WORLD HAPPINESS SUPERVISED LEARNING MODEL

Belen Ramirez, Ben Bucaj, Nguyen Le, and Ryan Rodriguez

WHAT'S THE PROBLEM?

- Key Question: What makes a country happy?

- The Problem:
  - Top countries like Finland, Denmark, and Switzerland consistently rank high.
  - Many nations face significant challenges in happiness.
  - Goal: Understand disparities and what drives happiness globally.

WORLD HAPPINESS REPORT: Uses the metric of happiness to rank countries. Governments and organizations use It to guide policy decisons.

ABOUT

Published annually since 2012 by Gallup, the UN Sustainable Development Solutions Network, and the Oxford Wellbeing Research Center.

BASED ON GALLUP WORLD POLL

Over 100,000 respondents from 130+ countries rate their lives on a 0-10 scale (Cantril ladder).

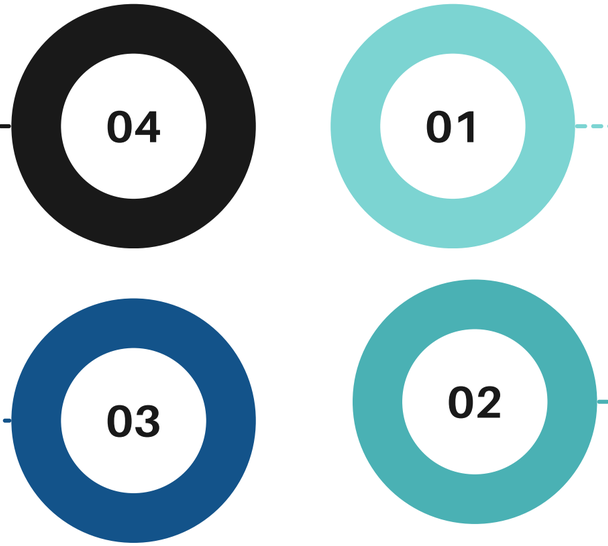
SIX EXPLANATORY FACTORS

GDP, social support, health, freedom, generosity, and trust (corruption).

WHY THE PROBLEM IS IMPORTANT

Policy Success

The United Nations urges member states to prioritize happiness in social and economic policies. Studies like the World Happiness Report have inspired countries to integrate well-being indicators into governance (e.g, New Zealand's Well-being Budget, UAE's Ministry of Happiness, etc.).



Global Impact

Happier countries have healthier populations, stronger economies, higher trust in institutions (e.g., Finland).

Mental Health Crisis

Depression costs the global economy $1 trillion annually from lost productivity (NIH).

Economic Productivity

Happier employees are 13% more productive (University of Oxford).

Our Project Goal:

- Analyze the World Happiness Report to identify factors driving happiness.
- Provide actionable insights for policymakers to improve well-being globally.

DATA DESCRIPTION

Source: Kaggle, World Happiness Report https://www.kaggle.com/datasets/unsdsn/world-happiness/data

Economy (GDP per Capita)
This predictor measures a country's economic strength in terms of GDP per capita, adjusted for Purchasing Power Parity (PPP) in constant international dollars.

Family (Social Support)
It indicates the strength of social networks and the availability of reliable relationships, which are key contributors to emotional and psychological well-being.

Health (Life Expectancy)
This measures the healthy life expectancy at birth, based on data from the World Health Organization (WHO).

Freedom (Freedom to Make Life Choices)
This variable reflects the extent to which individuals feel free to make personal life decisions.

We had **five csv files**, consisting of world happiness reports from 2015 to 2019.

Trust (Government Corruption)
Trust measures perceptions of corruption in government and business, based on binary responses to questions about whether corruption is widespread.

Generosity
Generosity is derived from responses to the question: "Have you donated money to a charity in the past month?"

Region
Region captures the geographical and cultural context in which a country exists.

Happiness Rankings
Based on answers to the main life evaluation question asked in the poll. This question, known as the Cantril ladder, asks respondents to think of a ladder with the best possible life for them being a 10 and the worst possible life being a 0 and to rate their own current lives on that scale.

PREPROCESSING

Stage 1. Combining the datasets

We had five csv files, consisting of world happiness reports from 2015 to 2019.

Stage 2. Dropping unnecessary or duplicate rows and columns

- Dystopia Residual, Lower Confidence Interval, Upper Confidence Interval, Whisker.high, and Whisker.low are columns that only exist in two datasets.
- Happiness Score and Standard Error are two unnecessary columns for our analysis. Happiness Score is a metric measured in 2015 by asking the sampled people the question: "How would you rate your happiness?", and Standard Error is the standard error of the hapiness score.
- Happiness.Score and Score are duplicates of the Happiness Score that was deleted. Dystopia.Residual is a duplicate of Dystopia Residual that was deleted.
- Rows where the Country is Northern Cyprus or Northern Macedonia was removed because only the 2018 and 2019 datasets have this rows.

Stage 3. Misclassification

One row got misclassified as Somalia. Somaliland is a self-declared independent state that considers itself separate from Somalia, though it is internationally recognized as part of Somalia.

Stage 4. Dealing with NaN's in region column

- Country has 465 missing values
- Trust has 1 missing value
- Country had duplicates
- Checking missing region data for each country

Stage 5. Final preprocessing

- Defining classes for training.
- Remove non-useful columns like 'Country', 'index'.

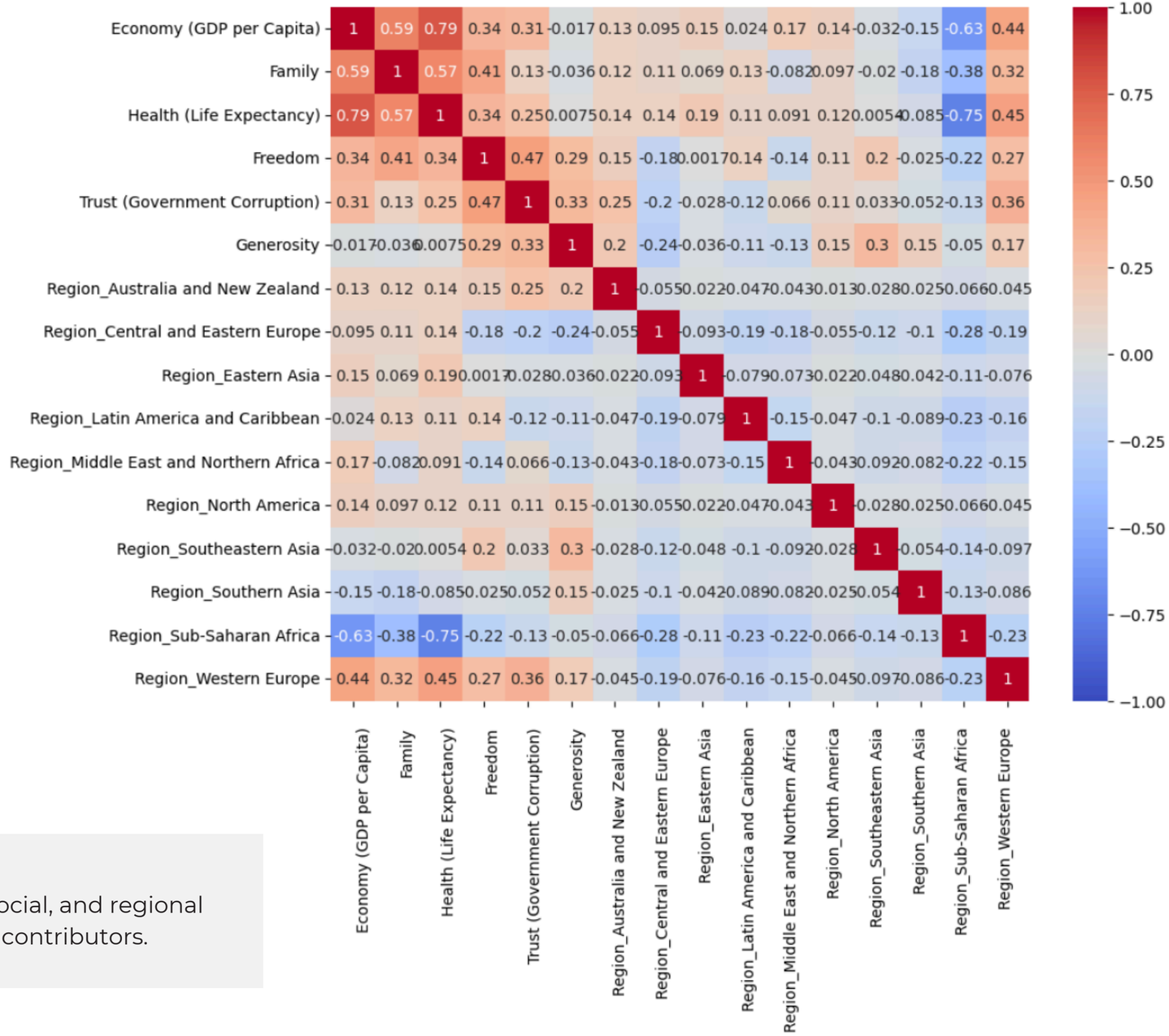
INSIGHTS FROM THE CORRELATION MATRIX

Positive Correlations:

- Economy (GDP per Capita): Strongly linked to Family and Health (Life Expectancy), showing that wealthier nations often enjoy better social support and health outcomes.
- Family: Positively correlated with Freedom, Health, and Economy, indicating the critical role of social support in fostering personal well-being.
- Freedom and Trust: Higher personal freedom often aligns with greater trust in institutions.

Negative Correlations:

- Region (Sub-Saharan Africa): Negatively correlated with Economy and Health, highlighting economic and healthcare challenges in this region.



Takeaway:
Happiness is driven by a multi-dimensional interplay of economic, social, and regional factors, where economic stability, social support, and health are key contributors.

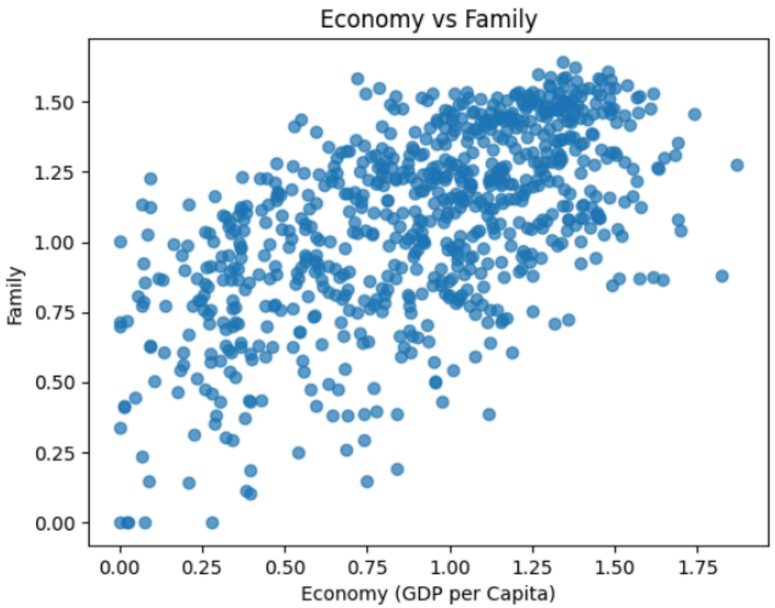
SCATTERPLOT ANALYSIS: ECONOMY VS. FAMILY & ECONOMY VS. HEALTH

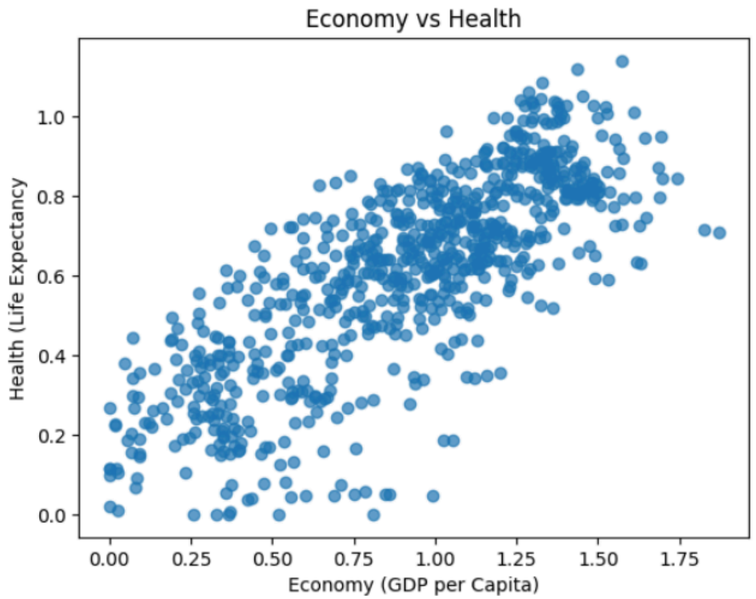
Key Observations:

- Economy vs. Family: Wealthier countries report stronger interpersonal support networks, but cultural and societal norms heavily influence this relationship, resulting in a bit of a weaker correlation compared to health but is still strong.
- Economy vs. Health: Wealthier nations tend to have higher life expectancy, driven by better access to healthcare, education, and resources. The positive trend highlights GDP's significant influence, though it is not the sole factor.

Supporting Evidence:

- Research from the Journal of Population Sciences and Opportunity Insights emphasizes the role of GDP in improving living conditions and fostering connected communities.





Takeaway:
Economic prosperity enhances health and social support but interacts with cultural and societal factors to shape these outcomes.

SCATTERPLOT ANALYSIS: TRUST VS FREEDOM

Key Observations:

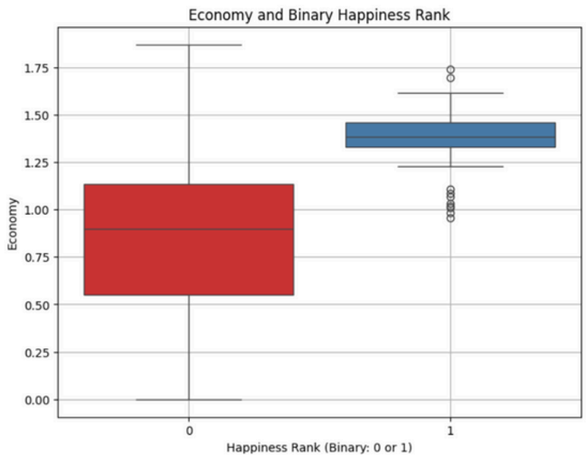
- Majority of countries show low trust in government (below 0.1), even when freedom ranges between 0.3–0.5.
- Slight positive trend: Higher trust is often associated with greater freedom.
- Insight: Trust in government is not a universal requirement for experiencing personal freedom.
- Cultural values or local governance policies can compensate for low trust.

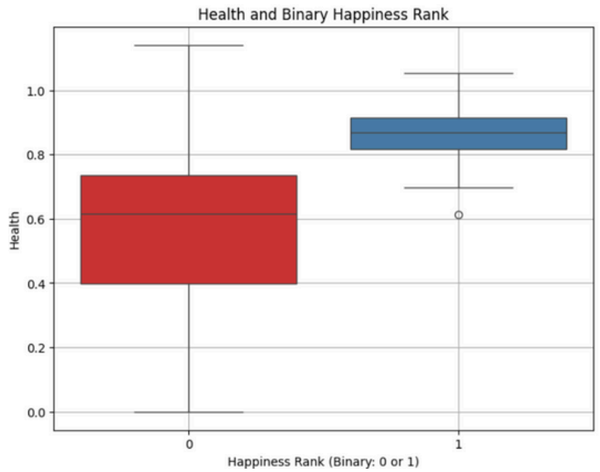
Supporting Evidence:

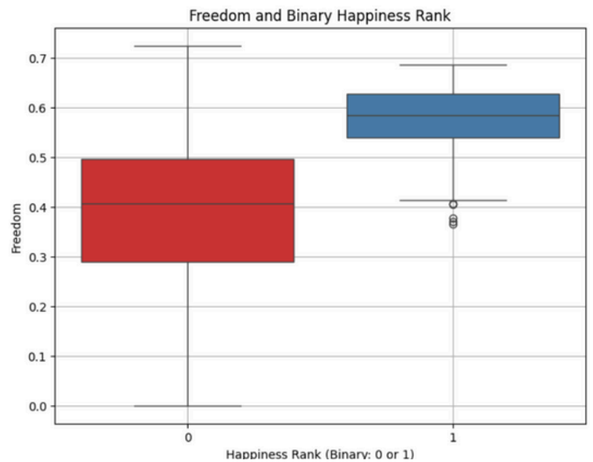
- Pew Research Center:
  - Established democracies (e.g., USA, France) show low public trust in government despite strong institutional freedoms.
  - Personal freedoms can persist independently of high trust in government due to ingrained democratic norms and liberties.

Takeaway:
Trust and freedom are linked but not interdependent; governance and cultural resilience play key roles in sustaining freedoms.

BOX PLOT ANALYSIS: FACTORS INFLUENCING HAPPINESS RANKINGS







Takeaway:
Social, economic, and governance factors collectively shape happiness, with top-performing countries excelling in all these areas.

Box plots compare countries in the Top 20 happiness rankings (binary: Top 20 = 1, Not = 0) across key factors, for example:

1. Family
   - Insight: Top 20 countries show higher median Family scores with fewer outliers, indicating strong social support as a key happiness driver.
2. Economy (GDP per Capita)
   - Insight: Top 20 countries consistently have higher GDP per capita, emphasizing economic prosperity as a major contributor to happiness.
3. Health (Life Expectancy)
   - Insight: Higher median health scores in the Top 20 reflect the critical role of access to healthcare and longevity in well-being.
4. Freedom
   - Insight: Top 20 countries report higher personal freedom, underlining the importance of autonomy in happiness rankings.

PREDICTED METHODS

Predictive Method	Decision Tree	Random Forest	Logistic Regression	KNN	Lasso Regression
Why chosen	Intuitive and easy to interpret	Ensemble method reducing overfitting, common issue with decision trees.	Simple yet effective model for classification problems	Capture complex relationships by considering proximity of data points	Feature selection by shrinking less important feature coefficients
Parameter Tuning	Testing various max_depth values and selecting the one yielding the highest accuracy.	Tuning involves adjusting the number of trees or the maximum depth.	We used the default settings, focusing on feature scaling to ensure the model's assumptions were met.	We used cross-validation to determine the optimal number of neighbors (k).	We tested a range of alpha values to find the one that minimized the MSE.
Performance	The decision tree performed well, but lowest accuracy of 92%.	This model achieved the highest accuracy of 97%.	It achieved an accuracy of around 96%, which was second highest.	The accuracy was second highest similar to logistic regression.	Comparable with Logistic and KNN

WHICH MODEL WAS MOST ACCURATE? WHY?

The Random Forest Classifier was the most accurate model, achieving an accuracy of approximately 97%.

Thanks to: the model's ensemble nature

Combines predictions from multiple decision trees.

Each trained on a different subset of the data and with different features.

Captures a wider range of patterns in the data.

Reduces overfitting + improves generalization to unseen data: a common problem with individual decision trees.

BIGGEST CHALLENGES

01 UNDERSTANDING AND CONFIGURING NESTED CV

Challenge:
Understanding how inner and outer loop of nested CV works and keeping track of how the folds interact

How we overcame:
Studying resources on nested CV, breaking the process into smaller, focused components, and testing with simple models before scaling up.

02 FINDING PROPER DATASET + REFINING VARIABLE DESCRIPTION

Challenge:
Finding a dataset that was not AI generated, clean enough to work with it, and that consisted of an intresting topic was difficult.

How we overcame:
Accessed several databases and websites

03 CLEANING THE DATA

Challenge
Dealing with 5 csv files with some inconsistent data was challenging. As well, just making sure the predictors were coded properly was time consuming.

04 TEAM WORKLOAD

Challenge:
Balancing tasks within the team was difficult at the beginning due to changes in our group size. and because of the workload.

How we overcame:
Delegated responsibilities based on strengths to ensure efficiency.

KEY TAKEAWAYS

- Predictive analysis shows GDP, Family, Health, and Freedom as strong contributors (in that order) to happiness
- Our analysis revealed that no single factor is sufficient to bring about happiness. In countries with a high happiness ranking, multiple factors (e.g., those listed above) work together in strong combination.

Takeaway: Look at top ranking countries in the Happiness Index and Identify policies that relate to the aforementioned predictors

For Example: High-ranking countries like Finland demonstrate these values in their policies:

- **Health Factor**: Finland has Universal Healthcare that Is publicly funded and available to Finnish residents (free or discounted healthcare services). Focuses on disease prevention and health promotion
- **Freedom Factor**: Finnish policies give citizens the feeling of control over their own lives. This comes from freedoms like speech, religion, and assembly along with policies like:
- **Family Factor** since government is providing the necessities required for family to support each other)
- **GDP Factor**: Finnish government provides sizable economic support toward citizens contributing toward one of the lowest poverty rates In the world

Policy Implications

- Countries can use happiness metrics as a blueprint for improving well-being.
- Invest in replicating relevant policies In countries with a high happiness ranking
- Shift in Policy Priorities: Governments might prioritize well-being over solely GDP growth, leading to sustainable development and global well-being.

Broader Impacts

- Improved Global Stability: Happier nations are often healthier, more productive, and more stable politically and economically.
- Social Equity: Policies that enhance happiness often reduce inequality, fostering more inclusive societies.
- Inspiration for Developing Nations: The index serves as a framework for nations striving to improve the quality of life despite economic limitations.