Anadolu İktisat ve İşletme Dergisi, 10 (2) 2026, 174-198
Anatolian Journal of Economics and Business, 10 (2) 2026, 174-198

Global Logistics Performance Before and After COVID-19: A Longitudinal Cluster-Based Analysis Using LPI Data

Hümeyra GÜNEYSU[1]
Mehdi SAFAEI[2]

Abstract

COVID-19 created major pressure on global supply chains, but it remains uncertain whether this coincided with changes in countries' logistics performance groups. This study investigates this issue using World Bank Logistics Performance Index data for 2016, 2018 and 2023. K-Means clustering was separately carried out on six LPI dimensions for 160 countries in both 2016 and 2018, and 139 countries in 2023. The 2023 edition reflects late-2022 survey responses. For transition analysis, 154 countries were matched in 2016–2018 and 134 in 2018–2023. In the primary three-cluster specification, 130 (84.42%) and 112 (83.58%), respectively, stayed in their former group; the latter period included 15 upward and seven downward movements. Silhouette favored two clusters, and alternative specifications changed some memberships. Five component means improved in the balanced 129-country panel in 2018–2023, while timeliness declined. The findings describe relative persistence and heterogeneous score changes, not a causal pandemic effect.

***Keywords**: Logistics Performance Index, K-Means clustering; COVID-19, longitudinal comparison, country classification*
***JEL Classification**: C38, L91, F14*

COVID-19 Öncesi ve Sonrası Küresel Lojistik Performans: LPI Verileriyle Boylamsal Kümeleme Analizi

Özet

COVID-19 salgını küresel tedarik zincirleri üzerinde ağır baskı oluşturmuştur; ancak bu durumun ülkelerin lojistik performans gruplarındaki değişimlerle örtüşüp örtüşmediği belirsizdir. Bu çalışma, 2016, 2018 ve 2023 Dünya Bankası Lojistik Performans Endeksi verilerini kullanarak bu soruyu incelemektedir. Altı LPI boyutu için K-Ortalamalar kümeleme analizi 2016 ve 2018 yıllarında 160'ar ülkeye, 2023 yılında ise 139 ülkeye ayrı ayrı uygulanmıştır. 2023 raporu, 2022 sonlarında gerçekleştirilen anketin yanıtlarını yansıtmaktadır. Geçiş analizi için 2016–2018 döneminde 154, 2018–2023 döneminde 134 ülke eşleştirilmiştir. Temel üç kümeli modelde sırasıyla 130 (%84,42) ve 112 (%83,58) ülke aynı grupta kalmıştır; ikinci dönemde 15 yükseliş ve yedi düşüş görülmüştür. Silüet ölçütü iki kümeyi desteklemiş, alternatif model seçimleri bazı üyelikleri değiştirmiştir. Dengeli 129 ülkelik panelde 2018–2023 arasında beş boyutun ortalaması yükselirken zamanında teslim puanı düşmüştür. Bulgular, göreli sürekliliği ve farklılaşan puan değişimlerini betimlemekte; pandeminin nedensel etkisini belirlememektedir.

***Anahtar Kelimeler**: Lojistik Performans Endeksi, K-Ortalamalar kümeleme, COVID-19, boylamsal karşılaştırma, ülke sınıflandırması*
***JEL Kodları**: C38, L91, F14*

[1] Graduate of Istanbul Gelisim University, Faculty of Economics, Administrative and Social Sciences, Department of Logistics Management, hguneysu@gelisim.edu.tr, orcid.org/0009-0008-5532-6688
[2] Asst. Prof., Istanbul Gelisim University, Faculty of Economics, Administrative and Social Sciences, Department of Logistics Management, msafaei@gelisim.edu.tr. orcid.org/0000-0003-3740-9486

Research Article / Araştırma Makalesi
Received / Geliş Tarihi: 10.07.2026
Accepted / Kabul Tarihi: 15.09.2026

1. Introduction

The COVID-19 pandemic could be perceived as one of the most significant challenges for the international logistics industry in the last decades. At first, it had been presented as a crisis in the field of healthcare, but shortly afterwards it had become apparent that its consequences were not only related to healthcare. The impacts concerned manufacturing processes, transportation capacity, ports functioning, border regimes, delivery of cargo, as well as freight costs (OECD, 2020). According to Queiroz et al. (2022), epidemic situations may affect all processes in supply chains, since procurement, production, transportation, storage, and customer services will all be simultaneously affected. This impact could be more obvious in the case of logistics services by sea, since in this case the impacts of the virus were unevenly distributed throughout the system (Notteboom et al., 2021).

However, strong operational disruption does not always mean that structural order is changed. Freight rate increase, delivery delays, congestion in ports, and temporary lack of capacity can affect logistics activities very seriously during crisis time. But these effects do not automatically transfer countries to other long-term performance groups. Logistics performance at country level reflects both current assessments and more stable capabilities. These include customs efficiency, transport and trade infrastructure, arranging international shipments easily, competence of logistics services, tracking and tracing systems, and also timeliness. These dimensions are measured in World Bank Logistics Performance Index (LPI), which is commonly used as a benchmark for comparing trade logistics performance among countries (Arvis et al., 2018; Arvis et al., 2023).

This difference is important, because many studies based on LPI are still mostly about ranking, cross-sectional comparison, or classification of one year. These studies can be useful to show relative position of countries; multi-year studies also exist, as discussed in Section 2.1. But they give limited evidence about whether countries stay structurally near to same peer groups after a large global disruption. The 2018 LPI gives the last complete benchmark before pandemic, and the 2023 LPI gives a comparable observation following pandemic onset, based on fieldwork in late 2022. Therefore, it provides a useful situation to examine if COVID-19 period was connected with changes in the classification of logistics performance groups, or if global logistics hierarchy stayed mostly stable despite serious disruption.

In the title, “before and after COVID-19” denotes observations before and after pandemic onset, not a comparison with a survey conducted after the pandemic ended. The 2023 release reflects late-2022 fieldwork; the study is descriptive rather than a causal before–after design.

This study responds to this issue through a longitudinal analysis based on clustering of the World Bank LPI data for 2016, 2018 and 2023. For each edition, K-Means method is conducted separately by using the six dimensions of LPI. Countries are afterward classified into High, Medium, and Low logistics performance groups.

The dataset includes 160 countries for 2016, 160 for 2018 and 139 for 2023. Common samples of 154 countries for 2016–2018 and 134 for 2018–2023 are formed for the transition comparisons; 129 countries appear in all three editions. Based on this matched group, movements of countries from one cluster to another are examined. The comparison describes changes in relative groups; it does not separate pandemic effects from other developments.

The main contribution of the paper comes from moving beyond a simple comparison of country rankings. It studies both stability and movement of logistics performance clusters across editions surrounding pandemic onset. Therefore, the question is not limited to which country received a higher or lower position in ranking. More importantly, the analysis considers whether countries changed their broader performance category. This provides a more structural understanding of global logistics performance. It combines identifier-matched transitions with original-scale score changes and sensitivity checks, without claiming that longitudinal LPI analysis itself is new or that pandemic effects are identified.

2. Literature Background

This section reviews the literature relevant to the comparative assessment of logistics performance across countries. The first subsection considers multi-year LPI analysis and performance benchmarking, while the second subsection reviews clustering and K-Means studies. This structure distinguishes studies examining changes in logistics performance over time from studies classifying countries according to multidimensional logistics characteristics.

2.1. Multi-year LPI Analysis and Performance Benchmarking

The Logistics Performance Index (LPI) remains one of the most widely used international benchmarks for assessing country-level trade logistics. Its six dimensions—customs, infrastructure, international shipments, logistics competence and quality, tracking and tracing, and timeliness—represent both physical and institutional capabilities of national logistics systems (Arvis et al., 2018; Arvis et al., 2023). However, recent research increasingly argues that a single composite ranking cannot fully explain the structural differences among countries. Gürler et al. (2024) showed that country rankings can be sensitive to the choice of indicators, weights, and decision rules, while Stević et al. (2024) demonstrated that alternative multi-criteria methods may produce different but internally consistent views of logistics performance. These studies support using classification and cluster-based approaches together with conventional LPI rankings.

Jonasíková et al. (2025) combined time-series and cluster analysis to examine the evolution and structure of LPI in selected countries over a longer period.

Özdil et al. (2025) compare seven LPI editions from 2007 to 2023 using LOPCOW weighting and EDAS ranking. This is a multi-year benchmarking precedent, not a matched-country cluster-transition analysis. Lukoszová (2023) compares EU countries' 2018 and 2023 LPI scores and percentage changes. Her results provide a

direct temporal comparator, but a change in score need not imply a change in relative cluster membership.

Martí et al. (2017) propose a DEA-based logistics performance index, providing an alternative efficiency-oriented measurement approach. Kabak et al. (2020) investigate interactions between competitiveness and logistics performance using Bayesian networks, partial least squares and importance–performance analysis. These studies motivate attention to measurement choices and national context; neither establishes that the present LPI-only clustering controls those conditions.

2.2. Clustering and K-Means Studies

Altıntaş (2022) classifies 160 countries in the 2018 LPI using K-Means, while Ulkhaq (2023) compares K-Means, K-Medoids and CLARA on that edition. These studies establish relevant single-edition precedents but do not resolve the stability of matched-country classifications across editions.

Recent studies published after the 2023 LPI release have expanded country-level logistics assessment in several directions. Pehlivan et al. (2024) combined TOPSIS and cluster analysis to evaluate G20 countries and showed that ranking and grouping provide complementary information. Kara et al. (2024) developed a decision-support framework for benchmarking the supply chain performance of 72 countries and emphasized delivery time and international tracking indicators as central macro-level performance dimensions. Stević et al. (2024) evaluated 118 countries using entropy weighting and alternative multi-criteria methods, with sensitivity analysis. Together, these contributions show that country performance should be interpreted through profiles and peer groups, not only through ordinal rank positions.

A second stream of recent literature applies clustering and dimensionality-reduction methods to reveal broader country typologies. Delahoz-Domínguez et al. (2025) integrated the six LPI components with governance and human-development indicators for 137 countries, using UMAP and K-Means to identify three statistically distinct country groups. The resulting clusters were supported by a high Hopkins statistic and multivariate validation tests. Visbal-Cadavid et al. (2025) reached a similar three-profile structure through multiple factor analysis and hierarchical clustering, linking logistics capacity with governance and socio-economic development. These studies link logistics performance to wider institutional and development conditions. Their validation results concern different feature sets and do not establish the validity of the six-variable clusters estimated here.

Yıldız et al. (2020) apply hierarchical single-linkage clustering with Euclidean distance to 90 economies selected using 2017 trade volume. They average the six LPI components across 2012, 2014, 2016 and 2018 and identify fourteen clusters. This supports multidimensional logistics profiling before the pandemic, while differing from the present study: averaging editions suppresses the edition-specific movements that matched transitions retain. Their method is neither K-Means nor Ward clustering.

The newest literature also connects logistics performance with environmental sustainability and digital capability. Yardımcı-Coşkun (2026) jointly clustered countries according to the 2023 LPI and 2024 Environmental Performance Index and identified a three-group solution representing high, moderate, and low combined logistics–environmental performance. Haralambides et al. (2026) examined the relationship between digital competitiveness and logistics performance and identified mismatches in which some countries perform better in logistics than their digital capability would predict, while others do not fully convert digital strength into logistics outcomes. These findings are relevant for interpreting tracking and tracing, customs digitalization, and service capability; the present study does not directly measure environmental outcomes or digital competitiveness.

Related evidence published in the Anatolian Journal of Economics and Business also connects transport activity with wider country-level conditions. Using panel data for G7 countries, Erşen and Aslan (2026) examine globalization, energy use and air-freight transportation together; their findings show why transport indicators should be interpreted within a broader economic and environmental setting rather than in isolation.

Post-pandemic resilience studies further clarify why a major operational shock may not immediately produce a complete structural reordering of countries. Pellegrino and Gaudenzi (2026), using evidence from 164 firms, found that digital resources contribute to resilience through operational and relational supply-chain capabilities. This result is consistent with the view that resilience depends on previously accumulated capabilities rather than on short-term crisis responses alone. At the country level, infrastructure, customs institutions, logistics competence, and shipment visibility may also reflect accumulated capacities. Serious disruption may therefore coexist with persistent country groups, but firm-level resilience evidence does not establish a country-level pandemic effect.

2.3. Novelty and Contribution of the Study

The present study addresses this narrower gap by applying the same six-dimensional K-Means design separately to 2016, 2018 and 2023, constructing matched samples of 154 and 134 countries, and evaluating both cluster stability and movement through a transition matrix and centroid comparison. Its contribution is therefore not another ranking of countries, but descriptive evidence on matched-country mobility, supplemented by original-scale comparisons and specification checks.

Despite this rapid development of literature, three limitations remain. First, many recent studies are still cross-sectional and use only the 2023 LPI. Second, the multi-year benchmarking studies discussed here examine trends or rankings rather than the matched transitions estimated in this paper. Third, multidimensional clustering studies often combine LPI with governance, environmental, or development indicators, which is useful for broader classification but addresses a broader construct than a classification based only on LPI components.

Table 1: Positioning of the Present Study within Recent LPI-Based Research

Study	Data	Method	Main contribution	Remaining limitation
Arvis et al. (2016; 2018; 2023)	Global LPI	Benchmark reports	Defines and compares six LPI dimensions	No cluster mobility analysis
Altıntaş (2022)	LPI 2018, 160 countries	K-Means	Single-year country classification	No longitudinal comparison
Ulkhaq (2023)	LPI 2018, 160 countries	K-Means, K-Medoids, CLARA	Compares clustering algorithms	No matched transition analysis
Pehlivan et al. (2024)	G20, LPI 2023	TOPSIS + clustering	Combines ranking and grouping	Restricted to G20 and one year
Gürler et al. (2024)	EU countries, 33 indicators	GA weighting + 11 MCDM methods	Shows sensitivity to indicators and weights	Ranking rather than mobility
Kara et al. (2024)	72 countries	MPSI-ARLON	Macro supply-chain benchmarking	No pre/post cluster tracking
Stević et al. (2024)	118 countries; LPI	Entropy + MCDM methods	Compares rankings and methodological sensitivity	No matched cluster transitions
Jonasíková et al. (2025)	Selected countries, 2007-2022	Time series + clustering	Examines LPI evolution over time	Does not test 2018-2023 matched mobility
Delahoz-Domínguez et al. (2025)	137 countries, 2023	UMAP + K-Means	Integrates logistics, governance, and HDI	Cross-sectional and multidimensional
Visbal-Cadavid et al. (2025)	Global LPI, WGI, HDI	MFA + hierarchical clustering	Identifies three development profiles	No pre/post transition matrix
Yardımcı-Coşkun (2026)	2023 LPI + 2024 EPI	K-Means + validity metrics	Links logistics with sustainability	Single-period joint index analysis
Özdil et al. (2025)	Seven LPI editions, 2007–2023	LOPCOW + EDAS	Multi-year EU logistics benchmarking	Ranking, not matched cluster transitions
Lukoszová (2023)	EU; LPI 2018 and 2023	Score and percentage-change comparison	Direct temporal comparison	Score change differs from cluster mobility
Yıldız et al. (2020)	90 economies; mean 2012/14/16/18 scores	Single-linkage; Euclidean distance	Fourteen pre-pandemic profile clusters	Pooled periods, not edition-specific transitions
Martí et al. (2017)	Logistics performance index	DEA	Alternative efficiency-based index	Not a matched mobility design
Kabak et al. (2020)	Competitiveness and LPI	Bayesian networks; PLS; IPMA	Examines national competitiveness context	Does not control context in this study
This study	LPI 2016, 2018 and 2023	K-Means + transition analysis	Tracks 154 and 134 matched countries; original-scale and sensitivity checks	Descriptive; sparse editions and specification sensitivity

3. Data and Methodology

3.1. Data and Sample

This study uses the country-level Logistics Performance Index (LPI) datasets published by the World Bank for 2016, 2018 and 2023. The year 2018 is treated as the last pre-pandemic benchmark, 2016 supplies an earlier comparison, and the 2023 edition reflects survey responses collected between 6 September and 5 November 2022 (Arvis et al., 2023). The analysis is based on the six standard LPI dimensions: customs, infrastructure, international shipments, logistics competence and quality, tracking and tracing, and timeliness. These dimensions provide a comparable basis for examining national logistics performance across the selected editions (Arvis et al., 2018; Arvis et al., 2023).

Countries with missing values in any one of the six LPI dimensions were removed from the data before doing clustering analysis. The final year-specific samples included 160 countries in 2016, 160 in 2018 and 139 in 2023. For the longitudinal comparison, only countries appearing in both editions of each pair were retained, using ISO3 identifiers rather than display names. This produced 154 matched countries for 2016–2018 and 134 for 2018–2023; the balanced panel contains 129. Matching removes unmatched countries from transition counts but does not itself remove coverage effects on fitted centroids, so common-sample refits are also examined.

The frozen data snapshot was retrieved from the World Bank API on 31 August 2026 and checked against the corresponding official report tables. API observations dated 2022 correspond to the 2023 LPI edition; the mapping is supported by the documented September–November 2022 survey period. Regional aggregates are excluded. Taiwan, China, which was absent from the API responses, is restored from the edition-specific report tables, not from multi-year aggregate scores. Source URLs, raw responses and the three restored rows are documented in the appendix reproducibility files.

The 2025 release is not appended to this survey panel. The World Bank's LPI 2.0 replaces the perception-based index with operational indicators of connectivity and time in maritime, aviation and postal logistics; its 2025 release draws on 2023–2024 operational data (World Bank, 2026). Its six core indicators are not the six earlier survey dimensions. Adding 2016 expands the comparable survey history, whereas pooling the 2025 operational indicators would change the measured construct. LPI 2.0 can support future external validation, but no artificial conversion is imposed here.

3.2. Data Preparation

The data preparation process included missing-value removal and z-score standardization. Potential outliers were examined using the interquartile range rule, but valid published scores were retained in the primary analysis. Replacing values outside the 1.5-IQR fences by the mean of values inside them would change the 2018 customs scores of Germany, Japan and Sweden to 2.6468; this earlier

adjustment is retained only for sensitivity and historical reconciliation. Each component is centered by its edition-specific sample mean and divided by its sample standard deviation (n−1 denominator). Cluster profiles are reported in z-scores and original 1–5 scores; min–max normalization is not required for this interpretation.

All 459 country-edition profiles were checked against the official report tables. Three 2018 component values differ by 0.01 between API and report; these differences and precision sensitivity are documented in Appendix Section A3. No missing component is imputed, regional aggregates are excluded, and the overall LPI score is not a seventh clustering variable.

3.3. K-Means Clustering

K-Means clustering was applied separately to the 2016, 2018 and 2023 datasets. Each country was represented by a six-dimensional standardized LPI vector. The algorithm partitions countries into non-overlapping clusters by minimizing the within-cluster sum of squared distances between each country and the centroid of its assigned cluster (MacQueen, 1967). The objective function is defined as follows:

$$\arg\min_{C_1,\ldots,C_k} \sum_{r=1}^{k} \sum_{z_i \in C_r} \|z_i - \mu_r\|^2 \quad (1)$$

where z_i denotes the standardized LPI vector of country i, C_r is cluster r, and μ_r is the centroid of cluster r. The algorithm assigns each country to the nearest centroid, updates the centroids, and repeats this process until the cluster memberships become stable.

The primary implementation uses Python 3.12.13 and scikit-learn 1.8.0, k-means++ initialization, the Lloyd algorithm, random_state=123, n_init=100, max_iter=300 and tol=0.0001. The best within-cluster objective is retained. These are explicit settings for the revised analysis, not inferred settings of the unavailable original R execution. Clustering uses all six standardized components; PCA is used afterward only for visualization.

3.4. Cluster Validation

The number of clusters was assessed using the Elbow method and Silhouette analysis. The Elbow method was used to inspect the reduction in within-cluster variation as the number of clusters increased from two to eight. Silhouette analysis was then used to assess cluster separation and internal consistency (Rousseeuw, 1987). For each country i, the silhouette coefficient is defined as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

where a(i) is the average distance between country i and all other countries in the same cluster, while b(i) is the minimum average distance between country i and countries in the nearest alternative cluster.

The average silhouette coefficient is calculated as follows:

$$\overline{s} = \frac{1}{n}\sum_{i=1}^{n} s\,(i) \quad (3)$$

where n is the number of countries included in the clustering analysis. Values near 1 show higher separation, values near 0 show overlapping clusters, and negative values indicate a closer alternative cluster. A three-cluster solution is retained as an interpretable Low–Medium–High resolution, not as the silhouette optimum. Two- and four-cluster solutions and Ward hierarchical clustering are examined as alternatives.

Robustness checks include common-country and balanced-panel refits, alternative scaling and preprocessing, fixed 2018 scaling and centers, 25 seeds, 200 country-bootstrap samples per edition, omission of each country, and 200 rounding-perturbation scenarios for 2023. Appendix section A2 gives the settings, adjusted Rand indices and country-level stability interpretation. These diagnose classification sensitivity, not survey sampling uncertainty or causal effects.

3.5. Cluster Labelling and Transition Analysis

After clustering, the groups were labelled as High, Medium, and Low logistics performance clusters according to their average original-scale LPI profiles. For each cluster r, the average original-scale profile is calculated as follows:

$$P_r = \frac{1}{6}\sum_{j=1}^{6} \bar{x}_{rj} \quad (4)$$

where $\bar{x}_{rj}$ is the average original-scale value of cluster r on LPI dimension j. The cluster with the highest value of P_r was labelled High, the cluster with the intermediate value was labelled Medium, and the cluster with the lowest value was labelled Low. Therefore, the labels were assigned after clustering and were not imposed in advance.

The transition matrix was constructed using the matched countries in each pair of editions. It compares each country's cluster membership in the earlier edition t with its cluster membership in the later edition u. Let S_{tu} be their common country set and g_i^t denote the cluster label of country i in edition t. Each element of the transition matrix is calculated as follows:

$$N_{ab}(t,u) = \sum_{i \in S_{tu}} 1(g_i^t = a \wedge g_i^u = b) \quad (5)$$

where a, b ∈ {High, Medium, Low}, $|S_{tu}|$ = 154 for 2016–2018 and 134 for 2018–2023, and 1(·) is an indicator function equal to 1 when the condition is true and 0 otherwise. The diagonal elements show countries that remained in the same logistics performance group, while the off-diagonal elements show countries that moved between clusters.

The two- and five-year edition gaps are unequal; transition shares are not annualized or interpreted as equal-duration risks. Original-scale changes are calculated separately for the same 129 countries, preventing shifts in independently standardized centers from being read as absolute score deterioration.

The empirical outputs are reported through two figures and four tables. Figure 1 presents the PCA visualization of the 2016, 2018 and 2023 cluster structures. Table 2 reports the cluster sizes and standardized cluster centers for all three editions. Figure 2 compares the cluster centers across the six LPI dimensions. Table 3 reports the transition matrices for the matched samples, and Table 4 reports original-scale changes in the balanced panel. Detailed diagnostics and historical reconciliation are provided in the Appendix.

4. Results

4.1. Cluster Validity

The cluster validation results show that three clusters provide a descriptive classification for 2016, 2018 and 2023, but do not establish three natural classes. The overall silhouette coefficients are 0.4849, 0.4489 and 0.4322, respectively. The corresponding two-cluster values are higher: 0.5773, 0.5767 and 0.5645. Figure A1 reports the Elbow and silhouette diagnostics for k=2–8. For logistics data at country level, countries do not form completely separated groups; they are located on a wider continuum of logistics performance. Results for alternative cluster resolutions are therefore reported in Section 4.4 and the Appendix.

This pattern can also be seen in Figure 1, where the K-Means solutions obtained from the six standardized LPI dimensions are projected onto the first two principal components for visualization. The first principal component captures 94.36% of total variation in 2016, 92.26% in 2018 and 90.78% in 2023. Cluster labels are assigned from the original-scale component means, not from PCA coordinates. More specifically, the Low, Medium, and High clusters occupy the left-to-right direction of this axis. The Medium cluster shows overlap and within-group heterogeneity. Nevertheless, the High and Low clusters can be easily distinguished by two opposite directions.

PCA axes are estimated separately, so their signs and coordinates are not a common longitudinal scale. Visual separation does not establish causal or statistical independence of clusters.

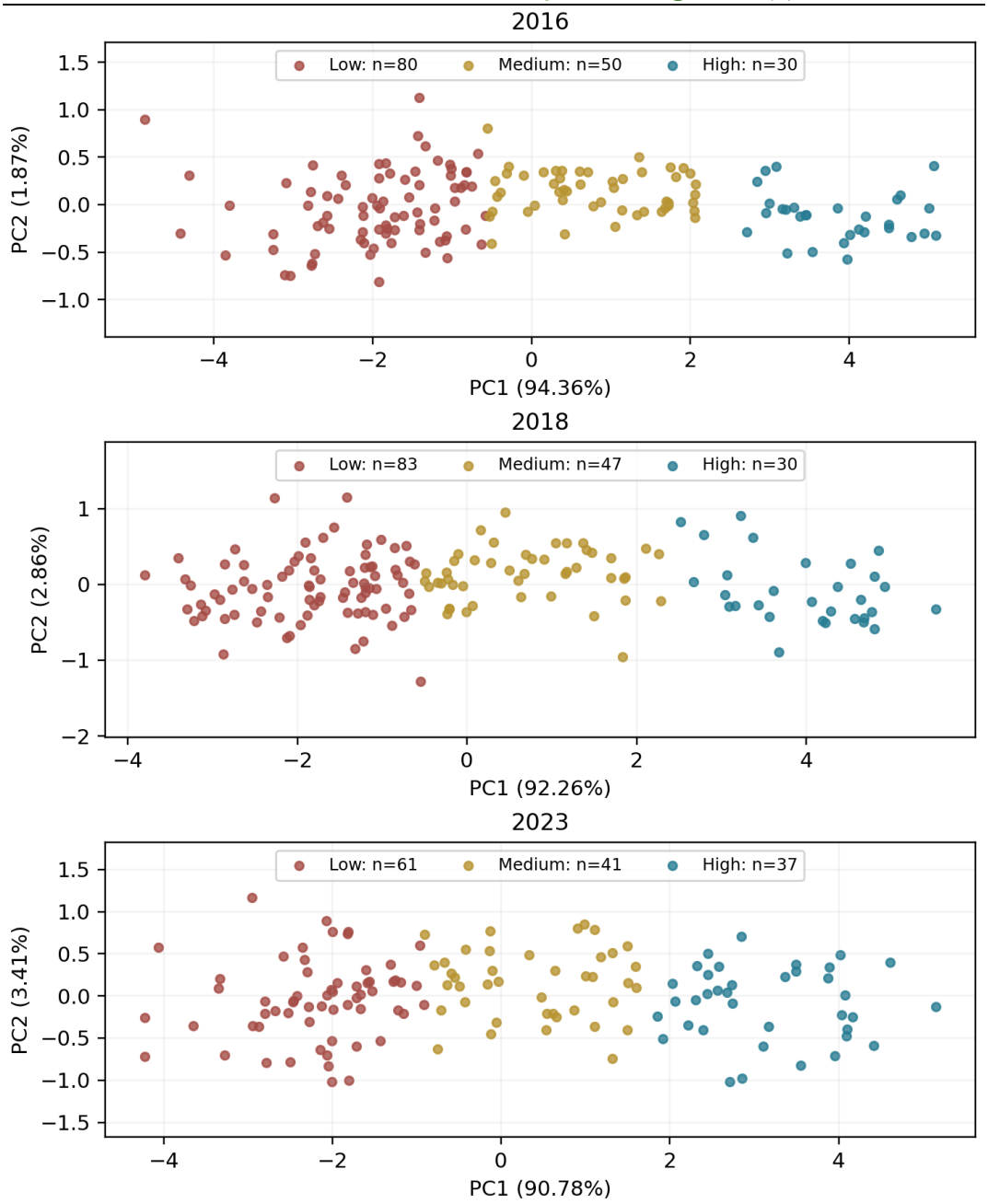


Figure 1. PCA Visualization of K-Means Clusters for 2016, 2018 and 2023

Source: Authors' calculations based on World Bank LPI data (Arvis et al., 2016; Arvis et al., 2018; Arvis et al., 2023). PCA visualizes the fitted six-dimensional classifications.

4.2. Cluster Structures and Cluster Centers

K-means clusters showed that there were three categories of logistics performance for each year under consideration. In 2018, the number of countries in the high, medium, and low clusters was 30, 47, and 83, respectively. In 2023, the High cluster became larger and reached 37 countries. The Medium cluster reduced to 41

countries, while the Low cluster also decreased to 61 countries. In 2016, the corresponding counts were 30, 50 and 80. These are full-sample group sizes, not matched-country movements, because country coverage differs between editions. However, the general three-level structure of global logistics performance remained visible in all three editions.

Table 2 reports the cluster sizes and standardized cluster centers for 2016, 2018 and 2023. The High cluster remained above the year-specific global average across all six LPI dimensions in all three editions, while the Low cluster remained below the average in all dimensions. The Medium cluster stayed close to the global average, with positive but smaller centers across all six indicators.

Table 2. Cluster Sizes and Standardized Average Profiles, 2016, 2018 and 2023

Edition	Group	n	CUS	INF	SHP	CMP	TRK	TIM
2016	High	30	1.604	1.647	1.508	1.651	1.569	1.463
2016	Medium	50	0.274	0.264	0.388	0.264	0.353	0.419
2016	Low	80	-0.773	-0.782	-0.808	-0.784	-0.809	-0.810
2018	High	30	1.651	1.689	1.522	1.697	1.632	1.532
2018	Medium	47	0.210	0.237	0.357	0.262	0.265	0.365
2018	Low	83	-0.715	-0.745	-0.752	-0.762	-0.740	-0.760
2023	High	37	1.323	1.346	1.165	1.333	1.309	1.245
2023	Medium	41	0.082	0.102	0.259	0.141	0.192	0.184
2023	Low	61	-0.858	-0.885	-0.881	-0.903	-0.923	-0.879

Note: Values are z-score standardized cluster centers. Positive values indicate above-average performance relative to the year-specific global mean, while negative values indicate below-average performance. CUS: customs; INF: infrastructure; SHP: international shipments; CMP: logistics competence and quality; TRK: tracking and tracing; TIM: timeliness. Source: Authors' calculations based on World Bank LPI data (Arvis et al., 2016; Arvis et al., 2018; Arvis et al., 2023).

Figure 2 shows that the High cluster declined relatively across all six LPI dimensions between 2018 and 2023, although it remained clearly above the global average. The largest relative declines were observed in logistics competence and international shipments. The Medium cluster showed lower standardized values across all six dimensions. The Low cluster moved further below the global average in all six dimensions, especially in tracking and tracing. Because edition samples and standardization differ, these are relative profile comparisons, not changes in the absolute performance of a fixed group of countries.

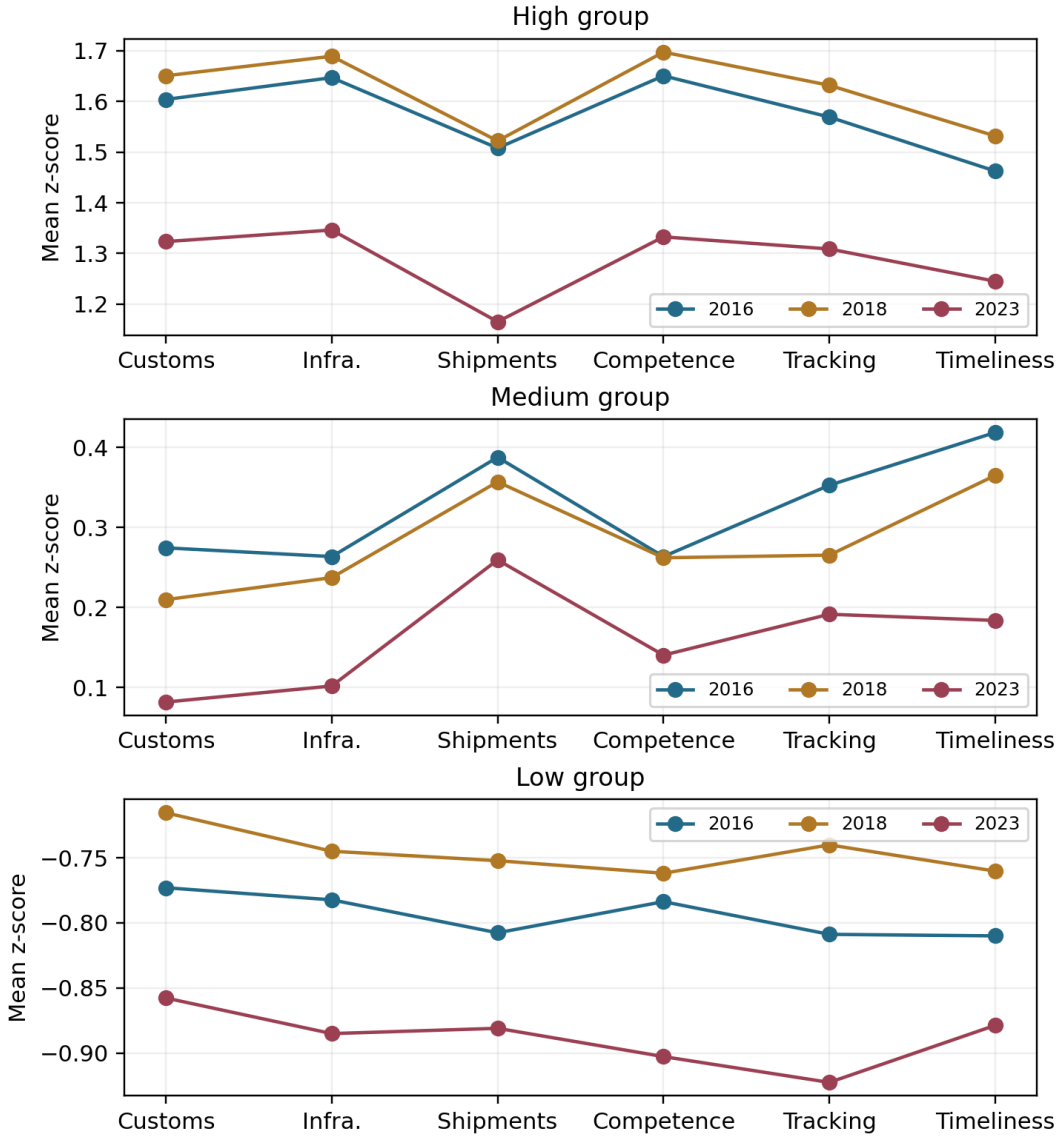


Figure 2. Cluster Center Comparison Across LPI Dimensions

Source: Authors' calculations based on World Bank LPI data (Arvis et al., 2016; Arvis et al., 2018; Arvis et al., 2023). Edition-specific samples are standardized separately.

Table 4 instead compares original-scale component means for the same 129 economies in all three editions. From 2018 to 2023, customs rise by 0.0809, infrastructure by 0.1246, international shipments by 0.0417, competence by 0.1419 and tracking by 0.0910, while timeliness falls by 0.0750. The official overall index rises by 0.0709. These descriptive changes contradict a blanket claim of absolute deterioration across all dimensions. They do not indicate uniform improvement: for example, 50 economies improve in timeliness and 79 decline.

4.3. Transition Analysis

The transition analysis provides the main evidence on changes in the classification of global logistics performance groups. This analysis is based on the 134 countries observed in both 2018 and 2023. Table 3 reports the transition matrix, together with the additional 154-country comparison for 2016–2018.

Table 3. Transition Matrix, 2016–2018 and 2018–2023, Matched Samples

Interval	Earlier group	Later High	Later Medium	Later Low	Total
2016–2018	High	27	3	0	30
2016–2018	Medium	3	35	9	47
2016–2018	Low	0	9	68	77
2016–2018	Total	30	47	77	154
2018–2023	High	28	2	0	30
2018–2023	Medium	9	30	5	44
2018–2023	Low	0	6	54	60
2018–2023	Total	37	38	59	134

Note: Diagonal cells show countries which stayed in the same logistics performance group. Off-diagonal cells show countries which moved between clusters. Rows denote the earlier edition and columns the later edition. Source: Authors' calculations based on World Bank LPI data (Arvis et al., 2016; Arvis et al., 2018; Arvis et al., 2023).

The transition matrix indicates high level of stability. From 134 matched countries, 112 countries remained in the same logistics performance group from 2018 to 2023. This means a stability rate of 83.58%. Twenty-two countries changed their cluster, which is equal to mobility rate of 16.42%. Some upward movements were observed. Nine countries moved from Medium to High cluster: Bahrain, Estonia, Greece, Iceland, Israel, Latvia, Malaysia, South Africa, and Thailand. Also, six countries moved from Low to Medium, including Honduras, North Macedonia, Peru, Solomon Islands, Sri Lanka, and Uruguay.

Downward movements were not very large. Czechia and Portugal moved from High to Medium. Also, Iran, Kazakhstan, Paraguay, Russian Federation, and Ukraine moved from Medium to Low. It is important to note that these differences reflect relative differences in logistics performance groups. They are not causal effects. Economic and geopolitical factors are possible explanations, but they are not tested here, and clustering does not imply causation.

In 2016–2018, 130 of 154 matched countries remained in the same group (84.42%), with 12 upward and 12 downward movements. In balanced-panel refits, stability is 109/129 (84.50%) for 2016–2018 and 107/129 (82.95%) for 2018–2023; the latter has 13 upward and nine downward movements. These estimates use their stated samples and do not replace the primary 2018–2023 estimate of 112/134 (83.58%).

Table 4. Original-Scale Means and Changes in the Balanced Panel (N=129)

Component	2016	2018	2023	Δ16–18	Δ18–23
Customs	2.7743	2.7548	2.8357	-0.0195	+0.0809
Infrastructure	2.8400	2.8320	2.9566	-0.0080	+0.1246
Shipments	2.9316	2.9133	2.9550	-0.0183	+0.0417
Competence	2.9061	2.9108	3.0527	+0.0047	+0.1419
Tracking	2.9535	2.9896	3.0806	+0.0361	+0.0910
Timeliness	3.3478	3.3362	3.2612	-0.0116	-0.0750
Overall	2.9619	2.9586	3.0295	-0.0032	+0.0709

Source: Authors' calculations for the same 129 countries in both intervals. Changes are later minus earlier and computed before display rounding. The overall index is descriptive only, not a clustering input.

Generally speaking, there are three empirical patterns that can be observed from these results. First, the three-cluster configuration was present in all three editions. Second, most matched countries stayed in their same performance group, and this shows limited mobility between clusters. Third, cluster centers changed in unequal way across the six LPI dimensions. There was relative decline in High and Low clusters, while the Medium cluster remained closer to the edition-specific mean. These findings make the empirical basis for the next discussion.

4.4. Robustness Checks

For 2018–2023, the common-pair refit reproduces the primary labels for all 134 countries. Persistence is 81.34% with fixed 2018 scaling and centers, 79.10% with median/IQR scaling, 89.55% with two K-Means clusters, 73.13% with four, and 58.96% with Ward's three-cluster partition. These results show material specification sensitivity rather than method invariance. Seed, bootstrap, country-omission and rounding diagnostics are reported in Tables A2–A3 and Figure A1; historical reconstruction is kept separately in Section A1. The 2023 original memberships and printed centers are preserved; changes to the 2018 fit follow the documented revised primary specification.

5. Discussion

The results make clear that there is an important difference between disruption in operations and changes in country classifications. The pandemic placed considerable pressure on global logistics systems. But still, based on the results, it seems that country performance groups according to LPI remained rather stable. This is not unexpected. LPI is not only showing short-term logistics conditions. It also considers customs efficiency, infrastructure quality, logistics skills and competence, tracking systems, and reliability of delivery (Arvis et al., 2018; Arvis et al., 2023).

As it was demonstrated by the case of COVID-19 pandemic, this issue was not restricted by transport sector alone. Instead, it became an issue of a much bigger system. In particular, it could impact multiple nodes of the logistics chain

simultaneously. From the availability of raw materials to managing inventory and finally delivering products to the end consumers, multiple issues could arise (Queiroz et al., 2022). In this way, it could be regarded as a full system shock. As for maritime logistics sector, one should admit the presence of asymmetric shocks. In particular, various ports experienced different levels of pressure. Thus, some ports operated under tremendous pressure due to the increased number of vessels arriving at them. At the same time, other ports faced less pressure. This consideration becomes important when considering pandemics (Notteboom et al., 2021).

Although reports from institutions like UNCTAD (2021) correctly mention many delays and inefficiencies in international trade, results of this study show another side. Under the primary three-cluster specification, 83.58% of matched countries remained in the same performance group between 2018 and 2023. This relative persistence can coexist with substantial operational disruption; it does not demonstrate that the pandemic had little effect or functioned only as a stress test. The 2016–2018 comparison supplies an earlier benchmark, not a counterfactual.

Interpretation of cluster centers requires caution. The High-Performance cluster remained clearly above the year-specific global average in all six LPI dimensions. However, its standardized centroid values were lower in 2023 than in 2018, indicating a narrower relative distance from the global average. Because the two years were standardized separately, this pattern should be interpreted as a relative change in cluster position rather than direct evidence of absolute deterioration caused by the pandemic. In contrast, the standardized centers of the Low-Performance cluster became more negative in all six dimensions, indicating that this group moved further below the year-specific global average. The analysis does not identify the causes of these relative differences. The Medium cluster remained closer to the mean, although its standardized centers declined in all six dimensions. Table 4 instead compares the same 129 countries on the original score scale: five component means increased in 2018–2023, while timeliness declined.

The transition results also should be read with caution. Some countries moved upward, and some others moved downward. But clustering results only indicate relative changes in logistics performance groups. They do not show causal mechanisms behind these movements. The cases of countries moving downward can be connected with different country-specific conditions, such as economic difficulties, geopolitical disturbances, limits in operation, or some other national factors. These possible explanations are not directly examined in this paper. Therefore, movement between clusters cannot be understood as causal evidence showing that COVID-19 alone produced these changes. The evidence provided here has mainly comparative and structural character. A large majority of countries stayed within their earlier performance category, although the internal characteristics of the clusters did not change equally across the six LPI dimensions.

The multi-year rankings of Özdil et al. (2025) and the 2018–2023 score comparisons of Lukoszová (2023) help interpret these results. Czechia and Portugal

move downward in the revised classification, consistent in direction with their score declines discussed by Lukoszová. Hungary remains Medium in both primary fits, illustrating that percentage score changes and relative category movements are not equivalent. Yıldız et al. (2020) instead group averaged pre-pandemic profiles; the present approach preserves edition-specific movements, but the Ward sensitivity shows that longitudinal comparison does not remove algorithm dependence.

From a logistics policy perspective, the results suggest that post-pandemic recovery should not be viewed only as rapid correction of the immediate weaknesses appearing in logistics systems. For countries in low and middle performance groups, more attention is needed for improving logistics performance through customs modernization, development of transport infrastructure, digital tracking capability, logistics service quality, and shipment reliability. Middle category countries may also focus on some specific dimensions where improvement is more possible, such as shipment planning systems and visibility of shipments. Countries in high performance group also need continuous adaptation of their logistics systems, rather than interpret their relative standing as protection against disruption.

These are conditional diagnostic priorities, not estimated policy effects or returns to investment. Local operational data, feasibility and each country's component profile should guide action; the balanced-panel fall in timeliness particularly motivates examining delivery reliability.

6. Conclusion

This study examined whether the period during COVID-19 pandemic was linked to changes in the classification of countries according to their logistics performance. For this purpose, World Bank LPI datasets for 2016, 2018 and 2023 were utilized to group the countries into three clusters based on their logistics performance on six LPI dimensions. K-means clustering technique was applied to each edition, and matched samples of 154 and 134 countries were used for the two transition comparisons.

The main result obtained is that most matched countries stayed in the same clusters under the primary three-cluster specification: 130 of 154 (84.42%) in 2016–2018 and 112 of 134 (83.58%) in 2018–2023. The latter comparison contains 15 upward and seven downward movements. At the same time, even without changing their logistics performance cluster, countries can experience score changes. The standardized centers of the High-Performance cluster moved closer to the year-specific global average while remaining clearly above it, whereas the Low-Performance cluster moved further below the year-specific global average. In the balanced 129-country panel, five original-scale component means improved in 2018–2023 and timeliness declined. These are descriptive findings, not statistically established pandemic effects; silhouette favors two clusters and alternative specifications change some memberships.

The contribution is the joint reporting of identifier-matched transitions, original-scale component changes and sensitivity to sample coverage and specification. This

extends the comparison beyond rankings without claiming that longitudinal LPI research or clustering itself is novel.

In conclusion, it needs to be highlighted that logistics policy can consider structural capacity building. Indeed, low-performing countries should focus on improving logistics performance through customs services modernization, development of transport/trade infrastructure and logistics service quality, as well as improving tracking, tracing, and shipments' reliability. Middle-performing countries could focus on upgrading one or several dimensions of logistics performance, while high-performing countries should continue with their efforts to adapt their logistics systems.

These priorities should be assessed against local operational evidence and feasibility; broad cluster membership alone cannot establish which intervention will improve performance. However, this study has some important limitations. First, it is limited to three sparse and unequally spaced LPI editions – 2016, 2018 and 2023; the last reflects late-2022 fieldwork. Second, the LPI itself is partially based on perception. Third, K-means clustering is subject to scaling issues and sensitivity to outliers and initial values. Country coverage and published score precision also differ. The balanced panel, alternative scaling and country-influence checks address some sensitivities but do not control omitted national economic or geopolitical factors or estimate survey confidence intervals. Finally, the paper identifies cluster shifts but cannot provide information on their causes. Future research may address all of these issues using explanatory panel designs, k-medoids clustering and further hierarchical comparisons, as well as other independent variables such as GDP per capita, income group, region, port performance, freight rate, digital customs clearance, logistics resilience indexes.

References

Altıntaş, F. F. (2022), "Examination of Countries in the Scope of Logistics Performance Index with K-Means Cluster Analysis", *Journal of Yaşar University*, 17(68): 878-891.

Arvis, J. F., Ojala, L., Shepherd, B., Ulybina, D., Wiederer, C. (2023), *Connecting to Compete 2023: Trade Logistics in an Uncertain Global Economy—The Logistics Performance Index and Its Indicators*, World Bank, Washington, DC.

Arvis, J. F., Ojala, L., Wiederer, C., Shepherd, B., Raj, A., Dairabayeva, K., Kiiski, T. (2018), *Connecting to Compete 2018: Trade Logistics in the Global Economy—The Logistics Performance Index and Its Indicators*, World Bank, Washington, DC.

Arvis, J.-F., Saslavsky, D., Ojala, L., Shepherd, B., Busch, C., Raj, A., Naula, T. (2016), *Connecting to Compete 2016: Trade Logistics in the Global Economy—The Logistics Performance Index and Its Indicators*, World Bank, Washington, DC. https://wb-lpi-media.s3.amazonaws.com/LPI_Report_2016.pdf, (31. 08. 2026).

Delahoz-Domínguez, E., Mendoza-Mendoza, A., Visbal-Cadavid, D. (2025), "Clustering of Countries Through UMAP and K-Means: A Multidimensional Analysis of Development, Governance, and Logistics", *Logistics*, 9(3): 108.

Erşen, D., Aslan, A. (2026), "G7 Ülkelerinde Küreselleşme, Enerji Tüketimi ve Hava Taşımacılığının CO_2 Emisyonları Üzerindeki Etkileri: Panel Veri Analizi", *Anadolu İktisat ve İşletme Dergisi*, 10(1): 42–59.

Gürler, H. E., Özçalıcı, M., Pamucar, D. (2024), "Determining Criteria Weights with Genetic Algorithms for Multi-Criteria Decision Making Methods: The Case of Logistics Performance Index Rankings of European Union Countries", *Socio-Economic Planning Sciences*, 91: 101758.

Haralambides, H., Akdamar, E., Akgül, E. F. (2026), "Revealing Patterns of Technology and Commerce: Evidence from Digital Competitiveness and Logistics Performance", *The Asian Journal of Shipping and Logistics*, 42(2): 69-83.

Jonasíková, D., Konečný, V., Zuzaniak, M. (2025), "Evolution of Logistics Performance Index and Their Structure in Selected Countries", *Transportation Research Procedia*, 87: 217-231.

Kabak, Ö., Önsel Ekici, Ş., Ülengin, F. (2020), "Analyzing Two-Way Interaction Between the Competitiveness and Logistics Performance of Countries", *Transport Policy*, 98: 238-246.

Kara, K., Yalçın, G. C., Simic, V., Baysal, Z., Pamucar, D. (2024), "The Alternative Ranking Using Two-Step Logarithmic Normalization Method for Benchmarking the Supply Chain Performance of Countries", *Socio-Economic Planning Sciences*, 92: 101822.

Lukoszová, X. (2023), *"Changes in Logistics Performance Indexes of European Union Countries for 2023 to the Period Before the Covid-19 Pandemic",* 17th International Scientific Conference INPROFORUM: Challenges and Opportunities in the Digital World: 71–77, České Budějovice, Czechia.

MacQueen, J. (1967), "Some Methods for Classification and Analysis of Multivariate Observations", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 281-297, University of California Press, Berkeley.

Martí, L., Martín, J. C., Puertas, R. (2017), "A DEA-Logistics Performance Index", *Journal of Applied Economics*, 20(1): 169-192.

Notteboom, T., Pallis, T., Rodrigue, J.-P. (2021), "Disruptions and Resilience in Global Container Shipping and Ports: The COVID-19 Pandemic versus the 2008-2009 Financial Crisis", *Maritime Economics & Logistics*, 23(2): 179-210.

OECD (2020), *"COVID-19 and International Trade: Issues and Actions",* OECD Policy Responses to Coronavirus (COVID-19), Policy Paper.

Özdil, Z., Kartal, C., Acar Karaboğa, M. (2025), "Analysis of Logistics Performance of EU 27 Countries with LOPCOW and EDAS Methods", *Third Sector Social Economic Review*, 60(1): 681–693.

Pehlivan, P., Aslan, A. I., David, S., Bacalum, S. (2024), "Determination of Logistics Performance of G20 Countries Using Quantitative Decision-Making Techniques", *Sustainability*, 16(5): 1852.

Pellegrino, R., Gaudenzi, B. (2026), "Supply Chain Resilience as a Viable Way to Cope with Disruptions: An Empirical Analysis of Italian Firms", *Review of Managerial Science*, 20(5): 1459–1489.

Queiroz, M. M., Ivanov, D., Dolgui, A., Fosso Wamba, S. (2022), "Impacts of Epidemic Outbreaks on Supply Chains: Mapping a Research Agenda amid the COVID-19 Pandemic through a Structured Literature Review", *Annals of Operations Research*, 319(1): 1159-1196.

Rousseeuw, P. J. (1987), "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis", *Journal of Computational and Applied Mathematics*, 20: 53-65.

Stević, Ž., Ersoy, N., Başar, E. E., Baydaş, M. (2024), "Addressing the Global Logistics Performance Index Rankings with Methodological Insights and an Innovative Decision Support Framework", *Applied Sciences*, 14(22): 10334.

Ulkhaq, M. M. (2023), "Clustering Countries According to the Logistics Performance Index", *JATISI: Jurnal Teknik Informatika dan Sistem Informasi*, 10(1): 1010-1018.

UNCTAD (2021), *Review of Maritime Transport 2021*, United Nations, Geneva.

Visbal-Cadavid, D., Delahoz-Domínguez, E., Mendoza-Mendoza, A. (2025), "A Multiple Factor Analysis and Hierarchical Clustering of Global Logistics Governance and Development", *Decision Analytics Journal*, 15: 100579.

World Bank (2026), *The New Logistics Performance Indicators 2.0 (LPI 2.0): Methodology and User Guide*, World Bank, Washington, DC. https://openknowledge.worldbank.org/server/api/core/bitstreams/5bd796fe-7716-4718-9466-375ba28a66e8/content, (31. 08. 2026).

Yardımcı-Coşkun, İ. (2026), "Towards Sustainable Logistics: A Clustering-Based Analysis of Global LPI and EPI Scores with a Focus on European Nations", *European Transport Research Review*, 18: 36.

Yıldız, A., Aydoğan, K., Kartum, G. (2020), "Türkiye'nin Uluslararası Lojistik Performans Endeksindeki Konumunun Kümeleme Analizi ile Araştırılması", *Turkish Studies–Social*, 15(3): 1659–1679.

Author Contributions

Hümeyra GÜNEYSU: Conceptualization, Literature review, Data curation, Methodology, Resources, Analysis, Validation, Visualization, Writing.

Mehdi SAFAEI: Conceptualization, Literature review, Data curation, Methodology, Resources, Analysis, Validation, Visualization, Writing.

Appendix

This appendix supports the revised manuscript with unchanged original title. It distinguishes the revised primary analysis from historical reconciliation and documents the full sensitivity evidence. The primary 2018–2023 result in the article is 112/134 (83.58%); no historical estimate replaces it.

A1. Historical Reconciliation

To preserve valid earlier findings, the original country lists were reconstructed separately from the revised primary fits. The 2023 reconstruction reproduces all 139 labels and all 18 printed center components, with silhouette 0.432161 and PC1 90.778%. Applying the disclosed old preprocessing to the original 2018 memberships gives silhouette 0.422486 and PC1 91.046%, consistent with the former rounded reports of 0.42 and 91.0%. Ten of the 18 printed 2018 center components match at three decimals; the maximum unrounded discrepancy is approximately 0.00109. The remaining discrepancies are recorded rather than assigned an unverified cause.

Keeping the original labels but correcting the name match yields 114 persistent economies among 134 (85.07%), with ten upward and ten downward movements. Its matrix is [30, 3, 0; 7, 32, 7; 0, 3, 52] in High–Medium–Low order. This is a historical reconciliation, not the primary Table 3 estimate. The former 112/132 result is recovered by exact-name matching, which excludes Türkiye and North Macedonia. Both were Medium-to-Medium under the original labels.

The original input file and executed R code were not recovered. The independently implemented old transform with the revised Python engine yields 32/51/77 groups in 2018, rather than the original 33/52/75. Consequently, exact replication of the historical program is not asserted. The revised primary analysis is fully specified and reproducible from the frozen official inputs; the original lists, unchanged manuscript, disclosed discrepancies and new outputs are retained in the audit files. This distinction also prevents attributing every difference solely to outlier treatment.

Table A1. Primary Estimate Versus Historical-Label Reconciliation

Comparison	Matched n	Stable	Up	Down	Stable %
Revised primary; raw inputs	134	112	15	7	83.58
Historical labels; corrected ISO match	134	114	10	10	85.07
Historical labels; original exact-name match	132	112	10	10	84.85

Source: Frozen membership files and original country lists. “Historical labels” means the original partition held fixed; only the matching rule differs. The original one-decimal report was 84.8%; the table uses two decimals. These are different specifications, not conflicting values of the same estimator.

A2. Specification and Computational Stability

Values nearer one indicate greater separation; values near zero indicate overlap, and negative values indicate a closer alternative group. The diagnostics are reported rather than used to conceal disagreement: silhouette favors two clusters in all three editions. Three-tier findings are consequently accompanied by two- and four-

cluster transitions and by Ward hierarchical clustering on the same standardized components.

Computational stability is evaluated using 25 seeds with 25 starts per seed. Sample-composition sensitivity is assessed with 200 country bootstrap samples per edition, refitting both the standardization and clustering, and by omitting each country in turn. Bootstrap partitions are aligned to the reference by maximum agreement. Adjusted Rand indices (ARI) compare partitions without depending on arbitrary cluster labels; country-level retention fractions identify uncertain boundaries. These diagnostics assess classification stability, not the survey sampling error of the LPI.

A further 200 scenarios perturb each 2023 component independently and uniformly by up to ±0.05, clipping to the 1–5 scale before refitting. This is a numerical rounding scenario, not a model of respondent error. The robustness set also includes raw versus IQR-mean preprocessing, median/IQR scaling, rounded 2016 values, common-country refits and classification against fixed 2018 scaling and centroids. The latter holds the reference system constant while allowing observed country profiles to change. Full settings and replicate outputs are supplied.

The common-country refit for the recent pair reproduces the primary labels for all 134 matched economies. In the balanced 129-country refits, 109 remain in the same group in 2016–2018 (84.50%) and 107 in 2018–2023 (82.95%). The latter comparison contains 13 upward and nine downward movements. Fixing the 2018 scaler and centroids gives 109 persistent labels among the 134 recent-pair economies (81.34%). Table A2 separates these estimands so that the primary result is not mixed with a differently fitted sample.

The numerical robustness checks do not support an unqualified claim of method invariance. For 2018–2023, persistence is 89.55% with two K-means clusters, 73.13% with four, and 58.96% with Ward's three-cluster partition. Ward and primary K-means have ARI values of 0.803, 0.338 and 0.654 for the three editions. Ward's group boundaries therefore differ substantially, especially in 2018. These differences are substantive limitations of any country narrative that depends on one classification algorithm.

Changing seeds leaves the 2018 and 2023 primary partitions unchanged in the 25-seed check, but not all 2016 fits coincide: its minimum ARI is 0.916. Median bootstrap ARIs are 0.877, 0.891 and 0.931, with minima of 0.358, 0.339 and 0.384. The numbers of economies retaining their reference label in fewer than 80% of bootstrap fits are 19, 17 and 10. Omitting a single economy changes at most six, ten and two remaining labels, respectively. The 80% cutoff is a descriptive reporting convention, not a calibrated confidence threshold.

Median/IQR scaling yields recent-pair persistence of 79.10%. Rounding the 2016 inputs to two decimals changes four primary labels, illustrating boundary sensitivity even to small score differences. Substitution of the three discrepant 2018 report values leaves all 160 primary memberships unchanged. The 2023 rounding scenarios have a median ARI of 0.977 and a minimum of 0.914. Table A3 summarizes uncertainty diagnostics; the distributions and country-level retention

values remain available rather than being reduced to a claim that every result is robust.

Table A2. Recent-Pair Transition Sensitivity, 2018–2023

Specification	n	Stable	Up	Down	Stable %
Primary: full-sample K-means, k=3	134	112	15	7	83.58
Common-pair refit, k=3	134	112	15	7	83.58
Balanced-panel refit, k=3	129	107	13	9	82.95
Fixed 2018 scale and centers	134	109	18	7	81.34
Median/IQR scale, k=3	134	106	18	10	79.10
Old transform; new Python fit	134	112	12	10	83.58
K-means, k=2	134	120	12	2	89.55
K-means, k=4	134	98	13	23	73.13
Ward hierarchical, k=3	134	79	2	53	58.96

Source: Authors' calculations. The balanced panel uses 129 economies; other rows use the 134-economy recent pair. Different k values change category resolution. The "old transform" row uses a new Python fit and is not the fixed historical-label estimate in Table A1.

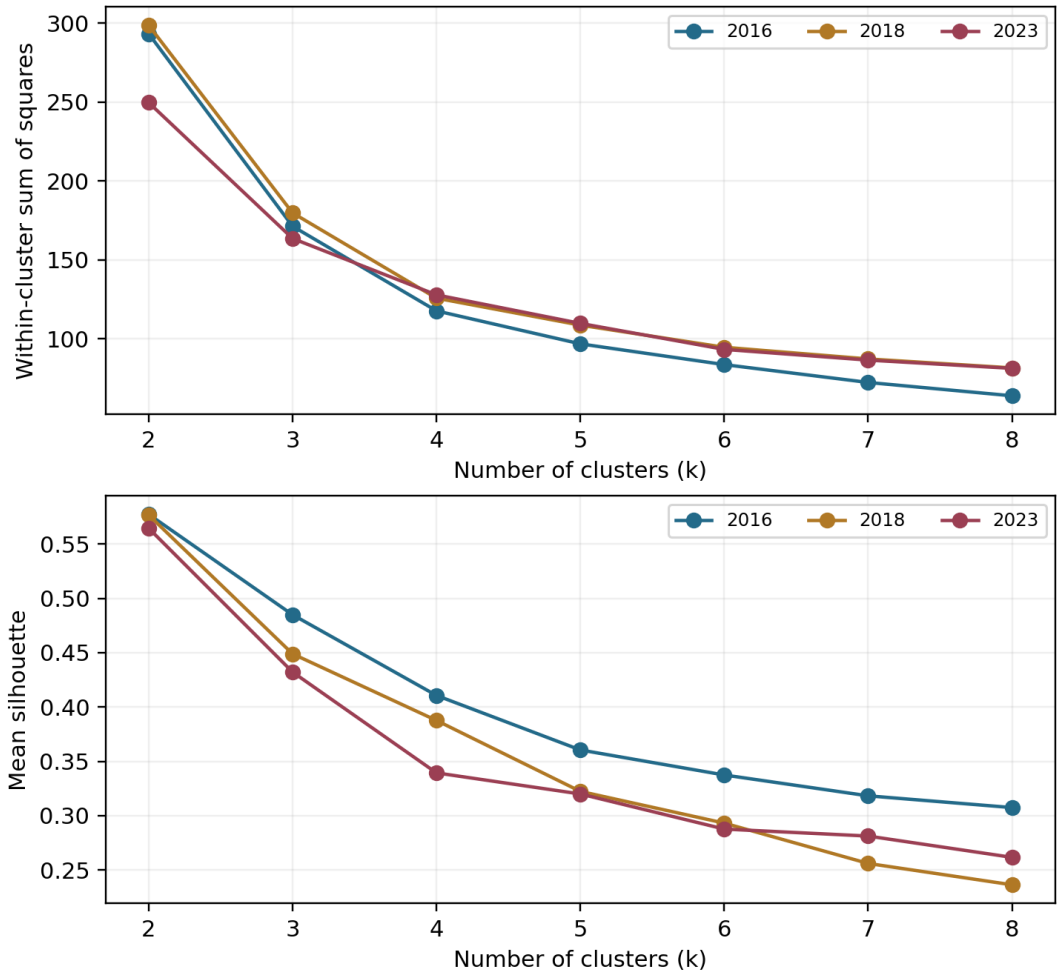


Figure A1. Cluster-Resolution Diagnostics for K=2–8

Source: Authors' calculations from the frozen World Bank data; identical feature definitions and estimation settings across the k grid.

Table A3: Stability and Country-Influence Diagnostics

Edition	Min seed ARI	Median boot. ARI	Min boot. ARI	Retention <80%	Max omission changes
2016	0.916	0.877	0.358	19	6
2018	1.000	0.891	0.339	17	10
2023	1.000	0.931	0.384	10	2

Source: Authors' calculations. Per edition: 25 seed fits and 200 country-bootstrap fits; every economy is omitted once. "Retention <80%" counts economies, and "Max omission changes" counts changed labels among the remaining economies. ARI: adjusted Rand index. Bootstrap retention is classification stability, not a survey confidence level.

A3. Data Verification and Provenance

All 459 country-edition score profiles were compared with the report tables. At displayed precision, three 2018 component values differ by 0.01 between the API and report: customs for Comoros and Iran is 2.62 versus 2.63, and tracking for Equatorial Guinea is 2.12 versus 2.13. Their cause is not established. The API values remain in the primary snapshot, and substitution of the report values is tested separately. The 2016 API retains greater precision than the printed report, whereas 2018 and 2023 scores have two and one decimal places, respectively. Rounding 2016 to two decimals is therefore an additional sensitivity check.

The data contains 459 complete economy-edition records (160+160+139), not 460. All were compared with the official report profiles in the earlier verification. As an additional manual check, six 2023 country profiles were compared again with the official World Bank report, Appendix 1, printed pages 32–33 (PDF pages 46–47). The dashboard route was unavailable; these checks use the official report, not a claim of a new live-API census. All seven displayed scores for these six countries agree with the frozen CSV.

Table A4: Additional Official-Report Spot Checks, 2023

Economy	Overall	CUS	INF	SHP	CMP	TRK	TIM
Singapore	4.3	4.2	4.6	4.0	4.4	4.4	4.3
Finland	4.2	4.0	4.2	4.1	4.2	4.2	4.3
Germany	4.1	3.9	4.3	3.7	4.2	4.2	4.1
United States	3.8	3.7	3.9	3.4	3.9	4.2	3.8
China	3.7	3.3	4.0	3.6	3.8	3.8	3.7
Türkiye	3.4	3.0	3.4	3.4	3.5	3.5	3.6

Source: Arvis et al. (2023), Appendix 1, pp. 32–33. CUS: customs; INF: infrastructure; SHP: international shipments; CMP: logistics competence and quality; TRK: tracking and tracing; TIM: timeliness. Columns here deliberately place tracking before timeliness, matching the analytical CSV.
Official report:
https://documents1.worldbank.org/curated/en/099042123145531599/pdf/P17146804a6a570ac0a4f80895e320dda1e.pdf. Source URLs and the API-to-edition mapping are also preserved in the reproducibility supplement. The spot checks complement, but do not replace, the full frozen-profile audit.

A4. Reproduction Files and Interpretation

The accompanying Reproducibility_Data_Code.zip contains frozen official API responses, the analytical CSV, explicit Python code, full memberships, component

centers, transitions, diagnostics and journal-figure code. The primary analyze.py uses only the official analytical CSV; historical membership material is confined to audit_legacy.py. The latter extracts the two original membership-table excerpts and writes the transcribed membership CSV; it does not supply the primary fit. No historical membership is injected into primary clustering.

The prior independent execution in a separate copy completed all six pipeline steps and reproduced 45 data/result/figure files byte-for-byte; the execution log is retained in Audit/Reproduction_Rerun.json. The present minimal-text revision does not change the frozen scientific inputs, model settings or computed results. The full numerical evidence is reused, not rerun merely to change prose.

The six components receive equal treatment after standardization. The official overall score is not a seventh feature. The original input file and executed R program remain unavailable; the new primary computation is reproducible, whereas exact historical-program replication is not asserted. Bootstrap, rounding and omission exercises evaluate sensitivity of classification, not confidence intervals for the underlying LPI survey.