
MASTER

ACTUARIAL SCIENCE

MASTER'S FINAL WORK

INTERNSHIP REPORT

# AN APPLICATION OF CLUSTERING ANALYSIS TO THE COMPARISON OF MORTALITY RATES

TUNG NGUYEN XUAN

SUPERVISION:

ANA SOFIA MANITALY SOUSA
ONOFRE ALVES SIMÕES

OCT - 2022

*To find, to seek, to strive, and not to yield*

*Alfred, Lord Tennyson*

## ACKNOWLEDGEMENTS

Someone had said: During the road, sometimes the destination is not the most important thing, but it is the journey. I can say the same with this thesis. The thesis itself is not the most important achievement. Still, it takes two years to stay in ISEG to learn many subjects, from the Mortality model to Experience Rating.

For that, I wish to show my appreciation to my professors, who helped me during the last two years. Professors Onofre, for your support and advice, professor Alexandra, professor Joao, and many other teachers have shown me the first bricks to build my career path as Actuary.

I hope to say thanks to my family: my grandparent, who provided me continued support; my father, uncle, and aunt, who encouraged me to study abroad; and my younger brothers and sisters, who always gave me the motivation to go forward.

Finally, I wish to give special thanks to LP, my best friend, who is always by my side. Without you, I cannot complete my journey.

TABLE OF CONTENTS

## List of Figures

## LIST OF TABLES

# AN APPLICATION OF CLUSTERING ANALYSIS TO THE COMPARISON OF MORTALITY RATES

ABSTRACT

This work provides the study of dissimilarity between mortality rates in 116 countries. Mean Shift Algorithm and Principal Component Analysis processed efficiently to classify countries into clusters, which show the reduction of the rate of mortality in the last twenty years. The result reveals an evident difference in the children and adult groups among researched countries, while it is more difficult to classify the rate for the seniors. Although the mortality rate will gradually reduce through time, there are two distinguishable patterns for developing countries having high rates: in Africa, where the value is higher for children, and in East Asia, where the survival probability for the adult is lower. The main reasons for the high mortality value in researched countries are natural disasters, global pandemics, and low-quality life. We also provided an application of the methodology by estimating the expected value and standard deviation of survival probability in Portugal and Vietnam in the next ten years.

KEYWORDS: Mortality rate; Life insurance; Clustering; Mean shift; Principal Component Analysis.

## 1 INTRODUCTION

### *1.1 Motivation*

Nowadays, there has been an expansion in using the mortality table in both private and public sectors. To ensure a high quality of life product, the Life insurance company and Private Pension Fund need high accuracy mortality tables, which requires continuously updating. Meanwhile, the mortality rate is essential in evaluating a country's development and its government system. During my Internship at Willis Towers Watson, I have an opportunity to learn about the tables from various regions in Europe. My tasks included analyzing the UK pension scheme data, researching their assumptions, and adjusting the related factors.

In detail, I have to evaluate different pension schemes, including thousands of members, by using a discount rate to value the amount of pension in various periods. In addition, I have to set the CPI and RPI, as well as the mortality table, and research the effect of late retirement and early retirement on the value of the pension scheme.

Comparing mortality tables between UK and Europe among long period, although all of them have similar curves, these tasks raises a question: if the Mortality is different

between periods in history and between countries and regions, how do we divide them into groups and use this information to estimate the growth in rate of mortality in the future?

### *1.2 Literature review*

Researching from multiple data: UK and Wales in Europe, Japan in Asia, and Chile in America, Omran A. R. (1971) had concluded that there are three stages of mortality in every society. The Age of Pestilence and Famine when mortality is high and fluctuating, thus precluding sustained population growth. In this stage, the average life expectancy is low and variable, fluctuating between 20 and 40 years. In the Age of Receding Pandemics, mortality decreases significantly since the epidemic peaks become less frequent or disappear. The life expectancy at birth fluctuates from 30 to 50 years. Finally, mortality reduces and becomes stable in the Age of Degenerative and Man-Made Diseases.

Olshansky, S. J., & Ault, A. B. (1986) developed this theory by adding the "fourth stage" during which the maximum point of convergence of life expectancies would seem to increase thanks to achievements in the treatment of cardiovascular diseases. Caselli, G & Vallin, J (2002) summarized other publications and suggested that the life expectancy at birth at stage four is 80. In addition, they provided a comparison between countries. While regions in Europe, Asia, and America are in the third and fourth stages, the mortality rate in Sub-Saharan Africa is at the unfinished second phase. In Figure 1, Caselli, G & Vallin, J classified them into six groups: countries having made rapid progress, countries having made steady progress, stagnating countries, Moderate or recent decrease, Deep regression, and countries at war. The main reason African countries are slower than others came from HIV/AIDS, war, and inefficient government systems.

McMichael, A & Mckee, M (2004) had recently researched the convergence and divergence in several regions. Among them, 42 countries show life expectancy at birth (both sexes) lower in 2001 than in 1960, 1980, or 1990, while others experienced the convergence. In addition, they showed that there is an increased heterogeneity between countries, here summarised as those achieving rapid gains, those achieving slower or plateauing gains, and those having frank reversals. The authors concluded that globally life expectancy has been on an extended uptrend. Still, the emerging picture of variable mortality trends and regional setbacks indicates that any general deterministic convergence process does not guarantee future health gains. The higher mortality rate are not only experienced higher in Developing countries rather than in Developed countries but also differences inside those regions. From these classifications, they developed to research about the relation between the economic development, population health, and technological change.

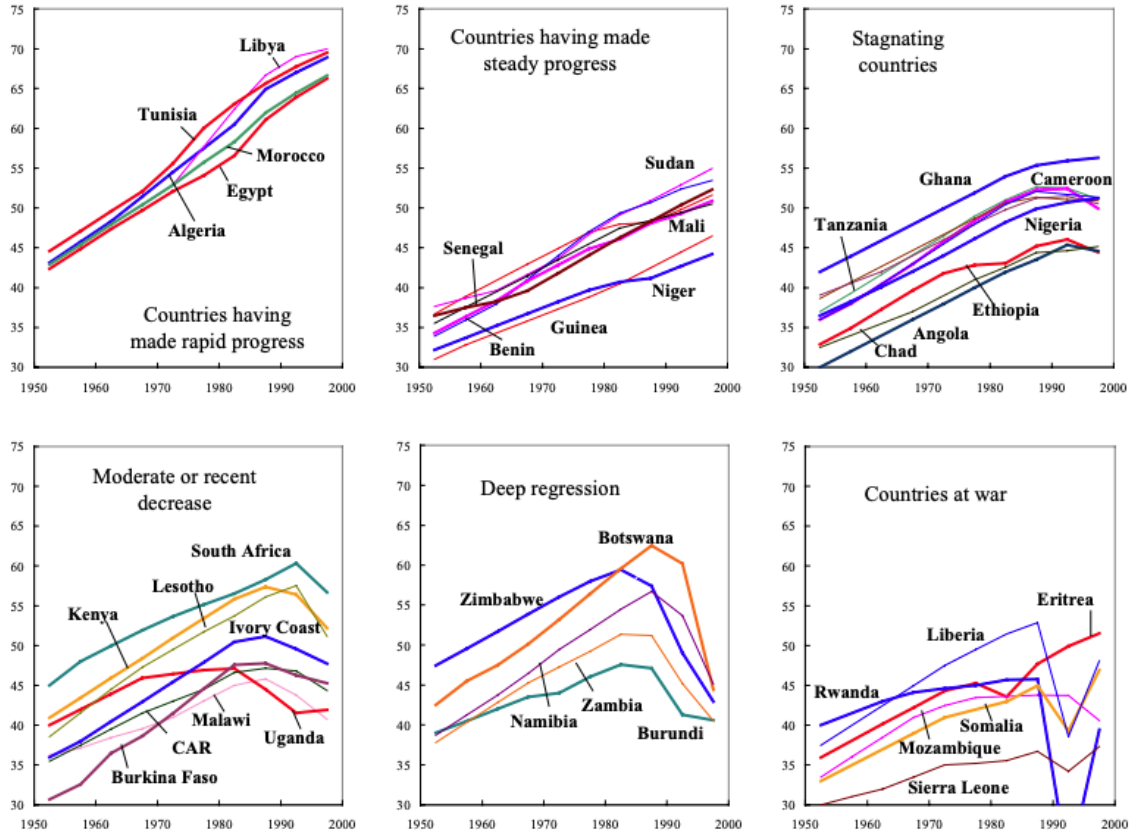


FIGURE 1: Life expectancy at birth: trends from 1950 to 2000 in all African countries
Source: Caselli, G & Vallin, J (2002)

Léger, Ainhoa-Elena & Mazzuco, Stefano (2020) used the Functional Data Analysis approach to cluster Human Mortality Database in 32 countries from 1960 to 2010. They used Component Practical Analyst to reduce the number of features. In Figure 2, there are similarities between the Mortality Rate of Males in Japan (2000) and France (2010), as well as between Sweden (2000), France (1990), and Japan (1990). On the other hand, Russia has a different pattern than others; the reason may come from the government system and life quality. Moreover, when applying distance-based cluster analysis, they found seVenezuela Mortality Rate groups. As shown in Figure 3, cluster 1 included regions with high infant mortality, and cluster 3 has a similar shape to cluster 1 but lower infant mortality. Group 2 has higher premature Mortality but a lower infant. Group 4 has a lower Mortality rate around modal age at death. Others have different curves and shift to the right. They concluded that homogenization appears in most considered countries, while they follow the same mortality trend through the clusters. Men from different countries belonged to the same groups in recent years, except population in Eastern Europe. There is homogeneity between the women in Northern, Western, Southern Europe countries, and the mortality curve of women in Central and Eastern Europe countries look not

change much in last 50 years.

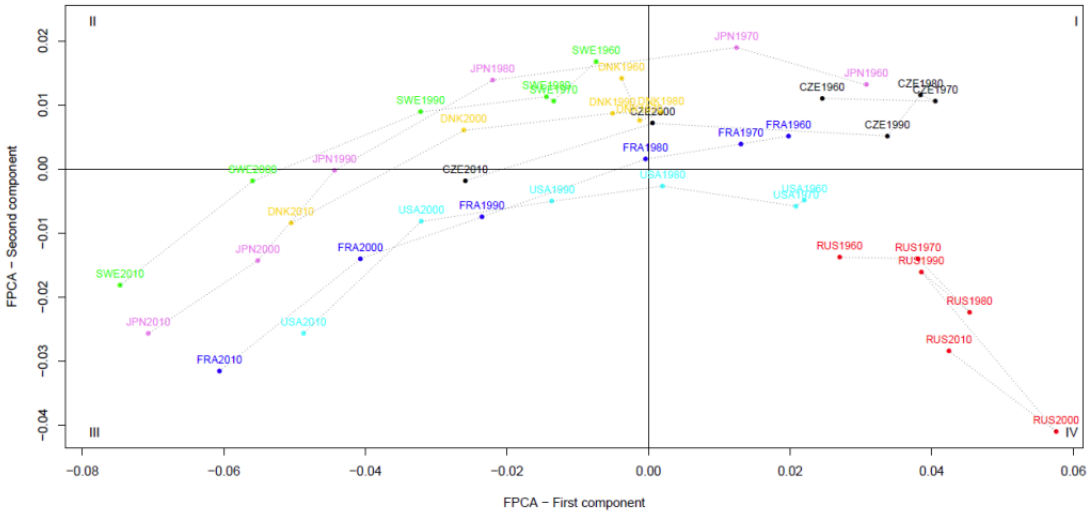


FIGURE 2: First principal subspace for Male

Source: Léger, Ainhoa-Elena  Mazzuco, Stefano (2020)

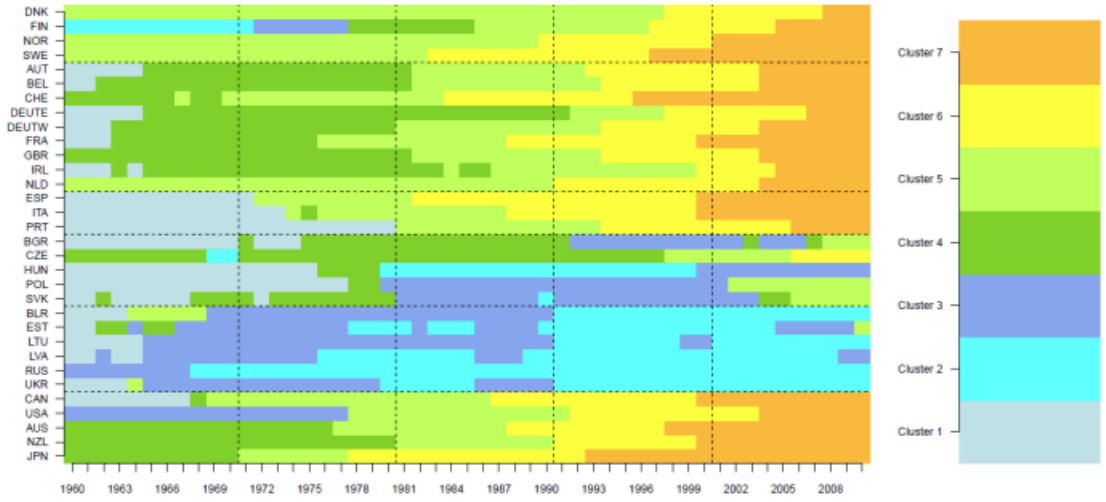


FIGURE 3: Model-based cluster analysis for women - Composition of the 7 clusters

Source: Léger, Ainhoa-Elena & Mazzuco, Stefano (2020)

In other cases, scholars focus on specific elements of mortality, ignoring the global pattern. In terms of extremely high ages at death, Medford (2019) compared the mortality rate between Denmark and Sweden for the cohorts born 1870 - 1904 and found that Danish centenarian lifespans are longer than those of Swedish people over 102 years old. As another example, Zanotto et al. (2020) focussed their study on premature mortality. They concluded that premature mortality has also evolved in the last years, with different

patterns for several countries. For instance, Zafeiris (2019) compared the mortality experience of 19 countries in Europe using the last available data between 2016 and 2015 by combining a modified Heligman–Pollard procedure and three cubic splines to smooth the life tables distribution and the estimate of several parameters. After applying the cluster methodology, the author found significant differences among the countries studied. Burke, M. (2016) researched surveys from 28 sub-Saharan African countries and concluded that the role of local, rather than national, is driving mortality patterns. Therefore they suggest a new approach to cluster with the district scale than a country. Curtin & Arias (2019) compared mortality trends by race and ethnicity among adults aged 25 and over in United States, in which they concluded the increase and decrease of age-adjusted death rate for different groups based on Ethnicity and Region.

In addition, in the on going pandemic, several publications cluster the mortality rate of victims between countries. Cerqueti & Ficcadenti (2022) analyzed the COVID-19 new deaths Peru million in 35 countries and applied rank-size models to cluster the regions. On a larger scale, Atsa'am & Wario (2020) using data from 206 countries, classified into three groups based on news cases, new deaths, and deaths because of the pandemic.

Lopez & Alan D (2001) completed research on the mortality table in 191 countries between 1980 and 2000. However, instead of using the mortality Rate as the feature in the models, they used a two-parameter logit life table system developed by Brazilss (1971) and Kraly  Norris (1978), which shows that different mortality related to each other by a logistic function. Lopez & Alan D (2001) concluded that there are differences in groups between developed and developing regions. Particularly in Africa, there is a similarity between Cameroon, Congo, and South Africa, which is higher than that figure in Benin, Senegal, and Togo. The reason come from different in economic development and social disparity.

In the first chapter in this report provided the research questions and briefly reviews the available publications about mortality clustering between countries. Section 2 gave overview of some cluster techniques and algorithms. Section 3 analyzed data and presented the logic behIndiafeature engineering. The results of the Cluster Model will be explained in Section 4. Finally, Section 5 will give the conclusion and some ideas for the following researchers.

## 2 METHODOLOGY

While processing the research with 32 countries, Léger & Mazzuco (2020) chose the age distribution of deaths as variables to cluster algorithms, instead of using age-specific rates found in some previous publications, since they wanted to focus on the mode and quantiles of data. Mazzuco, Scarpa  Zanotto (2018) have shown that the distribution of

deaths is more informative than the age-specific death rates, in some developed countries, regarding the shift transformation between ranges of ages. In addition, Basellini & Camarda (2019) illustrate that mortality rates, survival probabilities, and age distribution of deaths have strong relation.

In this report, to research the survival probabilities in a wide range of ages in various regions, including both developing and developed countries, Mortality rates will be chosen as observation for the clustering algorithm.

Jacques & Preda (2013) provided the overview of existing cluster algorithms. The Figure 4 summarizes the classical techniques: Raw - data clustering, model - based clustering and Machine Learning techniques.

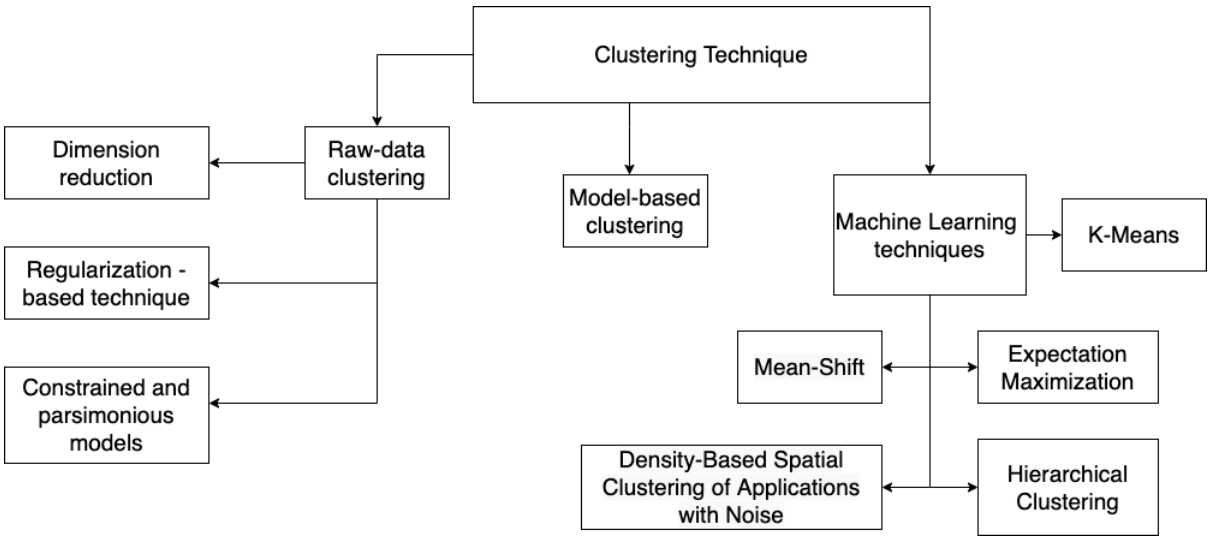


FIGURE 4: Summary the clustering techniques
Source: Summarize from Jacques & Preda (2013)

### *2.1 Raw-data clustering*

Raw-data clustering is the first technique to consider in any case study, since it is simple and easy to explain. The earliest approaches only deal with two-dimensional data, but recently many publications have provided an alternative method, which is helpful with the clustering of high-dimensional data. Bouveyron & Brunet (2013) summarize Raw-data clustering approaches: dimension reduction, regularization, and constrained and parsimonious.

The dimension reduction approach assumes that the number p of the independent variables is too large and suggests reducing to a lower dimension d, keeping the characteristic of original data. After the data is projected to low-dimensional space, it is possible to apply classical cluster algorithms, such as correlation comparison. Spline basis, which

was introduced by Wahba (1990) is the common choices because of their properties. One example is B-splines, proposed by F. Rossi (2004). However, the most popular dimension reduction method is principal component analysis (PCA), introduced by Pearson (1901), which we will describe briefly in next paragraph.

GiVenezuela one available data matrix Y [p x n], it is possible to transform Y to data matrix X [d x n], with d is smaller than p, by linear relation:

$$\mathrm{Y} = \tau X + \epsilon \tag{1}$$

Where $\tau$ is the y-intercept and $\epsilon$ is the slope of the line. These parameters were estimated by maximum likelihood of eigenvectors associated with the largest eigenvalues of the empirical covariance matrix of the data.

Another approach to cluster raw data in high dimensions is by measuring the covariance matrix between their variables. The covariance matrix formula for two variables x and y is:

$$cov_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N} \tag{2}$$

Friedman (1989) introduced regularized discriminant analysis (RDA) to numerically regularize the estimates of the covariance matrices before their inversion. Compared with Dimension reduction, the Regularization-based technique is more complex and suitable with time-series data. The reason is in time series problem, the variables is in favor of covariance measurements, which can show the relationship with previous data points.

### *2.2 Model-based clustering*

A other way to deal with the high dimension in clustering is to consider it a problem of over-parameterized modeling. Bouveyron & Brunet (2013) applied constrained Gaussian and parsimonious Gaussian models to reduce the number of free parameters. Although there are various choices of setting parameters and methodology to fit Gaussian models, this approach requires a strong assumption of the independence of the variables, which is unrealistic in several situations.

Developed from the Gaussian models approach above, a new set of methodology has been built for clustering, introduced by Madison & Lacroix (2013). This is a Bayes approach that assumed a density probability including a finite number of parameters to describe the clusters. The parameter is then to be used to cluster observations.

In this report, in the first step, Principal Component Analysis will be used to reduce the dimension of the mortality table. However, the drawback of classical clustering is that it reduces the information in the data set before processing. Nowadays, thanks to the

advanced technique of Machine Learning, we can approach this issue by clustering data without reducing their dimensions.

### *2.3 Machine Learning techniques*

Mitchell (1997) summarized from various papers that: Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is a part of artificial intelligence. Machine learning algorithms construct a model based on sample data, known as training data, to make predictions or decisions without being explicitly programmed to do so.

Machine learning approaches are divided into three broad categories, which correspond to learning paradigms, depending on the nature of the "signal" or "feedback" available to the learning system:

Supervised learning: The computer is presented with example inputs and their desired outputs, giVenezuela by the users, and the goal is to learn a general rule that maps inputs to outputs. There are two main algorithms in supervised learning: Classification and Regression.

Unsupervised learning: No labels are giVenezuela to the learning algorithm, leaving it on its own to fIndiastructure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data), or a means toward an end (feature learning). The main algorithms in Unsupervised learning are Clustering, which is analyzed in this report.

Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a specific goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

We compared five cluster algorithms before choosing the best approach for this report.

#### *2.3.1 K-Mean*

K-means clustering is a method of vector quantization to cluster observations into clusters where each observation belongs to the class with the nearest mean by minimizing the variances inside each group. Squared Euclidean distances always measure the distance. In detail, Lloyd (1957) and MacQueen (1967) provided the following formula for the K-Means method:

GiVenezuela a set of observations ($x_1$, $x_2$, ..., $x_n$), where each observation is a d-dimensional real vector, k-means will cluster the n observations into k (smaller than n) sets S = $S_1$, $S_2$, ..., $S_k$ to minimize the within-cluster sum of squares.

$$\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} (x - \mu_{Si})^2$$

Where $\mu_{Si}$ is the mean of the group $S_i$ and k is number of cluster.

K-Means required knowing the number of clusters in advance to process the algorithm; therefore, that is the main drawback of this methodology if it is not possible to know how many types of different mortality patterns. The advantages of K-Means are simple, high speed, and their algorithm follows the logic of our mortality cluster purpose, in which we want to group the country with the same expected value.

### *2.3.2 Mean shift*

Mean shift is a procedure to cluster discrete data sampled based on their density function. Overall, this approach will divide observations into groups having the highest number of the data point. In detail, Fukunaga and Hostetler (1975) provided the Gaussian Mean-Shift algorithm:

Let a kernel function $K(x_i - x)$ be given, which determines the weight of nearby points for re-estimation of the mean. For instance, the Gaussian kernel used for the distance of the current estimate will be chosen. The weighted mean of the density in the window determined by K is:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

where N(x) is the neighborhood of x, a set of points for which $K(x_i - x) \neq 0$. The Mean shift will replace x by m(x) until m(x) converges.

The best advantage of Mean shift is not to choose the number of clusters in advance; however, it takes more time to process the algorithm, and the density function approach is not straightforward as the expected value distance.

### *2.3.3 Density-Based Spatial Clustering of Applications with Noise*

Ester (1996) proposed another clustering technique, Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Overall, giVenezuela a data set with multiple points, DBSCAN will group points that are close-packed together, which ignores the mean of the dataset in some particular cases, and marks the points with high distance.

The most significant advantage of DBSCAN over K-means is that it works well with outlier data and does not require the number of clusters in advance. Compared with the Mean shift, DBSCAN can recognize the data points belonging to two different groups by having the same mean value. This algorithm's disadvantage is that it is complex and

requires a function called "region query," which needs time to estimate parameters and customize the model for particular cases.

### *2.3.4 Expectation–Maximization Clustering using Gaussian Mixture Models*

One drawback in both K-Mean and Mean-Shift is their naive use of the mean value for the cluster center, which leads them not to suitable while the mean of different cluster are near each other. In this case, another algorithm focusing on the data point's density model will be preferred.

The Gaussian Mixture Models (GMMs) approach assumes that the data points follow the Gaussian distribution rather than assuming the data points focus around the mean. By this premise, there are two parameters to describe the distribution: the average and the standard deviation. An optimization algorithm called Expectation-Maximization (EM) was used to estimate the parameter. After finding the mean and standard deviation, clustering would become simple.

The essential advantage of using GMMs is that they are more flexible about cluster covariance than K-Means, and standard deviation can be estimated. In addition, since GMMs is a probabilities distribution function, they can have multiple clusters for each data point, which makes this approach different from other algorithms, in which each data point only belongs to a particular group.

On the other hand, GMMs assume that the data points follow the Gaussian distribution, which is inappropriate for some types of data. Besides, the EM algorithm needs system resources and time to estimate the standard deviation, making this procedure slower.

### *2.3.5 Hierarchical Clustering*

Hierarchical Clustering is the method of building a hierarchy or a tree of the cluster; there are two types of Hierarchical Clustering: Agglomerative (bottom-up) or Divisive (top-down). Using Euclidean distance or Manhattan distance as a parameter to distinguish between two data points, users can customize the shape of the clusters. At start each point is counted as one group. The algorithm will reduce the number of clusters in each step, which will help approach the problems while it is not possible to know the number of clusters in advance. The drawback is that the result of cluster will be sensitive while building a hierarchy in a large data set.

## *2.4 Adopted Methodology*

After considering carefully between the above algorithms, we will choose Mean Shift as the methodology to cluster the mortality data. There are three reasons for this choice:

- It is not possible to know in advance the number of clusters, which is the drawback of K-means. Although in previous publications, many theories considered that a mortality rate in a country belongs to one of five different patterns it is not a solid foundation since our data set have several countries in twenty years, while the previous reports only focused on Europe and Africa. Therefore, the number of mortality clusters will not be assumed before running clustering.

- In our study case, it is convenient that two data points belong to the same class if they have same expectations. In detail, if two countries have the same expectation of mortality rate, they belong to same group. In reality, if two countries have the same mean and different standard deviations, they can come from dissimilar distributions, which make them belonged into different groups. However, in the scope of this report, that case will not be considered. From this assumption, we will not choose either DBSCAN, which ignores the mean of the data set, or GMM, which assumes data points follow Gaussian distribution.

- An algorithm that is simple and does not take time or computer resources to process data will be the priority. Beside that, the problem does not require complex solutions. Therefore we will not consider Hierarchical Clustering, GMM, or other advanced Machine Learning techniques.

### *2.5 Evaluation techniques*

It is necessary to build the parameter to determine the affection of the clustering technique. In other Machine Learning problems, such as classification or regression, there is the available, expected label that we want the algorithm to forecast or classify. For instance, when we want to classify a picture drawing a car or a river, we know that picture before letting the classification algorithm guess. Therefore when evaluating the result, we will provide a testing data set to measure how much the percentage algorithm can classify correctly. This is not true with clusters, in which we know in advance that there are some groups to which some data points belong, but we do not know exactly which data belong to which group; therefore, it is impossible to build testing data set. Another evaluation technique will be used, focusing on the distance between the group. The high distance between groups means that the clustering process is effective.

#### *2.5.1 Hopkins statistics*

Hopkins (1954) provided a statistical test to evaluate the efficiency of cluster procedures. Hopkins statistics measure the clustering tendency of a dataset. The null hypoth-

esis is that observations follow uniform distribution, which means data point distributed equally, therefore the cluster approaches do not have statistics.

In detail, let X be the set of n data points in d dimension. Generate a set Y of $m \leq n$ data points randomly from X, and define two distance measures by Euclid formula:

$u_i^d$, the distance of $y_i \in Y$ from its nearest neighbour in X, and

$w_i^d$, the distance of m number of randomly chosen $x_i$,$x_i \in X$ from its nearest neighbour in X.

The Hopkins statistic is defined as:

$$H = \frac{\sum_{i=1}^{m} u_i^d}{\sum_{i=1}^{m} u_i^d + \sum_{i=1}^{m} w_i^d}$$

A value close to 1 indicates that the data is highly clustered, close to 0.5 means the data is randomly distributed, and close to 0 means the data is uniformly distributed.

### *2.5.2 Silhouette coefficient*

Rousseeuw (1987) provided the Silhouette coefficient, a ratio calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each sample. The Silhouette Coefficient for a sample is

$$(\beta - \alpha)/max(\alpha, \beta)$$

In which, $\alpha$ is the mean intra-cluster distance and $\beta$ the mean nearest-cluster distance for each sample. The higher the Silhouette Coefficient, the farther away clusters are from each other. In data with a low dimension (less than 3), the result can be visualized so it is possible to evaluate the efficiency of the process. However, in case there is a higher dimension data which do not allow to visualize the result, Silhouette Coefficient will be used more effectively.

### *2.5.3 Variance Ratio Criterion*

Calinski and Harabasz (1974) built another score to determine the divergence of data points inside the cluster. It is also known as the Variance Ratio Criterion, which is the ratio of the between-clusters dispersion and inter-cluster dispersion for all clusters. The formula is:

$$\frac{BGSS}{k-1} / \frac{WGSS}{n-k}$$

Where BGSS is the between cluster sum-of-squares, WGSS the within cluster sum-of-squares, k the number of clusters and n the number of samples. Evaluating the ratio for

the different models with increasing k, the optimal clustering should be giVenezuela by the first local maximum of the ratios. Compared with Silhouette Coefficient, the Calinski-Harabasz index has advantages that relate to a standard concept of a cluster; this score is higher when clusters are dense and well separated, and it can be computed faster.

In Chapter 4, Hopkins statistics will be analyzed primarily to determine if the cluster is appropriate for the data, and in each cluster output, the Silhouette Coefficient and Calinski-Harabasz index will be calculated to evaluate the performance.

## 3 DATA PROCESSING

This report used the secondary data collection approach, a methodology in which data was previously gathered. The advantage of this approach is that a large amount of data can be accessible, while its drawback is lacking data, losing signal, or faking data. There are two sources of data: The Mortality Rate from "Mortality and global health estimates" of World Health Organization (WHO) Global Health Centre and Mortality Tables from Watson Willis Towers's system. Both sources are trustable: The data from WTW was published by the Institute and Faculty of Actuaries and similar organizations, while the data from WHO is continuously updated and has been used in previous researches. Moreover, using available life tables reduces the number of invalid data values and outliers.

In this report, 116 countries with the highest populations have been selected. China is the country with the highest population, about 1402 million, and Norway is the country with the lowest population, about 5 million. Because countries with a higher population are in Asia and Africa, these regions have the highest percentage of observations. On the other hand, only one country in Australia was collected.

Data from the last 19 years is available in each country, but only data points in 2000, 2005, 2010, 2015, and 2019 will be observed. Therefore in total there are 580 observations, for each one of them there are three groups of mortality rates for Man, Woman and Both, and 19 groups of age: '<1 year', '1-4 years', '5-9 years', '10-14 years', '15-19 years', '20-24 years', '25-29 years', '30-34 years', '35-39 years', '40-44 years', '45-49 years', '50-54 years', '55-59 years', '60-64 years', '65-69 years', '70-74 years', '75-79 years', '80-84 years', '85+ years'.

After collecting life tables from WHO and WTW, it is necessary to merge them into one dataset to clean and process model. In WHO´s tables, mortality rates was divided into age groups for each five cumulative years, while in WTW´s table those are unique ages. To transform data from WTW, the following formula will be applied:

$$ {}_5q_x = \frac{{}_5d_x}{l_x} $$

Where ${}_5q_x$ is the mortality rate between x and x + 5, and will be run for every five cumulative years, ${}_5d_x$ is the number of people who died, and $l_x$ is observation at the beginning of research process. Table I and Table II provide the sample data from each system.

| Country | Country_code | Year | Age Group | 15-19 years | 20-24 years |
|---|---|---|---|---|---|
| United Kingdom | GBR | 2010 | Both sexes | 0.000192 | 0.000213 |
| United Kingdom | GBR | 2010 | Male | 0.000202 | 0.000225 |
| United Kingdom | GBR | 2010 | Female | 0.000181 | 0.000211 |
| United Kingdom | GBR | 2005 | Both sexes | 0.000236 | 0.000238 |
| United Kingdom | GBR | 2005 | Male | 0.000275 | 0.000285 |
| United Kingdom | GBR | 2005 | Female | 0.000211 | 0.000214 |

TABLE I: Mortality sample data from WHO system

| Age x | *qx* |
|---|---|
| 16 | 0.000177 |
| 17 | 0.000185 |
| 18 | 0.000196 |
| 19 | 0.000208 |
| 20 | 0.000224 |
| 21 | 0.000243 |
| 22 | 0.000267 |
| 23 | 0.000295 |
| 24 | 0.000329 |

TABLE II: Mortality sample data from WTW system

The distribution of mortality rate will be calculated to evaluate the outlier value and clean the data. We calculated the average value of death rate between age groups in each countries, in each time period. There are 116 countries and 5 time points: 2000, 2005, 2010, 2015 and 2019, therefore total 580 data points will be collected. In each data points, we calculated the average value of mortality rate between 19 age groups. As the result, we will have 580 mean values. From there we will determine if there is any outlier values or not. The average death rate distribution is illustrated in box plot in Figure 5.

The rate for the Females is lower than the equivalent number for the Males. Meanwhile, the variance value for the Males is lower. There are some outlier values in both genders. These values are Haiti's average death rate in 2010, 27.04 percent for Males and 22.83 percent for Females. The reason for this outlier is Haiti's Earthquake, which led to an increase in the death rate for a short period. However, these values will not be removed from the database. They will be reviewed again after the clustering algorithm runs.

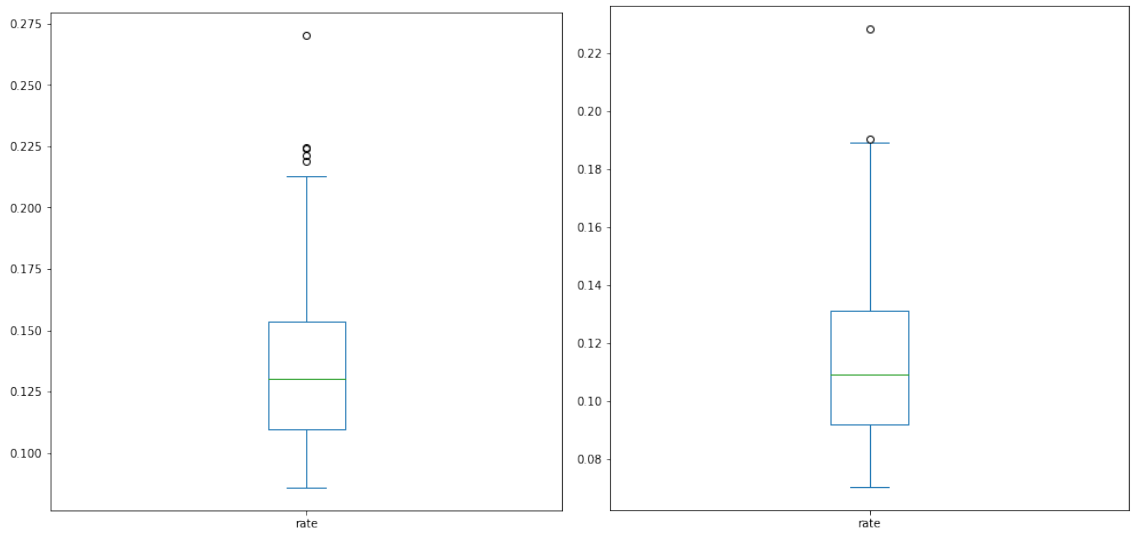


FIGURE 5: Average mortality rate distribution for males (left) and females (right)

Figure 6 shows the distribution of mortality rates in 116 countries in five time points. Therefore there are total 580 data points for each of 19 age groups. The x axis shows the age groups, while the y axis shows the mortality rate. Box plot was used to see the mean, the important quantiles, and the outliers.

There is an evident pattern of Mortality Rate when age increases. The rate was high in the first two groups ( <1 year and 1-4 years) before reducing continuously in the next 30 years. This period also records the high number of an outlier, but it is expected since those data points' value is appropriate 0. After 40 years old, the mortality rate increases slightly until 70, before rising sharply to 1 in group "85+ years". After considering carefully, only the value at "85+ years" will be removed since it is equal to 1 for every group.

Since Mean Shift used Euclid distance as a parameter to evaluate the density probability function, every data point should have a comparable value. It is essential to ensure a similar scale between data points in any Cluster Algorithm. As shown in Figure 6, the different periods have different data scales, leading to underfitting in Cluster Model. Therefore, it is necessary to divide the data set into small groups and research independently. Analyzing data, four groups will be considered:

- Mortality Data for children <1 year and 1-4 years: This group has a significantly high mortality rate in some developing countries, while extremely low in others.

- Mortality Data for people from 5 to 55 years old: extremely low in almost every region, with low variance and a high number of outlier values. In this range, data points have a small scale.

- Mortality Data for people 55 to 70 years old: the death rate increases with age, leading to the divergent of data points to a more extensive scale. Although the number of

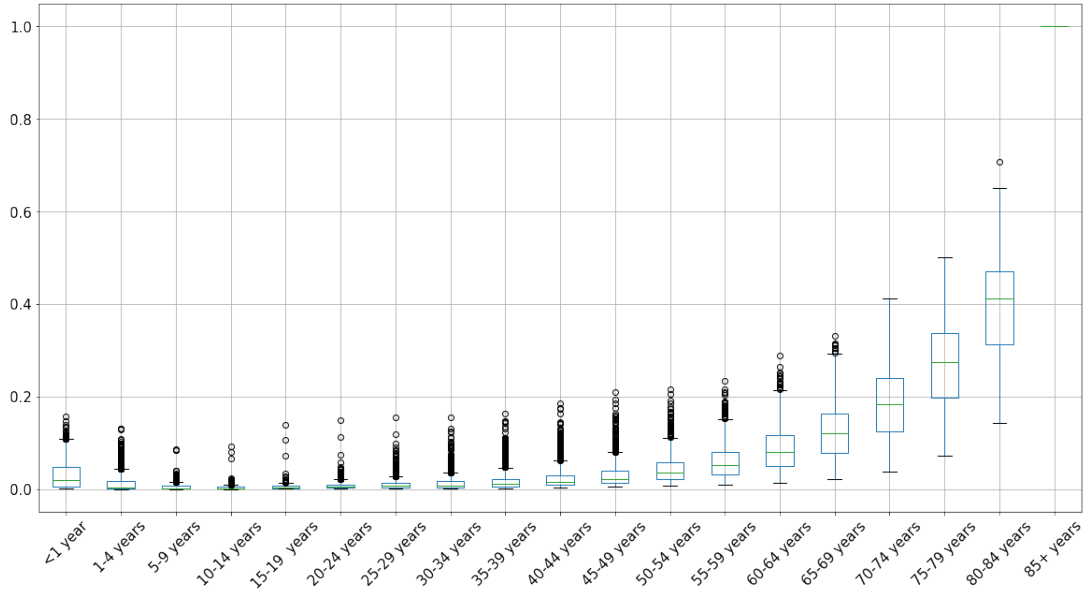


FIGURE 6: Distribution of mortality rates in 116 countries in 5 time point

outliers is lower, the divergence can cause the inefficient Cluster Algorithm.

- Mortality Data for the seniors older than 70 years old: High average value with high variance and do not have an outlier.

After analyzing, cleaning, and grouping data, they will be stored in the Database to prepare for running Cluster Algorithm. In the first step, PCA technique will be processed to cluster data overview and then focus on each group before researching important regions and providing a conclusion. Python in the Google-Collaboratory platform will be used to work with data. Available Machine Learning libraries, including Pandas, Scikit Learn, and Tensorflow, will be used to process workflow.

## 4 RESULT

### *4.1 Evaluation of cluster procedures*

Hopkins test was built to test the efficiency of cluster procedures. We use the hypothesis:

Null hypothesis (H0): Dataset comes from a random distribution and does not have statistically significant clusters.

Alternative hypothesis (H1): Dataset is significantly clusterable data.

Table III illustrates the Hopkins test ratio between five groups. Since all values are higher than 0.9 and near 1, we reject the null hypothesis and state that the dataset has statistically significant clusters. Among five groups, the cluster statistic characteristic is clearly in Senior and Adult Group A, while the PCA approach cannot visualize the

| Group evaluation | Hopkins test ratio |
|---|---|
| PCA Approach | 0.9372 |
| Children group | 0.9661 |
| Adult group A (5 - 55 years old) | 0.9771 |
| Adult group B (55 - 70 years old) | 0.9615 |
| Senior Group | 0.9752 |

TABLE III: Hopkins test ratio evaluation

difference between clusters.

### *4.2 Principal component analysis approach result*

In Figure 7, after visualizing, it is clear that the data point is not a random or unique distribution, which confirms the result of the Hopkins test. Different patterns exist for each part of the data point, such as concentration on the left side and dispersal on the right side of the distribution, meaning that the cluster will depend on the mean of observations. The Principal component analysis will be applied to analyze the distribution pattern, confirm the statistic of the clustering procedure again and review the efficiency of the Mean Shift algorithm. From these analyses, choosing the Mean Shift technique as the clustering algorithm is reasonable.

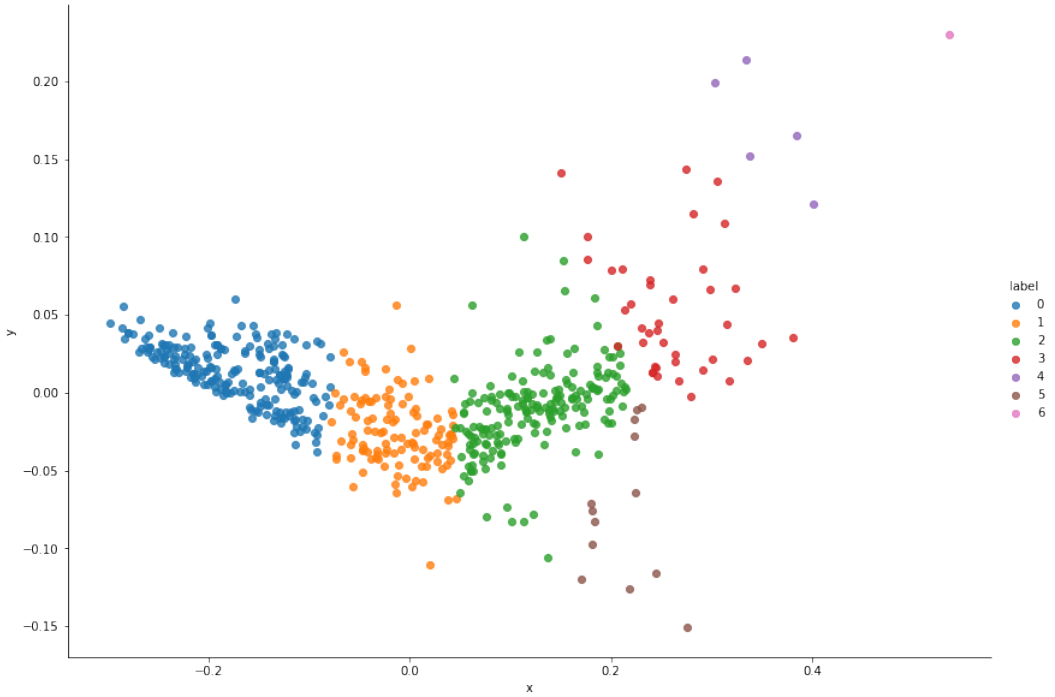


FIGURE 7: Cluster result from PCA approach

There is only one data point in cluster 6, Haiti, in 2010, with the highest mortality rate.

There is an apparent effect of the location on mortality rates: Middle East Asia countries such as Afghanistan, Azerbaijan, Tajikistan, and Syria are in cluster 5, while some African countries: Burundi, Malawi, Zambia, and Zimbabwe in cluster 4. Although both groups have high mortality rates, based on the scatter plot, it is clear that their points belonged to different patterns. This critical conclusion will be reviewed in detail regarding each group's age.

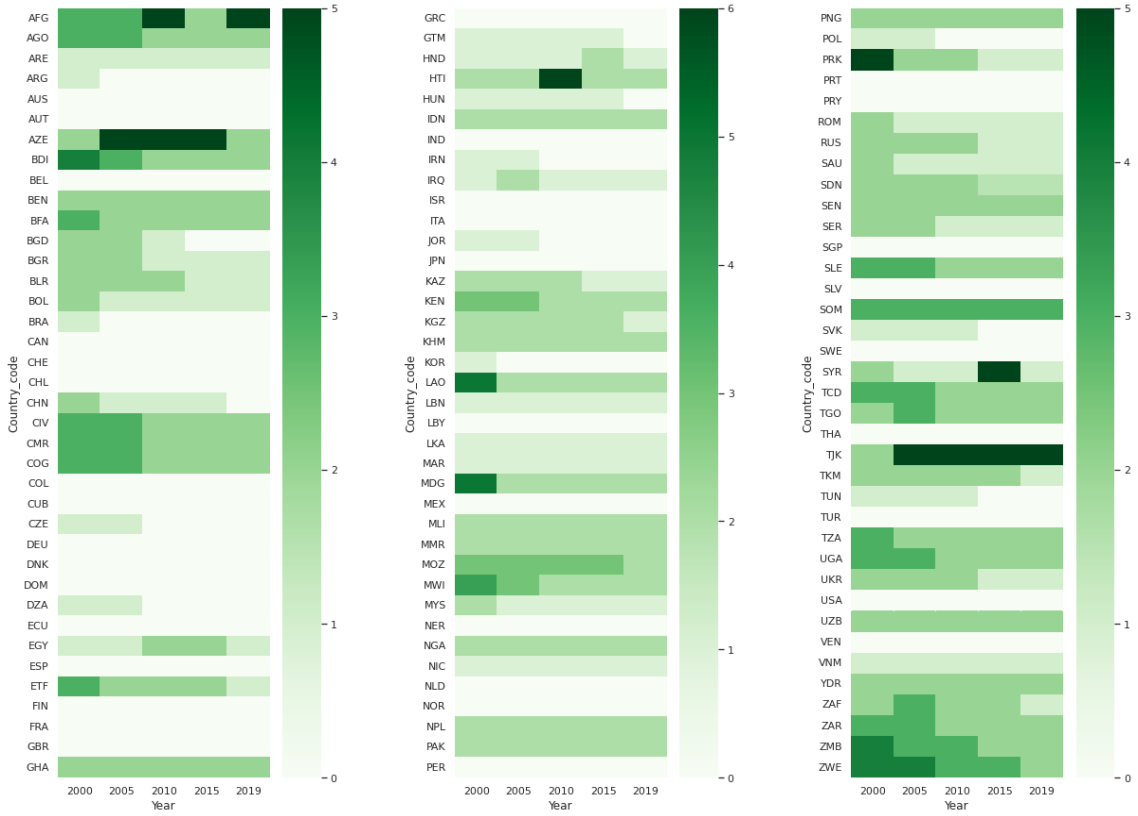


FIGURE 8: Heat map for cluster output of PCA Methodology

The heat map in Figure 8 illustrates the proceeding mortality rates between clusters. The shifting to cluster 3 from the higher group represents the reduction in death rate, which can be seen in some African and Asia countries. In Burundi, the death rate moved from cluster 4 to cluster 3 in 5 years, from 2000 to 2005, before reaching cluster 2 in 2010. Similar cases were experienced in Zimbabwe and Zambia in the last 20 years. In some Asia and Europe countries, the ratio shifted from group 2 to group 1, which has a lower value, and reached group 0, in which the lowest fatality rate was recorded. However, in the last 20 years, some countries have not changed their mortality groups, such as Argentina and Vietnam, which is still in cluster 1, or Marocco and Ghana, which is still in cluster 2. Since there are no countries still in cluster 4 or 5 in the same period, it is convenient to provide a conclusion that in a country having a low mortality rate, it is harder to reduce it.

Thirty-four counties belonged to cluster 0 in the last 29 years, most of which are

developed countries, which is the expected result. However, there are some developing countries, such as Indiaor Colombia in these groups, that question the cluster procedure's efficiency. This problem can be processed when deeply analyzing clusters in particular cases, by increasing the number of groups to detect the difference or by researching the difference in unique group ages.

### *4.3 Cluster result for children group ( < 5 years old)*

The analyzing mortality rate for children compares medical conditions between countries. The death rate for children is typically higher than that figure for adults. With the recent medical support and technological development, those rates are lower than in the past.

Figure 9 draws the scatter plot of clustering outcome with six different groups. On the y-axis is the mortality rate for children with less than one year, while the x-axis provides the equivalent number for children 1-4 years old. Overall, the pattern shows that the death rate for children in the (<1 year) group is higher than that number for children in (1-4 years). In almost developed country, those rates are approximately zero percent, while it is higher for developing countries.

There are two outliers in the scatter plot: Haiti in 2010 and Sierra Leone in 2000. While the first case is due to a natural disaster, the second case is due to the civil war. As the result, there is an increase in the death rate in the group [1-4 years], but not a rise in the equivalent figure in the group [< 1 year].

In normal conditions, the highest death rate was recorded virtually in African countries, such as Etitopia and Agola in the first ten years of this century, while the average value is between 10 and 12 percent for both groups. The average mortality rate decreased remarkably in clusters 3 and 2 and has the lowest value in cluster 1. The world's lowest death rate recorded for children is in Finland between 2010 and 2020, with 0.0017 and 0.0023 for each group, respectively. Other North Europe countries: Norway, Sweden, and Denmark, also experienced the highest survival opportunity for children in the last 20 years. In Asia, Japan and South Korea have the best chance of surviving in these groups, which is not surprising since both are developed countries with high levels of medical support and technology.

The heatmap 10 illustrates the development of survival opportunities for children between countries in the last 20 years. It is clear that there is a significant reduction in mortality rate in some regions, especially in Africa. In 2000, Agola and Sierra Leone were in the highest cluster [4], but now all are in the set [1], resulting from long-term development. The most impressive improvement can be shown in Senegal and UGA, from the cluster [3] 20 years ago; now, they are in the same set[0] as developed countries.

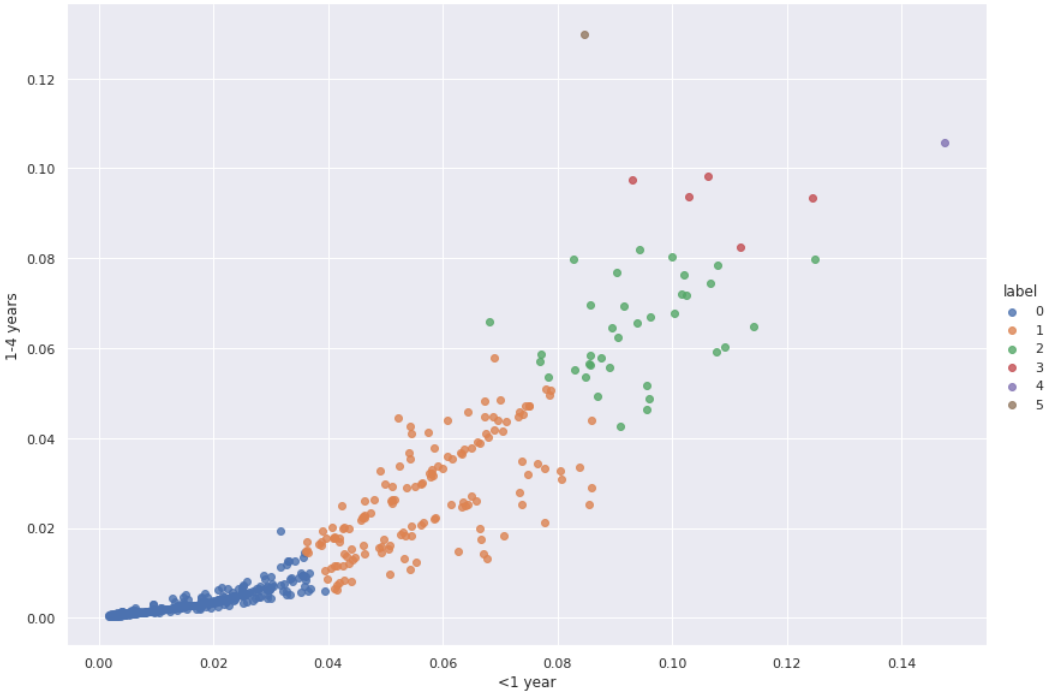


FIGURE 9: Cluster result for children group

Besides African countries, some nations in Asia also record a low chance of surviving children at the beginning of this century. For example, Yemen and Afghanistan have problems like low life quality or civil war. On the other hand, the developing countries in America recorded a lower death ratio. The highest figure was experienced in Bolivia from 2000 - 2010, in which they belong to cluster[1], before moving to set [0] in the last ten years. The reason may be that American countries are easier to receive medical support from developed countries than Africa because of their geographic location.

In conclusion, clustering data indicated that most countries today have a significantly lower mortality rate for children than in the past. Seventy percent of entire countries maintain a low ratio during the research period, not only in developed countries (United State, AUS, Germany) but also in developing countries (Brazil, TUN, Russia).

### *4.4 Cluster result for adult group A (5 - 55 years old)*

Adult group A, for people from 5 to 55 years old, includes the people with the lowest death ratio in their whole life. These groups have low mean, low variance, and a high number of outliers. Since there are ten dimensions in this case, which is equivalent to ten group ages, and no reduced dimension technique was processed, it is not convenient to visualize the scatter plot in a 2D or 3D diagram. With this group only, the Silhouette score and Clinski-Harabasz ratio will help evaluate the cluster procedure's effectiveness.

The Silhouette Coefficient illustrates the distance between different clusters. Table V

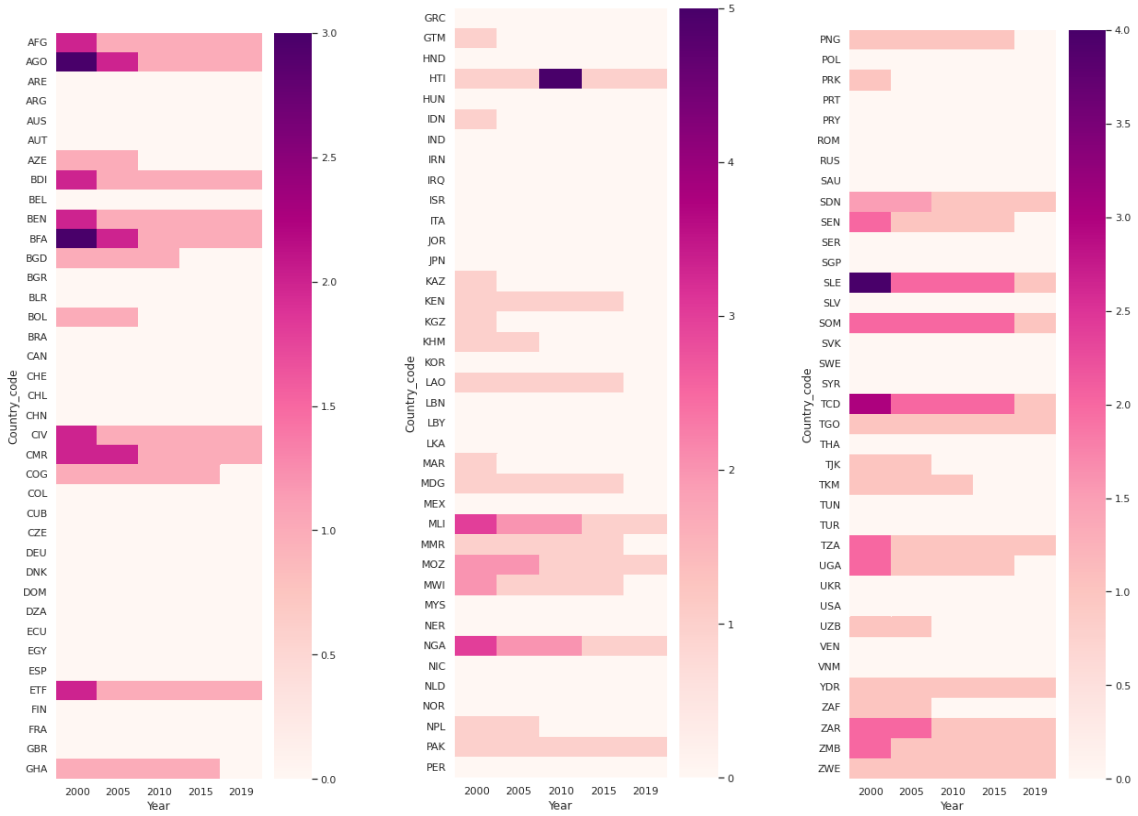


FIGURE 10: Heat map for cluster output of Children group

| Methodology | PCA approach | Children group | Adult group A (5-55) | Adult group B (55- 70) | Senior group |
|---|---|---|---|---|---|
| Silhouette score | -0.4160 | 0.6226 | 0.4541 | 0.6028 | 0.0608 |
| Calinski-Harabasz ratio | 0.9679 | 729.5351 | 558.3938 | 327.1250 | 154.2062 |

TABLE IV: Silhouette score and Clinski-Harabasz ratio comparison

shows that the Children group and Adult Group B have clear distance in cluster output, while Adult Group A is less obvious and hard to cluster than in previous cases. The reason can be the high effect on mortality rate from geography and politic in the children group, while this effect is lower for Adult Group A. Besides that, it is necessary to consider the effect of the mathematic scale on the data, while the observations with a low scale and near 0 tend to have a lower Silhouette score.

The Calinski-Harabasz index shows the distance between a data point and observations within and without the cluster. Analyzing this ratio shows that Adult group A has a high distance between the points from different clusters and a low distance between the points inside the same cluster. It is less apparent than the Children group in the last scatter plot but more precise than the PCA Approach. These analyses conclude that the cluster algorithm output from Adult group A is acceptable from statistical meaning.

The heatmap in Figure 11 shows the movement of mortality rate in the last 20 years. The movement in mortality cluster is not only from medical support or technology as

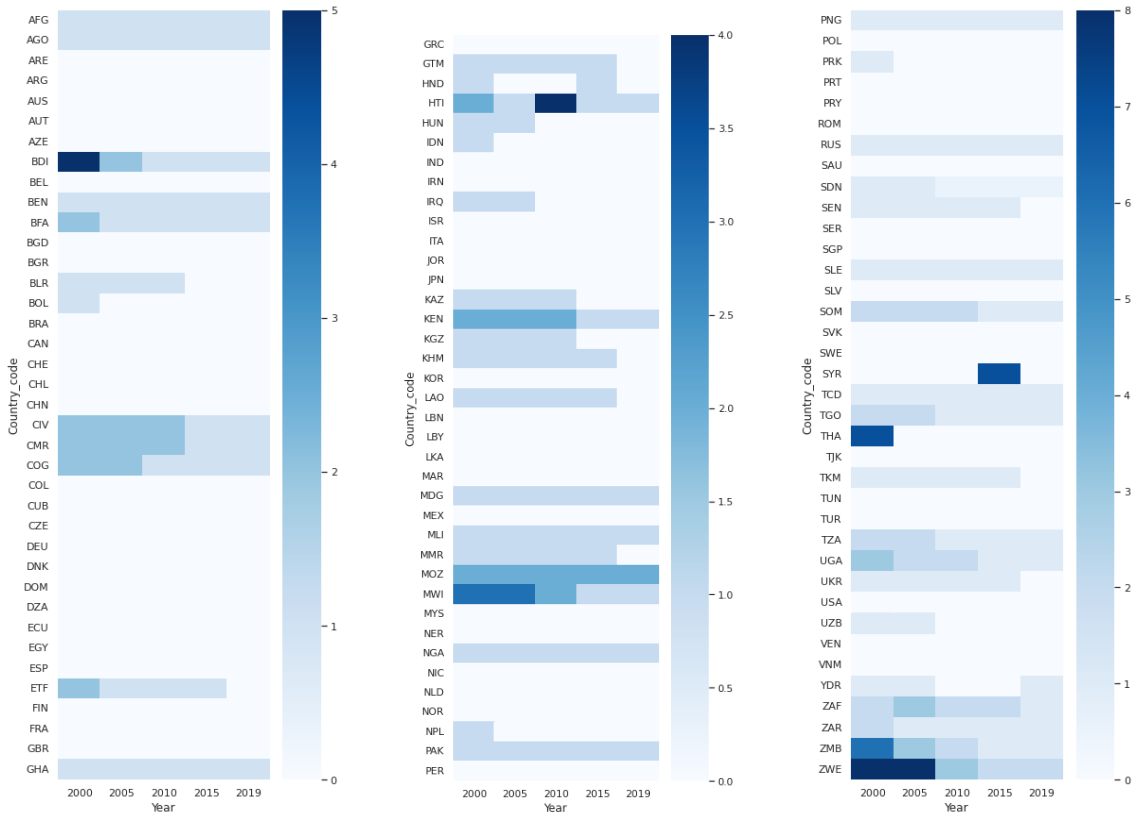


FIGURE 11: Heat map for cluster output of Adult Group A

children group, but also from the internal war, the pandemic, or the consequence of drawdown of economic and political in the earlier period. Some particular clusters have few members and high mortality rates, such as Haiti in 2010 or Burundi in 2000. Thailand recorded a higher mortality rate in 2000 for the adult group, at 1.7 percent, compared with their neighbor countries, such as 1.07 percent in Vietnam or 0.8 percent in Indonesia. The reason for this is Shifting Politics, Dragging the Economy, and Troubled Border in the ten final years of the last century. Syria in 2015 also recorded the lowest survival rate for 20 years when their civil war appeared.

There are two types of shifting from a high cluster to a lower one. In some countries with civil war or short-term drawdown of economic, such as Syria and Thaland, it takes them less than five years to return to their original mortality cluster. In addition, the pandemic and the economy's fall in the long term with the ineffective politic can take long-term to move to a better mortality group. Some developing countries in Africa, such as Mozambique and Kenya, only reduced their death rate slowly for adult people in whole last 19 years.

### *4.5 Cluster result for adult group B (55 - 70 years old)*

The data was split into groups A and B at 55 years old since, from this point, the mortality rate will increase gradually in all age groups. In the distribution of mortality rate diagram, these observations have a higher mean and variance than adult group A and a lower number of observations, which means there is a reduction in the difference in mortality rate between nations. Therefore it is harder for the Meanshift algorithm to identify a member in each cluster.

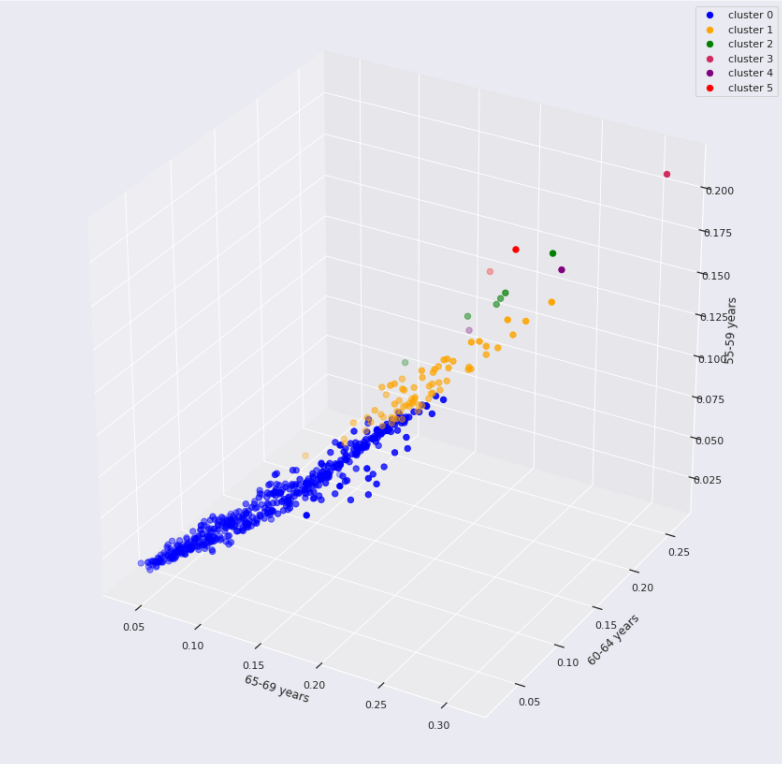


FIGURE 12: Cluster result for adult group B

Since adult group B is only group having three dimensions: the mortality rate from 55 - 59 years old, 60 - 64 years old, and 65 - 69 years old, this is the only group that we can use 3D scatter plot, which was provided in Figure 12

Cluster 0 is the class with the lowest mortality rate, but there is a difference between the distribution of data points in terms of dimensions, with some countries having higher survival opportunities in 55-59 but lower equivalent figures for people in 65-69. Japan and South Korea in 2019 recorded the lowest mortality rate globally for adults between 55 and 69 years old, with the smallest values for each age group of 1.6 percent, 2.6 percent, and 3.9 percent, respectively. In Europe, Switzerland rank above North Europe countries for the lowest death rate and lower than every country in America and Australia.

The most surprising result is that Nicaragua has a lower mortality rate than some

countries in northern Europe, such as Norway and Sweden, for people between 60 and 69 years, though their equivalent figure for children and adult group A is high. Egypt has the lowest mortality rate among countries in Africa. Almost other African countries belong to cluster 1 or 2, with an average death rate for people between 55 and 70 double as cluster 0, at about 11.29 percent. Another cluster: 3, 4, and 5 are outlier values with the pandemic event, such as HIV-AIDS in some Africa countries, leading to high mortality rate in that period.

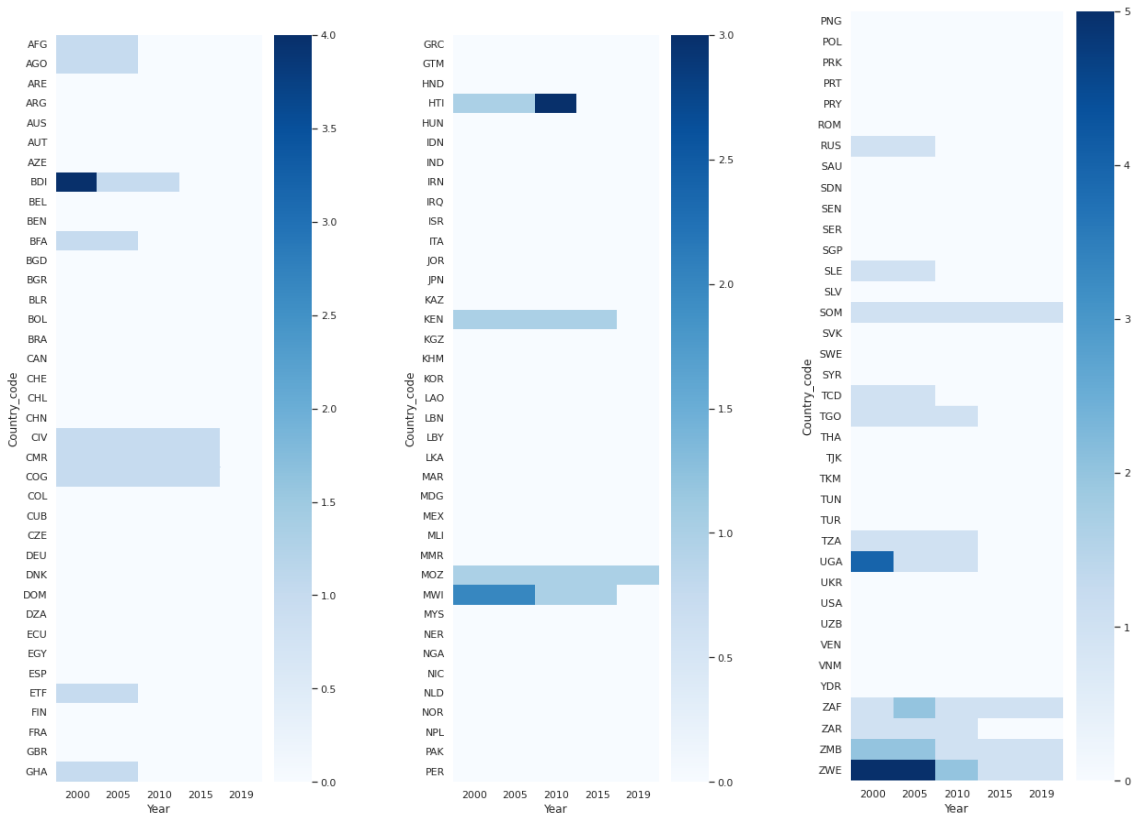


FIGURE 13: Heat map for cluster output of Adult Group B

The heat map on Figure 13 shows how the mortality rate for adult group B (55 - 70 years old) changes over time. Almost all countries in America, Europe, and Asia are in cluster 0, with a mortality rate of approximately zero in the whole 20 years. Without pandemic or catastrophic events, that ratio decreases gradually in every region. The explanation for this reduction is life's higher quality compared with the past and the development of medical support and technology. Although there was a low survival level for the researched group ages at the beginning of 21 century, Burundi, Malawi, and Zimbabwe had increased this figure in the last ten years to reach the same value as the developed countries. There is no development in some African nations, such as Mozambia, while their cluster is still in group 1 or 2 in the whole period. This analysis shows that despite the importance of geography and similarities between neighboring countries, the economic

condition of each nation in the long term still most significantly impacts the average death rate for people between 55 and 70 years old.

### *4.6 Cluster result for senior group (> 70 years old)*

After reaching 70 years old, the mortality rate of people will increase significantly. The difference in survival opportunities for senior people illustrates the health care system, the development of technology, and the natural condition. Since the adaption of humans to outside conditions decreases after 70 years old, natural condition has an essential role in people's health.

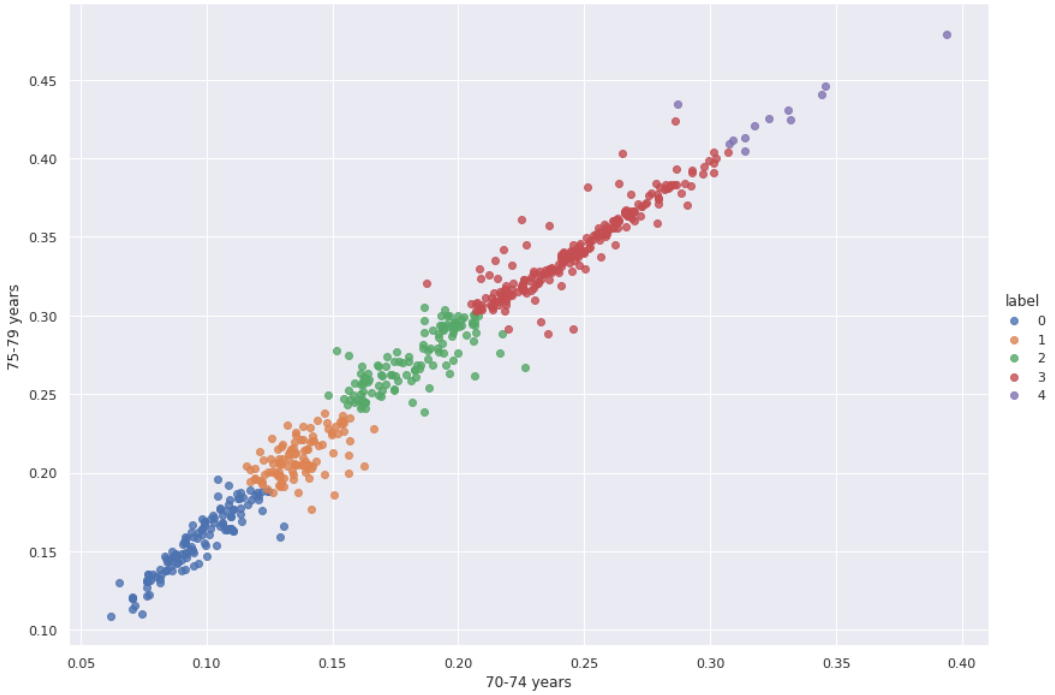


FIGURE 14: Cluster result for senior group

The scatter plot 14 shows the distribution of cluster output. Cluster observations are more challenging than other groups since every data point lies in a linear line, which means they have an identical pattern. In this situation, the Mean Shift algorithm only evaluates the mean value to determine which cluster an observation belongs to. There are five clusters, with the x-axis being 70-74 years old and the y-axis being 75-79 years old. In the highest cluster, with includes an outlier for Haiti-2010, the average death probability for group 70-74 years old is 35 percent, while the equivalent figure for group 75 - 79 years old is 45 percent. Almost all came from African countries in 2000. There is one particular case Tajikistan in Middle East Asia, in 2015, while this country experienced an earthquake. Other countries in Africa and Asia remain in clusters 2 and 3. In addition, with similar natural conditions, the health care system in countries is noteworthy.

Comparing South Korea and Pakistan, which are near each other, Pakistanistan belongs to cluster 3 in the 20 years, while South Korea lies in cluster 1 from 2000-2005, before reaching cluster 0 in 2005-2009.

The highest survival opportunity was recorded in developed countries in Europe, Asia, and America, with NOR, Japan, and CAN, respectively. While Norway and CAN have a low number of people, making it easy to manage the benefits of people, the high population in Japan shows the impression that the government health care system ensures a high quality of life. Some other countries in cluster 0 are Australia, France, Italia, Sweden.

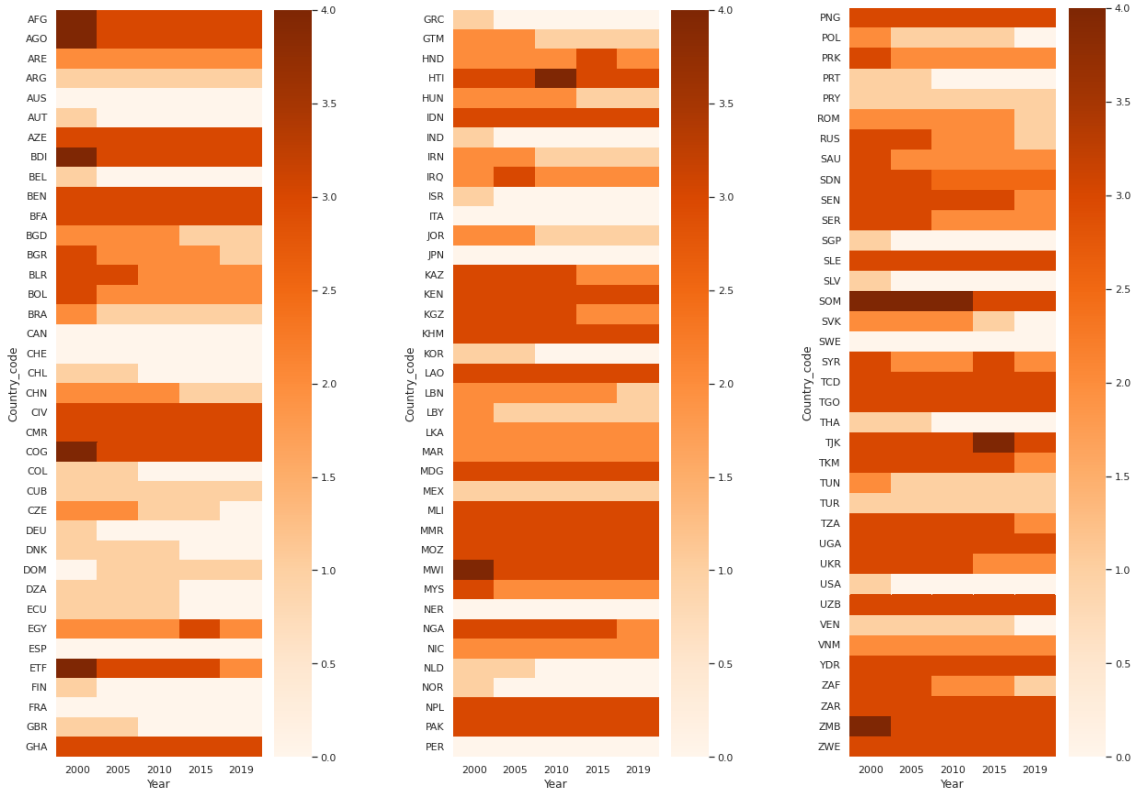


FIGURE 15: Heat map for cluster output of Senior Group

There is an increase in survival probability among countries in Figure 15. There is a growth significantly both in some developed countries from cluster 1 to cluster 0, such as United States, Isarel, or United State, and in some developing countries, such as Thailand and India, which shows that it is not necessary to trade-off between the economic development and the benefit of people. In some developing countries, this growth rises slightly, for instance, China, and Jordan, from cluster 2 to cluster 1. In addition, almost all developing countries in Asia and America belong to clusters 1 or 2, and some regions did not moved in the whole research period, such as Laos or Vietnam.

On the other hand, almost all developing countries in Africa belong to clusters 3 or 4. However, they recorded development in the last 20 years. In the 2000s, almost all

countries were in cluster 4 with the highest mortality, but today they are in cluster 3; some countries Venezuela reached cluster 2 in 2019, such as Senegal. North Africa and the Middle East experience some fluctuations, Egypt in 2015, moving from cluster 2 to 3 and returning after five years, or Syria and Iraq between 2000 and 2010.

### *4.7 Clustering future mortality at older ages in Portugal and Spain*

This section will provide an application of the Cluster Methodology of Mortality Rate to forecast the cluster of this ratio in the future. The difference with analysis, which required a low number of clusters, is that it is necessary to have a higher number since we need to clearly see how the Mortality Rate improved in the last 20 years. The group "Adult Age B" and "Senior" will be chosen to research. Figure 16 show the development of Survival Probability for Male in Venezuela European countries, from right to left, during 20 years, in which cluster 1 is the lowest Mortality Rate and cluster 14 is the highest. X axis is the rate for Adults aged (55 - 70), and the Y axis is the equivalent figure for seniors (More than 70 years old).

In 2000, almost all countries were in clusters 14 and 5, which is the lowest survival probability, and in the last 20 years, they have developed to a smaller set, with France in cluster 1 in both group ages. The countries in center Europe have the same pattern since Belgium also moved to cluster 1 in the senior group, while Spain and Germany have higher clusters and cannot reduce their Mortality rate significantly, especially in the Senior Age group.

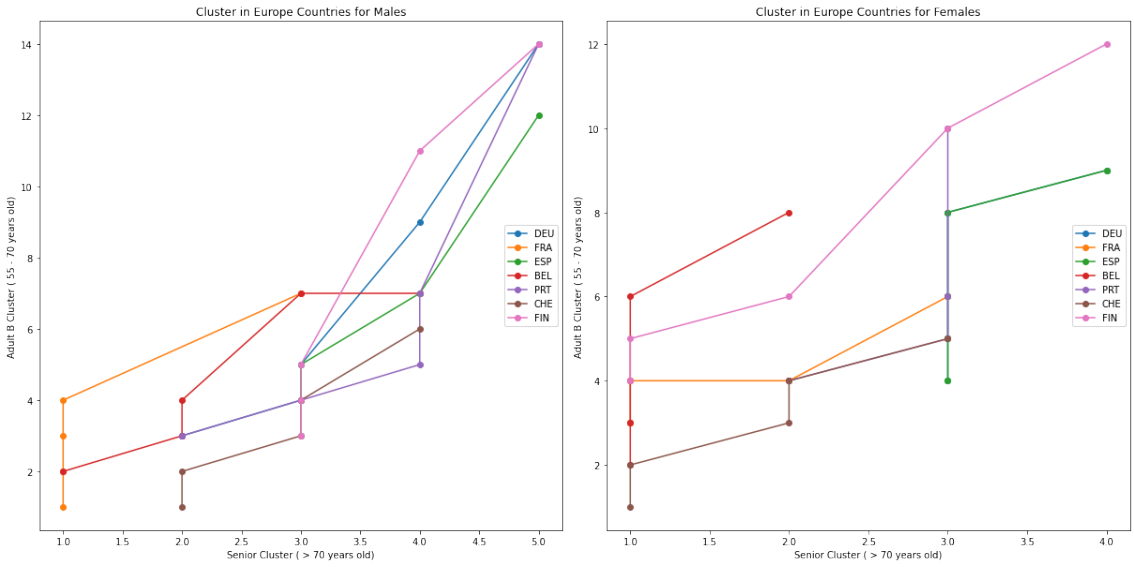


FIGURE 16: Development of Mortality rate in Europe

Figure 16 also shows the equivalent rate for females. While other countries, such as Finland, Germany, and Switzerland, moved to the lowest cluster for the Senior group,

Spain and Portugal remain in the higher cluster. In terms of Adult Group B mortality rate, all countries gradually reduced in the last 20 years.

This analysis provides the estimation for the cluster of Portugal in the next ten years (2020 - 2030). Table VI summarizes the estimation cluster and mean value of mortality for Portugal and Spain in the next ten years. Estimating the expected value and standard deviation from the observations in the same cluster is possible. In the Senior group, Portugal can move to cluster 1, which has the lowest mortality rate for males, and remain in cluster 3 for females. In the Adult group age B, Portugal can move to cluster 2 for males in the next five years, while the equivalent cluster for females is 3. The mortality rate for the male in the Senior group in Portugal is 19.12 percent, slightly lower than this value in Spain. In the Adult Group, the number for Portugal was also lower than Spain, at 5.49 and 5.64, respectively. For the death rate of the females, both countries would have an identical cluster, with the same mean in both Senior and Adult groups, at 15.79 and 2.48 percent, respectively.

| Group | Cluster | Mean | Standard_deviation |
|---|---|---|---|
| Senior_male_Portugal | 1 | 0.1912 | 0.0193 |
| Senior_female_Portugal | 3 | 0.1579 | 0.0252 |
| Adult_male_Portugal | 2 | 0.0549 | 0.0010 |
| Adult_female_Portugal | 3 | 0.0248 | 0.0015 |
| Senior_male_Spain | 2 | 0.1915 | 0.0266 |
| Senior_female_Spain | 3 | 0.1579 | 0.0252 |
| Adult_male_Spain | 3 | 0.0564 | 0.0027 |
| Adult_female_Spain | 3 | 0.0248 | 0.0015 |

TABLE V: Mortality rate estimation for Portugal and Spain in 2030

### *4.8 Clustering future mortality at older ages in Thailand and Vietnam*

This section estimates the mortality rate in some Asia countries in the next ten years. Figure 17 summarizes the difference in survival probability in Asia. In contrast with Europe, while there are similarities between countries, in Asia, there are notable differences between developed countries (Japan or Korea) and developing countries (Indonesia, Vietnam). Japan has the lowest mortality rate, and no other countries have had the same ratio for the last 20 years. There is a considerable development in some countries, such as Vietnam, China, or Thailand; however, since there is a significant difference between countries' mortality rates, it is difficult to cluster and estimate future ratios.

Table VII provides the expected value and standard deviation of the Mortality rate in Vietnam and Thailand in 2030. There is a similarity between the two countries, except for the higher survival probability of females in Adult Group B in Thailand compared with

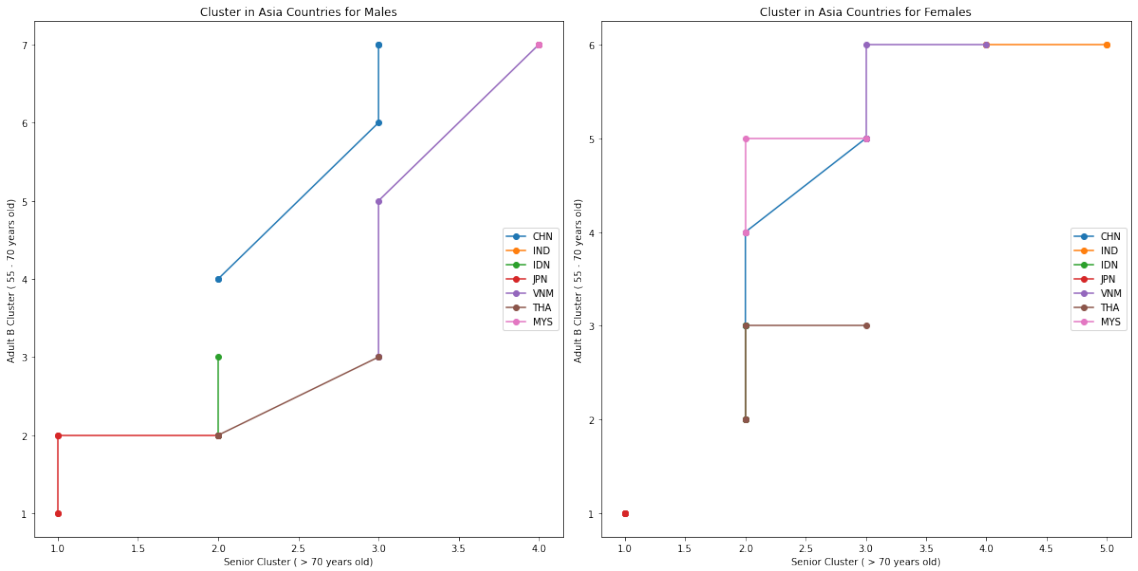


FIGURE 17: Development of Mortality rate in Asia

Vietnam. It is an estimation that in Vietnam, the cluster of the Senior group will be 2 in both sexes in the next ten years, while the cluster for Adult Group B will be 1 for Males and 4 for Females. in term of Thailand, the estimated cluster for both Senior and Adult group B will be 2 for both sexes in next ten years.

| Group | Cluster | Mean | Standard_deviation |
|---|---|---|---|
| Senior_male_ Vietnam | 2 | 0.2156 | 0.0135 |
| Senior_female_ Vietnam | 2 | 0.1927 | 0.0389 |
| Adult_male_Vietnam | 2 | 0.0596 | 0.0053 |
| Adult_female_ Vietnam | 4 | 0.0580 | 0.0020 |
| Senior_male_Thailand | 2 | 0.2156 | 0.0135 |
| Senior_female_Thailand | 2 | 0.1927 | 0.0389 |
| Adult_male_Thailand | 2 | 0.0596 | 0.0053 |
| Adult_female_Thailand | 2 | 0.0377 | 0.0094 |

TABLE VI: Mortality rate estimation for Vietnam and Thailand in 2030

## 5 CONCLUSION

In this report, Mean Shift, a machine learning algorithm, was used to analyze and estimate the cluster of Mortality rates in 116 countries. After processing, analyzing, and researching, we provide the following conclusion:

The Mean Shift algorithm can be used to cluster Mortality Rate, which is most suitable for the Children group. Senior and Adult groups do not show differences when clustered. The principal component analysis is a helpful technique to reduce the dimension of the

data set and support the initial analysis. Analyzing the history of a country´s cluster can provide the forecast for its future, as well as that country expected death rate and its standard deviation

There is a development in every country from high mortality rate cluster to lower one, but in high rate clusters, there are two types of pattern, one is familiar in Africa, and the other is familiar in East Asia. The main difference between them is that in Africa, the death rate for children is higher, but the ratio for Adults and seniors is lower. When the country is more developed, these patterns become a single cluster, with a low mortality rate in every age group.

Among countries with high mortality rates, there are two scenarios: Countries that experience a civil war or drawdown of the economy in the short term, these countries can revive after 2 or 5 years with appropriate policy; Or Countries that experience a national pandemic or government mismanagement in the long period, which leads to low life quality and higher periods of time to recharge.

It is well known that developed countries have lower mortality rates than developing ones. However, the developing countries can improve their survival experience by improving health support and living standards. The countries near each other will have more chances to have similar patterns of mortality reduction, but the gap between neighboring regions is higher in Asia than in Europe and Africa. One suggestion for the following research is to analyze the difference in mortality rate in Adult Group A (5 - 55 years old) across different regions in Asia since the clustering result shows that the dissimilarity between them is the main reason for the divergence in the ratio between neighbor countries.

Finally, I want to state my gratitude to Willis Tower Watson for allowing me to have this internship and for providing the necessary data and support. I also want to thank Ana Sousa, my supervisor, and mentor, for her support and help in reviewing these reports, and professor Onofre Simoes for his advice and comments.

## REFERENCES

Atsa'am, Donald and Wario, Ruth. (2020). Hierarchical cluster analysis of the morbidity and mortality of COVID-19 across 206 countries, territories and areas. International Journal of Medical Engineering and Informatics. 10.1504/IJMEI.2020.10033328.

Basellini, U., and Camarda, C. G. (2019). Modelling and forecasting adult age-at-death distributions. Population studies, 73(1), 119–138. https://doi.org/10.1080/00324728.2018.1545918

Bouveyron, Charles and Brunet, Camille. (2013). Model-Based Clustering of High-Dimensional Data: A review. Computational Statistics and Data Analysis. 71. 1-27. 10.1016/j.csda.2012.12.008.

Brass, W. (1971). On the scale of mortality, in W. Brass (ed.), Biological Aspects of Demography. London: Taylor Francis.

Brian Hopkins and J. G. Skellam (1954) A New Method for determining the Type of Distribution of Plant Individuals, Annals of Botany, Volume 18, Issue 2, April 1954, Pages 213–227,

Burke, M., Heft-Neal, S., and Bendavid, E. (2016). Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. The Lancet. Global health, 4(12), e936–e945. https://doi.org/10.1016/S2214-109X(16)30212-1

Caselli, Graziella and Vallin, Jacques. (2002). Epidemiologic transition theory exceptions. Genus. 58. 10.2307/29788712.

Cerqueti, R. and Ficcadenti, V. (2022). Combining rank-size and k-means for clustering countries over the COVID-19 new deaths Peru million. Chaos, Solitons and Fractals. 158, p. 111975. https://doi.org/10.1016/j.chaos.2022.111975

Curtin, Sally and Arias, Elizabeth. (2019). Mortality Trends by Race and Ethnicity Among Adults Aged 25 and over: United States, 2000-2017. NCHS data brief. Data brief. 1-8.

Ester, M.; Kriegel, H.-P.; Sander, J. and Xu, X. (1996), A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in 'Proc. of 2nd International Conference on Knowledge Discovery and' , pp. 226-231 .

F. Rossi, B. Conan-Guez, and A. El Golli. (2004) Clustering functional data with the Somalia algorithm. In Proceedings of ESANN 2004, pages 305–312, Bruges, Belgium, April 2004.

Fukunaga, Keinosuke and Larry D. Hostetler (1975). "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition". IEEE Transactions on Information Theory. 21 (1): 32–40.

G. Wahba. Spline models for observational data. SIAM, Philadelphia, 1990.

Giacofci, Madison Lambert-Lacroix, S. Marot, Guillemette and Picard, Franck. (2013). Wavelet-Based Clustering for Mixed-Effects Functional Models in High Dimension. Biometrics. 69. 10.1111/j.1541-0420.2012.01828.x.

Jacques, Julien and Preda, Cristian. (2013). Functional Data Clustering: A Survey. Advances in Data Analysis and Classification. 8. 231-255. 10.1007/s11634-013-0158-y.

Jerome H. Friedman (1989) Regularized Discriminant Analysis, Journal of the American Statistical Association, 84:405, 165-175, DOI: 10.1080/01621459.1989.10478752

Karl Pearson F.R.S. (1901) LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2:11, 559-572, DOI: 10.1080/14786440109462720

Konstantinos N. Zafeiris (2019) Mortality Differentials Among the Euro-Zone Countries: An Analysis Based on the Most Recent Available Data, Communications in Statistics: Case Studies, Data Analysis and Applications, 5:1, 59-73, DOI: 10.1080/23737484.2019.1579682

Kraly, E. P., and Norris, D. A. (1978). An Evaluation of Brazilss Mortality Estimates Under Conditions of Declining Mortality. Demography, 15(4), 549–557. https://doi.org/10.2307/2061206

Léger, Ainhoa-Elena and Mazzuco, Stefano (2020). What can we learn from functional clustering of mortality data? An application to HMD data. Applications (stat.AP), FOS: Computer and information sciences, FOS: Computer and information sciences, G.3; J.3, 62P25

Lloyd, S. P. (1957). Least squares quantization in PCM. Technical Report RR-5497, Bell Lab, September 1957.

Lopez, Alan D, Salomon, Joshua A, Ahmad, OMorocco B, Murray, Christopher J. L, Mafat and Doris. et al. (2001). Life tables for 191 countries : data, methods and results. World Health Organization.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam J. Neyman (Eds.), Proceedings of the fifth Berkeley

symposium on mathematical statistics and probability (Vol. 1, pp. 281–297). California: University of California Press.

Mazzuco, Stefano, Scarpa, Bruno and Zanotto, Lucia. (2018). A mortality model based on a mixture distribution function. Population Studies. 72. 1-10. 10.1080/00324728.2018.1439519.

McMichael, Anthony and Mckee, Martin, Shkolnikov, Vladimi, Valkonen, Tapani. (2004). Mortality trends and setbacks: Global convergence or divergence?. Lancet. 363. 1155-9. 10.1016/S0140-6736(04)15902-3.

Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

Medford, A., Christensen and K., Skytthe, A. et al. (2019) A Cohort Comparison of Lifespan After Age 100 in Denmark and Sweden: Are Only the Oldest Getting Older?. Demography 56, 665–677. https://doi.org/10.1007/s13524-018-0755-7

Olshansky, S. J., and Ault, A. B. (1986). The Fourth Stage of the Epidemiologic Transition: The Age of Delayed Degenerative Diseases. The Milbank Quarterly, 64(3), 355–391. https://doi.org/10.2307/3350025

Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. Milbank Quarterly, 39(4, Pt. 1), 509–538. https://doi.org/10.2307/3349375

Peter J. Rousseeuw (1987). “Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis”. Computational and Applied Mathematics 20: 53-65.

T. Calinski and J. Harabasz, (1974). “A dendrite method for cluster analysis”. Communications in Statistics

## GLOSSARY

**AFG** Afghanistan. 40

**AGO** Angola. 40

**ARE** United Arab Emirates. 40

**ARG** Argentina. 40

**AUS** Australia. 40

**AUT** Austria. 40

**AZE** Azerbaijan. 40

**BDI** Burundi. 40

**BEL** Belgium. 40

**BEN** Benin. 40

**BFA** Burkina Faso. 40

**BGD** Bangladesh. 40

**BGR** Bulgaria. 40

**BLR** Belarus. 40

**BOL** Bolivia. 40

**BRA** Brazil. 40

**CAN** Canada. 40

**CHE** Switzerland. 40

**CHL** Chile. 40

**CHN** China. 40

**CIV** Côte d'Ivoire. 40

**CMR** Cameroon. 40

**COD** DR Congo. 40

**COL** Colombia. 40

**CPI** Consumer price index. 40

**CUB** Cuba. 40

**CZE** Czech Republic (Czechia). 40

**DEU** Germany. 40

**DNK** Denmark. 40

**DOM** Dominican Republic. 40

**DZA** Algeria. 40

**ECU** Ecuador. 40

**EGY** Egypt. 40

**ESP** Spain. 40

**ETH** Ethiopia. 40

**FIN** Finland. 40

**FRA** France. 40

**GBR** United Kingdom. 40

**GDP** Gross domestic product. 40

**GHA** Ghana. 40

**GIN** Guinea. 40

**GMM** Gaussian Mixture Models. 40

**GRC** Greece. 40

**GTM** Guatemala. 40

**HND** Honduras. 40

**HOK** Hong Kong. 40

**HTI** Haiti. 40

**HUN** Hungary. 40

**IDN** Indonesia. 40

**IND** India. 40

**IRN** Iran. 40

**IRQ** Iraq. 40

**ISR** Israel. 40

**ITA** Italy. 40

**JOR** Jordan. 40

**JPN** Japan. 40

**KAZ** Kazakhstan. 40

**KEN** Kenya. 40

**KHM** Cambodia. 40

**KOR** South Korea. 40

**KYR** Kyrgyzstan. 40

**LAO** Laos. 40

**LBN** Lebanon. 40

**LBY** Libya. 40

**LKA** Sri Lanka. 40

**MAR** Morocco. 40

**MDG** Madagascar. 40

**MEX** Mexico. 40

**MLI** Mali. 40

**MMR** Myanmar. 40

**MOZ** Mozambique. 40

**MWI** Malawi. 40

**MYS** Malaysia. 40

**NER** Niger. 40

**NGA** Nigeria. 40

**NIC** Nicaragua. 40

**NLD** Netherlands. 40

**NOR** Norway. 40

**NPL** Nepal. 40

**PAK** Pakistan. 40

**PER** Peru. 40

**PHL** Philippines. 40

**PNG** Papua New Guinea. 40

**POL** Poland. 40

**Portugal** Portugal. 40

**PRK** North Korea. 40

**PRY** Paraguay. 40

**ROU** Romania. 40

**RPI** Retail price index. 40

**RUS** Russia. 40

**RWA** Rwanda. 40

**SAU** Saudi Arabia. 40

**SDN** Sudan. 40

**SEN** Senegal. 40

**SGP** Singapore. 40

**SLE** Sierra Leone. 40

**SLK** Slovakia. 40

**SLV** El Salvador. 40

**SOM** Somalia. 40

**SRB** Serbia. 40

**SSD** South Sudan. 40

**SWE** Sweden. 40

**SYR** Syria. 40

**TCD** Chad. 40

**TGO** Togo. 40

**THA** Thailand. 40

**TJK** Tajikistan. 40

**TKM** Turkmenistan. 40

**TUN** Tunisia. 40

**TUR** Turkey. 40

**TZA** Tanzania. 40

**UGA** Uganda. 40

**UKR** Ukraine. 40

**USA** United States. 40

**UZB** Uzbekistan. 40

**VEN** Venezuela. 40

**Vietnam** Vietnam. 40

**WHO** World Health Organization. 40

**WTW** Willis Tower Watson. 40

**YEM** Yemen. 40

**ZAF** South Africa. 40

**ZMB** Zambia. 40

**ZWE** Zimbabwe. 40

## A APPENDICES

- The link for full coding in google colaboratory can be found in here:

  https://colab.research.google.com/drive/1dIJMKbR04zkImHJnbblBPiJFa7WMJQ36?usp=sharing

- The detail clusters for principal component analysis approach:

| Country_code | 2000 | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|---|
| Afghanistan | 3 | 3 | 5 | 2 | 5 |
| Angola | 3 | 3 | 2 | 2 | 2 |
| United Arab Emirates | 1 | 1 | 1 | 1 | 1 |
| Argentina | 1 | 0 | 0 | 0 | 0 |
| Australia | 0 | 0 | 0 | 0 | 0 |
| Austria | 0 | 0 | 0 | 0 | 0 |
| Azerbaijan | 2 | 5 | 5 | 5 | 2 |
| Burundi | 4 | 3 | 2 | 2 | 2 |
| Belgium | 0 | 0 | 0 | 0 | 0 |
| Benin | 2 | 2 | 2 | 2 | 2 |
| Burkina Faso | 3 | 2 | 2 | 2 | 2 |
| Bangladesh | 2 | 2 | 1 | 0 | 0 |
| Bulgaria | 2 | 2 | 1 | 1 | 1 |
| Belarus | 2 | 2 | 2 | 1 | 1 |
| Bolivia | 2 | 1 | 1 | 1 | 1 |
| Brazil | 1 | 0 | 0 | 0 | 0 |
| CAN | 0 | 0 | 0 | 0 | 0 |
| Switzerland | 0 | 0 | 0 | 0 | 0 |
| Chile | 0 | 0 | 0 | 0 | 0 |
| China | 2 | 1 | 1 | 1 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Côte d'Ivoire | 3 | 3 | 2 | 2 | 2 |
| Cameroon | 3 | 3 | 2 | 2 | 2 |
| Congo | 3 | 3 | 2 | 2 | 2 |
| Colombia | 0 | 0 | 0 | 0 | 0 |
| Cuba | 0 | 0 | 0 | 0 | 0 |
| Czech Republic | 1 | 1 | 0 | 0 | 0 |
| Germany | 0 | 0 | 0 | 0 | 0 |
| Denmark | 0 | 0 | 0 | 0 | 0 |
| Dominican Republic | 0 | 0 | 0 | 0 | 0 |
| Algeria | 1 | 1 | 0 | 0 | 0 |
| Ecuador | 0 | 0 | 0 | 0 | 0 |
| Egypt | 1 | 1 | 2 | 2 | 1 |
| Spain | 0 | 0 | 0 | 0 | 0 |
| Ethiopia | 3 | 2 | 2 | 2 | 1 |
| Finland | 0 | 0 | 0 | 0 | 0 |
| France | 0 | 0 | 0 | 0 | 0 |
| United Kingdom | 0 | 0 | 0 | 0 | 0 |
| Ghana | 2 | 2 | 2 | 2 | 2 |
| Guinea | 2 | 2 | 2 | 2 | 2 |
| Greece | 0 | 0 | 0 | 0 | 0 |
| Guatemala | 1 | 1 | 1 | 1 | 0 |
| Honduras | 1 | 1 | 1 | 2 | 1 |
| Haiti | 2 | 2 | 6 | 2 | 2 |
| Hungary | 1 | 1 | 1 | 1 | 0 |
| Indonesia | 2 | 2 | 2 | 2 | 2 |
| India | 0 | 0 | 0 | 0 | 0 |
| Iran | 1 | 1 | 0 | 0 | 0 |
| Iraq | 1 | 2 | 1 | 1 | 1 |
| Israel | 0 | 0 | 0 | 0 | 0 |
| Italy | 0 | 0 | 0 | 0 | 0 |
| Jordan | 1 | 1 | 0 | 0 | 0 |
| Japan | 0 | 0 | 0 | 0 | 0 |
| Kazakhstan | 2 | 2 | 2 | 1 | 1 |
| Kenya | 3 | 3 | 2 | 2 | 2 |
| Kyrgyz | 2 | 2 | 2 | 2 | 1 |
| Cambodia | 2 | 2 | 2 | 2 | 2 |
| South Korea | 1 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Laos | 5 | 2 | 2 | 2 | 2 |
| Lebanon | 1 | 1 | 1 | 1 | 1 |
| Libya | 0 | 0 | 0 | 0 | 0 |
| Sri Lanka | 1 | 1 | 1 | 1 | 1 |
| Morocco | 1 | 1 | 1 | 1 | 1 |
| Madagascar | 5 | 2 | 2 | 2 | 2 |
| Mexico | 0 | 0 | 0 | 0 | 0 |
| Mali | 2 | 2 | 2 | 2 | 2 |
| MyanMorocco | 2 | 2 | 2 | 2 | 2 |
| Mozambique | 3 | 3 | 3 | 3 | 2 |
| Malawi | 4 | 3 | 2 | 2 | 2 |
| Malaysia | 2 | 1 | 1 | 1 | 1 |
| Niger | 0 | 0 | 0 | 0 | 0 |
| Nigeria | 2 | 2 | 2 | 2 | 2 |
| Nicaragua | 1 | 1 | 1 | 1 | 1 |
| Netherlands | 0 | 0 | 0 | 0 | 0 |
| Norway | 0 | 0 | 0 | 0 | 0 |
| Nepal | 2 | 2 | 2 | 2 | 2 |
| Pakistan | 2 | 2 | 2 | 2 | 2 |
| Peru | 0 | 0 | 0 | 0 | 0 |
| Philippines | 2 | 2 | 1 | 1 | 1 |
| Papua New Guinea | 2 | 2 | 2 | 2 | 2 |
| Poland | 1 | 1 | 0 | 0 | 0 |
| North Korea | 5 | 2 | 2 | 1 | 1 |
| Portugal | 0 | 0 | 0 | 0 | 0 |
| Paraguay | 0 | 0 | 0 | 0 | 0 |
| Romania | 2 | 1 | 1 | 1 | 1 |
| Russia | 2 | 2 | 2 | 1 | 1 |
| Saudi Arabia | 2 | 1 | 1 | 1 | 1 |
| Sudan | 2 | 2 | 2 | 1.5 | 1.5 |
| Senegal | 2 | 2 | 2 | 2 | 2 |
| Yugoslavia | 2 | 2 | 1 | 1 | 1 |
| Singapore | 0 | 0 | 0 | 0 | 0 |
| Sierra Leone | 3 | 3 | 2 | 2 | 2 |
| El Salvador | 0 | 0 | 0 | 0 | 0 |
| Somalia | 3 | 3 | 3 | 3 | 3 |
| Slovakia | 1 | 1 | 1 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Sweden | 0 | 0 | 0 | 0 | 0 |
| Syria | 2 | 1 | 1 | 5 | 1 |
| Chad | 3 | 3 | 2 | 2 | 2 |
| Togo | 2 | 3 | 2 | 2 | 2 |
| Thailand | 0 | 0 | 0 | 0 | 0 |
| Tajikistan | 2 | 5 | 5 | 5 | 5 |
| Turkmenistan | 2 | 2 | 2 | 2 | 1 |
| Tunisia | 1 | 1 | 1 | 0 | 0 |
| Turkey | 0 | 0 | 0 | 0 | 0 |
| Tanzania | 3 | 2 | 2 | 2 | 2 |
| Uganda | 3 | 3 | 2 | 2 | 2 |
| Ukraine | 2 | 2 | 2 | 1 | 1 |
| United State | 0 | 0 | 0 | 0 | 0 |
| Uzbekistan | 2 | 2 | 2 | 2 | 2 |
| Venezuela | 0 | 0 | 0 | 0 | 0 |
| Vietnam | 1 | 1 | 1 | 1 | 1 |
| Yemen | 2 | 2 | 2 | 2 | 2 |
| South Africa | 2 | 3 | 2 | 2 | 1 |
| Congo, Dem. Rep. | 3 | 3 | 2 | 2 | 2 |
| Zambia | 4 | 3 | 3 | 2 | 2 |
| Zimbabwe | 4 | 4 | 3 | 3 | 2 |

- The detail clusters for children group:

| Country_code | 2000 | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|---|
| Afghanistan | 2 | 1 | 1 | 1 | 1 |
| Angola | 3 | 2 | 1 | 1 | 1 |
| United Arab Emirates | 0 | 0 | 0 | 0 | 0 |
| Argentina | 0 | 0 | 0 | 0 | 0 |
| Australia | 0 | 0 | 0 | 0 | 0 |
| Austria | 0 | 0 | 0 | 0 | 0 |
| Azerbaijan | 1 | 1 | 0 | 0 | 0 |
| Burundi | 2 | 1 | 1 | 1 | 1 |
| Belgium | 0 | 0 | 0 | 0 | 0 |
| Benin | 2 | 1 | 1 | 1 | 1 |
| Burkina Faso | 3 | 2 | 1 | 1 | 1 |
| Bangladesh | 1 | 1 | 1 | 0 | 0 |
| Bulgaria | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Belarus | 0 | 0 | 0 | 0 | 0 |
| Bolivia | 1 | 1 | 0 | 0 | 0 |
| Brazil | 0 | 0 | 0 | 0 | 0 |
| CAN | 0 | 0 | 0 | 0 | 0 |
| Switzerland | 0 | 0 | 0 | 0 | 0 |
| Chile | 0 | 0 | 0 | 0 | 0 |
| China | 0 | 0 | 0 | 0 | 0 |
| Côte d'Ivoire | 2 | 1 | 1 | 1 | 1 |
| Cameroon | 2 | 2 | 1 | 1 | 1 |
| Congo | 1 | 1 | 1 | 1 | 0 |
| Colombia | 0 | 0 | 0 | 0 | 0 |
| Cuba | 0 | 0 | 0 | 0 | 0 |
| Czech Republic | 0 | 0 | 0 | 0 | 0 |
| Germany | 0 | 0 | 0 | 0 | 0 |
| Denmark | 0 | 0 | 0 | 0 | 0 |
| Dominican Republic | 0 | 0 | 0 | 0 | 0 |
| Algeria | 0 | 0 | 0 | 0 | 0 |
| Ecuador | 0 | 0 | 0 | 0 | 0 |
| Egypt | 0 | 0 | 0 | 0 | 0 |
| Spain | 0 | 0 | 0 | 0 | 0 |
| Ethiopia | 2 | 1 | 1 | 1 | 1 |
| Finland | 0 | 0 | 0 | 0 | 0 |
| France | 0 | 0 | 0 | 0 | 0 |
| United Kingdom | 0 | 0 | 0 | 0 | 0 |
| Ghana | 1 | 1 | 1 | 1 | 0 |
| Guinea | 2 | 2 | 1 | 1 | 1 |
| Greece | 0 | 0 | 0 | 0 | 0 |
| Guatemala | 1 | 0 | 0 | 0 | 0 |
| Honduras | 0 | 0 | 0 | 0 | 0 |
| Haiti | 1 | 1 | 5 | 1 | 1 |
| Hungary | 0 | 0 | 0 | 0 | 0 |
| Indonesia | 1 | 0 | 0 | 0 | 0 |
| India | 0 | 0 | 0 | 0 | 0 |
| Iran | 0 | 0 | 0 | 0 | 0 |
| Iraq | 0 | 0 | 0 | 0 | 0 |
| Israel | 0 | 0 | 0 | 0 | 0 |
| Italy | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Jordan | 0 | 0 | 0 | 0 | 0 |
| Japan | 0 | 0 | 0 | 0 | 0 |
| Kazakhstan | 1 | 0 | 0 | 0 | 0 |
| Kenya | 1 | 1 | 1 | 1 | 0 |
| Kyrgyz | 1 | 0 | 0 | 0 | 0 |
| Cambodia | 1 | 1 | 0 | 0 | 0 |
| South Korea | 0 | 0 | 0 | 0 | 0 |
| Laos | 1 | 1 | 1 | 1 | 0 |
| Lebanon | 0 | 0 | 0 | 0 | 0 |
| Libya | 0 | 0 | 0 | 0 | 0 |
| Sri Lanka | 0 | 0 | 0 | 0 | 0 |
| Morocco | 1 | 0 | 0 | 0 | 0 |
| Madagascar | 1 | 1 | 1 | 1 | 0 |
| Mexico | 0 | 0 | 0 | 0 | 0 |
| Mali | 3 | 2 | 2 | 1 | 1 |
| Myanmar | 1 | 1 | 1 | 1 | 0 |
| Mozambique | 2 | 2 | 1 | 1 | 1 |
| Malawi | 2 | 1 | 1 | 1 | 0 |
| Malaysia | 0 | 0 | 0 | 0 | 0 |
| Niger | 0 | 0 | 0 | 0 | 0 |
| Nigeria | 3 | 2 | 2 | 1 | 1 |
| Nicaragua | 0 | 0 | 0 | 0 | 0 |
| Netherlands | 0 | 0 | 0 | 0 | 0 |
| Norway | 0 | 0 | 0 | 0 | 0 |
| Nepal | 1 | 1 | 0 | 0 | 0 |
| Pakistan | 1 | 1 | 1 | 1 | 1 |
| Peru | 0 | 0 | 0 | 0 | 0 |
| Philippines | 0 | 0 | 0 | 0 | 0 |
| Papua New Guinea | 1 | 1 | 1 | 1 | 0 |
| Poland | 0 | 0 | 0 | 0 | 0 |
| North Korea | 1 | 0 | 0 | 0 | 0 |
| Portugal | 0 | 0 | 0 | 0 | 0 |
| Paraguay | 0 | 0 | 0 | 0 | 0 |
| Romania | 0 | 0 | 0 | 0 | 0 |
| Russia | 0 | 0 | 0 | 0 | 0 |
| Saudi Arabia | 0 | 0 | 0 | 0 | 0 |
| Sudan | 1.5 | 1.5 | 1 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Senegal | 2 | 1 | 1 | 1 | 0 |
| Yugoslavia | 0 | 0 | 0 | 0 | 0 |
| Singapore | 0 | 0 | 0 | 0 | 0 |
| Sierra Leone | 4 | 2 | 2 | 2 | 1 |
| El Salvador | 0 | 0 | 0 | 0 | 0 |
| Somalia | 2 | 2 | 2 | 2 | 1 |
| Slovakia | 0 | 0 | 0 | 0 | 0 |
| Sweden | 0 | 0 | 0 | 0 | 0 |
| Syria | 0 | 0 | 0 | 0 | 0 |
| Chad | 3 | 2 | 2 | 2 | 1 |
| Togo | 1 | 1 | 1 | 1 | 1 |
| Thailand | 0 | 0 | 0 | 0 | 0 |
| Tajikistan | 1 | 1 | 0 | 0 | 0 |
| Turkmenistan | 1 | 1 | 1 | 0 | 0 |
| Tunisia | 0 | 0 | 0 | 0 | 0 |
| Turkey | 0 | 0 | 0 | 0 | 0 |
| Tanzania | 2 | 1 | 1 | 1 | 1 |
| Uganda | 2 | 1 | 1 | 1 | 0 |
| Ukraine | 0 | 0 | 0 | 0 | 0 |
| United State | 0 | 0 | 0 | 0 | 0 |
| Uzbekistan | 1 | 1 | 0 | 0 | 0 |
| Venezuela | 0 | 0 | 0 | 0 | 0 |
| Vietnam | 0 | 0 | 0 | 0 | 0 |
| Yemen | 1 | 1 | 1 | 1 | 1 |
| South Africa | 1 | 1 | 0 | 0 | 0 |
| Congo, Dem. Rep. | 2 | 2 | 1 | 1 | 1 |
| Zambia | 2 | 1 | 1 | 1 | 1 |
| Zimbabwe | 1 | 1 | 1 | 1 | 1 |

- The detail clusters for Adult group A:

| Country_code | 2000 | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|---|
| Afghanistan | 1 | 1 | 1 | 1 | 1 |
| Angola | 1 | 1 | 1 | 1 | 1 |
| United Arab Emirates | 0 | 0 | 0 | 0 | 0 |
| Argentina | 0 | 0 | 0 | 0 | 0 |
| Australia | 0 | 0 | 0 | 0 | 0 |
| Austria | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Azerbaijan | 0 | 0 | 0 | 0 | 0 |
| Burundi | 5 | 2 | 1 | 1 | 1 |
| Belgium | 0 | 0 | 0 | 0 | 0 |
| Benin | 1 | 1 | 1 | 1 | 1 |
| Burkina Faso | 2 | 1 | 1 | 1 | 1 |
| Bangladesh | 0 | 0 | 0 | 0 | 0 |
| Bulgaria | 0 | 0 | 0 | 0 | 0 |
| Belarus | 1 | 1 | 1 | 0 | 0 |
| Bolivia | 1 | 0 | 0 | 0 | 0 |
| Brazil | 0 | 0 | 0 | 0 | 0 |
| CAN | 0 | 0 | 0 | 0 | 0 |
| Switzerland | 0 | 0 | 0 | 0 | 0 |
| Chile | 0 | 0 | 0 | 0 | 0 |
| China | 0 | 0 | 0 | 0 | 0 |
| Côte d'Ivoire | 2 | 2 | 2 | 1 | 1 |
| Cameroon | 2 | 2 | 2 | 1 | 1 |
| Congo | 2 | 2 | 1 | 1 | 1 |
| Colombia | 0 | 0 | 0 | 0 | 0 |
| Cuba | 0 | 0 | 0 | 0 | 0 |
| Czech Republic | 0 | 0 | 0 | 0 | 0 |
| Germany | 0 | 0 | 0 | 0 | 0 |
| Denmark | 0 | 0 | 0 | 0 | 0 |
| Dominican Republic | 0 | 0 | 0 | 0 | 0 |
| Algeria | 0 | 0 | 0 | 0 | 0 |
| Ecuador | 0 | 0 | 0 | 0 | 0 |
| Egypt | 0 | 0 | 0 | 0 | 0 |
| Spain | 0 | 0 | 0 | 0 | 0 |
| Ethiopia | 2 | 1 | 1 | 1 | 0 |
| Finland | 0 | 0 | 0 | 0 | 0 |
| France | 0 | 0 | 0 | 0 | 0 |
| United Kingdom | 0 | 0 | 0 | 0 | 0 |
| Ghana | 1 | 1 | 1 | 1 | 1 |
| Guinea | 1 | 1 | 1 | 1 | 1 |
| Greece | 0 | 0 | 0 | 0 | 0 |
| Guatemala | 1 | 1 | 1 | 1 | 0 |
| Honduras | 1 | 0 | 0 | 1 | 0 |
| Haiti | 2 | 1 | 4 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Hungary | 1 | 1 | 0 | 0 | 0 |
| Indonesia | 1 | 0 | 0 | 0 | 0 |
| India | 0 | 0 | 0 | 0 | 0 |
| Iran | 0 | 0 | 0 | 0 | 0 |
| Iraq | 1 | 1 | 0 | 0 | 0 |
| Israel | 0 | 0 | 0 | 0 | 0 |
| Italy | 0 | 0 | 0 | 0 | 0 |
| Jordan | 0 | 0 | 0 | 0 | 0 |
| Japan | 0 | 0 | 0 | 0 | 0 |
| Kazakhstan | 1 | 1 | 1 | 0 | 0 |
| Kenya | 2 | 2 | 2 | 1 | 1 |
| Kyrgyz | 1 | 1 | 1 | 0 | 0 |
| Cambodia | 1 | 1 | 1 | 1 | 0 |
| South Korea | 0 | 0 | 0 | 0 | 0 |
| Laos | 1 | 1 | 1 | 1 | 0 |
| Lebanon | 0 | 0 | 0 | 0 | 0 |
| Libya | 0 | 0 | 0 | 0 | 0 |
| Sri Lanka | 0 | 0 | 0 | 0 | 0 |
| Morocco | 0 | 0 | 0 | 0 | 0 |
| Madagascar | 1 | 1 | 1 | 1 | 1 |
| Mexico | 0 | 0 | 0 | 0 | 0 |
| Mali | 1 | 1 | 1 | 1 | 1 |
| MyanMorocco | 1 | 1 | 1 | 1 | 0 |
| Mozambique | 2 | 2 | 2 | 2 | 2 |
| Malawi | 3 | 3 | 2 | 1 | 1 |
| Malaysia | 0 | 0 | 0 | 0 | 0 |
| Niger | 0 | 0 | 0 | 0 | 0 |
| Nigeria | 1 | 1 | 1 | 1 | 1 |
| Nicaragua | 0 | 0 | 0 | 0 | 0 |
| Netherlands | 0 | 0 | 0 | 0 | 0 |
| Norway | 0 | 0 | 0 | 0 | 0 |
| Nepal | 1 | 0 | 0 | 0 | 0 |
| Pakistan | 1 | 1 | 1 | 1 | 1 |
| Peru | 0 | 0 | 0 | 0 | 0 |
| Philippines | 0 | 0 | 0 | 0 | 0 |
| Papua New Guinea | 1 | 1 | 1 | 1 | 1 |
| Poland | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| North Korea | 1 | 0 | 0 | 0 | 0 |
| Portugal | 0 | 0 | 0 | 0 | 0 |
| Paraguay | 0 | 0 | 0 | 0 | 0 |
| Romania | 0 | 0 | 0 | 0 | 0 |
| Russia | 1 | 1 | 1 | 1 | 1 |
| Saudi Arabia | 0 | 0 | 0 | 0 | 0 |
| Sudan | 1 | 1 | 0.5 | 0.5 | 0.5 |
| Senegal | 1 | 1 | 1 | 1 | 0 |
| Yugoslavia | 0 | 0 | 0 | 0 | 0 |
| Singapore | 0 | 0 | 0 | 0 | 0 |
| Sierra Leone | 1 | 1 | 1 | 1 | 1 |
| El Salvador | 0 | 0 | 0 | 0 | 0 |
| Somalia | 2 | 2 | 2 | 1 | 1 |
| Slovakia | 0 | 0 | 0 | 0 | 0 |
| Sweden | 0 | 0 | 0 | 0 | 0 |
| Syria | 0 | 0 | 0 | 7 | 0 |
| Chad | 1 | 1 | 1 | 1 | 1 |
| Togo | 2 | 2 | 1 | 1 | 1 |
| Thailand | 7 | 0 | 0 | 0 | 0 |
| Tajikistan | 0 | 0 | 0 | 0 | 0 |
| Turkmenistan | 1 | 1 | 1 | 1 | 0 |
| Tunisia | 0 | 0 | 0 | 0 | 0 |
| Turkey | 0 | 0 | 0 | 0 | 0 |
| Tanzania | 2 | 2 | 1 | 1 | 1 |
| Uganda | 3 | 2 | 2 | 1 | 1 |
| Ukraine | 1 | 1 | 1 | 1 | 0 |
| United State | 0 | 0 | 0 | 0 | 0 |
| Uzbekistan | 1 | 1 | 0 | 0 | 0 |
| Venezuela | 0 | 0 | 0 | 0 | 0 |
| Vietnam | 0 | 0 | 0 | 0 | 0 |
| Yemen | 1 | 1 | 0 | 0 | 1 |
| South Africa | 2 | 3 | 2 | 2 | 1 |
| Congo, Dem. Rep. | 2 | 1 | 1 | 1 | 1 |
| Zambia | 6 | 3 | 2 | 1 | 1 |
| Zimbabwe | 8 | 8 | 3 | 2 | 2 |

- The detail clusters for Adult group B:

| Country_code | 2000 | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|---|
| Afghanistan | 1 | 1 | 1 | 1 | 1 |
| Angola | 1 | 1 | 1 | 1 | 1 |
| United Arab Emirates | 0 | 0 | 0 | 0 | 0 |
| Argentina | 0 | 0 | 0 | 0 | 0 |
| Australia | 0 | 0 | 0 | 0 | 0 |
| Austria | 0 | 0 | 0 | 0 | 0 |
| Azerbaijan | 0 | 0 | 0 | 0 | 0 |
| Burundi | 5 | 2 | 1 | 1 | 1 |
| Belgium | 0 | 0 | 0 | 0 | 0 |
| Benin | 1 | 1 | 1 | 1 | 1 |
| Burkina Faso | 2 | 1 | 1 | 1 | 1 |
| Bangladesh | 0 | 0 | 0 | 0 | 0 |
| Bulgaria | 0 | 0 | 0 | 0 | 0 |
| Belarus | 1 | 1 | 1 | 0 | 0 |
| Bolivia | 1 | 0 | 0 | 0 | 0 |
| Brazil | 0 | 0 | 0 | 0 | 0 |
| CAN | 0 | 0 | 0 | 0 | 0 |
| Switzerland | 0 | 0 | 0 | 0 | 0 |
| Chile | 0 | 0 | 0 | 0 | 0 |
| China | 0 | 0 | 0 | 0 | 0 |
| Côte d'Ivoire | 2 | 2 | 2 | 1 | 1 |
| Cameroon | 2 | 2 | 2 | 1 | 1 |
| Congo | 2 | 2 | 1 | 1 | 1 |
| Colombia | 0 | 0 | 0 | 0 | 0 |
| Cuba | 0 | 0 | 0 | 0 | 0 |
| Czech Republic | 0 | 0 | 0 | 0 | 0 |
| Germany | 0 | 0 | 0 | 0 | 0 |
| Denmark | 0 | 0 | 0 | 0 | 0 |
| Dominican Republic | 0 | 0 | 0 | 0 | 0 |
| Algeria | 0 | 0 | 0 | 0 | 0 |
| Ecuador | 0 | 0 | 0 | 0 | 0 |
| Egypt | 0 | 0 | 0 | 0 | 0 |
| Spain | 0 | 0 | 0 | 0 | 0 |
| Ethiopia | 2 | 1 | 1 | 1 | 0 |
| Finland | 0 | 0 | 0 | 0 | 0 |
| France | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| United Kingdom | 0 | 0 | 0 | 0 | 0 |
| Ghana | 1 | 1 | 1 | 1 | 1 |
| Guinea | 1 | 1 | 1 | 1 | 1 |
| Greece | 0 | 0 | 0 | 0 | 0 |
| Guatemala | 1 | 1 | 1 | 1 | 0 |
| Honduras | 1 | 0 | 0 | 1 | 0 |
| Haiti | 2 | 1 | 4 | 1 | 1 |
| Hungary | 1 | 1 | 0 | 0 | 0 |
| Indonesia | 1 | 0 | 0 | 0 | 0 |
| India | 0 | 0 | 0 | 0 | 0 |
| Iran | 0 | 0 | 0 | 0 | 0 |
| Iraq | 1 | 1 | 0 | 0 | 0 |
| Israel | 0 | 0 | 0 | 0 | 0 |
| Italy | 0 | 0 | 0 | 0 | 0 |
| Jordan | 0 | 0 | 0 | 0 | 0 |
| Japan | 0 | 0 | 0 | 0 | 0 |
| Kazakhstan | 1 | 1 | 1 | 0 | 0 |
| Kenya | 2 | 2 | 2 | 1 | 1 |
| Kyrgyz | 1 | 1 | 1 | 0 | 0 |
| Cambodia | 1 | 1 | 1 | 1 | 0 |
| South Korea | 0 | 0 | 0 | 0 | 0 |
| Laos | 1 | 1 | 1 | 1 | 0 |
| Lebanon | 0 | 0 | 0 | 0 | 0 |
| Libya | 0 | 0 | 0 | 0 | 0 |
| Sri Lanka | 0 | 0 | 0 | 0 | 0 |
| Morocco | 0 | 0 | 0 | 0 | 0 |
| Madagascar | 1 | 1 | 1 | 1 | 1 |
| Mexico | 0 | 0 | 0 | 0 | 0 |
| Mali | 1 | 1 | 1 | 1 | 1 |
| MyanMorocco | 1 | 1 | 1 | 1 | 0 |
| Mozambique | 2 | 2 | 2 | 2 | 2 |
| Malawi | 3 | 3 | 2 | 1 | 1 |
| Malaysia | 0 | 0 | 0 | 0 | 0 |
| Niger | 0 | 0 | 0 | 0 | 0 |
| Nigeria | 1 | 1 | 1 | 1 | 1 |
| Nicaragua | 0 | 0 | 0 | 0 | 0 |
| Netherlands | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Norway | 0 | 0 | 0 | 0 | 0 |
| Nepal | 1 | 0 | 0 | 0 | 0 |
| Pakistan | 1 | 1 | 1 | 1 | 1 |
| Peru | 0 | 0 | 0 | 0 | 0 |
| Philippines | 0 | 0 | 0 | 0 | 0 |
| Papua New Guinea | 1 | 1 | 1 | 1 | 1 |
| Poland | 0 | 0 | 0 | 0 | 0 |
| North Korea | 1 | 0 | 0 | 0 | 0 |
| Portugal | 0 | 0 | 0 | 0 | 0 |
| Paraguay | 0 | 0 | 0 | 0 | 0 |
| Romania | 0 | 0 | 0 | 0 | 0 |
| Russia | 1 | 1 | 1 | 1 | 1 |
| Saudi Arabia | 0 | 0 | 0 | 0 | 0 |
| Sudan | 1 | 1 | 0.5 | 0.5 | 0.5 |
| Senegal | 1 | 1 | 1 | 1 | 0 |
| Yugoslavia | 0 | 0 | 0 | 0 | 0 |
| Singapore | 0 | 0 | 0 | 0 | 0 |
| Sierra Leone | 1 | 1 | 1 | 1 | 1 |
| El Salvador | 0 | 0 | 0 | 0 | 0 |
| Somalia | 2 | 2 | 2 | 1 | 1 |
| Slovakia | 0 | 0 | 0 | 0 | 0 |
| Sweden | 0 | 0 | 0 | 0 | 0 |
| Syria | 0 | 0 | 0 | 7 | 0 |
| Chad | 1 | 1 | 1 | 1 | 1 |
| Togo | 2 | 2 | 1 | 1 | 1 |
| Thailand | 7 | 0 | 0 | 0 | 0 |
| Tajikistan | 0 | 0 | 0 | 0 | 0 |
| Turkmenistan | 1 | 1 | 1 | 1 | 0 |
| Tunisia | 0 | 0 | 0 | 0 | 0 |
| Turkey | 0 | 0 | 0 | 0 | 0 |
| Tanzania | 2 | 2 | 1 | 1 | 1 |
| Uganda | 3 | 2 | 2 | 1 | 1 |
| Ukraine | 1 | 1 | 1 | 1 | 0 |
| United State | 0 | 0 | 0 | 0 | 0 |
| Uzbekistan | 1 | 1 | 0 | 0 | 0 |
| Venezuela | 0 | 0 | 0 | 0 | 0 |
| Vietnam | 0 | 0 | 0 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Yemen | 1 | 1 | 0 | 0 | 1 |
| South Africa | 2 | 3 | 2 | 2 | 1 |
| Congo, Dem. Rep. | 2 | 1 | 1 | 1 | 1 |
| Zambia | 6 | 3 | 2 | 1 | 1 |
| Zimbabwe | 8 | 8 | 3 | 2 | 2 |

- The detail clusters for Senior:

| Country_code | 2000 | 2005 | 2010 | 2015 | 2019 |
|---|---|---|---|---|---|
| Afghanistan | 4 | 3 | 3 | 3 | 3 |
| Angola | 4 | 3 | 3 | 3 | 3 |
| United Arab Emirates | 2 | 2 | 2 | 2 | 2 |
| Argentina | 1 | 1 | 1 | 1 | 1 |
| Australia | 0 | 0 | 0 | 0 | 0 |
| Austria | 1 | 0 | 0 | 0 | 0 |
| Azerbaijan | 3 | 3 | 3 | 3 | 3 |
| Burundi | 4 | 3 | 3 | 3 | 3 |
| Belgium | 1 | 0 | 0 | 0 | 0 |
| Benin | 3 | 3 | 3 | 3 | 3 |
| Burkina Faso | 3 | 3 | 3 | 3 | 3 |
| Bangladesh | 2 | 2 | 2 | 1 | 1 |
| Bulgaria | 3 | 2 | 2 | 2 | 1 |
| Belarus | 3 | 3 | 2 | 2 | 2 |
| Bolivia | 3 | 2 | 2 | 2 | 2 |
| Brazil | 2 | 1 | 1 | 1 | 1 |
| CAN | 0 | 0 | 0 | 0 | 0 |
| Switzerland | 0 | 0 | 0 | 0 | 0 |
| Chile | 1 | 1 | 0 | 0 | 0 |
| China | 2 | 2 | 2 | 1 | 1 |
| Côte d'Ivoire | 3 | 3 | 3 | 3 | 3 |
| Cameroon | 3 | 3 | 3 | 3 | 3 |
| Congo | 4 | 3 | 3 | 3 | 3 |
| Colombia | 1 | 1 | 0 | 0 | 0 |
| Cuba | 1 | 1 | 1 | 1 | 1 |
| Czech Republic | 2 | 2 | 1 | 1 | 0 |
| Germany | 1 | 0 | 0 | 0 | 0 |
| Denmark | 1 | 1 | 1 | 0 | 0 |
| Dominican Republic | 0 | 1 | 1 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Algeria | 1 | 1 | 1 | 0 | 0 |
| Ecuador | 1 | 1 | 1 | 0 | 0 |
| Egypt | 2 | 2 | 2 | 3 | 2 |
| Spain | 0 | 0 | 0 | 0 | 0 |
| Ethiopia | 4 | 3 | 3 | 3 | 2 |
| Finland | 1 | 0 | 0 | 0 | 0 |
| France | 0 | 0 | 0 | 0 | 0 |
| United Kingdom | 1 | 1 | 0 | 0 | 0 |
| Ghana | 3 | 3 | 3 | 3 | 3 |
| Guinea | 3 | 3 | 3 | 3 | 3 |
| Greece | 1 | 0 | 0 | 0 | 0 |
| Guatemala | 2 | 2 | 1 | 1 | 1 |
| Honduras | 2 | 2 | 2 | 3 | 2 |
| Haiti | 3 | 3 | 4 | 3 | 3 |
| Hungary | 2 | 2 | 2 | 1 | 1 |
| Indonesia | 3 | 3 | 3 | 3 | 3 |
| India | 1 | 0 | 0 | 0 | 0 |
| Iran | 2 | 2 | 1 | 1 | 1 |
| Iraq | 2 | 3 | 2 | 2 | 2 |
| Israel | 1 | 0 | 0 | 0 | 0 |
| Italy | 0 | 0 | 0 | 0 | 0 |
| Jordan | 2 | 2 | 1 | 1 | 1 |
| Japan | 0 | 0 | 0 | 0 | 0 |
| Kazakhstan | 3 | 3 | 3 | 2 | 2 |
| Kenya | 3 | 3 | 3 | 3 | 3 |
| Kyrgyz | 3 | 3 | 3 | 2 | 2 |
| Cambodia | 3 | 3 | 3 | 3 | 3 |
| South Korea | 1 | 1 | 0 | 0 | 0 |
| Laos | 3 | 3 | 3 | 3 | 3 |
| Lebanon | 2 | 2 | 2 | 2 | 1 |
| Libya | 2 | 1 | 1 | 1 | 1 |
| Sri Lanka | 2 | 2 | 2 | 2 | 2 |
| Morocco | 2 | 2 | 2 | 2 | 2 |
| Madagascar | 3 | 3 | 3 | 3 | 3 |
| Mexico | 1 | 1 | 1 | 1 | 1 |
| Mali | 3 | 3 | 3 | 3 | 3 |
| MyanMorocco | 3 | 3 | 3 | 3 | 3 |

| | | | | | |
|---|---|---|---|---|---|
| Mozambique | 3 | 3 | 3 | 3 | 3 |
| Malawi | 4 | 3 | 3 | 3 | 3 |
| Malaysia | 3 | 2 | 2 | 2 | 2 |
| Niger | 0 | 0 | 0 | 0 | 0 |
| Nigeria | 3 | 3 | 3 | 3 | 2 |
| Nicaragua | 2 | 2 | 2 | 2 | 2 |
| Netherlands | 1 | 1 | 0 | 0 | 0 |
| Norway | 1 | 0 | 0 | 0 | 0 |
| Nepal | 3 | 3 | 3 | 3 | 3 |
| Pakistan | 3 | 3 | 3 | 3 | 3 |
| Peru | 0 | 0 | 0 | 0 | 0 |
| Philippines | 2 | 2 | 2 | 2 | 2 |
| Papua New Guinea | 3 | 3 | 3 | 3 | 3 |
| Poland | 2 | 1 | 1 | 1 | 0 |
| North Korea | 3 | 2 | 2 | 2 | 2 |
| Portugal | 1 | 1 | 0 | 0 | 0 |
| Paraguay | 1 | 1 | 1 | 1 | 1 |
| Romania | 2 | 2 | 2 | 2 | 1 |
| Russia | 3 | 3 | 2 | 2 | 1 |
| Saudi Arabia | 3 | 2 | 2 | 2 | 2 |
| Sudan | 3 | 3 | 2.5 | 2.5 | 2.5 |
| Senegal | 3 | 3 | 3 | 3 | 2 |
| Yugoslavia | 3 | 3 | 2 | 2 | 2 |
| Singapore | 1 | 0 | 0 | 0 | 0 |
| Sierra Leone | 3 | 3 | 3 | 3 | 3 |
| El Salvador | 1 | 0 | 0 | 0 | 0 |
| Somalia | 4 | 4 | 4 | 3 | 3 |
| Slovakia | 2 | 2 | 2 | 1 | 0 |
| Sweden | 0 | 0 | 0 | 0 | 0 |
| Syria | 3 | 2 | 2 | 3 | 2 |
| Chad | 3 | 3 | 3 | 3 | 3 |
| Togo | 3 | 3 | 3 | 3 | 3 |
| Thailand | 1 | 1 | 0 | 0 | 0 |
| Tajikistan | 3 | 3 | 3 | 4 | 3 |
| Turkmenistan | 3 | 3 | 3 | 3 | 2 |
| Tunisia | 2 | 1 | 1 | 1 | 1 |
| Turkey | 1 | 1 | 1 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Tanzania | 3 | 3 | 3 | 3 | 2 |
| Uganda | 3 | 3 | 3 | 3 | 3 |
| Ukraine | 3 | 3 | 3 | 2 | 2 |
| United State | 1 | 0 | 0 | 0 | 0 |
| Uzbekistan | 3 | 3 | 3 | 3 | 3 |
| Venezuela | 1 | 1 | 1 | 1 | 0 |
| Vietnam | 2 | 2 | 2 | 2 | 2 |
| Yemen | 3 | 3 | 3 | 3 | 3 |
| South Africa | 3 | 3 | 2 | 2 | 1 |
| Congo, Dem. Rep. | 3 | 3 | 3 | 3 | 3 |
| Zambia | 4 | 3 | 3 | 3 | 3 |
| Zimbabwe | 3 | 3 | 3 | 3 | 3 |