Predicting the Clean Energy Future: Strategic Modeling for Renewable Transformation

Prepared by
Ahmed Elmasry

Contents:

Overview

- This project outlines a comprehensive, data-driven strategy to promote renewable energy as the core of future energy systems.
- It uses:
  - Historical data analysis from 2018 to 2023, and
  - Advanced machine learning forecasting.
- The evidence shows that renewable energy is:
  - Clean and emission-free
  - Increasingly reliable
  - Cost-efficient
  - Capable of providing sustainable, year-round energy supply
- A key element of the analysis is the development of a unified target variable:
  - renewable_strength_score
  - This score measures renewable energy performance in both operational and economic terms.

Objective

- A permanent and reliable electricity source.
- A clean alternative to fossil fuels with zero direct emissions.
- A cost-effective energy solution.
- Consistently available throughout the year via wind, solar, hydro, and more.
- A strategic investment for climate resilience and energy security.

Data and Methodology

The analysis used a cleaned dataset of hourly and seasonal energy data from 2018 to 2023.

Key stages included:

ry Data Analysis
dentify trends,
, and renewable
ewable patterns.

Feature engineering for extracting time, season, and energy-type attributes.

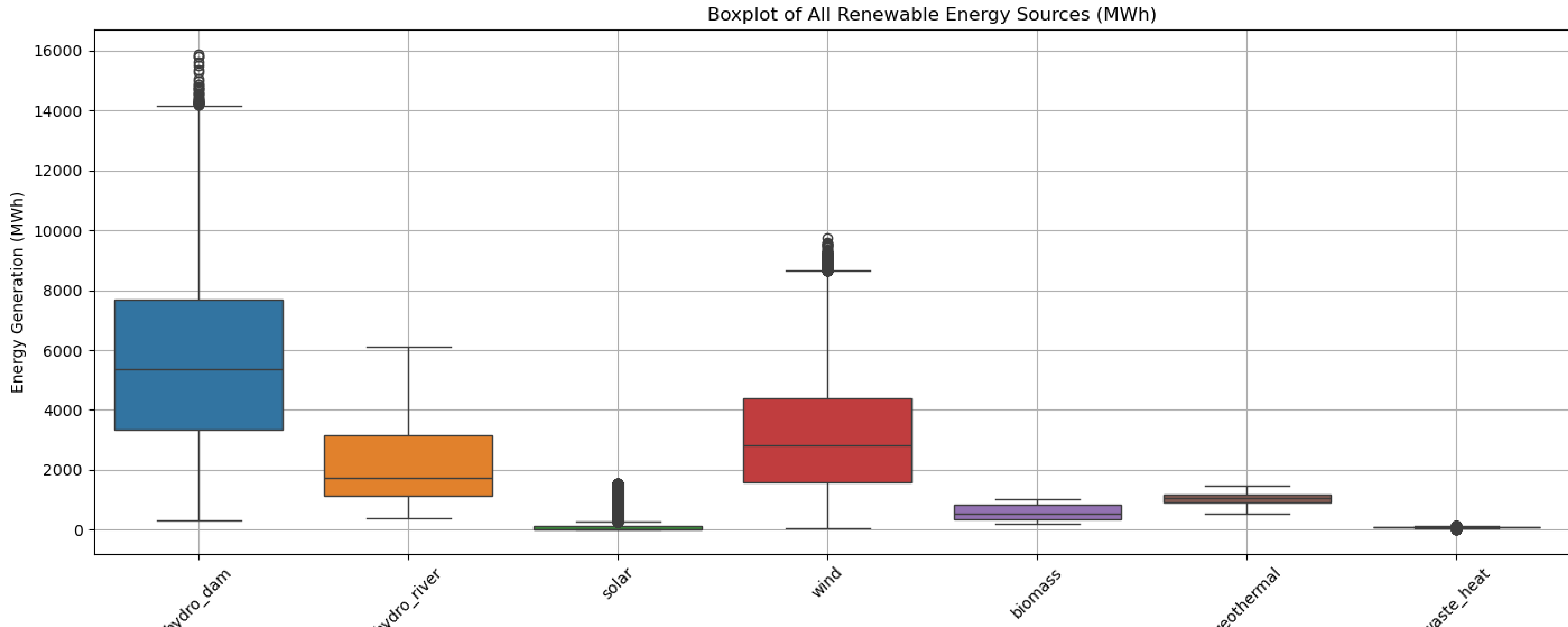
Development of a composite target variable (renewable_strength_score) to evaluate renewable energy quality.

Machine learning models (Linear Regression, Random Forest) trained to forecast this score and validate trends.

Visualizations using
Seaborn and Plotly f
clarity and commun

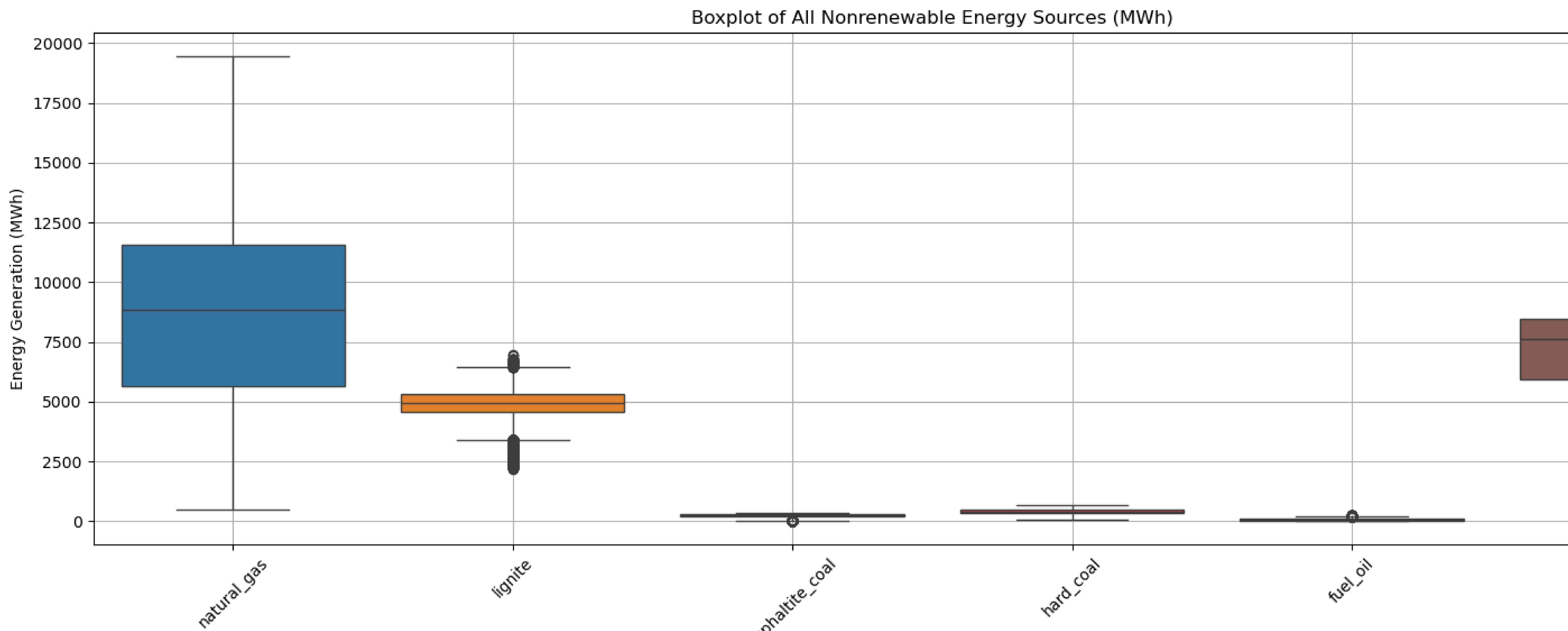
Outliers of Renewable Resources

- Solar shows the most extreme high-end outliers and right-skewed distribution.
- Wind, hydro_dam, and hydro_river have broader variability, reflecting natural fluctuations.
- Waste_heat, biomass, and geothermal are relatively stable with compact interquartile ranges.
- International shows moderate range and few anomalies.



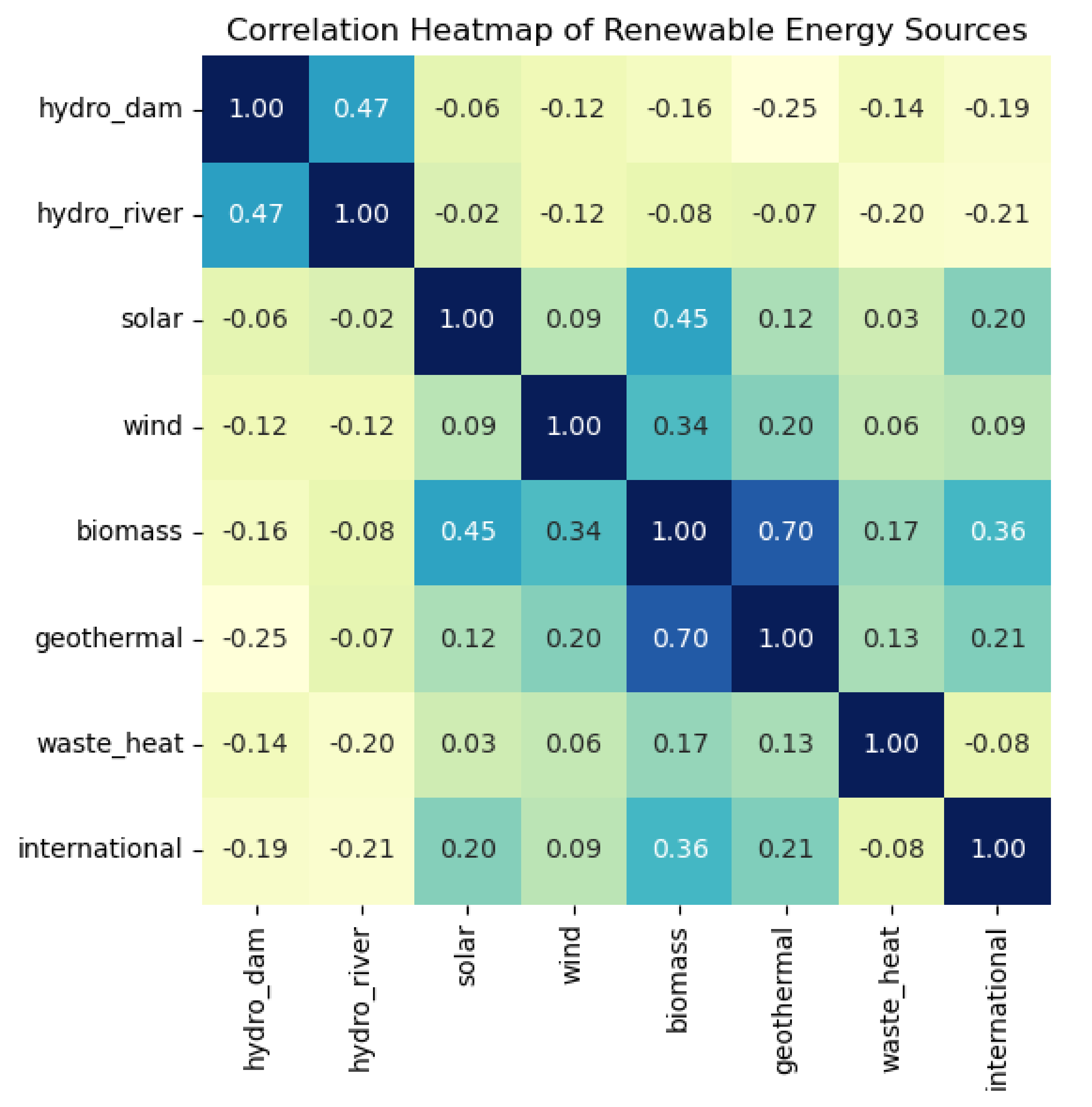
Outliers of Nonrenewable Resources

- Natural gas and lignite remain the dominant sources with broader distributions.
- Asphaltite coal, hard coal, fuel oil, and coal_imported show tighter ranges but with a few visible outliers.
- This version highlights the core fossil sources more clearly without skew from low-usage fuels.



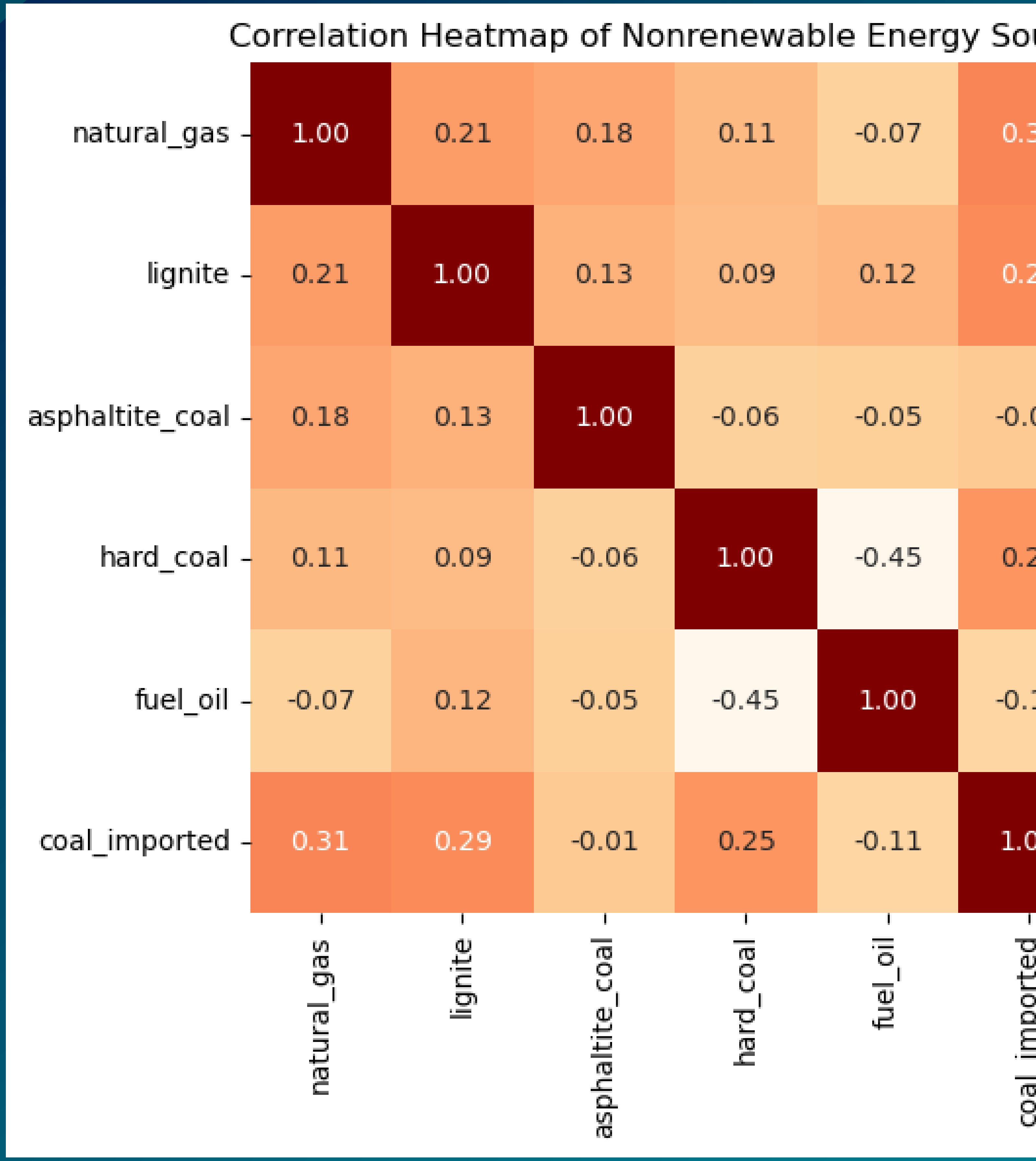
Correlation Relationships of Renewable Energy

- Hydro dam and hydro river are strongly correlated, indicating they often operate together (e.g., water flow dependence).
- Solar and wind are weakly correlated, supporting their complementary seasonal roles.
- Geothermal, biomass, and waste heat show weak correlations with other sources, suggesting they operate more independently or at consistent base loads.



Correlation Relationships of Nonrenewable Energy

- Hard coal, lignite, and coal_imported are moderately correlated, suggesting simultaneous usage patterns.
- Fuel oil and asphaltite coal have weak correlations with other sources, indicating more specialized or intermittent use.
- Natural gas shows weak correlation with others, supporting its role as a flexible or peaking source.



Why We Created the (renewable_strength_score)

The goal was to create a single metric that reflects the strategic qualities of renewable energy.

This target integrates:

- Total renewable energy output (MWh)
- Renewable share of total generation
- Electricity cost (TRY/MWh)

These components were normalized and combined with this weighted formula:

renewable_strength_score = 0.4 * scaled_renewables + 0.4 * scaled_share + 0.2 * (1 - scaled_price)

This allows the score to reflect high output, low cost, and a dominant share in the energy mix.

Interpretation of the Target

renewable_strength_score ranges from 0 to 1. Its interpretation:

0.80–1.00: 🌟 Excellent — high renewable output, dominant share, low cost

0.60–0.79: ✅ Strong — reliable and cost-effective

0.40–0.59: ⚠️ Moderate — good, but room for improvement

0.20–0.39: ❌ Weak — high fossil reliance or energy cost

0.00–0.19: 🚨 Very Poor — unsustainable conditions

Average score in the dataset: 0.511 This indicates moderate performance, validating the need for targeted investment and policy.

Machine Learning Validation

The score is ideal for regression:

- Continuous and normalized
- Model-friendly
- Stakeholder interpretable

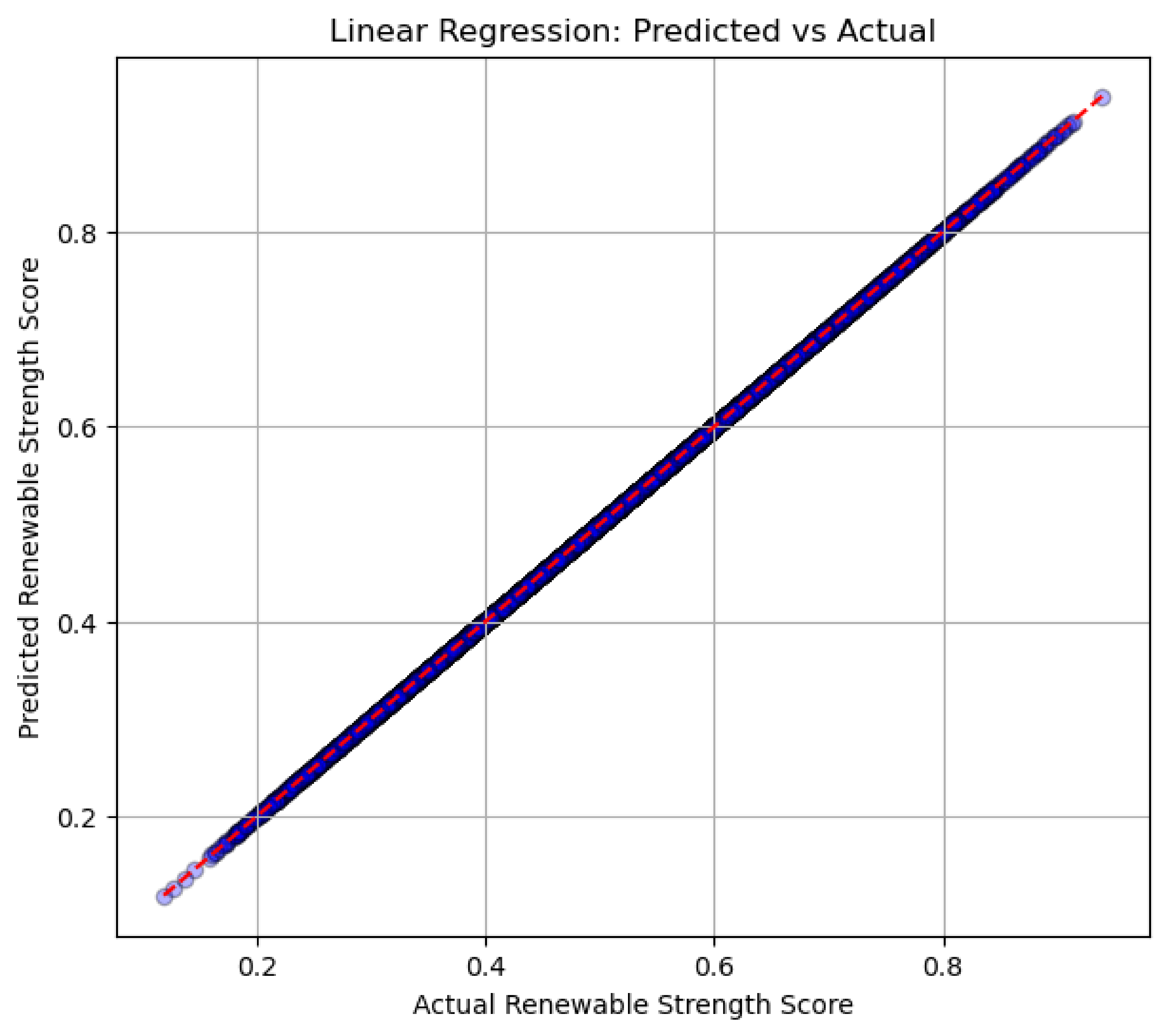
Model results:

- Linear Regression: $R^2 \approx 1.00$
- Random Forest: $R^2 \approx 0.998$

This confirms the score's validity and the models' ability to generalize patterns in the data.

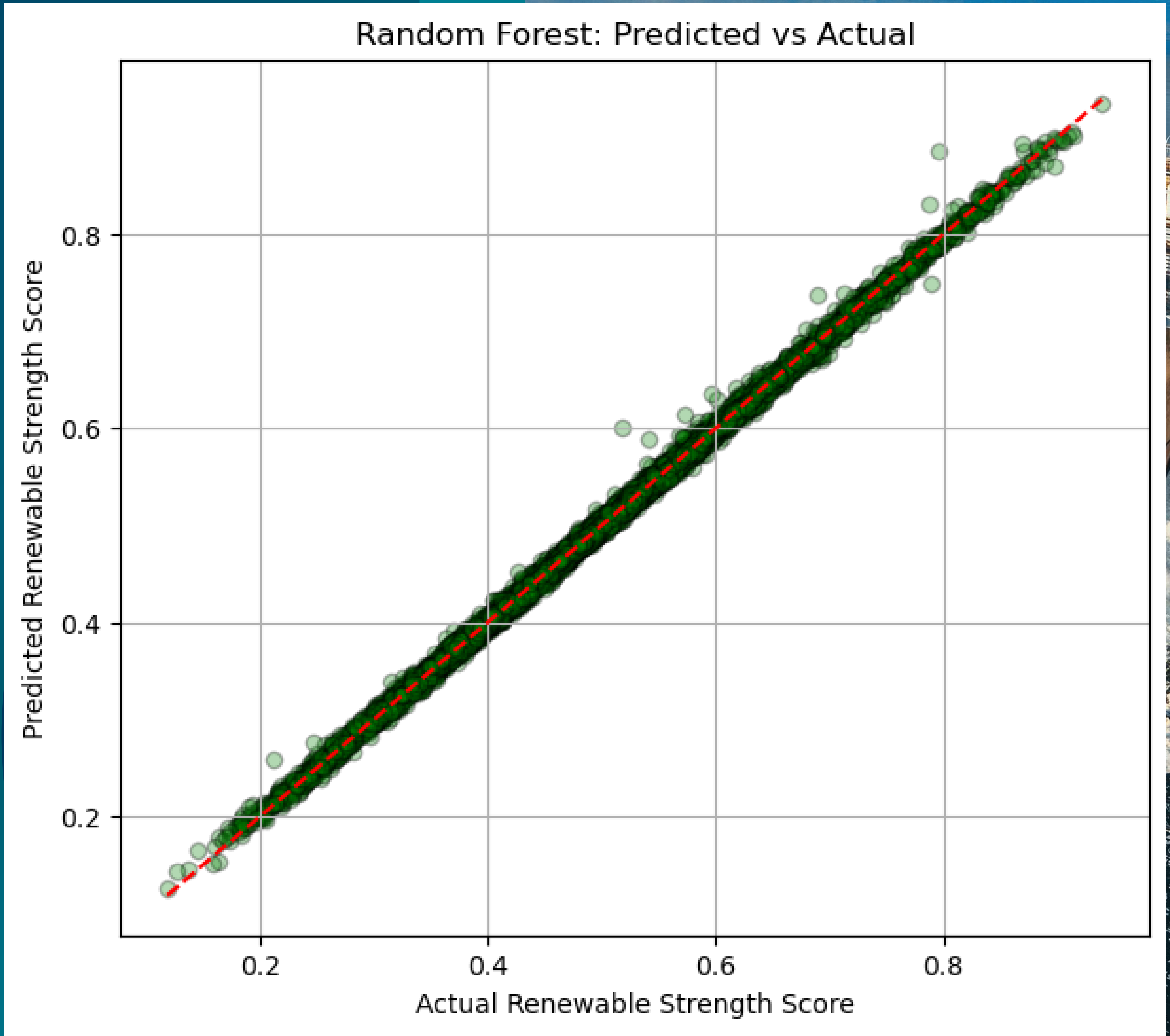
Linear Regression Model

- **R^2 Score: 1.00000**
  This means **100% of the variance** in the actual data is explained by the model. A perfect fit.
- **RMSE (Root Mean Squared Error): 0.00000**
  This indicates **no error** between predicted and actual values—again suggesting perfect prediction.



Random Forest Model

- **R^2 Score: 0.99863**
  This means that **99.86% of the variance** in actual scores is explained by the model—an **excellent** result indicating high model accuracy.
- **RMSE (Root Mean Squared Error): 0.00490**
  The average prediction error is very small, showing the model makes **very close predictions**.

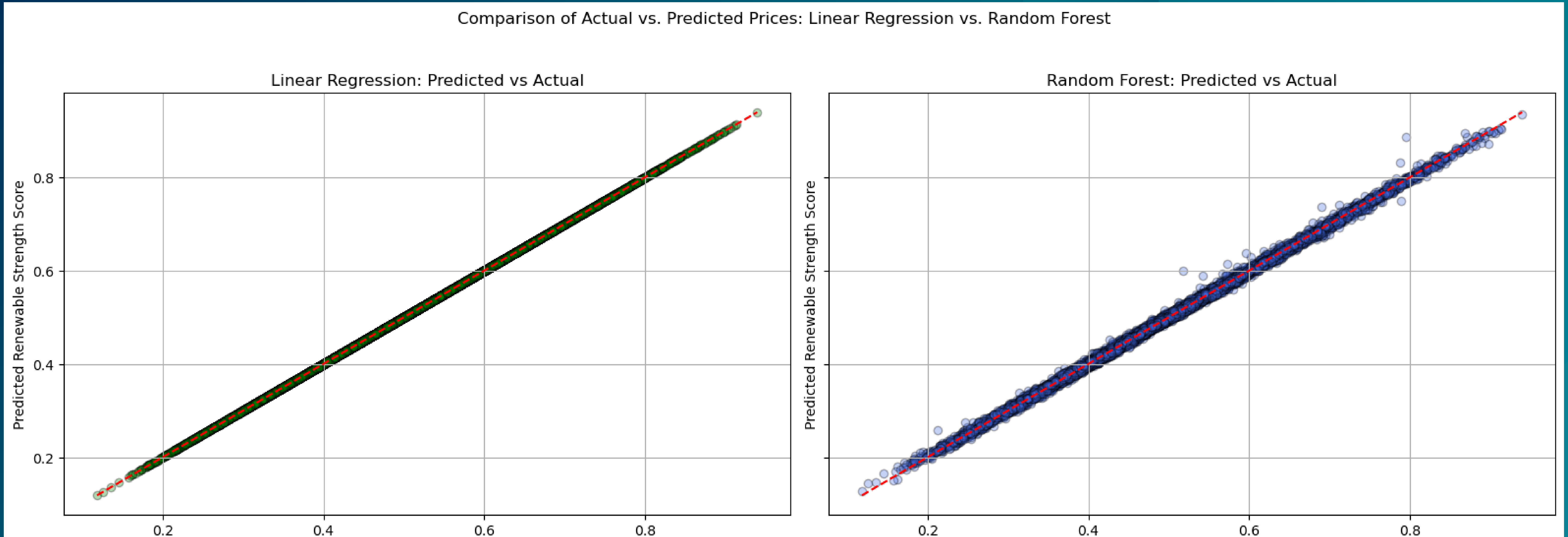


Comparison between Linear Regression and Random Forest

Overall Conclusion:

- **Linear Regression** fits the data **perfectly**, but may be too ideal if not properly validated.
- **Random Forest** shows **slightly less accuracy**, but is more likely to **generalize well** on unseen data due to its robustness and r capabilities.

Recommendation:
Use **cross-validation** or a **test set** to confirm whether the perfect score of the linear model is genuine or a case of overfitting.



Key Findings

- Steady increase in renewable output, especially wind and solar.
- Seasonal balance through wind, hydro, and solar.
- Decline in fossil fuel use, especially post-2020
- High accuracy in modeling renewable performance
- Reliable, well-prepared dataset with no missing or duplicate data

Investment and Policy Implications

Governments should prioritize grid modernization, storage, and incentives for renewables.

Private investors should consider solar, wind, and hybrid systems as scalable and profitable ventures.

(renewable_strength_score) can serve as a performance KPI for sustainability benchmarks.

Next Steps

- Integrate economic and weather data for sharper forecasting.
- Launch a public-facing dashboard to monitor trends.
- Extend the model for regional performance comparisons.
- Support education and advocacy based on the model's outputs.

Conclusion

Renewable energy is not just an alternative — it is the most logical, scalable, and sustainable energy path forward. With decreasing costs, year-round availability, and the ability to reduce emissions drastically, investing in renewables is a strategic move for both the public and private sectors. The use of renewable_strength_score empowers us to quantify, forecast, and advocate for this transition through sound data science and clear evidence.

Thank You For Your Attention