

Research Question and Data Sourcing

Research Question: What influences happiness across countries?

Relevance: Determinants of happiness are important for understanding because policymakers, researchers, and organisations seek to improve the quality of life. Moreover, governments and organisations can design better policies and run programs that help promote happiness by identifying which factors contribute most to happiness. In addition, this analysis can give information about how happiness is shaped by various socio-economic and cultural contexts, which might help us understand how to design interventions of happiness tailored to local needs.

Data Sourcing: For this analysis, the dataset is from the World Happiness Report (<https://worldhappiness.report/data/>); some indicators include the GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. The data runs for years across many countries, giving a global overview of happiness factors. All relevant variables for the analysis were merged from publicly available sources.

Data Preparation

In the data preparation process, the data was not clean, not consistent, and it was not ready for analysis, but we did several essential steps to prepare the data.

First, we loaded the CSV files from the World Happiness Report website in Google Colab. There were two separate CSV files, so we merged them into one dataset. We could clean up and analyse the data more effectively with this merger.

```
import pandas as pd

# Load the data
data_figure = pd.read_excel('/content/drive/MyDrive/DV/DataForFigure2.1+with+sub+bars+2024.xls')
data_table = pd.read_excel('/content/drive/MyDrive/DV/DataForTable2.1.xls')

# Merge the datasets
merged_data = pd.merge(data_figure, data_table, on=['Country name'], how='inner')
```

Managing missing values was one of the central things we tackled. In the process of merging, we mainly took pains to ensure that all missing values in numeric columns had zero. When dealing with categorical data, missing entries were labelled 'unknown' to avoid losing valuable information while doing a more simplified analysis.

```
# 1. Handle Missing Values
numeric_cols = ['Ladder score', 'upperwhisker', 'lowerwhisker', 'Explained by: Log 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',
                'Life Ladder', 'Log GDP per capita', 'Social support',
                'Healthy life expectancy at birth', 'Freedom to make life choices',
                'Generosity', 'Perceptions of corruption', 'Positive affect',
                'Negative affect']

for col in numeric_cols:
    merged_data[col] = merged_data[col].fillna(0)

for col in merged_data.select_dtypes(include=['object']).columns:
    if col not in numeric_cols:
        merged_data[col] = merged_data[col].fillna('Unknown')
```

Once we had dealt with missing values, we should have confirmed that each column contained the correct data type. This made this a vital step since the data cannot be manipulated due to incompatible formats.

```
# 2. Ensure Data Types are Correct
for col in numeric_cols:
    try:
        merged_data[col] = pd.to_numeric(merged_data[col], errors='coerce')
    except (ValueError, TypeError):
        print(f"Could not convert column '{col}' to numeric. Keeping as string.")
```

Moreover, while merging the two datasets, we saw some columns with similar data but different header names. In order to simplify our analysis, we spliced these columns into one coherent column, leaving ease of clarity in our dataset intact.

```
# Merge Similar Columns
column_mapping = {
    'Ladder score': 'Life Ladder',
    'upperwhisker': 'upperwhisker',
    'lowerwhisker': 'lowerwhisker',
    'Explained by: Log GDP per capita': 'Log GDP per capita',
    'Explained by: Social support': 'Social support',
    'Explained by: Healthy life expectancy': 'Healthy life expectancy at birth',
    'Explained by: Freedom to make life choices': 'Freedom to make life choices',
    'Explained by: Generosity': 'Generosity',
    'Explained by: Perceptions of corruption': 'Perceptions of corruption',
    'Dystopia + residual': 'Dystopia + residual',
    'Life Ladder': 'Life Ladder',
    'Log GDP per capita': 'Log GDP per capita',
    'Social support': 'Social support',
    'Healthy life expectancy at birth': 'Healthy life expectancy at birth',
    'Freedom to make life choices': 'Freedom to make life choices',
    'Generosity': 'Generosity',
    'Perceptions of corruption': 'Perceptions of corruption',
    'Positive affect': 'Positive affect',
    'Negative affect': 'Negative affect'
}

for original_col, new_col in column_mapping.items():
    if original_col != new_col and original_col in merged_data.columns and new_col in merged_data.columns:
        merged_data[new_col] = merged_data[[original_col, new_col]].mean(axis=1)
        merged_data = merged_data.drop(columns=[original_col])
```

We reordered the columns to improve readability and the 'year' column in the second row. It was easier to analyse trends over time with other variables.

```
3. Move 'Year' column to the second position
cols = list(merged_data.columns)
cols.insert(1, cols.pop(cols.index('year')))
merged_data = merged_data[cols]
```

Recognising that our dataset contained information for over 100 countries, we filtered the data to focus on a specific set of countries: India, Sri Lanka, South Africa, Germany, the United Kingdom, Mexico, the United States, New Zealand, Saudi Arabia, and the United Arab Emirates. With this targeted approach, we could focus on the countries of most interest.

```
# Filter for specific countries
countries_to_keep = ['Sri Lanka', 'India', 'South Africa', 'Germany', 'United Kingdom', 'Mexico',
                    'United States', 'New Zealand', 'Saudi Arabia', 'United Arab Emirates']
filtered_data = merged_data[merged_data['Country name'].isin(countries_to_keep)]
```

Finally, we stored the filtered dataset into a new csv file as it will be our base for further analysis.

```
# Save the filtered data
filtered_data.to_excel('/content/drive/MyDrive/DV/happiness_data_filtered.xlsx', index=False)
print("The cleaned, merged, and filtered data has been")
```

The cleaned, merged, and filtered data has been

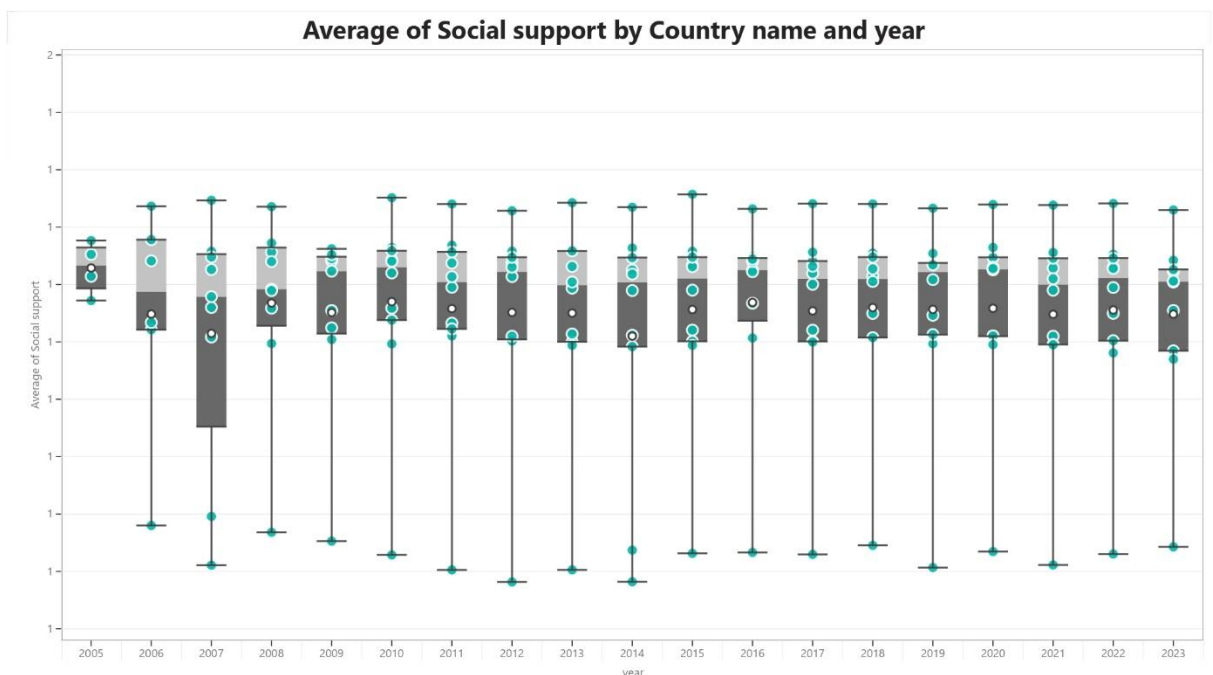
Exploratory Data Analysis

Exploratory Data Analysis (EDA) is when we analyse data in a visual sense when datasets exist and can find important characteristics through visual methods.

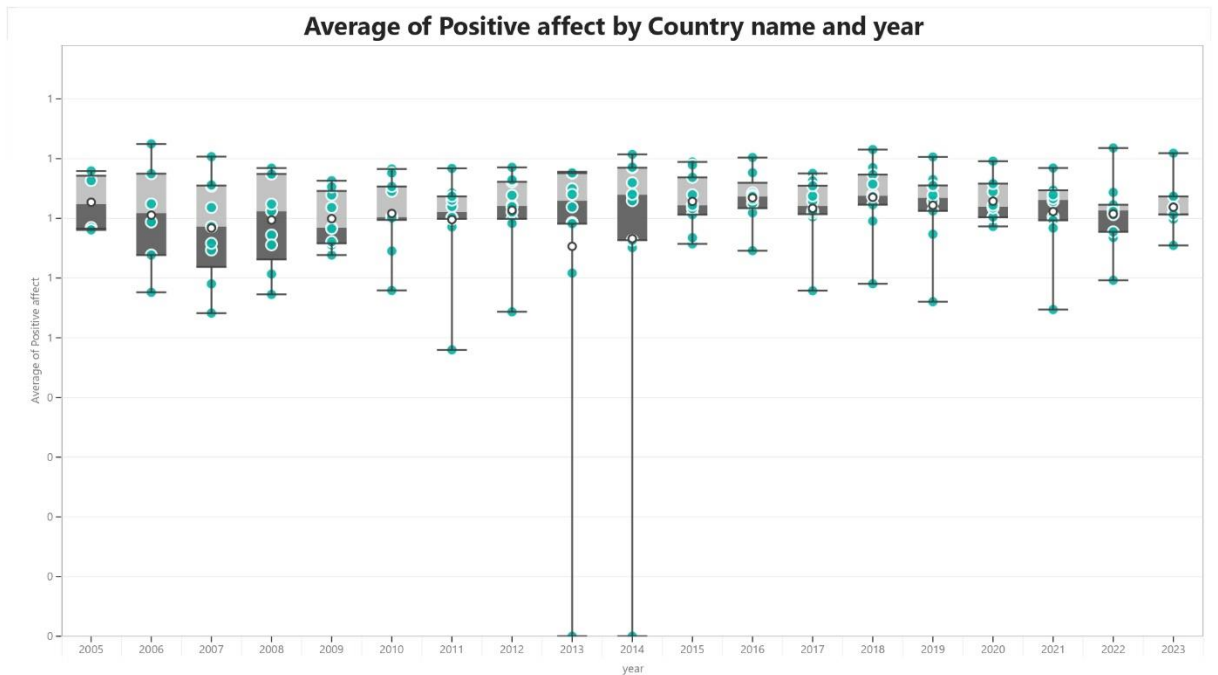
Univariate analysis involves taking out all the variables in statistical analysis except one at a time.

We created box plots for key metrics to reveal insights:

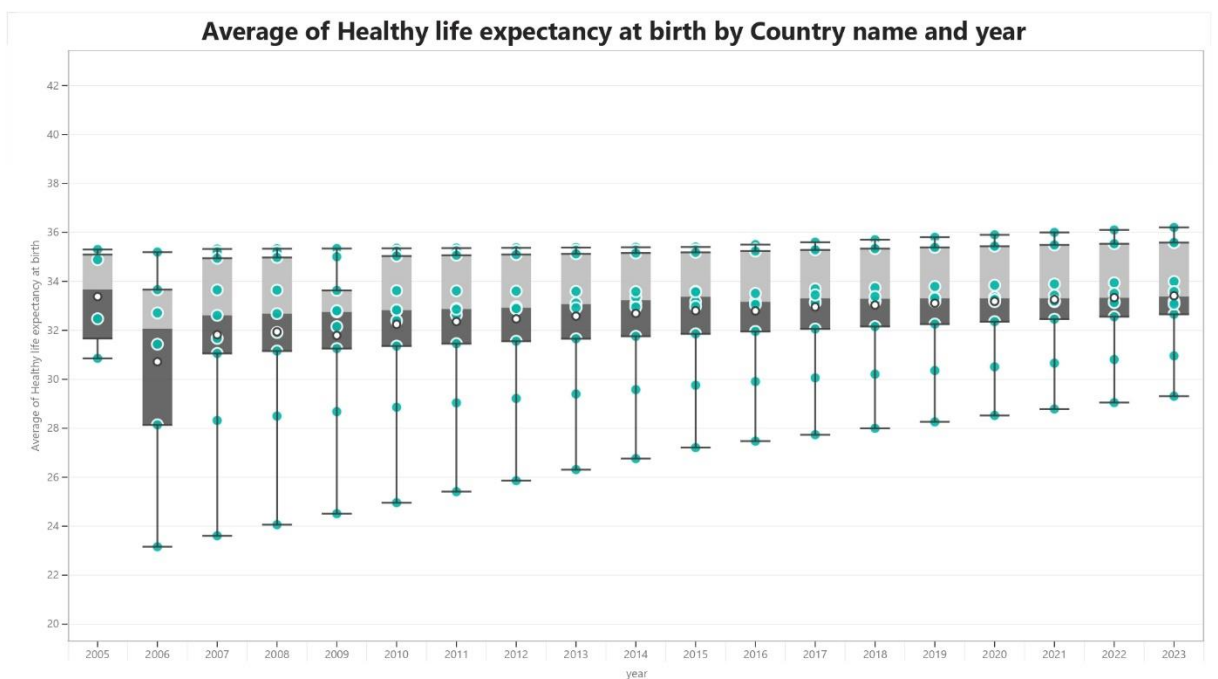
1. Average Social Support: The distribution of social support scores by country and year indicates that the support changes over time.



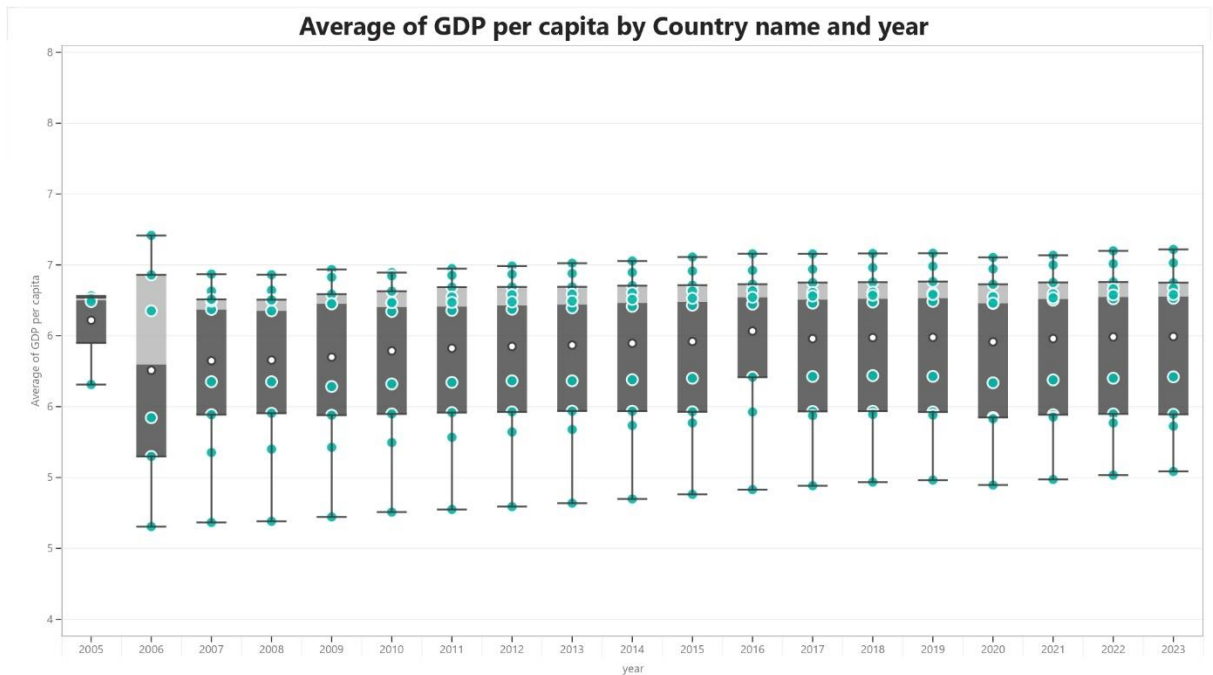
2. Average Positive Affect: It visualises the variation of positive emotions from country to year, the emotional states per region.



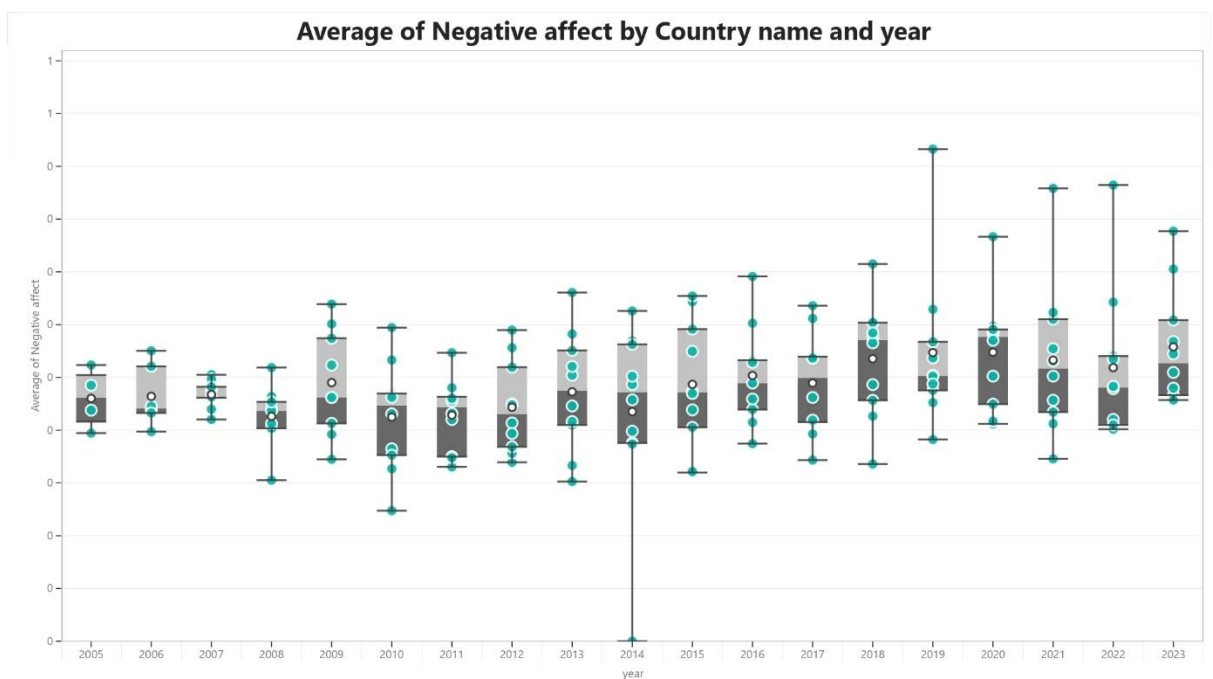
3. **Average Healthy Life Expectancy:** This plot carries healthy information on life expectancy differences by country over time, with differences in health.



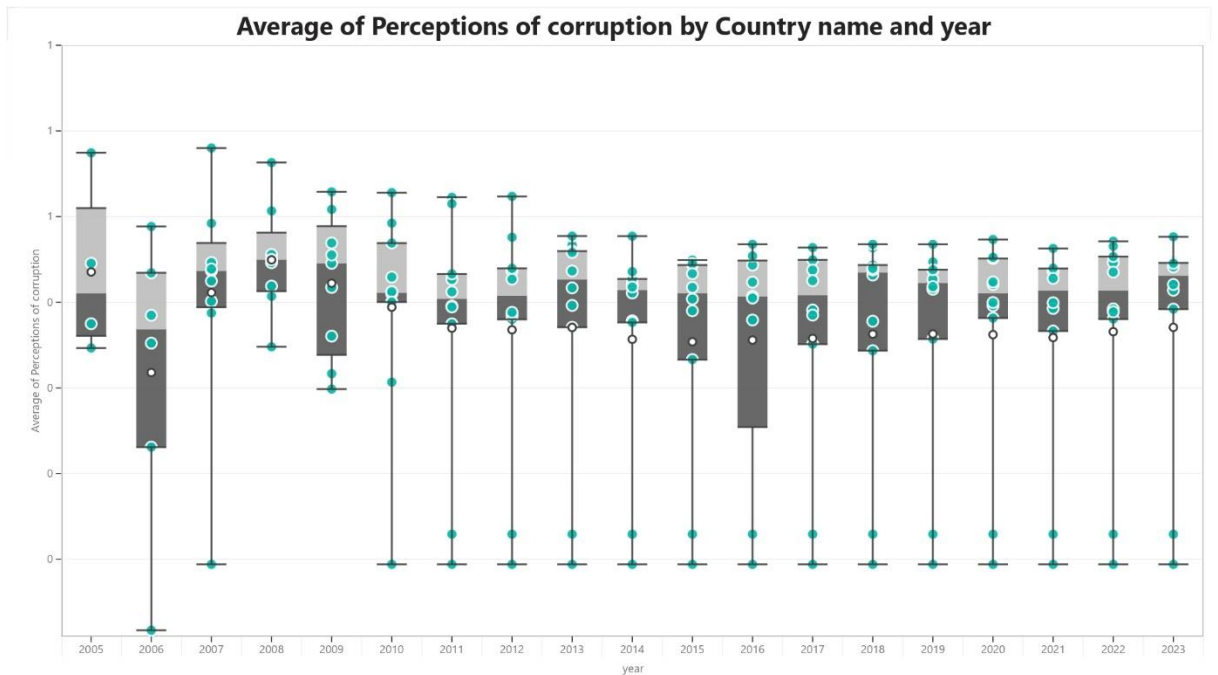
4. **Average GDP per Capita:** This representation shows how economically wealthy certain countries are.



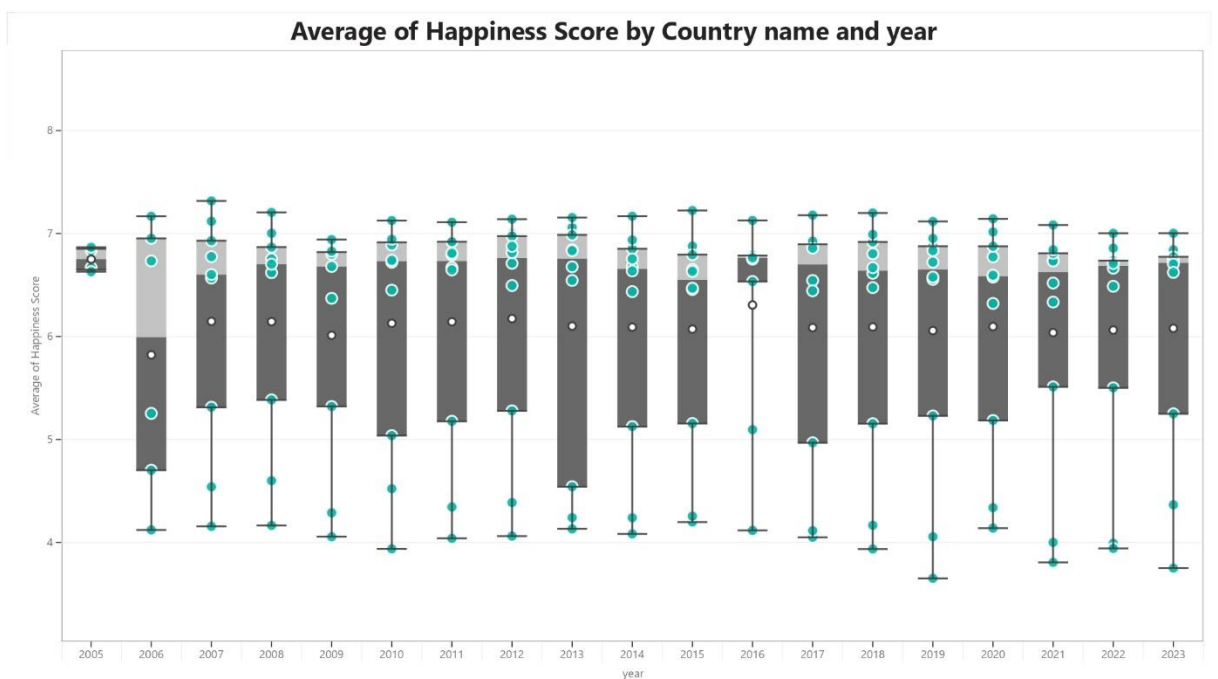
5. **Average Negative Affect:** The box plot shows negative emotions, which helps understand how often one feels not so good.



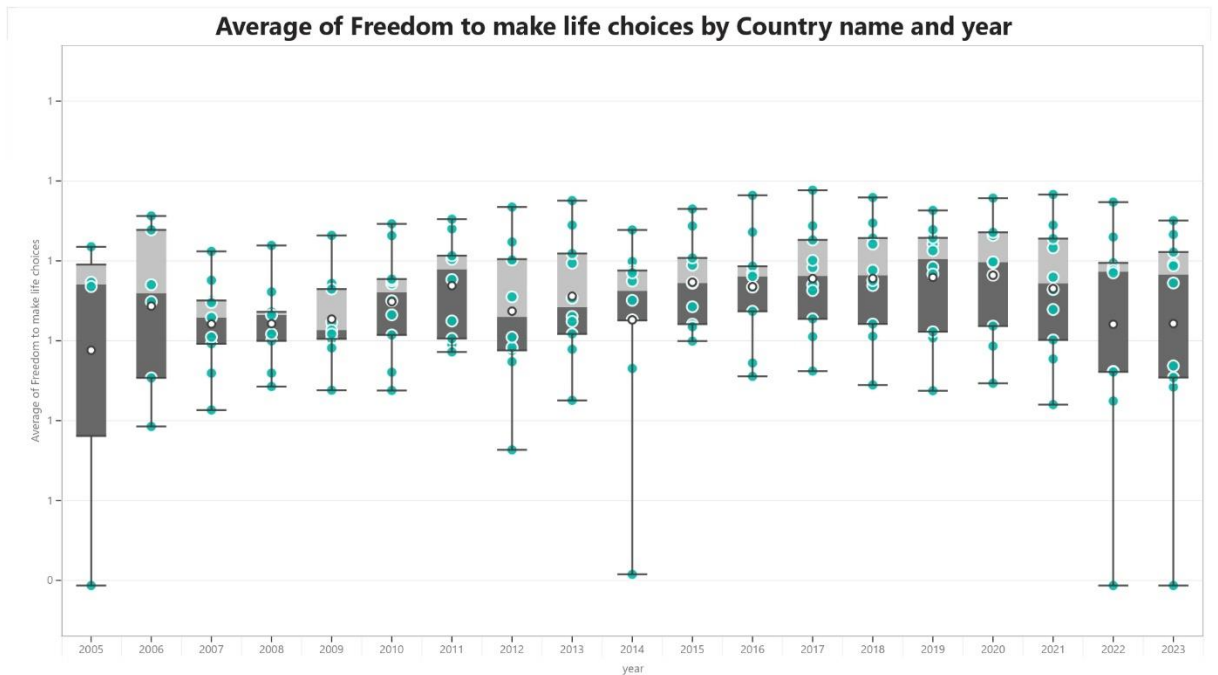
6. **Average Perceptions of Corruption:** Looking at the perception of global corruption, this plot helps us analyse the state of governance and integrity of countries.



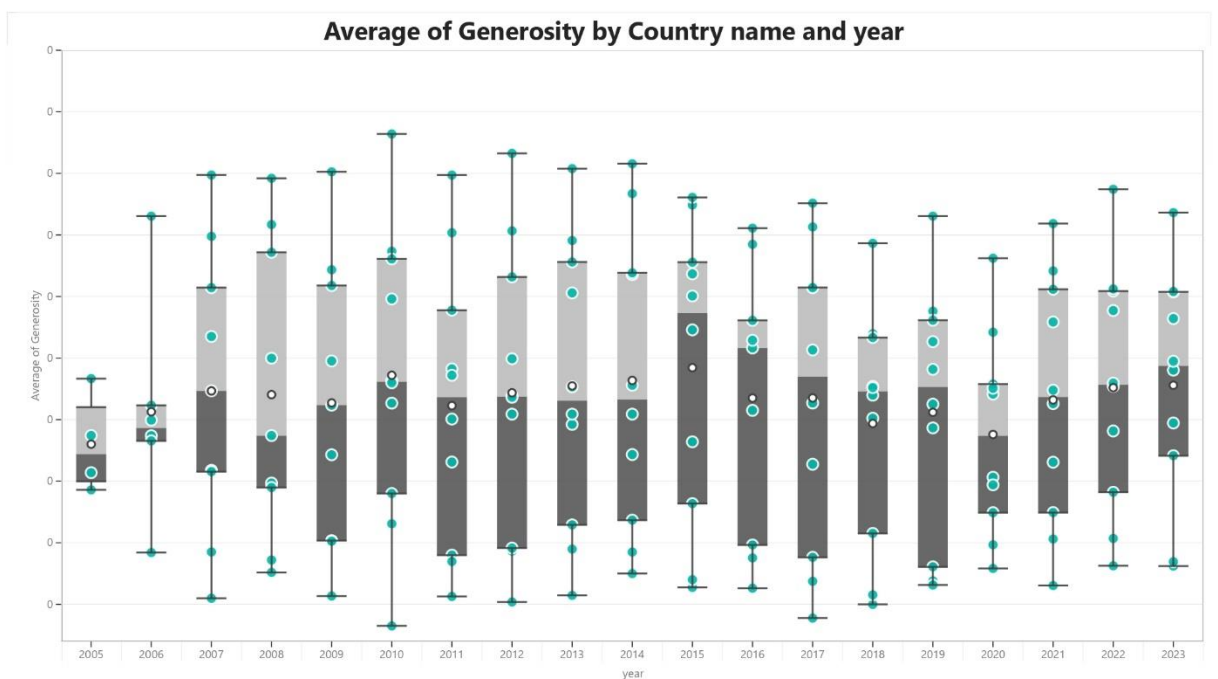
7. **Average Happiness Score:** This plot illustrates happiness scores, reflecting how the world views life as satisfying and well.



8. **Average Freedom to Make Life Choices:** It is known that happiness depends on personal freedom, and this chart demonstrates that.

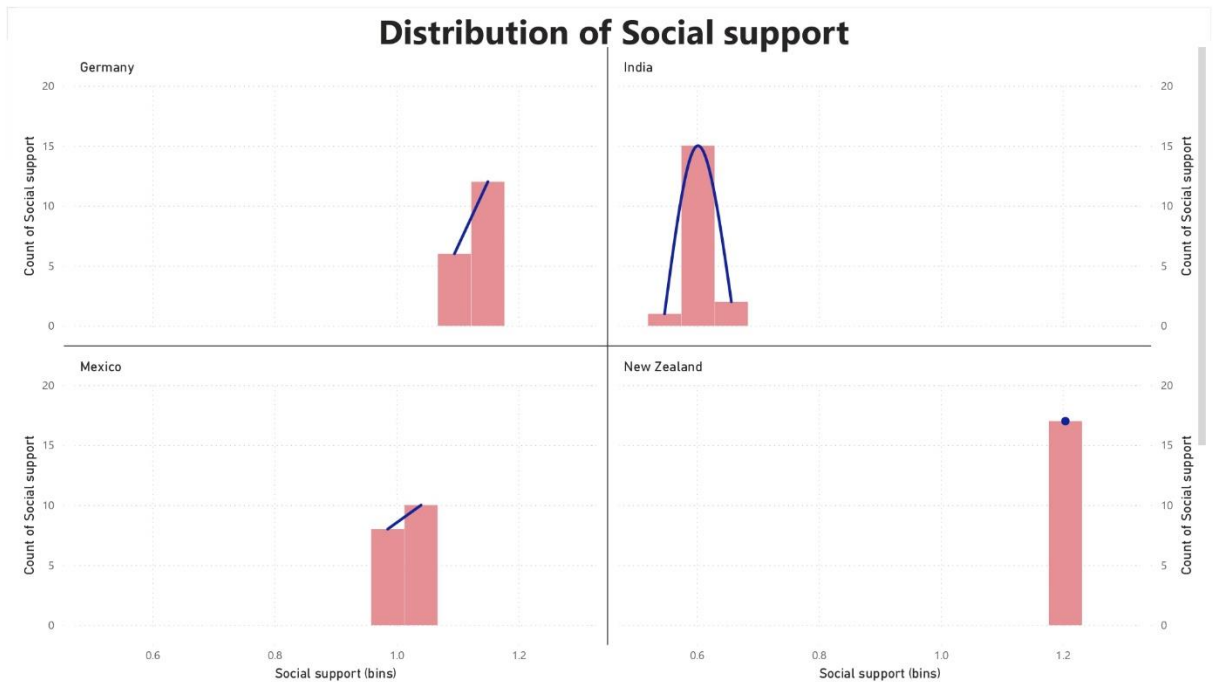


9. Average Generosity: Generosity scores are represented on this box plot, showing change with time and across regions.

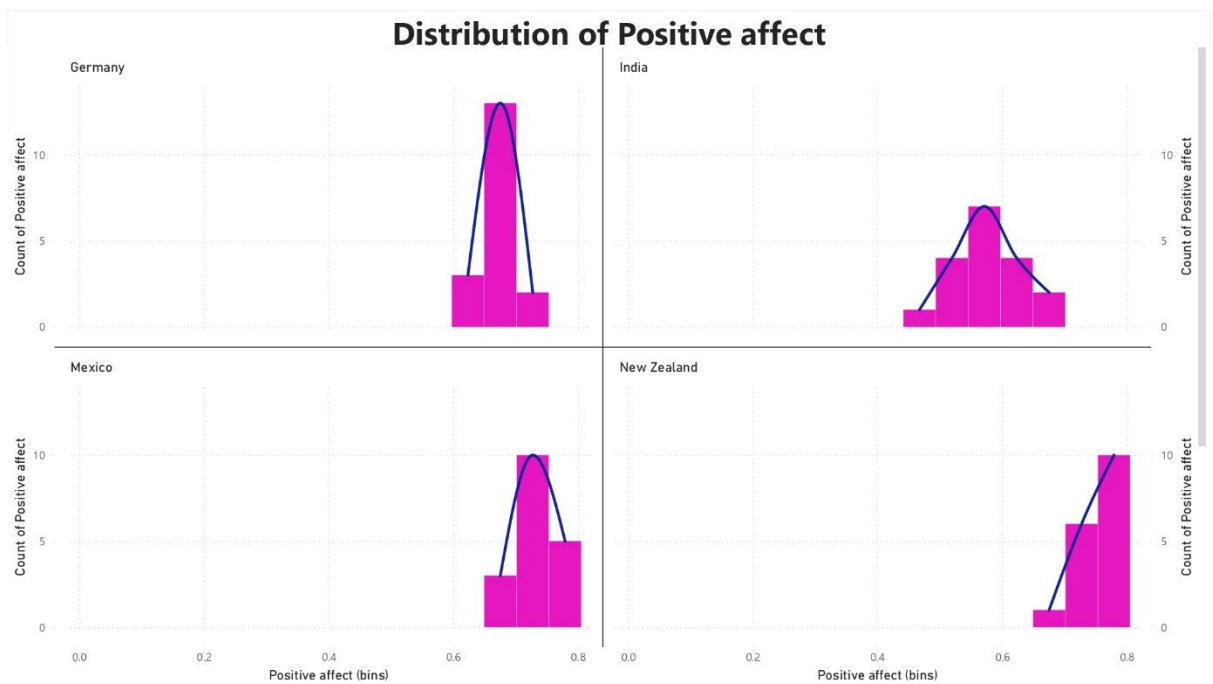


Histograms for our understanding of distributions:

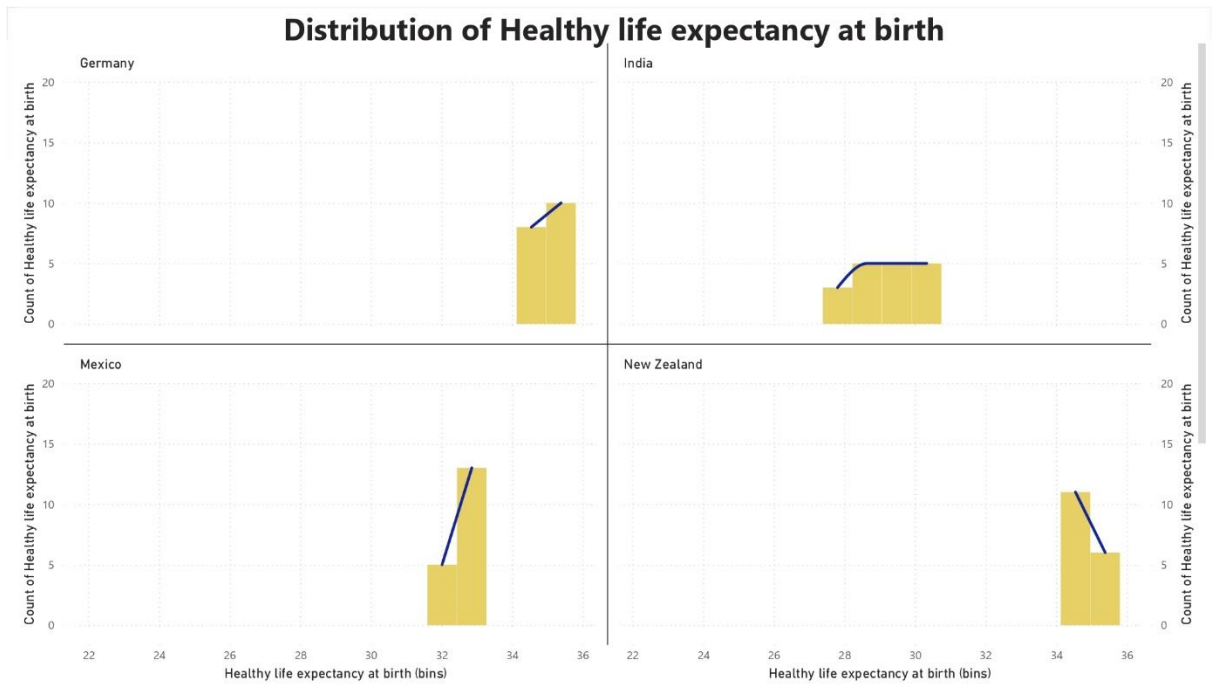
1. Social Support Distribution: Counties that have social support around them



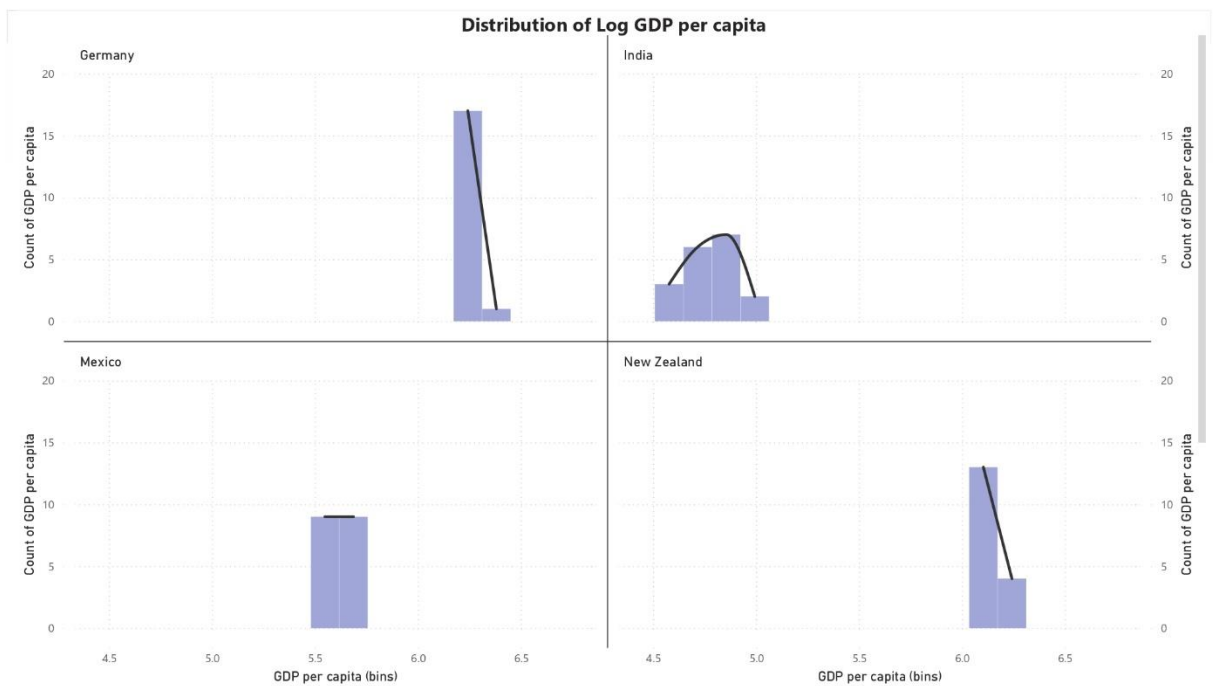
2. Positive Affect Distribution: Shows how frequently people experience positive emotions.



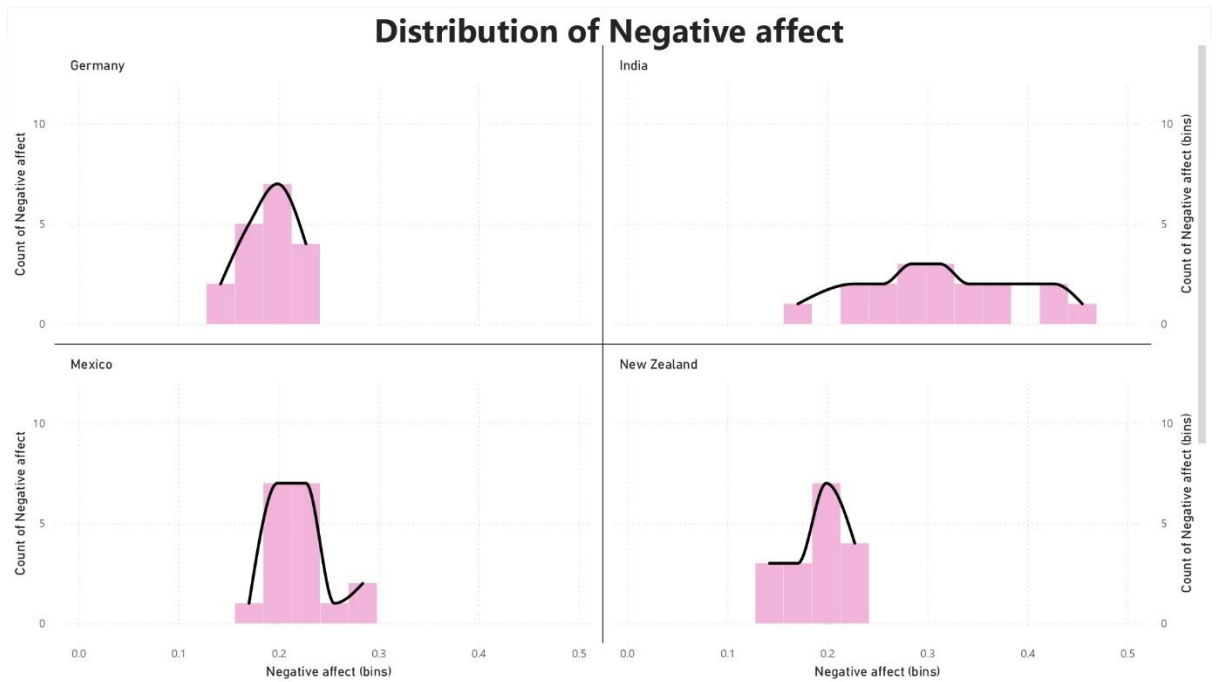
3. Healthy Life Expectancy Distribution: It captures variations in healthy life expectancy and shows health disparities.



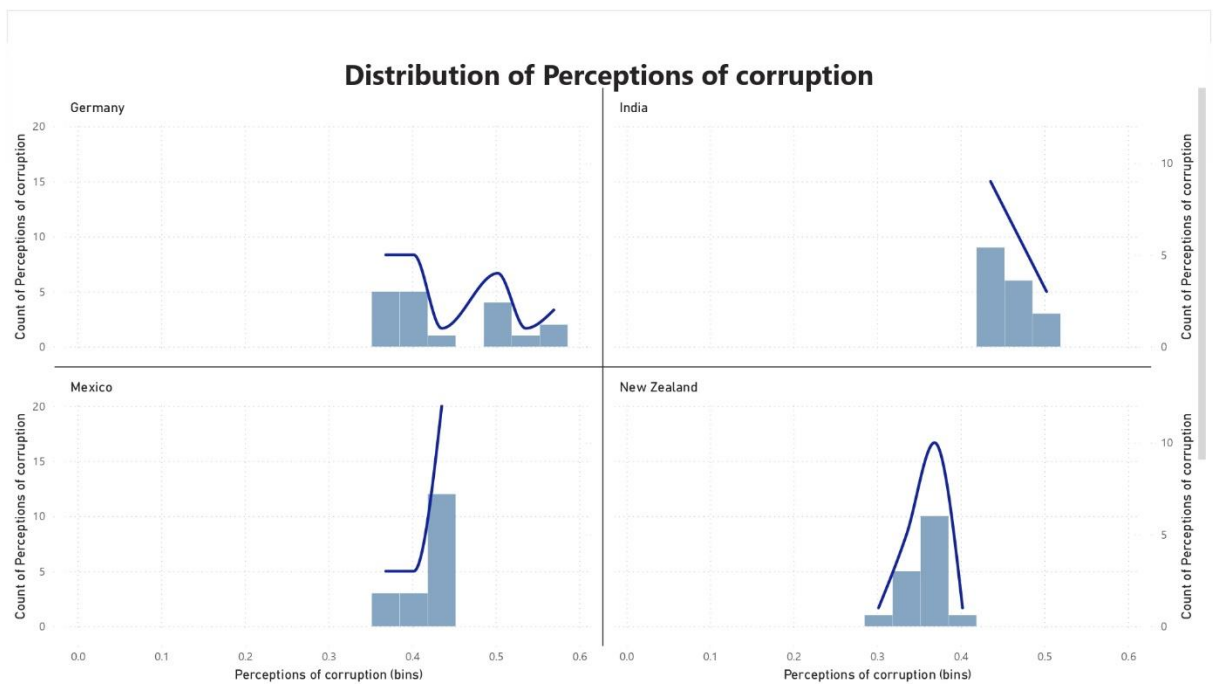
4. GDP per Capita Distribution: It helps to identify the economically prosperous and less prosperous countries.



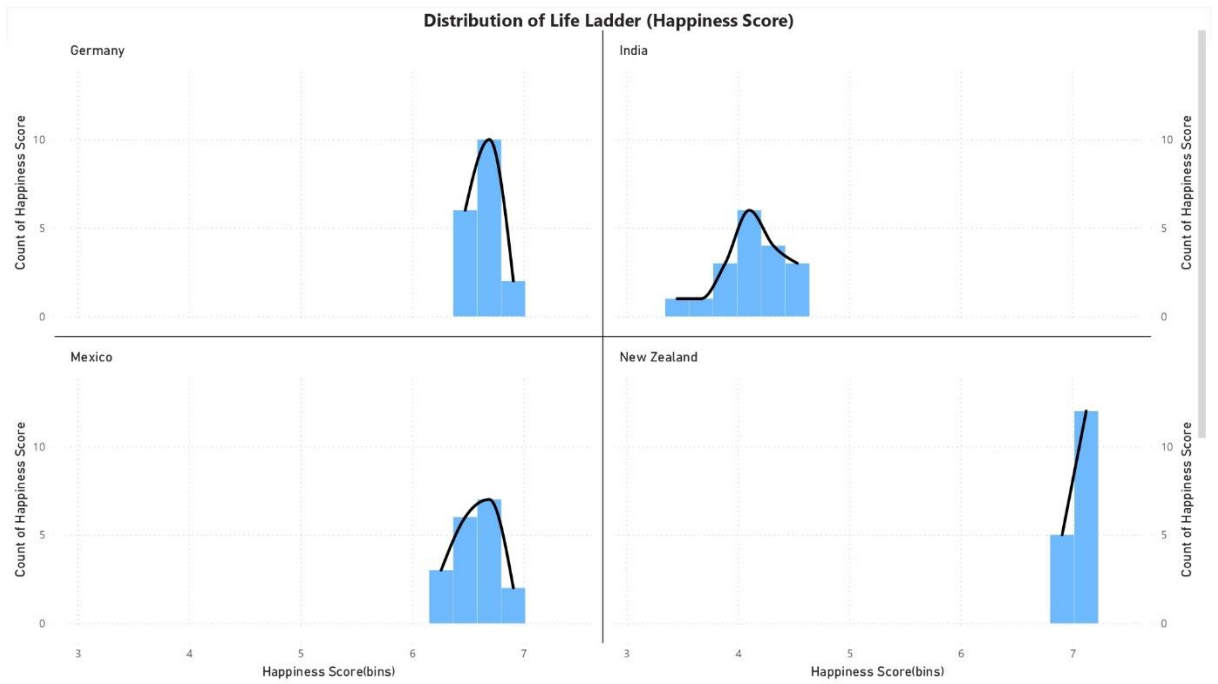
5. Negative Affect Distribution: This shows that negative emotions are pervasive among societies.



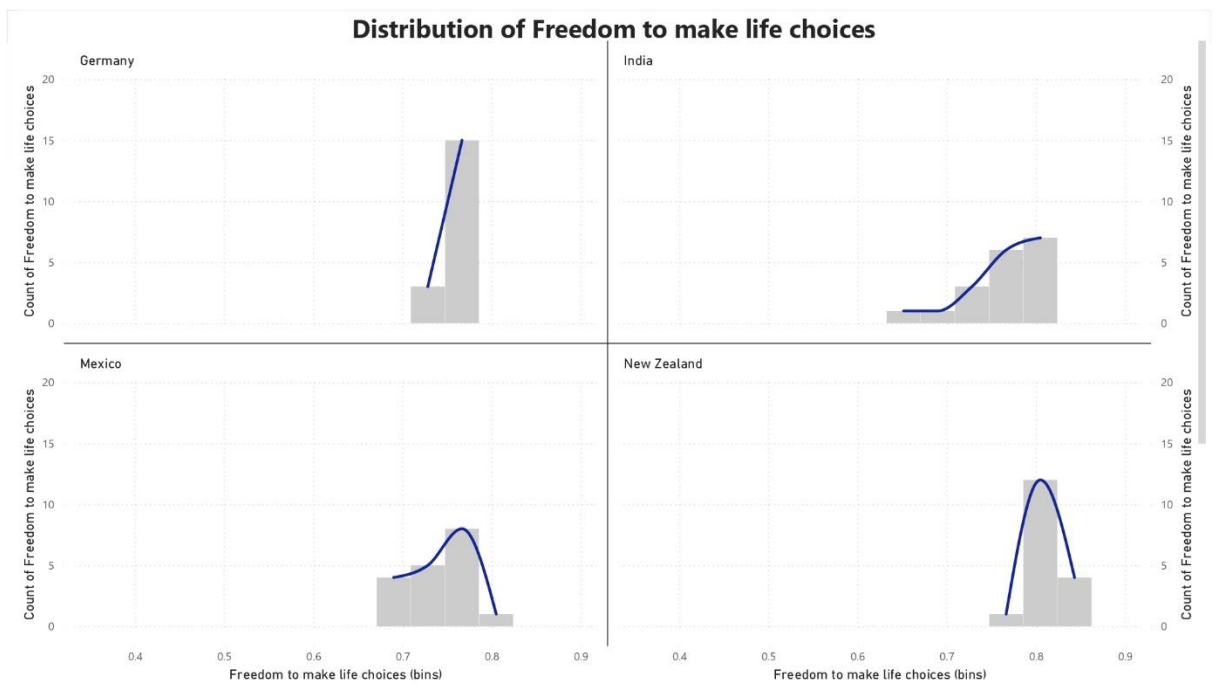
6. Corruption Perception Distribution: It shows how different countries perceive corruption.



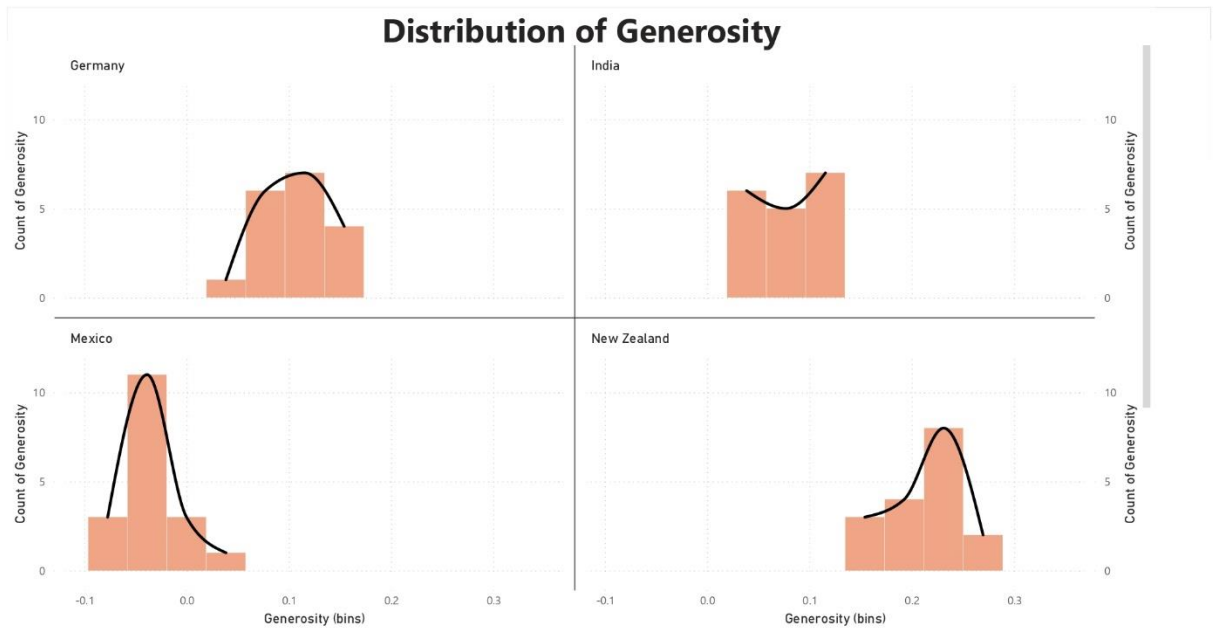
7. Happiness Score Distribution: It displays happiness scores distributed across nations.



8. Freedom to Make Life Choices Distribution: It portrays people's variation in how they perceive their freedom.

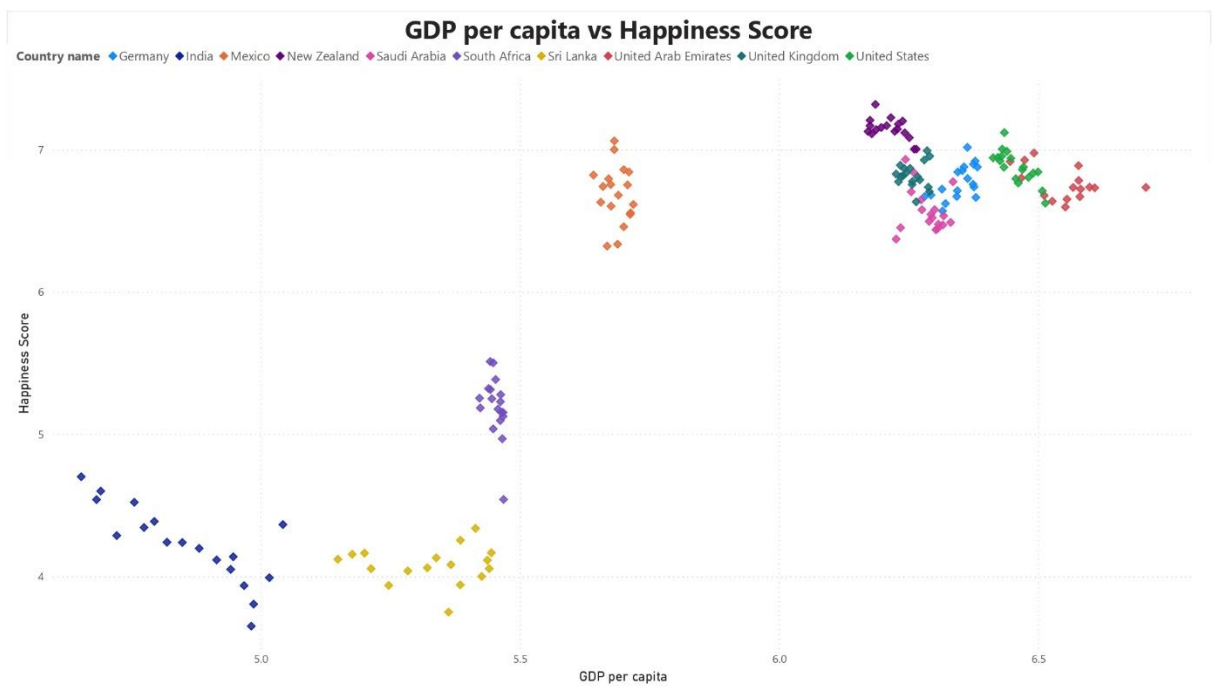


9. Generosity Distribution: Covers generosity levels of different regions.

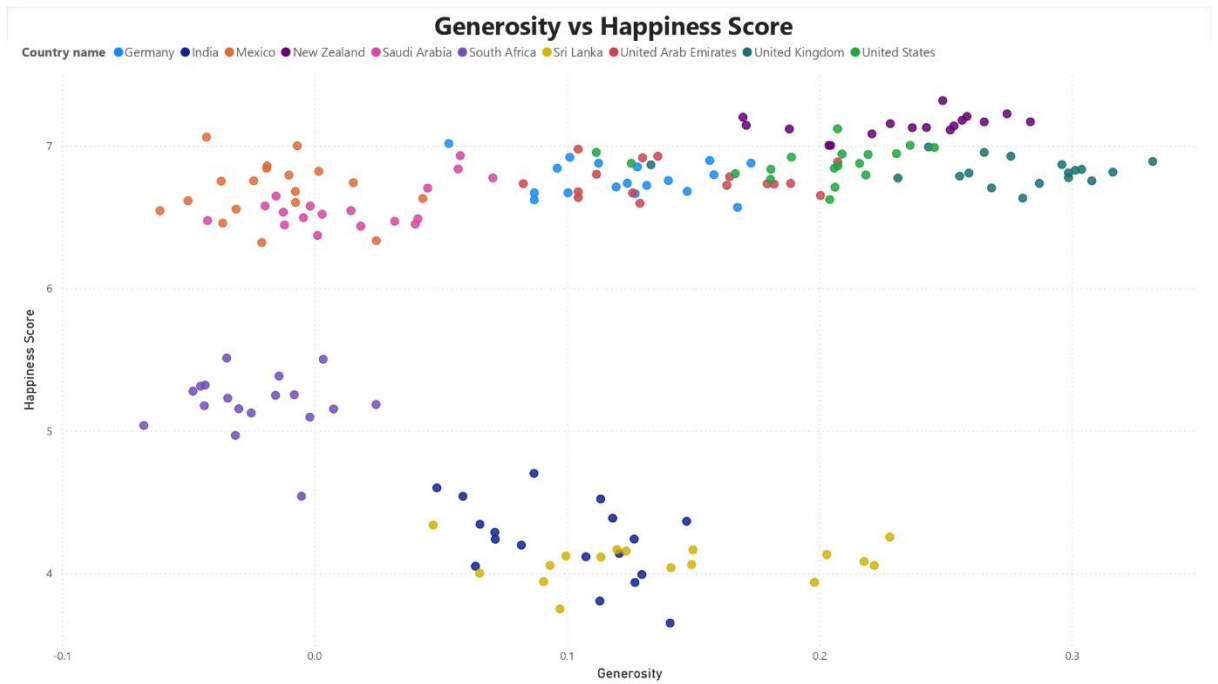


Multivariate analysis further explores relationships among multiple variables:

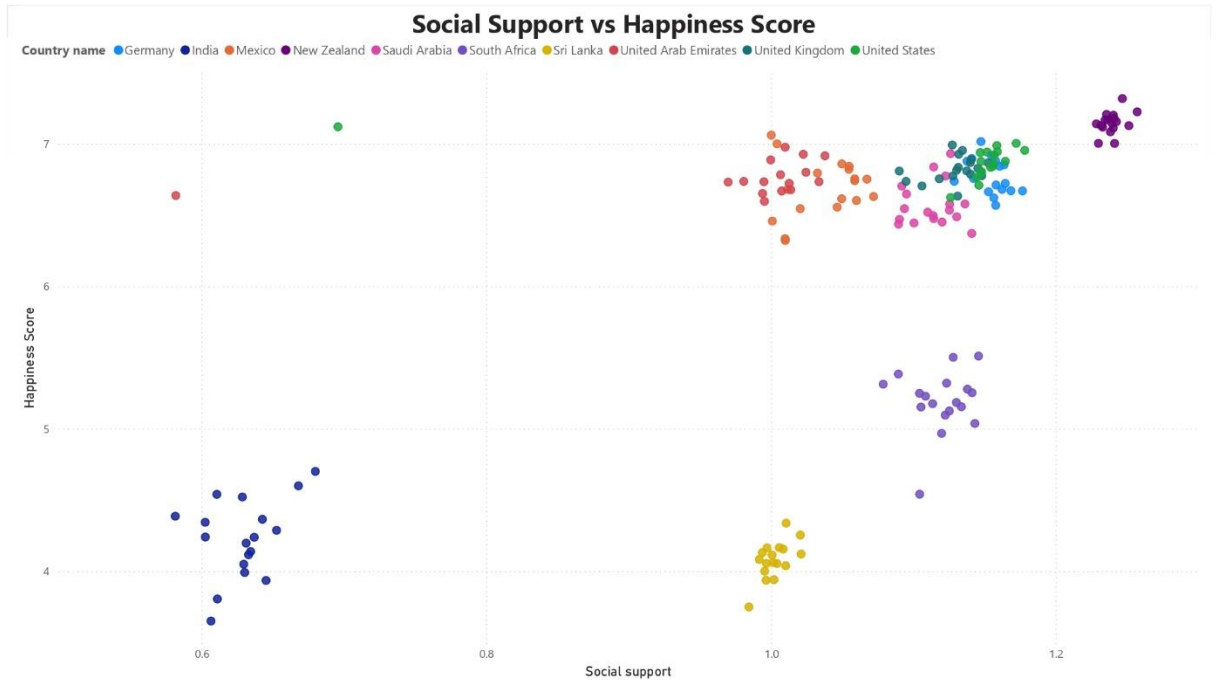
1. Happiness Score vs. GDP per Capita: A positive correlation shown on a scatter plot indicates that wealthier nations usually report higher happiness.



2. Happiness Score vs. Generosity: A positive relationship, suggesting that generosity does correlate with greater life satisfaction.



3. **Happiness Score vs. Social Support:** This is a correlation between social support and happiness, which is a strong positive correlation, which means that the more social support we have, the higher our happiness will be.



4. **Happiness Score vs. Healthy Life Expectancy:** It suggests increased happiness correlates with health outcomes.



5. Happiness Score vs. Freedom to Make Life Choices: The perceived freedom correlates with higher happiness.



6. Happiness Score vs. Corruption: Unveils the negative correlation that as corruption perceptions increase, so does unhappiness.



An analysis of these tells us what contributes to happiness in different countries.

Data Storytelling

INTRODUCTION

Have you ever wondered why people across different countries are happy or less happy?

This analysis dives into the key determinants of happiness, guided by the research question:

- How does the happiness of other countries come about?

Policymakers, researchers and organizations interested in improving the quality of life must understand these factors. When we determine these determinants, we can create better policies and programs to better people.

To do this analysis, we use the dataset from the World Happiness Report, which includes the following indicators:

- GDP per capita
- Social support
- Healthy life expectancy
- Freedom to make life choices
- Generosity

- Perceptions of corruption.

Reason?

Understanding these factors is crucial for policymakers, researchers, and organizations aiming to improve the quality of life globally. We can design better policies and programs to enhance well-being by identifying these determinants.

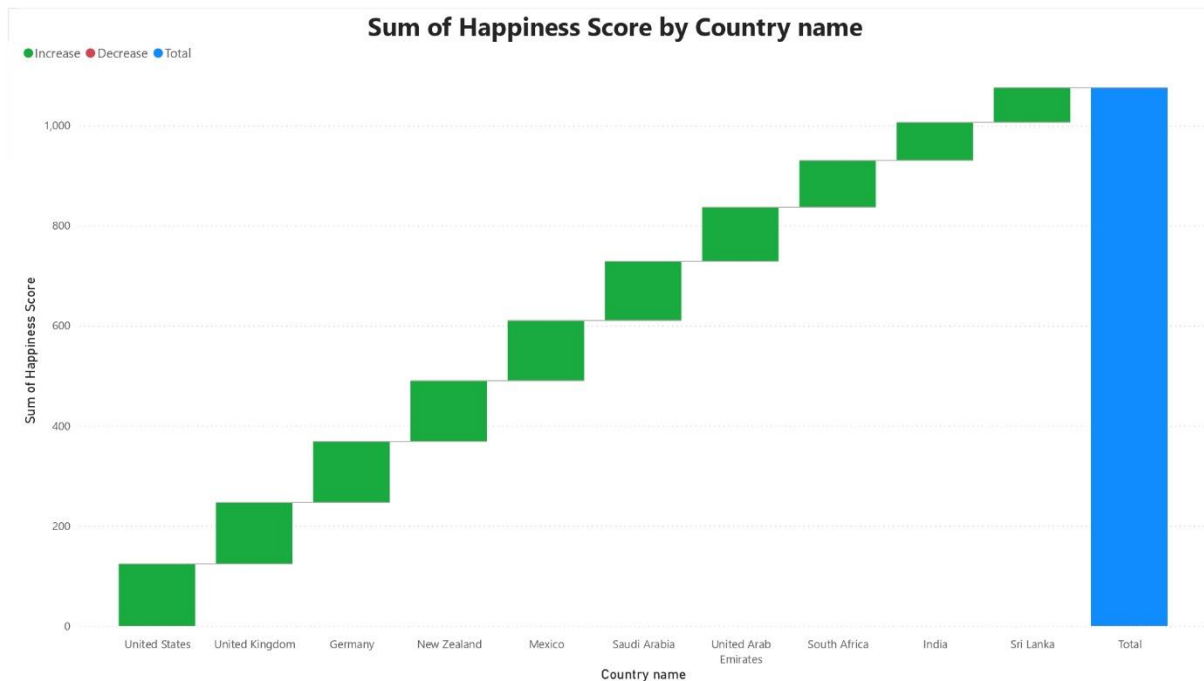
DATA OVERVIEW

For this analysis, we use the dataset from the World Happiness Report, which includes indicators such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. These indicators have been measured over several years across many countries, giving us a picture of what makes people happy.

Key Findings

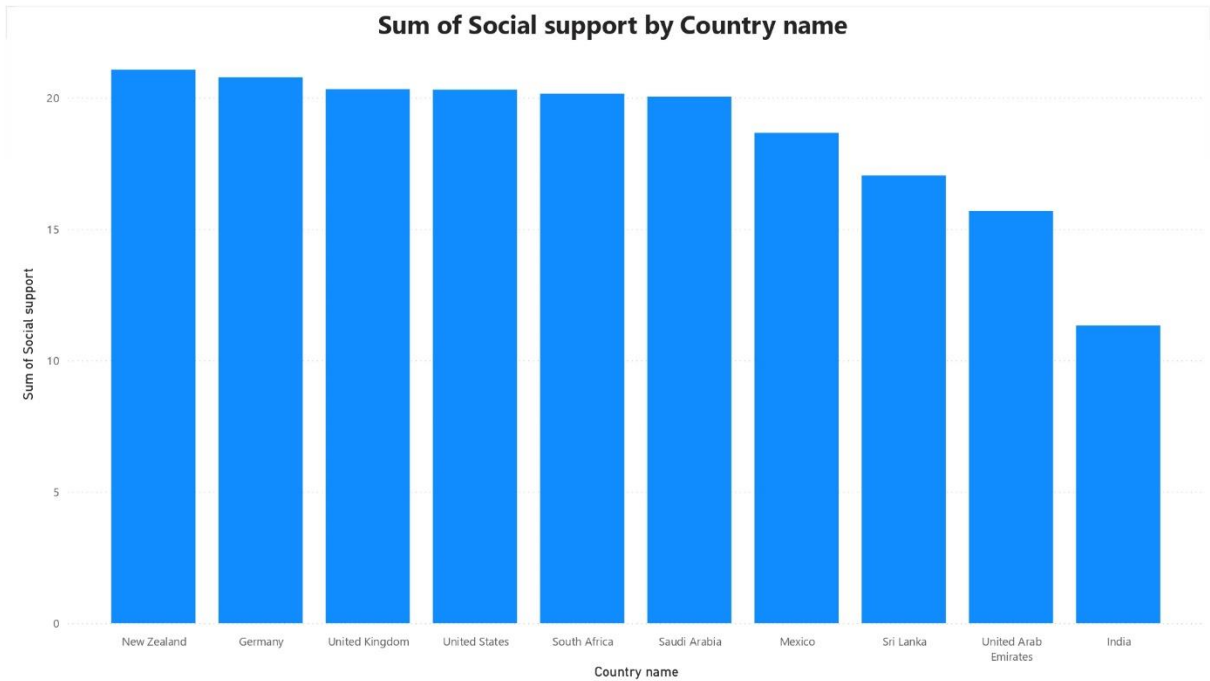
Sum of Happiness Score

Imagine a world map where New Zealand, the United Kingdom, and the United States are the brightest and most colourful countries. The cumulative happiness scores from these countries are the highest, and they're the happiest countries. This visualisation shows how these countries repeatedly report higher happiness levels.



Social Support Count by Country

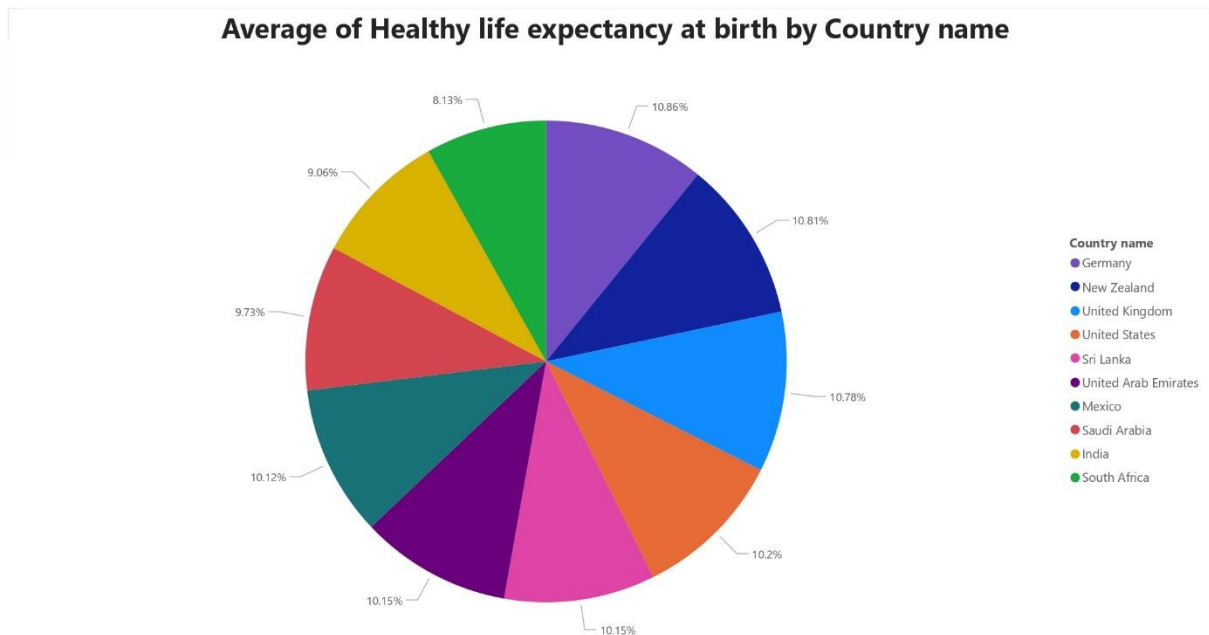
Imagine Germany and Mexico, with many social connections and community support. Countries that score higher on happiness do so with higher social support counts as well. Again, this illustrates how meaningful social connections are for joy.



Average Healthy Life Expectancy at Birth by Country

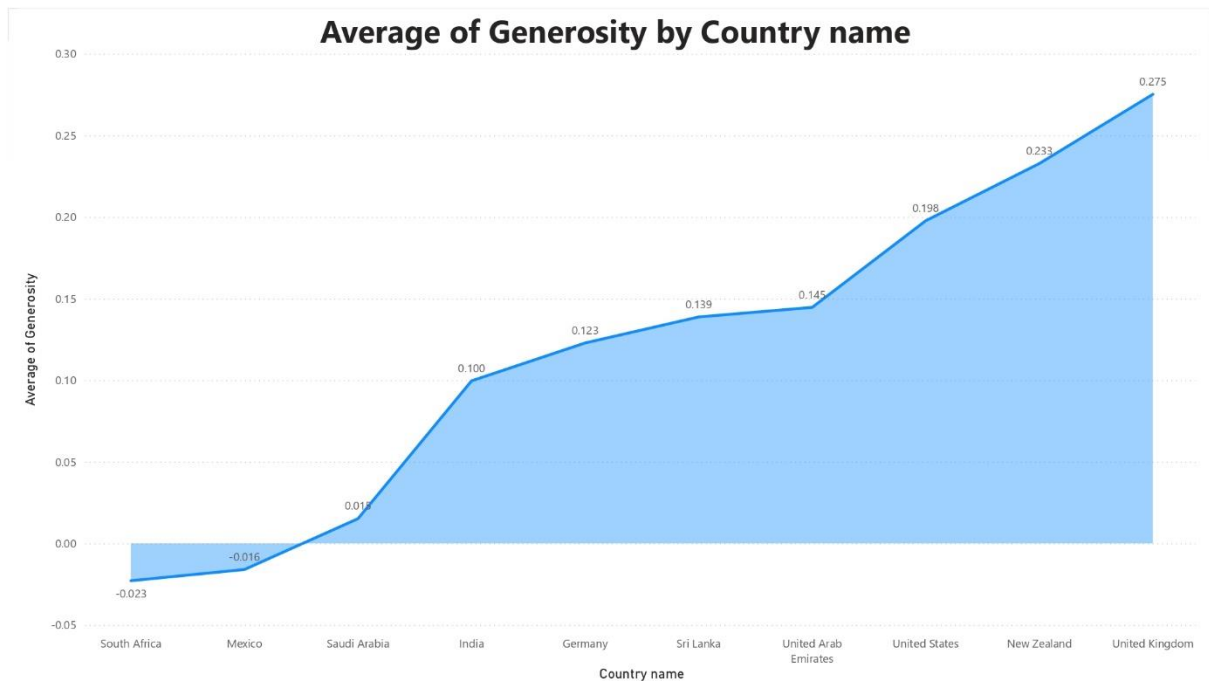
Countries like Korea, Germany, and the United Kingdom have a longer, healthier life. These nations also record above-average happiness scores, which suggests a direct

link between health and happiness.



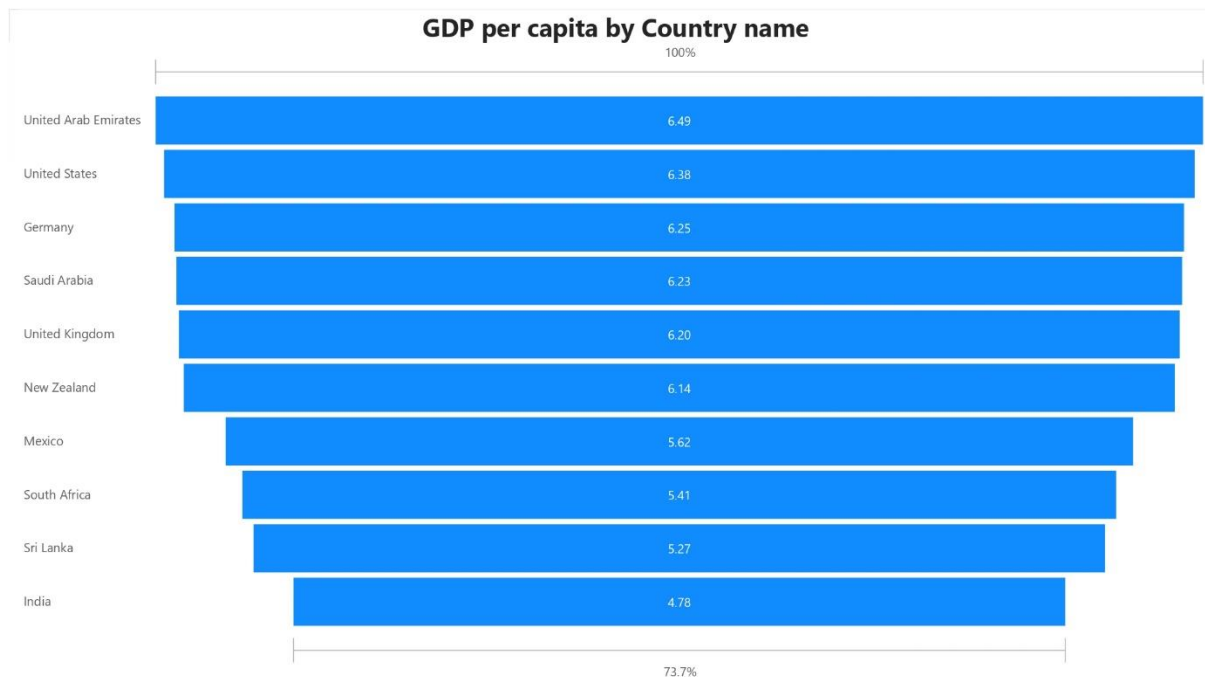
Average Generosity by Country

Or consider the United Arab Emirates and Saudi Arabia, where acts of generosity are commonplace. These countries seem prone to high rankings because of their ability to be generous and to be happy.



GDP per Capita by Country

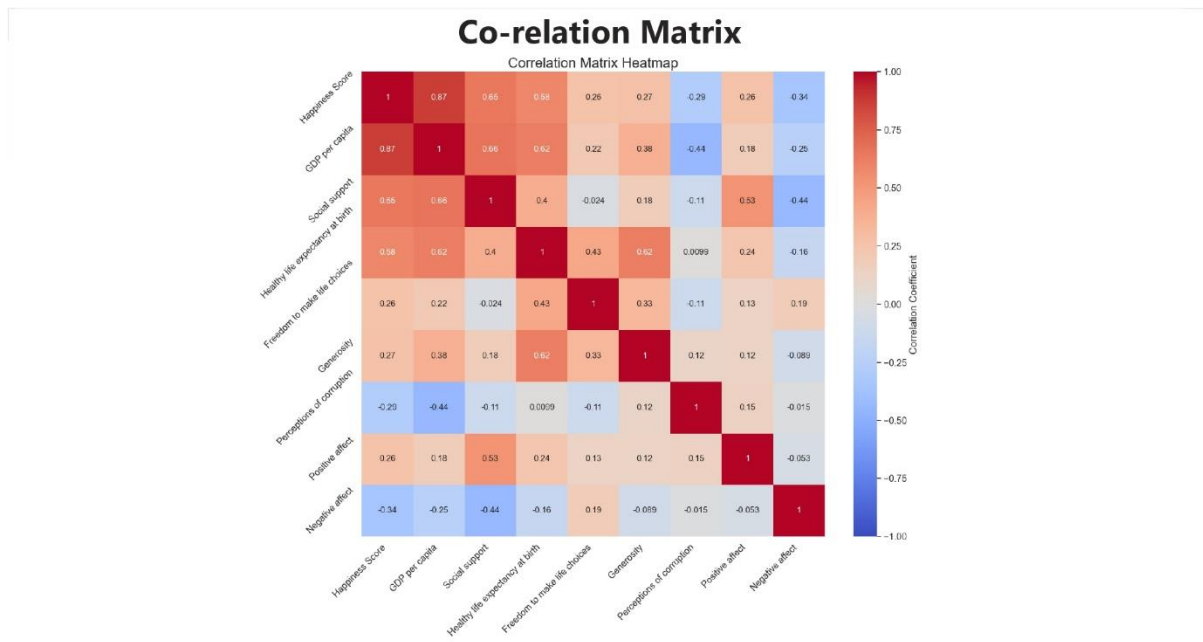
Still, happiness is not only determined by the economy: there is prosperity, but it is not the main factor. Happier are wealthier countries, such as the United States and Germany. However, there is a relationship between GDP per capita and happiness, and the GDP per capita against happiness graph suggests diminishing returns on the GDP per capita to happiness relationship beyond a certain point.



Correlation Matrix

Which factors correlate with each country: GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption? Positive correlations are found with happiness levels and these different measures: higher GDP per capita, social support, and healthy life expectancy, and negative

correlation with happiness levels and perceptions of corruption.



Conclusion

Lastly, I agree that no one factor can predict a notion of happiness for a population inside a particular group. Other factors leading to happiness are lifespan, GDP per capita, the quality of the social support system, personal freedom in life choices, levels of generosity and perceptions of corruption. These factors make life good for people and communities.

By using visualisations, we can better understand how each factor influences total life satisfaction and how they affect one another. For example, we can learn about the correlation between GDP per capita and levels of social support since, for example, the correlation between economic security and personal well-being can be found in this.

Understanding these relationships allows policymakers and organisations to shape targeted interventions to promote greater happiness — and broader well-being — on a global scale. An all-inclusive-inclusive approach can make positive differences by making policy changes that contribute to social equality, by writing programmes that encourage social interaction or by developing an environment that supports developing personal choice. In the end, finding our way to more happiness promotes a more harmonious society — which, after all, is what it boils down to.