

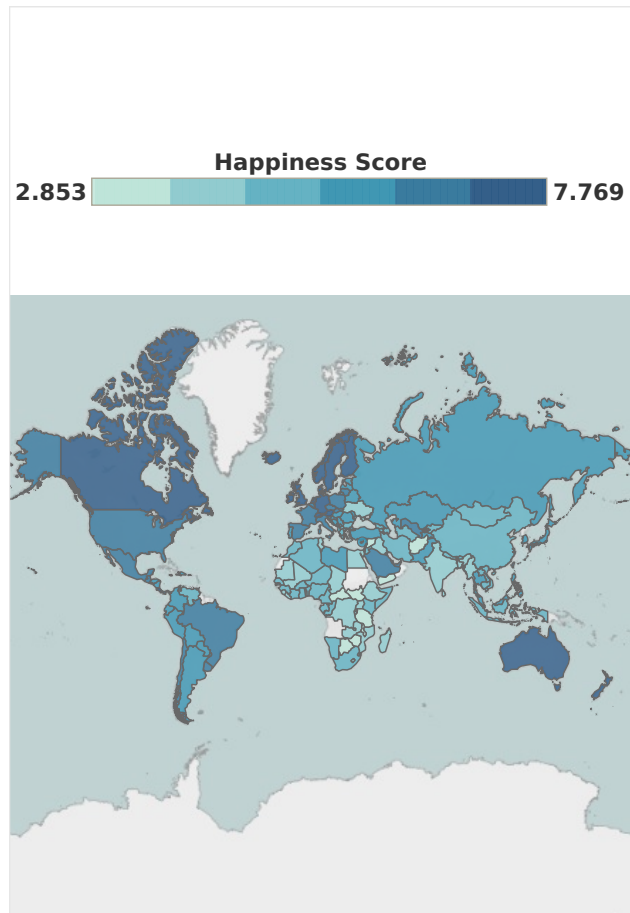


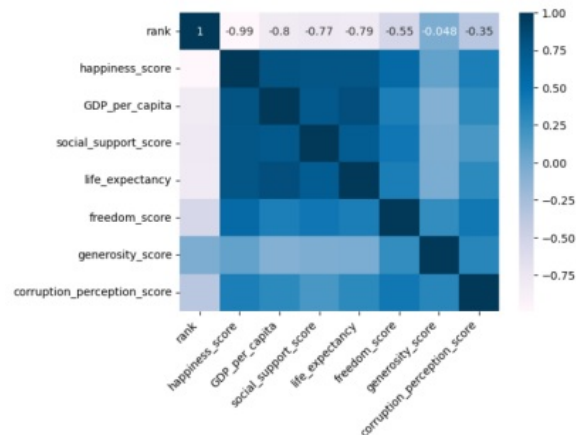
Introduction

The **World Happiness Report** is a collaborative effort involving Gallup, the Oxford Wellbeing Research Centre, and the UN Sustainable Development Solutions Network (SDSN). Its aim is to address the growing global interest in happiness and well-being as fundamental indicators for government policies. The life evaluations used in the annual happiness rankings are derived from the Gallup World Poll, which asks respondents to rate their own current lives on a scale from 0 to 10, with 10 representing the best possible life and 0 the worst.

Analysis Objectives

This analysis of the World Happiness Report 2019 is geared toward identifying trends and patterns in happiness scores, examining regional disparities, assessing correlations with various factors, conducting cluster analysis, and effectively communicating insights.



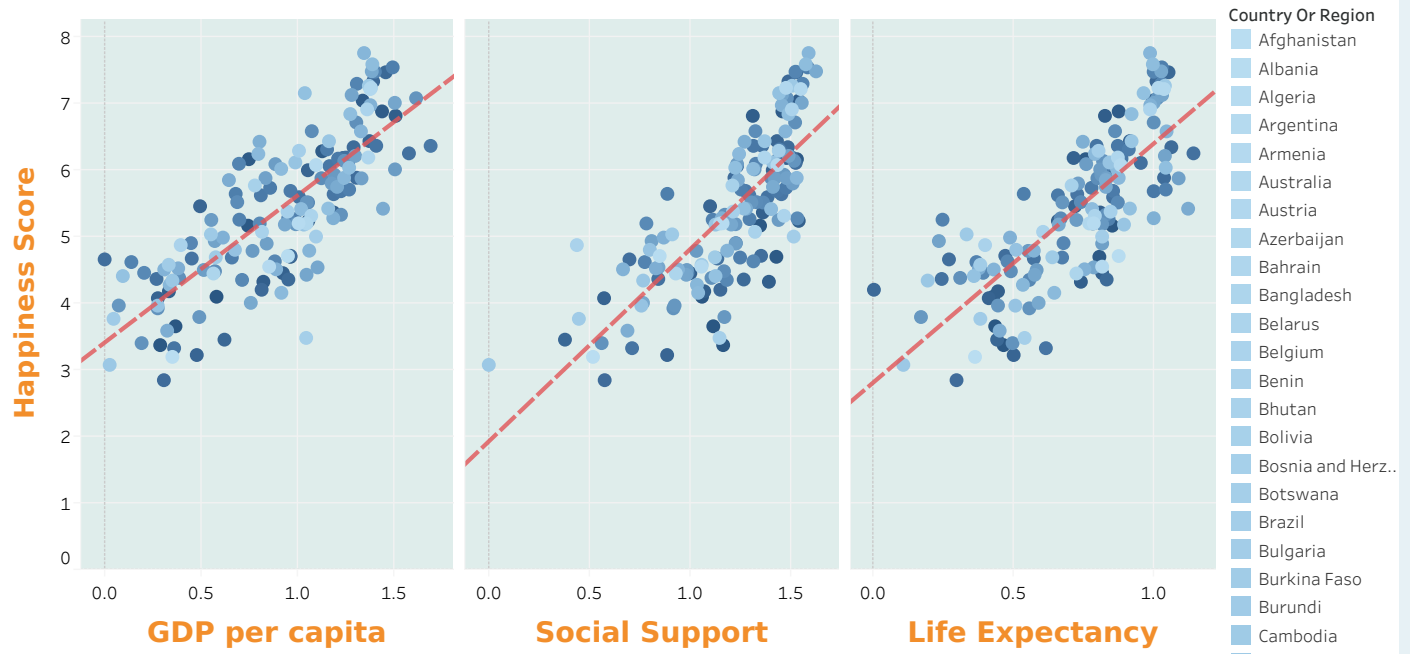


The exploratory phase of the analysis unveiled **strong positive correlations** between happiness and GDP per capita, social support, and life expectancy, as well as a moderately strong relationship between happiness and freedom. However, the correlations with generosity and corruption perception are **weaker**.

This led to the **hypothesis** that countries with higher GDP per capita, greater social support, longer life expectancy, and more freedom tend to have higher happiness scores.

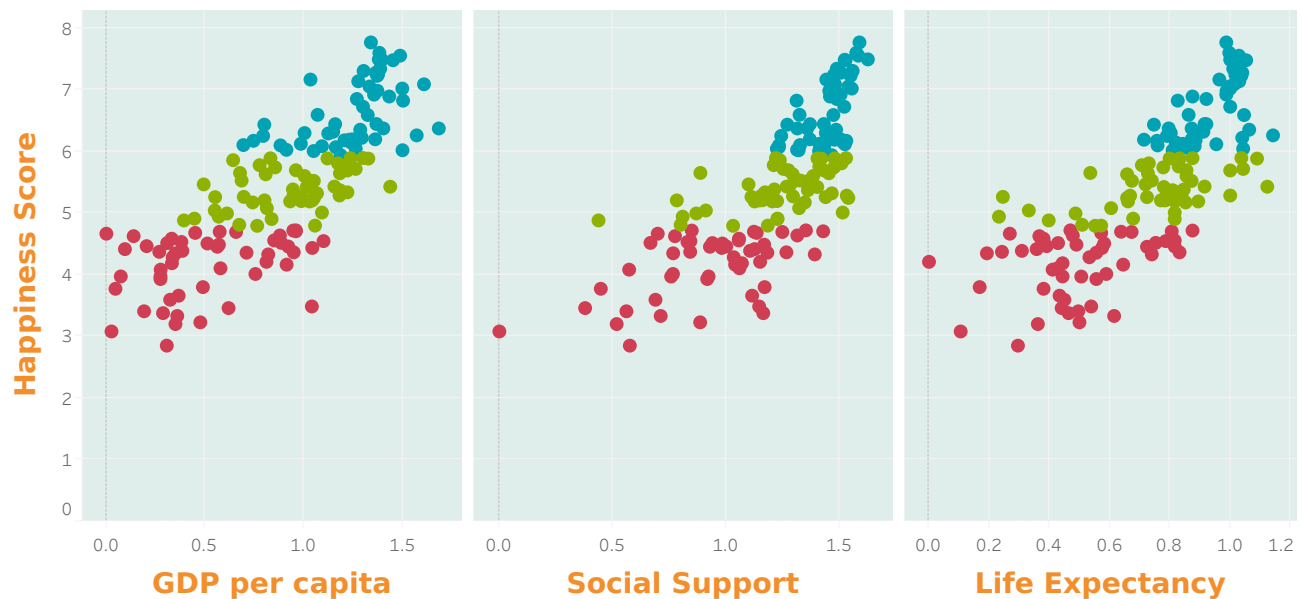


To investigate this hypothesis, a regression analysis was performed, indicating that these factors collectively account for **ca. 60%** of the variability in happiness scores.



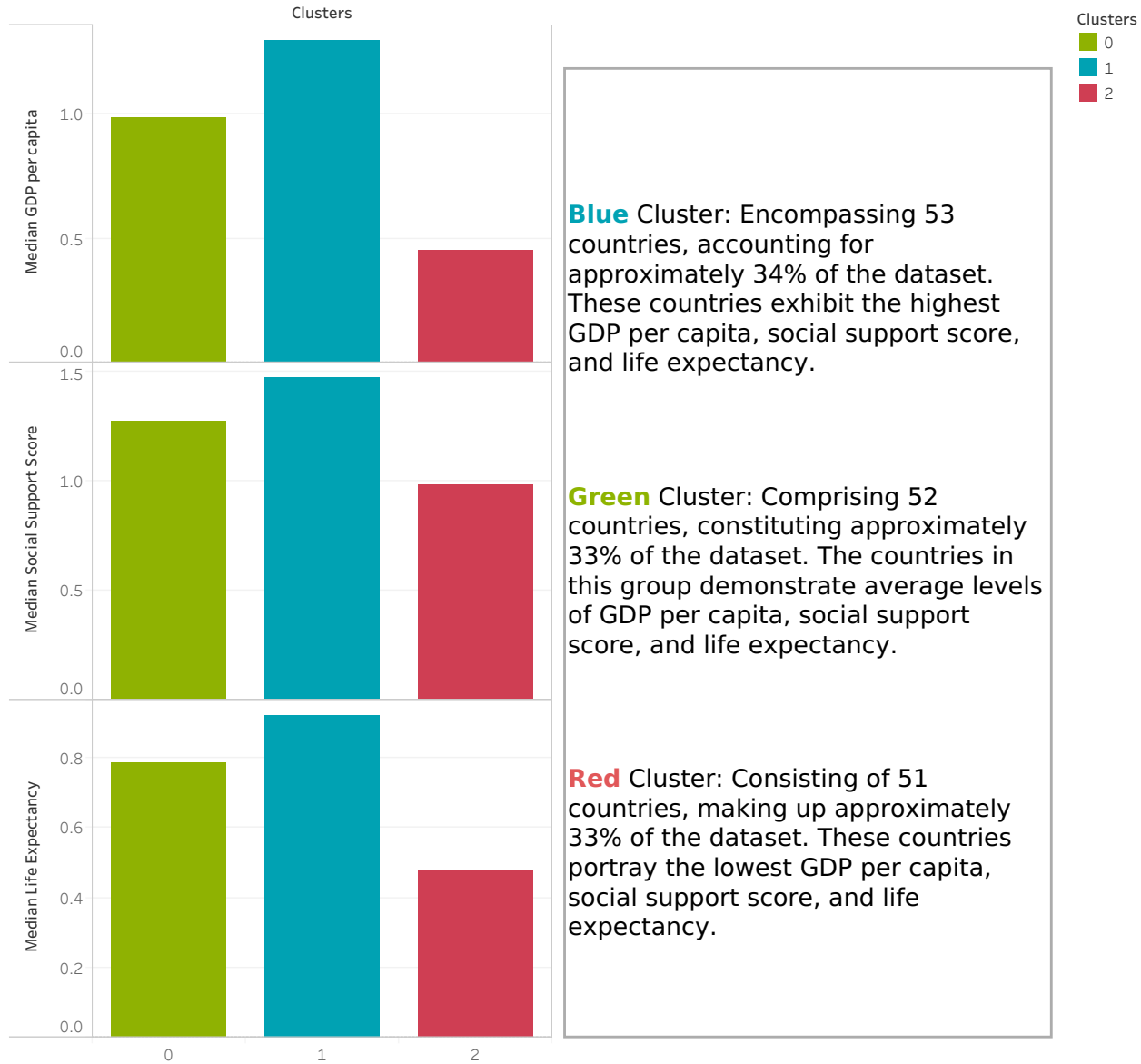
While this correlation is moderately strong, it suggests the need for additional exploration through non-linear analyses to fully understand the relationship between the predictor variables and happiness scores.

Hence, we opted to conduct a cluster analysis for a more comprehensive understanding of the data to group countries based on similarities in their socio-economic factors.

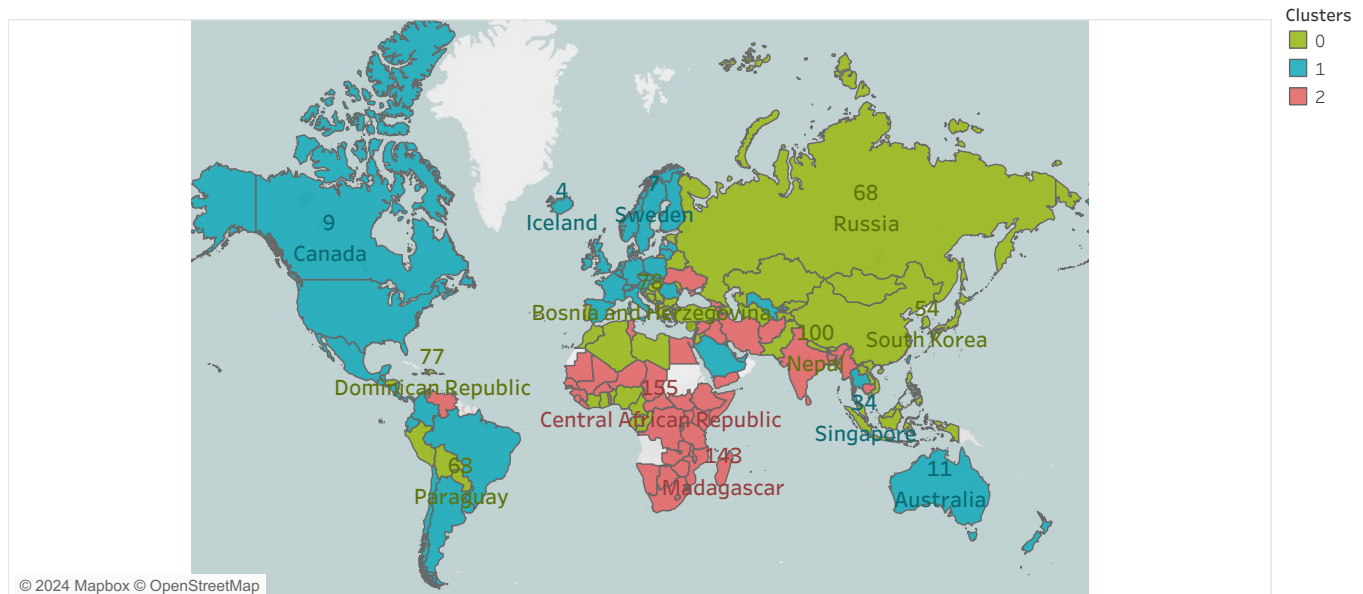


What's immediately noticeable from clearly delineated clusters is that while linear analysis identified GDP per capita as the best predictor with 63% accuracy, the densest clusters appear to form when happiness score is plotted against social support score and life expectancy.

This observation is particularly pronounced for countries belonging to the **blue** (high-prosperity) and **green** (average-prosperity) clusters.



Introduction and Analysis O..	Exploratory Analysis	Linear Regression	Cluster Analysis	Cluster Analysis Results	Final Results	Next Steps and Limitations
-------------------------------	----------------------	-------------------	------------------	--------------------------	---------------	----------------------------



Results

After clustering the data points, it became evident that higher socio-economic prosperity is indeed a reliable indicator of happiness across the globe.

The **blue** cluster, comprising the most prosperous countries, predominantly includes nations from Western Europe, North America, Australia, and New Zealand. Surprisingly, quite a number of countries from South America, as well as Romania in Eastern Europe and Uzbekistan in Central Asia, were also found within this cluster.

The **green** cluster represents a more dispersed group of countries ranging from Russia and China to North and West Africa, and from Central to South America.

In contrast, the least prosperous cluster (**red**) encompasses much of the Middle East and Africa but also includes regions further away, such as Venezuela, Albania, and Ukraine.

These insights suggest that while socio-economic factors reliably predict happiness scores, they do not account for the entire picture. Therefore, a comprehensive analysis should also consider factors with weaker correlations with happiness, such as **generosity** and **perceptions of corruption**.

Introduction and Analysis O..	Exploratory Analysis	Linear Regression	Cluster Analysis	Cluster Analysis Results	Final Results	Next Steps and Limitations
-------------------------------	----------------------	-------------------	------------------	--------------------------	---------------	----------------------------



± **Conducting further analysis** to explore the impact of additional variables on happiness scores, such as freedom to make choices, generosity and erceptions of corruption

± **Applying advanced machine learning techniques** to improve the accuracy of predictive models

± **Investigating temporal trends** by using time-series analysis techniques to identify significant changes or patterns.

± **Conducting surveys or qualitative research** to complement quantitative analysis, gaining insights into the subjective experiences and perceptions of happiness among populations.



§ **Limited Variables:** The analysis focused on a subset of socio-economic factors, potentially overlooking other influential variables such as cultural norms, political stability, and environmental conditions.

§ **Data Quality:** The accuracy and reliability of the data sources used for analysis may vary across countries, leading to potential biases or inaccuracies in the results.

§ **Generalization:** While clustering provides insights into broad trends, individual country experiences may vary, and generalizing findings to all countries should be done cautiously.

§ **Lack of Causality:** Correlation does not imply causation, and the analysis may not uncover the underlying causal relationships between socio-economic factors and happiness scores. Further research is needed to establish causality.