**PROJECT TITLE: ENERGY CONSUMPTION ANALYSIS IN KANO AND KATSINA STATES.**

TABLE OF CONTENTS

[**CHAPTER ONE** 1](#_Toc201344659)

[**INTRODUCTION** 1](#_Toc201344660)

[**1.0** **Introduction** 1](#_Toc201344661)

[**1.1 Background of the Study** 1](#_Toc201344662)

[**1.2 Statement of the Problem** 2](#_Toc201344663)

[**1.3 Aim and Objective of the Problem** 3](#_Toc201344664)

[**1.4 Research Question** 4](#_Toc201344665)

[**1.5 Significant of the Study** 4](#_Toc201344666)

[**1.6 Scope of the Study** 4](#_Toc201344667)

[**1.7 Operational Definition of Key Terms** 5](#_Toc201344668)

[**1.8 Chapter summary** 6](#_Toc201344669)

[**CHAPTER TWO** 7](#_Toc201344670)

[**LITERATURE REVIEW** 7](#_Toc201344671)

[**2.0 Introduction** 7](#_Toc201344672)

[**2.1 Reviewed Related Work** 7](#_Toc201344673)

[**2.2 Conceptual Framework** 9](#_Toc201344674)

[**2.2.1 Analyzing Electricity Consumption Trends and Patterns** 9](#_Toc201344675)

[**2.2.2 Identifying High-Consumption and High-Revenue Sectors** 10](#_Toc201344676)

[**2.2.3 Examining the Relationship among Pricing, Consumption and Revenue Generation** 11](#_Toc201344677)

[**2.2.4 Identifying Infrastructural Inefficiencies and Energy Losses** 11](#_Toc201344678)

[**2.3 Chapter summary** 12](#_Toc201344679)

[**CHAPTER THREE** 13](#_Toc201344680)

[**METHODOLOGY** 13](#_Toc201344681)

[**3.0 Introduction** 13](#_Toc201344682)

[**3.1 Research Design** 13](#_Toc201344683)

[**3.2 This study design is implemented using the following techniques:** 13](#_Toc201344684)

[**3.4 Data Analysis Methodology: Tools and Techniques** 16](#_Toc201344685)

[**3.4.1 Google Colaboratory (Colab)** 16](#_Toc201344686)

[**3.4.2 Jupyter Notebooks** 17](#_Toc201344687)

[**3.4.3 Python** 17](#_Toc201344688)

[**3.4.4 Power BI** 18](#_Toc201344689)

[**3.4.5 Comprehensive Discussion for Methodology Requirements** 18](#_Toc201344690)

[**3.5 Chapter summary** 19](#_Toc201344691)

[**CHAPTER FOUR** 21](#_Toc201344692)

[**RESULTS AND DISCUSSION** 21](#_Toc201344693)

[**4.0 Introduction** 21](#_Toc201344694)

[**4.1 Data Presentation and Analysis** 21](#_Toc201344695)

[**4.1.1 Data Loading and Preprocessing** 21](#_Toc201344696)

[**4.1.2 Electricity Consumption Trends** 22](#_Toc201344697)

[**4.1.3 High-Consumption and High-Revenue Sectors** 23](#_Toc201344698)

[**4.1.4 Relationship Between Electricity Price, Consumption, and Revenue** 24](#_Toc201344699)

[**4.1.5 Identification of Infrastructural Inefficiencies** 25](#_Toc201344700)

[**4.2 Summary of the Findings** 26](#_Toc201344701)

[**4.3 Chapter summary** 27](#_Toc201344702)

[**CHAPTER FIVE** 28](#_Toc201344703)

[**SUMMARY, CONCLUSION AND RECOMMENDATION** 28](#_Toc201344704)

[**5.0 Introduction** 28](#_Toc201344705)

[**5.1 Summary** 28](#_Toc201344706)

[**5.2 Conclusion** 29](#_Toc201344707)

[**5.3 Recommendation** 29](#_Toc201344708)

[**REFERENCES** 30](#_Toc201344709)

[**Appendix A** 32](#_Toc201344710)

[**A.1 Feeder Performance with Average Price** 33](#_Toc201344711)

[**A.2 Transformer Performance with Average Price** 33](#_Toc201344712)

[**A.3 Top Feeder Payments in 2021** 34](#_Toc201344713)

[**A.4 Top Feeder Payments in 2022** 34](#_Toc201344714)

[**A.3 Power BI Dashboard Year-2022** 35](#_Toc201344715)

[**A.6 Power BI Dashboard – Year 2021** 35](#_Toc201344716)

[**Appendix B** 36](#_Toc201344717)

[**Appendix C** 36](#_Toc201344718)

[**C.1 Software Environment** 36](#_Toc201344719)

[**C.2 Python Libraries and Packages** 36](#_Toc201344720)

## **CHAPTER ONE**

## **INTRODUCTION**

## **Introduction**

This study uses KEDCO’s dataset to analyze energy consumption trends in Kano and Katsina, aiming to improve grid efficiency and service delivery. The background of the study highlights energy’s role in economic growth, noting poor infrastructure and irregular supply issues. The statement of the problem addresses distribution inefficiencies due to rapid urbanization. The aims and objectives include analyzing consumption patterns, high-revenue sectors, and infrastructure performance, with the significance of the study emphasizing data-driven analytics. The Scope covers residential, commercial, and industrial sectors, with the definition of key terms and chapter summary focusing on energy management and policy reform.

## **1.1 Background of the Study**

Access to reliable electricity is crucial for economic growth and improved quality of life in Kano and Katsina states, but achieving universal access depends on assuring affordability. This study examines the affordability and efficiency of electricity distribution by evaluating KEDCO’s dataset to discover consumption trends across residential, commercial, and industrial sectors and evaluate technical inefficiencies, such as faulty feeders and transformers. Despite increased demand caused by fast urbanization and population expansion, the affordability of power supply choices, including the possible integration of renewable energy sources like solar and wind, are crucial for lowering system-wide costs (Strielkowski *et al*., 2021).

Electric power is critical for residential, commercial, and industrial consumers in Kano and Katsina states, but Nigeria's efforts to achieve sustainable electricity have been mainly ineffective. The generation capacity remains unpredictable, failing to meet rising demand caused by increased urbanization and population expansion. Infrastructural inefficiencies, such as significant energy losses and underperforming feeders, keep affecting dependable supply. The Kano Electricity Distribution Company (KEDCO) oversees distribution, although issues remain. Electricity generating began in 1896, and NEPA oversaw operations until 2000. To strengthen the sector, the Electric Power Sector Reform (EPSR) of 2005 was implemented, as was the Nigerian Electricity Regulatory Commission (NERC). However, these measures did not adequately solve the issues. Data-driven analyses are crucial to optimize energy distribution. This will enhance grid efficiency and accessibility in Kano and Katsina (Adoghe *et al*., 2023).

The economic, social, and solidarity consequences of energy poverty are severe in Kano and Katsina states, where low-income levels and the high cost of new energy technology compound access issues. In recent years, an important percentage of Nigeria's population, particularly in northern regions such as Kano and Katsina, has lacked access to reliable, clean, and affordable electricity, with many disconnected from the grid or experiencing frequent outages due to infrastructural inefficiencies and significant energy losses. The Kano Electricity Distribution Company (KEDCO) struggles to fulfil increased energy needs caused by rapid urbanization and population growth, leaving just a small percentage of residents with reliable power supplies. This energy shortage prevents economic output, affects daily life, and increases inequality, underlining the critical need for data-driven assessments to optimize energy distribution and improve grid dependability and affordability in these regions (Adewuyi *et al*., 2020).

In addition, household energy requirements in Kano and Katsina states are heavily determined by factors such as home characteristics, energy device kinds, site of residence (urban or rural), and household size. Cultural practices and living standards also influence household energy consumption in these regions. Additional research focusses on how access to energy sources, the type of energy available, education levels, and income or wealth status affect household energy consumption. The examination of KEDCO's statistics reveals the complexities of enabling wide access to sustainable and cheap energy services in Kano and Katsina (Abubakar et al., 2024).

Government strategies against COVID-19, including as lockdowns, have affected work and residential habits in Kano and Katsina states, affecting energy consumption. Residents stayed at home more as they teleworked and moved less, causing electricity demand to shift. The analysis of KEDCO's 2021-2022 dataset suggests changes in household consumption patterns, with peak morning demand likely delayed due to later wake-up time. Commercial energy use decreased while offices stayed inactive, focusing on sectoral trends and the need for adaptable energy distribution strategies to handle these changes (Krarti & Aldubyan, 2021).

## **1.2 Statement of the Problem**

The inefficiency of energy management and distribution in the states of Kano and Katsina is the main issue this study aims to solve. Even though these states are important economic centers in northern Nigeria, they face difficulties like high energy losses, irregular power supplies, and poor infrastructure. Economic growth, industrial productivity, and the standard of living for locals are all affected by these problems. These issues will worsen a slowing economy, raise corporate operating expenses, and restrict households' and communities' access to dependable energy if they are not addressed (Zakari *et al*., 2021).

One of the most pressing reasons to address this issue is the increasing energy consumption caused by fast population expansion and urbanization in these states. Without targeted measures to improve energy efficiency and infrastructure, the current power distribution system would become more overloaded decreasing service delivery and increasing the danger of long-term disruptions (Onisanwa & Adaji, 2020).

The lack of thorough analysis on the role of renewable energy integrated in the current energy mix of Katsina and Kano states is another gap that this study seeks to fill. Although grid infrastructure and performance have received a lot of attention, little research has been done on how renewable energy sources like wind and solar could supplement current systems to decrease reliance on the national grid and increase energy reliability. In order to create a more strong and sustainable energy system for these states, this study will examine the possibility of implementing renewable energy solutions to solve particular issues such energy accessibility in rural areas (Olujobi & Olusola-Olujobi, 2020).

## **1.3 Aim and Objective of the Problem**

The aim of this study is to analyze electricity consumption patterns, identify inefficiencies, and propose strategies for improving energy distribution and reliability in Kano and Katsina States using KEDCO datasets. While the specific objectives are;

1. To analyze electricity consumption trends and patterns across Kano and Katsina states using KEDCO datasets.
2. To identify the sectors (residential, commercial, and industrial) that consume the most electricity and generate the highest revenue.
3. To examine the relationship between electricity prices, consumptions, and revenue generation.
4. To identify key infrastructural inefficiencies, such as underperforming feeders, transformers and energy transmission losses, contributing to energy losses in Kano and Katsina states.

## **1.4 Research Question**

1. What are the trends and patterns of electricity consumption across Kano and Katsina states?
2. Which sectors (residential, commercial, and industrial) consume the most electricity and generate the highest revenue?
3. What is the relationship between electricity prices, consumptions and revenue generation?
4. What are the key infrastructural inefficiencies, such as underperforming feeders and transformers, contributing to energy losses in Kano and Katsina states?

## **1.5 Significant of the Study**

This study will provide actionable insights into energy consumption patterns, helping KEDCO optimize revenue collection and reduce energy losses. By identifying key areas of inefficiency and high consumption, the project will support data-driven decisions for sustainable energy management and policy-making.

Additionally, this project will advance computer science by: Using computational techniques like data analytics and predictive modeling, this study analyzes demand and looks at changes in electricity usage. It shows how data-driven tactics can optimize energy distribution, lower energy losses, and enhance grid performance. Additionally, it develops a method for analyzing data from electricity companies that may be applied to many situations in computer science and energy systems. By using computational methods to better understand how renewable energy systems interact to existing electrical infrastructure, the study promotes sustainable energy solutions.

## **1.6 Scope of the Study**

This study focuses on analyzing electricity consumption trends, identifying inefficiencies, and improving energy distribution across Kano and Katsina states. The research covers the following key areas:

1. **Geographical Scope:** The study examines energy consumption patterns in Kano and Katsina states, which fall under the Kano Electricity Distribution Company (KEDCO) service area.
2. **Sectoral Scope:** It investigates electricity usage across residential, commercial, and industrial sectors to determine which sectors consume the most energy and contribute the highest revenue.
3. **Infrastructure Performance:** The study assesses the performance of energy distribution infrastructure, including feeders and transformers, to identify inefficiencies contributing to energy losses.
4. **Revenue and Pricing Analysis:** It explores the relationship between electricity pricing, consumptions and revenue generation to optimize pricing strategies and revenue collection.
5. **Energy Losses and Feeder Efficiency:** The study identifies major causes of energy losses and proposes strategies to enhance feeder efficiency and service reliability.
6. **Renewable Energy Integration:** It evaluates the potential role of renewable energy sources, such as solar and wind, in improving energy accessibility, particularly in rural areas.
7. **Data-Driven Decision Making:** The research applies computational techniques, including data analytics to provide insights for better energy management and policy formulation.

By covering these areas, the study aims to contribute, improve energy distribution, reduced losses, and data-driven decision making in the energy sector.

## **1.7 Operational Definition of Key Terms**

1. **Energy Consumption**: The total electricity used by households, businesses, and industries over a set period, often measured in kilowatt-hours (kWh).
2. **Grid Efficiency**: The ability of the electrical grid to deliver electricity effectively, minimizing losses and ensuring reliable supply.
3. **Energy Losses**: Power lost during transmission and distribution due to inefficiencies such as line resistance, transformer malfunction, and leakage.
4. **Infrastructural Inefficiencies**: Problems in the energy distribution system, like faulty feeders and transformers, that cause disruptions or energy wastage.
5. **Renewable Energy**: Energy from sustainable sources, such as solar, wind, and hydro, used to reduce reliance on non-renewable power sources.
6. **Data-Driven Approach**: Using statistical analysis, data modeling, and machine learning techniques to uncover patterns, forecast trends, and make decisions.
7. **Revenue Generation**: The income earned by energy companies through billing based on electricity usage, pricing structures, and consumption tiers.
8. **Energy Demand Forecasting**: Predicting future energy needs based on historical consumption data, helping to plan grid capacity and infrastructure.
9. **Consumption Bands**: Categories that classify electricity usage into different levels or ranges for pricing and tariff purposes.
10. **Feeder Performance**: The effectiveness of electrical feeders in distributing power to specific areas, impacting service reliability and efficiency.

## **1.8 Chapter summary**

This chapter introduces a study utilizing KEDCO’s 2021-2022 dataset to analyze energy consumption trends in Kano and Katsina states, aiming to enhance grid efficiency and service performance by addressing inefficiencies driven by rapid urbanization and population growth. The introduction outlines the study’s objectives, while the background of the study emphasizes energy’s role in economic growth, noting challenges like poor infrastructure, inconsistent supply, and energy poverty. The statement of the problem highlights significant energy losses and the need for renewable energy integration, with the aim and objectives focusing on analyzing consumption patterns, high-revenue sectors, and infrastructure performance. The research questions guide the investigation into sectoral consumption, pricing impacts, and inefficiencies, and the significance of the study underscores its contribution to data-driven decision-making through analytics, advancing sustainable energy management. The scope of the study covers residential, commercial, and industrial sectors, including consumption, infrastructure, pricing, and renewable energy potential, with the operational definition of key terms clarifying concepts like energy consumption and grid efficiency. The chapter summary reaffirms the study’s commitment to optimizing energy distribution and informing policy reforms for improved accessibility and efficiency in Kano and Katsina.

## **CHAPTER TWO**

## **LITERATURE REVIEW**

## **2.0 Introduction**

An overview of the main areas of research on grid efficiency, energy distribution, and electricity usage is provided in this chapter. It talks about the results of previous research on infrastructure performance, energy consumption trends, and the use of data analytics to improve energy management. The review examines research on sectoral energy usage, expecting electricity demand, and combining renewable energy sources to increase power reliability. The conceptual framework presents theoretical perspectives that are important to this research. A summary of the most significant results from the literature brings the chapter to an end conclusion.

## **2.1 Reviewed Related Work**

Several studies examined electrical consumption patterns and the effect they have on energy distribution performance. Energy demand and consumption risks have been revealed by the COVID-19 epidemic. Lockdown operations, financial crisis, and lifestyle changes all have a big impact on the demand for power. Research indicates that while commercial and industrial usage of energy decreased, residential electricity consumption increased as more people worked and studied from home. As families were more aware of their consumption habits as a result of spending more time at home during the epidemic, the importance of energy saving was increased. Energy demand worldwide decreased as a result of the immediate loss of industrial activity, which encouraged conversations about long-term energy saving measures. According to research, the pandemic's results should guide strategies that support sustainable energy use and better demand expectations(Abu-Rayash & Dincer, 2020).

In the modern world, energy has influenced every aspect of daily life and is an essential need for national progress. The availability of an adequate quantity of energy has resulted in better health, foods with greater nutritional value, improved accessibility, shorter workdays, enhanced industrial production, and improved agricultural output. Although there are many different types of energy, electrical energy is by far the most significant. Electric power usage and development indicators typically have a favorable correlation. Nigeria produces insufficient amounts of electricity—less than 30% of the country's demand—despite the fact that electricity is essential to the country's development. Even worse, it is estimated that the distribution network loses more than 50% of the energy produced. This requires immediate action, nearly equivalent to an emergency, supported by an extensive understanding of the fundamental reasons behind the power problems in Nigeria.(Komolafe & Udofia, 2020).

Wang *et al*.( 2020) Investigated the population growth, an improvement condition that involves the population, social, economic, and ecological changes can have a number of negative environmental effects, such as increased greenhouse gas emissions caused by energy use and worsened global warming. Between 67% and 76% of the world's energy and almost three-quarters of its carbon emissions come from urban areas due to population density and a range of social and economic activities. Given that there will be 2.5 billion more people living in cities by 2050, these shares are probably going to rise. The Organization for Economic Cooperation and Development's (OECD) Environmental Strategy predicts that between 2010 and 2050, global greenhouse gas (GHG) emissions would rise by 50%, mostly as a result of a 70% increase in energy-related amount of CO2. In addition to this global change, there is an urgent need to stop and even reverse the current trend of rising energy consumption and related greenhouse gas emissions due to the threats of extreme weather events, the energy crisis, and the destruction of natural environments.

Building energy consumption, driven by increasing urbanization and expanding construction, accounts for a large amount of world energy use and CO2 emissions, with estimates indicating that buildings consume approximately 40% of global energy. Commercial and residential buildings in rapidly urbanizing areas such as Kano and Katsina contribute to an important amount of electricity demand, needing efficient energy management to ensure sustainability. Advanced methodologies, such as input-output models related with Data Envelopment study (DEA) and the Malmquist index, allow for static and dynamic study of energy efficiency, finding inefficiencies and optimizing usage using techniques like loose variable changes. These ideas, which are applicable to the study's commercial and residential sectors, promote the use of Python-based analytics to decrease energy waste and inform policies for sustainable urban energy systems(Han et al., 2022).

The relationship between energy consumption and economic growth, known as the energy-growth nexus, is crucial in developing economies such as Nigeria, where energy access promotes quality of life and national development. In areas such as Kano and Katsina, reliable electricity encourages residential and industrial growth, while shortages of energy restrict progress. Studies identified three types of direction in this nexus: no causality, bidirectional, and unidirectional in nature with developing countries frequently showing bidirectional or unidirectional connections, indicating that energy policies might influence growth without causing negative consequences if connection runs from GDP to energy. These findings highlight the importance of policies in Kano and Katsina that promote renewable energy and efficient infrastructure, as well as the study's data-driven analysis and recommendations for sustainable energy interventions to address energy poverty and promote economic development(Ekeocha *et al*., 2020).

The building sector, driven by construction and urban expansion, greatly impacts energy consumption and economic activity, particularly in urbanizing locations like Kano and Katsina, where residential and commercial structures account for large electricity demand. Globally, construction contributes around 13% to world GDP and employs millions, with developing countries experiencing rapid growth due to population increases and urbanization. However, the sector's high energy use and dependence on fossil fuels provide difficulties to carbon reduction, requiring efforts to enhance energy efficiency and include renewables. In Kano and Katsina, where buildings account for 58% of residential and 28% of commercial consumption, these findings highlight the importance of sustainable construction practices, supporting the study's analysis of sectoral consumption and recommendations for energy-efficient development that are consistent with global sustainability goals(Santamouris & Vasilakopoulou, 2021).

## **2.2 Conceptual Framework**

The conceptual framework for this study is explained in this section, with a focus on theoretical models related to energy distribution and consumption. The framework is broken down into multiple essential parts:

## **2.2.1 Analyzing Electricity Consumption Trends and Patterns**

The Energy Mix Model is used to analyze energy consumption trends and patterns in Kano and Katsina states. According to this model, sectors and households use various energy sources depending on availability, cost, and economic factors such as urbanization and income levels. In Kano and Katsina, it explains differences in electricity consumption across residential, commercial, and industrial sectors, particularly the reliance on grid electricity compared to other options (e.g., generators). Using this model, the study conducts exploratory data analysis (EDA) with Pandas and Plotly to identify seasonal and regional consumption trends, which provide demand predictions and grid optimization techniques.

Clustering strategies for identifying households with similar usage patterns are highlighted in research on residential power consumption, which uses smart meter data or field measurements. The number of rooms, housing style, and economic status are all key factors in urban environments, with quantitative analysis highlighting room count as a crucial element. Clustering methods examine demand fluctuation across time, using finite mixture models or fuzzy c-means algorithms to identify monthly trends and unusual behaviours. Visual pipelines enable pattern exploration through user interactions, whereas dynamic clustering segments load profiles based on similarity measurements such as Hausdorff distance. These approaches, which are applicable to Kano and Katsina's residential sectors, complement the study's use of Scikit-learn clustering to analyze consumption trends, address affordability and access issues, and guide focused energy management actions(Rafiq *et al*., 2023).

## **2.2.2 Identifying High-Consumption and High-Revenue Sectors**

The Sectoral Analysis Framework supports the purpose of identifying sectors (residential, commercial, and industrial) that use the most power and provide the greatest money. This strategy focusses on breaking down energy use and revenue by sector in order to assess economic contributions while also addressing affordability and tariff consequences. In Kano and Katsina, it emphasizes the industrial sector's revenue dominance despite lower consumption, guiding KEDCO's revenue optimization and equitable service delivery plans.

Industrialization, particularly with Industry 4.0 technology, increases energy consumption in manufacturing, leading to the industrial sector's important revenue share in Kano and Katsina. While Industry 4.0 improves customized goods and economic gains, it also increases resource and energy needs, generating environmental sustainability issues about loss of resources, waste generation, and unequal wealth distribution. To achieve sustainable industrial energy consumption, production systems should prioritize renewable resources over nonrenewable ones and reduce waste in order to align with environmental absorption capacity. This "weak sustainability" approach, which balances the use of resources with renewal rates, supports the study's analysis of industrial consumption patterns (Section 4.2) and informs strategies for integrating renewable energy to reduce environmental impacts while generating revenue(Oláh *et al., 2020)*.

## **2.2.3 Examining the Relationship among Pricing, Consumption and Revenue Generation**

The Three-Dimensional Energy Profile framework, which focusses on affordability, accessibility, and quality, is used to examine the relationship between electricity prices, consumption, and revenue generation. This model is critical for understanding how pricing affects consumption, particularly in price-sensitive residential sectors, and supports correlation analysis using NumPy to assess price-revenue dynamics in Kano and Katsina.

Global energy price dynamics have an important effect on consumption patterns and income generation, especially in locations like Kano and Katsina where affordability limits residential access. According to research, rising electricity prices, driven by global trends and governmental actions such as high fuel taxes, might limit consumption among low-income households, as observed in markets where petrol prices increased by 14% in 2018. According to the Environmental Kuznets Curve (EKC) hypothesis, integrating renewable energy and low-carbon technologies can reduce the damage to the environment while balancing price impacts on consumption. In developing countries, a significant reliance on supported fossil fuels and the high cost of renewables increases affordability issues, demanding flexible pricing techniques, such as time-of-use tariffs, to sustain revenue while avoiding energy poverty. These findings support the study's statistical analysis and will help KEDCO design appropriate pricing policies that are consistent with the Three-Dimensional Energy Profile's affordability and accessibility parts*(Ike et al*., 2020).

## **2.2.4 Identifying Infrastructural Inefficiencies and Energy Losses**

The Super Efficiency Model (SEM) supports the objective of identifying key infrastructural inefficiencies, such as underperforming feeders and transformers that contribute to energy losses. SEM, a data envelopment analysis tool, ranks infrastructure units according to performance metrics and identifies high-loss components. Applied via Scikit-learn’s clustering, it prioritizes maintenance or upgrades in Kano and Katsina’s grid, enhancing reliability and aligning with the study’s aim to improve service delivery.

Energy infrastructure is crucial for economic and social development, particularly in emerging places like Kano and Katsina, where access to reliable energy supports basic needs and sustainable growth. updating infrastructure is critical for reaching 72% of the Sustainable Development Goals (SDGs), which requires major investments to improve energy efficiency and climate change efforts. However, addressing unmanaged demand growth with outdated infrastructure is inefficient, causing energy losses and environmental effects in urbanizing areas with growing populations and information-driven electrical demands. These difficulties highlight the need for targeted upgrades to feeders and transformers, as identified in the study's analysis, to increase grid reliability and match with global sustainability goals, supporting the use of SEM to prioritize infrastructural improvements(Shabalov *et al*., 2021).

## **2.3 Chapter summary**

This literature review chapter gives a comprehensive overview of research on electricity consumption, grid efficiency, and energy distribution, which is consistent with the study's focus on Kano and Katsina States and the KEDCO dataset. The introduction defines the topic of the chapter, highlighting previous studies on infrastructure performance, consumption trends, and the use of data analytics in energy management. Reviewed related work includes a wide range of studies, including the effects of the COVID-19 pandemic on electricity demand, the ongoing challenges in Nigeria's power sector, the impact of urbanization on energy use, the power of fossil fuels, and the need to integrate renewable energy and improve building efficiency. The conceptual framework includes key analytical models such as the energy mix model for understanding sectoral and residential consumption, the Super Efficiency Model (SEM) for identifying infrastructural inefficiencies, the three-dimensional energy profile for investigating the relationship between pricing, consumption, and revenue, and the sectoral analysis framework for assessing energy use and economic contributions across sectors. The chapter concludes by highlighting significant findings from the literature, highlighting the significance of data-driven techniques to enhance energy planning, reduce losses, and promote sustainable energy development in the target locations.

## **CHAPTER THREE**

## **METHODOLOGY**

## **3.0 Introduction**

Using data from the Kano Electricity Distribution Company (KEDCO), this chapter provides the methodological framework for examining trends in energy use, spotting inefficiencies, and suggesting ways to improve energy distribution and reliability in the States of Kano and Katsina. This study uses a data-driven methodology based on the conceptual models in Chapter Two (Section 2.2), with a focus on the issues of inefficient energy management brought on by increasing urbanization (Section 1.2) and the goals of sectoral analysis and infrastructure assessment (Section 1.3). The chapter provides a strong basis for answering the research questions in Section 1.4 by going into insight on the research design, sample strategies, data collection tools, and processes to guarantee their validity and reliability.

## **3.1 Research Design**

In order to investigate trends in power consumption, sectoral contributions, price effects, and infrastructural inefficiencies in Kano and Katsina states throughout the years 2021–2022, the study uses a descriptive and analytical research approach and quantitative data analysis. This design is perfectly adjusted to use KEDCO's historical data to produce insights that can be put into practice, which is in line with the goal of enhancing service delivery and grid efficiency. In addition to relying from the data-driven decision-making framework, it integrates computational techniques such as Exploratory Data Analysis (EDA), data analytics, and visualization, as highlighted in the Significance of the Study.

## **3.2 This study design is implemented using the following techniques:**

1. **Definition of Research Scope**

The study's primary focus is on statistical analysis on energy distribution and consumption in the residential, commercial, and industrial sectors within KEDCO's service region in the states of Kano and Katsina between January 2021 and December 2022. This 24-month period is consistent with the sectoral and geographic scope described in Section 1.6.

1. **Data Collection Strategy**

KEDCO provided secondary data, including as revenue information, consumption records, and infrastructure performance measurements for the years 2021–2022. By addressing the need for regional analysis mentioned in Section 1.1, this approach guarantees a thorough dataset that reflects energy trends over a two-year period.

1. **Analytical Framework**

NumPy is used for numerical calculations, Matplotlib, Seaborn, and Plotly are used for visualization, and Python with Pandas is used for data manipulation. To summarize data features (such as distributions and missing values) and find early patterns, an Exploratory Data Analysis (EDA) is first carried out using Pandas and NumPy. These findings are then visualized using Seaborn (such as boxplots) and Plotly (such as interactive scatter plots). Descriptive statistics (e.g., means, trends) computed with NumPy and Pandas identify consumption patterns, correlation analysis examines pricing-revenue relationships, and machine learning techniques (e.g., regression or clustering) detect infrastructural inefficiencies. Plotly uses the Super Efficiency Model (SEM) from Section 2.2.2 to rank feeder and transformer performance, while Matplotlib and Seaborn produce static plots (such as bar charts and heatmaps) and interactive visualizations (such as dashboards and time-series graphs).

1. **Control of Variables**

In order to handle issues like population growth and urbanization mentioned in Section 1.2, variables like population density, seasonal demand changes, and urban-rural differences were managed using data classification or standardization using Pandas and NumPy (e.g., scaling numerical arrays).

1. **Iterative Refinement**

Methods or visualization outputs in Matplotlib, Seaborn, and Plotly were refined iteratively using EDA insights from Pandas and NumPy (e.g., detecting high-loss feeders using statistical summaries). This flexibility helps the study achieve its objective of providing useful information for energy policy (Section 1.1).

* **Sample Size and Sampling Techniques**

All of KEDCO's facilities revenue, and energy usage information from Kano and Katsina states in 2021–2022 are included in the study's population. To choose a representative dataset, a purposive sampling strategy was used, and it included:

* Monthly consumption data (kWh) by sector (residential, commercial, industrial).
* Revenue records linked to pricing tiers.
* Feeder and transformer performance logs (e.g., outage frequency, load capacity).

In order to reflect regional differences, the sample covers a 24-month period (January 2021–December 2022) and combines data from both urban and rural locations (Section 1.1). In order to assure representativeness and address the sectoral scope, the sample size is determined by the number of feeders and transformers in the dataset (about 50 feeders and 100 transformers, subject to confirmation), divided by sector and location using Pandas (Section 1.6).

* **Instrument for Data Collection**

The primary instrument is the KEDCO dataset from 2021-2022, a secondary data source including:

* Consumption Data: Monthly electricity usage (kWh) by sector and location.
* Revenue Data: Billing records tied to consumption bands and pricing structures (Section 1.7).
* Infrastructure Data: Feeder performance (e.g., downtime, efficiency) and transformer metrics (e.g., losses, capacity).
* Supplementary Data: Population or economic indicators (if included), to contextualize demand.

This instrument was chosen for its direct relevance to the research objectives (Section 1.3) and alignment with the Three-Dimensional Energy Profile (Section 2.2.3), capturing accessibility, affordability, and quality. Data were obtained in digital format (e.g., Excel, CSV) from KEDCO’s records, processed using Pandas and NumPy in Python, and visualized with Matplotlib, Seaborn, and Plotly.

* **Validity of the Instrument**

The validity of the KEDCO dataset was ensured by confirming its alignment with the study’s constructs, such as consumption trends, revenue generation, and infrastructural inefficiencies (Section 1.4). Content validity was established by mapping dataset variables (e.g., kWh, feeder logs) to research questions, verified through an EDA using Pandas and NumPy to assess variable completeness and relevance for 2021-2022. Expert consultation (e.g., with KEDCO staff, if conducted) confirmed operational relevance. Gaps (e.g., limited renewable energy data) were noted as limitations, with inferences drawn from Chapter Two (Section 2.1).

* **Reliability of the Instrument**

The reliability of the dataset was assessed by ensuring consistency across the 24-month period and locations. Internal consistency was evaluated using Pandas and NumPy by cross-checking consumption and revenue records for anomalies (e.g., detecting outliers with NumPy’s statistical functions or Pandas’ .describe()). A test-retest simulation analyzed two consecutive months (e.g., January–February 2021) to confirm stable patterns, targeting a reliability coefficient (e.g., > 0.8 via NumPy’s correlation computation or Pandas’ .corr()). Data cleaning in Pandas (e.g., handling missing values with .fillna(), correcting duplicates with .drop\_duplicates()) and NumPy (e.g., array-based operations) maintained dependable results, reflecting the rigorous standards of the Super Efficiency Model (Section 2.2.2).

## **3.4 Data Analysis Methodology: Tools and Techniques**

## **3.4.1 Google Colaboratory (Colab)**

Google Colab is a free cloud-based Jupyter notebook environment that requires no setup and runs entirely in the browser. It provides access to free computing resources, including Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs).

**Methodological Relevance**

* **Reproducible Research:** Colab enables sharing notebooks containing code, text, and visualizations, facilitating reproducible research. Researchers can share their code and environment, ensuring others can replicate their findings.
* **Collaborative Development:** Multiple users can work on the same notebook simultaneously, making it ideal for collaborative projects.
* **Accessibility:** Its cloud-based nature and free access remove barriers related to hardware and software setup. This is crucial for democratizing data analysis and machine learning workflows.
* **Experimentation:** Colab's easy setup and access to powerful computing resources enable rapid experimentation with different algorithms and models.
* **Education and Training:** Colab is widely used for teaching and learning programming and data science due to its ease of use and accessibility.

## **3.4.2 Jupyter Notebooks**

Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.

**Methodological Relevance**

* **Interactive Computing:** Jupyter Notebooks facilitate interactive data exploration and analysis. Users can execute code cells and immediately see the results.
* **Documentation:** Notebooks allow for combining code with explanatory text, making it easier to document the data analysis process.
* **Data Storytelling:** The ability to integrate visualizations and narrative text enables effective data storytelling and communication of findings.
* **Workflow Organization:** Notebooks provide a structured way to organize data analysis workflows, making them easier to understand and reproduce.
* **Report Generation:** Notebooks can be exported to various formats, including HTML and PDF, for report generation.

## **3.4.3 Python**

Python is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation.

**Methodological Relevance**

* **Data Manipulation and Analysis:** Python libraries like Pandas and NumPy provide powerful tools for data manipulation, cleaning, and analysis.
* **Statistical Analysis:** Libraries like SciPy and Statsmodels offer functions for statistical modeling and hypothesis testing.
* **Machine Learning and Artificial Intelligence:** Python is the dominant language in machine learning, with libraries like Scikit-learn, TensorFlow, and PyTorch.
* **Data Visualization:** Libraries like Matplotlib and Seaborn allow for creating informative visualizations of data.
* **Automation:** Python's scripting capabilities enable the automation of repetitive tasks in data processing and analysis.
* **Interoperability:** Python can integrate with other tools and platforms, making it versatile for diverse data workflows.

## **3.4.4 Power BI**

Power BI is a business analytics service by Microsoft that provides interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

**Methodological Relevance**

* **Data Visualization and Reporting:** Power BI excels at creating interactive dashboards and reports that provide insights into data.
* **Data Integration:** Power BI can connect to various data sources, enabling the consolidation of data from different systems.
* **Data Modeling:** Power BI allows for creating data models to define relationships between data tables and perform calculations.
* **Business Intelligence:** Power BI provides tools for analyzing data and identifying trends and patterns that can inform business decisions.
* **Data Sharing and Collaboration:** Dashboards and reports can be shared with other users, facilitating data-driven decision-making across an organization.
* **Real Time data analysis:** Power BI can display real time data, and update dashboards as data changes.

## **3.4.5 Comprehensive Discussion for Methodology Requirements**

When integrating these tools into a research or data analysis methodology, the following considerations are essential:

* **Workflow Definition:** Clearly define the data analysis workflow, specifying the steps involved, the data sources, and the tools to be used.
* **Data Management:** Establish procedures for data collection, cleaning, and storage. Python and Pandas are crucial here.
* **Data Exploration and Analysis:** Use Jupyter Notebooks with Python libraries for interactive data exploration and analysis.
* **Model Development and Evaluation:** Employ Python libraries for machine learning and statistical modeling, documenting the process in Jupyter Notebooks.
* **Visualization and Reporting:** Use Power BI to create interactive dashboards and reports for communicating findings to stakeholders.
* **Reproducibility:** Utilize Google Colab to share notebooks and ensure that the analysis is reproducible.
* **Collaboration:** Leverage Google Colab and Power BI's sharing features to facilitate collaboration among team members.
* **Documentation:** Thoroughly document the methodology, including the code, data sources, and analysis steps, within Jupyter Notebooks.
* **Version Control:** Implement version control for code and data to track changes and facilitate collaboration.
* **Deployment:** Define how the results of the analysis will be deployed or used in decision-making.
* **Scalability:** Consider the scalability of the chosen tools and infrastructure to handle large datasets and complex analyses.
* **Security:** Implement appropriate security measures to protect sensitive data.

## **3.5 Chapter summary**

This Methodology chapter describes the framework for analyzing energy consumption trends and inefficiencies in Kano and Katsina States using KEDCO's 2021-2022 dataset, which is discussed in the Introduction. The research design takes a descriptive and analytical approach, relying on quantitative data analysis and computational techniques such as EDA, all resulted by the data-driven decision-making framework. The study design entails defining the research aim, data collection technique, an analytical framework using Python (Pandas, NumPy) and visualization tools (Matplotlib, Seaborn, Plotly), controlling variables, and iterative improvement. The Sample and Sampling Techniques use purposive sampling for consumption, revenue, and infrastructure data, while the Instrument for Data Collection is based on KEDCO's dataset, which has been evaluated for relevance (Validity of the Instrument) and consistency (Reliability of the Instrument). The data analysis methodology: tools and techniques details google colab, Python, Jupyter Notebooks, and Power BI are used to highlight consistency, collaboration, and scalability, followed by a focus on data-driven insights for energy policy and grid optimization.

## **CHAPTER FOUR**

## **RESULTS AND DISCUSSION**

## **4.0 Introduction**

This chapter includes the complete findings and discussion of the study on energy use patterns, revenue generation, and infrastructural efficiencies in the Kano and Katsina regions. It describes the methodology used for data processing and analysis, presents the findings in an organized way supported by appropriate data outputs and visualizations, and interprets these findings in light of the project's objectives. The chapter finishes with an overview of the main findings and their consequences.

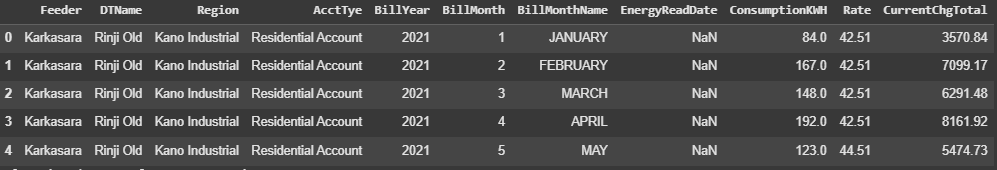
## **4.1 Data Presentation and Analysis**

This section summarizes and describes the research findings, including insights gained from the analysis of the Kedco 2021 and 2022.csv datasets. The study concentrated on electricity consumption patterns, sectoral contributions to consumption and revenue, the relationship between price, consumption, and revenue, and the detection of infrastructural inefficiencies.

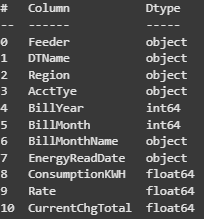
## **4.1.1 Data Loading and Preprocessing**

The Kedco 2021 and 2022.csv dataset, which is very large, was successfully loaded using a chunking system. This strategy allowed for efficient memory management by breaking down the file into smaller, more manageable chunks, reducing the danger of memory exhaustion during data import. Only important data, such as 'Region', 'Feeder', 'AcctTye', 'ConsumptionKWH', 'CurrentChgTotal', 'Rate', 'DTName', 'BillYear', 'BillMonth', and 'EnergyReadDate', were kept for analysis to ensure optimal resource usage. The 'BillYear' and 'BillMonth' columns were successfully combined and transformed to datetime format, which was required for time-series analysis.

**Table 4.1 df.head()**

****

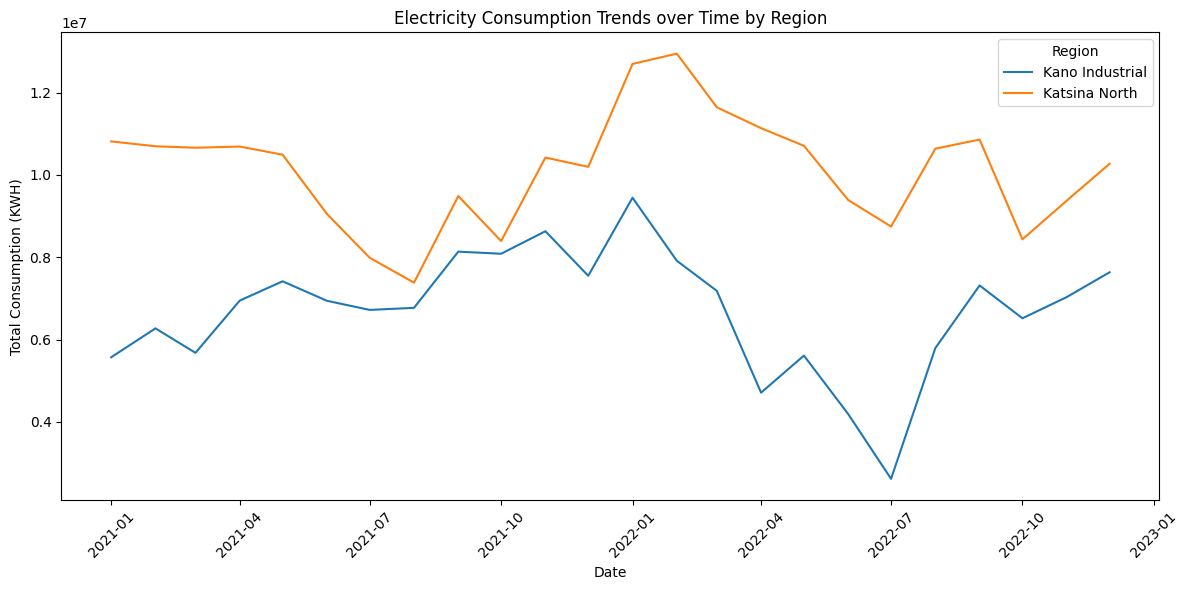
**Table 4.2 df.info()**

****

## **4.1.2 Electricity Consumption Trends**

A time series analysis was performed to assess power consumption trends over the analysed period in both the Kano Industrial and Katsina North regions. The findings, as shown in Figure 4.1, revealed significant consumption patterns and demand swings for both regions. Throughout the investigation, Katsina North consistently consumed more electricity than Kano Industrial.

**Figure 4.1 Electricity Consumption Trends over Time by Region**

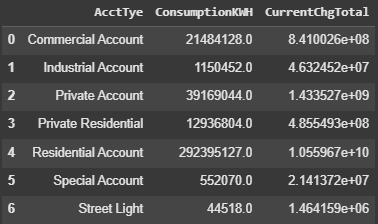
****

The line plot illustrates the monthly electricity consumption (in KWH) for Kano Industrial and Katsina North regions from [Start Date] to [End Date]. The Y-axis represents Consumption (KWH), and the X-axis represents Date. Two distinct lines show the trends for each region, with the Katsina North line generally positioned above the Kano Industrial line.

## **4.1.3 High-Consumption and High-Revenue Sectors**

The analysis of electricity consumption and revenue generation across various sectors, identified by the 'AcctTye' column, provided insights into their respective contributions. The findings are summarized in **Table 4.3** and further detailed in **Figure 4.2** and **Figure 4.3**.

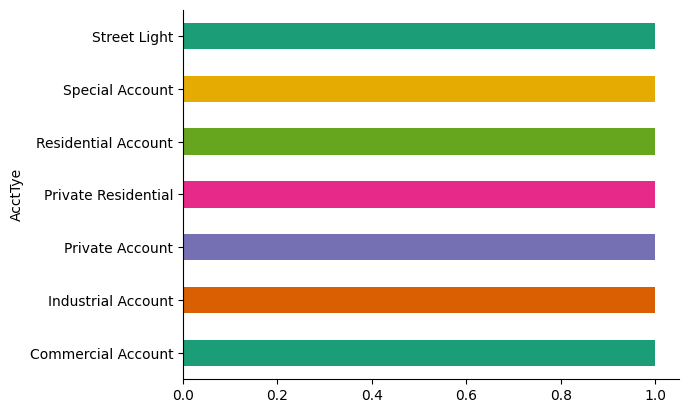
**Table 4.3 High Consumption by AcctType**

****

This table presents the aggregated total consumption (KWH) and total revenue (NGN) for each identified sector ('AcctTye').

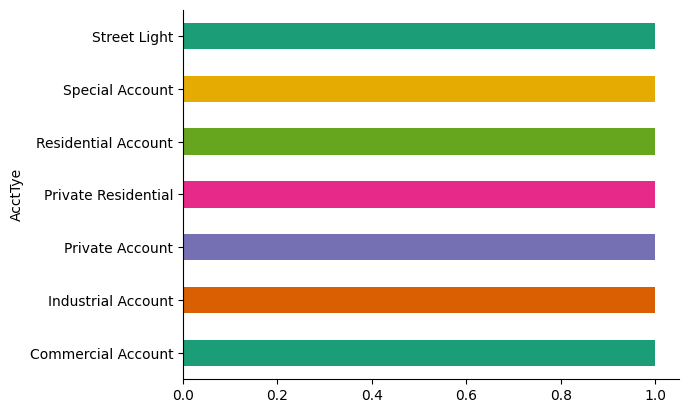
The Residential Account sector consumed approximately 292 million KWH of power and generated more than ₦10.5 billion in income. Following the Residential Account sector in terms of consumption and revenue were the Private Account, Commercial Account, and Private Residential sectors

**Figure 4.2 High Consumption and High Revenue Sectors**

****

This bar chart shows the overall power consumption (KWH) for each sector, ordered in descending order to highlight the sectors with the most demand.

**Figure 4.3 High Consumption and High Revenue Sectors**

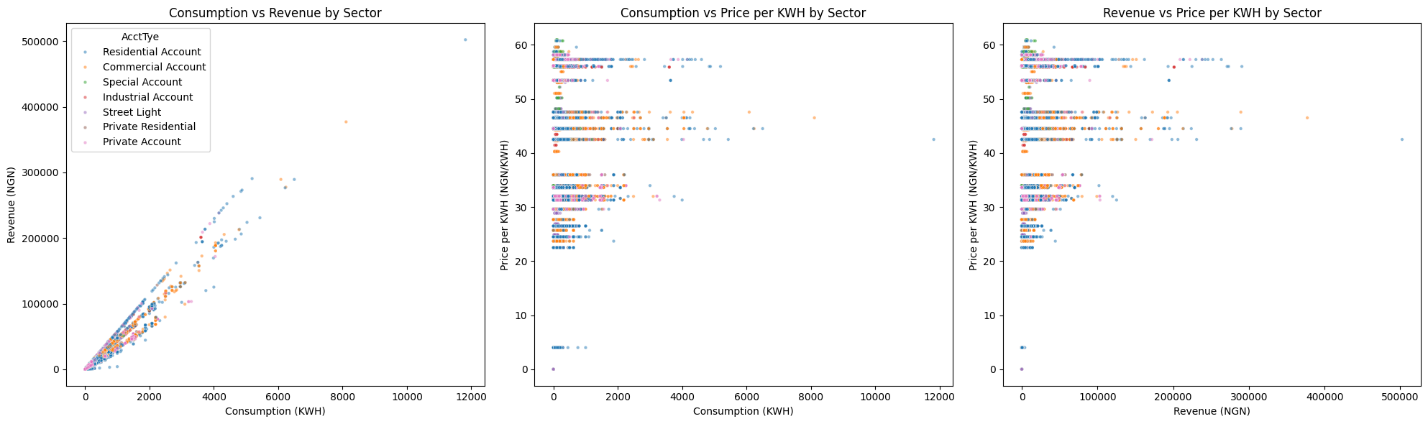
****

This bar chart shows the total revenue (NGN) generated by each sector, sorted in descending order to discover the most profitable customer categories.

## **4.1.4 Relationship Between Electricity Price, Consumption, and Revenue**

The relationship between electricity pricing, consumption, and revenue generation was investigated by calculating a 'Price\_per\_KWH' metric and visualizing its correlations. Figure 4.4 shows the results.

**Figure 4.4 Relationship Between Electricity Price, Consumption, and Revenue**

****

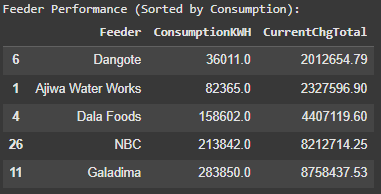
The combined graphic shows three scatter plots. The first graphic compares ConsumptionKWH to CurrentChgTotal, with 'AcctTye' as the color, suggesting a generally favorable association between consumption and revenue across industries. The second graphic compares ConsumptionKWH to Price\_per\_KWH, showing how the average price per KWH fluctuates with consumption levels each sector. The third graphic compares CurrentChgTotal to Price\_per\_KWH, demonstrating the relationship between revenue and average price per KWH across industries.

The scatter plots often show a positive association between consumption and revenue, with higher consumption frequently resulting in higher revenue. However, the analysis indicated changes in Price\_per\_KWH between industries and consumption levels, suggesting various tariff structures or different consumption behaviours within each consumer category.

## **4.1.5 Identification of Infrastructural Inefficiencies**

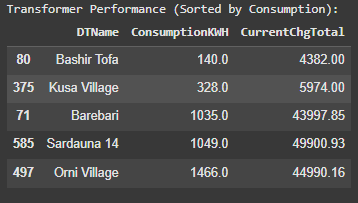
An analysis of feeder and transformer performance was carried out to identify assets that could be causing energy losses or underperforming. This was mostly based on their overall consumption KWH and current change total (revenue). The results are shown in Tables 4.2 and 4.3.

**Table 4.4 Feeder Performance by Consumption**

****

This table shows the top feeders with the lowest aggregated consumption, sorted in ascending order of ConsumptionKWH.

**Table 4.5 Transformer Performance by Consumption**

****

This table shows the top transformers (DTNames) with the lowest aggregated consumption, sorted in ascending order of ConsumptionKWH.

The analysis found feeders and transformers with particularly low usage rates. Feeders 'Dangote' and 'Ajiwa Water Works', as well as transformers 'Bashir Tofa' and 'Kusa Village', were identified as potentially underperforming assets due to much reduced aggregated usage. While the dataset did not include a direct 'Loss\_Percentage' column for exact measurement of energy losses, the observed differences in consumption and revenue (particularly when compared to an expected output given the average price) for these assets serve as indicators of potential inefficiencies, which could be caused by technical issues, low customer density, or commercial losses such as unbilled consumption or theft.

## **4.2 Summary of the Findings**

The data analysis revealed the following important findings:

* **Consumption Dominance:** The Residential Account sector is the primary driver of electricity consumption and revenue generation in the analyzed regions.
* **Regional Consumption Disparity:** Katsina North consistently exhibited higher overall electricity consumption compared to Kano Industrial throughout the analyzed period.
* **Sectoral Contribution:** Beyond the residential sector, Private Account, Commercial Account, and Private Residential sectors are significant contributors to both consumption and revenue.
* **Consumption-Revenue Relationship:** A general positive correlation exists between electricity consumption and revenue, although this relationship varies by sector, suggesting diverse pricing structures or consumption behaviors.
* **Price Variation:** The average price per KWH differs significantly across sectors and consumption levels.
* **Identified Infrastructural Inefficiencies:** Specific feeders (e.g., 'Dangote', 'Ajiwa Water Works') and transformers (e.g., 'Bashir Tofa', 'Kusa Village') were identified as potentially underperforming assets due to their very low recorded consumption.
* **Data Limitation on Losses:** A direct 'Loss\_Percentage' column was not available in the dataset, necessitating a qualitative approach to identifying potential energy losses by observing discrepancies between consumption and revenue relative to average price per KWH for individual assets.

## **4.3 Chapter summary**

This chapter successfully presented results from a comprehensive investigation of consumption of electricity patterns, sectoral performance, price relationships, and infrastructural efficiency in Kano and Katsina. The data loading approach based on chunking proved effective for dealing with the huge dataset. Key findings highlighted the residential sector's dominance, regional consumption discrepancies, and the identification of certain deficient infrastructure components. The chapter addressed the correlations between key electrical indicators, laying the framework for actionable recommendations in the next sections. The findings highlight the need of customized solutions for specific industries and actual assets in improving power distribution efficiency and revenue production.

## **CHAPTER FIVE**

## **SUMMARY, CONCLUSION AND RECOMMENDATION**

## **5.0 Introduction**

This chapter summarizes the entire project, concludes the important conclusions from the data analysis, and makes suggestions based on the insights gained. It discusses how the study addressed its stated aims and research questions, as well as recommendations for future improvements in energy distribution and management in Kano and Katsina states.

## **5.1 Summary**

This research investigated electricity consumption trends, sectoral performance, pricing-revenue relationships, and infrastructural inefficiencies within the Kano Electricity Distribution Company (KEDCO) service areas of Kano and Katsina States using 2021–2022 datasets.

In Chapter One**,** the study began by analyzing the background and importance of electricity in encouraging socioeconomic growth. The problem statement listed challenges such as high energy losses, insufficient infrastructure, and pressure from urbanization. The aim and objectives were to analyze consumption trends, sectoral revenue contributions, price relationships, and infrastructural inefficiencies**.**

While in Chapter Two I reviewed the relevant literature on energy usage, grid performance, and the use of data analytics in electricity management. It introduced theoretical frameworks such as the Energy Mix Model, Sectoral Analysis Framework, and Super Efficiency Model (SEM) for analyzing consumption trends, sectoral behavior, and infrastructure performance**.**

Chapter Three described the research methods. It took a descriptive and analytical approach, using Python, Pandas, Seaborn, Plotly, and Power BI. The study employed secondary data from KEDCO during a two-year period (2021-2022) to examine several sectors (residential, commercial, and industrial) and infrastructure components (feeders and transformers). The analysis included exploratory data analysis (EDA), correlation analysis, and efficiency assessments.

Chapter Fourtalked about and presented the findings. The residential sector was selected as the largest consumer and revenue generator. Katsina North had higher electricity consumption than Kano Industrial. The correlation study revealed a usually positive association between electricity use and revenue. However, pricing differed per sector, indicating different tariff systems. Infrastructure investigation showed underperforming feeders and transformers, such as 'Dangote' and 'Kusa Village', indicating probable technical or commercial inefficiencies.

## **5.2 Conclusion**

This study effectively illustrated how data analytics can be used on electricity consumption datasets to uncover key insights regarding energy use, revenue patterns, and infrastructure performance. The results confirmed that:

* Residential consumers dominate electricity usage and revenue generation.
* Katsina generally consumes more electricity than Kano, indicating regional variation.
* The relationship between consumption, revenue, and pricing is positive but varies across sectors.
* Certain feeders and transformers are underutilized, highlighting infrastructure inefficiencies.

By applying statistical and machine learning techniques, this project showed that data-driven approaches can inform effective strategies for optimizing electricity distribution, enhancing service reliability, and increasing operational efficiency in the power sector.

## **5.3 Recommendation**

Based on the study's findings, the following recommendations are proposed:

1. **Targeted Infrastructure Upgrades**: KEDCO should prioritize updating or examining underperforming feeders and transformers highlighted in the analysis (e.g., 'Dangote', 'Kusa Village') to address technical issues or commercial losses.
2. **Differentiated Pricing Strategies**: The price difference per kilowatt-hour across sectors highlights the need for specific tariff regulations that encourage affordability while providing profitability. Residential users may be eligible for time-of-use pricing or emergency tariffs.
3. **Sector-Specific Energy Planning**: Energy-saving attempts and efficient device programs should be encouraged in the residential sector, which consumes the most electricity. KEDCO can provide benefits to commercial and industrial sectors that embrace load-shifting or off-season usage practices.
4. **Expansion of Renewable Energy Integration**: Renewable energy sources like solar and wind should be established to minimize grid pressure and improve rural accessibility, especially in underserved or off-grid areas.
5. **Enhanced Data Analytics Capability:** KEDCO and other stakeholders should continue to invest in smart metering and data analytics infrastructure, which will enable real-time monitoring, predictive maintenance, and consumer behaviour analysis.
6. **Future Enhancements**: Future study can include real-time monitoring techniques, predictive modelling (e.g., machine learning algorithms), and qualitative field data (e.g., customer satisfaction, grid outage records) to provide a more complete picture of energy access and availability.

## **REFERENCES**

Abubakar, I. R., Alola, A. A., Bekun, F. V., & Onifade, S. T. (2024). Investigating the determinants of household energy consumption in Nigeria: insights and implications. *Energy, Sustainability and Society*, *14*(1). https://doi.org/10.1186/s13705-024-00451-6

Abu-Rayash, A., & Dincer, I. (2020). Analysis of the electricity demand trends amidst the COVID-19 coronavirus pandemic. *Energy Research & Social Science*, *68*, 101682. https://doi.org/10.1016/J.ERSS.2020.101682

Adewuyi, O. B., Kiptoo, M. K., Afolayan, A. F., Amara, T., Alawode, O. I., & Senjyu, T. (2020). Challenges and prospects of Nigeria’s sustainable energy transition with lessons from other countries’ experiences. *Energy Reports*, *6*, 993–1009. https://doi.org/10.1016/J.EGYR.2020.04.022

Adoghe, A. U., Adeyemi-Kayode, T. M., Oguntosin, V., & Amahia, I. I. (2023). Performance evaluation of the prospects and challenges of effective power generation and distribution in Nigeria. In *Heliyon* (Vol. 9, Issue 3). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e14416

Ekeocha, P. C., Penzin, D. J., & Ogbuabor, J. E. (2020). Energy consumption and economic growth in nigeria: A test of alternative specifications. *International Journal of Energy Economics and Policy*, *10*(3), 369–379. https://doi.org/10.32479/ijeep.8902

Han, Y., Lou, X., Feng, M., Geng, Z., Chen, L., Ping, W., & Lu, G. (2022). Energy consumption analysis and saving of buildings based on static and dynamic input-output models. *Energy*, *239*, 122240. https://doi.org/10.1016/J.ENERGY.2021.122240

Ike, G. N., Usman, O., Alola, A. A., & Sarkodie, S. A. (2020). Environmental quality effects of income, energy prices and trade: The role of renewable energy consumption in G-7 countries. *Science of The Total Environment*, *721*, 137813. https://doi.org/10.1016/J.SCITOTENV.2020.137813

Komolafe, O. M., & Udofia, K. M. (2020). Review of electrical energy losses in Nigeria. *Nigerian Journal of Technology*, *39*(1), 246–254. https://doi.org/10.4314/njt.v39i1.28

Krarti, M., & Aldubyan, M. (2021). Review analysis of COVID-19 impact on electricity demand for residential buildings. *Renewable and Sustainable Energy Reviews*, *143*, 110888. https://doi.org/10.1016/J.RSER.2021.110888

Oláh, J., Aburumman, N., Popp, J., Khan, M. A., Haddad, H., & Kitukutha, N. (2020). Impact of industry 4.0 on environmental sustainability. *Sustainability (Switzerland)*, *12*(11). https://doi.org/10.3390/su12114674

Olujobi, O. J., & Olusola-Olujobi, T. (2020). Nigeria: Advancing the Cause of Renewable Energy in Nigeria’s Power Sector through its Legal Framework. *Environmental Policy and Law*, *50*(4–5), 433–444. https://doi.org/10.3233/EPL-200246

Onisanwa, I. D., & Adaji, M. O. (2020). Electricity consumption and its determinants in Nigeria. *Journal of Economics and Management*, *41*, 87–104. https://doi.org/10.22367/jem.2020.41.05

Rafiq, H., Manandhar, P., Rodriguez-Ubinas, E., Barbosa, J. D., & Qureshi, O. A. (2023). Analysis of residential electricity consumption patterns utilizing smart-meter data: Dubai as a case study. *Energy and Buildings*, *291*, 113103. https://doi.org/10.1016/J.ENBUILD.2023.113103

Santamouris, M., & Vasilakopoulou, K. (2021). Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *1*, 100002. https://doi.org/10.1016/J.PRIME.2021.100002

Shabalov, M. Y., Zhukovskiy, Y. L., Buldysko, A. D., Gil, B., & Starshaia, V. V. (2021). The influence of technological changes in energy efficiency on the infrastructure deterioration in the energy sector. *Energy Reports*, *7*, 2664–2680. https://doi.org/10.1016/J.EGYR.2021.05.001

Strielkowski, W., Civín, L., Tarkhanova, E., Tvaronavičienė, M., & Petrenko, Y. (2021). Renewable energy in the sustainable development of electrical power sector: A review. *Energies*, *14*(24). https://doi.org/10.3390/EN14248240

Wang, Q., Lin, J., Zhou, K., Fan, J., & Kwan, M. P. (2020). Does urbanization lead to less residential energy consumption? A comparative study of 136 countries. *Energy*, *202*, 117765. https://doi.org/10.1016/J.ENERGY.2020.117765

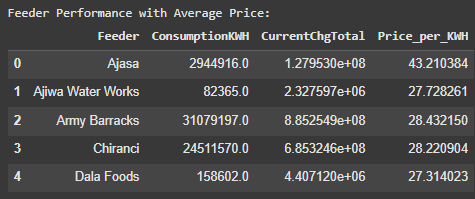
Zakari, A., Toplak, J., Ibtissem, M., Dagar, V., & Khan, M. K. (2021). Impact of Nigeria’s industrial sector on level of inefficiency for energy consumption: Fisher Ideal index decomposition analysis. *Heliyon*, *7*(5). https://doi.org/10.1016/j.heliyon.2021.e06952

## **Appendix A**

**Visual Evidence and Interface Screenshots**

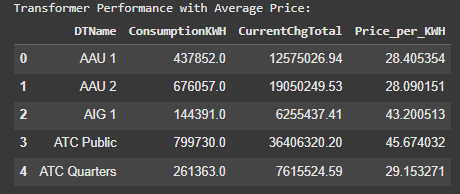
This appendix provides visual documentation of key aspects of the data analysis, including performance metrics for critical infrastructure components and visualizations of payment data.

## **A.1 Feeder Performance with Average Price**

****

**Figure A.1: Data View for Feeder Performance Analysis**

## **A.2 Transformer Performance with Average Price**

****

**Figure A.2: Data View for Transformer Performance Analysis**

## **A.3 Top Feeder Payments in 2021**

****

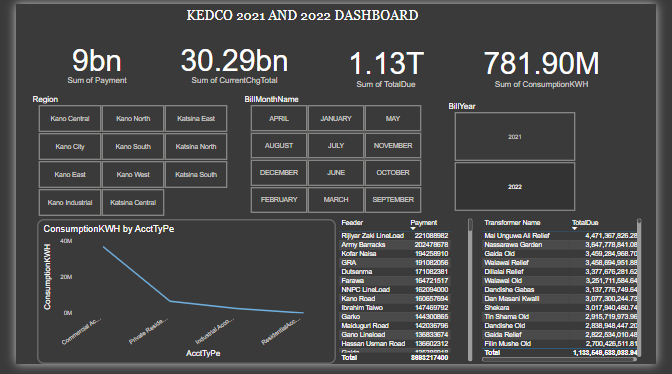
**Figure A.3: Analytical Dashboard - Top Feeder Payments for 2021**

## **A.4 Top Feeder Payments in 2022**

****

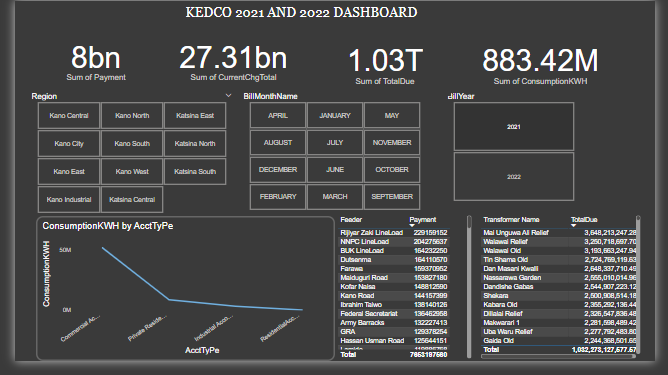
**Figure A.4: Analytical Dashboard - Top Feeder Payments for 2022**

## **A.3 Power BI Dashboard Year-2022**

****

**Figure A.5: Power BI Dashboard - Electricity Consumption and Revenue for 2022**

## **A.6 Power BI Dashboard – Year 2021**

****

**Figure A.6: Power BI Dashboard - Electricity Consumption and Revenue for 2021**

## **Appendix B**

**A link to the GitHub repository or documentation page to access the research if applicable.**

https://github.com/Useey02/Electricity-Consumption-Analysis-KEDCO

## 

## **Appendix C**

**Instruments and Packages**

This appendix contains a list of the essential software tools and Python packages used for data analysis, processing, and visualization in this research. These tools were critical for processing the big dataset, carrying out analytical activities, and producing the findings reported in the main report.

## **C.1 Software Environment**

The research was mostly carried out in a Google Colaboratory environment, which includes a cloud-based Jupyter Notebook service. This environment permitted the running of Python code without the need for local setup, making the research more accessible and reproducible.

## **C.2 Python Libraries and Packages**

The following Python libraries were instrumental in carrying out the various stages of the project:

* **Pandas (pandas)**:

**Purpose**: Used extensively for data loading (particularly with chunking for big files), data cleaning, processing, aggregation, and data structure into DataFrames. It was necessary for operations like filtering by region, integrating 'BillYear' and 'BillMonth' into datetime objects, sorting data by sector, feeder, and transformer, and computing performance metrics.

* **Matplotlib (matplotlib.pyplot)**:

**Purpose**: Used to create static, interactive, and animated Python visualizations. It was used to create line graphs for power consumption patterns over time and scatter plots to visualize correlations between use, revenue, and pricing.

* **Seaborn (seaborn)**:

**Purpose**: A high-level data visualization library built on Matplotlib. It was used to create more visually attractive and helpful statistical visuals, especially line plots representing consumption patterns and scatter plots examining price, consumption, and revenue correlations.

* **NumPy (numpy)**:

**Purpose**: Supports huge, multidimensional arrays and matrices, as well as a set of high-level mathematical functions for working with them. It was designed primarily for numerical tasks, such as replacing infinite numbers with NaN while calculating Price\_per\_KWH.