**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.

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 [Berkely Earth Data](https://berkeleyearth.org/data/). It consists of average temperature of all countries starting from 1750! This dataset consists of monthly averages of temperature for each country; we can select the year as 2018 and 2024 to obtain the average temperature.

Berkeley Earth’s data is licensed under [Creative Commons BY-NC 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/) for non-commercial use only. The raw dataset has been put together from reputable sources such as NASA’s GISTEMP, HadCrut, MLOST; making it a reputable source.

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.

**SECTION D: DATA CLEANING**

**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%)**