The database behind Facebook

Djurić, Sara

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Economics and Business / Sveučilište u Zagrebu, Ekonomski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:148:195735

University of Zagreb
Faculty of Economics and Business
Master in Managerial Informatics

The Database behind Facebook

Master Thesis

Sara Djurić
Mentor: Associate Professor Jovana Zoroja, Ph.D

Zagreb, September 2021

Sara Djurić

Name and family name of student

STATEMENT ON ACADEMIC INTEGRITY

I hereby declare and confirm with my signature that the Master thesis
(type of the paper)
is exclusively the result of my own autonomous work based on my research and literature published, which is seen in the notes and bibliography used.

I also declare that no part of the paper submitted has been made in an inappropriate way, whether by plagiarizing or infringing on any third person's copyright.

Finally, I declare that no part of the paper submitted has been used for any other paper in another higher education institution, research institution or educational institution.

Student:

In Zagreb,_____ _____
(date) (signature)

Abstract

We live in a world where a lot of private information through social media is provided to friends/followers. Databases are more and more powerful as technology advances. This paper will provide structured explanation of the databases and their types. As the Internet and the number of Internet users grows, the amounts of data grow exponentially. Consequently, the rise of Big Data happened. This paper introduces potential risks people are facing while sharing personal information, locations, and other content online, precisely on Facebook. However, they are not aware of possible risks. In this paper, intention is to find out are people aware of how their data is being used, where is it stored and who has access to it. The main goals of the work are:

- Collect and analyse available secondary data on databases in general, databases behind Facebook, big data, challenges and opportunities of big data, etc.
- Process in detail and present the database behind Facebook
- Conduct a primary research to analyse the attitudes of respondents, their confidence in the Facebook when it comes to sharing personal information. In the preparation of the thesis will be used secondary data collected from scientific and professional literature as a basis for conducting primary research. In addition, a primary research will be conducted using an online questionnaire via a Google form on a sample of approximately 100 randomly selected respondents accompanied by graphical representations.

In order to achieve research goal, literature review regarding databases and impact of Facebook database has on people's privacy will be presented. In addition, descriptive analysis will be applied to analyse collected data. Data will be collected via online questionnaire in order to investigate respondents' awareness when it comes to sharing information and any other type of content online. Questionnaire consists of following parts: (i) respondents' gender, age and demographic characteristics; (ii) respondents' awareness of where their data is being used (iii) respondents' opinion of the fact that there is a source which has so much information about them. Survey will be distributed via Internet, social media and online forums. The survey was carried out on a sample of mainly Croatian citizens who are Internet users.

The first chapter is Introduction which will include description of Topic and Goal of the Thesis. Furthermore, Data and Methodology will be presented. At the end of the Introduction, Structure of Thesis will be described.

The second chapter is introduction to Databases. This chapter will include description of Database models and each of them will be described in its subchapter. Subchapters will include: Flat model, Network model, Relative model and Dimensional model. Also, in the second chapter term Big Data will be explained along with the database behind Facebook.

The third chapter is about the protection of data on Facebook. This chapter will start off by explaining the protection of data in general, followed by terms of service designed by Facebook. Next subchapter will be data usage and privacy protection followed by the Facebook case and data protection in Croatia. The last subsection of the third chapter will present EU – USA privacy shied.

The fourth chapter is Analysis of Research Results regarding awareness of data protection, where the sub-chapters will explain and present the Research Methodology, Research Results followed by a Discussion.

After this, a Conclusion will be brought forth. Followed by the final three chapters, which are List of References, List of Graphs and Tables and the Appendix.

Key words: database, security, Facebook

Table of Contents

1. Introduction

1.1. Defining a problem

With over 8 million users across 2,000 college campuses, Facebook (www.facebook.com) is one of the most popular social networking services.[1] There are sure to be privacy risks with this much sensitive information arranged uniformly and aggregated in one place. Users may submit information without realizing it will be shared with advertising. Third parties may create and sell a database of Facebook data. Passwords or entire databases may be stolen from Facebook by intruders. Not to mention that user's data is being collected and used for all kinds of purposes.

1.2. Goal of the thesis

As the Internet grows so are the databases improving. Nowadays, one can store a huge amount of data in a small chip. In order to understand how social media collects and use our data it is important to introduce a concept of database and big data. Their importance is crucial for the entire IT sector, not only for collection of data on social media. Also, this paper will cover the protection of data and give some examples. The end goal is to properly present the concept of a database and to show the importance of being aware how data is collected and used.

1.3. Data and Methodology

The methods that will be used in the preparation of this paper are: method of analysis (which analyses individual complex concepts and/or conclusions to simpler components), method of description (to describe different measures, methods, procedures), method of proof (the purpose of which is to determine the accuracy of a cognition), method of graphical and tabular presentation of data, the method of generalization (as a method of generalization by which one particular the concept comes to a more general one that is higher in degree than other individual ones), method of classification (for the division of more complex thought formations into simpler ones comparison method, for comparison of certain data), method of synthesis (as a process of scientific research which explains reality through the synthesis of simple judgments into more complex ones) and method of comparison.

1.4. Structure of the thesis

This paper is composed of five parts. In the first part, the introduction, the problem of the work, the goals of the work, the methods used during the preparation of the work, the structure of the work and the explanation of the terms are defined. In the second part, an insight into the databases and big data is given. The third part deals with protection of the data. The empirical, fourth part of the paper introduces an online survey and discusses the questions answers. Lastly, in the 5th part, concluding remarks and reviews of the paper are presented.

2. Database

Database is a collection of records which are suited in a way that that a computer software can contact it when responding to a problem.[1] In other words, data is stored systematically and computer program can easily get to those data when needed. Every record is commonly recognized as a set of data elements which are sometimes called facts. Records are organized as a set of information elements for better return and sorting. Database management system is a software which is used to handle and test the database. The properties and design of a database system are included in the study of information science.[2]

For a given database, usually a structural description of the type of facts contained in that database can be found: that description is called a schema.[3] The diagram describes the objects and all the relationships between them that are being displayed in the database. There are a numerous of specified ways to organize a schema. For example, modelling a database structure which is called database models. The most popular model nowadays is the relative model, which is structured so that it presents all information in the form of multiple relative tables out of which, each consists of rows and columns. This model also shows all kinds of relationships by using values that are common to more tables. Other models suchlike as the hierarchical model and the network model use representations and relationships that are much more explicit.[4]

2.1. Database models

Different techniques are used to structure the data model. Most database structures are constructed toward only one specific data model.[5] Different physical implementations can be viable for any logical model, and maximum products will provide the person a few stages of control in tweaking physical implementation, because the selections made have an extensive impact on out - turn. This can be found in the relative model: all important performance of the relative model allow the creation of indexes, which then, allow quick access to the rows in the table. It is possible only if the values of certain columns are known.[6] A data model also defines a set of operations that may be carried out on information. The relative model, in this sense, determines operations including selection or choice, projection, and merging. However, those operations are usually not expressed in a specific question language, but they do they

offer the muse on which the question language is built.

2.1.1 Flat model

Flat Model (or tabular) includes a single, two-dimensional row of data elements, wherein all participants of a given column are assumed to have comparable values, and that each one participants of the row are associated with every other.[7] For example, name and password columns can be used as part of a database security system. Every line needs to contain a particular password which was previously associated with the each user. Next, table columns usually have a type joint with them which defines them as data labels, date or time information, a set, or floating-point numbers. Coincidentally, this model is also a spreadsheet database.

2.1.2. Network model

The network model is specific in a way that it organizes data using records and sets which are two fundamental builders or constructs. Records contain fields, sets define "one-to-all" relations among records which is one owner and many members. The record can be both member and owner in different number of gatherings. The network model navigation operations so that the program maintains the present position and manages from one record to another while following the relations in which the record operates. Nontheless, records can also be placed by supplying key values.[8]

2.1.3. Relational model

The relational model was introduced in 1970 by E. F. Codd in his academic paper as a method to create a database management system more autonomous of any other. It was introduced as a mathematical model which was defined in terms of predication logic and set theory. Even though the main idea of the database was very popular among mases, only a few people truly understood the mathematical definition, Moreover, even less of them knew how to apply vague SUBPs completely and without any additions.[9] For example, an Oracle database can be used in a purely relative way, but it also works with tables that allow double rows to be defined as — an appendix (or transgression) of the model in question.

In the common use of the Croatian language, a SUBP is called relational only if it supports relational operations. It is so regardless of notwithstanding it implements rigid devotion to a

formal relational model. The relational database is made up of multiple tables, which are all similar to the previously discussed "flat" database model. However, the big difference is that, tables are not linked by pointers which is not true in network databases. Instead of pointers keys are being used to crowd rows of data in different tables.[10]

Each column may be a key or more columns can be gathered into what is called a single key. It is not required to appoint every key in advance because the column can be used as a key even if it was not originally deliberated to be one. Unique key in one that is usually utilized to uniquely identify a row in a table. Usually, one of the unique keys is also a primary key.[11] When a key consists of data that has an external, real meaning, it is called a “natural” key. Consequently, if there is no natural key which is suitable at the moment, an arbitrary key may be assigned. However, almost all databases have both keys: generated and natural keys.

2.1.4.Dimensional model

The dimensional model is a distinctive processing of the relative model which is being used to display data. It is usually used in data warehouses. Data is easily summarized using an OLAP questionnaire. When it comes to a dimensional model, a database is consisted of one big table of facts which are described using dimensions and sizes.[12] Dimensions are often separate and hierarchical. For example, a place may comprise of a building, a region, and a state. Size is a quantity that describes income which will be a fact. Sizes can be significantly accumulated - for example, income from diverse places can be summed together. However, when it comes to the OLAP questionnaire, dimensions are selected and facts are gathered and combined together so they could all create a outline. It is non rare that dimensional model is implemented at the peak of the respective model by using a star scheme that consists of a single table containing facts and other tables containing dimensions. Furthermore, complex dimensions can in particular be shown using multiple tables, which will result in a flake scheme. A data store can be made with multiple star charts that share dimensional tables with each other, allowing them to be used together. Appearance with a normal set of dimensions is an crucial part of dimensional modeling.[13]

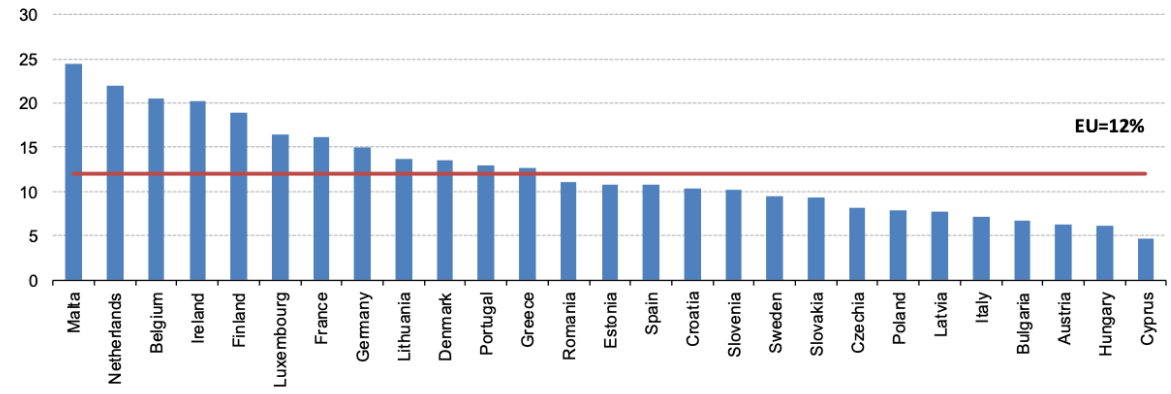
2.2. Big Data

Big data is a mixture of established, semi structured and even unstructured records amassed through businesses that may be mined for records and utilized in device gaining knowledge of projects, predictive modelling and different superior analytics applications.[14] It is a technology that is used to collect, process and analyse a big amount of data. Moreover, data is different - structured and unstructured, it is generated and comes at high speed at diverse times, which makes it very complicated to analyse. However, collecting and storing big amount of data isn't what makes big data technology. Rather, it is the ability to process and analyse this collected data for further use that makes this technology very valuable. Without the ability to analyse and the necessary tools (complex computer programs), it would be just a pile of collected data.[15]

The data that is being collected comes from the various devices or systems. Those are our smartphones, smart bracelets or watches, smart refrigerators, ATMs and so on. Simply put: smart devices enable data collection, data is sent for processing and analysis, then returned back in a form we understand and can use. For example, the motion sensor on a smart bracelet collects information about the movements a person is doing (running, jumping, walking, climbing, or some other function over a period of time).[16] It sends the same information via the Internet to the "database". The database is just a place, a repository where all this data is located. Next, Big Data technology takes over the job of processing and analysis. To simplify, in order to know how many steps, we have made and how many calories we have lost, Big Data technology must very quickly process all incoming data, analyse it and transform it into understandable and readable information, then send it back to the device via the Internet, in our example a smart bracelet. Thus, a person can immediately find out the result of his physical activity.[17]

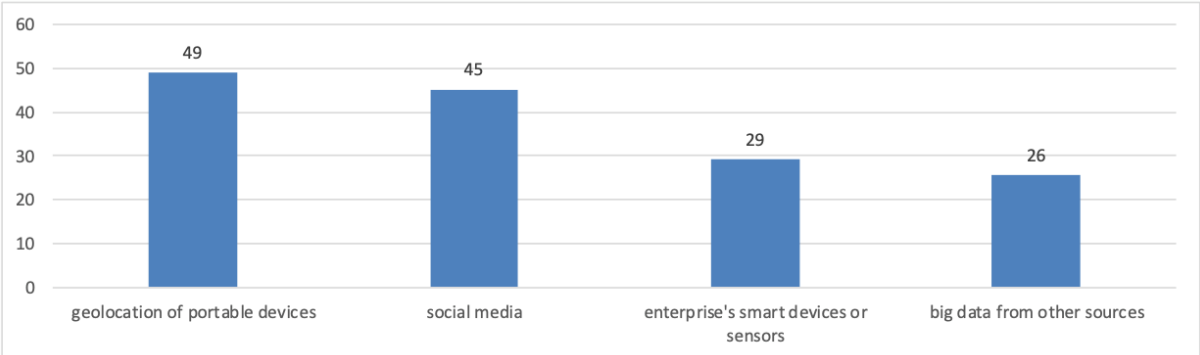
2.2.1. Use of Big Data technology in the EU

Lately, the amount of data created, stored and processed has grown largely. There was an analysis of Big Data conducted in 2018 by twelve percent of companies in EU which had at least ten employees. Results have shown that big data analyses were mainly done by big and medium-sized enterprises, with eight percent of the analyses being performed by the company's employees, while five percent of enterprises had an external company as an associate. In Croatia, only about ten percent of enterprises used the Big Data analysis and that puts us in an area below the average of all EU member states.



Picture 1. Use of Big Data analytics in companies in EU member states, 2018 (% of companies)
Source: Eurostat (2021)

Figure 2 represents the data sources of the EU big data analysis in 2018. Forty-nine percent of companies that use big data analytics analyse geolocation data collected using their mobile phones. Twenty-nine percent of the companies analysed the data collected using their own smart devices or sensors whereas data from other sources were analysed by twenty-six percent of companies.



Picture 2. Data sources for Big Data analytics in the EU, 2018 (% of companies using Big Data analysis)
Source: Eurostat (2021)

Data-driven digital platforms and business models are driving Europe's growth in industrial transformation and job creation. With the help of big data technology and the intelligence of digital platforms to access data, European businesses can increase efficiency, profitability and also strengthen their competitive advantage. Nonetheless, industries could also reduce risk and

pave the way for innovation and new jobs. As part of the digital single market strategy, the European Commission has launched a number of initiatives and projects to remove barriers to the use of large amounts of data and digital platforms, aiming to promote innovation in different industries and improve the quality of life of EU citizens.[18]

2.2.2. Examples of Big Data

As mentioned earlier, big data comes from many different sources, such as the Internet of Things (IoT) environment. [19] With customer data as a prime example, diverse branches of analysis can be performed with information from large data sets including benchmarking, social listening, marketing analytics, and analyse customer satisfaction. Comparative advantage means being able to examine metrics that measure user behaviour and customer engagement in real time to compare a company's products, services, and brand reputation. company with competitors. Listening to social media provides information regarding people's opinion on a particular business or product. Those information can be valuable when it comes to determining the targeted audience for marketing activities. Marketing analysis includes collecting informations that are being used to promote products and services. Customer satisfaction and sentiment analysis shows how customers are viewing the company or brand, how to maintain brand loyalty as well as how to further improve customer services.[20]

2.2.3. Storing and processing of Big Data

The necessity to deal with big data places special request on the computing infrastructure. It may happen that single server or group of servers may be overwhelmed by the necessary to process large quantities containing different kinds of information so quickly. To attain the requisite pace, companies must apply appropriate processing capacity to big data jobs. This could necessitate hundreds or thousands of servers capable of distributing processing tasks and cooperating in a clustered design, which is frequently based on Hadoop and Apache Spark. It's also difficult to achieve such speed while being cost-effective. A public cloud provider can store large amounts of data (petabytes) and scale up the number of servers required to execute a big data analytics in just a few hours.[21]

Ways of storing the Big Data in cloud environments is: HDFS (Hadoop Distributed File System), S3 (Amazon Simple Storage Service), NoSQL databases, and also in the previously discussed relational databases. They are all examples of low-cost cloud object storage.

Subsequently, to Hadoop and Spark, enterprises whose wish is to install on-premises big data systems can employ the following Apache open source technologies: the MapReduce programming framework, which is the most important part of Hadoop.[22]

2.2.4. Big data challenges

Aside from adaption ability and financial concerns, users face the additional barrier of developing a big data architecture. Big data chains should be created according to a company's specific demands, which a job of specifically educated people and developers who will cobble use variety of technologies in order to make the system function properly.[23] For instance, database administrators are people who mainly deal with rational software however, working with big data systems requires more knowledge, experience and abilities. Both of these difficulties can be alleviated by adopting controlled cloud service, but IT administrators must monitor cloud usage to ensure that expenses do not spiral out of control. Also moving data sets and processing workloads to the cloud is also a time-consuming procedure. To help analysts identify relevant data, IT and analytics teams are constantly trying to create data catalogues that include metadata management and data lineage services. However, data quality and data governance must be prioritized in order to ensure that collections of big data are clean, consistent, and used correctly.[23]

2.3. Database behind Facebook

MySQL is the most important database that Facebook uses for storing all collected social data. Facebook first used MySQL InnoDB database engine after which they designed MyRocksDB which eventually envolved in RocksDB MySQL database engine allowing software engineers to focus their energies on designing and implementing software declaration.[24] Also, it is a library that solves the problem of abstracting access to local stable storage. It currently runs some of the most demanding database workloads at Facebook and other equally challenging environments. Facebook uses: Apache Hadoop, HBase, Hive, Apache Thrift and PrestoDB for analysing, managing and warehousing the big data.[25]

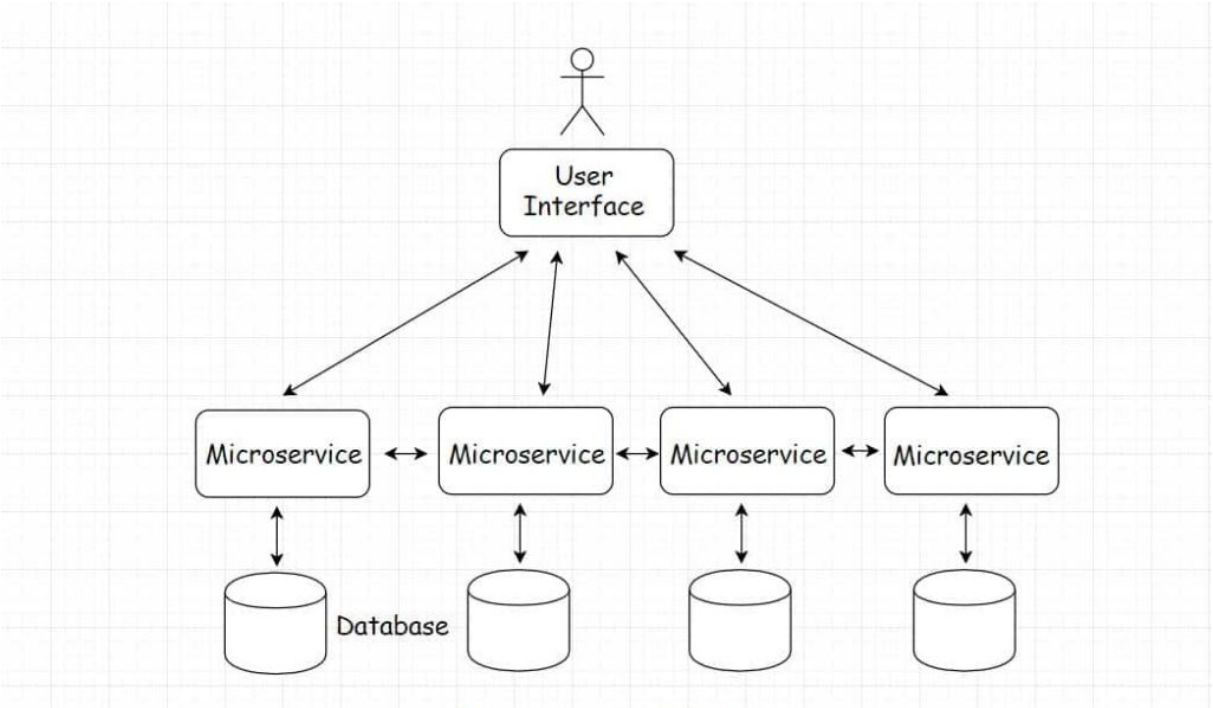
The Apache Hadoop software library is a platform that enables distributed processing of large datasets between groups of computers. It uses a uncomplicated technology and is designed to scale from single servers to thousands of machines, each providing computation and local

storage. It relies on hardware to provide high availability whereas, the library is made to recognise and deal with errors at the application level, thereby providing a high availability service.[25] HBase is defined as a column-oriented non-relational direction system which works on prime of Hadoop Distributed file system. HBase provides an error-tolerant manner of storing distributed data, that are similar to several massive data use cases.[26]

2.3.1. PolyGlot Persistence Architecture

Polyglot Persistence holds the meaning of using various databases while having their unique features and data models for the implementation of various use cases or business requirements. If we have ACID necessities like for a monetary transaction, MySQL Db might serve fine. On the opposite side, whilst we want to get admission to information fast, we'd pass for Memcache. In conditions wherein, we aren't a lot worried about information consistency however want a quick fantastically available, ultimately constant endurance system, a NoSQL answer will be the great answer.[27]

Facebook includes numerous specific openly coupled components comprised collectively like Lego blocks. For example, photo and different content material sharing, messenger, social graph, user post etc. are all one-of-a-kind loosely coupled microservices strolling together. Therefore, each microservice has a separate staying power layer to ensure it is simple to manage.[28]



Picture 3. Microservice Architecture
Source: https://www.researchgate.net/figure/Polyglot-persistence-of-e-commerce-platform_fig1_317395630

2.3.1. Database models used by Facebook

As previously mentioned, Facebook uses quite a lot of databases so it could keep a note of all the data that users share on Facebook. Some of them: MYSQL, HBase, Cassandra, Haystack, Memcached.

However, as the primary database, Facebook uses MySQL to store all data like different wall posts, different user information, their timeline, etc. This particular data flows between their various data centres. Also, Facebook has many reasons why choosing the MYSQL. To start with, it is not hard to manage a large quantity of MYSQL servers, so it becomes easy to provide high-quality services. It has flexible replication capabilities, mainly including the asynchronous replication process and other functions that protect data and help keep data intact.[29]

2.3.2. Storing and processing data by Facebook

In the previous years, Facebook united SMS, chat and email and by doing so formed an unique feature we today call Inbox. In HBase, all of the data/informations are fragmented and then,

marked as regions. Then, a particular region server modifies every region, and a selected region server is accountable for over one region.[30]

Cassandra is an open source NoSQL database which can deal with large quantities of data structure. It additionally gives outstanding service to its users. Cassandra especially makes use of performing the ductility and availability. However, it was designed to resolve the hassle of the storage situations of the Inbox search problem. Inbox search is a function that enables customers to go looking through messages that are located in Facebook Inbox.[31] Apache Cassandra is offering strong sustenance for commodity hardware groups. It was created for Facebook to improve the inbox search tool.[32]

Next, Haystack is a Facebook's most attractive feature. It is a photo's application. Since Facebook users have up until now uploaded more than 15 billion photos, it makes it the biggest content sharing websites of the present time. Facebook generates a 4 images which are all different sizes after each photo is uploaded. This kind of infrastructure combines photo serving tier and also the storage tier altogether into a replacement tier.[33]

3. Protection of data on Facebook

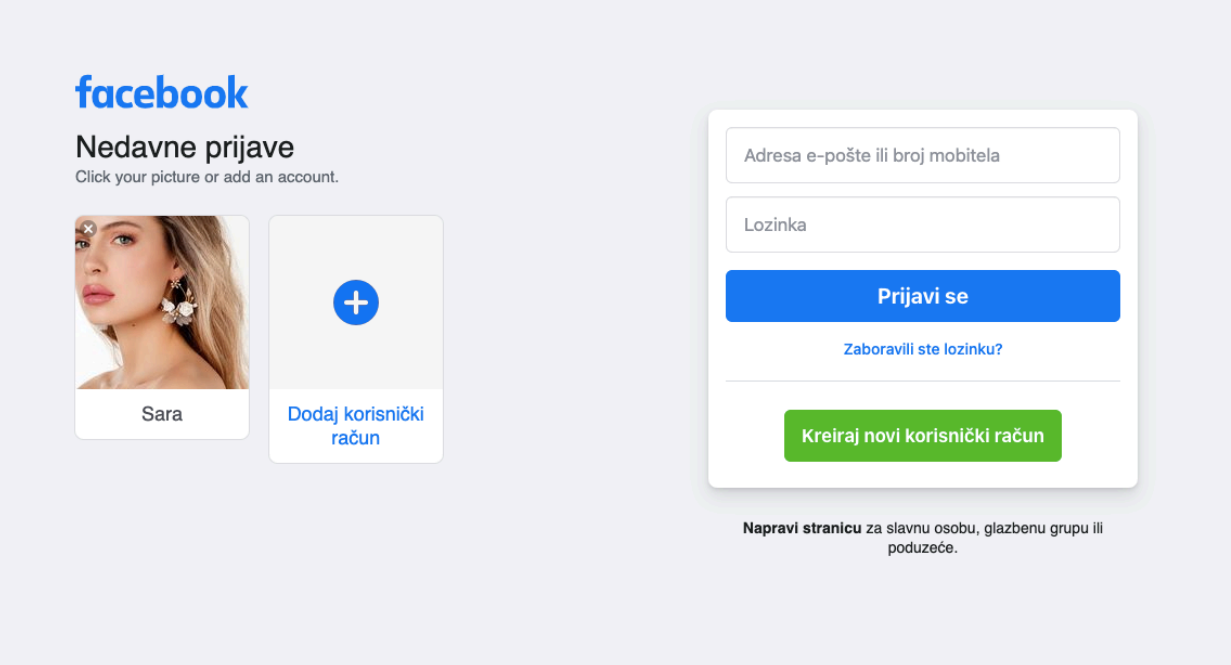
3.1. Protection of data in general

EU data safety regulations assure the safety of private data whenever it is collected, as an instance, whilst one has made an online purchase, has applied for a job vacancy, or has applied for a financial institution loan. The regulations apply to businesses and organizations whether they are public or private, withinside the EU or those primarily based totally outside the EU. However, supplying items or offerings withinside the EU, consisting of Facebook and Amazon, every time those groups are searching for or reuse private statistics of people withinside the EU. Regardless of the form of data, electronic or paper form, data protection rights must be respected.[34]

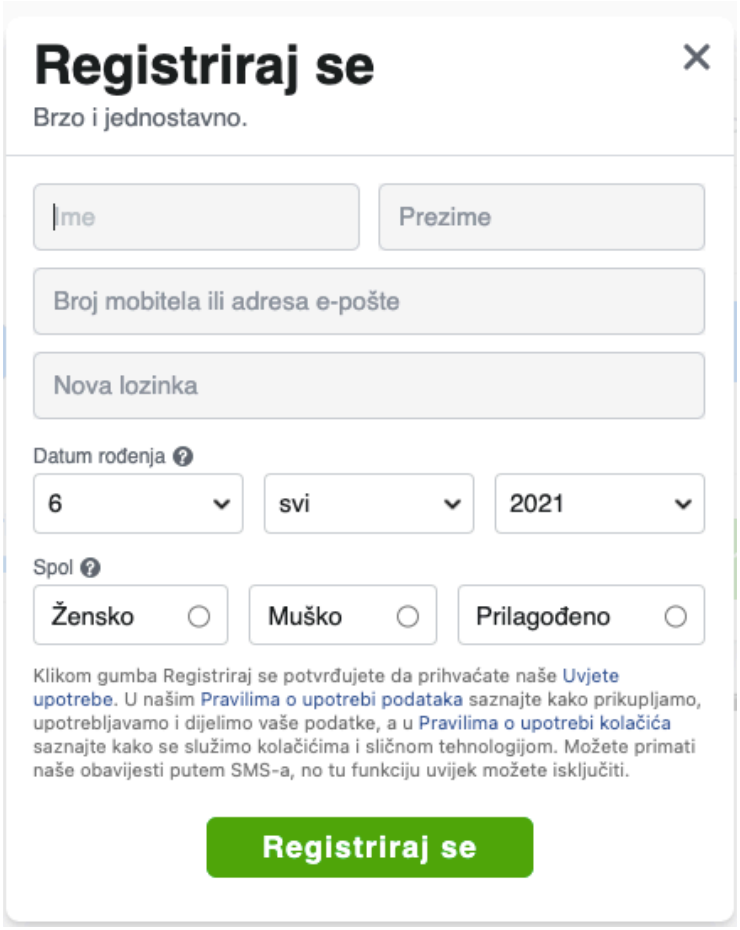
The EU data protection rules described the multiple situations in which companies and organizations may use user's personal data. Like, when some kind of contract was signed, for example on the delivery of goods or services (i.e. buying online) or an employment contract when where legal obligation exist, for example in cases where the use of personal data is legally required. Also, in cases where employer provides information on employee's monthly salary to the social security authority. Furthermore, personal data can be used for the tasks of public administration such as schools, hospitals and municipalities when there are legitimate interests. In all other situations, the company or organization must seek consent before they can collect or reuse your personal information.[35]

3.2. Terms of service

When creating a user account on Facebook, it is necessary to accept the Terms of Service. There are login windows on the home interface if the user has already registered on Facebook, while on the right there is a form for registration for a new user (Picture 1) (Picture 2).



Picture 4. Facebook interface
Source: facebook.com



Picture 5. Registration form for new user
Source: facebook.com

In order for a new user to register, he must accept the terms of use that are visible to every new registered user. The terms of use of Facebook are very extensive and almost no one reads them in full which will be confirmed by following research paper on the concrete sample at the end of the work. When we enter "Terms of use of facebook" in the search engine, the first result obtained is the terms of service, which differs from the previous title “Terms and conditions of use of the service Facebook ”. There are five main points in the Terms: 1. Our Services, 2. Financing of our services, 3. Your obligations to Facebook and our community, 4. Additional Terms and 5. Other terms and conditions that may apply to you (which include manage Facebook ads, privacy policy, cookie policy, policy on use of data, and additional sources are offered that offer a version of the Terms of Service for printing).

The first four points contain over 3000 words and can be read without logging in to Facebook while ten other terms and conditions require user login as parts may be edited or related specifically to an individual user. If a person is accessing from anonymous mode, Facebook offers explanations and education on privacy protection by clicking on the links under ‘Other’. The section that is available to everyone talks about using data for a personalized experience at Facebook, event suggestions, groups or pages, series to watch, or people we want to hang out with connect. That speaks to how much they monitor our behaviour in the Facebook environment. They point out how the ways people communicate is constantly developing and explain how they use the data for ejection as well relevant sponsored ads (content, products and services) that can use people’s data select by user.

This part also applies to tracking user clicks and movements in an online environment. They point out how they fight against harmful behaviour, so they offer options for blocking and removing unwanted content and promise security to the user. Advanced technologies are being used and created such as artificial intelligence, machine learning systems and improved reality to make it more comfortable for everyone to use their services (for example for those with disabilities, people in areas with limited internet access, etc.). To improve services and products, they have teams of people analysing all the data they collected. They also provide the ability to connect contacts and other important content within all products Facebook uses, and they mostly refer to Instagram and Messenger and other applications such as WhatsApp, Oculus,Moves,...

In the section on data usage rules, they briefly point to an additional link where collection and usage of the data is explained, and invite the user to select the level of privacy protection in settings, but this will be discussed more later in the paper. Users obligations towards Facebook and the community are also highlighted and it is defined who can there use their services, which user can share and work on Facebook (it is forbidden insertion of viruses, violation of other people's rights, community standards must be respected, etc.).

The part about the permissions that the user gives to Facebook is also interesting. For example, Facebook asks for permission to use names, profile pictures, and data regarding ads and sponsored content without the user benefiting from it. About the shared content it is stated that "... when a person shares, publishes or transmits content that is covered by the rights intellectual property (such as photographs or videos) connection with our products, preson/user grants us a non-exclusive, transferable license with the right to sublicense and free of charge 11 royalties valid worldwide for storage, use, distribution, modification, launch, copying, performing or displaying, translating and creating derivative works towards user (according to his/hers privacy permissions). Meaning that, if a user shares a content on Facebook, he/she gives permission for its photo or other content to be to stored, copied and shared with others (again, according to his/hers preferences). User may terminate this approval at any time by deleting the content or account. It should be clear that the content a person deletes may still exist, during limited period, in backups. Additionally, the content that has been deleted may still exist in some form if it has been shared with others, and they did not delete it. Almost no one thinks about that part of the story. Also, the user gives permission to update the software, download and install if using Facebook products. In the additional provisions they warn of a possible update of their Terms, but the user will be notified in a timely manner, and if he disagrees with the same he may delete his account. They can terminate the user's account if their terms and conditions are violated.

Disputes are sought to be avoided, but if they occur it is stated on which way of resolving (for the member states of the European Union, the competent courts of the countries are responsible or competent court in the Republic of Ireland where Facebook is headquartered in Europe).

3.3. Data usage rules and privacy protection

This part of the terms of use is most sensitive precisely because it considers users personal data. Only three sentences are written under the Terms of service emphasizing that they collect and use users 'personal information to provide the services described above and the link suggest more detailed information on use of data. They also call on the user to make their own data protection editing choices independently in settings. This point explains what types of data Facebook collects, in what way it uses that data, how data is shared, how Facebook companies work together, which data is to be processed, how the user can exercise his rights provided by the GDPR, information on data retention, deactivation and deletion of the account, etc. The rules on the use of data on Facebook alone contain over 4500 words.

3.4. The Facebook case

The General Data Protection Regulation (GDPR) is the General Data Protection Regulation that started to be present in April 2016 and has been present since 25 May 2018. The most important change that GDPR brings is to protect the personal data of users, clients or employees.[35] Organizations are required to know where what data is located and for what purpose it may be used at any time (GDPR2018, 2018).

An important term when it comes to the protection of personal data is consent. Consent should be given by clear affirmative action expressing the respondent's voluntary, specific, informed and unequivocal consent for the usage of personal data relating to each user, in a form of a written statement.[36] Silence, a pre-ticked field or lack of activity should not be considered as consent. In the case a person chooses to withdraw consent to the use of their personal data, organizations need to be able to do so within the set deadline (GDPR2018, 2018). Personal data informations include: name, address, e-mail, IP and MAC address, GPS location, RFID tags and cookies on web pages, telephone number, photo, videos of individuals, OIB, biometric data including fingerprint and iris imaging, genetic data, education and training data, salary data, credit data indebtedness, bank account information, health data, sexual orientation, voice and many more other.[37]

Person whose identity can be established is a person who can be identified, directly or indirectly with assistance of the previously listed identifiers, which, in addition to the above, also includes

publications on social media networks. Every company in Croatia that collects some of the above data and all foreign companies which collect and process data of European Union citizens are subject to this Regulation (GDPR2018, 2018). Until the appearance of the GDPR, the Law on Personal Data Protection was in force in Croatia To the Republic of Croatia, which ceased to be valid at the time of application of the GDPR.[38]

According to Facebook business, Facebook is well prepared to ensure that their products and services are in line with the GDPR as well as its subsidiaries. The same applies to Instagram as well. Their teams are working on expanding the tools to help users in managing their privacy. The Dublin Data Protection Team has also expanded whereas this regulation applies to citizens of the European Union. They also claim that it is important to their transparency, control and accountability, so in their Data Rules show everything necessary to understand the way people use and process personal data, provide users control the use of their data (simplify design settings privacy in the control centre), and point out that they are responsible for their privacy practices and GDPR compliance (Facebook Business, 2018).[39]

3.5. Data Protection Act in the Republic of Croatia

Prior to the GDPR, the Personal Data Protection Act was in force in Croatia. The consolidated text of the Personal Data Protection Act extends over more than 36 text cards. Article 1 states the following: "This Act regulates the protection of personal data and supervision over the collection, processing and use of personal data in the Republic Croatia''. The purpose of personal data protection is the protection of privacy and other human rights and fundamental freedoms in the collection, processing and use of personal data. Protection of personal data in the Republic of Croatia it is insured for every natural person regardless of citizenship and 19 residence and regardless of race, colour, sex, language, religion, political or other belief, national or social origin, property, birth, education, social status or other characteristics. "(Official Gazette 106/2012, 2012). The Data Protection Act in the Republic of Croatia ceased to be valid when it became General The Personal Data Protection Regulation (GDPR) entered into force on 25 May 2018.[40]

3.6. AZOP - Agency for Personal Data Protection

The agency in charge of implementing the GDPR in Croatia is AZOP, i.e. the Agency for the protection of personal data acting as an independent supervisory body with public powers that is responsible for its work to the Croatian Parliament (AZOP, 2018). As stated on the official website, the Personal Data Protection Agency is legal a person with public authority, who independently performs tasks within the scope of competencies determined by the Act on the Implementation of the General Regulation on Data Protection (Official Gazette), ensuring the implementation of Regulation.[41]

The Agency is established and operates independently of the executive and the legislature authorities, without receiving instructions and orders from any state body.[42] In short, the main task of AZOP is to fulfil all rights and obligations related to protection of personal data imposed on the Republic of Croatia as a full member of the European Union and the Council of Europe and increase the responsibility of all those involved in the processing or using personal data in within the legal framework for personal data protection in the Republic of Croatia. It is their permanent task raising public awareness of the importance of personal data protection, their rights and obligations. The Personal Data Protection Agency actually wants to achieve that privacy protection, respectively the protection of personal data as one of the fundamental human rights becomes a generally accepted principle the work of all those who collect, process and transmit personal data.[43]

3.7. EU – USA Privacy Shield

The disclosure of personal data is an important part of cooperation between the European Union and the United States. These disclosures includescollection and use of personal data. Given that data is mostly used for the global digital economy and social networks collection of personal data that allows identification such as name, date birth, home address, OIB, credit card number, e-mail, gender, marital status and the like. Such Data is very important to protect so that identity theft cannot happen. EU law requires that people's personal data continue to be processed under a high level of protection. This is exactly what the protocol for secure data transfer called “EU-US Privacy Shield” (AZOP, 2018). As for Facebook and Instagram, both social networks have a Privacy Shield certificate.[44] The Privacy Shield has the obligation of providing the user with certain rights. Also, companies are required to protect personal information in in accordance

with the Principles of Privacy.[45] A Privacy Shield Certified Company may use personal information only for that purpose for which the data were originally collected except in the case of subsequent authorization for use personal data for other purposes unless it deviates too much from the original. Also, the company must ensure that personal information is kept in a secure environment and that it is insured against unauthorized access, misuse, loss, destruction and alteration.[48] The user of the service has the right to request a certified company to give him access data and submit them to him with an explanation of the purpose for which they are processed, as well as request a change of personal data if necessary. He can also ask for the deletion of his own data if they are inaccurate, if they are outdated or processed contrary to the rules of the Privacy Shield.[47]

4. Analysis of Research Results

4.1. Research Methodology

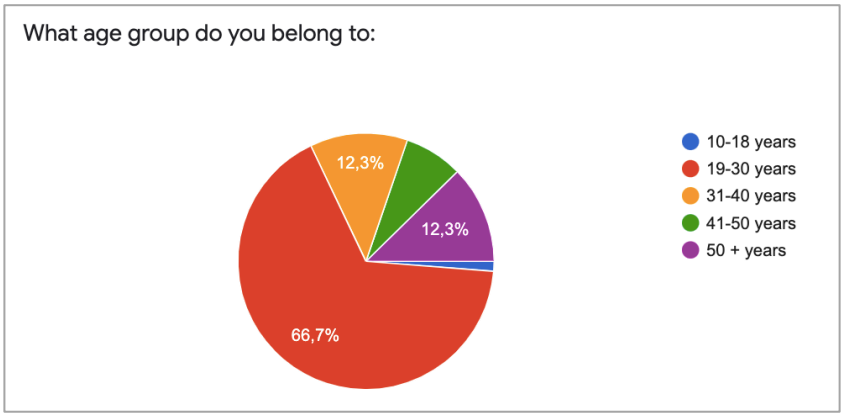
An online survey created with Google Forms was utilized as a research method in the next section of this thesis. One of the most used data-gathering methods is an online survey. It is a set of questions addressed to a specific group of people via email, social media, and websites. The responses of the respondents are saved, and the data can be analysed further. Online surveys are popular because they are simple to use, usually free, and considerably more accessible than traditional surveys. Data processing is simple and straightforward because the feedback is supplied in a digital format. This is a quantitative research technique.

This particular survey is made up of 30 questions divided into 3 sections. Questionnaire consists of following parts: (i) respondents' gender, age and demographic characteristics; (ii) respondents' awareness of where their data is being used (iii) respondents' opinion of the fact that there is a source which has so much information about them. The purpose of this survey is to collect numerical data from a group of respondents in order to get insight into the awareness of privacy regulations on social media, especially Facebook. The survey was completed by 81 participants.

4.2. Research Results

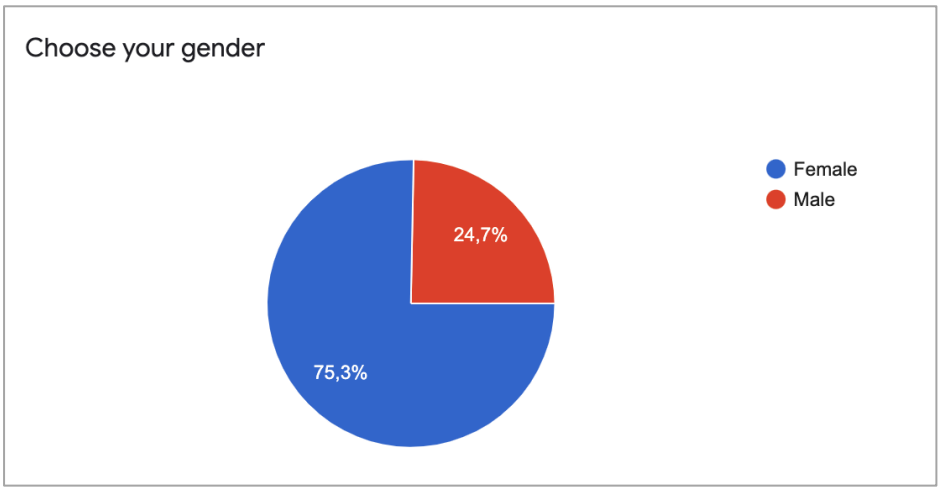
First part of the survey

The first part of questionnaire consists of 4 questions: 1. What age group you belong to, 2. Choose your gender, 3. What is your life status, 4. Where do you come from.



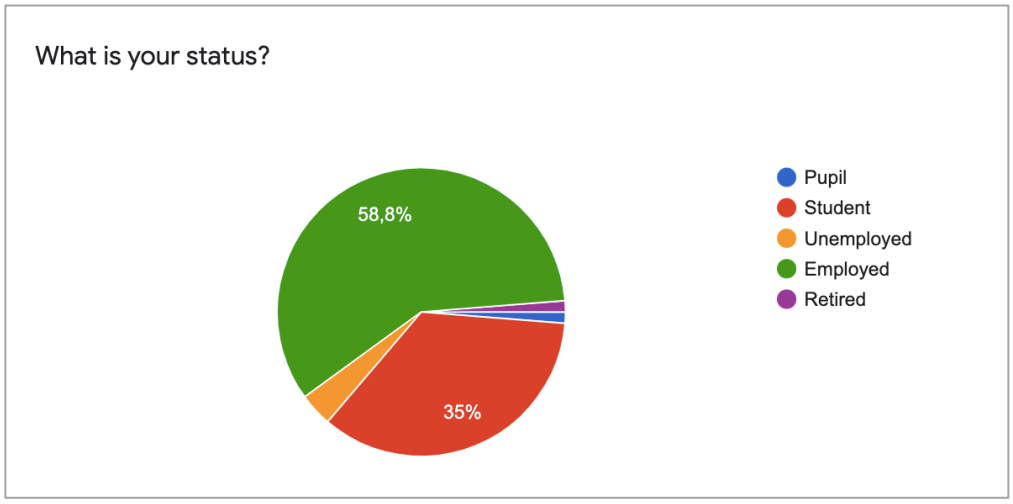
Picture 6. Age group of participants
Source: author's work (2021)

In the first question it can be observed that the majority of participants – 66,7% are between 19 and 30 years old. This is expected as this survey was distributed mostly to family, friends colleagues and other acquaintances. Out of the 81 participants, 1,2% of them are 10-18 years old, 12,3% were 31-40 years old, 7,4% were 41-50 years old and 12,3 were 50 + years old. (Picture 6)



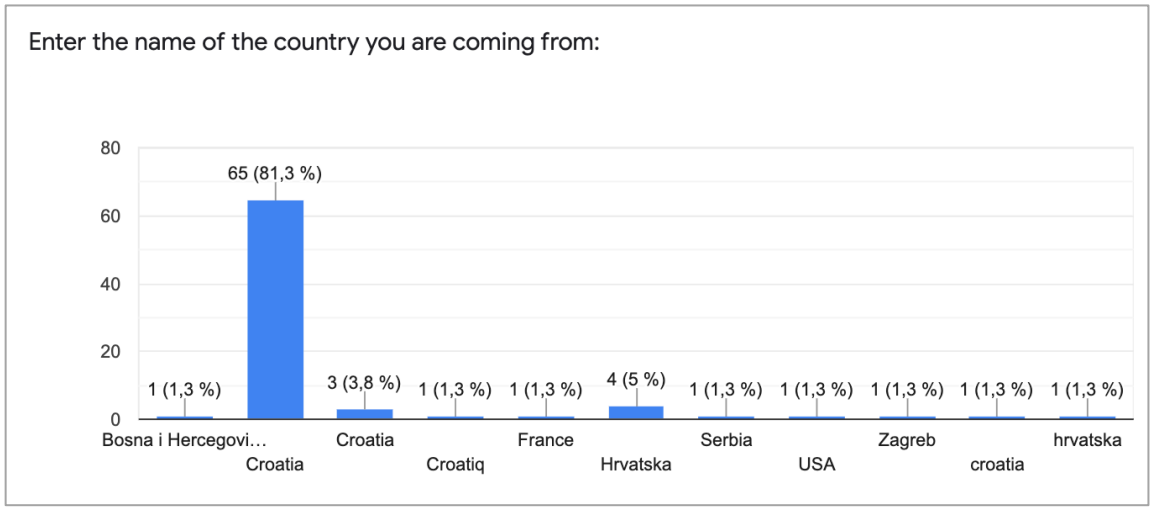
Picture 7. Gender of participants
Source: author's work (2021)

Out of eighty - one participants, 61 (75,3%) was female, the remaining twenty (24,7%) were male. (Picture 7)

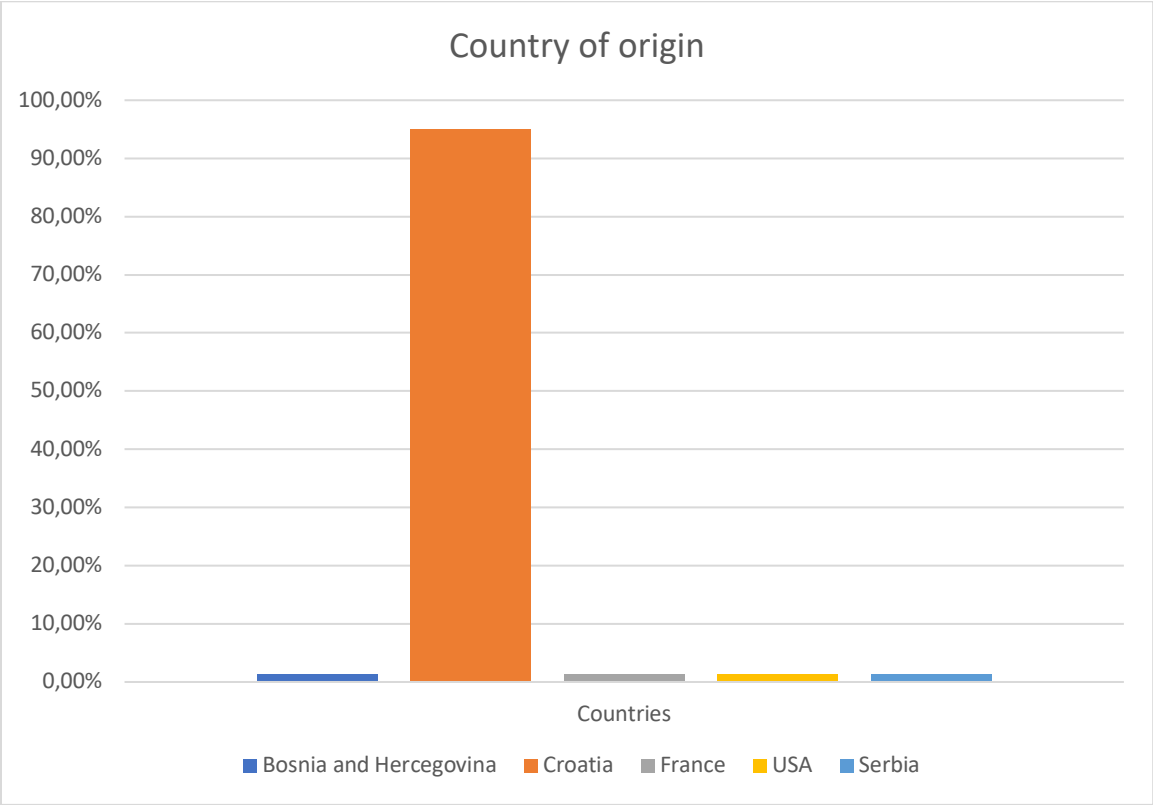


Picture 8. Status of participants.
Source: author's work (2021)

One participant is still pupil (1,3%), 28 participants are students (35%), 3 participants are unemployed (3,7%), 47 participants are employed (58,8%) and one participant is retired (1,2%). (Picture 8)



Picture 9. Country of origin
Source: author's work (2021)

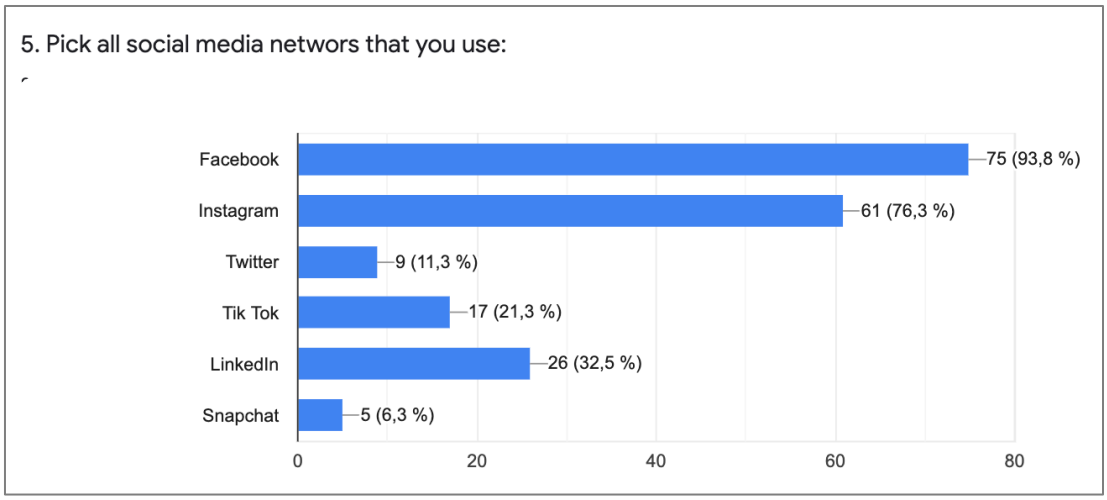


Picture 10. Country of origin
Source: author's work (2021)

One person comes from Bosnia and Hercegovina (1,25%), seventy - six participants (95%) come from Croatia, one person (1,25%) comes from France, one (1,25%) from Serbia and one (1,25%) from the USA. (Picture 10)

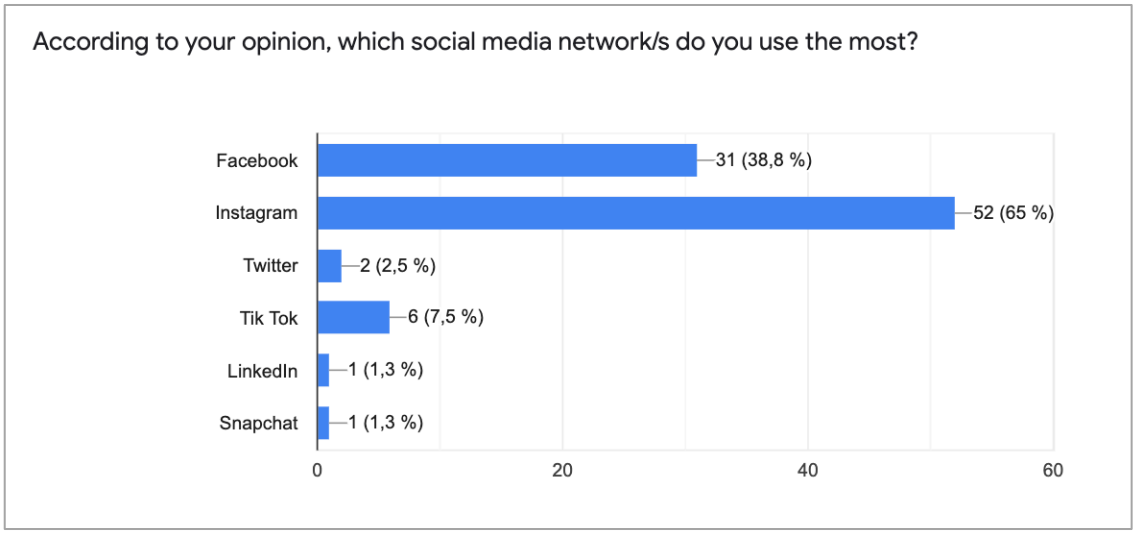
Second part of the survey

The second part of this survey collects data about usage of social media and awareness of the data collection that is happening online.



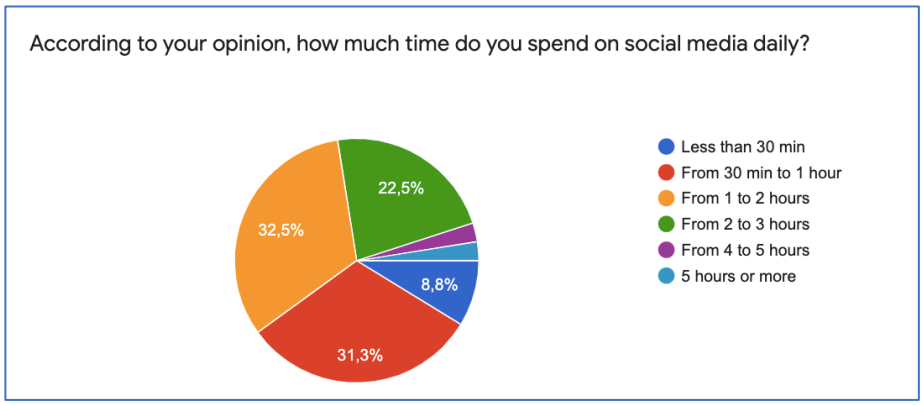
Picture 11. Usage of different social media platforms
Source: author's work (2021)

Here, the participants were asked to pick the social media platform/platforms they use the mist. The most used social media platform according to this survey is Facebook which is being used by 93,8% of the participants. Instagram is the second most used social media platform gathering 76,3% of the participants. LinkedInn is used by 32,5%, Tik Tok by 21,3%, Twitter by 11,3% and Snapchat is used by 6,3% participants. (Picture 11)



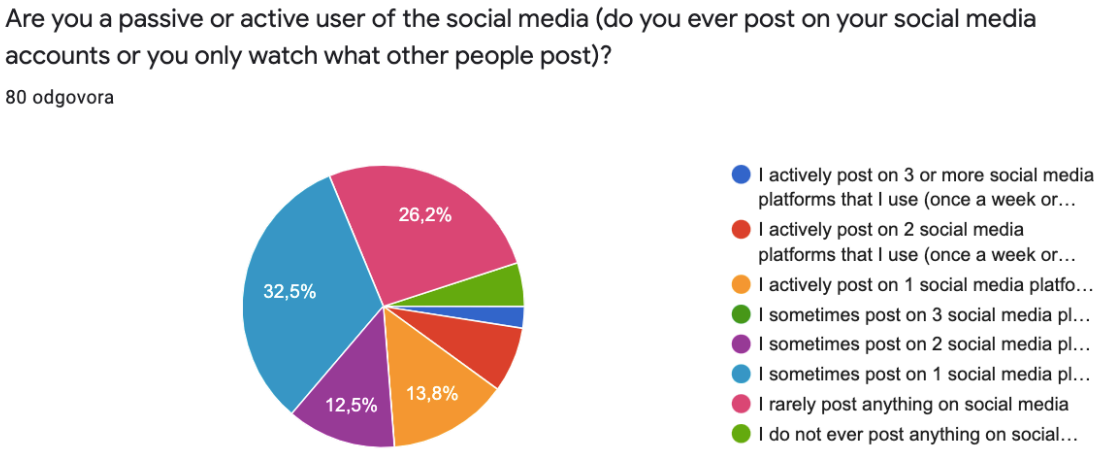
Picture 12. Usage of social media platforms
Source: author's work (2021)

Next, participants were asked which social media platform/platforms they use the most. According to this survey Instagram is the most used social media platform. Participants (65%) stated they use this particular platform most often. Next, 38.8% use Facebook most often and 7,5% use Tik Tok more than other platforms. Twitter, LinkedIn and Snapchat resulted in the lowest popularity among all social media platforms listed. Twitter with only 2,5%, LinkedIn and Snapchat 1,3% each. (Picture 12)



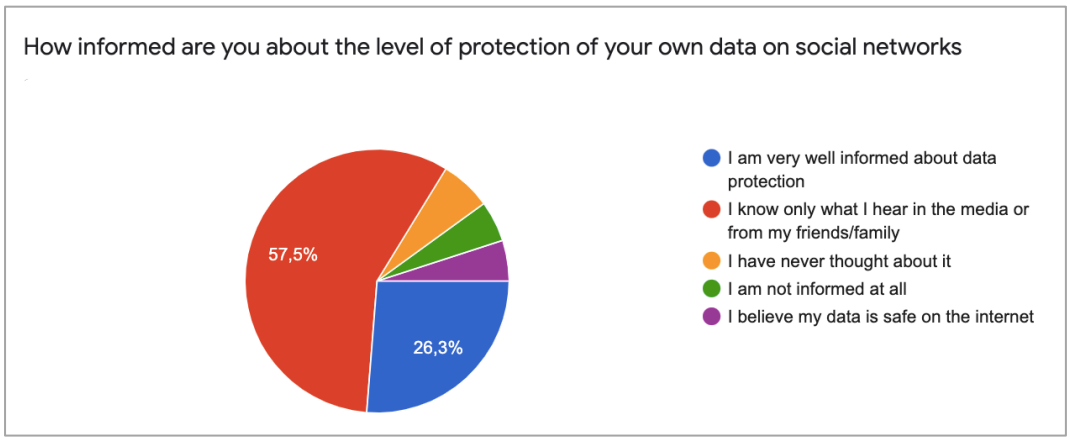
Picture 13. Average time on social media
Source: author's work (2021)

Here participants were asked how much time do they spend on social media daily and given range of answers was from less than thirty minutes to more than five hours. The most participants stated that they spend from one to two hours a day on social media (32,5%) . Next, 31,3% of participants spend a 30 min to 1 hour on social media while 22,5% spend from two to three hours a day on social media. Only 8,8% spend less than thirty min on social media. Lastly, 2,5% spend from four to five hours a day on social media and also 2,5% people spend 5 hours or more on social media. (Picture 13)



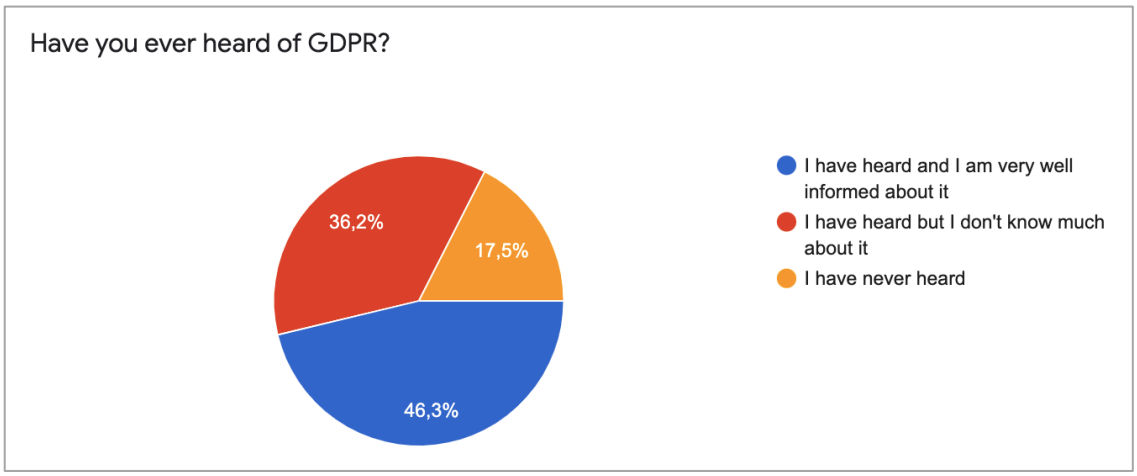
Picture 14. Passive or active users
Source: author's work (2021)

There are two types of users of social media platforms, active and passive. Active users are those users who regularly or unregularly post on platforms they use, while passive users are those who do not share content but only follow what others post. The majority of participants (32,5%) usually post content on one social media account once in a two to three weeks. Second most popular answer for this question was that participants rarely post anything on social media (26,2%). Next, 13,8% stated they actively post (once in a week or more) on one social media platform that they use, 12,5% sometimes post (once in a two to three weeks) on two social media platforms they have, 7,5% actively post (once in a week or more) on two social media platforms, 5% don't ever post on social media and two people 2,5% actively post (once a week or more) on three different social media platforms. (Picture 14)



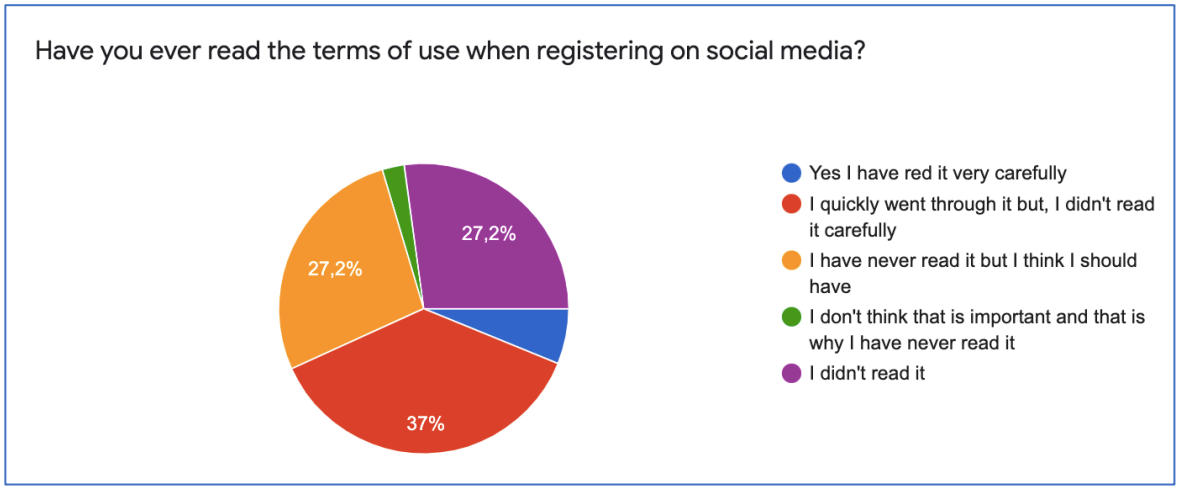
Picture 15. Data protection awareness
Source: author's work (2021)

In this question, 57,5% of participants stated they are informed about data protection only as much as they hear in social media or form friends or family, 26,3% participants are very well informed about the same issue, 6,3% participants have never thought about it while, 5% are both, either not informed at all or have never thought about it. (Picture 15)



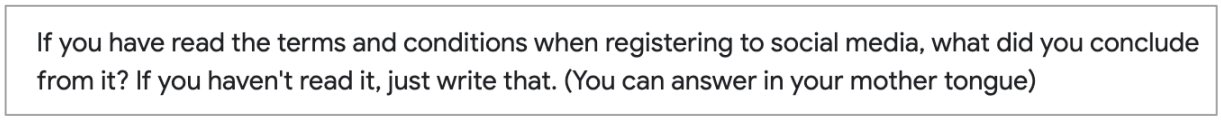
Picture 16. GDPR
Source: author's work (2021)

This question collects data about familiarity of participants with the term GDPR. Most of participants (46,3%) have heard of GDPR and are very well informed about it, 36,3% participants have heard about GDPR but don't know much about it while 17,5% participants have never heard about GDPR. (Picture 16)



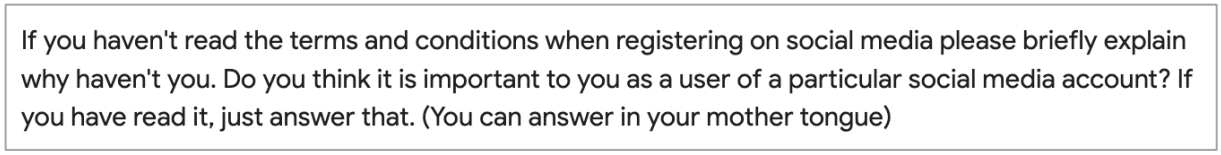
Picture 17. Terms of use
Source: author's work (2021)

The most of participants (37%) have quickly went through terms of use but didn't read it carefully, 27,2% of the participants didn't read the terms of use and the same number of them had never read it but thinks that they should have. Next, 6,2% participants have read the terms of use carefully while 2,5% participants didn't find it important and that is why they did not read it. (Picture 17)



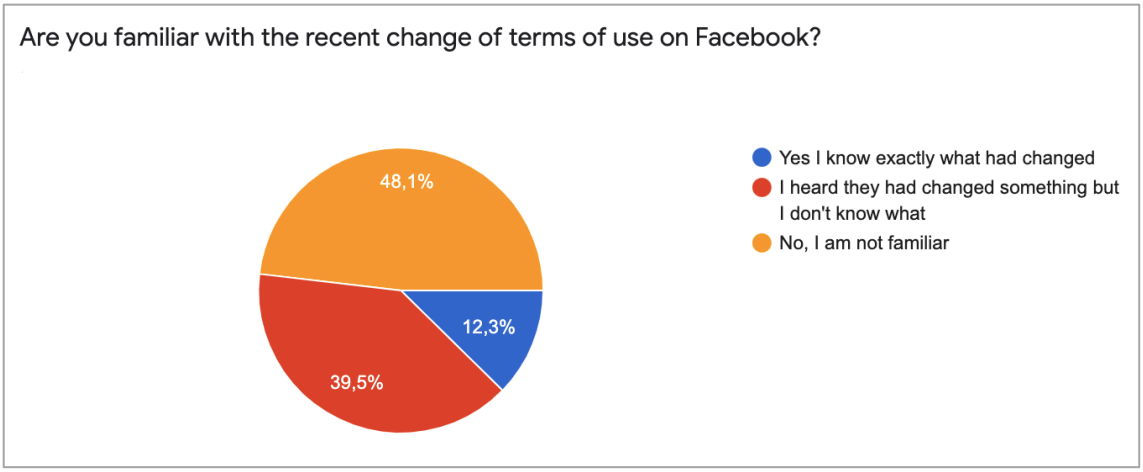
Picture 18. Conclusion after reading terms and conditions

Next question investigates what did participants conclude from terms of use. Some of the answers were:" I haven't read it", " Keep anything you want private to yourself and don't enter that information anywhere online", " I figured someone was following everything we clicked", " That if I don't accept the conditions they set I can't even use that kind of social network", " That the terms and conditions are acceptable to me", " That it is better not to know what we all agree to", " I have read that my data is confidential and as such are protected by these platforms. That my data will not be shared", " A lot of information, written in technical terms that are not clear to everyone", "Because It was boring". (Picture 18)



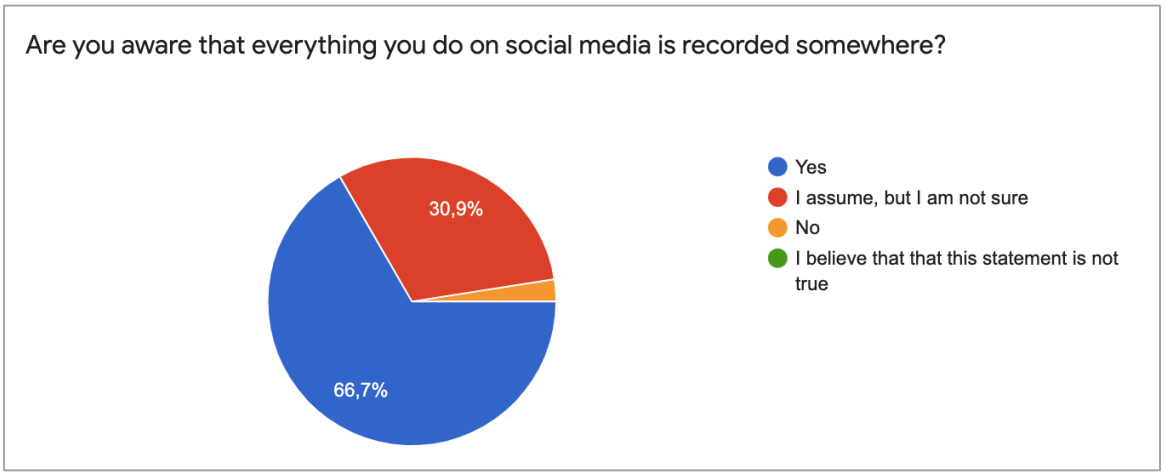
Picture 19 Reason for not reading terms and condition

Next question investigates the reason why people don’t read the terms of use while registering. Here, participants wrote their own answers instead of having a multiple choice question. Some of the answers were: " I don’t know. I’m probably not that aware of the fact that anything can happen on social media", " Because I don't have the patience to read that all those provisions and somehow I hope it can't hurt me too much", "I think that it’s important, but I’m too lazy to read it and considering the fact that everyone uses it, I somehow feel that it’s safe for me to use it too", " No time for that", " I think it’s wrong that I haven’t read the terms of use, but by no means do I catch enough time when registering to read everything. I think too much text was deliberately written so that most people would not be read it", " They violate privacy and analyze personal data anyways", " I read terms and conditions", "It is not important to me", " I didn’t read it because I didn’t feel like reading it and I don’t think I can influence it. I try not to give too detailed information about myself, where I am, etc.", " I haven't read it because I didn't think that it was important", " was too young when opening an account to know the importance of data privacy, so i skimmed the terms and conditions", " I have never been interested, they are too long and confusing", " It is extremely important to know that everything we write, every image, every link we open is followed and results in our profiling which is then used further into various, let's call them marketing purposes, although those purposes can be very different. We need to know what we are agreeing to, only very often we are not".



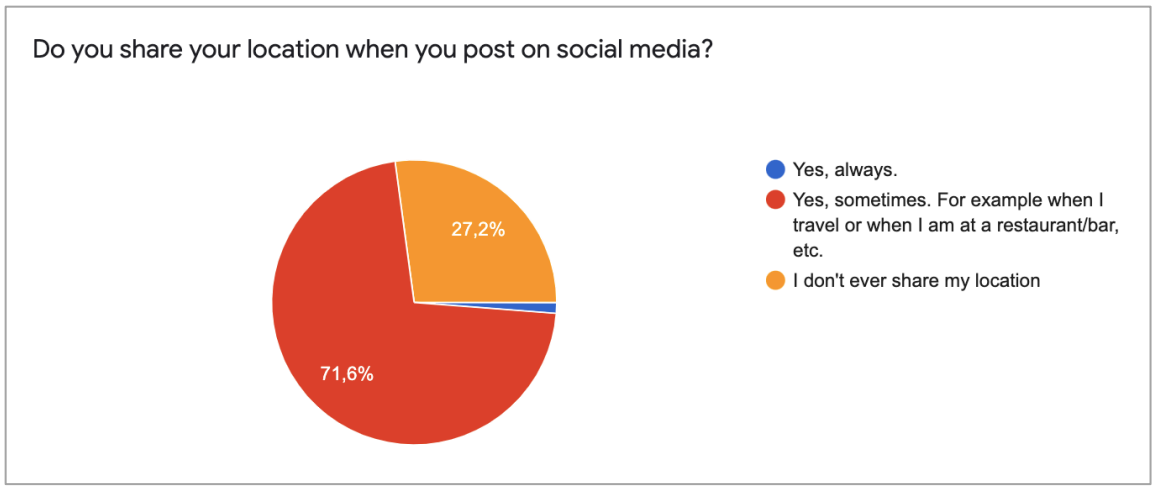
Picture 20. Changes of terms of use

The most of participants (48,1%) were not familiar with the recent change of terms, 19,5% heard something has changed but didn't know what, 12,3% is not familiar. (Picture 20)



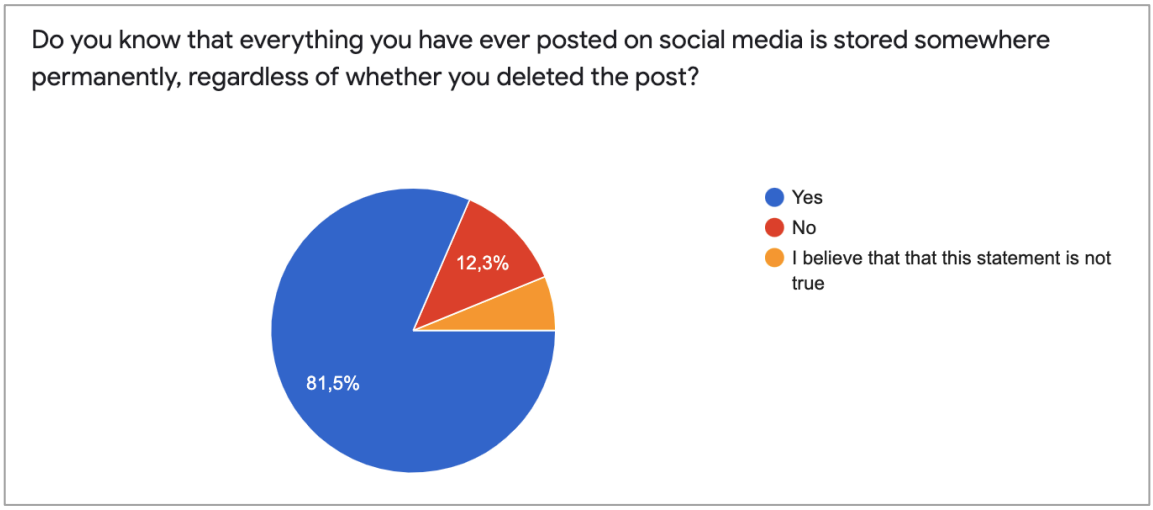
Picture 21. Collecting data awareness

The most participants (66,7%) are aware that their data is being stored, 30,9% assumes data is being stored but is not sure and 2,5% of them is not aware. (Picture 21)



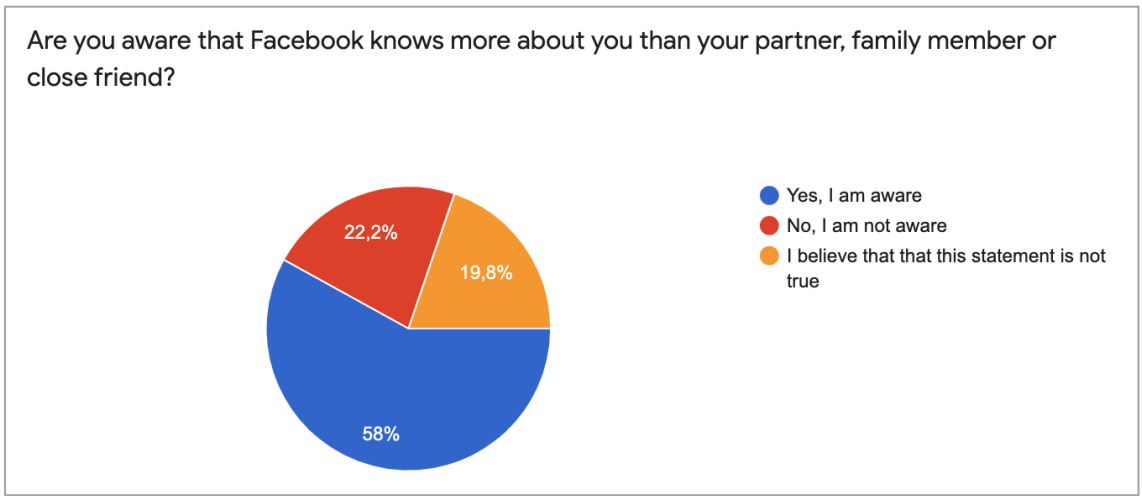
Picture 22. Location

Regarding location sharing when posting on social media, most of the respondents (71,6%) indicate that they share their location when they are traveling or having dinner at a restaurant. Around quarter of the respondents (27,2%) do not share their location. The smallest percentage of respondents always share their location.



Picture 23. Permanently stored data

The most of participants (81,5%) are aware that everything posted on social media is stored somewhere permanently, regardless of whether the post has been deleted, 12,3% of participants are not aware of that and 6,2% believe this is not true (Picture 23).

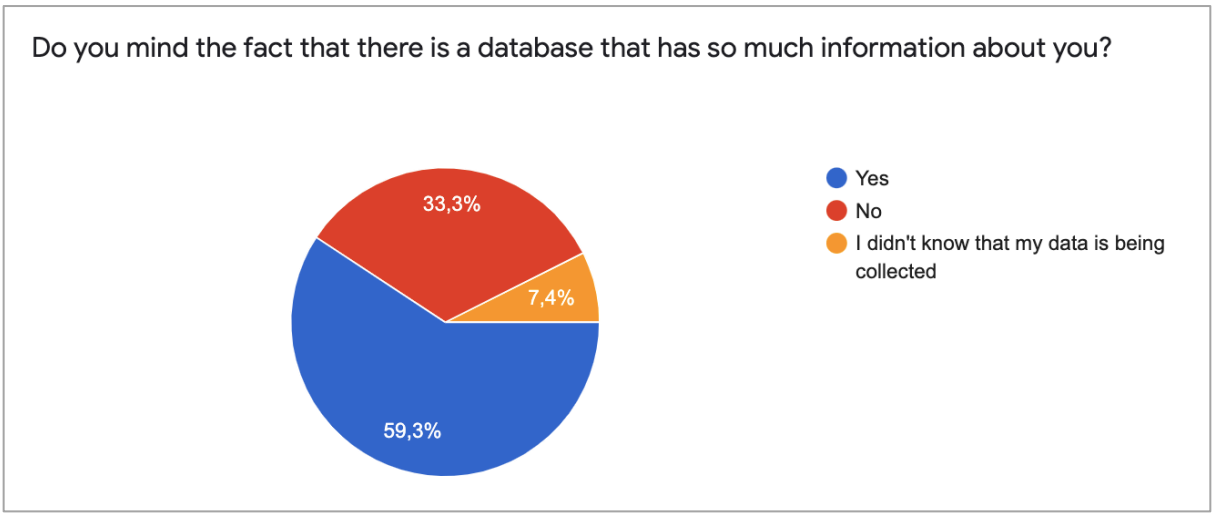


Picture 24. Amount of personal informations on Facebook

Most of the participants (58%) are aware of the fact that Facebook knows more about them than the closest family member or friend, 22,2% is not aware of that fact and 19,8% believes this statement is not true (Picture 24).

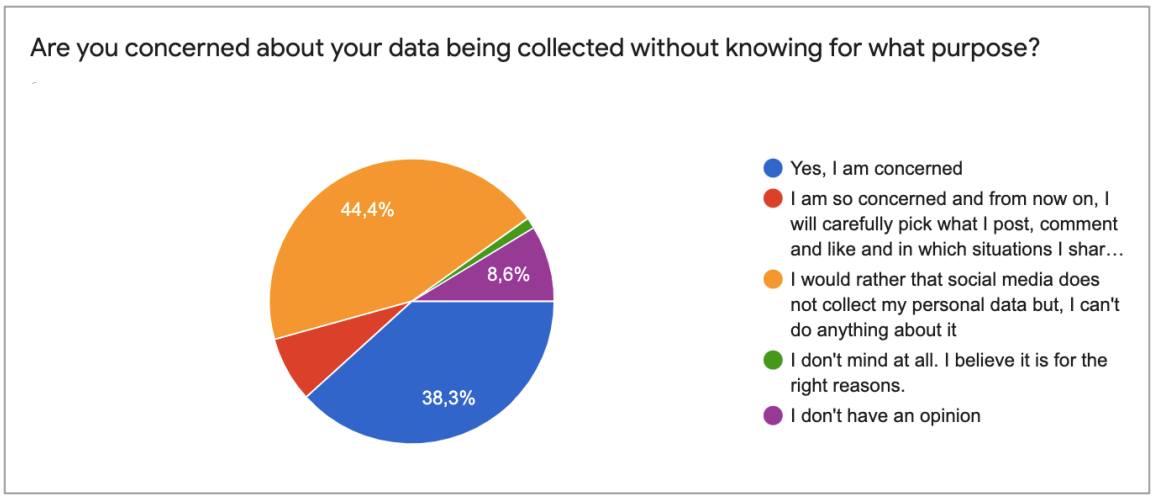
Third part of the survey

Third part of the survey is regarding participant's opinion regarding data collection.



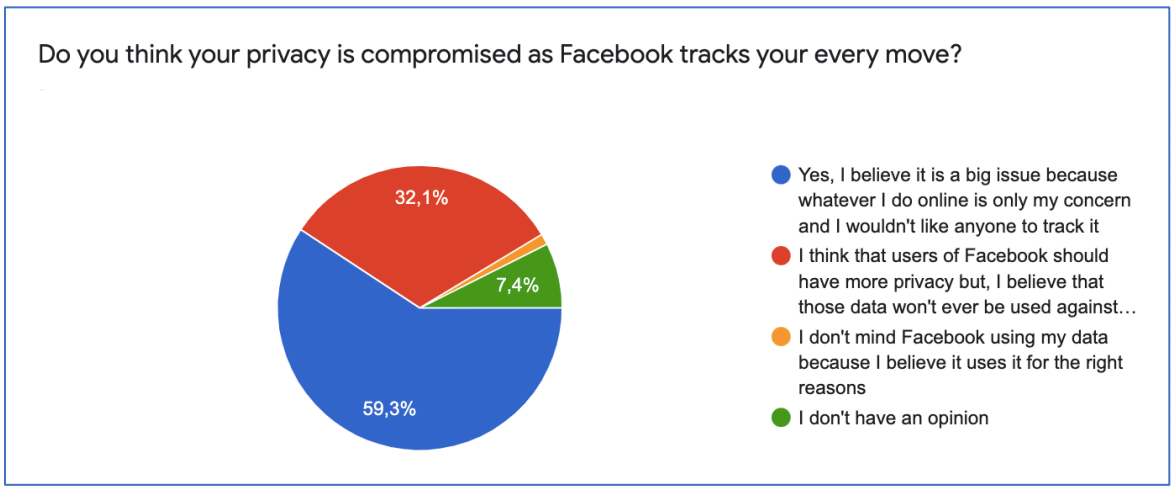
Picture 25. Databases storing personal informations

The most of participants (59,3%) are concerned about existence of database that collects their personal data, 33,3% is not concerned about their data being collected and six of them and 7,4% didn't know their data is being collected (Picture 25).



Picture 26.Collecting users data

The most participants (44,4%) would rather that their data is not being collected but they believe they cannot do anything about it, 38,3% is concerned, 8,6% did not have an opinion, 7,4% is so concerned that from now on they will carefully pick what they post, comment and like and in which situations they share their location, 1,2% don't mind personal data being collected and used without knowing for what purposes at all (Picture 26).



Picture 27. Participant's opinion on Facebook tracking their data

The most of participants (59,3%) believe that data collection is a big issue and that whatever they do online is their concern. Therefore, they wouldn't like anyone to track it, 32,1% believe that users of Facebook should have more privacy but, they believe those data wouldn't ever be used against them, 7,4% don't have opinion and 1,2% doesn't mind Facebook using and collecting his/hers personal data because he/she believes it is for the right reasons (Picture 27).

4.3. Discussion

Personal characteristics – 1st part

From the first part of the survey it can be concluded that majority of participants are between ninety and thirty years old, most of them are female, and currently employed. Nonetheless, majority of people come from Croatia, only four participants are from other countries. As previously mentioned, those outcomes are expected since the survey was distributed among family, friends and colleagues.

Usage of social media and awareness of the data collection – 2nd part

The most participants use Facebook and Instagram. The reason for this could be found in the age range of participants. Facebook was created in 2004, in 2005 it counted over 5,5 million users, in 2006 there were 12 million users, in 2007 more than 20 million, etc. Participants who completed the survey are mostly 19 years old or older. Younger populations use newer social media platforms like TikTok or Snapchat for two reasons. Firstly, there is an entertainment reason. TikTok and Snapchat are more of entertainment platforms. Snapchat is a platform where users record and share their pictures, tag themselves in locations they are at, chat with their friends, etc. TikTok is more of entertainment platform where users post and watch funny videos, dance videos, make up tutorials, etc. Generations that are older than 19 years are interested in more than just entertainment, they want to explore, maybe they use Facebook or Instagram for promotions of company they work for or themselves, they are maybe part of some important Facebook group where they get informations regarding their exams or studies, specific field they are interested in, etc. Secondly, population of 10+ got used comfort to social media platforms of their age and don't have enough motivation to switch to another. Since the entire era of social media platforms first started with Facebook and then Instagram, many of those 19+ age groups (from this survey) probably created their accounts on Facebook and Instagram and feel no need to switch to another social media platform.

On the other hand, when it comes to reading terms and conditions, only 6,2% of participants have read it, 37% quickly went through it and 27% have never read it. Terms and conditions consist of text that is very demanding and difficult to understand. It may be much for an average person to read just so he/she could register to desired social media platform. This is probably why people tend to skip it when registering.

The General Data Protection Rule (GDPR) 2016/679 is a European Union and European Economic Area regulation on data protection and privacy. From the survey it can be seen that 17,5% of people have never heard about GDPR, 36% have heard but don't know much about it and 46,3% have heard and is very well informed about it. GDPR is important regulation that all participants of social media should know about. From this numbers it can be assumed that people are not interested in their privacy online and where and how their data is being used.

Also, 2,5% of participants believe social media doesn't record what users are watching, liking, commenting and searching while being online while 31,7% isn't sure whether it is true or not. The terms of use of Facebook for instance, in the beginning of text state that users data is being used in order to connect a certain user with his/hers potential friends, groups he/she might like, places he might go to, commercials he might like, etc. From this it can be easily concluded that user's activity is being observed because if not, connecting an user with potential people or things he might like, buy or follow would not be possible. This gives us overview of how uninformed those 34,2% of participants are regarding their privacy online. In addition, 12,2% of participants were not aware that everything they post is stored somewhere, while 6,1% believe it is not true. This is especially dangerous for young people because a lot of their posts could be used against them in the future.

Respondents' opinion of the fact that there is a source which has so much information about them – 3rd part

Social media users mind that their data is being used and stored. They would rather that this isn't the case. However, according to this survey most users believe that collection of their data won't harm them in any way.

5. Conclusion

People’s desire to socialize always found new ways of interactions which, in conjunction with the rapid development of new technology, define sociability in a whole new level. Facebook is one of the most popular social networks which currently counts more than half billions of users. The authors of Facebook say it is their mission to make the world open and connected and that in the beginning of development of this network cared exclusively about attracting new users, neglecting safety aspects while doing so. As the network grows, its users share data with other parties and do not care how much and what information they really want to share with others.

To some extent, we can control the privacy of our own data through the settings we use. Privacy policy that defines Facebook, determine that name, surname and photo are next to profile picture, always publicly available. It is also third parties are prohibited from using the collected data in any their purpose. Although they have been introduced recently, many privacy enhancements and add-ons, Facebook can hardly have complete control over it.

Terms of use are widely accepted, and research has confirmed the assumption that almost no one reads it because people don't even care what they agree to. Privacy settings are something that the user can determine by its own criteria, but from the rules of data use we see that these networks collect everything about us. Privacy actually determines the protection of data on the outside, according to other users, but not to those who collect and use users data. It's GDPR in this sphere brought great positive changes, it would be better if users were more informed, but it is not necessary for their better protection.

In short, it seems that we cannot know what is being done with our personal data and how is it used. The rules on the use of data must specify what the data are used for and this should be adhered to, but who can confirm the truth of the above statements? Privacy on the internet is an abstract term and I would say that privacy on the internet does not exist. Perceptions of people on this topic varies depending on what people think is important to protect at all. Terms of use exist to keep platforms from liability for user privacy and how would provide guidance and explanations on how they work. I would end with a sentence that has often appeared on the internet when it comes to social networks and is one of Google’s first social media search results. This statement is by Pete Cashmore, founder and CEO of Mashable, a popular blog

about social media, it is very short and clear: “Privacy is dead and social media is holding a gun from which it smokes”.

6. List of References

1. C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, The Cambridge Structural Database, Acta Cryst. (2016). B72, 171-179

2. R Elmasri, SB Navathe, R Elmasri, SB Navathe Springer, Fundamentals of Database, 24th European Conference, ADBIS 2020, Lyon, France, August 25–27, 2020

3. Jian-Wei Lin, Yuan-Cheng Lai and Yuh-Shy Chuang, Timely Diagnostic Feedback for Database Concept Learning, Educational Technology & Society Vol. 16, No. 2, Grand Challenges and Research Directions in e-Learning of the 21th Century (April 2013)

4. Myung Jin Choi, Joseph J. Lim, Antonio Torralba, Alan S. Willsky, Exploiting hierarchical context on a large database of object categories, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition

5. Renzo Angles, A Comparison of Current Graph Database Models, 2012 IEEE 28th International Conference on Data Engineering Workshops, 10.1109/ICDEW.2012.31

6. Dorothea Koppisch, Jody Schinkel, Stefan Gabriel, Wouter Fransman, Erik Tielemans, Use of the Mega Exposure Database for the Validation of the Stoffenmanager Model, The Annals of Occupational Hygiene, Volume 56, Issue 4, May 2012

7. E. F. Codd, Data Models in Database Management, IBM Research Laboratory San Jose, California 95193

8. Mr. SANTOSH KUMAR Assistant Professor,Department of Computer Science & Engg. Dr APJ Abdul Kalam Women's Institute of Technology Lalit Narayan Mithila University, Darbhanga, Bihar, Data Models in DBMS

9. E. F. Codd, A Relational Model of Data for Large Shared Data Banks, M Research Laboratory, San Jose, California

10. Passeri, F., Comina, C., Foti, S. et al. The Polito Surface Wave flat-file Database (PSWD): statistical properties of test results and some inter-method comparisons. Bull Earthquake Eng 19, 2343–2370 (2021). https://doi.org/10.1007/s10518-021-01069-1

11. Jan Paredaens, Paul De Bra, Marc Gyssens, Dirk van Gucht, The Structure of the Relational Database Model

12. X. Lu; D. Colbry; A.K. Jain, Three-dimensional model based face recognition, IEEE, DOI: 10.1109/ICPR.2004.1334127, 26 Aug. 2004

13. Rolf Engelbrecht, Connecting Medical Informatics and Bio-informatics

14. Marx, V. The big challenges of big data. Nature 498, 255–260 (2013).

15. X. Wu, X. Zhu, G. -Q. Wu and W. Ding, "Data mining with big data," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, Jan. 2014, doi: 10.1109/TKDE.2013.109.

16. Jianqing Fan, Fang Han, Han Liu, Challenges of Big Data analysis, National Science Review, Volume 1, Issue 2, June 2014, Pages 293–314

17. S. Sagiroglu and D. Sinanc, "Big data: A review," 2013 International Conference on Collaboration Technologies and Systems (CTS), 2013, pp. 42-47, doi: 10.1109/CTS.2013.6567202.

18. S. Madden, "From Databases to Big Data," in IEEE Internet Computing, vol. 16, no. 3, pp. 4-6, May-June 2012, doi: 10.1109/MIC.2012.50.

19. Y. Edmund Lien, On the Equivalence of Database Models, Journal of the ACM Volume 29 Issue 2April 1982

20. Hakan Özköse, Emin Sertaç Arı, Cevriye Gencer, Yesterday, Today and Tomorrow of Big Data, Procedia - Social and Behavioral Sciences, Volume 195, 2015, Pages 1042-1050

21. Rubén Casado, Muhammad Younas, Emerging trends and technologies in big data processing, Volume27, Issue8 Special Issue: Combined Special Issues on The Internet of Things: shaping the new Internet space and Advances on cloud services and cloud computing 10 June 2015 Pages 2078-2091

22. Alexandros Labrinidis and H. V. Jagadish. 2012. Challenges and opportunities with big data. VLDB Endow 5, 12 (August 2012), 2032–2033. DOI:https://doi.org/10.14778/2367502.2367572

23. Alexandru Adrian, Big Data Challenges, Database Systems Journal vol. IV, no. 3/2013

24. Lior Abraham, John Allen, Oleksandr Barykin, Vinayak Borkar, Bhuwan Chopra, Ciprian Gerea, Daniel Merl, Josh Metzler, David Reiss, Subbu Subramanian, Janet L. Wiener, and Okay Zed. 2013. Scuba: diving into data at facebook. Proc. VLDB Endow. 6, 11 (August 2013), 1057–1067. DOI:https://doi.org/10.14778/2536222.2536231

25. Vinod Kumar Vavilapalli, Arun C. Murthy, Chris Douglas, Sharad Agarwal, Mahadev Konar, Robert Evans, Thomas Graves, Jason Lowe, Hitesh Shah, Siddharth Seth, Bikas Saha, Carlo Curino, Owen O'Malley, Sanjay Radia, Benjamin Reed, and Eric Baldeschwieler. 2013. Apache Hadoop YARN: yet another resource negotiator. In Proceedings of the 4th annual Symposium on Cloud Computing (SOCC '13). Association for Computing Machinery, New York, NY, USA, Article 5, 1–16. DOI:https://doi.org/10.1145/2523616.2523633

26. Dhruba Borthakur, Jonathan Gray, Joydeep Sen Sarma, Kannan Muthukkaruppan, Nicolas Spiegelberg, Hairong Kuang, Karthik Ranganathan, Dmytro Molkov, Aravind Menon, Samuel Rash, Rodrigo Schmidt, and Amitanand Aiyer. 2011. Apache hadoop goes realtime at Facebook. In Proceedings of the 2011 ACM SIGMOD International

Conference on Management of data (SIGMOD '11). Association for Computing Machinery, New York, NY, USA, 1071–1080. DOI:https://doi.org/10.1145/1989323.1989438

27. Carmi E. Rhythmedia: A Study of Facebook Immune System. Theory, Culture & Society. 2020;37(5):119-138. doi:10.1177/0263276420917466

28. Aakash Goel, Bhuwan Chopra, Ciprian Gerea, Dhruv Mátáni, Josh Metzler, Fahim Ul Haq, and Janet Wiener. 2014. Fast database restarts at facebook. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD '14). Association for Computing Machinery, New York, NY, USA, 541–549. DOI:https://doi.org/10.1145/2588555.2595642

29. Khine, P.P.; Wang, Z. A Review of Polyglot Persistence in the Big Data World. Information 2019, 10, 141. https://doi.org/10.3390/info10040141

30. Gašpar, D., & Mabić, M. (2017). NoSQL Databases as Social Networks Storage Systems. Proceedings of the ENTRENOVA - ENTerprise REsearch InNOVAtion Conference, 3(1), 367–373. Retrieved from https://hrcak.srce.hr/ojs/index.php/entrenova/article/view/14100

31. Jones, Milo and Silberzahn, Philippe. Constructing Cassandra: Reframing Intelligence Failure at the CIA, 1947–2001, Redwood City: Stanford University Press, 2013. https://doi.org/10.1515/9780804787154

32. S. Prasad and M. S. N. Sha, "NextGen data persistence pattern in healthcare: Polyglot persistence," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-8, doi: 10.1109/ICCCNT.2013.6726734.

33. M. Dawodi, M. H. Hedayati, J. A. Baktash and A. L. Erfan, "Facebook MySQL Performance vs MySQL Performance," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019,

pp. 0103-0109, doi: 10.1109/IEMCON.2019.8936259.

34. Ashish Thusoo, Zheng Shao, Suresh Anthony, Dhruba Borthakur, Namit Jain, Joydeep Sen Sarma, Raghotham Murthy, and Hao Liu. 2010. Data warehousing and analytics infrastructure at facebook. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, Association for Computing Machinery, New York, NY, USA, 1013–1020. DOI:https://doi.org/10.1145/1807167.1807278

35. Qi Huang, Ken Birman, Robbert van Renesse, Wyatt Lloyd, Sanjeev Kumar, and Harry C. Li. 2013. An analysis of Facebook photo caching. In Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles. Association for Computing Machinery, New York, NY, USA, 167–181. DOI:https://doi.org/10.1145/2517349.2522722

36. Goddard M. The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact. International Journal of Market Research. 2017;59(6):703-705. doi:10.2501/IJMR-2017-050

37. Y. Xu, E. Frachtenberg, S. Jiang and M. Paleczny, "Characterizing Facebook's Memcached Workload," in IEEE Internet Computing, vol. 18, no. 2, pp. 41-49, Mar.-Apr. 2014, doi: 10.1109/MIC.2013.80.

38. Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376321

39. Amelia Acker, Adam Kreisberg. (2020) Social media data archives in an API-driven world. Archival Science 20:2, pages 105-123.

40. J. Isaak and M. J. Hanna, "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection," in Computer, vol. 51, no. 8, pp. 56-59, August 2018, doi:

10.1109/MC.2018.3191268.

41. Puljak, L., Mladinić, A., Iphofen, R. i Koporc, Z. (2020). Before and after enforcement of GDPR: Personal data protection requests received by Croatian Personal Data Protection Agency from academic and research institutions. Biochemia Medica, 30 (3), 363-370. https://doi.org/10.11613/BM.2020.030201

42. Starčević, K., Crnković, B., & Glavaš, J. (2018). IMPLEMENTATION OF THE GENERAL DATA PROTECTION REGULATION IN COMPANIES IN THE REPUBLIC OF CROATIA. Ekonomski Vjesnik, 31(1), 163-176. Retrieved from https://www.proquest.com/scholarly-journals/implementation-general-data-protection-regulation/docview/2066622287/se-2?accountid=132154

43. Zuiderveen Borgesius, F., Poort, J. Online Price Discrimination and EU Data Privacy Law. J Consum Policy 40, 347–366 (2017). https://doi.org/10.1007/s10603-017-9354-z

44. B K Sovacool,D D Furszyfer Del Rio,S Griffiths,Policy mixes for more sustainable smart home technologies, Environmental Research Letters, 16, 5, (054073), (2021).

45. Mostert, M., Bredenoord, A., Biesaart, M. et al. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 24, 956–960 (2016). https://doi.org/10.1038/ejhg.2015.239

46. Voss, W. Gregory. “European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting.”,The Business Lawyer, vol. 72, no. 1, 2016, pp. 221–234. JSTOR, www.jstor.org/stable/26419118. Accessed 10 Sept. 2021.

47. A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua and S. Guo, "Protection of Big Data Privacy," in IEEE Access, vol. 4, pp. 1821-1834, 2016, doi: 10.1109/ACCESS.2016.2558446.

7. Appendix

Survey questions

What age group do you belong to:

- 10-18 years
- 19-30 years
- 31-40 years
- 41-50 years
- 50 + years

Choose your gender

- Female
- Male

What is your status?

- Pupil
- Student
- Unemployed
- Employed
- Retired

Enter the name of the country you are coming from:

Tekst kratkog odgovora

5. Pick all social media networs that you use:

- [ ] Facebook
- [ ] Instagram
- [ ] Twitter
- [ ] Tik Tok
- [ ] 3
- [ ] Snapchat

According to your opinion, which social media network/s do you use the most?

- [ ] Facebook
- [ ] Instagram
- [ ] Twitter
- [ ] Tik Tok
- [ ] LinkedIn
- [ ] Snapchat

According to your opinion, how much time do you spend on social media daily?

- ( ) Less than 30 min
- ( ) From 30 min to 1 hour
- ( ) From 1 to 2 hours
- ( ) From 2 to 3 hours
- ( ) From 4 to 5 hours
- ( ) 5 hours or more

Are you a passive or active user of the social media (do you ever post on your social media accounts or you only watch what other people post)?

- I actively post on 3 or more social media platforms that I use (once a week or more)
- I actively post on 2 social media platforms that I use (once a week or more)
- I actively post on 1 social media platform that I use (once a week or more)
- I sometimes post on 3 social media platforms that I use. (once in a two to three weeks)
- I sometimes post on 2 social media platforms that I use (once in a two to three weeks)
- I sometimes post on 1 social media platform that I use (once in a two to three weeks)
- I rarely post anything on social media
- I do not ever post anything on social media

How informed are you about the level of protection of your own data on social networks

- I am very well informed about data protection
- I know only what I hear in the media or from my friends/family
- I have never thought about it
- I am not informed at all
- I believe my data is safe on the internet

Have you ever heard of GDPR?

- I have heard and I am very well informed about it
- I have heard but I don't know much about it
- I have never heard

Have you ever read the terms of use when registering on social media?

- ◯ Yes I have red it very carefully
- ◯ I quickly went through it but, I didn't read it carefully
- ◯ I have never read it but I think I should have
- ◯ I don't think that is important and that is why I have never read it
- ◯ I didn't read it

If you have read the terms and conditions when registering to social media, what did you conclude from it? If you haven't read it, just write that. (You can answer in your mother tongue)

Tekst dugog odgovora

If you haven't read the terms and conditions when registering on social media please briefly explain why haven't you. Do you think it is important to you as a user of a particular social media account? If you have read it, just answer that. (You can answer in your mother tongue)

Tekst dugog odgovora

Are you familiar with the recent change of terms of use on Facebook?

- ◯ Yes I know exactly what had changed
- ◯ I heard they had changed something but I don't know what
- ◯ No, I am not familiar

Are you aware that everything you do on social media is recorded somewhere?

- ◯ Yes
- ◯ I assume, but I am not sure
- ◯ No
- ◯ I believe that that this statement is not true

Do you share your location when you post on social media?

- ◯ Yes, always.
- ◯ Yes, sometimes. For example when I travel or when I am at a restaurant/bar, etc.
- ◯ I don't ever share my location

Do you know that everything you have ever posted on social media is stored somewhere permanently, regardless of whether you deleted the post?

- ◯ Yes
- ◯ No
- ◯ I believe that that this statement is not true

Are you aware that Facebook knows more about you than your partner, family member or close friend?

- ◯ Yes, I am aware
- ◯ No, I am not aware
- ◯ I believe that that this statement is not true

Do you mind the fact that there is a database that has so much information about you?

- ◯ Yes
- ◯ No
- ◯ I didn't know that my data is being collected

Are you concerned about your data being collected without knowing for what purpose?

- ◯ Yes, I am concerned
- ◯ I am so concerned and from now on, I will carefully pick what I post, comment and like and in which situati…
- ◯ I would rather that social media does not collect my personal data but, I can't do anything about it
- ◯ I don't mind at all. I believe it is for the right reasons.
- ◯ I don't have an opinion

Do you think your privacy is compromised as Facebook tracks your every move?

- ◯ Yes, I believe it is a big issue because whatever I do online is only my concern and I wouldn't like anyone to…
- ◯ I think that users of Facebook should have more privacy but, I believe that those data won't ever be used a…
- ◯ I don't mind Facebook using my data because I believe it uses it for the right reasons
- ◯ I don't have an opinion

SARA DJURIĆ

PERSONAL

Name
Sara Djurić

Address
Zmajevac 18
10000 Zagreb

Phone number
0917343686

Email
sara.djuric@hotmail.com

LANGUAGES

Croatian ●●●●●

English ●●●●●

Italian ●○○○○

German ●○○○○

WORK EXPERIENCE

Fashion model Sep 2016 - Sep 2018
Women Management Paris, Paris

As a model I was doing fashion shows, shootings and presentations.

Public Relations Sep 2016 - Jul 2017
LA Bodega Zagreb, Zagreb

Public relations, organizing events.

Public Relations Jul 2017 - Sep 2017
La Bodega Dubrovnik, Dubrovnik

Public relations, organisation of events

Sales Manager Sep 2018 - Jul 2019
Nebo Modna Kuća, Zagreb

Sales, advisory, planning the future projects.

Student Assistent - Microeconomics Sep 2018 - Present
Faculty of Economics and Business, Zagreb

As I was very interested and engaged with the course of Microeconomics. I was honored to get an opportunity to work with my professor and serve as help by holding a repetition lectures, correcting the exams and homeworks as well as helping with all the tasks during the semester.

Student Assistent - Principle of Economics Sep 2018 - Present
Faculty of Economics and Business, Zagreb

As I was very interested and engaged with the course of Principle of Economics. I was honored to get an opportunity to work with my professor and serve as help by holding a repetition lectures, correcting the exams and homeworks as well as helping with all the tasks during the semester.

EDUCATION AND QUALIFICATIONS

Bachelor of Economics and Business Sep 2015 - Sep 2020
Faculty of Economics and Business, Zagreb, Zagreb

Secondary school Sep 2011 - Sep 2015
Gymnasium, Zagreb

Conference Jan 2014 - Feb 2014
Harvard University, Boston

Harvard Model United Nations

REFERENCES

References available on request.

SKILLS

Microsoft Office ●●●●●

PUBLICATIONS