Department of *Economics and Finance*

Chair of Marketing

Italian users' willingness to disclose sensitive data when browsing the Internet: an empirical research

Supervisor: Alberto Marcati

Candidate: Dalila Roma Toss
ID number: 215501

Academic year: 2019/2020

Um agradecimento especial à minha amada família, por serem meu porto seguro em todos os momentos da vida.

Table of Contents

Introduction

In the past two decades, worldwide concern regarding data privacy has been always subject of discussion since violation of users' data is a crime punishable by law The argument pertaining personal data protection and increasing awareness by individuals regarding issues linked to data violation, especially after events such as the Cambridge Analytica- Facebook scandal of leaked data, has led to the introduction of new laws that make up the foundation of data privacy. Recent events have led to the GDPR, which is a European regulatory framework that institutes rules regarding protection, circulation, and treatment of personal data of singular individuals in all EU countries.

Personal data protection is one of the features of the concept of privacy, and its relevancy acts as baseline when considering the progression of technology, with businesses established on the grounds of personal data assortment and big data collection and assessment. Data privacy is a relevant topic of discussion in the digital era because security breaches can expose personal data of millions of people and this can further develop into identity theft and blackmailing. Therefore, it is fundamental to find a balance between the necessity of personal data by companies and the right to seek privacy by users. Many studies concerning data privacy have been conducted in the past, and this thesis objective is to explore the implications of such theme in the Italian territory.

The aim of this thesis is to evaluate if Italians are conscious about their individual privacy when online surfing and to what extent Italian internet users are willing to self-disclose. By conducting a survey on a sample of 110 Italian internet users, I will verify the variables that affect Italian self-disclosure behavior and externalize the main findings that will emerge after the survey is conducted.

Chapter 1: ***Theoretical background***

In this chapter, I will analyze the concept of online data privacy and its theoretical background, by also assessing past works regarding the topic.

1.1 The importance of online data and the concept of privacy

In a digitalized world, data plays a significant role when assessing information. The majority of companies perform data analysis and evaluation when assessing their achievements and when they seek to upgrade their overall performance. The improvement of the enterprises' performance ultimately leads to greater market power and attract more shareholders who will have an appreciable interest to be linked to such enterprises. For that reason, one can affirm that data is nowadays the new currency of the globe, or "the new oil"[1], as it can be very valuable. Hence, online data is the corollary of radical changes in data treatment policies. Dissemination of information and online mass entertainment forms, such as social media platforms and online streaming services contribute to predict future regarding market trends. In order to meet consumers' demand with attentiveness, companies started to invest in online data analysis to optimize their sales and stimulate an efficient and effective work environment. By getting more acquainted with their target market through evaluation and correct interpretation of data, businesses are able to innovate to meet users' expectations accordingly. Furthermore, online data collection also contributes during the identification process of difficult-to-access population segments which otherwise would have been overlooked during the targeting process. During the creative process of a marketing campaign for instance, the awareness of the exact target group the company aims to attract is one of the most important features to take into account. Online data collection helps researchers and companies to gather information through samples encompassing thousands of people that depict the authentic representation of how the population as a whole behaves and makes choices.

[1] Knowledge@Wharton and Barratt, J. (2019) "Data as Currency: What Value Are You Getting?" [Online]. Available at: https://knowledge.wharton.upenn.edu/article/barrett-data-as-currency/

Understanding consumers' perspective in the decision-making process provides an overview of what works and what does not work out as planned by firms. As a consequence, data collection anticipate insights on what kind of innovation is more profitable and may settle down in the long run. The importance of online data analysis fundamentally changed the way in which businesses conduct their activities as it confers greater innovation levels thanks to the quicker acknowledgement of potential flaws and subsequent adjustments, optimization of schedules and cost-effectiveness of decisions. On top of that, online data study gives room for rigorous and competitive positioning, through which companies are well apprised of where they stand in the market and what sets them apart from their competitors.

Concurrently, data analysis prevents and minimizes financial risks the firms might face, in such a way to help firms to implement the right corporate decisions that assure to reach the desired objectives. All the strategies that could lead up to the success of a business are a result of a continuous analysis of factors that drive the competitiveness level of a particular firm, and therefore its market share. Through data assessment and risk management, managers and directors gain exclusive insights that contribute to road map the direction that the business should naturally follow and invest in.

Many scholars believe that technology is the driving force that changed firms' approach to data and made commercial markets more flexible, quick-moving and competitive. As the use of data plays a greater role in society's economic growth prospects, concerns about data privacy and security inevitably arise.

Privacy can be regarded as "the state of being alone, or the right to keep one's personal matters and relationships secret"[2], as defined by Cambridge dictionary. Broadly speaking, there are many ways to classify privacy according to different definitions. When defining privacy, one, for instance, could be referring to physical privacy or one could be referring to surveillance and information privacy, which is the category of privacy we are going to focus on this thesis. Before the advent of the internet, personal information could be more difficult to gather, but internet connection, web browsing, and social media usage have made humongous contribution to the way data is collected in modern times. Indeed, there are many interpretations and inferences attached to the concept of privacy that could be drawn and that may vary in different countries.

[2] Privacy. (n.d) In Cambridge Academic Content Dictionary. Available at: https://dictionary.cambridge.org/dictionary/english/privacy

Along with the proliferation of digital devices and the constant growth of the amount of information about consumers' habits that technology can now gather, it has never been more vital for companies to ensure that users' data were protected at all times. Beyond that, it is also fundamental that companies keep consumers informed and updated about how such data is used and to which extent they are being used. In order to make sure that consumers' feel safeguarded about their personal information, companies often adopt a transparency policy to guarantee they retain consumers' trust. In fact, companies must inform consumers about the reasons why data is being collected and how consumers can withdraw from assessment if they wish so. Indeed, transparency might invigorate data availability and quality, therefore behaving as a medium to obtain consumers' engagement.

In this digital era privacy can be viewed as a double-edged sword, as it is not always possible to conduct anonymization of data by disassociating personal data from individual profiles. There has been a drastic transition in the way that data is being seen, as in the present-day it is gathered by advanced databases and sold and resold to third parties. Personal data protection is one of the features of the concept of privacy, and its relevancy acts as baseline when considering the progression of technology, with many businesses established on the grounds of personal data assortment.

1.2 Big Data and processing of sensitive data

The term big data has its origins back in 1990s and it is used to describe a tremendous amount of data that once processed and interpreted, is able to produce revenue to businesses and organizations; its official definition according to Cambridge dictionary reads "very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods"[3]. International data corporation (IDC) estimates that the digital universe, which is composed by the data we create globally, will reach 44 zettabytes (44 trillion gigabytes) in 2020[4].

Nevertheless, big data is not only defined by its volume, which is greater than the traditional extent of storage, but also by other properties. According to Doug Laney, big data has the following characteristics, which are called the 3v's of big data[5]: volume, variety, and velocity*. Volume is associated to the quantity of information that is processed. For instance, clickstreams on a website or data arising from Facebook polls and quizzes add up to the volume of data that is elaborated. Velocity refers to the speed rate at which data is gathered and transferred. For example, many machines work in real time and produce immediate assessments. The concept of variety, instead, is attached to diverse types of data that are available, which can be structured and unstructured, such as texts, audio, videos, and files.

There is huge volume of data stemming from a variety of sectors being collected by enterprises to improve their profits; for instance, there is data being drawn out from climate, natural phenomena and disasters, surveillance, healthcare, IT, transportation, retail and so on and so forth.

Big data technology and tools help organizations to improve their revenue streams and to gain competitive advantage through technological devices, sensors, social media, web browsing; on

[3] Big Data. (n.d.) In Cambridge Advanced Learner's Dictionary and Thesaurus. Available at: https://dictionary.cambridge.org/dictionary/english/big-data

[4] "IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth Through 2022" (2019). Available at: https://www.idc.com/getdoc.jsp?containerId=prUS44998419

[5] Rezzani, A. (2018). Le tre V dei Big Data [Online]. Data skills understanding the world. Available at: https://www.dataskills.it/le-tre-v-dei-big-data/#gref

top of that, Internet of things (IoT) and Artificial intelligence (AI) contribute to generate new forms of complex and interrelated data which facilitate the automation of information by disclosing trends and social conduct.

There are diverse sources of big data deriving from differing sectors. Big data can arise from government institutions and database, and they encompass statistics regarding hospital records and results of medical examinations of the citizens, population census, electoral register etc. As a further matter, machines and equipment that make up the internet of things system (e.g. security systems, automobiles, sensors, smart electronic devices such as computers etc.) also play a significant role and are a big data game changer as they impressively augment the volume of such data. In addition, commercial data arise every time consumers interact and make transactions with organizations, which eventually collect data for performance measurement. Lastly, social media greatly affect big data as everyday meaningful information surface from user's pictures, posts, and likes. Namely, posts published on social networks like Facebook reveal information about personal preferences and are therefore exploited by Facebook's algorithm which will target users with a specific advertisement based off on what individuals have disclosed. As a consequence of large investments made by enterprises aimed at financing analytics and insights, revenues from big data and business analytics will be a cut above $270 billion by 2022, as forecasted by IDC[6]. Notably, social networks are massive beneficiaries of the data they trade with companies, to give an instance, 82% of Twitter's revenue in the second quarter of 2020 arise from advertising services promoted on the platform[7].

Over the last few years, many scholars have argued that the 3v's of big data are not the only ones to be taken into consideration when providing a definition to the term. Supposedly, there are two additional characteristics of big data that must be conveyed, one being veracity. Veracity refers to the appropriate data administration, and privacy apprehension and concerns. In the midst of the availability of so much information, it is necessary to identify the data's integrity and quality, meaning that reliable and trustworthy data are the pillars to the effective utilization of big data. In conjunction with veracity, value contributes as well to the denotation of big data. Value refers

[6] "IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth Through 2022" (2019). Available at: https://www.idc.com/getdoc.jsp?containerId=prUS44998419

[7] Reiff, N. (2020). How Twitter Makes Money [Online]. Investopedia. Available at: https://www.investopedia.com/ask/answers/120114/how-does-twitter-twtr-make-money.asp

to the precise value that big data generates to users and businesses. Big data has a capital value that can only be exploited if the approach to data collection is analytical and productive, meaning that among the myriad of information available it is possible to extrapolate only valuable information that will be strategically used to generate revenue growth for organizations. Collection of data in a chaotic manner without following a predetermined criterion can be confusing and could lead to misleading interpretation, which creates room for the implementation of wrong decisions. Therefore, it is fundamental to know how to correctly work with methodologies deriving from big data analytics in order to smooth the path of the decision-making process.

It is a given fact that big data affects organizations, but what about consumers? If we analyze the other spectrum of the topic, we can see that consumers can abundantly benefit from the correct use of big data: consumers preferences and demands are met in a more accurate manner, client relationship is more intimate and customized, and communication between consumers and firms is more straightforward. Various aspects of big data are determinants of how well data privacy is maintained. Security and living standard, for instance, are aspects that influence big data and that are also related to online data privacy. A person's web search history can reveal her personality, hobbies, political and religious preferences, among other things. Theoretically, big data is supposed to improve consumer's experience, however researchers agree that consumer's privacy could be compromised in the process as sensitive data privacy cannot always be ensured. In conclusion, with big data there also privacy and safety challenges that must be dealt by governments and corporations.

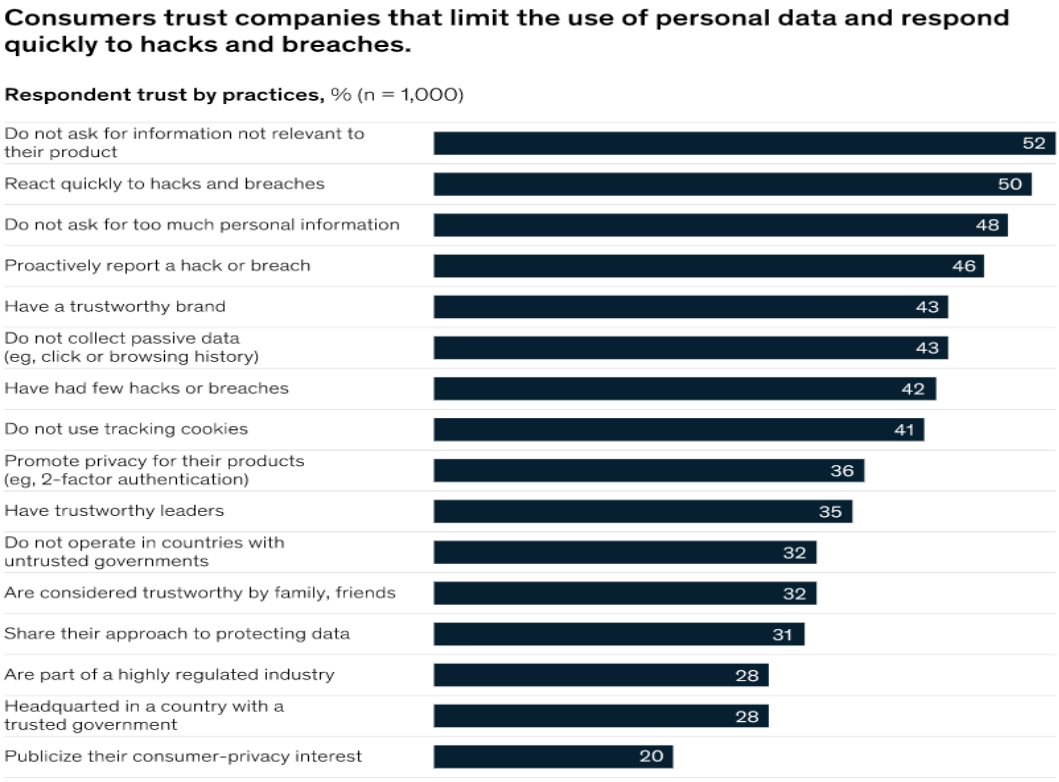
1.3 How data privacy breaches and threats affect users' choice

The absence of confidence by internet users to share their sensitive data – ethnic origin, political views, personal health information, sexual orientation, and geographic location, to name a few – is a consequence of the awareness of previous data breaches. Many websites and social media platforms have suffered data breaches throughout the digital age. To mention an example, a leaked data scandal, which was referred to as the "Facebook- Cambridge Analytica data breach" by the mainstream media, occurred when users' personal data were collected without consent for political purposes by Cambridge Analytica, a political consulting firm that . The breach affected millions of accounts and has supposedly affected United States' election strategies and outcomes in 2016.

But that was not the only occasion in which Facebook suffered data breach. In 2018, Facebook suffered a security breach through which hackers explored a bug found on the platform. The attackers had access to everything on a user's profile. Thus, it is a common practice of social media platforms to sell users' information to third parties and data breaches occur more often than not. For those reasons, many users' refrain from sharing information that they would not want to be publicly accessible if a breach occurred.

A recent study conducted by McKinsey & Co.[8] shows that users are becoming more mindful about what kind of data they share, and in particular, users are more prone to display information when such information treats healthcare and financial matters. However, when considering other online services sectors, users' trust levels are low. Moreover, an important repercussion of frequent data breaches is that one out of ten worldwide internet users avail themselves of softwares that impede companies from tracking their clickstream behavior and online activities. Ultimately, users are more inclined to trust a company's when the only type of information requested by the company is limited and relevant for the scope of the website.

[8] Anant, V., Donchak, L., Kaplan, J. and Soller, H. (2020) The consumer-data opportunity and the privacy imperative. McKinsey & Company. Available at: https://www.mckinsey.com/business-functions/risk/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative#



Source: McKinsey Survey of North American Consumers on Data Privacy and Protection, 2019

1.4 Existing regulations on the treatment of personal data- GDPR

With the rise of public awareness regarding the growing issue of data breach, internet users' demand for security and privacy have risen. Continuous episodes of data breaches have led to the implementation of new rules and regulations by Governments. For countries within the European Union, a regulation came into force, namely the General Data Protection Regulation (GDPR). The aim of this regulation is to set clear rules for all organizations working within the EU; it is a security law concerning data protection that allows citizens to have greater control of their personal data. Such laws regard privacy policies, informed consent, and sensitive data of the citizens. As technology progresses, it has become evident how inevitable episodes of data breaches are, and therefore this law ensures the legal collection of data. GDPR forbids unfair treatment and manipulation of data, and sets monetary sanctions if requirements are not observed by organizations.

1.5 Literature review on data privacy

Online data privacy and the associated risks attached to data breach is a newly common topic of debate because of the progression of the digital era. Nowadays, more and more people choose to engage with technology for multiple reason: work, academic reasons or simply leisure and entertainment. Regardless of the reason that attract internet users, there is a rising number of individuals that express concerns in relation to personal data processing, as it may comprehend sensitive information. Furthermore, some websites solicit personal information (e.g. date of birth, credit card number, mobile phone details, GPS coordinates etc.) and those requests influence customers' attitude towards consumption and data sharing. As reported by an EU research conducted in 2015[9], 3% of EU's users have declared that they felt that there had been some sort of online privacy violation which consequently led them to suffer a loss of money. Such violations often result in firms losing customers that no longer want to be associated with them and that deem them as untrustworthy. For those reasons, now more than ever, information and communication technologies involving online data have been under scrutiny of citizens and governments. Governments and organizations that collect data aim at avoiding any potential data breach and fraudulent conducts. Hence, this literature review objective is to investigate past research and papers regarding how worldwide users view online data privacy and their tendency to disclose personal information when surfing online to better understand consumers' behavior. As evidenced by multiple studies, the main risks that arise when one makes sensitive information known online are the unauthorized selling of such information, unwanted publicity and newsletter, and fraud.

According to Dinev and Hart, "Individuals make a tradeoff between the cost of disclosing personal information and the potential derived benefits".[10] Possible benefits could be obtained by customers in the form of coupons, free access to services, vouchers, gifts, customized gadgets

[9] "Safe surfing: A brief look on internet security" [Online]. Eurostat: Digital economy and society in the EU. Available at: https://ec.europa.eu/eurostat/cache/infographs/ict/bloc-3a.html

[10] Dinev, T. and Hart, P. (2003) An Extended Privacy Calculus Model for E-Commerce Transactions. Volume 17, Issue 1. Available at: https://doi.org/10.1287/isre.1060.0080

and so on. Indeed, it has been proved that individuals renounce to some degree of privacy by trading it off with what they believe is worth the threat of disclosing personal data. Thus, users' perception of vulnerability when revealing sensitive data and their trust or mistrust in a company merely depend on their attitude towards risks (and the perceived benefits), their past experiences, and their privacy literacy.

Moreover, past research has also pointed out how past online data privacy breaches affected worldwide consumer's behaviors when using the internet. As a matter of fact, data breaches are a frequent theme of discussion, as risks associated with breaches dissuade consumers from disclosing information online because of the existing chance of having their private data exploited or used in bad faith.

A 2003's study[11] highlighted the major factors that persuade the perceived consequences by consumers when online shopping, and those are cheaper prices, saving time, more efficient customer service, comparative shopping, and social influences.

Privacy violation does not seem to be a part of the perceived consequences by consumers when online shopping, although it is a driver of online purchasing because customers are more incline towards online shopping if they are not concerned about the dangers of data breaches.

Another survey conducted on the U.S. population[12] analyzed that after an episode of data breach, consumer's attenuation corresponded only to 11%. That is because the consequences of data loss and violation is not always clear to all consumers. On the other hand, many consumers demonstrate appreciation when enterprises that take accountability for data violations they go through. In fact, consumers continued to purchase or use services of such enterprises even after privacy scandals, and this may be a consequence of either customer loyalty or large costs of changing enterprise.

[11] Khalifa, M. and Limayem, M. (2003) Drivers of Internet shopping. Commun. ACM 46, 12, pp. 233–239. DOI: https://doi.org/10.1145/953460.953505

[12] Ablon, L., Heaton, P., Lavery, D.C., and Romanosky, S. (2016) Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information. Santa Monica, CA: RAND Corporation. Available at: https://www.rand.org/pubs/research_reports/RR1187.html.

Similarly, as outlined in Grabner-Kraeuter's article[13], in the course of time trust will be the definitive key for the growth or shutdown of e-commerce, and it is fundamental for companies to muster and retain consumer's trust. Since consumers are able to pick from myriad of options when purchasing goods and services online, and because of the broad competition retailers face on a regular basis, ensuring consumer's trust is an optimal strategy to boost sales. In fact, consumers are constantly looking for mechanisms that diminish uncertainty during the purchasing process, and of those mechanisms is trust. In conclusion, all the efforts attached to ensuring the security of clients will be beneficial to both consumers and retailers engaging in online commerce.

Ultimately, in past readings it has been made clear that there exists a relationship between online data and consumers' behavior.

As Adina Radulescu denoted in her reading[14], "promotion and prevention-associated behaviors concerning personal data adjust to the tendency of the antecedents." The antecedents under consideration which the author avails herself of are privacy violation and perceived control of personal data, which are mitigated by privacy comprehension and confidence in social media websites. Thus, internet user's behavior and their propensity to divulge information is strictly dependent to their knowledge on data privacy risks and awareness that what they are divulging might be used by third parties. For the reasons mentioned above, it is possible to identify a link between data privacy and consumers' behavior.

After an assessment of what has been studied and argued by scholars in this review, in the next chapter I will delineate new hypotheses which involve new further research regarding data privacy and consumers' behavior. Such hypotheses will lead to the development of a questionnaire that will be compiled by Italian internet users; thus, I will narrow down the geographic field of research restricting it to the Italian territory only.

[13] Grabner-Kraeuter, S. The Role of Consumers' Trust in Online-Shopping. Journal of Business Ethics 39, 43–50 (2002). https://doi.org/10.1023/A:1016323815802

[14] Rădulescu, Adina (2018). "Users' Social Trust of Sharing Data with Companies: Online Privacy Protection Behavior, Customer Perceived Value, and Continuous Usage Intention," Contemporary Readings in Law and Social Justice 10(1): 137–143.

Chapter 2: ***Research methodology***

From the theoretical background of the first chapter, it became evident that data privacy is a compelling topic that has been discussed by many scholars. Past research has been conducted from different countries and has taken into consideration different models and variables, which were measured using different scales. Therefore, following the guidelines of previous studies and models, this thesis will further explore the theme of internet users' behavior for what concerns data privacy, and in particular, it will focus on the Italian market.

The aim of this thesis is to investigate the driving factors that shape Italian users' attitude when having to disclose sensitive information online. In fact, previous studies attest that not all internet users attach the same level of concern towards data sharing. A second purpose of this paper is understanding to what extent Italian internet users are willing to disclose sensitive data. In studies conducted in other countries, it appeared that most people are more outraged with the illegal use of data, rather than being concerned with data-disclosure per se. In the following sections of this chapter, I will explain the structure of research model and the structure of the survey that will be conducted on a sample of Italian internet users. Moreover, I will discuss the hypotheses thanks to which the survey will be drawn upon. In addition, I will define the variables of the research model by distinguishing single-item from multi-item variables and stating the reasons why they are essential to conduct this research.

The research question is:

Are Italians interested in the safeguard of their individual privacy when online surfing and to what extent are Italian internet users willing to self- disclose?

2.1 The model

A survey will be presented to a sample of 110 Italian users. The survey is composed by 20 questions aimed at investigating Italian consumers' self-disclosure behavior when browsing on the web. The variables that respondents are going to value in the survey and that will be used to evaluate the hypotheses which I will develop later on are the following: website reputation, data privacy risk concerns, information asymmetry, perceived severity of a data leakage, and user's trust.

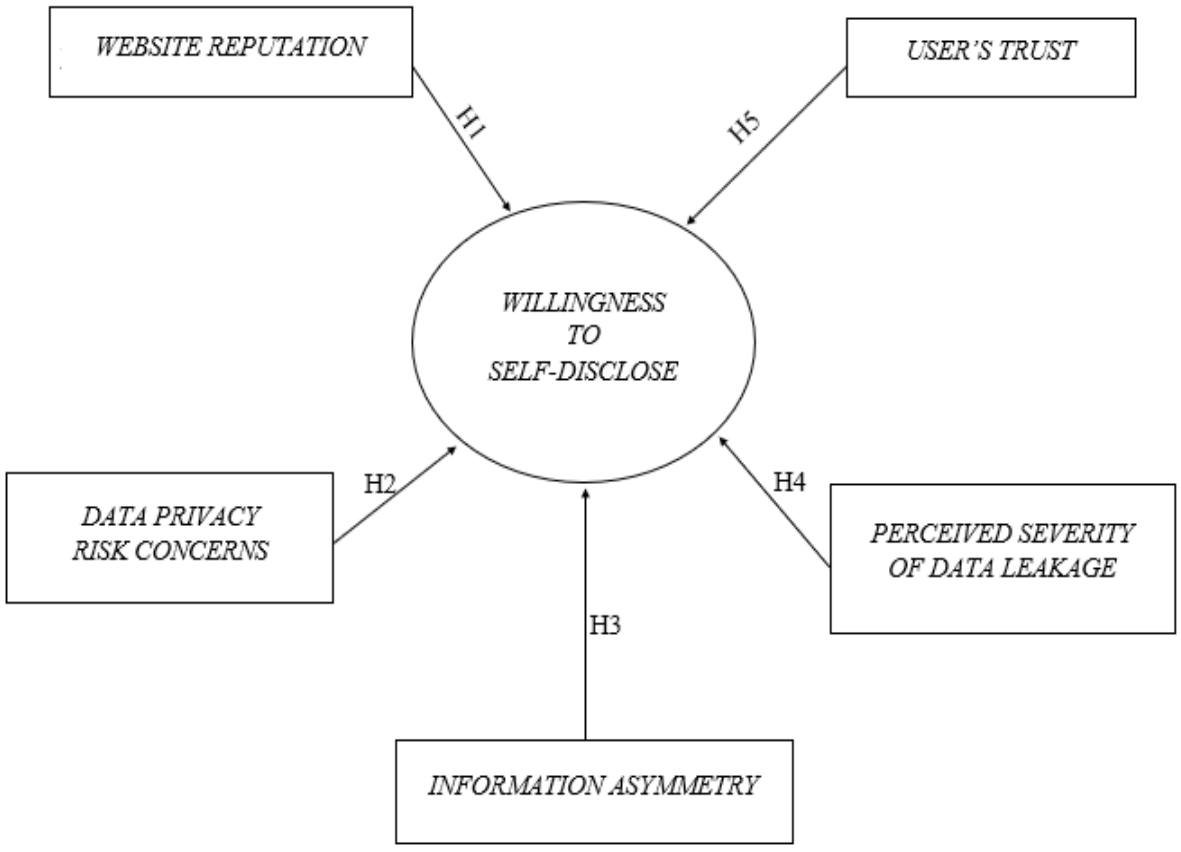


Figure 1: Research Model

2.2 Hypothesis development

After an attentive analysis of the existing literature on data privacy, the hypotheses that will be tested in the survey will allude to the variables previously mentioned. The hypotheses are the following:

Hypothesis 1 (H1): The more popular a website is considered by Italian users, the higher their inclination to share data.

Hypothesis 2 (H2): Data privacy is not a top-of-mind concept that concerns Italian internet users when navigating online.

Hypothesis 3 (H3): When terms of use (TOU) and/or the terms and conditions of a website are unmentioned or unclear, Italian users tend to restrain self-disclosure.

Hypothesis 4 (H4): Italian internet users are more propense to self-disclose when the website has no past history of data leakage.

Hypothesis 5 (H5): Italians are less attentive to self-disclose when navigating on a particular website when someone (a friend or an influencer) suggested it to them.

2.3 Variables description

INDEPENDENT VARIABLES

The independent variables of this model comprehend the analysis of the demographics of the respondents of the survey and users' attitudes. The independent variables that will be evaluated in this thesis are the following:

- Age – it will be measured by asking participants their age group
- Gender – it will be measured by asking participants their gender
- Education level – it will be measured by asking respondents their level of schooling
- Time spent on the internet – it will be measured by asking participants to choose amongst estimates of time frames

- Willingness to self-disclose – this variable will assess Italian internet user's *propensity to self-disclose*. From Li Z., Rau P.P., and Huang D. work[15], it results that worldwide internet users are more willing to disclose their personal preferences to Internet-of-things devices and conversational devices (e.g. Siri, Google home, Amazon Alexa, etc.) and least propense to disclose information regarding their financial situation.

DEPENDENT VARIABLES

The dependent variables are the most pertinent variables to assess, since the aftermath of the survey will be mainly determined by them, and they are:

1) Website reputation

Website reputation is a variable that must be assessed then analyzing users' behavior when it comes to information disclosure. As explained in Li Yuan study[16], the reputation perceived by users' regarding a website is a key element that affects individual's privacy concerns. Website reputation will be employed to analyze H1. This variable will be valued by scrutinizing three items: *online reviews*, *number of active users*, and company's website *social media following*. Online reviews (and ratings) shared by other users provide a feeling of reassurance to users who are interested in the website's reputation. The number of active users measures the number of users visiting a particular website over a certain period, and it can be thought of as a mean to verify the popularity of website, which consequently affect its reputation. Last but not least, website's social media following is also an item of keen relevance when evaluating the reputation of a website, and it measures the number of followers a company has on its social media platforms (e.g. Twitter, Facebook, Instagram, YouTube, etc.). In general, when a company has a big social media following, users tend to classify it as worthy of trust and safe.

[15] Li, Z., Rau, P.P. & Huang, D. (2019), "Self-Disclosure to an IoT Conversational Agent: Effects of Space and User Context on Users' Willingness to Self-Disclose Personal Information", Applied sciences, vol. 9, no. 9, pp. 1887.

[16] Li, Y. (2014), "The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns", Decision Support Systems, vol. 57, no. 1, pp. 343-354.

2) Data privacy risk concerns

This variable wants to measure the inclination of Italian users to disclose information. Privacy concerns can prevent internet users from using a particular service or product. As affirmed by Harborth D. and Pape S. in their paper[17], there is relationship among data privacy concerns, trust in online businesses, risk sentiment associated with online businesses' data treatment and users' behavior. Hence, I chose to consider two different items to assess this variable in the survey: *fear of fraudulent activities* and *sensitivity towards data collection*. Those items will be used to verify H2. Online threats such as malwares, scams, phishing, and identity theft are the most common categories of fraudulent activities and the extent of fear users' feel towards those threats help dictate the amount of information they will be willing to disclose when online. Sensitivity towards data collection also affects users' concerns regarding data privacy. In fact, it affects the amount of information users share online. It will be measured by asking participants how they feel towards data collection by companies (e.g. consulting firms, tech companies, etc.)

3) Information asymmetry

Asymmetric information occurs when one party (i.e. The website administrators) enjoys greater insights than the other party (i.e. Italian internet users)[18]. Research conducted by Jiang X, Hong J.I., and Landay J. A.[19] evidence how information asymmetry hinders data privacy; therefore, it is a variable that must be valued when discussing the topic.

This variable aims at measuring the predisposition of Italian internet users to disclose information when they are not fully aware of the terms of use (TOU) of a website (e.g. social networks), when there is a lack of understanding or of the terms and conditions of a website or when they are not mentioned (e.g. home banking websites, e-commerce, etc.). Asymmetric

[17] Harborth, D. & Pape, S. (2020) How Privacy Concerns, Trust and Risk Beliefs, and Privacy Literacy Influence Users' Intentions to Use Privacy-Enhancing Technologies. Data base. [Online] 51 (1), 51–69.

[18] Bloomenthal, A. (2020) Asymmetric information [Online]. Available at: https://www.investopedia.com/terms/a/asymmetricinformation.asp

[19] Jiang X., Hong J.I., Landay J.A. (2002) Approximate Information Flows: Socially Based Modeling of Privacy in Ubiquitous Computing. In: Borriello G., Holmquist L.E. (eds) UbiComp 2002: Ubiquitous Computing. UbiComp 2002. Lecture Notes in Computer Science, vol 2498. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45809-3_14

information is classified as a single-item variable in this thesis, and the item considered for its analysis is uninformed consent. *Uninformed consent* is a subject that has been discussed by several studies[20] and that I will fall back on to verify H3.

Specifically, the item will be measured by asking respondents if the (potential) experience of uninformed consent -through which websites have gathered their data- has ever discouraged them to share their personal data in their future online financial transactions or when surfing the web. In fact, it occurs that when privacy policies of a website are not well disclosed, users perceive a lack of transparency, which in turns leads them to refrain from sharing information.

4) Perceived severity of a data leakage

This variable wants to estimate how Italian internet users perceive data breaches. Perceived severity of data leakage will take into account two items: *internet literacy* and *awareness of previous data breach.* As Novak A.'s article[21] has evidenced, episodes of online data breach outraged the general public and had users question the legitimacy of data recruitment and raised fear of future breaches. Users feared future data leakage and the possibility of identity theft once they became aware that their sensitive data had been hacked. However, not all users discern the same degree of severity in regard to invasion of privacy; in fact, some individuals' lack of knowledge and awareness prevents their realization of the dangers of data breaches. Previous incidents involving leaked data contribute to increase user's concerns related to privacy and security. Internet literacy allows users to navigate the web's contents in a safe manner and ensures that they are conscious about their privacy protection. This item significantly affects user's perception of severity of a data leakage because it makes users' aware of all possible repercussions of such violation. Thus, I will use the two items above mentioned to demonstrate H4.

5) Users' trust

[20] Nunan, D. and Yenicioglu, B. (2013) 'Informed, Uninformed and Participative Consent in Social Media Research', International Journal of Market Research, 55(6), pp. 791–808. doi: 10.2501/IJMR-2013-067.

[21] Novak, A.N. & Vilceanu, M.O. (2019), ""The internet is not pleased": twitter and the 2017 Equifax data breach", The Communication Review, vol. 22, no. 3, pp. 196-221.

Users' trust in an essential variable to be studied when discussing online data privacy. Users' behavior when revealing information online depend on many factors, however, two key items can be distinguished*: word-of-mouth (WOM)* recommendation from people they know and *endorsement by influencers.* Multiple studies, including Jimenez-Castillo D. and Sanchez-Fernandez R.'s research[22], have proven that WOM and influencers' referral marketing campaigns can contribute to develop a feeling of trust and safety in consumers' mind. This variable will be used to assess H5.

DEPENDENT VARIABLE	**ITEM**

[22] Jiménez-Castillo, D. & Sánchez-Fernández, R. (2019), "The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention", International journal of information management, vol. 49, pp. 366-376.

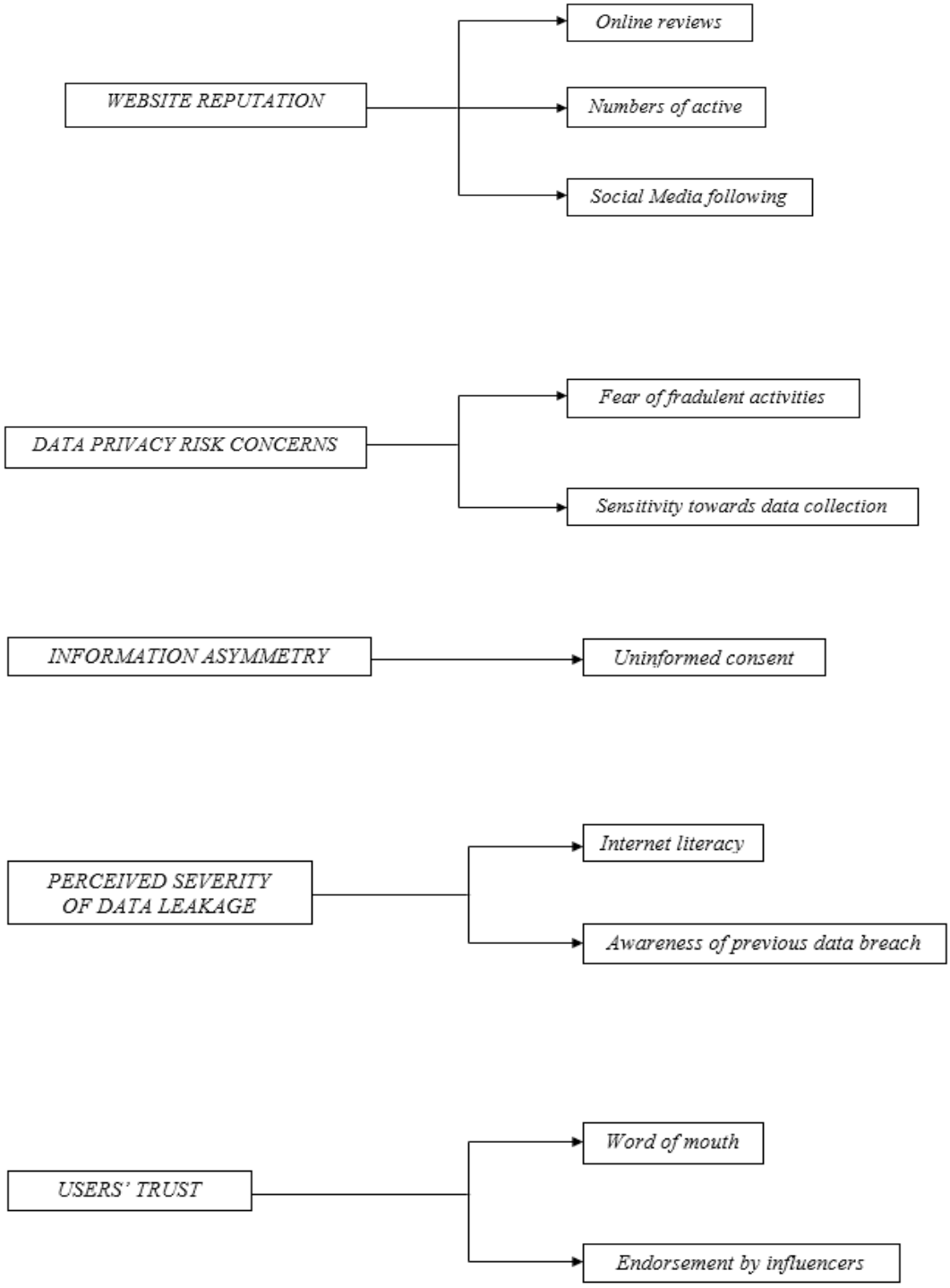
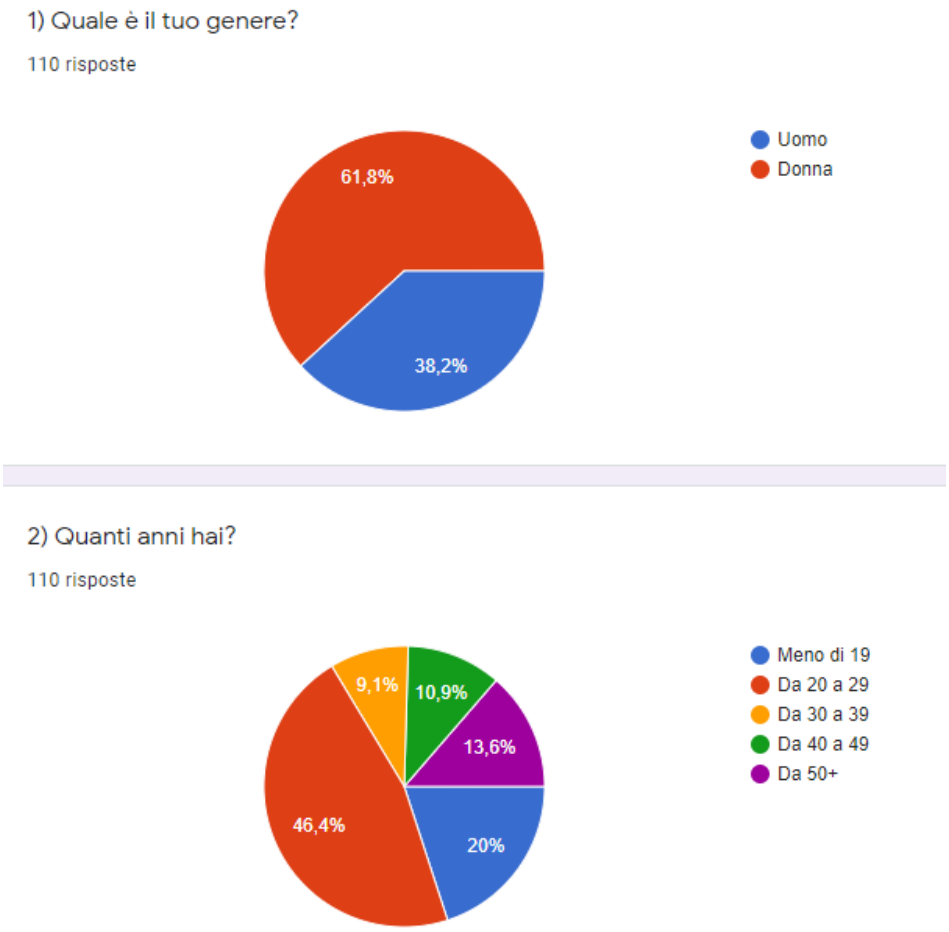


Figure 2: Dependent variables and their corresponding item/items

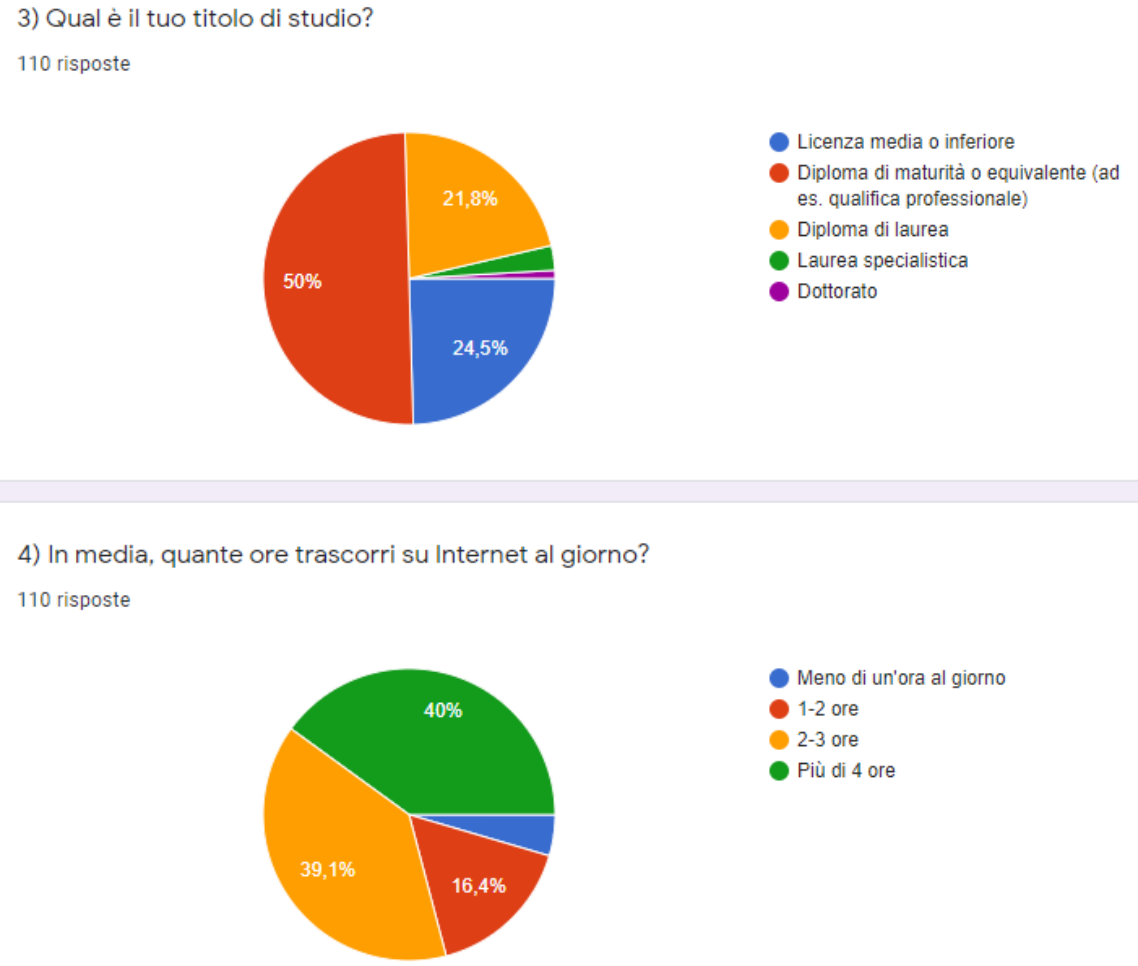
2.4 Sample demographics and data collection

The survey was distributed through social media platforms, specifically WhatsApp and Instagram direct messages. All the respondent's answers were anonymous and were collected via Google Forms. The sample size of the survey was 110 Italian internet users. Overall, the sample was mainly compiled by women. The survey was submitted by 68 females and 42 males. Moreover, the majority of participants age range fell under the group 20- 29 years old (51 respondents). Afterwards, the second largest age group that compiled the survey was made by participants under 19 years old (22 respondents). In addition, 15 respondents age range fell under the group of people over 50 years old. Finally, 22 respondents age range was from 30 to 49.



Figures 3 and 4: questions 1 and question 2 statistical graphs

Additionally, for what concerns education level, the majority of participants have obtained a high school diploma (55 participants). 27 participants have obtained a secondary school diploma or less. The remaining 28 participants have obtained at least a bachelor's degree.



Figures 5 and 6: question 3 and question 4 statistical graphs

From the data regarding time spent online, the majority of respondents declared spending more than 4 hours a day on the Internet (44 respondents). This is positive for the survey's outcome as the respondents would ideally have stronger opinions for what concerns information disclosure since they spend a lot of their time on the Internet. Only 5 participants declared spending less than one hour per day on the Internet.

Chapter 3: ***Data analysis***

In this chapter I will discuss the survey's structure, the statistics that were gathered from the survey and the key findings that emerged from the research.

3.1 The survey

The survey was completed by a sample of 110 Italian internet users and was carried out on the platform Google Forms. It was composed by seven parts. All questions, except questions regarding gender, age, education level, time spent on the internet, internet literacy and personal experience of privacy infringement, were assessed using a Likert-type scale ranging from strongly disagree to strongly agree. Each questions' objective was to assess the items presented in chapter 2 through interrelated questions in order to evaluate each variable.

The first four questions of the survey gathered the demographic characteristics of respondents (gender, age, and level of education) and time spent online through multiple choice questions. In the second part of the survey four questions regarding Italian users' general propensity and willingness to self-disclose were asked. For example, when asked to answer question six which reads, " I am comfortable with disclosing my personal information on the internet (e.g. sharing sensitive data in social media platforms, sharing credit/debit card information on a website, shopping on the internet, etc.)." respondents were required to express their propensity to disclose data by selecting only one among five choices which were presented on the form:

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

In the third part of the survey respondents were asked to evaluate website reputation. The fourth part of the survey determined the extent of respondent's data privacy risks concerns. Moreover, the fifth part of the survey evaluated information asymmetry. The sixth part of the survey concerned perceived severity of data leakage. The final part of the survey analyzed users' trust.

For confrontation purposes, I transformed the 5 choices Likert scale into a numeric scale for statistical analysis:

1= strongly disagree

2= somewhat disagree

3= neither agree nor disagree (neutral)

4= somewhat agree

5= strongly agree

3.2 Key findings

In this section, I will analyze the statistics regarding each variable and item to verify their relevance in the determination of Italian internet users' willingness and extent of self-disclosure. The following statistics were computed using STATA 14. The sample size is 110.

VARIABLE: ***WILLINGNESS TO SELF DISCLOSE*** (from Q5 to Q8)

```
summarize Q5 Q6 Q7 Q8
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q5	110	4.045455	.9897304	1	5
Q6	110	2.563636	1.344559	1	5
Q7	110	3.690909	1.114915	1	5
Q8	110	3.854545	1.02121	1	5

Table 1: individual propensity to self-disclose

Questions 5 through 8 aim was to identify Italian users' degree of individual propensity to self-disclose. Q5 which reads, "In my opinion, my online privacy is the most important thing to preserve" has a mean value above 3 (neither agree nor disagree) and this implies that the majority of respondents agreed or strongly agreed with such statement. Moreover, out of all the items assessed, Q5 displays the smallest standard deviation (inferior to 1) which demonstrates a low variation in the choices picked by respondents.

Q6 which reads, "I am comfortable with disclosing my personal information on the internet (e.g. sharing sensitive data in social media platforms, sharing credit/debit card information on a website, shopping on the internet, etc.)" has a mean value of 2.56 which indicates that the average Italian internet user is not propense to reveal her personal information on the Internet.

On the other hand, from the evaluation of the statistics of Q7 and Q8, it is possible to affirm that the participants of the survey were not extremely responsible or cautious when disclosing information online, but rather neutral. In conclusion, it is clear that the Italian users' propensity to self-disclose is not high, despite their lack of precautions when surfing on the Internet.

VARIABLE: ***WEBSITE REPUTATION*** (from Q9 to Q11)

```
summarize Q9 Q10 Q11
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q9	110	3.218182	1.191575	1	5
Q10	110	3.163636	1.184977	1	5
Q11	110	2.972727	1.207462	1	5

Table 2: website reputation

Q9, Q10 and Q11 assessed three different items to measure website reputation, respectively: online reviews, number of active users, and company's website social media following. It is evident that Italian internet users are neutral for what concerns disclosing information on a website that has a favorable reputation. In fact, even in the circumstances in which those websites have positive reviews, many active users and a considerable amount of social media followers, Italian internet users displayed an indifferent attitude towards data sharing. To sum up, it can be affirmed that website reputation is not variable that has a powerful effect on consumers' willingness to self-disclosure.

VARIABLE: ***DATA PRIVACY RISKS CONCERN*** (Q12 and Q13)

```
summarize Q12 Q13
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q12	110	4.009091	1.087843	1	5
Q13	110	3.4	1.242962	1	5

Table 3: data privacy risks concerns

The items evaluated in Q12 and Q13 were fear of fraudulent activities and sensitivity towards data collection. Due to the fact that question 12 mean equals to 4, it is clear that the great majority of the respondents are concerned about the possibility of fraudulent activities on the Internet, and therefore this leads to a lower predisposition to share sensitive data.

Q13 reveals that the average of users is not particularly conditioned by the existence of personal data collection by big companies when they browse online. Therefore, data privacy risks concern mainly arise due to the fear of fraudulent conduct.

VARIABLE***: INFORMATION ASYMMETRY*** (Q14, Q15 and Q16)

```
summarize Q14 Q15 Q16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q14	110	2.954545	1.266262	1	5
Q15	110	3.436364	1.215551	1	5
Q16	110	3.963636	1.180322	1	5

Table 4: information asymmetry

The item considered for information asymmetry was uninformed consent, which was assessed through three questions. By analyzing the statistics of Q14, it is possible to assert that the average Italian user disagrees that they would continue to navigate on a website's whose terms and conditions are not clear to them. Anyhow, Q15 statistics indicate that the average Italian users' attitude towards sharing personal information on a website that does not express their privacy policies is neither extremely positive nor extremely negative. Given that Q16's mean

(3.96) is closer to 4 and 5, there is a stronger tendency towards to not revealing information on a website that has gathered data through uninformed consent. For those reasons, one can assess that the existence of information asymmetry does have a great impact on Italian users' tendency to self-disclose when online.

VARIABLE: ***PERCEIVED SEVERITY OF DATA LEAKAGE*** (from Q17 to Q20)

```
summarize Q18
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q18	**110**	**3.909091**	**1.177351**	**1**	**5**

Table 5: perceived severity of data leakage

The items analyzed to evaluate perceived severity of data leakage were internet literacy and awareness of previous data breach. Q18 statistics show that there is a slight inclination by the average Italian users to avoid disclosing personal information on a website that has suffered hacking attacks in the past, since the mean (3.9) leans towards 4 ("somewhat agree"). Furthermore, from Q17 it is evident that the majority of the respondents (58.2%) had not received classes regarding how to safely navigate the internet. In addition, Q19 results show that out of 110 respondents, 34 respondents (30.9%) have suffered some form of online data violation, thus one can deduce that internet literacy and personal experience of data violation have a negative impact on users' willingness to self-disclose.

Figure 7: question 17 statistical graphics

Figure 8: question 18 statistical graphics

```
summarize Q20
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q20	110	3.836364	1.177209	1	5

Table 6: perceived severity of data leakage

From Q20 one can suggest that the average Italian internet user is more likely than not to reconsider sharing their information on a website that is vulnerable to hacking attacks – in fact, 35 users chose "somewhat agree" and 45 users chose "strongly agree" with Q20's statement. In conclusion, it is not possible to affirm that there is a strong correlation between the Italian users' perceived severity of data leakage and their willingness to self-disclose. However, it is more likely that Italian internet users will prefer to protect their private information, because although the mean of Q18 and Q20 is in the "neutral" range (value 3), 3.9 and 3.8 are both marginally on the side of "somewhat agree" (value 4) and "strongly agree" (value 5).

VARIABLE: ***USERS' TRUST*** (Q21 and Q22)

```
summarize Q21 Q22
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Q21	110	2.590909	1.287166	1	5
Q22	110	3.236364	1.270175	1	5

Table 7: users' trust

The items chosen to assess this variable were word-of-mouth (WOM) recommendations and endorsement by influencers.

From the statistics of Q21 and Q22, one can affirm that neither word-of-mouth recommendations nor endorsement by influencers are the main elements that affected the average Italian internet users' attitude to disclose personal information online.

Furthermore, the only dependent variable whose mean exceeds 4 ("somewhat agree") is the variable of question 12, which is data privacy risk concerns. This means that data privacy risk concerns have a significant impact on willingness to self-disclose. We can further proceed to confirm this theory by running a linear regression in which the independent variable will be the item presented on question 5, that is, propensity to self-disclose.

```
regress Q12 Q5
```

Source	SS	df	MS	Number of obs	=	110
				F(1, 108)	=	30.33
Model	28.2843957	1	28.2843957	Prob > F	=	0.0000
Residual	100.706513	108	.932467717	R-squared	=	0.2193
				Adj R-squared	=	0.2120
Total	128.990909	109	1.18340284	Root MSE	=	.96564

Q12	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Q5	.5146871	.0934516	5.51	0.000	.3294498	.6999244
_cons	1.926948	.389104	4.95	0.000	1.155676	2.698219

Table 8: linear regression

Assuming significance level α=5%, one can see that the p-value (0.000) is lower than α and thus statistically significant, and therefore one can reject the null hypothesis. That is, propensity to self-disclose is affected by different variables, and data privacy risk concerns is one of them.

R-squared measures how fit the linear regression can be deemed. The low value obtained from the regression for R-squared (0.2193) was to be expected since the sample size was small, and the thesis evaluates human behavior, which makes difficult to distinguish a common pattern.

Conclusion

The study conducted on this thesis has some limitations, mainly because the sample size considered was small (only 110 respondents) and because the number of variables that can affect Italian internet users' behavior regarding self-disclosure are wide-ranging.

However, from the items analyzed in each question, it appears that Italian internet users are willing to self-disclose, but only to a low the extent.

Additionally, website reputation does not have a meaningful impact in internet users' behavior when revealing information, regardless of the website's credibility and integrity. For this reason, H1 is not verified.

Data privacy risk concerns mainly arise due to fear of fraudulent activities and it is a variable that appears to have a considerable repercussion on users' willingness to self-disclose. Therefore, data privacy is a top-of-mind concept that appertain to inclination of self-disclosure. As noticeable by the linear regression computed in *Table 8*, we can see that privacy risks affect users' propensity to reveal their sensitive information and thus, H2 is verified. It can also be claimed that the main variable in this research affecting Italian users' actions when surfing on the web is concern of potentially fraudulent activities (i.e. scams, phishing, and identity theft).

It also appears that information asymmetry affects users' willingness to self-disclose. This means that Italian internet users are worried about understanding the terms and conditions of a website, and this confirms H3.

Moreover, perceived severity of data leakage is relevant factor to examine when estimating Italian users' disposition in regard to sensitive data exposure, since history of previous data breach negatively affects users' propensity to share information on a website, and thus H4 is truthful.

Lastly, from the statistical evaluation conducted to assess users' trust, it becomes evident that users' trust on a website does not affect users' willingness to self-disclose, and therefore H5 is not demonstrated.

To summarize, this thesis was built by taking into consideration the pre-existing literature regarding online data privacy and the research conducted in Italy revealed that Italian users' propensity to self-disclose when online is mainly determined by their fear of unlawful online conduct, their aversion towards websites that collect unauthorized data, and their lack of trust in websites that suffered data violations in the past.

APPENDIX

Survey's link:
https://docs.google.com/forms/d/e/1FAIpQLSd04yR4rllK1iC1ifz2MA3IHAbWKPN5AKBiRz_YpUIXA0BesA/viewform?usp=sf_link

Survey questions:

Questions on ***demographics***

Q1. What is your gender?

-Male

-Female

Q2. How old are you?

- Less than 19

- 20-29

- 30-39

- 40-49

- 50+

Q3. What is the highest level of education you have completed?

- secondary school diploma or less

- high school diploma or professional degree

- Bachelor's degree

- Master's degree

- PhD

Question on ***time spent on the Internet***

Q4. On average, how many hours per day do you spend on the internet?

- Less than 1 hour per day

- 1-2 hours

- 2-3 hours

- more than 4 hours a day

Questions 5 to 16, question 18, and questions 20 to 22 answer choices were the following:

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Questions regarding ***"Willingness to self-disclose"***

Q5. In my opinion, my online privacy is the most important thing to preserve.

Q6. I am comfortable with disclosing my personal information on the internet (e.g. sharing sensitive data in social media platforms, sharing credit/debit card information on a website, shopping on the internet, etc.).

Q7. Compared to others, I tend to be more responsible when sharing sensitive data online.

Q8. Compared to others, I am more cautious about potential violations of my personal privacy.

tabulation of Q5

Q5	Freq.	Percent	Cum.
1	2	1.82	1.82
2	6	5.45	7.27
3	21	19.09	26.36
4	37	33.64	60.00
5	44	40.00	100.00
Total	110	100.00	

tabulation of Q7

Q7	Freq.	Percent	Cum.
1	7	6.36	6.36
2	8	7.27	13.64
3	24	21.82	35.45
4	44	40.00	75.45
5	27	24.55	100.00
Total	110	100.00	

tabulation of Q6

Q6	Freq.	Percent	Cum.
1	34	30.91	30.91
2	21	19.09	50.00
3	24	21.82	71.82
4	21	19.09	90.91
5	10	9.09	100.00
Total	110	100.00	

tabulation of Q8

Q8	Freq.	Percent	Cum.
1	4	3.64	3.64
2	7	6.36	10.00
3	21	19.09	29.09
4	47	42.73	71.82
5	31	28.18	100.00
Total	110	100.00	

Questions regarding ***"Website reputation"***

Q9. I am more likely to disclose personal information on a website if it has many positive online reviews.

Q10. If a website has many active users, I find it reliable and therefore I am more incline to share information on it.

Q11. If a company has a substantial following on its social media platforms, I am more likely to disclose information in their website.

tabulation of Q9

Q9	Freq.	Percent	Cum.
1	15	13.64	13.64
2	12	10.91	24.55
3	28	25.45	50.00
4	44	40.00	90.00
5	11	10.00	100.00
Total	110	100.00	

tabulation of Q10

Q10	Freq.	Percent	Cum.
1	13	11.82	11.82
2	19	17.27	29.09
3	26	23.64	52.73
4	41	37.27	90.00
5	11	10.00	100.00
Total	110	100.00	

tabulation of Q11

Q11	Freq.	Percent	Cum.
1	18	16.36	16.36
2	20	18.18	34.55
3	26	23.64	58.18
4	39	35.45	93.64

ore—

Questions regarding ***"Data privacy risks concerns"***

Q12. I worry about fraudulent activities (e.g. identity theft, unauthorized access to my emails, credit card scams, etc.) when I am browsing the web.

Q13. I am conditioned by the existence of personal data collection by big companies when I surf on the Internet.

tabulation of Q12

Q12	Freq.	Percent	Cum.
1	3	2.73	2.73
2	9	8.18	10.91
3	19	17.27	28.18
4	32	29.09	57.27
5	47	42.73	100.00
Total	110	100.00	

tabulation of Q13

Q13	Freq.	Percent	Cum.
1	14	12.73	12.73
2	9	8.18	20.91
3	26	23.64	44.55
4	41	37.27	81.82
5	20	18.18	100.00
Total	110	100.00	

Questions regarding ***"Information Asymmetry"***

Q14. I would continue to navigate on website even if its terms and conditions are not clear to me.

Q15. I am less likely to share personal information on a website if the privacy policies of such website are not expressed.

Q16. I am less inclined to share personal information on a website if in the past the website has gathered my personal data without my explicit consent.

tabulation of Q14

Q14	Freq.	Percent	Cum.
1	19	17.27	17.27
2	20	18.18	35.45
3	31	28.18	63.64
4	27	24.55	88.18
5	13	11.82	100.00
Total	110	100.00	

tabulation of Q15

Q15	Freq.	Percent	Cum.
1	11	10.00	10.00
2	11	10.00	20.00
3	30	27.27	47.27
4	35	31.82	79.09
5	23	20.91	100.00
Total	110	100.00	

tabulation of Q16

Q16	Freq.	Percent	Cum.
1	7	6.36	6.36
2	4	3.64	10.00
3	24	21.82	31.82
4	26	23.64	55.45

Questions regarding ***"Perceived severity of data leakage"***

Q17. I have received lessons on how to safely navigate the Internet.

- Yes

- No

Q18. I would not disclose personal information on a website that has suffered hacking attacks in the past.

Q19. I have experienced some form of online data violation in the past (e.g. hacking attacks).

- Yes

- No

Q20. I will rethink sharing my information on a website if I sense that the website is vulnerable to hacking attacks or that it is not sufficiently safe.

tabulation of Q18

Q18	Freq.	Percent	Cum.
1	5	4.55	4.55
2	8	7.27	11.82
3	27	24.55	36.36
4	22	20.00	56.36
5	48	43.64	100.00
Total	110	100.00	

tabulation of Q20

Q20	Freq.	Percent	Cum.
1	10	9.09	9.09
2	3	2.73	11.82
3	17	15.45	27.27
4	45	40.91	68.18
5	35	31.82	100.00
Total	110	100.00	

Questions regarding ***"Users' trust"***

Q21. I believe that the websites endorsed by influencers can be deemed as trustworthy.

Q22. I believe that the websites suggested by the people I know (e.g. family, friends, acquaintances) are reliable.

tabulation of Q21

Q21	Freq.	Percent	Cum.
1	30	27.27	27.27
2	20	18.18	45.45
3	37	33.64	79.09
4	11	10.00	89.09
5	12	10.91	100.00
Total	110	100.00	

tabulation of Q22

Q22	Freq.	Percent	Cum.
1	17	15.45	15.45
2	11	10.00	25.45
3	27	24.55	50.00
4	39	35.45	85.45
5	16	14.55	100.00
Total	110	100.00	

References

"Big data analytics" [Online]. IBM. Available at: https://www.ibm.com/analytics/hadoop/big-data-analytics

"IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth Through 2022" (2019). Available at: https://www.idc.com/getdoc.jsp?containerId=prUS44998419

"Safe surfing: A brief look on internet security" [Online]. Eurostat: Digital economy and society in the EU. Available at: https://ec.europa.eu/eurostat/cache/infographs/ict/bloc-3a.html

"What is Big Data?" [Online]. Oracle. Available at: https://www.oracle.com/it/big-data/what-is-big-data.html

Abadeh, Maryam & Mirzaie, Mansooreh. (2020). DiffPageRank: an efficient differential PageRank approach in MapReduce. The Journal of Supercomputing. Available at: https://www.researchgate.net/publication/340281398_DiffPageRank_an_efficient_differential_PageRank_approach_in_MapReduce

Abhinav, R. (2020). What is Big Data – Characteristics, Types, Benefits & Examples [Blog]. Upgrad blog. Available at: https://www.upgrad.com/blog/what-is-big-data-types-characteristics-benefits-and-examples/

Ablon, L., Heaton, P., Lavery, D.C., and Romanosky, S. (2016) Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information. Santa Monica, CA: RAND Corporation. Available at: https://www.rand.org/pubs/research_reports/RR1187.html.

Anant, V., Donchak, L., Kaplan, J. and Soller, H. (2020) The consumer-data opportunity and the privacy imperative. McKinsey & Company. Available at: https://www.mckinsey.com/business-functions/risk/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative#

Big Data. (n.d.) In Cambridge Advanced Learner's Dictionary and Thesaurus. Available at: https://dictionary.cambridge.org/dictionary/english/big-data

Blazquez, D. & Domenech, J. 2018, "Big Data sources and methods for social and economic analyses", Technological forecasting & social change, vol. 130, pp. 99-113.

Bloomenthal, A. (2020) Asymmetric information [Online]. Available at: https://www.investopedia.com/terms/a/asymmetricinformation.asp

Dash, S., Shakyawar, S.K., Sharma, M. et al. (2019). Big data in healthcare: management, analysis and future prospects. J Big Data 6, 54. Available at: https://doi.org/10.1186/s40537-019-0217-0

Dinev, T. and Hart, P. (2003) An Extended Privacy Calculus Model for E-Commerce Transactions. Volume 17, Issue 1. Available at: https://doi.org/10.1287/isre.1060.0080

Grabner-Kraeuter, S. The Role of Consumers' Trust in Online-Shopping. Journal of Business Ethics 39, 43–50 (2002). https://doi.org/10.1023/A:1016323815802

Harborth, D. & Pape, S. (2020) How Privacy Concerns, Trust and Risk Beliefs, and Privacy Literacy Influence Users' Intentions to Use Privacy-Enhancing Technologies. Data base. [Online] 51 (1), 51–69.

Isaac, M. and Frenkel, S. (2018) Facebook Security Breach Exposes Accounts of 50 Million Users [Online]. The New York Times. Available at: https://www.nytimes.com/2018/09/28/technology/facebook-hack-data-breach.html

Jiang X., Hong J.I., Landay J.A. (2002) Approximate Information Flows: Socially Based Modeling of Privacy in Ubiquitous Computing. In: Borriello G., Holmquist L.E. (eds) UbiComp 2002: Ubiquitous Computing. UbiComp 2002. Lecture Notes in Computer Science, vol 2498. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45809-3_14

Jiménez-Castillo, D. & Sánchez-Fernández, R. (2019), "The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention", International journal of information management, vol. 49, pp. 366-376.

Khalifa, M. and Limayem, M. (2003) Drivers of Internet shopping. Commun. ACM 46, 12, pp. 233–239. DOI: https://doi.org/10.1145/953460.953505

Knowledge@Wharton and Barratt, J. (2019) "Data as Currency: What Value Are You Getting?" [Online]. Available at: https://knowledge.wharton.upenn.edu/article/barrett-data-as-currency/

Li, Y. (2014), "The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns", Decision Support Systems, vol. 57, no. 1, pp. 343-354.

Li, Z., Rau, P.P. & Huang, D. (2019), "Self-Disclosure to an IoT Conversational Agent: Effects of Space and User Context on Users' Willingness to Self-Disclose Personal Information", Applied sciences, vol. 9, no. 9, pp. 1887.

Lupyan, G. (2017), "Online data collection: The good, the better, and the advantages" [Online]. Available at: https://featuredcontent.psychonomic.org/online-data-collection-the-good-the-better-and-the-advantages/

Novak, A.N. & Vilceanu, M.O. (2019), ""The internet is not pleased": twitter and the 2017 Equifax data breach", The Communication Review, vol. 22, no. 3, pp. 196-221.

Nunan, D. and Yenicioglu, B. (2013) 'Informed, Uninformed and Participative Consent in Social Media Research', International Journal of Market Research, 55(6), pp. 791–808. doi: 10.2501/IJMR-2013-067.

Office of the Australian Information Commissioner. 2020. What is privacy? [Online] Available at: https://www.oaic.gov.au/privacy/your-privacy-rights/what-is-privacy/

Patgiri, R. and Ahmed, A. (2016). Big Data: The V's of the Game Changer Paradigm. IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NSW, 2016, pp. 17-24. Available at: https://www.researchgate.net/publication/311642627_Big_Data_The_V's_of_the_Game_Changer_ParaPara

Press, G. (2016). IoT Mid-Year Update From IDC And Other Research Firms [Online]. Forbes. Available at: https://www.forbes.com/sites/gilpress/2016/08/05/iot-mid-year-update-from-idc-and-other-research-firms/#1787c53b55c5

Privacy. (n.d) In Cambridge Academic Content Dictionary. Available at: https://dictionary.cambridge.org/dictionary/english/privacy

Rădulescu, Adina (2018). "Users' Social Trust of Sharing Data with Companies: Online Privacy Protection Behavior, Customer Perceived Value, and Continuous Usage Intention," Contemporary Readings in Law and Social Justice 10(1): 137–143.

Reiff, N. (2020). How Twitter Makes Money [Online]. Investopedia. Available at: https://www.investopedia.com/ask/answers/120114/how-does-twitter-twtr-make-money.asp

Rezzani, A. (2018). Le tre V dei Big Data [Online]. Data skills understanding the world. Available at: https://www.dataskills.it/le-tre-v-dei-big-data/#gref

Wong, J.C. (2019) The Cambridge Analytica scandal changed the world – but it didn't change Facebook [Online] The Guardian. Available at: https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook