Data Validation and Prediction for the Proposed JDI Index

Alfred Lerner College of Business and Economics
University of Delaware

Academic Advisor: Dr. Kalim Shah

Team Members:
Yahan Zhang
Tejaswini Ganti
Kamal K. Khanal

Fall 2020

DISCLAIMER

The **Data Validation and Prediction for the Proposed JDI Index** does not necessarily reflect the views or policies of the United States Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

Executive summary

Introduction

Jewelry Development Impact (JDI) index is a comparative country scoring system being developed within the framework of UN Indicators of Human Security. The aim of Data Validation and prediction is to provide a general scoring platform for the global jewelry sector to facilitate comparative socio-economic performance analysis of different countries by using the data that is already present from a few countries. Overall status of the sustainable performance of the jewelry sector is reflected in the cumulative score of the index where higher country score represents better performance of the sector. This research works aims to validate data frame collected and developed by the students of American University and develop the analytical method which will help in analyzing the current score and predict the score of the other countries.

Methodology

Regression Diagnosis for Data Validation

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors', 'covariates', or 'features').

In our analysis, first we run the regression using a dataset in which the total country score is calculated based on a mean of five indicators. Here the dependent variable is total score and all indicators are the independent variables. Since, we have total scores of ten countries, our data set is very small to run t-test. So, we decided to increase the dataset to 30 observations by randomly selecting each observation until we have 30 observations in a sample. Although the new sample passed all the linearity assumptions for multilinear regression the regression outcome failed to uniquely determine the regression coefficient for independent variables. Thus, if we are interested to know which indicators contribute more weightage to the total country score, it will be difficult to answer if the total country score is measured in terms of the mean of the indicators. To avoid this constraint of the dataset, we decided to calculate the total country score as the median of the indicators. Again we increase our dataset to 30 observations through random selection. The dataset tested for the linearity assumption and found no strong evidence that violates the linearity

assumptions. At this time we can uniquely identify the unique regression coefficients and decided to take the dataset with median total country score for further analysis.

Classification and Clustering of data

In continuation to the above-mentioned methodology and increasing the dataset to include 30 observations, we wanted to run regression of both classification and clustering in order to predict the future outcome based on the current dataset. Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Classification is an example of pattern recognition. Examples are assigning a given email to the "spam" or "non-spam" class, and assigning a diagnosis to a given patient based on observed characteristics of the patient (sex, blood pressure, presence or absence of certain symptoms, etc.).Using this classification methodology we aim to provide an approximate score for the country which has not been scored yet. This helps us in estimating the counties JDI score although it doesn't have the data about the indices. This is done solely by classifying the country into an existing classification and based on the group mean assign the predicted value. Classification and clustering are examples of the more general problem of pattern recognition, which is the assignment of some sort of output value to a given input value. The same can be done with the help of clustering which is an unsupervised learning part of the clustering. Clustering identifies the hidden patterns in the data. In this JDI development procedure, we have a few indices and sub-indices which will be made the independent variables and the overall JDI score of the country as the dependent variable. This helps us in the prediction of the overall score of the country in spite of having some missing numbers in the indices or the sub-indices. In the process of achieving this we had inspected various machine learning algorithms that can fit and model our current dataset. After running through a number of such models, we could achieve a superior accuracy score for the models called as the Logistic regression model, K-Nearest neighbor clustering model and hierarchical clustering model.

Conclusion

The regression model shows that the available data can be used to develop the JDI index for countries. The choice of dataset whether the total country score is calculated by taking the ‘mean’ or ‘median’ of the indicators solely depends upon the interest of the researcher and the problem

being addressed. If the researcher is interested in uniquely identifying the weightage of the different indicators in total country score, then taking the dataset having total country score as a median of the indicators is suggested.

The models developed above does include the accuracy score to support a strong mechanism and understand any future data outside of the current dataset and predict its outcome. Since our current data set is limited to just a few observations, any future data that might have many observations can equally reduce the accuracy of the predictions which is very common in the real world models of this kind.

Table of Contents

Introduction

Jewelry Development Impact (JDI) index is a comparative country scoring system being developed within the framework of UN Indicators of Human Security. The aim of Data Validation and prediction is to provide a general scoring platform for the global jewelry sector to facilitate comparative socio-economic performance analysis of different countries by using the data that is already present from a few countries. Overall status of the sustainable performance of the jewelry sector is reflected in the cumulative score of the index where higher country score represents better performance of the sector. This research works aims to validate data frame collected and developed by the students of American University and develop the analytical method which will help in analyzing the current score and predict the score of the other countries.

Literature Review

JDI (Jewelry Development Impact) Index

Jewelry Development Impact (JDI) index is a comparative country scoring system being developed within the framework of UN Indicators of Human Security. Overall status of the sustainable performance of the jewelry sector is reflected in the cumulative score of the index where higher country score represents better performance of the sector.

Framework

JDI index score includes the following components and their sub-components.

1. Governance

Sectors cannot achieve sustainable development if there is no proper transparency and accountability. Regulator, industry representative and labor class are three major stakeholders in the mining and jewelry industry and any kind of gap in the communication or information flow between them hinders smooth functioning of the regulatory body. Therefore, trust and Ethics need to be integrated in the governance of the mining, production and distribution of the jewelry products. They also need to be integrated with the regulation making and implementation process to ensure equal participation of different levels of stakeholders. [1]

2. Economy

Mining industry has significant influence on multiple fields, such as income, employment, and education level. To ensure sustainability and ethical practices different economic tools were applied in the mining sector which had variable influences on different socio-economic conditions of the countries. It is one of the biggest challenges to integrate economic wellbeing and sustainability in an ethical way. [1]

COVID-19 is having a direct impact on the livelihoods of ASM communities worldwide. Artisanal and small-scale mining (ASM) directly provides the livelihood to over 40 million people worldwide and tens of millions more family members and small business owners are reliant on the sector for their income. Strong social distancing and quarantine measures that will be in force for at least two more weeks, though it may be extended over several months. These measures have exposed the vulnerability of a large part of the population that earns their livelihood from the informal economy and has no guarantees against the interruption of their activities. [2]

Ultimately, as miners could not sell their production, many simply abandoned the artisanal mining sector, returning to their place of origin, or moved to agricultural activities in the area. Other miners, working in the extraction of cassiterite and wolframite, moved to gold or coltan mining sites, for which the selling price is higher.

3. Environment

Mining and further processing of jewelry making will emits toxic gases as a byproduct. Other negative influence will also occur with mining, for example, heavy metals are brought to the surface which get washed away with rain or natural water flow and enters the natural water body causing alerted pH and metal level. To sum up, raw material extraction as well as manufacturing stages of the jewelry have direct and indirect environmental consequence some of which have long term impacts upon environment. [1]

Environmental aspect is one of the most important aspects to be considered in the context of the extractive industry. It is a bitter truth that mining industries had caused environmental harm and has led to tremendous suspicion toward future projects. [3]

ASM plays an important role in a country’s economy, however the negative environmental and socio-economic effects on the host communities associated with it can overshadow these

economic gains [4]. Though ASM has been a source of employment and income-generation, there is an environmental burden due to land degradation and impact on health [5] (Yelpaala & Ali, 2005). Artisanal Small-Scale Mining (ASM) comes with attendant environmental impacts which affects humans, animals, climates and environment. Like any other mineral resource extraction, the artisanal mining method is associated with human intervention in the environment. The anthropogenic influence on environmental topography is an integral part of raw material extraction anywhere in the world. However, in developing countries like the Great Lake region (area surrounding seven large lakes and in the basin of two large rivers—the Congo and the Nile, comprising countries like Uganda, Kenya, Tanzania, Burundi, Rwanda), where there is less adherence to mandated practices and rules, the impact of mining on the landscape structure is even more significant.

ASM is characterized by a number of conditions, which include: Lack or much reduced degree of mechanization, great amount of physically demanding work, low level of occupational safety and health care, deficient qualification of the personnel on all level of the operation, inefficiency in the exploitation and processing of the mineral production (low recovery of values), exploitation of marginal and/or very small deposits, which are not economically exploitable by mechanized mining, low level of productivity, low level of salaries and income, periodical operation by local peasants or according to the market price development, lack of social security, insufficient consideration of environmental issues, chronically lack of working and investment capital, mostly working without legal mining titles

As claimed by the International Labor Office (ILO) report on artisanal mining, unlike conventional industrial mining, ASM is illegal in many cases, and thus degrades the environment [6]. Widespread environmental issues such as heavy metal pollution, indiscriminate vegetation removal and the destruction of farmlands, sedimentation of rivers, improper handling of waste, abandonment of excavated pits, and a lack of reclamation. These environmental issues are believed to exacerbate the socioeconomic conditions of the people living in these mining-affected communities.

Millions of people in the developing world depend on artisanal and small-scale mining (ASM) for their livelihoods. The wealth generated, most times, comes at a price with many associated environmental and occupational health issues, particularly when practiced informally or with

limited technical and material resources. The health and well-being of miners, their family members as well as nearby communities is often adversely affected. The Environmental component consists of sub-components like environmental regulation, pollution, biodiversity and post-production planning and remediation. The American University students have conducted a series of research work to collect information on the sub-components implementing a series of research questionnaires.

Mining and further processing of jewelry making will emit toxic gases as a byproduct. Other negative influences will also occur with mining, for example, heavy metals are brought to the surface which get washed away with rain or natural water flow and enter the natural water body causing altered pH and metal level. To sum up, raw material extraction as well as manufacturing stages of the jewelry have direct and indirect environmental consequences some of which have long term impacts upon the environment.

4. Human Health

To ensure a healthy workplace is of vital importance for sustainable operation of the mining and jewelry industry. Heavy machineries and harsh workplace environment result in ergonomic hazards to both men and women. Dust and noise are common issues in the mining operation. Low living standard and unhealthy workplace conditions cause Tuberculosis, HIV, silicosis, and respiratory diseases, exposure to toxic metals causes digestive system and colon cancers, and inhaling Cadmium fume causes decreased bone density and peripheral neuropathy. [1]

ASM presents physical hazards to workers in all kinds of aspects, such as physical risks from poorly constructed pits/shafts/tunnels prone to collapsed/landslides/flooding/lack of ventilation, and inappropriate use and maintenance of mechanical equipment. Diseases can be caused by the use of hazardous materials (ex: mercury, cyanide), poor waste management leading to water contamination, lack of potable water/latrines/sanitation facilities leading to gastrointestinal and other diseases, and lack of PPE (personal protective equipment) or training in proper use leading to silicosis and other health risks. There are also impacts related to dust/noise/exhaustive labor. [7]

5. Human Rights

Ethical practice of human right is a significant indicator of the healthy status of the jewelry sector around the world. Mining and jewelry industries have often been criticized for direct and

indirect inhuman practices. Poor rural and indigenous people are often exploited by rich and politically backed up miners, due to lack of awareness of their legal right. Use of heavy instruments and harsh workplace condition pose serious health hazards, resulting in a large number of workplace deaths. [1]

African nations are the major source of colored gemstones. Gemstone reserves are small and distributed in many locations. Artisanal and Small- Scale Mining operations are used to mine gemstones as operation of large -scale mining is not economically feasible. Rough- gemstone is smuggled mostly to Asian nations where it is processed further for adding value and then supplied around the globe. The people of African nations where extraction of rough gemstone takes place are not able to harness the full benefit from its trade as it's beneficiation takes place in other countries. Predominance of the illegal supply chain and its control by criminal groups is the obstacle for the economic prosperity of the region [8] (Hunter & Lawson, 2020). In addition, this has become a financial source for organized criminal activities, and armed conflict in the African continent.

Countries with major global suppliers of gemstone e.g. Thailand do not report the accurate data on gemstone transactions. Supply chain of gemstone is informal and illicit in nature and difficult to track the source of uncut gemstone. There is a need of research to understand the overall impact of the gemstones and other mining activities in the economy. Integration of this informal economy in the formal process will help in reduction of criminal activities and will be a backbone in sustainable development of the region.

Women and girls are the victims of human rights abuses in nations where conflict minerals are a source of funding for armed conflicts. A study done in Democratic Republic of Congo shows that the sexual violence was linked to exploitation and trade in conflict minerals (Buss, 2018) [9]. Research also shows that the prevalence of sexual violence is higher where the women and girls are living near artisanal and small-scale mining (Rustad et al., 2016) [10]. Incorporation of the risk factors associated with basic rights of women, children, and indigenous people is pivotal for measuring the overall human right condition of the nation in the form of index.

The methodologies adopted for development of JDI Risk Assessment Index considers the different elements of human rights. A case study done by the American University Students shows that they were using survey research methodology for collecting the data from the relevant

stakeholders. They have used qualitative measurement techniques which measure the level of risk the people of nations are facing due to the mining and trade activities related with Jewelry.

The Human Right component consists of sub-components like worker's rights, indigenous people's rights, women's rights, children's rights, freedom from violence. The American University students have conducted a series of research work to collect information on the sub-components implementing a series of research questionnaires.

(1) Diamond: Diamond is produced with exploitative labor (child labor, forced labor, forced child labor) in six countries namely Angola, Central African Republic, Democratic Republic of Congo, Guinea, Liberia, and Sierra Leone. Sierra Leone is the only country where forced child labor is used in diamond mining.

(2) Gold: Gold is produced with exploitative labor in 22 countries namely Bolivia, Burkina Faso, Colombia, Democratic Republic of Congo, Ecuador, Ethiopia, Ghana, Guinea, Indonesia, Mali, Mongolia, Nicaragua, Niger, Nigeria, North Korea, Peru, Philippines, Senegal, Sudan, Suriname, Tanzania, and Uganda. Amidst these countries Burkina Faso, and Democratic Republic of Congo engage in forced child gold mining

(3) Gems: India and Zambia use child labor in gems mining.

(4) Jade: Burma engages in child labor and forced child labor to produce Jade.

(5) Emeralds: Colombia use child labor in Emeralds mining.

(6) Sapphires: Madagascar use child labor in Sapphires mining.

Score Generating process

By Qualitative Risk Rating approach vulnerability of any sector or project is estimated and evaluated against standards. JDI is designed to be used as an evaluation approach of the jewelry sector of any country, state, or organization.

The score of each sub-component is obtained from questionnaires. The questionnaires are survey questions about risk assessment. The final index is the mean of overall sub-scores.

Mining Industry Situation

The original data set that we used to perform further analysis is made up with the JDI indexes of multiple products from 10 countries, generated by different teams of American University students. Their findings are summarized below.

1. Gold in Peru

The great diversity and a lack of navigable topography reduce the government's authority and make it difficult to govern some remote mining areas effectively. Peru has performed well in transparency, at least at the central government level. Peru is still struggling with high levels of informality and illicit gold mining operations.

There are informal and illicit sectors in Peru. Evidence indicates there is concerning evidence of conflict between mineral-rich and non-mineral-rich regions, rising clientelism, and a lack of accountability and transparency of the use of funds.

Currently, Peru has not identified a sustainable solution to rectify the mercury and deforestation issues that illegal alluvial small-scale mining has caused.

The government struggles to contain both illegal gold miners and the by-product consequences of mercury drain-off into many of the Madre De Dios communities. As a result, high concentrations of mercury in fish species as well as the local populations.

Chronic safety hazards, manipulation, and abuse are found in Peru's illicit and informal gold sectors. Peru also suffers from a sever child labor issues and sexual slavery. Besides, there is a significant presence of criminal organizations conducting gold mining operations with severe social and environmental impacts. [11]

2. Diamonds in Botswana

Botswana has democratic governance and stability for a long time. Because of the secrecy and the lack of transparency, the mining industry has been critiqued. Botswana has become one of the most successful partnerships in the world with its diamond industry partners.

Botswana has potentially unsustainable indirect government employment, but there are high levels of enforcement and regulation of worker's rights since the diamond industry has participated

in training and capacity building for its employees. In addition, there are many examples of positive diamond industry efforts to provide lasting positive benefits to Botswana.

The health issues have a direct influence on the populations and the respective mining operations in general. Botswana has already implemented the proper environmental safeguards in order to protect its environment, decreasing the negative impacts of the diamond industry.

Although restricted by the constitutional prohibition on striking, workers in Botswana are still offered great benefits. There is sufficient evidence of women or children's rights abuses in Botswana's diamond sector, as well as criminal organization operations. [11]

3. Rubies in Myanmar

Modest efforts have been made by the government to reconcile its exclusionary mining policies to support smaller, more local, businesses and non-state actors who formed as a result of said policies and poor governance.

Although Myanmar's economy is steadily growing and the government is more focused on development activities, what weighs more is the armed forces engaged in a conflict that is fueled by extractives – conflict minerals – and does not address smuggling or finding diplomatic solutions to the conflict.

The state of the environment is weak-moderate, the state of human health is weak, and the state of human rights is moderate, according to the rating scores. [12]

4. Lapis Lazuli in Afghanistan

Afghanistan is currently lacking governmental accountability and governing power over both Afghanistan broadly and the lapis lazuli mines. The prevalence of corrupt and non-transparent government officials, gaps in basic data on mining legislation, and a strong governing presence of non-state actors and terrorist organizations, all indicate that governance is very weak.

The economic security related to lapis lazuli mining gets in trouble. Non-state actors and terrorist groups generate revenue from trading lapis, causing conflict across the country. There have been no attempts from the Afghan government to address smuggling, corruption, and control of the lapis lazuli mines by non-state actors and terrorists.

The state of the environment is weak-moderate, the state of human health is very weak, and the state of human rights is very weak-weak, according to the rating scores. [12]

5. Platinum in South Africa

Fairly strong formal institutions are in place to monitor the industry and violators of the rule of law are usually held accountable. Despite a good track record of transparency in general, corruption country wide and specific to the industry remains a problem. Meanwhile, the illegal mining industry involves highly organized criminal and terrorist organizations, according to reports.

Highly organized informal platinum mining industry brings the highest risk to South Africa's economy. South Africa loses opportunities for taxation from the informal economy, and people working informally are typically subject to risky work environments and unfair payment of wages. Besides, reports show that some terrorist organizations, international criminal syndicates, and national criminal syndicates have a presence in this informal industry. Overall, South Africa's low to moderate risk ensures that is does reap a significant economic benefit from hosting the platinum jewelry industry.

A lack of enforcing the regulations in place is threatening the environment. Consequently, platinum mining occurs in highly biodiverse areas which have high pollution levels in these areas that are quite significant to the damage of the environment.

Platinum mining contains high-risk work for the miners, even though the mining companies address the health and safety issues. Living conditions near the mines, combined with environmental damage done by the platinum mining industry, risks both food and water security to people.

The mining industry is highly formalized, operated by large international companies, and fairly strictly regulated. As a result, workers are generally well treated and comparatively well paid. [13]

6. Sapphires in Madagascar

This country has the institutions and laws in place to potentially ensure industry regulations and accountability, which are not well enforced by the government unfortunately. Madagascar needs to take further measures to reduce the risk that the sapphire industry has on its governance.

Malagasy exports the stones in their raw form, when they are least valuable, because of lacking the ability to add value to the sapphires they mine. This results in the underrepresentation of Malagasy sapphires on the global market and a lost opportunity to build the country's reputation, which has the potential to dramatically increase the value of the gems.

The implementation of the regulations, which has been established by the government to oversee mining processes and its impact on the environment, is limited. There is little oversight of the mining sights including protected areas, while remediation efforts are also not enforced by the Malagasy government.

Artisanal miners typically work on sites that are unsecured and without proper safety equipment, due to the predominantly informal nature of the industry. The artisanal and small-scale mining of sapphires contributes more to localized contamination related to sediment and human waste, than reducing the supplement of food and water.

The highest risk to the human rights of workers comes from the predominantly informal nature of the industry. Thus miners are hardly protected by health and sanitation standards. They also lack the ability to voice concerns through unions. [13]

7. Emerald in Colombia

The emerald industry operated in a grey space between legality and illegality is challenging to govern, in the territories outside the control of the government. Risks to governance are caused by the lack of state presence in certain parts of the country, informal mining, and the military strength of certain armed groups.

Some of the major environmental concerns surrounding emerald mining are soil erosion and degradation, water contamination, loss of biodiversity, and excessive water and energy consumption. The specific regulations concerning the closure and remediation of mines is also inadequate.

Illegal or small-scale-artisanal mining arises the most significant impact on Colombian mines. People are forced to use their bare hands, increasing the risk of injury and illness.

Many miners are not formally employed and therefore they are not receiving a regular wage, because of the informal nature of emerald mining. Women are not always well received in some formal positions in the emerald mining industry, and child labor used in illegal emerald mining has been reported. [14]

8. Emerald in Zambia

Zambia is lacking the rule of law and strong governance institutions are evident when examining the economic, health, and social state of the country, which are magnified with the emerald mining industry in the country. Many of the reforms have included initiatives for transparent policymaking. However, the efforts have allegedly been continuously undermined by corruption.

Capital flight is the most salient risk to the Zambian economy that the emerald industry posits. Some mines in Zambia operating without licenses smuggle emeralds in their raw, driving the price on the market to fall.

The mining-induced environmental impacts, such as emissions of sulfur-dioxide, waste landfills and tailings, and lead and cadmium poisoning, are evident in communities near Zambian mines.

Injuries are common due to unsafe working conditions among artisanal miners. Additionally, chemical waste runoff generated by mining activities can contain high concentrations of heavy metals that contaminate soil, crops, ground water, and surface water. Poor sanitation or improper waste management at mining sites can increase the spread of disease, especially diarrheal diseases such as cholera.

There is a minimum wage for employees and regulates working conditions set by law, but many workers still consider their wages low. Workers' authorities do not adequately enforce legal protections. The Zambian government also does not investigate or prosecute companies for labor trafficking in the mining sector. The capacity to monitor these issues is inadequate, too. [14]

9. Amethyst in Brazil

Brazil government gives mining rights priority to the first party who applies for the mining rights, and also issues exploration permits. Additionally, Brazil laid the open government initiatives groundwork “for more ambitious actions” through these years.

As for the impact on environment, forests are often cut down or burned by miners for the access to gems, while abandoned Amethyst mines can also cause accidents. Debris from the mines and the miners’ waste contaminate the soil and streams, even kill the vegetation and wildlife.

Health risks differ based on individual miners. Injury and death occur more often in informal mines than formal ones, due to poor tunnel systems. However, chronic health problems, such as silicosis and pneumoconiosis, is common in both kinds of the mines.

Many of the laws which ensure the mining industry respects workers’ rights are not enforced in amethyst mines. Women usually participate in lower-wage positions in the amethyst mining industry, compared to men. Only about 20% of the amethyst mining industry is formal. Therefore, most miners work under informal contract arrangements, without any benefits like social security or retirement plans. [15]

10. Tanzanite in Tanzania

There are formal institutions that monitor Tanzania’s mining industry, as well as the Commissioner of Minerals and the Zonal Mining Officers that provide official administration. However, transparency remains low. Smuggling of raw tanzanite, the major concern, undercuts potential government revenue severely.

There is a general lack of monitoring pollution levels in and around the mines. Also, the lack of environmental regulation is on the ground, especially for artisanal and small-scale miners. Air pollution is significant, and huge risk of land degradation exists in the small-scale mining area. Besides, the unorganized shafts could result in land degradation and mine disasters.

The rates of mining injuries in Tanzania is one of the highest in the world. Harsh working conditions, pair with almost non-existent health care services, leads to acute and chronic health problems of miners. Tanzanite mining contributes to local water pollution, because the waste rock in piles left by artisanal miners can contaminate river water which is already unclean.

During the past, many artisanal and small-scale miners worked under other people's mining leases, so it was harder for the government to regulate the industry. Women benefit less from tanzanite mining than men, and child labor is used widely in tanzanite mines. [15]

Methodology

We are trying to validate data and develop analytical techniques to predict the JDI-Score of the countries using the dataset collected and developed by the American University Students from their series of research works. In addition, this analysis aims to understand how different components are uniquely correlated with the JDI-Score. Furthermore, we will develop the analytical method which will help in analyzing the current score and predict the score of the other countries with insufficient data.

The available dataset consists of 10 observations (10 countries) and their JDI-Score calculated as a mean of 5 different indicators, Namely, Risk to Governance, Risk to Economy, Risk to Environment, Risk to Health, Risk to Human Right. We use two methods for the analysis. The first one is the regression analysis and the second one is Clustering.

Regression Analysis: Country JDI-Score as the Mean of the Indicators

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors, 'covariates', or 'features' (Wikipedia)).

In our analysis, first we use the dataset with JDI-Score as the mean value of indicators. Here the dependent variable is country level JDI-Score and five different indicators are the independent variables. The regression process is described as below.

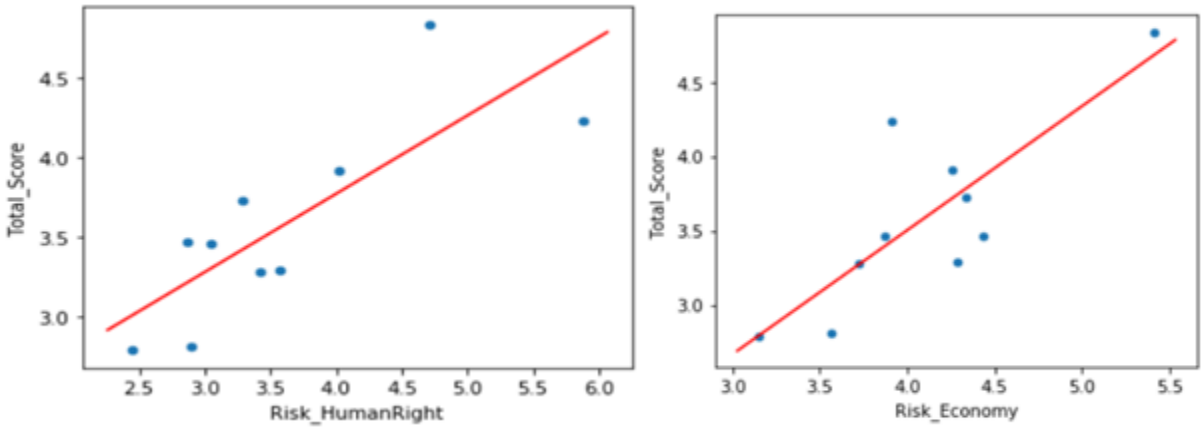
Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Governance	10.0	3.841585	0.967980	2.3920	3.0100	3.860444	4.699000	4.962963
Economy	10.0	4.099600	0.563448	3.3400	3.7900	4.006000	4.276000	5.404000
Environment	10.0	3.154892	0.763875	2.3680	2.6650	3.088000	3.187000	4.984000
Health	10.0	3.214356	0.511039	2.6320	2.8210	3.125778	3.517000	4.264000
Human_Rights	10.0	3.843262	1.058901	2.4880	3.1270	3.874000	3.971154	6.244000
Total_score	10.0	3.630739	0.603675	2.8816	3.3346	3.544375	3.800379	4.842400

Fig1: Descriptive Statistics of dataset having total score as the mean of the indicators

The above table shows that indicator Risk to Economy has the highest mean value 4.099 followed by Risk to Human right and Governance with mean value 3.84, followed by Risk to human health and Risk to environment with mean score of 3.21 and 3.15 respectively.

Identification of Outlier



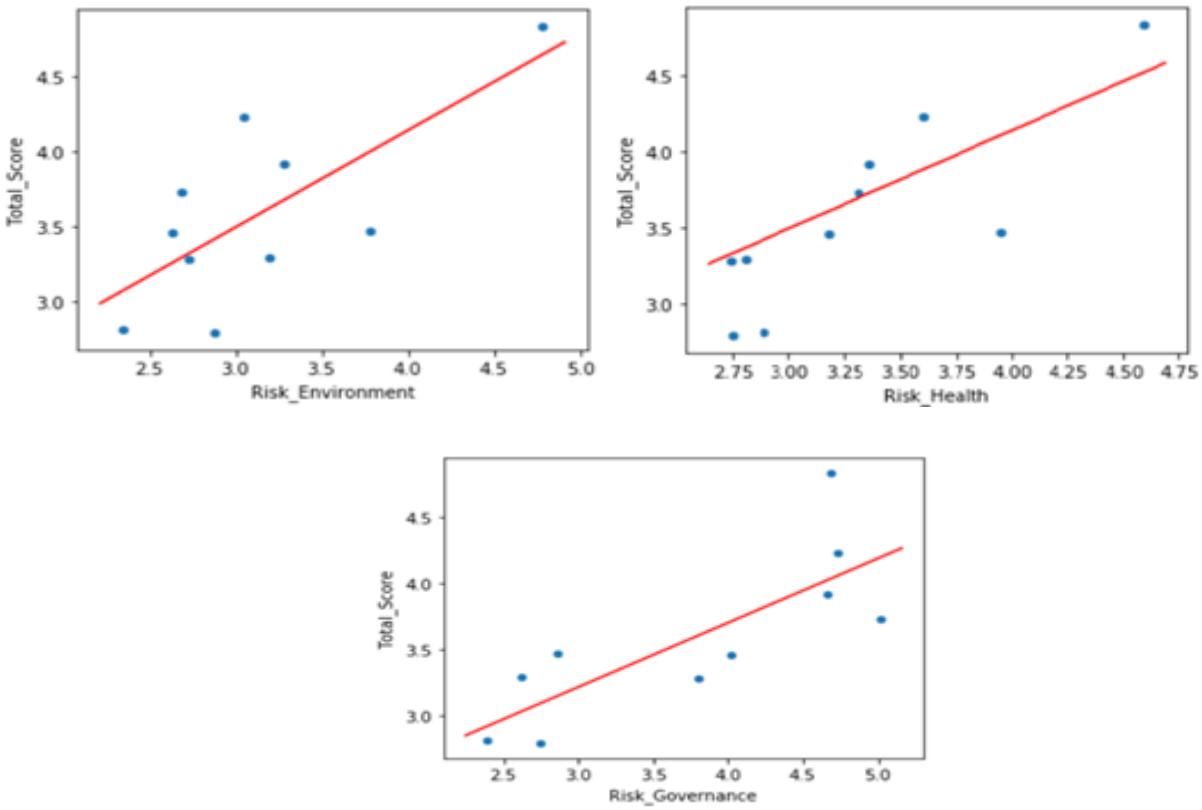


Fig 2: Scatter plot of total score against each indicator

The scatter plot shows at least we have one outlier. However, further diagnosis for potential outliers like studentized residuals, leverage value and influence statistics like DFFITS and Cook's D statistics shows there is no outlier. Therefore, we can conclude that the dataset does not have any outlier.

Multicollinearity Test

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	Intercept	1	1.43496E-14	0	Infty	<.0001	.	0
Human_Rights	Human_Rights	1	0.20000	0	Infty	<.0001	0.48270	2.07167
Governance	Governance	1	0.20000	0	Infty	<.0001	0.45865	2.18030
Health	Health	1	0.20000	0	Infty	<.0001	0.24479	4.08509
Economy	Economy	1	0.20000	0	Infty	<.0001	0.51679	1.93503
Environment	Environment	1	0.20000	0	Infty	<.0001	0.46084	2.16996

Fig 3: Multicollinearity test

The multicollinearity test table shows that the variance inflation is less than 10 for all indicators and the tolerance value is greater than 0.1. We can conclude that there is no multicollinearity.

Test for Normality of Residual

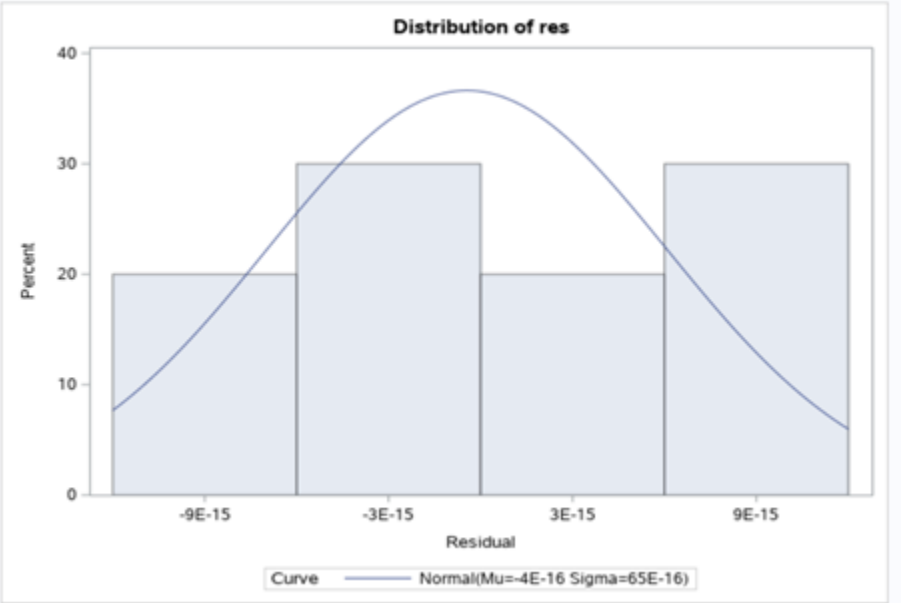


Fig 4: Test for Normality of Residual (Histogram)

Fig 4 shows that histogram is symmetric, there is no evidence of fat-tail.

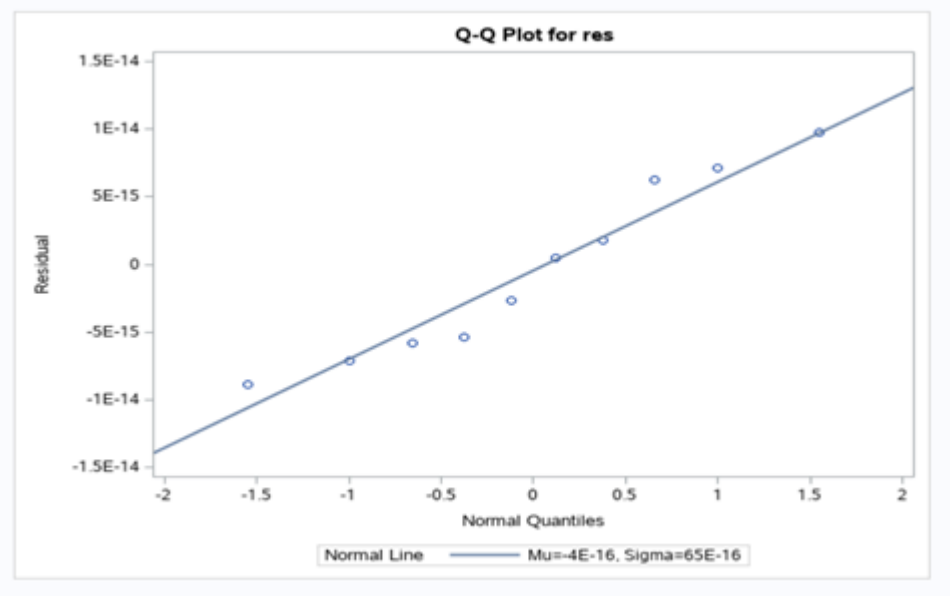


Fig 5: Test for Normality of Residual (Q-Q plot)

The Q-Q plot shows the residuals are almost normally distributed. We can conclude that residuals are normally distributed.

Test for Heteroscedasticity

We were not able to run the homoscedasticity test. This might be due to the limited number of observations in the dataset.

Multiple Linear Regression

Root MSE	0	R-Square	1.0000
Dependent Mean	3.63074	Adj R-Sq	1.0000
Coeff Var	0		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.43496E-14	0	Infty	<.0001
Human_Rights	Human_Rights	1	0.20000	0	Infty	<.0001
Governance	Governance	1	0.20000	0	Infty	<.0001
Health	Health	1	0.20000	0	Infty	<.0001
Economy	Economy	1	0.20000	0	Infty	<.0001
Environment	Environment	1	0.20000	0	Infty	<.0001

Fig6: Output Table of Multiple Linear Regression analysis

The multiple linear regression shows that the regression model is perfectly fitted as the R-Square and Adj R-Square value is 1. The regression coefficients show that it is not uniquely determined. One-unit change in each indicator value will change 1-unit change in the total score.

Therefore, to overcome the limitation of this model we tried to increase the sample size to 30 observations. The new sample of dataset is created by randomly selecting the row of observations each time by giving each row of observations an equal probability of being selected each time in the sample.

Multiple regression Analysis in 30 observations sample

The regression diagnosis process described from 1.2 to 1.6 is repeated.

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Score Total_Score

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	15.92076	3.18415	Infty	<.0001
Error	24	0	0		
Corrected Total	29	15.92076			

Root MSE	0	R-Square	1.0000
Dependent Mean	3.71266	Adj R-Sq	1.0000
Coeff Var	0		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	4.51063E-14	0	Infty	<.0001
Risk_Governance	Risk_Governance	1	0.20000	0	Infty	<.0001
Risk_Economy	Risk_Economy	1	0.20000	0	Infty	<.0001
Risk_Environment	Risk_Environment	1	0.20000	0	Infty	<.0001
Risk_Health	Risk_Health	1	0.20000	0	Infty	<.0001
Risk_HumanRight	Risk_HumanRight	1	0.20000	0	Infty	<.0001

Fig 7: Multiple Linear regression with 30 observations data sample

The regression analysis shows that we again cannot uniquely determine the coefficients for the independent variables and the model is perfectly fitted. The result is not similar with that of the 10 observations dataset.

Regression Analysis: Country JDI-Score as the Median of the Indicators

To overcome the limitation, we decided to calculate the new country JDI- score as a median value of 5 different indicators. And again, the same regression diagnosis process is repeated in the new

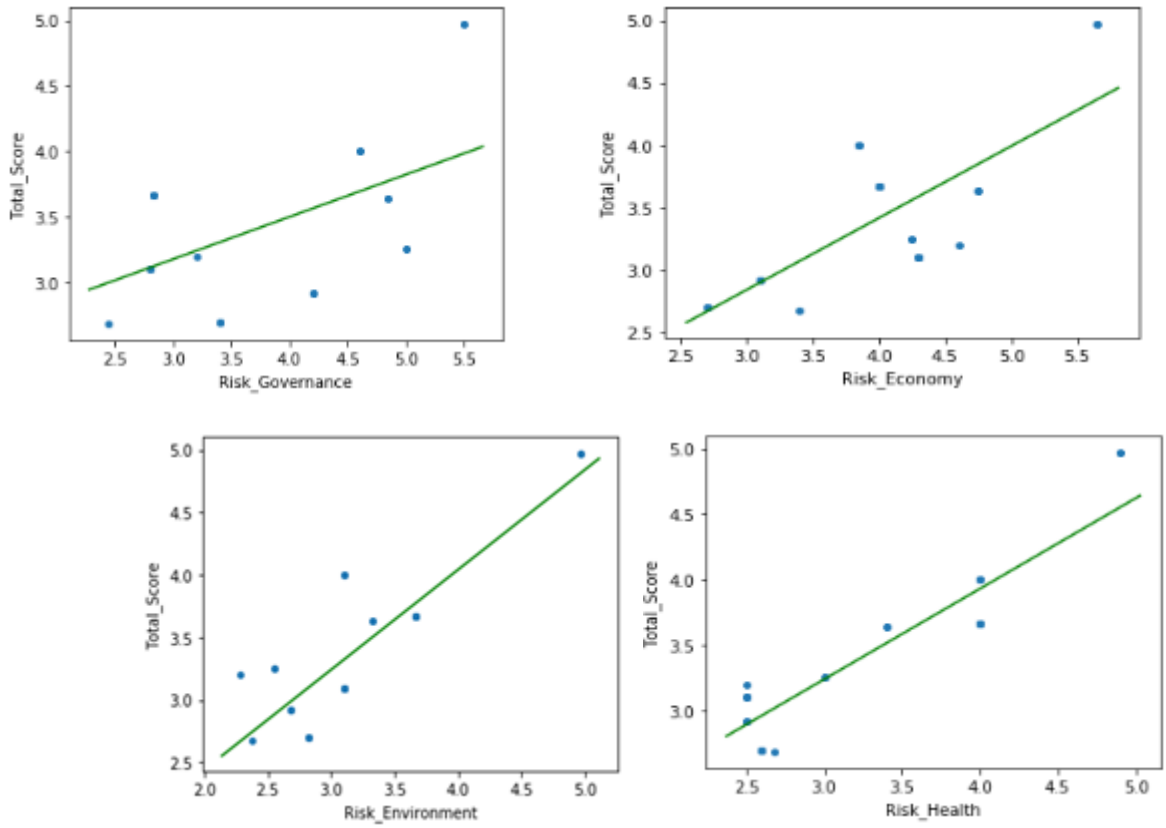
data set with 30 observations. The new sample of dataset is created by randomly selecting the row of observations each time by giving each row of observations an equal probability of being selected each time in the sample.

Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Risk_Governance	30.0	3.670000	0.968668	2.44	2.8300	3.400	4.60000	5.50
Risk_Economy	30.0	3.958333	0.769245	2.70	3.5125	4.000	4.30000	5.65
Risk_Environment	30.0	3.185500	0.627567	2.28	2.8250	3.100	3.57875	4.97
Risk_Health	30.0	3.222667	0.794268	2.50	2.5000	2.840	4.00000	4.90
Risk_HumanRight	30.0	3.429500	1.039282	2.34	2.8300	3.250	3.40000	6.04
Total_Score	30.0	3.391667	0.597087	2.68	2.9650	3.225	3.66500	4.97

Fig 8: Descriptive Statistics of Median Dataset

Identification of Outlier



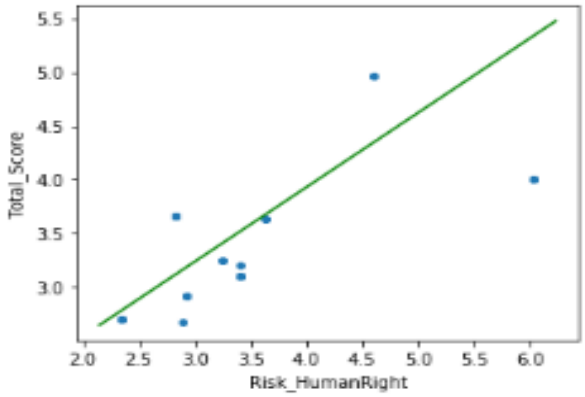


Fig 9: Scatter plot of total score against each indicator

The scatter plot shows at least we have one outlier.

Obs	country	Risk_Governance	Risk_Economy	Risk_Environment	Risk_Health	Risk_HumanRight	Total_Score
11	Botswana	5.50	5.65	4.97	4.9	4.600	4.97
25	Botswana	5.50	5.65	4.97	4.9	4.600	4.97

However, further diagnosis for potential outliers like studentized residuals, leverage value and influence statistics like DFFITS and Cook's D statistics shows there is no outlier. Therefore, we can conclude that the dataset does not have any outlier.

Multicollinearity Test

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	Intercept	1	0.19607	0.06891	2.83	0.0092	.	0
Risk_Governance	Risk_Governance	1	0.06488	0.01374	4.72	<.0001	0.68105	1.46833
Risk_Economy	Risk_Economy	1	0.17637	0.02044	8.63	<.0001	0.48773	2.05032
Risk_Environment	Risk_Environment	1	0.21973	0.03750	5.86	<.0001	0.21780	4.59140
Risk_Health	Risk_Health	1	0.35636	0.02866	12.43	<.0001	0.23273	4.29683
Risk_HumanRight	Risk_HumanRight	1	0.12012	0.01487	8.08	<.0001	0.50636	1.97479

Fig 10: Multicollinearity test output

The multicollinearity test table shows that the variance inflation is less than 10 for all indicators and the tolerance value is greater than 0.1. We can conclude that there is no multicollinearity.

Test for the Normality of Residual

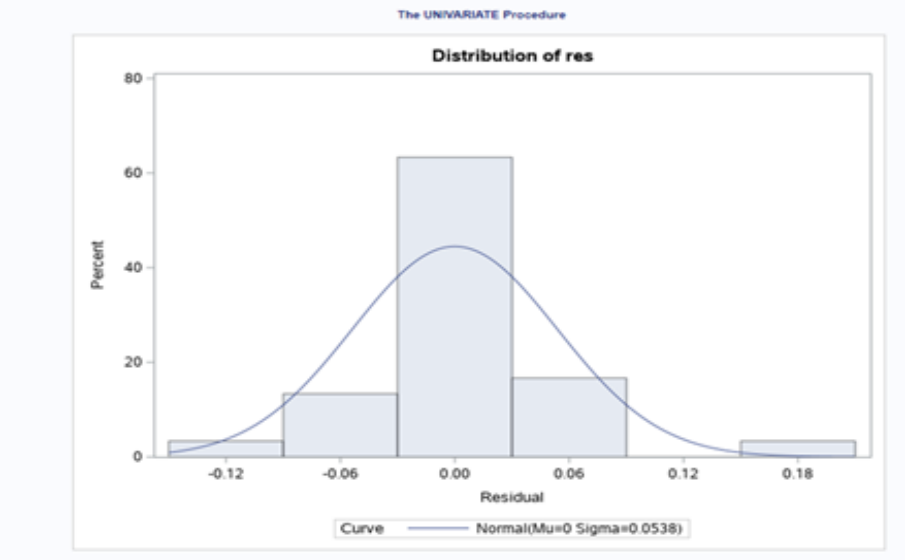


Fig 11: Test of Normality of Residual (Histogram)

The histogram is slightly skewed towards the right which is obvious given the nature of the dataset. However, it might not be a problem given the real-world data observations.

Test for Heteroscedasticity

One of the key assumptions of OLS regression is homoscedasticity.

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Score Total_Score

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
10	16.62	0.0832

Fig 12: White Test

White test shows that we fail to reject the null of homoscedasticity. So, we can conclude that the residuals have the constant variance.

Multiple Linear Regression JDI score (Median dataset with 30 observations)

Root MSE	0.05915	R-Square	0.9919
Dependent Mean	3.39167	Adj R-Sq	0.9902
Coeff Var	1.74391		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.19507	0.06891	2.83	0.0092
Risk_Governance	Risk_Governance	1	0.06488	0.01374	4.72	<.0001
Risk_Economy	Risk_Economy	1	0.17637	0.02044	8.63	<.0001
Risk_Environment	Risk_Environment	1	0.21973	0.03750	5.86	<.0001
Risk_Health	Risk_Health	1	0.35636	0.02866	12.43	<.0001
Risk_HumanRight	Risk_HumanRight	1	0.12012	0.01487	8.08	<.0001

fig 13: Output of regression analysis (JDI Score - Median dataset)

The regression analysis shows that the model is a good fit. All the indicators are significant and uniquely identified. Here, one-unit change is risk to health, risk to environment, Risk to economy, risk to human right and risk to governance will change 0.36, 0.22, 0.18, 0.12, and 0.06-unit change in the JDI-Score.

Limitation of regression analysis (Median dataset with 30 observations)

Though the difference between the mean and median score is not significant in this dataset. Cautious should be taken while calculating the total score when more data is available. The data set may skew if we increase the number of observations from smaller samples to the large sample. If we have sufficient real data points for the analysis, then increasing the data point is not suggested.

Classification

Classification in machine learning is considered as a task that learns how to assign a class label to examples from the problem domain. There are many different types of classification tasks that you may encounter in machine learning and specialized approaches to modeling that may be

used for each. Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).

For example, spam detection in email service providers can be identified as a classification problem. This is s binary classification since there are only 2 classes as spam and not spam. A classifier utilizes some training data to understand how given input variables relate to the class. In this case, known spam and non-spam emails have to be used as the training data. When the classifier is trained accurately, it can be used to detect an unknown email.

Classification belongs to the category of supervised learning where the targets also provided with the input data. There are many applications in classification in many domains such as in credit approval, medical diagnosis, target marketing etc. [16]

There are a lot of classification algorithms available now, but it is not possible to conclude which one is superior to another. It depends on the application and nature of the available data set. For example, if the classes are linearly separable, the linear classifiers like Logistic regression, Fisher's linear discriminant can outperform sophisticated models and vice versa.

From the dataset that we have we have considered the indices like Governance, Economy, environment, health, human right as the independent variable (X) and the total country score (y) as the dependent variable. The same can be expanded to include the sub-indices like Accountability mechanisms, Transparency etc. to include in the dependent variables.

By considering the above set of independent variables and dependent variables, we have inspected different models of classification in machine learning. Finally, we could achieve a considerable amount of accuracy with the model of Logistic Regression.

Logistic Regression for Classification

Logistic Regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent

variables. Sometimes logistic regressions are difficult to interpret; the Intellects Statistics tool easily allows you to conduct the analysis, then in plain English interprets the output.

The following are the steps included in applying the logistic regression to our existing dataset.

	Country	Governance	Economy	Environment	Health	Human_Rights	Total_score
0	Peru	3.832000	4.432	2.368000	3.256	3.028	3.380000
1	Afghanistan	2.752000	3.340	3.076000	2.752	2.488	2.881600
2	Columbia	4.600000	4.108	3.100000	3.196	3.880	3.776800
3	Madagascar	2.620000	4.288	3.196000	2.800	3.868	3.354400
4	Peru	3.832000	4.432	2.368000	3.256	3.028	3.383200
5	Columbia	4.600000	4.108	3.100000	3.196	3.880	3.776800
6	Afghanistan	2.752000	3.340	3.076000	2.752	2.488	2.881600
7	Brazil	3.888889	3.760	3.692308	3.700	4.000	3.808239
8	Zambia	3.784000	3.640	3.160000	2.632	3.424	3.328000
9	Botswana	4.852000	5.404	4.984000	4.264	4.708	4.842400
10	Madagascar	2.620000	4.288	3.196000	2.800	3.868	3.354400
11	Peru	3.832000	4.432	2.368000	3.256	3.028	3.380000
12	Myanmar	2.392000	3.880	2.536000	2.884	2.908	2.920000
13	Brazil	3.888889	3.760	3.692308	3.700	4.000	3.808239
14	Zambia	3.784000	3.640	3.160000	2.632	3.424	3.328000
15	Brazil	3.888889	3.760	3.692308	3.700	4.000	3.808239
16	Afghanistan	2.752000	3.340	3.076000	2.752	2.488	2.881600
17	Peru	3.832000	4.432	2.368000	3.256	3.028	3.380000
18	Myanmar	2.392000	3.880	2.536000	2.884	2.908	2.920000
19	Peru	3.832000	4.432	2.368000	3.256	3.028	3.380000
20	Madagascar	2.620000	4.288	3.196000	2.800	3.868	3.354400

Fig 14: Shows top 20 observations from the dataset

The above image shows the first 20 observations of the dataset. We have implemented all these analyses using a Python notebook. We have imported libraries like NumPy, Seaborn, pandas, matplotlib along with a few others.

Step 1: Once the dataset is loaded, we would like to describe the dataset to view the mean, median, standard deviation etc.

	Governance	Economy	Environment	Health	Human_Rights	Total_score
count	30.000000	30.000000	30.000000	30.00000	30.000000	30.000000
mean	3.632785	4.169600	3.287374	3.23200	3.581600	3.580245
std	0.861141	0.606779	0.786059	0.52395	0.690202	0.591327
min	2.392000	3.340000	2.368000	2.63200	2.488000	2.881600
25%	2.752000	3.760000	3.076000	2.80000	3.028000	3.334600
50%	3.832000	4.108000	3.130000	3.19600	3.868000	3.380000
75%	4.422222	4.432000	3.568231	3.58900	3.970000	3.800379
max	4.852000	5.404000	4.984000	4.26400	4.708000	4.842400

Fig 15: Descriptive analysis of the dataset

Step 2: Since logistic regression only takes the columns having only quantitative, we had to encode the countries to represent in numeric.

	Country	country_code
0	Peru	6
1	Afghanistan	0
2	Columbia	3
3	Madagascar	4
4	Peru	6
5	Columbia	3
6	Afghanistan	0
7	Brazil	2
8	Zambia	7
9	Botswana	1
10	Madagascar	4

Above picture shows the label encoding done to the country's column. The same label format will be used across the models moving forward.

Step 3: The immediate step after encoding is to split the dataset into training set (X) and test set (y). We have used 80% of our dataset to train the regressor and then test the rest 20% of the data with this regressor.

Step 4: After fitting the training set with the regressor, we have tried to fit and transform the dataset to predict the outcome. We could achieve 100% accuracy in this regard. This was possible only because we had a very small dataset. When extended to a real-world data the accuracy comes down which is expected to change.

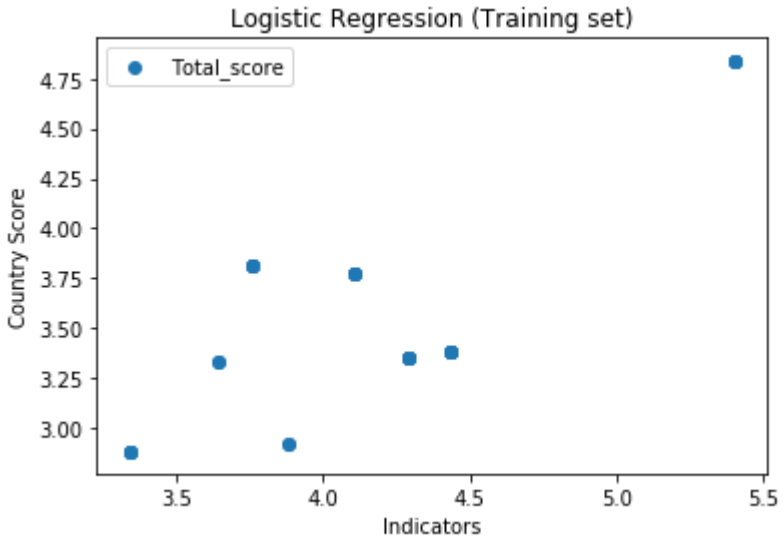


Fig 16: Scatter plot of the training dataset

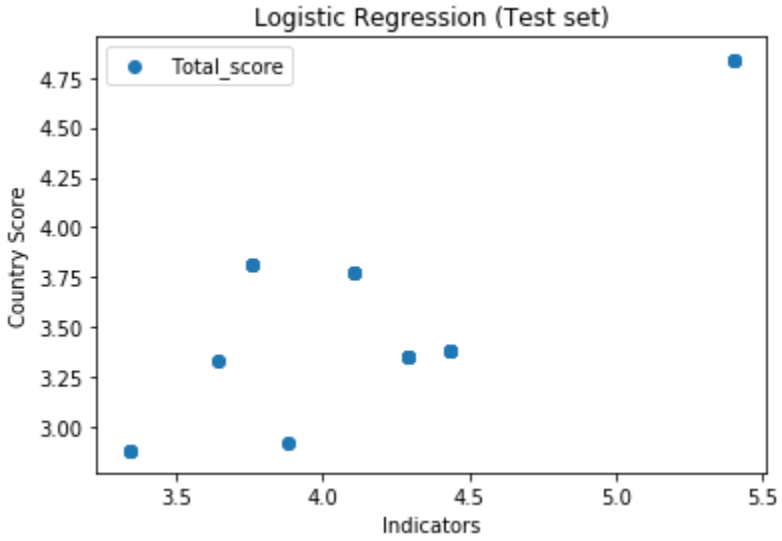


Fig 17: Scatter plot of the test dataset

The above two images show that the scatter plot of the training dataset and test dataset looks like mirror images, which says the accuracy is 100%.

Clustering

Clustering is a type of unsupervised learning. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labelled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

For example, the data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the below picture.

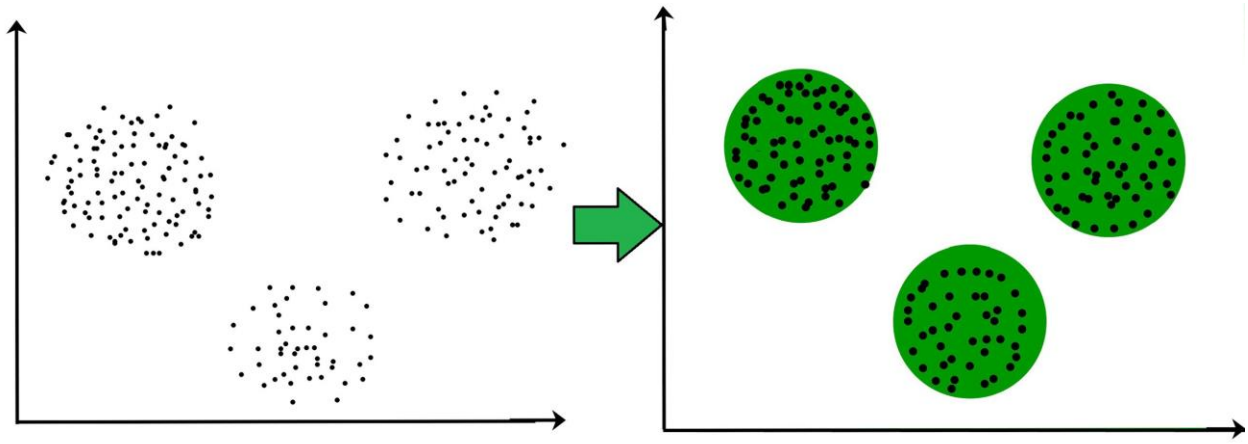


Fig 18: Sample showing the transforming clusters

It is not necessary for clusters to be spherical. Such as:

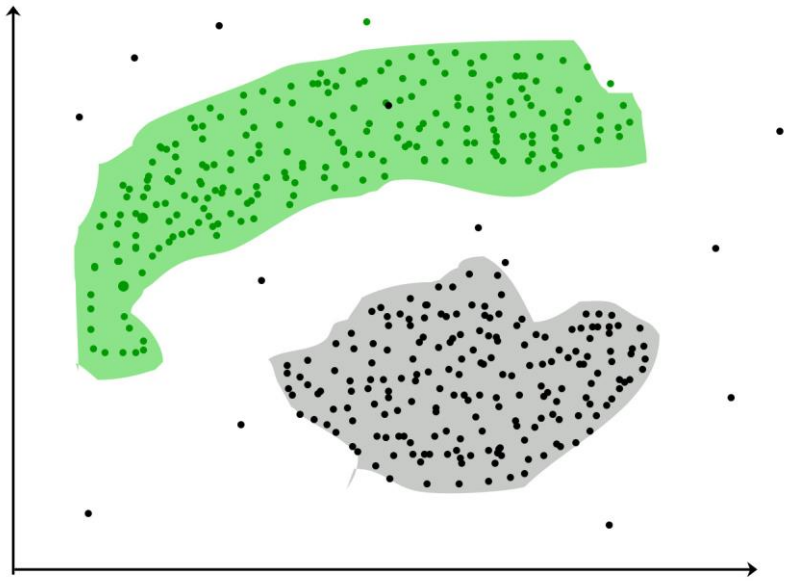


Fig 19: Examples of non-circular clusters

As an attempt to cluster the existing countries, we have implemented two techniques:

(1) K-means clustering

(2) Hierarchical clustering

K-means clustering

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes. A cluster refers to a collection of data points aggregated together because of certain similarities. You'll define a target number k, which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The 'means' in the K-means refers to averaging of the data; that is, finding the centroid. To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every

cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids. [17]

It halts creating and optimizing clusters when either:

(1) The centroids have stabilized — there is no change in their values because the clustering has been successful.

(2) The defined number of iterations has been achieved.

There are three main steps when using the K-means clustering technique:

Step 1: Use the elbow method, which can optimally define the number of clusters based on the dataset.

For our data, the optimal number of clusters was determined to be 6 by the elbow method.

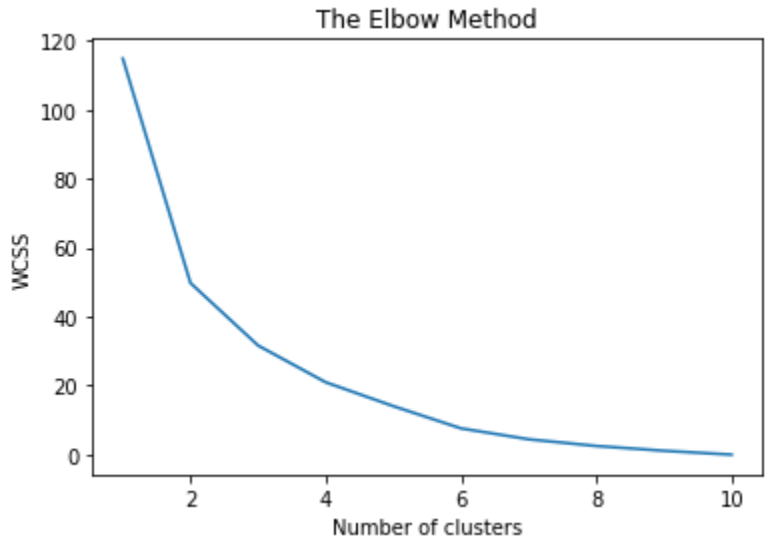


Fig 20: Graph of the WCSS and the number of clusters

Step 2: Once the optimal number of clusters is determined, we fit the training set with the k-means regressor and apply the regressor onto the test set to predict the accuracy.

Step 3: The final step is to plot the clusters onto a scatter plot to show the data points and their respective clusters.

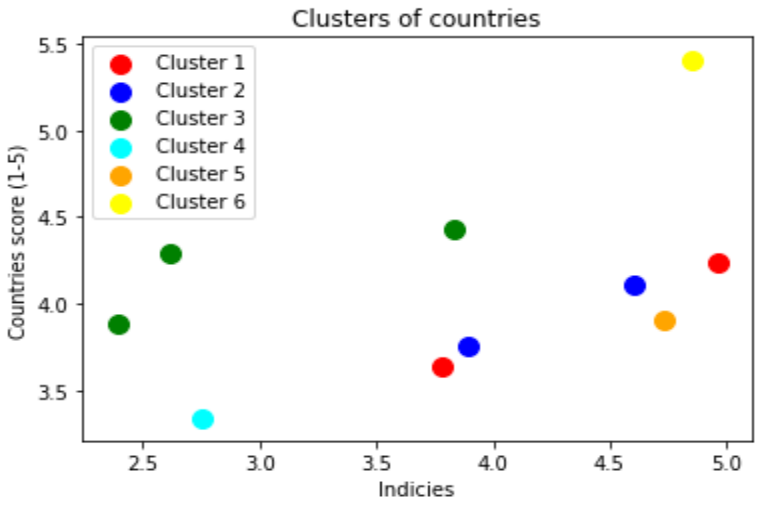


Fig 21: Scatter plot of the clustered countries versus scores

The 6 cluster that were formed by the existing 10 countries are as shown below:

- Cluster 1 - Tanzania and Zambia
- Cluster 2 - Brazil and Columbia
- Cluster 3 - Myanmar, Madagascar, and Peru
- Cluster 4 - Afghanistan
- Cluster 5 - South Africa
- Cluster 6 - Botswana

The accuracy of this model on the existing dataset was seen to be 100%. As explained in the previous model, the real-world accuracy of this model might as well come down, which is an expected scenario.

Hierarchical Clustering

Hierarchical Clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types [18]:

(1) Agglomerative: This is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

(2) Divisive: This is a "top-down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

The results of hierarchical clustering are usually presented in a dendrogram. In order to decide which clusters should be combined (for agglomerative), or where a cluster should be split (for divisive), a measure of dissimilarity between sets of observations is required. In most methods of hierarchical clustering, this is achieved by use of an appropriate metric (a measure of distance between pairs of observations), and a linkage criterion which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets.

In our analysis we have used the Agglomerative approach. In this technique, initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed. The basic algorithm of Agglomerative is straight forward:

(1) Compute the proximity matrix

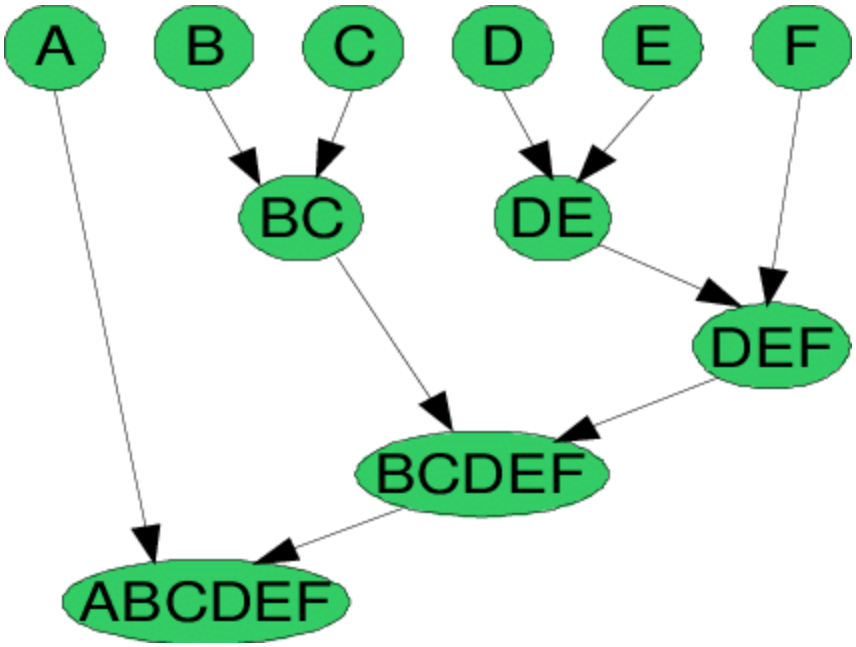
(2) Let each data point be a cluster

(3) Repeat: Merge the two closest clusters and update the proximity matrix

(4) Until only a single cluster remains

Key operation is the computation of the proximity of two clusters. To understand better let’s see a pictorial representation of the Agglomerative Hierarchical clustering Technique. Let's say we have six data points {A,B,C,D,E,F}.

Step 1: In the initial step, we calculate the proximity of individual points and consider all the six data points as individual clusters as shown in the image below.



Step 2: In step two, similar clusters are merged together and formed as a single cluster. Let's consider B,C, and D,E are similar clusters that are merged in step two. Now, we're left with four clusters which are A, BC, DE, F.

Step 3: We again calculate the proximity of new clusters and merge the similar clusters to form new clusters A, BC, DEF.

Step 4: Calculate the proximity of the new clusters. The clusters DEF and BC are similar and merged together to form a new cluster. We're now left with two clusters A, BCDEF.

Step 5: Finally, all the clusters are merged together and form a single cluster.

The hierarchical clustering technique can be visualized using a dendrogram.

A dendrogram is a tree-like diagram that records the sequences of merges or splits.

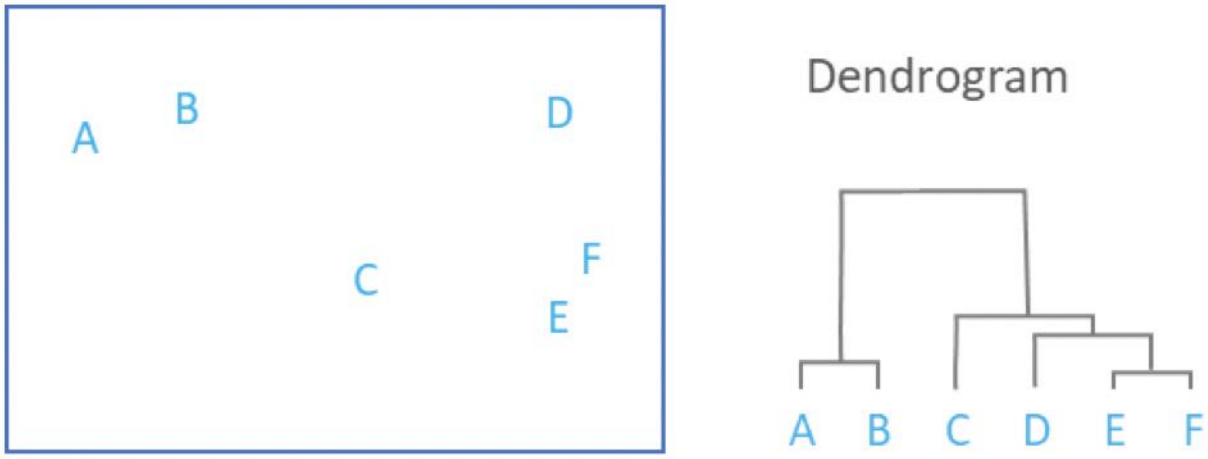


Fig 23: Example of Dendrogram and Scatterplot

After following the preprocessing steps like the K- means clustering and applying the regressor onto the test set, the following dendrogram was obtained.

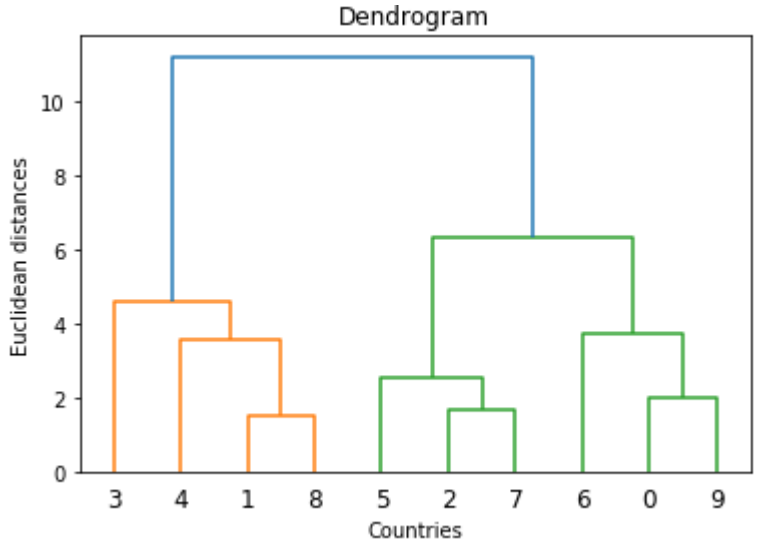


Fig 24: Dendrogram showing the countries clustered hierarchically

This dendrogram shows the countries clustered in clusters versus their Euclidean distances. Using these existing clusters, we can place a new country for which we need to determine the JDI score.

Limitations of Hierarchical clustering Technique

1. There is no mathematical objective for Hierarchical clustering.
2. All the approaches to calculate the similarity between clusters have its own disadvantages.
3. High space and time complexity for Hierarchical clustering. Hence this clustering algorithm cannot be used when we have huge data.

Conclusion

The regression model shows that the available data can be used to develop the JDI index for countries. The choice of dataset whether the total country score is calculated by taking the 'mean' or 'median' of the indicators solely depends upon the interest of the researcher and the problem being addressed. If the researcher is interested in uniquely identifying the weightage of the different indicators in total country score, then taking the dataset having total country score as a median of the indicators is suggested.

The analytical research done on the median dataset of the country scores, has several models of classification and clustering. This analysis has been 100% accurate with the above-mentioned dataset, when expanding this analysis to a larger dataset the accuracy is expected to come down. With these models we can predict the overall JDI scores of the countries with insufficient or missing data. Once the overall JDI score for a country is predicted, we can place the missing data in other indices either with the mean or the median scores.

References

[1] Shah, K., & Sharek, A., A Proposed Jewelry Development Impact (JDI) Index Methodology.

[2] Seccatore, J., Veiga, M., Origliasso C., Marin, T., deTomi, G. 2014. An estimation of the artisanal small-scale production of gold in the world. Science of the Total Environment. Pp. 662–667.

[3] Ali, S. H., &O'Faircheallaigh, C. (2005). Introduction: Extractive Industries, Environmental Performance and Corporate Social Responsibility. Greener Management International, 52, 5–16.

[4] 'Artisanal and small-scale gold mining and health Technical Paper #1: Environmental And Occupational Health Hazards Associated With Artisanal And Small-Scale Gold Mining' World Health Organization

[5] Yelpaala, K., & Ali, S. H. (2005). Multiple scales of diamond mining in Akwatia, Ghana: Addressing environmental and human development impact. Resources Policy, 30(3), 145–155. https://doi.org/10.1016/j.resourpol.2005.08.001

[6] https://www.mdpi.com/2071-1050/11/11/3027/htm

[7] https://en.wikipedia.org/wiki/Artisanal_mining#Health_and_safety

[8] Hunter, M., & Lawson, L.(2020).A-Rough-Cut-Trade-Africa's-Coloured-Gemstone-Flows-to-Asia-GITOC.pdf. Retrieved August 8, 2020, from https://globalinitiative.net/wp-content/uploads/2020/07/A-Rough-Cut-Trade-Africa%E2%80%99s-Coloured-Gemstone-Flows-to-Asia-GITOC.pdf

[9] Buss, D. (2018). Conflict Minerals and Sexual Violence in Central Africa: Troubling Research. Social Politics: International Studies in Gender, State & Society, 25(4), 545–567. https://doi.org/10.1093/sp/jxy03

[10] Rustad, S. A., Østby, G., & Nordås, R. (2016). Artisanal mining, conflict, and sexual violence in Eastern DRC. The Extractive Industries and Society, 3(2), 475–484. https://doi.org/10.1016/j.exis.2016.01.010

[11] “JEWELRY DEVELOPMENT IMPACT INDEX STUDY: A Comparative Case Study of Diamonds in Botswana and Gold in Peru”

[12] “JEWELRY DEVELOPMENT IMPACT INDEX STUDY: A Comparative Case Study of Rubies in Myanmar and Lapis Lazuli in Afghanistan”

[13] “JEWELRY DEVELOPMENT IMPACT INDEX STUDY: A Comparative Case Study of Platinum in South Africa and Sapphires in Madagascar”

[14] “JEWELRY DEVELOPMENT IMPACT INDEX STUDY: A Comparative Case Study of Emeralds in Colombia and Zambia”

[15] “JEWELRY DEVELOPMENT IMPACT INDEX STUDY: A Comparative Case Study of Amethyst in Brazil and Tanzanite in Tanzania”

[16] https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623

[17]https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[18] Hierarchical clustering,: https://en.wikipedia.org/wiki/Hierarchical_clustering

Contribution

1. Executive Summary: Tejaswini Ganti, Kamal Keshar Khanal
2. Introduction: Kamal Keshar Khanal
3. Literature Review:
   (1) JDI (Jewelry Development Impact) Index: Yahan Zhang, Kamal Keshar Khanal, Tejaswini Ganti
   (2) Mining Industry Situation: Yahan Zhang
4. Methodology:
   (1) Regression Analysis: Country JDI-Score as the Mean of the Indicators: Kamal Keshar Khanal
   (2) Regression Analysis: Country JDI-Score as the Mean of the Indicators: Kamal Keshar Khanal
   (3) Classification: Tejaswini Ganti
   (4) Clustering: Tejaswini Ganti
5. Conclusion: Tejaswini Ganti, Kamal Keshar Khanal
6. References: Kamal Keshar Khanal, Tejaswini Ganti, Yahan Zhang
7. Formatting and Finalization: Yahan Zhang, Kamal Keshar Khanal, Tejaswini Ganti