

# Exploring Global Energy Transition Trends

Dylan Kayyem  
Applied Computer Science  
University of Colorado  
Boulder, CO  
dyga6971@colorado.edu

## ABSTRACT

In the era of sustainable energy emphasis, understanding the intricacies of global energy consumption is vital. This research dives deep into the World Energy Consumption data set, shedding light on energy trends, renewable adoption rates, fossil fuel dependencies, and the correlation between energy consumption and economic growth. The goal is to provide invaluable insights for policymakers, researchers, and the global populace, helping to shape a more sustainable future. Drawing from numerous esteemed data repositories and publications, the research aims to not only echo the findings of previous studies but to add unique perspectives and insights to the global discourse on energy sustainability.

## 1 PROBLEM STATEMENT AND MOTIVATION

With the global focus shifting towards sustainable energy consumption, understanding the dynamics of global energy consumption and production becomes paramount. This project embarks on an exploratory journey into the *World Energy Consumption* (Poudel, 2020)<sup>1</sup> data set, aiming to uncover the profound shifts and trends that define our energy landscape. We intend to delve into specific facets such as renewable energy adoption rates, fossil fuel dependency, the relationship between per capita energy consumption and economic growth, and more. The motivation behind this endeavor is to offer key insights into environmental sustainability, economic ramifications of energy patterns, and the overarching energy transition narrative that unfolds worldwide.

Through this analysis, we hope to provide actionable insights for policymakers, researchers, and the global community at large. We aim to uncover insights into renewable energy adoption trends, fossil fuel dependency shifts, correlations between per capita energy consumption and GDP, carbon intensity of electricity production, and growth in hydropower and wind energy sectors.

## 2 PREVIOUS WORK

A plethora of work has been undertaken in the energy domain, ranging from comprehensive reports by giants like

the *International Energy Agency (IEA)* and the *U.S. Energy Information Administration (EIA)* to intensive academic research in fields such as energy economics. Additionally, governmental policy studies and data repositories maintained by renowned organizations, including the World Bank and NASA, have enriched the discourse on energy. These include forward-looking projections, pivotal policy suggestions, and incisive analysis.

### 2.1 LITERATURE SURVEY

IEA's renowned publication, the *World Energy Outlook* (IEA, 2021)<sup>2</sup>, paints a comprehensive picture, envisioning how the global energy landscape might shift in the upcoming decades. Parallely, the World Energy Council, through its *World Energy Trilemma Index* (World Energy Council, 2019)<sup>3</sup>, presents an evaluative measure of countries' energy performance, founded on the pillars of equity, security, and environmental fidelity. Notably, from a U.S. perspective, EIA's *Annual Energy Outlook* (U.S. EIA, 2023)<sup>4</sup> delineates projections for domestic energy markets until 2050, anchored by existing legislative frameworks.

### 2.2 ACADEMIC RESEARCH

The realm of energy has been an academic hotspot, marked by extensive studies spanning energy economics, environmental science, and sustainability—all showcased across various esteemed, peer-reviewed platforms.

In the sphere of Energy Economics, pivotal studies have been pursued, such as the exploration into the interplay between energy pricing and the economic growth trajectories across nations (Elsevier, 2023)<sup>5</sup>. From an Environmental Science perspective, research endeavors have delved into discerning the environmental ramifications of distinct energy origins (Zimmerman, 2022)<sup>6</sup>.

Furthermore, in the domain of Sustainability, scholarly contributions have shed light on the instrumental role of renewables in the quest to realize global sustainability objectives (MDPI, 2023)<sup>7</sup>.

## 2.3 GOVERNMENT AND POLICY STUDIES

To architect sound energy strategies, a gamut of governments and affiliated institutions have undertaken rigorous analyses, teasing out both economic and ecological outcomes linked to energy-oriented decisions. A case in point is the *European Commission's* delineated roadmap aiming to hit precise energy and climatic milestones by 2030 (European Commission, 2023)<sup>8</sup>.

## 2.4 DATA REPOSITORIES

The quest for data-rich repositories finds answers in esteemed entities like *The World Bank*, *NASA*, and the *U.S. National Renewable Energy Laboratory (NREL)*. These repositories offer a rich tapestry of datasets that fuel multifaceted research endeavors.

*The World Bank*, for instance, unveils datasets that delve into the intricacies of global energy dynamics—spanning production, consumption, and even intricate trade details (The World Bank, 2023)<sup>9</sup>. *NASA* brings to the table invaluable satellite-captured intel on solar radiation, becoming a cornerstone for solar energy-centric research (NASA, 2023)<sup>10</sup>. Meanwhile, the *U.S. National Renewable Energy Laboratory (NREL)* serves as a beacon for wind energy studies in the U.S., offering datasets that pinpoint wind potential, guiding turbine placements (NREL, 2023)<sup>11</sup>.

## 3 PROPOSED WORK

Our study aims to gain a deeper understanding of global energy trends through a systematic methodology. The primary source of our research will be the *World Energy Consumption* dataset sourced from Kaggle (Poudel, 2020)<sup>1</sup>, which is currently archived on Dylan Kayyem's local storage.

Transitioning from the collection phase to ensuring analytical integrity, we embark on an exhaustive data quality examination. This involves the intricate task of diagnosing and amending data discrepancies, anomalies, and missing entries. For a more specific focus, columns such as *biofuel\_cons\_change\_pct* (Poudel, 2020)<sup>1</sup> will undergo a detailed scrutiny to identify any data irregularities. Moving forward, an essential stage in our methodology is the preprocessing of this data. Central to this is the transformation phase where we ensure that every column in the dataset aligns with the desired format. This includes tasks like data type modifications, feature crafting - for instance, extrapolating the renewable energy consumption

percentage, and the normalization or scaling of various attributes.

Subsequently, our study adopts an analytical approach that is both versatile and integrative. The strategy revolves around aggregating data across different granularities, further enriched by potentially supplementing it with auxiliary datasets. This process ensures a comprehensive, layered perspective on the data at hand. Drawing a comparative analysis with extant literature, it's evident that while myriad organizations, academicians, and governmental entities have ventured into the realm of energy transition studies, our research carries distinctive nuances. Our preference for the *World Energy Consumption* data set from Kaggle offers a unique vantage point, potentially contrasting from popular datasets such as the BP Statistical Review of World Energy or the SHIFT Data Portal (Poudel, 2020)<sup>1</sup>. This choice promises an encompassing view of global energy dynamics.

We aim to unearth underlying patterns and correlations, for instance, the relationship between individual energy consumption and GDP or the nuances of carbon footprints. Given the transitory nature of energy paradigms, our study consciously anchors itself in the most recent datasets. This approach ensures that our insights are both timely and reflective of the current global energy landscape. Additionally, our research underscores the importance of data reshaping and integration. By merging our primary dataset with external sources, we anticipate deriving nuanced insights, potentially overlooked in preceding works. With a detailed documentation process in place, we aspire to ensure our methods are lucid, replicable, and serve as a reference point for future academic pursuits in this domain.

To ensure robust testing of our hypotheses, we will employ a multifaceted analytical framework that includes correlation matrices, trend analysis, and regression models. The correlation matrix will be utilized to assess the strength and direction of the relationship between renewable energy consumption, GDP per capita, and population. Trend analysis will be conducted to identify shifts in renewable energy consumption patterns across different countries. Furthermore, regression models will be applied to predict renewable energy consumption, incorporating a RandomForestRegressor to handle complex, non-linear data structures. This comprehensive approach will ensure that our insights are grounded in a rigorous, data-driven methodology.

## 4 DATA SET

Before delving into the research, it's imperative to have the appropriate dataset. Our chosen primary dataset, named *World Energy Consumption*, has been retrieved from Kaggle (Poudel, 2020)<sup>1</sup>. This dataset provides a comprehensive overview of energy consumption and production metrics for individual countries over various years. The data incorporates numerous facets such as primary energy, the composition of the energy mix, energy intensity, electricity consumption, and a myriad of other variables. Notably, these metrics are collected from diverse, reputable sources including the *BP Statistical Review of World Energy* (BP, 2022)<sup>12</sup>, *The SHIFT Data Portal* (The Shift Project, 2020)<sup>13</sup>, and *the Maddison Project Database*. The dataset can be found at this URL <https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption/data><sup>1</sup>.

Diving into the specifics of the dataset, one of its comprehensive areas covers primary energy consumption, its mix, and energy intensity. Such information is invaluable in assessing the energy consumption trends of countries and discerning the sources of their energy. Progressing from this, the dataset also sheds light on electricity consumption patterns and the composition of these energy sources. This data facilitates an analysis of the methods countries employ in producing electricity—be it from renewable energy, fossil fuels, or alternative sources.

An equally significant aspect is its detailing of annual variations in electricity generation. Specifically, the dataset differentiates between renewable avenues, such as wind and solar, and non-renewable means, offering a clear view of how countries' energy sources are evolving over time. Beyond just the numbers, the dataset acknowledges the broader implications of energy consumption. This is evident from columns like *greenhouse\_gas\_emissions*, which provide insight into the environmental repercussions tied to energy production and consumption (Poudel, 2020)<sup>1</sup>. Furthermore, the richness of this dataset isn't limited to just energy specifics. It also encompasses additional data points from esteemed entities like the United Nations, World Bank, Gapminder, and the Maddison Project Database. These markers offer demographic data, economic indicators, and other potential influencers of energy consumption.

The World Energy Consumption dataset, which forms the basis of our analysis, includes detailed records of renewable energy consumption trends, GDP per capita, and population data. These variables were central to our correlation and regression analyses, allowing us to uncover

patterns and draw insights into the energy transition narrative.

Concluding the overview of the data sources, it's worth noting that the energy consumption metrics derive primarily from the *BP Statistical Review of World Energy* (BP, 2022)<sup>12</sup> and *the SHIFT Data Portal* (The Shift Project, 2020)<sup>13</sup>. In parallel, metrics related to electricity consumption trace back to the *BP Statistical Review of World Energy* and *the EMBER – Global Electricity Dashboard* (Poudel, 2020)<sup>1</sup>.

## 5 EVALUATION METHODS

To effectively assess the *World Energy Consumption* dataset, it's imperative to utilize a blend of robust evaluation methods. These methods are meticulously chosen to offer a holistic and in-depth analysis of the data at hand. Beginning with foundational metrics, we opt for descriptive statistics as they provide insights into the core tendencies and fluctuations within the data. This method equips us with an understanding of the data's distribution, enabling us to discern central values and variations inherent to it.

Building on this foundation, we will incorporate data visualization tools. The rationale behind this choice is simple yet profound: the graphical representation of data can often unveil trends or anomalies that might remain obscured in a tabular format. By translating data into visual formats, we aim to provide a more intuitive understanding of the prevailing patterns.

Venturing into a deeper analytical terrain, correlation and regression analyses take center stage. While correlation analysis will spotlight potential relationships between variables, shedding light on how distinct data points might interplay, regression analysis will serve to model these relationships. The goal here is not just to identify associations, but to quantify and model them, offering predictions and insights into potential causal factors.

Our evaluation methods incorporate a detailed correlation matrix to identify the strength and direction of the relationship between renewable energy consumption, GDP per capita, and population. Additionally, we apply regression analysis using the *RandomForestRegressor* model, which has been rigorously tested against actual renewable energy consumption figures to ensure the reliability of our predictions.

Finally, we get to the step of data validation. Given the diverse sources of our dataset and the potential for discrepancies, it's paramount to ensure its accuracy. To

achieve this, we will validate our data against trusted external sources. This step underscores our commitment to rigorous research and reinforces the authenticity and reliability of our findings.

## 6 TOOLS

In our quest to effectively evaluate the *World Energy Consumption* dataset, the selection of appropriate tools is paramount. The tools we've chosen not only resonate with the nature of our dataset but also streamline our analytical workflow, ensuring precision and efficiency. At the very foundation of our analytical pipeline is data cleaning and preprocessing. For these foundational tasks, we primarily turn to the power of Python, bolstered by its renowned libraries: Pandas and NumPy. These tools offer robust functionalities that facilitate data wrangling, ensuring that our dataset is free from inconsistencies and is structured for subsequent analyses.

Transitioning from this preliminary stage, we delve into the heart of our evaluation: data integration and analysis. Here, visualization plays a pivotal role. To graphically represent our data in the most insightful ways, we have chosen a triad of libraries: Matplotlib, Seaborn, and Plotly. Each of these tools brings its unique strengths to the table, allowing us to paint a comprehensive picture of the prevailing energy consumption trends.

While our primary focus is on Python-based tools, we remain adaptable. Recognizing the potential complexities that might arise during our analysis, we are open to leveraging SQL databases, if deemed necessary. This ensures not only versatility in our analysis but also scalability, as SQL can efficiently manage larger datasets and more intricate operations.

## 7 MILESTONES

A structured timeline is vital for systematic research progression and timely completion. We've outlined key milestones for each phase. During Weeks 1 and 2, we'll focus on data acquisition, including its cleaning and preprocessing. The quality of this data will influence later analysis. In Weeks 3 and 4, our attention shifts to data integration and creating preliminary visualizations to highlight dataset trends. Weeks 5 and 6 will involve a detailed data analysis using regression modeling and comparative studies to identify patterns and make initial deductions. Finally, Weeks 7 and 8 are reserved for refining our findings, improving visualizations, and documenting our

work, ensuring it's replicable and beneficial for future research.

### 7.1 MILESTONES COMPLETED

In this research, we have successfully achieved several key milestones, essential for the advancement of our study on energy consumption. Initially, we focused on data cleaning and preparation, where we meticulously filled missing values with either zero or group medians, depending on the dataset's characteristics. This step was crucial to ensure data consistency, enabling us to align various parameters across different datasets for accurate subsequent analysis.

We have completed a comprehensive correlation analysis, revealing strong positive correlations between renewable energy consumption and population, and a weaker positive correlation with GDP per capita. Our regression analysis, conducted with a RandomForestRegressor model, has yielded an R-squared value of 0.745 and an MSE of 2199933.70, indicating a model that fits the data relatively well.

Further, we laid the groundwork for a regression analysis aimed at predicting renewable energy consumption. This entailed selecting relevant features and target variables, and meticulously preparing the data for the regression model. Handling missing values and splitting the dataset into training and testing sets were critical steps in this process.

Additionally, we conducted extensive additional data analyses, including correlation analyses focusing on greenhouse gas emissions and various types of energy consumption. We also analyzed trends in carbon intensity and the share of low-carbon and renewable energy sources, providing a comprehensive view of the energy landscape over time.

### 7.2 MILESTONES TODO

As we progress with our research, there are several critical analyses that we aim to conduct to deepen our understanding of energy consumption dynamics. A pivotal area of focus will be the detailed examination of annual variations in electricity generation. The dataset offers a unique opportunity to differentiate between renewable energy sources, such as wind and solar, and non-renewable means.

Analyzing this aspect will enable us to paint a clearer picture of how countries' energy sources are evolving over time. This analysis is not only significant for understanding the current energy landscape but also for forecasting future trends in energy generation and consumption.

In addition to this, we recognize the need to refine our year-on-year analysis of renewable energy consumption. The current analysis resulted in an R-squared value of 0.48, which, while indicative of some level of model fit, also highlights the potential for substantial improvement. We plan to refine our regression model, aiming to improve upon the current R-squared value. This may involve incorporating additional variables, tweaking the model parameters, or exploring other statistical models. Our goal is to enhance the predictive accuracy and reliability of our findings on renewable energy consumption trends.

Lastly, an essential upcoming milestone is the validation of our dataset. Data validation is crucial to ensure the accuracy and reliability of our findings. This process will involve rigorous checks for data integrity, consistency, and accuracy. By doing so, we aim to strengthen the foundation of our analysis and ensure that our conclusions are based on solid, verifiable data. This step is not only a best practice in data analysis but also a necessity to maintain the credibility and validity of our research outcomes.

8 RESULTS SO FAR

Our preliminary findings from the completed milestones offer insightful revelations into the patterns of energy consumption. The correlation analysis has been particularly revealing, showing a strong positive correlation of 0.8 between renewable energy consumption and population.

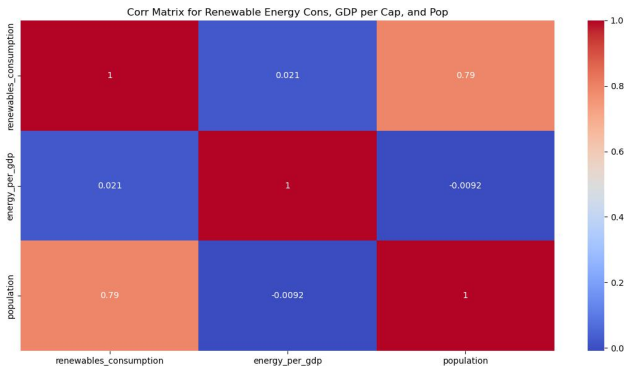


Figure 1: Correlation Matrix for Renewable Energy Consumption, GDP per Capita, and Population.

This suggests that countries with larger populations tend to consume more renewable energy. However, the analysis also indicated very weak correlations between renewable energy consumption and energy\_per\_gdp (0.022), and between population and energy\_per\_gdp (-0.0092), hinting at the complex dynamics of economic factors in energy consumption patterns (Figure 1).

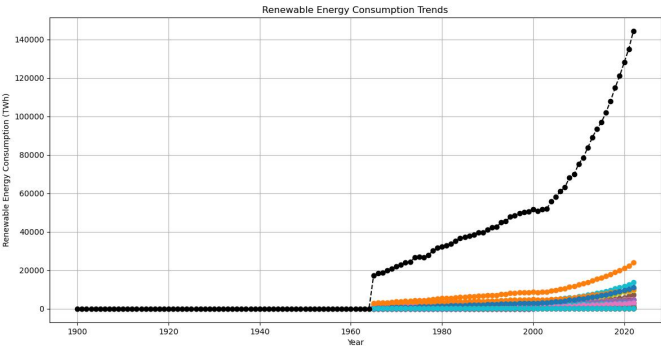


Figure 2: Renewable Energy Consumption Trends Grouped by Country.

The year-on-year analysis of renewable energy consumption uncovered a significant increase in renewable energy adoption in recent years. However, the R-squared value of 0.48 and the MSE of 450284.54 from our model indicate that there are still areas for improvement in our predictive capabilities (Figure 2).

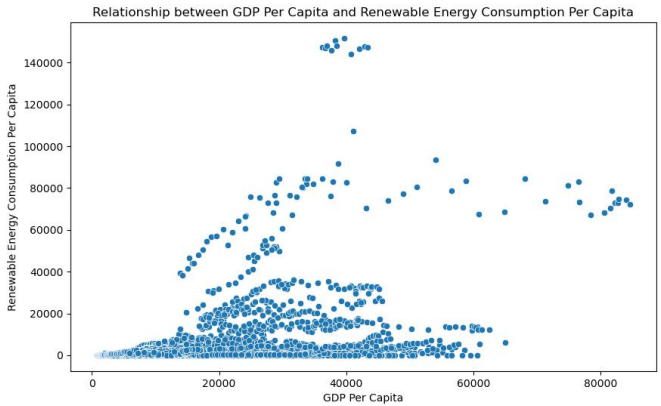
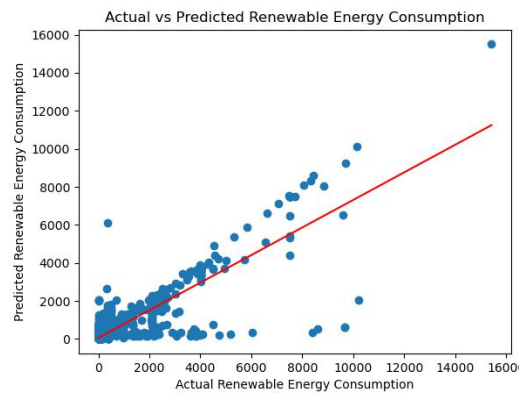


Figure 3: Relationship between GDP Per Capita and Renewable Energy Consumption Per Capita.

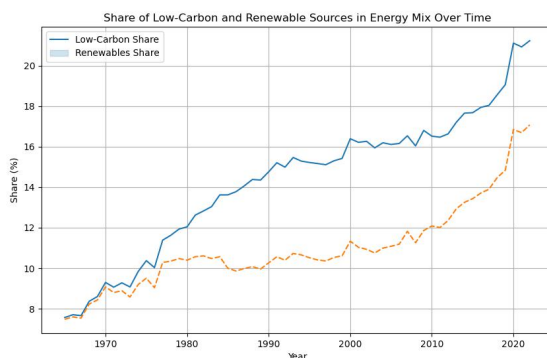
Our examination of the relationship between GDP per capita and renewable energy consumption per capita revealed a weak positive correlation of 0.398641289199194 (Figure 3). This implies that while there is some association between higher GDP per capita and increased renewable energy consumption, the relationship is not strongly predictive.





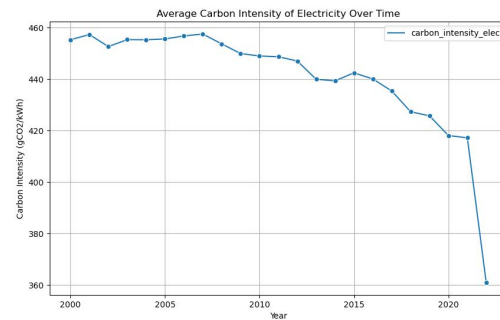
**Figure 4: Actual vs Predicted Renewable Energy Consumption Regression Analysis.**

The regression analysis, utilizing a RandomForestRegressor model, demonstrated an R-squared value of 0.745 and an MSE of 2199933.70 (Figure 4). These results suggest that the model fits the data relatively well, as also supported by the visualizations comparing actual versus predicted renewable energy consumption.



**Figure 5: Share of Low-Carbon and Renewable Sources in Energy Mix Over Time.**

Lastly, our additional analyses yielded low correlation coefficients between greenhouse gas emissions and types of energy consumption (Figure 5). The visualized trends in carbon intensity and renewable energy shares over time (Figure 6) provided a nuanced understanding of the changing landscape of energy consumption and its environmental implications.



**Figure 6: Average Carbon Intensity of Electricity Over Time.**

## REFERENCES

- [1] Poudel, P. (2020, September 9). *World Energy Consumption*. Consumption of energy by different countries. Retrieved October 30, 2020, from: <https://www.kaggle.com/datasets/pralabhoudel/world-energy-consumption/data>
- [2] IEA (2021), *World Energy Outlook 2021*, IEA, Paris <https://www.iea.org/reports/world-energy-outlook-2021>
- [3] World Energy Council (2019). *World Energy Trilemma Index | 2019*. Monitoring the sustainability of national energy systems. <https://www.worldenergy.org/publications/entry/world-energy-trilemma-index-2019>
- [4] U.S. Energy Information Administration (2023, March 16). *Annual Energy Outlook 2023*. Annual Energy Outlook Narrative. <https://www.eia.gov/outlooks/aeo/narrative/#casesdescriptors>
- [5] Elsevier (2023). *Energy Economics*. <https://www.sciencedirect.com/journal/energy-economics>
- [6] Zimmerman, J. (2022). *Environmental Science & Technology Journal*. An ACS Transformative Journal. <https://pubs.acs.org/journal/esthag>
- [7] MDPI. (2023). *Sustainability*. Monitoring the Performance of Sustainable Development Goals in the Italian Regions. <https://www.mdpi.com/journal/sustainability>
- [8] European Commission. (2023). Energy Portal. *European Commission's Energy Framework*. [https://energy.ec.europa.eu/index\\_en](https://energy.ec.europa.eu/index_en)
- [9] The World Bank. (2023). *Energy and Mining The World Bank*. <https://data.worldbank.org/topic/energy-and-mining>
- [10] NASA. (2023). *Earth Data*. <https://www.earthdata.nasa.gov/>
- [11] National Renewable Energy Laboratory. (2023). *Geospatial Data Science. NREL*. <https://www.nrel.gov/gis/data.html>
- [12] BP p.l.c. (2022). *Statistical Review of World Energy*. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>
- [13] The Shift Project. (2020). *The Shift Data Portal: Energy*. The Shift Project. <https://www.theshiftdataportal.org/energy>