UNIVERSITY OF LJUBLJANA

SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

ALESSIA GIAMBARTOLOMEI

EUROPEAN MASTER IN TOURISM MANAGEMENT

FACULTY OF HUMANITIES, UNIVERSITY OF SOUTHERN DENMARK
SCHOOL OF ECONOMICS AND BUSINESS, UNIVERSITY OF LJUBLJANA
FACULTY OF TOURISM, UNIVERSITY OF GIRONA

MASTER'S THESIS

THE IMPACT OF THE GENERAL DATA PROTECTION REGULATION (GDPR) ON TARGETED MARKETING IN THE TOURISM SECTOR

Ljubljana, July, 2023 ALESSIA GIAMBARTOLOMEI

AUTHORSHIP STATEMENT

The undersigned **Alessia Giambartolomei**, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEBLU), author of this written final work of studies with the title "**The impact of the general data protection regulation (GDPR) on targeted marketing in the tourism sector**", prepared under supervision of Prof. William Gartner and co-supervision of Prof. Tanja Mihalic

DECLARE

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the SEBLU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEBLU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work

could have for my status at the SEBLU in accordance with the relevant SEBLU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEBLU Study Information System;

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, July 15th 2023 Author's signature: 

Abstract

Using a double approach (systematic review of the existing academic literature and qualitative web content analysis) this study investigates the impact of the General Data Protection Regulation (GDPR), enforced by the European Union in 2018, on targeted marketing in the tourism sector. 18 different documents published by different experts in the sector (law firms, consultancies, business associations, marketing agencies, IT companies and news) were analyzed and coded with the support of the software Nvivo to 2 macro-categories and 21 sub-categories identified from the results of the systematic literature review. Nonetheless, a directed approach was used, meaning that as the content analysis progressed, new categories were added to the existing ones as they emerged through thorough reading and examination of the data. The results show that the GDPR has had a considerable impact on targeted marketing in the tourism industry due to its data and information-intensive nature. Another important insight is the emphasis put on the global validity of the GDPR, that applies to all companies processing EU citizens' data world-wide. Finally, several sources considered the GDPR as an opportunity to enhance customer loyalty, trust, and satisfaction by demonstrating transparency and protecting personal data, aspects explained by the permission-based model and theory acceptance model (TAM).

KEYWORDS: *GDPR; targeted marketing in tourism; transparency; permission-based model; theory acceptance model; web content analysis*

Table of content

List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
BT	Behavioral Targeting
DS	Data Science
FE	Feature Engineering
GDPR	General Data Protection Regulation
GIF	Graphic Interchange Format
HTML	Hypertext Markup Language
IA	Intelligence Augmentation
ICT	Information and Communication Technologies
IoT	Internet of Things
ML	Machine Learning
NLP	Natural Language Processing
OTA	Online Travel Agency
PII	Personal Identifiable Information
PNR	Passenger Name Record
RFID	Radio-Frequency Identification
RQ	Research Question
SEO	Search Engine Optimization
TAM	Theory Acceptance Model
UGC	User-generated content
EEA	European Economic Area
DMC	Destination Management Company
DPO	Data Protection Officer
SCC	Standard Contractual Clause
BRC	Binding Corporate Rule
CRM	Customer Relationship Management

List of figures

List of tables

1. Introduction

The revolution in information and communication technologies (ICTs) has had a deep effect on the tourism industry (Xu et al., 2017). Over the last decades ICTs have been tightly linked with the tourism experience ad management of tourism offer (Xiang et al., 2015). The result is that the concepts of smart tourism (ST) or e-tourism have been coined to refer to the application of ICTs aiming to develop innovative tools and mechanisms to improve tourism (Sotiriadis et al., 2020). Gretzel et al. (2015) define ST as

> A tourism supported by integrated efforts at a destination to collect and aggregate/harness data derived from physical infrastructure, social connections, government sources and human bodies/minds in combination with the use of advanced technologies to transform that data into on-site experiences and business value-propositions with a clear focus on efficiency, sustainability and experience enrichment. (p.181)

In the e-tourism system, digitalization is used in all processes of the travel industry. It aims also to increase the efficiency of information sharing, exchange and transaction processing (Liberato et al., 2018). The travel and tourism sector is an information-intensive industry (Buhalis & Amaranggana, 2015; Herdin & Egger, 2007). Therefore, a huge quantity and typology of data are available. Since we as humans are tracked and traced in both the real and the virtual world, each individual produces vast amounts of data per day without explicitly knowing it (Egger, 2022a). Everything on the Internet is measurable. Indeed, most websites are now able to track user actions such as how often they visit, which specific pages they visit and how long, where they went after finishing the visit and which products they bought (Miedema, 2018). With specific regards to tourism and travel, internet connectivity has simplified the task of consulting maps and travel information on mobile phones, access to transport services and paying via ticketing applications (Tapsell et al., 2018). This massive amount of data has become available electronically and can be used with different purposes in the tourism industry (Godinho et al., 2015). Most recent trends of e-tourism research are focusing on tourism marketing (Hamid et al., 2018). Indeed, many organizations operating in a competitive market and offering efficient services, are starting to rely on the aforementioned data in order to predict short- and long-term consumer trends by refining customers' profiles and marketing efforts (Miedema, 2018; Tapsell et al., 2018). Smart marketing has considerably reformed the traditional marketing patterns. It has introduced new creative digital media and communication technologies, enhanced the ability to identify and cater to the needs of tourists. The result: tourist destinations can be advertised more efficiently (Hamid et al., 2018).

Therefore, online users' personal data have acquired more and more importance in the last years, especially with the development of behavioral targeting (BT), that according to Goldfarb

(2014) has been accelerated by many benefits for companies such as customization, personalization, and measurements. Algorithmic profiling or targeting is an automatized process of tracking, mining, and using personal information to predict digital users' preferences based on statistical information and evaluations (Blass, 2019). With the help of machine learning technologies and algorithms, advertisers use different targeting techniques, including contextual and behavioral targeting, to ensure accurate personalization, customization, and efficient contextualization (Goldfarb, 2014). The sophistication of targeting and profiling processes in personalized advertising has led to a shift from a segmenting system based mainly on demographics and locations to tracking users' data related to personal, family, relationship, financial, location, education, entertainment, and relaxation digital activities (Basarudin & Raji, 2022). These processes not only benefit advertisers but also users, who can greatly benefit from BT advertisement since it reduces both cost and time of processing, searching, and buying on digital platforms (Basarudin & Raji, 2022).

At the same time, it must be acknowledged that many ethical issues are linked to BT for marketing purposes, regarding aspects data validity and impact on privacy. Individuals still have a low awareness of the fact that certain data have been collected from them and that these data can be used by decision-makers to make choices that can impact their life. This knowledge asymmetry makes individuals vulnerable, with limited resources to exercise their rights and freedom (Egger, 2022a).

To mitigate the aforementioned issues, the European Union has enforced in 2018 the General Data Protection Regulation (GDPR), which imposes rules about transparency, explicit consent of personal data collection and mining, and prohibits the processing of many categories such as racial or ethnic origin, political opinions or religious or philosophical beliefs (Basarudin & Raji, 2022).

I found evidence, in many scientific papers, of the analysis of the consequences of the enforcement of the GDPR and overall privacy rules from the point of view of users and customers. These range from the new rules for which they must express they explicit consent before accessing to a website (Karaduman, 2017); the complexity of the legal/ technical language of the rules they must explicitly agree on, which makes them hard to understand (Acquisti et al. 2015; O'Connor, 2020; McDonald & Cranor, 2008; Milne et al., 2006); the proposal of mechanisms for a safe and privacy-enhanced data collection and storage mechanisms able to comply with GDPR (Drosatos et al., 2015; Tapsell et al., 2018); and the ability of the higher level of transparency related to data collection, storage and use to create trust among users, eventually contributing to enhance the image of the company (O'Connor, 2020; Millet, 2022; Theocharidis et al., 2020).

Nonetheless, very few papers (Millet, 2022) exposed the effects of the practical implementation of the GDPR from the side of the companies and enterprises, especially the ones operating in the tourism field. However, many tourism stakeholders operating in different phases of the tourism experience are using BT for marketing purposes, such as online travel agencies (OTAs) (Privacy Statement, 2022; Custom Targeting for Travel Advertisers, 2022), hotels (O'Connor,

2020), and Airlines (Millet, 2022). Thus, the topic is acquiring more and more relevance in the tourism sector. Therefore, the purpose of this research will be to answer the following research question: *To what extent is the in the EU General Data Protection Regulation impacting the targeted marketing in the tourism sector?*

The aim of this research is to explore the consequences of the GDPR on targeted marketing in the tourism field, with a particular focus on the offer side of stakeholders, and thereby to contribute to the limited tourism literature on the topic. In order to answer to the research question, two main sets of data are utilized; data generated from the content analysis of the secondary data available online, and secondary data from the analysis of the existing academic literature mostly from the field of computer science and tourism.

The thesis starts with chapter 2 focusing on the methodology, where I will introduce the scientific paradigm and the methods used in this study, that is based on a systematic approach to select the existing academic literature and a web content analysis of the secondary data available online and published by experts such as marketing agencies, IT companies, sectoral journals, consultancies, and law firms. This will be followed by chapter 3 where I expose the findings of the systematic review I conducted. In this chapter I introduce the concepts of artificial intelligence, machine learning, data science, tracking and profiling techniques for marketing purposes and related ethical issues (privacy and data validity), and an overall view of the GDPR. Chapter 4 will follow, where I will analyze the material gathered and discuss the related findings. More precisely, the analysis and findings will be presented considering 2 aspects: a) the presentation of the main points of the GDPR, and b) the changes that businesses had to implement due to its implementation. Finally, chapter 5 will close my research and address the RQ, also exposing its limitations and suggestion for further research.

2. Methodology

The following section is about the methodology that lays the groundwork for this research. I will explain which paradigm I decided to adopt as a framework and the related epistemological and ontological connotations, as well as the research method.

2.1 Paradigm

Guba (1990) defines the paradigm as a "set of beliefs that guides action taken in connection with a disciplined inquiry" (p. 17). This helps researchers to position themselves in a complex research landscape (Munar et al., 2016). Creswell (2003) assumes that the inquiry process starts by thinking and stating what a researcher wants to learn during his/her inquiry and the process of how to learn it. With this regard, paradigms can provide a framework for thinking about the reality to be studied, how to study it, and the tools that can be used to do so (Hershberg, 2014). The same concept is also confirmed by Usher in Tribe (2001), that defines it as "an exemplary

way of working that functions as a model for what and how to do a research *(study*), what problem to focus on and work on" (p. 443). Considering that this thesis intends to construct knowledge and understanding the research conducted for the purpose of this thesis is led by the constructivist paradigm (Guba, 1990), which can be categorized in the naturalistic field of research and is rooted in constructivism (Rubin & Rubin, 2011).

Social constructivism is a paradigm grounded in social sciences that aims to understand, interpret and create perspectives (Jennings, 2015). As Wilson and Hollinshead (2015) explain, despite many critics of paradigms linked to paradigm methods of social sciences in the past, they can still open to important humanistic perspectives, especially in tourism. Indeed, as Jennings (2015) states: "tourism is a socially constructed and determined phenomena that is constantly being reframed and reinterpreted and reconstructed" (p. 6), and a better understanding of it "needs to include the cultural context of the image creation and communication process" (Young, 1999, p. 374).

The most important reason why I decided to choose this paradigm are the ontological, epistemological ad methodological proponents (Crotty, 1998; Guba, 1990; Jennings, 2001, 2009, 2015) that, interconnected, identify the paradigm, and will be discussed in the following sections.

2.2 Ontology

The ontological proponent of every paradigm deals with reality, defining its nature and structure and trying to understand the perception of the world in which the author conducts her research (Crotty, 1998; Guba, 1990; Jennings, 2001). Being informed by the constructivist paradigm, I adopt a relativistic ontology. That means that I am approaching my research considering the 'world' as consisted of multiple realities which exist and manifest in the form of different constructions and meaning about a specific phenomenon. These different interpretations are grounded in the social sphere and vary according to each person holding them and seeing them through the unique lenses of their prior experiences and understandings. (Creswell, 2003; Guba, 1990; Rubin & Rubin, 2011). Therefore, I acknowledge the existence of multiple and even conflicting explanations of a phenomenon under research, and not only one theory or casual relation able to explain it (Jennings, 2001; Rubin & Rubin, 2011). As a consequence, with my research, I do not aim to come up with a universal or generalizable theory about the impact the GDPR on digital marketing for tourism purposes. On the contrary, my purpose is to uncover and analyze different perspectives and interpretations of my research topic, able to ultimately create an informed knowledge and get a deep understanding of it through the process of their analysis (Candy, 1989; Creswell, 2003; Goldkuhl, 2012; Hershberg, 2014; Kamii, 2016). Another aspect that must be considered is that, as argued by Guba (1990), the knowledge and understanding that result from the interaction between the researcher and the researched topic can hardly ever be certified as definitive and complete, no

matter how many individuals the researcher interacts with: their realities and interpretations about the investigated phenomenon can change over time.

2.3 Epistemology

The epistemological proponent focuses on the study of knowledge, what can be known, how do we know what we know, and what we know (Jennings, 2015; Goodson & Phillimore, 2004; Tribe, 2004; Guba & Lincoln, 1994). In other words, epistemology describes the nature of the relationship between the researcher and his/her research subjects, whose interaction becomes a source of creation of knowledge (Creswell, 2003; Guba, 1990; Jennings, 2001; Kamii, 2016; Thomas, 2010). Considering that, as also stated in the ontological paragraph, knowledge is constructed by the interaction of individuals in the social world (Thomas, 2010), by adopting a constructivist paradigm I also adopt a subjective epistemological stance (Creswell, 2003; Jennings, 2001, 2009, 2015; Thomas, 2010). Indeed, the interactive process involving the researcher and the research participants will inevitably be shaped by the personal beliefs and preconceptions of the stakeholders, which will clearly influence the inquiry (Creswell, 2003). However, as a constructivist, I do not aim to uncover a definitive and universal truth or an objective way to understand to what extent the existing EU measures to protect users' data are affecting e-marketing strategies. On the contrary, thanks to the multiple-realities-approach that I am adopting, this can help me to unveil different perspectives about my research topic, element necessary for the construction of informed knowledge about it (Guba, 1990). Moreover, it is interesting to highlight that as Guba (1990) remarks, due to the nature of constructivism, the distinction between ontology and epistemology is obsolete. This is because the interaction between the inquirer and the inquired, and the resulting outcome that combines two and more existing truths makes them consubstantial.

2.4 Research method

This research has been conducted using a qualitative methodology. Indeed, under the constructivist paradigm researchers most commonly choose qualitative methods to successfully gain in-depth knowledge from the study setting (Jennings, 2009). Because of their characteristics, systematic literature review and web content analysis have been selected as approaches to conduct the present research.

2.4.1 Systematic review of the literature

Pahlevan-Sharif et al., (2019) explain that systematic approaches have been developed in the medical field, to enhance the quality and transparency of literature reviews by minimizing biases and omissions. These systematic reviews, along with associated methods like meta-analyses, employ a replicable and transparent process that aims to reduce bias through thorough

literature searches, providing a clear record of the reviewers' decisions, procedures, and conclusions. They utilize a range of methods to conduct research on existing studies (Briner & Walshe, 2014). The key feature that distinguishes systematic reviews from other types of reviews is the methodical synthesis of findings, ensuring unbiased searches with increased efficiency and quality (Liberati et al., 2009; Mulrow, 1994). For instance, Briner and Walshe (2014) explain that

> Traditional or narrative literature reviews, while useful in many ways, have rather different and often unclear aims, do not adopt an explicit or systematic method, cherry-pick research, may adopt a stance, and include only evidence that tends to support that position. (p. 417)

The term 'systematic' refers to a research protocol that safeguards objectivity by providing explicit descriptions of the steps taken, including specific research questions, focus, strategy, and inclusion/exclusion criteria (Davies & Crombie, 1998; Tranfield et al., 2003). As a result, systematic reviews support evidence-based practices and are considered a fundamental scientific activity in numerous disciplines, encompassing both science and social science (Mulrow, 1994), including tourism research.

A systematic review is valued for its ability to comprehensively synthesize diverse interdisciplinary research, conducted in various settings and utilizing different research designs (Pickering & Byrne, 2014). Given that this review focuses on analyzing the impact of the GDPR on tourism digital marketing, which encompasses a broad range of multi- and interdisciplinary literature from fields such as tourism, computer science, social science, economics, accounting, and risk research, a systematic quantitative approach is considered the most suitable methodology.

Additionally, the chosen method is well-suited for examining emerging trends within disciplines, as highlighted by Pickering and Byrne (2014), making it highly suitable for the objectives of this paper especially in order to identify the key categories for the coding process, which constitutes the primary component of the empirical research and serves the purpose of this thesis.

2.4.2 Web content analysis

The focus of my thesis lies in examining the impact of the GDPR on tourism digital marketing. To effectively gather data on this subject, a comprehensive analysis of digital content (e.g., blogs, briefings, articles) published online by experts in the field, such as marketing agencies, IT companies, sectoral journals, consultancies, and law firms deemed the most efficient approach. This approach is justified by the fact that since the introduction of commercial Internet applications in the early 1990s, businesses worldwide have embraced websites as a

means to conduct their operations, promote their brand, disseminate information, and engage with various stakeholders (Buhalis, 2003). The tourism industry, in particular, has fully embraced the potential of the World Wide Web (Buhalis, 2003). Consequently, there has been an exponential growth in the volume of data generated within the context of tourism. Herring (2010) highlights that the vast number of web pages, their diverse formats and functions, coupled with the ease of collecting and analyzing content using automated tools, present limitless opportunities for research. The communicative nature of websites and web pages inherently lends itself to content analysis (Weare & Lin, 2000).

Bauer (2000) provided a comprehensive definition of content analysis, characterizing it as a systematic technique employed for the purpose of coding symbolic content, such as text and images, that is present in various modes of communication. When content analysis is applied specifically to the analysis of web-based content, it is commonly referred to as web content analysis (Herring, 2010).

Berelson (1952) explained that the content analysis initially served as a methodological approach encompassing both qualitative and quantitative dimensions. Primarily utilized as a quantitative research method in its early stages, the content analysis involved the systematic coding of textual data into explicit categories, which were subsequently subjected to statistical analysis (Hsieh & Shannon, 2012). This quantitative application of content analysis gained relevance, particularly in the field of mass communication, facilitating the enumeration and manipulation of textual elements through various statistical techniques (Weber, 1990). However, the present study deviates from a primary focus on quantitative content analysis and instead adopts a qualitative content analysis approach. Qualitative content analysis is characterized as a research method based on subjectively interpreting the content of text data by employing a systematic classification process to identify themes or patterns (Hsieh & Shannon, 2005). Qualitative content analysis is an empirical, controlled analysis of texts within their communicative context, following content analytic rules and step-by-step models while avoiding hasty quantification (Mayring, 2000). Thus, the objective of this research methodology is to uncover meanings and patterns inherent in specific content types, such as open-ended survey questions, interviews, print media, websites, images, and others. Qualitative content analysis typically involves purposeful selection of texts that provide insight into the research questions at hand (Zhang & Wildemith, 2009). The outcomes of this approach often manifest as descriptive accounts or typologies (Hsieh & Shannon, 2005). Ultimately, the overarching aim of content analysis is to enrich knowledge and understanding of the phenomenon under investigation (Downe-Wamboldt, 1992).

According to Zhang & Wildemuth (2009), qualitative content analysis involves a systematic process of condensing raw data into meaningful categories or themes through valid inference and interpretation. The reasoning employed in this process is primarily inductive, wherein themes and categories emerge from the data through the careful examination and comparison conducted by the researcher. However, it is important to note that deductive reasoning can also be incorporated into the analysis process (Patton, 2002).

Berg (2001) argues that qualitative research, including content analysis, can benefit from generating concepts or variables based on existing theory or previous studies, particularly during the initial stages of data analysis. This approach provides a theoretical foundation and allows for a more focused exploration of the data.

Hsieh and Shannon (2005) outline three distinct approaches to qualitative content analysis, based on the degree of involvement of inductive reasoning. The first approach is conventional qualitative content analysis, where coding categories are derived directly and inductively from the raw data. The second approach is directed content analysis, which starts with initial coding guided by theory or relevant research findings. Themes and categories emerge during the data analysis phase as the researchers immerse themselves in the data. The purpose of this approach is often to validate or extend existing conceptual frameworks or theories. The third approach is summative content analysis, which begins with the counting of words or manifest content and subsequently extends the analysis to uncover latent meanings and themes.

In the present study, a directed content analysis approach was employed. The process commenced with a systematic review of the relevant theories on digital marketing for tourism and of the main principles of the GDPR from where the preliminary categories have been identified. As the content analysis progressed, new categories were added to the existing ones as they emerged through thorough reading and examination of the data. This directed content analysis approach was chosen to ensure a focused exploration of the data in line with the research objectives.

2.4.2.1 NVivo

Specialized computer programs now enable efficient content analysis, offering advantages over manual coding in terms of speed and consistency. Heinrich (1996) observed that these programs enhance efficiency and inter-subjectivity by explicitly defining coding rules and facilitating their application to various texts. NVivo is a prominent program used for qualitative text analysis (QSR International, 2014). Computer-assisted text analysis tools provide interactive guidance for human coders, assisting in coding selection and organizing extensive text data for interpretation (Gretzel et al., 2007). Additionally, Popping (2000) highlighted that these programs support the development and representation of conceptual schemes through node networks.

NVivo offers functions such as text editing, note-taking, coding, text retrieval, and category manipulation. It includes a visual presentation module that allows the visualization of category relationships. The program keeps a record of the coding history to track the analysis progress. Researchers can query the data in NVivo to answer questions based on the database (QSR International, 2014). The coded in NVivo categorize data, representing concepts, individuals, places, or relevant characteristics. Codes act as markers for significant text portions, similar to post-it notes. They have names and descriptions for easy searching. Three code types exist: free-standing codes, hierarchical codes linked to parent categories and subcategories, and case

codes for grouping cases with individual nodes for each case. For the present research I only used hierarchical codes, divided in parent and children categories. In the next paragraph I will explain further the steps followed in the research process.

2.5 Research process

In this paragraph I will explain step by step the process I followed to perform my research. It will develop in two main parts: 1) explaining the review process applied to my systematic review. The results, that will be presented in the next chapters, have set a solid and scientifically based foundation to identify the categories that are essential for the 2) coding process and related content analysis.

2.5.1 The systematic review process

The literature review relies on a systematic review of academic manuscripts conducted through a Boolean search strategy. This strategy involved using different keywords, including the main keywords from the research question, as well as similar and related terms. This approach is consistent with the methodology employed by Hamid et al. (2018) in a similar study.

The Scopus database was selected for this study due to its extensive coverage of scientific journals and its global reach (Baas et al., 2020). Scopus is known for encompassing publications from scientific publishers worldwide, making it a comprehensive source of information (Baas et al., 2020). It is particularly recognized for its strong coverage of tourism journals compared to other platforms (Wijesinghe et al., 2019). In a study by Yang et al. (2017), Scopus was identified as one of the most powerful databases, offering advanced search capabilities that surpass other databases in terms of functionality and generating a larger number of results. Furthermore, Scopus is a multidisciplinary database that covers various fields such as Life Sciences, Social Sciences, Physical Sciences, and Health Sciences, containing over 69 million records (Niñerola et al., 2019). Considering the multidisciplinary nature of this research, Scopus was therefore deemed the most suitable database for the study.

Following the methodology adopted in previous reviews (Dorcic et al., 2019; Wijesinghe et al., 2019), the review was conducted in two phases. The initial phase (Phase I) involved a search based on four specific criteria:

1. The inclusion criteria involved the identification of specific keywords and their combinations to select relevant papers. The keywords used for inclusion are:
   TITLE-ABS-KEY ("data-centric" OR "data -centric" OR "data - centric" OR "data-driven" OR "data -driven" OR "data" OR "data science" OR "big data" OR "data mining" OR "data profiling" OR "algorithmic model" OR "algorithmic profiling" OR "behavioural analysis" OR "behavioral analysis" OR "user-generated content" OR "ugc" OR "ugd" OR "user-generated data" OR "cookies" OR "cookie" OR "predictive

analytics" OR "artificial intelligence" OR "ai" AND "marketing" OR "e-marketing" OR "digital marketing" OR "personalization" AND "privacy" OR "ethics" OR "ethical issue" OR "risk" OR "security" AND "tourism" OR "travel" OR "traveller" OR "traveler" OR "travellers" OR "travelers" OR "booking" OR "reservation" AND "regulation" OR "regulations" OR "law" OR "protection" OR "GDPR" OR "General Data Protection Regulation")

Initially, all published papers that contained these keywords in any part of their content, including the title, abstract, keywords, main body, and references, were selected.

2. The last search was conducted in January 2023 and no time limit was applied.
3. Only articles written in English language were considered.
4. Articles, Conference Reviews, Conference Paper, and Book chapters were considered.

Considering these four inclusion criteria, 47 documents were taken into consideration. In the second phase (Phase II) I read and screen all the articles. In this process the following exclusion criteria were applied:

1. Manuscripts whose subject are is not related to the ones relevant for this research. The selection was made automatically by the option “filter by subject area” available in Scopus. Therefore, only the following subject area were included: Computer Science, Business Management and Accounting, Engineering, Social Sciences, Environmental Science, Decision Science, Economics, Econometrics and Finance. This led to the exclusion of 8 documents.
2. I performed a second manual screening, where I carefully read all the remaining documents to exclude the ones not fitting into the research aim and / or not covering the topic of this thesis. This included a careful screening of the conference papers included in the original selection, in order to select the ones relevant to the research. After this step, only 16 manuscripts were left.

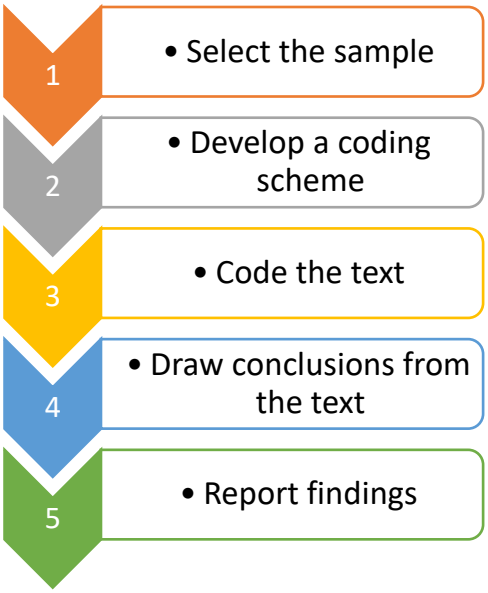
In order to gain additional insight from academia, I also included among the manuscripts used for the literature review relevant papers coming from the literature reviews of the original 16 documents selected in the systematic approach. Therefore, the final number of documents considered as foundation for the literature review is 126.

2.5.2 Qualitative content analysis

In order to perform the qualitative content analysis that constitutes the main empirical part of my research the steps proposed by Zhang and Wildemuth (2009) have been followed, with several adjustments, as summarized in figure 1.

Source: adapted from Zhang & Wildemuth (2009)

Figure 1 Research process: content analysis



Source: adapted from Zhang & Wildemuth (2009)

Step 1: Select the sample

The sample of the research included 18 documents published online between 2016 (year of adoption by the European Parliament of the GDPR) and 2020. This interval was considered in order to allow an analysis including the perceived impact before (2016-2017), in the moment of enforcement (2018) and after (2020) the GDPR. The documents have been analyzed between 30 April – 27 May 2023. The authors of the documents considered are blogs, interviews, briefings, articles, and reports published by companies operating in fields related to the topic of this research such as marketing agencies, IT companies, sectoral journals, consultancies, and law firms on the topic. Considering that the GDPR has a global validity, no restriction on the country of origin of the publishing company was applied. In order to ensure an analysis as much representative of the tourism sector as possible, not only documents discussing the impact on tourism overall (8 documents), but also ones with a specific focus on the 4 major tourism sub-sectors: Hotels (4 documents), Airlines (3 documents), cruise lines (2 documents) and Car rentals (1 document) (Sharma & Nicolau, 2020).

Step 2: Develop a coding scheme

In order to develop a coding scheme, it is essential also to define the unit of analysis considered to be coded. The unit of analysis in content analysis refers to the fundamental text unit that is categorized during the analysis process (Zhang & Wildemuth, 2009). While quantitative content analysis often employs physical linguistic units such as words, sentences, or paragraphs, qualitative content analysis, as described by Neuendorf (2002), typically adopts individual themes as the unit of analysis. These themes can be expressed through various textual components, including a single word, a phrase, a sentence, a paragraph, or even an entire document. When utilizing themes as coding units, researchers primarily focus on identifying expressions of ideas (Neuendorf, 2002). Consequently, a code may be assigned to

a text segment of any size, as long as it represents a coherent theme or category relevant to the research questions (Zhang & Wildemuth, 2009).

Categories and the coding scheme were partly derived from main point highlighted by the analysis of the systematic literature review. Two macro-categories were identified 1) a general presentation of the GDPR (that will be presented in paragraph 3.6 of the literature review), highlighting the essential points of the regulations including the importance of publicizing consumer rights, data controllers, data portability, data processors, data profiling, global validity, PII, the purpose of GDPR, the pseudonymization, the right of subject access, the right to be forgotten, special categories, and transparency. 2) changes that businesses had to implement after the enforcement of the GDPR. The principal sub-categories were identified in this case according to the steps of data processing for Machine Learning identified by Egger (2022b): data collection, data storage and data interpretation/use (presented in paragraph 3.2). More precisely, the elements taken into account were data collection, data storage and data interpretation/use changes before and after the GDPR, the role of third parties (paragraph 3.3.4), mentions of the direct impact on tourism and travel sector. Following these themes, parts of texts covering the selected topic were put into the appropriate code in NVivo. The coding scheme as developed in the research process is presented in figure 2.

Step 3: Code the text

The coding scheme created in step 2 served as the foundation for identifying pertinent themes in the content analysis. However, as the entire text was coded, it became apparent that additional themes needed to be incorporated as codes in NVivo. The predetermined categories initially established for analyzing both the presentation of the GDPR and the changes were expanded, and new codes were introduced (the principle of accountability for companies and penalties in case of violation of the GDPR, laws and regulations similar to GDPR, others). These codes were grouped and merged based on their content to enhance comprehension (refer to Figure 4 for visual representation).

Step 4: Draw conclusions from the text

In this step, the focus is on understanding and interpreting the identified themes or categories and their characteristics. It involves drawing inferences and reconstructing meanings from the data (Zhang & Wildemuth, 2009). The content associated with each individual category, as identified in the fourth step, was exported from NVivo and compiled into a Word document. During the first reading, repetitive content was eliminated, incorrectly coded content was repositioned to the appropriate section, and irrelevant information was removed. The second reading served as the foundation for establishing connections between categories, identifying patterns, and summarizing ideas.

Step 5: Report findings

Qualitative research is inherently interpretive, wherein interpretation reflects an individual's personal and theoretical comprehension of the phenomenon being studied (Zhang &

Wildemuth, 2009). Chapter 4 presents the outcomes of the content analysis and concludes with critical examination and discussion of the findings.

Figure 2 Coding scheme

Source: own elaboration

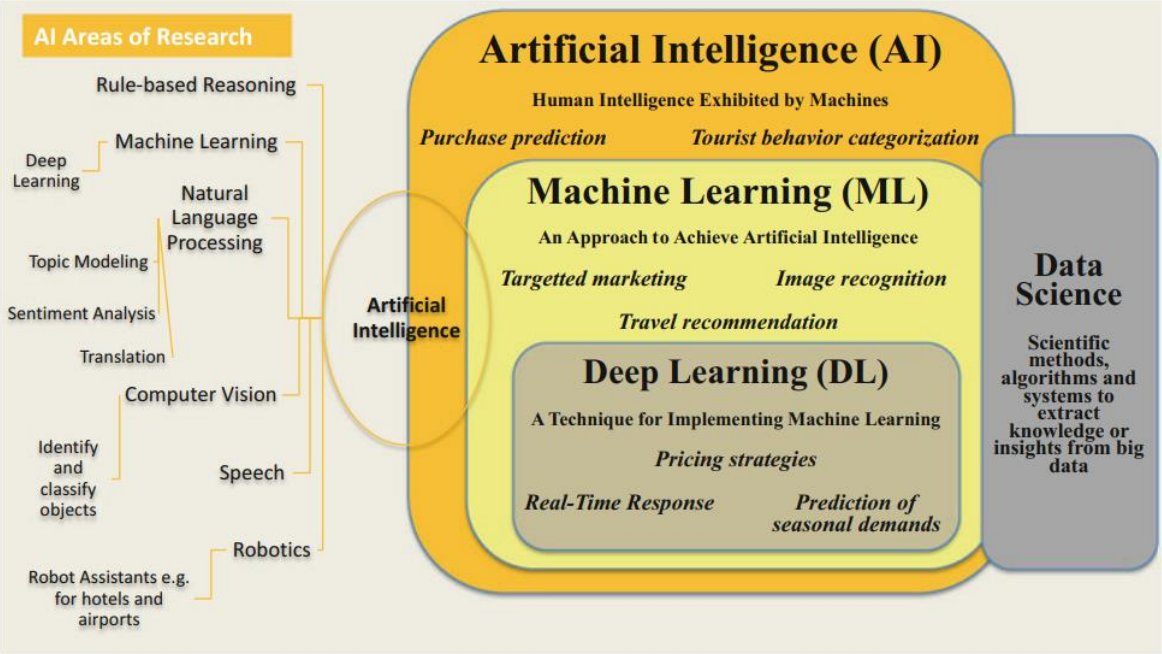
3. Literature review

In the following paragraphs, I will expose the findings of the systematic review that I conducted.

3.1 Artificial Intelligence: an umbrella concept

In order to understand data and their use in for marketing purposes in the tourism field it is important to define what artificial intelligence (AI) is. According to Egger (2022a), the rapid digital advance of our society with the development of powerful algorithms and the availability of vast amounts of data has created new concepts and analytical methods. The multitude of the latter and their respective names has caused confusion, making it difficult to distinguish one definition from the other. To reduce this confusion Egger (2022a) argues that AI can be considered as an umbrella term that includes more subject areas, including many more related to data such as data mining, big data, and smart data.

Figure 3 Artificial intelligence and its various research niches



Source: Egger (2022a)

The first definitions of AI date back to 1950s (McCarthy, 2007), all referring to the idea that AI enables computers to perform tasks and activities requiring human-level intelligence. Kok et al. (2009), go more in detail, defining AI as a field of study focusing on "the development of computers able to engage in human-like thought processes such as learning, reasoning, and self-correction" (p. 271). Examples include rule-based reasoning, machine learning (ML), natural language processing (NLP), computer vision, speech analytics, and robotics. The use of AI is growing exponentially and discussions about its application, performance, and impact are quotidian (Mich, 2022). Jordan (2019) goes beyond the concept of artificial intelligence and sums up the most recent theories in the term "Intelligence Augmentation" (IA), viewing

AI as a technology able to augment human intelligence and adopting a semi-automatic approach when AI is not able to completely replace human intelligence. AI can play a strategic role in fostering innovative solutions for businesses working in many sectors, including tourism.

Indeed, AI finds application in tourism because of its complex nature, where from one side the high number of stakeholders (service providers, intermediaries and customers requires a tuned system of communication and coordination (Boes et al,. 2015; Hofstaetter & Egger, 2009). On the other side, the goal to co-create memorable experiences for and with the customers, require a comprehensive understanding of the needs and wishes of the travelers to provide them with personalized recommendation in real-time (Xiang & Fesenmaier, 2017). With this regard, Gretzel et al. (2015) highlight that the combination of big data availability and the power of AI to enable new forms of value creation, sustainability, and well-being through technology has led to a new type of tourism known as "smart tourism". The same authors define three levels where AI finds application in smart tourism, all heavily relying on data: a) intelligent technologies able to understand specific situations and react accordingly based on data continuously recorded by sensors that track and trace our behavior. This poses the basis of any smart destination (Worden et al., 2003); b) AI that enhance smart experiences by personalizing and adapting experiences to guests' preferences. Again, this is possible thanks to the collection of context-sensitive and real-time data; c) the highly dynamic tourism ecosystem where data are collected, cleaned, processed, and shared between the stakeholders in order to be processed via AI (Egger, 2022a). Moreover, McKinsey (2018) insists that AI can be substantially relevant to economic sectors driven by sales and marketing, such as tourism. With this regard, Mich (2022) not only confirms this theory, but also highlight the lack of limit in terms of types of data sources and AI technologies that can usefully investigate and improve both the supply and demand side of tourism, including not only marketing strategies but also overtourism management and recovery from crisis periods.

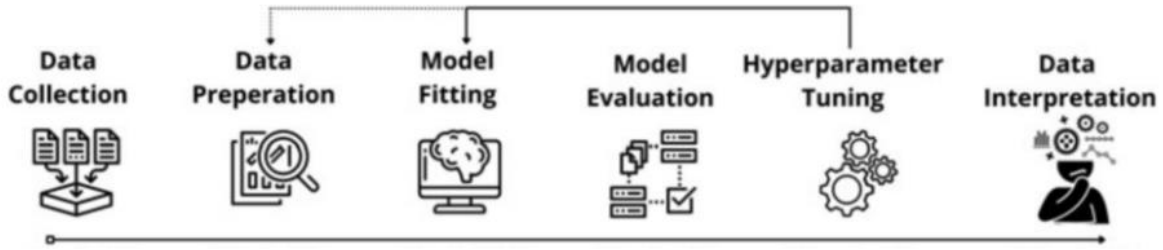
McStay (2020) highlights a specific sphere of AI whose application has started to be used in many sectors, including travel and tourism: emotional AI. Emotional AI refers to technologies that use AI techniques and affective computing to learn and interact with human emotional life in order to regulate and optimize the emotional experiences of spaces, namely travel infrastructures. The data used in emotional AI are words, images, facial expressions, gestures, voices, gaze direction, body temperature, heart rate, respiration, and electrical properties of skin. Despite the minimum efforts needed to process these data to see, read, feel, and learn about human emotional life, and the high human-tech interaction in the digital context, McStay (2020) argues that there is a lack of interest in understanding how human emotions factor into the latter.

3.2 Machine learning

As shown in fig.1 Egger (2022a) argues that ML is an important research niche of AI. According to Althbiti and Ma (2020), ML provides the central infrastructure for AI, and more and more often it lays the basis for data science (DS) projects, linking statistics and computer science to create and develop algorithms. The aim to achieve is high predictive performance and generalizability (Jordan & Mitchell, 2015). According to Mich (2022), ML and AI are terms that are often used as synonymous, because not only ML is considered one of the most relevant technological developments to shape our times (Jamal et al., 2018), but also because many of the recent advances in AI are due to ML (Mich, 2022).

There have been many attempts to define what ML is. According to Mich (2022), the first author who used the term ML is Arthur Samuel in 1959. He described it as "a field of study that gives computers the ability to learn without being explicitly programmed" (p. 86). The same Mich (2022) explains that the goal of ML is "to realize computational models that make computers learn "what to do" instead of having to tell them "how to do" it" (p. 86). Moreover, she insists on the different approach taken by ML compared to traditional stored programs. Stored programs register instructions on how to solve a problem in a memory and execute them on the input data to produce specific output data. Therefore, these programs work based on the translation of algorithms that use programming languages. On the contrary, ML is based on the concept of making computers learn from examples. Thus, computers do not need step-by-step explanations about how to solve a problem but learn autonomously which is the right output based on a given input. The definition of Akerkar (2019) is in line with the ones aforementioned, and additionally introduces the importance of data for ML defining it as "computational methods using the experience to improve the performance to make accurate predictions. [...] It is the study of algorithms that learn from examples and experiences instead of hardcoded rules" (p. 19). Indeed, the term "experience" refers to existing databases and their properties (the training data) which are used to learn and create a model based on patterns identified in the data (Mohri et al., 2018). These allow to better describe the data, increase performance, or make the most accurate possible prediction (Jamal et al., 2018) and personalized recommendation, which are the basis of marketing (Mich, 2022).

Figure 4 the ML process



Source: Egger (2022b)

The fig. 2 shows the typical ML process with all its stages. Although the steps can vary depending on the chosen algorithm, usually it starts with data collection of the training data and ends with the interpretation of the data (Mich, 2022).

Awad and Khanna (2015) highlight the importance, in the process of data collection, of selecting only specific subset of data or *features* and avoid using all the available data in order to solve the problem efficiently. These data can be either in structured or unstructured form and must be prepared before being inserted into an algorithm (Egger et al., 2022a). This means that they must be organized in a uniform format and incorrect or missing data must be removed. The features selected acquire strategic importance, because a model can only be as good as the data selected to be trained on (Sanchez, 2003). With this regard Duboue (2022) introduces the concept of Feature Engineering (FE), which is a human-centric process that involves both humans and computers working together to solve a problem using data. FE involves modifying the data based on human intuitions and experience to make problems easier to solve by computer. Differently from pure ML which focuses on solving problems using solely data, FE allows little human intervention to select and process features. The following step is to train the algorithm. For this purpose, three approaches are applicable:

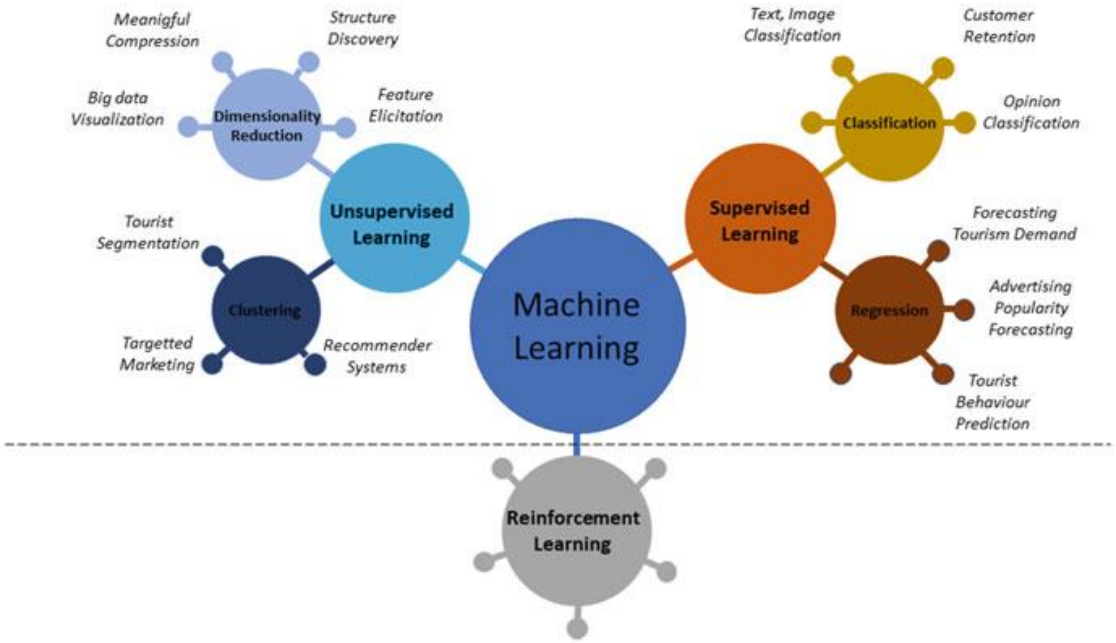
a. *Unsupervised learning*: This type of algorithm aims to identify common elements, creating patterns and recognizing useful structures and patterns starting from the input data without labeling them. (Mich, 2020, 2022). This means that unsupervised algorithms consider a set of predictors and analyze the relationship between them in order to either identify groups that show similar characteristics (*clustering*), or group features together according to the ones that best represent the data to make them more recognizable, reducing the quantity of data within a range that still remain reasonable (*dimensionality reduction)* (Egger, 2022b; Ozdemir, 2016; Provost & Fawcett, 2013; Bernstein & Kuleshov, 2014). In this case, the goal is not to make a prediction, but to learn something from the data (Arefieva et al., 2021). As shown in fig. 3 these methods find many applications in the tourism sector, especially *clustering*, used to typologize tourists' behavior and group homogeneously photos, reviews and destinations (Egger, 2022b), and used in targeted advertising. According to Provost and Fawcett (2013), one benefit of unsupervised learning algorithms is that there is no need to label the data, a fact that makes easier to find data material. On the other side, Egger (2022b) argues the predictive power of this kind of algorithm is lost, while Dy and Broadley explain that, due to the lack of a response variable, it is hard to evaluate the performance of the model without a subjective human judgement (Ozdemir, 2016).

b. *Supervised learning*: differently from unsupervised ones, these algorithms need data to be split into training data (to train the algorithm) and testing data (used to measure the performance of the algorithm). Therefore the target variable of the training data has a label, or a detailed description that can be used to train the algorithm (Egger, 2022b). from the labeled training data a model function is developed by trying to generalize the relationship between the input and output data (Awad & Khanna, 2015). In this case,

Jamal et al. (2018), highlight the importance of choosing the right features. Indeed, having a high number of them can confuse the learning algorithm. Considering that the aim is to train the algorithm so well that it is able to predict correct as accurately as possible class labels for a new, unseen dataset (Awad & Khanna, 2015). Therefore, the quality of a model and its performance highly depend on the training data and their labeling (Egger, 2022b). Supervised learning algorithms find two different applications depending on whether the target variable is continuous or discrete (Egger, 2022b). In case of continuous variable, the ML application used is *classification*. Classification is one of the most used ML applications whose goal is to divide similar data into different classes. Depending on which of the many classification algorithms have been used, different rules can be applied (Cleve & Lämmel, 2020). Arabie et al. (1996), argue that classification could appear similar to the clustering methods discussed in the previous paragraph, but the difference is the process of labeling that determines the structure of the groups, while clustering aims to reveal hidden feature similarities that will be the foundation of the segmentation in groups. In case of discrete target variable, the common ML application is *regression*. Shalev-Shwartz and Ben-David (2014) define regression as a set of statistical tools that describes the average relationship between numerical attributes by modeling the relationship between explanatory variables and a target variable. Fig. 3 shows that supervised learning algorithms find many applications in the tourism sector, especially in forecasting tourism demand and advertising popularity forecasting. Overall, one of the benefits of supervised learning it that they allow to measure the quality accurately (Provost & Fawcett, 2013), while Egger (2022b) argues that unsupervised learning algorithms are especially suitable for prediction purposes. At the same time, the author acknowledges that one of the issues related to this kind of algorithms is the risk of unavailability of the required labels in the dataset which, thus, need to be created first. Saeed et al. (2019) argue that this problem can be partially solved with unsupervised approaches, but to ensure data quality it is also need human labeling, which can become complex, expensive, and time-consuming.

c. *Reinforcement learning*: it is a form of automated, goal-oriented learning and decision making (Akerkar, 2019). Reinforcement learning do not focus on the learning algorithm itself, but on the characteristics of the learning problem. Thus, any kind of algorithm can be chosen to solve the problem as long as it is suitable for solving that specific problem (Skilton & Hovsepian, 2018). The system is given a task and, based on positive or negative feedback that aims to maximize a numerical reward value, it is supposed to learn and evolve (Ngyuen & Zeigermann, 2021). As argued by Egger (2022b), there are not many examples of application of reinforcement learning in tourism. One of these is provided by the study of Lu et al. (2021) who investigated the hesitation of airline passengers to choose a connecting airport when they have to choose among a large number of online sales options via many different channels.

Figure 5 Machine learning paradigms and application examples



Source: Egger (2022b)

The step that follows is model evaluation. As previously discussed, only supervised learning algorithms can be evaluated in terms of effectiveness and performance, an element that represents one of the major benefits of this typology (Egger, 2022b). According to Feurer and Hutter (2019) each ML system contains hyperparameters with settings that represent hints for optimizing the data processing. By changing these hyperparameters one can affect the algorithm's performance. This process of hyperparameters tuning, is very sensitive and require a deep knowledge of the effect of such change. Therefore, in order to identify the best setting tuning able to create the most efficient model an iterative process of the steps already mentioned takes place. Finally, the validated model is applied to an actual task and the results are interpreted and put in the original specific subject context (Egger, 2022b).

An important aspect that emerges from the analysis of the literature on ML is the essential role played by data. Mich (2022) highlights that despite the origin of ML dates back to 1950s, its huge potential has started to be understood only recently, when the amount of data has become impressive. With this regard, Kelm et al. (2020) highlight that ML is often mentioned in the discussion related to Big Data, and that the influence of ML in directly proportional to the amount of data available, while Mich (2022) supports the idea that it is the variety of sensors, input, sources, storage devices, the Internet, and networking systems which have contributed to the creation of large sets of data and the inability of traditional techniques for data processing to deal with them to boost the ML research. For this reason, in the next paragraph I will focus on the analysis of data and their sources.

3.3 Data, sources of data, and Data Science

As shown in the previous paragraphs, data and big data play a key role in AI and ML and their many applications. Hamid et al. (2021) define big data as "the massive amounts of data with various and high velocities measured in higher units than gigabytes". Li et al. (2018), in their literature review of big data in tourism research, argue that there is lack of uniform definition. Nonetheless, one of the most popular definition is the *3V: Volume, Variety* and *Velocity* originally coined by Laney (2001), and used by Li et al. (2018) and Mich (2022):

- *Volume:* a variety of sources and tools, such as business transactions, Internet of Things (IoT) and smart devices, industrial equipment, mobile devices, radio-frequency identification (RFID) readers, social networks, wireless sensor networks, etc. allows the collection of a high volume of data. This high number of available datasets goes along with easier and cheaper storage options and platforms.
- *Variety*: the data streams come in different formats. Few of them consist of structured data which can be managed with traditional databases. On the contrary, most of them come in form of unstructured data, such as texts, pictures, videos, sensor data, etc. This can be explained by the fact that usually a given input (for instance webpages), includes combinations of different formats, and by the lack of shared standards for many new data types.
- *Velocity*: data are produced and used at high speed. Widely used applications, such as chatbots and personal assistants, as well as cyber-physical systems which embed hardware and software like in driverless cars or robotics, produce fast data streams that need to be processed almost in real-time.

Mich (2022), also adds two more characteristics of big data, *Veracity* and *Variability:*

- *Veracity*: data have to pass a quality check. This acquires vital importance because, as also mentioned in the previous paragraphs, problems in the input data affect also the output. At the same time, the author acknowledges that the high volume and variety of data make their quality difficult to check.
- *Variability*: data flows can change often and vary greatly. The author explains that this is especially true in the tourism field where the change in tourist flows which depend on events, weather conditions or seasonality. These drastic changes need to be managed fast and adequately in order to efficiently deal with data peak in order to extract valuable knowledge.

On the other side, Gantz and Reinsel (2011) and Marr (2015) to the original 3 Vs add *Value:*

- *Value:* the data need to be interpreted, also considering their statistical or hypothetical nature. More precisely, Marr (2015) insist on the fact that the access to big data is useless, unless it cannot be translated into value.

At the same time, Mich (2022), explains that according to this definition, Value is not an intrinsic characteristic of big data, but a challenge to be faced by the companies which want to

use the data to extract value to ease the decision-making process. Therefore, it should not be included in the *Vs.*

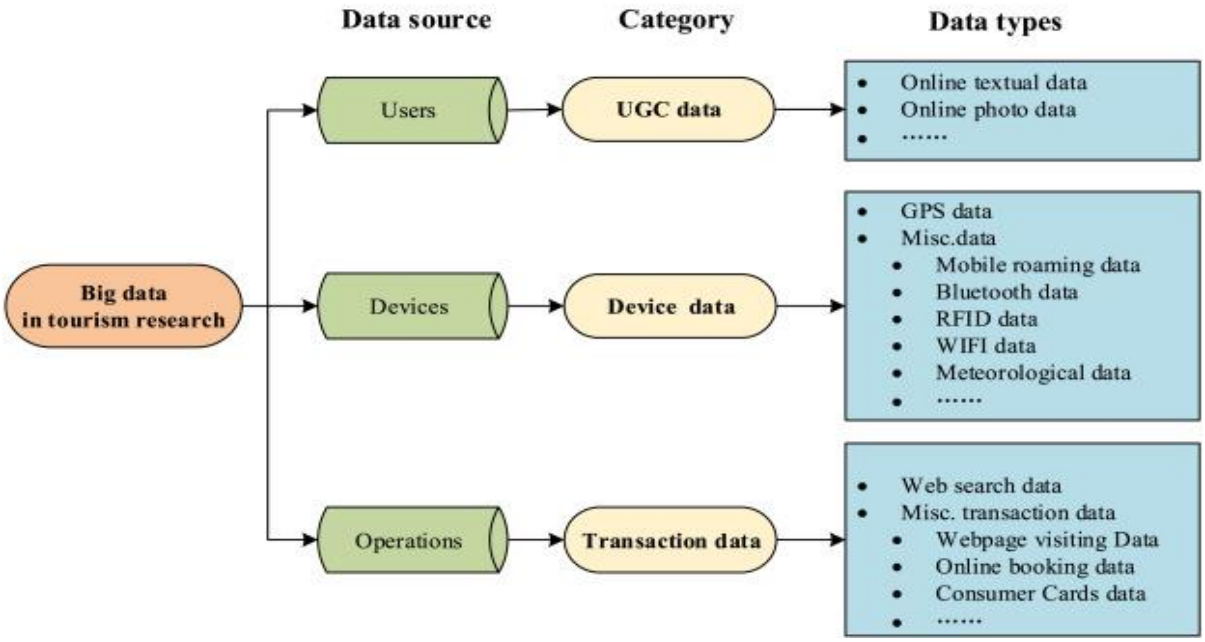
Many authors (Theocharidis et al., 2020; Tapsell et al., 2018) underline the role of consumers and users in the process of providing information that constitute the big data. Tapsell et al. (2018), introduces the concept of user-data:

> A set of data that represents and is associated with the identity, activities and service-offerings associated with a unique individual. Whether in an identifiable (non-anonymised) or non-identifiable (anonymised) form - collected/process/shared by an organisation (or its partners) to either provide/tailor a service to the respective individual. (p. 1380)

They also add that user-data are potentially able to capture a range of personal traits, activities and additional information to identify a unique individual. Therefore user-data also contain Personal Identifiable Information (PII): name, address, phone number, and e-mail address, etc. Theocharidis et al. (2020) explain that users accept and use technology in many different contexts following the Theory Acceptance Model (TAM) coined by Davis et al. (1989) and based on perceived ease of use and perceived usefulness. The TAM explains the determinants of conscious behaviors towards the use of computers and, according to Theocharidis et al. (2020) it can be applied and extended to any new technology introduced by using domain-specific theories. In the case of the authors' study analyzing the factors affecting the intention to book a hotel room through social media the TAM combines tourism and social media. On the contrary, in order to explain the relationship between users, digital world and their data, Sotidiaris et al. (2020) introduce the concept of *prosumption.* This term defines the "production by consumers" (p. 10) that has been fastened by the technological breakthrough, and especially by social media, which have given to people the possibility to register and share opinions, i.e. data, about products they used. The authors explain that this is valid for smart tourists too, who are actively creating and sharing information, trying to influence and persuade other users. Indeed, Femenia-Serra et al. (2019) define the smart tourist as "the tourist who, by being open to sharing his or her data and making use of smart technologies, interacts dynamically with other stakeholders, co-creating in this way an enhanced and personalised smart experience" (p. 14).

Mich (2022) discusses that being tourism a complex and multidimensional sector, and due to the high number of people, processes and activities involved, tourism data are often big data. In order to better expose the results of my systematic review, I will categorize the selected researches following the classification presented by Li et al. (2018) (fig. 4).

Figure 6 Analytical framework of the literature review proposed by Li et al.



Source: Li et al. (2018)

3.3.1 User-generated content (UGC)

Li et al. (2018) argue that the Internet has boosted the development of use of social media, which have become a spacious platform to share UGC data. With regards to tourism research, UGC include two types of data: 1) online textual data (such as product reviews posted on social media or blogs); 2) online photo data. This kind of data is considered in the studies of many researchers, with different purposes. Sotidiaris et al. (2020), in their paper investigating the influence of social networking sites and UGC on tourists to adopt a responsible and sustainable behavior, find out that this assumption is verified in the first two stages of the tourist experience: before consumption and during consumption. Secondly, they discuss the implications of this finding from a marketing management perspective, asserting the importance of an appropriate mix of social media to influence the behavior of visitors and guests, especially through the real-time interaction and the new types of technology-mediated social connections. Finally, they highlight that in social media content users are co-producers, co-marketers and co-consumers of tourism experiences. Theocharidis et al. (2020), in their study investigation the factors affecting customers' intentions to book a hotel room through social medias, highlight that UGC like photos, likes, recommendations, and positive comments can positively influence users and introduce a permission-based acceptance construct. This model is based on the trust, of users, to accept to receive marketing offers and announcements from a company. The authors insist that this model is effective only if the companies understand what push users to give their permission. Li et al. (2018) argue, specifically for online photo data, that they contain many useful metadata, i.e., information embedded in photos. For the sphere of tourism, the researchers declare useful: 1) user-related information

(ID photo and user ID), temporal information (dates of when the photos were taken and uploaded), geographical information latitude and longitude) and textual information (title, description, and tag). Abdul-Ghani (2020) and Fowler (2019) argue that in addition to posts and personal photo from UGC social media can collect name, date of birth, address, location, email, IP address, devices used, lists of contacts, web browsing records, interests and opinions, and online purchases of their users, elements that will be discussed in the next paragraphs.

3.3.2 Device data

In this category are included GPS data, mobile roaming data, Bluetooth data, RFID data, Wifi data and meteorological data. Their collection has been boosted by the development of IoT and sensors able to track tourists' movements and therefore providing a huge quantity of data in both structured and unstructured form to be used in travel and tourism decision making (Li et al., 2018; Shoval & Ahas, 2016). Chorus (2012) and Ceder and Jiang (2020) highlight that smartphone apps allow to save and learn from passengers' choices in order to predict passengers' personalized mobility paths. Indeed, the authors argue that smartphones, through their apps, are able to acquire real time journey planning information, ready to be re-used by public transport users. Passengers do not simply obtain a path from a specific origin to a given destination, but a path that also takes into account their preferences at the desired time of travel. McStay (2020) introduces another data that can be collected by devices, especially IoT that includes data about users' words, voices samples, images, facial expressions, gaze direction and biofeedback data (body temperature, heart rate, etc.). He explains that these data are the basis of a branch of AI, Emotional AI, that finds application in tourism to optimize the emotionality of travel infrastructure and restaurants, but also to provide emotional reactivity feedback to ease targeting of advertising to marketers.

3.3.3 Transaction data

Li et al. (2018) explain that transaction data record tourism-related operations like activities and events in the tourism market. For instance, in this typology, the researchers include web searching, online booking and purchasing, webpage visiting, etc. These data find applications that go from promoting tourism prediction, search engine optimization (SEO), tourism marketing, and tourism behavior understanding. In addition to the findings of Abdul-Ghani (2020) and Fowler (2019) discussed in the paragraph 3.3.1., O'Connor (2020) analyses the potentiality of loyalty or reward programs promoted by many hotel chains to gather transactional data about customer spending and behavior during their stay. This process is made possible by customer relationship management technologies which, combined with demographic data voluntarily given by customers, allow a better understanding of guests' preferences and behavior that can be used not only to improve the customer service but also marketing efforts. Drosatos et al. (2015) apply the use of transactional data as basis for the correct functioning of recommendation systems in tourism. They insist on the role of users to

enter their personal information, such as interests, previous visits, and ratings into a system. Then, they propose an approach for automatically creating a point-of-interest-user profile to be used as input data for a contextual suggestion model. On the contrary, other authors focus on the transactional data obtained from the reservation step. For instance, Gilbert (2008), discuss that, the increasing number of hotel reservations and travel planning online allow, through monitoring technologies such as cookies, web bugs and beacons, to obtain detailed information about people's browsing habits. These can be used to collect various information regarding spending patterns or users' interests that can be used to target marketing efforts more precisely. Millet (2022) discusses the role of the passenger name record (PNR) as a source of transactional data for airlines. Indeed, PNR contains: 1) identification information like name, email, phone number, postal address, passport number, frequent flyer number and the tier level for marketing; 2) customer behavior information that gives the history of changes, cancelations and disruptions of past bookings; 3) data about the specificity of the journey such as point of sale, flight numbers, fares and taxes. The author explains that starting from this information it is possible to push very simple marketing messages and offers adjusted to passengers' needs.

3.3.4 Aggregated data

In addition to the classification proposed by Li et al. (2018), there are various authors that consider the use of the aforementioned data in a combined and aggregated form. For instance, Drosatos et al. (2015) created Phytia, a model able to offer personalized contextual suggestion for tourism that is based on both data about the user's location (device data) and web searches (transaction data). Han et al. (2016), discuss about a Cyber-Physical social network, that indicates the combination of data coming from social network (UGC) and data collected from sensors present in mobile phones, tablets and IoT (device data). According to the authors, this system has a potential to achieve influence maximization, that is "targeting at maximizing the propagation of one product or viewpoint through social networks by word-of-mouth effect with limited initial resources" (p. 284), which has a significant role in targeted advertisements. Abdul-Ghani (2020) and discusses the concept of a "data ecosystem" that include consumers, big digital businesses (like Google and Facebook), marketers and advertising agencies, app developers, and "third party" data brokers, which are companies that collect and sell information about users that can be used by ad networks, other data buyers or audience buying platforms (Chester, 2012).

3.3.5 Targeted Marketing data

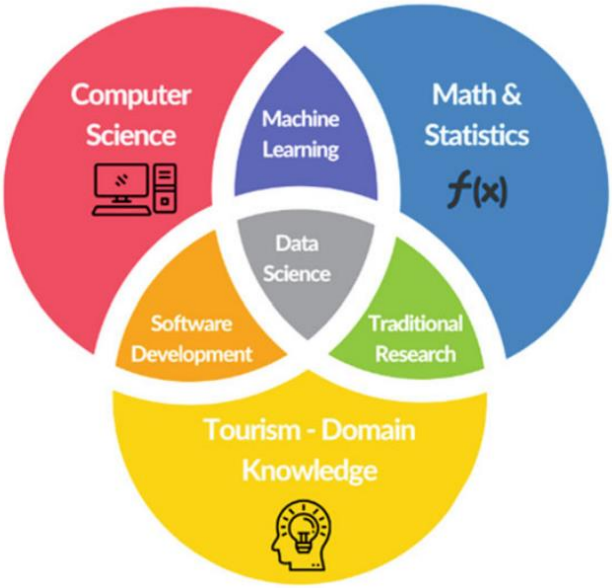
Shaw et al. (2021), in their study about the application of transportation data sources discuss another type of data: targeted marketing (TM) data. TM data are defined as "large databases that house hundreds to thousands of individual- and household-level variables that data providers (often these are credit reporting firms) either directly collect, purchase, or develop" (p. 151). TM data are collected with the purpose of being re-sold to companies that use specific

variables to target their specific audience in marketing campaigns. They constitute holistic datasets with variables that include socioeconomic, demographic, attitudinal and behavioral information. Moreover, the researchers explain that TM databases differ from big data because they come from both active and passive data sources, while big data are usually derived only from passive sources. Among the active data sources, Shaw et al. (2021) list surveys conducted by consumer research firms and individuals' answers to online quizzes, games, and questionnaires. At the same time, Erevelles et al. (2016) highlight that among the benefits of TM data, there are volume, variety, and velocity, that are shared with big data, and the fact that they are low-cost. As per the related challenges, many authors (LaValle et al., 2010; Lovelace et al., 2016) highlight that the veracity of TM data must be assessed before their use, usually by comparing them against benchmarks from other sources, like the Census or other surveys. The same is valid for their value. With this regard, Shaw et al. (2021) suggest integrating TM data into applications and observing the model predictions or performance with new data sources. The same authors underline also a second set of challenges: the development of the variables. Indeed, TM providers often use personal algorithms to develop and model many variables. This makes it difficult to evaluate the quality of TM variables and to interpret them. In addition, they add that modeled TM variables may be tweaked over time, therefore not allowing a consistent definition of the variables. In addition, the authors explain that TM data have enormous potential in the tourism field, especially to supplement travel demand modeling and forecasting efforts. This is possible thanks to their ability to identify and market to individuals who are more likely to be more receptive to a particular product, due to their variable richness.

3.3.6. Data Science

Mich (2022) and Murtagh and Devlin (2018), discuss that the development of AI and ML, and the application of big data in this field has contributed to a new research and application area: Data Science (DS). DS is a field whose purpose is to interpret available data and uncover hidden relationships and patterns. Egger (2022a) explains that being DS a significant aspect of AI, it can be defined as "a comprehensive set of methods, algorithms, systems that are applied to various sectors of an interdisciplinary field" (p. xxxiv). It combines computer science, mathematics, statistics, and domain-specific knowledge, machine learning, data visualization, communication, and presentation skills to get insights from sets of structured, semi-structured, and unstructured data (George et al. 2016; Mich, 2022; Chauhan & Sood, 2021). DS helps to explain and understand present processes and phenomena to predict the future ones (Egger, 2022a). DS is used in vast areas of modern society, where it has brought significant changes towards a more data-based decision-making (Steinberg & Aronovich, 2020; Power 2016; Provost & Fawcett, 2013).

Figure 7 DS - an interdisciplinary field



Source: Egger (2022a)

Egger and Yu (2022), highlight that due to the intrinsic high interdisciplinarity of tourism, that welcomes the research methods of each of the disciplines involved, DS has already started to be used to further understand and solve tourism issues. Egger (2022a) lists few examples of DS application in tourism that includes: route optimization, forecasting and predictive analysis, opinion mining and sentiment analysis, monitoring systems, image analysis, social network analysis and personalization and recommendation, especially useful in targeted marketing. Nevertheless, the same author acknowledges that the fullest potential of DS has not yet been understood by tourism scholars due to the vastness and unstructured nature of the data collected which contributes to creating confusion and uncertainty.

3.4 Tracking and profiling techniques for marketing purposes

In the previous paragraphs, the concept of smart tourism has been introduced. Hamid et al. (2021), in their paper "*How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management*" highlight that the most recent trends in smart tourism focus on the categories of tourism recommendation systems and tourism marketing. Rongrong (2017), define tourism marketing as the discipline aiming to attract visitors to a specific location that includes conference centers, hotels, cities, states, and consumer attraction. Smart marketing strategies help to grow a destination's share of the tourism market. Chester (2012) explains that in 1990s, when commercial Internet started to emerge, there was a great international enthusiasm for its democratic potential, but the ultimate impact of interactive advertising that it enabled was yet to be understood. In order to better structure this part, I will divide this paragraph according to some of the topics discussed by

Chester (2012) in his study “Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the “Big Data” Era”.

3.4.1 Behavioral targeting

Nowadays, digital marketers have at their disposal a powerful data-collecting, profiling, and targeting apparatus, based on techniques incorporating the latest development in fields like semantics, AI, social network analysis, neuroscience, and data mining (Chester, 2012). Moreover, Moore et al. (2015) and Sinclair (2016) argue that the most recent development of these technologies made the process of individual targeting and personalization almost instantaneous. It is possible to request an online ad as soon as an individual starts loading a webpage, and the ad can target that specific individual even before the page has completely loaded. This is possible through a complex algorithm-driven automatic analytic system, based on different mechanisms. Sometimes there is a *cookie* present on a user’s computer that allows to identify the user’s need according to recently visited pages, in other times the ad is determined by analytical profiling techniques based on data collected from massive databases that aggregate the history of click-streams of individuals and complex algorithms enabling a real time bidding for an ad space (Varnali, 2019; Försch & de Haan 2018; Qin, et al. 2017; Sinclair 2016). This process of displaying digital ads to consumers based on the history of their online behavior is called online behavioral advertising, or behavioral targeting (BT) (Aalberts et al., 2016; Boerman et al., 2017; Chen & Stallaert, 2014; Ham, 2017). Scholars do not have a common definition for BT. Ham (2017, p.632) defines it as “a type of digital advertising targeting method that tracks and compiles individual Internet users’ online behavioral data, such as what websites they visit, how long they stay there, and what they do (e.g., shopping; searching; surfing)”. For Smit et al. (2014, p.15) it is a process " adjusting advertisements to previous online surfing behavior”, while for McDonald and Cranor (2010, p. 2) is “the practice of collecting data about an individual’s online activities for use in selecting which advertisement to display”. Boerman et al. (2017, p. 364) extracted two common points from the different existing definitions that are “(1) the monitoring or tracking of consumers’ online behavior and (2) use of the collected data to individually target ads”. According to Aguirre et al. (2015), Chen and Stallaert (2014), and Alreck and Settle (2007), the use of BT is increasing due to its effectiveness and relatively low cost.

The system of BT count on three different actors: consumers, advertisers and publishers. A publisher offers an ad slot for sale. The slot can be filled with a) a traditional advertisement system, where the same ad is displayed to all customers; or b) BT, where targeting algorithms, using previous online behavioral data (searches, content views, clicks, purchases and other interactions) and user profile data (demographic information, segment or cluster information, e.g., if the user is an auto enthusiast), compute a fit score between a consumer and an ad. The publisher’s slot is usually sold by weighted unit-price auctions, where advertisers place cost-per-clicks bids and the winner is chosen depending on both the fit score and cost-per-click bids (Chen & Stallaert, 2014; Palos-Sanchez et al., 2019; Qin et al., 2017). The Federal Trade

Commission (FTC, 2014) also add as a player in BT data brokers, which are companies that collect consumers' personal information to resell or share them with others. This means that data brokers own huge datasets with information about individual consumers and are able to process them to create detailed consumer profiles. Most of the platforms offering targeting ads integrate one or more data brokerage companies (Ash, 2016). Nonetheless, Varnali (2019) argues that consumers are usually unaware of their data being collected and processed by data broken companies, opening an issue about lack of transparency that will be discussed more in detail in the next paragraphs.

Furthermore, Chester (2012) explains that not only web pages and sites we access compile users' data, but they are designed to ensure that visitor leave their digital fingerprints behind. Indeed, there are technologies that help to optimize the way users interact with web pages and help manage the users' online journey, in order to push the users to do whatever digital marketers want them to do. The design of a site includes an optimized placement for banners, buttons, videos, and various navigation tools that structure the user's journey. On top of that, the author adds that a range of neuromarketing-based applications, aiming to convince users to accept online marketing-based content, has added new forms of subliminal persuasion to data collection. Another set of tools is based on *eye-tracking* i.e., the evaluation of how the visitors' eyes move across a web page, in order to ensure they interact with the site. The goal is to "maximize conversions" i.e., lead a consumer to purchase (Hernández-Méndez & Muñoz-Leiva, 2015).

As highlighted by Chester (2012), one example of BT activities is provided by the company Experian. The author discusses that, despite its principal interest in credit reports, Experian has started to operate in online commerce. Experian Interactive collects information on consumers interested in loans or buying products through subsidiaries. These data are later reused to predict the location of consumers and will be inserted in an algorithm able to generate a score that considers information about a user's lifestyle and credit with the company.

3.4.2 Cookies

O'Connor (2004, p.9) defines cookies as "small files deposited by a web server onto a client's computer that facilitates the tracking of their activities", while Sipior et al. (2011) defines them as small text files that the web server inserts onto a visitor's hard drive. Furthermore, the authors explain that the original goal of cookies was to enhance user's interaction in the Web by enabling him/her to resume interaction in a site exactly where it was left off in the previous visit. Randall (1997) introduces the term "Magic Cookies" to indicate the new feature developed by the browser developer Netscape, that allows cookies to track and store users' online activities. The mechanism is based on the exchange of small strings of text that contain information about the user's interaction in a particular web site, from the site to the user's hard drive and back when the same user return to visit the site (Sipior et al. 2011). Therefore, a

cookie allows multiple page views to be linked together and, if combined with log-in or transactional data, to identify users (Charters, 2002).

Mayer-Schönberger (1998) and O'Connor (2020) discuss about third-party cookies, which are used by companies other than the one of the visited web site. Third-party cookies are considered to be even more powerful than simple ones, as once placed in one web site, they can be recognized by other websites on the same tracking network. As a consequence, users can be tracked and targeted not only on a single web site, but across many different sections on the Internet (Bruce et al., 2017).

The recent improvements in terms of precision and quality of data collected have led to a growth of BT and behavioral retargeting, which is the process of pursuing with ads consumers who have left a website without purchasing on web sites that they visit subsequently. (Gilbert, 2008; Kumar & Gupta, 2016). This approach has been proved to be highly effective: consumers receive a content that match their interests, while sellers can use a one-to-one approach that targets each consumer with customized marketing messages at minimal cost (Aguirre et al., 2015, O'Connor, 2007).

In the tourism sector, a study by O'Connor (2020) shows that the most of online travel agencies use these techniques based on collection of users' data to tailor their sales messages.

3.4.3 Beacons

Beacons, also called Web beacons, Web bugs, pixel tags, or clear gifs are single-pixel graphic interchange format (GIF) image tags in hypertext markup language (HTML) documents place in e-mails or Web sites with the purpose to monitor users' behavior (Martin et al., 2003). Beacons send information to a remote computer when the Web page is visualized. One major difference from cookies is that while for the latter users have the option to accept of decline them, beacons are placed in the HTML as small, often imperceptible graphic files, that generally have the same color as the background, therefore resulting transparent to the users' eyes. Their presence can be detected only if the user can find an IMG tag in the HTML, represented as <img> (Sipior et al., 2011). Angwin (2010) explains that beacons can track different aspects of a user interaction on a Web page, such as typed entries or mouse movements. And Sipior et al. (2011) add that they may be used to retrieve files saved in a hard drive, transmit images from the computer's camera, record conversation through the microphone, report site traffic, count unique visitors, personalize online user's experience and even audit and report on advertising. They can also collect click-stream data or take this information form a previously set cookie, tracking whether a precise Web page has been visualized or even whether an e-mail has been opened or forwarded. On top of simple cookies, they can use third-party cookies to track users over multiple different Web pages (Schoen, 2009). Martin et al. (2003) add that beacons can collect and transmit to third-party companies users' demographic data, banner ads seen before making a purchase, search terms, the browser

type and configuration, and PII. Additionally, online purchases can be linked to a specific ad that the user has seen before the purchase. (Sipior et al., 2011).

3.4.4. Mobile devices

Many of the techniques previously discussed for the Internet are also used in the mobile market. As discussed in the previous paragraphs mobile devices are able to collect device and transactional data, transforming them into portable behavioral tracking tools. By combining tracking tools with location information digital marketers can help to move a consumer to a particular store or shop, collect information about what has been bought by encouraging buyers to use QR or barcodes to check the price of a product and even use specific digital applications to encourage them to repeat this process (Chester, 2012). The Mobile Marketing Association (2011), has described an array of techniques aiming to foster a user to engage in data-driven advertising and other services such as: 1) Click-to-call: where a click on an ad starts a call to the content provider or advertiser; 2) Click-to-SMS: where a click initiates an SMS to send a keyword linking to a shortcode to get more information about a product or service; 3) Click-to-locate: click leads to a map enabled by location-based services where a user can find the closest service provider; 4) Click-to-buy: click leads to a Web page where the user can buy using a sort of mobile payment (i.e. credit card or operator bill); 5) Click-to-storyboard: click leads to another add that can provide additional actions.

3.4.5. Social Media

Hays et al. (2013) argue that social media has become a powerful and popular marketing channel thanks to its ability to include five properties: "1) information provision, 2) collaboration, 3) communication, 4) interactivity and finally its potential to conduct transactions" (p. 213). John (2020), adds that social media is widely used not only in advertisements but also in public relations, branding and customer management. The same author argue that an efficient use of this tool can bring many benefits such as automation, personalization, enhanced interaction and improved customer loyalty.

Batrinca and Treleaven (2015) argue that one of the popular uses of social media in marketing is for sentiment analysis. Sentiment analysis aims to determine a text writer or speaker's attitude about the topic or the overall contextual polarity of a document (Mejova, 2009). Sentiment analysis can be done by applying different approaches that range from sentiment polarity, degree of positivity, subjectivity detection, syntax, term presence, etc. (Pang & Lee, 2008). As Batrinca and Treleaven (2015), sentiment analysis has become a relevant field of research and business activity thanks to the emergence of web-based application programming interfaces (APIs) provided by many social media including Facebook and Twitter. Egger et al. (2022a) explain that, to obtain data from APIs, a request with description of the service the data are needed for must be submitted, and companies providing the APIs will decide whether to

approve or not the access to data. At the same Batrinca and Treleaven (2015) argue that although social media data is accessible though APIs, most of the sources (including Facebook and Google) due to the commercial value of the data, are making it more and more difficult to access to raw data.

Another important tool are *Social network media platforms*, which provide data mining and analytics on Twitter, Facebook and a range of other sources. These platforms target companies interested in monitoring sentiment about their brands or products. Few examples are Attensity and Salesforce Marketing Cloud and they measure demographics, influential topics and sentiments through text analytics on online consumer conversations. (Batrinca & Treleaven, 2015). All this considered, in the tourism sector the study by Theocharidis et al. (2020) discuss that hotel brands should tailor the information posted in their social media pages according to travelers' personal preferences or past behaviors. Additionally, Liang et al. (2009) explain that social media with fined targeting features allow hotels to identify ideal customers and to provide them with personalized advertising content and services. In return, users will perceive these customized offers as more useful and will be more likely to make a purchase.

This mechanism is also at the foundation of what Theocharidis et al. (2020) describe as *Permission-based* marketing model which refers to a strategy where consumers opt to receive marketing offers or ads from a brand after their explicit consent (Im & Ha, 2013). Although initially it was conceived for e-mails, this kind of marketing has evolved, and it is used in a range of social media platforms (Karimi et al., 2017). Permission-based marketing ensures that, since they voluntarily gave permission to receive marketing offers, customers not only pay more attention to the marketing content but also are inspired to be engaged in a long-term campaign (Theocharidis et al., 2020; Millet, 2022).

3.5 Security, privacy, and ethical issues

The literature previously analyzed has highlighted the relevance of data in many fields, including tourism and tourism marketing. Nevertheless, one must not forget that whenever data is used to make predictions or support decision-making processes, the decisions taken can significantly affect people in many ways (Barocas & Selbst, 2016). Although the fast development of the application of data analysis, the topic of ethical issues related to data usage has started to receive attention only in the recent times (Saltz & Dewar, 2019; Egger et al., 2022b). Ethics is defined as the moral evaluation of choices or the science of morals (LaFollette, 2007; Ulrich, 2008). As Saltz and Dewar (2019, p. 198) explain that its most foundational interpretation "refers to the perception of something being good or right". In the background of data science ethical principles tell what is the proper, right and socially appropriate approach to conducting research with the data (Egger et al., 2022b). Mingers and Walsham (2010) divide ethics into three categories: a) meta-ethics, describing ethical theories; b) normative ethics, describing the process of reaching moral conclusions; and c) applied ethics, focusing on their practical application in specific context. The case of applied ethics in

data science focuses on algorithm decisions, aiming to provide guidance on how to interpret data and, consequently, what actions to implement (Mittelstadt et al., 2016). Algorithms and their parameters are designed by developers with specific results and outcomes in mind. This means that they will necessarily prioritize specific values over others (Krarmer et al., 2011; Nakamura, 2013). At the same time, Mittelstadt et al. (2016), explain that operating with accepted parameters does not ensure that an ethically acceptable behavior has been adopted. Additionally, Barocas and Selbst (2016) and Egger et al. (2022b) argue that good algorithms depend on good quality data, and they have the same limitations as all data-processing methods. For instance, the authors showed how algorithms can be biased when it comes to minorities, because the data collected and used in the algorithm reflect the prejudices and biases that exist at societal level, making it difficult to identify the problem's original source.

In order to show the results of the review of the existing literature for the ethical issues in data science, I used the framework provided by Egger et al. (2022b), considering two major aspects: data validity and privacy.

3.5.1 Data validity

Lever et al. (2016) argue that errors in data analysis may have consequences concerning both a lack of validity and ethically problematic results with serious aftermaths. This happens because data are the foundation of different options for options and decisions to be taken. Therefore, any error during data collection, input, or processing steps can lead to results in the wrong direction. According to Egger et al. (2022b), problem in data validity can occur in three cases:

1. *Lack of validity of the data itself*: when data are used for calculations, like in algorithmic models, it is essential that the data quality is sufficient (Gao et al,. 2016) Nonetheless, the same authors explain that quality checks on the origin of these data is not always possible. With this regard, a study by John (2020) in digital marketing strategies for educational tourism show that source credibility, along with argument quality, are essential to the online user's acceptance and further adoption of online information. Source credibility is the extent to which an information source is perceived trustworthy, believable and considered an effective peripheral cue in the information adoption process (Petty et al., 2017). According to Stephenson et al. (2001) source credibility strengthens the message elaboration, thus influencing audience engagement. Indeed, source credibility have a positive effect on the attractiveness, charisma likeability and credibility of the message. Therefore, when an audience finds the quality of the message transmitted insufficient to form an attitude about a product or a service, they may rely on external aspect of the message such as the trustworthiness, reliability, integrity, and character of the source of the communication (John, 2020). Source credibility is also directly influencing both cognitive and behavioral engagement in social media marketing. Indeed, users are more likely to engage with online contents

whose source can be identified and verified, as shown by John (2020), who demonstrated that students looking for educational trips are very engaged in interacting with educational providers by commenting, expressing opinions and sharing contents with other communities. Another issue is related to the choice of representative samples when researchers can work only with data that is available to them (Seely-Gant & Frehill, 2015). Dindar and Yaman (2018), explain that in many studies Twitter posts are used as sample data to analyze public opinion due to their relative easiness to be obtained compared to Facebook or Instagram data. At the same time, the authors point out that Twitter users are typically younger, more technically savvy and usually have an higher income. Thus, this opens to a debate whether these data are actually representative of the entire population and question the validity of the results of the analysis of such data (Seely-Gant & Frehill, 2015; Egger et al., 2022);

2. *Lack of data processing:* this happens due to unfavorable decisions (Kwon et al., 2014), especially because many different approaches to deal with missing values exist (Ngiam & Khor, 2019). In addition, Pratama et al. (2016) and Egger et al. (2022) highlight that there are specific procedures that react very sensitively to the presence of missing values while other algorithms can handle them very well.

3. *Lack of validity of created models:* this aspect acquires particular importance in the moment of selection of features for machine learning (Cai et al., 2018). In order to train a model, it is essential to consider which features are available and which ones should be ultimately selected to train a model (Egger et al, 2022b). Additionally, the quality of training data plays a key role. If a model is trained using data from the past but the actual data used in the already trained model are different from the original training data, the result could be compromised (Raschka, 2018). Lever et al. (2016) highlight argue that the selection of the appropriate for a certain purpose must also reflect an ethical perspective. The authors insist that a researcher should consider whether a model has been evaluated, or whether there is a possibility of over- or underfitting the data, and even whether the hyperparameters (see 3.2) have been tuned. Nonetheless, choosing a model that fits the best the goal of the research but has not been accurately validated is unethical (Raschka, 2018). Additionally, a researcher must take into account other elements which can lead to serious consequences such as mistakes – from carelessness or ignorance, or subjective decisions (Kitchin, 2014). Thus, it is essential to manage data and models in a responsible and aware way.

3.5.2 Privacy

Privacy rights are a framework for the huge amount of unstructured data available and the protection of personal data and sensitive information (Schermer, 2013). Human beings have a need for privacy. They need that certain conversations, information, and thoughts are kept

private. Despite this privacy need being universal to all cultures, what needs to remain private (e.g., medical details, financial information, sexual matters) can change from one culture to the other (Abdul-Ghani, 2020; Acquisti et al., 2015).

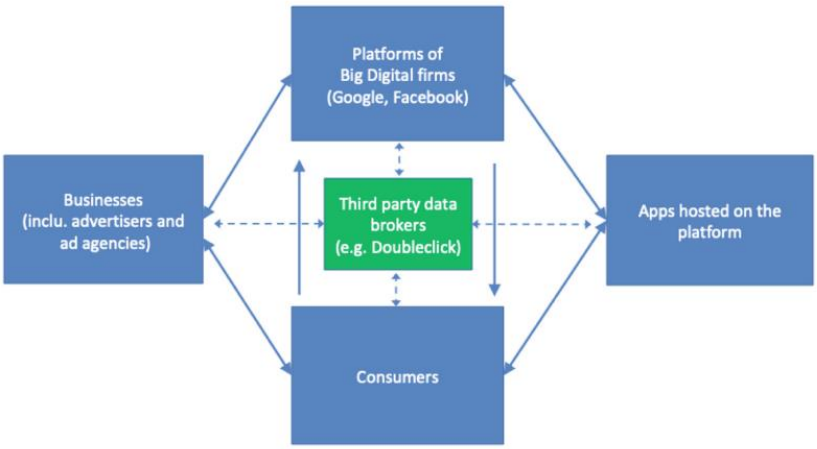
The literature available discuss many different points related to privacy and privacy rights. Smith et al. (2011) discuss three different point of views related to privacy: 1) privacy is a right, comparable to the right to be left alone; 2) privacy is a state, such as the state of limited access to people and their personal information; and 3) privacy is a form of control, like the possibility to control which individual information is shared with others. In addition to 1) privacy as a right, McStay (2020) and Floridi (2014) insist on the point that privacy is not only an individual right, but a group right, to be consider as collected good held by a group as a group and not by its members. Furthermore, the authors discuss also other privacy-related values such as respect for individuality, selfhood, autonomy, control, and dignity stating that privacy does include them but is not synonymous with them. However, with particular reference to Big data, McStay (2020) highlights that profiling is rarely based on targeting single individuals. Mateosian (2013) and Saltz and Dewar (2019) explain that privacy issues are related to how much control users have of their personal data, the ownership of data rights and the circumstances under which data can be accessed. Additionally, Birrer (2005) states that privacy issues also consider the process of data collection, the use of the collected data, the conclusions taken at the end of the data analysis, as well as the consequences of the conclusions.

Abdul-Ghani (2020) explains that the internet age has brought many issues related to the maintenance of consumer privacy. Through social media, users reveal information about their private life including photos, partners, children, and friends (UGC). While using online search engine consumers reveal their online web search behavior, their interests and wishes (transaction data, device data). Nonetheless, according to the author, the most ubiquitous issues are related to social and institutional privacy. "Social privacy refers to the revealing or sequestering of personal information from social display" (Abdul-Ghani, 2020, p. 41). According to Alkire et al. (2019) and Bartsch and Dienlin (2016), consumers are aware of issues related to social privacy and adopt measures to protect it by, for instance, adjusting the privacy settings on social media apps. On the other hand, "Institutional privacy, [...], is the issue of what private data on individual consumers is being collected online, stored and shared by commercial organisations such as Google and Facebook" (Abdul-Ghani, 2020, p. 41). In this case the author highlights a further need to establish the level of awareness among consumers of issues related to institutional privacy (Abdul-Ghani, 2020). With this regard, Martin and Murphy (2017) introduce the concept of "*privacy paradox"*, which refers to the tendency of consumers to have high concerns about online privacy while keeping sharing personal information online. This can be potentially explained by a low awareness of the users of how many information they are disclosing online (Martin & Murphy, 2017).

3.5.2.1 Surveillance capitalism

With reference to online institution privacy literacy, Abdul-Ghani (2020) introduces the concept of *Surveillance capitalism*. Surveillance capitalism refers to the business model of many digital businesses, including Google and Facebook, based on providing free-of-charge valuable services to consumers that includes online information source and access to social media platforms in exchange for personal information. Their revenue does not come from consumers but from advertisers, who pay the digital businesses to use consumers' personal information to drive effective advertising on their behalf. Thus, personal data on consumers is a valuable asset for digital businesses. Abdul-Ghani (2020) explains also that there are other actors in the ecosystem built around the process of sharing of personal consumer data, such as the app developers and third-party data brokers. (fig. 7). Digital marketers and marketing academics describe this gathering of personal data as a commercial exchange between a consumer and a firm where personal data are exchanged for services or improved services. (Abdul-Ghani, 2020). Martin and Murphy (2017) explain that supporters of consumer data harvesting argue that this process is legitim and can be theoretically referenced to social contract theory and social exchange theory. Nonetheless, if consumers are not aware of a) what information has been collected about them, b) what that information is used for and c) with whom the information has been shared with, the foundation of the social contract or social exchange does not exist. Therefore, if consumers' institutional privacy literacy is low, the data harvesting process risks becoming an unbalanced exchange based on information asymmetry conditions (Abdul-Ghani, 2020). Chester (2012) claims that nowadays the debate about online privacy has mainly focused on how data are collected without users' informed consent and less about the ultimate use of the collected information.

Figure 8 Commercial ecosystem developed around consumer data sharing



Source: Abdul-Ghani (2020)

Other supporters of data harvesting practices argue that this practice is relatively harmless, and it is only about offering to consumers more relevant advertisements (Chester, 2012). Others claim that this process is benefitting consumers, who can take more informed decisions than in the past (Chester, 2012). Chester (2012), discusses also about PII, explaining that many digital companies claim, that BT advertising based on data harvesting claim is innocuous, anonymous, and based on non-PII data collected from users. At the same time, the researcher also acknowledges that many of online advertisement businesses speak of their uses of unique personal data in their targeting equations.

3.5.2.2 Privacy enhancement

As previously discussed, there are several methods that allow consumers to protect their social privacy (Abdul-Ghani, 2020). Nonetheless, different tools to protect their information form institutional intrusion exist such as encryption, virtual private networks, incognito mode on search engines, and adjusting privacy settings on the browser of mobile phones (Abdul-Ghani, 2020). From the side of the users, Tapsell et al. (2018) argue that it would be possible to develop mechanisms that give consumers the control, traceability, management and auditing of their data, i.e., empowering them to control and observe how organization use their data. Tapsell et al. (2018) propose a model able to "enable a open, fair and transparent consumer data practices – building trust in the digital technologies and empowering the general public (users)" (p. 1381). The researchers also highlight the technological, commercial, and regulatory feasibility of the framework they propose, the "Consumer Oriented Data Control & Auditability" (CODCA), and its ability to balance between the user's privacy and the companies' goal to offer a valuable targeted and efficient service to their customers (Tapsell et al., 2018). Another example of user-centric concept is Pythia, which is an "architecture in which all user data are stored at the user-side. The contextual suggestion algorithms are also executed at the user-side. This choice offers the user a high level of control over the system" (Drosatos et al., 2015, p. 823). Therefore, Phytia ensures that any personal data is disclosed to any party, assuring strong privacy guarantees for users (Drosatos et al., 2015).

As per the organizational side, Palmatier and Martin (2019), suggest many ways in which businesses can behave ethically regarding the collection, storage, use and sharing of PII. For instance, they can adopt in-house consumer data privacy policy, minimize to the strict necessary the amount of data collected, regularly audit privacy practices, implement systems to protect stored data from eventual data breaches and even empower users through practices of data transparency and control. As Abdul-Ghani (2020) highlights, one mechanism designed to help users in this regard is the stated privacy policy on websites and apps. In order to get full access to the functionalities of a website or an app, consumers are asked to read and agree to the provider's privacy policy. Nevertheless, empirical research shows that consumers rarely read these policies before agreeing, and in case they do so, they find their language very difficult to understand (Acquisti et al. 2015; O'Connor, 2020).

At the same time, it must be acknowledged that there are privacy regulations that and organization has to abide by. The way these regulations are enforced depends on the single national governments or delegated data compliance or auditing authorities. These may range from audits to verify that companies are following regulations to simply committing to it (Tapsell et al., 2018; Shah, 2015). Nonetheless, the users do not have any tool to check autonomously or on-demand whether a specific organization is respecting the data privacy regulation. The only possible way is through local authorities (Tapsell et al., 2018). Examples of privacy regulations include General Data Protection Regulation (GDPR) enforced by the European Union (European Parliament, 2016).

3.6. The General Data Protection Regulation (GDPR)

The European Union General Data Protection Regulation (GDPR), is a regulation enforced from May 2018 that, through 99 articles, establishes rules aiming to make more transparent the processes related to personal data use, ensure higher level of security, and allow customers to control the data they agree or disagree to share (European Parliament, 2016; Millet, 2022; Abdul-Ghani, 2020).

GDPR is applicable to two stakeholders: data processors, i.e., the legal person or other body which processes the users' personal data (for instance Google or Facebook), and controllers, i.e., the person or body which defines the purposes of the data processing (Basarudin & Raji, 2022). GDPR includes in the definition of personal data any information related to a person's private life *stricto sensu*, referring not only to an individual's activity and routine, like financial data, social behavior, movements, preferences, and characteristics. Additionally, also different pieces of information which collected together can lead to the identification of a particular person (PII), are considered personal data (European Parliament, 2016; Basarudin & Raji, 2022). This aspect is also discussed in Article 30, dealing also with other identifiers, including cookies:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them. (European Parliament, 2016)

Furthermore, Article 9(1) discuss the processing of special categories of personal data:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. (European Parliament, 2016)

However, Article 9(2) lists ten cases where this rule does not apply, including when "the data subject as given explicit consent to the processing of those personal data for one or more specified purposes [...]" (European Parliament, 2016).

With specific regard to BT advertising, Article 4(4) of GDPR defines the concept of data profiling:

> 'Profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements. (European Parliament, 2016)

In the case of data profiling, Basarudin and Raji (2022) and Batista et al. (2020) discuss the importance for advertisers and data aggregators to tread carefully and avoid what Campbell et al. (2020) and Kietzmann et al. (2020) call hyper-personalization, i.e., the point where targeted users are irritated or psychologically stressed.

Another case when GDPR is not applicable is when personal data have undergone the process of "Pseudonymisation" (Article 26):

> 'Pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (European Parliament, 2016, Article 4(5))

Another point relevant to BT marketing is listed in Article 5 of GDPR, explaining the need to insert in personalized advertisements connotations and provisions in order to inform the users of the need to collect personal data, how it is collected and, how it will be used. All this is mandatory to comply with the rule on lawful and fairness (Article 5(1)(a)) and data minimization (Article (5)(1)(c)) (European Parliament, 2016; Basarudin and Raji, 2022).

Other salient points can be resumed in:

- *Right of subject access:* users have the right to ask to any organization to provide them with all the data it has collected about that specific individual. The organization has to serve this right within reasonable time and costless. Nonetheless, governments have the power to deny this right on national security grounds;
- *Right to be forgotten:* if a user requests it, an organization has to remove all relevant data related to that specific user from its storage;
- *Consumer rights must be publicized:* an organization must consider consumer rights when it conducts a 'Privacy Impact Assessment' for each new process involving users' data processing;
- *Data portability:* user data stored in an organization can be moved to any other organization upon the user's requests;
- *Global validity:* GDPR apply globally to any company collecting European citizens' data (European Parliament, 2016; Tapsell et al., 2018; Bu-Pasha, 2017; Millet, 2022; O'Connor, 2020).

3.6.1 The implementation of GDPR

The concrete implementation of GDPR translates into comprehensive and clear information that personal data will be collected, a clear explanation of the aims for which the collected data will be used that users will receive before accessing a website (Karaduman, 2017). Additionally, at the first access to a website, users must be informed and warned about any cookie use and privacy disclosure, ideally through a pop-up, slide over or similar technologies. Users must give their express consent and acknowledge being informed about all this before any data collection can begin (O'Connor, 2020).

Despite this mechanism, empirical studies show that overall privacy statements tend to be difficult to find (Wirtz et al., 2007), and are seldom read by users or written in legal or technical language that is hard to understand (Acquisti et al. 2015; O'Connor, 2020; McDonald & Cranor, 2008; Milne et al., 2006). O'Connor (2020) also highlights the users' paucity of knowledge about the functioning of online monitoring techniques. Research by Jensen et al. (2005) shows that despite the most of respondents claims to understand cookies (90.3%) only 15.5% demonstrated basic knowledge of them when further probed.

Considering the point of view of controllers, in the case of the tourism field the systematic review considers a study by Millet (2022) analyzing the effect of GDPR on airlines. The author

argues that enforcing GDPR is difficult for airlines, since the data collected, especially from the PNR, are many and have multiple sources, and thus data storage requires extra care. Though, she insists that the real challenge is to make users aware of the processing of personal data in an understandable and transparent way (Millet, 2022). Many authors argue that trust plays a key role in transparent communication with customers and the confidence that the data they are allowing to be collected will bring them true benefits (O'Connor, 2020; Millet, 2022; Theocharidis et al., 2020).

4. Findings and discussion

As also presented in the literature review, the enforcement of the GDPR has had a deep impact on all industries whose business relies considerably on data. The travel and tourism sector is a key example not only because every business works with users' and suppliers' information (AltexSoft, 2019; Neubauer; 2021), but also because of the role that data have in the creation of value for the destinations where companies operate in (Del Vecchio et al., 2018). This research presents the point of view of experts such as marketing agencies, IT companies, sectoral journals, consultancies, and law firms on the topic. More precisely, 18 different blogs, interviews, briefings, articles, and reports aiming to present and explain the changes due to the implementation of GDPR to companies operating in the tourism industry were analyzed.

As mentioned in the paragraph explaining the methodology used, the documents considered for the following content analysis have been published between 2016 (year of adoption by the European Parliament of the GDPR) and 2020. This interval was considered in order to allow an analysis including the perceived impact before (2016-2017), in the moment of enforcement (2018) and after (2020) the GDPR. With this regard, it is important to highlight that 7 out of the 18 documents included have been published in 2018, showing that greater emphasis was given to the topic in the year when the regulation became enforceable. The operating geographical area of the publishing companies was also considered. Indeed, the businesses were classified in 3 categories: a) EU (for companies operating only within the European Union), b) Non-EU (for companies operating only outside the European Union), and c) Global (for companies operating both inside and outside the European Union). This helped me to understand to which geographical extent the GDPR had an impact. Despite being an EU regulation, 16 out the 18 documents taken into account were written and published online by companies classified as b) Non-EU (11 out of 18) and c) Global (5 out of 18). This can be explained by the *Global validity* of GDPR, interesting every company collecting European citizens' data.

In order to gain in a more systematic way insights from the content analysis performed to answer to the research question posited, the empirical part of my master thesis is divided in three main parts. The first two parts will focus on the presentation of the findings coming from the samples considered in the analysis. In particular, the findings will be presented following the categorization used for the coding process and based on the literature review. Therefore,

firstly, findings from the presentation of the GDPR will be presented, including the importance of publicizing consumer rights, data controllers, data portability, data processors, data profiling, global validity, PII, the purpose of GDPR, the pseudonymization, the right of subject access, the right to be forgotten, special categories, laws and regulations similar to GDPR and transparency. The second part will focus on the changes that businesses had to implement due to the GDPR. Therefore, I will present the findings regarding data collection, data storage and data interpretation/use changes before and after the GDPR, as well as the principle of accountability for companies and penalties in case of violation of the GDPR, the role of third parties, and the direct impact on tourism and travel sector. In the final paragraph, I will evaluate the findings through a critical discussion.

Table 1 Sources of data for content analysis

Title	**Reference Type**	**Company**	**Year of publication**	**Company type**	**Area**	**Country**
GDPR - Three Ways to Make it Easier for Aviation Companies	Podcast notes	Aviation Business Consultants	2018	Marketing Agency	Non-EU	USA
GDPR And Its Impact On The Hotel Sector	Interview	Hospitality News	2018	News	Non-EU	Lebanon
The impact of GDPR on the hotel sector	Report	Crowe Horwath	2017	Consultant	EU	Ireland
GDPR Countdown: How will the new EU data protection laws affect hotel marketing?	Blog	Pegasus	2017	IT company	Global	Global
GDPR: Top 5 Post-Implementation Issues for Airlines	Blog	Covington & Burling	2018	Law Firm	Global	Global
Hotel Data Management: The Impact Of GDPR On	Blog	Stayntouch	2017	IT company	Global	Global

The Hotel Sector						
How GDPR will impact the travel sector	Interview	Data and Marketing Association	2017	Business association	Non-EU	UK
How to Comply with GDPR: Recommendations for the Travel Industry	Blog	Altexsoft	2018	IT company	Non-EU	Global
How will GDPR affect tour operators, DMCs and travel agents?	Blog	Tourwriter	2018	IT company	Non-EU	Global
Should Car Rental Companies Comply with GDPR (General Data Protection Regulation)	Blog	Rentall	2018	IT company	Non-EU	Global
Do you know what the GDPR will mean for your hotel? The keys	Blog	Amara Marketing	2018	Marketing agency	EU	Spain
The Impact of GDPR on the Hospitality Sector	Blog	Social Hospitality	2018	Marketing agency	Non-EU	USA
The Impact of the GDPR on the Hospitality Sector	Blog	Niel Harper	2018	Consultant	Non-EU	Barbados
What does GDPR mean	Blog	Wherewolf	2018	IT company	Non-EU	Global

for Tourism Operators?						
PREPARING FOR DATA PROTECTION: GDPR READINESS IN THE CRUISE INDUSTRY	Press release	Future Cruise	2017	News	Non-EU	Global
THE GDPR ICEBERG: DATA PROTECTION IN THE CRUISE INDUSTRY	Briefing	HFW	2017	Law Firm	Global	Global
GDPR Two Years On: What are the lessons learnt for the aviation industry?	Article	Bird & Bird	2020	Law Firm	Global	Global
GDPR and the Travel Industry	Blog	Mha	2018	Consultant	Non-EU	Global

Source: own elaboration

4.1 Presentation of the GDPR

A total of 155 units of analysis have been coded to this category, including all the sub-categories, while only 17 are classified under the parent category only. This paragraph shows how GDPR was presented in the different documents analyzed. The content below served mainly as an introduction for many of the references. An important common point in 10 of 13 references included (Aviation Business Consultants, 2018; Covington & Burling, 2018; Stayntouch, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018 Wherewolf, 2018; Future Cruise, 2017; Bird & Bird, 2020; Mha, 2018) is the presence of data that mentions the time of enforcement of the GDPR: the period of enforcement.

As per the specific content, for the documents discussing the impact of GDPR in general on the travel sector highlight two main points: the right of data subjects and the nature of the tourism and travel industry, which gather and process large amounts of personal and sensitive

data, making them directly affected by the regulation. For instance, Niel Harper (2018) discusses the rights of data subjects under the GDPR. This reference highlights the need for tourism businesses to align their digital marketing practices with the GDPR's regulations to ensure the protection of individuals' data rights. Other references (Altexsoft, 2018; Stayntouch, 2017) emphasize that travel companies, including those involved in digital marketing, gather and process personal and sensitive data, making them directly impacted by the GDPR's requirements.

From the point of view of sectorial documents, it can be highlighted that from the aviation field, Bird & Bird (2020) suggests that the GDPR's stringent privacy requirements may serve as a model for other jurisdictions, implying that its influence extends beyond the European Union. As per the cruise industry, Future Cruise (2017) enhances the importance of taking control of data and offers a survival guide for the cruise industry to navigate the GDPR's requirements.

These references collectively highlight the need for tourism businesses to adapt their digital marketing practices to comply with the GDPR. They emphasize the significance of data protection, the rights of data subjects, and the global impact of the GDPR's privacy requirements.

4.1.1 Purpose

Overall, a total of 10 units from 9 sources (Hospitality News, 2018; Crowe Horwath, 2017; Pegasus, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Future Cruise, 2017) were coded to this category, which, in many cases, was part of the introduction and presentation of the GDPR.

The analysis shows that the GDPR aims to strengthen and unify data protection for individuals within the European Union (EU) by providing guidelines for the collection, processing, usage, and storage of personal information (Hospitality News, 2018; Crowe Horwath, 2017; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018).

Different sources highlight different aspects of the goals of the GDPR. For instance, Altexsoft (2018) underlines that the GDPR is replacing an old directive and it introduces stricter rules:

> The main goal. The GDPR’s main goal is to replace the Data Protection Directive 95/46/EC 1998 and to introduce a single data protection law that increases privacy for individuals by enforcing stronger security rules for companies that handle personal data. (Altexsoft, 2018)

Future Cruise (2017) and Rentall (2018) has similar points but the last one adds also that the GDPR is seen as a response to the increasing transfer of data through online mediums and aims to harmonize data privacy laws and to give more control to EU Citizens over their data:

> The necessity of the GDPR emerged due to the multitude of data being transferred via online mediums.The EU GDPR portal states that the GDPR was designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organizations across the region approach data privacy. (Rentall, 2018)

Finally, Crowe Horwath insist on the need to prepare and implement for its enforcement:

> GDPR is a regulation to strengthen and unify data protection for individuals within the European Union. It was adopted in May 2016 and following a two-year implementation period will come into force on May 25th 2018. The legislation brings in a large number of changes, meaning that the level of effort involved in preparing for GDPR compliance is significant. (Crowe Horwath, 2017)

4.1.2 Global validity

A total of 16 documents (Aviation Business Consultants, 2018; Hospitality News, 2018; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018) and 30 units were coded in this category, i.e., the highest number among all the sub-categories.

The main point of the analysis shows the GDPR has a significant global impact and applies to companies operating in all fields, including aviation companies, hotels, tour operators, travel agents, car rental companies, and the hospitality sector, both within and outside the European Economic Area (EEA). It imposes robust privacy requirements and affects organizations that process personal data of individuals from the EU or have customers, guests, or employees based in the EEA. In addition to this general common points, Bird & Bird (2020) focuses goes

more in detail with the aviation industry. According to the law firm international data transfers play a crucial role in the aviation industry but are subject to restrictions under the GDPR. Transfers of data to non-EEA countries without adequate data protection measures are limited. Aviation companies based in such jurisdictions must utilize specific data transfer mechanisms like Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs) when transferring customer or employee data from the EEA. In the absence of appropriate safeguards, data transfers may be justified through derogations or specific exemptions, albeit limited to exceptional circumstances.

4.1.3 Personally Identifiable Information (PII)

A total of 13 documents and 19 units were coded to this category (Hospitality News, 2018; Crowe Horwath, 2017; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Mha, 2018).

According to Hospitality News (2018) personal data under the GDPR includes an individual's name, phone number, location data, online identifiers, physical appearance, political and religious beliefs, biometric data, genetic information, and sexual orientation. This definition is reinforced by HFW (2017) which states that online identifiers like cookies and IP addresses can also make an individual identifiable. Additionally, the GDPR requires organizations in the hotel and tourism sectors to review their data collection, processing, and communication practices, as highlighted by Mha (2018). This involves outlining guidelines for collecting and managing PII), as mentioned by Stayntouch (2017). Moreover, organizations must maintain stricter controls on the collection, recording, and storage of personally identifiable data of European citizens (Tourwriter, 2018).

4.1.4 Right of subject access

With the results of the analysis of the text coded to the category right of subject access, I start a section about a range of rights given to individuals by the GDPR. 10 text units from 8 sources were coded to this category (Crowe Horwath, 2017; Covington & Burling, 2018; Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020, Mha, 2018).

The data analysis reveals that several sources discuss the right of individuals to access their data, rectify incorrect information, object to direct marketing, and request erasure or deletion of their data (Mha, 2018; Bird & Bird, 2020; Covington & Burling, 2018; HFW, 2017; Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018). Indeed, the GDPR strengthens and expands individuals' rights, such as the right to make subject access requests, which require organizations to provide comprehensive information within a shorter response period (HFW, 2017). Altexsoft deepens into this aspect and focuses on specific points for which individuals have the right to access:

> Be ready to respond to user requests According to regulation rules, all users have the right to ask companies: List the data stored with them; Define data collection purposes and uses cases; Outline the time period for which the personal data will be stored; Send a copy of all their data that is held; Delete the data about them. Each company is obligated to supply this information and process such requests. (Altexsoft, 2018)

4.1.5 Right to be forgotten

Another category very related to right of subject access is the right to be forgotten. A total 12 text entries from 8 sources were coded to this category (Covington & Burling, 2018; Altexsoft, 2018; Tourwriter, 2018; Amara Marketing, 2018; Wherewolf, 2018; Future Cruise, 2017; Bird & Bird, 2020; Mha, 2018), with the highest contribution from sources analyzing the impact of GDPR generally on the tourism sector.

As mentioned in the previous paragraph, according to GDPR, the individuals' rights include the ability to access their data, rectify incorrect information, request erasure or deletion of data, withdraw consent, and receive a portable copy of their data (Mha, 2018; Bird & Bird, 2020). Amara Marketing (2018) specifically highlight that the GDPR emphasizes the importance of clear notice and explicit consent when collecting personal data, discouraging ambiguous language and favoring in particular opt-in approaches in order to ensure control of the data and the right of erasure upon request. Altexsoft (2018) takes into account also third parties and explains the role of the company which originally obtained travelers' data:

> Travel companies also need to ensure they can control the process of data deletion by third parties with access to existing information. For instance, when users book a trip, a travel portal transfers the information to a hotel or car rental provider. (Altexsoft, 2018)

HFW (2017), focuses on the importance of time and a quick reply after the consumers' request of erasure his/her data:

> Right to request deletion. The controversial "right to be forgotten" has been strengthened, specifying the circumstances where the data controller must on request erase personal data without "undue delay" (for example where processing is based on the individual's consent but the individual has decided to withdraw that consent). (HFW, 2017)

Overall, the sources analyzed explain that compliance with these rights and obligations is essential for travel operators, tour operators, Destination Management Companies (DMCs), travel agents, and hotels to meet GDPR requirements and protect customers' privacy (Tourwriter, 2018; Amara Marketing, 2018; Wherewolf, 2018).

4.1.6 Special categories

A total of 9 text units from 5 sources (Data and Marketing Association, 2017; Social Hospitality, 2018; Niel Harper, 2018; HFW, 2017; Bird & Bird, 2020) were coded to this category.

The first relevant point coming out from the analysis of the units of text is that understanding the difference between 'personal data' and 'sensitive data' (or special categories) and how they should be treated is essential for compliance with the GDPR in the hospitality sector (Neil Harper, 2018).

Other relevant contributions come from the aviation industry. Bird & Bird (2020) explains that, in this context the GDPR recognizes that in certain limited circumstances, consent may be appropriate to collect and process information about a passenger's medical condition or dietary requirements, which could eventually disclose their religion or state of health, which are considered special categories of data. With this regard the same source adds that, it is mandatory for the aviation industry to

> Appoint a Data Protection Officer ("DPO") if their core activities consist of large-scale processing of special categories of personal data (for example, health information, trade union membership or biometric data used to uniquely identify individuals) and of data relating to criminal offences or regular and systematic monitoring of individuals on a large scale. (Bird & Bird, 2020)

On the other side, sources from the cruise industry explains that the GDPR tightens the grounds for processing sensitive personal data, limiting the circumstances under which such data can be processed and that it introduces particularly strict requirement for processing sensitive personal data (HFW, 2017)

As per the sources from the general tourism and travel sector these mention in particular the booking step as a point of collection of data under the special category:

> The travel industry will be particularly affected due to the large volume of personal and sensitive data it processes about individuals. For example, personal information collected as part of the booking process, including 'special category' (i.e. sensitive) data such as health

and medical data [...]. (Data and Marketing Association, 2017)

4.1.7 Data portability

A total of 5 text units from 5 sources (Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018; HFW, 2017; Mha, 2018) were coded to this category.

The content analysis highlights the importance of data portability and the need for travel industry businesses to review their processes and systems in relation to data subjects' rights under the GDPR.

The GDPR introduced a new right known as data portability. This provision empowers individuals to request the transfer of their personal data from one electronic processing system to another, given that the processing relies on their consent or the execution of a contract. The requested data should be furnished in a format that is structured, widely adopted, and machine-readable, enabling the individual to transmit it to another data controller. (HFW, 2017; Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018). For this reason, in order to achieve GDPR compliance in the travel industry, it is crucial to conduct a thorough evaluation of processes and systems related to handling data subjects' rights. This includes reviewing the implementation of new rights such as data erasure, data portability, and the use of profiling (Mha, 2018). Wherewolf (2018) also explains that, in addition to being transferable they "need to be adequately protected – both from unauthorized access and from disasters such as a data-centre fire".

4.1.8 Consumer rights must be publicized

A total 2 of text units from 2 sources were coded to this category: Bird & Bird (2020) and Data and Marketing Association (2017). The 2 contributions come from the aviation industry (Bird & Bird, 2020) and the general tourism and travel sector (Data and Marketing Association, 2017).

The content analysis emphasizes the need for the aviation industry to publicize consumer rights and comply with the General Data Protection Regulation (GDPR). For instance Bird & Bird (2020) explains that the most efficient way to ensure this right are privacy notice:

> Privacy notices are the most typical customer and employee facing documents that you need to have in place in order to explain to passengers, flight bookers, crew members and other employees how you process their personal information. The GDPR sets out specific information requirements relating to such documentation [...]. (Bird & Bird, 2020)

Data and Marketing Association (2017) is more general explaining the importance of enhanced consent mechanism explaining the distinction between consent and legitimate interest when processing personal data. These features are essential to protect consumer privacy rights.

4.1.9 Data processors

This paragraph and the next one (Data controllers) are considerably related and are presented in almost all the references. A total of 5 text units from 4 sources (HFW, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018) were coded to this category. 3 out of 4 sources refer to general tourism, with only one referring to the aviation industry.

Altexsoft (2018) focuses on defining who data processors are: "The processor is a person (other than an employee of the data controller) or a company that processes the data on behalf of the controller".

The analysis additionally highlights that data processors are responsible for processing personal data on behalf of organizations. This includes actions such as storing personal data. However, it's important to note that data processors have fewer obligations compared to data controllers under the GDPR. If a data processor's actions lead to a breach of GDPR obligations, the organization on whose behalf they are processing the data may be held liable unless they can prove their lack of responsibility (HFW, 2017).

Data and Marketing Association (2017) emphasizes a point that is related to the global validity of the GDPR: physical location is not a determining factor for data processors to fall under the scope of the GDPR. This means that even entities outside the European Union can be subject to the regulations.

4.1.10 Data controllers

This category has received a total of 3 text units from 3 different sources (HFW, 2017; Data and Marketing Association, 2017; Altexsoft, 2018). As mentioned above, this category is closely related to the one of data processors.

The content analysis overall highlights the role of data controllers and emphasizes the applicability of the GDPR to entities outside the European Union (similarly to data processors and in line with the findings about the global validity) (Data and Marketing Association, 2017).

More precisely, HFW (2017) gives a definition of data controllers: ""data controllers" make decisions on how and why personal data is processed [...]". Altexsoft (2018) does the same, defining data controllers as "A person or company that determines the purposes and the means of processing data" and states that the GDPR affect both data controllers and data processors.

4.1.11 Pseudonymization

Only one text unit from one source (HFW, 2017) was coded to this category, making it the one with the lowest number of codes, all categories considered. The text unit comes from the cruise industry and presents pseudonymization as an option for cases when anonymization is not possible. It also additionally insists on the importance of collecting data only when really needed.

4.1.12 Data profiling

This category is the one with the most relevant contribution to the RQ of this study. A total of 8 text units from 5 sources (Hospitality News, 2018; Mha, 2018; Pegasus, 2017; Altexsoft, 2018; Rentall, 2018) were coded to this category.

Also for this category, the result of the analysis shows a close connection with the section of global validity. Hospitality News (2018) highlights that the GDPR applies to the handling of information on all EU citizens, regardless of their location. Hotels based outside the EU but actively using their data to market, sell products and services, or monitor EU citizens or customers are required to comply with the GDPR.

Mha (2018), Pegasus (2017), Altexsoft (2018) and Rentall (2018), highlight that data is a valuable asset for travel businesses, particularly in the form of customer information. The ability to generate repeat customer sales and the overall value of a travel business heavily depend on the data they acquire and hold. However, in line with the GDPR, safeguarding and securing customer data, as well as ensuring accuracy, are crucial aspects that directly impact a business's success. Additionally, it is equally important to handle staff data securely and accurately.

Obtaining appropriate consent for electronic marketing activities and clearly communicating the purposes of information use are also essential. Regular audits of information practices are recommended, especially for travel companies collecting data for personalized services (Altexsoft, 2018).

In addition, Mha (2018) explicitly mentions target marketing campaigns and online tracking tools but clearly states that it is up to the organization to review processes in relation to the data processing for this purposes:

> Many operators will hold extensive marketing databases containing personal information. This information will be collected through bookings and administration and online and offline marketing activities. Information will be collected directly from individuals, but also via intermediaries such as travel agents and travel search websites. User profiling and online

> tracking tools such as cookies can be used to help better target marketing campaigns. The GDPR requires organisations to review the information held as well as their processes and adopt new procedures in relation to why and how that information is collected and used (Mha, 2018).

From car rentals, Rentall (2018), looks at both pros and cons of the GDPR, in a sector where customization through data is becoming crucial to determine the success or the failure of a business:

> Through GDPR compliance the company inadvertently creates a bond with the customer, a pact if you will. The customer is aware that the data provided will be safe, therefore he/she would not object to provide more data.
>
> It must be noted that in an era where data is of paramount importance, this is a blessing in disguise. It provides the perfect opportunity to develop your service with data provided by your customers. The car rental industry is driven by the needs of the customer, therefore customizing the service provided with the help of the new data the company can acquire it can guarantee it's client better service, which will no doubt increase client satisfaction which will in turn make the business more profitable. (Rentall, 2018)

4.1.13 Similar to the GDPR

A total of 2 text units from 2 sources (Aviation Business Consultants, 2018; Data and Marketing Association, 2017) were coded to this category.

The analysis shows that the sources selected mentioned 2 regulations comparable to the GDPR that were already implemented or had to be implemented at the time of publication of the documents considered. More precisely the other example come for USA, Canada and UK. Aviation Business Consultants (2018) presents the Controlling the Assault of Non-Solicited Pornography And Marketing (CAN-SPAM) Act of 2003 that was "signed into law by President George W. Bush on December 16, 2003, established the United States' first national standards for the sending of commercial e-mail" (Aviation Business Consultants, 2018) and the Canadian SPAM act of 2017 that allow "individuals and organizations who receive email that violates CASL to sue the sender for both actual and punitive damages" (Aviation Business Consultants,

2018). Both cases directly concern marketing practices. The third case, from the UK is presented by Data and Marketing association which mention the UK government's consideration of a new UK Data Protection Bill that will implement GDPR. Despite Brexit, the bill aims to adapt the regulations to local requirements, and it is expected to become law before the GDPR enforcement date of 25th May 2018. It has actually been enforced in 2022 and it is currently under revision (*UK Government Proposes Key Changes to the UK GDPR | Perspectives & Events | Mayer Brown*, 2023)

4.1.13.1 Past regulations

A brief sub-paragraph is needed to present the predecessors of the GDPR, which were mentioned in 4 text units from 4 sources (HFW, 2017; Crowe Horwath, 2017; Amara Marketing, 2018; Niel Harper, 2018).

All the sources mention and, in many cases, describe the 1995 EU Data Protection Directive (95/46/EC). Before the GDPR, there was the 1995 EU Data Protection Directive (95/46/EC), a legislative act that set out a goal for all EU countries without defining the means to achieve it. With this regard, Amara Marketing (2018) adds that the GDPR emerged as a need to establish new rules for a new upcoming digital era.

4.1.14 Transparency

The end of this first part of the analysis will focus on a particular recurring aspect of the GDPR: transparency. 19 text units from 10 sources (Hospitality News, 2018; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018) were included in this category, making it one of the most coded ones.

The results of the analysis show that the GDPR requires companies to be transparent about their data collection, storage, and sharing practices, empowering consumers with greater control over their personal data (Hospitality News, 2018). Organizations are now accountable for demonstrating compliance to regulators and maintaining records of data protection management (Mha, 2018). The GDPR has also prompted an increase in reporting personal data breaches and emphasizes the need for robust security measures due to the growing sophistication of security attacks (Bird & Bird, 2020).

Bird & Bird (2020) and HFW (2017) highlight that in order to ensure transparency it is essential to start from the consent of collecting and dealing with consumers' data. Consent plays a crucial role under the GDPR. It must be obtained in a transparent, clear, specific, informed, and unambiguous manner, meeting the GDPR's high standards (Bird & Bird, 2020). Consent provisions should not be buried within lengthy texts and must reflect the individual's clear

indication of their wishes (HFW, 2017). Additionally, HFW (2017) explains that maintaining up-to-date data protection policies and ensuring transparent data processing are essential.

On the other side, Future Cruise (2017), Data and Marketing Association (2017), Altexsoft (2018), Rentall (2018) and Social Hospitality (2018) insist on the beneficial effects of the GDPR's transparency on consumers. Indeed, GDPR compliance can enhance customer loyalty, trust, and satisfaction by demonstrating transparency and protecting personal data. It also presents an opportunity for businesses to improve practices and stand out from competitors (Data and Marketing Association, 2017) Building trustful relationships with customers through clear communication and providing personalized experiences based on collected data can be advantageous (Altexsoft, 2018).

After this first part where GDPR and its specific aspects, I will now move on with the analysis of the second big category: the changes that businesses had to implement due to the GDPR.

4.2 Changes

A total of 221 text units from all 18 documents were coded to this category, including sub-categories, while only 48 text units from 15 sources (Aviation Business Consultants, 2018; Hospitality News, 2018; Crowe Horwath, 2017; Pegasus, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

Considering the vast number of references coded to this parent-category, I will summarize the most common patterns identified divided by the 4 macro categories: aviation, hotel, cruise and general travel industry. Car rentals could not be included because no text unit was coded to the parent category, but only to the sub-categories.

In the provided data, several references discuss the impact of GDPR on various industries, particularly aviation, hotels, the travel industry, and the cruise industry. Common patterns can be identified in terms of the issues addressed and the recommended actions for compliance. The references also highlight the importance of data mapping, data security, awareness, and training for GDPR compliance.

In the aviation industry, GDPR compliance is essential for aviation companies to adapt to changing circumstances and meet regulatory requirements. Aviation Business Consultants (2018) mentions reasons for changing online marketing, such as complying with association rules, changing demographics, or countering competition and discusses the use of flexible marketing tools to make changes quickly and easily. For instance, the same source emphasizes the advantages of using WordPress as a stable and flexible website platform for adapting to GDPR and other changes in content and navigation and highlights the importance of Customer

Relationship Management (CRM) systems in following GDPR requirements, including sending templated emails that meet the regulation's wording requirements.

The hotel sector also faces the challenge of GDPR compliance. Hospitality News (2018) provides a plan for hoteliers to ensure GDPR compliance, including creating awareness and acquiring buy-in from management. Additionally, emphasis is put on making customers aware of their rights under GDPR. Among the recommendations and suggestions auditing and reviewing current data processes for storage and handling are also very important. Another relevant insight from the hotel sector is highlighted by Stayntouch (2017), which mentions the need for technical and organizational records to prove data protection, and the relevance data mapping to understand data capture, storage, and usage, followed by data security assessments.

As per the cruise industry, what emerges from the analysis is that it also faces challenges in complying with GDPR. A relevant contribution comes from HFW (2017), which highlights the increased information that must be notified to individuals and the mandatory clauses, discusses the grounds for processing personal data under GDPR, including legitimate interests, contractual necessity, legal obligations, and consent, and emphasizes the importance of data audits, record-keeping, and personal data minimization. As also highlighted in the case of the hotel sector, both HFW (2017) and Future Cruise (2017) advise auditing personal data processing, updating outdated data, and minimizing data to comply with GDPR.

Last but not least, the references from the general travel industry highlight a need to adopt new measures and update internal processes to demonstrate GDPR compliance, as stated by Mha (2018), but also as in line with the outcomes of the sectorial analysis. In addition, the same source raises questions about data collection, privacy notices, and consent rules. Another recurring aspect discussed by both Mha (2018) and Data and Marketing Association (2017) is the suggestion to conduct a full data audit to understand the information currently held (present also in the hotel and cruise industry).

After this general overview, I will now proceed with the specific analysis of the sub-categories related to the one about changes.

4.2.1 Data collection changes

A total of 55 text units from 15 sources (Aviation Business Consultants, 2018; Hospitality News, 2018; Pegasus, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Associaiton, 2017; Altexsoft, 2018; Tourwriter, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020; Mha, 2018). This is the sub-category with the highest number of text units coded.

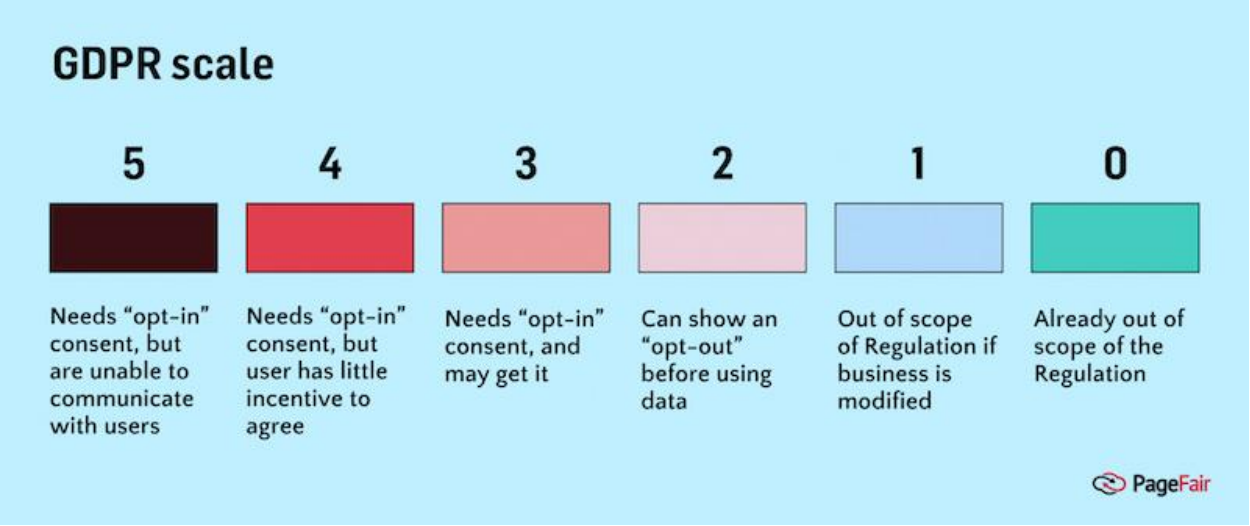
The first important result of the analysis discusses the changes related to consent and opt-in/opt-out methods. Mha (2018), Pegasus (2017) and Hospitality News (2018) highlight the

significant impact of GDPR on the marketing strategies of hotels. The implementation of GDPR introduces a shift from the widely used opt-out method to an opt-in approach for email marketing. Hotels, relying heavily on email communication with customers, now need explicit consent for data collection. This change may require hoteliers to obtain consent from customers during the check-in process, especially if data collection is involved. Loyalty programs also need to be examined to ensure compliance with consent requirements when using customer data:

> Customers will now have to opt in, or give consent to an email marketing service, unlike the opt-out method that has been widely used by companies in the past. This change may make it necessary for hoteliers to speak to customers at check-in, if explicit consent is required for any form of data collection. (Hospitality News, 2018)

In order to ease the understanding of GDPR and its impact on companies, Pegasus (2017) has created a GDPR scale that was used also to describe how several of Google's own ad products have been affected, including targeted advertising in the search engine, YouTube, Maps, and the Google Network.

Figure 9 The GDPR scale



Source: Pegasus (2017)

In addition, the explicit consent, the GDPR emphasizes that consent must be freely given, specific, informed, and unambiguous. This requirement aims to empower individuals by giving them control over their data and ensuring that their privacy rights are respected (Bird & Bird, 2020; Hospitality News, 2018).

Additionally, it is very important to know why data is being collected. GDPR mandates that businesses, including hotels, should have a clear understanding of the purpose behind data collection. This requirement ensures transparency (as also discussed in the previous

paragraphs) and accountability, enabling hotels to align their data practices with lawful processing conditions under GDPR (Bird & Bird, 2020, Hospitality News, 2018).

Hospitality News (2018) discussing the impact on hotel sector takes also into account the impact on payment procedures. It stresses the need for hotels to ensure that their payment processes comply with GDPR. Data protection regulations extend to payment transactions, requiring hotels to implement measures that safeguard customer payment information. By ensuring compliance with payment security standards, such as the Payment Card Industry Data Security Standard, hotels can protect sensitive data and maintain trust with customers.

Mha (2018) and Pegasus (2017) go beyond the role of consent and discuss the impact and the implications of GDPR for hotel marketing efforts, particularly in relation to Google's AdWords Customer Match. Customer Match enables hotels to create targeted advertising campaigns by using customer email addresses. However, GDPR requires explicit consent for using personal data in marketing. Hotels must explicitly inform guests from the EU that their collected emails may be used for advertising purposes and provide a clear opt-in mechanism for receiving marketing content.

> AdWords Customer Match is a form of remarketing that enables advertisers to upload customer email addresses to create highly specific PPC ads. These ads can then be displayed across the Search Network, Google Shopping, Gmail and YouTube. By segmenting email lists into demographics such as “family travelers” and “business travelers,” hotels can target customized campaigns at specific audiences, including existing customers and new prospects that share similar profiles. [...] While Google has been extremely cautious about the way Customer Match uses personal data (only accepting email addresses acquired through first-party means), new EU regulations will still require that consumers “opt-in” to receive marketing content through this tool. Hotel marketers will need to be very explicit in explaining to guests from the EU that any emails collected by the hotel may be used to advertise offers at a later date. (Pegasus, 2017)

Finally, another text unit coded to this category, also from Mha (2018) focuses on Facebook's Audience Network, which collects personal data from Facebook users for targeted advertising. This means of data collection may not be considered compatible with users' original activity, requiring explicit opt-in consent. GDPR restricts the use of "special" personal data, such as ethnicity, political opinion, sexual orientation, and religious beliefs, in personalized ads served through Facebook's platforms.

4.2.2 Data interpretation / use changes

A total of 54 text units from 14 sources were coded to this category, making it the second most coded one (Aviation Business Consultants, 2018; Crowe Horwath, 2017; Pegasus, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

With the regard of data implementation/ use, the impact of GDPR on the hotel sector is a topic that has garnered attention in various sources. Hospitality News (2018) emphasizes the importance of obtaining explicit consent and being transparent about data usage, similarly to data collection changes and inline with the results described in the paragraph about transparency. With this regard it links it to data collection and suggests that hotels can benefit from smarter data collection, which can lead to better customer experiences and targeted marketing messages. This approach can

result in a database of interested customers who are more likely to book and return to the hotel. In relation to transparency, Stayntouch (2017) highlights the requirements for providing detailed, clear and transparent information on data processing and retention policies. It emphasizes the need for organized data management to ensure compliance with GDPR regulations.

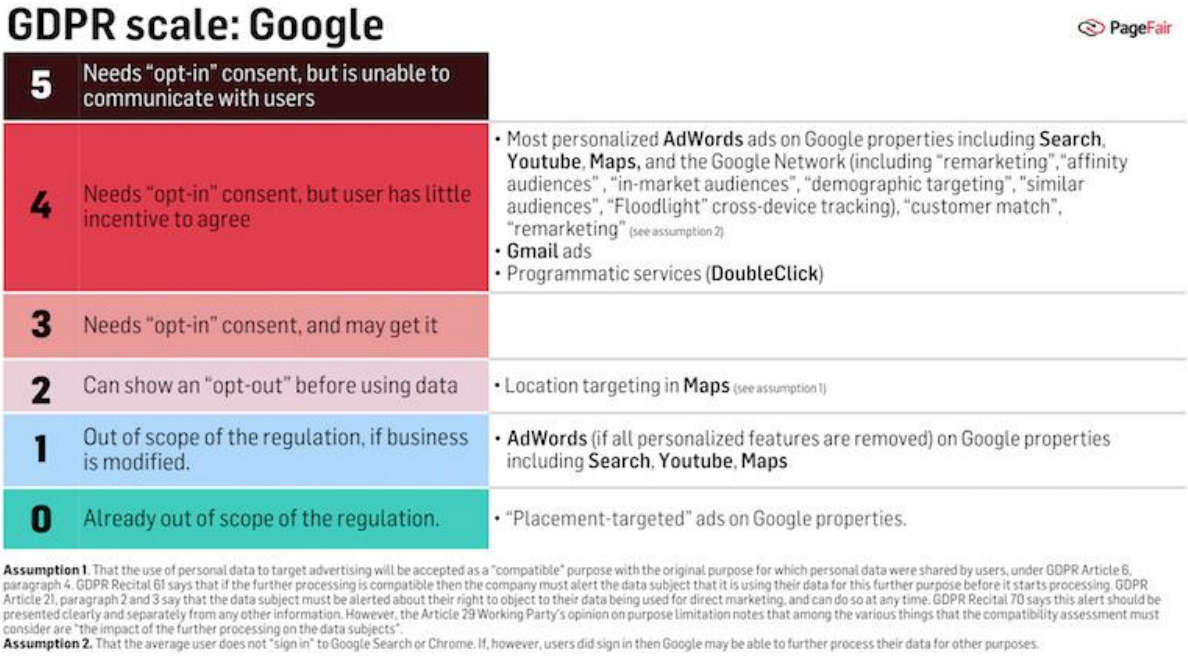
On the other hand, Mha (2018) focuses on the possible related downsides. It highlights the value of data for travel businesses and the potential negative effects on sales if data collection and management are not handled properly. It mentions the requirement for clear affirmative action, the end of pre-ticked boxes and bundled consents, and the need for expanded consent notices. The article emphasizes the importance of reviewing overall data processes and adopting new procedures to comply with GDPR:

> For a travel business, data is often the most valuable asset of the business - without a list of existing and past customers, the impact on a business's ability to generate repeat customer sales will have a seriously detrimental effect on its value when it comes to a potential sale. So how you acquire and hold data for existing and potential clients can be the difference between success and failure. It is also vital to make sure that the data you hold in respect of your staff is equally secure and accurate. (Mha, 2018)

As also presented in the paragraph about data collection changes, Pegasus (2017) addresses several aspects also with regard to data collection. The article mentions the requirement for explicit consent for personal data usage, the challenges faced by hotels when marketing to EU guests through platforms like Google and Facebook, and the restrictions on remarketing tools

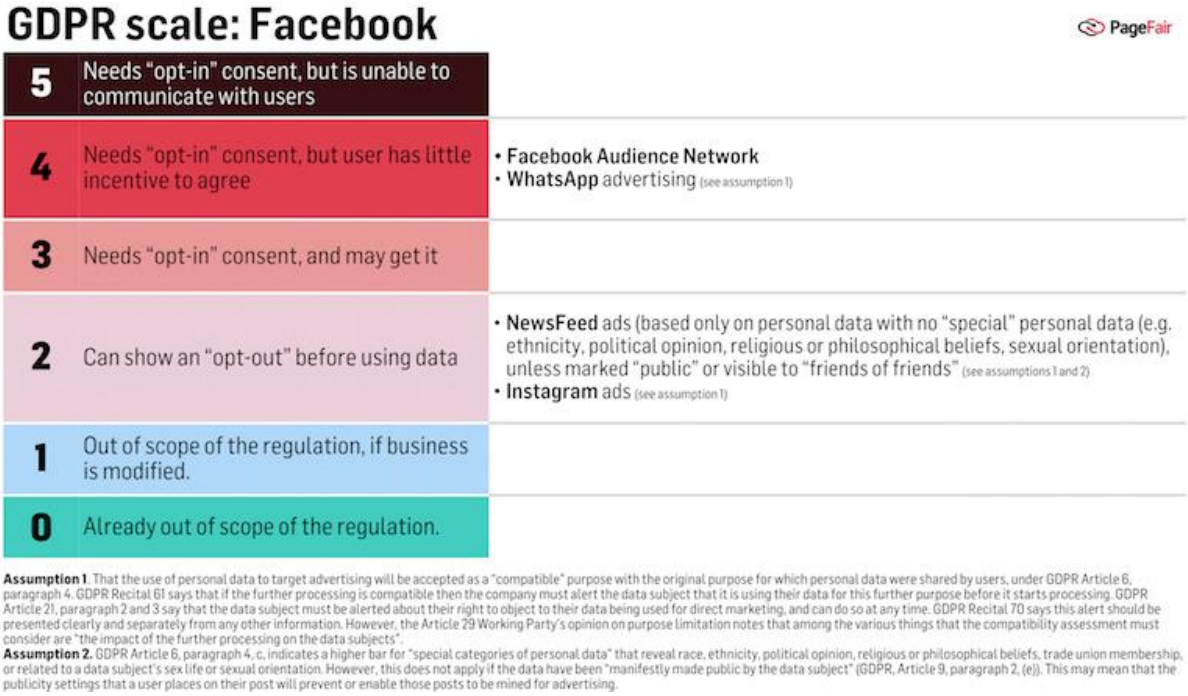
like Remarketing Lists for Search Ads and Customer Match. It also mentions the impact on Google and Facebook's advertising revenue due to GDPR restrictions.

Figure 10 GDPR scale: Google



Source: Pegasus (2017)

Figure 11 GDPR scale: Facebook



Source: Pegasus (2017)

The aviation industry point of view is discussed by Bird & Bird (2020) and Crowe Horwath (2017) who also mentions third parties. This source highlights the need for detailed data processing activity disclosure, including the legal basis for data processing and arrangements with third parties:

> Engagement with third parties
>
> The GDPR sets out specific requirements in relation to arrangements with data processors and other (joint) controllers. To the extent they haven't already, aviation companies should review their arrangements with third parties to ensure they meet such requirements. (Bird & Bird, 2020)

Bird & Bird (2020), Crowe Horwath (2017), Altexsoft (2018), Future Cruise (2017), HFW (2017) and Neil Harper (2018) highlight the need to appoint a Data Protection Officer (DPO) and the need for up-to-date contractual terms governing data sharing:

> Step 3 – Update GDPR policies and appoint data protection officer
>
> Hotels will need to review all current data protection policies such as their privacy policy, SARs (subject access request) policy, retention policy and other policies like shredding and breach management policy. Their policies relating to third party data contractors should be reviewed and consideration given to the appointment of a data protection officer (DPO). (Crowe Horwath, 2017)

4.2.3 Data storage changes

A total of 15 text units from 5 sources (Crowe Horwath, 2017; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018 Tourwriter, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Bird & Bird, 2020; Mha. 2018) were coded to this category. It is important to remark that many points are common with data collection and data interpretation / use changes, the three aspects being discussed contextually in many references.

Firstly, Social Hospitality (2018), Stayntouch (2017) and Mha (2018) highlight the significance of data in a travel business. According to Mha (2018), data is often the most valuable asset of a travel business. Without a list of existing and past customers, the ability to generate repeat customer sales is severely affected, impacting the business's overall value. With this regard,

the source emphasizes the importance of acquiring and securely holding data for existing and potential clients, as well as ensuring the security and accuracy of staff data. As already discussed in the previous paragraphs, also for data storage, consent management is another key aspect. Both Tourwriters (2018) and Mha (2018) highlight the new requirement for "clear affirmative action" and an end to pre-ticked boxes and bundled consents. Finally, both Mha (2018) and Bird & Bird (2020) also address data retention policies and practices. More precisely, it raises questions about how long information should be retained in marketing databases and emphasizes the need for a data cleansing policy. Without consent or a contractual requirement, businesses may be expected to destroy information related to completed travel arrangements:

> Data retention - how long do you retain information on your marketing databases? Do you have a data cleansing policy? Without consent you may be expected to destroy information after the travel arrangements have been completed and you no longer have a contractual requirement for it. (Mha, 2018)

Social Hospitality (2018) and Data and Marketing Association (2017) emphasize that data security is a shared concern. Social Hospitality (2018) suggests that hotels need to audit and review current data processes, implement secure data storage and handling methods, and educate employees on company-wide data security measures. It, therefore, gives emphasis also to the role of staff members in ensuring the correct data storage process. On the other side, Data and Marketing Association (2017) focuses on the importance of internal audits to ensure GDPR compliance in terms of data storage processes. It recommends conducting an internal audit of all personal data processing activities, including data collection purposes, legal basis, storage, security measures, and retention policies. Understanding these data processing activities is crucial for GDPR compliance.

These points align with the recommendations provided by Stayntouch (2017) which emphasizes the need for hotels to provide detailed information on the processing of personal data, including the planned duration of data retention. It highlights the importance of organized retention policies to ensure accurate data status. Finally, Stayntouch (2017), focuses on digital marketing and the collation of personal information, stressing the need for clear consent from the audience, specification of the desired data usage, and enabling guest access to modify or delete their information, with a clear mention to PII, and the right to be forgotten.

> When it comes to digital marketing and collating of personal information, Hotels need a section on their website that permits “opting in,” thus allowing hotels to store PII data. Hotels also must be able to prove that their audience has given consent for their data to be

used for marketing purposes, must also specify which data they wish to be used, and explain the process, enabling guests to access, modify and delete information. (Stayntouch, 2017)

4.2.4 Third parties

A total of 21 text units from 12 sources (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Stayntouch, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020; Mha, 2019) were coded to this category.

The analysis shows that the first common point across the references is the concept of shared responsibility for data processing. The GDPR emphasizes that both data controllers (such as hotels and businesses) and data processors (third-party providers) can be held accountable in the event of a data breach. This means that if a hotel outsources data processing to a third party that does not comply with GDPR regulations, both the hotel and the third-party processor can be held jointly responsible if a breach occurs (Social Hospitality, 2018; Mha, 2018; Bird & Bird, 2020; Covington & Burling, 2018; HFW, 2017; Crowe Horwath, 2017).

Another common point is the need for businesses to carefully review and assess their relationships with third-party partners. This applies to various aspects such as software providers, booking engines, CRM systems, payment processors, and marketing agencies. Businesses must ensure that these partners comply with GDPR regulations and adhere to the same data protection obligations. This review includes examining supplier arrangements, data processing agreements, and contracts to guarantee GDPR compliance throughout the data processing chain. Agreements must include specific terms and provisions related to data protection, usage restrictions, security measures, subcontractor restrictions, breach notification, and the return and deletion of data. (Social Hospitality, 2018; Mha, 2018; Bird & Bird, 2020; Covington & Burling, 2018; HFW, 2017; Stayntouch, 2017; Crowe Horwath, 2017; Altexsoft, 2018; Amara Marketing, 2018; Neil Harper, 2018).

In addition, Mha (2018) expressly mentions cookies and tracking tools that can be used, in combination with extensive marketing databases of personal information obtained from both individuals (direct collection) and via intermediaries (third parties) such as travel agents. Since the responsibility is shared by data collectors and processors, it is essential to review and adapt the processes to ensure GDPR compliance:

Many operators will hold extensive marketing databases containing personal information. This information will be collected through bookings and administration and online and offline marketing activities. Information will be collected directly from individuals, but also via intermediaries such as travel agents and travel search websites. User profiling and online

> tracking tools such as cookies can be used to help better target marketing campaigns. The GDPR requires organisations to review the information held as well as their processes and adopt new procedures in relation to why and how that information is collected and used. (Mha, 2018)

Another aspect related to the role of third parties and to transparency in data collection is the importance of clarifying and ensuring that privacy notices are provided to individuals at the time of data collection. This includes not only specifying the purpose of information use, but also informing individuals about data transfers to third parties and enabling them to exercise their data protection rights (Mha, 2018; Bird & Bird, 2020; Covington & Burling, 2018; Stayntouch, 2017; Crowe Horwath, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Wherewolf, 2018).

4.2.5 Accountability and penalties

This category was added to the ones identified from the literature review because of a recurring topic emerging from the sources analyzed. This can also be explained by the high number of text unit (28) and sources (14) (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018) coded to this category.

The first important point emerging from the analysis of the text units concerns financial penalties. Non-compliance with GDPR can lead to significant financial penalties. Businesses found to be in breach of the rules may be required to pay fines of up to €20 million or 4% of their global annual turnover, whichever is higher. (Hospitality News, 2018; Mha, 2018; HFW, 2017; Stayntouch, 2017; Crowe Horwath, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Future Cruise, 2017; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018). In addition, Hospitality News (2018), Mha (2018), Stayntouch (2017), Crowe Horwath (2017), Amara Marketing (2018) and Future Cruise (2017) mention another type of non-monetary penalty: reputational damage. Non-compliance with GDPR can result in a damaged reputation in the hospitality industry and adverse publicity, potentially leading to a loss of customer trust.

> Moreover, just as cruising is becoming even more popular and cruise ship order books are at an all-time high, financial penalties for GDPR non-compliance are a scary prospect. Those who fail to deliver face a fine of € 20 million or 4% of their global annual turnover (whichever of the two is higher). For most companies, this could signal irreparable damage

to their corporate reputation or even put them out of business altogether. (Future Cruise, 2017)

Another aspect that emerged from the analysis is the one related to accountability requirements. The GDPR emphasizes the principle of accountability, requiring organizations to not only comply with the regulations but also be able to demonstrate compliance. This includes maintaining records of processing activities and having data protection policies and procedures in place. (Mha, 2018; Bird & Bird, 2020; Stayntouch, 2017).

4.2.6 Direct impact of GDPR on tourism sector

A total of 46 text units from 14 sources (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018 Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018) were coded to this category.

Despite the high number of text units and sources, the findings about this category can be grouped under one main point that is shared among many of the references: the GDPR has had a severe impact on the tourism and travel sector due to its high level of processing on personal data. Table 2 presents all the text units coded to this category and mentioning this specific aspect. In many cases, these text units are presented by the sources as in relation to risk of data breaches and security risks (whose management is included in the GDPR and that will be discussed more thoroughly in the next paragraph).

Table 2 Sources relating high data volumes processing and GDPR

Reference	Source
Many operators will hold extensive marketing databases containing personal information. This information will be collected through bookings and administration and online and offline marketing activities. Information will be collected directly from individuals, but also via intermediaries such as travel agents and travel search websites.	Mha (2018)
Airlines, airport operators and their service providers, such as ground handling companies, routinely process large amounts of personal data: information about passengers, crew and other employees as well as personal data relating to suppliers and other business contacts. The highly regulated environment in which aviation players operate and the international character of their operations add another layer of complexity in respect of data protection compliance.	Bird & Bird (2020)

Airlines are uniquely affected by the GDPR with passenger data being at the heart of their business and international operations. As new technologies allow airlines to pursue new and innovative uses of customer data, it is imperative that airlines continue to conduct their operations with GDPR compliance in mind, particularly given the financial and other reputational issues that can arise for a failure to meet the GDPR's strict requirements.	Covington & Burling (2018)
Cruise lines control a lot of personal data. They collect and store information about their passengers' identities, preferences and health requirements. They hold information on their large workforces (whether employed or contracted). They have immigration law obligations in numerous jurisdictions. They conduct consumer-facing marketing campaigns. All of this information is likely to cross national borders and be exposed from time to time to physical and cyber security risk. The need to ensure data protection is already essential. Once the GDPR applies, and risk of large fines and reputational damage increases, a breach of the data protection rules could potentially sink the business (or at least cause it to take on water).	HFW (2017)
GDPR and the hotel sector When it comes to data security, there are few sectors as vulnerable to threats as the hotel industry. With the volume of processed personal and credit card information being handed over to hotels on a daily basis the hotel industry is currently one of the most vulnerable to data breaches (Verizon 2016 Data Breach Investigations). It is no surprise that the industry accounted for the second largest share of security breaches in 2016. With the enforcement deadline for the GDPR looming closer, it is imperative that hotels upgrade their data protection processes, or they face the risk of large financial penalties.	Crowe Horwath (2017)
The travel industry will be particularly affected due to the large volume of personal and sensitive data it processes about individuals. For example, personal information collected as part of the booking process, including 'special category' (i.e. sensitive) data such as health and medical data. [...] Travel companies also use data in marketing new promotions to people, as well as sharing large volumes of data with overseas suppliers, such as accommodation and excursion providers. All related activities must be reviewed and brought in line with the new regulations.	Data and Marketing Association (2017)
Travel industry perspective. Booking.com, the largest flight, and accommodation OTA, collects a broad spectrum of personal details,	Altexsoft (2018)

including names, travel purposes (leisure or work), travel with children, emails, payment data, etc. [...] Travel industry perspective. As OTAs, hotels, and airlines collect and store much of identifying personal data, from names to children's information, ensuring the right response to breaches becomes critical.	
Cruising is today's fastest growing leisure industry and with that accolade comes responsibility for managing infinite numbers of passenger records as well as crew member details.	Future Cruise (2017)
Why is the Hotel sector more sensitive to the GDPR then many others? Unlike the majority of sectors, the Hotel industry is extremely vulnerable to data security threats. The volume of sensitive personal data and credit card information, collected and processed makes the Hospitality industry one of the most vulnerable to data breaches (Verizon 2016 Data Breach Investigations). Online booking systems and multiple points of payment make the hotels an easy target for cyber attacks. According to the report, the industry accounted for the largest number of cyber incidents in 2016.	Amara Marketing (2018)
With regards to data security, there are few sectors more vulnerable to data-related threats than the hospitality sector. The volume of processed personal and credit card information being handed over to hotels, restaurants, etc. on a daily basis makes the sector extremely vulnerable	Neil Harper (2018)

Source: own elaboration

4.2.7 Others

This category was added to the ones identified from the literature review to include any text unit dealing with specific topics, but whose content did not fit into any other category previously mentioned. A total of 28 text units from 11 sources (Aviation Business Consultants, 2018; Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Stayntouch, 2017; Altexsoft, 2018; Social Hospitality, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

The main general points included in this category are:

1) Security concerns and data protection: Multiple sources emphasize the increasing importance of data security and privacy in the aviation, drone, hotel, travel, and cruise industries. Companies are adapting their marketing strategies and implementing measures to comply with data protection regulations like GDPR (Aviation Business Consultants, 2018; Altexsoft, 2018; Bird & Bird, 2020; Future Cruise, 2017).

2) Personal data breach reporting: GDPR requires timely reporting of personal data breaches within 72 hours. This requirement applies to various industries, including aviation, travel,

hotels, and airlines. Prompt incident response and notification are crucial to mitigate the potential impact of data breaches (Mha, 2018; Birds & Birds, 2020, Covington & Burling, 2018; HFW, 2017; Stayntouch, 2017; Altexsoft, 2018; Future Cruise, 2017).

3) Training and awareness: Many sources highlight the significance of training employees on personal data breaches and GDPR compliance. Industries such as hotels, airlines, and cruises recognize the importance of educating their staff members to handle personal data securely and responsibly (Hospitality News, 2018; Birds & Birds, 2020; Crowe Horwath, 2017; Future Cruise, 2017; Social Hospitality, 2018).

4) Ongoing compliance and monitoring: GDPR compliance is an ongoing process that requires regular updates to policies and procedures. Industries such as travel and hotels need to establish a proactive and adaptable approach to maintain compliance in a rapidly evolving regulatory landscape (Stayntouch, 2017; Crowe Horwath, 2017; Future Cruise, 2017).

4.3 Discussion of the findings

What emerges from the findings of the content analysis is that, considering how many text units were coded to the part of the presentation of the GDPR, the stakeholders included in the analysis recognize the significant importance of presenting the regulation and its specific aspects to the companies and/or stakeholders operating in the tourism and travel sector (target of the documents). The authors of the sources analyzed include business associations, law firms consultancies, and specialized press, whose aim is to support and ease the job of the businesses. Their documents can be seen as an attempt to explain in more easily comprehensible way what the GDPR is and how it will actually impact the businesses. As Hartley (2000) argues, in many cases laws are written in such ways that they cannot be understood by lay persons. Therefore, this can explain why the sources considered include a large part focusing on the presentation of the Regulation.

Another important insight related to the company type is its origin. 16 out of 18 companies analyzed operate in Non-EU or global (EU and Non-EU) level and this indicates that the reach of the GDPR goes beyond the borders of the EU. Indeed, this can be explained by the fact that despite being an EU regulation, the GDPR applies not only to companies within the EU but also to organizations outside the EU that process the data of EU citizens. The high number of text units coded to the category “Global validity” shows that many authors put a lot of emphasis on this point and indicates the importance of the GDPR's impact also outside the EU. The fact that Non-EU companies are often more concerned about GDPR enforcement compared to companies within the EU suggests that they put particular attention to the Regulation and potential consequences. This heightened concern may be due to many reasons, such as the unfamiliarity of Non-EU companies with EU data protection regulations, the potential impact on their business operations, the liability or the potential risk of penalties for non-compliance. Additionally, as emerges from the paragraph “Similar to the GDPR” the global impact of the GDPR can also be attributed to its role as a catalyst for global data protection reforms. The

GDPR has set a high standard for data protection and privacy regulations worldwide. In response to this framework, several countries and regions have introduced or revised their own data protection laws to align with the GDPR rules, such as the UK and its UK Data Protection Bill (Custers et al., 2018). Therefore, the global influence of the GDPR goes beyond its direct application and contributes to shape data protection practices worldwide.

Another aspect that must be considered is that the aspects of GDPR presented do not apply exclusively to the tourism sector but also to all the industries and sectors. Even though it may not have been explicitly designed and created for the tourism industry, its principles and requirements have a broad application and have an impact on all businesses that handle and deal with customers or staff's personal data. For this reason, although explicit mention of the impact on marketing and targeted/personalized marketing may be lacking in some of the documents considered in the content analysis, it is essential to consider the data and information-intensive nature of the tourism industry (Theocharidis et al., 2020; Tapsell et al., 2018; Li et al., 2018; Buhalis & Amaranggana, 2015; Herdin & Egger, 2007) and the importance of targeted marketing strategies in this sector (Shaw et al., 2021). Targeted marketing requires a deep understanding of the target audience, which involves collecting and using considerable amounts of data and information about individuals, that, as explained by Li et al. (2018), can include UCG data, device data and transaction data.

Going deeply into the findings about specific aspects of the presentation of the GDPR, one interesting insight is that pseudonymization, which is a mechanism that transforms personally identifiable information (PII) into non-PII data, received the lowest number of coded text units among the analyzed sources. Pseudonymization is a technique that can be used to enhance data privacy and security by dissociating individuals' identities from their data while still allowing for analysis and processing (European Parliament, 2016). The limited mention of pseudonymization in the analyzed sources could be attributed to several factors. Firstly, it is important to note that while pseudonymization is a technique that can contribute to data protection, it does not remove the data from the scope of the GDPR. The GDPR recognizes that pseudonymized data still falls within the realm of personal data if there is a possibility of re-identification. Authors such as Mourby et al. (2018) and Scheuing & Niininen (2022) emphasize the fact that pseudonymized data are still subject to the GDPR's regulations. This may lead to caution and reluctance among companies when employing pseudonymization as a data processing technique. Despite its potential benefits, there is a perceived risk associated with using pseudonymized data due to the possibility of re-identification and the potential implications for compliance with the GDPR. Moreover, the limited mention of pseudonymization could also stem from the fact that the GDPR places a stronger emphasis on anonymization as a means of achieving data protection. Anonymization involves irreversibly removing any identifying information from the data, ensuring that individuals cannot be re-identified. In contrast, pseudonymization retains the potential for re-identification if the necessary additional information is available. Considering the high level of scrutiny and strict regulations imposed by the GDPR, companies may opt for anonymization techniques rather than pseudonymization to minimize the risks associated with handling personal data.

Anonymization provides a more robust approach in terms of mitigating the potential re-identification of individuals, thereby offering a stronger safeguard for GDPR compliance (Hintze & El Emam, 2018).

Moving to the category "changes", an important insight to be considered is the high number of text units coded to the category "Data collection changes". More precisely, one relevant point is the new opt-in approach for marketing services (in contrast with the previous opt-out one), as highlighted by several sources. This specific point of the GDPR has two important consequences. The first one is indeed the enhanced level of transparency for users. This, as also highlighted in the homonymous paragraph by many sources, is in line with the GDPR purpose to empower individuals by providing them with clear and easily understandable information about how their personal data is collected, processed, and used. Therefore, thanks to the GDPR individuals should be more aware of the purposes for which their data is going to be used, who is involved in processing their data, and their rights about their personal information. In addition, transparency also helps, despite only partially, to address information asymmetry between individuals and organizations. It ensures that individuals have the all the information needed to make informed choices and understand the potential risks of sharing their data. By providing clear and accessible information, organizations can bridge the information gap and ensure that individuals have a fair understanding of how their data is being handled (European Parliament, 2016). Nonetheless, an important point to be considered and that was presented in the literature review, even though the organizations collecting data are obliged to inform the users about the purposes of data processing, many authors (Wirtz et al., 2007; Acquisti et al. 2015; O'Connor, 2020; McDonald & Cranor, 2008; Milne et al., 2006) highlighted the difficulty to find or understand the privacy statements. As van de Waerdt (2020) argues, the GDPR places the burden of responsibility on consumers, expecting them to gather information, understand privacy policies, and exercise their rights. However, this approach is unrealistic and fails to ensure transparency in the face of Data-Driven companies, like the ones in the tourism and travel sector.

The second point to be considered is that, despite the additional measures required to be implemented by companies (naming a DPO, including privacy statements, verifying the compliance of third parties, etc.), these can bring an added value to the business. Indeed, many sources highlight that GDPR compliance can enhance customer loyalty, trust, and satisfaction by demonstrating transparency and protecting personal data. This can be explained by the permission-based model and theory acceptance model (TAM).

As explained by Theocharidis et al. (2020), the permission-based model empowers individuals by giving them control over their personal data. By explicitly seeking their consent, organizations demonstrate respect for individuals' autonomy and privacy rights. This transparency in seeking permission fosters trust as individuals feel they have a say in how their data is used. Besides the model requires organizations to provide clear and accessible information about data processing activities. This transparency in communication helps individuals understand how their data will be handled, who will have access to it, and for what

purposes. Clear communication builds trust by reducing uncertainty and enabling individuals to make informed decisions about sharing their data. Finally, the permission-based model aligns with the accountability principle of the GDPR. Organizations are responsible for demonstrating compliance with data protection regulations and ensuring that consent is obtained and recorded. This accountability enhances trust as individuals see organizations taking responsibility for their data and adhering to legal obligations.

In a similar way, as discussed by Davis et al. (1989), the TAM explains how individuals' acceptance and trust in technology are influenced by perceived usefulness and ease of use. The TAM suggests that when individuals perceive that sharing their data is beneficial and will result in value-added services or personalized experiences, trust is enhanced. If organizations are able to communicate the benefits of processing consumers' data, individuals are more likely to trust that their data will be used in their best interest. This aspect is ensured by the GDPR thanks to the informed consent required by users and the opt-in approach. Moreover, the TAM highlights the importance of individuals perceiving data processing activities as easy to understand and engage with. When organizations provide clear and user-friendly information about data practices, individuals can comprehend how their data will be processed and any associated risks. This ease of understanding builds trust as individuals feel more confident in their interactions with online services. Again, this is another relevant aspect intrinsic to the informed consent needed from the users. Finally, trust is a central element in the TAM. When individuals trust that organizations will handle their data responsibly and protect their privacy, as required by the GDPR, they are more likely to accept and engage with online services. Transparent data practices, such as clear privacy policies, secure data handling procedures, and compliance with regulations, contribute to building trust and fostering individuals' acceptance of data processing activities.

Last but not least, another point that deserves a specific mention concerns 2 categories that were not chosen from the findings of literature review but emerged from the sources: "accountability and penalties and "others". Actually, the result of the analysis shows the two are very related and the link lies in their connection to GDPR compliance and its impact on financial penalties, reputational damage, accountability, data protection measures, security concerns, and personal data breach reporting. These aspects collectively underscore the significance of adhering to the GDPR's regulations and addressing privacy and security considerations in industries affected by the regulation. Furthermore, the link is also empirically proven by Wolff and & Atallah (2021) who found out that 2 years after the implementation of the GDPR, that the majority of fines imposed under the GDPR are a result of privacy violations, however it is noteworthy that the largest fines have been associated with security incidents. On average, security violations tend to attract higher fines compared to privacy violations. This indicates that while privacy breaches are a significant concern, security incidents have the potential for more severe consequences and thus attract more substantial penalties. Therefore, adhering to GDPR regulations and addressing both privacy and security considerations are crucial for organizations operating in industries impacted by the regulation.

5. Conclusions

The purpose of the thesis was to explore the consequences of the enforcement of the GDPR on targeted advertisement on the tourism field, with a particular focus on the point of view of the companies and businesses that operates in the tourism fields and put in act strategies of targeted marketing.

The qualitative analysis of the 18 documents selected was structured in 2 parts. The first part focused on the presentation of the GDPR, where 14 subcategories were identified from the literature review or from recurring topics in the documents analyzed.

The findings for the category "Purpose" revealed that the GDPR aims to enhance data protection for individuals within the EU by providing guidelines for collecting, processing, and storing personal information. It replaces an old directive, harmonizes data privacy laws, and empowers EU citizens over their data. Organizations need to prepare for GDPR compliance before its enforcement in May 2018 (Hospitality News, 2018; Crowe Horwath, 2017; Pegasus, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Future Cruise, 2017).

The category "Global validity" shows that the GDPR has a global impact, applying to companies in all tourism industries. It imposes strict privacy requirements and affects organizations handling personal data of EU individuals or having customers, guests, or employees with EU citizenship, no matter where they are located (Aviation Business Consultants, 2018; Hospitality News, 2018; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

The analysis revealed that personally identifiable information (PII) under the GDPR includes various data such as name, phone number, location, online identifiers, appearance, beliefs, biometric and genetic information, and sexual orientation. Organizations in the hotel and tourism sectors need to review their data practices and establish guidelines for collecting and managing PII. Stricter controls are required for the collection, recording, and storage of personally identifiable data of European citizens (Hospitality News, 2018; Crowe Horwath, 2017; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Mha, 2018).

As per "Right of subject access", the content analysis revealed that under the GDPR, individuals have rights to access their data, rectify incorrect information, object to direct marketing, and request data erasure. Organizations must respond to these requests by providing comprehensive information and deleting data when requested. The GDPR strengthens individual rights and imposes obligations on companies (Crowe Horwath, 2017; Covington & Burling, 2018; Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020, Mha, 2018).

The content analysis found that the “Right to be forgotten” is a significant aspect of GDPR. Individuals have the right to request erasure or deletion of their data, and companies must ensure control over third parties that have access to the data. Compliance with these rights is crucial for operators in tourism and travel sector to meet GDPR requirements and safeguard customer privacy (Covington & Burling, 2018; Altexsoft, 2018; Tourwriter, 2018; Amara Marketing, 2018; Wherewolf, 2018; Future Cruise, 2017; Bird & Bird, 2020; Mha, 2018).

Several sources also discussed what “Special Categories” of data are under the GDPR. Differentiating between personal data and special categories of data is crucial for GDPR compliance in the hospitality sector. The aviation industry may require a Data Protection Officer for large-scale processing of special categories of personal data, such as health information. The cruise industry faces tighter restrictions on processing sensitive personal data, while the general tourism sector highlights the collection of sensitive data, like health information, during the booking process (Data and Marketing Association, 2017; Social Hospitality, 2018; Niel Harper, 2018; HFW, 2017; Bird & Bird, 2020).

The content analysis found that “Data portability” is an important aspect of GDPR compliance in the travel industry. Individuals have the right to request the transfer of their personal data between electronic processing systems, provided the processing relies on their consent or the execution of a contract. Businesses need to evaluate their processes and systems to ensure they can fulfill data portability requests while also protecting the data from unauthorized access or disasters (Altexsoft, 2018; Tourwriter, 2018; Wherewolf, 2018; HFW, 2017; Mha, 2018).

Publicizing consumer rights is highlighted as another important aspect of the GDPR. Privacy notices serve as an efficient way to inform passengers, flight bookers, and employees about how their personal information is processed. Additionally, there is an emphasis on the need for enhanced consent mechanisms and distinguishing between consent and legitimate interest to protect consumer privacy rights in the tourism and travel sector (Bird & Bird, 2020; Data and Marketing Association, 2017).

The findings also show that data processors, who process personal data on behalf of organizations, have fewer obligations compared to data controllers under the GDPR. They are responsible for actions such as storing personal data, and if their actions lead to a breach, the organization may be held liable. The GDPR's scope includes entities outside the EU (due to the principle of global validity) meaning that physical location is not a determining factor for data processors (HFW, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018).

As per data controllers, different sources emphasize that the GDPR applies to entities outside the EU, similar to data processors, and in line with the global validity of the GDPR. Data controllers are defined as individuals or companies that determine the purposes and means of processing data, and they are subject to GDPR regulations along with data processors (HFW, 2017; Data and Marketing Association, 2017; Altexsoft, 2018).

The content analysis found only one text unit from HFW (2017) discussing pseudonymization, which is the lowest among all the categories. It highlights pseudonymization as an alternative when anonymization is not feasible, emphasizing the importance of collecting data only when necessary.

One of the most relevant findings for the RQ of this study is "Data profiling". The content analysis reveals a strong connection to the global validity section, where the GDPR applies to handling information of EU citizens regardless of their location. The analysis emphasizes the value of data for travel businesses and the need to safeguard, secure, and accurately handle customer and staff data. Obtaining consent for electronic marketing and conducting regular information practice audits are crucial, particularly for personalized services. The car rental industry sees GDPR compliance as an opportunity to develop better services through data customization, enhancing customer satisfaction and business profitability (Hospitality News, 2018; Mha, 2018; Pegasus, 2017; Altexsoft, 2018; Rentall, 2018).

In the category "Similar to the GDPR" two comparable regulations to the GDPR were mentioned: the CAN-SPAM Act of 2003 in the USA and the Canadian SPAM Act of 2017, both focusing on commercial email practices. Additionally, the UK considered a new Data Protection Bill to implement GDPR, aiming to adapt regulations to local requirements despite Brexit (Aviation Business Consultants, 2018; Data and Marketing Association, 2017; *UK Government Proposes Key Changes to the UK GDPR | Perspectives & Events | Mayer Brown*, 2023) Predecessors of the GDPR, such as the 1995 EU Data Protection Directive (95/46/EC), were mentioned in four sources, with the GDPR emerging as a response to the need for new rules in the digital era (HFW, 2017; Crowe Horwath, 2017; Amara Marketing, 2018; Niel Harper, 2018).

The content analysis reveals that the GDPR has led to an increase in reporting personal data breaches and emphasizes the need for robust security measures. Transparency plays a crucial role, and consent should be obtained in a clear, specific, and informed manner. GDPR compliance can enhance customer loyalty, trust, and satisfaction by demonstrating transparency and protecting personal data, providing an opportunity for businesses to improve practices and build trustful relationships with customers (Hospitality News, 2018; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

The second part focused on the changes that tourism and travel companies had to implement due to the enforcement GDPR. 7 subcategories were identified from the literature review or from recurring topics in the documents analyzed.

The content analysis found that data collection changes are a significant aspect of GDPR compliance. The shift from opt-out to opt-in methods for email marketing has impacted hotels, requiring explicit consent from customers. GDPR emphasizes that consent must be freely given, specific, informed, and unambiguous. Businesses, including hotels, must have a clear

understanding of the purpose behind data collection and ensure compliance with payment security standards. The use of customer data in targeted advertising, requires explicit opt-in consent (Aviation Business Consultants, 2018; Hospitality News, 2018; Pegasus, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Associaiton, 2017; Altexsoft, 2018; Tourwriter, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

As per data interpretation / use, the analysis reveals that GDPR has brought several changes. Obtaining explicit consent and ensuring transparency about data usage are emphasized. Smart data collection can lead to better customer experiences and targeted marketing. Proper data management and clear information on data processing and retention policies are crucial for compliance. Data's value for travel businesses is highlighted, emphasizing the need for secure and accurate data handling. Challenges arise in marketing to EU guests through platforms like Google and Facebook due to GDPR requirements. Disclosure of data processing activities and arrangements with third parties, appointment of a Data Protection Officer (DPO), and up-to-date contractual terms are also important considerations (Aviation Business Consultants, 2018; Crowe Horwath, 2017; Pegasus, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

The findings related to data storage changes in the context of the GDPR show that data is identified as a valuable asset for travel businesses, and acquiring and securely holding customer data is crucial. Consent management is emphasized, with a focus on clear affirmative action and the end of pre-ticked boxes. Data retention policies and practices, including data cleansing and destruction of information, are discussed. Data security measures, internal audits, and educating employees on data security are highlighted. Detailed information on data processing, including retention duration, is recommended, along with providing clear consent options for digital marketing and enabling guest access to modify or delete their information (Crowe Horwath, 2017; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018 Tourwriter, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Bird & Bird, 2020; Mha. 2018).

The analysis also reveals key points regarding third-party involvement in data processing under GDPR. Shared responsibility between data controllers and processors is emphasized, with both being accountable in case of data breaches. Businesses must carefully review relationships with third-party partners, ensuring their compliance with GDPR and data protection obligations. This includes examining supplier arrangements, data processing agreements, and contracts. The importance of privacy notices, specifying data use and transfers to third parties, and enabling individuals to exercise their data protection rights is also highlighted (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Stayntouch, 2017; Altexsoft, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Wherewolf, 2018; HFW, 2017; Bird & Bird, 2020; Mha, 2019).

The analysis of the category "Accountability and penalties" revealed important findings. Non-compliance with GDPR can lead to significant financial penalties, with fines of up to €20 million or 4% of global annual turnover. Reputational damage is another consequence, impacting customer trust and the hospitality industry. Additionally, the GDPR emphasizes the principle of accountability, requiring organizations to demonstrate compliance through record-keeping and data protection policies (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Stayntouch, 2017; Data and Marketing Association, 2017; Altexsoft, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020; Mha, 2018).

The GDPR has had a significant impact on the tourism and travel sector due to its extensive processing of personal data. The focus is often on the risk of data breaches and security concerns, highlighting the need for compliance with GDPR regulations (Hospitality News, 2018; Crowe Horwath, 2017; Covington & Burling, 2018; Data and Marketing Association, 2017; Altexsoft, 2018; Tourwriter, 2018; Rentall, 2018; Amara Marketing, 2018; Social Hospitality, 2018; Niel Harper, 2018; Future Cruise, 2017; HFW, 2017; Bird & Bird, 2020 Mha, 2018).

The category “Others” includes various topics not fitting into other categories. The main points addressed are data security and privacy concerns, reporting personal data breaches, employee training on GDPR compliance, and the need for ongoing compliance and monitoring to adapt to changing regulations (Aviation Business Consultants, 2018, Altexsoft, 2018; Bird & Bird, 2020; Future Cruise, 2017; Mha, 2018; Covington & Burling, 2018; HFW, 2017; Stayntouch, 2017; Hospitality News, 2018; Crowe Horwath, 2017; Social Hospitality, 2018).

Overall, the analysis shows that the impact of the enforcement of this regulation on tourism companies has been considerable. At the same time, the fact that almost all the sources considered for the analysis have a section presenting all the main points of the regulation enforced in 2018 and that does not necessarily apply only and specifically to the tourism industry opens the probability that the GDPR has had a consistent and general impact on all the industries. Nonetheless, a major link that could eventually translate into a higher impact for the tourism industry is its information-intensive nature (Buhalis and Amaranggana, 2015; Herdin and Egger, 2007) that put more emphasis on ethical issues such as data validity and impact on privacy. Individuals are still too little aware of the data that companies and third party collect about them and reuse to target specific advertisement to them, therefore this makes the enforcement of the GDPR more relevant and the measures to implement, somehow more impactful (including the fines and sanctions in case of violation).

At the same time, the analysis highlighted that the implementation of the GDPR can benefit the businesses processing data. One of the purposes of the GDPR is empower consumers and ensure higher transparency on how the data collected are going to be used by data controllers and data processors. As highlighted by Davis et al. (1989) and Theocharidis et al. (2020) respectively in their TAM and permission-based model, the opt-in data collection consent approach supported by the GDPR contributes to build trust between individuals and companies.

Indeed, by explicitly seeking their consent, organizations demonstrate respect for individuals' autonomy and privacy rights that eventually lead to let individuals feeling more confident in their interactions with online services.

5.1 Research limitations

The master thesis has also several limitations that should be acknowledged. These limitations are mainly related to the methodology used, the type data sources, and the consequences for the research question and the content analysis I conducted.

One of the main limitations of my study is that it is based on secondary data. The use of secondary data only, can have, as a consequence, that the information obtained may not be as precise and targeted as it would have been by using primary data. Methodologies based on primary data, such as surveys or interviews, allow researchers to collect data specifically tailored to their research objectives.

Additionally, in my thesis the literature review is based on information collected from a single database, Scopus. This limitation is mainly due to time constraints, which reduced my ability to explore multiple databases for a comprehensive systematic literature review. The selection of relevant information was done solely by me as a researcher, which may lead to subjectivity and potential bias in the process of collection and selection of the sources to include in my systematic literature review. With this regard, Puljak (2017) and Systematic Reviews (2020), emphasize the importance of involving multiple authors in the selection process to ensure objectivity, using several databases, aspect that require significant amount of time (usually 18 months).

Another limitation is the reliance on data sources that mainly represent the perspectives of stakeholders supporting or advising businesses in the implementation of the GDPR, such as consultancy companies, law firm or business associations. Indeed, these sources provide valuable insights and expertise, but at the same time they do not directly talk about the concrete experiences or viewpoints of the businesses that had to implement the GDPR themselves. As a result, the analysis may not fully show the specific challenges and issues faced by businesses in complying with the GDPR for tourism targeted marketing. Nonetheless, the secondary data I selected can still contribute relevant information and insights to my research question, as shown in the analysis.

5.2 Recommendations and future research

Based on the identified limitations of my master thesis, the following recommendations can be made for future research.

First of all, in order to address the limitation of relying on secondary data only, a study that uses primary data methods can be recommended. This could involve surveys, interviews, or

focus groups to gather data directly from businesses operating in the tourism field and that have implemented the GDPR targeted marketing. By collecting primary data, the research can obtain more targeted and focused results that directly address RQ. This would provide valuable insights into the challenges, successes, and specific strategies put in action by businesses in the GDPR.

Another recommended action is to complement this qualitative content analysis, with a quantitative analysis. By incorporating a quantitative approach, the researcher can obtain a larger and more comprehensive picture of the current situation. This can be achieved by using statistical software to analyze data, allowing for the identification of patterns, trends, and relationships between variables. The inclusion of quantitative analysis would also help to reduce subjectivity and bias that may arise when a single author is conducting the analysis, as in the case of this study and as highlighted in the paragraph about limitations.

Also, to enlarge the scope of my research and get a broader picture of the impact of data protection regulations and laws on targeted marketing in tourism, I suggest conducting a comparative analysis. Such a study could analyze the impact of the implementation of data protection regulations enforced not only by the EU but also by other countries such as the United States, Canada, and the United Kingdom, which, as underlined in paragraph 4.1.13 have developed similar strategies to protect the data of their citizens. By comparing different legislative frameworks and approaches for the implementation, the research could provide valuable insights highlighting the similarities, the differences, and the potential lessons that can be learned from various jurisdictions.

Finally, another related future research that could be worth to conduct is understanding how the GDPR could contribute and/or guarantee the transparency and protection of consumers' data in targeted marketing in tourism in sight of what are, at the time I am writing this thesis, the current scientific and technological developments of AI, that are raising further ethical questions among policy-makers world-wide (European Commission, 2023).

REFERENCES

1. Aalberts, R. J., Nill, A., & Poon, P. S. (2016). Online behavioral targeting: What does the law say?. *Journal of Current Issues & Research in Advertising*, *37*(2), 95-112.
2. Abdul-Ghani, E. (2020). Consumers' Online Institutional Privacy Literacy. *In: Martínez-López, F., D'Alessandro, S. (eds) Advances in Digital Marketing and eCommerce. Springer Proceedings in Business and Economics*. Springer, Cham. https://doi.org/10.1007/978-3-030-47595-6_6
3. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509-514.
4. Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unravelling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34–49.
5. Akerkar, R. (2019). Machine learning. In R. Akerkar (Ed.), SpringerBriefs in business. Artificial intelligence for business (pp. 19–32). *Springer International*. https://doi.org/10.1007/978-3-319- 97436-1_2
6. Alkire, L., Pohlmann, J., & Barnett, W. (2019). Triggers and motivators of privacy protection behavior on Facebook. *Journal of Services Marketing*, *33*(1), 57-72.
7. Althbiti, A., & Ma, X. (2020). Machine learning. In L. A. Schintler & C. L. McNeely (Eds.), Encyclopedia of big data (pp. 1–5). *Springer International*. https://doi.org/10.1007/978-3-319- 32001-4_539-1
8. Amara Marketing. (2018, January 16). Do you know what GDPR will mean for your hotel? The keys. - Amara, ingeniería de marketing. *Amara, ingeniería de marketing*. Retrieved June 26, 2023, from https://www.amara-marketing.com/en/travel-blog/gdpr-for-hotels/?hs_amp=true
9. Angwin, J. (2010). The web's new gold mine: Your secretsa journal investigation finds that one of the fastest-growing businesses on the internet is the business of spying on consumers; first in a series. *Wall Street Journal*.
10. Annells, M. (1996). Grounded theory method: Philosophical perspectives, paradigm of inquiry, and postmodernism. *Qualitative health research*, *6*(3), 379-393.
11. Ash, M. (2016, March 17). Why Is A Data Management Platform Essential In 2016? *Forbes*. Retrieved July 15, 2023, from https://www.forbes.com/sites/teradata/2016/03/17/why-is-a-data-management-platform-essential-in-2016/
12. Awad, M., & Khanna, R. (Eds.). (2015). Books for professionals by professionals. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. *Apress Springer Science+Business Media*. https://doi.org/10.1007/978-1-4302-5990-9
13. Baas, J., Schotten, M., Plume, A., Côté, G. and Karimi, R. (2020), "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies", *Quantitative Science Studies*, Vol. 1 No. 1, pp. 377-386.
14. Backman, K., & Kyngäs, H. A. (1999). Challenges of the grounded theory approach to a novice researcher. *Nursing & health sciences*, *1*(3), 147-153.

15. Barocas, S., & Selbst, A. (2016). Big data's disparate impact. California Law Review, 104, 671.
16. Bartsch, M., & Dienlin, T. (2016). Control your Facebook: An analysis of online privacy literacy. *Computers in Human Behavior*, *56*, 147-154.
17. Basarudin, N. A., & Raji, R. A. (2022). Implication of Personalized Advertising on Personal Data: A Legal Analysis of the EU General Data Protection Regulation. *Environment-Behaviour Proceedings Journal*, *7*(22), 109.
18. Batista, M., Fernandes, A., Ribeiro, L. P., Alturas, B., & Costa, C. P. (2020, June). Tensions between privacy and targeted advertising: Is the general data protection regulation being violated?. In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-5). IEEE.
19. Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*, 89-116.
20. Bauer, M. (2000). Classical content analysis: A review. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image, and sound: A practical handbook, (pp. 131151).* London:Sage.
21. Berelson, B. (1952). *Content analysis in communication research*. New York: Free Press
22. Berg, B. L (2001). *Qualitative research methods for the social sciences*. Pearson/Allyn & Bacon
23. Bernstein, A., & Kuleshov, A. (2014). Dimensionality reduction in statistical learning. In ICMLA '14: Proceedings of the 2014 13th international conference on machine learning and applications (pp. 330–335). IEEE Computer Society. https://doi.org/10.1109/ICMLA.2014.59
24. Birrer, F. A. (2005). Data mining to combat terrorism and the roots of privacy concerns. *Ethics and Information Technology*, *7*, 211-220.
25. Blass, J. (2019). Algorithmic advertising discrimination. *Nw. UL Rev.*,114, 415.
26. Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. *Journal of advertising*, *46*(3), 363-376.
27. Boes, K., Buhalis, D., & Inversini, A. (2015). Conceptualising smart tourism destination dimensions. In Information and communication technologies in tourism 2015 (pp. 391–403). Cham: Springer. https://doi.org/10.1007/978-3-319- 14343-9_29
28. Briner, R. B., & Walshe, N. D. (2014). From passively received wisdom to actively constructed knowledge: Teaching systematic review skills as a foundation of evidence-based management. *The Academy of Management Learning and Education*, 13(3), 415-432
29. Briner, R. B., & Walshe, N. D. (2014). From passively received wisdom to actively constructed knowledge: Teaching systematic review skills as a foundation of evidence-based management. *Academy of Management Learning & Education*, *13*(3), 415-432.
30. Bruce, N. I., Murthi, B. P. S., & Rao, R. C. (2017). A dynamic model for digital advertising: The effects of creative format, message content, and targeting on engagement. *Journal of Marketing Research*, April, 202–218.

31. Buhalis, D., & Amaranggana, A. (2015). Smart tourism destinations enhancing tourism experience through personalisation of services. In *Information and communication technologies in tourism 2015* (pp. 377–389). Springer. https://doi.org/10.1007/978-3-319-14343-9_28
32. Buhalis, D., (2003). eTourism: information technology for strategic tourism management. Pearson (Financial Times/Prentice Hall), London.
33. Bu-Pasha, S. (2017, September). Cross-border issues under EU data protection law with regard to personal data protection. *Information and Communications Technology Law*, *26*(3), 213–228.
34. Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70–79.
35. Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y. J., & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, 63(2), 227-243
36. Candy, C. P. (1989). Alternative paradigms in educational research. *The Australian Educational Researcher*, 16(3), 1-11.
37. Ceder, A. A., & Jiang, Y. (2020). Route guidance ranking procedures with human perception consideration for personalized public transport service. *Transportation Research Part C: Emerging Technologies*, *118*, 102667
38. Charters, D. (2002). Electronic monitoring and privacy issues in business-marketing: The ethics of the doubleclick experience. *Journal of Business Ethics*, 35, 243–252.
39. Chauhan, P., & Sood, M. (2021). Big data: Present and future. Computer, 54, 59–65. https://doi.org/ 10.1109/MC.2021.3057442
40. Chen, J., & Stallaert, J. (2014). An economic analysis of online advertising using behavioral targeting. *Mis Quarterly*, *38*(2), 429-A7.
41. Chester, J. (2012). Cookie wars: How new data profiling and targeting techniques threaten citizens and consumers in the "big data" era. *European Data Protection: In Good Health?*, 53-77.
42. Chiovitti, R. F., & Piran, N. (2003). Rigour and grounded theory research. *Journal of advanced nursing*, *44*(4), 427-435.
43. Chorus, C. G. (2012). Travel information: time to drop the labels?. *IEEE Transactions on Intelligent Transportation Systems*, *13*(3), 1235-1242.
44. Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications
45. Creswell, J. W. (2003). Research design: Qualitative, quantitative, and mixed methods approaches (2nd ed.). *Thousand Oaks*, CA: Sage.
46. Crotty, M. (1998). The Foundations of Social Research Meaning and Perspective in the Research Process. (pp. 1-17)
47. Custers, B., Dechesne, F., Sears, A. M., Tani, T., & Van der Hof, S. (2018). A comparison of data protection legislation and policies across the EU. *Computer Law & Security Review*, *34*(2), 234-243.
48. *Custom Targeting for travel advertisers | Expedia Group Media Solutions. (2022, August 22). Expedia Group Media Solutions. Retrieved July 15, 2023, from https://advertising.expedia.com/solutions/targeting/*

49. Davies, H. T., & Crombie, I. K. (1998). Getting to grips with systematic reviews and meta-analyses. *Hospital Medicine (london, England: 1998)*, *59*(12), 955-958.
50. Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), 982-1003.
51. Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management*, *54*(5), 847-860.
52. Dindar, M., & Yaman, N. D. (2018). #IUseTwitterBecause: Content analytic study of a trending topic in Twitter. *Information Technology and People*, 31(1), 256–277.
53. Dorcic, J., Komsic, J. and Markovic, S. (2019), "Mobile technologies and applications towards smart tourism–state of the art", *Tourism Review*, Vol. 74 No. 1, pp. 82-103
54. Drosatos, G., Efraimidis, P. S., Arampatzis, A., Stamatelatos, G., & Athanasiadis, I. N. (2015). Pythia: A privacy-enhanced personalized contextual suggestion system for tourism. In *2015 IEEE 39th Annual Computer Software and Applications Conference* (Vol. 2, pp. 822-827). IEEE.
55. Duboue, P. (2022). Feature Engineering: Human-in-the-Loop Machine Learning. In *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 109-127). Cham: Springer International Publishing.
56. Dy, J., & Brodley, C. (2004). Feature selection for unsupervised learning. Journal of Machine Learning Research, 5, 845–889. https://doi.org/10.1007/SpringerReference_302701
57. Egger, R. (2022a). Introduction: Data science in tourism. In R. Egger (Ed.), Tourism on the verge. Applied data science in tourism (pp. ix–xxiv). Springer.
58. Egger, R. (2022b). Machine Learning in Tourism: A Brief Overview: Generation of Knowledge from Experience. *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, 85-107.
59. Egger, R., & Yu, C.-E. (2022). Data Science and Interdisciplinarity. In R. Egger (Ed.), Tourism on the verge. Applied data science in Tourism (pp. 35–49). Cham: Springer
60. Egger, R., Kroner, M., & Stöckl, A. (2022a). Web scraping. In R. Egger (Ed.), Tourism on the verge. *Applied data science in tourism* (pp. 67–84). Springer
61. Egger, R., Neuburger, L., & Mattuzzi, M. (2022b). Data Science and Ethical Issues: Between Knowledge Gain and Ethical Responsibility. In Applied *Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 51-66). Cham: Springer International Publishing.
62. European Commission. (2023, June 30). *A European approach to artificial intelligence*. Shaping Europe's Digital Future. Retrieved July 7, 2023, from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
63. European Parliament. (2016, April). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. Eur-Lex. Retrieved March 11, 2023, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

64. Femenia-Serra, F., Neuhofer, B., & Ivars-Baidal, J. A. (2019). Towards a conceptualisation of smart tourists and their role within the smart destination scenario. *The Service Industries Journal*, *39*(2), 109-133.
65. Floridi, L (2014) Open data, data protection, and group privacy. *Philosophy & Technology* 27: 1–3.
66. Försch, S., and E. de Haan. 2018. "Targeting Online Display Ads: Choosing Their Frequency and Spacing." *International Journal of Research in Marketing* 35: 661–672.
67. Fowler, G.A.: It's the middle of the night. Do you know who your iPhone is talking to? Washington Post. https://www.washingtonpost.com/technology/2019/05/28/its-middle-night-do-you-know-who-your-iphone-is-talking/. (Accessed 18 February 2023).
68. FTC. 2014. "Data Brokers: A Call for Transparency and Accountability." Report. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf
69. Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC review*, *1142*(2011), 1-12.
70. Gao, J., Xie, C., & Tao, C. (Eds.). (2016). Big data validation and quality assurance—Issues, Challenges, and Needs. In 2016 IEEE Symposium on Service-Oriented System Engineering (SOSE). IEEE
71. *GDPR - Three Ways to Make it Easier for Aviation Companies*. (2018). Aviation Business Consultants. Retrieved June 26, 2023, from https://aviationbusinessconsultants.com/2018/05/gdpr-three-ways-to-make-it-easier/
72. *GDPR And Its Impact On The Hotel Sector*. (2018, August 11). Hospitality News. Retrieved June 26, 2023, from https://www.hospitalitynewsmag.com/gdpr-and-its-impact-on-the-hotel-sector/
73. *GDPR and the Travel Industry*. (2018, March 15). Mha.
74. *GDPR Countdown: How will the new EU data protection laws affect hotel marketing? | Pegasus*. (2017, November 14). Pegasus. Retrieved June 26, 2023, from https://www.pegs.com/blog/gdpr-countdown-how-will-the-new-eu-data-protection-laws-affect-hotel-marketing/
75. *GDPR Two Years On: What are the lessons learnt for the aviation industry?* (2020). Bird & Bird. Retrieved June 26, 2023, from https://www.twobirds.com/en/insights/2020/global/gdpr-two-years-on-what-are-the-lessons-learnt-for-the-aviation-industry
76. *GDPR: Top 5 Post-Implementation Issues for Airlines*. (2018). Covington & Burling. Retrieved June 26, 2023, from https://www.insideprivacy.com/airlines/gdpr-top-5-post-implementation-issues-for-airlines/#:~:text=GDPR%3A%20Top%205%20Post-Implementation%20Issues%20for%20Airlines%201,4%204.%20New%20Products%2C%20Apps%2C%20and%20Services%20
77. George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, 59 (5), 1493–1507. https://doi.org/10.5465/amj.2016.4005

78. Gilbert, F. (2008). Beacons, bugs and pixel tags. Do you comply with the FTC behavioural marketing principles and foreign law requirements? *Journal of Internet Law*, May, 3–10.
79. Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the journal of the British sociological association*, *12*(1), 27-49.
80. Godinho, P., Silva, M., & Moutinho, L. (2015). Tourism site location based on a genetic algorithm. *Tourism Analysis*, *20*(2), 159-172.
81. Goldfarb, A. (2014). What is different about online advertising?. *Review of Industrial Organization*, 44(2), 115-129
82. Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135-146.
83. Goodson, L., & Phillimore, J. (2004). The inquiry paradigm in qualitative tourism research. In Qualitative research in tourism (pp. 48-63). Routledge.
84. Goulding, C., & Saren, M. (2010). Immersion, emergence and reflexivity: Grounded theory and aesthetic consumption. *International Journal of Culture, Tourism and Hospitality Research*.
85. Gretzel, U., Reino, S., Kopera, S., & Koo, C. (2015). Smart tourism challenges. *Journal of Tourism*, *16*(1), 41-47.
86. Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179–188. https://doi.org/10.1007/ s12525-015-0196-8
87. Gretzel, U., Xiang, Z., Wober, K., & Fesenmaier, D.R. (2007). Deconstructing Destination Perceptions, Experiences, Stories and Internet search: Text Analysis in Tourism Research. In Woodside, A. & Martin, D. (2007). *Tourism management analysis, behaviour, and strategy* (pp.339-357). Wallingford, Oxfordshire Cambridge, MA: CABI Pub.
88. Guba, E. G. (1990). The Paradigm Dialog. Newbury Park: SAGE Publications
89. Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.
90. Ham, C. D. (2017). Exploring how consumers cope with online behavioral advertising. *International Journal of Advertising*, *36*(4), 632-658.
91. Hamid, R. A., Albahri, A. S., Alwan, J. K., Al-Qaysi, Z. T., Albahri, O. S., Zaidan, A. A., ... & Zaidan, B. B. (2021). How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. *Computer Science Review*, *39*, 100337.
92. Hardy, A. (2005). Using grounded theory to explore stakeholder perceptions of tourism. *Journal of tourism and cultural change*, *3*(2), 108-133.
93. Hartley, J. (2000). Legal ease and 'legalese'. *Psychology, Crime & Law*, *6*(1), 1-20.
94. Heinrich, H.A. (1996). Traditional versus computer aided content analysis. A comparison between codings done by raters as well as by Intext'. In F. Faulbaum and W. Bandilla (eds), SofiStat '95. *Advances in Statistical Software 5* (pp. 327-33), Stuttgart: Lucius & Lucius.
95. Herdin, T., & Egger, R. (2007). *Tourismus: Herausforderung: Zukunft*. LIT Verlag.

96. Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in human Behavior*, *50*, 618-625.
97. Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, M. Allen, & L. Klastrup (Eds.). The International Handbook of Internet Research, (pp. 233-249). Berlin: Springer Verlag. Retrieved June 10, 2023, from http://ella.slis.indiana.edu/~herring/webca.preprint.pdf
98. Hershberg, R. (2014). Constructivism. In D. Coghlan & M. Brydon-Miller (Ed.), The SAGE Encyclopaedia of Action Research (pp. 183-186).
99. Hershberg, R. (2014). Constructivism. In D. Coghlan & M. Brydon-Miller (Ed.), The SAGE Encyclopaedia of Action Research (pp. 183-186).
100. Hintze, M., & El Emam, K. (2018). Comparing the benefits of pseudonymisation and anonymisation under the GDPR. *Journal of Data Protection & Privacy*, 2(2), 145-158.
101. Hofstaetter, C., & Egger, R. (2009). The importance and use of weblogs for backpackers. Information and communication technologies in tourism 2009 (pp. 99–110)
102. *Hotel Data Management: The Impact Of GDPR On The Hotel Sector*. (2017). Stayntouch. Retrieved June 26, 2023, from https://www.stayntouch.com/blog/hotel-data-management-the-impact-of-gdpr-on-the-hotel-sector/
103. *How GDPR will impact the travel sector | DMA*. (2017). Data and Marketing Association. Retrieved June 26, 2023, from https://dma.org.uk/article/how-gdpr-will-impact-the-travel-sector
104. How to Comply with GDPR: Recommendations for the Travel Industry. (2018, March 15). *Altexsoft*. Retrieved June 26, 2023, from https://www.altexsoft.com/blog/business/how-to-comply-with-gdpr-recommendations-for-travel-industry/
105. How will GDPR affect Tour Operators, DMCs and Travel Agents? (2018). *Tourwriter*. Retrieved June 26, 2023, from https://www.tourwriter.com/gdpr-for-tour-operators-dmcs-and-travel-agents/
106. Hsieh, H.F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–88.
107. Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated machine learning. Springer International. https://doi.org/10.1007/978-3-030-05318-5
108. Im, H., & Ha, Y. (2013). Enablers and inhibitors of permission-based marketing: A case of mobile coupons. *Journal of Retailing and Consumer Services*, *20*(5), 495-503.
109. Jamal, S., Goyal, S., Grover, A., & Shanker, A. (2018). Machine learning: What, why, and how? In A. Shanker (Ed.), Bioinformatics: Sequences, structures, phylogeny (pp. 359–374). Springer Singapore. https://doi.org/10.1007/978-981-13-1562-6_16
110. Jennings, G. (2001). Tourism research. Milton, Qld.: Wiley Australia.
111. Jennings, G. (2009). Methodologies and methods. In T. Jamal & M. Robinson The SAGE handbook of tourism studies (pp. 673-693). London: SAGE Publications Ltd.
112. Jennings, G. (2015). Qualitative Research for University Sector: Paradigms that inform Research. Imagine Consulting Group International Pty Ltd

113. John, S. P. (2020). Digital Marketing Strategies in Educational Tourism: A Social Media Perspective. In *Advances in Digital Marketing and eCommerce: First International Conference, 2020* (pp. 114-121). Springer International Publishing.
114. Jordan, M. I. (2019). Artificial intelligence—The revolution hasn't happened yet Harvard. Data Science Review, 1(1). https://doi.org/10.1162/99608f92.f06c6e61
115. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science (New York, N.Y.), 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415
116. Kamii, C. (2016). Constructivism. In D. Couchenour & J. K. Chrisman (Ed.), The SAGE Encyclopedia of Contemporary Early Childhood Education (pp. 327-330).
117. Karaduman, O. (2017). The general data protection regulation: Achieving compliance for EU and non-EU companies. *Business Law International*, *18*(3), 225
118. Karimi, S., Stoev, Y., & Zander, O. (2017). Ethical Issues in E-Permission Marketing: A study of how consumer behaviour is affected by unethical practices involving E-Permission Marketing.
119. Kelle, U. (1995). Theories as heuristic tools in qualitative research. *Openness in research: The tension between self and other*, 33-50.
120. Kelle, U. (2007). " Emergence" vs." forcing" of empirical data? A crucial problem of" grounded theory" reconsidered. *Historical Social Research/Historische Sozialforschung. Supplement*, 133-156.
121. Kelm, O., Gerl, K., & Meißner, F. (2020). Machine learning. In I. Borucki, K. Kleinen-von Königslöw, S. Marschall, & T. Zerback (Eds.), Handbuch Politische Kommunikation (pp. 1–9). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-26242-6_55-1
122. Kensbock, S., & Jennings, G. (2011). Pursuing: A grounded theory of tourism entrepreneurs' understanding and praxis of sustainable tourism. *Asia Pacific Journal of Tourism Research*, *16*(5), 489-504.
123. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?. Business Horizons, 63(2), 135-146
124. Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data and Society, 1(1), 2053951714528481.
125. Kok, J. N., Boers, E. J. W., Kosters, W. A., van der Putten, P., & Poel, M. (2009). Artificial intelligence: Definition, trends, techniques and cases. Artificial Intelligence, 1, 270–29
126. Kramer, A. D., Guillory, J. E., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America, 111(24), 8788–8790.
127. Kumar, V., & Gupta, S. (2016). Conceptualising the evolution and future of advertising. *Journal of Advertising*, 45(3), 302–317.
128. Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387–394.
129. LaFollette, H. (2007). Ethics in practice. Blackwell.

130. Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, *6*(70), 1.
131. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2010). Big data, analytics and the path from insights to value. *MIT sloan management review*.
132. Lever, J., Krzywinski, M., & Altman, N. (2016). Points of significance: Model selection and overfitting. Nature Methods, 13(9), 703–704.
133. Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism management*, *68*, 301-323.
134. Liang, T. P., Chen, H. Y., & Turban, E. (2009, August). Effect of personalization on the perceived usefulness of online customer services: A dual-core theory. In *Proceedings of the 11th International Conference on Electronic Commerce* (pp. 279-288).
135. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, *151*(4), W-65.
136. Liberato, P., Liberato, D., Abreu, A., Alén-González, E., & Rocha, Á. (2018). Generation Y: the competitiveness of the tourism sector based on digital technology. In *Information Technology Science* (pp. 227-240). Springer International Publishing.
137. Lovelace, R., Birkin, M., Cross, P., & Clarke, M. (2016). From big noise to big data: Toward the verification of large data sets for understanding regional retail flows. *Geographical Analysis*, *48*(1), 59-81.
138. Lu, J., Meng, Y., Timmermans, H., & Zhang, A. (2021). Modeling hesitancy in airport choice: A comparison of discrete choice and machine learning methods. Transportation Research Part A:Policy and Practice, 147, 230–250. https://doi.org/10.1016/j.tra.2021.03.006
139. Marr, B. (2015). *Big data: Using smart big data, analytics and metrics to make better decisions and improve performance*. Wiley
140. Martin, D., & Woodside, A. G. (2008). Grounded theory of international tourism behavior. *Journal of Travel & Tourism Marketing*, *24*(4), 245-258.
141. Martin, D., Wu, H., & Alsaid, A. (2003). Hidden surveillance by Web sites: Web bugs in contemporary use. *Communications of the ACM*, *46*(12), 258-264.
142. Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, *45*, 135-155.
143. Mateosian, R. (2013). Ethics of big data. IEEE Micro, 33(2), 60–61.
144. Mayer-Schönberger, V. (1998). Internet privacy-The internet and privacy legislation: cookies for a treat?. Computer Law & Security Review: *The International Journal of Technology Law and Practice*, 3(14), 166-174.
145. Mayring, P. (2000). Qualitative Content Analysis [28 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1(2), Art. 20, Retrieved June 10, 2023, from http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204.
146. McCarthy, J. (2007). What is artificial intelligence?.
147. McDonald, A., & Cranor, L. (2008). The cost of reading privacy policies. *Information Science*, 4(3), 543–567.

148. McKinsey Global Institute analysis (2018). Artificial intelligence (AI) has the potential to create value across sectors. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020
149. McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, *7*(1), 2053951720904386.
150. Mehmetoglu, M., & Altinay, L. (2006). Examination of grounded theory analysis with an application to hospitality research. *International Journal of Hospitality Management*, 25(1), 12-33.
151. Mich, L. (2020). Artificial intelligence and machine learning. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), Handbook of e-tourism (pp. 1–21). Springer International. https://doi.org/10. 1007/978-3-030-05324-6_25-1
152. Mich, L. (2022). AI and Big Data in Tourism: Definitions, Areas, and Approaches. In *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 3-15). Cham: Springer International Publishing.
153. Miedema, T. (2018). Consumer protection in cyber space and the ethics of stewardship. *Journal of Consumer Policy*, 41(1), 55–75.
154. Millet, H. (2022). Leveraging booking information for offer personalization and GDPR consequences. *Journal of Revenue and Pricing Management*, 1-5.
155. Milne, G., Culnan, M., & Green, H. (2006). A longitudinal asssessment of online privacy notice readability. *Journal of Public Policy and Marketing*, *25*(2), 238–249.
156. Mingers, J., & Walsham, G. (2010). Toward ethical information systems: The contribution of discourse ethics. Mis Quarterly, 34(4), 833–854.
157. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data and Society, 3(2), 2053951716679679.
158. Mobile Marketing Association. (2011). Mobile advertising guidelines. Mobile Advertising Guidelines (mmaglobal.com) (accessed February 24, 2023).
159. Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. In Adaptive computation and machine learning. Cambridge Massachusetts (2nd ed.). The MIT Press
160. Moore, R. S., M. L. Moore, K. J. Shanahan, and B. Mack. 2015. "Creepy Marketing: Three Dimensions of Perceived Excessive Online Privacy Violation." Marketing Management 25: 42–53.
161. Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., ... & Kaye, J. (2018). Are 'pseudonymised'data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, *34*(2), 222-233.
162. Mulrow, C. D. (1994). Systematic reviews: rationale for systematic reviews. *Bmj*, *309*(6954), 597-599.
163. Munar, A. M., & Jamal, T. (2016). What are paradigms for? (pp. 1-16) Emerald Group Publishing Limited.
164. Murtagh, F., & Devlin, K. (2018). The development of data science: Implications for education, employment, research, and the data revolution for sustainable

development. Big Data and Cognitive Computing, 2(2), 14. https://doi.org/10.3390/bdcc2020014
165. Nakamura, L. (2013). Cybertypes: Race, ethnicity, and identity on the internet. Routledge.
166. Neubauer, S. (2021, November 24). *Hotel Data Management: The Impact Of GDPR On The Hotel Sector*. Stayntouch. https://www.stayntouch.com/blog/hotel-data-management-the-impact-of-gdpr-on-the-hotel-sector/
167. Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, Calif: Sage Publications.
168. Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. The Lancet Oncology, 20(5), e262–e273
169. Ngyuen, C. N., & Zeigermann, O. (2021). Machine learning: Kurz & gut (2. Auflage). dpunkt. verlag, O'Reilly, Preselect.media Gmbh.
170. Niel Harper. (2018, October 27). *The Impact of the GDPR on the Hospitality Sector*. Retrieved June 26, 2023, from https://nielharper.com/2018/06/26/the-impact-of-the-gdpr-on-the-hospitality-sector/
171. Niñerola, A., Sánchez-Rebull, M. V., & Hernández-Lara, A. B. (2019). Tourism research on sustainability: A bibliometric analysis. *Sustainability*, *11*(5), 1377.
172. Niñerola, A., Sánchez-Rebull, M. V., & Hernández-Lara, A. B. (2020). Quality improvement in healthcare: Six Sigma systematic review. *Health Policy*, *124*(4), 438-445.
173. Nunkoo, R., & Ramkissoon, H. (2016). Stakeholders' views of enclave tourism: A grounded theory approach. *Journal of Hospitality & Tourism Research*, *40*(5), 557-558.
174. O'Connor, P. (2004, January). Privacy and the online travel customer: An analysis of privacy policy content, use and compliance by online travel agencies. In *ENTER* (pp. 401-412).
175. O'Connor, P. (2007). Online consumer privacy. Cornell Hotel and Restaurant Administration Quarterly, 48(2), 183–200.
176. O'Connor, P. (2020). Online Tracking Using Cookies and Similar Technologies: An Analysis of Hotel Company Practices. In *The Emerald Handbook of ICT in Tourism and Hospitality* (pp. 33-44). Emerald Publishing Limited.
177. Ozdemir, S. (2016). Principles of data science: Learn the techniques and math you need to start making sense of your data. Packt. Retrieved from http://proquest.tech.safaribooksonline.de/ 9781785887918
178. Pahlevan-Sharif, S., Mura, P., & Wijesinghe, S. N. (2019). A systematic review of systematic reviews in tourism. *Journal of Hospitality and Tourism Management*, *39*, 158-165
179. Palmatier, R. W., & Martin, K. D. (2019). *The intelligent marketer's guide to data privacy: The impact of big data on customer trust*. Cham, Switzerland: Springer International Publishing.
180. Palos-Sanchez, P., Saura, J. R., & Martin-Velicia, F. (2019). A study of the effects of programmatic advertising on users' concerns about privacy overtime. *Journal of Business Research*, *96*, 61-72.

181. Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism management*, *32*(2), 215-224.
182. Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: Sage.
183. Petty, R. E., Briñol, P., Teeny, J., & Horcajo, J. (2017). The elaboration likelihood model: Changing attitudes toward exercising and beyond. *Persuasion and communication in sport, exercise, and physical activity*, 22-37.
184. Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development, 33(3),* 534e548. http://dx.doi.org/10.1080/07294360.2013.841651.
185. Popping, R. (2000). *Computer-assisted text analysis*. London Thousand Oaks, Calif: Sage Publications.
186. Power, D. J. (2016). Data science: Supporting decision-making. Journal of Decision Systems, 25(4), 345–356. https://doi.org/10.1080/12460125.2016.1171610
187. Pratama, I., Permanasari, A. E., Ardiyanto, I., & Indrayani, R. (Eds.). (2016). A review of missing values handling methods on time-series data. In 2016 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE.
188. *Preparing for data protection: GDPR readiness in the cruise industry*. (2017). Future Cruise. Retrieved June 26, 2023, from https://future-cruise.nridigital.com/future_cruise_mar18/preparing_for_data_protection_gdpr_readiness_in_the_cruise_industry
189. *Privacy Statement*. (2022). Booking.com. https://www.booking.com/content/privacy.en-gb.html?aid=397645
190. Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking (1st ed.). Sebastopol, CA: O'Reilly Media
191. Puljak, L. (2017). If there is only one author or only one database was searched, a study should not be called a systematic review. *Journal of Clinical Epidemiology*, *91*, 4.
192. Qin, R., Y. Yuan, and F. Wang. 2017. "Exploring the Optimal Granularity for Market Segmentation in RTB Advertising via Computational Experiment Approach." *Electronic Commerce Research & Applications* 24: 68–83. doi:10.1016/j.elerap.2017.07.001.
193. QSR International (2014). Retrieved June 10, 2023 from http://www.qsrinternational.com/
194. Randall, N. (1997). The new Cookie Monster. PC Magazine 16 (8).
195. Raschka, S. (2018, November 13). Model evaluation, model selection, and algorithm selection in machine learning. Cornell University. https://arxiv.org/abs/1811.12808
196. Rongrong, Y. (2017, October). A mobile smart tourism and marketing system design for harbin. In *2017 International Conference on Robots & Intelligent System (ICRIS)* (pp. 12-14). IEEE.

197. Rubin, H. J., & Rubin, I. S. (2011). Qualitative interviewing: The art of hearing data. sage.
198. Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology*, 21, 197–208.
199. Sanchez, J. S. (2003). Analysis of new techniques to obtain quality training sets. *Pattern Recognition Letters*, 24(7), 1015–1022. https://doi.org/10.1016/S0167-8655(02)00
200. Schermer, B. (2013). Risks of profiling and the limits of data protection law. In B. Custers, T. Calders, B. Schermer, & T. Zarsky (Eds.), Discrimination and privacy in the information society (pp. 137–152). Springer.
201. Scheuing, S., & Niininen, O. (2022) GDPR guidelines for academic research in marketing. *Contemporary Issues in Digital Marketing*, 139.
202. Schoen, S. (2009). New cookie technologies: Harder to see and remove, widely used to track you. Electronic Frontier Foundation. Technical Analysis. New Cookie Technologies: Harder to See and Remove, Widely Used to Track You | Electronic Frontier Foundation (eff.org) (accessed February 23, 2023).
203. Schreiber, R. S. (2001). Theory: Avoiding the Pitfalls. *Using grounded theory in nursing*, 55.
204. Seely-Gant, K., & Frehill, L. M. (2015). Exploring bias and error in big data research. *Journal of the Washington Academy of Sciences*, 101(3), 29–38.
205. Shah, R. (2015). Law enforcement and data privacy: a forward-looking approach.
206. Sharma, A., & Nicolau, J. L. (2020). An open market valuation of the effects of COVID-19 on the travel and tourism industry. *Annals of Tourism Research*, 83, 102990. https://doi.org/10.1016/j.annals.2020.102990
207. Shaw, F. A., Wang, X., Mokhtarian, P. L., & Watkins, K. E. (2021). Supplementing transportation data sources with targeted marketing data: Applications, integration, and internal validation. *Transportation Research Part A: Policy and Practice*, *149*, 150-169.
208. Should Car Rental Companies Comply with GDPR (General Data Protection Regulation). (2018). *RENTALL*. Retrieved June 26, 2023, from https://www.rentallsoftware.com/blog/should-car-rental-companies-comply-with-gdpr-general-data-protection-regulation/
209. Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, *18*(5), 587-606.
210. Sinclair, J. (2016). Shift or Stasis| Advertising and Media in the Age of the Algorithm. *International Journal of Communication*, *10*, 14.
211. Sipior, J. C., Ward, B. T., & Mendoza, R. A. (2011). Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons. *Journal of Internet Commerce*, 10(1), 1–16. doi:10.1080/15332861.2011.558454
212. Skilton, M., & Hovsepian, F. (2018). Machine learning. In M. Skilton & F. Hovsepian (Eds.), The 4th industrial revolution (pp. 121–157). Springer International. https://doi.org/10.1007/978-3-319-62479-2_5

213. Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS quarterly*, 989-1015.
214. Sotiriadis, M., Shen, S., & Zhou, Q. (2020). Influence of Social Networks on Responsible Behaviour by Smart Tourists. In *Advances in Digital Marketing and eCommerce: First International Conference, 2020* (pp. 9-16). Springer International Publishing.
215. Steinberg, D., & Aronovich, E. (2020). Thoughts on data science in business and industry. Applied Stochastic Models in Business and Industry, 36.1, 36–40. Retrieved from https://www.researchgate.net/profile/eddie-aronovich-2/publication/338692827_thoughts_on_data_science_in_business_and_industry/links/ 5e7955674585158bd501b45a/thoughts-on-data-science-in-business-and-industry.pdf
216. Stephenson, M. T., Benoit, W. L., & Tschida, D. A. (2001). Testing the mediating role of cognitive responses in the elaboration likelihood model. *Communication Studies*, *52*(4), 324-337.
217. Stumpf, T. S., Sandstrom, J., & Swanger, N. (2016). Bridging the gap: Grounded theory method, theory development, and sustainable tourism research. *Journal of Sustainable Tourism*, *24*(12), 1691-1708.
218. *Systematic Reviews*. (2020, June). Centers for Disease Control and Prevention. Retrieved July 7, 2023, from https://www.cdc.gov/library/researchguides/systematicreviews.html
219. Tapsell, J., Akram, R. N., & Markantonakis, K. (2018, August). Consumer centric data control, tracking and transparency–a position paper. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 1380-1385). IEEE.
220. *THE GDPR ICEBERG: DATA PROTECTION IN THE CRUISE INDUSTRY*. (2017, October). HFW. Retrieved June 26, 2023, from https://www.hfw.com/Avoiding-the-GDPR-iceberg-data-protection-in-the-shipping-industry-September-2017
221. *The Impact of GDPR on the Hospitality Sector - Social Hospitality*. (2018, February 14). Social Hospitality. Retrieved June 26, 2023, from https://socialhospitality.com/2018/02/the-impact-of-gdpr-on-the-hospitality-sector/
222. The impact of GDPR on the hotel sector. (2017). In *Crowe Horwath*. Retrieved June 26, 2023, from https://www.crowe.com/ie/insights/the-impact-of-gdpr-on-the-hotel-sector
223. Theocharidis, A. I., Argyropoulou, M., Karavasilis, G., Vrana, V., & Kehris, E. (2020). An approach towards investigating factors affecting intention to book a hotel room through social media. *Sustainability*, *12*(21), 8973.
224. Thomas, P. Y. (2010). Towards developing a web-based blended learning environment at the University of Botswana (Doctoral dissertation).
225. Thornberg, R. (2012). Informed grounded theory. *Scandinavian journal of educational research*, *56*(3), 243-259.

226. Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, *14*(3), 207-222.
227. Tribe, J. (2001). Research paradigms and the tourism curriculum. *Journal of travel research*, *39*(4), 442-448.
228. Tribe, J. (2004). Knowing about tourism: Epistemological issues. In *Qualitative research in tourism* (pp. 64-66). Routledge.
229. *UK Government proposes key changes to the UK GDPR | Perspectives & Events | Mayer Brown*. (2023, March). Retrieved July 8, 2023, from https://www.mayerbrown.com/en/perspectives-events/publications/2023/03/uk-government-proposes-key-changes-to-the-uk-gdpr
230. Ulrich, P. (2008). Integrative economic ethics: Foundations of a civilized market economy. Cambridge University Press.
231. van de Waerdt, P. J. (2020). Information asymmetries: recognizing the limits of the GDPR on the data-driven market. *Computer Law & Security Review*, 38, 105436.
232. Varnali, K. (2019). Online behavioral advertising: An integrative review. Journal of Marketing Communications, 1–22. doi:10.1080/13527266.2019.1630664
233. Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative health research*, 16(4), 547-559.
234. Weare, C., & Lin, W. Y. (2000). Content analysis of the World Wide Web—Opportunities and challenges. *Social Science Computer Review*, 18(3), 272-292.
235. Wherewolf. (2018, August 15). What does GDPR mean for Tourism Operators? *Wherewolf*. Retrieved June 26, 2023, from https://getwherewolf.com/what-does-gdpr-mean-for-tourism-operators/
236. Wijesinghe, S.N., Mura, P. and Bouchon, F. (2019), "Tourism knowledge and neocolonialism – a systematic critical review of the literature", *Current Issues in Tourism*, Vol. 22 No. 11, pp. 1263-1279
237. Wilson, E., & Hollinshead, K. (2015). Qualitative tourism research: Opportunities in the emergent soft sciences. *Annals of Tourism Research*, *54*, 30-47.
238. Wirtz, J., Lwin, M., & Williams, J. (2007). Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management*, *18*, 326–341.
239. Wolff, J., & Atallah, N. (2021). Early GDPR penalties: Analysis of implementation and fines through May 2020. *Journal of Information Policy*, 11, 63-103.
240. Worden, K., Haywood, J., & Bullough, W. A. (2003). Smart technologies. River Edge, NJ: World Scientific.
241. Xiang, Z., & Fesenmaier, D. R. (2017). Analytics in Smart Tourism Design. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44263-1
242. Xiang, Z., Tussyadiah, I., & Buhalis, D. (2015). Smart destinations: Foundations, analytics, and applications. *Journal of Destination Marketing and Management*, *4*(3), 143-144.
243. Xu, F., Buhalis, D., & Weber, J. (2017). Serious games and the gamification of tourism. *Tourism management*, *60*, 244-256.

244. Yang, E. C. L., Khoo-Lattimore, C., & Arcodia, C. (2017). A systematic literature review of risk and gender research in tourism. Tourism Management, 58, 89–100. doi:10.1016/j.tourman.2016.10. 011
245. Young, M. (1999). *The Social Construction of Tourist Places. Australian Geographer, 30(3), 373–389.* doi:10.1080/00049189993648
246. Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science*, (pp. 308-319). Westport, CT: Libraries Unlimited. Retrieved June 10, 2023 from https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf