

# Project THE – Times Higher Education University Ranking

Sergey Khrapak

## I. Data

### 1. Data Sourcing

For the analysis I chose THE (Times Higher Education) University Rankings 2016-2025 dataset. The dataset can be found at

<https://www.kaggle.com/datasets/raymondtoo/the-world-university-rankings-2016-2024>

The dataset provides a list of universities along with their rankings and various other metrics according to the Times Higher Education report.

Further details about Times Higher Education can be found at

<https://www.timeshighereducation.com/>

### 2. Data Content

The file contains 14 columns and include:

Rank, Name, Country, Student Population, Students to Staff Ratio, International Students, Female to Male Ratio, Overall Score, Teaching, Research Environment, Research Quality, Industry Impact, International Outlook, Year

The content of the dataset is semi obvious from the column names. The name and country column datatypes are string. Year is integer. The rest is expressed in numerical format. Note, however, that International students are expressed in % and the Female to Male ratio is expressed as xx:xx. As such the string datatype is used for these columns.

### 3. Data relevance

This dataset is of relevance for students searching for the best institutions that suit their learning preferences before making their application. The dataset provides an easy way to see how a university of interest ranks globally (or in its native country) and then

comparing it with another university. The dataset is also relevant for researcher, when choosing where to continue their research/teaching carrier. The dataset can be used by funding organization in various situations (e.g. considering individual student applications, research proposals, etc.). The dataset can provide a basis for important governmental decisions, e.g. spending more money on educational programs and supporting certain universities. My personal interest in this data is related to my over 25 years career in academia as well as working relations with several universities. I am also contributing personally to the THE ranking. So, I am looking forward to becoming more familiar with this dataset.

#### 4. Limitations and Ethics.

Among the limitations of this dataset is that it ranks the universities on a general basis, not considering their specifics. As a result, smaller universities, specializing in certain disciplines, can sometimes have lower ranking than they deserve. I can see no ethical issues related to the present dataset.

#### 5. Tentative questions

- How highly ranked universities are distributed over countries?
- Whether this distribution changes with time?
- What is an “average” university rank for each country and how to define it?
- Does rank correlates with the size of the university in terms of the number of students?
- How are Research Quality and Industry Impact correlated?
- What is the number of universities in top 100 for each country?
- Does this change with time?
- How Students to Staff Ratio and Research Quality are correlated?