Global Happiness Index Analysis & Visualization Project Report

World Happiness Report Analysis (2020–2024)
Prepared by: *Aslican Alacal*
Major: *M.Sc. Data Science*
Date: *08/07/2025*

1. Introduction

• **Objective:** This project aims to explore and visually represent the data from the World Happiness Report spanning 2020 to 2024. The goal is to identify evolving trends over the years, pinpoint the most influential factors contributing to happiness, and examine how different countries rank and compare based on these factors.

• **Dataset:** The dataset is from World Happiness Report website and was assembled by merging multiple CSV files representing each year's report from 2020 to 2024.To ensure analytical consistency, the data underwent thorough cleaning and preprocessing steps.

The column names are:

- Country
- Year
- Explained by: GDP per capita
- Explained by: Social support
- Explained by: Healthy life expectancy
- Explained by: Freedom to make life choices
- Explained by: Generosity
- Explained by: Perceptions of corruption
- Dystopia + residual

2. Data Preparation

• **Loading:** All five annual datasets were consolidated into a single, unified DataFrame to allow seamless year-over-year comparisons.

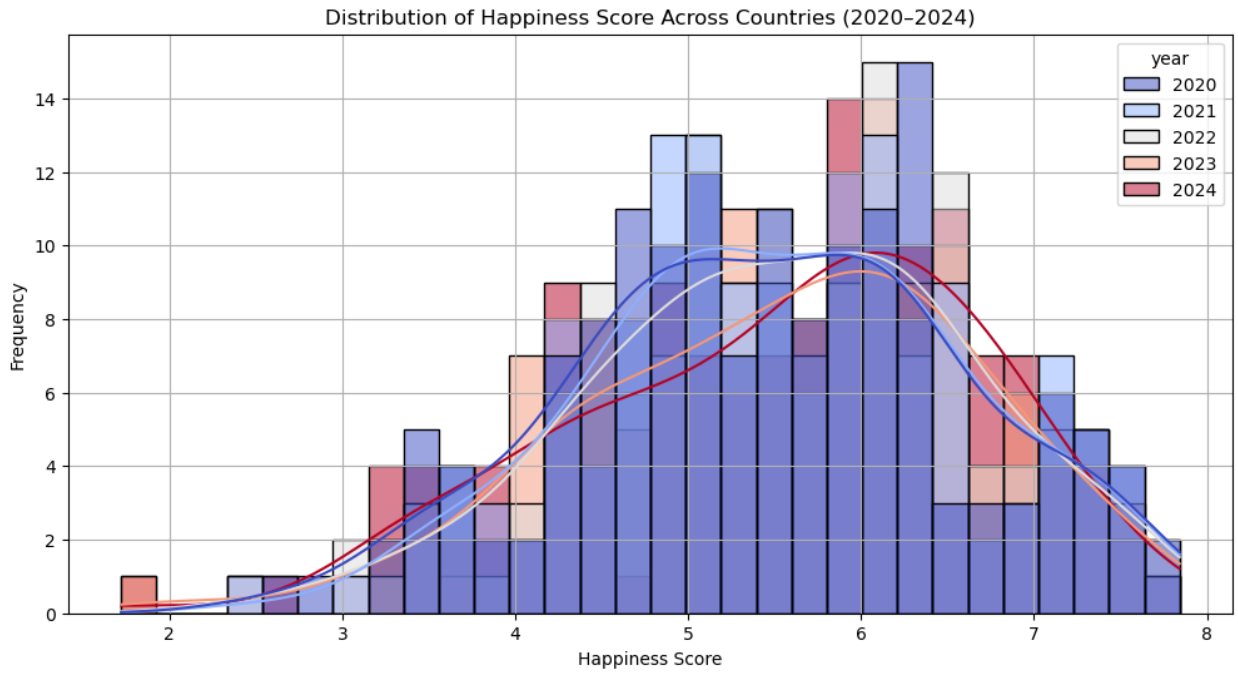
• **Cleaning:** Column names were standardized for uniformity, missing values were appropriately addressed, and discrepancies in country names were resolved to enable accurate geographical mapping and analysis.

• **Feature Engineering:** A new metric was introduced, total happiness score, calculated as the sum of all individual contributing factors such as GDP, social support, life expectancy, freedom, generosity, and perceptions of corruption.

3. Exploratory Data Analysis

Distribution of Happiness Scores

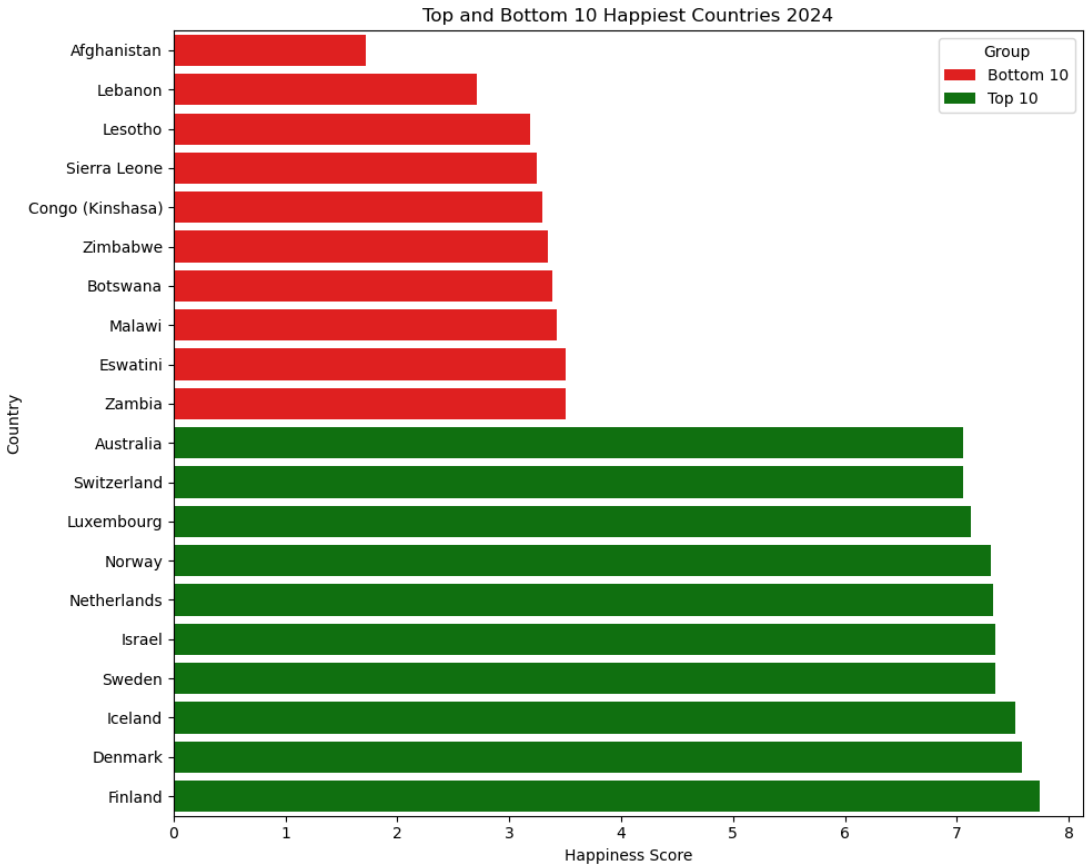
- **Histogram:** Analysing the distribution of the happiness score across countries using histogram.



- **Insights:** The majority of countries tend to cluster within a happiness score range of 5 to 6, indicating a global average level of well-being. However, there are a few notable outliers, some significantly happier and others markedly less so. Interestingly, the data for higher-scoring countries predominantly comes from the years 2020 and 2021, suggesting that these years may have seen improved conditions or more consistent reporting. In contrast, countries with lower happiness scores have data distributed across the entire range from 2020 to 2024, pointing to persistent challenges or consistent underperformance over time.

Top & Bottom Countries

- **Bar Chart:** Identifying the top and bottom 10 countries by happiness score using horizontal bar chart.



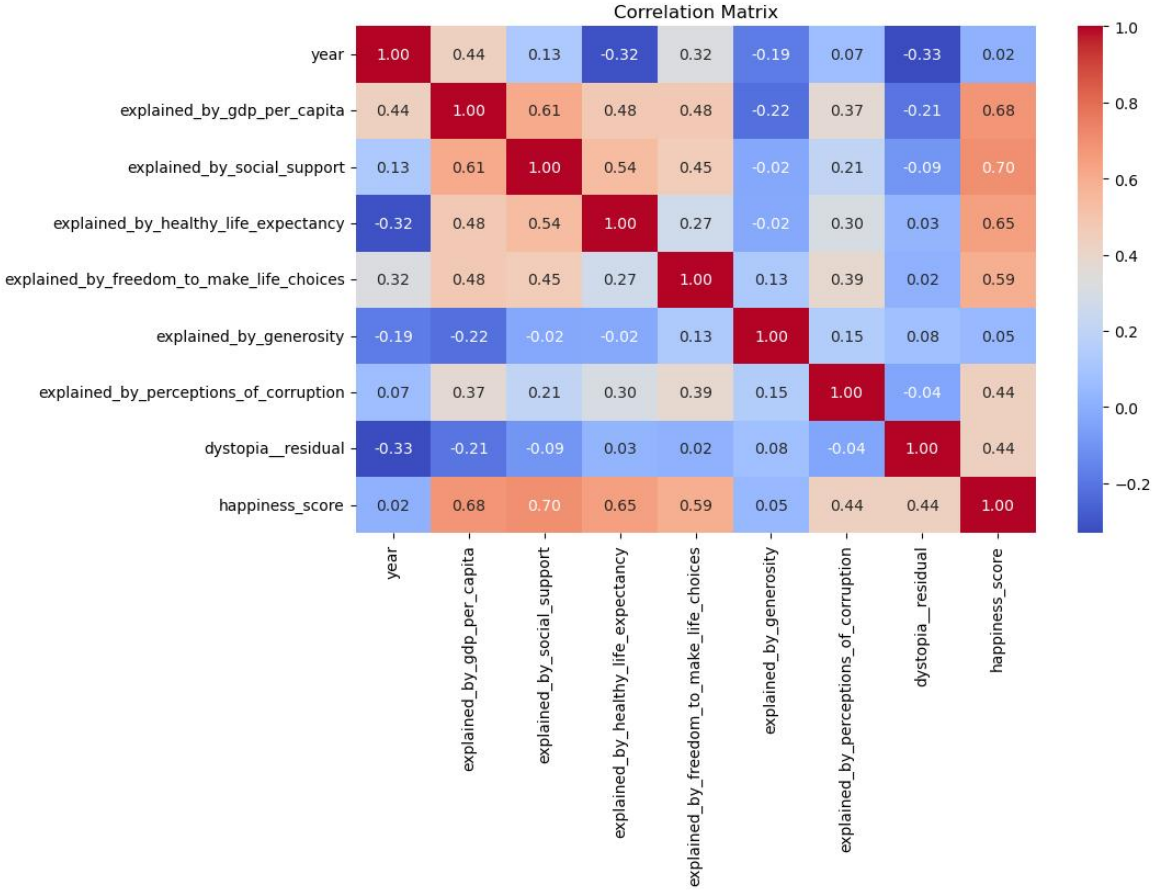
- **Observation:** A consistent pattern emerges when examining the extremes of the World Happiness Report rankings. Nordic countries, known for their strong social systems, economic stability, and high quality of life, continue to dominate the top positions year after year. In contrast, nations experiencing conflict, economic hardship, or political instability tend to fall toward the bottom of the rankings.

  Specifically, the ten lowest-ranking countries in terms of happiness scores are, in ascending order: **Afghanistan, Lebanon, Lesotho, Sierra Leone, Congo, Zimbabwe, Botswana, Malawi, Eswatini**, and **Zambia**. These countries often face compounding challenges such as poverty, limited access to healthcare and education, or prolonged conflict and all of which significantly impact overall well-being.

  On the other end of the spectrum, the ten highest-ranking countries for happiness are: **Australia, Switzerland, Luxembourg, Norway, Netherlands, Israel, Sweden, Iceland, Denmark**, and **Finland**. These nations typically benefit from strong governance, high levels of trust in institutions, generous social support systems, and robust public services factors that consistently correlate with greater life satisfaction.

Correlation Analysis

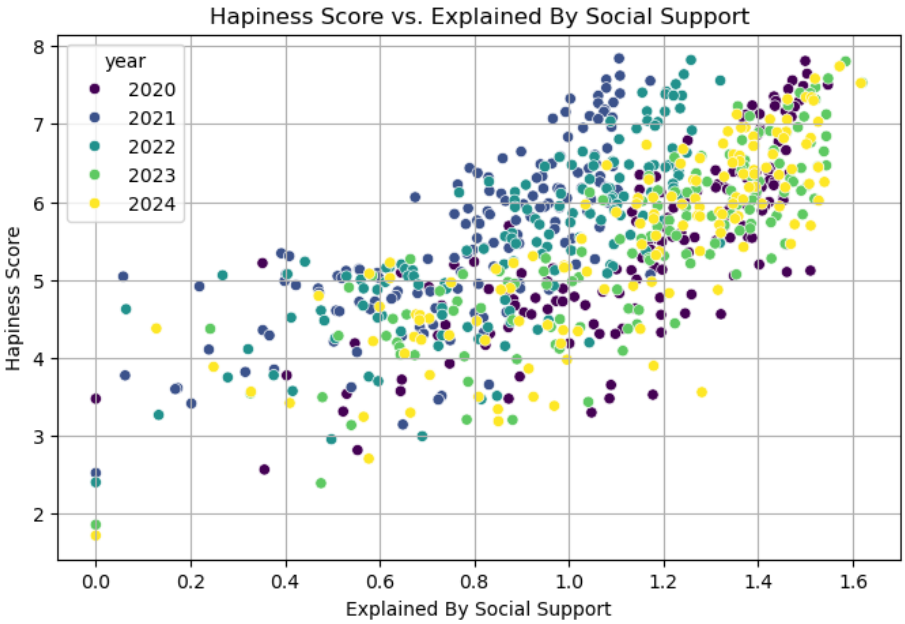
- **Heatmap:** Revealing strong positive correlations between happiness and factors like GDP per capita, social support, and healthy life expectancy.

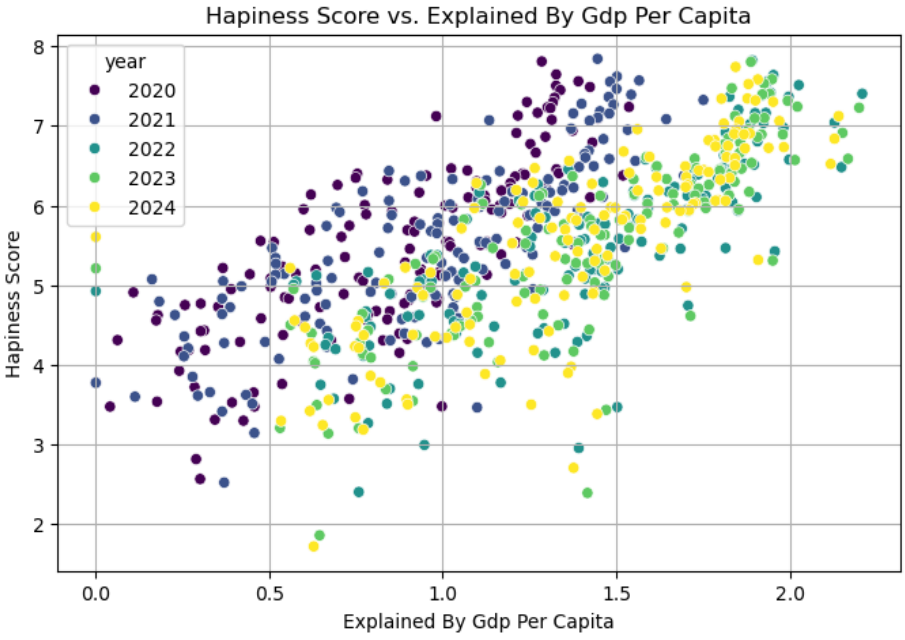


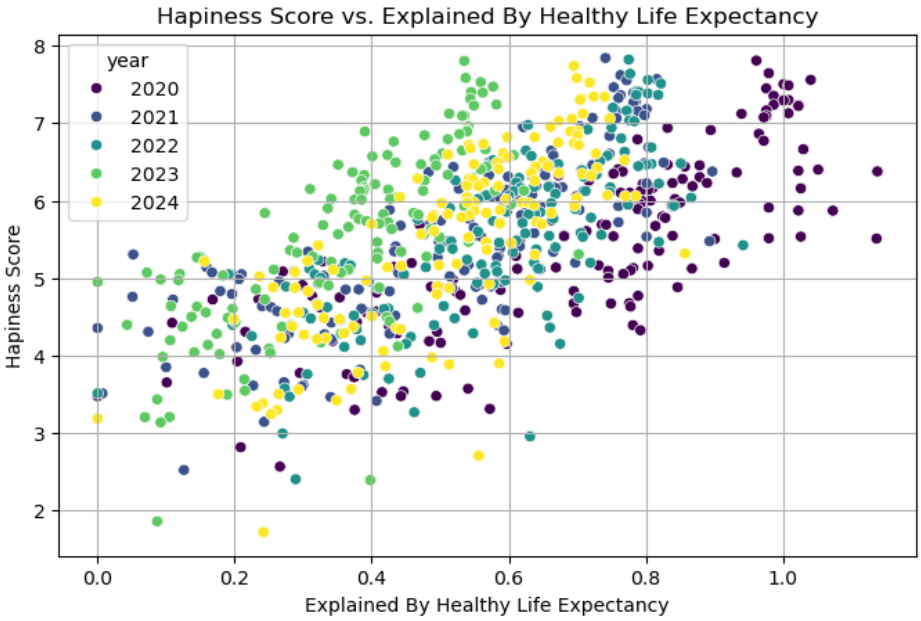
- **Insights:** According to our correlation matrix, the top three features most strongly associated with the happiness score and ranked in ascending order, are: healthy life expectancy, GDP per capita, and social support. These variables demonstrate the highest positive correlation with overall happiness, indicating that countries with better health outcomes, stronger economies, and more robust social support systems tend to report higher levels of life satisfaction.

  Recognizing the importance of these key factors, are being selected for further exploration through scatter plot visualizations. These plots will help illustrate the nature and strength of the relationships between each feature and the happiness score, providing a clearer picture of how these elements contribute to a nation's well-being.

- **Scatter Plots:** Finding correlations between happiness and other indicators using scatter plot.





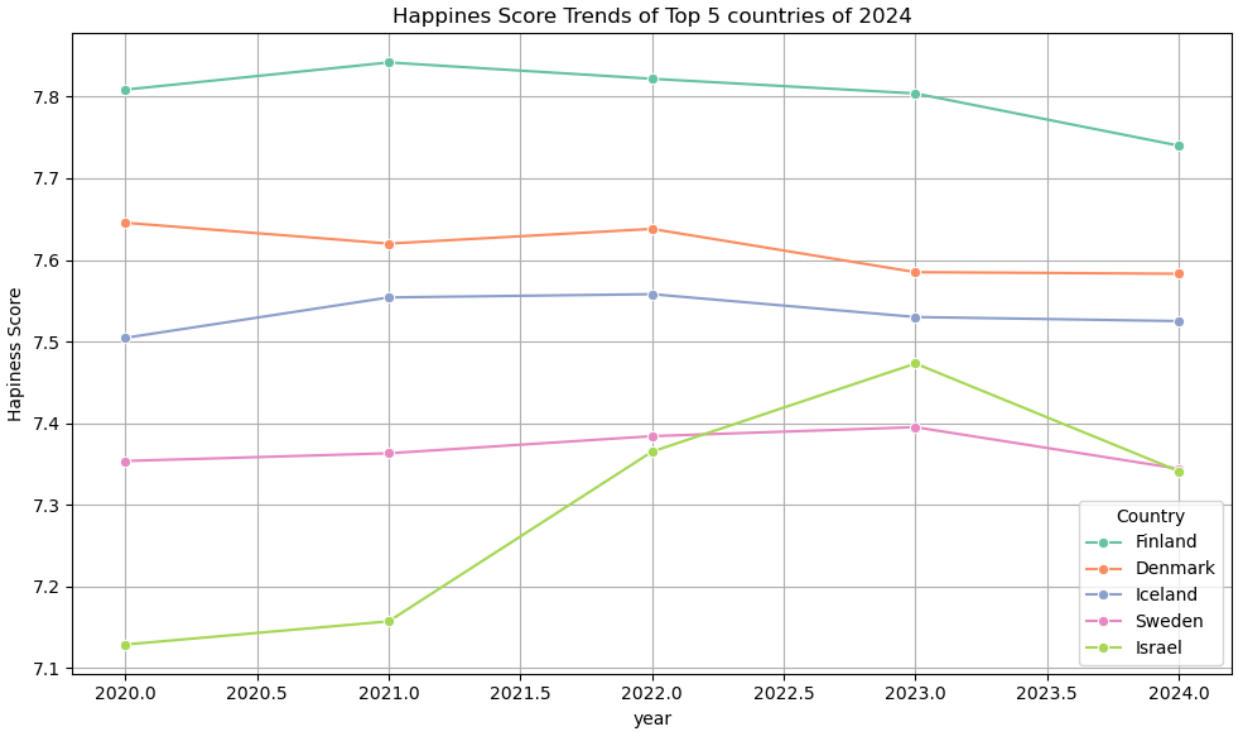


Insights: Based on the three scatter plots, the most clearly linear relationship appears in the second plot: Happiness Score vs. GDP per Capita. This visualization shows a strong positive trend, suggesting that as a country's GDP per capita increases, so does its happiness score. However, there are a few notable outliers, particularly from the 2023 and 2024 data, where some countries with relatively high GDP levels report lower happiness scores than expected. This indicates that while economic prosperity is a strong predictor of happiness, it is not the sole determinant.

In contrast, the plots for Healthy Life Expectancy and Social Support reveal more complex and less linear patterns. For instance, there are countries with low life expectancy values that still report moderate or even high happiness scores. Similarly, in the case of social support, some nations with limited support systems still score relatively high in happiness, while others with strong support networks rank unexpectedly low. These deviations suggest that while health and social structures are important, cultural, psychological, and contextual factors may also significantly influence overall well-being.

Trends Over Time

- **Line Plot:** Analysing how the rankings of top countries changed over the last five years using Line plot.

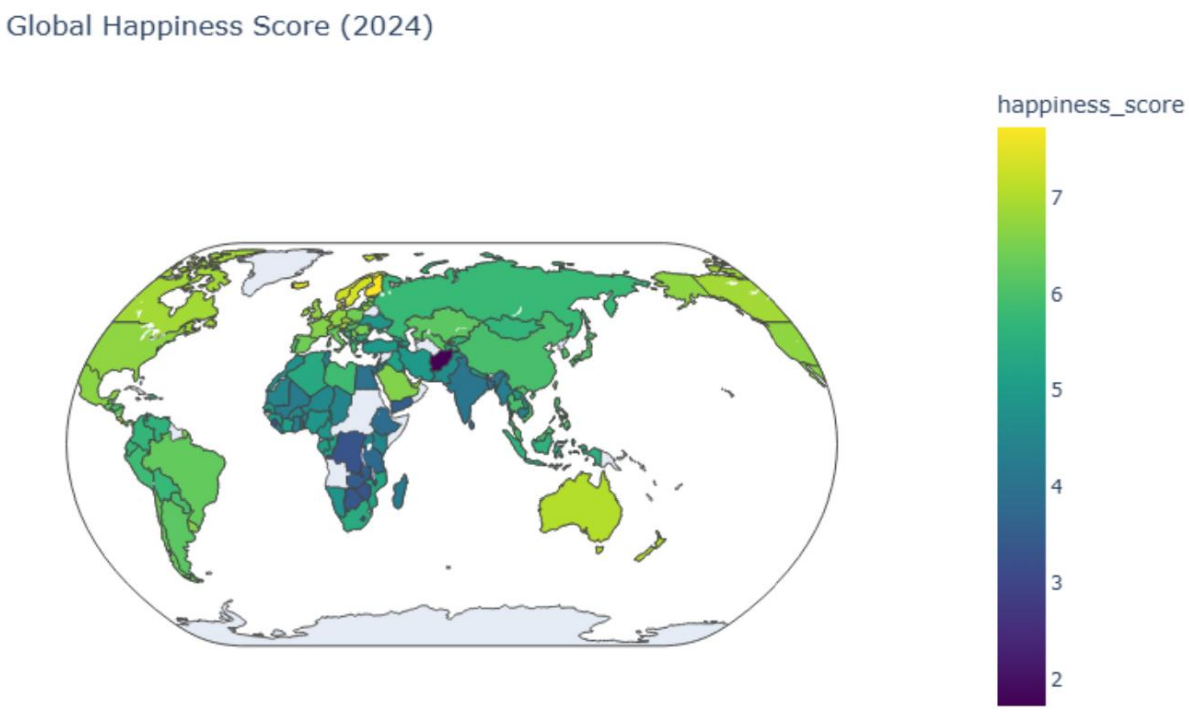


- **Findings:** When examining the top four happiest countries, Finland, Denmark, Iceland, and Sweden, one noticeable trend is the remarkable stability in their happiness scores over the years. The fluctuations are minimal, typically within a narrow range of just 0.1 points, highlighting a consistent sense of national well-being. This steadiness can largely be attributed to factors discussed earlier, such as economic stability, trust in government, strong social support systems, and high-quality public services, all of which contribute to a sustained level of life satisfaction.

  Israel, however, presents a slightly different story. Its happiness score saw a notable rise, placing it among the top five happiest countries at one point. Yet, beginning in 2023, a downward trend became evident. One possible contributing factor to this recent decline could be the ongoing war, which may have significantly impacted public sentiment, sense of safety, and overall well-being. This case underscores how sociopolitical events can quickly impact national happiness, even in countries that previously ranked very high.

Geographic Visualization

- **Choropleth Map:** Ploting happiness scores on a world map using Choropleth map.



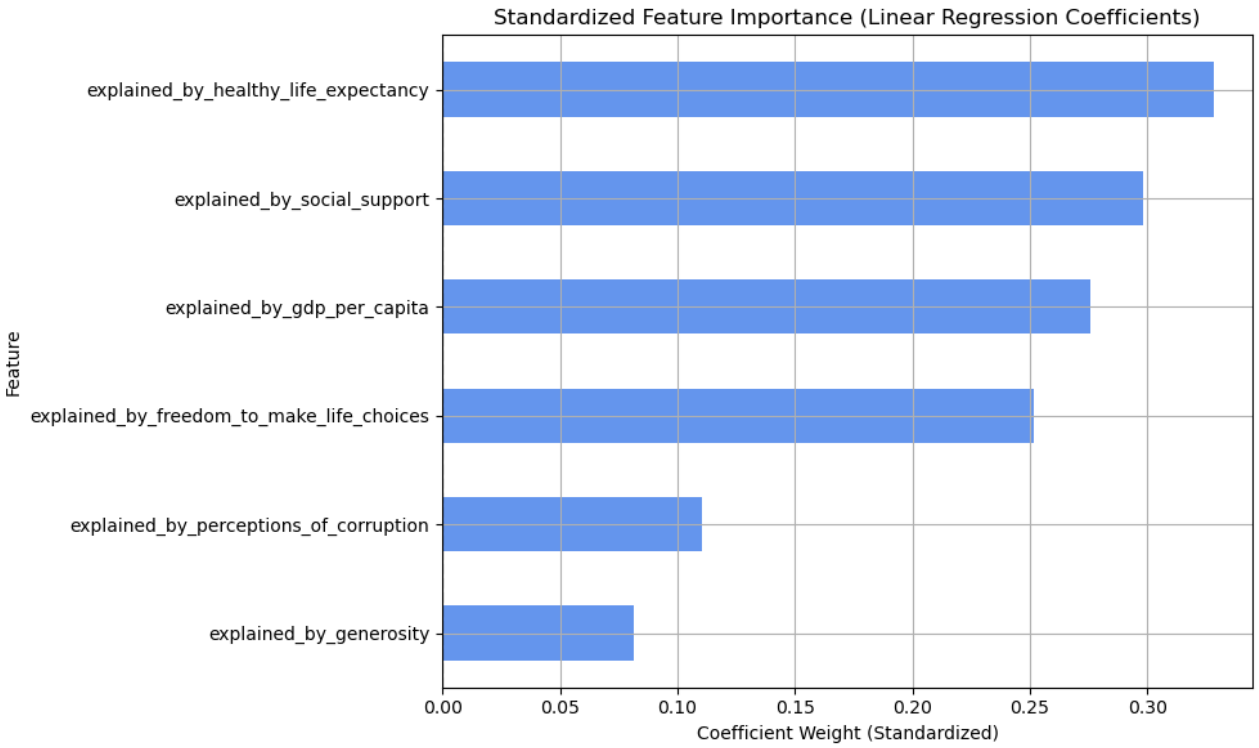
- **Insights:** Many of the visualizations in this report are based on data from the most recent available year, selected to enhance clarity, relevance, and interpretability. This decision was made to prioritize clarity and simplicity, ensuring the visual insights are both current and easy to interpret.

  Looking specifically at the choropleth map, several regional patterns emerge. Wealthier regions, including Europe, North America, and Oceania, tend to consistently score higher in happiness. This aligns with earlier findings that link economic prosperity, stable governance, and strong social infrastructure with greater overall well-being.

  As expected, Greenland has no available data, which is common across many global datasets due to its limited survey coverage. On the other end of the spectrum, Afghanistan continues to hold the lowest happiness score, reflecting its ongoing struggles with conflict, instability, and humanitarian challenges.

Factor Analysis

- **Linear Regression:** Performing linear regression or feature importance to understand which factors most affect happiness.

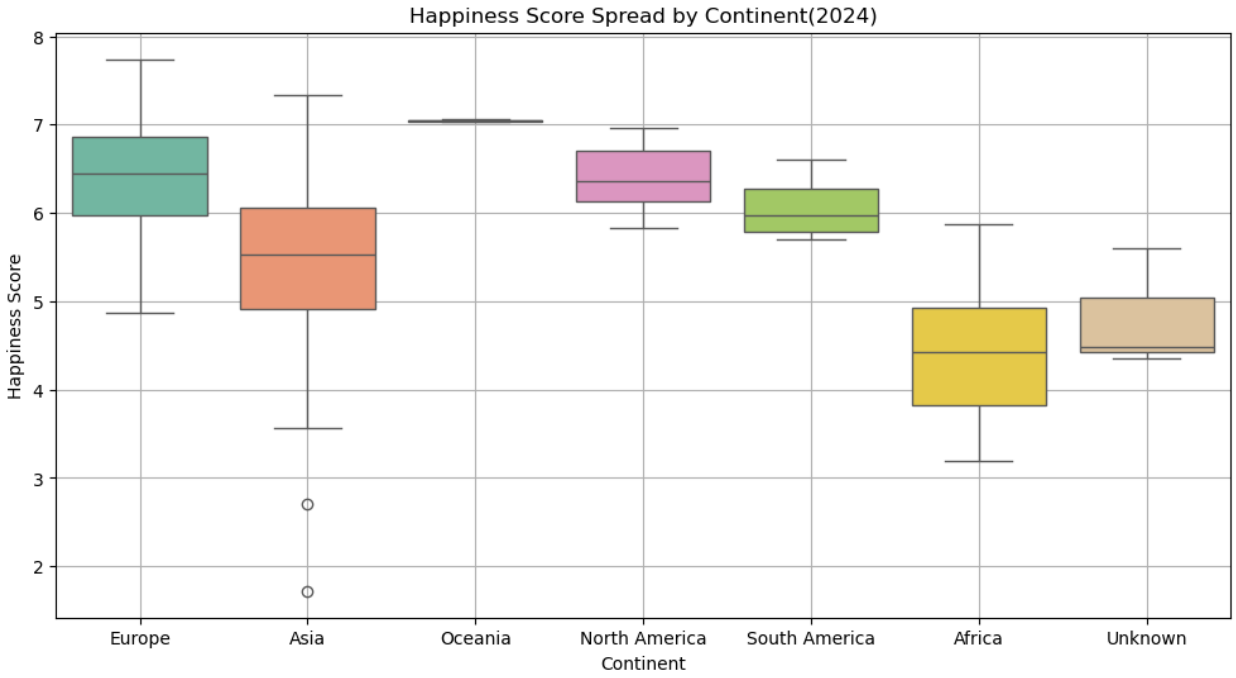


- **Insights:** According to the horizontal bar plot, healthy life expectancy emerges as the most important factor influencing the happiness score. It is followed by social support, GDP per capita, freedom to make life choices, perceptions of corruption, and generosity. This ranking provides a clear view of which elements contribute most significantly to national well-being.

  It's worth noting that the data used for this visualization spans all available years in the dataset (2020–2024), which adds weight to the results. By considering the full timeframe, the analysis captures consistent patterns rather than temporary fluctuations, offering a more reliable perspective on what drives happiness globally.

Group & Cluster Analysis

- **Box Plot by Continent:** Grouping countries by continent and compute average happiness using box plots.



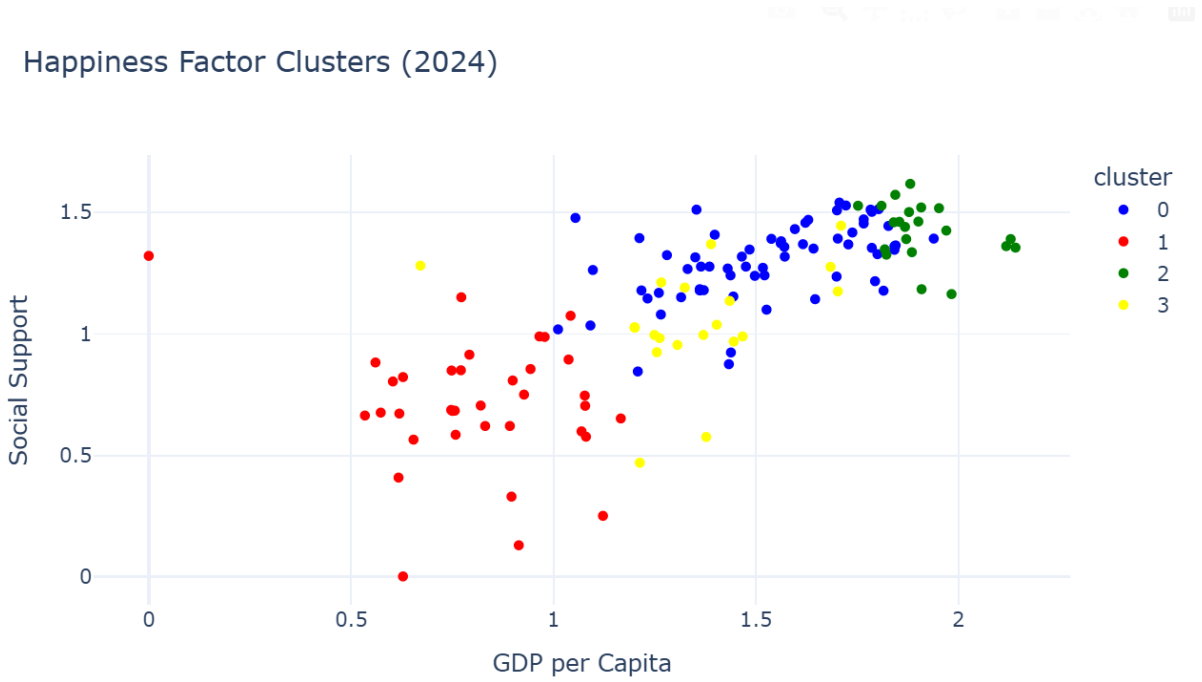
- **Insights:** Based on the box plots for the 2024 data, regions like Europe and North America show a relatively balanced distribution in happiness scores. Both continents exhibit a well-centred median and a symmetrical spread, indicating consistent levels of well-being across countries within these regions.

  While Oceania reports the highest median happiness score, the box plot lacks visible quartiles or a spread range, likely due to the small number of countries in the region. As a result, only the median is displayed, limiting deeper insights into variability.

  The category labelled "Unknown" continent likely includes countries from the Middle East. This group shows a wider and uneven spread, along with a slightly lower median, reflecting greater disparity in happiness levels. Similarly, Africa demonstrates a mild imbalance in its distribution but still presents a clear median of around 4.5, which is close to the median of the "Unknown" region.

  Asia, on the other hand, shows a slightly higher overall happiness score compared to Africa and the Middle East, but with two significant outliers. One outlier falls near a score of 2.5, likely representing Lebanon, while the second, closer to 1.5, almost certainly corresponds to Afghanistan, both countries consistently ranked among the lowest in global happiness.

K-Means Clustering: Applying clustering algorithms (K-Means) and visualizing in 2D scatter plot.



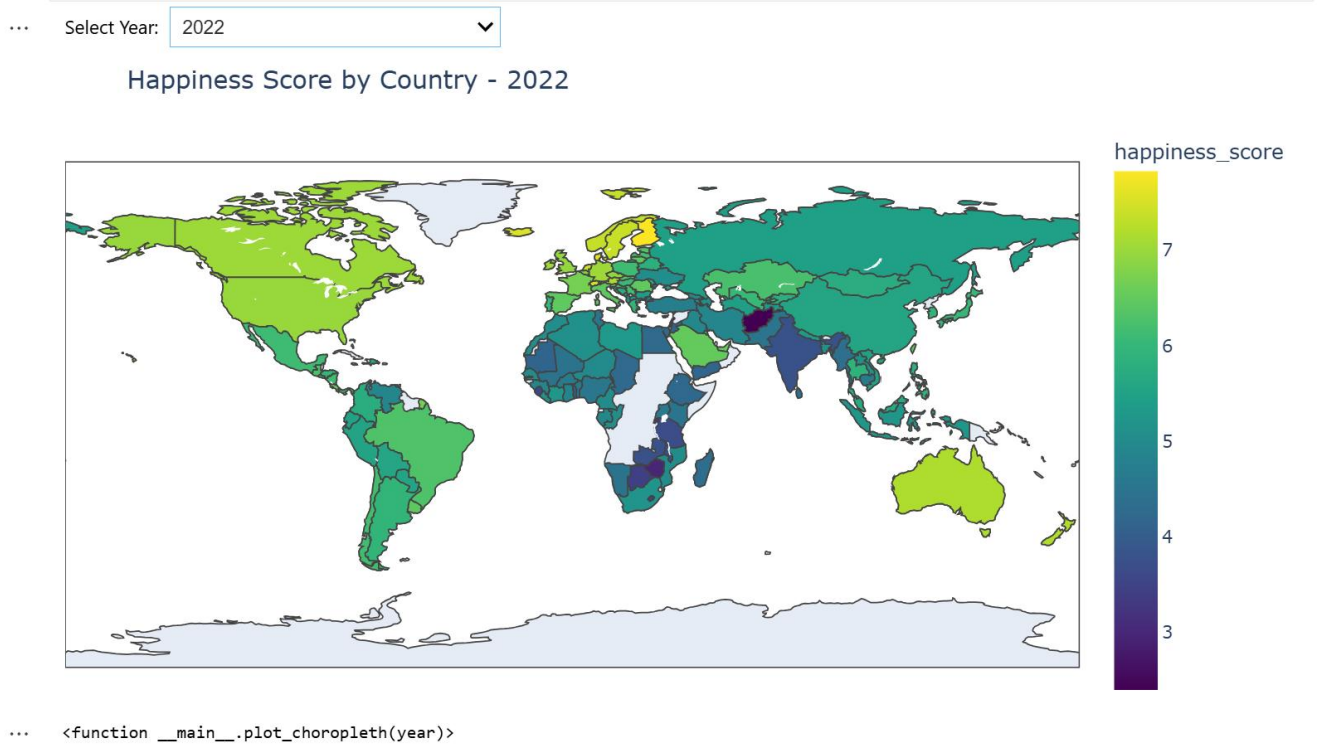
- **Insights:** After applying the K-Means clustering algorithm with n = 4, four distinct clusters emerged based on countries' GDP per capita and social support levels. Among them, Cluster 0 appears most frequently, suggesting that a significant number of countries fall into this grouping. It's important to note that Plotly Express automatically orders legend items based on their appearance in the dataset, not by their numerical cluster labels, so the order in the legend does not necessarily reflect the actual cluster rankings.

  According to the scatter plot, Cluster 0 is characterized by relatively high GDP per capita and strong social support, which may explain its prevalence. Cluster 2, although less populated, includes countries with the highest GDP and social support levels overall, typically representing wealthy, high-performing nations with smaller populations.

  Cluster 3 consists of countries with moderate GDP and moderate to high levels of social support, placing them somewhere between developing and developed status. Lastly, Cluster 1 generally includes countries with low GDP per capita and low to moderate social support. However, there are some exceptions, countries like Venezuela, Uganda, and Nigeria fall into this cluster despite not entirely fitting the pattern, possibly due to unique economic or social conditions affecting their placement.

Interactive Dashboards

- **Year Selector:** Interactive maps for any year.



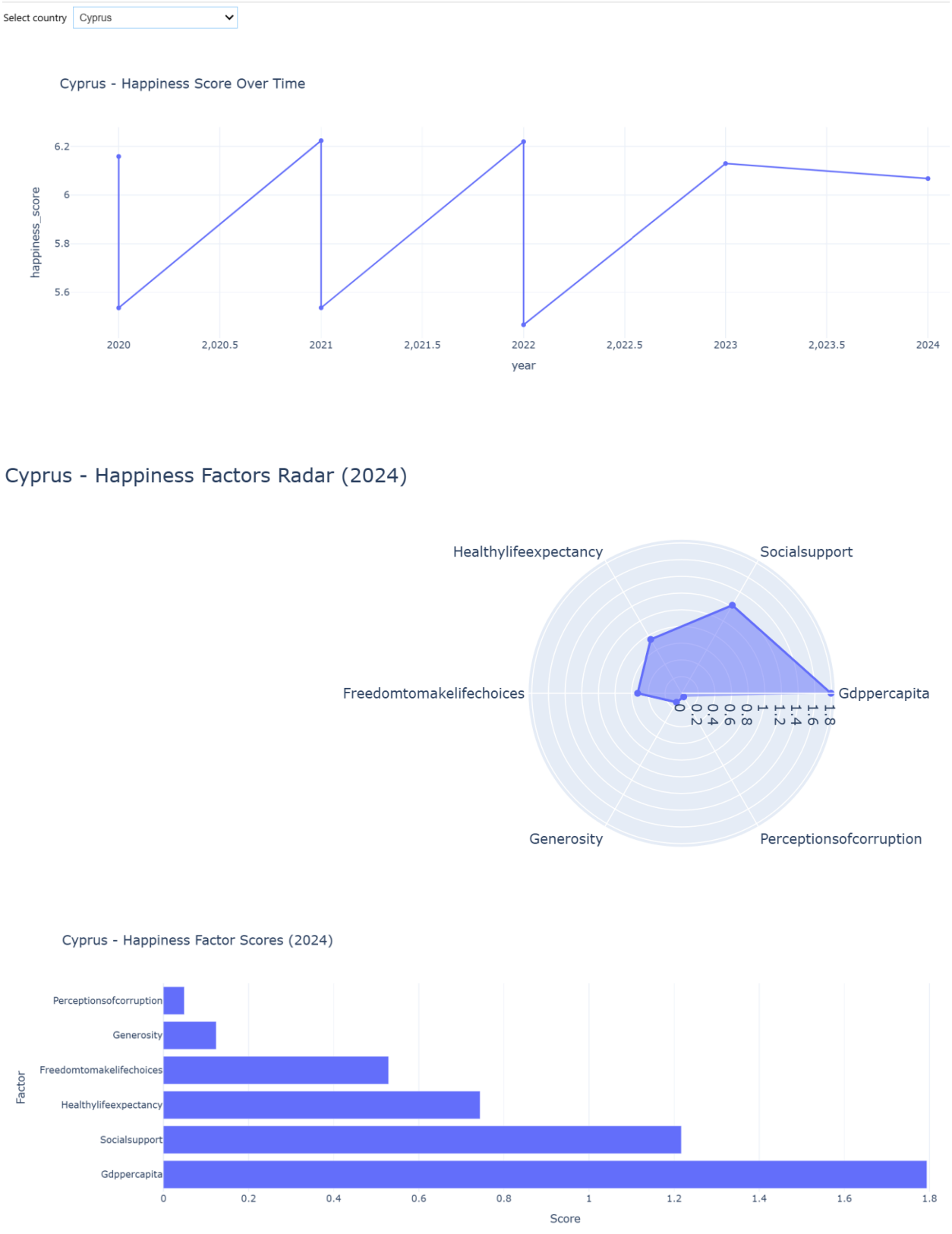

```
<function __main__.plot_choropleth(year)>
```

- **Insights:** At the top of the interactive dashboard, a dropdown menu allows users to select the year, displaying happiness scores by country for each selected time point. This feature enables an intuitive, year-by-year exploration of how happiness has evolved globally.

  From the data, it's fair to observe that regions such as Europe, North America, and Oceania have shown remarkable consistency in their happiness scores across the five-year period. These regions tend to maintain stable levels of well-being, likely due to sustained economic development, strong institutions, and resilient social systems.

  In contrast, countries facing economic or political instability, such as Turkey, have experienced more noticeable fluctuations, with a general decline in happiness scores over the years. This trend reflects how national well-being can be sensitive to economic downturns, inflation, or governance issues, highlighting the importance of long-term stability in maintaining happiness.

- **Country Profile:** Selecting a country to view its happiness trend, radar chart of factors, and factor bar chart.



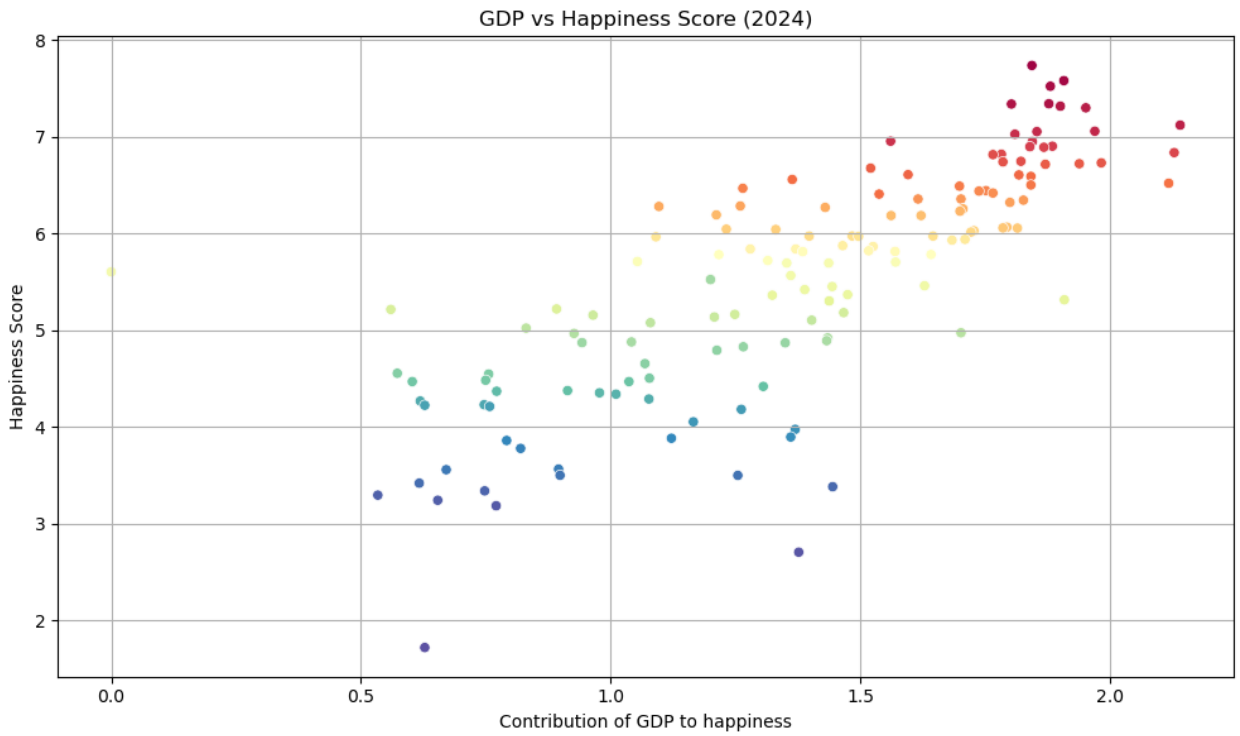
Findings: In the combined interactive dashboard, Cyprus was selected as a case study due to its notable year-to-year fluctuations in happiness scores. The trendline for Cyprus forms a clear zigzag pattern, reflecting significant changes almost every year. These sharp ups and downs suggest varying national conditions or public sentiment during the period. However, between 2023 and 2024, the trend becomes more stable, indicating a potential return to consistency or recovery from past instability.

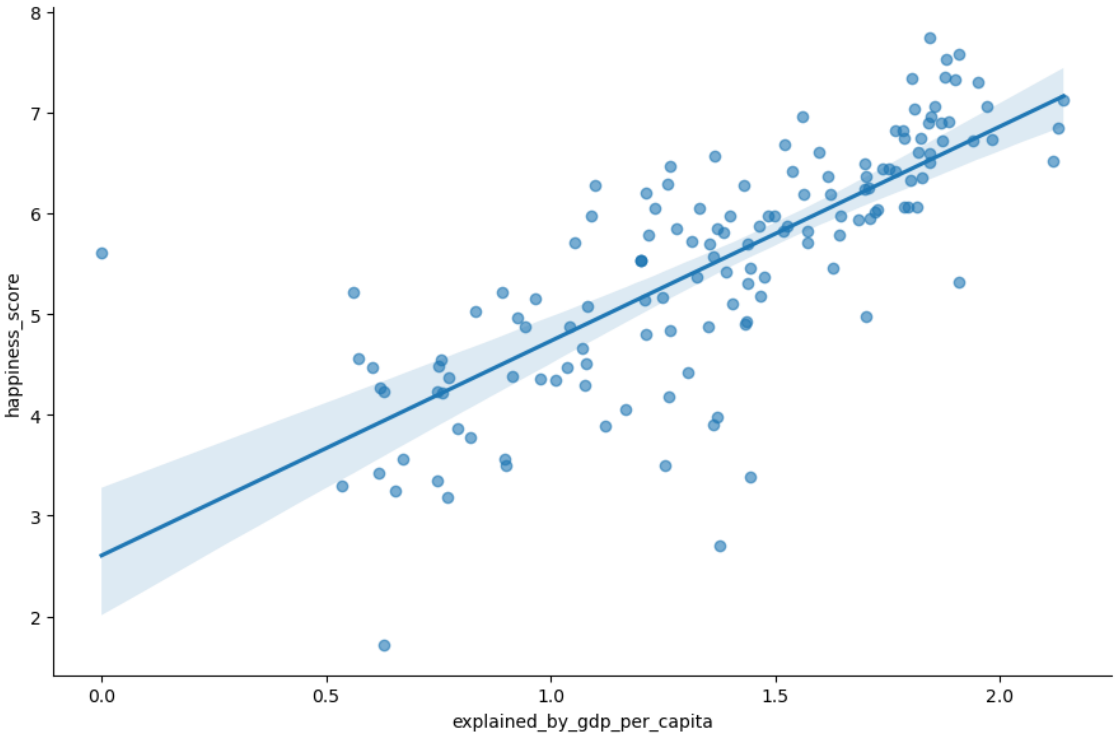
The radar chart provides deeper insight into the underlying factors contributing to Cyprus's happiness score. It shows that GDP per capita and social support are the most influential factors, closely tied to the overall well-being of the population. In contrast, generosity and perceptions of corruption appear to have a less significant impact.

To complement the radar chart, the horizontal bar chart presents the same data in a clearer, more linear format, reinforcing the relative importance of each factor. Together, these visualizations offer a comprehensive view of what drives happiness in Cyprus and how it compares year over year.

Special Analyses

- **GDP vs. Happiness: Analysing the relationship between GDP per capita and happiness score using Scatter Plot**



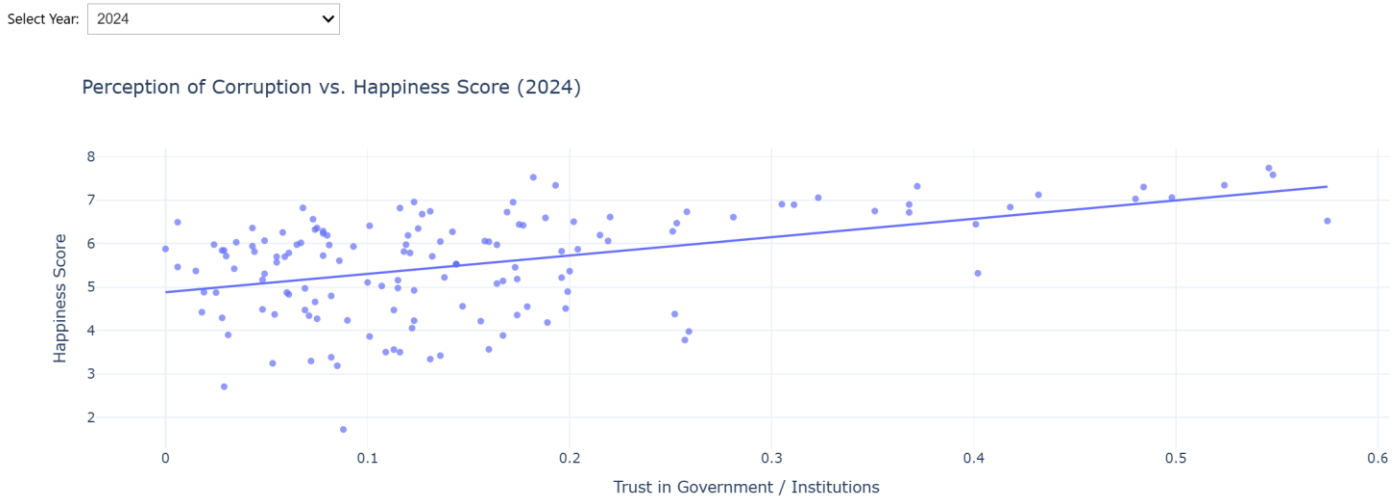


Findings: The data used for this visualization is drawn from the 2024 dataset, offering the most recent snapshot for understanding these trends.

The scatter plot with a regression line reveals a strong positive relationship between the selected variables, indicating that as one increases, the other tends to rise as well. The clear upward trend suggests a good fit, reflecting a meaningful correlation in the data.

However, since this plot is not part of an interactive dashboard, it is more challenging to identify and explore specific outliers directly. Still, a few data points clearly deviate from the regression line, signaling that some countries differ notably from the overall pattern, either performing better or worse than expected based on the variables.

- **Corruption vs. Happiness: Examining how perception of corruption correlates with happiness with plotly express**

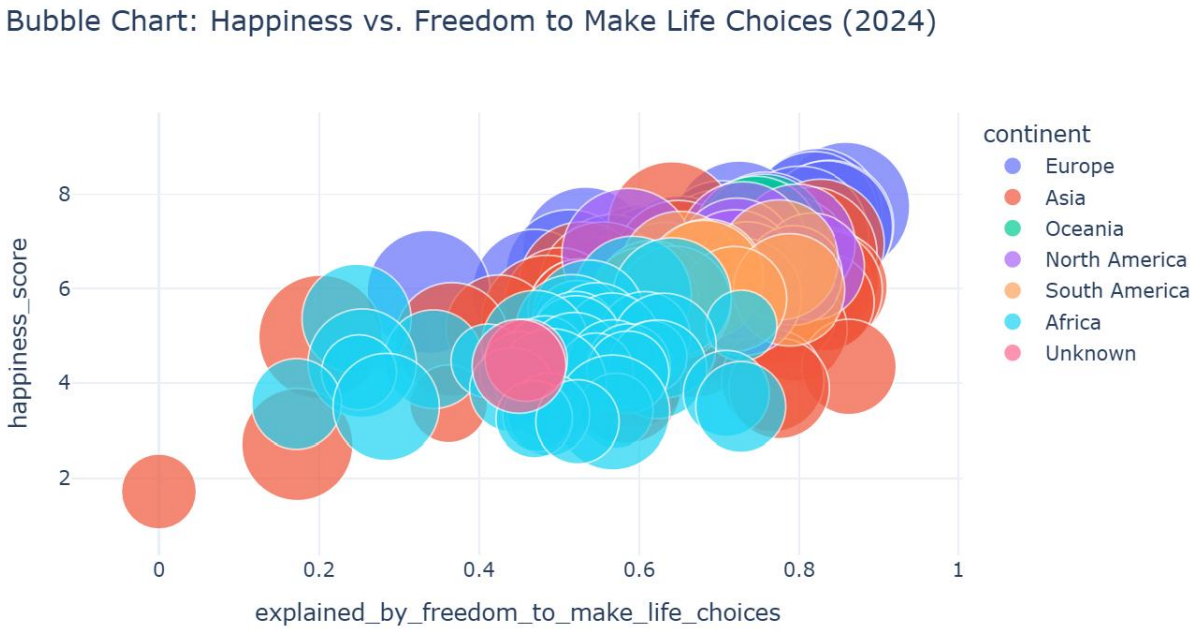


Findings: In this interactive dashboard, the relationship between trust in government and happiness score is analyzed, revealing a weak and slightly negative correlation. This suggests that, while trust in institutions is often considered important, it may not have a strong direct impact on happiness when viewed in isolation, or its effect may vary widely between countries.

The scatter plot also reveals the presence of several outliers, countries where high or low trust levels do not align with the expected happiness scores. These exceptions highlight the complexity of national well-being and the influence of additional social, economic, or cultural factors.

Using the year dropdown menu, users can switch between different years to observe how this relationship evolves over time. Each data point represents a country, allowing for detailed, country-level insights and comparisons across years.

- **Freedom vs. Happiness: Exploring the relationship between the freedom to make life choices and happiness score using Bubble chart**



Findings: Using this interactive scatter plot, users can explore the relationship between freedom to make life choices and happiness score, with additional context provided by country, continent, and GDP per capita. Although the dataset does not include population figures, the bubble sizes represent GDP per capita, offering a visual cue for each country's economic standing.

The plot reveals a general positive trend, countries with greater freedom to make life choices tend to report higher happiness scores. At the lowest end of the spectrum, Afghanistan is positioned with both low freedom and low happiness, reflecting ongoing challenges. In contrast, Finland appears in the upper right corner, combining high levels of personal freedom and happiness, consistent with its position as one of the world's happiest nations.

This visualization allows for a deeper, more interactive understanding of how freedom and economic context may jointly influence national well-being.

Animated Visualization

- **Choropleth Animation: Creating an animation showing how global happiness has changed over the last five years**

Findings: This animation illustrates how happiness scores have changed globally over the past five years, offering a dynamic view of shifts in well-being across all countries. Regions like Europe, North America, and Oceania have generally maintained stable happiness levels, with little variation over time, highlighting their resilience and consistency.

In contrast, some countries such as Turkey, Pakistan, and Lebanon have experienced noticeable declines in happiness. These shifts are reflected in the map turning "greener" for these nations, indicating lower happiness scores compared to previous years. The animation effectively captures these year-by-year transitions, making it easy to observe both global patterns and regional changes in well-being.

11. Conclusions

• **Key Drivers:** Healthy life expectancy, social support, and GDP per capita consistently emerged as the most influential factors driving happiness scores. Regression analysis and feature importance rankings confirmed these as the top predictors, while factors like generosity and perceptions of corruption showed weaker correlations.

• **Regional Trends:** Regions such as Europe, North America, and Oceania maintained stable happiness scores across 2020–2024. Nordic countries, in particular, continued to lead the rankings with minimal fluctuation. In contrast, countries experiencing economic or political instability, such as Turkey, Lebanon, and Afghanistan showed declining or highly variable trends.

• **Policy Implications:** Findings suggest that improving public health, strengthening social support systems, and fostering economic resilience are critical for boosting national happiness. Additionally, addressing trust in government and freedom of choice may help certain countries close the well-being gap seen in regional outliers.

12. Limitations & Future Work

Limitations & Future Work

While this analysis provides valuable insights into global happiness trends, several limitations should be considered. One key constraint is the use of **derived indicators** rather than raw values, for example, features like **"Explained by GDP per capita"** represent the contribution of GDP to happiness rather than actual economic output. This distinction limits the ability to interpret absolute socioeconomic conditions and may introduce bias in understanding the real-world impact of these variables. Additionally, the dataset lacks certain dimensions such as **population size**, which could have added context to visualizations like bubble charts. Some countries, like **Greenland**, were excluded due to missing data, and overall country-level happiness may be affected by **unmeasured cultural or geopolitical factors**.

For future work, incorporating **actual economic, health, and social indicators**, as well as additional variables like **mental health, education, or environmental sustainability**, could yield a more comprehensive picture. Employing more advanced methods, such as **time-series forecasting, causal inference models, or dynamic clustering**, could also enhance the depth and predictive power of the analysis.

Appendix

- **Code:** All code and visualizations are available in the attached Jupyter Notebook.
- **Dataset Source:** World Happiness Report (https://worldhappiness.report/)