Text

Description automatically generated

**MSc Information Technology**

**COMP11108 – Data Analysis & Visualization**

**Coursework Report**

**Exploring and Visualizing University Ranking Data**

**By Group-26**

|  |  |  |
| --- | --- | --- |
| **NAME** | **BANNER ID** | **CONTRIBUTION** |
| **Prajwalaradhya Shivakumaraswamy Kesaramadu** | **B01759301** | **4,5,6,7** |
| **Stephen Kwaku Pometsey** | **B01757368** | **2,3,4,5** |
| **Sayyar Hayat** | **B01772768** | **1** |
| **Aman Misra** | **B01746656** | **6,7** |
| **Sreeraj Karuvanthodi Ramachandran** | **B01764963** | **6,7** |
| **Muhammed Ali Panthalingal** | **B01755979** | **5** |
| **Qaiser Sajawal** | **B01794675** | **1** |

Table of Contents

[Task 1: Data Selection 3](#_Toc195552018)

[Key insights features 3](#_Toc195552019)

[Task 2: Data Exploration 5](#_Toc195552020)

[Data Structure Inspection 5](#_Toc195552021)

[Summary Statistics and Missing Data Analysis 5](#_Toc195552022)

[Task 3: Data Cleaning and Transformation 7](#_Toc195552023)

[Task 4: Visualisation Design 9](#_Toc195552024)

[Types of Charts Used 9](#_Toc195552025)

[Visualizations 9](#_Toc195552026)

[Task 5: Interactive Dashboard 14](#_Toc195552027)

[Charts in the Dashboard 14](#_Toc195552028)

[Dashboard Interactivity and Features 14](#_Toc195552029)

[Task 6: Data Storytelling 17](#_Toc195552030)

# Task 1: Data Selection

The chosen dataset focuses on 2011-2016 global university rankings and includes a comprehensive set of indicators such as the number of international students, student-to-staff ratio, teaching and research scores, citation impact, income, total score, and year. These variables provide a well-rounded view of institutional performance and make the dataset ideal for both analysis and visualization.

This dataset was selected due to its relevance in today’s global educational landscape, where universities are evaluated not only on academic quality but also on their international outlook and operational efficiency. It offers opportunities to explore meaningful patterns, trends over time, and correlations between variables—for example, how research output relates to teaching quality or how international diversity affects overall ranking.

Its quantitative and time-based nature makes it particularly suitable for creating compelling visualizations, enabling clear communication of complex insights. Moreover, the data is sourced from a reputable and real-time ranking system, ensuring both credibility and practical relevance. Overall, this dataset supports the goals of the course by allowing the application of data analysis techniques to a real-world, socially significant topic.

## Key insights features

**Top 15 Universities by Total Score:** This feature identifies the top-performing universities globally based on the overall score, which is a composite metric reflecting multiple indicators like teaching, research, citations, income, and international outlook.

* These universities often include globally recognized institutions from countries such as the United States, the United Kingdom, and Switzerland.
* The high total scores reflect excellence across multiple dimensions, not just academics.
* The top 15 can serve as benchmarks for quality and best practices in higher education.
* There may be recurring universities year after year, indicating consistency in performance and prestige.

**Average Total Score of Universities by Country:** This calculates the mean total score of all universities from each country, providing insight into national-level performance rather than individual excellence.

* Countries like the UK, US, and Switzerland tend to have higher average scores, reflecting strong, well-funded education systems.
* Countries with fewer but high-performing universities may show high averages (e.g., Singapore, Netherlands).
* This metric highlights not only global leaders but also emerging countries improving in education quality.
* It helps differentiate between countries with one top institution and those with a broader base of quality education.

**Correlation Between Teaching and Research Scores:** This analysis explores the relationship between teaching quality and research output, determining whether institutions strong in one area are also strong in the other.

* A **positive correlation** suggests that investment in research often goes hand-in-hand with high-quality teaching environments.
* However, some institutions may focus more on research (e.g., producing publications) than on teaching, leading to outliers.
* This insight can help students and policymakers understand whether a strong research reputation also translates into a better learning experience.
* If correlation is weak, it may highlight the need for balanced development in universities.

**Top 15 Universities by International Student Percentage:** This feature ranks universities based on the proportion of international students enrolled, highlighting global diversity and institutional appeal.

* Universities in the UK, Australia, and Switzerland often have high international student populations due to language, visa policies, and global outreach.
* A high percentage of international students often reflects strong global branding, multicultural learning environments, and internationally relevant programs.
* It also suggests economic factors, as international students are often a source of revenue for universities.
* Students seeking culturally diverse campuses or global networks may find this information particularly useful.

**Student/Staff Ratio per Country:** This metric compares the number of students to academic staff in universities, averaged per country, to evaluate potential learning support and teaching quality.

* Lower ratios typically indicate smaller class sizes, better access to faculty, and potentially higher teaching quality.
* Countries with well-funded education systems often show lower ratios (e.g., Germany, Netherlands).
* Higher ratios in some countries may reflect overcrowded universities or underfunded systems.
* It’s an important metric for prospective students as it impacts their academic experience directly.

# Task 2: Data Exploration

## Data Structure Inspection

The preliminary phase of Exploratory Data Analysis (EDA) on the original dataset, is essential to inspect the structure and composition of the dataset to understand its scope, variables, and potential data quality issues. The dataset under review is extracted from the Times\_University\_ranking\_Data.xlsx file. This sheet comprises a total of 2,603 observations (rows) and 14 distinct attributes (columns), each representing various metrics related to global university performance.

The variables include categorical fields such as university\_name and country, alongside quantitative indicators like teaching, research, citations, and total\_score. Variables such as num\_students, student\_staff\_ratio, and international\_students provide further insights into the institutional characteristics. The year column enables longitudinal analysis across different time periods.

A preliminary inspection of the first five records reveals that while the core metrics appear numerical and well-structured, some fields such as female\_male\_ratio contain inconsistent values — including percentages, ratios like “33 : 67”, and even time stamps. Similarly, the income field occasionally contains the placeholder '-', which indicates missing or unavailable data that must be addressed during cleaning.

This structural assessment serves as a foundational step in the data analysis pipeline, allowing us to identify potential anomalies and prepare for subsequent steps such as summary statistics, missing value profiling, and visualization. By clarifying the data types and previewing content at this stage, we can move on to cleaning and analysis process in the following stages of the EDA.

## Summary Statistics and Missing Data Analysis

This step provides a quantitative overview of the data’s distribution, variability, and completeness, which are critical for determining appropriate cleaning strategies.

The dataset includes several numeric fields, such as teaching, research, citations, num\_students, and student\_staff\_ratio. Summary statistics reveal that the average teaching score across all universities is 37.80, while the mean research score is 35.91 and the mean citations score is relatively higher at 60.92. The average number of students per institution is approximately 23,874, and the average student-staff ratio is 18.45. Furthermore, the mean proportion of international students is 15.44%, indicating moderate global representation across institutions.

In terms of data completeness, four fields exhibit missing values. The female\_male\_ratio column has the highest rate of missing data, with 233 missing values (8.95%), followed by international\_students (2.57% missing), num\_students (2.27%), and student\_staff\_ratio (2.27%). All other fields, including core performance indicators like teaching, research, citations, and total\_score, are fully populated.

The identification of missing values in these specific columns signals the need for strategic handling. For example, female\_male\_ratio, being non-numeric and inconsistently formatted, may be a candidate for removal or categorical transformation. On the other hand, fields like num\_students and international\_students, which contain a small proportion of nulls, could be imputed using statistical methods such as mean or median substitution.

This statistical and missing value assessment equips us with a clearer understanding of the data's quality and guides the decisions in the subsequent cleaning and transformation steps of the analysis pipeline.

A screenshot of a graph

AI-generated content may be incorrect.

There’s a huge bar labelled “Null” at the top — this means there are missing or non-numeric values in the total\_score field that were converted to Null during cleaning.

The EDA shows that there a lot of nulls and non-numeric values that need to be looked at before any meaningful insight could be drawn. This is handled in task 3. Below.

# Task 3: Data Cleaning and Transformation

As indicated in task 2 above, our data needs to be cleaned, structured, and transformed to analysis-ready format, a comprehensive data cleaning process was carried out using Tableau Prep. This essay outlines the methodology, transformation logic, and scripting techniques employed to prepare the dataset for insightful university performance analysis.

The process began with importing the dataset into Tableau Prep, where an initial visual review was conducted. It was immediately evident that the data required several layers of cleaning. Issues identified included numeric fields containing text or symbols, such as "=124" in rank fields, or "-" placeholders used to indicate missing values. Some numeric fields were formatted as strings, such as "2,011" for the year, and others had symbols like commas and percentage signs that would interfere with aggregation and filtering.

The first cleaning task involved renaming and standardizing the column headers. For instance, university\_name was renamed to university, income to Industry Income, and international\_students to International Students to improve readability and consistency across the dataset. This helped streamline further processing and dashboard labeling in Tableau.

Next, formatting issues were tackled through calculated fields in Tableau Prep. To remove unwanted characters, simple REPLACE() functions were used. One such example was cleaning

Rank entries like “=124”:

REPLACE([WorldRank],” =”, “”)

This transformation stripped away any equals signs while preserving the numerical rank value. Similarly, to clean fields containing commas or percentages, such as Number of Students and

International Students, calculated fields were created:

FLOAT(REPLACE[Number of Students], “,”, “”))

FLOAT(REPLACE[International Students], “%”, “”)).

This ensured that these fields could be correctly interpreted as numerical values and used in charts

or aggregations without error.

Another critical step in the data preparation process involved handling null and placeholder values. Tableau Prep automatically recognized the placeholder "-" as a string, and once the column was converted to a numeric data type, these entries were automatically marked as null. The strategy for dealing with null values depended on the importance of the field to the core analysis. Fields like Total Score and Industry Income, which are essential performance indicators, were treated strictly: all rows with nulls in these fields were filtered out using Tableau Prep's built-in filtering.

Filter > Null Values.

On the other hand, contextual fields such as Student-Staff Ratio, Number of Students, and International Students were preserved by replacing nulls with calculated average values. These averages were derived externally using Python (Pandas) to ensure statistical accuracy. The scripts

Used to calculate and fill in these fields were:

IF ISNULL([Student-Staff Ratio]) THEN 18.45 ELSE [Student-Staff Ratio] END

IF ISNULL([Number of Students]) THEN 23873 ELSE [Number of Students] END

IF ISNULL([International Students]) THEN 0.1544 ELSE [International Students] END

These filled values ensured data completeness while maintaining the integrity of the dataset for visualization purposes. In addition to cleaning, some transformations were also performed to enhance the analytical capability of the dataset. For instance, rank ranges like "201–250" were split into two parts and converted into numeric start and end ranks. A new calculated field, Rank Midpoint, was introduced to provide a single representative value for analysis purposes. This was especially useful when analysing rank distribution trends over time. Some fields were intentionally removed from the final dataset due to high volumes of missing values or low relevance to the main objectives of the analysis. For example, the Female-Male Ratio field, which had over 200 null entries, was excluded to prevent skewing the dataset. By prioritizing core performance indicators, the data remained focused and robust. At the end of the cleaning process, the dataset was significantly refined. The row count decreased from 2,603 to 1,061, not due to arbitrary deletion, but through purposeful filtering based on the completeness and analytical value of each record. All essential fields were properly typed, formatted, and filled, and unnecessary clutter was removed to enable smoother analysis and dashboard design. In conclusion, the data cleaning process in Tableau Prep successfully transformed the raw Times University Ranking dataset into a structured, high-quality dataset ready for performance analysis. Through a combination of thoughtful filtering, precise string manipulation, calculated fields, and statistical imputation, the dataset now supports accurate and insightful visualizations. This process demonstrates how robust data preparation is not only foundational but indispensable to reliable data analytics.

# Task 4: Visualisation Design

In this analysis, we explored global university rankings using Tableau, leveraging various chart types such as bar charts, scatter plots, bubble charts, and a world map. The dashboards include interactive elements like Year parameters to filter views dynamically. We also used calculated fields to refine metrics and enhance insights.

## Types of Charts Used

1. **Bar Chart:** A bar chart is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.
2. **Scatter Plot:** A graphical representation that displays the relationship between two numerical variables.
3. **Bubble Chart:** A bubble chart is a variation of a scatter chart in which the data points are replaced with bubbles, and an additional dimension of the data is represented in the size of the bubbles.
4. **World Map:** A visual representation of the Earth's surface, used to display data related to geographical locations.

## Visualizations

1. **Top 15 Universities by Total Score (Global Ranking)**

This bar chart presents the Top 15 universities globally, ranked by their Total Score. Each bar represents a university, with the height indicating overall performance across indicators like teaching, research, citations, and international outlook. The bars are coloured by Country, allowing quick visual comparison between nations. The United States and the United Kingdom dominate the list, as seen from the orange and blue colour distribution. A year filter is applied, enabling users to dynamically explore rankings for different years. The use of a Top N Set ensures only the best-performing universities for the selected year are displayed. The chart is sorted in descending order for clarity, highlighting institutions like Caltech, Stanford, and Oxford at the top. Tooltips provide exact score values for each university. This visualization effectively communicates global academic excellence. It’s ideal for benchmarking and identifying leaders in higher education.

**A graph of a number of people

AI-generated content may be incorrect.**

1. **Average total score of Universities by Country**

The second bar chart illustrates the average total score of universities by country, giving insight into how countries perform academically on average. Each bar represents a country, with the height indicating the average score of its universities. The color gradient, ranging from dark to light blue, visually encodes performance — darker bars represent higher averages. A key addition is the count of universities per country, visible in the tooltip, which provides context behind each average score. For example, Singapore has an average score of 65.06 based on 12 universities, highlighting strong, consistent performance. This prevents misinterpretation of averages by exposing how many data points contribute to each. The chart uses a filter for year, allowing dynamic analysis over time. It helps analysts compare performance fairly, whether a country has many institutions or just a few elite ones. This visualization combines clarity and context, making it ideal for regional benchmarking.

**A graph of different colored squares

AI-generated content may be incorrect.**

1. **Correlation Between Teaching and Research Scores**

This scatter plot illustrates the correlation between Teaching and Research scores for universities worldwide. Each dot represents a single university, plotted by its Teaching score (x-axis) and Research score (y-axis). The chart reveals a strong positive relationship — as teaching scores increase, and research scores tend to increase as well. This suggests that universities excelling in teaching often also excel in research. Each point is color-coded by country, helping to spot regional trends or clusters. For instance, certain countries like the United States or United Kingdom may dominate the top-right quadrant, indicating top performance in both areas. The spread of points also highlights outliers — schools that are strong in one metric but not the other. The Year filter allows dynamic exploration over time. This visualization supports comparative analysis and reveals how tightly linked teaching and research quality are in global higher education.

**A graph with many colored circles

AI-generated content may be incorrect.**

1. **Top 15 Universities by International Student Percentage**

This bar chart displays the Top 15 universities with the highest percentage of international students, serving as an indicator of global engagement. Each bar represents a university, with the height showing the proportion of its student body from outside its home country. The chart reveals that universities like École Polytechnique Fédérale de Lausanne and Imperial College London lead in attracting international students. Institutions from the United Kingdom, Switzerland, Australia, Hong Kong, and the United States are prominent in this ranking. Bars are color-coded by country to highlight geographic patterns. The Year filter allows dynamic exploration of changes in international presence over time. A high international student percentage may indicate a strong global reputation, diverse campus culture, and active recruitment abroad. This visualization complements performance rankings by showing how internationally connected universities are. It also helps identify strategic leaders in global education.

**A close-up of a graph

AI-generated content may be incorrect.**

1. **Average Score vs Student Population by Country**

This bubble chart visualizes the average university ranking scores of countries based on the TIMES dataset, with each bubble representing a country. The size of each bubble corresponds to the total number of students in that country’s universities, while the label inside displays the country name along with its average score. The United States stands out with the largest bubble and the highest average score of 65.18, indicating both a large student population and strong academic performance. Canada, Australia, the UK, and the Netherlands follow, each showing a solid combination of student size and ranking scores. This chart effectively highlights the global distribution of top-performing countries in higher education, reflecting both academic excellence and institutional scale.

A screenshot of a cellphone

AI-generated content may be incorrect.

1. **Global Distribution of Ranked Universities by Country**

This world map chart highlights the geographical spread of universities featured in the TIMES ranking dataset. Each country is shaded and labelled based on the number of universities it has in the ranking. The United States leads with 327 institutions, followed by the UK with 186, showcasing their dominance in global higher education. Germany, the Netherlands, Australia, and Canada also have significant representation, with 68, 66, 44, and 43 universities respectively. This visualization offers a clear picture of which countries contribute the most to the global academic landscape in terms of university count.

A map of different countries/regions

AI-generated content may be incorrect.

A map of the world

AI-generated content may be incorrect.

# Task 5: Interactive Dashboard

The dashboard presents a multi-dimensional analysis of university performance based on the TIMES World University Ranking dataset. It brings together four key visualizations to explore scores, global comparisons, and trends in higher education. Built in Tableau, the dashboard is fully interactive, allowing users to engage with the data in real time.

## Charts in the Dashboard

1. **Top 15 Universities by Total Score (Global Ranking)**

Displays the top 15 universities globally, ranked by their total score. Users can observe how institutional rankings change over time using the Year Selector, and if USA or UK is selected in the Country Selector, universities from the selected country are highlighted for focused comparison.

1. **Average Total Score of Universities by Country**

Shows the average total score of all ranked universities in each country. A gradient colour legend visually represents score ranges, enhancing clarity. Selecting a country from this chart also triggers an action filter on the scatter plot to show only data points from that country. The chart responds to both Year and Country selectors, allowing focused analysis on time and geography.

1. **Correlation Between Teaching and Research Scores**

This chart presents the relationship between teaching and research scores of universities, with each point representing an institution. The chart is responsive to:

* Year selection, filtering results to a particular period.
* Country selection from the bar chart, to show only that country's universities.
* Country Selector (USA/UK), to highlight respective universities for comparison.

1. **Top 15 Universities by International Student Percentage**

Highlights universities with the highest percentage of international students. Like other charts, it updates based on the Year Selector and highlights universities from USA or UK when selected via the Country Selector. This allows users to examine which global institutions are the most internationally diverse.

## Dashboard Interactivity and Features

The dashboard is designed with user-friendly interactive elements that allow dynamic exploration of the dataset. It uses dropdown filters, country selectors, and dashboard actions to provide personalized, real-time insights across all visualizations.

* **Year Dropdown Filter:** A central interactive control that updates all four charts simultaneously based on the selected academic year.
* **Cross-Chart Filtering:** Selecting a country in the average score chart filters the scatter plot to show only that country’s universities.
* **Dynamic Calculated Fields:** Parameters and calculated values drive chart interactivity, ensuring real-time updates and accurate insights.
* **Gradient Colour Legend:** Enhances the average score chart, giving a clear visual representation of performance levels.
* **Country Selector (USA/UK):** Highlights universities in all charts that belong to the selected country, making it easy to compare USA and UK-based institutions across multiple dimensions.
* **Dashboard Action (Filter):** A custom Tableau action filter is implemented on the Average Total Score by Country chart. When a user clicks on a country bar, it triggers a filter that updates the Teaching vs Research scatter plot with only that country’s universities—enabling deeper, country-specific insight.

**Default Dashboard View (No Filters Applied)**

The dashboard loads with all data visible. It displays the global overview of university performance without any filter or parameter applied. All charts reflect the complete dataset across all years and countries.

A screenshot of a graph

AI-generated content may be incorrect.

**Year Selection Using Dropdown Filter**

Upon selecting a specific year, all four charts update dynamically to show data for that year only. This is implemented using a parameter and calculated fields to filter data accordingly. This allows users to view year-wise trends across global universities.

A screenshot of a graph

AI-generated content may be incorrect.

**Country Selection via Chart**

A dashboard action filter is triggered when a country bar is clicked. This updates all charts to show only universities from the selected country (Australia in this case). Other charts remain unchanged, maintaining full context, while providing focused insights into one country's academic profile.

A screenshot of a computer

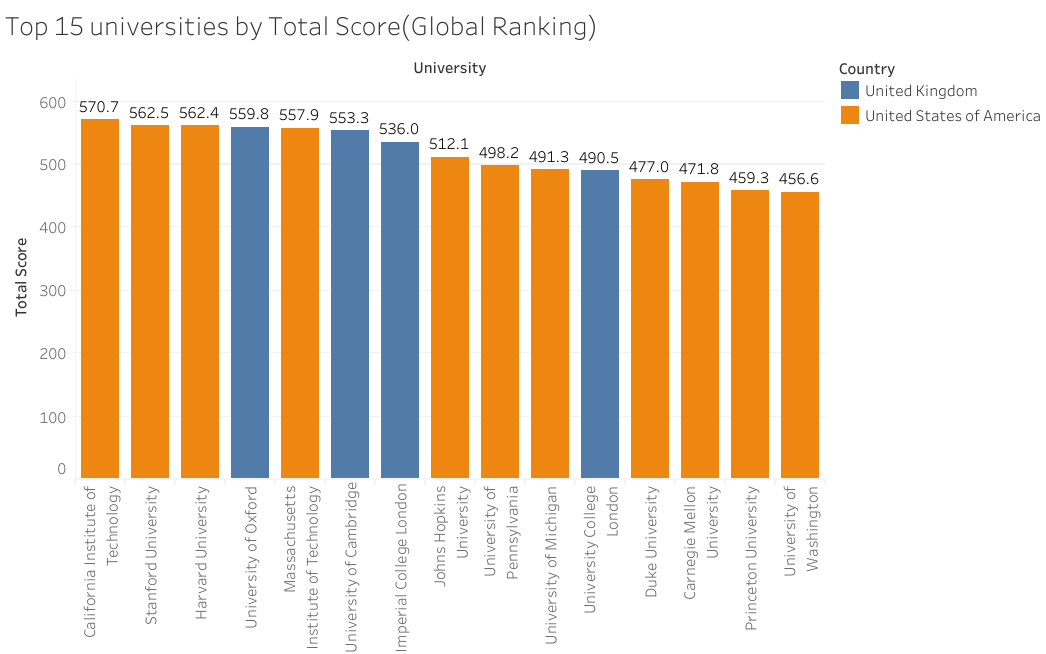
AI-generated content may be incorrect.

The interactive Tableau dashboard provides a dynamic and user-friendly platform to explore global university rankings. Using filters, parameters, and dashboard actions, users can gain detailed insights by selecting specific years and countries. These interactive elements enhance data exploration and enable focused comparisons across various academic metrics. Overall, the dashboard successfully combines functionality with visual clarity for effective data storytelling.

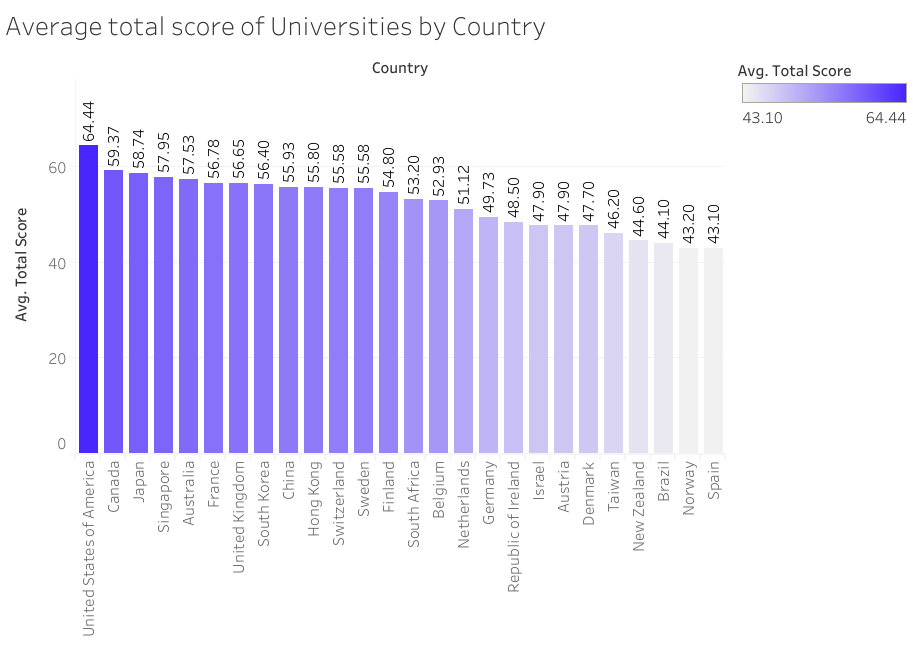
# Task 6: Data Storytelling

This data storytelling section guides the viewer through key insights drawn from global university rankings using six interactive visualizations. It highlights top-performing institutions, compares countries based on academic metrics, and explores trends in teaching, research, international diversity, and student populations. Together, the charts reveal how different regions contribute to the global landscape of higher education.

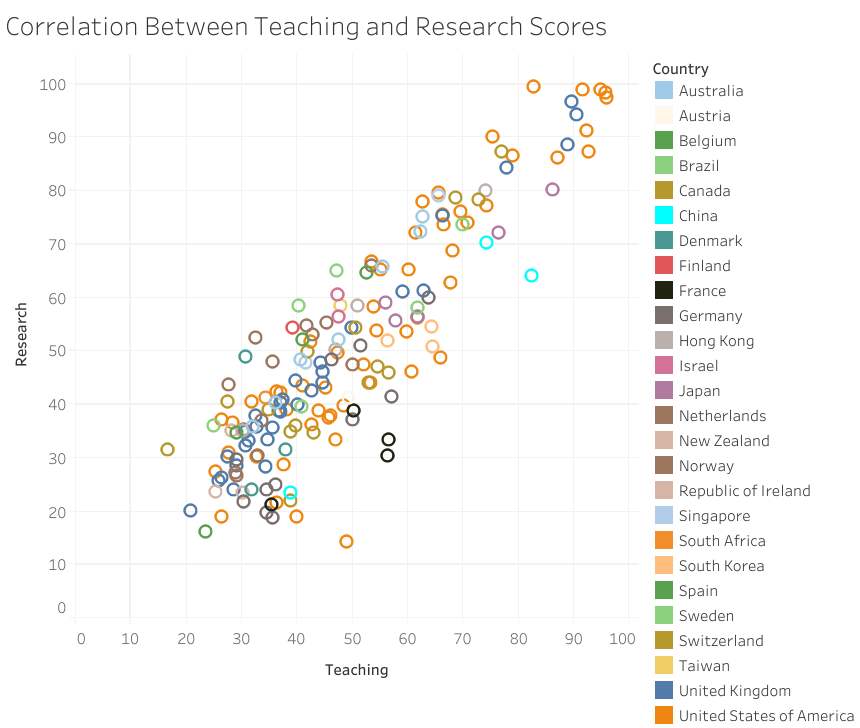
The **Top 15 Universities by Total Score** chart reveals that the United States clearly dominates the global stage, with 11 out of the 15 top-ranked universities. The remaining four come from the United Kingdom, displaying their shared strength in higher education. At the top of the list, the California Institute of Technology leads with a remarkable total score of 570.7, closely followed by Stanford University and Harvard. Even those at the lower end of the top 15, such as Princeton, Carnegie Mellon, and the University of Washington, maintain incredibly competitive scores above 450, confirming the ambitious standards maintained by these institutions.



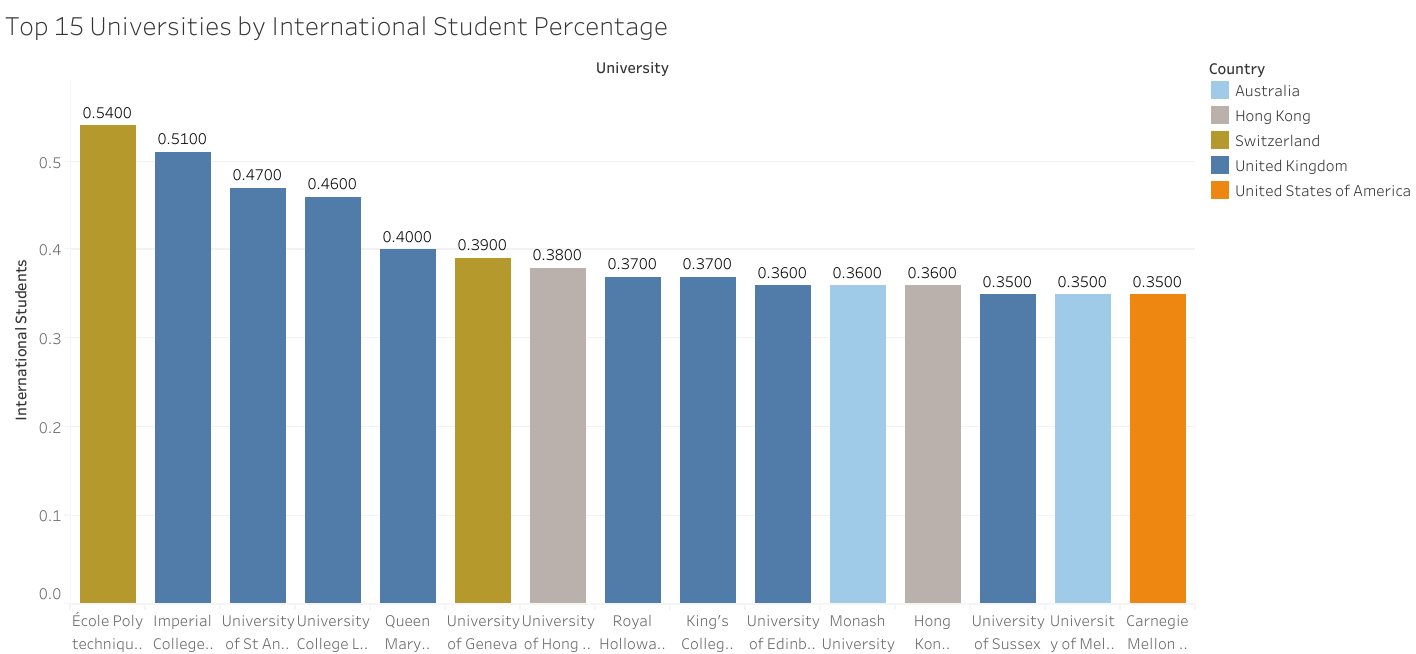
Looking at the **Average Total Score of Universities by Country**, countries like Singapore (65.6), the USA (65.18), and Canada (61.34) stand out with the highest averages, indicating both academic quality and consistency. In contrast, countries such as Israel, Norway, and Brazil score lower, suggesting that while they may have individual strong universities, their national averages are modest. This chart emphasizes how academic excellence is distributed globally and where the highest concentration of top-performing institutions can be found.



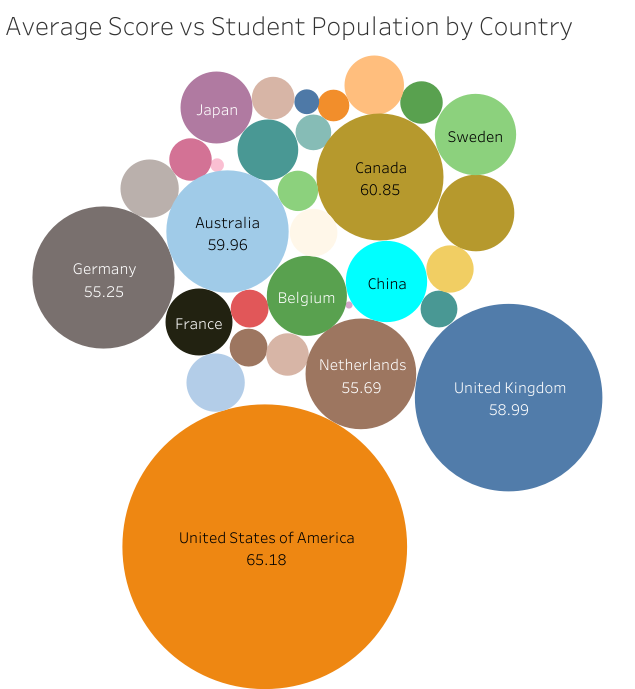
The **Correlation Between Teaching and Research Scores scatter plot** shows a tight and consistent cluster for UK universities, reflecting a balanced performance in both areas. The US universities, though slightly more spread out, still exhibit strong correlation, and include some of the best performers globally. Institutions like Caltech, Harvard, Stanford, and Princeton from the US, along with Oxford, Cambridge, and Imperial College from the UK, sit prominently in the upper-right region of the chart—signifying excellence in both teaching and research.



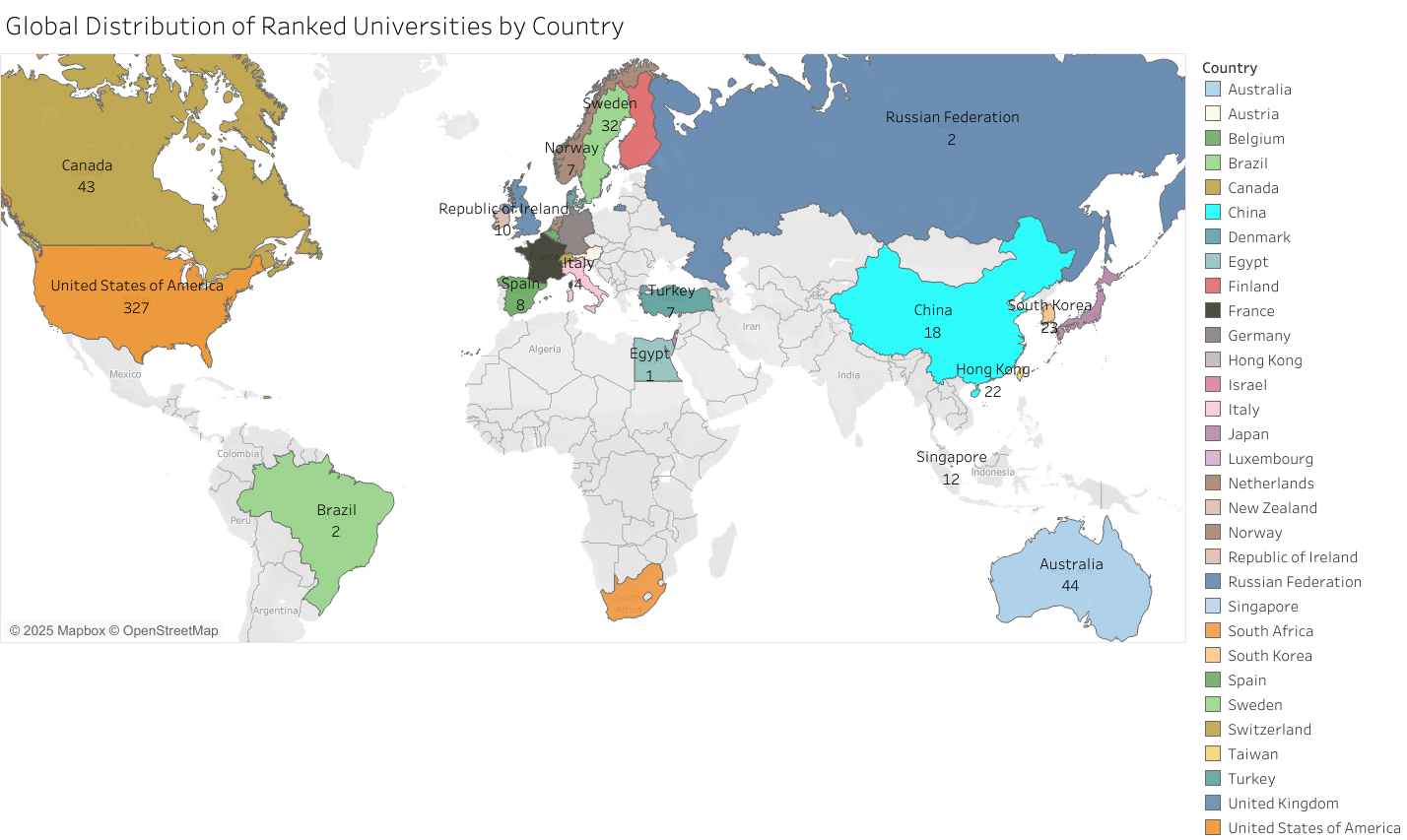
In terms of international engagement, the **Top 15 Universities by International Student Percentage chart** highlights the appeal of European and UK institutions to students worldwide. Leading this chart is the Swiss university École Polytechnique Fédérale de Lausanne with a percentage score of 3.2, followed closely by Imperial College London and the University of St Andrews from the UK. Carnegie Mellon University in the US and the University of Melbourne in Australia also feature in this list, showing that global mobility in education is thriving in select regions.



The **Average Score vs Student Population by Country bubble chart** further reveals the scale of higher education systems. The USA has the largest student population at over 8.1 million, paired with a strong average score of 65.18. The UK and Australia follow with 3.5 million and 1.49 million students respectively, each maintaining respectable average scores. Countries like Luxembourg and Italy, despite having smaller student populations, still demonstrate solid academic averages, suggesting quality over quantity in their educational systems.



Finally, the **Global Distribution of Ranked Universities map** brings it all together by showcasing how these institutions are spread around the world. The United States again leads with a total of 327 ranked universities, followed by the UK with 186 and Germany with 68. Australia, Canada, and the Netherlands also have notable representation, reflecting their strong presence in global academia.



Overall, this data story reveals that while the USA dominates in terms of volume and average performance, other countries like the UK, Singapore, and Switzerland hold significant positions of strength in research quality and internationalization. The interactive dashboard effectively guides the viewer through these insights, offering both high-level overviews and detailed explorations of university performance across various dimensions.