

Data Monetization and Inequality: Evidence from Case Studies

Zayna Ahsan

Received April 05, 2024

Accepted August 04, 2024

Electronic access August 15, 2024

This paper delves into the increasing significance of data as a driving force for economic growth, drawing parallels with the historical role of oil. Utilizing a qualitative framework of research, it examines the complex relationship between data monetization and social as well as economic inequality through a comprehensive literature review. Case studies featuring prominent firms such as Spotify and Facebook showcase a range of data monetization strategies, shedding light on the challenges faced by smaller enterprises. Furthermore, an analysis of sectoral trade networks reveals notable disparities in data monetization capabilities, emphasizing the necessity of understanding sector dynamics for fostering inclusive growth in the digital era. The findings underscore the multifaceted implications of data monetization on economic and social structures, advocating for nuanced policy interventions. By synthesizing insights from diverse disciplines, this paper enhances our understanding of how data-driven processes shape inequality and economic evolution. It calls for policies that promote equitable access to data resources and foster inclusive growth across sectors, intending to bridge the gap between data-rich and data-poor entities in the digital age.

Keywords: inequality, data monetization, big data, sectoral networks, economic growth, social class dynamics, labor mobility, and inclusive growth

Introduction

In recent years, it seems that data has become the oil of this generation as it ‘fuels’ the economy as well as our daily lives (Hirsch). Like oil, data needs to be collected and refined to be of any value, and the impacts of this refinement process i.e. data monetization is what I will explore in this paper. As per Gartner’s IT glossary, data monetization refers to the process of using data to obtain quantifiable economic benefits (“Definition of Data Monetization - Gartner Information Technology Glossary”) Fred, J. (2017)¹. Data monetization techniques include using data to make measurable business performance improvements and inform decisions. The global pandemic of COVID-19 has only added to the value of data mainly due to increased digitization and a massive shift in consumer behavior towards digital platforms, e-commerce, and online activities.

The vast amounts of information generated daily have led to data becoming an integral part of firms’ business strategies and societal interactions, meaning its role in the economy is continuously increasing in significance.

It is forecasted that the data monetization market will reach a value of up to \$15.4 billion by 2030, with a compound annual growth rate of 19.98%. To understand the growth, it should be considered that in 2019, the market was valued at around only \$1.61 billion. These statistics provide context to the phenomena of data-derived activities. The increasing prevalence of data monetization brings about a need to highlight its consequences as it touches upon broader issues of fairness, access, and the

distribution of benefits in the digital age.

Prior research has explored the different techniques and implications of data monetization, shedding light on its potential consequences for privacy, individual autonomy, and corporate practices. However, the specific linkages between data monetization and social and economic inequality remain underexplored. McKinsey & Company published a survey² in 2017 discussing how data monetization fuels growth and briefly introduced the possibility of it as a differential, but there is much academic work to be done regarding this idea. This paper will discuss how data monetization and inequality are related using a thorough literature review. In addition, to add to the main literature review, there is a case study that will revolve around different companies that are key players in this field to identify a pattern.

In the case studies, companies like Spotify, Facebook, Mastercard, Progressive Insurance, and AB InBev showcase successful data monetization strategies, enhancing their revenue streams and market competitiveness. Data analysis reveals that sectors with higher intra-sectoral trade, like services, benefit more from data monetization, exacerbating economic disparities, while limited trade connections hinder the leveraging of data monetization by sectors like manufacturing.

This paper aims to explore the extent to which data monetization contributes to economic and social inequality and the underlying mechanisms shaping this relationship. Our hypothesis is that data monetization exacerbates existing inequalities due to disparities in access and capacity to utilize

data resources effectively. By synthesizing evidence from existing literature and case studies, this paper seeks to provide a nuanced understanding of how data monetization strategies impact economic and social structures, advocating for policies that promote equitable access to data resources and foster inclusive growth across sectors.

The rest of the paper is structured as follows. Section II covers methodology of the paper, which includes selection criteria for the literature review and case studies, as well as an explanation of the calculation of sector involvement and so on. Data limitations are addressed as well. Section III provides an in-depth literature review of academic materials based on data monetization and inequality while developing links to the research question at hand. Section IV delves into the case study with examples of firms using data monetization to generate revenue. Section V uses data analysis to determine inequality by sectors i.e. how involved they are in the economic network, and relating it to their involvement in the data market. Network effects are to be discussed in depth in Section V. Section VI discusses the findings of the main three sections and their implications and concludes the paper.

Methodology

This study utilizes a qualitative framework, synthesizing evidence from existing literature and case studies of prominent firms such as Spotify, Facebook, Mastercard, Progressive Insurance, and AB InBev. The selection criteria for case studies included market influence, diversity of sectors, and availability of data. Literature sources were chosen based on their relevance, publication in high-ranking journals, and citation frequency on platforms like Google Scholar.

Data for the analysis was sourced from publicly available input-output tables, detailing product trade between sectors. The specific datasets used include the U.S. Use of Commodities by Industry table, covering recent years up to the present. Statistical methods involved calculating sectoral trade shares and network effects to understand sector involvement in data monetization.

Data limitations include proprietary restrictions and the need for data aggregation from multiple sources. These limitations prevent a full quantitative analysis, highlighting the necessity for future research with comprehensive datasets. Discussing these limitations in detail ensures transparency and clarity about the scope and constraints of the current analysis.

Literature Review

This paper brings together two strands of literature, 1) data as an asset and how it can become a source of economic benefit and revenues, and 2) the causes and mechanisms of economic

and social inequality. The paper aims to connect and correlate these two strands in the later sections.

First, the literature presents a multifaceted exploration of data monetization, examining its different techniques and how it operates. Big data monetization is about deriving economic value from vast amounts of data, and it involves several key stages known as the big data value chain. These stages include data collection, storage, processing, analysis, and monetization. Firstly, data collection involves gathering information from diverse sources like social media, sensors, and transactions. Various techniques such as web scraping, APIs, and IoT devices are used to collect this data, ensuring it's of high quality and complies with privacy regulations. Once collected, the data needs to be stored efficiently for easy access and analysis. Different storage technologies like data warehouses, data lakes, or cloud-based solutions are utilized for this purpose, each with its implications for data monetization. Next, data processing involves cleaning, transforming, and preparing the data for analysis. Technologies like Hadoop and Spark are commonly used here, along with techniques such as data cleansing and normalization to ensure the data is suitable for analysis. In the data analysis stage, insights and patterns are extracted from the data using techniques like statistical analysis and machine learning. These insights can be used to inform decision-making and drive value creation. Finally, in the data monetization stage, organizations leverage the insights gained from data analysis to generate revenue. This can involve selling data to third parties, offering data-driven products or services, or using data to optimize internal processes and drive cost savings. By understanding and effectively managing each stage of the big data value chain, organizations can unlock the economic potential of their data assets and create new opportunities for revenue generation and growth (Liu and Chen, 2015)³.

Ofulue and Benyoucef (2022)⁴ provide a technology-enabled perspective through a literature review, offering insights into how businesses leverage data as an asset. The paper not only delineates existing research but also identifies gaps in the field, particularly in ethical considerations such as privacy and security. Delving into methods and approaches, their work offers valuable insights into challenges, opportunities, and best practices within the evolving landscape of data-driven strategies.

Moving on to the literature on social inequality, Polacko's (2021)⁵ overview of income inequality provides a foundational understanding of the broader economic and social context in which issues of inequality manifest, such as tax policy and the spread of non-standard jobs due to technology. This knowledge serves as a crucial backdrop for examining the specific impacts of data monetization on inequality. Similarly, Kraus, Park, and Tan (2017)⁶ explore the manifestations of economic inequality in daily life through the lens of social class. Though primarily focused on social class dynamics, their work establishes a connection to discussions about making money from data,

facilitating a holistic perspective on the interplay between economic inequality and technological trends.

Furthermore, DiMaggio and Garip's (2012)⁷ investigation into network effects and social inequality sheds light on how social networks influence opportunities, resources, and information distribution. This exploration not only uncovers the root causes of social inequality but also informs potential interventions. The network effect becomes a crucial mechanism for understanding how data monetization may contribute to inequality between companies, underlining the interconnected nature of economic structures and technology trends. Collectively, these works provide a nuanced understanding of data monetization, economic inequality, and societal structures, but their relationship is yet to be identified.

The research question, exploring how data monetization contributes to social and economic inequality and the mechanisms shaping this relationship, aligns well with existing literature, providing a unique perspective on the interaction of data monetization, social class, and economic inequality. The insights shared by Ofulue and Benyoucef (2022)⁴ into data monetization discuss business-centric usage. My research complements this by addressing potential inequalities arising from the different ways firms use data. Polacko's (2021)⁵ overview of income inequality lays a foundational understanding. The collaborative work of Kraus, Park, and Tan⁶, concentrating on social class dynamics, serves as a bridge to my review. Understanding the signs they discuss is crucial for investigating how data monetization may influence or perpetuate these social class markers. Chien-Hung Liu and Chuen-Lun Chen's review³, highlighting strategic dimensions of data monetization, aligns with my exploration of economic impacts, contributing insights into the broader impact of data-driven strategies on social and economic inequality. This alignment positions my study as a valuable contribution, offering a deeper understanding of the societal implications of data monetization in the context of social class and economic disparities.

By integrating these areas of study, this review aims to uncover the potential impacts of data monetization on societal equity and economic inequality. The insights shared by Ofulue and Benyoucef (2022)⁴ into business-centric data usage complement our research by addressing potential inequalities arising from differential data utilization across firms. Additionally, discussing types of journals and highly cited articles related to data monetization will highlight the significance of existing studies. I provide further insights by doing a holistic analysis of what the literature pertaining to data monetization and inequality has to offer.

I add to the existing literature by carrying out a series of case studies that contribute to our understanding of how data monetization may worsen or alleviate income disparities.

Case Study

While current research has shown that the data monetization market is rapidly expanding, it is important to understand the distribution across different sectors by analyzing the key data monetization players. Let's examine some current companies that utilize data as a source of revenue.

Case Study I: Spotify

Firstly, we can consider the work being done by Spotify, a music streaming company. It collects data continuously about user behavior and preferences e.g. their listening habits and music preferences. This data is analyzed to identify patterns and commonalities using pre-developed algorithms. Once this data is analyzed and refined, Spotify uses it in a two-pronged manner. It creates a personalized experience for all users through tailored recommendations and advertisements, enhancing consumer relations and attracting a larger customer base. Additionally, it sells this data to artists and labels so that they, in turn, may use it to understand the current trends and make informed decisions about tours and marketing. This can be used to maximize overall success for both Spotify and its music partners.

Case Study II: Facebook

Secondly, it is important to discuss Facebook, a key player in the social media domain. Social media runs off of sharing and collecting data so its monetization was inevitable. Facebook monetizes data primarily through its extensive collection of user information. Through various means like user interactions, demographic details, location data, and third-party sources, Facebook puts together detailed profiles of its users. These profiles enable advertisers to target specific audience segments. Advertisers utilize Facebook's advertising tools to create tailored ad campaigns, specifying criteria like age, gender, location, and interests. Ads are displayed across Facebook's platform, including the News Feed, Stories, and Instagram, and on external websites and apps through the Facebook Audience Network. An auction-based system determines which ads are shown, with advertisers bidding for ad space. Revenue is generated through advertisers paying for ad impressions or actions. Additionally, Facebook may monetize data through partnerships or by licensing anonymized or aggregated data to third parties. In essence, Facebook's business model revolves around leveraging user data to deliver targeted advertising, constituting a significant portion of its revenue stream.

Case Study III: Mastercard

From the financial sector, Mastercard has also delved into data monetization through its division Mastercard Advisors. They collect data about credit card transactions globally and analyze it

using machine learning algorithms. Patterns are identified based on geographical locations, transaction amounts, and merchant categories. Mastercard uses this data as a product for its clients to aid them in overcoming business challenges. This entails strategic recommendations for marketing and even new products. This way it benefits Mastercard as well as the companies that are its clients. It is important to acknowledge Mastercard's unique access to transaction data and its industry knowledge and expertise, maximizing the value derived from the data.

Case Study IV: Progressive Insurance

Another example is a leading auto insurance company, Progressive Insurance. It has launched an innovative data monetization effort through its Snapshot program. Utilizing telematics devices, Snapshot gathers driving data from enrolled customers, enabling Progressive to provide tailored discounts and rewards for safe driving habits. This pricing approach not only enables them to derive value from the collected data but also enhances Progressive's risk assessment capabilities, creating mutual benefits for both the company and its policyholders.

Case Study V: AB InBev

While the above examples are all part of the services sector, it is also possible for manufacturing sectors to participate in data monetization. AB InBev, the world's leading brewing company, has been relying on data-driven insights since 2019. They planned to design and launch a worldwide data platform with the aims of expanding analytics, automating operations, and creating new revenue streams. The different functions of this platform include optimizing operations by leveraging microeconomic data related to soil, water, and weather conditions to support local producers and brewers. This data helps in optimizing demand forecasts and product development lifecycles, leading to operational efficiencies and cost savings. Secondly, it works on minimizing downtime by utilizing data to minimize machine downtime by identifying potential issues before they occur. This proactive approach improves operational efficiency and reduces maintenance costs.

Moreover, AB InBev is using data to identify opportunities for expanding into new markets. By launching new lines of business supported by data insights, they can target specific consumer segments and geographic regions more effectively. By creating a unified data platform using Talend's data integration platform, AB InBev has streamlined data management and eliminated silos. This enables better analytics capabilities and decision-making across regions and brands, improving operational performance and potentially new revenue streams.

Keeping these examples of firms successfully monetizing data in mind, we can discuss their similarities to identify a pattern. All the companies discussed above can be classified as

successful companies in their fields with large-scale operations. Spotify is valued at around \$45.2 billion. Facebook (Meta Platforms) has a market cap of \$1.2 trillion, and Mastercard is worth approximately \$428 billion. Lastly, AB InBev has a net worth of \$123.5 billion. The reason behind this is the high costs associated with data monetization e.g. in the collecting of data in servers and databases, investment in cybersecurity measures, development of analytical technology, and creation of monetization strategies. It can also be attributed to the fact that only big companies have access to a large enough volume of data to be able to extract valuable information from it. Smaller companies, on the other hand, end up paying for the finished data product as they cannot afford to gather and monetize data themselves. This exacerbates economic inequality between larger and smaller firms due to inequality of opportunity.

Furthermore, most of the key players in data monetization tend to be from the tertiary sector, except AB InBev. In recent times, we have seen a shift in the global economy from the manufacturing sector to the services sector. Therefore, all these companies appear to be thriving and can invest in data monetization whereas primary and secondary (less valued than tertiary) have been unable to directly derive value from data to date. Services also require a higher degree of direct interaction with the consumer hence there is more data available to the firms. Their product in itself depends on data collection and analytics, which go hand in hand with data monetization. AB InBev's success and innovation in this matter may be on account of their large enterprise size and control over 500 different brands, allowing access to more data and a wider consumer base. Generally, we can conclude that data monetization is restricted to the tertiary sector and large companies. This sectoral composition contributes to existing economic and social inequality as economically disadvantaged firms cannot afford to utilize data as an asset.

These examples illustrate successful data monetization strategies but also underscore disparities faced by smaller firms. Larger firms benefit from extensive data resources and advanced analytics capabilities, while smaller firms struggle to compete, exacerbating economic inequality. If including smaller firms is not possible due to lack of public information, this limitation should be explicitly stated, emphasizing that data from struggling firms is often not released, contributing to a skewed analysis.

Descriptive Statistics of Data Monetization and Network Effects

The previous section looked into several individual companies, but let's consider the network between them in detail. For this purpose, I used data from the US's The Use of Commodities by Industry table and the figures of products traded between

sectors. I calculated the share of products that are consumed by other sectors for each industry respectively. The higher the percentage share, the more the sector trades within themselves in relative terms. This shows their level of involvement within the sector.

Analyzing the network structure can validate the assertion about the relative advancement of the services sector. One notable aspect is that the services sector tends to engage in more intra-sectoral trade, while sectors like agriculture have lower inter-sectoral trade activity. An example of this is Manufacturing with a percentage of 56.6% whereas Other services except government have a percentage of 96.8%. This means that only 56.6% of manufacturing products were demanded and consumed by other sectors whereas for other services it was 96.8%.

This observation is crucial in understanding why the manufacturing sector often lacks robust data monetization strategies. Unlike the services sector, which frequently interacts with other industries and is in high demand by most of them, manufacturing sectors typically have limited trade connections with services. As a result, they may struggle to leverage data monetization effectively, as they lack the necessary interconnectivity with sectors where such strategies are prevalent.

The absence of strong trade ties between manufacturing and services exacerbates economic inequality, as discussed previously. This disparity is further compounded by network effects, which widen the gap between sectors benefiting from interconnectedness and those left behind.

Considering the workforce in these industries, it's essential to explore the concept of 'induced' employment and economy. 'Induced' refers to individuals employed directly by a specific sector. These individuals have restricted opportunities to move to another sector where they may take advantage of data monetization and get opportunities for better earnings and potential growth.

In conclusion, the complex interplay between sectoral networks, data monetization capabilities, and labor mobility underscores the challenges of economic inequality. There is a positive relationship between sector involvement and data monetization capabilities. These in turn negatively impact economic equality due to unequal opportunities for sectors relatively less integrated into the network. Understanding these dynamics is crucial for formulating policies that address disparities and promote inclusive growth.

To better understand these dynamics, future research should explore how inter-sectoral trade relationships and network effects contribute to the disparity in data monetization capabilities. Quantitative analysis using comprehensive datasets can provide deeper insights into these relationships and inform more effective policy interventions.

Conclusion and Discussion

This paper examines how data, likened to the oil of our generation, is increasingly fueling the economy through processes like data monetization. Reviewing the literature, analyzing sector involvement, and studying case examples delves into the impact of data monetization on social and economic inequality, highlighting its implications for privacy, autonomy, and corporate practices in the digital age. The literature review synthesizes two key strands: the monetization of data and research on economic and social inequality. It explores the process of data monetization, from data collection to revenue generation, and investigates its intersection with issues such as income distribution and social class dynamics. By integrating these areas of study, this review aims to uncover the potential impacts of data monetization on societal equity and economic inequality. The case study section explores how various companies, such as Spotify, Facebook, Mastercard, Progressive Insurance, and AB InBev, utilize data monetization strategies across different sectors. These examples underscore the significant role of data in revenue generation and highlight the challenges and disparities smaller firms face in accessing and leveraging data for economic benefit, contributing to existing economic and social inequality. The analysis examines sectoral interconnectivity using data from the US's commodity consumption table, highlighting the services sector's higher intra-sectoral trade compared to sectors like manufacturing and agriculture. This disparity in trade connections underscores challenges for manufacturing in leveraging data monetization effectively, exacerbating economic inequality and emphasizing the importance of understanding sectoral networks for policy formulation aimed at promoting inclusive growth.

Future research should aim to conduct quantitative analyses, leveraging comprehensive datasets to establish causal relationships. This would involve using indicators like the Gini coefficient to measure economic inequality. Our findings advocate for policies that promote equitable data access and foster inclusive growth, bridging the gap between data-rich and data-poor entities.

No existing paper directly answers our research question. This study infers potential answers using literature and case studies, contributing to a deeper understanding of the societal implications of data monetization. Additionally, framing our qualitative analysis as synthesizing evidence from literature and case studies of real firms helps to understand how data monetization strategies may be linked to inequality. This approach underscores the importance of our research in the broader academic context.

I plan to follow up on this paper by researching deeper into the relationship between data monetization and inequality to identify whether they have a causal relationship, or simply a correlated one. Causation means that one variable directly influences

the other. On the other hand, correlation indicates a statistical relationship. For example, higher education is associated with higher income levels but does not necessarily mean more education causes more income. To draw this conclusion for data monetization, and socioeconomic inequality, sufficient quantitative data analysis will have to be conducted, perhaps across countries using indicators like the Gini Coefficient. When data and results about the employment of data monetization and its effects are more readily available, it will be possible to conduct this research and find a definite conclusion in the context of correlation versus causation.

Acknowledgments

I would like to thank Kabir Hiranandani for his infallible support and my mentors, Tacye Hiong and Hammad Shaikh, for their guidance and patience throughout my work. Secondly, this work would not have been possible without my parents' constant encouragement and belief in me.

References

- 1 J. Fred, *Master's thesis*, 2017.
- 2 J. Gottlieb and K. Rifai, *Fueling Growth Through Data Monetization*, 2017, <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Fueling%20growth%20through%20data%20monetization/Fueling-growth-through-data-monetization.pdf>, Accessed: 2024-08-11.
- 3 C. H. Liu and C. L. Chen, 2015.
- 4 J. Ofulue and M. Benyoucef, *Management Review Quarterly*, 2022, 1–45.
- 5 M. Polacko, *Statistics, Politics and Policy*, 2021, **12**, 341–357.
- 6 M. W. Kraus, J. W. Park and J. J. Tan, *Perspectives on Psychological Science*, 2017, **12**, 422–435.
- 7 P. DiMaggio and F. Garip, *Annual Review of Sociology*, 2012, **38**, 93–118.