**Cyber Security Indexes Analysis**

## **1. Title Page**

* **Project Title:** Cyber Security Indexes Analysis
* **Name:** Ammar Khan

## **2. Abstract**

The Dataset "Cyber Security Indexes" includes four indicators which illustrate the current cyber security situation around the world. The data is provided on 193 countries and territories, grouped by five geographical regions - Africa, North America, South America, Europe and Asia-Pasific.

*The Cybersecurity Exposure Index* **(CEI)** defines the level of exposure to cybercrime by country from 0 to 1; the higher the score, the higher the exposure (provided by [10guard](https://10guards.com/en/articles/global-cybersecurity-exposure-index-2020/)). The indicator was last updated in 2020.

*The Global Cyber Security Index* **(GCI)** is a trusted reference that measures the commitment of countries to cybersecurity at a global level – to raise awareness of the importance and different dimensions of the issue (provided by the International Telecommunication Union - [ITU](https://composite-indicators.jrc.ec.europa.eu/explorer/explorer/indices/GCI/global-cyber-security-index)). The indicator was last updated in 2021.

*The National Cyber Security Index* **(NCSI)** measures a country's readiness to address cyber threats and manage cyber incidents. It is composed of categories, capacities, and indicators (provided by [NCSI](https://ncsi.ega.ee/ncsi-index/)). The indicator was last updated in January 2023.

*The Digital Development Level* **(DDL)** defines the average percentage the country received from the maximum value of both indices (provided by [NCSI](https://ncsi.ega.ee/ncsi-index/)). The indicator was last updated in January 2023.

The dataset can be used for practising data cleaning, data visualization (on maps and round/bar charts), finding correlations between the indexes and predicting the missing data.

The data was used in the analytical article research *[The Geography of Cybersecurity: Cyber Threats and Vulnerabilities](https://intersog.com/blog/geography-of-cyber-security/)*

## **3. Introduction**

### **Overview of the Problem Statement**

Cybersecurity threats are increasing worldwide, and understanding global readiness is crucial. This project aims to explore cybersecurity indexes and their interrelationships.

### **Purpose and Significance**

By analyzing these indexes, we can assess exposure to cybercrime, cybersecurity preparedness, and digital development trends globally.

### **Dataset Description and Relevance**

The dataset includes cybersecurity indicators for 193 countries, grouped by region, and provides insights into global cyber readiness.

## **4. Dataset Description**

### **Source of the Dataset**

The dataset is obtained from ITU, 10Guard, and NCSI databases.

### **Dataset Structure**

The dataset contains the following key features:

* **Country**: Name of the country
* **Region**: Continent/Geographical grouping
* **CEI**: Cybersecurity Exposure Index (0-1)
* **GCI**: Global Cyber Security Index (0-1)
* **NCSI**: National Cyber Security Index (0-100)
* **DDL**: Digital Development Level (0-100)

### **Summary Statistics & Key Observations**

* CEI: Measures cybersecurity risk exposure.
* GCI: Assesses a country's commitment to cybersecurity.
* NCSI: Evaluates cybersecurity readiness.
* DDL: Measures digital development based on CEI & GCI.

### **Preprocessing Steps**

* Handling missing values using visualization (Missingno matrix, bar charts)
* Checking unique values for categorical data
* Generating descriptive statistics

## **5. Feature Selection and Engineering**

* **Selected Features:** CEI, GCI, NCSI, DDL
* **Feature Engineering:** Created a ‘Size’ feature for better visual representation in maps.

## **6. Methodology**

* **Exploratory Data Analysis (EDA)** using statistical methods and visualizations.
* **Visualization Techniques:** Histograms, box plots, and scatter plots.
* **Predictive Modeling:** Regression analysis for missing values.

## **7. Model Development**

* Implemented missing value imputation using statistical techniques.
* Developed visualizations using Matplotlib, Seaborn, and Plotly for analysis.

## **8. Model Evaluation**

* **Metrics Used:** Correlation heatmaps to analyze relationships between indexes.
* **Findings:** High correlation between GCI and DDL, suggesting cybersecurity commitment aligns with digital development.

## **9. Predictions and Insights**

* High CEI values indicate high exposure to cyber threats.
* Countries with high GCI tend to have better cybersecurity preparedness.
* Regional variations show Asia-Pacific and Europe performing better in cybersecurity preparedness.

## **10. Visualizations**

* Cybersecurity exposure by region (Box Plots)
* Global cybersecurity commitment (Scatter Geo maps)
* Readiness and digital development levels across regions

## **11. Challenges and Limitations**

* Missing data in some countries affected accuracy.
* Model assumptions for linear relationships may not fully capture non-linear cybersecurity trends.

## **12. Conclusion**

* Cybersecurity commitment is crucial for national security.
* Strong correlation between digital development and cybersecurity readiness.
* Future work: More advanced predictive modeling and policy recommendations.

## **13. References**

* ITU Global Cyber Security Index Report
* 10Guard Cybersecurity Exposure Index Data
* National Cyber Security Index Database
* Python Libraries: Pandas, Matplotlib, Seaborn, Plotly

## **14. Appendices (Optional)**

* Additional visualizations
* Sample dataset records