UNIVERSITAT POLITÈCNICA DE VALÈNCIA

School of Informatics

Study of the social, enviromental and economic policies of each business in their outcome and benefits

End of Degree Project

Bachelor's Degree in Data Science

AUTHOR: Salelles Roman, Clara

Tutor: Morillas Gómez, Samuel

External cotutor: KOKSVIK, GITTE

ACADEMIC YEAR: 2021/2022

Study of the social, environmental and economic policies of each business in their outcome and benefits

Bachelor's Thesis

Data Science

Author: Clara Salelles Román

Tutors:

Samuel Morillas Gómez

Gitte Hanssen Kosvik

Resumen

El proyecto de investigación tiene como objetivo estudiar las políticas sostenibles y éticas de las empresas para generar un ranking global mundial teniendo en cuenta no solo su impacto social y medioambiental sino el entrono en el que se desarrollan. Para ello se analizan tanto las políticas de las diferentes compañías como las de los gobiernos que rigen en los países donde se localizan. Se estudian 15600 empresas y 201 países. Tras analizar los datos éticos de cada país, se determina un indicador del score para cada país y utilizando este se calcula la nueva ponderación global mundial ética de las empresas. Por último, los resultados son visualizados en mapas de histogramas clasificando las empresas por cuartiles para así reflejar las similitudes entre países del mismo continente y facilitando al inversor la lectura del ranking creado.

Palabras clave: responsabilidad social, ética empresarial, estadística,

Resumen

4

El projecte d'investigació té com a objectiu estudiar la polítiques sostenibles i ètiques de les empreses per a generar un rànquing global tenint en compte no sols el seu impacte social i mediambiental sinó l'entronitze en el que es desenvolupen. Per a això s'analitza tant les polítiques de les diferents companyies com els governs que regeixen els països on es localitzen. S'estudien 15600 empreses i 201 països, mitjançant un clúster K-*means i un *PCA, respectivament. Es calcula la nova ponderació ètica de les empreses en funció de la bondat del seu país mitjançant *Data *Mapping. Finalment, els resultats són visualitzats en mapes aconseguint així estudiar les similituds entre països del mateix continent i facilitant a l'inversor la lectura del rànquing creat.

:

Abstract

5

The research project aims to study the sustainable and ethical policies of companies to generate a global ranking considering not only their social and environmental impact but also the framework in which they operate. To this end, both the policies of the different companies and the governments that regulate the countries where they are located are analyzed. A total of 15,600 companies and 201 countries are studied, using a K-means cluster and a PCA, respectively. The new ethical rating of the companies is calculated according to the quality of their country by means of Data Mapping. Finally, the results are visualized on maps, making it possible to study the similarities between countries on the same continent and facilitating the investor's reading of the ranking created.

Index

Index

Figure Index

1. **Introduction**

The importance of ethics in business is not new. The topic is covered with the frequency of every manager's management classes, and HR specialists' monologues are filled with emphasis on communication, employee well-being, and best practices. Although these areas have long been subject to management's interventions, the appearance of these public discussions confirms the need for additional involvement from all stakeholders.

Ethical behaviour is not simply a mere compliance requirement. It is a complex concept that means that your actions have a positive or negative impact on the people that work in the company and the society the company works for. It also means that you can fulfill both professional and personal commitments. A professional relationship with people you respect, whose contribution is needed for the organisation to succeed, means you need to treat them fairly and behave ethically. However,

1.1 Goals

The main objective of this study is to generate a global ranking by ordering companies according to their ethical commitment both internally and externally.
In order to achieve this objective, the following specific objectives are proposed:

- To study how sustainable and involved in social impact the different governments of the world's countries are and how they behave in this respect. To analyze the evolution over time.
- To investigate how companies are grouped according to their environmental, social and governmental impact scores. Create a new weighting according to the country in which the business operates.

1.2 Structure

The study is divided into four parts. First, the procedure followed to obtain the appropriate data that best fit our objective and a brief exploratory analysis of the data are explained. Next, two

studies will be carried out, one for each selected dataset. A PCA for the country governments and a k-means cluster for the Companies performance. We show loading scores in bar plots and world maps to an easy comprehension.

A new global world weighting will be calculated and the ranking of the companies will be obtained according to it. For further analysis, data will be group by Regions and Subregions. World maps will be created to explain the different positions achieved by the companies in the different countries.

2. Theoretical framework

For the development of our work, a measure of sustainability and business ethics is needed. Over the years, attempts have been made to score seemingly countless variables.

In 1990, the first Environmental, Social and Governance Index called The MSCI KLD 400 was published. It was initially designed so that investors interested in social responsibility could study not only the economic variables of the companies in which they chose to participate, but also the social and environmental factors.

The measurement has evolved over the following three decades with the participation of a multitude of organizations that collect and analyze new data related to ESG topics. Thanks to this, it is possible to weight the sustainable policies and behaviors of different companies. The following figure shows which companies and what role they play in the development of the scores and rating.
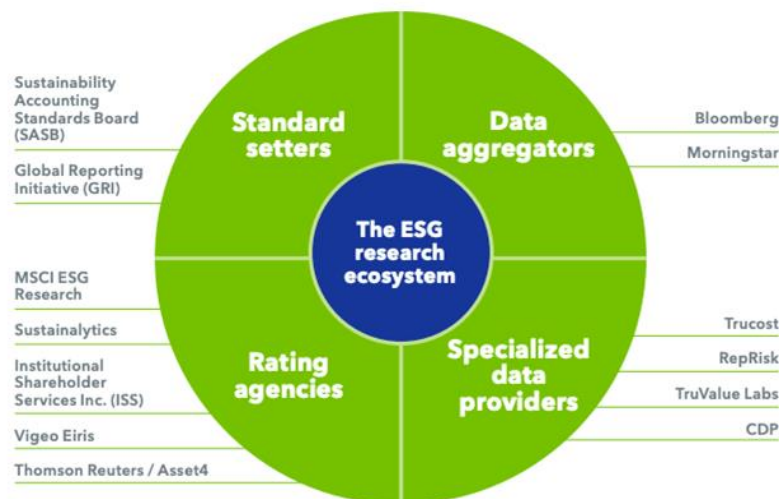
Figure 1 Table of the ESG research ecosystem

The analysis and calculation of the different variables has been refined, but the purpose and the desired outcome is the same. "Increasingly investors seek to combine certain non-financial objectives such as ethical, religious, political, cultural, societal values and preferences without hampering financial objectives" (Georg Inderst and Fiona Stewart, 2018).

However, there are still discrepancies regarding a definitive list of ESG Scores. One of the main criticisms is that the data processed to obtain the score varies depending on the environment in which the company is located and develops. But in the ESG Scores this is not yet contemplated.

"It looks impossible to agree on. Markets, technologies, policies, policies, values and social preferences change all the time, and vary from region to region, country to country and even within countries." (Georg Inderst and Fiona Stewart, 2018) That is why this study focuses not only on the ethics of companies but also of the governments of the countries where they operate. "Most scholars, policymakers, aid donors, and aid recipients recognize that good governance is a fundamental ingredient of sustained economic development" (Daniel Kaufmann & Aart Kraay, 2008). There is a lot of information to take into account regarding the assessment of a government, but for our sustainable objective we highlight ethical features. But, can the sustainable behavior of a country be measured? This is another factor that has been the main focus of much research. "Key pillars of the 'art of governance' emphasize justice, ethics, and anti-autocratic tendencies." (Daniel Kaufmann & Aart Kraay, 2008) .

One of the most recognized and widely used indicators today is The Worldwide Governance Indicators. Its definition according to its author Daniel Kaufman "are a long-standing research project to develop cross-country indicators of governance. The WGI consists of six composite indicators of broad dimensions of governance covering over 200 countries since 1996". It stands out for its quantitative assessment of qualitative characteristics such as a country's peace and security, citizens' rights and participation, as well as the quality of policy implementation. This project will analyze this database with this information, which will be described in detail below.

3. Data collection procedure and exploratory analysis

3.1 Data acquisition

The study will be based on the analysis of two datasets. One with the World Governance Indicators to be able to analyze the policies of the different countries and the second one containing the ESG scores of the companies. The process for obtaining the final datasets and their description is explained below.

One of the most important concerns when studying the behavior of governments in different countries is to quantify both the effectiveness and the ethics of the policies that are established. We use the Data World Bank tool, a platform that stores different time series data on a multitude of topics and provides the option to create your own queries.

There is a section on global indicators, and it has been decided to make the first consultation onEnvironment, Social and Governance data. The ESG database provides a set of indicators for sustainable investment analysis, based on a preliminary framework of seventeen key sustainability criteria. All countries and sixty-seven variables available are selected. We chose all of them for study and analysis. In terms of time series, we will work with data from 2005 to 2019.

We obtain an Excel file with which we will work in the R language. An exploratory analysis of the file is carried out and when preprocessing the data, we discover an abundance of NAs. All the rows present more than 30% of the variables without information. As for the columns, there are 39 variables with more than 30% missing data and 35 variables with more than 50% missing data.

Figure 2 Missing data in rows

By eliminating those variables with more than 50% we lose the score estimates we needed, and we are left with variables that do not contribute the desired value to our final objective. In addition, I consider that to make an imputation of Na's when the percentage is so high would be far from reality. That is why it is decided to keep looking for more databases.

We ran 4 more queries in which when we performed data cleaning we found that important variables had more than half of the data missing. Until we found Worldwide Governance Indicators. This will be explained in section 3.2.

Regarding corporate ethics, we wanted to look for a database in which a quantification of social and environmental impact could be found. There are many organizations and companies focused on investors that provide this information, but as a student we received a negative response..

At NasDaq, a data collection website. I applied for a license to work with Upright Net Impact Data. It provided us with a score of companies over time updated quarterly since 1991. It measured various aspects such as the company's impact on society and the environment. It was interesting because it broke down each score into different variables that made it possible to study the behavior of the different companies.

company_name	legal_entity_identifier	isin	net_impact_ratio
Infineon Technologies	TSI2PJM6EPETEQ4X1U25	DE0006231004	3482
Skyworks Solutions	549300WZ3ORQ2BVKBD96	US83088M1027	6397
Incyte	549300Z4WN6JVZ3T4680	US45337C1027	6836
Swatch	529900NTES2LEON1JZ55	CH0012255151	424
Seattle Genetics	549300UJFWSN15PSBY77	US8125781026	6565
Merck	529900OAREIS0MOPTW25	DE0006599905	2877
Alexion Pharmaceuticals	M1YXUUZR0EIMU8T0EM75	US0153511094	6834
Fresenius Medical Care	549300CP8NY40UP89Q40	DE0005785802	6243
Partners	549300I3IVVWZWDHI788	CH0024608827	1418
Bayer	549300J4U55H3WP1XT59	DE000BAY0017	4064

Figure 3 Capture Sample Dataset Upright Net Impact

score_s	score_k	score_h	score_e	score_sp	score_kp	score_hp	score_ep	score_sn	score_kn	score_hn	score_en
22293	2758	32	-10782	23323	13953	2771	1021	103	11195	2739	11803
21346	25038	2315	-4512	2219	36572	7722	2593	844	11534	5407	7105
13554	9938	60795	-4529	18995	30376	672	105	5441	20438	6404	4634
20402	-4073	3023	-7427	20707	2799	4276	34	305	6872	1253	7767
11538	-8615	71265	-4562	18837	12044	75	18	7299	20659	3735	4742
9695	-9711	28602	-11663	14092	5664	36169	2897	4396	15375	7568	14561
15781	-5437	66782	-4267	22881	13343	70224	164	71	1878	3442	4431

Figure 4 Capture Sample Dataset Upright Net Impact

The license was rejected and I asked for multiple other databases such as WVB Global Fundamentals Data ,RepRisk ESG Risk Metrics and Analytics. They had similar information but it was not accepted either. Also, I wrote emails to authors of papers that addressed similar topics but did not even receive a response. Finally, after many attempts and searches in Kaggle we found a public dataset with ESG scores measured by MSCI. It will be studied in detail in point 3.1.2 below.

3.1.1 Countries DataBase

We did a query about Worldwide Governance Indicators. The 206 available countries and 40 variables are selected over a 14-year time series from 2005 to 2019. The 6 main variables are estimated on the aggregate indicator, in units of a normal distribution, from -2.5 to 2.5. They are:

Control of Corruption (CC): captures the public's perception of the extent to which there are illicit and illegal practices by those in positions of power in the government for their own benefit.

Government Effectiveness (GE): captures the image that the population has of the quality of both the administrative organization and public services. The quality of policies, as well as the confidence they have in their fulfillment. It considers whether the government is independent or yields to political pressures.

Political Stability and Absence of Violence/Terrorism (PA): refers to the degree of political instability and politically motivated violence as perceived by the population.

Regulatory Quality (RQ): raise awareness of the government's ability to develop and implement appropriate policies and regulations to facilitate and promote private sector development.

Voice and Accountability (VA): measures the freedom perceived by citizens, in the media, in the election of the president, freedom of expression and association. It represents the level of voice and vote of the country's inhabitants.

Rule of Law (RL): refers to insights into the extent to which agents trust and follow social norms, in particular the quality of contract enforcement, property rights, police and courts, and their capacity to cause harm.

For each of these variables we have the score estimate, the number of sources used for its calculation, the percentile rank, the lower limit of the percentile rank, the upper limit of the percentile rank and the standard error of the estimate.

We grouped each main variable CC, GE, PV, RQ and VA with all the available information for each one and made a correlation matrix. It is observed that all of them have a high correlation between 0.97 and 0.99. Therefore, the method will be performed only with the 6 main ones.
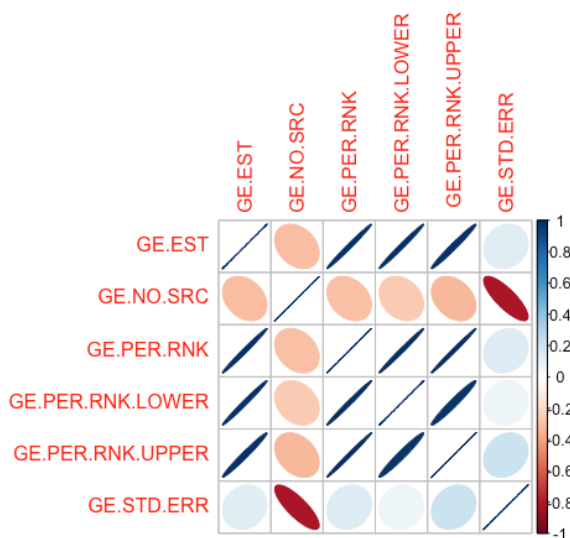

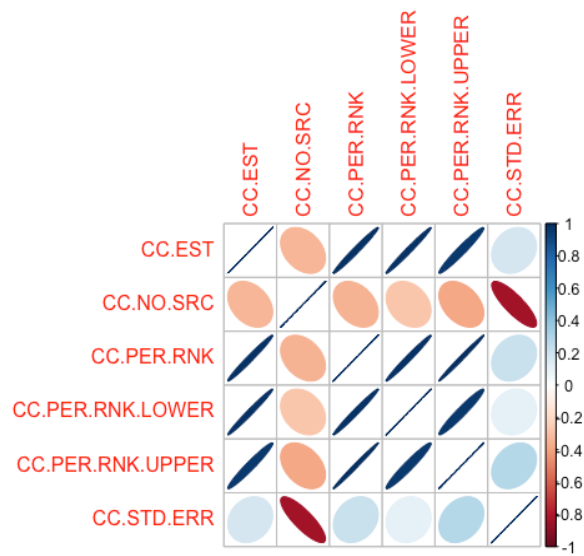*Figure 5 Correlation Matrix Govern Effectiveness*

Figure 6 Correlation Matrix Control Corruption

3.1.2 Companies DataBase

MSCI is an American company dedicated to rating companies according to different parameters focused on investment markets. As we have studied in the state of the art, it has been quantifying these variables since 1990. In Kaggle, we find a dataset that collects the ESG scores assigned by MSCI to a multitude of companies. The year 2019 is chosen to follow the same line of research as the governments of the world. These weightings allow us to do research on the sustainability of companies.

The dataset is made up of 15645 companies whose name, country, Environmental Score, Social Score, Governance Score, Overall ESG rating, Overall ESG score are indicated. There are 3 qualitative and 4 quantitative variables. They are defined as follows

Environmental Score: considers the company's performance as a protector of nature. Evaluates the impact of the company in a direct or indirect way with the environment. It measures the quality of the policies of the business in terms of ecology and sustainability. It is measured if the enterprises mitigate the possible negative effects of their activity and if they undertake actions that generate a direct positive impact.

Social Score: evaluates the impact that a company has on the community, i.e., on its social environment. An assessment is made of whether human resources are promoting equality and diversity among the workforce and guaranteeing a healthy environment, for all stakeholders involved in.

Governance Score: grade how the leading administration of the firm performs tasks such as considering the composition and diversity of management councils, developing ethical and best practice guidelines, examining supply chains to ensure regulatory compliance, and providing transparent disclosure of tax information in their accounts.

Overall ESG Rating: companies are rated in 3 groups depending on how they manage each important aspect shown in the following Figure.

Figure 7 Key Issues Framework ESG rating

Leader: group where companies accomplish impeccably, the most important Key Issues for all the 3 pillars. This group is tagged with AAA and AA.

Average: entities that manage to satisfactorily comply, but not always, with the requirements of the Environmental, Social and Governance Score. Its labeled A, BBB and BB.

Laggard: those that do not adequate to the conditions proposed in each of the branches and are not capable of dealing with ESG risks. Label as B and CCC.

Overall ESG Rating is a qualitative variable, so in order to work with it we decided to recode them. Those labelled with AAA and AA, which would be the leaders, with a 1. Those with A, BBB and BB, that is, the average with a 2 and finally with a 3 to those belonging to the Laggard group, tagged with B and CCC.

Overall ESG Score: is the mean of the three first variables Environmental Social and ESG Score.

4. Principal component analysis (PCA)

4.1 Application Method

In this analysis we will analyze the 206 countries that have been studied for the 6 variables described above. Once we have done the cleaning we are going to work with the centered data.

PCA is performed on the centered database, where we have the 6 quantitative variables. The PCA model is generated on the centered and scaled data and we select the number of principal components from the scree-plot. The knitr library is loaded to obtain the table showing the percentage that can explain each dimension. Three principal components (PC) are selected. With the first PC, which suggests the first "elbow", 86% of the variability of the data is explained. The 3 PCs explain more than 95% and are close to the eigenvalue criterion which is represented by the red line in the scree plot. It cuts dimension 3, i.e. it suggests 3 principal components.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	12751.9947	86.6707825	86.67078
Dim.2	834.0130	5.6684904	92.33927
Dim.3	502.2734	3.4137741	95.75305

Figure 9 Table percentage of variance for each dimension
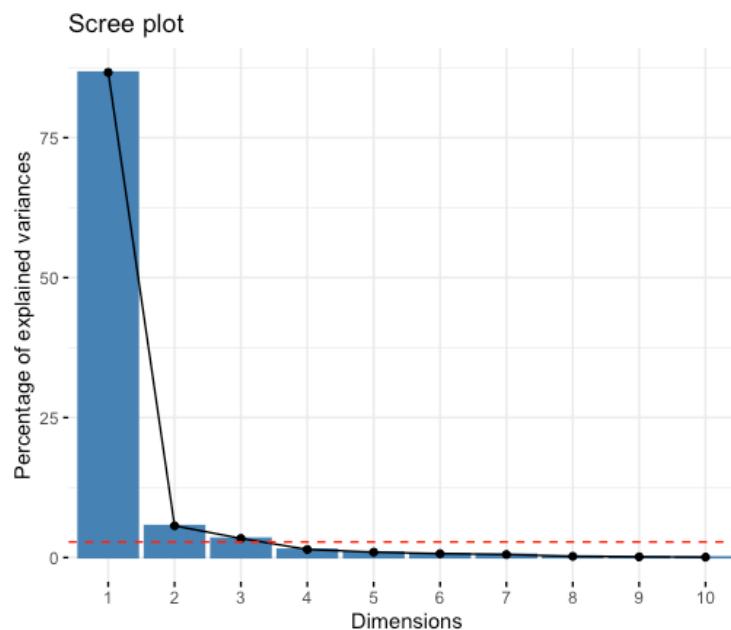
Figure 8 Scree Plot

4.2 Numerical and graphical results

Once the PCA has been calculated, the loading plots are used to analyse which variables contribute most to each principal component and, therefore, in which variables the countries indicated show different behaviours. The loading scores to be worked with are those multiplied by the standard deviation of the corresponding component, i.e. the square root of the associated eigenvalue.
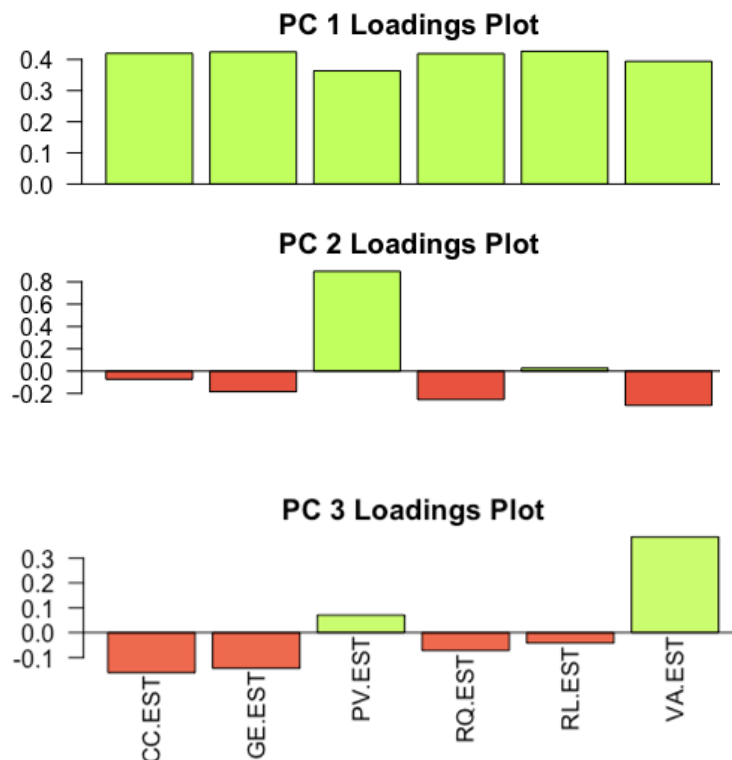
Figure 10 Loadings PC 1, 2 Y 3

The loadings scores of each variable in the different principal components are plotted. In this way, it is possible to analyse how each variable behaves and its correlation in each dimension. Principal component 1 presents all 6 variables in positive, so all variables in our database are positively correlated in dimension 1. If a country has a very high positive contribution in principal component 1, it means that it has obtained very high values in all global governance indicators.

In principal component 2, the variable PV is positively correlated with dimension 2, as well as Rule of Law which has a positive value very close to 0. The remaining variables Control of Corruption, Government Effectiveness, Regulatory Quality, and Voice and Accountability are negatively correlated in component 2. In this case, if a country's contribution score is very high and positive it means that the country has political stability and is unlikely to be destabilised by

unconstitutional means. But the government will not be effective, there will be no quality of policies in place, no control of corruption, no freedom of expression and no quality of policies in place.

In the last dimension we find the variable Voice and Accountability with a positive value, i.e. it is positively correlated with principal component 3, as is PV but to a lesser extent. While Control of Corruption, Government Effectiveness, Regulatory Quality, and Rule of Law are negatively correlated as their loading scores are less than 0. Countries with high contributions to this dimension will be those that do not have a prosperous government but citizens perceive a certain degree of freedom.

Once the importance of the variables in each component is understood, we proceed to study the individuals, the countries. The figure 11 represents the scores of the countries, i.e., their projections in the new space formed by the first two dimension. The countries are colored by their contribution to the components. First of all, it is notable that there are some countries with very high scores in some of the components, behaviors that we will study later.
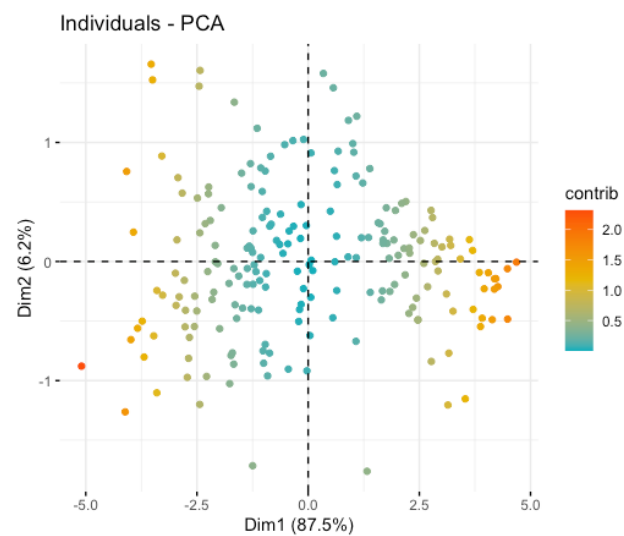
Figure 11 Countries scores in PC 1 and 2

The graphs of the loadings will allow us to interpret the meaning of the principal components and discover the correlation between the variables of the process. To begin with, we plot dimension 1 and 2 for Loading Plot and Score Plot.
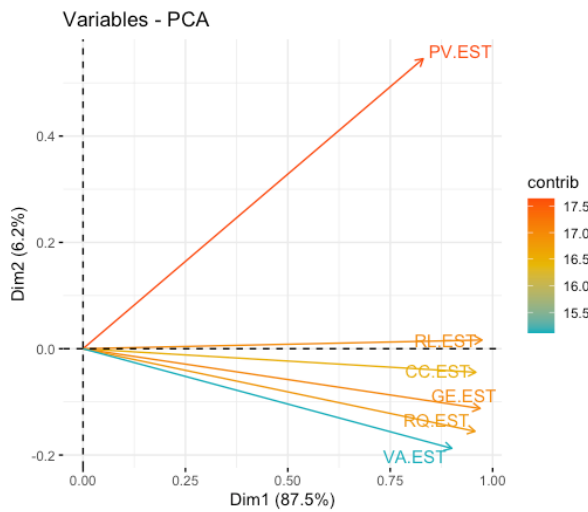
Figure 12 Contributions of variables in PC 1 and 2
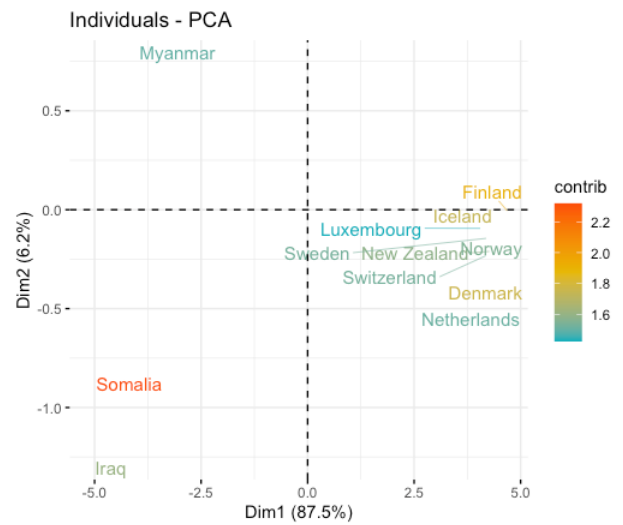
Figure 13 Countries contribution in PC 1 and 2

In Figure 13 Political Stability is significantly positively correlated with the first and second components. Rule of Law also appears in this quadrant but with a lower contribution. On the other side, the remaining 4 variables are found in the fourth quadrant. This means that they present a positive correlation with dimension 1 but are negatively correlated with dimension 2. If we look at Figure 12 it is observed the behavior of the countries in the different components. Due to Finland's position, we would highlight a contribution in the Rule of Law variable. It could be said that Finnish citizens trust and follow the policies established by their government. Furthermore, we find Norway, Iceland, the Netherlands, Denmark, Switzerland and Luxembourg in the fourth quadrant. Therefore, we can say that these countries have high scores in Control of Corruption, Government Effectiveness, Regulatory Quality, and Voice and Accountability.

It is worth mentioning that Somalia has a very high but negative contribution in both principal component 1 and component 2, so we would say that it is a country with negative estimates in all 6 variables. We would be talking about an underdeveloped country.

4.3 Visualization

After calculating the PCA analysis to our data we get a list of matrices containing all the results of the active variables: coordinates, correlation between variables, axes, cosine squared and contribution. Using resulsts.pcaindcoord we access the coordinates of each country in each dimension. Therefore, we can analyze the behavior of all countries in the different components. A world map is made using ggplot() and each country is filled with its coordinates, i.e. the contribution in the component.
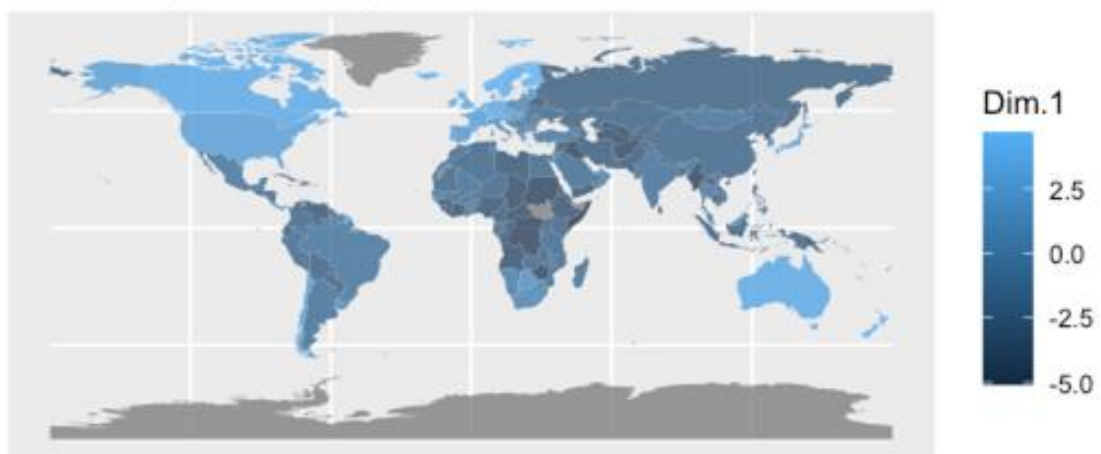
Figure 14 World Map 2005 contribution countries PC 1

World Map PC1

There are several countries with contributions above 2.5, Europe, especially the Scandinavian countries such as Finland, Norway and Sweden; as well as North America and Australia. In this component all the variables have a positive correlation. This means that these countries in 2005 had achieved stable governments based on the implementation of quality policies, with a control of corruption and favoring the freedom of citizens. In contrast, those countries that are represented by the darker blues have negative contributions. In other words, they are governments that lacked political stability, adequate regulation of the country and freedoms. These would be most of the countries in Africa, notably Sudan, the Republic of Congo and Somalia; in Asia, Myanmar, Syria, Turkmenistan, Kazakhstan and Uzbekistan.

World Map PC2

World Map of Loadings Scores

Figure 15 World Map 2005 contribution countries PC 2

In this map of dimension 2, the countries with the highest contribution are represented by lighter blues. It is identified Libya, Belarus, Turkmenistan and North Korea. This means that they have high political stability. In 2005 all these countries were governed by dictators who exercised a totalitarian regime. So due to the strict control they managed to avoid the possibility of being overthrown by violent means. But as we have studied before, the variables Control of Corruption, Government Effectiveness, Regulatory Quality, and Voice and Accountability are negatively correlated in this component. So they are stable governments in which there is a lack of freedom, quality in policies and public services as well as regulation of the private sector. In contrast, in the lowest scores distinguish Colombia, which was threatened by armed opposition formed by criminal gangs and organized armed groups (OAGs). Russia in 2005 was also threatened by Ukraine with the Orange Revolution and by Kyrgyzstan with the Tulip Revolution.

We could found in the article "Governance Matters VII: Aggregate and Individual Governance Indicators" justified how internal conflicts are measured in the positive variable of this 2 CP "The highest scores go to countries without armed opposition, and where the government does not indulge in arbitrary, direct or indirect violence. The lowest scores go to countries torn by civil war." (Daniel Kaufmann, p. 66, 2007).

World Map PC3

World Map of Loadings Scores

Figure 16 World Map 2005 contribution countries PC 3

In this case, the VA variable is positively correlated, so countries with negative values in this dimension have very low scores in Voice and Accountability. China and Saudi Arabia are the most notable. This means that their citizens perceive a lack of freedom despite they are effective governments.

4.4 Evolution in time

After analysing the year 2005, it is studied whether the data obtained in the different years present the same principal components or whether they vary over time.

4.4.1 Components

To carry out this study, PCA is performed on the database where the entire time series 2005-2019 appears. The contribution matrix is calculated using "pcavarcoord" and it is obtained that the variable Time has a loading score of 0.00670 in PC1, - 0.05 in PC2 and 0.03 in PC3. It is used the "kable" library for a better visualisation of the matrix.

The contribution of the variable Time is practically non-existent, so it could be said that the importance of the variables in the different components does not vary over time. However, to check this, we have calculated the PCA in the year 2010 and 2019 separately.
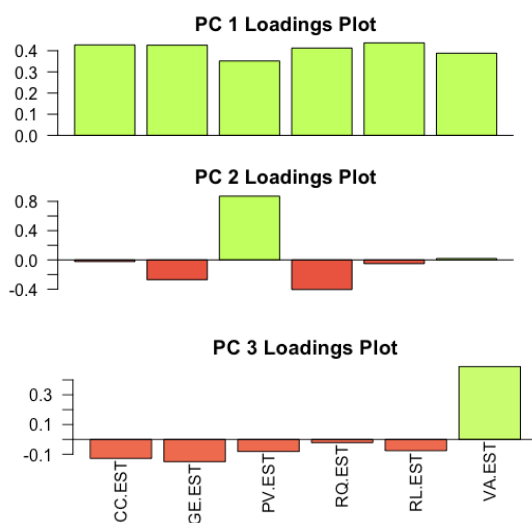
Figure 18 Loadings scores 2010
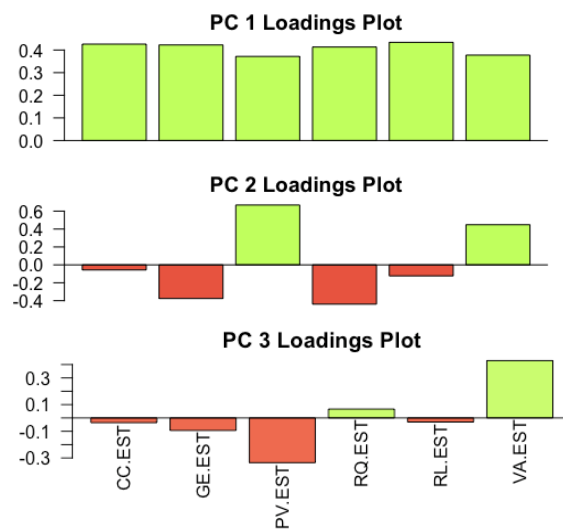
Figure 17 Loadings scores 2019

In both graphs, the correlation of the variables in the dimensions is practically the same as that studied in 2005, but it is necessary to highlight some changes. In component 2 in 2019, the variable Voice and Accountability has a positive value of 0.36 while in 2010 it was 0 and in 2005 it was -0.21. Furthermore, in dimension 3 the Regulatory Quality variable continues to have a value very close to 0 but in 2010 it is negative and in 2019 it is positive. And Political Stability has a much higher negative value.

4.4.2 Individuals – Countries

This study aims to investigate whether the dynamics of the governments of the countries studied in 2005 have evolved. We analyze whether they continue to report in the same way in the different dimensions according to the new correlations of the variables. The contibution matrix of the individuals of the PCA calculated from the 2019 dataset is visualized in maps. These are the plots obtained.

Figure 19 World Map 2019 contribution countries PC 1

Countries that had a high contribution when all variables are positively correlated in dimension 1 are the same in 2019. This means that the countries that were developed are maintained over time. In addition, countries such as Syria and Colombia have higher negative values. After 14 years, their governments have suffered conflicts and political instability, resulting in a worse weighting than that obtained in 2005.
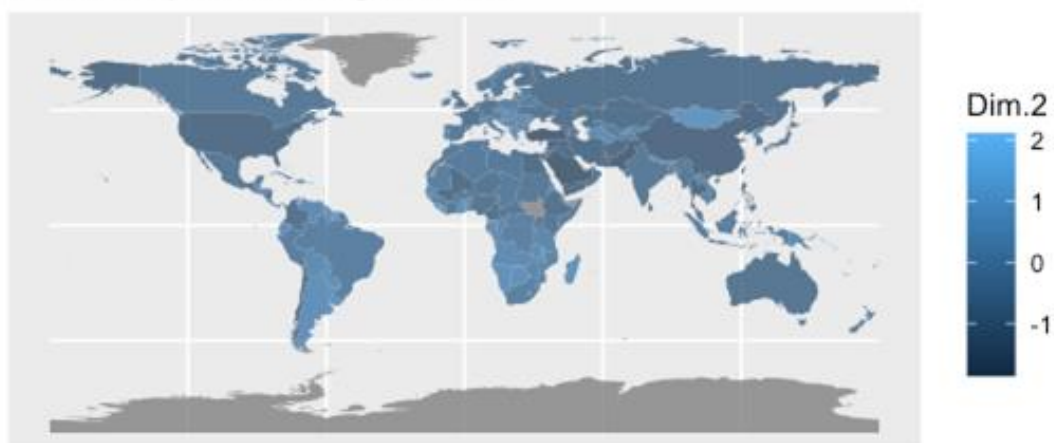
Figure 20 World Map 2019 contribution countries PC 2

In this dimension 2 there is great change in the behavior of individuals, since in this case the Voice and Accountability variable takes positive values. Those countries that had a high contribution in 2005 due to a great political stability are now the same countries with the most negative values in this component. This shows that this stability is governed by a totalitarian regime that generates a lack of freedoms for citizens.

After studying this map in Figure 5.9, we can say that there are no governments that are stable and free but do not have an effective government, a control of corruption or a regulation of the quality of laws.

Figure 21 World Map 2019 contribution countries PC 3

In the case of the third component, we note that compared to 2005 there is a clear increase in the contribution in most of the countries. Over the years and with the development of society, citizens are demanding their rights, and this is reflected in the 2019 map with an increase in the contribution from 0 to 1. However, the Political Stability variable has a negative contribution of -0.33. So those countries with a very high contribution in this component mean that citizens have a certain freedom, but the government is not stable and there is a high presence of politically motivated terrorism. These would be countries such as Mali, Ukraine, Iraq and Afghanistan. On the other hand, China and Saudi Arabia still have a negative contribution. In other words, countries that had no freedoms, but a stable government remain in place over time.

5.Companies

Now with the companies data set it is going to be studied the way the different individuals are associated with others. Analyse which are the characteristics that create different clusters and which are their similarities.

5.1 Cluster K-means

K-means clustering is an unsupervised learning technique to classify unlabeled data by grouping them by features. There is a variable K which represents the number of groups or categories created. The goal is to split the data into K different clusters and report the location of the center of mass for each cluster. Then, a new data point is assigned a cluster based on the closed center of mass. Different measures can be used to find the centroids. Since we want to study companies with similar characteristics, we choose the Euclidean distance. First, a centroid is selected at random, then the Euclidean distance from the current centroid to each of the points is calculated and the average is obtained. The mean of the x0 and the mean of the x1 will be the new centroid coordinate. To determine the variable K, i.e. the number of clusters, we will use two methods.

5.1.2 WSS

It is called Elbow method because after calculating the sum of squared errors within the cluster (WSS) for different K, we must choose the one where the WSS starts to decrease and this is represented in the graph as an elbow. In the graph we can see that the Elbow method shows a differential elbow at K = 3.
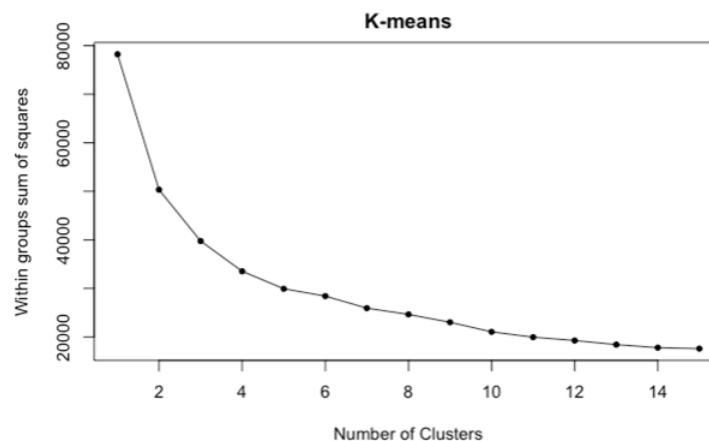
Figure 22 WSS plot

5.1.3 Silhouette

The mean silhouette method is very similar to the Elbow method, except that instead of reducing the sum of squares between groups, the silhouette coefficients are averaged. Approach. This parameter determines the allocative quality of an observation by comparing how similar it is to the rest of the observations in its group with those in other groups. Its value can range from -1 to 1, with high values indicating that the score is assigned to the correct group.
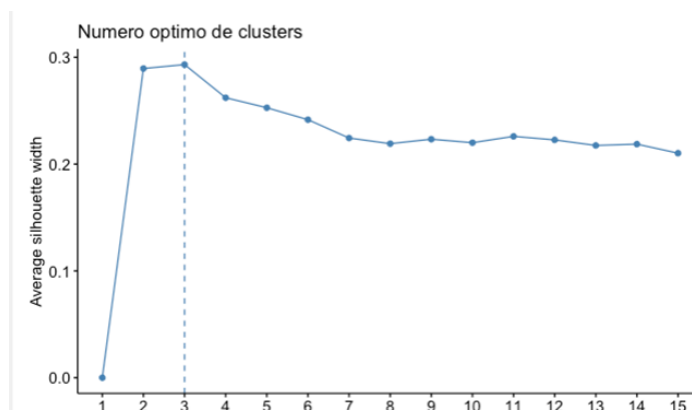
Figure 23 Optimal number of clusters by silhouette

The average silhouette method considers the optimal number of clusters to be the one that maximises the average silhouette coefficient of all observations, in this case 3.

5.1.4 3-Means

Once the variable K = 3 is selected, the clustering is created. We plot each company with different colours grouped by the cluster they belong and draw a ellipse around each association calculated. The following graph is obtained in Figure 22. The clusters are clearly grouped by dimension 1 as the cuts are vertical

Figure 24 Cluster K-means

In order to study which variables are important in dimension 1 and therefore create the different associations, we applied a PCA on the same data used for the cluster. We plot the loadings scores in the first two dimensions and In component 1, the 3 ESG Scores: Governance, Social and Environmental are positively correlated. Therefore, the further to the left on the X-axis, the lower the scores of the companies. Thus, the blue cluster 1 will contain the leading companies in the 3 ESG scores.

Figure 25 PCA ESG Scores Companies

This coincides with the rating in which they are divided as follows:

Figure 26 Capture of MSCI ESG ratings groups

Meanwhile in component 2 there is a clear differentiation between the Governance variable which maintains a positive correlation while Social and Environmental Scores are below the X-axis. They have negative values so they would be negatively related to dimension 2.

If we analyze these contributions in the cluster, we can understand their behavior. The companies that are in the first quadrant are those that are politically legal and public have a transparent image of them due to their policies. Also means having a good organization of decision making and shareholders. But on the contrary, they would not be trying to improve the impact in the nature and in people they work with. These would be considered not ethical as they have negative Social and Environmental scores. On the other hand, in the fourth quadrant we would find those companies that are Ethical but with governance at a minimum to be able to operate.

6. Global World Score

The characteristics of different countries and the performance of 15,000 companies have been analysed. To take into account both the company's policies and the environment in which it operates, it is proposed to calculate a Global World Score. This would be the score a company would obtain for its impact in social and environmental terms, but it would also evaluate the policies of the country in which the company is located. In this way, it would be possible to study how the government favors or penalizes different companies. This new rating is a heuristic that we have designed based on the results of all previous analyses.

6.1 Calculation

First, we obtain the values of the loadings of each of the variables in the first component, which is the one that indicates the general ethics of a country. For each country, we calculate the following linear combination

$$Ethics_C = CC_C X_1 + GE_C T X_2 + PV_C X_3 + RQ_C X_4 + RL_C X_5 + VA_C X_6$$

, where CC, GE, PV, RQ, RL, VA are the scores obtained by each country c and Xi corresponds to the weight of each variable i in the first component of the PCA.

The following formula is then applied to each company:

$$GlobalScore_b = ESG_b * Ethics_C$$

,where for each business I multiply its Overall ESG Score by the Ethics score of the country in which the company operates.

The ESG score of each company is multiplied by the Ethics score of its country. To do this Data Mapping techniques are used. Companies database is scrolled through and for each row the program calls a function 'get_ethics' that compares the two-letter country code with the WGI database where the name has been encoded in the same way. It finds the country it matches and returns the ethics score for that country. It then multiplies that value by its ESG score.
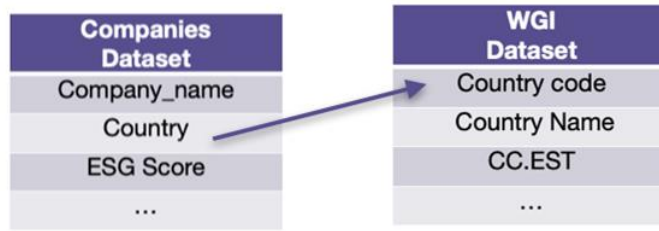
Figure 27 Scheme Data Mapping by country

6.2 Ranking Analysis

The calculated ranking is analyzed. The database is divided into quartiles and 4 datasets are obtained with 25% of the companies in each one. As the database is ordered according to the GlobalScore, the best-ranked companies will be found in the first quartile. So on down to the fourth quartile. Using python, it is created a function that calculates the frequency of countries in each quartile. A counting of all the companies appearing from each country is performed and stored in a dictionary. This dictionary is converted to a dataframe to be able to manage it later. This function is applied to the 4 quartiles and the following datasets are obtained, now the first 10 countries of each one are shown.

Number of companies of each country in each quartile

	Country	N_Companies_1Q
0	NZ	73
1	NO	78
2	FI	59
3	CH	163
4	LU	132
5	DK	65
6	NL	207
7	SG	97
8	CA	330
9	AU	364

	Country	N_Companies_2Q
0	US	2058
1	AU	68
2	BE	23
3	IE	27
4	JP	418
5	NL	36
6	GB	226
7	FR	78
8	AT	11
9	LU	31

	Country	N_Companies_3Q
0	US	1760
1	AU	31
2	KY	92
3	SE	185
4	VG	74
5	MO	12
6	JE	23
7	CA	48
8	CH	9
9	JP	410

	Country	N_Companies_4Q
0	GR	8
1	KR	272
2	AE	19
3	FR	4
4	KY	22
5	VG	32
6	US	261
7	UY	1
8	MY	94
9	QA	28

1st   **2nd**   **3rd**   **4th**

Figure 28 Capture Dataset for each ranking quartile

However, it should be remembered that out of the 15,000 companies, there are not the same number for each country. Therefore, in order to compare the frequencies, it is necessary to

calculate the percentage of the number of companies of a country in a quartile in relation to the total number of companies of that country in the global dataset.

The number of companies from each country in the global database is calculated. And a dataset is created. From this dataset, for each country we obtain the percentage of presence of businesses in each quartile. With a .head() part of the results obtained are shown.

	Country	Number_Companies	%FirstQuartil	%SecondQuartil	%ThirdQuartil	%FourthQuartil
1	NO	86	90.697674	6.976744	2.325581	0.000000
2	FI	65	90.769231	9.230769	0.000000	0.000000
3	CH	195	83.589744	11.794872	4.615385	0.000000
4	LU	198	66.666667	15.656566	14.646465	4.040404
5	DK	72	90.277778	8.333333	1.388889	0.000000
6	NL	273	75.824176	13.186813	11.721612	0.732601

Figure 29 Dataset %freq companies of each country in each quartile

6.3 World Map visualization

In an attempt to understand the results of the newly created dataset, a map visualization is performed with ggplot2 in R. The corresponding tables of information are also presented, ordered from highest to lowest percentage. We are going to represent each quartile on a map in purpose of analyzing if there is a pattern of behavior of the countries. It will be observed whether all the companies of the same country are in the same quartile or on the contrary they are dispersed throughout the ranking.

Percentage of Companies in 1st Quartile

Country	%FirstQuartil
AW	100.000000
LI	100.000000
FI	90.769231
NO	90.697674
DK	90.277778
NZ	89.024390
CH	83.589744
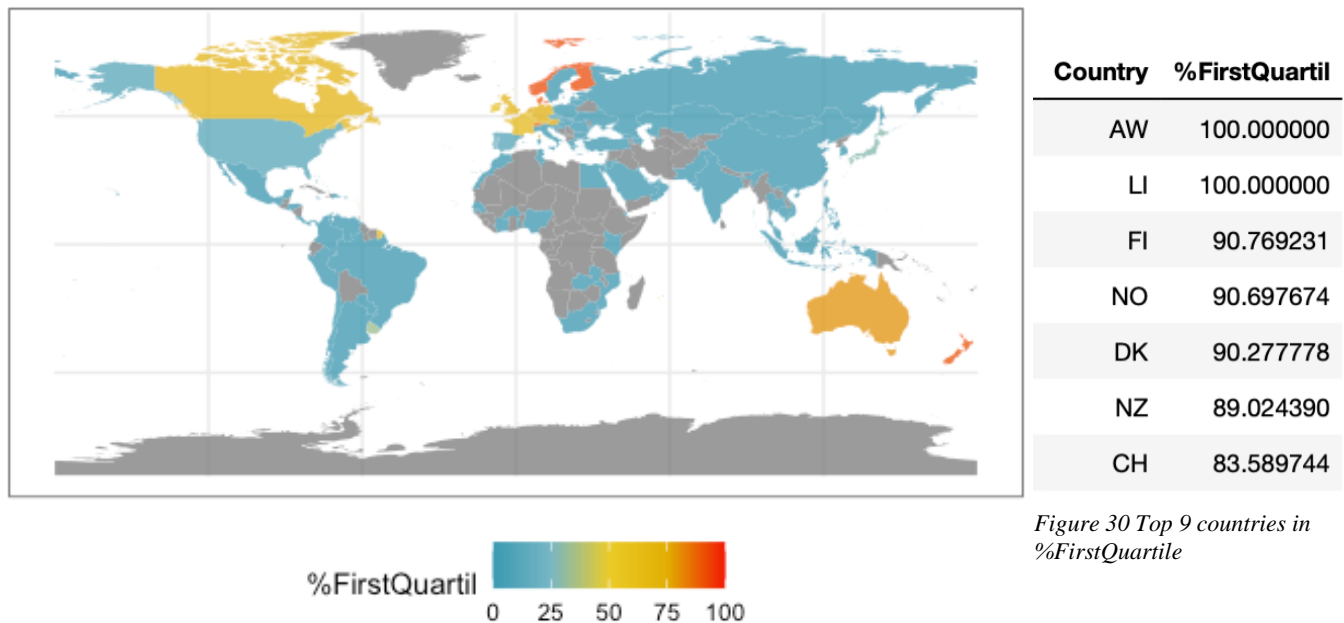
Figure 30 Top 9 countries in %FirstQuartile

Figure 31 World Map of percentage of Companies in 1st Quartile

Norway, Finland, Denmark, Liechtenstein, Aruba, Switzerland and New Zealand stand out with more than 83% of their companies in the top 25% of the ranking. This means that both their policies and those of the country are ethical and efficient. Also, Australia, Canada, France, Germany and Guiana are in the 70-75% range.

They are followed in green by Uruguay with 33.3% and Portugal with 24.13%. As for the rest of the countries, their presence in this case is 0%, except for the United States, which reaches 13.25%.

Figure 32 World Map of percentage of Companies in 2nd Quartile

Country	%SecondQuartil
SI	100.000000
CZ	52.941176
SK	50.000000
LT	50.000000
IM	50.000000
ES	49.650350
PT	48.275862
TW	45.731707
US	43.936806

Figure 33 Top 9 countries in %SecondQuartil

In the second quartile the only country that has 100% of the companies is Slovenia; with a large difference with respect to the following countries that range between 43-53% which are Slovakia, Lithuania, Spain, Portugal, Taiwan, United States and Chile. It is also worth mentioning Canada that reaches 22.2% while the rest of the countries have a very low presence of between 0-14%.

It can be seen that in the second quartile the countries have companies distributed in different positions of the ranking, since here the top 5 have only half of them. In contrast, the percentage of enterprises from the top 10 countries in the 1st quartile is almost 100%. This could be explained by the fact that the countries in the first quartile both the government and the company implement and enforce policies designed not only from an economic point of view but also taking into account the social and environmental impact.

Percentage of Companies in 3rd Quartile

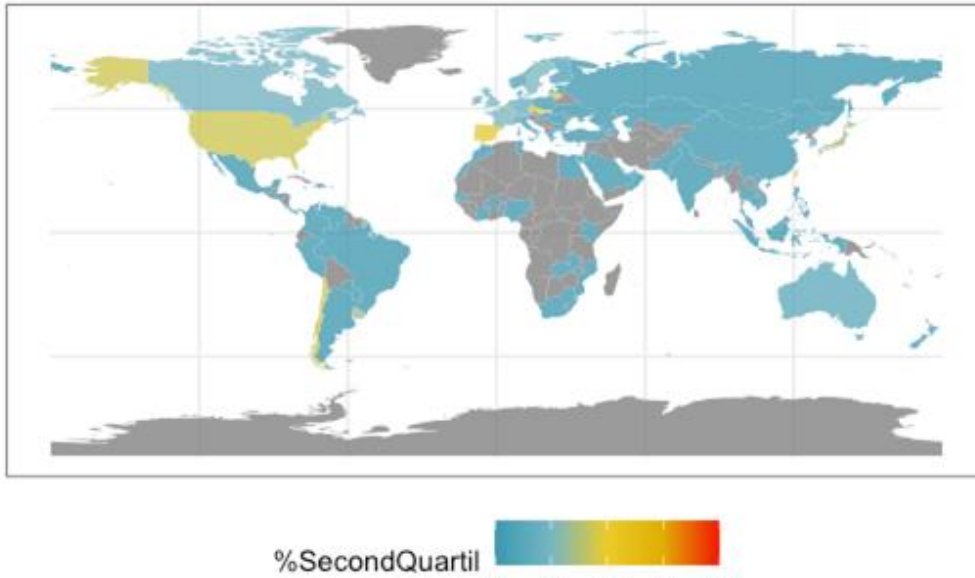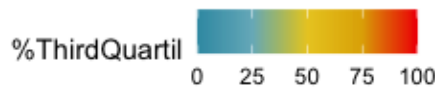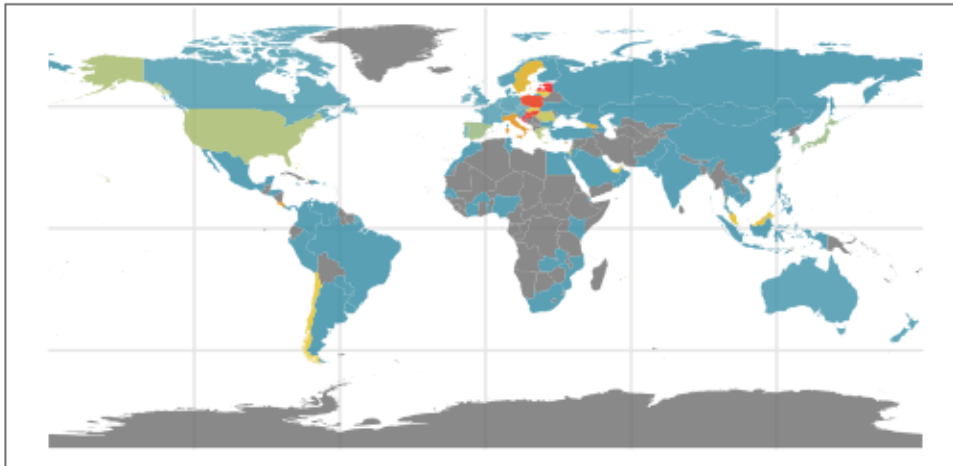| Country	%ThirdQuartil
FO	100.000000
BB	100.000000
LV	100.000000
MO	100.000000
HR	100.000000
EE	100.000000

Figure 34 Top 7 countries in %ThirdQuartil

Figure 35 World Map of percentage of Companies in 3rd Quartile

There are the Baltic countries Estonia, Lithuania and Croatia; and islands such as Barbados and the Foe Islands with 100% of their companies in this quartile. They are followed by Hungary and Poland with more than 90%. And Sweden with 70.5%. We notice that the US and Spain again have a percentage of between 35-50% on this map. So the companies of these countries are distributed throughout the world ranking.

Percentage of Companies in 4th Quartile
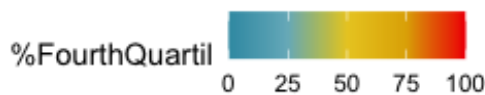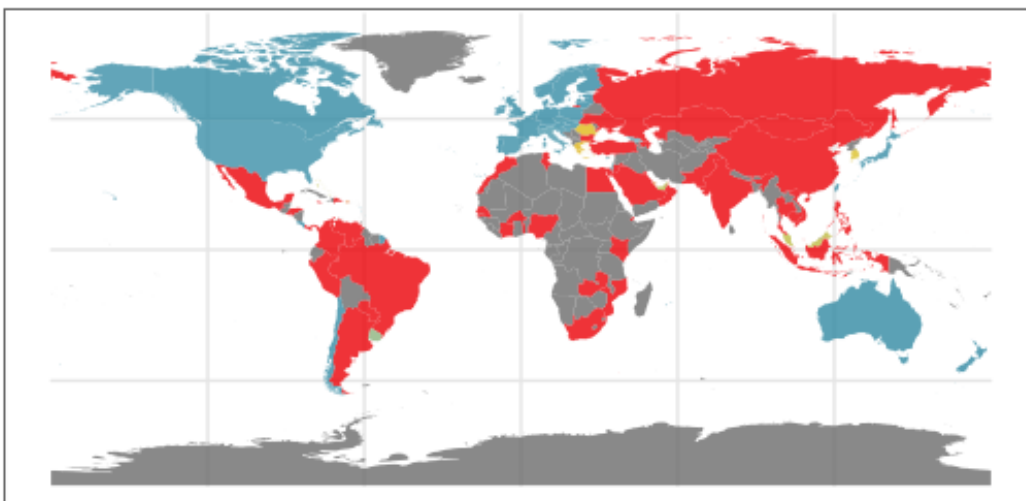
Country	%FourthQuartil
VE	100.0
PE	100.0
SN	100.0
MN	100.0
JM	100.0
PA	100.0
OM	100.0
ZA	100.0

Figure 36 Top 10 countries in %FourthQuartile

Figure 37 World Map of percentage of Companies in 4th Quartile

This figure shows a big difference with the rest of the maps. All the countries that have ranged between 0-8% in the first 3 quartiles have 90-100% of companies in the bottom 25% of the world ranking. This would be a large part of Asia and South America, as well as Africa. After analyzing the latest world maps we detected that there is a common behavior among the countries of the same continent. This will be studied in more detail.

6.4 Ranking divided by continents

In order to group the ranking by continents we first proceed to look for a database containing both the country code and the Region and Sub-Region. We found one on the Kaggle website. Once the new database is obtained, a small cleaning of the database is performed. We focus on the Region and Subregion columns and detect 0% null values. It can be worked with.

The variables Region and Subregion are added to the Ranking database. To do this, we use Data Mapping to compare and integrate both datasets. The Ranking is processed by rows and the country code that both datasets have in common is compared. When it matches, the region of that same row is returned and added to the Ranking dataset. In this way, the region and subregion in which they are located are added to the 15,000 companies.

The companies are then grouped by continent using the same technique as for the countries. A first count of the total number of companies belonging to each region, then a count of the companies from each continent in each quartile. A data frame is obtained for each quartile containing the number of companies. Then, the percentage is calculated using the global dataset of the ranking and the new one we have obtained.

	Regions	Number_Companies	%FirstQuartil	%SecondQuartil	%ThirdQuartil	%FourthQuartil
0	Oceania	545	80.183486	13.211009	6.605505	1.100917
1	Europe	3547	54.468565	20.411615	20.947279	5.102904
2	Asia	5067	10.440103	16.400237	20.307874	53.068877
3	Americas	6305	15.352895	36.494845	32.529738	15.876289
4	Africa	179	0.000000	2.234637	3.910615	93.854749

Figure 38 Dataset companies group by Region

	Subregions	%FirstQuartil	%SecondQuartil	%ThirdQuartil	%FourthQuartil
0	Australia and New Zealand	80.478821	13.259669	6.629834	0.736648
1	Northern Europe	56.265060	21.807229	21.024096	1.927711
2	Western Europe	70.721049	18.208303	10.924982	1.310998
3	South-eastern Asia	15.179969	5.790297	25.508607	53.521127
4	Northern America	18.238994	41.776253	34.991424	5.298266
5	Eastern Asia	12.304187	22.301339	22.614640	43.093136
6	Southern Europe	6.896552	28.160920	53.735632	11.206897
7	Eastern Europe	1.807229	8.433735	34.337349	55.421687
8	Latin America and the Caribbean	1.039698	10.302457	20.321361	68.336484
9	Sub-Saharan Africa	0.000000	2.666667	4.666667	92.666667
10	Western Asia	0.000000	3.514377	23.003195	73.482428
11	Micronesia	0.000000	0.000000	0.000000	100.000000
12	Southern Asia	0.000000	0.000000	0.000000	100.000000
13	Northern Africa	0.000000	0.000000	0.000000	100.000000
14	Central Asia	0.000000	0.000000	0.000000	100.000000

Figure 39 Dataset Companies group by Subregion

A stacked bar chart is generated for easier and more efficient analysis. The first shows the percentage of companies in each quartile by region. It is understood that Oceania and Europe have 80% and 54%, respectively, of their companies at the top of the ranking. Taking into account that the total number of businesses in Oceania is 545 and Europe 3547. It could be said that more than 1200 European enterprises are among the most ethical and sustainable in the world.

America has a total of 6305 corporations, of which 60% are evenly distributed between the second and third quartile. In other words, the majority are positioned in the middle of the total ranking.
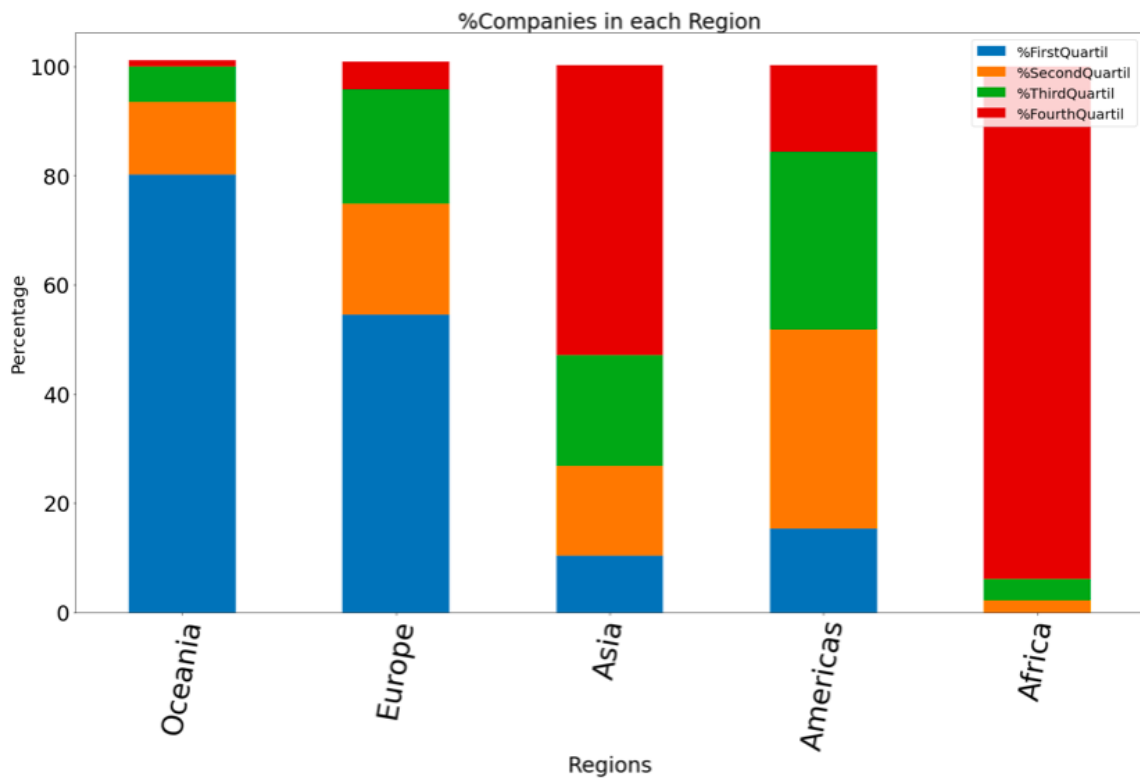

Figure 40 %Companies in each quartile group by Region Stacked Bar Chart

The remaining 40% is divided between the first and fourth quartiles. Being a continent made up of very different cultures between North and South America, there are both leading companies and companies that are still underdeveloped.

In contrast to America, Asia, despite also having firms in all quartiles, 50% of them are positioned in the last 25% of the global ranking data. The remaining 50% are distributed among the first three quartiles. Finally, Africa obtains practically 100% of the rankings at the bottom of the table, which means that neither industries nor the environment in which they are located are favorable. Furthermore, no African company appears in the top 3000.

To provide an even more concrete view, we have applied the same procedure but with the subregions. Figure 41 provides a multitude of conclusions. If we look at Europe, we find that the North and West are where the best-ranked companies are. Whereas Eastern Europe has more than 50% of companies at the bottom of the table and no company in the first quartile of the ranking. Therefore, its companies are not yet aware of and concerned about their social and environmental impact. Southern Europe would be more distributed between the second and third quartile.

If we remember America had both excellent and developing companies, they appeared throughout the table. Now we understand why. South America has more than 70% of companies in the last positions while North America has a similar percentage in the 3 quartiles and barely 5% in the fourth quartile. South and East Asia do have ethical companies as they have approximately 20% of their companies in the first quadrant of the ranking, while West and South have all enterprises in red. It could be said that South an East Asia are more developed in terms of business ethics. Finally, Africa has no subregion with more prosperous companies, maintaining the positions obtained in Figure 40, with almost 100% of the companies in the fourth quartile.
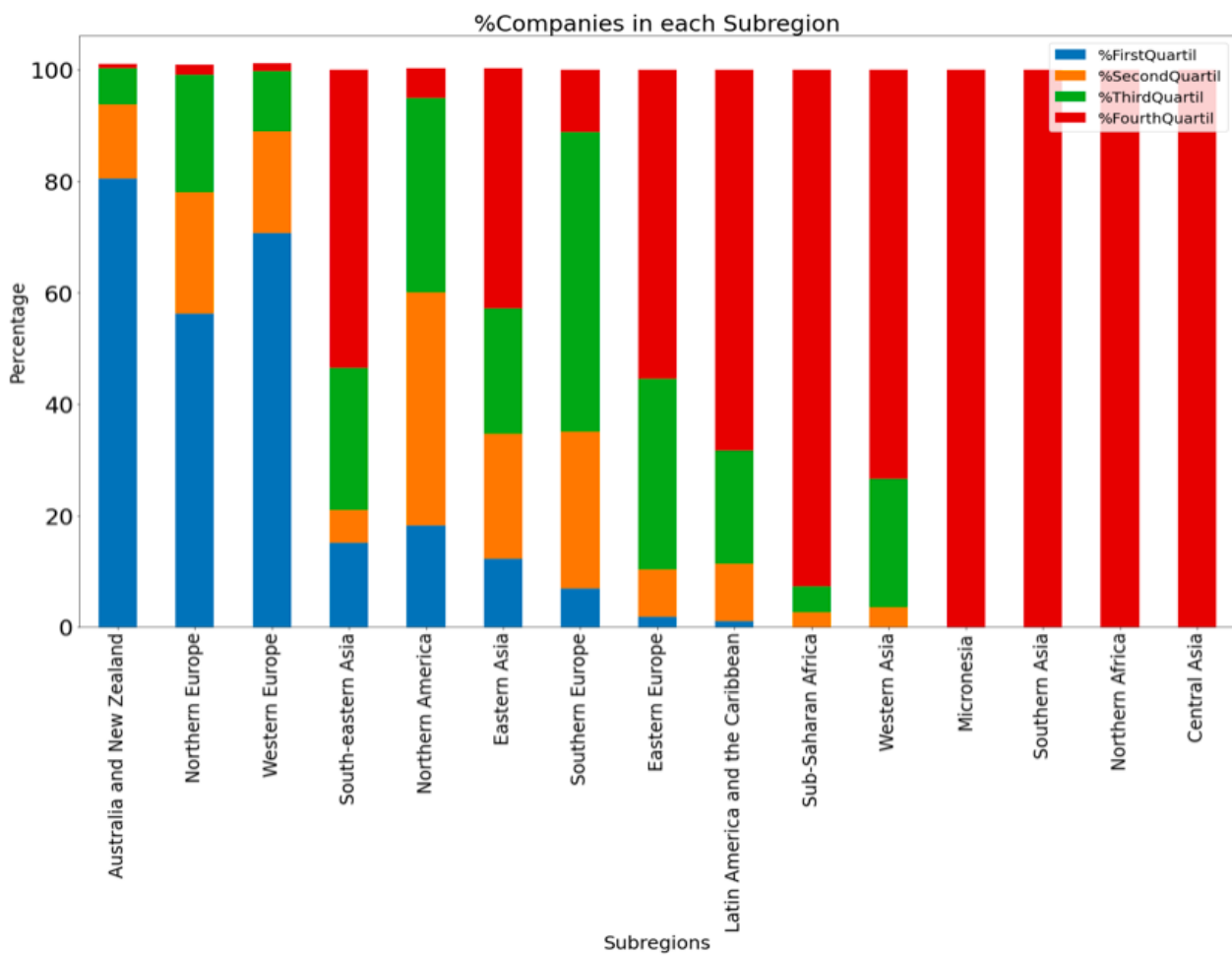

Figure 41 %Companies in each quartile group by Subegion Stacked Bar Chart

7. Conclusions

To sum up, the global ranking has been successfully generated by ordering the most ethical companies according to the Global World Score. Investors studying ESG scores can also consider whether the country, region or sub-region in which the firm is located is sustainable and in line with their beliefs regarding social responsibility.

The specific objectives have also been achieved. Thanks to the Principal Component Analysis in the World Governance Indicators dataset, it has been possible to study the behavior of the world's governments from a moral and sustainable perspective. Also, after the cluster analysis in the second ESG scores dataset, it was possible to understand the data to be used for the new weighting.

A Global World Score was calculated using Data Mapping techniques in order to integrate the data from the different datasets according to a common variable. For all of them, the common variable has been the country code. However, it should be noted that there are drawbacks since in the WGI dataset its scores are associated with errors, so when we calculate the new score we assume that the error is carried over.

Finally, the visualization of the data obtained using maps of the world has allowed us to discover similarities in the performance of companies on the same continents. We have continued in the research line of data analysis to be able to assess intangible characteristics.

8. Future projects

We have studied that business ethics is an indispensable part of a company. It is about having ethical principles by which a business can achieve a social commitment. But it cannot always be regulated by law but depends on the human being himself. I would suggest two new lines of research following this project.

In the first place, the calculation of the variables that measure the ethical traits of the company can be improved from the beginning because, being such a subjective aspect, the results may vary depending on the variables chosen for the study. An improvement for the future would be to take into account the region, not later as we have done, but in the calculation itself. Depending on the place where the different firms are located, there should be a predetermined weighting.

It would be interesting to add variables that capture the management policies of covid following the line of ethical investigation such as the research in the journal American Behavioral Scientist, the article "Consumer ethicality perception and legitimacy: competitive advantages in COVID-19 crisis". To perform a study in which, depending on the policies and the recovery process carried out, they receive a score. It would be possible to analyze which companies have managed to maintain their position in the ranking and which were more affected. Also, comparing whether it was a country that was hit sooner and how governments tried to mitigate infections.

9. Bibliography

Gjølberg, M. (2009). Measuring the immeasurable? Constructing an index of CSR practices and CSR performance in 20 countries. Scandinavian journal of management, 25(1), 10-22.

Gjølberg, M. (2009). The origin of corporate social responsibility: global forces or national legacies? . Socio-economic review, 7(4), 605-637.

Inderst, G., & Stewart, F. (2018). Incorporating environmental, social and governance (ESG) factors into fixed income investment. World Bank Group publication, April.

Issah, O., & Rodrigues, L. L. (2021). Corporate Social Responsibility and corporate tax aggressiveness: A scientometric analysis of the existing literature to map the future. Sustainability, 13(11), 6225. https://doi.org/10.3390/su13116225

Kaufmann, D., & Kraay, A. (2008). Governance indicators: where are we, where should we be going?. The World Bank Research Observer, 23(1), 1-30.

Kjellberg, S., Pradhan, T., & Kuhn, T. (n.d.). Chinaesg-pa2f.com., from https://www.chinaesg-pa2f.com/upload/file/20210313/20210313211157575757.pdf

MSCI – Powering better investment decisions. (n.d.). Msci.com. 2022, from https://www.msci.com/

Nasdaq Data Link. (n.d.). Nasdaq.com. Retrieved Feburary 12, 2022, from https://data.nasdaq.com/

Our ethical ratings. (2018, June 1). Ethical Consumer. https://www.ethicalconsumer.org/about-us/our-ethical-ratings

Payne, G., Blanco-González, A., Miotto, G., & del-Castillo, C. (2021). Consumer ethicality perception and legitimacy: competitive advantages in COVID-19 crisis. American Behavioral Scientist, 00027642211016515.

S&P global. (n.d.). Spglobal.com 2022, from https://www.spglobal.com/esg/scores/results?cid=3001792

10. ANEXO ODS

OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. **Fin de la pobreza.**		X		
ODS 2. **Hambre cero.**		X		
ODS 3. **Salud y bienestar.**			X	
ODS 4. **Educación de calidad.**		X		
ODS 5. **Igualdad de género.**		X		
ODS 6. **Agua limpia y saneamiento.**		X		
ODS 7. **Energía asequible y no contaminante.**	X			
ODS 8. **Trabajo decente y crecimiento económico.**	X			
ODS 9. **Industria, innovación e infraestructuras.**		X		
ODS 10. **Reducción de las desigualdades.**		X		
ODS 11. **Ciudades y comunidades sostenibles.**	X			
ODS 12. **Producción y consumo responsables.**	X			
ODS 13. **Acción por el clima.**	X			
ODS 14. **Vida submarina.**				X
ODS 15. **Vida de ecosistemas terrestres.**				X
ODS 16. **Paz, justicia e instituciones sólidas.**	X			
ODS 17. **Alianzas para lograr objetivos.**	X			

Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados.

Mi TFG parte de la base de estudiar la implicación de las empresas en ser sostenibles y éticas. Por lo que está muy relacionado y puede ir de la mano de los ODS. En nuestro caso, tratamos de analizar e identificar cuáles siguen políticas y están gobernadas por administraciones que están interesadas y preocupadas por la responsabilidad social.

Hemos estudiado la evolución de los gobiernos de 207 países, por lo que hemos podido localizar qué países tienen un gobierno eficiente, con políticas igualitarias que proporcionan libertad a los ciudadanos y capacidad para elegir a sus líderes. Por ello, nos acercamos a objetivos como:

- Paz, justicia e instituciones sólidas
- Producción y consumo responsable
- Reducción de las desigualdades

Mediante nuestro trabajo podemos observar qué gobiernos ya están en el camino de alcanzar este bienestar y cuáles todavía están en desarrollo. De forma que podemos centrarnos en reforzar aquellos países en los cuales todavía no se ha logrado. Además, se han estudiado las características éticas de 15000 empresas. Para ello se ha creado un ranking, cuyos resultados hemos visualizado con mapas ya que teníamos la información de donde estaba situada cada industria. De esta manera, conocemos qué países tienen más industrias de consumo sostenible, donde el gobierno y la empresa van de la mano. Siguen el mismo objetivo respecto la importancia del impacto qué pueden tener en la naturaleza y los recursos disponibles. Comparando con la ODS estaríamos hablando de objetivos tales como:

- Ciudades y comunidades sostenibles
- Industria, innovación e infraestructuras
- Acción por el clima
- Agua limpia y saneamiento.

Nuestro estudio ha sido más general y dentro de la efectividad de un gobierno no se ha entrado en detalle en aspectos concretos como cuán pobre es un país, si se pasa hambre o si tienen una educación de calidad. No obstante, creo que de forma indirecta si podemos detectar estas características. Si un país no tiene un gobierno que implemente políticas de calidad ni lleva a cabo ningún tipo de regulación. Lo más probable, es que estemos hablando de un país precario.

Como conclusión, nuestro trabajo fin de grado está muy relacionado con los ODS porque ofrece la oportunidad de ver cómo actúan los gobiernos respecto a gran parte de los objetivos y permite tomar decisiones que favorezcan a aquellos que todavía están en desarrollo.