6.1: Sourcing Open Data

THE World University Rankings 2016-2023 Data Set

Summary and Source: The THE World University Rankings provide the definitive list of the world's best universities, with an emphasis on the research mission. It is the only global university league table to judge research-intensive universities across all their core missions: teaching (the learning environment); research (volume, income and reputation); citations (research influence); industry income (knowledge transfer) and international outlook (staff, students and research). It uses 10 carefully calibrated performance indicators to provide the most comprehensive and balanced comparisons. Besides the World University Ranking systems, there are also information about the schools and countries, about the education expenditure (per country and year) – OECD Education resources - Public spending on education - OECD Data. All these data sources can be classified as trustworthy.

Original data, as well as ranking methodology described per each year, is available on the official website World University Rankings | Times Higher Education (THE).

Collection: Data was collected and aggregated from the departments of public health from hundreds of U.S. counties and states. The data from THE World University Rankings comes from different sources. **Institutional data** – self-submitted on the THE Portal. The **bibliometric data** supplier Elsevier examined more than 121 million citations to 15.5 million journal articles, article reviews, conference proceedings, books and book chapters published over five years. **Academic reputation survey** - A survey was sent to a sample of academics selected by THE, in which they asked them to nominate the universities that they perceive to be the best for Teaching and/or Research in their field. **Reference data** THE incorporates reference datasets into its model to convert country-level data provided by institutions via the portal (e.g. research income in a local currency) to a single comparable dataset for all institutions. The OECD Data comes from the countries members of OECD and there is official data. The both data sets comes from external data source.

Contents: Contains two data sets

Table: THE_World_Universities_2016-2023_rankings.cvs

Column Name	Description	Time Variant/Invariant	Structured/ Unstructured	Quantitative /Qualitative	Nominal/Ordinal /Discrete/Continuous
rank	Rank of University (smaller values are better)	Invariant	Structured	Quantitative	Ordinal/ Discrete
name	Name of University	Invariant	Structured	Qualitative	Nominal
location	Location of University	Invariant	Structured	Qualitative	Nominal
scores_overall	Overall (total_score)	Invariant	Structured	Quantitative	Continuous

scores_citations	Citations (citations/research influence)	Invariant	Structured	Quantitative	Continuous
scores_industry_income	Research income from industry & commerce / Academic Staff	Invariant	Structured	Quantitative	Continuous
scores_ international outlook	International Outlook (international/ staff, students and research)	Invariant	Structured	Quantitative	Continuous
scores_research	Research (volume, income and reputation	Invariant	Structured	Quantitative	Continuous
scores_teaching	Teaching (teaching/the learning environment)	Invariant	Structured	Quantitative	Continuous
stats_number_students	Number of FTE Students	Invariant	Structured	Quantitative	Continuous
stats_student_staff_ratio	Number of Students per Staff	Invariant	Structured	Quantitative	Continuous
stats_pc_intl_students	International Students	Invariant	Structured	Quantitative	Continuous
stats_female_male_ratio	Female-male ratio	Invariant	Structured	Quantitative	Continuous
Year	Year of ranking	Invariant	Structured	Quantitative	Ordinal

Tabel: Exp_public_privat_ education.cvs

Column Name	Description	Time Variant/Invariant	Structured/ Unstructured	Quantitative /Qualitative	Nominal/Ordinal /Discrete/Continuous
LOCATION	**Name of Country**	Invariant	Structured	Qualitative	Nominal
INDICATOR	**Type of Educational Spending**	Invariant	Structured	Qualitative	Nominal
SUBJECT	**Type of Education**	Invariant	Structured	Qualitative	Nominal
MEASURE	**% of GDP**	Invariant	Structured	Quantitative	Continuous
FREQUENCY	**How often the results are published**	Invariant	Structured	Quantitative	Continuous
TIME	**Year of publication**	Invariant	Structured	Quantitative	Continuous
Value	**Indicator Value**	Invariant	Structured	Quantitative	Continuous

Limitations: In this ranking are taken into consideration only the institutions that teach at an undergraduate level, usually indicated by having more than zero undergraduate degrees awarded. Postgraduate-only institutions are therefore not in the ranking. Also they must not be focused on a single narrow subject area (more than 80% of their publication output is from one subject area). They must not have more than two of the critical values (academic staff, international academic staff, research staff, students, international students, undergraduate degrees awarded, doctorates awarded, institutional income, research income, research income from industry and commerce) as null (either marked by the institution as "unavailable" or "withheld"). Null values will cause any metric based on that value to also be null. Institutions that have requested not to participate in the ranking or that are not eligible for other institution-specific reasons have been excluded.

Ethics: The dataset does not contain any HIPPA-related information.

Relevance: I believe this data set meets the necessary requirements for this project as it is open source, includes a geospatial component, and meets the size and variable requirements. Also this data sets presents a new data and the newest World University Rankings 2023.

Data Cleaning Measures

Wrangling Steps

Columns dropped	Columns renamed	Columns' type changed	Comment/Reason
FREQUENCY			Unnecessary column for analysis.
	LOCATION	Country	For consistency
	INDICATOR	Spending_Indicator	For clarity
	SUBJECT	Type od Education	For clarity
	MEASURE	Measuring Spending	For clarity
	TIME	Year	For consistency
	Change code country to full name		For clarity
		Year	Changed the type of column to string.
		scores_teaching	Changed the type of column to Int.
		scores_citations	Changed the type of column to Int.
		scores_research	Changed the type of column to Int.
		scores_industry_income	Changed the type of column to Int.
		scores_international_outlook	Changed the type of column to Int.
	Created 2 new columns:		
	male_%		For clarity
	female_%		For clarity

Consistency Checks

Dataset	Missing values	Missing values treatment	Duplicates
Ranking	There are 996	These were replaced with 0	There were no duplicates.
Spending	There are no missing		There were no duplicates.

Questions to Explore the Analysis

1. What are the Top Universities in the World?
2. What Countries are Top Universities located in?
3. What factors are most important in ranking of the Top Universities?
4. What other indicators and measurements are important in University ranking and there values in the tom Universities?
5. How much the countries are spending in primary, secondary and tertiary education and ranking of the top countries?
6. Is there connection and relationship between the location of the top universities and the spending for education in those countries?