**FINAL PROJECT: GLOBAL AI INDEX ANALYSIS**

INFO 5709 Section 003 - Data Visualization and Communication

Department of Information Science, University of North Texas

Course Instructor: Dr. Gahangir Hossain

Submitted by Vemula Padmaja

26th July 2024

**Introduction**

Globally, artificial intelligence (AI) is significantly impacting how we live our lives, from healthcare to transportation. Researchers, corporations, and politicians must comprehend how various nations are equipped to utilize artificial intelligence technologies. The goal of this project is to analyze a dataset that offers information on how prepared different nations are for AI. Important parameters including talent, infrastructure, operating environment, R&D, government strategy, and commercial activities are included in the dataset. The primary goal in analyzing these characteristics is to find patterns and trends that illustrate the advantages and disadvantages of AI development and adoption in various geographical areas.

The dataset's structure gives users a thorough understanding of each nation's AI environment, including key details about their tactics and capabilities. This study will assist in determining which areas need greater focus and funding and which are at the forefront of AI development. To provide a comprehensive image of the competitive AI environment, this project intends to answer important questions regarding the worldwide level of AI with intelligent visualizations and exploratory data analysis (EDA).

**Dataset**

For this project, I used a dataset from Kaggle that details AI index metrics for various countries. This dataset offers a thorough analysis of each nation's artificial intelligence readiness and capabilities. It comprises important indicators including skill, resources, environment of operation, R&D, government policy, and business activity. Understanding the AI landscape and the relative positions of various nations regarding AI development and adoption depends heavily on these indicators. The dataset consists of sixty-two rows and thirteen columns, representing sixty-two countries and thirteen different attributes related to their AI readiness.

**Kaggle Dataset link:**

<https://www.kaggle.com/datasets/katerynameleshenko/ai-index?resource=download>

**Attributes:**

1. Country: The nationality of the nation.
2. Talent: A score indicating the quantity and caliber of AI talent in the nation.
3. Infrastructure: A grade that represents the quantity and caliber of AI-related infrastructure in the nation.
4. Operating Environment: A grade that considers business and regulatory factors to evaluate the general environment for AI operations.
5. Research: A grade that reflects the nation's ability and productivity in AI research.
6. Development: A tally of the nation's advancements and investments in AI research & development.
7. Government Strategy: An assessment of the efficiency and thoroughness of the national government's AI policies and strategy.
8. Commercial: A grade that represents the nation's level of commercial AI activity and business adoption.
9. Total Score: A combined score that reflects how prepared the nation is for AI in general.
10. Region: The geographical region where the country is located.
11. Cluster: The division of the nation into distinct groups or clusters according to their level of AI readiness.
12. Income Group: The national income group classification (high, midrange, and, low income).
13. Political System: The type of political regime in the country such as democracy or autocracy.

**Software Tools Used:**

* Python
* Tableau

**Data Cleaning and Preprocessing:**

The dataset looks to be clean, with no missing values or evident problems, based on the preliminary examination. To guarantee the dataset is fully prepared for analysis, it is usually a promising idea to conduct simple data cleansing procedures. Here is an overview of my findings and the lowest amount of cleaning that was done:

Initial Examination:

* There are thirteen columns and sixty-two rows in the dataset.
* None of the columns contain any missing data.
* The data types of the columns are appropriate (floats for numerical values and objects for categorical values).
* Minimal Cleaning steps including duplicate checks and data type checks were done using Google Colab using the Python code as mentioned below.

A screenshot of a computer

Description automatically generated

**Exploratory Data Analysis (EDA)**

Exploratory data analysis (EDA) is a crucial initial step in determining the dataset's underlying patterns, trends, and relationships. The EDA for this dataset includes generating research questions, looking into relationships between variables, and presenting distributions.

Five visualizations are created using Tableau software to analyze, understand, and identify the existing relationships between the attributes of the selected dataset.

1. **Which political regimes have the highest average AI scores?**

A graph of political regime

Description automatically generated

**Explanation:**

The highest average AI scores are found in “liberal democracies,” suggesting a substantial relationship between political freedom and AI advancement. “Electoral democracies,” on the other hand, have the lowest average AI ratings, indicating that partial political liberties might not be as helpful for AI development.

1. **What is the distribution of AI development scores across different income groups?**

**A graph with a bar chart

Description automatically generated with medium confidence**

**Explanation:**

From the box plot above it is evident that the greatest and most stable AI development scores are found in high-income nations, demonstrating their superior AI capabilities. On the contrary, low-income nations exhibit the most inconsistent and lowest ratings, underscoring the considerable differences in AI development.

1. **How do commercial AI activities compare across countries within each region?**

**A screenshot of a computer screen

Description automatically generated**

**Explanation:**

The above treemap shows the scores for commercial AI activity in 62 nations scattered over five regions: Africa, the Americas, Asia-Pacific, Europe, and the Middle East. It is simple to determine which countries are at the top of each region by looking at the larger, more vibrantly colored squares that indicate those with better scores for commercial AI activities. Clear and comprehensive information is quickly provided by labels that display the nation, region, and total amount of commercial activity.

1. **What is the percentage distribution of AI research scores within each region, and how do these distributions compare across different income groups?A screenshot of a computer screen

   Description automatically generated**

**Explanation:**

The distribution of AI research scores by region and income group is shown in the above visualization as pie charts, where each slice represents the percentage of a given income group's research score that goes toward the overall score in that region. Most contributions to AI research come from high-income groups, particularly in Europe and the Americas. The contributions of lower and upper-middle-class groups are smaller but still noteworthy, especially in Asia-Pacific and Africa.

1. **How do the total AI scores of countries compare across different regions and political regimes for the top-ranked countries?**

**A graph of different colored squares

Description automatically generated**

**Explanation:**

Each region's total AI score is displayed in the provided stacked bar chart with colors denoting the various political systems (closed autocracy, electoral autocracy, electoral democracy, liberal democracy). The rank of the overall score and the region are written on each bar. Liberal democracies predominate in the Americas and Asia-Pacific, which have the highest overall AI scores. The distribution of political regimes within each location emphasizes how governance affects the advancement of AI.

**Hypothesis**

Based on EDA the following hypotheses were made.

1. Countries with higher talent ratings also tend to have higher overall AI scores, suggesting that a developed talent pool is necessary for advanced AI capabilities. This correlation indicates that investing in talent development can significantly increase a country's AI performance.
2. Regions with better infrastructure are expected to have higher commercial AI activity because robust infrastructure supports the development and deployment of AI technologies
3. The research and development scores of nations with robust government policies for AI development are higher.

**How do the average total AI scores compare for countries with high, medium, and low talent scores?**

Visualization Type: Bar Chart

A graph with blue and red squares

Description automatically generated

Explanation:

The bar graph above, colored based on talent category, shows the average total AI scores for each talent category (High, Medium, and Low). The figure indicates that countries with high talent scores have the highest average total AI scores (100.00), followed by countries with medium talent scores (45.67) and low talent scores (22.54). This supports Hypothesis 1 by showing a clear positive correlation between higher talent ratings and stronger AI performance.

**Is there a statistically significant correlation between talent ratings and total AI scores?**

Visualization Type: Scatter plot

**A graph with a line and numbers

Description automatically generated**

**Explanation:**

A robust positive association (R-squared of 0.993) has been shown in the scatter plot "Talent Ratings vs. Total AI Scores," suggesting that nations with higher talent ratings typically have better overall AI performance. The statistical significance of this association is confirmed by the significant p-value (0.0002334), thereby providing credence to the premise that investing in talent development is crucial in augmenting a nation's AI capabilities.

**How do infrastructure scores and commercial AI activity scores vary across different regions and income groups?**

Visualization Type: Heatmap

**A screenshot of a graph

Description automatically generated**

**Explanation:**

The resulting heatmap shows the total infrastructure and commercial AI activity scores for various areas and income brackets. Larger squares represent higher commercial AI activity scores, which are correlated with higher infrastructure scores (shown by darker hues), especially in affluent locations such as the Americas and Europe. This demonstrates that areas with superior infrastructure typically have more commercial AI activity with very few exceptions, which partially validates Hypothesis 2.

**How do infrastructure and commercial AI activity scores compare across different regions and political regimes?**

Visualization Type: Stacked Bar

**A graph of different colored squares

Description automatically generated**

**Explanation:**

The stacked bar chart, which is colored by the political regime, shows the total quantity of commercial AI activity for each place. Regional differences in infrastructure scores and commercial AI activity values, particularly in the Americas and Europe, are consistent with Hypothesis 2, which maintains that robust infrastructure fosters higher levels of commercial AI activity. The distinction in color based on political regime serves as an additional example of how government influences AI development.

**Which countries have the highest government strategy scores and how do their research scores compare?**

Visualization Type: Treemap

**A screenshot of a computer screen

Description automatically generated**

The larger and darker squares on the graph represent the countries with the greatest government strategy scores. Most of these countries like the USA, and Singapore also generally display better research ratings. Strong government policies for AI development are linked to greater research and development scores, as this positive association demonstrates in support of Hypothesis 3. This interaction between different countries is effectively highlighted by the above treemap.

**How do research and development scores vary across countries with different government strategy scores and income levels?**

Visualization Type: Grouped Bar chart

**A screenshot of a graph

Description automatically generated**

**Explanation:**

The total of each nation's research and development scores are shown in the provided grouped bar chart, with distinct colors denoting different income groups. The third hypothesis, according to which strong government policies foster the development of AI, is supported by the higher research and development scores found in nations with strong government strategies. It is also evident from the differentiation by income groups that better results from research and development are correlated with higher income levels.

**Conclusion**

Analyzing the global AI index dataset allowed for a better understanding of how infrastructure, talent, and political regulations influence AI development globally. The analysis confirmed that nations with higher talent and infrastructure rankings had a higher chance of achieving both high AI performance and commercial success. Successful government measures were also linked to higher research and development ratings. Creating various graphics, such as scatter plots and bar charts, with Tableau made it easier to understand these correlations. This effort helped me better grasp data analysis and visualization by exposing the essential elements that drive AI progress globally.

**References**

*AI Global Index*. (2023, April 26). Kaggle. <https://www.kaggle.com/datasets/katerynameleshenko/aiindex?resource=download&select=AI_index_db.csv>

*Artificial Intelligence Index*. (n.d.). <https://aiindex.stanford.edu/>

Kateryna, & Kateryna. (2024, July 16). Global AI race: dominant players and aspiring challengers. *Intersog*. <https://intersog.com/blog/development/ai-dominant-players-and-aspiring-challengers/>