

6.1 Sourcing Open Data

World Happiness Report from 2015-2022

Executive Summary

This dataset represents the World Happiness data from the Gallup World Poll surveys from 2019 to 2022. They are based on answers to the main life evaluation question asked in the Cantril ladder (it asks respondents to think of a ladder, with the best possible life for them being a ten and the worst possible life being a zero). These survey questions show the estimated extent to which each of the six factors: levels of GDP, life expectancy, generosity, social support, freedom, and corruption, affects people in different countries.

1. Data:

Data Source:

This data is publicly available open-source data. It was downloaded from Kaggle.com.

Link: [World Happiness Report up to 2022](#)

Data Contents:

The dataset includes happiness values regarding GDP, life expectancy, generosity, social support, freedom, and corruption for 163 countries.

Data Profile:

- The dataset was divided into eight individual CSV files, which were merged. Before merging, a column called 'Year' was created in each dataset, representing the year the data was recorded. In addition, variables that were not in common between datasets were dropped to allow the merging process. After merging, the data set presented ten columns and 1231 rows. After the Data Quality check and the spatial variable 'Region' generation, the final dataset contains 11 columns and 1230 rows.
- Data consistency and modifications, such as change of column position and removing duplicates, were done using Python.

Summary of Data Cleaning and Data Consistency Checks

- Consolidation of 8 CSV files into 1
- Mismatched columns were deleted before the merge
- Variables Year and Region were created
- Check for duplicate data
- Check for mixed data type

Column	Description	Data Type	Time Variant
Country	Spatial characteristic	Nominal, Ordinal	No
Region	Spatial characteristic	Nominal, Ordinal	No
Year	Year data was taken	Quantitative, Discrete	Yes
Happiness_Rank	Ranking according to Happiness Score	Quantitative, Discrete	Yes
Happiness_Score	From zero to ten, 10 represents the best possible life, and zero is the worst.	Quantitative, Discrete	Yes
GDP_per_Capita	The influence of the economic part on Happiness	Quantitative, Continuous	Yes
Social_support	The influence of social support part in Happiness	Quantitative, Continuous	Yes
Life_Expectancy	The influence of health on Happiness	Quantitative, Continuous	Yes
Freedom	The influence of freedom on Happiness	Quantitative, Continuous	Yes
Generosity	The influence of generosity on Happiness	Quantitative, Continuous	Yes
Corruption	The influence of corruption on Happiness	Quantitative, Continuous	Yes

2. Limitations and Ethics:

- Small amount of available data (1230 rows).
- Region column was missing in many datasets; therefore, it was manually created.
- Several columns were dropped due to a mismatch.
- In terms of ethics, the data does not contain personal information.

3. Questions to explore:

- Is there a change in the happiness score over the years?
- Did COVID influence since 2019 have an effect on the happiness score as well as in the other variables?
- Are similar happiness ranks located in the same regions?
- What are the main factors influencing the happiness score over the years?
- Do the factors influencing the happiness score to vary spatially?
- Is there a primary factor influencing the happiness score?