Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

EduVision DV: Visual Analytics for University Rankings and Education Indicators

Vaibhav Kumawat

kumawatvaibhav22@gmail.com

Swami Vivekanand Group of Institutes

Research Article

Keywords: university rankings, data integration, visual analytics, data quality, Tableau, education indicators

Posted Date: October 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11335593/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

EduVision DV: Visual Analytics for University Rankings and Education Indicators

Vaibhav Kumawat
Department of Computer Science and Engineering, Swami Vivekanand College of Engineering, Indore, India (affiliated to Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal)
Corresponding author: kumawatvaibhav22@gmail.com | ORCID: 0009-0007-8259-2597

Abstract

University rankings from different providers publish finalised indicator values under different schemas, coverage and scoring conventions, which makes comparison across ranking systems, and against national education statistics, difficult. This paper presents EduVision DV, an integrated data pipeline and interactive visual analytics workbook that links three university-ranking datasets (QS World University Rankings 2025 [QS], Times Higher Education World University Rankings 2024 [THE], and a 2023 World University Rankings table) and World Bank EdStats indicators to a single master university dimension. Cleaning and identifier harmonisation are performed in Python, and the resulting four-table extract drives five coordinated Tableau dashboards with shared parameters and dashboard actions. A coverage audit shows that all 867 THE 2024 records and all 812 records of the 2023 table match the 1,503-university QS master, that 761 universities appear in both, 918 in at least one, and that 104 of the 106 QS countries have World Bank data. The audit also quantifies data-quality conditions hidden from a dashboard user: 60.1% of QS overall scores are estimated (at rank 601 and below), 82.6% of the 2023 overall scores are formula-reconstructed, and 721 of 867 THE 2024 records carry an identical constant overall score. In exploratory analysis of the 600 published QS scores, academic reputation has the highest rank correlation with the overall score (Spearman 0.83), and QS 2024 and 2025 ranks are highly stable (Spearman 0.97). Twenty functional tests passed. The contribution is a documented, reproducible integration workflow with explicit data-quality disclosure; no predictive model or user study is reported.

Keywords: university rankings, data integration, visual analytics, data quality, Tableau, education indicators

1. INTRODUCTION

Global university rankings are widely consulted by prospective students, academic researchers, university administrators and policymakers. Their usefulness is limited, however, by the way they are published. Moed [1], in a comparison of five ranking systems, argues that current systems remain one-dimensional: they provide finalised, seemingly unrelated indicator values rather than a dataset and tools with which patterns in multi-faceted data can be observed. Selten et al. [2] show that the three major rankings are stable over time but differ from each other, and that the variables used to construct them mainly reflect reputation and research performance. Pietrucha [3] shows that country-level factors are associated with ranking position, which suggests that institutional results are best read alongside national context.

A practitioner who wishes to combine several public ranking releases with national education statistics faces a practical problem before any analysis can begin. The sources use different column names, identifiers and coverage; institutions are named differently across releases; some scores are published while others are estimated or reconstructed; and national indicators cover a different period from the rankings. If these conditions are hidden inside a dashboard, aggregate figures can be misread.

This paper reports EduVision DV, a completed Major Project that addresses this practical problem with a documented integration pipeline and a Tableau workbook of five coordinated dashboards. All Tableau dashboards and worksheets were built by the author. The paper makes the following contributions:

1. a documented integration design that links QS 2025, THE 2024, a 2023 World University Rankings table and World Bank EdStats to a master university dimension, with quantified coverage and overlap,

2. an explicit audit of estimated, reconstructed, constant and imputed values in the integrated data, with an assessment of how they affect the aggregate indicators shown in the dashboards,
3. a four-view interactive dashboard design (plus a control page) in a widely available business-intelligence tool, with a functional test record, and
4. exploratory descriptive findings computed reproducibly from the packaged workbook.

The remainder of the paper is organised as follows. Sections 2 and 3 state the problem and objectives, Sections 4 and 5 review related work and identify the gap, Sections 6 to 9 describe the methodology, architecture, data and implementation, Sections 10 to 12 present results, discussion and limitations, and Sections 13 and 14 conclude and outline future work.

2. PROBLEM STATEMENT

Publicly available university-ranking datasets and national education statistics are published with different schemas, identifiers, coverage and score-completeness conventions. Without a systematic integration process and explicit data-quality flags, it is difficult to compare institutions across ranking systems, to relate institutional results to national education indicators, and to judge how far aggregate indicators (for example an average overall score) reflect published rather than estimated values. The problem addressed here is how to integrate such heterogeneous sources into a single, validated, interactive analytical environment while keeping their limitations visible.

3. OBJECTIVES

The main objective is to develop a unified Tableau workbook, backed by a repeatable data pipeline, that supports analysis of university rankings, research performance, student characteristics and country-level education indicators using publicly available datasets. The specific objectives are:

1. to profile, clean and integrate four public sources against a master university dimension and validate the resulting identifiers,
2. to engineer six key performance indicators (KPIs) in Python before loading them into Tableau,
3. to build University Overview, Research Analytics, Student Analytics and Country Comparison dashboards and a home dashboard,
4. to provide interactivity through filters, parameters and dashboard actions,
5. to test the workbook for functional correctness, and
6. to audit and disclose the coverage and data-quality characteristics of the integrated data.

4. LITERATURE REVIEW

The literature relevant to this work falls into four groups: comparative analyses of ranking systems, country-level determinants of ranking position, data integration for university indicators, and visual analysis and dashboard design.

4.1 Comparative analyses of ranking systems

Moed [1] links ARWU (the Academic Ranking of World Universities), the CWTS (Centre for Science and Technology Studies) Leiden Ranking, THE, QS and U-Multirank at the level of individual institutions and analyses the overlap in institutional and geographical coverage, how indicators are calculated from raw data, the skewness of indicator distributions and the correlations between indicators. Its central argument, that ranking systems provide finalised values rather than a dataset and tools for observing patterns, is the closest statement of the problem that EduVision DV addresses. Selten et al. [2] study ARWU, THE and QS over several editions using principal component analysis and exploratory factor analysis. They report that rankings are stable over time but differ between systems, that the constituent variables mainly capture two underlying factors (reputation and research performance), and that the variables may not capture the concepts they claim to measure.

4.2 Country-level context

Pietrucha [3] uses cross-sectional regression across countries to show that national economic potential, research and development spending, political stability and government effectiveness are associated with the position of universities in world

rankings. This motivates the inclusion of a country-level layer in EduVision DV, although the present work does not attempt regression or causal analysis.

4.3 Data integration

Daraio and Bonaccorsi [4] show, for European universities, that integrating existing data from heterogeneous sources into an open, linked-data platform supports the generation of new indicators. Their platform is considerably heavier than the file-based pipeline used here. Entity resolution, which underlies the matching of institution names across sources, is surveyed by Papadakis et al. [7], who review blocking and filtering techniques for reducing the comparisons needed to match records.

4.4 Visual analysis and dashboards

Gratzl et al. [5] introduce LineUp, a visual technique for exploring multi-attribute rankings that includes slope graphs for comparing rankings; a qualitative study reported that users solved complex ranking tasks quickly, and the technique was demonstrated on the QS 2012 dataset. Sarikaya et al. [6] review dashboard examples and literature to derive a design space and to classify dashboards by design goal and level of interaction. In the educational setting, Bodily and Verbert [8] review 93 articles on student-facing learning analytics systems and call for needs analyses, justified information selection and usability testing. That review concerns student-level rather than institution-level dashboards and is used here only to frame the evaluation limitations in Section 12.

Table 1 (next page) compares the reviewed studies.

Table 1. Comparison of the reviewed literature

Paper	Year	Methodology	Dataset	Key findings	Limitations	Relation to this project
Moed [1]	2017	Institution-level linkage of five ranking systems; overlap, indicator calculation, skewness and correlation analysis	ARWU 2015, Leiden 2016, QS and THE 2015-16, U-Multirank 2016	Systems give finalised indicator values, not a dataset and tools for observing patterns in multi-faceted data	Static statistical analysis; no interactive tool	Primary base paper. Same institution-level linkage and coverage audit; EduVision DV addresses the tooling gap it identifies
Selten et al. [2]	2020	Longitudinal comparison; PCA and exploratory factor analysis	ARWU, THE and QS, several annual editions	Rankings are stable over time but differ; variables mainly capture reputation and research performance	Statistical only; no interactive views or country indicators	Informs the indicator-association analysis in Section 10.3; PCA is not performed here
Pietrucha [3]	2018	Cross-sectional regression of ranking position on country-level factors	Ranking results and national indicators (42 to 71 countries, depending on ranking)	Economic potential, R&D spending, political stability and government effectiveness affect ranking position	Regression rather than exploratory tooling; country-level data only	Justifies the country-context layer (Country Comparison dashboard)
Daraio and Bonaccorsi [4]	2017	Linking heterogeneous data sources in an open linked-data platform	European universities, multiple existing sources	Integrated data supports generation of new indicators	Heavier platform than a file-based pipeline	Closest precedent for the integration objective
Gratzl et al. [5]	2013	LineUp visual technique for multi-attribute rankings; qualitative user study	QS 2012 dataset (demonstration)	Users solved complex ranking tasks quickly	Abstract does not mention country indicators or data-quality flags	Visual-analysis precedent; unlike [5], no user study was conducted here
Sarikaya et al. [6]	2019	Review of dashboard examples and literature; design space	Dashboard examples and literature	Dashboards can be classified by design goal and level of interaction	General, not education-specific	Frames how the five dashboards are classified
Papadakis et al. [7]	2020	Survey of blocking and filtering techniques for entity resolution	Survey; no single dataset	Reviews techniques for scalable record matching	Not specific to universities	Contextualises name-based institution matching
Bodily and Verbert [8]	2017	Review of 93 articles on student-facing learning analytics systems	Literature	Calls for needs analysis, justified information selection and usability testing	Student-level, not institution-level	Supports the evaluation-rigour argument in Section 12

5. RESEARCH GAP

Within the reviewed set, Moed [1] argues that ranking systems lack tools for exploring multi-faceted data, and Selten et al. [2] show that rankings differ and that their variables may not measure what they claim. Pietrucha [3] shows that country factors matter, and Daraio and Bonaccorsi [4] show what linking heterogeneous data can offer. Gratzl et al. [5] offer a visual technique for rankings, and Papadakis et al. [7] survey entity resolution. Among these works we did not find a documented, reproducible workflow that combines (i) institution-level indicators from several ranking releases linked to country-level education indicators, (ii) explicit flags for estimated, reconstructed and imputed values, and (iii) interactive coordinated views in a widely available business-intelligence tool.

This is a gap within the reviewed set and not a claim about the whole literature.

6. PROPOSED METHODOLOGY

The methodology follows the pipeline documented in the project files. The datasets are Kaggle-hosted copies of public ranking and education statistics [9]–[11] (see References). The exact Kaggle listing for the THE 2024 dataset could not be located and is therefore not cited as a reference (see Section 12).

1. **Collection and profiling.** The four source tables were collected with a Python script and profiled for completeness and structure.
2. **Cleaning and standardisation.** Duplicate institution and country records were removed; university and country names were standardised against a single reference list; ranking and score fields were normalised to comparable numeric ranges. Ambiguous name matches were resolved using country, region and fuzzy name matching, with manual review of low-confidence matches. One cross-country false match and a character-encoding corruption (accented letters) were identified and corrected.
3. **Integration to a master dimension.** The QS 2025 table (1,503 universities) was used as the master university dimension. The THE 2024 table, the 2023 table and the World Bank EdStats table were synchronised to it, so that every retained record carries a valid master university or country identifier. The World Bank table was reduced from 55,918 raw observations to 33,450 by retaining only countries present in the master.
4. **Research-metric cascade.** For research and student measures, THE 2024 values were used as the primary source and the 2023 table as the fallback.
5. **KPI engineering.** Six KPIs (Table 5) were computed in Python with pandas and NumPy and written to a Tableau-ready file, keeping the KPI logic outside Tableau and under version control.
6. **Dashboard construction and testing.** A packaged workbook with an embedded extract was built in Tableau and tested with a written checklist and a 20-case interaction test plan.

Handling of estimated scores. Some sources do not publish an overall score for every institution. For these institutions the overall score was estimated by applying the provider's official indicator weightings; Section 10.2 independently tests this against the official published weightings in Table 2. In the integrated data, each estimated or reconstructed value is identifiable through a flag (overall_score_is_estimated, overall_score_source, female_pct_imputed), which is what makes the audit in Section 10 possible.

Table 2. Indicator weightings used to validate overall scores

Ranking table	Indicator (weight, %)
QS 2025 [12]	Academic reputation (30), employer reputation (15), faculty-student (10), citations per faculty (20), international faculty (5), international students (5), international research network (5), employment outcomes (5), sustainability (5)
2023 table [13]	Teaching (30), research (30), citations (30), international outlook (7.5), industry income (2.5) - pre-2024 (pre-WUR 3.0) methodology
THE 2024 table [14]	Teaching (29.5), research environment (29), research quality (30), international outlook (7.5), industry (4) - WUR 3.0 methodology, confirmed against exact-rank rows (Table 7)

THE revised its weightings for the 2024 edition (WUR 3.0 methodology). Table 2 therefore lists the pre-2024 weights that apply to the 2023 table and the WUR 3.0 weights that apply to the THE 2024 table separately; both were confirmed against the official methodology pages, and the THE 2024 validation in Section 10.2 and Table 7 was recomputed using the correct WUR 3.0 weights.

7. SYSTEM ARCHITECTURE

Figure 1 shows the architecture. Four public sources feed a cleaning and standardisation stage, an integration stage that synchronises all tables to the master university dimension, and a KPI-engineering stage. The output is a combined Excel dataset and a Tableau extract with four logical tables, which drive five dashboards.

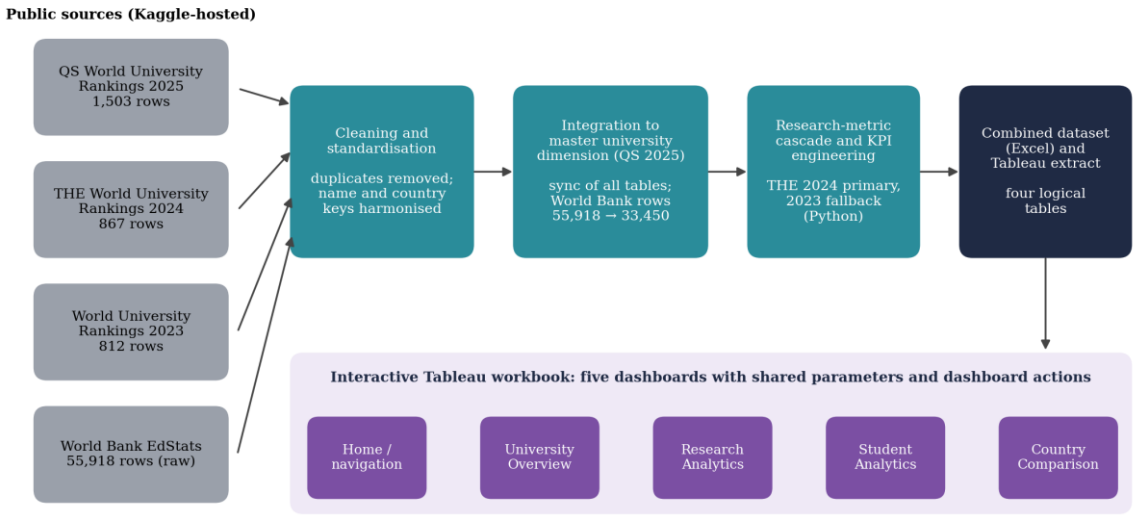


Fig. 1 EduVision DV data pipeline and dashboard architecture. Row counts are taken from the packaged workbook; the World Bank table is reduced from 55,918 raw to 33,450 rows after synchronisation to the master

The data model has a hub-and-spoke structure. The University Overview table (one row per QS university) is the hub. The Research Analytics (THE) and Student Analytics (2023 table) tables relate to the hub many-to-many, matched on university_id and cross-checked on country_id; the Country Comparison table is a long-format table (country, year, indicator, value) that relates to the hub through country_id and has no institution-level rank field. Table 3 summarises the dashboards.

Table 3. Dashboards in the workbook

Dashboard	Purpose	Grain
EduVision DV (home)	Landing page with the Top N, Select Education Indicator and DASHBOARDS parameter controls and a region selector	Control layer
University Overview	Global ranking, KPI cards, regional distribution, academic-reputation analysis, ranking tables for 2024 and 2025	One row per university (QS)
Research Analytics	Teaching, research, citations and industry-income scores; citation-versus-research scatter plot by region	One row per university (THE)
Student Analytics	Enrolment, international student share, student-staff ratio, gender split	One row per university (2023 table)
Country Comparison	Country-level education indicators over time, ranked by a selected indicator	One row per country, year and indicator

8. DATASET AND DATA SOURCES

Table 4 lists the four sources with the record counts observed in the packaged workbook. All are secondary copies obtained from Kaggle. The provider of the 2023 table's underlying ranking data is not stated on its Kaggle page (uploader “alitaqi000”, dataset titled “World University Rankings 2023,” see Reference [10]). The dataset page describes 1,799 universities across 104 countries, more than the 812-university subset embedded in this workbook.

Table 4. Data sources and record counts (packaged workbook)

Source	Records	Scope	Used by
QS World University Rankings 2025 [9]	1,503 universities	106 countries; regions: Asia 502, Europe 497, Americas 418, Oceania 46, Africa 40	University Overview
THE World University Rankings 2024 (source listing not located; see Section 12)	867 universities	88 countries; teaching, research, citations, industry income, international outlook. Rank type: 146 exact, 682 banded, 39 unranked reporters	Research Analytics
World University Rankings 2023 [10]	812 universities	Student numbers, student-staff ratio, international student share, gender split	Student Analytics
World Bank EdStats [11]	33,450 observations (55,918 raw)	104 countries, 7 World Bank regions, 5 income groups, 13 indicators, 1970-2016	Country Comparison

The Final Documentation (Sections 1.1 and 3) reports 868 research records, 842 student records and 216 countries; these are the pre-cleaning, raw counts. Rows with missing or null values in key fields were removed during cleaning, leaving the 867, 812 and 104 figures used throughout this paper, which are the figures that appear in the packaged workbook. The two sets of figures therefore describe different pipeline stages rather than conflicting with each other.

9. IMPLEMENTATION

The workbook was built in Tableau Desktop (file saved by build 2026.2.2) and saved as a packaged workbook (.twbx) with an embedded extract [15]. It contains 47 worksheets arranged on five final dashboards; six further dashboard tabs (draft copies, an INFO page and upper-case duplicates) were deliberately retained for traceability and are documented as housekeeping items. All dashboards share a single dark-theme formatting standard applied at workbook level.

Controls. Three parameters are exposed on the home page: Top N (default 15), which limits the ranking tables; Select Education Indicator, which selects one of three World Bank indicators for the Country Comparison dashboard (tertiary gross enrolment ratio, government expenditure on education as a percentage of GDP, and adult literacy rate); and DASHBOARDS, which labels the intended destination dashboard. Fifteen dashboard actions provide filtering and highlighting, for example selecting a university row to filter related charts, selecting a region in a legend to highlight it across a dashboard, and selecting a country to filter university-level views where a mapping exists.

KPIs. Table 5 lists the six KPIs as documented in the project, taken from the dataset's own documented source columns. They were computed in Python (generate_education_kpis.py) rather than as Tableau calculated fields.

Table 5. Documented KPIs

KPI	Documented definition	Source fields	Formula
Global Ranking Score	QS 2025 overall position and score, shown as rank and score out of 100 (direct pass-through of the QS Overall Score; not blended with THE)	rank_2024_numeric, rank_2025_numeric, overall_score	Direct pass-through of the QS 2025 Overall Score (no blending with THE).
Research Impact Score	Citation strength, from THE 2024 (fallback: 2023 table)	scores_citations (THE 2024, primary); citations_score (2023 table, fallback)	Direct pass-through of the Citations Score: THE 2024 for 867 universities, 2023-table fallback for 51 universities not covered by THE 2024 (918 total).
Faculty-to-Student Ratio	Students per faculty member, shown as a ratio such as 1 to 17.3	student_staff_ratio, total_students	Direct pass-through of the Student/Staff Ratio: THE 2024 for 867 universities, 2023-table fallback for 51 universities (918 total), displayed as a ratio, e.g. 1 to 17.3.

KPI	Documented definition	Source fields	Formula
International Student Percentage	Share of enrolled students who are international	international_students_percentage, total_students	Direct pass-through of International Students %: THE 2024 for 867 universities, 2023-table fallback for 51 universities (918 total).
Academic Reputation Score	Survey-based peer regard, an input to the QS methodology	academic_reputation	Direct pass-through of the QS 2025 Academic Reputation Score (not blended with Employer Reputation).
Research Productivity Index	Weighted blend of research volume, citation strength and international outlook (50/30/20)	scores_research, scores_citations, scores_international_outlook	Derived: 0.50 × Research Score + 0.30 × Citations Score + 0.20 × International Outlook Score, from THE 2024 for 867 universities, 2023-table fallback for 51 universities (918 total).

The six formulas above are taken from the dataset's own documented source columns. The university counts they cite (867 from THE 2024, 51 falling back to the 2023 table, 918 in total) were cross-checked against the packaged workbook and match exactly; within that same 918-university subset, the overall_score_is_estimated flag splits 486 True to 432 False, which also matches the workbook. One implementation note: although the Research Productivity Index is documented as a 50/30/20 blend of Research, Citations and International Outlook scores, the corresponding dashboard worksheet (Research Productivity Trends) plots the Citations Score and Research Score components separately rather than the blended index itself, so a dashboard viewer does not see the index directly. The KPI formulas are reported as documented in the project files and were not independently re-derived from the pipeline scripts for this paper.

10. RESULTS AND ANALYSIS

Results are labelled by type: **observed** (recorded in the project documents), **derived** (computed by us from the data embedded in the packaged workbook), and **interpretation**. All derived figures were recomputed from the workbook's Hyper extract by a single script (reproduce_results.py, provided alongside this paper;), which also produced Figures 2 to 4. No expected result is presented as an achieved result.

10.1 Coverage and overlap (derived)

Every THE 2024 and 2023-table record matches a QS master identifier (867 of 867 and 812 of 812). Table 6 and Figure 2 show the overlap. Of the 1,503 master universities, 761 (50.6%) appear in both supplementary tables, 918 (61.1%) in at least one, and 585 (38.9%) in neither; 106 appear only in THE 2024 and 51 only in the 2023 table. Of the 106 QS countries, 104 have World Bank rows; Northern Cyprus and Taiwan do not. Matching in the final extract is recorded as exact key with country for all 867 THE records, and as exact key with country (710) or unique exact key (102) for the 2023 table.

Table 6. Coverage and overlap relative to the QS master (n = 1,503)

Subset	Universities	Share of master
THE 2024 table	867	57.7%
2023 table	812	54.0%
In both THE 2024 and the 2023 table	761	50.6%
In at least one (research-metric set)	918	61.1%
In neither	585	38.9%

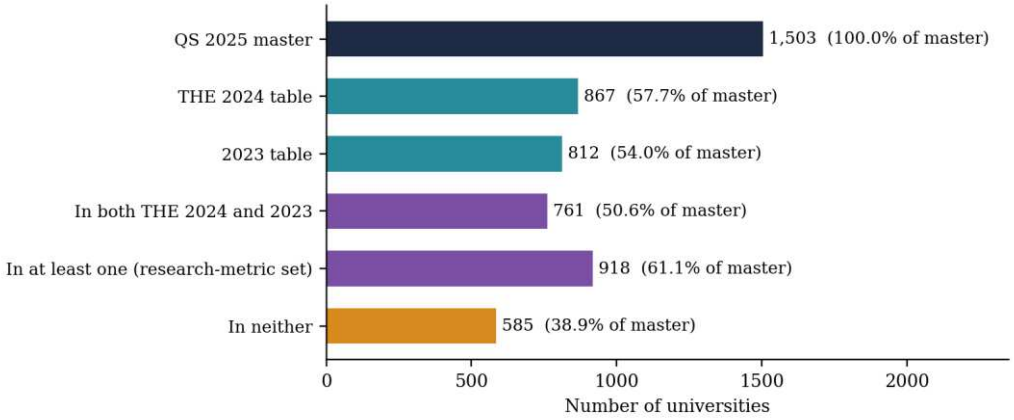


Fig. 2 Coverage and overlap of the supplementary tables relative to the QS 2025 master (derived)

Interpretation. Because 38.9% of master universities lack both research and student measures, the Research Analytics and Student Analytics dashboards describe a subset of the universities shown in University Overview. Users comparing counts across dashboards should expect different denominators.

10.2 Data-quality composition and validation of estimated scores (derived)

Figure 3 shows the composition of the overall-score fields. In the QS table, 903 of 1,503 overall scores (60.1%) are flagged as estimated, all at rank 601 and below (the highest published rank is 596 and the lowest estimated rank is 601). In the 2023 table, 671 of 812 overall scores (82.6%) are formula-reconstructed and 65 gender values (8.0%) are imputed. In the THE 2024 table, 721 of 867 rows (83.2%), namely the banded and unranked rows, carry the identical overall score of 66.45. Band-valued ranks are stored as the lower bound of the band (100% agreement between the band and the numeric rank field), so numeric ranks for banded institutions are approximate.

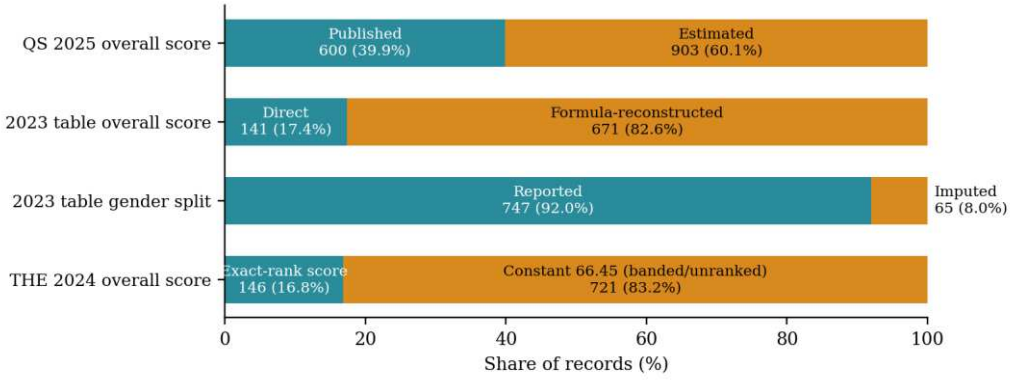


Fig. 3 Composition of overall-score and gender fields by provenance (derived)

To assess the estimated and reconstructed scores, each overall score was recomputed from its component indicators using the weightings in Table 2 and compared with the stored value (Table 7). For the 2023 table the recomputed and stored scores agree closely for both direct (mean absolute error 0.029) and formula-reconstructed rows (0.025), so the reconstruction is consistent with the published weighting. For QS, the recomputed and stored values are almost perfectly correlated for both groups, but the stored estimated scores lie a small, nearly constant amount below the weighted sum of the component scores (mean difference 1.39, range 1.06 to 1.68), and published scores also differ from the weighted sum (mean absolute error 0.68, range -0.94 to 3.89). For the 146 THE 2024 rows with an exact rank, recomputing with the correct WUR 3.0 weights (Table 2) gives a mean absolute error of 0.03, matching the precision obtained for the 2023 table and confirming that the estimation procedure holds once the correct per-edition weights are used. For the 721 banded or unranked rows the recomputed mean is 37.9, far from the constant

66.45, so the constant is not a usable overall score for those rows.

Table 7. Recomputation of overall scores from component indicators (derived)

Table and group	n	Agreement with stored overall score
QS 2025, published	600	Pearson r = 0.99993; MAE 0.68; mean difference 0.57 (min -0.94, max 3.89)
QS 2025, estimated	903	Pearson r = 0.99998; MAE 1.39; mean difference 1.39 (min 1.06, max 1.68)
2023 table, direct	141	MAE 0.029; maximum absolute difference 0.078
2023 table, formula-reconstructed	671	MAE 0.025; maximum absolute difference 0.050
THE 2024, exact rank	146	MAE 0.03 (WUR 3.0 weights)
THE 2024, banded or unranked	721	Recomputed mean 37.9 versus stored constant 66.45

The estimated scores are also much lower than the published ones: the mean QS overall score is 41.84 for the 600 published scores, 10.87 for the 903 estimated scores and 23.24 across all 1,503 rows. The pooled value of 23.24 is the figure shown by the Average Overall Score KPI card. *Interpretation*. This pooled average describes neither group and depends mainly on the share of estimated scores, so it should be reported by provenance or restricted to published scores.

10.3 Descriptive findings (derived)

Rank stability. The Spearman correlation between QS 2024 and 2025 numeric ranks is 0.969; 477 universities improved, 611 declined and 415 were unchanged. Because banded ranks are stored as lower bounds, the counts of movement are coarse. This is consistent with the stability reported in the multi-edition study of Selten et al. [2], although the two analyses differ in scope and method.

Regional and country composition. The QS top 100 comprises 34 European, 33 American, 23 Asian and 10 Oceanian universities and none from Africa. By country, the United States has 25, the United Kingdom 15, Australia 9, and Mainland China and Hong Kong SAR 5 each.

Indicator association. Among the 600 published QS scores, Spearman correlation with the overall score is highest for academic reputation (0.83) and employer reputation (0.70) and lowest for the faculty-student score (0.28); see Figure 4. Among THE 2024 pillar scores, teaching and research are strongly correlated (Pearson r = 0.91), followed by research and industry income (0.76). In the 2023 table, the international student percentage has a Spearman correlation of 0.62 with the overall score, whereas the student-staff ratio (-0.05) and total enrolment (0.11) show little association.

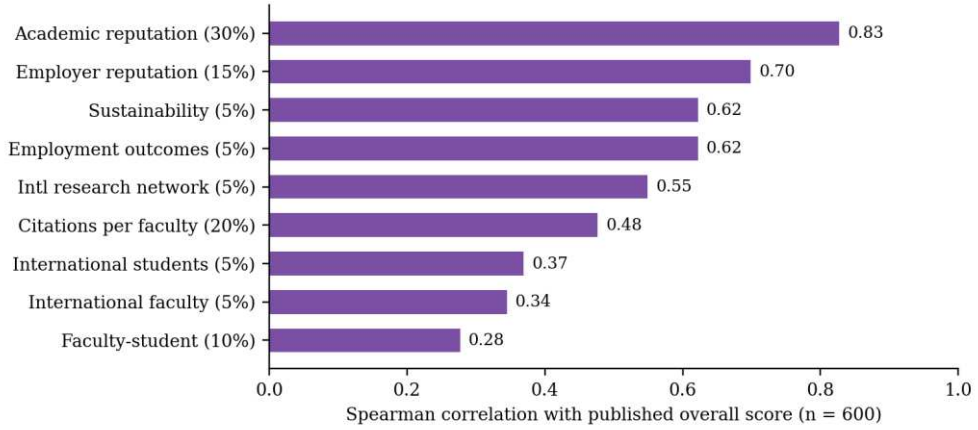


Fig. 4 Spearman correlation of QS indicator scores with the published overall score (n = 600; derived)

Interpretation. These coefficients describe the structure of the composite scores, in which reputation indicators carry the largest weights. They do not indicate causal effects, and the

correlations of components with an aggregate that includes them are partly definitional.

10.4 Functional testing (observed, documented)

The Dashboard Testing Report records 20 test cases across the five dashboards covering filters, parameters, dashboard actions, cross-chart highlighting and layout at two screen resolutions (1366 x 768 and 1920 x 1080). All 20 passed. One defect was found and resolved: on the home dashboard, floating chart zones from the analytical dashboards rendered simultaneously and overlapped, because no action controlled their visibility; the layout was rebuilt so that the home page holds only the header and controls, and navigation uses the dashboard tab strip. Sampled KPI checks showed no discrepancies. A separate QA checklist covers KPI accuracy, ranking calculations, interactions, educational-metric ranges and formatting. These results are reported as documented; the tests were not re-executed for this paper, and they are functional tests performed by the project team rather than an evaluation with users. The five dashboards are shown in Figs. 5–9.

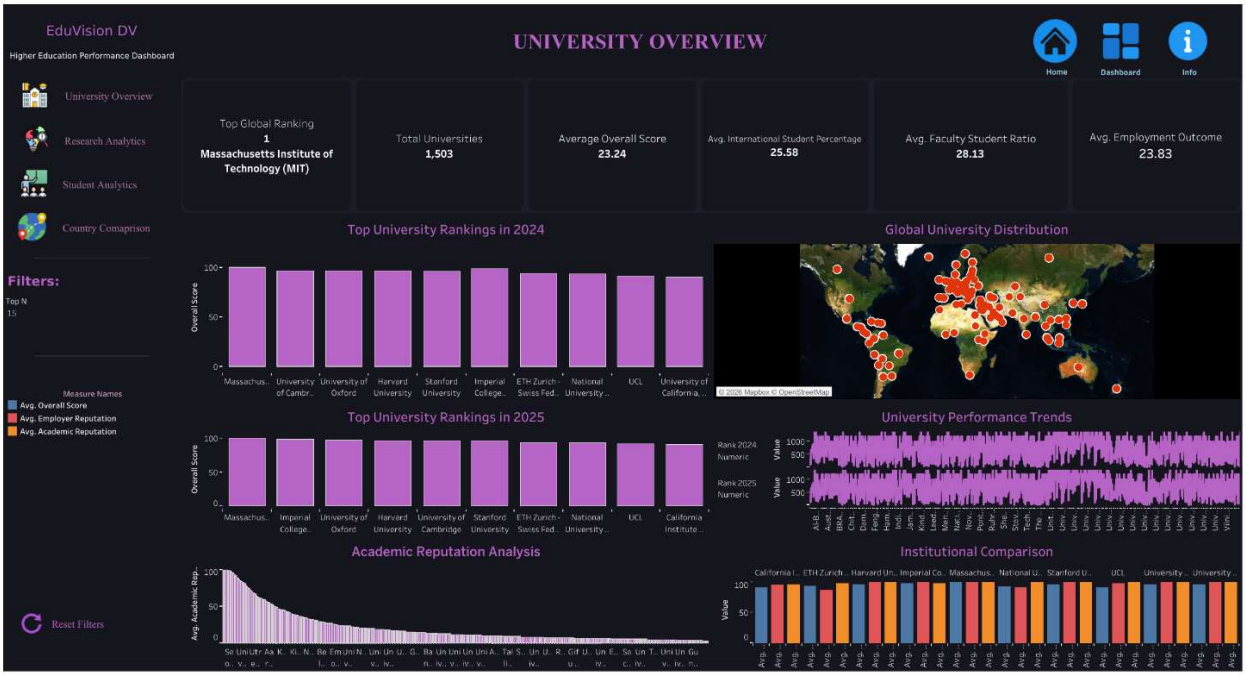


Fig. 5 University Overview dashboard

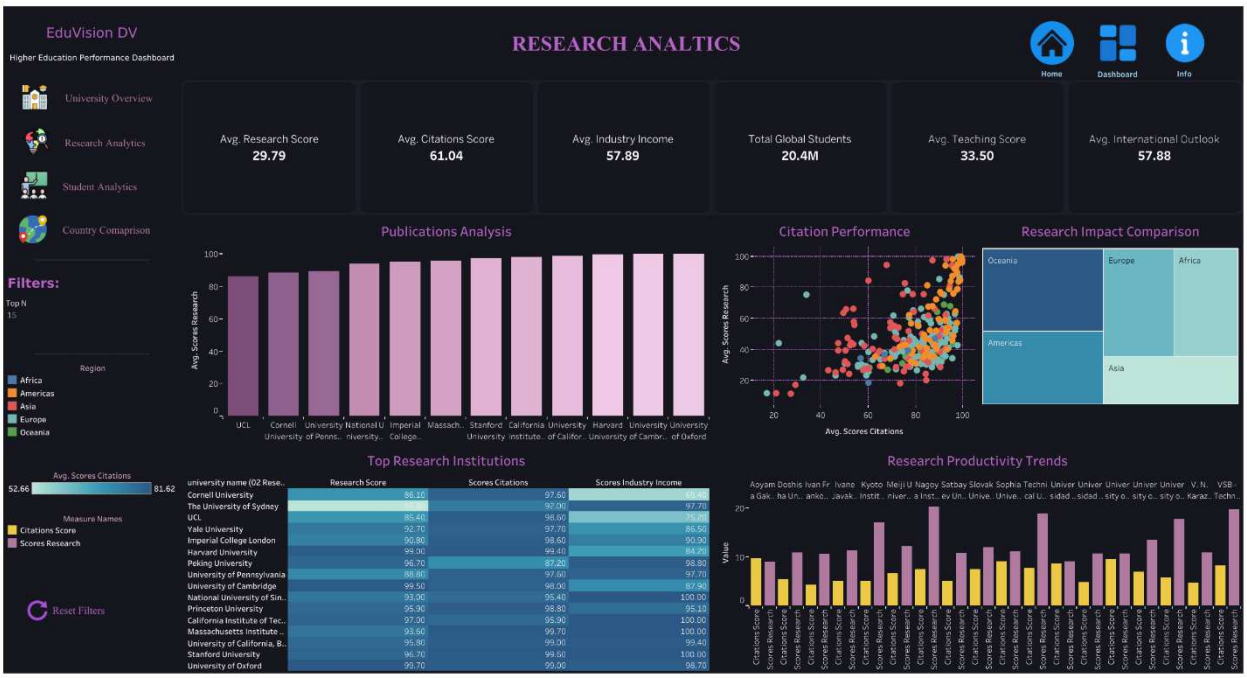


Fig. 6 Research Analytics dashboard

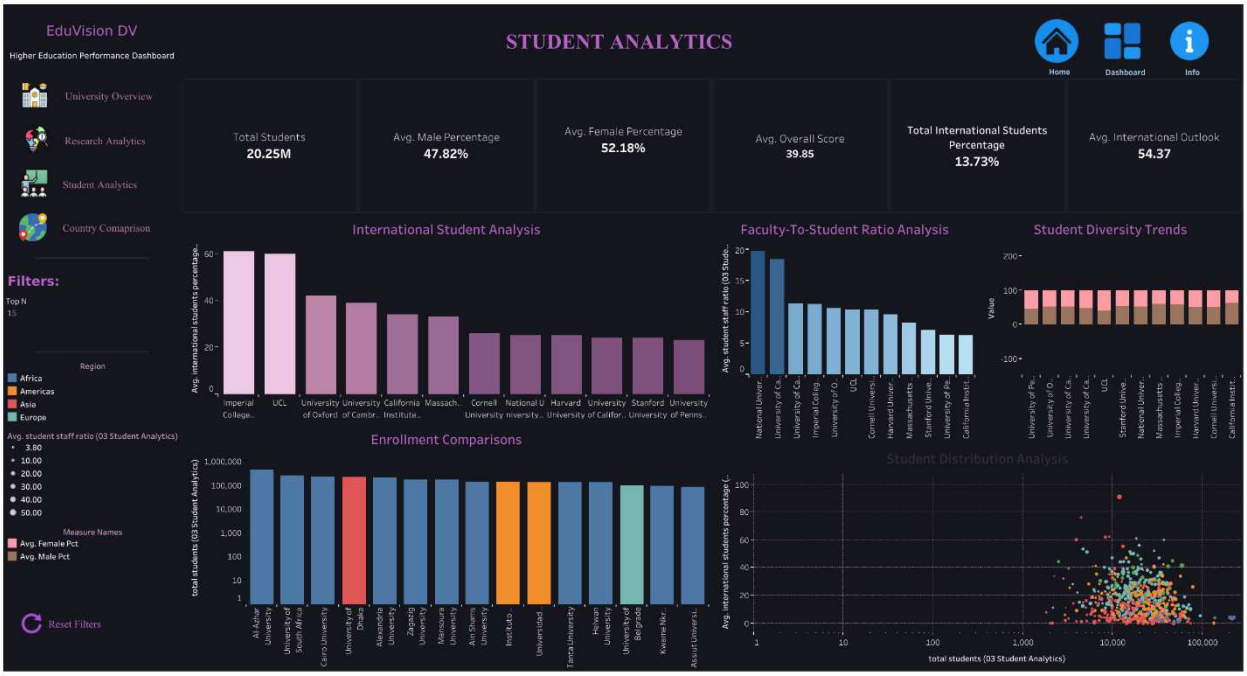


Fig. 7 Student Analytics dashboard

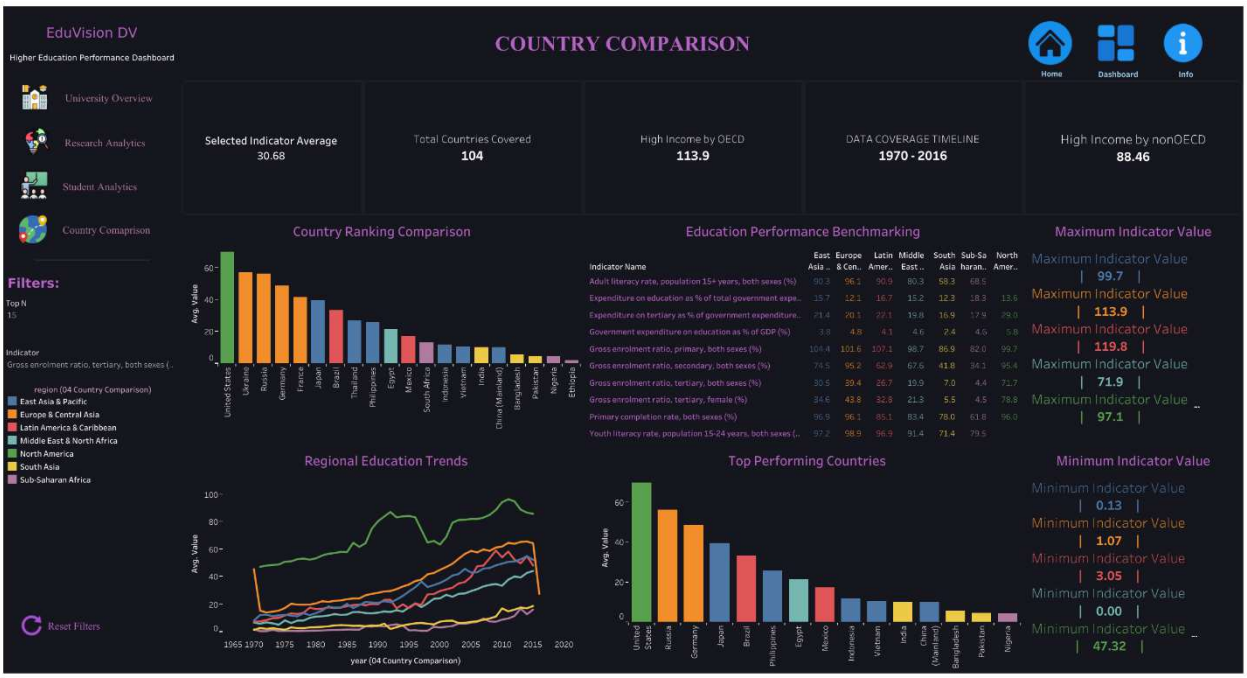


Fig. 8 Country Comparison dashboard

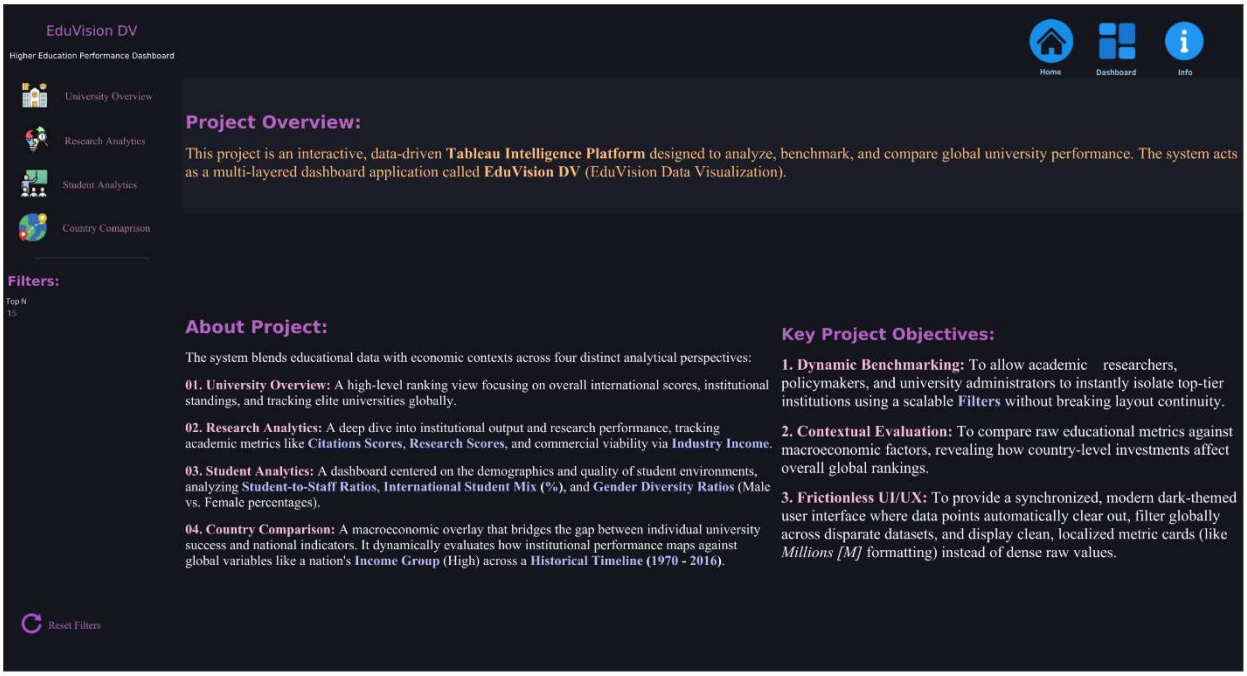


Fig. 9 Project overview (information) page of the workbook

11. DISCUSSION

The audit shows that the value of the integration lies as much in what it makes visible as in the merged data itself. First, a large part of the QS score field is estimated, and the estimated scores are systematically lower than the published ones; any dashboard average that pools them mixes two populations. Second, the THE 2024 overall score cannot be used for most rows, which is why the research analytics rely on pillar scores. Third, coverage differs across tables, so cross-dashboard comparisons involve different sets of universities. These observations are in line with the argument of Moed [1] that users of ranking systems are given finalised values without the means to examine how they arise, and with the caution of Selten et al. [2] about what ranking variables measure.

The country layer should be read as context rather than explanation. The World Bank data end in 2016 whereas the ranking editions are 2023 to 2025, so the two can be placed side by side but not linked causally. Pietrucha [3] relates country factors to

ranking position with regression on contemporaneous data, and no comparable inference is drawn here.

For dashboard design, the results suggest displaying provenance flags next to aggregate indicators, for example by splitting the Average Overall Score card into published and estimated values. In terms of the design space of Sarikaya et al. [6], the workbook is an exploratory analytical dashboard with filter, highlight and parameter interaction; its interaction design has not been evaluated with users.

12. LIMITATIONS

- **Estimated and reconstructed values.** 60.1% of QS, 82.6% of 2023-table and 83.2% of THE 2024 overall-score values are estimated, reconstructed or constant; conclusions based on overall scores are therefore approximate for a large share of institutions.
- **Rank approximation.** Banded ranks are stored as band lower bounds.
- **Time mismatch.** World Bank data cover 1970 to 2016, the rankings 2023 to 2025.
- **Entity matching.** Matching is name-based; residual errors cannot be excluded although identifiers passed referential-integrity checks.
- **Secondary sources.** The datasets are Kaggle-hosted copies; QS 2025, the 2023 table and World Bank EdStats are identified with their Kaggle uploader, title and URL (References [9]–[11]). The exact listing of the THE 2024 source, and version numbers, licences and download dates for all four datasets, are not recorded in this paper; the licence terms of each source should be checked before the data are reused.
- **Evaluation.** Only functional testing by the project team was performed; there was no user study, so usability and usefulness are not established (cf. Bodily and Verbert [8]).
- **Unverified components.** The six KPIs were reported as documented by the dataset's source columns but not independently reproduced from the pipeline scripts for this paper, and the Research Productivity Index, as noted in Section 9, is not plotted directly in the dashboard.
- **Scope.** The work is a data-integration and visual-analytics case study; it contains no predictive model and supports no causal claims.

13. CONCLUSION

EduVision DV integrates three university-ranking datasets and World Bank education indicators into a single validated master-keyed data model and exposes them through five coordinated Tableau dashboards. The coverage audit and data-quality analysis quantify how far the integrated data are published rather than estimated, reconstructed or constant, and show that aggregate indicators such as the pooled average overall score can be misleading if these differences are hidden. The contribution is a documented and reproducible integration workflow with explicit disclosure of data-quality conditions, together with exploratory descriptive findings. It does not establish predictive validity, causal relationships or user-tested usability.

14. FUTURE SCOPE

- Show provenance flags (published, estimated, reconstructed, imputed) directly in the dashboards and report KPI cards by provenance.
- Update the World Bank data beyond 2016 or align them with the ranking years, so that country context and rankings cover the same period.
- Add further ranking systems (for example ARWU or CWUR) and improve entity matching using the techniques surveyed in [7].
- Extend the indicator analysis to principal component or factor analysis, as in [2].
- Add Tableau navigation buttons to the home dashboard, as recommended in the project test report.
- Conduct a usability study with the target audiences (students, researchers, administrators, policymakers) using a standard instrument.

STATEMENTS AND DECLARATIONS

Funding. No funding was received to assist with the preparation of this manuscript.

Competing interests. The author has no relevant financial or non-financial interests to disclose.

Ethics approval and consent. Not applicable. The study used only publicly available secondary data and involved no human participants or animals.

Data availability. The source datasets are publicly available on Kaggle [9]–[11]. The integrated Tableau workbook (EduVision_DV.twbx) is published on Tableau Public [15]. The project repository, including the data-processing scripts, is maintained at https://github.com/springboardmentor1234r/Higher-Education-Intelligence-System-for-Institutional-Performance-and-Academic-Excellence-Analysis (branch: Vaibhav_Kumawat). The script used to reproduce the derived results in Section 10 (reproduce_results.py) is provided alongside this submission as supplementary material.

Author contributions. Vaibhav Kumawat conceived the project, built the data pipeline and Tableau dashboards, performed the analysis, prepared all figures and tables, and wrote the manuscript.

Acknowledgements. The author thanks Ruchi Shriwas, Assistant Professor, Department of Computer Science and Engineering, Swami Vivekanand College of Engineering, Indore, for guidance during the Major Project.

REFERENCES

[1] Moed, H.F.: A critical comparative analysis of five world university rankings. Scientometrics 110(2), 967–990 (2017). https://doi.org/10.1007/s11192-016-2212-y

[2] Selten, F., Neylon, C., Huang, C.-K., Groth, P.: A longitudinal analysis of university rankings. Quant. Sci. Stud. 1(3), 1109–1135 (2020). https://doi.org/10.1162/qss_a_00052

[3] Pietrucha, J.: Country-specific determinants of world university rankings. Scientometrics 114(3), 1129–1139 (2018). https://doi.org/10.1007/s11192-017-2634-1

[4] Daraio, C., Bonaccorsi, A.: Beyond university rankings? Generating new indicators on universities by linking data in open platforms. J. Assoc. Inf. Sci. Technol. 68(2), 508–529 (2017). https://doi.org/10.1002/asi.23679

[5] Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., Streit, M.: LineUp: Visual analysis of multi-attribute rankings. IEEE Trans. Vis. Comput. Graph. 19(12), 2277–2286 (2013). https://doi.org/10.1109/TVCG.2013.173

[6] Sarikaya, A., Correll, M., Bartram, L., Tory, M., Fisher, D.: What do we talk about when we talk about dashboards? IEEE Trans. Vis. Comput. Graph. 25(1), 682–692 (2019). https://doi.org/10.1109/TVCG.2018.2864903

[7] Papadakis, G., Skoutas, D., Thanos, E., Palpanas, T.: Blocking and filtering techniques for entity resolution: A survey. ACM Comput. Surv. 53(2) (2020). https://doi.org/10.1145/3377455

[8] Bodily, R., Verbert, K.: Review of research on student-facing learning analytics dashboards and educational recommender systems. IEEE Trans. Learn. Technol. 10(4), 405–418 (2017). https://doi.org/10.1109/TLT.2017.2740172

[9] Monfared, M.: QS World University Rankings 2025. Kaggle (2025). https://www.kaggle.com/datasets/melissamonfared/qs-world-university-rankings-2025. Accessed 4 Oct 2026

[10] alitaqi000: World University Rankings 2023. Kaggle (2023). https://www.kaggle.com/datasets/alitaqi000/world-university-rankings-2023. Accessed 4 Oct 2026

[11] The World Bank: Education Statistics. Kaggle. https://www.kaggle.com/datasets/theworldbank/education-statistics. Accessed 4 Oct 2026

[12] QS Quacquarelli Symonds: QS World University Rankings: Methodology (2025 edition). https://www.topuniversities.com/qs-world-university-rankings/methodology

[13] Times Higher Education: World University Rankings 2022: Methodology (pre-WUR 3.0 weighting, applicable to the 2023 table). https://www.timeshighereducation.com/node/717460

[14] Times Higher Education: World University Rankings 2024: Methodology (WUR 3.0 weighting, applicable to the THE 2024 table). https://www.timeshighereducation.com/world-university-rankings-2024-methodology

[15] Tableau Software: Tableau Desktop (build 2026.2.2) and EduVision DV workbook (EduVision_DV.twbx). https://public.tableau.com/views/UniversityOverview_17890783104670/UNIVERSITYOVERVIEW

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- EduVisionDV.twbx