Milestone 1: Data Overview and Preliminary Analysis for Global Well-Being

Andrea Miele, David Revaz, Zeynep Tandogan

Github link: **Click Here**

1 Dataset

We are integrating complementary Kaggle datasets to create a comprehensive view of national well-being. The first dataset, "Quality of Life Index by Country" evaluates various living standards and socio-economic indicators from 2015 to 2022. The second, the "Human Development Index (HDR) Dataset", covers 1990 to 2022 and details progress in education, income, life expectancy and other human development factors. Finally, the "Global Happiness Scores and Factors" dataset adds insights into worldwide happiness levels and its contributing factors. Linking these datasets allows us to explore the relationships among quality of life, human development, and happiness to build a multi-dimensional picture of national well-being.

While the Quality of Life Index and Global Happiness datasets are already cleaned, the HDR dataset required reformatting. Originally, metrics were formatted in columns (e.g., hdi_1990, hdi_1991, etc.), so we pivoted the data to create a dedicated Year column with one column per metric and multiple rows for each year. Additionally, we filled incomplete entries, such as missing regions (using "EU" for Europe). Even when focusing on the period 2015–2022—when the first two datasets are complete—we found that 65 out of 193 countries have at least one missing value. On average, each of these 65 countries is missing data for 7.37 unique years (median 8.00 out of 8), and the overall mean number of missing columns per country is 9.11 (median 8.00). These gaps will require further cleaning before robust visualizations can be developed.

2 Problematic

2.1 Overview

The project aims to combine happiness and quality of life data with human development index (HDR) dataset to create a comprehensive perspective on global well-being. In this project, global well-being will be analyzed in the scope of external life conditions such as socio-economic factors. By merging these datasets, our aims are to identify and understand the factors influencing well-being globally and investigate the connections among important determinants. Through visualizations, our goal is to highlight trends, disparities, and correlations, providing valuable insights into global well-being.

2.2 Motivation & Target Audience

Our motivation is that global well-being is a critical issue that directly affects the quality of life of individuals and the sustainable development of societies. Today, there are various factors that affect well-being levels in many ways such as economic fluctuations, educational inequalities. Therefore, understanding the fundamental external dynamics that determine well-being is not only beneficial but crucial for:

- **Policymakers:** To create evidence-based policies that improve people's quality of life.

- **Civil Society Organizations & NGOs:** To recognize and resolve the disparities in well-being.
- **Researchers & Academics:** To better understand the factors affecting global well-being for future studies.
- **Individuals & General Public:** To raise awareness on well-being.

2.3 Our Goals with Visualizations

- Ranking countries on indicators of happiness, quality of life, and human development metrics.
- Compare how happiness correlates with specific quality of life components (e.g., Safety Value, Purchasing Power Value) and HDR measures (e.g. life expectancy, education).
- Analyzing improvements or declines in happiness, quality of life, and HDR indicators across regions over time.
- Identifying patterns of countries that consistently perform well or poorly on various measures of well-being.
- Detecting countries that deviate significantly from typical relationships observed in the dataset (e.g., a country with relatively low HDI but high happiness).

3 Exploratory Data Analysis

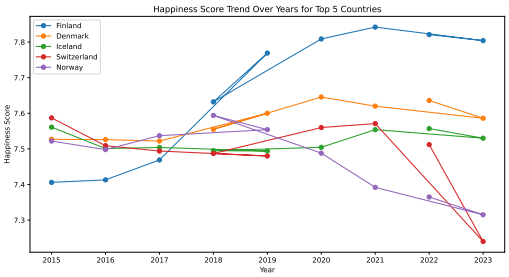
The datasets we have chosen for our analysis are already well-curated and sanitized. In particular, the Quality of Life Index by Country and the Global Happiness Scores and Factors datasets come with no missing entries, while the Human Development Index dataset required only minor pre-processing (such as pivoting metrics with years embedded in their column names into a long, more user-friendly format). Our final combined data allows us to explore various dimensions of well-being, development, and happiness across different countries.

Beginning with happiness scores, Figure 1a illustrates the trend of average happiness scores over time, revealing interesting fluctuations that warrant further investigation. Figure 1b highlights the correlations among various socio-economic indicators across our datasets. Trends among the top five countries in terms of happiness scores are detailed in Figure 2a, while average happiness score for the top 10 happiest countries are shown in 2b. Further analysis of the contributing factors are shown in Figure 3a and 3b which help to uncover the relative influence and highlight which dimensions of wellbeing are most strongly associated with higher happiness scores.

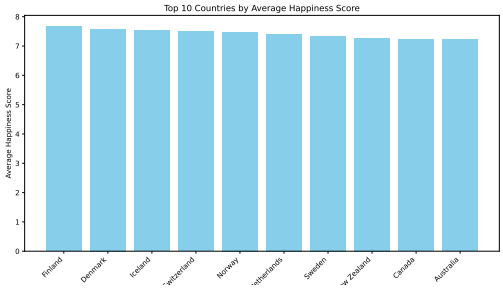
(a) Average happiness score over years.

(b) Correlation among various socio-economic indicators.

Figure 1: Happiness: Trends and Correlations (Part I)

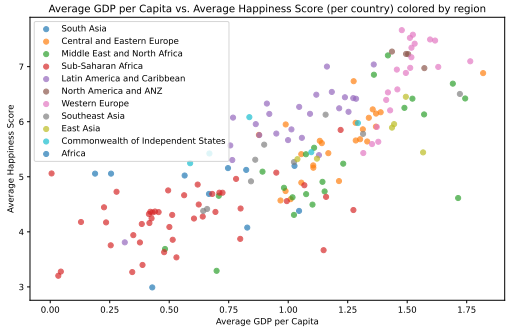


(a) Happiness score trend of the top 5 countries.

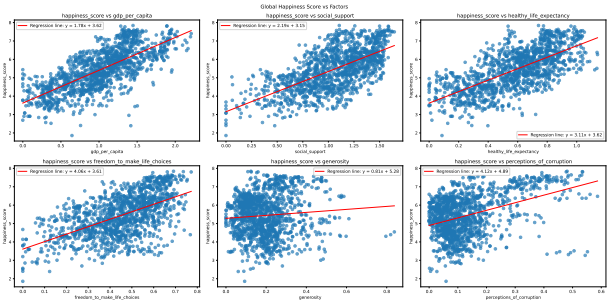


(b) Top 10 countries by happiness score.

Figure 2: Happiness: Top Country Analysis (Part II)



(a) GDP per capita vs. Happiness score by region.



(b) Happiness score versus different factors.

Figure 3: Happiness: Economic and Factor Analysis (Part III)

Turning to quality of life, Figure 4 shows the interrelationships between the different quality of life dimensions with a correlation matrix, while Figure 5a provides a comprehensive view of the data distribution across countries. Figure 5b reveals how different quality of life factors interrelate. The trend for the top 5 countries based on quality of life indices and the average quality of life indices of the top 10 countries are presented in 6a and 6b, respectively.

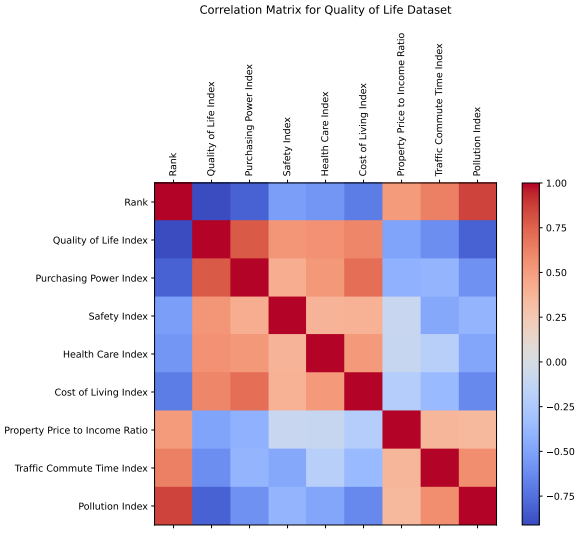
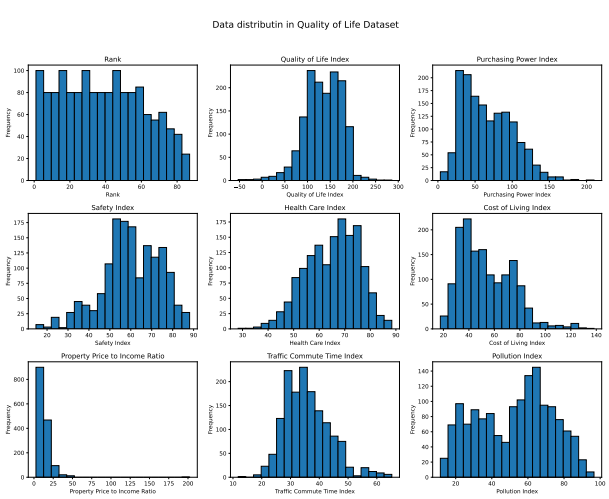
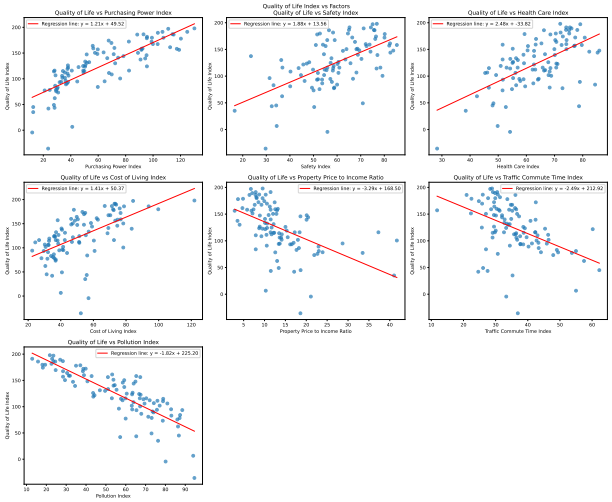


Figure 4: Correlation matrix on different quality of life indicators

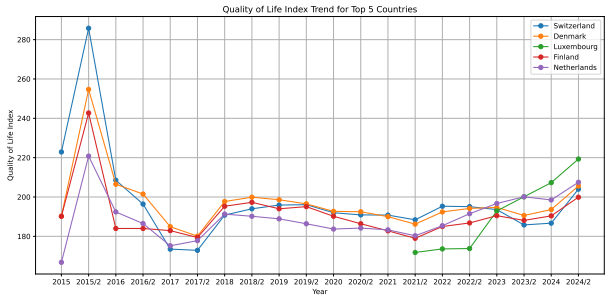


(a) Data distribution of the different quality of life indicators

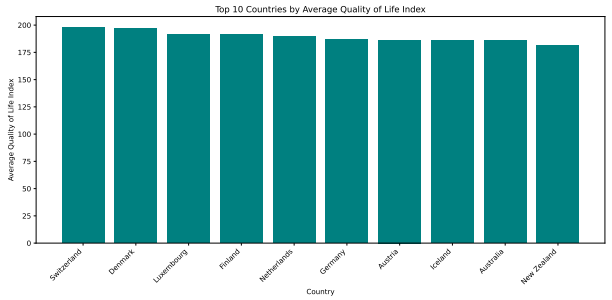


(b) Quality of life versus factors

Figure 5: Quality of life data and factor comparisons.



(a) Evolution of quality of life over time for the top 5 countries



(b) The top 10 countries by quality of life

Figure 6: Quality of life rankings and trends.

Focusing on human development, temporal trends in human development are captured in Figures 8a and 8b, while the GNI vs. HDI relationship for the year 2020 is illustrated in Figure 7. More detailed comparisons for the year 2022 are presented through multiple chars. Figure 9a shows the HDI against GNI per capita by regions, while population-based comparisons is presented in Figure 9b. Broader variable interactions can be found in Figure 10b and Figure 11a. These 2022-specific plots allow for a richer, multi-dimensional exploration of how HDI relates to various indicators. Figure 10a further highlights disparities in life expectancy across regions. Further analysis can be extended based on zooming in a country as in Figure 11b.

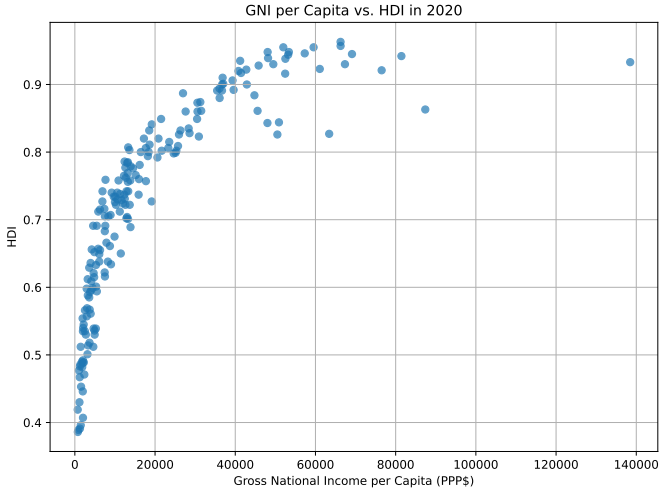
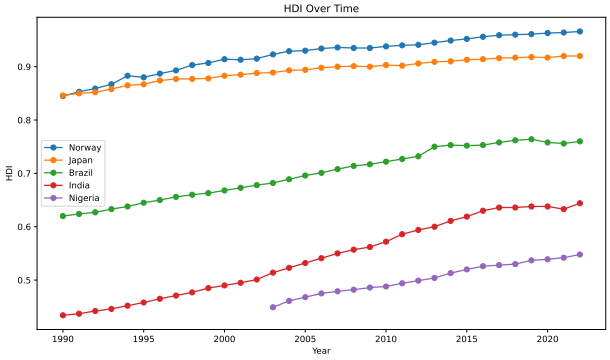
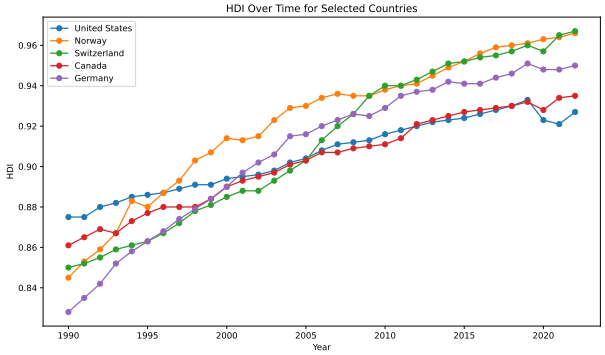


Figure 7: GNI vs HDI

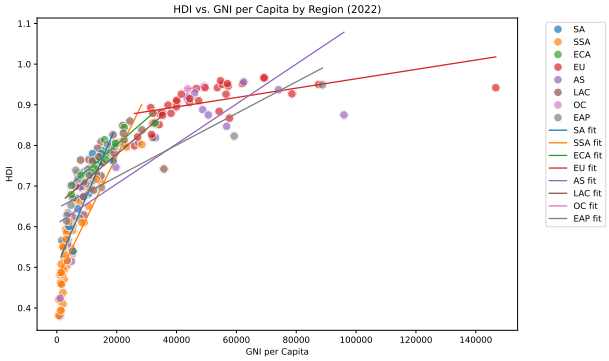


(a) HDI over time for a selection of countries

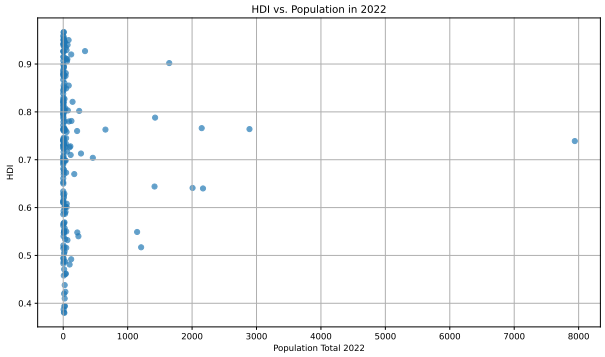


(b) HDI over time for the 5 best countries in 2022

Figure 8: HDI trends over time.

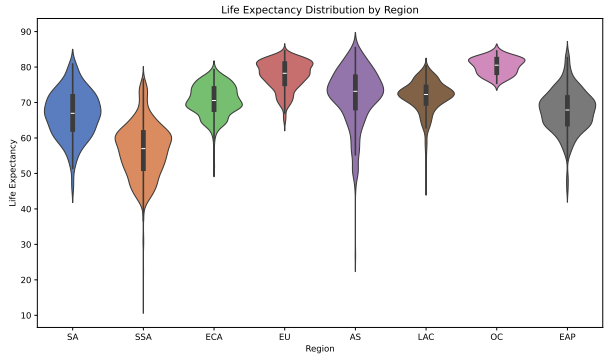


(a) HDI vs GNIPC per region

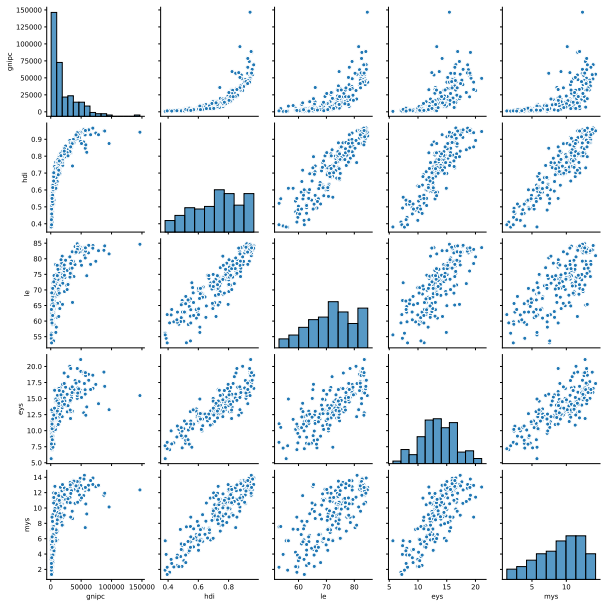


(b) HDI vs Population in 2022

Figure 9: Relationships between HDI and other economic/demographic factors.

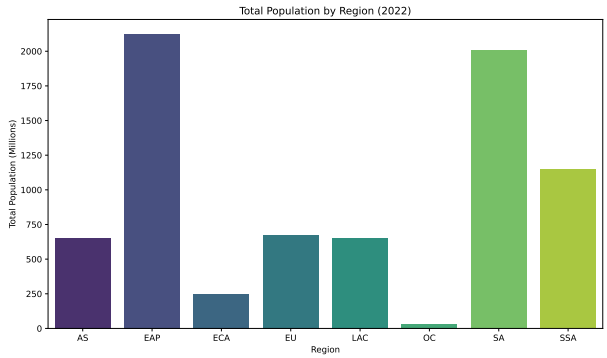


(a) Life expectancy grouped by region

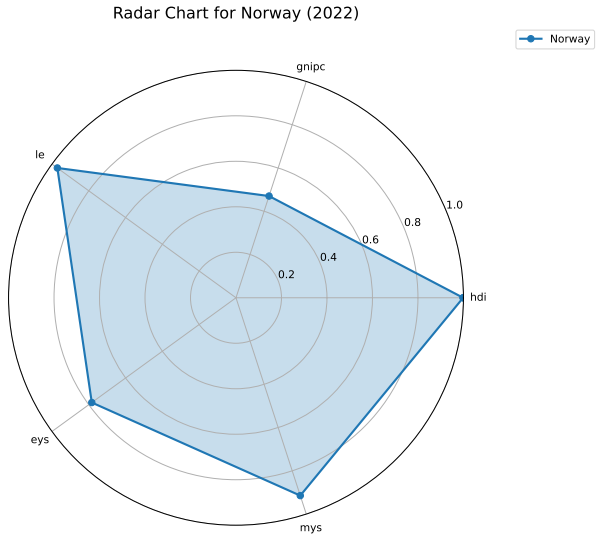


(b) Pairwise comparison of key indicators for 2022.

Figure 10: Combined figures: (a) Life expectancy by region and (b) Pair plot for 2022.



(a) Population by region in 2022



(b) Radar chart for Norway showing several indicators

Figure 11: Combined figures: (a) Population by region (2022) and (b) Radar chart for Norway.

4 Related Work

Previous studies have examined well-being through various national-level indices, often focusing on one or two dimensions at a time. The Quality of Life Index (QOLI), such as Numbeo's, aggregates objective factors like safety, healthcare, and cost of living, and correlates strongly with HDI tiers, particularly in countries with "Very High" HDI ([Shu et al., 2022]; [Koohi et al., 2017]).

The Human Development Index (HDI), which combines education, longevity, and income, is widely used to benchmark national development. HDI aligns closely with many social outcomes and correlates strongly with happiness scores ([Perkins et al., 2021]; [Nations, 2014]). Happiness scores, measured via surveys like the Gallup World Poll, have been popularized through the World Happiness

Report, which links happiness to income, health, and social support ([Ortiz-Ospina and Roser, 2017]). Studies using these scores often treat happiness as the outcome variable explained by objective measures; for instance, Jannani et al. found that GDP per capita was the strongest predictor of happiness in a random forest model ([Jannani et al.,]).

Some research has attempted to compare multiple indices, often using correlation or regression analyses between pairs such as HDI and happiness ([Nations, 2014]; [contributors, 2025]). A few have gone further: Perkins et al. examined four well-being metrics simultaneously ([Perkins et al., 2021]). Most literature focuses on isolated or paired indicators.

Our approach differs by fully integrating these three dimensions into a single dataset. Rather than explaining one measure with others, it treats all indicators as complementary components of a broader well-being profile. This allows for multidimensional analysis, such as identifying clusters of countries with mismatched well-being indicators.

References

[contributors, 2025] contributors, W. (2025). Quality of life.

[Jannani et al.,] Jannani, A., Sael, N., and Benabbou, F. Machine learning for the analysis of quality of life using the World Happiness Index and Human Development Indicators.

[Koohi et al., 2017] Koohi, F., Nedjat, S., Yaseri, M., and Cheraghi, Z. (2017). Quality of Life among General Populations of Different Countries in the Past 10 Years, with a Focus on Human Development Index: A Systematic Review and Meta-analysis.

[Nations, 2014] Nations, U. (2014). Getting Serious about Happiness.

[Ortiz-Ospina and Roser, 2017] Ortiz-Ospina, E. and Roser, M. (2017). Happiness and life satisfaction.

[Perkins et al., 2021] Perkins, D. D., Ozgurer, M. R., Lupton, A., and Omidvar-Tehrani, S. (2021). Well-Being as Human Development, Equality, Happiness and the Role of Freedom, Activism, Decentralization, Volunteerism and Voter Participation: A Global Country-Level Study. *Frontiers in Psychology*, 12.

[Shu et al., 2022] Shu, Z., Carrasco, R. A., García-Miguel, J. P., and Sánchez-Montañés, M. (2022). Multiple scenarios of quality of life index using fuzzy linguistic quantifiers: the case of 85 countries in Numbeo. *Mathematics*, 10(12):2091.