Title: World College Ranking

Course: DSC 520

Author: Blandon Lee

Date: 03/01/2022

Introduction

Seeking higher education can be a daunting task and one that should be thoroughly researched prior to committing. Determining what you want to do is the first hurdle one must cross. However, when that has been achieved the next one is to decide which school best suits your needs and wants. There are many reasons individuals choose a school such as family, political affiliation, or location. However, often it just comes down to what school is the best. There are numerous national and international rankings available such as the Center for World University Rankings than make this decision easier. However, does it matter which school you really attend? Is the education attained from a public university, private university, or an ivy league school so vastly different?

Summarize the problem statement you addressed.

The focus of this project was to see if quality education can only be garnered from top-ranked universities. In an attempt to answer this, I created a Null and an alternate hypothesis.

Null: The quality of education has no effect on world ranking

Alternate: The quality of education does have a strong connection to world rankings

Based on the analysis conducted the determination was made to reject the Null hypothesis and accept the alternate hypothesis. This indicates that there is indeed a strong connection between them. There were some substantial factors deemed to be significate that contributed to the world rankings such as graduation Rates and student retention. However, other major contributors found throughout the research was quality of education, citations, and the amount/type of research conducted by the schools.

Analyzing the different variables, it was found that citation and world rankings and quality of education and quality of the school's facilities returned strong correlations .94 and .76 respectively. Also, the

p_values for these two models were near zero and this made accepting the alternative hypothesis much easier.

Summarize how you addressed this problem statement (the data used and the methodology employed, including a recommendation for a model that could be implemented).

To address the problem statement, I collected four datasets that were found on Kaggle.com and proceeded to clean them in preparation for the project. I was then able to merge some of them to make the data a bit easier to manipulate. Utilizing regressions (linear and multiple) helped determine which variables had the greatest effect on the problem statement . Through the correlation tests and calculations of p-values and other summary statistics additional information because apparent. Creating models helped identify variables that had a significant impact on rankings.

Dataset 1: CWUR.2020

```
str(CWUR.2020)
```

'data.frame': 2000 obs. of 9 variables: $ world_rank : int 1 2 3 4 5 6 7 8 9 10 ... $ institution : chr "Harvard University" "Massachusetts Institute of Tech”.. $ country : chr "USA" "USA" "USA" "United Kingdom" ... $ national_rank : int 1 2 3 1 2 4 5 6 7 8 ... $ quality_of_education: chr "2" "1" "9" "4" ... $ alumni_employment : chr "1" "10" "3" "19" ... $ quality_of_faculty : chr "1" "2" "3" "5" ... $ research_performance: chr "1" "5" "2" "11" ... $ score : num 100 96.7 95.2 94.1 93.3 92.6 92 91.6 91.1 90.7 ...

Dataset 2: timesData

```
str(timesData)
```

'data.frame': 2603 obs. of 14 variables: $ world_rank : chr "1" "2" "3" "4" ... $ institution : chr "Harvard University" "California Institute of Tech"... $ country : chr "United States of America" "United States of America"... $ teaching : num 99.7 97.7 97.8 98.3 90.9 90.5 88.2 84.2 89.2 92.1 ... $ International_Outlook : chr "72.4" "54.6" "82.3" "29.5" ... $ research : num 98.7 98 91.4 98.1 95.4 94.1 93.9 99.3 94.5 89.7 ... $ citations : num 98.8 99.9 99.9 99.2 99.9 94 95.1 97.8 88.3 91.5 ... $ income : chr "34.5" "83.7" "87.5" "64.3" ... $ score : chr "96.1" "96.0" "95.6" "94.3" ... $ num_students : chr "20,152" "2,243" "11,074" "15,596" ... $ student_staff_ratio : num 8.9 6.9 9 7.8 8.4 11.8 11.6 16.4 11.7 4.4 ... $ international_students: chr "25%" "27%" "33%" "22%" ... $ female_male_ratio : chr "" "33 : 67" "37 : 63" "42 : 58" ... $ year : int 2011 2011 2011 2011 2011 2011 2011 2011 2011 2011 ...

Dataset 3: University_world_rank_Data

```
str(University_world_rank_Data)
```

'data.frame': 2200 obs. of 13 variables: $ world_rank : int 1 2 3 4 5 6 7 8 9 10 ... $ institution : chr "Harvard University" "Massachusetts Institute of Tech"... $ country : chr "USA" "USA" "USA" "United Kingdom" ... $ national_rank : int 1 2 3 1 4 5 2 6 7 8 ... $ quality_of_education: int 7 9 17 10 2 8 13 14 23 16 ... $ alumni_employment : int 9 17 11 24 29 14 28 31 21 52 ... $ quality_of_faculty : int 1 3 5 4 7 2 9

12 10 6 ... $ publications : int 1 12 4 16 37 53 15 14 13 6 ... $ influence : int 1 4 2 16 22 33 13 6 12 5 ... $ citations : int 1 4 2 11 22 26 19 15 14 3 ... $ patents : int 5 1 15 50 18 101 26 66 5 16 ... $ score : num 100 91.7 89.5 86.2 85.2 ... $ year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...

Dataset 4: world.university.ranking.2020

str(world.university.rankning.2020)

'data.frame': 1396 obs. of 15 variables: $ world_rank : int 1 2 3 4 5 6 7 8 9 10 ... $ institution : chr "University of Oxford" "California Institute of Tech"... $ country : chr "United Kingdom" "United States" "United Kingdom"... $ Number_students : chr "20,664" "2,240" "18,978" "16,135" ... $ Numb_students_per_Staff: num 11.2 6.4 10.9 7.3 8.6 8.1 9.2 5.4 5.7 11.7 ... $ international_students : chr "41%" "30%" "37%" "23%" ... $ Percentage_Female : chr "46%" "34%" "47%" "43%" ... $ Percentage_Male : chr "54%" "66%" "53%" "57%" ... $ teaching : num 90.5 92.1 91.4 92.8 90.5 90.3 89.2 92 89.1 84.5 ... $ research : num 99.6 97.2 98.7 96.4 92.4 96.3 98.6 94.8 91.4 87.6 ... $ citations : num 98.4 97.9 95.8 99.9 99.5 98.8 99.1 97.3 96.7 97 ... $ Industry_Income : num 65.5 88 59.3 66.2 86.9 58.6 47.3 52.4 52.7 69.9 ...

Summarize the interesting insights that your analysis provided.

After completing the analysis, it was determined that the variables that would be expected to provide an extensive correlation had little to no effect on the actual university rankings. Variables like international students or international outlook did not affect the rankings at all. The most interesting thing from this project was that the Null hypothesis was rejected. One would think that quality of education would be one if not the one variable that had the most impact on rankings. However, according to the analysis it was not a real contributor.

Summarize the implications to the consumer (target audience) of your analysis.

Many people seek locations, facility quality, affordability, scholarships, and so on. However, it does not seem that quality of education is high on the list and has no impact on the school ranking. According to the analysis it is possible to attend a top 50-100 school and still receive quality education (see figure 1. below). If someone is looking at for things such as research and teaching at a university then it is good to note that these attributes have little effect on the overall ranking (see figure 2. below).

Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.

The data collected did present some roadblocks that could have been key to greater insights into the project and better answered the question. Having information from other countries would have forwarded the opportunity to test the variables against that information. It would have been interesting to see how each country compares to one another. Having a way to assign a number to the country would offer additional significance to the project.

Images

Figure 1. World Ranking for Quality of Education

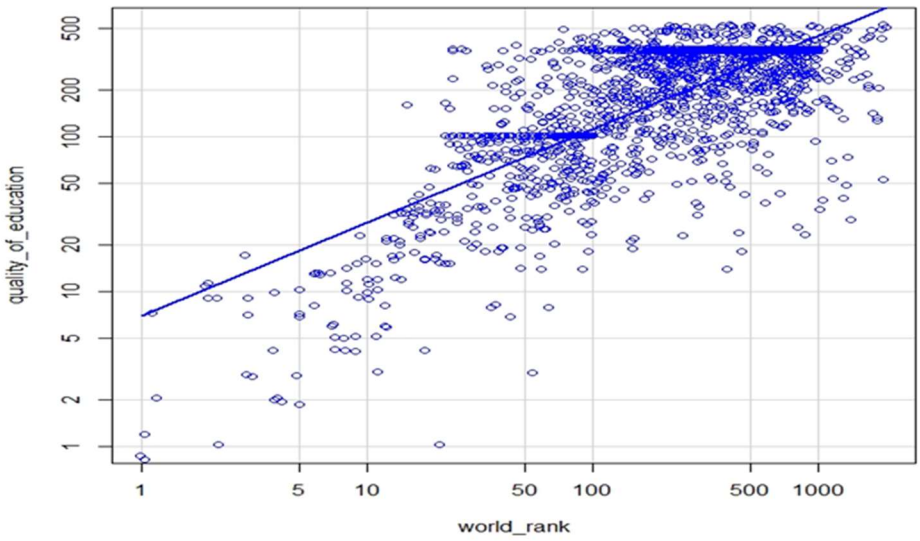


Figure 2. World Ranking for Teaching and Research

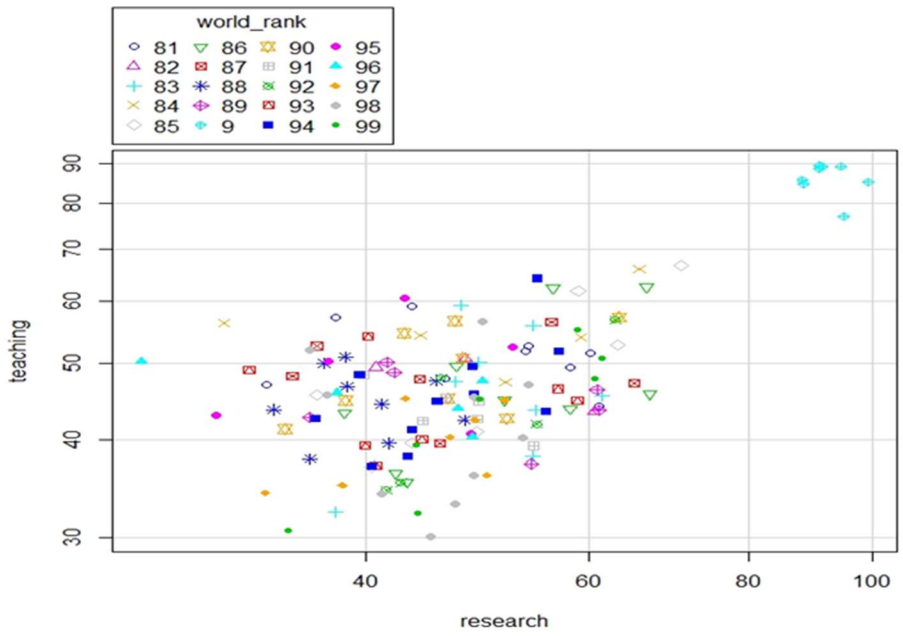


Figure 3. Research Ranking (bins = 40)

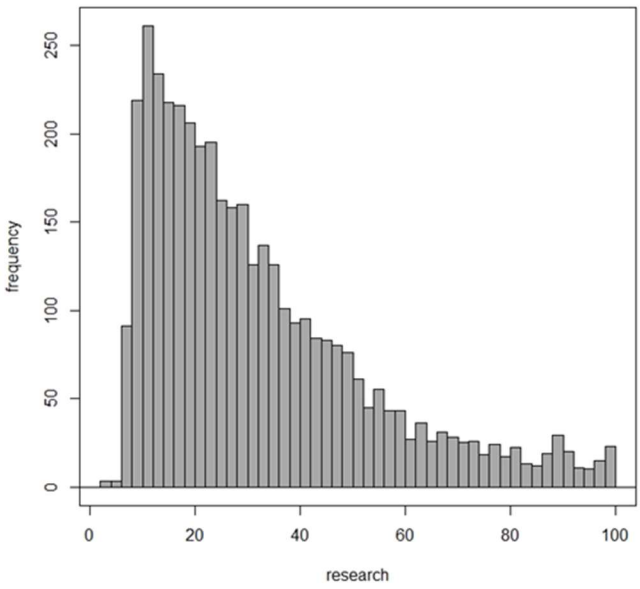
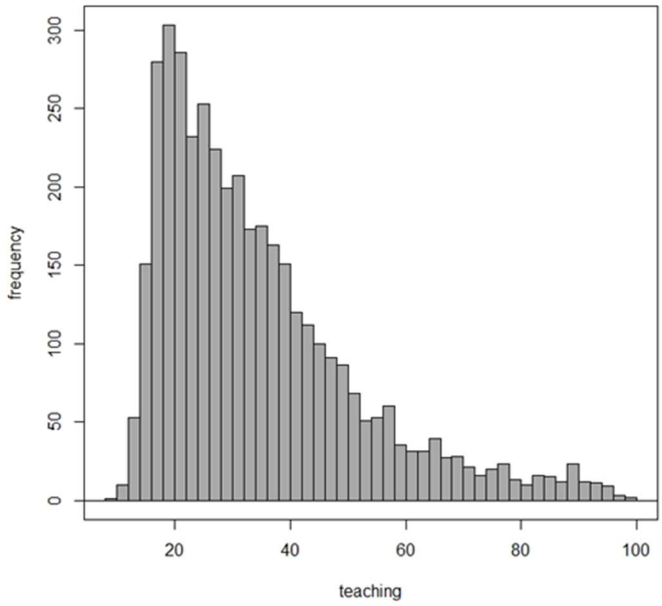


Figure 4. Teaching Ranking (bins = 40)



Concluding Remarks

The goal of this project was to analyze numerous variables from the datasets found for this project and determine what constituted university world rankings. After doing so it began to show top schools do provide the best education. However, the initial research did indicate that quality of education has little to no impact on the rankings. Variables such as citations, male to female ratio, student to teacher ratio, and research were noncontributors in establishing rankings.