VISUALIZATION AND ANALYSIS OF ELECTRICITY DATA FOR SECTOR REVAMP: A CASE STUDY OF LAGOS STATE, NIGERIA.

BY
CHIOMA PERE-AFUN

Submitted to the
University of Roehampton

In partial fulfillment of the requirement for the degree of
Masters of Data Science

TABLE OF CONTENT

ABSTRACT

This case study examines the visualization and analysis of electricity data as part of a comprehensive sector revitalization initiative in Lagos State, Nigeria. The study's primary focus is on understanding collective sector-wide electricity consumption, rather than individual usage, and identifying opportunities for growth and improvement within Lagos' electricity sector. The report underscores the critical role of uninterrupted energy supply in driving economic growth, ensuring security, and safeguarding public health, emphasizing the pivotal contribution of data visualization in enabling strategic and well-informed decision-making.

The study encompasses the collection of data from both unmetered and metered customers, leveraging data visualization techniques to present energy consumption patterns in a manner that is intuitive and user-friendly. Furthermore, the report places significant emphasis on addressing the legal, social, and ethical considerations inherent in the process of data collection and analysis. It underscores the importance of informed consent, anonymity, and data confidentiality.

Within the report, the findings and insights derived from the visualization and analysis of electricity data are thoroughly discussed. Additionally, the study explores potential applications of the methodologies and techniques employed in this case study, offering a comprehensive perspective on the implications and opportunities arising from this data-driven approach.

DECLARATION

I hereby state that this report is a product of my own efforts. Whenever I have incorporated language, ideas, expressions, or writings from external sources, I have diligently provided proper references and attribution. I affirm that this report represents original work that has not been previously submitted for the attainment of any other academic degree in any other educational institution.

DATE: SEPTEMBER 06, 2023

Signed(apply signature below)

ACKNOWLEDGEMENT

My sincere appreciation goes to Almighty God for His grace and enablement all through the course of this research work. I wish to appreciate my supervisor Rabail Tahir for her unflinching support and guidance during this project. Your advice and contributions are the reason for the successful completion of this work. Many thanks to my darling husband, Pere Afun for being my strong support system throughout the program and my children- Praise, Amarachukwu, and Favour, for all their sacrifices and patience during this research, not forgetting my course mates who were very helpful in the course of my research, particularly Alireza and Oreoluwa.

This project is dedicated to my amazing children: Praise Pere-Afun, Amarachukwu Pere-Afun, and Favour Pere-Afun

1. INTRODUCTION

Energy is a crucial factor in the economic growth, progress, and development of any nation, as well as poverty eradication and security [1]. The availability of uninterrupted energy supply is a vital issue for all countries today. Future economic growth depends on the long-term availability of energy from sources that are affordable, accessible, and environmentally friendly. Security, economic growth, and public health are closely interrelated with energy. In recent times, data has been a vital tool in making strategic and informed decisions by individuals, private organizations, and government bodies. Data visualization has been shown to be an effective tool for helping electricity users understand and anticipate energy usage [5]. Traditional energy usage data presented in tables or graphs can be difficult for people to understand and interpret. Hence, data visualization provides a more intuitive and user-friendly way of presenting energy consumption data. In a country like Nigeria however, there are not enough studies focused on sensitizing households and businesses to electricity consumption patterns hence a need to design visualizations to help them understand their collective energy consumption patterns. This study is looking to explore collective sector consumption, not individual consumption, exploring their interaction with energy with the aim to identify room for growth and opportunity in the electricity sector of Lagos Nigeria.

In Nigeria, there are two channels for consumers to access electricity: unmetered customers and metered customers. Unmetered customers are those that draw current and are connected to the Distribution Network without a meter recording their energy consumption, while metered customers subscribe to buying as much energy as they can afford because they have a meter that can account for their energy usage. However, the ownership of prepaid meters is very low, and only 37% of the population has access to it, hence a larger percentage of consumers are charged based on electricity bills by distribution companies' estimation or direct billing method [8]. Unmetered users pay the highest as a result of estimated billing, and 57% of Nigerian electricity consumers are exploring other options due to being overcharged. Also, due to frequent power outages that can span for days or weeks in certain areas, customers don't get access to electricity. Frequently, users still get charged for unused electricity during these periods, especially unmetered users. These factors contribute to a downfall in customer retention and loss of revenue for distribution companies as customers opt for other energy options.

An analysis of data from the second quarter 2022 report of the Nigerian Electricity Regulatory Commission and its first quarter 2022 report showed a marginal reduction in the number of unmetered power consumers across the country [9]. Indicating that unmetered power users dropped from 7,802,427 in the first quarter of last year to 7,744,909 in the second quarter, showing a reduction of 57,518. This study will look to identify patterns in regions where unmetered and metered customers are situated to not only identify if there are decreases but also why, and also make recommendations that would improve electricity distribution. With the visualization, the study hopes to help the consumers understand their regional consumption flow over the years spanning across various industries in their location, the times of the year when they should anticipate minimal supply, and also consistent supply.

In a study to understand energy consumption and key industrial growth in China, Zou, G. (2022)[10] found that energy consumption could be linked to growth in the retail sector and that in the short run, a 1% growth in the industry would result in energy reduction by 0.48%. This confirms that there is in fact a link between sector growth and electricity consumption. In this research work, we will be exploring the visualization and analysis of electricity consumption data of households as well as industries like education, healthcare, manufacturing, hospitality, and corporate/public organizations or bodies in Lagos Nigeria, identifying lapses, areas for growth, and opportunities, as well as making consumption forecasts for the next one years.

1.1 Research Questions or Problems that will be Addressed

1. Are seasonality, trend, and cyclical components present in the data?
2. If yes, what periods of the month do we observe said trends?
3. Which industry consumes more electricity?
4. Which industry consumes less electricity?
5. Are there underserved regions in the state? If yes, why?
6. Can this study provide recommendations that corroborate the insights from the data?

1.2 Aims

The aims of this study are:

1. To visualize the Lagos State electricity data in an easily comprehensible, informative, and user-friendly way.
2. To identify patterns and trends that can be useful to the decision-making process for households and also other sectors within the states.
3. To simplify the electricity consumption data so the residents of Lagos State, Nigeria, both households and other sectors, can better understand their consumption habits over the years. This will help them stay prepared for any changes in the transmission of electricity and also changes in payment tariffs in relation to trends like seasonality.
4. To help both the consumers and the distribution companies transact better as both parties will become aware of their consumption and distribution habits over the years, thus helping them anticipate and better prepare for changes such as a reduction in the amount of transmitted energy as well as reduction or increase in tariffs in relation to trends like seasonality.
5. To help the distribution company identify areas for growth and opportunities within the state.
6. To forecast the consumption trend for the next 1 year.

1.3 Objectives

1. This study will adopt Jupyter Notebook, a web-based open-source platform used for interactive computing across multiple programming languages, and Python, a programming language for data cleaning, time series, and machine learning forecast. Tableau, an interactive data visualization software company focused on business intelligence, will be used for the visualization.
2. Python libraries like SkLearn, Panda, Numpy, Tensorflow, Statsmodels, will be used for data cleaning, time series forecast, and visualization, machine learning forecast.
3. Data cleaning and preprocessing will be done. Given the sensitivity of the data, information such as company names as well as the distribution company will be excluded and then cataloged into industries as opposed to individual household or organization electricity information.
4. This study will make projections for the next two years using time series models, ARIMA and SARIMA, juxtaposing them to identify which model performs better in terms of accuracy.
5. This study will visualize the data in two (2) outputs, and present them to residents of Lagos states in the bid to access which output clearly communicates and also retains information.

1.4 Legal, Social, Ethical, and Professional Considerations

This focuses on the analysis of the societal implications of, but not limited to, cutting-edge biomedical research and technologies[1]. This study focuses on the – (i) legal ramifications of improper data use and unauthorized access of data, (ii) informing society of their electricity consumption over time across diverse sectors, and proposing ideas and solutions that are beneficial to the distribution company and revamp of the electricity sector at large, (iii) ethical confidentiality of private business data, as well as individuals, and animosity of participants in the opinion survey.

1.4.1 LEGAL CONSIDERATIONS

According to Solicitors Regulation Authority, UK, a legal issue is an event that occurs and has legal implications which may require the help of a legal counsel to sort out. Such events may include disputes, conflicts of interest, and violations of laws and agreements. Hence, legal concerns in the context of this study may arise from the unauthorized access to the electricity data and improper use of said data. Being that electricity distribution is managed by privately owned companies, unauthorized access to their business data is a strong ground for litigation as well as divulgence of their data outside the scope of the study. The data for this study was obtained legally from the Public Relations unit of the company, which as a requirement for the release of their data, will be named Alexis Distribution Company.

1.4.2 SOCIAL CONSIDERATIONS

Many users believe that they are overcharged for electricity consumption, however, they do not have in-depth knowledge of the energy consumption of their region over the years and why they are charged that way. They think that the distribution companies are to be blamed for the exorbitant rates they pay but if they understand that having unmetered manufacturing companies, hospitals, and other establishments that rely heavily on electricity in their environment would cause a spike in their energy consumption, they better understand how to address the issue. So presenting the electricity data in an easily comprehensible and retentive way would be beneficial to society at large.

1.4.3 ETHICAL CONSIDERATIONS

Ethical considerations in research are a set of principles that guide the research designs and practices. It is imperative for researchers and scientists to adhere to a certain code of conduct when collecting data from people. This study involves data collection constraining sensitive information about private businesses and government bodies, its crucial that there is informed consent, animosity, and confidentiality. To avoid a breach of trust between Alexis Distribution Company and its customers, the names, addresses (location) and details of its customers were removed. Each customer was cataloged based on the industry they fall into [2]. Only findings and questions in relation to consumption will be explored in this study using their data. Also, full animosity will be implored when undergoing the opinion survey in the visualization evaluation phase. Participants, within the age range of 22-60, will be required to fill out a content form documenting their willingness to partake in the opinion survey in the form of a questionnaire. Full animosity will be guaranteed as no personal information would be required from them.

Finally, this project will draw inspiration from other academic works focused on data visualization and storytelling as a tool for communication and retention, as well as electricity consumption and forecasting. No intellectual property will be violated and all references used will be acknowledged and properly cited.

1.5 Report Overview

Chapter 1: This section comprises the introduction, aims, objectives, and research questions that would guide the direction of the project. It also contains the legal, social, and ethical considerations that were explored during the course of the research.

Chapter 2: The section entails the literature review and theoretical approach of data visualization as a tool for communication and information retention, and time series forecasting. Previous academic journals in these areas will be explored highlighting areas of convergence and divergence and what this study aims to do differently.

Chapter 3: This section expresses in detail the design and methodology flowchart and the technology review.

2. LITERATURE REVIEW

It is crucial to understand the concept of visualization to develop a communication tool for information dissemination and strategic decision-making. Thus, the first part of this chapter reviews the related concepts of data visualization. The second part previews literature about time series forecasting techniques implementable in electricity analysis. Also, literature on machine learning-based forecasting approaches applicable to electricity consumption data will be reviewed. Finally, the technologies and tools to be used will be reviewed.

2.1 Data Visualization

The use of data visualization has become increasingly popular in recent years as a means of communicating complex information in a more accessible and engaging way. In today's world where there is information overload, communicating with the intent to ensure information retention is as important as making said information easily comprehendible. In a study conducted by Kernbach et al(2010) [7], the researchers sought to compare the effectiveness of visualization and text in conveying business strategy. They conducted an experiment involving 74 managers who were exposed to a presentation outlining the simplified strategy of BMW Financial Services. The researchers manipulated the visual support provided, using three distinct types: text presented through PowerPoint, visualization in the form of a visual metaphor, and a roadmap. Each participant was exposed to only one of the three types of visual support. The researchers measured the impact on attention, recall, and attitude toward the strategy as their primary evaluation criteria. They found statistical evidence showing that the utilization of visualizations in the form of a visual metaphor and a temporal diagram resulted in significantly more favorable perceptions of the presenter compared to the condition where only text in the form of bulleted lists was used.

In terms of electricity consumption, because of the complexity and vastness of electricity data, just any kind of viaulization will not aid comprehension and retention. The choice of visualization also matters. In research by Herrmann et al (2018)[6] found that the choice of data visualization has a significant impact on people's ability to interpret domestic energy usage data. In their experiment, they used three different visualization techniques: aggregated, disaggregated, and normalized visualization. The aggregated visualization showed the total energy consumption of the household, while the disaggregated visualization showed the energy consumption of individual appliances. The normalized visualization showed the energy consumption of individual appliances as a percentage of the total energy consumption. They found that the normalized condition yielded the best results, as area-based graphs are more suitable for summarizing consumption over time than line graphs, hence recommending that summary overviews are better than time-series data visualizations to enable people to understand their domestic energy usage data. The study by Kontokosta et al(2017)[11] also corroborates Herrman et al[6] study that summarized overviews are most efficient in communicating electricity consumption. In Kontokosta et al's work [11], the choice of visualization, interactive map, scatter plot, and histogram that display summary statistics about the selected property was based on the need to provide building stakeholders with a user-friendly and interactive tool that allows for accurate comparisons to be made between different properties. Based on these literature, it is evident that visualization of complex data helped with comprehension, and using visual representations such as dashboard summary overviews allows residents and businesses to understand energy data better and also easily identify patterns and trends in their energy usage, and make informed decisions. Hence, this research will be using descriptive statistic charts/summary overview charts to visualize electricity consumption data. Charts like piecharts, histograms, bar charts, heatmaps, distribution charts, and scatterplots. This research will comprise a time series analysis, it will not be used to visualize the data.

2.2 Time Series Forecast

Time series forecasting involves the analysis of time series data through statistical methods and modeling to make predictions, enabling strategic decision-making. In various industries, electricity plays a critical and defining role in their progress, necessitating the need for efficient planning of energy accessibility and demand by producers, distributors, and end-users. Electricity consumption forecasting, commonly known as Load Forecasting, can be categorized into four groups: very short-term load forecasting (VSTLF), short-term load forecasting (STLF), medium-term load forecasting (MTLF), and long-term load forecasting (LTLF) [12, 13, 14]. VSTLF is applicable in real-time control and predicts energy demand within minutes to one hour ahead. STLF, on the other hand, focuses on forecasting within a range of one hour to seven days or months ahead. MTLF is utilized for predicting periods spanning from one week to one year. Finally, LTLF serves to forecast energy demand beyond a year, up to 20 years ahead and proves valuable in predicting new generation construction, strategic planning, and changes in the electric energy supply and delivery system [12, 14]. One of the objectives of this study, forecasting for two years, we will concentrate on the literature pertaining to LTLF.

The majority of previous studies focused on forecasting the total consumption demand (both residential and commercial) on distribution companies. However, in this particular study [12, 15], the researchers proposed a long-term (10 years) load forecasting approach using Neural Networks and Autoregressive Integrated Moving Average (ARIMA) to predict the Electrical Energy Demand (EED) of Kuwait. For this forecasting model, several attributes were considered as independent variables, including weather temperature and humidity, average salary, gross domestic product, oil price, population, residence, currency earning rate, and economic factors such as total import and export in USD. The study's findings revealed that Neural Networks (NN) outperformed ARIMA in terms of forecast accuracy. Additionally, the researchers discovered that weather parameters played a more significant role in the forecasting process compared to average salary, gross domestic product, and oil price.

To enhance the accuracy of electricity consumption forecasting, researchers incorporated historical electricity data along with Twitter data as input variables into a hybrid forecast model combining Artificial Neural Network (ANN) and Support Vector Machine (SVM). This model was utilized to forecast electricity consumption in Dutch [12,16]. The study conducted a comparison between ANN and SVM, with the findings indicating that ANN outperformed SVM in terms of forecast accuracy. However, the SVM exhibited improved performance in long-term forecasting. Despite this, the researchers acknowledged that including weather data as an input did not lead to an increase in the model's overall performance.

In their research, Sulandari et al. [12,16] introduced a hybrid model combining Artificial Neural Network (ANN) and a Fuzzy algorithm, along with a recurrent formula (LRF), to predict electricity demand in Indonesia. The findings indicated that this hybrid approach performed favorably, showing low values of Root Mean Square Error (RMSE). Similarly, in another study [12,17], a hybrid model was proposed, utilizing a clustering technique (K-means) in conjunction with the Autoregressive Integrated Moving Average (ARIMA) forecasting model to predict electricity demand for university buildings. The paper revealed that this hybrid model outperformed the standalone ARIMA model as a forecasting tool.

In another study [18], ANN model was used, specifically the Curve-Fitting Neural Network (CFNN) showed high efficiency for commercial long-term load forecasting for 10 years. The evaluation method used to justify the accuracy of the forecast involved calculating the Percentage Error along with Root Mean Square Error (RMSE) and Mean Square Error (MSE). In the study[19], Long Short-Term Memory - Recurrent Neural Networks hybrid model was constructed and the results were evaluated using a Mean Absolute Percentage Error (MAPE). Across all the papers, Neural networks, and regression models, either as stand-alone or hybrid were the most effective load forecasting models. The accuracy and error measures used were Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). However, there was a difference in the independent variables as that is subjective to the attributes in the data as well as the focus of the study. This study will be working with ARIMA, ANN, and SVM models. The independent variables will be Location, Industry, and Connection Type (metered and unmetered).

2.3 Tools

Most of the literature reviewed in the study used standard web technologies: HTML, CSS, and Javascript, for data visualization. This study considered using Microsoft PowerBI and Tableau, however, only Tableau Public will be used as the main visualization tool. Python will be used for data cleaning and the time series forecast.

Tableau Public: According to Tableau website, it is a business intelligence, data visualization, and analytics platform aimed at making it easier for people to explore and manage data, and faster to discover and share insights that can change businesses and the world. It connects to various data sources, including file formats like CSV, JSON, XML, and MS Excel, relational and nonrelational data systems such as PostgreSQL, MySQL, as well as cloud systems like AWS, Oracle Cloud, and Microsoft Azure [22]. Its uniqueness is that it has data blending features that allow for the seamless integration of data from multiple sources. Considering the opinion survey to be conducted, Tableau was the most preferable tool as it allows for sharing of visualization files hence setting it as a great and flexible tool for research-focused works. Also, Tableau is very responsive and solemnly hangs when working with large datasets compared to other visualization tools, and given the vastness of the dataset to be used in this study, it was the most viable option. Also, it slows for multiple dashboard designs on the same workbook, and that will come in handy in this study as we would be exploring consumption in relation to various industries, also connection types, and locations. Multiple dashboards will be created in relation to each section. Also Tableau public is free for academic use.

Microsoft PowerBI: Microsoft PowerBI, is also a data visualization tool. Microsoft Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence. It performs relatively the same purpose and has similar features to the Tableau public like sharing files, working with diverse data types, and live collaboration. It is also integrated with AI services hence making it a very great visualization tool. However, the app is only effective when using the paid version. The free PowerBI tool is not as responsive as Tableau and takes a longer time to load. The live collaboration is also limited on the free application which will not be beneficial to the opinions survey section of the study. Overall, it is expensive to manage.

Python: Python is a high-level, general-purpose programming language that prioritizes code readability with the use of significant indentation. Python is home to over 300 libraries of which this study will be using five of them.

Pandas: is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language[24]. Its functionality includes analyzing, cleaning, exploring, and manipulating data. It will be used in this study to eliminate irrelevant attributes, clean the data, treat missing values, change data types, and transform the data into a format suitable for the time series and machine learning models.

Numpy: Numpy is a library used for working with numerical data in Python and is relevant when working with other libraries like Pandas, SciPy, Matplotlib, and scikit-learn (SkLearn). Being that this study will be implementing machine learning models hence using sci-kit learn, the NumPy library is vital for our study[25].

Sklearn: This is another open-source machine learning library in Python that provides a wide range of tools for various machine learning tasks, including regression modeling. This study will be making use of the regression model, Support vector machine hence it's crucial to the study.

Tensorflow: TensorFlow is an open-source machine learning framework developed by the Google Brain team, primarily released in November 2015. It has since become one of the most popular and widely used libraries for building and training deep learning-based machine learning models. This study will be using the TensorFlow library to design the multilayered artificial neural network[27].

Statsmodels: Statsmodels is a Python library that provides classes and functions for the estimation of many different statistical models, hypothesis tests, and statistical data exploration. Statsmodels has the ability to handle time series data and implement ARIMA models, which are widely used for time series forecasting. ARIMA stands for AutoRegressive Integrated Moving Average, and it is a popular approach for modeling and forecasting time series data. ARIMA models are especially useful for non-stationary time series, where the mean and variance change over time[26].

3. METHODOLOGY OR DESIGN

3.1 Data Collection and Cleaning

In this academic study, the electricity distribution data of Lagos state from 2018-2022, which is managed by private organizations, will be used for analysis. It is important to note that due to the sensitivity of the data, it was not publicly accessible. Acquiring the data legally took a period of two weeks, and strict conditions were adhered to, prohibiting the disclosure of the company name, customer names, and specific locations. The dataset comprises various attributes, but we will be limited to the following:

1. **Business Unit**, representing the region to which the electricity feeder belongs. The dataset consists of six Business Units, namely Shomo, Egba, Wonjo, Eja, Orodu, and Hodi, which are not their actual names for confidentiality.
2. **The Feeder Nomenclature** attribute provides specific feeder addresses, usually associated with transformers' locations.
3. **The Utility Unit** attribute indicates the subregion to which the feeders belong.
4. **Metering Status** differentiates between customers using either prepaid meters or estimated billing.
5. **The Capacity** attribute focuses on the number of customers supplied data via a feeder, representing the number of houses or businesses served by a specific transformer.
6. **The Consumption (kWh)** attribute reveals the total electricity consumed by a customer in a given month.
7. Additionally, the **Industry attribute** categorizes customers into different sectors, such as manufacturing, healthcare, education, hospitality management, households, and corporate and public organizations.

To prepare the data for analysis, the pandas library will be utilized for data cleaning. This process involves addressing issues such as missing values, correcting data types, and removing irrelevant attributes. Furthermore, to safeguard the company's data from competitors, variables or attributes will be renamed in compliance with legal considerations. In conclusion, the data will undergo preprocessing to ensure its suitability for further analysis and visualization in accordance with academic standards and ethical considerations.

3.2 Building The Visualization

In this research, the visualization of the data will be presented through seven distinct dashboards. The first dashboard will offer an overview and summary of energy consumption in each business unit, providing insights into the performance of consumption in these units, the progression of total capacities, distribution of metering status, utility units, and industries. This comprehensive section aims to provide viewers with a holistic view of the attributes interacting with energy consumption (kWh). The other six dashboards will focus individually on each business unit, showcasing how the remaining attributes interact with it. This approach allows viewers to choose whether to explore the entire visualization or concentrate on their specific region of interest.

To analyze the data, both univariate and multivariate descriptive statistics charts will be utilized, exclusively in this section of the study[29]. Univariate descriptive statistics charts present visual summaries of data for a single variable, capturing key characteristics of the variable's distribution while maintaining its original integrity. Examples of univariate charts include stem-and-leaf displays, distribution/line charts, box plots, and violin plots. On the other hand, multivariate descriptive statistics charts are employed to summarize data for multiple variables, displaying trends and relationships between them. These charts are particularly useful for exploring complex data sets. However, in cases where there are numerous variables or a large number of individuals to be displayed, multivariate charts can become cluttered and challenging to interpret[29]. In such situations, using numerical summary statistics, such as averages or correlations, in tabular form would provide a more efficient summary. Examples of multivariate charts include scatterplots, heat maps, and parallel coordinate plots.

Throughout this study, both univariate and multivariate visualization options will be explored to depict the variables effectively and gain valuable insights from the data. By employing various visualization techniques, the research aims to provide a comprehensive understanding of the relationships and patterns within the dataset.

3.3 Forecasting.

ARIMA

ARIMA models are commonly used in electricity demand forecasting due to their ability to capture the temporal dependencies and seasonality in the data. ARIMA models are typically implemented using statistical software packages such as R or Python, which provide functions for fitting and evaluating ARIMA models. ARIMA (AutoRegressive Integrated Moving Average) is a popular time series model used for forecasting. It is a combination of three components: autoregression (AR), differencing (I), and moving average (MA). The AR component refers to the use of past values of the variable being forecasted to predict future values. The order of the AR component (p) specifies the number of past values used in the model. The I component refers to differencing the time series data to make it stationary. Stationarity means that the statistical properties of the data (such as mean and variance) do not change over time. Differencing involves subtracting the current value from the previous value to remove trends and seasonality. The order of differencing (d) specifies the number of times the data is differenced. The MA component refers to the use of past forecast errors to predict future values. The order of the MA component (q) specifies the number of past errors used in the model.

The ARIMA model is denoted as ARIMA(p,d,q). The parameters p, d, and q are determined by analyzing the autocorrelation and partial autocorrelation plots of the time series data. The autocorrelation plot shows the correlation between the time series and its lagged values, while the partial autocorrelation plot shows the correlation between the time series and its lagged values after removing the effects of intermediate lags.
The ARIMA model is mathematically expressed as follows:

$y(t)=c+a1yt-1+\ldots+apyt-p+et(1)$

$y(t)=\mu+ut+mlut-1+\ldots+mqut-q(2)$

In these equations, $a_1,\ldots,a_p$, and $m_1,\ldots, m_q$ are the parameters representing the autoregressive (AR) and moving average (MA) portions respectively, and c is the constant term. The orders of the AR and MA portions are denoted by p and q, respectively. The term *e* represents white noise or the error term, and μ represents the expectation of *y(t)*[11]. The equation (2) indicates the integration of the two models using the same training data, resulting in the final form of the ARIMA (p,q) model:

$y(t)=c+alyt-1+\ldots+apyt-p+ut+m1ut-1+\ldots+mqut-q(3)$

Support Vector Machine

SVM is a machine learning algorithm that can be used for regression and classification tasks. In the context of electricity demand forecasting, SVM can be used to predict future electricity demand based on historical data and other relevant variables such as weather data, time of day, and day of the week. For this study, we will be using SVM for the regression task. The dependent variable in this case is the Consumption (kWh), while the other attributes will be serving as the independent variable. Being that the data passed the cleaning and preprocessing phase, the next step would be to split the data into training and testing sets and train them on an SVM model on the training data using appropriate kernel functions and hyperparameters. Finally, we evaluate the performance of the model on the testing data and if the evaluation metric show great promise, it could be then used to forecast.

Artificial Neural Network

Artificial Neural Networks (ANNs) are machine learning algorithms inspired by the human brain's structure and function. They consist of interconnected nodes, or neurons, and can be used for tasks like classification and regression. In electricity demand forecasting, ANNs predict future demand using historical data and relevant variables as they handle complex relationships well. ANNs have input, hidden, and output layers. The input layer receives the input data, which is then processed by one or more hidden layers, and finally, the output layer produces the predicted output. The number of layers and neurons in each layer can vary depending on the complexity of the task and the amount of data available. Activation functions add nonlinearity to the model. Activation functions are mathematical functions that are applied to the output of each neuron in the network. Activation functions introduce nonlinearity into the network, allowing it to model complex relationships between input and output variables. Common activation functions include the sigmoid function, the hyperbolic tangent (tanh) function, and the rectified linear unit (ReLU) function. Choosing the right function and layer configuration is crucial for optimal performance. This study would experiment with 10 total layers, and also the ReLU and Sigmod activation functions to identify which is optimal. The ANN process is similar to the SVM approach, however, the distinction is that it is built on the Tensorflow and keras libraries which allow for multidimensional or layered analysis.

3.4 Evaluation Plans.

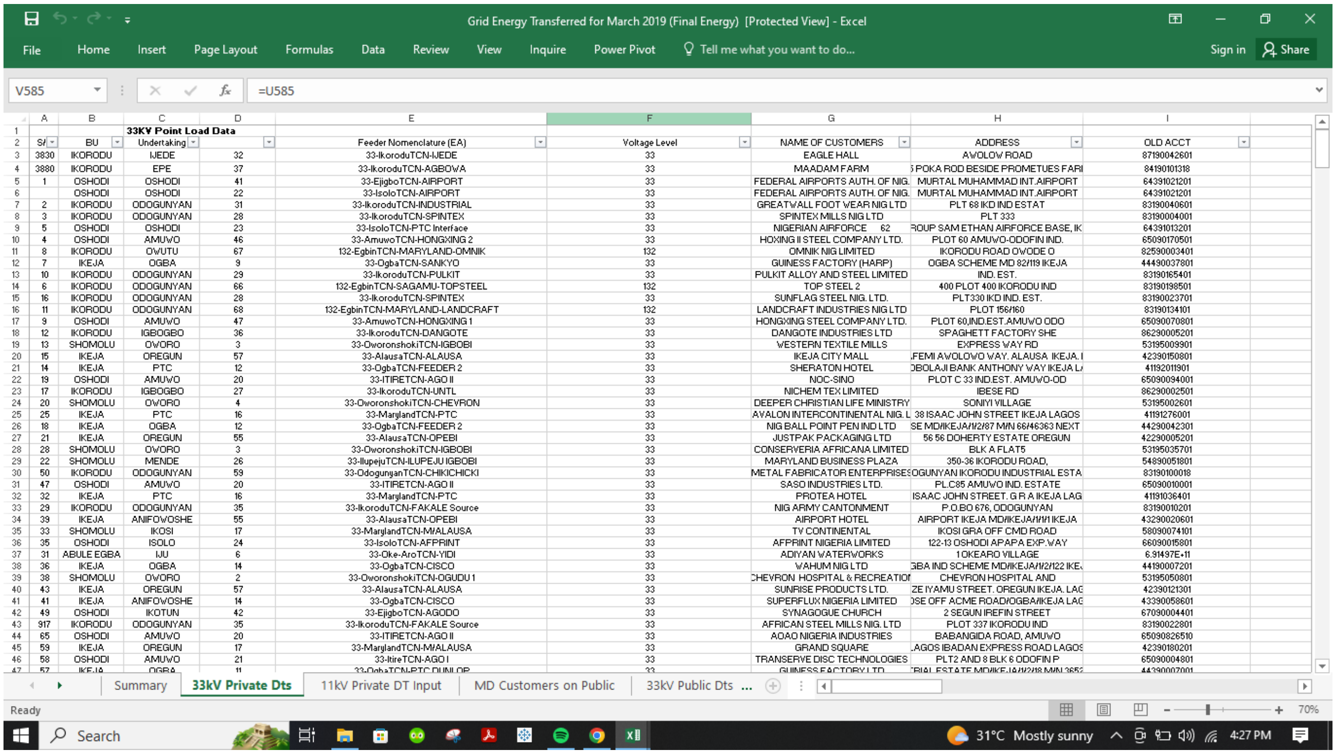
The forecast models will be evaluated using evaluation metrics such as MAPE, MAE, and RMSE.

4. IMPLEMENTATION OF RESULT

4.1 Forecasting

Data Cleaning

The raw data contained information on the companies in each local government as opposed to the industry. First, cross-checking of the companies was done, and they were then sectioned into the industry they belonged to. The industries created were Housing, Healthcare, Education, Consulting, Manufacturing/Production, and Hospitality. This was done using Excel as the dataset was large and contained multiple sheets for each month.



33KV Point Load Data

S/	BU	Undertaking		Feeder Nomenclature (EA)	Voltage Level	NAME OF CUSTOMERS	ADDRESS	OLD ACCT
3830	IKORODU	IJEDE	32	33-IkoroduTCN-IJEDE	33	EAGLE HALL	AWOLOW ROAD	87190042601
3880	IKORODU	EPE	37	33-IkoroduTCN-AGBOWA	33	MAADAM FARM	5 POKA ROD BESIDE PROMETUES FARI	84190101318
1	OSHODI	OSHODI	41	33-EjigboTCN-AIRPORT	33	FEDERAL AIRPORTS AUTH. OF NIG.	MURTAL MUHAMMAD INT.AIRPORT	64391021201
	OSHODI	OSHODI	22	33-IsoloTCN-AIRPORT	33	FEDERAL AIRPORTS AUTH. OF NIG.	MURTAL MUHAMMAD INT.AIRPORT	64391021201
2	IKORODU	ODOGUNYAN	31	33-IkoroduTCN-INDUSTRIAL	33	GREATWALL FOOT WEAR NIG LTD	PLT 68 IKD IND ESTAT	83190040601
3	IKORODU	ODOGUNYAN	28	33-IkoroduTCN-SPINTEX	33	SPINTEX MILLS NIG LTD	PLT 333	83190004001
5	OSHODI	OSHODI	23	33-IsoloTCN-PTC Interface	33	NIGERIAN AIRFORCE 62	ROUP SAM ETHAN AIRFORCE BASE, IK	64391013201
4	OSHODI	AMUWO	46	33-AmuwoTCN-HONGXING 2	33	HOXING II STEEL COMPANY LTD.	PLOT 60 AMUWO-ODOFIN IND.	65090170501
8	IKORODU	OWUTU	67	132-EgbinTCN-MARYLAND-OMNIK	132	OMNIK NIG LIMITED	IKORODU ROAD OWODE O	82590003401
7	IKEJA	OGBA	9	33-OgbaTCN-SANKYO	33	GUINESS FACTORY (HARP)	OGBA SCHEME MD 82/119 IKEJA	44490037801
10	IKORODU	ODOGUNYAN	29	33-IkoroduTCN-PULKIT	33	PULKIT ALLOY AND STEEL LIMITED	IND. EST.	83190165401
6	IKORODU	ODOGUNYAN	66	132-EgbinTCN-SAGAMU-TOPSTEEL	132	TOP STEEL 2	400 PLOT 400 IKORODU IND	83190198501
16	IKORODU	ODOGUNYAN	28	33-IkoroduTCN-SPINTEX	33	SUNFLAG STEEL NIG. LTD.	PLT330 IKD IND. EST.	83190023701
11	IKORODU	ODOGUNYAN	68	132-EgbinTCN-MARYLAND-LANDCRAFT	132	LANDCRAFT INDUSTRIES NIG LTD	PLOT 156/160	83190134101
9	OSHODI	AMUWO	47	33-AmuwoTCN-HONGXING 1	33	HONGXING STEEL COMPANY LTD.	PLOT 60,IND.EST.AMUWO ODO	65090070801
12	IKORODU	IGBOGBO	36	33-IkoroduTCN-DANGOTE	33	DANGOTE INDUSTRIES LTD	SPAGHETT FACTORY SHE	86290005201
13	SHOMOLU	OWORO	3	33-OworonshokiTCN-IGBOBI	33	WESTERN TEXTILE MILLS	EXPRESS WAY RD	53195009901
15	IKEJA	OREGUN	57	33-AlausaTCN-ALAUSA	33	IKEJA CITY MALL	.FEMI AWOLOWO WAY. ALAUSA IKEJA. I	42390150801
14	IKEJA	PTC	12	33-OgbaTCN-FEEDER 2	33	SHERATON HOTEL	OBOLAJI BANK ANTHONY WAY IKEJA L/	41192011901
19	OSHODI	AMUWO	20	33-ITIRETCN-AGO II	33	NOC-SINO	PLOT C 33 IND.EST. AMUWO-OD	65090094001
17	IKORODU	IGBOGBO	27	33-IkoroduTCN-UNTL	33	NICHEM TEX LIMITED	IBESE RD	86290002501
20	SHOMOLU	OWORO	4	33-OworonshokiTCN-CHEVRON	33	DEEPER CHRISTIAN LIFE MINISTRY	SONIYI VILLAGE	53195002601
25	IKEJA	PTC	16	33-MarylandTCN-PTC	33	AVALON INTERCONTINENTAL NIG. L	38 ISAAC JOHN STREET IKEJA LAGOS	41191276001
18	IKEJA	OGBA	12	33-OgbaTCN-FEEDER 2	33	NIG BALL POINT PEN IND LTD	SE MD/IKEJA/1/2/87 M/N 66/46363 NEXT	44290042301
21	IKEJA	OREGUN	55	33-AlausaTCN-OPEBI	33	JUSTPAK PACKAGING LTD	56 56 DOHERTY ESTATE OREGUN	42290005201
28	SHOMOLU	OWORO	3	33-OworonshokiTCN-IGBOBI	33	CONSERVERIA AFRICANA LIMITED	BLK A FLAT5	53195035701
22	SHOMOLU	MENDE	26	33-IlupejuTCN-ILUPEJU IGBOBI	33	MARYLAND BUSINESS PLAZA	350-36 IKORODU ROAD,	54890051801
50	IKORODU	ODOGUNYAN	59	33-OdogunyanTCN-CHIKICHICKI	33	METAL FABRICATOR ENTERPRISES	OGUNYAN IKORODU INDUSTRIAL ESTA	83190100018
47	OSHODI	AMUWO	20	33-ITIRETCN-AGO II	33	SASO INDUSTRIES LTD.	PL.C85 AMUWO IND. ESTATE	65090010001
32	IKEJA	PTC	16	33-MarylandTCN-PTC	33	PROTEA HOTEL	ISAAC JOHN STREET. G R A IKEJA LAG	41191036401
29	IKORODU	ODOGUNYAN	35	33-IkoroduTCN-FAKALE Source	33	NIG ARMY CANTONMENT	P.O.BO 676, ODOGUNYAN	83190010201
39	IKEJA	ANIFOWOSHE	55	33-AlausaTCN-OPEBI	33	AIRPORT HOTEL	AIRPORT IKEJA MD/IKEJA/1/1/1 IKEJA	43290020601
33	SHOMOLU	IKOSI	17	33-MarylandTCN-M/ALAUSA	33	TV CONTINENTAL	IKOSI GRA OFF CMD ROAD	58090074101
35	OSHODI	ISOLO	24	33-IsoloTCN-AFPRINT	33	AFPRINT NIGERIA LIMITED	122-13 OSHODI APAPA EXP.WAY	66090015801
31	ABULE EGBA	IJU	6	33-Oke-AroTCN-YIDI	33	ADIYAN WATERWORKS	1 OKEARO VILLAGE	6.91497E+11
36	IKEJA	OGBA	14	33-OgbaTCN-CISCO	33	WAHUM NIG LTD	GBA IND SCHEME MD/IKEJA/1/2/122 IKE.	44190007201
38	SHOMOLU	OWORO	2	33-OworonshokiTCN-OGUDU 1	33	CHEVRON HOSPITAL & RECREATIOI	CHEVRON HOSPITAL AND	53195050801
43	IKEJA	OREGUN	57	33-AlausaTCN-ALAUSA	33	SUNRISE PRODUCTS LTD.	ZE IYAMU STREET. OREGUN IKEJA. LAG	42390121301
41	IKEJA	ANIFOWOSHE	14	33-OgbaTCN-CISCO	33	SUPERFLUX NIGERIA LIMITED	OSE OFF ACME ROAD/OGBA/IKEJA LAG	43390058601
49	OSHODI	IKOTUN	42	33-EjigboTCN-AGODO	33	SYNAGOGUE CHURCH	2 SEGUN IREFIN STREET	67090004401
917	IKORODU	ODOGUNYAN	35	33-IkoroduTCN-FAKALE Source	33	AFRICAN STEEL MILLS NIG. LTD	PLOT 337 IKORODU IND	83190022801
65	OSHODI	AMUWO	20	33-ITIRETCN-AGO II	33	AOAO NIGERIA INDUSTRIES	BABANGIDA ROAD, AMUWO	65090826510
59	IKEJA	OREGUN	17	33-MarylandTCN-M/ALAUSA	33	GRAND SQUARE	LAGOS IBADAN EXPRESS ROAD LAGOS	42390180201
58	OSHODI	AMUWO	21	33-ItireTCN-AGO I	33	TRANSERVE DISC TECHNOLOGIES	PLT2 AND 8 BLK 6 ODOFIN P	65090004801

The obtained results were transformed into a CSV (Comma-Separated Values) file, which subsequently served as the foundation for forecasting and visualization tasks. To uphold the legal and ethical commitments outlined in the research, the business units were anonymized and replaced with pseudonyms: 'Egba,' 'Wonjo,' 'Eja,' 'Orodu,' 'Hodi,' and 'Shomo.' Furthermore, the dates were converted into the DateTime data type for accurate timestamp representation. I also obtained the statisrical properties of the data using the describe function.

To facilitate the analysis process and effectively manage data points associated with each of the aforementioned pseudonymous business units, a series of dataframes were created. These dataframes were instrumental in streamlining data analysis and ensuring that data related to each group could be easily and effectively collected and analyzed.

```
[ ] business_unit_mapping = {
        'ABULE EGBA': 'Egba',
        'AKOWONJO': 'Wonjo',
        'IKEJA': 'Eja',
        'IKORODU': 'Orodu',
        'OSHODI': 'Hodi',
        'SHOMOLU': 'Shomo'
    }

    # Replace the values in the "Business Unit" column
    data['Business Unit'] = data['Business Unit'].replace(business_unit_mapping)

    # Display the modified DataFrame
    data.head()
```

	Date	Business Unit	Housing	Healthcare	Education	Manufacturing/Production	Hospitality	Consulting
0	1/1/2019	Egba	29005.64	313.78	117.24	582.731775	150.737272	66.99
1	1/1/2019	Wonjo	42967.86	334.70	635.17	621.583662	816.646951	362.95
2	1/1/2019	Eja	32039.02	1918.92	6533.44	3563.704638	8400.132287	3733.39
3	1/1/2019	Orodu	47547.38	4462.64	511.07	8287.769101	657.091238	292.04
4	1/1/2019	Hodi	46534.40	5705.02	1512.07	10595.032060	1944.092169	864.04

When analyzing time series data, the time component or attribute of the data must be made the index hence I first made the Date attribute the index of the data. After this was done, I went on to treat the forecasting and timeline visualization of each Business Unit.

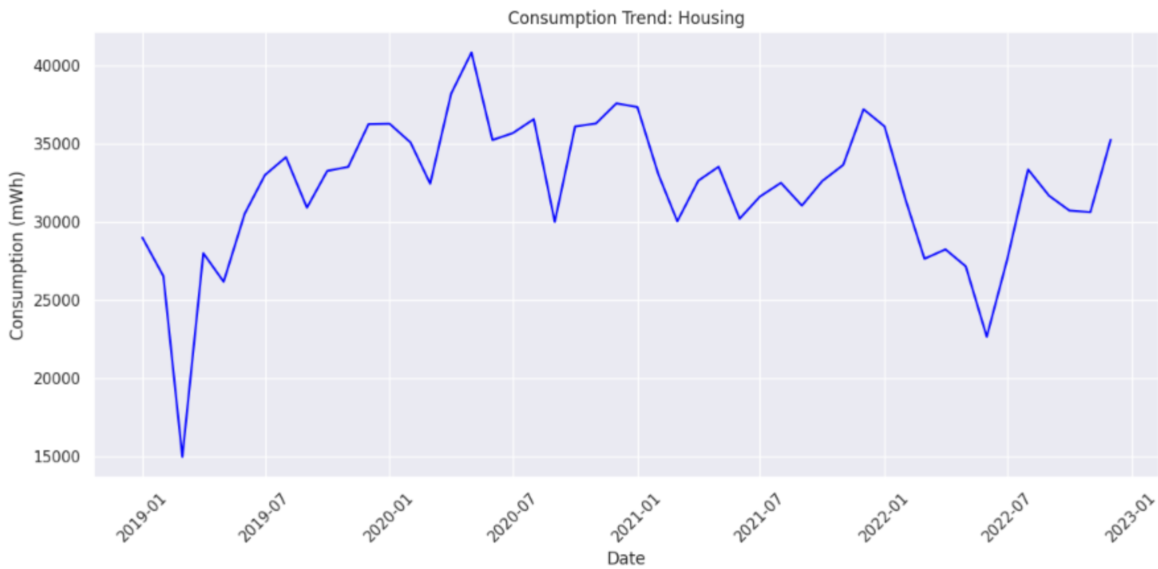
```
data.Date=pd.to_datetime(data.Date)
data.set_index('Date', inplace=True)
data.head()
```

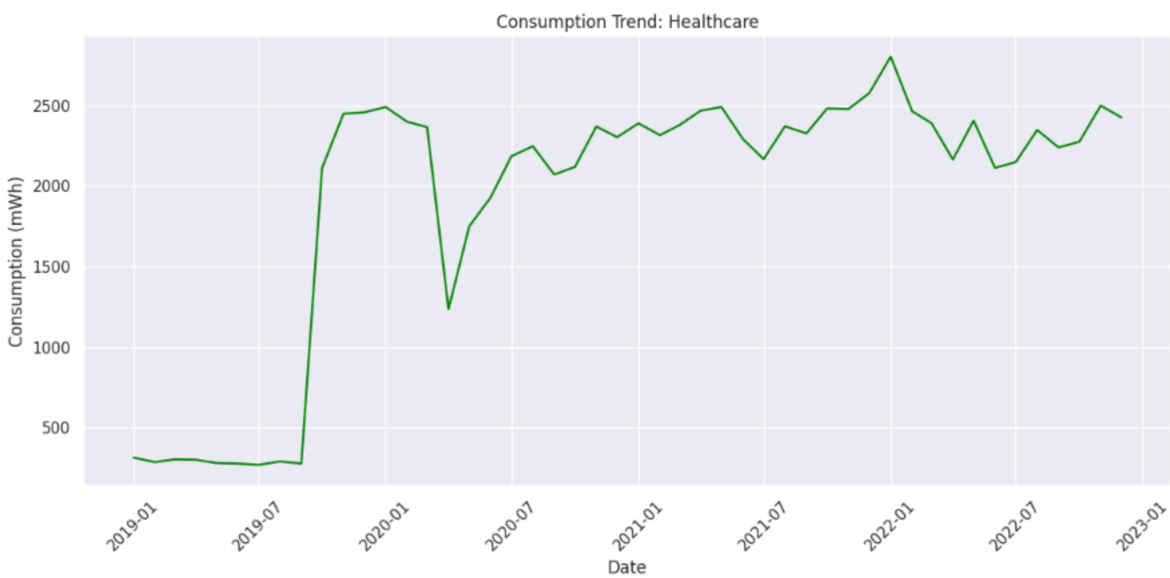
Date	Business Unit	Housing	Healthcare	Education	Manufacturing/Production	Hospitality	Consulting
2019-01-01	Egba	29005.64	313.78	117.24	582.731775	150.737272	66.99
2019-01-01	Wonjo	42967.86	334.70	635.17	621.583662	816.646951	362.95
2019-01-01	Eja	32039.02	1918.92	6533.44	3563.704638	8400.132287	3733.39
2019-01-01	Orodu	47547.38	4462.64	511.07	8287.769101	657.091238	292.04
2019-01-01	Hodi	46534.40	5705.02	1512.07	10595.032060	1944.092169	864.04

Egba Business Unit

Initially, I created a line chart or timeline graph that displays the variations in energy consumption over the years. This visualization served a dual purpose: firstly, it allowed for a visual representation of the data, and secondly, it facilitated an initial assessment for the presence of seasonality, trends, cyclical patterns, or irregular components. While additional statistical tests will be conducted for a more comprehensive analysis, this visualization offers valuable insights and provides a preliminary direction for the ensuing time series analysis.

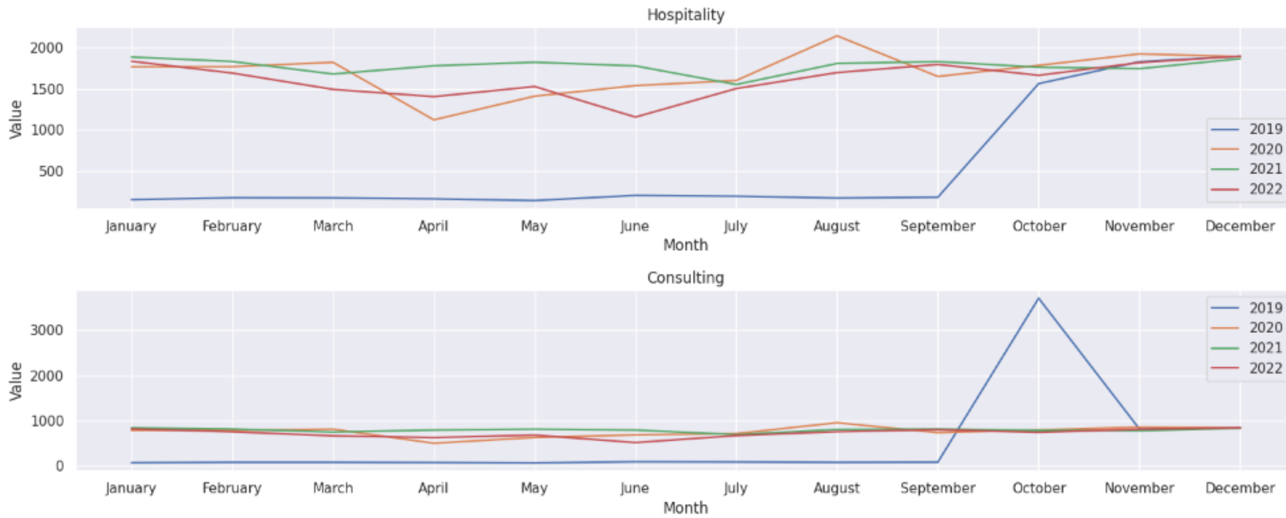
```
# Columns to plot (excluding the index column)
columns_to_plot = dataEGBA.columns[0:]

# Set up colors for the lines
colors = ['blue', 'green', 'red', 'purple', 'orange', 'cyan']

# Iterate through each column and create a separate line chart
for idx, col in enumerate(columns_to_plot):
    plt.figure(figsize=(12, 6))  # Adjust the figure size as needed
    plt.plot(dataEGBA.index, dataEGBA[col], color=colors[idx])  # Use the index for x-axis
    plt.xlabel('Date')
    plt.ylabel('Consumption (mWh)')
    plt.title(f'Consumption Trend: {col}')
    plt.xticks(rotation=45)  # Rotate x-axis labels for better visibility
    plt.tight_layout()

    plt.show()
```





Other visualizations can be found in the code file.

I also was interested in how all the business units grew over the years hence i plotted a multivariate chart.



To perform time series forecasting effectively, it is crucial to ensure that the dataset meets the requirement of stationarity. Stationarity implies that the statistical properties of the data remain consistent over time. Specifically, for a dataset to be considered stationary, the following conditions must hold:

1. **Constant Mean:** The mean (average) of the data remains constant over time, indicating that there is no significant upward or downward trend in the series.
2. **Constant Variance:** The variance (spread or dispersion) of the data remains constant over time, indicating that the data points are not becoming more or less spread out as time progresses.
3. **Constant Covariance:** The covariance between data points at the same time lag remains constant. In other words, the relationship between data points at different time periods is consistent. This implies that the series does not exhibit seasonality or cyclical patterns.

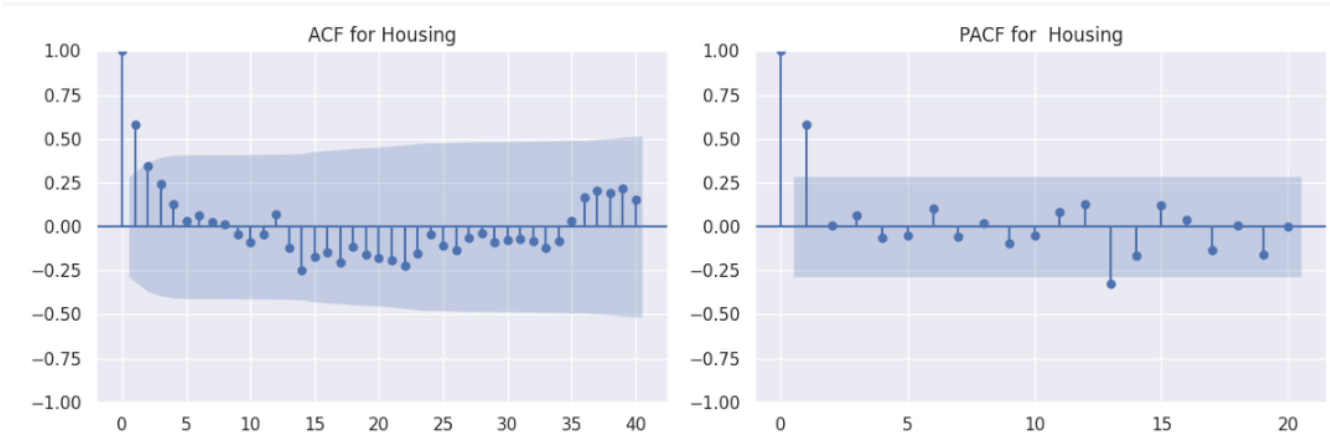
As a result, I undertook an assessment of stationarity. To achieve this, I employed the Augmented Dickey-Fuller (ADF) algorithm, a diagnostic tool used to ascertain the stationarity of a dataset. The ADF algorithm accomplishes this by evaluating whether the three aforementioned conditions hold true. According to the ADF test, a dataset is considered stationary when its p-value is less than 0.05. Upon applying this test to the Egba Business Unit data, the results indicated that three datasets were non-stationary, while three others exhibited stationarity.

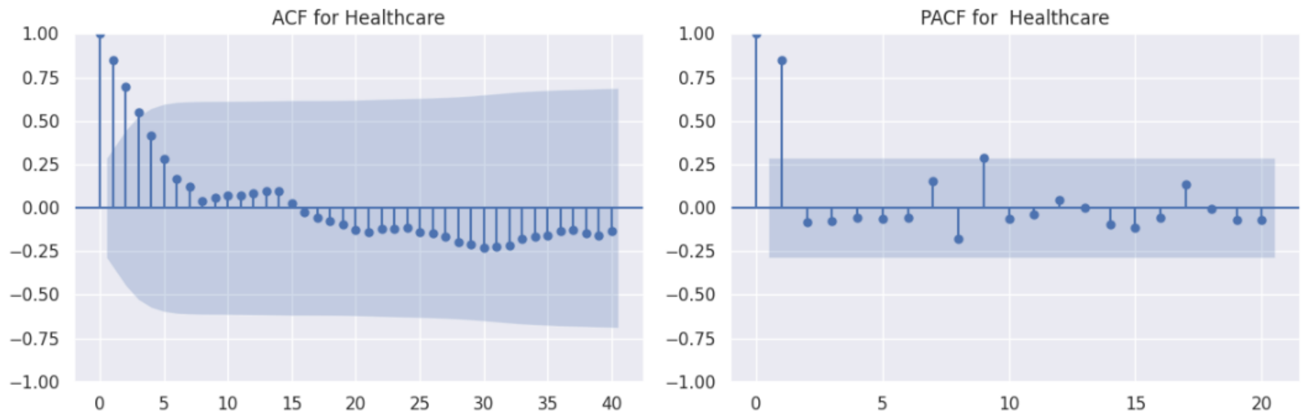
```
Column: Housing
ADF Statistic: -3.337467274382218
p-value: 0.013284529656477435
Critical Values:
        1%: -3.578
        5%: -2.925
        10%: -2.601
Reject Ho - Time Series is Stationary
=================================================
Column: Healthcare
ADF Statistic: -3.600762307749795
p-value: 0.005746470775813004
Critical Values:
        1%: -3.601
        5%: -2.935
        10%: -2.606
Reject Ho - Time Series is Stationary
=================================================
Column: Education
ADF Statistic: -2.515850837179189
p-value: 0.11166512426953107
Critical Values:
        1%: -3.578
        5%: -2.925
        10%: -2.601
Failed to Reject Ho - Time Series is Non-Stationary
=================================================
```

```
Column: Manufacturing/Production
ADF Statistic: -3.6007573785378617
p-value: 0.005746564796403063
Critical Values:
        1%: -3.601
        5%: -2.935
        10%: -2.606
Reject Ho - Time Series is Stationary
=================================================
Column: Hospitality
ADF Statistic: -2.175985821240822
p-value: 0.21507636366676747
Critical Values:
        1%: -3.578
        5%: -2.925
        10%: -2.601
Failed to Reject Ho - Time Series is Non-Stationary
=================================================
Column: Consulting
ADF Statistic: -2.126005698006376
p-value: 0.23418695348620017
Critical Values:
        1%: -3.621
        5%: -2.944
        10%: -2.610
Failed to Reject Ho - Time Series is Non-Stationary
=================================================
```

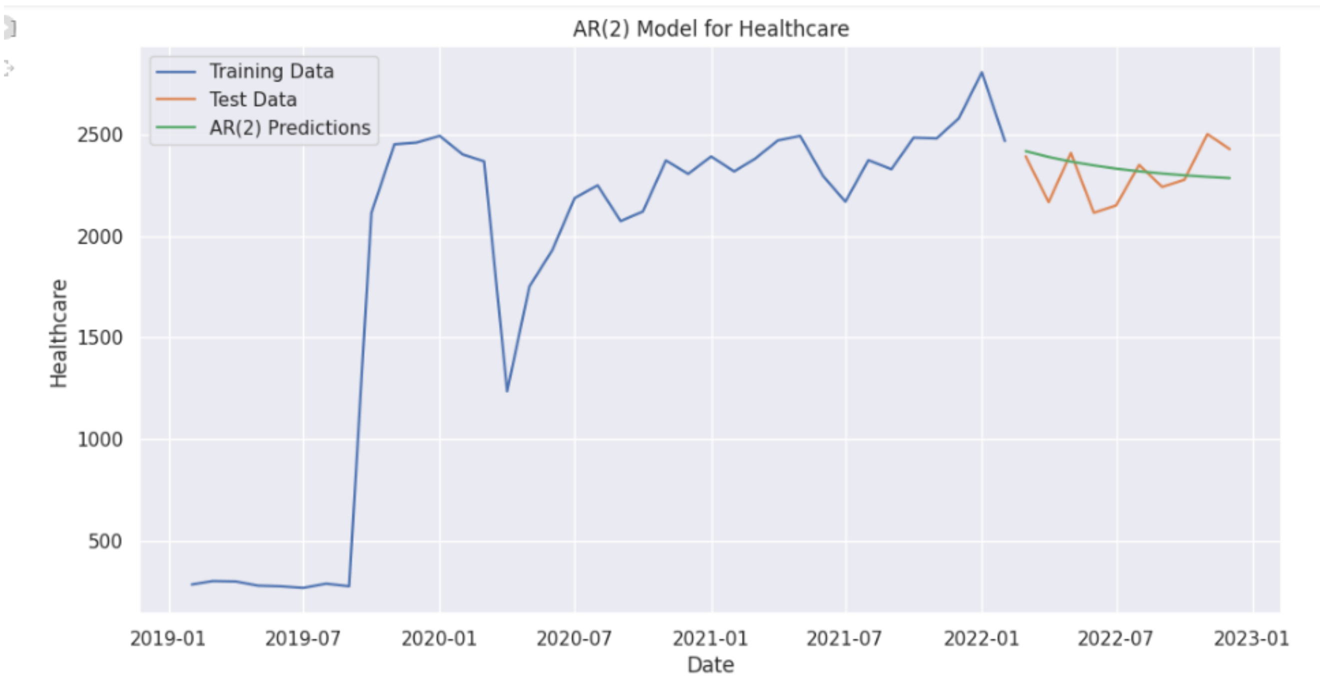
Subsequently, I converted the non-stationary data into a stationary form, motivated by the necessity to adhere to the prerequisites for conducting time series forecasts, which include achieving stationarity. Employing the differencing technique, I successfully transformed the data into a stationary state. Differencing, as a method for addressing non-stationary data, involves the creation of a new series derived from the differences between each data point and its preceding data point. Through the application of this differencing procedure, performed twice in this instance, all data points within the time series were ultimately rendered stationary.

I proceeded to examine the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These functions consist of covariance values at various time lags and, when graphically depicted, offer insights into selecting an appropriate model for the data and determining its order. In the case of the Egba Business Unit data, the visual representations indicated that it was unsuitable for AutoIntegrated Moving Average (ARIMA) and Seasonal AutoIntegrated Moving Average (SARIMA) algorithms. Although the initial intention was to explore these forecasting methods, the visualizations clearly revealed patterns indicative of Autoregressive (AR) and Exponential Smoothing models. Consequently, the research shifted its focus to the Autoregressive Model.





The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) consistently displayed a gradual decline starting from lag points 1 and 2. This pattern suggests that the most suitable AutoRegressive (AR) model is AR(2), indicating an AutoRegressive Model of order two. Subsequently, I proceeded to divide the dataset into training and testing subsets in chronological order, maintaining an 80:20 ratio, respectively. The training data was employed to train the AutoRegressive model, and the resulting prediction errors or residuals were collected. To assess the model's performance, I visualized the predicted values and compared their proximity to the actual values using the test dataset.



Upon concluding the training phase, I applied the trained model to perform a one-year forecast. The results of this forecasting process will be detailed in the Results section. The same procedure was iteratively carried out for the remaining five business units, and their respective outcomes will also be presented and discussed in the Results section.

4.2 Evaluation

The evaluation metrics employed in this analysis included the Mean Square Error (MSE), Mean Absolute Error (MAE), and the R-Square (R^2) value. These metrics are commonly used for assessing the performance of regression models, including the AutoRegressive (AR) model. In all cases, lower scores indicate superior model performance, making them valuable indicators for evaluating the accuracy and effectiveness of the forecasts.

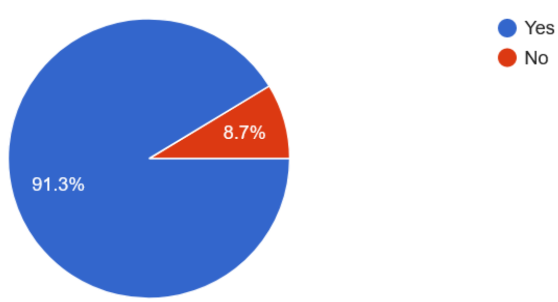
```
"""
                            AutoReg Model Results
==============================================================================
Dep. Variable:     Manufacturing/Production   No. Observations:           37
Model:                           AutoReg(2)   Log Likelihood        -278.394
Method:                     Conditional MLE   S.D. of innovations    688.947
Date:                      Wed, 06 Sep 2023   AIC                    564.787
Time:                              15:15:21   BIC                    571.009
Sample:                          04-01-2019   HQIC                   566.935
                               - 02-01-2022
==================================================================================================
                                coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                       668.4471    286.777      2.331      0.020     106.375    1230.519
Manufacturing/Production.L1   0.8863      0.169      5.251      0.000       0.555       1.217
Manufacturing/Production.L2  -0.0461      0.164     -0.281      0.778      -0.367       0.275
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.2035          +0.0000j            1.2035            0.0000
AR.2           18.0289          +0.0000j           18.0289            0.0000
-----------------------------------------------------------------------------


""", 'Mean Squared Error (MSE)': 27671.453829574777, 'Mean Absolute Error (MAE)': 138.1683664523933, '
"""

'R-squared (R2)': -0.1116862317797708 8, 'Mean Absolute Percentage Error (MAPE)': 95.9673949692792
```

Pre-Assessment Questionnaire/Interview

This section aims to gather insights into participants' understanding of electricity patterns in their respective areas, with a specific focus on residents within the chosen business unit. The following questions were posed to participants:

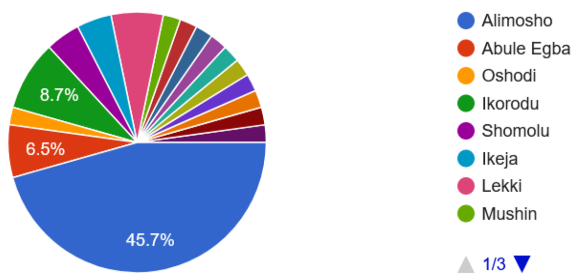
Do you live in Lagos, Nigeria.

46 responses



Do you reside in one of these local government?

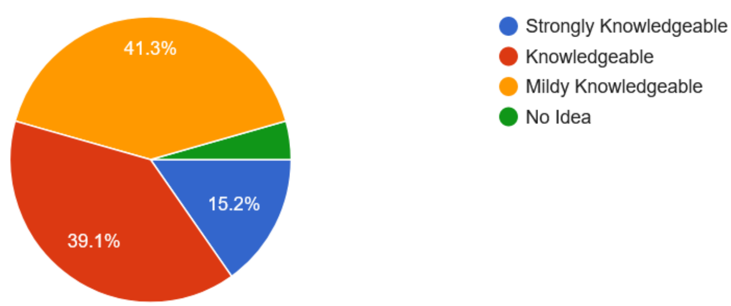
46 responses



This question was asked to ensure they fulfill the condition of being residents in our areas of interest. 65% of the population are residents of Alimosho

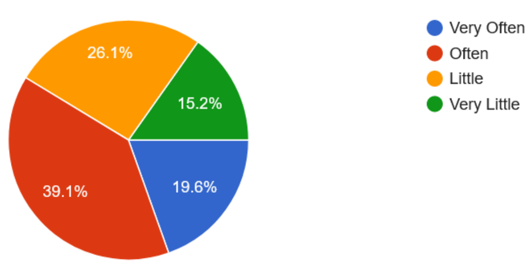
How knowledgeable are you about the electricity information of your residential area?

46 responses



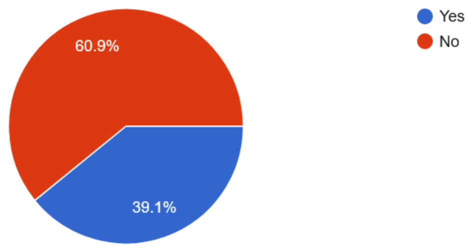
How often do you get access to electricity

46 responses



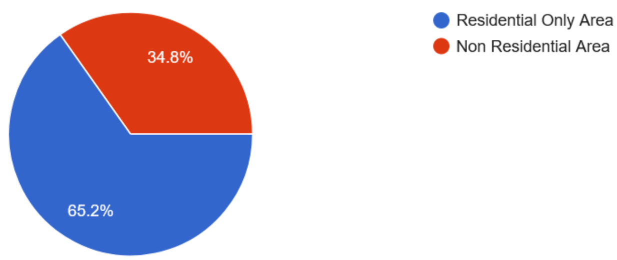
Do you feel that you are getting your money's worth?

46 responses



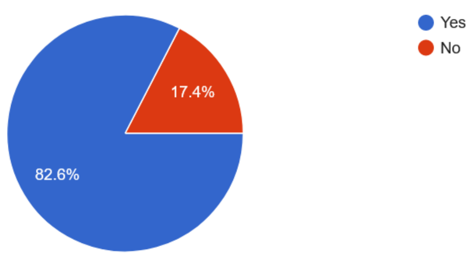
Do you in a Residential Only area or do Non-Residential players exist in your area eg (schools, hospitals, hotels, companies etc.)

46 responses



Are there times in the year where you consume more electricity or pay more for electricity?

46 responses



In summary, the outcomes derived from the questionnaire revealed a notable degree of dissatisfaction among users concerning the quality of service they had received.

5. CONCLUSION

In summary, data visualization proves to be a powerful means of imparting knowledge and insights about intricate datasets. The survey participants have gained a deeper understanding of consumption patterns in their respective areas through this approach. Additionally, time series data serves as a valuable tool for forecasting, although it presents challenges when it comes to model selection and interpreting the significance of ACF and PACF results.

Further research is warranted to develop methods that facilitate the identification of the most suitable models and their appropriate orders. Such advancements would make it more accessible for individuals without technical expertise to harness time series analysis in their everyday decision-making processes.

REFERENCE

1. *Ogbogu, U., & Ahmed, N. (2022). Ethical, Legal, and Social Implications (ELSI) Research: Methods and Approaches. Current Protocols, 2(1), e354. https://doi.org/10.1002/cpz1.354*
2. *Yip, C., Han, L. R., & Sng, B. L. (2016). Legal and ethical issues in research. Indian Journal of Anaesthesia, 60(9), 684-688. https://doi.org/10.4103/0019-5049.190627*
3. *Ramchandra P, Boucar D: Green Energy and Technology. Springer, London Dordrecht Heidelberg New York; 2011*
4. *Eberhard, K. The effects of visualization on judgment and decision-making: a systematic literature review. Manag Rev Q 73, 167–214 (2023). https://doi.org/10.1007/s11301-021-00235-*
5. *Melanie R. Herrmann, Duncan P. Brumby, Tadj Oreszczyn & Xavier M. P.Gilbert (2018) Does data visualization affect users' understanding of electricity consumption?, Building Research & Information, 46:3, 238-250, DOI: 10.1080/09613218.2017.1356164*
6. *Kernbach, Sebastian & Eppler, Martin. (2010). The Use of Visualization in the Communication of Business Strategies: An Experimental Evaluation. http://www.alexandria.unisg.ch/Publikationen/62537. 52. 10.1109/IV.2010.55.*
7. Zou, G. (2022). The relationships between energy consumption and key industrial sector growth in China. *Energy Reports*, *8*, 924-935. https://doi.org/10.1016/j.egyr.2022.07.112
8. https://www.icirnigeria.org/57-of-electricity-consumers-overcharged-with-estimated-billing/
9. *https://punchng.com/unmetered-power-consumers-drop-to-7-74-million/*
10. *Kontokosta, Constantine & Tull, Christopher & Marulli, David & Pingerra, Renate & Yaqub, Maha. (2015). Web-Based Visualization and Prediction of Urban Energy Use from Building Benchmarking Data.*
11. *Goswami, Kakoli & Kandali, Aditya. (2021). Forecasting Electricity Demand Using Statistical Technique for the State of Assam. 10.1007/978-981-16-1119-3_27.*
12. *Nti, I. K., Teimeh, M., & Adekoya, A. F. (2020). Electricity load forecasting: A systematic review. Journal of Electrical Systems and Information Technology, 7(1), 1-19. https://doi.org/10.1186/s43067-020-00021-8*
13. *Nti IK, Asafo-Adjei S, Agyemang M (2019) Predicting monthly electricity demand using soft-computing technique. Int Res J Eng Technol 06:1967–1973*
14. *Hammad MA, Jereb B, Rosi B, Dragan D (2020) Methods and models for electric load forecasting: a comprehensive review. Logist Sustain Transp 11:51–76. https://doi.org/10.2478/jlst-2020-0004*
15. *Zakarya S, Abbas H, Belal M (2017) Long-term deep learning load forecasting based on social and economic factors in the Kuwait region. J Theor Appl Inf Technol 95:1524–1535*
16. *Sulandari W, Subanar MH, Lee PC (2020) Rodrigues, Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. Energy 190:116408. https://doi.org/10.1016/j.energy.2019.116408*
17. *Nepal B, Yamaha M, Yokoe A, Yamaji T (2020) Electricity load forecasting using clustering and ARIMA model for energy management in buildings. Jpn Archit Rev 3:62–76. https://doi.org/10.1002/2475-8876.12135*
18. *FAgboola, Olasunkanmi & Esobinenwu, Chizindu. (2023). Long Term Load Forecast Using Artificial Neural Network Method: Rainbow-Elekahia Commercial 33kV Feeder in Port Harcourt, Nigeria. 7. 2965-3325.*

19. *Lukong, Terence & Nganyu Tanyu, Derick & Tamo Tatietse, Thomas & Schulz, Detlef. (2022). Long Term Electricity Load Forecast Based on Machine Learning for Cameroon's Power System. Energy and Environment Research. 12. 45. 10.5539/eer.v12n1p45.*
20. *Bharati, Subrato, Podder, Prajoy, and Mondal, M. Rubaiyat Hossain. 'Visualization and Prediction of Energy Consumption in Smart Homes'. 1 Jan. 2020 : 81 – 97. (https://content.iospress.com/articles/international-journal-of-hybrid-intelligent-systems/his200283)*
21. *Zeng, A., Liu, S., & Yu, Y. (2019). Comparative study of data-driven methods in building electricity use prediction. Energy and Buildings, 194, 289-300. https://doi.org/10.1016/j.enbuild.2019.04.029*
22. *Journal, I. J. C. S. M. C., & Amer, A. (2019). Tableau Big Data Visualization Tool in the Higher Education Institutions for Sustainable Development Goals. IJCSMC, 8(7), 71–78.*
23. *Lee, Y.W. & Gaik, Tay & Yaan Yee, Choy. (2018). Forecasting Electricity Consumption Using Time Series Model. International Journal of Engineering and Technology(UAE). 7. 218-223. 10.14419/ijet.v7i4.30.22124.*
24. *https://pandas.pydata.org/docs/#:~:text=pandas%20is%20an%20open%20source,for%20the%20Python%20programming%20language.&text=New%20to%20pandas%3F,and%20links%20to%20additional%20tutorials.*
25. *https://www.tensorflow.org/api_docs/python/tf/keras/Model*
26. *https://numpy.org/doc/stable/user/absolute_beginners.html*
27. *https://www.statsmodels.org/stable/index.html*
28. *https://en.wikipedia.org/wiki/Feeder_line_(network)#:~:text=Feeder%20transmits%20power%20from%20Generating,well%20as%20the%20receiving%20end.*
29. *Cooksey, R. W. Descriptive Statistics for Summarising Data. Illustrating Statistical Procedures: Finding Meaning in Quantitative Data, 61-139. https://doi.org/10.1007/978-981-15-2537-7_5*