Sezgin Savaş, Emre Ergen, "The Use of the Personal Data Collected Through Digital Footprints by Corporations in Understanding the Target Audience: An Analysis on Dot-com Companies", **Istanbul Gelisim University Journal of Social Sciences,** 10 (2), October 2023, pp. 668-689.

The Use of the Personal Data Collected Through Digital Footprints by Corporations in Understanding the Target Audience: An Analysis on Dot-com Companies

Sezgin SAVAŞ[*], Emre ERGEN[**]

Abstract

This study aims to reveal the categories in which companies collect data the most as well as the changes occurring in terms of these categories. The study, within this context, explores the connexion between the tendency to collect data, revenue, and customer satisfaction in order to lay bare the contribution of data collection to understanding the customer. The findings of the study suggest that the considerable amount of personal data collected by companies has no more than little bearing on customer satisfaction. The study also shows that companies request more data in certain categories. Furthermore, there appears to be an ambiguity in terms of the link between the tendencies of data collection and revenue/customer satisfaction. Another striking point that the findings of the study make is that the data that the companies request to collect vary by sectors with little difference.

Keywords: Digital footprint, passive footprint, active footprint, personal data, target audience

Kurumların Hedef Kitlelerini Tanıma Aşamasında Dijital İzler Yardımıyla Toplanan Kişisel Verilerin Kullanımı: İnternet Şirketleri Üzerinden Bir Değerlendirme

Öz

Bu çalışma, kurumların bireyler hakkında depoladığı kişisel veri tercihlerinin hangi kategorilerde yoğunlaştığı ve ilgili kategorilerin sektörel anlamda değişimlerini ortaya koymayı amaçlamaktadır. Bu amaçla bağlantılı olarak çalışma, veri toplama eğilimi ile ciro ve memnuniyet arasındaki bağlantıyı da irdelemekte ve veri toplamanın tüketiciyi tanıma aşamasında sağladığı katkıyı tartışmaktadır. Çalışmanın sonuçlarına göre kurumların önemli düzeyde kişisel veri toplamasına rağmen bunların tüketiciler üzerindeki memnuniyet yansımasının sınırlı olduğu görülmüştür. Çalışmada, kurumların özellikle kişisel veri talep etme anlamında bazı kategorilerde yoğunluk gösterdiği tespit edilmiştir. Bununla birlikte şirketin sahip olduğu ciro ve müşteri memnuniyeti ile veri

Özgün Araştırma Makalesi (Original Research Article)
Geliş/Received: 19.11.2021
Kabul/Accepted: 04.01.2022
DOI: https://dx.doi.org/10.17336/igusbd.1025833

[*] Assoc. Prof. Dr., Istanbul Gelisim University, Faculty of Fine Arts, Istanbul, Türkiye.
E-mail: ssavas@gelisim.edu.tr ORCID https://orcid.org/0000-0003-2141-1055
[**] Res. Assist., Istanbul Gelisim University, Faculty of Economics, Administrative and Social Sciences Istanbul, Türkiye. E-mail: eergen@gelisim.edu.tr ORCID https://orcid.org/0000-0002-8214-0204

toplama eğilimleri arasındaki bağlantının belirsiz olduğu ve sektörel anlamda kurumların elde etmeyi talep ettiği verilerde küçük ölçekli de olsa değişiklik bulunduğu da çalışmanın diğer önemli sonuçları arasında yer almaktadır.

Anahtar Kelimeler: Dijital iz, pasif iz, aktif iz, kişisel veri, hedef kitle

1. Introduction

Digitalisation caused a significant transformation in people's lives. One of the most significant outcomes of digitalisation, at a macro scale, is the issue of privacy. The issue of privacy through the use of the social media applications is being discussed nowadays in general terms, but this discussion should focus particularly on digital footprints. It is safe to assert that the individual, who has to go beyond the boundaries of his/her traditional life style, would be stepping into the uncharted territories of digital world where s/he might face certain challenges. In other words, stepping into the digital world would mean another arena for the individual to protect his/her personal data— that is, his/her digital footprints.

Over the course of their digital lives, individuals leave behind their digital footprints as long as they are online. Whether the individual leaves these trails of information behind intentionally is one thing, but the storage of these data by a third party is another thing. Collecting and storing digital footprints are done for certain purposes and have certain consequences. These purposes and consequences can be viewed from the individual perspective as well as from the institutional vantage point. The personal data, whose collecting made possible via digital footprints, are proven to be quite beneficial for institutions. To be more specific, digital footprints play a significant role in identifying the thoughts, feelings, and behaviours of the consumers and thus in designing products as well as providing services for the consumers.

This study, accordingly, explores corporations' tendency to collect and store the personal data of their customers via the digital footprints that the customers leave behind. By putting 100 dot-com companies operating in Turkey under scrutiny, the study will focus on the following: under which categories companies collect information, industry-specific changes in the tendency of data collection, and the connexion of the collected data with revenue and customer satisfaction rates.

2. Digital Footprint and Digital Privacy of Personal Data

New media, as a medium, emerged as a result of the technological developments that has been occurring at present times. This has made the traces that individuals left behind open to further discussion. As the lives of the people have been digitalised, so have the traces that they left behind. Consequently, these traces have come to be explored in the digital world. The concept of digital footprint, accordingly, emerged as an extension of this development.

Digital footprint can be described as the trail of data that individuals leave behind in the digital world. To put it differently, digital footprints are the data which documents an individual's interactions with digital devices or services (Jungherr, 2018). Alternatively, digital footprints can also be described as the traces generated out of a particular human behaviour (Gerhard & Hepp, 2018, pp.687). Furthermore, digital footprints, in a broader sense, represent the information and data that individuals intentionally or unintentionally create when they are online (Buchanan et al., 2017, pp.275). The comments that a person make on the internet, reading newspapers and/or

blogs online, instant messaging, browsing history, the use of wearable technologies (i.e. smart watches and sensors), accessing films and songs, and attending an online event leave specific trails of data behind (Buchanan et al., 2018, pp.54; Gerhard & Hepp, 2018, pp.687). A condition for existence is a must in order to leave digital footprints behind. To exist means the accumulated experience of having been to a place (Benyon et al., 2014, pp.522). This suggests that digital footprints spread through and accumulate within a particular timeline.

People leave behind traces in every stage of their lives. Similarly, online interactions and activities leave behind digital footprints. The form of the trail of data that is left behind is also important within this context. An individual can leave behind traces knowingly or unknowingly. These traces are called active digital footprints and passive digital footprints. Active digital footprint stands for the web sites that an individual intentionally visit and the social media posts that an individual intentionally shares (Hussein, 2018, pp.153). Passive digital footprint, on the other hand, represents activities and interactions that an individual unknowingly takes part in. Passive digital footprints can be cookies, search history, location information, and so on that emerge due to the daily usage of mobile network (Hengstler, 2011, pp.93; Sürmelioğlu & Seferoğlu, 2019, pp.49; Girardin et al., 2008, pp.37). Therefore, it would be safe to argue that individuals share information about themselves via each online interaction and activity.

In comparing these two types of digital footprints, one can argue that passive digital footprints are much more important than the active ones because individuals have control over the active digital footprints. The individual has the option to share only a part of his/her views in the process of leaving behind active digital footprints. The information shared through the active digital footprints may not be based on true information about the individual, who left these trails of data behind, at all. In the process of passively leaving behind digital footprints, however, the information shared about the individual would be much more substantial. Since the individual is not aware of the fact that s/he is leaving behind trails of data, s/he tends to unveil the thoughts and behaviour that s/he would like to keep to himself/herself. The increased use of smartwatches and smartphones, which track the activities of the individual wearing them, have caused people to leave behind more digital footprints. Whereas some people are aware of the fact that they leave behind certain trails of data, most of the people who use digital devices that track their activities have no idea about the amount of digital footprint they leave behind (Gerhard & Hepp, 2018, pp.683).

The trails of data that the individuals leave behind compose their digital profile. Every bit of data about the individual is stored and used to generate the whole picture of the individual's profile. This picture of the individual, which can also be called the digital shadow, represents what kind of person that individual is. With every piece of data that is added to this picture, the scope of the information obtained about the individual widens and much more data can be collected about the individual's thoughts and behaviour, which s/he thinks to keep to himself/herself.

Digital footprints are associated with acquiring information about people and social phenomena (Sen et al., 2021, pp.399). This begs the question why the digital footprints that belong to individuals are being collected. This process of data collection has certain consequences at both macro and micro scale. Digital footprints are analysed in terms of cyberbullying, surveillance, and data privacy. These analyses usually point out negative aspects of digital footprints. Collecting digital footprints for the purpose of cyberbullying would usually include sending threatening/disturbing messages or mails, sharing a mail that was not meant to be shared with anyone else, creating a website that causes someone embarrassment, publishing a photo or a video to insult someone, constantly ignoring or targeting someone in a chat room, and so on (Campbell, 2005,

pp.70). Tracing digital footprints, therefore, are crucial in order to know how to intimidate and abuse an individual. Within the context of surveillance, collecting an individual's trails of data plays an important role in monitoring and overseeing an individual's online interactions. If one takes Foucault's (1984, pp.206) concept of panopticon (which he borrowed from Jeremy Bentham) into consideration, one would come to realise that storing individuals' digital footprints would allow those individuals to be monitored and controlled by the governing system. In broader terms, digital footprints would mean a threat to an individual's data security. In addition to their thoughts and habits, people share valuable information about themselves online. Amongst these are credentials, date of birth, information about credit cards, and so on (Dülger, 2016, pp.117-118). Leaking these details about the individual would violate the individual's privacy. What's more, it can also give rise to various financial problems.

Thanks to the trails of data an individual leaves behind due to his/her digital activities, s/he might find himself/herself in one of these above-mentioned situations. Since digital footprints can be stored and kept for a long period of time because nothing in the digital world can ever be removed completely and for good, individuals can face these harsh consequences. Everything that a person thinks s/he deleted continues to exist in the digital world. This means that the digitalised data of an individual will outlive that individual.

Storing digital footprints are carried out by certain individuals and groups as well as by governments and corporations. Certain individuals store the digital footprints of others for cyberbullying and fraud, whereas governments for surveillance. Corporations, on the other hand, track and store digital footprints to find out about customer taste. Within the context of this study, identifying target audience and customer taste will be put under scrutiny.

Personal data, as mentioned above, are collected through digital footprints. Little by little, digital footprints in time create the whole picture of an individual. This process allows the information about an individual's potential behaviour to be collected. This prediction of the individuals' possible behaviours bears serious consequences especially in politics and marketing. Therefore, digital privacy of personal data has become important. Furthermore, collecting personal data is proven to have serious consequences for the individual.

3. The Use of Personal Data to Identify Target Audience

The discussions about the use of personal data in today's world are mainly centred around institutions and corporations. Many government agencies and private corporations collect and store personal data in order to make use of them. This process of storage has gone beyond the traditional boundaries and spread into the digital world. Thus storing digital footprints of people has gained further significance. Surely personal data can be collected for various reasons, but for private corporations one of the main reasons for tracking and collecting personal data is to understand the marketing environment. This brings the concept of target audience to the fore.

Target groups or target audience stands for a group of people sharing a common interest (Skodvin et al., 2010, pp.855). In other words, it represents certain groups that share similar expectations and affiliations with a company or institution (Peltekoğlu, 2004, pp.143). Target audience, accordingly, is necessary for the longevity of corporations.

Target audience might sound like a small group of people at first glance, but its scope may expand or narrow down depending on the resources that a corporation has (Bülbül, 2004, pp.63). To put it differently, corporations, based on their resources, may

target a small group of people or a large group of people. Who the target audience is or whether its scope is to be widened or not are determined by the activity carried out by the corporation. Thus knowing the target audience is a permanent necessity.

Target audience governs the resources that the policymakers need (Skodvin et al., 2010, pp.855). The resources can be vote or money. Corporations make products for consumers to buy, therefore; it would be safe to say that the target audience controls the resources. In order to understand how to obtain these resources from the consumers, corporations must know the target audience. It is crucial to understand the values, needs, preferences, and behaviours of this certain group of people to succeed at this endeavour (Zenker, 2009, pp.23).

A corporate's existence depends upon whether the target audience prefers to buy its product and/or service or not (Elden, 2013, pp.367). In order to maintain customer satisfaction, corporations need to provide either the best service or the best product. Therefore knowing the needs and the expectations of the target audience is crucial (Gonzalez, 2019, pp.867). Although the concepts of need and expectation sound similar, these two concepts have different meanings. Consumer needs are divided into three parts: must-be needs, one-dimensional needs, and attractive needs. Must-be needs are that which are fundamental to the product or the service that is offered. Although their presence would not necessarily make the customers feel completely satisfied with the product or the service, lack of them would make the customers extremely unsatisfied. One-dimensional needs are the ones that are demanded by the customers. They are closely associated with customer satisfaction. Attractive needs, which are also called latent needs, are the ones that customers do not realise that they need. Attractive needs ensure customer satisfaction. Furthermore, since latent needs are a complex mixture that includes customer needs and expectations, this means that fulfilling these needs means exceeding all expectations. In other words, latent needs refer to the rather more complex expectations that are located in the subconscious instead of the ones that are conscious, accessible, and specific (Bonfanti, 2016, pp.889-890).

Accordingly, the problem of identifying the needs and the expectations of customers by corporations has gained significance. One of the possible solutions to this problem is to access personal data through tracing digital footprints. It is claimed that traditional approaches not only fall short of identifying customer needs but are also costly. This failure occurs particularly in the identification of the latent needs of customers (Yang, 2013, pp.1333). Because the passive digital footprints are thought to represent the actual reality of the customer, they are believed to be more advantageous in understanding the customer. Thus digital footprints are of great importance in identifying customer needs and ensuring sales success. The trails of data left behind by the individual generate the whole picture of the individual's profile, which includes all kinds of information about the individual. This improves customer predictability. To be more specific, it allows the individual's psychological traits as well as personalities to be predicted (Lambiotte & Kosinski, 2014, pp.1934). These predictions are essential in maintaining customer satisfaction and thus corporations attempt to collect personal data to make such predictions possible. Costumer expectations, however, are influenced by the factors like word of mouth communication, personal needs, previous experiences, and external communication (Tseng & Hung, 2013, 175). That being said, various market segments have various needs and expectations (Matzler et al., 1996, pp.14). Therefore, it is important to know the overall impact of customer needs and expectations. The quality and the use of the collected footprints are also significant.

It has been observed that individuals' tendencies of leaving behind digital footprints have changed due to social, personal, and safety factors (Muhammad et al., 2018). This makes it difficult for corporations to collect digital footprints. The trails of

data left behind by each individual is different and thus unique. When online behavioural tendencies change, so do the advertisements that the individual comes across (Kumar and Raj, 2020, pp.17). Corporations, accordingly, use the data collected through the tracing of digital footprints to make the products and/or services more appealing to the customers. Moreover, these data are also used in advertisements made specifically for a certain individual.

Taking into consideration all that has been discussed so far, corporations collecting personal data via digital footprints can be regarded as something both positive and negative. For the former, the collection of personal data by corporations would be beneficial in making products and services more appealing and satisfactory to the customers. What's more, it also makes the advertisements of the products and the services that the customer wants to buy come his/her way. For the latter, collecting and using personal data violate privacy and can also be used as a tool of manipulation in order to make the customer buy the products that s/he does not actually need.

4. An Analysis of the Corporations' Tendency to Categorically Store Personal Data

4.1 The Problem that the Study Intends to Solve

Corporations collect the personal data of their customers through digital footprints. Nowadays the collection of personal data is no longer a hidden process. Rather, it occurs through the consciousness of the customer. With the emergence of smartphones, the interest in mobile phones has increased and many dot-com companies, accordingly, started to rely on mobile applications in tracing and collecting personal data. Mobile phones being constantly used by people has made corporations trace the footprints of their customers much easier. This also made the intended uses of the digital personal data stored by corporations to be questioned. Besides the very act of collecting personal data, the types of personal data that is most collected are also quite significant within this context. This study, accordingly, explores the collection of personal data via digital footprints in terms of these above-mentioned aspects.

4.2 The Aim of the Study

This study aims to lay bare under which categories companies collect information the most and the industry-specific changes of these categories. In relation to this, the connexion of the tendency of collecting data with revenue and the customer satisfaction rates is also put under scrutiny.

4.3 The Scope of the Study and the Identified Gap in the Scholarship

The scope of the data that will be presented by this study are limited to the dot-com companies located in Turkey. AppStore's classification is used in the categorisation of personal data in the study. The reason why is that AppStore's categorisation of personal data is much more detailed compared to others. Seventy-nine of the companies, which are included in Fast Company's top 100 list, have applications on AppStore. This study, accordingly, focuses on these corporations and their customers as target audience.

The gap that this study aims to fill in the scholarship is that it focuses on the consequences of using the collected personal data rather than whether the collected personal data are used or not. To be more specific, analysing storing and using the digital personal data in terms of the interaction between the customer and the company

(revenue growth and customer satisfaction) will be the unique contribution of this study to the scholarship.

4.4 The Universe and Sample of the Study

The study's universe encompasses all of the dot-com companies operating in Turkey, whereas its sample consists of Fast Company's top 100 list. Because there are no other available lists on the dot-com companies in Turkey, Fast Company's list is used in this study, which conducts 'purposive sampling'. Purposive sampling is conducted when all of the elements in the universe share similar characteristics (Ergin, 1994, pp.93).

4.5 Method of Research

The study implements content analysis. Content analysis means the analysis of the content of the object of study by dividing the content into categories. Because there are no standardised or predetermined categories in content analysis, the categories are formed through the materials of analysis found in every study (Çilingir, 2017, pp.151; Elo & Kyngas, 2007, pp.109). In this study, accordingly, the personal data collected by the corporations are subjected to a certain process of categorisation. As mentioned above, AppStore's classification is used in the categorisation of personal data in this study. The categories formed are as follows:

DATA CATEGORY	CONTENT
Contact Information	Name, address, email address, phone number etc.
Health and Fitness	Medical history: HealthKit, movement disorder, researches with a guinea pig, clinical/general data on health, data on fitness and exercise
Financial Information	Payment methods, account information, salary, income, assets, debts etc.
Location	Information on location in the form of the exact or approximate latitude and longitude
Sensitive Information	Ethnicity, sexual orientation, date of birth, religious belief, union membership, political information etc.
Phone Book	Contact list, directory etc.
User-generated Content	Email and texts, pictures and videos, recordings, contents produced in games, customer support data etc.
Browsing History	Various viewed contents such as web sites, which are not part of an application
Search History	Information on in-app searches
Identifiers	Nicknames, account identity and assigned identity, account defining identities, device id, customer number etc.
Purchase History	Information on purchases made and on purchase intention
Usage Data	Application launching, swipes, data on music/video, product interaction, advertisement etc.
Diagnostics	Time of launch, energy use, performance data, malfunction data etc.
Other Data	Types of data not mentioned above

Source: APPLE (n.d.). Privacy definitions and examples, Date of access: 30.10.2021, https://apps.apple.com/story/id1539235847''

Table 1. Categories of Personal Data

Following the determination of categories, the study examines the categories that the applications of the biggest (in terms of revenue) dot-com companies in Turkey request the data of. Afterwards in the study, the extent and the characteristic of the personal data

required by these 100 companies from their consumers are identified. Through this identification, the meaning of collecting personal data for companies (as well as for the overall industry) and the importance of the tendency to collect personal data in understanding the consumer are explored. Furthermore, the possible consequences of requesting and storing personal data on revenue and customer satisfaction are discussed in the study.

5. Findings

The study includes the analyses of seventy-nine companies from various categories. The industries that these companies are in are classified based on their line of business in table 2.

Category	Overall Rate
Shopping	30
Travel	13
Finance & Business	10
Technology & Software	8
Sports	7
Lifestyle & Entertainment	6
Social Network	5
Total	79

Table 2. A Review of the List on the Basis of Industry

Table 2 shows that thirty companies in the biggest 100 dot-com companies in Turkey list operate in shopping industry. It is followed by the categories of travel, finance & business, technology & software, sports, lifestyle & entertainment, and social network.

Data Category	Overall Rate
Diagnostics	58
Identifiers	56
Contact Information	54
Usage Data	52
User-generated Content	34
Location	33
Financial Information	27
Purchase History	27
Search History	23
Other Data	12
Phone Book	8
Sensitive Information	4
Health and Fitness	3
Browsing History	2

Table 3. Categories in which the Most Data is Collected by the Companies

When one puts under scrutiny the data obtained through the study in terms of the categories of data collection, one realises that there are certain differences between these categories. It is displayed that the category in which the most data is collected by companies is 'diagnostics'. It is followed by the categories of identifiers, contact

information, usage data, user-generated contents, location, financial information, purchase history, search history, other data, phone book, sensitive information, health and fitness, and browsing history. Table 3 shows that the overall rates drop significantly following the category of 'usage data'.

Total Number of Categories Whose Information is Requested	**Company Name**
14	Facebook
13	Amazon
12	LinkedIn
11	Google
10	-
9	Trendyol, Yemeksepeti, Getir, Sahibinden, Çiçekspeeti, Banabi, Bitaksi, Hopi
8	Nesine.com, N11.com, Peak Games, Netflix, Dolap, Yolcu360
7	Bilyoner, Misli, Modanisa, En Uygun, ETA Borusan, Armut, Supplementler, Narcade Games
6	Gittigidiyor, Ebebek, Tuttur, Sefamerve, Obilet, İstegelsin, İyzico, Modacruz, Moov, Gamegos Game, Paynet, Maçkolik, Meditopia
5	Gram Games, Rollic Games, Mynet, Voltlines, Martı, Türk Para, BluTv, Apsiyon
4	İkinciyeni, Tapu.com, Kitapyurdu, PayTR
3	Masomo Games, Good Job Games, OtelZ, Fırsat Bu Fırsat
2	EPTTAvm, Papara, Garajsepeti, Biletall, Connected2me, Tatilbudur, Ucuzabilet, Odamax, Lidyana
1	Hepsiburada, Morhipo, Rentiva
None	Otonet, Oley, Tatilsepeti, Turna, Kitapseç, Nadirkitap, App Samurai
Ones that Provide no Information on the Data Collected	İncehesap, Birebin, MenaPay, Birleşik Ödeme

Table 4. Categories of Data Collection by Companies

Table 4 illustrates that whereas most of the companies tend to collect data from the categories which are ranked average, companies like Facebook, Amazon, LinkedIn, and Google tend to collect data in many categories. Moreover, table 4 displays that while the companies that tend to collect data in many categories are US-based, the Turkey-based companies do not collect information more than in 9 categories in total.

Data Category	Overall Rate
Contact Information	10
Diagnostics	9
Identifiers	9
Usage Data	9
User-generated Content	9
Location	7
Financial Information	7
Purchase History	6
Search History	5
Other Data	4
Phone Book	3
Browsing History	2
Sensitive Information	1
Health and Fitness	1

Table 5. The Most Collected Data Categories by the Top 10 Companies

The information on table 5 is similar to the information on table 4, which details out the whole picture. Whereas the list, from top to down, goes on like this on table 4: Diagnostics, Identifiers, Contact Information, Usage Data, and so on; it goes on like this on table 6: Contact Information, Diagnostics, Identifiers, Usage Data, and so on. The comparison of the aforementioned tables demonstrate that companies find the category of 'user-generated content' quite significant.

Company Name	Total Number of Data Request Categories
Trendyol	9
Hepsiburada	1
Nesine.com	8
N11.com	8
Bilyoner	7
Gittigidiyor	6
Yemeksepeti	9
Google (Youtube)	11
Peak Games (Toon Blast)	8
Facebook	14

Table 6. Largest Companies by Revenue and Data Requests

Table 6 provides accurate and consistent information to explore the connexion between data request and revenue. Although there is not much gap between the revenues of the top 10 companies, the total number of categories of data request differs considerably. What's more, the companies that have the same amount of data request attempts as the ones at the top differs from them considerably in terms of revenue. Hepsiburada is significant in analysing the link between data collection and revenue mainly because of two reasons: (1) it requests the data in only one category and (2) still manages to be ranked second in terms of revenue.

It is also possible to establish an industry-specific approach in the analysis of data request categories. Companies operating in various industries need various personal data. An industry-specific comparison, accordingly, will also be conducted below.

Company Name	**Total Number of Data Request Categories**
Trendyol	9
Hepsiburada	1
N11.com	8
Gittigidiyor	6
Yemeksepeti	9
Getir	9
Sahibinden	9
Ebebek	6
EPTTAvm	2
Modanisa	7
Çiçeksepeti	9
Otonet	0
Morhipo	1
Amazon (Amazon Shopping)	13
Banabi	9
Sefamerve	6
İstegelsin	6
Garaj Sepeti	2
Dolap	8
Rentiva (Garajyeri)	1
Gardrops	0
Tapu.com	4
Modacruz	6
Kitapyurdu	4
Kitapseç	0
Supplementler	7
Nadir Kitap	0
Hopi	9
Fırsat Bu Fırsat	3
Lidyana	2

Table 7. Data Request Numbers in the Category of Shopping (Rank by Revenue)

Amazon, through 'Amazon Shopping' application, is the company with the most requests (in 13 categories) for access to personal data. The companies such as Otonet, Nadir Kitap, and Gardrops are seen to be not requesting any access to any forms of data.

Data Category	Overall Rate
Contact Information	30/23
Diagnostics	30/21
Identifiers	30/20
Usage Data	30/18
Search History	30/15
User-generated Content	30/14
Location	30/13
Purchase History	30/12
Financial Information	30/11
Other Information	30/4
Phone Book	30/3
Health & Fitness	30/1
Sensitive Information	30/1

Table 8. The Type of Data Requested the Most in the Category of Shopping

Table 8 lays bare that the data categories listed here are consistent with the general picture provided so far. However, it is interesting that the category of search history, which is ranked 8th on table 3, is ranked 5th on table 8. When this list is narrowed down on the basis of the top 10 companies in the shopping category, the category of 'search history' appears to go up as far as the 2nd rank with the score of 10/8.

Company Name	Total Number of Data Request Categories
En Uygun	7
Obilet	6
Biletall	2
Bitaksi	9
OtelZ	3
Tatilbudur	2
Tatilsepeti	0
Moov	6
Ucuzabilet	2
Turna	0
Yolcu360	8
Martı	5
Odamax	2

Table 9. Data Request Numbers in the Category of Travel (Rank by Revenue)

'Bitaksi' appears to be the company with the most requests (in thirteen categories) for access to personal data in the category of travel. Companies such as Tatilsepeti and Turna seem to be not requesting any access to any forms of data.

Data Category	Overall Rate
Diagnostics	13/10
Contact Information	13/7
Identifiers	13/7
Usage Data	13/7
Purchase History	13/5
Financial Information	13/4
Location	13/4
Search History	13/4
Other Data	13/2
Sensitive Information	13/1
User-generated Content	13/1

Table 10. The Type of Data Requested the Most in the Category of Travel

Table 10 shows that the data categories listed here are consistent with the general picture provided so far. Although there is not much difference to be observed between the categories in general, it can be seen that the category of 'purchase history' goes up and there are changes to be observed in the order of the first four categories. When this list is narrowed down on the basis of the top 10 companies in the shopping category, there is not much difference to be observed at all.

Company Name	Total Number of Data Request Categories
ETA Borusan	7
Papara	2
İyzico	6
Voltlines	5
Paynet	6
Türk Para (Param)	5
MenaPay	-
PayTR	4
Apsiyon	5
Birleşik Ödeme (Gazmatik)	-

Table 11. Data Request Numbers in the Category of Finance & Business (Rank by Revenue)

'ETA Borusan' appears to be the company with the most requests (in seven categories) for access to personal data in the category of finance & business. Companies like MenaPay and Birleşik Ödeme do not appear to be sharing any information related to data categories.

Data Category	Overall Rate
Contact Information	10/8
Identifiers	10/8
Usage Data	10/6
User-generated Content	10/5
Diagnostics	10/5
Financial Information	10/4
Location	10/3
Phone Book	10/1

Table 12. The Type of Data Requested the Most in the Category of Finance & Business

Table 12 illustrates that the data categories listed here are consistent with the general picture provided so far. However, it can be observed that whereas the category of 'user-generated content' goes up, the category of 'diagnostics' does down when comparing table 12 with the general picture. Another striking thing about table 12 is that it displays that there are no data collected in the categories of health and fitness, sensitive information, browsing history, search history, purchase history, and other data.

Company Name	Total Number of Data Request Categories
Peak Games (Toon Blast)	8
Gram Games (Merge Dragons)	5
Rollic Games (Hair Challenge)	5
Masomo (Kafa Topu 2)	3
Good Job – Games (Fun Race 3D.)	3
Gamegos Game (Cafeland - World Kitchen)	6
App Samurai (Storyly)	0
Narcade Games (Trick Me: Logical Brain Teaser)	7

Table 13. Data Request Numbers in the Category of Technology & Software (Rank by Revenue)

'Peak Games' appears to be the company with the most requests (in eight categories) for access to personal data in the category of technology & software. App Samurai seems to be not requesting any access to any forms of data.

Data Category	Overall Rate
Identifiers	8/7
Usage Data	8/7
Diagnostics	8/7
Purchase History	8/5
User-generated Content	8/4
Location	8/3
Other Data	8/2
Contact Information	8/1
Phone Book	8/1

Table 14. The Type of Data Requested the Most in the Category of Technology & Software

Table 14 demonstrates that the data categories listed here are consistent with the general picture provided so far. However, it could be seen that the category of 'purchase history' goes up when compared to the general picture provided so far. Table 14 also shows that technology & software companies do not collect data in the categories of health and fitness, financial information, sensitive information, browsing history, and search history.

Company Name	Total Number of Data Request Categories
Nesine.com	8
Bilyoner	7
Misli	7
Tuttur	6
Oley	0
Birebin	-
Maçkolik	6

Table 15. Data Request Numbers in the Category of Sports (Rank by Revenue**)**

'Nesine.com' is the company with the most requests (in eight categories) for access to personal data in the category of sports. Oley seems to be not requesting any access to any forms of data.

Data Category	Overall Rate
Contact Information	7/5
User-generated Content	7/5
Identifiers	7/5
Usage Data	7/5
Diagnostics	7/5
Financial Information	7/4
Location	7/4
Purchase History	7/1

Table 16. The Type of Data Requested the Most in the Category of Sports

Table 16 displays that the data categories listed here are consistent with the general picture provided so far. However, it can be observed that the category of 'user-generated content' rises up in terms of rank when compared to the general picture provided so far. Table 16 also illustrates that sports companies do not collect data in the following categories: health and fitness, sensitive information, phone book, browsing history, search history, and other data.

Company Name	Total Number of Data Request Categories
Netflix	8
İkinciyeni	4
İncehesap	-
Armut	7
BluTV	5
Meditopia	6

Table 17. Data Request Numbers in the Category of Lifestyle & Entertainment (Rank by Revenue)

'Netflix' is the company with the most requests (in eight categories) for access to personal data in the category of lifestyle & entertainment. There is no company in the category of lifestyle & entertainment that does not collect personal data.

Data Category	Overall Rate
Contact Information	6/5
Identifiers	6/5
Usage Data	6/5
Identifiers	6/5
Location	6/2
User-generated Content	6/2
Purchase History	6/2
Health and Fitness	6/1
Financial Information	6/1
Search History	6/1
Other Data	6/1

Table 18. The Type of Data Requested the Most in the Category of Lifestyle & Entertainment

Table 18 demonstrates that the data categories listed here are consistent with the general picture provided so far. Table 18 also shows that lifestyle & entertainment companies do not collect data in the following categories: sensitive information, phone book, and browsing history.

Company Name	Total Number of Data Request Categories
Google (Youtube)	11
Facebook	14
LinkedIn	12
Mynet	5
Connected2me	2

Table 19. Data Request Numbers in the Category of Social Network (Rank by Revenue)

'Facebook' is the company with the most requests (in fourteen categories) for access to personal data in the category of social network. There is no company in the category of social network that does not collect personal data.

Data Category	Overall Rate
Contact Information	5
Diagnostics	5
Location	4
Identifiers	4
Usage Data	4
Financial Information	3
Phone Book	3
User-generated Content	3
Search History	3
Other Data	3
Sensitive Information	2
Browsing History	2
Purchase History	2
Health and Fitness	1

Table 20. The Type of Data Requested the Most in the Category of Social Network

Table 20 illustrates that the data categories listed here are consistent with the general picture provided so far. However, it can be seen that the category of 'location' goes up in terms of rank when compared to the general picture so far provided. It is also significant that social network companies collect data in each data category.

The tendency to collect data can also be analysed through customer satisfaction. It is common knowledge that companies access personal data in order to provide the products and the services that would appeal to the overall taste of the customer. Accordingly, the tendencies of the companies (with at least 1.000 users) that scored at least 4.5 out of 5 to collect data are identified in order to explore the connexion between data collection and customer satisfaction.

Company Name	Total Number of Data Request Categories	Score and Number of Participants
Trendyol	9	4.6 • 2.2M
Hepsiburada	1	4.7 • 754.1K
Nesine.com	8	4,5 • 169,3K
N11.com	8	4,7 • 570,9K
Gittigidiyor	6	4.7 • 292.7K
Yemeksepeti	9	4.7 • 1.4M
Google (Youtube)	11	4,5 • 1,7B
Peak Games (Toon Blast)	8	4.7 • 62.1K
Getir	9	4.8 • 470.9K
Sahibinden	9	4.7 • 1.7M
Gram Games (Merge Dragons)	5	4.8 • 5.9K
Modanisa	7	4,6 • 21,9K
Çiçeksepeti	9	4.6 • 459K
En Uygun	7	4.8 • 46.7K
Morhipo	1	4.6 • 65K
Amazon (Amazon Shopping)	13	4.6 • 16.5K
Banabi	9	4.7 • 1.4M

Obilet	6	4,7 • 49,5K
Papara	2	4,8 • 159,7K
Armut	7	4.7 • 45.6K
Dolap	8	4.6 • 209.8K
Biletall	2	4,6 • 4,5K
Mynet	5	4.5 • 37.6K
Bitaksi	9	4.8 • 169.2K
Gardrops	0	4,7 • 166,6K
Modacruz	6	4,6 • 58,9K
Yolcu360	8	4,7 • 6,9K
Martı	5	4,7 • 53,7K
Gamegos Game (Cafeland - World Kitchen)	6	4.5 • 10.3K
Supplementler	7	4,8 • 17,4K
BluTV	5	4,5 • 105,7K
Hopi	9	4,6 • 19,7K
Lidyana	2	4.6 • 2.7K
Meditopia	6	4.8 • 62.2K

Table 21. A Comparison of Companies' Scores (4.5+) and Data Category Numbers (Rank by Revenue)

Table 21 lays bare that there is ambiguity in terms of the link between data collection and customer satisfaction. When the scores that the users give are taken into consideration, both the companies with the highest and the lowest number of data request categories enjoy considerably high rates of participation in customer satisfaction surveys. The companies such as Hepsiburada, Morhipo, Papara, Biletall, Gardrops, and Lidyana, which collect either a very little amount of data or not at all, cause the link between data collection and customer satisfaction to be questioned. Moreover, many companies on the list appear to collect data at an average level.

Data Category	**Overall Rate**
Contact Information	34/30
Identifiers	34/29
Diagnostics	34/29
Usage Data	34/27
Purchase History	34/22
Location	34/19
User-generated Content	34/19
Search History	34/17
Financial Information	34/16
Other Data	34/5
Phone Book	34/4
Health and Fitness	34/2
Sensitive Information	34/2
Browsing History	34/1

Table 22. Data Request Numbers of the Companies with the Highest Score

Table 22 shows that the data categories collected are consistent with the general picture provided so far. There are little changes to be observed in the order of a few

categories, whereas the category of 'purchase history' seems to considerably go up in rank. The data provided in the category of 'purchase history' appear to be an important category for the companies that enjoy the highest rates of participation in customer satisfaction surveys. This begs the question whether the data category of 'purchase history' has anything to do with customer satisfaction or not.

6. Conclusion

The data collected by companies can be approached from two different vantage points: individual and institutional. Whereas individuals regard this process of collecting data as an issue of privacy, institutions view it as an attempt of understanding the customer. However, when the tendency of data collection by corporations is taken into consideration, the total amount of data collected outweighs the intended use of these data. The process of data collection by companies contains various categories of personal data.

According to the results of the study, the category of 'diagnostics' is the type of data requested the most by the companies. The category of diagnostics is related to performance and malfunction data of the product and/or the device that is offered by a company. Particularly during the coronavirus outbreak, accordingly, technology & software companies experienced revenue growth. The second in rank is the category of 'identifiers', which means that corporations are quite interested in their customers' account identity, customer number, and so on. It also appears that most of the companies tend to collect data in the categories which are ranked average. Only one company collects data in all of the categories. The corporations that collect data in many categories are observed to be US-based or to operate on a global scale.

The study also suggests that there is simply not enough data to explore the link between the companies' tendency to collect data and revenue. While the general picture illustrates that all of the top ten companies, with the exception of Hepsiburada, collect data in more than six categories, there is still not enough data to work on the connexion between revenue and data collection. The study furthermore states that regardless of the amount of revenue, all corporations tend to collect data.

Application owners on AppStore have the right to not to provide information about the data they access. The study shows that the less the revenue is, the more the tendency to not provide information on data collection. To illustrate, the first company on the lists that does not share information on data collection is İncehesap, which is ranked last on the list. İncehesap is followed by Birebin, MenaPay, and Birleşik Ödeme, which likewise do not provide any information on the data collected. A company not sharing any information on data collection might have the potential to make that company look unreliable for customers. Thus, the companies with the highest revenue provide exact and detailed information on data collection. The negative response of the customers to the companies that do not provide information on data collection is manifested in customer satisfaction surveys. The companies that do not share information on the data collected (except for MenaPay) suffer in terms of both the total number of votes and the overall score: İncehesap with 2.8/117, Birebin with 2.5/117, MenaPay with 4.5/54, and Birleşik Ödeme with 1.7/24.

It is also important to notice that eight companies do not collect data through their applications on Apple Store. An interesting fact is that four of these are shopping companies. Amongst the hundred companies on the list, only three are applications on selling books. Two of these companies, strikingly, do not collect any data. The scores that the companies which do not collect data receive from customers appear to be high. One striking company in terms of that is Gardrops with 4.7/166.6. Therefore not collecting

data does not mean that the potential to manifest the products and/or to provide the services that satisfy the customer would diminish.

When the data provided by this study are analysed from the vantage point of an industry-specific approach, the sectoral data appear to be consistent with the general picture. The tendency to collect data increases in terms of the category of 'search history' in the shopping category, 'purchase history' in the category of travel, 'user-generated content' in the category of finance & business, 'purchase history' in the category of technology & software, 'user-generated content' in the category of sports, and 'location' in the category of social network.

Last but not least, the study puts forward that some of the categories of data collection are regarded unimportant by the corporations. Two of these are the categories of health and fitness and sensitive information. There are only three dot-com companies (Facebook, Amazon, and Mediopia) that collect data on health and fitness. Similarly, there are only four dot-com companies (Facebook, Amazon, LinkedIn, and Bitaksi) that collect data on sensitive information. Browsing history is the other category whose data is not collected by the companies.

Taking into consideration the findings of the study, it is safe to say that the advantages that the tendency to collect data and the scale of the data collected provide in understanding the customer is not clear. The findings of the study suggest that there is no linear-by-linear association between the tendency to collect data, the scores that the companies receive in surveys (except for the category of purchase history), and the revenue. Whereas some of the companies that enjoy the highest rates of participation in customer satisfaction surveys or the largest ones by revenue request high amounts of data, the others request very little amount of data. Companies tend to identify the target audience in order to offer products or services that appeal to the customer. Moreover, the tendency to collect data has certain consequences on companies as its repercussions can be observed on customer satisfaction surveys. All in all, it is impossible to talk about the whole picture as some parts of the puzzle are still missing. Hence, the companies' tendency to collect data should be analysed in a much more comprehensive manner.

REFERENCES

APPLE (n.d). Privacy definitions and examples, Date of access: 30.10.2021, https://apps.apple.com/story/id1539235847

BENYON, D., QUIGLEY, A., O'KEEFE, B. & RIVA, G. (2014). Presence and digital tourism. *AI & Society, 29(4)*, 521-529.

BONFANTI, A. (2016). Customers' needs and expectations on servicescape surveillance management. *The TQM Journal, 28(6)*, 887-906.

BUCHANAN, R., SOUTHGATE, E., SMITH, S. P., MURRAY, T. & NOBLE, B. (2017). Post no photos, leave no trace: children's digital footprint management strategies. E-*Learning and Digital Media, 14(5)*, 275-290.

BUCHANAN, R., SOUTHGATE, E., SCEVAK, J. & SMITH, S. P. (2018). Expert insights into education for positive digital footprint development. *Scan: The Journal for Educators, 37(2)*, 49-64.

BÜLBÜL, R. (2004). *Halkla ilişkiler*. Ankara: Nobel Yayınları.

CAMPBELL, M. A. (2005). Cyber bullying: an old problem in a new guise?. *Journal of Psychologists and Counsellors in Schools, 15(1)*, 68-76.

ÇİLİNGİR, A. (2017). İletişim alanında içerik analizi yöntemi kullanılarak yapılan yüksek lisans ve doktora tezleri üzerine bir inceleme. *Erciyes İletişim Dergisi, 5(1)*, 148-160.

DÜLGER, M. V. (2016). Kişisel verilerin korunması kanunu ve türk ceza kanunu bağlamında kişisel verilerin ceza normlarıyla korunması. *İstanbul Medipol Üniversitesi Hukuk Fakültesi Dergisi, 3(2)*, 101-168.

ELDEN, M. (2013). *Reklam ve reklamcılık*. İstanbul: Say Yayınları.

ELO, S., KYNGÄS, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62(1)*, 107-115.

ERGİN, D. Y. (1994). Örnekleme türleri. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi, 6(6)*, 91-102.

FASTCOMPANY (2021). En büyük 100 internet şirketi, Date of access: 30.10.2021, https://fastcompany.com.tr/calisma-hayati/en-buyuk-100-internet-sirketi/

FOUCAULT, M. (1984). Panopticism, in the foucault reader. New York: Pantheon Books.

GERHARD, U., HEPP, A. (2018). Digital traces in context| appropriating digital traces of self-quantification: contextualizing pragmatic and enthusiast self-trackers. *International Journal of Communication, 12*, 683-700.

GIRARDIN, F., CALABRESE, F., DAL FIORE, F., RATTI, C. & BLAT, J. (2008). Digital footprinting: uncovering tourists with user-generated content. *IEEE Pervasive computing, 7(4)*, 36-43.

GONZALEZ, M. E. (2019). Improving customer satisfaction of a healthcare facility: reading the customers' needs. *Benchmarking: An International Journal, 26(3)*, 854-870.

HENGSTLER, J. (2011). Managing your digital footprint: ostriches v. eagles. *Education For A Digital World, 2.0*, 89-139.

HUSSEIN, H. A. (2018). A passive digital footprint for extranet environment. *AL-Qadisiyah Journal of pure Science, 23(1)*, 152-158.

JUNGHERR, A. (2018). Normalizing digital trace data. Natalie Jomini Stroud, Shannon Mcgregor (Ed.) *In Digital discussions (ss. 9-35)*. Routledge.

KUMAR, H., RAJ, P. (2020). An indagation on experiences and awareness of digital footprint among pupils of higher education. *Academic Research International*, 11(3), 16-31.

LAMBIOTTE, R., KOSINSKI, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE, 102(12)*, 1934-1939.

MATZLER, K., HINTERHUBER, H. H., BAILOM, F. & SAUERWEIN, E. (1996). How to delight your customers. *Journal of Product & Brand Management, 5(2)*, 6-18.

MUHAMMAD, S. S., DEY, B. L. & WEERAKKODY, V. (2018). Analysis of factors that influence customers' willingness to leave big data digital footprints on social media: a systematic review of literature. *Information Systems Frontiers, 20(3)*, 559-576.

PELTEKOĞLU, F. B. (2004). Halkla ilişkiler nedir?. İstanbul: Beta Yayınları.

SEN, I. & FLÖCK, F., WELLER, K., WEIß, B. & WAGNER, C. (2021). A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly, 85(S1)*, 399-422.

SKODVIN, T., GULLBERG, A. T. & AAKRE, S. (2010). Target-Group influence and political feasibility: the case of climate policy design in europe. *Journal of European Public Policy, 17(6)*, 854-873.

SURMELIOGLU, Y., SEFEROGLU, S. S. (2019). An examination of digital footprint awareness and digital experiences of higher education students. *World Journal on Educational Technology: Current Issues, 11(1)*, 48-64.

TSENG, S. C., HUNG, S. W. (2013). A framework identifying the gaps between customers' expectations and their perceptions in green products. *Journal of Cleaner Production, 59*, 174-184.

YANG, C. C. (2013). An analytical methodology for identifying the latent needs of customers. *Total Quality Management & Business Excellence, 24(11-12)*, 1332-1346.

ZENKER, S. (2009). Who's your target? The creative class as a target group for place branding. *Journal of Place Management and Development, 2(1)*, 23-32.

Özet

Bu çalışma, bireylerin dijital ortamda bırakmış oldukları izlerin depolanması ve depolanan izlerin kullanım amaçlarına irdelemektedir. Günümüzde dijital izlerin pek çok farklı alanda kullanıldığı bilinmektedir. Özellikle kurumsal açıdan konu değerlendirildiğinde kurumların bu izleri paydaşlarını ve paydaşları özelinde tüketicilerini tanımak için depoladıkları düşünülmektedir. Bu doğrultuda bu araştırma da kurumların bireyler hakkında depoladığı kişisel veri tercihlerinin hangi kategorilerde yoğunlaştığı ve ilgili kategorilerin sektörel anlamda değişimlerini ortaya koymayı amaçlamaktadır. Araştırmada Fast Company Dergisi tarafından gerçekleştirilmiş index referans alınarak Türkiye'de faaliyet gösteren en büyük yüz internet şirketi değerlendirmeye alınmıştır. İlgili kurumların dijital izler yardımıyla kişisel veri toplama eğilimleri çeşitli veri kategorileri üzerinden tespit edilmiş ve veri toplama eğiliminin hem kurumsal hem de sektörel açıdan yansımaları irdelenmiştir.

Araştırma sonuçlarına göre kurumlar tarafından en çok toplanan veri kategorileri göz önüne alındığında "tanılar" kategorisinin birinci sırada yer aldığı görülmektedir. Bu kategoriyi sırasıyla kimlik tanıtıcılar, iletişim bilgileri ve kullanım verileri kategorilerinin takip ettiği söylenebilmektedir. Kurumların veri toplama eğilimleriyle ciroları arasındaki bağlantı incelendiğinde ise bu tür bir bağlantı kurabilme aşamasında yeterli veri olmadığı görülmektedir. Cirosu hem düşük hem de yüksek olan kurumların birbiriyle benzer olmayan bir şekilde veri topladığı görülmektedir. Ancak araştırma kapsamında özellikle kurumun cirosu düştükçe veri toplama hakkında bilgi vermeme eğiliminde artış görülmeye başlandığı da belirtilebilmektedir. Bu doğrultuda ilk bilgi vermeyen kurumun otuz altıncı sırada yer aldığı görülmektedir. Araştırma kapsamında veri toplama işlemi hakkında bilgi vermeyen kurumların yanında veri toplamayan kurumlara da bir parantez açmak gerekmektedir. İlgili listede toplam sekiz adet kurumun veri toplamadığı görülmektedir. Araştırma sonuçları sektörel açıdan incelendiğinde, genel olarak sektörel verilerin genel tablo ile uyumlu olduğu görülse de sektörel anlamda bazı veri kategorilerinin de öne çıktığı söylenebilmektedir. Dolayısıyla kurumların hangi veri kategorilerine daha fazla önem verdiğinin ilgili kurumun hangi sektörde bulunduğuyla bir bağlantısı olduğunu söylemek mümkündür. Araştırma kapsamında değinilmesi gereken bir diğer nokta ise özellikle bazı veri kategorilerinin kurumlar tarafından önemli görülmediğidir. Bu doğrultuda "sağlık ve fitness", "hassas bilgiler" ve "göz atma geçmişi" gibi kategorilerinin oldukça az toplanan veri türleri olduğu söylenebilmektedir. Araştırma sonuçları göz önüne alındığında veri toplama eğilimi veya tüketiciler hakkında toplanan verilerinin boyutunun müşteriyi tanıma anlamında kesin çıkarımlar sağlamadığı söylenebilmektedir. Kurumların veri toplama eğilimleri ile tüketicilerinden aldıkları oylar veya ciroları değerlendirildiğinde doğrusal bir bağlantının yokluğu göz çarpmaktadır. Sonuçlar, yüksek düzeyde ciroya veya yüksek oranda beğenilme düzeyine sahip kurumların bazılarının yüksek oranda veri talep ederken bazılarının çok az veya hiç veri talep etmediğini göstermektedir. Bu anlamda dijital izlerin depolanması ve tüketiciyi tanıma anlamında kullanımının yansımalarının tüketici memnuniyeti yaratma açısından tartışılması gerektiği söylenebilmektedir.