

BIG DATA VISUAL ANALYTICS (CS661)

PROJECT PROPOSAL REPORT

GROUP- 06

World University Rankings Analysis & Visualization

1. Introduction

University rankings serve as a crucial benchmark for academic performance, helping students, policymakers, and researchers assess the quality of institutions worldwide. This project aims to analyze and visualize trends in university rankings from **2016 to 2025**, focusing on factors such as teaching quality, research impact, student diversity, industry collaboration, etc.

We will create an **interactive Web-based system** to explore dynamic ranking trends using Python-based tools (Matplotlib, Seaborn, Plotly, Dash, Streamlit).

2. Data Source

The dataset will be sourced from [Times Higher Education](#) and also available on [Kaggle](#). It contains university rankings from **2016 to 2025** with the following attributes:

SNo	Attributes	Descriptive information
1	University Rank & Name	The global ranking of institutions
2	Country	The country of each university
3	Student Population	Total number of students enrolled
4	Student-to-Staff Ratio	Faculty availability for students
5	International Students	Percentage of international students
6	Female-to-Male Ratio	Gender distribution of students
7	Overall Score	Composite ranking score
8	Teaching Score	Rating based on teaching quality.
9	Research Environment Score	Research facilities & output.
10	Research Quality Score	Research impact in citations.
11	Industry Impact Score	University collaboration with industries.
12	International Outlook Score	Diversity in faculty & students.
13	Year (2016-2025)	Year-wise rankings.

3. Key Tasks & Visualizations

We will perform **exploratory data analysis (EDA)** of the dataset and generate **interactive visualizations** for meaningful insights.

We will tentatively perform the following tasks:

1. Global Ranking Trends Over Time

- a. **Visualization:** Line chart showing the ranking progression of top universities from 2016 to 2025.
- b. **Insights:** Identifies **consistent top performers** and **fluctuating universities**, helping in trend forecasting.

2. Country-Wise University Count & Performance

- a. **Visualization:** Bar chart displaying the **number of ranked universities per country** & their average overall score.
- b. **Insights:** Highlights **which countries dominate** university rankings and their overall academic performance.

3. Top 10 Universities Over Time

- a. **Visualization:** Interactive table with year-wise **ranking filters**.
- b. **Insights:** Provides a **quick comparison** of top universities across different years.

4. University Rankings by Region

- a. **Visualization:** Interactive world map showing **ranking distribution by continent**.
- b. **Insights:** Helps analyze **regional performance variations** in higher education.

5. Student Population vs. Rank Correlation

- a. **Visualization:** Scatter plot showing **student population vs. rank**.
- b. **Insights:** Helps understand **if more prominent universities** tend to rank higher or if size impacts academic excellence.

6. Student-to-Staff Ratio Impact on Teaching Score

- a. **Visualization:** Correlation heatmap and scatter plot.
- b. **Insights:** Examines **whether lower student-to-staff ratios lead to better teaching scores**, aiding universities in optimizing faculty size.

7. International Student Distribution

- a. **Visualization:** Interactive **world map** with color-coded density of international students.
- b. **Insights:** This shows which universities attract the most **global talent** and are helpful for international applicants.

8. Female-to-Male Ratio Trends

- a. **Visualization:** Pie/bar charts for **gender distribution across years**.
- b. **Insights:** Helps monitor progress in **gender diversity** and identifies universities excelling in gender balance.

9. Research Quality vs. Overall Score

- a. **Visualization:** Regression analysis or scatter plot.
- b. **Insights:** Determines whether universities with **higher research quality rank higher**, providing insight into **research-driven rankings**.

10. Industry Impact Score vs. Research Environment Score

- a. **Visualization:** Bubble chart for displaying **industry collaborations vs. research output**.
- b. **Insights:** Shows whether universities with strong industry ties **also have robust research environments**.

11. Multi-Filter Dashboard for Custom Analysis

- a. **Visualization:** Dashboard with filters (country, rank range, scores).
- b. **Insights:** Allows users to **explore data** based on their interests dynamically.

12. Some Unsupervised ML Tasks

- a. **Visualization:** Cluster analysis based on different features, 3D Scatter plot after PCA, etc

4. Implementation Summary

Data Processing Steps

- **Data Cleaning:** Handling missing values, standardizing formats, etc.
- **Feature Engineering:** Creating derived metrics like score-to-rank correlation, etc.
- **Data Aggregation:** Summarizing by country, region, year, etc.

Visualization Libraries & Tools

- **Python** Pandas, NumPy (data processing), Matplotlib & Seaborn (Visualizations)
- **Python** scikit-learn (ML tasks)
- **Plotly & Dash / Streamlit** (Interactive charts & Visualization dashboards)
- **Django / Flask** (backend development), **SQL / PostgreSQL** (Database if needed)

5. Interface Design

We will create an interactive web-based system mainly using Python-based tools (Matplotlib, Seaborn, Plotly, Dash, Streamlit, etc.) to showcase different visualization tasks and explore dynamic ranking trends (hosted on a suitable platform as per requirements).

Crafting an Engaging Data Story

Creating a compelling narrative using the visualizations to explore global university ranking trends from 2016 to 2025 highlighting key insights, including ranking fluctuations, regional performance, the impact of student-to-staff ratios on teaching quality, and research quality correlations with overall scores etc. This will help students, policymakers, and researchers gain deeper insights into higher education dynamics worldwide.

6. Task Distribution (8 Members)

Tentative task distribution is as follows:

Data Collection & Cleaning, Report & Documentation

- Senthil Ganesh (241110089)
- Rajan Kumar (241110087)

Exploratory Data Analysis (EDA) and Static Visualizations

- Hritik Chouhan (241110030)
- Divyanshu (241110023)
- Richik Majumder (241040068)

Interactive Visualization and Dashboard Development: Frontend/Backend Tasks

- Khushwant Kaswan (241110034)
- Rishit Kumar (241110056)
- Deepak Soni (241110018)