

1

IS DATA A SOURCE OF COMPETITION OR CONCERN IN THE BRAZILIAN RETAIL MARKET?¹

Dados são uma fonte de competição ou preocupação no mercado varejista brasileiro?

Bernardo Gouthier Macedo²

LCA Consultores – São Paulo/SP, Brasil

Anna Olímpia de Moura Leite³

LCA Consultores – São Paulo/SP, Brasil

Paulo Henrique Oliveira⁴

LCA Consultores – São Paulo/SP, Brasil

STRUCTURED ABSTRACT

Context: businesses have always used data to make decisions, but the amount, variety, and speed of data have increased dramatically in recent years. This fact led to the emergence of big data, which in turn intensified innovation, led to new services and products such as real-time traffic information and personalized marketing and delivery services. While big data has led to efficiency gains, it has also sparked a new debate in antitrust: to what extent, if any, could data become a competitive concern?

Objective: this article is a case study on the Brazilian retail market based on recent literature on barriers to entry fostered by data-based activities and the importance of data as an asset capable of creating market power and giving rise to anti-competitive practices.

Method: a bibliographical review was carried out regarding the importance of data as a possible asset that creates barriers to entry, as well as the interpretation of the Brazilian antitrust authority – Conselho Administrativo de Defesa Econômica (Cade), in recent mergers, in addition to an exploratory analysis of data on access and revenue of the main retail players in different market segments and business types.

Editor responsável: Prof. Dr. Victor Oliveira Fernandes, Conselho Administrativo de Defesa Econômica (Cade), Brasília, DF, Brasil. **Lattes:** <http://lattes.cnpq.br/5250274768971874>. **ORCID:** <https://orcid.org/0000-0001-5431-4142>.

1 **Recebido em:** 20/02/2024 **Aceito em:** 28/05/2024 **Publicado em:** 19/06/2024

2 Mestre e doutor em economia pela Unicamp. Sócio-Diretor da LCA Consultores. Conselheiro do IBRAC. Foi diretor de economia do IBRAC. Sediado em São Paulo, SP, Brasil.

E-mail: bernardo.macedo@lcaconsultores.com **ORCID:** <https://orcid.org/0000-0002-9891-6038>

3 Mestre e graduação em economia pela Faculdade de Economia e Administração da USP. Gerente da LCA Consultores. Coordenadora de Conduas Unilaterais do Comitê de Concorrência do IBRAC desde 2022. Sediada em São Paulo, SP, Brasil.

E-mail: anna.olimpia@lcaconsultores.com **Lattes:** <http://lattes.cnpq.br/6799293897895906>

ORCID: <https://orcid.org/0009-0000-6909-2562>

4 Mestrando em Administração Pública e Governo na FGV-EAESP e graduado em Ciências Econômicas pela UNIFESP. Pesquisador associado ao Observatório do Poder Econômico (OPE), membro do Comitê de Mercados Digitais do IBRAC e economista sênior da LCA Consultores. Sediado em São Paulo, SP, Brasil.

E-mail: paulo.oliveira@lcaconsultores.com **Lattes:** <http://lattes.cnpq.br/7317015365816360>

ORCID: <https://orcid.org/0009-0002-9976-8695>



Conclusions: the article concludes that, taking into account the case study, it is not evident that big data leads to highly competitive advantages in the retail sector in Brazil and that competitive concerns do not extend to all data-intensive markets, as data can be segmented and provide different competitive advantages that do not imply relevant barriers to the Brazilian retail sector.

Keywords: data-driven markets; competition; retail; consumer behavior; multihoming.

RESUMO ESTRUTURADO

Contexto: as empresas sempre utilizaram dados para tomar decisões, mas a quantidade, a variedade e a velocidade dos dados aumentaram drasticamente nos últimos anos. Este fato conduziu ao aparecimento da big data, que por sua vez intensificou a inovação, levou a novos serviços e a produtos como a informação de tráfego em tempo real e a serviços personalizados de marketing e entrega. Embora os grandes volumes de dados tenham conduzido a ganhos de eficiência, também suscitaram um novo debate no domínio antitruste: até que ponto podem os dados tornarem-se uma preocupação concorrencial?

Objetivo: este artigo é um estudo de caso sobre o mercado varejista brasileiro a partir da literatura recente sobre barreiras à entrada fomentado por atividades baseadas em dados e a importância dos dados como um ativo capaz de criar poder de mercado e ensejar práticas anticompetitivas.

Método: foi realizado uma revisão bibliográfica acerca da importância dos dados como possível ativo criador de barreiras à entrada, bem como a interpretação da autoridade antitruste brasileira – Conselho Administrativo de Defesa Econômica (Cade), em atos de concentração recentes, além de análise exploratória dos dados de acesso e de receita dos principais players varejistas em diferentes segmentos de mercado e modalidade de negócios.

Conclusões: o artigo conclui que, tendo em conta o estudo de caso, não é evidente que a big data conduza a vantagens altamente competitivas no setor do varejo no Brasil e que as preocupações concorrenciais não se estendem para todos os mercados intensivos em utilização de dados, pois os dados podem ser segmentados e proporcionam diferentes vantagens competitivas que não implicam em relevantes barreiras à no setor varejista brasileiro.

Palavras-chave: mercados baseados em dados; competição; varejo; comportamento do consumidor; multihoming.

Classificação JEL: L11; L4; L81.

Summary: 1. Introduction; 2. Big data: characteristics, efficiency gains and competitive concerns; 2.1. Literature review; 2.2. Overview of the Brazilian jurisdiction; 3. Access and use of data in the Brazilian retail market; 3.1. Collection of data; 3.2. Usage of data; 4. Final remarks: the multiple factors determining competition in the Brazilian retail market; References; Appendix.

1. INTRODUCTION

Firms have long used data to make strategic decisions. However, the volume, variety and velocity of data have increased dramatically in recent years. This has led to the emergence of big data (MONTEIRO, 2017)⁵.

Big data has provided innovation through new services and products such as real-time traffic information, personalized marketing, delivery services etc. Nevertheless, despite the efficiency gains linked to its usage, it has sparked a new debate in antitrust: could data become a competitive concern? If so, to what extent? Some academics and policy makers understand that big data, due to some of its characteristics, could lead to an unfair competitive advantage and harm consumers (FURMAN, 2019). Contrarily, Andres Lerner, as detailed in Sokol and Comerford (2016), argues that real-world evidence may not support these concerns. The author argues that feedback loops in big data aren't as impactful as often believed, and the collection of online data can lead to improved services at lower costs. Lerner also contends that no single firm dominates data control, and without clear evidence of anticompetitive effects, rigorous antitrust actions could hinder competition and innovation. This debate becomes even more complex when it is considered that the importance and ability to generate barriers to entry through the use of big data can vary significantly from market to market.

This paper studies the context of the Brazilian retail market, an especially interesting case of an industry which has shown an important growth trend and with a large plurality of players that use data in the most diverse ways. This is a complex market in which some players are marketplaces (Magalu, Mercado Livre, Americanas, Netshoes, Amazon), while others are not (Assaí, Boticário, Renner); some are moving into physical stores, while others are investing in omnichannel strategies. Thus, the Brazilian market presents a special context of coexistence of large and small players and distinct business models that operate in a fiercely competitive environment especially fascinating for the analysis of antitrust tools in digital markets (RANKING..., 2023a; RANKING..., 2023b)⁶.

To address these points, this study is organized into three sections, in addition to this Introduction. Section 2 describes the recent discussion about the particularities of big data, their efficiency gains, and potential competition concerns, and how the Brazilian antitrust authority has been considering big data in its analyses. It presents the research held by Rubinfeld and Gal (2017), in which the authors investigate the characteristics and barriers to entry related to data at all stages of its value chain, showing that these barriers generate varying impacts based on the significance of data in each individual market. Therefore, data-driven markets should not be considered as a single group, but rather as a collection of sub-groups with different characteristics. Section 3, in turn, analyzes the diverse data sources available to retailers in Brazil and its uses in different markets. It illustrates that companies possess a wide range of information resources and employ different technologies to use

5 "The term big data reflects the tendency to collect, acquire, storage, and process big volumes of digital data to create economic value" (MONTEIRO, 2017, p. 9).

6 In 2021, the largest retailers in Brazil in terms of revenue were: Carrefour (R\$ 77,75 bi), Pão de Açúcar (R\$ 51.29 bi), Assaí (R\$ 41.9 bi), Magazine Luiza (R\$ 35.28 bi), Via Varejo (R\$ 30.9 bi), Raia Drogasil (R\$ 24.13 bi), Lojas Americanas (R\$ 22.7 bi), Mercado Livre (R\$ 21.09 bi), Big (R\$ 20.12 bi), Grupo Mateus (R\$ 15.88 bi), Drogaria DPSP (R\$ 11.17 bi), Havan (R\$ 9.56 bi), Cencosud (R\$ 8.15 bi), Farmácias Pague Menos (R\$ 7.53 bi) and Grupo Saga (R\$ 6.83). In terms of the most visited shopping websites in May 2023, the top five were Mercado Livre, Amazon, Magalu, OLX and Shopee. The top five most used shopping apps, between January and June 12th, 2023, were Mercado Livre, Shopee, Shein, AliExpress, OLX, Magalu, Amazon, Americanas, Casas Bahia, Lojas Renner, Netshoes, Loja Online Samsung, O Boticário, Riachuelo, Enjoei, Dafiti, and Centauro.



data according to the needs presented in each sector. Finally, the concluding remarks discuss factors that play an important role in shaping competition in the Brazilian retail market.

2. BIG DATA: CHARACTERISTICS, EFFICIENCY GAINS AND COMPETITIVE CONCERNS

2.1. Literature review

The effects of big data in the competition environment are a complex and multifaceted topic. Rubinfeld and Gal (2017) note that big data's capacity of compiling great amounts of relevant information can help firms to create and/or improve their products and services, which can lead to lower prices and higher diversity and quality for consumers. The authors point out, however, that some scholars have raised concerns, although without empirical evidence to support this claim, that it can also be used to engage in anti-competitive behavior, such as artificially increasing barriers to entry. Others, such as Sokol and Comerford (2016), attest that the overall effect of big data on competition depends on the specific circumstances and the particularities of each market⁷.

In the first sense, some studies have claimed, based on theoretical assumptions, that markets in which data is a valuable asset or input can become highly concentrated and lead to anti-competitive practices. This is the case of the OECD (2015) report *Data-driven innovation: Big Data for Growth and Well-Being* and studies like Furman (2019). These studies conclude that data-driven markets are often characterized by a positive feedback loop, in which incumbents have a competitive advantage over entrants by gaining more access to data, which leads to higher quality of products and services, which attracts more users, which leads to more data, and so on. The feedback loop can perpetuate itself over time, leading to concentration of market power in the hands of a few large firms.

Other studies – also based on theoretical assumptions – have focused on the characteristics of data-driven indirect network effects, concluding that markets with this kind of features tend to tip towards monopoly. Prüfer and Schottmüller (2017) highlight that data-driven indirect network effects are the main competitive advantage of first movers, which cause tipping effects to be persistent. In their words, data-driven markets are “[...] markets where the marginal cost of quality production is decreasing in the amount of machine-generated data about user preferences or characteristics (henceforth: user information)” (PRÜFER; SCHOTTMÜLLER, 2017). In the authors' economic model, these network effects are expressed by the decreasing marginal costs of innovation that are driven by user demand.

Rubinfeld and Gal (2017), however, challenge the consistency of the body of studies that follow in this direction. They argue that they lack in-depth economic analysis and fail to define big data markets holistically, without considering the particularities of each subgroup of data markets. The authors also state that the existence of high barriers to entry in data-driven markets does not necessarily mean that potential harm is expected:

⁷ Sokol and Comerford (2016, p. 1141) emphasizes the need to understand fully the ways online platforms use big data, particularly in the context of two-sided platforms. For instance, while social media platforms offer free access to users on one side, they generate revenue from advertising services to businesses on the other. Furthermore, the author argues that a comprehensive antitrust analysis should consider competition on each side of such platforms and assess the overall benefits and harms to both sides of the platform.

[...] due to the unique characteristics of at least some big data markets, the mere existence of high entry barriers into these markets, by itself, does not automatically lead to the conclusion that social welfare will be harmed. (RUBINFELD; GAL, 2017, p. 34).

Another analytical perspective is focused on the non-rivalrous and non-exclusive⁸ character of data, meaning that it can be replicated by multiple firms, thus making it easier for new companies to enter data-driven markets as they can produce or purchase their own data (TUCKER; WELLFORD, 2014; SOKOL; COMERFORD, 2016). In the same sense, collecting a piece of information does not prevent other companies from collecting the same data themselves. For example, if a user is searching for a dress in different apps – say a social media platform, a search engine, and a marketplace – then all of these apps will have access to similar information about that user's purchasing preferences. The replicability of data is confirmed by the statistics provided by Statista, that estimates that, in 2020, 90% of the global data was replicated and only 10% was unique (STATISTA, 2020)⁹.

Additionally, as pointed by Hagiu and Wright (2020), data has a limited lifespan and its value decreases over time. Therefore, should the data lose its relevance swiftly, all else being equal, it facilitates market entry for competitors, as they aren't required to rival the incumbent's extended period of data-driven learning. Furthermore, the authors point out that the rate at which the value of the collected data deteriorates is specific to each market and application.

Indeed, old data becomes less useful as consumers change their habits and preferences. This means that entrants can enter the market and compete with incumbents as newer data is more valuable. Bilbao-Osorio, Dutta and Lanvin (2014, p. 35) points out that 90% of the available data were created in the last two years, showing that technological advances mean that data is being produced at an ever-increasing rate, effectively reducing the relevance of data produced in an earlier period as it is several levels smaller than recent production.

It is also important to consider the additional benefits derived from additional data. Some scholars recognize that the benefits of data analysis eventually diminish as the accumulated volume increases. Hagiu and Wright (2020) and Iansiti (2021) discuss a series of factors that limit the value and competitive differentiation that data access can generate. Among these factors, economies of scale, economies of scope and data specificity are important variables.

Economies of scale¹⁰ are related to how business performance or product quality improves, reducing the average cost based on the volume of accumulated data. Indeed, according to Bajari *et al.* (2019), access to data improves the quality of forecasts made by the firm, but with diminishing

8 A good is characterized as "non-rivalrous" when one individual's consumption doesn't prevent another's simultaneous use, nor does it reduce in quantity or quality. Furthermore, a good is said to be non-exclusive when parties cannot be excluded from using it. Duch-Brown, Martens and Mueller-Langer (2017, p. 25) argues that: "Data are non-rival and non-excludable by their very nature: many parties can use them at the same time without any loss of utility for any of the parties. If they are not made excludable by law they become a public good. Free-riding by other users would take away incentives for private agents to invest in their production."

9 Statista (2020) categorizes data in two types: unique and replicated. Unique data refers to data that is originally created and captured, while replicated data pertains to data that is duplicated and used. The percentages of both unique and replicated data were determined by Statista, utilizing ratios furnished by the original source.

10 A firm's average costs may remain constant, rise, or fall as its output expands. If average cost falls as output increases, the firm is said to have economies of scale (or increasing returns to scale) (CARLTON; PERLOFF, 1994, p. 35-45). Concurrently, "when it is cheaper to produce two products together (joint production) rather than separately, there is an economy of scope" (PANZAR; WILLIG, 1977; BAUMOL, PANZAR; WILLIG, 1982).

marginal returns. In other words, as more data is accumulated, the additional gains decrease. In this sense, after a certain point (which varies depending on the market), players of diverse characteristics may compete despite initial advantage and previous data accumulation.

On the other hand, economies of scope are related to the relevance of accumulated data in one segment to other available data. The complementarity of databases can increase the value of an existing database, allowing it to be used multiple times and for multiple purposes. However, the study by Bajari *et al.* (2019) concluded that having access to a larger portfolio of products, even in the same category, does not have a significant effect on forecast accuracy. This indicates that data produced in certain retail segments is not crucial for predicting consumer preferences in other segments, resulting in limited economies of scope¹¹.

All these characteristics act as disciplining elements of the potential anti-competitiveness of the accumulation and use of data. Specifically, as presented above, the following stand out: i) data's economies of scale and scope; ii) its non-rivalrous/non-exclusive nature (which means that the same bundle of data can be used by multiple agents at the same time without diminishing its value); and iii) the limited lifespan of data (which creates opportunities for new entrants to compete in data-driven markets). For these reasons, the large volume of data of generalized firms might not be entirely useful on specific segments, thus specialized firms would not be at a disadvantage in terms of data. This would lead to a scenario in which tipping – the dominance of incumbents – is not necessarily true.

Rubinfeld and Gal (2017) agree that the non-rivalry, the non-exclusiveness and the limited lifespan of data are all characteristics that indicate low barriers to entry in data-driven markets. However, they emphasize the need to analyze all barriers associated with each link of the data value chain, which the authors classify by collection (extraction of data); storage (organizing data in a database); synthesis and analysis (integration of different types of data and the analytical processing of data); and usage (use in decision-making).

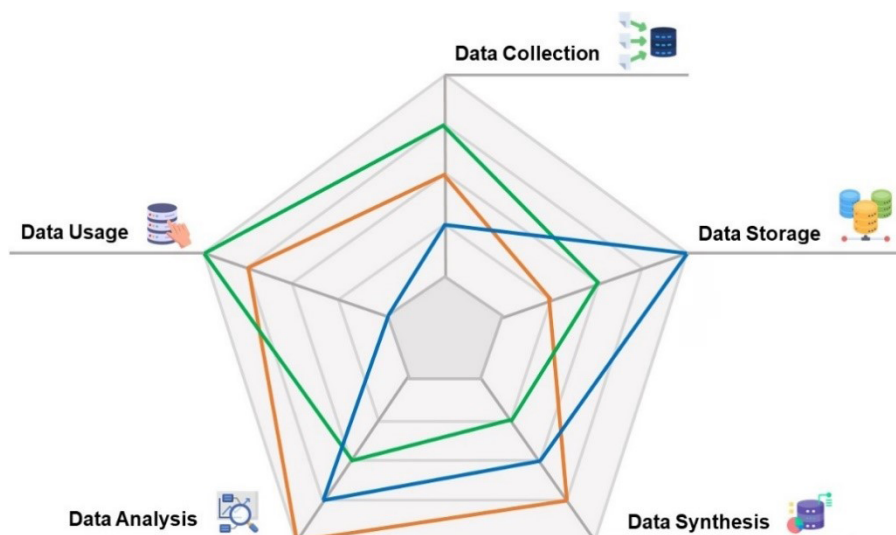
The authors argue that barriers established in relation to each of the aforementioned links would determine the effect on competition and welfare and serve to delineate markets for the antitrust analysis. Their conclusions are based on a comprehensive literature review on technological and legal barriers and interviews of players in markets heavily impacted by the use of big data tools. Barriers cited as examples from the authors include technological, data protection laws, data compatibility and interoperability, network effects, and switching costs¹².

In their theoretical model, Rubinfeld and Gal (2017) argue that barriers to entry can arise at any stage of the data value chain. This means that there is a potential for cumulative negative effects on competition, or that high barriers to entry in one stage of the chain could be overcome by particularities of the relations established in another stage. For example, a firm with limited access to data due to lower economies of scope than its competitors could overcome this barrier by obtaining a good algorithm that would allow it to better understand and use the data it does have (RUBINFELD; GAL, 2017). These situations are illustrated in the figure below.

11 Durand and Williams (2022, p. 4) agree on that characteristic. The authors argue that data must be relevant to be useful. In the specific case of retailers, they affirm that data is useful “when they feed into data analytics that improve the retailers’ performance, notably in terms of sale of products and/or services to customers”. As an example, the authors cite a multi-product firm that will have access to a large volume of data, but not necessarily data from one segment of the business will be helpful elsewhere.

12 For a detailed analysis, see Rubinfeld and Gal (2017, p. 11-31).

Figure 1 - Data value chain and hypothetical situations



Source: Rubinfeld and Gal (2017). Elaborated by the authors. Note: the image provides a visualization of the varying intensity of data usage across the value chain for five hypothetical sectors. Each sector showcases unique data demands and applications, which are distinctly illustrated by the intensity of each link within the chain. This representation offers insights into how different industries prioritize and utilize data at various stages.

Given this broad chain of connections between various spheres of data use, the authors concluded that data-driven markets should be analyzed considering their particularities, specifically bearing in mind the relevance of data in each specific market. This is essential to identify the optimal balance between the efficiencies gained from the usage of data and its potential harm to competition, qualifying the actions of the antitrust authority in each specific context. Furthermore, it should be noted that the efficiencies gains derived from big data have brought unprecedented changes to people's daily activities. Illustrating such argument, Sokol and Comerford (2016) list the following examples:

- i. Monetization of data subsidizes free products for consumers: targeted advertising sales provides the means for data-driven companies to offer free products to consumers, such as real-time traffic updates;
- ii. Improved quality and enhanced innovation:

Big data allows for regularized customization of decision-making, reducing risk and improving performance. It also changes corporate ecosystems by moving data analytics into core operational and production functions, and it enables the introduction of new products (RUBINFELD; GAL, 2017, p. 2).

2.2. Overview of the Brazilian jurisdiction

The discussion of data as an asset or input with important competitive value – and even as an essential facility – has not been fully addressed in Brazil. The Brazilian antitrust authority, *Conselho Administrativo de Defesa Econômica (Cade)*, has not yet fully considered data as an asset in its merger

analysis (MONTEIRO, 2017)¹³. However, Cade's report on Digital Platforms gives some examples where data was considered in the analysis (BRASIL, 2021).

In the merger case n. 08700.006084/2016-85 (Microsoft/LinkedIn)¹⁴, data was part of the economic rationales presented by the Parties. It was stated that users' data from LinkedIn would enhance Microsoft product offers. Cade analyzed the use of LinkedIn's data by the CRM solution offered by Microsoft. The Brazilian authority concluded that data from LinkedIn couldn't be considered an essential facility for two reasons: (i) players that offer CRM solutions do not use data from LinkedIn; and (ii) products are complementary and not vertically related. Cade's decision in the LinkedIn/Microsoft case suggests that there was no consideration of the use of data as an essential facility within the framework of merger analysis. However, the report on Digital Platforms (BRASIL, 2021) indicates that Cade is open to consider data as an asset with competitive value in other contexts.

The analysis of the merger case n. 08700.001796/2020-94 (OLX/Zap Imóveis) also considered that the operation could raise the possibility of limiting third-party CRM solutions by restricting access to data. The analysis showed that, while the merged firm could technically engage in this practice, it would likely lead to a negative rating of the merged firms' platforms, as their performance is related to their capacity to generate data. Therefore, there was no incentive for the merged firm to do so. In this case, Cade analyzed the likelihood of the merged firm to engage in an anticompetitive practice. Its conclusion showed that the authority is indeed concerned about the impacts of data as an input, but that data's competitive relevance is being analyzed on a case-by-case basis.

In summary, Cade's approach is consistent with the literature that suggests that big data could be an issue to be considered in a competition assessment but not necessarily an anti-competitive condition *per se*. In other words, even though there could be some potential competition concerns that need to be considered – such as the existence of barriers to entry and the competitive relevance of data in each specific market –, data have *prima facie* pro-competitive effects as a means to convey value to the final consumer. For this reason, it is important to analyze each case individually.

3. ACCESS AND USE OF DATA IN THE BRAZILIAN RETAIL MARKET

Data is considered a strategic resource for companies seeking competitive advantages. However, not all sectors benefit equally from data collection, storage, analysis, and usage. Likewise, not all types of data are equal and, as discussed earlier, several steps beyond data access are relevant to assess its effectiveness as a competitive advantage. As Félix, Tavares and Cavalcante (2018) points out, the retail sector is one of the segments that uses big data tools the most in Brazil, thus it was chosen as a case analysis to identify the ways in which this industry (i) collects and (ii) processes and uses data, emphasizing the changes that the massification of internet access has brought to the industry and the means through which retailers could take advantage of data. The evidence collected is based on a review of literature from business schools, newspapers, and business journals. This information provides insight into how Brazilian retail businesses access and use data. It is important

¹³ Although the importance of the use of data has been discussed laterally by Cade in its jurisprudence, there are no signs of the authority constructing clear theories of harm in the sense of understanding the use of data as the sole condition for market closure, regardless of the case-by-case analysis.

¹⁴ Editor's note: all the Cade cases mentioned in this article can be consulted at <https://tinyurl.com/y7obr4z5>.

to note that there is limited academic research on this topic, so the understanding of the market must be built using various types of sources.

The existence of stores and markets has always been directly related to the interest merchants have in accessing, collecting, and evaluating information about their customers. This set of data accounts for everything from shopping preferences to demographic characteristics, allowing vendors to better understand their target audience, customize products and services, create competitive differentiators, and build consumer loyalty. Levy, Weits and Grewal (2023) summarize this point:

Retailers have always gathered data, in more or less formal ways. Good salespeople might keep notes about what their favorite customers like, and many retailers have extensive data about members of their loyalty programs (LEVY; WEITS; GREWAL, 2023, p. 17).

The widespread use of big data has led to more sophisticated analysis of consumer behavior and market trends. This made it possible to customize options and promotions for individuals or groups of customers, as well as to improve the quality of products and services. For example, retailers can now offer products in an omnichannel model, which means that customers can buy online or in-store and choose how to receive their purchases. Additionally, retailers can use big data to gain deeper insights into consumer behavior¹⁵, providing relevant inputs for a better strategic decision-making and leading to improvements in their operations.

3.1. Collection of data

The literature review on the impacts of big data on competition, discussed in Section 2, highlights the importance of analyzing the entire value chain of data. This includes the collection, storage, processing, analysis, and use of data. For the Brazilian retail market, we first analyzed how retailers access data. This is then followed by an analysis of how data is processed and used in the Brazilian retail market.

Brazilian retailers have been collecting data for a long time (FÉLIX; TAVARES. CAVALCANTE, 2018) and a well-known example is Casas Bahia (QUEIROZ, 2006). This retailer developed solutions that allowed low-income consumers – and therefore those with less access to credit – to finance products of greater value. To make this strategy viable, Casas Bahia and companies that replicated its

¹⁵ It is also important to consider the possibility of manipulating consumers' behavioral patterns such as in the impact of dark patterns in mischievous data usage. Dark patterns have been originally defined as “tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something”. In other words, dark patterns undermine user autonomy by subverting or influencing decisions made by the users. Calo and Rosenblat (2017) argue that marketplace business models, including sharing economy business, for instance, ease market manipulation specially through the use of personal data. One very prominent example of a dark pattern is the “attention grabber”, where unrelated signs or ads are strategically located within the field of view to attract the users click. These “attention grabbers” can be easily suited to a given user's preferences through extensive use of their personal data. Competition-wise, dark patterns in the form of “obstructions”, where the interface or even the rules of engagement in a membership hinder the cancellation or changes in preference of users, can result in high switching costs for users, keeping rivals from accessing this market (MATHUR; KSHIRSAGAR; MAYER, 2021). Such a comprehensive definition demonstrates the need to consider specific analyzes on a case-by-case basis, since, like the specific use of data solutions, the design of dark patterns tends to be very specific to the business model and the particularities of the intermediated market, in addition to being an element addressed by consumer protection, data protection and competition defense mechanisms.

model sought to collect as much information as possible from their customers. Therefore, retailers in Brazil have been developing financial products heavily based on the ability to analyze consumers' personal data within the retail operation itself (QUEIROZ, 2006). Also, many retailers have partnered with financial institutions to offer financing options and installment purchases, which requires obtaining personal data from consumers such as credit history, income, and address (QUEIROZ, 2006).

More recently, open banking¹⁶ expanded the ability of retailers in Brazil to gain better insights into consumer spending habits. An extensive report published by the consulting firm Ernest Young (CREMA; AMARAL, 2022) shows that, by sharing their financial data, consumers can benefit from personalized offers and discounts, while retailers can improve their marketing and risk assessment. Moreover, according to AFRAC, the Brazilian technological association for commerce, open banking will enable retailers to provide better customer service, develop new products and services, and improve their overall efficiency (AFRAC, 2020; GARCIA, 2021).

In fact, retailers are incorporating financial services and payment methods into their offerings. This can provide a competitive advantage and an additional source of data. For example, Magalu, a major Brazilian retailer, registered 5,2 million digital accounts in 2022 (TAUHATA, 2022). The entry of retailers can also be noted within the payment systems segment. Mercado Livre, the largest online marketplace in Latin America, has partnered with Vtex, a specialized tech firm in virtual payments, to develop its own payment system. As a consequence of this transformation, traditional credit bureaus are expanding their data solutions to accommodate the influx of information from open banking. These products offer sophisticated modeling structures and focus on credit risk mitigation to a wide range of retailers in Brazil (RIBEIRO, 2023).

The use of data is also functional as an element of exploring the competitive landscape of the market. Major retailers rely on market data and surveys provided by third-party companies such as Nielsen, Euromonitor, Kantar etc. to gain a detailed view of the market beyond their own reach. These firms conduct research and gather information on consumer trends, market segmentation, consumer behavior, and other relevant data for the market. Retailers also purchase custom research from research institutes to learn more about their customers and their rivals' customers¹⁷. As in the case of financial data, this type of information is not new for large retailers. However, it remains relevant even as big data is increasingly used for analytical purposes. This dynamic is related to the fact that internal data cannot provide a complete picture of the market. Additionally, the way research is conducted has changed over time, one example is role of surveys, that are now often held online, which can accelerate results and lower costs¹⁸.

New data mining companies surged, and existing data companies expanded their businesses.

16 Open Banking, or the open financial system, allows customers of financial products and services to share their information with different institutions authorized by the Central Bank. This service streamlines the relationship between banks and client, standardizes data sharing in the National Financial System, and enhances competitiveness while reducing service costs (TEM..., 2022).

17 Custom research studies are designed to meet the specific needs of the retailer, such as understanding consumers in a particular locality, age group, or gender. In general, these analyses could involve questionnaires, interviews, and direct observations aimed at understanding consumer preferences, needs, and motivations. The results obtained are subsequently used to guide marketing strategies, product selection, and market segmentation.

18 "[...] a survey that took one month to be completed is now being completed in one week, with a cost reduction of 50%." said Lucas Melo, cofounder of MindMiners." (POR QUE..., 2017).

According to data available at Crunchbase¹⁹, Brazil is home to 42 data mining companies 73 data science companies (an average founding year of 2010), and 103 data analytics companies. Therefore, data became available for retailers of all sizes. For example, Scantech is a data startup founded in 1992 in Uruguay that only started growing rapidly in 2016. The company developed a business model in which it connects its API to retailers of all sizes for free and offers personalized reports that show how their sales compare to those of other retailers in a specific region. Scantech then sells the data gathered from its API to the manufacturers (SOUZA, 2023). Another example of a data mining company is Zyte, an European-based firm founded in 2010 that provides services to Brazil. The company specializes in web scraping, which is the process of extracting data from websites. Zyte's clients include a wide range of companies (HOFFMAN, 2015)²⁰.

In another sphere, it's noteworthy that public surveys have also become more comprehensive and accessible to the public, including the Monthly Services Survey (PMS), the Monthly Trade Survey (PMC), and even surveys that map population data, such as the Continuous National Household Sample Survey (PNADc)²¹, which analyzes information on income and consumption of the population in each municipality. Additionally, there has been a significant increase in the offering of free available data, such as Google Trends. These enable the mapping of consumer trends and company performance.

In the same vein, it is worth mentioning that retailers will collect data through their own apps and websites, as well as through their performance on marketplaces. The amount of data will vary depending on the size of the retailer, but technologies to collect, store and process it are available. Companies such as Nuvemshop offer services for creating and managing websites and app stores for retailers. Brazilian retailers have access to a large amount of data and the means to collect it. The retail market in Brazil is composed of businesses of various sizes, each with its own strategy for creating and accessing data. Additionally, several large retailers coexist, all of which have full access to a wide range of data.

Digitalization has led to the emergence of new data companies that offer data mining and analysis services. Additionally, some companies find value in the data generated by retailers and purchase it from them in order to provide market assessments, making it easier for smaller players to access data. Social media, marketplaces, and search engines are also sources of information for retailers of all sizes.

19 Crunchbase collects and organizes data from more than 3,700 global investment firms that submit monthly portfolio updates. The company uses artificial intelligence, machine learning, and manual data validation and curation to provide accurate data.

20 Data mining is a powerful tool to spread data access to different sizes of retailers. In this vein, other companies that should be noted are Experian, Opinion Box, Horus, and Neogrid. They have gained greater relevance in the Brazilian retail market by providing more accurate market research and studies.

21 All the mentioned databases are produced by the Brazilian Institute for Geography and Statistics (IBGE). The PNADc (Continuous National Household Sample Survey) aims to track quarterly workforce fluctuations and long-term socioeconomic development in Brazil. Launched in January 2012, it provides quarterly workforce indicators and annual indicators on supplementary topics like employment types, caregiving, and technology use, among others, investigated either quarterly or yearly. The survey's unit of investigation is the household, and it has been designed to yield data for various regional and administrative divisions across Brazil, progressively broadening the scope of disclosed indicators since its inception. The PMC generates indicators to monitor the retail sector's performance in Brazil, examining the gross resale revenue of formally established companies with 20 or more employees, primarily engaged in retail. Initiated in January 1995 in the Rio de Janeiro Metropolitan Region, it provides real and nominal revenue, employment, and wages and other remuneration indicators. Finally, the PMS provides indicators to track the performance of the non-financial services sector in Brazil, analyzing the gross service revenue of formally established companies with 20 or more employees, excluding the health and education sectors.



3.2. Usage of data

Brazilian retailers use data in a variety of ways to improve their operations and profitability. Some of these ways include: (i) personalizing marketing campaigns; (ii) implementing dynamic pricing algorithms; (iii) making strategic decisions; (iv) forecasting sales (MACHADO, 2018).

Retailers can use data to personalize their marketing campaigns by targeting their ads more effectively and tracking the effectiveness of their marketing campaigns. For example, they can use data to identify which customers are most likely to be interested in a particular product and then target those customers with ads for that product. This enables them to personalize their marketing strategies and develop targeted campaigns that effectively reach and engage their desired audience (HUI *et al.*, 2013). They can also use data to track how many people click on their ads and how many make a purchase after clicking on an ad.

Data is also utilized by retailers to implement dynamic pricing algorithms, which involve adjusting prices based on factors such as demand, supply, competition, and other relevant variables. This enables retailers to automate price adjustments, leading partially or fully to revenue maximization, optimized profit margins, and competitive pricing for a diverse customer base. A study conducted by Poel (2020) examined the effects of dynamic pricing on Netshoes, a Brazilian retailer of sports products. The findings revealed that the retailer pricing model is particularly beneficial for companies with a large inventory volume, enhancing their operational efficiency. However, dynamic pricing algorithms are available also for smaller retailers through specialized companies. An example is Infoprice, a startup focused on collecting data of prices to transform information for retailers.

Data is essential for strategic decision-making in the retail market. Through data analytics, retailers can effectively identify business expansion opportunities, determine the viability of opening new stores, evaluate the potential of entering new markets or segments, and make informed choices regarding product launches or discontinuations. Such analysis is commonly utilized in the retail sector, with the Huff Model being one example of instrument to predict a store's area of influence, aiding in the selection of optimal locations for expansion (LIMA; TEIXEIRA, 2008).

The field of logistics and infrastructure management has seen remarkable advancements in productivity through the extensive use of data (NAZÁRIO, 1999; CESAR; MOORI, 2021). An example is Via Varejo, a leading retail chain in Brazil, is a prime example of this. The company has achieved significant productivity gains by leveraging user data and modeling it within a cloud-based system. This transformative experience began in the company's largest distribution center, which is also the largest in Latin America. Logistics management is also available for small and medium-size retailers. Note that companies like Nuvemshop offer solutions that help businesses of all sizes in Brazil integrate their sales with delivery companies, promising a freight costs reduction by up to 40%, according to the company.

Data also provides valuable insights into competitors' performance, enabling retailers to closely monitor the market trends and adjust their strategies accordingly (FÉLIX; TAVARES. CAVALCANTE, 2018). Making strategic decisions based on data-driven insights is crucial for achieving success and maintaining competitiveness in the retail industry. Sales forecasting and inventory optimization are essential applications of data in the retail market. By utilizing data analysis techniques such as time series models and machine learning algorithms, retailers can accurately predict future demand

based on historical sales data, seasonal patterns, market trends, and other relevant information. The use of data-driven forecasting in the Brazilian retail industry is extensive, covering a wide range of applications across different segments.

The dynamics discussed in the preceding paragraph can be clearly seen in the Brazilian drugstore industry. Raia Drogasil is a prime example of a company that counted on data to develop its strategic decision. It has built a comprehensive database of customer preferences by operating a network of brick-and-mortar stores. This database turned out to be a useful tool for deciding its digital and physical expansion. The retailer will also use information collected from stores to personalize commercial strategies for online sales (BRANDÃO, 2022).

Retailers can also use data to forecast sales. This forecasting often relies on historical data on sales and data related to economic conditions of each market. A notable method of gathering such information is through platforms like Google Trends, as mentioned above. To illustrate this possibility, Azevedo (2021) successfully employed data from Google Trends to anticipate product demand in retail, yielding commendable outcomes. Effective sales forecasting enables retailers to adeptly manage inventory and staffing, preventing both product shortages and excess stock. Efficient inventory management not only reduces warehousing expenses but also enhances operational efficacy. By leveraging data for forecasting and optimization purposes, retailers can make informed decisions about inventory management, leading to improved customer satisfaction and cost-effectiveness in their operations.

Félix, Tavares and Cavalcante (2018) investigated the development of big data analytics at Magazine Luiza in a case study. The goal of the study was to identify the main obstacles to the development of this area within the company. One interesting finding was that the author did not mention a lack of data as a barrier, but rather the need for significant management changes to enable physical stores to share data digitally. Other challenges mentioned were the availability of data scientists and the development of more sophisticated tools to enable a faster response (the interviewees indicated that this should be overcome when the use of big data becomes more intensive - a learning by doing process). Note that the availability of data scientists should improve over time as the demand for the job increases and more courses are offered to train workers.

The development of internet access has revolutionized data collection and processing in the retail market in Brazil. As a result, the retail industry is experiencing a remarkable transformation, enabling retailers to make data-driven decisions and enhance their overall performance. In fact, according to a survey held by SBVC (Brazilian Society for Retail and Consumption), 64% of retailers, large and small, are willing to invest more in digital transformation, where online sales and omnichannel appear as top priorities (SBVC, 2023).

4. FINAL REMARKS: THE MULTIPLE FACTORS DETERMINING COMPETITION IN THE BRAZILIAN RETAIL MARKET

While data can contribute to various aspects of retail activities, it is important to recognize that competitiveness extends beyond data alone. Retail's end activity is delivering products and services to customers, so it drastically differs from sectors where data solutions are the final product. Price, product range, quality, service levels, store design, branding, and promotional efforts are all



crucial dimensions of competition in retail. These factors can affect consumers' choices regarding different retailers and brands, and their importance should not be underestimated as they determine the competitive dynamics of the retail market. The dynamic nature of retailing allows innovative new entrants to target unfulfilled opportunities, especially with the growth of online shopping.

Furthermore, numerous successful online retail start-ups have demonstrated that pre-existing data is not a prerequisite for success and that data access could be achieved for new entrants through existing services (HAGIU; WRIGHT, 2020)²². Factors like scale, efficient logistics, purchasing economies, and other non-data-related aspects can also contribute to a retailer's competitive advantage. It is evident that a retailer's relative competitive performance cannot be solely attributed to data asymmetries, as there are numerous conventional reasons for a retailer's success or failure that have nothing to do with data in the first place.

In addition, it is fundamental to consider the transformations in the behavioral patterns of the Brazilian consumer. A survey conducted by Google/Ipsos between October 2021 and January 2022 reaffirms that most Brazilian consumers are researching online even when buying physically – the study indicates 80% of consumers will research online before going to a physical store (MADEIRA; FRANZÃO, 2022). In the same sense, Brazilians are also considering other issues beyond price while researching online. Evidence on that is the number of times the terms “best” and “cheap” were searched at Google Shopping. By the end of 2021, consumers in Brazil started to research more for “best” than “cheap”²³.

Rennie *et al.* (2020) reinforce this point citing experts and studies on consumer behavior that argue consumers are researching on the internet not only for comparing prices, but also for “comparing everything”, which also shows the complexity behind consumers' decision-making²⁴. This evidence also reinstates the influence of multihoming²⁵ in the Brazilian scenario, showing that there is an important influx of access and usage of various comparison tools, retailers, marketplaces, and, fundamentally, different purchase channels, whether online or offline. According to a survey conducted by Akman (2022) with 11,151 users from 10 countries, Brazilians used an average of four retail platforms over a 12-month period – one of the highest rates in the world. This multihoming-prone behavior can be attributed to the highly heterogeneous nature of retail activities and products in Brazil, as well as the consumer behavior of actively seeking out the best deals available.

That said, the behavioral transformations described run parallel to broader trends in the Brazilian scenario, most notably the dispersed nature of the national retail market and the income

22 Hagiu and Wright (2020) argue that, as data is proprietary, it's possible it can be purchased from other sources. Although the use of unique and proprietary data can provide a competitive edge, technological progress can undermine this position. The authors support their point with an example: Nuance's Dragon solutions once dominated the speech-recognition software market, relying on users to train the system with their individual voice patterns and storing this information within its database. However, recent developments have introduced speaker-independent systems. These modern models, which require minimal to no user training, utilize public speech data sets. As a result, various companies have been able to launch innovative speech-recognition applications, posing a challenge to Nuance's market dominance.

23 See Appendix, Figure 2.

24 “We know more about advertising performance than ever before and can measure outcomes with amazing granularity. And yet, understanding consumer decision-making is more difficult than it's ever been. In 2020, following the outbreak of coronavirus and subsequent restrictions on physical retail, the proportion of purchases happening online has risen to record levels. And while most purchases are still made offline, the media and information that inform those purchases are increasingly online, and the complexity of potential decision-making pathways has grown considerably” (RENNIE *et al.*, 2020, p. 6).

25 Belleflamme and Pleitz (2019) define multihoming as the use of multiple platforms for the same purpose during the purchasing process.

restrictions faced by most consumers. Regarding the consumer's budget constraints, although the greater sophistication of the search pattern is noticeable, price remains a central variable. An analysis based on Google Trends data shows that “price” is searched much more often on Google Shopping than terms like “best”, indicating that price is a major factor that Brazilians consider when researching products^{26, 27}. Price centrality is another vector towards greater multihoming, as switching costs are weighed down by lower reserve prices. Therefore, the budget constraint condition leads the average consumer to derive greater utility from obtaining low prices, increasing the propensity to search for different sellers and therefore widening the space for new retailers and entrants.

As for the organization of the Brazilian online retail market, the sector is marked by fragmentation, with a high incidence of small and independent providers (not directly linked to a large chain) that operates centrally through the structure provided by marketplaces. In this scenario, the popularization of online consumption and the existence of a wide range of alternatives for these small sellers implies a reduction in barriers to entry, allowing atomized suppliers to offer their products without incurring in the high entry and maintenance costs associated with the need for stores and physical structures. In the same sense, technological evolution expands the potential demand, increasing the consumer market for small players. Finally, the possibility of access to data analysis tools provided by marketplaces qualifies the performance of these merchants, increasing their productivity and reducing costs.

In conclusion, multi-homing, retailer differentiation, the characteristics of data such as diminishing returns of its value, and the widespread access and usage of data from retailers in Brazil are all factors that limit the competitive advantage gains derived exclusively from data. Marketplaces, which are widely known as data-driven retailers, began operating in Brazil in 2000, and by 2010, many other retailers switched to this business model. Thirteen years later, old and new companies continue to fiercely compete in the Brazilian market.

As a final remark, it should be noted that this study was an exploratory one, based on publicly available examples and data. There is little research on the topic, so the analysis employed here included not only academic studies, but also news and articles related to business practices. This dynamic is also important because the topic is quite new and therefore under-researched in academic literature. As a result, the most effective descriptions of the relevance of data in the practice of different industries and businesses come directly from the thinking developed by the people involved in those businesses. Hence, this paper contributes to consolidating a literature review on barriers to entry due to big data and indicates that it is not straightforward that big data leads to high levels of competitive advantage between companies in the same industry, serving as a basis for other future investigations.

REFERENCES

AKMAN, Pinar. A Web of Paradoxes: Empirical Evidence on Online Platform Users and Implications for Competition and Regulation in Digital Markets. **Virginia Law and Business Review**, v. 16, n. 2, 2022. Available at: <http://dx.doi.org/10.2139/ssrn.3835280>. Accessed on: July 10, 2023.

26 Google Trends data for the last 5 years was used.

27 See Appendix, Figure 3.



ASSOCIAÇÃO BRASILEIRA DE TECNOLOGIA PARA O COMÉRCIO E SERVIÇOS (AFRAC). **Orientação do nosso VP de Relações Institucionais, Edgard de Castro, sobre a tecnologia Open Banking, é citada em artigo da Folha de São Paulo.** São Paulo, November 30, 2020. Available at: <https://www.afrac.com.br/blog-full-content.php?id=64>. Accessed on: July 10, 2023.

AZEVEDO, Matheus Pasche. **Previsão do volume de vendas do comércio varejista com o uso do Google Trends.** 2021. Trabalho de Conclusão de Curso (Bacharelado em Ciências Econômicas) - Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021. Available at: <https://pantheon.ufrj.br/handle/11422/15085>. Accessed on: July 15, 2023.

BAJARI, Patrick; CHERNOZHUKOV, Víctor; HORTAÇSU, Ali; SUZUKI, Junichi. The Impact of Big Data on Firm Performance: An Empirical Investigation. **AEA Papers and Proceedings**, v. 109, p. 33–37, May 1, 2019. Available at: <https://www.aeaweb.org/articles?id=10.1257/pandp.20191000>. Accessed on: July 2, 2023.

BAUMOL, William J.; PANZAR, John C.; WILLIG, Roberto D. Contestable markets: An uprising in the theory of industry structure: Reply. **The American Economic Review**, v. 73, n. 3, p. 491-496, 1983. Available at: <https://www.jstor.org/stable/1808145>. Accessed on: July 12, 2023.

BELLEFLAMME, Paul; PEITZ, Martin. Platform competition: Who benefits from multihoming? **International Journal of Industrial Organization**, v. 64, p. 1-26, 2019. Available at: <https://doi.org/10.1016/j.ijindorg.2018.03.014>. Accessed on: July 15, 2023.

BILBAO-OSORIO, Beñat; DUTTA, Soumitra; LANVIN, Bruno. **The Global Information Technology Report 2014: Rewards and Risks of Big Data.** Geneva: World Economic Forum, 2014. Available at: <https://tinyurl.com/2xl8khj3>. Accessed on: July 12, 2023.

BRANDÃO, Raquel. A estratégia da Raia Drogasil (RADL3) para ser a farmácia do futuro. **Exame**, November 24, 2022. Available at: <https://tinyurl.com/28ezqffo>. Accessed on: July 12, 2023.

BRASIL. Conselho Administrativo de Defesa Econômica. **Mercados de Plataformas Digitais: Cadernos do Cade.** Brasília: Cade, 2021. Available at: <https://tinyurl.com/2juujeq4>. Accessed on: July 12, 2023.

CALO, Ryan; ROSENBLAT, Alex. The taking economy: Uber, information, and power. **Columbia Law Review**, v. 117, n. 6, 2017. Available at: <https://tinyurl.com/24zuw7wt>. Accessed on: July 20, 2023.

CARLTON, Dennis W.; PERLOFF, Jeffrey M. **Modern industrial organization.** 2. ed. New York: HarperCollins College Publishers, 1994.

CESAR, Dirceu Gonçalves; MOORI, Roberto Giro. O Efeito do Big Data Analytics na Relação entre Gestão Estratégica da Logística em Serviços e Desempenho Operacional. **Gestão & Regionalidade**, v. 37, n. 112, 2021. Available at: <https://doi.org/10.13037/gr.vol37n112.6348>. Accessed on: July 5, 2023.

CREMA, Juliana; AMARAL, Cristiane. **Future Consumer Index 2022: Preço é essencial, mas consumidor brasileiro quer mais.** **EY Parthenon**, Brazil, Oct. 3 2022. Available at: <https://tinyurl.com/2chtlvaw>. Accessed on July 10, 2023.

DUCH-BROWN, Nestor; MARTENS, Bertin; MUELLER-LANGER, Frank. **The economics of ownership, access and trade in digital data:** JCR Digital Economy Working Paper 2017-01. Seville: European Commission, 2017. Available at: <https://tinyurl.com/2coggnr2>. Accessed on: July 10, 2023.

DURAND, Benoît; WILLIAMS, Iestyn. Are data a source of unassailable competitive advantage in retailing? **RBB Economics**, [S. l.], 2022. Available at: <https://tinyurl.com/2yq8bbr4>. Accessed on: July 11, 2023.

FÉLIX, Bruno Muniz; TAVARES, Elaine; CAVALCANTE, Ney Wagner Freitas. Fatores críticos de sucesso para adoção de Big Data no varejo virtual: estudo de caso do Magazine Luiza. **Revista Brasileira de Gestão de Negócios**, São Paulo, v. 20, p. 112-126, 2018. Available at: <https://doi.org/10.7819/rbgn.v20i1.3627>. Accessed on: July 5, 2023.

FURMAN, Jason. **Unlocking digital competition**: Report of the Digital Competition Expert Panel. London: Crown, 2019. Available at: <https://tinyurl.com/ynxmx6gu>. Accessed on: July 15, 2023.

GARCIA, Larissa. Open Banking permite juntar serviços de vários bancos em plataforma única. **Folha de São Paulo**, São Paulo, Jan. 29 2021. Available at: <https://tinyurl.com/27qt462e>. Accessed on: July 16, 2023.

HAGIU, Andrei; WRIGHT, Julian. When Data Creates Competitive Advantage: And when it doesn't. **Harvard Business Review**, Boston, Jan. 2020. Available at: <https://tinyurl.com/qpwa485>. Accessed on: July 5, 2023.

HOFFMAN, Pablo. The history of Scraping. **Zyte**, [S. l.], Mar. 16 2015. Available at: <https://tinyurl.com/25snvykc>. Accessed on: July 5, 2023.

HUI, Sam K.; HUANG, Yanliu; SUHER, Jacob; INMAN, Jeffrey. Deconstructing the "First Moment of Truth": Understanding Unplanned Consideration and Purchase Conversion Using In-Store Video Tracking. **Journal of Marketing Research**, v. 50, n. 4, p. 445-462, August, 2013. Available at: <https://doi.org/10.1509/jmr.12.0065>. Accessed on: July 10, 2023.

IANSITI, Marco. The Value of Data and Its Impact on Competition. **Harvard Business School NOM Unit Working Paper**, v. 22, n. 2, July 20, 2021. Available at: <http://dx.doi.org/10.2139/ssrn.3890387>. Accessed on: July 9, 2023.

LEVY, Michael; WEITZ, Barton A.; GREWAL, Dhruv. **Retailing management**. 11. ed. [S. l.]: McGraw Hill Education, 2023.

LIMA, Andre. G. P.; TEIXEIRA, Dalton. J. Localização Varejista: Um Estudo sobre a Utilização do Modelo de Huff para a Tomada de Decisões sobre Localização. **Revista Gestão & Tecnologia**, v. 8, n. 1, p. 1-19, 2008. Doi: <https://doi.org/10.20397/2177-6652/2008.v8i1.211>. Accessed on: July 10, 2023.

MACHADO, Felipe. N. R. **Big data o futuro dos dados e aplicações**. São José dos Campos: Érica, 2018.

MADEIRA, Luiza; FRANZÃO, Thiago. O trajeto das compras de fim de ano: como mobilizar o desejo do consumidor com ajuda do Waze. **Think with Google**, Brazil, Nov. 2022. Available at: <https://tinyurl.com/29x2qprv>. Accessed on: July 12, 2023.

MATHUR, Arunesh.; KSHIRSAGAR, Mihir.; MAYER, Jonathan. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In: CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2021, Yokohama, Japan. **Proceedings** [...]. Yokohama, Japan. New York: Association for Computing Machinery, 2021. Available at: <https://doi.org/10.1145/3411764.3445610>. Accessed on: July 8, 2023.

MONTEIRO, Gabriela Reis Paiva. **Big data e concorrência**: Uma avaliação dos impactos da exploração de big data para o método antitruste tradicional de análise de concentrações econômicas. Dissertação de Mestrado (Mestrado Acadêmico em Direito da Regulação) – Escola de Direito do Rio Janeiro, Fundação Getúlio Vargas, Rio de Janeiro, 2017. Available at: <https://hdl.handle.net/10438/20312>. Accessed on: July



11, 2023.

NAZÁRIO, Paulo. A importância de sistemas de informação para a competitividade logística. **Revista Tecnológica**, São Paulo, v. 5, p. 31, 1999. Available at: <https://tinyurl.com/2xkx55tt>. Accessed on: July 12, 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Data-driven innovation: Big data for growth and well-being**. Paris: OECD Publishing, 2015. Available at: <https://doi.org/10.1787/9789264229358-en>. Accessed on: July 15, 2023.

PANZAR, John C.; WILLIG, Robert D. Economies of scale in multi-output production. **The Quarterly Journal of Economics**, v. 91, n. 3, p. 481-493, 1977. Available at: <https://doi.org/10.2307/1885979>. Accessed on: July 5, 2023.

POEL, Naiche Van Der. **Precificação dinâmica como uma ferramenta para administrar preços e vendas no varejo on-line**: um estudo de caso da Netshoes. 2020. Dissertação (Mestrado Profissional em Gestão para a Competitividade) – Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2020. Available at: <https://hdl.handle.net/10438/29493>. Accessed on: July 5, 2023.

POR QUE é importante conhecer os consumidores de sua marca? **Exame**, São Paulo, April 24 2017. Available at: <https://tinyurl.com/28wx963u>. Accessed on: July 7, 2023.

PRÜFER, Jens; SCHOTTMÜLLER, Christoph. Competing with big data. TILEC **Discussion Paper**, Tilburg, n. 2017-006, 2017. Available at: <http://dx.doi.org/10.2139/ssrn.2918726>. Accessed on: July 11, 2023.

QUEIROZ, Renata S. Bernardes de. A importância dos modelos de Credit Scoring na concessão de crédito ao consumidor no varejo. In: SEMINÁRIOS DE ADMINISTRAÇÃO FEA USP, 9., 2006, São Paulo. **Anais [...]**. São Paulo: Faculdade de Economia, Administração, Contabilidade e Atuária, 2006.

RANKING dos Principais Apps. **Similarweb**, São Paulo, 2023. Available at: <https://tinyurl.com/2adgvht4>. Accessed on: July 5, 2023a.

RANKING dos Sites Principais. **Similarweb**, São Paulo, 2023. Available at: <https://tinyurl.com/28cpt9w5>. Accessed on July 5, 2023b.

RENNIE, Alistair; PROTHEROE, Jonny; CHARRON, Claire; BREATNACH, Gerald. **Decoding decisions: Making sense of the messy middle**. [S. l.], Think with Google, 2020. Available at: <https://tinyurl.com/y6aafv5r>. Accessed on: July 12 2023.

RIBEIRO, Mariana. Mercado Pago e VTEX lançam iniciador de pagamento. **Valor Econômico**, São Paulo, May 17 2023. Available at: <https://tinyurl.com/2ns5pk76>. Accessed on: July 12 2023.

RUBINFELD, Daniel L.; GAL, Michal. Access barriers to big data. **Arizona Law Review**, v. 59, 2017. Available at: <http://dx.doi.org/10.2139/ssrn.2830586>. Accessed on: July 15 2023.

SOCIEDADE BRASILEIRA DE VAREJO E CONSUMO (SBVC). **Transformação Digital no Varejo Brasileiro**. 5th ed. São Paulo: SBVC, 2023. Available at: <https://tinyurl.com/24m8veqf>. Accessed on: July 12, 2023.

SOKOL, D. Daniel; COMERFORD, Roisin. Antitrust and regulating big data. **The George Mason Law Review**, v. 23, 2016. Available at: <https://tinyurl.com/235ohjlv>. Accessed on: July 7, 2023.

SOUZA, Karina. Scanntech, startup de dados para varejo e indústria, capta R\$ 210 mi com Warbug Pincus. **Exame**, São Paulo, January 19 2023. Available at: <https://tinyurl.com/2xkwpudw>. Accessed on

July 10, 2023.

STATISTA. Share of unique data and replicated data in the global datasphere in 2020 and 2024. **Statista**, New York, 2020. Available at: <https://tinyurl.com/2btypq3p>. Accessed on: July 13, 2023.

TAUHATA, Sérgio. Magalu lança cartão para empresas e avança no crédito. **Valor Econômico**, São Paulo, May 12 2022. Available at: <https://tinyurl.com/25ey76z5>. Accessed on: July 11, 2023.

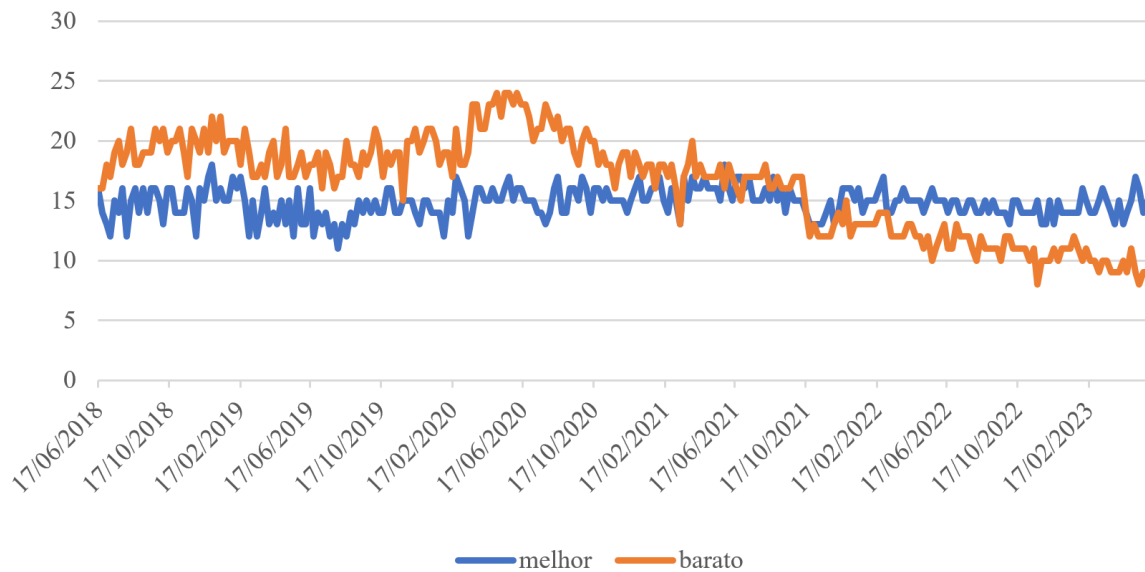
TEM início nova fase do Open Banking no Brasil. **Serviços e Informações do Brasil**, Brasília, October 31 2022. Available at: <https://tinyurl.com/2agyqb6z>. Accessed on: October 17, 2023.

TUCKER, Darren S.; WELLFORD, Hill. Big mistakes regarding big data. **Antitrust Source**, Chicago, December, 2014. Available at: <https://ssrn.com/abstract=2549044>. Accessed on: May 29, 2024.



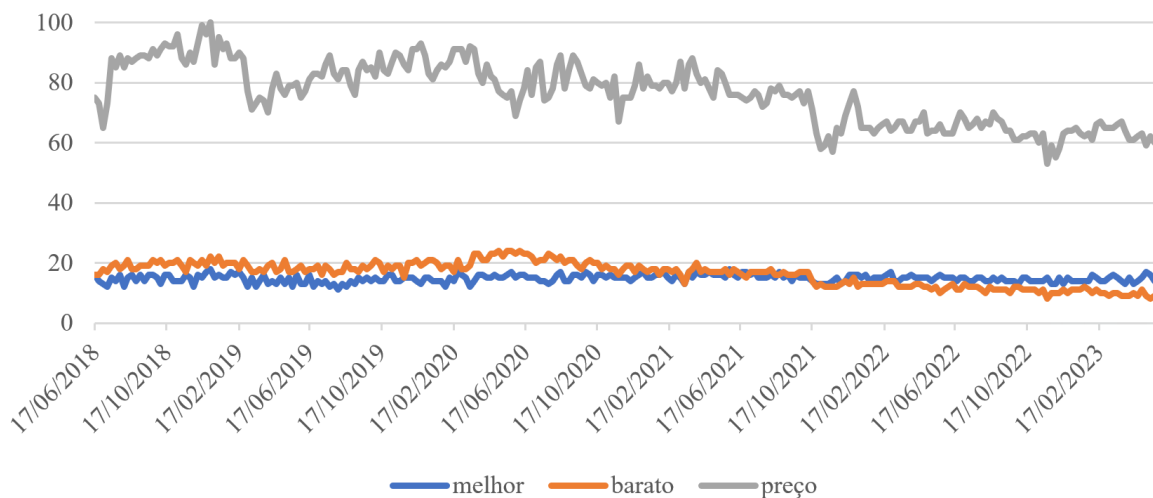
APPENDIX

Figure 2 - Interest over time in searches in Google Shopping for best (*melhor*) and cheap (*barato*) in the last 5 years



Source: Google Trends. Elaborated by the authors.

Figure 3 - Interest over time in searches in Google Shopping for best (*melhor*), cheap (*barato*) and price (*preço*) in the last 5 years



Source: Google Trends. Elaborated by the authors. Note: Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.