University Analysis

Rahma Touaibi

2025-12-10

```
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(tidyr))
suppressPackageStartupMessages(library(tidyverse))
suppressPackageStartupMessages(library(lubridate))
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(Hmisc))
```

```
## Warning: le package 'Hmisc' a été compilé avec la version R 4.5.2
```

1.Introduction

This project analyzes the top universities worldwide based on their overall scores and citations. It visualizes the rankings and explores the relationship between university performance and research impact.

2.Data Cleaning

The dataset contains global university rankings.

We will:

-Remove duplicates

-Handle missing values

-Check and correct data types

-Prepare data for visualization

Upload the data

```
rank<-read.csv("C:\\Users\\admin\\Desktop\\Data Analytics\\Projects For Portfolio\\MiniProjetR\\cwurData
```

Make a copy of data

```
rank_clean<-rank
```

Check data

```
head(rank_clean)
```

```
##   world_rank                          institution        country national_rank
## 1          1                    Harvard University            USA             1
## 2          2 Massachusetts Institute of Technology            USA             2
## 3          3                   Stanford University            USA             3
```

```
## 4          4               University of Cambridge United Kingdom             1
## 5          5    California Institute of Technology            USA             4
## 6          6                  Princeton University            USA             5
##   quality_of_education alumni_employment quality_of_faculty publications
## 1                    7                 9                  1            1
## 2                    9                17                  3           12
## 3                   17                11                  5            4
## 4                   10                24                  4           16
## 5                    2                29                  7           37
## 6                    8                14                  2           53
##   influence citations broad_impact patents  score year
## 1         1         1           NA       5 100.00 2012
## 2         4         4           NA       1  91.67 2012
## 3         2         2           NA      15  89.50 2012
## 4        16        11           NA      50  86.17 2012
## 5        22        22           NA      18  85.21 2012
## 6        33        26           NA     101  82.50 2012
```

Check missing values

```
colSums(is.na(rank_clean))
```

```
##           world_rank          institution              country
##                    0                    0                    0
##        national_rank quality_of_education    alumni_employment
##                    0                    0                    0
##   quality_of_faculty         publications            influence
##                    0                    0                    0
##            citations        broad_impact              patents
##                    0                  200                    0
##                score                 year
##                    0                    0
```

Check dublicates

```
anyDuplicated(rank_clean)
```

```
## [1] 0
```

Clean data

```
rank_clean$broad_impact[is.na(rank_clean$broad_impact)]<-mean(rank_clean$broad_impact,na.rm=TRUE)
rank_clean$broad_impact<-round(rank_clean$broad_impact,2)
rank_clean <- rank_clean %>%
  distinct(institution, .keep_all = TRUE)
```

3.KPIs
Average score

```
avg_score <- rank_clean %>%
  summarise(Average_Score = mean(score, na.rm = TRUE))
avg_score
```

```
##   Average_Score
## 1      46.50258
```

Total Citations

```
total_citations <- rank_clean %>%
  summarise(Total_Citations = sum(citations, na.rm = TRUE))
total_citations
```

```
##   Total_Citations
## 1          458926
```

Insights

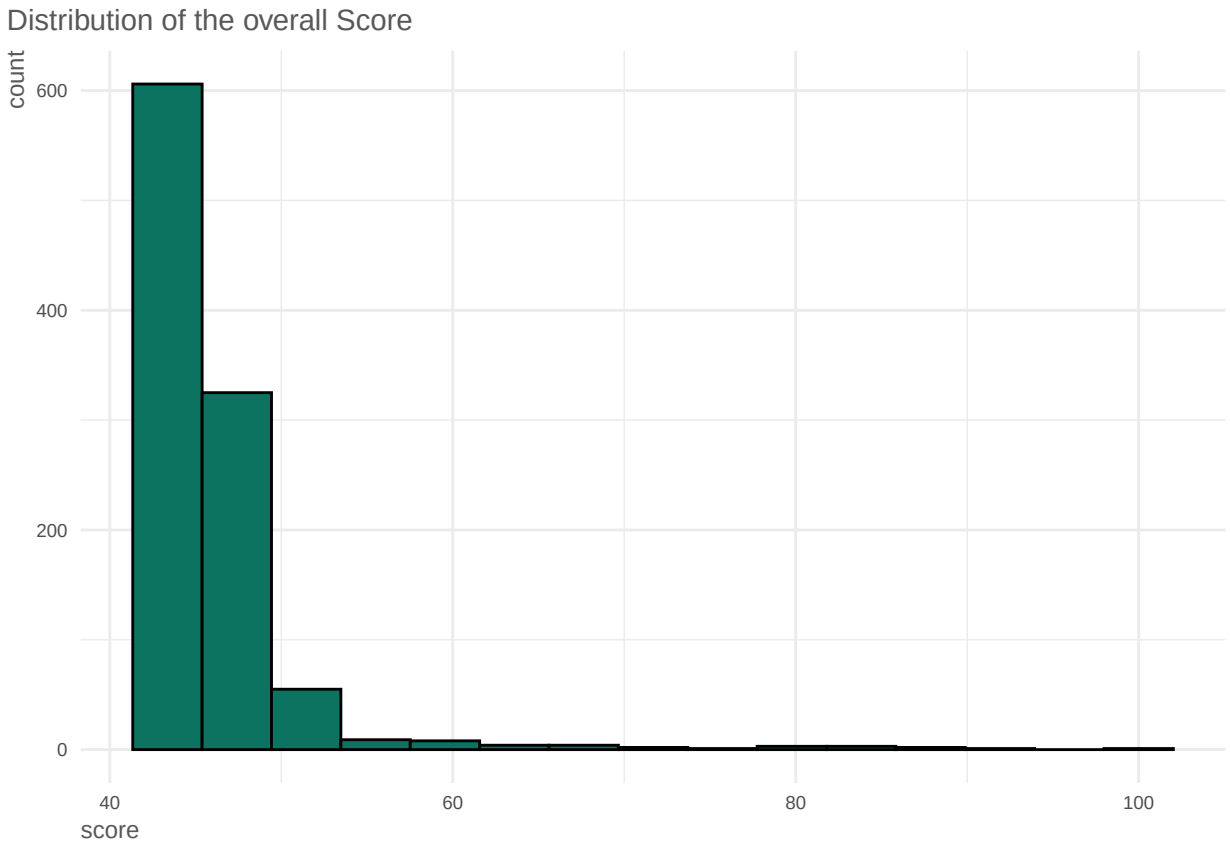
Average Score: The global average university score is 46.5, showing moderate overall performance.

Total Citations: Universities have a total of 458,926 citations, reflecting their research impact.

4.Visualizations

4.1 Distribution of the overall Score

```
ggplot(rank_clean,aes(score))+
geom_histogram(bins=15,fill="#0B735F",color="black")+
ggtitle("Distribution of the overall Score")+
theme_minimal()+
theme(plot.title =element_text(size=12,color ="#585858"))+
theme(
  plot.title = element_text(size=11, hjust=-0.1, color="#585858"),
  axis.title.x = element_text(size=9,hjust=0, color="#585858"),
  axis.title.y = element_text(size=9,hjust=1, color="#585858"),
  axis.text.x = element_text(size=7),
  axis.text.y = element_text(size=7)
)
```

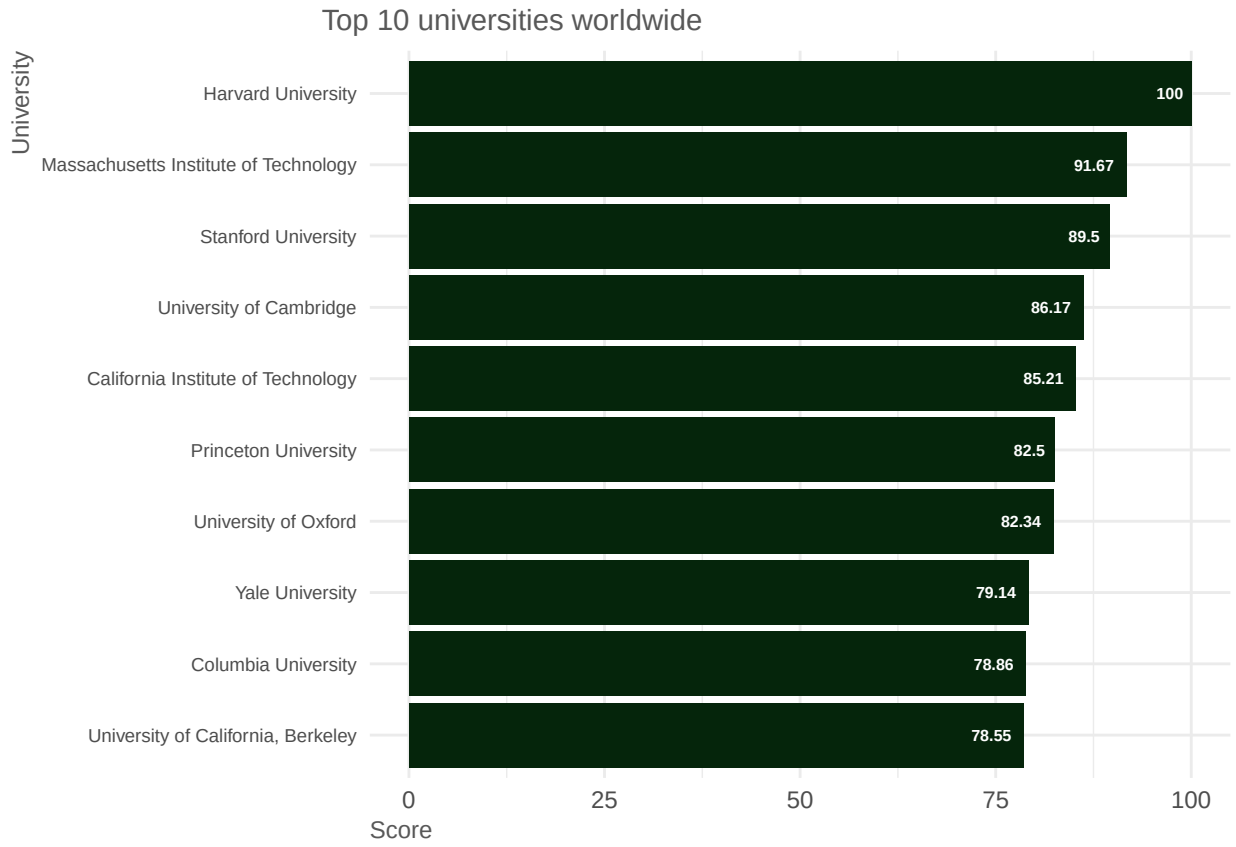


Insights

Most universities score between 40 and 50, with very high scores being rare.

4.2 Top 10 universities worldwide

```
rank_clean%>%
  arrange(desc(score))%>%
  slice(1:10)%>%
  mutate(institution=factor(institution,levels = institution))%>%
  ggplot(aes(x=reorder(institution,score),y=score))+
  geom_col(fill="#05250B")+
  coord_flip()+
  geom_text(aes(label = score),
            hjust = 1.3,  # moves text to the left of the bar
            color = "white",
            size = 2,
            fontface="bold") +
  theme_minimal()+
  labs(title="Top 10 universities worldwide",x="University",y="Score")+
  theme(
    plot.title = element_text(size = 11, hjust =-0.1,color="#585858"),
    axis.title.x = element_text(size = 9,hjust =0,color="#585858"),
    axis.title.y = element_text(size = 9,hjust =1,color="#585858"),
    axis.text.y = element_text(size = 7),
    legend.position = "bottom"
  )
```



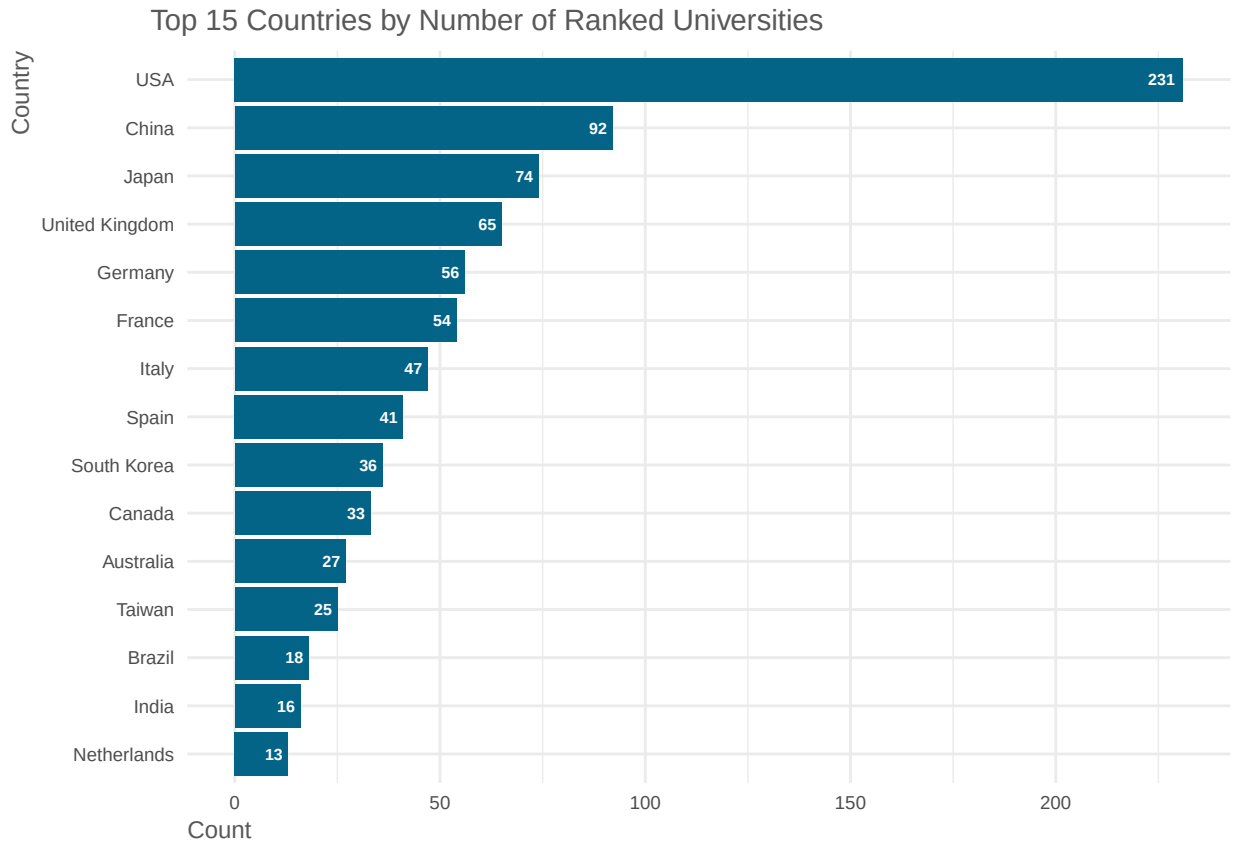
Insights

The top 10 universities achieve very high scores, led by Harvard with a perfect 100.

Most of these top institutions are based in the USA, showing its dominance in global rankings.

4.3 Top 15 Countries by Number of Ranked Universities

```
rank_clean %>%
  count(country, sort=TRUE) %>%
  slice(1:15) %>%
  mutate(country = factor(country, levels = country)) %>%
  ggplot(aes(x = reorder(country, n), y = n)) +
  geom_col(fill = "#046488") +
  geom_text(aes(label = n), hjust = 1.3, color = "white", size=2,fontface="bold") +
  coord_flip() +
  theme_minimal() +
  labs(title="Top 15 Countries by Number of Ranked Universities", x="Country", y="Count") +
  theme(
    plot.title = element_text(size=11, hjust=-0.1, color="#585858"),
    axis.title.x = element_text(size=9,hjust =0, color="#585858"),
    axis.title.y = element_text(size=9,hjust =1, color="#585858"),
    axis.text.x = element_text(size=7),
    axis.text.y = element_text(size=7)
  )
```

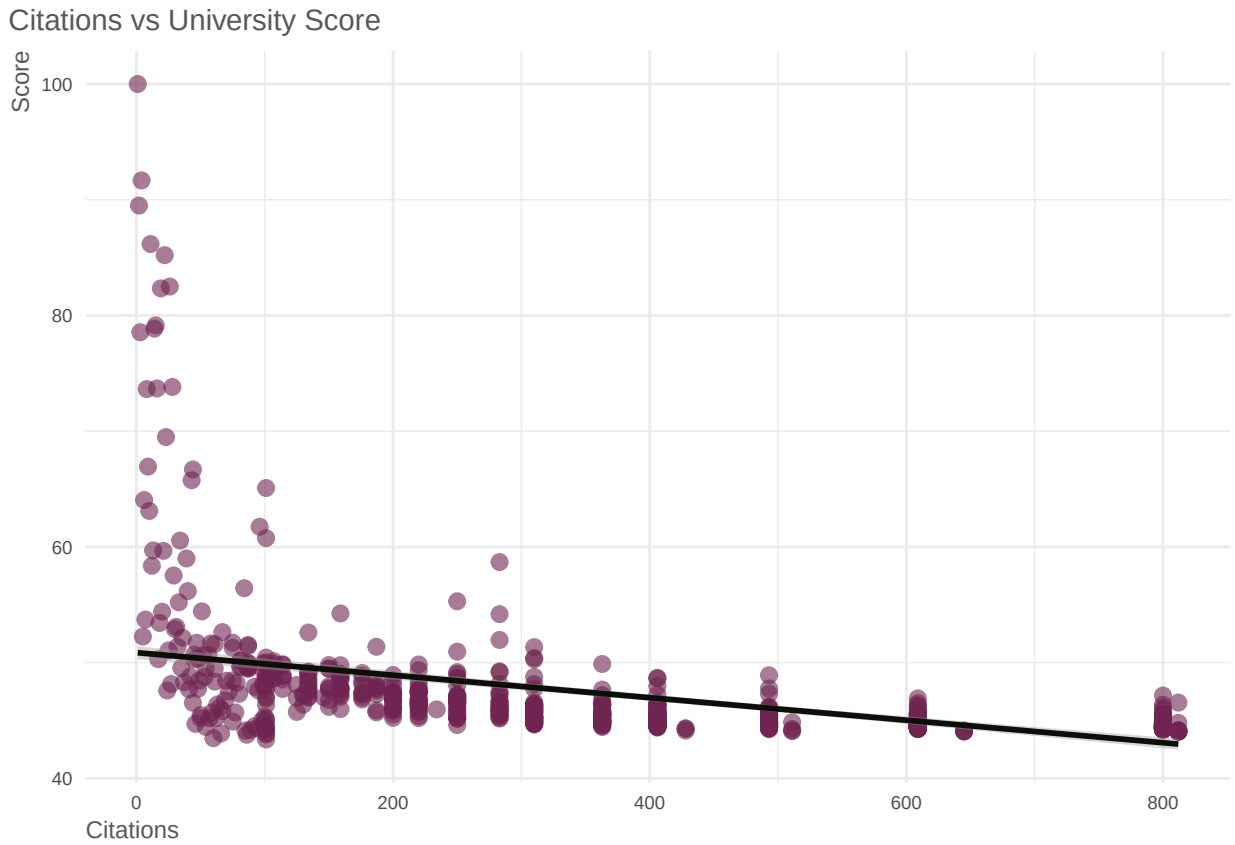


Insights *The USA clearly leads with 231 ranked universities, far more than any other country. China follows with 92, Japan with 74, and the UK with 65, showing a big gap between the USA and the rest.*

4.4 Citations vs University Score

```
ggplot(rank_clean, aes(x = citations, y = score)) +
  geom_point(color="#6F2451", alpha=0.6, size=2.5) +
  geom_smooth(method = "lm", se = TRUE, color = "#0B0F08")+   # adds trendline
  theme_minimal() +
  labs(title="Citations vs University Score", x="Citations", y="Score") +
  theme(
    plot.title = element_text(size=11,hjust=-0.1, color="#585858"),
    axis.title.x = element_text(size=9,hjust=0, color="#585858"),
    axis.title.y = element_text(size=9,hjust=1, color="#585858"),
    axis.text.x = element_text(size=7),
    axis.text.y = element_text(size=7)
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



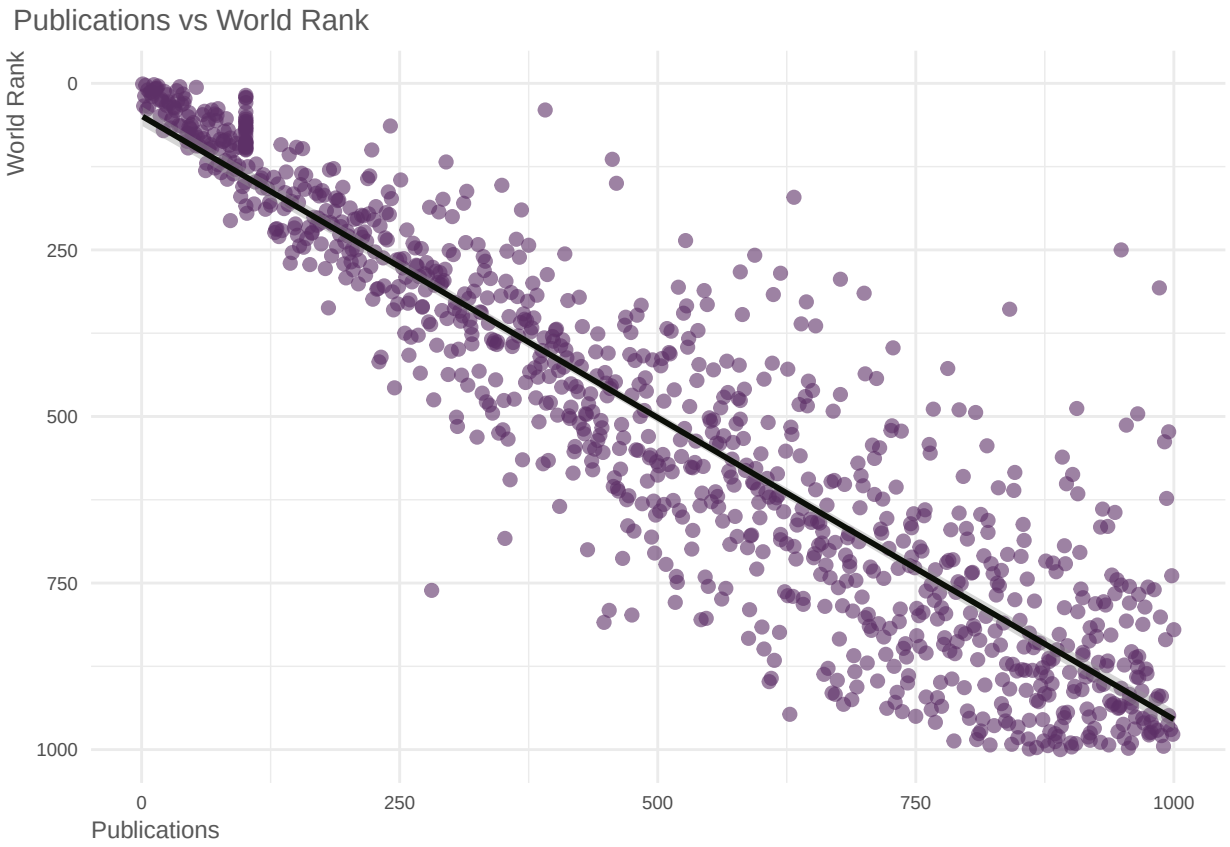
Insights

Universities with higher citation counts do not necessarily achieve higher overall scores, as the trend line slopes downward. This suggests that citations alone are not a strong predictor of ranking performance, and other factors play a bigger role in determining a university's score.

4.5 Publications vs World Rank

```
ggplot(rank_clean, aes(x = publications, y = world_rank)) +
geom_point(color="#5D3067", alpha=0.6, size=2) +
geom_smooth(method = "lm", se = TRUE, color = "#0B0F08") +  #              adds trendline
scale_y_reverse() +  # Rank 1 is top
theme_minimal() +
labs(title="Publications vs World Rank", x="Publications", y="World Rank") +
  theme(
    plot.title = element_text(size=11,hjust=-0.1, color="#585858"),
    axis.title.x = element_text(size=9,hjust=0, color="#585858"),
    axis.title.y = element_text(size=9,hjust=1, color="#585858"),
    axis.text.x = element_text(size=7),
    axis.text.y = element_text(size=7)
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Insights

Top-ranked places or people publish a lot more work. This high output helps them stay at the top and be well-known

5.Insights & Recommendations

Focus on research and publications to improve rankings.

Improve teaching, collaboration, and innovation, not just citations.

Learn from top universities to see what works.

Support smaller universities with funding and partnerships.

Conclusion

The analysis shows that top universities are mostly concentrated in a few countries, with the USA leading. While citations contribute to rankings, overall performance depends on multiple factors like research output, teaching quality, and innovation.