Revealing the Complex Interactions between the Human Development Index, the Sustainable Development Goals and Hidden Factors: A Machine Learning-driven Analysis

D. Sierra-Porta[1,a], Mairene Tobón-Ospino[2,a], and Humberto Marbello Peña[1]

[1]Facultad de Ciencias Básicas. Universidad Tecnológica de Bolívar, Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco, Cartagena de Indias, 130010, Colombia.
[2]Escuela de Trabajo Social. Universidad Industrial de Santander. Grupo de Investigación Prometeo. Car 27 #9, Bucaramanga 680001, Santander, Colombia
[*]Corresponding author: dporta@utb.edu.co

June 24, 2023

Abstract
The Sustainable Development Goals (SDGs), established by the United Nations Organization in 2015, have emerged as a comprehensive framework for evaluating the potential of nations and their societies to forge a path towards improving the lives of people worldwide, guided by the principle of "leaving no one behind." Alongside this, the Human Development Index (HDI), developed by the United Nations Development Program, has served as a tool for categorizing countries based on their potential for human development, assuming a scenario without inequality. This study aims to investigate the correlation between these two indices by employing modern machine learning techniques. Specifically, Pearson's correlation and linear regression analyses are utilized to examine the relationship between the two indicators. Furthermore, dimensionality reduction methods, notably the machine learning algorithm UMAP, are explored in this research.

Keywords: Sustainable Development Goals, Human Development Index, UMAP, Machine Learning, Data Science.

1 Introduction

1.1 Measuring Countries' Development

In 2015, the United Nations General Assembly (UNGA) established the Sustainable Development Goals (SDGs), also known as the 2030 Agenda. These goals serve as a collective and individual framework for countries to strive towards achieving by the year 2030 (UNGA, 2016; Lee et al., 2016). Comprising 17 interconnected and interdependent objectives, the SDGs aim to foster a more sustainable and prosperous future for all, encapsulating a global aspiration of "leaving no one behind." The 17 goals encompass various areas, including poverty eradication, zero hunger, health and well-being, quality education, gender equality, clean water and sanitation, affordable and clean energy, decent work and economic growth, industry and innovation, reducing inequalities, sustainable cities and communities, responsible production and consumption, climate action, marine life preservation, terrestrial ecosystem preservation, peace, justice and strong institutions, and partnerships for the goals.

Recognizing the interdependence and correlation among these goals as well as their global significance, the United Nations (UN) established indicators and evidence targets in 2017 to facilitate measurement and assessment of goal achievement (UNGA, 2017; MacFeely, 2020; Janoušková et al., 2018). These indicators enable the quantification of countries' progress toward achieving the SDGs and are tailored to measure specific aspects of the targets. They employ various scales of measurement, including percentages, population sizes, and other relevant metrics.

The SDGs have evolved into a significant conceptual framework, not only driving research across multiple disciplines but also playing a crucial role in advancing countries' development toward a better world. Scholars and researchers have utilized the SDGs to explore diverse aspects of associated indicators, addressing specific challenges within individual countries as well as on a global scale. The Sustainable Development Goals consist of 17 Goals, 169 Targets, and 232 distinct non-repetitive Indicators. While ongoing discussions may surround some of these indicators, several organizations have already embarked on measuring the SDGs using this framework.

In parallel, the United Nations Development Programme (UNDP) introduced the Human Development Index (HDI) in 2010, aiming to classify and quantify a country's level of development based on three dimensions. Unlike the comprehensive nature of the SDGs, the HDI focuses on three main dimensions: (1) longevity and a healthy life, (2) knowledge, and (3) decent living standards. The Life Expectancy Index (LEI) evaluates the first dimension, measuring life expectancy at birth. The second dimension involves assessing average years of schooling and expected years of schooling as indices. Lastly, the third dimension considers gross national income (GNI) per capita. In contrast to the SDGs, the HDI is measured on a scale from 0 to 1. The HDI was developed based on the research work of Pakistani economist Mahbub ul Haq in 1990, drawing heavily from the ideas developed by Amartya Sen (Ul Haq & United Nations Development Programme, 1989; Ul Haq et al., 1992).

One of the pressing priorities on the international agenda is establishing a link between the measurement system of the Human Development Index (HDI) and its relationship with the Sustainable Development Goals (SDGs). This connection is also reflected in the UNDP Human Development Reports, where the statistical data used in the 2019 Report highlight the alignment with the SDGs (UNDP, 2019,0).

However, it is crucial to consider the validity of the targets and indicators at the national and local levels. Debates often arise regarding the validity of 80 of the indicators (Diaz-Sarachaga et al., 2018; Dawes, 2020; Jain & Jain, 2020; Dalampira & Nastis, 2020). In the Turkey Situation Analysis Report, released by the Ministry of Development (https://sustainabledevelopment.un.org/content/documents/107102030%20Agenda%20Turkey%20Report.pdf), it is stated that 14 of the sub-targets do not apply to Turkey. For instance, the report rightly emphasizes that Turkey has already surpassed the extreme poverty line of $1.25. These minimum standards, typically designed for underdeveloped countries, require periodic updates until 2030.

Another significant issue concerns the structure of national statistical data used as the basis for constructing indicator sets. While many data are available at the national level, they may not be accessible at the regional level. In some countries, measuring certain indicators may not be feasible due to a lack of information or non-existent data. An analysis of data gaps and a literature review conducted for the UN SDG report (UNDP, 2020) demonstrate poor data quality (Al-Salim et al., 2022; Hagel et al., 2020) and reveal a logical gap in data quality. The gaps identified in the UN SDG 2020 report highlight various data quality issues, including incomplete data from many countries, lengthy data processing times by official national statistical institutes (NSIs), insufficient data from many low-income countries, reliance on traditional data analysis methods, outdated data leading to inaccurate perceptions or decisions, and the use of data from different sources without a consistent data quality framework.

1.2 Relating HDI and SDGs

The Human Development Index (HDI) has been widely used since 1990 as a measure of socioeconomic development. However, the integration of socioeconomic development with environmental considerations has led to the formulation of the Sustainable Development Goals (SDGs). The interconnectedness between human development and sustainable development is essential for creating a better society.

Human development focuses on assessing basic human needs such as health, education, and livelihood within the context of the economy. However, given the interdependence of the SDGs, unregulated economic growth can have detrimental effects on the environment. The concept of integrating the economy and the environment has given rise to the SDGs. The HDI aims to ensure individuals' quality of life and economic well-being, while human development and sustainable development share a common vision of empowering people by enhancing their capabilities and capacities.

Numerous indices have been developed to specifically measure human development, including the Human Poverty Index, the Social Progress Index, and the HDI. Among these, the HDI, which was first calculated for UN member states in 1975, is the most widely used index. It provides a consistent framework for calculating the HDI for 182 countries. The HDI's broad coverage can be attributed to its utilization of easily understandable and measurable variables derived from national statistics. However, the HDI has faced criticism on various fronts, including its failure to account for inequality and environmental considerations, methodological flaws in reflecting within-country variations, the subjective selection of indices, and its overly simplistic and idealized nature (Hou et al., 2015; Sagar & Najam, 1998; Ranis et al., 2006). Despite these criticisms, the HDI remains a valuable criterion for comparing human development across countries, although it does not capture within-country variations such as income inequality, standard of living, and health disparities (Adam et al., 2015; Blum, 2013).

Certain variables measured in the SDGs are likely to be correlated with the HDI and may influence it through causal feedback relationships. For instance, an increase in GDP may impact specific HDI parameters, or vice versa. However, the extent and nature of the correlations between the HDI and the dimensional indices of the SDGs remain unclear, as existing studies have either produced inconclusive results or focused on narrow scopes of investigation (Mehta, 2014; Ouedraogo, 2013; Ranis et al., 2000), although some correlations are evident. For instance, the relationship between life expectancy and access to safe drinking water has been established, as access to clean and safe water reduces the transmission of waterborne diseases, improves overall health, and increases life expectancy (Greene, 2001; Mehta, 2014). Nevertheless, the causality between the HDI and other measurement variables in the context of the SDGs requires further exploration.

1.3 Recent Studies

While certain causal relationships between economic and environmental variables and human development are well-established (e.g., the impact of access to safe drinking water on human health), there are still aspects that require further exploration.

Previous studies have explicitly examined the relationship between the HDI and the SDGs. While most of these studies focus on understanding the internal consistencies and relationships among the SDGs themselves, only a few have extended their analysis to the connection with the human development index. For instance, a study by (Wang et al., 2018) explores the relationship between economic growth, renewable energy consumption, and the HDI in the case of Pakistan. The findings suggest that the consumption of renewable energy does not necessarily lead to improvements in the country's development process. In the area of health, there are also initiatives to examine related indices using multivariate techniques, as seen in studies by (Lim et al., 2016) and (Sachs et al., 2016).

Recent research has delved into the relationships between specific SDGs and their corresponding HDI components. For example, studies by (Division, 2002) and (Arriani et al., 2021) employ quantitative regression and correlation methods to analyze the underlying connections between goal 1 (poverty reduction) and the HDI, revealing an inverse relationship between the two variables. These studies also explore the relationships with goal 8 (decent work) under the hypothesis that increased economic development leads to a decrease in unemployment rates and the promotion of more decent jobs.

Research conducted in Central Kalimantan (Hukom, 2015) and Indonesia (Hidayat & Bariyah, 2020) highlights the negative impact of poverty on the HDI, indicating that higher levels of poverty are associated with lower HDI values. Similarly, a study in West Seram Regency, Maluku Province in 2018 demonstrates the negative correlation between HDI and poverty, while also noting a positive correlation with the dependency ratio (Amaluddin et al., 2018). These findings imply that an increase in poverty levels negatively affects the HDI.

Other interesting works have investigated the relationship between the HDI and the SDG associated with water quality (goal 6) (Kirschke et al., 2020; Carvalho et al., 2019). These studies highlight the importance of incorporating qualitative data on the challenges of water quality monitoring to gain a better understanding of the mechanisms underlying the negative relationship between the HDI and the application of water quality parameters.

1.4 Aims and scope

In this paper, our main aim is to explore the correlation and causality between the Human Development Index (HDI), its dimension indices, and various national economic, water, energy, and other variables from the Sustainable Development Goals (SDGs) measurement framework. We will analyze globally consistent time series data on the HDI and the SDGs to assess the existence and strength of correlations between these variables. Furthermore, we aim to investigate if there is any underlying causality between the variables and, if so, to understand the direction and magnitude of this causality.

Our approach differs from previous studies in that we utilize available data from both the HDI and SDGs to establish a set of latent variables using machine learning techniques, particularly dimension reduction methods. By deriving these latent variables, we aim to uncover underlying patterns and relationships between the SDGs, HDI, and other variables. This approach allows us to explore more comprehensive and nuanced associations among the variables and provides insights into the complex interactions between human development and sustainable development.

By employing advanced statistical and machine learning techniques, our study aims to provide a deeper understanding of the interplay between the HDI, SDGs, and related variables. We believe that this approach will contribute to the existing body of knowledge and provide valuable insights for policymakers, researchers, and practitioners working towards achieving sustainable development and improving human well-being.

2 Data and methods

2.1 Data

The data used in this research comes from multiple sources to examine the relationship between the Human Development Index (HDI) and the Sustainable Development Goals (SDGs). The information related to the SDGs is obtained from the SDG-Tracker initiative system, which provides data from various indicators using official statistics from the United Nations (UN) and other international organizations. The SDG-Tracker (https://sdg-tracker.org/) is a free, open-access publication that tracks global progress towards the SDGs and helps hold governments accountable for achieving these goals. The data obtained from the SDG-Tracker covers the period from 2015 to the present and is regularly updated with the latest information.

Regarding the HDI, it is a composite measure that reflects achievements in three dimensions of human development: a long and healthy life, access to education, and a decent standard of living. The HDI is calculated based on indicators such as life expectancy at birth, average years of schooling, expected years of schooling, and gross national income (GNI) per capita. The data for the HDI is available in the Human Development Report, published by the United Nations Development Programme (UNDP, 2020) (https://ourworldindata.org/human-development-index). The HDI values used in this study are those reported for the year 2021.

To provide an overview of the HDI variation across different regions of the world, Figure 1 visually presents the HDI values reported for the year 2021. Each bar represents the HDI value for a specific region, while the colored lines indicate the minimum and maximum HDI values reported for the countries within each region or subregion.

By utilizing these comprehensive datasets, we aim to analyze the relationship between the HDI and the SDGs and uncover any correlations or causal links between these variables. The data collected from these sources will enable us to examine the interplay between human development and sustainable development, contributing to a better understanding of the progress made towards achieving the SDGs and improving human well-being worldwide.

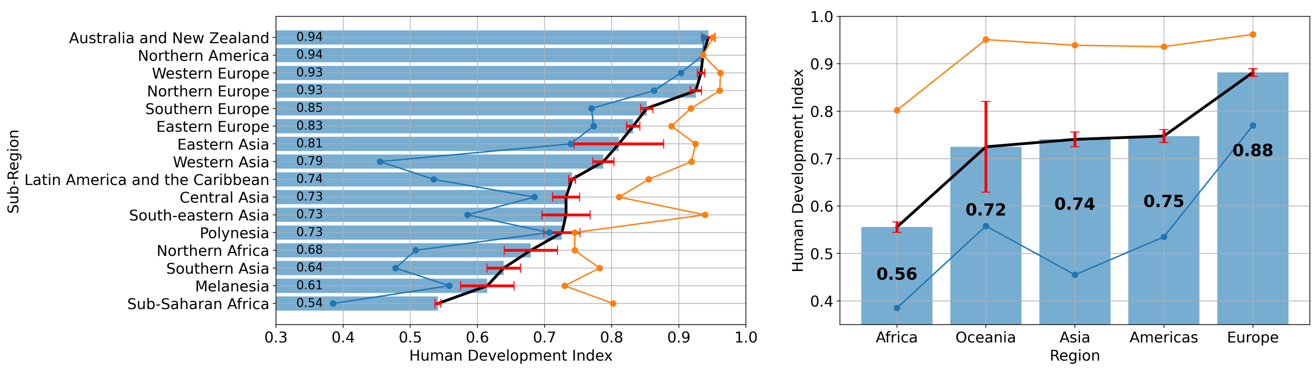

Figure 1: Visualization of HDI data variation across world regions in 2021, with bars representing calculated average HDI values and colored lines indicating the minimum and maximum indices within each region.

According to the UNDP, Human Development Report 2021 (UNDP, 2022,0,0), countries with very high human development (between 0.800 and 1.000) (Cooke et al., 2007; Mitrou et al., 2014) include Norway, Switzerland, Australia, Germany, Canada, while those with high human development (between 0.700 and 0.799) are for example Argentina, Chile, Uruguay, United Arab Emirates, Estonia. For medium performance (Nundy et al., 2022; Datta & Singh, 2019) (between 0.550 and 0.699) are India, South Africa, Egypt, Bolivia, Indonesia, and finally for those with low human development (below 0.550): Niger, Chad, Central African Republic, Burundi, Mali (Sarkodie & Adams, 2020; Otekunrin et al., 2020).

2.2 Data mining for data

From the data collected, a data engineering process is introduced (see figure 2) to ensure the widest possible representation of data and also of countries and variables with complete data.

The data preparation process for this research involves several steps to create a comprehensive dataset that combines the SDG indicators, HDI data, and additional indexes. Here is an overview of the data engineering process:

1. Data Collection: The SDG indicators are collected from the SDG-Tracker initiative system, which provides official statistics from the UN and other international organizations. The HDI data is obtained from the Human Development Report published by the UNDP. Additional indexes and variables are sourced from relevant sources.
2. Variable Selection: Only numerical quantitative variables are retained from the collected data. This helps in focusing on variables that can be analyzed and compared quantitatively.
3. Dataset Creation: A single dataset is created that includes all the indicators across all SDG targets. This dataset consists of 233 instances, representing 190 countries, as well as country clusters, associations, territorial organizations, and unions of confederations or countries. The dataset contains a total of 622 variables,

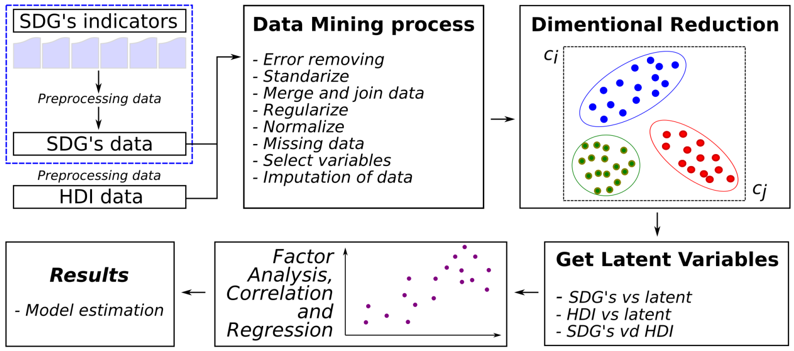

Figure 2: Logical framework of the data mining process and methodology analysis.

representing the information for all 17 SDG objectives.

4. Data Cleaning: Variables with missing or no information are eliminated from the dataset. Additionally, variables with a high number of missing data, exceeding a specified critical limit, are also removed. This step helps in ensuring that the dataset contains meaningful and informative variables.
5. Missing Data Imputation: To recover variables with missing data, a machine learning methodology is employed. Specifically, Random Forest, kNN (k-Nearest Neighbors), and Gradient Boosting techniques are utilized to impute missing values. These algorithms utilize patterns and relationships within the data to predict and fill in missing values.
6. Dataset Concatenation: Finally, the complete SDG dataset is concatenated with the HDI data and other relevant indexes. This step optimally organizes the dataset, allowing for comprehensive analysis and examination of the relationships between the HDI, SDGs, and other variables.

By following this data engineering process, a unified and comprehensive dataset is created that incorporates the SDG indicators, HDI data, and additional indexes. This dataset serves as the basis for further analysis and exploration of the correlations and causality between the HDI, SDGs, and other measurement variables.

2.3 Clustering analysis

To comprehensively examine the relationships between HDI and SDGs, we propose the following methodology. Utilizing the previously reduced dataset, we employ a clustering and dimensional reduction technique to extract latent variables. These variables are then correlated at two different levels: (a) correlations between the latent variables obtained through dimensional reduction and the SDG variables, and (b) correlations between the latent variables and the HDI variables. By doing so, we aim to identify causal relationships between the variables and ultimately uncover explicit associations between the variables of SDGs and HDI.

For the task of dimensionality reduction, we employ the Uniform Manifold Approximation & Projection (UMAP) technique (McInnes et al., 2018; Becht et al., 2019). Dimensionality reduction is a powerful tool used by machine learning practitioners to visualize and comprehend large, high-dimensional datasets. While t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008; Wattenberg et al., 2016) is a widely used technique for visualization, it faces challenges when applied to large datasets and may not deliver optimal performance. In contrast, UMAP, introduced by McInnes et al., offers several advantages over t-SNE, including faster processing speed and better preservation of the overall data structure.

UMAP operates by constructing a topological representation of the high-dimensional data, modeling it as a fuzzy topological simplicial set. It initiates the process by constructing a weighted k-nearest neighbor graph from the data and subsequently optimizes the embedding in a low-dimensional space, aiming to preserve the topological structure. The optimization process involves minimizing the cross-entropy between the fuzzy simplicial set representation of the high-dimensional data and the low-dimensional embedding.

The UMAP algorithm has gained significant popularity due to its ability to capture intricate patterns and structures present in the data, all while providing rapid computation times in comparison to other dimensionality reduction techniques such as t-SNE. Its effectiveness has been demonstrated across various domains, including image analysis (Vermeulen et al., 2021; Giraldo et al., 2021; Ternes et al., 2022), genomics (Dorrity et al., 2020; Diaz-Papkovich et al., 2019), natural language processing (Low et al., 2020; Tynecki et al., 2020), and the visualization of high-dimensional datasets (Allaoui et al., 2020; Probst & Reymond, 2020), among others.

2.4 Factor analysis

Factor analysis is a statistical method employed to explore the relationships and variability among observed and correlated variables by identifying a smaller set of unobserved variables known as factors. It aims to determine whether variations in the observed variables can be attributed to variations in the underlying factors. In essence, factor analysis seeks to uncover the shared variance between variables and explain it through latent variables. The observed variables are modeled as linear combinations of these latent factors, along with error terms, making factor analysis a type of error model for variables (Jöreskog, 1983).

The factor loading of a variable represents the extent to which the variable is associated with a specific factor (Bandalos, 2018). By examining the patterns of factor loadings, factor analysis provides insights into the underlying structure of the data and can aid in reducing the dimensionality of a dataset. This makes it a valuable tool in various fields such as psychometrics, personality psychology, biology, marketing, product management, operations research, finance, and machine learning. Factor analysis is particularly useful when working with datasets that consist of numerous observed variables that are believed to reflect a smaller number of latent variables. It is widely utilized as an interdependence technique to uncover the latent factors that contribute to the shared variation among the variables.

3 Results and discussion

The UMAP algorithm was employed to perform dimensional reduction on the high-dimensional dataset, aiming to uncover its underlying structure and facilitate visualization. Prior to dimensional reduction, several preprocessing techniques were applied to ensure data quality. The preprocessing phase involved implementing a workflow that addressed missing values in the data. Specifically, the "mean" strategy was utilized to impute missing values by calculating the mean of the available data. Additionally, the SplineTransformer was applied to adjust the shape and distribution of the data, ensuring completeness and adherence to an appropriate distribution prior to dimensional reduction.

Following the preprocessing steps, the preprocessed data underwent dimensional reduction using the UMAP algorithm. UMAP, as a nonlinear dimensional reduction technique, mapped the high-dimensional data to a lower-dimensional space while preserving the inherent topological structure. For this analysis, the desired number of components for the reduction was set to five.

The results obtained from the dimensional reduction using UMAP were highly satisfactory, as demonstrated in Figure 3. The visualization clearly reveals a distinct separation and clustering of the data in the lower-dimensional space, with each cluster representing one of the five generalized components in terms of HDI (as indicated by the vertical colorbar in Figure 3). This outcome indicates that UMAP effectively captured the underlying structures and relationships present in the original data. By reducing the dimensionality of the data, UMAP enabled a more interpretable and comprehensible representation of the dataset, facilitating the identification of patterns and structures within it.

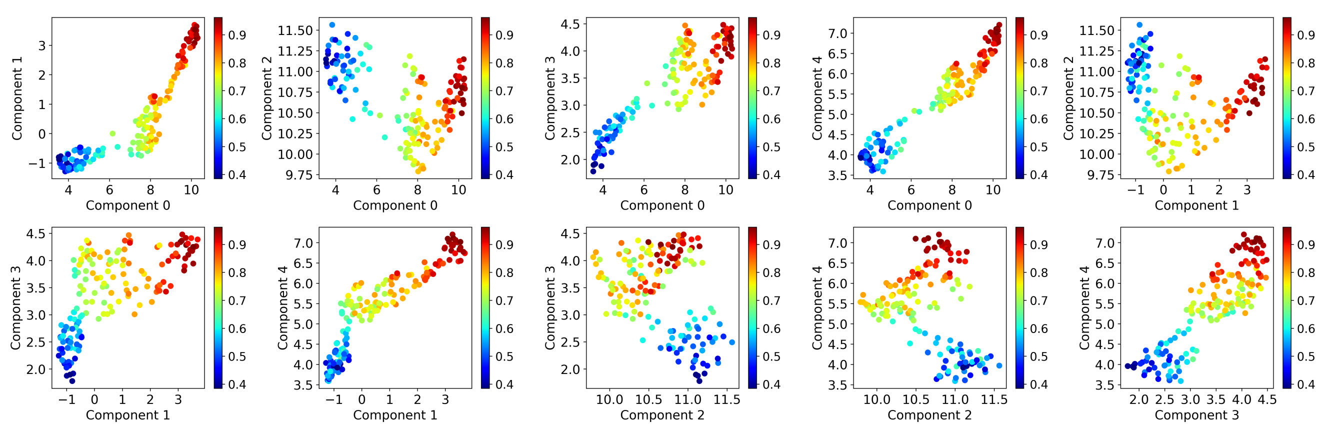

Figure 3: The results of the dimensional reduction with UMAP, a clear separation and clustering of the data in the lowest dimensional space is observed for the five generalized components.

For example from figure 3 we can see that for the case of the generalized components 0 and 1, low values of HDI are highly correlated with low values of both components and also the same case for high values of HDI. A similar situation can be observed with the visualization of the other cross components.

3.1 Analysis of generalized components in terms of HDI

Next, we proceed to determine the regression between HDI and each of the generalized components obtained from the dimensional reduction. The goal is to assess whether the HDI variable can be explained in terms of these components and examine the influence of each component. Initially, it is crucial to establish the linear correlation between each component and HDI. Upon analysis, the correlation coefficients are as follows: 0.959 for C_0, 0.875 for C_1, -0.392 for C_2, 0.868 for C_3, and 0.946 for C_4.

To further explore the relationship, we employ the least squares fit method and summarize the results in Table 1. This analysis provides insights into the regression coefficients, standard errors, t-statistics, and p-values associated with each component. By examining these statistical measures, we can evaluate the significance and strength of the relationship between the components and HDI.

Table 1: Regression analysis by Ordinary Least Squares (OLS) regression model between HDI and generalized components from dimensional reduction. No. Observations: 159, Df Residuals: 153, R-squared: 0.937, Adj. R-squared: 0.935, F-statistic: 458.0, AIC: -571.6, BIC: -553.2.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	0.5725	0.181	3.169	**0.002**	0.216	0.929	*
Component C_0	0.0252	0.009	2.955	**0.004**	0.008	0.042	*
Component C_1	0.0326	0.008	3.905	**0.000**	0.016	0.049	*
Component C_2	-0.0283	0.014	-2.050	**0.042**	-0.056	-0.001	*
Component C_3	0.0508	0.012	4.107	**0.000**	0.026	0.075	
Component C_4	0.0131	0.014	0.931	0.353	-0.015	0.041	

In the regression analysis conducted, an Ordinary Least Squares (OLS) model was employed to explore the relationship between the dependent variable HDI and the independent variables labeled as C_0, C_1, C_2, C_3, and C_4. Let's analyze the key statistics and findings from the regression results. The R-squared value, a measure of the proportion of variance in the dependent variable explained by the independent variables, is 0.937. This indicates that approximately 93.7% of the variability in the Human Development Index can be explained by the regression model. The adjusted R-squared value, accounting for the number of independent variables, is 0.935.

The F-statistic, with a value of 458.0, assesses the overall significance of the regression model. The associated probability (Prob(F-statistic)) is extremely small (4.20e-90), indicating that the model as a whole is highly significant.

Turning to the individual coefficients, they represent the estimated impact of each independent variable on the dependent variable, holding other variables constant. The constant term (constant) has a coefficient of 0.5725, implying that when all independent variables are zero, the predicted value of the Human Development Index is 0.5725. Among the independent variables, C_1 has a coefficient of 0.0326, suggesting that a one-unit increase in C_1 is associated with an estimated increase of 0.0326 in the Human Development Index. This coefficient is statistically significant at a high level of confidence ($P < 0.001$). C_0 has a coefficient of 0.0252, indicating a positive relationship with the Human Development Index, and it is statistically significant at a 0.004 level.

On the other hand, C_2 has a coefficient of -0.0283, implying a negative relationship with the Human Development Index. This coefficient is statistically significant at a 0.042 level. Similarly, C_3 has a coefficient of 0.0508, indicating a positive relationship with the Human Development Index. It is also statistically significant at a high level of confidence ($P < 0.001$).

Lastly, the coefficient for C_4 is 0.0131, suggesting a negligible relationship with the Human Development Index. This coefficient is not statistically significant as the associated p-value ($P > |t|$) is 0.353. The regression analysis indicates that C_0, C_1, C_2, and C_3 have statistically significant relationships with the Human Development Index, while C_4 does not exhibit a significant association.

After performing the dimensional reduction and obtaining the generalized components, we proceeded to analyze the total variance explained by each variable related to the sustainable development goals in relation to these components. Specifically, we focused on the components that were found to be statistically significant in the previous analysis. This involved calculating the total variance explained by each component in terms of the sustainable development goals present in the dataset.

To identify the most influential variables, we applied a filtering process based on the total explained variance. We selected variables that had a total explained variance greater than 0.4. As a result, we identified 21 variables for component C_0, 15 variables for component C_1, 4 variables for component C_2, 11 variables for component C_3, and 12 variables for component C_4.

This filtering approach allowed us to focus on the most significant variables that contributed to the explanation of the sustainable development goals in the context of the dimensional reduction. By considering the variables with higher total explained variance, we aimed to identify the key factors driving the relationship between the generalized components and the sustainable development goals. In this case we calculate the total variance explained between each of the generalized components of the model and the variables from the SDGs. We select the variables whose total explained variance is greater than 0.4 and then we calculate several models for each of the components in terms of the SDGs variables. The results are shown below.

3.2 Analysis of generalized components in terms of SDGs

Based on the observation that the C_i components are significantly related to HDI (see Table 1), we then perform a regression analysis to determine the specific weight of each of them related with sustainable development goals in a linear model. The results of this evaluation can be seen in Tables 2, 3, 4, 5 and 6. Of all the variables considered in the regression for the C_i components that have been selected and filtered because their corresponding explained variances are larger, it is important to highlight that this component can be explained with significant statistics by six goals and their corresponding indicators (see explicit names in appendix A).

Particularly for target 2, appearing significantly in C_0, C_1 and C_4, the indicator "Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)" is an important factor in the model. However, undernourishment and severe food insecurity are factors that can significantly affect human development and people's quality of life (ur Rehman et al., 2022; Burchi et al., 2012; Burchi & De Muro, 2016). Although undernutrition and severe food insecurity are not directly included in the construction of the HDI, they are problems that can have a negative impact on people's health and well-being, and therefore can indirectly influence human development. Lack of access to adequate food can lead to health problems, malnutrition, decreased productivity and difficulties in access to education.

The connection between total official flows directed towards agriculture in developing nations and the UN HDI (Target 2, indicator a.1) can be comprehended by considering their contribution to investment in agricultural development. These funds can be utilized to enhance agricultural productivity, foster value chains, promote the adoption of sustainable practices, ensure food security, and bolster the resilience of rural communities. Moreover, sustainable agricultural development plays a crucial role in advancing human development. By providing resources and support for agriculture in developing countries (Long et al., 2020), numerous benefits can be achieved, including increased food production, the creation of rural employment opportunities, poverty reduction, access to adequate food, improved health and education, among others.

The relationship between aid flows to agriculture and the HDI is intricate and influenced by several factors, such as the effectiveness of fund utilization, governance, national policies, civil society engagement, and other contextual elements, which may vary based on different measurements and targets of the SDGs. However, on the whole, augmenting aid flows to agriculture in developing countries can establish a solid foundation for enhancing human development through the promotion of sustainable agriculture and rural development.

Regarding SDG Target 3 indicators, appearing significantly in C_0, C_3 and C_4, such as Neonatal Mortality Rate (3.2.1), Mortality Rate attributed to cardiovascular diseases, cancer, diabetes or chronic respiratory diseases (3.4.1), and Mortality Rate attributed to unintentional poisoning (3.9.3), they are not directly included in the calculation of the HDI. However, these indicators are related to key aspects of human development and can influence a country's level of development.

The indicators of neonatal mortality (Khazaei et al., 2016a; Anele et al., 2021) and mortality attributed to cardiovascular diseases (Amini et al., 2021; Soares et al., 2016; Masaebi et al., 2021; Costa et al., 2018; Zhu et al., 2016), cancer (Khazaei et al., 2016b), diabetes or chronic respiratory diseases (Goodarzi et al., 2021; Xie et al., 2020) are related to the health dimension of the HDI. A lower neonatal mortality rate and lower mortality attributed to chronic diseases indicate better access to medical care, the quality of health services and the general well-being of the population .

On the other hand, the mortality rate attributed to unintentional poisoning is related to safety and quality of life, which are aspects considered in the decent standard of living dimension of the HDI. Although these indicators are not directly included in the HDI calculation, their improvement is fundamental to human development. By reducing neonatal mortality, controlling and preventing chronic diseases and ensuring the safety of the population, an enabling environment is created for sustainable development and the overall well-being of society.

The relationship between the parity indices (female/male, rural/urban, lower/upper wealth quintile) and other education-related breakdowns (Goal 4, indicator 4.5.1), appearing significantly in C_0, C_1, C_2, C_3 and C_4, and the HDI is noteworthy but indirect. The education-related parity indices disaggregate these indicators based on various factors, including gender, rural/urban areas, and wealth quintiles (Boutayeb & Helmert, 2011; Dancer & Rammohan, 2007). These breakdowns facilitate the examination of disparities in educational access and the involvement of different demographic groups in the education system.

Promoting greater equity in education is crucial for fostering human development, as it guarantees equal opportunities for all individuals to obtain education and unleash their potential (Maniyalath & Narendran, 2016; Mills, 2010). Education issues rightly define one of the fundamental pillars of HDI, but it is interesting that gender disparity (Goal 10, indicator 5.1) in this aspect of education appears in this model as a significant relationship.

In other words, the annual growth rate of real GDP per capita, Goal 8 appearing significantly only in C_0, is directly linked to the UN Human Development Index (HDI) as it relates to the dimension of a decent standard of living. The HDI utilizes per capita income as an indicator to evaluate the economic well-being and standard of living within a population. A higher rate of real GDP per capita growth typically signifies an expansion in economic output and potentially an improvement in income levels and overall economic well-being for individuals.

Finally, the relationship between the variables of the SDG indicators and the UN HDI can be understood by considering the following points, particularly in relation to Target 12, appearing significantly in C_0 and C_3. The indicators mentioned, such as materials footprint, materials footprint per capita, materials footprint per GDP, per capita waste generation, and recycling rates, are all connected to resource consumption, efficiency in material use, and waste management within the economy. Achieving a lower materials footprint per capita and per unit of GDP indicates a more efficient use of resources and greater environmental sustainability (Hickel, 2020; Jiang et al., 2022). This, in turn, can have indirect effects on the HDI, as sustainable resource management contributes to environmental protection, ecosystem conservation, and climate change mitigation, ultimately leading to improved quality of life.

Similarly, lower per capita waste generation and higher rates of recycling reflect better waste management practices and a more sustainable approach to resource utilization. This can positively impact the HDI by reducing environmental pollution, promoting public health, and mitigating the negative effects of waste on communities. Additionally, higher recycling rates can stimulate job creation in the circular economy sector and foster sustainable economic development.

4 Conclusions

This paper aimed to explore the correlation and causality between the Human Development Index and various national economic, water, energy, and other variables from the Sustainable Development Goals measurement framework. A unique approach was adopted by using machine learning and dimensional reduction techniques to establish latent variables and analyze their relationships with the HDI and SDGs.

The results obtained from the dimensional reduction using the UMAP algorithm provided valuable insights into the underlying structure of the data. The visualization of the data in a lower-dimensional space revealed clear separation and clustering of the data for the five generalized components. This indicated that the UMAP algorithm effectively captured the intrinsic structures and relationships present in the original data.

The regression analysis between the HDI and the generalized components derived from dimensional reduction yielded significant findings. The analysis demonstrated that the components C_0, C_1, C_2, and C_3 had statistically significant relationships with the HDI, while C_4 did not exhibit a significant association. The regression results indicated that approximately 93.7% of the variability in the HDI could be explained by the regression model, with C_0 and C_1 contributing the most to the explanation.

Furthermore, the analysis of the generalized components in terms of the SDGs revealed the key variables that contributed to the explanation of the sustainable development goals. Several variables were found to be significantly associated with the generalized components, indicating their influence on the relationship between the components and the SDGs. For example, indicators related to food security, agricultural development, health, and education showed significant relationships with the components.

This study provided valuable insights into the correlation and causality between the HDI, SDGs, and other variables. The use of machine learning and dimensional reduction techniques allowed for a comprehensive analysis of the data, uncovering meaningful patterns and relationships. The findings can contribute to a better understanding of the factors influencing human development and the interplay between different dimensions of sustainable development.

It is important to note that this study has some limitations. The analysis was based on available data, and the results are subject to the quality and completeness of the data sources. Additionally, the causality inferred from the regression analysis should be interpreted with caution, as it is based on observational data. Further research and analysis are needed to validate and expand upon these findings.

With this contribution we believe we complement to the existing literature by providing a unique approach to analyzing the correlation and causality between the HDI, SDGs, and other variables. The findings highlight the importance of considering multiple dimensions of sustainable development and their interconnectedness. The insights gained from this study can inform policymakers and stakeholders in designing and implementing strategies for promoting human development and achieving the SDGs.

A Descriptions of SDGs code names used in this research

In this section we explicitly put the names of the SDG targets and indicators that have been important in describing the results of this research. For a more complete list please check: https://sdgs.un.org/goals.

- Goal 2 - End hunger, achieve food security and improved nutrition and promote sustainable agriculture.
  - Indicator 2.1.1_SN_ITK_DEFC: Prevalence of undernourishment.
  - 2.a.1_AG_PRD_ORTIND: The agriculture orientation index for government expenditures.
  - 2.a.1_AG_XPD_AGSGB: Total official flows (official development assistance plus other official flows) to the agriculture sector.
- Goal 3 - Ensure healthy lives and promote well-being for all at all ages.
  - 3.2.2_SH_DYN_NMRT: Neonatal Mortality Rate.
  - 3.4.1_SH_DTH_NCOM: Mortality Rate Attributed to Cardiovascular Disease, Cancer, Diabetes, or Chronic Respiratory Disease.
  - 3.9.3_SH_STA_POISN: Mortality Rate Attributed to Unintentional Poisoning.
  - 3.3.4_SH_HAP_HBSAG: Hepatitis B surface antigen (HBsAg) prevalence among children.
- Goal 4 - Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.
  - 4.5.1_SE_AGP_CPRA: Parity Indices (Female/Male, Rural/Urban, Bottom/Top Wealth Quintile) for Achievement in Reading and Mathematics at the End of Primary Education.
  - 4.5.1_SE_ALP_CPLR: Parity Indices (Female/Male, Rural/Urban, Bottom/Top Wealth Quintile) for Completion Rates in Education.
  - 4.5.1_SE_AWP_CPRA: Parity Indices (Female/Male, Rural/Urban, Bottom/Top Wealth Quintile) for Early Childhood Development.
- Goal 8 - Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.
  - 8.2.1_SL_EMP_PCAP: Annual Growth Rate of Real GDP per Capita.
- Goal 10 - Reduce inequality within and among countries.
  - 10.5.1_FI_FSI_FSERA: Financial Soundness Indicators for Financial Institutions in the Financial Soundness and Early Warning System.
- Goal 12 - Ensure sustainable consumption and production patterns.
  - 12.2.2_EN_MAT_DOMCMPG: Material Footprint, Material Footprint per Capita, and Material Footprint per GDP.
  - 12.4.2_EN_EWT_GENPCAP: Generation of waste per capita and recycling rates.

B Results for linear regression for generalized components

References

Adam, A., Kammas, P., & Lapatinas, A. (2015). Income inequality and the tax structure: Evidence from developed and developing countries. *Journal of Comparative Economics*, 43(1), 138–154. https://doi.org/https://doi.org/10.1016/j.jce.2014.05.006

Al-Salim, W., Darwish, A. S. K., & Farrell, P. (2022). Analysing data quality frameworks and evaluating the statistical output of united nations sustainable development goals' reports. *Renewable Energy and Environmental Sustainability*, 7, 17. https://doi.org/https://doi.org/10.1051/rees/2022003

Allaoui, M., Kherfi, M. L., & Cheriet, A. (2020). Considerably improving clustering algorithms using umap dimensionality reduction technique: a comparative study. *Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings 9*, 317–325. https://doi.org/https://doi.org/10.1007/978-3-030-51935-3_34

Table 2: Regression analysis by Ordinary Least Squares (OLS) regression model between C_0 generalized component and SDGs variables from dimensional reduction. No. Observations: 159, Df Residuals: 138, R-squared: 0.953, Adj. R-squared: 0.946, F-statistic: 139.9, AIC: 249.6, BIC: 314.0.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	7.4457	0.441	16.869	**0.000**	6.573	8.318	*
10.5.1_FI_FSI_FSERA	-0.1255	0.045	-2.804	**0.006**	-0.214	-0.037	*
12.2.2_EN_MAT_DOMCMPG	0.0779	0.128	0.610	0.543	-0.175	0.331	
12.4.2_EN_EWT_GENPCAP	0.1062	0.014	7.484	**0.000**	0.078	0.134	*
2.1.1_SN_ITK_DEFC	-0.0276	0.007	-3.972	**0.000**	-0.041	-0.014	*
2.a.1_AG_PRD_ORTIND	-0.0034	0.028	-0.122	0.903	-0.059	0.052	
2.a.1_AG_XPD_AGSGB	-0.0246	0.019	-1.271	0.206	-0.063	0.014	
2.c.1_AG_FPA_CFPI	-0.0800	0.054	-1.477	0.142	-0.187	0.027	
3.2.2_SH_DYN_NMRT	-0.0656	0.008	-7.897	**0.000**	-0.082	-0.049	*
3.3.1_SH_HIV_INCD	-0.0185	0.032	-0.572	0.568	-0.082	0.045	
3.3.4_SH_HAP_HBSAG	-0.0389	0.023	-1.683	0.095	-0.085	0.007	
3.4.1_SH_DTH_NCOM	0.0215	0.011	2.034	**0.044**	0.001	0.042	*
3.6.1_SH_STA_TRAF	-0.0151	0.008	-1.904	0.059	-0.031	0.001	
3.8.2_SH_XPD_EARN25	-0.0202	0.019	-1.049	0.296	-0.058	0.018	
3.9.3_SH_STA_POISN	-0.2107	0.064	-3.278	**0.001**	-0.338	-0.084	*
4.5.1_SE_AGP_CPRA	0.2813	0.369	0.763	0.447	-0.448	1.011	
4.5.1_SE_ALP_CPLR	-0.5998	0.319	-1.878	0.062	-1.231	0.032	
4.5.1_SE_AWP_CPRA	1.0145	0.314	3.227	**0.002**	0.393	1.636	*
6.b.1_ER_H2O_RURP	0.0674	0.049	1.371	0.173	-0.030	0.165	
7.3.1_EG_EGY_PRIM	0.0057	0.016	0.353	0.724	-0.026	0.038	
8.2.1_SL_EMP_PCAP	0.0479	0.018	2.614	**0.010**	0.012	0.084	*

Amaluddin, A., Payapo, R. W., Laitupa, A. A., Serang, M. R., et al. (2018). A modified human development index and poverty in the villages of west seram regency, maluku province, indonesia. *International journal of economics and financial issues*, 8(2), 325–330.

Amini, M., Zayeri, F., & Salehi, M. (2021). Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *BMC Public Health*, 21(1), 1–12. https://doi.org/https://doi.org/10.1186/s12889-021-10429-0

Anele, C. R., Hirakata, V. N., Goldani, M. Z., & da Silva, C. H. (2021). The influence of the municipal human development index and maternal education on infant mortality: an investigation in a retrospective cohort study in the extreme south of brazil. *BMC Public Health*, 21(1), 1–12. https://doi.org/https://doi.org/10.1186/s12889-021-10226-9

Arriani, R. et al. (2021). *The Correlation of SDG 1 and 8 and Spatial Effect of Human Development Index in Central Java*, volume 940-1. IOP Publishing. https://doi.org/http://doi.org/10.1088/1755-1315/940/1/012063

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W., Ng, L. G., Ginhoux, F., & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using umap. *Nature biotechnology*, 37(1), 38–44. https://doi.org/https://doi.org/10.1038/nbt.4314

Blum, M. (2013). The influence of inequality on the standard of living: Worldwide anthropometric evidence from the 19th and 20th centuries. *Economics & Human Biology*, 11(4), 436–452. https://doi.org/https://doi.org/10.1016/j.ehb.2012.12.002

Boutayeb, A. & Helmert, U. (2011). Social inequalities, regional disparities and health inequity in north african countries. *International Journal for Equity in Health*, 10, 1–9. https://doi.org/https://doi.org/10.1186/1475-9276-10-23

Burchi, F. & De Muro, P. (2016). From food availability to nutritional capabilities: Advancing food security analysis. *Food policy*, 60, 10–19. https://doi.org/https://doi.org/10.1016/j.foodpol.2015.03.008

Burchi, F., De Muro, P., et al. (2012). A human development and capability approach to food security: Conceptual framework and informational basis. *Background paper*, 8. https://doi.org/https://www.undp.org/sites/g/files/zskgke326/files/migration/africa/Capability-Approach-Food-Security.pdf

Carvalho, L., Mackay, E. B., Cardoso, A. C., Baattrup-Pedersen, A., Birk, S., Blackstock, K. L., Borics, G., Borja, A., Feld, C. K., Ferreira, M. T., et al. (2019). Protecting and restoring europe's waters: An analysis of the

Table 3: Regression analysis by Ordinary Least Squares (OLS) regression model between C_1 generalized component and SDGs variables from dimensional reduction. No. Observations: 159, Df Residuals: 144, R-squared: 0.690, Adj. R-squared: 0.660, F-statistic: 22.86, AIC: 427.9, BIC: 473.9.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	0.7414	0.492	1.507	0.134	-0.231	1.714	
10.5.1_FI_FSI_FSERA	-0.1978	0.079	-2.507	**0.013**	-0.354	-0.042	*
12.2.2_EN_MAT_DOMCMPG	-0.3309	0.216	-1.530	0.128	-0.758	0.096	
2.a.1_AG_PRD_ORTIND	0.1369	0.047	2.942	**0.004**	0.045	0.229	*
2.a.1_AG_XPD_AGSGB	-0.0542	0.033	-1.644	0.102	-0.119	0.011	
2.c.1_AG_FPA_CFPI	0.1267	0.094	1.345	0.181	-0.059	0.313	
3.3.1_SH_HIV_INCD	-0.0670	0.054	-1.241	0.217	-0.174	0.040	
3.3.4_SH_HAP_HBSAG	0.0056	0.039	0.142	0.887	-0.072	0.083	
3.8.2_SH_XPD_EARN25	-0.0358	0.034	-1.058	0.292	-0.103	0.031	
3.9.3_SH_STA_POISN	-0.0916	0.103	-0.892	0.374	-0.295	0.111	
4.5.1_SE_AGP_CPRA	-1.1372	0.623	-1.826	0.070	-2.368	0.094	
4.5.1_SE_ALP_CPLR	0.0966	0.538	0.180	0.858	-0.967	1.160	
4.5.1_SE_AWP_CPRA	3.2856	0.475	6.923	**0.000**	2.348	4.224	*
6.b.1_ER_H2O_RURP	-0.0789	0.084	-0.945	0.346	-0.244	0.086	
7.3.1_EG_EGY_PRIM	0.0293	0.028	1.032	0.304	-0.027	0.085	

Table 4: Regression analysis by Ordinary Least Squares (OLS) regression model between C_2 generalized component and SDGs variables from dimensional reduction. No. Observations: 159, Df Residuals: 155, R-squared: 0.229, Adj. R-squared: 0.215, F-statistic: 15.38, AIC: 139.1, BIC: 151.4.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	10.9723	0.140	78.202	**0.000**	10.695	11.250	*
4.5.1_SE_AGP_CPRA	-0.9606	0.220	-4.359	**0.000**	-1.396	-0.525	*
4.5.1_SE_ALP_CPLR	1.1613	0.211	5.516	**0.000**	0.745	1.577	*
4.5.1_SE_AWP_CPRA	-0.6350	0.162	-3.917	**0.000**	-0.955	-0.315	*

future development needs of the water framework directive. *Science of the Total Environment*, 658, 1228–1238. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.12.255

Cooke, M., Mitrou, F., Lawrence, D., Guimond, E., & Beavon, D. (2007). Indigenous well-being in four countries: an application of the undp's human development index to indigenous peoples in australia, canada, new zealand, and the united states. *BMC international health and human rights*, 7(1), 1–11. https://doi.org/https://doi.org/10.1186/1472-698X-7-9

Costa, S. d. M., Lima, C. d. A., Nobre, A. L. C. S. D., Vieira, D. d. M. A., & Leal, A. L. R. (2018). Hypertension bearers with high risk/big risk of cardiovascular diseases and socioeconomic and health indicators. *Revista da Associação Médica Brasileira*, 64, 601–610. https://doi.org/https://doi.org/10.1590/1806-9282.64.07.601

Dalampira, E.-S. & Nastis, S. A. (2020). Mapping sustainable development goals: A network analysis framework. *Sustainable Development*, 28(1), 46–55. https://doi.org/https://doi.org/10.1002/sd.1964

Dancer, D. & Rammohan, A. (2007). Determinants of schooling in egypt: The role of gender and rural/urban residence. *Oxford Development Studies*, 35(2), 171–195. https://doi.org/https://doi.org/10.1080/13600810701322041

Datta, S. K. & Singh, K. (2019). Variation and determinants of financial inclusion and their association with human development: A cross-country analysis. *IIMB Management Review*, 31(4), 336–349. https://doi.org/https://doi.org/10.1016/j.iimb.2019.07.013

Dawes, J. H. (2020). Are the sustainable development goals self-consistent and mutually achievable? *Sustainable Development*, 28(1), 101–117. https://doi.org/https://doi.org/10.1002/sd.1975

Diaz-Papkovich, A., Anderson-Trocmé, L., Ben-Eghan, C., & Gravel, S. (2019). Umap reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS genetics*, 15(11), e1008432. https://doi.org/https://doi.org/10.1371/journal.pgen.1008432

Diaz-Sarachaga, J. M., Jato-Espino, D., & Castro-Fresno, D. (2018). Is the sustainable development goals (sdg)

Table 5: Regression analysis by Ordinary Least Squares (OLS) regression model between C_3 generalized component and SDGs variables from dimensional reduction. No. Observations: 159, Df Residuals: 148, R-squared: 0.803, Adj. R-squared: 0.789, F-statistic: 60.23, AIC: 116.4, BIC: 150.2.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	3.6420	0.170	21.449	**0.000**	3.306	3.978	*
10.5.1_FI_FSI_FSERA	-0.0186	0.029	-0.636	0.526	-0.076	0.039	
12.2.2_EN_MAT_DOMCMPG	-0.2539	0.073	-3.468	**0.001**	-0.399	-0.109	*
2.c.1_AG_FPA_CFPI	-0.0620	0.035	-1.760	0.080	-0.132	0.008	
3.3.1_SH_HIV_INCD	-0.0197	0.020	-0.973	0.332	-0.060	0.020	
3.3.4_SH_HAP_HBSAG	-0.0414	0.015	-2.817	**0.006**	-0.070	-0.012	*
3.9.3_SH_STA_POISN	-0.3053	0.039	-7.926	**0.000**	-0.381	-0.229	*
4.5.1_SE_AGP_CPRA	0.1979	0.231	0.857	0.393	-0.258	0.654	
4.5.1_SE_ALP_CPLR	-0.1091	0.204	-0.536	0.593	-0.511	0.293	
4.5.1_SE_AWP_CPRA	0.4950	0.176	2.812	**0.006**	0.147	0.843	*
6.b.1_ER_H2O_RURP	-0.0031	0.031	-0.099	0.921	-0.065	0.059	

Table 6: Regression analysis by Ordinary Least Squares (OLS) regression model between C_4 generalized component and SDGs variables from dimensional reduction. No. Observations: 159, Df Residuals: 147, R-squared: 0.799, Adj. R-squared: 0.784, F-statistic: 53.19, AIC: 218.7, BIC: 255.6.

Variable	**coefficient**	**std error**	**t-statistic**	$P > \|t\|$	**[0.025**	**0.975]**	**Signif.**
const	5.8719	0.237	24.727	**0.000**	5.403	6.341	*
10.5.1_FI_FSI_FSERA	-0.0788	0.040	-1.963	0.052	-0.158	0.001	
12.2.2_EN_MAT_DOMCMPG	-0.1699	0.102	-1.666	0.098	-0.371	0.032	
2.a.1_AG_XPD_AGSGB	-0.0452	0.017	-2.678	**0.008**	-0.079	-0.012	*
2.c.1_AG_FPA_CFPI	-0.0272	0.049	-0.561	0.576	-0.123	0.069	
3.3.1_SH_HIV_INCD	-0.0222	0.028	-0.795	0.428	-0.077	0.033	
3.3.4_SH_HAP_HBSAG	-0.0262	0.020	-1.289	0.200	-0.066	0.014	
3.9.3_SH_STA_POISN	-0.2717	0.053	-5.112	**0.000**	-0.377	-0.167	*
4.5.1_SE_AGP_CPRA	-0.2142	0.318	-0.674	0.501	-0.842	0.414	
4.5.1_SE_ALP_CPLR	-0.5733	0.280	-2.047	**0.042**	-1.127	-0.020	*
4.5.1_SE_AWP_CPRA	1.9438	0.246	7.886	**0.000**	1.457	2.431	*
6.b.1_ER_H2O_RURP	-0.0643	0.044	-1.476	0.142	-0.151	0.022	

index an adequate framework to measure the progress of the 2030 agenda? *Sustainable Development*, 26(6), 663–671. https://doi.org/https://doi.org/10.1002/sd.1735

Division, U. N. P. (2002). *International migration report 2002*. https://digitallibrary.un.org/record/490239/files/InternationalMigration2019_Report.pdf?ln=en

Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., & Trapnell, C. (2020). Dimensionality reduction by umap to visualize physical and genetic interactions. *Nature communications*, 11(1), 1537. https://doi.org/https://doi.org/10.1038/s41467-020-15351-4

Giraldo, N. A., Berry, S., Becht, E., Ates, D., Schenk, K. M., Engle, E. L., Green, B., Nguyen, P., Soni, A., Stein, J. E., et al. (2021). Spatial umap and image cytometry for topographic immuno-oncology biomarker discoveryspatial umap and image cytometry. *Cancer immunology research*, 9(11), 1262–1269. https://doi.org/https://doi.org/10.1158/2326-6066.CIR-21-0015

Goodarzi, E., Sohrabivafa, M., Darvishi, I., Naemi, H., & Khazaei, Z. (2021). Epidemiology of mortality induced by acute respiratory infections in infants and children under the age of 5 years and its relationship with the human development index in asia: an updated ecological study. *Journal of Public Health*, 29, 1047–1054. https://doi.org/https://doi.org/10.1007/s10389-020-01195-1

Greene, V. W. (2001). Personal hygiene and life expectancy improvements since 1850: Historic and epidemiologic associations. *American Journal of Infection Control*, 29(4), 203–206. https://doi.org/https://doi.org/10.1067/mic.2001.115686

Hagel, C., Paton, C., Mbevi, G., & English, M. (2020). Data for tracking sdgs: challenges in capturing neonatal data from hospitals in kenya. *BMJ Global Health*, 5(3), e002108. https://doi.org/http://dx.doi.org/10.1136/bmjgh-2019-002108

Hickel, J. (2020). The sustainable development index: Measuring the ecological efficiency of human development in the anthropocene. *Ecological Economics*, 167, 106331. https://doi.org/https://doi.org/10.1016/j.ecolecon.2019.05.011

Hidayat, E. W. & Bariyah, N. (2020). Human development index, unemployment and poverty rate in kalimantan barat. *Seminar Akad Tahun Ilmu Ekonomi dan Studi Pembangunan (SATIESP 2020). Published online*, 12–23.

Hou, J., Walsh, P. P., & Zhang, J. (2015). The dynamics of human development index. *The Social Science Journal*, 52(3), 331–347. https://doi.org/https://doi.org/10.1016/j.soscij.2014.07.003

Hukom, A. (2015). Effect of capital expenditures, economic growth and poverty on human development in central kalimantan. *Iosrjournals*, 6(6), 19–27.

Jain, P. & Jain, P. (2020). Are the sustainable development goals really sustainable? a policy perspective. *Sustainable Development*, 28(6), 1642–1651. https://doi.org/https://doi.org/10.1002/sd.2112

Janoušková, S., Hák, T., & Moldan, B. (2018). Global sdgs assessments: Helping or confusing indicators? *Sustainability*, 10(5), 1540. https://doi.org/https://doi.org/10.3390/su10051540

Jiang, M., Behrens, P., Yang, Y., Tang, Z., Chen, D., Yu, Y., Liu, L., Gong, P., Zhu, S., Zhou, W., et al. (2022). Different material footprint trends between china and the world in 2007-2012 explained by construction-and manufacturing-associated investment. *One Earth*, 5(1), 109–119. https://doi.org/https://doi.org/10.1016/j.oneear.2021.12.011

Jöreskog, K. G. (1983). Factor analysis as an error-in-variables model. *Principles of modern psychological measurement*, 185–196.

Khazaei, S., Ayubi, E., Nematollahi, S., & Khazaei, S. (2016a). Variations of infant and under-five child mortality rates around the world, the role of human development index (hdi). *Int J Pediatr*, 4(5), 1671–1677.

Khazaei, S., Rezaeian, S., Soheylizad, M., Khazaei, S., & Biderafsh, A. (2016b). Global incidence and mortality rates of stomach cancer and the human development index: an ecological study. *Asian Pacific Journal of Cancer Prevention*, 17(4), 1701–1704. https://doi.org/http://doi.org/10.15167/2421-4248/jpmh2021.62.1.1606

Kirschke, S., Avellán, T., Bärlund, I., Bogardi, J. J., Carvalho, L., Chapman, D., Dickens, C. W., Irvine, K., Lee, S., Mehner, T., et al. (2020). Capacity challenges in water quality monitoring: understanding the role of human development. *Environmental monitoring and assessment*, 192(5), 1–16. https://doi.org/https://doi.org/10.1007/s10661-020-8224-3

Lee, B. X., Kjaerulf, F., Turner, S., Cohen, L., Donnelly, P. D., Muggah, R., Davis, R., Realini, A., Kieselbach, B., MacGregor, L. S., et al. (2016). Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy*, 37(1), 13–31. https://doi.org/https://doi.org/10.1057/s41271-016-0002-7

Lim, S. S., Allen, K., Bhutta, Z. A., Dandona, L., Forouzanfar, M. H., Fullman, N., Gething, P. W., Goldberg, E. M., Hay, S. I., Holmberg, M., et al. (2016). Measuring the health-related sustainable development goals in 188 countries: a baseline analysis from the global burden of disease study 2015. *The Lancet*, 388(10053), 1813–1850. https://doi.org/https://doi.org/10.1016/S0140-6736(16)31467-2

Long, X., Yu, H., Sun, M., Wang, X.-C., Klemeš, J. J., Xie, W., Wang, C., Li, W., & Wang, Y. (2020). Sustainability evaluation based on the three-dimensional ecological footprint and human development index: A case study on the four island regions in china. *Journal of Environmental Management*, 265, 110509. https://doi.org/https://doi.org/10.1016/j.jenvman.2020.110509

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research*, 22(10), e22635. https://doi.org/https://doi.org/10.2196/22635

MacFeely, S. (2020). Measuring the sustainable development goal indicators: An unprecedented statistical challenge. *Journal of official statistics*, 36(2), 361–378. https://doi.org/https://doi.org/10.2478/jos-2020-0019

Maniyalath, N. & Narendran, R. (2016). The human development index predicts female entrepreneurship rates. *International Journal of Entrepreneurial Behavior & Research*, 22(5), 745–766. https://doi.org/https://doi.org/10.1108/IJEBR-11-2015-0258

Masaebi, F., Salehi, M., Kazemi, M., Vahabi, N., Azizmohammad Looha, M., & Zayeri, F. (2021). Trend analysis of disability adjusted life years due to cardiovascular diseases: results from the global burden of disease study 2019. *BMC Public Health*, 21, 1–13. https://doi.org/https://doi.org/10.1186/s12889-021-11348-w

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Mehta, L. (2014). Water and human development. *World development*, 59, 59–69. https://doi.org/https://doi.org/10.1016/j.worlddev.2013.12.018

Mills, M. (2010). Gender roles, gender (in) equality and fertility: An empirical test of five gender equity indices. *Canadian Studies in Population [ARCHIVES]*, 37(3-4), 445–474. https://doi.org/https://doi.org/10.25336/P6131Q

Mitrou, F., Cooke, M., Lawrence, D., Povah, D., Mobilia, E., Guimond, E., & Zubrick, S. R. (2014). Gaps in indigenous disadvantage not closing: a census cohort study of social determinants of health in australia, canada, and new zealand from 1981–2006. *BMC Public Health*, 14, 1–9. https://doi.org/https://doi.org/10.1186/1471-2458-14-201

Nundy, S., Kakar, A., Bhutta, Z. A., Nundy, S., Kakar, A., & Bhutta, Z. A. (2022). The status of biomedical research in some developing countries. *How to Practice Academic Medicine and Publish from Developing Countries? A Practical Guide*, 47–62. https://doi.org/https://doi.org/10.1007/978-981-16-5248-6_6

Otekunrin, O. A., Otekunrin, O. A., Sawicka, B., & Ayinde, I. A. (2020). Three decades of fighting against hunger in africa: Progress, challenges and opportunities. *World Nutrition*, 11(3), 86–111. https://doi.org/https://doi.org/10.26596/wn.202011386-111

Ouedraogo, N. S. (2013). Energy consumption and human development: Evidence from a panel cointegration and error correction model. *Energy*, 63, 28–41. https://doi.org/https://doi.org/10.1016/j.energy.2013.09.067

Probst, D. & Reymond, J.-L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics*, 12(1), 1–13. https://doi.org/https://doi.org/10.1186/s13321-020-0416-x

Ranis, G., Stewart, F., & Ramirez, A. (2000). Economic growth and human development. *World development*, 28(2), 197–219. https://doi.org/https://doi.org/10.1016/S0305-750X(99)00131-X

Ranis, G., Stewart, F., & Samman, E. (2006). Human development: beyond the human development index. *Journal of Human Development*, 7(3), 323–358. https://doi.org/https://doi.org/10.1080/14649880600815917

Sachs, J., Schmidt-Traub, G., Kroll, C., Durand-Delacre, D., & Teksoz, K. (2016). *Sdg index & dashboards: A global report*. https://s3.amazonaws.com/sustainabledevelopment.report/2016/2016_sdg_index_and_dashboards_report.pdf

Sagar, A. D. & Najam, A. (1998). The human development index: a critical review. *Ecological economics*, 25(3), 249–264. https://doi.org/https://doi.org/10.1016/S0921-8009(97)00168-7

Sarkodie, S. A. & Adams, S. (2020). Electricity access, human development index, governance and income inequality in sub-saharan africa. *Energy Reports*, 6, 455–466. https://doi.org/https://doi.org/10.1016/j.egyr.2020.02.009

Soares, G. P., Klein, C. H., Silva, N. A. d. S., & Oliveira, G. M. M. d. (2016). Progression of mortality due to diseases of the circulatory system and human development index in rio de janeiro municipalities. *Arquivos brasileiros de cardiologia*, 107, 314–322. https://doi.org/https://doi.org/10.5935/abc.20160141

Ternes, L., Dane, M., Gross, S., Labrie, M., Mills, G., Gray, J., Heiser, L., & Chang, Y. H. (2022). A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. *Communications biology*, 5(1), 255. https://doi.org/https://doi.org/10.1038/s42003-022-03218-x

Tynecki, P., Guziński, A., Kazimierczak, J., Jadczuk, M., Dastych, J., & Onisko, A. (2020). Phageai-bacteriophage life cycle recognition with machine learning and natural language processing. *BioRxiv*, 2020–07. https://doi.org/https://doi.org/10.1101/2020.07.11.198606

Ul Haq, M. et al. (1992). *Human development in a changing world*. https://core.ac.uk/download/pdf/6248876.pdf

Ul Haq, M. & United Nations Development Programme, U. (1989). *Human development report*. UnitedNationsDevelopmentProgramme,https://hdr.undp.org/system/files/documents/hdr1990encompletenostatspdf.pdf

UNDP, U. N. D. P. (2019). *Human development report 2019*. https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf

UNDP, U. N. D. P. (2020). *The sustainable development goals report 2020*. https://unstats.un.org/sdgs/report/2020/The-Sustainable-Development-Goals-Report-2020.pdf

UNDP, U. N. D. P. (2022). *Human development report 2021-22*. https://hdr.undp.org/system/files/documents/global-report-document/hdr2021-22pdf_1.pdf

UNGA, 71st session 2016-2017 2nd Committee, U. N. G. A. (2016). *Operational activities for development: operational activities for development of the united nations system. report of the 2nd committee. general assembly, 71st session*. https://digitallibrary.un.org/record/1291226/files/A_RES_71_313-EN.pdf?ln=en

UNGA, 71st session 2016-2017 2nd Committee, U. N. G. A. (2017). *Work of the statistical commission pertaining*

to the 2030 agenda for sustainable development : resolution / adopted by the general assembly. https://digitallibrary.un.org/record/1291226/files/A_RES_71_313-EN.pdf?ln=en
ur Rehman, M., Ahmed, A., Khan, A., & Khan, A. (2022). The role of poverty, food security, and rapid population growth on human development in pakistan. *NUST Journal of Social Sciences and Humanities*, 8(3), 89–105.
van der Maaten, L. & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html
Vermeulen, M., Smith, K., Eremin, K., Rayner, G., & Walton, M. (2021). Application of uniform manifold approximation and projection (umap) in spectral imaging of artworks. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 252, 119547. https://doi.org/https://doi.org/10.1016/j.saa.2021.119547
Wang, Z., Zhang, B., Wang, B., et al. (2018). Renewable energy consumption, economic growth and human development index in pakistan: evidence form simultaneous equation model. *Journal of cleaner production*, 184, 1081–1090. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.02.260
Wattenberg, M., Viégas, F., & Johnson, I. (2016). How to use t-sne effectively. *Distill*. https://doi.org/10.23915/distill.00002
Xie, M., Liu, X., Cao, X., Guo, M., & Li, X. (2020). Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017. *Respiratory research*, 21(1), 1–13. https://doi.org/https://doi.org/10.1186/s12931-020-1291-8
Zhu, K.-F., Wang, Y.-M., Zhu, J.-Z., Zhou, Q.-Y., & Wang, N.-F. (2016). National prevalence of coronary heart disease and its relationship with human development index: a systematic review. *European journal of preventive cardiology*, 23(5), 530–543. https://doi.org/https://doi.org/10.1177/2047487315587402