Money Can't Buy Happiness, or Can It? Understanding the Relationship Between GDP and Global Well-Being

# Data Sourcing

## Data Rational

## I chose these datasets with the goal of cross-examining the perception of happiness and reported worldwide GDP by country. The objective of this analysis is to investigate past World Happiness reports in relation to the tangible measures derived from reported GDP from the same time period, with the aim seeing if any meaningful insights can be drawn.

## Additional Information on data sourcing can be found in the following documents:

* + [World Happiness Ratings](../Data/Original%20Data/Data%20Source/Data%20Info,%20World%20Happiness%20Report.docx)
  + [GDP by Country from 1999-2022](../Data/Original%20Data/Data%20Source/data%20Info,GDP%20by%20Country%201999-2022.docx)

# Data Profiling

## Original Data Profiling Report

All initial descriptive statistics on original data sets can be located in Python Script named “1. Data Exploration GDP and WHR 2015 to 2019 CSV Files” in the scripts sub-folder.

|  |  |  |
| --- | --- | --- |
| **File - GDP by Country 1999-2022.CSV** | | |
| **Rows- 180 x Columns-24** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1999  …2022 | All object | 180 countries are listed by row under each column depicting their overall reported GDP for the year (in US$ billions) |

|  |  |  |
| --- | --- | --- |
| **File – 2015.CSV** | | |
| **Rows- 158 x Columns- 12** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1.        Country | object | Nominal data denoting the name or identity of a country. |
| 2.        Region | object | Nominal data indicating the geographic or political region to which a country belongs. |
| 3.        Happiness Rank | int64 | Ordinal data indicating the relative position of countries in terms of happiness. |
| 4.        Happiness Score | object | Continuous data representing the overall happiness level of countries from a scale of 1-10 |
| 5.        Standard Error | float64 | Quantitative data indicating the variability in happiness score measurements. |
| 6.        Economy (GDP per Capita) | float64 | Quantitative data measuring a country's economic performance. Measure is gathered from participants from 1-10 and pre-calculated to depict this measures overall impact on the country’s happiness rating. |
| 7.        Family | float64 | Continuous data reflecting the perceived strength of social support and family bonds. Measure is gathered from participants from 1-10 and pre-calculated to depict this measures overall impact on the country’s happiness rating. |
| 8.        Health (Life Expectancy) | float64 | Continuous data representing the perceived life expectancy of a country’s citizens. Measure is gathered from participants from 1-10 and pre-calculated to depict this measures overall impact on the country’s happiness rating. |
| 9.        Freedom | float64 | Continuous data reflecting the perceived freedom to make life choices. Measure is gathered from participants from 1-10 and pre-calculated to depict this measures overall impact on the country’s happiness rating. |
| 10.     Trust (Government Corruption) | float64 | Continuous data indicating the perceived level of trust in government institutions. Measure is gathered from participants from 1-10 and pre-calculated to depict this measures overall impact on the country’s happiness rating. |
| 11.     Generosity | float64 | Continuous data measuring the perceived generosity of a country's citizens. |
| 12.     Dystopia Residual | float64 | Continuous data representing an imaginary country used as a benchmark for low happiness. |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **File – 2016.CSV** | | |
| **Rows- 157 x Columns- 13** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1. Country | object | Nominal data denoting the name or identity of a country. |
| 1. Region | object | Nominal data indicating the geographic or political region to which a country belongs. |
| 1. Happiness Rank | int64 | Continuous data representing the overall happiness level of countries. |
| 1. Happiness Score | float64 | Quantitative data indicating the variability in happiness score measurements. |
| 1. Lower Confidence Interval | float64 | Quantitative data representing the lower bound of a confidence interval for a statistical estimate. |
| 1. Upper Confidence Interval | float64 | Quantitative data representing the upper bound of a confidence interval for a statistical estimate. |
| 1. Economy (GDP per Capita) | float64 | Quantitative data measuring a country's economic performance. |
| 1. Family | float64 | Continuous data reflecting the perceived strength of social support and family bonds. |
| 1. Health (Life Expectancy) | float64 | Continuous data representing the perceived life expectancy of a country’s citizens. |
| 1. Freedom | float64 | Continuous data reflecting the perceived freedom to make life choices. |
| 1. Trust (Government Corruption) | float64 | Continuous data indicating the perceived level of trust in government institutions. |
| 1. Generosity | float64 | Continuous data measuring the perceived generosity of a country's citizens. |
| 1. Dystopia Residual | float64 | Continuous data representing an imaginary country used as a benchmark for low happiness. |

|  |  |  |
| --- | --- | --- |
| **File – 2017.CSV** | | |
| **Rows- 155 x Columns- 12** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1. Country | object | Nominal data denoting the name or identity of a country. |
| 1. Happiness.Rank | Int64 | Continuous data representing the overall happiness level of countries. |
| 1. Happiness.Score | float64 | Quantitative data indicating the variability in happiness score measurements. |
| 1. Whisker.high | float64 | Quantitative data representing the upper bound of a confidence interval for a statistical estimate. |
| 1. Whisker.low | float64 | Quantitative data representing the lower bound of a confidence interval for a statistical estimate. |
| 1. Economy..GDP.per.Capita. | float64 | Quantitative data measuring a country's economic performance. |
| 1. Family | float64 | Continuous data reflecting the perceived strength of social support and family bonds. |
| 1. Health..Life.Expectancy. | float64 | Continuous data representing the perceived life expectancy of a country’s citizens. |
| 1. Freedom | float64 | Continuous data reflecting the perceived freedom to make life choices. |
| 1. Generosity | float64 | Continuous data measuring the perceived generosity of a country's citizens. |
| 1. Trust..Government.Corruption. | float64 | Continuous data indicating the perceived level of trust in government institutions. |
| 1. Dystopia.Residual | float64 | Continuous data representing an imaginary country used as a benchmark for low happiness. |

|  |  |  |
| --- | --- | --- |
| **File – 2018.CSV** | | |
| **Rows- 156 x Columns- 9** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1. Overall Rank | int64 | Ordinal data indicating the relative position of countries in terms of happiness |
| 1. Country or Region | object | Nominal data denoting the name or identity of a country. |
| 1. Score | object | Continuous data representing the overall happiness level of countries. |
| 1. GDP per Capita | float64 | Quantitative data measuring a country's economic performance. |
| 1. Social Support | float64 | Continuous data reflecting the perceived strength of social support and family bonds. |
| 1. Healthy life expectancy | float64 | Continuous data representing the perceived life expectancy of a country’s citizens. |
| 1. Freedom to make life choices | float64 | Continuous data reflecting the perceived freedom to make life choices. |
| 1. Generosity | float64 | Continuous data measuring the perceived generosity of a country's citizens. |
| 1. Perceptions of corruption | float64 | Continuous data indicating the perceived level of trust in government institutions. |

|  |  |  |
| --- | --- | --- |
| **File – 2019.CSV** | | |
| **Rows- 156 x Columns- 9** | | |
| **Column Name** | **Data Type** | **Data Structure** |
| 1. Overall Rank | int64 | Ordinal data indicating the relative position of countries in terms of happiness |
| 1. Country or Region | object | Nominal data denoting the name or identity of a country. |
| 1. Score | object | Continuous data representing the overall happiness level of countries. |
| 1. GDP per Capita | float64 | Quantitative data measuring a country's economic performance. |
| 1. Social Support | float64 | Continuous data reflecting the perceived strength of social support and family bonds. |
| 1. Healthy life expectancy | float64 | Continuous data representing the perceived life expectancy of a country’s citizens. |
| 1. Freedom to make life choices | float64 | Continuous data reflecting the perceived freedom to make life choices. |
| 1. Generosity | float64 | Continuous data measuring the perceived generosity of a country's citizens. |
| 1. Perceptions of corruption | float64 | Continuous data indicating the perceived level of trust in government institutions. |

## Limitations and Ethical Considerations

• The insights from this analysis should not be used to draw conclusive evidence, as complex factors beyond the scope of this analysis may impact a country's GDP and happiness perception.

• Cultural sensitivity should be at the forefront of the analysis and interpretation of insights. Different cultures have varying perspectives on happiness, and GDP may not capture all aspects of well-being. It's crucial to approach the comparison with cultural sensitivity, recognizing diverse interpretations of happiness.

• GDP is just one indicator of a country's well-being and development. Balancing economic indicators with social and environmental considerations is essential for a comprehensive understanding.

• Limitations in reporting for both reports may impact available data for some years and influence a country's overall report scores.

• Due to the available timeframe of this analysis, only a country's overall total GDP can be analyzed. Future considerations may include the addition of other economic factors when drawing insights across this data.

# Data Cleaning

## Wrangling/Cleaning Steps

For consistency each of the 5 World Happiness Reports required renaming of columns, dropping of variables that were not consistent across all tables, and aggregating of variables to make all tables consistent for concatenation into one cohesive dataset. Sets are outlined below taken to each dataset. Files 2018 and 2019 were used for primary modeling due to their limited columns available for analysis.

The following column names were used as standards across all tables: county, region, happiness\_rank, happiness\_score, social\_support, logged\_GDP\_per\_capita, ealthy\_life\_expectancy, freedom\_to\_make\_life\_choices, generosity, perceptions\_of\_corruption, Dystopia\_residual(if applicable), upper\_confidence\_interval, lower\_confidence\_interval, & standard\_error (for rows that this applies to).

Scripts containing wrangling and concating steps located in Python scripts folder labeled:

2. Wrangling and Concating 2015 to 2019 CSV files

3. 3. Wranging and concating GDP with WHR CSV

|  |  |  |  |
| --- | --- | --- | --- |
| **FILE NAME** | **Columns Renamed** | **Columns Dropped** | **Column Data Type Changed** |
| 2015.CSV | Country  Region  Standard Error  Happiness Rank  Happiness Score  Economy (GDP per Capita)  Family  Health (Life Expectancy)  Freedom  Trust (Government Corruption)  Generosity  Dystopia Residual  Standard Error | None | No mixed types found |
| 2016.CSV | Country  Region  Lower Confidence Interval  Upper Confidence Interval  Happiness Rank  Happiness Score  Economy (GDP per Capita)  Family  Health (Life Expectancy)  Freedom  Trust (Government Corruption)  Generosity  Dystopia Residual | None. | No mixed types found |
| 2017.CSV | Country  Happiness.Rank  Happiness.Score  Whisker.high  Whisker.low  Economy..GDP.per.Capita.  Family  Health..Life.Expectancy.  Freedom  Generosity  Trust..Government.Corruption.  Dystopia.Residual |  | No mixed types found |
| 2018.CSV | Overall Rank  Country or Region  Score  GDP per Capita  Social Support  Healthy life expectancy  Freedom to make life choices  Generosity  Perceptions of corruption | None | No mixed types found |
| 2019.CSV | Overall Rank  Country or Region  Score  GDP per Capita  Social Support  Healthy life expectancy  Freedom to make life choices  Generosity  Perceptions of corruption | None | No mixed types found |
| GDP by Country 1999-2022.CSV  \*Renamed GDP\_2015\_2019 | CSV was initially manipulated in Excel to aggregate year and country both into columns for easier manipulation in Python. The data set was limited to 2015 to 2019 to better align with concating files in later steps.  Countries required matching names between datasets prior to joining tables and the process for finding different county names and the countries that were renamed can be found in Python script “3”. | None | No mixed types found |

## Consistency Checks

|  |  |  |  |
| --- | --- | --- | --- |
| **FILE NAME** | **Duplicates** | **Missing Values** | **Notes** |
| 2015.CSV | 0 | 0 | No concerns found upon initial analysis. |
| 2016.CSV | 0 | 0 | No concerns found upon initial analysis. |
| 2017.CSV | 0 | 0 | No concerns found upon initial analysis. |
| 2018.CSV | 0 | 0 | No concerns found upon initial analysis. |
| 2019.CSV | 0 | 0 | No concerns found upon initial analysis. |
| GDP by Country 1999-2022.CSV | 0 | 69 rows marked 0 | No NaN values, but 69 cells were marked 0. This is consistent with original data from the world bank as some countries did not report some years, were dissolved, or other factors impacting reporting. Values will be left 0 for this analysis.  Countries with 0 listed:  Afghanistan 5  Bahamas, The 5  Cape Verde 5  Congo (Kinshasa) 5  Congo (Brazzaville) 5  Egypt 5  Gambia, The 5  Iran 5  Laos 5  Macedonia 5  Syria 5  Timor-Leste, Dem. Rep. of 5  Yemen 5  Pakistan |
| DF\_WHR.CSV | 0 | 0 | No concerns found upon initial analysis. |
| WHR\_GDP\_2015to2019.csv |  | 69 rows marked 0  AND  28 rows with NAN created during merge with WHR and GDP tables | The same rows identified above remained in the analysis.  28 rows were lost in the inner merge of tables due to being unable to match the following countries with a WHR equivalent  Kosovo 5  Montenegro 5  Palestinian Territories 5  Iraq 5  Somalia 4  Somaliland region 1  Somaliland Region 1  North Macedonia 1  Gambia |

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **New column** | **Column/s it was Derived From** | **Conditions** |
| DF\_WHR | region | Derived from conditions in the 2015,2016, and 2017 region column by aggregating countries into different regions of the world. This condition was applied over the existing region column for consistency across the dataset. | DF\_all.loc[DF\_all['Country'].isin(['Mauritius', 'Nigeria', 'Zambia', 'Somaliland region', 'Mozambique', 'Lesotho', 'Swaziland', 'South Africa', 'Ghana', 'Zimbabwe', 'Liberia', 'Sudan', 'Congo (Kinshasa)', 'Ethiopia', 'Sierra Leone', 'Mauritania', 'Kenya', 'Djibouti', Botswana', 'Malawi', 'Cameroon', 'Angola', 'Mali', 'Congo (Brazzaville)', 'Comoros', 'Uganda', Senegal', 'Gabon', 'Niger', 'Tanzania', 'Madagascar', 'Central African Republic', 'Chad', Guinea', 'Ivory Coast', 'Burkina Faso' 'Rwanda', 'Benin', 'Burundi', 'Togo']),  'region'] = 'Sub-Saharan Africa'  DF\_all.loc[DF\_all['Country'].isin(['Czech Republic', 'Uzbekistan', 'Slovakia', 'Moldova', 'Kazakhstan', 'Slovenia', 'Lithuania', 'Belarus', 'Poland', 'Croatia', 'Russia', 'Kosovo', 'Turkmenistan', 'Estonia', Kyrgyzstan', 'Azerbaijan', 'Montenegro', 'Romania', 'Serbia', 'Latvia', 'Macedonia', 'Albania', 'Bosnia and Herzegovina', 'Hungary', 'Tajikistan', 'Ukraine', 'Armenia', 'Georgia', 'Bulgaria']),  'region'] = 'Central and Eastern Europe'  DF\_all.loc[DF\_all['Country'].isin(['Costa Rica', 'Mexico', 'Brazil', 'Venezuela', 'Panama', 'Chile', 'Argentina', 'Uruguay', ‘'Colombia', 'Suriname', 'Trinidad and Tobago', 'El Salvador', 'Guatemala', 'Ecuador', 'Bolivia', 'Paraguay', 'Nicaragua', 'Peru', 'Jamaica', 'Dominican Republic', 'Honduras',  'Haiti']),  'region'] = 'Latin America and Caribbean'  DF\_all.loc[DF\_all['Country'].isin(['Switzerland', 'Iceland', 'Denmark', 'Norway', 'Finland', 'Netherlands', 'Sweden', 'Austria', 'Luxembourg', 'Ireland', 'Belgium', 'United Kingdom', 'Germany', 'France', 'Spain', 'Malta',  'Italy', 'North Cyprus', 'Cyprus', 'Portugal', 'Greece']),  'region'] = 'Western Europe'  DF\_all.loc[DF\_all['Country'].isin(['Israel', 'United Arab Emirates', 'Oman', 'Qatar', 'Saudi Arabia', 'Kuwait', 'Bahrain', 'Libya', 'Algeria', 'Turkey', 'Jordan', 'Morocco', 'Lebanon', 'Tunisia', 'Palestinian Territories', 'Iran', 'Iraq', 'Egypt', 'Yemen', 'Syria']),  'region'] = 'Middle East and Northern Africa'  DF\_all.loc[DF\_all['Country'].isin(['Singapore', 'Thailand', 'Malaysia', 'Indonesia', 'Vietnam', 'Philippines', 'Laos', 'Myanmar', 'Cambodia']),  'region'] = 'Southeastern Asia'  DF\_all.loc[DF\_all['Country'].isin(['Bhutan', 'Pakistan', 'Bangladesh', 'India', 'Nepal', 'Sri Lanka', 'Afghanistan']),  'region'] = 'Southern Asia'  DF\_all.loc[DF\_all['Country'].isin(['Taiwan', 'Japan', 'South Korea', 'Hong Kong', 'China', 'Mongolia']),  'region'] = 'Eastern Asia'  DF\_all.loc[DF\_all['Country'].isin(['Canada', 'United States']),  'region'] = 'North America'  DF\_all.loc[DF\_all['Country'].isin(['New Zealand', 'Australia']), 'region'] = 'Australia and New Zealand' |
| WHR\_GDP\_2015to2019.csv | \_merge | Created when doing a left join on WHR\_all and GDP by Country 1999-2022. | DF\_all= pd.merge(DF\_WHR, DF\_GDP, on=['Country', 'year'], how='left', indicator=True) |
|  |  |  |  |
|  |  |  |  |

# Project Definition

## Defining Questions

1. How strongly is a country's GDP correlated with its happiness score?
2. Are the wealthiest countries by year ranked the happiest by the WHR?
3. Do any indicators of happiness correlate with the growth or decline of a countries overall GDP?
4. Are there certain regions of the world that are happier and/or wealthier based on these reports? How does regional economic development relate to regional happiness levels?
5. As a county’s GDP increased or decreases do any specific indicators of happiness show correlated changes?
6. Can a country’s GDP predict a certain percentage of a happiness indicator will be achieved when it is at or above a specific threshold (E.G. the top 5% of the worlds GDP contributors)? Are there any strong correlations found in the data?
7. Are there any outliers where a country's happiness is disproportionately high or low given its GDP?
8. How have the relationships between GDP and happiness changed over time?

# Exploratory Insights

* Correlations and Trends
* Factors and Implications
* Influences on Happiness
* Comparative Perspectives
  + Regional Contrasts

# Conclusions

# Limitations and Future Research

# References

* Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022). World

Happiness Report 2022. New York: Sustainable Development Solutions Network.

* MOTH (2022). GDP By Country 1999 – 2022. Retrieved [12/16/23] from

[https://www.kaggle.com/datasets/alejopaullier/-gdp-by-country-1999-2022/]

* Sustainable Development Solutions Network (2019). World Happiness Report. Retrieved [12/16/23] from

[https://www.kaggle.com/datasets/unsdsn/world-happiness]

* World Bank Group (2023). World Bank, World Development Indicators (GDP by US$).

Retrieved from [https://databank.worldbank.org/]

# Appendices