Clustering Countries According to Global Innovation Index 2022 Using Fuzzy Clustering Analysis

Esra POLAT (espolat@hacettepe.edu.tr)
Department of Statistics, Faculty of Science, Hacettepe University, Ankara, Turkey

ABSTRACT

In the dynamic landscape of global competition, characterized by the escalating significance of technology, innovation emerges as a pivotal determinant for nations seeking to enhance and sustain their competitiveness. In this research, the dataset encompassing seven subcategories within each primary indicator of both innovation input and output subscales, as delineated in the 2022 Global Innovation Index (GII) report, was employed for clustering 132 countries with a fuzzy c-means clustering algorithm. Cluster 1 encompasses a total of 97 countries, while Cluster 2 comprises 35 countries. Following the analysis, the countries with high-income levels in Cluster 2 ranked first. These countries are also positioned among the foremost countries in the GII rankings, which means the ones exhibiting high-income levels attain leading positions similarly across innovation indicators. However, all of the low-income countries and all low-middle-income countries except India clustered in Cluster 1. The cluster analysis results and index rankings are parallel for the countries with high and low GII values. The top countries in GII rankings clustered in Cluster 2. The countries at the bottom of GII rankings clustered in Cluster 1. The fuzzy c-means clustering algorithm revealed the power of the GII to reflect the data.

Keyword: *Countries, Global innovation index, innovation, Fuzzy Clustering Analysis*

JEL Classification: *B41, C13, C22, R21, R29*

1. INTRODUCTION

Innovation indices aspire to methodically assess the impact of innovation on diverse variables across technological, macro, micro, and other dimensions. Global Innovation Index (GII) is one of the evaluators of innovation performance and efficiency. The GII may be considered as a paramount metric of a country's ability for innovation. The GII is derived by computing the mean of the sub-index values about innovation output and innovation input. Hancioglu (2016) observed that the GII can facilitate the computation of innovation efficiency performance values for countries. The determination of countries' innovation efficiency performance involves the computation of the ratio between innovation output sub-index values and innovation input sub-index values. This ratio elucidates the extent to which outputs can generated per unit of input (Aytekin et al., 2022).

The overarching objective of the GII is to enhance the precision of innovation measurement, thereby contributing to a more comprehensive depiction of global innovation ecosystems (WIPO, 2022).

The comprehensive ranking within the GII hinges upon two pivotal sub-components, the Innovation Output Sub-Component and the Innovation Input Sub-Component, both of equal significance in delineating a comprehensive overview of innovation. As a result, the computation necessitates the derivation of three distinct indices (WIPO, 2022).

- Innovation Input Sub-Component: 5 input components encapsulate facets of the economic framework that foster and facilitate innovative endeavors.
- Innovation Output Sub-Component: Innovation outputs manifest as outcomes of inventive activities within the economic sphere. Despite the Output Sub-Component incorporating solely two components, its significance equals that of the Input Sub-Component in the computation of the overarching GII scores.
- The aggregate GII score is calculated as the mean of the Output and Sub-Components, serving as the basis for the generation of GII economy rankings.

The 2022 GII Report, authored by the World Intellectual Property Organization (WIPO), a Switzerland-based agency operating under the auspices of the United Nations (UN), has been released.

There are studies on analyzing GII report data sets using cluster analysis. Some of them are listed below.

Jankowska et al.(2017) utilized k-means cluster analysis on GII (2015) data to discern countries exhibiting varying levels of innovation inputs, delineated as high, medium, or low, thereby reflecting their capacity to generate innovation output. Then, they wanted to identify countries deviating from expected patterns, i.e., even though they were well (poorly, moderately) equipped, performed better (or worse) than foreseen. Furthermore, they conducted a focused examination of Poland and Bulgaria to ascertain the underlying reasons for their challenges in sustaining innovations.

In their study, Unlu (2019) empirically examined variations in innovation performance efficiency across middle-income countries. They used Ward's agglomerative hierarchical technique for cluster analysis. Subsequently, cluster analyses were performed individually for both input and output indicators. Additionally, discriminant analysis was employed to ascertain the determinants of efficiency discrepancies. The study encompassed 54 countries classified based on the World Bank's income categorization, comprising 31 upper-middle-income and 23 lower-middle-income nations. The dataset is taken from the 2018 GII. The results substantiate the presence of inefficiency issues regarding innovation performance within middle-income economies.

Gurtuna and Polat (2020) examined the three subcategories associated with each primary indicator within the innovation output and input subscales of the 2018 GII report. The dataset comprised 126 countries, which underwent analysis using the clustering method. This investigation employs Ward's Technique and the k-means method. Their first aim was to assign the countries into 3 and 5 clusters using their GII values. They mentioned that sorting countries by GII values was possible, however, the issue of determining clusters was uncertain. Cluster analysis of countries made it possible to cluster countries such as Low - Medium - High or Low - Low Medium - Medium - Medium High – High. The second purpose of the analysis was to use the 21 variables when creating the index to determine similar countries in terms of innovation. This target was accomplished by using various cluster numbers, such as 3, 4, and 5, and different methods, such as Ward's Technique and the k-means method. Although ranking countries according to cluster analysis results or according to GII values were consistent with each other in some situations, it also observed that they behaved differently at some points.

Famalika and Sihombing (2021) employed the k-medians and k-means techniques to cluster countries using the GII 2018 dataset in their research. The sub-component within the GII comprises 7 components: Human Capital and Research, Institutions, Market sophistication, Infrastructure, Creative Outputs, Knowledge and Technology Outputs, and Business Sophistication. The clustering analyses applied to these seven variables. Upon conducting the research, the derived clustering outcomes employing both the k-medians and k-means methods revealed that k-medians outperformed the k-means technique, evidenced by the smaller variance value associated with k-medians. In each method, 3 clusters were created. In the k-means method, Cluster 1 comprises 48 countries, Cluster 2 includes 45 countries, and Cluster 3 encompasses 33 countries. Notably, Cluster 1 exhibits a relatively high average value across seven variables. However, Cluster 2 demonstrates a low average value for these variables, while Cluster 3 manifests the highest average value among the 3 clusters. Transitioning to the k-medians method, Cluster 1 encompasses 33 countries, Cluster 2 involves 53 countries, and Cluster 3 includes 40 countries. Cluster 1, in this context, displays the highest average value across the seven variables. Cluster 2 demonstrates a relatively high average value, whereas Cluster 3 exhibits a low average value for the mentioned variables.

In their study, Eren and Gelmez (2022) ranked 132 countries based on the GII (2021) report dataset, employing ARAS and COPRAS methods across seven criteria. The ENTROPY weighting method was applied as the primary approach for ranking countries based on their innovation performance. After ranking innovation performances of 132 countries within the index, they categorized into clusters based on their innovation indicators. The cluster analysis was applied utilizing the WEKA program. Switzerland, Sweden, and the USA emerged as the nations with the most favorable rankings concerning innovation indicators, as determined through the ARAS and COPRAS techniques. Conversely, Benin, Angola, and Guinea were identified as the countries with the least favorable rankings. The outcome of the clustering analysis conducted using the WEKA program revealed the subdivision of these countries into eight distinct clusters.

In their investigation, Alqararah and Alnafrah (2023) utilized a multi-dimensional innovation-driven clustering methodological analysis for the data set of GII for the year 2019. k-means and hierarchical cluster analysis approaches were employed, utilizing diverse sets of distance matrices to unveil and scrutinize discrete innovation patterns. They categorized 129 countries into 4 clusters: Advanced, Specials, Primitives, and Intermediates. Each cluster demonstrates distinct weaknesses and strengths concerning innovation performance. The Specials cluster demonstrates notable proficiency in knowledge commercialization and institutions, whereas, the Advanced cluster exhibits strengths in education and ICT-related services, albeit with a weakness apparent in patent commercialization. The Intermediates cluster exhibits strengths in venture capital and labor productivity, while simultaneously manifesting weaknesses in R&D expenditure and the quality of higher education. The Primitives cluster demonstrates proficiency in creative actions but it presents deficiencies in training, education, and digital skills. Moreover, they specified 35 indicators characterized by minimal variance parts across nations (Alqararah and Alnafrah, 2023).

The countries can be ranked using innovation indices. However, countries could be similar or different regarding innovation indicators, and this may not have reflected in the indexes. One of the goals of this study is to examine the similarities and differences between countries with each other within the scope of innovation performances and evaluate how much GII index values reflect these similarities and differences. By applying fuzzy cluster analysis, it is aimed to bring together countries with similar characteristics.

The comparison of the results of fuzzy cluster analysis with the results of an innovation measurement index could be used to measure the consistency of the index for future studies.

The following sections of this study are organized as in below. In Section 2, a brief theory for Fuzzy Clustering based on the fuzzy c-means clustering algorithm is presented. In Section 3, presentation of the methodology and variables used for clustering the 132 countries by fuzzy clustering analysis is presented. Finally, general comments and a summary of the results are presented in the last section.

2.MATERIAL AND METHOD

2.1.Fuzzy Clustering

The Fuzzy Clustering technique is recognized as a generalized variant that incorporates elements from both the medoids and k-means clustering techniques, both of which exemplify non-hierarchical clustering approaches. The Fuzzy Clustering technique involves the separation of n units into k clusters, allowing for the non-compulsory inclusion of units in clusters and permitting their divergence. In traditional clustering methodologies, units are unequivocally allocated to a specific cluster. Nevertheless, within the fuzzy clustering technique, it is necessary to compute the membership coefficient and membership probability of units across various clusters. In clustering analysis, the allocation of units to a cluster is examined within three distinctive scenarios: probabilistic, fuzzy, and absolute. In the paradigm of absolute clustering, units exhibit an exclusive affiliation wherein they are either a member or not a member of a single cluster. Conversely, in fuzzy clustering, elements can concurrently belong to multiple clusters. In probabilistic clustering, a unit is assigned to a cluster or not. Nevertheless, the allotment of a unit to a cluster is contingent upon the underlying probability distribution (Alptekin and Yesilaydin, 2015). The definitiveness inherent in traditional clustering methodologies occasionally gives rise to inaccuracies in results. In instances where observational units are equidistant from each homogeneous cluster, ambiguity arises concerning the assignment of these units to specific clusters. This scenario underscores the significance of the conception of the probability of membership to clusters (Bulbul and Camkiran, 2018). Given that the fuzzy clustering technique facilitates membership determination based on the degree of affiliation with clusters, it often yields more robust and natural outcomes compared to conventional methods (Cebeci and Yildiz, 2015).

Fuzzy clustering affords a nuanced exploration of data, offering more detailed insights. However, challenges arise when summarizing and classifying information when dealing with an abundance of units and clusters, leading to an excess of generated outputs (Zorlutuna and Erilli, 2018).

The predominant technique in fuzzy clustering is the fuzzy c-means clustering algorithm, initially introduced by Bezdek and Hathaway (1987) and subsequently refined by Kaufman and Rousseeuw (1990). As an alternative approach to the conventional k-means method, where each unit is exclusively assigned to a single cluster, fuzzy clustering assigns each unit a probability of belonging to every cluster individually and distinctively from other clusters. The Fuzzy c-means algorithm addresses situations where units are positioned in a manner that makes it difficult to determine the optimal center to which they should belong. This challenge arises when the distances between a unit and neighboring centers are almost identical to each other. Fuzzy c-means determines centroids according to these probabilities. The applied procedures for iteration, termination, and initialization are identical to the ones used in the k-means algorithm. It is discerned that fuzzy c-means and k-means diverge in their treatment of assigning probabilities to individual data points, with k-means assigning a probability of 1 if the unit is closest to a centroid and 0 otherwise. Challenges arise in case of the distances between a unit and neighboring centers are almost identical to each other (Al Rahhal and Rencber, 2022).

This algorithm was designed to minimize the cost function, computed based on cluster memberships and distances. The cost function is presented in Eqn [1] (Bagdatli Kalkan, 2019).

$$J = \left(U, c_1, \cdots, c_C\right) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \qquad [1]$$

In Eqn [1], U represents the Membership Matrix consisting of uij, denoting membership probabilities. These probabilities range between zero and one, with the sum of membership probabilities for each point equating to one. ci is the cluster center of fuzzy group i; and dij signifies the Euclidean distance between the ith cluster center and the jth unit. The parameter m serves as a weighting exponent. The necessary circumstances for Eqn [1] to achieve its minimum are as follows:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \qquad [2]$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(n-1)}} \qquad [3]$$

The Fuzzy c-means algorithm encompasses four distinct steps. The initial step involves the initialization of the membership matrix U with subjective values ranging between zero and one. In the second step, cluster centers are calculated using Eqn [2]. In the third step, Eqn [1] is used for the calculation of the cost function. Stop if either cost function falls below a specified acceptance value or its betterment over former iterations below a specific threshold. The final step involves the formulation of the U matrix utilizing Eqn [3], subsequently iterating back to the second step (Saravananathan and Velmurugan, 2018). Given that the outcome of this algorithm is contingent upon the initially created random values, various algorithms have been and continue to be developed to address challenges arising from inherent randomness (Zorlutuna and Erilli, 2018).

3. RESULTS AND DISCUSSION

GII measures the innovation of countries by employing a multitude of indicators that have an impact on innovation. In this study, using data obtained from the GII (2022) Report, seven criteria (Institutions, Infrastructure, Market sophistication, Knowledge and technology outputs, Human capital and research, Business sophistication, and Creative outputs), 132 countries were analyzed by fuzzy clustering analysis. The data was analyzed using a fuzzy c-means clustering algorithm.

This study utilized secondary data obtained from collaborative efforts involving the World Intellectual Property Organization in conjunction with INSEAD and Cornell University. These three institutions assessed a nation's global innovation standing based on seven components, as given in Table 1 (Famalika and Sihombing, 2021; Aytekin et al., 2022).

Global Innovation Index components

Table 1

Input innovation	**Output innovation**
Institutions	Knowledge and technology outputs
Human capital and research	Creative outputs
Infrastructure	
Market sophistication	
Business sophistication	

Initially, validity indices are employed for the determination of the suitable number of clusters. The validity values are presented for various numbers of clusters in Table 2.

Fuzzy C-Means Clustering Validity values

Table 2

	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Partition Entropy Index	0.350	0.606	0.873	1.055	1.228	1.371	1.497	1.607	1.703
Partition Coefficient	0.790	0.659	0.529	0.453	0.386	0.343	0.307	0.279	0.252
Modified Partition Coefficient	0.579	0.489	0.372	0.316	0.263	0.234	0.208	0.189	0.169
Fuzzy Silhouette Index	0.780	0.676	0.568	0.504	0.403	0.427	0.416	0.402	0.381

Validity indices are commonly employed for determination the optimal number of clusters; however, they cannot inherently furnish definitive insights into the quality of clustering outcomes. The computation of the Partition Coefficient Index involves the utilization of the clustering degrees matrix (U), to achieve a maximum value. The Modified Partition Coefficient Index is characterized as a linear transformation of the Partition Coefficient, with its values constrained within the range of 0 to 1. The Modified Partition Coefficient Index is characterized as a linear transformation of the Partition Coefficient, with its values constrained within the range of 0 to 1. The Fuzzy Silhouette Index is a more sophisticated metric in comparison to other indices, leveraging a broader spectrum of information. The objective is to maximize this value (Ferraro and Giordani, 2015; Bagdatli Kalkan, 2019). Consequently, several indices listed in Table 2 do not serve as conclusive evidence for the quality of clustering. Nevertheless, the current quantity is computed to be the most optimal among alternative cluster numbers. It is important to note that no validity index produces definitive outcomes, thereby necessitating ongoing developments in the refinement of these indices. According to these indexes, the number of clusters was determined as 2. After fuzzy c-means clustering, obtained 2 clusters of countries were based on the GII 2022. Membership values of countries to clusters are shown in Table 3.

Membership values of countries to clusters

Table 3

	Country	Cluster 1 Membership Degree	Cluster 2 Membership Degree		Country	Cluster 1 Membership Degree	Cluster 2 Membership Degree
C-1	Albania	0.956	0.044	C-67	Lithuania	0.324	0.676
C-2	Algeria	0.947	0.053	C-68	Luxembourg	0.090	0.910
C-3	Angola	0.920	0.080	C-69	Madagascar	0.925	0.075
C-4	Argentina	0.896	0.104	C-70	Malaysia	0.261	0.739
C-5	Armenia	0.957	0.043	C-71	Mali	0.925	0.075
C-6	Australia	0.048	0.952	C-72	Malta	0.105	0.895
C-7	Austria	0.042	0.958	C-73	Mauritania	0.909	0.091
C-8	Azerbaijan	0.949	0.051	C-74	Mauritius	0.546	0.454
C-9	Bahrain	0.765	0.235	C-75	Mexico	0.822	0.178
C-10	Bangladesh	0.962	0.038	C-76	Mongolia	0.902	0.098
C-11	Belarus	0.796	0.204	C-77	Montenegro	0.783	0.217
C-12	Belgium	0.056	0.944	C-78	Morocco	0.921	0.079
C-13	Benin	0.931	0.069	C-79	Mozambique	0.927	0.073
C-14	Bosnia and Herzegovina	0.825	0.175	C-80	Myanmar	0.936	0.064
C-15	Botswana	0.899	0.101	C-81	Namibia	0.942	0.058
C-16	Brazil	0.701	0.299	C-82	Nepal	0.940	0.060
C-17	Brunei Darussalam	0.802	0.198	C-83	Netherlands	0.072	0.928
C-18	Bulgaria	0.417	0.583	C-84	New Zealand	0.042	0.958
C-19	Burkina Faso	0.950	0.050	C-85	Nicaragua	0.931	0.069
C-20	Burundi	0.907	0.093	C-86	Niger	0.922	0.078
C-21	Cote d’Ivoire	0.950	0.050	C-87	Nigeria	0.940	0.060
C-22	Cambodia	0.954	0.046	C-88	North Macedonia	0.817	0.183
C-23	Cameroon	0.916	0.084	C-89	Norway	0.056	0.944
C-24	Canada	0.058	0.942	C-90	Oman	0.861	0.139
C-25	Chile	0.595	0.405	C-91	Pakistan	0.952	0.048
C-26	China	0.068	0.932	C-92	Panama	0.937	0.063
C-27	Colombia	0.862	0.138	C-93	Paraguay	0.969	0.031
C-28	Costa Rica	0.877	0.123	C-94	Peru	0.796	0.204
C-29	Croatia	0.541	0.459	C-95	Philippines	0.838	0.162

C-30	Cyprus	0.054	0.946	C-96	Poland	0.400	0.600
C-31	Czech Republic	0.178	0.822	C-97	Portugal	0.151	0.849
C-32	Denmark	0.049	0.951	C-98	Qatar	0.575	0.425
C-33	Dominican Republic	0.979	0.021	C-99	Republic of Korea	0.084	0.916
C-34	Ecuador	0.965	0.035	C-100	Republic of Moldova	0.841	0.159
C-35	Egypt	0.986	0.014	C-101	Romania	0.654	0.346
C-36	El Salvador	0.984	0.016	C-102	Russian Federation	0.568	0.432
C-37	Estonia	0.088	0.912	C-103	Rwanda	0.909	0.091
C-38	Ethiopia	0.935	0.065	C-104	Saudi Arabia	0.527	0.473
C-39	Finland	0.078	0.922	C-105	Senegal	0.957	0.043
C-40	France	0.047	0.953	C-106	Serbia	0.706	0.294
C-41	Georgia	0.843	0.157	C-107	Singapore	0.120	0.880
C-42	Germany	0.066	0.934	C-108	Slovakia	0.628	0.372
C-43	Ghana	0.972	0.028	C-109	Slovenia	0.190	0.810
C-44	Greece	0.562	0.438	C-110	South Africa	0.865	0.135
C-45	Guatemala	0.949	0.051	C-111	Spain	0.054	0.946
C-46	Guinea	0.891	0.109	C-112	Sri Lanka	0.938	0.062
C-47	Honduras	0.956	0.044	C-113	Sweden	0.110	0.890
C-48	Hong Kong	0.161	0.839	C-114	Switzerland	0.125	0.875
C-49	Hungary	0.258	0.742	C-115	Tajikistan	0.960	0.040
C-50	Iceland	0.044	0.956	C-116	Thailand	0.587	0.413
C-51	India	0.464	0.536	C-117	Togo	0.947	0.053
C-52	Indonesia	0.893	0.107	C-118	Trinidad and Tobago	0.957	0.043
C-53	Iran (Islamic Republic of)	0.646	0.354	C-119	Türkiye	0.463	0.537
C-54	Iraq	0.905	0.095	C-120	Tunisia	0.898	0.102
C-55	Ireland	0.054	0.946	C-121	Uganda	0.920	0.080
C-56	Israel	0.099	0.901	C-122	Ukraine	0.803	0.197
C-57	Italy	0.106	0.894	C-123	United Kingdom	0.097	0.903
C-58	Jamaica	0.883	0.117	C-124	United Arab Emirates	0.142	0.858
C-59	Japan	0.044	0.956	C-125	United Republic of Tanzania	0.959	0.041
C-60	Jordan	0.895	0.105	C-126	United States of America	0.143	0.857

C-61	Kazakhstan	0.900	0.100	C-127	Uruguay	0.809	0.191
C-62	Kenya	0.966	0.034	C-128	Uzbekistan	0.933	0.067
C-63	Kuwait	0.828	0.172	C-129	Viet Nam	0.674	0.326
C-64	Kyrgyzstan	0.934	0.066	C-130	Yemen	0.866	0.134
C-65	Lao People's Democratic Republic	0.946	0.054	C-131	Zambia	0.945	0.055
C-66	Latvia	0.437	0.563	C-132	Zimbabwe	0.932	0.068

It is clear that from Table 4, 97 countries are assigned to Cluster 1, 35 countries are assigned to Cluster 2. The ranks of countries according to their GII values are given in Table 4 in parenthesis.

Clustering results of countries and GII ranks

Table 4

Cluster	Countries
1	Albania (84), Algeria (115), Angola (127), Argentina (69), Armenia (80), Azerbaijan (93), Bahrain (72), Bangladesh (102), Belarus (77), Benin (124), Bosnia and Herzegovina (70), Botswana (86), Brazil (54), Brunei Darussalam (92), Burkina Faso (120), Burundi (130), Cote d'Ivoire (109), Cambodia (97), Cameroon (121), Chile (50), Colombia (63), Costa Rica (68), Croatia (42), Dominican Republic (90), Ecuador (98), Egypt (89), El Salvador (100), Ethiopia (117), Georgia (74), Ghana (95), Greece (44), Guatemala (110), Guinea (132), Honduras (113), Indonesia (75), Iran (Islamic Republic of) (53), Iraq (131), Jamaica (76), Jordan (78), Kazakhstan (83), Kenya (88), Kuwait (62), Kyrgyzstan (94), Lao People's Democratic Republic (112), Madagascar (106), Mali (126), Mauritania (129), Mauritius (45), Mexico (58), Mongolia (71), Montenegro (60), Morocco (67), Mozambique (123), Myanmar (116), Namibia (96), Nepal (111), Nicaragua (108), Niger (125), Nigeria (114), North Macedonia (66), Oman (79), Pakistan (87), Panama (81), Paraguay (91), Peru (65), Philippines (59), Qatar (52), Republic of Moldova (56), Romania (49), Russian Federation (47), Rwanda (105), Saudi Arabia (51), Senegal (99), Serbia (55), Slovakia (46), South Africa (61), Sri Lanka (85), Tajikistan (104), Thailand (43), Togo (122), Trinidad and Tobago (101), Tunisia (73), Uganda (119), Ukraine (57), United Republic of Tanzania (103), Uruguay (64), Uzbekistan (82), Viet Nam (48), Yemen (128), Zambia (118), Zimbabwe (107)
2	Australia (25), Austria (17), Belgium (26), Bulgaria (35), Canada (15), China (11), Cyprus (27), Czech Republic (30), Denmark (10), Estonia (18), Finland (9), France (12), Germany (8), Hong Kong (14), Hungary (34), Iceland (20), India (40), Ireland (23), Israel (16), Italy (28), Japan (13), Latvia (41), Lithuania (39), Luxembourg (19), Malaysia (36), Malta (21), Netherlands (5), New Zealand (24), Norway (22), Poland (38), Portugal (32), Republic of Korea (6), Singapore (7), Slovenia (33), Spain (29), Sweden (3), Switzerland (1), Türkiye (37), United Kingdom (4), United Arab Emirates (31), United States of America (2)

The countries with the highest GII values are in Cluster 2. These countries in Cluster 2 are mostly upper-income or upper-middle-income countries. In Cluster 2, only India is a low-middle-income country, and Bulgaria, China, Malaysia, and Türkiye are upper-middle-income countries. Therefore, the countries with high-income levels, as well as rank high in terms of innovation indicators. This result is consistent with the study of Eren and Gelmez (2022) that clustered countries by using the GII (2021) data set. The countries in Cluster 1 consist of mostly lower-middle and low-income countries. Therefore, countries with low-income levels are at the bottom regarding innovation indicators. However, countries in Cluster 1 such as Albania, Armenia, Azerbaijan, Argentina, Belarus, Bahrain, Bosnia and Herzegovina, Brazil, Botswana, Brunei Darussalam, Chile, Colombia, Croatia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Georgia, Greece, Guatemala, Indonesia, Iraq, Jamaica, Kazakhstan, Kuwait, Mauritius, Mexico, Montenegro, Namibia, North Macedonia, Oman, Panama, Paraguay, Peru, Republic of Moldova, Romania, Qatar, Saudi Arabia, Russian Federation, Serbia, South Africa, Slovakia, Thailand, Trinidad and Tobago are upper income or upper-middle-income countries.

Türkiye is located in Cluster 2, and Türkiye ranked 37th according to the 2022 GII rankings. Türkiye rose four places compared to the previous year. Per the findings in the report, while Türkiye had the highest performance in the human capital and research index, it showed the lowest performance in the institutions sub-component.

Final cluster prototype

Table 5

Sub-indexes	Cluster 1	Cluster 2
Institutions	50.38571	73.68434
Human capital and research	23.14634	50.74311
Infrastructure	35.81554	57.66978
Market sophistication	25.98683	48.68099
Business sophistication	23.16648	49.61126
Knowledge and technology outputs	15.23638	42.46534
Creative outputs	12.64314	39.36598

Upon examining Table 5, it becomes evident that all variables exhibit their highest values within Cluster 2. Consequently, countries that are members of the second cluster demonstrate superior performance regarding the GII. This result verifies the results presented in Table 4.

Summary statistics of the variables on the clusters is presented in Table 6.

Summary statistics of the variables on the clusters

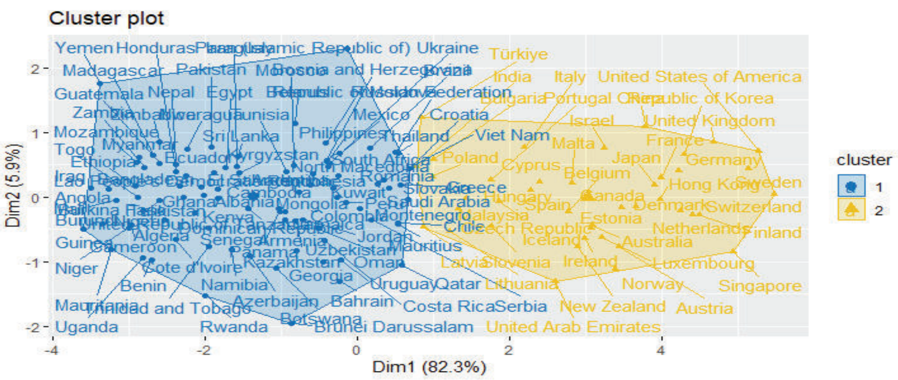
Table 6

Variable	Mean±SD	Median	Min-Max
Institutions	58.07273±14.90591	56.05	17.5 - 95.9
Human capital and research	32.62424± 15.47464	30.7	6-66.4
Infrastructure	43.50455± 12.9991	43.4	17.5- 95.9
Market sophistication	33.8697 ± 14.89748	32.45	4.4-80.8
Business sophistication	31.85379 ± 14.22105	27.15	10.2-69.8
Knowledge and technology outputs	24.46742 ± 15.51748	20.75	1.6-67.1
Creative outputs	21.70379 ± 15.26518	19.4	0.3-56.3

Figure 1 illustrate the distribution of countries in the two clusters.

Distribution of countries into two clusters

Figure 1



4.CONCLUSION AND DISCUSSION

Upon reviewing the clustering outcomes for the countries, it is clear that the two clusters consistently align with the rankings of the GII for the year 2022. Consequently, the reliability of the analytical findings coincides with our analysis. Furthermore, upon scrutinizing the clusters in conjunction with country profiles, it was evident that the employed analyses complemented each other. Countries characterized by high-income levels in Cluster 2 attained the top ranking. These countries also feature prominently among the leading countries in the GII. This observation underscores the correlation between high-income countries and their prominent positions regarding innovation indicators. Cluster 1 comprises primarily low-income and lower-middle countries, which illustrates countries with lower income levels are ordered similarly in the lower echelons of innovation indicators. Türkiye, our country, is in Cluster 2, characterized by high-income and upper-middle-income countries. Türkiye was positioned 41st according to the GII for the year 2021. Türkiye rose four places in 2022 to 37th place. Türkiye entered the top 40 for the first time, climbing 14 places in the Index in the last two years. Türkiye also maintained its 4th place among 36 upper-middle-income countries.

Since fuzzy cluster analysis evaluates the whole data set, it has the chance to reveal some similarities that indices expressing a single numerical value cannot reveal. The cluster analysis results and index rankings are parallel for the countries with high and low GII values. The top countries in GII rankings clustered in Cluster 2. The countries at the bottom of GII rankings clustered in Cluster 1. Cluster analysis is a method based on whole data having the chance to reveal some similarities that indices based on a single numerical value could be incapable. This type of clustering analysis shows the power of the index to reflect the data. Our study reveals the consistency of the rankings according to the GII index.

Future studies may consider comparing the results of these analyses through the application of additional or alternative quantitative methodologies for assessing the innovation performances of countries. Moreover, the measurement of innovation performances could be examined for the diverse categories of countries, such as according to their income levels or other categories that could logically have a relationship or connection with their innovation levels. Furthermore, the various clustering techniques can be compared to each other for different numbers of clusters.

REFERENCES

1. Alqararah, K. and Alnafrah, I. (2023), "Innovation-Driven Clustering for Better National Innovation Benchmarking", *Journal of Entrepreneurship and Public Policy*, 2045-2101. DOI 10.1108/JEPP-01-2023-0007.
2. Al Rahhal, T. and Rencber, O. F. (2022), *Comparisons of K-Means and Fuzzy C-Means Algorithm by Clustering Countries*, Ankara, Turkey: Iksad Publications.
3. Alptekin, N. and Yesilaydin, G. (2015), "OECD Ülkelerinin Sağlık Göstergelerine Göre Bulanık Kümeleme Analizi İle Sınıflandırılması", *Journal of Business Research Turk*, 7(4), pp. 137-155.
4. Aytekin, A., Ecer, F., Korucuk, S. and Karamasa, C. (2022), "Global innovation efficiency assessment of EU member and candidate countries via DEA-EATWIOS multi-criteria methodology", *Technology in Society*, 68, 101896. https://doi.org/10.1016/j.techsoc.2022.101896.
5. Bagdatli Kalkan, S. (2019), "Classification of European Union Member and Candidate Countries According to Ease of Doing Business Index Using Fuzzy Clustering", *Eurasian Econometrics*, *Statistics & Emprical Economics Journal*, 12, pp.25-35. http://dx.doi.org/10.17740/eas.stat.2019-V12-03.
6. Bezdek, J. C. and Hathaway, R. J. (1987), "Clustering with Relational c-Means Partitions From Pairwise Distance Data", *Mathematical Modelling*, 9(6), pp.435-439. https://doi.org/10.1016/0270-0255(87)90509-4.

7. Bulbul, S. and Camkiran, C. (2018), "Bankalarin Klasik ve Bulanik Yaklaşimlarla Siniflandirilmasi", *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 20(2), pp.367-385. https://doi.org/10.26468/trakyasobed.464442.
8. Cebeci, Z. and Yildiz, F. (2015), *"Bulanık C-Ortalamalar Algoritmasının Farklı Küme Büyüklükleri için Hesaplama Performansı ve Kümeleme Geçerliliğinin Karşılaştırılması",* paper presented at the 9. Ulusal Zootekni Bilim Kongresi, Konya, pp.227-239.
9. Eren, H. and Gelmez, E. (2022), "Ülkelerin İnovasyon Performansına Göre Kümelenmesi; Entropi, Copras ve Aras Yöntemleriyle Değerlendirilmesi", *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 12(3), pp.1546-1565. https://doi.org/10.30783/nevsosbilen.1153211.
10. Famalika, A. and Sihombing, P. R. (2021), "Implementation of k-means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018", *Advance Sustainable Science, Engineering and Technology (ASSET)*, 3(1), pp.0210107-1 ~ 0210107-10. https://doi.org/10.26877/asset.v3i1.8461.
11. Ferraro, M. B. and Giordani, P. (2015), "A toolbox for fuzzy clustering using the Rprogramming language", *Fuzzy Sets and Systems*, 279, pp.1-16. http://dx.doi.org/10.1016/j.fss.2015.05.001.
12. Gurtuna, F. and Polat, U. (2020), "Küresel İnovasyon Endeksi Verilerinin Kümeleme Analizi ile Değerlendirilmesi", *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 35(2), pp.551-565. https://doi.org/10.21605/cukurovaummfd.792484.
13. Hancioglu, Y. (2016), "Küresel İnovasyon Endeksini Oluşturan İnovasyon Girdi ve Çıktı Göstergeleri Arasındaki İlişkinin Kanonik Korelasyon Analizi ile İncelenmesi: OECD Örneği", *AİBÜ Sosyal Bilimler Enstitüsü Dergisi*, 16(4), pp.131-157. https://doi.org/10.11616/basbed.vi.455411.
14. Jankowska, B., Matysek-Jędrych, A. and Mroczek-Dąbrowska, K. (2017), "Efficiency of National Innovation Systems – Poland and Bulgaria in The Context of the Global Innovation Index", *Comparative Economic Research*, 20(3), pp.77-94. https://doi.org/10.1515/cer-2017-0021.
15. Kaufman, L. and Rousseeuw, P.J. (1990). *Findings Groups in Data: An Introduction to Cluster Analysis*. John Wiley: USA.
16. Saravananathan, K. and Velmurugan, T. (2018), "Cluster Based Performance Analysis for Diabetic Data", *International Journal of Pure and Applied Mathematics*, 119(16), pp.399-410.
17. Unlu, F. (2019), "Differences in Efficiency of Innovation Performance among Middle-income Countries: An Empirical Approach", *International Journal of Economics and Innovation*, 5(2), pp.213-229. https://doi.org/10.20979/ueyd.567962.
18. Zorlutuna, S. and Erilli, N. A. (2018), "Sosyo-Ekonomik Verilere Göre İllerin Bulanık C-Ortalamalar Yöntemi ile Sınıflandırılması: 2002-2008-2013 Dönemleri Karşılaştırması", *İktisadi Yenilik Dergisi*, 5(2), pp.13-31.
19. WIPO (2022). *Global Innovation Index 2022*. [online] Avaliable at: https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2022-en-main-report-global-innovation-index-2022-15th-edition.pdf [Accesed 22 October 2023].