World University Ranking

Info – 5709 Data Visualization and Communication

Name: Tejesh Bathula

EUID: 11610736

Introduction:

"Is the Ranking System Fair?" An assessment using the World University Rankings Dataset**.** World University Rankings are yearly rankings of universities worldwide from years between 2012 and 2015 based on a variety of indicators and criteria such as academic reputation, research output, quality of education, quality of faculty, scores of universities, and international diversity. This research provides an overview of the World University Ranking dataset.

Background:

The World University Rankings project is a global endeavor that ranks institutions based on academic achievement, reputation, and research output. Many organizations, including Times Higher Education, QS World University Rankings, and Academic Ranking of World Universities, collaborated on this study. To establish a complete rating of institutions globally, the project collects data from a variety of sources, including academic surveys, employer surveys, citation indices, and research outputs, among others. Students, professors, and institutions use the rankings to judge the quality of education and research offered by various universities and to make educated judgments. This dataset may be used by researchers, analysts, and academics to obtain insights into the global higher education scene, analyze changes in institution rankings over time, and compare performance across different regions and nations.

Research Problem:

The World Ranking University dataset is being analyzed and visualized in the following ways: Choosing a university may be a tough prospect for both students and parents. To assist in selecting, I assess, and display university rankings based on many aspects such as facilities, professors, and educational quality. This gives students guidelines to help them make decisions while keeping their academic and personal goals in mind. These recommendations can also assist institutions in improving their offerings while better addressing their students requirements. Finally, this data-driven method aids candidates in navigating the difficult university application process and picking the best educational school for them.

Related Work:

The selection of a good university for higher education is a crucial decision that depends on various factors, including the faculty, the university's rank, and its facilities. The research paper written in the year 2020 by author “O. Loyola-González”, “A Contrast Pattern-Based Scientometric Study of the QS World University Ranking" analyzed the top 200 universities based on the QS rankings to rank them accordingly. The study revealed that universities with better rankings have dedicated webpages highlighting their position in various university rankings, which increases their attractiveness to funding agencies, researchers, faculty members, and other universities seeking collaboration. The paper also compared the strengths and weaknesses of various education systems and ranked universities accordingly. Overall, the paper provides valuable insights into the factors that influence university rankings and can help students, policymakers, and university administrators make informed decisions about selecting the right university for higher education.

According to the author Zoljargal Dembereldorj,2018 “Review on the Impact of World Higher Education Rankings: Institutional Competitive Competence and Institutional Competence”. The research suggests that rankings have become an important factor in the competitiveness of universities, both in advanced economies and developing countries. While rankings can provide benefits such as increased visibility and funding, they can also create a sense of competition that may prioritize ranking over quality education and research. The paper calls for a more nuanced approach to rankings that considers the diversity of higher education systems and university missions. The author argues that most top-ranked higher education institutions in the global university rankings are from countries with advanced economies. Overall, the study emphasizes the significance of university rankings in transforming higher education institutions throughout the world, as well as the need of a balanced approach to rankings that favors excellent education and research above competitiveness.

Methods:

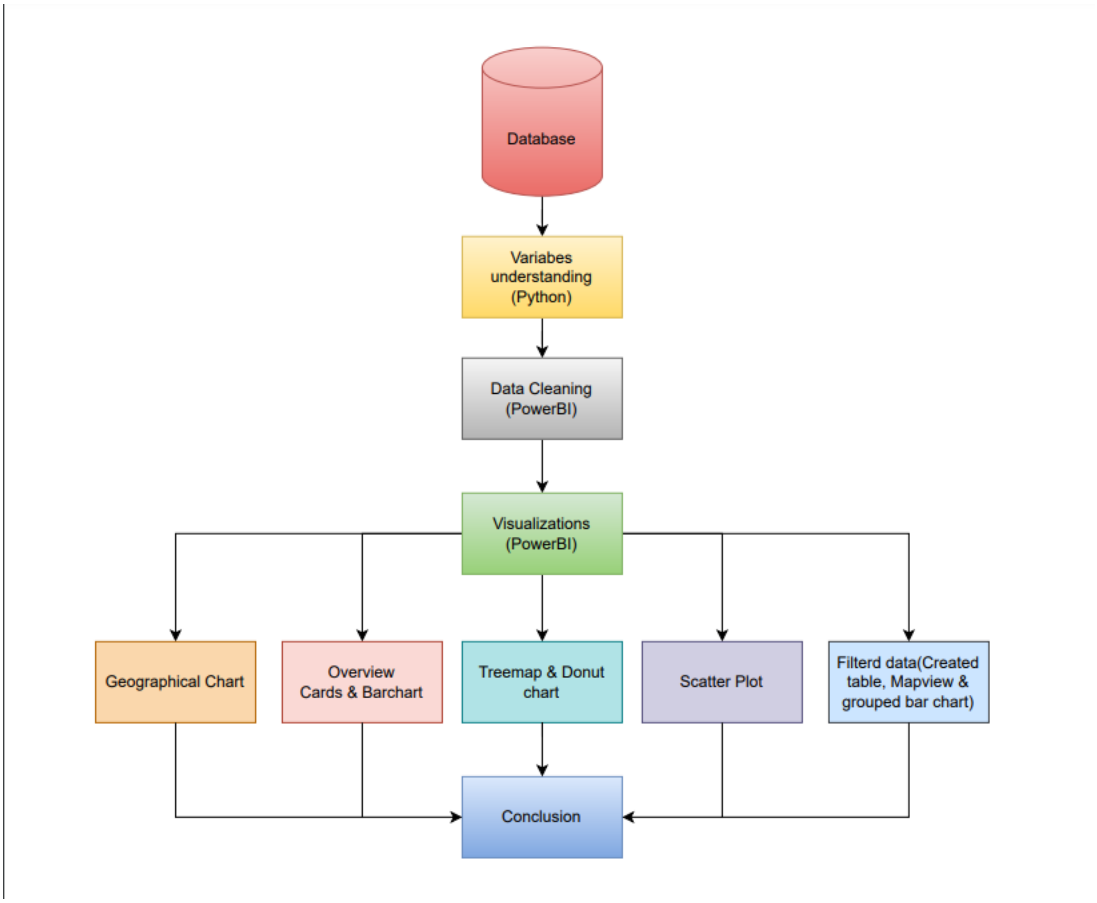
The methods used in this research are “Python” and “Power BI”. In this research “Python” is used for the information about the variable datatypes in the dataset and for finding missing values in the dataset. Whereas “Power BI” is used

for replacement of null values in the data and performing exploratory data analysis for the data.

Python is an object-oriented programming language. It is a popular and user-friendly programming language. It has simple syntax that is easy to learn and understand. Now a days python is used widely in many industries like web development, data analysis and artificial intelligence.

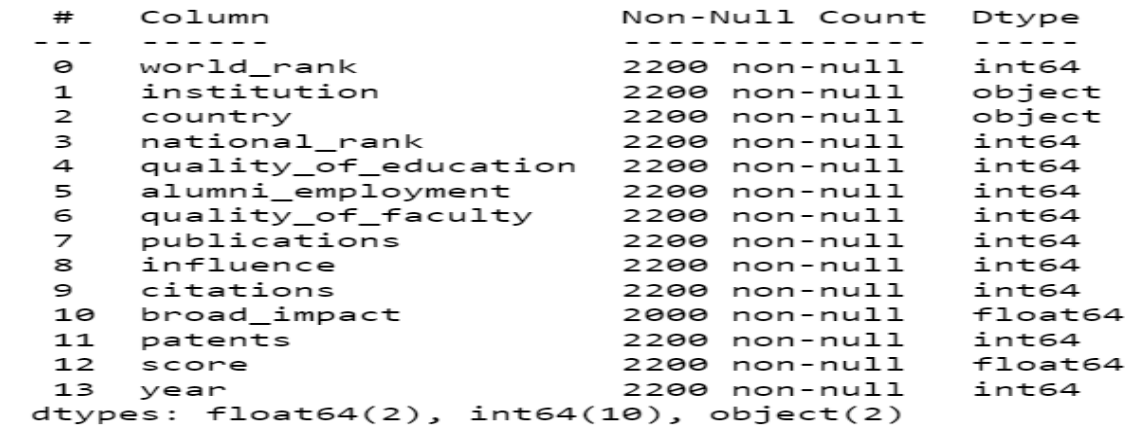
Microsoft's Power BI is a business analytics tool. It enables users to interact with and see data in a more engaging and dynamic manner. Power BI may be used to produce reports, dashboards, and other visualizations to assist users in extracting insights from their data. It is simple to use, even for novices, because it has a drag-and-drop interface for making charts and graphs. Power BI can also connect to a variety of data sources, including Excel spreadsheets, SQL databases, and cloud-based platforms like Azure. Power BI is an excellent tool for anybody trying to obtain insights from their data.

Workflow:



About Dataset:

The dataset for understanding about the universities is taken from the Kaggle Repository which is “cwurData.csv”. The dataset explains about the standing of the universities among all the universities in the world. The database contains 14 columns and 2200 records.



```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   world_rank            2200 non-null   int64
 1   institution           2200 non-null   object
 2   country               2200 non-null   object
 3   national_rank         2200 non-null   int64
 4   quality_of_education  2200 non-null   int64
 5   alumni_employment     2200 non-null   int64
 6   quality_of_faculty    2200 non-null   int64
 7   publications          2200 non-null   int64
 8   influence             2200 non-null   int64
 9   citations             2200 non-null   int64
 10  broad_impact          2000 non-null   float64
 11  patents               2200 non-null   int64
 12  score                 2200 non-null   float64
 13  year                  2200 non-null   int64
dtypes: float64(2), int64(10), object(2)
```

The above image is a clear explanation of the data. It shows the variables in the data, null values in the data and the datatypes of the variables.

Variables Explanation:

- World_rank - world rank for university.
- Institution - name of university.
- Country - country of each university.
- national_rank - rank of university within its country.
- quality_of_education - rank for quality of education
- alumni_employment - rank for alumni employment
- quality_of_faculty - rank for quality of faculty
- Publications - rank for publications.
- Influence - rank for influence
- Citations - number of students at the university

Data Cleaning:

In the variable “broad_impact” there are missing values of about 200 values which is 9.5%. As the null values percentage are less instead of dropping the column replaced missing values with 0 using ‘PowerBI’ tool.

Hypothesis:

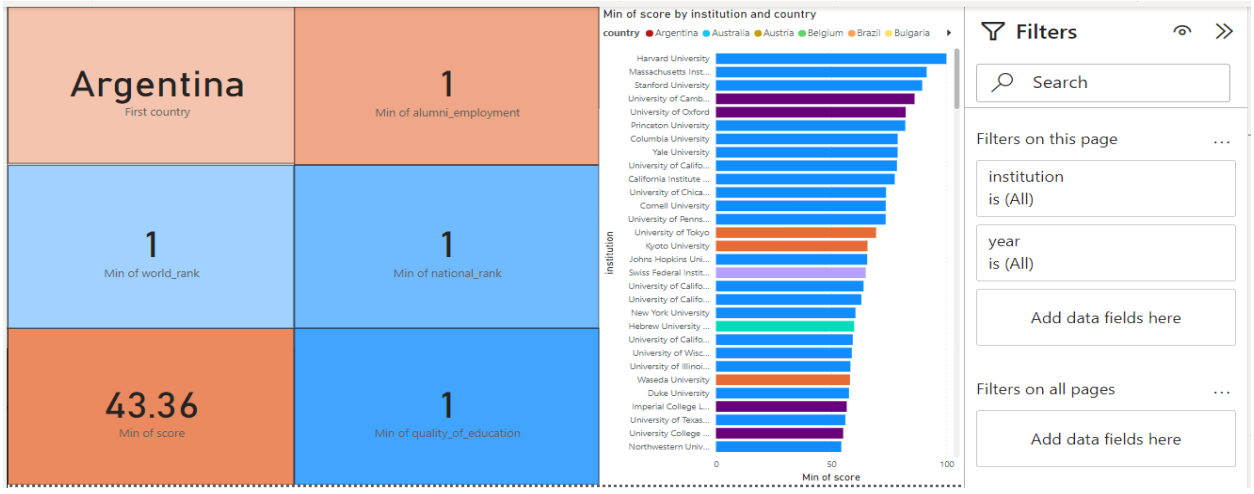
- Show me the educational institutions from which parts of the world have been considered?
- Show me the overview of your research?
- Explain which countries have more universities and which countries have less universities?
- Is there any relation between quality_of_education of the university and world_rank of the university?
- Is there any relation between quality_of_faculty of the university and citations in the university?
- Is there any relation between World_rank of the university and citations in the university?
- Is there any relation between quality_of_education of the university and quality_of_faculty in the university?
- Considering all the criteria of the universities show me the universities having their score greater than 75%?

Results:

Geographical Chart:

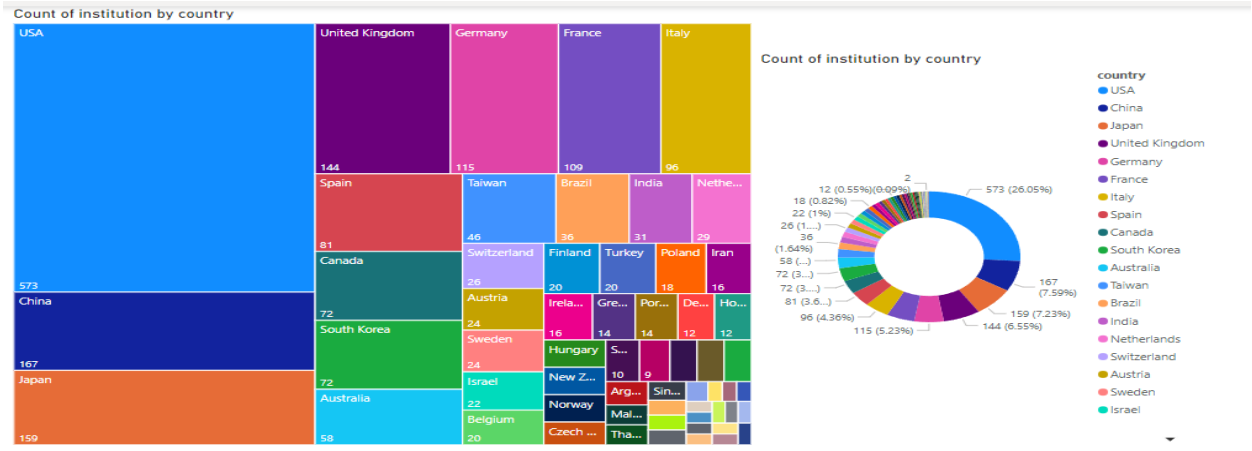
- The map shown above is the geographical map. Essentially, this graph depicts the number of universities in various nations. The larger circle denotes the greater number of institutions, whereas the smaller circle reflects the lesser number of institutions. According to the graphic, the "USA" country has a bigger number of colleges.

Overview of the universities:



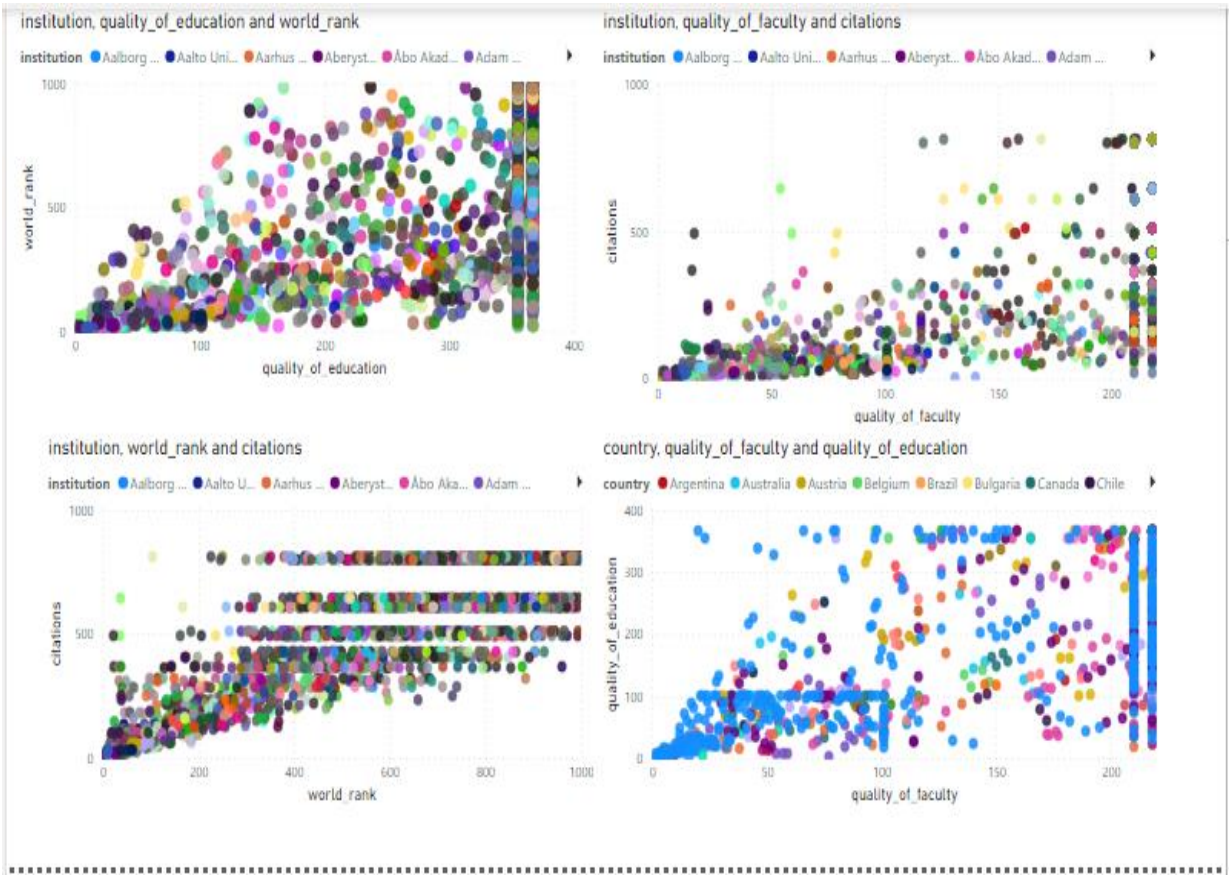
- The graphic above depicts a high-level overview of the institutions. You can see on the left side of the picture that cards are used to represent "Country", "Alumni_employment", "World_rank", "National_rank", "Score of the universities", "quality_of_education" and a bar graph to show the university names and scores. Right to the image filters are given the university name and year. We may choose a certain university and year by utilizing the filter pane. The data will automatically alter based on the inputs.

Tree map and Donut chart:



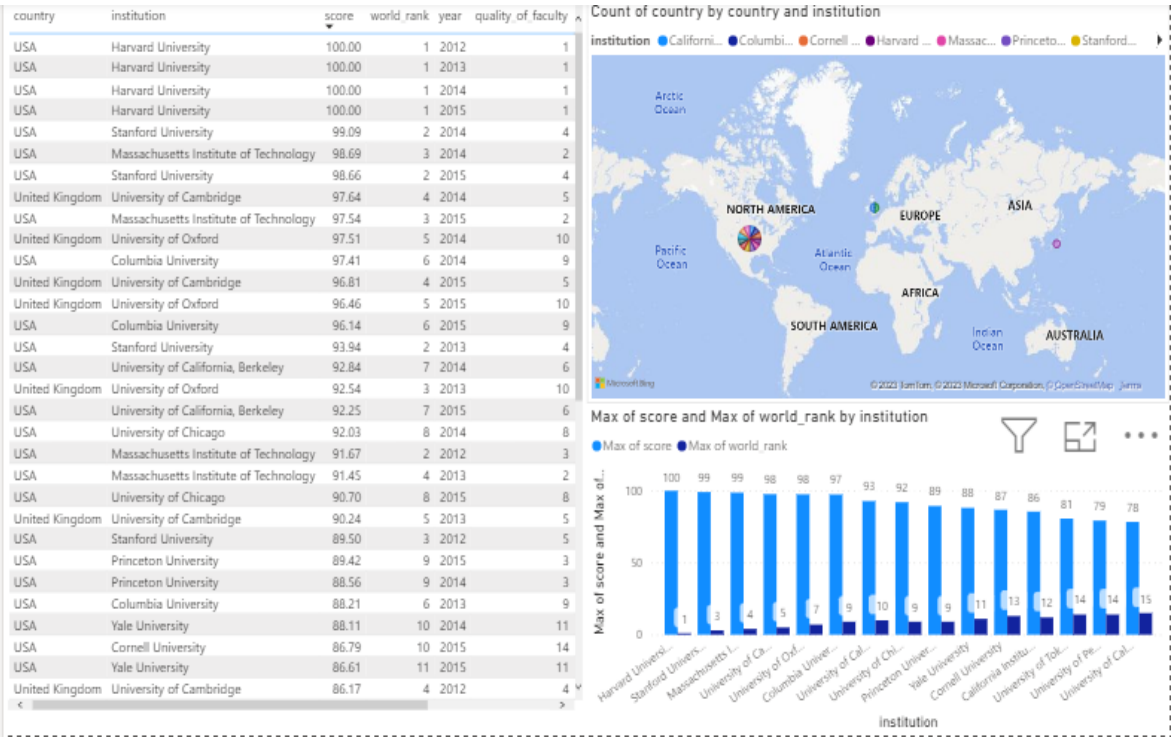
- The two charts shown above are a tree map and a donut chart. The tree map depicts the number of institutions in the nation, while the donut chart depicts the proportion of universities in each country in comparison to other countries. The "USA" country has 573 universities, which is 26.06% more than the other countries, followed by China with 167 universities, which is 7.59% more, and Japan with 159 universities, which is 7.23% more. Uruguay, Serbia, Iceland, and other countries have two universities, accounting for 0.09% of the total number of universities in the globe.

Scatter Plot:



- The scatter plots depicted in the above charts demonstrate the connection between variables. The accompanying graph depicts the relationship between educational quality and global ranking. As the quality of education at the university improves, its global ranking will be good. We can also discover a link between professor quality and citation. If a university's staff is of high quality, a higher number of students will be interested in enrolling there. Furthermore, if a university's global ranking is high, most students will be interested in enrolling there. Furthermore, if the quality of faculty at the university is high, the quality of education will be high as well. Universities will be ranked using all these factors.

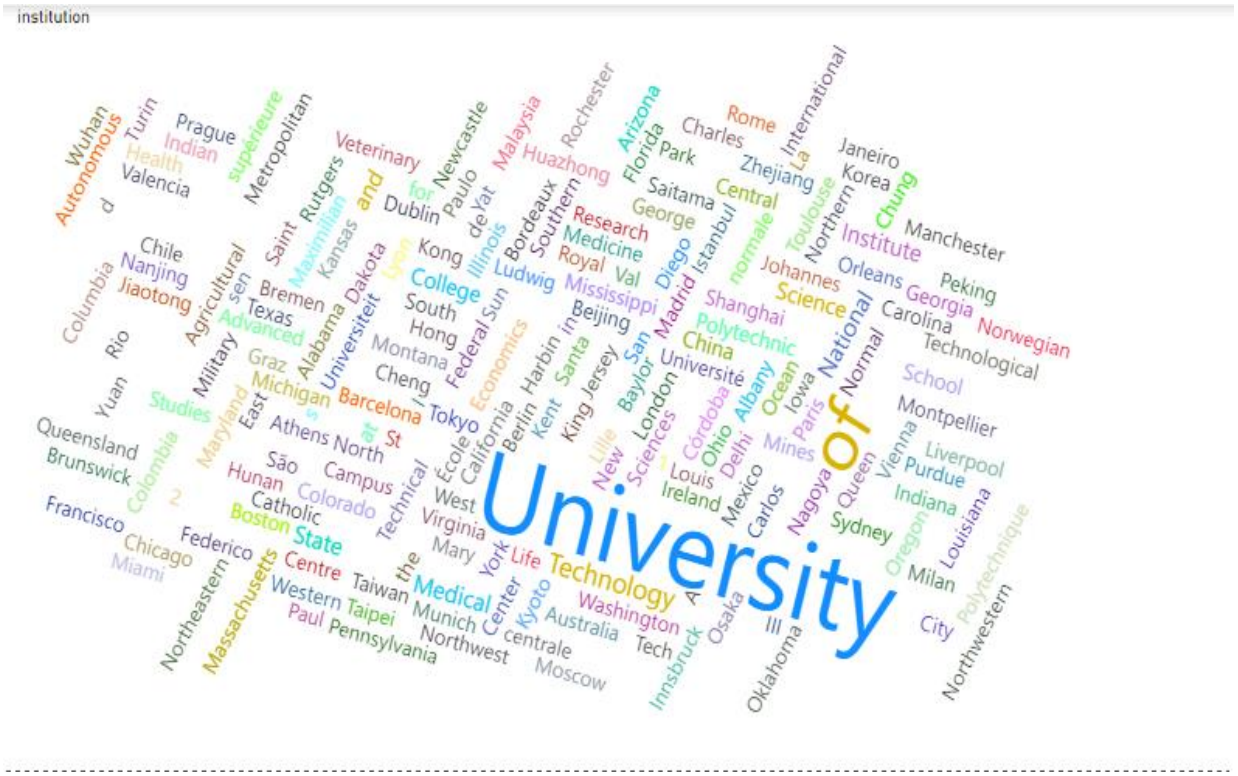
Filtered Data:



country	institution	score	world_rank	year	quality_of_faculty
USA	Harvard University	100.00	1	2012	1
USA	Harvard University	100.00	1	2013	1
USA	Harvard University	100.00	1	2014	1
USA	Harvard University	100.00	1	2015	1
USA	Stanford University	99.09	2	2014	4
USA	Massachusetts Institute of Technology	98.69	3	2014	2
USA	Stanford University	98.66	2	2015	4
United Kingdom	University of Cambridge	97.64	4	2014	5
USA	Massachusetts Institute of Technology	97.54	3	2015	2
United Kingdom	University of Oxford	97.51	5	2014	10
USA	Columbia University	97.41	6	2014	9
United Kingdom	University of Cambridge	96.81	4	2015	5
United Kingdom	University of Oxford	96.46	5	2015	10
USA	Columbia University	96.14	6	2015	9
USA	Stanford University	93.94	2	2013	4
USA	University of California, Berkeley	92.84	7	2014	6
United Kingdom	University of Oxford	92.54	3	2013	10
USA	University of California, Berkeley	92.25	7	2015	6
USA	University of Chicago	92.03	8	2014	8
USA	Massachusetts Institute of Technology	91.67	2	2012	3
USA	Massachusetts Institute of Technology	91.45	4	2013	2
USA	University of Chicago	90.70	8	2015	8
United Kingdom	University of Cambridge	90.24	5	2013	5
USA	Stanford University	89.50	3	2012	5
USA	Princeton University	89.42	9	2015	3
USA	Princeton University	88.56	9	2014	3
USA	Columbia University	88.21	6	2013	9
USA	Yale University	88.11	10	2014	11
USA	Cornell University	86.79	10	2015	14
USA	Yale University	86.61	11	2015	11
United Kingdom	University of Cambridge	86.17	4	2012	4

- The data shows the filtered scores and worldwide rankings of universities whose ranks are greater than 75%. Harvard University had the highest score of 100% and was ranked first globally. Stanford University and Massachusetts Institute of Technology followed with scores of 98.66% and 97.54% respectively. The University of Tokyo had the lowest score of 76.23%. The map only displays universities in the USA, UK, and Japan. The bar graph shows the scores and rankings, with Harvard having the highest score and UCLA having the lowest. The maximum score and total world rank are negatively correlated.

  Max of score and Max of world_rank diverged the most when the institution was Harvard University, when Max of score were 99 higher than Max of world_rank. Across all 15 institutions, Max of score ranged from 78.35 to 100 and Max of world_rank ranged from 1 to 15.



Discussion:

- Selecting a decent university for the student's bright future is critical to the student's professional advancement. But, based on what sources, should a student or parent choose a university? The above visuals clearly demonstrate the solution to this query.
- This research employs a geographical chart to display the number of highly ranked universities in various countries, allowing students to apply filters to identify universities with good scores, global and local rankings, and quality of education. The tree map and donut chart provide an overview of the number of highly ranked universities per country, while the scatter plot highlights the relationship between different factors such as quality of education, faculty, citations, and world rank. The study recommends universities with scores higher than 75% for admission, considering the quality of faculty, education, global rankings, and other relevant criteria. By analyzing the charts, students can compare the number of highly ranked universities across different countries and identify the best universities based on their personal preferences and goals.

Future Work:

- Performing Time – Series analysis on the dataset. The dataset covers the university ranking from years between 2012 and 2015. Time series analysis technique will identify the patterns and trends in the ranking data and will predict the future ranking of the university**.**
- Can also perform machine learning algorithms like Linear Regression, Random Forest, and Support vector machine. The linear regression algorithm will explain the linear relation between the predictor variables and target variable and will provide the importance of different features. Whereas Random Forest model combines multiple decisions trees to increase the accuracy of the model.

References

Dembereldorj1, Z. (n.d.). 1 Center for Foreign languages, Division of Humanities, School of Arts and Sciences, National University of Mongolia, Ikh surguuliin gudamj-1 P.O. Box – 46A/523, 210646, Ulaanbaatar, Mongolia.

O. Loyola-González, M. A.-P.-K. (2020). "A Contrast Pattern-Based Scientometric Study of the QS World University Ranking," in IEEE Access, vol. 8, pp. 206088-206104, 2020, doi: 10.1109/ACCESS.2020.3037665.