

Università degli studi Milano Bicocca



Data Science Lab

Project Report:

**Exploring the harmony between Freedom and Happiness worldwide**

Colombaro Daniel 897062

Ferretti Federico 897902

Fracchia Camilla 898235

Salvatori Ilaria 898009

## Abstract

*This research delves into the intricate interplay between political freedom and happiness, with a specific focus on understanding how one influences the other. Leveraging data from the 'Freedom in the World' research conducted by Freedom House and the World Happiness Report spanning a decade (2013-2022), we employ advanced clustering techniques to categorize countries based on their levels of political rights, civil liberties, and overall happiness.*

*The analysis uncovers five distinct clusters for both sets of indicators, providing a comprehensive insight into where countries stand within these dimensions. Interestingly, the study reveals that while some nations witness concurrent improvements in both freedom and happiness, there are instances where these factors do not align.*

*Furthermore, the study illuminates variations in these relationships over time, identifying countries that have undergone similar trajectories in terms of freedom and happiness over the past decade.*

*Despite the challenges presented by the pandemic and global instability, this research offers a nuanced perspective on the complex interaction between these two dimensions, underscoring that they do not always coexist in a straightforward manner.*

## Keywords

*Freedom, Happiness, Cluster analysis, K means, Countries*

---

## Index

|   |    |
|---|----|
| 1. INTRODUCTION.....  | 1  |
| 1.1. Objectives of the Analysis and Research Questions..... | 1  |
| 2. DATA DESCRIPTION.....                                    | 2  |
| 2.1. Freedom in the World .....                             | 2  |
| 2.2. Happiness in the World .....                           | 3  |
| 3. DATA PREPARATION.....                                    | 4  |
| 3.1. Data Cleaning .....                                    | 4  |
| 4. DATA EXPLORATION.....                                    | 6  |
| 4.1. Freedom.....   | 6  |
| 4.2. Happiness.....   | 7  |
| 5. CLUSTERING.....  | 9  |
| 5.1. Choosing clustering technique .....                    | 9  |
| 5.2. K-Means implementation .....                           | 11 |
| 5.2.1. Clustering for Freedom in 2013 and 2022 .....        | 12 |
| 5.2.2. Clustering for Happiness in 2013 and 2022 .....      | 14 |
| 5.2.3. Clustering on Variability.....                       | 16 |
| CONCLUSIONS .....   | 17 |
| BIBLIOGRAPHY AND SITOGRAPHY .....                           | 19 |

## 1. INTRODUCTION

In our research, we explore the relationship between political freedom and happiness, specifically examining how the former influences the latter. To conduct this analysis we utilized data from two reputable resources: the 'Freedom in the World' research conducted by the Freedom House<sup>1</sup> and the World Happiness Report. The objective of this project is to perform clustering analyses on these two data sources separately and in conjunction, while also studying their variations over a ten-year period (2013-2022).

The 'Freedom in the World' research conducted by the Freedom House provides a comprehensive assessment of political freedom and civil liberties in countries worldwide. This assessment is conducted by external analysts who employ a combination of fieldwork, interviews with locals, information from news articles, NGOs, governments, and various other sources. The analyst's conclusions are then scrutinized by knowledgeable advisors and local experts. The consensus among analysts, advisors, and Freedom House staff is reflected in the final report: in the contemporary era, in fact, evaluating the political and civil rights of national communities is a fundamental element in understanding the global landscape of human rights and fundamental freedoms.

On the other hand, the World Happiness Report measures subjective well-being and happiness levels based on a range of factors, including social support, healthy life expectancy, the logarithm GDP per capita, generosity, perceptions of corruption, and freedom to make life choices. Although the World Happiness Report is based on a wide variety of data, the most important source has always been the Gallup World Poll, unique in its range and comparability of global annual surveys. The poll includes over 100 questions that are both universally applicable and region-specific. It covers various global indices, such as

law and order, access to food and shelter, the state of institutions and infrastructure, the availability of good jobs, overall well-being, and the concept of "brain gain." Gallup collaborates with institutions, communities, governments, and nations to design unique products and indices that gather data on specialized topics of interest.

### 1.1. Objectives of the Analysis and Research Questions

Through the analysis of these two data sources, our objectives are to uncover underlying patterns or associations between political freedom and happiness. Additionally, we aim to explore how these relationships have evolved, offering insights into potential trends and changes during the period under review. Our research seeks to answer the following questions:

- How can we categorize countries based on their levels of political rights and civil liberties from 2013 to 2022?
- How can we categorize countries based on their level of happiness in the period spanning from 2013 to 2022?
- How can we jointly cluster countries based on both political rights, civil liberties, and happiness? What is the interplay between these factors?

To achieve these objectives, we applied several clustering techniques including DBScan, Mixture Models, and K-means. Clustering, an unsupervised machine learning method, groups similar data points based on their shared features or characteristics.

By applying clustering algorithms on both individual and combined datasets, we can identify similarities and differences among countries and potentially reveal meaningful clusters that shed light on the relationship between political freedom and happiness.

---

<sup>1</sup> A non-profit organization known for political advocacy surrounding issues of democracy, political freedom, and human rights.

The findings of this project can enhance our understanding of the intricate interplay between freedom and happiness, offering valuable insights for policymakers, researchers, and individuals interested in the social sciences. Furthermore, the results may illuminate variations in these relationships over time, potentially indicating groups of countries that have experienced similar developments in terms of political freedom and happiness over the past decade. Additionally, the clustering algorithms were initially fitted on 2013 data and subsequently applied to 2022 data, to verify whether and to what extent the countries were assigned to different clusters over time. Consequently, for every data source, we analyzed the situation at both the beginning and at the end of the period, as well as clustering for the dynamics across time.

In the following sections of this report, we will delve into the methodology used for data collection, preprocessing, and clustering analysis. We will present and interpret the data and the results obtained through clustering with appropriate charts and visualizations. Finally, we will discuss the implications of our findings, potential limitations of the study, and future research directions.

## 2. DATA DESCRIPTION

Both happiness and political freedom are intricate and multi-faceted concepts with nuances that can vary based on cultural and personal perspectives or background.

To address this complexity, we based our analysis on data from globally recognized organizations that provide clear and transparent frameworks for defining these two factors objectively.

We chose to focus on the 2013-2022 timeframe as both data sources were available for this period. Data on political freedom was available from 2013 onwards, and happiness data for 2023 had not been published at the time of our work.

### 2.1. Freedom in the World

The Freedom in the World Report tracks the development of freedom in 195 countries and 15 territories from 2013 to 2022. The report's methodology uses as a main source the Universal Declaration of Human Rights and operates on the premise that freedom for all people is best realized in liberal democratic societies. Consequently, the report's main objective is to assess political rights and civil liberties worldwide and categorize countries as Free, Partly Free, or Not Free.

This is possible through a scoring process that evaluates responses to questions related to political rights and civil liberties. There are 25 questions in total, representing indicators for 10 political rights and 15 civil liberties. Each country or territory<sup>2</sup> is assigned a score ranging from 0 to 4 for each question, with 0 representing the lowest degree of freedom and 4 signifying the highest degree of freedom. Therefore, the highest overall score that can be awarded for political rights is 40 (4 points for each of the 10 questions), while civil liberties can receive a maximum score of 60 (4 points for each of the 15 questions). Moreover, these questions are grouped into subcategories as follows:

- The political rights questions are grouped into three subcategories:
  - A. Electoral Process:** Three questions about the freedom and fairness of elections and electoral laws;
  - B. Political Pluralism and Participation:** Four questions related to the possibility of freely creating political groups, the fair opportunity to increase power by opposition, free political choices, fair political and electoral opportunities for minorities;
  - C. Functioning of Government:** Three questions concerning the possibility of acting by those freely elected, the presence of effective safeguards against corruption, and government transparency.

---

<sup>2</sup> Territory: explained in paragraph 3.1, see later.

- The civil liberties questions are grouped into these subcategories:
  - D. Freedom of Expression and Belief:** Four questions regarding freedom of expression in the fields of media, religion, the academic world and personal views;
  - E. Associational and Organizational Rights:** Three questions about freedom of assembly and freedom for nongovernmental organizations and labour organizations;
  - F. Rule of Law:** Four questions on judiciary independence, the prevalence of dual process, the existence of protection from the illegitimate use of physical force, ensuring fair treatment of minorities.
  - G. Personal Autonomy and Individual Rights:** Four questions about freedom of movement, right to own property, right to personal social freedom, equality of opportunity and freedom in the economic field.

Finally, each question gives a certain score to a country or territory and the combination of the overall score awarded for political rights and civil liberties, after being equally weighted, is converted into a status (free/partly free/not free) thanks to the following table.

| Status                |       | Political Rights score |      |       |       |       |       |       |
|-----------------------|-------|------------------------|------|-------|-------|-------|-------|-------|
|                       |       | 0-5*                   | 6-11 | 12-17 | 18-23 | 24-29 | 30-35 | 36-40 |
| Civil Liberties score | 53-60 | PF                     | PF   | PF    | F     | F     | F     | F     |
|                       | 44-52 | PF                     | PF   | PF    | PF    | F     | F     | F     |
|                       | 35-43 | PF                     | PF   | PF    | PF    | PF    | F     | F     |
|                       | 26-34 | NF                     | PF   | PF    | PF    | PF    | PF    | F     |
|                       | 17-25 | NF                     | NF   | PF    | PF    | PF    | PF    | PF    |
|                       | 8-16  | NF                     | NF   | NF    | PF    | PF    | PF    | PF    |
|                       | 0-7   | NF                     | NF   | NF    | NF    | PF    | PF    | PF    |

F = Free, PF = Partly Free, and NF = Not Free

Figure 1: Table to assign Status according to weighted scores

The final status of each country or territory is determined by both its aggregate political rights score (on a scale of 0-40) and its aggregate civil liberties score (on a scale of 0-60). In some cases, scores can fall below 0 (ranging from -1 to -4) if a

country or territory receives mostly all zeros for political rights questions and a sufficiently negative score.

This dataset includes the following variables for each country or territory:

- Region of belonging
- Type of location (country or territory)
- Edition of the report from 2013 to 2022
- Status (Free -F-, Partly Free -PF-, Not Free -NF-)
- Political rights (PR) rating from 1 to 7 (according to the table presented)
- Civil liberties (CL) rating from 1 to 7 (according to the table presented)
- Subcategory scores (e.g., A1, A2, A3)
- Total scores for each category (e.g., A)
- The final total score, obtained by summing CL and PR scores

## 2.2. Happiness in the World

The World Happiness Report is a publication of the Sustainable Development Solutions Network<sup>3</sup> that delves into the state of happiness globally. A nationally representative sample of 1,000 individuals from each country is asked life evaluation questions, which they evaluate on a Cantril ladder from 0 (lowest possible standard of livin) to 10 (highest).

The Gallup World Poll questionnaire measures 14 areas within its core questions: business & economics, citizen engagement, communications & technology, diversity (social issues), education & families, emotions (well-being), environment & energy, food & shelter, government and politics, law & order (safety), health, religion & ethics, transportation, and work.

The World Happiness Report condenses these 14 areas into six variables<sup>4</sup>:

<sup>3</sup> An NGO launched by United Nations in 2012 to promote the fulfillment of the sustainable development solutions by UN at a national and international level.

<sup>4</sup> *GDP*: measure of country's economic, taking into account its population; *Social support*: index that take into account presence of

other people that help in times of crisis; *Healthy life expectancy*: average number of years of 'full-health'; *Freedom of making choices*: autonomy; *Generosity*: being kind and generous; *Perception of corruption*: trust in government.

- Logarithm of the real GDP per capita
- Social support
- Healthy life expectancy
- Freedom to make life choices
- Generosity
- Perceptions of corruption

Each country is compared against a fictitious nation called Dystopia. Dystopia is characterized by the worst average for each one of the six indicators and it serves as a benchmark for the regression determining the happiness score for all countries. The same approach is used for residuals, reflecting the extent to which the six variables either over- or under-explain average life evaluations from 2014-2016. These residuals have an average value of approximately zero over the whole set of countries. The six metrics do not affect the overall score presented for each country; rather, they are used to explain how much each of these elements is thought to contribute to raising life satisfaction when compared to the hypothetical country of Dystopia.

The use of a single item as an indicator of subjective well-being in the World Happiness Report differs fundamentally from more conventional index approaches that use a variety of indicators, such as the Human Development Index of the United Nations, the Better Life Index of the OECD from 2011 or the Social Progress Index of 2013. The use of single-item and multi-item scales as indicators of life satisfaction has also been the subject of ongoing discussion.

The variables utilized are consistent with what has generally been discovered in the study literature to explain variations in life assessments at the national level. Due to the lack of comparable international statistics for the whole sample of nations, certain noteworthy variables, such as unemployment or inequality, do not show up. Since some of the data are taken from the same survey sources, the variables are meant to show significant correlation rather than to indicate pure causal estimations as some of the data are gathered from the same articles and studies.

### 3. DATA PREPARATION

Upon loading the dataset, the initial step involves assessing the need for preprocessing before delving into a more comprehensive analysis. At this stage, the data will be cleaned and reorganized to achieve better results for our analysis during the modelling phase. Our objective is to merge the two distinct data. However, prior to merging, we address inconsistencies and missing values within each data source separately.

#### 3.1. Data Cleaning

##### Freedom

In the data from the Freedom House, 15 territories were included that are not officially recognized as independent nations<sup>5</sup>. This distinction is justified by a stark difference in political rights and civil liberties compared to the situation of the country to which they belong. A territory is considered separate in the following cases:

- Whether conditions on the ground for political rights and civil liberties are significantly different from those in the rest of the relevant country or countries, meaning a separate assessment is likely to yield different ratings;
- Whether the territory is the subject of enduring popular or diplomatic pressure for autonomy, independence, or incorporation into another country;
- Whether the territory's boundaries are sufficiently stable to allow an assessment of conditions for the year under review,
- Whether they can be expected to remain stable in future years so that year-on-year comparisons are possible;
- Whether the territory is large and/or politically significant.

Freedom House typically maintains a neutral stance on territorial and separatist disputes,

---

<sup>5</sup> An example: Somaliland, with respect to Somalia or South Sudan asks for independence from Sudan.

focusing instead on assessing political rights and civil liberties within specific geographic regions.

Given that the territories are considered separately in the study by the Freedom House, we decided to maintain their independent status for the purposes of our analysis. However, we integrated these territories with international identification codes referring to the country they are associated with. This integration allows for a relative comparison between the territory and the country through a general consensus.

The .csv file with all the codes was obtained by merging:

- the [Wikipedia ISO 3166-1 article](#) for alpha and numeric country codes.

ISO 3166-1 is a standard defining codes for the names of countries, dependent territories, and distinctive locations. It is the first part of the ISO 3166 standard published by the International Organization for Standardization<sup>6</sup> (ISO).

The most popular of the three two-letter country codes, [ISO 3166-1 alpha-2](#), is used (with a few exceptions) for the majority of country code top-level domains on the Internet.

Compared to alpha-2 codes, [ISO 3166-1 alpha-3](#) nation codes provide for a greater visual correlation between the codes and the country names.

With the advantage of script (writing system) independence, [ISO 3166-1 numeric](#) three-digit country codes are identical to those created and maintained by the United Nations and are therefore helpful for persons or systems utilizing non-Latin scripts.

- the [UN Statistics site](#) for countries' regional, and sub-regional codes. It is a dataset provided by the Statistics Department of the

United Nations (UN), known as the M49 standard.

All these codes have been associated with the name of each country. In some cases, where countries have more than one common denomination (e.g. United States and U.S.), we manually imputed a unique denomination to ensure consistency. We encountered this problem upon integration with political freedom data with happiness data as well and we proceeded in the same way. The only territory that was still left without ISO-codes was Kosovo, as it is commonly recognized as an independent State, but it still lacks standard country codes.

The optional question 'Add A' was removed, as it relates to traditional monarchies that have some form of public consultation but no substantial political parties or electoral systems. Given its limited applicability to the majority of countries and the potential for disruptive effects due to a high proportion of missing values, it was deemed best to remove this question.

### Happiness

The happiness score was not measured in all editions for all countries, resulting in the presence of some missing values. We imputed such values with the most probable replacement, by defining a custom rule to substitute the values. After having identified the rows with missing values, we replaced them with the average of the missing variable for the specific country measured in the other editions. If a certain indicator was never measured for a particular country, the missing value remained.

### Transformations on merged data

The two data sources did not share an identical list of countries (and territories), therefore for the purposes of our analysis, we only considered the ones present in both. Redundant and irrelevant columns were removed, such as the national ISO codes (only one version was kept and the other formats were deleted), information about the

---

<sup>6</sup> An independent, non-governmental international organization that brings together experts to share knowledge and develop voluntary, consensus-based, market relevant International Standards that support innovation and provide solutions to global challenges.

intermediate region and duplicates of the country name.

To measure the variation from 2013 to 2022 across the world, we did not want to lose all the information regarding the intermediate years (simply subtracting the data relating to 2013 from those relating to 2022) but rather we tried to make the dynamic component of time 'static'. We achieved this by creating an additional column for each relevant clustering variable and by calculating the percentage change with respect to the previous year. This means that if for a certain country we had 'Log GDP per capita' to be equal to '100' in 2013 and '110' in 2014, 'Log GDP per capita % change' in 2014 would be:

$$(110 - 100) / 100 = 10\%$$

As we are interested in variability, we have calculated the square of all the percentage variations in such a way to avoid large positive and negative changes that compensate each other by bringing the new variable values to almost zero. Consequently, we obtained a measure of the squared cumulative percent change for each relevant variable and country across consecutive years from 2013 to 2022.

## 4. DATA EXPLORATION

The data exploratory phase is crucial to understand what information you are working with. In our case it focused on inspecting the scores of Freedom and Happiness among the different nations available.

In order to facilitate the reader in the understanding of what has been found, we present some of the graphs elaborated almost aggregated for continent and not for single country because of the high number of these last ones in the datasets.

Instead, we leave the faculty to interested readers to explore the individual state situations through the interactive graphs found within the shared Python codes.

### 4.1. Freedom

As elucidated earlier, the Freedom score comprises two primary constituents: political and civil rights. The greater the divergence in scores, the higher a country is rated on the freedom scale, which ranges from 0 to 100.

As expected, *Figure 2* illustrates that the trend in the average score per continent during the analyzed period is higher in Western countries, which have traditionally leaned towards greater freedom. Conversely, Africa and the East (Eurasia and the Middle East) are marked by more authoritarian regimes.

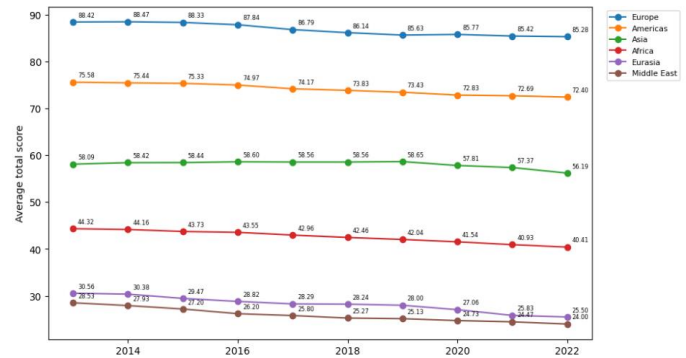


Figure 2: Average total score by continent in each year

Examining the graph, we can observe that, across all continents, there are no significant changes in the average total scores derived from the summation of political and civil liberties scores over the course of these 10 years. For all continents, we note a linear decrease in scores, and considering the 2022 scores, it becomes evident that this year presents the lowest scores among all ten years for all continents. We also observe that the overall score of the Americas is three times that of the Middle East and that the latter has a trend similar to Eurasia. The most efficient way to represent and understand the situation of individual countries is to plot the scores on a choropleth, a thematic map in which areas are colored based on the legend depicted in *Figure 3*: lighter colors correspond to higher scores, indicating freer countries, while darker colors represent more oppressed countries.



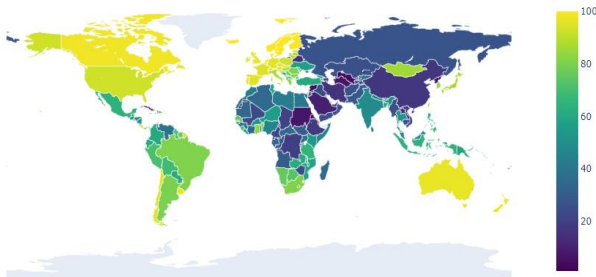


Figure 3: Choropleth of Freedom score, 2013

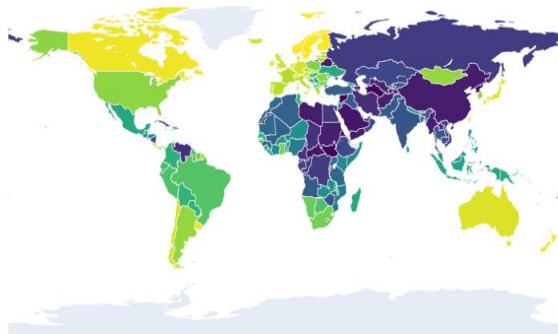


Figure 4: Choropleth of Freedom score, 2022

In North America, we observe that the USA has experienced a decrease in its total score. In 2022, it is represented in green on the map, whereas in 2013, the color of the state tended more towards yellow, so we can assert that the situation reflects the trend shown in *Figure 2*: total freedom in individual countries, as well as in continents, has decreased over the course of a decade.

Comparing *Figure 3* with *Figure 4*, it's evident that the majority of countries are shaded with a darker hue. This is once again indicative of a decline in the final score, the causes of which are not straightforward and easy to point out. However, it is likely that political tensions related to conflicts have played a fundamental role in perceiving reduced freedom.

The most significant deterioration, as one might expect, is observed in the most unstable countries: in Central Africa, the Middle East, and Asia, we can see a dense cloud of dark blue, notably larger than the one of 2013, reflecting a score below 20 out of 100.

## 4.2. Happiness

Happiness is a profoundly significant and intricate indicator of well-being, with its origins in philosophy and history. It has been acknowledged as a fundamental right, as exemplified by documents such as the United States Declaration of Independence in 1776.

In this context, as mentioned, happiness is measured using six key indicators. Clearly, this method is not exhaustive or comprehensive in capturing the full scope of this phenomenon, but Gallup World Poll Report is considered one of the most objective and dependable means of measurement available.

As for freedom, we try to explain the available data through the presentation of some graphs. In *Figure 5* we can observe the time series of all continent scores over the decade under examination.

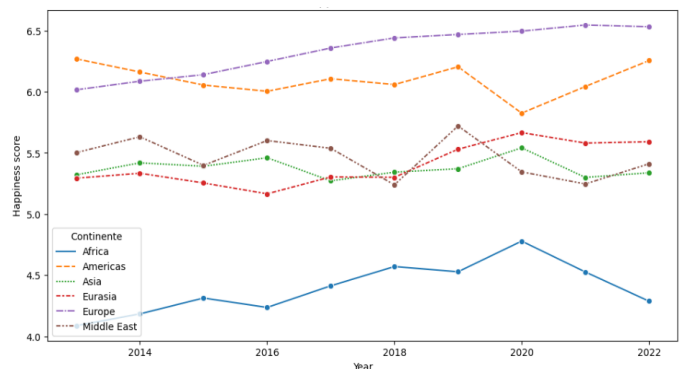


Figure 5: Time Series of Happiness score for each continent

We can observe that no continent has remained constant over time, and levels of happiness have increased and decreased. This marks both an improvement and a worsening of living conditions, similar to what has already been demonstrated in the case of freedom. Again, countries within the 'western' continents show the highest scores.

It's interesting to note that in 2013, the Americas had the highest score of happiness among continents. However, this score declined, particularly in 2020, as evidenced by a sharp drop in the chart. In fact, in 2022 the Americas are no longer the happiest continent as they have slipped

into second place<sup>7</sup>, with Europe taking the lead. In fact, Europe has consistently shown a slight but steady improvement in happiness over the past 10 years.

The Middle East, Eurasia and Asia maintained relatively similar happiness and performance scores until 2018; then, in particular, the Middle East experienced a more significant decline in happiness from that date onwards.

Africa, on the other hand, although placed in last place, until the year of the pandemic from Covid-19 shows an almost increasing trend; subsequently the score has rapidly decreased, returning to the level of the year of 2016.

Even in the case of happiness, the reasons for these changes are to be found in the socio-economic conditions of the various countries: we tried to justify it in the last paragraph of this report.

Upon scrutinizing the cartograms depicting the state of affairs in each country<sup>8</sup>, we observe conditions that mirror what we found when analyzing freedom maps. Over the past decade, the United States has seen a decline in living conditions and happiness, yet the most challenging living conditions continue to be concentrated in Africa and Asia.

These first conclusions make us think that the two aspects, happiness and freedom, go hand in hand but the results of the analysis have shown us that this is not always true.

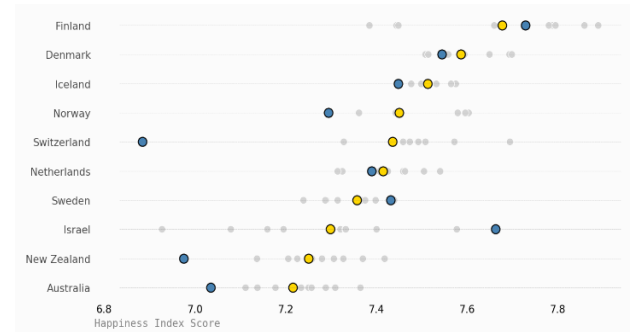


Figure 6: Happiness Index Scores through the years (top 10 countries)

Thanks to the chart presented above, we can identify the top 10 countries based on their happiness scores. The blue points represent the scores for 2022, while the yellow ones represent the average scores calculated over the past decade for each respective country.

Finland clearly stands out as the dominant country in this ranking, boasting both the highest average happiness score over the decade and the highest happiness score in 2022 compared to all other nations<sup>9</sup>.

Furthermore, even within this subset of the happiest countries, it's apparent that the majority (7 over 10 countries) experienced a decline in their average happiness scores in the last year under analysis. This highlights that the sense of unease pervading our society is being felt even within the traditionally 'happy' nations.

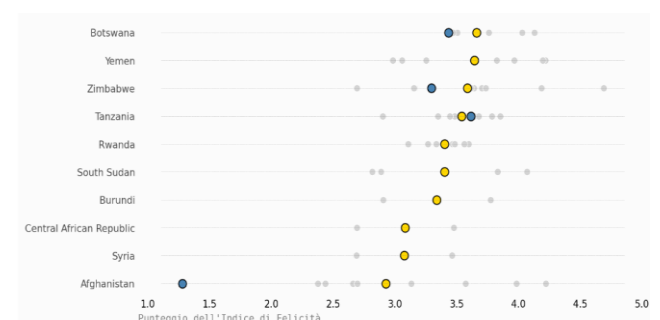


Figure 7: Happiness Index Scores through the years (worst 10 countries)

<sup>7</sup> An explanation can be traced both in the loss of exclusive global economic driving strength both in the corresponding appearance of fierce economic competitors, and also the conditions imposed with the pandemic.

<sup>8</sup> These details, while not included here to avoid overloading the text with nearly identical information, are accessible within the provided code.

<sup>9</sup> But it remains to verify the continuity of the same trend, in the light of the Russian-Ukrainian conflict and the political implications related to it.

When focusing on the 10 countries with the lowest happiness scores, Afghanistan consistently emerges as the nation reporting the lowest happiness scores throughout the decade under analysis, as evident from the gray data points on the background representing the scores for each year.

The situation notably deteriorated in 2022, as the total score shifted significantly to the left of the average score (represented by the yellow dot)<sup>10</sup>.

Throughout the decade in question, the ten least happy countries exhibited some degree of variability in their total scores, with the exception of Tanzania and Rwanda. However, their scores consistently remained at low levels, never exceeding 5.0.

Moreover, when comparing *Figure 7* to *Figure 6* and considering happiness as a quantitative indicator, it becomes apparent that, on average, the least happy countries were not even half as happy as the top 10 countries with the most favorable conditions. This underscores the significant disparities within our society.

## 5. CLUSTERING

### 5.1. Choosing clustering technique

This analysis aims to uncover concealed patterns within the data by implementing an unsupervised learning technique. In the context of 'unsupervised learning,' we refer to a machine learning approach that clusters data using specific algorithms based on the inherent information contained within them. These algorithms strive to identify similarities and differences among observations that are not assigned 'labels' determining their membership in particular groups. The underlying theory of these algorithms suggests that similar observations will share analogous characteristics and properties, facilitating their grouping.

Within the data mining technique known as '*clustering*', there are various types that render the approach flexible and adaptable to diverse contexts. This enables the management and grouping of raw, unclassified data objects based on structures or patterns present in the information.

Among the different types implemented, we find:

- Exclusive and Overlapping Clustering
- Probabilistic Clustering
- Density-Based Clustering

The goal of this article is not to provide a complete and exhaustive explanation of the used algorithms but rather a brief explanation for proper contextualization of the topics. Interested readers can find useful resources to delve deeper into the knowledge of the algorithms used in the provided bibliographic and online sources.

For the first technique, we implemented the K-Means algorithm, a widely known and conceptually straightforward method. This algorithm groups data points based on their similarity or proximity to one another. In simpler terms, it identifies data points with similar values and arranges them into the same cluster. This similarity is measured by a low Euclidean distance, indicating that observations closer to each other are placed in the same section. The K-Means process begins with carefully selecting a target number 'K,' which represents the number of centroids. Centroids are either imaginary or real positions that serve as the center of each cluster. Ideally, these centroids should be as far apart as possible to create well-separated clusters. The algorithm initializes by randomly selecting 'K' centroids. In subsequent iterations, the centroids' positions are refined by recalculating them each time a new observation becomes part of the cluster associated with the 'k-th' centroid. A new observation is assigned to the nearest cluster, i. e. the cluster in which is minimized the sum of squared distances within the cluster. The algorithm continues iterating until it reaches

---

<sup>10</sup> Deeper analysis on 5.2.3 paragraph

convergence. Convergence occurs when either all data units have been assigned to clusters, and there are no further changes in assignments, or when the maximum predefined number of iterations has been reached.

Regarding the second technique, the Gaussian Mixture Model has been implemented, which offers greater flexibility compared to K-Means (which fails when clusters do not have circular shapes) but requires the basic assumption about the data distribution. It is necessary for the data to be normally distributed to identify unknown subpopulations in a heterogeneous population. Being a probabilistic clustering, for each data unit, the probability of belonging to all groups is estimated, and the unit is assigned to the group where the probability is maximized. To accomplish this, the algorithm utilizes the Expectation Maximization (EM) algorithm. EM begins by estimating the parameters (mean and variance) of each cluster and then iteratively optimizing these parameters in subsequent iterations to improve the model's fit to the data.

For the last technique, we implemented DBSCAN, which stands for Density-Based Spatial Clustering of Applications with Noise. This approach starts by selecting an arbitrary starting point within the dataset. It then calculates the distance between this starting point and all other nearby points. If a point's distance is less than a predefined value known as epsilon ( $\epsilon$ ), it is considered a 'neighbor' of the starting point. If a sufficient number of points, as determined by the 'minPoints' parameter, fall within this neighborhood, the clustering process begins. The current data point becomes the first point in a new cluster. However, if not enough points are within this neighborhood, the point is classified as 'noise' and not assigned to any cluster. Subsequently, the search for neighboring points continues until the algorithm reaches convergence, meaning all points within the  $\epsilon$  neighborhood of the cluster have been visited and classified either into a cluster or as noise. DBSCAN offers several advantages, such as not requiring prior knowledge of the number of

clusters and its robustness in handling outliers, which are simply classified as 'noises'. However, it faces limitations when dealing with variations in data density, as the  $\epsilon$  distance and 'minPoints' parameter for identifying neighboring points will vary from cluster to cluster depending on the density.

After a data pre-processing phase, which included normalization <sup>11</sup> and standardization <sup>12</sup>, for all three clustering models, the appropriate number of clusters was chosen, where applicable, based on the "elbow" method (minimizing the Within-Cluster Sum of Squares or WCSS) and maximizing the silhouette coefficient.

These two mentioned techniques use the screeplot, which represents the total sum of intra-cluster squared distances (WCSS) as a function of K (the number of clusters). As K increases, the sum of squares generally decreases, but this rate of decrease diminishes. The optimal K is typically identified at the 'elbow' of the curve, where it levels off. Meanwhile, the silhouette coefficient is a measure of how similar a data point is to the data in its cluster (cohesion) and how dissimilar it is to data in the nearest cluster, i.e., the cluster whose average distance from the data point is the lowest (separation). A silhouette value close to 1 implies that the data point is in an appropriate cluster, while a value close to -1 indicates that the data point is in the wrong cluster.

To decide which algorithm to present, we analyzed, for each of the three techniques, the Internal Validation Indexes that do not require a priori knowledge of partitions already known.

In *Figure 8*, the results of internal validation with Silhouette, Calinski-Harabasz (also known as the Variance Ratio Criterion) and Davies-Bouldin metrics are presented.

The second metric is calculated as the ratio between the sum of the dispersion between clusters and the sum of the dispersion within clusters for all clusters (where the dispersion is the

---

<sup>11</sup> If bell distribution of the data is required or to remove the effects of potential outliers.

<sup>12</sup> When magnitude of data is different.

sum of the square distances): a high CH value indicates a better grouping, since the observations within each cluster are closer to each other (denser), while the clusters themselves are more distant from each other (well separated). The third index, instead, is calculated as the average similarity of each cluster with the cluster more similar to it: the index represents the average "similarity" of the clusters, where the similarity is a measure that relates the distance between the clusters to their size. A model with a lower Davies-Bouldin index has better separation between clusters.

| ALGORITHM                      | Silhouette coefficient     | Calinski-Harabasz | Davies-Bouldin |
|--------------------------------|----------------------------|-------------------|----------------|
| K-Means (Freedom clustering)   | 0.53 (2013)<br>0.52 (2022) | 631.87<br>562.90  | 0.59<br>0.59   |
| GMM (Freedom clustering)       | 0.54<br>0.53               | 312.93<br>135.73  | 0.49<br>0.56   |
| DBSCAN (Freedom clustering)    | 0.43                       | 117.58            | 1.27           |
| K-Means (Happiness clustering) | 0.26 (2013)<br>0.32 (2022) | 48.84<br>50.46    | 1.15<br>1.31   |
| GMM (Happiness clustering)     | 0.19<br>0.14               | 35.38<br>28.23    | 1.76<br>1.57   |
| DBSCAN (Happiness clustering)  | 0.16<br>-0.16              | 4.63<br>5.09      | 3.2<br>2.0     |

Figure 8: Internal validation indices of the algorithms. It can be observed that the GMM outperforms the K-Means only for the clustering of the Freedom dataset, but in all other cases, K-Means performs better. Therefore, it has been chosen as the algorithm to be presented in the paper.

Hence, the choice to present the results has been oriented towards K-Means, not only due to its superior performance in terms of internal validation but also because of its simplicity and clarity in interpreting the results. In fact, each data point is exclusively assigned to one cluster, unlike

GMM where a probability of membership is estimated, making the understanding more complex and less objective. Furthermore, the greater scalability and the absence of constraints regarding assumptions on data distribution and density made this algorithm preferable over the others.

## 5.2. K-Means implementation

The clustering operation aims to group countries around the world in order to better understand how and why some countries are considered more similar than others in terms of both political rights, civil liberties and the perception of freedom.

This process enables us to conduct a spatio-temporal analysis: spatial, because it helps us comprehend similarities and differences among countries despite their diverse geographical locations; and temporal, because it allows us to compare the clustering patterns at the beginning of the period (2013) and the end (2022).

To ensure comparable results, K-means has been implemented in the same way looking at both freedom data and happiness data. In fact, if we had not intervened with some costrictions (such as the choice of the number of clusters and the increasing ordering; the first cluster in fact corresponds to that of the less happy or free countries) the algorithm would have been automatically and independently executed in the years under analysis, making the results difficult to compare.

First of all, data has been analyzed singularly for 2013 and 2022, in order to allow the comparison between the analysis period. Subsequently, we normalized the data to ensure all variables shared the same scale and magnitude, ranging from 0 to 1, creating uniformity. This step is crucial to prevent any single variable from dominating the analysis. Specifically, we applied normalization to the freedom dataset, while the happiness data required standardization to address variations in measuring scales.

In the subsequent paragraphs, we present the interpretation of the results. These results were

obtained after determining the optimal number of clusters using the above mentioned techniques and conducting internal validation of the algorithm. This validation process is essential, albeit intricate, to evaluate the quality of the clustering: it represents a critical phase in such analysis.

### 5.2.1. Clustering for Freedom in 2013 and 2022

We selected the variables Political Rights and Civil Liberties as our clustering variables. These variables represent the cumulative values of specific subcomponents: A-C for Political Rights and D-G for Civil Liberties, which we utilized to analyze the phenomenon. We opted for these aggregated variables because the individual subcomponents had very similar values, typically falling within the range of 0 to 4 points. Using the subcomponents made it challenging for the algorithm to identify distinct patterns. Thus, aggregating these components helped resolve this issue.

To address potential discrepancies in magnitude that might disrupt the clustering process, we normalized the data. You can observe the results in *Figure 9*.

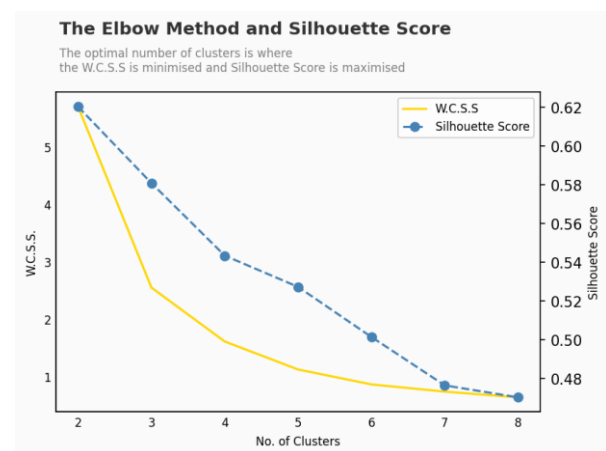
|       | count | mean      | std       | min | 25%       | 50%       | 75%        | max   |
|-------|-------|-----------|-----------|-----|-----------|-----------|------------|-------|
| index | 135.0 | 67.000000 | 39.115214 | 0.0 | 33.500000 | 67.000000 | 100.500000 | 134.0 |
| A     | 135.0 | 0.641975  | 0.344055  | 0.0 | 0.333333  | 0.750000  | 1.000000   | 1.0   |
| B     | 135.0 | 0.635648  | 0.311242  | 0.0 | 0.375000  | 0.687500  | 0.937500   | 1.0   |
| C     | 135.0 | 0.537654  | 0.295720  | 0.0 | 0.250000  | 0.583333  | 0.750000   | 1.0   |
| D     | 135.0 | 0.677531  | 0.279446  | 0.0 | 0.466667  | 0.733333  | 0.933333   | 1.0   |
| E     | 135.0 | 0.635802  | 0.304530  | 0.0 | 0.375000  | 0.666667  | 0.916667   | 1.0   |
| F     | 135.0 | 0.504630  | 0.293611  | 0.0 | 0.250000  | 0.437500  | 0.750000   | 1.0   |
| G     | 135.0 | 0.574815  | 0.259207  | 0.0 | 0.400000  | 0.600000  | 0.800000   | 1.0   |
| PR    | 135.0 | 0.616079  | 0.302278  | 0.0 | 0.341463  | 0.658537  | 0.902439   | 1.0   |
| CL    | 135.0 | 0.580159  | 0.282016  | 0.0 | 0.357143  | 0.571429  | 0.830357   | 1.0   |
| Total | 135.0 | 60.748148 | 27.922258 | 4.0 | 39.000000 | 62.000000 | 86.000000  | 100.0 |

*Figure 9: Normalized results. Here are presented all variables, including also the total, but for the clustering we used only 'PR' and 'CL' variables.*

The chart presented in *Figure 10*, corresponding to the year 2013, facilitated our selection of the optimal number of clusters. Through careful observation and analysis of the trade-off between minimizing the within-cluster sum of squares and maximizing the silhouette coefficient, we

determined that the ideal number of clusters is 5. This choice ensures that our clusters exhibit both strong internal cohesion and clear external separation, with a silhouette coefficient exceeding 0.50. Additionally, other internal validation indices, such as Calinski-Harabasz and Davies-Bouldin, also returned acceptable values.

Remarkably, we obtained nearly identical results for the year 2022. Consequently, to streamline this report, we have omitted the graph for 2022, but it remains accessible within the provided Python code



*Figure 10: Screeplot useful for choosing the proper number of clusters, year 2013*

The composition of each single cluster allow us to say that:

- *Freedom Strugglers*, cluster 0, includes countries with the poorest political and civil conditions; in particular, characterized by scores falling within the first quantile, totaling 25 points or less.
- *Harmonious Progressors*, cluster 1, consists of countries whose total score is around 40 (with a moderate civil situation and a low to medium political situation). Most of these countries belong to Asia and North Africa;
- *Balanced Achievers*, cluster 2, encompasses those countries that have average scores. These are mostly South American or East Asian countries, with total scores ranging between 60 and 70;



- *Freedom Champions*, cluster 3, includes those countries with a medium-high of political and civil conditions, achieving a total score of approximately 80.
- *Liberty Pioneers*, cluster 4, is where the freest countries are; these are typically Western nations but here are also included some countries of different continents, such as Uruguay and Taiwan in 2013 and Costa Rica in 2022.

The choropleth maps, presented in *Figure 11* and *12*, display the outcomes of this analysis for the years 2013 and 2022, respectively: this has enabled us to assess the conditions of the countries under consideration.

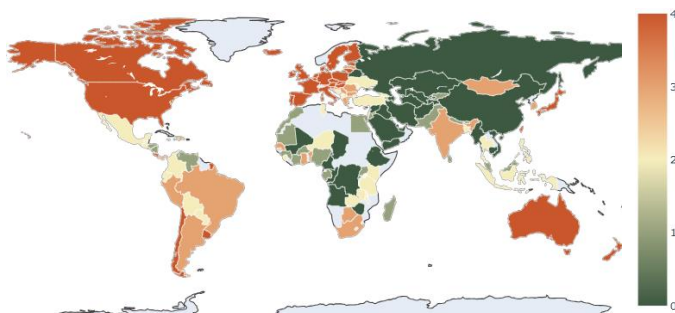


Figure 11: State of Political rights and Civil Liberties in 2013

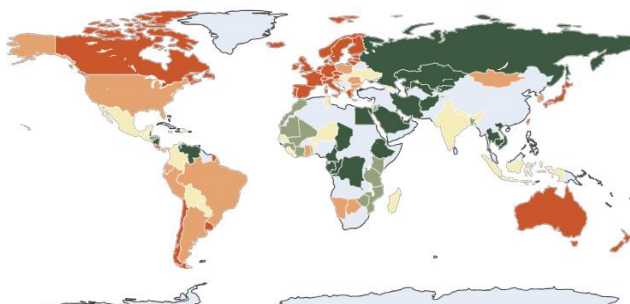


Figure 12: State of Political rights and Civil Liberties in 2022

Several countries are not included in our analysis due to data integration issues between the two datasets focusing on freedom and happiness. However, this doesn't hinder our ability to observe a general decline in global freedom levels from 2013 to 2022. In fact, looking Continent by Continent it is possible to notice some differences: in North America, it's evident that there has been a decline in the perceived levels of freedom, both in the United States and Venezuela in 2022. Moreover, both in 2013 and 2022 Africa has a really fragmented situation, with the majority of

countries experiencing poor conditions of freedom, since most of them belong to the two worst clusters.

Despite that, it is possible to notice some improvements involving for example Madagascar and Mali, but unfortunately also some worsening for example in Egypt, Kenya and Tanzania. Moving to a different continent, we find that Europe and Oceania have experienced relatively stable situations. However, the Asian continent exhibits some stability, considering the countries for which we have observed data changes over the years, though India has witnessed a decline

The general aggravation in terms of freedom lived by the world is certainly also due to the pandemic and to all that followed in terms of threat to health, political and economic stability.

Through the following map, it is possible to better interpret how the situation changed through the years.

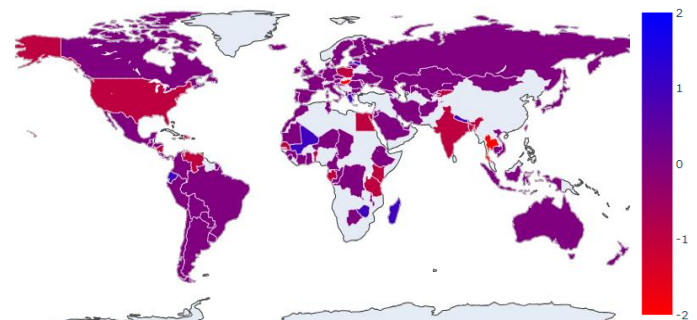


Figure 13: Choropleth map of clustering differences about freedom between 2013 and 2022

Countries that have maintained their status quo throughout the specified period are highlighted in purple, while those that have shown improvement are marked in blue, and those that have seen a decline are depicted in reddish hues. It is possible to notice that most of the countries haven't changed in freedom conditions from 2013 to 2022. However, there have indeed been some alterations, although unfortunately in most cases they have been negative.

Countries like the USA, Nicaragua, Venezuela, the Dominican Republic, Senegal, Benin, Gabon, Tanzania, Kenya, Egypt, Poland, Kyrgyzstan, India, and Bangladesh have witnessed a deterioration in

their situations. Notably, Thailand and Hungary, with the conservative turn of Orbán's policy<sup>13</sup>, have experienced particularly severe declines.

On a more positive note, Ecuador, Mali, Zimbabwe, Madagascar, Greece ( compared to previous years, characterized by financial default), Latvia, and Nepal have managed to improve their situations.

Inspecting those worst countries, it is possible to find some explaining factor of this aggravation. For instance, Hungary has seen a gradual but distinct departure from democracy in recent years, while Thailand has been under military dictatorship since 2014.

### 5.2.2. Clustering for Happiness in 2013 and 2022

Indicating the level of happiness in the countries, six variables are considered: GDP per capita, social support, healthy life expectancy at birth, freedom to make life choices, generosity, perception of corruption. Since these variables have different scales, standardization has been applied to ensure that the k-means algorithm remains robust and is not unduly influenced by variables with the least or most prominent features

Again, through the inspection of the trade-off between WCSS and Silhouette score it has been possible to discover the optimal number of clusters.

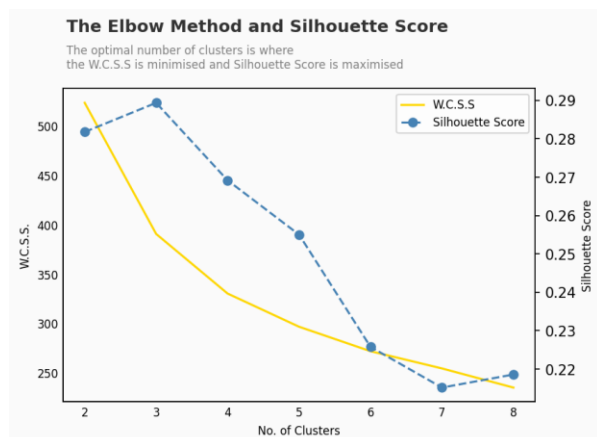


Figure 14: Optimal number of clusters for Happiness dataset in 2013

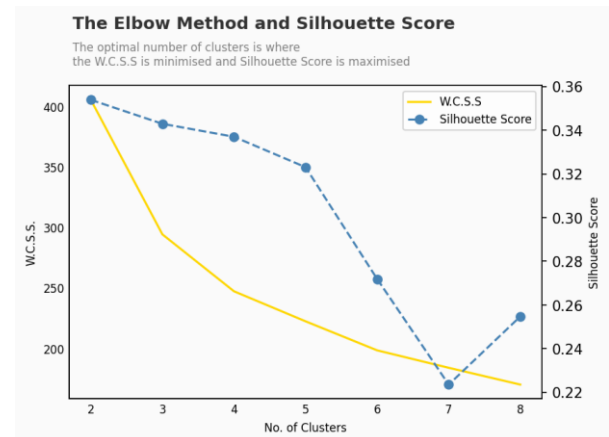


Figure 15: Optimal number of clusters for Happiness dataset in 2022

Therefore, 5 is the proper number of clusters in both years, that ensures a lower WCSS and a quite good Silhouette score that is not so much lower than the one obtained with 4 clusters. In fact, maintaining the number of clusters fixed on 5 as it has been possible till now, is a good way to allow the comparison between clusters in every different field of analysis; for instance, thanks to this, it is possible to compare which countries are inside the same group considering freedom clusters and happiness clusters.

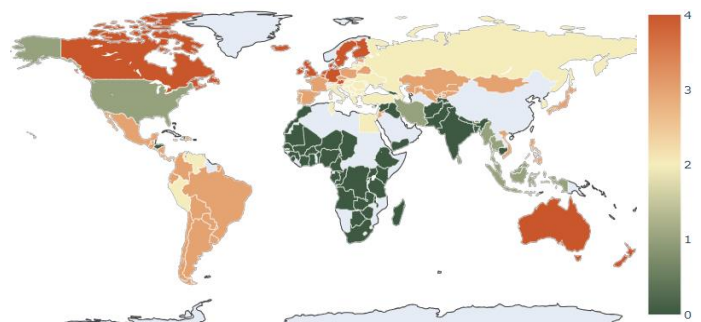


Figure 16: State of happiness in 2013

Looking at Figure 16 it is evident the deep state of unhappiness of Africa, where only Egypt and Tunisia are in a cluster better than 0. Moreover, it is quite curious that the United States are set only in cluster 1, indicating a low medium level of general happiness in one of the most developed countries in the world. In particular, it is a country with one of the highest GDP level of the world but

<sup>13</sup> <https://it.euronews.com/2023/05/24/in-ungheria-la-democrazia-e-in-regressione-sotto-la-guida-di-viktor-orban>;  
<https://www.gisreportsonline.com/r/hungary-orban-political/>



it reaches also a very low score in trust in governance.

Conversely, as expected, Canada, Oceania and North European areas demonstrate a uniform pattern, with countries all grouped in the happiest cluster. At the same time, the Asian continent seems to be quite heterogeneous, with countries like India and its neighboring nations with the lowest happiness levels and countries such as Japan and Thailand that are in cluster 3.

Therefore, cluster 2 and 3 are those with the most heterogeneous condition, with countries from all around the world. For example, Italy is in cluster 2 with Peru, Egypt and Tunisia, even if in our mind these countries are very far away from Italy. The common clustering is explained by the fact that Italy has a limited level of the variable “freedom to make life choices”, that brings its general condition in a low group, where there are also those countries like Peru and Tunisia with low levels of every variable.

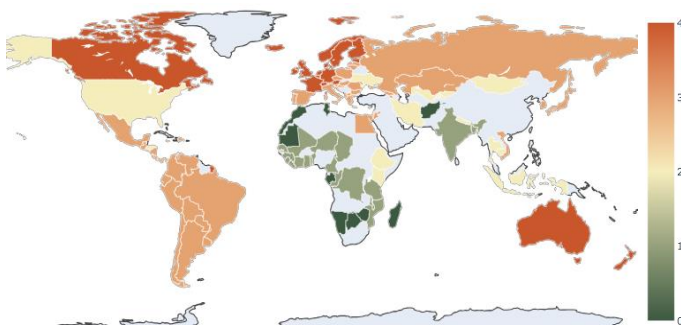


Figure 17: State of happiness in 2022

From a first quick glance, it is possible to see a clear greater presence of orange, indicating that many countries are in the clusters with the highest levels of happiness. Therefore, we can say that from 2013 to 2022 the state of happiness has improved on average all over the world, even if in Africa this improvement didn't lead any country to the top cluster.

In the following picture, it is possible to appreciate the change experienced by individual countries from 2013 to 2022 in happiness score.

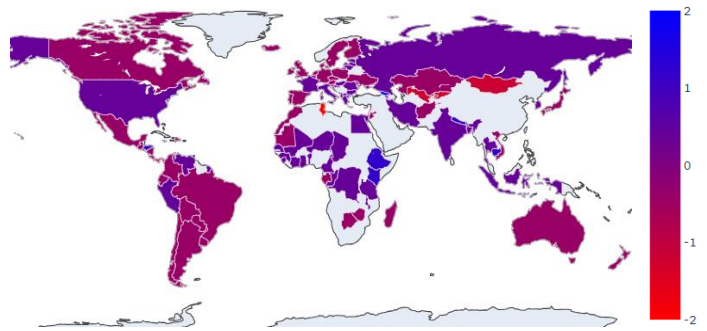


Figure 18: Choropleth map of clustering differences about happiness between 2013 and 2022

From this table it is possible to see that some countries including the United States, Italy, India, Tanzania and others have seen improvements in their happiness levels as they have transitioned to the top-tier cluster compared to their 2013 rankings. In particular, the greatest improvements were experienced by Honduras, Ethiopia, Kenya, Georgia, Nepal and Cambodia.

Conversely, some countries have experienced a decline in their happiness levels, with notable examples being Mongolia, Kyrgyzstan, Uzbekistan, and particularly Tunisia. These improvements or deterioration can often be attributed to challenging political and economic circumstances, that are factors that erode people's trust in governance, lead to lower GDP per capita levels, and so forth.

When examining both happiness and freedom clustering, a clear distinction emerges. In the freedom clustering colors are brighter, meaning that here countries are more concentrated in the most extreme clusters. In contrast, the happiness clustering displays paler colors, with the majority of countries falling into the low or high middle clusters.

Is clear that only a select few countries exhibit a similar pattern in terms of happiness and freedom; in other words, they belong to the same category in both cases. Notably, this group includes Canada, Oceania, and Northern European countries, all of which fall into the cluster characterized by the highest levels of both freedom and happiness (Cluster 4). In these regions, people genuinely perceive themselves as living in states of freedom,

and correspondingly, they experience elevated levels of happiness. Consequently, it's safe to conclude that happiness and freedom do not necessarily align hand in hand.

### 5.2.3. Clustering on Variability

During the analysis, an interesting research question emerged regarding countries that have maintained a stable situation over time and those that have undergone significant changes. To explore this issue, we had to perform clustering. However, considering that time dynamically influences the variables under consideration, we had to find a way to disentangle the analysis from temporal factors. Therefore, for each year, we calculated the percentage variation of the variables of interest, and to account for both positive and negative effects, we summed the squares of these variations. Subsequently, we applied the K-Means algorithm to cluster the data, obtaining a grouping of countries with the most significant variations over time. The results of this analysis are described in the next paragraph.

To better understand the results, we ordered the clusters in ascending order: Group 0 called 'Steadfast Equilibrium' includes countries whose happiness and freedom values remained almost constant over time, both positively for the happiest and freest countries, and negatively. On the other hand, Group 4 includes countries that experienced the most significant variations ('Dynamic Shifters' nations).

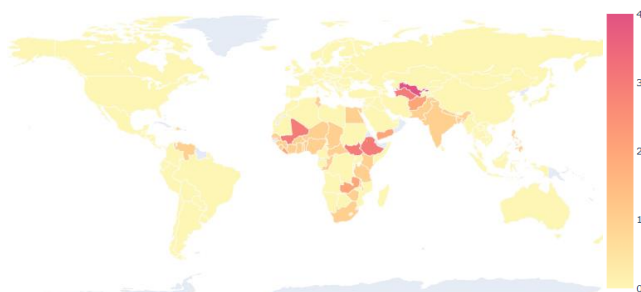


Figure 19: Coropleth of Variations

The results obtained from this analysis highlight that most countries have maintained a consistent trend over time, without undergoing significant political or social shocks that would revolutionize

the values of the indicators comprising the freedom and happiness scores.

As expected, all Western countries remained constant over time, while Asian and sub-Saharan African countries showed more variability.

Cluster 1 that can be called 'Happiness Shapers' contains countries that experienced variations mainly in the happiness component during the analyzed period, while political and civil rights remained more stable. In this case, the analysis returned African countries and India, which remained unchanged from a political perspective but experienced changes in the social components that influence the happiness score.

In the next group (Cluster 2), the emblematic cases of Afghanistan and Zambia showed an even more pronounced variation in the happiness indicator compared to the previous group. In both countries (despite different socio-economic and political situations), the indicator decreased by 2 points over time, signaling a worrying deterioration in living conditions, though less pronounced in Zambia compared to other countries in this cluster.

In particular, for Afghanistan, a drastic decline was also observed in the freedom components: political rights experienced a significant contraction between 2021 and 2022, going from 13/40 to 1/40, while civil rights dropped by half compared to the previous decade (from 16 to 9 out of a total of 60). The happiness components that were most affected by the political and social changes, following the U.S. withdrawal, were generosity (altruism and volunteering), freedom of life choices, and social support.

The Taliban's takeover of Kabul in August 2021 raised serious concerns about human rights and civil liberties in the country. The Taliban's military offensive, which began in May 2021, erased the modest gains in the rights of ethnic and religious minorities in a little over two months. With the restoration of the Islamic Emirate of Afghanistan by the Taliban, violence and restrictions on freedom of expression and women's rights became widespread, casting a shadow over prospects for progress in the country.

The penultimate cluster('Liberty Transformers' group), instead, contains countries in which the happiness score varied the most in terms of civil and political rights components. It includes Ethiopia, Turkmenistan, and Mali. Ethiopia is an example of positive variation: the score for political rights almost doubled in 10 years, while civil rights, after a sharp decline in the 2017-2018 biennium, remained almost unchanged over time and fixed at 11. Although the country is not considered free (with a total score of only 23 out of 100), the changes shown seem to gradually push the country towards a slow political opening promoted by the new Prime Minister Abiy Ahmed, marking a moment of significant political change, despite significant ethnic tensions and demands for autonomy from different regions of the country<sup>14</sup>.

In contrast, Mali and Turkmenistan experienced a sudden decline in the freedom indicator over 10 years: the former became the scene of clashes between Islamist groups, resulting in the worsening of civil and political rights limited due to the lack of effective democratic representation and a climate of widespread insecurity.

In the second country, political rights were zeroed out, while civil rights decreased by over half. The explanation for this lies in the rise to power of the Democratic Party of Turkmenistan (DPT)<sup>15</sup>, which won the majority in the 2018 political elections, establishing a dictatorial government that suppressed opposition and granted itself full powers, worsening the already precarious conditions of the country. The repression of opponents, the limitation of expression and assembly rights, and human rights violations documented by international organizations gave rise to a climate of repression that contributed to the overall deterioration of rights in Turkmenistan during the period under consideration.

Finally, we conclude the paper with a slight note of hope, highlighting how the country that recorded the most significant variation

(Uzbekistan) unexpectedly showed a positive trend. Despite not starting from a situation of freedom (in 2012, the score was 4 out of 11), both civil and political rights doubled (from 4 to 9 and from 0 to 2 points, respectively) over 10 years.

The greater wave of freedom occurred between 2017 and 2018 when the second president, Shavkat Mirziyoyev<sup>16</sup>, came to power and launched a series of political and social reforms aimed at leading the country towards greater political openness and better protection of human rights. The country has indeed experienced a phase of political openness and dialogue with the opposition, reforms have been undertaken to improve the efficiency and transparency of government and judicial institutions, and there has been an expression of willingness to abolish the slavery of cotton and the systematic use of child labor. However, despite these advancements, the country remains dominated by an authoritarian regime and continues to be classified as "partially free" by Freedom House. There are still significant challenges to address to achieve the features of democracies in free countries.

## CONCLUSIONS

The results obtained lead us to understand that at the beginning of the analysis period, the world was primarily characterized by two opposing situations: on one hand, the "Liberty Pioneers," where individuals lived in democracy and did not have to fear expressing themselves freely, and on the other hand, the "Freedom Strugglers", where socio-economic conditions were often poor, and apparent democracies concealed, in fact, authoritarian regimes. However, by the end of the decade, the clear distinction had completely disappeared, giving way to a more complex situation that highlights a general deterioration, which also affected some nations previously

---

<sup>14</sup>[https://it.wikipedia.org/wiki/Caduta\\_di\\_Kabul\\_\(2021\)](https://it.wikipedia.org/wiki/Caduta_di_Kabul_(2021))  
<https://onwardforafghanwomen.org/policy/timeline-of-taliban-assault-on-womens-and-human-rights/>

<sup>15</sup> [https://it.wikipedia.org/wiki/Turkmenistan#Ordinamento\\_dello\\_Stato](https://it.wikipedia.org/wiki/Turkmenistan#Ordinamento_dello_Stato)

<sup>16</sup> <https://it.wikipedia.org/wiki/Uzbekistan#Politica>

considered "Liberty Pioneers," such as the United States.

Regarding happiness during the same analysis period, excluding Canada, Oceania, and countries in Northern Europe, the data reveal a widespread sense of discontent, particularly pronounced and confirmed in African nations. Another anomalous situation arises in South America where happiness levels are surprisingly good despite widespread poverty and political uncertainty, with most countries residing in cluster 3, denoting a high to medium level of happiness.

However, in 2022, global happiness levels showed a positive trend; specifically, this improvement affected countries such as the United States, Russia, Southern European countries, and some Third World states, albeit starting from a moderately low level. Nevertheless, when analyzing the joint variation of both aspects, significant changes were not observed in most countries, except for the situations highlighted in the previous paragraph.

In conclusion, we can assert that freedom and happiness, although they may appear interconnected and influenced by a country's socio-political circumstances, do not always go hand in hand. A striking example of this is Russia: this country finds itself in the worst cluster (0) when it comes to political and civil freedom, yet despite this, the levels of happiness within its borders are relatively high, as evidenced by its belonging to cluster 3.

Conversely, consider the example of the United States, which, despite having favorable socio-political conditions, cannot be said to be equally positive in terms of happiness.

Therefore, further developments of this analysis include monitoring the trend over time, especially in light of the Russo-Ukrainian conflict, which has had repercussions worldwide, along with the new policies associated with it.

## BIBLIOGRAPHY AND SITOGRAPHY

<https://freedomhouse.org/countries/freedom-world/scores>

<https://worldhappiness.report/about/>

<https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>

[https://en.wikipedia.org/wiki/ISO\\_3166-1#Officially\\_assigned\\_code\\_elements](https://en.wikipedia.org/wiki/ISO_3166-1#Officially_assigned_code_elements)

<https://unstats.un.org/unsd/methodology/m49/overview>

[https://www.jstor.org/stable/pdf/2346830.pdf?casa\\_token=8ZrsNqfJ390AAAA:l-nJDlIpkiVi-dRmt2A\\_ahePK6zkHTwMON42n7vEP0AuQQyUSrJtGhSXh7yy2iKa5SOa9Tkdfiy75yxLDLw1k7U68](https://www.jstor.org/stable/pdf/2346830.pdf?casa_token=8ZrsNqfJ390AAAA:l-nJDlIpkiVi-dRmt2A_ahePK6zkHTwMON42n7vEP0AuQQyUSrJtGhSXh7yy2iKa5SOa9Tkdfiy75yxLDLw1k7U68)

[0oSJe1HnxRxBjmtsIOCdJWm0r1 \(Algorithm AS 136: A K-Means Clustering Algorithm Author\(s\): J.A. Hartigan and M. A. Wong](#)

[Source: Journal of the Royal Statistical Society. Series C \(Applied Statistics\), 1979, Vol. 28, No. 1](#)

[\(1979\), pp. 100-108 Published by: Wiley for the Royal Statistical Society\)](#)

### K Means

<https://www.analyticsvidhya.com/blog/2021/02/simple-explanation-to-understand-k-means-clustering/#>

<https://towardsdatascience.com/k-means-clustering-explain-it-to-me-like-im-10-e0badf10734a>

### GMM

<https://towardsdatascience.com/gmm-gaussian-mixture-models-how-to-successfully-use-it-to-cluster-your-data-891dc8ac058f>

<https://behesht.medium.com/unsupervised-learning-clustering-using-gaussian-mixture-model-gmm-c788b280932b>

<https://towardsdatascience.com/clustering-out-of-the-black-box-5e8285220717>

<https://bobrupakroy.medium.com/gaussian-mixture-model-gmm-clustering-497c090a8316>

[http://leap.ee.iisc.ac.in/sriram/teaching/MLSP\\_16/refs/GMM\\_Tutorial\\_Reynolds.pdf](http://leap.ee.iisc.ac.in/sriram/teaching/MLSP_16/refs/GMM_Tutorial_Reynolds.pdf) (Douglas Reynolds MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA)

### DBSCAN

<https://medium.com/@elutins/dbscan-what-is-it-when-to-use-it-how-to-use-it-8bd506293818>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6814687>

<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

<https://medium.com/towards-data-science/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>

<https://www.ibm.com/it-it/topics/unsupervised-learning>

### Metrishe

<https://towardsdatascience.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python-4feFeb2988e>

[https://heinonline.org/HOL/Page?handle=hein.journals/helsnk14&div=24&q\\_sent=1&casa\\_token=8yVbbTVdO1kAAAA:VeCZ3SIAPHyG6jRmojuerGZn73i96Tclp5omEh2G0Fw9Pd39ej5HiuORtHiSdIsbQFWpeT4YOA&collection=journals](https://heinonline.org/HOL/Page?handle=hein.journals/helsnk14&div=24&q_sent=1&casa_token=8yVbbTVdO1kAAAA:VeCZ3SIAPHyG6jRmojuerGZn73i96Tclp5omEh2G0Fw9Pd39ej5HiuORtHiSdIsbQFWpeT4YOA&collection=journals)