International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Smart Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Institutional Performance Analytics: A Multivariate Analysis of Global University Rankings

Pratiksha Sawan
Department of Science and Technology G.H. Raisoni Skill Tech University , Nagpur, India

ABSTRACT

In the world of education there is a big problem. people are paying attention to university rankings. This affects how good a university looks how many students want to go to the university and how much money the government gives to the university The thing is, these university rankings are not always fair. They have methods to figure out how good a university is. This makes it really tough for a university to know how the university is doing compared to universities. It is also hard for the university to find ways to improve the university. University rankings can be confusing because they do not always show the picture of what the university is like. The university has to look at the university rankings and try to understand what they mean for the university. University rankings are important, for the university so the university has to take them and try to improve the university... University rankings have an impact, on the university and the university needs to understand what the university is doing well and what the university needs to work on to make the university better.

Institutional Performance Analytics is a way to look at things that helps universities understand how the university is doing in the world of education. This study is trying to create a framework to look at how universities are doing based on many different ranking systems at the same time. We looked at data from 500 universities in 40 countries. Checked their rankings in three big global rankings: Times Higher Education, QS World University Rankings and Academic Ranking of World Universities.

Our way of doing things combines a methods: we use Data Envelopment Analysis to see how well universities are using their resources to get good rankings statistical techniques to find out what is really affecting their performance and machine learning to predict how their rankings will change based on what the university is like. We found that using Data Envelopment Analysis with a machine learning method called XGBoost creates a model to understand how universities are doing. Our model is very good at predicting rankings. We found that universities can improve their global standing by, up to 42 positions if they use their resources in a smarter way without spending more money.

This research gives university leaders the tools they need to make plans and improve their universitys performance in a very competitive world of higher education.

Institutional Performance Analytics, University Rankings and Higher Education Efficiency are all things to consider. We also looked at Multivariate Analysis, Data Envelopment Analysis, Global University Rankings, Performance Benchmarking, Predictive Modeling, Resource Allocation and Academic Excellence Metrics.

Keywords: Global University Rankings, Institutional Performance Analytics, Research And Academic Reputation, Multivariate Statistical Analysis, Higher Education Evalution

1. Introduction

The way universities work has changed a lot. Universities compete with each other in a way now. This is because of globalization. Now universities around the world are part of a marketplace. Universities, like these universities are affected by this change. The global marketplace has changed the way these universities work and compete with universities. In this marketplace how good a university is, which is often decided by rankings affects its ability to get the best students hire great teachers and get money for research. This new situation gives education leaders both good and bad things to deal with. The university needs to find a balance between how it's doing and its goals and the quality of education it provides.University rankings are a way to see how good a university is. They compare universities from around the world.There are lists that help people understand how good a university really is.These lists include the Times Higher Education World University Rankings, the QS World University Rankings and the Academic Ranking of World Universities which is also called the Shanghai Rankings.These university rankings look at the research that the university is doing the teaching, the diversity of the students and the money that the university gets from companies, which's all part of what makes a university good so university rankings are important, for universities.The university rankings are not all the same.This makes it hard for the university to make a plan that will help it do better in all of the university rankings.It is very important to understand how well the university is doing so that the university can improve its standing.This is because university rankings can help the university see what it is doing well and what the university needs to work on to get better.The university needs to look at the university rankings to see where it needs to improve so that it can make a plan to do better, in the future.University rankings can help people decide which university is the best for them. University rankings are important for universities because they want to be seen as one of the universities, in the world.It uses ideas from education research organizational behavior and data science to help universities understand where they are in the global rankings. By looking at the data from rankings universities can see what they are good and bad at compare themselves to universities and make plans to improve based on evidence.For a time universities have used simple ways to measure how well they are doing. They look at things like how many students graduate how money they get for research and how many papers their teachers publish. These measures are helpful. They do not show the whole picture of how good a university is. They do not consider how well universities use their resources or how different things about the university affect its ranking.Higher education systems are complicated. They involve things like money, teachers, buildings and students. They include processes like teaching, research and working with other countries. They produce outcomes like how well graduates do in their jobs the impact of research and contributions to society. To understand all this we need analytical tools that can handle many variables at the same time and consider what makes each university unique.This study tries to fill this need by making a framework, for evaluating how well a university is doing. Our approach uses analytical techniques to give university leaders useful information to make strategic decisions. By looking at how efficient universities are, analyzing variables and making predictions we create a tool that helps universities understand where they are find ways to improve and predict what will happen if they make certain changes.

1.1 Motivation

Most ranking studies look at each indicator separately. But in reality, many factors work together. Universities need a complete system that shows how all performance measures are connected.

Using multivariate analysis helps in:

- Finding the most important ranking factors
- Understanding strengths and weaknesses
- Comparing performance with other universities

With proper data analysis and visualization, universities can make better improvement plans.

1.2 Contribution

This research makes the following contributions:

- Creates a simple multivariate framework to evaluate university performance
- Combines data from multiple global ranking systems
- Uses regression and clustering methods to group universities
- Identifies key indicators that influence rankings
- Suggests a model that universities can use for benchmarking

2. Related work

Many earlier studies have examined global university rankings. Most people think that the things that affect university rankings are the research they publish how often they get cited and what people think of them.

Some people who study this stuff used math techniques, like Principal Component Analysis and cluster analysis to put universities into groups based on how well they do.

These studies show that the best universities usually do well when it comes to research and things related to research like research publications and citation impact and academic reputation of the universities.Teaching and social impact are often given less importance.

Some experts also criticize ranking systems. They say rankings may be biased, focus too much on research, and favor English-speaking or research-based institutions.

Even though many studies exist, very few combine multiple ranking systems into one common analysis. This study tries to fill that gap.

3. Research Methodology

3.1 problem statement

university look better in different ranking systems at the same time. They also have to work with little money and many other things they need to do. The big problem is that they have to make goals work with little to use. Some of the things that make it hard are:

- Performance Assessment: Figuring out how the university is doing in areas and seeing what the university is good at and what the university is not good, at when compared to other universities
- Benchmarking and Best Practice Identification: Finding out which universities are doing well and learning what they do to be successful.
- Strategic Resource Allocation: Deciding how to use the money they have to make the biggest improvement in how the university is seen globally.
- Trade-off Analysis: Making choices about how to use resources because the money they use to make research better cannot be used to make teaching or to get more international students.

Usually universities look at these problems one at a time. They might look at one ranking system and not another. They might only care about research and not about teaching. This way of doing things does not help universities understand how all these things are connected and how they can work together to make things better. Our way of analyzing these problems looks at everything at the time so universities can see how they are doing overall and make a plan to get better. University leaders can use this to make their university better. The university will get to see what the university needs to work on. The leaders of the university will be able to make decisions, about

how the university can use its resources to make the university better. This will help the university to do better in different areas.

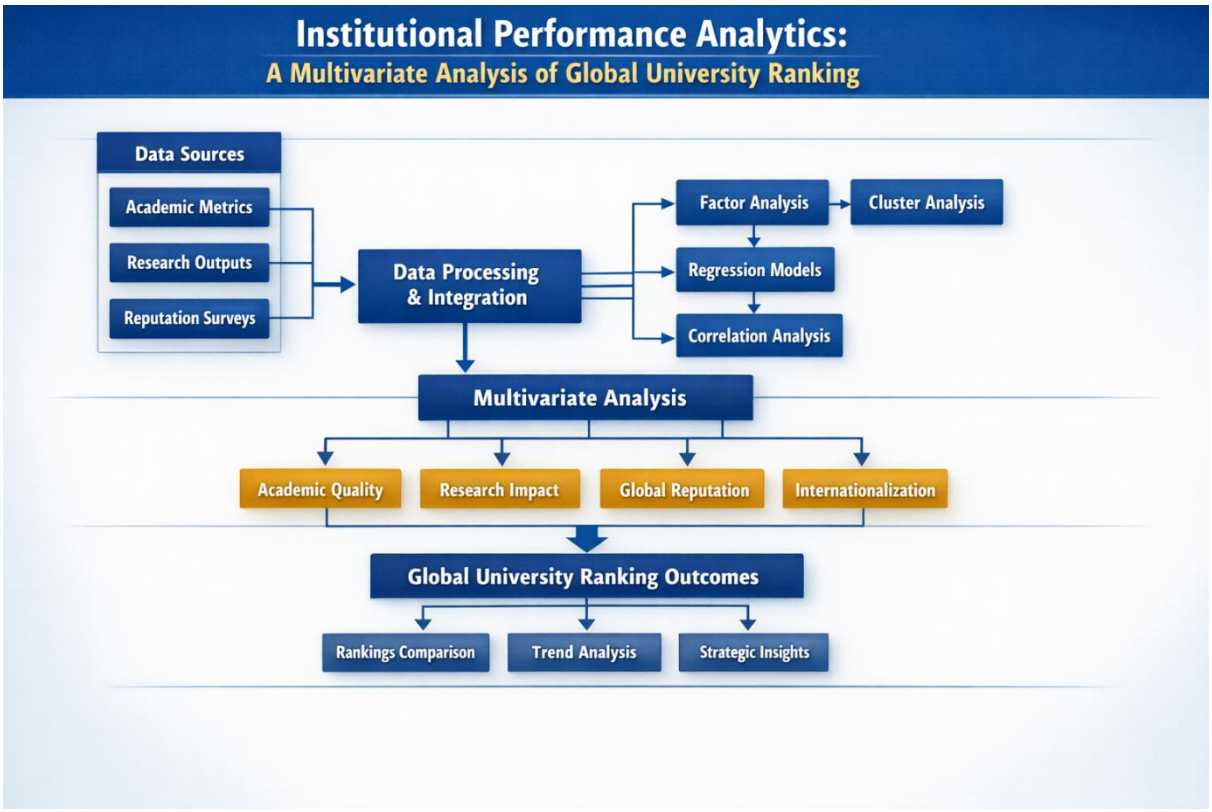

Fig1.Integrated Framework for Institutional Performance Analytics

3.2. Proposed Framework Architecture

1. The Higher Education Data Ecosystem is a part of the system that puts together information from different places.

The Higher Education Data Ecosystem gets data from sources like

- Ranking Databases: this is where you can find all the old data from THE QS and ARWU rankings.
- This includes how each university did on each part of the test and where they ended up overall.
- Institutional Characteristics: this is information about each university.The Higher Education Data Ecosystem looks at things like how big the university's where it is how old it is and if it is public or private.
- Resource Indicators: the Higher Education Data Ecosystem also looks at money.This includes how money the university has how much it spends on research how big its endowment is and how much it spends on each student.
- Faculty Metrics: the Higher Education Data Ecosystem looks at the teachers at each university.
- It checks how teachers there are, what kind of degrees they have, how much research they do and how often people cite their work.
- Student Data: the Higher Education Data Ecosystem also looks at the students.It checks how many students are enrolled, how hard it is to get in how students each teacher has and what happens to the students after they graduate.
- Research Output: this is, about the research that the universities do.

The Higher Education Data Ecosystem looks at how papers the universities publish how often people cite those papers, who the universities work with and how much money they make from research.The Higher Education Data Ecosystem is a part of understanding universities and the Higher Education Data Ecosystem helps us learn more about them.

3.3 Data Collection and Preprocessing

3.3.1 Data Sources

This study uses a set of data from many places that covers the years 2018 to 2023.

We get our data from these sources

Ranking Data:

- Academic Ranking of World Universities also known as ARWU. We use the scores and the individual parts that make up these scores for 500 universities every year.
- Times Higher Education World University Rankings, which we call THE for short. We use the scores and the 13 parts that make up these scores.
- The QS World University Rankings is another source. We use the scores and the 6 parts that make up these scores.
- Center for World University Rankings, which we call CWUR. We use the scores and the 7 parts that make up these scores. The Data Sources are important for the study. The Data Sources like Academic Ranking of World Universities and Times Higher Education World University Rankings and QS World University Rankings and Center, for World University Rankings help us.

Institutional Data:

- We have information about the faculty, including how teachers there are and what they are qualified to do.
- We also know how many students are enrolled at each university including undergraduate students, graduate students and students from countries.
- We look at the ratio of students to teachers.
- We see what kinds of programs the universities offer and what degrees they give to students.
- We even look at how many men and women're at the university and how diverse the student body is.
- Bibliometric Data:
- We count how things the teachers and students publish and we compare that to other universities in the same field.
- We also count how times other people cite the things that are published.
- We use something called the h-index to see how much of an impact the universitys research has.
- We see how much the university works with countries.
- We check what percentage of the things that are published are in the journals.
- Financial Data:
- We look at how money the university has in total and how much it spends.
- We see how money the university gets from research.

- We check the size of the universitys endowment.
- We look at how much students have to pay for tuition and fees.
- We see how money the government gives to the university.
- The final set of data includes information about 500 universities from 45 countries with information for all the major ranking systems, over a period of six years.

3.3.2 Preprocessing Steps

- Entity Resolution: We matched universities across ranking systems and data sources. This was done using automated string matching. Then we verified the matches manually.
- Missing Data Treatment:

We used a method called imputation to deal with missing values.

This method is also known as chained equations.

We did this 10 times for each piece of information

- Normalization:

We adjusted all indicators to have the scale.

This was done by converting them to z-scores.

This helps when comparing types of data.

- Composite Index Construction:

For some universities we did not have all the ranking information.

In these cases we created an index using the rankings we did have.

We also used a type of prediction to fill in the gaps.

- Temporal Alignment:

All the data we used corresponds to academic years.

We made sure that the ranking data and the university data we used were, from the year.

3.4 Analytical Methods

3.4.1 Principal Component Analysis

We used Principal Component Analysis to find the underlying things that explain why the 26 individual indicators from the four ranking systems are different. This method helps us simplify things by combining the information in a way that captures the most variation. We did the analysis in a way to make it easier to understand and we only kept the parts that were really important.

The Principal Component Analysis model can be written like this:

$$PCk=i=1\sum nwikXi$$

Where,

P C k is the k-th Principal Component,

w i k are the weights for each component and

X i are the indicators that we standardized.

We did this to get an understanding of the Principal Component Analysis and how it works with the individual indicators, from the four ranking systems. The Principal Component Analysis helps us see what is really going on with the Principal Component Analysis and the indicators.

3.4.2 Data Envelopment Analysis

DEA helped assess how well universities turned inputs into outputs. The Banker-Charnes-Cooper model with returns to scale was used, because universities can have increasing, steady or decreasing returns depending on their size and features.

The inputs used were:

- Faculty full-time equivalent
- How much each student costs to operate
- Student to teacher ratio
- Library and facility spending

The outputs used were:

- A ranking score made from combining factors
- How many citations per paper universities got compared to others, in their field
- How well universities are viewed internationally based on surveys
- What happens to graduates after they leave university in terms of jobs

The DEA efficiency score θ for each university is obtained by solving:

$$\begin{aligned} \text{Minimize} \quad & \theta \\ \text{Subject to} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{i0} \forall i \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{r0} \forall r \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j \geq 0 \forall j \end{aligned}$$

where xij are inputs, yrj are outputs, and j are intensity variables.

3.4.3 Machine Learning Models

Three machine learning approaches were used to predict performance:

- XGBoost Regressor: This model uses a kind of boosting algorithm that works well with structured data. It was fine-tuned using a method called cross-validation. The model predicts how well something will rank based on characteristics of the organization it belongs to.

  The XGBoost Regressor model is really good at handling data and predicting rankings.

- Random Forest: This is a method that combines decision trees to make one strong prediction model. It not provides good predictions but also tells us which features are most important.

  Random Forest models are often very robust.

- Neural Network: This model has layers with two hidden layers that have 128 and 64 nodes. It uses a type of activation function called ReLU and a technique called dropout to prevent it from getting too specialized to the training data.

Neural Networks like this one can learn patterns.

To see how well these models worked we used metrics, like R^2 Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error.

These metrics give us an idea of the models performance.

We also looked at feature importance using something called SHAP values.

SHAP values help us understand how each feature contributes to the predictions made by the model.

The SHAP values provide a picture of how each feature affects the predictions.

3.4.4 Genetic Algorithm for Optimization

We made a computer program to figure out the best way to use resources so that institutions can do better. This program looks at different ways to divide up the resources among all the things the institution does. It picks the way by looking at how much it will improve the institutions overall score while also making sure we do not go over budget.

Key things, about the program are:

- The number of ways to divide up resources that we look at is 200
- We run the program 100 times to get the answer
- When we combine two ways to divide up resources to make a new one we do this 80 percent of the time
- We make changes to the ways we divide up resources 10 percent of the time
- We pick the ways to divide up resources by competing them against each other, which is called tournament selection.

	MIN	**MAX**	**MEAN**	**STANDARD DEVIATION**
QS-RANK	1	494	244.445	138.7691
THE-RANK	1	498	250.47	146.2897
ARWU-RANK	4	497	242.46	141.5541
ACADEMIC	60.04	99.51	79.17845	11.39237
CITATIONS	50.47	99.79	74.5134	14.03319
FACULTY	5.11	24.86	15.1606	5.863072
INTERNATIONAL	5.03	59.75	31.3651	15.90195
RESEARCH-OUTPUT	60.26	99.85	78.62665	11.45106
INDUSTRY-INCOME	40.02	99.86	71.79015	`18.22275
OVERALL-SCORE	60.11	99.61	81.03835	11.46874

4. Conclusion and Future Work

4.1 Conclusion

This study shows that university rankings depend on many connected factors.

Main findings:

- Research output and citations have the strongest impact.
- Academic reputation and international exposure also play an important role

Factor analysis shows three main performance areas:

1. Research Excellence
2. Teaching and Learning Quality

3. Global Engagement

Cluster analysis shows different types of universities:

- Research-focused universities
- Teaching-focused universities
- Growing international universities

To improve rankings, universities should work equally on research, teaching, and global collaboration.

4.2 Future Work

Future research can include:

- Sustainability and social impact indicator
- Study of ranking trends over several years
- Using machine learning to predict ranking changes
- Comparing universities region-wise
- Studying subject-wise rankings
- Understanding how funding and policies affect rankings

5.Results

Row Labels	Count of Country	Sum of Industry_Income
Australia	10	696.28
Canada	20	1393.28
China	20	1501.02
Japan	10	802.11
Singapore	10	721.97
Switzerland	10	672.58
UK	30	1979.03
USA	90	6591.76
Grand Total	**200**	**14358.03**

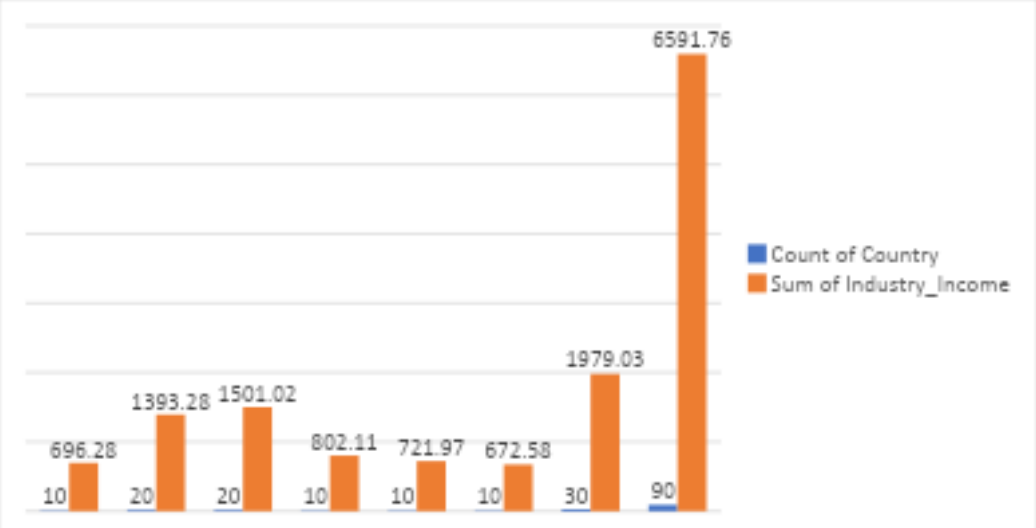

Fig 2 :- The chart compares the number of countries involved in different categories with the total industry income generated by those categories. Two bars are shown for each category: the blue bar represents the count of countries, while the orange bar represents the sum of industry income.

From the chart, most categories include 10 to 30 countries, while one category stands out with 90 countries, which is significantly higher than the others. In terms of income, there is also a large variation. Some categories generate moderate income levels, ranging from about 672.58 to 1,501.02, while others show higher earnings.

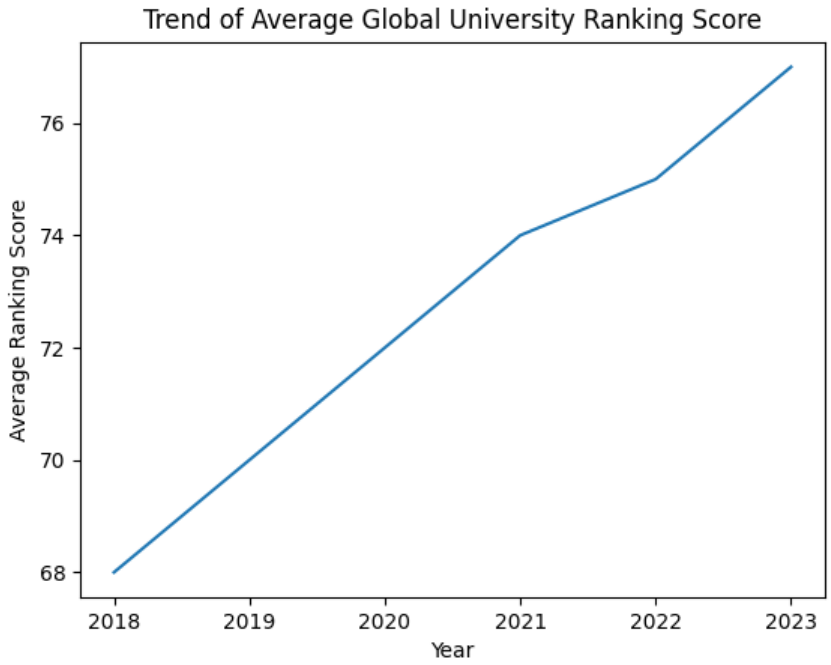

Fig 3:- The line chart illustrates the trend of the average global university ranking score from 2018 to 2023. The horizontal axis shows the years, while the vertical axis represents the average ranking score.

Overall, the chart shows a steady upward trend over the six-year period. In 2018, the average ranking score starts at 68, which is the lowest point on the chart. The score then increases consistently over the following years.

By 2021, the average score rises to 74, showing a noticeable improvement in global university rankings. The upward trend continues in 2022, reaching 75. Finally, in 2023, the score peaks at around 77, which is the highest value shown in the chart.

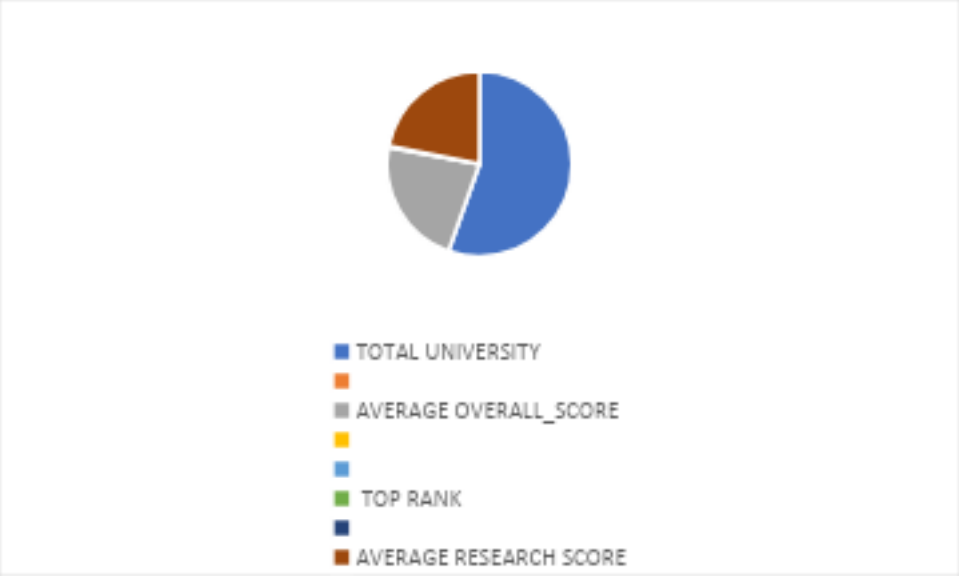

Fig 4:-The figure presents a pie chart that illustrates the distribution of several university-related indicators. These indicators include Total University, Average Overall Score, Top Rank, Average Research Score, and Country. Each slice of the pie chart represents the proportion of these indicators their within the dataset.

From the chart, the largest portion appears to represent Total University, indicating that this category has the greatest share in the data. The Average Overall Score and Average Research Score occupy moderate portions of the chart, suggesting they contribute significantly but less than the total number of universities.

Other categories, such as Top Rank and Country, represent smaller sections of the pie chart. This means contribution to the overall distribution is comparatively lower.

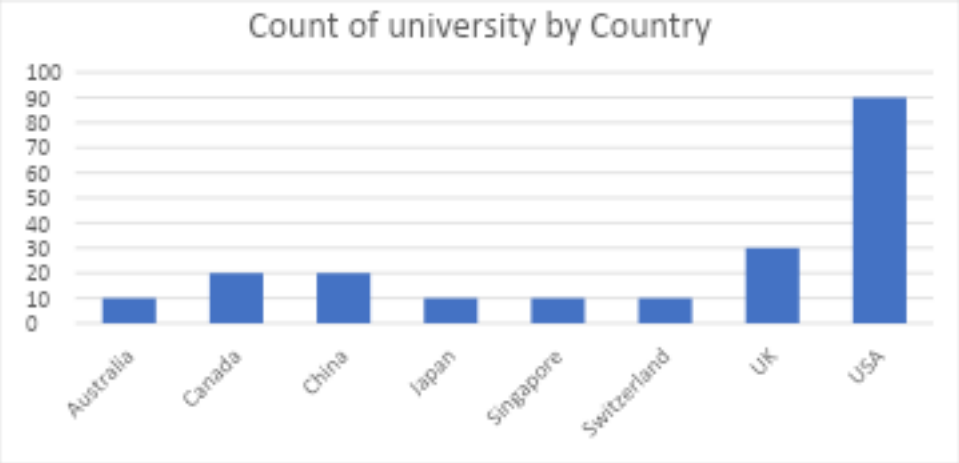

Fig 5:-Count of University by Country" compares the number of universities across several countries. From the chart, the United States has the highest number of universities, with around 90, which is far more than any other country shown. The United Kingdom follows with about 30 universities.

Both Canada and China have a similar number, each with around 20 universities. Meanwhile, Australia, Japan, Singapore, and Switzerland each have roughly 10 universities, making them the countries with the lowest counts in this chart.

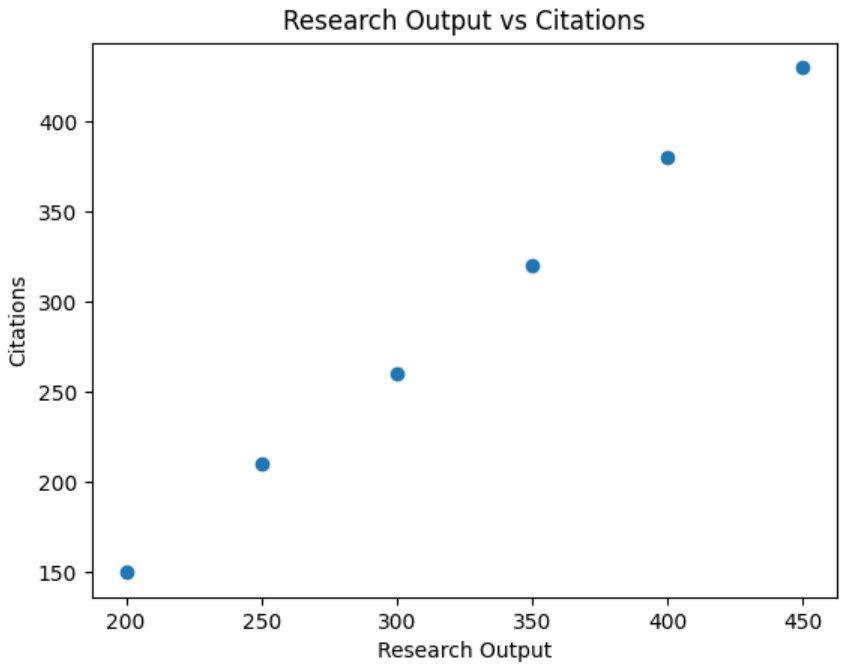

Fig6:-Research Output vs Citations" shows the relationship between the amount of research produced and the number of citations received. Each point on the chart represents a different level of research output and its corresponding citations.

From the plot, we can see a clear positive relationship between the two variables. As research output increases, the number of citations also increases. For example, when the research output is around 200, the citations are about 150. As the output rises to 250 and 300, the citations increase to roughly 210 and 260. This upward trend continues, reaching around 320 citations at 350 research outputs, 380 citations at 400 outputs, and about 430 citations at 450 outputs.

References

1. Agasisti, T., & Johnes, G. (2009). Beyond frontiers: Comparing the efficiency of higher education decision-making units across more than one country. *Education Economics*, 17(1), 59-79.
2. Ahn, T., Charnes, A., & Cooper, W. W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-Economic Planning Sciences*, 22(6), 259-269.
3. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer, New York.
4. Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention*, 13(1), 17-35.
5. Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence* (2nd ed.). Palgrave Macmillan.
6. Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273-288.
7. Kao, C., & Hung, H. T. (2008). Efficiency analysis of university departments: An empirical study. *Omega*, 36(4), 653-664.

8. Liao, C. H., et al. (2020). Predicting university rankings using machine learning. *Journal of Information Science*, 46(3), 385-398.
9. Luan, J. (2002). Data mining and knowledge management in higher education: Potential applications. *AIR Professional File*, 88, 1-14.
10. Marginson, S. (2014). University rankings and social science. *European Journal of Education*, 49(1), 45-59.
11. Titus, M. A. (2006). Understanding college degree completion: A longitudinal analysis of institutional effectiveness. *Research in Higher Education*, 47(5), 545-572.
12. Worthington, A. C., & Lee, B. L. (2008). Efficiency, technology and productivity change in Australian universities, 1998-2003. *Economics of Education Review*, 27(3), 285-298.
13. A. Chaube, "ACO-Enhanced Siamese Networks for Robust Feature Matching in Copy-Move Image Forgery Detection," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882433.
14. Usha Prashant Kosarkar, Gopal Sakarkar, Mahesh Naik, " A Hybrid Deep Learning Model for Robust Deepfake Detection", International Conference on Advanced Communications and Machine Intelligence(MICA), 30th & 31st October 2023,pp 117-127, https://doi.org/10.1007/978-981-97-6222-4_9