University Ranking Analysis

Asmita Bamma, Dikshya Upreti, Alisha Baral

Anderson College of Business and Computing, Regis University

MSDS 670: Data Visualization

John Koenig

October 13, 2022

Abstract

This project utilizes the worldwide university ranking dataset that was extracted from Kaggle. The data is by The Center for World University Rankings (CWUR) which has been providing policy advice, strategic insights and consulting services to governments and universities to improve educational and research outcomes. This dataset carries information on various universities globally where they are ranked in several aspects like research, faculty, employability, education, and more. This week's project would be to analyze the factors that affect the rankings of universities around the world. This research would help students looking for universities and researchers to make their decisions based on the ranking of the university depending on different factors.

What is Data Visualization?

Data visualization is the presentation of data in an illustrated or graphical format. It enables decision-makers to see analytics presented visually, so it grasps difficult concepts or identifies new patterns. With the help of interactive visualization, analyzing data could be further by using different techniques to make graphs for detailed, interactive visuals across teams. The visuals created of the complex data is easier to understand the data relationships and data-driven insights.

As the amount of data collected today is more than ever, it is very important to have accessible ways to view and understand data. Businesses and employees at any level need to understand the impact of any trend that might affect their company's output. To bring the best out of any data, it is important to use the right visualization tool. Microsoft Excel is a great basic visualization tool along with Python's graphing library, Tableau, and Dashboard.

Among these platforms, Python offers several plotting libraries, namely Matplotlib, Seaborn, and many other packages. This allows python to create explanatory, customizable, and appealing plots which will help to present data in the most effortless way.

The two libraries that are used as python libraries are Matplotlib and Seaborn for visualization. It enables users to generate visualizations like histograms, scatterplots, bar charts, and much more.

About the data

For our final project, we will be using the World University Ranking data from the year 2022-2023 in python to make the analysis of the data. This data is published by the Center of World University Ranking(CWUR) which assesses the quality of education, alumni employment, quality of faculty, and research performance.

GLOBAL 2000 LIST BY THE CENTER FOR WORLD UNIVERSITY RANKINGS

Institution	Location	National Rank	Education Rank	Employability Rank	Facul
-	-	-	-	-	-
2000 unique values	USA 17% China 15% Other (1363) 68%	1 335	- 79% 1 0% Other (429) 21%	- 50% 1330 0% Other (997) 50%	- 102 Other (2
Harvard University	USA	1	1	1	1
Massachusetts Institute of Technology	USA	2	4	12	2
Stanford University	USA	3	11	4	3
University of Cambridge	United Kingdom	1	3	25	4

We will be using Matplotlib and Seaborn which acts as the backbone of data visualization through Python. They are libraries used for plotting and excelling graphs with the help of other libraries like Numpy and Pandas.

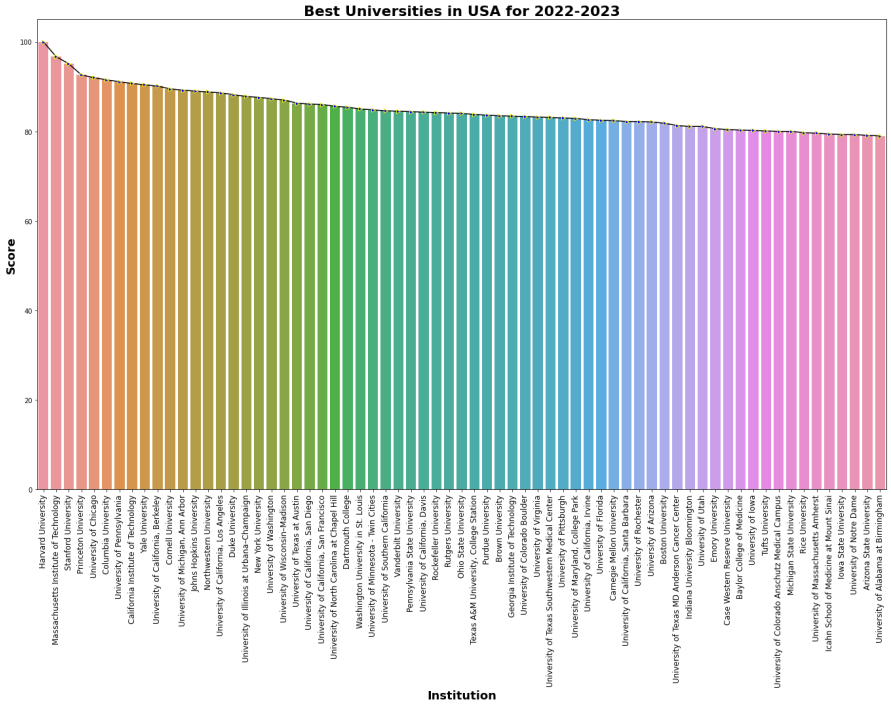
The university data we are using has 2000 rows and 9 columns. It is big data so we will start by performing some EDA before building our plots. This process will ensure that the results produced will be valid and we will be using a cleaner dataset for the analysis. The first step

would be to clean our data and replace the null value with 0 so it does not affect our analysis as the value will be filled rather than empty rows or columns.

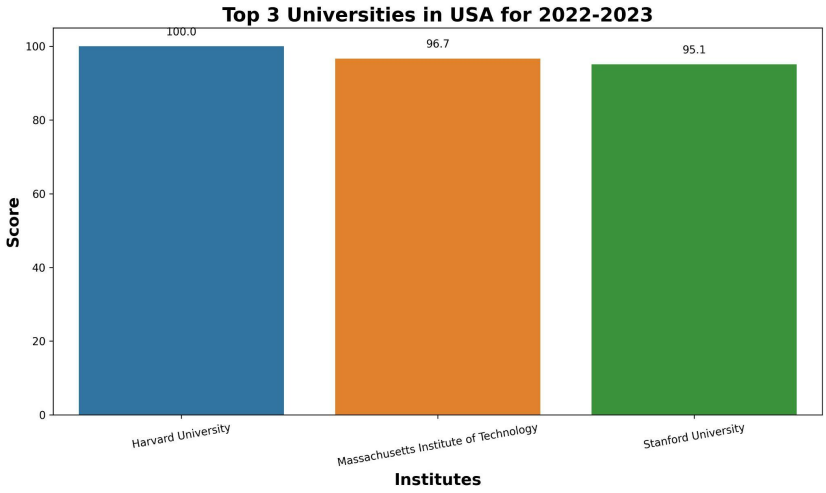
Creating our plots

For a better analysis of the data, we will be making some plots in python using Seaborn and Matplotlib. Below is a total of 4 different types of graphs and the analysis of the pattern of the graph.

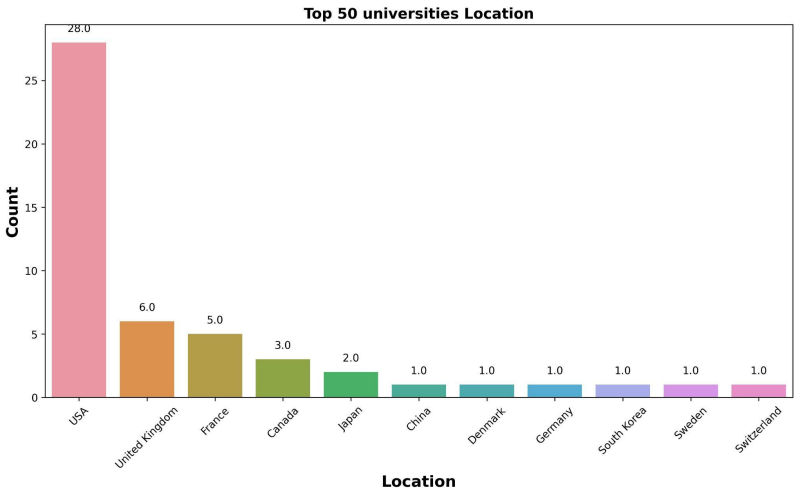
Our first graph will be a bar plot showing the top 50 ranked universities in the USA. The score of the university is 80 and above which means they are at the top level and very competitive among themselves. Harvard University is the highest ranked on the list and has a perfect score of 100. The cooler to warmer color in this graph represents the highest rank to the lowest.



Secondly, we will plot another bar chart showing only the top 3 universities in the USA for the year 2022-2023. This will give us a better visualization of the exact score and the difference. Three colors represent three different universities. Harvard with a perfect score of 100 is in blue, MIT with 96.7 is in orange and Stanford with 95.1 is in green.

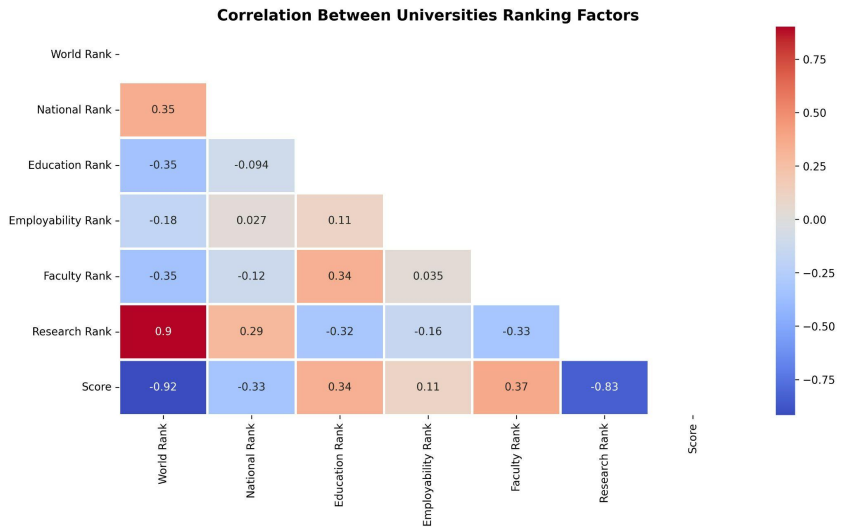


This data has university data from around the world. To know the location and the count of universities in that location, we will use a bar plot to plot the graph.

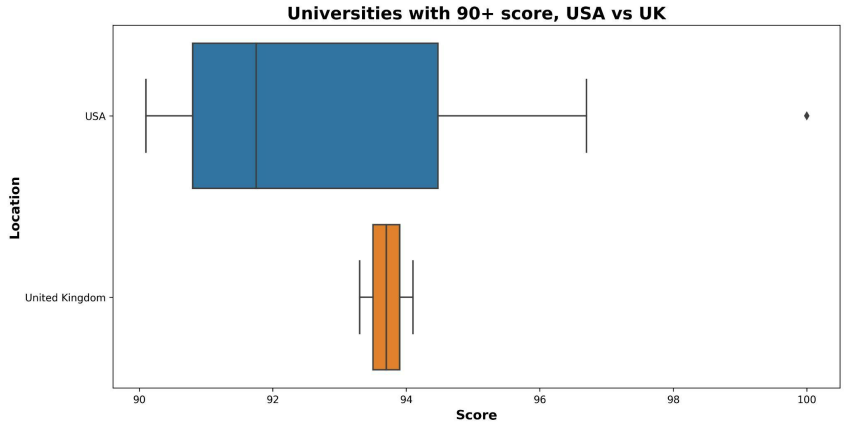


In the above plot, we could see that the USA in pink has the highest number with a count of 28, and the second is the United Kingdom with a count of 6. The difference as we could see between the highest and the second highest is nearly 5 times. We could say that the USA has the most number of great universities. Other countries in the plot lie 5 or below.

Next, we will be plotting a heatmap, where we will see the correlation between universities and their ranking factors. Correlation is positive if the value of the correlation is 1 or close to 1. In the below figure, we could see that Research Rank and World Rank have the highest correlation for worldwide university ranking with a z score of 0.9. The other closest but very minimal correlation would be National Rank to World Rank followed by Score to Faculty Ranking.



Lastly, we have a boxplot to show the countries with 90+ scores.



According to this plot, there are only two countries with universities that score above 90. Blue represents the USA and orange represents the UK. Again, the USA has the highest number of over 90 scoring universities. If we look at the plot we could see an outlier at a score of 100 which is the highest. The median of the USA is somewhere around 91.5 whereas the lowest value is 90.5. On the other hand, the highest and the lowest value for the UK is very close to each other which is why the plot is very small and congested. Everything lies between 93.5(the lowest value) and 94.3(the highest value) where 93.8 could be the median.

Conclusion

The purpose of this analysis was to analyze the factors that affect the overall ranking of the universities. Research rank has the highest correlation to world rankings which means that research is one factor that affects the rankings of Universities. According to our analysis, the number one university in the whole world is Harvard with the highest score of 100. Secondly, the USA is the country among all the countries to have the highest number of most high ranked

universities followed by the UK. This could bring us to the conclusion that the USA would be the best destination for students to get the best education and for research purposes.

References

Data visualization: What it is and why it matters. SAS. (n.d.). Retrieved October 16, 2022, from https://www.sas.com/en_us/insights/big-data/data-visualization.html

Simplilearn. (2022, August 29). *Data Visualization in Python: Overview, libraries & graphs: Simplilearn*. Simplilearn.com. Retrieved October 16, 2022, from https://www.simplilearn.com/tutorials/python-tutorial/data-visualization-in-python

Chauhan, A. (2022, September 30). *World University Ranking 2022 - 2023*. Kaggle. Retrieved October 17, 2022, from https://www.kaggle.com/datasets/whenamancodes/world-university-ranking-2022-2023

Sasaki Tetsuya. (October,2022). *World Universities Clustering & PCA Analysis*. Retrieved from https://www.kaggle.com/code/sasakitetsuya/world-universities-clustering-pca-analysis

Patki Saumit. (October,2022). *World University Ranking Regression.* Retrieved from https://www.kaggle.com/code/saumitgp/world-university-ranking-regression#EDA

Jatin.(June,2022). *Data Analysis world's Best Universities.* Retrieved from https://www.kaggle.com/code/jkanthony/data-analysis-world-s-best-universities