CBS

INTERNATIONAL BUSINESS SCHOOL

LEVERAGING BIG DATA FOR ENHANCED FINANCIAL INCLUSION IN NIGERIA: A COMPREHENSIVE ANALYSIS OF STRATEGIES, IMPACT, AND CHALLENGES IN DEVELOPING ECONOMIES.

Master of Arts Thesis

in partial fulfilment of the requirements for the degree of

MASTER OF ARTS (MA)

in International Business

with specialization in Digital Transformation Management

Kosisochukwu Cypran Akubude

Student-No. 1195501031

Advisor: Prof. Dr. Geoffrey Writes

Cologne, May 2024

Acknowledgement

Romans 8:18: For I consider that the sufferings of this present time are not worth comparing with the glory that is to be revealed to us.

My greatest thanks to God for his unwavering blessings and grace upon my life. I would like to thank my dad, Chief Cyprian Akubude for his financial and guiding support throughout these challenging years of my life. I would not have made it this far if not for his sacrifices to ensure that I pursue the best education possible and to bring a positive light to the family. Many thanks to my awesome siblings, who are always there to cheer me up. Also, thanks to my company and work colleagues at A.H.T. Syngas Technology N.V. for their support and giving me the opportunity to restructure my work schedule to make room for my Master Thesis. I immensely appreciate my friends and my partner, Chiara Panschar, for their emotional support till the completion of this thesis.

Special thanks to:

Prof. Dr. Geoffrey K. Writes, Professor of Finance, and Lecturer at CBS International Business School for choosing to support and advise me on this research topic. Also, thanks to the Google Data Analytics Professional Certification team for providing me with the skills required in the statistical analysis of this paper.

In Loving Memory:

Mrs. Jacinta Akubude

Abstract

Financial inclusion has been a battle the Nigeria has been trying to combat for over two decades but with the growing demand for digital devices, it is imperative that the opportunities from accessing these devices could be leveraged by financial institutions to better understand the secluded market and offer financial products and services which are more geared towards such market. The purpose of this study is to evaluate the relationship between big data and financial inclusion. A secondary examination adopted in the study is to assess the relationship between economic data performance and financial inclusion. In the process, two models were devised to effectively analyze the chosen variables for the models. This study collected data for all African countries based on availability. Only data for the year of 2021 were collected due to inconsistent data over the previous years and because the current data were updated in 2021. Sampled data consists of 40 countries and 35 countries for Model I and II, respectively. Model I examined the relationship between big data and financial inclusion whilst Modell II examined the relationship between economic statistical data performance and financial inclusion. The study utilized cross-sectional OLS regression model to evaluate the relationship between the variables for the models. The findings from Model I in the study show that there is a significant relationship between big data and financial inclusion. Model II findings also show that there is a statistically significant relationship between economic data performance and financial inclusion.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
FI	Financial Inclusion
AI	Artificial Intelligence
ML	Machine Learning
GDP	Gross Domestic Product
MDGs	Millenium Development Goals
SDGs	Sustainable Development Goals
SMEs	Small Medium Enterprises
UN	United Nation
AU	African Union
Ecommerce	Electronic Commerce
OECD	Organization for Economic Co-operation and Development
BDA	Big Data Analytics
CBN	Central Bank of Nigeria
TAM	Technology Acceptance Model
NGOs	Non-Governmental Organizations
USAID	United States Agency for International Development
DFID	Department for International Development - UK
FDI	Foreign Direct Investment
FinTech	Financial Technology
SPI	Statistical Performance Indicator
OLS	Ordinary Least Square Regression

1. Introduction

1.1. Background and Context

Prior to my earlier experience in the field of financial inclusion, I researched and authored a paper on the topic of financial exclusion existing in the Ghanaian economy, more specifically in the rural or small communities. The paper also emphasized the importance of the existence of microfinance institutions which play a critical role towards bridging that gap between the neglected citizens and access to finance. This was an important topic considering that the motive was to help contribute to fulfilling Goal 1.A of the MDGs which was aimed at eradicating extreme poverty and hunger in half by 2015. Extreme poverty was described as people living on less than $1.25 per day. The MDGs consisted of 8 goals with 18 targets in total (Ritchie & Roser, 2018). My interest was mostly stuck with the first goal. As a finance student, it just made sense to focus on that since my learnings could be well applicable in that area.

According to the United Nations Department of Economic and Social Affairs (2015), it was reported that over 1 billion people have been lifted out of extreme poverty since 1990. This represents almost half of the total population of people in the poverty line and was achieved 5 years ahead of the 2015 expected deadline. Despite how impressive the result seems; it was also reported that about 800 million people are still living in poverty at the global level. That is still a substantial amount of people living in abject poverty and something needs to be done to change the status quo.

There are considerable amounts of criticisms surrounding the implementation and organization of the MDGs, several of which include statistical concerns. A general critique is the apparent lack of analytical strength and rationale for the selected goals. Following their initial adoption, some indicator definitions, baselines, and targets were modified to give the impression that progress had been more rapid than it had been (Hickel, 2014).

Financial inclusion in emerging countries can be looked at in several dimensions in the sense that there are various complex factors that come into play to determine the extent to which people have access to financial services. Some of the factors include regulatory, technological, and socio-economic dynamics. Emerging economies are plagued by both opportunities and challenges due to certain variables such as changes in demographics, increasing urbanization, and technological advancements. Despite the massive adoption of smartphones over the past decade, there are certain hindrances that are preventing the delivery of innovative financial services some of which include limited physical infrastructure, low financial literacy rates, and regulatory constraints (Błach, 2020).

As someone who has lived in developing countries like Nigeria, Togo, and Ghana, I have witnessed firsthand how underdeveloped traditional banking infrastructure is, most especially in rural areas. According to Chase (2024), a traditional bank is described as an institution with a license to operate in a physical location and capable of issuing loans, operating savings, and checking accounts, and offering investment services. There are quite a number of requirements from banks to offer basic financial services. For example, a credit score is required to grant loans by traditional banks. In a country that does not have a robust credit scoring system, this automatically poses a threat to individuals or SMEs' ability to access those loans. This is one of the major reasons why there is a rampant trend for microfinance institutions to fill in those gaps as they tend to be more flexible. Traditionally, banks rely on things like credit scores and proof of income to decide who gets these services. But what if a high population of people do not have a formal credit history or a regular job? That is where big data comes into play.

There is an outstanding amount of data being generated every second from our smart devices through the consumption of social media, ecommerce, and other digital activities. This kind of data is called big data because it is endless and increases rapidly. Gathering data such as someone's mobile usage pattern like paying bills, recharging frequency, online purchases on credit or transaction records, to mention but few, we should be able to analyze and generate an estimate of one's credit worthiness without necessarily needing a traditional credit score. Big data is not only applicable to estimating credit worthiness but also could help in gaining valuable insights into consumer behavior, develop innovative products and identify underserved populations to tailor to their needs.

Big data analytics has so much potential to promote financial inclusion, but it is still met by a considerable number of barriers for a successful and effective implementation (Bharadwaj, 2019). Regulatory impediments, limitations in technological infrastructure, and most importantly, data privacy concerns are some of the barriers that pose widespread adoption of big data in developing countries. Also, there is a lack of comprehensive research of how big data can be leveraged for enhanced financial inclusion leading to the motive of this thesis.

Nigeria being the most populous nation in Africa means that it will be met with a lot of challenges towards enhancing even access to financial services. More than 85% of

Nigerians have access or owns a mobile phone which makes Nigeria the hub of innovation in Africa (Statista, 2023). With the widespread adoption of mobile money solutions, the limitations of having a formal bank account have become outdated as people can conveniently send and receive money with their phones (Akintola et al., 2020). The series of financial initiatives confirms that the government is not taking a passive role. Specific regulations imposed by the Central Bank of Nigeria (CBN) to regulate mobile money providers highlights the interest to actively promote financial inclusion and increasing access to system infrastructure (CBN, 2020). Despite the efforts, there are roadblocks that should be expected. Many people still face difficulties understanding technology and adjusting to changes in that realm, especially in the rural communities (Akintola et al., 2020). Nigeria does not stop trying when you consider the introduction of the eNaira, Africa's first digital currency (CBN, 2021). Though the eNaira is still in its infancy stage and we cannot predict how well it is going to pan out eventually, this development still shows how progressive the country is. We can gain deeper understanding of the continent if we look at how Nigeria approaches financial inclusion. Nigeria serves as a test case for us, demonstrating what approaches to financial access using big data and technology work and what does not. It comes from actual experience rather than being concocted in a lab.

The major purpose of this paper is to address the gaps by conducting an analysis of the potential of big data in promoting financial inclusion and overcoming barriers in emerging countries like Nigeria.

An examination of how big data can help determine and resolve the bottlenecks of financial inclusion should provide conclusive actionable insights for key stakeholders such as financial institutions on the forefront, policymakers, and development agencies such as the UN and AU at the backdoor. Imagine communities in developing countries where people cannot easily open a bank account, get a small loan, or insure their businesses. That is the reality of financial exclusion. Big data might hold the key to changing this. This research dives into whether big data can really build better financial systems for everyone. We will look at how it is being used, the difference it could make, and the hurdles we will need to clear along the way.

1.2. Problem Statement

Consider big data as a potent instrument that has the potential to provide financial services to millions of underserved people in developing nations. The problem lies in the fact that we are still figuring out how to make it function well in areas with less dependable infrastructure and a widespread lack of regulations compared to wealthy nations. Too frequently, research only provides us with a limited understanding of the applications of big data, concentrating on credit checks or in a specific area. The bigger picture is lacking: how is this technology genuinely impacting people's lives on the ground, what functions well and poorly, and what obstacles prevent it from realizing its full potential? Not only is big data essential for a more inclusive community but the government also has a role to play with regards to data. Depending solely on utilizing big data for strategic decision making will be less impactful if national statistical data is not prioritized and included in the study. The government has a responsibility to maintain quality economic data and ensure that access to such data is provided accordingly. The performance of national statistical data should have an impact on the level of reach of financial products and services. Examples of such economic statistical data include economic growth metrics, national surveys on individuals and organizations and administrative data such as taxes and health data. Periodically maintaining such data as an economy is essential not just for the internal stakeholders, but also for the external players seeking to leverage these data for investment purposes.

1.3. Research Questions

Based on our overview of the problem statement and the significance of the study, we formulate the following research questions as follow:

I. Does the relationship between big data and financial inclusion in Nigeria hold a positive significance?
II. Does economic data performance have a positive and significant impact on financial inclusion in Nigeria?
III. What are the determinants of financial inclusion in Nigeria?

1.4. Objectives of the Study

The primary objective of this study is to examine the relationship between big data and financial inclusion in Nigeria. The secondary objective is to examine the impact of economic data performance on financial inclusion. Economic data performance is measured by source data statistical capacity and five main pillars: Data Use, Data

Services, Data Products, Data Sources, and Data Infrastructure. The following specific objectives will guide the study as well as help to:

I. Identify the various measures of financial inclusion in Nigeria.
II. Identify the effect and significance of internet access on financial inclusion in Nigeria.
III. Identify the effect and significance of mobile utility payment on financial inclusion in Nigeria.
IV. Identify the effect and significance of digital payments on financial inclusion in Nigeria.
V. Identify the effect and significance of mobile phone ownership on financial inclusion in Nigeria.
VI. Identify the effect and significance of mobile bill payment on financial inclusion in Nigeria.
VII. Identify the effect and significance of data capacity on financial inclusion in Nigeria.
VIII. Identify the effect and significance of data use on financial inclusion in Nigeria.
IX. Identify the effect and significance of data services on financial inclusion in Nigeria.
X. Identify the effect and significance of data products on financial inclusion in Nigeria.
XI. Identify the effect and significance of data sources on financial inclusion in Nigeria.
XII. Identify the effect and significance of data infrastructure on financial inclusion in Nigeria.

1.5. Research Hypotheses

The hypotheses of this study are stated in alternative form:

Ha1: A high degree of big data from all digital sources has an overall strong impact on financial inclusion in Nigeria.

Ha2: Big data from internet usage could improve overall financial inclusion in Nigeria.

Ha3: Big data collected from mobile utility payment could improve financial inclusion in Nigeria.

Ha4: Big data from all digital payments could improve financial inclusion in Nigeria.

Ha5: A high degree of mobile phone usage could improve financial inclusion in Nigeria through sourced usage data.

Ha6: A high degree of mobile bill payment could improve financial inclusion in Nigeria through collected data.

Ha7: An increase in the general statistical performance of a country can improve financial inclusion in Nigeria.

Ha8: A high degree of source data capacity could improve financial inclusion in Nigeria.

Ha9: Economic data use score has a positive influence on financial inclusion in Nigeria.

Ha10: Availability of quality and accessible economic data services has a positive influence on financial inclusion in Nigeria.

Ha11: A high degree of producible economic data score has a positive influence on financial inclusion in Nigeria.

Ha12: A high degree of provisional data sources could improve financial inclusion in Nigeria.

Ha13: The increase in data infrastructure could improve overall financial inclusion in Nigeria.

1.6. Scope of the Study

By assessing the impact of big data on financial inclusion in developing economies, this paper limits its scope to Nigeria. The study will be achieved by gathering and analyzing available data for all developing countries in Africa which are representative of the chosen variables. The familiarity of the economy is what primarily drove our interest to study this area. Also taking into consideration is the increased level of economic instability crippling the banking industry over the past five years especially with regards to the regulatory uncertainties around FinTech. The unavailability of consistent data over the years among developing countries has resulted in data collected and analyzed only for the period of 2021. Based on the selected variables to be evaluated, the paper only used the following indexes and measurements: ACC, IA, MUP, DP, MMA, MPO, MBP, SDC, DUS, DSS, DPS, DSC, and DIS. It should be noted that the indexes may not be the accurate measure of the tested variables and are limited by the difficulties of accessing an extensive set of datasets more in line with the topic of the study. There is a need for further research to build upon this limitation.

1.7. Significance of the Study

The research believes that this paper could alter the situation at hand. Solving a delicate problem requires the involvement of all academic personnel to devise and share knowledge from the findings. Theories will remain theories until they are applied in the real world. Consider the people who are most affected by financial exclusion: small businesses that cannot source loans to purchase machineries to enhance efficiency, or individuals who cannot invest and earn passive income. Big data has the potential to help individuals and organizations just like these if it is used morally and proactively.

Although prior research has been conducted on the topic of big data and financial inclusion, it has primarily concentrated on individual case studies or aspects (Bharadwaj, 2019). We hope to fill a fraction of this gap with this research study. It is about drawing up a road map for success in the real world: which big data strategies have proven effective, where are the roadblocks, and how can we avoid unforeseen issues?

1.8. Structure of the Thesis

The tasks in this study are divided into six (6) chapters. A summary of each chapter is discussed below.

Chapter 1: This chapter proceeded with an introduction as well as a background of the study, statement of the problem, objectives of the study, research questions, research hypothesis, significance of the study, scope of the study and finally organization of the study.

Chapter 2: This chapter comprises of a review of previous related works resulting from a literature review. Relevant information pertaining to the research area such as what people have written, their key findings/results, key solutions and arguments are extracted.

Chapter 3: This chapter focused on the methodology that was used in conducting the research. It describes the model used, sources of data and methods of data collection, techniques of data analysis and statistical methods.

Chapter 4: This chapter encompasses the findings of the result and provides a detailed analysis of the collected data.

Chapter 5: Seeks to interpret the findings of the paper, limitations from the study, recommendations necessary for future research, and conclusions are drawn from summarizing the key findings of the study and providing remarks.

2. Literature Review

2.1. Introduction to Literature Review

This chapter necessitates the review of academic studies conducted within similar fields. The importance of this is to look at what has been done in the past and analyze the problems and solutions devised towards the problem statement. The goal of conducting literature review is to dive deeper into the theories surrounding the topic, empirical data used, and the foundational concepts within the discipline. The gaps from the reviewed literature will also be stated out for the purpose of future studies. The first section starts with the conceptual review which consists of but not limited to understanding of relevant concepts such as big data, financial inclusion, and digital financial inclusion. An examination of these concepts will give us a comprehensive outlook of the technical terms and the theoretical framework upon which the foundation of our entire research is served. Afterwards, our focus is proceeded to the theoretical review where key models and theories are analyzed to derive insights into the spread and adoption of new innovations in the field of financial products and services.

2.2. Conceptual Review

The various concepts underpinning this research topic are reviewed and discussed in the following subsections.

2.2.1. Big Data

Imagine for example you want to extract and analyze the web activity data or PayPal data of transactions over a period of 10 years for your company. You download the data which is mostly in .csv file formats. You open Microsoft Excel to read the data and then realize how slow Excel is. You count the number of rows, and it is running in the 5 – 6 digits. That is the reality of big data. Excel, although being an intelligent and useful tool, instantly becomes useless in this situation and you are left with no choice than to figure out alternative data processing tools. Big data as the name implies refers to data that is so huge and complex that traditional data processing software or applications simply cannot manage to derive insights from (Ozili, 2023).

Various industries have undergone radical changes since John Mashey introduced Big Data in the late 1990s (Fernando et al. 2021). Nearly 87% of businesses surveyed in

2014 across a range of industries expect Big Data analytics to change the game for their industry's competitors in the next three years (Columbus, 2014). Furthermore, 89% of the firms polled believe that companies risk falling behind the competition and losing steam if they do not have a strategy to leverage Big Data analytics within the next year. There is an almost equal distribution of opportunities and challenges that every industry is facing with regards to the level of data in different structures generated daily. It is no surprise that some industries are growing at a pace faster than their counterparts and that is mostly contributed by their early adoption of AI and Machine Learning (Ozili, 2023). The finance industry predominantly in Wall Street has already capitalized on the use of big data as evident in application areas such as cryptocurrency, wealth management, stock predictions, asset management, and insurance.

It is the job of data scientists to synthesize data to remove impurities and produce more understandable outputs. The OECD (2019) states that data sources, software, analytics, programming, and statistics all make up the big data ecosystem. In the context of big data, the '4Vs' suggest defining features. Among these features are qualities like exhaustivity, extensionality, and complexity, in addition to volume (the size of the data), velocity (the speed of processing and analysis of streaming data), variety (the heterogeneity of the data), and veracity (the certainty of the data, the reliability of the source, and the honesty) (IBM, 2020).

The financial sector's data production and consumption rates are notoriously difficult to regulate (Fernando et al., 2021). To ensure the data's consistency and security, a strong data governance architecture is crucial. Another requirement that the financial sector must fulfil is conformity with regulatory rules that govern the reliability and accessibility of financial data utilized in reporting. Consumers of financial services are vulnerable to biased or discriminatory decisions made by artificial intelligence systems trained on inaccurate or poorly quality data. If the decision-making process of an AI is impossible or difficult to explain, it may be difficult to comply with the current supervisory and regulatory standards, which have not considered the increasing deployment of AI.

Let us dive deeper into the Vs of big data. As previously mentioned in the article from IBM (2020) there consists of 4 Vs of big data which include volume, velocity, variety, and veracity. Not saying they are wrong, but it appears that there is more to that. A simple google search attempt will show that there are about 10 of them. It is probable that there has been a major development in that regard. As big data continues to evolve, it would also lead to advancement in its definition and what it entails to fully capture every essential. For this paper, we will limit our examination to just 5 Vs. Based on a study by Husamaldin & Saeed (2019), there are various characteristics of big data which has been categorized into 5 Vs as well as their definitions as follows:

- **Volume**: The first V and maybe the most important basis of big data. This refers to the size of data which has evidently increased over the years thereby producing a storage problem. This challenge has single handedly led to the development of a new IT sector called cloud technology.
- **Variety**: The second V refers to the generation of data from diverse sources, for example: sensors, mobile devices, social networks, etc. These data can come in structured, semi-structured, and unstructured formats (Husamaldin & Saeed, 2019). Databases and spreadsheets are platforms to keep organized data to offer a complete view of data classification. Unstructured data, which might include media types such as text, audio, images, and video, does not always have the necessary organizational structure for analysis. Semi-structured data falls in between structured and unstructured data.
- **Velocity**: The third V is all about how fast data is produced and delivered (Perez, 2023). The single term to describe it is “speed”. Data can be consistently processed in real time in an efficient manner. The higher the velocity, the more companies can react in real time to issues and trends emerging from different demographics thereby proving them with a competitive advantage.
- **Veracity**: The fourth V refers to the concern that many data sources (such as social media sites) intrinsically contain some amount of ambiguity and unreliability which highlights the significance of data quality and trustworthiness. It is essential to make sure that informed decision making is not affected due to relying on inaccurate and unreliable data.
- **Value**: Big data contains a lot of hidden values which could be neglected during processing and analysis. The fifth V recognizes the importance of fishing out these untapped values from the large dataset to enhance decision making. Using advanced analytical tools can help unearth these values.

2.2.2. Data Analytics

Several career paths have spawned out of big data alone such as data scientists, data engineers, machine learning engineers, data analysts, etc. (Wang et.al, 2018). These career roles are somewhat applicable in every single sector or domain. We have data professionals in finance who process and analyze financial metrics. There are those in the healthcare sector, agriculture, engineering, to mention but a few. This shows how flexible those roles can be applied in every field. There is data growing every second from daily activities. These data can be nurtured, extracted, and analyzed to gain valuable information and awareness of potential demands and improvement in product and service delivery.

As previously discussed in section 2.2.1, data comes in three main formats: structured, semi-structured, and unstructured formats. According to a study by Edge Delta (2024), 80% of all data generated are accounted to be unstructured data. Meaning that structured and semi structured data are the minority. This metric alone depicts how important it is to dissect and analyze unstructured data so that meaning can be made from it. Despite structured data being a minority, it is mentioned in the study that it still receives 60% of total expenditure on data accounting for majority of the investment. It is a no brainer that this is the case. The data is straightforward and easy to analyze, requiring minimum effort. There are several software tools that are available to extract structured data from unstructured data. The study also stated that 95% of businesses perceive that even though data is a valuable asset to them, it still poses a significant problem. Businesses would have to incorporate solutions into their business model to store, secure and manage this data which requires a huge investment to begin with. The reason is that they will keep pumping more money into structured data.

Processing unstructured data from call logs, mobile banking transactions, online user-generated content such as blog posts and tweets, online searches, and images is the essence of BDA. This data can be transformed into valuable business information through the application of computational techniques, which reveal patterns and trends between datasets (Zakir, Seymour, & Berg, 2015).

BDA is a reflection of the difficulties that may arise when dealing with data that is too large, too unstructured, and too rapidly changing to be managed using conventional methods. It is currently common practice for organizations, ranging from businesses and research institutes to governments, to generate data of an unprecedented scope and complexity systematically. Organizations all over the world are placing a greater emphasis on the importance of extracting relevant information and competitive advantages from vast amounts of data (Gandomi & Haider, 2015). Attempting to successfully extract useful insights from such data sources in a rapid and easy manner is a tough endeavor. Consequently, analytics has become an inseparable component to fully grasp the potential of Big Data to enhance their business performance and significantly expand their market share. Over the past few years, there has been a significant improvement in the tools that are available to manage the volume, velocity, and variety of big data. The cost of these technologies is not prohibitively expensive, and a significant portion of the software is available for free (Zakir, Seymour, & Berg, 2015).

2.2.3. Financial Inclusion

Ozili (2020) described financial inclusion as a basic human right that every citizen should be entitled to in the sense that every member of an adult population should be able to access basic financial products and services. Traditionally, access to financial services is measured by the ownership of a formal bank account (Ozili, 2023; Allen et al, 2016). There are several policies that have been put in place by different nations to spearhead the adoption of financial inclusion across all sectors. More than 20 national authorities including ones from Africa such Kenya National Financial Inclusion Strategy, South African National Financial Inclusion Strategy (NFIS), and the Financial Inclusion initiatives through the Central Bank of Nigeria have publicly announced a national financial inclusion strategy in an effort to ensure that every individual and household has access to formal financial services (World Bank Group 2022). Some of the various tools that policy makers have implemented to achieve this goal are through financial literacy programs, payment system solutions like Mobile Money in Nigeria, delimited bank account access and registration, mobile banking services and digital finance schemes (Ozili, 2023).

In the first chapter of this paper, we mentioned the concept of microfinance institutions. This concept was pioneered by Muhammad Yunus, and it has played an incredible role in the rural and urban communities. The establishment of microfinance institutions can be constituted as one of the early developments with regards to financial inclusion. The model involves the opening of a microfinance bank in a community, dispersing credit officers all over the town to scout and present the services of the institution to the people mostly through face-to-face interaction with potential customers, and finally granting the service request of the customer. The model is designed such that more emphasis is focused on economic development other than just profit making. However, whenever an opportunity presents itself, there will always be those ready to take advantage of it. A trend which resulted in the collapse of a number of microfinance institutions in Ghana. Antwi (2015) mentioned 3 reasons which led to their collapse such as diversion of funds, high interest rates and overtrading, and regulatory non-compliance. Some of the institutions ended up transforming into a full-blown traditional bank as their size grew and was able to meet up with the Bank of Ghana minimum reserve requirement, whilst others merged. Establishing microfinance institutions as a sole means of enhancing financial inclusion is not sufficient. There are other means that should be considered towards the development of financial products and services that can meet the needs of poor people and those who cannot access formal loans from traditional financial institutions. Digitalization has led to increased opportunities in this regard. Certain areas that were initially non-existent have become possible because of big data. This serves the foundation of this paper and evaluating these opportunities gives us an insight into how we can tap into these opportunities to promote inclusivity among all individuals.

2.2.4. Digital Financial Inclusion

It is a no brainer that the future of financial inclusion is digital just like every other industry. As already mentioned, digital financial inclusion serves as one of the means of aiding countries in achieving its SDGs. The steps towards the promotion of available resources, equal opportunities, and access to traditional financial products and services towards the masses can be achieved through digital financial inclusion. Increasing accessibility to financial services such as loans, insurance, and other key financial services to the underbanked population mostly in rural communities and in developing countries at large is the main purpose of financial inclusion (Bruhn & Love, 2014). Having a bank account with a local bank or financial institution is the starting point of formal inclusivity. A bank account opens so many opportunities for the holder such as the opportunity to build a credit history, easy access to other financial services, ability to receive payments and transfer of funds for personal or business purposes, savings, financial security, and much more. This access can open so many doors to the owner offering them the freedom to transact and engage in financial activities with others. As a result, the goal of financial inclusion is to ensure equality among everyone irrespective of their race, demographic, tribe, gender, etc. have access to financial services, which is directly related to the rate of economic expansion in a country (Hasan et al., 2021).

Currently, financial inclusion is accomplished through the utilization of digital technology that is incorporated into mobile phones, banking applications, or on digital devices such as automated teller machines, point-of-sale terminals, among other things. Early initiatives introduced to combat financial inclusion in Africa can be traced back to the late 1990s/early 2000s when South Africa introduced experimental projects like Wizzit and MTN Mobile Money. The aim was to offer basic banking services like a financial account which enabled money transfer, bill payments, airtime top-up, account management, and financial education (Rouse & Verhoef, 2017; Nan, 2018). The success of MTN Mobile Money in South Africa was automatically replicated in other MTN markets like Benin, Ghana, Côte d'Ivoire, Liberia, Rwanda and more recently Nigeria, MTN's largest customer base. Kenya also had its own financial inclusion revolution with the success of M-Pesa which was launched in 2007. Its quick adoption highlights the strong demand for accessible financial services via mobile phones (Jack & Suri, 2014). The early 2010s sparked the revolution of policy and infrastructure development by the government which are already discussed in section 2.2.3. The aspect of mobile money which the author holds to utmost importance is the fact that it does not require a smartphone to function. In emerging economies where dumbphones are predominantly used necessitates the need for a financial technology that works across any kind of mobile device if there is a sim card. The growth of mobile phones in Africa during the early 2010s alongside increased internet access laid the foundation for growth of further digital financial services (World Bank, 2023).

By bringing formal finance into the hands of people through their mobile (smart) phones, the use of digital technology to achieve financial inclusion is superior to the microfinance model. This is because it avoids the high fixed costs that microfinance institutions incur when expanding financial services to underserved communities with the help of digital technology. Payment service banks and payment service providers have also emerged because of the digital technology revolution. These players play a key role in the delivery of payment services to underserved communities to increase financial inclusion. Insurance Technology, Registration Technology, Property technology, and financial technology (Fintech) players have also emerged because of this revolution (Ozili, 2018). Fintech credit increases credit supply in the local community, low cost of financial services, fast and efficient financial services, convenience to users, time savings, and the low fixed cost incurred are some of the benefits of using digital technology to promote financial inclusion.

In an ever-growing competitive market, some financial institutions have formed alliances with technology companies to enhance the reach of their services to more regions. An example is the strategic alliance between MTN and Ecobank to stifle growth in mobile banking and mobile payments (African Market Entry Consulting Ltd, 2023). This is because of the benefits that these partnerships offer. There are a number of potential risks associated with the future of financial inclusion through digital technology. These risks include concerns regarding privacy, the unauthorized use of customer data, cybersecurity threats, high rates of fraud, and high transaction costs (Ozili, 2018). Because of these dangers, fresh approaches are being formulated to achieve financial inclusion in a manner that is both more effective and more efficient in the future.

2.2.5. Financial Inclusion in Nigeria

The Central Bank of Nigeria (CBN) (n.d.) defined financial inclusion as a state when adult Nigerians have easy access to a broad range of formal financial services that meet their needs. It is possible that the rate of growth and development could be accelerated if a sizable portion of the population of any country had access to financial services. The Nigerian government among other scholars have emphasized the significance of bringing millions of the unbanked population into the financial network and increased access to financial products and services (Iwedi, 2023). With a population of over 200 million people, Nigeria is the most populous country in Africa and one of the reasons why they are referred to as the Giant of Africa. According to the 2023 Access to Finance (A2F) survey which was conducted by Enhancing Financial Innovation and Access (EFInA), an organization which exclusively reports on Nigeria's Financial Inclusion targets, it was reported that financial inclusion has increased to 74%, an increase by 8.82% from 2020 (Central Bank of Nigeria, n.d.; Fakoyejo, 2023). The financial exclusion rate is reduced by 26%, just 1% above the target that the CBN aims to achieve in 2024. Considering we are only 3 months in 2024, it is no surprise that this target will be achieved before the end if they continue through this trajectory. This growth is fueled by major gains in non-bank formal adoption and growth in the banked population. The World Bank (2022) population estimate of Nigerians between the ages of 15 – 64 is 54% of total population. According to Statista (2023), total population of Nigeria as of 2022 is around 216 million. So roughly, with the 26% exclusion rate, about 56 million Nigeria citizens between the ages of 15 – 64 are still financially excluded. That is still a lot of people and initiatives need to be put in place to drive a stringent financial inclusion.

The dynamic of Nigeria is predominantly an informal economy where most of the population is engaged in entrepreneurship in various sectors such as trading, agriculture, handy jobs, etc. In an economy where there is no proper infrastructure and incentives designed to enhance availability of job opportunities, most of the populace would have to create one for themselves. The Igbo ethnic group of Nigeria has an apprenticeship system which is championed as the world's largest business incubator (Havard Business Review, 2024) has a tradition of developing ventures and incentivizing people mostly family relatives into entrepreneurship. This eventually serves as a means of reducing poverty within the eastern region of Nigeria whilst championing education and prosperity.

Looking at the level of financial exclusion by regions in Nigeria, the Northen Region is the most affected. The 2023 Access to Finance (A2F) survey shows that financial exclusion is at 38% in the North-East and 47% in the North-West which is significantly higher than the South-West and the South-South at 5% and 10% respectively (Fakoyejo, 2023). One of the stated factors for this gap are population sizes compared to the southwest, poor network connectivity, and insecurity which leads to the complication in providing traditional bank services. Other stated impeding factors are religious and cultural differences.

In 2012, the National Financial Inclusion Strategy (NFIS) was born after the CBN joined the global network policy makers by signing the Maya Declaration in 2010 with a commitment to reduce financial exclusion in Nigeria by 2020 (CBN, n.d.). As of 2012 during the conception of the NFIS, financial exclusion stood at 46.3%. significant strides have been made to reduce that over the years. The stated targets to achieve by 2020 were to scale payments to 70%, credit, insurance, and payments by 40%, and savings by 60%. Other targets were to scale up Deposit Money Banks (DMB) branches by 7.6, microfinance bank branches by 5.0, cash machines by 203.6, point of sale devices by 850, and mobile agents by 62.0 all per 100,000 population respectively (CBN, n.d.). The focus of the NFIS was centered around several key areas that were considered as critical consideration and action points such as convenience, diversity, options, accessibility, affordability, and usage. During the implementation of this initiative, five obstacles were identified namely: illiteracy levels around financial services, high cost of services, lack of sufficient income, long distance to access locations, and complex client-onboarding processes by banks. Most of these obstacles can be reduced through effective application of digital technology and big data has a crucial role to play in this field.

2.2.6. Technological Innovation Trends in Financial Inclusion

Section 2.2.3, 2.2.4, and 2.2.5 have more like mentioned some technological trends in the field of financial inclusion. This section aims at summarizing all that in a more concise and structured manner. A widespread inclusive financial access will be unattainable without the application of technology. Here are some of the key technological trends shaping financial inclusion in emerging markets.

Mobile Money and Digital Wallets

The wide penetration and adoption of mobile phones in emerging markets such as Nigeria has enabled the possibility of mobile money services. This has allowed the underbanked population the opportunity of accessing basic financial services like saving, sending, and receiving money from all around the country thereby increasing convenience (Maurer, 2012). Examples include MTN Mobile Money operating in most MTN locations in Africa, etc., M-Pesa from Kenya, and T-Money from Togo. Mobile money initially started with enabling payments within the country, but advancements have been made to enable transfer internationally as is proven by the success of MTN. Mobile money comes at a reduced cost of operation compared to traditional banks which is more suitable to people in rural communities with low-income levels. The reduced cost of operation also benefits the financial institution or network operator as there is not much need for operating physical branches all over the country (Suri & Jack, 2016). Further services can also be unlocked through mobile money accounts such as microloans, micro investments/savings, and microinsurance (Demirgüç-Kunt et al., 2018).

Agent Banking and Branchless Banking

The non requirement of operating branches across the country as compared to traditional banks is one of the benefits of technology towards financial inclusion. Financial institutions can extend their services using physical agents often located in shops or kiosks to target and meet the demands of clients in remote locations (Klapper & Singer, 2017). This model is similar to what logistic companies like DHL and Hermes have adopted across Europe. Instead of investing millions building packet shops/packet stations across the country, why not just partner with shops which are more widespread and can be found within walking distance of any location. These agents can also be an incentive to offer financial literacy to users and serve as a spot to provide personalized assistance to clients with technical problems (Demirgüç-Kunt et al., 2018).

Cloud Based Technology and Big Data

Prominent cloud platforms such as Amazon Web Services (AWS), Microsoft Azure and Google Cloud offer organizations the ability to scale their business rapidly and efficiently to accommodate a growing customer base. Huge costs can be saved as compared to having an in-house database management system. AWS offers a pay as you go service which offers an incentive to minimize cost for organizations with fluctuating demands or limited budget (Gupta, 2022). With the growing number of data generated through digital activities, big data offers an opportunity to the financial institutions to process, analyze and make data-driven decisions like assessing credit worthiness, analyzing the market for potential customers in underserved regions, and developing targeted products and services (Kumar & McKay, 2020).

AI and ML

Clients do not always have to interact with company staff to resolve their problems. AI and ML can help lessen that time. Incorporating AI-powered chatbots can help ease the account opening process and offer basic assistance to frequently asked questions (FAQ). This also has the potential of making the website more user-friendly for people with limited financial literacy (Pillai et al., 2020). With the use of ML algorithms, financial institutions can analyze user data (ethically) to make predictions and design personalized products and services for clients and potential clients (Kumar & McKay, 2020). Alternative credit scoring methods can be developed through AI and ML, thereby offering a means to determine creditworthiness for the unbanked population. Data related to payment of bills, online purchases, and mobile account transactions should be able to give an insight into one's creditworthiness to access loans and other financial services (Bharadwaj, 2019).

Biometric Authentication and Digital Identity

One of the fundamental prerequisites to accessing financial services is the need for personal identification. Nigeria has had its share of complications with the introduction of digital identity. Issues relating to corruption, limitations in infrastructure, and privacy concerns have plagued its implementation. Despite all that, the government still has a commitment to leverage digital technologies to improve the identification process. Modern smartphones have biometric authentication like fingerprint and facial recognition which financial institutions can leverage to enhance security to their clients when accessing their financial platform (Gideon et al., 2019). So, there is no need to build one internally. All that is required is to make sure that this biometric authentication is available within the platform.

2.3. Theoretical Review

This section provides theoretical explanations which will provide an assistive understanding of the context of this research paper. The theories range from aspects such as technology, government, and literacy levels to how they influence or predict the development of financial inclusion. There are several theories but for the sake of this paper, only seven theories are reviewed as we believe they are more integral to the topic of discussion.

2.3.1. Technology Acceptance Model Theory of Financial Inclusion

Acceptance towards the use of innovative technologies can bring instant and long-term advantages to individuals and organizational entities such as cost efficiencies, improved performance, and financial gains. The Technological Acceptance Model (TAM) is popularly seen as one of the most important models for evaluating how people react to new technological developments (Marikyan & Papagiannidis, 2023). Initially proposed by Fred Davis in 1989, the purpose of the TAM model was to investigate how people's perceptions of the value and ease of use of a particular technology influence their likelihood of using that technology. People react differently to change, most especially in the aspect of technology and it is also dependent on the age and impact technology has on the environment. People above the age of forty are more likely to be resistant to change (Kunze et al., 2013). Younger people are more capable of adapting to change and it is reasonable considering modern technological innovations such as the internet and social media scaled up over the past 3 decades. Sectors such as healthcare, hospitality, finance, and transportation are reliant on technology to ease the various processes involved in the management of their daily activities.

There are two main factors that decide whether users will accept technology: perceived usefulness and the ease of use of the technology. User's subjective view on technology is universally recognized by the TAM theory as it places emphasis about the usefulness of a product based on how the users can interact with the said technology. Looking at the mobile banking technology in Nigeria as an example, the TAM posits that the usefulness and ease of use of the technology have a significant impact on its adoption and usage (Iwedi, 2023). In the aspect of software development, ease of use in this regard would be measured by the user experience. How readable, accessible, and intuitive is the platform? The underserved population is usually dominated by people with low financial literacy. Designing a product that can be used irrespective of their literacy levels would determine the potential of a product being a big hit in the market. When environmental stimuli are exposed to people, cognitive reactions are triggered, which in turn trigger affective responses, which in turn influence behavior regarding usage. In accordance with TAM, these mental reactions consist of how individuals evaluate the technology in terms of how helpful it is and how simple it is to use. On the other hand, subsequent research has shown that people's attitudes toward technology have a significant impact on their intentions to use it and the frequency with which they do so.

TAM offers valuable insight into the factors that influence individuals to use mobile banking services, which is important in the aspect of financial inclusion. Financial institutions can increase the likelihood of gaining the untapped market by offering tailored financial services that meet the needs and requirements of such market. This in turn will lead to significant adoption and usage rates among users.

2.3.2. Financial Literacy Theory of Financial Inclusion

The search for more theories surrounding the aspect of financial inclusion by Ozili (2020) has brought us to the theory of financial literacy. The theory is of the notion that financial literacy has the potential to increase people's willingness to enroll into the formal financial sector. Financial literacy according to the European Commission (2024) is defined as the knowledge and skills needed to make important financial decisions. According to the theory, individuals who are financially excluded will be more willing to incorporate themselves into the formal financial sector if they have a better understanding of financial literacy. Consequently, this indicates that the goal of achieving financial inclusion can be accomplished through education that raises the level of financial literacy among inhabitants. Once those who are financially excluded acquire the knowledge necessary to understand their financial situation, they will look for financial products and services wherever they can find them.

Ozili (2020) delved more into the advantages of financial literacy which have been summarized in four key points: awareness, adoption, financial stability, and lower campaign cost. Awareness refers to the ability of people to recognize and understand the series of financial products and services that are being offered by the financial institutions. One of the major problems with financial inclusion could be traced to the fact that most underserved citizens have no idea or have little knowledge of how to operate an account. The older generation are still stuck to saving and storing money in outdated methods such as under the bed or in unsecured safe boxes. This increases the risk of having their money physically, especially in situations when there is a fire outbreak. Having a formal bank account offers them a lot of protection. Adoption is described in this situation as taking advantage of other financial products such as investment and mortgages. Having a means to invest your money and earn some passive income should be an opportunity everyone should have access to, not just the rich but the poor. Increased investment in financial products such as stocks, bonds, securities should provide people with some level of financial stability and being self-sufficient. They should be able to manage their budgets by figuring out the difference between their wants and needs. And lastly, financial literacy is also an advantage to the government in the sense that it is much easier and cost effective to educate people on financial management instead of taking the upper hand towards building the entire infrastructure necessary for promoting financial inclusion across the nation. Instead of spending billions on infrastructure and products, companies can be incentivized to offer such services by introducing favorable policies such as lowering the minimum reserve requirement and the license cost.

There are also drawbacks stated in his paper on financial literacy theory of financial inclusion. The theory focuses on 'willingness' and not 'capacity'. Financial literacy could potentially increase literacy rate among societies but there is a major limitation with regards to how far they can go with applying that knowledge in the financial sector. Investing in financial assets requires money and not everyone is privileged to have enough money set aside for investment. They are willing but not capable. This is one of the important reasons why income inequality should be combated to promote a well-distributed and inclusive access to all opportunities.

2.3.3. Information Asymmetry Theory

When one party has more information than the other, there is inequality which could potentially lead to market failure (Ruan, 2019). This imbalance is categorized as information asymmetry. In relation to financial inclusion, information asymmetry exists in a case where the financial service providers have less information about the market or are unable to collect accurate information about the market segment. This lack of access leads to difficulties in figuring out the right products and services to market to such a region, leading to hindered access to financial services. There are 2 outcomes of information asymmetry which are adverse selection and moral hazard (Chen et al., 2019). Adverse selection is the situation whereby individuals who are riskier or have low credit scores are more likely to seek out financial services. Evaluating credit worthiness is an essential part of offering credit to applicants and a situation where there is insufficient data of such clients could greatly affect decision making. With respect to financial institutions who are eager to gather a large customer base, they might lose sight of this important requirement and end up offering said credit to the client. The result could be a loan default eventually, which will affect the financial performance of the institution. This is considered a moral hazard. Behavioral patterns tend to be consistent among adults and offering financial services to people who have a consistent pattern of unpaid bills would almost equate to a loan default. There are lower margins from offering financial services to remote locations or undeserved communities and strict measures must be put in place to ensure that the institution continues to operate into the unforeseeable future. By reviewing ML technologies and applications, Chen et al. (2019) demonstrates that this problem can be alleviated by engaging in data analytics activities for credit check assessment. Estimating credit worthiness could be achieved if data from diverse areas such as mobile phone usage patterns, digital payments, and mobile utility payments are assessed and processed.

2.3.4. Theory of Financial Inclusion and Income Inequality

Kling et al. (2020) discussed the theory that looks at the relationship between financial inclusion to income inequality. Financial inclusion in the article is measured by two metrics: access to formal loans and access to a bank account. When there is an unequal distribution of income among citizens, that is income inequality. For example, Nigeria has high inequality rate. For a country with the largest population and largest economy in Africa, the country is ranked 11th in West Africa and 100th out of 163 countries in the world in terms of wealth inequality (Uduu, 2022). There is a general misconception that financial exclusion would lead to continued inequality and this theory seeks to prove otherwise (Kling et al., 2020). The assertion from the theory suggests that even though financial inclusion may potentially not lead to reduced income inequality but could decrease underinvestment in education through the provision of financial services to the underbanked population.

One aspect of the theory looks at the relative importance of initial endowments towards shaping income inequality. Initial endowments according to Smith (2020) includes human capital (education and skills), financial assets and other inherited advantages that individuals possess at the start of economic journey. It posits that individuals with higher initial endowment may have a competitive edge over others in terms of income generation and wealth accumulation. Simply speaking, investing further in education may not lead to additional increases in income. Financial assets on the other hand continue to earn returns irrespective of amounts invested individually. The economic concept, the law of diminishing marginal returns comes into play here implying that financial assets are not subject to the law unlike investment in human capital.

The theoretical prediction of Kling et al. (2020) was assessed using data from the China Household Finance Survey (CHFS) for the years ranging between 2011 – 2013. An analysis of the real-world data helps to determine if the predictions hold through. The result of the findings showed a positive relationship between financial inclusion and underinvestment in education. The findings also showed a concerning result where it observed that reliance on formal and informal loans would increase income inequality. This result throws an imbalance of the expected relationship between financial inclusion and income distribution.

2.3.5. Access and Empowerment Theory

The core argument behind access and empowerment theory is that big data can play a vital role towards promoting financial inclusion by offering tailored financial products and services. This can be more effectively achieved if the financial institutions can determine the specific markets where such products can be offered. Exploiting data driven insights can increase access to savings, credit, and other financial assets and can empower people to participate in a more inclusive financial system.

A study by Ashraf et al. (2016) shows that financial inclusion is closely related to economic development and poverty reduction. It can be deduced that this theory is comparable to the Information Asymmetry Theory as one of the bases of the argument lies on access to the right information to guide decision making and enable better product and service offerings. The differentiating factor between the separate theories is empowerment. Whilst the information asymmetry theory focuses primarily on risk assessment, this theory takes it a notch further by looking at the broader societal benefits at large instead of profit and corporate efficiency. Financial institutions should also view the economy at large on how their services are contributing to increased economic growth, reduced income inequality, and enhanced social stability. Policy makers also play a crucial role in making sure that they create an enabling economy for the financial institutions to prosper. There should be a strong corporation between the two stakeholders to effectively promote access to formal financial services. Financial institutions running on its own without close supervision would lead to a deviation from their core duties to the society and would rather invest more efforts towards capitalizing on the marginalized market for their own gain or to meet shareholders' expectations.

2.3.6. Financial Friction Theory

There are people who do not want to invest any efforts towards going to a financial institution to access the available products and services. In most cases, it could be that they already have first-hand experience of the required deliverables, and they are not willing to go through all those processes due to the expectation of invested time. Lengthy application processes, high fees, and lack of awareness create friction that deter people from accessing financial services (Beshears et al., 2015). The financial friction theory evaluates the barriers which are predominantly imposed by financial institutions denying interests of majority people from low-income backgrounds. Financial institutions need to understand that different markets require a certain level of product differentiation. It is not a sound business decision to take products and services which are sold in urban or major cities to be sold to people in rural communities. There are different tastes and purchasing power that each customer has which determines the value of a product or service. The value of service required is determined by the income levels of everyone. Minimal effort towards product differentiation leads to a perpetuating cycle of financial exclusion and limited access to economic opportunities.

Demirgüç-Kunt et al. (2018) highlighted the role of big data towards enhancing financial inclusion as evidenced by the Global Findex Database which highlights the impact of fintech innovations on expanding access to financial services. It is imperative that big data analytics are nurtured and applied in the financial sector to reduce these frictions. Leveraging data driven insights gives financial institutions the ability to modernize application processes based on required documents, personalized offerings based on target market, and improve financial literacy through targeted educational initiatives. The basic principle of the Financial Friction Theory stresses the importance of identifying these barriers that could hinder financial access by introducing technological innovations such as big data. Minimizing these frictions will potentially aid policy makers and financial institutions to make financial services readily available and accessible whilst also making sure that they are affordable to people in the marginalized communities. The result after attaining these barriers would be an increased economic opportunity and social inclusion.

2.3.7. Intervention Fund Theory of Financial Inclusion

The last theory on this paper is anchored on the intervention fund theory of financial inclusion which was developed by Ozili (2020). It is of the argument that the success of a well-established and evenly spread financial inclusion activities and programs highly depends on the interventions of a series of special funders instead of reliance on the government who use taxpayers' money for these initiatives. Special funders in this regard consists of philanthropists (Bill & Melinda Gates Foundation, Tony Elumelu Foundation, Mo Ibrahim Foundation, etc.), NGOs (Financial Alliance for Women, Accion, Kiva, etc.), and foreign governments (USAID, DFID, World Bank, etc.). These special funders often have a more structured way of operation which eliminates bias. Government organizations as widely seen in Africa or most especially in Nigeria are plagued by high levels of corruption. Diversion of funds for personal gains is something that the Nigerian economy continues to experience. Dependence on these governmental organizations could hinder the growth of financial services access. Having access to funding from overseas countries significantly increases the overall capital required to invest in financial inclusion initiatives. Policy makers should create an enabling environment for them to develop and implement their solutions. By having multiple establishments, these organizations can introduce varying financial inclusion programs thereby promoting competition in targeting potential customers in the underserved population.

There are three main advantages that come with the intervention fund theory of financial inclusion which Ozili (2020) stated in his paper. Firstly, the allocation of funds would not have to pass through the time-consuming political bureaucracy which plays an integral role of reducing the time in implementing financial inclusion projects. Secondly, the freedom to operate independently gives the special funders access to the domestic and foreign capital markets, thereby offering them high flexibility with investment projects. Thirdly, the special funders have the capacity to create new sub institutions to develop and solve specialized issues. An example could be the establishment of an IT institution created to aid research in digitalized products whilst outside of the core operation.

Ozili (2020), also mentioned some disadvantages associated with the intervention fund theory of financial inclusion. To begin with, since special funders operate without the intervention of the government, they would have to devise methodological approaches to determine which segment of the economy is financially marginalized. Secondly, the methodology that the special funders must devise may not accurately determine which segment of the population is marginalized. This would inherently lead to failed financial inclusion programs and a financial loss which would be difficult to recover from. Lastly, the use of funds from foreign donors or government could create a negative representation of the domestic government. It would be perceived that the local government are incapable of solving its financial exclusion problems. A country having a negative reputation can lead to a series of effects such as difficulty attracting FDI, reduced tourism, increased borrowing costs, damaged relationships with other countries, etc. (Kahn, 2016; Center for Global Development, 2019). A more comprehensive approach to mitigate demerits of this theory would be a coexistence between the government and the special funders. But then again, that will in turn change the merits of this theory into demerits.

2.4. Empirical Review

There have been a considerable number of studies in the field of financial inclusion, and they focused on certain areas such as the impact of AI, ML, digital technology, innovation, blockchain, and including big data. A review of findings from this past research will be presented in this section.

There is a positive correlation between the adoption of AI and big data on financial inclusion according to the findings of Ozili (2023). Their finding purports that better products and services can be developed by financial institutions targeted at the underbanked population if these technologies are efficiently and accordingly implemented. Their paper does not clearly state the methodology or research approach used in the study, but a general overview indicates finding based on review of previous research within the topic of big data and financial inclusion. Strong emphasis was made to evaluate the merits and challenges of big data and AI on financial inclusion of which several findings were made. Among the benefits of AI and big data on financial inclusion especially towards improving efficiency and risk management process include the provision of smart financial products and services to qualified adults and the simplification of the account registration/opening process for the unbanked population. The issues found in the paper include the cumbersome procedure towards training AI algorithms to process large data generated from customers of financial services providers. Shortage of skilled labor with regards to AI is another problem which was identified. The annual salary range for AI experts in Nigeria is between 18 – 20 million Naira (*Intel Region*, 2023). That is a significant amount of money to pay whilst considering other key roles like data scientists, data engineers, and data analysts who all co-exist to produce data driven insights. For a financial institution targeting the unbanked population, the profit margin is low making it difficult for such institutions to be able to afford these experts. Another problem identified in the paper is that the adoption of AI will lead to increased unemployment in the financial sector. AI models and applications are also found to be biased in their design which could lead to financial exclusion towards vulnerable people such as the elderly and disabled people. Lastly, privacy concerns are one of the issues that needs to be addressed by introducing strict privacy laws which could further hinder the potential of AI's growth.

Using a combination of traditional credit scoring models and innovative big data sources, Óskarsdóttir et al. (2019) studied the impact of using mobile phone data and social network analytics on enhancing financial inclusion. They examined the use of call-detailrecords (CDR) data. Alternatively speaking, the paper focused on one determinant of financial inclusion which is credit score. By monitoring call logs, models can be designed that help to predict the creditworthiness of applicants from the unbanked population. The research also touched base on the importance of applying strict care with the data generated to avoid breach of privacy laws and infringement of customer rights.

Agarwal et al. (2019) examined the role of big data and machine learning on financial inclusion and how they can produce alternative credit scoring for elderly people without a traditional credit score. Quite akin to the study by Óskarsdóttir et al. (2019), their paper adopted a statistical approach with data obtained from a top fintech company in India (no specific company mentioned). Datasets which include loan default rates were collected and analyzed. The use of simple logit regression and advanced machine learning algorithm was applied in the study to predict individual mobile footprint and deeper social footprint based on call logs. The findings show that mobile and social footprint factors have a significant effect in evaluating credit risk. Also, the findings of the study indicated that fintech lenders can alternatively generate credit scores using these alternative data thereby expanding access to credit to people with little or no credit history.

Despite the study of Richard & Shabir (2023) having no mention of financial inclusion in the paper, it is still important to review their findings as it has a close relation to the topic of this research. Their paper seeks to evaluate the importance of big data in the provision of financial services. To conveniently analyze the data and concepts of the topic, they employed a mixed-method research approach. Both qualitative and quantitative methods were incorporated to produce a comprehensive overview of the relationship between big data and innovation in the delivery of financial services. The findings of their paper shed light on 3 key segments: decision making through strategic means, client-driven innovation, and the identified challenges. Looking at strategic decision making, the overall finding shows that big data can help enhance the quality of decision making. On the second aspect, strict focus on customer needs can help unlock new products and services through the involvement of data analytics. Personalized financial products and services can be created, onboarding processes can be improved, and the institutions can easily adapt to changing market demands. Lastly, the challenges identified in the paper include data privacy, regulatory hurdles, and the need for skilled labor.

Qureshi (2020) conducted a thorough review of previous research to investigate why data is relevant for further development in financial inclusion. The paper illustrates the need to explore ways in which new markets are being created with regards to products of data analytics. How data is extracted, analyzed, and commoditized should be of major concern to financial institutions. The findings show that big data can address several social economic issues such as poverty, food security, and climate disasters efficiently. Involvement of big data in the financial sector to address the barriers associated with the underserved population could also lead to potential barriers such as a data divide where the affected population are left out from financial products and services from the solutions devised from big data. Data models are reliant on algorithms and there is a tendency for these algorithms to deviate from their intended purpose thereby creating an even further barrier.

With the aim of assessing the digital disruption from big data of the South African financial services sector focusing on financial inclusion, Mungai & Bayat (2018) used a mixed method approach in their study. The collection of data from semi-structured interviews with the participation of key informants in the financial services sector and related fields of knowledge were conducted for the qualitative method. For the quantitative approach, secondary data within the years 2011, 2014 and 2017 were collected from the World Bank Findex database to examine trends in the financial activities of adults across a specified number of countries. The result of their findings shows that financial inclusion can be measured through the ownership of different financial channels such as debit cards, credit cards, and mobile money accounts. On the aspect of big data, it shows that it has a significant impact on financial inclusion with a score of four out of seven. It posits that alternative data sources enable the extension of financial services to individuals and households that were previously considered to be un-bankable. The main reason based on the findings why the un-bankable have no access to financial services is not because they are bad clients but simply because the financial service providers have low visibility on such population due to data insufficiencies.

How et al. (2020) adopted a Bayesian approach to model the performance levels of systems which are proven engineering techniques to measure the possibility of relationships between theoretical constructs. Their paper's primary objective is to show how a financial inclusive product or service provider in a developing economy could use AI to analyze data from an existing legacy database to find potential client segments to approach with financial inclusion offerings. Their findings revealed that data driven predictive modelling using AI can help financial service providers to learn more about their prospective customers, help towards the provision of meaningful insights and the delivery of relevant financial products. Also, AI fosters economic growth, societal progress, and enhancements of human financial stability.

By using a conceptual and documentary analysis method, Mhlanga (2020) researched the effect of AI on financial inclusion. Findings of the study show that AI poses a strong impact on financial inclusion depending on which field it is applied. Such fields include risk detection, management, measurement, information asymmetry, fraud detection, cybersecurity, and attending to customer needs via chatbots. The paper suggested that all key stakeholders responsible for enhancing financial inclusion such as financial and non-financial institutions and the governments across the globe should invest towards scaling up the application of AI in finance for the benefits of maximizing access to financial services, especially in the unbanked populations.

Mhlanga (2024) also conducted a recent study where they evaluated the importance of big data in FinTech and its impact on enhancing financial inclusion. Research on this topic necessitated the review of related materials of which various sources were incorporated such as industry reports, online content, and most importantly academic papers. The overall result of their findings shows that big data plays an essential role in the development of new financial products and services in the sense that financial institutions can analyze customer behavior to identify patterns which helps in making informed decisions about their worthiness in accessing certain financial services that they offer. Not only will they be able to create new products and services targeted to the marginalized population, but financial institutions will also be able to improve efficiency in operations and better manage risk which are crucial factors to enhancing a more financial inclusive society.

By reviewing academic materials related to the topic of big data and financial inclusion, Adeoye et al. (2024) stated the success stories of Tala, M-PESA, and JUMO in their respective markets which was attained from their adoption of AI and data analytics into the business process. There result shows evidence that big data is integral in the advancement of financial products and services access in the underserved population. Although AI and big data show significant efforts towards improving financial inclusion, it is faced with a few challenges. How do financial institutions ensure that customers data are used in the right manner? Is there clear communication between the two parties on what kind of data is accessible or not? What about the biases that may arise from AI algorithms? Addressing these questions goes a long way to ensuring that AI and big data goes a long way towards ethically offering sustainable insights which are essential for management decision making in products and service offerings.

Using an empirical review of related literature, the study of Falaiye et al. (2024) looked at the role of technology on financial inclusion. various technological advancements in finance such as mobile banking, blockchain, and digital wallets are said to be one of the rising trends in technology which have aided in the further developments of delivering financial products and services to the neglected population by traditional financial institutions. Among these technological advancements, their findings shows that mobile banking in the form of mobile money services is the key factor fostering easy access to finance. This is because of the ease and convenience of using mobile devices irrespective of what capability (smart or dumb) has in conducting basic transactions such as money transfer, payment of bills, and the application of loans.

Finally on the section of our empirical review, Hollanders (2020) evaluated the joint project between the Committee on Payment and Market Infrastructures (CPMI) and the World Bank to investigate the potential of FinTech for improving access to and use of transaction accounts. They focused on reviewing the Payments Aspect of Financial Inclusion (PAFI). The findings show a positive relationship between access to payments and financial inclusion. Also, among the findings emphasize the role of big data in helping fasten the process of onboarding new customers through screening processes or biometric authentication methods.

2.5. Gaps in the Literature

From the empirical review, it shows that there is minimal research that relates to big data and financial inclusion. The reviewed papers mostly delved into topics like AI, FinTech and Digital Technology. This paper has taken a more specific approach. Whilst big data is a sub section of AI, it is importance that this paper look at this area as it reduces the scope of the study and give a direct approach to how big data plays a significant factor towards enhancing financial inclusion in the underserved markets. Studies of the literature reviewed mostly covers other countries with economic structure different from that of Nigeria or other developing economies in Africa. Whilst the insight from their studies is still beneficial for this paper, it would be more proper to focus on only studies around countries of the same region to conveniently draw comparisons across different performance metrics. Most of the literature reviews did not clearly state the empirical methods used nor did they conduct any empirical analysis. This poses a problem of statistically measuring results across findings.

3. Methodology

3.1. Introduction

From the previous chapter we discussed the theoretical and empirical facts of the relationship between big data and financial inclusion based on previous research. The results from a review of the literature are used to establish expectations for the relationship between these variables. Therefore, the purpose of this chapter is to present the methods of analyzing the research questions and hypotheses, the underlying principles of research methodology and the choice of the appropriate research method for the thesis. The chapter is organized as follows: Section 3.2 presents the research design adopted. Section 3.3 discusses the sources of data and method of data collection. Section 3.4 presents the description of variables and their measurements. Section 3.5 presents the specification of the model adopted. Finally, section 3.6 looks at the method of data analysis of the experimental relationship.

3.2. Research Design and Approach

For this study, we will be adopting solely a quantitative research approach. More specifically, an observational research design will be used to examine the relationship between big data and financial inclusion. The choice of quantitative method is because datasets relating to variables associated with financial inclusion and big data will be analyzed which requires the need for statistical tools. The nature of this study requires mathematical interpretation of the relationship between big data and financial inclusion. Quantitative research methods are objective in nature making it a more reliable means of researching the occurrence of a phenomenon. Professionals such as corporate managers, sociologists, doctors, scholars, etc. rely more on quantitative research to make sound and testable decisions (Heale & Twycross, 2015).

3.2.1. Study Population and Sampling Technique

The purpose of this study is to focus on the Nigerian economy. However, due to the limited availability of consistent yearly data, the study population has been expanded to developing countries in Africa including Nigeria of which statistical inference will be drawn based on the findings to explain the relationship of variables for Nigeria.

To empirically evaluate and analyze the hypothesis relating to the research topic, this paper will rely solely on secondary data. Secondary data refers to data that was originally collected for purposes different from the current study objective. The data is readily available and accessible for others to use via various platforms. The major advantage to secondary data as stated by (Hox & Boeije, 2005), is the large-scale format of datasets much larger than one could primarily collect themselves.

This paper will employ convenience sampling, a type of non-probability sampling. Subjects are chosen for convenience sampling based on how easy it is for the researcher to reach them. For this paper, all African countries have been selected as the target group. The number of countries for each analysis is majorly dependent on the availability of key data variables on financial inclusion. We will be analyzing the relationship between financial inclusion and big data indicators such as access to the internet, access to mobile phones, digital utility payments. Statistical performance indicators (SPI) such as data use, data services, data product, data sources and data infrastructure of a country will also be analyzed to explain its impact on financial inclusion.

3.2.2. Time Frame

As previously mentioned about the unavailability of consistent yearly data, this paper will adopt a cross-sectional data analysis. According to Thomas (2023), the major advantage of cross-sectional analysis is due to its low-cost access to data as variables for a single period are more easily accessible as compared to longitudinal data. Based on the datasets obtained so far, the newest available data are from 2021, hence why the study will be focusing on only variables from that year.

3.3. Data Sources, Data Collection Methods, and Type of Study

Based on the topic of financial inclusion, this paper provides an approach to understand the relationship between big data analytics and financial inclusion from the datasets hosted by the World Bank DataBank. There are two main databases that serve as a foundation for the required datasets extracted for analysis: the Global Financial Inclusion (Global Findex) Database and the SPI Database. The datasets were last updated in 2021 which sets the timeframe for the variables analyzed. The Global Findex dataset comprised 1,450 rows whilst the SPI dataset extracted has 625 rows. Not all data are used due to missing values for certain countries. To maintain reliability with the results, countries with missing values were omitted for the analysis. After a series of data cleaning and restructuring, we are left with 40 countries for analysis on Global Findex variables and 35 countries for SPI variables.

Regarding the study methodology, the nature of this paper requires adopting the mono method of data collection and analysis. With the mono method, the analysis of the variable is sufficient since it is quantitative data. It is considered as a more convenient method to evaluate the research questions and hypotheses (Ojebode et al., 2018).

Table 1: Global Findex DataFrame Codebook and Variable Descriptions

Labels	**Description**
internet_access	% of respondents above the age of 15 who has access to the Internet
mobile_utility_payment	% of respondents above the age of 15 who made a utility payment: using a mobile phone
digital_payment	% of respondents above the age of 15 who made or received a digital payment
mobile_money_account	% of respondents above the age of 15 who has a mobile money account
mobile_phone_ownership	% of respondents above the age of 15 own a mobile phone
mobile_bill_payment	% of respondents above the age of 15 who used a mobile phone or the internet to pay bills

Source: Framed by Author

Table 2: SPI DataFrame Codebook and Variable Descriptions

Labels	Description
account	% of respondents above the age of 15 who has any form of financial account
source_data_capacity	Source data assessment of statistical capacity (score 0 - 100)
data_use	SPI Indicator: Pillar 1 data use score (0-100)
data_services	SPI Indicator: Pillar 2 data services score (0-100)
data_products	SPI Indicator: Pillar 3 data products score (0-100)
data_sources	SPI Indicator: Pillar 4 data sources score (0-100)
data_infrastructure	SPI Indicator: Pillar 5 data infrastructure score (0-100)

Source: Framed by Author

3.4. Variables Description and Measurement

The subsequent sections present the selected dependent variable as a proxy of financial inclusion and the independent variables as a proxy of big data sources and economic data performance.

3.4.1. Dependent Variable

Financial Inclusion is one of the ultimate goals towards ensuring equality and access to all resources in an economy. A deviation in the financial system where certain regions or demographics are indirectly disallowed access to basic financial needs such as money transfer, payments, savings or investment, credit to mention but few could pose an economic challenge where there is income inequality and a negative effect on GDP per Capita. It is the duty of all key stakeholders within the public and private sector to ensure that access to finance is recognized as a fundamental right.

The dependent variable that will be studied in this paper is financial inclusion. Variables for measuring financial inclusion were obtained from The World Bank Group (2024) Global Financial Inclusion Index simply named as the Global Findex database. The

Findex database is a valuable tool to assess all data related to the access and usage of financial inclusion products and services across different economies. It also considers both digital and non-digital metrics of financial inclusion. This paper relies more on the digital aspect of financial inclusion as we are evaluating the importance of big data which is data generated from usage of digital devices. After a review of relevant literature, we can deduce that financial inclusion can be measured by different indicators explained below:

I. Ownership of Financial Account: The use of formal bank account to access all kinds of financial services such as loan, insurance, savings, etc.
II. Access to Credit: The proportion of people who have access to loans or other lines of credit like mortgages.
III. Mobile Money Account: A sub-section of financial accounts that describes access to mobile money accounts targeted at the underbanked population due to their ease of use and minimum credential requirements.
IV. Access to Financial Digital Payments: Encompasses all aspects of payments made through digital channels like mobile money, digital wallets, credit, or debit cards, etc.
V. Savings at a Financial Institution: The proportion of individuals who have deposits or saved in any form of financial institution.

There are two main financial inclusion indicators which will be adopted in this research. The reason being that we will be evaluating relationship in two spectrums: big data source variables which serve the key foundation on this paper, and economic data availability variables which measure the quality and accessibility of economic data.

Dependent Variable – Model I: Mobile Money Account

The study of Khera et al. (2021) used mobile money account as a proxy for financial inclusion. The introduction of mobile money as an alternative source of finance was to drive access to basic financial services in rural areas. Therefore, it is essential to use the adoption of this as a metric to measuring financial inclusion. The Global Findex database provides data on the adoption rate of mobile money account measured as a percentage of a country's population. The data variable was obtained from the study of adults from the age of 15 and above.

Dependent Variable – Model II: Account Ownership

Account ownership is defined as individuals who have access to or own any type of financial account be it via traditional or non-traditional financial institutions (World Bank, 2024). Traditional financial institutions consist of commercial banks whilst the non-traditional financial institutions consist of mobile money service providers (Alade, 2017; Daasi, 2012). The proxy for measuring financial inclusion according to Ozili (2021) is access to an account. The Global Findex database provides data on the access of accounts measured as a percentage of a country's population. The data variable was obtained from the study of adults from the age of 15 and above.

3.4.2. Independent Variables

This subsection describes the independent variables that are used in the empirical model to estimate relationship with the dependent variables. Following prior research into the determinants of financial inclusion, the independent variables are classified into two categories: big data source variables and economic data variables. To better evaluate the relationship between them, the variables will be split into two models. The selected proxy variables for Model I are access to the internet, utility payments with mobile phones, usage of mobile phones, bill payment with mobile phone or internet, and digital payments across different channels. The selected proxy variables for Model II are assessment of statistical capacity, data use score, data services score, data products score, data sources score, and data infrastructure score.

Independent Variables – Model I

This model encompasses various access to big data sources. Digitalization and innovation have led to the generation of big data from the daily use of digital devices like mobile phones, personal computers, smart TVs, smart wearables, e-readers, etc. The ability to collect user data ethically from these devices gives organizations an advantage towards developing targeted products that better suit the needs of its customers and potential customers. According to data from the World Bank (2024) Findex Database, these are the selected digital factors that influence financial inclusion. All data variables were obtained from the study of adults from the age of 15 and above.

I. Access to the Internet

Having access to the internet is one of the fundamental signs of a progressive society. It gives people the ability to exchange information and source knowledge from across the globe. In the process of actively participating on the internet, various data are generated be it streaming contents from Netflix, monitoring your health activity with smart watches, purchasing products and services from ecommerce platforms like Amazon, or your usual social media activity on platforms like X. This metric necessitates the need to include in the empirical model which is devised for Model I. The Global Findex database provides data on internet access measured as a percentage of respondents who report having access to the internet.

II. Utility Payment with Mobile Phone

Ozili (2022) and How et.al (2020) in their paper, demonstrated the importance of mobile phone usage as an influencing factor of digital financial inclusion. This variable focuses on one aspect of payment which is utility. It is of the expectation that by monitoring utility payment data from mobile devices, financial institutions can determine the credit worthiness of potential customers. The Global Findex database provides data on utility payment with mobile phones measured as a percentage of respondents who report making regular payments for trash collection, water supply, or electricity using a mobile phone.

III. Incoming or Outgoing Digital payments

Aiforgoodstg (2023) stated the importance of mining data from digital financial service providers using AI. Leveraging big data could help to automate the business process and provide a means to tackle certain obstacles in providing financial services to the masses. We can analyze the effect that digital payments have on financial inclusion. The Global Findex database provides data on the usage of digital payments measured as a percentage of respondents who report using a mobile phone or the internet to make a payment from an account.

IV. Ownership of a Mobile Phone

The Global Findex database provides data on access to a mobile phone measured as a percentage of respondents who report owning a mobile phone. This variable is peculiar for the study as it encompasses a variety of different metrics such as mobile money access, internet access, social media access, payments, etc. Owning a mobile phone serves as a foundation to a digitalized financial inclusion initiative.

V. Bill Payment with Mobile Phone or the Internet

The Global Findex database provides data on the usage of mobile phone or the internet to pay bills measured as a percentage of respondents who report using a mobile phone or the internet to pay bills in the past year. Only adults from the age of 15 and above are considered in the estimation.

Independent Variables – Model II

Big data can be accessed and processed directly by financial service providers using AI and ML tools. However, there are other forms of data which are necessary to better understand the financial market of a country and discover discrepancies in the level of financial products and services offered across the nation. As mentioned in the previous chapter, there are different stakeholders responsible for enabling an inclusive financial system. The government is a key player in this regard. The dissemination of financial products and services would not be viable if the government does not have quality data available for analysis such as investment in financial infrastructure, financial literacy levels, mobile money usage, regulatory environment, remittance flows, resource allocation, etc. These are secondary data which are essential to accompany study on big data variables.

To avoid broadening the scope of this study, we will be looking at the state of data readiness and quality which is measured by the Statistical Performance Indicators (SPI). The SPI is a quantitative means of depicting the ability of countries to provide efficient and testable national statistics systems, including an indicator that measures country's overall statistical capacity (World Bank, 2024). The indicators which are used in this paper as obtained from the World Bank SPI database are as follow:

I. **Statistical Capacity**

The source data assessment of statistical capacity measures the availability of administrative system data and whether data collection activities follow internationally recommended periodicity which are both reflected in the source data indicator. The score is generated from calculating the weighted average of the individual pillars which constitutes data use, data services, data products, data sources, and data infrastructure. Therefore, the data capacity gives an overall score of the statistical performance of a nation with all the specific individual measures considered.

II. **Data Use**

Demand for national statistical data is measured by evaluating the data use score. The data use pillar comprises five distinct user groups: legislators, governors, civil society (including sub-national entities), academics, and international organizations. To gauge progress, there would be corresponding metrics for each dimension. In contrast to a less developed system, which would have lower scores along certain dimensions, a fully developed system would perform admirably on all of them. The range of scores is between 0 – 100 whereby 0 is the lowest score and 100 the best score.

III. Data Services

This non-financial information is mostly sourced from the public data hosted by the bureau of statistics of a nation. The data services score for SPI: Pillar 2 is a composite that considers dimensions related to data access, online access, advisory and analytical services, and data releases. The range of scores is between 0 – 100 whereby 0 is the lowest score and 100 the best score.

IV. Data Products

The third pillar of the SPI places more focus on the SDGs. The score depicts the performance of a nation towards producing relevant indicators required to track the performance of SDG initiatives. Using the SDG typology, the data products (internal process) pillar is divided into four sections: social, economic, environmental, and institutional. The range of scores is between 0 – 100 whereby 0 is the lowest score and 100 the best score.

V. Data Sources

The data source overall score measures the extent to which data is available from sources such as administrative data, census and surveys, geospatial data, private sector, and citizen generated data. The range of scores is between 0 – 100 whereby 0 is the lowest score and 100 the best score.

VI. Data Infrastructure

An effective statistical system requires hard and soft infrastructure, which the data infrastructure pillar score measures. Laws and governance cover the existence of laws and a functioning institutional framework for the statistical system; standards and methods address compliance with recognized frameworks and concepts; skills include statistical literacy; partnerships reflect the need for an inclusive and coherent statistical system; and finance mobilized both. The range of scores is between 0 – 100 whereby 0 is the lowest score and 100 the best score. The World Bank SPI Database provides access to this data across all countries.

Table 3: Model I - Variables and Their Expected Relationship

Variables	**Notation**	**Expected Relationship with FI**
Dependent Variable (Financial Inclusion)		
mobile_money_account	MMA	N/A
Independent Variables (Big Data Sources)		
internet_access	IA	(+)
mobile_utility_payment	MUP	(+)
digital_payment	DP	(+)
mobile_phone_ownership	MPO	(+)
mobile_bill_payment	MBP	(+)

Source: Framed by Author

Table 4: Model II - Variables and Their Expected Relationship

Variables	Notation	Expected Relationship with FI
Dependent Variable (Financial Inclusion)		
account	ACC	N/A
Independent Variables (Data Readiness)		
source_data_capacity	SDC	(+)
data_use	DUS	(+)
data_services	DSS	(+)
data_products	DPS	(+)
data_sources	DSC	(+)
data_infrastructure	DIS	(+)

Source: Framed by Author

3.5. Model Specification

The multiple regression method was employed to assess the relationship between the predicting variables and financial inclusion. Modeling is based on cross-sectional data analysis technique. The specification of a model is based on the available information relevant to the study in question. This is to say, the formulation of an empirical model is dependent on available information on the study as embedded in standard theories and other major empirical works, or else, the model would be theoretical. The general model to be estimated is the following linear forms which are modified from Ozili (2022).

$$\Pi i = \alpha + \Sigma \beta k \, Xni + \varepsilon i$$

Where:

Πi is the dependent variable measured by the financial inclusion index,

i represent country,

α is a constant term or the intercept (the value of financial inclusion when all independent variables are 0),

Σβ is the summation of coefficients for the dependent variables,

Xni are k explanatory variables, superscript n denotes the number of observations, and

εi is the disturbance or error term.

To effectively account for the models developed for this study, the following equations are proposed based on the specified variables:

Model I Equation:

$$MMAi = \alpha + \beta 1(IAi) + \beta 2(MUPi) + \beta 3(DPi) + \beta 4(MPOi) + \beta 5(MBPi) + \epsilon i$$

Model II Equation:

$$ACCi = \alpha + \beta 1(SDCi) + \beta 2(DUSi) + \beta 3(DSSi) + \beta 4(DPSi) + \beta 5(DSCi) + \beta 6(DISi) + \epsilon i$$

3.6. Data Analysis Method

Python programming language will be used as the data analysis tool for this paper. The statistical software tool used is Anaconda and Jupyter Notebook. Anaconda as a software package provides access to foundational open-source data science packages without having to worry about individually managing the packages. The two fundamental programming languages used in data science are Python and R which are already in-built into the software. Jupyter Notebook on the other hand is a software that provides an IDE or a text editor environment for codes to be created. The choice of using these tools is to minimize the size of files used for the analysis since the data gathered are exported in CSV file formats.

3.6.1. Regression Analysis

For this study, the regression analysis known as Ordinary Least Squares (OLS) will be used to estimate the relationship between financial inclusion and the data variables. The regression coefficient will be analyzed for the two models to evaluate the extent to which changes in the dependent variable are explained or predicted by the independent variables.

3.6.2. Pearson Correlation Test

To measure the association between all tested variables, coefficient of the variables is obtained through the inclusion of the Pearson Correlation Test (Liu et al., 2020). The correlation coefficient ranges from -1 to 1, where -1 explains a perfect inverse relationship, 0 explains no relationships at all, and +1 explains a perfect direct relationship.

3.6.3. Multicollinearity Test

When two or more independent variables in a regression model are highly correlated against each other, there is a sign of multicollinearity. A situation like this makes it difficult to assess the individual impact of each variable on the dependent variable (Bhandari, 2024). To detect the presence of multicollinearity, the Variance Inflation Factor (VIF) is employed. The criteria for determining if multicollinearity exists depends on evaluating the values of VIF if they fall within a safe range or exceed the range. The VIF ranges from 1 and has no limits. Therefore, the maximum value can be endless. A value of 1 indicates that there is no sign of multicollinearity, a value within the range of 5 – 10 shows moderate signs of multicollinearity but is acceptable, and a value exceeding 10 shows troubling signs of multicollinearity. The formular for estimating VIF is as follow:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Where:

R^2_j is the multiple R^2 value obtained by regression estimate, and

j is the specific independent variable.

3.6.4. Descriptive Statistics

Descriptive statistics were used to summarize and analyze the cross-sectional data collected for the period 2021. Elements of the descriptive statistics employed in this research are the mean, standard deviation, standard error, skewness, and kurtosis.

3.6.5. Hypotheses Testing

Finally, the hypotheses of this paper will be assessed based on the statistical results of the previous section. Inferences will be made to determine if there is a relationship between financial inclusion and big data in the Nigerian market. Estimates whether to accept or reject the alternative hypothesis will be made.

4. Analysis and Findings

4.1. Introduction

The preceding chapter presented the research methods adopted in the study. The purpose of this chapter is to present the results of the different statistical tests conducted. Appropriate tables are presented to facilitate the discussion of the results. The current chapter has five sections which present results of the statistical tests for Model I and Model II. The chapter begins by analyzing the regression results to express the relationship between the variables. The result of the correlation test between the dependent and independent variables will follow. The correlation between the independent variables through a multicollinearity test will be analyzed. The results of descriptive statistics will follow and finally, hypothesis testing.

4.2. Regression Analysis

The following regression outputs exhibits whether the coefficient is negative or positive. The coefficients indicate each variable’s level of influence on the dependent variable. R squared (R^2) values indicate the explanatory power of the model and in this study Adjusted R squared (Adjusted R^2) value which considers the loss of degrees of freedom associated with adding extra variables were inferred to see the explanatory powers of the models.

As presented in the third chapter the empirical model used in the study to identify the relationship between the dependent and independent variables is provided below for the two models:

Model I: $MMAi = \alpha + \beta 1(IAi) + \beta 2(MUPi) + \beta 3(DPi) + \beta 4(MPOi) + \beta 5(MBPi) + \epsilon i$

Model II: $ACCi = \alpha + \beta 1(SDCi) + \beta 2(DUSi) + \beta 3(DSSi) + \beta 4(DPSi) + \beta 5(DSCi) + \beta 6$ $(DISi) + \epsilon i$

4.2.1. Model I Regression

Model I measure the effects of big data through internet access, mobile utility payments, digital payments, mobile phone ownership, and mobile bill payment on financial inclusion (measured by ownership of mobile money account). The results of the regression model are illustrated in table 5.

Table 5: Regression Summary For Model I

OLS Regression Results

Dep. Variable:	MMA	R-squared:	0.858
Model:	OLS	Adj. R-squared:	0.837
Method:	Least Squares	F-statistic:	41.18
Date:	Tue, 16 Apr 2024	Prob (F-statistic):	1.77e-13*
Time:	17:40:38	Log-Likelihood:	-135.17
No. Observations:	40	AIC:	282.3
Df Residuals:	34	BIC:	292.5
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-7.4068	7.749	-0.956	0.346	-23.155	8.341
IA	-0.5527	0.125	-4.411	0.000*	-0.807	-0.298
MUP	0.7147	0.415	1.721	0.094**	-0.129	1.559
DP	0.6621	0.119	5.579	0.000*	0.421	0.903
MPO	0.2944	0.161	1.824	0.077*	-0.034	0.622
MBP	0.0607	0.396	0.153	0.879	-0.744	0.865

Omnibus:	3.825	Durbin-Watson:	2.053
Prob(Omnibus):	0.148	Jarque-Bera (JB):	2.562
Skew:	-0.484	Prob(JB):	0.278
Kurtosis:	3.775	Cond. No.	618.

Source: Computed using Python 3.12 (2024). *Note*: * and ** indicate significance at 5% and 10%, respectively.

We begin by assessing the overall significance of the regression model. The overall p-value of the model alternatively represented as the Prob (F-Statistics) is 0.0000 which is below the 0.05 significance level. This means that the model is statistically significant as the independent variables collectively are good explanatory variables of financial inclusion. The R-squared and the adjusted-R squared of the model are 85.8% and 83.7%, respectively. Based on the result of adj. R, it indicates that the changes in the independent variables explain 83.7% of the changes in the dependent variable. The remaining 16.3% of changes were explained by other factors which are not included in the model. All independent variables except internet access have a positive regression coefficient. Meaning that a change in the value of the positive variable will lead to a direct change in financial inclusion. Internet access on the other hand will cause an inverse change in the value of financial inclusion.

4.2.2. Model II Regression

Model II measures the effects of economic data performance through source data capacity, data use, data services, data products, data sources, and data infrastructure on financial inclusion (measured by account ownership). The results of the regression model are illustrated in table 6.

Table 6: Regression Summary For Model II

OLS Regression Results

Dep. Variable:	ACC	R-squared:	0.612
Model:	OLS	Adj. R-squared:	0.529
Method:	Least Squares	F-statistic:	7.353
Date:	Fri, 19 Apr 2024	Prob (F-statistic):	8.64e-05*
Time:	17:55:23	Log-Likelihood:	-136.66
No. Observations:	35	AIC:	287.3
Df Residuals:	28	BIC:	298.2
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.8971	25.628	0.269	0.790	-45.600	59.394
SDC	-0.5816	0.214	-2.720	0.011*	-1.020	-0.144
DUS	-0.2143	0.243	-0.882	0.385	-0.712	0.284
DSS	-0.7636	0.228	-3.343	0.002*	-1.231	-0.296
DPS	0.7052	0.485	1.455	0.157	-0.288	1.698
DSC	1.1690	0.273	4.276	0.000*	0.609	1.729
DIS	0.8336	0.289	2.880	0.008*	0.241	1.427

Omnibus:	2.268	Durbin-Watson:	2.512
Prob(Omnibus):	0.322	Jarque-Bera (JB):	1.155
Skew:	-0.237	Prob(JB):	0.561
Kurtosis:	3.753	Cond. No.	1.72e+03

Source: Computed using Python 3.12 (2024). Note: * and ** indicate significance at 5% and 10%, respectively.

The overall p-value of the model is 0.0001 which is below the 0.05 significance level. This means that the model is statistically significant as the independent variables collectively are good explanatory variables of financial inclusion. The R-squared and the adjusted-R squared of the model are 61.2% and 52.9%, respectively. Based on the result of adj. R, it indicates that the changes in the independent variables explain 52.9% of the changes in the dependent variable. The remaining 47.1% of changes were explained by other factors which are not included in the model. Source data capacity, data use, and data services all have a negative coefficient whilst data products, data sources and data infrastructure have a positive coefficient. Meaning that a change in the value of the positive variables will lead to a direct change in financial inclusion. Negative variables on the other hand will cause an inverse change in the value of financial inclusion.

4.3. Pearson Correlation Test

The purpose of a correlation test is to assess the possibility of a linear relationship between the tested variables.

4.3.1. Model I Correlation

The correlation coefficients between the dependent and independent variables are highlighted in table 7.

Table 7: Correlation Matrix for Model I

	IA	MUP	DP	MPO	MBP	MMA
IA	1					
MUP	0.2754	1				
DP	0.5075	0.7370	1			
MPO	0.8565	0.3466	0.5740	1		
MBP	0.3893	0.9405	0.8093	0.4409	1	
MMA	0.0929	0.7992	0.8060	0.2790	0.7883	1

Source: Computed using Python 3.12 (2024)

From a glance, it is evident that there is a positive correlation between all variables in the model. The correlation coefficient between the Mobile Money Account (MMA) and the independent variables; Internet Access (IA) and Mobile Phone Ownership are 0.0929 and 0.2790, respectively. As IA and MPO increases, financial inclusion will also increase. However, this suggests a weak positive correlation since they are below the 0.5 range. Mobile Utility Payment, Digital Payments and Mobile Bill Payments all have a strong positive correlation on MMA with a correlation coefficient of over 0.7. This indicates a strong influence between the variables and financial inclusion.

4.3.2. Model II Correlation

The correlation coefficients between the dependent and independent variables are highlighted in table 8.

Table 8: Correlation Matrix for Model II

	SDC	DUS	DSS	DPS	DSC	DIS	ACC
SDC	1						
DUS	0.3689	1					
DSS	0.1998	0.2434	1				
DPS	0.4825	0.5854	0.5694	1			
DSC	0.6059	0.2682	0.5821	0.5460	1		
DIS	0.6540	0.5397	0.3783	0.4397	0.5924	1	
ACC	0.3154	0.2046	0.1318	0.3098	0.6144	0.5279	1

Source: Computed using Python 3.12 (2024)

It is also evident that there is a positive correlation between all variables in the model. However, only Data Sources (DSS) and Data Infrastructure (DIS) have a strong correlation with financial inclusion with a correlation coefficient of over 0.5. The other variables Source Data Capacity (SDC), Data Use (DUS), Data Services (DSS), and Data Products (DPS) have a weak correlation due to their correlation coefficient ranging from 0.1 – 0.3. Hence, an increase in the variables will slightly lead to an increase in financial inclusion (ACC)

4.4. Test for Multicollinearity

Testing for multicollinearity typically requires accessing the correlation between multiple independent variables with no considerations about the dependent variable. The correlation coefficients estimated in the previous section can give us a glimpse of which independent variables have multicollinearity against each other. But to simplify the process of evaluating each correlation coefficient, we will be utilizing the Variance Inflation Factor (VIF). There is no universally accepted VIP range but for the purpose of this paper, we will be adopting the range specified by Pennsylvania State University (n.d.). The basic rule of thumb is that a VIF < 5 indicates weak multicollinearity which is acceptable, VIF > 5 indicates multicollinearity issues which requires further investigation, and VIF > 10 indicates serious multicollinearity problem that requires immediate action.

4.4.1. Model I Multicollinearity Test

Table 9: Variance Inflation Factor (VIF) For Model I

Variable	VIF
IA	11.393105
MUP	20.275091
DP	20.113419
MPO	17.854881
MBP	31.157105

Source: Computed using Python 3.12 (2024)

All independent variables consisting of Internet Access, Mobile Utility Payment, Digital Payment, Mobile Phone Ownership, and Mobile Bill Payment for model I exceeds the VIF value of 10 which poses a problem to the study. It is an indication that all the independent variables are highly correlated which could potentially inflate the standard errors. Alternatively, it shows the level of unreliability in the variables when predicting the outcome of financial inclusion. The main goal of this paper is to make predictions between the variables and not understand the exact relationship between the dependent and independent variables. Remedies to resolve this problem will not be considered in this paper and will be mentioned in the limitations of the study and recommendations for future research in chapter 5.

4.4.2. Model II Multicollinearity Test

Table 10: Variance Inflation Factor (VIF) For Model II

Variable	VIF
SDC	2.453183
DUS	2.039501
DSS	2.109660
DPS	2.651508
DSC	2.616190
DIS	2.604678

Source: Computed using Python 3.12 (2024)

Source Data Capacity (SDC), Data Use (DUS), Data Services (DSS), Data Products (DPS), Data Sources (DSC), and Data Infrastructure (DIS) all fall below the value of 5. This means that all independent variables in the model have low multicollinearity which is acceptable and dependable for predicting effects on financial inclusion.

4.5. Descriptive Statistics

The study horizon for this paper is for a single year, which is 2021. For this reason, not all descriptive statistics metric will be analyzed and discussed in this paper as they are less useful for the study goal. The underlying statistics that will be discussed are mean, standard deviation, standard error, skewness, and kurtosis.

4.5.1. Model I Descriptive Statistics

There are 40 African countries studied in model I, including Nigeria. The variables studied consists of Mobile Money Account (MMA), Internet Access (IA), Mobile Utility Payment (MUP), Digital Payment (DP), Mobile Phone Ownership (MPO), and Mobile Bill Payment (MBP). The summary of the descriptive statistics is illustrated in table 11.

Table 11: Summary of Descriptive Statistics For Model I

	MMA	IA	MUP	DP	MPO	MBP
Count	40.00	40.00	40.00	40.00	40.00	40.00
Mean	28.77	37.36	9.64	41.87	72.71	13.41
Std	19.11	19.45	9.14	19.43	15.61	10.89
Min	0.00	5.22	0.00	4.81	32.16	1.45
25%	9.78	23.12	3.14	26.74	63.49	4.94
50%	31.87	31.06	6.23	42.99	75.67	9.86
75%	42.48	49.79	13.87	54.18	83.21	18.96
Max	68.66	82.91	36.06	80.81	100.00	44.91
Skewness	0.09	0.69	1.34	0.24	-0.68	1.40
Kurtosis	-1.06	-0.18	1.17	-0.65	0.11	1.68
Standard Error	3.02	3.07	1.45	3.07	2.47	1.72
Mode	29.38	5.22	13.61	33.74	82.76	1.45
S. Variance	365.00	378.13	83.54	377.72	243.54	118.52
Range	68.66	77.69	36.06	76.00	67.84	43.46
Sum	1150.90	1494.36	385.42	1674.61	2908.47	536.50

Source: Computed using Python 3.12 (2024). Note: **Std** stands for Standard and **S.** stands for Sample.

Starting with analysis of the mean, all 40 African countries in the model have a combined average of 28.77 percent for mobile money account. This means that only 28.77 percent of adults from the age of 15 have access to or own a mobile money account. While it may appear to be a poor performance, it can be influenced by countries with a high penetration rate compared to those with a lower rate. On average, approximately 37.36 percent of adults have internet access. During the period, only 9.62 percent of the study population made utility payments. 41.87 percent of adults made all kinds of payments using digital devices. 72.71 percent of correspondents have access to or own a mobile phone, and just 13.41 percent paid bills with their mobile phones.

For standard deviation values, all the variables appear to have moderate variability around the mean except for mobile utility payment with a standard deviation of 9.14. Compared to the mean of 9.64, this suggests that the variability is relatively low.

Looking at skewness, all variables except for mobile phone ownership (MPO) are positively skewed. Depicting that the mean is skewed to the right side of the distribution. MPO has a skewness of -0.68 which is moderately skewed to the left side of the distribution. Mobile money account has a near zero skewness of 0.09 showing that there is a relatively symmetrical distribution. MUP and MBP have skewness exceeding 1 which shows that they are highly skewed. In the real sense, a positive skewness indicates that while the variables are improving, there are outliers or certain segments of the country where usage of digital devices are more dominant. Negative skewness with respect to MPO shows a more balanced usage or ownership of mobile phones across all demographics.

Kurtosis values for all variables fall within the range of -2 to +2 indicating no sign of excess kurtosis. Therefore, the distribution is neither too peaked (heavy-tailed) nor too flat (light-tailed). In other words, the data used are moderately appropriate and do not display extreme tails that could impact the shape of the distribution.

MUP and MBP have relatively small standard errors (1.45 and 1.72 respectively) which signifies precision of estimates of the population mean. IA and DP on the other hand both have standard errors of 3.07 showing that the variability is slightly high. MMA and MPO both have moderate standard errors showing moderate levels of variability in the sample means.

4.5.2. Model II Descriptive Statistics

There are 35 African countries studied in model II including Nigeria. The variables studied consist of Account (ACC), Source Data Capacity (SDC), Data Use (DUS), Data Services (DSS), Data Products (DPS), Data Sources (DSC), and Data Infrastructure (DIS). Just a quick reminder, the independent variables are measured by scores ranging from 0 – 100 with 0 indicating lowest score and 100 indicating highest score. The summary of the descriptive statistics is illustrated in table 12.

Table 12: Summary of Descriptive Statistics For Model II

	ACC	SDC	DUS	DSS	DPS	DSC	DIS
Count	35.00	35.00	35.00	35.00	35.00	35.00	35.00
Mean	48.73	55.14	76.85	61.06	78.39	32.93	51.86
Std	19.55	16.87	13.53	14.64	7.74	13.62	12.84
Min	5.83	20.00	40.00	20.60	53.92	9.04	25.00
25%	34.56	45.00	70.00	60.30	76.54	25.59	42.50
50%	49.49	60.00	80.00	63.87	78.96	29.18	50.00
75%	61.69	65.00	90.00	67.20	82.45	38.62	60.00
Max	90.53	80.00	100.00	86.67	89.39	70.28	80.00
Skewness	-0.05	-0.38	-0.62	-1.17	-1.19	0.97	0.10
Kurtosis	-0.08	-0.43	0.42	2.00	1.96	1.24	-0.36
Std Error	3.31	2.85	2.29	2.47	1.31	2.30	2.17
Mode	5.83	60.00	80.00	57.03	53.92	9.04	40.00
S. Variance	382.34	284.54	183.05	214.37	59.85	185.54	164.83
Range	84.70	60.00	60.00	66.07	35.47	61.24	55.00
Sum	1705.60	1930.00	2689.80	2137.27	2743.63	1152.45	1815.00

Source: Computed using Python 3.12 (2024). Note: Std stands for Standard and S. stands for Sample.

On average, almost half (48.73 percent) of adults from the study population own or have access to any form of financial account. This indicates a partial penetration rate of financial inclusion. SDC, which measures a country's ability to source and maintain statistical data, has an average of 55.14 score for the study population. The average of DUS is high at 76.85 score showing a strong demand for statistical data among all countries. DSS has an average of 66.01 score showing a moderate data services score among all countries. With a DPS mean of 78.39, the study population has a strong score with regards to the quality of available data products. The quality and reliability of data sources used by all countries is low with a mean DSC of 32.93. Data infrastructure score is moderate at an average of 51.86 for all tested countries.

There is relatively high standard deviation among all the variables except for DPS which has a standard deviation of 7.74. This indicates that the other variables are less consistent around the mean whilst DPS has less variability in the data points.

All tested variables except DSC and DIS have a negative skewness indicating that the tail of the distribution is skewed to the left side of the curve. DSC and DIS are skewed to the right as they have positive skewness of 0.97 and 0.10, respectively. ACC has a near zero skewness of -0.05 showing that there is a relatively symmetrical distribution. None of the variables have a skewness exceeding +1 which shows that they are not highly skewed to the right. DSS and DPS have skewness less than -1 indicating that they are highly skewed to the left.

Kurtosis values for all variables fall within the range of -2 to +2 indicating no sign of excess kurtosis. Therefore, the distribution is neither too peaked (heavy-tailed) nor too flat (light-tailed). In other words, the data used are moderately appropriate and do not display extreme tails that could impact the shape of the distribution.

DPS has the lowest standard error at 1.31 showing that the variable has the highest precision in estimating the sample mean. SDC has a slightly lower variability with a standard error of 2.85. DUS, DSS and DSC have relatively lower variability with standard errors of 2.29, 2.47, and 2.30, respectively. ACC and DIS respectively have standard errors of 3.31 and 2.17 indicating moderate variability in the sample means.

4.6. Hypotheses Testing

To test for the specified hypothesis, 2 metrics have been adopted: hypothesis testing using p-values and hypothesis testing using correlation coefficients. The summary of the hypotheses testing using p-values and correlation coefficients will be presented in table 13 and 14, respectively. It should be noted that all numerical analysis conducted are based on a 5% and 10% significance level which serves as a basis for deciding whether to accept or reject the null hypothesis.

4.6.1. Model I Hypotheses Test

The results of the correlation test have already been discussed in section 4.3 so less will be discussed about each individual variable coefficient. As previously mentioned, all correlation coefficient obtained are positive signifying that there is a direct relationship between all the variables on financial inclusion. An increase in big data will potentially lead to an increase in financial inclusion. However, not all variables have a strong impact on financial inclusion. only internet access and mobile phone ownership have a weak relationship whilst mobile utility payment, digital payments, and mobile bill payments have a strong positive relationship.

Testing the hypothesis from a different dimension by using p-values, we get different results. Internet access, mobile utility payments, digital payments and mobile phone ownership are statistically significant with a p-value of less than 0.05 or 0.1. Hence, the alternative hypothesis is accepted. On the other hand, only mobile bill payment exceeds the critical value of 0.05 or 0.1 with a p-value of 0.879 showing statistical insignificance. Therefore, the alternative hypothesis emphasizing significant impact of the variables on financial inclusion will be rejected.

4.6.2. Model II Hypotheses Test

The correlation coefficients for all variables in Model II also are positive proving that an increase in source data capacity, data use, data services, data products, data sources, and data infrastructure will lead to an improved growth in financial inclusion. By looking at only the correlation coefficients as a means of testing the hypothesis, all the stipulated alternative hypotheses would be accepted.

However, to justify the acceptance rule we must look at the p-values for each variable as a backup on whether the hypothesis stays true or not. The p-values of data use and data products exceed the critical point of 0.05 or 0.10 showing that they are statistically insignificant. Therefore, the alternative hypothesis must be rejected. Looking at the other variables (that is, source data capacity, data services, data sources, and data infrastructure), their p-values fall below the critical point of 0.05. This means that the variables are statistically significant, and that the alternative hypothesis will be accepted.

Table 13: Summary of Hypothesis Test Using P-Values

Hypothesis	**Result**	**Justification**
Model I		
H_{a1}: A high degree of big data from all digital sources has an overall strong impact on financial inclusion.	Accept	Statistically Significant (Prob F = 0.000)
H_{a2}: Big data from internet usage could improve overall financial inclusion.	Accept	Statistically Significant (p-value = 0.000)
H_{a3}: Big data collected from mobile utility payment could improve financial inclusion.	Accept	Statistically Significant (p-value = 0.094)
H_{a4}: Big data from all digital payments could improve financial inclusion.	Accept	Statistically Significant (p-value = 0.000)
H_{a5}: A high degree of mobile phone usage could improve financial inclusion through sourced usage data.	Accept	Statistically Significant (p-value = 0.077)
H_{a6}: A high degree of mobile bill payment could improve financial inclusion through collected data.	*Reject*	Statistically *Insignificant* (p-value = 0.879)
Model II		
H_{a7}: An increase in the general statistical performance of a country can improve financial inclusion.	Accept	Statistically Significant (Prob F = 0.000)

H_{a8}: A high level of source data capacity could improve financial inclusion.	Accept	Statistically Significant (p-value = 0.011)
H_{a9}: Economic data use score has a positive influence on financial inclusion.	*Reject*	Statistically *Insignificant* (p-value = 0.385)
H_{a10}: Availability of quality and accessible economic data services have a positive influence on financial inclusion.	Accept	Statistically Significant (p-value = 0.002)
H_{a11}: A high degree of producible economic data score has a positive influence on financial inclusion.	*Reject*	Statistically *Insignificant* (p-value = 0.157)
H_{a12}: A high degree of provisional data sources could improve financial inclusion.	Accept	Statistically Significant (p-value = 0.000)
H_{a13}: The increase in data infrastructure could improve overall financial inclusion.	Accept	Statistically Significant (p-value = 0.008)

Source: Framed by Author. *Note:* **Prob F** stands for Probabilistic F-Statistics (Significance F).

Table 14: Summary of Hypothesis Test Using Correlation Coefficients

Hypothesis	**Result**	**Justification**
Model I		
H_{a1}: A high degree of big data from all digital sources has an overall strong impact on financial inclusion.	N/A	N/A
H_{a2}: Big data from internet usage could improve overall financial inclusion.	Accept	Positive Correlation (r = 0.092)

H_{a3}: Big data collected from mobile utility payment could improve financial inclusion.	Accept	Positive Correlation (r = 0.799)
H_{a4}: Big data from all digital payments could improve financial inclusion.	Accept	Positive Correlation (r = 0.806)
H_{a5}: A high degree of mobile phone usage could improve financial inclusion through sourced usage data.	Accept	Positive Correlation (r = 0.279)
H_{a6}: A high degree of mobile bill payment could improve financial inclusion through collected data.	Accept	Positive Correlation (r = 0.788)
Model II		
H_{a7}: An increase in the general statistical performance of a country can improve financial inclusion.	N/A	N/A
H_{a8}: A high level of source data capacity could improve financial inclusion.	Accept	Positive Correlation (r = 0.315)
H_{a9}: Economic data use score has a positive influence on financial inclusion.	Accept	Positive Correlation (r = 0.205)
H_{a10}: Availability of quality and accessible economic data services have a positive influence on financial inclusion.	Accept	Positive Correlation (r = 0.132)
H_{a11}: A high degree of producible economic data score has a positive influence on financial inclusion.	Accept	Positive Correlation (r = 0.310)
H_{a12}: A high degree of provisional data sources could improve financial inclusion.	Accept	Positive Correlation (r = 0.614)

H_{a13}: The increase in data infrastructure could improve overall financial inclusion.	Accept	Positive Correlation (r = 0.528)

Source: Framed by Author. *Note:* **r** stands for correlation coefficient.

5. Discussions, Recommendations and Conclusion

5.1. Introduction

The study goal was to analyze the relationship between big data and financial inclusion. Out of this goal are specific research objectives which were laid out for investigation. Two models were developed to look at data from different angles. First is big data from primary sources and then economic data performance, which we believe is important when making decisions towards an inclusive financial industry. To achieve the objectives, cross-sectional data analysis was conducted by employing statistical methods such as regression analysis, correlation analysis, multicollinearity test, and descriptive analysis. A summary of the results and findings will be presented in this chapter. Study limitations, recommendations for future research, and conclusion will also be finalized in this chapter.

5.2. Summary of Findings

The research objectives were successfully achieved, which entails analyzing the relationship and significance of big data on financial inclusion. The big data variables employed in Model I are internet access, mobile utility payment, digital payments, mobile phone ownership, and mobile bill payments. The proxy variables for measuring financial inclusion for model I in this study is mobile money account. The stated alternative hypothesis for Model I is that there is a positive relationship between all the variables. For model II, the independent variables include source data capacity, data use score, data services score, data products score, data sources score, and data infrastructure score. Financial inclusion for model II was measured by ownership of an account. The stated alternative hypothesis is that there is a positive relationship between all tested variables for model II. The major findings from the study are elaborated below:

I. Results from the OLS regression models show that mobile utility payment, digital payments, mobile phone ownership, data sources, and data infrastructure have a significant increasing impact on financial inclusion.

II. Internet access, source data capacity, and data services demonstrated a significant decreasing impact on financial inclusion based on the regression models.

III. The results from the Pearsion correlation test demonstrated a positive relationship between all tested variables on financial inclusion. However, only mobile utility payment, digital payment, mobile bill payment, data sources, and data infrastructure have a dominant or strong relationship with financial inclusion.

IV. All stated hypotheses in the study will be accepted if we focus on the correlation estimates. However, by using significance level, hypotheses 1-5, 7-8, 10, 12, and 13 will be accepted. Therefore, only hypotheses 6, 9, and 11 are rejected in this study. Overall, the distribution of accepted and rejected hypotheses are 10 and 3, respectively.

5.3. Study Limitations

The results of this study depend on secondary information analyses. The study results are subjected to the limitations of the datasets available to the public. The initial goal of the study was to examine the Nigerian market, but after a series of attempts to gather relevant data, it had to be expanded to include other developing countries, more specifically in Africa due to relatedness and proximity with economic structure. The study therefore focuses on all developing economies whereby statistical inference will be made about the Nigerian market. As a result, the estimates of the findings may not give a full picture of the nature of financial inclusion in Nigeria and its influencing factors. Only data for a single year (2021) were pooled for this study leading to incapacity to analyze time series data which could have given more insight in the analysis for all studied variables. Furthermore, there are other relevant variables which could have been tested and could potentially have strong explanatory power on financial inclusion that were not included in this study.

5.4. Recommendations for Future Research

Given the aforementioned limitations of this study, the suggestions for future studies are stipulated as follow:

I. The study aimed at assessing the nature of financial inclusion in Nigeria. However, the study was hindered by the availability of limited data required to capture the full picture of the economy. Therefore, future studies should try alternative means of getting the required data. This will require obtaining a portion of the data firsthand or adopting primary research methods.

II. The time horizon of this study encapsulates only the year 2021. Future research should consider adopting cross-sectional time series data analysis which will better give a depiction of trend in the selected variables across time.

III. The VIF estimates show high signs of multicollinearity among the independent variables for model I. This suggests that there is a high correlation among the variables. For future research interest in a related field, researchers should adopt certain modification techniques such as removing independent variables with high VIF values entirely. However, doing so might result in not having any variables to study as this study produces. Another consideration on fixing this issue would be to transform the variables into different format to help further reduce multicollinearity. Regularization methods can also be adopted.

IV. Further studies could also benefit if ML and AI techniques are adopted into the research to precisely identify areas where big data could be of benefit for financial institutions to adopt and implement in secluded markets.

5.5. Recommendations

Based on the empirical results of this study, the findings produce the following recommendations which should be considered across all key stakeholders in the financial industry:

I. Financial service providers should incorporate data analytics into their business model as it has been shown to have significant impact on financial inclusion. Data from sources such as mobile utility payment, digital payments, mobile phone ownership and mobile bill payments can be tracked and analyzed to get better insights into customer behavior. Analyzed data could produce alternative credit scoring metrics which can be used to access the credit worthiness of customers in the unbanked population.

II. The government should introduce initiatives to ensure that relevant data across all economic facets are consistently collected, stored, and appropriately made accessible if not the public, but to the financial institutions as it has shown that such data have statistically significant impact on financial inclusion. This data is essential as it accompanies big data towards getting an immersive overview of the economic situation and helps make better data driven decisions which over the long term should foster growth in financial inclusion.

5.6. Conclusion

Concluding this chapter requires a summary of all that has been accomplished from the previous chapters. The study focused on investigating the impact of big data on financial inclusion. We began by reviewing relevant literature surrounding big data, data analytics, financial inclusion and its current state in Nigeria, innovations trends in financial inclusion, and finally the key theories underpinning the topic. Reviewing previous materials like the topic gives us a valuable understanding of what is expected from the paper and the expected deliverables. Review of pervious materials show that digitalized solutions like big data analytics through the application of AI and ML has a significant impact on increasing access to finance. However, there were not enough statistical analysis conducted by previous researchers to analyze in numeric terms the level of significance big data has on financial inclusion which is one of the foundations underpinning this paper.

To comply with the objectives of the research, the study used an appropriate OLS regression model to estimate the coefficient of the variables. The quantitative data were obtained from the World Bank Databank, more specifically the Global Findex and SPI databases. The gathered data was forwarded for statistical analysis using Python programing language version 3.12. The analysis constituted conducting regression analysis, Pearson correlation test, multicollinearity test, descriptive statistics, and hypothesis testing. Two models were developed for analysis: Model I for big data analysis and Model II for economic data quality and accessibility analysis. For assessing the specified hypothesis, a sample of 40 and 35 developing countries in Africa were collected for Model I and II, respectively. Out of the 13 hypotheses formulated, 10 were approved and only 3 were rejected. The empirical findings show a significant relationship between big data and economic data quality on improving access to finance. The 10 hypotheses accepted explain that a high degree of financial inclusion is explained by big data and data quality. In other words, the results shows that financial access in Nigeria can be improved if high considerations were made towards analyzing big data obtained from various sources such as internet usage, mobile utility payments, digital payments, and phone usage data. Analyzing this data should help determine the areas of finance that need improvement and create targeted financial products and services for people with no or little access to finance. The result also shows that other data variables like data collection, data service centers, platforms to access key econometric data, and improvement in data infrastructure can significantly contribute to improved financial inclusion. We can therefore conclude that this paper has successfully accomplished exploration of the research questions and serves as a foundation for further studies to be conducted.

This study has contributed to a considerable amount of knowledge as it is among the first to have examined the impact of internet access, mobile utility payments, digital payments, mobile phone ownership and mobile bill payments on financial inclusion. Also more importantly, the study added to knowledge by analyzing the impact of SPI indicators which are seen to be relevant towards a better data driven decision making, especially with regards to fostering growth in financial inclusion.

Word Count: 21,543

(excluding cover page, acknowledgement, abstract, list of abbreviations, table of contents, table of figures, list of tables, references, appendix, and affidavit)

References

Adeoye, O. B., Addy, W. A., Ajayi-Nifise, A. O., Odeyemi, O., Okoye, C. C., & Ofodile, O. C. (2024, March 9). *Leveraging Ai and Data Analytics for Enhancing Financial Inclusion in Developing Economies*. Finance & Accounting Research Journal. https://doi.org/10.51594/farj.v6i3.856

African Market Entry Consulting Ltd (Citizens & Expats Agency). (2023). *Forming Strategic Alliances as a Market Entry Option into Africa*. Link

Agarwal, S., Alok, S., Ghosh, P., & Gupta, S. (2019). *Fintech and Credit Scoring for the Millennials: Evidence using Mobile and Social Footprints*. Social Science Research Network. https://doi.org/10.2139/ssrn.3507827

Aiforgoodstg. (2023). *Use of big data and AI for inclusive fintech*. AI For Good. Link

Akintola, D. D., Adedoyin, A. A., & Onashoga, B. A. (2020). Factors influencing the adoption of mobile money services in Nigeria. *Journal of Economic Cooperation and Development*, 45(3), 101-122. (https://www.sesric.org/publications-jecd-volumes.php)

Alade, S. O. (2017). The Nigerian Financial System at a Glance. Central Bank of Nigeria, Monetary Policy Department.

Allen, F., Demirguc-Kunt, A., Klapper, L., & Martinez Peria, M. S. (2016). The foundations of financial inclusion: Understanding ownership and use of formal accounts. *Journal of Financial Intermediation, 27*, 1-30. https://doi.org/10.1016/j.jfi.2015.12.003

Antwi, B. O. (2015). *The failure of microfinance in Ghana: A case study of Noble Dream Microfinance Limited*. Kwame Nkrumah University of Science and Technology.

Ashraf, N., Chandrasekhar, A., & Klapper, L. (2016). Financial inclusion and development. The World Bank Research Observer, 31(2), 162-188. https://doi.org/10.1093/wbro/lkw001

Beshears, D., Choi, J. J., DellaVigna, S., & Laibson, D. (2015). Making credit cheaper for low-income borrowers. *The American Economic Review*, 105(4), 1284-1310.

Bhandari, A. (2024). *Multicollinearity | Causes, effects and detection using VIF*. Analytics Vidhya. Link

Bharadwaj, A. (2019). Big data and credit scoring in developing economies. *Journal of Financial Regulation and Compliance*.

Błach, J. (2020). Barriers to Financial Innovation—Corporate Finance Perspective. *Journal of Risk and Financial Management*, 13(11), 273. https://doi.org/10.3390/jrfm13110273

Bruhn M., & Love I. (2014). The real impact of improved access to finance: Evidence from Mexico. *The Journal of Finance*, *69*(3), 1347–1376. https://doi.org/10.1111/jofi.12091

Central Bank of Nigeria (CBN). (2020). Revised regulatory framework for agent banking in Nigeria. Link

Central Bank of Nigeria (CBN). (2021). eNaira: A digital currency for Nigeria. https://www.cbn.gov.ng/currency/enaira.asp

Central Bank of Nigeria. (n.d.). Financial Inclusion. Retrieved March 23, 2024, from the Central Bank of Nigeria website: [Central Bank of Nigeria :: Development Finance (cbn.gov.ng)]

Central Bank of Nigeria. Cardoso Calls for Stakeholders' Commitment to 95% Financial Inclusion Target. (n.d.). Link

Centre for Global Development. (2019). Rethinking Aid: When does foreign assistance make countries more dependent?

Chen, H., Guo, Z., & Zhou, L. (2019). Big data analytics for credit risk assessment: A review of machine learning technologies and applications. Journal of Economic Surveys, 33(4), 1085-1134.

Columbus, L. (2014). 84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes In The Next Year. *Forbes*. Link

Daasi, G. (2024). Traditional Financial Institutions and Rural Enterprises in Nigeria: The Case of Ogoni Land.

Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). The Global Findex database 2017: Measuring financial inclusion and the fintech revolution. The World Bank.

Ekekwe, N. (2024). *A Nigerian Model for Stakeholder Capitalism*. Harvard Business Review. https://hbr.org/2021/05/a-nigerian-model-for-stakeholder-capitalism

European Commission. (2024). Financial literacy and the EU: Empowering citizens with financial knowledge and skills. Financial literacy - European Commission (europa.eu)

Fakoyejo, O. (2023). *Report: Nigeria's financial inclusion rose to 74% in 2023, northern region*. TheCable. Link

Falaiye, T., Elufioye, O. A., Awonuga, K. F., Ibeh, C. V., Olatoye, F. O., & Mhlongo, N. Z. (2024). Financial Inclusion Through Technology: A Review of Trends in Emerging Markets. *International Journal of Management & Entrepreneurship Research*, *6*(2), 368–379. https://doi.org/10.51594/ijmer.v6i2.776

Fernando, M., Shaji, G. A., & Krishnamoorthy, K. (2021). Applications & implications of big data analytics and AI in finance. *IJARCCE, 10*, 67-71. https://doi.org/10.17148/IJARCCE.2021.101111

Gandomi, A. H., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics*. International Journal of Information Management. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gideon, T., Hsu, F., & Mutua, S. (2019). Digital identity: A literature review. World Bank.

Gupta, S. (2022). Role of cloud computing in driving financial inclusion of emerging economies. Procedia Computer Science, 203, 669-673.

Hasan, M., Le, T., & Hoque, A. (2021). How does financial literacy impact on inclusive finance? *Financial Innovation*, *7*(1). https://doi.org/10.1186/s40854-021-00259-9

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Hickel, J. (2014). Exposing the great 'poverty reduction' lie. Al Jazeera. Link

Hollanders, M., Former Special Adviser on Financial Infrastructure, Bank for International Settlements, Switzerland. (2020). FinTech and financial inclusion: Opportunities and challenges. In *Journal of Payments Strategy & Systems* (No. 4; Vol. 14, p. 315). Link

How, M. L., Cheah, S. M., Khor, A. C., & Chan, Y. J. (2020). *Artificial Intelligence-Enhanced Predictive Insights for Advancing Financial Inclusion: A Human-Centric AI-Thinking Approach*. Big Data and Cognitive Computing. https://doi.org/10.3390/bdcc4020008

Hox, J. J., & Boeije, H. R. (2005). Data Collection, Primary vs. Secondary. *Encyclopedia of Social Measurement*, 593–599. https://doi.org/10.1016/b0-12-369398-5/00041-4

Husamaldin, L., & Saeed, N. (2019). *Big Data Analytics Correlation Taxonomy*. Information. https://doi.org/10.3390/info11010017

IBM. (2020). The Four V's of Big Data. *IBM Big Data & Analytics Hub*. Retrieved from https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Intel Region. (2023). Top 11 highest paying tech jobs in Nigeria | Intel Region. Intel Region. Link

Iwedi, M. (2023). Digital Banking Technology and Financial Inclusion in Nigeria. *DS Journal of Digital Science and Technology, 2*(3), 9-16. https://doi.org/10.59232/DST-V2I3P102

Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. American Economic Review, 104(1), 183-223. https://doi.org/10.1257/aer.104.1.183

Khan, A. (2016), The negative impact of foreign aid on recipient countries. * Journal of International Development.

Khera, P., Ng, S. Y., Ogawa, S., & Sahay, R. (2021). Measuring Digital Financial Inclusion in Emerging and Developing Economies: A New Index. IMF Working Papers, 2021(090), A001. Retrieved Apr 12, 2024, from https://doi.org/10.5089/9781513574196.001.A001

Klapper, L., & Singer, D. (2017). Strengthening financial inclusion through agent banking. Focus Note - Consultative Group to Assist the Poor, 101.

Kling, G., Pesqué-Cela, V., Tian, L., & Luo, D. (2020). A theory of financial inclusion and income inequality. *The European Journal of Finance*, 28, 137-157. https://doi.org/10.1080/1351847X.2020.1792960

Kumar, K., & McKay, J. (2020). Big data and financial inclusion. Journal of Technology and Development, 13(2), 114-133.

Kunze, F., Boehm, S. A., & Bruch, H. (2013). Age, resistance to change, and job performance. Journal of Managerial Psychology, 28(7/8), 741–760. https://doi.org/10.1108/jmp-06-2013-0194

Lohr, S. (2013). The Origins of 'Big Data': An Etymological Detective Story. *The New York Times*. Link

Marikyan, D. & Papagiannidis, S. (2023). *Technology Acceptance Model: A review*. In S. Papagiannidis (Ed), TheoryHub Book. Available at https://open.ncl.ac.uk / ISBN: 9781739604400

Maurer, B. (2012). Mobile money: Communication, consumption and change in the payments space. Journal of Development Studies, 48(5), 589-604.

McArthur, J., & Rasmussen, K. (2017). How successful were the millennium development goals?[1] *The Guardian*. Link

Mhlanga, D. (2020). *Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion*. International Journal of Financial Studies. https://doi.org/10.3390/ijfs8030045

Mhlanga, D. (2024). *Big Data and Financial Technology (Fintech) Towards Financial Inclusion*. Social Science Research Network. https://doi.org/10.2139/ssrn.4697365

Mungai, K., & Bayat, A. (2018). *The Impact of Big Data on the South African Banking Industry*. ResearchGate. Link

Nan, W. V. (2018). *Mobile Money and Socioeconomic Development: A Crosscountry Investigation in Sub-Saharan Africa.* Document - Gale Academic OneFile. Link

Neumann, C. M., & Berndt, T. (2020). Data Analytics in der Financial Due Diligence – Status quo und künftige Entwicklungen. *ResearchGate*. Link

OECD (2019). OECD Business and Finance Outlook 2019: Strengthening Trust in Business, OECD Publishing, Paris, https://doi.org/10.1787/af784794-en.

OECD. (2021). *Artificial intelligence, machine learning and big data in finance: Opportunities, challenges, and implications for policy makers*. Link

Ojebode, A., Ojebuyi, B. R., Oladapo, O. A., & Oyedele, O. J. (2018). Mono-Method Research Approach and Scholar–Policy Disengagement in Nigerian Communication Research. *The Palgrave Handbook of Media and Communication Research in Africa*, 369–383. https://doi.org/10.1007/978-3-319-70443-2_20

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). *The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics*. Applied Soft Computing. https://doi.org/10.1016/j.asoc.2018.10.004

Ozili, P. (2023). Big Data and Artificial Intelligence for Financial Inclusion. https://doi.org/10.1201/9781003125204-1

Ozili, P. K. (2018). Impact of digital finance on financial inclusion and stability. *Borsa Istanbul Review*, 18(4), 329-340. https://doi.org/10.1016/j.bir.2017.12.003

Ozili, P. K. (2020). Theories of financial inclusion. In *Uncertainty and Challenges in Contemporary Economic Behaviour* (pp. 1-2). DOI: 10.1108/978-1-80043-095-220201008

Pennsylvania State University. (n.d.). *10.7 - Detecting multicollinearity using variance inflation Factors* | *STAT 462*. https://online.stat.psu.edu/stat462/node/180/

Perez, M. (2023). *The 5 Vs in Big data: what they are and how to apply them*. SMOWL Proctoring | Supervision System for Online Exams. https://smowl.net/en/blog/big-data-5v/

Peronto, R. (2024). What percentage of data is unstructured? 3 must-know statistics. Edge Delta. Link

Pillai, R., Sivathanu, B., & Sreejith, S. (2020). Adoption of mobile banking services at BOP in India–an application of change management theories. Journal of Financial Services Marketing, 25(1), 59-73.

Qureshi, S. (2020). *Why Data Matters for Development? Exploring Data Justice, Micro-Entrepreneurship, Mobile Money, and Financial Inclusion*. Information Technology for Development. https://doi.org/10.1080/02681102.2020.1736820

Richard, J., & Shabir, G. (2023). *Revolutionizing Financial Services: The Power of Big Data and Analytics*. ResearchGate. https://doi.org/10.13140/RG.2.2.22090.95685

Ritchie, H., & Roser, M. (2018). Now it is possible to take stock - did the world achieve the Millennium Development Goals? *Our World in Data*. Link

Rouse, M., & Verhoef, G. (2017). Mobile banking in Sub-Saharan Africa: Setting the way towards financial development. Retrieved from https://mpra.ub.uni-muenchen.de/78006/

Ruan, K. (2019). *Case Study: Insuring the Future of Everything*. Elsevier eBooks. https://doi.org/10.1016/b978-0-12-812158-0.00010-7

Smith, J. (2020). The impact of initial endowments on economic mobility. Journal of Economic Inequality, 28(2), 145-165.

Sodipo, J. A., Ikue, N. J., Enegesi, L. B., Oraemesi, C. N., Yisa, D. N., & Ejinkonye, J. C. (2023). The Effect of Financial Inclusion on Poverty in Nigeria. *Diamond Bridge Economics and Business Journal*, 3(3). https://doi.org/10.60089/dbebj.2023.3.3.12

Statista. (2023). Mobile phone penetration in Nigeria from 2014 to 2027. Link

Statista. (2023). *Nigeria: population 1950-2023*. Link

Suri, T., & Jack, W. (2016). The long-run poverty and gender impacts of mobile money. Science, 354(6317), 1288-1292.

The World Bank (2024). Global Financial Inclusion. Link

The World Bank (2024). Statistical Performance Indicators (SPI). Link

Uduu, O. (2022). *Nigeria's Wealth Inequality Score is 35.1 and its 11th in West Africa*. Dataphyte. Link

United Nations Department of Economic and Social Affairs. (2015). *The state of the global partnership for development*. United Nations Millennium Development Goals

United Nations. (2023). Sustainable development goals. United Nations. https://www.un.org/sustainabledevelopment/

Wang, Y., Kung, L., Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13. https://doi.org/10.1016/j.techfore.2015.12.019

World Bank Group. (2022). National Financial Inclusion Strategies Resource Center. Retrieved from Financial Inclusion Strategies Resource Center (worldbank.org)

World Bank Open Data. (2022). World Bank Open Data. https://data.worldbank.org/indicator/SP.POP.1564.TO.ZS?locations=NG

World Bank. (2023). Mobile cellular subscriptions (per 100 people) - Sub-Saharan Africa. https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=ZG

Zakir, J., Seymour, T., & Berg, K. (2015). Big data analytics. *Issues in Information Systems, 16*(II), 81-90. https://doi.org/10.48009/2_iis_2015_81-90

Appendix I: Accessing Python Code for Statistical Analysis on GitHub

GitHub Repository

Information on how to access the codes used for the statistical analysis of this research paper with python is explained in this appendix.

- **Repository Name:** Master-Thesis-Statistical-Analysis
- **Link:** https://github.com/CypranA/Master-Thesis-Statistical-Analysis

Code Access Description

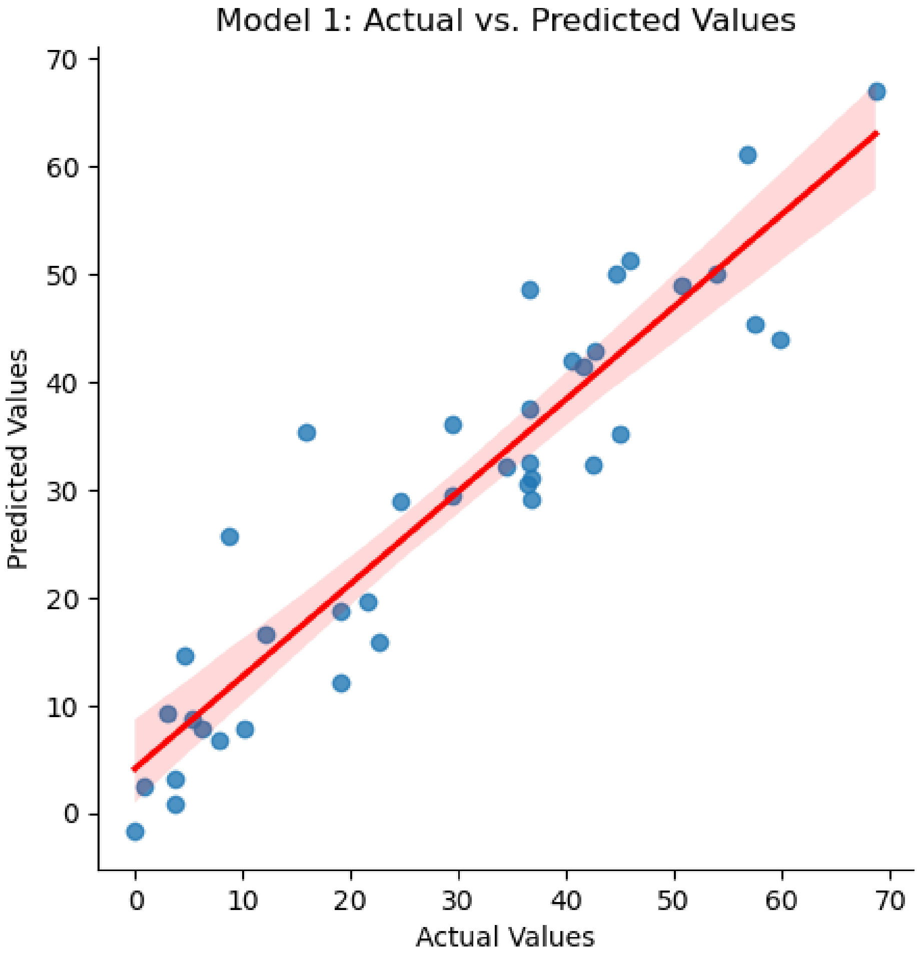
1. **Repository:** Click on the provided link or navigate to the repository directly on GitHub.
2. **Files in the Repository:** The repository contains the following files/folders:
    - Jupyter Notebook Python file (.ipynb) which contains the runnable code cells.
    - README.md file which has the same content as the .ipynb file but has been converted into a .md file so the repository can directly display the codes within the main page.
    - CBS Logo
    - Model I and II .csv files containing the datasets for each model.
    - Images of data visualization created with Python.
3. **Requirements:** To run the code, you will need:
    - Programming tool: Anaconda/Jupyter Notebooks
    - Programming language version: Python 3.11+
    - Required libraries: The required libraries used in the study have already been imported into the python file. All that needs to be done is to run the cell containing the specified codes; Pandas, NumPy, Matplotlib, Seaborn, Statsmodels, and SciPy.

Additional Notes

- **Code Updates:** There might be minor changes or updates with the code to reflect and visualize other metrics not discussed in the paper.
- **Data Availability:** The data is publicly accessible.
- **Comments/Questions:** Please feel free to contact me via my email address if you require further assistance.

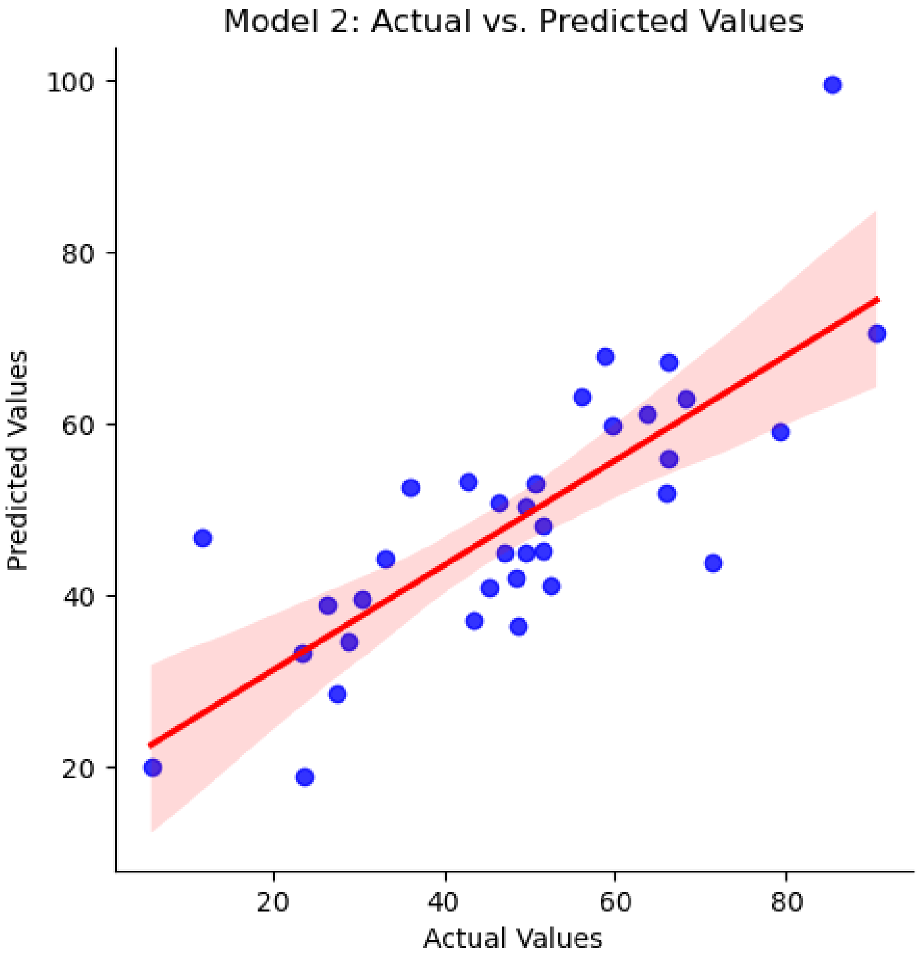
Appendix II: Other Materials

Figure 1: Visualization of Predictive Modelling for Model I



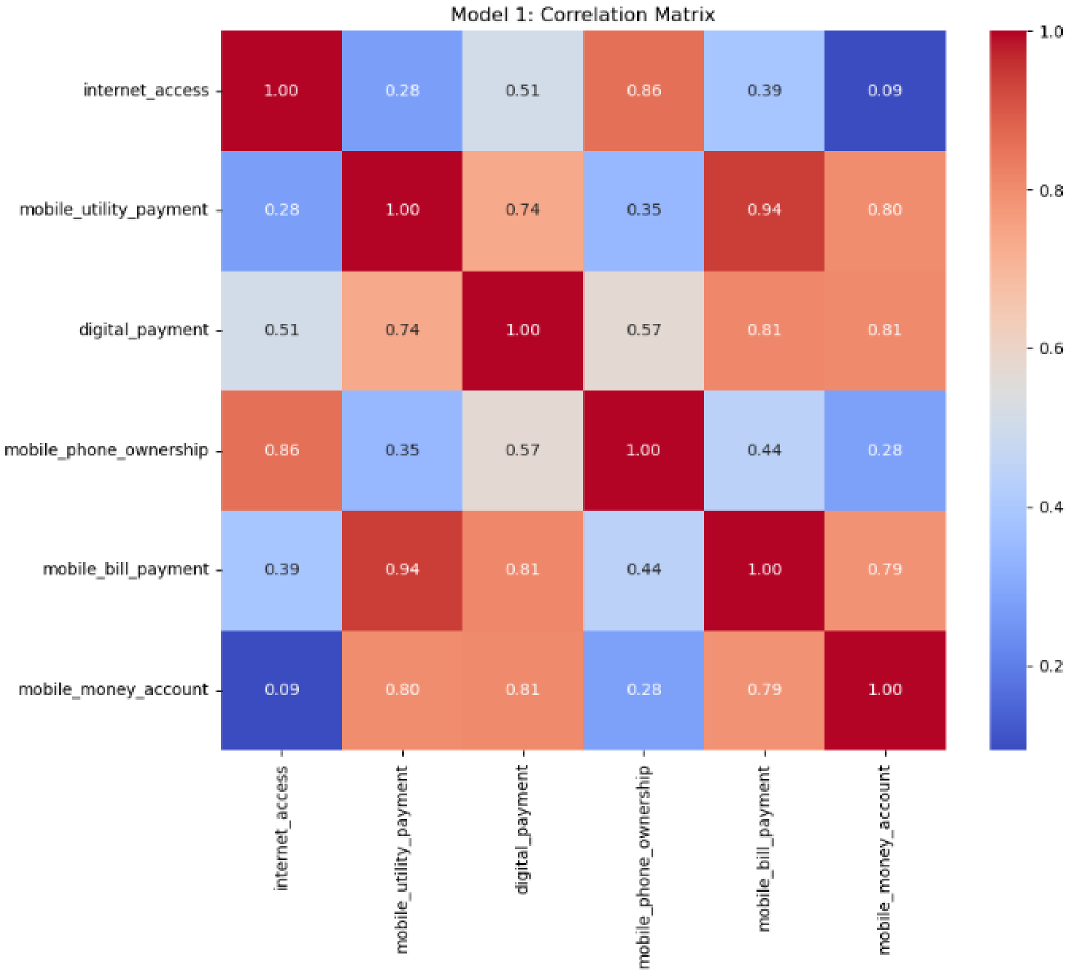
Source: Computed using Python 3.12. Note: This plot was created after creating predictive values for the models using predictive modelling techniques to depict the relationship between financial inclusion and all independent variables.

Figure 2: Visualization of Predictive Modelling for Model II



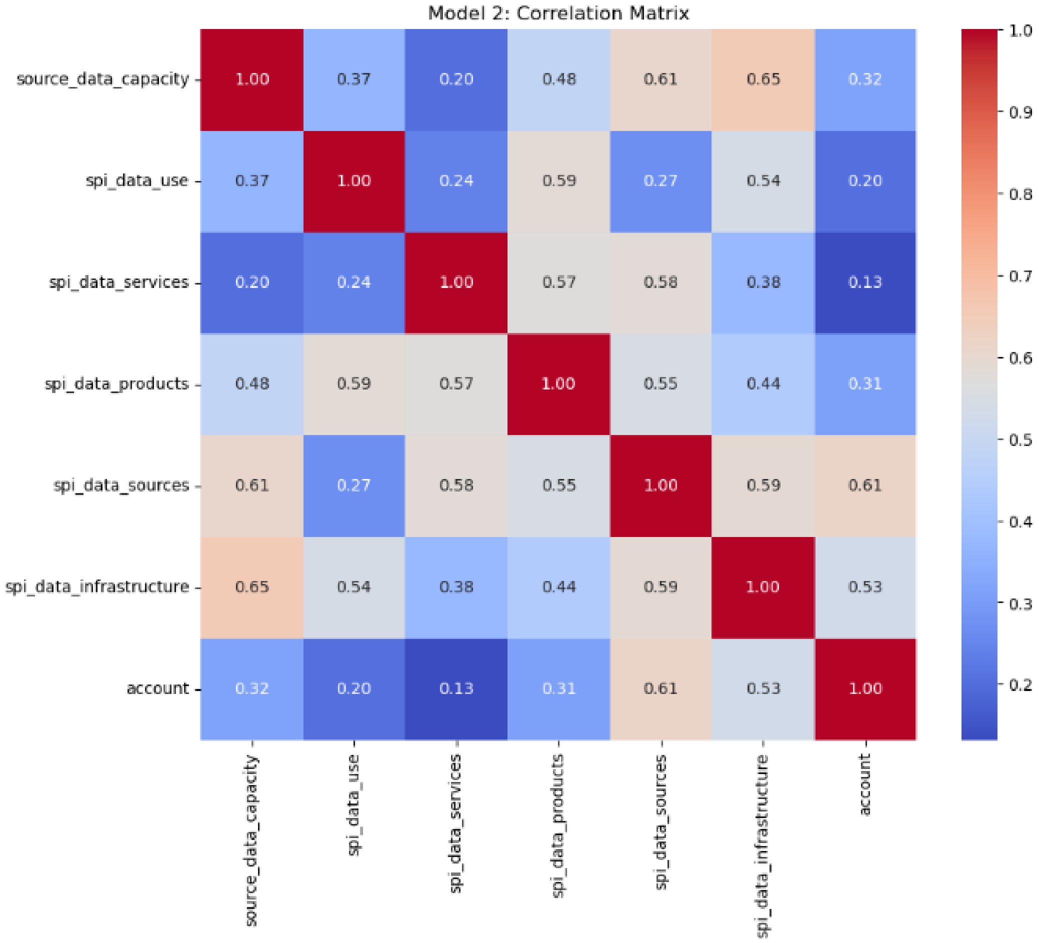
Source: Computed using Python 3.12. Note: This plot was created after creating predictive values for the models using predictive modelling techniques to depict the relationship between financial inclusion and all independent variables.

Figure 3: Correlation Matrix For Model I



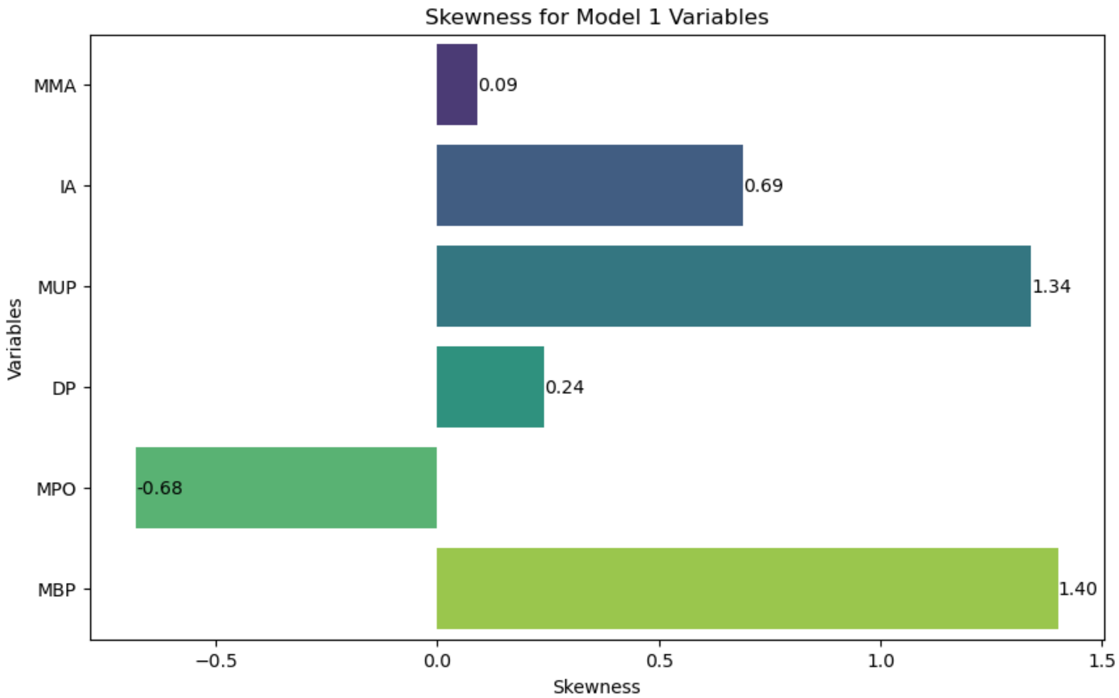
Source: Computed using Python 3.12.

Figure 4: Correlation Matrix For Model II



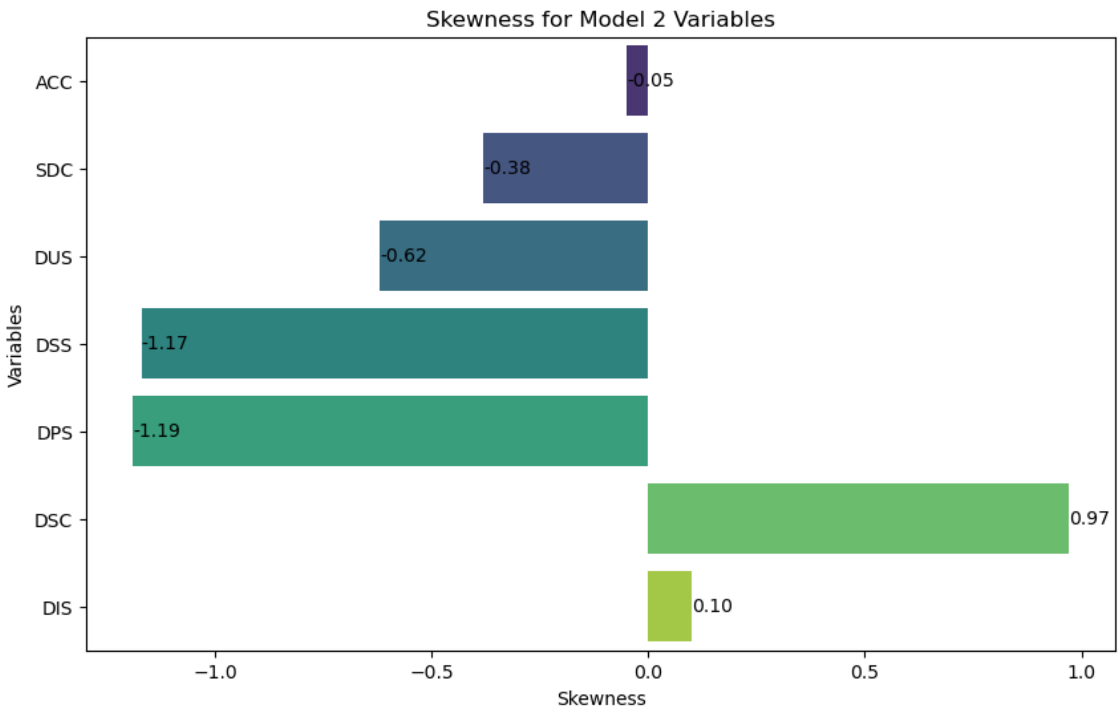
Source: Computed using Python 3.12

Figure 5: Skewness Plot For Model I



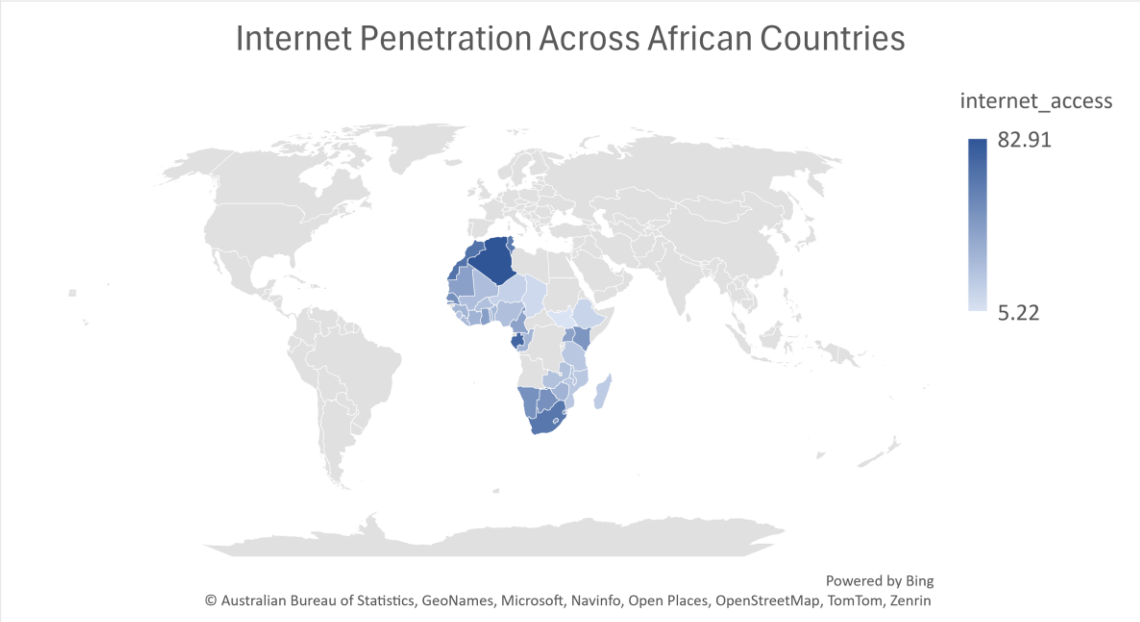
Source: Computed using Python 3.12

Figure 6: Skewness Plot For Model II



Source: Computed using Python 3.12

Figure 7: Internet Penetration Across All Studied African Countries



Source: Computed using Excel

Affidavit

Affirmation

I hereby affirm that I have written this academic paper independently and without the help of others. In doing so, I have used only the sources and means mentioned. I have not used works or parts of texts that are protected by copyright without changing or having reworked them. Passages in the paper where the wording or meaning is taken from other works or sources – including those taken from the internet – have been identified as quotations, citing the exact source.

Granting of Rights of Use

To check the paper for violations of copyright and plagiarism, CBS uses, in addition to the manual examination, a so-called web-based anti-plagiarism software. To carry out this examination of the paper, I grant CBS and their external service providers the right to reproduce the paper electronically, to save and temporarily store it, as well as to consult it for an indefinite period for comparative purposes when checking other papers. I therefore give my consent for my paper to be stored and used for a plagiarism assessment and to be passed on to a provider of a web-based plagiarism software here or abroad, which may only process and use the paper for this purpose.

Data Privacy Declaration

An examination paper can also contain statements about the personal and factual circumstances of the student or other persons concerned. The collection, storage and use of such data is only possible with the consent of the person concerned. In this context, I assure that all persons concerned (e.g., interview partners, if given) have consented to a publication and that I have clarified whether anonymization is desired, and that all personal data have been anonymized for those persons who have consented to publication only in anonymous form.

Cologne, *01.05.2024*

Kosisochukwu Cypran, Akubude, MA IB19