

Data Source: The dataset, sourced from **World Happiness Report**, offers valuable insights into global well-being trends and informs policy decisions for governments and organizations seeking to improve quality of life.

Summary of the Dataset Source

- The **World Happiness Report** dataset, **hosted on Kaggle**, compiles happiness scores for various countries based on economic, social, and health factors. The report has been published annually since 2012, using data from the **Gallup World Poll** to measure global happiness levels.

Data Collection and Methodology

- The happiness rankings are derived from survey responses to the **Cantril ladder** question, where individuals rate their life from **0 (worst) to 10 (best)**.
- Scores represent **nationally representative samples** and are weighted using **Gallup methodology** to ensure accuracy.
- Six major factors contribute to happiness scores:
 - **Economic Production (GDP per capita)**
 - **Social Support (Family & Community)**
 - **Health (Life Expectancy)**
 - **Freedom (Personal & Political)**
 - **Generosity (Charitable Behavior)**
 - **Trust (Absence of Corruption)**
- A hypothetical benchmark, Dystopia, represents the lowest observed national values across all six categories. Countries are ranked by comparing their scores against Dystopia

Dataset Limitations

- **Time Constraints:**
 - The dataset only covers 2015-2019, limiting its ability to reflect long-term trends or recent shifts in happiness rankings.
 - It does not account for major global events after 2019, such as pandemics, economic downturns, or political shifts that could dramatically affect happiness levels.
- **Sampling & Representation Issues**
 - Happiness scores rely on Gallup World Poll responses, which are nationally representative but may not fully capture certain populations, such as minority communities or remote areas.
 - Differences in survey methodologies across countries can lead to inconsistencies in responses, making direct comparisons difficult.
- **Interpretation of Happiness Factors**
 - Economic production, social support, health, freedom, generosity, and corruption do not directly contribute to the happiness score but are used to explain variations across countries.
 - Cultural differences influence responses: Some societies may value happiness differently, making their scores not directly comparable with other countries.

- **Residuals & Unexplained Variations**

- The dataset includes residual scores, which capture unexplained components of happiness not covered by the six main factors.
- These residuals can be large and inconsistent, making it difficult to pinpoint exact reasons why certain countries rank higher or lower.

- **Possible Biases in Reporting**

- Happiness scores are based on self-reported evaluations, meaning responses may be influenced by short-term emotions or external factors (e.g., seasonal effects or political events).
- Countries with higher media coverage and more transparent reporting may appear to have higher or lower scores based on how freely citizens express their opinions.

My rationale for choosing this dataset:

Happiness is a fundamental human aspiration, yet it remains elusive for many, with individuals often struggling to pinpoint the precise factors that influence their well-being. Recognizing this complexity, I have chosen the **World Happiness Report dataset**, as it provides a structured, data-driven approach to understanding the determinants of happiness across different nations.

With my background in **international development**, I am particularly interested in how societal structures, economic conditions, governance, and social support systems contribute to overall well-being. This dataset allows for a systematic examination of these influences, providing an opportunity to **move beyond subjective interpretations and derive insights rooted in measurable data**.

By analyzing patterns in happiness rankings and their contributing factors, my goal is **to identify meaningful trends, offer data-backed insights, and explore potential solutions** that could enhance quality of life at both individual and policy levels. Through this approach, I hope to contribute to the broader discussion on **effective interventions for improving happiness and well-being worldwide** while developing and honing my skills as a data analyst.

Ethical Considerations for the World Happiness Report Dataset

The World Happiness Report dataset provides valuable insights into global well-being, but its use comes with ethical responsibilities. Here are key considerations:

1. **Data Bias & Representation**

- a. The dataset relies on **self-reported** happiness scores, which can be influenced by **cultural differences** in how people perceive and express happiness.
- b. Some **regions or demographics** may be underrepresented due to **survey limitations**, affecting the accuracy of global comparisons.

2. **Subjectivity & Interpretation**

- a. Happiness is **subjective**, and the dataset's reliance on the **Cantril ladder** (self-rating from 0 to 10) may not fully capture **complex emotional and psychological factors**.
- b. The six contributing factors (economic production, social support, life expectancy, freedom, generosity, and corruption) **do not directly determine happiness** but are used to explain variations.

3. Data Privacy & Consent

- a. The dataset is based on **survey responses**, raising concerns about **how data is collected, stored, and used**.
- b. Ensuring **informed consent** and **data protection** for respondents is crucial, especially when analyzing personal well-being.

4. Potential for Misuse

- a. Policymakers and organizations may **misinterpret** or **oversimplify** findings, leading to **ineffective or biased policies**.
- b. The dataset could be used to **rank countries unfairly**, ignoring **historical, cultural, and socioeconomic contexts** that shape happiness.

5. Ethical Use in Research & Policy

- a. Researchers must **acknowledge limitations** and **avoid deterministic conclusions** about happiness.
- b. Policymakers should **use the data responsibly**, ensuring that interventions **enhance well-being** rather than impose **one-size-fits-all solutions**.

Analytical & Exploratory Questions

- **Descriptive Analysis Questions (What is happening?)**
 - What are the happiest and unhappiest countries each year?
 - How does global happiness vary year by year?
 - What is the **average happiness score globally** for each year?
 - Are there upward or downward trends in global happiness?
 - Which country improved or declined the most over time?
- **Geographic Insights Questions (Where is it happening?)**
 - How does happiness vary by region or continent?
 - Are wealthier countries happier?
 - Where on the map does the data visually show global happiness?
- **Factor-Based Questions (Why is it happening?)**
 - What factors are most strongly associated with happiness?
 - Which factors vary the most among countries?
 - Which factor shows the most inequality globally?
 - What combination of factors best predicts happiness?
- **Temporal Trends (How is it changing over time?)**
 - How do happiness factors evolve over time?
 - Are countries getting better at freedom, generosity, or life expectancy?
 - Which factors have increased/decreased over the years?
 - Did specific events influence global happiness?
- **Strategic Questions**
 - What can countries learn from others?
 - What are high-performing countries doing differently?
 - Can we cluster countries based on their happiness profiles?
 - What is the 'return on investment' for happiness?
 - How much does a unit increase in GDP or life expectancy affect happiness?

- Are some factors more “efficient” at boosting happiness?
- **Sociopolitical Questions**
 - Is happiness more influenced by internal (freedom, generosity) or external (GDP, life expectancy) factors?
 - Do democratic or less corrupt nations report higher happiness?

Data Profile

Dataset Cleaning

- Add a **Year** column to each dataset. This is necessary for filtering, trends, and merging
- Add **Region** column for grouping
- Income group based on World Bank classifications (e.g., Low, Middle, High income)
- Standardize, and remove extra characters from, column names across all years, as shown in the table below:

Current Column Names	New Standardized Names
Country, Country or region	Country
Happiness Rank, Overall rank	Happiness Rank
Happiness Score, Score	Happiness Score
Economy (GDP per Capita), Economy.GDP.per.Capita	Economy (GDP per Capita)
Family, Social support	Family support
Health (Life Expectancy), Health..Life.Expectancy. Healthy life expectancy	Health (Life Expectancy)
Freedom, Freedom to make life choices	Freedom
Generosity	Generosity
Trust (Government Corruption), Trust. Government. Corruption.Perceptions of corruption	Trust (Government Corruption)

- Match column order by reordering the columns consistently for easier merging later
- Remove unnecessary columns like "Standard Error", "Lower Confidence Interval", "Upper Confidence Interval", "Whisker.high", "Whisker.low", and "Dystopia Residual" because of their inconsistencies across the years and because of the need to achieve a consistent dataset for analysis.
- **Handling Missing Values:**
 - Replace 'N/A' with a numerical NaN (Not a Number) in the "Perceptions of corruption" column
 - Two possible strategies for handling these NaN values:
 - Imputation (e.g., mean, median, mode of the column, if appropriate).
 - Exclusion of rows with NaN values (if the number of affected rows is small and won't significantly impact the analysis).
- Clean and standardize country names based on a reference list. For example:

- **United States of America** to **United States**
 - **Russian Federation** to **Russia**
 - **Republic of Korea** to **South Korea**
- Convert data types and ensure all numerical columns are of a numeric data type (e.g., float) after cleaning, especially after handling 'N/A' values
- Remove duplicates by ensuring that after merging, there are no duplicates.
- Save the final cleaned version as:
 - **File name:** Global_Happiness_2015_2019_Cleaned.xlsx
 - **Sheet name:** Happiness Data