WORLD HAPPINESS 2015-2022

Open-source Data Project

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document. When you're ready to add your content, just click here and start typing.]

Iryna Smologonova

September 2022

Executive Summary

The World Happiness Report is a landmark survey that measures and ranks the happiness of individuals in 158 countries around the globe. Happiness is increasingly considered an important and useful way to guide public policy and measure its effectiveness.

The World Happiness Report 2015 was launched by the Sustainable Development Solutions Network (SDSN). The World Happiness Report 2017 was released at the United Nations at an event celebrating International Day of Happiness on March 20th.

We will explore the geographical and temporal trends in the happiest countries/regions, impact of Covid-19 outbreak on the happiness score to have a better understanding and to make a prediction for the following year.

Data Sourcing

External data source provided by SDSN, a global initiative for the United Nations. The SDSN Association is an independent non-profit organization in the United States and a non-profit Association 1901 in France. The SDSN Association serves as the institutional home of the SDSN and ensures that the work of the SDSN complies with all US and international laws. Since SDSN complies with the US and international laws, this source of data is considered to be trustworthy.

Data is available here: https://www.kaggle.com/datasets/mathurinache/world-happiness-report

Merging the datasets based on country results in a prepared dataset which covers the key project requirements and allows to conduct spatial and time series analysis.

Data Collecting

The World Happiness Report uses data from the Gallup World Poll survey based on answers to the main life evaluation question asked in the poll. The typical annual sample for each country is 1,000 people. However, many countries have not had annual surveys. If a typical country had surveys each year, the sample size would be 3,000. It is used responses from the three most recent years to provide an up-to-date and robust estimate of life evaluations.

Data contents

The data is categorized by country for 2015-2022 and contains six variables:

1. GDP per capita, i.e. purchasing power
2. Social support (having someone to count on in troubled times)
3. Healthy life expectancy
4. Freedom to make life decisions,
5. Generosity (donations to charity)
6. Perceptions of corruption

The happiness ranking of countries was determined by asking individuals to evaluate their lives as a whole, according to the variables. The Happiness Score is a national average of the responses to the main life evaluation question asked in the Gallup World Poll (GWP). Respondents scored them on a scale of 0 to 10.

Dystopia is an imaginary country that has the world's least-happy people. The purpose in establishing Dystopia is to have a benchmark against which all countries can be favorably compared in terms of each of the six key variables, thus allowing each sub-bar to be of positive width. The lowest scores observed for the six key variables, therefore, characterize Dystopia. Since life would be very unpleasant in a country with the world's lowest incomes, lowest life expectancy, lowest generosity, most corruption, least freedom and least social support, it is referred to as "Dystopia". The Dystopia Residual metric actually is the Dystopia Happiness Score(1.85) + the Residual value or the unexplained value for each country as stated in the previous answer.

Data limitations and Ethics

The data is collected via survey based on random sample of respondents. Therefore, it might lead to sample errors.

It was the damaging effects of social media use on the happiness and self-image of adolescents, mainly based on data from the United States in the World Happiness Report 2019. This runs parallel to evidence from earlier Reports showing that in-person friendships support happiness, while online connections do not. But COVID-19 and its limitations on in-person meetings offered a chance for electronic connections to develop their potential for creating and maintaining the social bonds that support happiness.

There is no PII in the dataset and therefore it is considered as low-risk in terms of ethics. The measurement was conducted based on a country level; thus, group of people and small communities can't be identified.

Data relevance

The data is relevant because we can identify happiness rank by country and by time period 2015-2019, measurements of six key variables by countries and we can conduct spatial and time series analysis to make a prediction of happiness level.

Data cleaning steps

The wrangling and cleaning steps have been conducted in the script "Task 6.1_1 wrangling procedure":

- Dropping the columns that are irrelevant for the analysis
- Renaming and shuffling the columns to have a consistent naming and sequence convention
- Adding the column "Year" for each file before merging and changing datatype for the column "

Data merging and consistency check

The merging and consistency check steps have been conducted in the script "Task 6.1_2 merging & consistency check":

- Merging seven files into one
- Renaming and checking country names to have a consistent naming convention
- Mapping regions to countries based on created dictionary
- Checking the missing values, imputed one missing value based on the mean (United Arabian Emirates, 2018, the value 'trust_government_corruption' – 0.277306)
- No duplicates have found
- No mixed data have found

Data profile

The prepared dataset has 1229 entries, 0 to 1228, and 13 columns.

Column name	Description	Datatype	Time Variant
country	Country of measurement	Qualitative, nominal	no
region	Region of measurement	Qualitative, nominal	no
Year	Year of measurement	Quantitative, discrete	yes
happiness_rank	The country rank based on a happiness score	Quantitative, discrete	no
happiness_score	The happiness score is a national average of the responses to the main life evaluation question	Quantitative, continuous	yes
economy_GDP_per_capita	The score of GDP per capita, i.e. purchasing power	Quantitative, continuous	yes
family	The score of social support (having someone to count on in troubled times)	Quantitative, continuous	yes
health_life_expectancy	The score of healthy life expectancy	Quantitative, continuous	yes
freedom	The score of freedom to make life choices	Quantitative, continuous	yes
trust_government_corruption	The score of corruption perception	Quantitative, continuous	yes
generosity	The score of generosity (donations to charity)	Quantitative, continuous	yes
dystopia_residual	The metric of happiness (1.85) + the residual value of the imaginary country Dystopia	Quantitative, continuous	yes

Questions to explore

- Which region/top 10 countries are the happiest?
- What geographical and temporal trends could be observed?
- Could we predict the following year the happiest countries/region?
- How was the score affected by COVID-19 outbreak?