



Education, Economy, and Society:

Correlations Between Universities and Global Development

Rishika Deshmukh
Adiba Hasnat
Ananya Sai

Data Selection

Objective: Investigate the relationship between university ranking, count, and international student percentage in countries and various global development scores.

Datasets (from Kaggle):

- **University Variables:** university name, country, international student percentage, rankings.
- **Global Development Variables:** country, different societal scores including safety and security, personal freedom, governance, social capital, economic quality, living conditions, health, education, natural environment, and the average score of all these metrics.

Rationale :

- Universities play a key role in economic and educational development, attracting international students and contributing to global mobility.
- Global development indicators provide insight into how a country's overall well-being correlates with its educational system.

Expected Insights

1. Do countries with a higher number of ranked universities tend to have better societal well-being?
2. Do countries in high university count also have higher average societal scores?:
3. Do economic quality and education scores correlate with university count for the top 10 countries?
4. Investigate whether economic factors correlate to the percentage of international students at universities.

ETL Process

EXTRACTION

- Data sourced from Kaggle via CSV files

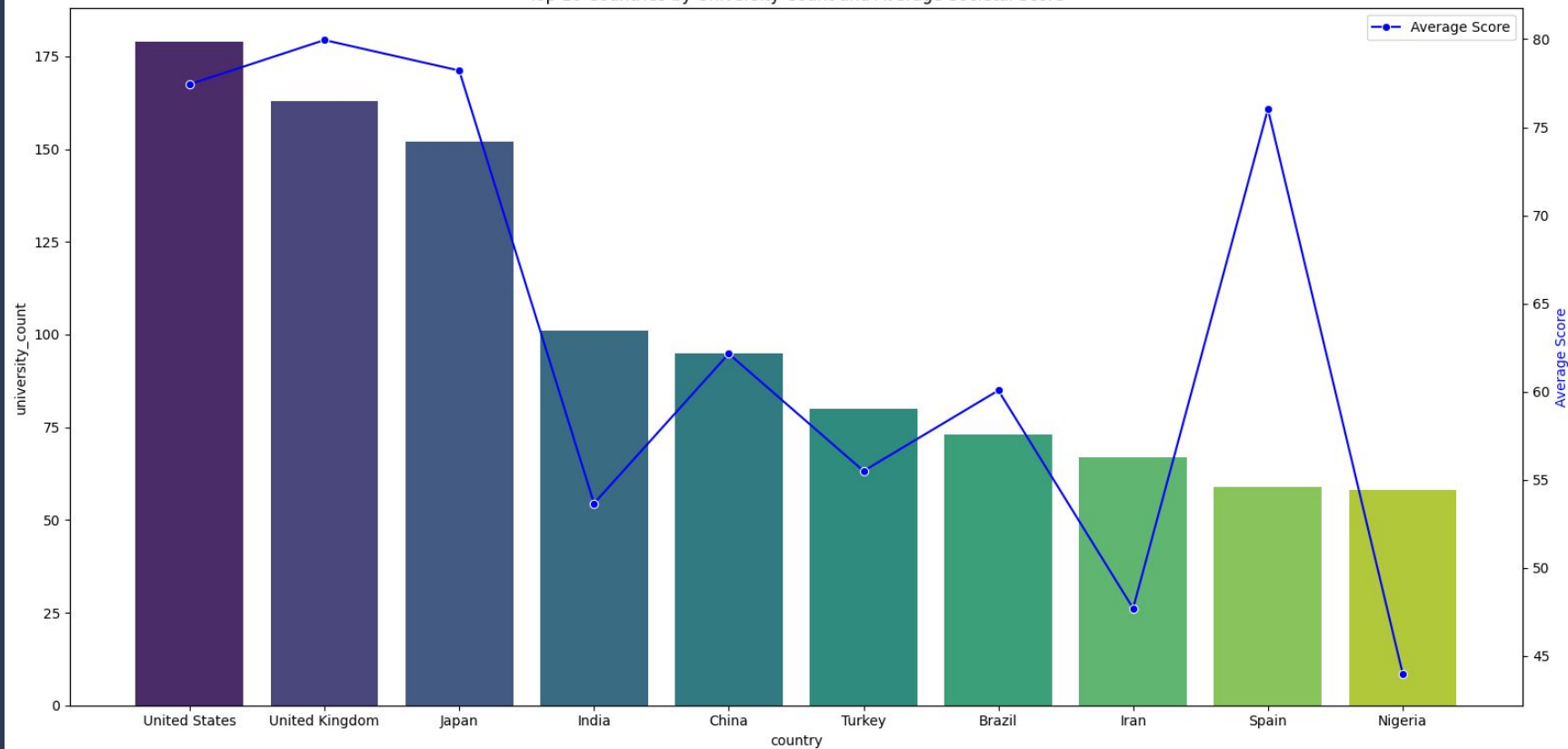
TRANSFORMATION

1. Removed missing and irrelevant data.
2. Focused on relevant variables through conducting EDA: country, university rankings, societal scores.
3. Filtered top universities by international student percentage and university count.
4. Merged datasets together with the country column.

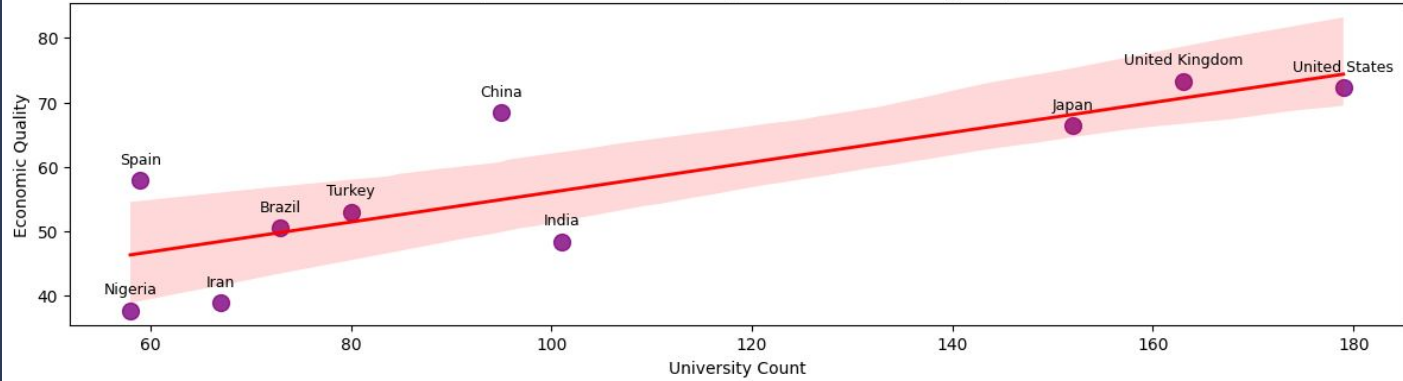
LOAD

1. Cleaned data sets stored in CSV files
2. Loaded data into MongoDB for easy access
3. Set up Google Cloud storage for transformed data

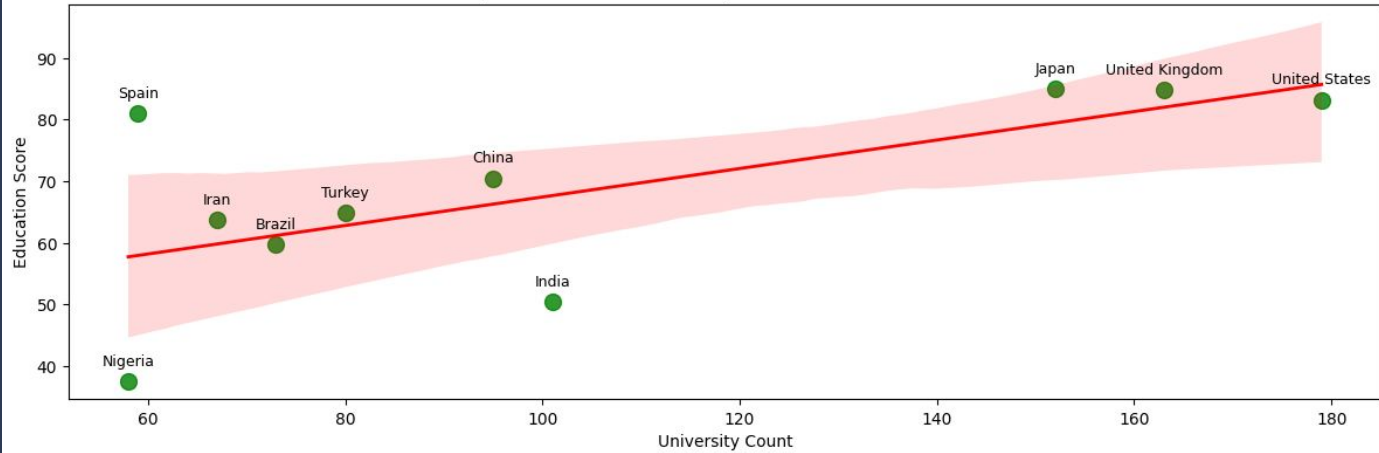
Top 10 Countries by University Count and Average Societal Score



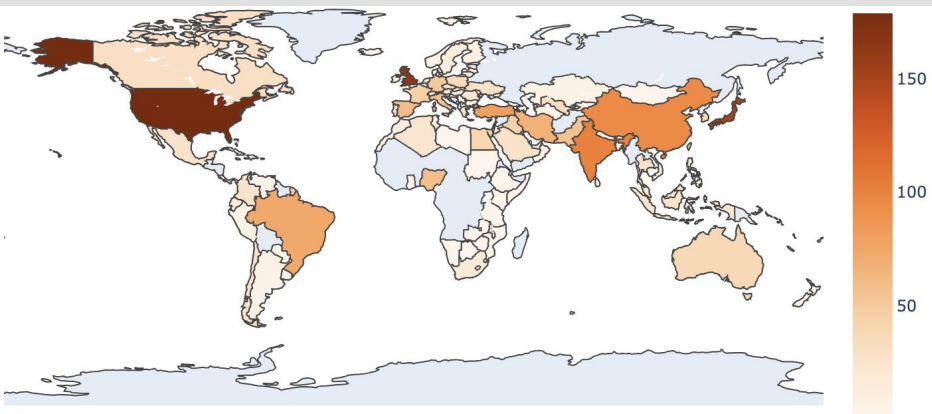
Top 10 Countries: University Count vs Economic Quality



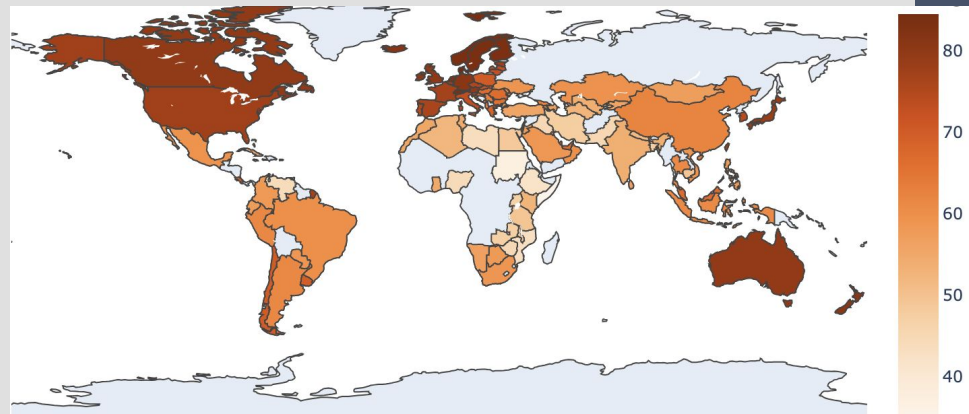
Top 10 Countries: University Count vs Education Score



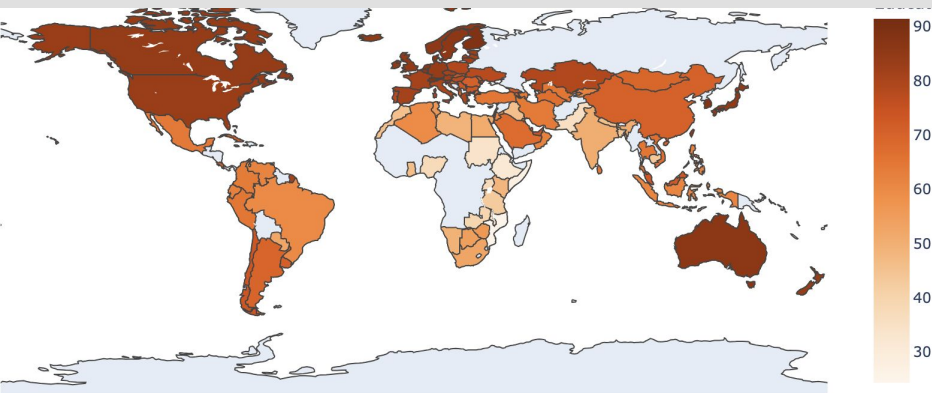
Number of Ranked Universities



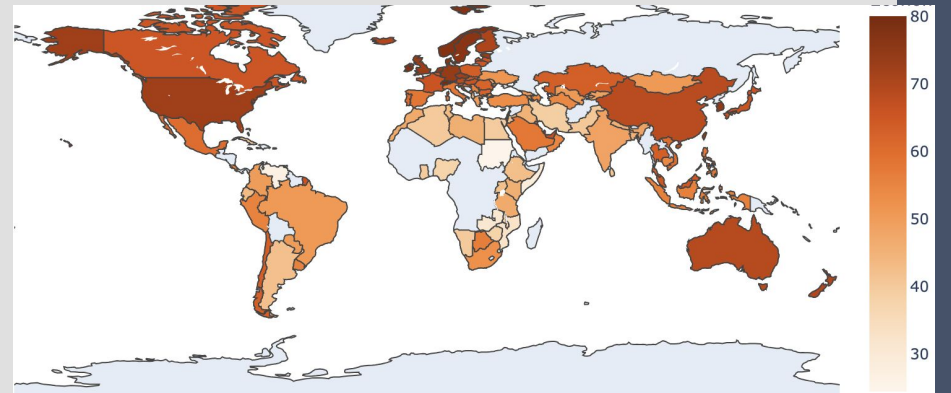
Average Development Score

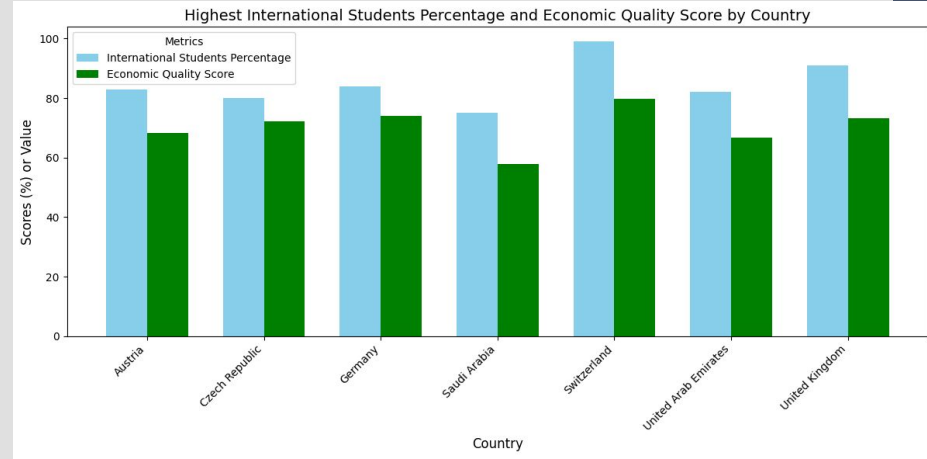
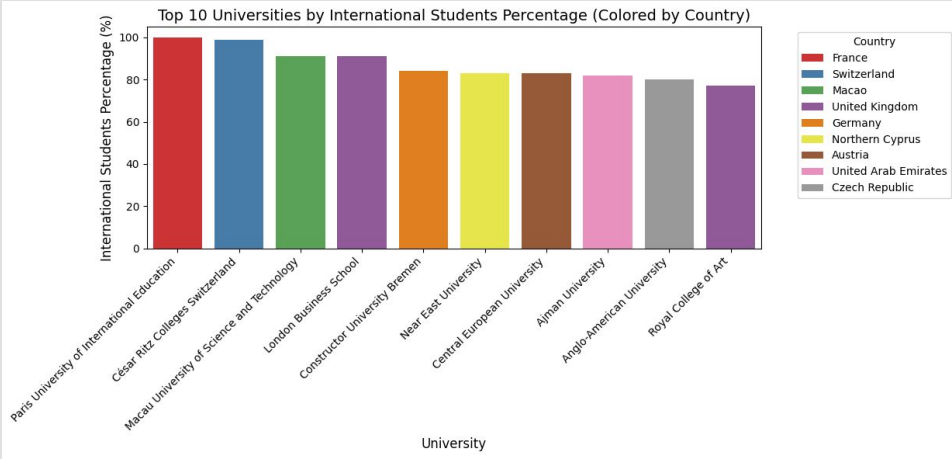


Education Score



Economic Quality Score





Cloud Storage

- Created a **Google Cloud bucket** named "global_datasets" to store cleaned datasets
- Set up a **service account** with Storage Admin role and downloaded the JSON key file for secure access
- Uploaded the **JSON key** file to Google Colab for authentication
- Used Python (google-cloud-storage) to **authenticate and connect** to the bucket
- Uploaded cleaned datasets to the bucket for centralized storage and sharing
- Tested programmatic access by listing and downloading files from the bucket

Analysis

University Count and Societal Score

- More ranked universities often correlate with higher societal well-being (e.g., US, UK, Japan).
- Exceptions like Spain (high societal scores, fewer universities) and India (many universities, lower scores) suggest other factors, like educational quality and socio-economic conditions

University Count vs. Economic and Education Scores

- Positive correlation between university count and economic/education scores—countries investing in education tend to have stronger economies
- Discrepancies in countries like India and Nigeria highlight that quantity of universities alone doesn't guarantee high-quality outcomes; factors like governance, resources, and quality disparity matter.

International Students Analysis

- Strong economies attract more international students, shown by high economic quality and student percentages.
- Economic scores are slightly lower than student percentages, but the trend still aligns

Challenges

Technical

- Selecting datasets that give meaningful insights for visualization
 - With a special emphasis on the amount of variables and type of variables
 - The type of data we wanted to focus on
- Handling missing or inconsistent data across multiple datasets was challenging especially when merging datasets (especially finding columns to merge two independent datasets on)
- Mounting the datasets on MongoDB

Analytical

- While correlations between university count, societal scores, and economic factors were observed, several outliers and exceptions made it challenging to establish clear relationships.
- Societal well-being and university success are influenced by numerous factors beyond the datasets, such as political stability, cultural aspects, and regional disparities, complicating the analysis and interpretation of results.

Key Takeaways

- Ensuring relevant data is clean and structured was crucial for meaningful analysis
- Choosing a dataset with a good number of features enabled us to be able to not be overwhelmed in our analysis but also gave us room to compare and contrast different variables that could challenge our preconceived notions
- It is good to have a static container to hold data in so that you can access it
- While a positive correlation between university count and societal scores was observed, exceptions highlighted that other factors like economic and educational quality also play significant roles in societal development.
- Our analysis showed that strong economies attract more international students (top ten schools for international students), suggesting there is a possibility for a mutual relationship where international students contribute to, as well as benefit from, a country's economic growth and educational infrastructure.

Thank you!

