Megha Shakya
Aayusha Lamichhane
DS 325
May 2, 2025

Exploring the Relationship Between Government Effectiveness and Renewable Energy Adoption Across Countries

Introduction

As countries work to meet sustainability goals, the adoption of renewable energy has become a global priority. A key factor in this transition is a government's ability to create and enforce policies that support clean energy development. However, not all countries make equal progress. We hypothesize that government effectiveness (a measure of quality of public services, independence of the civil service, effectiveness of policy making and implementation) plays a significant role in influencing renewable energy use across nations.

To test this, we used World Bank panel data covering over 100 countries and examined the relationship between renewable energy consumption and variables such as government effectiveness index, CO_2 emissions, foreign investment, GDP growth, and fossil fuel use. We also group countries by income level to explore broader patterns. We applied linear regression, elastic net regression (with grid search), and random forest regression to model the relationship and compare predictive performance.

Our results show that while linear and elastic net models provide reasonable baseline predictions ($R^2 \approx 0.89$), the random forest model performed significantly better ($R^2 \approx 0.98$), suggesting strong nonlinear dynamics. These findings highlight that regulatory quality is a strong predictor of renewable energy use.

Method:

To explore how regulatory effectiveness impacts renewable energy adoption, we use cross-country panel data from the World Bank. Our target variable is renewable energy consumption (% of total final energy use). The explanatory variables includes: regulatory quality index, CO_2 emissions (% change from 1990), GDP growth (% annual), fossil fuel energy consumption (% of total), foreign direct investment, control of corruption, energy imports (% of energy use), energy use per GDP, and renewable freshwater resources. Our primary goal is to predict renewable energy usage and understand which governance or economic factors influence it the most.

Data preprocessing involved multiple steps. We began by individually reshaping two different datasets from wide-format CSVs into long-format data for consistency and then merging them. We cleaned the datasets by converting all variables (except country names) to numeric, handling different formats, and renaming columns using a consistent dictionary. We also generated a correlation matrix, showing moderate correlations, but none of them showing high enough correlation that required excluding variables.

Due to extensive missing data, especially in the regulatory environment rating and certain years post-2021, we dropped those variables and years. To ensure model accuracy, we filtered the dataset to only include countries with complete observations across all selected variables and across all 21 years. We further mapped countries into four income groups—Low, Lower Middle, Upper Middle, and High—based on IMF classification and encoded them using an ordinal encoder.

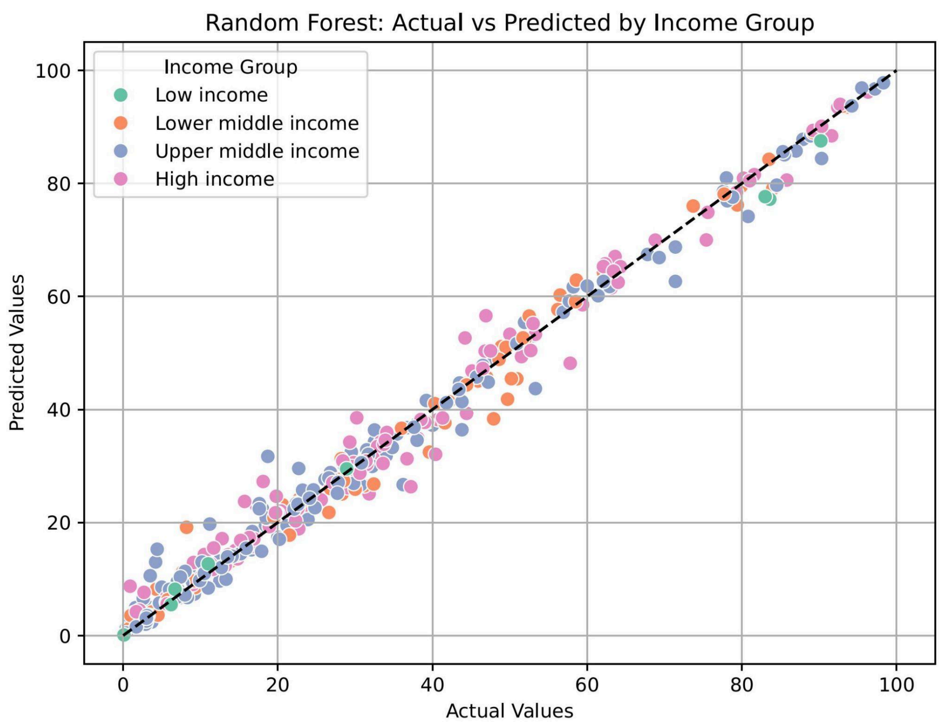
We tested multiple models. First, we applied linear regression as a baseline to measure basic relationships. Next, we used Elastic Net regression, tuning hyperparameters (alpha and l1_ratio) through grid search for improved generalization. Lastly, we applied Random Forest regression, which can handle complex, nonlinear patterns, and further optimized it using a grid search on hyperparameters like tree depth, number of estimators, and sample splits. All features were scaled using StandardScaler prior to modeling. Comparing these models allowed us to evaluate both linear and nonlinear relationships

Results

To evaluate model performance, we used Mean Squared Error (MSE) and R^2 (coefficient of determination) for all regression models. The baseline linear regression model achieved an R^2 of

0.89.84 and an MSE of 67.89, indicating that about 89.84% of the variance in renewable energy consumption could be explained by the predictors. The elastic net regression, which combines L1 and L2 regularization, slightly improved performance with an R^2 of 0.8983 and an MSE of 67.90. Grid search identified the best hyperparameters as $\alpha = 0.01$ and l1_ratio = 0.9.

The most effective model was the random forest regressor. After tuning hyperparameters through grid search, the model achieved an R^2 of 0.9880 and an MSE of 8.05, demonstrating its ability to capture complex, non-linear relationships in the data. A predicted vs. actual scatter plot by income group below shows the model's accuracy, with most points closely aligned with the diagonal line of perfect prediction.



Discussion

Our project reinforces an important insight: government quality plays a crucial role in promoting sustainability. Specifically, countries with more effective, transparent, and stable institutions are more likely to adopt and invest in renewable energy. This suggests that beyond technical or financial capacity, governance structures are key enablers of clean energy transitions.

However, getting to these findings wasn’t straightforward. One of the biggest challenges we faced was missing data. A significant portion of variables, especially regulatory indices and environmental measures, had gaps across countries and years. This made it stressful and time-consuming, especially since models like linear regression and elastic net cannot handle missing values. We tried various fixes like dropping columns, filtering countries and ultimately restricted our analysis to countries with complete data.

Looking back, we could have run a preliminary check on data availability and balance before finalizing our topic. This would have saved time and avoided surprises mid-project. We also recognize that our analysis does not fully address endogeneity like reverse causality between governance and renewable use which would be important in future work.

If we had more time, we would have experimented with converting renewable energy consumption into categories like low, medium, and high, and used classification models instead. Exploring additional models such as gradient boosting or XGBoost could also reveal new patterns.

Citations:

CitationsOpenAI. (2025). *ChatGPT (May 2 Version)*.https://chat.openai.com ChatGPT was used to help with idea clarification and coding assistance

Deniz Gunay. “Missing Values - Deniz Gunay - Medium.” *Medium*, 15 Aug. 2023, medium.com/@denizgunay/missing-values-6742e535196b. Accessed 2 May 2025.

GeeksforGeeks. “Using Dictionary to Remap Values in Pandas DataFrame Columns.” *GeeksforGeeks*, 23 Jan. 2019, www.geeksforgeeks.org/using-dictionary-to-remap-values-in-pandas-dataframe-columns/#. Accessed 2 May 2025.

“Pandas.to_numeric — Pandas 1.4.2 Documentation.” *Pandas.pydata.org*, pandas.pydata.org/docs/reference/api/pandas.to_numeric.html.

Tomáš Konyárik. “How to Convert Datetime to Integer in Python.” *Stack Overflow*, 2024, stackoverflow.com/questions/28154066/how-to-convert-datetime-to-integer-in-python.

Figures:

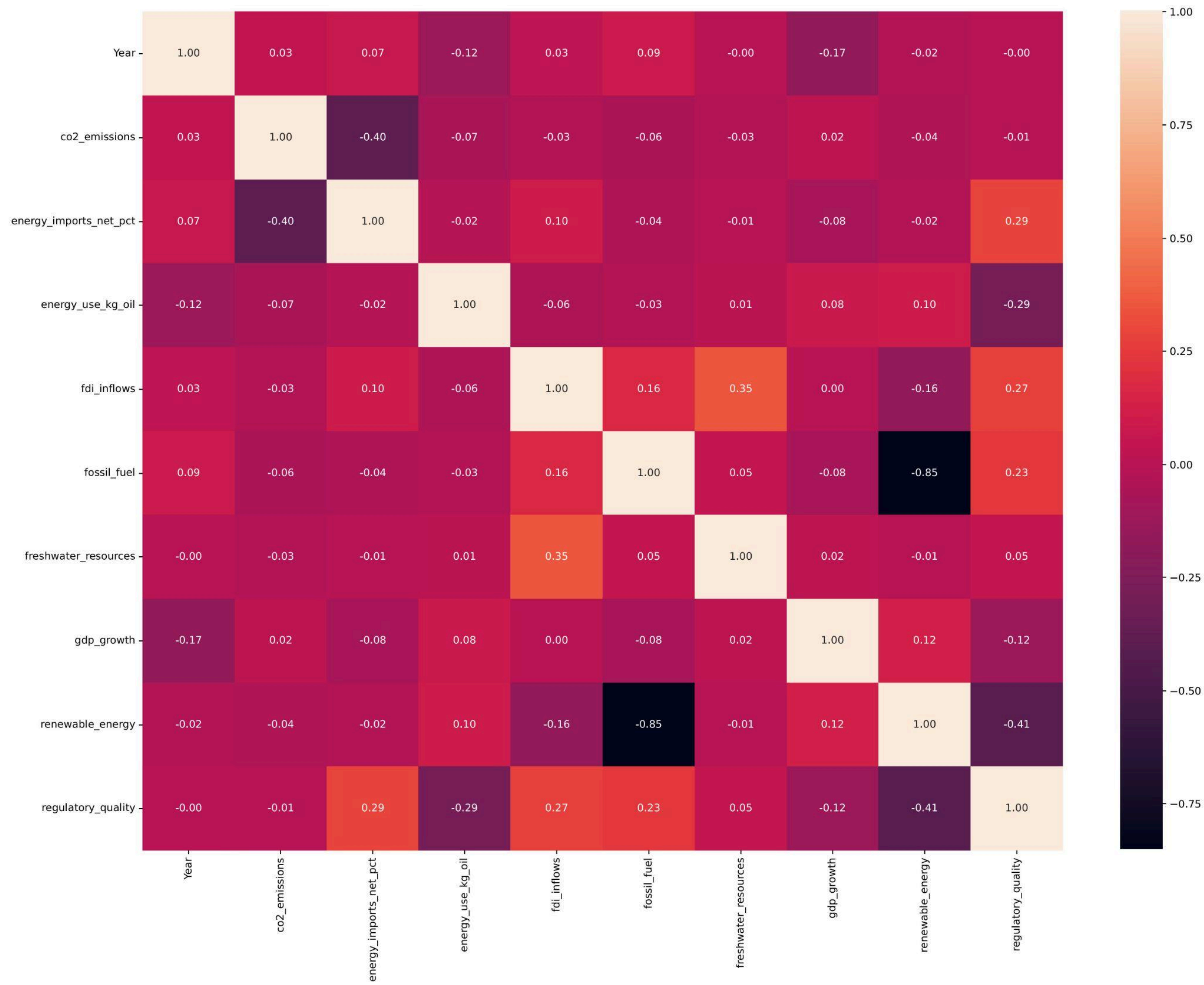


Figure 1: Correlation matrix

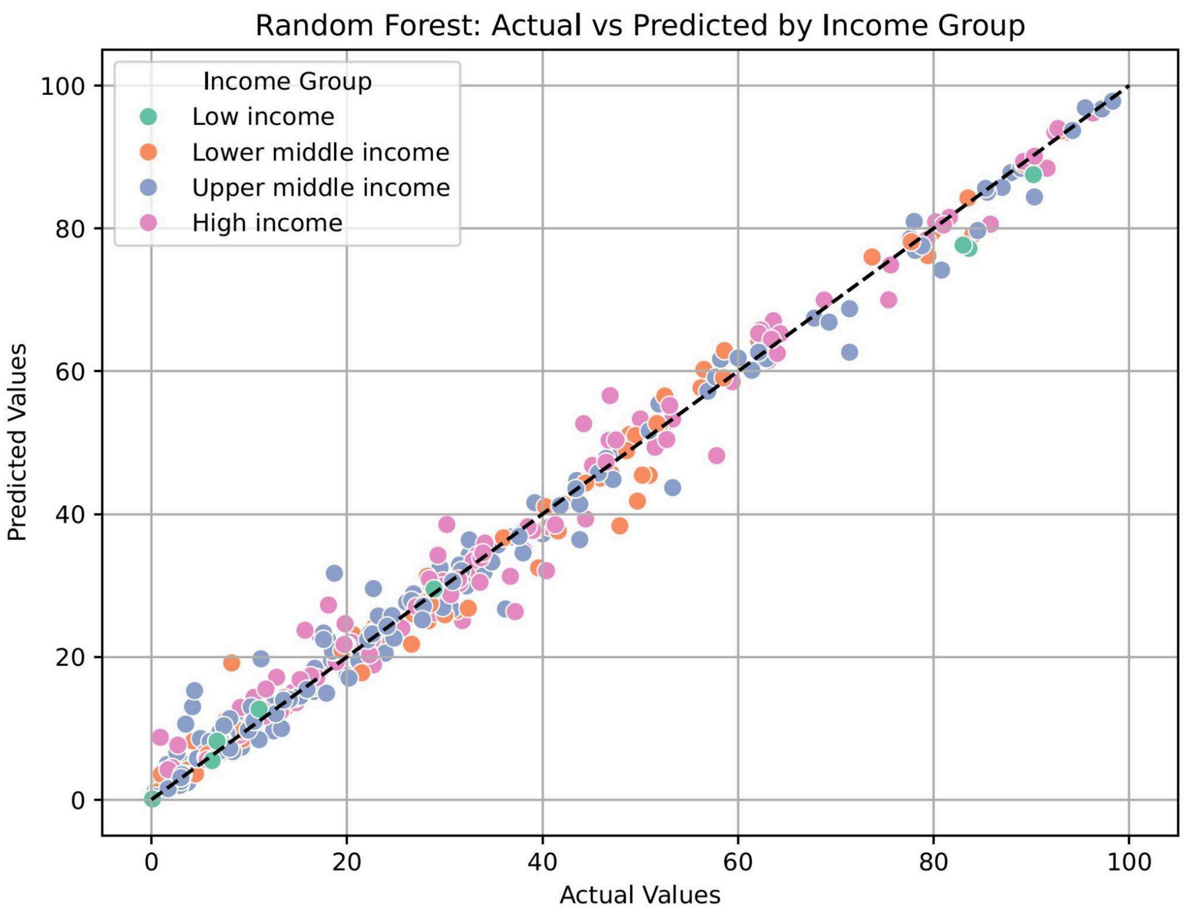


Figure 2: Scatter Plot of actual vs. predicted for random forest regressor