# SABANCI UNIVERSITY DSA 210 TERM PROJECT PHASE 2

ZİŞAN YEŞİL 32587

# The Data Behind Women Leaders and National Well-Being

1. **INTRODUCTION**

This study aims to explore the connection between women's political empowerment and national happiness. The central question addressed is: Do countries with higher levels of women's leadership tend to be happier and more developed? Drawing from data science techniques such as correlation analysis, hypothesis testing, and visualization, this project seeks to uncover meaningful patterns and offer a data-driven narrative around the value of female leadership.

1. **MOTIVATION**

I believe that political empowerment is not just a matter of equality but also national prosperity. This project allows me to combine technical skill with purpose, exploring the message that diverse leadership fosters better societies. The analysis is intended to inform future policy decisions and highlight that supporting woman in leadership is a powerful lever for broader societal progress.

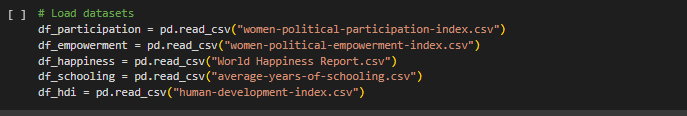
1. **DATA COLLECTION AND CLEANING**

This analysis uses five datasets from publicly available global sources:

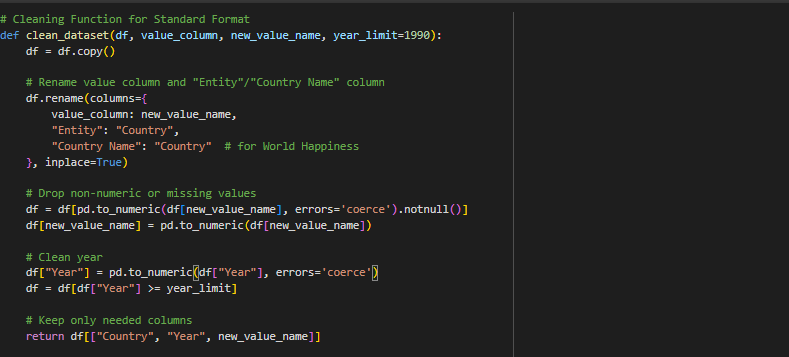
* World Happiness Report
* Women's Political Empowerment Index
* Women's Political Participation Index
* Human Development Index (HDI)
* Average Years of Schooling

The datasets were downloaded from reputable platforms such as Our World in Data and the World Happiness Report website. All of these accessed in CSV format and uploaded Google Colab for processing.



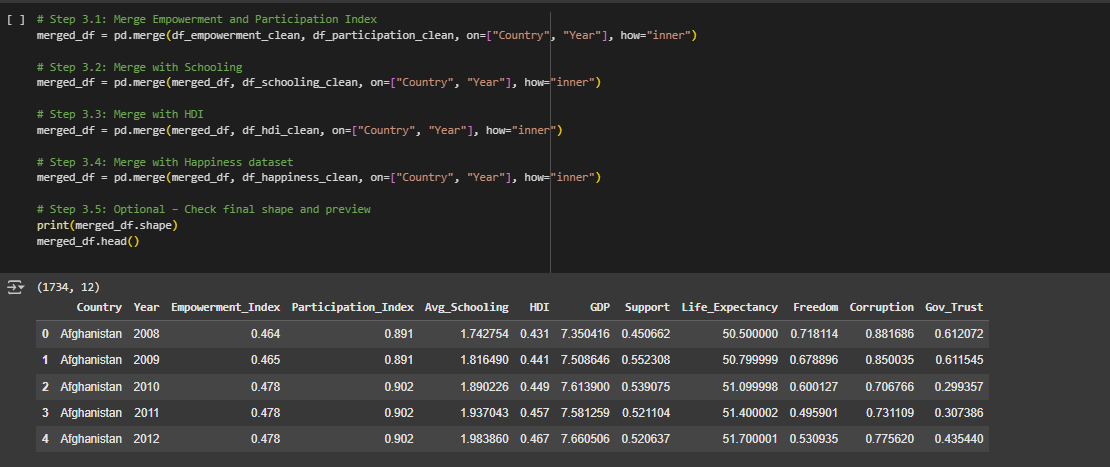


I know that, it is not good to use the raw data as it is. So that, to prepare the datasets for analysis, I used a standardized cleaning function and applied across all datasets. This function first unified inconsistent columns names such as "Entity" or "Country name" into a standard "Country" label and renamed the selected value column to a consistent name. It then handled missing and non-numeric values by assigning to NaN and converted all relevant columns to numeric types. Only rows from 1900 onward were retained to focus on more recent data. For each dataset, only the essential columns were preserved. This process ensured that all datasets shared a uniform structure, enabling them to be accurately merged and compared in merging.

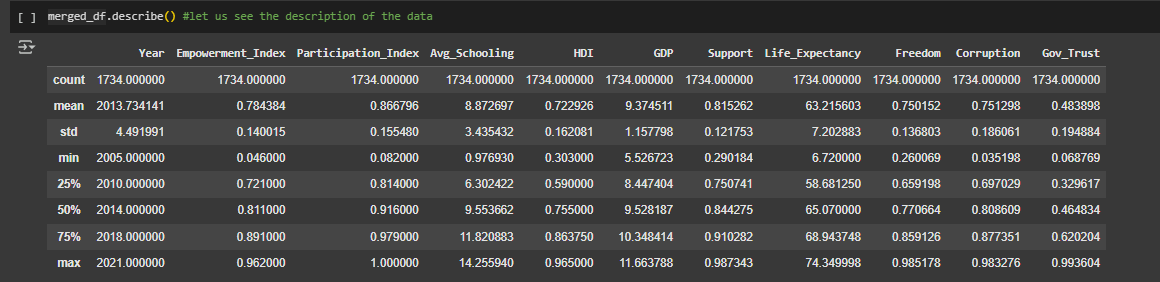




All datasets were then merged using inner joins on Country and Year, resulting in a combined dataset of 1,734 observations across 12 variables, ready for exploratory and inferential analysis.



1. **EXPLORATORY DATA ANALYSIS**



To begin the analysis, the merged dataset was first summarized using descriptive statistics. This step revealed the central tendencies (means and medians) and variability (standard deviations) of key variables such as women's empowerment, schooling, HDI, GDP, and happiness-related indicators: **Women's Political Empowerment Index**

Mean: 0.78

Range: 0.05 – 0.96

Explanation: Most countries demonstrate moderate to high levels of political empowerment for women, with the median value at 0.81. However, the wide range of values across country-year pairs indicates that political empowerment for women has varied significantly over time and across countries.

**Women's Political Participation Index**

Mean: 0.87

Range: 0.08 – 1.00

Explanation: 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.

**Average Years of Schooling**

Mean: 8.87 years

Range: ~1 – 14.3 years

Explanation: The data reflects significant inequality in education. This can also be seen with the standard deviation (approximately 3.43). While some populations receive more than 14 years of schooling on average, others barely exceed 1 year. Education will be analyzed as a structural factor possibly affecting both empowerment and well-being.

**Human Development Index (HDI)**

Mean: 0.72

Range: 0.30 – 0.96

Explanation: Most countries fall within the medium to high development range. HDI is expected to be an important control variable in later analyses, especially when modeling happiness or trust in governance.

**GDP (Log Scale)**

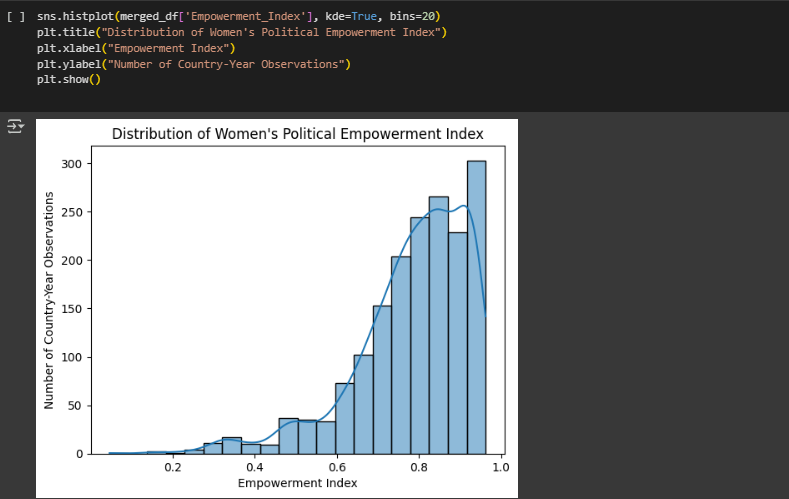
Mean: 9.37

Range: 5.53 – 11.66

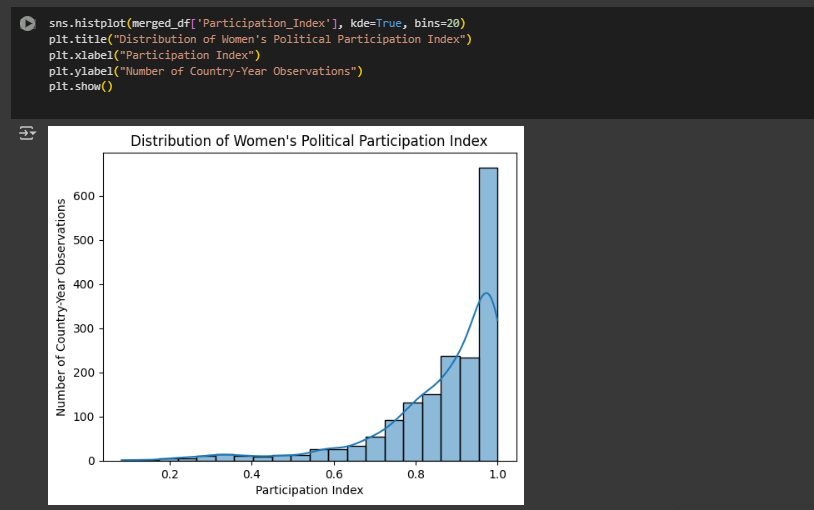
Explanation: There is a skewness toward higher-income countries, with a few wealthy nations significantly raising the global average. Since GDP influences well-being, it will be included in predictive models to isolate the effect of women's empowerment.

Histograms were then generated for each variable to explore their distribution shapes. These plots revealed whether variables were normally distributed, skewed, or had outliers.

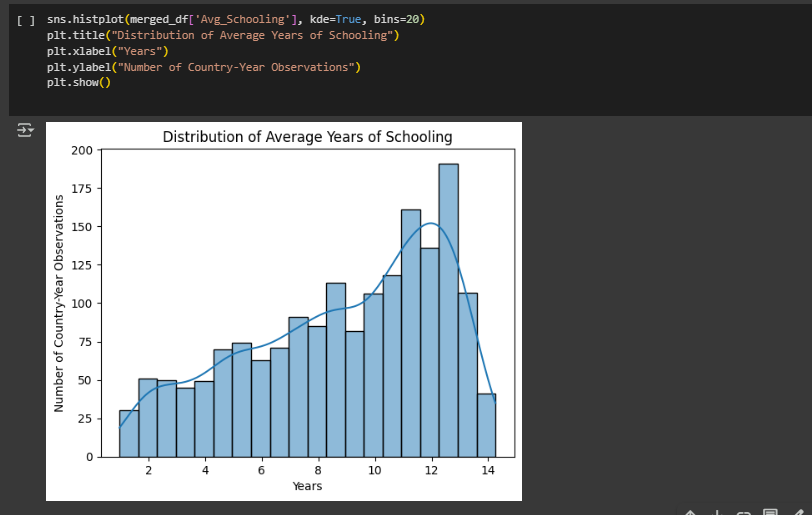
* The distribution of Women’s political empowerment highlights a global trend toward greater women's political empowerment, although there are still outlier cases with low scores. In addition to that, most values are concentrated between 0.7 and 0.95, indicating that in many countries and years, women have had relatively high political empowerment.



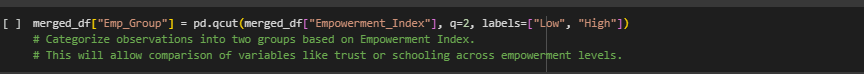
* The distribution of Women’s Political Participation Index shows that there is a strong right skew, with the majority of values clustering between 0.85 and 1.0. This suggests that in many cases, women are formally participating in political processes. Compared to the previous histogram on Empowerment Index, this distribution is more concentrated at the top end, implying that political participation is more widespread than true political empowerment.

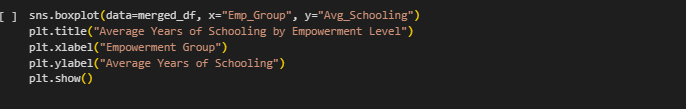


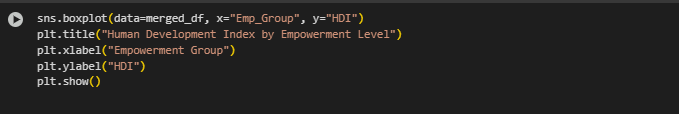
* The distribution of Average Years of schooling indicates that while a large portion of the world seems to be achieving moderate to high educational levels, the data still reflects structural inequalities. This uneven distribution may be a key factor moderating the relationship between women's empowerment and broader societal well-being.

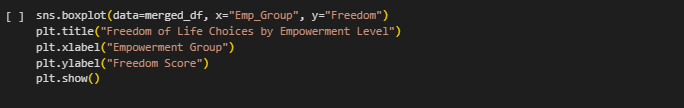


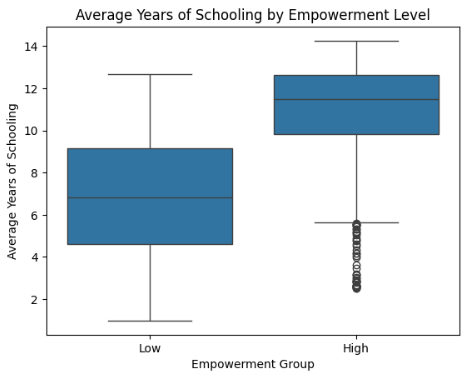
Boxplots were created to compare distributions of key indicators across countries grouped by empowerment levels (low vs. high). These visualizations illustrated not just median differences, but also the variability and presence of outliers.



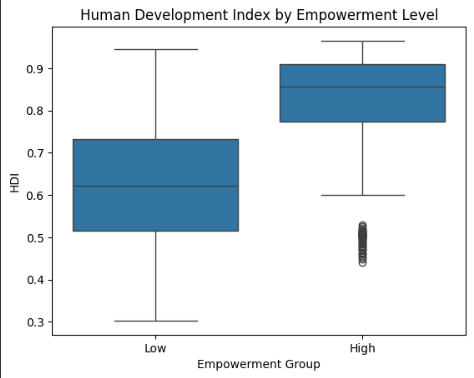




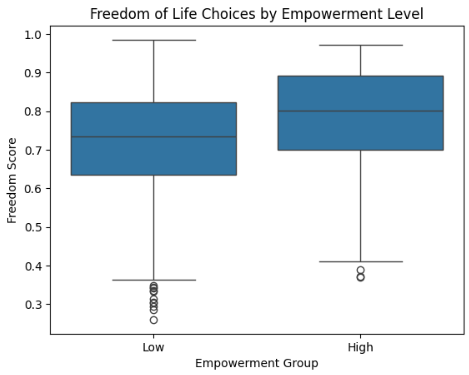




This supports the idea that education plays a foundational role in fostering women’s political empowerment. Countries with higher schooling levels are more likely to empower women in meaningful political roles, reinforcing the link between education and women's empowerment level.

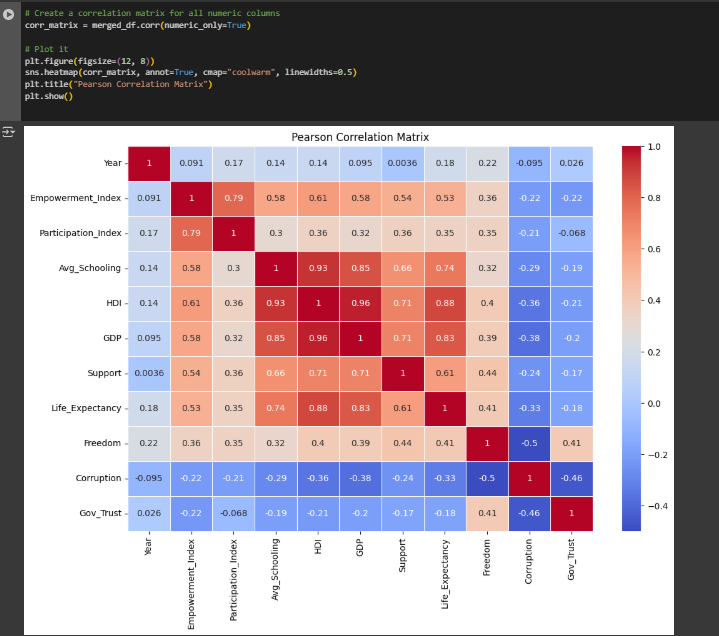


This bar chart confirms a strong positive relationship between women’s political empowerment and national development. Countries that politically empower women tend to perform significantly better on core human development indicators, supporting the idea that empowering women contributes to overall well-being



This visual suggests that higher political empowerment of women is positively associated with perceived personal freedom in a society. The high empowerment group has a higher median freedom score, centered around 0.85, and most values lie within a narrower, upper range.

To assess relationships between variables, a Pearson correlation matrix was calculated and visualized using a heatmap.



**Correlation Analysis Summary**

The pearson correlation matrix highlights several key relationships between women's political empowerment and national well-being indicators:

1- Empowerment has strong positive correlations with:

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

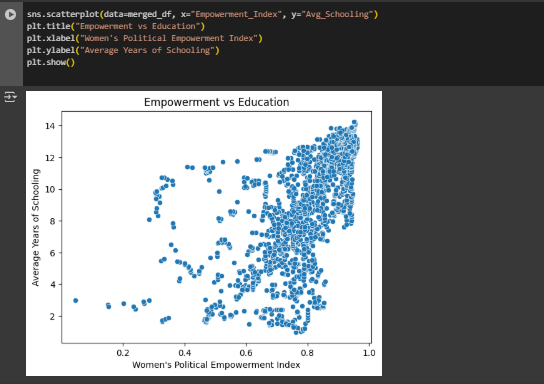
2- Empowerment has moderate positive correlations with:

* Life Expectancy (0.53)
* Freedom (0.36)

3- Empowerment has moderate negative correlations with:

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

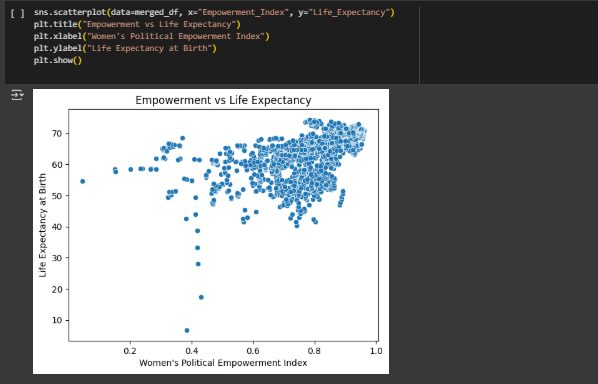
Scatter plots were also used to visually confirm these associations, particularly between Empowerment and HDI, and Empowerment and GDP. These plots showed upward trends, indicating that increases in empowerment are generally associated with improvements in national well-being indicators.



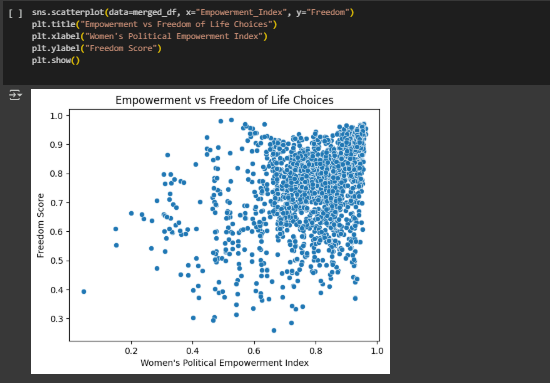
Countries with higher women's empowerment levels tend to have significantly higher average years of schooling. This supports the idea that education is not just an outcome of development, but also a foundation for political inclusion and long-term women's empowerment.



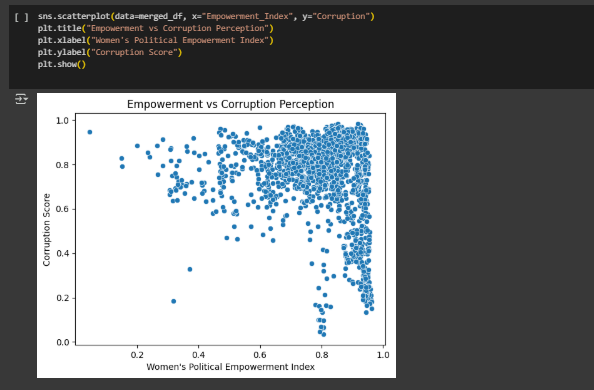
Both HDI and GDP are strongly and positively correlated with empowerment. The scatter plots reveal that countries with more empowered women tend to be more economically prosperous and socially developed.



Life expectancy increases with women's empowerment, which may reflect better access to healthcare, stability, and long-term policy focus in more inclusive societies.



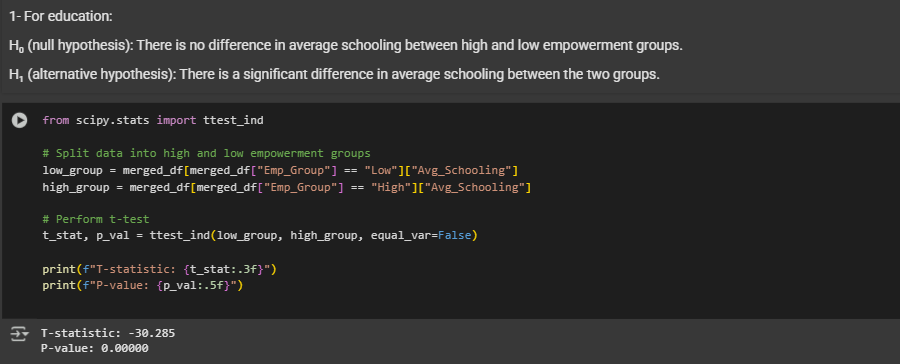
While the relationship is not as strong as the other indicators, there's a positive trend which suggests that political empowerment may also influence perceived personal freedom.



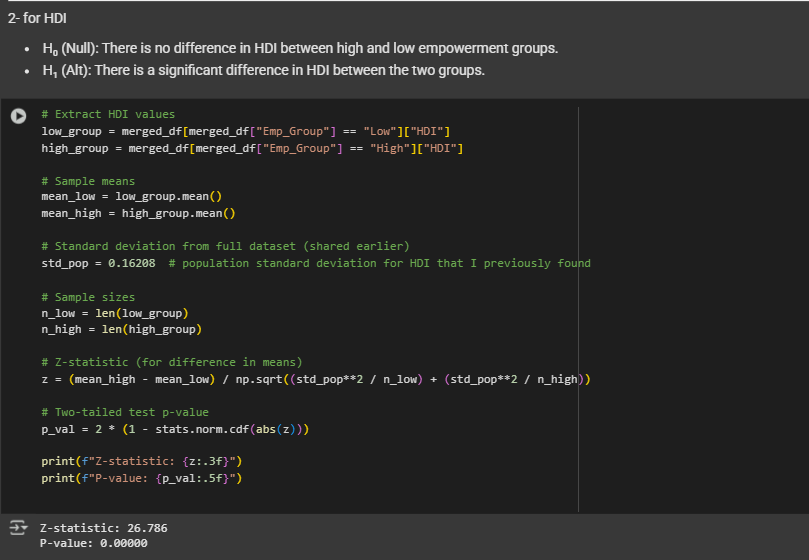
The moderate negative trend indicates that more women's empowered countries tend to report lower levels of perceived corruption showing that as women's empowerment increases, perceived corruption tends to decrease, pointing to a healthier governance environment.

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



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 leds us to reject the null hypothesis and say there is a significant difference in average schooling between 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. CONCLUSION

This project provides robust evidence that women's political empowerment is positively associated with national development and components of happiness. Countries with higher empowerment consistently demonstrate better outcomes in education, development, freedom, and economic prosperity. While happiness was measured indirectly through its components, the patterns support the idea that empowering women is an effective path toward building stronger, more balanced nations.

1. Reflection and Future Work

Regression models and machine learning methods will be included in future extensions of the study to explore predictive insights and model variable importance in shaping national well-being. Besides, additional variables such as gender-based policy efforts, political regime types, or cultural dimensions could also be included to enrich the analysis and uncover new layers of insight.