Paper 1663-2014

SAS® Visual Analytics deliverers insights into the UK University League Tables – Working Paper

Authors: R J Self, S Berry, C Foyle and D Voorhis, University of Derby, UK

Abstract

Universities in the UK are now subject to League Table reporting by a range of providers. The criteria used by each League Table differ. Universities, their Faculties and individual Subject areas want to understand how the different tables are constructed and calculated and what is required in order to maximise their position in each league table in order to attract the best students to their institution, in order to maximise recruitment and student related income streams. The School of Computing and Maths at the University of Derby is developing the use SAS® Visual Analytics to analyse each league table to provide actionable insights as to actions that can be taken to improve their relative standing in the league tables and also to gain insights into feasible levels of targets relative to the peer groups of institutions. This paper outlines the approaches taken and some of the critical insights developed that will be of value to other Higher Education institutions in the UK and will suggest useful approaches that may be valuable in other countries.

Introduction

Universities in the UK are subject to comparison by three main League tables, The Complete University Guide, The Guardian and the Times / Sunday Times leagues. Each one has a different construction, based on differing criteria and weighting factors, together with different algorithms for consolidating the results from each criterion into the final league table position. Each provider publishes details of their own algorithms. At least two of the tables use the Z-Transform on most or all criteria which results in considerable difficulty in interpreting the overall relative positions at subject or institutional level.

In addition, there are several data sets collected via the Higher Education Statistics Agency (HESA) and the Higher Education Funding Council for England (HEFCE) which are published independently and are the source of much of the data included in the League Tables referred to above.

As a result, Universities and Schools find it very difficult to understand the “raw” data provided in the various tables. They want to be able to gain insights into the way that they can develop and improve the factors under their control, in order to be able to improve their comparative standing, as part of their organisational development strategies. The construction of the league tables, as published is found to be confusing and opaque.

This project is being undertaken to gain an understanding of the construction process of the league tables and to develop the necessary and feasible actionable insights that can be gained from each League table and from a broader consideration of all relevant League Tables. It is a

work-in-progress and this session will provide the latest status and achievements of the research and insights into how to maximise the value of SAS Visual Analytics during this process.

Data Sources

The three key independent league tables use different factors and weightings from each other and may also use a different set of factors in their Institutional level analysis from that used in the subject analysis. These are:-

- The Complete University Guide[1]
- The Times Good University Guide[2]
- The Guardian University Guide[3]

Additional sets of data are sourced from the UK Government agencies, HEFCE and HESA, such as the National Student Survey (from HESA) and the source data for a range of aspects such as staff numbers, student destinations (Employability factors), expenditure per student, etc. from HEFCE.

Definitions and Algorithms

Each of the three league table providers publishes details of their algorithms for calculating the final position of each institution in the Overall table and in the subject group tables. They include details of the relative weightings of each factor, together with the precise definition of each factor and its derivation.

Definitional Example

As an example, the Student Satisfaction factor is calculated as the mean score of the relevant students in two of the tables and as the percentage of students answering either "somewhat agree" or "strongly agree" in a five point Likert scale response in the other table. The latter interpretation ignores all the mid-point responses and ignores the impact of overall distribution. It also has significant impact on the managerial and academic consequences and potential actions, in that the response to the latter is to engage students in such a way that they only choose the top two response positions (potentially gaming the system), whereas the averaging algorithms may lead to a broader academic approach to address the more fundamental causes of the dissatisfaction.

In addition to this, some of the league tables publish the %ge Satisfied score (the top two Likert scale points) in the table but actually use the average value in the calculations.

Algorithm Example

Two of the League tables are constructed using the Z-transform to standardise the raw data into the normalised Z-space prior to the weighted summation into the overall score within the table and thence to the relative standing of the institution in the overall table. The other one defines a process (the S-Score) which appears similar to the Z-transform but applies a filter to outlier data points which exceed three sigma.

---

[1] To be found at http://www.thecompleteuniversityguide.co.uk/
[2] To be found at http://www.thetimes.co.uk/tto/public/gug/
[3] To be found at http://www.theguardian.com/education/universityguide

All three providers transform the Z-transformed factors into a defined space, often between 0 and 100 which is then the subject of the weighted summation process into the final total score which is then normalised into a defined space between 0 and 100 (applied to the top institution).

It is clear that this normalisation process is effective in identifying the relative positions of the institutions vis-à-vis each other. The process, however, makes it extremely hard to link the relative position of an institution to the raw (or processed) data scores that are provided in the tables and thence to the appropriate managerial and academic actions needed to be undertaken in order to improve the position of an institution.

First Steps

It is well know that re-purposing data collected for one purpose can pose problems (Yorke, 2011) particularly where large datasets are involved. These range from issues with missing data and incorrect data to the precise definitions of the parameters used in the models. As a result a range of analyses were undertaken to identify any potential issues that could be identified resulting from data cleanliness and the opacity of the algorithms on the outcome compared to the data.

Visual Exploration

As a result, the first step of the research was to explore the data visually, using scatter plots relating the individual league position of institutions versus each particular league table attribute separately. This identified a few instances where there was an indication of incomplete or incorrect data in easily identifiable outlier points which were far off the broad band of roughly correlated points. This is important if the outlier point is either one's own institution or a competitor or bench-mark institution.

Visual exploration also identified some parameters where there appeared to be sub-populations. This was indicated visually by what could be called a "kinked line" in the data with differing gradients, which appears to imply different weighting factors are used for the two sub-populations, but is likely to be an artefact of the various transforms applied to the data and also the substitutions.

Statistical Exploration

This approach is undertaken to identify the strengths and level of relationships between the displayed factors and the league position tables using multi-variate regression analysis.

A crucial result from this analysis is that some of the factors appear to have considerably greater impact on the overall league table position than would be expected from the specified weighting factors and vice-versa.

Next Steps

Having identified some of the issues with the data and the published algorithms that inhibit the development of the relevant managerial and academic strategies, the project is developing a range of tools in Base SAS and in SAS Visual Analytics that will provide a rapid analysis of each league table as soon as they are published and provide management and academics with the ability to identify critical insights for their own institution. These will include using both the SAS VA Explorer mode and SAS VA Dash Board.

All available sets of historical league table data have been obtained so that time-series analysis can be carried out, in order to evaluate achievable levels of improvement based on changes to various factors. In addition, the project will source as much of the HEFCE and HESA primary data as possible, again attempting to obtain as much of the historical data as possible.

Conclusions

The presentation will demonstrate the power of visual analytics in the exploration of publicly available data to identify critical insights into suitable managerial and academic actions to improve institutional positions in UK Higher Education University League Tables.

The approaches presented should be transferrable to other contexts where public transparency data is made available to citizens in order to improve decision making.

Bibliography

Yorke, M. (2011) Analysing existing datasets: some considerations arising from practical experience, **International Journal of Research & Method in Education**, 34:3, 255-267, DOI: 10.1080/1743727X.2011.609549

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Richard J Self

Organization: School of Computing and Maths, University of Derby

Address: Kedleston Road,

City, State ZIP: Derby, DE22 1GB, England

Work Phone: +44 1332 591150

Email: r.j.self@derby.ac.uk

Web: Home Page: http://computing.derby.ac.uk/wordpress/people-2/richard-j-self/

LinkedIn: http://uk.linkedin.com/pub/richard-self/a/93a/829