

University World Ranking Analysis

PROJECT GOALS

To compare and contrast three distinguished university ranking systems by exploring various factors used to determine a university's ranking and to determine if specific factors are good indicators for determining if a university is top-ranking.

DATA SETS

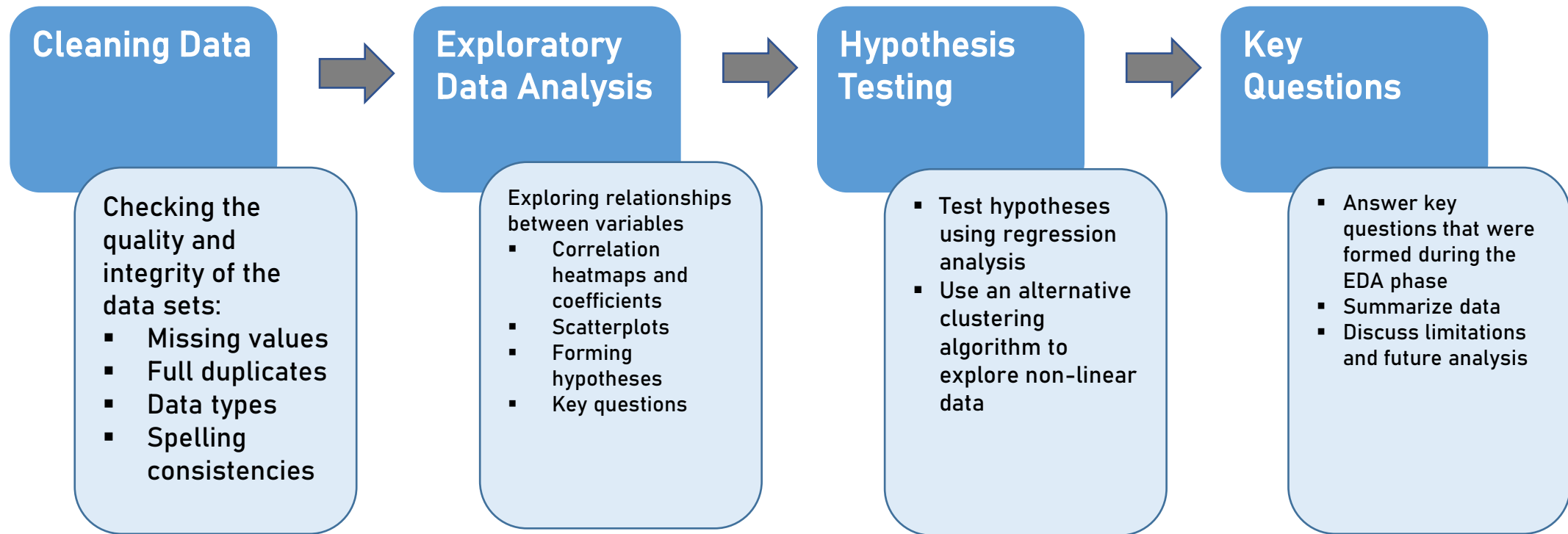
There were three ranking systems used: Center for World University Rankings (CWUR), Academic Ranking of World Universities (ARWU), and the Times Higher Education (Times). Each data set had a variety of factors/variables used for ranking universities.

DATA SOURCE and ANALYSIS

All three data sets were downloaded from the open-source resource Kaggle*. The following analyses were used in this achievement: correlation heatmaps and coefficients, scatterplots, regression analysis, cluster analysis, spatial analysis, and time series analysis.

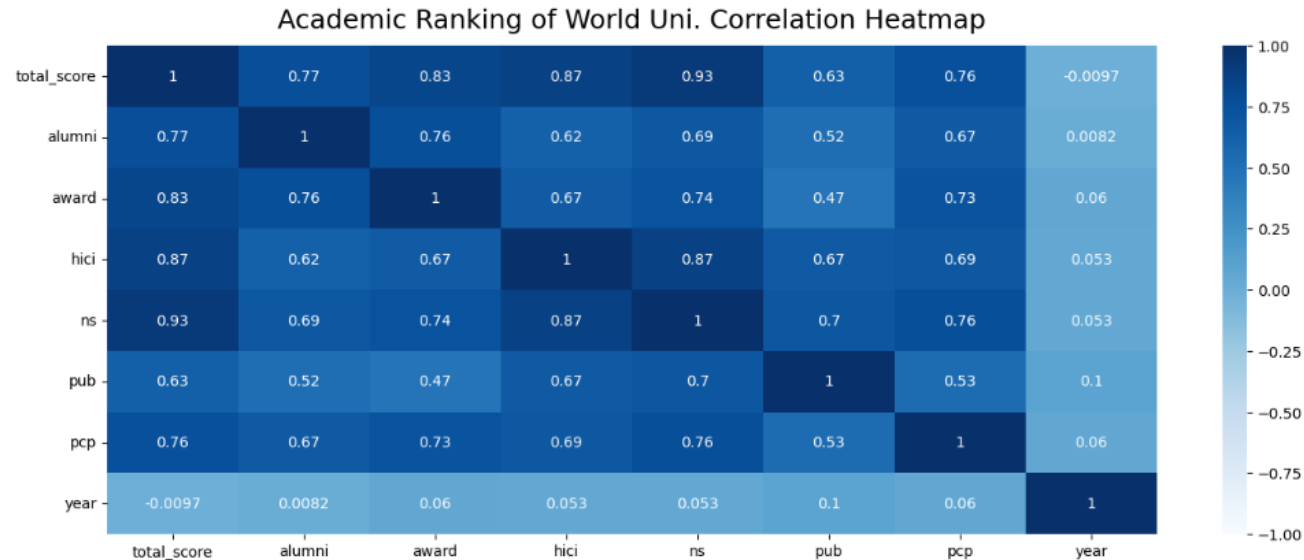
*<https://www.kaggle.com/datasets/mylesoneill/world-university-rankings>

University World Ranking Analysis



University World Ranking Analysis

EDA: Correlation heatmaps and coefficients



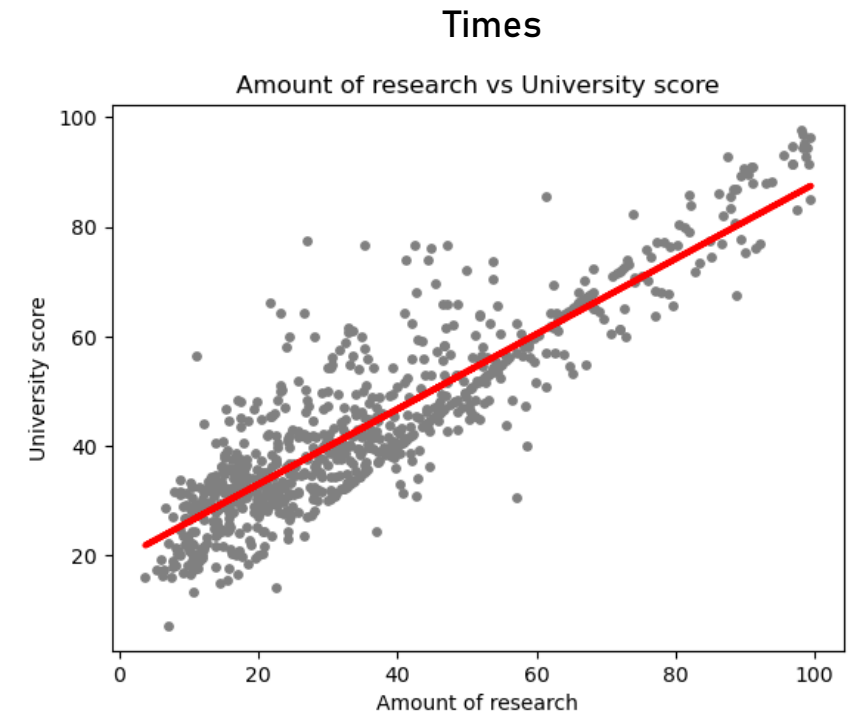
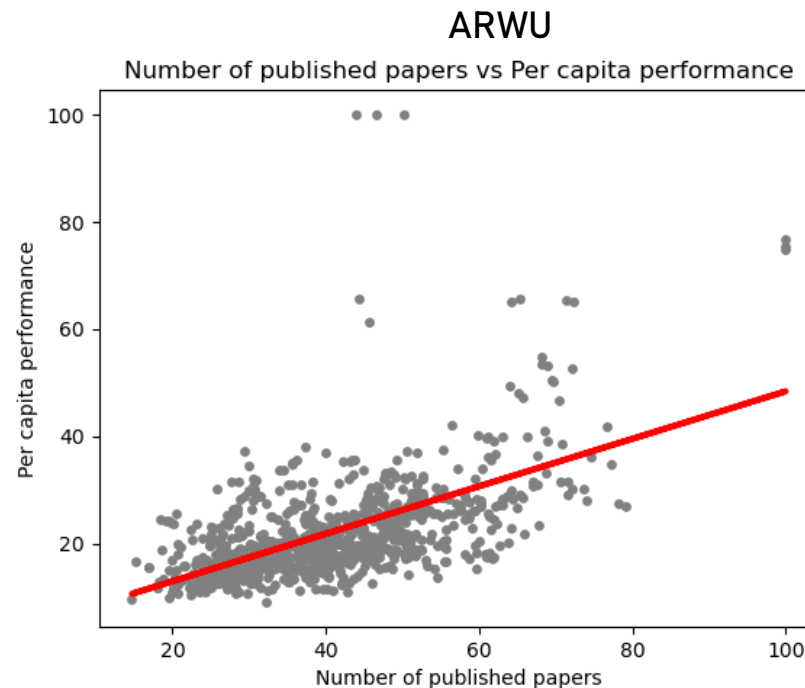
The correlation coefficients closest to 1 indicate a possible strong, positive, linear relationship. For example, the total score and the “ns” variable (which refers to the number of published articles in the Nature and Science journal), have a correlation coefficient of 0.93. If this coefficient is accurate in predicting linearity, then this means that as the number of published articles in the Nature and Science journal increase, the total score will also increase. Similarly, other variables such as “award” and “hici”, also had strong, positive correlations with the total score.

To test and see if this was a true linear relationship, a regression analysis was used.

University World Ranking Analysis

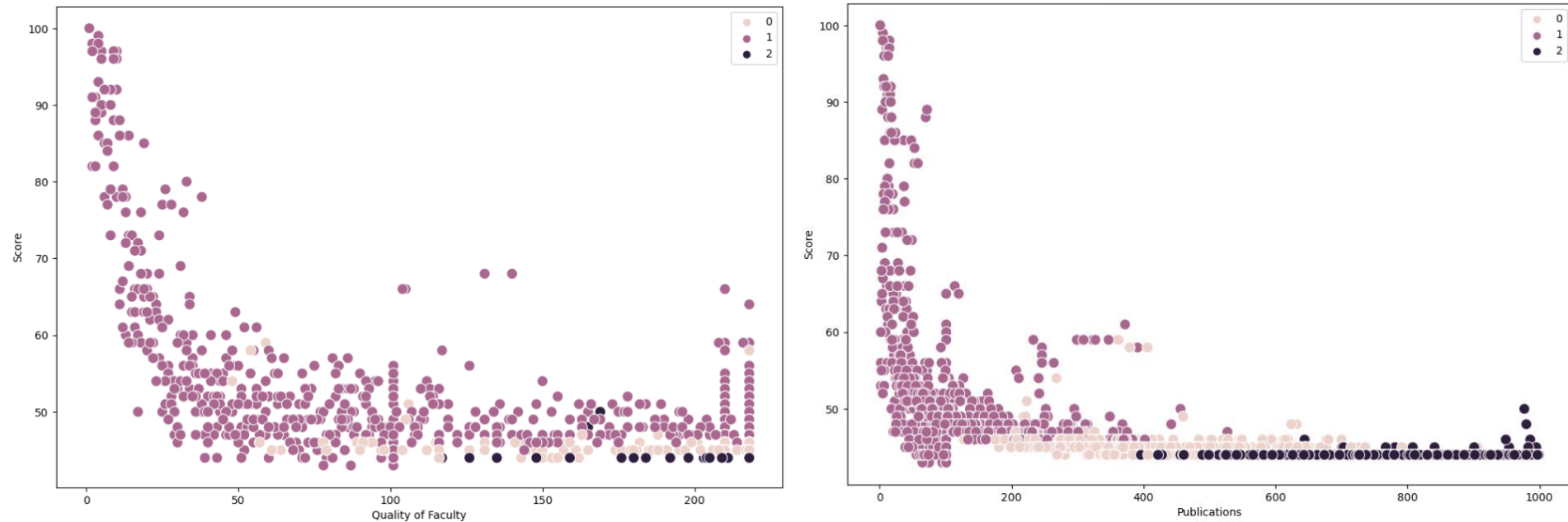
Regression Analysis

- A regression analysis was used to determine if the variables of interest were linearly related and to further test the hypothesis:
 - Ho: If the university has a high research output, then the university score will also be high, resulting in a top world rank.
- Times
 - The data follows a general positive, upward trend and the data points do not fit tightly to the regression line
 - The R2 score is 0.796 (or 80%) indicating that about 80% of the variance in the university score can be explained by the amount of research.
- ARWU
 - The linear regression model did not fit as well as it did for the Times variables.
 - The R2 score is 0.322 (or 32%), indicating that 32% of the per capita performance score can be explained by the number of published papers.
- The mean squared error for both of these regression model examples was around 62, which is extremely high for this data. This error indicates that the data points are very far from the regression line.



University World Ranking Analysis

Cluster Analysis



A cluster analysis was used to describe the CWUR variable relationships because of the non-linearity of the scatterplots. This analysis provided insights into how the data was organized. For example, after the number of publications exceeds ~150, the university score stays approximately the same (or plateaus). Similarly, after the quality of faculty exceeds ~50, majority of universities will have scores less than 60. Additionally, the cluster analysis further emphasized that the variables explored were not necessarily the best indicators for a top-ranking university. For example, a university with a score of 100, had 0 publications.

University World Ranking Analysis

Key Questions

- 1) **Are the top 10 universities the same for each ranking system?**
 - a) There is some overlap of universities when looking at the top 5, 10, and 15 for each ranking system.
 - b) For example, Harvard University, Stanford University, University of Oxford, and Princeton University are present in the top 10 average university scores for each ranking system (CWUR, ARWU, and Times).
- 2) **Do all ranking systems include the same countries?**
 - a) Each ranking system had a different number of countries.
 - b) CWUR = 59 countries
 - c) ARWU = 42 countries
 - d) Times = 70 countries
- 3) **How does the university score change over time?**
 - a) The ARWU data was used to explore the change over time because this data set provided data circa 2005 to 2015.
 - b) From 2005 to 2011, the average university score did not change. After 2011, the university score began to increase and peaked in 2013.
 - c) After 2013, the university score began to decrease.

University World Ranking Analysis

Final Project Deliverable

A [GitHub repository](#) was created to showcase the scripts created in Python. A [tableau storyboard](#) was also created to present the findings of this analysis. Lastly, an Excel workbook was created to summarize the data profile, sources, variables, and key questions.

Summary of Results

I fail to reject the null hypothesis. Depending on the ranking system, a high research output is associated with a high university score. Both the ARWU and Times ranking systems had moderate to strong linear relationships. Despite the CWUR ranking system having a weak linear relationship, all three ranking systems had similar top universities present in the top ranks.

Further Analysis

I think conducting an analysis that considers the actual weights for each variable would be interesting. Each ranking system has a different methodology, and each variable will contribute to the overall score/world ranking by different amounts.

Additionally, conducting an analysis on data that is more up-to-date would also be beneficial and more accurate.