Submitted: September 2025

Data Science Lab Project

Università degli Studi di Milano-Bicocca

CdLM in Data Science

Keywords

panel data, fixed effects, robust standard errors, political freedom, well-being, World Bank, Freedom House

University ID Numbers

Negri - 809625
Licciardello - 889935
Biswas - 925349
Issa - 922689

Francesca Negri, Gianluca Licciardello, Arnab Biswas, Farabi Issa

The Correlation between Well-Being and Freedom: Empirical Evidence and Forecasting

Abstract

This project aims to study and exploit the association between political freedom and human well–being using a panel of 193 countries from 2006 to 2023. After harmonization and imputation, we construct a composite well–being index and assess its link to Freedom in the World scores through correlations, regression models, and predictive forecasting. Correlation tests (Pearson $r \approx 0.75$, $p < 0.001$) reveal a strong positive association, robust to partial correlations controlling for socio–economic indicators. Pooled OLS explains 58% of the cross–country variance ($\beta \approx 0.035$, $p < 0.001$), while two–way fixed effects confirm a smaller yet significant within–country effect ($\beta \approx 0.008$, $p = 0.004$). Predictive models further support these findings: a simple linear regression explains 31% of WB variance ($R^2 = 0.307$), polynomial Elastic Net improves accuracy to 45% ($R^2 = 0.447$), and histogram gradient boosting achieves competitive performance ($R^2 = 0.413$). Dynamic panel forecasts to 2025 and 2030 suggest that countries maintaining higher freedom are likely to sustain higher well–being levels, whereas those with persistent restrictions risk stagnation. While observational in nature, the evidence consistently points to freedom as a robust predictor and correlate of well–being. By using the same tools we built in this study for monitoring trends in political freedom, policymakers can recognize it as a critical factor that may influence future well-being outcomes and incorporate it into long-term planning and decision-making.

Contents

Introduction

Goal & Motivation

Well-being and **democratic freedom** are two key topics when talking about a country, and they can be analyzed from different perspectives such as income, work, housing, health and access to services, but also environment, education, security, civic engagement, governance and community. Both can, on their own, provide a perfect understanding of the main topic they cover (the Well-being composite speaks to lived outcomes, while the other one captures rights and institutions), but what happens when you combine them? Maybe, **together they could offer a socially relevant lens that citizens, journalists, and policymakers actually use**. Our motivation is pragmatic rather than causal: to summarize the empirical link between Well-being and Freedom and to provide a simple, maintainable tool that can turn future Freedom information releases into transparent, scenario-based expectations for Well-being.
With these objectives in mind, we formulated proper research questions.

Research Questions

- Do freedom and political rights correlate with well-being outcomes?
- Which well-being dimensions are most strongly associated with freedom?
- If such associations exist, how well can we predict near-term well-being from freedom indicators, and what do scenario-based forecasts for 2025 and 2030 suggest?

These questions are later translated into hypotheses and models, and we will see whether they will be confirmed or refuted.

Datasets overview

World Bank's ESG Data

The World Bank's *Environment, Social and Governance* (ESG) dataset (later referred as "WB") provides annual indicators related to **environmental sustainability, social development and governance performance** for over **200 countries**. The **Environment pillar** measures the sustainability of a country's economic performance given its natural resource endowment, management, its risk or resilience to climate change and other natural hazards. The **Social pillar** quantifies the sustainability of a country's economic performance with regard to its efficacy in meeting the basic needs of its population. The **Governance pillar** describes the sustainability of a country's economic performance in the context of its institutional capacity to support long-term stability, growth and poverty reduction.

Chosen features To identify the most relevant variables for the analysis, a feature selection process was carried out combining theoretical considerations and exploratory data analysis. Starting from the full set of **71 ESG indicators** provided by the World Bank, we applied a hybrid approach involving **domain-driven filtering and unsupervised techniques** (variance-based screening, principal component analysis (PCA), and hierarchical feature clustering). This allowed us to retain a **subset of 10 indicators** that are both representative of the environmental, social, and governance dimensions and exhibit meaningful variability and informational value across countries and years. The process is fully explained in details in the *Appendix: Supplementary Methodological Notes*, and the chosen features are reported on Table 1. Descriptive statistics of the 10 indicators are reported on Table 2.

Some facts that emerge about those features:

Original Feature	*Renamed as*	Description
School enrollment, primary (% gross)	*Education*	Measures access to primary education across the population.
Unemployment, total (% of labor force)	*Unemployment*	Indicates overall labor market health and opportunities.
Gini index	*Income*	Reflects income inequality within a country.
Rule of Law: Estimate	*Safety*	Captures confidence in legal systems and contract enforcement.
Political Stability and Absence of Violence/Terrorism: Estimate	*Political Stability*	Assesses risk of political instability and violence. +
Life expectancy at birth, total	*Life Expectancy*	Summarizes overall health and development outcomes.
PM2.5 air pollution, mean annual exposure	*Pollution*	Measures exposure to fine particulate air pollution.
Voice and Accountability: Estimate	*Civic Engagement*	Indicates citizens' participation and freedom of expression.
Individuals using the Internet (% of population)	*Digital Access*	Represents digital access and technological inclusion.
Government Effectiveness: Estimate	*Community*	Evaluates public service quality and policy implementation capacity.

Feature	Unit
Education	%
Unemployment	%
Income (Gini index)	Index (0–100)
Safety	Index (-2.5 – 2.5)
Political Stability	Index (-2.5 – 2.5)
Life Expectancy	Years
Pollution	$\mu g/m^3$
Civic Engagement	Index (-2.5 – 2.5)
Digital Access	%
Community	Index (-2.5 – 2.5)

Table 1. Selected indicators with renamed labels and descriptions. Below, the measure units: there are 3 percentage features (*Unemployment, Digital Access, Education*), one indicator is expressed in years (*Life Expectancy*), another is a physical quantity (*Pollution*) and the remaining 5 are indices, four from -2.5 to 2.5 and one (*Income*) from 0 to 100.

1) **Civic Engagement, Community, Confidence** and **Political Stability** have mean ~ 0 and standard deviation ~ 1, meaning they are already standardized.
2) **Education** is a percentage but the mean exceeds 100: that's because this feature is actually a **Gross enrollment ratio**, that is, a ratio between the number of registered people (regardless of age or origin) and the population in the official age group (times 100). That means that the final value can exceed 100 because, for example, there can be foreigner students or enrolled students above or below the age group.
3) Even though it's not explicitly reported in table 2, indicators also exhibit **varying degrees of completeness**, with Income having the highest rate of missing values ($\approx 78\%$) and Unemployment the lowest ($\approx 6\%$). These were handled during preprocessing.

In Figure 1, the correlation matrix and four scatterplots of correlations are reported.

Indicator	Mean	Std	Min	25%	50%	75%	Max
Civic Engagement	-0.05	1.00	2.31	0.89	0.00	0.85	1.77
Pollution	29.87	18.18	5.16	15.80	24.07	43.16	95.24
Income	35.64	7.72	23.20	29.50	34.30	40.20	57.50
Education (%)	101.94	11.44	44.18	97.92	101.34	106.32	151.73
Life Expectancy (years)	70.44	8.10	42.91	64.65	71.91	76.44	84.56
Community	-0.08	0.99	-2.44	-0.78	-0.20	0.55	2.47
Confidence	-0.08	0.99	-2.59	-0.82	-0.24	0.62	2.12
Political Stability	-0.07	0.99	-3.31	-0.69	0.01	0.78	1.67
Unemployment (%)	7.55	5.28	0.10	4.08	6.06	9.48	36.39
Digital Access (%)	42.94	30.04	0.18	14.99	39.81	70.32	100.00

Table 2. Descriptive statistics of selected ESG indicators

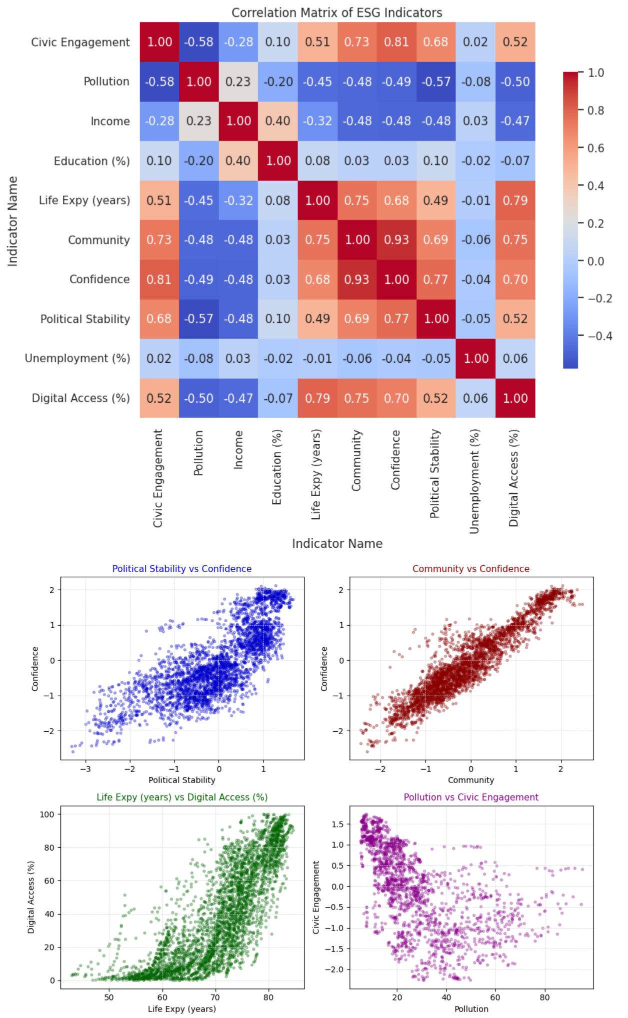

Figure 1. Above, correlation matrix shows that **Community** and **Confidence** have the most influence - strong and positive correlation with **Life Expectancy**, **Political Stability**, **Civic Engagement** and between each other, and also moderate and negative with **Pollution** and **Income**. **Education** and **Unemployment**, instead, show a corr. index of ~ 0 meaning that they are uncorrelated with the remaining features. Below, scatterplots of selected strong correlations in the dataset. **Confidence shows clear positive associations with Political Stability, Community, and Life Expectancy (via Digital Access)**, as reflected by the narrow, upward-sloping data clouds. The more the index goes toward 1, the narrower is the distribution. In contrast, **Civic Engagement is moderately and negatively correlated with Pollution**: lower engagement corresponds to higher pollution levels.

Freedom in The World

The *Freedom in the World* report (later referred as "FW"), the most widely read and cited of its kind, is edited by **Freedom House**, an American organization devoted to the support and defense of democracy around the world, and tracks global trends in political rights and civil liberties for over **50 years**. it's composed of numerical ratings and supporting descriptive texts for **195 countries and 13 territories**, collected by using a combination of on-the-ground research, consultations with local contacts, information from news articles, nongovernmental organizations, governments, and a variety of other sources. For each country and territory, Freedom in the World analyzes the **electoral process, political pluralism and participation, the functioning of the government, freedom of expression and of belief, associational and organizational rights, the rule of law, and personal autonomy and individual rights**.

Dataset structure The dataset presents a hierarchical structure that can be divided in two main categories (or "Aggregates"), **Civil Liberties (CL)** and **Political Rights (PR)**. Below them, there are the subcategories, with codes A-G: subcategories **A, B, C** are related to PR, subcategories **D, E, F, G** to CL. Each of them has a score, which can reach 12 or 16 depending on the subcategory. The particular thing about the dataset is that the subcategories' indicators are not separate: **each subcategory contributes to the aggregate score of its main category** (the so-called "PR Aggregate" and "CL Aggregate"). Aggregate PR can reach up to 40 points, while Aggregate Cl can reach 40. It goes without saying that the *Total Aggregate* has, therefore, a range 0-100. In table 3 all subcategories, along with their main category, their description and their score, are summarized.

Table 3. Freedom in the World (FW): Subcategories and Aggregates.

Code	Subcategory	Points	Category	Brief Description
A	Electoral Process	0–12	PR	Elections, suffrage, competitiveness, transfer of power.
B	Political Pluralism & Participation	0–16	PR	Parties, opposition space, minority participation.
C	Functioning of Government	0–12	PR	Government accountability, corruption, policy effectiveness.
D	Freedom of Expression & Belief	0–16	CL	Media, academic, religious freedom; personal expression.
E	Associational & Organizational Rights	0–12	CL	Assembly, civil society, labor rights.
F	Rule of Law	0–16	CL	Judiciary independence, due process, equality before law.
G	Personal Autonomy & Individual Rights	0–16	CL	Movement, property, social freedoms, personal choices.
Discretionary questions		0–4 ea.	PR/CL	*Add Q* (B) and *Add A*: applied when warranted.
PR aggregate		A + B + C (+ discretionary), range 0–40 (obs. min ≈ –3)		
CL aggregate		D + E + F + G, range 0–60		
Total aggregate		PR + CL, range 0–100		

FW report also provides two additional rating columns, based on the score achieved in each of the two categories. They are called **PR Rating** and **CL Rating** and the value inside them are calculated for every country according to the PR/CL Aggregate they reached, using the scale reported in table 4.

Table 4. Conversion of aggregate scores into 7–point ratings for Political Rights (PR) and Civil Liberties (CL). Lower ratings indicate higher freedom.

Interpretation	Rating	PR aggregate score	CL aggregate score
Most free (highest rights/liberties)	1	36–40	53–60
High level of rights/liberties	2	30–35	44–52
Moderate rights/liberties	3	24–29	35–43
Mixed situation	4	18–23	26–34
Limited rights/liberties	5	12–17	17–25
Very restricted rights/liberties	6	6–11	8–16
Least free (lowest rights/liberties)	7	0–5	0–7

Descriptive statistics of the dataset are reported in Table 5, and there are no missing values, reflecting a well-maintained dataset. It is certainly worth noting that the average PR/CL Rating is around 3.4: **the world is closer to "partially free" than to "completely free"**. Furthermore, if we analyze the PR subcategories, the B (Political Pluralism) has the highest mean ($\sim$ 10 out of 16), while the C (Government Functioning) has the lowest ($\sim$ 6 out of 12), and that leads us to suppose a **persistence of formal democratic institutions that are not always matched by effective governance**. In general, both PR and CL show high variances ($\sim$ 13 and $\sim$ 17 respectively), meaning that there are large regional differences (see chart 4).
In the end, the initial descriptive statistics of the dataset reveal a **polarized global landscape**: while the average country scores around 60 out of 100, values range from nearly zero to the maximum, indicating **substantial disparities between consolidated democracies and authoritarian regimes** (see chart 2).
There are also few negative values that are probably outliers or input errors, and that is addressed in a preprocess step.

Table 5. Descriptive statistics of FW indicators (2006–2023)

Indicator	Mean	Std	Min	25%	50%	75%	Max
Edition	2015	5.47	2006	2010	2015	2020	2023
PR Rating	3.43	2.18	1.0	1.0	3.0	6.0	7.0
CL Rating	3.30	1.89	1.0	1.0	3.0	5.0	7.0
A	7.59	4.38	0.0	3.0	9.0	12.0	12.0
B	9.94	5.29	0.0	5.0	11.0	15.0	16.0
C	6.35	3.72	0.0	3.0	7.0	10.0	12.0
Add Q	0.10	0.52	0.0	0.0	0.0	0.0	4.0
Add A	0.04	0.34	0.0	0.0	0.0	0.0	4.0
PR	23.82	13.17	-3.0	11.0	27.0	36.0	40.0
D	10.99	4.53	0.0	8.0	12.0	15.0	16.0
E	7.61	3.88	0.0	4.0	8.0	11.0	12.0
F	8.23	4.84	0.0	4.0	8.0	13.0	16.0
G	9.56	4.10	0.0	6.0	10.0	13.0	16.0
CL	36.39	16.73	0.0	23.0	37.0	53.0	60.0
Total	60.20	29.62	-1.0	34.0	64.0	89.0	100.0

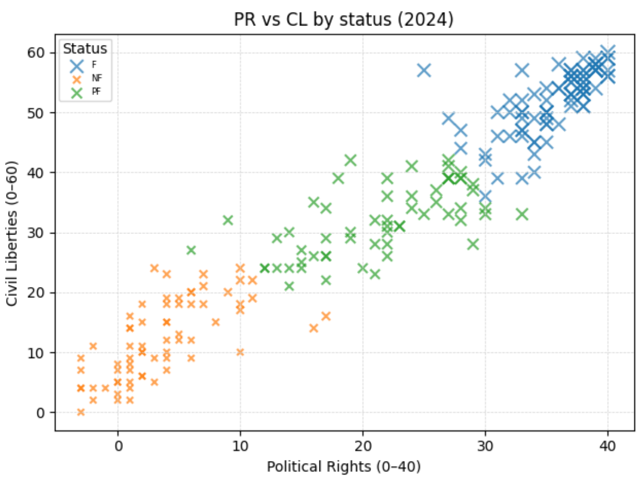

Figure 2. Here the polarization of the countries based on PR and CL ratings: the Not-Free ones (NF) have ratings lower than 20, while the Free (F) ones are concentrated in the last 10 points of both scales. The partially Free (PF) are distributed in the middle.

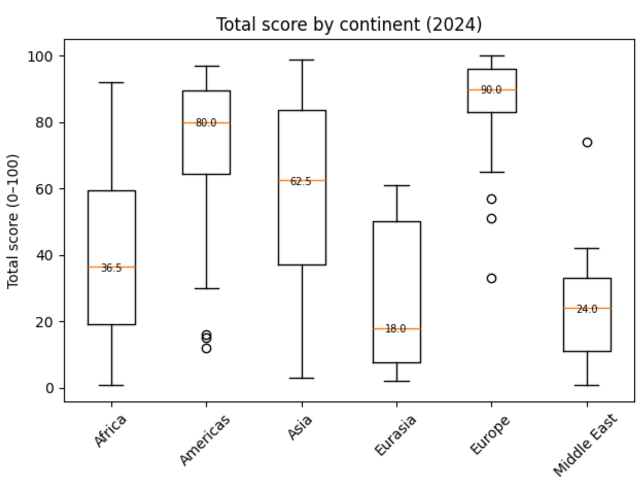

Figure 3. This boxplot chart show the difference in distributions of the Total Aggregate across the continents. The orange line represents the median, and its value is printed above the line. Given that Total is PR+CL, the chart clearly reflects the difference among various areas of the globe in the rating of the 2 categories.

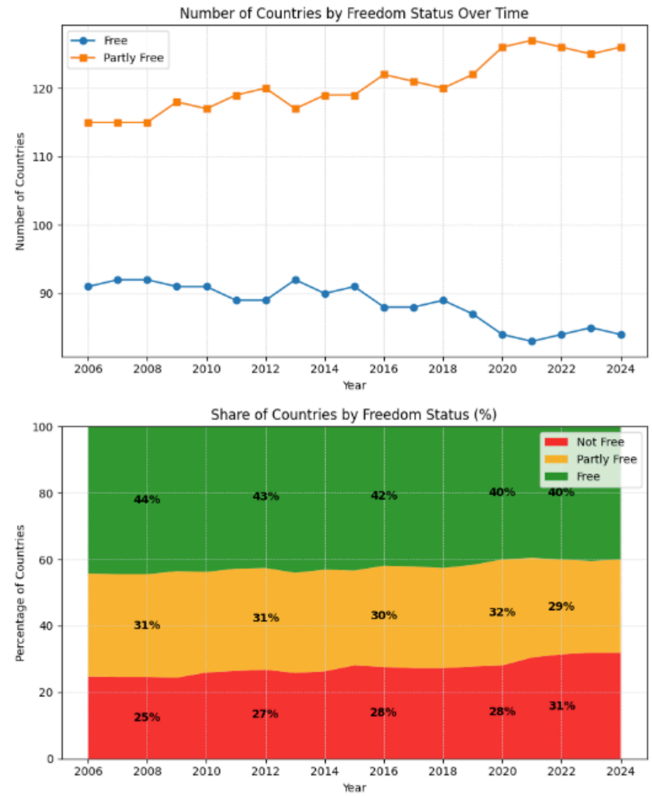

Figure 4. Trends in global freedom status. Above, absolute number of countries classified as *Free* and *Partly Free*. The count of Free countries steadily declined after 2010, while Partly Free countries increased, especially after 2015. Below, relative share of countries in each category (*Free*, *Partly Free*, *Not Free*). The proportion of Free countries dropped from 44% in 2008 to 40% in 2024, whereas Not Free countries rose from 25% to 31%, indicating a global decline in political and civil liberties.

Data Preparation

In this section we quickly summarize all the actions performed on the two data sources to obtain the final dataset on which to test the hypotheses and build the models.

Feature handling on both data sources

WB Imputations The World Bank dataset shows **substantial missingness**, varying by indicator. We therefore adopted a hybrid imputation strategy:

1) **Linear interpolation + median back–fill (Education, Unemployment):** interpolated within each country over time, with forward/backward fill at edges; if an entire series is missing, replaced by the global median. This preserves temporal continuity while ensuring full coverage.
2) **Iterative chained equations (MICE) (Income, Pollution):** each variable is imputed conditionally on the others through iterative regressions, preserving multivariate dependencies and avoiding distortions from single-variable imputations.
3) **k–Nearest Neighbors (KNN) (remaining indicators):** missing values replaced by the average of the k most similar countries, exploiting cross–sectional similarity for sporadic gaps.

These three imputation techniques are detailed in the *Appendix: Supplementary Methodological Notes.*
This hybrid approach combines temporal dynamics, multivariate structure,

and cross–sectional information, providing robust imputations across heterogeneous indicators.

WB Feature Normalization We applied **Min–Max scaling ($[0, 10]$)** to harmonize heterogeneous ESG indicators in the WB dataset, which were expressed in different units (percentages, years, indices, physical measures). This approach was preferred over standardization (e.g., z-scores) exactly because of the **miscellaneous nature of the data**: the scaling makes variables directly comparable and prevents those with larger ranges from dominating the analysis, while standardization would've just rescaled variables to have mean 0 and standard deviation 1. In our dataset, governance indices such as *Community*, *Confidence*, *Political Stability*, and *Civic Engagement* were already provided in a standardized scale ($[-2.5, 2.5]$) and were therefore left unchanged. While the Min-Max Scaling method is **sensitive to outliers** and reduces absolute meaning of the original units, it is preferable anyway because it **preserves distribution shape and interpretability**. (See figure 5).

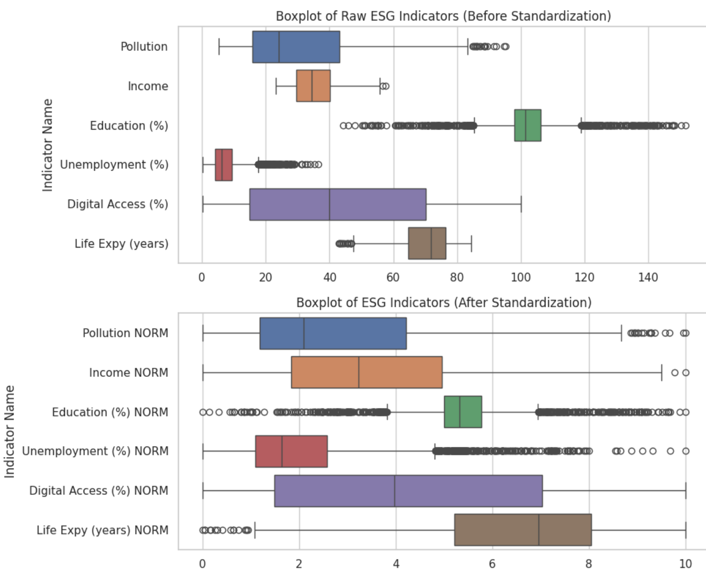

Figure 5. Boxplot of ESG indicators before and after MinMax scaling: the process enables clear visual comparison across the heterogeneous indicators. As shown, the percentages and the years get "lost", but an overall comparable scale is acquired.

FW's chosen features FW included an additional discretionary question labeled **Add A**, that we decided to **remove from the dataset**. Its exploratory analysis revealed that over 90% of its values were zeros, missing values could only be imputed with the same value, and correlations with other indicators were negligible. Retaining it would not have added explanatory power but only noise.

Countries & years selection

To ensure compatibility, country names were harmonized across the two datasets by applying a **manual mapping of alternative labels** (e.g., "Czechia" $\rightarrow$ "Czech Republic", "Russian Federation" $\rightarrow$ "Russia"). After renaming, we retained only the intersection of countries present in both sources, reducing the datasets to a common sample of 193 countries.
Then, since FW dataset covers only data from 2006 to 2023, also WB dataset has been limited to these years.

Merging the sources

The final dataset was obtained by merging Freedom in the World (FW) and World Bank (WB) indicators **on common country–year keys**, retaining only overlapping observations. This inner join ensured **perfect alignment between political freedom scores and normalized well–being measures across 193 countries (2006–2023).** The merged FW–WB panel comprises:

Metric	**Value**
Observations	3450
Countries	193
Years	18
Span	2006–2023
Median missing (%)	0.0
IQR missing (%)	0.0–0.0

Construction of the Well–being Index

The Well–being Index (**WB Index**) aggregates the 10 features of the dataset (six socio–economic variables and four governance perception indicators) in a composite index, which offers two main advantages: first, it **reduces the dimensionality of the analysis** by synthesizing multiple correlated variables into a single interpretable score, while still retaining their common signal; second, it **allows for a direct and parsimonious comparison with other high–level constructs**, such as Freedom in the World (FW), which is itself expressed as an aggregate score.
Formally, a composite index can be represented as a weighted sum of K standardized indicators:

$$I_{ct} = \sum_{k=1}^{K} w_k \, x_{k,ct},$$

where I_{ct} denotes the index for country c in year t, $x_{k,ct}$ is the normalized value of indicator k, and w_k is its weight. In this study, we adopt **equal weights** ($w_k = 1/K$) to avoid introducing arbitrary emphasis on specific dimensions and to maximize interpretability across audiences. All inputs were first aligned in direction: **higher values consistently denote better outcomes**. Accordingly, negative variables (Unemployment, Pollution, and Income measured by the Gini index) were inverted. For the governance indicators we kept the original standardized scale of $[-2.5, 2.5]$ in the dataset, but for the purpose of computing the composite WB Index they were temporarily rescaled to [0,10] to **ensure comparability with normalized socio–economic indicators**. After aggregation, the rescaled copies were discarded and the raw values retained. The final WB Index was computed as the **unweighted mean across the ten components**, yielding a synthetic well–being measure on a $[0, 10]$ scale.

Table 6. Descriptive statistics of the Well–being Index (WB Index).

Mean	**Std**	**Min**	**25%**	**50%**	**75%**	**Max**
6.23	1.36	2.63	5.31	6.18	7.18	9.30

The constructed WB Index, scaled to the interval $[0, 10]$, goes across **3450 country–year observation**s. Values range **from 2.63 (lowest well–being) to 9.30 (highest well–being),** with the interquartile range **spanning 5.31–7.18**. This indicates that **most countries cluster around the middle–high part of the scale**, while only a few exhibit very low or very high levels of overall well–being.

Methodologies

After obtaining the final dataset, we want to outline the statistical techniques that link our research questions to empirical evidence, and the architecture of the chosen models.

Correlation Study

Visual Correlation Analysis

As a first step, a visual correlation analysis explores **potential relationships between variables** using scatterplots, boxplots, and time–series plots. This is a **descriptive approach**: it highlights patterns of association without implying statistical significance or causality. Association reflects whether two variables co–vary, formally measured by the covariance:

$$\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),$$

which is positive when variables increase together and negative when one increases as the other decreases. In this study, the analysis examines whether higher freedom levels are linked to higher well–being, and the results are listed in section ***Visual Correlation***.

Correlation Tests

Correlation coefficients quantify the association between two continuous variables:

- **Pearson's** r measures linear dependence:

  $$r = \frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y},$$

  with $r \approx +1$ for strong positive, $r \approx -1$ for strong negative, and $r \approx 0$ for no linear association.
- **Spearman's** ρ is based on ranked values, capturing monotonic (not necessarily linear) relationships:

  $$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}.$$

  It is more robust to outliers and nonlinear patterns.
- **Kendall's** τ counts concordant vs. discordant pairs, and is especially reliable for small or skewed samples.

All coefficients will be **tested statistically under H_0 (no association)**. Low p–values (e.g. < 0.05) suggest the observed correlation is unlikely due to chance. In our study, we apply these measures to evaluate the relationship between indicators of political freedom and well–being outcomes, as discussed in Section **Inferential statistics**.

Partial Correlation Test

Partial correlation quantifies the **linear association between two variables** while controlling for one or more others. For X and Y controlling for Z:

$$r_{XY\cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1-r_{XZ}^2)(1-r_{YZ}^2)}},$$

where r_{XY}, r_{XZ} and r_{YZ} are Pearson correlations. By **removing the influence of** Z, this test reveals whether the **link between** X **and** Y **is genuine** or explained by a third factor. It is especially useful in observational studies with potential confounders, as in ours.
Our practical application can be found in Section **Inferential statistics**.

Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) regression estimates the **coefficients** $\hat{\beta}$ **that minimize the sum of squared residuals**:

$$\hat{\beta} = \arg\min_{\beta}\sum_{i=1}^{n}(y_i - X_i\beta)^2.$$

The fitted model is

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where β represents the average change in y for a one–unit change in x, holding other variables constant. Significance is typically assessed via t–tests ($H_0 : \beta = 0$), while R^2 indicates explained variance, and criteria such as AIC and BIC enable model comparison.

OLS assumes linearity, independence of errors, homoskedasticity, and normal residuals; violations mainly affect inference rather than coefficient estimates. Heteroskedasticity–robust standard errors (e.g., HC3) are often employed, like in our case (see Paragraph *OLS regression* under Section **Inferential statistics**).

Panel Regressions with Fixed Effects

Panel data combine repeated **observations for multiple entities over time**, allowing one to control for **unobserved heterogeneity**. A common specification is the two-way fixed effects (TWFE) model:

$$y_{it} = \beta x_{it} + \alpha_i + \gamma_t + \varepsilon_{it},$$

where α_i capture entity-specific effects and γ_t time effects. The coefficient β is identified from within-entity variation over time, net of time-invariant heterogeneity and common shocks. Standard errors are clustered at the unit level to account for heteroskedasticity and serial correlation. Compared to pooled OLS, fixed effects rely only on within-unit variation, making estimates more conservative but less prone to omitted-variable bias from time-invariant or common factors.

Predictive Modeling

In this study, we employ both **linear models** (Linear Regression, Polynomial Regression with Elastic Net) and **non–linear models** (Dynamic Panel Regression, Histogram Gradient Boosting Trees) to **forecast the WB index,** comparing their performance on held–out data.

Linear Regression

The *linear regression* is a predictive method which assumes that the target variable depends linearly on the predictors:

$$\hat{y} = \beta_0 + \sum_{j=1}^{p}\beta_j\,\tilde{x}_j,$$

where $\tilde{x}_j$ are standardized features and β_j the coefficients estimated via ordinary least squares. The estimator minimizes the sum of squared residuals, providing the best linear fit under the assumptions of linearity, independence, and homoscedasticity of errors. Once fitted, the model generates predictions for the test period by applying the same standardization and regression coefficients.
Performance on the held-out test set is quantified by:
- **Mean Squared Error (MSE),** the average of squared prediction errors:

$$\mathrm{MSE} \;=\; \frac{1}{n_{\text{test}}}\sum_{i\in\text{test}}\big(y_i-\hat{y}_i\big)^2.$$

- Root Mean Squared Error (RMSE):

$$\text{RMSE} \;=\; \sqrt{\text{MSE}},$$

which is in the same unit as the target and penalizes larger errors more heavily;

- Coefficient of Determination (R^2), the proportion of variance in the target explained by the model:

$$R^2 \;=\; 1 - \frac{\sum_{i\in\text{test}}\big(y_i - \hat{y}_i\big)^2}{\sum_{i\in\text{test}}\big(y_i - \bar{y}_{\text{test}}\big)^2},$$

where $\bar{y}_{\text{test}}$ is the test-set mean of y. The statistic $R^2 \in (-\infty, 1]$ measures the fraction of variance explained by the model, with higher values indicating better fit on the test data. These metrics provide complementary views on accuracy and explanatory power, allowing a clear evaluation of predictive quality on unseen data. This contistutes the first trained model in our experiments and can be found in Section **Linear Regression Training & Outcome**.

Non-Linear Regression (Polynomial Features + Elastic Net)

This approach augments the feature space to capture non-linear relationships while retaining a linear estimator in the transformed space. Numerical predictors are expanded with polynomial basis functions of degree d (without bias term), and then standardized to zero mean and unit variance; categorical variables are encoded via one-hot indicators. Let $\phi(\mathbf{x}) \in \mathbb{R}^q$ denote the concatenation of the polynomial expansion of numeric features and the one-hot vectors of categorical features. The model is linear in $\phi(\mathbf{x})$ but non-linear in the original inputs.

Estimator: Elastic Net. Parameters are learned with the elastic net, which combines ℓ_1 (lasso) and ℓ_2 (ridge) regularization to promote sparsity and stabilize estimates under multicollinearity:

$$\min_{\beta_0,\,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - \beta_0 - \phi(\mathbf{x}_i)^\top\boldsymbol{\beta}\Big)^2 \;+\; \lambda\left[(1-\eta)\,\frac{\|\boldsymbol{\beta}\|_2^2}{2} + \eta\|\boldsymbol{\beta}\|_1\right],$$

where $\lambda > 0$ controls the overall strength of regularization and $\eta \in [0,1]$ balances between ridge ($\eta = 0$) and lasso ($\eta = 1$). Hyperparameters (λ, η) are selected by K-fold cross-validation on the training set.

Prediction and evaluation. Given a new input $\mathbf{x}_*$, the prediction is $\hat{y}_* = \hat{\beta}_0 + \phi(\mathbf{x}_*)^\top\hat{\boldsymbol{\beta}}$, using the same fitted preprocessing steps. Performance on the held-out test set is again summarized by Mean Squared Error (MSE), Root MSE (RMSE), Mean Absolute Error (MAE), and R^2. We trained this model in our experiments in the hope of improving the outcome of the first model, and the results can be found in Section **Polynomial Regression with Elastic Net - Training & Outcome**, with its forecasting presented in Section **Forecasting 2025–2030 with Polynomial Elastic Net**.

Dynamic Panel Regression with Recursive Forecasting

This specification models the target index dynamically and heterogeneously across countries. Let $y_{i,t}$ be the target for country i in year t, and let $x_{i,t}$ denote an exogenous driver. The model includes: (i) a lagged dependent variable to capture persistence, (ii) country fixed effects to absorb time-invariant heterogeneity, (iii) continent dummies for broader regional shifts, and (iv) a linear time trend:

$$y_{i,t} \;=\; \alpha_i \;+\; \gamma_{g(i)} \;+\; \rho\, y_{i,t-1} \;+\; \beta_1\, x_{i,t} \;+\; \beta_2\, t \;+\; \varepsilon_{i,t},$$

where α_i are country effects, $\gamma_{g(i)}$ are continent effects, and $\varepsilon_{i,t}$ is an idiosyncratic error. Parameters are estimated by OLS on a training window using all available countries, after constructing $y_{i,t-1}$ via a one-period lag within each country.
Future paths of the driver $x_{i,t}$ are obtained via a local linear trend fitted on the most recent W years per country. Let $(t, x_{i,t})$ over a rolling window define the regression; if the in-window fit quality (e.g., R^2) is insufficient, a conservative fallback (e.g., last observed value) is used. Since $x_{i,t}$ represents a bounded index, projected values are truncated to a plausible range.

Recursive multi-step forecasting. With $\widehat{\theta} = \{\hat{\alpha}_i, \hat{\gamma}_g, \hat{\rho}, \hat{\beta}_1, \hat{\beta}_2\}$ and initial condition $y_{i,T}$ from the last training year T, forecasts proceed recursively:

$$\hat{y}_{i,T+1} = \hat{\alpha}_i + \hat{\gamma}_{g(i)} + \hat{\rho}\, y_{i,T} + \hat{\beta}_1\, \hat{x}_{i,T+1} + \hat{\beta}_2\,(T+1),$$

$$\hat{y}_{i,T+h} = \hat{\alpha}_i + \hat{\gamma}_{g(i)} + \hat{\rho}\, \hat{y}_{i,T+h-1} + \hat{\beta}_1\, \hat{x}_{i,T+h} + \hat{\beta}_2\,(T+h),$$

$$h = 2, \ldots, H.$$

This one-step-ahead recursion propagates the lag term using previously forecast values, yielding a full path up to the desired horizons.

We wanted to add this specification to see if that can lead to an improvement in the prediction results, and our conclusions about it can be found in Section **Dynamic Forecasting 2025–2030**.

Non-Linear Regression (Histogram Gradient Boosting Trees)

This model captures non-linearities and feature interactions using an ensemble of decision trees trained by gradient boosting. Let $f^{(m)}$ denote the predictor after m boosting stages. Starting from a constant $f^{(0)}$, each stage fits a shallow regression tree $h^{(m)}$ to the negative gradient of the loss on the current residuals and updates

$$f^{(m)}(x) \;=\; f^{(m-1)}(x) \;+\; \nu\, h^{(m)}(x),$$

where $\nu \in (0, 1]$ is the learning rate controlling the step size. With squared-error loss, this is equivalent to stage-wise least-squares fitting on residuals.

Prediction and evaluation. Given a new input x_*, the prediction is the sum of the ensemble's outputs, $\hat{y} = f^{(M)}(x_*)$. Generalization on the held-out test period is summarized by RMSE, MAE, and R^2.

A non-linear model was essential in order to figure out the best possible approach to our research questions. We present the experiment in Section **HGBR - Training & outcome** and the relative forecast in Section **Forecasting 2025–2030 with HGBR**.

Correlation Results

Visual Correlation

First of all, the relationship between political freedom (FW) and well–being (WB) is explored through visual correlation analysis. Three complementary plots are presented in Figures 6, 7 and 8. Overall, the three graphs provide converging evidence of a strong positive cross-sectional association between freedom and well-being, while also pointing to a global temporal divergence that complicates the picture: yes, there appears to be a strong correlation across countries, but a global temporal divergence is a huge warning: well-being can hold up for a while even in the absence of freedom, but how is it sustainable in the long run?

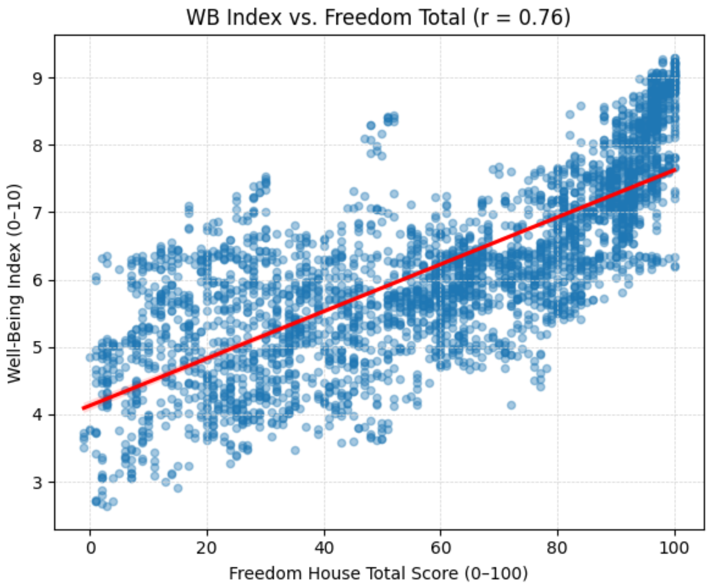

Figure 6. Scatterplot of the WB Index against the FW Total Aggregate

The scatterplot of the WB Index against the FW Total Aggregate score shows a clear positive association: countries with higher levels of political freedom tend to achieve higher levels of well–being. The data cloud aligns closely with an upward–sloping regression line, confirming a strong positive co–movement at the cross–sectional level.

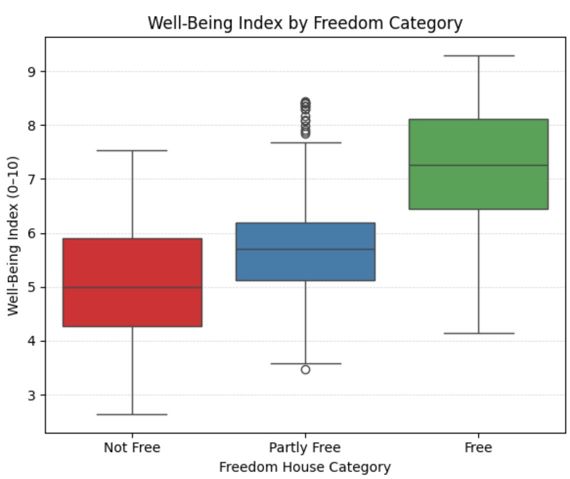

Figure 7. Boxplot of WB Index by FW status

When categorizing countries into *Free*, *Partly Free*, and *Not Free* a marked gradient appears: the median well–being is highest among free countries and lowest among Not Free ones. This further illustrates that freedom status is associated with substantial differences in socio–economic and governance outcomes.

The time–series comparison of the global averages reveals a divergence: while the WB Index remains relatively stable around 6.2, the FW score exhibits a persistent decline since 2006. The shaded gap highlights how well–being has not deteriorated as rapidly as freedom, suggesting that improvements in health, education, and digital access may temporarily offset declines in political rights. This raises the question of whether such stability can be sustained in the long term under reduced freedom.

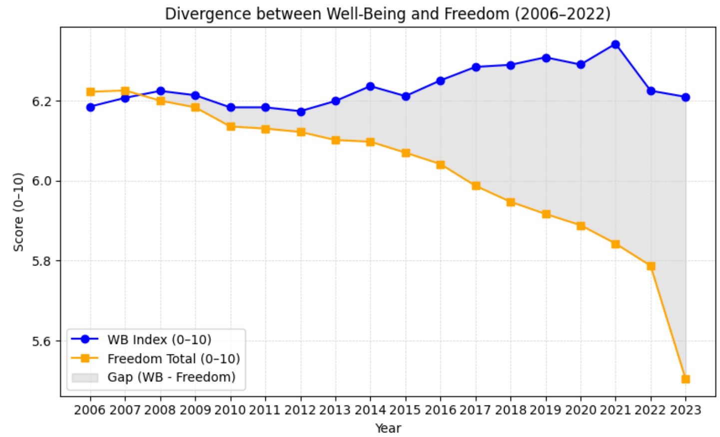

Figure 8. Line charts of WB index and FW Total Aggregate, with their difference highlighted.

Inferential statistics

Pearson & Spearman correlations

The analysis yields a Pearson correlation of $r \approx 0.75$ ($p < 0.001$) and a Spearman correlation of $\rho \approx 0.72$ ($p < 0.001$). Both coefficients confirm a strong, positive, and statistically significant association between political freedom and well–being. This supports our initial hypothesis: countries with higher levels of political rights and civil liberties also tend to achieve higher levels of well–being. Nevertheless, correlation does not imply causation, and further causal inference would be necessary to determine the direction and nature of this relationship.

Kendall's τ

The Kendall rank correlation between the *WB index* and the *Freedom Total* is $\tau = 0.607$ with a p–value < 0.001. This indicates a strong and statistically significant positive monotonic association: countries with higher well–being scores also tend to register higher levels of political freedom. Since Kendall's τ is non–parametric, this result is robust to outliers and non–linearities in the data distribution.

Partial correlations

We compute partial correlations between the WB index (X) and the freedom score (Y), controlling one covariate at a time from the dataset of the merged sources. For each control Z, we estimate $r_{XY \cdot Z}$ on matched observations (listwise deletion), reporting the sample size, the partial coefficient, and its p–value (with false–discovery–rate correction across tests).

Table 7. Partial correlations between WB index and Freedom Total, controlling for single well–being indicators. All coefficients are significant at $p < 0.001$ (FDR–adjusted).

Control variable	Partial r	95% CI
Unemployment (NORM)	0.775	[0.76, 0.79]
Education (NORM)	0.759	[0.74, 0.77]
Income (NORM)	0.743	[0.73, 0.76]
Life Expectancy (NORM)	0.719	[0.70, 0.73]
Digital Access (NORM)	0.713	[0.70, 0.73]
Pollution (NORM)	0.707	[0.69, 0.72]
Political Stability	0.540	[0.52, 0.56]
Community	0.540	[0.52, 0.56]
Confidence	0.356	[0.33, 0.38]
Civic Engagement	-0.173	[-0.20, -0.14]

Table 7 reports the partial correlation coefficients between the WB index and the Freedom Total score, controlling for each well–being indicator in turn. Results remain consistently high ($r \approx 0.71$–0.78) when condition-ing on core socio–economic indicators such as education, income, life expectancy, unemployment, digital access, and pollution. The association weakens when controlling for governance dimensions such as political stability and community ($r \approx 0.54$), and becomes modest with confidence ($r \approx 0.36$). Interestingly, when controlling for civic engagement, the correlation turns negative ($r \approx -0.17$). These results suggest that while the WB–Freedom link is robust, some institutional and social trust indicators capture overlapping variation that partially mediates the relationship.

OLS regression

We estimated a simple linear regression of the form

$$WB_i = \alpha + \beta\, FW_i + \varepsilon_i,$$

where WB_i is the WB index and FW_i the FW Total Aggregate score.
The baseline OLS model yields a positive and highly significant association ($\beta = 0.035$, $t \approx 69$, $p < 0.001$), with $R^2 = 0.58$, meaning that freedom explains more than half of the variance in well–being. To account for possible heteroskedasticity in cross–country data, we re–estimated the model using heteroskedasticity–robust standard errors (HC3). The coefficient remains virtually unchanged ($\beta = 0.035$), and significance persists ($z \approx 64$, $p < 0.001$), confirming that results are not driven by violations of the homoskedasticity assumption. Thus, both conventional and robust specifications strongly support a positive link between freedom and well–being.

A note about OLS results: The model coded also showed data about the Information criteria (AIC=8901, BIC=8914), which appear large in absolute terms. Usually, a great values for AIC and BIC is an alarm-bell, but these measures are only meaningful in relative comparisons: lower values indicate better fit when models are compared on the same data. Thus, AIC and BIC will become informative when evaluating alternative specifications such as panel regressions with fixed effects. Right now, they are not concerning.

Specification	Coef.	Std. Err.	t/z	p
OLS (conventional SE)	0.035	0.001	68.94	0.000
OLS (HC3 robust SE)	0.035	0.001	64.46	0.000
R^2		0.580		
N		3450		

Table 8. OLS regression of Well–being index on Freedom Total, with conventional and robust (HC3) standard errors.

Panel Regression Results

Model	Predictor	Coef.	p-value	N	R^2
OLS	const	5.650^{***}	0.000	3450	0.580
OLS	Freedom Total	0.0349^{***}	0.000	3450	0.580
OLS–HC3	const	5.650^{***}	0.000	3450	0.580
OLS–HC3	Freedom Total	0.0349^{***}	0.000	3450	0.580
FE (TW)	const	—	—	3450	0.004 (within)
FE (TW)	Freedom Total	0.0078^{***}	0.004	3450	0.004 (within)

FE: two-way fixed effects (country & year), SE clustered by country.
Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9. Comparison of pooled OLS, OLS with HC3 robust SE, and two-way fixed effects. While pooled OLS indicates a strong cross-sectional association between freedom and well-being, the FE estimate is smaller yet significant, pointing to a modest within-country effect once unit and time heterogeneity are removed.

Table 9 compares pooled OLS, OLS with HC3 robust standard errors, and two-way fixed effects (FE) with country- and year-specific intercepts. The panel includes 193 countries over 18 years (2006–2023), for a total of 3,450 observations.

Pooled OLS suggests a strong positive association: a one-point increase in the Freedom Total score is linked to a 0.035 rise in the WB Index ($p < 0.001$), with an R^2 of 0.58. This association remains highly significant when using HC3 robust errors.

Once unobserved country heterogeneity and common time shocks are controlled for via two-way FE, the estimated coefficient falls to 0.0078 (SE clustered by country = 0.0027, $p = 0.004$). The within-R^2 is modest (0.004), indicating that most of the explanatory power comes from cross-sectional differences rather than within-country temporal variation. Nevertheless, the FE results confirm a statistically significant, though smaller, within-country relationship.

We cluster standard errors at the country level to allow for arbitrary heteroskedasticity and serial correlation within units; with large N and moderate T, this yields reliable inference. As a robustness direction, two-way clustering or Driscoll–Kraay errors could address cross-sectional dependence.

Overall, the results indicate that while much of the observed freedom–well-being association is driven by cross-country differences, there remains a modest but robust positive within-country effect.

Predictive Modeling Results & Forecast

Linear Regression Training & Outcome

We train a linear regression model on the WB index using the feature *Freedom Total* as predictor. The dataset is split chronologically, with observations up to 2019 as training data and those from 2020 onwards as test data, to mimic a forecasting scenario but also to have data for comparison and results check. Before fitting, the predictor is standardized to zero mean and unit variance. The regression coefficients are estimated by ordinary least squares, and predictions for the test period are obtained by applying the same standardization.

Table 10. Linear Regression – specifications for reproducibility

Parameter	Value
Target	WB index
Features	Total
Train period	≤ 2019
Test period	≥ 2020
Preprocessing	StandardScaler
Estimator	LinearRegression (OLS)
Pipeline	Scaler $\rightarrow$ LinReg
Split policy	Chronological by Year
Output file	`pred_wb_2020_2023.csv`
Metrics	RMSE, MSE, R^2

Model performance is evaluated on the held–out test set using the standard metrics RMSE, MSE, and the coefficient of determination R^2.

Table 11. Linear regression results on the test period (2020–2023).

Metric	Value
RMSE	0.909
R^2	0.307

The results show that the linear model achieves an RMSE of 0.909 and explains approximately 30.7% of the variance in the target, laying the foundation for subsequent models with improving outcomes as a goal.

Polynomial Regression with Elastic Net - Training & Outcome

After the first, simple regression, we extended that linear specification by enriching the feature space with polynomial transformations of degree three for the numerical predictors and by one–hot encoding the categorical variables. To prevent data leakage, preprocessing and model fitting are integrated into a single pipeline. Model parameters are estimated with the *Elastic Net*, combining ℓ_1 (lasso) and ℓ_2 (ridge) regularization. The optimal penalty strength α and mixing ratio l_1 are selected via 5–fold cross–validation, that led to an $\alpha = 0.0001$ and a l_1 ratio $= 1.0$.

As in the previous setup, the data are split chronologically (train ≤ 2019, test ≥ 2020). Predictions for the test period are then compared against the true values of the WB index.

Table 12. Polynomial Elastic Net: specifications for reproducibility

Parameter	Value
Target	WB index
Features	Total, Year, Continent
Train period	≤ 2019
Test period	≥ 2020
Preproc num	Poly(deg=3) + Scaler
Preproc cat	OneHotEncoder
Estimator	ElasticNetCV
Grid search	$l1\_ratio \in \{0.1, 0.5, 0.9, 0.95, 1\}, \alpha \in (-4, 1, 80)$
CV folds	5
Pipeline	Preproc $\rightarrow$ ElasticNet
Output file	`pred_wb_2020_2023.csv`
Metrics	RMSE, MAE, R^2

Model performance is summarized using RMSE, MAE, and R^2. The results show that **the polynomial expansion combined with regularization improves predictive accuracy** compared to the simple linear regression, explaining about 44.7% of the variance in the target.

Table 13. Polynomial Elastic Net results on the test period (2020–2023).

Metric	Value
RMSE	0.812
MAE	0.614
R^2	0.447

Overall, the findings suggest that introducing non–linearities through polynomial features allows the model to capture more complex patterns in the data, achieving a higher explanatory power than the baseline linear regression.

Forecasting 2025–2030 with Polynomial Elastic Net

As a next step, we retrained the polynomial Elastic Net model using all available data up to 2022 (we left out 2023 because its data are not full released by the time we performed our project) and applied it to generate medium–term forecasts for the years 2025 and 2030. The feature space is the same as in the previous experiment (polynomial degree three for numeric predictors, one–hot encoding for categorical variables), and preprocessing is handled within the same pipeline. Future values of the predictor *Total* are obtained under a *trend scenario*, where a linear regression is fitted on the last five observed years for each country; if insufficient history is available, the last observed value is carried forward. The trained model then produces predictions of the WB index for each country at the selected horizons.

Table 15 reports examples of the forecasted values for 2025.

These results illustrate the model's use in scenario analysis: **by combining projected trajectories of explanatory variables with the trained regression, it is possible to obtain medium–term forecasts of the WB index at the country level**.

Table 14. Polynomial Elastic Net (Future Forecast): specifications for reproducibility

Parameter	Value
Target	WB index
Features	Total, Year, Continent
Train period	≤ 2022
Forecast yrs	2025, 2030
Scenario	Trend on last $W = 5$ yrs, fallback=hold
Preproc num	Poly(deg=3) + Scaler
Preproc cat	OneHotEncoder
Estimator	ElasticNetCV ($l1\_ratio \in \{0.9, 1.0\}, \alpha \in (-4, 1, 80)$)
Pipeline	Preproc $\rightarrow$ ElasticNet
Output file	`pred_wb_future_2025_2030.csv`
Metrics	– (no ground truth)

Table 15. Polynomial Elastic Net forecasts for 2025 under the trend scenario.

Country	Total (proj.)	WB_pred	Continent
Afghanistan	7.4	5.12	Asia
Albania	65.2	6.54	Europe
Algeria	29.4	4.99	Africa
Andorra	90.5	7.91	Europe
Angola	34.0	5.06	Africa
Antigua and Barbuda	86.9	7.06	Americas
Argentina	85.0	6.92	Americas
Armenia	63.8	6.55	Eurasia
Australia	93.5	8.06	Asia
Austria	92.2	8.06	Europe

Dynamic Forecasting 2025–2030

As our specification model, we implemented a dynamic panel regression to produce recursive forecasts of the WB index up to 2030. The model is estimated on the full panel of countries up to 2022, including a one–period lag of the dependent variable, country and continent fixed effects, and a linear time trend. Future values of the predictor *Total* are generated using a local linear trend over the last five available years for each country, with fallback to the last observation when insufficient history is available. Forecasts are obtained recursively: the 2025 prediction is based on the last observed values, while the 2030 forecast uses the 2025 predictions as lag input. This setup reflects real–time forecasting conditions, where future lags are not directly observable.

Table 16. Dynamic Panel Forecast: specifications for reproducibility

Parameter	Value
Target	WB index
Train period	≤ 2022
Forecast years	2025, 2030 (recursive)
Data prep	Sort by {Country, Year}, add WB_lag
Missing handling	Drop NA on WB_lag
Model	OLS: $WB \sim$ Country FE + WB_lag + Total + Year + Continent
FE structure	Country dummies + Continent dummies
Estimator	`statsmodels OLS`
Total projection	Trend last $W = 5$ yrs; fallback=hold-last
R^2 threshold	$R^2_{\text{min}} = 0.2$
Bounds	Total $\in [0, 100]$
Forecasting	Year-by-year recursion (2023–2030)
Output file	`pred_wb_dynamic_2025_2030.csv`
Metrics	– (no ground truth)

Table 17 reports examples of predicted values for 2025, and Table 18 for 2030.

These results highlight how the **dynamic specification, by incorporating lagged effects and country heterogeneity, produces medium–term forecasts that differ from the purely cross–sectional elastic net approach**, particularly in countries with strong persistence or distinct regional patterns.

HGBR - Training & outcome

With this *Histogram Gradient Boosting Regression (HRGB)* model, we evaluated a non–linear ensemble approach based on histogram gradient boost-

Table 17. Dynamic panel regression forecasts for 2025 under the trend scenario.

Country	Total (proj.)	WB_pred	Continent
Afghanistan	7.4	3.63	Asia
Albania	65.2	6.41	Europe
Algeria	29.4	5.33	Africa
Andorra	90.5	8.40	Europe
Angola	34.0	4.88	Africa
Antigua and Barbuda	86.9	7.20	Americas
Argentina	84.0	6.41	Americas
Armenia	63.8	6.59	Eurasia
Australia	93.5	8.38	Asia
Austria	92.2	8.37	Europe
Azerbaijan	6.9	5.54	Eurasia
Bahamas	91.0	7.02	Americas

Table 18. Dynamic panel regression forecasts for 2030 under the trend scenario.

Country	Total (proj.)	WB_pred	Continent
Afghanistan	0.0	3.54	Asia
Albania	63.2	6.47	Europe
Algeria	25.4	5.32	Africa
Andorra	87.0	8.52	Europe
Angola	38.0	4.91	Africa
Antigua and Barbuda	89.4	7.35	Americas
Argentina	84.0	6.50	Americas
Armenia	75.8	6.58	Eurasia
Australia	90.0	8.52	Asia
Austria	91.2	8.54	Europe

ing, which can capture complex interactions between numeric and categorical predictors. To guarantee comparison, the model is trained on data up to 2019 and tested on the period 2020–2023, just like we did for the linear approach. Categorical features are encoded through one–hot encoding, while numeric features are passed unchanged, allowing the tree–based learner to exploit both cross–sectional and temporal variation.

Table 19. Histogram Gradient Boosting (Batch): specifications for reproducibility

Parameter	Value
Target	WB index
Features	Total, Year, Continent
Train period	≤ 2019
Test period	≥ 2020
Preproc	ColumnTransformer[OneHotEncoder on Continent; remainder=passthrough]
Estimator	HistGradientBoostingRegressor
Hyperparams	loss=squared_error; learning_rate=0.05; max_depth=6; max_iter=600; l2_regularization=0.0; early_stopping=True; random_seed=42
Pipeline	Preprocess $\rightarrow$ HGBR
Output file	`pred_wb_2020_2023_hgbr.csv`
Metrics	RMSE, MAE, R^2

Model accuracy is again measured using RMSE, MAE, and R^2. The results show that HGBR achieves an RMSE of 0.837 and explains 41.3% of the variance in the WB index, **outperforming the simple linear regression but slightly below the polynomial Elastic Net in terms of explanatory power.**

Table 20. HGBR results on the test period (2020–2023).

Metric	Value
RMSE	0.837
MAE	0.621
R^2	0.413

Overall, the boosting model provides **a flexible (though slightly worse) alternative to polynomial regression**, delivering competitive predictive performance with automated handling of non–linearities.

Forecasting 2025–2030 with HGBR

Out of curiosity, we extended the HGBR model to generate medium–term forecasts up to 2030. The model is retrained on all available data up to 2022, and future values of the predictor *Total* are constructed under a *hold* scenario, where the last observed values are carried forward. The trained boosting ensemble is then applied to obtain well–being forecasts for 2025 and 2030.

Table 21. Histogram Gradient Boosting (Future Forecast, hold): specifications for reproducibility

Parameter	Value
Target	WB index
Features	Total, Year, Continent
Train period	≤ 2022
Forecast yrs	2025, 2030
Scenario	Hold-last for Total
Preproc	ColumnTransformer[OneHotEncoder(Continent); remainder=passthrough]
Estimator	HistGradientBoostingRegressor
Hyperparams	loss=squared_error; lr=0.05; max_depth=6; max_iter=600; l2=0.0; early_stopping=True; random_seed=42
Pipeline	Preprocess $\rightarrow$ HGBR
Output file	`pred_wb_future_hgbr_2025_2030.csv`
Metrics	– (no ground truth)

Table 22 reports examples of predicted values for 2025.

Table 22. HGBR forecasts for 2025 under the hold scenario.

Country	Total (proj.)	WB_pred	Continent
Afghanistan	10.0	5.35	Asia
Albania	67.0	6.51	Europe
Algeria	32.0	5.01	Africa
Andorra	93.0	7.93	Europe
Angola	30.0	4.89	Africa
Antigua and Barbuda	85.0	6.68	Americas
Argentina	84.0	6.71	Americas
Armenia	55.0	6.44	Eurasia
Australia	95.0	7.92	Asia
Austria	93.0	7.93	Europe

These results show that the boosting model can be effectively used for scenario analysis, providing forecasts of the WB index consistent with country–specific predictors and historical patterns.

Limitations and Future Work

The predictive experiments achieved **reasonable accuracy given the data structure, with out–of–sample RMSE around 0.8–0.9 on a 0–10 scal**e. This corresponds to an **average prediction error below one index point**, which is **congruent with the cross–country heterogeneity of the dataset**. While satisfactory, several improvements could be explored in future work.

First, **feature enrichment could include additional socio–economic and institutional indicators from the dataset** (e.g., digital access, education, pollution), as well as lagged values, interaction terms, or temporal differences. Such engineering may help capture non–linear or dynamic relationships not fully exploited here.

Second, **model tuning could be refined**: Elastic Net performance is sensitive to the choice of α and l_1 ratio, while gradient boosting models may benefit from optimized learning rates, tree depths, and early stopping strategies.

Third, **alternative algorithms** (e.g., Random Forests, XGBoost, LightGBM, or shallow neural networks) **could be tested** for their ability to capture complex interactions. Finally, target transformations (such as log or robust loss functions) and time–series cross–validation could further stabilize results and reduce the impact of outliers.

Overall, while current models already provide meaningful forecasts, these enhancements offer promising directions to further reduce prediction error and improve robustness.

Policy and Practical suggestions for the future. Initially, well-being was often treated as a purely normative assessment of a particular country.

However, based on the results obtained in this study, we can propose a practical tool for policymakers: an early-warning system that signals how the well-being of citizens is closely linked to the level of political freedom. By monitoring trends in political freedom, policymakers can recognize it as a critical factor that may influence future well-being outcomes and incorporate it into long-term planning and decision-making.

Conclusions

This study investigated the relationship between political freedom (FW) and well–being (WB) by combining data from Freedom House and the World Bank into a unified panel dataset (2006–2023). After addressing missing values through tailored imputation strategies and constructing a composite well–being index, we explored three guiding research questions.

1. Do freedom and political rights correlate with well-being outcomes? Descriptive scatterplots and visual correlations suggested a positive association, which was confirmed statistically. Pearson, Spearman, and Kendall coefficients all showed strong and highly significant positive correlations between freedom and WB (e.g., Pearson $r \approx 0.75$, $p < 0.001$). This indicates that countries with higher levels of political rights and civil liberties also tend to display higher well–being outcomes.

2. Which well-being dimensions are most strongly associated with freedom? Partial correlation tests, controlling for individual WB indicators, confirmed that the FW–WB relationship remains robust across specifications. The strongest associations emerged when controlling for unemployment ($r = 0.775$) and education ($r = 0.759$), while civic engagement showed a weak or even negative association ($r = -0.173$). This suggests that freedom is systematically linked with core socio–economic well–being dimensions, although institutional and trust factors mediate the strength of the relationship.

3. If such associations exist, how well can we predict near-term well-being from freedom indicators, and what do scenario-based forecasts for 2025 and 2030 suggest? Predictive modeling provided quantitative evidence on forecasting performance. A simple linear regression explained 30.7% of the variance in WB for 2020–2023 ($R^2 = 0.307$, RMSE=0.909). Polynomial regression with Elastic Net improved explanatory power to 44.7% ($R^2 = 0.447$, RMSE=0.812, MAE=0.614). Histogram Gradient Boosting achieved competitive accuracy ($R^2 = 0.413$, RMSE=0.837, MAE=0.621). Dynamic panel regressions, incorporating lagged WB and fixed effects, extended predictions to 2025 and 2030. Scenario–based forecasts under both trend and hold assumptions consistently predict rising WB levels in countries with sustained freedom and stagnation or decline where freedom remains limited.

Overall assessment. Across correlation tests, partial associations, regression models, and predictive forecasts, the evidence consistently points to a robust positive relationship between freedom and well–being. While causality remains outside the scope of this study, the results highlight that freedom not only correlates with WB historically, but also improves our ability to predict near–term WB trajectories and to construct meaningful medium–term scenarios.

Supplementary Methodological Notes

APPENDIX

This appendix provides extended descriptions of the preprocessing techniques applied in the study. It includes the theoretical background, mathematical formulations, and a discussion of advantages and drawbacks for the following steps:

i) Feature Selection
ii) Min-Max Scaling - normalization
iii) Imputation Techniques

Feature Selection Technique

To reduce dimensionality and focus the analysis on the most informative and policy-relevant indicators, a multi-step feature selection process was applied to the full set of 71 Environment, Social and Governance (ESG) indicators provided by the World Bank.

First, we conducted an initial *exploratory data analysis* (EDA) to assess the statistical properties of each indicator, including missing value rates, standard deviation, and distributional shape. Indicators with excessive sparsity or limited variance across countries and years were discarded.

Subsequently, we performed a *correlation analysis* to identify and remove highly collinear variables. Pearson correlation coefficients were computed across all numerical variables, and in cases of strong linear dependence (e.g., $|r| > 0.9$), redundant features were excluded to reduce multicollinearity.

To further refine the selection, we applied a combination of unsupervised methods:

- **Principal Component Analysis (PCA)** was used to examine the underlying latent structure of the dataset. Components explaining the highest variance were examined to identify key contributing features.
- **Mutual information** scores were computed to evaluate the non-linear dependencies between variables, helping to prioritize indicators with high informational value.
- **Hierarchical clustering** on the feature space was used to group similar indicators, allowing for representative features to be chosen from each cluster.

Finally, domain expertise played a critical role in selecting variables that are interpretable and relevant to the goals of the analysis. This hybrid approach led to the selection of the following 10 indicators, which span key ESG dimensions such as education, labor, governance, inequality, environmental exposure, and digital access.

Min-Max Scaling Technique

To harmonize heterogeneous indicators, we applied Min–Max scaling to map each variable into $[0, 10]$:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \cdot 10.$$

This was preferred over z-scores given the mix of percentages, years, physical measures, and bounded governance indices.

The method ensures comparability across features and preserves distributional shape, but is sensitive to outliers and removes original units. In our study, scaling was applied only to raw variables, while governance indicators (already normalized) were left unchanged to retain interpretability.

Imputation Techniques

Linear Interpolation + Median Back–Fill (Education, Unemployment)

For time–series variables with moderate gaps, we applied linear interpolation within each country:

$$x_{i,t}' = x_{i,t-1} + \frac{x_{i,t+1} - x_{i,t-1}}{t + 1 - (t - 1)} (t - t_{-1}),$$

followed by forward/backward fill at the edges. If an entire series was missing, values were replaced with the global median across countries. In practice, this was implemented with `pandas.DataFrame.interpolate()` and fallback to `median()` imputation. This ensured temporal continuity and avoided dropping any country.

Iterative Chained Equations (MICE) (Income, Pollution)

For variables with strong multivariate structure, we used Multiple Imputation by Chained Equations (MICE). Each variable with missing data was regressed on the others, updated iteratively until convergence:

$$x_j^{(m+1)} \sim f_j\big(x_{-j}^{(m)}\big), \quad j = 1, \dots, p,$$

where f_j is the conditional model for variable j. Implementation was based on scikit–learn's `IterativeImputer` with random seed fixed for reproducibility. This preserved correlations between Income, Pollution, and the rest of the ESG indicators.

k–Nearest Neighbors (KNN) (remaining indicators)

For sporadic gaps, we applied KNN imputation. Each missing entry was replaced by the mean of its k nearest neighbors:

$$x_i' = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} x_j,$$

where $\mathcal{N}_k(i)$ are the k most similar countries (based on Euclidean distance in the feature space). We used `KNNImputer` from scikit–learn with $k = 5$, applied after scaling the features to ensure comparability. This exploited cross–sectional similarity while keeping the method non–parametric.