

6.1 World Happiness Project (2015-2022)

Rajit Sikka

Data Source

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

This dataset contains the World Happiness Report from the years 2015 – 2022. The dataset contains a happiness score for approximately 150 countries alongside additional factors (such as GDP per Capita, Healthy Life Expectancy, Social Support, etc.) to explain the score. The report is sourced externally from the Gallup World Poll Survey. The dataset was provided by a user on Kaggle.

Why I chose this set

I chose the world happiness report because I found the topic interesting, as I want to see what factors influence “happiness” for a given country. Furthermore, I specifically chose this dataset as it has information up to 2022 so I can factor in how COVID impacted each country.

Variables

Column Name	Description	Variable Type
Year	Year that Happiness report was taken	Categorical
Country	Name of Country	Categorical
Happiness Score	Average “Life as a whole” rating	Quantitative
GDP per Capita	Economic performance of a country.	Quantitative
Social Support	Extent to which people feel they can rely on others.	Quantitative
Healthy Life Expectancy	Quantity and Quality of life score for healthy life.	Quantitative
Freedom to make life choices	Satisfaction level with choice of freedom	Quantitative
Generosity	National Average of amount donated to charity	Quantitative
Perceptions of Corruption	The extent in which people see corruption in government and businesses.	Quantitative

Full Definitions for variables here: <https://happiness-report.s3.amazonaws.com/2024/Ch2+Appendix.pdf>

Data Profile:

Full list of changes made to dataset (s) (Documentation is also available in Python Workbook):

Data Cleaning

1. Removing Unnecessary Columns and Creating Column Consistency
 - a. I only kept the variables listed above. Each Yearly Report had some different variables so I just removed them.
 - b. Renamed columns from each year the same for consistency
 - c. I re-ordered the columns for each workbook so they would be in the same order.
(Year, Country, Happiness Score, GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to make life choices, Generosity, Perceptions of Corruption)
2. Data Quality Checks
 - a. Mixed Type Data: None found
 - b. Duplicate Data: None found
 - c. Two instances of missing data: One in the 18 WHR (World Happiness Report), and one in the 22 WHR. For the 18 WHR it was a single value under United Arab Emirates for Perceptions of Corruption, so I just took the median of that column for the year and imputed it. As for the 22 WHR it was an unnecessary row of missing data (with only xx for country), so I just deleted the entire row.

Data Wrangling

1. Adding Year Columns for each dataset (to help for the merged dataset)
2. Making a merged dataset
 - a. I merged all World Happiness Reports (2015 – 2022) into a single workbook

Data Limitations:

This dataset is only up till 2022, so it may not be completely up to date with current country happiness. Additionally, this dataset doesn't contain all the countries in the world for each year, so there are slight inconsistencies.

Ethics:

Since this dataset was created through survey data, there are some inherent biases. As per [Gallup](#), surveys are conducted through either telephone or face to face. While Gallup admittedly does seem to try their best against bias, there can be potential differences when conducting telephone interviews vs. in-person (in person may seem more pressured/willing to give an easy answer depending on circumstance). Also the sample collected may not be fully representative of the population in situations when they can't reach rural areas/conflict areas to conduct a survey. Finally there can also be data collection issues, depending on who is actively conducting the surveys.

Key Questions:

- What countries are the happiest? Least Happy?
- What are the rankings for each of the other six factors (GDP per capita, Social support, etc.)
 - How have they changed over time for each country? And relative for each year?
- What countries have improved their happiness score over time, worsened their happiness scores.
- What factors have the greatest impact on a change in happiness?
- Are certain continents/regions happy than others?
 - Additionally do factors like social support/ perceptions of corruption also are consistent between certain regions?