IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

SNNO: 2347–3657

Volume 9,Issue3 sep ,2021

Grouping of Nations according to Development-Influencing Factors and a Comparison through Time

Mrs. Mukka Shirisha, Mr.Syed Shah Gulam Mujtaba Quadri, Mr. Abdul Rajak Shaik

Abstract:

There are a number of factors that might effect a country's growth and economy in today's globalised world. The future of nations depends on a thorough examination of these issues. In this context, nations are categorised as undeveloped, transitional, developing, and developed countries, among other things. Countries that are categorised this manner and are in the same class have a wide range of commonalities. Using elements that influence different degrees of development, this research aims to highlight the changes in the economies of Balkan and former Soviet Union nations during the previous two decades. According to the years in which the socio-economic variables were obtained, missing data imputation techniques were employed to identify the values of the variables that were not present. Cluster analysis is used to identify nations that are comparable in terms of their degree of development, and these countries are then grouped into clusters. For 1995 and 2015, the same processes are used, but the nations' locations are indicated to have changed.

INTRODUCTION

Countries' economic progress has been a central theme in the history of the world. In addition, there are numerous other factors that may be used to gauge a country's social and cultural progress, such as its GDP. Countries may be compared based on these factors. Developing nations, transition countries, and industrialised countries may all be used as examples of literature's degree of development. In this perspective, it is clear that nations with equal levels of development have the same names. As is well known, access to data has become simpler in the modern day. But there are still a few missing numbers in the data Due to the importance of full observations in data analysis, many statistical approaches are utilised to overcome the issue of missing data. A nation cannot be compared to other countries if it does not have all of the necessary facts to do so. The 54 nations included in this research were compared using 14 distinct factors. Comparisons between 1995 and 2015 are the primary objective of this research. Missing value imputation techniques were used to fill in the blanks in our data set. After that, factors were grouped together in a clustering analysis to determine which nations were at different stages of development. For the years 1995 as well as 2015, these trades were made. One of the goals is to track down how each country has evolved over these last two decades. The following steps will be taken to complete this project: Section II reviews past research, Section III defines data and variables, Section IV examines methods, and Section V offers an interpretation of the findings. Section VI will wrap up the research.

1,2,3 Assistant Professor
1,2,3 Department of CSE&(AI & ML)
1,2,3 Global Institute of Engineering and Technology Moinabad, Ranga Reddy District, Telangana State.

PREVIOUS WORKS:

A panel of 23 OECD nations was employed by Carree, Van Stel, Thurik, and Wennekers [1] for their research. The equilibrium points between economic growth and company ownership have been identified. A comparison of GDP Per Capita in developed and developing countries was performed by Maddison, A [2]. Compared to the factor analysis based on the socioeconomic development levels of Turkey and the EU countries, Taş,. K. & zel, S.. [3] We used hierarchical cluster analysis of the welfare regimes in advanced nations to identify the clusters of the welfare regimes based on a set of quantitative social indices. Low technical efficiency has created an economic growth gap between wealthy nations like the United States and low-income countries like China, according to research by Hulten, R. and A. Isaksson [5]. Empirical conclusions based on human capital and foreign direct investment (FDI) in developing countries were discovered by Noorbakhsh, Paloni, and Youssef, A. It was shown that the real exchange rate between developing countries is strongly linked to the interaction between international commerce, foreign direct investment, and foreign direct investment. According to Williamson and Boehmer [8], gender stratification theory may be used to explain disparities in female life expectancy across nations that are less developed. In 2005 and 2012, Grzebyk, M., and Stec, M. [9] used statistical analysis to assess EU nations' progress in sustainable development.

DATA AND VARIABLES:

Analysis of data is A. A variety of websites were explored to gather the information needed for the study. Each year's data was collected independently, beginning in 1995 and concluding in 2015. By employing several missing value imputation approaches, the data was able to fill in the blanks. Both datasets had to be standardised since the variables were measured in different units of measurement. After that, the k-means approach was used to do clustering analysis utilising standardised variables. The B. Variables The following are the factors that indicate a country's advancement. Investment in US Dollars at current prices and exchange rates in millions of FDI logarithms is referred to as FDI. • GDP: The logarithm of GDP in constant 2005 U.S. dollars. (Import + Export) / GDP • Real Effective Exchange Rate • Trade Openness • CL: Civil Liberty Index: 1: Very High Civil Liberty, 7: Very Low Civil Liberty. • KOF: Globalization Index From one to a hundred, globalisation is either: Inflation: Consumer Prices on a Per Capita Basis • PR: Political Right Index 1: a high level of political right, 7: a low level of political right 0 on the Economic Freedom Index means that there is no freedom. 100: No Restraints. Internet users: the proportion of the working-age population who are educated at the secondary and tertiary levels. Secondary: the proportion of persons of working age who are educated. Tertiary: (per 100 people) There are millions of ways to say "absolute number." Economic output per worker (in constant 1990 US dollars at PPP) is a measure of labour productivity.

METHODS:

It is important to note throughout the research that the proper imputation procedures between missing data were used to ascertain the missing values. As a result of the various measures, the data must be normalised. Using standardised variables, a clustering analysis was performed as a final step. MICE, Amelia, missForest, Hmisc, and mi packages and the kmeans() function in the r programming language were employed in our research for missing data imputation. A. Imputation of Missing Values The imputation of missing data has received a lot of attention in the statistical community [10]. There are a variety of ways to fill in the blanks when it comes to missing values. As an example, there are a number of techniques that may be used for a variety of purposes. If you're going to get the best results, it's critical to use the suitable procedure for the data structure. To put it simply, regression analysis is used to estimate the values of one or more dependent variables using one or more independent variables. Simple or multiple linear regression analysis is used to detect the missing variable in numerical data, whereas logistic regression is used to estimate the value in binary qualitative data [11]. On the other hand, in our analysis, we treated the data as a time series since we had values from 1995 to 2015. Whenever an observation's missing value is discovered, the dependent variable is transformed into the year, which is then used as an independent

variable in the trend equation. Using this trend equation, the imputation of a missing value is achieved. Clustering with K-Means Clustering methods include k-means clustering, k-medoids clustering, hierarchical clustering, and density-based clustering, to name a few examples. K-Means is the most often used clustering technique since it is simple to implement. The procedure has been designed to accommodate a certain number of sets (k). The procedure may be broken down into three steps [12]. First, find the centre points. Centroids may be used to measure the distance between objects. Cluster the objects in the smallest possible distance. The technique is then used to reduce this function to the smallest possible value [13].

$$J = \sum_{i=1}^{k} \sum_{i=1}^{n} \left\| X_j^i - c_i \right\|^2$$

where x_j^i and c_i show the j-th data point and the i-th centroids. $\left\| x_j^i - c_i \right\|$ means the L^2 norm of $(x_j^i - c_i)$.

RESULTS:

Table I and Table II provide the raw data for the factors that determine the nations' development levels. These charts show the mean, standard deviation, and lowest and maximum values for each country between 1995 and 2015. The raw status of these variables is presented to better interpret the data. After this, the data is normalised so that it may be clustered.

TABLE I. DESCRIPTIVE STATISTICS FOR VARIABLES RELATED TO THE LEVEL OF DEVELOPMENT OF COUNTRIES IN 1995

	1995 (N=54)			
Variables	Minimum	Maximum	Mean	Std. Deviation
FDI	3.47	11.52	7.32	2.08
GDP	7.80	14.20	10.66	1.74
Internet	0.00	2.90	0.24	0.58
ExchangeRate	53.46	407.22	97.33	52.56
PR	1.00	7.00	4.06	2.02
CL	1.00	7.00	4.20	1.66
Freedom	23.30	72.00	51.01	11.71
KOF	27.87	69.48	49.31	10.66
Labor Productivity	6.12	10.64	8.93	1.06
Trade Openness	15.08	177.36	74.18	37.39
Pop	5.94	14.03	9.64	1.63
Inflation	1.06	2672.23	142.66	400.44
Secondary	13.00	91.20	56.32	23.97
Tertiary	1.24	39.13	10.97	7.87
Energy	3.48	7.86	6.05	1.04

TABLE II. DESCRIPTIVE STATISTICS FOR VARIABLES RELATED TO THE LEVEL OF DEVELOPMENT OF COUNTRIES IN 2015

	2015 (N=54)			
Variables	Minimum	Maximum	Mean	Std. Deviation
FDI	7.62	14.02	10.53	1.47
GDP	8.71	16.00	11.52	1.70
Internet	11.60	89.65	52.55	22.31
ExchangeRate	62.74	529.72	112.56	61.40
PR	1.00	7.00	3.54	2.10
CL	1.00	7.00	3.50	1.79
Freedom	34.30	76.80	59.85	8.94
KOF	37.43	84.20	65.22	11.81
Labor Productivity	6.97	10.98	9.50	0.94
Trade Openness	23.04	276.23	87.51	48.57
Pop	6.06	14.15	9.82	1.72
Inflation	0.13	910.00	36.81	152.60
Secondary	21.78	98.12	71.70	23.58
Tertiary	2.51	57.71	16.38	11.32
Energy	4.81	7.67	6.37	0.74

In the following table, we provide the findings of our k-means clustering study. The k-means cluster approach was selected using the principal component analysis. The variance explanations for 1995 and 2015 were found to be 67.823 and 66.750, respectively, after the study. As a consequence of this investigation, the k values for both are set to 3. Table III and Table IV exhibit the analysis findings for 1995 and 2015, respectively. According to K-means clustering, there are three clusters of sizes 18, 20, and 16 for 1995, and three clusters of sizes 13, 23, and 18 for 2015. There was a significant difference between the two years' worth of clusters in the ANOVA analysis of the variables. Three clusters of 14 variables are also distinct.

Table III. RESULTS OF THE K-MEANS METHOD FOR 1995

1995		
Cluster 1 (N=18)	Cluster 2 (N=20)	Cluster 3 (N=16)
Algeria	Argentina	Albania
Brazil	Bulgaria	Angola
Cameroon	Croatia	Armenia
China	Cyprus	Azerbaijan
Cote d'Ivoire	Czech Republic	Belarus
Egypt	Estonia	Bosnia and Herzegovina
Ghana	Hungary	Ethiopia
India	Latvia	Georgia
Indonesia	Lebanon	Kazakhstan
Kenya	Lithuania	Kyrgyzstan
Mexico	Malaysia	Macedonia
Morocco	Malta	Moldova
Nigeria	Poland	Tajikistan
Pakistan	Romania	Turkmenistan
Peru	Russia	Ukraine
Senegal	Slovakia	Uzbekistan
Tunisia	Slovenia	
Turkey	South Africa	
	South Korea	
	Venezuela	

TABLE IV. RESULTS OF THE K-MEANS METHOD FOR 2015

2015		
Cluster 1 (N=13)	Cluster 2 (N=23)	Cluster 3 (N=18)
Argentina	Albania	Algeria
Brazil	Armenia	Angola
China	Bosnia and Herzegovina	Azerbaijan
India	Bulgaria	Belarus
Indonesia	Croatia	Cameroon
Kazakhstan	Cyprus	Cote d'Ivoire
Mexico	Czech Republic	Egypt
Morocco	Estonia	Ethiopia
Peru	Georgia	Ghana
Russia	Hungary	Kenya
South Africa	Latvia	Kyrgyzstan
Turkey	Lebanon	Nigeria
Ukraine	Lithuania	Pakistan
	Macedonia	Senegal
	Malaysia	Tajikistan
	Malta	Turkmenistan
	Moldova	Uzbekistan
	Poland	Venezuela
	Romania	
	Slovakia	
	Slovenia	
	South Korea	
	Tunisia	

In the first cluster, nations like Brazil, Senegal, and Morocco are considered developing countries when we look at all three groups in 1995. The Czech Republic, Poland, and Romania may be found in the second cluster, as can Poland and the Czech Republic. The third cluster includes nations like Turkmenistan, Uzbekistan, and Angola. A close examination of its findings reveals that it is rational and consistent. After looking at the data from 2015, it's easy to see why nations like Belarus, Turkey, and Malta have remained stable. It may also be stated that nations such as Egypt and Pakistan are shifting from industrialised countries to transitional ones. The list of developed nations also includes Argentina and Russia in 1995. It may be argued that it belongs to the group of developing nations based on the findings of 2015. For this analysis, it is critical that the nations in a given cluster have certain features. If in 1995 the countries of Hungary and Estonia were grouped together, then the countries of Tajikistan and Turkmenistan were in the same cluster in 2015. This illustrates that Venezuela's growth during the last two decades has been negative.

CONCLUSIONS:

For the development of nations, 14 factors are considered in this research. It was necessary to compile information on 54 countries, including both industrialised and developing nations as well as those in transition. Foremost among them are the former Soviet Union member states, as well as those of the Balkans and South America. In addition, statistics for 1995 and 2015 were gathered for comparison. Methods for assigning a value to an unobserved data point are used first. Standardization and inclusion in the analysis followed. K-means was used to evaluate the data between 1995 and 2015. The ANOVA study found that the 14 variables were different among the generated clusters of variables employed. Using the results of clustering analysis, we were able to identify three distinct groups. The 1995 and 2015 analyses were compared. The findings show that certain nations have advanced or regressed in the last two decades. For nations to grow or decline in terms of development over the course of 20 years is common. According to the findings, Turkey has maintained its position at the top of its class. Furthermore, in 1995, Venezuela was in the same cluster as developing nations, but in 2015, it can be claimed that it was in the same cluster as transitional countries. Different assessments of nations may be made based on outcomes, of course.

REFERENCES :

[1] Carree, M., Van Stel, A., Thurik, R., &Wennekers, S. (2002). Economic development and business ownership: an analysis using data of 23 OECD countries in the period 1976–1996. Small business economics, 19(3), 271-290.

[2] Maddison, A. (1983). A comparison of levels of GDP per capita in developed and developing countries, 1700–1980. The Journal of Economic History, 43(1), 27-41.

[3] Taş, Ç. K., & Özel, S. Ö. (2017). FaktöranaliziyöntemiileTürkiyeveAvrupaBirliğiÜyesiÜlkelerinsosyo-ekonomikgöstergelerbakımındangelişmişlikdüzeylerininkarşılaştırılması. Journal of the Cukurova University Institute of Social Sciences, 26(3), 60.

[4] Saint-Arnaud, S., & Bernard, P. (2003). Convergence or resilience? A hierarchical cluster analysis of the welfare regimes in advanced countries. Current Sociology, 51(5), 499-527.

[5] Hulten, C. R., & Isaksson, A. (2007). Why development levels differ: The sources of differential economic growth in a panel of high and low-income countries (No. w13469). National Bureau of Economic Research.

[6] Noorbakhsh, F., Paloni, A., & Youssef, A. (2001). Human capital and FDI inflows to developing countries: New empirical evidence. World development, 29(9), 1593-1610.

[7] Goldberg, L. S., & Klein, M. W. (1997). Foreign Direct Investment, Trade and Real Exchange Rate Linkages in Developing Countries (No. w6344). National Bureau of Economic Research.

[8] Williamson, J. B., & Boehmer, U. (1997). Female life expectancy, gender stratification, health status, and level of economic development: A cross-national study of less developed countries. Social Science & Medicine, 45(2), 305-317.

[9] Grzebyk, M., & Stec, M. (2015). Sustainable development in EU countries: concept and rating of levels of development. Sustainable Development, 23(2), 110-123.

[10] Demirhan, H., & Renwick, Z. (2018). Missing value imputation for short to mid-term horizontal solar irradiance data. Applied Energy, 225, 998-1012.

[11] Alpar, R. (2011). Çok Değişkenlİ statistiksel Yöntemler, Ankara: Detay Yayıncılık.

[12] Dalatu, P. I., Fitrianto, A., & Mustapha, A. (2017). Hybrid distance functions for K-Means clustering algorithms. Statistical Journal of the IAOS, 33(4), 989-996.

[13] Wang, Q., Wang, Y., Niu, R., & Peng, L. (2017). Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility Mapping in the Three Gorges Area, China. Remote Sensing, 9(9), 938.