NDMVP SAMAJ'S
K.T.H.M. COLLEGE, NASHIK
(K.R.T. Arts, B.H. Commerce and A.M. Science College.)

A PROJECT ON

" COMPARATIVE STUDY OF ONLINE AND OFFLINE SHOPPING "

SUBMITTED TO

University of Pune

BY

1) Tamboli Kashish
2) Sonawane Anushka
3) Kolhe Aditi
4) Jain Tanvi
5) Shewale Prachi

Under the guidance of

Miss .M.D.Hiray

(Department of Statistics)

Year 2022-2023

MARATHA VIDYA PRASARAK SAMAJ'S

KRT ARTS, BH COMMERCE & AM SCIENCE (KTHM) COLLEGE,NASHIK

CERTIFICATE

This is to certify that a project report entitled " **Comparative Study Of Online And Offline Shopping** " is benifited work carried out by,

1) Tamboli Kashish
2) Sonawane Anushka
3) Kolhe Aditi
4) Jain Tanvi
5) Shewale Prachi

Students of B.Sc. (statistics) program under my guidance and supervision during the academic year 2022-2023.

Miss. M. D. Hiray	**Examiner**	**Dr. G.S.Phad**
(Project guide)		(Head of department)

DECLARATION BY STUDENT

We declare that the project entitled " **Comaparative Study Of Online And Offline Shopping** ", submitted by us, for the partial fulfillment of our bachelor degree of science in statistics during 2022-2023 in our work.

We further declare that analysis has been carried out based on primary data collected by us.

1) Tamboli Kashish
2) Sonawane Anushka
3) Kolhe Aditi
4) Jain Tanvi
5) Shewale Prachi

Date: / / 2023

Place : Nashik

INTRODUCTION

Online shopping and offline shopping are two different methods of purchasing goods and services. Offline shopping, also known as brick-and-mortar shopping, involves physically visiting a store or a shopping center to buy products or services. On the other hand, online shopping, also known as e-commerce, involves buying products or services over the internet through a website or mobile application.

The comparison of online and offline shopping has been a topic of interest for researchers and business owners for several years. Both online and offline shopping have their advantages and disadvantages, and the choice between them largely depends on the customer's preferences, needs, and circumstances.

Online shopping offers several benefits such as convenience, a wider range of products, and the ability to compare prices and read reviews before making a purchase. It also eliminates the need to physically visit a store, saving time and transportation costs. On the other hand, online shopping also has some drawbacks such as the inability to touch and feel products before buying them, the risk of fraud or data theft, and the delay in receiving the product due to shipping.

Offline shopping, on the other hand, provides the customer with the opportunity to physically inspect and try on products before purchasing them, which can lead to a more satisfying shopping experience. It also allows for immediate possession of the product, eliminating the need to wait for shipping. However, offline shopping can be time-consuming and can limit the range of products available to the customer.

Overall, the comparison of online and offline shopping involves analyzing the advantages and disadvantages of each method and understanding the factors that influence customer preferences. This analysis can help businesses make informed decisions about their sales strategies and provide customers with a better shopping experience.

ACKNOWLEDGEMENT

Presentation inspiration and motivation have always played key role in success of any venture.

We like to thank our Principal, Head of department Dr.G.S.Phad and other statistical department staff of K.T.H.M. college.

We would like to express our special thanks of gratitude to Miss.M .D. Hiray, who gave us golden opportunity to do this project on the topic “Comparative Study OF Online And Offline Shopping” , which also helped us in doing a lot of research .

Last but not the least our parents and friends are also an important inspiration for us. So with due regards, we express our gratitude to them.

INDEX

OBJECTIVE

- To examine the factor influencing the consumer to switch from the offline shopping to online shopping and online shopping to offline shopping.
- The factor influencing the consumer to shop solely online and solely offline.
- To analysis whether the qualification of consumer affect the online shopping and offline purchasing.
- To understand the difference in the consumer behaviour for the online vs offline shopping
- To analyse what consumer are preferring now a days (online shopping or Instore shopping)
- To determine the factors affecting online buying .

METHODOLOGY

The main objective of our project is statistical analysis of "Comparative Study of Online and Offline Shopping". So, for this we decided to collect raw data . For this , a questionnaire was deployed and over 162 responses were valid. Chi square test, Proportion test and graphs were used to understand factors associated with Online and Offline Shopping.

Basics of Diagrammatic Presentation

Concept of Diagrammatic Presentation

- It is a technique of presenting numeric data through pictograms, cartograms, bar diagrams, and pie diagrams. It is the most attractive and appealing way to represent statistical data. Diagrams help in visual comparison and they have a bird's eye view.
- Under pictograms, we use pictures to present data. For example, if we have to show the production of cars, we can draw cars. Suppose the production of cars is 40,000, we can show it by a picture having four cars, where 1 car represents 10,000 units.
- Under cartograms, we make use of maps to show the geographical allocation of certain things.
- Bar diagrams are rectangular and placed on the same base. Their heights represent the magnitude/value of the variable. The width of all the bars and the gaps between the two bars are kept the same.
- Pie diagram is a circle that is subdivided or partitioned to show the proportion of various components of the data.
- Out of the given diagrams, only one-dimensional bar diagrams and pie diagrams are there in our scope.

Types of One-Dimensional Diagram

One-dimensional diagram is a diagram in which only the length of the diagram is considered. It can be drawn in the form of a line or various types of bars.

The following are the types of one-dimensional diagram.

(1) Simple bar diagram

Simple bar diagram consists of a group of rectangular bars of equal width for each class or category of data.

(2) Multiple bar diagram

This diagram is used when we have to make a comparison between two or more variables like income and expenditure, import and export for different years, marks obtained in different subjects in different classes, etc.

(3) Subdivided bar diagram

This diagram is constructed by subdividing the bars in the ratio of various components.

(4) Percentage bar diagram

The subdivided bar diagram presented on a percentage basis is known as the percentage bar diagram.

(5) Broken-scale bar diagram

This diagram is used when the value of one observation is very high as compared to the other.

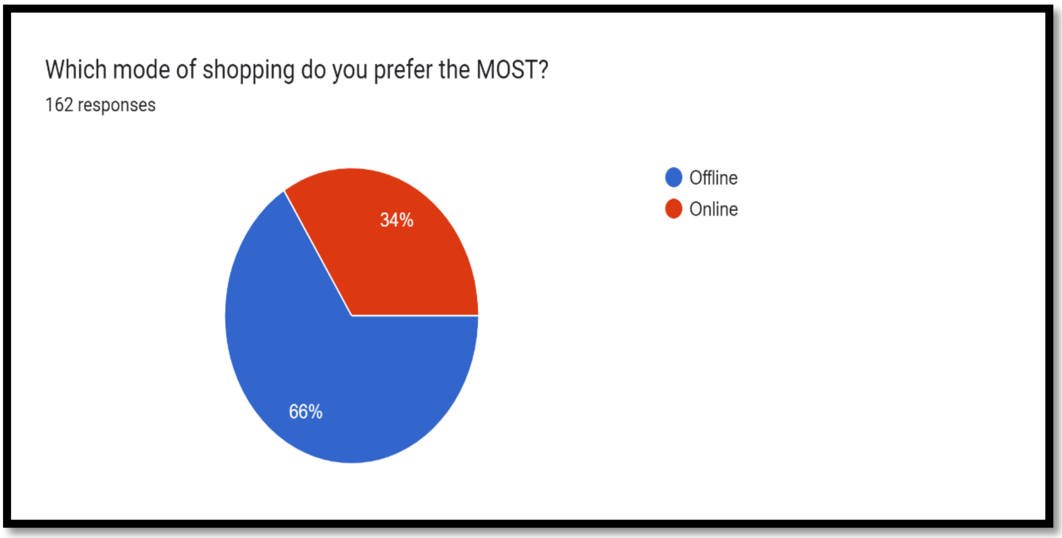
To gain space for the smaller bars of the series, the larger bars may be broken.

The value of each bar is written at the top of the bar.

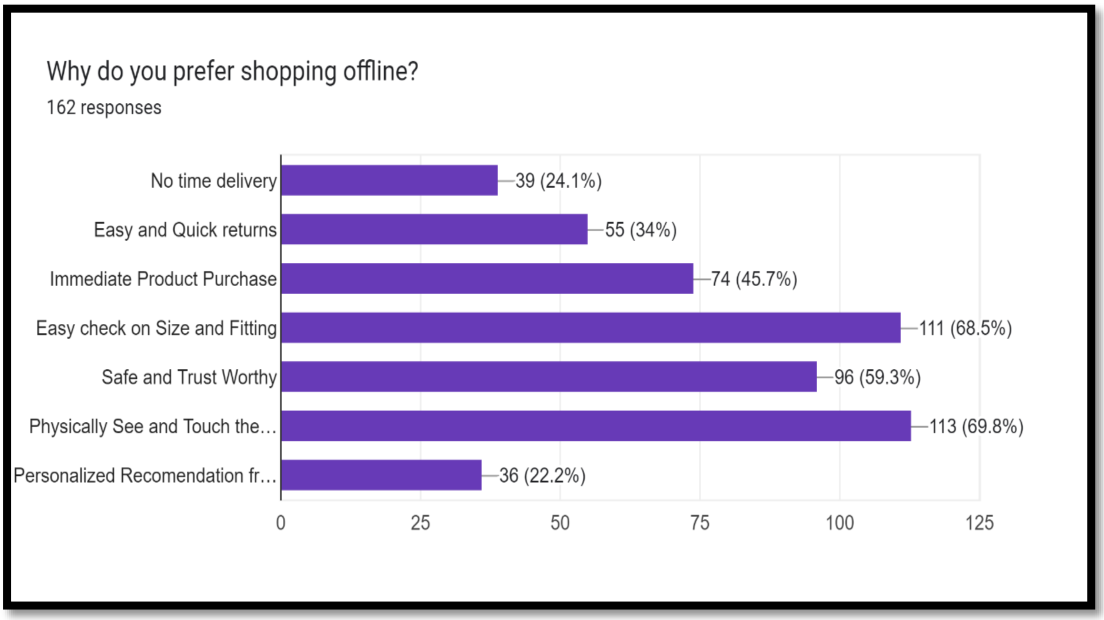
(6) Deviation bar diagram

Deviation bars are used to represent net changes in the data like net profit, net loss, net exports, net imports, etc.

Diagramatic Representation



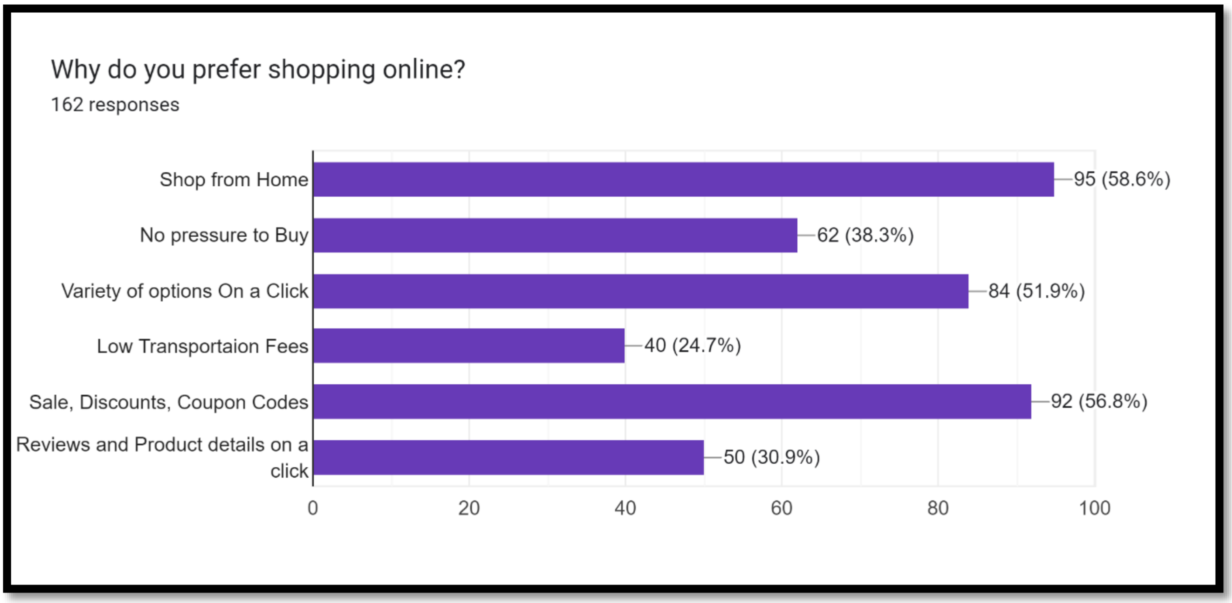
Interpretation : From the above graph, we conclude that offline shopping is preferrable than online shopping.



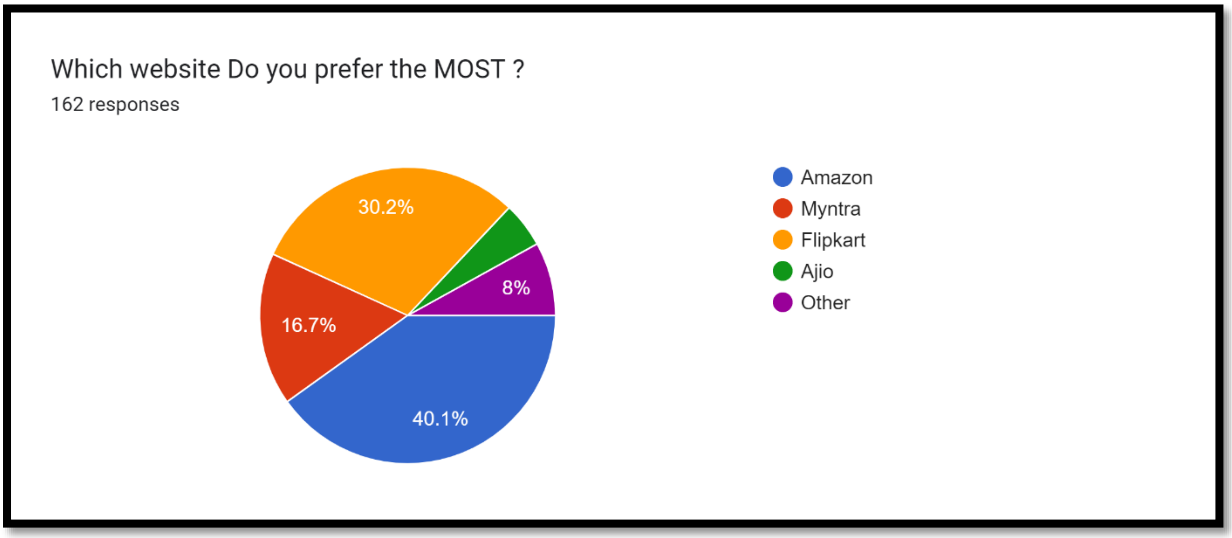
Interpretation: This graph represents, that most people prefer

Shopping offline because they are physically able to touch and see

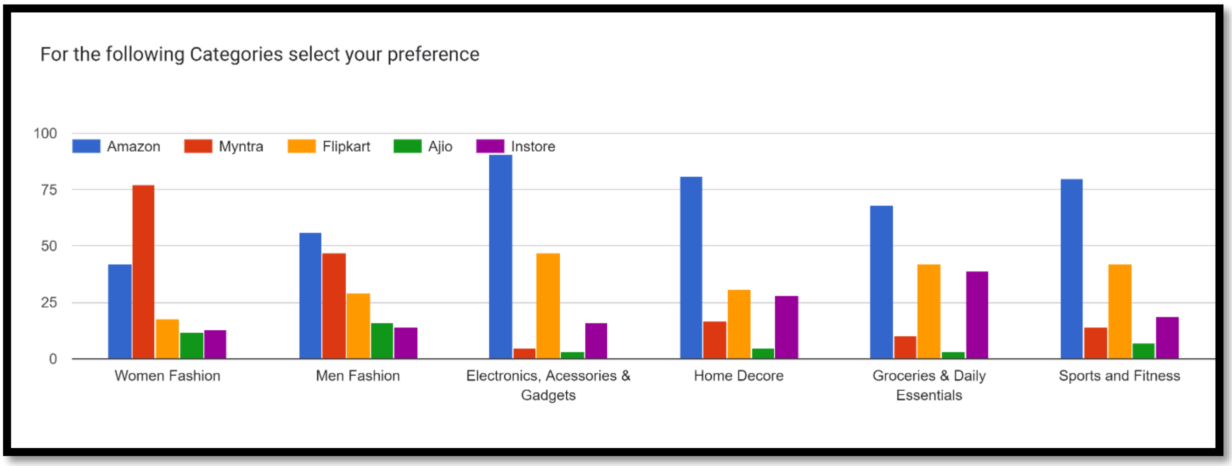
Product.



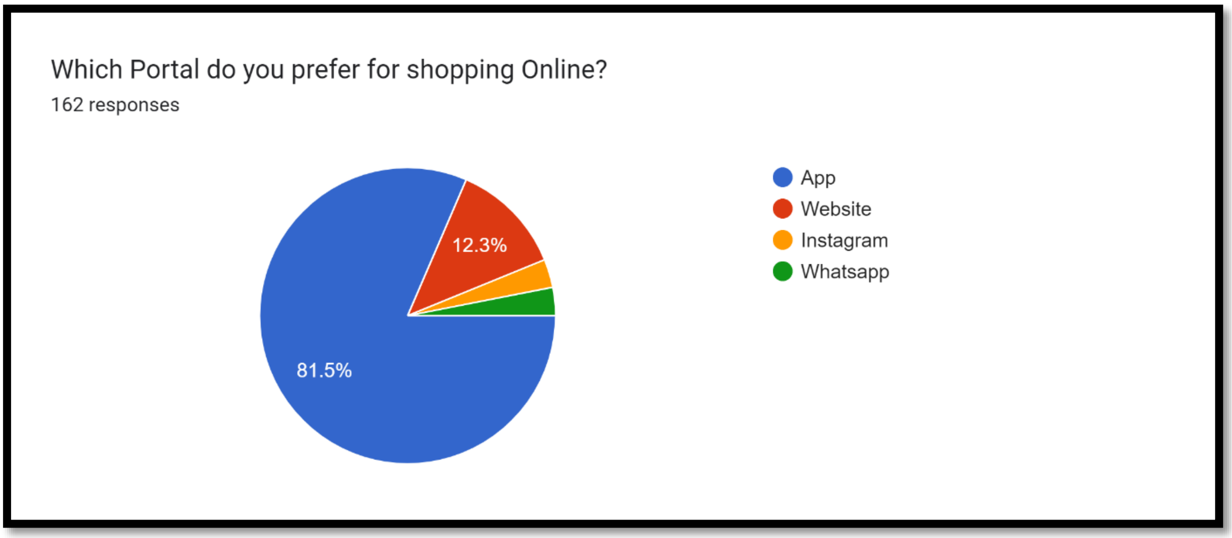
Interpretation : From the graph we observe, that most people prefer shopping online because, they are able to shop from home and also due to sale, discounts and coupon codes.



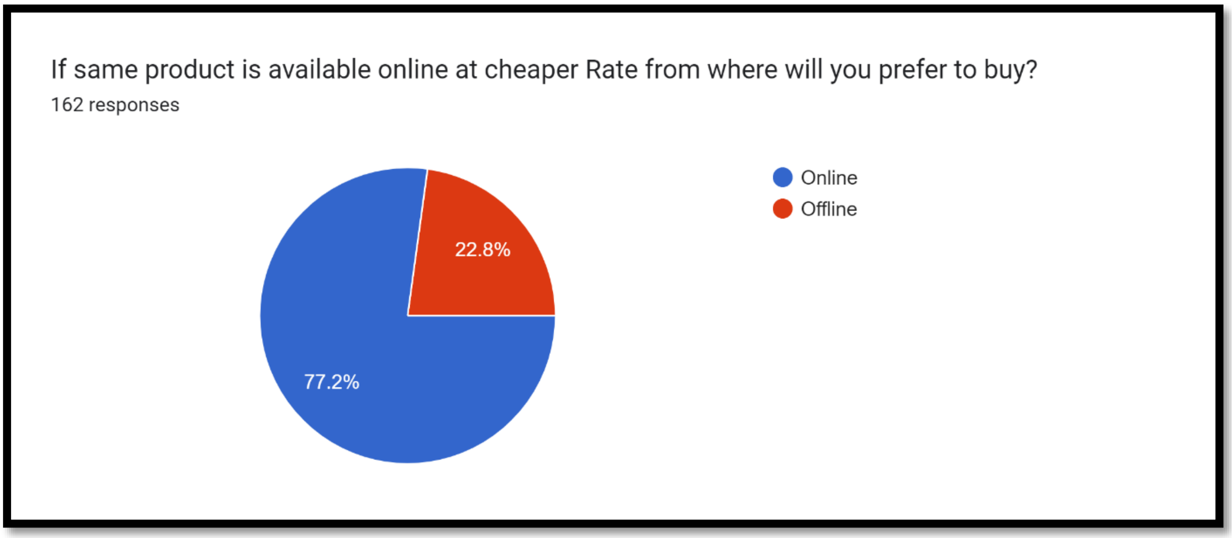
Interpretation : The above graph represents, that most people prefer Amazon for online shopping.



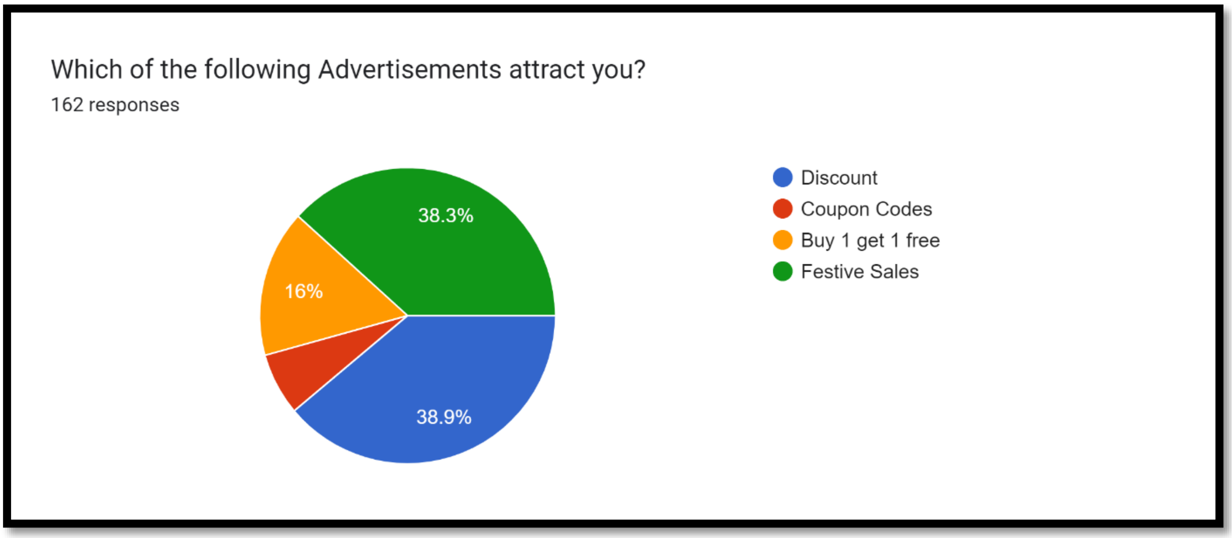
Interpretation : From the above we observe, that the app of online shopping is different for different preferences, such as for women's fashion people use amazon ,etc.



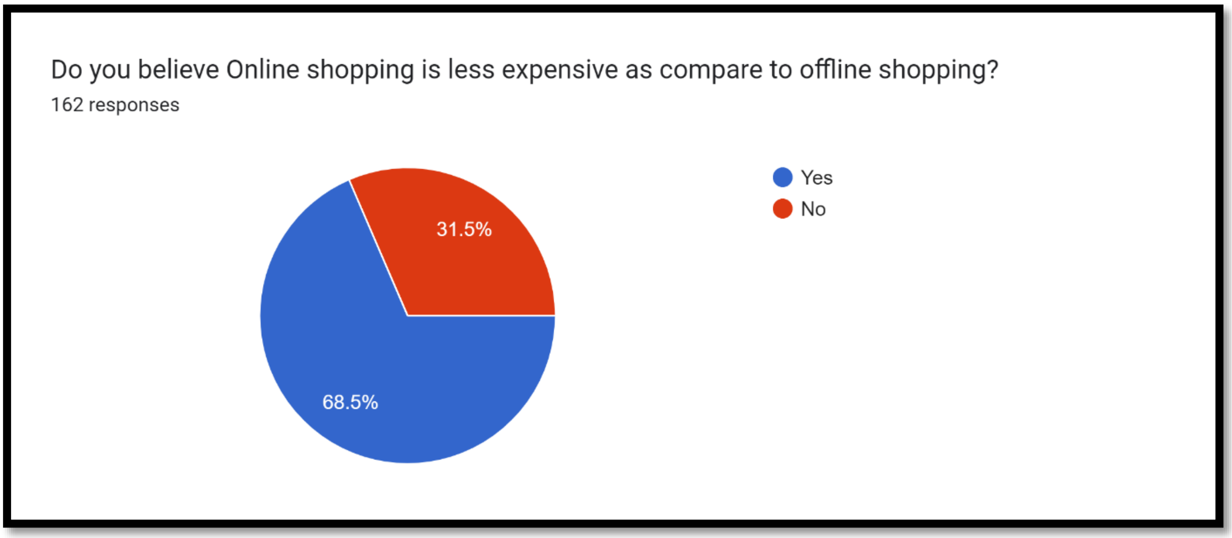
Interpretation: From the graph we conclude, that most of the people prefer app while shopping online.



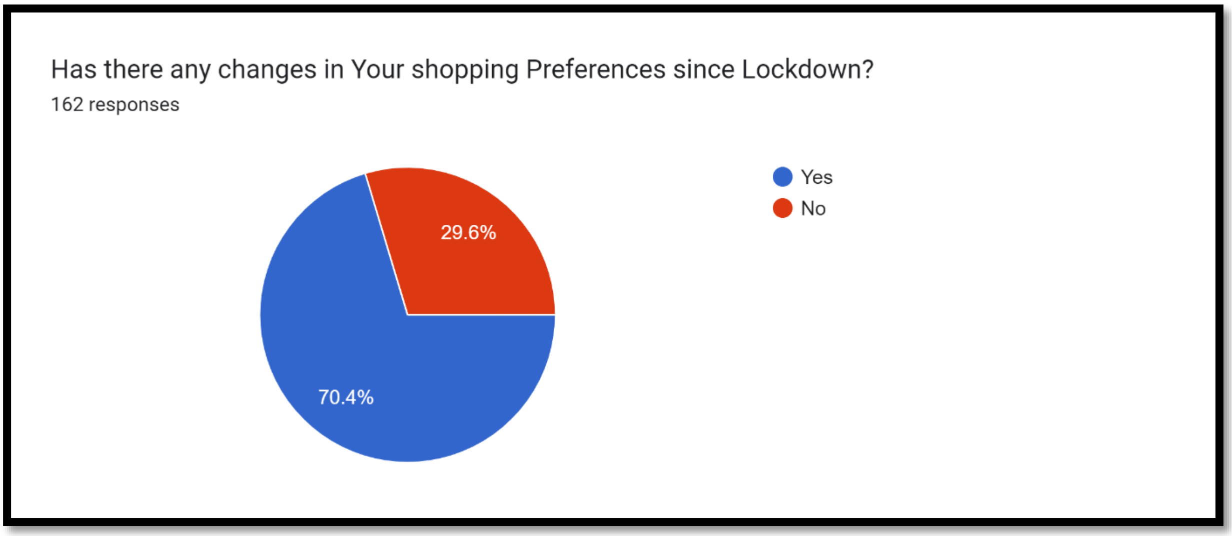
Interpretation: The above graph interprets, that if same product is available online at cheaper rate , then people prefer online over offline shopping.



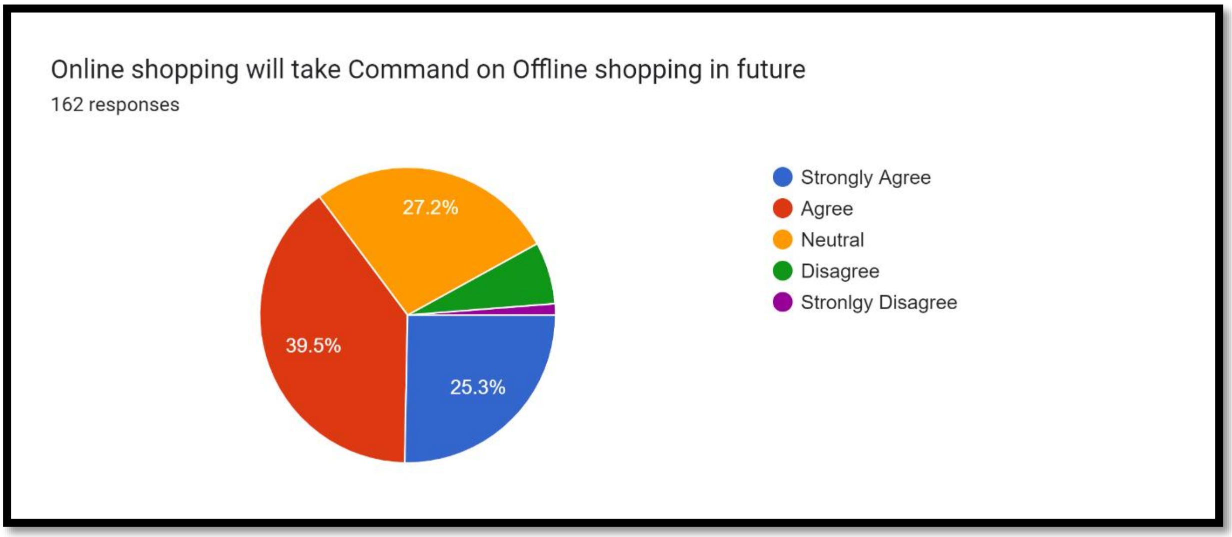
Interpretation: From the graph we conclude, that discount and festive sales attracts most of the people.



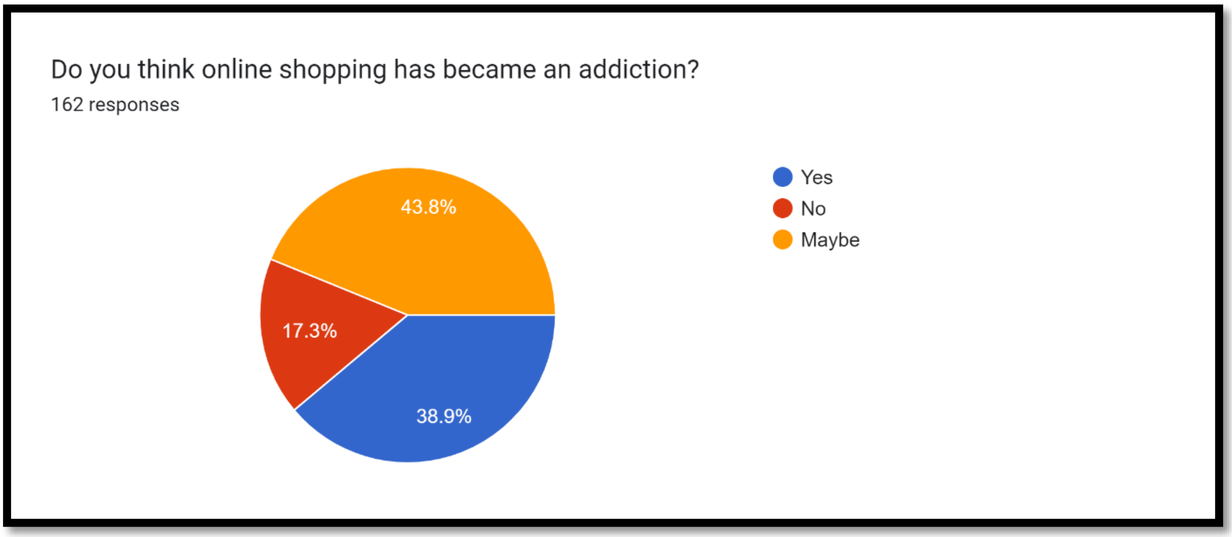
Interpretation: From the graph we conclude, most of the people think online shopping is less expensive as compare to offline shopping.



Interpertation: From the graph ,it is assumed that peoples preferences have changed after lockdown .



Interpretation: From the graph we interpret that, many people thinks that online shopping will take command on offline shopping in future.



Interpretation: From the graph we observe that, 38.9% people thinks that online shopping has become an addiction.

Interpretation: The above graph represents that, 30.2% people prefer local markets for offline shopping.

Data Analysis And Interpretation

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

The process of data analysis

Analysis, refers to dividing a whole into its separate components for individual examination. *Data analysis*, is a process for obtaining raw data, and subsequently converting it into information useful for decision-making by users. *Data*, is collected and analyzed to answer questions, test hypotheses, or disprove theories.

Statistician John Tukey, defined data analysis in 1961, as:

"Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."

Techniques for analyzing quantitative data

- Check raw data for anomalies prior to performing an analysis;
- Re-perform important calculations, such as verifying columns of data that are formula driven;
- Confirm main totals are the sum of subtotals;
- Check relationships between numbers that should be related in a predictable way, such as ratios over time;
- Normalize numbers to make comparisons easier, such as analyzing amounts per person or relative to GDP or as an index value relative to a base year;
- Break problems into component parts by analyzing factors that led to the results, such as DuPont analysis of return on equity.

For the variables under examination, analysts typically obtain descriptive statistics for them, such as the mean (average), median, and standard deviation. They may also analyze the distribution of the key variables to see how the individual values cluster around the mean.

Analysts may use robust statistical measurements to solve certain analytical problems. Hypothesis testing is used when a particular hypothesis about the true state of affairs is made by the analyst and data is gathered to determine whether that state of affairs is true or false. For example, the hypothesis might be that "Unemployment has no effect on inflation", which relates to an economics concept called the Phillips Curve. Hypothesis testing involves considering the likelihood of Type I and type II errors, which relate to whether the data supports accepting or rejecting the hypothesis.

Regression analysis may be used when the analyst is trying to determine the extent to which independent variable X affects dependent variable Y (e.g., "To what extent do changes in the unemployment rate (X) affect the inflation rate (Y)?"). This is an attempt to model or fit an equation line or curve to the data, such that Y is a function of X.

Chi-squared test

A **chi-squared test** (also **chi-square** or **χ^2 test**) is a statistical hypothesis test used in the analysis of contingency tables when the sample sizes are large. In simpler terms, this test is primarily used to examine whether two categorical variables (*two dimensions of the contingency table*) are independent in influencing the test statistic (*values within the table*). The test is valid when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test and variants thereof. 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. For contingency tables with smaller sample sizes, a Fisher's exact test is used instead.

In the standard applications of this test, the observations are classified into mutually exclusive classes. If the null hypothesis that there are no differences between the classes in the population is true, the test statistic computed from the observations follows a χ^2 frequency distribution. The purpose of the test is to evaluate how likely the observed frequencies would be assuming the null hypothesis is true.

Test statistics that follow a χ^2 distribution occur when the observations are independent. There are also χ^2 tests for testing the null hypothesis of independence of a pair of random variables based on observations of the pairs.

Chi-squared tests often refers to tests for which the distribution of the test statistic approaches the χ^2 distribution asymptotically, meaning that the sampling distribution (if the null hypothesis is true) of the test statistic approximates a chi-squared distribution more and more closely as sample sizes increase.

Proportional test

A two-proportion Z-test is a statistical hypothesis test used to determine whether two proportions are different from each other. While performing the test, Z-statistics is computed from two independent samples and the null hypothesis is that the two proportions are equal. In other words, the two samples are coming from the same population. In order to be able to use the two-sample z-test, the following conditions must be met:

- The two populations must be normal or approximately normal
- The two samples must be randomly sampled from the two populations
- The two proportions must be independent

The sample consists of n independent and identically distributed Bernoulli trials, where each trial has a probability p of success.

The null hypothesis is that the true population proportion of successes is equal to a hypothesized value p0.

The alternative hypothesis is that the true population proportion of successes is not equal to p0.

The test statistic is a z-score, which measures the difference between the observed proportion and the hypothesized proportion, taking into account the sample size and the variance of the binomial distribution.

The formula for the test statistic is:

z = (p - p0) / sqrt(p0 * (1 - p0) / n)

where:

p is the observed proportion of successes in the sample.

p0 is the hypothesized proportion of successes under the null hypothesis.

n is the sample size.

The test statistic follows a standard normal distribution under the null hypothesis, so we can calculate the p-value using the standard normal cumulative distribution function (CDF) or a table of standard normal probabilities.If the p-value is less than the significance level alpha, we reject the null hypothesis in favor of the alternative hypothesis. Otherwise, we fail to reject the null hypothesis. The significance level alpha is typically set to 0.05, but it can be adjusted depending on the specific application and desired level of certainty.

Explanatory Data Interpretation

1)

P1=proportion of males

P2=proportion of females

H0:p1=p2

Vs

H1:p1<p2

```
> x1=24
> x2=30
> n1=77
> n2=85
> p1=x1/n1
> p1
[1] 0.3116883
> p2=x2/n2
> p2
[1] 0.3529412
> n=n1+n2
> n
[1] 162
> x=x1+x2
> x
[1] 54
> p=x/n
> p
```

[1] 0.3333333

> q=1-p

> q

[1] 0.6666667

> prop.test(x,n,p=NULL,alternative="greater",correct=TRUE)

1-sample proportions test with continuity correction

data: x out of n, null probability 0.5

X-squared = 17.34, df = 1, p-value = 1

alternative hypothesis: true p is greater than 0.5

95 percent confidence interval:

0.2726941 1.0000000

sample estimates:

p

0.3333333

Decision :We accept H0.

Conclusion: Proportion of males is equal to proportion of females in online shopping.

2)

	Offline	Online
Female	55	30
Male	53	24

H0: Shopping preference is independent of gender.

Vs

H1: Shopping preferences is dependent on gender.

```
data=matrix(c(55,30,53,24),ncol=2,byrow=TRUE)
> data
     [,1] [,2]
[1,]   55   30
[2,]   53   24
> colnames(data)=c("Offline","Online")
> rownames(data)=c("Female","Male")
> data=as.table(data)
> data
       Offline Online
Female      55     30
Male        53     24
> chisq.test(data)

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

data:  data
X-squared = 0.1516, df = 1, p-value = 0.697
```

Decision : we accept H0.

Conclusion : Shopping preferences is independent of gender.

3)

	Urban	Rural	Sub-Urban
Online	35	9	12
Offline	50	37	19

H0 : There is no association between shopping preferences and resedential place.

Vs

H1: There is association between shopping preferences and resedential place.

```
m=cbind(c(9,12,35),c(37,19,50))

> m

     [,1] [,2]

[1,]    9   37

[2,]   12   19

[3,]   35   50

> chisq.test(m,correct=TRUE,simulate.p.value=FALSE)


                Pearson's Chi-squared test


data:  m

X-squared = 6.4539, df = 2, p-value = 0.03968
```

Decision criteria: Reject null hypothesis.

Conclusion: So ,we can say that there is association between shopping preferences and residential places.

4)

	15-30	30-60
Online	37	17
Offline	76	30

H0: There is no association between shopping preferences and age .

Vs

H1:There is association between shopping preference and age.

```
x=cbind(c(37,17),c(76,30))

> x

     [,1] [,2]

[1,]   37   76

[2,]   17   30

> chisq.test(x,correct=TRUE,simulate.p.value=FALSE)


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


data:  x

X-squared = 0.054758, df = 1, p-value = 0.815
```

Decision : We accept H0.

Conclusion: There is no association between shopping preference and age.

Conclusion

- Proportion of males is equal to proportion of females in online shopping.
- Shopping preferences is independent of gender.
- There is association between shopping preferences and residential places.
- There is no association between shopping preference and age.
- 30.2% people prefer local markets for offline shopping.
- 38.9% people thinks that online shopping has become an addiction.
- From the graph we interpret that, many people thinks that online shopping will take command on offline shopping in future.
- From the graph, we conclude that offline shopping is preferrable than online shopping.
- The graph represents, that most people prefer Amazon for online shopping.
- From the graph ,it is assumed that peoples preferences have changed after lockdown .

References

- Tests of Significance and Statistical Methods, Nirali Publication.
- Statistical Computing Using R Software by Vishwas R. Pawgi, Nirali Publication.
- https://www.tutorialspoint.com/diagrammatic-presentation-of-data
- https://en.wikipedia.org/wiki/Chi-squared_test
- https://sixsigmastudyguide.com/two-sample-test-of-proportions