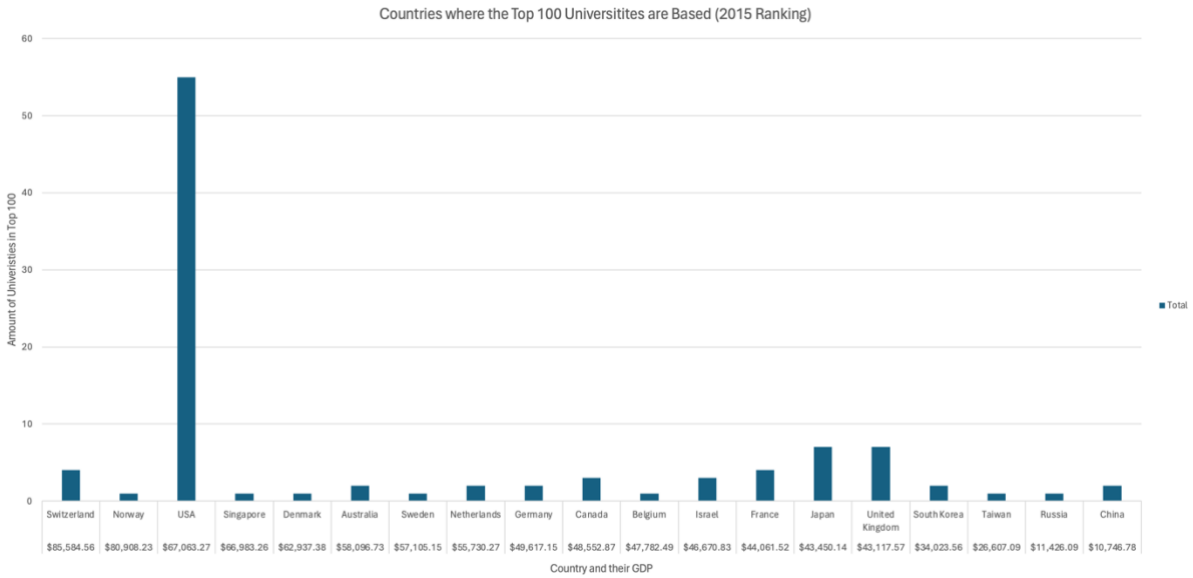
Global Education and Innovation Analysis Project Report

1. Project Summary

Question 1: Do countries with a higher-than-average GDP per capita have more universities ranked in the top 100 in 2015?

I found that the average GDP Per Capita for all the countries in the provided dataset is $31638.03.

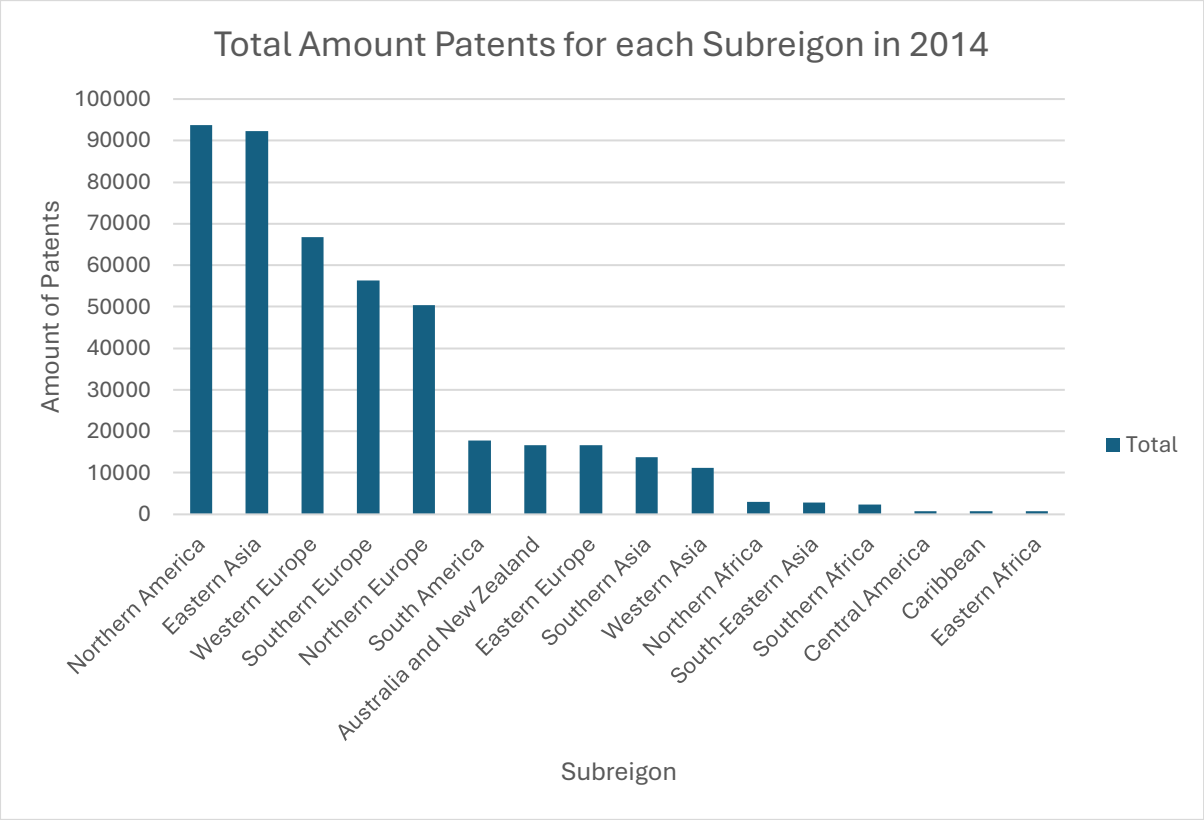


Based on the analysis of the provided graph, there is a clear correlation between a country's GDP and the presence of its universities in the top 100 rankings. Countries with higher GDPs tend to have a greater number of universities in these prestigious rankings. Specifically, only three countries with GDPs lower than the global average have universities in the top 100: Russia and Taiwan with one university and China with two. This represents a mere 4% of the top 100 universities.

The graph highlights that out of 19 countries represented, 40 countries did not have any universities in the top 100 in the 2015 rankings. Notably, the United States accounts for over 50% of the top 100 universities, while also having the third highest GDP among the countries on the graph.

In conclusion, the data clearly indicates that countries with a higher GDP per capita are more likely to have a greater number of top-ranked universities in 2015.

Question 2: What Subregion Produced the Most Patents in 2014?

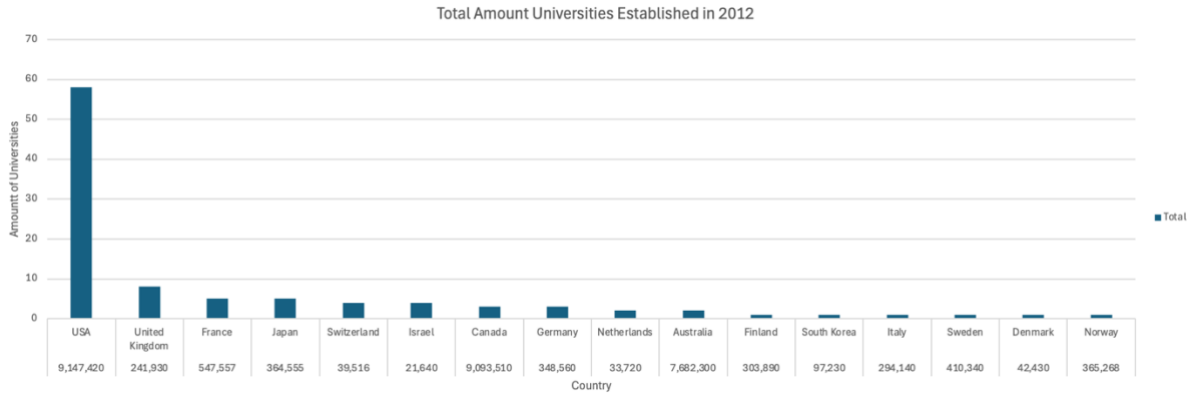


The provided data reveals that Northern America and Eastern Asia are leading in global innovation, with 93,685 and 92,267 patents filed respectively in 2014. Western Europe also shows significant activity with 66,728 patents, highlighting robust R&D efforts in these regions.

Emerging regions like Southern Asia and Eastern Europe, with 13,778 and 16,691 patents respectively, present new opportunities for market entry and expansion. These regions are showing notable growth in innovation, making them attractive for businesses seeking to capitalize on new markets. Monitoring these trends can help anticipate future competitive landscapes and technological advancements.

Subregions with lower patent activities, such as Northern Africa and Southern Africa, indicate potential for development and growth. Companies looking for untapped markets may find strategic advantages by focusing on these areas, fostering innovation, and contributing to their economic development.

Question 3: Do Countries with a Higher Land Area have more Universities established in 2012?



The analysis shows that while there is some correlation between a country's land area and the number of universities established, it is not a definitive rule. The USA, with its vast land area, has the highest number of universities, suggesting that larger countries often have more universities due to greater resources and population distribution needs. However, smaller countries like the United Kingdom and France also have many universities, indicating that economic factors, population density, and historical development significantly influence university establishment. Thus, while land area plays a role, it is not the sole determinant in the number of universities a country establishes.

Final Summary:

Based on our dataset analysis, we have successfully addressed the three key questions we posed. First, we observed a clear trend that countries with a higher-than-average GDP per capita tend to have more universities ranked in the top 100 globally in 2015. Second, our data shows that in 2014, Northern America and Eastern Asia significantly outperformed all other subregions in patent production, leading in innovation outputs. Third, while there is a correlation between larger land area and the number of universities established in 2012, land area is not the most decisive factor influencing university establishment. These insights highlight important relationships between economic indicators, innovation, and educational infrastructure on a global scale.

2. Wrangling Details

Link to Data Set Used: https://docs.google.com/spreadsheets/d/16PTQVf-zlWaO0FAYOUpMPX069grMPr8-/edit?usp=sharing&ouid=111040291671863362570&rtpof=true&sd=true

Dataset: World University Rankings

Link: https://www.kaggle.com/datasets/mylesoneill/world-university-rankings
General Characteristics:
Format: csv file
Structure: Tabular

Columns: 14 Columns, world_rank, institution, country, national_rank, alumni_employment, quality_of_faculty, publications, influence, citations, broad_impact, patents, score and year
Number of records: 2200
Initial Observations: I noticed that any record that has a year of 2013 does not have anything stored in the broad_impact field. I may also need to filter the data if I require a specific year.

Dataset: GDP Among World

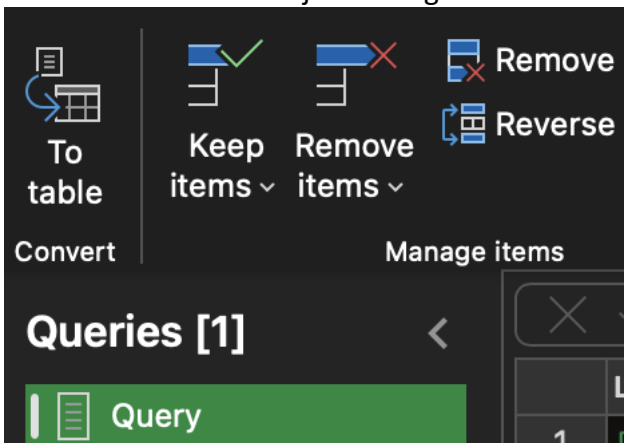
Link: https://www.kaggle.com/datasets/darknez/gdp-among-worldGeneral Characteristics:
Format: JSON file
Structure: JSON Array of objects
Columns: Country, Population Rank, Growth Rate, World Percentage, Density, Land Area, 2020 Population Rank, 2020 World Percentage, 2020 Growth Rate, Area, Region, Subregion, Anthem, Government, GDP (IMF), GDP (UN), GDP Per Capita
Number of records: 184
Initial Observations: Some of the numeric values have ","'s or "%"'s in them so they will need to be removed. IMF stands for International Monetary Fund. UN stands for United Nations.

Steps to Combine the Dataset.

1. Load the World University Rankings dataset into Excel using Power Query, set the delimiter to comma.

2. Using Power Query import the World GDP JSON File.

3. Convert the JSON objects using the "To table" button and set the delimiter to "None".



4. Expand the columns so that we can see all the values for each country. This is done by clicking the button with two arrows. Include all columns and select the setting use column name as prefix.

5. Change GDP Per Capita to a decimal number.

6. Close and load the data into excel.

7. In the GDP dataset change "United States" to USA in the Country column.

8. Create a column called Country Population Rank in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Population Rank]],2, FALSE).

9. Create a column called Country GDP Growth Rate in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Growth Rate]],3, FALSE).

10. Create a column called Country World Percentage in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [World Percentage]],4, FALSE).

11. Create a column called Country Density in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Density]],5, FALSE).

12. Create a column called Country Land Area in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Land Area]],6, FALSE).

13. Create a column called Country Total Area in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Area]],10, FALSE).

14. Create a column called Country's Continent in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Area]],10, FALSE)).

15. Create a column called Country's Subregion in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [Subregion]],13, FALSE)).

16. Create a column called Country's GDP (IMF) in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [GDP (IMF)]],16, FALSE).

17. Create a column called Country's GDP (UN) in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [GDP (UN)]],17, FALSE).

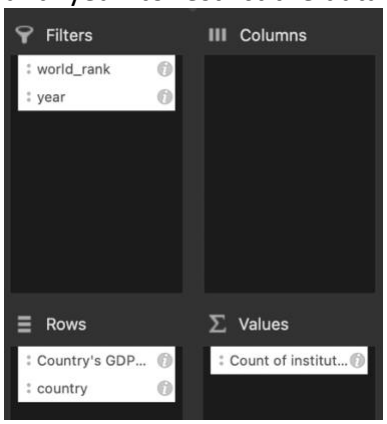
18. Create a column called Country's GDP (UN) in the rankings spreadsheet and enter this formula in the first row =VLOOKUP (C2, Query[[Country]: [GDP Per Capita]],18, FALSE).

3. Questions and Answers

Question 1: Do countries with a higher-than-average GDP per capita have more universities ranked in the top 100 in 2015?

The initial step involved determining the average GDP per capita. To achieve this, I extracted all unique values from the 'Country's GDP Per Capita' column, accounting for the presence of multiple universities per country, which could lead to duplicate GDP values. I utilized the formula =UNIQUE (FILTER (Y2:Y2201, Y2:Y2201 <> "")) to isolate these unique values while excluding any blank entries, as some countries had not reported their GDP. This process enabled the calculation of the average GDP per capita, resulting in a value of $31,638.03 (to 1 decimal place).

The next step was to construct a pivot chart using the refined dataset to facilitate a succinct summary. In the rows field, I included the country names and their corresponding GDP per capita values. The values field comprised the count of institutions. To tailor the chart to the specific query, I incorporated two filters: 'world_rank' to display only the top 100 institutions and 'year' to restrict the data to 2015, as specified in the question.



Upon generating the pivot chart, I ensured the axes and title were appropriately labelled as previously outlined in the report. However, I observed that several countries present in the dataset were not represented in the chart. To address this, I applied a similar method to the one used for calculating the average GDP values. I extracted all unique country names from the 'Country' column using the formula =UNIQUE (C2:C2201, FALSE, FALSE) and subsequently counted the distinct entries.

Question 2: What Subregion Produced the Most Patents in 2014?

The initial step entailed the creation of a pivot table utilizing the final dataset, facilitating a comprehensive and easily interpretable summary of the data. In the pivot table fields list, the Rows field was designated to contain the 'Subregions' attribute, thereby structuring the data according to geographic subdivisions. For the Values field, the configuration was set to

compute the sum of all patents, enabling an aggregate analysis of patent distribution across different subregions. This setup was crucial for obtaining a clear and concise summary of the total patent counts within each subregion. To further refine the dataset, the 'Years' attribute was placed in the Filters field. This allowed for the application of a temporal filter, specifically isolating the data for the year 2014. By filtering for this particular year, the analysis was focused on a discrete temporal segment, ensuring the relevance and specificity of the insights derived. This methodical approach ensured that the pivot table not only summarized the data effectively but also allowed for targeted analysis of patent distribution across subregions for the year 2014, providing valuable insights into temporal and spatial trends within the dataset.

Question 3: Do Countries with a Higher Land Area have more Universities established in 2012?

The preliminary phase involved incorporating the 'Land Area' variable into the Rows field of the pivot table, establishing the foundational structure for subsequent data analysis. Following this configuration, the Values field was meticulously defined to represent the count of institutions, enabling a quantitative assessment of distribution across various land areas. To refine the dataset, a filter was strategically applied to isolate universities established specifically in the year 2012. This temporal constraint was crucial for ensuring the relevance and specificity of the analysis, allowing for a focused examination of institutions within a particular establishment period. After the filtering process, the data was organized in descending order based on the count of institutions. This sorting methodology was employed to ensure a rational and logical presentation of the data, facilitating the identification of significant patterns and trends. By adopting this structured approach, the analysis aimed to uncover insights into the spatial distribution and establishment trends of universities, thereby providing a comprehensive understanding of the underlying data dynamics.