

World Happiness Data Analysis Project

Lee Heng Chuah, Junior Data Analyst

Executive Summary

This dataset can help us understand which countries are the happiest in the world and why. Identifying the factors that contribute to happiness can help us create a blueprint for a happier future.

The world happiness dataset is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 155 countries by their happiness levels, was released at the United Nations at an event celebrating the International Day of Happiness on March 20th.

As described in data source, this study continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. The study reviews the state of happiness today and shows how the new science of happiness explains personal and national variations in happiness. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations.

The happiness scores and rankings use data from the Gallup World Poll. The scores are 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 current lives on that scale. The scores are from nationally representative samples and use the Gallup weights to make the estimates representative. The columns following the happiness score estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to making life evaluations higher in each country than they are in Dystopia. This hypothetical country has values equal to the world's lowest national averages for each of the six factors. They do not impact the total score reported for each country, but they explain why some countries rank higher than others.

Data Sources

The data for this project is an open-source data downloaded from:

<https://www.kaggle.com/datasets/unsdsn/world-happiness>

Limitations and Ethics

1. Data set available are from 2015 to 2019, which makes it difficult to accurately predict the current happiness trends year especially during the Covid19 and after Covid19.
2. The scores are from nationally representative samples. It could be resulting in sample bias. If there is a lack of diversity within that sample, it could be biased.

3. The scores are based on answers to the main life evaluation question asked in the poll. It could be also the measurement bias occurred when there is a problem with the humans doing the measuring.

Data Cleaning and Consistency Check

- Investigate the data
- Rename the columns based in year, and standardise the columns name in small letter
- Check for mixed data type
- Check for missing data
- Check for duplicate data

Data Profile

Number of rows and columns in the cleaned dataset is the same as original dataset:

- 2015_clean_data.csv (158, 12)
- 2016_clean_data.csv (157, 13)
- 2017_clean_data.csv (155, 12)
- 2018_clean_data.csv (156, 9)
- 2019_clean_data.csv (156, 9)

Notes:

1. There are 195 countries in the world today, but the dataset only has 155-158 countries. The dataset doesn't represent the full picture of happiness score.
2. There are different numbers of columns in each dataset. Different factors had been previously taken into consideration for each year but there are also same factors for each year had been considered e.g., happiness rank, happiness score, healthy life expectancy, freedom, generosity, and trust. These datasets had been kept for further investigation.
3. The merge / combination of dataset will be happened at the next stage for full data analysis.

Column

Column	Column Description	Data Type	Time Variant
Country	This column represents the country.	Qualitative, Nominal	no
Region	This column represents the region.	Qualitative, Nominal	no
happiness_rank_year	This column represents the rank of the happiness.	Quantitative, Continuous	no

happiness_score_year	This column represents the score of the happiness.	Quantitative, Continuous	no
standard_error_year	This column represents the standard error of the happiness.	Quantitative, Continuous	no
lower_confidence_year	This column represents the lower confidence of the happiness.	Quantitative, Continuous	no
upper_confidence_year	This column represents the upper confidence of the happiness.	Quantitative, Continuous	no
whisker_high_year	This column represents the whisker high of the happiness.	Quantitative, Continuous	no
whisker_low_year	This column represents the whisker low of the happiness.	Quantitative, Continuous	no
economy_gdp_year	This column represents the happiness score based on economy (GDP per capita).	Quantitative, Continuous	no
family_year	This column represents the happiness score based on family score.	Quantitative, Continuous	no
health_life_expectancy_year	This column represents the happiness score based	Quantitative, Continuous	no

	on health life expectancy.		
freedom_year	This column represents the happiness score based on freedom.	Quantitative, Continuous	no
trust_year	This column represents the happiness score based on trust (government corruption)	Quantitative, Continuous	no
generosity_year	This column represents the happiness score based on generosity.	Quantitative, Continuous	no
dystopia_residual_year	This column represents the happiness score based on dystopia residual.	Quantitative, Continuous	no
social_support_year	This column represents the happiness score based on social support.	Quantitative, Continuous	no

Note: year in column represents the year for each dataset.

Questions To Explore

1. What countries or regions rank the highest in overall happiness?
2. Did any country experience a significant increase or decrease in happiness for the past 5 years (from 2015 to 2019)?
3. What factors that contribute to happiness? What can we learn from these countries?