**Report on the Analysis of South Africa's Digital Growth and Cybersecurity Readiness**

**Course:** NDTA631  
**Group Members:**

* Hope Mnguni (202203558)
* Gomolemo Walaza (202200387)
* Gift Nemakonde (202304063)
* Keotshepile Modise (202321837)
* Ikho Nogemane (202301434)

**Date:** March 2025

**Page 1: Executive Summary**

This report presents a data-driven analysis of the relationship between South Africa's digital growth and its cybersecurity preparedness, utilizing datasets from the World Bank Data360 portal. The primary research question guiding this investigation was: **"Is South Africa's cybersecurity infrastructure keeping up with its growing digital footprint?"**

**Key Findings:**

* **Digital Adoption has Grown Exponentially:** Internet user penetration increased from 5.35% in 2000 to approximately 75% in 2024, characterized by a mobile-first revolution.
* **Cybersecurity Posture Shows Mixed Results:** The overall Global Cybersecurity Index (GCI) score improved from 78.46 (2020) to 86.25 (2024). However, a critical component, Capacity Development, declined from 15.37 to 12.83.
* **A Significant Data Gap was Identified:** A core constraint of this analysis is the disparity in data availability. Robust ICT data spans 2000-2024, while cybersecurity data is only available for 2020 and 2024, preventing a longitudinal correlation analysis.

**Conclusion:** South Africa has achieved remarkable digital inclusion, predominantly through mobile technology. However, the development of human capital in cybersecurity is not keeping pace with technical and legal advancements. The analysis concludes that while foundational cybersecurity frameworks are strong, urgent investment in skills development is required to ensure future resilience. The methodologies developed herein provide a robust framework for ongoing monitoring.

**Page 2: Introduction & Methodology**

**1.1 Project Objective**  
This project aims to quantitatively assess the relationship between the expansion of Information and Communication Technology (ICT) infrastructure and adoption in South Africa and the nation's concurrent maturity and preparedness in cybersecurity.

**1.2 Data Sources**  
Two primary datasets were obtained from the World Bank Data360 portal:

1. **ICT Indicators Database (**gci\_ICT.csv**):** Contains metrics measuring digital connectivity and infrastructure (e.g., mobile and broadband subscriptions, internet usage, international bandwidth).
2. **Global Cybersecurity Index (**gci\_Cybersecurity.csv**):** Captures country-level scores on cybersecurity readiness across dimensions including legal, technical, organizational, capacity development, and cooperation.

**1.3 Analytical Tools and Environment**  
The analysis was performed in Python using a standard data science stack:

* **pandas & NumPy:** For data loading, cleaning, reshaping, and numerical computation.
* **Matplotlib & Seaborn:** For creating static, visual, and statistical graphics.
* **SQLite:** For database integration and querying.

This environment was chosen to ensure reproducibility, scalability, and the ability to handle real-world data inconsistencies.

**Page 3: Data Preparation and Cleaning**

**3.1 Data Loading and Inspection**  
The raw CSV files were loaded into Pandas DataFrames:

python

ict = pd.read\_csv("data/gci\_ICT.csv") *# Initial shape: (204 rows, 50 columns)*

cyb = pd.read\_csv("data/gci\_Cybersecurity.csv") *# Initial shape: (30 rows, 20 columns)*

The data was immediately filtered to isolate records for South Africa using the country code ZAF.

python

ict\_sa = ict[ict["REF\_AREA"] == "ZAF"].copy()

cyb\_sa = cyb[cyb["REF\_AREA"] == "ZAF"].copy()

**3.2 Data Reshaping (Wide to Long Format)**  
The datasets were provided in a wide format where each year was a separate column. For effective time-series analysis, this was transformed into a long format using the melt() function.

python

ict\_years = [c for c in ict\_sa.columns if c.isdigit()]

ict\_long = ict\_sa.melt(

id\_vars=["REF\_AREA","REF\_AREA\_LABEL","INDICATOR","INDICATOR\_LABEL"],

value\_vars=ict\_years,

var\_name="Year",

value\_name="Value"

)

This resulted in a tidy data structure with columns: [REF\_AREA, INDICATOR\_LABEL, Year, Value].

**3.3 Handling Missing Values**  
Missing values (NaN) in the Value column were addressed using a combination of forward-fill (ffill) and backward-fill (bfill) methods, grouped by indicator.

python

ict\_long["Value"] = ict\_long.groupby("INDICATOR\_LABEL")["Value"].transform(

lambda x: x.ffill().bfill()

)

This imputation strategy preserved the continuity of time-series data without introducing artificial variance. Post-cleaning, both datasets contained zero missing values.

**Page 4: Numerical Analysis and Findings**

**4.1 Analytical Approach**  
The cleaned, long-format data was pivoted to create a Year × Indicator matrix, enabling vectorized calculations.

* **Year-over-Year (YoY) Change:** Calculated to understand annual momentum and identify shocks or acceleration periods.

python

internet\_yoy\_pct = (np.diff(internet\_users) / internet\_users[:-1]) \* 100

* **Compound Annual Growth Rate (CAGR):** Computed to smooth volatility and understand long-term trends.

python

years\_between = len(years) - 1

internet\_cagr = ((internet\_users[-1] / internet\_users[0]) \*\* (1/years\_between) - 1) \* 100

* **Correlation Analysis:** Attempted to measure the linear relationship between ICT and cybersecurity metrics. This returned NaN due to insufficient cybersecurity data points.

**4.2 Key ICT Findings**

* **Internet Adoption:** Grew from **5.35%** of the population in **2000** to approximately **75%** in **2024**, with a CAGR of **~8.2%**.
* **Mobile-First Nation:** Mobile broadband subscriptions (**72.3 per 100 inhabitants**) vastly outpace fixed broadband (**5.3 per 100 inhabitants**), indicating a strategy of technological leapfrogging.
* **Infrastructure Capacity:** International bandwidth per capita saw exponential growth, from **7.4 bps (2000)** to **~31,090 bps (2023)**.

**4.3 Key Cybersecurity Findings**

* **Overall Improvement:** The aggregate GCI score improved from **78.46 (2020)** to **86.25 (2024)**.
* **Component Analysis:** The 2024 score comprises:
  + Cooperative Score (CS): **20.00/20** (Perfect)
  + Technical Score (TS): **18.57/20** (Strong)
  + Organizational Score (OS): **14.84/20** (Good)
  + Capacity Development Score (CDS): **12.83/20** (Concerning)
* **Critical Decline:** The Capacity Development score regressed from **15.37 to 12.83**, indicating a decline in training and human capital development.

**Page 5: Database Integration and Advanced Analysis**

**5.1 Database Schema and Implementation**  
A SQLite database (GCIDatabase.db) was created to ensure data integrity, facilitate complex queries, and provide a persistent storage solution. Separate tables were designed for ICT and Cybersecurity data to reflect their differing structures.

python

cursor.execute('''

CREATE TABLE IF NOT EXISTS gci\_ICT (

REF\_AREA TEXT,

REF\_AREA\_LABEL TEXT,

INDICATOR TEXT,

INDICATOR\_LABEL TEXT,

"2000" REAL,

...,

"2024" REAL

)

''')

Data was programmatically inserted from the CSVs, and operations were performed to update values and delete irrelevant records, demonstrating full CRUD (Create, Read, Update, Delete) functionality.

**5.2 Advanced Analytical Functions**  
Conditional formatting and significance testing were implemented to automate insight generation.

* **Conditional Formatting:** Values were categorized as 'High', 'Normal', or 'Low' based on their position relative to statistical thresholds (e.g., quartiles).
* **Significance Testing:** YoY changes were filtered to highlight only those exceeding a 10% change or a 5-point absolute change, separating meaningful trends from minor fluctuations.

**5.3 Core Limitation: Data Disparity**  
The database structure highlights the project's main constraint. The ICT table contains columns for years **2000-2024**, while the Cybersecurity table effectively contains only **2020 and 2024**. This fundamental mismatch in temporal resolution is the primary reason a correlative analysis between the two domains could not be conclusively performed.

**Page 6: Discussion of Integrated Insights**

**6.1 Interpreting the Disconnect**  
The inability to calculate a correlation coefficient is, in itself, a significant finding. It indicates that the time-series data required to rigorously answer the research question is not yet available. This underscores a need for more consistent and frequent reporting of cybersecurity metrics.

**6.2 Convergent Trends**  
Despite the data gap, convergent trends can be inferred:

* **Positive Alignment:** The period of rapid digital adoption (post-2010) coincides with the development of a strong cybersecurity legal and technical framework, as evidenced by the high 2024 scores in these areas.
* **Mobile Security:** The mobile-dominated growth pattern necessitates a cybersecurity strategy focused on mobile network security and consumer device protection.

**6.3 Divergent Trends - The Capacity Gap**  
The most clear and alarming divergent trend is the **decline in the Capacity Development score** against a backdrop of soaring digital adoption. This indicates that the human capital required to operate and maintain cybersecurity systems is not being developed at a sufficient rate. A nation cannot secure a growing digital ecosystem with a shrinking pool of experts.

**6.4 Strategic Recommendations**  
**Immediate Actions (0-6 months):**

* Launch national cybersecurity awareness campaigns focused on mobile users.
* Establish public-private partnerships to address the capacity development gap urgently.

**Medium-term Actions (6-18 months):**

* Integrate cybersecurity fundamentals into national educational curricula.
* Invest in securing critical infrastructure, particularly undersea cable landing points (reflected by high international bandwidth).

**Long-term Strategy (18+ months):**

* Create a national cybersecurity innovation fund to develop indigenous solutions.
* Position South Africa as a regional cybersecurity hub for Africa.

**Page 7: Conclusion and Future Work**

**7.1 Conclusion**  
South Africa's journey toward digital inclusion has been remarkable, largely driven by mobile technology that has allowed it to leapfrog traditional infrastructure barriers. The analysis confirms strong digital growth.

The country has concurrently developed robust cybersecurity capabilities in legal, technical, and cooperative dimensions. However, this progress is threatened by an emerging crisis in human capacity development. The decline in training and skills development creates a significant vulnerability. Therefore, while cybersecurity infrastructure is advancing, it is not keeping pace with the scale of digital growth in terms of human capital.

**7.2 Future Work**  
This project developed a full data analysis pipeline. The immediate future work is to populate this pipeline with more complete data. Specifically:

1. **Source Additional Data:** Acquire cybersecurity data for the intervening years (2000-2019, 2021-2023) to enable a proper longitudinal study.
2. **Implement Forecasting:** Use the clear ICT growth trends to build a time-series model that forecasts the required cybersecurity capacity for future years.
3. **Expand the Framework:** Apply this same analytical framework to other nations in the Southern African Development Community (SADC) for a comparative regional analysis.

The foundation for a continuous monitoring system is now in place. The critical next step is to feed it more data to generate the insights needed to secure South Africa's digital future.