

Report

Analysis of University Rankings

Julian BACKÉ
Tobias SALZER
Ajayvir SINGH

January 21, 2020

1 Objective

The objective of this project is to analyse, visualise and interpret data concerning university rankings. Certain characteristics of universities, such as their publication history, research output, number of students, budget, ... are considered for this task. Furthermore, the characteristics of countries and regions these universities are located in, are analysed to determine how those characteristics influence the rankings of said universities.

2 Workflow

1. Problem definition:

How do university rankings change over time? Which characteristics of universities contribute most to good rankings, or to large changes in the ranking position? Are there predictors for increases or decreases in the rankings?

2. Data acquisition:

- a) The main source of data for the rankings itself based on certain characteristics of the university: **World University Rankings**
- b) Although missing a substantial amount of data, the **Shanghai dataset** for world rankings was useful, because it contained data starting from 2005.
- c) Similarly, the **Times dataset** contained data not in the previous two datasets, allowing us to retrieve information about the number of students, ratio to staff etc.
- d) To gain information about expenditure, we found it necessary to use the **Expenditure dataset**.
- e) Since previous datasets only contained data about characteristics of the universities itself, the following data set contains data about a characteristic of the country itself: **Human Development Index**
- f) To analyse rankings over regions, a way to order countries by region was necessary. The following dataset of countries ordered by region was used: **Countries of the world**
- g) Another feature that was planned to be considered was a measure of corruption by country: **Corruption Perceptions Index**

3. Data loading:

The next step was to preload data into separate dataframes:

4. Preprocessing data:

One of the challenges of this analysis was to bring all the different datasets into a unified format. This included renaming certain columns and values, imputing missing values, removing unnecessary columns and stripping white-spaces.

a) Renaming:

One issue was different naming schemes for certain data. Analysing and renaming such values was imperative before merging data.

b) Missing data:

Similarly, we discovered that some datasets had missing data. Shanghai dataset contained a column, which had over 75% missing

values. Due this specific characteristic (score) being contained in another dataset already, it was decided to completely remove this column.

Our main dataset, the **World University Rankings**, seemed relative complete, with the exception of some missing values for *broad_impact* characteristic. By analysing how many rows were missing and the correlation to ranking of universities, we came to the conclusion, that multivariate imputation would be an optimal solution here, due to good correlation with other characteristics and only a small amount of rows being required to be imputed, leading to introduction of almost no bias. For instance, we see that *broad_impact* characteristic has a very good correlation with the rankings:

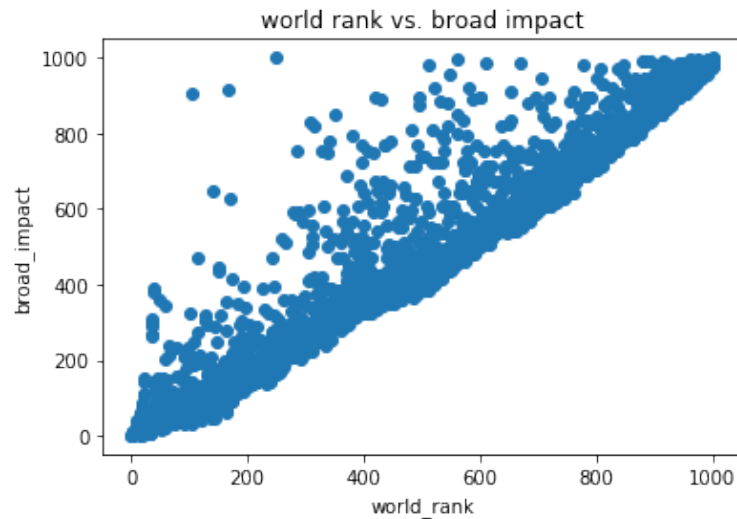


Figure 1: world rank vs broad impact.

c) ...

5. Data Exploration

6. Data Modelling

3 Data Exploration

3.1 Correlation Matrix for World University Rankings

We found it effective to construct a heatmap to get a general overview of how characteristics of universities correlate with their rankings and which correlate the most:

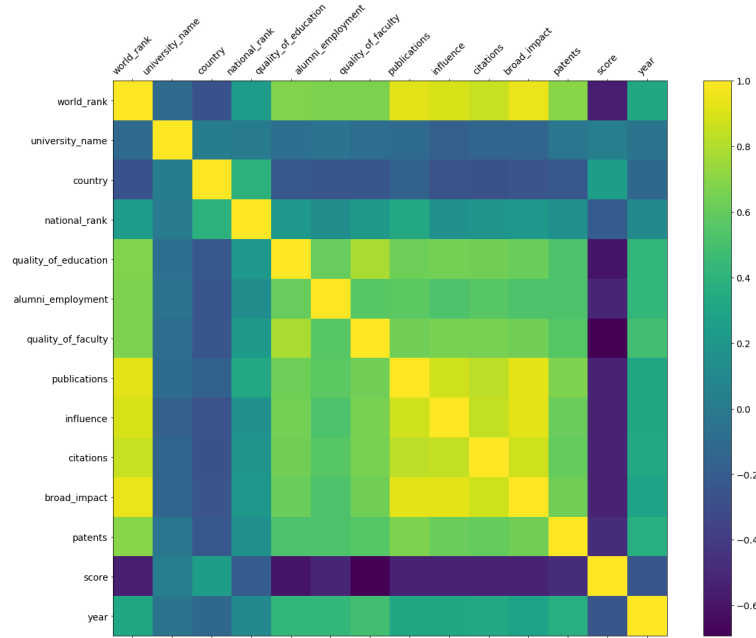


Figure 2: World University Rankings Heatmap.

As evident, while certain characteristics contribute less to a good rank, others, like *broad_impact*, have a surprisingly good correlation with the rank. It is evident, that imputing *broad_impact* could have introduced some bias to the data, but as shown in Figure 1, the correlation between these two characteristics was already good before.

4 ...

5 ...

6 Results and Conclusions

References

Smith, J. M. and Jones, A. B. (2012). *Chemistry*. Publisher, 7th edition.