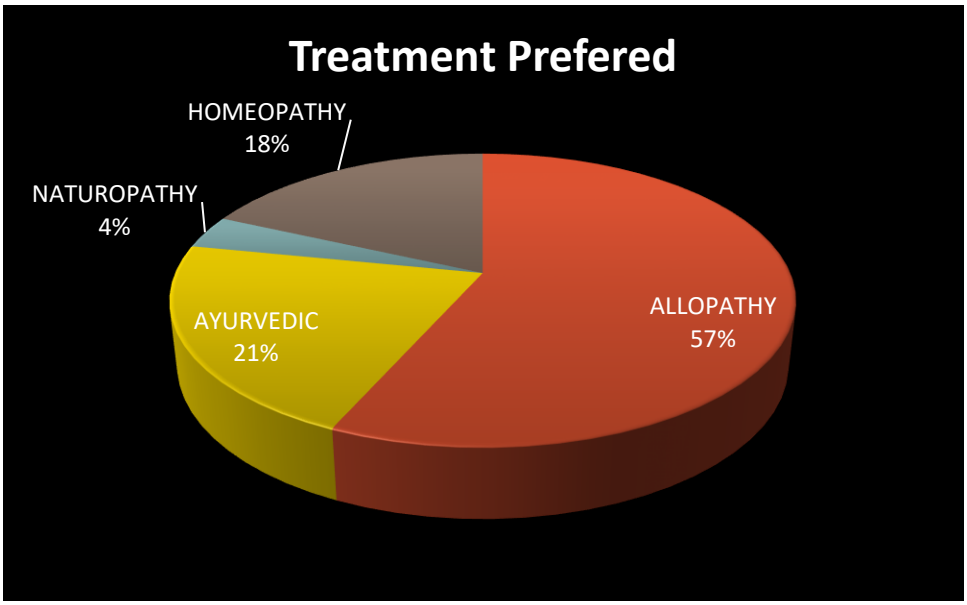
Reason for shopping	No of people
FEAR OF COVID	81
TIME SAVING	106
EASY TO USE	105
SALE & CASHBACK	79

Conclusion:-

We can observe people prefer online shopping, mostly because it is time-saving followed by it is easy to use.

iv. Treatment Prefered

Type of treatment	No of people using treatment
ALLOPATHY	140
AYURVEDIC	53
NATUROPATHY	9
HOMEOPATHY	45



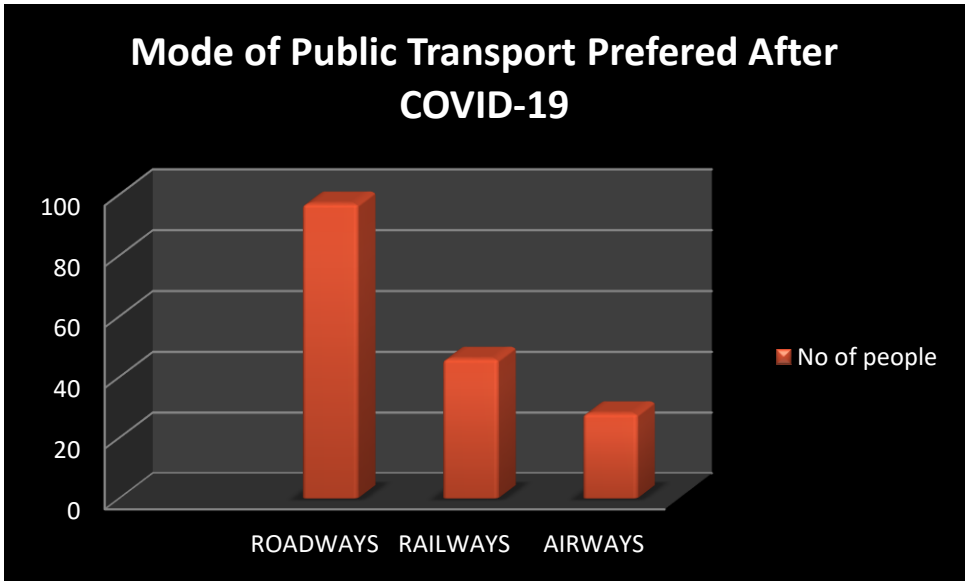
Conclusion:-

From the above diagram, we can observe that most of the people prefer allopathy treatment.

2. Simple Bar Diagram:

i. Mode of Public Transport used after COVID-19

PUBLIC TRANSPORT	No of people
ROADWAYS	97
RAILWAYS	46
AIRWAYS	28

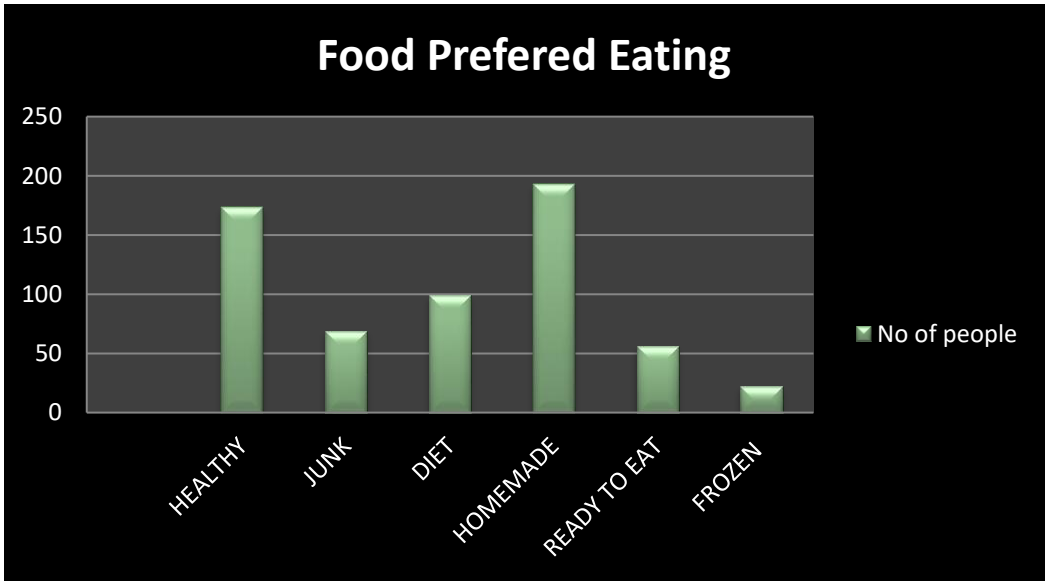


Conclusion:-

From the above graph, we can observe that people preferring public transport mostly use roadways for traveling followed by railways.

ii. Food Prefered Eating

FOOD PREFERED EATING	No of people
HEALTHY	173
JUNK	69
DIET	99
HOMEMADE	193
READY TO EAT	56
FROZEN	22

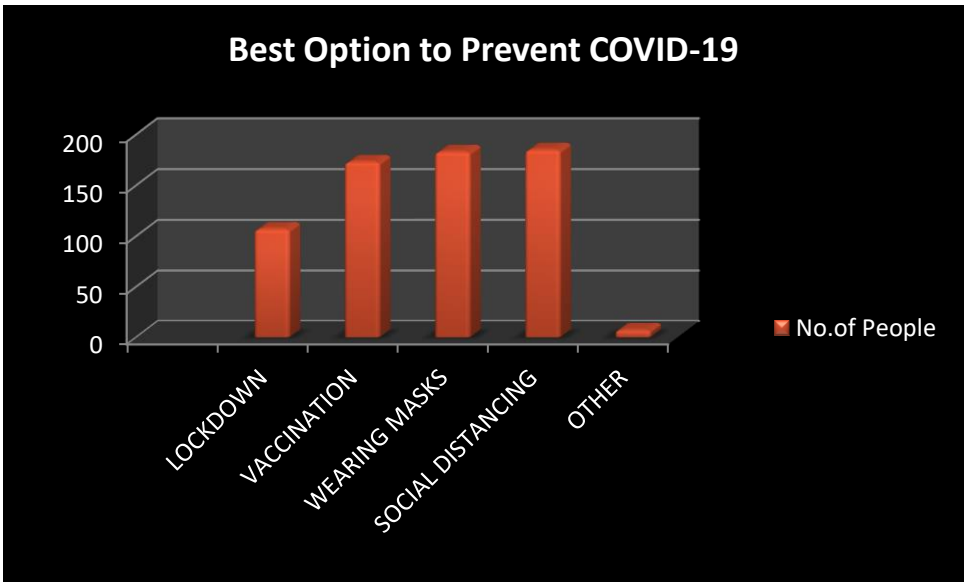


Conclusion:-

From the above bar diagram, we observe that most of the people prefer homemade food followed by healthy food for eating.

iii. Best Option to Prevent COVID-19

BEST OPTION TO PREVENT COVID-19	No.of People
LOCKDOWN	107
VACCINATION	173
WEARING MASKS	183
SOCIAL DISTANCING	185
OTHER	7



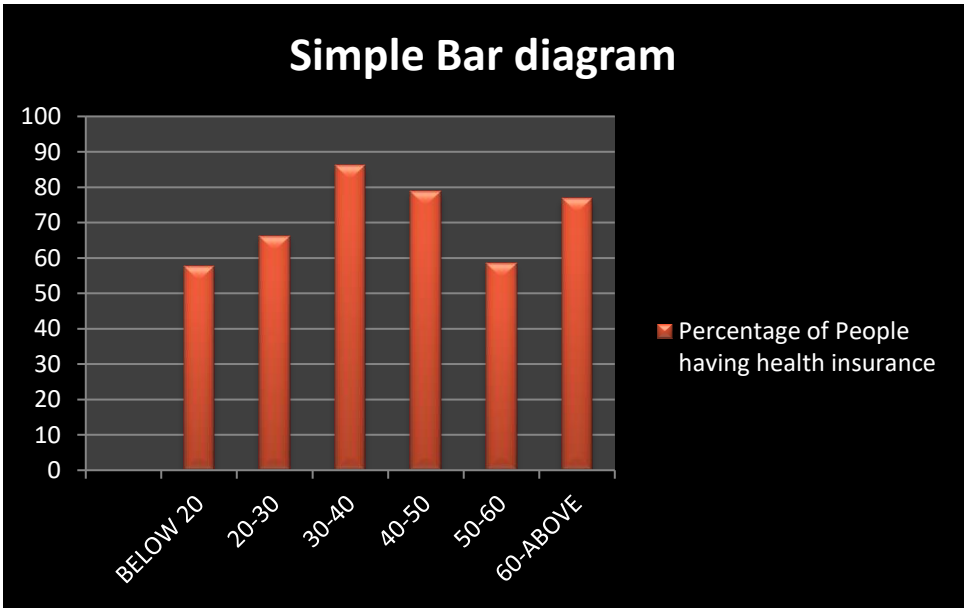
Conclusion:-

According to people's opinion, Social distancing is the best option to prevent COVID-19 followed by wearing masks. Also, we may observe that number of people preferring a lockdown is considerably less.

iv. Percentage of people having health insurance age-wise

AGE	No. of People Having Health Insurance	Total Responses
BELOW 20	30	52
20-30	45	68
30-40	50	58
40-50	37	47
50-60	7	12
60-ABOVE	10	13

AGE	Percentage of People having health insurance
BELOW 20	57.69
20-30	66.18
30-40	86.21
40-50	78.72
50-60	58.34
60-ABOVE	76.92



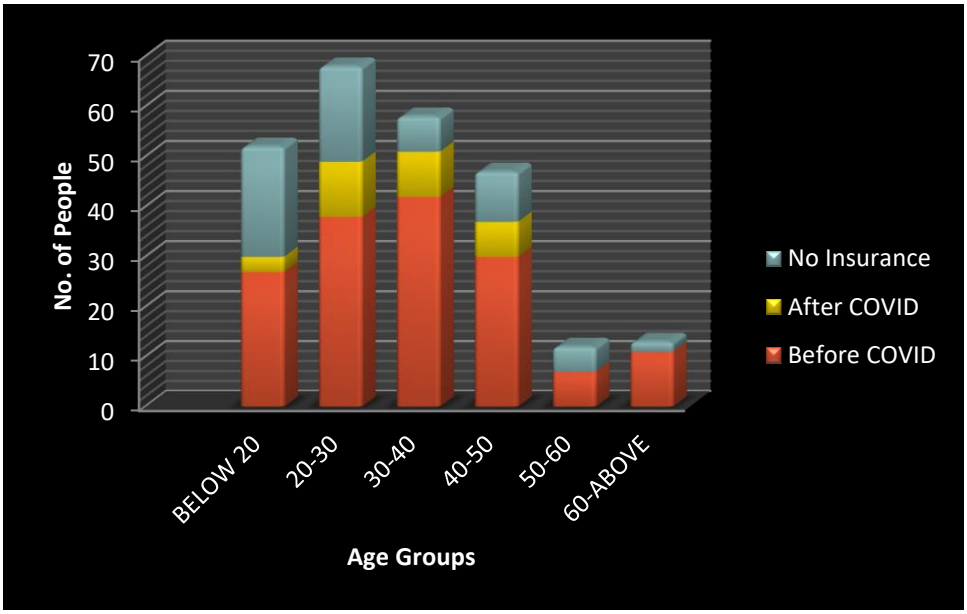
Conclusion:-

From the above graph, we may observe that percentage of people having health insurance is much higher in the age-group 30-40 followed by 40-50.

3. Sub-divided Bar diagram

i. Time of getting health insurance age-wise

AGE	Before COVID	After COVID	No Insurance
BELOW 20	27	3	22
20-30	38	11	19
30-40	42	9	7
40-50	30	7	10
50-60	7	0	5
60-ABOVE	11	0	2



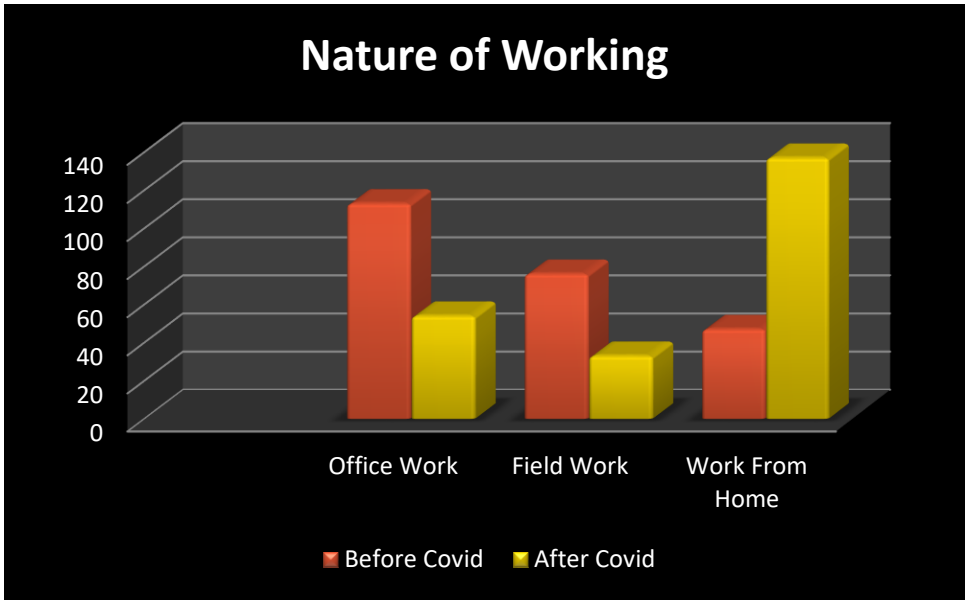
Conclusion:-

From the above graph, most of the people had their health insurance before COVID-19 pandemic. People opting for Health insurance after COVID-19 are mostly of age group 20-30. Also, people who has not opted for health insurance yet are mostly found in the age group below-20.

4. Multiple Bar Diagram:

i. Nature of Working(Before and After COVID-19)

Nature Of Working	Before Covid	After Covid
Office Work	113	54
Field Work	76	33
Work From Home	47	137



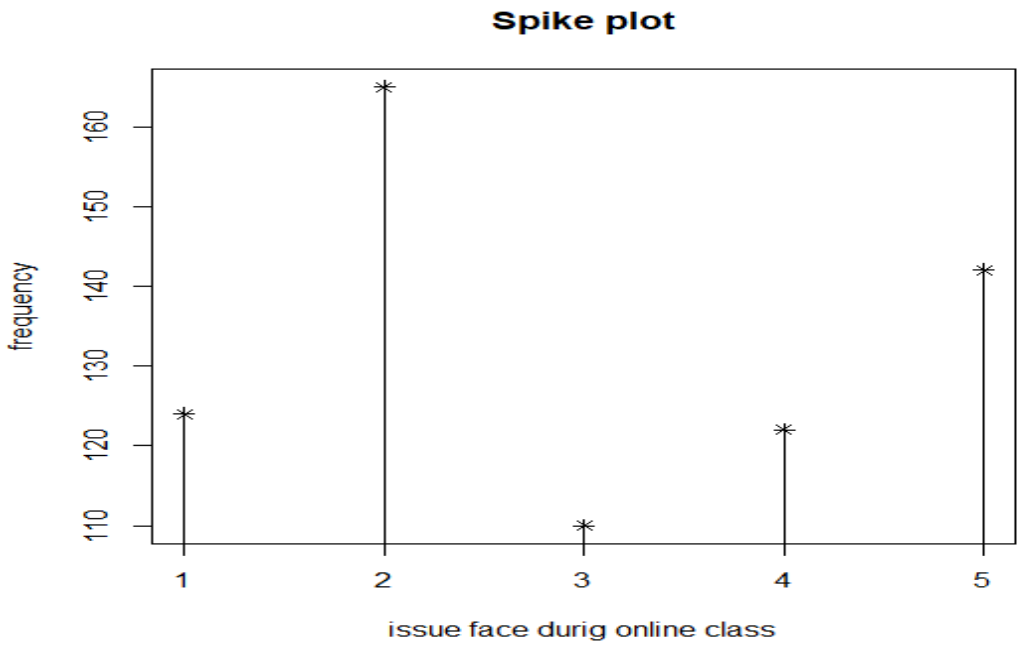
Conclusion:-

From the above graph, we can observe a gradual decrease in no. of people working in office or field before COVID-19. Also , there is increase in no. of people working from home after COVID-19.

5. Rod Plot

i. Issues faced in Online Teaching

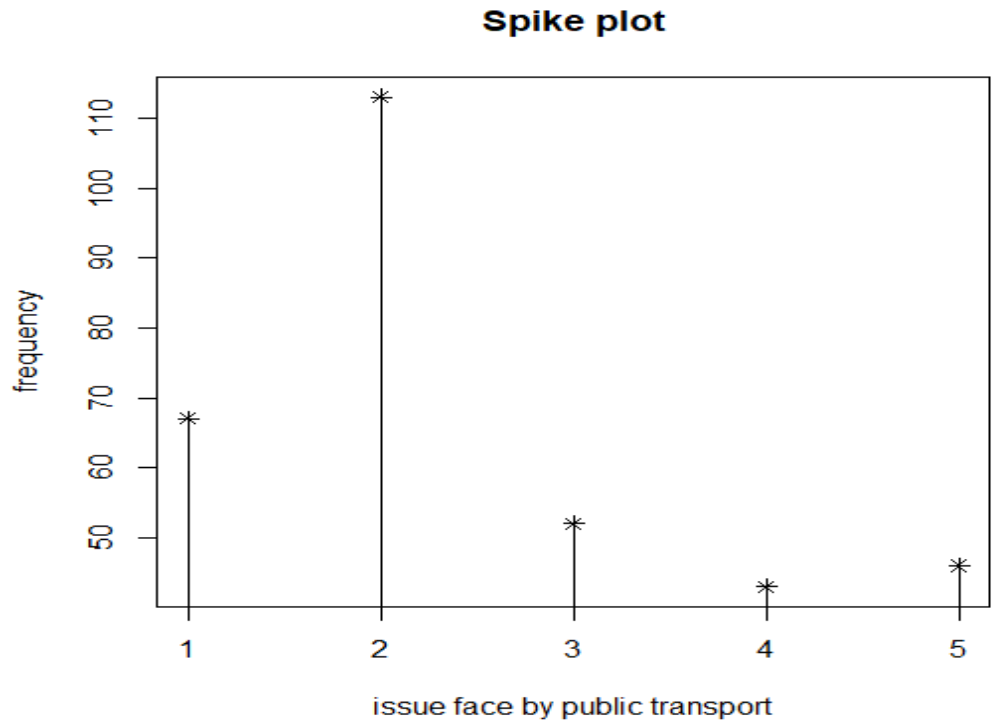
ISSUES FACED ONLINE	Frequency	CODE
COMMUNICATION	124	1
NETWORK	165	2
LACK OF PROBLEM SOLVING	110	3
CLASSROM ENVIRONMENT	122	4
LACK OF FOCUS	142	5



From the above plot, we may observe that the most common issue faced by students in online teaching is network issue. Also a significant no. of people think that online teaching causes lack of focus among students

ii. Issues faced traveling publicly

ISSUE FACED PUBLIC		CODE
RESERVATION	67	1
SOCIAL DISTANCING	113	2
DELAY DUE SECURITY	52	3
STAFF INSUFFICIENT	43	4
TIME FOR TRAVEL	46	5



Conclusion:-

From the above plot, we may observe that most of the people face issues regarding social distancing is faced while traveling followed by reservation issues.

FITTING OF NORMAL

We want to check whether fitting of Normal Distribution is good for the following data or not.

So here age group wise distribution of people having health insurance is tabulated as below,

Age Groups	People having Health Insurance
10-20	30
20-30	45
30-40	50
40-50	37
50-60	7
60-70	10

Fitting of Normal distribution can also be done with the help of R-software.

Following are corresponding commands in R with the output:

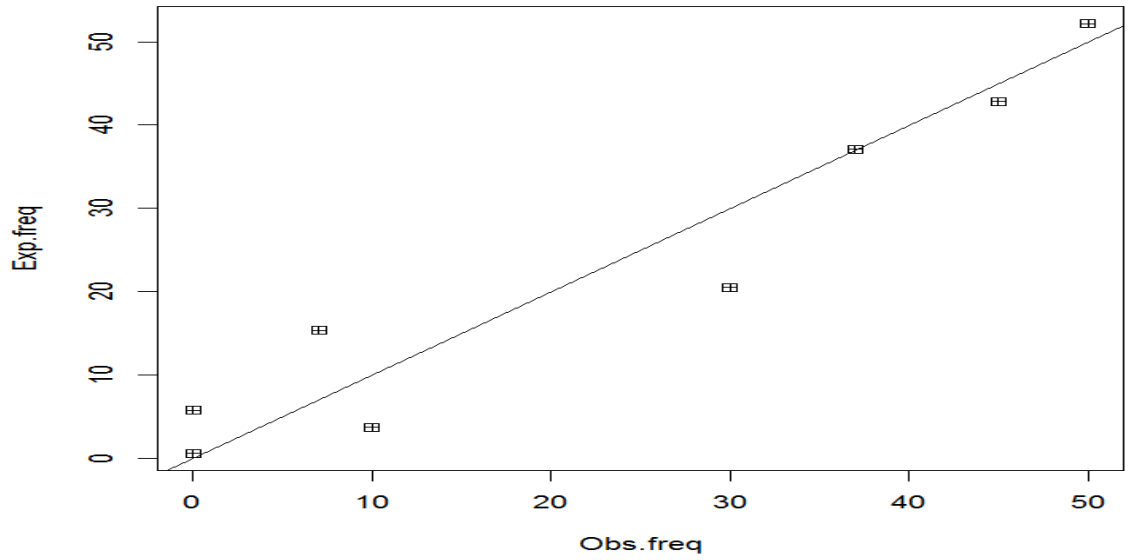
```
lb=c(10,20,30,40,50,60)
> ub=c(20,30,40,50,60,70)
> f=c(30,45,50,37,7,10)
> x=(lb+ub)/2
> x
[1] 15 25 35 45 55 65
> n=sum(f)
> k=length(f)
> n
[1] 179
> k
[1] 6
> mx=sum(f*x)/n
> mx
```

```
[1] 33.65922
> v=sum(f*(x-mx)^2)/n
> sd=v^0.5
> v
[1] 176.9733
> sd
[1] 13.30313
> lb1=c(0,lb,70)
> ub1=c(10,ub,80)
> cp=pnorm(lb1,mx,sd)
> p=round(diff(cp),6)
> p=round(c(p,1-cp[k+2]),6)
> f=c(0,f,0)
> ef=round(n*p,2)
> d=data.frame("LB"=lb1,"UB"=ub1,"Frequency"=f,"Cummulative
Probability"=cp,"Probability"=p,"Expected Frequencies"=ef)
> print(d)
```

SR no.	L.B.	U.B.	Frequency	Cummulative Probability	Probability	Expected Frequency
1	0	10	0	0.0057003	0.031963	5.72
2	10	20	30	0.0376633	0.114602	20.51
3	20	30	45	0.1522650	0.239368	42.85
4	30	40	50	0.3916333	0.291557	52.19
5	40	50	37	0.6831901	0.207150	37.08
6	50	60	7	0.8903405	0.085810	15.36
7	60	70	10	0.9761507	0.020699	3.71
8	70	80	0	0.9968500	0.003150	0.56

```
>plot (f,ef,xlab=”Obs. freq”,ylab=”Exp. freq”,”p”,pch=12)

>abline(0,1)
```



Conclusion:

This plot shows that points do not lie along a straight line. Hence, the assumption of normality is not satisfied.

- **χ^2 test for goodness of fit:**

  Ho: Fitting of Normal distribution is good v/s
  H1: Fitting of Normal distribution is not good
  >f=c(30,45,50,37,7,10)
  > chisq.test(f,p=rep(1/6,6))

  Chi-squared test for given probabilities

  data: f
  X-squared = 53.726, df = 5, p-value = 2.386e-10

  Conclusion:

  Here, l.o.s is greater than p-value. Hence, we may reject Ho at 5% los. That means fitting of normal distribution is not good.

- **Skewness and Kurtoisis:**
  For skewness:

```
> lb=seq(10,60,10)
> ub=seq(20,70,10)
> h=10
> f=c(30,45,50,37,7,10)
> x=(lb+ub)/2
> n=sum(f)
> y=rep(x,f)
> lcf=cumsum(f)
> mx=mean(y)
> mu2=sum(f*(x-mx)^2)/n
> mu3=sum(f*(x-mx)^3)/n
> cmu2=mu2-h^2/12
> b1=mu3^2/cmu2^3
> g1=sqrt(b1)
> g1
[1] 0.525124
```

Conclusion: Here, ϒ1>0, hence distribution is **positively skewed.**

For Kurtoisis:

```
> lb=seq(10,60,10)
> ub=seq(20,70,10)
> h=10
> f=c(30,45,50,37,7,10)
> x=(lb+ub)/2
> n=sum(f)
```

```
> y=rep(x,f)
> mx=mean(y)
> mu2=sum(f*(x-mx)^2)/n
> mu3=sum(f*(x-mx)^3)/n
> mu4=sum(f*(x-mx)^4)/n
> cmu2=mu2-h^2/12
> cmu4=mu4-(h^2/2)*mu2+(7/240)*h^4
> b2=cmu4/(cmu2^2)
> g2=b2-3
> g2
[1] -0.2361043
```

Conclusion: Here, ϒ2<0, hence distribution is **platykurtic.**

PAIRED T TEST

X: Nature of Work of people before COVID-19 pandemic.
Y: Nature of Work of people after COVID-19 pandemic.

To test, Ho: μd=0 v/s
H1: μd≠0.

R output

```
>x=c(113,76,47)
> y=c(54,33,137)
> t.test(x,y,paired=T)

        Paired t-test

data:  x and y
t = 0.084698, df = 2, p-value = 0.9402
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -199.1988  207.1988
sample estimates:
mean of the differences
                     4
```

Conclusion: Hear we observe that p-value is greater than l.o.s=5%. Hence we accept Ho. Also the 95% confidence interval contains value μd=0, so we accept Ho. We conclude that there is no change in the nature of work of people.

PROPORTION TEST (P=Po)

Here we want to test whether the proportion of people using allopathy treatment is greater than 50%.

P=Proportion of people using allopathy in population.
n= Sample size of people using allopathy = 247
X=Number of people using allopathy in a sample of size n = 136
p=X/n=Proportion of people using allopathy in sample
Po=Specified value of P = 50% = 0.05
To test,
Ho: P = 0.5 against
H1:P > 0.5
We use commands in R-Software for testing P=Po as follows,

Output of R:

```
>x=136
>n=247
> prop.test(x,n,alternative="greater")
  1-sample proportions test with continuity correction
data:  x out of n, null probability 0.5
X-squared = 2.332, df = 1, p-value = 0.06337
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.4962547 1.0000000
sample estimates:
      p
0.5506073
```

Conclusion: Here l.o.s. =0.05 < p-value=0.06337, so we may accept Ho. That means we observe that among all the data collected , 50% people prefer allopathy treatment.

PROPORTION TEST (P=Po)

Here we want to test whether the proportion of people working from home is greater than 50%.

P=Proportion of people is working from home
n= Sample size of people working from home=137
X=Number of people working=224
p=X/n=Proportion of people is working from home
Po=Specified value of P=50%=0.05
To test,
Ho: P=0.5 against
H1: P >0.5
We use commands in R-Software for testing P=Po as follows,

Output of R:

```
x=137;n=224
> prop.test(x,n,alternative="greater")
        1-sample proportions test with continuity correction
data:  x out of n, null probability 0.5
X-squared = 10.719, df = 1, p-value = 0.0005303
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5547548 1.0000000
sample estimates:
        p
0.6116071
```

Conclusion: Here l.o.s. =0.05>p-value=0.00053, so we may reject Ho. That means we observe that among all the data collected, people working from home are greater than 50% after COVID-19.

PROPORTION TEST (P1=P2)

Here we want to test whether proportion of people having health insurance before COVID-19 is greater than those having H.I. after COVID-19

H_0=Proportion of People Having Health Insurance Before COVID-19 Pandemic is Same As Proportion Of People Having Insurance After The Pandemic ($P_1=P_2$)

H_1=Proportion of People Having Health Insurance Before COVID-19 Pandemic Is Greater Than Proportion Of People Having Insurance After The Pandemic ($P_1>P_2$)

Here,

n_1=177 X_1=155 Before COVID-19

n_2=40 X_2=40 After COVID-19

By Solving In R Software We Get

```
>prop.test (x=c(155,30), n=c(177,40), alternative="greater")
2-sample test for equality of proportions with continuity correction
data:  c(155, 30) out of c(177, 40)
X-squared = 3.162, df = 1, p-value = 0.03769
alternative hypothesis: greater
95 percent confidence interval:
 -0.009393386  1.000000000
sample estimates:
   prop 1    prop 2
0.8757062 0.7500000
```

Decision: As p-value is less than 0.05, we may accept H1 at 5% los

Conclusion: The proportion of people purchasing Health insurance before COVID-19 pandemic is greater compared to people purchasing Health insurance after COVID-19 pandemic.

PROPORTION TEST (P1=P2)

Here we want to test whether proportion of people using private transport after the pandemic is greater than the proportion of people using public transport.

H_0:-Proportion of People using private transport is same as proportion of people using public transport ($P_1=P_2$)

H_1:- Proportion of People using private transport is greater than proportion of people using public transport ($P_1>P_2$)

Here,

n_1=200 X_1=108 Private Transport

n_2=200 X_2=82 Public Transport

By Solving In R Software We Get

```
>prop.test (x=c(108,82), n=c(200,200), alternative="greater")

2-sample test for equality of proportions with continuity
correction

data:  c(108, 82) out of c(200, 200)

X-squared = 6.2657, df = 1, p-value = 0.006155

alternative hypothesis: greater

95 percent confidence interval:

 0.04355898 1.00000000

sample estimates:

prop 1 prop 2

  0.54   0.41
```

Decision :- As p-value is less than 0.05,We may accept H_1 At 5% los.

Conclusion:- **:-** Proportion of people using private transport is greater than proportion of people using public transport.

PROPORTION TEST (P1=P2)

Here we want to test whether proportion of people preferring offline teaching is greater than proportion of people preferring online teaching.

H_0:-Proportion of people preferring offline teaching is same as proportion of people preferring online teaching (P1=P2)

H_1:- Proportion of people preferring offline teaching is greater than proportion of people preferring online teaching ($P_1>P_2$)

Here,

n_1=250 X_1=170 Offline Teaching

n_2=250 X_2=80 Online Teaching

By Solving In R Software We Get,

```
>prop.test (x=c(170,80), n=c(250,250), alternative="greater")
2-sample test for equality of proportions with continuity correction
data:  c(170, 80) out of c(250, 250)
X-squared = 63.368, df = 1, p-value = 0.0008574
alternative hypothesis: greater
95 percent confidence interval:
 0.2873719 1.0000000
sample estimates:
prop 1 prop 2
  0.68   0.32
```

Decision:- As p-value is less than 0.05,We may accept H_1 At 5% los.

Conclusion:- Proportion of people preferring offline teaching is greater than proportion of people preferring online teaching.

CHI SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

(1) χ^2 Test for Independence of Attributes:

To test whether pre-existing disease is independent of insurance or not.
Attributes to be tested for independence are:
A: Pre-existing Disease
B: Insurance Done Before COVID-19 or After COVID-19

Insurance Done / Pre existing Disease	Insurance Done before Covid-19	Insurance Done after Covid-19	Total
Yes	79	6	85
No	70	18	88
Total	149	24	N = 173

To test,

H0 : Attributes A & B are independent V/s

H1 : Attributes A & B are not independent.

Test Of Independence of Attributes on R software

```
> Pre_existing_disease=matrix(c(79,6,70,18),byrow=T,nrow=2)
> Pre_existing_disease
     [,1] [,2]
[1,]   79    6
[2,]   70   18
> rownames(Pre_existing_disease)=c("Yes","No")
>
colnames(Pre_existing_disease)=c("Insurance_Before_COVID","Insur
ance_After_COVID")
> Pre_existing_disease
    Insurance_Before_COVID Insurance_After_COVID
Yes                     79                    6
No                      70                   18
> model=chisq.test(Pre_existing_disease)
> model

        Pearson's Chi-squared test with Yates' continuity correction

data:  Pre_existing_disease
X-squared = 5.4208, df = 1, p-value = 0.0199

> model$expected
    Insurance_Before_COVID Insurance_After_COVID
Yes               73.20809              11.79191
No                75.79191              12.20809
> model$residuals
    Insurance_Before_COVID Insurance_After_COVID
Yes              0.6769274             -1.686668
No              -0.6652888              1.657669
```

Here, p-value from R is 0.0199 < l.o.s is 0.05

Decision : Here we may reject Ho at 5% l.o.s.

Conclusion : This means that the two attributes A and B are dependent of each other i.e. time of having health insurance depends upon having any pre-existing disease.

(2) χ^2 Test for Independence of Attributes:

To test whether type of treatment people prefer for a particular disease is independent of monthly income or not.

We divided our observation of monthly income (in Rs.) according to same specific classes. We considered the class as below:

Attributes to be tested for independence are:

A: Class according to monthly income of a person

B: Type of Treatment

Treatment preferred / Monthly Income(in lakhs Rs.)	Allopathy	Ayurvedic	Naturopathy	Homeopathy	Total
0-4	17	13	9	6	45
4-8	21	6	36	12	75
8-10	12	5	18	25	60
10-above	20	21	12	7	70
Total	70	45	75	60	250

To test,

Ho: Attribute A & B are independent

H1: Attribute A & B are not independent

Test of Chi Sq Independence of Attribute on R Software

```
>
Disease=matrix(c(17,13,9,6,21,6,36,12,12,5,18,25,20,21,12,17),byro
w=T,nrow=4)
> Disease
    [,1] [,2] [,3] [,4]
[1,]   17   13    9    6
[2,]   21    6   36   12
[3,]   12    5   18   25
[4,]   20   21   12   17
> rownames(Disease)=c(0-4,4-8,8-10,10-20)
>
colnames(Disease)=c("allopathy","ayruvedic","naturopathy","homeo
pathy")
> Disease
   allopathy ayruvedic naturopathy homeopathy
-4        17       13          9         6
-4        21        6         36        12
-2        12        5         18        25
-10       20       21         12        17
> model=chisq.test(Disease)
> model

        Pearson's Chi-squared test

data:  Disease
X-squared = 44.157, df = 9, p-value = 0.0032
```

>
> model$expected
   allopathy ayruvedic naturopathy homeopathy
-4   12.6   8.1   13.5   10.8
-4   21.0   13.5   22.5   18.0
-2   16.8   10.8   18.0   14.4
-10   19.6   12.6   21.0   16.8

Here, P value from R is 0.0032 < l.o.s = 0.05

Decision : Here we may reject Ho at 5% l.o.s

Conclusion :
This means that the two attributes A and B are dependent of each there i.e. monthly income of a person depends on the type of treatment they prefer.

LOGISTIC REGRESSION

A: Age of Person

O: Opinion i.e O=0(No for lockdown) & O=1(Yes for Lockdown)

To test,

Ho: b=0 against H1: b‡0

R commands are as follows:

>a=scan("clipboard") #ages

Read 250 items

>a

[1] 26 33 19 20 20 23 20 21 28 23 43 22 52 42 20 20 23 20 20 26 20
20 21 21 20
 [26] 21 21 21 44 44 28 22 32 20 33 20 21 21 31 22 53 23 20 38 20 19
21 22 21 20
 [51] 20 21 23 21 21 25 20 21 35 25 21 34 24 21 21 21 22 20 21 29 19
54 20 21 20
 [76] 20 20 22 46 20 41 19 20 22 52 44 36 42 49 24 64 21 20 20 20 20
20 20 27 25
[101] 28 25 30 23 27 45 66 28 37 31 45 45 32 44 25 35 35 45 16 47 40
46 48 41 48
[126] 49 44 42 48 41 48 22 28 28 20 25 29 64 53 36 38 40 31 15 31 19
15 33 35 33
[151] 12 36 39 16 18 35 36 39 14 64 16 25 27 20 29 19 29 28 23 44 38
45 32 38 34
[176] 34 30 37 33 42 36 40 45 25 19 13 15 33 33 44 48 35 31 36 43 48
39 38 40 41
[201] 47 38 39 36 50 46 38 32 47 36 39 32 31 38 43 45 35 47 49 46 37
47 47 32 51
[226] 38 64 66 62 63 68 66 66 64 61 46 19 17 26 18 28 29 24 49 51 54
59 57 56 52

```
>o=scan("clipboard")  #opinion
Read 250 items
>d
[1] 1 1 1 0 1 1 1 1 0 1 1 0 0 0 0 0 1 1 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 1 0
1
 [38] 1 1 1 1 0 0 1 0 1 1 1 1 0 0 0 0 0 1 1 0 1 0 0 0 1 0 1 1 0 1 0 0 0 1 0
1 0
 [75] 0 1 1 1 0 0 1 1 0 0 0 0 0 1 1 1 1 0 0 1 0 0 0 0 0 0 1 1 1 0 1 1 0 1 0 0
0 1
[112] 0 1 1 0 1 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 1 1 1 1 1 1 1
1 0
[149] 0 1 1 0 0 1 1 1 0 0 1 0 0 1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1
0 1
[186] 1 1 1 0 0 0 1 0 1 0 0 1 0 0 1 1 1 0 1 1 0 0 0 1 1 1 1 0 1 1 1 1 0 0 1
1 1
[223] 0 1 0 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 1 0 1 1 1
> lfit=glm(d~a,family=binomial)
> summary(lfit)
```

Output of R:

```
Call:
glm(formula = d ~ a, family = binomial)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.286  -1.280   1.074   1.077   1.080

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.2294159  0.3459344   0.663    0.507
a           0.0003547  0.0097140   0.037    0.971

(Dispersion parameter for binomial family taken to be 1)
```

Null deviance: 342.96 on 249 degrees of freedom
Residual deviance: 342.96 on 248 degrees of freedom
AIC: 346.96

Number of Fisher Scoring iterations: 3

Conclusion: The estimates of a and b are 0.2294159 and 0.0003547 respectively. The equation of logistic regression model is = 0.2294159+0.0003547X .

To test, Ho: b=0 against the two sided alternative.
From the results we can see that the

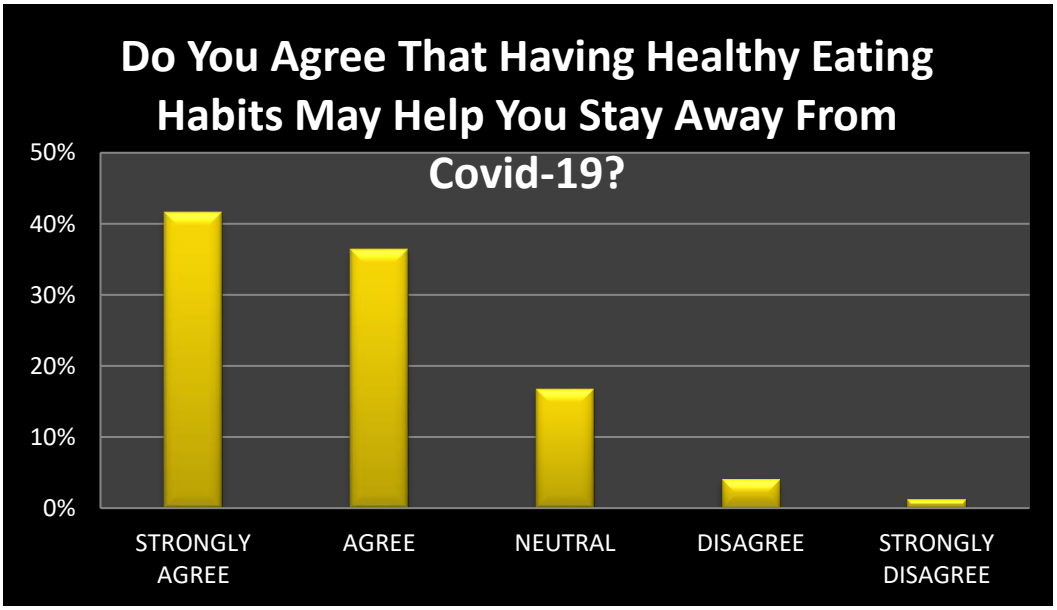
p-value=0.971 is greater that 0.05. Hence, we may accept Ho at 5% l.o.s.

That means we can say that x does not adds significantly to the model i.e. the result of the test is that people's opinion about having lockdown does not depend on age of the person.

LIKERT SCALE

(1) **Having Healthy Eating Habits May Help You Stay Away From Covid-19.**

STRONGLY AGREE	104
AGREE	91
NEUTRAL	42
DISAGREE	10
STRONGLY DISAGREE	3

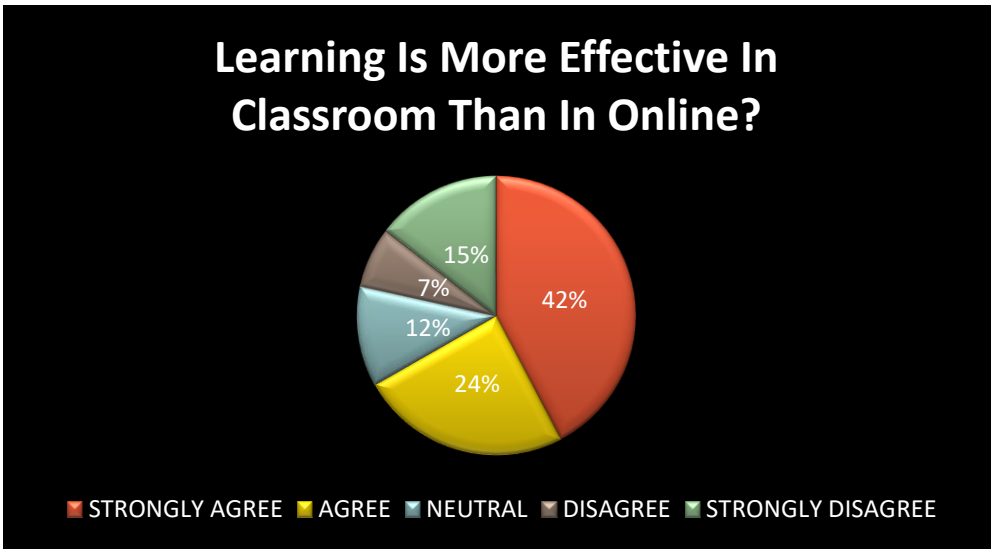


Conclusion:

From the above graph we may observe that above 40% of people strongly agree to the statement that having healthy eating habits may help stay away from COVID-19.

(2)**Learning Is More Effective In Classroom Than In Online.**

STRONGLY AGREE	102
AGREE	59
NEUTRAL	28
DISAGREE	17
STRONGLY DISAGREE	35



Conclusion:

From the above pie chart, we may observe that about 42% people strongly agree to the statement that learning is more effective in classroom than in online.

(3)**COVID-19 Pandemic has affected lives Of people.**

STRONGLY AGREE	128
AGREE	90
NEUTRAL	26
DISAGREE	3
STRONGLY DISAGREE	3

STRONGLY AGREE	51%
AGREE	36%
NEUTRAL	10%
DISAGREE	1%
STRONGLY DISAGREE	1%

Conclusion:

From the above table, we may observe that about 51% people strongly agree that COVID-19 pandemic has affected lives of people.

(4)**Lockdown has really helped preventing spread of COVID-19 pandemic.**

STRONGLY AGREE	65
AGREE	85
NEUTRAL	83
DISAGREE	13
STRONGLY DISAGREE	4

STRONGLY AGREE	26%
AGREE	34%
NEUTRAL	33%
DISAGREE	5%
STRONGLY DISAGREE	2%

Conclusion:

From the above table, we may observe that about 34% of people agree to the above statement wherein 33% share neutral opinion about the same.

CONCLUSION

- **Graphs and Charts:**

Graph Type	Conclusion
Pie Chart	Most of the people prefer both online & offline method of payment.
Pie Chart	Most of the people prefer both online and offline shopping.
Pie Chart	People prefer online shopping, mostly because it is time-saving followed by it is easy to use.
Pie Chart	Most of the people prefer allopathy treatment.
Simple Bar Diagram	People preferring public transport mostly use roadways for traveling followed by railways.
Simple Bar Diagram	Most of the people prefer homemade food followed by healthy food for eating.
Simple Bar Diagram	According to people's opinion, Social distancing is the best option to prevent COVID-19 followed by wearing masks. Also, the number of people preferring a lockdown is considerably less.
Simple Bar Diagram	Percentage of people having health insurance is much higher in the

	age-group 30-40 followed by 40-50.
Subdivided Bar Diagram	Most of the people had their health insurance before COVID-19 pandemic. People opting for Health insurance after COVID-19 are mostly of age group 20-30. Also, people who has not opted for health insurance yet are mostly found in the age group below-20.
Multiple Bar Diagram	There is gradual decrease in no. of people working in office or field before COVID-19. Also, there is increase in no. of people working from home after COVID-19.
Spike or Rod Plot	The most common issue faced by students in online teaching is network issue. Also a significant no. of people think that online teaching causes lack of focus among students
Spike or Rod Plot	Most of the people face issues regarding social distancing is faced while traveling followed by reservation issues.

- **Fitting Of Normal Distribution:**

We tried to fit Normal distribution to age group wise distribution of people having health insurance. But we observe that distribution is not good for the considered data. Hence, we find coefficient of skewness and kurtoisis and observe that distribution is positively skewed and it is platykurtic.

- **Testing of Hypothesis**
  - **Paired T test:**

Test	Hypothesis to be Tested	Conclusion
To test difference in means of nature of work of people before and after COVID-19	Ho: μd=0 v/s H1: μd≠0	There is no significant difference in nature of work of people before and after COVID-19

  - **Tests for Proportions:**

Test	Hypothesis to be tested	Conclusion
To test whether the proportion of people using allopathy treatment is greater than 50%.	Ho: P = 0.5 against H1:P > 0.5	Among all the data collected , 50% people prefer allopathy treatment.
To test whether the proportion of people working from home is greater than 50%.	Ho: P = 0.5 against H1:P > 0.5	Among all the data collected, people working from home are greater than 50% after COVID-19.
To test whether proportion of people having health insurance before COVID-19 is greater than those having H.I. after COVID-19	Ho: P1=P2 against H1:P1 >P2	The proportion of people purchasing Health insurance before COVID-19 pandemic is greater compared to people purchasing Health insurance after COVID-19 pandemic.

To test whether proportion of people preferring offline teaching is greater than proportion of people preferring online teaching.	Ho: P1=P2 against H1:P1 >P2	Proportion of people preferring offline teaching is greater than proportion of people preferring online teaching
To test whether proportion of people using private transport after the people using pandemic is greater than the proportion of public transport	Ho: P1=P2 against H1:P1 >P2	Proportion of people using private transport is greater than proportion of people using public transport.

- **Chi-Square Test for Independence of Attributes:**

  Hypothesis under consideration are:

  Ho: Two attributes are independent v/s

  H1: Two attributes are not independent.

Test	Conclusion
To test whether pre-existing disease is independent of insurance or not.	The two attributes A and B are dependent of each other i.e. time of having health insurance depends upon having any pre-existing disease.
To test whether type of treatment people prefer for a particular disease is independent of monthly income or not.	the two attributes A and B are dependent of each there i.e. monthly income of a person depends on the type of treatment they prefer

- **Logistic Regression**

  From the result of the test we can conclude that people's opinion about having lockdown does not depend on age of the person.

- **Likert Scale:**

Statement	Conclusion
Having Healthy Eating Habits May Help You Stay Away From Covid-19.	Above 40% of people strongly agree to the statement that having healthy eating habits may help stay away from COVID-19.
Learning Is More Effective In Classroom Than In Online	About 42% people strongly agree to the statement that learning is more effective in classroom than in online.
COVID-19 Pandemic has affected lives Of people.	About 51% people strongly agree that COVID-19 pandemic has affected lives of people.
Lockdown has really helped preventing spread of COVID-19 pandemic	About 34% of people agree to the above statement wherein 33% share neutral opinion about the same.

SUGGESTIONS

According to the results that we have obtained by using different statistical tests, we want to suggest these important tips to stay fit and disease free for living a better life. So, these are some suggestions:

1.Exercise regularly. Keeping fit goes hand in hand with exercising. Take at least thirty minutes a day to do some exercises. You can choose to go to the gym or do it at the comfort of your home. The best time to exercise is early in the morning when the body is fresh. Exercising improves the general health of the body.

So we think Health and fitness go hand in hand, for you to be healthy you haveto be fit. It is important to take some time away from your regular schedule to engage in fitness activities. Also, watch what you eat. It is said that you are what you eat.

2. Eat healthy foods: Our diet should be balanced. We should take proper food. Ordinary, food that we take is enough for health, if we take it fresh and in proper quantity. Too much eating in extremely bad for health. The food that we eat is to be taken in regular hours. Food taken hurriedly is not well digested. What we choose to eat plays a very major role in our fitness. Proper food selection leads to

eating healthy. Some of the best foods to eat so as to remain fit are vitamins found in vegetables and fruits, proteins found in whole grains and clean meat. Eat foods that have less oil, since too much oil can increase cholesterol levels which is harmful to the body.

3. Proper rest and sleep: Another hing about health is that we should rest and sleep in time. Early to bed and early to rise is the golden rule of health. A person, who labours too hard may ruins health soon.

4. Keep anxieties away: The essential thing for staying in good health is to free the mind from cares and anxieties as far as possible. A care-worn man cannot enjoy his rest or sleep soundly at night.

5. Participate in yoga. Yoga has got a lot of fitness benefits such as flexibility, burn calories, increase blood flow, boosts immunity, lowers blood sugar and so much more. All these help to keep the body in good shape.

6. **Get a COVID-19 Vaccine Soon:**

7.Say no to junk food: Fast food is rich in trans fat, sugar, spices and artificial preservatives. Constantly consuming junk food will expand your waistline and cause serious health problems in the long run such as high cholesterol, diabetes and heart problems. Since fast food is rich in bad fat, it raises the bad cholesterol in the body.

Hence, switch to a healthier diet and protect yourself from weight gain and other serious wellbeing problems.

8.Avoid alcohol and smoking. These two play a major role in putting our bodies at risk. Smoking caused lung diseases, whereas alcohol can affect the liver. They also affect our sleep, which is important for the body to remain healthy and in good shape.

9.Importance of regular check-ups: Regular check-ups are an important part of prevention routine and can help people to keep their life on perfect track. They also help to find health issues before they become a problem. This will take an important steps toward longer and healthier life.

LIMITATIONS

1. This data or this survey is limited to 250 observations. Therefore, whatever results or conclusions we obtained are only for this sample of 250 respondents. Hence these results are not accepted for worldwide purpose.
2. If the number of observation increase or decrease the above results may be different.
3. Also there is one more limitation that some people do not have any kind of disease.
4. We collected the information for age 10 years and above.
5. The sample size is not too large. As for such surveys, sample should be very large.
6. The results may vary from sample to sample.

SURVEY EXPERIENCE

Collecting the required information for the project is a different experience for all of us. It was a time consuming task to go to different areas in Nashik city.

Sometimes it becomes a difficult task to convince the informant to fill up the questionnaire.

Most of the people ignored us, not even ready to listen for a while. This may happened because many such surveys was faced by them. Or it is just a part of human behaviour. On the other hand, some people found this survey very useful and accordingly answered all the questions very seriously.

Some of the people are also interested in knowing the results of the project and they also encouraged us to perform these kind of projects.

Overall our experience was good and satisfactory.

BIBLIOGRAPHY

- Text Books:

(1) Descriptive Statistics (Nirali Prakashan)

(2) Sampling Distributions and Inference (Nirali Prakashan)

(3) Statistical Methods and Use of R-software (Nirali Prakashan)

(4) Statistical Computing Using R-software (Nirali Prakashan)

- Software:

(1)Microsoft Excel

(2)R-Software

- Websites:

(1)www.encyclopedia.com

(2)www.wikipedia.com

(3)www.google.com

STAY HOME. STAY SAFE. STAY HEALTHY