E-Commerce

by 919 ..

Submission date: 19-May-2020 01:12PM (UTC+0530)
Submission ID: 1327603889
File name: 2k18_mba_919_rohit_For_plag_check.docx (2.91M)
Word count: 11346
Character count: 57811

EXECUTIVE SUMMARY

E-Commerce is one of the growing industries which is under a lot of innovation these days. The e-commerce market has changed the way of doing business. This industry came under a lot of limelight after the emergence of Analytics field in IT world. Huge amount of data is being collected using the customer accounts, online buying behaviour, loyalty cards, card transactions, etc. This project talks about the various aspects of E-Commerce Industry. It explains the use of the user's data and analytics in the E-Commerce industry. Also explains how an online retailer uses the data to build a recommender engine that suggest products to its online customers, how customers get real time updates for a newly launched products, customised deals and promotions.

In an online platform, every single piece of data from your search till your final purchase and after sales experience, can easily be captured. Aim of this project is to find all the possible data points that a retailer can easily collect to improve its business and user experiences. This project will focus on the global leader of E-commerce industry, i.e. **Amazon**, to know how the retailers use the click-stream data and historical purchase data of customers for showing customised results on customized web pages on their platform, in order to provide their customers best experience with their services. Also, the objective of this research is to know the real customers' perception towards online shopping in India which have been done in this research by collecting data from primary source. The report additionally consists recommendations and conclusion in keeping with my point of view, which I assume would improve the business enterprises.

CHAPTER 1: INTRODUCTION

Most of people today use online websites and portals to shop at times. People prefer to shop online more rather than going out personally to buy products. With the increase in internet users in India, number of people using online portals to buy products is also increasing. With more discounts, lowered price range and increased variety in products, the scope for online shopping has increased vaguely. Amazon is one such online shopping site which is based in Seattle. It focuses on E-commerce, Artificial Intelligence, Digital Streaming and Cloud Computing. Amazon was launched in India in February 2012 as shopping website.

E-commerce as industry has a lot of growth and innovation these days. It has changed the look of business. Huge amount of data is collected in e-commerce. An online platform takes in a person's personal data like name, contact details and even card details. They save it further for future purchases. Not only this website, like these also save preferences or viewed products and personalise the whole shopping for an individual. This way the whole shopping process becomes more personalized and catered to each person individually.

The aim behind this research is to find out how satisfied customers are by the services provided to them by the company and what all data driven technologies are used in today's e-commerce industry. We want to know the rate of people who actually prefer shopping online more than real time shopping.

1.1 INDUSTRY ANALYSIS

1.1.1 Industry Profile

E-commerce is the process of buying and selling of products, services on web or other networks. A physical store (brick and mortar) can convert into an e-commerce business just by adding a few things: a virtual storefront which includes a web catalogue in it. Most of the time, e-business refers exclusively to Internet business, but it also refers to any kind of business that uses Internet to enhance their productivity and profitability. The e-commerce has transformed the way of doing business in India. The Indian e-commerce market was US$ 38.5 billion as of 2017, it is predicted to grow to US$ 200 billion by 2026 [Source: IBEF Report, 2019]. The growth of the industry is followed by the increase in the internet and smartphone usage in India. According to IBEF report, the current digital transformation within the country is predicted to increase with the rate of India's total Internet user base to 829 million by 2021, which was around 636.73 million in FY19. Due to the support of e-commerce, India's internet economy is additionally expected to double from US$ 125 billion as of April 2017 to US$ 250 billion by 2020. India's E-commerce Revenue is expected to grow with an annual rate of 51 percent, becoming the highest in the world, from US$ 39 billion as of 2017 to US$ 120 billion by 2020.

1.1.2 Market Size

As per a report, the Indian E-commerce market is predicted to grow from US$ 38.5 billion in 2017 to US$ 200 billion by 2027 due to the rising smartphone penetration, increase in number of internet users, the launch of 4G networks and increasing consumer income. The sales of Online retailers are also expected to grow by 31% to reach the mark of US$ 32.70 billion mark in 2018, which would be led by Amazon India, Flipkart, Snapdeal etc. The smartphone shipment in India grew 9.9 percent year-by-year to 36.9 million shipments, during the April-June quarter 2019. It is expected to cross 200 million in 2020. Electronics is currently the biggest contributor to online retail sales in India as of now with a share of 48 percent, followed closely by apparel at 29 percent [Source: IBEF report on E-commerce, 2019].

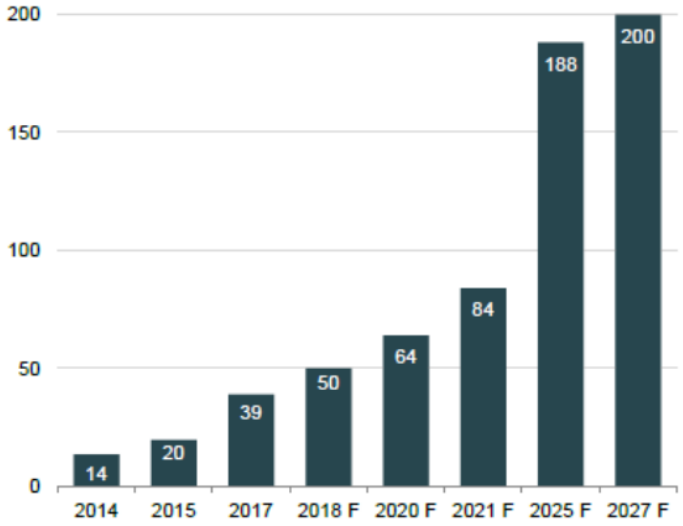


Fig: 1.1 Forecasted Growth in Indian Ecommerce Industry

[*Source: Growth of ecommerce, IBEF*]

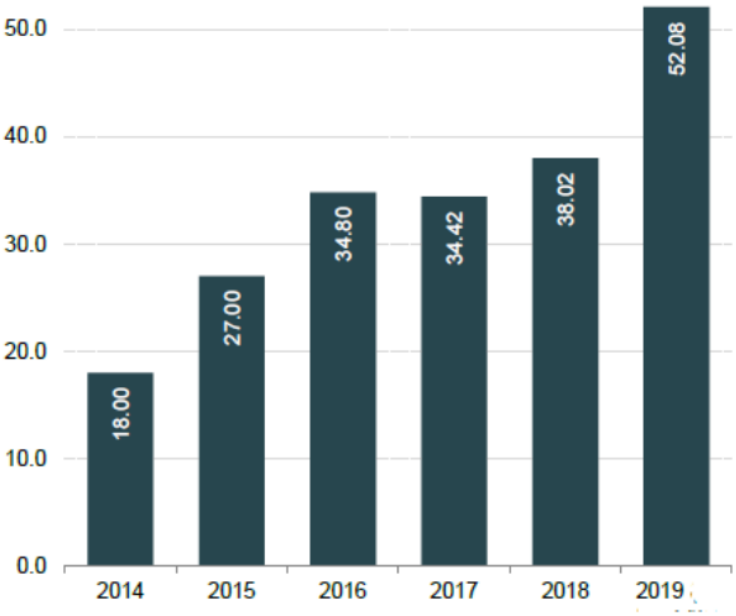


Fig:1.2 Internet Penetration in India (%)

[*Source: Economics Times*]

In 2007, Internet penetration in India was only 4 percent but now in 2020, it is more than 52.08% internet users in India. According to a report, there were 665.31 million internet users in India by the end of December 2019. Rising number of Internet users is a good sign for E-commerce industry. Refer to Fig 1.1, 1.2.

Fig:1.3 Expected Market share of Online Retail out of Total retail in India

There are plenty of opportunities for online retailers in India to capitalize, considering growing internet penetration among Indian customers. In the year 2016-17, online retail Contribution was 1.5 percent of the total retail market in India and it was expected to contribute 2.9 percent in 2018. In 2020, market share of online retail is expected to be 5% of overall retail market. Refer to fig 1.3.

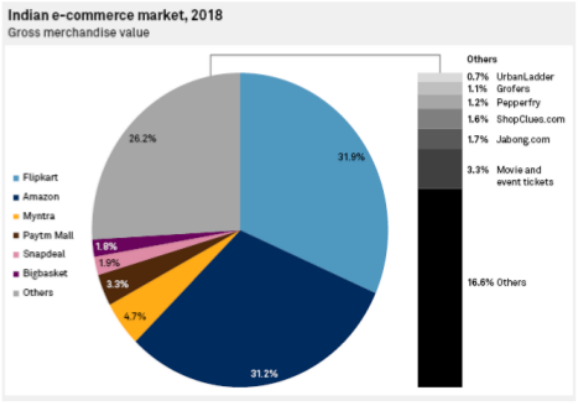


Fig:1.4 Indian E-Commerce Market in 2018

[*Source: Forrester Analytics Online Retail Forecast, 2018 to 2023 (Asia Pacific)*]

In Indian e-commerce market, there are 2 market leaders in the industry with a gross market share of more than 60 percent as of 2018. Flipkart is still at the top position (fig 1.4) followed by Amazon India. Amazon has 31.2 percent market share in the E-Commerce market of India.

1.1.3 Advantage for E-commerce in India

1). Growth in demand

- One of the fastest growing market in India is the E-commerce sector, the market for the same is expected to grow by around 1200% by 2026.
- Amazon India has worked on the Amazon Marketplace App Store which provides solutions to vendors. It registered a healthy growth in organized retail sector in 2017.
- The Indian E-commerce market is expected to reach 84 billion US dollar by 2021.

2). Attractive Opportunities

- India's Blue-Chip PE firms that used to previously avoid investing in the Indian e-commerce sector are now looking forward for investing opportunities in India.
- The start-up ecosystem present in India which are supported by Government initiatives and the rising internet penetration in the country have seen a rapid growth.

3). Increase in Investment

- The Recent growth in digital literacy has increase the flow of investment in e-commerce firms, leveling the market open for new players, while also churning out new patterns to disrupt of functioning.
- In 2018, the E-commerce companies in India got more than 7 billion US dollar in private equity and the venture capital.

4). Policy Support

- B2B e-commerce in India allows 100% FDI.
- The new FDI policy states that online entities occurring through foreign investments cannot offer products which are sold by retailers and hold an equity stake.
- The new FDI guidelines of e-commerce also state that 100 percent FDI is allowed under the automatic route in the marketplace model of e-commerce.
- Government of India has also made a huge investment by commissioning the fiber network 5G which in return will also help in boosting the e-commerce industry in India.

1.1.4 Online Retailing Business Model

India's E-commerce industry has 2 major types of E-tailing business models.

1). Marketplace Model

- The marketplace model follows the standards and instructions of a zero-inventory model. For example, Amazon.
- In this e-commerce marketplace model, merchants don't warehouse the inventory. However, they allow local seller to sell their products using company's brand name.
- Marketplace also provide merchants with shipment, distribution and payment support by collaborating with select logistics companies and financial institutions.
- The new FDI policy rules and regulations in the e-commerce market have allowed 100 percent FDI in the e-commerce marketplace model under the automatic route.

2). Inventory-led Model

- In these models, shopping websites, where online shoppers select products owned by an online shopping website, making the entire process end-to-end, along with the purchase, storage, and deliver of the product.
- For Example: Jabong, Yepme.

There are 7 key players in Indian E-commerce market currently, Flipkart leads the market with highest market share of 31.9% followed by amazon with 31.2% then Snapdeal, Myntra, Paytm Mall, Shop Clues and also Nykaa plays major role in e-commerce market.

This Project report will mainly focus on Amazon as an online retailer in India. After this industry analysis it is found that Amazon India has to follow Market Place model due to the Government regulations for foreign companies. Also, there is huge scope for Amazon to expand its business in India. The reason behind choosing Amazon India is the positive growth rate of the company it has since 5 years and it is expected that amazon will capture a good market share to become a market leader in India very soon.

1.2 ORGANIZATION ANALYSIS

1.2.1 Organisation Profile

Amazon is an US based E-Commerce company which was started by Jeff Bezoz on the 5th of July in 1994 in Bellevue, Washington. In 1994, it was initially started as an online market for books under the initial name of "Cadabra.com". But later on, it expanded more to sell electronics, software, toys, apparel, furniture, jewellery, groceries and food. It is the world's largest online market place. Amazon in the present day, provides their services in 58 different countries including China, Netherlands, United states, China, United Kingdoms, Italy, India, Japan, etc. Amazon's Prime Videos is Amazon's recent product which streams online video and audio services for entertainment purposes around the globe. It is presently the world's largest online sales company and also provides the largest cloud infrastructure through Amazon Web Services It is guided by four principles: commitment to long-term thinking, customer obsession instead of competitive focus, operational excellence and passion for invention.

Amazon had no infrastructure in India six year ago, but today it dominates the Indian Markets. At the beginning it was perceived by the investors that it will not work in India like China back in the year 2004. When Amazon initially entered in China, it didn't see much success there with Alibaba, the Chinese Competitor which dominated the e-commerce market. After the bad run in China, Bezos went for India. The facts collected to form Amazon India were based on the huge numbers of headcount which is 1.3 billion. It is four times as big as the U. S's and more double than that of Europe's. Out of these, 520 million, around 40 percent are internet users. According to the researches the yearly growth of internet users is highest in India, around six million users join monthly.

1.2.2 Amazon's Journey

Here is the Timeline of Amazon's Journey.

1994: "Cadabra.com" is built by Jeff Bezos (Founder of Amazon) in his garage, situated in Bellevue, Washington, USA.

1995: "Cadabra.com" was changed into "Amazon", which is famous today.

1996: Amazon increases its employee count to 11 and shift from the garage into a small warehouse, which later becomes Amazon's 2nd official headquarters.

1997: In this year, Amazon issues its first IPO of stock, at $18 per share.

1998: Amazon's avidity increases, which results in the takeover of various other companies, like "Drugstore.com", "Overstock.com" and "Pets.com".

1999: Jeff Bezos gets the titled "Person of the Year" by Times Magazine, which gave the company a national spotlight.

2000: Amazon has updated its logo by introducing the curved arrow pointing from A to Z.

2001: Amazon turns its first profit of $5 million on revenues of more than 1 billion US dollar.

2002: Amazon launches a new platform Amazon Web Services (AWS) for developers so that they could also include some features of amazon into their own sites.

2003: Amazon launches a new feature (Search Inside the Book) that allows customers to search for keywords in the full text of books that are listed on amazon site.

2004: Amazon's electronics item sales increased more than book sales for the very first time.

2005: Amazon's Prime membership service was launched. Today, it is a company's most popular membership program.

2006: Amazon Simple Storage, an online storage service was launched in this year.

2007: New services like Amazon Fresh, Amazon Music and Amazon Kindle were launched.

2008: Company releases a paid search feature known as Product Ads.

2009: Amazon Basics, a private label brand of amazon for electronic goods and accessories.

2010: Sales of Kindle e-books surpass sales of printed books for the first time.

2011: Subscribe & Save program was launched to offer special discounts on items delivered monthly. Kindle Fire also launched on this e-commerce platform.

2012: Amazon Supply, an online marketplace for industrial and scientific goods was launched.

2013: Amazon entered the Indian E-Commerce market.

2014: Amazon Pantry, was introduced which is a groceries and dry goods delivery service but it launched only for its prime members.

2015: Amazon completes 20 in 2015. And started focusing more on its amazon prime services.

2016: Amazon's private label line entered in FMCG categories.

2017: Amazon Prime Day sets a new record and amazon announces the acquisition of Whole Foods.

2018: Company grows excellently in this year. It reaches 100 million prime members.

2019: The company's growth and expansion continue.

1.2.4 Amazon's Business Model

Amazon has a Vague business model that relies on several revenue streams. At its base, the online store is still the primary revenue streams. Followed by Physical stores, Third Party Seller

Service, Amazon Web Services (AWS) Subscription Service, and Advertising revenues. Refer to fig 1.2.1.

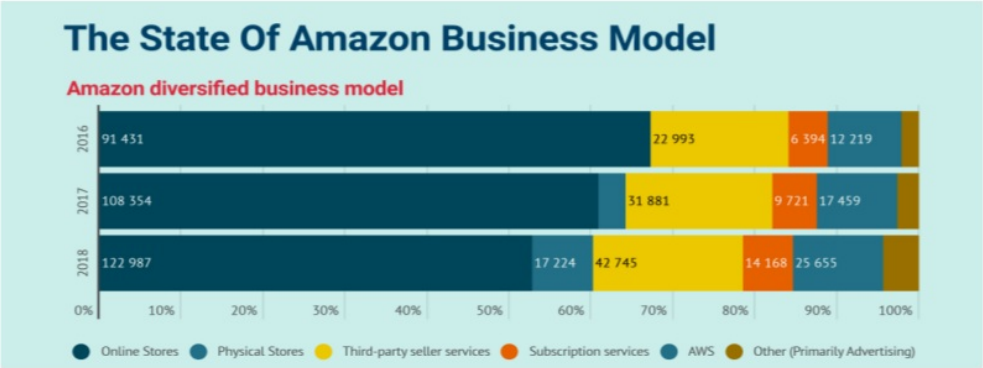


Fig:1.2.1 The State of Amazon Business Model

[*Source: Annual report, Amazon (2015-2017)*]

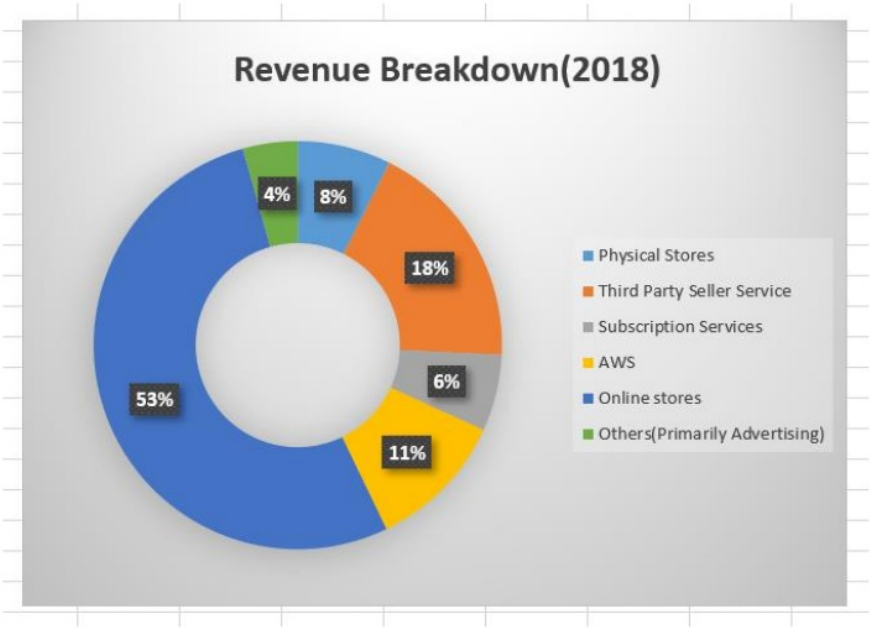


Fig:1.2.2 Revenue Breakdown of Amazon in 2018

[*Source: Annual report, Amazon (2015-2017)*]

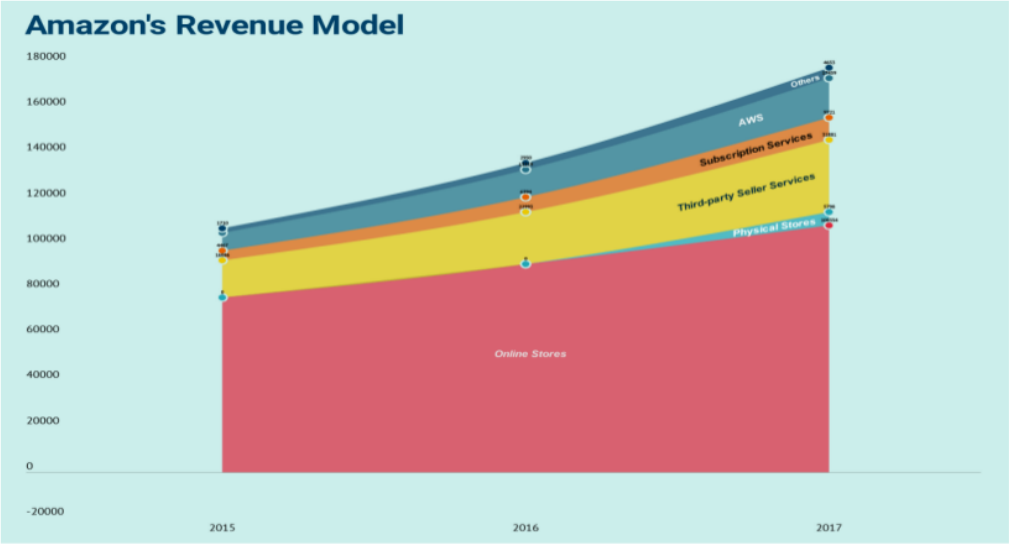


Fig:1.2.3 Amazon's Revenue Model (2018)

[Source: Annual Reports 2015-2017]

According to the annual report of Amazon, you can see that the major source of Amazon's revenue comes from its Online Stores, followed by Third Party seller service. Amazon's 53% of total revenue comes only from its main business of e-tailing. Refer to fig-1.2.2 and fig-1.2.3.

1.2.5 Swot Analysis

This Amazon SWOT analysis is to identify the company's strength, weakness, opportunities and threat to reveals how this online e-commerce giant used its competitive advantages to become a major player in the retail industry around the world and in India.

Strength

- Amazon is world's leading online retailer which enjoys the top of mind recall from worldwide consumers.
- Amazon has a good brand image and strong position in the market because of being a global giant.
- The key strengths of Amazon are cost leadership, differentiation and focus.
- Amazon has a strong IT infrastructure which is still a challenge for many e-commerce companies
- Amazon has a very strong logistics and distribution system

- Being an online retailer, Amazon has lots of customer data which can be used to for making data driven marketing strategies.
- Amazon has the largest number of third-party sellers on its platform

Weakness

- In India, Amazon works at a very low margin due to the different consumer behaviour observed among Indian consumers.
- In order to offer free shipping or low-cost shipping to its Indian customers amazon might not be able to optimize costs.
- Also, Amazon has no physical store in India.

Opportunities

- Amazon can focus more on its private label product line and increase the number of products under its own brand name rather than becoming a forwarding site for third party products
- Amazon can also increase the portfolio of its product offerings to translate sales into higher revenue.
- Entering in offline-retail market is a biggest opportunity for Amazon India. Company may open more physical stores in emerging markets which allows people to purchase from its physical store.
- For higher market share, amazon can acquire some e-commerce companies of India like shop clues, big basket, etc. This will help the company to reduce competition level in the market.

Threats

- Physical retail stores and local market places are still a biggest threat to amazon in India.
- Increase in people's concern about identity theft and data stealing over online shopping
- In the recent time, increase in Cyber Attacks on big companies' IT infrastructure is also a biggest threat. Amazon need to safeguard its customer data along with its own IT network.
- Due to its aggressive pricing strategies, amazon may have to face lawsuits from its rival in online retail industry.

Conclusion

The SWOT analysis clarifies the current state of Amazon. Some necessary reforms need to be made to reduce scarcity and strengthen its market position. Amazon needs to strengthen its core areas, mitigate its weaknesses, take advantage of opportunities and face threats to future growth.

Some recommendations are given below:

1. Amazon can Strengthen its market dominance by increasing their marketing efforts, promotional activities and competitive advantages.
2. Amazon needs to efficiently manage the features of its app to reduce the negative publicity in the market and attract more customers.
3. Opening physical stores may led to increase its limited presence outside US. This will increase the brand's popularity and market penetration.
4. Increase competitive edges and widen the gap between Amazon and its competitors.
5. To address the issues of fake sales and cyber-crime, Amazon can upgrade technology measures.
6. Amazon can enhance network security systems to protect the rights of consumers.

1.3 OBJECTIVE OF THE STUDY

In an online platform, every single piece of data from your search till your final purchase and after sales experience, can easily be captured. Objective of this project is to find all the possible data points that a retailer can easily collect to improve its business and user experiences. This project will focus on the global leader of E-commerce industry, i.e. **Amazon**, to know how the retailers use the click-stream data and historical purchase data of customers to show customised results on customized web pages on their platform in order to provide their customers best experience with their services. (This project focuses more on managerial applications rather than technical aspects.)

Objective of the Project

1. To study the functions supported by Analytics in E-Commerce Industry
2. To know the ways in which Big Data analytics are helping e-commerce companies
3. To find out trends prevailing in E-commerce industry in India
4. To identify all the possible data points an online retailer can collect and use
5. To study analytics tools and techniques that can be used to attain the objective of maximizing sales (market basket, recommender system, etc.)
6. To find how Analytics can be used to enhance Business
7. To study the Consumer Perspective towards Online Shopping

Scope of the project

This research project will help us study the major changes in the e-commerce industry going about today and it will also help us identify the major reasons behind it.

CHAPTER 2: LITERATURE RIVIEW

"The study of big data analysis in E-commerce" is a research paper compiled by Pavithra b, Dr. Niranjanmurhy M, Kamal Shaker J, Martein Sylvester mani F. The research was carried out in the month of October 2016. Big data here refers to a large amount of data which is being collected by E-commerce companies which can be later on used for evolution of their business. The information that a company receives is in huge amounts the data needs to be processed so that it can be mined out at the time of need. Big data analytics has changed the influence on business, all sectors are benefiting at the most with the help of the information of the data being collected on daily basis.

Many of the retailers use the same system to predict the customers choices and preferences when a customer searches for the products. This way the whole process of shopping becomes even more personalized for the customers. The big data usually consists of three types of data,

1). Structured data

2). Unstructured data

3). Semi structured data

Structured data refers to the basic and nominal information like the name, surname and address of a person. Whereas unstructured data refers to other type of data which gets collected through social media through advertisements. It also includes videos. This data helps E-commerce businesses and retailers that operate online a chance to understand the environment of the market and attract more customers.

The utilization of Big data in E-commerce sector is as-

- The vital role of big data is to give the customer a better experience while using the website, and by trying to satisfy the users' needs by showing them relevant searches.
- To predict a user's interest and behaviours, the predictive analytics is used to showcase required products to the user to satisfy the demands by proper advertisement based on the predictive data.
- It also is used to personalize the user's information like mail id and address in order to increase the rate of conversation.
- By using real time analytics, the prices of products are changed in order to compete better with other retailers.

The Data which is retracted from different sources, application and devices are emerged together into big data. The data which is retrieved from social media such as Facebook, twitter etc, holds data that is posted by millions of users across the world such data is called as social media data. Along with that transportation data, search engine data, stock exchange data are all also retrieved into big data. The data which is collected is of 3 types- structured data, semi structured data and unstructured data.

The author Shahriar Akter and Samuel Fosso Wamba has discussed in the paper about the different types of big data that are used in Ecommerce. The first type is online transactional data which deals with goods selling an providing other services such as Amazon, E-Bay, Expedia etc. where these transactions make use of data which is further broadly classified into categories.

Web Mining leads to some issues when it applies to e-commerce. One of the vital parameter which is used for implementation in retailers are automatic timeouts of user session, they need to be considered as one of the issues because without the time systems the data mining intelligent algorithms can't be applied and the user end applications need to be based the organizations constraints which varies from organization to organization and the type of users. One of the other issues is that the clients lose their shopping carts because of timeouts of the standard time session. Other issue is that there are millions of transactions happening frequently to keep a track of those is a costly exercise. Another issue is designing the user interface form because the form has to be designed user friendly. Mining of data has to be done at the certain level, or else the chances of the result being incorrect is high.

There are many big data analytics tools available like- descriptive analytics, survival analysis, predictive analytics, logistics regression, linear regression, neutral networks and support vector machines are some methods which are already in use. The whole process of marketing can be automated using any one of the digital marketing tools which help in extracting the data for analysis.

Online Stores are using big data for better customer relationship and giving out better services. In this way, big data benefits helps e-commerce companies.

By distributing something valuable, having more personalized item, it helps to build more accurate predictions, it decreases the shopping cart rejection rate, and provide customers with a better experience. For example, Amazon, its online grocery sales were no.1 with 22% share

in 2015 while Wal-Mart is 13%. This helps a company to find out their status and how to work in particular areas for sales growth.

Some of the challenges that big data faces are capturing related information, storage spaces for all data, searching of related information, sharing, transferring, analysis, presentation big data issued widely across the world it can be used to give more accurate analysis and help to give out best decision to compete with other organization. E-commerce can utilize big data to build a better customer relationship and gain customer loyalty. To predict analytics of a customer's choice. Personalization using emails to increase conversations and also by taking expertise opinion the pricing can be changed to fit in the market.

Big Data and E-commerce go hand in hand, both are necessary in decision making for businesses. The use of Big Data has grown in e-commerce. All the functions are used to improve individual business. This paper helps in identifying different applications of big data in e-commerce to let people know the importance of big data. It makes better understanding on the usage of big data and its application as well as its components.

CHAPTER 3: RESEARCH METHODOLOGY

Aim: To study the usage and analyse the popularity of online shopping among the people.

Objectives:

- To find if there any gender significance on mode of shopping preference.
- To find out the most popular online retailer amongst the youth.
- To find out the reason of preference of their favourite online retailer.
- To find out if youth is comfortable with online shopping.
- To find out if youth prefers online shopping over physical shopping.

Operational Definitions

1. **Online Shopping** – It is a form of e-commerce which allows customers to directly buy goods or services from a seller over the Internet using a web browser or a mobile application of the merchant.
2. **Offline Shopping** – It is a traditional way of buying goods and services by directly visiting to the store/shop/ in a market.
3. **Personal Data** - Personal data is any information that relates to an identified or identifiable person.
4. **Physical Store** - Physical store shopping is one of the traditional shopping channels where consumers need to visit physical stores in person to do their shopping activities.
5. **Preference** - It is a greater liking of one alternative over others

Research Hypothesis

- There is no significance between Gender and the preference for mode of shopping.

Explanation

A major change is taking place in the youth's preferences of shopping. They want more ease in purchasing items from home using internet technology that is more easily available and gives them more variety. They are striving for quality products, something that is more durable and available at low prices. Companies design their online stores (Website, Mobile app interface) in such a way that more and more customer would like to use their services and exclusive offers. Product line is also evolving to meet the user preferences. Online shopping has boosted a lot in the past 4-5 years with lots of lucrative discount offers being available on the platform and not at traditional physical stores, all because of Big Data and Machine Learning. This research is to study this change, why is there such a proportional change? do people really prefer to shop online? how popular has online shopping gotten in comparison to shopping in malls or local markets? how many actually like to shop online from these online retailers rather than going to a market for shopping?

3.1 Methodology

- **Area / field-** Delhi-NCR and some other region.
- **Population-** College going students, Working Professionals, etc.
- **Sampling-** Convenient sampling

Sample Size: Sample size taken for the analysis is 100 students from different location of Delhi-NCR and some other states. Out of these students from different region, sample are their family members and relatives.

3.2 Method of Collecting Data

42

The data were collected from both the primary and secondary data sources. The primary data is collected by using the technique of digital questionnaire [Fig 3.1]. The secondary data were collected from different sources given below:

- Annual Report
- Research Papers
- Journals
- Business Magazines
- Company's official website
- Data from open source websites

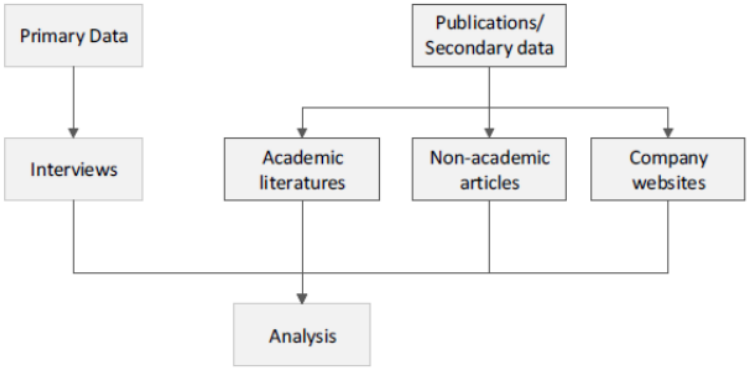


Fig: 3.1 Data Collection Method

3.3 Research Instrument

Questionnaire

A questionnaire is a research tool that consist of a series of questions aimed at gathering information from respondents. The questionnaires is considered as a type of written interview. It can be done face to face in person, over the telephone, computer or by post. Questionnaires provide an inexpensive, quick and efficient way of gathering large amounts of information from a large sample of people. Data can be collected quickly because the researcher does not need to be present while completing the. This is useful for large populations when interviews would be impractical.

Statistical Tools: Tools used for the analysis of collected data are Tables, Bar diagram, Histogram, Pie-chart, Chi-Square test, K-mean Clustering and Python Libraries.

CHAPTER 4: CASE STUDY ON ONLINE RETAIL

As we have seen that the major source of revenue for Amazon is its online business(E-tailing). This is the reason amazon focus more on its online stores and look forward to improve its business with the help of real time data generated through its own online platform, when a user visit or purchase from its website or mobile app. In order to know its customer and remarketing the product, amazon collects some data points and use it further for its business growth.

A case study was undertaken for this project to get the deep knowledge of how online retailers work. The purpose of this case study is to find out the answer of below mentioned questions.

1. What all possible data points online retailer can collect and how do they use it?
2. Which all analytical tools and techniques can be used to attain the objective of maximizing sales?
3. Which tool will help in which way?
4. How retail Analytics can be used to enhance Business for ecommerce companies?

4.1 List of possible data points that Amazon can collect

In an online platform, every single piece of data can easily be collected. Amazon also collects data which could be used to improve its business and user experiences. Some of the possible data points that amazon collects in the background are as follow:

1. View History
    - to recommend the similar type of products to the visitor
    - to recommend the most visited item by others
    - to track conversion rate: what you have viewed and purchased

2. At what time you visit the website; when do you view?
    - to know at what time you become more active on their sites
    - to give promotional offers at the busiest time of a day to get easily noticed

3. Search Data
    - to know specific item, you are looking for

4. Transaction Data
    - date of purchase

- item purchased
- day and time of transaction
- mode of payment

5. Personal Data
    - Name, Age, etc.
    - your shipping addresses
    - contact number, etc.

6. User Experience Data
    - your review information (star rating, comments, etc)
    - to your sentiments and user experience

7. How much time or number of days a customer takes to write a review?

8. How are you likely to click on the link in the Amazon's promotional email?
    - to figure out an alternate or a better way to send customised mails

9. Information regarding Prime Membership of Amazon
    - In order to compare the user experience between prime and non-prime customers

10. Information related to the mode you prefer to access the website
    - For example: Phone or Laptop, Website or Mobile App
    - to know the user's browsing pattern

11. For a Seller Account
    - Information of items sold by the seller
    - Customer's review/feedback information and use it for seller ranking

12. How does a user access the website?
    - Type the website link e.g. "amazon.in" directly or redirecting from an online add or a link

- to know the major source of traffic coming to the online store

4.2 Tools and Techniques that Amazon Uses

4.2.1 Amazon Recommender System

At some point, it might have come to your mind that how these online shopping sites like Amazon recommends you to buy the product. The simple answer to this is Big Data and Analytics. Amazon uses its customer's data to recommend the products to its customers. So, how does Big Data, Analytics and Machine Learning help the recommendation system of amazon?

Basically, there are 3 stages in the entire recommendation process. Three stages of recommendation are as follow:

1. Event
2. Rating
3. Filtering

1. Events: Amazon tracks and stores data on all customers' behaviours and on-site activity. A record of the event is entered in the database after each and every click made by the shopper. The entry is stored as something like "User A clicked product X details once". Events are captured for all kind of actions e.g.

- Adding product to the cart
- Purchasing a product
- User liking a product

2. Rating: Ratings are important because they reveal what a user feels about a Product. Recommendation system can assign implicit values on different kinds of user actions. The maximum rating is 5. For example

- Purchase - 5 stars,
- Click - 4 stars
- Like - 3 stars,

Recommendation system can also take into account rating and feedback users provide. For example, "I like the product", "Not so good", "So Happy", "Poor Quality", etc.

3. Filtering: Filtering products based on rating and other user data**.** Recommendation system uses three type of filtering

1. Collaborating Filtering
2. User-Based Filtering
3. Hybrid Approach

Collaborating Filtering: In Collaborative filtering the visitor's choice is compared and they get a recommendation. For Example: If user 'X' like Product A, B, C and D & user 'Y' like products A, B, C, D and E. Then it's likely user 'X' will also like Product E.

User-Based Filtering: In user-based filtering, the users browsing history, likes, purchase and rating are taken into account before providing recommendation.

Hybrid Approach: Many companies also have a hybrid approach which consist of both the above-mentioned approaches.

This is how Amazon's Recommender system works and big data generated helps shoppers have a great time at Amazon.

4.2.2 Churn Model

- Churn model helps in identifying the customers with the most probability of switching to other e-commerce platforms.
- Churn rate helps in calculating the percentage loss in customers and in total business, etc.

How it helps-

In a survey it has been found that 80% of revenue comes form 20 % of company's loyal customers. Thus, customer retention is so important for a company a company like Amazon for its growth and business expansion because in e-commerce, word of mouth marketing plays a crucial role for the online retailer.

4.2.3 Customer Predictive Lifetime Value (LTV) Modelling

- The lifetime value of customer is the prediction of the net profit that a customer is expected to bring to the company.
- Its purpose is to treat customers to buy anything to predict their future activities.
- This model collects, categorize and clean up the data of customer's purchases, needs, etc.

- After pre-processing the data, data science algorithm is used to identify interdependencies between customer choice and behavior by assuring better understanding of customers.

How it helps-

It helps a company in formulating strategies as per its customers' behaviour like up cost for customer's purchase.

4.2.4 Customer Sentiment Analysis

- The simple way to use the customer sentiment analysis tool is by collecting feedbacks of different customers from website or from different social media platforms.
- The brand customer sentiment analysis is done based on Text Analysis, NLP, NPS, etc.
- The feedback data from different sources like social media post, feedback forms, surveys, etc. are used for extracting sentiments out of the post to analyse.

How it helps-

Manual feedback has become an old-fashioned and time-consuming way to know people's sentiments. Digital Platforms like social media sites are the new source for the retailers to get feedback from customers and analysing customer's post on social media site using Data Analytics, Machine Learning and Artificial Intelligence.

4.2.5 Fraud Detection

- Being a Digital Store, security is a major concern for the e-commerce companies.
- AI and Artificial Neural Networks (ANN) are used to detect fraudulent patterns to protect the company from fraudsters.

How it helps-

This technique is being used to detect a fraud and prevent occurring new frauds as preventing a fraud is the biggest challenge in front of e-commerce companies like Amazon.

4.2.6 Inventory Management

- Lack of inventory may be a hindering factor to retain customers. Therefore, it is necessary to meet customers' needs on time.
- Predictive analysis, forecasting and machine learning algorithms helps in inventory management.
- Analysis helps detect patterns in the most demanding parameters that defines inventory strategies using ML algorithms.

How it helps-

The maintenance in the supply chain has become very complex in the today's market. Using inventory data analytics, problem of inventory shortage at peak time can be avoided.

4.2.7 Market basket analysis

- It is a modeling technique that defines that if a customer buys something, he can also buy another thing.
- Here the algorithm works by association rule mining and in this way, it identifies a customer's product basket and product association rules.
- Apriori algorithm is best suited to identify the items purchased frequently.

How it helps-

E-commerce websites always encourage its customers to buy more, using recommendation engine that offers most relevant products in display which uses the market basket analysis outcome.

4.3 Use of Analytics to grow business

Ecommerce retailers can be benefited from big data. In today's time E-commerce companies building more advance Machine Learning models that uses big data and implementing it in their websites to provide better services to the online shoppers. Personalized recommendations, offers and additional discounts. Having a large customer base including more than 100 million Amazon Prime members, amazon uses big data about its customers to predict customer purchases, as well as make personalized recommendations of offerings and to optimize the supply chains. It also uses big data to offer daily and weekly promotional offers during festive seasons like "Great Indian Festival Sales". Amazon's website models are based on machine learning algorithms that collect data from each transaction contributing to bigger the training data set. "*Bigger the training data set, Better the algorithm's learning*" that later used to translate predicted sales.

Customer Relationship Management: An Integrated Customer Relationship Management Database is critical to business success, so many retailers Investing in building such a database. Retailers are also focused on analytical tools and techniques to figure out new sources of revenues to optimise its business and profitable CRM. In today's time, to take full advantage of big data in online retail sales environment CRM strategies should be time, location channel and customer specific

CHAPTER 5: DATA ANALYSIS

5.1 Pilot Study Report

The pilot study was conducted on 30th April 2020. The digital questionnaire was distributed through a messaging app "WhatsApp" for the most optimal results. In the span of the 2.5 hours

15 responses were recorded. There were no problems in understanding the questions but one issue did come up. I realized that a 'both of the above" option had to be included for the ones who equally like to go for online and offline shopping or the questionnaire wouldn't be submitted. So, I added the 'Both of the above' option in the following questions:

- What is your preferred mode of shopping?
- What is your main concern for not preferring online shopping?

This allowed the participants to clearly input their choices without any errors from the survey. It allowed us to not lose out the collection of precious information and analysis an honest answer sheet from our participants.

5.2 Execution

I started the execution of the research on 2nd of May 2020. With the pilot study. The questionnaire was disturbed through a messaging software for optimal results. Within two and half hours of the distribution the form there were 14 responses recorded. With the answers were recorded it was found that there was no problem in understanding the question, however I found out an error in the form I hadn't put in the 'others' option. which didn't help the samples to answer fully. Later I rectified the mistake which then individually updated every form. This made the questionnaire a lot more approachable.

At the end of the first day I only had 20 responses. The following day I had to forward the link again and asked people to fill, by the end of the second day I had reached up to 66 responses. I had sent the questionnaire in every one of my college groups. I also requested my friends to forward it to their friends to gain more response.

At first there were not many responses, but at the beginning of third day I had 102 responses. I knew in the beginning that collecting responses would be hard and would also be time consuming but at the same time I were also determined to get in the responses as soon as possible which would help me to analyse faster, and would also give me more time for the analysis process. I have learned one this that is initially the responses come in slowly but at after a day or two it is easy to reach your mark.

5.3 Primary Data Analysis

1. Chi-Square Test

Gender	Online	Offline	Both (Onl	Total
Male	40	7	15	62
Female	2	2	34	38
Total	42	9	49	100

Hypothesis: For this instance, the Null and Alternative hypothesis are as follow,

H_0: Gender and Shopping Mode Preferences are independent

H_1: Gender and Shopping Mode Preference are not independent

For this analysis, the level of significance is 0.05 and the test used is *Chi-Square Test*.

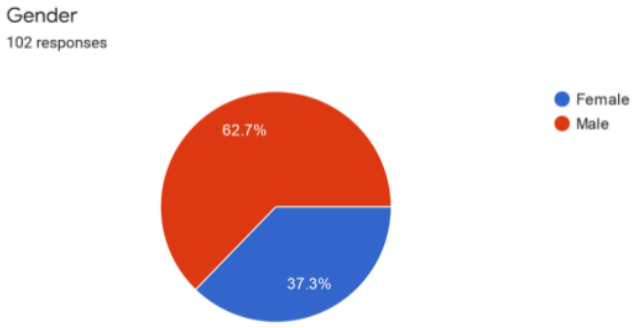
Results						
	Online	Offline	Online & offline			*Row Totals*
Male	40 (26.04) [7.48]	7 (5.58) [0.36]	15 (30.38) [7.79]			62
Female	2 (15.96) [12.21]	2 (3.42) [0.59]	34 (18.62) [12.70]			38
Column Totals	42	9	49			100 (Grand Total)

The chi-square statistic is 41.1355. The *p*-value is < 0.00001. The result is significant at $p < .05$.

Fig: 5.3.1 Chi Square Test result

Interpretation: Here, we observed that the calculated P-Value 0.00001 is less than the significance level of 0.05. So, we reject the null hypothesis. Therefore, we conclude that there is a relationship between Gender and shopping mode preference.

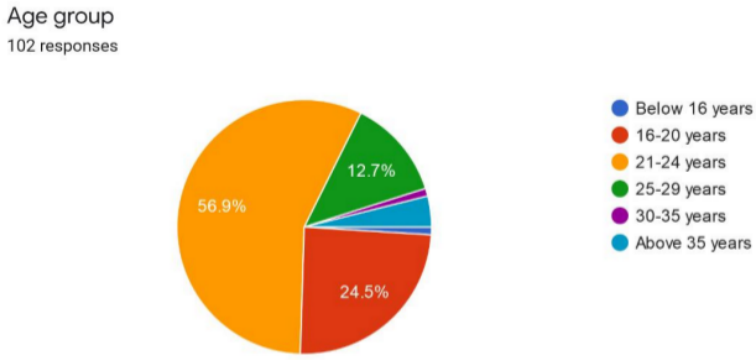
2. Gender



ANALYSIS:

The survey was distributed among many people, out the 102 responses received, 62.7% responses were of males and the rest of 33.7% were recorded by females.

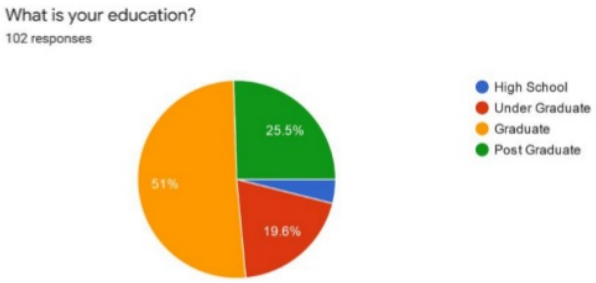
3.



ANALYSIS:

The population for the research were people between the age group 16 to 30. The reason behind choosing this age group was as they have internet and are also independent enough to shop on their own. Which increases the scope of knowing what the preferences of the people are. Out of the responses received, the majority of the answers were from the age group 21 to 24 years- 56.9%. People from age group 16 to 20 also responded, up to 24.5% people fall in this category. The rest of the responses were from 25 to 29 years- 12.7% and up to 5% people were above 35 years. About 1% people were from below 16 years of age group and above 35 years.

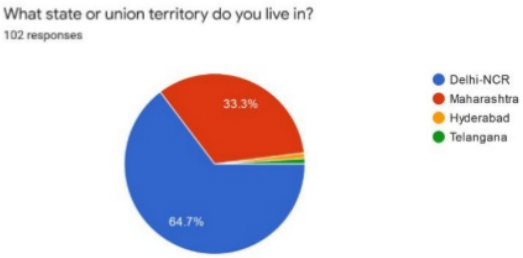
4.



ANALYSIS:

Out of the whole population, majority of the people were under the graduate category being 51%. Followed by the second category 25.5% people had completed their post-graduation.19.6% people had been still at the under graduate level. And the rest of the sampling consisted of high school students.

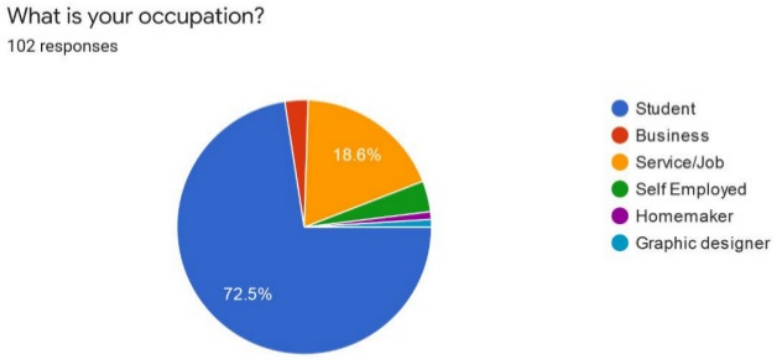
5.



ANALYSIS:

My hometown being Delhi- NCR, more than half responses were taken from the same area about 64.7% of the total population. The rest of the 33.3% responses were from Maharashtra. Remaining of 1% were taken from Hyderabad and Telangana.

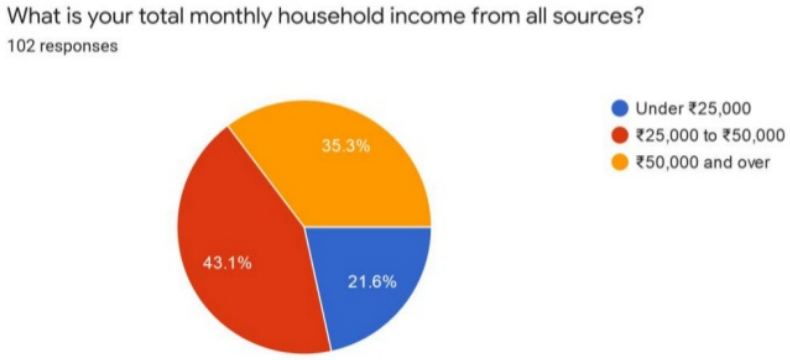
6.



ANALYSIS:

As the age group of the samples were from 16 to 30 years old, majority of these people are still studying. More than half of the people who responded were students – 72.5%. the rest of the 18.6% were working in a service or were doing job. The rest of the responses were from people who were self-employed, had businesses, were home makers, and also graphic designers.

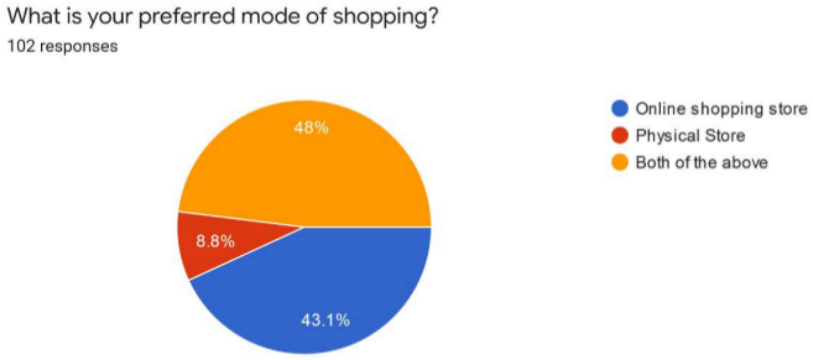
7.



ANALYSIS:

When the question was asked to the audience about their family income, 43.1% people had their monthly household income between Rs.25,000 to Rs. 50,000. The next category were people whose monthly household income were from Rs.50,000 and above. -35.3%. The rest of the 21.6% had their monthly household income under Rs.25,000.

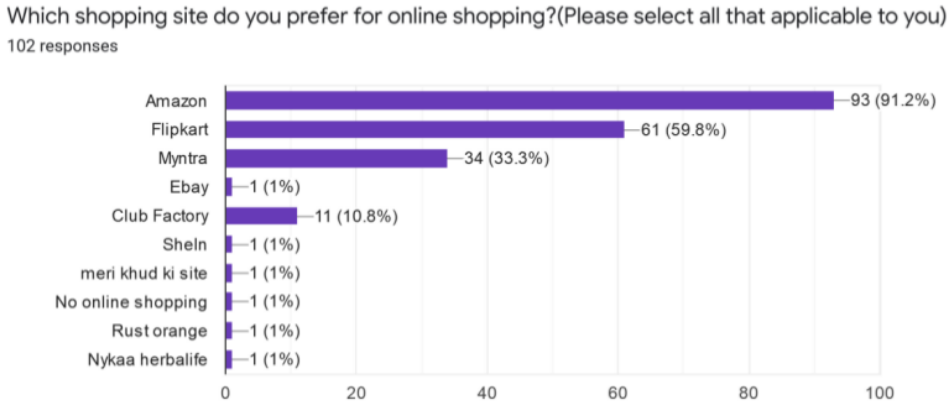
8.



ANALYSIS:

According to the findings, it was known that around 43.1% people prefer preferred shopping online. Whereas 8.8% people preferred to shop in a physical store. But almost half of the people preferred both shopping online as well as shopping in a physical store the percentage of the people were 48%.

9.



ANALYSIS:

It was important to know which shopping site did they prefer for shopping; people could select multiple options for their preferences. According to my findings, 91.2% people preferred Amazon. Whereas 59.8% people preferred Flipkart. Followed by Myntra which was 33.3%.

club factory was next with 10.8%. Only 1% people preferred sites like EBay, Shein, Rust Orange, Nykaa or Herbal life.

10.

ANALYSIS:

To know more about the reasons behind why people preferred to shop online. The Audience could select all the options that were applicable to them. Around 85.3% people said that online sites had wide range of products. 76.5% people said that they preferred to shop online due to the discounts available on the e-commerce sites. Around 62.7% people said that they liked the exclusive offers the sites had to offer them. 57.8% people said that they liked that they didn't actually have to go to a shop to buy stuff, all the shopping could be done by sitting at home. 54.9% people said that easy payments were the reason they preferred to shop online. The rest of 16.7% people stated that the liked the personalised emails and messages a brand sends to them which in return attracts them to shop online.

11.

ANALYSIS:

This is to know which platform do they preferred to shop online. 47.1% people said they preferred to shop using mobile app. Whereas 47.1% people stated to prefer desktop or laptop. The remaining 5.8% people stated to prefer to use phones browser.

12.

ANALYSIS:

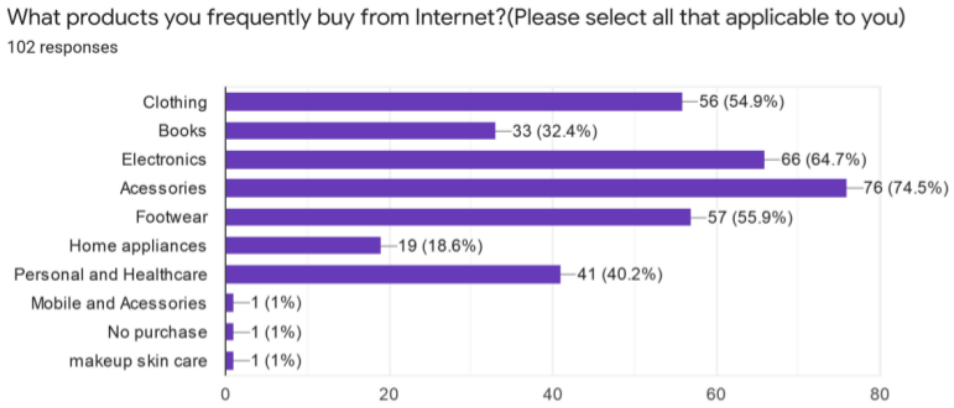
When asked the people about their preferred mode of payment when shopping online. Majority (43.1%) of people stated to be using cash on delivery. 33.3% said they prefer debit cards. 14.7% people said to be using mobile wallet like (Paytm, Google pay, etc.). the remaining crowd preferred UPI payment and credit card.

13.

ANALYSIS:

The next question was to know if people saved their card details in their personal shopping accounts. According to my analysis of the response received was that 42.2% people had saved their card details. Whereas 41.2% people said they hadn't saved their card details. Around 16.7% people said they weren't sure of it.

14.



ANALYSIS:

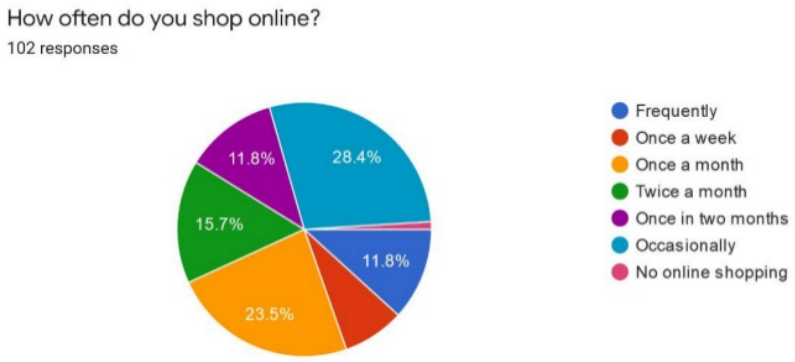
The next question asked was about the products that they preferred to buy from online sites. The majority of the audience said that they usually bought Accessories, Electronics, Footwear, clothing, personal and healthcare, books and home appliances in the order of preference. very few people purchased mobile and accessories, makeup and skin care.

15.

ANALYSIS:

It was equally important to know the reason behind why the sampling chose a particular brand for shopping. The question was asked as a multiple option question. The Reponses received stated that 81.4% people shopped in a brand because of its name and establishment. 56.9% people said that they often chose a brand based on the specials offers they had on their sites. Whereas 52% also said that personal recommendation also plays an important role in selecting a brand. Not only that 44.1% people stated that price aggregator sites was also important. 37.3% people said that social media remunerations were also helpful in selecting a brand. 35.3% people said they choose a brand based on its social media advertisements. 15.7% people chose a brand based on the search engine and 1% people said that they chose a brand based on the reviews it had.

16.



ANALYSIS:

According to analysis of the response, 28.4% people stated that they shopped frequently. 23.5% people stated they shopped once a month. 15.7% stated they shopped twice a month from online sites. 11.8% people stated they shopped twice a month. And the remaining stated that they shopped once a week.

17.

ANALYSIS:

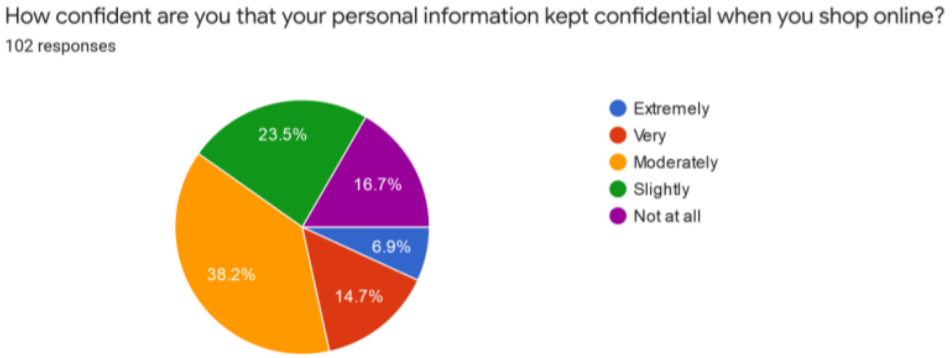
The next question to the people was to know how many products they bought at a time from an Online store. The responses were such that 41.6% people bought two or more product at a time. Whereas 33.7% people bought two products at a time. 24.8% people stated that they only bought one product at a time.

18.

ANALYSIS:

The next was to find out whether people found the recommended products on a site to be attractive or no. 59.8% people said that they sometimes found the recommended products to be attractive. 27.5% people state that they always find recommended products attractive. Whereas only 12.7% people said they never like recommended products.

19.



ANALYSIS:

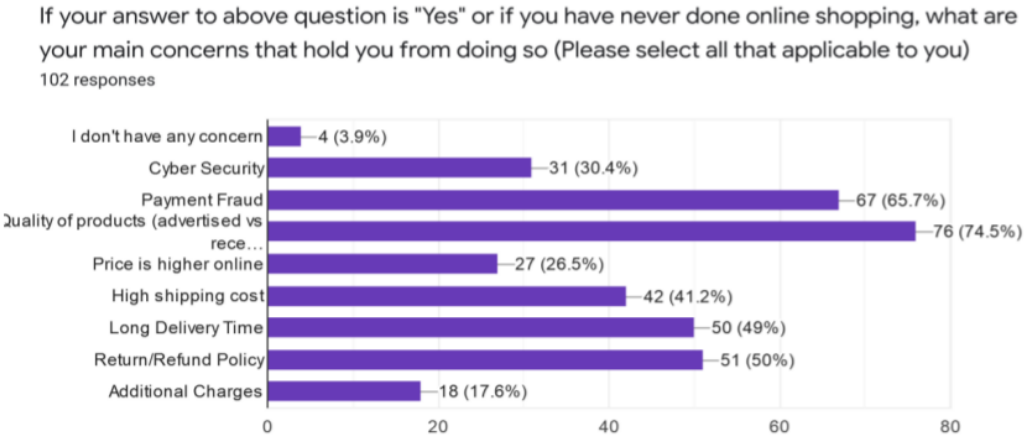
38.2% people were moderately confident that their personal information will be kept confidential while shopping. 23.5% people state they were slightly confident. 16.7% people are not at all sure of the safety of their personal details. Whereas 14.7% and 6.9% are very and extremely confident about the confidentiality of their personal details.

20.

ANALYSIS:

While asking the people about the fears or doubt they might have had while shopping, 8.3% people agree to have fear/ doubt while shopping online. Whereas 16.7 people state they have no fear while shopping online.

21.



ANALYSIS:

After coming to that people had fears/ doubts while shopping online, the next question was to know more about the people's fear when they shop online, the people were asked a few common doubts a person has while shopping. In responses t these options. 74.5% people said they had doubts about the quality of the products, that is they would receive something different

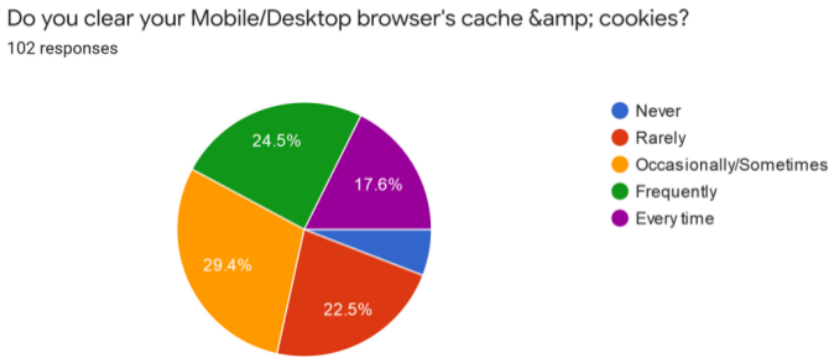
than what they saw or ordered. 65.7% people said they feared payment fraud. 50% people said they had doubts about a company's return/ refund policy. 49% people had issues with long delivery time. 41% people had issues with additional and high shipping cost the company imposed. 30.6% people had concerns with cyber security. 26.5% people stated that websites actually had higher prices than normal. 17.6% people had issues with additional charges and 3.9% people stated they had no concerns.

22.

ANALYSIS:

To track a person's overall experience while shopping online, 44.1% people said to have a moderately satisfied experience. 40.2% said they were very satisfied. The amount of people having and extremely satisfied experience was very less only 9.8%.

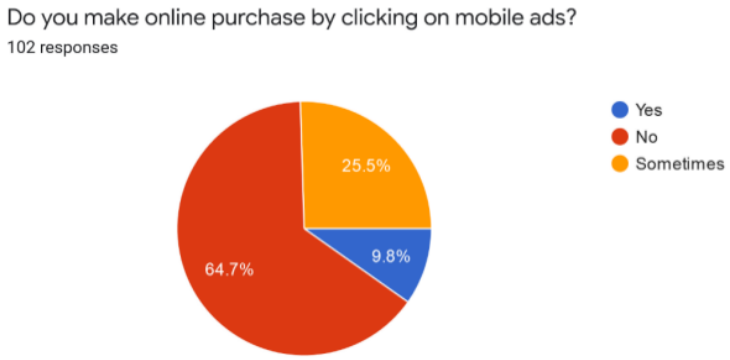
23.



ANALYSIS:

To know more about their usage of intent. 29.4% people stated the occasionally/ sometimes cleared their mobile/ desktop's cookies and cache. 24.5% said they cleared everything frequently. 22.5% people said the rarely cleared everything. Whereas 17.6% people said they cleared everything every time.

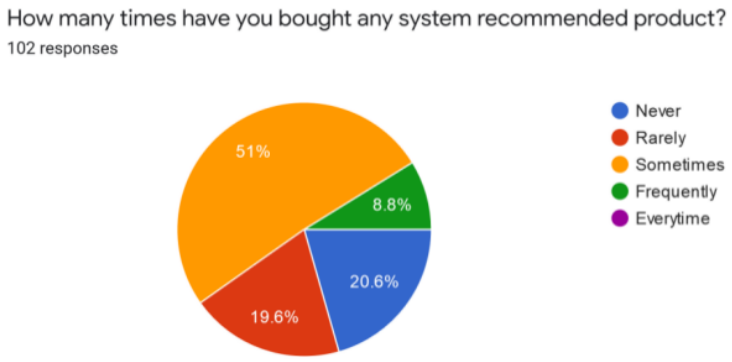
24.



ANALYSIS:

The analysis of this question said that 64.7% did not purchase items by clicking on mobile ads. Whereas 9.8% said to have purchased item. 25.5% said they sometimes have bought stuff by ad.

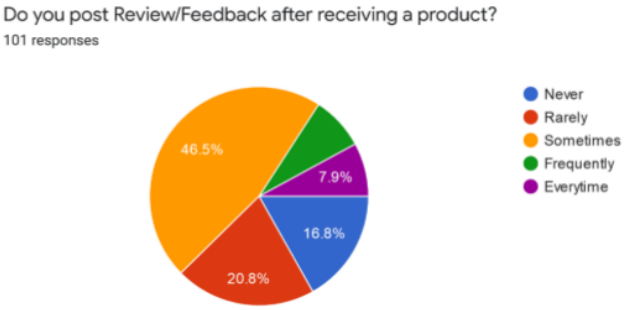
25.



ANALYSIS:

System recommend products are a way to knowing how much influence a brand has on its customer. 51% people said to have bought system recommended products. 20.65 people said they have nerve bought a recommended product. 19.6% people said they rarely buy anything system recommended. Whereas 8.8% said the frequently bought system recommended products.

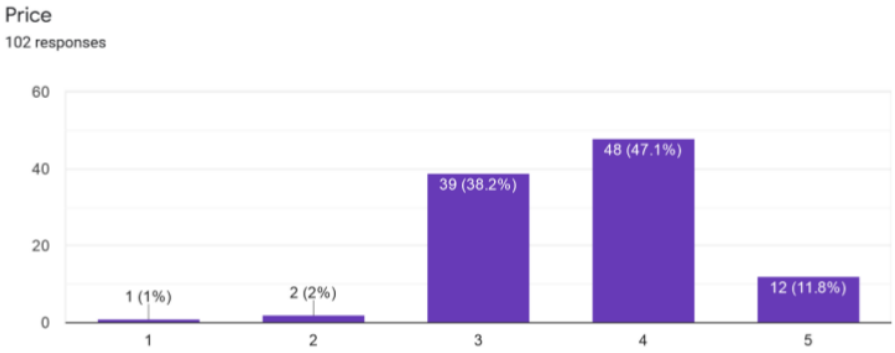
26.



ANALYSIS:

According to the analysis of this question, 46.5% people have stated they sometimes post a review online. 20.8% said they rarely posted an review. 16.8% said they have never posted a review about the products. Whereas 7.9% said they post reviews every time after they purchase a product.

Please rate the satisfaction level for the item you frequently buy online

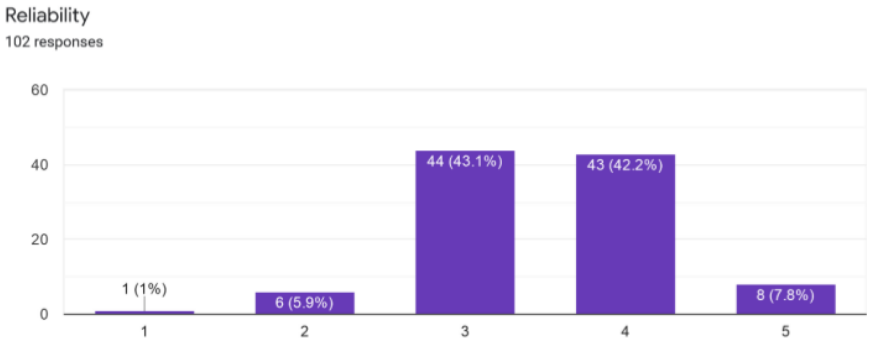


27.

ANALYSIS:

In context to the prices online, 11.8% people said they are extremely satisfied with the prices. 47.5% they are very satisfied with the prices. 38.2% said they are somewhat satisfied with the prices online. 2% said they are moderately satisfied. Whereas 1% said they are not at all satisfied with the prices.

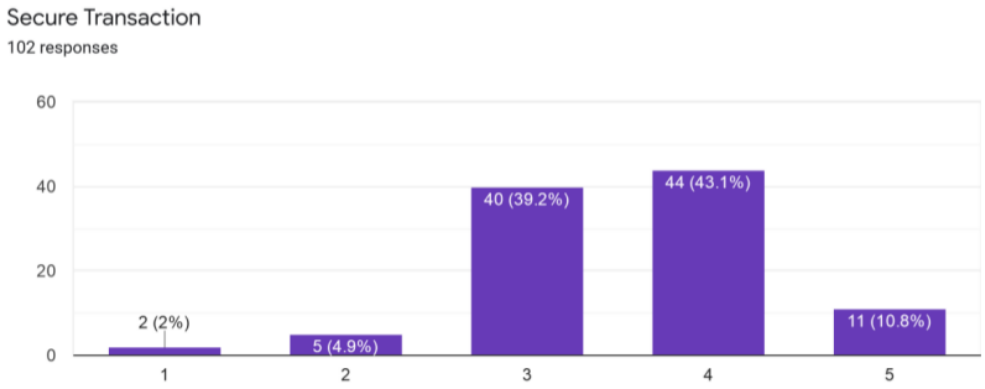
28.



ANALYSIS:

To know the reliability of the products that people often buy online, 7.8% people said that the products that they bought were extremely reliable. 42.2% said the products were very reliable. 43.1% said the products were somewhat reliable. 5.9% said the products were moderately reliable. Whereas 1% said the products were not at all reliable.

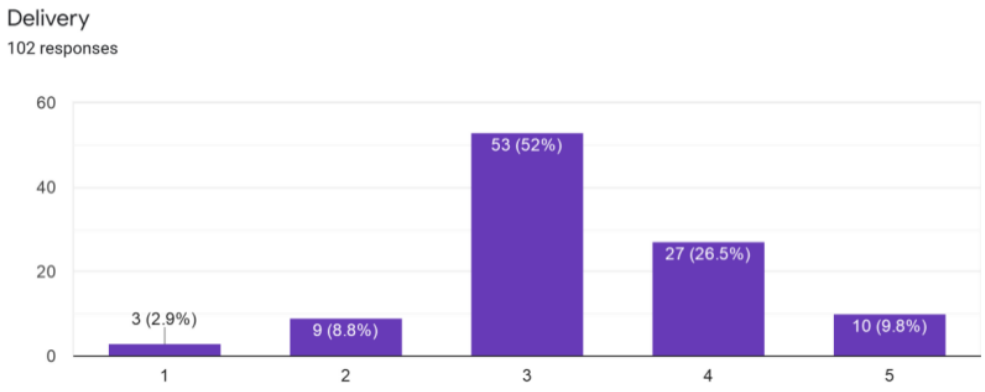
29.



ANALYSIS:

To know the people's security while doing transactions. 10.8% people said there felt extremely secure with their transactions. 43.1% said they felt very secure with online transactions. 39.2% said they felt somewhat secure. 4.95 said they felt moderately secure. Whereas 2% did not feel secure at all.

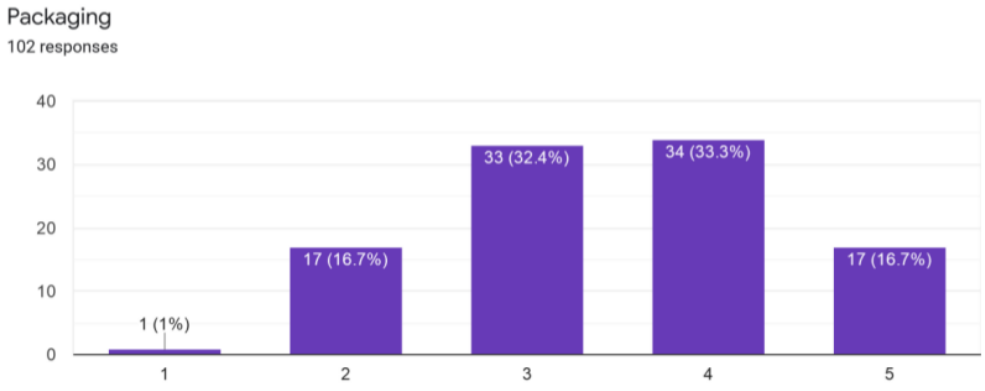
30.



ANALYSIS:

To know the delivery of the products. 9.8% were extremely happy with the delivery. 26.5% were very happy with the delivery. 52% were somewhat happy with the delivery. 8.8% were moderately happy with their delivery. Whereas 2.9% were not at all happy with the delivery of their products.

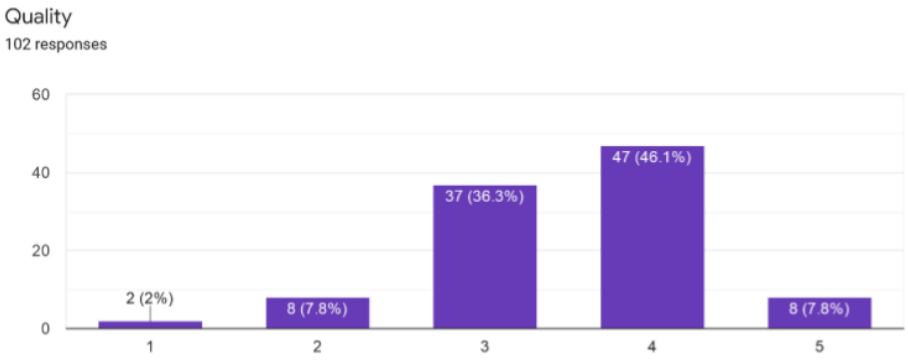
31.



ANALYSIS:

16.7% people were extremely happy with the packaging of their products. 33.3% were very happy. 32.4% were somewhat happy with the packaging. 16.7% were moderately happy with packaging. Whereas 1% felt that they were not at all happy with the packaging.

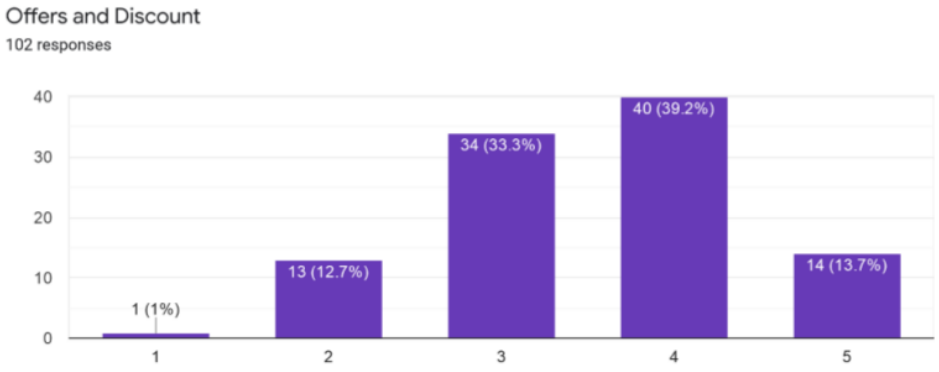
32.



ANALYSIS:

7.8% people were extremely happy with the quality of the product. 46.1% were very happy with the quality. 36.3% were somewhat happy with the quality, 7.8% were moderately happy and 2 % were not at all happy with the quality that they received.

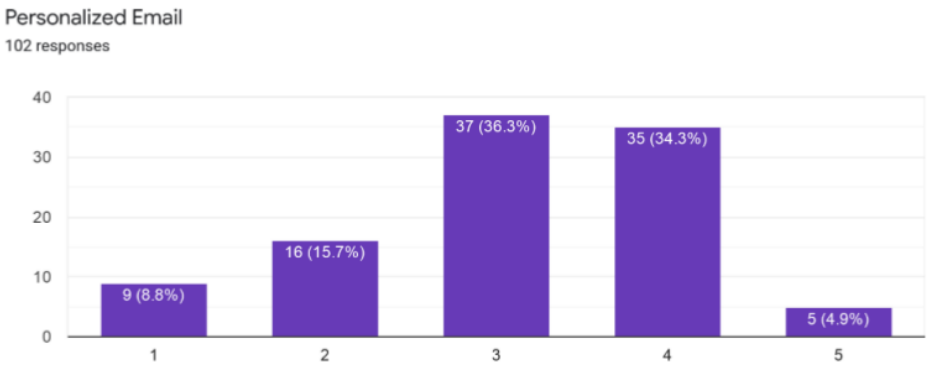
33.



ANALYSIS:

The offers and discounts that each site offers is usually what attracts the customer in the first place. 13.7% people are extremely happy with the discounts and offers they get on online websites. 39.2% are very happy with the discounts. 33.3% are somewhat happy with the discounts. 12.7% are moderately happy where as 1 % are not at all happy with the discounts and offers being given by the company.

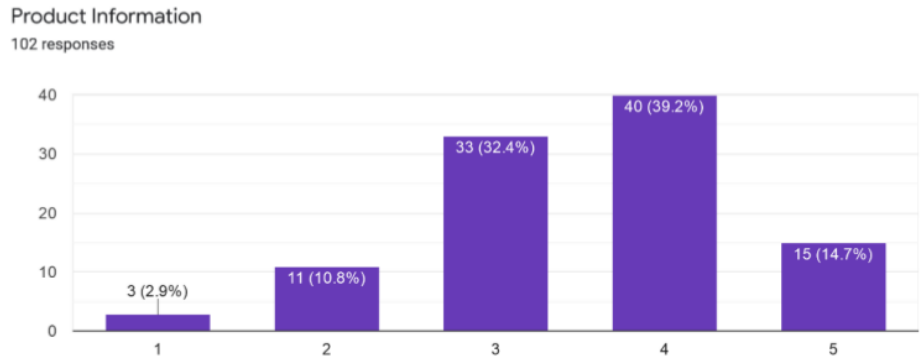
34.



ANALYSIS:

4.9% people said they extremely liked the personalized emails they get from a company. 34.3% said they very much like the emails. 36.3 % said they somewhat like it. 15.7% said they moderately like it. Whereas 8.8% said they did not like the emails at all.

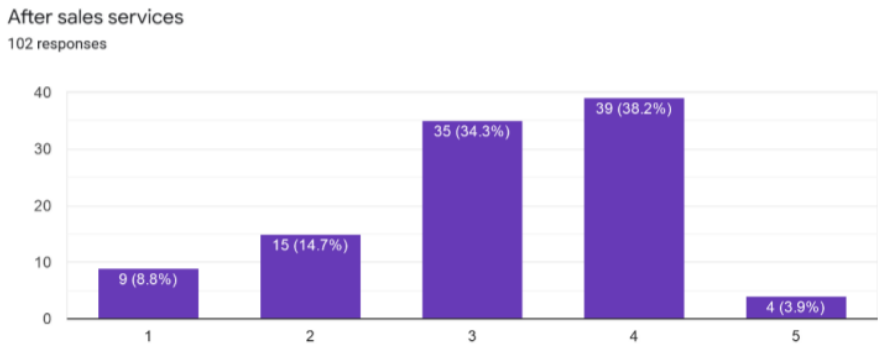
35.



ANALYSIS:

When a product is being put up, there's a certain information that is put up along with it. It is important to know if the user find that information useful or no. 14.7% people find the product information to be extremely useful. 39.2% find the information to be very useful. 32.4% find the product information to be somewhat useful. 10.8% find the information to be moderately useful, whereas 2.9% find that the information is not at all useful.

36.



ANALYSIS:

To know the rate of satisfaction of the after-sale services a person had after a purchase. 3.9% said they were extremely satisfied with the after sales service. 38.2% said they were very satisfied with the services. 34.3% said they were somewhat satisfied. 14.7% said they were moderately satisfied. Whereas 8.8 5 said that they were not at all satisfied with the company's after sale services.

37.

ANALYSIS:

The last question i asked was to know if the people were aware that their information could be easily collected and used when they go and shop online, to which the responses were that. 84.2% people were aware of this fact. Whereas 11.9% people were not aware of this fact. The rest of the people thought it didn't matter.

38.

Gender	Two	One	Two+	Total
Male	24	19	18	61
Female	10	6	22	38
Total	34	25	40	99

Analysis: As per this graph, most of the participants(40% approax) buy more than 2 products in a single transaction.

5.4 Customer Segmentation Using Python

Customer Segmentation: Customer Segmentation is a technique which is used by the retailers for dividing their customers into groups, based on the similar characteristics among them. Segmentation is used for making different marketing strategies to target similar group of customers. These groups are also known as "clusters".

In this section, analysis is being done for customer segmentation based on the customer's spending score. Due to the unavailability of spending score for all the survey participants, I have assumed and calculated the spending score manually(using the available data points like their monthly income, number of items purchased in single transaction and their buying habits).

Code:

```
In [4]: import pandas as pd
        import numpy as np
        import sklearn
        import matplotlib.pyplot as plt
        import seaborn as sns
        import seaborn as sns
        %matplotlib inline

In [5]: from sklearn.cluster import KMeans
        from sklearn.decomposition import PCA
        from sklearn.preprocessing import MinMaxScaler

In [6]: import os
        print(os.listdir("C:/Users/Rohit Singh/Desktop/Python Analysis"))

        ['.ipynb_checkpoints', 'e1.ipynb', 'n9.csv']

In [7]: import extra_graphs

In [8]: df = pd.read_csv('C:/Users/Rohit Singh/Desktop/Python Analysis/n9.csv')

In [9]: df.head()

Out[9]:
```

	S.NO.	Gender	Age group	What is your total monthly household income from all sources?	Score
0	1	Male	23	75	80
1	2	Male	25	44	50
2	3	Female	19	48	79
3	4	Female	17	43	5
4	5	Female	15	23	28

Observation: For the customer segmentation, first of all we need to import the data set. Here, a customer dataset with a file name 'n9.csv' is used which contains 5 variables:

- S. No: It is used as a unique customer ID.

- Gender: Customer's Gender.
- Age: Age of the customer.
- Monthly income: Monthly income of the customer.
- Score: It is the customer spending score based of their buying behaviour (assumed).

Data is not in the proper format so first we need to clean the data. We are using Matplot lib, Scilit-learn, K-Means algorithm and PCA for this analysis.

Data Pre-processing

Renaming the columns for easy interpretation.

```
In [10]: df.rename(columns={"S.NO.":"CustomerID","Age group ":"Age","What is your total monthly household income from all source
```

```
In [11]: df.head()
```

Out[11]:

	CustomerID	Gender	Age	Salary	Score
0	1	Male	23	75	80
1	2	Male	25	44	50
2	3	Female	19	48	79
3	4	Female	17	43	5
4	5	Female	15	23	28

Column's name have been changed

Exploratory Data Analysis

Here, we check for the missing value in the dataset.

```
print(f"Checking for missing values:\n{df.isnull().sum()}")
```

```
Checking for missing values:
CustomerID    0
Gender        0
Age           0
Salary        0
Score         0
dtype: int64
```

Observation: There was no missing value found in the data set.

Next, we check for the Duplicate rows.

```
print(f"Checking for duplicate rows: {df.duplicated().sum()}")
```

```
Checking for duplicate rows: 0
```

Observation: There was no duplicate row found in the data set

Next, check for the data type of all the variables in the dataset.

```
: print(f"checking for data type \n{df.dtypes}")

checking for data type
CustomerID     int64
Gender        object
Age            int64
Salary         int64
Score          int64
dtype: object
```

Observation: Out of 5 variables 4 were already in integer format. Later, we change the data type of variable "Gender".

Statistical Analysis

Here is the statistics for the variable 'Score'.

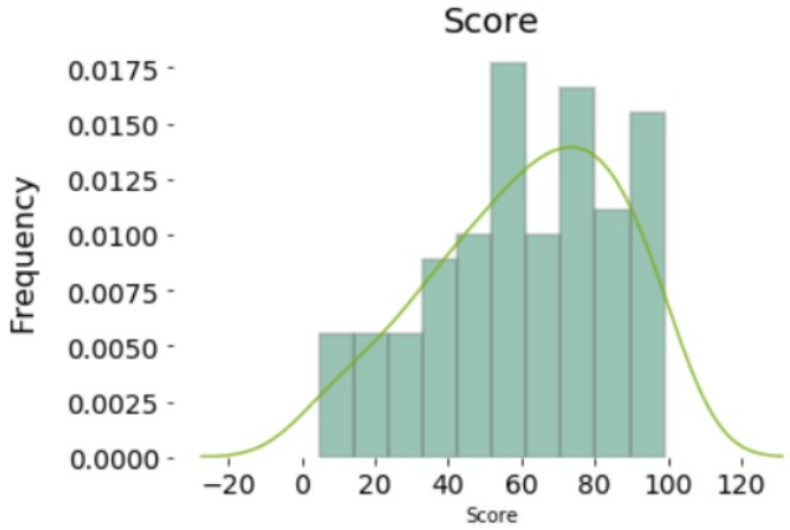
```
: spending_score = df["Score"]

: statistics(spending_score)
```

:

Variable	Mean	Standard Deviation	Median	Variance
Score	61.260417	25.173252	63.0	633.6926

Observation: Mean score value is 61.2 and the Median is 63.0. where as the SD is 25.17 and Variance is 633.67.

```
graph_histo(spending_score)
```

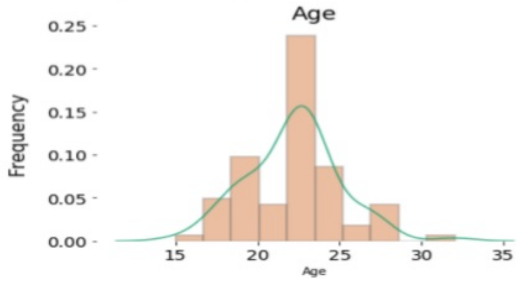


Observation: This is the frequency distribution of Score. Most of the spending score lies between 50-60.

```
age_stats = df["Age"]
statistics(age_stats)
```

Variable	Mean	Standard Deviation	Median	Variance
Age	22.197917	2.808543	22.0	7.887912

```
graph_histo(age_stats)
```

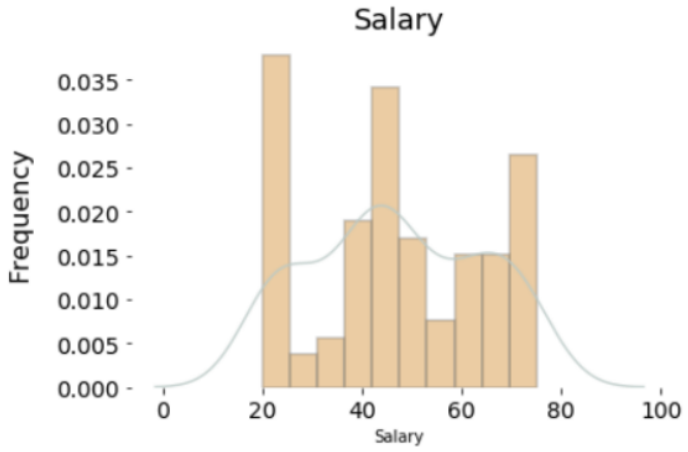


The average age in the data set is 22.19. As we have seen in the previous pie chart analysis, that most of the participants were ranging from 19 to 25.

```
salary = df["Salary"]
statistics(salary)
```

Variable	Mean	Standard Deviation	Median	Variance
Salary	46.875	16.76321	45.0	281.005208

```
graph_histo(salary)
```

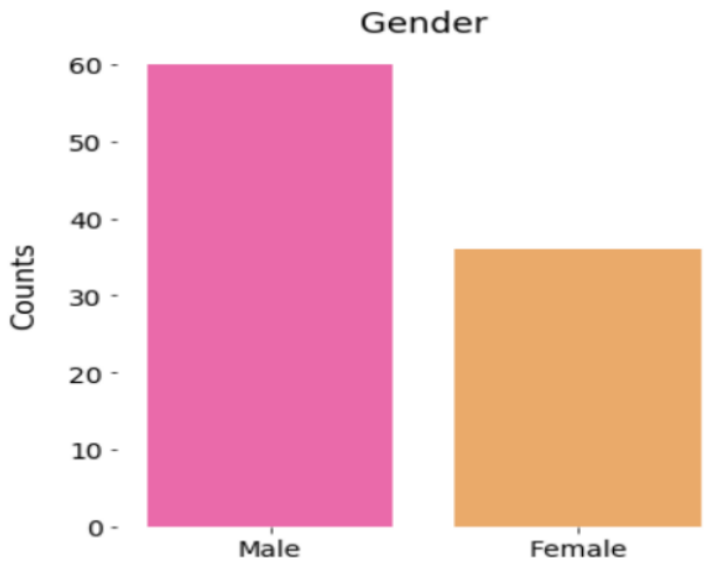


Observation: The average monthly salary is 46.87(thousand rupees).

```
gender = df["Gender"]
statistics(gender)
```

	Gender
Male	60
Female	36

```
graph_histo(gender)
```

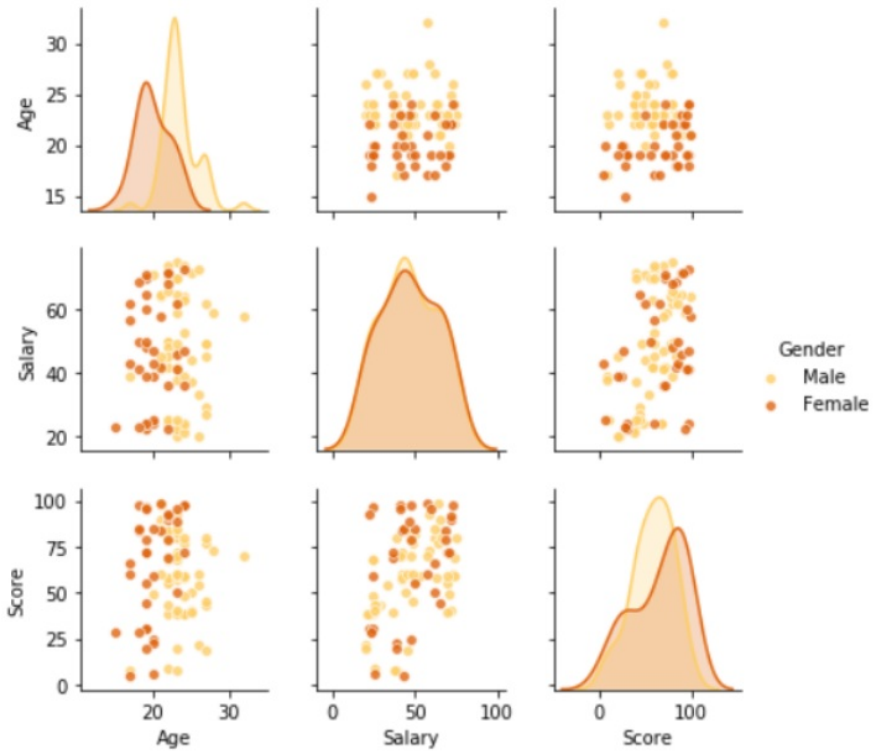


Observation: After cleaning the data and removing the outliers, this data set contains the data of total 95 customers. Out of 95, 60 customers are male and 35 customers are female.

Correlation Check

To check the correlation between 2 variables, pair plot is used here.

```
sns.pairplot(df, x_vars = ["Age", "Salary", "Score"],
                 y_vars = ["Age", "Salary", "Score"],
                 hue = "Gender",
                 kind= "scatter",
                 palette = "YlOrBr",
                 height = 2,
                 plot_kws={"s": 35, "alpha": 0.8});
```



So far, we observed that all the variables in the data set were normally distributed. Also, no correlation between two variables were found. There is no pattern in the above pair plot graphs thus, we conclude that there is no correlation between any of the variable. We can see that age and score is not the right components for the clustering so we choose the Salary and Score for clustering of customers.

Now, we use the Principle Component analysis to find out the main components in the dataset. We start with transforming the categorical value (Gender) into binary form.

PCA: Principal Component Analysis

```
Here i have assigned numerical values to the Gender; 1 for True and 0 for False
```

```
df["Male"] = df.Gender.apply(lambda x: 1 if x == "Male" else 0)
```

```
df["Female"] = df.Gender.apply(lambda x: 1 if x == "Female" else 0)
```

```
X = df.iloc[:, 2:]
```

```
X.head()
```

	Age	Salary	Score	Male	Female
0	23	75	80	1	0
1	25	44	50	1	0
2	19	48	79	0	1
3	17	43	5	0	1
4	15	23	28	0	1

The 'Gender' is splitted into 2 binary categories, Male and Female. Where 1 is used for True and 0 for False.

```
# PCA and features selection
pca = PCA(n_components=2).fit(X)
```

```
print(pca.components_)
```

```
[[-0.00527781 -0.38972206 -0.92091277  0.00206566 -0.00206566]
 [-0.00345139 -0.92091454  0.38973043 -0.00271093  0.00271093]]
```

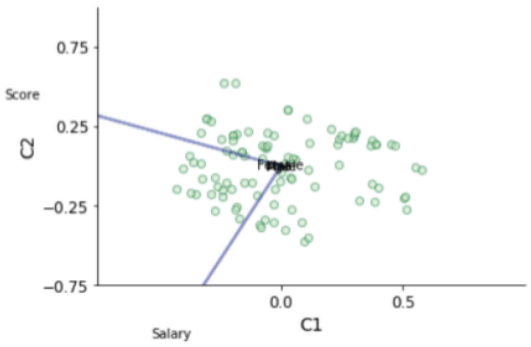
```
print(pca.explained_variance_)
```

```
[718.14033673 206.21716767]
```

The output for PCA is given above, which will be used for the fitting process by the model. This principle component defines the direction of the vectors whereas the square length of vector is defined by explained variance, calculated above. This will represent the principal axis and importance of that axis in describing the distribution.

```
# Transform samples using the PCA fit
pca_2d = pca.transform(X)
extra_graphs.biplot(pca_2d[:,0:2], np.transpose(pca.components_[0:2, :]), labels=X.columns)
plt.xlabel('C1')
plt.ylabel('C2')
```

```
Text(0, 0.5, 'C2')
```

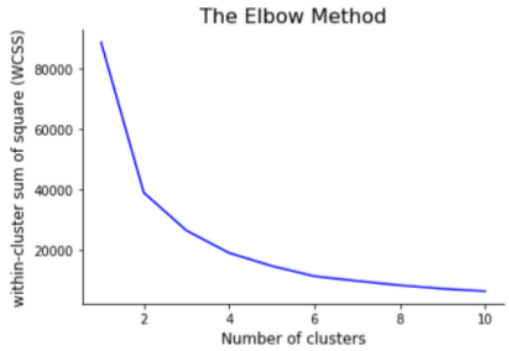


From the above graph, we observed that the Salary and the Score component came out as an important component in the dataset. So, we use the Score and Salary of the customer for the clustering.

5.4.1 K-Mean Clustering

For cluster the data, we need to determine whether the two data points are the same. Proximity measurement is characterized by similarity or dissimilarity between objects. Here K-mean clustering is used for the clustering purpose. Before applying k-mean algorithm we need to find the value of k. K value can be obtained using the Elbow curve, so elbow method is used here to get the value of k for clustering.

```
wcss = []
for i in range(1,11):
    km = KMeans(n_clusters=i,init='k-means++', max_iter=300, n_init=10, random_state=0)
    km.fit(X)
    wcss.append(km.inertia_)
plt.plot(range(1,11),wcss, c="Blue")
plt.gca().spines["top"].set_visible(False)
plt.gca().spines["right"].set_visible(False)
plt.title('The Elbow Method', size=16)
plt.xlabel('Number of clusters', size=12)
plt.ylabel('within-cluster sum of square (WCSS)', size=12)
plt.show()
```



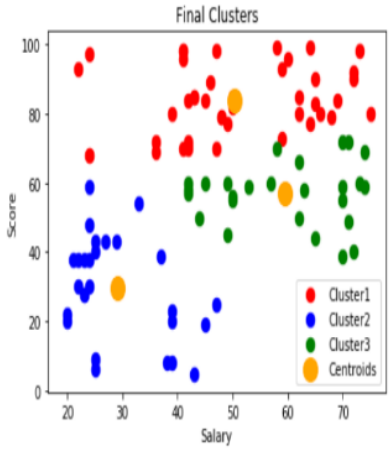
To implement K-Means clustering, we need to find the value of k using at the Elbow curve.

The elbow method is useful to determine the number of clusters in k-means clustering. For K-mean clustering, main idea is to select centroid, one for each cluster. We compute the Euclidean distance between each of the points and the cluster centroid. We assign those points to the cluster center where the distance is minimal. After that, we repeat the calculation for the cluster center. We select the point that is in the middle of each cluster as the new center. And we start again, calculating distances, assigning clusters, calculating new centers. When will we stop? When the center will not change anymore.

Observation: It is clear from the above elbow curve, that we should take the k value equal to 3. The slope of the curve is not steep after 3 so we should take 3 clusters for clustering.

Final Clusters

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 70, c = 'red', label = 'Cluster1')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 70, c = 'blue', label = 'Cluster2')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 70, c = 'green', label = 'Cluster3')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 200, c = 'orange', label = 'Centroids')
plt.title('Final Clusters')
plt.xlabel('Salary')
plt.ylabel('Score')
plt.legend()
plt.show()
```



Observation: There are 3 clusters in Blue, Red and Green colour. The Yellow spot is the centroid of each of the cluster.

- Cluster 1(RED): It is the group of those customers who have high spending score and salary is medium.
- Cluster 2(BLUE): It is a group of those customers who have low spending score and low monthly salary.
- Cluster 3(GREEN): It is a group of those customers who have high salary but medium spending score.

Now, for the retailers it will be easy to formulate marketing strategies for different groups of customers to target them effectively. In Our case retailers need to focus more on cluster 3 as they have the potential to buy more and their spending score may also increase if proper marketing strategies used.

5.4.2 Prediction for a new customer

Here, we enter the values for Age, Salary, Score and Gender and the model will predict for the entered value of a new customer and assign it in a cluster.

```
centroids = pd.DataFrame(kmeans.cluster_centers_, columns = ["Age", "Salary", "Score", "Male", "Female"])
```

```
centroids.index_name = "ClusterID"
```

```
centroids["ClusterID"] = centroids.index
centroids = centroids.reset_index(drop=True)
```

```
centroids
```

	Age	Salary	Score	Male	Female	ClusterID
0	22.461538	59.538462	57.000000	0.730769	0.269231	0
1	22.209302	50.372093	83.651163	0.558140	0.441860	1
2	21.925926	29.111111	29.703704	0.629630	0.370370	2

```
centroids
```

	Age	Salary	Score	Male	Female	ClusterID
0	22.461538	59.538462	57.000000	0.730769	0.269231	0
1	22.209302	50.372093	83.651163	0.558140	0.441860	1
2	21.925926	29.111111	29.703704	0.629630	0.370370	2

```
#prediction for a new customer

customer_X = np.array([[21, 62, 62, 1, 0]])

new_customer = kmeans.predict(customer_X)
print(f"The customer X will belong to cluster: {new_customer[0]}")

customer_Y = np.array([[21, 20, 30, 1, 0]])

new_customer = kmeans.predict(customer_Y)
print(f"The customer Y will belong to cluster: {new_customer[0]}")
```

```
The customer X will belong to cluster: 0
The customer Y will belong to cluster: 2
```

The data enters for new Customer X will belongs to cluster 0 and the data of Customer Y belongs to cluster 2.

CHAPTER 6: CONCLUSION

E-commerce is a sector which has been growing over the years. It is also one such industry which infect is going to grow immensely over the years. Surprisingly, the industry has also flourished in India and has spread the roots all over the country. This research has brought front many aspects which will help the company's grow their business. The very first being, Brand name has a lot of importance in this country. If the brand is well established and has a lot of network the trust people have in the brand increases automatically.

The second thing that was discovered was maintaining customer relationships becomes equally important while building the brand. Even though people prefer to shop from a physical store, they equally perhaps enjoy the discounts more that the e-commerce sites have to offer. Along with the discounts, people also enjoy the wide range of products, they like that they don't have to travel anywhere to shop, personalized emails, offers and easy payment. These are some reasons which incline people more towards online shopping.

Social media advertisement also plays a vital role in attracting the youth towards the sites. Majority of the people who were a part of the audience agreed to have clicked an advertisement. Many have agreed that they have purchased an item from and advertisement, or even a system recommended product. These methods help to convert potential customers into consumers.

One more thing that came up in the analysis of the research was that people also prefer to shop online due to easy payment, but yet they fear or have doubts while doing online transactions. Many people are aware that their card details do get saved and can be later be pulled up by the companies. Some of the other fears/ doubts/ issues that the consumers have is that they fear buying online because they could get something different than what they ordered, return of refund policy, high shipping costs, additional chargers etc. This proves that there are some aspects which need to be changed other than which E-commerce is still at height, but by gaining more customer trust they can acquire more customers.

CHAPTER 7: FINDINGS AND RECOMMENDATIONS

1). Online shopping in India has changed drastically. The number of people that love to shop online have increased over the past few years and the number is still increasing. More and more people are shifting their preference from traditional shopping methods to online shopping, but still there are lots of people who prefer both the methods. In this research, it has been found that there is a relation existing between gender and mode of their shopping preference (online, offline and both equally). Most of the male participants prefer to shop online but female still like both the methods of shopping.

2). During the research, we also identified the major source of customer data and how easy it is for online merchants to collect customer data. We have seen the importance of mining correct information from the available data for better decision making for business growth.

3). 33% of people prefer to pay using their card and 42% have saved their card details in their online shopping account or website. Online retailers use customer's card details and transaction data to learn more about their customers. Information about their purchase behaviour, lifestyle of the card holder, and what services and goods he / she likes can be easily obtained which can be used for clustering the customer base or target individually.

4). In this study, it has been observed that 75% of the people usually prefer to buy accessories from online stores, followed by electronics items which is 66% of the people. Popularity for electronics items has increased over the years, since online retailers began selling exclusive items on their online stores like mobile phones, Health bands, etc. Using the same strategy for clothing's and other items may attract customers to purchase exclusive items online.

5). In a country like India, where people are less confident in making digital payments; there is always a high expectation among the customers for good quality products at very low prices, brand name matters the most. This study has also seen that people are more loyal towards selected ecommerce brands i.e. Amazon, Flipkart and Myntra. And less towards Club Factory and E-bay due to negative feedback about fake product delivery.

6). Most of the people buy online products occasionally. Then there is another set of people who buy at least once in a month. Still the good news for the e-companies is that most of the people buy two or more products in a single purchase. Until now E-commerce companies were engaged in building their brand and customer base by working at low profit margin but soon, they will start making profit from their customers.

7). E-commerce companies need to frame their marketing strategies to attract to more people to buy online.

8). 84% of the online shoppers aware about their personal data is being collected at online sites but most of them are less confident that their data is secure. Low confidence among the customers about data safety could be a threat to the e-commerce companies.

9). Companies need to make different set of marketing strategies for different groups of customers. Using customer segmentation, retailers can target their customers to sell specific product to the target audience based of the customers' interest.

CHAPTER 9: LIMITATION OF STUDY

Due to the outbreak of COVID-19, and the imposition of lockdown, the circumstances reduced the ability to go out and actually interview people for this research. At the same time, due the same reasons the internet speed had reduced variably all over which in all made the whole research process longer as everything acquired for this project was done digitally.

E-Commerce

ORIGINALITY REPORT

18%	10%	1%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijarcce.com Internet Source	3%
2	www.ibef.org Internet Source	1%
3	Submitted to Intercollege Student Paper	1%
4	Submitted to Symbiosis International University Student Paper	1%
5	www.analyticsindiamag.com Internet Source	1%
6	www.slideshare.net Internet Source	1%
7	www.channeladvisor.com Internet Source	1%
8	Submitted to University of Bradford Student Paper	1%
9	Submitted to Institute of Technology, Nirma University	1%

	Student Paper	
10	Submitted to Saint Conleths Community College Student Paper	<1%
11	Submitted to Study Group Australia Student Paper	<1%
12	www.simplypsychology.org Internet Source	<1%
13	Submitted to Christ University Student Paper	<1%
14	Venky Shankar. "Big Data and Analytics in Retailing", NIM Marketing Intelligence Review, 2019 Publication	<1%
15	Submitted to University of Central Lancashire Student Paper	<1%
16	Submitted to Central Queensland University Student Paper	<1%
17	Submitted to Jaipuria Institute of Management Student Paper	<1%
18	Submitted to Embry Riddle Aeronautical University Student Paper	<1%
19	Submitted to DeVry, Inc. Student Paper	<1%

20	Submitted to Oaklands College Student Paper	<1%
21	Submitted to Oxford Brookes University Student Paper	<1%
22	www.igi-global.com Internet Source	<1%
23	Submitted to Keller Graduate School of Management Student Paper	<1%
24	Submitted to The Art Institute Online Student Paper	<1%
25	Submitted to Amity University Student Paper	<1%
26	Submitted to North South University Student Paper	<1%
27	Submitted to Citrus Hill High School Student Paper	<1%
28	Submitted to VHS Virtual High School Student Paper	<1%
29	Submitted to University of North Texas Student Paper	<1%
30	Submitted to Cardiff University Student Paper	<1%

31	Submitted to University of Northumbria at Newcastle Student Paper	<1%
32	Submitted to Colorado State University, Global Campus Student Paper	<1%
33	Submitted to St. Xavier's College Student Paper	<1%
34	Submitted to The University of the West of Scotland Student Paper	<1%
35	Submitted to Hertford Regional College Student Paper	<1%
36	Submitted to Hawaii Pacific University Student Paper	<1%
37	Submitted to University of Nairobi Student Paper	<1%
38	Submitted to University of Reading Student Paper	<1%
39	Submitted to European University Student Paper	<1%
40	www.petnology.com Internet Source	<1%
41	Submitted to TechKnowledge	

	Student Paper	<1%
42	www.emeraldinsight.com Internet Source	<1%
43	Submitted to University of Strathclyde Student Paper	<1%
44	Submitted to Laureate Education Inc. Student Paper	<1%
45	Submitted to University of North Carolina, Greensboro Student Paper	<1%
46	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
47	www.apmas.org Internet Source	<1%
48	Submitted to Waikato Institute of Technology Student Paper	<1%
49	www.geeksforgeeks.org Internet Source	<1%
50	Submitted to Husson University Student Paper	<1%
51	Submitted to The University of Manchester Student Paper	<1%

52	Submitted to Wawasan Open University Student Paper	<1%
53	Submitted to Delhi University Student Paper	<1%
54	Submitted to S.P. Jain Institute of Management and Research, Mumbai Student Paper	<1%
55	Submitted to RMIT University Student Paper	<1%
56	Submitted to University of Hong Kong Student Paper	<1%
57	Submitted to Pace University Student Paper	<1%
58	Submitted to University of New South Wales Student Paper	<1%
59	Submitted to Indus International School, Pune Student Paper	<1%
60	Submitted to University of the Arts, London Student Paper	<1%
61	Irfan Bashir, Chendragiri Madhavaiah. "Consumer attitude and behavioural intention towards Internet banking adoption in India", Journal of Indian Business Research, 2015 Publication	<1%

62	Submitted to University of Southern California Student Paper	<1%
63	Submitted to Imperial College of Science, Technology and Medicine Student Paper	<1%
64	Submitted to University of New York in Tirana Student Paper	<1%
65	Submitted to Chetana's R.K. Institute of Management and Research Student Paper	<1%
66	Submitted to Institute of Management Technology, Hyderabad Student Paper	<1%
67	Submitted to Queensland Academy for Health Sciences Student Paper	<1%
68	Submitted to London School of Business and Finance Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		