School-to-Work Transition in the Youth Labor Market in Central and Eastern Europe: A Cluster Analysis Approach

Tomislav Korotaj
University of Zagreb, Faculty of Economics & Business, Croatia
James Ming Chen
Michigan State University, College of Law, United States of America
Nataša Kurnoga
University of Zagreb, Faculty of Economics & Business, Croatia

Abstract

Background: This study analyzes education, training, and the youth labor market in central and eastern Europe. **Objectives:** This study aims to evaluate similarities and differences in youth labor markets among eleven central and eastern European countries from 2008 to 2021. It specifically examines three aspects: wage ratios, early departure from education or training, and the share of the population not in employment, education, or training. **Methods/Approach:** This study applies hierarchical clustering and multidimensional scaling to panel data. The complete-link method organizes countries into clusters. This study combines three-dimensional Cartesian projections and two-dimensional projections based on multidimensional scaling with dendrograms and heatmaps, to graphically illustrate the "school-to-work" transition across this region. **Results:** Clustering highlights the Visegrád countries, the Baltics, and the Balkans as zones with internally homogeneous yet externally heterogeneous challenges for the youth generation. As the outliers in each of these regions, Poland, Estonia, and Bulgaria support clustering solutions that deviate from conventional understandings of central and eastern Europe. **Conclusions:** Historical and geographical ties continue to define this region's youth labor markets across political and economic dimensions. Clustering analysis identifies triumphs and struggles in policymaking in some of the poorest and most politically challenging member-states of the European Union.

Keywords: hierarchical cluster analysis; complete-link method; time series; youth population; wage ratio; NEET; early departures from education; central and eastern Europe

JEL classification: C38, E24, I21, I28, J4
Paper type: Research article
Received: Apr 11, 2024
Accepted: Aug 23, 2024

Citation: Korotaj, T, Chen, J., M., & Kurnoga, N. (2024). School-to-Work Transition in the Youth Labor Market in Central and Eastern Europe: A Cluster Analysis Approach. *Business Systems Research, 15*(2), 100-139.
DOI: doi.org/10.2478/bsrj-2024-0020

Introduction

Today's youth population grapples with considerable societal pressures that exact varying expectations from different groups. Graduating from school marks a significant early milestone. Yet the journey does not end there. Some young people advance to higher levels of the educational system. Others directly enter the labour market. Another cohort finds itself in a challenging position: early dropouts from compulsory schooling. Education, an indispensable tool for enhancing the quality of life, may prove elusive for young people who lack motivation, family support, or equal opportunities. Navigating these challenges demands a determined collective response and a societal commitment to providing proper guidance and support to vulnerable young people.

International organizations (such as the International Labour Organization (ILO), Organization for Economic Co-Operation and Development (OECD)) and statistical offices (such as Eurostat) regularly report on the obstacles encountered by young Europeans. In 2021, the Council of the European Union (EU) set two crucial thresholds for educational progress (Council of the EU, 2021): (a) reducing the rate of early departure from education or training (EDET) to less than 9 percent and (b) increasing the share of tertiary-educated individuals aged 25-34 to at least 45 percent. Both targets apply across the entire EU, with a deadline of 2030. These targets build on the earlier Europe 2020 strategy (which covered the period 2010-2020) in which the thresholds (European Commission, 2010) were set at 10 percent for EDET and 40 percent for 30-to-34-year-olds with tertiary education.

This paper extends previous analysis of youth labor markets and the school-to-work transition in central and eastern Europe (Korotaj, Chen, and Kurnoga, 2023). In addition to emphasizing developments over the past decade, this paper shifts its focus beyond tertiary education and the cohort of youth wholly excluded from training, education, and employment. This new study examines the ratio of minimum to average wages and early departures from education or training alongside labor market failures. These three variables – the Kaitz ratio of minimum to average wages, the EDET factor revealing early departures from education or training, and the NEET factor identifying young people not participating in education, employment, or training – present a comprehensive summary of inputs and outcomes in the youth labor market, without presuming or relying on any specific causal mechanism connecting the variables.

This paper's motivation lies in uncovering the relative successes and failures of eastern and central European countries in addressing their youth labor markets. Numerous studies of youth labor markets have elaborated the relationship between educational shortcomings and eventual unemployment. Although causal relationships among educational inputs and labor market outcomes are latent in the data, this paper's reliance on *unsupervised* learning produces results that differ from the testable hypotheses of multiple regression and other forms of *supervised* learning (Valkenborg et al., 2023; Watson, 2023). "Unsupervised learning involves the use of training datasets without dependent variables. In this scenario, the goal is to explore the data and draw conclusions about, for instance, hidden patterns or structures that are present in the data" (Burzykowski et al., 2023, p. 733). "Unlike supervised learning, unsupervised learning methods cannot be directly applied to a regression or a classification problem …." (El Bouchefry & de Souza, 2020, p. 227). "[C]lustering algorithms," generally speaking, "do not test against the null hypothesis" (Watson, 2023, p. 27).

By "using datasets without clear notice of the dependent (response) variable," unsupervised learning directs a "machine or computer [to] learn patterns from the data without referring to any specific response" (Valkenborg et al., 2023, p. 877). "Unsupervised learning aims to explore the data structure and generate a hypothesis rather than to test any hypothesis by statistical methods or to construct prediction or classification models on the basis of a set of conditions and a specified response" (Valkenborg et al., 2023, p. 877). Ultimately, "unsupervised learning is ontologically fundamental," or at least "more so, at any rate, than supervised or reinforcement learning algorithms, which necessarily presume some a priori division of the world into endogenous inputs … and exogenous outputs" (Watson, 2023, p. 29).

Consistent with the "common" use of "unsupervised learning … to find hidden patterns or groupings in the data" (El Bouchefry & de Souza, 2020, p. 228), the clustering analysis in this paper aims to identify patterns in policymaking and policy success or failure among the nations of central and eastern Europe. Indeed, as a review of the relevant literature will demonstrate, clustering analysis has figured prominently in multiple comprehensive examinations of youth labor markets. Although clustering can serve as a form of preprocessing preceding null hypothesis significance testing through regression (Sakellariou et al., 2012), clustering as the primary form of unsupervised machine learning can stand as its own exclusive research method in a scientific study. One such study relying exclusively on clustering without naming, confirming, or invalidating a formal research hypothesis, has been published by this journal (Krpan et al., 2023).

The research hypothesis of this article relies upon unsupervised machine learning to achieve a taxonomy or ontology of countries according to mathematical relationships latent in economic data. In the sense of the ancient Greek word, ἀποκάλυψις (apokálupsis), as a "revelation" or "disclosure," clustering and manifold learning can produce mathematically informed, principled interpretations drawn directly from data. Specifically, this article evaluates similarities and differences in youth labor markets from 2008 through 2021 among eleven countries in central and eastern Europe. The three variables indicating the health of those labor markets, as young people complete the school-to-work transition, emphasize the ratio of minimum to society-wide average wages, the rate of early departures from education or training, and the total rate at which young people fail entirely to achieve employment, education, or training. Through the application of hierarchical clustering and multidimensional scaling to country-by-country time series capturing those three labor market variables, this article seeks to compare the performance of those eleven countries.

Conventional understandings of central and eastern Europe, informed by history that spans the Roman Empire through the world wars and intraregional conflicts of the twentieth and early twenty-first centuries, classify this distinctive and volatile part of Europe according to three geopolitical subregions: the Visegrád Group, the Baltic states, and the Balkan peninsula. "[A] page of history is worth a volume of logic" (Supreme Court of the United States, 1921, p. 349): From Diocletian's decision to split the Roman Empire to the Molotov-Ribbentrop Pact that precipitated World War II and the more recent politics of the European Union, these three subregions have played distinctive roles in parts of Europe beyond the Carolingian Empire and the "Inner Six" signatories of the Treaty of Rome (Bokros, 2013; Ghica, 2008; Nič, 2016). The identification of distinct clusters among central and eastern European nations then reveals relationships among wage ratios, education, and labor market outcomes. Those relationships follow distinct patterns in different subregions across central and eastern Europe. This paper ultimately illuminates relative successes and struggles among some of the poorest and most politically challenging member-states of the European Union.

Based upon results from clustering and manifold learning, this article tests whether this informal taxonomy can withstand rigorous mathematical evaluation. The extent to which clusters reported by unsupervised learning reinforce or depart from conventional understandings delivers an ontology that can guide the formulation of labor policies throughout this part of Europe, particular for countries that over- or underperform their immediate neighbors and historical counterparts.

After reviewing the relevant literature, this paper will provide a concise description of its data and research methodology. This study conducts hierarchical clustering analysis of youth labor market time series with the complete-link algorithm. Results will be presented through three-dimensional projections and their corresponding two-dimensional manifolds, dendrograms, and heatmaps, as well as verbal descriptions of plausible cluster solutions. The discussion section reviews the reasons that support possible groupings of central and eastern European countries. Concluding thoughts highlight the limitations of our research and suggest areas for further investigation.

Literature Review

In 1996, Hungarian law raised the age at which students could stop attending school from 16 to 18 years (Adamecz, 2023). Hungary adopted this measure to combat early departures from primary and secondary school. Despite increasing the required years of schooling, the reform did not reduce dropouts or improve employment outcomes for 20- and 25-year-olds. In schooling systems that force failing students to repeat a grade, obligatory education should aim to keep students in school until they obtain a secondary school diploma. Different regression discontinuity design models estimated the effects of extended compulsory schooling in Hungary.

Focusing on convergence among EU countries, Cuestas, Monfort, and Ordóñez (2021) comprehensively evaluated the Europe 2020 strategy. Their study unveils the presence of convergence clubs (Bernard & Durlauf, 1995, 1996) within the framework, albeit with varying paces of progress. Employing the Philips-Sul approach to analyze crucial educational variables emphasized in Europe 2020, the research identified three clubs for EDET and six for tertiary education.

These findings highlight a lack of convergence among EU countries. Notably, central and eastern European countries are dispersed across multiple clubs. Croatia stands out as the best EDET performer without club convergence. These results suggest that EU countries will probably fall short of achieving educational convergence within Europe 2020's anticipated timeframe.

Other sources have criticized Europe 2020. One multivariate analysis of non-stationary time series compared the EDET rates in Czechia and EU-28 countries between 2005 and 2018 (Blatná, 2020). Czechia performs significantly better than the EU-28. A mismatch in the integration order of the EDET indicator`s time series prevented the discovery of a meaningful statistical relationship between Czechia and the EU-28 countries. Nevertheless, the results for Czechia showed that higher social benefits and increased opportunities for well-paying jobs can potentially overcome incentives to leave education early.

An evaluation of the Scottish School Leavers Surveys (SSLS) delivered one of the first assessments of the NEET concept (Furlong, 2006). NEET describes persons who have "no employment, education, or training." The author investigated the reasons that young people leave school early and ultimately lack employment, education, and training. Many young people could not find a suitable job or course, reflected indecision about career choices, or lacked additional qualifications or skills for employment. Heterodox analysis of the extremely diverse backgrounds of young people is essential to the identification of subgroups among youth and the prescription of specific policies for each subgroup. A "one size fits all" approach promises little success in reducing either dropouts or unemployment. Moreover, forcing employment or training in any job against personal inclinations might do more harm than good.

Factors influencing the NEET population in Italy and Spain between 2007 and 2017 emerged in a time-varying correlation model that analyzed changing patterns in NEET and EDET rates (De Luca et al., 2020). Spain exhibited the highest EDET rates among EU countries, while Italy faced the highest NEET. Gender-specific patterns in Italy revealed that the vast majority of women progressed from EDET to NEET status.

A similar but less severe gender relationship was observed in Spain. Regression analysis confirmed a statistically significant influence of the EDET indicator on the NEET indicator in Italy but not in Spain. NEET rates depend on business cycles, unemployment, and the amount of time spent in education. Monitoring early departures from school can mitigate the NEET problem through remedial measures that address EDET. Policymakers can, therefore, tackle two challenges with a single intervention.

A multilevel logit model has investigated educational and socioeconomic reasons for early school departures in the EU (Lavrijsen & Nicaise, 2015). Data for persons aged 20 to 30 in 2009 showed a significant influence on parental background and financial situation. The children of poorly educated families and families with material deprivation experienced higher EDET rates. Vocational education has proved a suitable mechanism for keeping young people in school longer (at least until they finish high school). The results recommend educational reforms that should be implemented alongside improved social policies. Everything starts with equal opportunities, and a more equal society is a prerequisite for positive changes in education.

Panel analysis of the development of young people in Germany ten years after they left an apprenticeship without completing it revealed differences in work experience and wages of that cohort relative to peers who completed their apprenticeships (Patzina & Wydra-Somaggio, 2020). The timing of a departure from an apprenticeship is significant. Wage growth is more pronounced among individuals who drop out later (more than two years of training) than those who drop out in the middle (between one and two years of training) or early (up to one year of training) stages of training. Therefore, preventing early dropout must be a priority. Dialogue between employers, training companies, and political institutions is necessary for the success of apprenticeships.

Different reasons motivated dropouts to return to education in Spain (Portela Pruaño et al., 2022). In a qualitative study based on data from an education center in Ceuta, most respondents stated that the primary motivation for returning to education was inactivity, acquiring qualifications, and increasing their chances of finding a suitable job. The majority said they would leave the program if offered a job. The support of family, friends, and teachers enabled young people to return to and stay in education. Communal influence and a supportive environment can encourage young people to improve themselves through training. In cooperation with the government, training centers should provide clear financial barriers to education, training, and self-improvement.

EDET and NEET rates converged in EU regions from 2003 to 2015 (Rambla & Scandurra, 2021). This study combined data at the NUTS (Nomenclature des Unités Territoriales Statistiques) 1 and NUTS 2 levels. Geographic differences related to EDET decreased.

The picture is more complicated for NEET. Regional GDP appears to accelerate convergence among wealthier regions, but it also reinforces slower development for economically disadvantaged regions. No significant convergence effects were observed in post-socialist EU countries. Between 2007 and 2016, the NEET gap decreased in post-socialist countries as regional GDP increased. Still, the mitigation of regional disparities was not as significant as the EU's average reduction in the NEET rate.

In their analysis of the impact of minimum wage increases on youth employment in various EU NUTS 2 regions, Vodă, Bercu, and Sebestova (2021) used panel data analysis for the period 2008-2014. They found a negative link between relative minimum wage as measured by the Kaitz index, which is the ratio of the minimum to the average wage (Brown et al., 1982; Kaitz, 1970), and youth employment. The employment of young individuals aged 15 through 24 decreased as a result of a higher minimum wage.

Minimum wages affect youth employment in the Visegrád Group countries (Fialová & Mysíková, 2020). According to this analysis of panel data from 2003 to 2016, youth employment declined in Hungary from 2008 to 2011 and in Czechia from 2003 to 2007. However, the changes in relative minimum wage had an overall non-negative impact across all the observed countries.

A closer examination of Albania and North Macedonia uncovered not only persistent youth unemployment but also a subpopulation of young people who are resistant or unresponsive to education (Mehmetaj & Zulfiu Alili, 2020). These social pathologies appear peculiarly characteristic of these countries in ways that differ from the rest of the Balkan sub-region, to say nothing of the remainder of central and eastern Europe.

Multiple works have applied clustering analysis to the school-to-work transition in central and eastern Europe. One application of hierarchical clustering identified distinct school-to-work transition subregimes within this post-socialist region (Dingeldey & Buttler, 2023). The variables in this study described the broader economic backdrop, legal protections for employment, education and vocational training, and labor market policies across 28 European countries and Israel.

Hierarchical clustering has also illuminated the relationship between tertiary and labor market outcomes across the European Union (Krpan et al., 2023). This study distinguished sharply between tertiary education as a labor market input and a range of indicators for labor market outputs. The authors found an imperfect match between countries that attained the highest shares of adults with tertiary education, on one hand, and countries that realized the best outcomes, such as the highest average employment and income benefits arising from tertiary education. Limitations on data availability constrained the ability to draw stronger conclusions about the impact of tertiary education on labor market outcomes.

One study has combined clustering analysis with conventional linear regression of educational, labor-related, and developmental factors across Europe. Tudor et al. (2023) made broader use of variables such as gross domestic product and societal investment in education. These authors found three distinct clusters of EU countries with low, average, and high levels of educational investment. Although these authors did find that a higher rate of educational dropouts diminished compensation, hours worked, and productivity, the impact of education on wages and productivity does differ according to country-wide levels of investment.

Like Dingeldey and Buttler (2023), Krpan et al. (2023), and Tudor et al. (2023), this study contributes to the academic understanding of the similarities and differences in the experiences of young people as they navigate schooling, training, and employment. This study shares Dingeldey and Buttler's (2023) focus on central and eastern Europe. Unlike Krpan, Gardijan Kedžo, and Žmuk (2023), who confined their clustering analysis to exactly two years (2012 and 2021), this study has accounted for the effects of changes in policy and labor conditions across 14 years (2008 through 2021 inclusive).

At the methodological level, this study aligns more closely with Dingeldey and Buttler (2023) and with Krpan, Gardijan Kedžo, and Žmuk (2023) in emphasizing clustering and unsupervised learning as standalone methods of economic analysis that allow data to reveal insights without the stipulation of a formal research hypothesis. Tudor et al. (2023) did deploy clustering as a step in preprocessing *en route* to a more explicit exercise in causal inference. Whereas Tudor et al. consciously conducted linear regression, studies relying exclusively on unsupervised learning (including this one) are more cogently understood as indirect exercises in classification. This article extends the use of unsupervised learning as an indirect classification technique, routinely applied to images and computer vision (Olaode et al., 2014; Schmarje et al., 2021), to economics and other social sciences. Clustering analyses strive to identify differences in policies and performance across jurisdictional boundaries.

In other words, sorting the countries of central and eastern Europe into a cogent geopolitical ontology according to Euclidean distances in standardized time series representing the Kaitz wage ratio, the EDET rate, and the NEET rate is this study's research objective. Ontological or taxonomic outcomes identify possible differences in socioeconomic conditions and/or public policy among these "newer" members of the European Union. Studies relying exclusively on unsupervised machine learning – whether in economics, cognate social sciences, or seemingly remote domains such as medicine or computer vision – demonstrate that the identification of differences in economic performance and/or political choices is a legitimate and worthy research objective, even in the absence of answers to questions more conventionally addressed through generalized linear regression methods or other forms of supervised learning.

Like their counterparts in natural science and visual representations of physical reality, economic and political patterns can be detected by clustering. In light of this broader region's shared experience with socialism and (in some instances) Soviet domination after the Second World War, the discovery of mathematically meaningful distinctions in performance and policy adds to the academic understanding of youth labor markets across the region. This is especially true if those distinctions reinforce or modify the conventional, qualitative classification of central and eastern Europe into three distinct subregions known as the Visegrád Group, the Baltic states, and the Balkan peninsula.

This study's taxonomy of the post-socialist member-states of the European Union also differs from the conclusions of Dingeldey and Buttler (2023) in key respects. Although those authors place "[a]ll post-communist states [within] one cluster with the exception of Hungary" (*ibid.*, p. 167), their "results do not conform to the popular grouping" that recognizes the "Baltic states and the Visegrád Group … as homogeneous clusters" (*ibid.*, p. 168).

By contrast, the analysis in this study finds that conventional geopolitical boundaries do define (or come close to defining) cogent clusters within central and eastern Europe. Dingeldey and Buttler ultimately treat Poland and Czechia, Latvia, and Bulgaria as representative countries that serve as proxies for the three large subregions of central and eastern Europe: the Visegrád Group, the Baltics, and the Balkans "as a southern group of … countries with less developed economies" (*ibid*., p. 169). This study aspires to inform policymaking in the European Union by facilitating fruitful comparisons across national boundaries within the EU (Krpan et al., 2023, p. 209).

Methodology

Analytical Blueprint

A comprehensive analytical framework to youth labor markets has been applied, aiming to uncover distinct patterns based on wage ratios, early departures from education or training, and rates of nonparticipation in employment, education, or training, or NEET. The complete-link algorithm and Euclidean distances provide a robust mechanism for grouping data points based on their maximum distances. Heatmaps and dendrograms will facilitate the interpretation of clusters within the data.

Data were drawn from Eurostat databases and covered an observation period from 2008 to 2021. Standardization of the data ensures comparability and the accurate representation of social desirability in all variables. Three-dimensional (3D) projections, alongside dendrograms and heatmaps, provide a dynamic visualization of changes in labor markets. These results provide a nuanced understanding of progress and setbacks in youth labor markets across central and eastern Europe.

Data

This study's sample comprises 11 countries: Bulgaria (BG), Croatia (HR), Czechia (CZ), Estonia (EE), Hungary (HU), Latvia (LV), Lithuania (LT), Poland (PL), Romania (RO), Slovakia (SK) and Slovenia (SI). This study investigates the youth labor market across 14 years, from 2008 to 2021 inclusive. Data preprocessing and cluster analysis, as conducted in Python, aligned all three variables in the study in the same direction by reversing the NEET and EDET rate, and expressing them as "$100 - NEET$" and "$100 - EDET$", respectively. This transformation expresses rising values in each variable as socially beneficial. Specifically, higher values for all three indicators are preferable.

NEET is reported as the percentage of young individuals engaged in some form of employment, education or training. Likewise, EDET is reported as the percentage of young people with at least a high school diploma. It represents the group that stays in compulsory education or training for the full duration. Key indicators for this research are briefly explained below.

The Kaitz index, which is the ratio of minimum to average wages, is a crucial indicator of income inequality and poverty. "The Kaitz index has the advantage of summarizing a great deal of information about the minimum wage law in a single variable" (Brown et al., 1982, p. 500). The Kaitz index is often used as a relative measure of the minimum wage within national labor markets (Arpaia et al., 2017; Askenazy, 2003; Lenhart, 2017; O'Higgins & Moscariello, 2017; Williams & Mills, 1998).

The NEET rate refers to the percentage of individuals aged 25-34 who are not engaged in employment, education or training. It can be further divided into NEET rate for youth who are unemployed and NEET rate for youth who are inactive. Eurostat provides this metric for various age ranges, including 15-17, 15-19, 15-24, 15-29, 15-34, 18-24, 20-24, 20-34, and 25-29. This study used an inverted version of the NEET rate so that higher values indicate a positive social impact.

The EDET rate indicates young individuals aged 18-24 who have left compulsory school or training early. It is a group of youth which have completed primary school (lower secondary school according to Eurostat), but dropped out from secondary school (upper secondary school according to Eurostat) or were not involved in any kind of further education or training. As with NEET, this study used the inverted version of the EDET rate so that higher values reflect positive social impact.

Additional Methodological Considerations

This overview of this study's dataset warrants a few additional methodological observations. This dataset contains 14 annual observations for 11 countries. Consequently, $n = 154$. Among the three variables in the design matrix – the Kaitz wage index and modified versions of the EDET and NEET rates – NEET may be regarded as an implied target variable. By contrast with the Kaitz ratio as an indicator of background market conditions and EDET as a gauge of education and other labor market inputs, NEET measures labor market outcomes. If all three variables are treated equally, the cardinality of the dataset is quite low: $p = 3$. Treating NEET as the implied target variable reduces p even further to a value of 2.

Studies with deeper data, such as Tudor et al. (2023) and Vasilescu, Stănilă, Popescu, Militaru, and Marin (2024), have successfully deployed clustering analysis as a preprocessing step in anticipation of conventional linear regression. This study's low values *of n and p* motivate a simpler approach, one that relies exclusively on clustering. Unsupervised learning, unaccompanied by linear regression or any other apparatus for supervised learning, can still provide useful insights where both n and p are low. A dataset so small that it might struggle to support properly supervised learning, especially according to best practices that would split the data into training and test set instances and thereby reduce N for training purposes even further (Müller & Guido, 2017, pp. 17-18), may still offer useful information through clustering.

Unlike many other exercises in clustering and other applications of machine learning, this study does not suffer from a surfeit of variables or the curse of dimensionality (Evangelista et al., 2006; Marimont & Shapiro, 1979). The opposite is true: Scarcity of information forecloses resort to familiar techniques for reducing dimensionality. In order to preserve the richness of the information in the dataset, this study bypasses principal component analysis (PCA) and proceeds directly to cluster analysis. Although PCA can help simplify data, it can also lead to information loss (Jolliffe, 2002; Geiger & Kubin, 2012). Instead, this study engages in a direct evaluation of hierarchical clustering results and highlights the impact of political and economic history on the composition of clusters.

Clustering and Visualization Methods

The complete-linkage method calculates the distance between two objects, a and b, in two different clusters, A and B, according to this formula (Bezdek, 2021):

$$\boldsymbol{\delta_{CL}(A,B) = \max_{i\epsilon A,\, j\epsilon B}\{d_{ij}\}} \qquad (1)$$

Complete linkage uses the maximum distance between any single data point in the first cluster (A) and any single data point in the second cluster (B) to determine the distance between two clusters. This method tends to produce more compact clusters with greater similarity within clusters, making it suitable for identifying tightly knit groups within a dataset.

Hierarchical clustering can also use average linkage (Nielsen, 2016, p. 225; Xu et al., 2021, pp. 6010-6011):

$$\boldsymbol{\delta_{AL}(A,B) = \frac{1}{|A||B|}\sum_{i\in A}\sum_{j\in B}\{d_{ij}\}} \qquad (2)$$

The average-linkage method is embedded by default in heatmaps produced by the Seaborn graphical package for Python. Average linkage, therefore, provides an alternative approach to the complete-linkage method of hierarchical clustering.

Euclidean distance measures the shortest path between two data points in Euclidean space (Hair, 2018), serving as a standard metric for assessing the similarity between data points. Within an *N*-dimensional space, the distance between two objects, a and b, is computed based on the square root of the sum of the squared differences between corresponding coordinates of the two data points in Euclidean space:

$$\boldsymbol{d_{ij}} = \|\boldsymbol{a} - \boldsymbol{b}\|_2 = \sqrt{(\boldsymbol{a_1} - \boldsymbol{b_1})^2 + (\boldsymbol{a_2} - \boldsymbol{b_2})^2 + \cdots + (\boldsymbol{a_n} - \boldsymbol{b_n})^2} \quad (\mathbf{3})$$

This research uses Euclidean distance for hierarchical clustering, primarily due to its ability to produce closely knit cluster solutions, its efficiency in handling clusters of varying sizes, and its computational speed. At the same time, this approach is limited by the irreversible nature of cluster agglomeration, reduced suitability for large datasets, and a pronounced vulnerability to outliers. The selection of Euclidean distance is recommended for lower-dimensional datasets, where the "curse of dimensionality" is diminished (Aggarwal, 2001; Domingos, 2012; Kriegel et al., 2009). This study's focus on a three-dimensional dataset encompassing three economic variables further substantiates the choice of the Euclidean metric.

In a three-dimensional Euclidean space, a Cartesian projection specifies a unique point for every triplet of numbers (*x*, *y*, *z*). This mapping allows for the precise positioning and visualization of data within a three-dimensional framework. 3D projections visually represent multidimensional data along three-dimensional Cartesian coordinates. They indicate the spatial distribution of clusters, revealing patterns and possible outliers. Moreover, 3D plotting makes complex relationships more visually accessible. Interpretation relies upon the identification of groupings and Euclidean distances between data points in three-dimensional space. This approach, therefore, enables a visual understanding of multidimensional data. The weaknesses of this approach include potential overlap and occlusion, which may make some data points or relationships less visible.

Two-dimensional representations of these same spaces accompany the 3D plots. Dimensionality reduction through multidimensional scaling (MDS) enables the visualization of the data in this study in two dimensions (Cox & Cox, 2008; Hout et al., 2013). A 2D projection based on multidimensional scaling replaces all underlying variables. It reduces them to two arbitrary dimensions, which in turn can be understood and visualized as the *x*- and *y*-axes of a conventional Cartesian projection. Although the application of MDS results in the loss of data and the application of a numerical scale not directly linked to the underlying data, MDS does display relationships in a more readily interpreted 2D format.

Dendrograms visually represent the arrangement of clusters formed by hierarchical clustering. They indicate the sequence of cluster mergers and the distance at which clusters merge. The tree-like diagram reveals cluster hierarchy and similarity. Dendrograms often inform decisions about possible and final cluster solutions. Using dendrograms in cluster analysis offers deep insights into relationships and hierarchies within clusters, improving the understanding of data grouping. However, the complexity of analyzing large datasets through dendrograms can be challenging as numerous branches can complicate the determination of a cluster solution.

Heatmaps display data through variations in coloring. They utilize color intensity to signify the magnitude of values and illustrate the level of observation similarity within a dataset. Interpretation involves assessing color gradients to identify patterns, correlations, or anomalies. Darker colors typically represent lower values or intensities, while brighter colors reflect higher values or intensities. Changes in heatmap colors reflect standardized Euclidean distances between observations. The advantages of heatmaps include a straightforward visual comparison of large data matrices. Nevertheless, challenges may arise in distinguishing subtle shades, potentially complicating the analysis of closely related values.

These graphical tools collectively enhance the understanding of clustering results. Each tool offers unique insights and interpretive considerations. The next section reports clustering results for eleven central and eastern European countries from 2008 to 2021.

Results

Descriptive analysis

The following tables, 1 through 3, present descriptive statistics for the three variables underlying this analysis – the Kaitz index (*wages*), (one hundred minus) the rate of early departures from education or training (*edet*), and (one hundred minus) the rate of young persons not in employment, education, or training (*neet*). These means, standard deviations (SD), and five-quantile statistics (minimum and maximum values and the 25th, 50th, and 75th percentiles) address all years covered from 2008 to 2021.

Table 1

Descriptive statistics for *wages*, the Kaitz index, by country

	Mean	SD	Min	25%	50%	75%	Max
Bulgaria	40.536	2.635	35.500	39.125	40.800	43.100	43.300
Czechia	35.636	3.265	31.600	32.925	34.700	37.925	41.400
Estonia	37.093	2.925	33.000	34.975	36.800	38.450	42.600
Croatia	40.671	3.233	37.000	37.925	40.200	42.250	46.600
Latvia	42.643	2.405	37.400	41.600	42.500	44.550	46.400
Lithuania	45.393	2.900	40.200	42.850	46.400	46.875	50.600
Hungary	41.593	2.284	38.000	39.225	42.450	43.300	44.500
Poland	45.000	3.161	39.100	42.525	45.400	46.250	50.500
Romania	40.429	6.490	31.300	35.025	39.500	47.300	48.400
Slovenia	51.171	3.323	43.400	51.400	51.950	52.725	55.200
Slovakia	37.850	3.193	33.600	35.775	36.500	39.400	44.400

Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

Table 2

Descriptive statistics for *edet*, or (one hundred minus) the EDET rate; by country

	Mean	SD	Min	25%	50%	75%	Max
Bulgaria	86.907	0.903	85.200	86.300	87.250	87.475	88.200
Czechia	94.029	0.781	92.400	93.450	94.100	94.575	95.100
Estonia	88.607	1.599	86.000	88.000	88.700	89.625	91.500
Croatia	96.300	1.132	94.800	95.125	96.800	97.200	97.800
Latvia	89.771	2.531	84.500	88.650	90.150	91.475	92.800
Lithuania	93.900	1.358	91.300	92.825	94.250	94.675	96.000
Hungary	88.186	0.475	87.500	87.925	88.200	88.475	89.200
Poland	94.657	0.298	94.100	94.450	94.650	94.800	95.200
Romania	82.757	1.385	80.700	81.900	82.450	83.975	84.700
Slovenia	95.536	0.580	94.700	95.025	95.600	95.800	96.900
Slovakia	93.214	1.460	90.700	92.250	93.200	94.525	95.300

Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

Table 3
Descriptive statistics for *neet*, or (one hundred minus) the NEET rate; by country

	Mean	SD	Min	25%	50%	75%	Max
Bulgaria	75.457	3.441	70.300	72.300	75.850	78.100	80.400
Czechia	80.393	1.406	78.700	79.525	80.000	81.375	83.200
Estonia	82.007	3.466	75.600	80.050	82.550	84.650	87.000
Croatia	78.586	3.373	73.700	76.050	78.750	81.225	83.600
Latvia	80.750	3.596	74.000	79.075	82.100	83.325	85.200
Lithuania	82.950	3.543	75.600	80.625	83.400	86.025	87.400
Hungary	78.071	4.175	72.500	74.150	78.100	81.200	86.700
Poland	81.014	1.739	78.400	79.600	80.800	82.425	84.300
Romania	78.486	2.245	75.600	76.725	78.050	79.400	83.600
Slovenia	87.643	2.402	83.300	85.975	88.400	89.325	91.700
Slovakia	75.779	2.905	71.800	73.200	76.150	77.600	82.100

Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

It may reveal even more when examining all three variables by year rather than by country. Tables 4, 5, and 6 examine, respectively, the Kaitz wage ratio (*wages*), (one hundred minus) the EDET rate of early departures from education or training (*edet*), and (one hundred minus) the NEET rate of young people not in employment, education, or training (*neet*), all arranged by the years 2008 through 2021 rather than by country. Table 4 reveals that the *wage variable generally increased* from 2008 to 2021. This suggests a region-wide trend toward greater income equality.

Table 4
Descriptive statistics for *wages*, the Kaitz wage ratio, by year

Year	Mean	SD	Min	25%	50%	75%	Max
2008	37.436	3.509	31.300	34.900	37.600	39.650	43.400
2009	38.791	3.345	34.300	36.050	38.300	41.750	44.200
2010	38.864	5.265	32.400	35.750	38.000	41.950	50.500
2011	39.073	5.701	32.400	35.450	37.500	42.150	51.700
2012	39.345	6.044	31.600	34.850	37.900	42.950	52.200
2013	40.745	6.284	32.600	36.150	39.200	43.950	53.200
2014	41.255	5.805	32.800	37.200	40.500	44.700	52.800
2015	42.373	5.243	34.400	38.850	41.600	45.950	52.400
2016	43.400	5.060	35.500	40.100	43.400	45.800	51.700
2017	43.973	4.224	37.100	41.900	43.200	47.050	51.300
2018	43.991	3.959	38.200	41.100	43.700	46.250	51.700
2019	43.736	4.116	38.500	41.150	43.200	46.300	52.500
2020	45.200	4.287	39.500	42.200	43.600	47.400	53.600
2021	44.745	5.013	37.900	41.400	44.400	47.450	55.200

Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

The *edet* variable changes very modestly over this period (Table 5). Year-by-year differences in the average EDET rate across central and eastern Europe were no greater than 1.636 percent. Indeed, the minimum annual value in region-wide average *edet* was observed in 2008, while the maximum value was observed in 2021. Because all variables have been oriented so that higher values are socially preferable, we see that early departures became less troublesome, even if only marginally, from 2008 to 2021.

Table 5
Descriptive statistics for *edet*, or (one hundred minus) the EDET rate; by year

Year	Mean	SD	Min	25%	50%	75%	Max
2008	90.409	4.766	84.100	85.600	92.500	94.650	95.600
2009	90.418	4.613	83.400	86.100	91.300	94.700	95.100
2010	90.936	4.705	80.700	88.200	92.100	94.900	95.300
2011	91.300	4.358	81.900	88.500	92.600	94.950	95.800
2012	91.318	4.244	82.200	88.800	93.500	94.600	95.600
2013	91.518	4.154	82.700	89.150	93.600	94.500	96.100
2014	91.491	4.584	81.900	88.300	93.300	94.550	97.200
2015	90.964	4.936	80.900	87.500	93.100	94.600	97.200
2016	91.109	4.766	81.500	88.100	92.600	94.950	97.200
2017	91.136	4.558	81.900	87.850	91.400	94.800	96.900
2018	91.491	4.337	83.600	87.750	91.700	95.300	96.700
2019	91.573	4.165	84.700	88.500	91.700	95.100	97.000
2020	91.936	4.002	84.400	89.700	92.400	94.500	97.800
2021	**92.045**	**4.004**	**84.700**	**89.100**	**92.700**	**94.400**	97.600

Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

Finally, the central and eastern European region experienced higher levels of young people with bad labor market outcomes after the global financial crisis from 2009 onward. This trend is reflected in Table 6's year-by-year summary of the *neet* variable. Full regional recovery in the NEET rate did not occur until 2019. Variability among countries tended to decline throughout the recovery – until the COVID-19 pandemic struck in 2020.

Table 6
Descriptive statistics for *neet*, or (one hundred minus) the NEET rate; by year

Year	Mean	SD	Min	25%	50%	75%	Max
2008	82.100	4.040	75.800	80.550	82.000	83.600	91.700
2009	79.027	4.558	73.800	75.900	78.400	81.200	89.400
2010	77.291	4.507	72.500	74.200	75.600	79.200	88.400
2011	77.545	4.576	71.200	74.750	77.700	79.200	88.400
2012	77.418	4.241	71.500	73.800	78.800	79.350	86.200
2013	77.709	4.467	70.300	74.450	78.400	81.150	83.900
2014	78.355	3.982	71.800	76.450	79.600	81.050	83.300
2015	79.736	3.866	74.400	76.450	79.600	83.150	85.400
2016	80.236	3.899	74.000	76.850	80.800	82.500	85.900
2017	81.827	3.735	76.600	78.200	82.200	84.300	87.500
2018	82.636	3.526	77.700	80.250	82.700	84.750	88.400
2019	83.018	3.267	77.900	81.000	82.500	85.100	89.500
2020	81.491	3.481	77.300	79.400	80.400	83.000	89.100
2021	83.055	4.229	75.600	80.500	82.400	86.450	89.900

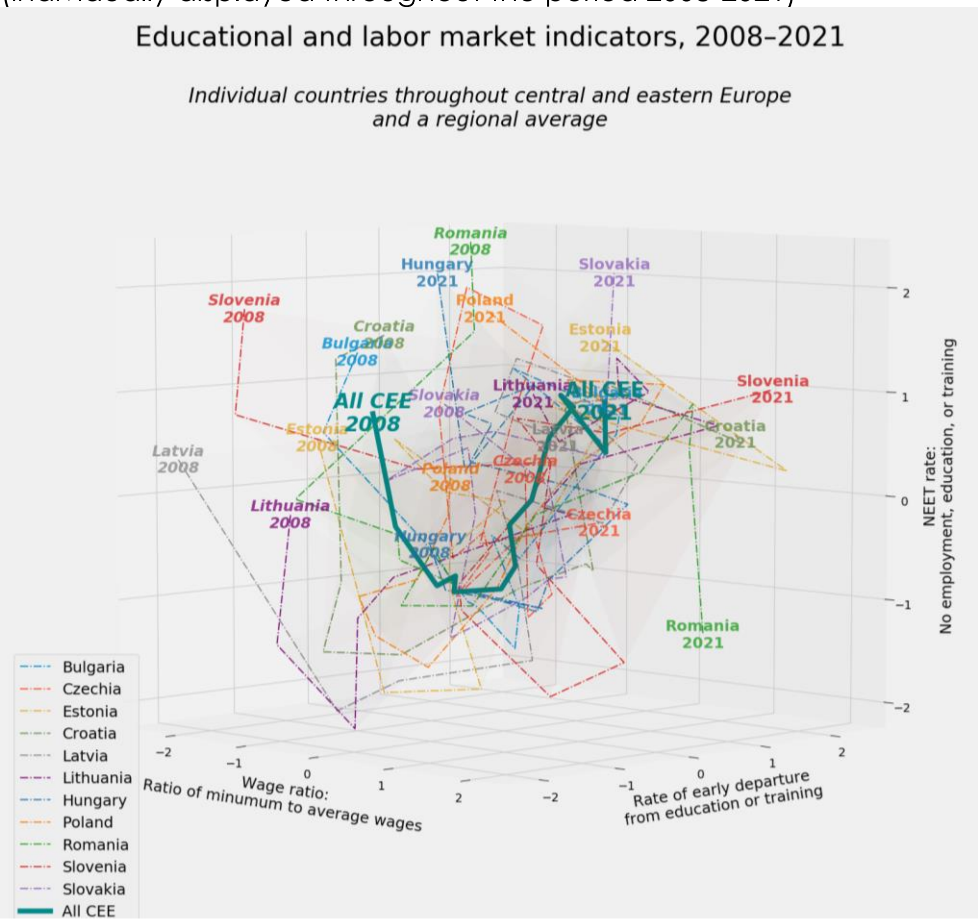
Note: SD = Standard Deviation; Min = Minimum, 25% = First Quartile; 50% = Median; 75% = Third Quartile; Max = Maximum; *Source:* Authors' work

The contrast between geographic and temporal overviews of the data also warrants commentary. Tables 1 through 3 organize the variables by country. Orderly rearrangement of the data is neither intuitive nor immediately evident. By contrast, temporal ordering by years in Tables 4 through 6 flows naturally. Even without elaborate computational intervention, some insights and patterns over time are visible to the naked eye. The clustering analysis that follows this exploratory description of the data reflects both geographic and temporal elements in the data. Above all, it must be remembered that all three variables are structured as *time series*. These series reflect autocorrelation within individual countries and geopolitical effects across borders. Therefore, clustering analysis should evaluate a continuous cross-section of data by multiple years, if at all possible, rather than individual years in isolation.

Three-dimensional projections

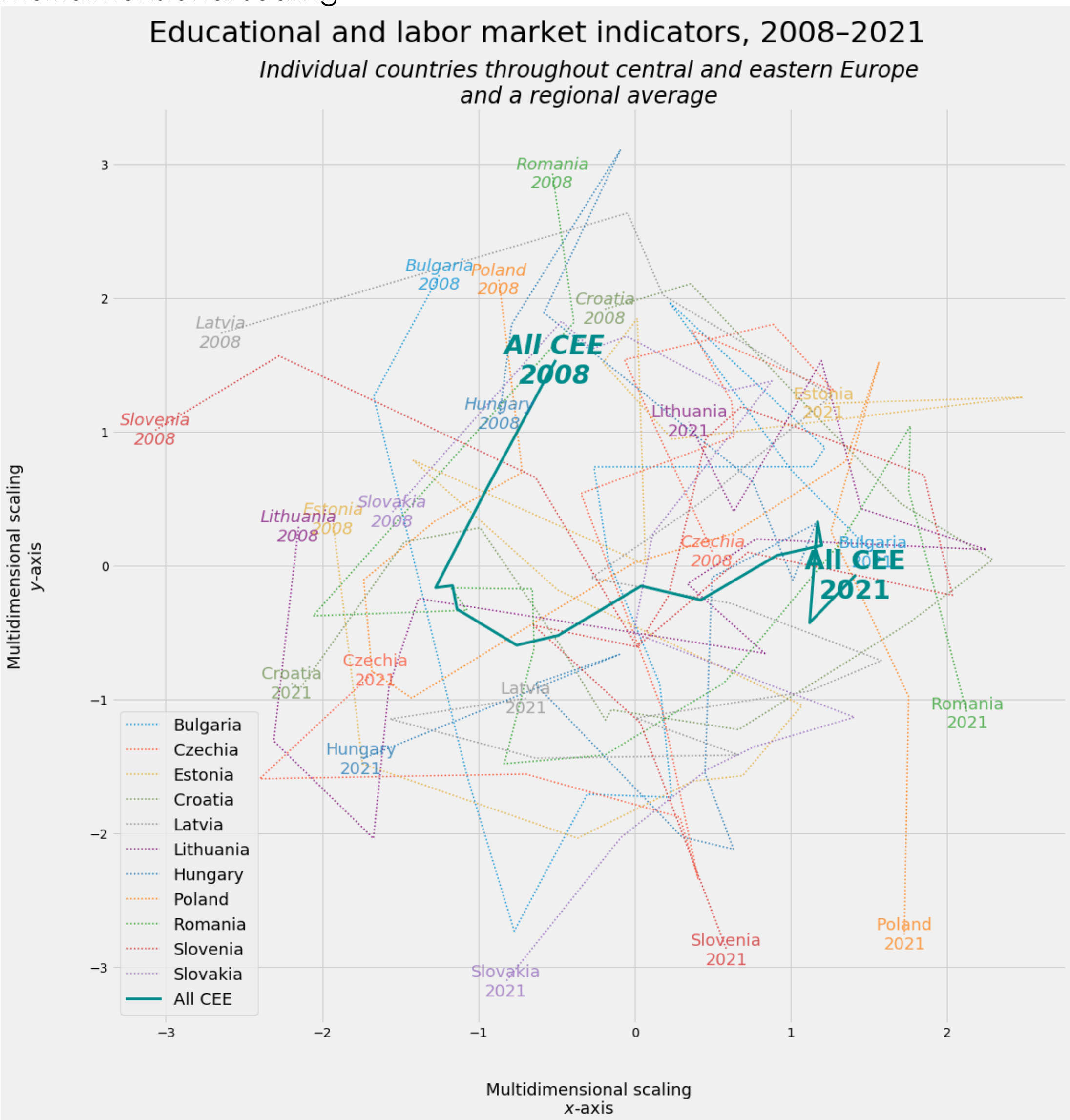
Analysis of the youth labor market begins with 3D projections. In the initial step, the standardization of raw values created a 3D projection of each country's performance in all three educational and labor market predictors (Figure 1). The average of all Cartesian coordinates representing these countries' Kaitz wage ratios, EDET rates, and NEET rates produces a global centroid, which is designated as the "All CEE" time series in Figure 1. The corresponding 2D manifold, enabled by multidimensional scaling, appears immediately afterward as Figure 1A.

Figure 1
3D projection of combined indicators for central and eastern European countries (individually displayed throughout the period 2008-2021)



Source: Authors' illustration

Figure 1A
2D projection of combined indicators for central and eastern European countries (individually displayed throughout the period 2008-2021), obtained with multidimensional scaling



Source: Authors' illustration

In Figure 1, standardized values are presented along three axes: (i) *wages:* The Kaitz wage ratio on the x-axis, (ii) *edet*: (One hundred minus) the EDET rate on the *y*-axis, and (iii) *neet*: (One hundred minus) the NEET rate on the z-axis.

These three variables do not appear in Figure 1A's MDS-enabled 2D manifold that encodes and visualizes the same information. Instead, the two-dimensional MDS manifold reports all three axes in Figure 1A – standardized values for *wages*, *edet*, and *neet* – as two arbitrarily designated x- and *y*-axes. As mathematical "translations" of the fuller 3D visualizations, this article's 2D manifolds are not intended to add analytical insights beyond the information contained in all three dimensions in the data. Rather, the 2D manifolds present clustering results in a way that is less visually demanding, and therefore perhaps more intuitively understandable.

Although this study's exclusive focus on clustering and unsupervised learning does not, strictly speaking, support distinctions between independent and dependent variables, let alone a search for causal inferences among variables, the NEET rate occupies the position on the z-axis that would ordinarily be reserved for a true target variable. To the extent that causal mechanisms could be hypothesized, sources such as Tudor et al. (2023) treat ultimate labor market outcomes such as employment, wages, and productivity as functions of macroeconomic and educational inputs. The Kaitz wage ratio measures the market for unskilled labor, while EDET is a surrogate for the overall educational attainment of the workforce. NEET, therefore, presents the strongest case for treatment as a target variable.

Values along each axis range roughly from –3 to +3, as should be expected of standardized data.

Countries moving toward higher values on all axes across this time period, such as Slovenia, indicate a favorable labor market with higher wage ratios and improved NEET and EDET conditions. Conversely, countries such as Bulgaria move toward lower values. These countries face suggesting challenges such as declining wage ratios or worsening NEET and EDET rates. With trajectories crossing the midsection of this 3D projection, Poland and Slovakia appear to have achieved moderate performance amid fluctuating labor markets.

In all three-dimensional plots, the upper right corner, representing higher values on all axes, is more desirable. Positive values on all axes indicate a more favorable youth labor market, with higher wage ratios and lower rates of NEET and EDET. The best performers exhibit upward trends across all axes, while the worst decline or stagnate. Moderate performers might show improvement in some areas, but not uniformly. Trends are influenced by economic policies, educational reforms, and broader socioeconomic factors. For instance, an improving trend could result from effective youth employment policies or educational programs reducing early departures from school or training. Stagnation might reflect persistent economic challenges or inadequate policy responses. More details follow the presentation of clustering results.

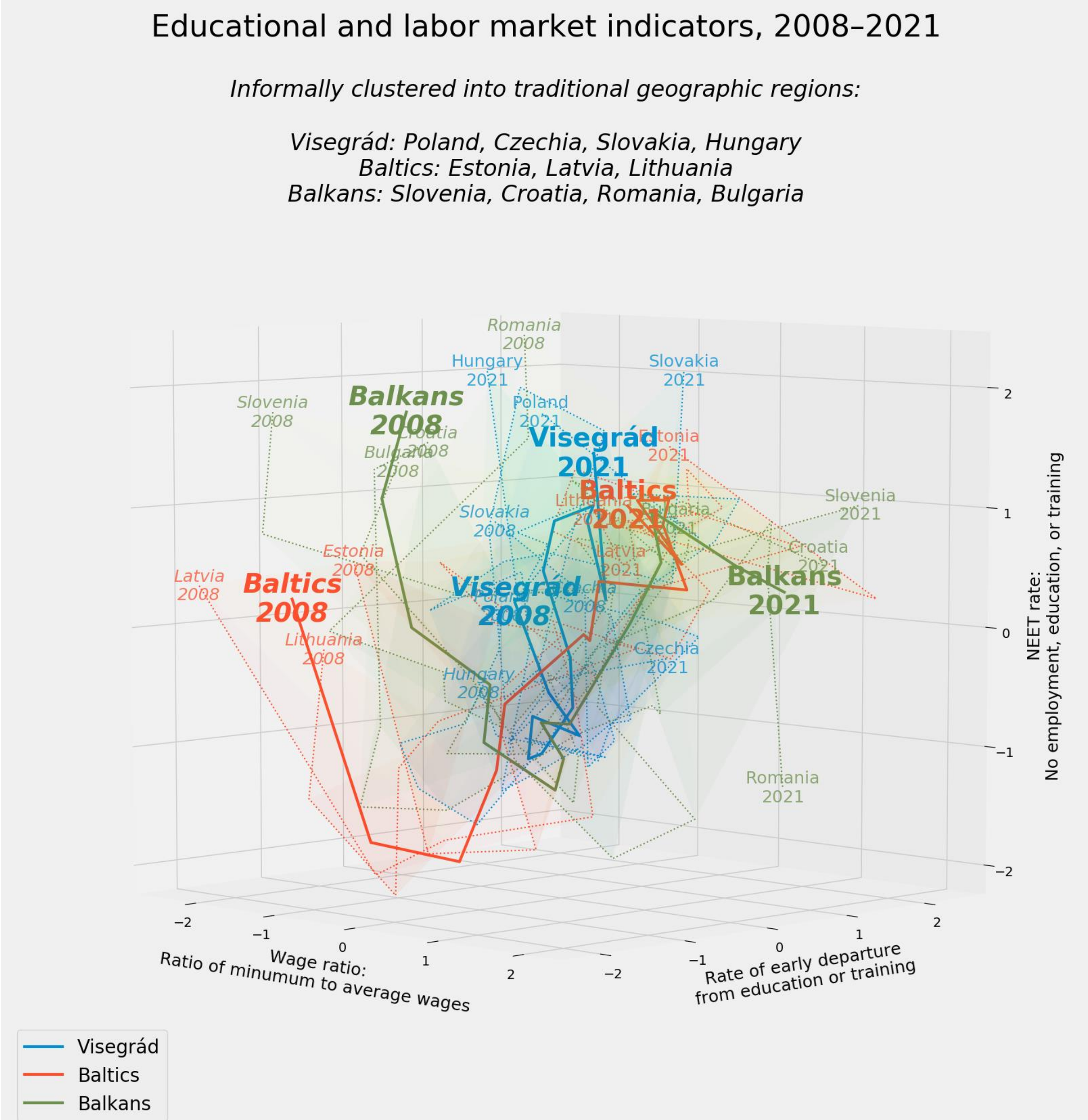
As in connection with Figure 2, an MDS manifold presents the same information in two dimensions (Figure 2A).

The second 3D projection (Figure 2) reports informal groupings of countries based on labor market variables – and, critically, on expert human judgment rather than strictly machine-derived definitions of clusters within central and eastern Europe. As in Figure 1, mean values for countries within a grouping produce a centroid in three-dimensional Cartesian space.

The Visegrád countries (Czechia, Hungary, Poland, Slovakia) are closely clustered, potentially indicating similar economic policies or labor market conditions influenced by their shared history and regional cooperation. The Baltics (Estonia, Latvia, Lithuania) also appear closely grouped, which could reflect their geographic proximity and comparable post-Soviet economic transitions. The Balkans (Bulgaria, Croatia, Romania, Slovenia) display a more dispersed pattern, perhaps due to varied economic conditions and differences in their levels of EU integration. As implied by the English word *balkanization*, this large and mountainous peninsula is one of Europe's most diverse regions.

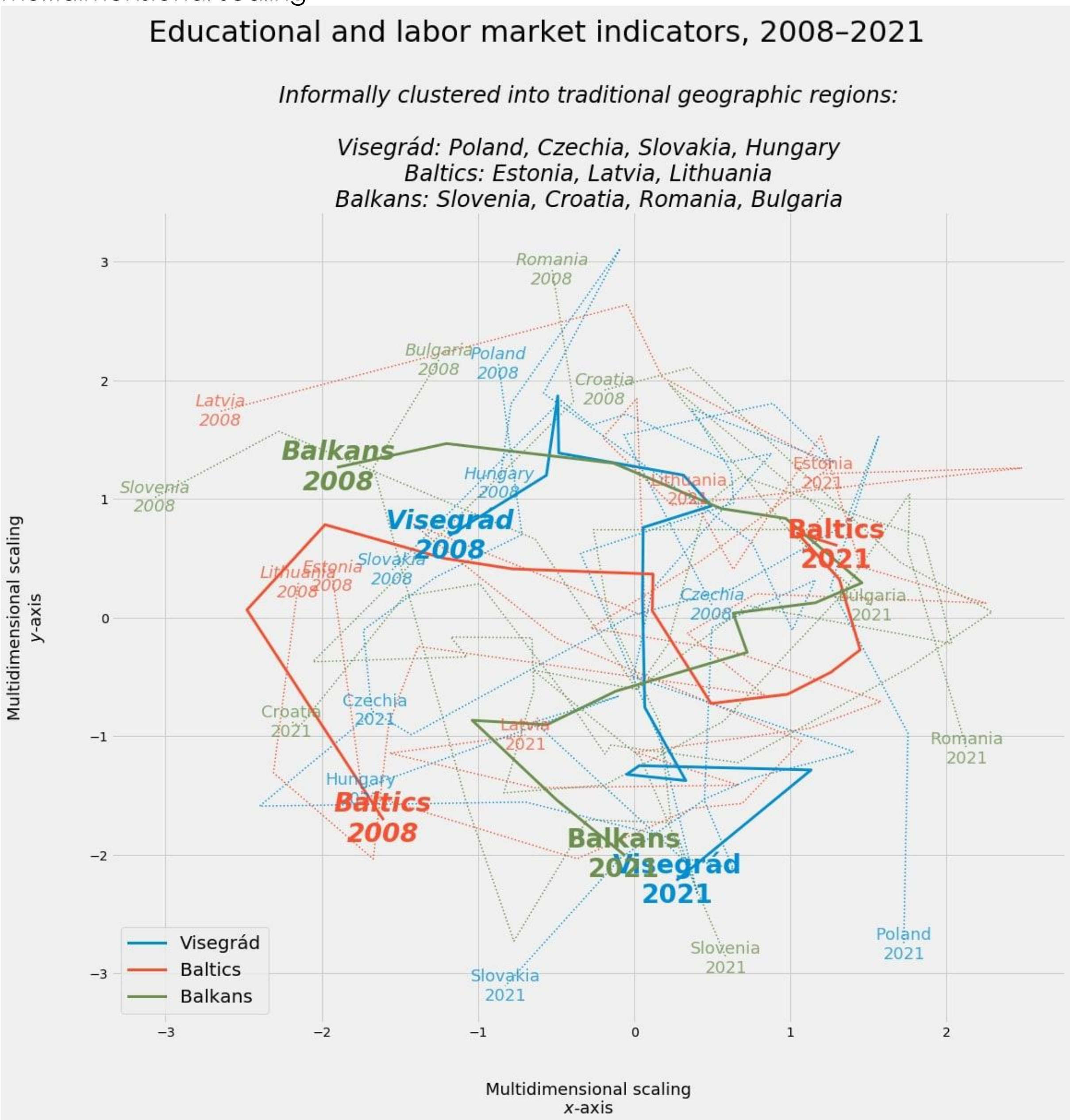
These informal clusters may reveal underlying historical, geographical, and political affinities, as well as shared economic experiences or reforms that influence labor market outcomes (Kunić, 2022). It will be instructive to compare qualitative or even intuitive human judgment against results drawn strictly from unsupervised machine learning.

Figure 2
3D projection of combined standardized indicators for central and eastern European countries (informally clustered throughout the period 2008-2021)



Source: Authors' illustration

Figure 2A
2D projection of combined standardized indicators for central and eastern European countries (informally clustered throughout the period 2008-2021) obtained with multidimensional scaling

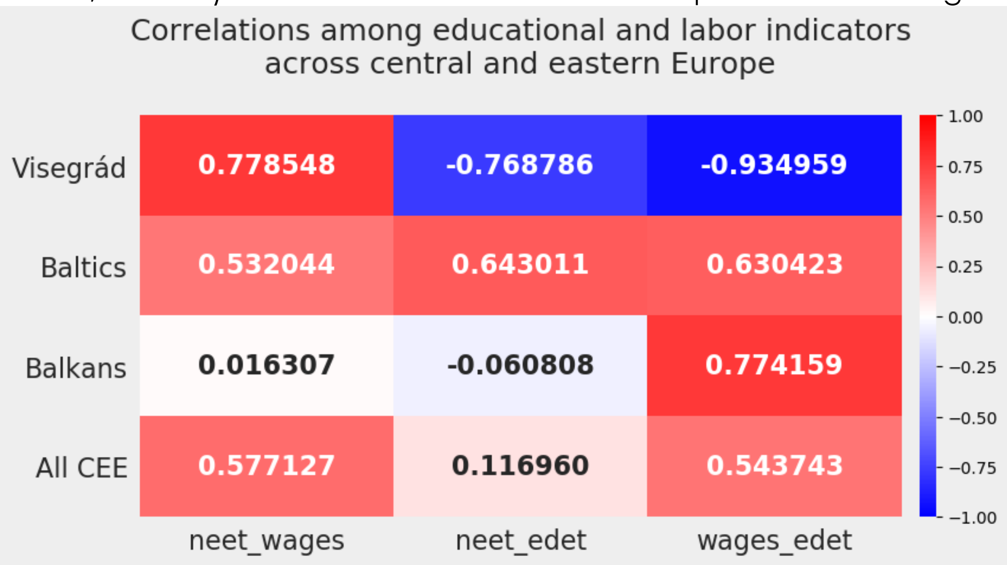


Source: Authors' illustration

A table of correlations for each informal geographic cluster shows how the three variables (the Kaitz wage ratio [or *wages*], EDET, and NEET) have starkly different relationships to one another by region (Figure 2B). The correlations in this table, and in other correlation tables throughout this article, cover all years from 2008 to 2021 inclusive. In principle, a favorable ratio of minimum to average wages, minimizing early departures from education and training, and avoidance of catastrophic NEET outcomes among young people should be mutually reinforcing. But a virtuous cycle of high wages for unskilled labor, persistence in education, and fuller employment may prove elusive. The effect of wages or education on ultimate labor outcomes may experience a lag. Those effects almost certainly differ by country and region.

Central and eastern Europe as a whole does exhibit a positive, mutually reinforcing relationship between the wage ratio, a positive interpretation of EDET, and a positive interpretation of NEET. All three correlations (NEET to wages, NEET to EDET, and wages to EDET) are mildly to modestly positive across these countries collectively. But only the Baltic region reflects the same, uniformly positive relationship. In the Balkans, neither wages nor EDET has any bearing on NEETs. At the same time, the wage ratio and this study's measure of EDET correlate very positively in the Balkans.

Figure 2B
Correlations among variables for each informal, conventional cluster (Visegrád, Baltics, Balkans) and for central and eastern Europe as an entire region



Source: Authors' illustration

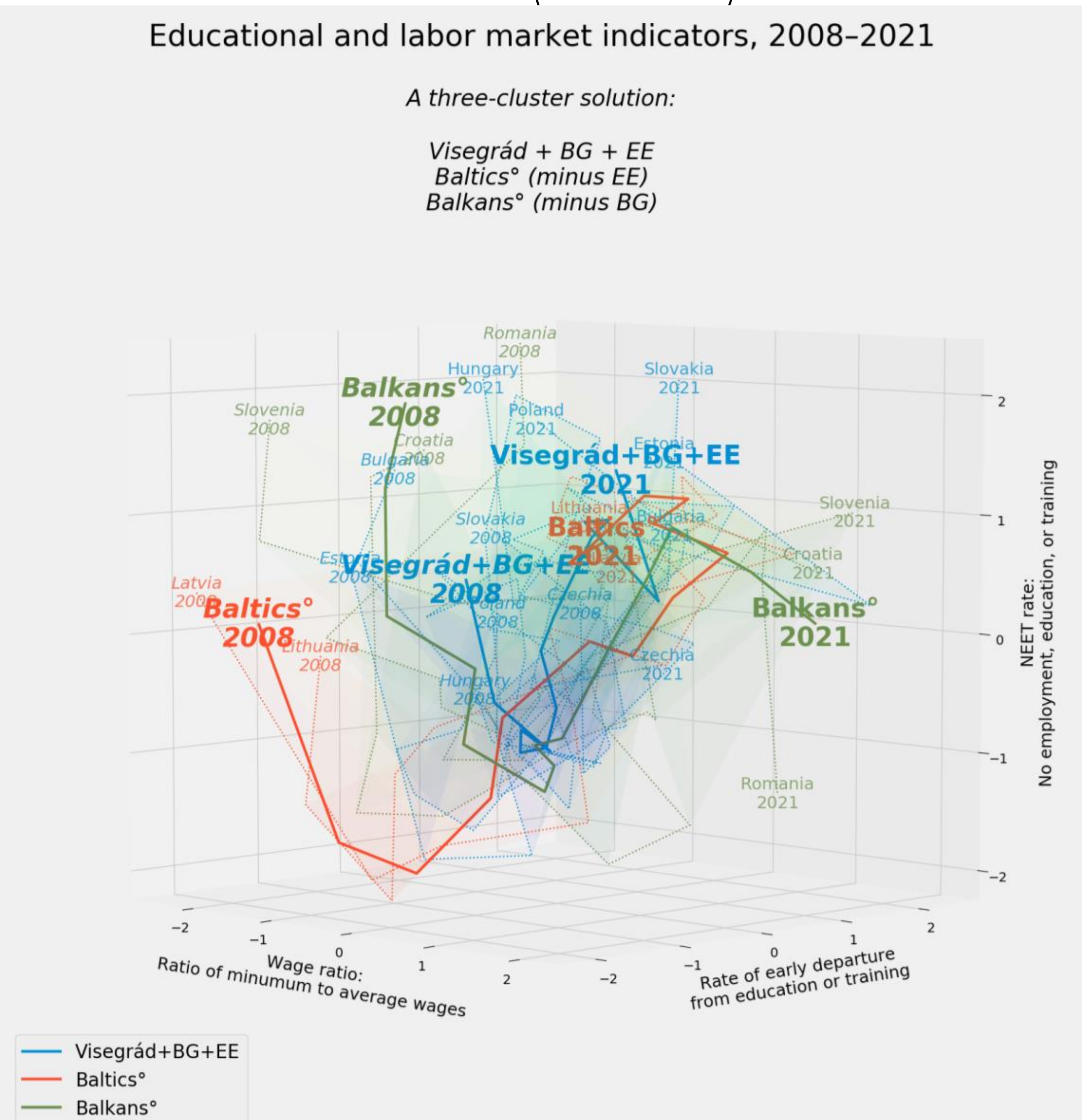
By contrast, the Visegrád countries exhibit a strongly negative relationship between EDET, on one hand, and either wages or NEET, on the other hand. Though at odds with the central and eastern European experience at large, a negative relationship could arise from the low opportunity cost of training or education during recessions and other disruptions of the lower-wage labor market. Cheap, abundant education or training may also lower the number of young people without employment, education, or training. The relative economic and political strength of the Visegrád countries, especially by contrast with poorer Balkan nations such as Bulgaria and Romania, may account for the different correlations.

Such stark differences may arise from diverse economic conditions and differences in policy. From 2008 to 2021, the Visegrád countries strengthened their cooperation in areas including labor policies, benefiting from EU funding and initiatives to improve youth employment (Bieszk-Stolorz & Dmytrów, 2020; Krzaklewska, 2013). The Baltic States, following EU accession, implemented reforms to address youth labor market integration, often focusing on vocational education and training enhancements (European Commission, 2017). The Balkan countries, with varied levels of EU integration, faced diverse challenges. Slovenia and Croatia, as EU members, took advantage of EU youth employment initiatives (Youth Wiki 2020, 2023a). Bulgaria and Romania worked to align with EU standards, addressing high NEET rates through national reforms and EU-supported programs (Institute for Market Economics, 2019; Neagu, Lendzhova, & Keranova, 2021; Toderiță, Damian, & Meiroşu, 2019).

Countries in all of these regions often increased minimum wages to improve living standards while combating youth unemployment and early departures from education through tailored policies, including labor market incentives and educational improvements. This 3D plot suggests that traditional country groupings may reflect more precisely mathematically prescribed agglomeration patterns revealed in subsequent 3D visualizations.

This study now shifts its focus to two formally selected cluster solutions. Cluster solution "A" is a three-cluster solution, and cluster solution "B" is a five-cluster solution. The former is displayed in Figure 3, and the latter, in Figure 4. Corresponding tables of correlations are also shown (Figure 3B and 4B).

Figure 3
3D visualization of Cluster solution "A" (three clusters)



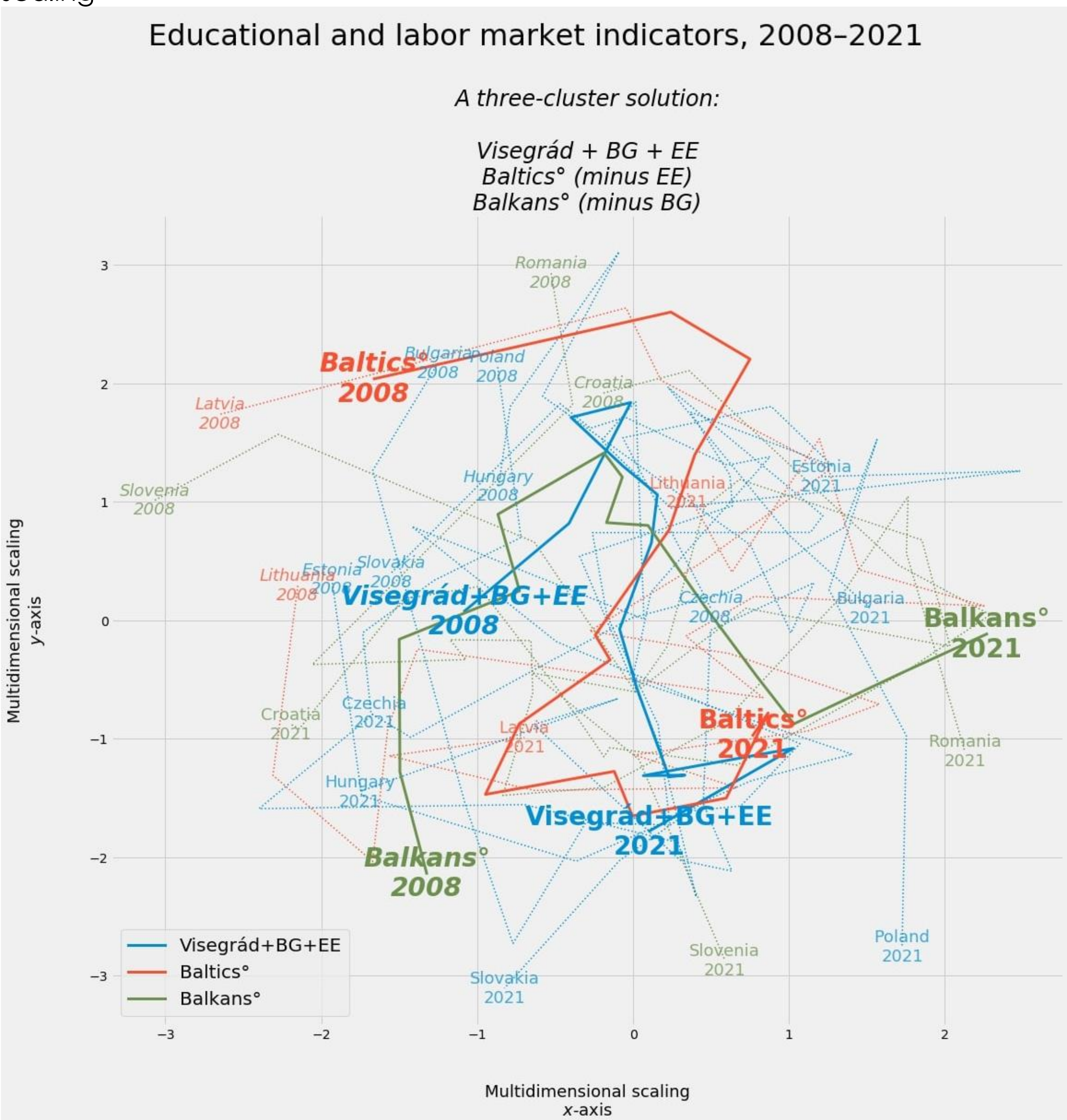
Source: Authors' illustration

The three-cluster solution also appears in two-dimensional form in Figure 3A. Over time, the Visegrád countries, joined by Bulgaria and Estonia, show a cluster that generally trends toward higher values on the wage ratio axis, suggesting relatively stronger wage growth compared to the other two clusters. Relative to the rest of central and eastern countries, minimum wages within this Visegrád-centered cluster have kept better pace with overall wage growth.

The Baltic cluster, excluding Estonia, displays a diverse trajectory, with some movement toward lower NEET and EDET rates. Latvia and Lithuania, the members of this shrunken Baltic cluster, could be considered moderate performers. Their most salient trait is modest improvement in educational performance, one not yet matched by wage growth for their lowest-earning workers.

The Balkan cluster, excluding Bulgaria, tends to have lower values across all variables. Uniformly negative movement across all three dimensions indicates persistent challenges in the labor market. The shrunken Balkan cluster of Croatia, Romania, and Slovenia appears to present the harshest educational and labor market conditions for young people in central and eastern Europe.

Figure 3A
2D visualization of Cluster solution "A" (three clusters) obtained with multidimensional scaling

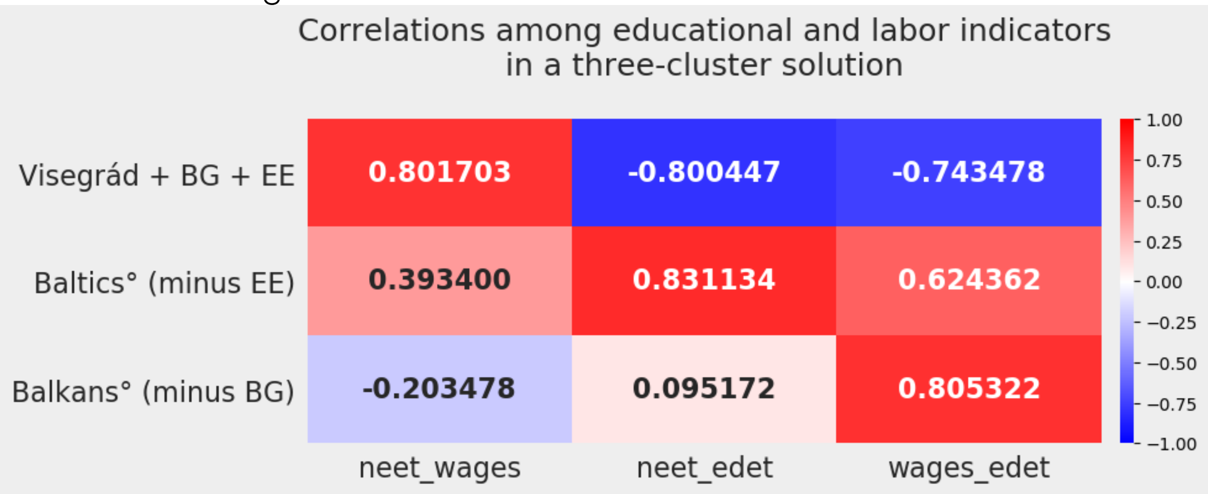


Source: Authors' illustration

Correlations among the three educational and labor market variables in the three-cluster solution "A" closely resemble their correlations in the informal division of central and eastern Europe into its traditional geographic components of the Visegrád Group, the Baltics, and the Balkans. The removal of Bulgaria from the Balkan cohort does have the effect of reversing the sign on the correlations between wages and NEET and between EDET and NEET. However, neither of those correlations had been large, in either direction, even with Bulgaria.

The reassignment of Estonia to an expanded group dominated by Visegrád does not alter the fundamental nature of either Visegrád or the Baltics. The reduced Baltic region of Latvia and Lithuania remains the one corner of eastern and central Europe where all three variables are positively correlated. The addition of Bulgaria and Estonia to an expanded Visegrád Group does not alter the unusual relationship among wage ratios, EDET, and NEET in the dominant countries of (north) central Europe.

Figure 3B
Correlations among variables within the three-cluster solution "A"



Source: Authors' illustration

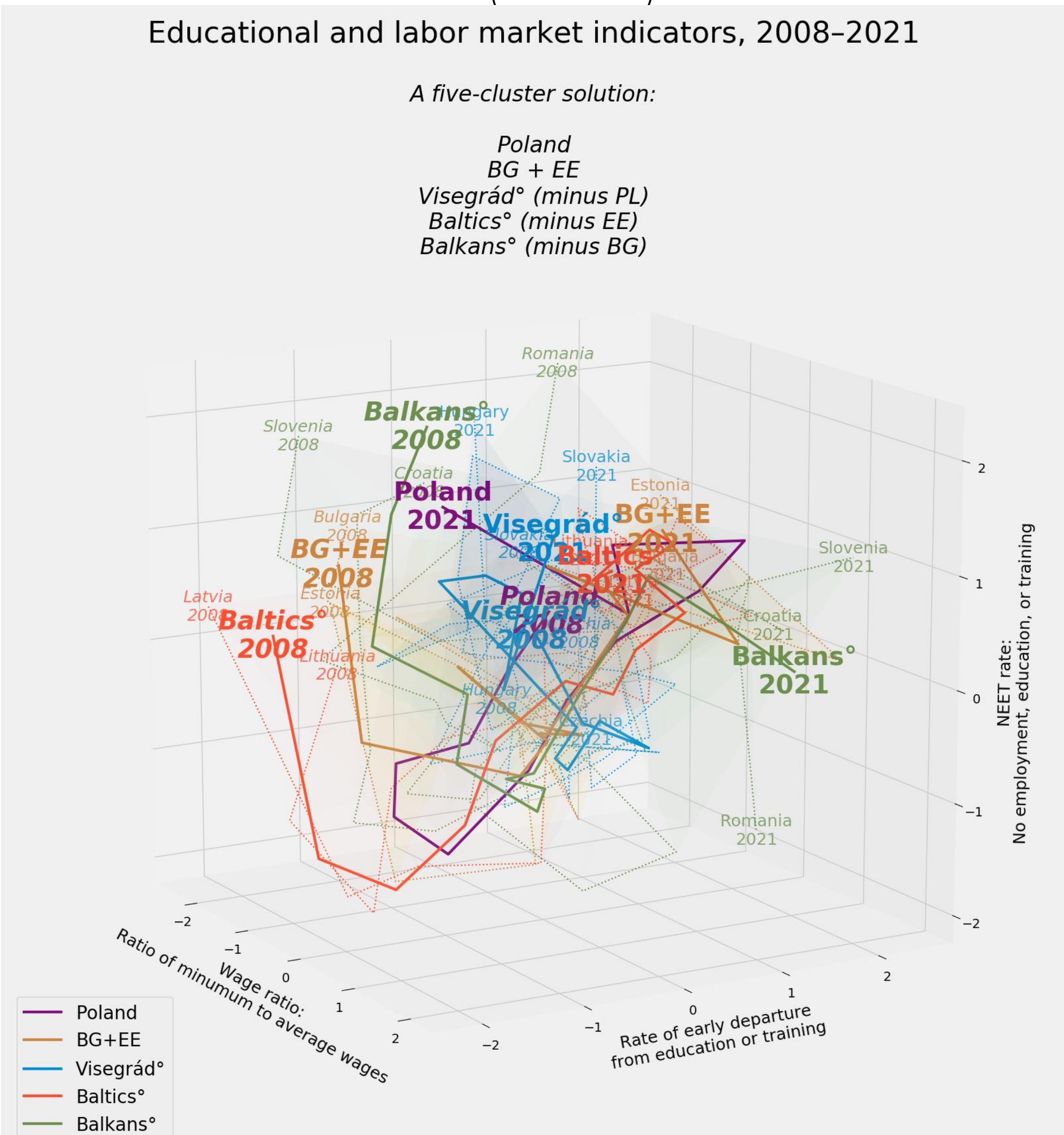
The results of the fourth 3D projection (Figure 4) provide a more granular view of central and Eastern Europe. By presenting five rather than three formal clusters, Poland stands out as a cluster of its own, suggesting unique labor market dynamics or policy outcomes relative to other countries. The presence of a singleton, in any clustering exercise, invites closer examination of that lone member's underlying traits.

The "BG+EE" cluster suggests similarities between Bulgaria and Estonia that isolate them from their respective regional groups. Despite their geographic separation, they share common pioneering trends in youth labor market reforms.

The remaining Visegrád countries, excluding Poland, still cluster together. The removal of Poland shows how Czechia, Hungary, and Slovakia have followed a distinct path relative to Poland. The mathematical definition of separate clusters supports the inference that these three Visegrád countries, whatever their internal variations, resemble each other more than Poland or, for that matter, other countries throughout central and eastern Europe.

The three-cluster solution ("A") and the five-cluster solution ("B") both deviate from conventional definitions of the three regions of central and eastern Europe. The reassignment of Bulgaria and Estonia, either to a larger cluster dominated by Visegrád, or to a new cluster consisting of those two countries, has the unavoidable effect of reducing the original, informal Baltic and Balkan clusters. In three-cluster solution A, the traditional regions of the Baltics and the Balkans each lose a member. Five-cluster solution B goes further. It isolates Poland entirely from the other members of the Visegrád Group.

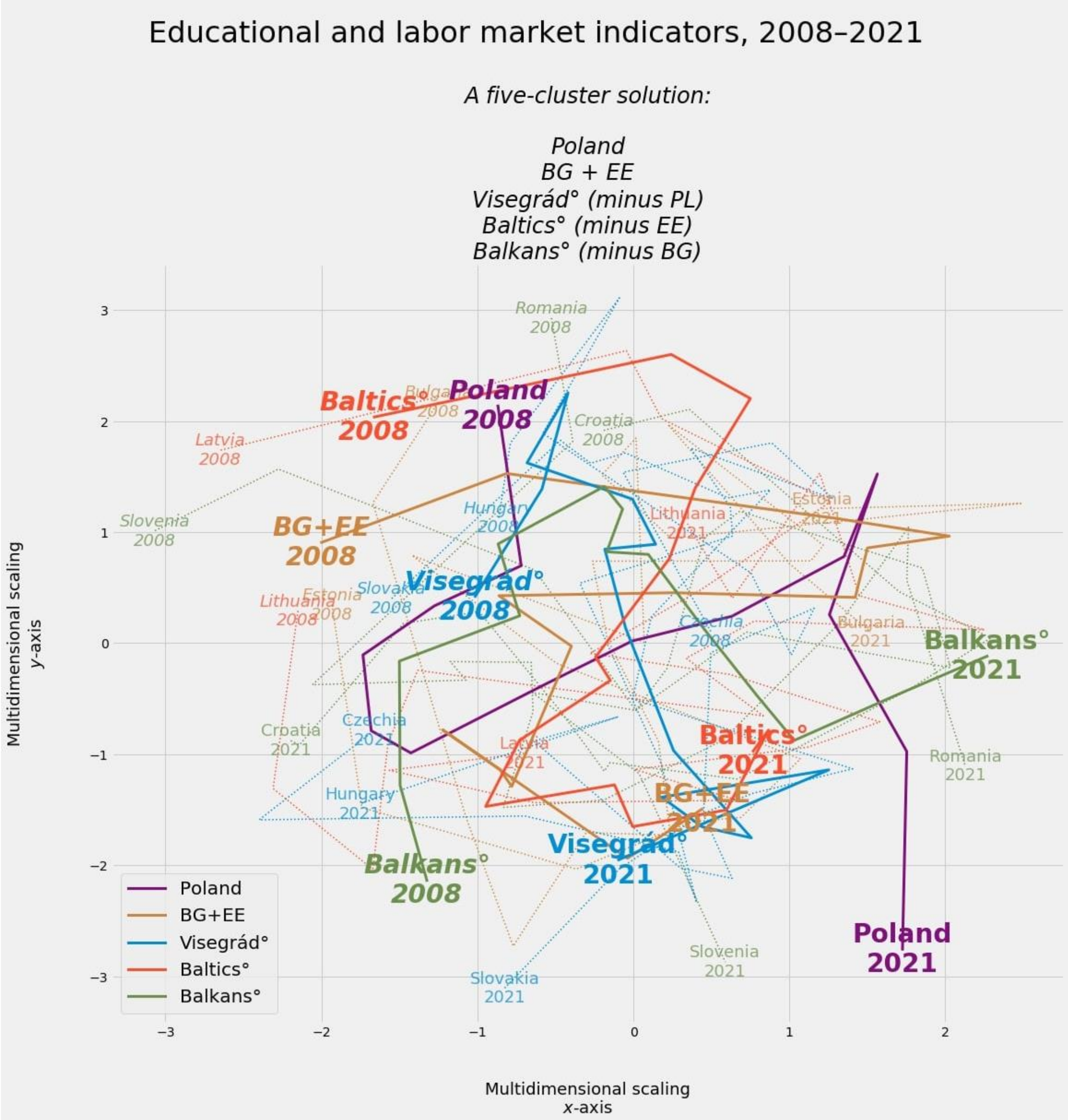
Figure 4
3D visualization of Cluster solution "B" (five clusters)



Source: Authors' illustration

A 2D variant of Figure 4 appears in Figure 4A. In cluster solution B, as depicted in Figures 4 and 4A, the Baltics (excluding Estonia) and the Balkans (excluding Bulgaria) maintain their closeness, perhaps due to shared regional characteristics or responses to economic conditions. The Baltic nations of Latvia and Lithuania show modest progress toward improving the school-to-work transition for young people, whereas a complex set of challenges peculiar to the Balkan region appears to have beset Romania and the ex-Yugoslav member-states of the EU.

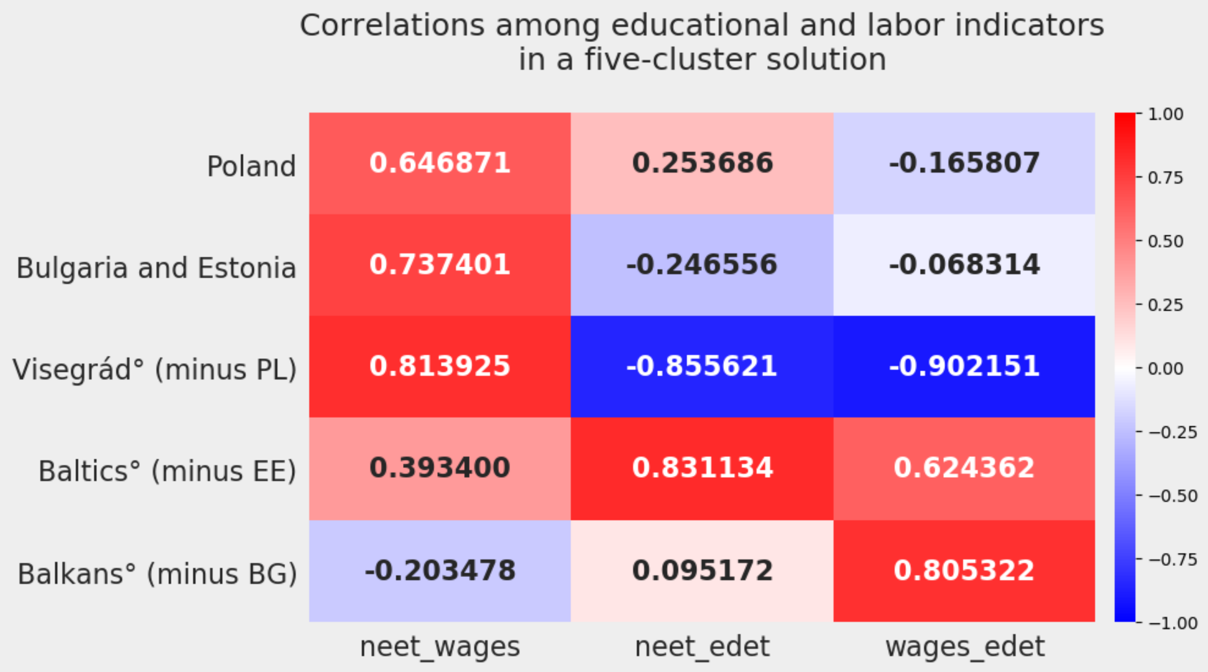
Figure 4A
2D visualization of Cluster solution "B" (five clusters) obtained with multidimensional scaling



Source: Authors' illustration

Correlations among variables within the five-cluster information realistically add only three new sets of information rather than five. The reduced Baltic° and Balkan° clusters from three-cluster solution A remain intact. Visegrád minus Poland retains its distinctive pattern of a negative correlation between EDET on one hand and the wage and NEET variables on the other hand. Poland shows a positive relationship between EDET and NEET. Bulgaria and Estonia, isolated on their own and removed from the expanded Visegrád cluster to which three-cluster solution A had assigned them, do reflect the Visegrád countries' collectively negative correlation between EDET and the wage and NEET variables. But the correlation between EDIT and NEET is very mildly negative, and between EDET and wages even more modestly negative.

Figure 4B
Correlations among variables within the five-cluster solution "B"



Source: Authors' illustration

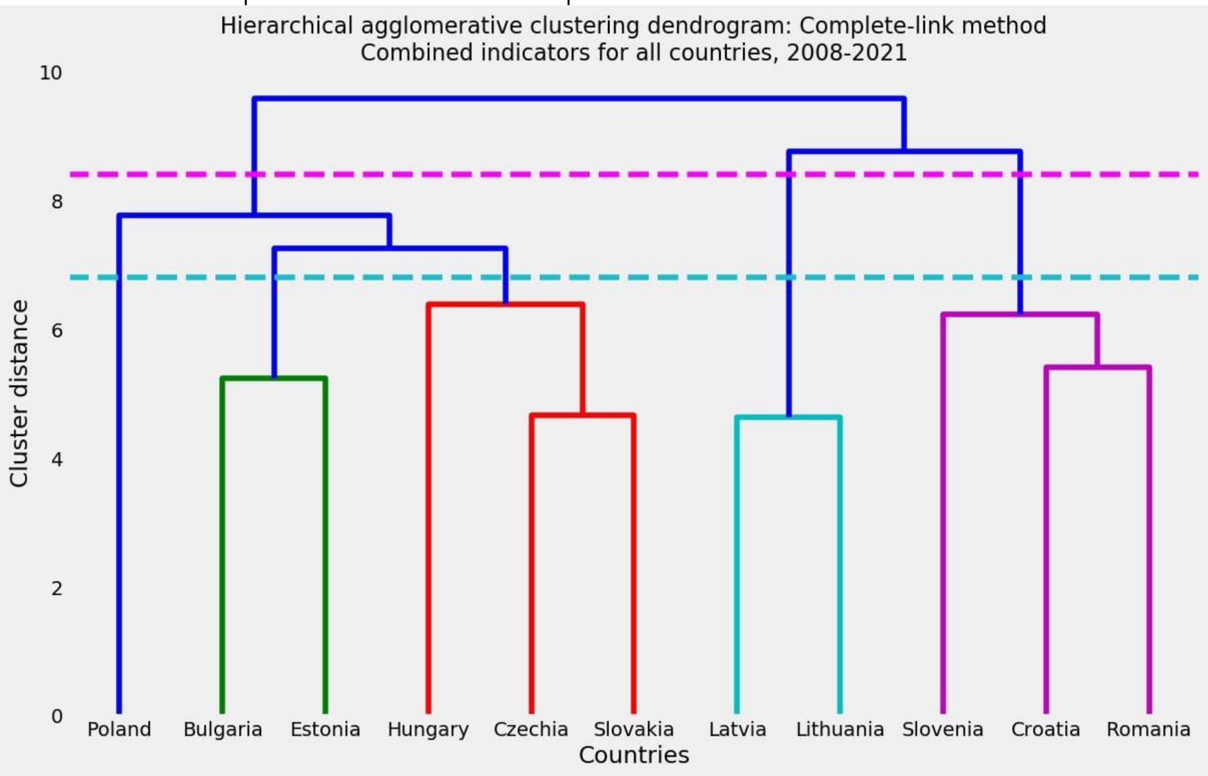
Dendrogram

The lone dendrogram in this study visualizes the results of hierarchical agglomerative clustering using complete linkage based on standardized Euclidean distances. Figure 5 shows how central and eastern European countries are arranged within three- and five-cluster solutions. Although these three- and five-cluster solutions have already been presented in Figures 3 and 4, the dendrograms in this section illustrate those results in a more mathematically formal way. Figure 5 also reveals a two-cluster solution not previously discussed. Thanks to their sheer simplicity, these visualizations also avoid the pitfalls that can beset complicated three-dimensional plots.

Figure 5 presents the three- and five-cluster solutions depicted in earlier 3D projections (Figures 3, 4). Within a five-cluster solution, Poland stands alone. Its status as a singleton indicates a distinct labor market profile from the rest of the group. Bulgaria and Estonia form a separate cluster, based on similarities in their labor market indicators. A larger cluster combines Hungary, Czechia, and Slovakia, implying these Visegrád countries share common labor market characteristics. Latvia, Lithuania, Slovenia, Croatia, and Romania cluster together at a higher distance, forming a plausible supercluster distinct from its counterpart, a Visegrád-dominated supercluster consisting of Poland, Czechia, Hungary, Slovakia, Bulgaria, and Estonia (indicated in Figure 4 as "Visegrád + BG + EE"). Dividing the non-Visegrád supercluster into two distinct components, one containing the Balkan nations of Croatia, Romania, and Slovenia, and other combining Latvia and Lithuania, yields a three-cluster solution.

Combining 3D projections with dendrograms enriches the analysis by providing a macroscopic and microscopic view of the data. While 3D projections give a spatial sense of the data, dendrograms offer detailed, formal insights into the hierarchical relationships and clustering process. The final clustering solutions are determined by considering the cluster distance at which clusters merge, with a larger vertical distance signifying less similarity. Significant increases in the dendrogram's vertical distances suggest natural divisions between clusters.

Figure 5
Dendrogram of combined indicators (wage ratios, EDET and NEET rates) for central and eastern European countries for the period 2008-2021



Source: Authors' illustration

Figure 5 presents at least two and possibly three potential clustering solutions. Notwithstanding previously presented three-cluster and five-cluster solutions, an even more simplified two-cluster solution distinguishes the Visegrád countries, Bulgaria, and Estonia from the rest of the region – namely most of the Baltic and Balkan subregions. Consequently, this analysis affirms previous selections and maintains both the three-cluster solution (displayed as "Cluster solution A" in Figure 3) and the five-cluster solution (displayed as "Cluster solution B" in Figure 4).

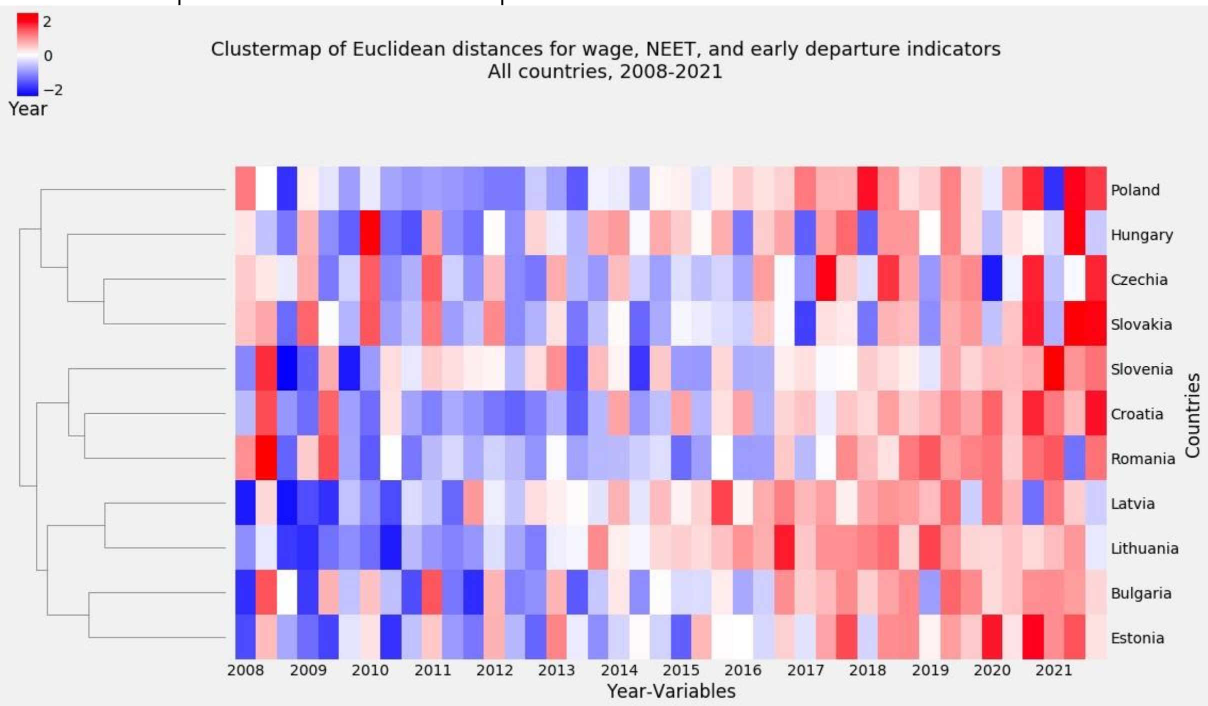
Heatmap

Figure 6 presents a heatmap, the last graphical tool in this study. Colors on a heatmap reflect the level of similarity or difference. Colors closer to blue indicate smaller standardized Euclidean distances and, therefore, denote higher similarity. Colors closer to red represent larger distances and, therefore, indicate lower similarity.

The color comparison does not directly separate countries from one another. Rather, they indicate distance from the Gaussian mean for the time series representing each feature. A blue reading for Poland, for instance, indicates a lower value for that country and for a particular year during the range from 2008 to 2021, relative to values for the Kaitz wage ratio, for instance. A red reading for a higher-than-average result for that variable in a country at a specific moment in time.

The resulting spatial, or distance-based, arrangement among central and eastern European countries appears along the vertical axis of each heatmap, while the horizontal axis shows the one-way progression of years from 2008 through 2021. Each row therefore represents a single country's performance throughout this period. Each column shows country-by-country differences during a single year.

Figure 6
Heatmap of combined indicators (wage ratios, EDET, and NEET rates) for central and eastern European countries for the period 2008-2021



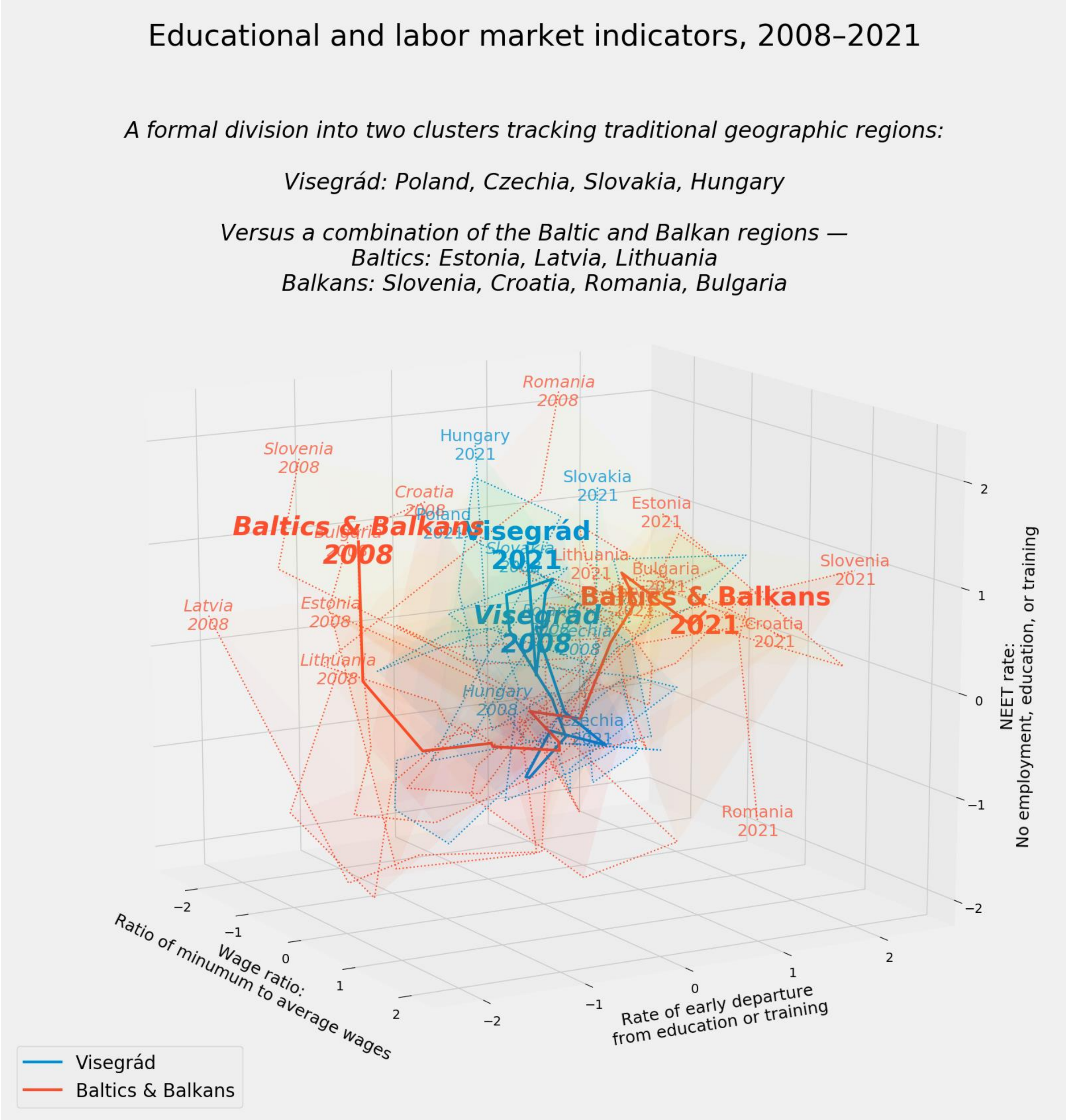
Source: Authors' illustration

The heatmap in Figure 6 also incorporates a dendrogram. This is an artifact of the Seaborn package for Python, which performs hierarchical clustering while generating heatmaps. Because the default linkage for hierarchical clustering within seaborn heatmaps is average rather than complete linkage, this heatmap has the incidental benefit of revealing clustering solutions that might differ from those prescribed by the dendrogram in Figure 5.

Figure 6 enables the distinction between two and five clusters. The former divides central and eastern Europe into two blocs: the Visegrád Group on one side and a grand union of the Baltics and Balkans on the other. The Baltic/Balkan union within Figure 6 does not distinguish perfectly between those regions. However, Bulgaria is closer to Estonia and other Baltic nations than it is to other Balkan countries.

Figure 6 also supports a five-cluster solution identical to that identified in Figure 4 (Cluster solution "B"). Under any clustering solution, Poland stands at a greater distance from other central and eastern European countries. The two-cluster configuration implied by the heatmap in Figure 6 is compelling. It consists simply of the Visegrád Group on one side and a grand north-south alliance of the Baltic and Balkan nations on the other side.

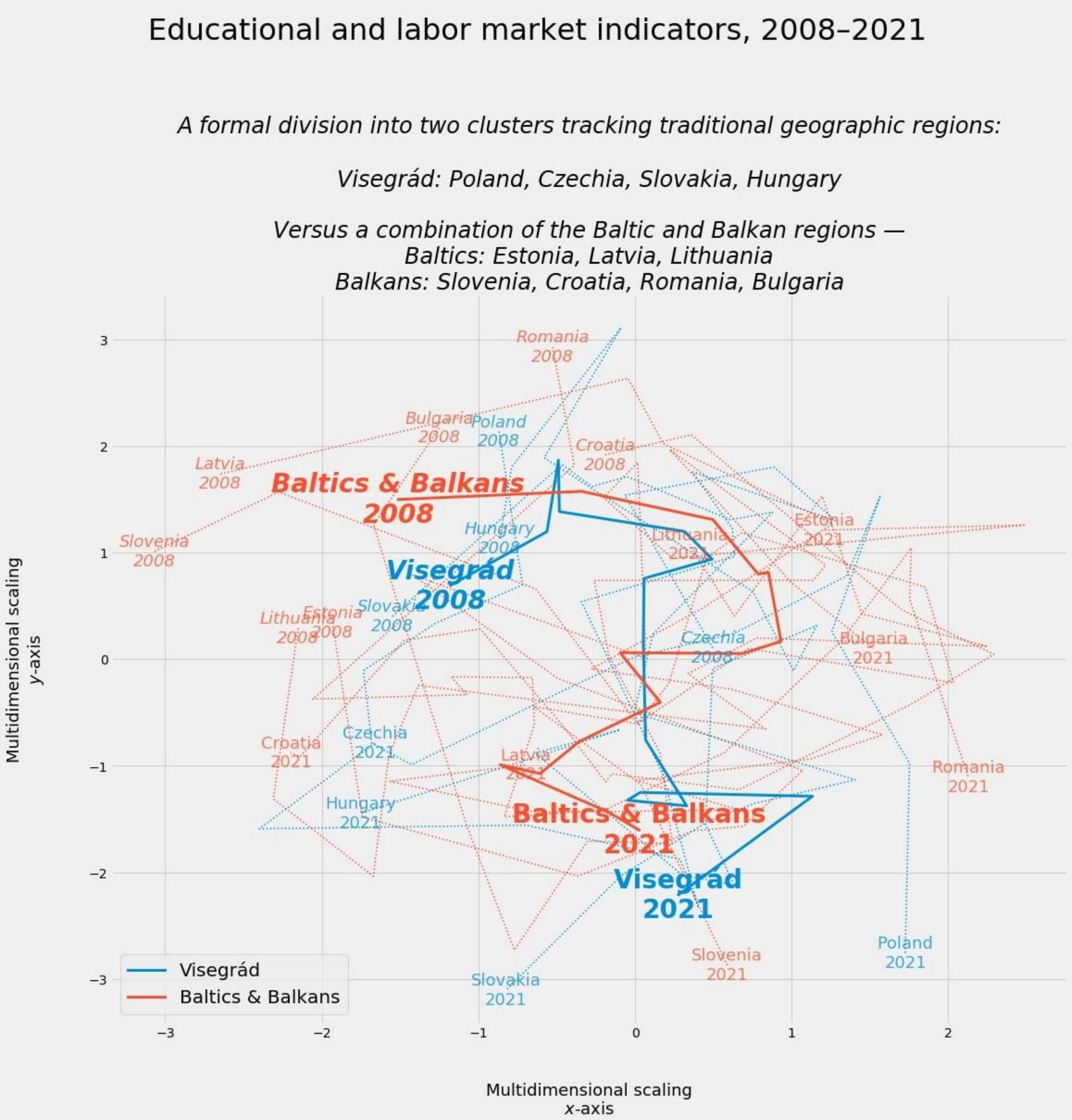
Figure 7
3D visualization of Cluster solution "C" (two clusters), as suggested by the average-linkage hierarchical clustering implied by the heatmap in Figure 6



Source: Authors' illustration

Figure 7 revisits the idea of a 3D projection, this time reduced to this two-cluster solution. The alignment of central and eastern Europe along one of the most geographically, culturally, and historically satisfying boundaries in the continent's history is an intuitive and satisfying result. Figure 7A presents the same information as a two-dimensional, MDS-enabled manifold.

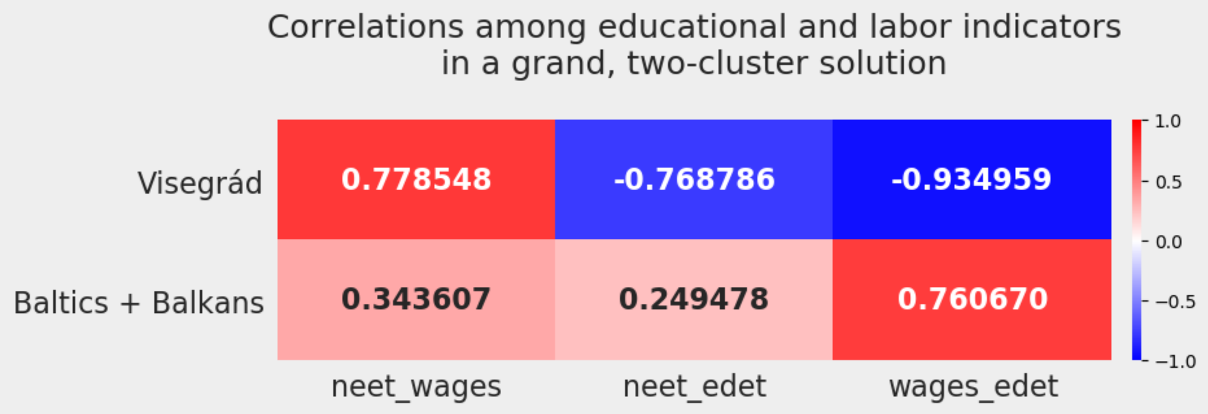
Figure 7A
2D visualization of Cluster solution "C" (two clusters), as suggested by the average-linkage hierarchical clustering implied by the heatmap in Figure 6



Source: Authors' illustration

Even with no more than two cohorts of countries, a correlation table reveals how the three underlying time series relate to one another (Figure 7B), which summarizes the correlations among the Kaitz, EDET, and NEET variables in the two clusters. The seven EU member-states along the Baltic Sea and on the Balkan Peninsula report positive correlations among all three variables. The four large states in the Visegrád Group report a negative relationship between EDET and the other variables. In microcosm, contemporary central and eastern Europe. The contrast between the Visegrád countries at the geographic and historic core of central and eastern Europe, on one hand, and the Baltics and Balkans, on the other hand, is reminiscent of the historical rivalry between the "inner six" nations that signed the original Treaty of Rome and the "outer seven" members of the European Free Trade Association (Kaiser, 1997).

Figure 7B
Correlations among variables within the grand two-cluster solution "C"



Source: Authors' illustration

Discussion

Having presented all the results of its clustering analysis, this paper now discusses the forces that may propel the youth labor market dynamics of central and eastern Europe. Again, this paper draws upon the reports of national and international organizations and national and EU policies aimed at education and school-to-work transition.

From the Baltic to the Black and Adriatic seas, central and eastern Europe stands out as a dynamic player within the European Union. Despite a common socialist history, the countries within this region exhibit diverse policies and responses to labor market challenges. Human judgment, uninformed by formal mathematical analysis, might be inclined to align these countries into three groups according to conventional geographical and political understandings: the Visegrád Group, the Baltics, and the Balkans.

However, hierarchical clustering enables deeper than superficial definitions by exploring economic reasons tied to labor market policies. This study has focused on three variables: wage policies and educational outcomes measurable by EDET and NEET rates. This research uses 3D projections to enhance our discussion of competing solutions, which offer two-, three-, and five-cluster solutions describing how countries in this region have strived to keep young people in school and to help them avoid unproductive labor market outcomes.

Wage ratios

Central and eastern European countries have actively increased legal minimum wages between 2008 and 2021 (ILOSTAT Explorer, 2024). Czechia experienced stagnation in minimum wage growth during the global financial crisis of 2008-2013 (Grossmann, 2021). Freezing wages at 8,000 CZK negatively affected the ratio of minimum to mean wages. Nonetheless, from 2008 to 2021, Czechia saw a 90 percent increase in the minimum wage.

Slovakia and Poland recorded the largest increases in wage ratio. Increases exceeding 10 percentage points can be attributed to significant hikes in minimum wages by 132 percent and 149 percent, respectively. Poland, in particular, exhibited the highest wage ratio throughout the period, exceeding 50 percent in 2021.

Among central and eastern European countries, Poland arguably distinguished itself with the best policy response during the global economic crisis by implementing countercyclical measures (Epstein et al., 2012; Piatkowski, 2015). Poland was the only

EU economy to avert a recession during that crisis. These reasons support clustering solutions that treat Poland as a singleton.

Latvia and Lithuania exhibited similar increases in their wage ratios, around 4 percentage points each during the observed period. In contrast, the Balkan nations of Slovenia, Croatia, and Romania, alongside Poland, ranked favorably in terms of wage ratio. Notably, Slovenia led all of central and eastern Europe with a 55.2 percent wage ratio, surpassing the EU's minimum wage directive threshold of 50 percent (European Parliament, 2022). Slovenia implemented two significant increases in the minimum wage in 2010 and 2018 (Slovenian Press Agency – STA, 2020).

The EDET rate

Early departures from compulsory schooling merit further discussion under a three-cluster solution (Figure 3). From 2008 to 2021, all four countries of the Visegrád Group witnessed an increase in EDET rates. On the other hand, Czechia, Slovakia, and Poland have already fallen below the EU's 2030 threshold of 9 percent. Hungary continues to struggle with a persistent rate of approximately 12 percent. Despite the implementation of programs and strategies aimed at reducing dropout rates among compulsory school attendees and providing support for disadvantaged students (such as the “Mid-Term Strategy on Early School Leaving 2014-2020”, the “Springboard Programme”, and the “Tanoda Programme”) (Lénárt, 2021; Youth Wiki, 2023f), Hungary has not made significant progress.

Bulgaria and Estonia have managed to reduce their EDET rates, with more substantial reductions during times of crisis. This can be partially attributed to the lower opportunity costs of education during economic downturns, as finding employment becomes more challenging (Adamopoulou & Tanzi, 2017; Borjas, 2013; Long, 2014). Nonetheless, these countries still need to meet the EU's 2030 threshold.

Latvia and Lithuania also show promising results in NEET indicators. Notably, Latvia has reduced its dropout rate by more than half from 2008 to 2021, a success that can be linked to its long-term and sustainable growth strategy in anticipation of 2030. Both of these Baltic countries meet the EU threshold thanks to the development of early warning systems, such as electronic school diaries for parental monitoring of school absences, efforts to track and support low achievers, and youth homes for disadvantaged students.

Croatia and Slovenia boast the lowest EDET rates in all of central and eastern Europe, around 2 to 3 percent in 2021. Unsurprisingly, these countries lack a national strategy specifically targeting this issue.

Romania, conversely, has faced the highest EDET rates throughout the period from 2008 to 2021. Difficulties with dropouts arise largely from social exclusion affecting primarily four groups (Youth Wiki, 2023g): (a) young people aged 18-24, (b) young people from low-income families, (c) young people in rural areas, and (d) ethnic minorities such as the Roma. Despite Romania’s efforts to reduce the EDET rate within the observed timeframe, it continues to lag the rest of Europe. Romania’s EDET rate stands at 15.3 percent as of 2021.

The EU’s Youth Guarantee program is aimed as much at solving the EDET issue as it is at addressing NEET challenges. The two concerns overlap considerably. To some extent, preventing early departures from education or training can also mitigate future failures to find work, training, or education. In other words, policies targeting the EDET rate may reduce the future incidence of NEET. Therefore, it is crucial to accurately identify the problem, develop tailored programs that respect the diversity of vulnerable groups, and actively implement and regularly update these programs to reintegrate as many young people as possible into the education system, labor market, and society at large.

The NEET rate

The countries of central and eastern Europe are working toward an EU-level goal of reducing the NEET rate to below 9 percent by 2030. Since its introduction in the 1990s in the United Kingdom, where it replaced the term "Status Zer0" (Istance, Rees, & Williamson, 1994, as cited in Mascherini, 2018), the NEET concept has evolved, particularly regarding the age groups it encompasses. While the EU target focuses on the 15-29 age group, our analysis concentrates on the 25-34 age group, where being NEET represents a more significant concern.

Cluster solution "A" identifies varied challenges related to NEET. Despite fluctuations, three of the four Visegrád countries (Hungary, Poland, and Slovakia) had reduced their NEET rates by 2021. Hungary showed the most significant decrease of 10.9 percentage points. By contrast, Czechia experienced stagnation, as its NEET rate stayed persistently close to 20 percent.

Bulgaria and Estonia are addressing similar challenges in retaining young people in secondary vocational education and training. School dropouts, especially from a track consciously designed to serve students who do not intend to continue their education after high school, are a crucial component of the NEET phenomenon. Secondary school dropouts often join the NEET population. Despite vastly different NEET rates (22 percent in Bulgaria and 13 percent in Estonia in 2021), both countries face common structural issues and have pursued similar national policies. Bulgaria has initiated programs (Youth Wiki, 2023b) to support the long-term unemployed and young people who are not taking part in the labor market ("Program for Training and Employment of Long-Term Unemployed Persons" and "Activation of Inactive Persons"). Estonia has launched similar initiatives (Youth Wiki, 2023c) such as "My First Job" and "Youth Prop Up."

Latvia and Lithuania experienced similar dynamics, with a NEET rate that peaked in 2009-2010. A general decrease in NEET rates ensued, thanks to successful national intervention projects (Youth Wiki, 2023d, 2023e) such as "Know and Do!" (Latvia) and "Discover Yourself" and "New Start" (Lithuania).

The Balkan trio of Croatia, Slovenia, and Romania saw an increase in NEET rates during 2008-2013. As of 2021, Slovenia is closest to the 2030 EU target with a 10.1 percent NEET rate, while Romania lags at 24.4 percent. Croatia falls in between. All three countries report higher NEET rates in 2021 compared to 2008.

This analysis of NEET rates justifies the preference for a three-cluster solution. However, it is important to consider other factors shaping the youth labor market, such as income distribution and the quality of the educational system. Incorporating more variables could yield further insights.

All of the countries in this study have embraced the EU's "Youth Guarantee" program (adopted in 2013; Council of the EU, 2013) aimed at addressing NEET challenges after the global financial crisis. The "Reinforced Youth Guarantee" (2020, Council of the EU, 2020) responded to the COVID-19 crisis, serving as EU guidelines for developing national programs tailored to local market needs. The long-term impact of this strategy on young people in the European labor market remains to be seen.

Conclusion

This paper has examined central and eastern Europe during a period marked by two global economic shocks: the global financial crisis (2008-2013) and the COVID-19 pandemic (2020-2021). It focuses on the challenges faced by young people. This paper has explored these challenges by analyzing wage ratios and EDET and NEET rates. This research makes a distinctive contribution through its novel graphical exploration of the youth labor market. Through 3D projections, dendrograms, and heatmaps, this study offers new insights into the dynamics underlying that market.

Hierarchical clustering with complete linkage, as applied to time series data spanning 2008 through 2021, reveals multiple cluster solutions. Two of those solutions appeared illustrative of the youth labor market's dynamics: a narrower three-cluster solution that divides this part of Europe into (1) the Visegrád Group plus Bulgaria and Estonia, (2) the remaining Baltic countries, and (3) the remaining Balkan countries.

A five-cluster solution further subdivides the first, Visegrád-dominated cluster into a Polish singleton, the geographically remote pair of Bulgaria and Estonia, and the remaining Visegrád countries of Czechia, Hungary, and Slovakia. The remaining two clusters consist, as they do in the three-cluster solution, of the Baltic states (minus Estonia) and the Balkan states (minus Bulgaria).

An alternative average-linkage approach to hierarchical clustering supports a historically and geographically intuitive bifurcation of central and eastern Europe into the Visegrád Group and a grand north/south coalition of the Baltic and Balkan countries.

The analysis highlighted standout performers: Slovenia excels in combating EDET challenges (alongside Croatia) and achieving high wage ratios (alongside Poland). Slovenia also boasts the region's lowest NEET rates. The Baltic countries have shown respectable progress.

These findings suggest that a strategic shift might be beneficial. Rather than solely promoting tertiary education, policymakers should place heightened focus on preventing dropout from compulsory education. This appears to be an especially significant issue in Bulgaria, Hungary, and Romania. Enhancing vocational education and training programs is crucial. As EU members, all countries in this study are aligned with common labor market strategies, especially the Youth Guarantee program. The enduring influence of political and economic legacies is apparent. Geopolitical proximity influences regional clustering and the unique challenges each bloc faces.

The primary limitations of this study arise from the same traits that supply its leading strengths. As a species of unsupervised machine learning, clustering analysis cannot directly support causal inferences or validate hypotheses regarding drivers of youth labor markets. Those research questions are better addressed by generalized linear methods and related forms of supervised machine learning. Studies such as Tudor et al. (2023) do combine clustering analysis with OLS regression.

A study combining unsupervised and supervised learning should be able to achieve at least two tasks beyond those attainable through clustering alone. First, studies incorporating regression should aspire to predict future labor market outcomes. Second, such studies should extract causal inferences related to hypotheses on the relationship of education, training, and other labor market inputs with ultimate labor market outcomes.

The identification of specific clusters within part or all of a multinational union such as the European Union is the primary contribution of a study such as this one. Predictions, causal inferences, and policy prescriptions may all be improved if researchers, guided by clustering analysis, can tailor their conclusions and recommendations according to geopolitical differences revealed by the data.

Other limitations of this research include the choice of average wage over median wage for calculating wage ratios due to incomplete data across several countries. In line with studies such as Dingeldey and Buttler (2023); Krpan, Gardijan Kedžo, and Žmuk (2023); and Tudor et al. (2023), future research should consider additional variables such as income distribution measures, socioeconomic indicators, and indicators of educational quality. More comprehensive collection and evaluation of distinct aspects of educational inputs and labor market outcomes could overcome some of the limitations of this study. In particular, a more detailed analysis of the NEET rate, distinguishing between active and inactive youth, could further refine policymakers' understanding.

Consistent with its stated research hypothesis, this article has classified similarities and differences in the school-to-work transition in central and eastern Europe. Conventional historical and geopolitical understandings sort the eleven countries in this article into the following three groups: the Visegrád countries, the Baltics, and the Balkans. Rigorous mathematical evaluation of three youth labor market variables, attained through hierarchical clustering and multidimensional scaling, supports two-, three-, and five-cluster solutions that partially reflect the conventional geopolitical taxonomy, but ultimately recommend more carefully tailored ontologies that should guide the formulation of labor policy in this vital region within the European Union.

Understanding the current labor market situation, as shaped by past governmental decisions, is vital to effective policymaking. The dynamic environment, evolving political landscape, and lingering historical influences all challenge the EU's convergence goals. Advanced economies are converging at a faster pace, while less developed ones are facing divergence pressures (Rambla & Scandurra, 2021). The European Union must ask whether convergence is a realistic goal, or whether the EU should strive instead for regional convergence. Central and eastern Europe stands at a crossroads, facing the challenge of meeting the EU's 2030 goals amid these complexities.

References

1. Adamecz, A. (2023). Longer schooling with grade retention: The effects of increasing the school leaving age on dropping out and labour market success. *Economics of Education Review*, *97*, 1-33. https://doi.org/10.1016/j.econedurev.2023.102487
2. Adamopoulou, E., & Tanzi, G., M. (2017). Academic Drop-Out and the Great Recession. *Journal of Human Capital, 11(1)*, 35-71. Retrieved from https://doi.org/10.1086/690650
3. Aggarwal, C., C., Hinneburg, A., & Keim, D., A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. *Database Theory – ICDT 2001. Lecture Notes in Computer Science, Springer*, Berlin/Heidelberg, 1973, 420-434. https://doi.org/10.1007/3-540-44503-X_27
4. Arpaia, A., Cardoso, P., Kiss, A., Van Herck, K., & Vandeplas, A. (2017). Statutory Minimum Wages in the EU: Institutional Settings and Macroeconomic Implications. *IZA Policy Paper No. 124, Institute of Labor Economics*, 1-49.
5. Askenazy, P. (2003). Minimum wage, exports, and growth. *European Economic Review, 47(1)*, 147-164. https://doi.org/10.1016/S0014-2921(02)00187-3
6. Bernard, A., B., & Durlauf, S., N. (1995). Convergence in international output. *Journal of Applied Econometrics*, *10*(2), 97-108. https://doi.org/10.1002/jae.3950100202
7. Bernard, A., B., & Durlauf, S., N. (1996). Interpreting tests of the convergence hypothesis. Journal of Econometrics, *71*(1-2), 161-173. https://doi.org/10.1016/0304-4076(94)01699-2
8. Bezdek, J., C. (2021). *Elementary Cluster Analysis: Four Basic Methods that (Usually) Work*. Gistrup, Denmark: River Publishers.
9. Bieszk-Stolorz, B., & Dmytrów, K. (2020). Influence of Accession of the Visegrad Group Countries to the EU on the Situation in Their Labour Markets. *Sustainability 2020, 12(16)*, 6694. https://doi.org/10.3390/su12166694
10. Blatná, D. (2020). Analysis of the Relationship between the Share of Early Leavers from Education and Training and Selected Socio-Economic Indicators in the Czech Republic in the Period 2005 – 2018. The 14th International Days of Statistics and Economics, Prague, September 10-12, 2020, 136-145.

11. Bokros, L. (2013). *Accidental Occidental: Economics and Culture of Transition in Mitteleuropa, the Baltic and the Balkan Area*. Budapest: Central European University Press.
12. Borjas, G., J. (2013). *Labor Economics* (6th ed.). New York, United States: McGraw-Hill Companies, Inc.
13. Brown, C., Gilroy, C., & Kohen, A. (1982). The effect of the minimum wage on employment and unemployment. *Journal of Economic Literature*, *20*(2), 487-528. Retrieved from https://www.jstor.org/stable/2724487
14. Burzykowski, T., Rousseau, A.-J., Geubbelmans, M., & Valkenborg, D. (2023). Introduction to Machine Learning. *American Journal of Orthodontics & Dentofacial Orthopedics, 163(5)*, 732-734, https://doi.org/10.1016/j.ajodo.2023.02.005
15. Council of the European Union (2013). Council Recommendation of 22 April 2013 on establishing a Youth Guarantee. Retrieved from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32013H0426%2801%29
16. Council of The European Union (2020). Council Recommendation of 30 October 2020 on A Bridge to Jobs – Reinforcing the Youth Guarantee and replacing the Council Recommendation of 22 April 2013 on establishing a Youth Guarantee 2020/C 372/01. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AC%3A2020%3A372%3ATOC&uri=uriserv%3AOJ.C_.2020.372.01.0001.01.ENG
17. Council of the European Union (2021). Council Resolution on a strategic framework for European cooperation in education and training toward the European Education Area and beyond (2021-2030), 2021/C, 66/01. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021G0226%2801%29&qid=1617706323424
18. Cox, M., A., A., & Cox, T., F. (2008). Multidimensional scaling. In Chen, C.-H., Härdle, W., & Unwin, A. (Ed.). *Handbook of Data Visualization* (pp. 315-347). Berlin: Springer. https://doi.org/10.1007/978-3-540-33037-0_14
19. Cuestas, J., C., Monfort, M., Ordóñez, J. (2021). The education pillar of the Europe 2020 strategy: a convergence analysis. *Empirica*, *48*(4), 1113-1129. https://doi.org/10.1007/s10663-021-09504-x
20. De Luca, G., Mazzocchi, P., Quintano, C., & Rocca, A. (2020). Going Behind the High Rates of NEETs in Italy and Spain: The Role of Early School Leavers. *Social Indicators Research, 151*(1), 345-363. https://doi.org/10.1007/s11205-020-02370-3
21. Dingeldey, I., & Buttler, D. (2023). School-to-work transition regimes in post-socialist countries: An introduction. In Buttler, D., Ławrynowicz, M., & Michoń, P. (Ed.). *School-to-Work Transition in Comparative Perspective* (pp. 156-173.). Cheltenham, UK, and Northampton, Massachusetts, USA: Edward Elgar Publishing. https://doi.org/10.4337/9781800370111.00015
22. Domingos, P. (2012). A Few Useful Things to Know about Machine Learning. *Communications of the ACM, 55*(10), 78-87. https://doi.org/10.1145/2347736.2347755
23. El Bouchefry, K., & de Souza, R., S. (2020). Learning in Big Data: Introduction to Machine Learning. In Škoda, P., & Adam, R. (Ed.), *Knowledge Discovery in Big Data from Astronomy and Earth Observation: AstroGeoInformatics*, pp. 225-249. St. Louis, Mo.: Elsevier. https://doi.org/10.1016/B978-0-12-819154-S.00023-3
24. Epstein, N., Goretti, M., Llaudes, R., & Velculescu, D., (2012). Poland: A Beacon of Resilience in Europe. In Bakker, B., B., & Klingen, C. (Ed.). *How Emerging Europe Came Through the 2008/09 Crisis* (pp. 155-162). International Monetary Fund. Retrieved from https://www.elibrary.imf.org/display/book/9781616353810/ch012.xml
25. European Commission (2010). EUROPE 2020 – A strategy for smart, sustainable and inclusive growth. Retrieved from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52010DC2020
26. European Commission (2017). Baltic Alliance for Apprenticeships. Enhancing the attractiveness of VET systems in the Baltic states through work-based learning and apprenticeships. Retrieved from https://ec.europa.eu/social/BlobServlet?docId=18440&langId=en

27. European Parliament, Council of the European Union (2022). Directive (EU) 2022/2041 of the European Parliament and of the Council of 19 October 2022 on adequate minimum wages in the European Union. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022L2041
28. Evangelista, P., F., Embrechts, M., J., & Szymanski, B., K. (2006). Taming the Curse of Dimensionality in Kernels and Novelty Detection. In: Abraham, A., de Baets, B., Köppen, M., Nickolay, B. (Ed.) Applied Soft Computing Technologies: The Challenge of Complexity. Advances in Soft Computing, Vol 34. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-31662-0_33
29. Fialová, K., & Mysíková, M. (2020). Minimum Wage and Youth Employment in Regions of the Visegrád Countries. *Eastern European Economics, 59*(1), 82-102. https://doi.org/10.1080/00128775.2020.1816474
30. Furlong, A. (2006). Not a very NEET solution: representing problematic labour market transitions among early school-leavers. *Work, Employment and Society, 20*(3), 553-569. https://doi.org/10.1177/0950017006067001
31. Geiger, B., C., & Kubin, G. (2012). Relative Information Loss in the PCA. *Proceedings of the 2012 IEEE Information Theory Workshop*, Lausanne, Switzerland, pp. 562-566. https://doi.org/10.1109/ITW.2012.6404738
32. Ghica, L.-A. (2008). Rhetorical Strategies, Institutional Dilemmas: The Visegrád Group and the Baltic Cooperation Facing the EU and NATO Accession Process. *Analele Universității din Bucureşti. Seria Ştiințe Politice*, *10*(10), 75-86. Retrieved from https://www.ceeol.com/search/article-detail?id=89903
33. Grossmann, J. (2021). The Effects of Minimum Wage Increases in the Czech Republic. *CERGE-EI Working Paper Series No. 679*. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3784096
34. Hair, J., F., Black, W., C., Babin, B., J., & Anderson, R., E. (2018). *Multivariate Data Analysis* (8th ed.). Andover, Hampshire, United Kingdom: Cengage Learning EMEA.
35. Institute for Market Economics (2019). *Assessment of the people not in employment, education and training (NEETs) in Bulgaria and policy measures to effectively address their integration*. European Commission, Luxembourg: Publications Office of the European Union, 1-70. https://doi.org/10.2767/768177
36. Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *WIREs Cognitive Science*, *4*(1), 93-103. https://doi.org/10.1002/wcs.1203
37. International Labour Organization. Statutory nominal gross monthly minimum wage – Annual. ILOSTAT Explorer. Retrieved from https://www.ilo.org/shinyapps/bulkexplorer31/?lang=en&id=EAR_4MMN_CUR_NB_A [January 31, 2024]
38. Istance, D., Rees, G., & Williamson, H. (1994). *Young People Not in Education, Training or Employment in South Glamorgan*. Cardiff, UK: South Glamorgan Training and Enterprise Council.
39. Jolliffe, I., T. (2002). *Principal Component Analysis* (2nd ed.). New York, United States: Springer-Verlag New York, Inc.
40. Kaiser, W. (1997). Challenge to the community: the creation, crisis and consolidation of the European Free Trade Association, 1958-72. *Journal of European Integration History*, *3*(1), 7-33.
41. Kaitz, H. (1970). Experience of the Past: The National Minimum. *Youth unemployment and minimum wages*, Bulletin 1657, United States Department of Labor, Bureau of Labor Statistics, 30-54.
42. Korotaj, T., Chen, J., M., & Kurnoga, N. (2023). Hierarchical Clustering of CEE Countries According to Educational and Labour Market Indicators. *Proceedings of the 17th International Symposium on Operational Research* (SOR'23, Bled, Slovenia), 270-276.
43. Kriegel, H.-P., Kröger, P., & Zimek, A. (2009). Clustering high-dimensional data. *ACM Transactions on Knowledge Discovery from Data, 3*(1), 1-58. https://doi.org/10.1145/1497577.1497578
44. Krpan, M., Gardijan Kedžo, M., & Žmuk, B. (2023). Exploring the link between education length and employment outcomes among youth in Europe: A hierarchical clustering

approach. *Business Systems Research*. *14*(2), 190-213. https://doi.org/10.2478/bsrj-2023-0019
45. Krzaklewska, E. (2013). Visegrad Youth – Comparative review of the situation of young people in the V4 countries. *Council of Europe*. Warszava, Poland.
46. Kunić, J. (2022). The Future of the European Union and the European Integration. In Dimitrijević, D., & Mileski, T. (Ed.). *International Organizations: Serbia and Contemporary World* (1 (2022), 448-457). https://doi.org/10.18485/iipe_ioscw.2022.1.ch26
47. Lénárt, A. (2021). Hungary. Profile commissioned by EASNIE for the Global Education Monitoring Report 2021 – Central and Eastern Europe, the Caucasus and Central Asia – Inclusion and education: All means all. Retrieved from https://gem-report-2020.unesco.org/wp-content/uploads/2021/02/Hungary.pdf
48. Lenhart, O. (2017). The impact of minimum wages on population health: evidence from 24 OECD countries. *European Journal of Health Economics, 18*(8), 1031-1039. https://doi.org/10.1007/s10198-016-0847-5
49. Long, B., T. (2014). The Financial Crisis and College Enrollment: How Have Students and Their Families Responded? In Brown, J., R., & Hoxby, C., M. (Ed.). *How the Financial Crisis and Great Recession Affected Higher Education* (pp. 209-233). Chicago, United States: University of Chicago Press. Retrieved from https://www.nber.org/system/files/chapters/c12862/c12862.pdf
50. Macedo, E., Nairz-Wirth, E., Araújo, H. C., & Szalai, J. (2020). Drawing lessons from early school leavers in a social justice context: Introductory comments. *European Educational Research Journal, 19(5)*, 387-397. https://doi.org/10.1177/1474904120951658
51. Marimont, R., B.; & Shapiro, M., B. (1979). Nearest neighbour searches and the curse of dimensionality. *IMA Journal of Applied Mathematics*. *24(1)*, 59-70. https://doi.org/10.1093/imamat/24.1.59
52. Mascherini, M. (2018). Origins and future of the concept of NEETs in the European policy agenda. In O'Reilly, J., Leschke, J., Ortlieb, R., Seeleib-Kaiser, M., & Villa, P. (Ed.). *Youth Labor in Transition: Inequalities, Mobility, and Policies in Europe* (pp. 503-529). Oxford Academic. Retrieved from https://academic.oup.com/book/26877/chapter/195940054
53. Mehmetaj, N., & Zulfiu Alili, M. (2020). Youth Graduates Employment Country Studies of Albania and North Macedonia in Economics Programs. *ENTRENOVA — ENTerprise Research InNOVAtion,* 6(1), 275-290. Retrieved from: https://econstore.eu/bitstream/10419/224695/1/26-ENT-2020-Mehmetaj-275-290.pdf
54. Müller, A., C., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. Sebastopol, Cal.: O'Reilly Media.
55. Neagu, G., Lendzhova, V., & Keranova, D. (2021). The Specifics of the NEETs Population in Bulgaria and Romania. *Postmodernism Problems, 11*(2), 112-123. https://doi.org/10.46324/PMP2102112
56. Nič, M. (2016). The Visegrád Group in the EU: 2016 as a Turning-Point? *European View*, *15*(2), 281-290. https://doi.org/10.1007/s12290-016-0422-6
57. Nielsen, F. (2016). Hierarchical clustering. In *Introduction to HPC with MPI for Data Science* (195-211). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-21903-5_8
58. O'Higgins, N., & Moscariello, V. (2017). *Labour market institutions and youth labour markets: Minimum wages and youth employment revisited*. Employment Working Paper No. 223. Geneva, Switzerland: International Labour Office, Employment Policy Department.
59. Olaode, A., Naghdy, G., & Todd, C. (2014). Unsupervised Classification of Images: A Review. *International Journal of Image Processing*, *8*(5), 325-342. https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJIP-918
60. Patzina, A., & Wydra-Somaggio, G. (2020). Early Careers of Dropouts from Vocational Training: Signals, Human Capital Formation, and Training Firms. *European Sociological Review, 36*(5), 741-759. https://doi.org/10.1093/esr/jcaa011

61. Piatkowski, M. (2015, June 12). *Four ways Poland's state bank helped it avoid recession*. Retrieved from https://www.brookings.edu/articles/four-ways-polands-state-bank-helped-it-avoid-recession/
62. Portela Pruaño, A., Rodríguez Entrena, M., J., Torres Soto, A., & Nieto Cano, J., M. (2022). Why vulnerable early school leavers return to and re-engage with education: Push and pull reasons underlying their decision. *Intercultural Education, 33*(2), 156-172. https://doi.org/10.1080/14675986.2021.2018209
63. Rambla, X., & Scandurra, R. (2021). Is the distribution of NEETs and early leavers from education and training converging across the regions of the European Union? *European Societies, 23*(5), 563-589. https://doi.org/10.1080/14616696.2020.1869282
64. Sakellariou, A., Sanoudou, D. & Spyrou, G. (2012). Combining multiple hypothesis testing and affinity propagation clustering leads to accurate, robust and sample size independent classification on gene expression data. *BMC Bioinformatics, 13*, 270. https://doi.org/10.1186/1471-2105-13-270
65. Schmarje, L., Santarossa, M., Schröder, S.-M., & Koch, R. (2021). A Survey on Semi-, Self- and Unsupervised Learning for Image Classification. *IEEE Access*, *9*, 82146-82168. https://doi.org/10.1109/ACCESS.2021.3084358
66. The Slovenian Press Agency – STA (2020, August 24). *Slovenia Has Smallest Difference Between Minimum & Average Wages in EU*. Retrieved from https://www.total-slovenia-news.com/business/6818-slovenia-has-smallest-difference-between-minimum-average-wages-in-eu
67. Supreme Court of the United States (1921). New York Trust Company v. Eisner. *United States Reports*, *256*, 345-350.
68. Toderiţă, A., Damian, A., & Meiroşu, C. (2019). *Access of young vulnerable NEETs to Youth Guarantee measures in Romania. Current status and way forward.* CRPE Policy Memo no. 76, 2019. Retrieved from https://www.crpe.ro/wp-content/uploads/2019/09/CRPE-Policy-Memo76_Access-of-young-NEETs-coming-from-vulnerable-groups-to-YG-1.pdf
69. Tudor, S., Cilan, T., F., Năstase, L., L., Ecobici, M., L., Opran, E., R., Cojocaru, A., V. (2023). Evolution of interdependencies between education and the labor market in the view of sustainable development and investment in the educational system. *Sustainability*. *15*(5), 3908. https://doi.org/10.3390/su15053908
70. Valkenborg, D., Rousseau, A.-J., Geubbelmans, M., & Burzykowski, T. (2023). Unsupervised learning. *American Journal of Orthodontics & Dentofacial Orthopedics, 163*(6), 877-882. https://doi.org/10.1016/j.ajodo.2023.04.001
71. Vasilescu, M., D., Stănilă, L., Popescu, M., E., Militaru, E.; & Marin, E. (2024). Using panel data clustering regression analysis to revisit income inequalities in the European Union. *Applied Economics Letters*, 1-6. https://doi.org/10.1080/13504851.2024.2358186
72. Vodă, A., I., Bercu, A-M., & Sebestova, J. (2021). Is a higher minimum wage associated with a higher youth employment rate? A panel data analysis. *European Journal of International Management, 16*(4), 631-646. https://doi.org/10.1504/EJIM.2021.118569
73. Watson, D., S. (2023). On the Philosophy of Unsupervised Learning. *Philosophy & Technology*, *38*, 28. https://doi.org/10.1007/s13347-023-00635-6
74. Williams, N., & Mills, J., A. (1998). Minimum wage effects by gender. *Journal of Labor Research, 19*(2), 397-414. https://doi.org/10.1007/s12122-998-1022-2
75. Xu, N., Finkelman, R., B., Dai, S., Xu, C., & Peng, M. (2021). Average Linkage Hierarchical Clustering Algorithm for Determining the Relationships between Elements in Coal. *ACS Omega*, *6*(9), 6206-6217. https://doi.org/10.1021/acsomega.0c05758
76. Youth Wiki (2020). Youth policies in Croatia – 2020. European Commission. Retrieved from https://national-policies.eacea.ec.europa.eu/sites/default/files/2021-09/Croatia_2020_0.pdf
77. Youth Wiki (2023a). Youth Policy Governance, National youth strategy: Slovenia. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/slovenia/13-national-youth-strategy
78. Youth Wiki (2023b). Employment and Entrepreneurship, Integration of Young People in the Labour Market: Bulgaria. Retrieved from https://national-

policies.eacea.ec.europa.eu/youthwiki/chapters/bulgaria/36-integration-of-young-people-in-the-labour-market
79. Youth Wiki (2023c). Employment and Entrepreneurship, Integration of Young People in the Labour Market: Estonia. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/estonia/36-integration-of-young-people-in-the-labour-market
80. Youth Wiki (2023d). Employment and Entrepreneurship, Integration of Young People in the Labour Market: Latvia. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/latvia/36-integration-of-young-people-in-the-labour-market
81. Youth Wiki (2023e). Employment and Entrepreneurship, Integration of Young People in the Labour Market: Lithuania. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/lithuania/36-integration-of-young-people-in-the-labour-market
82. Youth Wiki (2023f). Education and Training, Preventing early leaving from education and training (ELET): Hungary. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/hungary/63-preventing-early-leaving-from-education-and-training-elet
83. Youth Wiki (2023g). Education and Training, Preventing early leaving from education and training (ELET): Romania. Retrieved from https://national-policies.eacea.ec.europa.eu/youthwiki/chapters/romania/63-preventing-early-leaving-from-education-and-training-elet

About the authors

Tomislav Korotaj is a Teaching and Research Assistant at the Department of Statistics, Faculty of Economics and Business, University of Zagreb, Croatia. He holds a bachelor's degree in business economics and a master's degree in accounting and auditing from the University of Zagreb. He teaches the following courses: Statistics, Business Statistics, and Methods of Multivariate Analysis. He is pursuing a PhD in Economics, focusing on the youth labour market, wage policies, and income inequality. In addition to the above, his research interests include cluster analysis, panel data models, and school-to-work transition. The author can be contacted at **tkorotaj@efzg.hr**.

James Ming Chen is Professor of Law and Justin Smith Morrill Chair in Law at Michigan State University, U.S.A. He holds a degree in law, magna cum laude, from Harvard University and a master's degree in data science from Northwestern University. He is a member of the American Law Institute and a senior fellow of the Administrative Conference of the United States. He is licensed to practice law in Virginia and the District of Columbia. His research interests include law (particularly administrative law, competition law, and economic regulation), machine learning, mathematical and behavioural finance, and computational social science. The author can be contacted at **chenjame@law.msu.edu**.

Nataša Kurnoga is a Full Professor (Tenured) at the Department of Statistics, Faculty of Economics and Business, University of Zagreb, Croatia. She teaches the following courses at different study levels: Statistics, Business Statistics, Quantitative Methods, and Methods of Multivariate Analysis. She has participated in various scientific research projects and published numerous papers in scientific journals, conference proceedings, and books. She is a reviewer for numerous international scientific journals. Her research interests include business and economics statistics, especially multivariate data analysis methods. The author can be contacted at **nkurnoga@efzg.hr.**