**World’s Happiest Countries**

**INTRODUCTION:** The report focuses on studying the various factors that influence the world happiness index which in turn determines the rankings of the world’s happiest countries. The motivation for finding the various factors comes from performing sentiment analysis on a video titled ‘World’s happiest countries: Explained’.

The video mentions some key points like how the economy of a country is not a main predictor of the happiness index, it emphasizes on how the countries with the highest economies do not top the list. Various economists argue that overall well-being is more important to happiness than having a good economy. This serves as a key point in our research indicating what data is required.

There are also mentions of how climate impacts the happiness of people living in a country. The list often contains Nordic/European countries which have colder climates compared to the rest of the world.

A screenshot of a graph

AI-generated content may be incorrect.

**Fig 1.1: Top 10 Happiest countries in the world**

All the data that is being used here is from 2018, the reason this year was selected is to align with the findings from the sentiment analysis. The video as mentioned before, is from 2018. There was no suitable video that had a decent number of moderated comments that were suitable for sentiment analysis. We can be more robust in this report and include findings from both 2018 and 2024 to see the changes in the rankings and what caused them.

**Analysis Questions:**

1. What is the average climate of the countries at the top and bottom of the list of world’s happiest countries?
2. What is the economy of the countries at the top and bottom of the list
3. What is a more influential factor, country’s GDP or GDP per capita?
4. How have the rankings changed from 2018 to 2024?
5. How does healthcare impact the overall quality of life?

**DATA SOURCING AND JUSTIFICATION –** The ‘main’ dataset has been obtained from [the Gallup World Poll](https://worldhappiness.report/data-sharing/), it consists of a list the world’s happiest countries. The dataset is available for free and there are no restrictions for public usage. This is the best source to acquire the data from as the survey is conducted from the Gallup, making it the original source of the dataset.

The columns are as follows:

Year: The data consists of happiness index of countries from 2011 to 2024

Rank: The rank of countries in the list

Country name: Name of the country, a categorical variable

Ladder score: Happiness index of the country

upperwhisker: Upper-bound CI of the score

lowerwhisker: Lower-bound CI of the score

Explained by: Log GDP per capita: how GDP per capita explains the happiness index

Explained by: Social support: how social support per capita explains the happiness index

Explained by: Healthy life expectancy: how healthy life expectancy explains the happiness index

Explained by: Freedom to make life choices: how freedom explains the happiness index

Explained by: Generosity: how generosity explains the happiness index

Explained by: Perceptions of corruption: how perception of corruption explains the happiness index

Dystopia + residual: The deviance from the ladder score

The complete dataset consists of 1969 rows and 13 columns. We will be using only a few of these columns and combining them with other dataset which will be acquired from similar sources. Examples include climate data, GDP data…

Climate Data - The second data source comes from [World Bank Group](https://climateknowledgeportal.worldbank.org/download-data#htab-1497). It consists of average temperature of all countries starting from 1901 This dataset consists of annual averages of temperature for each country.

Data presented on CCKP is disseminated by the World Bank under its Open Data Policy.

A close-up of a list of text

AI-generated content may be incorrect.

**Fig 3.1: Filters applied to obtain climate data**

GDP and GDP per capita – These 2 datasets were obtained from [World Bank Group](https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG?end=2023&start=2018). The World Bank Group makes data publicly available according to [open data standards](http://opendefinition.org/) and licenses datasets under the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0) (CC-BY 4.0).

The [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0) allows users to copy, modify and distribute data in any format for any purpose, including commercial use. Users are only obligated to give appropriate credit (attribution) and indicate if they have made any changes, including translations.

The excel file was modified to only include GDP of countries from 2018; to align with the findings from the sentiment analysis. It consists of every country’s GDP which can be filtered out to obtain the desired results for our research.

**DATA CLEANING –** Since there are multiple datasets to work with, we can use Power BI to transform the data and create relationships between each of these datasets based on the Country name.

Since we do not require data prior to 2018 we can simplify filter out the rows to include data only from 2018.

First, we start by cleaning the dataset World’s happiest countries:

A screenshot of a computer

AI-generated content may be incorrect.

**Fig 4.1: Missing values of dataset ‘Worlds’ Happiest Countries’**

There are a lot of missing values, especially for data after 2018. Replacing the data where almost 50% is missing does not make sense. These columns were not to be included in the analysis either way, so we can delete these columns.

A screenshot of a black background with numbers and numbers

AI-generated content may be incorrect. **Fig 4.2: Countries ranked from dataset ‘Worlds’ Happiest Countries’**

Once these columns have been removed and the years have been filtered from 2018 to 2023, we can check for missing values:

A screenshot of a computer

AI-generated content may be incorrect. **Fig 4.3: Column profile of happiness index from dataset ‘Worlds’ Happiest Countries’**

There were no missing values in either of these columns (Images of other columns have been excluded to avoid repetition).

Cleaning the second dataset ‘GDP by Country’: There a lot of missing values for GDP for different countries across the year.

A screenshot of a computer code

AI-generated content may be incorrect. **Fig 4.4: List of countries with missing data from dataset ‘GDP by Country’**

These countries are not in the list that we need for our analysis. So instead of replacing them we can remove them. But, before we can remove the missing data, it is key to look at what the dataset is organized as:

A screenshot of a computer

AI-generated content may be incorrect. **Fig 4.5: Columns displaying the spread of data across the years from dataset ‘GDP by Country’**

We can unpivot these columns and rename the column to year to match the format of the other datasets.

A screenshot of a computer

AI-generated content may be incorrect. **Fig 4.7: Un-pivoted columns and changed data type to year from dataset ‘GDP by Country’**

In the above step the columns 2018 – 2023 were un-pivoted, renamed to year and changed data type to year. The other columns ‘Country code’, ‘Indicator name’, ‘Indicator code’ were removed as they are not required for our analysis. The GDP column was filtered to remove missing data from the list of countries which are not part of the analysis as shown in the image of the Python code above (Fig 4.4).

The next dataset ‘GDP per Capita’ will follow the same steps as the dataset ‘GDP by Country’

A screenshot of a computer

AI-generated content may be incorrect.

**Fig 4.8: Un-pivoted columns and changed data type to year from dataset ‘GDP per Capita’**

The final dataset consists of weather data for different countries from 1901 to 2023.

A screenshot of a computer

AI-generated content may be incorrect.

**Fig 4.9: Climate data across the years for different countries from dataset ‘Climate data’**

The data looks similar to previous datasets; we can use roughly the same steps to clean the data.

A screenshot of a computer

AI-generated content may be incorrect.

**Fig 4.10: Un-pivoted columns and changed data type to year from dataset ‘Climate data’**

Now that all the data has been cleaned, we can create relationships for the dataset based on the country name.

A screenshot of a computer

AI-generated content may be incorrect. **Fig 4.11: Creating Many-to-Many relationships between multiple tables based on column country**

It is important to create this relationship as the other datasets will help explain the relationship between World Happiness index and other factors. With this, we conclude the data cleaning/pre-processing part. We can now move on to visualizing the data.

**SECTION E: VISUALIZATIONS AND FINDINGS(5%) - Should focus on Exploratory Data Analysis (EDA), Statistical Visualization, Classification, Cluster and Sentiment Analysis.**

**SECTION F: FILE ORGANIZATION- (1%)**