National Taipei University of Technology

Computer Science and Information Engineering

Data Science Principles with Applications on

Educational Data

Spring 2025

Semester Group Project Report

**教育數據中的關鍵因素探索：以大學排名為例**

**Exploring Key Factors in Educational Data:**

**A Case Study on University Rankings**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name: | 張育丞, | 張字青, | 周姿妤, | 黃詳諺, | Duong Van Nhat Quang |
| Sid: | 113598043, | 113598032, | 113598090, | 113598091, | 113998411 |

Date:06/11/2025

**Content**

**Abstract**

1. **Introduction**

近年來，全球大學排名的重要性日益提升，對高等教育政策、資源分配與大學發展策略產生了深遠影響。對於希望提升國際競爭力與社會影響力的學術機構而言，深入理解排名背後的評估指標已成為不可或缺的一環。

大學排名通常根據多項教育指標進行評估，包括學術產出、創新能力、研究影響力、校友成就及機構聲譽等。隨著教育數據日趨豐富，這些資訊為我們提供了理想的基礎，以分析各項指標之間的關聯性，並探索它們如何共同影響全球排名表現。

本專案旨在透過資料科學方法，探討各項教育指標與大學排名之間的綜合關係，挖掘其中的關鍵模式，進而提出具體的策略建議，協助學術機構掌握優化方向。

本研究的主要目標包括：

* 辨識並分析對全球大學排名影響最顯著的關鍵教育指標。
* 整合多項特徵（如引用量、專利數、論文發表數量），探討其綜合效應。
* 透過視覺化方式呈現指標之間的關聯性與潛在模式，提升資料可解讀性。
* 辨識並分析對全球大學排名影響最顯著的關鍵教育指標。
* 提供以數據為基礎的分析結果，協助理解並評估大學的整體競爭力。

1. **Literature review and related works**

本研究將使用資料集為Kaggle 平台上公開之「World University Rankings」的資料集，該資料集具有2014與2015年的全球大學排名資訊，並提供多元且詳盡的教育評估指標，進而探討大學排名與較於實力間的關聯性，這也為本報告提供良好的資料基礎，我們將於表一進行欄位的說明。

表一 欄位及欄位說明 （來源：Kaggle平台）

|  |  |  |
| --- | --- | --- |
| 欄位名稱 | Kaggle 說明 | 說明 |
| world\_rank | world rank for university. | 世界大學排名 |
| institution | name of university | 學校名稱 |
| contry | country of each university | 國家 |
| national\_rank | rank of university within its country | 國家排名 |
| quality\_of\_education | rank for quality of education | 教育品質排名 |
| alumni\_employment | rank for alumni employment | 校友就業表現指標排名 |
| quality\_of\_faculty | rank for quality of faculty | 師資品質排名 |
| publication | rank for publications | 論文發表量排名 |
| influence | rank for influence | 學術影響力排名 |
| citations | number of students at the university | 大學學生人數 |
| broad\_impact | rank for broad impact  (only available for 2014 and 2015) | 廣泛影響力排名  （範圍：2014與2015年） |
| patents | rank for patents | 專利數量排名 |
| score | total score, used for determining world rank | 綜合分數（用於排名） |
| year | year of ranking | 年份排名 |

「World University Ranking」資料整合了教育成果、創新能力與學術榮譽等多個層面的指標並且具有多個年份，相較於其他年份區間，在2014與2015年的資料較為完整，以兩年完整資料進行比較，將具備高度的研究價值，因此成為本報告進行分析兩者數據的重要依據。

1. **Problem statement**

現今已有多個全球排名，包含QS世界大學排名與THE高能教育排名等，對於各世界大學排名在評估時，評估指標的使用與其對應的權重都有不同的見解及分析方式。對於本報告將使用「World University Rankings」作為基底，團隊將對此整理表一的13項指標，並規劃出6大核心問題進行分析，並於表二呈現6大核心議題。

表二 核心議題

|  |  |  |
| --- | --- | --- |
| 題目 | 問題說明 | 方法 |
| 1. 論文引用量是否影響世界排名？ | 被引用越多，世界排名是否越前？ | 相關分析 + 線性迴歸 |
| 2. 校友就業力是否推動得分？ | 有高就業力的學校，是否得分也會高的？ | 相關分析 + 迴歸 |
| 3. 哪個指標最影響世界排名？ | 教育品質、師資、引用量、專利，誰最影響排名？ | 關係性分析 (關係權重) |
| 4. 國家對排名的影響力？ | 特定國家是否在相同分數下排名更高？ | 群組分析 (GroupBy) |
| 5. 高專利產出 vs 高學術影響力大學比較 | 多專利的大學和多引用的大學，誰的排名高？ | 分組比較 |
| 6. 學術 vs 創新型大學的分類 | 根據論文引用、專利將大學分類 | 分群 (Clustering) |

1. **Proposed models (approaches)**

本報告旨在探討13項教育指標對於世界大學排名之影響，針對資料集的處理將依據圖Ｘ，進行資料集預處理、13項欄位的關聯性探討，接著進行分析轉為教育指標，並對其數據並可視化。

1. 資料預處理（Data Pre-processing）

這步驟是個相當重要的一個環節，除了在進行資料的檢視，經由檢視將異常的資料進行過篩，也在將資料轉換成為較容易處理的狀態，避免後續藉此降低後續資料取得與分析的門檻，其操作將依循以下三步驟。

* 1. 檢視與解析資料內容。
  2. 移除缺漏資料。
  3. 清除不一致或抵觸的資料。

1. 變數關聯性分析（Correlation Analysis）
   1. 探討各13項教育指標與排名之間的相關程度。
   2. 使用皮爾森係數與熱力圖進行初步視覺化。
   3. 線性與多元回歸分析。
   4. 分析單一及多個教育指標對於大學排名的解釋能力。
2. 關聯權重與重要度排序（Feature Importance Ranking）
   1. 使用回歸係數、逐步回歸。
   2. 繪製特徵圖示。
3. 國家與地區群組分析（Grouping by Country）
   1. 探討
4. 分組比較分析（Group Comparison）
5. 分群模型建構（Clustering）

* - 國家與地區群組分析（Grouping by Country）
* -探討是否特定國家在相同得分下排名更高
* -使用 GroupBy 技術比較各國平均排名與指標平均值
* -分組比較分析（Group Comparison）
* - 將大學依據高專利數與高引用量分類為不同群組，進行排名差異比較（t-test 或 ANOVA）
* - 分群模型建構（Clustering）
* - 依據引用量、專利數、發表數等特徵，使用 K-Means 或階層式分群（Hierarchical Clustering）將大學劃分為「學術導向型」與「創新導向型」等類別

- - 分析各群在排名與其他指標上的特性差異

1. **Experiments**

(放結果圖，並將結果圖進行稍微的分析)

1. 論文引用量是否影響世界排名？(黃)

(相關性分析 ＋線性回歸)

1. 校友就業力是否推動得分？(周)

(相關性分析 ＋ 回歸)

1. 哪個指標最影響世界排名？(黃)

(關係性分析-關係全中)

1. 國家對世界排名的影響力？(Quang)

(群組分析 Group By)

1. 高專利 vs. 高學術影響力大學比較 (青)

(分組比較)

1. 學術 vs. 創新型大學的分類 (丞)

(分群)

1. **Conclusion**

Please include what the project has done and whether the objectives are achieved. What do you get in the project and what can be done in the future.

1. **Others（Future Work）**

Please state the workload and role of each member in your team for the project. Besides, show the timeline for the project and check whether all the proposed works have been done.

(文字雲....)

**Reference**