

**Important information from University of Roehampton**

Feedback provided by the Studiosity service is designed to help you to improve your academic writing skills by indicating areas for improvement related to structure, writing style, language use, spelling and grammar.

This service does not assess the quality of your response to the assessment task and how this meets the assessment criteria and therefore any feedback received here does not provide any indication of your final mark/ grade - this can only be determined by your module tutor.

Studiosity can only offer general referencing guidance. For course-specific referencing advice you should always go to the University of Roehampton Referencing Guides at <https://library.roehampton.ac.uk/referencing>

If you would like any further advice visit the Learning Skills Hub on Moodle for resources and information about drop-ins, workshops and other academic skills events.

**File:** Chioma\_Pere\_Project\_Work\_\_1\_.docx

**Student:** Chioma Lois Pere-Afun - PER21526408pereafuc@roehampton.ac.uk

**Word count:** 5810

**Assessment task:** Case study

**Feedback areas:**

- Structure
- Choice of language
- Use Of Sources

**State of document:** Nearly there

**Submission ID:** a8738522-7b83-4b07-a7b7-6ac52e8316af

## Glossary

[Click here to see a full glossary of English writing terms and their definitions.](#)

# Chioma\_Pere\_Project\_Work\_\_1\_.docx

Please note that the specialist has not proofread your work but has highlighted examples of the types of errors you need to look out for. Please review your entire document with these examples in mind before you submit your assignment.



**Meagan W**

Joined in 2016

 ON, Canada

## General Comments

Greetings! Thank you for submitting your paper to Studiosity for review. It looks like you have put a lot of time and effort into your paper which is great to see. However, there are some improvements that can be made to your work moving forward.

Regarding your structure, try to avoid the use of paragraphs that are only one or two sentences long. When editing your language, focus on crafting more concise writing by avoiding issues like unnecessary language and repetition. Additionally, you should ensure you have included an in-text citation wherever you have referenced information from another source. I will talk about these issues more under the relevant headings.

## Structure

I was glad to see that you broke your work down into specific sections that were labelled with clear headings. You also made sure to clearly outline your aims and objectives at the beginning of this piece. However, there are some structural improvements that can be made to your work moving forward. For instance, you should pay close attention to your paragraph format and ensure each paragraph is an appropriate length (i.e., longer than a sentence or two) and contains a clear linking/concluding sentence. Additionally, I noticed that your paper is lacking a conclusion section. You should check your assignment guidelines to determine if a conclusion needs to be included in your paper.

### Avoiding single-sentence or very short paragraphs

Generally, in order to form a complete paragraph, you need to use several sentences. A full paragraph must include a topic sentence, elaborating details, examples and analysis, and a concluding or transition sentence at the end; one or two sentences are rarely able to make a complete paragraph. When you find yourself with a very short paragraph, try either integrating it into a surrounding paragraph where you are discussing similar topics or adding more information to it to form a complete paragraph of its own.

### Using linking sentences

A paragraph should have a concluding sentence that links back and reinforces the main idea of the paragraph. Sometimes, it may also introduce the idea of the next paragraph. For example, a paragraph with the topic sentence 'Genetic engineering can be very beneficial in some areas' may have the concluding sentence 'Therefore, the application of genetic engineering has benefitted not only science but also medicine and agriculture'.

For more information, watch this: <https://vimeo.com/825655718/4b59f8cbc9?share=copy>.

### Formatting headings consistently

It appears that you have included some of your headings by using a different format. The format of your main headings should be consistent throughout your document. For example, if you are going to use bold for some of the headings, you should use bold for all of the headings on that same level.

### Structuring a conclusion

The conclusion of a piece of writing is usually one paragraph that draws all the ideas together. In general, the first sentence consists of a signal phrase that restates or readdresses the thesis or focus question. The next few sentences should summarise the points in the body paragraphs. The concluding sentence should identify the main idea of your paper or next steps, or both.

For more information, watch this: <https://vimeo.com/725501927/42daffca8c>

## Choice of language

---

Reading over your work, I found that you included appropriate and relevant language to your discussion topic throughout your writing; this was good to see. However, there are some language issues in your writing that are impacting your work, such as repetition, contractions, unnecessary language, and informal language. These issues are elaborated on below so that you can apply any necessary revisions to your writing.

### Using personal pronouns

Our system found some personal pronouns such as I, you, or we in your writing. While this is widely accepted, it's a good idea to ensure that this is permitted in this particular task. It is possible that the use of personal pronouns is outlined in your task description or references, in which case please disregard this comment.

### Avoiding repetition

Be mindful of using similar language too frequently. In a highlighted instance of your text, you have included the repeated words quite close together. An alternative wording choice may be required to minimise reader confusion and ensure a more natural flow.

### Using contractions

Academic texts are generally read as more formal without the use of contractions. For example, 'Communication strategies should not be ignored' is stronger than 'Communication strategies shouldn't be ignored'. Be sure to keep your contraction usage minimal throughout your academic writing.

### Using casual or informal language

To improve the overall quality of your writing, you will need to use more sophisticated and academic wording. For instance, rather than writing 'give', try using a word such as 'provide'. These simple changes can have a very positive impact on the quality of your writing.

For more information, watch this: <https://vimeo.com/725502858/b3f32856c5>

### Avoiding unnecessary language

It is usually best to keep your wording and sentence structure simple and concise in order to make sure your reader can easily understand your ideas, being mindful that language must be appropriate for the context and audience. By using precise words and eliminating wordiness and unnecessary repetition, your writing will be clear, concise, and easy to understand. For example, 'The investment had some costly results' is much more concise than 'The investment had some bad results that cost the firm a lot of money'. To eliminate wordiness, try to substitute long phrases with precise words.

## Spelling and grammar

---

## Use of sources

---

It is good to see that you included some references in your paper to help support your work. However, our citation assist picked up many areas throughout your writing that did not include citations. Please review your writing carefully and ensure you include an in-text citation wherever you have referenced information from another source.

Our Citation Assist feature has detected a text match with an uncited source in your document. To avoid unintentional plagiarism, it is important to reference any material that comes from another source by using in-text citations, footnotes or endnotes (depending on your style guide), along with a reference list at the end that includes all relevant information about each source. Please make sure that you follow the instructions outlined for the referencing style chosen by your education provider.

### Citing all sources

When using information from sources, it is important that you cite the source every time and use the appropriate format specified by your referencing style.

For more information, watch this: <https://vimeo.com/725502018/b948d2fcb3>

### Avoiding inconsistencies in reference list

It is important to make sure that there are no inconsistencies in the formatting of your reference list. You should aim to check each reference against your institution's referencing style guide for accuracy before you submit your assignment for marking.

---

### **Final comments**

---

Good luck as you continue working on your paper!

---

### **Roehampton resources**

When producing academic assignments, you will need to ensure you are acknowledging the work of others by correctly referencing what you've researched. The Library has provided access to online referencing guidance for you to not only learn how to reference but to apply the correct referencing conventions to your work in your discipline.



<https://library.roehampton.ac.uk/referencing/introduction>

---

[View Reviewed Document](#)

**Chioma\_Pere\_Project\_Work\_\_1\_.docx**

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

Submitted to the  
**University of Roehampton**

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

[TABLE OF CONTENT] [It is good to see that you have included a clear table of contents here at the beginning of your paper so the reader can easily navigate through your work. ]

INTRODUCTION 3

Research Questions or Problems that will be Addressed 4

Aims 4

Objectives 5

Legal, Social, Ethical, and Professional Considerations 5

LEGAL CONSIDERATIONS 5

SOCIAL CONSIDERATIONS 5

ETHICAL CONSIDERATIONS 6

Report Overview 6

LITERATURE REVIEW 7

Data Visualization 7

Time Series Forecast 8

Tools 9

METHODOLOGY OR DESIGN 11

Building The Visualization 11

Predicting The Energy Consumption And Impacting on Housing, 11

Evaluation Plans, 11

REFERENCE 12

## 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. [C1, 10 matches found, It appears that you have used a source here, but you have not cited it. Make sure to clearly reference any information taken from other sources in your paper using in-text citations. ] 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 [C2, 1 match found] 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 [You have repeated this word close together in the sentence. Consider an alternative word choice (e.g., "investigating", "analysing", etc.) or a concise revision of your phrasing to avoid repetition.] 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 [C3, 1 match found] [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 [To maintain an academic tone, avoid using contractions. Instead, write this as "do not". ] 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. [C4, 7 matches found] 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 [Some institutions may prefer it if you avoid the use of personal pronouns in academic writing.] 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 organization or bodies in Lagos Nigeria, identifying lapses, areas for growth, and opportunities, as well as making consumption forecasts for the next two years.

## 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?

## 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 [2] [Typically, numbers under 10 should be written out as words. However, refer to your style guide on this.] years.

## 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,] [C5, 3 matches found, It looks like you have also used a source here but not cited it. Please see the Feedback Summary for more information about this.] 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 as well as machine learning models, Support Vector Machine, and Artificial Neural Networks, 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.

## 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.

### 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.] [This is another example of where you may have used a source without citing it. Check your work carefully for any other possible cases of this.] 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.

### 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 the society at large.

### 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.] [C6, 9 matches found] 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.] [Try to avoid the use of paragraphs that are only one or two sentences long. Please refer to the feedback under Structure.]

## 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.





## 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. [C7, 2 matches found] 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.

### 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 comprehensible. 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 visualization will not aid comprehension and retention. The choice of visualization also matters. [In research by] [Try to avoid the use of unnecessary words because it leads to wordiness and poor readability. For instance, this highlighted text is not necessary here and could be removed to express your point in a more clear and direct manner.] 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.

### 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 \(ANN\)](#) **[You have already identified this acronym earlier in your work. Once you have done this, you should carry on only using the abbreviated term (in this case, 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).

## 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.](#) **[C8, 1 match found]** 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](#) **[Your tone is too informal here. Try to choose words that are more formal (e.g., "be useful").]** 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.](#) **[C9, 11 matches found]** 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\[24\].](#) **[C10, 7 matches found]** 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](#) **[C11, 1 match found]** modeling. This study will be making use of the regression model, Support vector machine hence [it's](#) **[This is another contraction, which makes your tone less academic. Please see my notes in the Feedback Summary section for more help with this issue.]** crucial to the study.

[Tensorflow: TensorFlow is an open-source machine learning framework developed by](#) **[C12, 5 matches found]** 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](#) **[C13, 2 matches found]** 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].

## METHODOLOGY OR DESIGN

### 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.

### 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. [C14, 1 match found] Examples of multivariate charts include scatterplots, heatmaps, and parallel coordinate plots. [This paragraph requires a linking or concluding sentence to link back to the main idea of the paragraph or lead into the following paragraph.]

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. [This is another example of a very short paragraph.]

Forecasting. [The formatting of your headings should be consistent throughout your document. For instance, in previous headings on this level you did not place a full stop at the end like you have here. Ensure you are following a consistent format. ]

#### 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. [C15, 5 matches found] It is a combination of three components: autoregression (AR), differencing (I), and moving average (MA). [C16, 1 match found] 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. [C17, 1 match found] 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:

$$y(t)=c+a_1y_{t-1}+...+a_p y_{t-p}+e(t)$$

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

In these equations,  $a_1, \dots, a_p$ , and  $m_1, \dots, m_q$  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  $\mu$  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:

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

### Support Vector Machine

SVM is a machine learning algorithm that can be used for regression and classification tasks. [C18, 2 matches found] 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 would 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. [C19, 2 matches found] 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. [C20, 3 matches found] Activation functions add nonlinearity to the model. Activation functions are mathematical functions that are applied to the output of each neuron in the network. [C21, 3 matches found] Activation functions introduce nonlinearity into the network, allowing it to model complex relationships [C22, 1 match found] between input and output variables. Common activation functions include the sigmoid function, the hyperbolic tangent (tanh) function, and the rectified linear unit (ReLU) function. [C23, 5 matches found] 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 Sigmoid 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.

### Evaluation Plans.

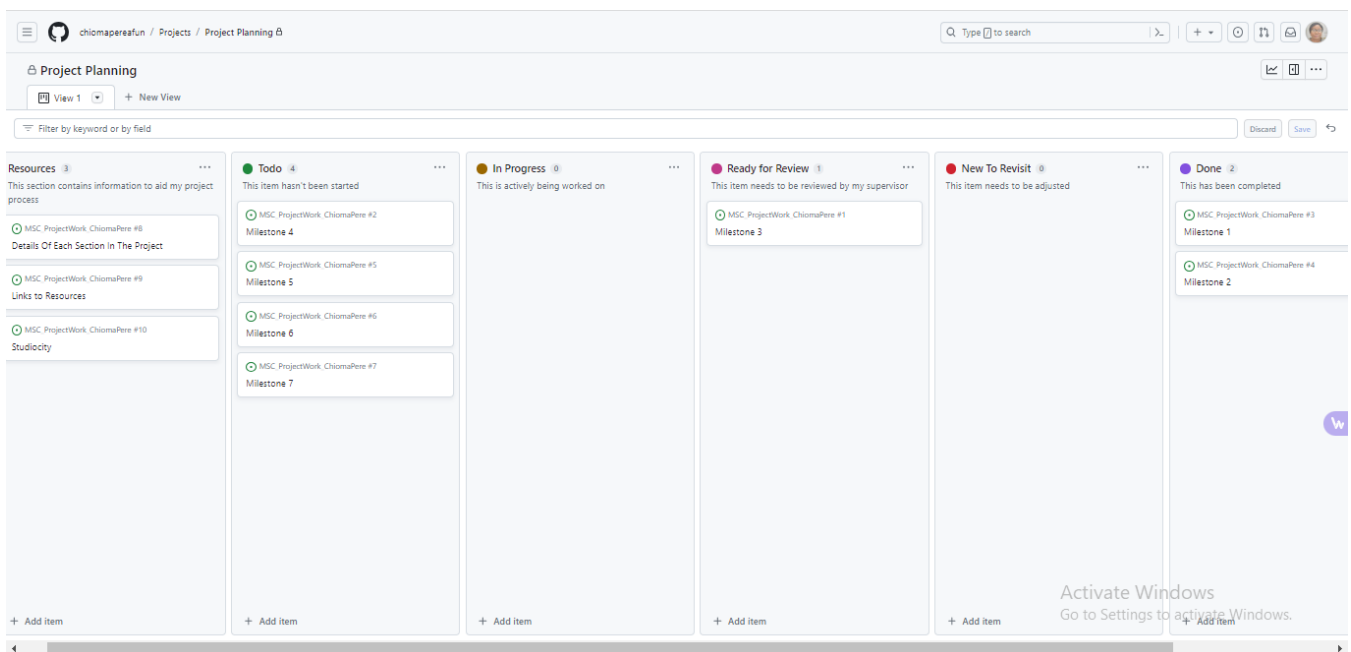
The forecast models will be evaluated using evaluation metrics such as MAPE, MAE, and RMSE. [Your paper is lacking a conclusion section. See the feedback summary for more information on conclusions to determine if one should be included in your paper.]

IMPLEMENTATION OF RESULT.

## 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. [\[Ramchandra P, Boucar D\]](#)*[You will need to double-check your reference list because it contains inconsistencies that need to be reviewed. For instance, in previous citations you included an ampersand between the last two authors' names, but you have not done so here. Be sure to use the referencing guide provided by your institution.]* 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/s11301-021-00235-8>
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/ijst-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](https://www.tensorflow.org/api_docs/python/tf/keras/Model)
26. [https://numpy.org/doc/stable/user/absolute\\_beginners.html](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.](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](https://doi.org/10.1007/978-981-15-2537-7_5)



Your draft has been checked for text that matches existing public sources.

Finished reviewing your submission?

## Appendix: Citation Assist

- C1. *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.*

10 matches found

<https://wseas.org/wseas/cms.action?id=6938>  
<https://mwnation.com/malawi-look-wind-energy-solution/>  
<https://www.goldstandard.org/blog-item/2021-energy-risk-and-governance-summit>  
<https://www.projectpapers.net/design-and-fabrication-of-an-egg-incubator-2/>  
<https://conrema.org/2018/08/13/should-malawi-look-to-wind-for-energy-solution/>  
<https://energysustainoc.biomedcentral.com/articles/10.1186/2192-0567-2-15>  
[https://www.linkedin.com/posts/jairus-gayus-801163244\\_making-a-differences-for-the-planet-activity-7048723536038760449-bitZ](https://www.linkedin.com/posts/jairus-gayus-801163244_making-a-differences-for-the-planet-activity-7048723536038760449-bitZ)  
<https://businessday.ng/analysis/article/nigeria-not-ripe-for-renewable-energy/>  
<https://academy.sustain-cert.com/event/2021-energy-risk-and-governance-summit/>  
<https://casade.org/academic-papers/energy-and-sustainable-development-in-nigeria-the-way-forward/>

- C2. *Data visualization has been shown to be an effective tool*

1 match found

<https://visually/community/Infographics/technology/future-challenges-computer-science>

- C3. *electricity bills by distribution companies' estimation or direct billing method*

1 match found

<https://www.icirnigeria.org/57-of-electricity-consumers-overcharged-with-estimated-billing/#~:text=Installing%20prepaid%20meters.&text=NIGERIAN%20electricity%20consumers%20are%20yet,consumers%20are%20still%20being%20overcharged.>

- C4. *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.*

7 matches found

<https://von.gov.ng/nerc-reports-marginal-decrease-in-unmetered-electricity-users/>  
<https://www.rotana-hotel.com/unmetered-power-consumers-drop-to-7-74-million/>  
<https://punchng.com/unmetered-power-consumers-drop-to-7-74-million/>  
<https://ntm.ng/2023/01/15/unmetered-power-consumers-drop-to-7-74-million/>  
<https://oraclenews.ng/number-of-unmetered-electricity-consumers-in-nigeria-drops-nerc/>  
<https://ntm.ng/2023/01/15/unmetered-power-consumers-drop-to-7-74-million/amp/>  
<https://www.thetidenewsonline.com/2023/01/16/unmetered-power-consumers-drop-to-7-74m/>

- C5. *Tableau, an interactive data visualization software company focused on business intelligence,*

3 matches found

<https://business.cornell.edu/hub/2021/11/24/hospitality-hackathon-promotes-interdisciplinary-data-driven-solutions/>  
<https://www.hospitalitynet.org/news/4108100.html>  
<https://news.cornell.edu/stories/2021/12/hospitality-hackathon-promotes-interdisciplinary-data-driven-solutions>

- C6. *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.*

9 matches found

[https://advance.sagepub.com/articles/preprint/Ethical\\_Considerations\\_For\\_Discrete\\_Choice\\_Experiments\\_with\\_Caregiving\\_Populations/19125350](https://advance.sagepub.com/articles/preprint/Ethical_Considerations_For_Discrete_Choice_Experiments_with_Caregiving_Populations/19125350)  
[https://www.theseus.fi/bitstream/handle/10024/784876/Truong\\_Pham\\_2022.12.02.pdf?sequence=2](https://www.theseus.fi/bitstream/handle/10024/784876/Truong_Pham_2022.12.02.pdf?sequence=2)  
<https://testbook.com/question-answer/the-ethical-considerations-in-research-include--627ef499fdb6c646393ede84>  
<https://www.scribbr.com/methodology/research-ethics/>  
<https://www.scribbr.com/frequently-asked-questions/what-are-ethical-considerations-in-research/>  
[https://www.theseus.fi/bitstream/handle/10024/793757/Kallunki\\_Erika.pdf?sequence=2&isAllowed=y](https://www.theseus.fi/bitstream/handle/10024/793757/Kallunki_Erika.pdf?sequence=2&isAllowed=y)  
<https://www.bda.uk.com/practice-and-education/research-in-practice/conducting-research.html>  
[http://www.santetropicale.com/ostelec/uk/research\\_ethics.asp?inscription=ok](http://www.santetropicale.com/ostelec/uk/research_ethics.asp?inscription=ok)  
<https://testbook.com/question-answer/in-any-research-involving-human-participants-main--6176a626d2713841fcc78048>

- C7. *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.*

2 matches found

<http://www.diva-portal.org/smash/get/diva2:1480413/FULLTEXT01.pdf>  
<https://repository.bilkent.edu.tr/server/api/core/bitstreams/1af403dc-72d9-44c5-9af3-3125b9c2f436/content>

- C8. *it easier for people to explore and manage data, and faster to discover and share insights that can change businesses and the world.*

1 match found

<https://library.geneseo.edu/c.php?g=1302911&p=9668352>

- C9. *Microsoft Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence.*

11 matches found



[https://en.wikipedia.org/wiki/Microsoft\\_Power\\_BI](https://en.wikipedia.org/wiki/Microsoft_Power_BI)  
<https://medium.com/@gurmolsfly12/coding-for-data-scientists-329e47cb4b24>  
[https://www.wikiwand.com/en/Microsoft\\_Power\\_BI](https://www.wikiwand.com/en/Microsoft_Power_BI)  
[https://wiki2.org/en/Power\\_BI](https://wiki2.org/en/Power_BI)  
[https://www.linkedin.com/posts/sanika-mahimkar\\_data-software-power-activity-7066740340069666816-BcjT](https://www.linkedin.com/posts/sanika-mahimkar_data-software-power-activity-7066740340069666816-BcjT)  
[https://dbpedia.org/page/Microsoft\\_Power\\_BI](https://dbpedia.org/page/Microsoft_Power_BI)  
[https://www.linkedin.com/posts/dishant-salunke-b67a58136\\_data-powerbi-work-activity-7071822533703966720-ezjd](https://www.linkedin.com/posts/dishant-salunke-b67a58136_data-powerbi-work-activity-7071822533703966720-ezjd)  
<https://texadasoftware.com/power-bi/>  
[https://www.linkedin.com/posts/kiran-sagar-a00a761b2\\_powerbi-datavisualization-pythonlife-activity-7070608425105178624-32Hj](https://www.linkedin.com/posts/kiran-sagar-a00a761b2_powerbi-datavisualization-pythonlife-activity-7070608425105178624-32Hj)  
<https://www.linkedin.com/pulse/microsoft-power-bi-beginner-haridharan-m>  
<https://library.geneseo.edu/c.php?g=1302911&p=9668352>

C10. *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].

7 matches found

<https://pandas.pydata.org/pandas-docs/version/2.0.0/>  
<https://pandas.pydata.org/pandas-docs/version/2.0.2/>  
<https://pandas.pydata.org/docs/>  
<https://pandas.pydata.org/pandas-docs/stable/index.html?highlight=str%20split>  
<https://pandas.pydata.org/pandas-docs/version/0.22/overview.html>  
<https://pandas.pydata.org/pandas-docs/version/0.23/overview.html>  
<https://pandas.pydata.org/pandas-docs/version/0.21/overview.html>

C11. *provides a wide range of tools for various machine learning tasks, including regression*

1 match found

<https://unogeeks.com/python-for-data-analysis/>

C12. *Tensorflow: TensorFlow is an open-source machine learning framework developed by*

5 matches found

<https://medium.com/aimonks/top-10-ai-tools-you-need-to-know-about-54520cc41a77?source=rss-----ai-5>  
<https://medium.com/@olaoluwaishreal02/certainly-here-are-7-awesome-and-free-ai-tools-you-can-use-39dba0e31392>  
<https://medium.com/@jamieswears/what-are-best-ai-tools-4bec167b06fd?source=rss-----1>  
<https://jaswdr.dev/posts/>  
<https://hinative.com/questions/24157522>

C13. *ARIMA stands for AutoRegressive Integrated Moving Average, and it is a*

2 matches found

<https://assignment.ignou-service.in/2023/04/time-series-analysis-is-one-of-most.html>  
<https://efzn.othersensesen.com/arima-model-for-time-series-forecasting-in-python-github/35381121>

C14. *In such situations, using numerical summary statistics, such as averages or correlations, in tabular form would provide a more efficient summary.*

1 match found

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7221239/>

C15. *ARIMA (AutoRegressive Integrated Moving Average) is a popular time series model used for forecasting.*

5 matches found

<https://datpen.com/tag/statistics/>  
<https://prodataanalyst.com/faq-copy/>  
<https://medium.com/@shreyasarpal26/time-series-forecasting-55771b2fa401>  
<https://www.javatpoint.com/sales-prediction-using-machine-learning>  
<https://prodataanalyst.com/category/statistics/>

C16. *It is a combination of three components: autoregression (AR), differencing (I), and moving average (MA).*

1 match found

<https://www.linkedin.com/pulse/power-autoregressive-integrated-moving-average-arima-models-pandey>

C17. *the correlation between the time series and its lagged values,*

1 match found

<https://www.javatpoint.com/sales-prediction-using-machine-learning>

C18. *SVM is a machine learning algorithm that can be used for regression and classification tasks.*

2 matches found

<https://vitalflux.com/category/career/interview-questions/page/2/>  
<https://scholarworks.calstate.edu/downloads/wp988r98m>

C19. *Artificial Neural Networks (ANNs) are machine learning algorithms inspired by the human brain's structure and function.*

2 matches found

<https://inclusioncloud.com/insights/blog/artificial-neural-networks-the-future-of-machine-learning/>  
<https://www.mdpi.com/2076-3417/13/5/3186>

C20. *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.*

3 matches found

<https://saturncloud.io/blog/why-is-a-simple-2layer-neural-network-unable-to-learn-00-sequence/>

<https://btu.edu.ge/wp-content/uploads/2023/04/Introduction-to-Machine-Learning-.docx.pdf>

<https://www.tutorialspoint.com/adaline-and-madaline-network>

C21. *Activation functions are mathematical functions that are applied to the output of each neuron in the network.*

3 matches found

<https://medium.com/tech-talk-with-chatgpt/what-you-need-to-know-before-diving-into-transformers-f8007227a3f?source=rss-----1>

<https://saturncloud.io/blog/how-to-restrict-output-of-a-neural-net-to-a-specific-range/>

<https://deveshsingh226.medium.com/deep-learning-series-cnn-103956e1d059>

C22. *introduce nonlinearity into the network, allowing it to model complex relationships*

1 match found

[https://www.linkedin.com/pulse/neural-networks-activation-function-priya-iragavarapu?trk=public\\_post](https://www.linkedin.com/pulse/neural-networks-activation-function-priya-iragavarapu?trk=public_post)

C23. *Common activation functions include the sigmoid function, the hyperbolic tangent (tanh) function, and the rectified linear unit (ReLU) function.*

5 matches found

[https://dev.to/\\_akshaym/neural-network-basics-the-perceptron-3c4](https://dev.to/_akshaym/neural-network-basics-the-perceptron-3c4)

<https://lazyprogrammer.me/mlcompendium/deeplearning/ann.html>

<https://makshay.com/neural-network-basics-the-perceptron/>

<https://devforum.roblox.com/t/devlog-1-creating-a-neural-network-to-play-fps-shooters-on-roblox/2313806>

[https://www.linkedin.com/pulse/neural-networks-activation-function-priya-iragavarapu?trk=public\\_post](https://www.linkedin.com/pulse/neural-networks-activation-function-priya-iragavarapu?trk=public_post)

