Gregor Gurski

Career Foundry Project 6

Sourcing Open Data

23.11.2023

World University Rankings

Source: The source of this data is from [Kaggle](). Which is formed of a community of data scientists and developers. It is the world's largest data science community with over one million registered users. Members can talk about development models, explore data sets, or network across 194 separate countries around the world. I will only focus on The Times Higher Education World University Ranking Data, which is widely regarded as one of the most influential and widely observed university measures.

Collection: The data was compiled and scraped from the times higher education website, so it should be dependable. They write on their website about Data Collection: "Institutions provide and sign off their institutional data for use in the rankings. On the rare occasions when a particular data point is not provided, we enter a low estimate between the average value of the indicators and the lowest value reported: the 25th percentile of the other indicators. By doing this, we avoid penalizing an institution too harshly with a "zero" value for data that it overlooks or does not provide, but we do not reward it for withholding them."

Contents: The Times Higher Education World University Rankings are global performance tables that judge research-intensive universities across all their core missions: teaching, research, knowledge transfer and international outlook. They use 13 performance indicators to provide comprehensive and balanced comparisons, trusted by students, academics, university leaders, industry, and governments.

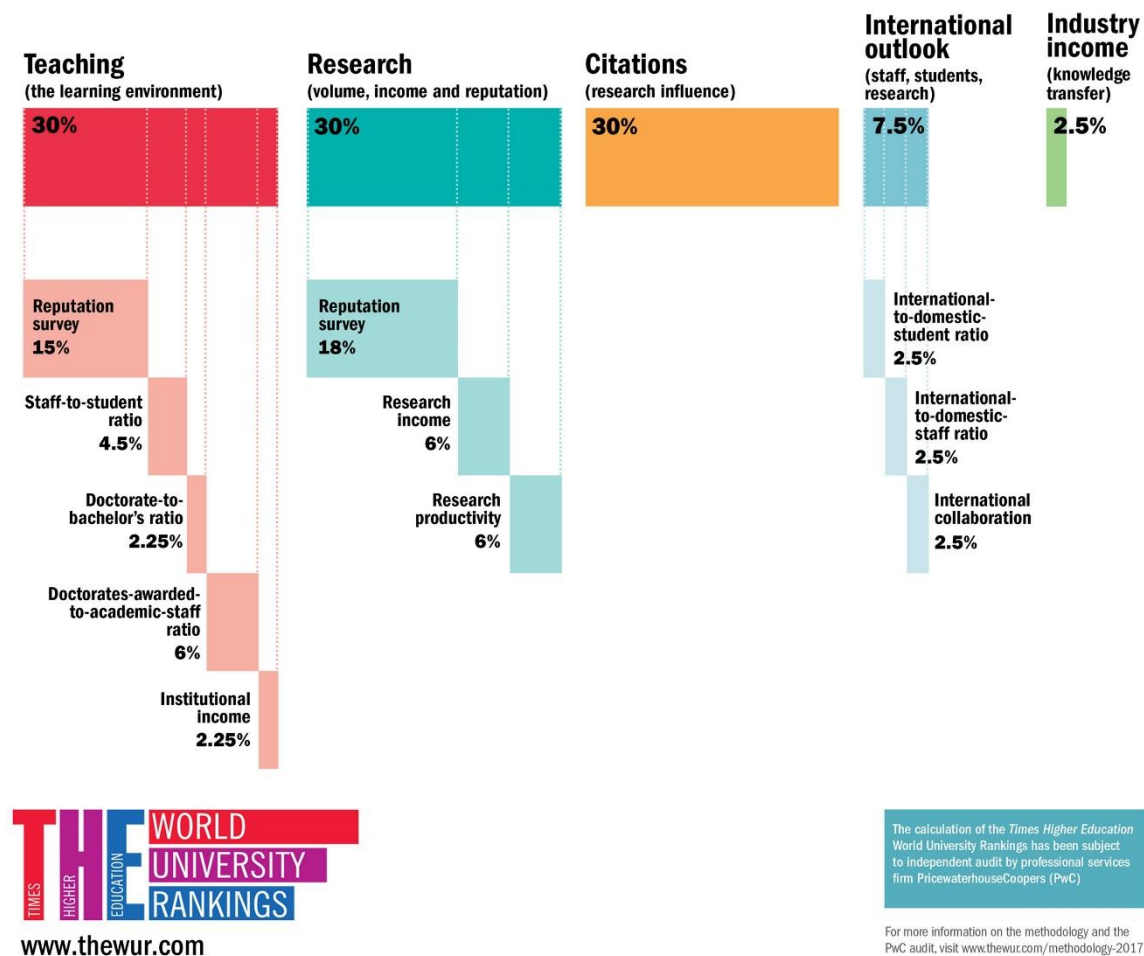The performance indicators are grouped into five areas:

Teaching (the learning environment)

Research (volume, income, and reputation)

Citations (research influence)

International outlook (staff, students, and research)

Industry income (knowledge transfer)

Teaching (the learning environment)	Research (volume, income and reputation)	Citations (research influence)	International outlook (staff, students, research)	Industry income (knowledge transfer)
30%	30%	30%	7.5%	2.5%

Teaching:
- Reputation survey 15%
- Staff-to-student ratio 4.5%
- Doctorate-to-bachelor's ratio 2.25%
- Doctorates-awarded-to-academic-staff ratio 6%
- Institutional income 2.25%

Research:
- Reputation survey 18%
- Research income 6%
- Research productivity 6%

International outlook:
- International-to-domestic-student ratio 2.5%
- International-to-domestic-staff ratio 2.5%
- International collaboration 2.5%

THE WORLD UNIVERSITY RANKINGS
TIMES HIGHER EDUCATION
www.thewur.com

The calculation of the *Times Higher Education* World University Rankings has been subject to independent audit by professional services firm PricewaterhouseCoopers (PwC)

For more information on the methodology and the PwC audit, visit www.thewur.com/methodology-2017

Limitations & Exclusions: Universities are excluded from the World University Rankings if they do not teach under-graduates or if their research output amounted to fewer than 1,000 articles between 2011 and 2015 (and a minimum of 150 a year). Universities can also be excluded if 80 per cent or more of their activity is exclusively in one of the eight subject areas.

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. The project is a little older than 3 years old but was found in the career foundry brief.

Ethics:

As this data is open-source I don't think there is an issue of data ethics.

Data Profile

Variable	Description	Time Variant/ Invariant	Structured/ Unstructured	Quantitative/Qualitative	Nominal/Ordinal /Discrete/Continuous
world_rank	Worldwide Rank of the University.	variant	structured	quantitative	discrete
university_name	Name of the University.	invariant	structured	Qualitative	nominal
country	Country of the University	invariant	structured	qualitative	nominal
teaching	Teaching score	variant	structured	quantitative	continuous
international	International score	variant	structured	quantitative	continuous
research	Research score	variant	structured	Quantitative	continuous
citations	Citations score	variant	structured	Quantitative	continuous
income	Income score	variant	structured	Quantitative	continuous
total_score	Total score	variant	structured	Quantitative	continuous
num_students	Number of students at the University	variant	structured	Quantitative	discrete
student_staff_ratio	The ration between the number of students divided by the number of staff members	variant	structured	Quantitative	continuous
international students	Percentage of international students at the University	variant	structured	Quantitative	discrete
female_male_ratio	The ratio between female students compared to male students	variant	structured	quantitative	continuous
year	The year of the ranking	variant	structured	quantitative	discrete

Questions to Explore

- How does my university fare against the world?
- Are there certain biases? For example, The Times Higher Education World University Ranking Data has been criticized for its commercialization and for undermining non-English-instructing institutions.
- Of all the universities in the world, which are the top 10?
- How are the performance indicators defined and measured?
- Why are certain categories weighted higher than others?
- Are these weightings reasonable?
- How many students attend the top 10 Universities?
- Which of the top 10 Universities has the highest/lowest amount of students?
- Are there more men or more women attending the top 10 Universities? What does this result mean in terms of gender equality?