

**Dynamic Analysis of Energy Consumption, Production, and Economic Growth on CO2 Emissions Across Countries**

**Siddharth Kr Jaiswal**

**40425082**

**Word Count: 7417**

**Project Submitted in Partial Fulfillment of Degree of Masters of Science in Business Analytics**

**September 2024**

**Queen’s Business School**

# DECLARATION

I, **Siddharth Kr Jaiswal**, hereby declare that this is my original work being submitted for assessment and has not been submitted to any other institution of higher learning for assessment for the Master's Degree. The work conducted in this paper has followed academic ethics. Any work belonging to another author in this dissertation, be it quotes, images or phrases has been well-cited and included in the referencing section at the end of this report.

**…………………. ……………….. …………. ………….**

**Student Name StudentNumber Signature Date**

**Siddharth Kr Jaiswal 40425082 Siddharth kr jaiswal 04/09/2024**

**Supervisors’ Name Signature Date**

# ACKNOWLEDGEMENTS

I sincerely show my gratitude to my supervisor, for helping me through my dissertation from the start to the end. I am very grateful he has been helpful and always challenged me to produce better work. I also thank the whole department at Queen's Business School, both staff and fellow students. I take this opportunity also to thank my family in a special way forever being there for me and providing moral support. God bless you abundantly.

# ABSTRACT

This study investigated the dynamic relationships between energy consumption, production, economic growth, and CO2 emissions across various countries. The study aims to understand how these factors interact and influence each other to provide knowledge of effective strategies for reducing carbon emissions while supporting economic development.

The study incorporated exploratory data analysis, statistical modeling, and predictive analytics. The EDA explored trends and patterns in energy consumption and CO2 emissions to offer initial insights into the data's structure and relationships. For modeling, multiple techniques were employed: Multiple Linear Regression was used to assess the impact of predictors on CO2 emissions, Time Series Forecasting was done using ARIMA models to predict future trends, and K-means Clustering to identify groups of countries with similar profiles.

The study used datasets from kaggle.com, including data on sustainable energy and CO2 emissions spanning several decades. The predictive models were evaluated using performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Adjusted R-squared. The regression model and ARIMA demonstrated robust forecasting accuracy.

The results showed a significant relationship between energy consumption, economic growth, and CO2 emissions. The findings demonstrate the need for tailored energy policies that balance economic growth with environmental sustainability. Recommendations include implementing targeted renewable energy policies, enhancing energy efficiency, and considering carbon pricing mechanisms. The study faced limitations such as data quality and the scope of the datasets. Future studies should be conducted from more granular data and advanced modeling techniques to refine predictions and support more effective policy-making.

TABLE OF CONTENTS

[DECLARATION 2](#_Toc177007324)

[ACKNOWLEDGEMENTS 3](#_Toc177007325)

[ABSTRACT 4](#_Toc177007326)

[TABLE OF CONTENTS 5](#_Toc177007327)

[LIST OF TABLES 8](#_Toc177007328)

[LIST OF FIGURES 9](#_Toc177007329)

[1. INTRODUCTION 10](#_Toc177007330)

[1.1 Background of Study 10](#_Toc177007331)

[1.2 Research Problem 12](#_Toc177007332)

[1.3 Research Objective 13](#_Toc177007333)

[1.4 Research Questions 13](#_Toc177007334)

[1.5 Significance of the Study 14](#_Toc177007335)

[1.6 Scope of the Study 14](#_Toc177007336)

[1.7 Structure of the Study 15](#_Toc177007337)

[2. LITERATURE REVIEW 16](#_Toc177007338)

[2.1 Introduction 16](#_Toc177007339)

[2.2 Environmental Kuznets Curve (EKC) 16](#_Toc177007340)

[2.3 Economic Growth and Environmental Degradation 18](#_Toc177007341)

[2.4 Sustainable Energy Policies 19](#_Toc177007342)

[2.4.1 Promoting Renewable Energy Sources 19](#_Toc177007343)

[2.4.2 Enhancing Energy Efficiency 19](#_Toc177007344)

[2.4.3 Implementing Carbon Pricing Mechanisms 20](#_Toc177007345)

[2.5 Empirical Studies on Energy, Economic Growth, and CO2 Emissions 21](#_Toc177007346)

[2.6 Predicting Future CO2 Emissions 21](#_Toc177007347)

[2.6 Gaps in the Literature 22](#_Toc177007348)

[3. METHODOLOGY 22](#_Toc177007349)

[3.1 Business Understanding 23](#_Toc177007350)

[3.2 Data Understanding 24](#_Toc177007351)

[3.3 Data Preparation 26](#_Toc177007352)

[3.4 Exploratory Data Analysis 27](#_Toc177007353)

[3.5 Modeling 27](#_Toc177007354)

[3.5.1 Multiple Linear Regression Analysis 27](#_Toc177007355)

[3.5.2 Time Series Forecasting 29](#_Toc177007356)

[3.5.3 Cluster Analysis 29](#_Toc177007357)

[3.6 Evaluation 29](#_Toc177007358)

[4. FINDINGS 30](#_Toc177007359)

[4.1 Exploratory Data Analysis 30](#_Toc177007360)

[4.1.1 Descriptive statistics 30](#_Toc177007361)

[4.1.2 Data Visualization 31](#_Toc177007362)

[4.2 Modeling 40](#_Toc177007363)

[1.2.1 Multiple Linear Regression 40](#_Toc177007364)

[1.2.2 Cluster Analysis 42](#_Toc177007365)

[1.2.3 Time Series Forecasting 45](#_Toc177007366)

[5. DISCUSSION OF FINDINGS 48](#_Toc177007367)

[5.2 Limitations of the Study 50](#_Toc177007368)

[6. Conclusions 51](#_Toc177007369)

[7. Recommendations 52](#_Toc177007370)

[References 54](#_Toc177007371)

[Appendix 2: Python Codes 65](#_Toc177007372)

# LIST OF TABLES

[Table 1: The Global Data on Sustainable Energy (2000-2020) 20](#_Toc175856470)

[Table 2**:** Energy and CO2 Emissions Dataset (1980-2020). 22](#_Toc175856471)

[Table 3: Linear Regression Model Performance Metrics 38](#_Toc175856472)

[Table 4: ARIMA Model Performance Metrics 43](#_Toc175856473)

# LIST OF FIGURES

Figure 1:  Environmental Kuznets curve (Stern, 2018) 17

Figure 2: Environmental Kuznets curve for sulfur emissions (Huang et al., 2009) 18

Figure 3: CRISP-DM framework(Taylor, 2018) 23

Figure 5: Summary Statistics 30

Figure 6: Variables’ Correlation Matrix 32

Figure 7: Main Sources of Electricity 33

Figure 8: Global Change in Access to Electricity (2000-2020) 33

Figure 9: Top 10 countries with highest CO2 emissions 34

Figure 10: Top 10 Countries by Electricity Consumption from Fossil Fuels 35

Figure 11: CO2 Emissions (2000–2020) 35

Figure 12: Energy Consumption and CO2 Emissions by Income Category 36

Figure 13: GDP and Energy Production by Income Category 37

Figure 14: Distribution of GDP by Income Category 37

Figure 15: GDP and CO2 emissions by Income category 38

Figure 16: Distribution of Total CO2 Emissions by Income Category 39

Figure 17: Energy Production vs. CO2 Emissions by Income Category 39

Figure 18: Summary of Linear Regression Model 41

Figure 19: Elbow Method Showing the Optimal Number of Clusters 43

Figure 20: Clusters of Countries by Energy Consumption and CO2 Emissions 44

Figure 21: Clusters of Countries by Energy Production and CO2 Emissions 45

Figure 22: Forecasted CO2 emissions (2020-2029) 46

# 1. INTRODUCTION

## 1.1 Background of Study

One of the primary principles of the global climate change conversation is the complex relationship between energy consumption, production, and economic growth, and how these factors affect CO2 emissions. As countries work towards economic development, their dependence on energy, especially from fossil fuels, has increased CO2 emissions, which have a major impact on long-term global temperature rise and environmental degradation(Osobajo et al., 2020). This relationship shows how important it is to understand how different countries manage their economic activity and energy resources to find a balance between environmental sustainability and growth.

Energy is an essential part of sustaining business, manufacturing, and household activities, and it is a basic driver of economic growth. Modern economies cannot run properly without it, as it makes everything from domestic amenities to industrial manufacturing processes possible. Economic expansions usually result in higher energy demand, which raises consumption and, ultimately, CO2 emissions. The main causes of these emissions are the production and consumption of energy, especially those that come from fossil fuels like coal, oil, and natural gas. Therefore, promoting economic growth while reducing negative environmental effects of energy use, is a challenge for policymakers globally(Picano et al., 2022).

Technological developments and the adoption of greener energy sources have reduced CO2 emissions per GDP unit and increased energy efficiency in high-income countries. These countries have imposed strict environmental restrictions that limit emissions and have invested in cleaner technology, such as renewable energy sources (wind, solar, hydro, and biomass). These nations have therefore witnessed a decoupling of economic growth from environmental degradation, proving that economic growth and carbon footprint reduction are compatible (M. M. Rahman et al., 2022). Increased energy efficiency and the incorporation of sustainable practices across multiple industries are common indicators of this decoupling.

However, because fossil fuels are readily available and affordable, a lot of low-income nations continue to primarily rely on them for the production of energy. As these economies grow, this dependence has caused carbon emissions to rise. These countries’ shortage of financial resources and technological know-how makes it difficult to make a shift to cleaner energy sources. The difficulties in reducing CO2 emissions are made even worse by the fact that economic growth in these areas frequently puts short-term developmental requirements ahead of long-term environmental sustainability. It is important to have fair and practical climate policies that take into account the particular conditions of these countries. In the absence of such policies, international efforts to reduce climate change might not be successful (Azam et al., 2016)

Growth in the economy can have varying effects on CO2 emissions over time and between nations(González-Álvarez & Montañés, 2023). Environmental deterioration first tends to rise with industrialization and economic development in a nation. Higher CO2 emissions and energy use are characteristics associated with this period. But when countries get wealthier, they frequently make investments in greener technology and adopt stricter environmental laws. As a result of this expenditure, per capita emissions decrease. This is a phenomenon referred to as the Environmental Kuznets Curve (EKC). According to Grossman and Krueger (1995), the EKC indicates that, while economic expansion initially causes environmental deterioration, it eventually leads to improvements in the environment once a particular income level is reached.

Creating sustainable energy policies and programs requires an understanding of the dynamic relationships that exist between energy production, consumption, economic growth, and CO2 emissions(Gershon et al., 2024). This study aimed to explore these relationships across different countries, identifying patterns and forecasting future trends to provide knowledge that helps policymakers design strategies that promote economic growth while minimizing CO2 emissions. These strategies include investing in renewable energy technologies, enhancing energy efficiency, and implementing policies that encourage sustainable practices across various sectors.

This study also points out how important it is for nations to work together to reduce CO2 emissions globally. Since the effects of climate change are global in scope, cooperation is essential to reducing their effects. Large economies have a major part to play in helping low-income nations make a shift to cleaner energy sources by giving them financial and technological support.

## 1.2 Research Problem

Although a study has been done in the past on the variables influencing CO2 emissions, no thorough studies have been done that concurrently consider energy consumption, production, and economic growth in a wide number of nations. The differences between high, middle, and low-income countries' economies were not adequately taken into consideration. By offering a dynamic analysis of these interrelated elements and their combined impact on CO2 emissions, this study aimed to close this gap. This study aimed to fill this research gap by offering an in-depth analysis of economic growth, energy production, and consumption across many countries. This provides a deeper understanding of these factors' combined impact on CO2 emissions by looking at them together while considering how they interact in various economic settings. This analysis is essential for creating focused and practical climate policies that can take into account the diverse economic status of many countries and support more sustainable global climate initiatives.

## 1.3 Research Objective

The primary aim of this research is to examine the relationship between changes in energy production and consumption and CO2 emissions in the context of global economic expansion.

Specific Objectives:

1. To determine the correlation between energy consumption and CO2 emissions among high, middle, and low-income countries.
2. To determine how fluctuations in economic growth influence energy production and subsequent CO2 emissions.
3. To develop models that enable future CO2 emissions to be predicted based on current trends in energy consumption and the level of economy in a country.

## 1.4 Research Questions

1. How does energy consumption correlate with CO2 emissions across high, middle, and low-income countries?
2. How do fluctuations in economic growth influence energy production and subsequent CO2 emissions across different nations?
3. What are the predictive relationships between current trends in energy consumption, economic growth, and future CO2 emissions across different countries?

## 1.5 Significance of the Study

This study is very significant because it provides important insights into the complex relationship between energy production, consumption, economic growth, and CO2 emissions in various nations. By addressing the limitations in the literature, and the absence of a thorough analysis across different income categories, this study offers a deeper understanding of the relationship between economic status and CO2 emissions. It assists in the development of focused climate policies by constantly examining the relationships between important variables to determine how these factors affect emissions over time. Predictive modeling supports proactive mitigation strategies by providing precise forecasts of future CO2 emissions. Moreover, by improving knowledge of the Environmental Kuznets Curve and guiding international initiatives, this study supports efforts to address climate change on a global level. The results of this study help researchers, organizations, and governments create fairer and successful climate action programs that support sustainable development while achieving a balance between environmental sustainability and economic growth.

## 1.6 Scope of the Study

This study investigated the dynamics of energy consumption, energy production, economic growth, and CO2 emissions using a combination of analytical models and secondary data. Three models were applied: Multiple Linear Regression, ARIMA modeling, and K-Means Clustering. The Multiple Linear Regression model was used to explore relationships between the variables and their collective impact on CO2 emissions. The ARIMA model was employed to forecast future CO2 emissions based on historical trends. K-Means Clustering was used to categorize countries into clusters based on their energy and economic profiles. The study relied on secondary data sourced from Kaggle.com and employed quantitative research methods for data processing and analysis using statistical software, Jupyter Notebook.

## 1.7 Structure of the Study

The study is structured into five chapters: Chapter One covers the introduction, including the background, problem statement, research objectives, questions, significance, scope, and structure of the study. Chapter Two provides a literature review of relevant past research. Chapter Three details the data collection and analysis techniques. Chapters Four and Five present the results analysis, and discussion of the findings respectively. Chapter Six concludes with key findings, and Chapter 7 gives recommendations based on the study's results.

# 2. LITERATURE REVIEW

## 2.1 Introduction

This section examines the research-relevant theoretical and empirical literature. Discussion Topics include the Environmental Kuznets Curve (EKC), the link between environmental deterioration and economic growth, and the effect of renewable energy policy on CO2 emissions. The foundation of this research is highlighted in the review.

## 2.2 Environmental Kuznets Curve (EKC)

Numerous studies have already looked closely at the connection between CO2 emissions, output, energy consumption, and economic growth in the framework of sustainable development. In this field, the EKC hypothesis is a basic idea. There seems to be an inverted U-shaped relationship between environmental degradation and economic growth. Environmental Deterioration initially rises with economic expansion, but after GDP per capita reaches a certain point, more economic growth promotes improvements in environmental quality. This is shown in Figure 1 below.

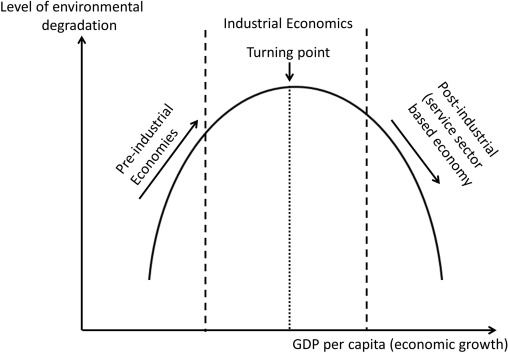


Figure 1:  Environmental Kuznets curve (Stern, 2018)

Grossman and Krueger (1991) developed this seminal concept by utilizing data on sulfur dioxide and smoke to demonstrate that pollution levels rise with income to a certain point and then fall as seen in Figure 2 below.

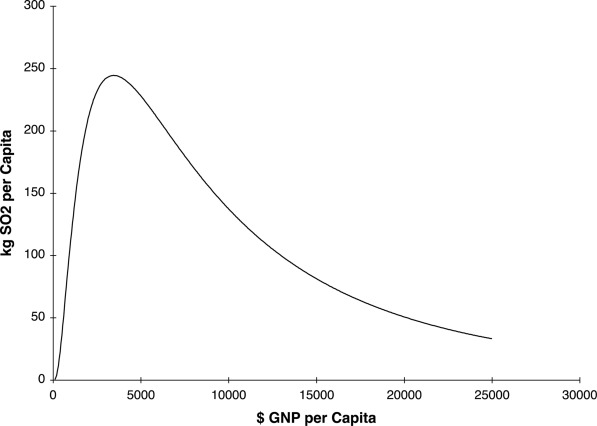


Figure 2: Environmental Kuznets curve for sulfur emissions (Huang et al., 2009)

Stern (Stern, 2004) critically assessed the EKC literature. He emphasized the need for careful empirical testing and considering factors such as trade, technology, and policy interventions. He highlighted that the EKC may not hold universally and that its shape and turning points can vary significantly across different pollutants and regions. Also, Dinda (Dinda, 2004) conducted a review of EKC experiments and made the argument that although the EKC hypothesis offers a helpful framework, it is oversimplified.

## 2.3 Economic Growth and Environmental Degradation

The correlation between environmental degradation and economic growth varies based on the economic structure and development level (Acheampong & Opoku, 2023). Stern et al. (1996) studied the connection between different markers of environmental quality together with income. They discovered mixed evidence for the EKC across many contaminants. They discovered that certain environmental indicators improved with income while others did not. This implies a more intricate relationship. Tashtamirov (2023) emphasized the impact of policy and institutional elements on the correlation link environmental quality and economic prosperity. He maintained that strong institutions and the right policies are essential to ensuring that economic expansion results in better environmental outcomes. Dasgupta et al. (2002) looked into the relationship between environmental quality and economic expansion. They discovered that economic growth frequently results in higher pollution levels in the absence of strict environmental controls. However, growth can become more sustainable with the right policies.

## 2.4 Sustainable Energy Policies

Policies promoting sustainable energy seek to boost economic expansion while lowering CO2 emissions. Among these actions are the promotion of renewable energy sources, improvements in energy efficiency, and the implementation of carbon pricing schemes (Xue et al., 2022). The effectiveness of these measures can significantly impact the country's ability to reduce its carbon emissions and transition to a sustainable energy system.

### 2.4.1 Promoting Renewable Energy Sources

Renewable energy sources help in lowering CO2 emissions(Chandra Voumik et al., 2023). Subsidies, tax breaks, and requirements such as renewable portfolio standards (RPS) are examples of policies that encourage the use of various energy sources(Dilanchiev et al., 2023). Encouraging policies for renewable energy has significantly increased the share of renewables in their energy mix (Munro et al., 2020). Laws have aided in lowering the cost of renewable energy technologies, increasing their parity with fossil fuels.

### 2.4.2 Enhancing Energy Efficiency

An additional crucial element of sustainable energy policy is increasing energy efficiency. Energy-saving techniques can cut CO2 emissions and energy use by a large amount without slowing down economic expansion. Establishing energy performance requirements, providing financial incentives for energy-saving devices, and launching awareness campaigns to promote energy-saving behaviors are some examples of policies to improve energy efficiency(Linares & Labandeira, 2010).

The significance of energy efficiency in accomplishing global climate targets is emphasized by the *Global Energy Review 2020*(2020). According to the agency's data, increases in energy efficiency have accounted for around 40% of the recent decrease in CO2 emissions associated with energy. all essential areas for improving energy efficiency. For example, repairing old structures can significantly cut energy use, whilst stricter building codes and standards can ensure that new projects are energy-efficient.

Industry can save a lot of energy and cut emissions by implementing energy-efficient practices and technologies(Cai et al., 2019). Additionally, promoting electric cars, fuel-efficient cars, and public transportation can reduce the amount of energy used in the transportation industry(Linares & Labandeira, 2010).

### 2.4.3 Implementing Carbon Pricing Mechanisms

By internalizing the external costs of carbon emissions, pricing schemes such as carbon taxes and cap-and-trade programs can also help corporations reduce their carbon footprint economically. These policies incentivize the adoption of environmentally friendly technologies and behaviors by putting a price on carbon(Stella Emeka-Okoli et al., 2024).

Research conducted by the World Bank ( 2020) found that many nations that have implemented the price of carbon schemes have seen notable drops in their CO2 emissions. In this instance, the European Union Emissions Trading System (EU ETS) has been effective in reducing emissions from power plants and other significant industrial sources. The scheme, which restricts overall emissions and allows firms to buy and sell emission allowances, creates a financial incentive to reduce emissions.

## 2.5 Empirical Studies on Energy, Economic Growth, and CO2 Emissions

Several empirical studies have been conducted to investigate the links between energy consumption, production, economic growth, and CO2 emissions in different countries. Ang (2007) used time series data to investigate the relationship between CO2 emissions, energy consumption, and economic growth in France. It establishes a relationship between these parameters that maintain long-term equilibrium. Soytas et al. (2007) made policy suggestions based on their research into the causal relationships between energy consumption, economic growth, and CO2 emissions in the United States. According to their findings, there is a need for integrated policy responses because energy consumption and economic growth are significant drivers of carbon emissions.

## 2.6 Predicting Future CO2 Emissions

Predicting future CO2 emissions is an important area of research, particularly in the context of global efforts to mitigate climate change. Accurate forecasting models are essential for policymakers to design effective strategies that align with sustainable development goals(Nahar, 2024). The complexity of predicting CO2 emissions arises from the need to account for multiple variables, including economic growth, energy consumption, technological advancements, and policy interventions(Jin et al., 2024). Predicting future CO2 emissions has obtained substantial attention in environmental economics and data science, driven by the need to understand and mitigate climate change impacts(Hussin et al., 2023). Early forecasting efforts primarily employed **Time Series Models**, such as ARIMA (Autoregressive Integrated Moving Average), to predict emissions based on historical data(Lai & Dzombak, 2020).

(A. Rahman & Hasan, 2017) utilized ARIMA models to track CO2 emissions and demonstrated that while these models effectively captured historical trends, their predictive accuracy declined when extended over longer horizons due to their reliance on linear assumptions and historical patterns alone. More recent advances have seen a shift towards **Machine Learning Techniques**, which offer greater flexibility and accuracy in forecasting complex, non-linear relationships. (Acheampong & Boateng, 2019) applied Artificial Neural Networks (ANNs) to predict CO2 emissions and found that ANNs outperformed traditional linear models. They provide more accurate forecasts by learning intricate patterns from large datasets.

## 2.6 Gaps in the Literature

Despite a wealth of study, there are still gaps in our knowledge of the dynamic linkages between production, energy use, economic stage, and CO2 release, especially when considering various countries. Comparative studies spanning a broad variety of nations with differing degrees of economic development have not been widely undertaken. This gap emphasizes how more studies that take into account the various economic situations and their particular possibilities and difficulties are needed. Furthermore, sophisticated prediction models are required which can project future CO2 emissions by utilizing data on economic performance and energy consumption trends. Policymakers would be able to create forward-thinking and successful plans with the help of these models

# 3. METHODOLOGY

This chapter describes the approaches that were followed in the study to meet the main objective. The study followed the Cross-Industry Standard Process for Data Mining, structure. The six phases of the CRISP-DM framework are business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. To guarantee that the study is thorough, organized, and repeatable, each stage is crucial(Schröer et al., 2021). The structure of the framework is shown in Figure 3 below. This methodological approach was chosen because of its flexibility and effectiveness in guiding complex data mining projects.

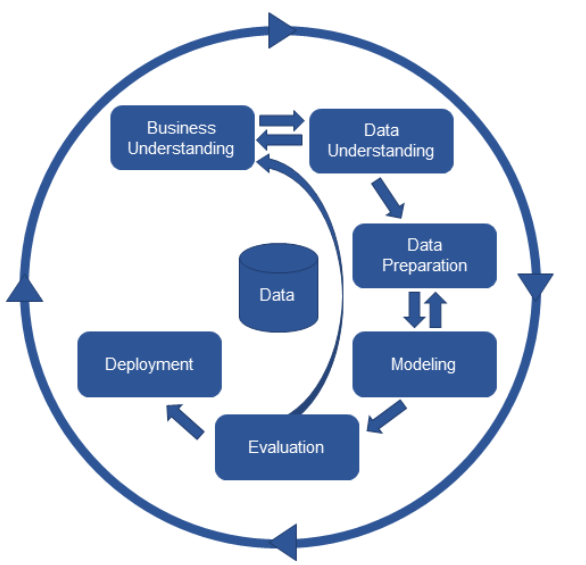


Figure 3: CRISP-DM framework(Taylor, 2018)

## 3.1 Business Understanding

Understanding the requirements and goals of the research from a business point is the first step in data analysis(Ejuma Martha Adaga et al., 2024). The main aim of this study was to explore the dynamic relationships between energy consumption, production, level of economy, and CO2 emissions across different countries (Low, middle, and high-income countries). This study provides knowledge that informs energy policies and strategies for sustainable development. Specifically, the research objectives were to:

* To determine the correlation between energy consumption and CO2 emissions among high, middle, and low-income countries.
* To determine how fluctuations in economic growth influence energy production and subsequent CO2 emissions.
* To develop models that enable future CO2 emissions to be predicted based on current trends in energy consumption and level of economy in a country?

## 3.2 Data Understanding

The data for this study were sourced from Kaggle, a reputable platform for obtaining diverse datasets. Two primary datasets were utilized. The Global Data on Sustainable Energy (2000-2020) and the Energy and CO2 Emissions Dataset (1980-2020). The Global Data on Sustainable Energy dataset has 3649 rows and 21 columns and it was used to conduct exploratory analysis while the Energy and CO2 emission data has 55440 rows and 11 columns. This dataset was used to build predictive models. Table 1 and 2 below shows the variable's descriptions. Both datasets were selected because of their relevance to the research objectives and their extensive coverage of energy.

Table 1: The Global Data on Sustainable Energy (2000-2020)

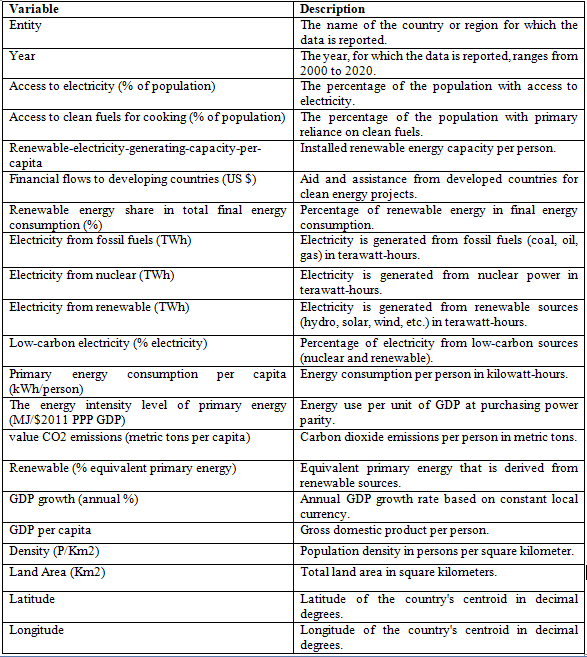
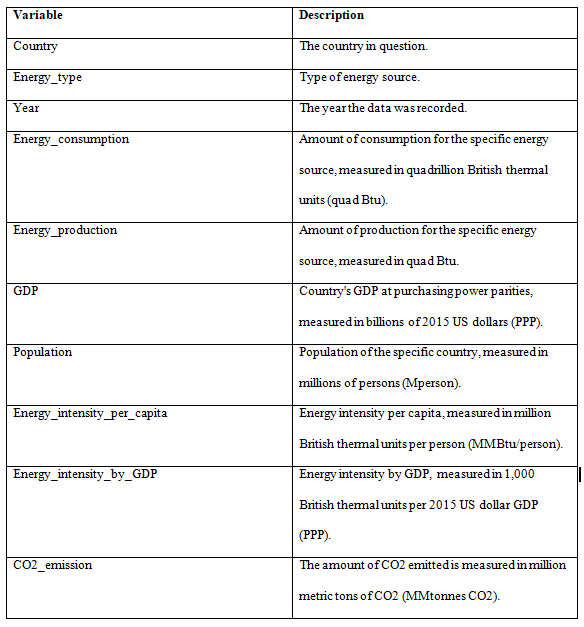


Table 2**:** Energy and CO2 Emissions Dataset (1980-2020).



## 3.3 Data Preparation

Data preparation is an important step in data analysis(Abdallah et al., 2017). After the datasets were downloaded from kaggle.com, they were subsequently loaded into Jupyter Notebook for the cleaning process. The first step involved inspecting the datasets for missing values and duplicates. The rows with missing values were identified and removed to ensure data completeness. Missing data in data leads to biased results, reduced statistical power, impaired model performance, and challenges in interpretation which can significantly affect the validity and generalizability of the findings(Padgett et al., 2014). Duplicate rows were also checked to avoid redundancy. The Year variable was then converted to a date-time format to facilitate time series analysis. Also, the data was standardized to ensure that features were on the same scale before conducting cluster analysis. Moreover, the data was split into training and testing sets for training the model and evaluating the performance of the model respectively.

## 3.4 Exploratory Data Analysis

Exploratory Data Analysis is conducted as a foundational step to understand the underlying patterns, trends, and relationships within the data(Ho Yu, 2010). EDA involves visualizing the distribution of key variables and examining correlations between energy consumption, production, GDP, and CO2 emissions. Visualizations, including bar plots, and scatter plots were used to visualize the data.

## 3.5 Modeling

Various models were applied to the prepared data to analyze the relationships and make predictions of CO2 emissions. This research used the following analytical methods:

### 3.5.1 Multiple Linear Regression Analysis

Multiple linear regression models were used to explore the relationship between CO2 emissions and independent variables, Energy Consumption, Energy Production, GDP, Population, Energy Intensity per Capita, and Energy Intensity by GDP. Linear regression is used to assess the relationship between dependent and independent variables(Uyanık & Güler, 2013). The equation of the multiple linear regressions is shown below.



Where *y* is the target variable, X1, X2, Xn is the predictor variables, β0 is the intercept, β1, β2… βn are the coefficients of the independent variables, and ϵ is the error term.

Multiple regression models were used to identify the effect of each predictor's variable on CO2 emissions. The general form of the multiple regression equation used is:

Where:

* CO2 Emission is the dependent variable.
* Energy Consumption, Energy Production, GDP, Population, Energy Intensity per Capita, and Energy Intensity by GDP are the predictors
* Β0 ​ is the intercept term.
* β1, and β2, are the coefficients for the predictors, and
* ϵ is an error term.

### 3.5.2 Time Series Forecasting

Time series forecasting involves using previous data to anticipate future values (Ratnadip Adhikari & R. K. Agrawal, 2013). This study used the ARIMA model to forecast the total CO2 emissions for 2020 up to 2029. ARIMA models explain a particular time series using its previous values ​​(autoregressive terms), past errors (moving average terms), and a difference in raw observations (integrated part) to render the time series stationary. The ARIMA model was chosen because of its robustness in predicting future values of CO2 emissions based on historical data. This allows for accurate and reliable forecasts of CO2 emissions.

### 3.5.3 Cluster Analysis

Cluster analysis was conducted using the K-means clustering algorithm. This unsupervised machine learning technique was used to group countries with similar energy usage, production, and CO2 emissions profiles. K-means Clustering is an unsupervised machine learning approach for dividing a dataset into k distinct, non-overlapping clusters. Each data point is assigned to the cluster with the closest mean, known as the cluster centroid(Zhao & Zhou, 2021).

## 3.6 Evaluation

The performance of the models was rigorously evaluated to ensure that they met the research objectives. Key performance metrics used included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Adjusted R-squared. MAE is used to measure the average magnitude of errors in a set of predictions. RMSE measures the square root of the average squared differences between predicted and actual values. Adjusted R-squared indicates how well the independent variables explain the variation in the dependent variable, adjusted for the number of predictors in the model(Hodson, 2022).

# 4. FINDINGS

## 4.1 Exploratory Data Analysis

### 4.1.1 Descriptive statistics

Summary statistics of the variables in the data are shown in Figure 5 below. The dataset contains 33,065 observations for each variable. On average, energy consumption and production are close to each other, with means of approximately 0.945 quad Btu and 0.929 quad Btu, respectively. This shows that energy usage is roughly balanced with production in general. The GDP has a mean of $470.83 billion but with a large standard deviation of $1,577.16 billion. This shows a wide range of economic output among countries. Population size also varies greatly, with a mean of approximately 38.58 million people and a maximum of over 1.4 billion.

Energy intensity per capita has a mean of 87.14 MMBtu per person, while energy intensity by GDP has a mean of 4.77 MMBtu per PPP GDP. This indicates differing energy efficiency across countries. CO2 emissions have a mean of 56.71 MMtonnes, with a wide range from -0.005 MMtonnes to 10,732 MMtonnes. This shows substantial differences in emissions levels globally, likely due to factors such as industrialization and the types of energy sources used.

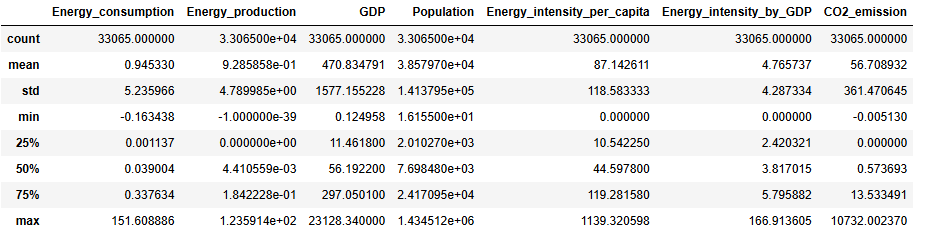


Figure 5: Summary Statistics

### 4.1.2 Data Visualization

Data visualization involves the use of charts, tables, and graphs to display the data. It provides insights that can be used for making data-driven decisions(Zhou, 2023). It was important to determine the correlation between the variables of interest. From Figure 6 below, it can be observed that there is a strong positive correlation between energy consumption and energy production. This shows that countries with high energy consumption also tend to have high energy production. Both energy consumption and production show strong positive correlations with CO2 emissions which suggest that increased energy activities contribute significantly to CO2 emissions. GDP and population are also positively correlated with CO2 emissions, although to a lesser extent. Interestingly, energy intensity per capita and energy intensity by GDP show weaker correlations with CO2 emissions. This shows that these measures of energy efficiency are less directly related to emission levels.

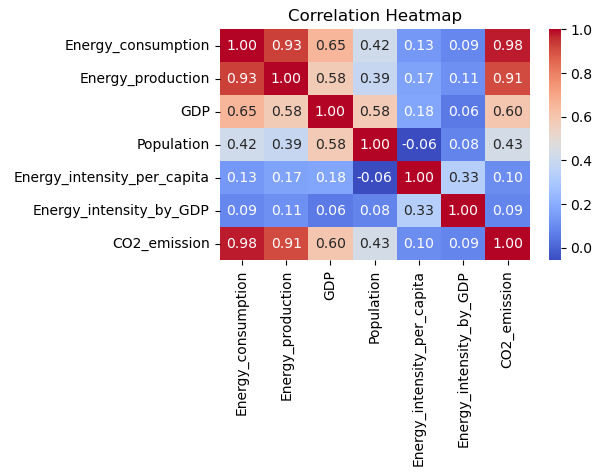


Figure 6: Variables’ Correlation Matrix

It was important to determine the main source of electricity globally. It can be observed from Figure 7 below that, fossil fuel is a primary energy source, accounting for 72.7% of total electricity production. 24.6% is from renewable energy sources and 2.71% is from nuclear energy.

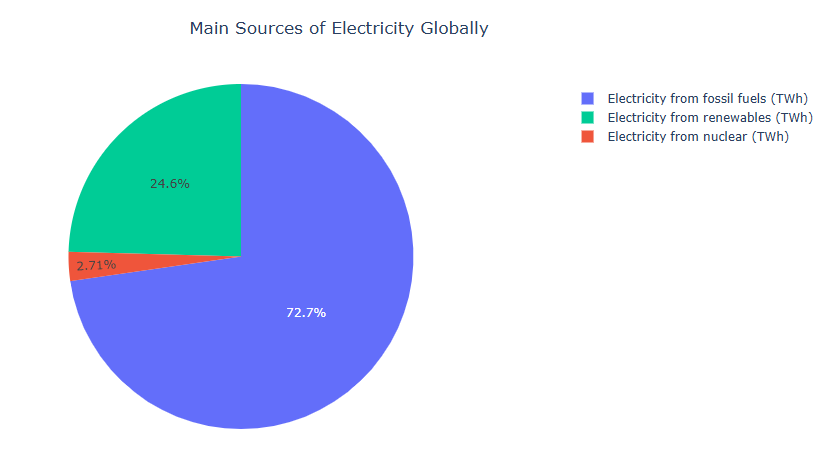


Figure 7: Main Sources of Electricity

How access to electricity has changed between the years 2000 and 2020 is shown in Figure 8 below. It can be observed that global access to electricity has consistently increased. This uptrend suggests that more populations, especially in developing regions, gained access to electricity.

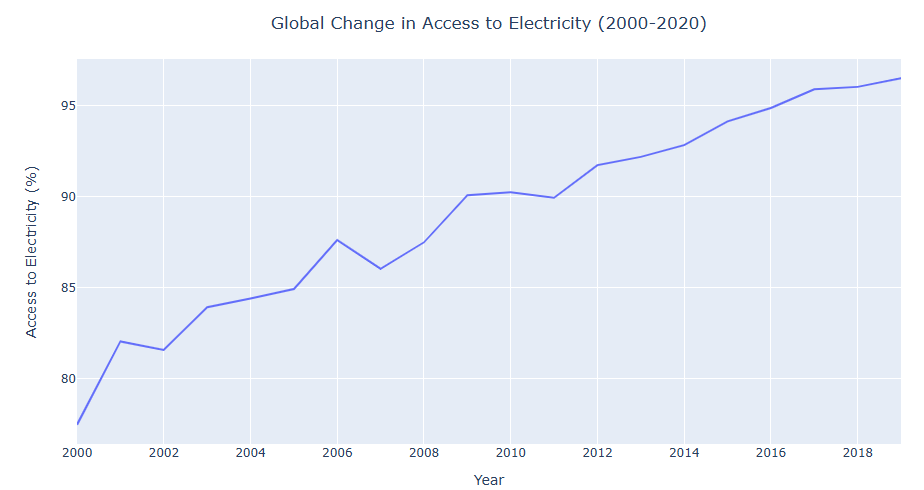


Figure 8: Global Change in Access to Electricity (2000-2020)

The top 9 countries with the highest CO2 emissions are shown in Figure 8 below. It can be observed that China and the USA have significantly higher CO2 emissions compared to the rest of the world. It is followed by India, Russia, Japan, Germany, South Korea, Iran, Saudi Arabia, and the UK. China contributes significantly to global CO2 levels due to its vast industrial sector and reliance on coal for energy.

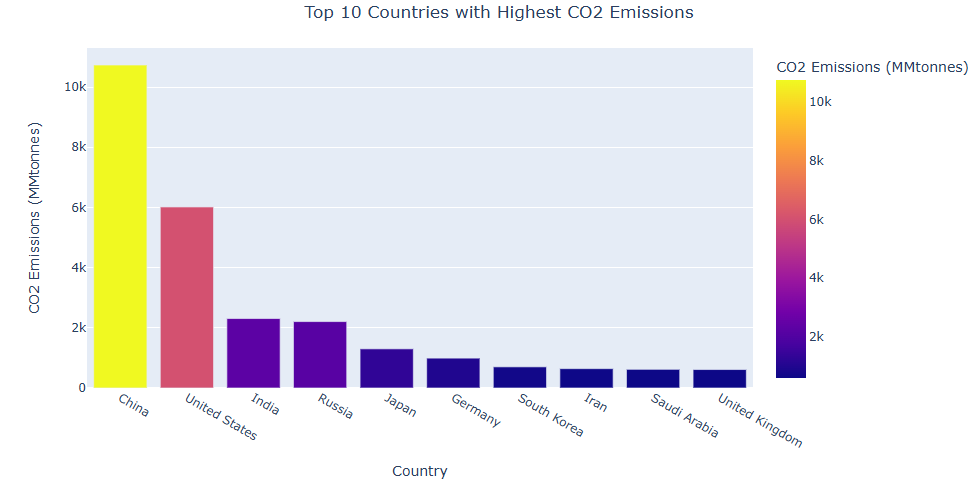


Figure 9: Top 10 countries with highest CO2 emissions

Figure 10 below shows the top 10 countries with the highest CO2 emissions. China is leading followed by India, South Africa, Mexico, Thailand, Argentina, Brazil, Pakistan, the Philippines, and Algeria. There is a significant concentration of emissions in developing and emerging economies. China, as the leading emitter, has a vast industrial base and high reliance on coal, contributing massively to global CO2 levels. India, with its large population and growing industrial sector, also contributes heavily. Mexico, Thailand, Argentina, and Brazil represent rapidly industrializing nations with growing energy demands.

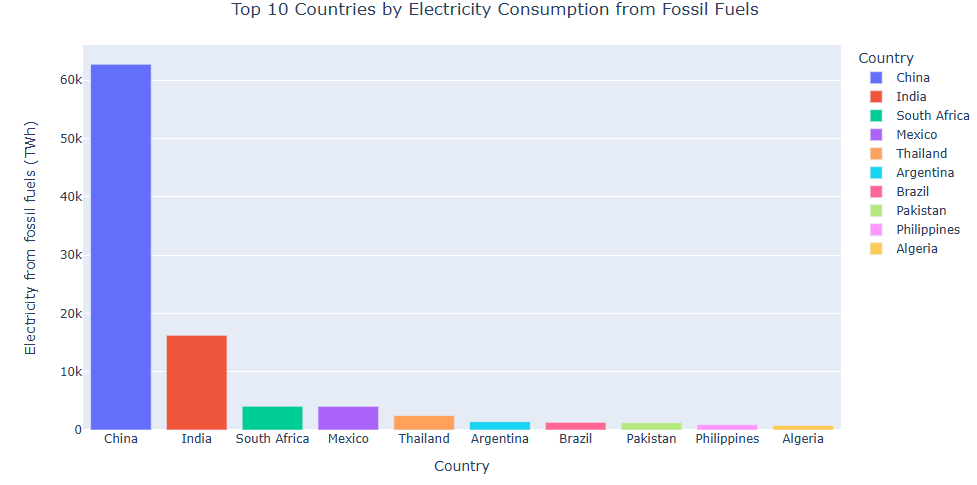


Figure 10: Top 10 Countries by Electricity Consumption from Fossil Fuels

The trend in global CO2 emissions has been on the rise as shown in figure 11 below. This shows a continuous growth of industrial activities, energy consumption, and population expansion worldwide. Economic development in emerging economies, coupled with industrialization and urbanization, has driven a significant increase in CO2 emissions.

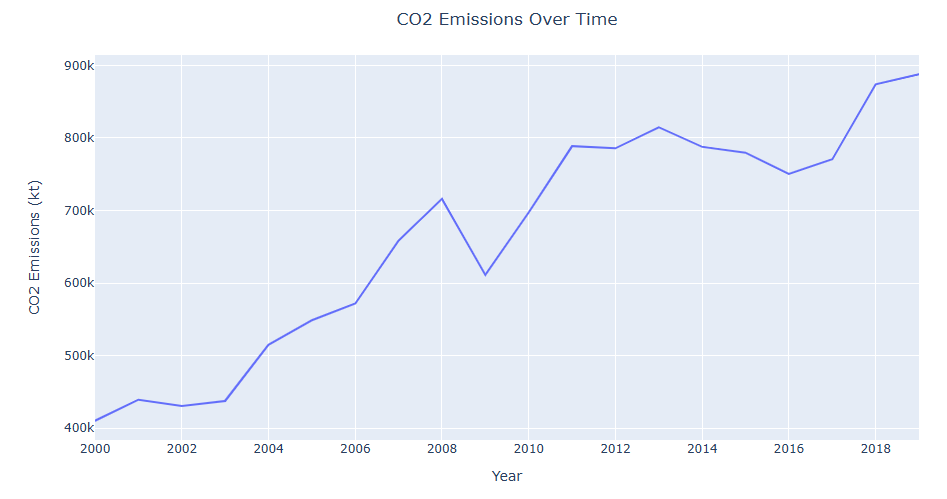


Figure 11: CO2 Emissions (2000–2020)

The scatter plot in Figure 12 below shows the relationship between Energy Consumption and CO2 Emissions across different income categories. It can be observed that there is a consistent trend across all income groups. As energy consumption increases, CO2 emissions also tend to rise. This means that higher energy usage generally leads to higher CO2 emissions.

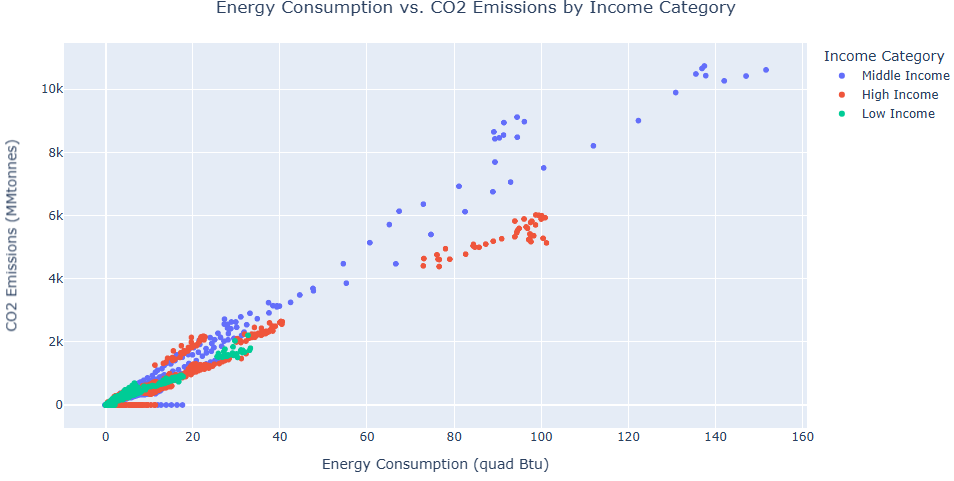


Figure 12: Energy Consumption and CO2 Emissions by Income Category

How economic growth influences energy production in different income categories of countries was determined using the scatter plot in Figure 13 below. It can be observed that there is a positive correlation across different income categories. As GDP increases, energy production also tends to increase in high, middle, and low-income nations.



Figure 13: GDP and Energy Production by Income Category

It was also important to determine the distribution of GDP in low, high, and middle-income nations. From Figure 14 below, it is evident that high-income nations have the highest total GDP of approximately $5.4M. It is followed by middle-income and low-income nations with $3.7M and $1.1M respectively.

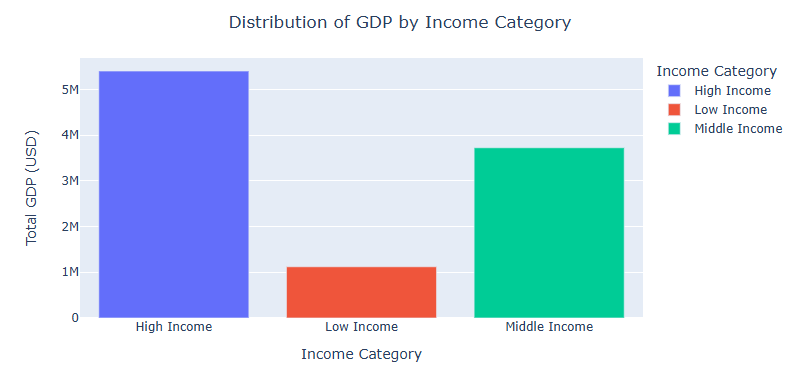


Figure 14: Distribution of GDP by Income Category

Figure 15 below shows the relationship between GDP and CO2 emission in low, middle, and high-income countries. It can be observed that there is a positive relationship in all three income categories. As GDP increases, CO2 emissions tend to rise correspondingly.

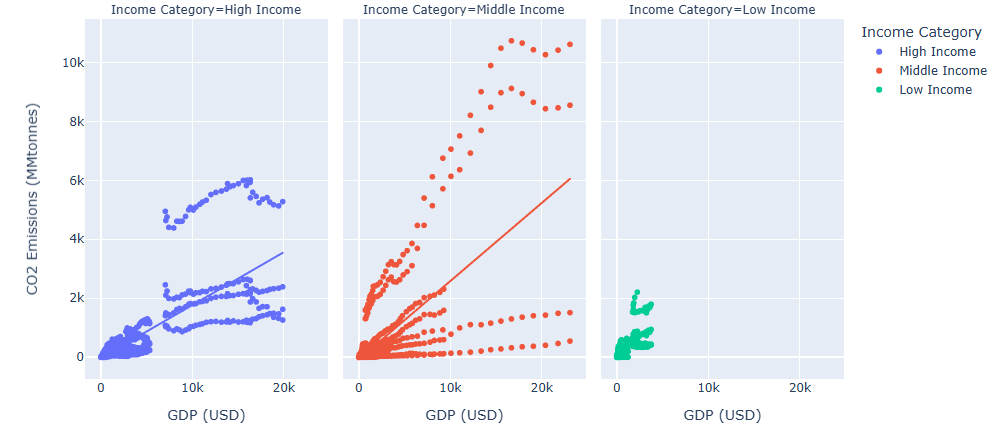


Figure 15: GDP and CO2 emissions by Income category

The distribution of CO2 emissions in low, middle, and high-income countries is shown in Figure 16 below. High-income countries have the highest total CO2 emission of approximately 862.8K MMtonnes. It is followed by middle and low-income nations with 730.4K MMtonnes and 229.9K MMtonnes respectively.

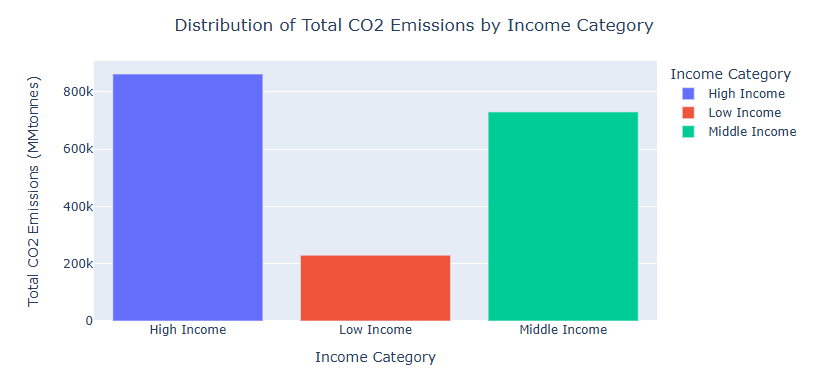


Figure 16: Distribution of Total CO2 Emissions by Income Category

Figure 17 below shows the relationship between energy production and CO2 emissions in low, middle, and high-income nations. It can be observed that in all the income categories, energy production and CO2 emissions have positive relationships. An increase in energy production leads to an increase in CO2 emissions.

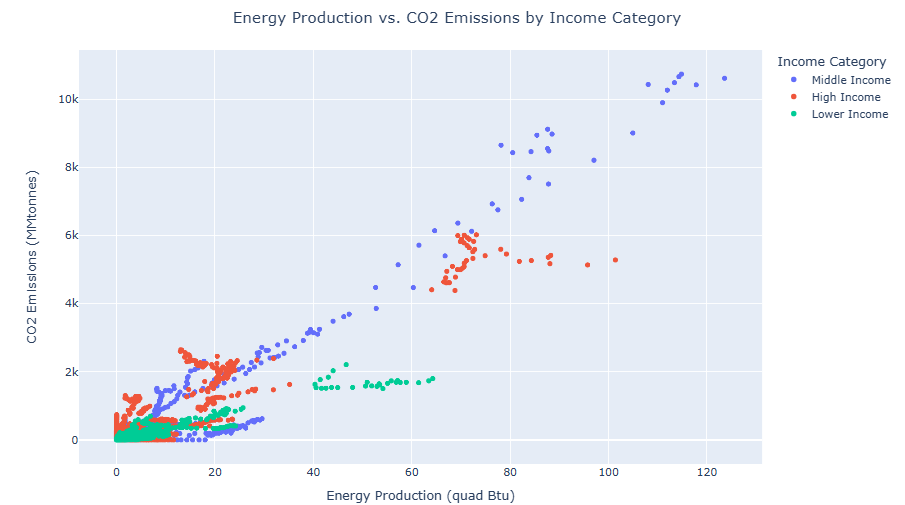


Figure 17: Energy Production vs. CO2 Emissions by Income Category

## 4.2 Modeling

Models were built to predict CO2 emissions based on current trends in energy consumption, energy production, and economic activity. A multiple Linear Regression model was used to assess the effect of each predictor's variable on CO2 emissions. ARIMA Model was used to forecast future CO2 emissions. Cluster Analysis, K-means clustering was used to group countries with similar profiles regarding their energy use, production, economic growth, and CO2 emissions.

### Multiple Linear Regression

A linear regression model was developed and used to assess the relationship between CO2 emissions and predictor variables, Energy Consumption, Energy Production, GDP, Population, Energy Intensity per Capita, and Energy Intensity by GDP. This helps in determining how changes in the predictors affect CO2 emissions and to quantify their impact. The summary of the obtained model is shown in Figure 18 below. It can be observed that energy consumption has a strong positive effect on CO2 emissions. It has a coefficient of 68.98. This means that for every additional quad Btu of energy consumption, CO2 emissions are expected to increase by approximately 68.98 MMtonnes. This relationship is highly statistically significant since its p-value < 0.05 alpha significance level.

Also, energy production has a positive impact on CO2 emissions, with a coefficient of approximately 1.76. This indicates that for every additional quad Btu of energy production, CO2 emissions rise by about 1.76 MMtonnes. This effect is also statistically significant because its p-value < 0.05 alpha significance level. GDP has a very weak and statistically significant negative effect on CO2 emissions, with a coefficient of -0.0235. This shows that an increase in GDP up to a certain level leads to a decrease in CO2 emissions. Also, Population has a small but significant positive effect. The produced coefficient value is 0.0002, meaning that population growth marginally contributes to higher emissions. Energy intensity per capita also has a significant negative coefficient of -0.0353. This indicates that more efficient energy use per person reduces CO2 emissions by 0.353 MMtonnes. Energy intensity by GDP shows a negative but insignificant effect on emissions, P-value > 0.05 alpha significance level. This means that its impact on CO2 emissions may not be substantial in this model.

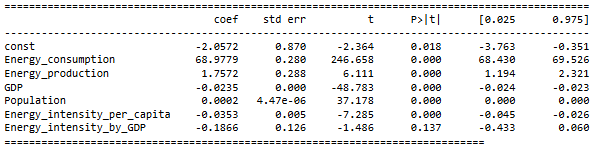


Figure 18: Summary of Linear Regression Model

The performance of the linear regression model was assessed using the MAE, RMSE, R-squared, and Adjusted R-squared. The model demonstrated a strong ability to predict CO2 emissions based on energy consumption, energy production, and GDP, with minimal error and high explanatory power. The model has an MAE of 21.14 MMtonnes. This means that, on average, the model's predictions deviate from the actual CO2 emissions by about 21.14 MMtonnes. This value is low and it suggests that the model's predictions are quite accurate. The RMSE of the model is 80.57 MMtonnes. This is the average squared difference between predicted and actual values, with larger errors being penalized more heavily. Despite some larger deviations, the model generally performs well, with most errors falling within a reasonable range. The R-squared value is 0.959, meaning that approximately 96% of the variance in CO2 emissions is explained by the model. This high value indicates that the model has excellent explanatory power and fits the data very well. The adjusted R-squared is 0.959, which is very close to the R-squared value. This metric accounts for the number of predictors in the model and adjusts for the possibility of overfitting. This value is high and it shows that the model is well-specified and that the predictors (Energy Consumption, Energy Production, and GDP) significantly contribute to explaining the variance in CO2 emissions. Below is a table showing the model metrics.

Table 3: Linear Regression Model Performance Metrics

|  |  |
| --- | --- |
| **Metric** | **Value (MMtonnes)** |
| Mean Absolute Error (MAE) | 21.14 |
| Root Mean Squared Error (RMSE) | 80.57 |
| R-squared | 0.959 |
| Adjusted R-squared | 0.959 |

### Cluster Analysis

Cluster analysis is an unsupervised machine learning technique used to group data points into clusters based on their similarity. The goal is to ensure that data points within the same cluster are more similar to each other than to those in other clusters. K-means clustering was used to identify groups of countries with similar profiles in terms of energy consumption, production, economic growth, and CO2 emissions. Elbow Method was used to find the optimal number of clusters. In Figure 19 below, it can be observed that the rate of decrease in inertia sharply slows down at k=3. Therefore, the optimal number of clusters is 3.

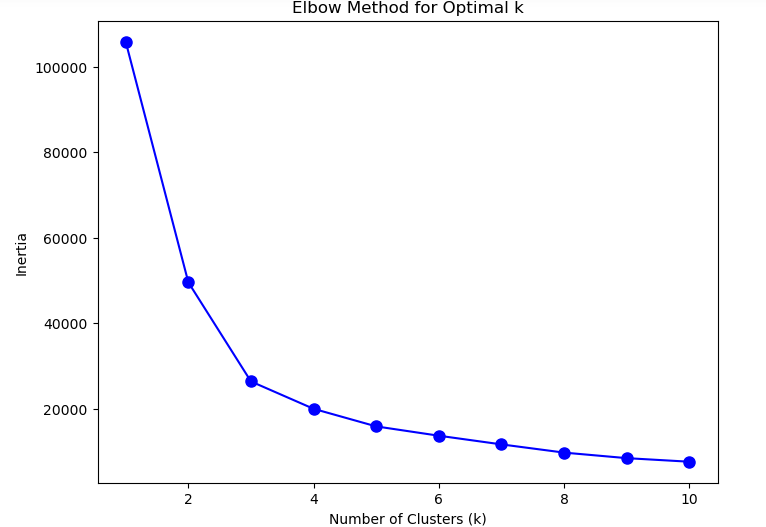


Figure 19: Elbow Method Showing the Optimal Number of Clusters

The k-means algorithm was applied with the selected number of clusters to assign each data point to a cluster. Figure 20 below shows the obtained clusters of countries by energy consumption and CO2 emissions. It can be observed that cluster 1 is characterized by relatively low energy consumption below 20 quad Btu and corresponding CO2 emissions below 1000 MMtonnes. Cluster 2 includes countries with moderate energy consumption mostly between 20 and 60 quad Btu and CO2 emissions between 1000 and 4000 MMtonnes. Cluster 3 consists of countries with very high energy consumption exceeding 60 quad Btu and the highest CO2 emissions over 4000 MMtonnes.

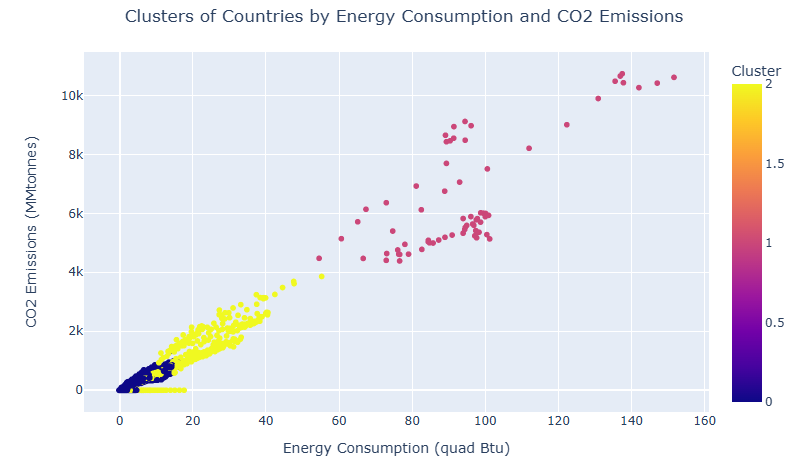


Figure 20: Clusters of Countries by Energy Consumption and CO2 Emissions

Clustering was also done based on the energy production. It can be observed that cluster 1 represents countries with low energy production and low CO2 emissions below 30 quad Btu and 2000 MMtonnes respectively. **Cluster 2** consists of countries with moderate energy production (between between 20 and 60 quad Btu) and corresponding CO2 emissions between 1000 MMtonnes and 4000 MMtonnes. **Cluster 3 r**epresents high energy production and high CO2 emissions above 60 quad Btu and 4000 MMtonnes.

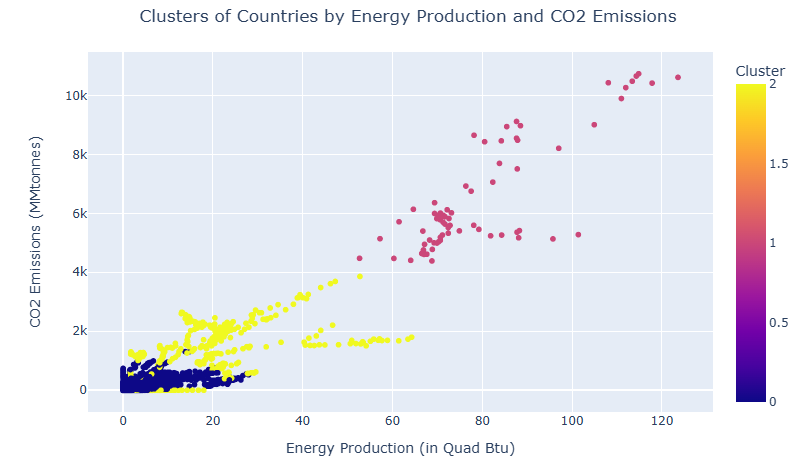


Figure 21: Clusters of Countries by Energy Production and CO2 Emissions

### Time Series Forecasting

Time series forecasting was used to predict future CO2 emissions based on historical data, taking into account trends, seasonality, and other temporal patterns. ARIMA model was used. The model combines autoregressive (AR) terms, integrated (I) terms, and moving average (MA) terms. They are used to understand and predict future values in a time series based on its past values (AR), the difference between values (I), and a combination of past error terms (MA). The ARIMA model parameters p, d, and q are then set to 1, 1, and 0, respectively. The model was then built and used to forecast the total CO2 emissions in the next 10 years. From Figure 22 below, the forecasted CO2 emissions from 2020 to 2029 show a trend of gradual increase over the decade. In 2020, emissions are projected at 69502.96 MMtonnes, rising slowly each year to reach 69982.33 MMtonnes by 2029. The annual increases are modest, with the forecasted emissions in 2021 at 69753.13 MMtonnes and climbing slightly each subsequent year. By 2024, emissions are expected to be 69957.74 MMtonnes, and the rate of increase continues to decelerate, reaching just 69982.33 MMtonnes in 2029.

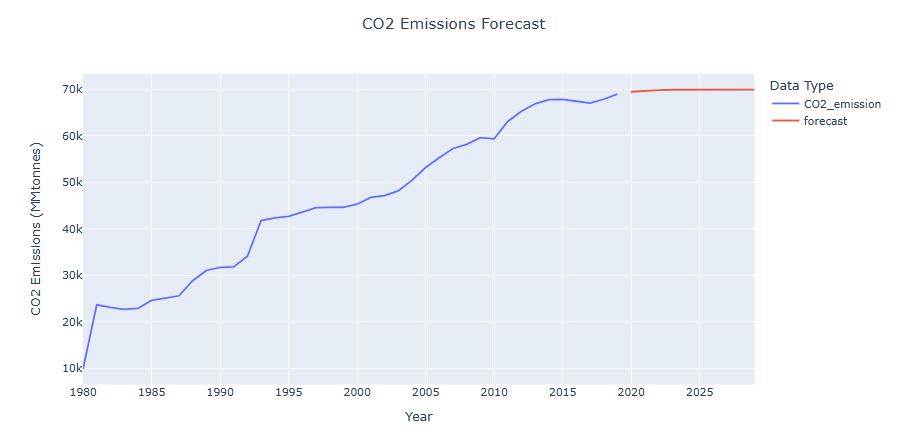


Figure 22: Forecasted CO2 emissions (2020-2029)

The performance of the ARIMA model in forecasting CO2 emissions was also assessed using the MAE, RMSE, R-squared, and Adjusted R-squared. From Table 4 below, the Mean Absolute Error is 481.716 MMtonnes. This shows that the average magnitude of errors in the model's predictions is around 481.716 MMtonnes. This is a moderate level of forecast accuracy. The Root Mean Squared Error is 548.221. The R-squared value is 0.691. This means that the model explains approximately 69.1% of the variance in the observed data which is a reasonably good fit to the data. The Adjusted R-squared is 0.640. This accounts for the number of predictors in the model and slightly reduces the explained variance to 64.0%. Overall, the ARIMA model provides a decent fit and forecast accuracy.

Table 4: ARIMA Model Performance Metrics

|  |  |
| --- | --- |
| **Metric** | **Value (MMtonnes)** |
| MAE | 481.716 |
| RMSE | 548.221 |
| R-squared | 0.691 |
| Adjusted R-squared | 0.640 |

# 5. DISCUSSION OF FINDINGS

Exploratory data analysis, cluster analysis, and predictive modeling were successfully conducted in this study. The analysis revealed a strong positive correlation between energy consumption and CO2 emissions. This aligns with existing literature on the environmental impacts of energy use. The results showed that countries with higher energy consumption generally experience greater CO2 emissions, consistent with findings by (Davis & Socolow, 2014) Davis, who observed that increased energy consumption, particularly from fossil fuels, leads to higher greenhouse gas emissions. The Multiple Linear Regression model further supported this relationship. It indicated that each additional quad Btu of energy consumption correlates with an increase of approximately 69.02 MMtonnes in CO2 emissions. This finding is consistent with the work of (Jeon, 2022), who reported that energy consumption significantly drives CO2 emissions in industrialized nations.

Also, energy production is positively correlated with CO2 emissions, though the effect is less compared to energy consumption. The regression results show that an additional quad Btu of energy production is associated with a 2.07 MMtonnes increase in CO2 emissions. This finding gains support from the results of (Garba & Abdulrahman, 2024), who noted that while energy production contributes to emissions, its impact is moderated by the type of energy sources used. The lower coefficient for energy production compared to consumption shows that energy efficiency and production technologies play an important role in emissions results.

The relationship between GDP and CO2 emissions showed complex dynamics. The scatter plot showed a positive relationship between the two variables while the model produced a negative coefficient. This weak negative coefficient suggests a potential reduction in emissions as GDP increases, these finding contrasts with studies that have found a strong positive relationship between economic growth and CO2 emissions. There is a strong positive relationship between economic growth and CO2 emissions(Onofrei et al., 2022). This divergence may indicate that economic development can indeed lead to enhanced energy efficiency and the adoption of cleaner technologies, which can prevent the growth of emissions. This is consistent with the findings of (Milindi & Inglesi-Lotz, 2023), who reported that higher GDP is often associated with technological advancements that help reduce emissions intensity. These results mean that while economic growth has the potential to increase emissions, it also provides opportunities for investing in greener technologies that can offset this impact.

The small but statistically significant positive coefficient for **population** supports the argument made by Rehman et al. (2021), who emphasize that population growth leads to increased energy demand, which in turn elevates CO2 emissions, but in small increments. This reflects the direct connection between a rising population and the subsequent need for more energy to sustain consumption patterns, even if the per capita emissions increase is marginal. The significant negative coefficient for energy intensity per capita shows that improvements in energy efficiency lead to a decrease in emissions. More efficient energy use per person translates into lower CO2 emissions, as modern technologies and practices optimize energy usage, reducing waste and promoting sustainability.

The distribution of CO2 emissions across low, middle, and high-income countries indicates the disparities in emissions contributions. High-income countries contribute the most to global CO2 emissions. This is consistent with the findings of (Tucker, 1995), who attributed high emissions levels to advanced industrial activities and energy consumption patterns in wealthier nations. Middle and low-income countries also showed significant emissions. This shows their growing industrial activities and energy needs.

The ARIMA model forecasted a gradual increase in CO2 emissions over the next decade, projecting a rise from 69502.96 MMtonnes in 2020 to 69982.33 MMtonnes by 2029. This trend indicates the global trajectory observed by the (International Energy Agency, 2012), which has reported steady increases in emissions due to ongoing industrialization and energy consumption. The model’s forecast provides the urgency of implementing effective climate policies to curb this upward trend. The K-means clustering results revealed three distinct clusters of countries based on energy consumption, production, and CO2 emissions. Cluster 1 comprises countries with low energy consumption and emissions, while Cluster 3 includes those with high levels. This clustering aligns with the findings of (Nematchoua et al., 2021). They identified similar groupings based on energy profiles and emissions.

The knowledge gained from this analysis is consistent with the broader literature, reinforcing the importance of continued research and policy efforts in managing global CO2 emissions and achieving sustainable development goals.

## 5.2 Limitations of the Study

The project design and methods employed faced some limitations. This includes the reliance on publicly available datasets that have had incomplete data, potentially affecting the accuracy of the analysis. Also, the chosen models may not fully capture complex, nonlinear relationships in the data.

# 6. Conclusions

This study explored the dynamic relationships between energy consumption, energy production, economic growth, and CO2 emissions across different countries, to understand how these factors interact and influence each other. The research followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework which encompasses Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The study aimed to answer the question of how energy use and economic activities contribute to CO2 emissions to inform sustainable energy policies. The problem was formulated around the global challenge of balancing economic growth with environmental sustainability. The obtained results demonstrated significant relationships between the variables, with energy consumption, production, and GDP showing a strong influence on CO2 emissions. These findings directly addressed the research questions and met the study's objectives by providing a comprehensive understanding of the dynamics between energy use and CO2 emissions.

This study contributes to both theory and practice. Theoretically, it deepens the understanding of the relationships between economic and environmental factors. Practically, it offers actionable insights for policymakers, suggesting tailored strategies for different countries to achieve sustainable energy use and economic growth. In conclusion, this study effectively addressed the problem, answered the research questions, met the objectives, and made meaningful contributions to both academic research and real-world policy development.

# 7. Recommendations

Despite the comprehensive analysis presented, this study acknowledges several limitations that could be addressed in future research. One of the limitations is the scope of the datasets used, which, while extensive, may not capture all subtle differences in energy consumption, economic growth, and CO2 emissions across different regions and periods. Future investigations could benefit from incorporating more granular and up-to-date data, including regional variations within countries and emerging trends in energy technology and policy. Also, future research should explore the causal relationships between variables using advanced econometric techniques or experimental designs to better understand the direction and strength of these relationships. Longitudinal studies that track changes over longer periods could provide deeper insights into how shifts in policy or economic conditions impact the dynamic relationship between energy use and CO2 emissions.

Based on the findings, several actionable recommendations are made. In policy implementation, governments should develop and implement policies that promote energy efficiency and the adoption of renewable energy sources. These policies could include subsidies for renewable energy technologies, stricter energy efficiency standards for buildings and vehicles, and carbon pricing mechanisms such as carbon taxes or cap-and-trade systems. The implementation of these policies should be prioritized to mitigate the impact of CO2 emissions and support sustainable economic growth. Also, policymakers should design targeted strategies for different regions based on their specific energy profiles and economic conditions. For instance, countries with high energy consumption but low renewable energy adoption could focus on increasing the share of renewable, while others with significant economic growth but high emissions might prioritize energy efficiency improvements.

Encouraging collaboration between governments, industries, and research institutions can drive innovation in energy technologies and sustainable practices. Initiatives that support research and development in energy efficiency and renewable technologies should be promoted, with funding and incentives provided to accelerate advancements in these areas. Moreover, robust mechanisms for monitoring and evaluating the effectiveness of implemented policies should be established. Regular assessments and adjustments based on empirical evidence can ensure that policies remain effective and relevant in addressing energy and environmental challenges. This will foster a more sustainable and economically viable future.

# References

Abdallah, Z. S., Du, L., & Webb, G. I. (2017). Data Preparation. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 318–327). Springer US. https://doi.org/10.1007/978-1-4899-7687-1\_62

Acheampong, A. O., & Boateng, E. B. (2019). Modelling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production*, *225*, 833–856. https://doi.org/10.1016/j.jclepro.2019.03.352

Cai, W., Lai, K., Liu, C., Wei, F., Ma, M., Jia, S., Jiang, Z., & Lv, L. (2019). Promoting sustainability of manufacturing industry through the lean energy-saving and emission-reduction strategy. *Science of The Total Environment*, *665*, 23–32. https://doi.org/10.1016/j.scitotenv.2019.02.069

Chandra Voumik, L., Ridwan, M., Hasanur Rahman, Md., & Raihan, A. (2023). An investigation into the primary causes of carbon dioxide releases in Kenya: Does renewable energy matter to reduce carbon emission? *Renewable Energy Focus*, *47*, 100491. https://doi.org/10.1016/j.ref.2023.100491

Davis, S. J., & Socolow, R. H. (2014). Commitment accounting of CO 2 emissions. *Environmental Research Letters*, *9*(8), 084018. https://doi.org/10.1088/1748-9326/9/8/084018

Dilanchiev, A., Nuta, F., Khan, I., & Khan, H. (2023). Urbanization, renewable energy production, and carbon dioxide emission in BSEC member states: Implications for climate change mitigation and energy markets. *Environmental Science and Pollution Research*, *30*(25), 67338–67350. https://doi.org/10.1007/s11356-023-27221-9

Dinda, S. (2004). Environmental Kuznets Curve Hypothesis: A Survey. *Ecological Economics*, *49*(4), 431–455. https://doi.org/10.1016/j.ecolecon.2004.02.011

Ejuma Martha Adaga, Gold Nmesoma Okorie, Zainab Efe Egieya, Uneku Ikwue, Chioma Ann Udeh, Donald Obinna DaraOjimba, & Osato Itohan Oriekhoe. (2024). THE ROLE OF BIG DATA IN BUSINESS STRATEGY: A CRITICAL REVIEW. *Computer Science & IT Research Journal*, *4*(3), 327–350. https://doi.org/10.51594/csitrj.v4i3.686

Garba, N., & Abdulrahman, B. (2024). Renewable Energy Sources, Sustainability and Environmental Protection: A Review. *European Journal of Theoretical and Applied Sciences*, *2*(2), 449–462. https://doi.org/10.59324/ejtas.2024.2(2).39

Gershon, O., Asafo, J. K., Nyarko-Asomani, A., & Koranteng, E. F. (2024). Investigating the nexus of energy consumption, economic growth and carbon emissions in selected african countries. *Energy Strategy Reviews*, *51*, 101269. https://doi.org/10.1016/j.esr.2023.101269

*Global Energy Review 2020*. (2020).

González-Álvarez, M. A., & Montañés, A. (2023). CO2 emissions, energy consumption, and economic growth: Determining the stability of the 3E relationship. *Economic Modelling*, *121*, 106195. https://doi.org/10.1016/j.econmod.2023.106195

Ho Yu, C. (2010). Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research*, *3*(1), 9–22. https://doi.org/10.21500/20112084.819

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Huang, Z., Tonooka, Y., Sekiguchi, K., Wang, Q., & Sakamoto, K. (2009). Long-term Sulfur Emissions and Environmental Kuznets Curves: Comparison and Implications. *Asian Journal of Atmospheric Environment*, *3*(1), 19–26. https://doi.org/10.5572/ajae.2009.3.1.019

Hussin, F., Md Rahim, S. A. N., Hatta, N. S. M., Aroua, M. K., & Mazari, S. A. (2023). A systematic review of machine learning approaches in carbon capture applications. *Journal of CO2 Utilization*, *71*, 102474. https://doi.org/10.1016/j.jcou.2023.102474

International Energy Agency. (2012). *CO2 Emissions from Fuel Combustion 2012*. OECD. https://doi.org/10.1787/co2\_fuel-2012-en

Jeon, H. (2022). CO2 emissions, renewable energy and economic growth in the US. *The Electricity Journal*, *35*(7), 107170. https://doi.org/10.1016/j.tej.2022.107170

Jin, Y., Sharifi, A., Li, Z., Chen, S., Zeng, S., & Zhao, S. (2024). Carbon emission prediction models: A review. *Science of The Total Environment*, *927*, 172319. https://doi.org/10.1016/j.scitotenv.2024.172319

Lai, Y., & Dzombak, D. A. (2020). Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation. *Weather and Forecasting*, *35*(3), 959–976. https://doi.org/10.1175/WAF-D-19-0158.1

Linares, P., & Labandeira, X. (2010). ENERGY EFFICIENCY: ECONOMICS AND POLICY. *Journal of Economic Surveys*, *24*(3), 573–592. https://doi.org/10.1111/j.1467-6419.2009.00609.x

Milindi, C. B., & Inglesi-Lotz, R. (2023). The relationship between technology and emissions: Evidence from different income level countries and economic sectors. *Energy Reports*, *10*, 2900–2916. https://doi.org/10.1016/j.egyr.2023.09.094

Nahar, S. (2024). Modeling the effects of artificial intelligence (AI)-based innovation on sustainable development goals (SDGs): Applying a system dynamics perspective in a cross-country setting. *Technological Forecasting and Social Change*, *201*, 123203. https://doi.org/10.1016/j.techfore.2023.123203

Nematchoua, M. K., Sadeghi, M., & Reiter, S. (2021). Strategies and scenarios to reduce energy consumption and CO2 emission in the urban, rural and sustainable neighbourhoods. *Sustainable Cities and Society*, *72*, 103053. https://doi.org/10.1016/j.scs.2021.103053

Onofrei, M., Vatamanu, A. F., & Cigu, E. (2022). The Relationship Between Economic Growth and CO2 Emissions in EU Countries: A Cointegration Analysis. *Frontiers in Environmental Science*, *10*, 934885. https://doi.org/10.3389/fenvs.2022.934885

Osobajo, O. A., Otitoju, A., Otitoju, M. A., & Oke, A. (2020). The Impact of Energy Consumption and Economic Growth on Carbon Dioxide Emissions. *Sustainability*, *12*(19), 7965. https://doi.org/10.3390/su12197965

Padgett, C. R., Skilbeck, C. E., & Summers, M. J. (2014). Missing Data: The Importance and Impact of Missing Data from Clinical Research. *Brain Impairment*, *15*(1), 1–9. https://doi.org/10.1017/BrImp.2014.2

Picano, E., Mangia, C., & D’Andrea, A. (2022). Climate Change, Carbon Dioxide Emissions, and Medical Imaging Contribution. *Journal of Clinical Medicine*, *12*(1), 215. https://doi.org/10.3390/jcm12010215

Rahman, A., & Hasan, M. M. (2017). Modeling and Forecasting of Carbon Dioxide Emissions in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) Models. *Open Journal of Statistics*, *07*(04), 560–566. https://doi.org/10.4236/ojs.2017.74038

Rahman, M. M., Alam, K., & Velayutham, E. (2022). Reduction of CO2 emissions: The role of renewable energy, technological innovation and export quality. *Energy Reports*, *8*, 2793–2805. https://doi.org/10.1016/j.egyr.2022.01.200

Ratnadip Adhikari & R. K. Agrawal. (2013). *An Introductory Study on Time series Modeling and Forecasting*. LAP Lambert Academic Publishing. https://doi.org/10.13140/2.1.2771.8084

Rehman, A., Ma, H., Ahmad, M., Irfan, M., Traore, O., & Chandio, A. A. (2021). Towards environmental Sustainability: Devolving the influence of carbon dioxide emission to population growth, climate change, Forestry, livestock and crops production in Pakistan. *Ecological Indicators*, *125*, 107460. https://doi.org/10.1016/j.ecolind.2021.107460

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Stella Emeka-Okoli, Christiana Adanma Otonnah, Tochukwu Chinwuba Nwankwo, & Ekene Ezinwa Nwankwo. (2024). REVIEW OF CARBON PRICING MECHANISMS: EFFECTIVENESS AND POLICY IMPLICATIONS. *International Journal of Applied Research in Social Sciences*, *6*(3), 337–347. https://doi.org/10.51594/ijarss.v6i3.891

Stern, D. I. (2004). The Rise and Fall of the Environmental Kuznets Curve. *World Development*, *32*(8), 1419–1439. https://doi.org/10.1016/j.worlddev.2004.03.004

Stern, D. I. (2018). The Environmental Kuznets Curve ☆. In *Reference Module in Earth Systems and Environmental Sciences* (p. B9780124095489092782). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09278-2

Taylor, J. (2018). CRISP-DM & Decision Management. *Decision Management*.

Tucker, M. (1995). Carbon dioxide emissions and global GDP. *Ecological Economics*, *15*(3), 215–223. https://doi.org/10.1016/0921-8009(95)00045-3

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

World Bank. (2020). *Global Economic Prospects, June 2020*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1553-9

Zhao, Y., & Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research. *Journal of Physics: Conference Series*, *1873*(1), 012074. https://doi.org/10.1088/1742-6596/1873/1/012074

Zhou, L. (2023). An Introduction to Data Visualization. *Highlights in Science, Engineering and Technology*, *31*, 60–63. https://doi.org/10.54097/hset.v31i.4813

Acheampong, A. O., & Boateng, E. B. (2019). Modeling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production*, *225*, 833–856. https://doi.org/10.1016/j.jclepro.2019.03.352

Cai, W., Lai, K., Liu, C., Wei, F., Ma, M., Jia, S., Jiang, Z., & Lv, L. (2019). Promoting sustainability of the manufacturing industry through the lean energy-saving and emission-reduction strategy. *Science of The Total Environment*, *665*, 23–32. https://doi.org/10.1016/j.scitotenv.2019.02.069

Chandra Voumik, L., Ridwan, M., Hasanur Rahman, Md., & Raihan, A. (2023). An investigation into the primary causes of carbon dioxide releases in Kenya: Does renewable energy matter to reduce carbon emission? *Renewable Energy Focus*, *47*, 100491. https://doi.org/10.1016/j.ref.2023.100491

Davis, S. J., & Socolow, R. H. (2014). Commitment accounting of CO2 emissions. *Environmental Research Letters*, *9*(8), 084018. https://doi.org/10.1088/1748-9326/9/8/084018

Dilanchiev, A., Nuta, F., Khan, I., & Khan, H. (2023). Urbanization, renewable energy production, and carbon dioxide emission in BSEC member states: Implications for climate change mitigation and energy markets. *Environmental Science and Pollution Research*, *30*(25), 67338–67350. https://doi.org/10.1007/s11356-023-27221-9

Dinda, S. (2004). Environmental Kuznets Curve Hypothesis: A Survey. *Ecological Economics*, *49*(4), 431–455. https://doi.org/10.1016/j.ecolecon.2004.02.011

Ejuma Martha Adaga, Gold Nmesoma Okorie, Zainab Efe Egieya, Uneku Ikwue, Chioma Ann Udeh, Donald Obinna DaraOjimba, & Osato Itohan Oriekhoe. (2024). THE ROLE OF BIG DATA IN BUSINESS STRATEGY: A CRITICAL REVIEW. *Computer Science & IT Research Journal*, *4*(3), 327–350. https://doi.org/10.51594/csitrj.v4i3.686

Garba, N., & Abdulrahman, B. (2024). Renewable Energy Sources, Sustainability, and Environmental Protection: A Review. *European Journal of Theoretical and Applied Sciences*, *2*(2), 449–462. https://doi.org/10.59324/ejtas.2024.2(2).39

Gershon, O., Asafo, J. K., Nyarko-Asomani, A., & Koranteng, E. F. (2024). Investigating the nexus of energy consumption, economic growth, and carbon emissions in selected African countries. *Energy Strategy Reviews*, *51*, 101269. https://doi.org/10.1016/j.esr.2023.101269

*Global Energy Review 2020*. (2020).

González-Álvarez, M. A., & Montañés, A. (2023). CO2 emissions, energy consumption, and economic growth: Determining the stability of the 3E relationship. *Economic Modelling*, *121*, 106195. https://doi.org/10.1016/j.econmod.2023.106195

Ho Yu, C. (2010). Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research*, *3*(1), 9–22. https://doi.org/10.21500/20112084.819

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Huang, Z., Tonooka, Y., Sekiguchi, K., Wang, Q., & Sakamoto, K. (2009). Long-term Sulfur Emissions and Environmental Kuznets Curves: Comparison and Implications. *Asian Journal of Atmospheric Environment*, *3*(1), 19–26. https://doi.org/10.5572/ajae.2009.3.1.019

Hussin, F., Md Rahim, S. A. N., Hatta, N. S. M., Aroua, M. K., & Mazari, S. A. (2023). A systematic review of machine learning approaches in carbon capture applications. *Journal of CO2 Utilization*, *71*, 102474. https://doi.org/10.1016/j.jcou.2023.102474

International Energy Agency. (2012). *CO2 Emissions from Fuel Combustion 2012*. OECD. https://doi.org/10.1787/co2\_fuel-2012-en

Jeon, H. (2022). CO2 emissions, renewable energy, and economic growth in the US. *The Electricity Journal*, *35*(7), 107170. https://doi.org/10.1016/j.tej.2022.107170

Jin, Y., Sharifi, A., Li, Z., Chen, S., Zeng, S., & Zhao, S. (2024). Carbon emission prediction models: A review. *Science of The Total Environment*, *927*, 172319. https://doi.org/10.1016/j.scitotenv.2024.172319

Lai, Y., & Dzombak, D. A. (2020). Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation. *Weather and Forecasting*, *35*(3), 959–976. https://doi.org/10.1175/WAF-D-19-0158.1

Linares, P., & Labandeira, X. (2010). Energy Efficiency: Economics And Policy. *Journal of Economic Surveys*, *24*(3), 573–592. https://doi.org/10.1111/j.1467-6419.2009.00609.x

Milindi, C. B., & Inglesi-Lotz, R. (2023). The relationship between technology and emissions: Evidence from different income level countries and economic sectors. *Energy Reports*, *10*, 2900–2916. https://doi.org/10.1016/j.egyr.2023.09.094

Nahar, S. (2024). Modeling the effects of artificial intelligence (AI)-based innovation on sustainable development goals (SDGs): Applying a system dynamics perspective in a cross-country setting. *Technological Forecasting and Social Change*, *201*, 123203. https://doi.org/10.1016/j.techfore.2023.123203

Nematchoua, M. K., Sadeghi, M., & Reiter, S. (2021). Strategies and scenarios to reduce energy consumption and CO2 emission in urban, rural, and sustainable neighborhoods. *Sustainable Cities and Society*, *72*, 103053. https://doi.org/10.1016/j.scs.2021.103053

Onofrei, M., Vatamanu, A. F., & Cigu, E. (2022). The Relationship Between Economic Growth and CO2 Emissions in EU Countries: A Cointegration Analysis. *Frontiers in Environmental Science*, *10*, 934885. https://doi.org/10.3389/fenvs.2022.934885

Osobajo, O. A., Otitoju, A., Otitoju, M. A., & Oke, A. (2020). The Impact of Energy Consumption and Economic Growth on Carbon Dioxide Emissions. *Sustainability*, *12*(19), 7965. https://doi.org/10.3390/su12197965

Padgett, C. R., Skilbeck, C. E., & Summers, M. J. (2014). Missing Data: The Importance and Impact of Missing Data from Clinical Research. *Brain Impairment*, *15*(1), 1–9. https://doi.org/10.1017/BrImp.2014.2

Picano, E., Mangia, C., & D’Andrea, A. (2022). Climate Change, Carbon Dioxide Emissions, and Medical Imaging Contribution. *Journal of Clinical Medicine*, *12*(1), 215. https://doi.org/10.3390/jcm12010215

Rahman, A., & Hasan, M. M. (2017). Modeling and Forecasting of Carbon Dioxide Emissions in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) Models. *Open Journal of Statistics*, *07*(04), 560–566. https://doi.org/10.4236/ojs.2017.74038

Rahman, M. M., Alam, K., & Velayutham, E. (2022). Reduction of CO2 emissions: The role of renewable energy, technological innovation, and export quality. *Energy Reports*, *8*, 2793–2805. https://doi.org/10.1016/j.egyr.2022.01.200

Ratnadip Adhikari & R. K. Agrawal. (2013). *An Introductory Study on Time series Modeling and Forecasting*. LAP Lambert Academic Publishing. https://doi.org/10.13140/2.1.2771.8084

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Stella Emeka-Okoli, Christiana Adanma Otonnah, Tochukwu Chinwuba Nwankwo, & Ekene Ezinwa Nwankwo. (2024). REVIEW OF CARBON PRICING MECHANISMS: EFFECTIVENESS AND POLICY IMPLICATIONS. *International Journal of Applied Research in Social Sciences*, *6*(3), 337–347. https://doi.org/10.51594/ijarss.v6i3.891

Stern, D. I. (2004). The Rise and Fall of the Environmental Kuznets Curve. *World Development*, *32*(8), 1419–1439. https://doi.org/10.1016/j.worlddev.2004.03.004

Stern, D. I. (2018). The Environmental Kuznets Curve ☆. In *Reference Module in Earth Systems and Environmental Sciences* (p. B9780124095489092782). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09278-2

Taylor, J. (2018). CRISP-DM & Decision Management. *Decision Management*.

Tucker, M. (1995). Carbon dioxide emissions and global GDP. *Ecological Economics*, *15*(3), 215–223. https://doi.org/10.1016/0921-8009(95)00045-3

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

World Bank. (2020). *Global Economic Prospects, June 2020*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1553-9

Zhao, Y., & Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research. *Journal of Physics: Conference Series*, *1873*(1), 012074. https://doi.org/10.1088/1742-6596/1873/1/012074

Zhou, L. (2023). An Introduction to Data Visualization. *Highlights in Science, Engineering and Technology*, *31*, 60–63. https://doi.org/10.54097/hset.v31i.4813

# Appendix 2: Python Codes

