Technical Report: World Happiness

Jolly Ranchers
University of Virginia
STAT 4996

1. Motivation

Research Question: Which country-wide economic, health, and social factors are most important in predicting self-reported happiness across the world?

The motivation for this project, put simply, is to determine what factors influence world happiness. This is certainly a challenging task, as a large number of factors contribute to the happiness of people within their respective nations. While we initially had a general idea about what we wanted to focus on, narrowing down the question was an important step we needed to take. Our research focused on attempting to predict the self-reported happiness score of a random sample of citizens in a country using education, health, economic, and social data. This happiness score, or "Life Ladder" score, is taken from the World Happiness Report, where citizens are asked to rate their overall well-being on a scale from one to ten. This was our response variable. Additionally, we used the median value of Life Ladder to split countries into two groups, considered happy countries if their life ladder score was greater than the median and unhappy countries if their scores were less than the median.

We collected various data points from each country for indicators we presumed would be significant, pulling from the World Bank. These included vaccination rates, educational attainment data, economic indicators, and many additional factors. We thought this would be an interesting way to look at world happiness because most research we found measured happiness directly through surveys. While surveys are likely the most accurate way to reflect happiness, there has not been as much research tying different real-world indicators beyond GDP to survey results.

Ideally, this research will help policymakers around the globe determine where to focus their efforts to improve the well-being of their citizens. We expect that this research could be valuable to many different stakeholders, including governments, the United Nations, NGOs, and others working to create happier communities worldwide.

2. Bibliography Annotation

General Overview:

The following articles provide valuable information about the factors that influence the happiness of people around the globe when viewed individually. When observed together, they offer significantly more insight into happiness trends than they do alone. A main theme across the literature is the strong link between the economic prominence of a country and the happiness of its people. This can be considered straightforward since many would expect that a nation’s population would become happier as their country grows wealthier. Additionally, a higher life expectancy is highly correlated with higher happiness even though this relationship could have multicollinearity since wealth also correlates with life expectancy.

One particularly interesting finding from the literature is the U-shape of happiness over a lifetime. This means that people tend to be least happy during midlife. While one article clearly demonstrated this U-shaped curve, another article plotting age against well-being had wider confidence intervals which makes it harder to draw conclusions. Another important finding was

the role of less tangible factors such as community involvement and cultural harmony. While it's harder to quantify the idea of "community," it was clear from different studies that a strong sense of community increases overall well-being.

There were some differences that arose from the literature despite most of them broadly agreeing. For example, one analysis found that higher perceived levels of freedom, trust, and generosity were significantly correlated with happiness, while another did not find this connection. Despite these differences, the overall takeaway is that happiness is influenced by a complex mix of economic, health, social, and psychological factors. Our research builds on these findings by exploring a wide range of indicators to better understand what drives happiness and how governments and international organizations can focus their efforts.

Bibliography:

Blanchflower, D.G. — *Is happiness U-shaped everywhere?*

This article examined the relationship between age and happiness across 145 countries. David Blanchflower, a professor of economics at Dartmouth College, found that well-being is U-shaped when compared to age, meaning happiness is lowest in midlife and higher at the beginning and end of life. This trend was consistent across countries and did not vary by language, region, or life expectancy. Blanchflower hypothesized that midlife unhappiness could be due to financial struggles, deaths of despair, or increased mental illness. However, he did not offer an explanation for why happiness increases again later in life. It would be interesting to know his thoughts on that. Overall, this article offers valuable insight into how age patterns influence well-being across nations.

Jaswal, Vidushi et al. — *Understanding the determinants of happiness through the Gallup World Poll*:

This article used the Gallup World Poll from 2016 to 2017, which is the same dataset we planned to use but from an earlier time period. The authors, from various medical and psychology departments in Chandigarh, India, found that GDP per capita, family, and life expectancy were significantly correlated with happiness. However, they didn't explain clearly what the "family" variable included, which would have been interesting to know. The study concludes that more research is needed, and while it does a good job of identifying broad contributors to happiness, it would be even more helpful to explore specific policy approaches to improve these factors.

Behera, Rahut, Padmaja, Dash — *Socioeconomic determinants of happiness: Empirical evidence from developed and developing countries*:

This article analyzed happiness across 166 countries between 2005 and 2020 and found that per capita income, social support, and freedom positively impact happiness. It also found that air pollution negatively affects happiness. Gender inequality did not show a significant relationship with happiness and their findings align with the Easterlin Paradox which states that more income boosts happiness only up to a point. This article offered insights about how factors

affect developed and developing nations differently but did not give many specific policy recommendations for different regions which would have been valuable.

Kshtriya, Singh, Valk — *Health, Hope, and Harmony: A Systematic Review of the Determinants of Happiness across Cultures and Countries*:

This article grouped happiness factors into three major categories: Health (mental, emotional, and physical well-being), Hope (purposeful life, social relationships, self-care), and Harmony (alignment with cultural and community contexts). This article was interesting because it took a broader and more holistic view than the World Happiness Report which only often focuses on individual factors. Grouping variables into different categories offers a clearer framework for understanding how different aspects of life contribute to happiness. One question the article fails to answer is how economic factors fit into this framework since this overlap wasn't fully addressed.

Volunteer Organization — *Abraham Maslow's Psychology: Pioneering the Pursuit of Happiness through the Hierarchy of Needs*:

This article reviewed Maslow's Hierarchy of Needs and how it relates to well-being and further emphasized the importance of factors such as physiological needs, safety, love and belonging, esteem, and self-actualization. The idea that self-actualization is influenced even when lower-level needs aren't fully met was particularly interesting. This framework is useful for choosing variables like food security, education, and crime rates to study happiness. However, the article doesn't dive deep into how we could quantitatively measure each level which would have been helpful for our research.

Ruggeri — *Well-being is More Than Happiness and Life Satisfaction: A Multidimensional Analysis of 21 Countries*:

This article emphasized that well-being is more complex than just happiness or life satisfaction. Ruggeri showed that psychological, social, and community dimensions all contribute to well-being by using the European Social Survey. This approach highlights the importance of considering multiple variables and not just relying on happiness scores alone. The study's insights about how well-being varies by gender, education, and employment were valuable. However, it didn't discuss cultural context which could be important when comparing countries.

OpenStax College (Rice University) — *Perspectives on Happiness*:

This lesson provided a broad overview of happiness from psychological perspectives, including the concepts of the pleasant life, the good life, and the meaningful life. Positive psychology and the idea of "flow" which is being fully absorbed in activities were also discussed. While the overview was very helpful, it didn't go deeply into how happiness might vary across cultures which would add more depth to the discussion.

Anya (UCLA DataRes) — *Quantifying Happiness: Exploring Factors Affecting Happiness*:

This article emphasized the importance of quantifying happiness using World Happiness Report metrics. It identified economic indicators, social support, life expectancy, freedom, generosity, and low corruption as key factors. While the article offered a great concise overview,

it mainly focused on correlations and did not explore causal mechanisms or cultural variation, leaving room for deeper research.

3. Data Source

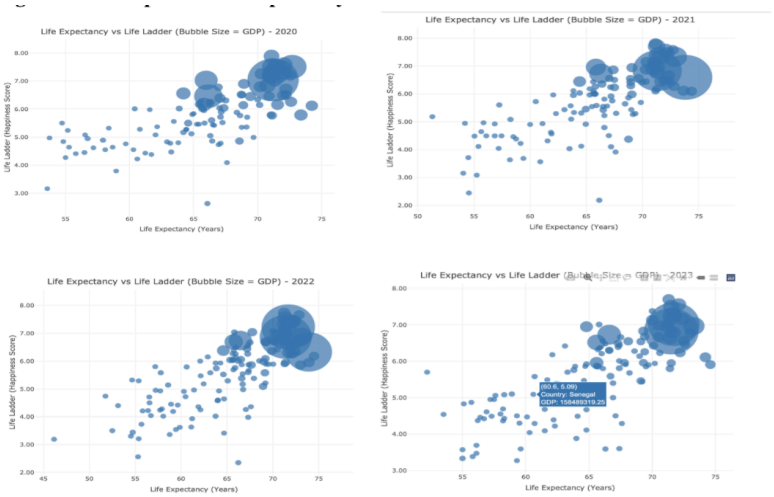
We gathered and merged country-level data from the World Happiness Report and the World Bank to explore how happiness can be measured. The "Life Ladder" score which ranges from 1 to 10 represents people's self-reported level of well-being and served as our response variable. We compiled a broad range of explanatory variables across five key areas to help explain this score: health (e.g., life expectancy, immunization rates), education (e.g., school enrollment), economic indicators (e.g., GDP per capita, inflation), social factors (e.g., food insecurity rates, net migration), and infrastructure (e.g., internet access, sanitation services).

Our study draws on two referenced and highly reliable sources which are the World Bank and the World Happiness Report. The World Happiness Report uses data from the Gallup World Poll to measure subjective well-being across nations. The World Bank provides internationally standardized indicators across various areas of development. There were some factors such as political stability and sense of community belonging that we would have liked to include but could not find consistently measured values across countries even though our dataset was large and included many important variables.

We standardized the variables and handled missing values by filling them in with the median value for each variable to prepare the data. While this approach allowed us to maintain a usable dataset, it might be more accurate to impute missing values using similarities between countries rather than relying purely on median values in future studies. This is because it would better account for regional or contextual patterns for when data is missing.

Overall, the steps we took in collecting and cleaning the data gave us a strong foundation to explore how different aspects of a country's conditions relate to the happiness of its population.

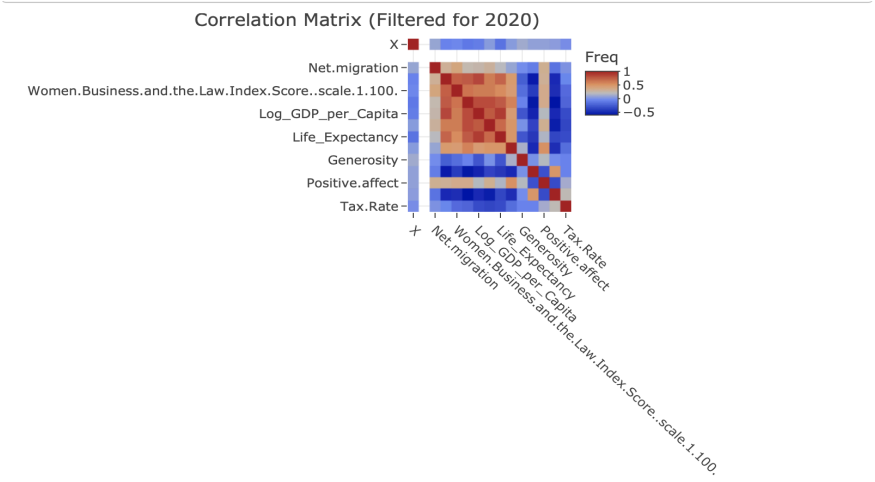
4. Data Visualizations



1. Life Expectancy vs. Life Ladder with GDP as Bubble Size (2020-2023)

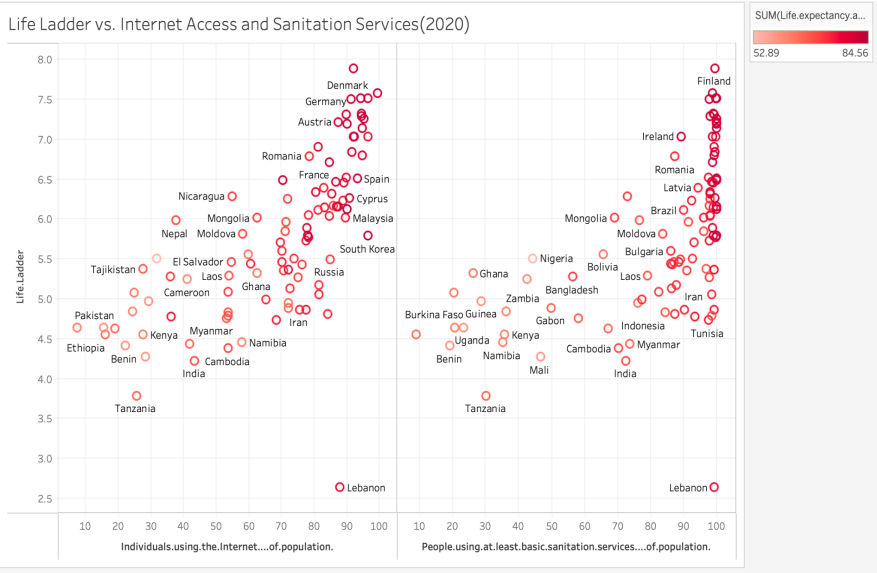
These scatterplots demonstrate the relationship between life expectancy and self-reported happiness across countries. The bubble size represents GDP per capita, so the bigger the country, the wealthier it is. All four years show a clear positive correlation and this is particularly strong in years 2020 and 2022. The clusters of larger bubbles in the top right

corner demonstrates that economic well-being is heavily linked with longer lives and greater happiness. These findings suggest that countries with higher GDP and life expectancy tend to report higher happiness levels and reinforce the strong relationship between economic resources, public health, and societal well-being.



2. Correlation Matrix (2020)

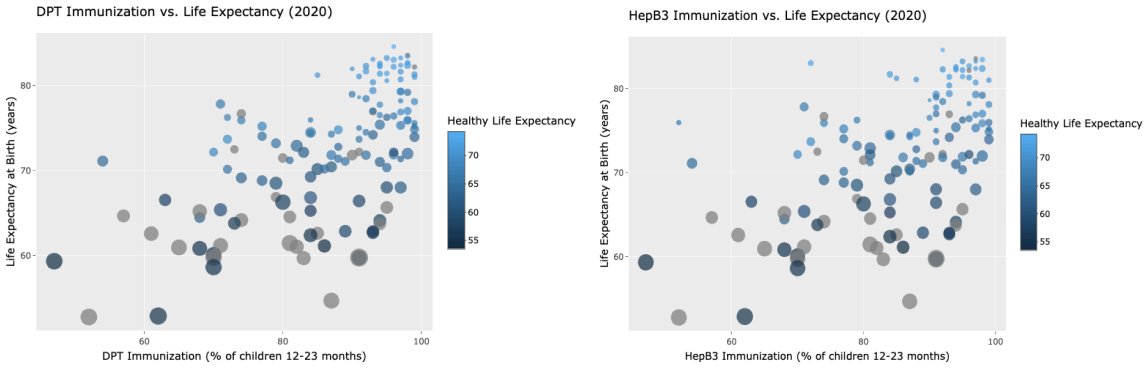
Our 2020 correlation matrix provides an overview of relationships between key numeric variables. It demonstrates strong positive correlations between log GDP per capita and life expectancy and between generosity and positive affect. They are both indicators of strong social cohesion. It is important to note that perceived corruption is negatively correlated with happiness and social support. This highlights the importance of institutional trust. These patterns helped guide our later decisions about which variables to prioritize in our modeling.

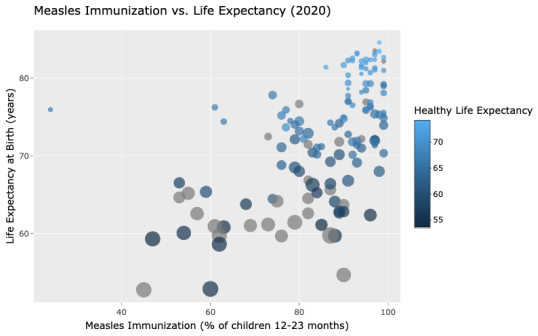


3. Scatterplots of Happiness vs. Internet Access and Sanitation Services (2020)

These twin scatterplots highlight an important contrast. While sanitation access is nearly universal across many countries and shows only a limited relationship with happiness beyond a certain threshold, internet access demonstrates a more continuous positive correlation with happiness. This suggests that in a modern context, digital connectivity is playing an increasingly important role in well-being. The case of Lebanon, which has strong infrastructure access but low happiness scores, shows that political instability or other factors can override basic infrastructure benefits.

4. Immunization Rates vs. Life Expectancy (DPT, Measles, HepB3)





These bubble charts show immunization coverage for DPT, measles, and HepB3 plotted against life expectancy. Across all three vaccines, higher immunization rates are consistently associated with longer life expectancy and lower infant mortality which are represented by the smaller bubbles. These visualizations underline the important role of public health initiatives in improving overall national well-being especially in lower-income countries.

Summary of Key Findings from Visualizations

Our data visualizations revealed several important trends. First, life expectancy and GDP per capita show strong positive relationships with happiness. However, the impact of GDP levels off starting from about $50,000 per capita. This suggests that economic prosperity isn't enough to fully explain variances in well-being.

Health factors such as vaccination rates and lower infant mortality were consistently associated with longer life expectancy and higher happiness scores. This pattern highlights the importance of public health investments. Infrastructure measures such as sanitation access appeared to play a role up to a point while internet access showed a more strong and consistent positive relationship with happiness. This reflects the increasing importance of digital connection for economic and social inclusion.

Social and political factors also played a key role where countries with lower perceived corruption and greater personal freedoms reported higher happiness levels. The correlation matrix reinforced this idea where trust, generosity, and institutional quality were heavily linked to happiness levels.

Altogether, these findings suggest that economic wealth provides an important foundation for happiness and factors such as healthcare access, digital infrastructure, institutional trust, and social cohesion are also crucial especially once basic needs are met. These findings guide our modeling choices as we focus on health, education, infrastructure, and governance indicators alongside economic indicators.

5. Regression Methods

Linear Model:

One of the first tools we used was a basic linear regression model to see how different variables contributed to the Life Ladder score. We regressed the Life Ladder which is our response variable on all the explanatory variables we selected from the World Bank dataset. This simple model served as a baseline to compare with our more advanced approaches. We chose linear regression because it is a classical and widely-used method for modeling quantitative response variables, meaning outcomes that are numerical and continuous, like the Life Ladder score which ranges from 0 to 10.

Surprisingly, the linear model performed very well. Although many individual predictors were not significant at the five percent level, the model as a whole produced a test root mean squared error (RMSE) of 0.449 which is smaller than the errors for any other methods we tested. The model also produced an R-squared value of 0.77 which means that it explains about 77% of the variation in Life Ladder scores. It is important to acknowledge the limitations even though this is impressive. The linear model assumes a strictly linear relationship between predictors and happiness which may not fully capture the true complexity. Also, including many correlated predictors can risk overfitting. However, the strong test performance suggests that overfitting was not a major issue in this case.

The linear model highlighted several important factors for happiness which included infant mortality rate, inflation, and access to basic drinking water. These findings are consistent with the literature review since health security, economic stability, and basic needs satisfaction were major contributors to well-being.

Composite Scores:

Another tool we used was the creation of composite scores to group our explanatory variables into broader categories. These composite scores were created to simplify our large number of variables into groups of variables related to each other. Based on our initial literature review and exploratory data analysis, we identified five main themes: Economic Variables, Social Variables, Health Variables, Education Variables, and Infrastructure Variables. Condensing the variables into these categories not only simplified our model but also enabled viewers to more easily interpret the broad categories of explanatory variables we were considering.

Since the original variables were in different units, we first standardized each variable by calculating its z-score. We then created correlation matrices between each explanatory variable

and Life Ladder scores to help decide which variables to include. Variables with higher correlations were weighted more heavily in their composite score, while weaker variables were either downweighted or eliminated. For example, in the Health composite score, we initially considered seven variables but narrowed it down to six after reviewing correlations and group discussions. Variables like Tuberculosis Treatment Success Rate were removed due to their relatively low association with happiness.

This method allowed us to represent broader aspects of a country's society through single indicators, simplifying the modeling process. However, one limitation was that assigning variables to categories involved some subjective judgment, and weighting decisions were partially based on correlation strength rather than deeper causal modeling. Despite these limitations, regression analysis using the composite scores showed that each broad factor, especially health and economic strength significantly contributed to higher Life Ladder scores.

Principal Component Analysis (PCA)

Given the large number of original variables (47), we also used Principal Component Analysis (PCA) to reduce dimensionality. PCA is an unsupervised learning technique that transforms original variables into a smaller number of principal components (PCs) that capture most of the original variation without using the response variable.

We chose to apply PCA to address two important challenges which are high dimensionality and potential multicollinearity among our predictors. Reducing the number of variables simplifies the model, makes it less prone to overfitting, and makes it easier to interpret. Summarizing many correlated variables into a smaller set of components increases model stability and reduces variance in predictions as well. This is crucial in our project since many of our predictors such as health and economic indicators are often highly correlated.

Before applying PCA, we prepped the data by replacing missing values with median values (to allow PCA to run properly) and dropping non-predictive columns like country name and year. We split the transformed data into an 80% training set and a 20% testing set. Using the prcomp() function in R, we centered and scaled the variables before performing PCA.

The first five principal components accounted for approximately 67% of the variance in Life Ladder scores. Looking at PC1, many of the weighted factors were health-related, such as infant mortality rate, fertility rate, and tuberculosis treatment success rate, similar to the groupings we found when constructing composite scores manually. The model achieved a test RMSE of 0.511 and a test R-squared of 0.767 which indicates a strong predictive accuracy and how the reduced set of components successfully preserved most of the information relevant to Life Ladder scores.

While PCA helped simplify the dataset and identify important underlying themes, its limitations include assuming linear relationships among predictors and making interpretation more difficult, since principal components are combinations of many variables.

Principal Component Regression (PCR):

After PCA, we decided to go a step further by using Principal Component Regression (PCR), which combines dimensionality reduction with supervised learning. Unlike PCA, which only looks at variance among predictors, PCR also considers how well the principal components predict the outcome variable (Life Ladder score). This method was chosen because we had 47 variables and wanted to reduce dimensions. This would make model prediction easier because a smaller set of principal components would still retain most of the information.

As with PCA, we handled missing data by filling with medians and dropped irrelevant columns. We trained the PCR model on 80% of the data and tested on 20%. Cross-validation was used to assess how many components to keep for best performance.

The first five principal components explained about 73% of the variance in Life Ladder scores, and the model achieved a test RMSE of approximately 0.57. These results were encouraging and suggested that even a small number of broad "themes" could predict happiness scores fairly well.

However, PCR shares some limitations with PCA: it assumes linear relationships, it selects predictors based on variance rather than direct importance to the outcome, and it can be sensitive to outliers. Nonetheless, PCR helped confirm that key areas like health, infrastructure, and governance could be bundled into meaningful dimensions that predict happiness.

Table 1: Results of Regression Methods

Comparison of Regression Models		
Models	**Test RMSE**	**Test R-squared**
Linear Regression	0.449	0.77
Composite Scores	0.417	0.757
Principal Component Analysis	0.511	0.767
Principal Component Regression	0.57	0.714

6. Classification Methods

Classification Using Logistic Regression:

After obtaining regression results from our different methods, we also thought it could be valuable to attempt to classify countries as happy or unhappy based on the median value of Life Ladder score. After separating observations into their respective categories, we performed logistic regression using the first four Principal Components generated from PCA. We performed logistic regression because it works well to classify binary response variables based on either quantitative or categorical predictors. We also used this as a baseline to compare with our more advanced tree models and random forest models. This resulted in an accuracy of 0.86, with the

precision metric was 0.90 and the recall was also 0.86. This indicates that our logistic regression model was very good at predicting whether a country would be classified as happy or unhappy based on the initial predictor values and the Principal Components.

Classification Using Decision Tree:

We performed two decision tree classifications where one used the original predictor variables and the other used the first ten PCs generated through PCA. Decision trees are useful for discovering non-linear and interaction effects among variables since they don't assume linear decision boundaries. Rather, they separate the data into rectangular regions based on variable thresholds. This is important because it makes them flexible when modeling complex patterns. It is important to note that this flexibility decision trees offer also can lead to overfitting especially when the true decision boundaries are closer to linear.

The first decision tree resulted in an accuracy of 0.84, a precision of 0.87, and a recall of 0.79. Important drivers of happiness were identified which included Log GDP per capita, Freedom to make life choices, Social support, and Healthy life expectancy at birth. The second decision tree was built to improve performance and reduce dimensionality. This PCA-based decision tree slightly outperformed the previous model as it resulted in an accuracy of 0.87, a precision of 0.86, and a recall of 0.95 which is a way higher value. Even though the PCs aren't directly interpretable themselves, this strong result proves how effective the model was at capturing the overall structure of the data to better classify happiness.

Overall, the first tree offered better interpretability regarding which factors drive happiness while the second tree offered stronger classification performance as it condensed the information from the original predictors.

Classification Using Random Forest:

We wanted to use Random Forest because it takes the averages of many decision trees to try and reduce overfitting and improve generalization. Using random forest to classify countries as happy or unhappy based on median Life Ladder score, we fine tuned the model to optimize for the best outcome possible. At each split, eight variables are randomly selected. This helps with overfitting and makes the tree more diverse. Variables were dropped looking at their Mean Decrease Accuracy values to increase model performance. The final random forest model achieved an accuracy score of 0.80, precision score of 0.70, and a recall score of 0.86. The model was able to correctly identify most positive cases.

Table 2: Results of Classification Methods

Results of Classification Analysis Methods			
Models	**Accuracy**	**Precision**	**Recall**
Logistic Regression	86%	90%	86%
Decision Tree	87%	86%	95%
Random Forest	80%	70%	86%

Modeling Summary

Using a variety of modeling approaches: linear regression, composite scores, PCA, and PCR allowed us to validate and cross-check our findings. Across all methods, public health, economic strength, freedom, and social trust consistently emerged as key drivers of happiness. These conclusions align well with the existing literature, such as Ruggeri's multidimensional analysis of well-being and Kshtriya et al.'s emphasis on Health, Hope, and Harmony.

When looking at classification methods, we used accuracy, precision, and recall as metrics to assess success. Since it is important to be able to correctly classify happy and unhappy countries equally, high accuracy is favorable. Decision tree was the best model when it came to classifying countries as happy or unhappy. It had the highest accuracy score of 0.87, which means that it was able to correctly match the true happiness of the countries.

The variety of models also emphasized that happiness cannot be captured by any one single metric; rather, it is shaped by a broad constellation of interconnected factors. Our early exploratory visualizations pointed toward these same trends, and our modeling efforts confirmed their importance.

7. Conclusions

Our project confirmed that happiness cannot be explained by just one factor. While economic strength and wealth are important foundations, our analysis showed that once basic needs are met, other dimensions like health, safety, personal freedom, and social trust play a much larger role in shaping well-being.

Across all of our models — whether it was simple linear regression, composite scoring, PCA, or PCR - the major drivers of happiness remained consistent. Health indicators, trust in institutions, freedom to make life choices, and digital connectivity emerged as key components alongside traditional economic measures.

Ultimately, building happier societies is not just about boosting GDP. It's about creating environments where people feel safe, healthy, connected, and empowered. Happiness is complex, and that complexity is what makes it such an important and fascinating area of study.

Appendix A: References

Blanchflower, D.G. (2021). Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. Journal of Population Economics, 34, 575–624. https://doi.org/10.1007/s00148-020-00797-z

Jaswal, V., et al. (2020). Understanding the determinants of happiness through the Gallup World Poll. Journal of Family Medicine and Primary Care, 9(9), 4826–4832. https://doi.org/10.4103/jfmpc.jfmpc_156_20

Behera, D.K., Rahut, D.B., Padmaja, M., & Dash, A.K. (2024). Socioeconomic determinants of happiness: Empirical evidence from developed and developing countries. Journal of Behavioral and Experimental Economics, 109, 102187. https://doi.org/10.1016/j.socec.2024.102187

Kshtriya, S., Singh, S., & Valk, R. (2023). Health, Hope, and Harmony: A Systematic Review of the Determinants of Happiness across Cultures and Countries. International Journal of Environmental Research and Public Health, 20(4), 3306. https://doi.org/10.3390/ijerph20043306

Pursuit of Happiness Organization. (2024). Abraham Maslow's Psychology: Pioneering the Pursuit of Happiness through the Hierarchy of Needs. Retrieved from https://www.pursuit-of-happiness.org/history-of-happiness/abraham-maslow/

Ruggeri, K. (2024). Well-being is More Than Happiness and Life Satisfaction: A Multidimensional Analysis of 21 Countries. Retrieved from [source not fully provided — if you have a DOI or link, I can add it]

OpenStax College (Rice University). (n.d.). Perspectives on Happiness. Retrieved from https://oertx.highered.texas.gov/courseware/lesson/2165/overview

Anya (UCLA DataRes). (n.d.). Quantifying Happiness: Exploring Factors Affecting Happiness. Retrieved from https://ucladatares.medium.com/quantifying-happiness-exploring-factors-affecting-happiness-e9376d32c847

Appendix B: Data Dictionary

Variable	Definition	Type
Access to electricity (% of population)	Access to electricity is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources.	Numerical
Adequacy of social insurance programs (% of total welfare of beneficiary households)	Adequacy of social insurance programs is measured by the total transfer amount received by the population participating in social insurance programs as a share of their total welfare. Welfare is defined as the total income or total expenditure of beneficiary households. Social insurance programs include old age contributory pensions (including survivors and disability) and social security and health insurance benefits (including occupational injury benefits, paid sick leave, maternity and other social insurance). Estimates include both direct and indirect beneficiaries.	Numerical
Adequacy of social protection and labor programs (% of total welfare of beneficiary households)	Adequacy of social protection and labor programs (SPL) is measured by the total transfer amount received by the population participating in social insurance, social safety net, and unemployment benefits and active labor market programs as a share of their total welfare. Welfare is defined as the total income or total expenditure of beneficiary households. Estimates include both direct and indirect beneficiaries.	Numerical
Adequacy of social safety net programs (% of total welfare of beneficiary households)	Adequacy of social safety net programs is measured by the total transfer amount received by the population participating in social safety net programs as a share of their total welfare. Welfare is defined as the total income or total expenditure of beneficiary households. Social safety net programs include cash transfers and last resort programs, noncontributory social pensions, other cash transfers programs (child, family and orphan allowances, birth and death grants, disability benefits, and other allowances), conditional cash transfers, in-kind food transfers (food stamps and vouchers, food rations, supplementary feeding, and emergency food distribution), school feeding, other social assistance programs (housing allowances, scholarships, fee waivers, health subsidies, and other social assistance) and public works programs (cash for work and food for work). Estimates include both direct and	Numerical

	indirect beneficiaries.	
Adolescent fertility rate (births per 1,000 women ages 15-19)	Adolescent fertility rate is the number of births per 1,000 women ages 15-19.	Numerical
Control of corruption: percentile rank	Control of Corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. Percentile rank indicates the country's rank among all countries covered by the aggregate indicator, with 0 corresponding to lowest rank, and 100 to highest rank. Percentile ranks have been adjusted to correct for changes over time in the composition of the countries covered by the WGI.	Numerical
Coverage of social insurance programs (% of population)	Coverage of social insurance programs shows the percentage of population participating in programs that provide old age contributory pensions (including survivors and disability) and social security and health insurance benefits (including occupational injury benefits, paid sick leave, maternity and other social insurance). Estimates include both direct and indirect beneficiaries.	Numerical
Coverage of social protection and labor programs (% of population)	Coverage of social protection and labor programs (SPL) shows the percentage of population participating in social insurance, social safety net, and unemployment benefits and active labor market programs. Estimates include both direct and indirect beneficiaries.	Numerical
Coverage of social safety net programs (% of population)	Coverage of social safety net programs shows the percentage of population participating in cash transfers and last resort programs, noncontributory social pensions, other cash transfers programs (child, family and orphan allowances, birth and death grants, disability benefits, and other allowances), conditional cash transfers, in-kind food transfers (food stamps and vouchers, food rations, supplementary feeding, and emergency food distribution), school feeding, other social assistance programs (housing allowances, scholarships, fee waivers, health subsidies, and other social assistance) and public works programs (cash for work and food for work). Estimates include both direct and indirect beneficiaries.	Numerical
Domestic general government health expenditure (% of general	Public expenditure on health from domestic sources as a share of total public expenditure. It indicates the priority of the	Numerical

government expenditure)	government to spend on health from own domestic public resources.	
Domestic general government health expenditure per capita (current US$)	Public expenditure on health from domestic sources per capita expressed in current US dollars.	Numerical
Domestic private health expenditure (% of current health expenditure)	Share of current health expenditures funded from domestic private sources. Domestic private sources include funds from households, corporations and non-profit organizations. Such expenditures can be either prepaid to voluntary health insurance or paid directly to healthcare providers.	Numerical
Domestic private health expenditure per capita (current US$)	Current private expenditures on health per capita expressed in current US dollars. Domestic private sources include funds from households, corporations and non-profit organizations. Such expenditures can be either prepaid to voluntary health insurance or paid directly to healthcare providers.	Numerical
Educational attainment, at least Bachelor's or equivalent, population 25+, total (%) (cumulative	The percentage of population ages 25 and over that attained or completed Bachelor's or equivalent.	Numerical
Educational attainment, at least completed upper secondary, population 25+, total (%) (cumulative)	The percentage of population ages 25 and over that attained or completed upper secondary education.	Numerical
Educational attainment, at least Master's or equivalent, population 25+, total (%) (cumulative)	The percentage of population ages 25 and over that attained or completed Master's or equivalent	Numerical
Educational attainment, Doctoral or equivalent, population 25+, total (%) (cumulative)	The percentage of population ages 25 and over that attained or completed Doctoral or equivalent.	Numerical
Fertility rate, total (births per woman)	Total fertility rate represents the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year.	Numerical
Food Production Index (2014-2016 = 100)	Food production index covers food crops that are considered edible and that contain nutrients. Coffee and tea are excluded because, although edible, they have no nutritive value.	Numerical

Foreign direct investment, net (BoP, current US$)	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows total net FDI. In BPM6, financial account balances are calculated as the change in assets minus the change in liabilities. Net FDI outflows are assets and net FDI inflows are liabilities. Data are in current U.S. dollars.	Numerical
Foreign direct investment, net inflows (% of GDP)	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows net inflows (new investment inflows less disinvestment) in the reporting economy from foreign investors, and is divided by GDP.	Numerical
Foreign direct investment, net outflows (% of GDP)	Foreign direct investment refers to direct investment equity flows in an economy. It is the sum of equity capital, reinvestment of earnings, and other capital. Direct investment is a category of cross-border investment associated with a resident in one economy having control or a significant degree of influence on the management of an enterprise that is resident in another economy. Ownership of 10 percent or more of the ordinary shares of voting stock is the criterion for determining the existence of a direct investment relationship. This series shows net outflows of investment from the reporting economy to the rest of the world, and is divided by GDP.	Numerical
Foreign direct investment, net outflows (BoP, current US$)	Foreign direct investment refers to direct investment equity flows in an economy. It is the sum of equity capital, reinvestment of earnings, and other capital. Direct investment is a category of cross-border investment associated with a resident in one economy having control or a significant degree of influence on the management of an enterprise that is resident in another economy. Ownership of 10 percent or more of the ordinary shares of voting stock is the criterion for	Numerical

	determining the existence of a direct investment relationship. This series shows net outflows of investment from the reporting economy to the rest of the world. Data are in current U.S. dollars.	
Forest area (% of land area)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.	Numerical
Forest area (sq. km)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems	Numerical
GDP (current US$)	Gross domestic product (GDP) at prices of the current reporting period, expressed in terms of current prices in USD.	Numerical
GDP per capita (current US$)	GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.	Numerical
GDP, PPP (current international $)	This indicator provides values for gross domestic product (GDP) expressed in current international dollars, converted by purchasing power parity (PPP) conversion factor. GDP is the sum of gross value added by all resident producers in the country plus any product taxes and minus any subsidies not included in the value of the products. PPP conversion factor is a spatial price deflator and currency converter that eliminates the effects of the differences in price levels between countries. From April 2020, “GDP: linked series (current LCU)” [NY.GDP.MKTP.CN.AD] is used as underlying GDP in local currency unit so that it’s in line with time series of PPP conversion factors for GDP, which are extrapolated with linked GDP deflators.	Numerical
Gini index	The Gini index measures the extent to which the distribution of income or consumption among individuals or households within an economy deviates from a perfectly equal distribution. A Gini index of 0 represents perfect equality,	Numerical

	while an index of 100 implies perfect inequality.	
Government expenditure on education, total (% of GDP)	General government expenditure on education (current, capital, and transfers) is expressed as a percentage of GDP. It includes expenditure funded by transfers from international sources to government. General government usually refers to local, regional and central governments.	Numerical
Government expenditure on education, total (% of government expenditure)	General government expenditure on education (current, capital, and transfers) is expressed as a percentage of total general government expenditure on all sectors (including health, education, social services, etc.). It includes expenditure funded by transfers from international sources to government. General government usually refers to local, regional and central governments.	Numerical
High-technology exports (current US$)	High-technology exports are products with high R&D intensity, such as aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery. Data are in current U.S. dollars.	Numerical
Households and NPISHs Final consumption expenditure, PPP (current international $)	This indicator provides values for households and NPISHs final consumption expenditure expressed in current international dollars converted by purchasing power parity (PPP) conversion factor. Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households. It excludes purchases of dwellings but includes imputed rent for owner-occupied dwellings. It also includes payments and fees to governments to obtain permits and licenses. Here, household consumption expenditure includes the expenditures of nonprofit institutions serving households, even when reported separately by the country. PPP conversion factor is a spatial price deflator and currency converter that eliminates the effects of the differences in price levels between countries. From July 2020, "Households and NPISHs final consumption expenditure: linked series (current LCU)" [NE.CON.PRVT.CN.AD] is used as underlying expenditure in local currency unit so that it's in line with time series of PPP conversion factor, private consumption (LCU per international $), which are extrapolated with linked CPI.	Numerical

Immunization, DPT (% of children ages 12-23 months)	Child immunization, DPT, measures the percentage of children ages 12-23 months who received DPT vaccinations before 12 months or at any time before the survey. A child is considered adequately immunized against diphtheria, pertussis (or whooping cough), and tetanus (DPT) after receiving three doses of vaccine.	Numerical
Immunization, HepB3 (% of one-year-old children)	Child immunization rate, hepatitis B is the percentage of children ages 12-23 months who received hepatitis B vaccinations before 12 months or at any time before the survey. A child is considered adequately immunized after three doses.	Numerical
Immunization, measles (% of children ages 12-23 months)	Child immunization, measles, measures the percentage of children ages 12-23 months who received the measles vaccination before 12 months or at any time before the survey. A child is considered adequately immunized against measles after receiving one dose of vaccine.	Numerical
Income share held by highest 10%	Percentage share of income or consumption is the share that accrues to subgroups of population indicated by deciles or quintiles.	Numerical
Individuals using the Internet (% of population)	Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.	Numerical
Inflation consumer prices	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used.	Numerical
Intentional Homicides per 100000 people	Intentional homicides are estimates of unlawful homicides purposely inflicted as a result of domestic disputes, interpersonal violence, violent conflicts over land resources, intergang violence over turf or control, and predatory violence and killing by armed groups. Intentional homicide does not include all intentional killing; the difference is usually in the organization of the killing. Individuals or small groups usually commit homicide, whereas killing in armed conflict is usually committed by fairly cohesive groups of up to several hundred members and is thus usually excluded.	Numerical

International tourism, expenditures (% of total imports)	International tourism expenditures are expenditures of international outbound visitors in other countries, including payments to foreign carriers for international transport. These expenditures may include those by residents traveling abroad as same-day visitors, except in cases where these are important enough to justify separate classification. For some countries they do not include expenditures for passenger transport items. Their share in imports is calculated as a ratio to imports of goods and services, which comprise all transactions between residents of a country and the rest of the world involving a change of ownership from nonresidents to residents of general merchandise, goods sent for processing and repairs, nonmonetary gold, and services.	Numerical
International tourism, expenditures (current US$)	International tourism expenditures are expenditures of international outbound visitors in other countries, including payments to foreign carriers for international transport. These expenditures may include those by residents traveling abroad as same-day visitors, except in cases where these are important enough to justify separate classification. For some countries they do not include expenditures for passenger transport items. Data are in current U.S. dollars.	Numerical
Life expectancy at birth, total (years)	The average number of years that a newborn could expect to live, if he or she were to pass through life exposed to the sex- and age-specific death rates prevailing at the time of his or her birth, for a specific year, in a given country, territory, or geographic area.	Numerical
Literacy rate, adult total (% of people ages 15 and above)	Adult literacy rate is the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life.	Numerical
Maternal mortality ratio (modeled estimate, per 100,000 live births)	Maternal mortality ratio is the number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births. The data are estimated with a regression model using information on the proportion of maternal deaths among non-AIDS deaths in women ages 15-49, fertility, birth attendants, and GDP measured using purchasing power parities (PPPs).	Numerical
Maternal mortality ratio (national estimate, per 100,000 live births)	Maternal mortality ratio is the number of women who die from pregnancy-related causes while pregnant or within 42 days of	Numerical

	pregnancy termination per 100,000 live births.	
Military expenditure (% of GDP)	Military expenditures data from SIPRI are derived from the NATO definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces; defense ministries and other government agencies engaged in defense projects; paramilitary forces, if these are judged to be trained and equipped for military operations; and military space activities. Such expenditures include military and civil personnel, including retirement pensions of military personnel and social services for personnel; operation and maintenance; procurement; military research and development; and military aid (in the military expenditures of the donor country). Excluded are civil defense and current expenditures for previous military activities, such as for veterans' benefits, demobilization, conversion, and destruction of weapons. This definition cannot be applied for all countries, however, since that would require much more detailed information than is available about what is included in military budgets and off-budget military expenditure items.	Numerical
Military expenditure (current USD)	Military expenditures data from SIPRI are derived from the NATO definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces; defense ministries and other government agencies engaged in defense projects; paramilitary forces, if these are judged to be trained and equipped for military operations; and military space activities. Such expenditures include military and civil personnel, including retirement pensions of military personnel and social services for personnel; operation and maintenance; procurement; military research and development; and military aid (in the military expenditures of the donor country). Excluded are civil defense and current expenditures for previous military activities, such as for veterans' benefits, demobilization, conversion, and destruction of weapons. This definition cannot be applied for all countries, however, since that would require much more detailed information than is available about what is included in military budgets and off-budget military expenditure items.	Numerical

Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%)	Mortality from CVD, cancer, diabetes or CRD is the percent of 30-year-old-people who would die before their 70th birthday from any of cardiovascular disease, cancer, diabetes, or chronic respiratory disease, assuming that s/he would experience current mortality rates at every age and s/he would not die from any other cause of death (e.g., injuries or HIV/AIDS).	Numerical
Mortality rate, infant (per 1,000 live births)	Infant mortality rate is the number of infants dying before reaching one year of age, per 1,000 live births in a given year.	Numerical
Multidimensional poverty headcount ratio (UNDP) (% of population)	the percentage of a population that is multidimensionally poor	Numerical
Multidimensional poverty headcount ratio (World Bank) (% of population)	The percentage of people who are multidimensionally poor	Numerical
Net migration	Net migration is the net total of migrants during the period, that is, the number of immigrants minus the number of emigrants, including both citizens and noncitizens.	Numerical
Patent applications, nonresidents	Patent applications are worldwide patent applications filed through the Patent Cooperation Treaty procedure or with a national patent office for exclusive rights for an invention--a product or process that provides a new way of doing something or offers a new technical solution to a problem. A patent provides protection for the invention to the owner of the patent for a limited period, generally 20 years.	Numerical
Patent applications, residents	Patent applications are worldwide patent applications filed through the Patent Cooperation Treaty procedure or with a national patent office for exclusive rights for an invention--a product or process that provides a new way of doing something or offers a new technical solution to a problem. A patent provides protection for the invention to the owner of the patent for a limited period, generally 20 years.	Numerical
People using at least basic drinking water services (% of population)	The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services. Basic drinking water services is defined as drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes or tubewells, protected dug wells, protected springs, and packaged or delivered water.	Numerical

People using at least basic sanitation services (% of population)	The percentage of population using at least basic sanitation services, that is, improved sanitation facilities that are not shared with other households. This indicator encompasses both people using basic sanitation services as well as those using safely managed sanitation services. Improved sanitation facilities include flush/pour flush toilets connected to piped sewer systems, septic tanks or pit latrines; pit latrines with slabs (including ventilated pit latrines), and composting toilets.	Numerical
PM2.5 air pollution, population exposed to levels exceeding WHO guideline value (% of total)	Percent of population exposed to ambient concentrations of PM2.5 that exceed the WHO guideline value is defined as the portion of a country's population living in places where mean annual concentrations of PM2.5 are greater than 10 micrograms per cubic meter, the guideline value recommended by the World Health Organization as the lower end of the range of concentrations over which adverse health effects due to PM2.5 exposure have been observed.	Numerical
Prevalence of moderate or severe food insecurity in the population (%)	The percentage of individuals in the population who have experienced food insecurity at moderate or severe levels during the reference period.	Numerical
Proportion of population pushed below the 60% median consumption poverty line by out-of-pocket health expenditure (%)	The proportion of people pushed further into poverty by out-of-pocket health expenses is equivalent to the proportion of poor individuals who spend on healthcare. Poverty status is determined based on household income or consumption, including health expenditures, measured per capita per day. Household sample weights ensure population representativeness. A country-specific relative poverty line, set at 60% of the median daily per capita income, highlights the link between reducing poverty (SDG 1.2) and achieving Universal Health Coverage (SDG 3.8).	Numerical
Proportion of seats held by women in national parliaments (%)	Women in parliaments are the percentage of parliamentary seats in a single or lower chamber held by women.	Numerical
Suicide mortality rate (per 100,000 population)	Suicide mortality rate is the number of suicide deaths in a year per 100,000 population. Crude suicide rate (not age-adjusted)	Numerical
Total greenhouse gas emissions excluding LULUCF per capita (t CO2e/capita)	a measurement of the average amount of greenhouse gas emissions per person in a region or country. It's calculated by dividing the total emissions by the population.	Numerical

Trademark applications, nonresident, by count	Trademark applications filed are applications to register a trademark with a national or regional Intellectual Property (IP) offices and designations received by relevant offices through the Madrid System. A trademark is a distinctive sign which identifies certain goods or services as those produced or provided by a specific person or enterprise. A trademark provides protection to the owner of the mark by ensuring the exclusive right to use it to identify goods or services, or to authorize another to use it in return for payment. The period of protection varies, but a trademark can be renewed indefinitely beyond the time limit on payment of additional fees. Non-resident application refers to an application filed with the IP office of or acting on behalf of a state or jurisdiction in which the first-named applicant in the application is not domiciled. Class count is used to render application data for trademark applications across offices comparable, as some offices follow a single-class/single-design filing system while other have a multiple class/design filing system.	Numerical
Trademark applications, resident, by count	Trademark applications filed are applications to register a trademark with a national or regional Intellectual Property (IP) offices and designations received by relevant offices through the Madrid System. A trademark is a distinctive sign which identifies certain goods or services as those produced or provided by a specific person or enterprise. A trademark provides protection to the owner of the mark by ensuring the exclusive right to use it to identify goods or services, or to authorize another to use it in return for payment. The period of protection varies, but a trademark can be renewed indefinitely beyond the time limit on payment of additional fees. Resident application refers to an application filed with the IP office of or acting on behalf of the state or jurisdiction in which the first-named applicant in the application has residence. Class count is used to render application data for trademark applications across offices comparable, as some offices follow a single-class/single-design filing system while other have a multiple class/design filing system.	Numerical
Tuberculosis treatment success rate (% of new cases)	Tuberculosis treatment success rate is the percentage of all new tuberculosis cases (or new and relapse cases for some countries) registered under a national tuberculosis control	Numerical

	programme in a given year that successfully completed treatment, with or without bacteriological evidence of success ("cured" and "treatment completed" respectively).	
UHC service coverage index	The coverage of essential health services (defined as the average coverage of essential services based on tracer interventions that include reproductive, maternal, newborn and child health, infectious diseases, non-communicable diseases and service capacity and access, among the general and the most disadvantaged population). The indicator is an index reported on a unitless scale of 0 to 100, which is computed as the geometric mean of 14 tracer indicators of health service coverage. The tracer indicators are as follows, organized by four components of service coverage: 1. Reproductive, maternal, newborn and child health 2. Infectious diseases 3. Noncommunicable diseases 4. Service capacity and access.	Numerical
Unemployment, total (% of total labor force) (modeled ILO estimate)	Unemployment refers to the share of the labor force that is without work but available for and seeking employment.	Numerical
Unemployment, total (% of total labor force) (national estimate)	Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.	Numerical
Voice and Accountability: Percentile Rank	Voice and Accountability captures perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media. Percentile rank indicates the country's rank among all countries covered by the aggregate indicator, with 0 corresponding to lowest rank, and 100 to highest rank. Percentile ranks have been adjusted to correct for changes over time in the composition of the countries covered by the WGI. Percentile Rank Upper refers to upper bound of 90 percent confidence interval for governance, expressed in percentile rank terms.	Numerical
Women Business and the Law Index Score (scale 1-100)	The index measures how laws and regulations affect women’s economic opportunity. Overall scores are calculated by taking the average score of each index (Mobility, Workplace, Pay, Marriage, Parenthood, Entrepreneurship, Assets and Pension), with 100 representing the highest possible score.	Numerical

Women making their own informed decisions regarding sexual relations, contraceptive use and reproductive health care (% of women age 15-49)	Proportion of women ages 15-49 years (married or in union) who make their own decision on all three selected areas i.e. can say no to sexual intercourse with their husband or partner if they do not want; decide on use of contraception; and decide on their own health care. Only women who provide a “yes” answer to all three components are considered as women who “make her own decisions regarding sexual and reproductive”.	Numerical