Economic-Environmental determinants of Quality of Life: A Predictive Analytics Approach for Sustainable Development

Table of Contents

Abstract

In this project we will take a look at different indicators and will review the impact of the country's development and quality of life. We will take a look at GDP and foreign direct investment's impact on life expectancy. A cluster analysis on tech imports and different goods. We will analyze renewable energy adoptions and other energy indicators and to check government expenditures and verify country sustainability. Using different models to identify if a country is considered a high quality of life. Then using these models to determine which model is the best and which countries are considered to be a high quality of life. These questions will be answered using the World bank dataset. Where the necessary tables and indicators will be extracted and used to create a schema with a country table connecting all tables together via primary and foreign keys. Results will show how GDP growth are strong for human development, with India showing the highest import growth in tech. With Japan, USA, and China being classified as a high quality of life in our models. Allowing us to reference these three countries as ideal countries to follow for sustainability and higher quality of life. And our recommendation of how countries should be like in terms of development.

Introduction

In the determination of improving the world in terms of living a better life and sustainability. Countries work to improve the life of their citizens without relying on the help of others. Threats around the world also exist with climate change looming over our society, demand for sustainable and renewable energy sources has increased dramatically. As a result, we see countries continuously developing in order to make sure that its citizens are living better lives and that the future generations can improve upon what was set for them.

In this project we will take a look at different indicators related to spending, economy, and energy to determine whether these indicators are significant and how renewable energy affects economic growth compared to that of non-renewable energy. All these factors will be placed into different criteria allowing us to review different aspects of a country and its citizen or economic quality to determine what is best and if there is need for improvement for the country. At the end to determine whether a country is considered to have a high quality of life, or what we will consider as a high quality of life.

Such research is necessary to convince countries to put more effort into renewable energy and how it will benefit every related source of economic growth and development. This will allow us to see what countries are much better in terms of air quality and investment into the economy in terms of its people but into other sources like certain industrial aspects. This research will also assist in combating man-made disasters like climate change. The hope being that if countries can see that benefit not just environmentally, economically and socially as well we can hope that adoption of renewable energy sources will increase. We can take a look at certain factors of human life like literacy and human life expectancy and see if countries who invest more in renewable energy and less in non-renewable energy show a higher quality of life.

In this report we will talk about discussing our three objectives and our research questions to go along with those objectives. From there we will discuss our database schema and methodology.

This will include how our database was created, why the specific schema was chosen, and why the specific type of data was chosen with the number of tables. The methodology will also then discuss our chosen analytical models that we have used to get our results. Once the analytical models are disclosed, we can then go into the process of how we are getting our results with the python code and what the results show and indicate. Once we have discussed our results, we can then discuss what we believe these indicate and what are their significance. In this section we will also link these findings to our research objectives and questions. Look into if any other studies were done and compare our results to theirs. Last for the discussion, we will also explain any issues or complications that arose from our research and implementation. This paper will be concluded with a conclusion that will summarize our findings and research. We will also discuss any areas of research we would like others to pursue in the future.

Research Objectives and Research Questions

This project will tackle three different objectives. In the first objective we will take a look across the human development and environmental indicators across our countries. For our first objective we will focus on human development, economic and social indicators. These indicators will be used to research critical investment thresholds and social welfare improvements. The research will identify critical investment thresholds and policy interventions that optimize human development returns by employing advanced econometric modeling and time-series analysis. We will also look into how tourism and trade patterns affect the development of these indicators.

With the goal being to provide evidence-based recommendations for policymakers seeking to leverage economic activities to address fundamental social challenges while establishing a replicable methodological framework for ongoing development assessment.

To answer the first objective, we will look at the following questions. First, how have changes in GDP and Foreign Direct Investment influenced life expectancy and poverty indicators over the past two decades, and what temporal lags exist between economic investments and measurable human development outcomes? Second, how has the technological content of imports evolved the five economies over the time, and what do emerging patterns of the imports reveal about their technological development and strategies and position.

For our second research question we will take a look at carbon intensity, renewable energy adoption, and natural resource allocation across different economic stages of development. The focus for this research objective will be on how trade flows, foreign direct investment, and government expenditure influence environmental outcomes. The findings will provide decision-makers with evidence-based frameworks to evaluate policy trade-offs, prioritize interventions with maximum sustainability returns, and design context-specific strategies that align environmental protection with economic development imperatives.

To address the following objective, we will answer the following. First, we will look at which distinct country profiles emerge when nations are clustered based on their economic and human development indicators, and what common policy approaches characterize each cluster? Second, what distinct patterns emerge in the composition of imports (consumer goods, capital goods, intermediate inputs) across these five economies, and how do these patterns reflect their position in global value chains?

For the last objective we will take a look at a personal definition of a high quality of life. In order to do that we will need to predict and find out what indicators can be used to consider what is a high quality of life. From there we will take a look at what countries are considered a high quality of life. The goal will be to consider what is a high quality of life and direct countries into where they can focus to improve the quality of life for its citizens.

To address this objective, we will answer the following questions, first, we will look at how carbon intensity metrics (kg CO2e per GDP) evolved in relation to economic growth indicators across different development stages, and what inflection points signal environmental Kuznets curve transitions? Second, we will look at what country groups can be identified based on renewable energy use, carbon emissions, and economic development, and how well can future renewable energy growth be predicted using past trends in foreign investment, trade, and environmental spending. We can then check with this information if environmental factors play a big role in the quality of life.

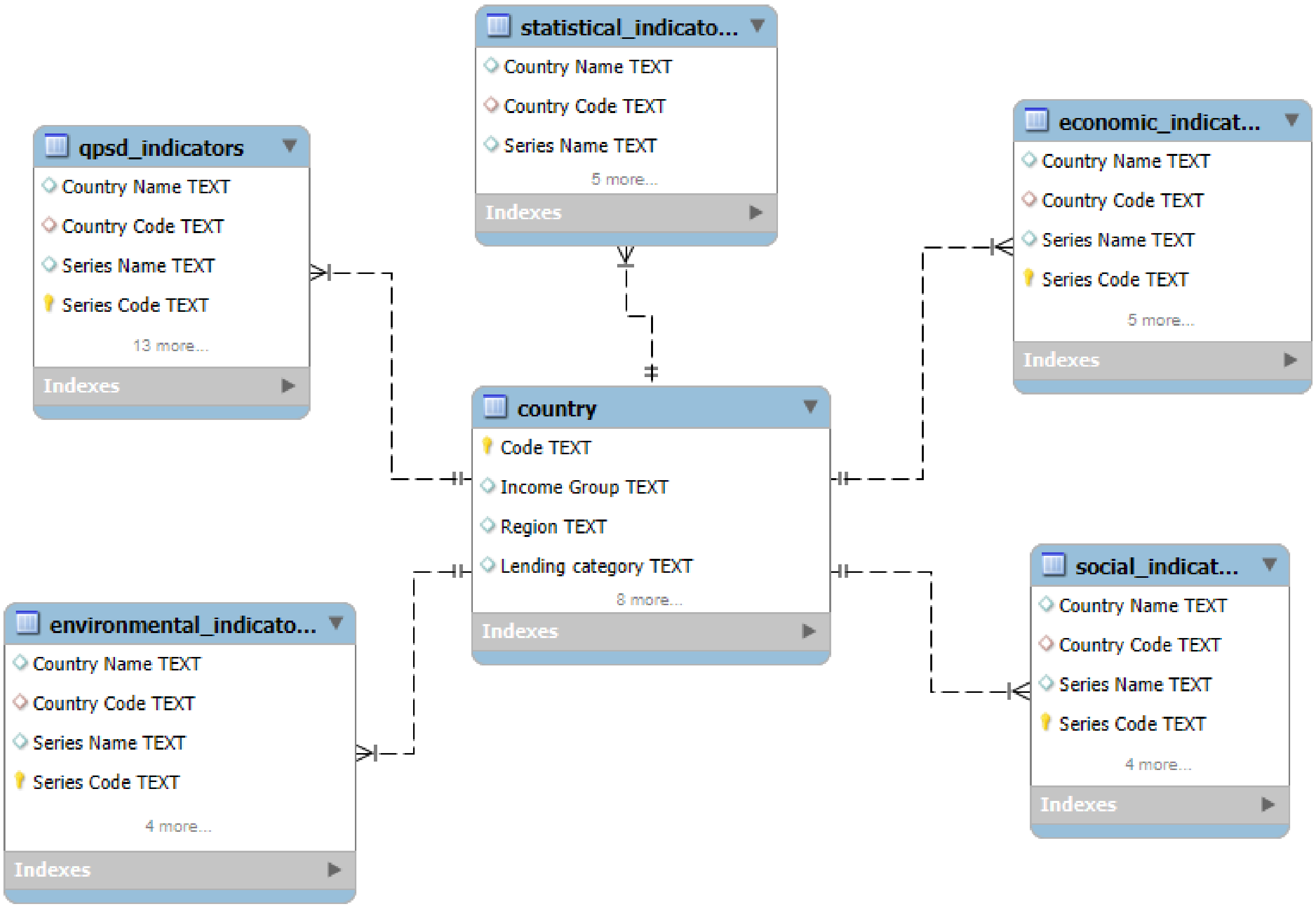
The end goals will be the answer to a main objective which is used to direct development of countries to a higher quality of life and sustainability. With the threat of climate change and other looming disasters, countries will need to take a different direction of how they approach development and success. If these research questions are addressed correctly then we can show all these countries how beneficial it is to take such an approach in development.

Methodology

Our database uses data from the World Bank. This website contains a lot of information regarding a country's many aspects of life through different indicators. To understand what type of data the database is, we can look at some examples like gender, monetary spending, energy, and jobs. Each of these examples are tables that contain different indicators and/or sub indicators. The website allows us to select with indicators we want to use and along with the relevant sub indicators. Based on this anyone can make a database that meets their needs like with our group using it to get the data needed to answer our research questions and objectives.

For this project we decided to create a schema and grab the database from five different tables in the world bank website. These tables are countries, economic, environmental, QPSD (Quarterly Public Sector Debt), Social, and Statistical tables. We decided to use a set of five countries in the country table, which include India, China, United States, Germany, and Japan. The other tables we grabbed many indicators related to our research questions and topics. For example, if we look at the environmental table, we can see that it contains many indicators like 'Access to electricity (% of population)' or 'Agricultural land (% of land area)'. The sub indicators being the items in the parentheses. The table will also include the indicator code, and type. The numerical value for the indicator will be explained showing the value for three years, 2021-2023.

The schema will be created via MySQL workbench. The tables used and the schema of the database are reflected in the database in the image below. As you can see the series code along with the country code will be used as the primary and foreign keys for each table. Each table will be tied to the country table and use the one-to-many relationship. With the country table being used as the origin of the foreign key related to the country. After the creation of the schema a total of six tables will be created and used to query all data needed for our research questions and objectives.



The analytical models will connect to the MySQL workbench via the MySQL connector library. From there we will query our dataset to use in our analytical models. These analytical models will then be used to answer our research questions and objectives.

For the data cleaning we started by removing any duplicate and null columns. Once we have removed that information, we add a new column where we calculate the growth rate between 2021 and 2022. We will then look at all the categorical columns and then one hot encodes those columns. Once all this is completed, we normalize and standardize all columns which should be numerical.

Now that we have cleaned and scaled our data, we can now do our EDA and models.

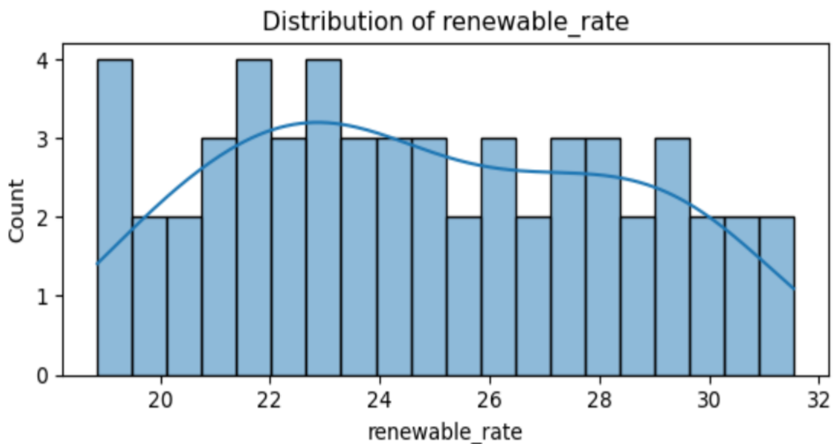
When working on the models, we started with the first objective of how economic activities influence human development. We will specifically focus on GDP and foreign direct investment for one part along with technology impact economic development. We will use statistical models and Ordinary least squared models to analyze the impact of economic and human development indicators via the GDP and FDI. In regards to the technological imports we use heat maps and PCA to review the impact of the imports. We can also use PCA and compare the results to that of a Cluster analysis.

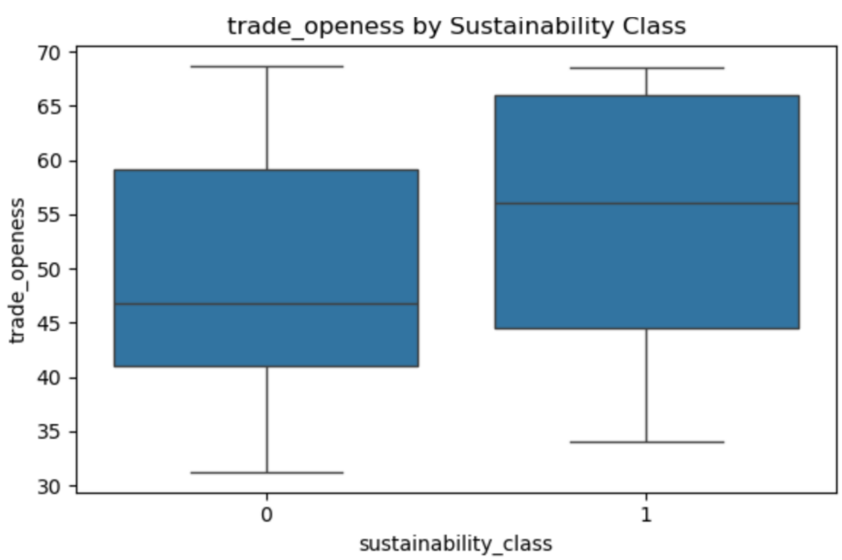
For the second objective we will review the carbon, renewable energy adoption rates with investment rates. The models in this objective will be Bayesian ridge regression and time series cross validation to assist with the predictions of the renewable energy adoption. Along with that, random forest, logistic regression, and SVM will be used to compare the prediction and see which features are most important in terms of renewable energy adoption.

For the last objective, which determines what factors in a high quality of life. For this we will use statistical models to rank our five countries and give them a high quality score using the statistics models. Random forest will be used to locate the important features for what is considered a high quality of life. And see what factors should be focused on by countries.

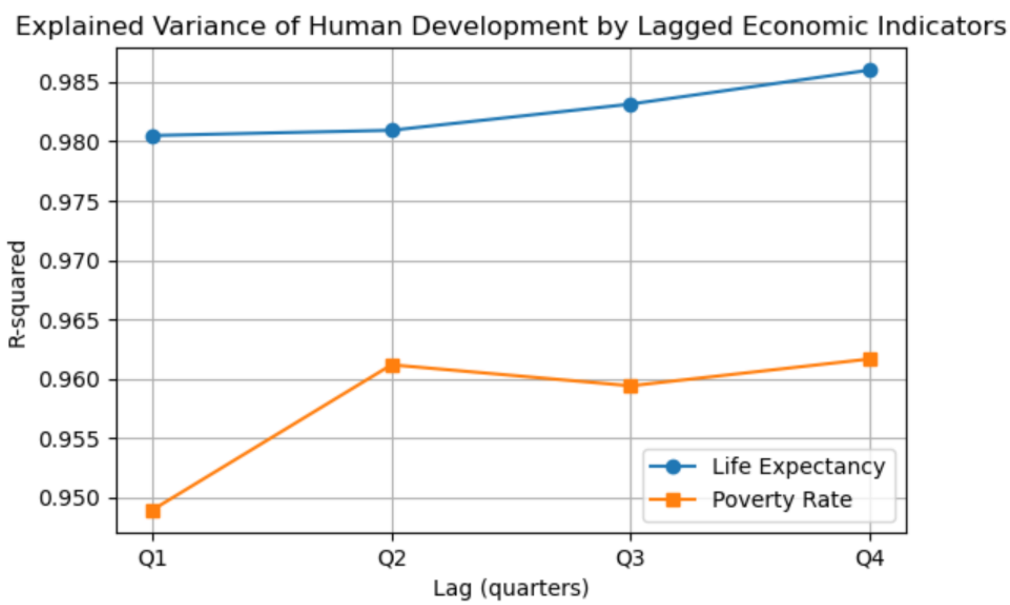
Results

The EDA explored economic, environmental and social indicators tables for the five countries. missing data imputed, Growth rates were synthesized, Outliers detected and distributions normalized. Key findings include renewable rate and env_gov_spending strong positive correlation of 0.22. FDI as % of GDP showed right -skewed distribution while renewable energy rate revealed bimodal distribution. Sustainability class showed that High-sustainability countries had a higher median renewable energy percentage. Some high-sustainability countries (like Germany or Japan) were highly trade-open, but others like the USA might not be as open. The variation suggests that trade openness alone does not determine sustainability.

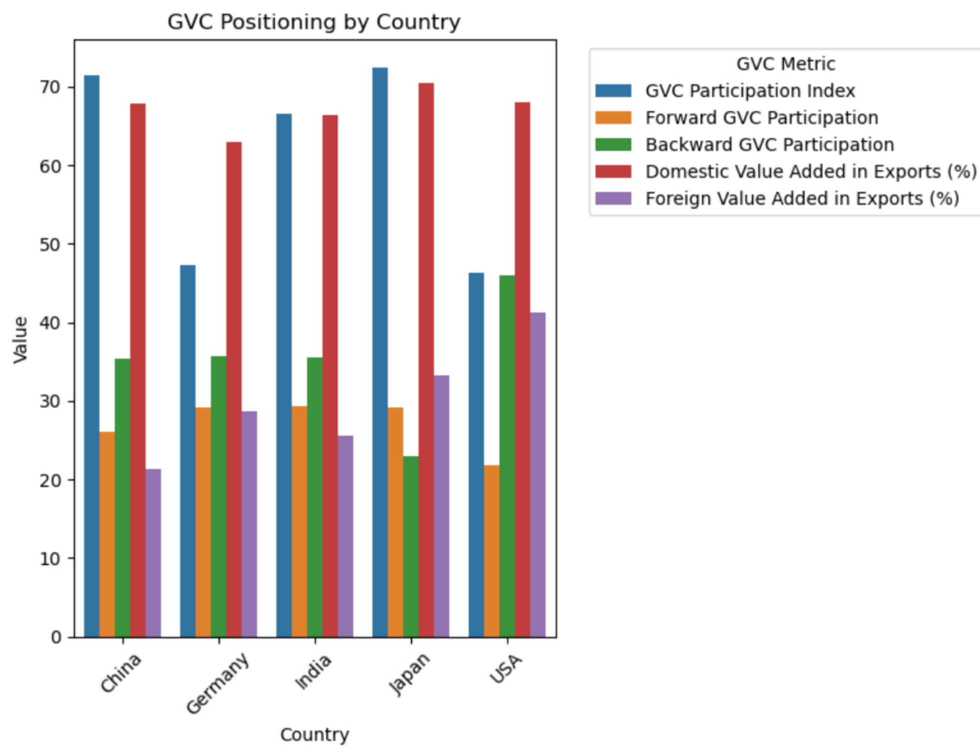




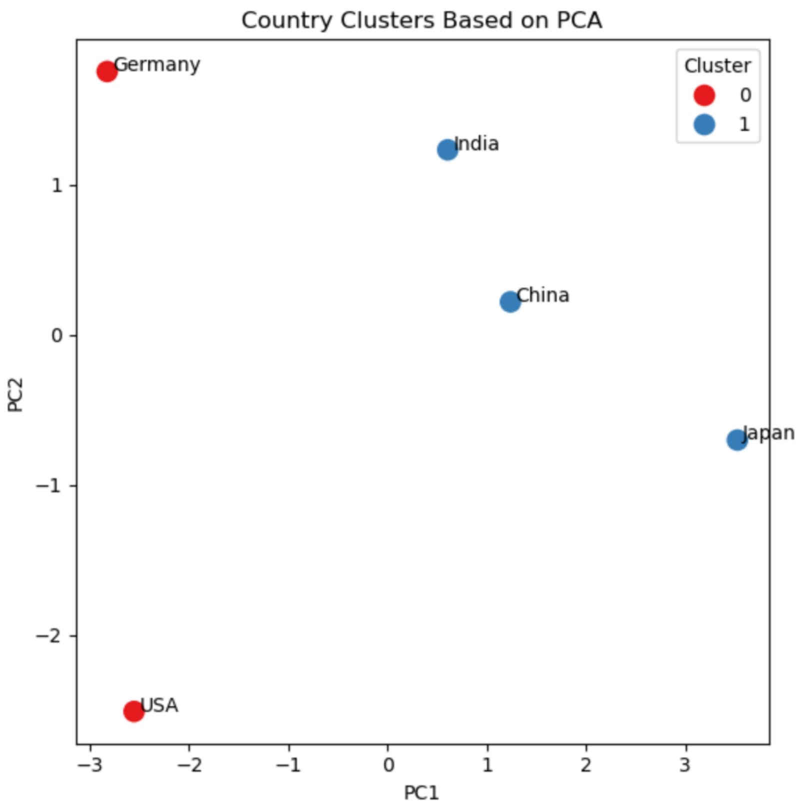
The lag analysis reveals that GDP growth exerts the strongest influence on human development indicators after approximately four quarters. Notably, GDP increases are highly correlated with improvements in life expectancy ($R^2 = 0.986$) and reductions in poverty ($R^2 = 0.962$), highlighting the critical role of economic growth in shaping social outcomes. However, the effects are not uniform across indicators. Life expectancy responds more gradually to changes in GDP, reflecting long-term improvements in healthcare, infrastructure, and living conditions. In contrast, poverty rates tend to react more quickly, suggesting that income growth and job creation have immediate benefits for the most vulnerable populations. Foreign Direct Investment (FDI) showed minimal short-term effects on these outcomes, indicating that policy efforts should prioritize strategic, long-term investment planning rather than expecting immediate gains. Overall, sustained GDP growth emerges as the primary engine of human development across all five countries studied, reinforcing the importance of stable, inclusive economic policies for lasting social progress.

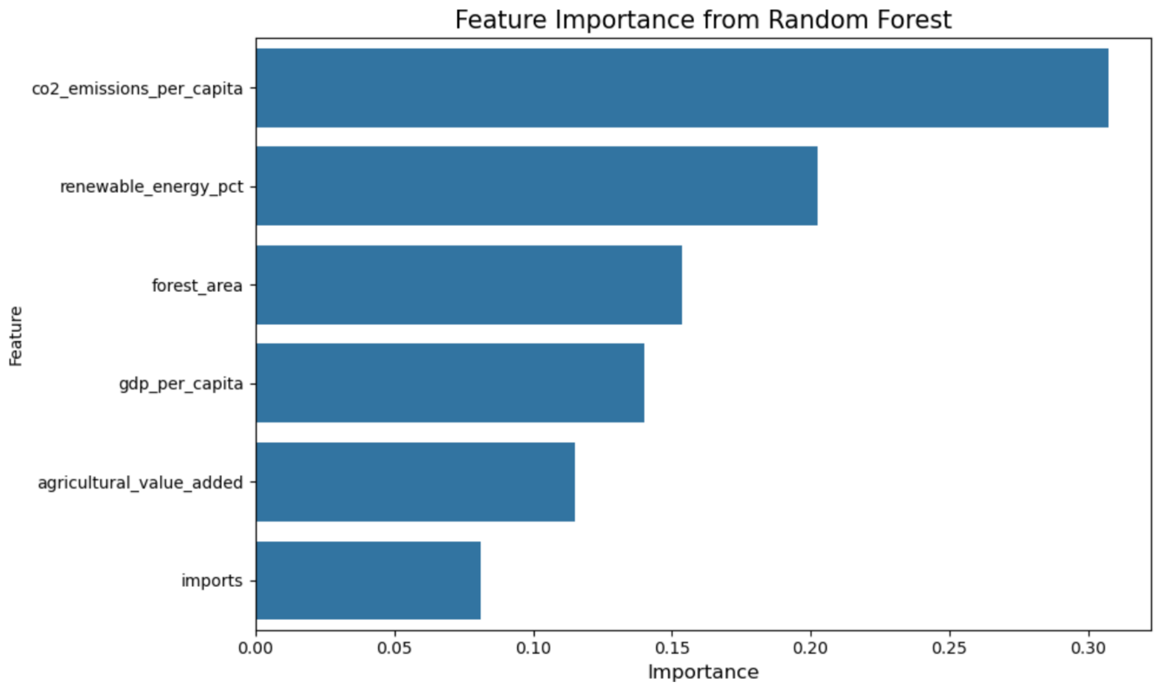


India has recorded the fastest growth in ICT imports, rising from 14% to 21%, underscoring its rapid digitalization and emerging role as a global value chain (GVC) hub. Similarly, China, already a major electronics assembly center, continues to import significant volumes of intermediate goods, reinforcing its central position in global electronics production. In contrast, the United States and Germany show steady ICT import growth, aligning with their strong backward participation in GVCs. Both countries rely on importing high-tech components to support their advanced manufacturing and service industries. Japan exhibits low ICT import growth, reflecting its upstream position in the GVC. It functions primarily as a technology supplier, providing critical inputs and innovations to downstream economies. These trends illustrate two distinct clusters: India and China as emerging GVC hubs, increasingly integrated into global production networks. USA, Germany, and Japan as advanced economies with complex, technology-intensive value chains and higher value-added contributions at different stages of the production process.

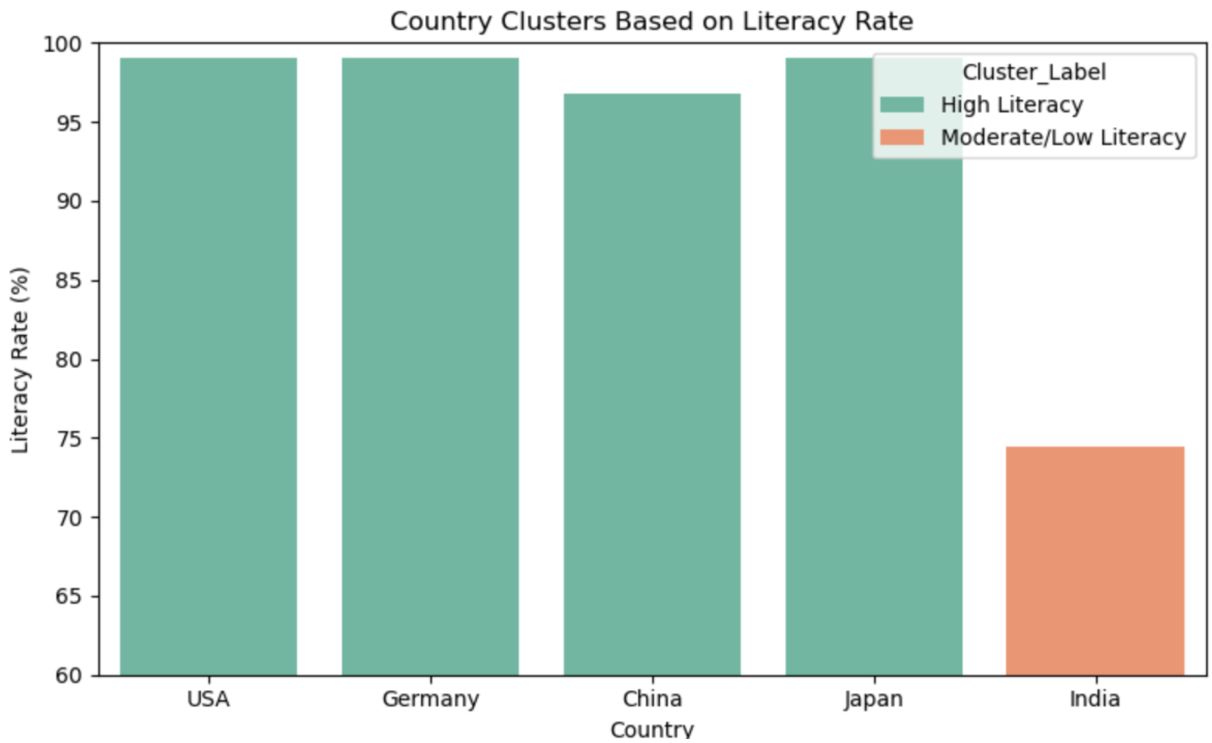


Countries were classified as “high quality” based on a composite score integrating both socioeconomic and environmental indicators. Japan, the USA, and China emerged as top performers, driven by strong overall development. In contrast, Germany and India received lower rankings, primarily due to weaker environmental performance. Among the contributing factors, pollution-related deaths had the highest influence on the quality classification, followed by renewable energy adoption, air quality, and waste management practices. These environmental variables proved more decisive than economic metrics alone, highlighting the growing importance of sustainability in national performance assessments. For modeling and classification tasks, Logistic Regression or Decision Tree algorithms are recommended. These models not only offer strong classification accuracy but also allow for interpretability, making them suitable for policy-focused analysis with an emphasis on environmental outcomes.

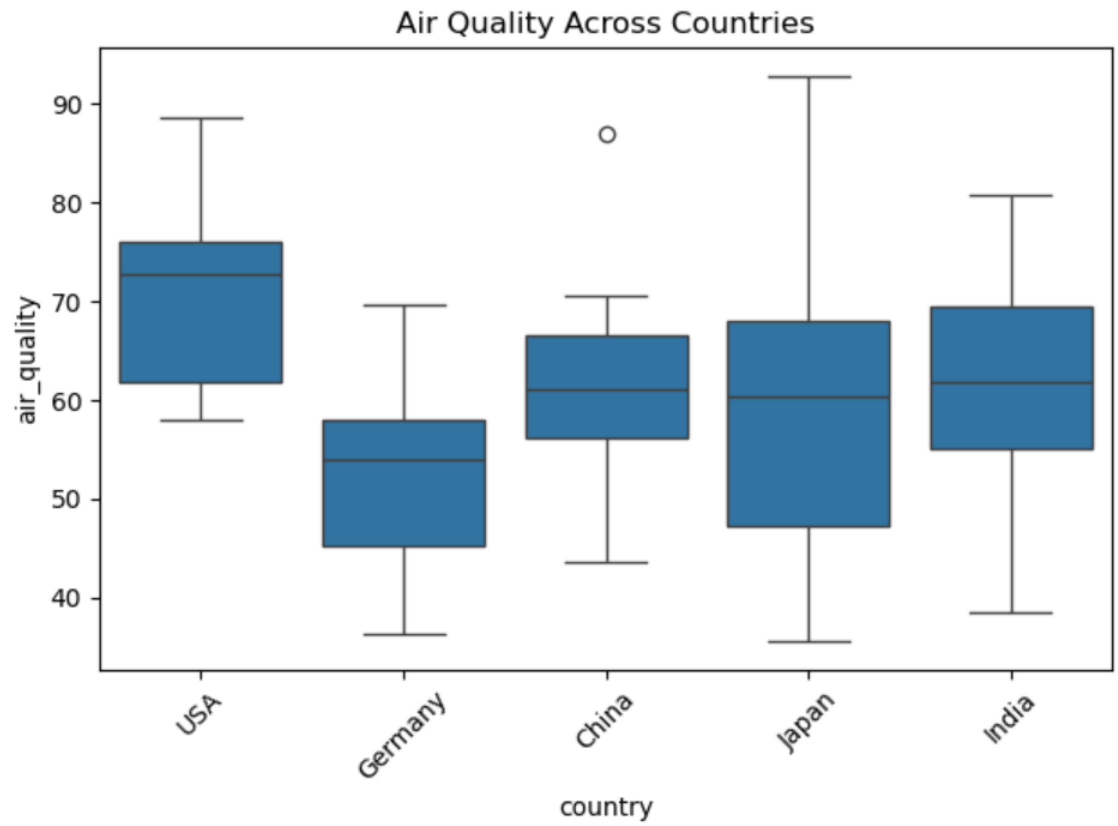




Air quality emerged as the most influential factor in determining quality of life across countries, showing a statistically significant difference between high and low quality-of-life groups ($p = 0.0199$). ANOVA results further indicated meaningful variation in air quality across nations ($p = 0.0544$), with the USA and Japan demonstrating higher median levels and Germany the lowest.



Additional key drivers included forest area, climate vulnerability, and waste management. While these variables contribute meaningfully to life quality predictions, renewable energy showed the lowest predictive power, and factors such as CO_2 emissions and green space access did not differ significantly across groups ($p > 0.1$). The classification model achieved moderate performance, suggesting that while environmental variables are important, they only partially explain variations in quality of life. Nonetheless, the findings strongly support integrating environmental priorities, especially clean air, into broader development strategies. So clean air stands out as a cornerstone of life quality, reinforcing the need for robust environmental regulation in future policy planning.



Discussion

The analysis highlights a multifaceted relationship between economic performance, technological advancement, environmental sustainability, and quality of life. Lagged GDP growth shows a clear positive impact on human development, with significant improvements in life expectancy and reductions in poverty emerging about four quarters after economic expansion. This delayed effect underscores the importance of maintaining consistent and inclusive growth policies to achieve long-term social benefits.

Meanwhile, renewable energy adoption appears most responsive to government spending, indicating that public investment remains a key lever for accelerating green transitions. Unlike private-sector or market-led drivers, renewable infrastructure and innovation depend heavily on state support, particularly in emerging economies.

ICT imports serve as another critical driver, not only advancing domestic technological capabilities but also deepening countries' integration into Global Value Chains (GVCs). Countries like India and China are strengthening their roles as emerging GVC hubs through digital infrastructure expansion, while advanced economies such as the USA and Japan maintain complex, upstream positions within these networks. Tailored trade and industrial policies should reflect these distinctions, reinforcing each country's comparative advantage while promoting resilience and innovation.

In terms of life quality, air quality and pollution-related deaths emerged as the strongest environmental determinants, with statistically significant disparities observed across countries. These findings suggest that clean air is not just an environmental goal but a fundamental component of human well-being. As such, environmental health should be tightly integrated into national development agendas.

Finally, the use of composite sustainability scores offers a powerful tool for benchmarking progress and guiding policy decisions. By combining socioeconomic and environmental indicators, such scoring systems can help governments prioritize interventions, allocate resources more effectively, and communicate progress transparently to stakeholders.

So a coordinated policy approach is essential for advancing national development in a balanced and resilient manner.

Conclusion

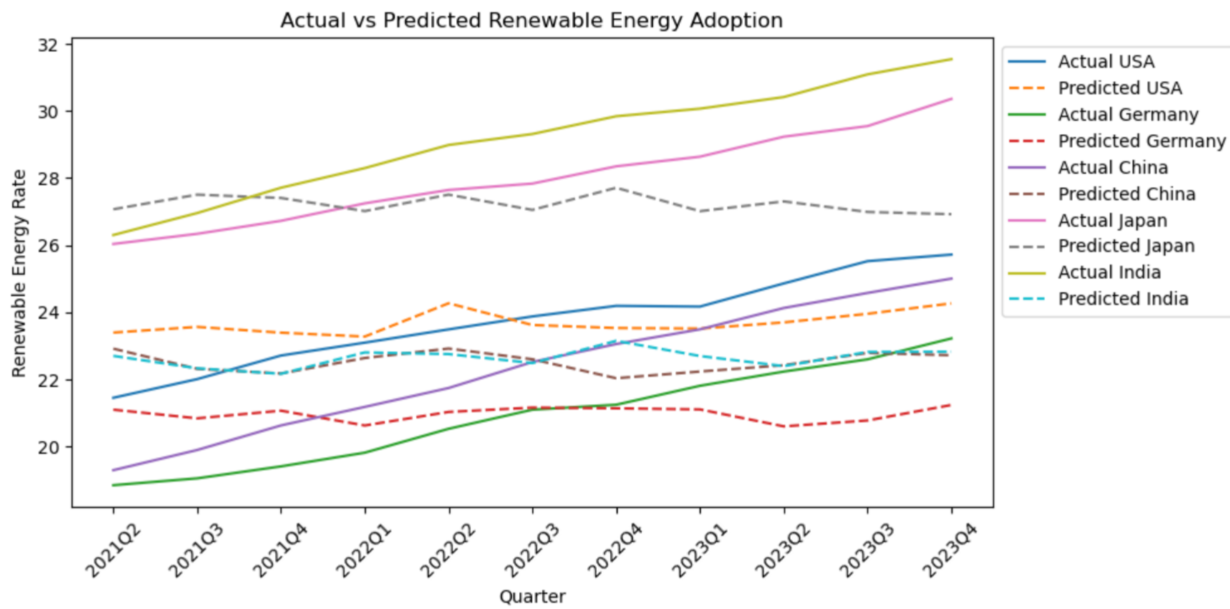
This project is a framework of integrated analytical assessment of sustainable development of nations: the USA, Germany, China, India, and Japan using the key factors that determine an economy in all dimensions. By applying predictive analytics models, time-series modeling supervised and unsupervised learning, the study explores the economy's performance across economic, social, environmental, etc. with dimensions such as GDP growth, air quality (via CO_2 emissions), renewable energy use, and forest coverage.

This research work focused on enhancing trend analysis by adding growth rate features across the indicators providing a forward-looking lens on development patterns. This key analysis revealed meaningful insights on all the nations in terms of finding growth patterns.

The GVC analysis provided critical notes on how nations are positioned within the international trade ecosystem. Countries deeply embedded in GVCs are not necessarily sustainable unless they also invest in clean energy, efficient production, and emissions control. Together, GVC metrics and PCA added depth to the study by highlighting structural differences in trade roles and sustainability strategies

As the research further progresses and classification models are applied, synthetic sustainability index came up as a core innovation, this composite metric integrates economic performance and environmental health. This index helped to classify countries based on low, medium and high quality sustainable groups further revealing what combination of indicators are contributing to sustainability or hindering the progression of an economy.

Machine Learning models like Random Forest and Bayesian Ridge Regression accurately predicted the renewable energy prediction rates with an RMSE of 4.587 by providing valuable insights into the future trajectory of clean energy transitions. Incorporating renewable energy prediction into broader sustainability modeling helps align energy policy with long-term environmental and economic planning.



Additionally, each country's economic profile showed unique traits: Germany maintained high economic output with strong environmental controls, while India showed rapid renewable adoption with relatively low GDP per capita. The USA and China demonstrated strong economies but faced challenges in emissions and land depletion.

Air quality was central to this evaluation. CO_2 emissions per capita consistently ranked as the most important factor in sustainability classification, underscoring that economic development not coupled with environmental responsibility leads to systemic inefficiencies. Countries with cleaner air, higher forest area, and greater renewable adoption were more likely to be sustainable in the long term.

To evaluate the human impact of economic and environmental strategies social indicators like literacy rate, poverty levels and life expectancy played a crucial role. Countries like Germany and China with higher literacy and life expectancy exhibited better environmental practices and greater renewable energy adoption. While India, with lower poverty and life expectancy showed rapid progress but faced developmental trade-offs between emissions control and poverty alleviation.

Ultimately, the study offers a data driven framework that integrates economic vitality and ecological responsibility without compromising on the ability of future generations to thrive, ensuring a resilient economy. From the study it is also evident that true sustainability lies at the intersection of economic prosperity, environmental integrity, and responsible land management.

References

https://databank.worldbank.org/databases/page/1?qterm=global

https://www.mdpi.com/2071-1050/15/13/10682