Sustainability Around The World

Aradhya Mathur
Data Science
University of Rochester
Rochester, United States
amath12@ur.rochester.edu

Ozlem Gunes
Data Science
University of Rochester
Rochester, United States
ogunes@ur.rochester.edu

***Abstract*— Investing in policies and programs which lead to achieving a climate neutral world is more important than ever. It is significant since providing clean drinking water and sanitation, reducing air pollution, and responding to public health crises yield large returns for human well-being. The world economy has a significant impact on human well-being. There are direct relations between environmental sustainability, the world economy, and social inclusion. In this research, these complex relations are studied to see the effect of economic, social, and other external factors on our physical environment and well-being. Using the time-series data, the effect of COVID on our environment is analyzed, and the pandemic favors sustainability. However, environmental sustainability performance has rebounded to pre-pandemic levels almost everywhere, as have many countries' air pollution, troubles in preserving biodiversity, and water pollution.**

Keywords—sustainability, world economy, time series analysis, covid

I. Introduction

Sustainable development is a serious challenge of our times. Climate change intimidates our world and our world is under a lot of strain. It is a world of extreme poverty and enormous wealth. Inequalities are increasing. These impacts are getting so enormous that the earth is undergoing irreversible changes. At this stage, sustainable development interprets interactions of human and environmental systems.

In the near future, we will be facing the challenges to make environmental neutrality secured, economic prosperity achieved; however, what are specifically these challenges first of all? How can we figure them out? Do we have solutions already or these problems are just being discussed and hung in the air?

First of all, environmental neutrality can be associated with wealth (GDP per capita), meaning that economic prosperity makes it possible for countries to invest in policies that lead to desirable outcomes. Second, the pursuit of economic prosperity - industrialization and urbanization - often means more pollution and other strains on ecosystem vitality, especially in the developing world, where air and water emissions remain significant. Meaning that, there are complex relations between the world economy, physical environment (climate change, environmental health, and ecosystem vitality) and social inclusion.

In this project, we will list the issues and factors associated with these relations such as air quality, waste management, climate change mitigation , biodiversity habitat index, GDP per capita, population, life expectancy, global health security index and so on.Analyzing these indicators of climate change performance, environmental health, and ecosystem vitality, we will try to explain the interaction between human and environmental systems.

We will also be doing time series analysis using linear regression models and ARIMA to determine whether or not the countries have addressed the environmental challenges they faced in past years and how they will.

II. Data

We have three datasets. First dataset "Fig1" is about the Environmental Performance Index and is taken from NASA Sedac[1] . Second database "Fig2" consists of external factors like GDP, Life Expectancy. Third database "Fig3" is a time series dataset [2] consisting of attributes for past and present years for all the countries.

A. Data Description

code	iso	country	region	EPI.new	HLT.new	AIR.new
4	AFG	Afghanista	Southern Asia	43.6	16	15.5
8	ALB	Albania	Eastern Europ	47.1	40	37.5
12	DZA	Algeria	Greater Middl	29.6	42	39.4
24	AGO	Angola	Sub-Saharan A	30.5	20.5	23.1
28	ATG	Antigua an	Latin America	52.4	55.8	56.5
32	ARG	Argentina	Latin America	41.1	56.3	52
51	ARM	Armenia	Former Soviet	48.3	40.7	32.1
36	AUS	Australia	Global West	60.1	86.4	91.1
40	AUT	Austria	Global West	66.5	81.7	75
31	AZE	Azerbaijan	Former Soviet	38.6	30.7	22.1

Fig 1. There are 57 attributes in our EPI dataset for 180 countries.

Country	Region	EPI	Population	GDP_PC	GHS_Inde	Life_Expe	Air
Afghanista	Southern A	43.6	41128771	489.1013	28.8	62.879	15.5
Albania	Eastern Eu	47.1	2842321	6424.342	45	76.833	37.5
Algeria	Greater M	29.6	44903225	3741.004	26.2	77.129	39.4
Angola	Sub-Sahara	30.5	35588987	2038.467	29.1	61.929	23.1
Antigua an	Latin Amer	52.4	98728	14900.8	30	79.236	56.5
Argentina	Latin Amer	41.1	45510318	10799.59	54.4	76.064	52
Armenia	Former So	48.3	2780469	4985.196	61.8	73.372	32.1
Australia	Global We	60.1	26177413	58930.95	71.1	83.579	91.1
Austria	Global We	69.4	8939617	53367.22	56.9	82.412	75
Azerbaijan	Former So	38.6	10358074	5273.391	34.7	73.488	22.1

Fig2. There are 15 attributes in our external factors dataset.

year	iso	country	region	EPI	AIR	H2O	BDH	WRS
2022	AFG	Afghanist	Southern	43,6	15,5	28,1	30,7	0
2022	ALB	Albania	Eastern E	47,1	37,5	54,1	63,9	1,9
2022	DZA	Algeria	Greater M	29,6	39,4	53,3	22,7	33,1
2022	AGO	Angola	Sub-Saha	30,5	23,1	12,8	30,1	0
2022	ATG	Antigua a	Latin Ame	52,4	56,5	50,1	54,2	15,7
2022	ARG	Argentina	Latin Ame	41,1	52	64,8	42,4	5,9
2022	ARM	Armenia	Former So	48,3	32,1	57,3	73,3	4,5
2022	AUS	Australia	Global We	60,1	91,1	87,1	82,1	92,9
2022	AUT	Austria	Global We	66,5	75	94,7	86	94
2022	AZE	Azerbaija	Former So	38,6	22,1	45,6	46,2	3,9

Fig3. Time series database

B. Glossary of Terms, Abbreviations and Acronyms

In this paper we will be using acronyms such as EPI, GDP, HLT and ECO. For clear understanding, glossary can be found in "Fig 4".

Term	Description
EPI	Environmental Performance Index
PCC	Climate Change Mitigation
HLT	Environmental Health
ECO	Ecosystem Vitality
GDP	Gross Domestic Product
GHS	Global Health Security
BHI	Biodiversity Health Index
ARIMA	AutoRegressive Integrated Moving Average.
CDA	CO2 growth rate
PMD	PM2.5 Exposure
HAD	Household Solid Fuels
VOE	VOC Exposure
COE	CO Exposure
SOE	SO2 Exposure
NOE	Nox Exposure
OZD	Ozone Exposure
AIR	Air Quality
H2O	Sanitation and Drinking Water
HMT	Heavy Metals
WMG	Waste Management
BDH	Biodiversity and Habitat
ECS	Ecosystems Services
FSH	Fisheries
ACD	Acid Rain
AGR	Agriculture
WRS	Water Resources

Fig4. Glossary

III. Background

A. *Literature Survey*

In this section we discuss the literature survey work which was carried out previously. Wang, Yang, Yin and Zhang [3] proposed a non-radial and non-oriented biennial generalized DDF to consider all types of factors and using these they successfully solved the infeasibility dilemma. They also examined the impact and mechanism of environmental regulation on Chinese green TFP from the dual decomposition viewpoint. They employed three Data Envelopment Analysis (DEA) models to accomplish a better, more robust estimation of efficiency. They developed a three-stage procedure to fit the time series model to the time series data. And then they developed a three-layer BPNN model to determine the forecasting model for the non-linear patterns; it was evident that the time series–ANN makes the prediction results more consistent. In the end they found out that the development capacity of the ecological economic system has a negative relation with ecological footprint. They also found a positive relation between ecological economic system and ecological footprint diversity.

Matsumoto, Makridou and Doumpos [4] have used Data Envelopment Analysis approach and global Malmquist-Lemberger index to evaluate the environmental performance of European Union countries. They used a DEA window approach that considered desirable and undesirable results and energy and non-energy inputs to evaluate the environmental performance. DEA window approach helped in eliminating the issue of reviewing efficiency over time. It also increased the discrimination capability and provided more accurate results. They also applied the GML index to measure performances of 27 countries. They also incorporated air pollutants into the analysis. This step was necessary to investigate the environmental performance. In the end they evaluated Economic–energy–environmental effects on environmental performance and found out that economic and environmental variables both significantly impacted performance.

Kanmani, Obringer, Rachunok and Nateghi [5] proposed a new data-driven framework to evaluate the environmental sustainability of countries accurately by using unsupervised learning theory. They used Self-Organized Maps to group countries on the basis of their characteristic environmental performance metrics and monitor development in terms of shifts within clusters over time. This new method helped countries make more informed decisions by understanding effective and detailed pathways towards improving their environmental sustainability. Self Organized maps represent underlying similarity between the respective environmental performance of countries. Results obtained by them demonstrate substantial inconsistencies between the EPI rankings and cluster association. The result indicated that the EPI might not be the ideal measure of environmental sustainability.

B. *Problem Statement*

We will be doing country and region wise analysis to determine trends, frequent patterns and some key information..We will spot the problems of each country in terms of environmental sustainability and how much they have addressed the environmental challenges they faced and how they will in the future.

We will also analyze the relation between environmental performance, GDP per capita, life expectancy and GHS Index and BHI Index for countries.

We will also be analyzing the impact of COVID on environmental sustainability by mining data from pre, during and post COVID times.

C. *Objectives*

- To determine relationships between EPI and external factors like GDP per capita, Life Expectancy and GHS Index.
- Analyze the impact of COVID-19 on EPI scores. Also, analyzing the post COVID-19 period.

IV. Exploratory Data Analysis

Exploratory data analysis is an important step and helps in determining trends, relations and insights within attributes. We will be doing country-region based analysis, binning and frequent pattern mining to generate insights.

A. *Country-Region Based Analysis*

From the analysis conducted below we can generate insights and confirm some pre known relations between

attributes. Heatmap and scatterplots are great methods to determine distributions (region wise), correlations and trends.

All the countries are divided into 8 regions, namely:

- Global West
- Eastern Europe
- Southern Asia
- Greater Middle East
- Sub-Saharan Africa
- Asia-Pacific
- Former Soviet States
- Latin America and Caribbean

Most of the countries in the Global West region have good EPI scores as well as PCC scores which is observed in "Fig5".

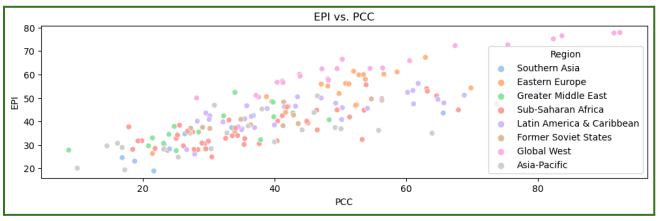


Fig5. EPI vs PCC

Climate change mitigation score is positively related with CDA and GHN (GHG emissions). It can be observed in "Fig6".

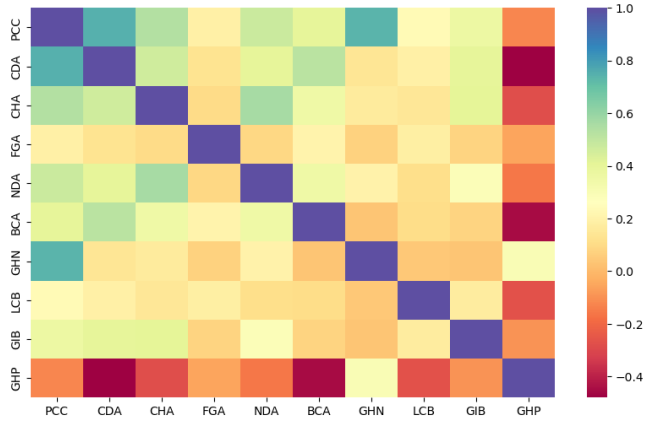


Fig6. Heat map for PCC and attributes

All the Global West countries have a great HLT score, followed by Eastern European countries. Countries in the Sub-Saharan African region have poor scores as observed in "Fig7.".

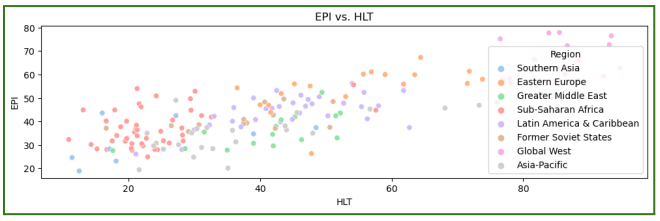


Fig7. EPI vs HLT

Environmental Health score is positively related to air quality, water quality and other attributes as shown in "Fig8". .

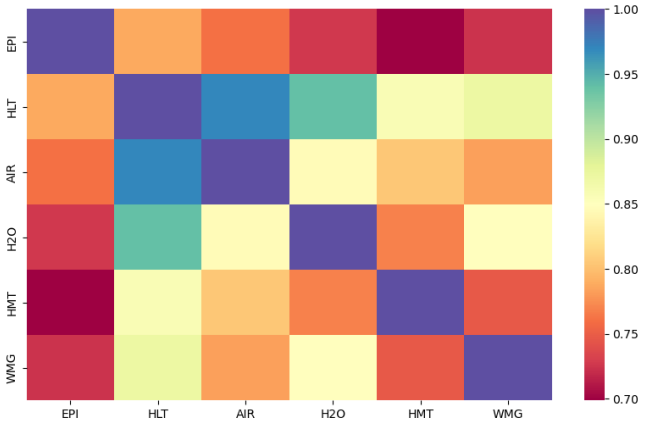


Fig8. Heat map for HLT and attributes

Similar observations can be made from the above figure. Countries lying in the Global West and Eastern European region have the best ECO scores.

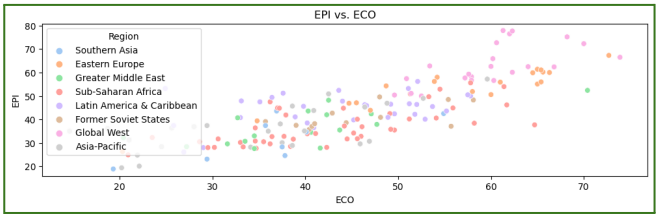


Fig9. EPI vs ECO

In "Fig10" we can observe that Ecosystem Vitality is positively related to water resources and acid rain.

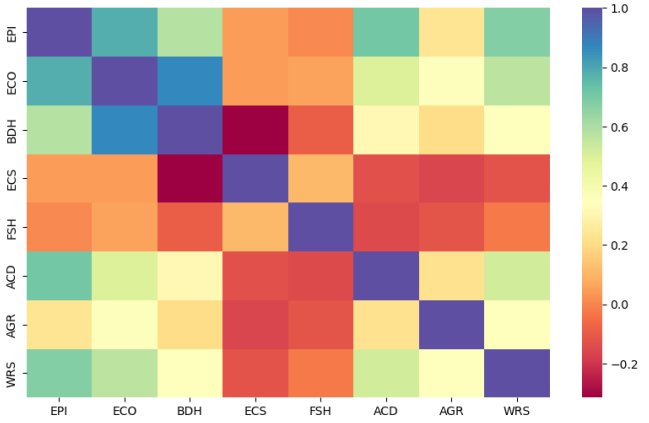


Fig10. Heat map for ECO and attributes

It can be inferred that countries lying in the Global West region have the best air quality scores. while countries lying in Southern Asian regions have poor scores

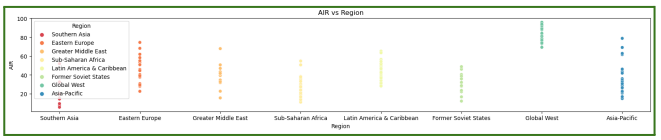


Fig11. Air Quality vs Regions

It can also be inferred that countries lying in the Global West, Asia-Pacific and Eastern Europe regions have better water quality scores. while countries lying in Sub-Saharan African regions have poor scores

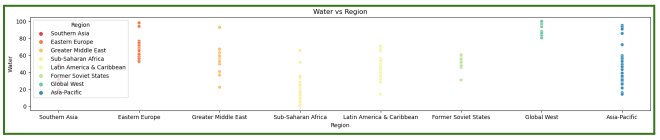


Fig12. Water Quality vs Region

B. *Binning*

It can be difficult to analyze numerical variables, it can be easily done after binning as numerical variables are converted to categorical variables.

Most of the countries lying in Eastern Europe and Global West are binned in above average, good and very good bins. Also, quite a few countries in the Asia-Pacific region lie in very poor bin which can be seen in "Fig13".

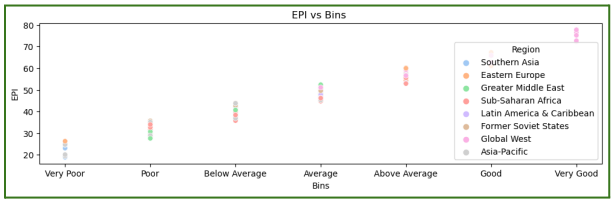


Fig13. EPI scores vs Bins

Countries lying in the Global West region were binned based on their HLT score. It was observed that Sweden, Finland, Norway and Iceland lie in the High HLT score bin.

	Country	Region	EPI	HLT	AIR	H2O	HMT	WMG	Binned
	Sweden	Global West	72.7	93.1	94.0	98.6	96.9	70.8	High
	Finland	Global West	76.5	93.4	93.5	100.0	100.0	69.6	High
	Norway	Global West	59.3	92.2	92.4	100.0	93.0	70.7	High
	Iceland	Global West	62.8	94.7	96.0	100.0	95.1	73.9	High

Fig14. HLT scores vs Bins

Similarly, Countries lying in the Global West region were binned based on their ECO score. It was observed that Luxembourg, Austria, Germany and Malta have the best ECO score among all the countries in that region.

	Country	Region	EPI	ECO	BDH	ECS	FSH	ACD	AGR	WRS	Binned
95	Luxembourg	Global West	72.3	70.0	84.8	18.1	0.0	100.0	55.9	98.0	High
8	Austria	Global West	66.5	73.9	86.0	28.0	10.4	100.0	70.6	94.0	High
62	Germany	Global West	62.4	66.8	88.5	17.9	26.9	100.0	60.9	97.0	High
101	Malta	Global West	75.2	68.2	72.9	100.0	47.8	100.0	28.3	0.0	High

Fig15. ECO scores vs Bins

C. *Frequent Pattern Mining*

Apriori algorithm is one of the most used frequent pattern mining techniques. Frequent patterns are mined based on regions, air qualities and other factors.
Dataset was converted from numerical scores to binned groups for every attribute which can be seen in "Fig16" .

The Apriori algorithm is easy to understand. The results are intuitive and easy to infer. This algorithm can be efficiently used here as the dataset is not large.

	Country	Region	AIR_BIN	HAD_BIN	PMD_BIN	OZD_BIN	NOE_BIN	SOE_BIN	COE_BIN	VOE_BIN
0	Afghanistan	Southern Asia	Poor AIR	Poor HAD	Poor PMD	Poor OZD	Average NOE	Average SOE	Average COE	Average VOE
1	Albania	Eastern Europe	Average AIR	Average HAD	Average PMD	Average OZD	Poor NOE	Average SOE	Average COE	Average VOE
2	Algeria	Greater Middle East	Average AIR	Great HAD	Poor PMD	Average OZD	Poor NOE	Poor SOE	Average COE	Poor VOE
3	Angola	Sub-Saharan Africa	Poor AIR	Poor HAD	Poor PMD	Average OZD	Poor NOE	Average SOE	Poor COE	Poor VOE
4	Antigua and Barbuda	Latin America & Caribbean	Average AIR	Great HAD	Average PMD	Great OZD	Great NOE	Average SOE	Great COE	Great VOE

Fig16. Apriori Algorithm dataset

After implementing the algorithm, results can be seen in "Fig17." and we can say that there is a good chance that if a country has Poor VOE, Poor HAD, Poor AIR, Average NOE it lies in Sub-Saharan Africa. Similarly, if a country has Poor VOE, Poor HAD, Poor AIR, Poor PMD it lies in Sub-Saharan Africa

```
L4
Number of patterns=3

[['Poor HAD', 'Average NOE', 'Sub-Saharan Africa', 'Poor AIR', 'Poor VOE'], ['Poor
HAD', 'Sub-Saharan Africa', 'Poor AIR', 'Poor VOE', 'Poor PMD'], ['Average VOE',
'Average COE', 'Average OZD', 'Poor NOE', 'Average SOE']]

L5
Number of patterns=0
```

Fig17. Frequent pattern result

V. METHODS

In this section, different methods are used to search the impacts of external factors on the environmental sustainability performance of countries, regions, and the world. These external factors consist of the components of the world economy, social inclusion, or unusual (unexpected) factors such as COVID.

A. *Searching for External Factors Having a Significant Effect on Sustainable Development*

It can be observed that GDP per capita, Life expectancy, Global Health Security Index and Biodiversity Habitat Index have positive relations with the EPI. Population has a negative relation which seems true as higher the population, more will be stress on the environment.

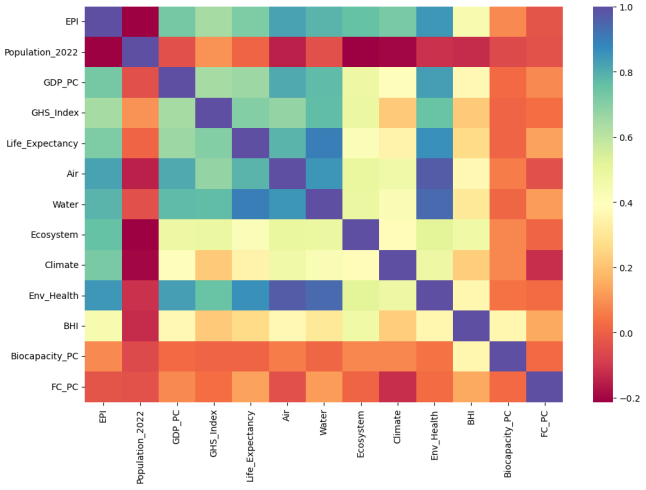


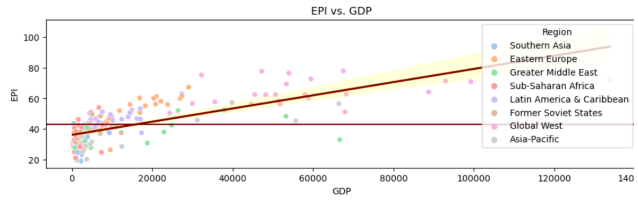
Fig18. Heat map for EPI and external factors

1. *Simple Linear Regression*

Simple linear regression is a model which analyzes one dependent and one independent variable and provides insight about their relationship. Using simple linear regression, we can observe the general relation between external factors and EPI.

a) *EPI vs GDP per capita*

A strong positive correlation can be seen for EPI vs GDP per capita [6]. Spearman correlation is 0.778 and the p-value is $2.312e^{-35}$ also confirms the same.



SpearmanrResult(correlation=0.7781825131823187, pvalue=2.3129598993934406e-35)

Fig19. Linear regression plot for EPI and GDP per capita

A good R-squared value of 0.529 can be observed.

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.529
Model:	OLS	Adj. R-squared:	0.526
Method:	Least Squares	F-statistic:	186.4
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	6.27e-29
Time:	13:04:11	Log-Likelihood:	-609.76
No. Observations:	168	AIC:	1224.
Df Residuals:	166	BIC:	1230.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	36.0659	0.870	41.472	0.000	34.349	37.783
GDP_PC	0.0004	3.16e-05	13.651	0.000	0.000	0.000

Omnibus:	0.698	Durbin-Watson:	2.085
Prob(Omnibus):	0.705	Jarque-Bera (JB):	0.434
Skew:	-0.107	Prob(JB):	0.805
Kurtosis:	3.127	Cond. No.	3.38e+04

Fig20. OLS regression plot for EPI and GDP per capita

b) *EPI vs Life Expectancy*

A strong positive correlation can be seen for EPI vs Life Expectancy. Spearman correlation is 0.724 and the p-value is $1.487e^{-28}$ also confirms the same.

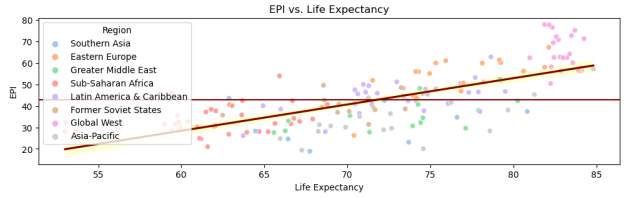


Fig21. Linear regression plot for EPI and Life Expectancy

R-squared value of 0.509 can be observed along with 1.2231 coefficient value.

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.509
Model:	OLS	Adj. R-squared:	0.506
Method:	Least Squares	F-statistic:	172.4
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	1.85e-27
Time:	13:04:12	Log-Likelihood:	-613.16
No. Observations:	168	AIC:	1230.
Df Residuals:	166	BIC:	1237.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.97
const	-45.0233	6.741	-6.679	0.000	-58.332	-31.7
Life_Expectancy	1.2241	0.093	13.128	0.000	1.040	1.4

Omnibus:	0.361	Durbin-Watson:	2.045
Prob(Omnibus):	0.835	Jarque-Bera (JB):	0.127
Skew:	-0.031	Prob(JB):	0.938
Kurtosis:	3.120	Cond. No.	675.

Fig22. OLS regression plot for EPI and Life Expectancy

c) *EPI vs Global Health Security Index*

A positive relation can be seen for EPI vs Global Health Security Index. Spearman correlation is 0.602 and the p-value is $5.313e^{-18}$ also confirms the same.

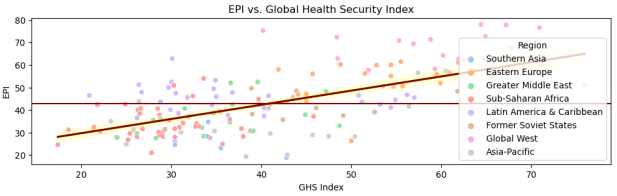


Fig23. Linear regression plot for EPI and Global Health Security Index

R-squared value of 0.410 can be observed using the OLS regression method as shown below in "FIg24"..

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.410
Model:	OLS	Adj. R-squared:	0.407
Method:	Least Squares	F-statistic:	115.6
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	8.57e-21
Time:	13:04:13	Log-Likelihood:	-628.59
No. Observations:	168	AIC:	1261.
Df Residuals:	166	BIC:	1267.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.1410	2.529	6.778	0.000	12.148	22.134
GHS_Index	0.6288	0.058	10.751	0.000	0.513	0.744

Omnibus:	1.726	Durbin-Watson:	2.033
Prob(Omnibus):	0.422	Jarque-Bera (JB):	1.330
Skew:	0.188	Prob(JB):	0.514
Kurtosis:	3.220	Cond. No.	138.

Fig24. OLS regression plot for EPI and Global Health Security Index

.

d) *EPI vs Biodiversity Habitat Index*

A positive relation can be seen for EPI vs Biodiversity Habitat Index. Spearman correlation is 0.412 and the p-value is $2.783e^{-08}$ also confirms the same.

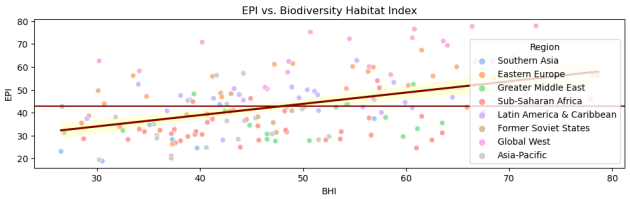


Fig25. Linear regression plot for EPI and Biodiversity Habitat Index

A decent R-squared value of 0.194 can be observed in "Fig26".

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.194
Model:	OLS	Adj. R-squared:	0.189
Method:	Least Squares	F-statistic:	39.93
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	2.33e-09
Time:	13:10:22	Log-Likelihood:	-654.87
No. Observations:	168	AIC:	1314.
Df Residuals:	166	BIC:	1320.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	19.2386	3.867	4.976	0.000	11.605	26.872
BHI	0.4923	0.078	6.319	0.000	0.338	0.646

Omnibus:	6.392	Durbin-Watson:	1.829
Prob(Omnibus):	0.041	Jarque-Bera (JB):	6.463
Skew:	0.449	Prob(JB):	0.0395
Kurtosis:	2.657	Cond. No.	207.

Fig26. OLS regression plot for EPI and Biodiversity Habitat Index

e) EPI vs Population

A negative relation can be seen for EPI vs Population. Spearman correlation is -0.292 and the p-value is 0.0001 also confirms the same.

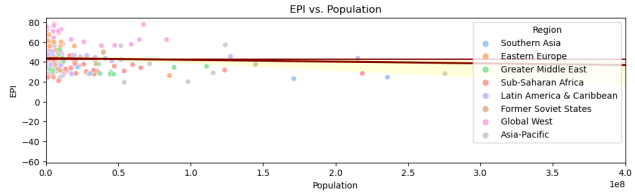


Fig27. Linear regression plot for EPI and Life Expectancy

A very low R-squared value of 0.044 can be observed in the "Fig28" below. Negative coefficient of -1743e^{-08} can also be seen.

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.044
Model:	OLS	Adj. R-squared:	0.038
Method:	Least Squares	F-statistic:	7.578
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	0.00657
Time:	13:10:23	Log-Likelihood:	-669.23
No. Observations:	168	AIC:	1342.
Df Residuals:	166	BIC:	1349.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	43.7600	1.050	41.695	0.000	41.688	45.832
Population_2022	-1.743e-08	6.33e-09	-2.753	0.007	-2.99e-08	-4.93e-09

Omnibus:	9.617	Durbin-Watson:	1.850
Prob(Omnibus):	0.008	Jarque-Bera (JB):	10.288
Skew:	0.592	Prob(JB):	0.00584
Kurtosis:	2.738	Cond. No.	1.72e+08

Fig28. OLS regression plot for EPI and Population

f) Environmental Health vs Life Expectancy

A strong positive relation can be seen for Environmental Health vs Life Expectancy. Spearman correlation is 0.879 and the p-value is 2.547e^{-55} also confirms the same.

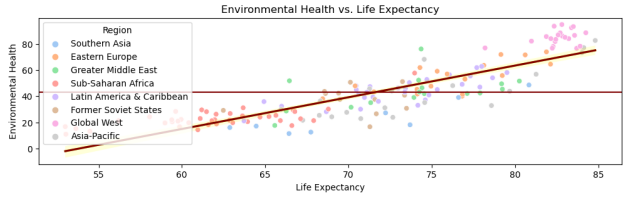


Fig29. Environmental Health vs Life Expectancy

A very good R-squared value of 0.738 can be observed.

OLS Regression Results

Dep. Variable:	Env_Health	R-squared:	0.738
Model:	OLS	Adj. R-squared:	0.736
Method:	Least Squares	F-statistic:	466.8
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	4.19e-50
Time:	13:10:23	Log-Likelihood:	-644.03
No. Observations:	168	AIC:	1292.
Df Residuals:	166	BIC:	1298.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.97
const	-130.3468	8.100	-16.091	0.000	-146.340	-114.3
Life_Expectancy	2.4209	0.112	21.605	0.000	2.200	2.6

Omnibus:	0.994	Durbin-Watson:	1.836
Prob(Omnibus):	0.608	Jarque-Bera (JB):	1.006
Skew:	-0.056	Prob(JB):	0.605
Kurtosis:	2.638	Cond. No.	675.

Fig30. OLS regression plot for EPI and Life Expectancy

2. Multi Linear Regression

We will be using several explanatory variables to predict the outcome of a responsible variable.

We will be using our four main external factors namely GDP per capita, Life Expectancy, Global Health Security Index and Biodiversity Health Index to analyze and determine relation with EPI

We observed a very good R-squared value of 0.666 with very small p-values as shown above

OLS Regression Results

Dep. Variable:	EPI	R-squared:	0.666
Model:	OLS	Adj. R-squared:	0.657
Method:	Least Squares	F-statistic:	81.14
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	9.18e-38
Time:	13:10:25	Log-Likelihood:	-580.95
No. Observations:	168	AIC:	1172.
Df Residuals:	163	BIC:	1188.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-17.9789	7.580	-2.372	0.019	-32.946	-3.012
Life_Expectancy	0.5786	0.118	4.890	0.000	0.345	0.812
GHS_Index	0.1372	0.066	2.064	0.041	0.006	0.268
GDP_PC	0.0002	3.94e-05	5.179	0.000	0.000	0.000
BHI	0.2172	0.055	3.984	0.000	0.110	0.325

Omnibus:	0.240	Durbin-Watson:	2.089
Prob(Omnibus):	0.887	Jarque-Bera (JB):	0.199
Skew:	-0.083	Prob(JB):	0.905
Kurtosis:	2.971	Cond. No.	3.46e+05

Fig31. OLS regression plot for EPI and (GDP per capita, Life Expectancy, GHS Index and BHI)

It is interesting to observe that Life Expectancy has the highest coefficient 0.5786 which made us analyze the Life expectancy factor further.

A very high R-squared value of 0.807 is observed. Significant coefficient values and very small p-values can be observed.

OLS Regression Results

Dep. Variable:	Life_Expectancy	R-squared:	0.807
Model:	OLS	Adj. R-squared:	0.804
Method:	Least Squares	F-statistic:	344.6
Date:	Mon, 05 Dec 2022	Prob (F-statistic):	1.22e-59
Time:	13:10:25	Log-Likelihood:	-444.23
No. Observations:	168	AIC:	894.5
Df Residuals:	165	BIC:	903.8
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	59.6373	0.577	103.291	0.000	58.497	60.777
Air	0.0323	0.024	1.373	0.172	-0.014	0.079
Water	0.2245	0.018	12.708	0.000	0.190	0.259

Omnibus:	3.495	Durbin-Watson:	1.792
Prob(Omnibus):	0.174	Jarque-Bera (JB):	3.053
Skew:	-0.310	Prob(JB):	0.217
Kurtosis:	3.226	Cond. No.	157.

Fig32. OLS regression plot for Life Expectancy and (Air Quality and Water Quality)

B. Time Series Analysis to See COVID and Other Factor Effects on Sustainability

Time series analysis is conducted to see the impact of COVID on the environmental sustainability performance of countries. Models are built to predict the EPI of countries during the COVID period and after the COVID period.

Higher prediction errors for the COVID period mean that there was an unusual and different than expected value of EPI during COVID.

First, the simple linear regression model is built using data from 2010 to 2018 to fit a line to predict the EPI score for 2020 (during COVID) (Figure 1).

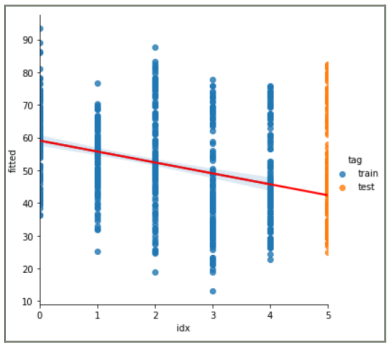


Fig33. EPI prediction for 2020 (Simple linear regression)

Indexes represent the years. 0 stands for 2010 and 5 stands for 2020. The mean squared error for this model is 320 which is a benchmark. Then, a multiple linear regression model is built using AIR, H2O, BDH, and WRS values as additional attributes (Figure 2). The mean squared error for this model is 20.

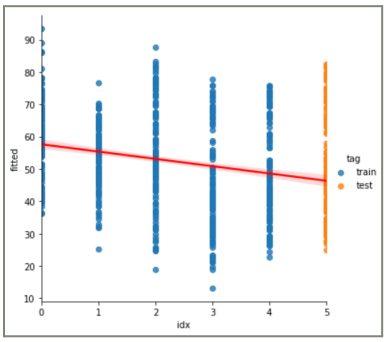


Fig34. EPI prediction for 2020 (Multiple linear regression)

Since the multiple linear regression model is better compared to the single linear regression model in terms of mean squared error, the next model to predict 2022 is built using multiple linear regression.

Then, a multiple linear regression model is built using data from 2010 to 2020 to predict the EPI for 2022 (Figure 3). It is expected to see lower error compared to the predictions for 2020 due to the COVID effect.

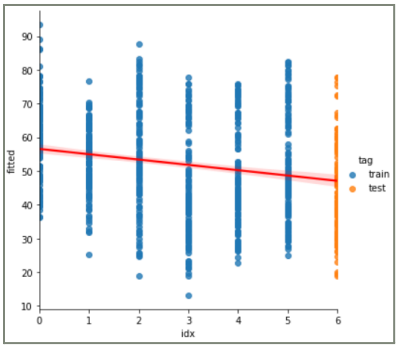


Fig35. EPI prediction for 2022 (Multiple linear regression)

Index 6 stands for the year 2022. The mean squared error of this model is 77. This is not expected. Then, a sample country is taken to see the detailed reason for this. Norway is chosen as the sample country. Multiple linear regression models are built to predict EPI for 2020 and 2022 years respectively.

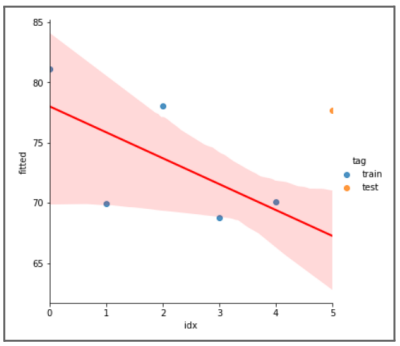


Fig36. EPI prediction for 2020 Norway (Multiple linear regression)

As to the prediction of EPI for 2020, it seems as an outlier point and hard to predict. That's why we have a high mean squared error which is 97 (Figure 4).

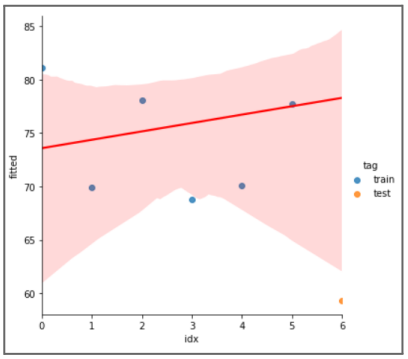


Fig37. EPI prediction for 2022 Norway (Multiple linear regression)

EPI for 2022 also seems an outlier since this value is too low compared to the previous year 2020 and other years. The mean squared error for this case is even higher, which is 708 (Figure 5).

It is observed that linear regression cannot capture the trends and other patterns in the data. Then, it was decided to build an ARIMA model [7][8]. The average of EPI is taken at each time point.

The first ARIMA model is built using data from 2010 to 2018 to predict the EPI of 2020 (Figure 6).

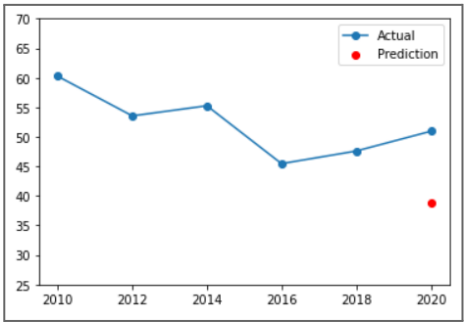


Fig38. EPI prediction for 2020 (ARIMA)

The mean squared error of this model is 148.

The ARIMA model is built using data from 2010 to 2020 to predict the EPI of 2022 (Figure 7).

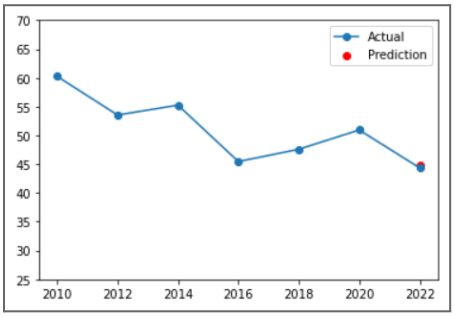


Fig39. EPI prediction for 2022 (ARIMA)

The mean squared error of this model is 0.16.

It is observed that the error of predictions for 2022 is less than the one for 2020. That is, there was an anomaly - an unusual pattern during the COVID period. It can be seen from the EPI graph that EPI (environmental sustainability performance) increased during the COVID period and this increase is significantly shown by the higher error in 2020 predictions. Mean squared errors for each model and test year can be seen on Table 3.

It can be seen that there was an increasing trend from 2016 which was followed by COVID. It is found that the Paris Agreement[9] was signed in 2016 which consists of policies to achieve a climate-neutral world.

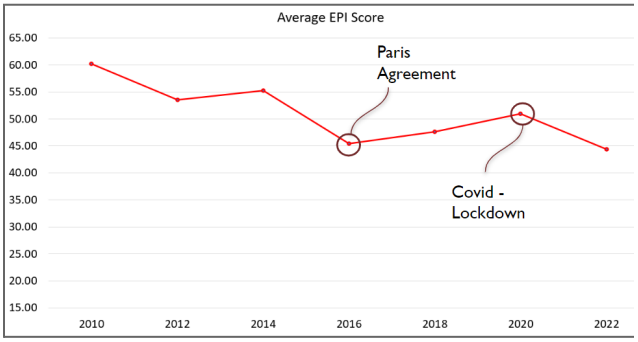


Fig40. Average EPI values from 2010 to 2022

COVID and Paris Agreement had a positive impact on the environmental sustainability performance of countries from 2016 to 2020. However, a decreasing trend has continued after COVID has lost its effect.

RESULTS

It can be observed that GDP per capita, Life Expectancy, Global Health Security Index and Biodiversity Habitat Index are important factors and have good correlation and R-squared values.

Population has a very low R-squared value and hence can be ignored for multi linear regression. Table I below shows the correlation and R-squared values.

TABLE I. SIMPLE LINEAR REGRESSION

Attributes	**Correlation**	**R-Squared**
EPI vs GDP per Capita	0.778	0.529
EPI vs Life Expectancy	0.723	0.509
EPI vs GHS Index	0.602	0.410
EPI vs BHI	0.412	0.194
EPI vs Population	-0.292	0.044
Environmental Health vs Life Expectancy	0.879	0.738

All four external factors together can be used for multi linear regression. Life expectancy in general can be explored further and analyzed with multiple variables like Air Quality and Water Quality. R-squared value can be referred from Table II given below.

TABLE II. SIMPLE LINEAR REGRESSION

Attributes	**R-Squared**
EPI vs (GDP per Capita, Life Expectancy, GHS Index and BHI)	0.666
Life Expectancy vs (Air Quality and Water Quality)	0.807

We have used Simple Linear Regression, Multi Linear Regression and ARIMA and found out ARIMA has the lowest MSE value for the year 2022. Table III below shows the values for each model and case we considered.

TABLE III. TIME SERIES MODELS

Model	**Country**	**Test Year**	**Mean Squared Error**
Simple Linear Regression	All countries	2020	320.6
Simple Linear Regression	All countries	2022	183.7
Multiple Linear Regression	All countries	2020	20.8
Multiple Linear Regression	All countries	2022	77.7

Multiple Linear Regression	Norway	2020	97.7
Multiple Linear Regression	Norway	2022	708.3
ARIMA	**All countries**	**2020**	**147.5**
ARIMA	**All countries**	**2022**	**0.2**

Conclusion

Economic and societal setbacks deriving from the COVID pandemic continue to add to the challenge of meeting sustainability requirements. However, significant improvements in air quality followed early lockdowns and fundamental shifts in economic activities.

Policymakers now have a chance to rebuild their economies and societies on a more sustainable basis that keeps the pandemic-induced benefits in environmental health and ecosystem vitality, but the latest data shows that this opportunity is being wasted across most of the world.

Future scope

We are waiting for the Global governance index to be released for 2022, this can be an important external factor to be examined.

ARIMA can also be used for predicting values for 2023 and comparing them next year once the dataset is released.

References

[1] Socioeconomic Data and Applications Center (SEDAC). https://sedac.ciesin.columbia.edu/data/collection/epi/sets/browse

[2] Environmental Performance Index. https://epi.yale.edu/

[3] Zhaohua Wang, Lin Yang, Jianhua Yin & Bin Zhang. (2018). Assessment and prediction of environmental sustainability in China based on a modified ecological footprint model. Resources, Conservation and Recycling (Volume 132), 301-313.

[4] Ken'ichi Matsumoto, Georgia Makridou & Michalis Doumpos. (2020). Evaluating environmental performance using data envelopment analysis: The case of European countries. Journal of Cleaner Production (Vol. 272, p. 122637)

[5] Aiyshwariya Paulvannan Kanmani, Renee Obringer, Benjamin Rachunok & Roshanak Nateghi. (2020). Assessing Global Environmental Sustainability Via an Unsupervised Clustering Framework. Sustainability, MDPI (Vol. 12(2)), 1-12.

[6] Fatma Tektüfekçi & Nilgün Kutay. (2016). The Relationship Between EPI and GDP: An Examination on Developed and Emerging Countries. Journal of Modern Accounting and Auditing, May 2016, Vol. 12, No. 5, 268-276

[7] Auto Arima. pmdarima 2.0.2. alkaline-ml

[8] Jose Marcial Portilla. (2018). Using Python and Auto ARIMA to Forecast Seasonal Time Series

[9] The Paris Agreement. (2016). United Nations Climate Change..