Data Science Principles with Applications on Educational Data
GROUP PROJECT REPORT
Exploring Key Factors in Educational Data: A Case Study on University Rankings

Group: 06

Name: 張育丞, 張宇青, 周姿妤, 黃詳諺, Duong Van Nhat Quang

Sid: 113598043, 113598032, 113598090, 113598091, 113998411

Date: 06/11/2025

National Taipei University of Technology
Computer Science and Information Engineering

TABLE OF CONTENTS

01

Introduction

INTRODUCTION

- Objective
  - Understand how critical academic indicators shape university competitiveness and influence rankings globally.
- Integrated Analysis
  - Combining these features reveals deeper insights into how universities perform in a global context.
- Visualization
  - Graphs and correlation maps enhance interpretability and strategic planning.
- Outcome
  - Data-driven strategies can help institutions improve policies, optimize resources, and elevate their international standing.

02

Literature Review And Related Works

LITERATURE REVIEW AND RELATED WORKS

- Dataset from Kaggle: "World University Ranking."
- Selected 13 headers for analysis

  (List in two columns)

  - World Rank
  - Institution
  - Country
  - Quality of Faculty
  - Influence
  - Broad Impact
  - Score
  - National Rank
  - Quality of Education
  - Alumni Employment
  - Publications
  - Citations
  - Patents
  - Year

03

Problem Statement

PROBLEM STATEMENT

Topic	Research Question	Method
1. Citations vs. World Rank	Does more citations lead to better rank?	Correlation + Linear Regression
2. Key Indicators Analysis	Which indicator influences rank most?	Relation Weight Analysis
3. Alumni Employment vs. Score	Does better employment lead to higher score?	Correlation + Regression
4. Country Impact	Do specific countries rank higher?	Group Analysis
5. Patents vs. Academic Impact	Which type of university ranks higher?	Comparative Analysis
6. University Type Classification	Classify by citations and patents	Clustering

04

Proposed Models (Approaches)

PROPOSED MODELS (APPROACHES)

1. Data Pre-processing
   - Handled missing & inconsistent values
   - Ensured logical and format consistency
2. Correlation Analysis
   - Assessed 13 indicators vs. rankings
   - Used Pearson correlation, regression models, and heatmaps
3. Feature Importance Ranking
   - Ranked indicators by regression weight
   - Visualized key features impacting rankings
4. Country-Level Comparison
   - Analyzed same-score universities across countries
   - Identified regional biases in rankings
5. Output-Based Grouping
   - Grouped by citations & patents
   - Compared ranking impacts with statistical tests
6. Clustering
   - Applied K-Means & hierarchical clustering
   - Profiled institutions based on research output

05

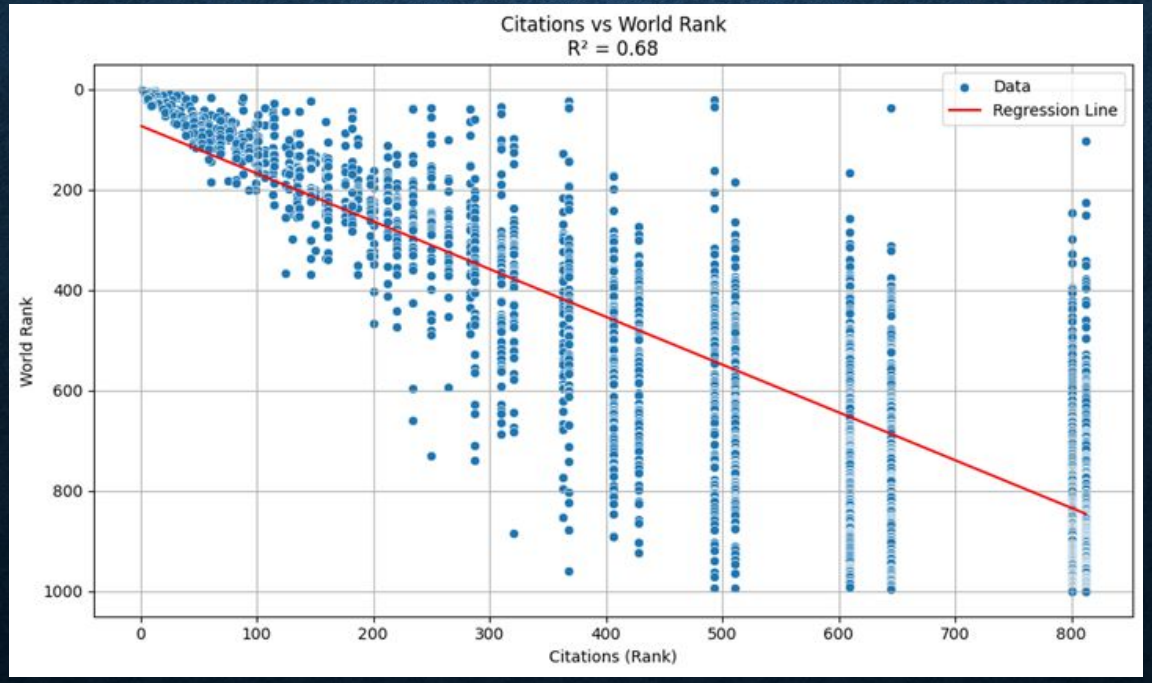
Experiments

EXPERIMENTS

1. Does more citations lead to better rank?
2. Which indicator influences rank most?
3. Does better employment lead to higher score?
4. Do specific countries rank higher?
5. Which type of university ranks higher?
6. Classify by citations and patents.

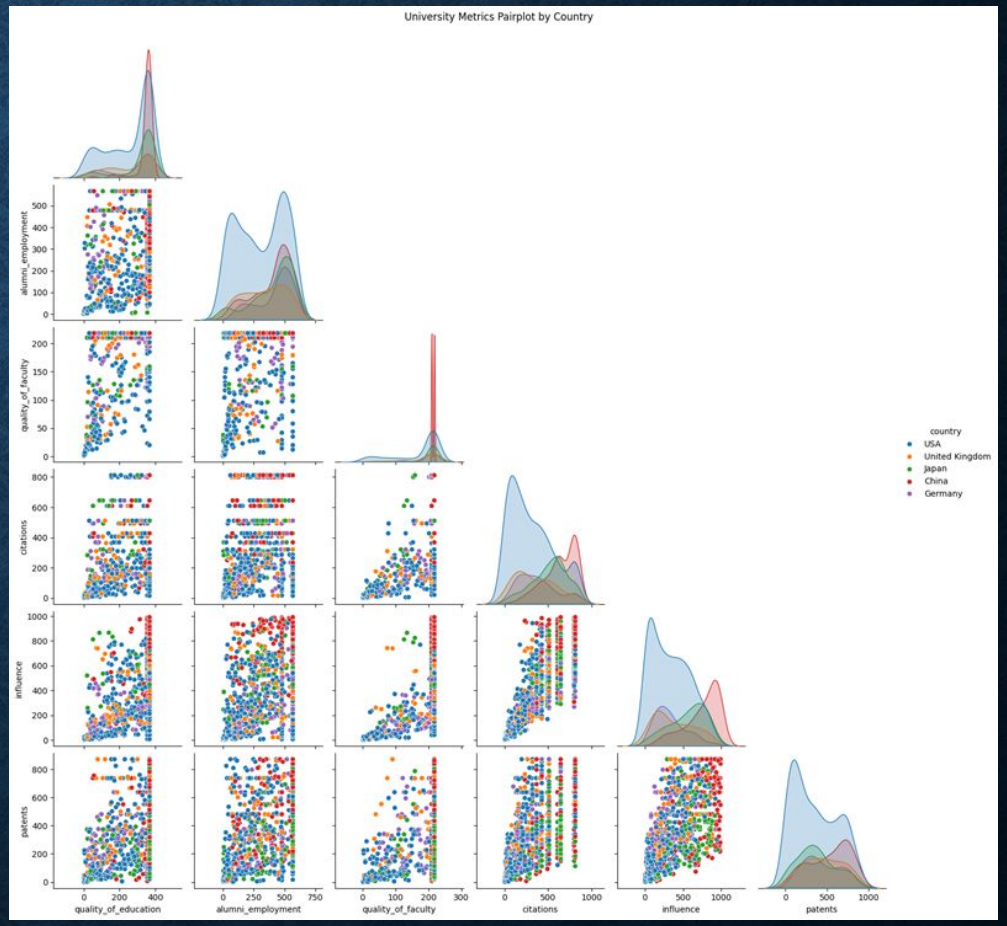
1. Does More Citations Lead To Better Rank?

- **Method: Correlation Analysis + Linear Regression**
  By using **citations** as the independent variable and **world_rank** as the dependent variable, our regression analysis revealed
  - A **significant negative correlation** between citation count and world ranking.
  - In short, **the more citations a university receives, the higher it tends to rank** globally.

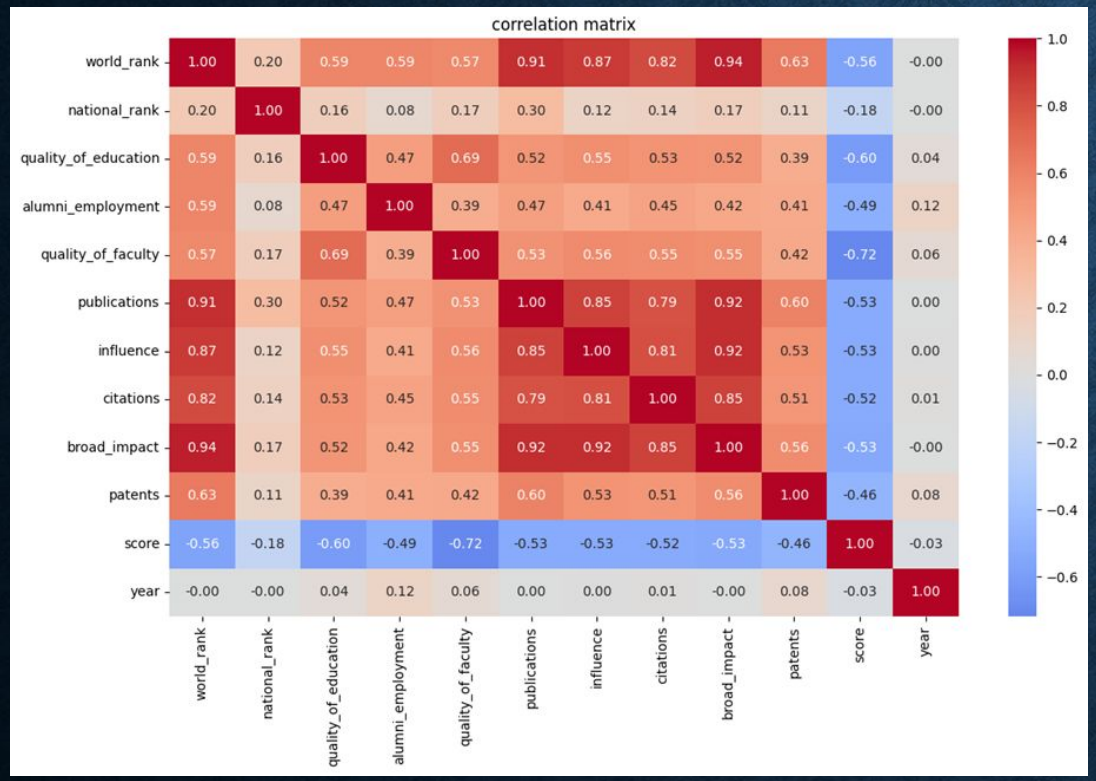


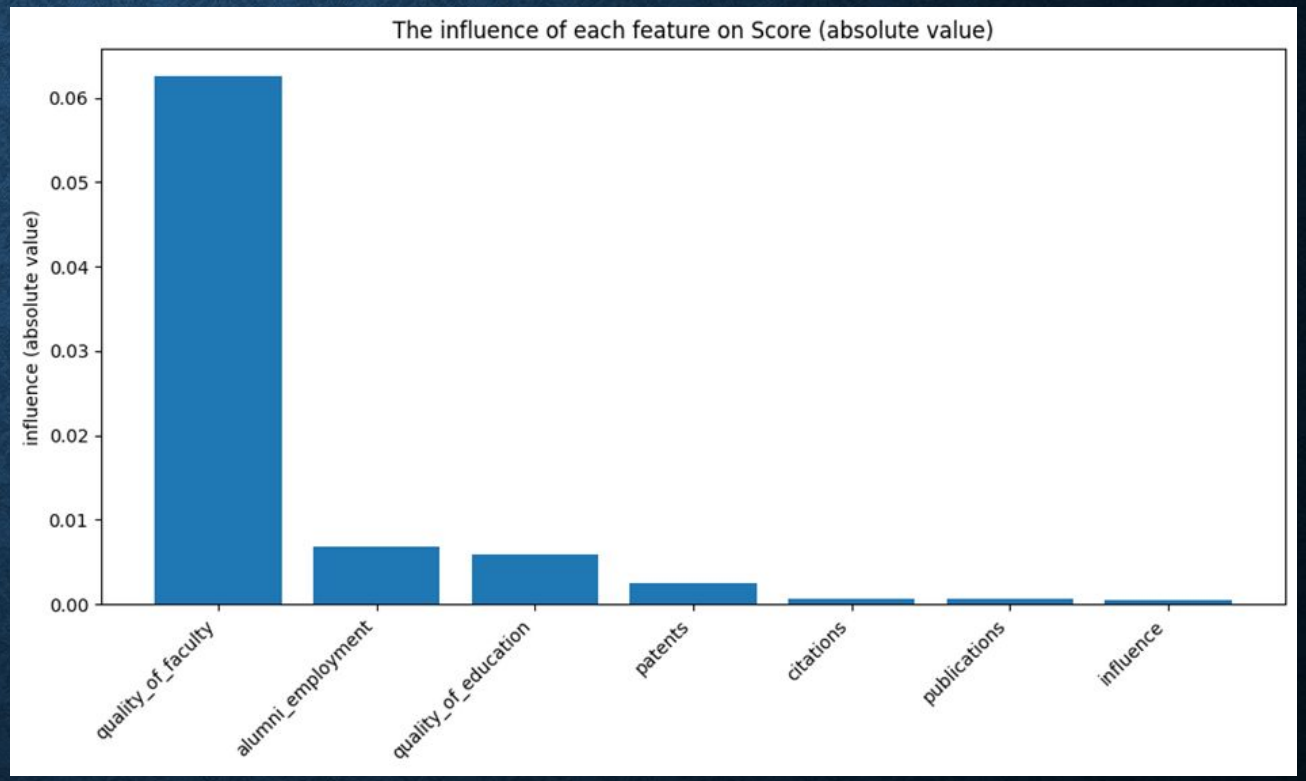
2. Which Indicator Influences Rank Most?

- **Method: Relationship Analysis (Regression & Correlation Coefficients)**
  - **Most influential factor**
    Quality of Faculty stands out as the top indicator, with a significantly stronger impact than others.
  - **Next in line**
    Alumni Employment and **Quality of Education** also show strong positive influence.
  - **Less impactful**
    Traditional academic indicators like **Citations**, **Publications**, and **Influence** are relatively weaker in this model.

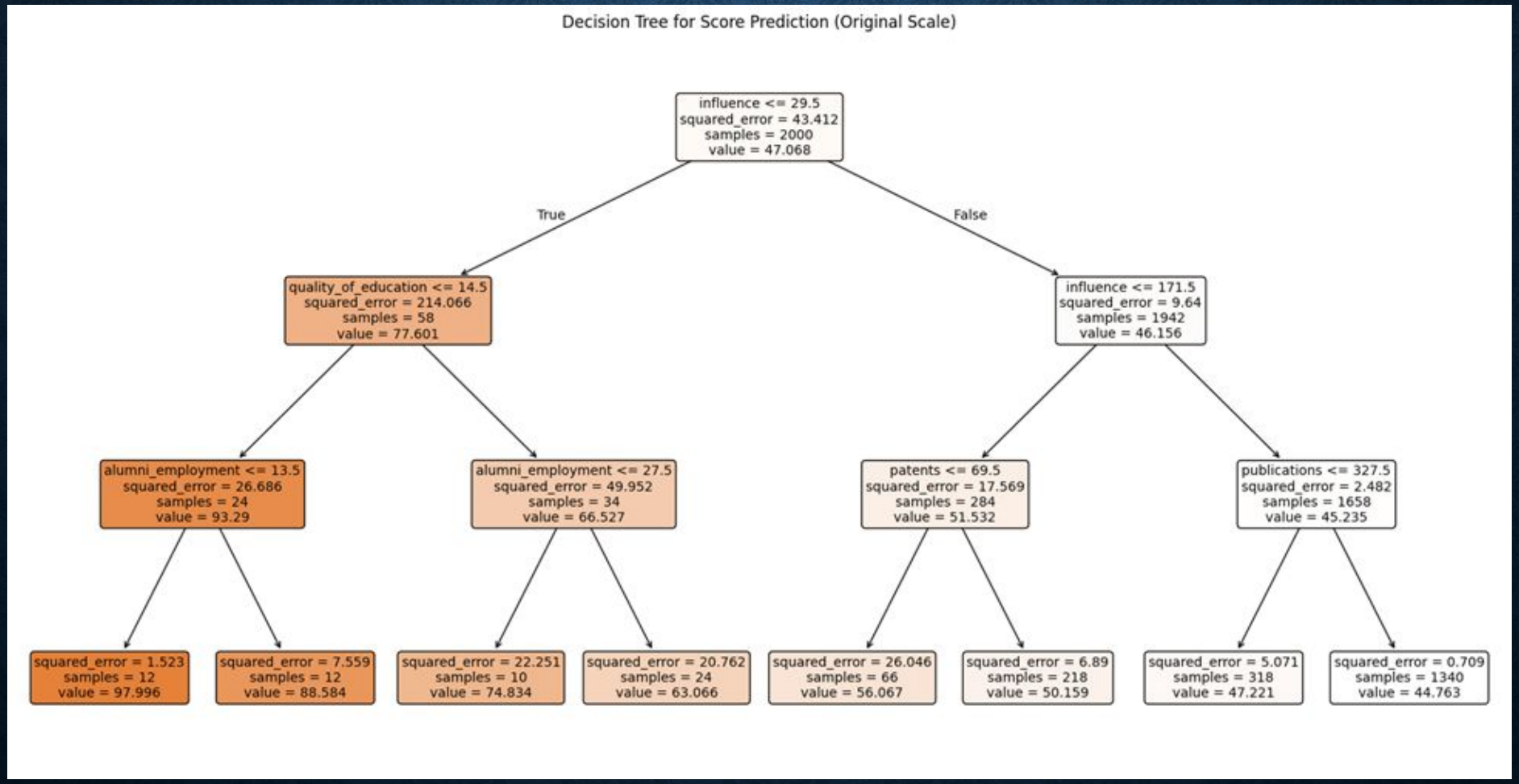


2. Which Indicator Influences Rank Most?



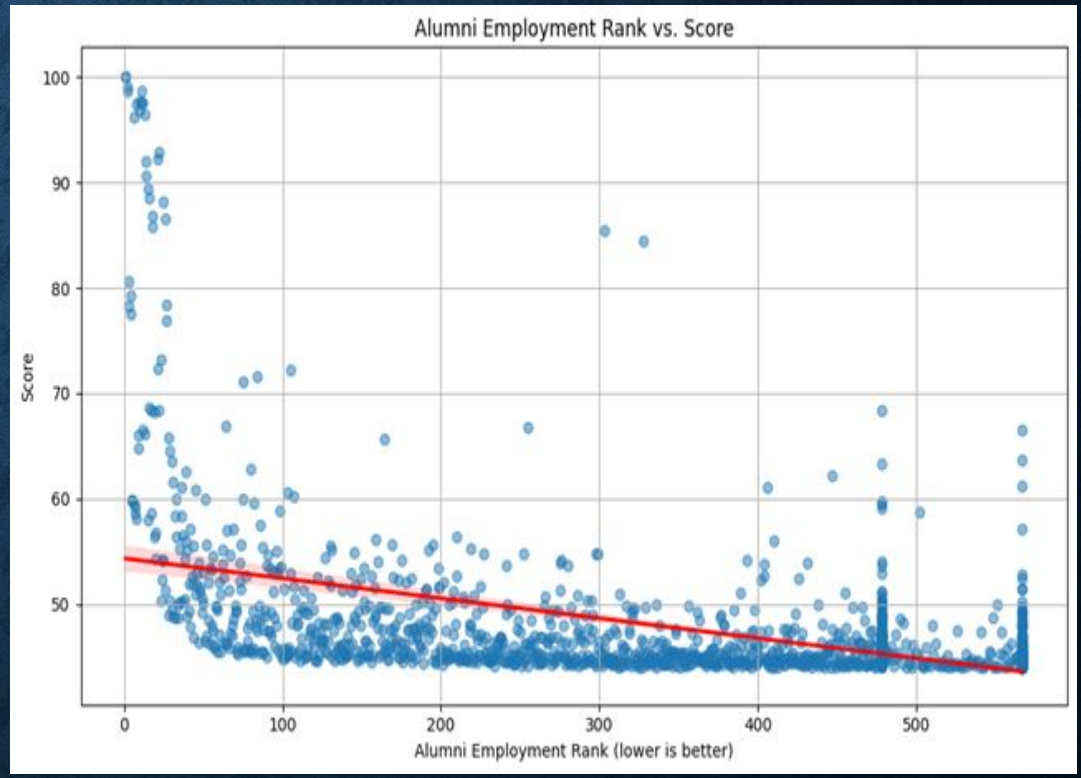


2. Which Indicator Influences Rank Most?



3. Does Better Employment Lead To Higher Score?

- **Method: Correlation + Regression Analysis**
  - **Independent Variable:** Alumni Employment Ranking
  - **Dependent Variable:** Total Score
  - **Finding**
    Regression results show a **significant negative correlation**, meaning
    → Better alumni employment → Higher overall score

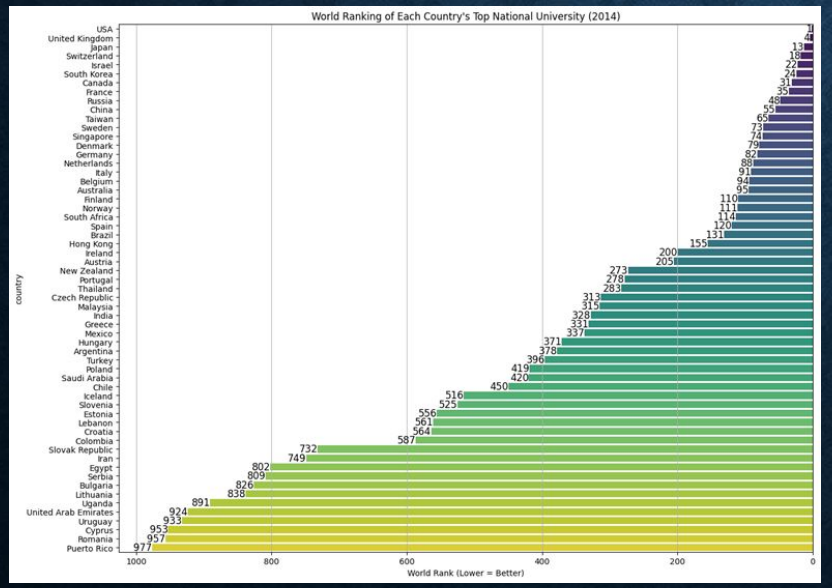


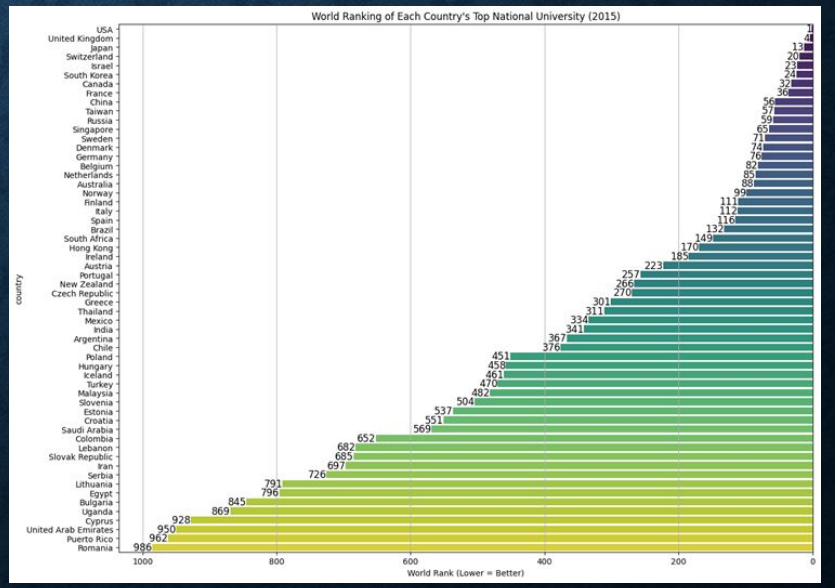
4. Do Specific Countries Rank Higher?

- **Method: Data manipulation and visualization**

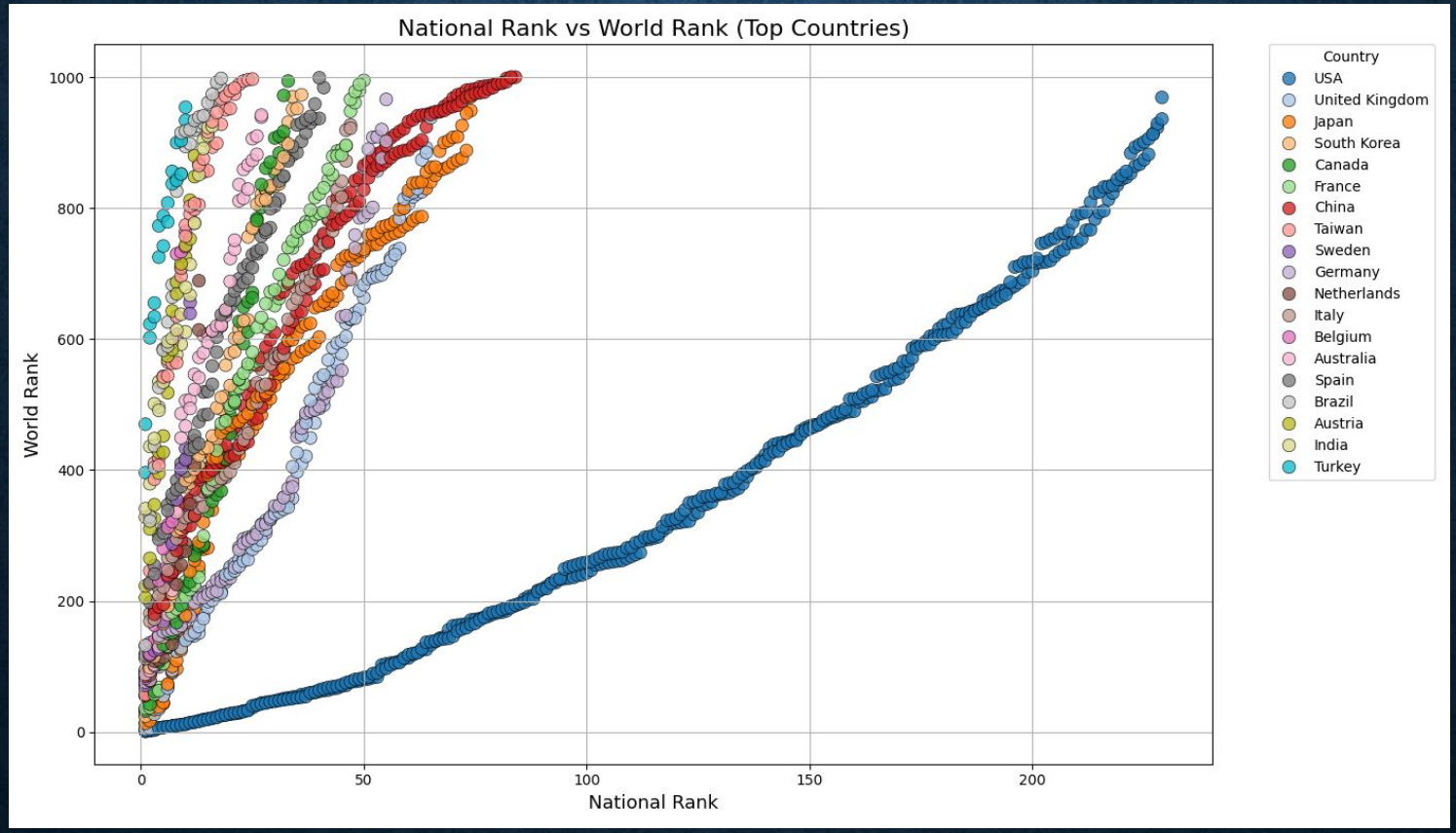
  By grouping universities by country, select and compare top university of each country, we found

  - **US and UK** dominate in both average ranking and citations, showing global academic leadership.
  - **Asian countries** like China and Japan have similar indicator levels but slightly lower rankings.
  - This suggests that even with comparable educational input, **visibility and systemic factors** still influence ranking outcomes.



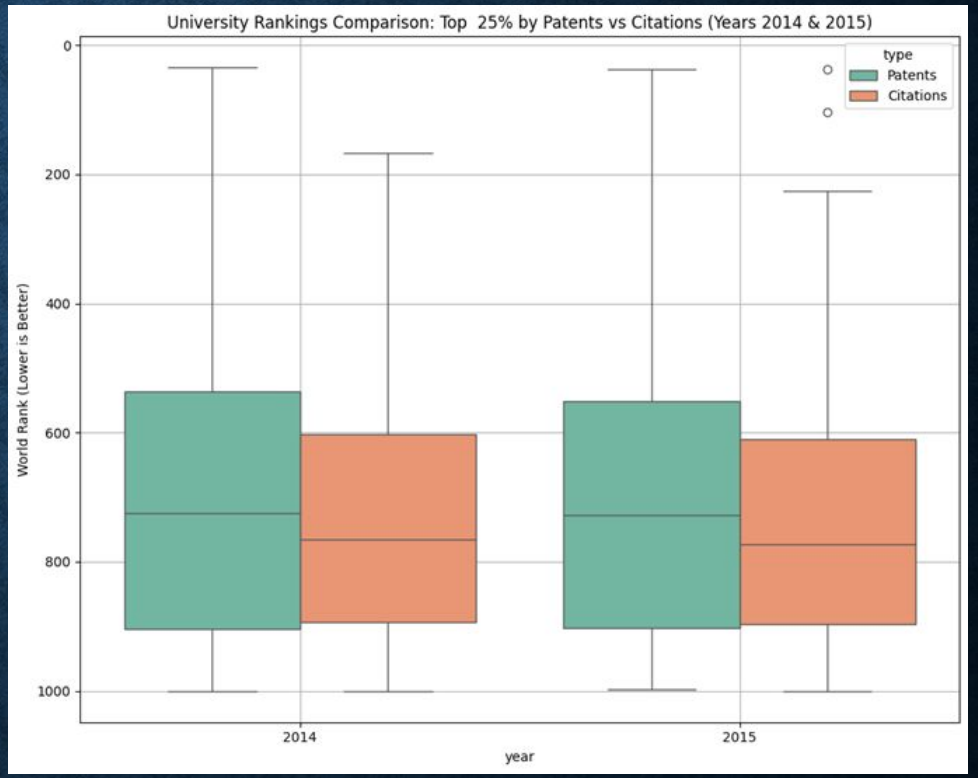


4. Do Specific Countries Rank Higher?



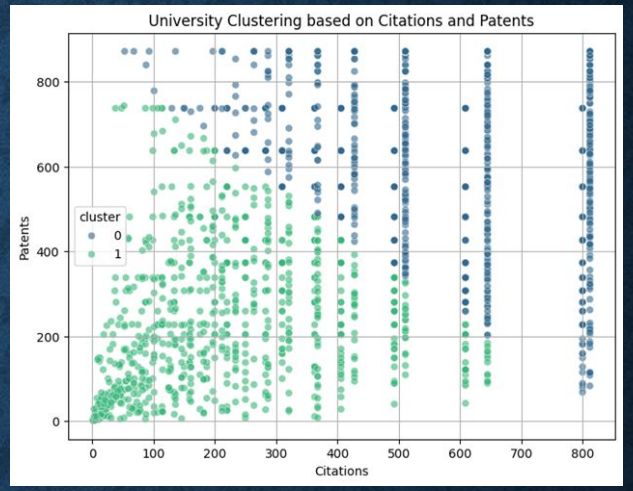
5. Which Type Of University Ranks Higher?

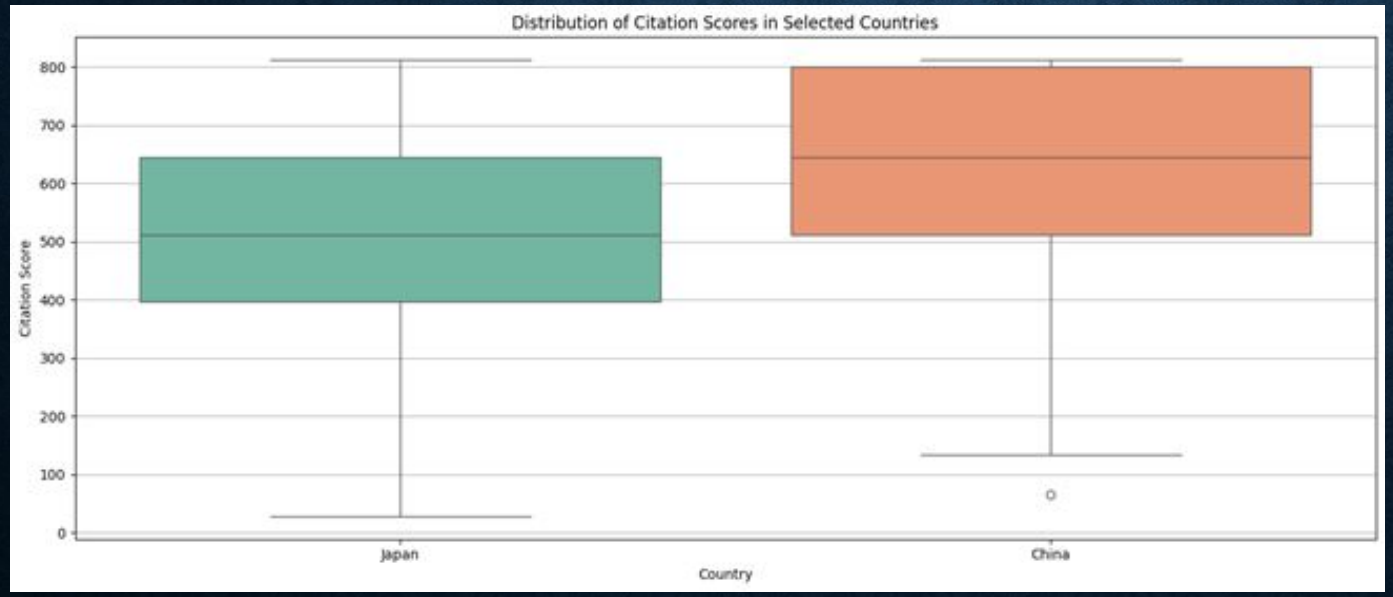
- **Method: Comparative Group Analysis**
  - Universities grouped by **high citations** vs. **high patents**
  - **Findings**
    - **High-citation universities** rank significantly higher globally → Traditional academic output remains dominant.
    - **High-patent universities** excel in innovation and tech-transfer metrics → But have limited impact on overall rankings.

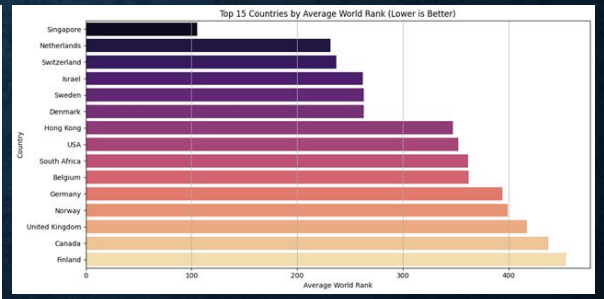


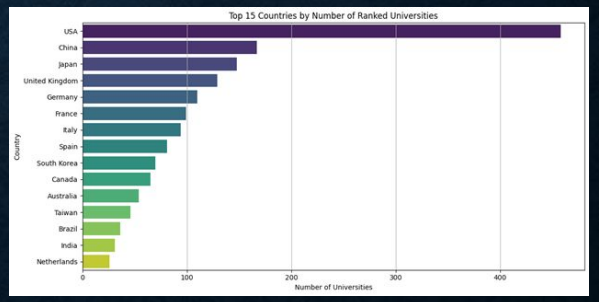
6. Classify By Citations And Patents.

- **Method: K-Means & Hierarchical Clustering**
  - Features: **Citations**, **Patents**, **Publications**
  - Identified 3 distinct university types:
    - **Academic-Oriented** (high citations & publications)
    - **Innovation-Oriented** (high patents)
      Balanced Type (moderate across all metrics)









06

Conclusions

CONCLUSIONS

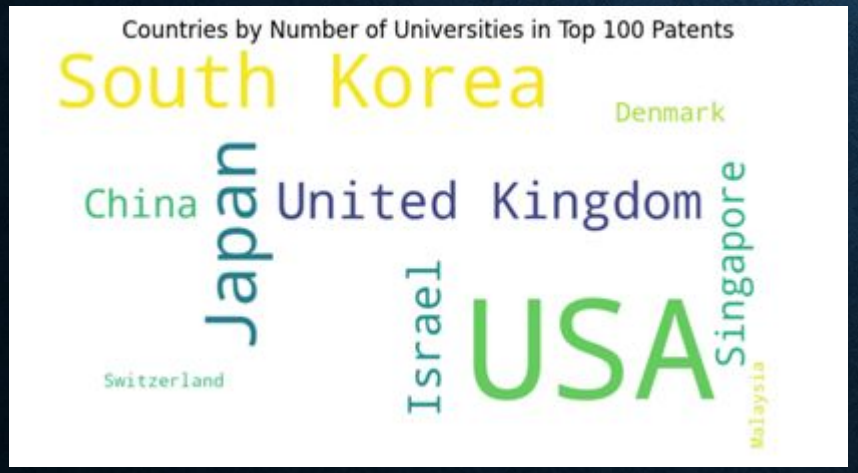
- **Citations, faculty quality, and alumni employment** are the strongest predictors of global rankings.
- **Data visualization** uncovers meaningful national and institutional differences.
- **Structural biases**—such as language and publication venue—may skew rankings.
- **Clustering reveals** distinct university types: academic, innovative, and balanced.
- **Takeaway:**
  Data science not only answers the research questions but also exposes hidden biases in global ranking systems, offering guidance for strategic decision-making.

07

Others (Bonus)

OTHERS (BONUS)

- To enhance the depth and creativity of our analysis, we added a **keyword cloud** to visualize the visibility of different academic themes across countries.
- We also experimented with integrating **Large Language Models (LLMs)**, combining Wikipedia API, web scraping, and translation tools to build a semantic-enhanced system. Users can input a university name and receive both English and Chinese summaries.
- Results show that combining LLMs with open APIs breaks the limitations of static data and adds semantic depth and flexibility to the analysis.




```
🔸 Massachusetts Institute of Technology
📘 The Massachusetts Institute of Technology (MIT) is a private research university in Cambridge, Massachusetts, United States. Establishe
In response to the increasing industrialization of the United States, William Barton Rogers organized a school in Boston to create "useful

🔸 Harvard University
📘 Harvard University is a private Ivy League research university in Cambridge, Massachusetts, United States. Founded October 28, 1636, an

🔸 Stanford University
📘 Summary not available for Stanford University: Page id "standard university" does not match any pages. Try another id!
```

```
🔸 大學名稱: Massachusetts Institute of Technology
   (中文翻譯: 麻省理工學院)
📘 摘要內容:
The Massachusetts Institute of Technology (MIT) is a private research university in Cambridge, Massachusetts, United States. Established j
In response to the increasing industrialization of the United States, William Barton Rogers organized a school in Boston to create "useful
   (中文翻譯: 麻省理工學院（MIT）是一所私立研究型大學，位於美國麻省劍橋市。麻省理工學院成立於 1861 年，在現代技術和科學的許多領域的發展中扮演了重要的角色。
為了因應美國日益增長的工業化，William Barton Rogers 在波士頓組織了一所學校，以創造 「有用的知識」。該學院最初由聯邦土地補助金資助，採用理工學院模式，強調應用科學

🔸 大學名稱: Harvard University
   (中文翻譯: 哈佛大學)
📘 摘要內容:
Harvard University is a private Ivy League research university in Cambridge, Massachusetts, United States. Founded October 28, 1636, and r
   (中文翻譯: 哈佛大學（Harvard University）位於美國麻薩諸塞州劍橋市，是一所私立常春藤盟校研究型大學。哈佛大學成立於 1636 年 10 月 28 日，以其第一位恩人、清教
```

Thank You

If you have any questions, please feel free to contact us by email.

t { 113598043, 113598032, 113598090, 113598091, 113998411 } @ntut.org.tw