**Final Project Report**

Eric Li, Yu Ding

After the theme of the world became peace and development, education has been regarded as one of the most vital part of a country. As the elementary stage of the higher education, universities around the world are ranked by many different organizations worldwide by their own standards, like QS, US News, and so on. These rankings have earned a lot attentions, and millions of students are dreaming about attending the best of them. This project will focus on the top 30 universities around the world ranked by Quacquarelli Symonds (QS), which is a British company that dedicates to analyze the educational institute around the world.

Overall, QS company ranks the universities and colleges each year based on six aspects: International student ratio, international faculty ratio, faculty to student ratio, citations per faculty, academic reputation, and employer reputations. QS will give each university scores based on their own evaluations and finalize to an overall score. All participated institution will be ranked based on the final score. Officially, QS publishes that the final score is computed based on the following standards:

Academic Reputation: 40%

Faculty/Student Ratio: 20%

Employer Reputation: 10%

Citations per Faculty: 20%

International Student Ratio: 5%

International Faculty Ratio: 5%

However, after preliminary calculation, the overall scores do not equal to the sum of the product of the weight of each aspect and their scores. Then, how does get the overall score starts to interest people. Thus, this project will commit to find out the rule of top university ranking or at least find out the most significant factor that affects overall score of top 30 university.

To attain the raw data, we copy the most up to date information from QS website about top 100 university worldwide 2021. Then, six scatter plots were drawn for us to get a big picture of the relationship between six evaluated aspects and the overall scores. We organized these plots as a dashboard with interaction with users.

To begin with, even though the overall score is not exactly the sum the product, the numbers are relatively close. Therefore, we guessed that there may be a possible base score for the ranking, or the published weights of each aspect are just a set of approximation. Then, we decide to use the linear regression method to simulate the model. Based on the dataset, we get: . Purely based on the model, the international student ratio seems to be the most significant factor due to its largest absolute value of coefficient, but it only weights 5% according to the QS official report. Moreover, the coefficients of faculty student ratio, academic reputation, and employer reputation are negative, which means that if the university has higher scores on these fields, it will result in lower overall scores. This is definitely not a reasonable and desired outcome.

After linear regression model, we realized it is hard to draw a definite conclusion based on these data about how QS actually ranks the university. Then, we decide to find the most correlated factor with the overall score. The correlation coefficients between overall score all six aspects are calculated and drawn in a heatmap. The map shows the faculty student ratio has the largest correlation coefficient, 0.54, which indicates the strongest correlation with the overall score. However, faculty student ratio only occupies 20% in the official report. After browsing the information of the university that the project emphasized, the possible explanation of this contradiction is that only top 30 universities are used. These educational institutes enjoy extraordinary reputation all around world. Except for some special cases, like the international student ratio of the Stanford University, the scores of all six aspects are comparatively close for these university. For academic reputation, the average value is 96.86. This actively elucidates that for top 30 university around the world, the reputation cannot be the most essential factors, but this will not deny the importance of the academic reputation for a university generally. We believe expanding the dataset to top 100 or even top 500 university will reveal the fact that much closer to the QS’s report.

The entire project is uploaded to [GitHub](https://github.com/zewen0531/MA346-Final-Project) and [Deepnotes](https://deepnote.com/@zewen-li/Final-Project-m3ftzP4XQJaGkjPAP5ZY7Q), so please feel free to access this information whenever you need.