Carnegie Mellon University

AFRINIC
Progress Report 3

APRIL 21ST, 2021

Researchers: Pasqua Ruggiero, Esther Kamau, Blaise Viateur Niyigena, Isaac Manzi
Faculty Advisor: Dr. Assane Gueye
Client: Dr. Amreesh Phokeer

Agenda

1. Introduction
2. Data
3. Methods
4. Results
5. Next Steps

Introductions

CMU Student Team: Researchers

Isaac Manzi

CMU Africa – MSIT Program

Blaise Viateur Niyigena

CMU Africa – MSIT Program

Esther Kamau

CMU Africa – MSIT Program

Pasqua Ruggiero

CMU Pittsburgh – MSP Program

CMU Student Team Advisor: Dr. Assane Gueye

Profession: Assistant Professor at CMU Africa

Research Interests:

- Cybersecurity
- Connectivity in Rural and Under-Served Areas
- Machine Learning and Artificial Intelligence

Client: Dr. Amreesh Phokeer

Previous Profession: Research Manager at AFRINIC

Current Profession: Internet Measurement and Data Expert at Internet Society

Research Interests:

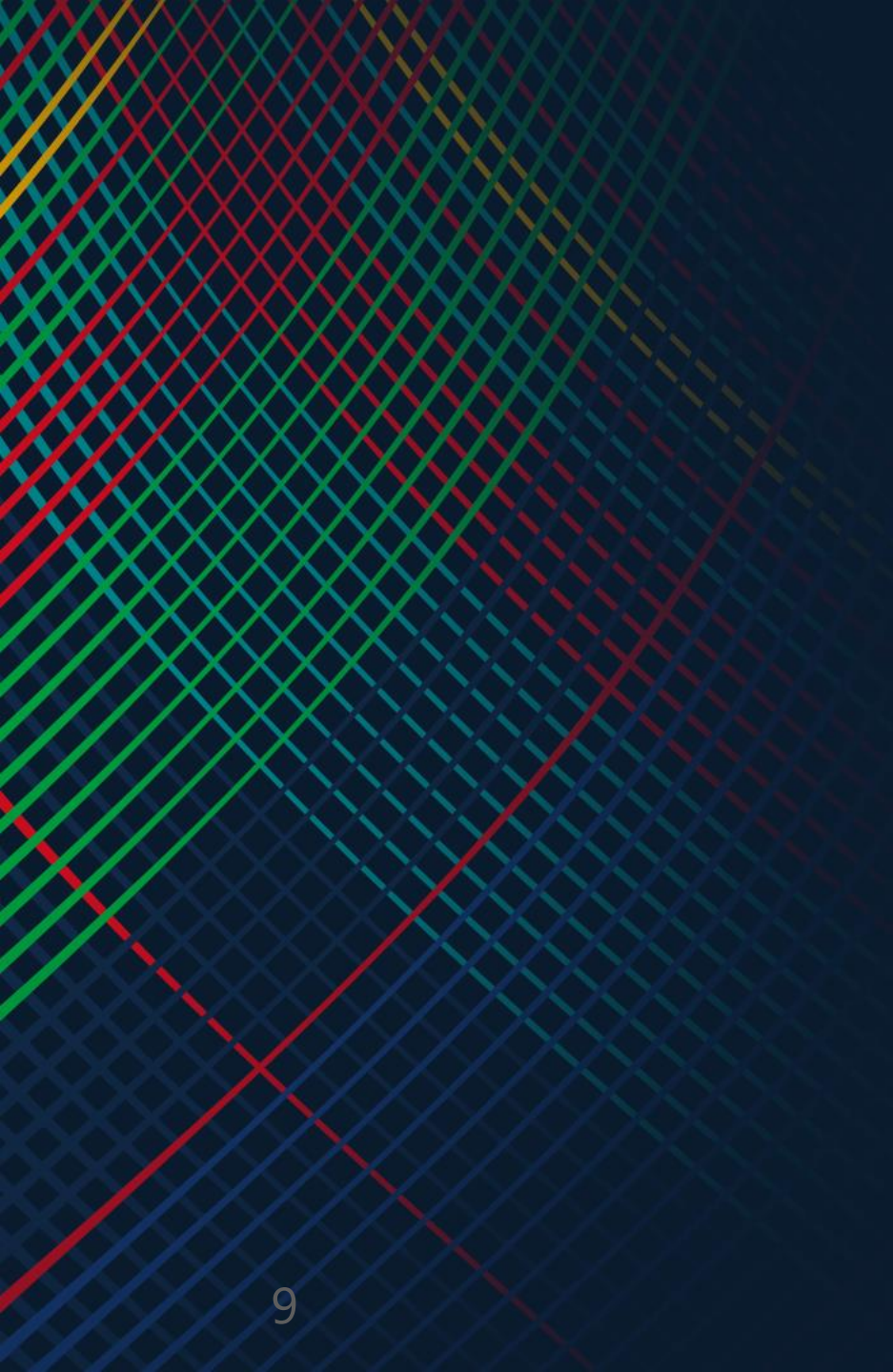
- Interdomain routing
- Network Security
- Internet Measurements
- Software Design

Background

- AFRINIC(Regional Internet Registry for Africa) – allocate internet number resources to network providers
- MIRA(Measuring Internet Resiliency in Africa) – survey current state of Internet in African countries and create framework that evaluates capability to provide reliable means of Internet connectivity during times of crises
  - Provide recommendations to providers of low-scoring countries to help them achieve higher resilience

Questions

1. Which metrics are most relevant and readily available to rank network resilience?
2. How do we use the above metrics to create an aggregate index representation to rank internet resiliency among different spatial extents of Africa?
3. Is there a way that we can best automate the re-calibration process of model parameters?
4. How can we best visualize internet resiliency in Africa to stakeholders and end users?



Data

Data overview

- 15* datasets used – one for each metric of interest
  - All datasets obtained
  - Open source, AFRINIC or other internal collections
- Preprocessing performed on most datasets

*Not including Geojson and ISO code data needed to create dashboard

Category	Metric	Measurement
QoS	Throughput	Fluctuation in throughput
	Latency	Latency to local services (ms)
	IPv6 capability	IPv6 capability of the ISP network (count)
Security	Routing	% of prefixes covered by IRR object
	AS hegemony	Compute the AS dependency of network
	MANRS Score	The consolidated MANRS score
	DDos Potential	Level of risks posed to other countries
	Spam Infection	%
Infrastructure	IXPs per 10M	Number of IXPs per 10M inhabitants
	IXP efficiency	% of ASes present at the IXP
	Upstream	Number of upstream providers
	Cable landing stations	Number of cable landing stations per capita/km2
	reach	% of population within 10-Km reach
	degree distribution	Degree distribution of cable entering/leaving a country/city
Affordability	Affordability	How affordable is Internet services in this country ($)

Data Description

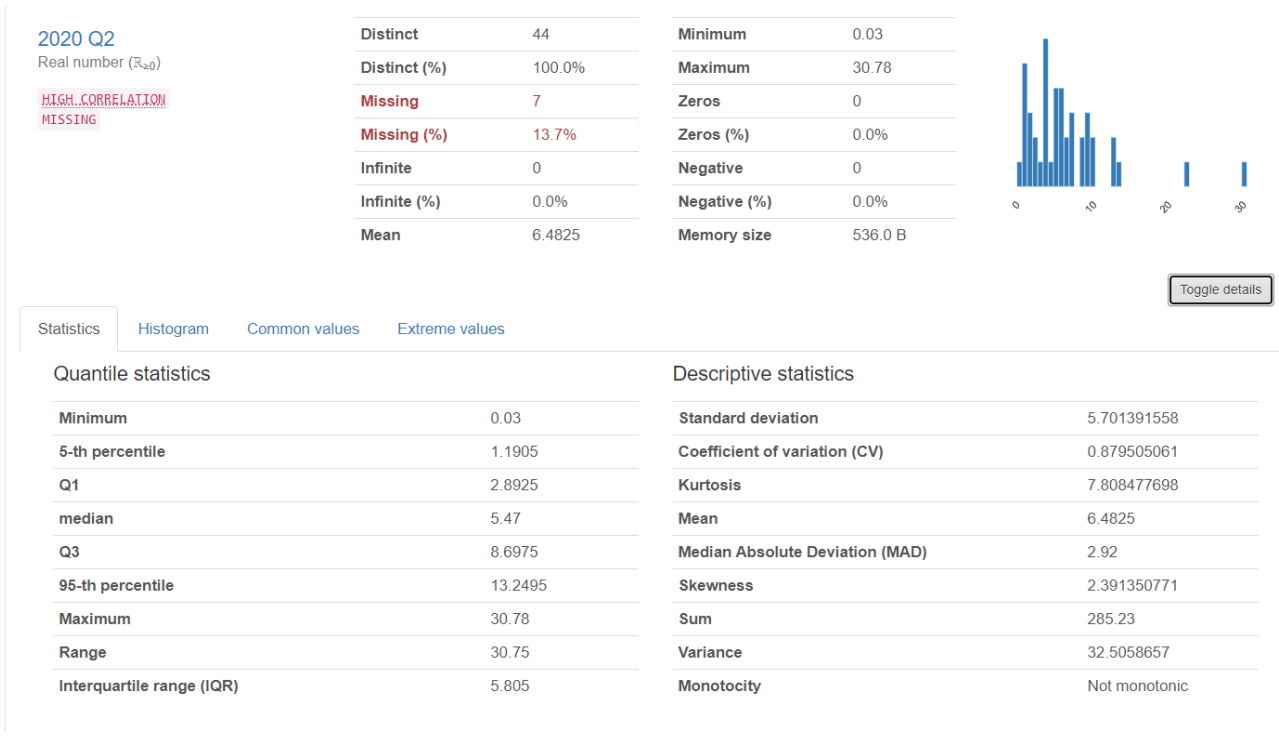
- Various formats – json, csv, text files
  - All standardized to pickle files
- Between 37 and 57 countries represented per dataset
- Measurements obtained between years of 2019-2021

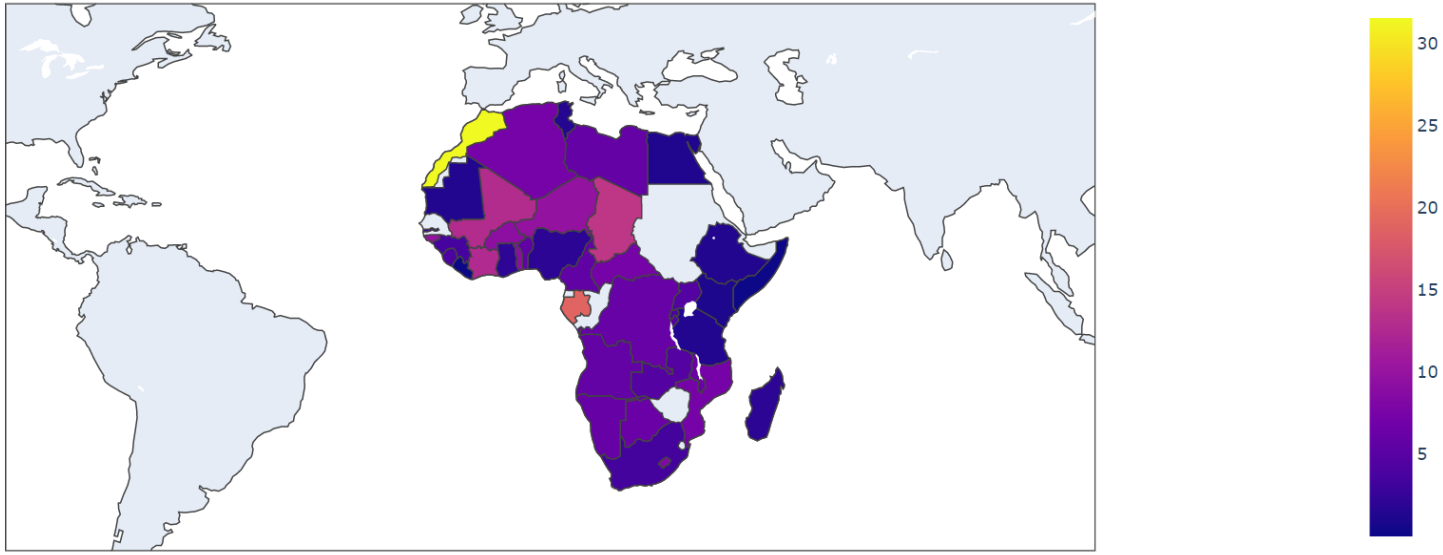
Data Preprocessing

1. Tidy – one row per country
2. High quality – reliable sources
3. Highly representative of African countries
4. Standardized measurements

Data Exploration

- Pandas Profiling Library
- Choropleth maps using Python Plotly
- Correlation analysis to prevent double penalization





Methodology

Methodology Overview

1. Select metrics based on relevance, independence, and data accessibility
2. Use a combination of research findings, intuition and subject matter opinions for weighing/ranking
3. Use Python dash library to create dashboard, aim to deploy in Apache Superset
4. Store data in MySQL database created by team to facilitate recalibration and model updates in the future

Step 1: Research of Metrics

- Research potential metrics relating to Internet security, Quality of Service, Infrastructure and Affordability
  - White papers, websites, reports, previous research by AFRINIC, etc.
  - Conversations with client based on their expertise

Step 2: Selection of Metrics Guidelines

1. Real measurements easily attainable
2. Majority coverage of African countries
3. Recent data available
4. Independent metrics (i.e. not heavily correlated to others selected)
5. Indicative of certain aspect of country's Internet resiliency

Step 3: Weighing Metrics

- Grouped based on type of measurement
- Categories primarily influenced by MIRA white paper, The Economist's "The Inclusive Internet Index 2020 Methodology report"
  1. **Internet Availability**
  2. **Internet Affordability**
  3. Internet Accessibility
  4. Internet Readiness
- Data coverage, availability and importance to describing each category

Step 4: Displaying Scores

- Calculate score per country based on formula below – 2 level weighted average
- Translate score into qualitative representation (ex. Low, medium, high)
- Facilitate comparison through dashboard
  - Currently using Python Dash and Plotly in notebooks, will be transferring to Apache Superset
  - Easy to integrate

$$Y = w_{c_1} * (w_{m_1} * m_1 + \ldots) + w_{c_2} * (w_{m_1} * m_1 + \ldots) + \ldots$$

Step 5: Recalibration and Parameter Updates

- MySQL database
- More to come!

Results

Current Results – Metric Weights

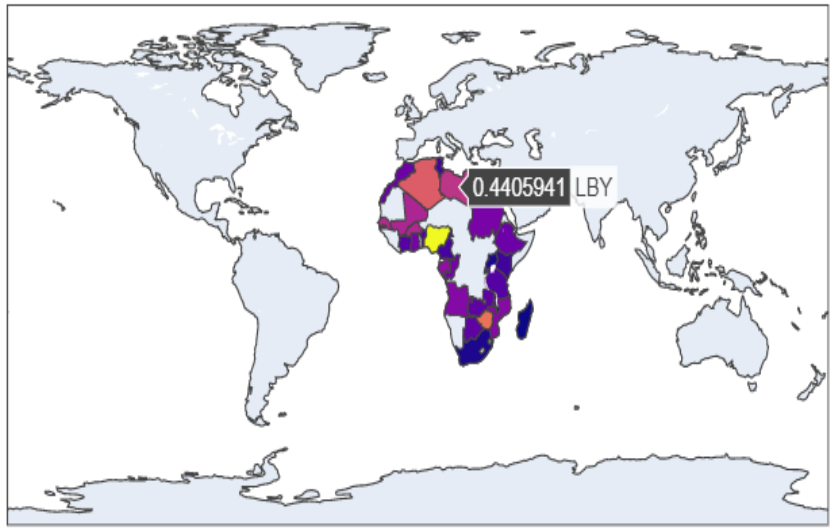
- Ad-hoc weighing scheme in Internet Lifecycle: Availability -> Affordability -> Relevance -> Readiness
- 4 Categories: Quality of Service, Security, Infrastructure, Affordability
- 1-6 metrics per category
- Discussion/awaiting approval of client
    - Discussions occurring within AFRINIC

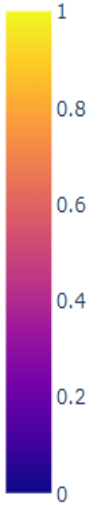
Category	Metric	Proposed Individual Metric Weight According to Internet Lifecycle	Proposed Individual Metric Weight	Measurement
QoS - 25%	Throughput	Availability - Quality (very important)	33.30%	Fluctuation in throughput
	Latency	Availability - Quality (very important)	33.30%	Latency to local services (ms)
	IPv6 capability	Availability - Quality (very important)	33.30%	IPv6 capability of the ISP network (count)
Security - 25%	Routing	Availability - Quality (very important)	20%	% of prefixes covered by IRR object
	AS hegemony	Availability - Quality (very important)	20%	Compute the AS dependency of network
	MANRS Score	Availability - Quality (very important)	20%	The consolidated MANRS score
	DDos Potential	Availability - Quality (very important)	20%	Level of risks posed to other countries
	Spam Infection	Availability - Quality (very important)	20%	%
Infrastructure - 35%	IXPs per 10M	Availability - Infrastructure (very important)	12.50%	Number of IXPs per 10M inhabitants
	IXP efficiency	Availability - Infrastructure (very important)	12.50%	% of ASes present at the IXP
	Upstream	Availability - Quality (very important)	25%	Number of upstream providers
	Cable landing stations	Availability - Infrastructure (very important)	12.50%	Number of cable landing stations per capita/km2
	reach	Availability - Infrastructure (very important)	25%	% of population within 10-Km reach
	degree distribution	Availability - Infrastructure (very important)	12.50%	Degree distribution of cable entering/leaving a country/city
Affordability - 15%	Affordability	Affordability - Price	100%	How affordable is Internet services in this country ($)

Current Results – Dashboard

- Currently, can view African countries with hover effect
- Features to be included:
    - User selection of weights per category and metric
    - Incorporation and display of Internet Resiliency scores
    - Qualitative assignment of scores to display
    - Deployment to AFRINIC site

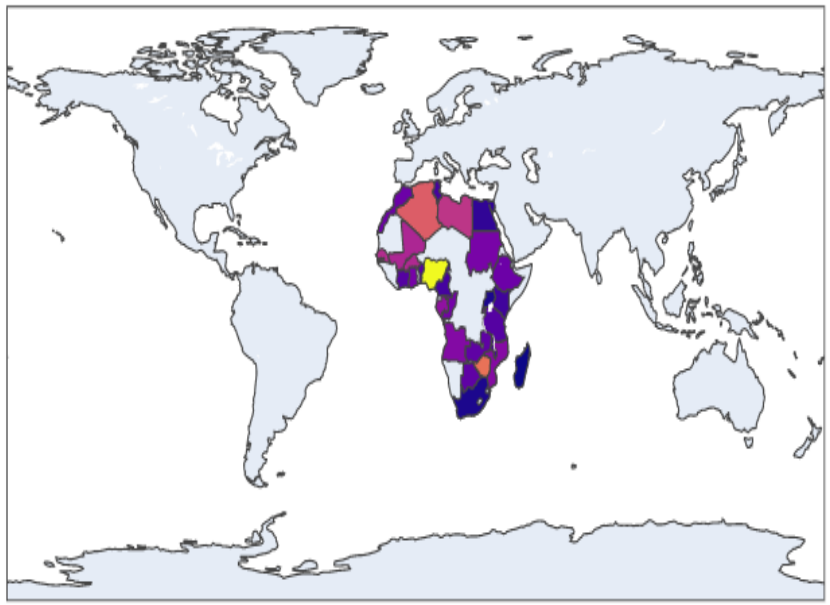
Weights:

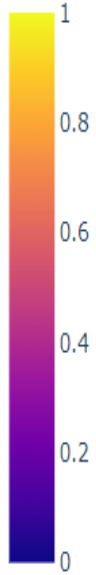




Metrics:

○ standardized 10-km Range ○ standardized affordability ○ standardized amplified count ○ standardized links per node ○ standardized ipv6 counts ○ standardized irr ○ standardized landing stations ◉ standardized spam





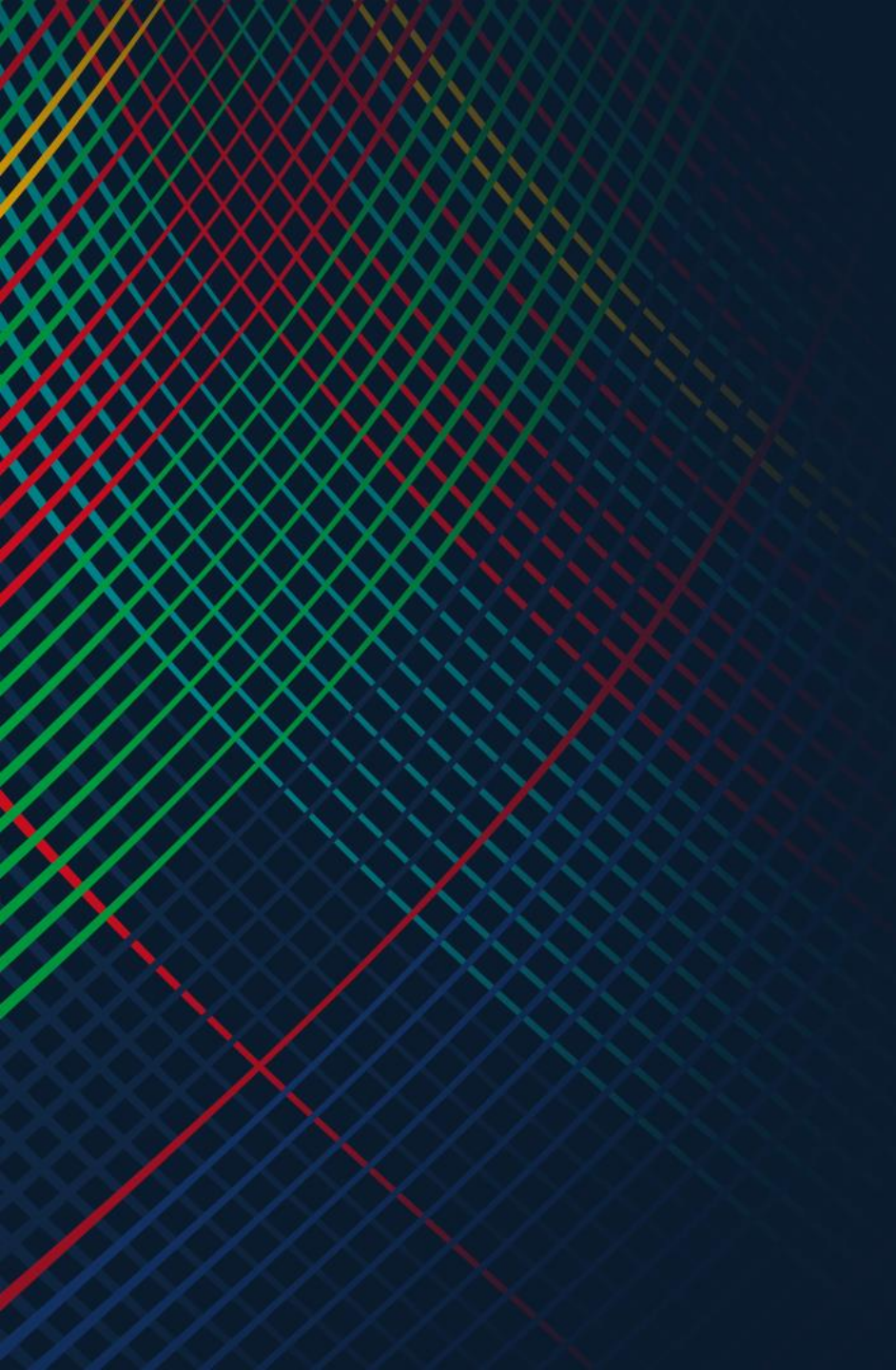
Next Steps

Products expected

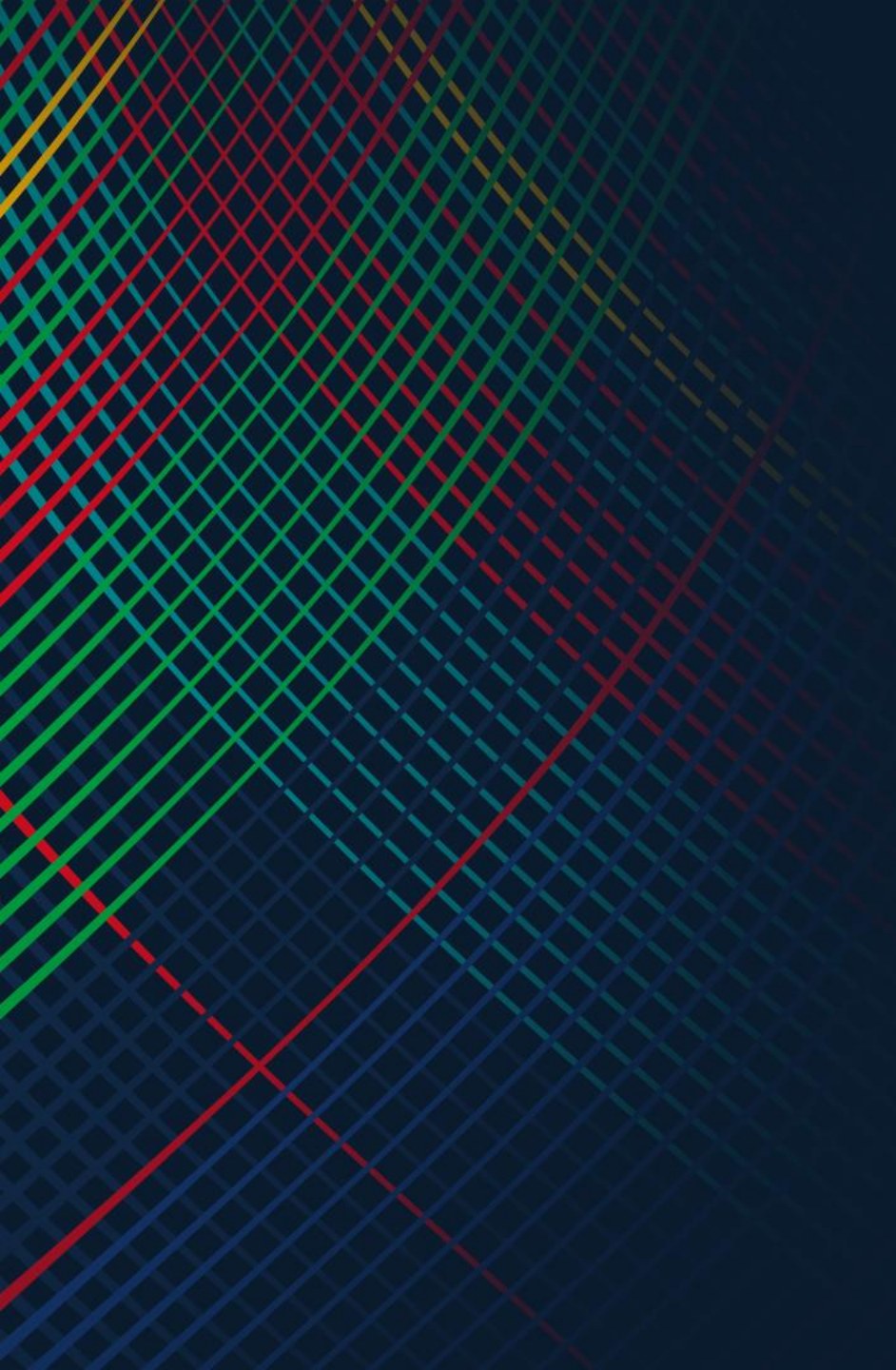
1. ~~Exploratory data analysis of various metrics~~
2. ~~Aggregation Index of Internet Resiliency measures in Africa~~
3. ETL pipeline to automate data extraction and model recalibration
4. Interactive dashboard that not only displays our score, but allows users to select weights per metric
5. Research Paper/Final Presentation

Timeline

1. All incomplete deliverables are currently being worked on in parallel
2. Currently have large focus on merging datasets, completion of dashboard and creation of database



Thank you!



Questions?