Carnegie Mellon University

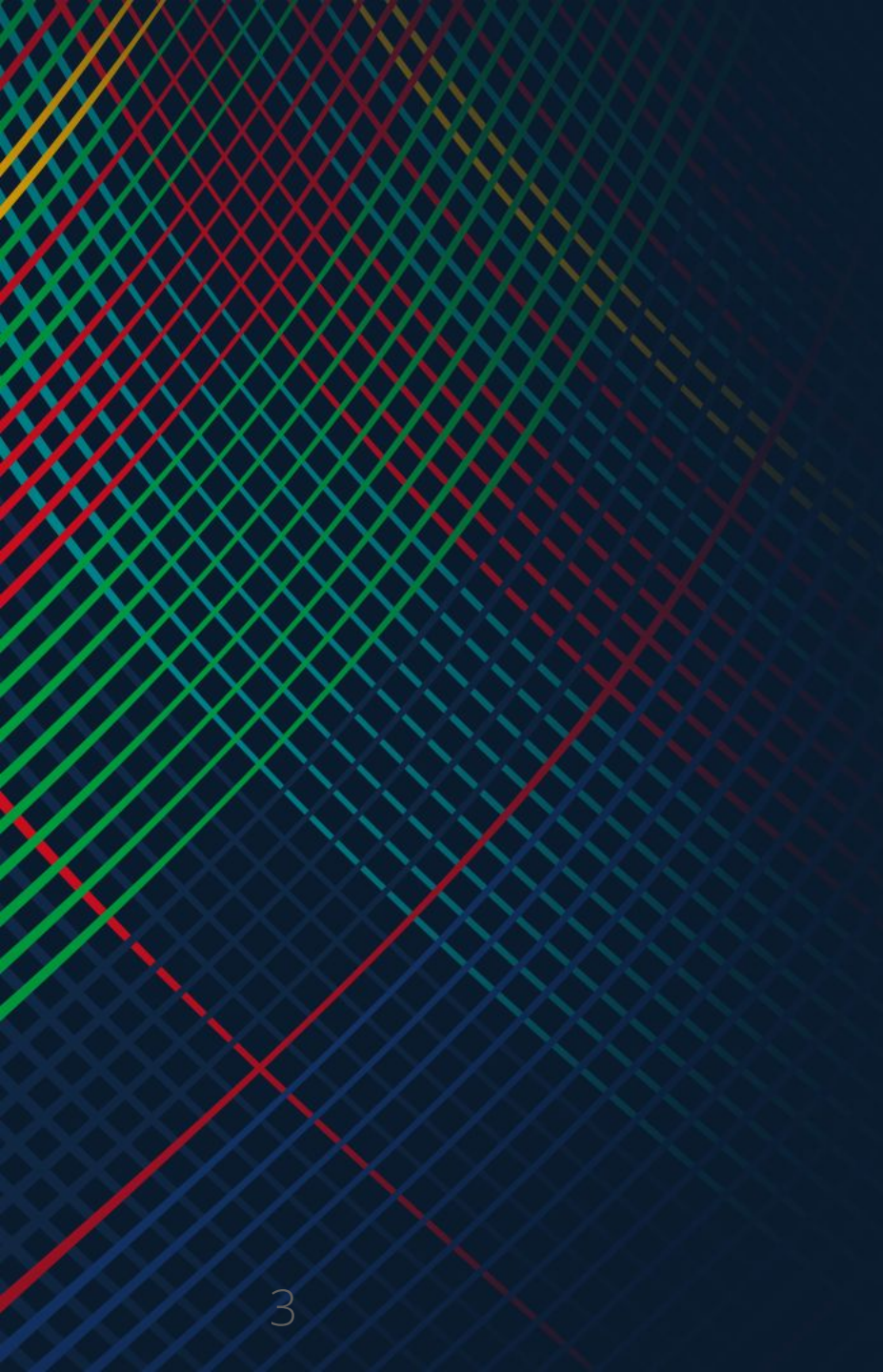
AFRINIC
Final Presentation

MAY 7TH, 2021

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

Agenda

1. Introduction
2. The Problem
3. The Solution
4. The Demo
5. Evaluation
6. Future Work



Introductions

CMU Student Team: Researchers

Isaac Manzi

CMU Africa – MSIT Program

Esther Kamau

CMU Africa – MSIT Program

Blaise Viateur Niyigena

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

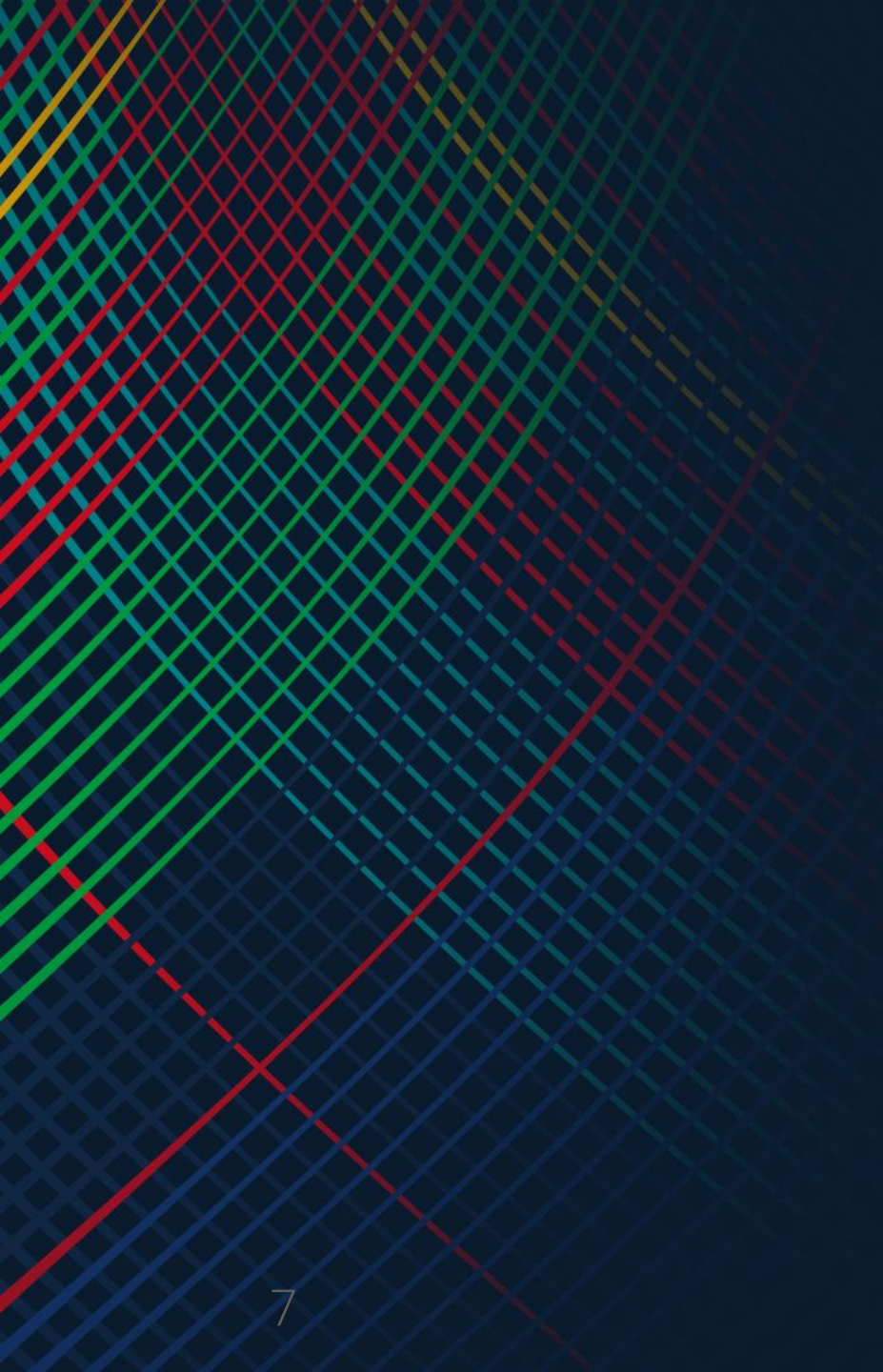
Client: Dr. Amreesh Phokeer

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



The Problem

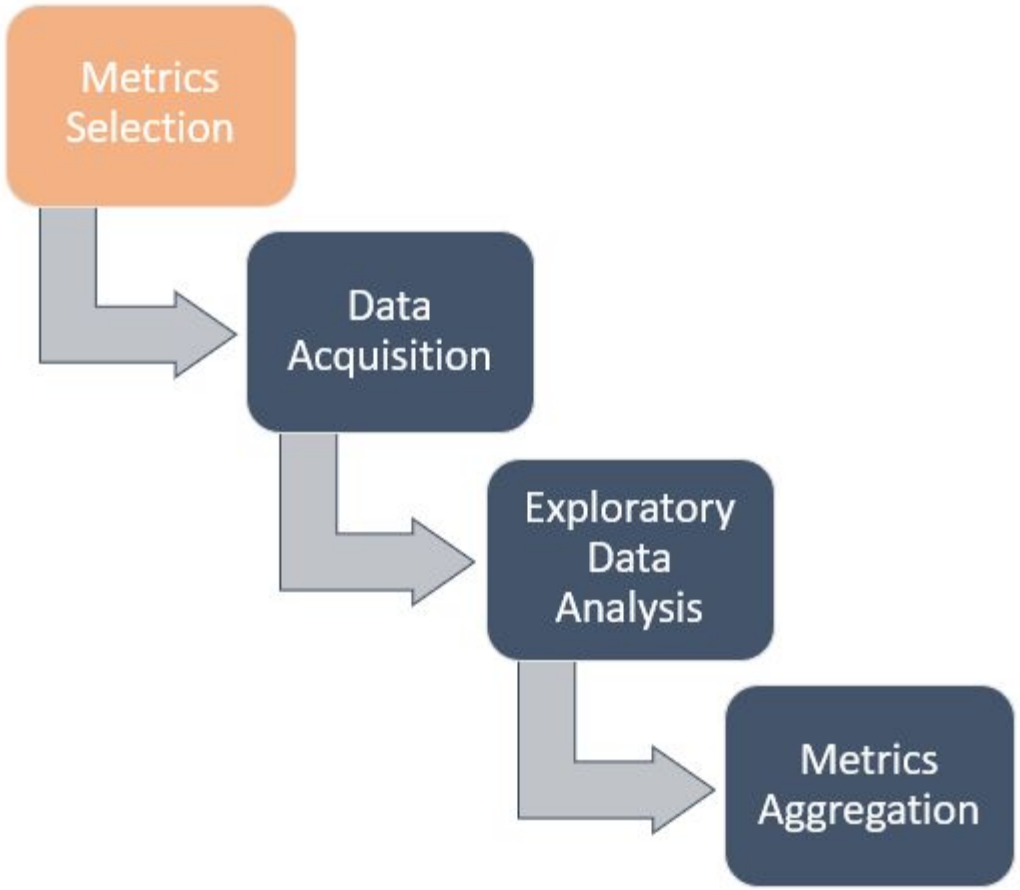
The Problem

- The AFRINIC organization currently lacks a proper metric aggregation method that qualitatively represents and communicates the state of a country's or region's network resilience to its end users and stakeholders.
- To achieve this, a metrics aggregation method will be delivered as a way for AFRINIC to efficiently gauge and inform network operators, ISPs, regulators, and other end users of the network resilience in any city, country or a region.

The Solution

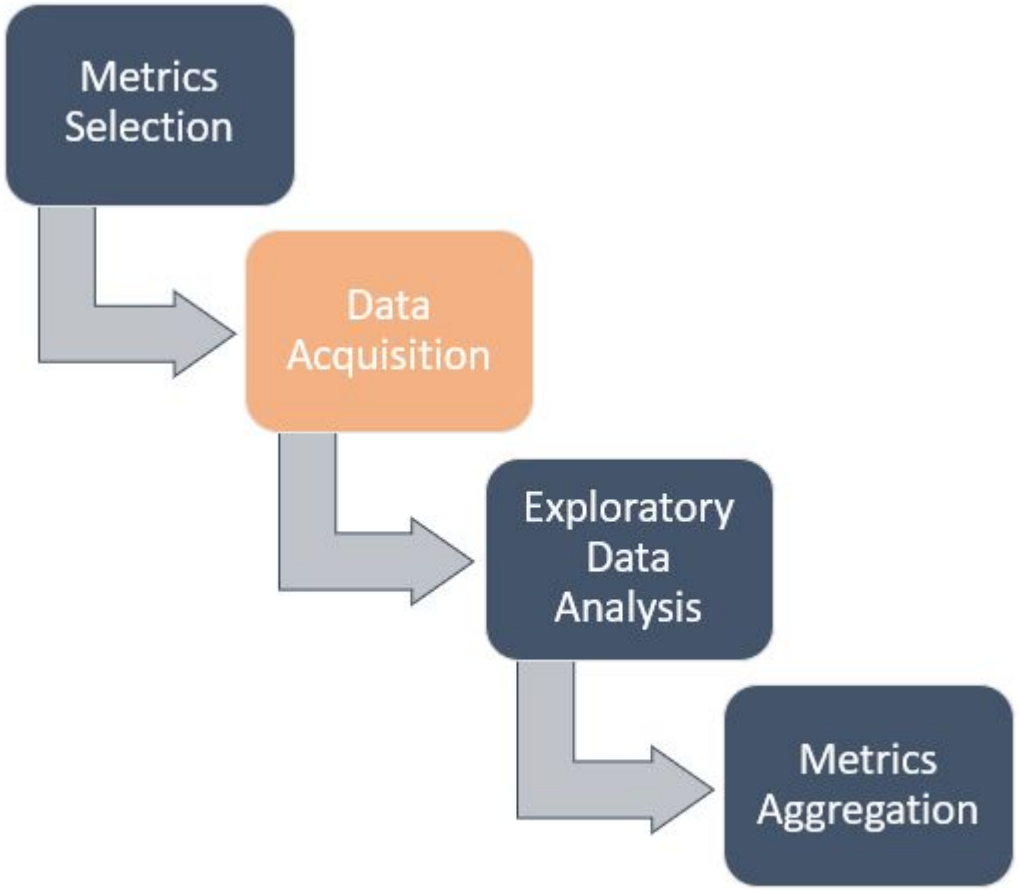
Metric Selection

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



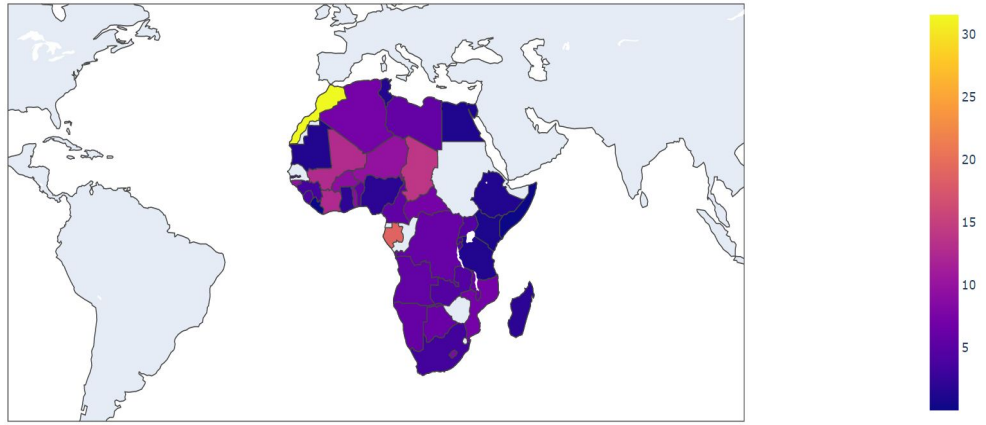
Data Acquisition

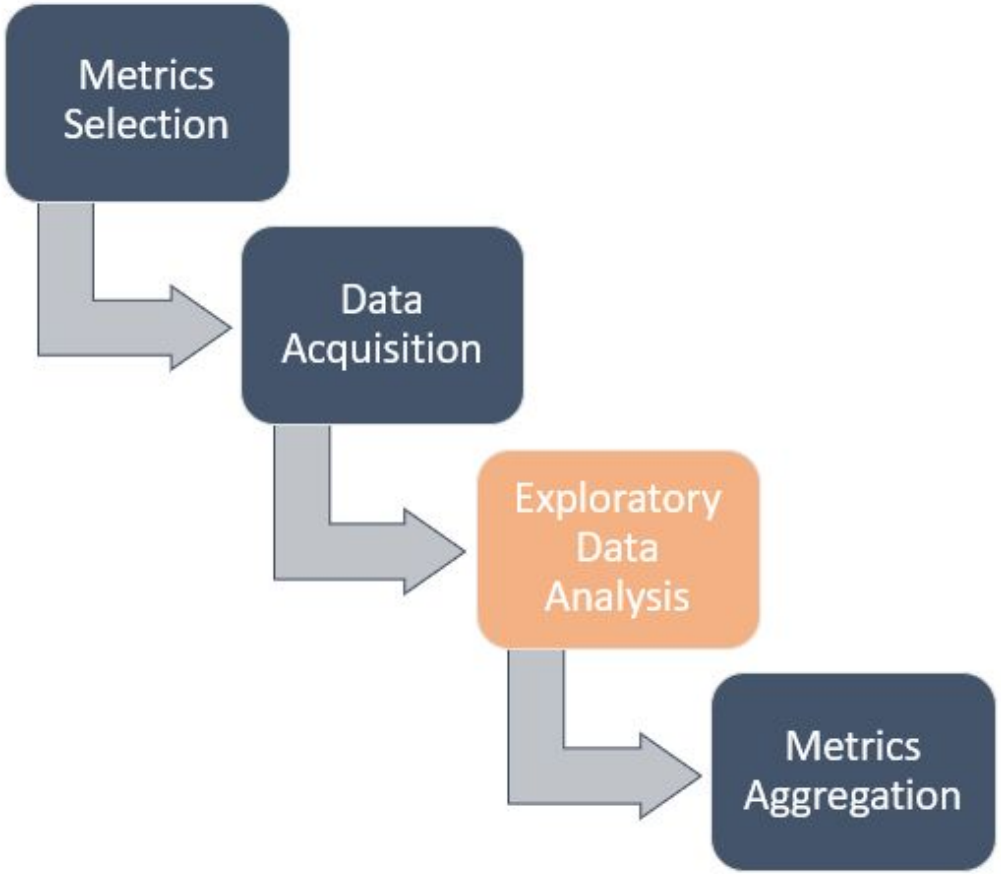
- Open source, AFRINIC or other internal collections
- Various formats – json, csv, text files
  - All standardized to pickle files
- Between 37 and 57 countries represented per dataset
- Measurements obtained between years of 2019-2020



Exploratory Data Analysis

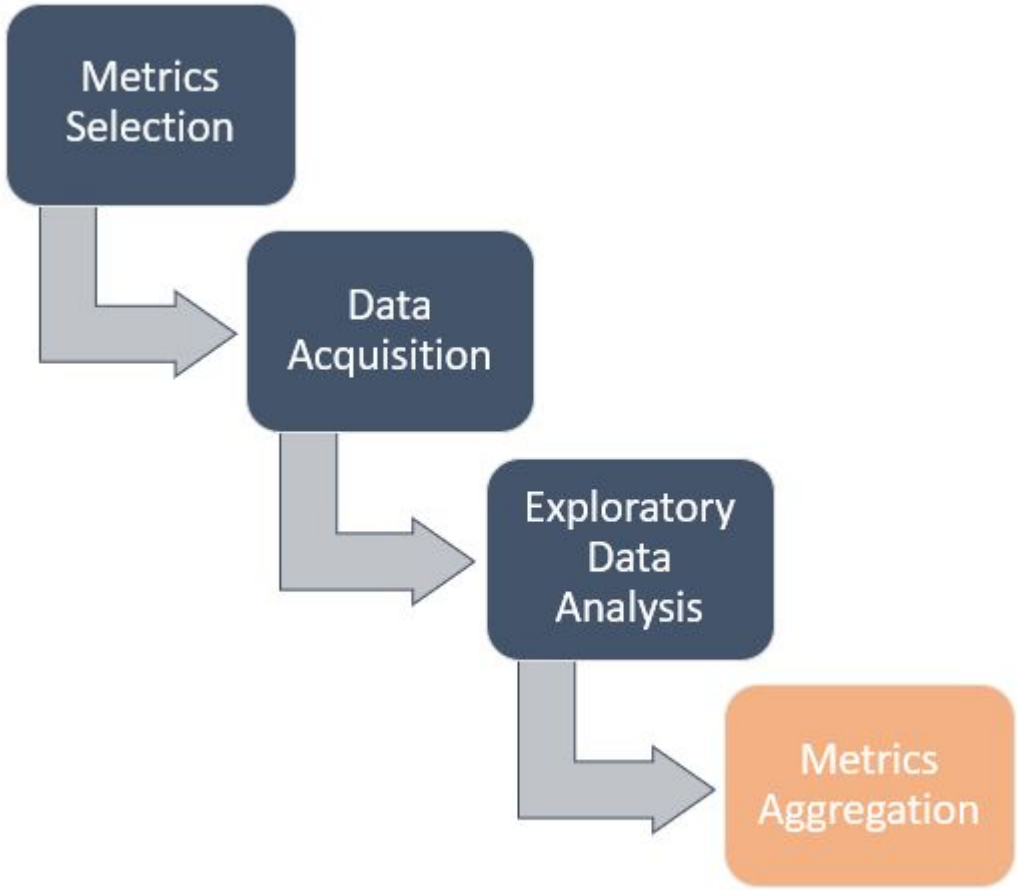
- Pandas Profiling Library
- Choropleth maps using Python Plotly





Metrics Aggregation

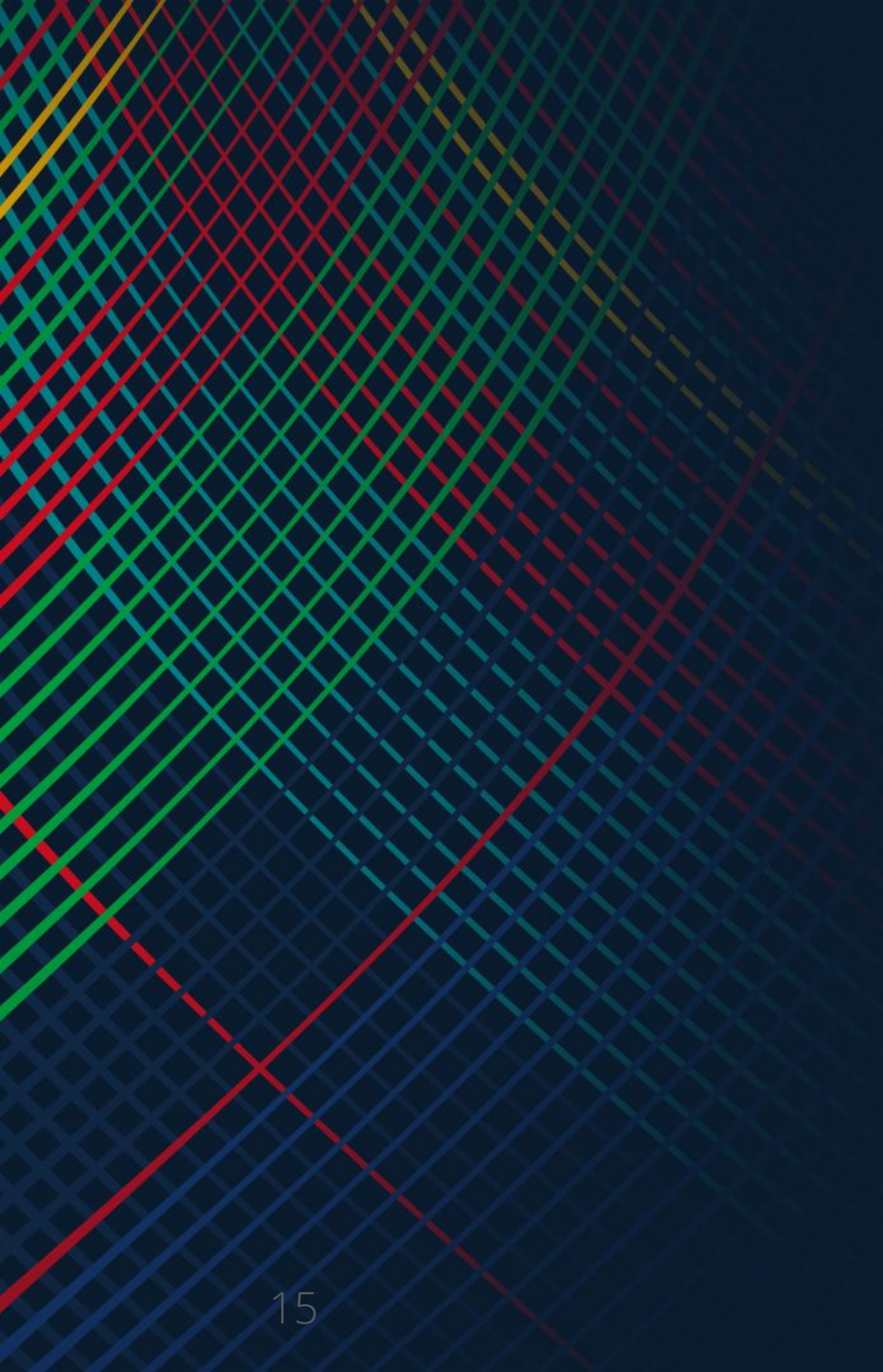
- Grouped based on type of measurement
- Categories primarily influenced by MIRA white paper, The Economist's "The Inclusive Internet Index 2020 Methodology report"
- Data coverage, availability and importance to describing each category



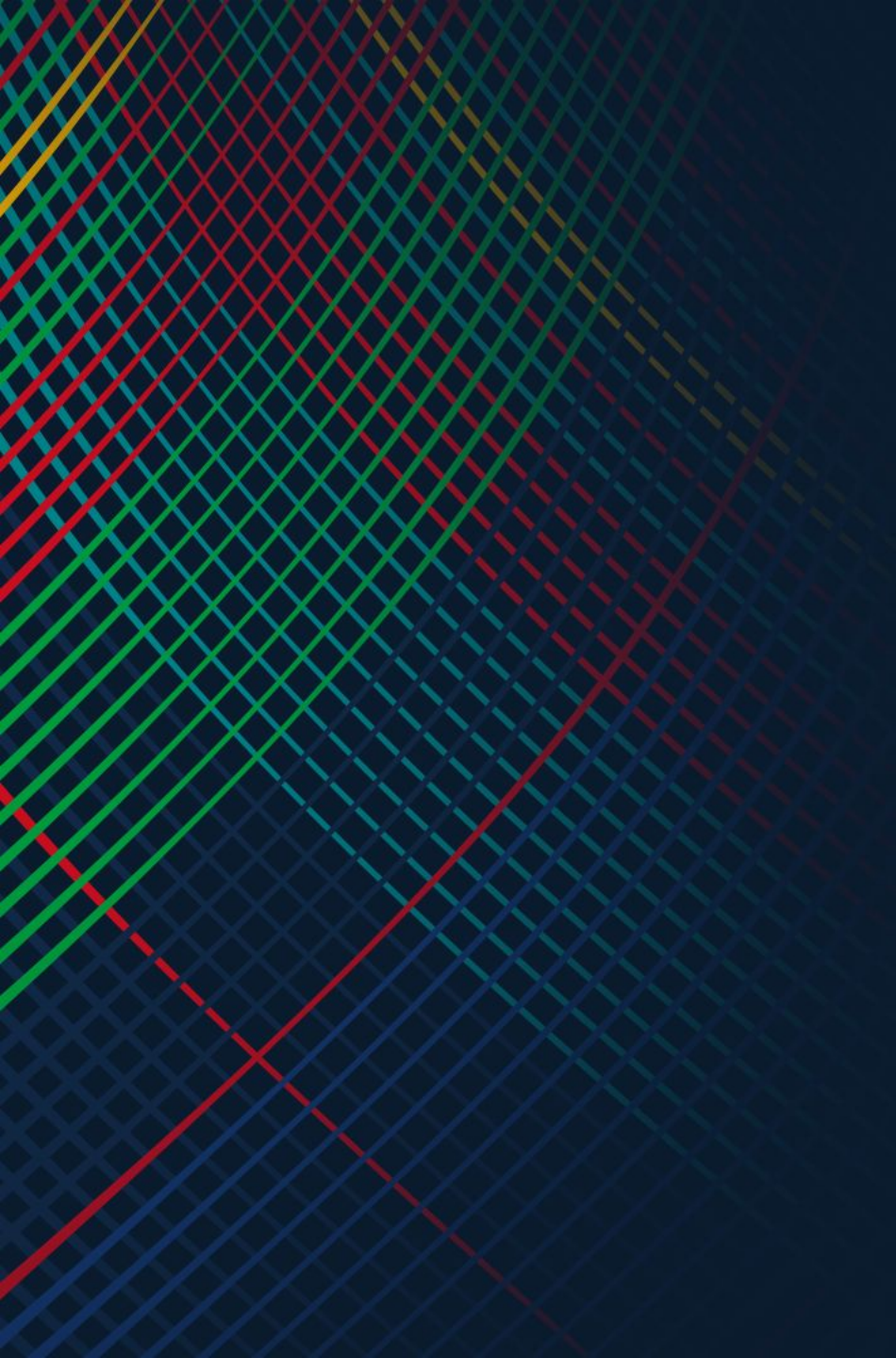
Simulations

- Python Dash, Plotly, Excel
- 4 Categories: Quality of Service, Security, Infrastructure, Affordability
- 1-6 metrics per category
- Geometric aggregation formula of sub-indicators

$$Y = w_{c_1} * (w_{m_1 1} * m_1 2 + \dots) + w_{c_2} * (w_{m_2 1} * m_2 1 + \dots) + \dots$$



The Demo



Evaluation

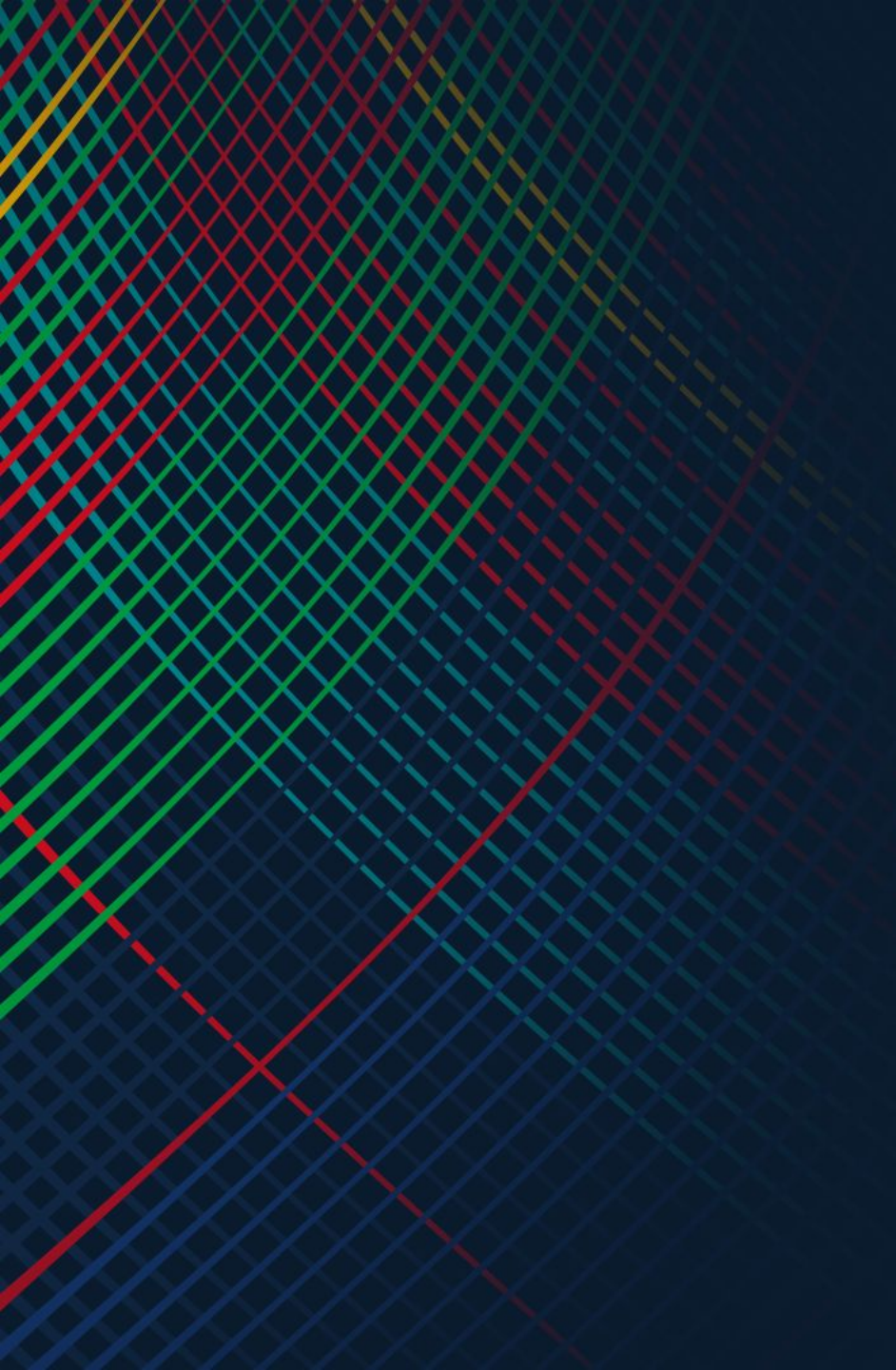
Evaluation

- Past research work

  We used some of the publications on past work done on measurement of Internet resilience as a benchmark for our results.

- Expert opinion

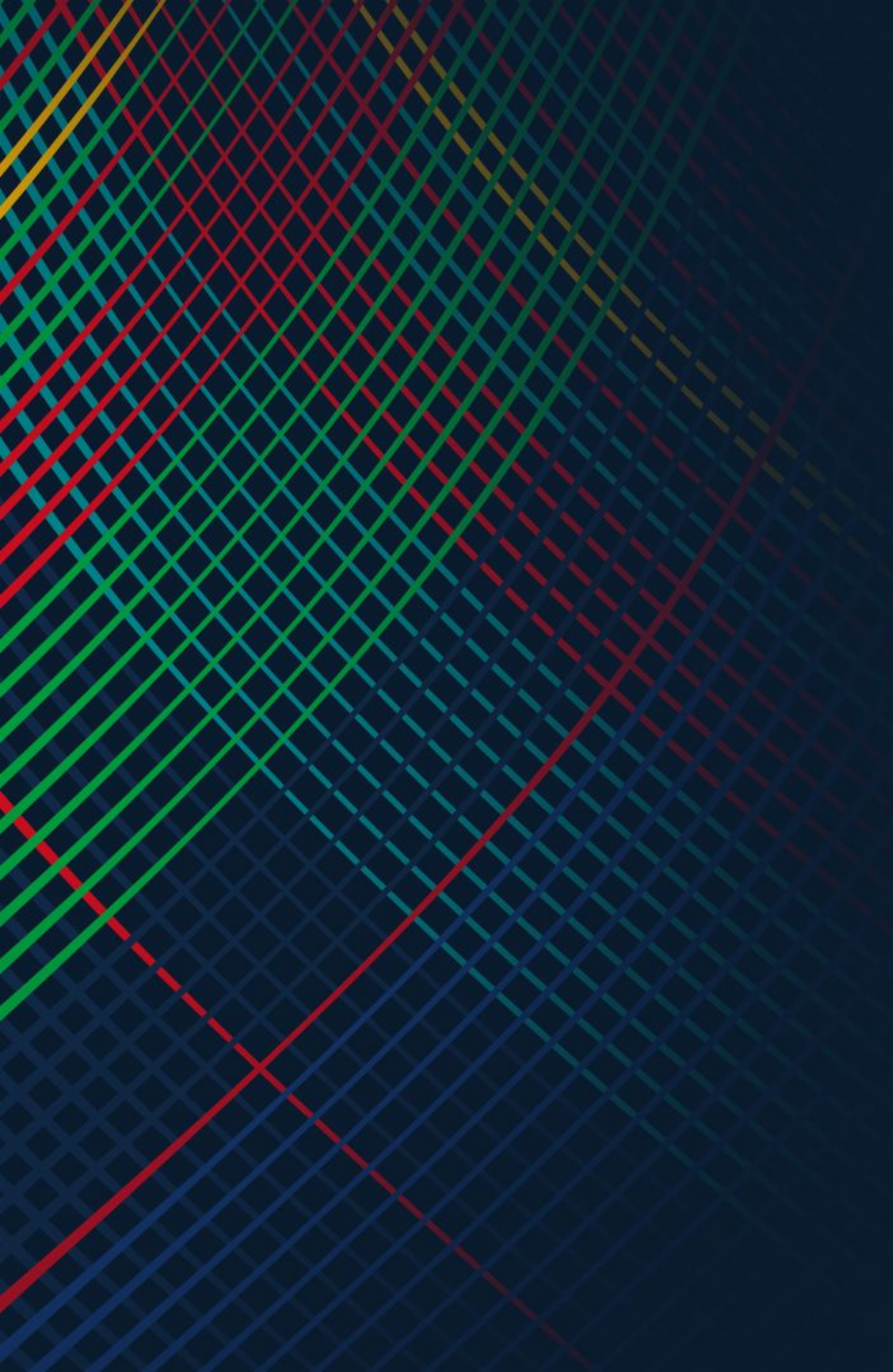
  We had discussions with our client, Dr. Amreesh, and his team on the different stages of our project



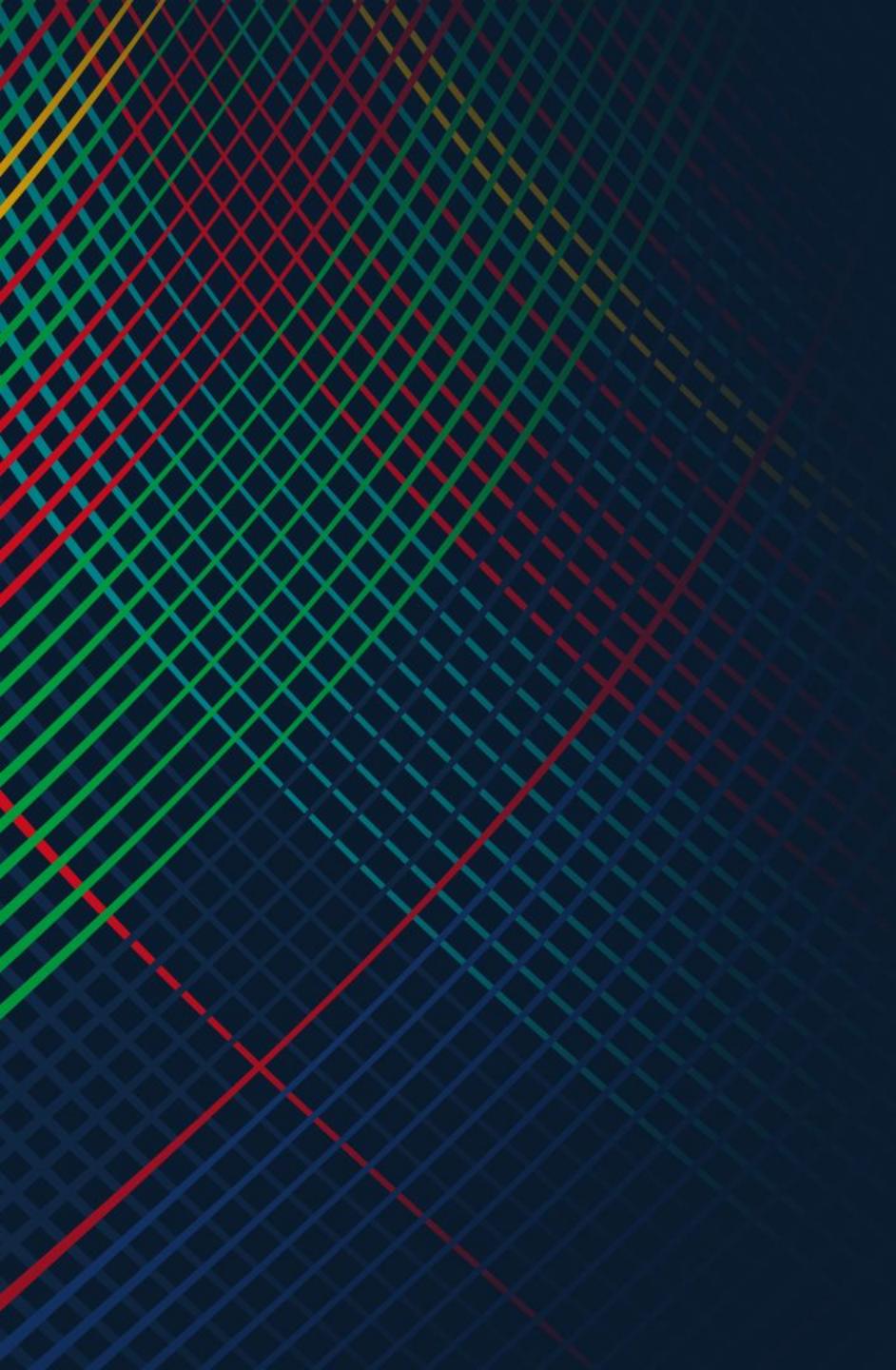
Future Work

Future Work

- Update Subject Expert Harnessing tool to receive feedback from Subject Matter Experts and to help validate final model
- Collection of unavailable data per selected metric using Internet Probes
- Expanding the usage of this tool to countries in other continents
- Discussion of further study of Internet Measurement Probes and benefits, other measurements that can be incorporated
- Implementation of database for potential recalibration and updating of model parameters and data



Thank you!



Questions?

Technical Appendix

Demo Cloud Access

Link:
https://drive.google.com/file/d/1GHTEofDt4wMaFaz4d8K8l0Goc34z61XN/view?usp=sharing

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

Category	Metric	Proposed Individual Metric Weight According to Internet Lifecycle	Proposed Individual Metric Weight	Measurement
QoS - 25%	Throughput - Download Speed Fluctuations		8.35%	
	Throughput - Download Speed	Availability - Quality (very important)	8.35%	Fluctuation in throughput - download speeds
	Throughput - Upload Speed Fluctuations		8.35%	
	Throughput - Upload Speed	Availability - Quality (very important)	8.35%	Fluctuation in throughput - upload speeds
	Latency Fluctations		16.70%	
	Latency	Availability - Quality (very important)	16.70%	Latency to local services (ms)
	IPv6 capability	Availability - Quality (very important)	33.30%	IPv6 capability of the ISP network (count)
Security - 25%	MANRS score (Routing regulations)	Availability - Quality (very important)	25%	% of prefixes covered by IRR object
	AS hegemony	Availability - Quality (very important)	25%	Compute the AS dependency of network
	DDos Potential	Availability - Quality (very important)	25%	Level of risks posed to other countries
	Spam Infection	Availability - Quality (very important)	25%	%
Infrastructure - 35%	IXP efficiency	Availability - Infrastructure (very important)		% of ASes present at the IXP
	Upstream	Availability - Quality (very important)	33%	Number of upstream providers
	Cable landing stations	Availability - Infrastructure (very important)	16.70%	Number of cable landing stations per capita/km2
	reach	Availability - Infrastructure (very important)	33%	% of population within 10-Km reach
	degree distribution	Availability - Infrastructure (very important)	16.70%	Degree distribution of cable entering/leaving a country/city
Affordability - 15%	Affordability	Affordability - Price	100%	How affordable is Internet services in this country ($)

Data Preprocessing

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

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

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

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

Category	Metric	Proposed Individual Metric Weight According to Internet Lifecycle	Proposed Individual Metric Weight	Measurement
QoS - 25%	Throughput - Download Speed Fluctuations		8.35%	
	Throughput - Download Speed	Availability - Quality (very important)	8.35%	Fluctuation in throughput - download speeds
	Throughput - Upload Speed Fluctuations		8.35%	
	Throughput - Upload Speed	Availability - Quality (very important)	8.35%	Fluctuation in throughput - upload speeds
	Latency Fluctations		16.70%	
	Latency	Availability - Quality (very important)	16.70%	Latency to local services (ms)
	IPv6 capability	Availability - Quality (very important)	33.30%	IPv6 capability of the ISP network (count)
Security - 25%	MANRS score (Routing regulations)	Availability - Quality (very important)	25%	% of prefixes covered by IRR object
	AS hegemony	Availability - Quality (very important)	25%	Compute the AS dependency of network
	DDos Potential	Availability - Quality (very important)	25%	Level of risks posed to other countries
	Spam Infection	Availability - Quality (very important)	25%	%
Infrastructure - 35%	IXP efficiency	Availability - Infrastructure (very important)		% of ASes present at the IXP
	Upstream	Availability - Quality (very important)	33%	Number of upstream providers
	Cable landing stations	Availability - Infrastructure (very important)	16.70%	Number of cable landing stations per capita/km2
	reach	Availability - Infrastructure (very important)	33%	% of population within 10-Km reach
	degree distribution	Availability - Infrastructure (very important)	16.70%	Degree distribution of cable entering/leaving a country/city
Affordability - 15%	Affordability	Affordability - Price	100%	How affordable is Internet services in this country ($)