ევროპის ცენტრალური
უნივერსიტეტი

ეკონომიკური პროფილი
ECONOMIC PROFILE

ტომი 20, #2 (30), 2025
VOLUME 20, ISSUE #2(30), 2025

ევროპის ცენტრალური უნივერსიტეტის რეფერირებადი და
რეცენზირებადი საერთაშორისო სამეცნიერო-პრაქტიკული
ჟურნალი

ჟურნალში გამოქვეყნებული სტატიის სიზუსტეზე
პასუხისმგებელია ავტორი

ამასთან, მისი პოზიცია შესაძლოა არ ემთხვეოდეს ჟურნალის სარედაქციო კოლეგიის
მოსაზრებებს

REFEREED AND PEER-REVIEWED
INTERNATIONAL SCIENTIFIC-PRACTICAL
JOURNAL OF
CENTRAL UNIVERSITY OF EUROPE

Each author is responsible for the accuracy of the article in
this journal. His/her position may not coincide with the
opinions of the editorial board

http://economicprofile.org/
ISSN 1512-3901
eISSN 2587-5310
DOI: https://10.52244/ep

2025

Machine Learning and DEA Application to Assess Railway Efficiency: Evidence from Uzbekistan

Komil Turaev
PhD Student, Silesian University of Technology, Poland. komil.turaev@polsl.pl

Aleksander Sladkowski
Professor, Silesian University of Technology, Poland. aleksander.sladkowski@polsl.pl

KEY WORDS: Railway efficiency; Data Envelopment Analysis; Clustering methods; Regional disparities; Uzbekistan
J.E.L. Classification: R42; O21; C800 DOI: https://doi.org/10.52244/ep.2025.30.07

For Citation: Turaev K., Sladkowski A., (2025) Machine Learning and DEA Application to Assess Railway Efficiency: Evidence from Uzbekistan. Economic Profile, Vol. 20, 2(30), p. 94-108. DOI: https://doi.org/10.52244/ep.2025.30.07

Annotation: *Railways are critical for economic integration and regional development, yet many emerging economies face uneven infrastructure performance and limited tools for monitoring efficiency. This study develops a replicable framework to measure the efficiency of Uzbekistan's railway sector using Data Envelopment Analysis and clustering methods. Three DEA models were constructed to capture complementary aspects of efficiency, linking population and investment inputs with outputs such as railway length, organizational activity, industrial production, and retail turnover. The results reveal strong regional disparities: Tashkent City and Karakalpakstan consistently perform above average, while Khorezm and Sirdarya lag behind. Efficiency trends further show that infrastructure availability does not always align with economic activity, underscoring mismatches in investment allocation. Clustering analysis grouped regions into three categories—high performers, emerging regions, and lagging areas—providing a practical framework for policymakers to design differentiated interventions. Beyond the case of Uzbekistan, the study demonstrates how open-source data and transparent Python-based DEA can be used to support transport policy in data-limited contexts.*

1. Introduction

Railway transport remains a cornerstone of sustainable mobility and economic integration, particularly in landlocked and emerging economies. Efficient rail systems reduce logistics costs, support industrial production, and enable regional development. In countries such as Uzbekistan, railways also hold strategic importance within Eurasian trade corridors, serving as a bridge between Asia and Europe. Despite this potential, regional disparities in infrastructure and service performance raise concerns about whether railway investments are being translated into balanced economic growth.

At the policy level, the Presidential Decree No. 329 (2023) [1] identified the railway sector as a national strategic priority, emphasizing modernization, efficiency, and the reduction of regional disparities. A major challenge highlighted in the decree is that investments are often allocated broadly at the regional level rather than specifically to rail infrastructure. This complicates the evaluation of how financial resources translate into tangible transport outcomes. Against this backdrop, rigorous methods are required to assess regional railway performance and provide actionable insights for policymakers.

Data Envelopment Analysis (DEA) has become one of the most widely used approaches for measuring transport efficiency, as it allows for benchmarking decision-making units (DMUs) without assuming a predefined functional form [2]. In the railway sector, DEA has been applied to passenger and freight performance [3], to incorporate service quality measures [4], and to examine broader drivers of efficiency across multiple contexts [5], [6]. However, most applications focus on advanced economies or concession-based freight operators, while less attention has been paid to emerging economies with constrained and aggregated datasets.

This study responds to these gaps by applying DEA to Uzbekistan's regional railway sector and extending the analysis with clustering. Using open-source socio-economic and infrastructure data, three DEA models are developed to reflect different perspectives: infrastructure accessibility, economic productivity, and investment-driven performance. Clustering is then applied to the DEA scores to group regions with similar efficiency profiles, allowing for clearer identification of strong performers, emerging regions, and lagging areas. While DEA has been extensively applied in the railway literature internationally [6], the integration of DEA with clustering remains underexplored in this domain. More importantly, no prior research has applied this combined framework to the Uzbek context. The contribution of this study is therefore twofold: (1) it demonstrates how an open-source DEA clustering approach can be used in a data-limited environment, and (2) it generates policy-relevant insights directly aligned with Uzbekistan's modernization priorities under Decree No. 329.

2. Literature review

DEA has been widely adopted in transport studies to evaluate efficiency across multiple dimensions. In the railway sector, early applications emphasized production efficiency, often linking inputs such as staff, rolling stock, or track length to outputs like passenger-kilometers or freight tonnage [3]. More recent studies have expanded the scope, integrating service quality indicators to capture user-oriented efficiency or focusing on high-speed rail and freight transport performance [4], [7]. A key finding across this literature is that efficiency outcomes are sensitive to variable selection and contextual conditions, underscoring the need for models tailored to local realities.

Railway capacity and infrastructure adequacy remain central concerns in efficiency studies. Khadem Sameni & Moradi (2022) [8] reviewed methods for capacity evaluation, showing that indicators such as operational track length and network density are critical for assessing accessibility and congestion. These insights justify including infrastructure-related outputs in DEA models, even when their correlation with socio-economic indicators is weak, since they capture structural constraints on system development.

Transport efficiency is increasingly analyzed in terms of resilience to disruptions and adaptability to uncertainty. A study proposed a fuzzy multi-layer DEA framework to evaluate international transport resilience under uncertain conditions, highlighting the importance of robustness in performance evaluation. While this study does not explicitly model resilience, it shares the motivation of identifying systemic vulnerabilities that undermine long-term efficiency [9].

To enhance interpretability, DEA results are often combined with second-stage methods such as regression or clustering. Marchetti & Wanke (2017) [10] applied DEA with clustering to Brazilian freight rail concessionaires, enabling regulators to design cluster-specific policy measures. Blagojević et al. (2020) [5] further demonstrated how hybrid DEA approaches can integrate fuzzy AHP and other tools to account for contextual factors. These studies highlight the added value of combining DEA with complementary methods, a practice followed in this research by applying K-Means clustering to DEA scores.

Despite extensive applications of DEA in transport, three gaps remain. First, limited attention has been paid to emerging economies with constrained data availability, where open-source methods could play a transformative role. Second, the use of DEA in Central Asia, particularly in the context of Uzbekistan's strategic role in Eurasian rail corridors, is almost absent in the literature. Third, few studies translate DEA results into cluster-based categories that can guide policymakers directly. Addressing these gaps, this study applies an output-oriented DEA and clustering framework to Uzbekistan's regional railway sector, offering methodological and practical contributions for both scholars and decision-makers.

3. Methodology

3.1. Data Source and Timeframe

The analysis is based on regional-level socio-economic and railway-related indicators for Uzbekistan, covering the period 2013–2022. The dataset includes 14 regions (12 provinces, the Republic of Karakalpakstan, and Tashkent City), observed over ten years, resulting in a balanced panel of 140 decision-making units (DMUs). Data was obtained from official national statistical sources, which provide open access to socio-economic and infrastructure indicators [11].

The variables capture demographic, economic, and infrastructure characteristics relevant to railway performance presented in Tab.1. Since detailed railway-specific investment data was not available, total regional investment was used as a proxy for development potential. This limitation reflects common challenges in applying DEA in emerging economies, where open-source data may be aggregated across sectors (Marchetti & Wanke, 2019) [3].

3.2 Variable Selection and Correlation Analysis

Since the dataset contains variables measured in different units (e.g., population, GDP, investment, railway length), normalization was applied to ensure comparability and avoid scale effects in the DEA analysis. The Min-Max Scaling method was used, transforming all variables into the $[0,1]$ range:

$$X_{scaled} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

where X is the original value, $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the variable, and X_{scaled} is the normalized value. This procedure preserves the relative differences between observations while placing all indicators on a common scale for correlation analysis and DEA modeling. Moreover, approach is consistent with best practices in DEA literature, where normalization is recommended to balance heterogeneous variables and avoid dominance by scale-intensive indicators [3].

	Variable	Description
1	Pop	Total population
2	Emp_Pop	Employed population
3	GDP	Gross Domestic Product
4	GDP_PerCap	GDP per capita
5	ORG	Number of operating organizations

6	Industry_Prod	Regional industrial output
7	Good_Prod_PerCap	Goods produced per capita
8	Retail_Turover	Retail turnover
9	Inv	Total capital investment
10	Lengths	Operational railway length

Table 1. Regional level socio-economic and railway-related indicators

3.3 Variable Selection and Correlation Analysis

To avoid redundancy and ensure meaningful model specification, Pearson correlation analysis was conducted (Fig. 1). The results confirmed strong positive correlations among several economic indicators (GDP, Industry_Prod, ORG, and Retail_Turover), while railway length showed weak or negative correlations with most socio-economic variables. This finding suggests that infrastructure distribution does not automatically follow population or economic growth, underscoring the need for DEA to benchmark efficiency.

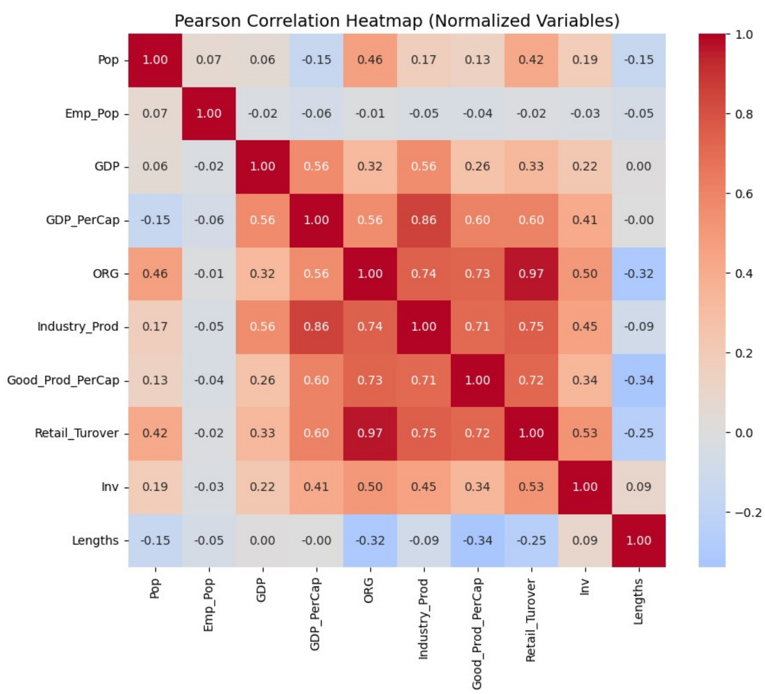

Figure 1. Peason Correlation Heatmap of Variables

These findings echo results from previous study who showed that excessive correlations among inputs and outputs can distort DEA outcomes by reducing discriminatory power. In line with their recommendations, highly collinear variables were not included together in the same model. Instead, the variables were grouped logically into complementary scenarios, capturing infrastructure capacity, economic productivity, and investment-driven growth.

Three DEA models were constructed: (1) evaluating economic efficiency in infrastructure development, (2) assessing demand-based investment efficiency, and (3) measuring the socio-economic impact of railway infrastructure. This structured approach ensures logical, data-driven variable selection aligned with the study's objectives.

3.4. DEA Model Construction

Data Envelopment Analysis (DEA) is a non-parametric linear programming technique that evaluates the relative efficiency of DMUs by comparing weighted outputs to weighted inputs. An output-oriented constant returns to scale (CRS) model was adopted, as the focus is on maximizing railway-related outputs given available resources. This follows the classical Charnes, Cooper, and Rhodes (CCR) formulation [2] .

Mathematical Formulation of the CCR Output-Oriented DEA Model:

In this study, efficiency is measured using the output-oriented DEA model. For each decision-making unit (DMU), the goal is to maximize the weighted outputs relative to the weighted inputs. The model can be written as [2]:

$$\max \theta_k = \sum_{r=1}^{s} u_r y_{rk} \quad (2)$$

subject to:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \quad \forall j - 1, \dots, N \quad (3)$$

$$\sum_{i=1}^{m} v_i x_{ik} = 1 \quad (4)$$

$$u_r, v_i \geq \epsilon \quad (5)$$

where:

- x_{ij} and y_{rj} are the observed inputs and outputs for each DMU j.
- u_r and v_i are the weights assigned to outputs and inputs, chosen by the optimization.
- The normalization condition ensures that inputs of the evaluated DMU are scaled to one.
- θ_k is the efficiency score for DMU k, where values closer to 1 indicate higher efficiency.

This means the model tries to find the best weights for each region so that its efficiency looks as high as possible, but without letting it exceed what other regions could achieve under the same weights. Scores close to 1 show that the region is efficient, while lower values mean there is room for improvement. The choice of an output-oriented model follows the logic used in previous railway DEA studies, which often prioritize output maximization given fixed infrastructure and demographic inputs [5] [4].

Model Design and Scenario Development:

To capture different perspectives of railway sector efficiency, three DEA model scenarios were developed Tab.2. Each model combines inputs and outputs in a way that reflects complementary aspects of regional development.

- Model A (Inputs: Pop, Inv; Outputs: Lengths, ORG) evaluates whether demographic and financial resources translate into greater railway accessibility and organizational capacity. Although investment (Inv) refers to total regional capital investment and not railway-specific spending, its inclusion serves as a proxy for infrastructure development potential. The weak correlations between Inv and railway length emphasize the usefulness of DEA in identifying regions that convert available resources into infrastructure and service coverage more efficiently.
- Model B (Inputs: Pop, Inv; Outputs: GDP, Industry_Prod) measures the economic productivity of railways, focusing on whether population and investment

generate stronger GDP and industrial output. GDP and industrial production are strongly correlated, ensuring that this model captures a coherent dimension of regional economic outcomes.

- Model C (Input: Inv; Outputs: ORG, Retail_Turover, Industry_Prod) focuses on investment-driven commercial and organizational activity. While Inv is again general regional investment, it is positively correlated with the chosen outputs, which themselves are strongly interrelated. This scenario therefore measures how efficiently investment is transformed into organizational growth and retail–industrial activity.

Model	Objective	Inputs	Outputs	Justification
Model A	Railway Accessibility and Infrastructure Capacity	Pop, Inv	Lengths, ORG	Evaluates whether population and investment result in greater railway infrastructure and service coverage.
Model B	Economic Productivity of Railway Infrastructure	Pop, Inv	GDP, Industry_Prod	Assesses how effectively economic resources and demand convert into regional economic outputs facilitated by railways.
Model C	Return on Investment via Economic and Commercial Activity	Inv	ORG, Retail_Turover, Industry_Prod	Measures how efficiently investment is transformed into measurable economic outcomes driven by railway services.

Table 2. Summary of DEA Models Rational

All models were implemented in Python using open-source libraries for data processing, statistical analysis, and optimization, including pandas (McKinney, 2010), NumPy (Harris et al., 2020), scikit-learn (Pedregosa et al., 2011), Seaborn (Waskom, 2021), Matplotlib (Hunter, 2007), PuLP (Mitchell & OSI, 2023) [12, 13, 14, 15, 16, 17].

3.5. Clustering Analyses

To complement the DEA results, clustering analysis was applied in order to group regions with similar efficiency profiles. The K-Means algorithm, a widely used unsupervised machine learning method, was chosen for this task. K-Means partitions the dataset into a predefined number of clusters by minimizing within-cluster variance and maximizing between-cluster differences [18].

The algorithm operates iteratively using two main steps:

Assignment step (distance calculation): Each region is assigned to the nearest cluster centroid based on Euclidean distance:

$$d(x,c)=\sqrt{(x_1-c_1)^2+(x_2-c_2)^2+\cdots+(x_n-c_n)^2} \quad (6)$$

where x is a data point (here, the DEA scores of a region), and c is a cluster centroid.

Centroid calculation: Cluster centroids are recalculated as the mean of all data points assigned to the cluster:

$$\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x \quad (7)$$

where C_i is the set of points in cluster i, and N_i is the number of points in that cluster.

This process is repeated until cluster assignments stabilize. In this study, $k = 3$ clusters were selected, balancing interpretability with the observed variability in DEA scores.

Clustering was performed using the efficiency scores from all three DEA models (A, B, C). This approach allows regions to be categorized into performance groups: low-efficiency, emerging, and high-efficiency clusters. Such classification helps identify regional typologies and supports the design of differentiated policy interventions for railway sector development.

The use of clustering as a second-stage analysis is consistent with Marchetti & Wanke (2017) [10], who applied DEA and clustering to Brazilian railways to identify operator groups, and with Blagojević et al. (2020) [5], who demonstrated the added value of hybrid approaches for transport efficiency studies. By applying clustering, Uzbek regions were grouped into high-efficiency, emerging, and low-efficiency clusters, enabling clearer policy implications than DEA scores alone.

4. Results

4.1 DEA Results for All Models

The DEA analysis produced efficiency scores for 14 regions over the period 2013–2022 across three models. Figure below presents efficiency trends averaged across regions for Models A, B, and C (Fig.2).

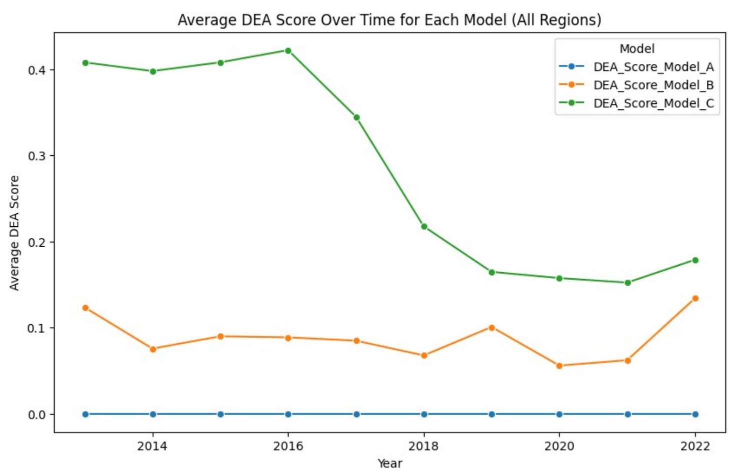

Figure 2 Average DEA Score Over Time for Each Model

The results reveal important differences across the models. Model A, which links population and investment to infrastructure outputs, consistently produced very low values close to zero. This outcome reflects the weak correlation between investment and railway length, confirming that general regional investment is not automatically directed toward railway infrastructure. Nevertheless, the few fluctuations observed highlight short-term spikes where infrastructure development projects aligned with organizational activity.

In contrast, Model B (economic productivity) shows more stable and moderate efficiency levels, with clear improvements in certain years, suggesting that regions with larger populations and investments tend to generate stronger economic outcomes through industrial production and GDP. Model C (investment-driven commercial outcomes) demonstrates the highest relative scores, with some regions reaching close to full efficiency. This indicates that, even though investment is not rail-specific, it translates more directly into organizational activity and retail turnover.

Overall, these patterns show that efficiency in Uzbekistan's railway sector is uneven: while infrastructure development lags relative to inputs, economic and organizational outputs respond more strongly to investment.

To capture spatial and temporal variation, Figure 3 presents a comparative heatmap of DEA efficiency scores for all three models across the 14 regions over the study period. The heatmaps make clear that efficiency levels are not evenly distributed: Tashkent and Karakalpakstan emerge as relatively stronger performers, while Khorezm and Sirdarya frequently appear in the low-efficiency range.

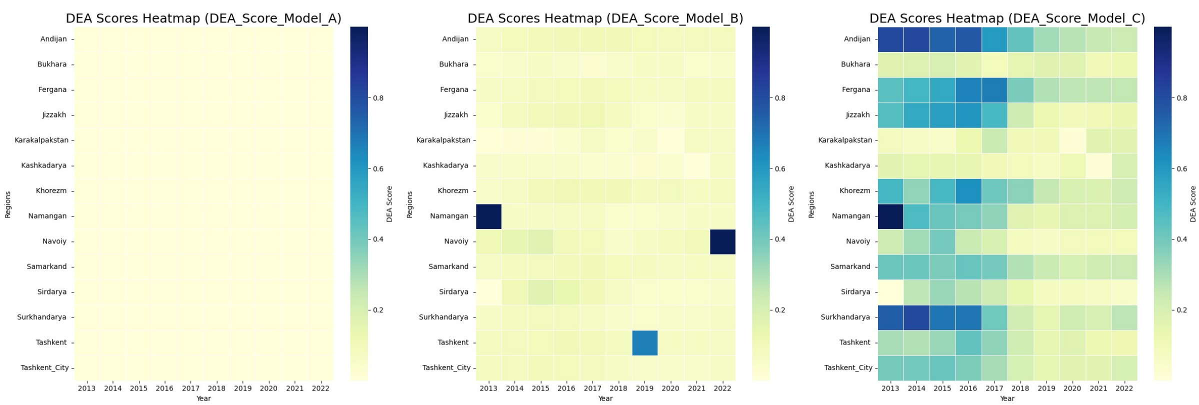

Figure 3. Heatmap of average efficiency scores by region and model

Model A's heatmap appears almost entirely yellow, indicating values very close to zero. This pattern reflects the fact that Model A's scores are on the order of 10^{-6}. The very small magnitude does not mean the model failed, rather, it shows that when population and investment are linked to railway length and organizational activity, efficiency values become highly compressed, highlighting only subtle differences between regions.

Model B's heatmap shows isolated peaks of efficiency in certain regions and years, suggesting temporary advantages in converting resources into GDP and industrial output. In contrast, Model C's heatmap reveals more variation across regions and a clearer temporal trend: efficiency was relatively higher in the early years but declined steadily after 2017 for most regions. These visual patterns emphasize the value of using multiple models, since each captures a different aspect of efficiency and produces different insights into regional performance.

To further analyze regional performance, Figures 4, 5 and 6 present regional DEA efficiency trends separately for Models A, B, and C. Model A (Fig. 4) shows unusual patterns due to its very small efficiency values (in the order of 1e–6). Despite this, relative differences are visible: Tashkent City and Karakalpakstan gradually rise, while Khorezm and Surkhandarya remain consistently low.

Model B (Fig. 5) highlights spikes in efficiency for a few regions such as Namangan and Tashkent, which likely reflect short-term economic gains or statistical outliers, while most other regions cluster at low values. Model C (Figure 5) shows clearer, smoother trajectories, with regions such as Andijan and Jizzakh initially performing well but declining after 2017, while Tashkent maintains a relatively higher

Taken together, these five figures provide complementary insights. The average model comparison reveals overall dynamics; the heatmaps highlight the spatial distribution of performance; and the regional trend plots trace individual trajectories over time. This efficiency compared to other regions.

combination of visuals ensures that the DEA results are not only statistically sound but also easily interpretable for policymakers, making efficiency disparities across Uzbekistan's railway sector more transparent.

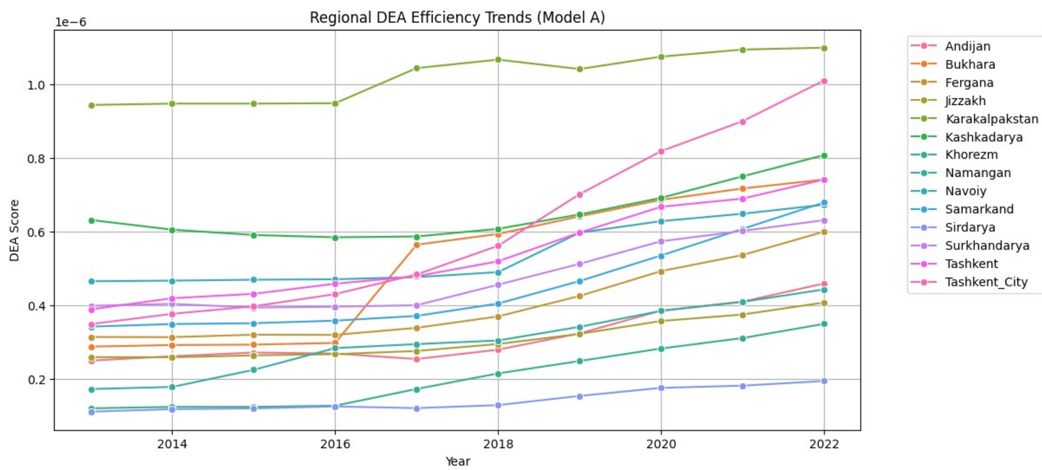

Figure 4. Model A – Regional Efficiency Trends

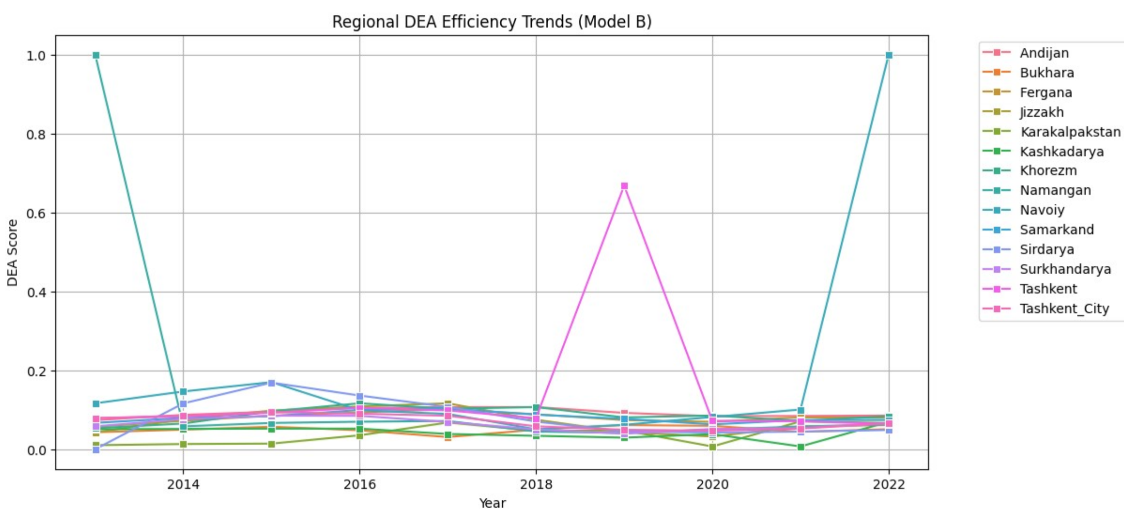

Figure 5. Model B – Regional Efficiency Trends

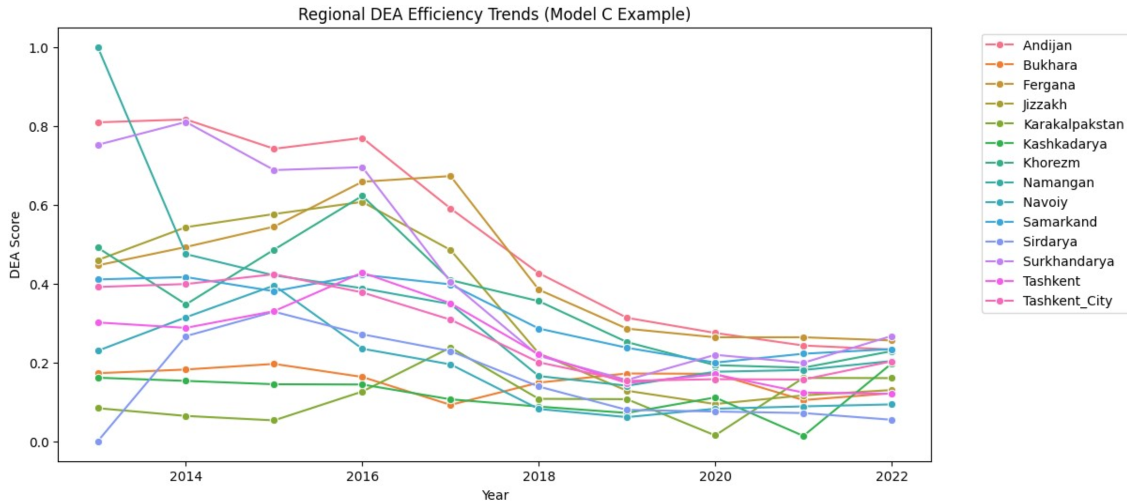

Figure 6. Model C – Regional Efficiency Trends

4.2 Clustering Results

To group regions by efficiency, a k-means clustering analysis was conducted. The elbow method (Fig. 7) suggested three optimal clusters. The 3D clustering plot (Fig. 8) illustrates how regions were grouped based on their DEA scores across Models A, B, and C.

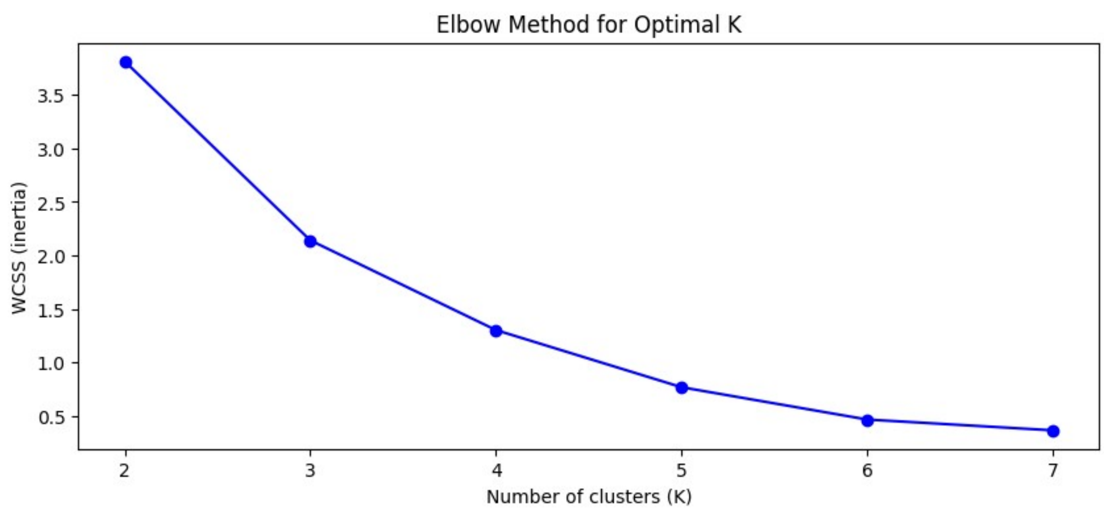

Figure 7. Elbow Method to Select Optimal K

Cluster 0 includes most regions with consistently low scores, showing structural weaknesses in efficiency. Cluster 1 captures regions with moderate performance, while Cluster 2 consists of very few regions with relatively high efficiency. In practice, this classification shows that Uzbekistan's railway efficiency is highly uneven, with only a small group of regions consistently performing well.

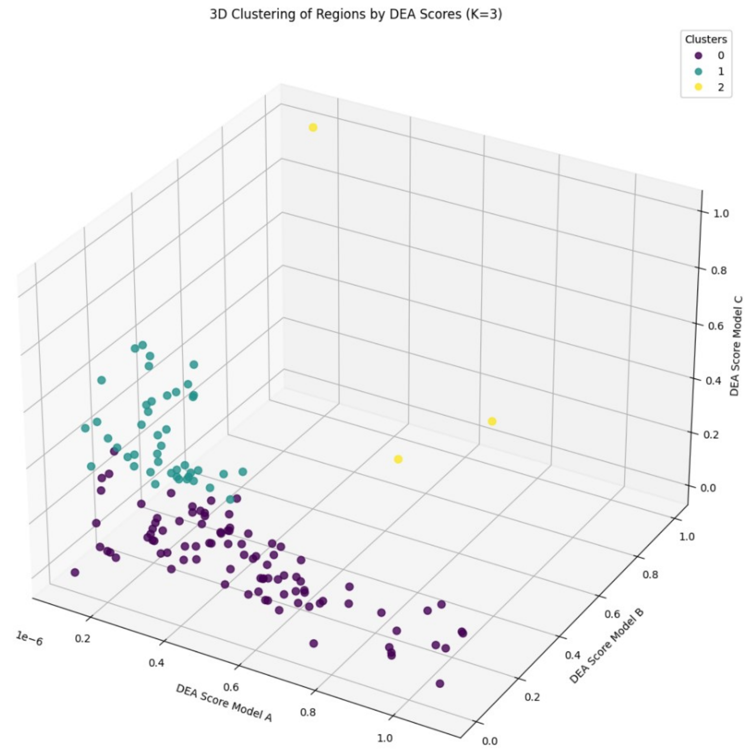

Figure 8. 3D Graph of Clustering Regions by DEA Models

5. Discussion and Recommendations

The findings show that model design strongly influences the interpretation of railway efficiency. Model A, based only on population and investment, produced extremely low and uniform scores. This indicates that general investment and population size are poor predictors of railway efficiency when used alone. Policymakers should therefore avoid relying solely on broad demographic or economic indicators when assessing transport performance.

Model B, using GDP and industrial production as outputs, offered some variation but still failed to capture consistent regional differences. This suggests that macroeconomic indicators, while relevant, cannot fully explain transport efficiency. They need to be complemented with sector-specific measures.

Model C provided the clearest picture, as it incorporated transport-related outputs such as organization of railway activity and retail turnover. This model revealed strong regional disparities: Tashkent, Fergana, and Andijan showed relatively high efficiency, while regions such as Karakalpakstan and Sirdarya consistently lagged behind. These results demonstrate that transport efficiency is closely tied to both infrastructure capacity and local economic activity.

Clustering analysis reinforced these patterns by showing that most regions belong to the low-efficiency group. Only a handful of regions achieved moderate or high efficiency. For policymakers, this highlights the urgent need to reduce the performance gap between core regions and the periphery.

5.1 Recommendations

The findings of this study show that the interpretation of railway efficiency is highly sensitive to model design. Model A, based on population and investment as inputs, produced extremely low and almost uniform scores. This suggests that broad demographic and aggregate investment indicators alone are insufficient to explain variations in railway efficiency. Policymakers should therefore avoid relying solely on general economic or population measures when evaluating transport performance.

Model B, which replaced infrastructure outputs with GDP and industrial production, introduced some variation but still failed to capture stable regional differences. While macroeconomic indicators are relevant for assessing regional potential, they cannot by themselves account for transport efficiency, as they are influenced by wider economic structures.

Model C, which included more sector-specific outputs such as organizational activity and retail turnover, provided the clearest picture of regional disparities. Regions such as Tashkent, Fergana, and Andijan emerged as relatively efficient, while Karakalpakstan and Sirdarya consistently underperformed. This indicates that transport efficiency is closely tied not only to infrastructure availability but also to the intensity of local economic activity and the extent to which railways are integrated into regional economies. Clustering analysis reinforced these findings by grouping most regions into a low-efficiency cluster, with only a handful in moderate or high categories.

Building on these results, several recommendations can be drawn. First, investment strategies should be more narrowly targeted toward railway infrastructure itself. Because the investment variable in this study captured total regional capital spending rather than rail-specific investment, the analysis revealed that general spending often failed to translate into efficiency gains. Directing funds into rail modernization, electrification, rolling stock upgrades, and intermodal hubs would likely yield clearer improvements.

Second, railway planning should be better aligned with regional economic development. As Model C highlighted, efficiency is strongly connected to industrial output and commercial activity. Policymakers should ensure that industrial zones, logistics hubs, and retail centers are strategically linked to the rail network, enabling multiplier effects across regional economies.

Third, reducing regional disparities should be prioritized. While Tashkent and Karakalpakstan have performed relatively well, peripheral regions such as Khorezm and Sirdarya remain persistently inefficient. Targeted interventions — from modernizing basic infrastructure to subsidizing passenger services or incentivizing freight operators — can help close this gap. Lessons from more efficient regions could be adapted and scaled to weaker areas.

Fourth, improving data collection and transparency is essential. The reliance on aggregated regional investment data underscores the need for more detailed transport-specific statistics. Establishing a national monitoring system for railway efficiency that integrates financial, operational, and service quality indicators would strengthen the basis for evidence-driven policy.

Finally, Uzbekistan should pursue international cooperation and modernization efforts. By adopting global best practices — such as digital traffic management, predictive maintenance, and integrated corridor planning — the railway sector can become more efficient and resilient. These recommendations align with national policy priorities, as emphasized in Presidential Decree No. 329 (2023), which calls for modernization and a reduction of regional disparities in railway development.

6. Conclusion

This study applied Data Envelopment Analysis and clustering to evaluate the efficiency of the railway sector across Uzbekistan's regions from 2013 to 2022. The results show that efficiency outcomes vary substantially depending on the choice of indicators: infrastructure-based efficiency remains very low, while models incorporating socio-economic activity reveal more pronounced but uneven regional performance. Clustering confirmed that most regions remain in the low-efficiency group, highlighting the structural challenges of the sector.

By using open-source data and free analytical tools, the study demonstrates a replicable framework that can be adopted for ongoing performance monitoring. The policy implications point to the importance of targeted investments, better integration of railways with regional economies, improved data systems, and international cooperation. Together, these measures can help ensure that railway development supports more balanced and inclusive regional growth in Uzbekistan.

References

1. President of the Republic of Uzbekistan. (2023, October 10). On measures for the fundamental reform of the railway transport sector of the Republic of Uzbekistan (Presidential Resolution No. RP-329). Official legal act. Retrieved from https://lex.uz/ru/docs/7560452
2. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8
3. Marchetti, D., & Wanke, P. F. (2019). Efficiency in rail transport: Evaluation of the main drivers through meta-analysis with resampling. *Transportationa Research Part A: Policy and Practice*, *120*, 83–100. https://doi.org/10.1016/j.tra.2018.12.005
4. Link, H. (2019). The impact of including service quality into efficiency analysis: The case of franchising regional rail passenger serves in Germany. *Transportation Research Part A: Policy and Practice*, *119*, 284–300. https://doi.org/10.1016/j.tra.2018.11.019
5. Blagojević, A., Vesković, S., Kasalica, S., Gojić, A., & Allamani, A. (2020). The application of the fuzzy AHP and dea for measuring the efficiency of freight transport railway undertakings. Operational Research in Engineering Sciences: Theory and Applications, 3(2), 1–23. https://doi.org/10.31181/oresta2003001b
6. Kapetanovic, M., Milenkovic, M., Bojovic, N., & Avramovic, Z. (2017). Evaluation of European railway companies efficiency: Application of a two-stage analysis. *Tehnika*, *72*(3), 403–410. https://doi.org/10.5937/tehnika1703403k
7. Zhou, Y., Wang, J., & Yang, H. (2019). Resilience of Transportation Systems: Concepts and Comprehensive Review. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4262–4276. https://doi.org/10.1109/TITS.2018.2883766
8. Khadem Sameni, M., & Moradi, A. (2022). Railway capacity: A review of analysis methods. In *Journal of Rail Transport Planning and Management* (Vol. 24). Elsevier B.V. https://doi.org/10.1016/j.jrtpm.2022.100357
9. Amoozad Mahdiraji, H., Yaftiyan, F., Abbasi-Kamardi, A., Vrontis, D., & Gong, Y. (2024). Disentangling the resiliency of international transportation systems under uncertainty by a novel multi-layer spherical fuzzy decision-making framework: Evidence from an emerging

economy. *Transportation Research Part A: Policy and Practice*, *186*. https://doi.org/10.1016/j.tra.2024.104151

10. Marchetti, D., & Wanke, P. (2017). Brazil's rail freight transport: Efficiency analysis using two-stage DEA and cluster-driven public policies. *Socio-Economic Planning Sciences*, *59*, 26–42. https://doi.org/10.1016/j.seps.2016.10.005
11. Agency on Statistics under the President of the Republic of Uzbekistan. (2023). Statistical Yearbook of the Republic of Uzbekistan 2013-2022. Tashkent: Agency on Statistics. Retrieved from https://stat.uz/en/publications/3627-statistical-yearbook-of-the-republic-of-uzbekistan
12. McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), Proceedings of the 9th Python in Science Conference (pp. 51–56). https://doi.org/10.25080/Majora-92bf1922-00a
13. Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2
14. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. http://scikit-learn.sourceforge.net.
15. Waskom, M. L. (2021). seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/joss.03021
16. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
17. Mitchell, S., O'Sullivan, M., & Dunning, I. (2011). PuLP: A linear programming toolkit for Python [Software manual]. COIN-OR. https://coin-or.github.io/pulp/
18. Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489
19. Sameni, M. K., & Preston, J. M. (2012). Value for railway capacity. *Transportation Research Record*, *2289*, 134–144. https://doi.org/10.3141/2289-18
20. Chenrrim, P., & Relvas, S. (2025). A data envelopment analysis approach for evaluating high-speed freight transportation efficiency. In *Transportation Research Interdisciplinary Perspectives* (Vol. 29). Elsevier Ltd. https://doi.org/10.1016/j.trip.2025.101338

მანქანური სწავლება და DEA აპლიკაციის გამოყენება რკინიგზის ეფექტურობის შესაფასებლად: უზბეკეთის შემთხვევა

კომილ ტურაევი
დოქტორანტი, სილეზიის ტექნოლოგიური უნივერსიტეტი, პოლონეთი, komil.turaev@polsl.pl

ალექსანდრე სკლადკოვსკი
სილეზიის ტექნოლოგიური უნივერსიტეტის პროფესორი
პოლონეთი, aleksander.sladkowski@polsl.pl

საკვანძო სიტყვები: რკინიგზის ეფექტურობა; მონაცემთა კონვერტაციის ანალიზი; კლასტერიზაციის მეთოდები; რეგიონული უთანასწორობა; უზბეკეთი
J.E.L. Classification: R42; O21; C800 DOI: https://doi.org/10.52244/ep.2025.30.07

ციტირებისთვის: Turaev K., Sladkowski A., (2025) Machine Learning and DEA Application to Assess Railway Efficiency: Evidence from Uzbekistan. Economic Profile, Vol. 20, 2(30), p. 94-108. DOI: https://doi.org/10.52244/ep.2025.30.07

ანოტაცია. *რკინიგზა კრიტიკულად მნიშვნელოვანია ეკონომიკური ინტეგრაციისა და რეგიონული განვითარებისთვის, თუმცა ბევრი განვითარებადი ეკონომიკისათვის პრობლემად შეიძლება იქცეს არათანაბარი ინფრასტრუქტურული მუშაობა და ეფექტურობის მონიტორინგის შეზღუდულ ინსტრუმენტები. წარმოდგენილი კვლევა ავითარებს განმეორებად ჩარჩოს უზბეკეთის რკინიგზის სექტორის ეფექტურობის გასაზომად მონაცემთა კონვერტაციის ანალიზისა და კლასტერიზაციის მეთოდების გამოყენებით. ეფექტურობის დამატებითი ასპექტების ასახვისთვის შეიქმნა სამი DEA მოდელი, რომლებიც აკავშირებდა მოსახლეობისა და ინვესტიციების ხარჯებს ისეთ გამომავალ რესურსებთან, როგორიცაა რკინიგზის სიგრძე, ორგანიზაციული აქტივობა, სამრეწველო წარმოება და საცალო ბრუნვა. შედეგები რეგიონულ დისბალანსს ავლენს: ტაშკენტი და ყარაყალპაკეთი მუდმივად საშუალოზე მაღალ მაჩვენებლებს აფიქსირებენ, ხოლო ხორეზმი და სირდარია ჩამორჩებიან. ეფექტურობის ტენდენციები ასევე აჩვენებს, რომ ინფრასტრუქტურის ხელმისაწვდომობა ყოველთვის არ შეესაბამება ეკონომიკურ აქტივობას, რაც ინვესტიციების განაწილების შეუსაბამობაზე მიუთითებს. კლასტერული ანალიზის შედეგად რეგიონები სამ კატეგორიად დაჯგუფდა - მაღალი მაჩვენებლების მქონე, განვითარებადი და ჩამორჩენილი რეგიონები - რაც პოლიტიკის შემქმნელებს დიფერენცირებული ინტერვენციების შემუშავების პრაქტიკულ ჩარჩოს უქმნის. უზბეკეთის შემთხვევის გარდა, კვლევა აჩვენებს, თუ როგორ შეიძლება ღია კოდის მონაცემებისა და გამჭვირვალე Python-ზე დაფუძნებული DEA-ს გამოყენება სატრანსპორტო პოლიტიკის მხარდასაჭერად მონაცემთა შეზღუდულ კონტექსტში.*