# SABANCI UNIVERSITY DSA 210 TERM PROJECT FINAL REPORT

ZİŞAN YEŞİL 32587

# The Data Behind Women Leaders and National Well-Being

1. **Abstract**

This study investigates the relationship between women’s political empowerment and national well-being across countries by combining data science techniques with a values-driven perspective. Using merged datasets from sources like the World Happiness Report, Human Development Index, and Women’s Political Empowerment Index, a composite well-being score was created from six key indicators. Through statistical tests and supervised machine learning models, including linear regression, decision trees, and random forests, the analysis reveals that women’s empowerment is not only positively correlated with development but also serves as a strong predictor of well-being.

1. **Indicators and Data Used**

To investigate the relationship between women’s political empowerment and national well-being, I used a combination of five publicly available datasets. Each dataset contributed unique indicators that were either used as input features or integrated into a composite well-being score. Below are the datasets and the variables selected, along with their justifications:

* **Human Development Index (HDI)**

HDI is a composite measure reflecting health, education and standard of living. It is used as a benchmark for national development.

* **Average Years of Schooling – Our World Data**

This indicator measures the average number of years of education received by people ages 25 and older. It reflects structural access to education and was included as a key control variable, given its known impact on both empowerment and well-being.

* **World Happiness Report – [UN Sustainable Development Solutions Network]**

Variables Used: Life expectancy, Freedom to Make Life Choices, Social Support, Trust in Government (Inverted Corruption)

These variables are core components of subjective well-being. I selected them to build a **composite well-being score** that goes beyond GDP, offering a more multidimensional view of human flourishing. The corruption indicator was inverted so that higher values always represent better governance.

* **Women’s Political Empowerment Index**

This index measures the actual influence women have in politics. It captures the substantive, rather than symbolic aspect of female political inclusion, and was a central feature in all models.

* **Women’s Political Participation Index**

This index reflects the degree to which women formally participate in political institutions. While it captures representation, it doesn’t always imply power.

1. **Project Process**

This project began with a personal and intellectual curiosity: *Does empowering women in politics lead to happier, more developed societies?* As someone passionate about gender equality and data science, I wanted to explore this question not just through intuition or ideology, but through empirical evidence. That’s why I came up with this project idea.

Then, I collected five publicly available datasets from trusted global sources like, UNDP, Our World in Data etc. Each dataset was selected to represent a critical dimension of political inclusion, development, or subjective well-being. The goal was to combine both *structural indicators* (like education and GDP) and *social indicators* (like trust, freedom, and support).

After collecting the data, I wanted to make the data usable, that is why I applied a data cleaning and choosing a sample process:

* I standardized country and year formats across datasets.
* I unified inconsistent column names such as “Country Name” to “Country”
* I handled non-numeric values and missing entries using filtering.
* For the composite well-being score, I used Min-Max normalization to scale all features between 0 and 1.
* The corruption variable was inverted so that higher values consistently reflected better conditions.
* I choose the sample as the years after 1990.
* GDP score was already log-transformed so that I used that transformed values.

After that, I merged all datasets on Country and Year, resulting in a unified dataset with 12 columns and over 1700 observations. This structured dataset became the foundation for all visual, statistical, and machine learning analyses.

To model national well-being in a multidimensional way, I followed the approach inspired by Martinez-Martin’s *Composite Global Well-Being Index (CGWBI)* (2016), which emphasizes the integration of both objective and subjective indicators of development. Based on this framework, I constructed a composite score using six variables: Human Development Index (HDI), Life Expectancy, Freedom, Social Support, Government Trust, and Corruption. These indicators were chosen to reflect a holistic view of human well-being, going beyond economic output.

Exploratory Data Analysis (EDA): I performed visual and statistical analysis to:

* Understand distributions of key variables
* Explore correlations between empowerment and well-being indicators
* Compare groups
* Calculate and visualize a Pearson correlation matrix

This step helped me identify patterns, outliers, and potential features for modeling.

Hypothesis Testing: To test whether differences across empowerment groups were statistically significant, I ran: Independent samples tests on HDI, schooling, and composite well-being.

Machine Learning Models: I trained three supervised learning models to predict the composite well-being score: Linear Regression, Decision Tree, Random Forest.

1. **Findings**
2. **Descriptive Analysis of Key Indicators:**

Before diving into hypothesis testing and predictive modeling, I conducted a descriptive analysis of the variables used in this project. This step was essential for understanding the underlying distributions, central tendencies, and variability across countries and years.

**Women's Political Empowerment Index**

Mean: 0.78 Range: 0.05 – 0.96

Most countries show moderate to high empowerment levels, though the wide range reveals strong variation across time and countries.

**Women's Political Participation Index**

Mean: 0.87 Range: 0.08 – 1.00

Participation levels are generally high across the globe. While this suggests active inclusion of women in political processes, the gap between participation and empowerment indicates that presence alone may not translate to influence. Women may be present in political spaces, but presence alone doesn’t mean power. It raises the possibility that many of these roles are symbolic. That’s not just a statistical detail; it’s a reflection of deeper issues around gender equality. *This realization stays with me. While I didn’t dive into this specific dynamic in this project, it’s something I’m eager to explore in the future.*

**Average Years of Schooling**

Mean: 8.87 years Range: ~1 – 14.3 years

Educational access varies greatly. Some populations exceed 14 years, while others barely reach 1, indicating deep global inequality.

**Human Development Index (HDI)**

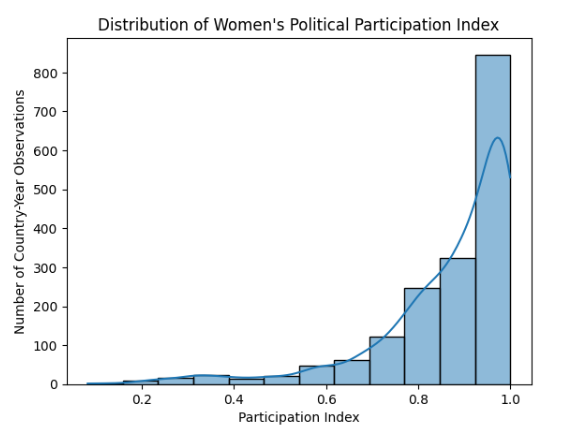
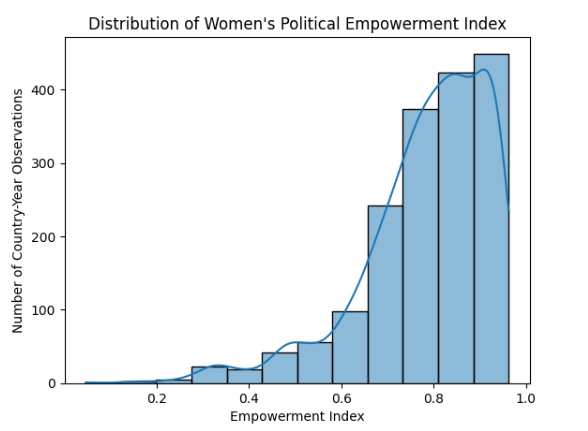
Mean: 0.72 Range: 0.30 – 0.96

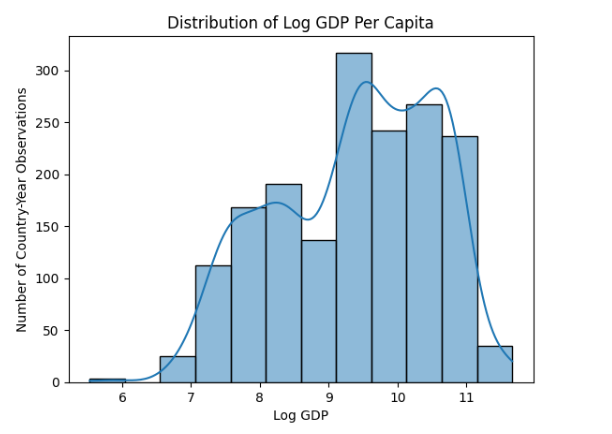
Most countries fall within medium to high development levels. HDI serves as a crucial variable when analyzing national well-being.

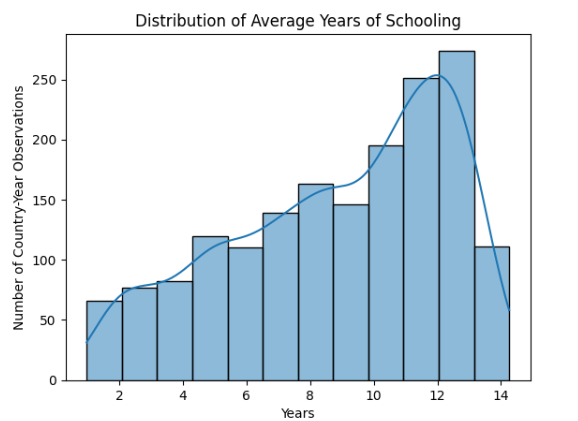
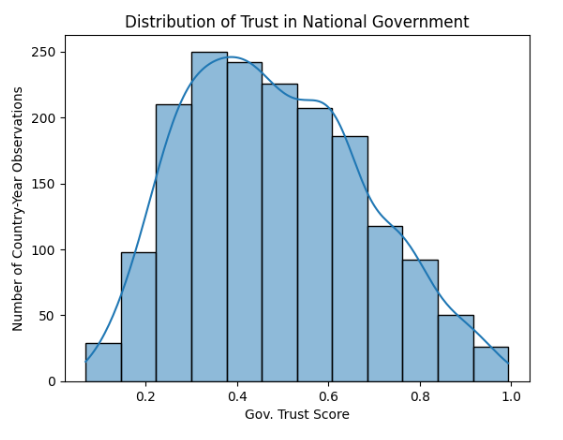
**GDP (Log Scale)**

Mean: 9.37 Range: 5.53 – 11.66

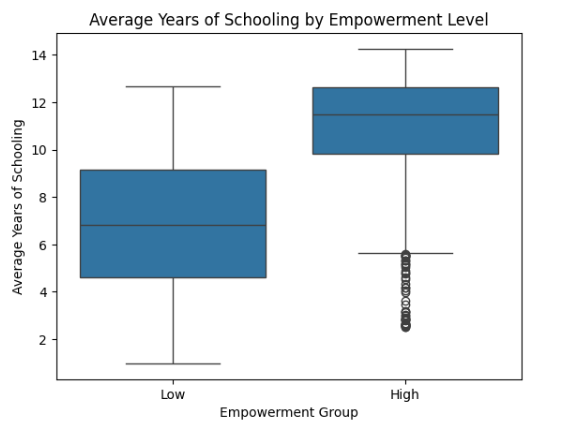
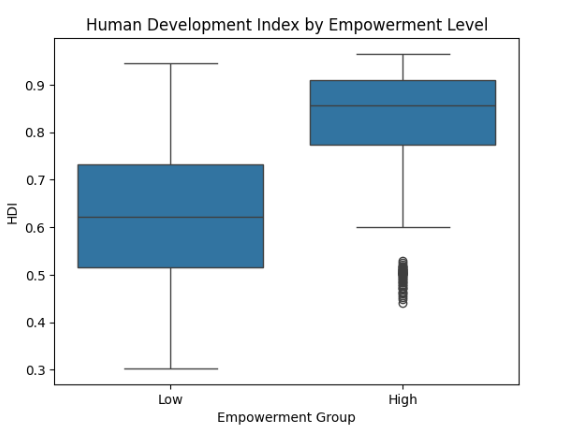
GDP is skewed by a few wealthy countries. It’s used as a control to better isolate the effect of women’s empowerment on well-being.

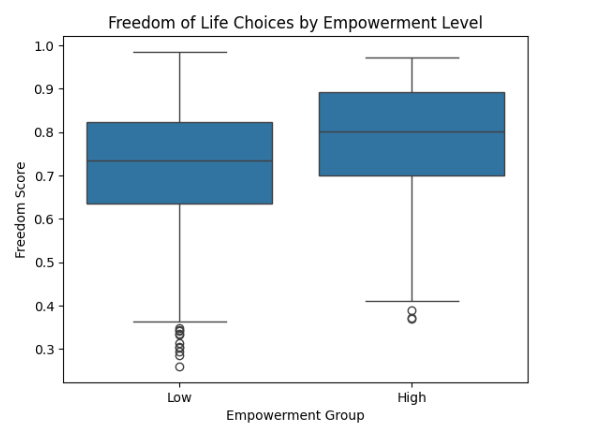
1. Histograms

The distributions reveal a global trend toward increased women’s political empowerment, with most values between 0.7 and 0.95. However, outliers with low scores remain. Political participation, on the other hand, is even more concentrated near 1.0, suggesting that while women are widely included in political processes, true empowerment is less common. This raises an important question: are women being symbolically included without real authority? I aim to explore how this gap may relate to national well-being outcomes like trust in government or corruption in the future of this project.



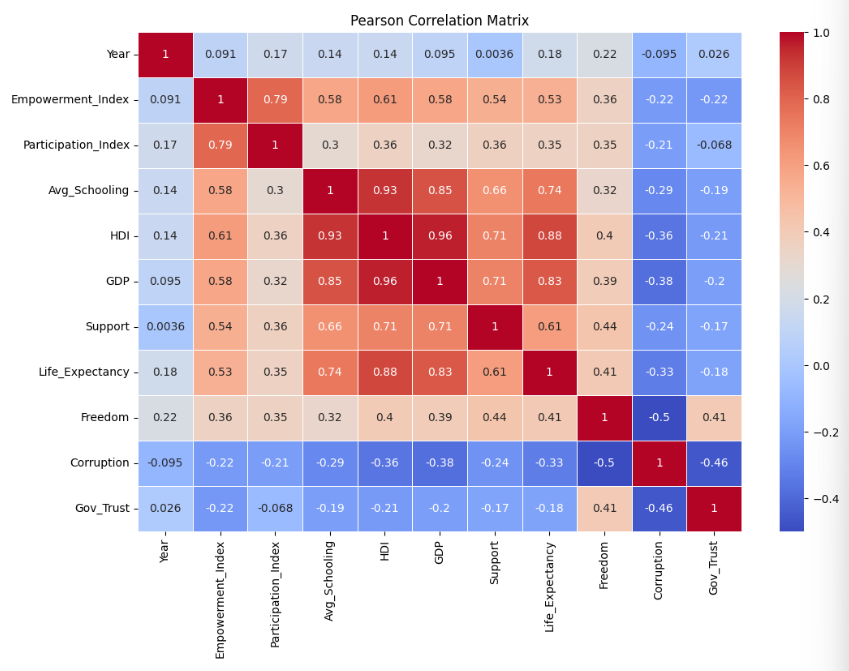
The distribution of **average years of schooling** shows that while many countries have reached moderate to high education levels, global inequalities persist, highlighting education’s role as a potential moderator between women’s empowerment and well-being. The **log GDP distribution** reflects wide variation in economic conditions, emphasizing the need to control for income levels in the analysis. Lastly, **trust in government** follows a bell-shaped distribution centered around moderate trust levels (0.4–0.5), suggesting that most countries fall in the middle. Investigating how trust interacts with empowerment, education, and GDP will be crucial in understanding their combined effect on national well-being.

1. Boxplots



The boxplots show that countries with higher women’s political empowerment also tend to have significantly better outcomes in **education**, **human development**, and **freedom of life choices**. Higher empowerment groups display higher medians and narrower ranges in all three indicators, suggesting not only better conditions but also more consistency. These patterns support the idea that empowering women politically aligns with broader societal well-being and reinforces the importance of education and development as foundational factors in fostering meaningful political inclusion.

1. Correlation Analysis



To better understand the relationships between women’s political empowerment and national well-being, I conducted a Pearson correlation analysis across key indicators. This section highlights the strength and direction of these associations.

1- Empowerment has strong positive correlations with:

* Participation Index (0.79)
* Average Years of Schooling (0.58)
* HDI (0.61) and GDP (0.58)

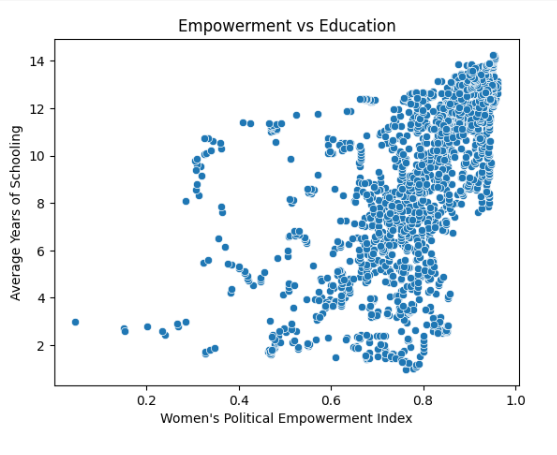
2- Empowerment has moderate positive correlations with:

* Life Expectancy (0.53)
* Freedom (0.36)

3- Empowerment has moderate negative correlations with:

* Corruption (-0.22)
* Government Trust (-0.22)

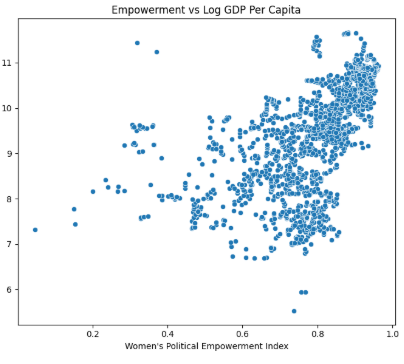
These results suggest that higher women’s political empowerment tends to align with stronger development, better education, and increased personal freedoms—while also being linked to lower corruption and distrust in government. This reinforces the idea that empowerment is deeply connected to broader societal well-being.

1. Scatter Plots

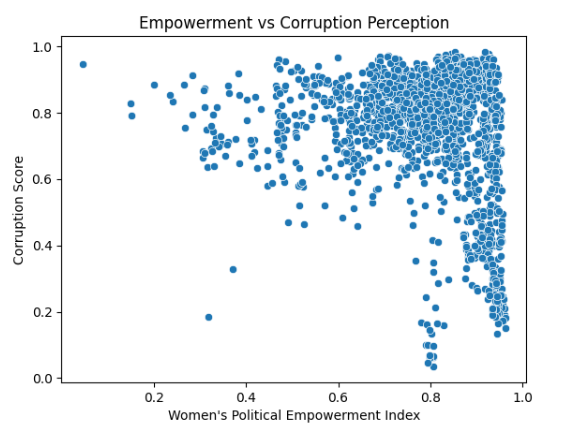
Countries with higher women's empowerment levels tend to have significantly higher average years of schooling. This supports the idea that education is not just an outcome of development, but also a foundation for political inclusion and long-term women's empowerment



Higher women’s political empowerment is strongly associated with higher HDI scores. This suggests that empowering women contributes to broader development outcomes and supports the view that inclusive leadership drives national progress.



This scatter plot shows a clear positive relationship between women’s political empowerment and log GDP per capita, suggesting that more empowered countries also tend to be more economically developed.



The moderate negative trend indicates that more women's empowered countries tend to report lower levels of perceived corruption showing that as women's empowerment increases, perceived corruption tends to decrease, pointing to a healthier governance environment.

Correlation analysis and scatter plots reveal that women's political empowerment is positively correlated with key development indicators such as education, HDI, and GDP. The visualizations show upward trends, confirming that countries with higher levels of women's empowerment tend to be more developed, educated, and economically stable. These relationships provide strong statistical and visual support for the idea that empowering women in politics contributes to broader national well-being.

1. Hypothesis Testing

* For Education:

H0 (Null Hypothesis): There is no difference in average schooling between high and low empowerment groups.

Ha (Alternative Hypothesis): There is a significant difference in average schooling between the two groups.



An independent samples t-test was conducted to determine whether the average years of schooling differs significantly between countries with low and high levels of women's political empowerment. The test yielded a t-statistic of -30.285 and a p-value < 0.005, which led us to reject the null hypothesis and say there is a significant difference in average schooling between two groups.

* For HDI:

Ho (Null Hypothesis): There is no difference in HDI between high and low empowerment groups.

Ha: There is a significant difference in HDI between the two groups.



The result is highly statistically significant, allowing us to confidently reject the null hypothesis. Since the high empowerment group has a higher average HDI, we conclude that countries with greater political empowerment for women also tend to have significantly higher levels of human development.

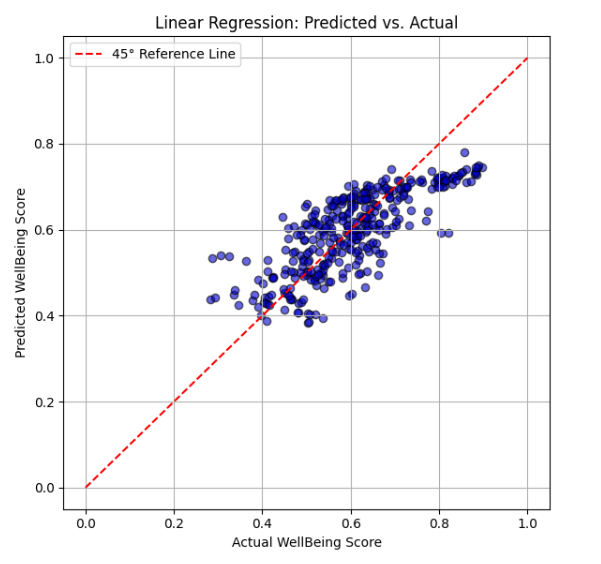
1. Machine Learning Analysis

* **Linear Regression Model**

****

A multiple linear regression model was trained to predict the composite well-being score using four features: Empowerment\_Index, Participation\_Index, Avg\_Schooling, and GDP. The dataset was split into training and test sets (80/20), and the model’s performance was evaluated on the test set using the R² score and Mean Squared Error (MSE).

The resulting R² score of approximately 0.59 which indicates that the model explains about 59% of the variance in national well-being. This suggests that women's leadership, when combined with structural factors like education and economic performance, has a meaningful association with overall well-being.

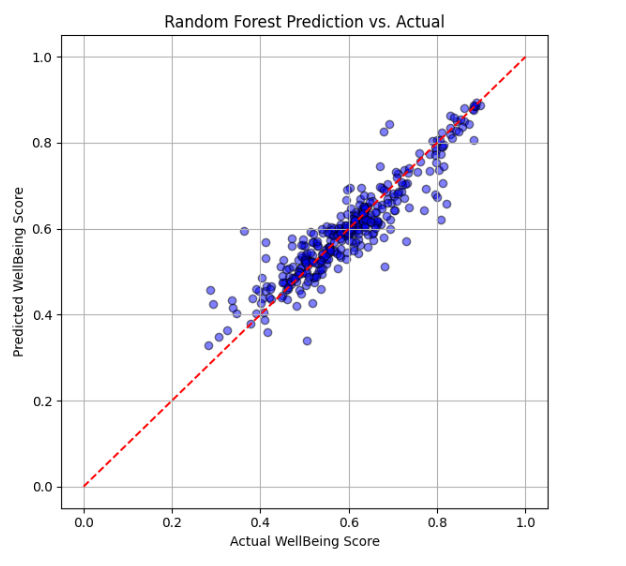


* **Random Forest Model**

****

To capture potential nonlinear relationships between features and well-being, a RandomForest Regressor was trained using the same four predictors: Empowerment\_Index, Participation\_Index, Avg\_Schooling, and GDP. Random Forests, being ensemble models based on decision trees, are well-suited for modeling complex interactions without requiring explicit feature transformations.

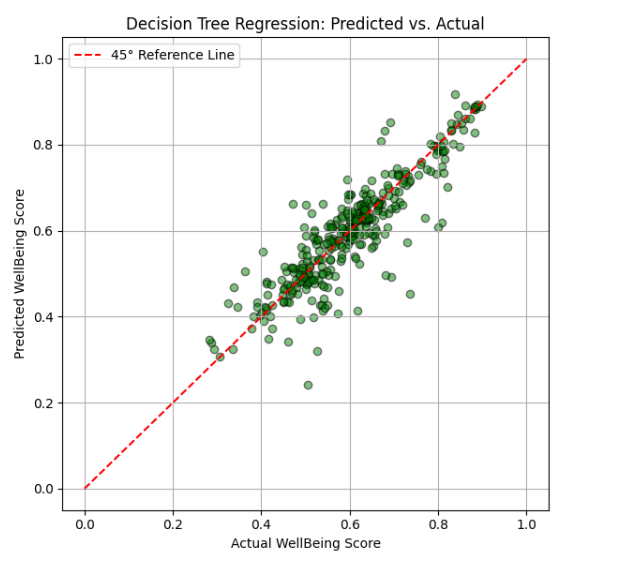
The model achieved a high R² score of approximately 0.85, significantly outperforming the linear regression model. This indicates that 85% of the variance in the well-being score can be explained by the selected features, suggesting a strong predictive relationship, especially when allowing for nonlinear effects.



* Decision Tree



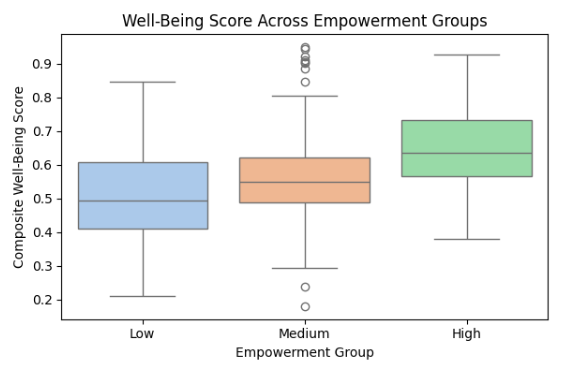
This evaluates the performance of the Decision Tree Regressor by predicting the well-being scores for the test set and calculating the R² Score and Mean Squared Error (MSE). An R² Score of approximately 0.77 indicates that the model explains a substantial portion of the variance in well-being scores, while the low MSE confirms good prediction accuracy.

****

The results from all three models demonstrate that women’s political empowerment, along with education and economic development, plays a significant role in predicting national well-being. While the linear regression model confirmed a meaningful linear relationship (R² ≈ 0.59), the decision tree and random forest models captured more complex patterns, with the random forest achieving the highest performance (R² ≈ 0.85).

1. Comparing Well-Being Across Levels of Women’s Empowerment

To examine how well-being or freedom varies with different levels of women’s empowerment, the continuous Empowerment\_Index was categorized into three ordinal groups: Low (≤ 0.6), Medium (0.6–0.8), and High (> 0.8). This transformation makes it easier to compare well-being outcomes across clearly defined empowerment levels.



This boxplot visualizes how the Composite Well-Being Score varies across Low, Medium, and High empowerment groups.

-As empowerment increases, both the median and distribution of well-being rise.

-The boxplot supports the statistical findings: higher empowerment aligns with greater national well-being.

1. **Conclusion**

This project set out to explore whether women's political empowerment contributes meaningfully to national well-being, and the results provide strong support for that hypothesis. Across multiple stages of analysis, from descriptive statistics and hypothesis testing to advanced machine learning models, a consistent pattern emerged: countries that politically empower women tend to be more educated, economically developed, and socially stable. Ultimately, this study supports the argument that political empowerment of women is more than a matter of representation or fairness, it is a strategic lever for national progress. Empowering women leads to better outcomes not only for individuals but for societies as a whole.

1. **Future Work**

Future research could build on this study by using longitudinal data to explore how changes in women’s empowerment over time impact national well-being. Adding variables like political regime type, gender-related policy measures, or media freedom could also deepen the analysis.

**Another key area is the gap between political participation and political empowerment. In the future I will examine this distinction in more detail, using qualitative or case-based approaches.**

Lastly, applying this framework to regional or sector-specific could reveal more targeted insights and policy implications.