

Data Sourcing

Dataset(s): [World Happiness Report 2015-2019 hosted on Kaggle](#)

Data Source: The data source for the World Happiness Report datasets is external, originating from the Sustainable Development Solutions Network (SDSN) a non-profit organization funded by the United Nations. It was first released at the United Nations and uses rankings and data from the Gallup World Poll. The data is provided on Kaggle, a platform for sharing and discovering datasets. As the creator of the datasets is a reputable organization backed by the United Nations, the data can be considered trustworthy.

Data Collection: The World Happiness Report datasets are based on data from the Gallup World Poll, a global survey conducted by the Gallup organization. The survey involves asking individuals to rate their happiness on a scale known as the Cantril ladder, where 0 represents the worst possible life and 10 represents the best possible life. The datasets include variables such as happiness scores, GDP per capita, family, health-life expectancy, freedom, trust-government corruption, and generosity. The data collection method involves administering surveys to individuals across various countries.

Contents: The datasets consist of five separate files corresponding to the World Happiness Reports from 2015 to 2019. Each file contains similar variables related to happiness scores, country information, and factors influencing happiness. These variables include country, region (present in some years), happiness rank, happiness score, GDP per capita, family, health-life expectancy, freedom, trust-government corruption, generosity, as well as additional variables such as standard error, confidence intervals, and whiskers.

Data Relevance: The World Happiness Report datasets are relevant for various reasons. They provide valuable insights into factors influencing happiness within countries, allowing for comparisons between countries and regions to identify trends and patterns in happiness levels over time.

Data Limitations: There are several limitations to consider when using the World Happiness Report datasets. Firstly, happiness is a subjective measure influenced by cultural, societal, and individual differences, which may not be fully captured in the data. Data accuracy may be affected by the survey based nature of the data collection process. Bias could potentially be introduced, as not all countries may have

comprehensive data available for all factors included in the report. Lastly, the time frame of the data is limited, as the most recent report is a five years old.

Data Choice: I'm interested in using the World Happiness Report datasets because I believe it is an interesting topic to analyse, especially considering the last report was 2019, before the world was affected by Covid-19 in a way we never expected. These datasets raise awareness about the importance of happiness and well-being as essential components of human development. Analysing these datasets provides an opportunity to gain a deeper understanding of the complex interplay between various socio-economic factors and subjective well-being on a global scale.

Data Profile *More can be found in the Jupyter notebook 6.1 *

2015 Descriptive analysis

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity
count	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000
mean	79.493671	5.375734	0.846137	0.991046	0.630259	0.428615	0.143422	0.237296
std	45.754363	1.145010	0.403121	0.272369	0.247078	0.150693	0.120034	0.126685
min	1.000000	2.839000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	40.250000	4.526000	0.545808	0.856823	0.439185	0.328330	0.061675	0.150553
50%	79.500000	5.232500	0.910245	1.029510	0.696705	0.435515	0.107220	0.216130
75%	118.750000	6.243750	1.158448	1.214405	0.811013	0.549092	0.180255	0.309883
max	158.000000	7.587000	1.690420	1.402230	1.025250	0.669730	0.551910	0.795880

2016 Descriptive analysis

	Country	Region	Happiness Rank	Happiness Score	Lower Confidence Interval	Upper Confidence Interval	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
0	Denmark	Western Europe	1	7.526	7.460	7.592	1.44178	1.16374	0.79504	0.57941	0.44453	0.36171	2.73939
1	Switzerland	Western Europe	2	7.509	7.428	7.590	1.52733	1.14524	0.86303	0.58557	0.41203	0.28083	2.69463
2	Iceland	Western Europe	3	7.501	7.333	7.669	1.42666	1.18326	0.86733	0.56624	0.14975	0.47678	2.83137
3	Norway	Western Europe	4	7.498	7.421	7.575	1.57744	1.12690	0.79579	0.59609	0.35776	0.37895	2.66465
4	Finland	Western Europe	5	7.413	7.351	7.475	1.40598	1.13464	0.81091	0.57104	0.41004	0.25492	2.82596

2017 Descriptive analysis

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Generosity	Trust (Government Corruption)
count	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000
mean	78.000000	5.354019	0.984718	1.188898	0.551341	0.408786	0.246883	0.123120
std	44.888751	1.131230	0.420793	0.287263	0.237073	0.149997	0.134780	0.101661
min	1.000000	2.693000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	39.500000	4.505500	0.663371	1.042635	0.369866	0.303677	0.154106	0.057271
50%	78.000000	5.279000	1.064578	1.253918	0.606042	0.437454	0.231538	0.089848
75%	116.500000	6.101500	1.318027	1.414316	0.723008	0.516561	0.323762	0.153296
max	155.000000	7.537000	1.870766	1.610574	0.949492	0.658249	0.838075	0.464308

2018 Descriptive analysis

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Generosity	Trust (Government Corruption)
count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	155.000000
mean	78.500000	5.375917	0.891449	1.213237	0.597346	0.454506	0.181006	0.112000
std	45.177428	1.119506	0.391921	0.302372	0.247579	0.162424	0.098471	0.096492
min	1.000000	2.905000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	39.750000	4.453750	0.616250	1.066750	0.422250	0.356000	0.109500	0.051000
50%	78.500000	5.378000	0.949500	1.255000	0.644000	0.487000	0.174000	0.082000
75%	117.250000	6.168500	1.197750	1.463000	0.777250	0.578500	0.239000	0.137000
max	156.000000	7.632000	2.096000	1.644000	1.030000	0.724000	0.598000	0.457000

2019 Descriptive analysis

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Generosity	Trust (Government Corruption)
count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000
mean	78.500000	5.407096	0.905147	1.208814	0.725244	0.392571	0.184846	0.110603
std	45.177428	1.113120	0.398389	0.299191	0.242124	0.143289	0.095254	0.094538
min	1.000000	2.853000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	39.750000	4.544500	0.602750	1.055750	0.547750	0.308000	0.108750	0.047000
50%	78.500000	5.379500	0.960000	1.271500	0.789000	0.417000	0.177500	0.085500
75%	117.250000	6.184500	1.232500	1.452500	0.881750	0.507250	0.248250	0.141250
max	156.000000	7.769000	1.684000	1.624000	1.141000	0.631000	0.566000	0.453000

Limitations:

Data Completeness: The World Happiness Report datasets may contain missing variables for certain countries or years. This could be due to variations in data collection methods across regions or insufficient tracking of certain metrics by the Gallup World Poll. Despite this limitation, the datasets offer a comprehensive overview of happiness scores and associated factors across multiple countries and years.

Correlation vs. Causation: Given the generalised nature of the data, it's essential to distinguish between correlation and causation when analyzing relationships between variables. For instance, strong correlations observed between factors like GDP per capita and happiness scores should not be interpreted as causal relationships, as various confounding variables may influence the outcomes.

Collection Bias: The World Happiness Report datasets span multiple years and originate from diverse sources, potentially leading to variations in data collection methods over time and across regions. While efforts have been made to standardize measurements, there remains a possibility of collection bias affecting the comparability of data across different contexts. The data collected for the World Happiness Report comes from government entities, which may introduce bias in certain regions and time periods due to political agendas. This bias could influence the reporting of happiness scores and associated factors, impacting the accuracy and reliability of the data. Some

of the data in the World Happiness Report may be derived from surveys, raising concerns about self-selection bias among respondents or potential pressure to respond in certain ways due to political or societal factors. This could affect the validity of the happiness scores and other subjective measures included in the datasets.

Ethical Considerations:

Privacy & Personally Identifiable Information (PII): The World Happiness Report datasets do not contain personally identifiable information, alleviating concerns about privacy and individual privacy rights. However, we as data analysts should still handle the data responsibly and ensure that any analysis or findings are presented in a manner that respects privacy concerns.

Cultural Sensitivity: Given that the data includes variables related to religion and other culturally influenced behaviors, it's crucial to approach the analysis with cultural sensitivity. Any interpretations or insights should be objective and avoid making assumptions about cultural motivations behind the observed trends or patterns.

Data Usage - Discrimination: Insights derived from the World Happiness Report datasets should not be used to generalise or discriminate against specific populations or cultures negatively. It's essential to use the data responsibly and refrain from making sweeping generalisations that could perpetuate stereotypes or biases.

Data Usage - Health: Any insights related to health outcomes should not be construed as medical advice and should be further scrutinised if used in authoritative capacities. While the data may provide valuable insights into the relationship between happiness and health, caution should be exercised to avoid misinterpretation or misapplication of the findings.

Questions to explore

How does the distribution of happiness scores vary among countries within each region?

Are there any countries that are consistently ranking high or low in happiness over the multiple years?

What is the relationship between GDP growth rate and changes in happiness scores over time?

How do countries with similar GDPs differ in terms of happiness scores?

Is there a correlation between environmental factors (such as air quality or access to green spaces) and happiness scores?

How does the level of income inequality within a country affect its overall happiness score?

Does access to healthcare services and infrastructure impact a country's happiness score?

How do subjective well-being measures, such as self-reported life satisfaction, align with objective indicators of happiness?

Is there a relationship between educational attainment levels and happiness scores across different nations?