Received: 18 JAN 2026 | Accepted: 25 JUNE 2026 | Published Online: 30 JUNE 2026
Recoletos Multidisciplinary Research Journal Vol. 14 no. 1 (June 2026)
DOI: https://doi.org/10.32871/rmrj2614.01.05

Original Article

Uncovering Conditional Patterns in the THE Ranking Criteria: An Apriori Analysis of Global Universities

Rex Villavelez[1]*, Giovanni Pable[1]

[1] School of Education, University of San Jose-Recoletos, Cebu City, Philippines

*Correspondence:
rex.villavelez@usjr.edu.ph /
rex.villavelez@gmail.com

Abstract

Background: *Global university rankings have become an increasingly crucial metric for measuring the quality and performance of universities. This study analyzed the interrelationships among the criteria used in the Times Higher Education (THE) World University Rankings, identifying underlying patterns across the five criteria: teaching, research, citations, industry income, and international outlook, using data from 1,799 universities across 104 countries.*

Methods: *The study utilized association rule mining with WEKA's Apriori algorithm to identify non-random, co-occurring patterns among the ranking criteria, using preprocessed data from the publicly available THE 2023 ranking of institutions as high- or non-high-performing. Pearson correlation was also used to assess linear relationships among the criteria.*

Results: *Correlation analysis showed positive correlations among all the criteria. Interestingly, Apriori results identified an international outlook as a key factor associated with high citation performance. Strong bidirectional association rules indicate that internationally oriented universities are substantially more likely to achieve high research visibility. Additional rules reveal that a strong international outlook and high citation performance co-occur among institutions that do not necessarily have strong teaching ratings.*

Conclusion: *While teaching and research remain critical for institutional quality, strong international linkages are closely associated with greater global visibility and research impact amongst highly ranked universities.*

Keywords

apriori algorithm, Times Higher Education, university rankings, association rule mining, higher education metrics, UN SDG 4

INTRODUCTION

In recent years, global university rankings have become increasingly influential in shaping higher education policy, institutional planning, funding allocation, and international reputation. Governments, accreditation agencies, students, and university leaders frequently rely on ranking systems to evaluate institutional performance and competitiveness. These systems provide comparative measures of institutional quality based on indicators of teaching, research, internationalization, and innovation (Yüksel et al., 2023). One such metric is the Times Higher Education (THE) World University Rankings, widely regarded as a trustworthy benchmark for identifying top universities. Various indicators form the foundation for ranking universities

worldwide (THE, 2023). In today's tech-driven world, the surge of data has never been more significant, and the ability to analyze all types of data has become more crucial, if not essential. With the abundance of data from the THE Rankings, analyzing it can reveal new insights into their subtleties through data mining.

The growing availability of higher education datasets has also created opportunities to apply data analytics and educational data mining techniques to understand institutional performance better. While conventional statistical methods are useful for identifying linear relationships among ranking indicators, data mining approaches can uncover hidden patterns, conditional relationships, and co-occurrence structures that may not be evident through traditional analyses alone. Such approaches can provide deeper insights into how institutional characteristics interact within complex ranking systems.

Data mining involves applying various algorithms to large datasets to identify trends and patterns (Han, Pei, and Kamber, 2011). Common data mining techniques include clustering, classification, outlier detection, and association rule mining (Witten & Frank, 2005).

While many studies have already conducted data mining and analysis of university rankings, there is limited research that has explored the interrelationships among the indicators in the THE World University rankings using an association-rule-based data mining technique. Knowing how indicators co-occur can provide deeper insights into the characteristics associated with top-ranking universities.

The study may help determine how universities develop or adjust their programs and policies to better align with the THE World University Rankings. This will encourage higher education institutions to focus on delivering quality education and boosting their reputation within the expanding international academic community. Using association-rule-based data mining, this study examines co-occurrence patterns among specific university attributes. It assesses how the presence or absence of specific performance characteristics is associated with institutional performance in the THE World University Rankings.

METHODS

Study design, data source and study variables

This study employed a secondary data analysis approach using association rule mining and correlation analysis to examine relationships among indicators in the THE World University Rankings. The dataset is from the publicly available THE World University Rankings 2023, sourced from the THE database. The original THE World University Rankings 2023 dataset contained 2,345 universities. Before analysis, the dataset underwent a cleaning process to ensure completeness and consistency. Universities with missing values in one or more of the five ranking indicators (teaching, research, citations, industry income, and international outlook) were excluded using listwise deletion.

Listwise deletion was selected because the study focused on identifying associations among all five ranking criteria simultaneously. Retaining records with incomplete information or imputing missing values could introduce artificial patterns and potentially bias the association rule mining process. Duplicate records were checked, and none were identified. All indicator scores were verified to be within the expected THE scoring range, and no coding inconsistencies were identified.

The THE World University Rankings collects data from multiple sources, including institutional data submitted by universities themselves, bibliometric data from Elsevier's Scopus database, global academic reputation surveys, and measures of knowledge transfer, such as citations and international engagement. These data are then combined using a weighted methodology to generate a score per university across the five major performance criteria.

The study examined five institutional performance indicators used in the THE World University Rankings, namely:

- Teaching – Measures the learning environment through reputation surveys, staff-to-student ratios, doctoral education indicators, and student outcomes.
- Research – Assesses research productivity, research income, and research reputation.
- Citations – Measures research influence through citation impact and the extent to which published work contributes to global scholarship.

- Industry Income – Evaluates knowledge transfer by measuring research income generated from industry partnerships.
- International Outlook – Assesses internationalization through the proportion of international students, international faculty members, and international research collaborations.

Data Analysis

For association rule mining, the continuous indicator scores were converted to categorical variables because the Apriori algorithm operates on discrete transaction data. Each indicator was initially classified into six ordinal categories: A (90.00–100.00), B (80.00–89.99), C (70.00–79.99), D (60.00–69.99), E (50.00–59.99), and F (below 50.00). To facilitate the identification of meaningful patterns among high-performing institutions and reduce data sparsity, categories A and B are combined and coded as "High," whereas categories C, D, E, and F are combined and coded as "Not High."

This preprocessing step did not modify the original numerical values of the ranking indicators. The continuous scores were retained for descriptive statistics and Pearson correlation analysis, while the binary variables were used exclusively for Apriori association rule mining.

Association rule mining was performed using the Apriori algorithm implemented in WEKA data mining software. Apriori rule mining is used for its ability to identify frequent and meaningful co-occurrence patterns among categorical variables. In this study, association rules are interpreted as conditional patterns (if–then relationships) rather than as measures of correlation or causation.

To identify meaningful associations, a two-step evaluation procedure was applied. First, rules were generated using a minimum support threshold of 0.01, equivalent to approximately 1% of the analytical sample (18 universities), to account for the relative rarity of high-performing institutions within the dataset. A support threshold of 0.01 was selected to retain meaningful but relatively infrequent patterns while avoiding overly rare rules that may lack practical significance. Second, the resulting rules were evaluated using confidence and lift metrics. A minimum confidence threshold of 0.25 was used to retain nontrivial association rules, while only rules with lift values of at least 1.50 were considered for interpretation. To facilitate interpretation and emphasize the most meaningful patterns, the final output was limited to the three strongest association rules.

Association rule mining was conducted alongside descriptive statistics and Pearson correlation analyses of the original continuous variables. While correlation analysis was used to examine overall linear relationships among THE ranking criteria, the Apriori algorithm was employed to identify conditional and asymmetric relationships that may not be apparent through conventional statistical methods. This complementary approach provided a more comprehensive understanding of the relationships among institutional performance indicators.

Ethical Considerations

The study utilized publicly available institutional-level data obtained from the Times Higher Education World University Rankings database. No human participants were involved, no personally identifiable information was collected, and no intervention was conducted. In addition, formal ethical approval was not required. The study adhered to principles of responsible data use and academic integrity throughout the research process.

RESULTS

After removing 546 institutions (23.28% of the original dataset), the final dataset yielded an analytical sample of 1,799 universities from 104 countries. The results are presented in three stages. First, descriptive statistics are provided to summarize the distribution of the five THE ranking indicators. Second, Pearson correlation analysis is used to examine linear relationships among the indicators. Finally, association rule mining is employed to identify meaningful co-occurrence patterns among high-performing institutions.

Table 1 shows descriptive statistics (mean, standard deviation, minimum, and maximum) scores for the five THE ranking criteria in the 2023 dataset. The results reveal very high variability among institutions, as

shown by a very high standard deviation, especially in citation scores and international outlook. Teaching and research scores have more moderate variation, reflecting relatively consistent institutional capabilities in core academic functions. These descriptive statistics offer a baseline context for the following correlation and association rule analyses.

Table 1. *Descriptive Statistics of THE Score (n = 1799)*

	Mean score	Std. Deviation	Minimum	Maximum
Teaching	27.02	13.28	11.60	94.80
Research	23.02	16.76	7.40	99.70
Citations	48.50	27.97	0.80	100.00
Industry income	47.10	15.09	36.90	100.00
International Outlook	46.88	22.58	14.10	99.70

Table 2 shows the Pearson correlation coefficients between the five THE criteria. The results indicate that all Pearson correlation coefficients were positive and statistically significant at the 0.001 level. The correlation coefficients ranged from 0.183 to 0.882, indicating varying degrees of association among the THE ranking criteria. There is a very strong positive correlation between teaching and research (r = .882), suggesting that those institutions with stronger teaching scores tend to have stronger research scores as well. The results also indicate moderate correlations between research and citations (r = .514), research and international outlook (r = .538), and citations and international outlook (r = .536). Industry income has weaker but statistically significant correlations with the other criteria.

These results confirm expected linear relationships among the ranking indicators and align with previous studies on university performance. However, correlation analysis reflects overall trends across institutions and does not identify conditional or non-random co-occurrence patterns among high-performing universities.

Table 2. *Pearson Correlation Coefficients (r) among THE Ranking Criteria*

Criteria	1	2	3	4	5
1. Teaching	—				
2. Research	.882***	—			
3. Citations	.451***	.514***	—		
4. Industry income	.493***	.583***	.183***	—	
5. International outlook	.385***	.538***	.536***	.218***	—

Note. Values represent Pearson correlation coefficients (r).
*Positive values indicate positive associations between criteria. ***p < .001.*

Table 3 presents the association rules that identify meaningful co-occurrence patterns beyond overall linear relationships. Association rule mining was conducted using the Apriori algorithm on binary categories of high and non-high performance. The rules were evaluated using confidence and lift measures, focusing on those with a lift greater than 1.5 to identify stronger, more meaningful non-random co-occurrence patterns.

The strongest association rules reveal two main patterns: a bidirectional relationship between international outlook and citation impact. Universities with a high international outlook are more than twice as likely to demonstrate high citation performance compared to random expectation (confidence = 0.44, lift = 2.55). Similarly, highly cited universities also tend to exhibit a strong international outlook (confidence = 0.31, lift = 2.55). A third rule indicates that universities with a high international outlook and non-high teaching performance frequently co-occur with high citation performance, suggesting that international engagement is associated with greater research visibility.

The Apriori results reveal patterns among criteria that are not visible when looking only at correlations. Correlation analysis shows which indicators generally move together across universities, but it does not explain how specific combinations of strengths help some universities become more visible globally. The Apriori analysis fills this gap by showing which characteristics tend to appear together among universities with high citation performance. One clear finding is the strong link between international outlook and citations. Universities with a high international outlook are much more likely to achieve high citation impact than expected by chance. When universities collaborate with international partners, publish with global researchers, and participate in international research networks, their work is more likely to be read and cited by scholars worldwide.

Another key insight is that high citation scores do not always translate into high scores across other areas. Some universities achieve high citation impact even if they do not score highly under the teaching or research criteria. The findings suggest that international engagement is strongly associated with increased research visibility, even when other areas remain less strong. These findings also clarify why international outlook stands out in the Apriori results, even though it shows only moderate correlations with other criteria. As a result, international outlook may not appear dominant overall, but it becomes highly important in distinguishing globally visible universities from others. Lastly, the results show that, though teaching and research are the foundation of universities, they alone do not guarantee high research visibility. International connections extend the reach of research and attract greater attention. Universities that maintain strong global connections position themselves more effectively to translate their research into global impact, as reflected in higher citation performance.

Table 3. *Key Association Rules among THE Ranking Criteria*

Rule	If (Antecedent)	Then (Consequent)	Confidence	Lift
1	International Outlook = High	Citations = High	0.44	2.55
2	Citations = High	International Outlook = High	0.31	2.55
3	Teaching = Not High & International Outlook = High	Citations = High	0.40	2.54

Note. Antecedent (If) is the condition; consequent (Then) is the outcome. Confidence is the proportion of cases in which the consequent occurs given the antecedent, and lift is the strength of the association relative to random expectation.

DISCUSSION

Foundational Relationships Among Teaching and Research

The correlation analysis showed a strong positive relationship between teaching and research scores, indicating that institutions with strong teaching capabilities also have strong research capabilities. This finding aligns with the long-held theoretical and empirical perspectives that view teaching and research as closely intertwined academic functions that require the same set of institutional, academic, and organizational resources (Prosser & Trigwell, 1999; Sisavanh, 2003). In global rankings, this is to be expected because both factors contribute to the institution's reputation and academic quality (Delgado-Márquez et al., 2012; Hazelkorn, 2011). Nevertheless, these strong correlations indicate structural relationships among academic functions, but they do not explain why some institutions become more visible than others in the international research community. The correlation analysis primarily emphasizes the overall linear relationships among institutions but does not provide information on the patterns that distinguish high-performing institutions within the ranking system.

Additionally, the research criteria of THE rankings primarily capture research activity, income, and reputation, while citations capture research impact and the global spread of knowledge (Aksnes et al., 2019; THE, 2023). The present findings suggest that international engagement may contribute to greater research visibility, as increased international research collaboration and engagement with external partners can shape research activity and visibility (Azoulay et al., 2009; Etzkowitz & Zhou, 2017; Geuna & Nesta, 2006).

Existing research about university-industry collaboration suggests that engagement outside the academic environment can shape research productivity and innovation (Calcagnini et al., 2015; Park & Leydesdorff, 2010), albeit in a context-dependent manner (Banal-Estanol et al., 2008; Mark et al., 2014). Although industry income and research capacity are generally positively correlated (Bothwell & Staufenberg, 2016; Cordeiro et al., 2020), the present findings emphasize the critical role of the international outlook in distinguishing universities with high citation impact.

Bidirectional Associations between International Outlook and Citations as Key Non-Random Associations

The results of association rule mining offer further insights, revealing a strong bidirectional relationship between international outlook and citation performance. The two strongest Apriori rules show that universities with a high international outlook are much more likely to attain high citation impact, and that highly cited universities are also more likely to be internationally oriented. The high lift values associated with these rules suggest that this co-occurrence is significantly greater than expected by chance, emphasizing international engagement as a key factor in global research visibility.

This finding aligns with the existing literature, which highlights the importance of international collaboration, cross-border academic networks, and global research dissemination in shaping citation patterns (Azoulay et al., 2009; Bornmann & Leydesdorff, 2013). Citations reflect not just research productivity but also the visibility and utilization of research within the global academic community. Universities with stronger international outlooks are more likely to participate in collaborative research, publish in internationally recognized outlets, and engage with broader scholarly networks, thereby boosting their citation impact.

International Reputation, Citation Impact, and the Limits of Ranking Indicators

Findings from the Apriori analysis show that a high citation impact does not necessarily associate with a high score in teaching and research criteria. This suggests that some institutions with a strong international outlook score highly in citation impact, even though they do not score highly in the teaching criteria. This result highlights the distinction between research capacity and research impact in the THE methodology. These results also support the argument presented by Haddad-Adaimi et al. (2025) that certain world ranking metrics are unreliable predictors of good student learning outcomes and tend to lead institutions to devote more time, effort, and resources to improving their scores and image rather than their academic standards.

Results also show that internationally established and prominent universities, those with an internationally recognized reputation, are more likely to climb in the rankings. In contrast, universities considered good on the other criteria but not yet prominent internationally might find it extremely difficult to rise in the THE rankings (Bowman & Bastedo, 2011; Sorz et al., 2015).

Correlation and Association Rule Mining

Together, the correlation matrix and Apriori findings highlight the benefits of integrating traditional statistical methods with data mining approaches. Correlation analysis verifies the anticipated structural relationships among the criteria and provides foundational validation of the dataset (Saisana et al., 2005). Meanwhile, association rule mining reveals non-random, conditional patterns that linear correlations may miss (Frank et al., 2016; Han et al., 2006; Witten & Frank, 2005).

By revealing how specific performance states cluster among universities, especially at higher levels of achievement, the Apriori analysis provides deeper insights into the multidimensional nature of institutional excellence within global rankings. These findings support the idea that, although teaching and research form the academic core of universities, international engagement plays a crucial role in shaping the visibility and impact of research in the global higher education landscape (Marginson, 2022).

Overall, the findings suggest that, within the context of the ranking, teaching and research capacity form essential institutional foundations. However, an international outlook is strongly associated with greater research influence and global citation impact.

CONCLUSION

This study demonstrates the complexities of global university rankings, in which multiple areas of university performance interact to shape academic visibility and reputation. Understanding the interactions can greatly help universities to improve their competitiveness in an increasingly internationalized higher education environment. From a methodological standpoint, the study demonstrates that educational data mining techniques can complement traditional statistical analysis, uncovering additional findings that can be obtained only by data mining, such as co-occurring patterns that cannot be determined by correlation or even regression analysis. Beyond their use in university rankings, these approaches offer valuable opportunities to investigate complex relationships within large educational datasets. Future studies may extend this work by examining longitudinal ranking data, comparing multiple ranking systems, and exploring how institutional characteristics evolve. Such efforts can contribute to a more nuanced understanding of excellence, visibility, and performance in global higher education.

Author Contributions

R. Villavelez: Conceptualization, Methodology, Software, Data curation, Writing-Original draft preparation, Visualization, Investigation, Supervision, Validation; Writing-Reviewing and Editing; **G. Pable:** Writing-Original draft preparation, Writing-Reviewing and Editing, Conceptualization

Funding

This research received no external funding.

Ethical Approval

Not applicable since the study is data mining.

Competing interest

The authors declare no conflicts of interest.

Data Availability

Data will be made available by the corresponding author on request.

Declaration of Artificial Intelligence Use

In this work, the author(s) used artificial intelligence (AI) tools and methodologies, including Elicit AI to streamline literature research and ChatGPT to refine content. After using this tool/service, the author(s) evaluated and revised the content as necessary and take full responsibility for the published content.

REFERENCES

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. SAGE Open, 9(1). https://doi.org/10.1177/2158244019829575

Azoulay, P., Ding, W. and Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output*. The Journal of Industrial Economics, 57: 637-676. https://doi.org/10.1111/j.1467-6451.2009.00395.x

Banal-Estañol, A., Jofre-Bonet, M., & Lawson, C. (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. Research Policy, 44(6), 1160–1175. https://doi.org/10.1016/j.respol.2015.02.006

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. Journal of Informetrics, 7(2), 286–291. https://doi.org/10.1016/j.joi.2012.12.003

Bothwell, E. (2023, July 24). The world's top universities for attracting industry funding. Times Higher Education (THE). https://www.timeshighereducation.com/world-university-rankings/funding-for-innovation-ranking-2016

Bowman, N. A., & Bastedo, M. N. (2011). Anchoring effects in world university rankings: exploring biases in reputation scores. Higher Education, 61(4), 431–444. https://doi.org/10.1007/S10734-010-9339-1

Calcagnini, G., Favaretto, I., Giombini, G. et al. The role of universities in the location of innovative start-ups. Journal of Technology Transfer 41(4), 670–693 (2016). https://doi.org/10.1007/s10961-015-9396-9
Cordeiro, L. G., Lievore, C., & Pagani, R. N. (2020). Instituições de Ensino Superior no ranking THE: uma análise sobre ensino, pesquisa e renda industrial. Revista Stricto Sensu, 5(1). https://doi.org/10.24222/2525-3395.2020v5n1p043
Delgado-Márquez, B. L., Bondar, Y., & Delgado-Márquez, L. (2012). Higher education in a global context: Drivers of top-universities' reputation. Problems of Education in the 21st Century, 40, 17–26. https://www.proquest.com/scholarly-journals/higher-education-global-context-drivers-top/docview/2344208486/se-2
Etzkowitz, H., & Zhou, C. (2017). The Triple Helix: University–Industry–Government Innovation and Entrepreneurship (2nd ed.). Routledge. https://doi.org/10.4324/9781315620183
Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA workbench: Online appendix for "Data mining: Practical machine learning tools and techniques" (4th ed.). Morgan Kaufmann. https://ml.cms.waikato.ac.nz/weka/Witten_et_al_2016_appendix.pdf
Geuna, A., & Nesta, L. J. (2006). University patenting and its effects on academic research: The emerging European evidence. Research policy, 35(6), 790-807. https://doi.org/10.1016/j.respol.2006.04.005
Haddad-Adaimi, M., Daou, R. a. Z., & Ducq, Y. (2025). Assessing institutional reputation beyond quality rankings. Corporate Reputation Review, 28(2), 175–197. https://doi.org/10.1057/s41299-025-00219-4
Han, J., Pei, J., & Kamber, M. (2006). Data mining: Concepts and techniques (2nd ed.). Morgan Kaufmann Publishers.
Hazelkorn, E. (2011). Rankings and the reshaping of higher Education: The battle for world-class excellence. Tertiary Education and Management, 17, 373-375. https://doi.org/10.1080/13583883.2011.601753
Marginson, S. (2022). What is global higher education? Oxford Review of Education, 48(4), 492-517. https://doi.org/10.1080/03054985.2022.2061438
Mark, M., Jensen, R. L., & Norn, M. T. (2014). Estimating the economic effects of university-industry collaboration. International Journal of Technology Transfer and Commercialisation, 13(1-2), 80-106. https://doi.org/10.1504/IJTTC.2014.072687
Park, H. W., & Leydesdorff, L. (2010). Longitudinal trends in networks of university–industry–government relations in South Korea: The role of programmatic incentives. Research Policy, 39(5), 640–649. https://doi.org/10.1016/j.respol.2010.02.009
Prosser, M., & Trigwell, K. (1999). Understanding learning and teaching: The experience in higher education. Society for Research into Higher Education & Open University Press. https://eric.ed.gov/?id=ED434542
Saisana, M., Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x
Sisavanh, K. (2003) Educational research, policy and practice: National report on secondary education in Lao PDR. Educational Research for Policy and Practice 2, 3–11 (2003). https://doi.org/10.1023/A:1024417419791
Sorz, J., Wallner, B., Seidler, H., & Fieder, M. (2015). Inconsistent year-to-year fluctuations limit the conclusiveness of global higher education rankings for university management. 3. https://uscholar.univie.ac.at/view/o:448488
Times Higher Education. (2023). World University Rankings 2023. The Higher Education (THE). https://www.timeshighereducation.com/world-university-rankings/2023/world-ranking
Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques (2nd ed.). Morgan Kaufmann Publishers.
Yüksel, F. Ş., Kayadelen, A. N., & Antmen, F. (2023). A systematic literature review on multi-criteria decision making in higher education. International Journal of Assessment Tools in Education, 10(1), 12-28. https://doi.org/10.21449/ijate.1104005

How to cite this article:
Villavelez, R., & Pable, G. (2026). Uncovering conditional patterns in the THE ranking criteria: An Apriori analysis of global universities. *Recoletos Multidisciplinary Research Journal, 14*(1), 61–68. https://doi.org/10.32871/rmrj2614.01.05