

Major Project Report on

**A study on users' awareness and perception on
companies collecting their data for providing
customized marketing**

Submitted By:

Sonakshi Varshney

2K18/EMBA/538

Under the guidance of:

Mr. Yashdeep Singh

Asst. Professor, DSM



Delhi School of Management

Delhi Technological University

Bawana Road, Delhi-110042

DECLARATION

I, Sonakshi Varshney, student of MBA (Executive) 2018-20 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42, declare that Dissertation Report on “**A study on users' awareness and perception on companies collecting their data for providing customized marketing**” submitted in fulfilment of Degree of Master of Business Administration (Executive) is the original work conducted by me.

The information and data given in the report is authentic to the best of my knowledge. This report is not being submitted to any other university for award of any other degree, diploma and fellowship.

Place: New Delhi

Sonakshi Varshney

Date: 24-05-2020

CERTIFICATE

This is to certify that Sonakshi Varshney, Roll No: 2K18/EMBA/538, student of Masters of Business Administration (Executive 2018-2020) at Delhi Technological University, Delhi has accomplished the project titled “**A study on users' awareness and perception on companies collecting their data for providing customized marketing**” under my guidance and to the best of my knowledge completed the project successfully, for the fulfilment of the course Executive MBA.

Dr. Rajan Yadav
Head of Department
Delhi School of Management
Delhi Technological University

Mr. Yashdeep Singh
Asst. Professor
Delhi School of Management
Delhi Technological University

ACKNOWLEDGEMENT

I Sonakshi Varshney, wish to extend my gratitude to Mr. Yashdeep Singh, Asst. Professor, Delhi School of Management (DSM), Delhi Technological University, for giving me all the guidance and valuable insights to take up this Major Project.

I would also like to pay my sincere gratitude to the interview candidates who took out time from their busy schedule to help me with this study.

I also take this opportunity to convey sincere thanks to all the faculty members and friends for directing and advising during the course.

Sonakshi Varshney

2K18/EMBA/538

ABSTRACT

This project aims to study awareness levels and perception of users about data being collected through web tracking over the internet so as to provide them with personalized marketing contents, sometimes through AI as well. The aim is to try to find out whether the users give more importance to privacy or to comfort or there is going to be a trade-off between the two.

With the growing demand for data analytics in order to provide data driven solutions for business problems, companies have been trying really hard to gather as much information from their customers as they can. Not only this but using the data, analytics and AI, various e-commerce players have implemented different features such as recommendation widget in their apps. These features have actually become the bare minimum features that are expected to be incorporated in any e-commerce app. However, this has some positive as well as negative aspects as well. On one hand it is helping companies to know their customers better and server them as per their need and saving customers from getting overloaded with the information that they don't need while on the other hand it also puts the customers into the risk of various privacy related issues. Gathering a lot of customer data and using them for finding solutions to different business problems raises a lot of privacy related concerns among the customers. These days internet users, specially the ones using internet over smartphone, knowingly or unknowingly share their data with various companies.

This project tries to understand the perception of various internet users on this subject and also understand their awareness level for the same. The study has been conducted using descriptive research methodology with both qualitative and quantitative data analysis. The data has been collected through personal interviews and online questionnaire survey. The quantitative analysis has been done using R and PowerBI.

As a conclusion of this study, it can be said that internet users are aware of web tracking and their comfort level in letting the companies collect their data varies according to the context.

TABLE OF CONTENTS

Chapter	Title	Page
Chapter 1	Introduction	1-5
1.1	Project Overview	1
1.2	Introduction to web tracking	1
1.3	Web tracking and AI: changing the way companies market	4
Chapter 2	Literature Review	6-14
2.1	Mechanisms for Web Tracking	6
2.2	Web tracking and privacy concerns	9
2.3	Choices for users to control web tracking	11
2.5	Data protection and privacy laws in India	12
Chapter 3	Research Methodology	15-20
3.1	Problem description and scope of the study	15
3.2	Research approach	15
3.3	Data collection	16
3.4	Analysis methodology	19
Chapter 4	Analysis and Discussion	21-39
4.1	Themes from the PI data	21
4.2	Analysis of the survey	22
4.3	Major findings	37
Chapter 5	Challenges and Limitations	39
5.1	Challenges encountered	39
5.2	Limitations of the study	39
Chapter 6	Conclusion and Research Implications	40
	References	42
	Annexures	44

TABLE OF FIGURES

FIGURE 1: WHY DIGITAL MARKETING ANALYTICS IS IMPORTANT	5
FIGURE 2: DEMOGRAPHICS OF PI CANDIDATES	17
FIGURE 3: AGE-WISE DISTRIBUTION OF SAMPLE	17
FIGURE 4: GENDER-WISE DISTRIBUTION OF SAMPLE.....	18
FIGURE 5: PROFESSION-WISE DISTRIBUTION OF SAMPLE.....	18
FIGURE 6: ANALYSIS METHODOLOGY	20
FIGURE 7: MEANING OF CODES USED IN THE VISUALS -1	22
FIGURE 8: INTERNET USAGE TRAITS OF THE RESPONDENTS	23
FIGURE 9: PERSONALIZATION EXPERIENCE	24
FIGURE 10: MEANING OF CODES USED FOR VISUALS -2	25
FIGURE 11: GENERAL PERCEPTION OF USERS.....	26
FIGURE 12: AWARENESS OF DATA TRACKING (GENERATION-WISE DISTRIBUTION).....	27
FIGURE 13: AWARENESS OF DATA TRACKING (GENDER-WISE DISTRIBUTION)	28
FIGURE 14: AWARENESS OF DATA TRACKING (PROFESSION-WISE DISTRIBUTION)	29
FIGURE 15: COMFORT LEVEL IN SHARING DIFFERENT TYPE OF DATA.....	30
FIGURE 16: AWARENESS ABOUT DATA PROTECTION BILL (GENERATION-WISE DISTRIBUTION)	31
FIGURE 17: AWARENESS ABOUT DATA PROTECTION BILL (GENDER-WISE DISTRIBUTION)	32
FIGURE 18: AWARENESS ABOUT DATA PROTECTION BILL (PROFESSION-WISE DISTRIBUTION).....	33
FIGURE 19: RESULTS OF DECISION TREE FOR PERSONAL DATA	34
FIGURE 20: DECISION TREE FOR PERSONAL DATA	34
FIGURE 21: RESULTS OF DECISION TREE FOR FINANCIAL DATA.....	35
FIGURE 22: DECISION TREE FOR FINANCIAL DATA	35
FIGURE 23: RESULTS OF DECISION TREE FOR ONLINE ACTIVITIES	36
FIGURE 24: DECISION TREE FOR ONLINE ACTIVITIES	36

CHAPTER 1: INTRODUCTION

1.1. Project Overview

The growth of technology has completely changed the way businesses are being run in today's era. Technology has impacted every aspect of a business such as efficiency of operations, communication with stakeholders, research capacities and more.

The internet technology has completely transformed the way organizations conduct their marketing activities. Companies have started moving from conventional marketing to digital marketing due to many benefits it provides such as better targeting, reaching masses with low cost options and many more.

At present, companies have started investing in data analytics, artificial intelligence etc. to provide customized marketing solutions to their customers for which they track the web activities of their potential or actual customers via different means.

This activity of web tracking poses a great threat to the privacy of the users. There may be some internet users who are not even aware that they are being tracked over the web.

The intent is to study the users' awareness and perception towards web tracking by different companies for the purpose of providing customized marketing experience to its customers. The aim is to understand how much awareness is there among the users about web tracking and the extent to which users are ready to let companies invade their privacy for providing customized marketing experience.

1.2. Introduction to web tracking

1.2.1. What is web tracking?

Web tracking is a way to track the activities of users over the web.

Websites track each of their users for their online behaviour. They collect different user information, store it and sometimes even share it

with other parties for different purposes. The tracking is done using specialised software tools.

Types of web tracking:

- **First-party web tracking:** When the tracking is done by the website itself which the user is using, it is known as first-party web tracking. For example: If you are reading an article on nytimes.com, New York Times is aware which article you are reading. In this case, New York Times is the first-party and the tracking done by it is called first-party tracking.
- **Third-party web tracking:** When an external entity has been allowed to embed its tracking tool by a company on its website, the tracking done in this scenario is known as third-party web tracking. For example: New York Times allows a third-party tracker such as doubleclick.net to be embedded into its website for providing targeted advertising, let's say. This will be called as third-party tracking.

1.2.2. Why web tracking is done?

There are multiple reasons for tracking the users' online activities. Some of them are discussed as below:

- **Website performance:** Web tracking is essential for a lot of the website features. Some of the features that won't work without tracking are as follows:
 - Keeping the user logged in while he/she is browsing through different pages of the website.
 - Next video recommendation on YouTube.
 - Saving the products in the shopping cart on an e-commerce website.
- **Advertising:**

With the emerging trend of digital marketing, companies have started tracking internet users' activities to provide them with

“**targeted advertising**”. Data from different sources such as user’s search history, time of the day, location etc. is collected and adverts based on this data is shown to the users. This helps companies to better target their customers, showing them only what they are interested in and not overloading them with adverts about the products or services that they either don’t need or have no interest in.

A common practice called “**retargeting**” is also the reason for tracking the users. In this practice, a user’s data about the sites he/she has recently visited is collected and adverts on this basis are shown to him/her.

Web tracking helps in adding the element of *customization* into the advertising and customized adverts are believed to bring in more value to the company.

- **Web Analytics:**

Using analytics tools over websites has become very common now-a-days. Companies keep track of their users’ activities over their websites in order to collect as much information about them as they can. They use their demographic information, browsing history and other information. All this information help companies take better business decisions and design personalized solutions for their customers. This information also helps them to optimize their websites according to how users use it. Optimization in turn increases the performance of the website and hence, more customer is more satisfied.

With the boom of digital marketing and use of artificial intelligence in digital marketing, the companies have started coming up with real time content generation and marketing offers for their customers. The marketing strategies have become highly customized these days because of AI and this leads to a greater customer satisfaction and retention.

1.2.3. Controversy around web tracking

The controversy associated with web tracking is the “**Privacy of Users**”. The web tracking requires users’ data to be collected. Most of the time the users are not even aware that their activities are being tracked. These days, companies do mention over their websites that they are using cookies. This way they are actually just informing the users rather than seeking their permission. So, it becomes necessary for them to share their information if they want to use the website. Also, it may happen that users are not even aware of what cookies are and for what purpose they are used.

IT industry experts have a mixed perception about it. Some of them feel that collecting the users’ data by tracking them is ethically wrong even if they are informing the users as users may have objection in sharing any of their information over the web and there should be a mechanism to seek their permission before collecting their data. On the other hand, some of them felt that the information that is being collected is not so private that it can cause them any sort of harm. They believe that this tracking is actually providing benefits to the users and hence even if users are made aware, they will be comfortable in sharing such information in order to receive better services.

Because of the mixed reviews from the experts on this a clear conclusion could not be drawn on this issue. Hence, this study is being conducted to know the users’ perspective on this issue as at the end they are the owner of the information that is being collected through web tracking.

1.3. Web tracking and AI: changing the way companies market

The companies these days are trying to achieve competitive advantage by optimizing their marketing strategies. In order to achieve this goal, digital marketers are using digital marketing analytics.

Digital marketing analytics is all about generating insights from the data collected through consumer behaviour and using these insights to formulate

future marketing strategies. When paired with artificial intelligence (AI), it gives an extra edge to the marketing solutions and strategies.

With the growing benefits of data analytics, it has become important for marketers to acquire right data and understand how they are going to analyse this data.



Figure 1: Why digital marketing analytics is important
(Source: <https://www.lyfemarketing.com/blog/digital-marketing-analytics/>)

The digital marketing analytics and AI broadly provide following benefits to the marketers:

- It can help marketers gain a complete view of their customers over different digital media channels and hence with the help of AI they can personalise their marketing strategies.
- It can help marketers to predict customer behaviour and hence they can take actions proactively.
- Analytics along with AI in making real time decisions and hence gives a competitive edge to the organization.
- AI helps in increasing the productivity as it automates the process and hence it improves ROI for a firm. Also, with analytics the performance of AI solutions become better.

CHAPTER 2: LITERATURE REVIEW

2.1. Mechanisms for Web Tracking

There are different mechanisms that can be used to track users over the web.

These mechanisms can be broadly divided into following categories:

- Session-only
- Storage-based
- Cache-based
- Fingerprinting
- Other mechanisms

Each mechanism has multiple techniques to track the users. Techniques under each mechanism will be discussed in the following paragraphs.

2.1.1. Session-only mechanisms:

In these mechanisms only the information related to the series of requests from the same browser are tracked. This mechanism includes different techniques which are discussed below:

- I. *Session Identifiers:* These are the unique numbers assigned by a web server to a particular user session. Whenever a user visits a website, a unique session ID or identifier is assigned to that session.
- II. *Explicit Form Authentication:* Some websites require the user to register on the website before they can have access to its content. This way, websites authenticate the users and monitor their activities over the website irrespective of the user's browser, place, computer or OS.
- III. *Document Object Model (DOM) property:* The DOM is a programming interface that defines the logical structure for HTML and XML documents and the way to access and manipulate them. There is a DOM property called `window.name` that can hold a large amount of data as string and can then be stored on user's computer. This basically stores session related data for the user.

2.1.2. Storage-based mechanisms:

The techniques under these mechanisms involve storing the data on user's computer explicitly. Some of the different techniques under storage-based mechanisms are as follows:

- I. *HTTP Cookies:* These are the files with small pieces of data containing a unique identifier for the user and is usually stored in user's computer. They can be used alone or combined with other methods for tracking purpose such as cookies along with web-form authentication.
- II. *Persistence Service and Flash Cookies:* Flash cookies are the text files sent to the client by a web server for the content that is supported by Adobe Flash. These are usually used in advertisements of the website. There are also Java JNLP Persistence Services that store the data on the client systems locally and can be used for even those applications that are executing in untrusted environments.
- III. *LocalConnection object:* There is another method for web tracking which is Flash LocalConnection object that is used for communication between different flash instances that are executing on either normal or private windows.
- IV. *Silverlight-Isolated Storage:* It can be used to store about 100 KB of information on a site related to the user as flash cookies in his profile. This storage is not enabled for private mode and can only be cleaned manually.
- V. *SQL and Indexed Databases:* Sometimes, a relational SQLite database is used at the client side to store information instead of using local file based system. Also, there is HTML5 indexedDB that is not a relational database, but a key-value pair based database and it has the same privacy impact as the local storage.

2.1.3. Cache-based mechanisms:

The cache-based tracking mechanisms also use the storage at client side but apart from storing data, these mechanisms also use other prospects such as finding out the websites that have been visited previously. Some of the techniques under this category are:

- I. *Web Cache*: Whenever a user visits a website, the browser stores the information in its cache in order to access it faster when the user visits the website again. This way browsing history of a user can be stored.
- II. *DNS Cache*: There are caches that stores DNS information for the previously visited websites. Whenever a user visits a website the first lookup is done in this cache and if the entry exists in the cache, the lookup time is reduced significantly.
- III. *Operational Cache*: These caches store the information related to the operations performed by a browser such as information about credentials for authentication, redirects etc.

2.1.4. Fingerprinting mechanisms:

Fingerprinting is a collection of different methods that use a wide variety of technologies in order to track users on different websites he is visiting. This way, fingerprinting covers a wide range of properties belonging to user's online activities. Some of the fingerprinting techniques are discussed as follows:

- I. *Location and Network fingerprinting*: In this method, the global network address and the location of the user is extracted from the HTTP request's header and using different tools user's ISP and domain can be identified.
- II. *Device fingerprinting*: This method collects the information about user's time zone, IP address, plugins, screen resolution, language settings and more. This method basically collects information related to the device that user is using for online activities.
- III. *Browser fingerprinting*: This method is used to collect the type and version of the browser that a user is using to visit websites. The websites use this method to identify a distinct user and his online behaviour.
- IV. *OS fingerprinting*: This method involves identification of version and architecture of the device's operating system used by the user for visiting the websites. This is done by JavaScript and Flash both.

2.1.5. Other mechanisms:

Apart from the mechanisms discussed above, there are many other methods like *clickjacking*, where a sensitive element of the website is presented out of the context so that user also starts acting out of context. There are *timing attacks* also using which information like a user's login status and if he has an account on a website or not can be determined.

Some websites also collect data by collaborating with unconscious users or by capturing the phone's metadata through mobile apps.

2.2. Web tracking and privacy concerns

As the companies are moving to online platforms with their own websites, social media pages or handles and more, their customers are being tracked in one way or the other. This tracking over online platforms raise a lot of privacy concerns for the users. Some of these privacy concerns are discussed below:

2.2.1. Data leakage:

When a data is transferred from an organization to an external entity in an unauthorized manner, it is called as data leakage. This can be done physically or electronically. There can be various reasons for data leak, some of them are as follows:

- I. The data leaks can be *accidental*. Most of the data leaks are usually accidental and not malicious or intended. However, the penalties are no less than intended or malicious ones.
- II. The data can be leaked by *a disappointed or an impudent employee* even though an agreement has been signed by him with a company. This is known as *data exfiltration*.
- III. There are also malicious attacks on electronic media of communication such as *phishing attacks*. The attackers pretend to be legitimate users and contain malicious links in the email that can ask employees to share secret data such as important credentials.

However, companies have to take necessary precautions so as to prevent any data leakage.

Usually companies that host third party content over their website, the third party usually gets to know about the URL and hence about the details that can be extracted from HTTP request URL as well.

2.2.2. Identifiability:

The data extracted from browser search history can help identifying an individual. This may help you to know about user's interests, his personality, gender, sexual orientation and a lot more. This is a clear intrusion to someone's privacy. Not everyone is comfortable to reveal their identity to anyone. Each individual has different comfort level regarding this. Apart from this, there are scenarios where a first party can intentionally or unintentionally share this identifiable information of the users with a third party. Scenarios like this pose greater threat to the privacy of the customers.

2.2.3. Different views regarding tracking:

Different people may have different views regarding the web tracking and the policies related to it. How much tracking to be done and where to put the limit are all subjective matters and hence the perception varies from one person to another and also from policy makers of one country to another country. For example, according to European policy makers believe that by default there should be no tracking while on the other hand individuals from advertising groups believe in making web tracking a default process.

2.2.4. Harmful risks to users:

The web tracking also makes users prone to different harmful scenarios. People with malicious intent may use this data for harming the users or some employee in the greed of money may sell the users' data to rivals. A lot of possible harmful scenarios may happen if users' data is not well protected.

While looking into the harmful aspects, there are four variables that needs to be focused upon –

- I. *Actor*: Someone who causes the harm. It can be a hacker, an employee or any government agency etc.
- II. *Access medium*: A mean for actor to use the data collected from tracking.
- III. *Action*: It is something that causes the actual harm. It consists of the efforts that an actor puts in to cause harm.
- IV. *Harm*: It is the actual damage caused. The harm can be economic in nature, it can be physical as well as psychological.

2.3.Choices for users to control web tracking

A web user has a lot of options through which he can limit this tracking. Some of the different ways through which a user can control web tracking are:

2.3.1. Private Mode of Browsing:

The browsers come with a private mode browsing, for example, in Google Chrome there is an incognito mode of browsing while Firefox has private browsing feature. In the private browsing, the browser doesn't store any cookies, history, form-based data etc.

Apart from that, there are search engines available that do not store any of your search history as well. For example, "duckduckgo" is one such search engine.

2.3.2. Opt-out Cookies:

These are the cookies used by companies over their website for the users who don't want the website to install the cookie in future and hence this way a user can prevent third parties from tracking his preferences.

2.3.3. Browser-based Plugins and Blocking:

The web browsers also have extensions or settings that can be used to block tracking by the websites. For example, Firefox and Chrome provide an extension called "Adblock Plus" that is used to block unwanted advertisements. Another example is that of Internet Explorer's "Tracking Protection List" that prevents cross browser tracking by third parties.

However, sometimes the extensions can also be malicious and hence it is recommended to download these extensions from reliable sources.

2.3.4. DNT (Do Not Track):

It is a web browser setting that a user can enable to stop being tracked. If a user has enabled this setting, then a signal is being sent by web browser to different websites and third party services that the user has opted out from being tracked.

2.4.Data protection and privacy laws in India

When there is a talk about fundamental *right to privacy*, then Indian Constitution doesn't grant this right evidently to the citizens, but it has been read out by the courts into other fundamental rights such as freedom of speech and expression and the right to life and personal liberty.

The data protection related laws in India are discussed in the following sections:

2.4.1. The Personal Data Protection Bill, 2019:

On December 11, 2019 Mr. Ravi Shankar Prasad (Ministry of Electronics and Information Technology) introduced this bill in Lok Sabha. The bill attempts to provide security to personal data of individuals and setup an authority for the same.

The key highlights of this bill are as follows:

- I. The bill categorizes financial data, religious or political beliefs, biometric data, caste etc. as the sensitive personal data.
- II. It also controls how the personal data can be processed by government, Indian companies and foreign companies in India.
- III. It also defines the responsibilities of a data fiduciary (someone who decides the ways and reasons of processing the data). For example, Data fiduciaries are required take necessary steps to maintain transparency and accountability.
- IV. The bill also lays down the rights of individuals such as the right to limit the continuation of disclosure of their personal data by

the fiduciary or right to get confirmation on whether the data has been examined by the fiduciary and more.

- V. The bill allows the fiduciary to process data only when the consent has been provided by the individual except for certain cases such as for legal proceedings or medical emergency etc.
- VI. It defines social media intermediaries and their obligations as well.
- VII. The bill also sets up an authority for this - *Data Protection Authority* and defines its responsibilities such as preventing misuse of data, ensuring compliance with the bill and taking measures to protect individuals' interests.
- VIII. It also lays down the policies related to transfer of data outside India.
- IX. The bill also explains the exemption rules for central government agencies.
- X. It also defines offenses to this bill, amendments to IT Act and policy about sharing of non-personal or anonymised personal data with the government.

2.4.2. Laws related to privacy in India before 2019:

Before the personal data protection bill was introduced, the data protection was dealt under the Information Technology Act, 2000 and the Contract Act, 1872.

The IT act has provisions to deal with the issues regarding the compensation and punishment in case of misuse and unfair disclosure of personal data and infringement of terms of contract that are related to personal data.

However, the rules under this law handled issues related to only sensitive personal data such as passwords, financial information (such as bank account details, credit or debit card information etc.), biometric details, medical records, sexual orientation etc.

Under section 43A of IT Act, a body who fails to implement and maintain appropriate security measures for the data, may have the

liability to pay the damages to the affected person and there is no defined upper limit for the compensation in this regard.

The section 72A of IT Act has the provisions for punishing the act of information disclosure (intentional or unintentional) without the concerned individual's consent.

The section 69 of this act lists out the scenarios under which there is an exception to the privacy rules such as in case of matters related to sovereignty or integrity of India, security of State, public order, investigation of any offense etc.

In 2008, a lot of amendments were made to this act such as provision for punishment in case of privacy violation, identity theft, cyber terrorism etc. Even the section 43A was added in this amendment.

CHAPTER 3: RESEARCH METHODOLOGY

3.1. Problem description and scope of the study

The project aims at studying the awareness level and perception of internet users about tracking over the internet. With the growing internet usage and companies going digital, leveraging various new technologies such as AI to gain competitive advantage over their competitors through digital marketing, the privacy of the user's data becomes a very critical concern. As the data collected through web tracking is increasing, the cybercrimes and frauds are also increasing.

In one of the earlier studies on “how AI is transforming digital marketing” conducted as part of semester project for this course, it was concluded that tracking user over internet is a highly subjective matter and opinion differed from person to person within a group of IT experts. Hence, this study is being conducted to conclude a generic opinion on this matter with people having different demographic background.

This study aims to understand how much aware the users are about this web tracking mechanism and how much comfortable they are to share their personal and financial data with the companies for personalized marketing.

This study also tries to find out the user's awareness about laws related to protection of personal data in India.

The study is conducted for internet users only. Hence, the scope of the study is limited to the internet using population of urban area.

3.2. Research approach

The research approach followed for this study is a descriptive research approach where in the answer for how much aware and comfortable a user is about sharing his data for personalized marketing.

The rationales behind selecting this approach are:

- The research problem was well defined.
- This approach would give a holistic view of the problem.
- The research could be conducted in its natural environment.

- The aim of the study was to generalize an opinion.

3.3. Data collection

3.3.1. Data collection procedure:

This study has been conducted on the data collected through *primary data collection* procedures. The data collection has been done through two different methods:

- I. *Personal Interview (PI)*: Telephonic PIs were conducted to perform qualitative analysis so that it can provide a direction for further analysis on a larger sample.
- II. *Online Survey through Questionnaire*: An online questionnaire was used to collect data from a larger sample in order to provide a generalized conclusion.

3.3.2. Sample size:

Sample size for the personal interviews was 5. The properties of this sample have been discussed in one of the upcoming sections.

For the online questionnaire, the sample size was of 300 respondents and general properties of this sample is also discussed in the same section as that of the PI. Before starting with the data collection procedure for the questionnaire, a pilot test on 10 respondents was run in order to determine if the questionnaire actually captures the required information well or not. These respondents were not part of actual sample considered for the survey.

3.3.3. Sampling technique:

The sampling technique used for this study is the *convenience sampling* wherein the PI candidates were identified from the people known to the researcher and questionnaire was also shared with acquaintances, who further shared it with their contacts.

3.3.4. Sample properties:

- I. **Demographics of PI Candidates:**

The demographic details include the respondent's gender, age and profession. The demographic profile of the respondents has been captured in the table below:

Candidate	Gender	Age Group	Profession
Candidate 1	Male	18-24 Years	Student
Candidate 2	Female	45-54 Years	Homemaker
Candidate 3	Male	25-34 Years	Private job in banking sector
Candidate 4	Male	35-44 Years	IT Professional
Candidate 5	Female	25-34 Years	Business

Figure 2: Demographics of PI Candidates

II. Demographics of Survey Sample:

The demographic details of the sample have been shown using pie charts.

The charts below show the distribution of the sample on the basis of 3 parameters which are:

- Age
- Gender
- Profession

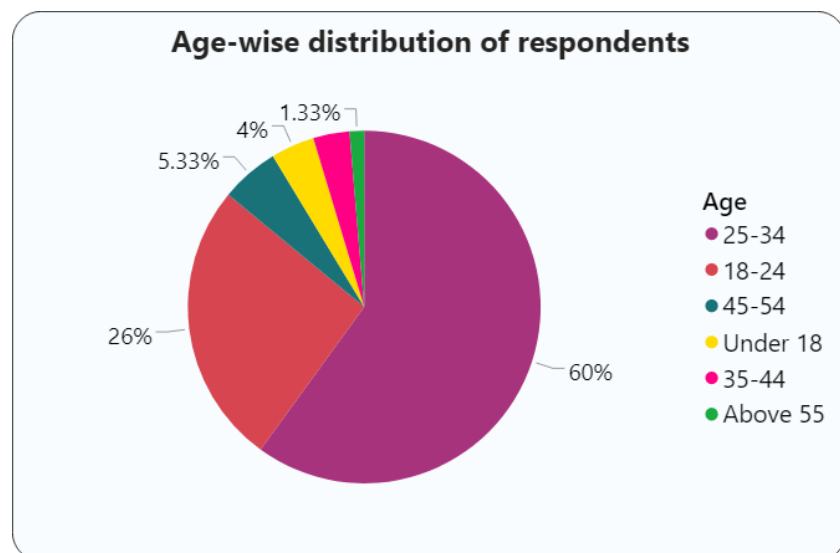


Figure 3: Age-wise distribution of sample

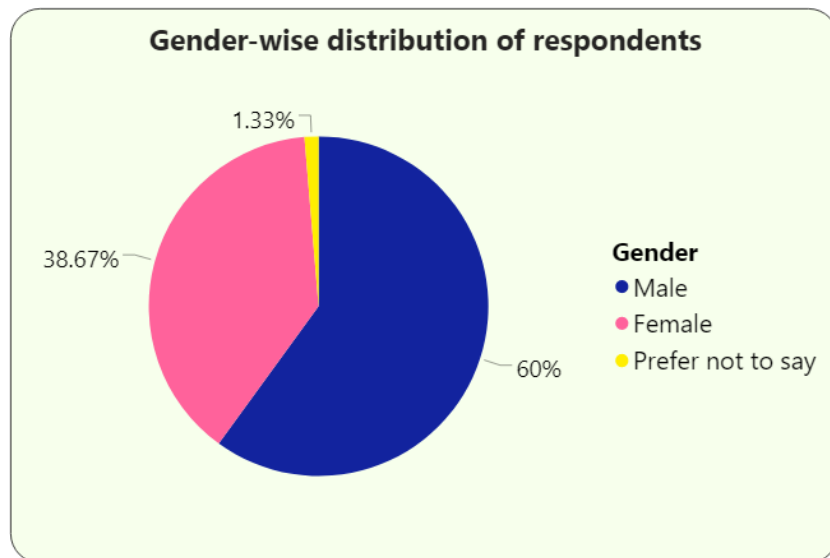


Figure 4: Gender-wise distribution of sample

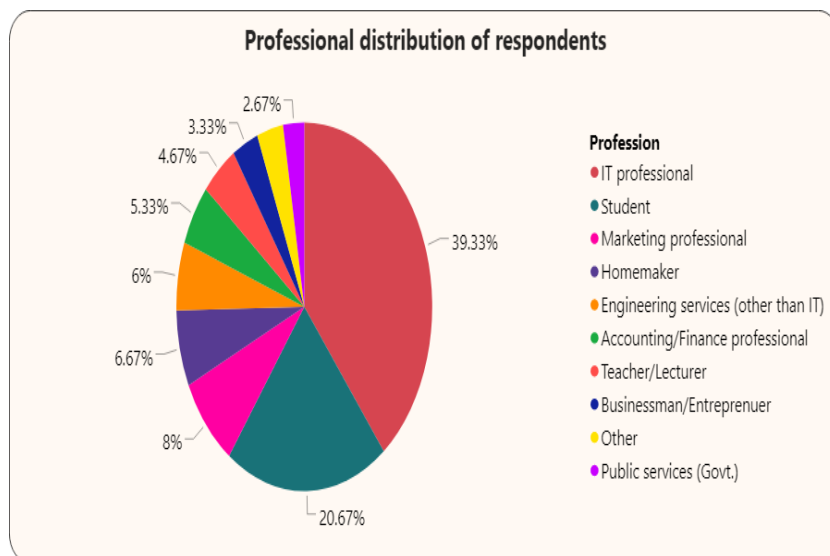


Figure 5: Profession-wise distribution of sample

The demographics of the sample as shown in the chart can be summarized as follows:

- The majority of the respondents belonged to the age group of 25-34 years, which is mainly the working class of the society.
- The sample largely comprised of the male population. 60% of the respondents were male while rest 40% belonged to female and other category.

- If profession-wise distribution is looked into, the sample largely consisted of the IT professionals, followed by the students.

3.4. Analysis methodology

The study has been conducted using both qualitative and quantitative analysis.

- *Qualitative Analysis:* The data collected from PI was analysed in a qualitative fashion in order to generate an insight into the problem which was further used for quantitative analysis to check if these findings can be generalized to a larger sample or not.
- *Quantitative Analysis:* The data collected from questionnaire has been analysed in quantitative manner using various data exploration and visualization techniques along with decision tree analytics method.

Before analysing the data, data cleaning was done to remove any duplicate entry and post that data transformation was done using Excel tools and formulas depending upon some criteria which has been mentioned in the next section. Data transformation has been done to convert the data into the format that is convenient for analysis.

The analysis has been done using Power BI visualization techniques and R software. The data has been analysed in two phases:

- I. Phase 1- Data exploration:* In this phase, exploratory insights have been generated using visualization techniques. These insights have been generated by following the themes that have been uncovered from the analysis of the PI.
- II. Phase 2- Predictive Analytics:* In this phase, decision tree analytics has been used to make a conclusion on whether a user will consider sharing the data for personalized market or not depending on the certain parameters such as – demographic profile of the user and the kind of data that has been demanded.

Overall analysis methodology is shown in the figure below:

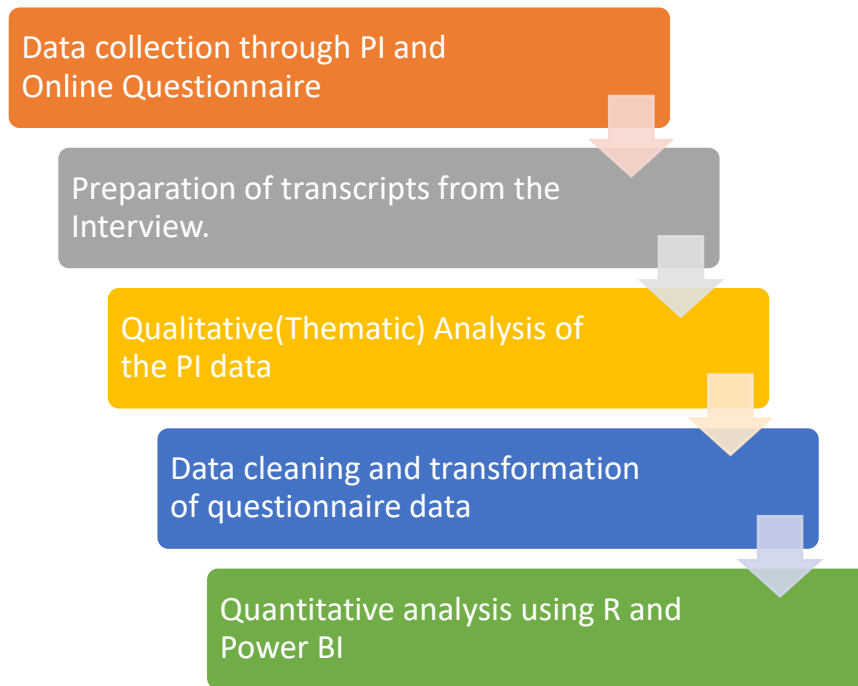


Figure 6: Analysis Methodology

CHAPTER 4: ANALYSIS AND DISCUSSIONS

4.1. Themes from the PI data

From the data collected via interview, following themes emerged while analysing it:

- Awareness about personalization over the web and data collection by companies for this.
- Kind of information candidates are comfortable in sharing
- Awareness about the legal rights available in India for data protection

The data from PI reveals that, all the 5 candidates experienced personalization while using internet either in the form of content recommendation or targeted ads over various platforms. While 3 candidates were clearly aware that this personalization is being offered at the cost of their data that has been collected by the companies in one way or the other, one of them was unconsciously aware that all these personalized recommendations and adverts are using browsing history and other data and how to some extent any online activity is an invasion into the privacy of an individual, while on the other hand one of the candidate was totally unaware of this fact. A major point of differentiation between those three candidates who were very well aware about data collection by companies and other two candidates was that of their professional environment. Those three candidates were actually spending their time on internet for their professional work as well while among the other two, one was a housemaker and another one was a college student. So, it may be possible that their awareness about data collection is somewhat attributed to their profession.

When it comes to the kind of information that users are comfortable to share for any sort of personalization over the internet, four out of five candidates were comfortable in sharing their basic details such as age, gender, their interests etc., while one of them who was an IT professional wasn't comfortable in sharing his personal details and rather expected companies to give a confirmation that even the historical data will not be kept or used. Apart from this, not even a single candidate was comfortable in providing the

financial details even if they are to receive any extraordinary benefits out of it. However, one of the candidates working in banking sector was comfortable to share details like PAN card number only on reliable portals such as the one at the workplace or for some govt. authority. Again, within this theme as well, we can see that profession of the candidate may have some sort of impact on their perception. Also, the candidates who were ready to share their basic details were using internet for socializing as well while the candidate who was not comfortable in sharing any of the data was using the internet mostly for official purpose or for Netflix.

For the awareness of the legal rights, there was only one candidate who knew very well about the data protection bill in India because the same was being implemented as part of the project in the organization. However, rest of the candidates were not much aware about it. Two of the candidates were not even aware if any such bill exists or not. While rest of the two knew that there exists a bill for data protection but nothing more than that. From this, we may assume that users may not actually be aware of their rights as a citizen of India that they have for protection of their personal data.

However, all these insights need to be analysed for a larger sample using some quantitative techniques which has been covered up in the next section.

4.2. Analysis of Survey

4.2.1. General Traits of the Sample:

The general traits of the sample are studied on the basis of two categories:

- The internet usage traits
- Experience of personalized marketing/features

The visuals in this section have certain codes which can be read as:

Code	Meaning	Colour in the visual
0	No	Pink
1	Yes	Blue

Figure 7: Meaning of codes used in the visuals -1

I. *Internet usage traits:* Figure 8 shows the various internet usage related traits of the sample. The following traits can be seen from the visuals:

- There are very few respondents who use the internet for less than an hour in a day.
- A lot of respondents use internet for entertainment (More than 80%) followed by social networking (More than 75%). On the other hand, the usage of internet for gaming purpose among the respondents was the lowest of all, i.e., 36%.

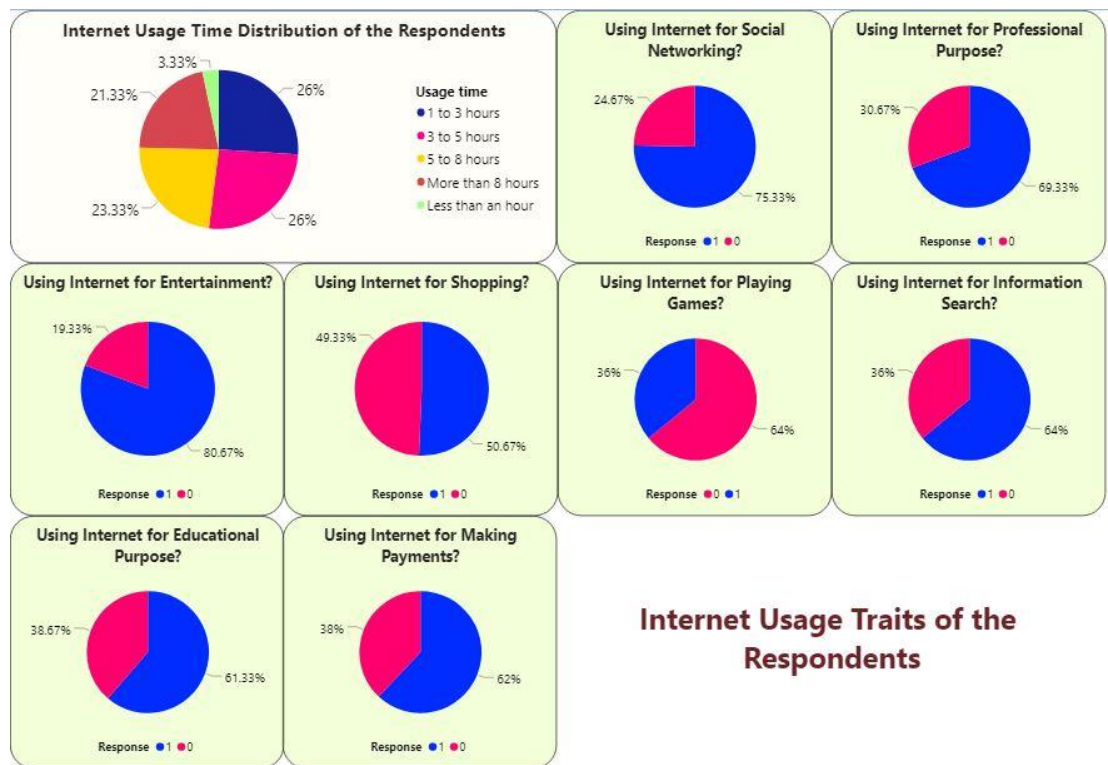


Figure 8: Internet Usage Traits of the Respondents

II. *Experience of personalized marketing/features:* Figure 9 shows the kind of personalization that has been experienced by the respondents in the sample. The following findings can be seen from the visuals:

- There are only 4% of the respondents who didn't experience any sort of personalization while using the internet.
- A large number of respondents have experienced personalization in the form of recommended videos (More

than 78%) and personalized ads or content over the websites (78%).

- Experience related to exclusive personalized offers and discounts was also towards lower side (41.33%).

Experience of Personalized Marketing

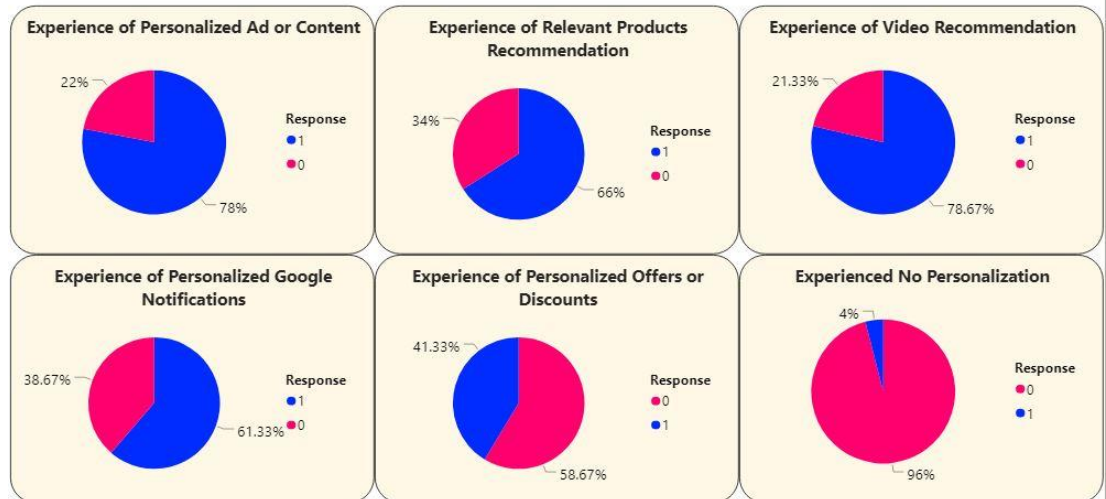


Figure 9: Personalization Experience

4.2.2. Data Exploration:

From the findings of above section, it can be seen that respondents have been using internet mainly for the purpose of entertainment or social networking, which is a very common trend these days and most of them have experienced personalized marketing in one form or the other. The kind of personalization (Ad/Content and Video recommendation) that has been experienced by most of the respondents is seen mainly on the websites like Facebook, Instagram, YouTube etc., and these sites are actually meant for the purpose of social networking and entertainment only.

It is a well known fact that personalization is offered by collecting different type of the data of a user. In this section, the focus will be on studying the awareness about data tracking and data protection bill along with the levels of comfort in sharing their data for personalization among the respondents.

The visuals under this section use different codes that have been assigned to the values of different variables. The codes can be read as follows:

Variable	Code	Meaning
Generation (Age-Group)	G1	Above 55
	G2	45-54
	G3	35-44
	G4	25-34
	G5	18-24
	G6	Under 18
Profession	P1	IT professional
	P2	Marketing professional
	P3	Teacher/Lecturer
	P4	Businessman/Entrepreneur
	P5	Accounting/Finance professional
	P6	Public services (Govt.)
	P7	Engineering services (other than IT)
	P8	Student
	P9	Homemaker
	P10	Other

Figure 10: Meaning of codes used for visuals -2

For the data exploration, the focus has been on the three themes that were uncovered from the PI. The three themes are as follows:

- Awareness about data collection among respondents according to their demographic profile.
- Comfort level of respondents in sharing different types of data with the companies for personalized marketing
- Awareness about the data protection bill among the respondents according to their demographic profile.

Figure 11 shows a general perception of the users based on the three themes. The following insights can be generated from the visuals:

- Most of the respondents in the sample were aware that data tracking is done by companies to provide personalization experience (More than 92%).
- While there was a good number of respondents who knew about data protection bill in India (More than 45%), there were

also many respondents who felt that there was a bill like that but were not certain about it (More than 36%).

- Also, there were more than 59% of respondents who would not consider sharing their data for any personalized marketing experience. But still some respondents would actually consider sharing of data for having personalized marketing experience.

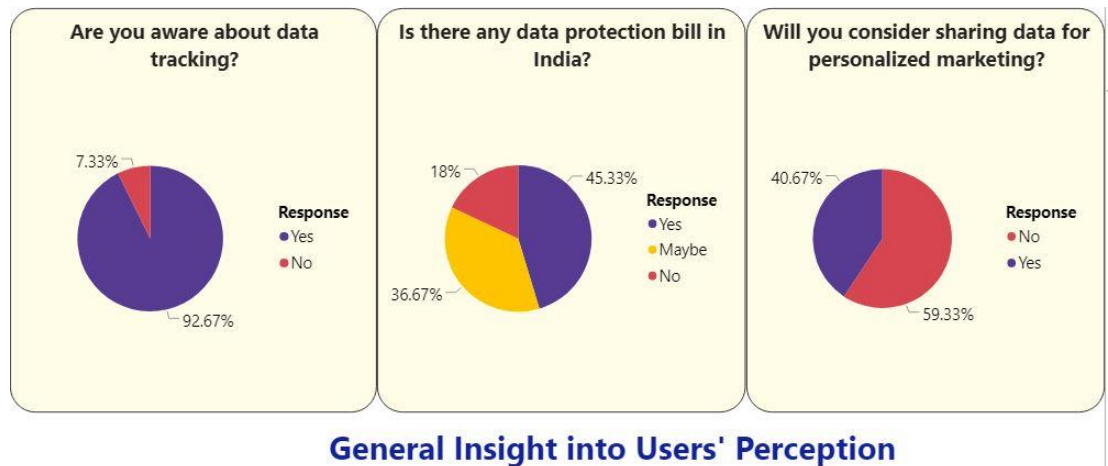


Figure 11: General Perception of Users

Now, looking into the insights coming from each of the three themes along with the demographic profile of the respondents.

I. Awareness about data collection among respondents according to their demographic profile:

Figure 12 shows the generation-wise distribution of the awareness levels of respondents. The following findings can be seen in the visual:

- Most of the respondents were actually aware about data tracking (80% or more in all the generation categories).
- All the respondents of generation G1 and G6 were aware of data tracking, however, this could be due to the low percentage of respondents belonging to this category (see, figure 3).
- Most of the unaware users belonged to the generation G3 which is the group of respondents having age in the age group 35-44 years.

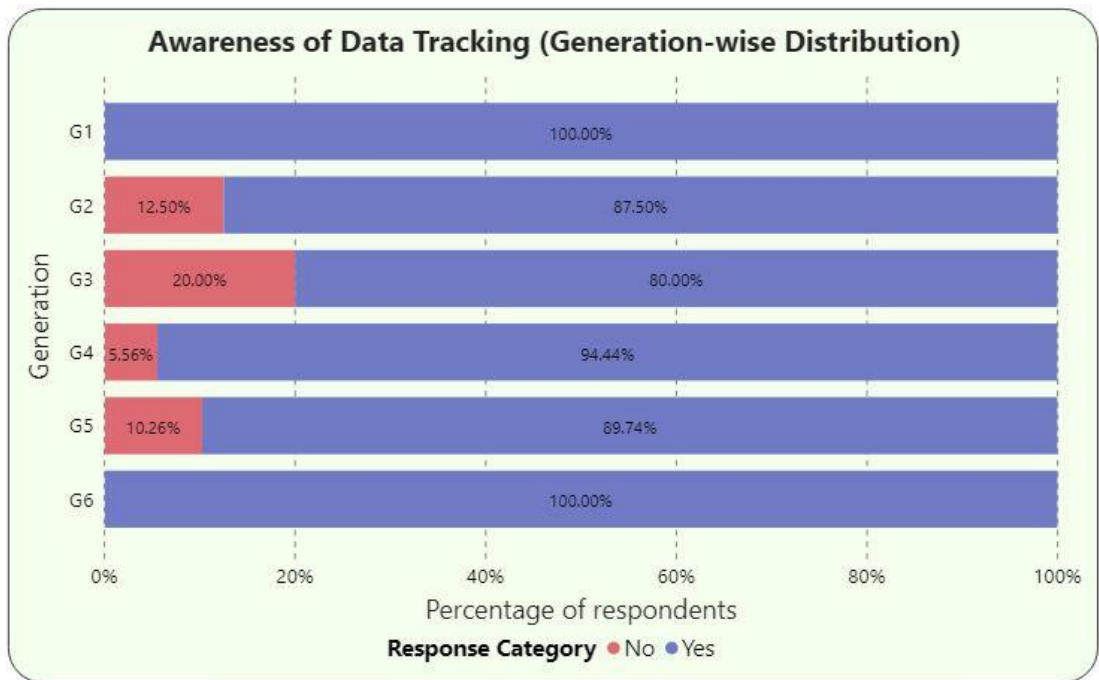


Figure 12: Awareness of Data Tracking (Generation-wise Distribution)

Figure 13 shows the Gender-wise distribution of the awareness levels of respondents. The findings that can be seen from the visual are as follows:

- The 95% of males were aware that companies have been tracking them for personalizing marketing experience for them.
- The maximum number of unaware respondents belonged to the female category.
- The finding that 100% respondents belonging to category other than male or female, have awareness of data tracking can't be attributed to their gender because the number of respondents belonging to this category is extremely low as compared to the whole sample.

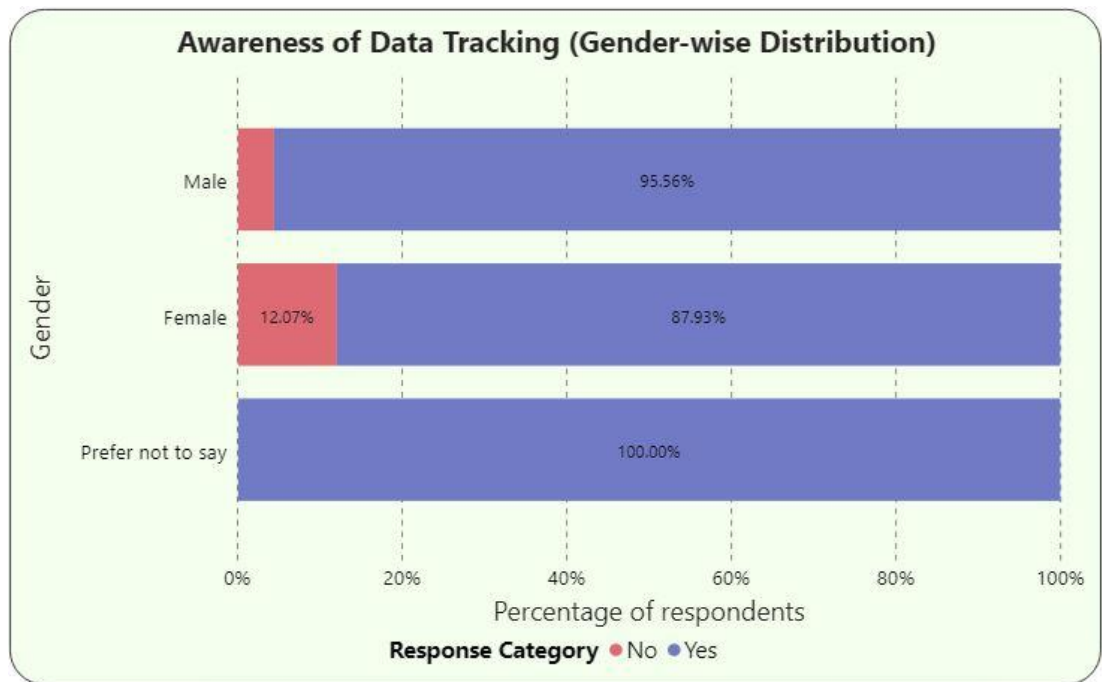


Figure 13: Awareness of Data Tracking (Gender-wise Distribution)

Figure 12 shows the Profession-wise distribution of the awareness levels of the respondents. The findings from this visual are as follows:

- The unaware respondents belonged to the profession of category P1 (IT Professional), P2 (Marketing Professional), P8 (Student), P9 (Housemaker).
- The maximum number of unaware respondents belonged to the category of profession P9 which is the category belonging to the housemakers.
- It is somewhat expected that students and housemakers may not be aware of data tracking but is highly unexpected that respondents working in IT or Marketing are not aware that companies are collecting the data to provide customized marketing because the collaboration of marketing and IT skills provides such customized solutions.

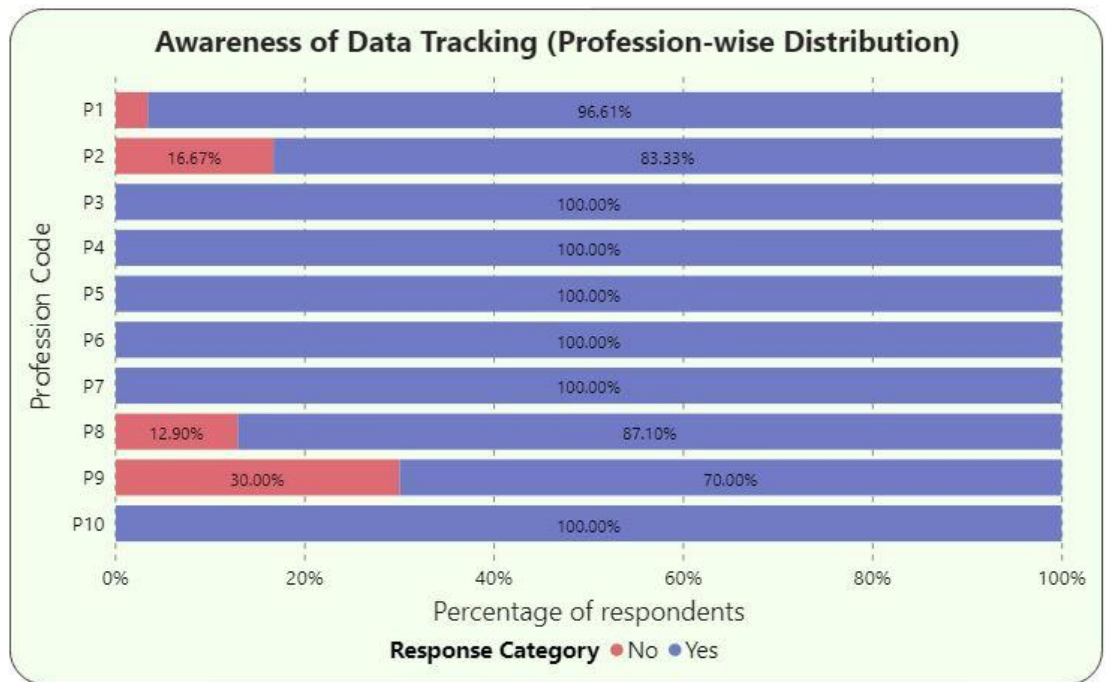


Figure 14: Awareness of Data Tracking (Profession-wise Distribution)

II. *Comfort level of respondents in sharing different types of data with the companies for personalized marketing:*

There have been 5 levels of comfort defined for this study. Level 1 means *not at all comfortable* while level 5 means *extremely comfortable*.

The type of data has also been divided into 3 categories:

- *Personal Data:* This category covers the basic details of the users such as: name, age, gender, email id etc.
- *Online Activities:* This category covers the data related to the activities done by a user on the internet, such as the data collected through browsing history, cookies, fingerprinting etc.
- *Financial Data:* This category includes data related to a person’s income, EMIs, banks where a user is having the account etc.

Figure 13 shows the comfort level of respondents for sharing different types of data that is collected by the companies. The following insights can be generated from this visual:

- This can be seen that respondents are somewhat comfortable in sharing their personal data.
- The comfort level decreases a little when they have to share the track of their online activities.
- There is a large reduction in the comfort level when the respondents have to share their financial data. As can be seen that for financial data a lot of respondents have responded that they are not at all comfortable (Level 1) in sharing this type of information.

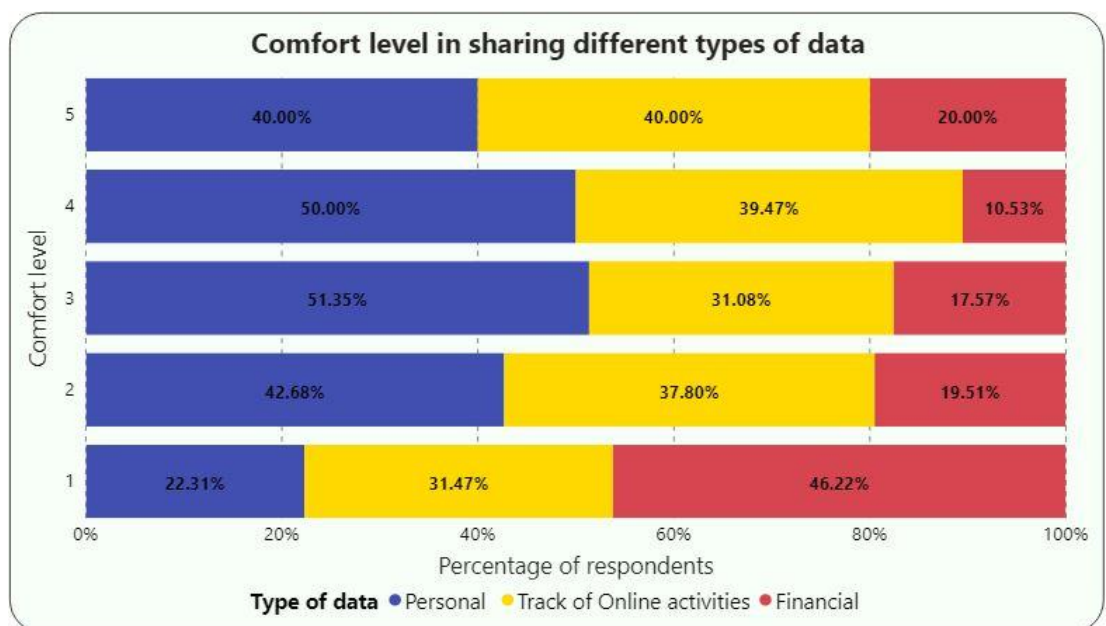


Figure 15: Comfort level in sharing different type of data

III. Awareness about the data protection bill among the respondents according to their demographic profile:

Figure 14 shows the generation-wise distribution of awareness levels of the respondents about data protection bill. Following insights can be generated from this visual:

- There are 80% of respondents belonging to generation G3, who are confirm that India has a data protection bill.
- All of the respondents from G1 category are aware about data protection bill but this category is only 1.33% of the total sample (see figure 3)

- If data of G4 generation is looked at (category which has the maximum number of respondents in the sample), majority of the respondents were either confirmed or partially confirmed that there is a data protection bill.
- For G5, the number of unaware respondents is least.
- Looking at an overall picture, it can be concluded that respondents are somewhat aware about the existence of data protection bill in India.

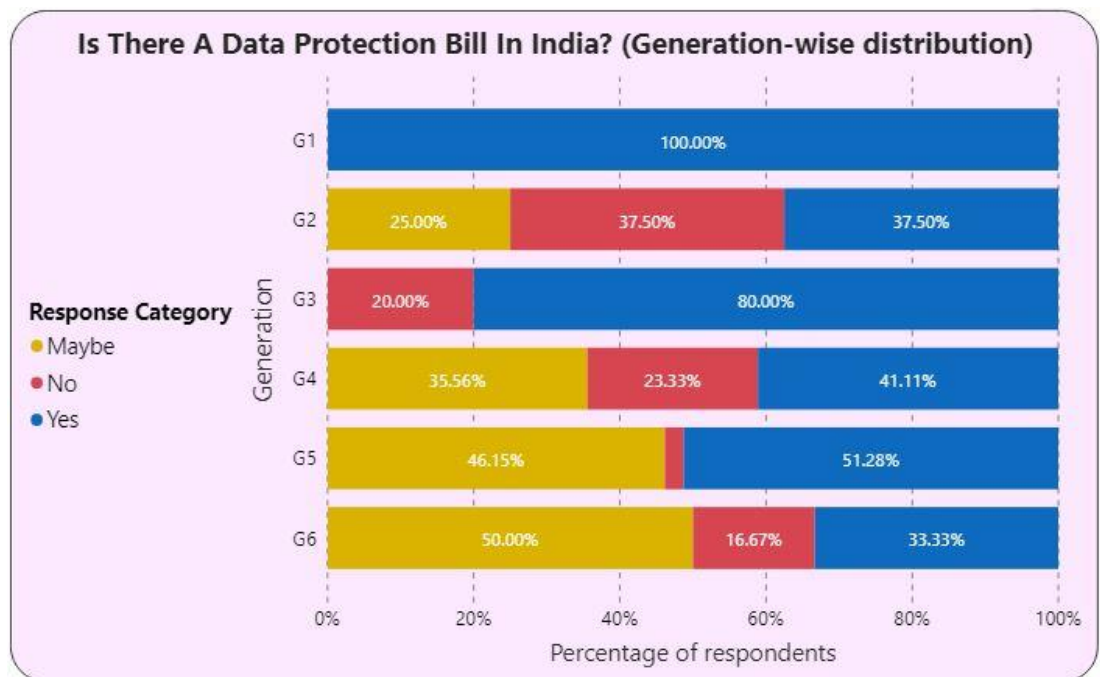


Figure 16: Awareness about Data Protection Bill (Generation-wise Distribution)

Figure 15 shows the gender-wise distribution of awareness levels of the respondents about data protection bill. From the visual, it can be seen that there had been almost similar results of awareness within males and females, wherein females appear to be more aware about the bill's existence as compare to males. The data for respondents belonging to category other than male or female can't be used to draw in a conclusion because the sample comprises of only 1.33% of respondents belonging to this category. However, it can be concluded that gender-wise distribution also shows that respondents were somewhat aware about the data protection bill irrespective of their gender.

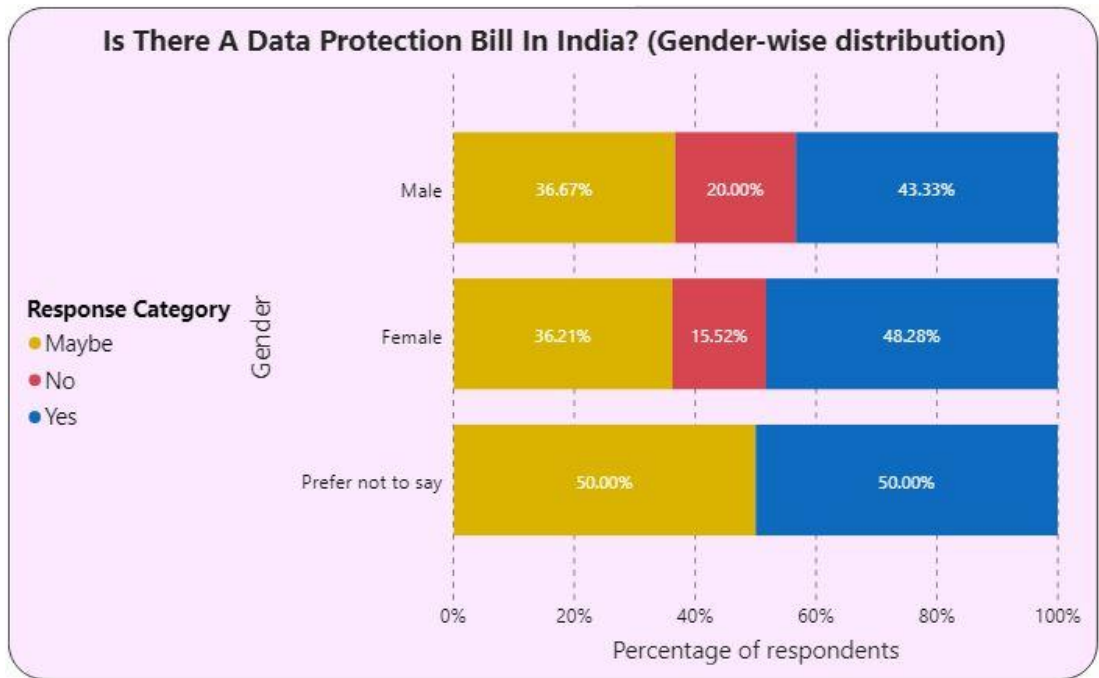


Figure 17: Awareness about Data Protection Bill (Gender-wise Distribution)

Figure 16 shows the profession-wise distribution of awareness levels of the respondents about data protection bill. It can be seen in the visual that profession-wise distribution of aware and unaware respondents is different from what was seen in the previous two visuals. This distribution shows that people under category P4 (Businessman/Entrepreneur) are the most unaware about the existence of data protection bill. However, the respondents of this category are expected to be very well aware about such bill as they have been dealing with the customer data a lot but there was not a single respondent in this category who was confirm about the existence of this bill. Also, it can be seen that number of unaware respondents were the lowest among the students (P8). Also, the highest number of confirmed respondents belonged to the P5 (Accounting/Finance professional) category. However, the overall picture for this distribution also says that respondents are somewhat aware about the existence of legal rights for data protection.

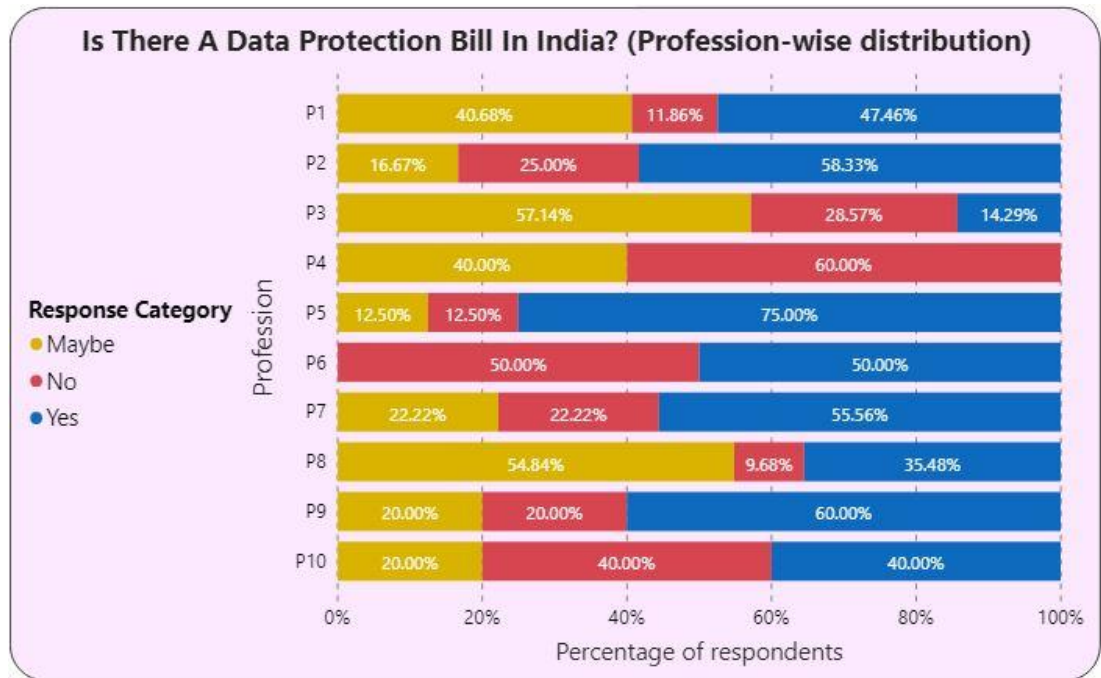


Figure 18: Awareness about Data Protection Bill (Profession-wise Distribution)

4.2.3. Predictive Analytics:

In this phase of analysis, decision tree model has been used to identify the segments of customers that will consider sharing the data for personalized marketing purpose.

Before getting into the discussion of the analysis, there has to be a clear definition of the problem statement that has been tried to solve using predictive analytics.

Problem statement:

To find out the traits of those respondents who will consider sharing their data for personalized marketing and those who will not consider it based on the type of data asked, i.e. personal, financial and online activities. Also, find out the probability that respondents will consider sharing data for each case.

Analysis:

For the purpose of building the decision tree for the scenario, the gender of the respondents was also coded as follows:

- 0 – Male
- 1 – Female
- 2 – Prefer Not to Say

3 decision trees were generated for each of the following categories of the data: Personal, Financial, Online Activities.

I. *Personal Data:*

The decision tree model in this case has generated 11 terminal nodes. Each of the node has following findings:

Node	Traits				Probability
	Age	Gender	Comfort	Profession	
7	G3-G6	0	1	P4, P5	0.75
8	G3-G6	0	1	P2, P8	0.25
9	G3-G6	0	> 1	P2, P4, P5, P8	0.833
10	G3-G6	0	1, 2	P1, P10, P3, P7	0.273
11	G3-G6	1, 2	1,2	-	0.273
12	G1-G2	-	1,2	-	0
14	G1-G3	-	> 2	-	1
16	G5	-	> 2	-	0.571
18	G4, G6	-	> 2	P2, P4, P5, P6, P10	0.75
20	G4, G6	-	> 2	P1, P7	0.444
21	G4, G6	-	> 2	P8, P9	0

Figure 19: Results of Decision Tree for Personal Data

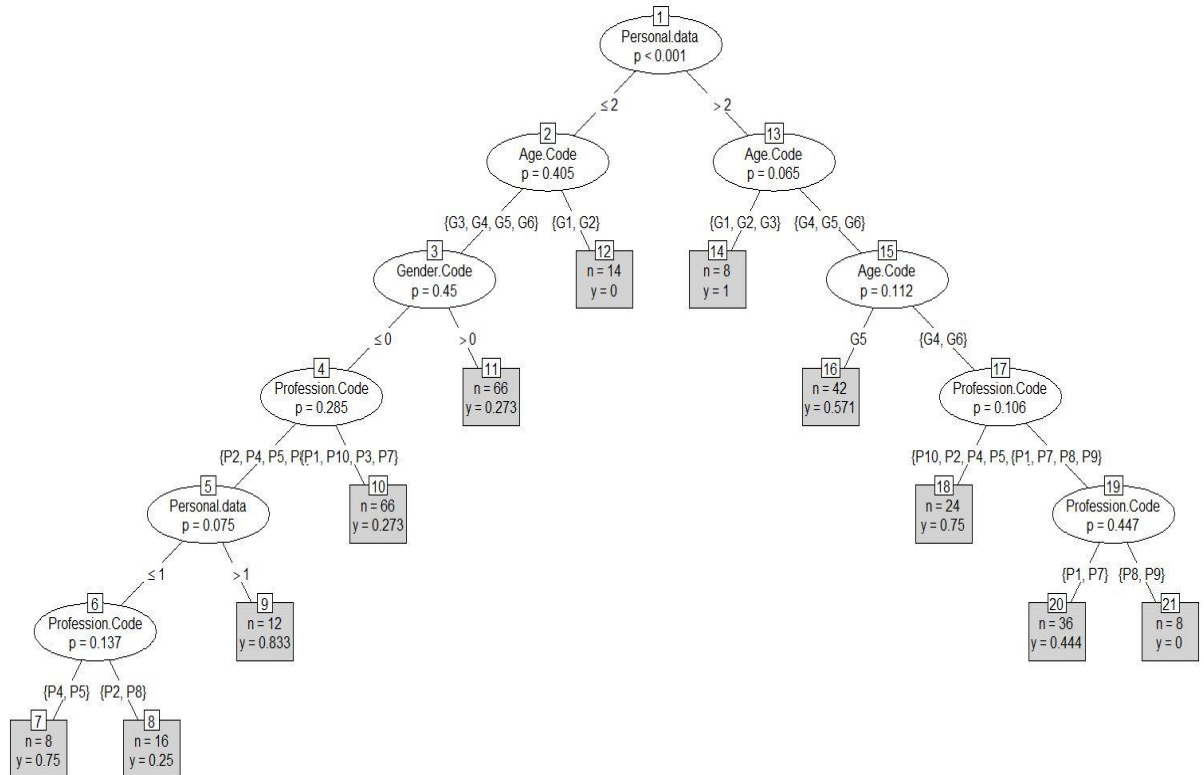


Figure 20: Decision Tree for Personal Data

II. Financial Data:

The decision tree model in this case has generated 9 terminal nodes.

Each of the node has following findings:

Node	Traits				Probability
	Age	Gender	Comfort	Profession	
4	-	0	1	P2, P5, P6, P8, P10	0.6
7	G5	0	1	P1, P3, P4, P7	0.6
8	G4	0	1	P1, P3, P4, P7	0.333
9	G1-G3	0	1	P1, P3, P4, P7	0
10	-	1, 2	1	-	0.271
13	-	0	1, 2	-	0.25
14	-	0	> 2	-	0.5
16	-	1. 2	> 1	P1, P5, P8, P9	1
17	-	1, 2	> 1	P2, P6	0.5

Figure 21: Results of Decision Tree for Financial Data

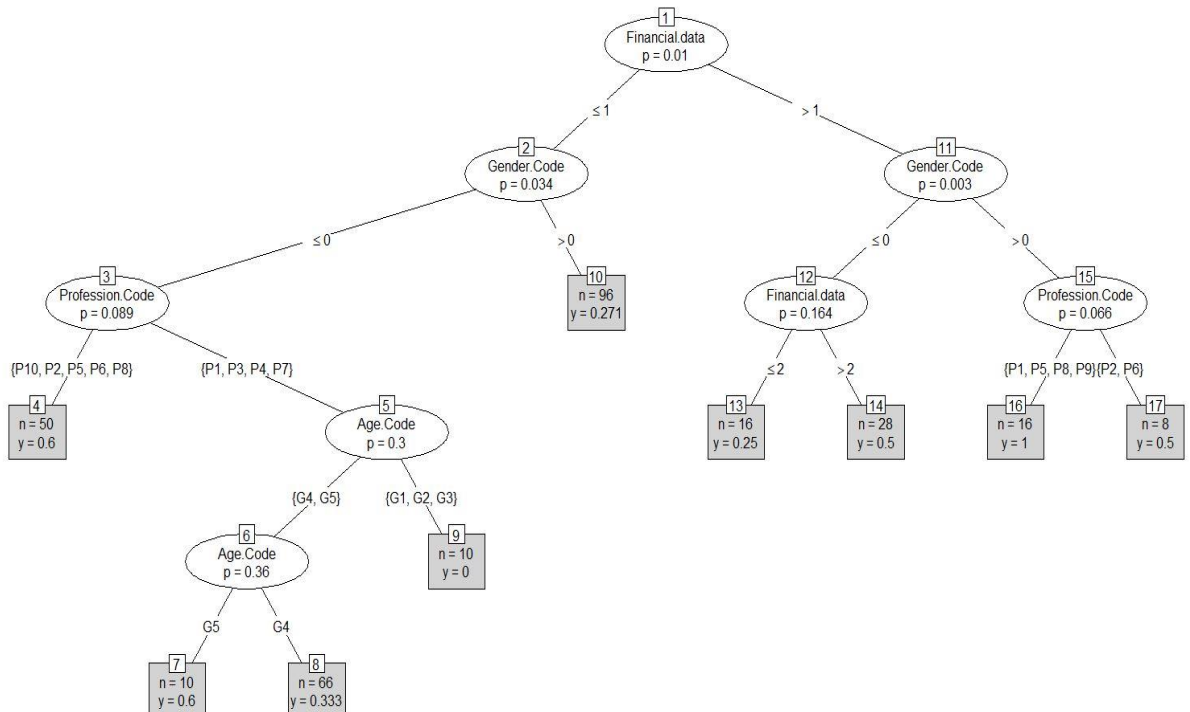


Figure 22: Decision Tree for Financial Data

III. Online Activities:

The decision tree model in this case has generated 10 terminal nodes. Each of the node has following findings:

Node	Traits				Probability
	Age	Gender	Comfort	Profession	
6	G3-G5	0	1	P2, P4, P5, P8	0.6
7	G3-G5	0	2	P2, P4, P5, P8	0.6
8	G3-G5	0	1, 2	P1, P3, P7, P10	0.333
11	G3, G4	1, 2	1	-	0.271
12	G5	1, 2	1	-	0.25
14	G3-G5	1, 2	2	P8	0.5
15	G3-G5	1, 2	2	P1, P5, P6, P9	1
16	G1, G2, G6	-	1, 2	-	0.5
18	-	-	> 2	P2, P5, P6, P7, P8, P9, P10	0.762
19	-	-	> 2	P1, P4	0.368

Figure 23: Results of Decision Tree for Online Activities

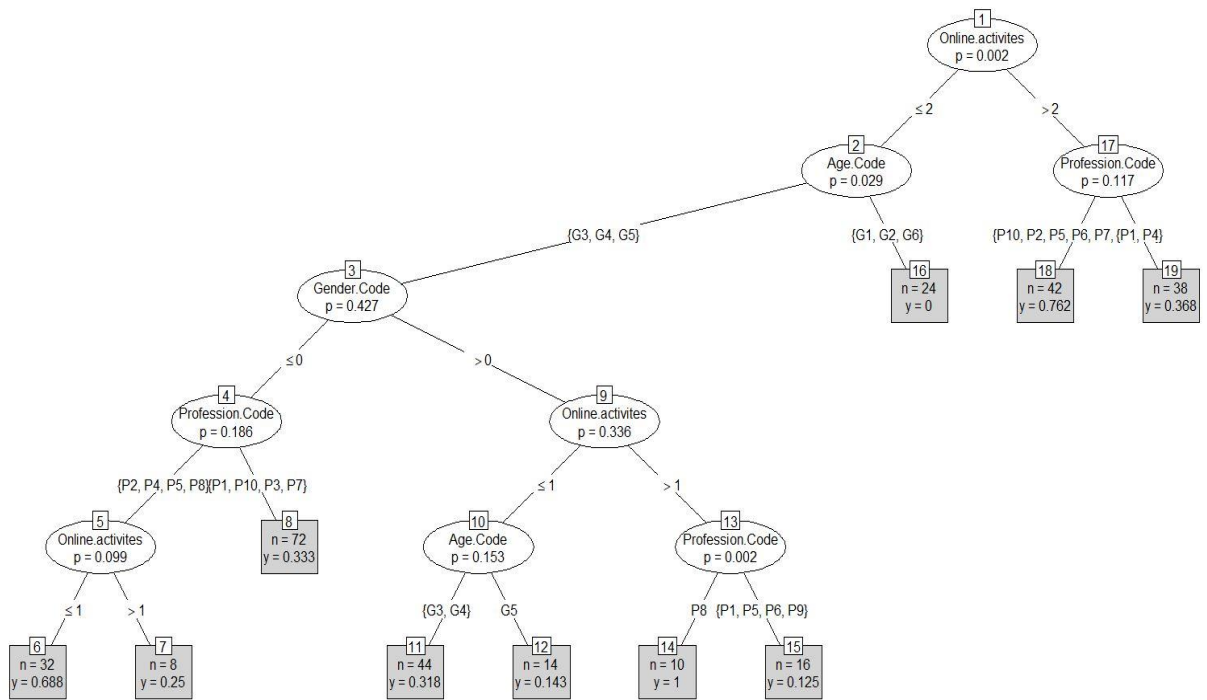


Figure 24: Decision Tree for Online Activities

4.3. Major Findings

The major findings from the complete analysis are:

1. 80% or more respondents from each generation category were aware that their data is being collected over the internet.
2. The highest number of unaware respondents belonged to generation G3 (35-44 Years)
3. More than 87% of respondents from each gender category were aware of the data tracking.
4. The highest number of unaware respondents belonged to female category.
5. The unaware respondents belonged to the P1 (IT Professional), P2 (Marketing Professional), P8 (Student), P9 (Housemaker) categories of the profession with highest numbers in the P9 category.
6. It is highly unexpected that respondents working in IT or Marketing are not aware that companies are collecting the data to provide customized marketing because the collaboration of marketing and IT skills provides such customized solutions.
7. The level of comfort in sharing the data is more when the data asked for belongs to personal category while the comfort level is less for financial data.
8. The respondents of different generation and gender categories are somewhat aware about the existence of data protection bill in India.
9. The respondents falling under category P4 (Businessman/Entrepreneur) are the most unaware about the existence of data protection bill which is quite unexpected due to the reason that they must have to deal with a lot of customer related data.
10. Students were among the least unaware respondents while Accounting/Finance professionals were among the most aware respondents for awareness about existence of data protection bill.
11. There is a high probability (more than 80%) that respondents with following traits will consider sharing their personal data:
 - Comfort level > 2 and Generation G1 or G2 or G3

- Comfort level > 1, Gender 0, Generation G3, or G4 or G5 or G6 and Profession P2 or P4 or P5 or P8
12. There is a very less probability (less than 25%) that respondents with following traits will consider sharing their personal data:
- Comfort level 1 or 2 and Generation G1 or G2
 - Comfort level > 2, Generation G4 or G6, Profession P8 or P9
 - Comfort level 1, Gender 0, Generation G3, or G4 or G5 or G6 and Profession P2 or P8
13. There is some probability (more than 60%) that respondents with following traits will consider sharing their financial data:
- Comfort level 1, Gender 0 and Profession P2 or P5 or P6 or P8 or P10
 - Comfort level 1, Gender 0, Generation G5 and Profession P1 or P3 or P4 or P7
 - Comfort level > 1, Gender 1 or 2 and Profession P1 or P5 or P8 or P9
14. There is a very less probability (less than 25%) that respondents with following traits will consider sharing their financial data:
- Comfort level 1 or 2 and Gender 0
 - Comfort level 1, Gender 0, Generation G1, or G2 or G3 and Profession P1 or P3 or P4 or P7
15. There is some probability (more than 76%) that respondents with following traits will consider sharing their online activities:
- Comfort level 2, Gender 1 or 2, Generation G3 or G4 or G5 and Profession P1 or P5 or P6 or P9
 - Comfort level > 2, and Profession P2 or P5 or P6 or P7 or P8 or P9 or P10
16. There is a very less probability (less than 25%) that respondents with following traits will consider sharing their online activities:
- Comfort level 1, Gender 1 or 2, Generation G5

CHAPTER 5: CHALLENGES AND LIMITATIONS

5.1.Challenges encountered

The following challenges were encountered during this study:

- Interviewing people for qualitative analysis during lockdown.
- Finding the diverse range of respondents for data collection during lockdown.
- Decision tree analysis requires a large dataset to generate good results. Data collection to keep sample size to around 300 during lockdown using only internet as the medium to reach out to people was a challenge in itself.

5.2.Limitations of the study

This study has following limitations:

- The sampling technique used (convenient sampling) would have introduced some sort of bias as there are sample imbalances as can be seen in the demographic properties of the sample (see section 3.3.4).
- The data may not be free from any respondent specific bias such as
 - Respondent may not be honest in answering the questionnaire.
 - Respondent's answers may be influenced by current state of emotions.
 - Respondent may have provided the response without even understanding the question being asked properly.

There can be more of such limitations.

- The results generated in predictive analytics may not generate similar results when applied to even larger dataset. For better prediction models, very large datasets are required.

CHAPTER 6: CONCLUSION AND RESEARCH IMPLICATIONS

6.1. Conclusion

After all the analysis following conclusions can be made about user's perception towards data collection by companies for providing personalized marketing solutions:

- Users are quite aware of web tracking. Also, it might be possible that the female users of age group 35-44 years who are housemaker are most likely to be unaware of this.
- Users may somewhat be comfortable to share personal data or online activities but not financial data.
- In general, users are reluctant to share their financial information with the companies even though they receive extraordinary benefits because of that.
- Users are more sensitive towards their financial data than any other data.
- There might be very few users who are extremely comfortable in sharing their data for personalized marketing. The normal comfort level lies somewhere in between.
- Some users might be aware of the existence of the data protection bill, but users need to have more awareness about their legal rights with respect to data protection.
- The personality as well as demographic traits play a role in determining whether a user will consider sharing his data for personalized marketing experience or not.
- The comfort level of a person plays an important role in determining whether a user will consider sharing his/her data with the company for personalized marketing experience or not.
- The probability that a user will consider sharing the data can be represented using the relationship:

$$P(\text{Personal Data}) > P(\text{Online Activities}) > P(\text{Financial Data})$$

where, $P(x)$ = Probability/Chance that user will consider sharing the data belonging to category x.

6.2. Research Implications

This research can be helpful in following scenarios:

- For marketers who are planning to move to digital marketing from traditional marketing.
- For digital marketers who are planning to provide customization experience to their customers.
- For marketers to understand the traits of the market and find out their target market for providing personalization experience.

With the current growing trend of personalized marketing it is important for companies to stay ethical while indulging in using and sharing user's data. This is study will help marketers to understand their customer's perception in a better way and hence they can plan their marketing strategies accordingly.

REFERENCES

1. Princiya (2018, Apr 23). Web Tracking: What You Should Know About Your Privacy Online. Retrieved from: <https://www.freecodecamp.org/news/what-you-should-know-about-web-tracking-and-how-it-affects-your-online-privacy-4293535525/>
2. Here's Why the Internet Can Sometimes 'Read Your Mind'. Retrieved from: <https://whatismyipaddress.com/web-tracking>
3. How do websites track users?: Technologies and methods: CCPA and GDPR compliance. Retrieved from: <https://www.cookiebot.com/en/website-tracking/>
4. Crawford, Eliza (2019, Oct 16). Website Tracking: Why and How Do Websites Track You? Retrieved from: <https://www.cookiepro.com/blog/website-tracking/>
5. Hudson, Elissa (2019, Feb 25). How to Blend Web Analytics and Digital Marketing Analytics to Grow Better. Retrieved from: <https://blog.hubspot.com/marketing/digital-marketing-analytics>
6. Sherman (2019, Oct 8). Understanding the Importance of Digital Marketing Analytics. Retrieved from: <https://www.lyfemarketing.com/blog/digital-marketing-analytics/>
7. Balla, John (2014, Sep 17). Seven benefits from using marketing analytics. Retrieved from: <https://blogs.sas.com/content/customeranalytics/2014/09/17/seven-benefits-from-using-marketing-analytics/>
8. Bujlow, Tomasz & Carela-Español, Valentín & Solé-Pareta, Josep & Barlet-Ros, Pere. (2015). Web Tracking: Mechanisms, Implications, and Defenses. Retrieved from: https://www.researchgate.net/publication/280590332_Web_Tracking_Mechanisms_Implications_and_Defenses
9. Session Tracking. Retrieved from: <https://www.cs.tut.fi/lintula/manual/java/tutorial/servlets/client-state/session-tracking.html>
10. Rouse, Margaret. session ID. Retrieved from: <https://searchsoftwarequality.techtarget.com/definition/session-ID>
11. Robie, Jonathan and Research, Texcel. What is the Document Object Model? Retrieved from: <https://www.w3.org/TR/WD-DOM/introduction.html>
12. Rouse, Margaret. Flash cookie. Retrieved from: <https://whatis.techtarget.com/definition/Flash-cookies>
13. Zawadziński, Maciej and Wlosik, Michal. What Is Device Fingerprinting And How Does It Work? Retrieved from: <https://clearcode.cc/blog/device-fingerprinting/>
14. Browser Fingerprinting. Retrieved from: <https://pixelprivacy.com/resources/browser-fingerprinting/>

15. What is Data Leakage? Data Leakage Defined, Explained, and Explored. Retrieved from: <https://www.forcepoint.com/cyber-edu/data-leakage>
16. Manage Cookies. Retrieved from: <https://www.allaboutcookies.org/manage-cookies/opt-out-cookies.html>
17. Hoffman, Chris (2017, Jul 5). How Private Browsing Works, and Why It Doesn't Offer Complete Privacy. Retrieved from: <https://www.howtogeek.com/117776/htg-explains-how-private-browsing-works-and-why-it-doesnt-offer-complete-privacy/>
18. Parchisanu, Daniel (2018, Nov 22). How to enable Do Not Track (DNT) in Chrome, Firefox, Edge, Opera and Internet Explorer. Retrieved from: <https://www.digitalcitizen.life/enable-do-not-track-dnt-chrome-firefox-edge-opera-internet-explorer>
19. Dalmia, Vijay Pal (2017, Dec 13). India: Data Protection Laws In India - Everything You Must Know. Retrieved from: <https://www.mondaq.com/india/data-protection/655034/data-protection-laws-in-india--everything-you-must-know>
20. The Personal Data Protection Bill, 2019. Retrieved from: <https://www.prsindia.org/billtrack/personal-data-protection-bill-2019>
21. Pinto, Saloni Lerisa (2018) Privacy and Data Protection: A Study on Awareness and Attitudes of Millennial Consumers on the Internet - An Irish Perspective. Retrieved from: <http://trap.ncirl.ie/3386/>
22. Nair, Aswini K and S.S., Vidhya (2016, July). A survey on importance of user awareness of web tracking. Retrieved from: https://www.ijarse.com/images/fullpdf/1467900473_11_Research_Paper.pdf
23. Bhat, Adi. Descriptive research: definition, characteristics, methods, examples and advantages. Retrieved from: <https://www.questionpro.com/blog/descriptive-research/>
24. McCombes, Shona (2019, May 15). Descriptive research. Retrieved from: <https://www.scribbr.com/methodology/descriptive-research/>
25. Hothorn, Torsten. Ctree. Retrieved from: <https://www.rdocumentation.org/packages/partykit/versions/1.2-7/topics/ctree>
26. Hothorn, Torsten. Ctree_control. Retrieved from: https://www.rdocumentation.org/packages/partykit/versions/1.2-7/topics/ctree_control
27. Blake, Christina (2018, Nov 29). How to reorder the legend in PowerBi. Retrieved from: <https://www.seerinteractive.com/blog/reorder-powerbi-legend/>
28. Microsoft Power BI Guided Learning. Retrieved from: <https://docs.microsoft.com/en-us/power-bi/guided-learning/>

ANNEXURES

ANNEXURE A: Summary of PI

ANNEXURE B: Questionnaire for Survey

ANNEXURE C: Plagiarism Report

(NOTE: The annexures have been appended at the end of this document.)

ANNEXURE - A SUMMARY OF PI

Questions	Responses				
	Candidate 1	Candidate 2	Candidate 3	Candidate 4	Candidate 5
Age Group	18-24 Years	45-54 Years	25-34 Years	35-44 Years	25-34 Years
Gender	Male	Female	Male	Male	Female
Profession	Student	Homemaker	Private job in banking sector	IT professional	Business
Average internet usage time	Around 12 hours	Around 6-7 hours	Around 12 hours	12-14 hrs	8-10 hrs
Purpose of using internet	Online games, Reading/Watching content of interest, Social Media	Social Media – FB, WhatsApp, YouTube. Searching Information.	Official purpose, online shopping, social media	Official work and Netflix	Official work, entertainment, online shopping
What all type of personalization has been experienced over various apps and websites?	Personalized content and adverts over social media	Personalized content on social media	Personalized content and adverts over social media and shopping apps	Personalized adverts	Personalized content over Netflix, Amazon Prime and shopping apps
Are you aware that companies are collecting your data for providing personalization experience?	Unconsciously aware of the fact that companies are collecting data and it is somewhat an invasion to the privacy	No	Yes	Yes	Yes
Are you comfortable in sharing your data to get personalized marketing experience?	Yes because the information is not too personal and can be shared easily with anyone.	Yes comfortable in sharing search history and language preferences etc.	Yes	No	Yes as long as there are not unnecessary notifications or calls
What kind of information can you share comfortably with the companies to receive personalized marketing experience?	Basic details like age, gender etc. along with area of interest.	Basic details that has been shared with consent.	Preferences, interests, basic personal detail like name, age, gender, email id etc.	Just the needed at that particular time, and confirmation that historical data will not be kept or used	Basic details like name, email id, age, gender etc.
Are you comfortable in sharing your financial data to get personalized offers or discounts?	No	No	No but can share PAN details etc on reliable portals like the one at workplace or for some govt. authority.	No	No
Are you aware how you can limit companies to track you over internet?	Some basic stuff like “not allowing” the cookies if a website asks.	No	Some basic details like if cache tracking asked by a company then will deny.	There are some standard tricks which one can use but beyond that it is more about companies keeping their word	Not much aware
Are you aware of there is any law in India for data protection? How much do you know about data protection policies in India?	Maybe yes. Didn't read about it.	Yes but don't know the details	Yes. Working on implementing the bill at the workplace. Fully aware of it.	Doesn't know much	Don't know anything about it.

ANNEXURE - B

User's Awareness and Perception on data collection by websites for personalized marketing

This questionnaire is created to conduct an academic study on how much users are aware about web tracking and what is their perception about companies collecting their information online to provide them with customized marketing experience.

Disclaimer: The data collected from this form will be used solely for the purpose of academic study and will not be shared with anyone.

* Required

1. Email address *

Demographics
Information

This section collects the demographic details of the respondents

2. Name

3. Gender *

Mark only one oval.

Female

Male

Prefer not to say

Other: _____

4. Age *

Mark only one oval.

- Under 18
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- Above 55

5. Profession *

Mark only one oval.

- IT professional
- Marketing professional
- Teacher/Lecturer
- Businessman/Entrepreneur
- Medical professional
- Accounting/Finance professional
- Public services (Govt.)
- Engineering services (other than IT)
- Student
- Homemaker
- Other: _____

Web Usage Information

This section collects the internet usage related information of the respondents

6. How much time in a day you spend online? *

Mark only one oval.

- Less than an hour
- 1 to 3 hours
- 3 to 5 hours
- 5 to 8 hours
- More than 8 hours

7. For what purposes you use internet? *

Check all that apply.

- For connecting with friends and families (social networking)
- For professional purposes (emails, meetings, working remotely etc.)
- For entertainment or passing time (watching movies/series, listening to songs, reading articles etc.)
- For online shopping
- For playing online games
- For searching different information (food recipes, job search, places for vacations etc.)
- For educational purpose (studying for exams, preparing content for a lecture, conducting online class etc.)
- For digital payments/transactions (UPI, netbanking, e-wallet etc.)

Other: _____

8. Which of the following devices you use for internet access? (Select which all are applicable) *

Check all that apply.

- Smartphone
- Tablet
- Laptop/Desktop

Exposure to
personalized content
over web

This section tries to understand how much exposure (knowingly/unknowingly) the user has to personalized content over web

9. Have you experienced any of the following? (Select what all applicable) *

Check all that apply.

- Saw an ad/content on FB/YouTube/Google about something that you were searching on other apps
- Saw recommendations for relevant products over e-commerce app
- Saw recommendations of videos based on previously watched videos on Youtube/Netflix
- Saw google notification asking to review the place where you currently are or visited sometime back
- Received offers for online shopping that were not available for other people
- None of them

10. Do you know that websites or apps collect your data for the features mentioned above? *

Mark only one oval.

- Yes *Skip to question 17*
- No *Skip to question 11*

Understanding the perception of the user

11. Are you comfortable in sharing your data with the apps/websites if you get to know that it will be used to provide you special discounts and offers? *

Mark only one oval.

- Yes
- No
- Maybe

12. Are you comfortable in getting your app/website activities tracked for getting personalized content on it?

Mark only one oval.

- Yes
 No
 Maybe

13. Are you comfortable in sharing your data if you know that it will be shared to some other app for showing personalized ads? *

Mark only one oval.

- Yes
 No
 Maybe

14. Are you comfortable in sharing your financial data such as your salary, monthly emi, monthly bills etc if you know that you will be given special offers or discounts by the company using that information? *

Mark only one oval.

- Yes
 No
 Maybe

15. Are you comfortable in sharing your data with apps/websites if you are not receiving any special benefits like extra discounts from it? *

Mark only one oval.

- Yes
 No
 Maybe

16. Do you think we have an act/bill in India for user's data protection? *

Mark only one oval.

- Yes *Skip to question 21*
- No *Skip to question 30*
- Maybe *Skip to question 30*

Understanding the user's awareness

17. Out of the following, what all information do you think websites or apps are collecting? *

Check all that apply.

- Browser history
- Personal information entered into an app/website for registration
- Your location data
- Your activities on social media apps

Other: _____

18. Are you aware that you can choose to not share your information with the companies online? *

Mark only one oval.

- Yes
- No

19. Have you ever looked into what all permissions you are giving to an app that you have just installed? *

Mark only one oval.

- Yes
- No

20. Do you think we have an act/bill in India for user's data protection? *

Mark only one oval.

- Yes *Skip to question 21*
- No *Skip to question 24*
- Maybe *Skip to question 24*

Understanding of legal aspects of data protection

21. Do you know the name of the act that is meant for policies related to data protection in India? (If Yes, please write it in the other section) *

Mark only one oval.

- Yes
- No
- Maybe
- Other: _____

22. Have you read the act properly? *

Mark only one oval.

- Yes
- No
- Maybe

23. How well aware are you about the data protection policies in India? *

Mark only one oval.

- 1 2 3 4 5
-
- Not at all aware Very well aware
-

Understanding the user's comfort level

24. How comfortable are you with the personalized marketing by companies using the data collected from users? *

Mark only one oval.

1 2 3 4 5

Not at all comfortable Extremely comfortable

25. How much comfortable are you in sharing your personal data such as your location, birthdate, profession, qualifications etc.? *

Mark only one oval.

1 2 3 4 5

Not at all comfortable Highly comfortable

26. How much comfortable are you to get your online activities such as your google search, social media activities etc being tracked for receiving personalized content? *

Mark only one oval.

1 2 3 4 5

Not at all comfortable Extremely comfortable

27. How comfortable are you when you know that an app that you are using has shared your data with some other app for displaying you personalized ads from that app? *

Mark only one oval.

1 2 3 4 5

Not at all comfortable Extremely comfortable

28. How much comfortable are you in sharing your financial data such as your salary, monthly emi, monthly bill amounts etc. for receiving customized ads? *

Mark only one oval.

	1	2	3	4	5	
Not at all comfortable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Extremely comfortable

29. How much comfortable are you in sharing your data if you know that you are going to receive special discounts or offers because of it? *

Mark only one oval.

	1	2	3	4	5	
Not at all comfortable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Extremely comfortable

Understanding the user's attitude

30. After knowing that different websites are tracking your activities and collecting some personal data, how your online activities are going to change? *

Mark only one oval.

- No change, I am fine with sharing my information with the companies
- I am fine with sharing basic information but still be more cautious during online activities
- I will take necessary measures to aware myself and ensure my data is not collected without my consent
- I am not comfortable with sharing my data at all, i will stop my online activities immediately

This content is neither created nor endorsed by Google.

Google Forms

ORIGINALITY REPORT

8%

SIMILARITY INDEX

6%

INTERNET SOURCES

1%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1	docs.google.com Internet Source	2%
2	dspace.dtu.ac.in:8080 Internet Source	1%
3	Submitted to University of Leeds Student Paper	1%
4	kassmfawaz.com Internet Source	1%
5	Submitted to Universiti Teknologi MARA Student Paper	<1%
6	D.P. Goyal. "Information systems planning practices in Indian public enterprises", Information Management & Computer Security, 2007 Publication	<1%
7	Submitted to Institute of Technology, Tralee Student Paper	<1%
8	Submitted to University of London External System	<1%

9 Submitted to National Law School of India University, Bangalore
Student Paper <1%

10 ulspace.ul.ac.za
Internet Source <1%

11 eprints.utar.edu.my
Internet Source <1%

12 Submitted to Monash University
Student Paper <1%

13 tcc.ac.in
Internet Source <1%

14 Submitted to University College London
Student Paper <1%

15 Submitted to Symbiosis International University
Student Paper <1%

16 brage.bibsys.no
Internet Source <1%

17 Submitted to Ateneo de Manila University
Student Paper <1%

18 baadalsg.inflibnet.ac.in
Internet Source <1%

19 Submitted to University of the Highlands and Islands Millennium Institute <1%

20 pt.scribd.com <1 %
Internet Source

21 Submitted to Zionsville Community School District <1 %
Student Paper

22 Parul Jain Gupta, Pradeep Kumar Suri. "Analysing the Influence of Improved Situation, Capability Level of Actors and Flexible Process Workflow on Public Value of E-Governance Projects in India", Global Journal of Flexible Systems Management, 2018 <1 %
Publication

23 Submitted to London School of Business and Finance <1 %
Student Paper

24 id.123dok.com <1 %
Internet Source

25 Submitted to Higher Education Commission Pakistan <1 %
Student Paper

26 Submitted to University of KwaZulu-Natal <1 %
Student Paper

27 www.digitalcitizen.life <1 %
Internet Source

Exclude quotes On

Exclude matches < 10 words

Exclude bibliography On