

A SURVEY ON “AN ANALYSIS OF ONLINE SHOPPING BEHAVIOUR AMONG BHU STUDENTS”

PRESESNTED TO
THE DEPARTMENT OF STATISTICS,
INSTITUTE OF SCIENCE,
BANARAS HINDU UNIVERSITY,
VARANASI

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF
BACHELOR’S DEGREE IN STATISTICS

UNDER SUPERVISION OF:

DR. ABHIMANYU SINGH YADAV

ASSISTANT PROFESSOR

DEPARTMENT OF STATISTICS

BANARAS HINDU UNIVERSITY

SUBMITTED BY:

ADITYA NARAYAN SINGH

B.A.(HONS.)

FACULTY OF SOCIAL SCIENCES

ENROLLMENT NO - 439659

EXAM ROLL NO - 21221STA013

Certificate

This is to certify that the project report “An Analysis of Online Shopping Behaviour Among BHU Students” submitted to the department of statistics, Banaras Hindu University, in partial fulfillment for the award of Bachelor Degree in Statistics is a record of bona fide work carried out by Mr. Aditya Narayan Singh (Roll No-21221STA013), under my supervision and guidance.

Dr Abhimanyu Singh Yadav

Assistant Professor

Department of Statistics

Banaras Hindu University

Acknowledgement

I like to share our sincere gratitude to all those who helped me in completion of this project. During the work I faced many challenges due to my lack of knowledge and experience but these people helped me to get over from all the difficulties and in final compilation of our ideas to shape sculpture.

I would like to show my gratitude to my project supervisor Dr. Abhimanyu Singh Yadav Sir for his continuous governance and guidance, because of which I was able to learn the minute aspects of a project work.

I wish to place on record our sincere thanks to Prof. G.P. Singh, Head of Department, Department of Statistics, Institute of Science, BHU for extending the facilities existing in the department during the work.

I am thankful to Dr. Vikas Sharma, Dr. Arun Kaushik and all the faculties and staff of Department of Statistics, BHU for their help and support during our Under-graduation.

I am thankful to my batch mates and most of all to my family members who were always there for me.

Table of Contents

Introduction	5
Need For Study	7
Problem Statement.....	8
Objectives	9
Methodology	11
• Sampling Theory for getting response from BHU population:.....	11
• Statistical Tools.....	13
1. The Chi-Square test:.....	13
2. Frequency table:	14
3. Pie charts:.....	14
• Study Area: Banaras Hindu University.....	16
• Data Collection	18
• Descriptive Data Analysis.....	20
• Chi-Square Data Analysis	27
Conclusion	36
Field Experience and Difficulties	39
Suggestions.....	40
References and Bibliography.....	41
Appendix:.....	42
• Questionnaire	42

Introduction

In the digital era, the landscape of commerce has undergone a significant transformation with the advent of online shopping platforms. These platforms have revolutionized the way consumers browse, purchase, and interact with products and services. Understanding the dynamics of online shopping behavior and the factors influencing consumer choices is crucial for businesses to tailor their strategies and enhance the overall shopping experience.

This project aims to conduct a comprehensive statistical study of online shopping platforms, focusing on the relationship between online shopping frequency and various demographic factors such as age, gender, residence, education level, and income level. By analyzing data collected through a structured questionnaire, this study seeks to uncover patterns, preferences, and trends among online shoppers, providing valuable insights for both businesses and researchers in the e-commerce domain.

The proliferation of e-commerce platforms has led to a plethora of choices for consumers, ranging from global giants like Amazon and Flipkart to niche platforms catering to specific product categories. Understanding the usage patterns and preferences for these platforms among different demographic groups can help businesses tailor their marketing strategies, product offerings, and user experiences to better serve their target audience.

Furthermore, investigating factors influencing consumer satisfaction and loyalty towards online shopping platforms is essential for fostering brand advocacy and long-term customer relationships. By examining the impact of factors such as pricing, discounts, customer service, and website usability on overall satisfaction and likelihood to recommend, businesses can identify areas for improvement and differentiate themselves in a competitive market.

Through the application of statistical techniques such as chi-square tests, Fisher's exact test, this project aims to uncover meaningful insights into the complex interplay between demographic variables and online shopping behavior. By leveraging the power of data analytics, businesses can make

informed decisions to optimize their online platforms and enhance the overall shopping experience for consumers.

Overall, this project endeavors to contribute to the growing body of knowledge in the field of e-commerce analytics, shedding light on the intricacies of consumer behavior in the digital marketplace and providing actionable insights for businesses to thrive in an increasingly digitized world.

Need For Study

- Understanding the demographics, preferences, and behaviors of potential markets helps online shopping platforms identify target audiences and tailor their offerings to meet specific consumer needs. By analyzing real-life data, businesses can gain insights into untapped market segments and opportunities for expansion.
- Real-life data analysis allows online shopping platforms to segment their target markets based on factors such as age, income level, geographic location, and purchasing behavior.
- Real-life data analysis provides online shopping platforms with actionable insights to inform strategic decision-making. By leveraging data-driven decision-making processes, platforms can make informed choices regarding market entry strategies, resource allocation, pricing strategies, and strategic partnerships to achieve their business objectives and drive sustainable growth.
- Learning to survey real life data and how to handle it. Also perform some data analysis tools to infer conclusions.

Problem Statement

The rapid growth of e-commerce has transformed the retail landscape, offering consumers unprecedented convenience and choice in shopping for goods and services. However, amidst the abundance of online shopping platforms, businesses face the challenge of understanding the complex interplay between demographic factors and consumer behavior to effectively target and engage their audience. This project seeks to address this challenge by conducting a statistical study of online shopping platforms, aiming to identify the relationship between demographic variables (such as age, gender, residence, education level, and income level) and online shopping frequency. By analyzing data collected through a structured questionnaire, this study aims to uncover patterns, preferences, and trends among online shoppers, providing valuable insights for businesses to optimize their marketing strategies, product offerings, and user experiences. Through the application of statistical techniques and data analysis methods, this project endeavors to contribute to the understanding of consumer behavior in the digital marketplace and offer actionable recommendations for businesses to enhance their online shopping platforms and foster long-term customer relationships.

Objectives

- **To Investigate Demographic Trends:** Examine the demographic characteristics (such as age, gender, residence, education level, and income level) of online shoppers to identify trends and patterns in online shopping behavior.
- **To Analyze Online Shopping Frequency:** Determine the frequency of online shopping among different demographic groups to understand variations in consumer engagement with e-commerce platforms.
- **To Explore Platform Preference:** Investigate the preferred online shopping platforms among different demographic segments to discern platform popularity and usage trends.
- **To Assess Satisfaction Levels:** Evaluate customer satisfaction with the most recent online shopping experience and analyze the factors influencing satisfaction across demographic groups.
- **To Determine Likelihood to Recommend:** Measure the likelihood of consumers recommending their preferred online shopping platform to friends or family members and identify demographic factors influencing this likelihood.
- **To Identify Factors Influencing Shopping Decisions:** Investigate the factors influencing consumers' decisions to shop on specific online platforms, including pricing, discounts, customer service, and website usability.
- **To Examine Product Satisfaction:** Assess customer satisfaction with the products received from online purchases and explore any variations across demographic groups.
- **To Provide Actionable Insights:** Synthesize the findings of the statistical analysis to offer actionable recommendations for businesses to optimize their online shopping platforms and enhance the overall customer experience.
- **To Contribute to Knowledge in E-commerce Analytics:** Contribute to the existing body of knowledge in the field of e-commerce analytics by providing insights into the relationship between demographic variables and online shopping behavior.

- **To Enhance Decision-making for Businesses:** Empower businesses operating in the e-commerce domain with data-driven insights to make informed decisions regarding marketing strategies, product offerings, and user experiences.

Methodology

For the completion of the project, I have developed a google form, which contained some questions related to online shopping experiences. Thus, i have used:

Sampling Theory for getting response from BHU population:

Sampling is a process of selecting a subset of individuals or items from a larger population for the purpose of making statistical inferences about the population. Here are the general steps involved in any sampling technique:

1. **Define the population:** The first step in sampling is to define the population that you are interested in studying. This population could be people, animals, plants, or any other type of entity.
2. **Determine the sampling frame:** Once you have defined the population, you need to determine the sampling frame, which is the list of individuals or items that you will use to select your sample. The sampling frame should be representative of the population and should include all individuals or items in the population.
3. **Choose the sampling method:** There are many different sampling methods, including simple random sampling, stratified sampling, cluster sampling, and systematic sampling. Each method has its advantages and disadvantages, and the choice of sampling method depends on the specific research question and the characteristics of the population.
4. **Determine the sample size:** Once you have chosen a sampling method, you need to determine the sample size, which is the number of individuals or items that you will select from the population. The sample size should be large enough to provide a representative sample of the population but small enough to be manageable.
5. **Select the sample:** Once you have determined the sample size, you need to select the sample using the chosen sampling method. For example, if you are using simple random sampling, you would randomly select individuals or items from the sampling frame until you reach the desired sample size.

6. **Data Validation and Cleaning:** Conduct a thorough review of the collected data to identify any inconsistencies, missing values, or outliers. Implement validation checks to ensure data integrity and accuracy during the data entry process. Use data cleaning techniques, such as outlier detection, imputation of missing values, and consistency checks, to rectify errors and inconsistencies in the dataset.
7. **Analyze the data:** Once you have collected data from the sample, you can use statistical methods to analyze the data and make inferences about the population. The validity of these inferences depends on the representativeness of the sample and the quality of the data collection process.

Statistical Tools

The Chi-Square test:

The chi-square test is a statistical hypothesis test used to determine whether there is a significant association between two categorical variables. It is used to compare the observed frequencies of each category with the expected frequencies under the null hypothesis. The Chi-square test assumes that the data are independent and that the sample size is large enough to apply the central limit theorem. The test is commonly used in medical research, social sciences, and quality control.

To perform a chi-square test, you need to follow these steps:

- i. State the null and alternative hypothesis. The null hypothesis states that there is no significant association between the two variables, while the alternative hypothesis states that there is a significant association.
- ii. Determine the significance level (alpha) for the test. This is typically set to 0.05.
- iii. Create a contingency table, which displays the observed frequencies for each category of the two variables.
- iv. Calculate the expected frequencies for each category under the null hypothesis. The expected frequency for each category is equal to the product of the row total and column total divided by the total sample size.
- v. Calculate the chi-square statistic, which measures the difference between the observed frequencies and the expected frequencies, weighted by the expected frequencies. The formula for the chi-square statistic is:

$$\chi^2 = \sum (\text{Observed} - \text{Expected})^2 / \text{Expected}$$

- vi. Determine the degrees of freedom, which is equal to the product of the number of categories minus 1 for each variable.
- vii. Determine the critical value of the chi-square distribution, based on the degrees of freedom and the significance level.
- viii. Compare the test statistic to the critical value. If the test statistic is greater than the critical value, then you can reject the null hypothesis and conclude that there is a significant association between the two

variables. If the test statistic is less than the critical value, then you fail to reject the null hypothesis and cannot conclude that there is a significant association.

- ix. Calculate the p-value, which is the probability of obtaining a test statistic as extreme as the one observed, assuming that the null hypothesis is true. If the p-value is less than the significance level, then you can reject the null hypothesis and conclude that there is a significant association between the two variables.

Frequency table:

Frequency refers to the number of times an event occurs or a value occurs. A frequency table is a table that lists items and shows the number of times the item occurs.

Pie charts:

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size of the data, and it is a type of pictorial representation of data. A pie chart requires a list of categorical variables and numerical variables. Here, the term “pie” represents the whole, and the “slices” represent the parts of the whole. The “pie chart” is also known as a “circle chart” dividing the circular statistical graphic into sectors or sections to illustrate the numerical problems. Each sector denotes a proportionate part of the whole. To find out the composition of something, Pie-chart works the best at that time. In most cases, pie charts replace other graphs like the bar graph, line plots, histograms, etc.

The pie chart is an important type of data representation. It contains different segments and sectors in which each segment and sector of a pie chart forms a specific portion of the total(percentage). The sum of all the data is equal to 360° . The total value of the pie is always 100%. To work out with the percentage for a pie chart, follow the steps given below:

- i. Categorize the data
- ii. Calculate the total
- iii. Divide the categories
- iv. Convert into percentages
- v. Finally, calculate the degrees Therefore, the pie chart formula is given as

$$\text{(Given Data/Total value of Data)} \times 360^\circ$$

Study Area: Banaras Hindu University

Location:

Banaras Hindu University (BHU) is located in Varanasi, Uttar Pradesh, India, one of the oldest continuously inhabited cities in the world.

The university campus spans over 1,300 acres along the southern banks of the river Ganges, providing a picturesque setting for academic and extracurricular activities.

Historical Significance:

BHU, founded in 1916 by Pandit Madan Mohan Malaviya, is one of the largest residential universities in Asia and holds immense historical and cultural significance.

The university's sprawling campus is home to several renowned institutions, faculties, and research centers, offering a diverse and vibrant academic environment.

Academic Excellence:

BHU is recognized for its excellence in teaching, research, and innovation across various disciplines, including humanities, sciences, engineering, medicine, and management.

The university attracts students, faculty, and researchers from across India and around the world, fostering a rich and multicultural academic community.

Student Life:

The BHU campus is bustling with student activities, clubs, societies, and cultural events, providing ample opportunities for personal growth, networking, and social engagement.

Due to large area of campus, students are often accustomed to online shopping platforms.

Community Engagement:

BHU is integral parts of the Varanasi community, contributing to local development, social welfare, and cultural heritage preservation initiatives.

The university serves as a hub for academic, cultural, and artistic exchanges, enriching the lives of students, faculty, staff, and residents alike.

Economic Impact:

BHU plays a significant role in the local economy, generating employment opportunities, supporting small businesses, and driving innovation and entrepreneurship in the region.

The presence of BHU attracts visitors, tourists, and investors to Varanasi, contributing to the city's economic growth and development.

Conclusion:

The study area of Banaras Hindu University (BHU) offers a unique and dynamic setting for research on online shopping behavior among students, faculty, staff, and alumni. By focusing on this vibrant academic community, the research aims to provide valuable insights into the preferences, habits, and satisfaction levels of online shoppers within the BHU ecosystem.

Data Collection

Data represents information collected in the form of numbers and text. Data collection is generally done after the experiment or observation. Primary data and Secondary data are helpful in planning and estimating. Data collection is either qualitative or quantitative.

The two types of data are as follows.

1. Primary Data:

When an investigator collects data himself with a definite plan or design in his/her way, then the data is known as primary data. Generally, the results derived from the primary data are accurate as the researcher gathers the information. But, one of the disadvantages of primary data collection is the expenses associated with it. Primary data research is very time-consuming and expensive.

2. Secondary Data:

Data that the investigator does not initially collect but instead obtains from published or unpublished sources are secondary data. Secondary data is collected by an individual or an institution for some purpose and are used by someone else in another context. It is worth noting that although secondary data is cheaper to obtain, it raises concerns about accuracy. As the data is second-hand, one cannot fully rely on the information to be authentic.

Primary Data Collection Method:

Primary or raw data is obtained directly from the first-hand source through experiments, surveys, or observations. The primary data collection method is further classified into two types, and they are given below:

- Quantitative Data Collection Methods
- Qualitative Data Collection Methods

I created a well-structured questionnaire and then form it in Google forms as this process was more accessible for the responses. Nearly every respondent have smartphones. So, it is relatively easy to use google forms.

Questionnaire Development:

A structured questionnaire was developed using Google Forms, consisting of multiple-choice questions, Likert-scale items, and open-ended prompts.

The questionnaire was designed to capture demographic information, online shopping behavior, satisfaction levels, and factors influencing platform preferences.

Data Collection Procedures:

The Google Forms questionnaire was distributed to the target population via university email lists, social media platforms, and departmental communication channels.

Participants were provided with a link to the online survey and instructed to complete it at their convenience within a specified timeframe.

Anonymity and confidentiality were ensured, and participants were informed of their right to withdraw from the study at any time.

Monitoring and Follow-Up:

Data collection progress was monitored in real-time using Google Forms analytics, allowing for timely follow-up with potential respondents and adjustments to distribution strategies as needed.

Reminder emails and social media posts were sent periodically to encourage survey participation and maximize response rates.

Descriptive Data Analysis

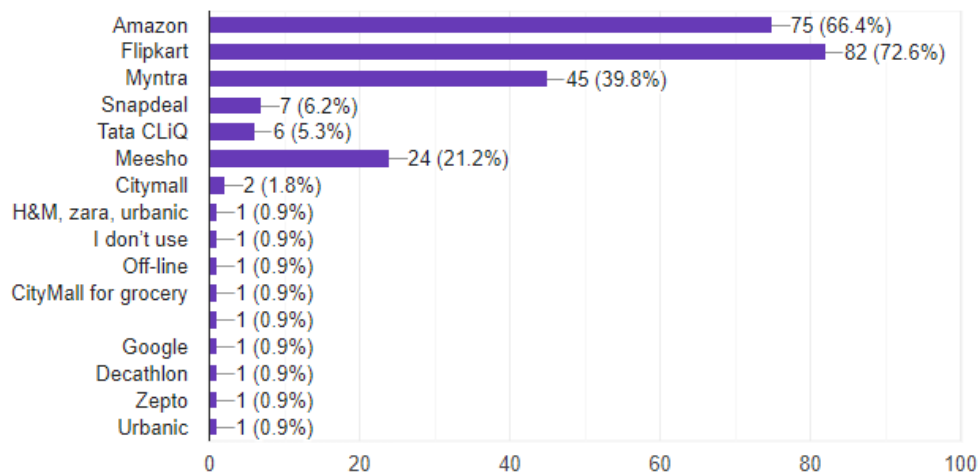
To know that which online shopping platform is mostly used in BHU, I have asked a question regarding that in my questionnaire.

The description of that question is shown below.

The following data has multiple selection option.

Which online shopping platforms do you use regularly? (Check all that apply)

113 responses



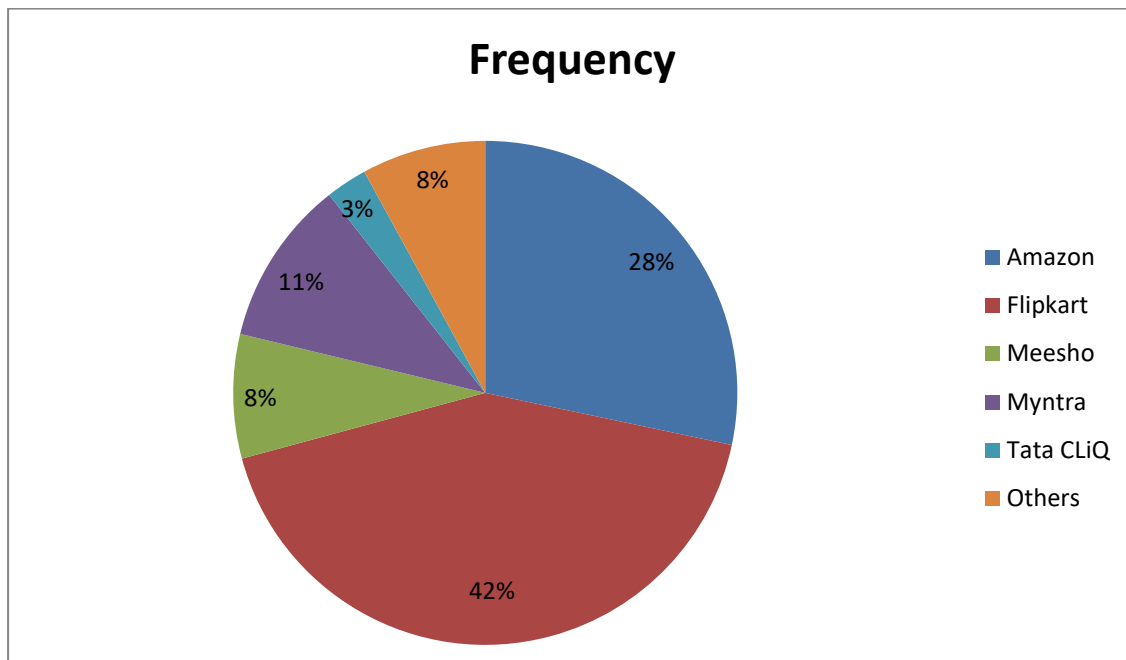
This bar graph describes that most of the BHU students are using Flipkart and Amazon for shopping online.

Another question is asked about last used platform, and that data is also important to know online shopping platform usage data.

1. The frequency table of last used online shopping platform is given below.

Platforms	Frequency
Amazon	32
Flipkart	48
Meesho	9
Myntra	12
Tata CLiQ	3
Others	9
Total	113

For the better understanding of this frequency table, pie chart will be more suitable.



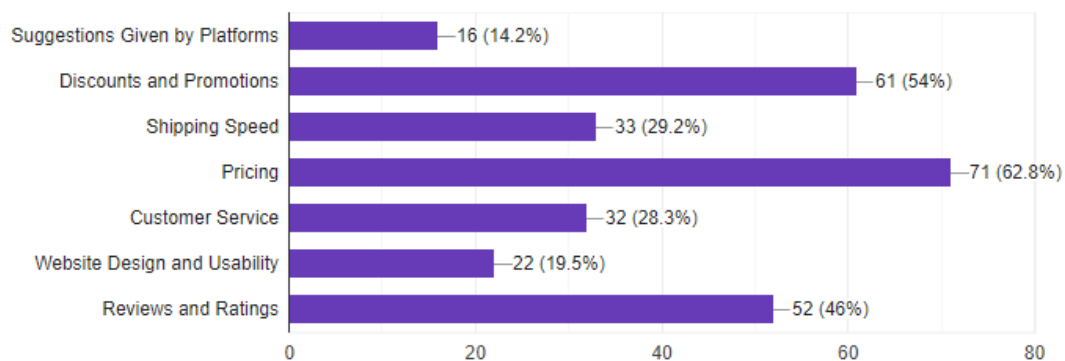
The pie chart is also showing that Flipkart and Amazon are widely used in BHU in comparison of other online shopping platforms.

2. What factors influenced the decision of respondent to shop on the online platforms (Check all that apply)

The following data has multiple selection option. It means one respondent can select multiple options.

What factors influenced your decision to shop on the online platform you used most recently? (Check all that apply)

113 responses

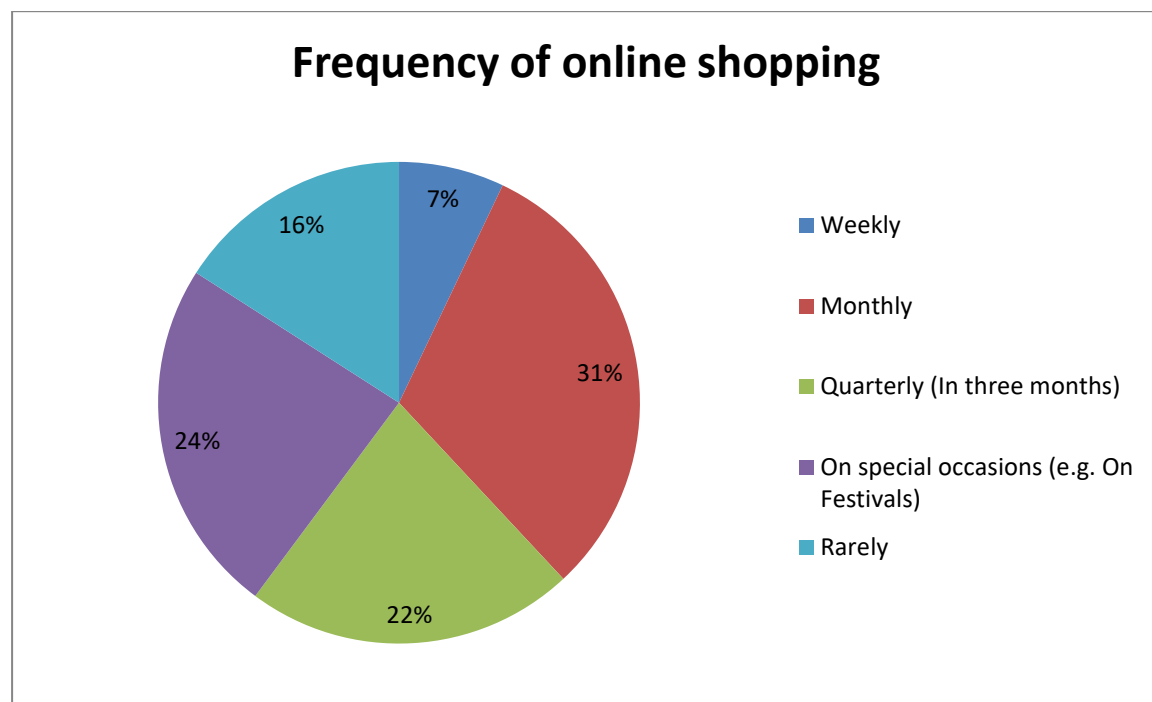


Pricing, Discounts and Promotions, Reviews and Ratings plays important role in influencing the decision to shop on the online shopping platforms. Other factors are also important.

3. The frequencies of online shopping of the respondents are shown in following frequency table.

Intervals	Frequency of online shopping
Weekly	8
Monthly	35
Quarterly (In three months)	25
On special occasions (e.g. On Festivals)	27
Rarely	18
Total	113

To provide a clear emphasis on short look, I am providing a pie chart for same frequency table.

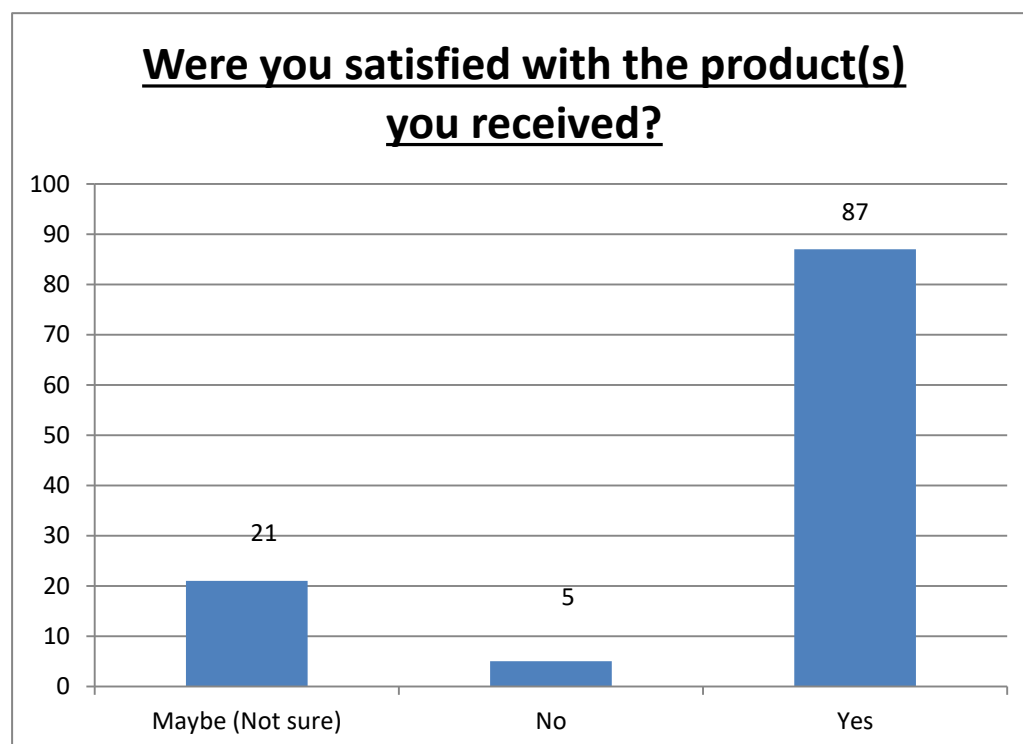


In BHU, the frequency of online shopping is less as compared to tier-1 or tier-2 cities. Mostly people buy in a month or on special occasions. It shows that, still there are significant people who are not shopping online on frequent manners. Nearly **16%** respondents rarely do online shopping and this is a quite significant figure.

4. Were you satisfied with the product(s) you received?

The respondents were asked whether they were satisfied with the product they received from online shopping or not.

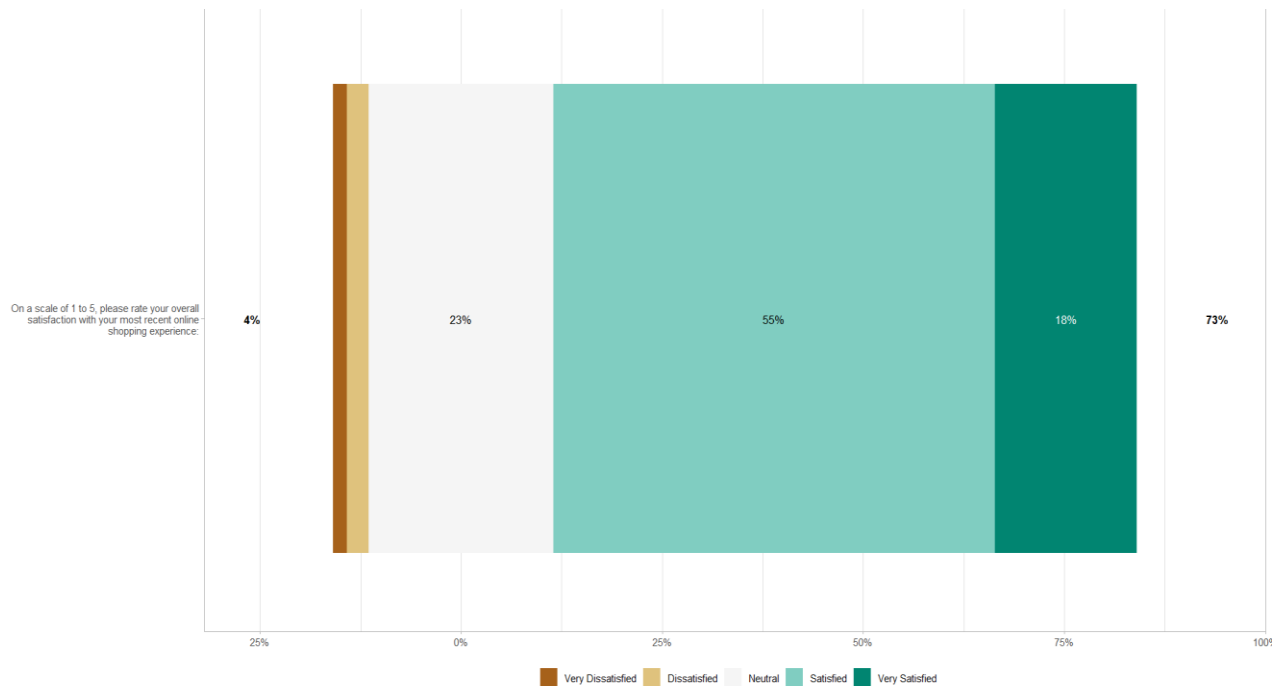
Row Labels	Count of Were you satisfied with the product(s) you received?
Maybe (Not sure)	21
No	5
Yes	87
Grand Total	113



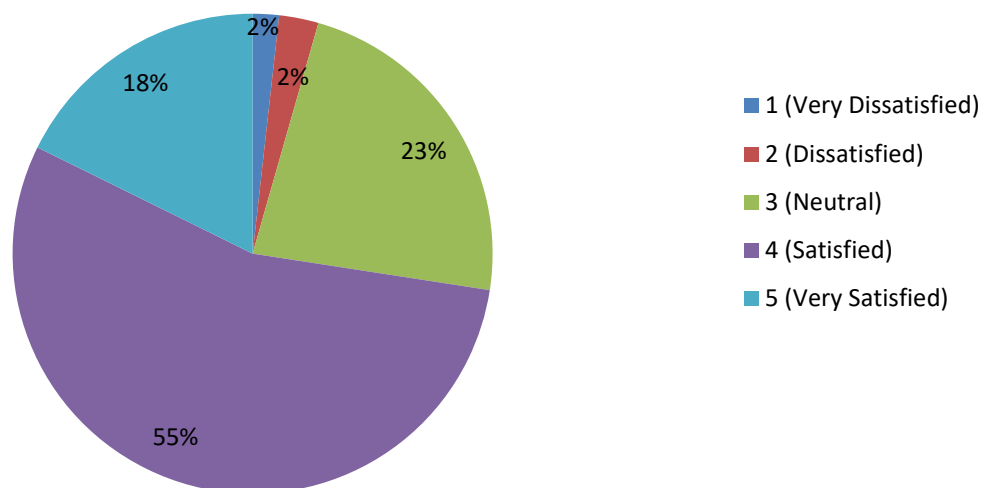
Nearly all respondents are satisfied with their product. It describes that the quality of goods provided by online shopping platform is quite good. While, a significant number of people are still not sure.

5. On a scale of 1 to 5, please rate your overall satisfaction with your most recent online shopping experience:

To explain this, I am providing a **Likert scale** that is a rating scale that quantitatively assesses opinions, attitudes, or behaviors.



Count of On a scale of 1 to 5, please rate your overall satisfaction with your most recent online shopping experience:

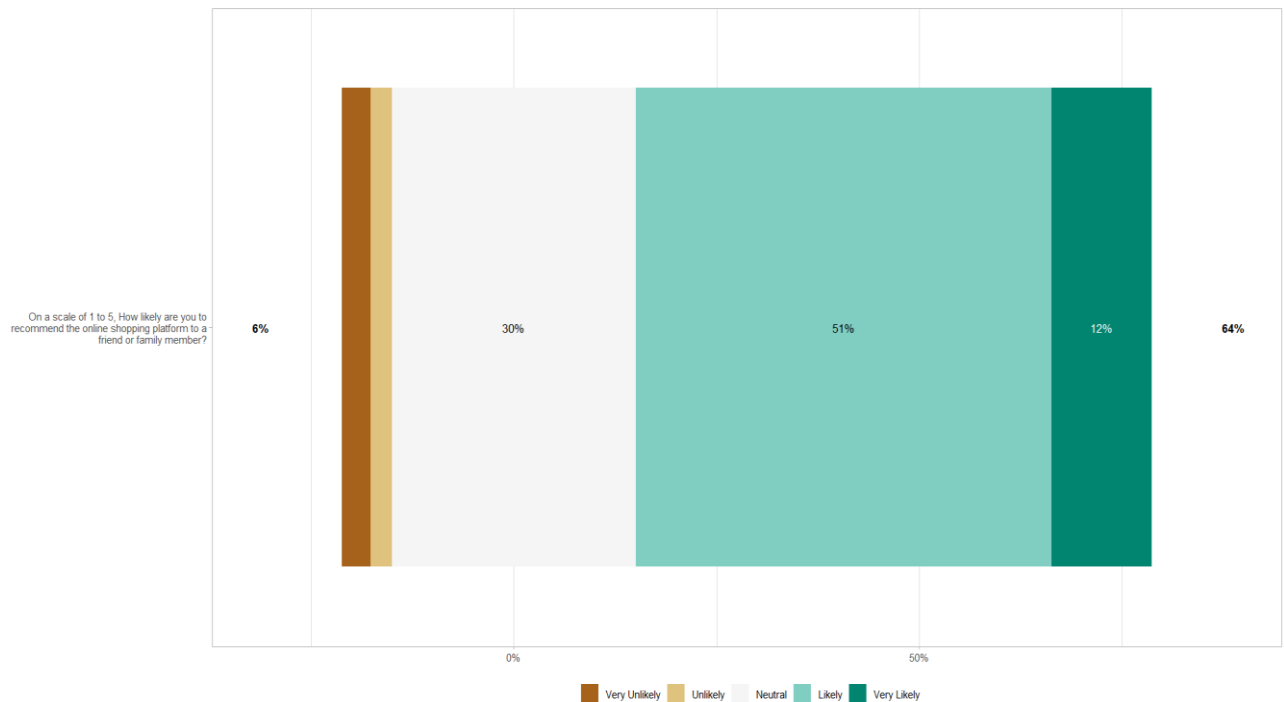


Most of the respondents were satisfied with the product and some are neutral.

Very less proportion of respondents were dissatisfied.

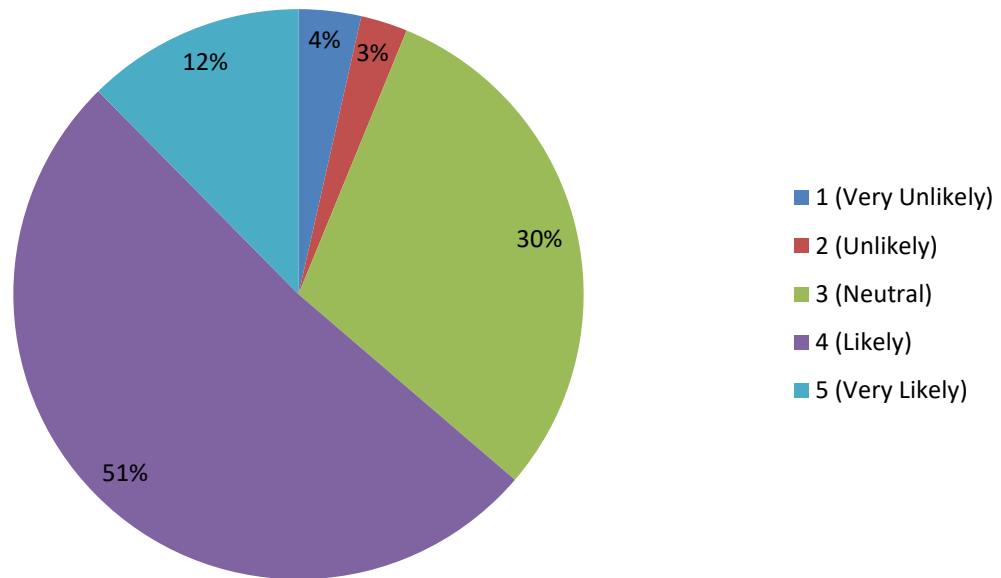
6. On a scale of 1 to 5, How likely are you to recommend the online shopping platform to a friend or family member:

To explain this, I am providing a Likert scale that is a rating scale that quantitatively assesses opinions, attitudes, or behaviors.



The image of likert scale is not much clear, so in addition a pie chart is provided on the next page.

Count of On a scale of 1 to 5, How likely are you to recommend the online shopping platform to a friend or family member?



Most of the respondents are likely to recommend the online shopping platforms and some are neutral.

Very less proportion of respondents are not likely to recommend the online shopping platform.

These are some of the important description of the data collected from the questionnaire.

Chi-Square Data Analysis

Chi square test independence of attributes: (Shopping frequency with gender)

- **H₀:** There is no significant association between frequency of online shopping and gender.
- **H₁:** There is significant association between frequency of online shopping and gender.

Observed Frequencies

Gender	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
Female	13	10	6	4	4	37
Male	22	17	19	14	4	76
Grand Total	35	27	25	18	8	113

Expected Frequencies

Gender	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
Female	11.46017699	8.840707965	8.185840708	5.894	2.61947	37
Male	23.53982301	18.15929204	16.81415929	12.11	5.38053	76
Grand Total	35	27	25	18	8	113

Degree of freedom = $(5 - 1) \times (2 - 1) = 4$
without pooling, the degree of freedom is 4.

But, pooling will change (reduce) the degree of freedom.

Now, here the expected frequency of females who shop weekly is less than 5.

So, it is necessary to pool with the data of females who shop rarely.

Now, the reduced degree of freedom = $(4 - 1) = 3$

calculations for chi square

Class	Observed freq(f_i)	Expected freq(e_i)	$(f_i - e_i)$	$(f_i - e_i)^2 / e_i$
Female Monthly	13	11.46017699	1.539823009	0.206895138
Female On special occasions	10	8.840707965	1.159292035	0.152019276
Female Quarterly	6	8.185840708	-2.185840708	0.583678546
Female Rarely and Weekly	8	8.513274336	-0.513274336	0.030945854
Male Monthly	22	23.53982301	-1.539823009	0.100725264
Male On special occasions	17	18.15929204	-1.159292035	0.074009384
Male Quarterly	19	16.81415929	2.185840708	0.284159292
Male Rarely	14	12.10619469	1.89380531	0.29625317
Male Weekly	4	5.380530973	-1.380530973	0.354215184
Total	113	113		2.082901108

Calculated value of $\chi^2 = 2.0829$

Degree of freedom = 3

Level of Significance = 0.05 or 5%

Critical value of $\chi^2 = 7.815$

Since calculated value of χ^2 is very less than critical value of χ^2 , there is no sufficient information to reject H_0 .

Chi square test independence of attributes: (Shopping frequency with age)

- **H₀:** There is no significant association between frequency of online shopping and age.
- **H₁:** There is significant association between frequency of online shopping and age.

Contingency Table

Age	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
16-21	22	12	14	14	4	66
22 and above	13	15	11	4	4	47
Grand Total	35	27	25	18	8	113

Conducting chi-square test with r-programming gives:

```
> # Create a matrix with the counts of shopping frequency by age group
> data <- matrix(c(22, 12, 14, 14, 4,
+               13, 15, 11, 4, 4),
+               nrow = 2,
+               dimnames = list(
+                 "Age" = c("16-21", "22 and above"),
+                 "Shopping Frequency" = c("Monthly", "On special occasions", "Quarterly", "Rarely", "Weekly")))
>
> # Perform chi-square test of independence
> chi_square_result <- chisq.test(data)
Warning message:
In chisq.test(data) : Chi-squared approximation may be incorrect
>
> # Print the result
> print(chi_square_result)

Pearson's Chi-squared test

data: data
X-squared = 8.1159, df = 4, p-value = 0.08742

>
> fisher.test(data)$p.value
[1] 0.08404268
```

The p-value obtained from Fisher's Exact Test is 0.08404268, which is greater than the significance level of 0.05.

With a p-value greater than the significance level, we fail to reject the null hypothesis (H_0).

Therefore, there is insufficient evidence to conclude that there is a significant association between the frequency of online shopping and age based on Fisher's Exact Test.

Chi square test independence of attributes: (Shopping frequency with residence)

- H_0 : There is no significant association between frequency of online shopping and residence.
- H_1 : There is significant association between frequency of online shopping and residence.

Contingency Table

Residence	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
Day Scholar	11	10	13	6	3	43
Hosteller	24	17	12	12	5	70
Grand Total	35	27	25	18	8	113

```

> # Create a matrix with the counts of shopping frequency by residence type
> data <- matrix(c(11, 10, 13, 6, 3,
+                 24, 17, 12, 12, 5),
+               nrow = 2,
+               dimnames = list(
+                 "Residence" = c("Day Scholar", "Hosteller"),
+                 "Shopping Frequency" = c("Monthly", "On special occasions", "Quarterly", "Rarely", "Weekly")))
>
> # Perform chi-square test of independence
> chi_square_result <- chisq.test(data)
>
> # Print the result
> print(chi_square_result)

      Pearson's Chi-squared test

data: data
X-squared = 22.697, df = 4, p-value = 0.0001455

>
> fisher.test(data)$p.value
[1] 6.653235e-05

```

The p-value obtained from Fisher's Exact Test is 6.653235e-05 (or approximately 0.0000665), which is much smaller than the significance level of 0.05.

With a p-value smaller than the significance level, we reject the null hypothesis (H_0).

Therefore, there is sufficient evidence to conclude that there is a significant association between the frequency of online shopping and residence based on Fisher's Exact Test.

Chi square test independence of attributes: (Shopping frequency with degree)

- **H_0 :** There is no significant association between frequency of online shopping and degree.
- **H_1 :** There is significant association between frequency of online shopping and degree.

Contingency Table

Degree	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
Bachelor's Degree	23	16	15	16	5	75
Higher Degree	1	2	3	0	0	6
Intermediate or Equivalent	9	3	6	2	1	21
Master's Degree	2	6	1	0	2	11
Grand Total	35	27	25	18	8	113

```

> # Create a matrix with the counts of shopping frequency by degree
> data <- matrix(c(23, 16, 15, 16, 5,
+               1, 2, 3, 0, 0,
+               9, 3, 6, 2, 1,
+               2, 6, 1, 0, 2),
+               nrow = 4,
+               dimnames = list(
+                 "Degree" = c("Bachelor's Degree", "Higher Degree", "Intermediate or Equivalent", "Master's Degree"),
+                 "Shopping Frequency" = c("Monthly", "On special occasions", "Quarterly", "Rarely", "Weekly")))
>
> # Perform chi-square test of independence
> chi_square_result <- chisq.test(data)
Warning message:
In chisq.test(data) : Chi-squared approximation may be incorrect
>
> # Print the result
> print(chi_square_result)

Pearson's Chi-squared test

data: data
X-squared = 29.001, df = 12, p-value = 0.003939

>
> fisher.test(data, simulate.p.value = T)$p.value
[1] 0.004997501

```

The p-value obtained from Fisher's Exact Test is 0.004997501, which is smaller than the significance level of 0.05.

With a p-value smaller than the significance level, we reject the null hypothesis (H_0).

Therefore, there is sufficient evidence to conclude that there is a significant association between the frequency of online shopping and degree based on Fisher's Exact Test.

Chi square test independence of attributes: (Shopping frequency with Annual household income)

- **H₀:** There is no significant association between frequency of online shopping and annual household income.
- **H₁:** There is significant association between frequency of online shopping and annual household income.

Contingency Table

Annual Household Income	Monthly	On special occasions (e.g. On Festivals)	Quarterly (In three months)	Rarely	Weekly	Grand Total
Less than Rs. 50,000	4	6	3	3	0	16
Rs. 50,000 - Rs. 99,999	5	6	5	6	0	22
Rs. 1,00,000 - Rs. 1,49,999	5	4	2	5	1	17
Rs. 1,50,000 - Rs.1,99,999	1	2	4	1	1	9
Rs. 2,00,000 - Rs. 4,99,999	5	4	8	2	1	20
Rs. 5,00,000 - Rs. 11,99,999	12	4	1	1	1	19
Rs. 12,00,000 and above	3	1	2	0	4	10
Grand Total	35	27	25	18	8	113

```

> # Create a matrix with the counts of shopping frequency by annual household income
> data <- matrix(c(4, 6, 3, 3, 0,
+                 5, 6, 5, 6, 0,
+                 5, 4, 2, 5, 1,
+                 1, 2, 4, 1, 1,
+                 5, 4, 8, 2, 1,
+                 12, 4, 1, 1, 1,
+                 3, 1, 2, 0, 4),
+               nrow = 7,
+               dimnames = list(
+                 "Annual Household Income" = c("Less than Rs. 50,000", "Rs. 50,000 - Rs. 99,999", "Rs. 1,00,000
- Rs. 1,49,999",
+                 "Rs. 1,50,000 - Rs.1,99,999", "Rs. 2,00,000 - Rs. 4,99,999", "R
s. 5,00,000 - Rs. 11,99,999",
+                 "Rs. 12,00,000 and above"),
+                 "Shopping Frequency" = c("Monthly", "On special occasions", "Quarterly", "Rarely", "Weekly")))
>
> # Perform chi-square test of independence
> chi_square_result <- chisq.test(data)
Warning message:
In chisq.test(data) : Chi-squared approximation may be incorrect
>
> # Print the result
> print(chi_square_result)

        Pearson's Chi-squared test

data: data
X-squared = 41.027, df = 24, p-value = 0.01654

>
> fisher.test(data, simulate.p.value = T)$p.value
[1] 0.003998001

```

The p-value obtained from Fisher's Exact Test is 0.003998001, which is smaller than the significance level of 0.05.

With a p-value smaller than the significance level, we reject the null hypothesis (H_0).

Therefore, there is sufficient evidence to conclude that there is a significant association between the frequency of online shopping and annual household income based on Fisher's Exact Test.

Chi square test independence of attributes: (Satisfaction with Different Gender)

- **H₀:** There is no significant association between the satisfaction on product received (online shopping) and gender.
- **H₁:** There is significant association between the satisfaction on product received (online shopping) and gender.

Contingency Table

Gender	Maybe (Not sure)	No	Yes	Grand Total
Female	9	1	27	37
Male	12	4	60	76
Grand Total	21	5	87	113

```
> # Create a matrix with the counts of responses by gender and answer
> data <- matrix(c(9, 1, 27,
+               12, 4, 60), # Note: NA added for the missing value in the Male row
+               nrow = 2,
+               byrow = TRUE,
+               dimnames = list(
+                 "Gender" = c("Female", "Male"),
+                 "Response" = c("Maybe (Not sure)", "No", "Yes")))
>
> # Remove the last column containing NA values
> data <- data[, -ncol(data)]
>
> # Perform chi-square test of independence
> chi_square_result <- chisq.test(data)
Warning message:
In chisq.test(data) : Chi-squared approximation may be incorrect
>
> # Print the result
> print(chi_square_result)

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

data: data
X-squared = 0.18726, df = 1, p-value = 0.6652

>
> fisher.test(data)$p.value
[1] 0.6169048
```

The chi-square test statistic obtained is 0.18726, with 1 degree of freedom.

The p-value associated with this test statistic is 0.6652.

With a p-value of 0.6652, which is greater than the significance level of 0.05, we fail to reject the null hypothesis (H_0).

Therefore, there is insufficient evidence to conclude that there is a significant association between the satisfaction on product received (online shopping) and gender based on Pearson's Chi-Square Test.

Conclusion

Usage of Online Shopping Platforms at BHU:

- The bar graph and pie chart illustrate that Flipkart and Amazon are the most widely used online shopping platforms among BHU students.
- These platforms are preferred over others, indicating their popularity and dominance in the online shopping market within the BHU community.

Factors Influencing Online Shopping Decisions:

- Pricing, discounts and promotions, as well as reviews and ratings, are identified as key factors influencing the decision to shop on online platforms.
- This suggests that BHU students are sensitive to pricing strategies, seek value through discounts, and rely on peer feedback when making purchasing decisions.

Frequency of Online Shopping:

- BHU students exhibit a lower frequency of online shopping compared to residents of tier-1 or tier-2 cities.
- Most purchases occur on a monthly basis or during special occasions, indicating a less frequent shopping behavior among the BHU population.
- Additionally, a significant proportion of respondents rarely engage in online shopping, highlighting potential areas for increasing online shopping adoption and frequency.

Satisfaction with Online Shopping Experience:

- Overall, respondents express satisfaction with the quality of goods received through online shopping platforms.
- While a majority is satisfied, a notable proportion is unsure or neutral about their satisfaction, indicating a need for further investigation into their experiences and concerns.
- Dissatisfaction with online shopping experiences is relatively low, suggesting that online platforms generally meet or exceed expectations in terms of product quality and service.

Likelihood of Recommending Online Shopping Platforms:

- Most respondents express likelihood to recommend online shopping platforms to others, indicating a positive perception of their overall shopping experiences.
- However, a notable proportion are neutral about recommending, suggesting a need for platforms to enhance customer satisfaction and loyalty to drive advocacy.

Hypothesis Testing Conclusions:

- The conclusions drawn from hypothesis testing indicate mixed results regarding the association between the frequency of online shopping and demographic factors such as age, residence, degree, annual household income, and gender.
- While significant associations are found between the frequency of online shopping and residence, degree, and annual household income, no significant association is observed with gender.
- We do not have enough evidence to support the claim that there is a significant association between the satisfaction on product received

(online shopping) and gender. This suggests that gender may not be a significant factor influencing satisfaction levels with online shopping products among the respondents in the study.

Overall, the combination of descriptive analysis and hypothesis testing results provides a comprehensive understanding of online shopping behavior within the BHU community, offering valuable insights for both academic research and practical application in marketing and business strategies.

Field Experience and Difficulties

It is common practice that the difficulties are always encountered when one starts to do some new work. My experience was also same but I took them lightly. A brief description of the difficulties which I faced while collecting and analysing the data are given below,

- I faced many problems while collecting data through Google form. For user convenience, I made available Google form link in Quick response (QR) code form but many students did not fill the form because form did not serve any purpose for them. Participants may ignore or overlook survey invitations, leading to low participation rates and potentially biased results. Additional outreach efforts may be required to encourage participation and increase response rates.
- The most common difficulty related to Google form data collection was that we got many incomplete forms, which we had to separate from other forms.
- Designing an effective online survey requires careful consideration of question wording, response options, and survey flow. Poorly designed surveys may lead to confusion, respondent fatigue, and inaccurate or incomplete responses. I did pilot test my surveys to identify and address any design flaws before launching the full-scale data collection.

Due to these difficulties, my project work became much more interesting and challenging for me. I got a totally new experience. I met many types of people during my project and came to know about different kinds of problems faced while applying theory in real life.

I think, it was totally new experience to work with real life data and to make reliable interpretation from it. I am thankful to my supervisor who helped and guided me at each and every step of our project.

Suggestions

- **Targeted Marketing Strategies:** Given the popularity of Flipkart and Amazon among BHU students, online shopping platforms can tailor their marketing strategies to better engage this demographic. Utilizing targeted advertisements, promotions, and discounts on these platforms may further increase their appeal and usage among BHU students.
- **Enhanced User Experience:** To capitalize on the factors influencing online shopping decisions such as pricing, discounts, and reviews, platforms should prioritize enhancing the overall user experience. This can include optimizing website design and usability, streamlining checkout processes, and providing personalized recommendations based on user preferences.
- **Promotion of Online Shopping Benefits:** Despite the satisfaction expressed with online shopping experiences, there remains a significant proportion of respondents who are not frequent online shoppers. Educating individuals about the benefits of online shopping, such as convenience, variety, and cost-effectiveness, may help increase adoption and frequency of online shopping among BHU students.
- **Customer Feedback and Improvement:** Platforms should actively seek and incorporate customer feedback to continuously improve their services. Monitoring and addressing customer concerns, enhancing product quality, and refining customer service processes can help maintain high levels of satisfaction and loyalty among online shoppers.

By implementing these suggestions, researchers and practitioners can enhance the effectiveness of their online data collection efforts, improve customer satisfaction and loyalty, and better meet the needs and preferences of online shoppers within the BHU community.

References and Bibliography

Ithnin, M. S. S., Abdullah, N. A., & Iskandar, M. L. (2022). Relationship between Perceived Usefulness and Online Shopping Behaviour among University Students. *Journal of International Business, Economics and Entrepreneurship*, 7(1), 62-62.

Daroch, B., Nagrath, G., & Gupta, A. (2021). A study on factors limiting online shopping behaviour of consumers. *Rajagiri Management Journal*.

Kuswanto, H., Pratama, W. B. H., & Ahmad, I. S. (2020). Survey data on students' online shopping behaviour: A focus on selected university students in Indonesia. *Data in brief*, 29, 105073.

Goon A.M., Gupta M.K. and Das Gupta B. (1980), *An outline of statistical theory Volume 1*, The world press publishers pvt. Ltd.

Appendix:

Questionnaire

“AN ANALYSIS OF ONLINE SHOPPING BEHAVIOUR AMONG BHU STUDENTS”

Dear respondent,

This is Aditya Narayan Singh, and I am a student of Banaras Hindu University. This survey aims to conduct a statistical study of online shopping platforms to understand the relationship between online shopping frequency and demographic variables (age, income level, gender, residence, and education level).

Your privacy will be maintained and kept confidential during the research process and publication.

Thank you for participating in my survey. Your feedback is valuable in helping us understand and improve the online shopping experience. Please take a few minutes to answer the following questions honestly and to the best of your ability.

Age:

- ☐ 16-21
- ☐ 22 and above

Gender:

- ☐ Male
- ☐ Female
- ☐ Other: _____

Residence:

- ☐ Hosteller
- ☐ Day Scholar

Education Level:

- ☐ Intermediate or Equivalent
- ☐ Bachelor's Degree
- ☐ Master's Degree
- ☐ Higher Degree

Annual Household Income:

- ☐ Less than Rs. 50,000
- ☐ Rs. 50,000 - Rs. 99,999
- ☐ Rs. 1,00,000 - Rs. 1,49,999
- ☐ Rs. 1,50,000 - Rs. 1,99,999
- ☐ Rs. 2,00,000 - Rs. 4,99,999
- ☐ Rs. 5,00,000 - Rs. 11,99,999
- ☐ Rs. 12,00,000 and above

How frequently do you shop online?

- ☐ Weekly
- ☐ Monthly
- ☐ Quarterly (In three months)
- ☐ On special occasions (e.g., On Festivals)
- ☐ Rarely
- ☐ Never

Which online shopping platforms do you use regularly? (Check all that apply)

- ☐ Amazon
- ☐ Flipkart
- ☐ Myntra
- ☐ Snapdeal
- ☐ Tata CLiQ
- ☐ Meesho
- ☐ Other: _____

Which online shopping platforms did you use recently? (For your last shopping)

- ☐ Amazon
- ☐ Flipkart
- ☐ Myntra
- ☐ Snapdeal
- ☐ Tata CLiQ
- ☐ Meesho
- ☐ Other: _____

On a scale of 1 to 5, please rate your overall satisfaction with your most recent online shopping experience:

- ☐ 1 (Very Dissatisfied)
- ☐ 2 (Dissatisfied)
- ☐ 3 (Neutral)
- ☐ 4 (Satisfied)
- ☐ 5 (Very Satisfied)

What factors influenced your decision to shop on the online platform you used most recently? (Check all that apply)

- ☐ Suggestions Given by Platforms
- ☐ Discounts and Promotions
- ☐ Shipping Speed
- ☐ Pricing
- ☐ Customer Service
- ☐ Website Design and Usability
- ☐ Reviews and Ratings

Were you satisfied with the product(s) you received?

- ☐ Yes
- ☐ No
- ☐ Maybe (Not sure)

On a scale of 1 to 5, How likely are you to recommend the online shopping platform to a friend or family member?

- ☐ 1 (Very Unlikely)
- ☐ 2 (Unlikely)
- ☐ 3 (Neutral)
- ☐ 4 (Likely)
- ☐ 5 (Very Likely)

Thanks!

Thank you for your time and feedback. Your input will help us improve the structure of online shopping platforms.