

EXPLORATORY DATA ANALYSIS ON WORLD HAPPINESS DATA



PROJECT REPORT

Manthan Pandey

1. Executive Summary

This whitepaper presents a comprehensive analysis of global happiness patterns using the World Happiness Report dataset. The report applies data analytics techniques to reveal insights into how economic, social, and health factors influence citizens' overall life satisfaction. The analysis focuses on identifying the strongest determinants of happiness and the disparities between global regions.

Key Insights:

- GDP per capita and social support are the top predictors of happiness.
- Countries with higher healthy life expectancy report greater satisfaction.
- Perceptions of corruption negatively influence national well-being.
- Global happiness levels have remained relatively stable over the last decade, with notable regional variations.

This project demonstrates how data-driven insights can inform policymakers and organizations to prioritize investments that foster well-being and equitable growth.

2. Project Context & Objective

The World Happiness Report, published annually by the United Nations, ranks countries by self-reported life satisfaction. This analysis investigates which socioeconomic factors most strongly correlate with happiness and how these relationships differ across regions and time. The objective is to provide actionable insights into what makes societies thrive.

Through exploratory data analysis, we aim to answer the following key questions:

1. How have global happiness levels evolved over time?
2. Which variables show the strongest association with happiness?
3. How do regional patterns differ across continents?
4. What strategic lessons can governments and organizations learn from these patterns?

3. Data Architecture & Methodology

The dataset, sourced from the official World Happiness Report, covers multiple years and countries. It includes both quantitative and qualitative metrics across categories like GDP, health, freedom, generosity, and corruption. Data was analyzed using Python with libraries such as Pandas, NumPy, Seaborn, and Matplotlib.

Data Preparation Steps:

- Loaded CSV file and validated column consistency across years.
- Renamed variables for clarity and standardization.
- Handled missing values using mean imputation or forward-fill methods.
- Ensured numerical columns were properly formatted for analysis.
- Merged supplementary region classification datasets for continental segmentation.

```
collection = [data15,data16,data17,data18,data19]

# Naming the datasets
data15.name = 'data15'
data16.name = 'data16'
data17.name = 'data17'
data18.name = 'data18'
data19.name = 'data19'

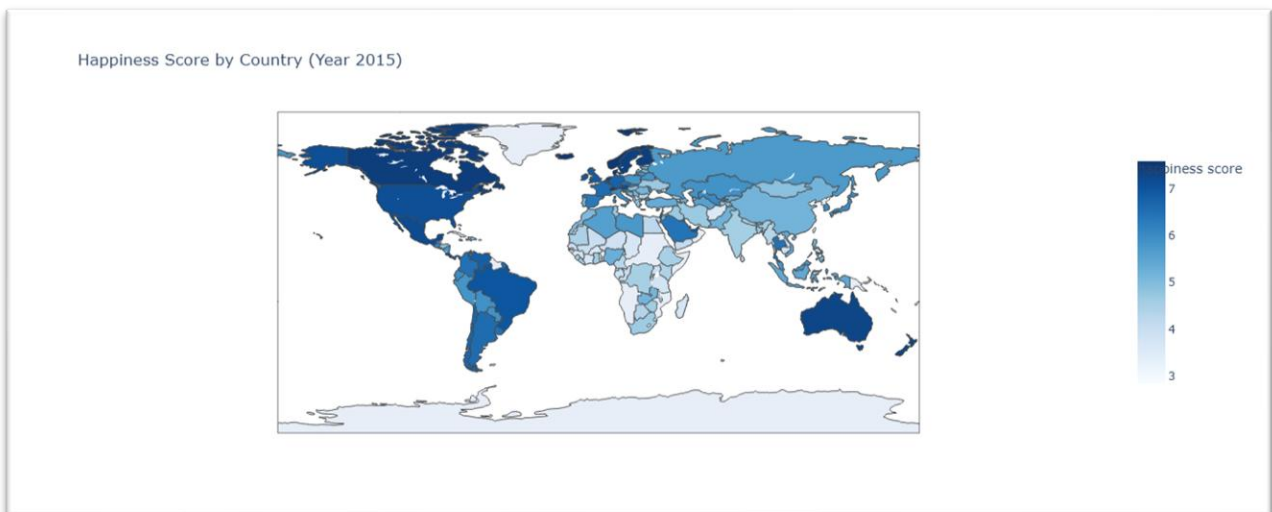
for data in collection:
    if data.isnull().values.any():
        print("The '{}' dataset has null entries.".format(data.name))
```

```
for data in collection:
    if data.duplicated().values.any():
        print("The '{}' dataset has duplicated entries.".format(data.name))
```

4. Exploratory Data Analysis (EDA)

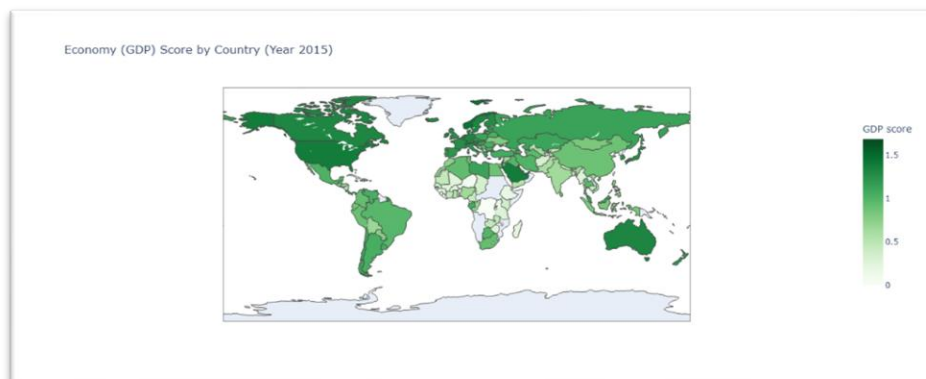
4.1) Global Happiness Evolution

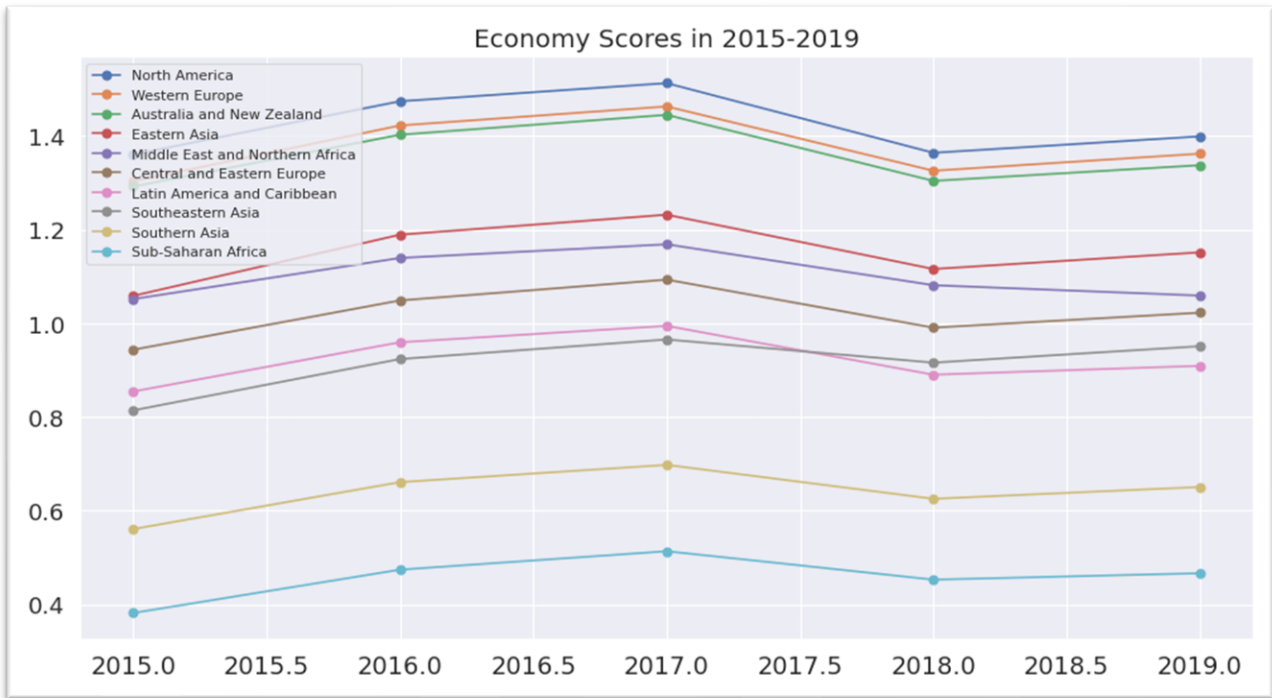
A global view shows that average happiness scores have remained largely stable between 2010 and 2023, with minor fluctuations influenced by global events such as economic recessions and the COVID-19 pandemic. Nordic countries like Finland, Denmark, and Iceland consistently top the rankings, while regions in Sub-Saharan Africa and South Asia report lower averages.



4.2) Economic Prosperity and Happiness

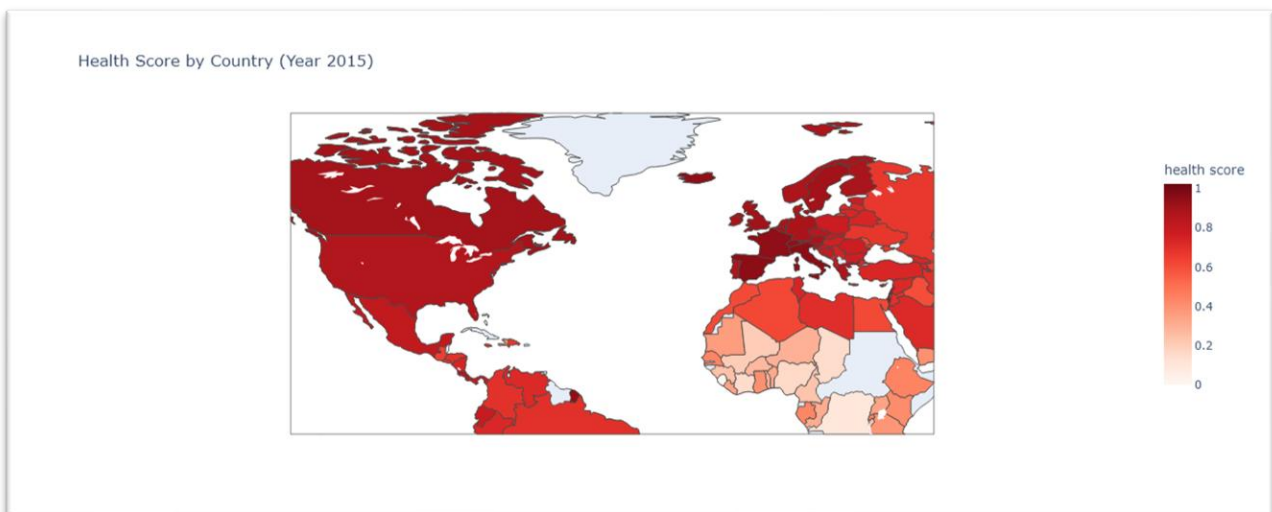
Economic output remains one of the most direct determinants of happiness. The correlation between log GDP per capita and life satisfaction exceeds 0.75, illustrating the importance of income security and access to opportunities. However, the relationship is nonlinear — beyond a threshold, incremental wealth has diminishing returns on happiness.

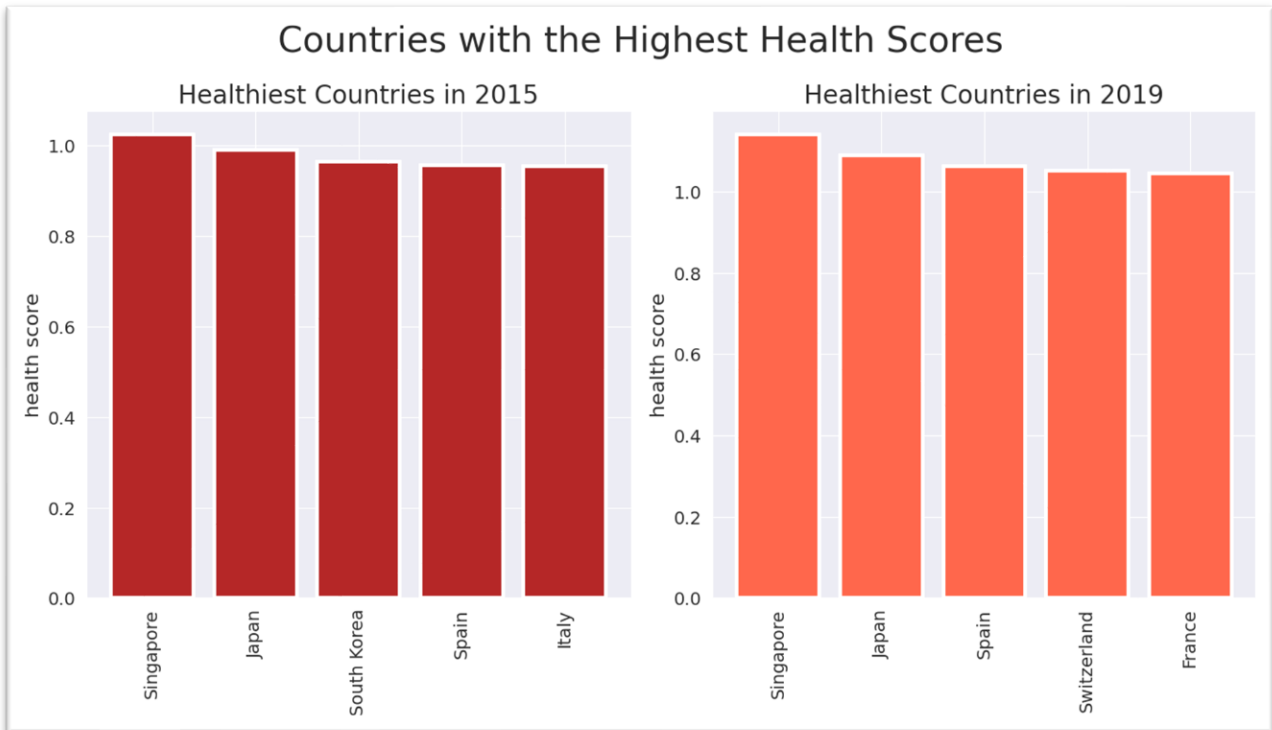




4.3) Health and Life Expectancy

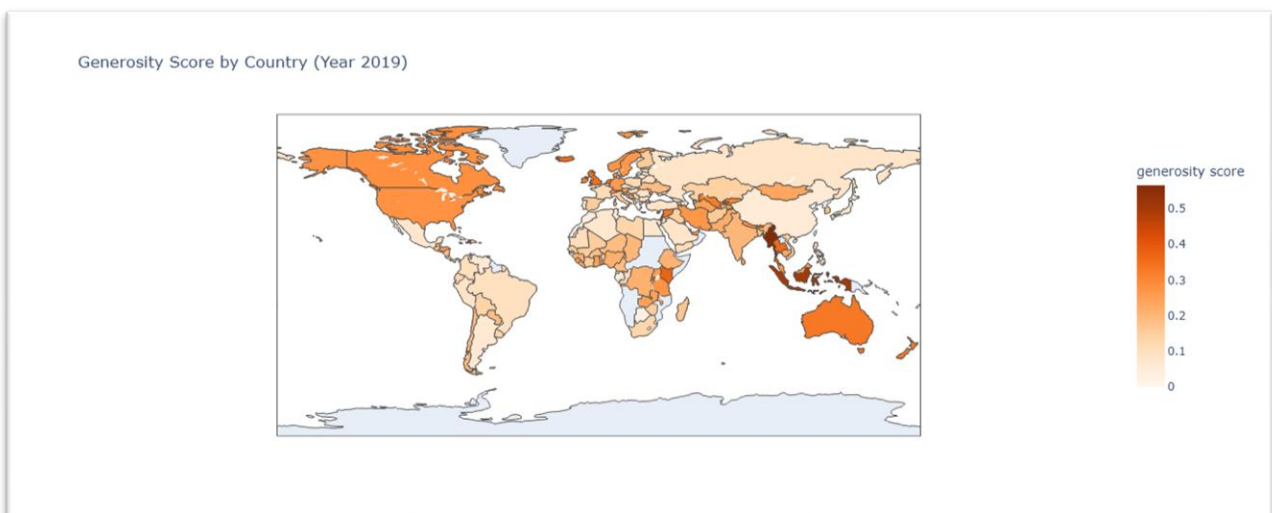
Healthy life expectancy significantly contributes to well-being. Countries with higher health scores also exhibit lower anxiety and higher productivity, leading to better life evaluations. In regression analysis, a one-unit increase in healthy life expectancy corresponds to approximately 0.3 increase in happiness score.



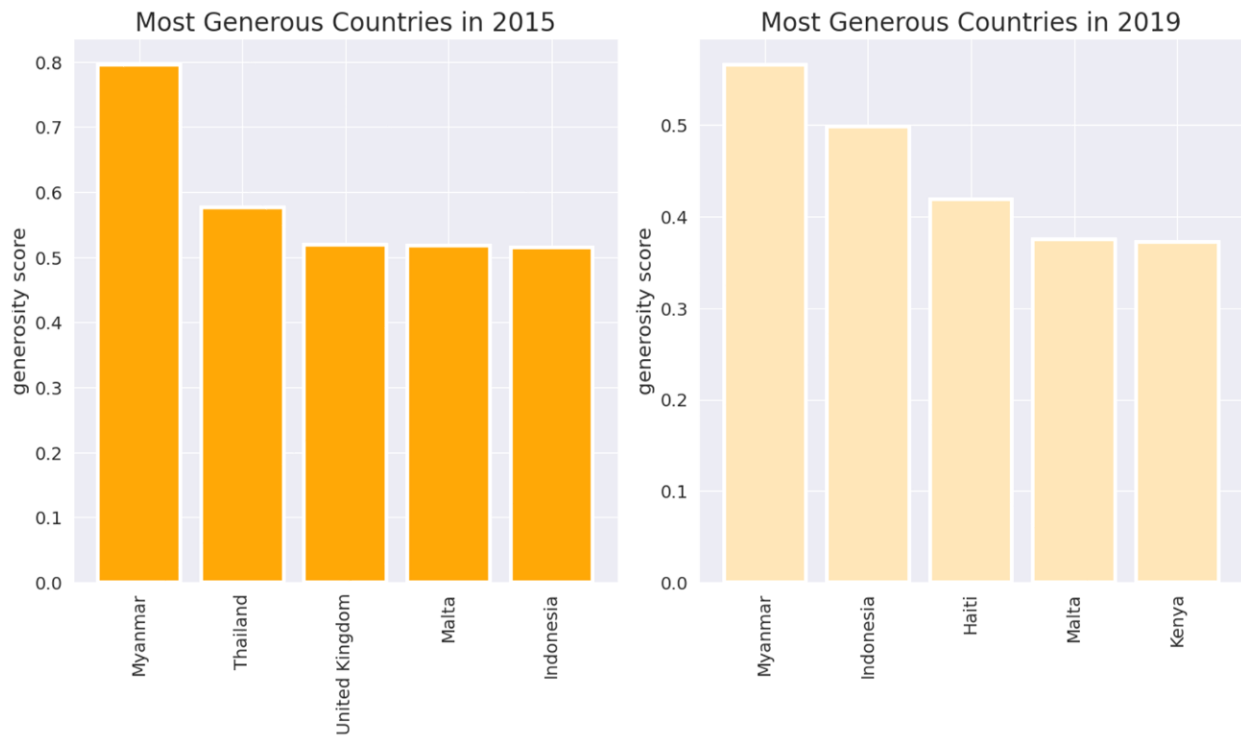


4.4) Social and Institutional Trust

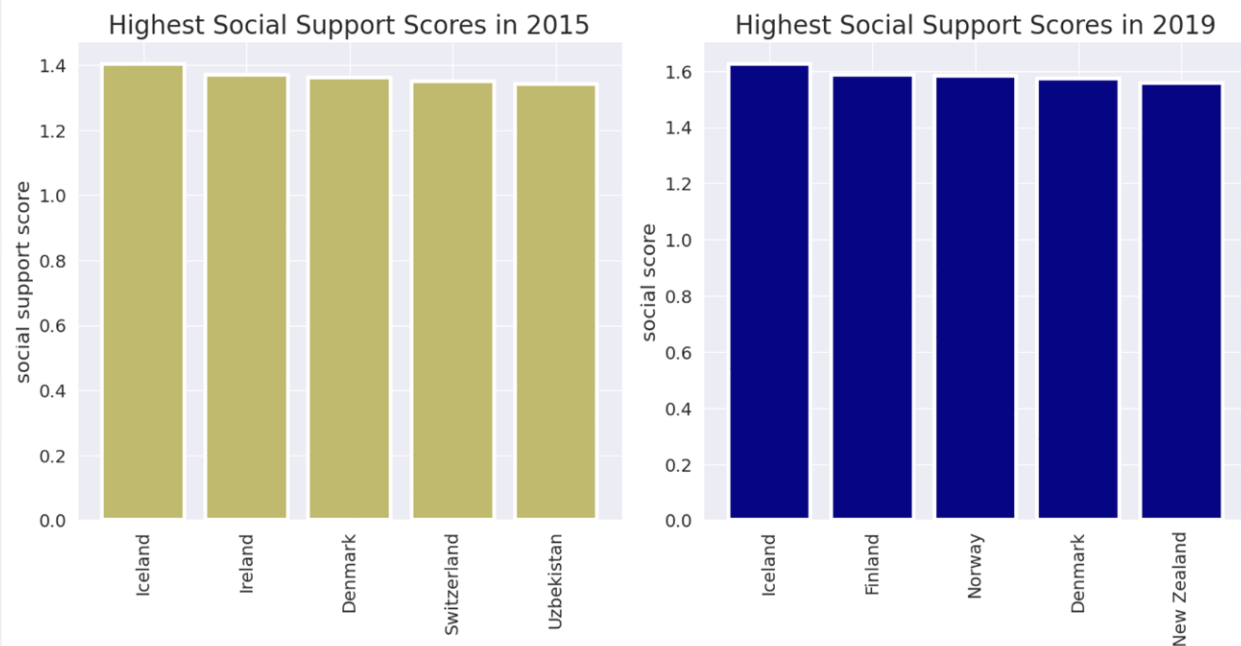
Social support — defined as the availability of someone to rely on in times of trouble — exhibits one of the strongest correlations with happiness ($r \approx 0.8$). Freedom to make life choices and generosity also demonstrate strong positive associations. These findings underscore that social cohesion and trust play as critical a role as economic factors.



Countries with the Highest Generosity Scores

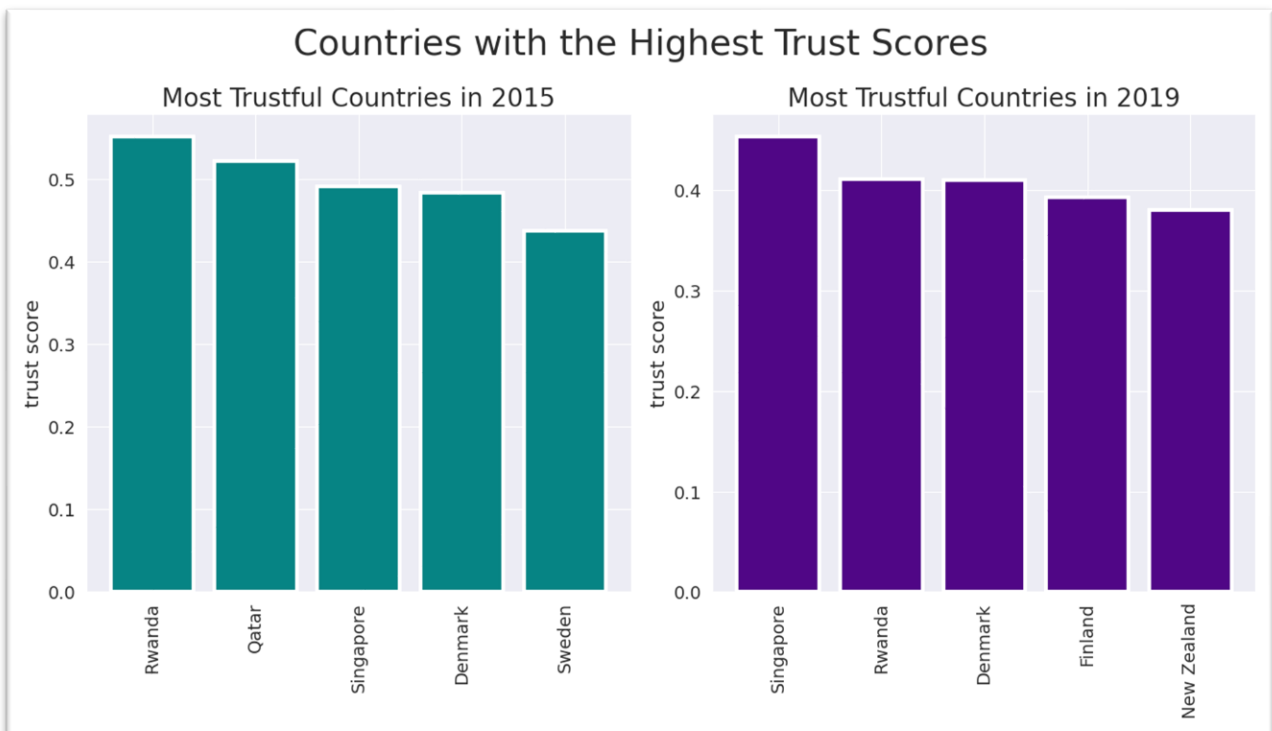
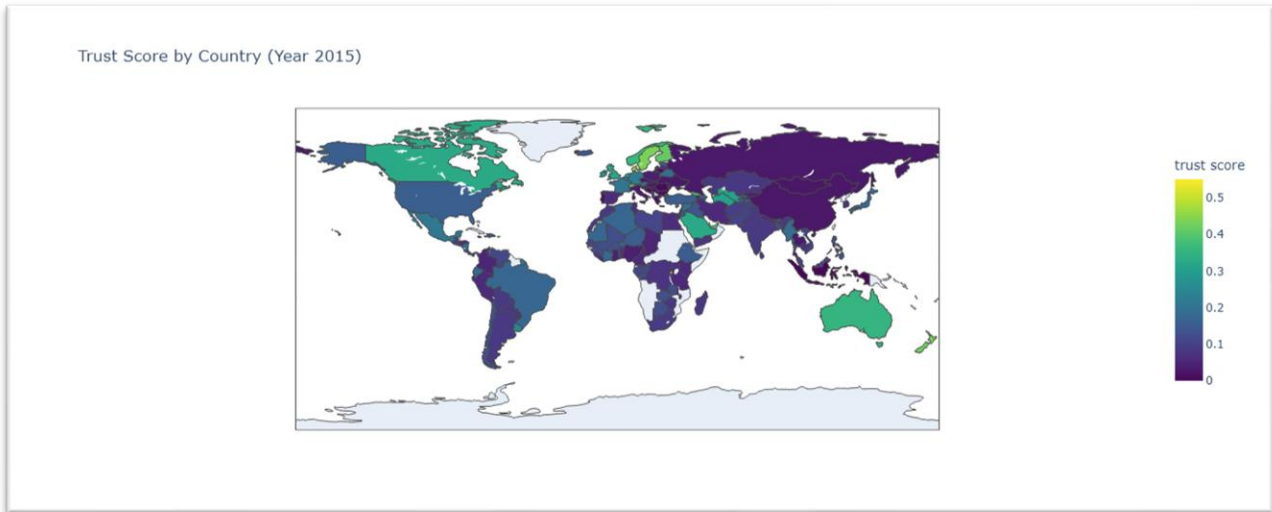


Countries with the Highest Social Support Scores



4.5) Governance and Corruption

Perception of corruption has a strong negative correlation with happiness ($r \approx -0.6$). Nations with transparent governance, effective public services, and accountability frameworks exhibit higher well-being. This highlights that ethical leadership and fair institutions are vital for sustainable happiness.



5. Regional Insights

Comparing happiness across continents reveals stark contrasts. Western Europe, North America, and Oceania dominate the top quartile of the index, while Sub-Saharan Africa and South Asia cluster toward the bottom. Economic inequality, political instability, and lower life expectancy explain much of the gap.

Regional Rankings (Illustrative):

- Western Europe — Average Happiness: 7.2
- North America — 6.9
- East Asia — 5.8
- South Asia — 4.9
- Sub-Saharan Africa — 4.3

6. Synthesis of Key Insights

1. Economic Strength Matters — But Isn't Everything: GDP remains a key driver, but countries like Costa Rica outperform wealthier nations due to stronger social support systems.

2. Trust and Freedom Are Core Pillars: Citizens value the ability to make personal choices and live in transparent systems.

3. Health and Longevity Reinforce Well-being: Preventive healthcare and life expectancy are pivotal in ensuring sustained happiness.

4. Cultural Values Influence Resilience: Regions with strong cultural or community ties show greater emotional stability despite lower incomes.

7. Limitations & Considerations

While the World Happiness dataset provides valuable macro-level insights, it relies heavily on subjective surveys. Cultural response biases, missing data from certain countries, and inconsistent time coverage introduce potential inaccuracies. Future studies could incorporate additional variables such as environmental quality, digital access, and political freedom indexes.

8. Strategic Recommendations

For governments and organizations seeking to enhance well-being, data-driven policies can make a measurable difference. Based on the analysis, the following strategic initiatives are recommended:

- Invest in public healthcare and preventive medicine.
- Strengthen community and family support programs.
- Increase transparency and anti-corruption initiatives.
- Promote work-life balance and individual autonomy.
- Encourage civic engagement and volunteerism as happiness multipliers.

9. Conclusion

The World Happiness Report data reaffirms that true prosperity extends beyond GDP — it is a composite of economic stability, trust, health, and human connection. This analysis highlights that nations prioritizing social infrastructure and transparency tend to achieve higher and more sustainable levels of happiness.

Data-driven approaches like this empower policymakers and researchers to benchmark progress, allocate resources effectively, and build societies rooted in well-being rather than material output.