Carnegie Mellon University

AFRINIC
Progress Report 2

MARCH 31ST, 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 & Roadblocks

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

Profession: Research Manager at AFRINIC

Research Interests:

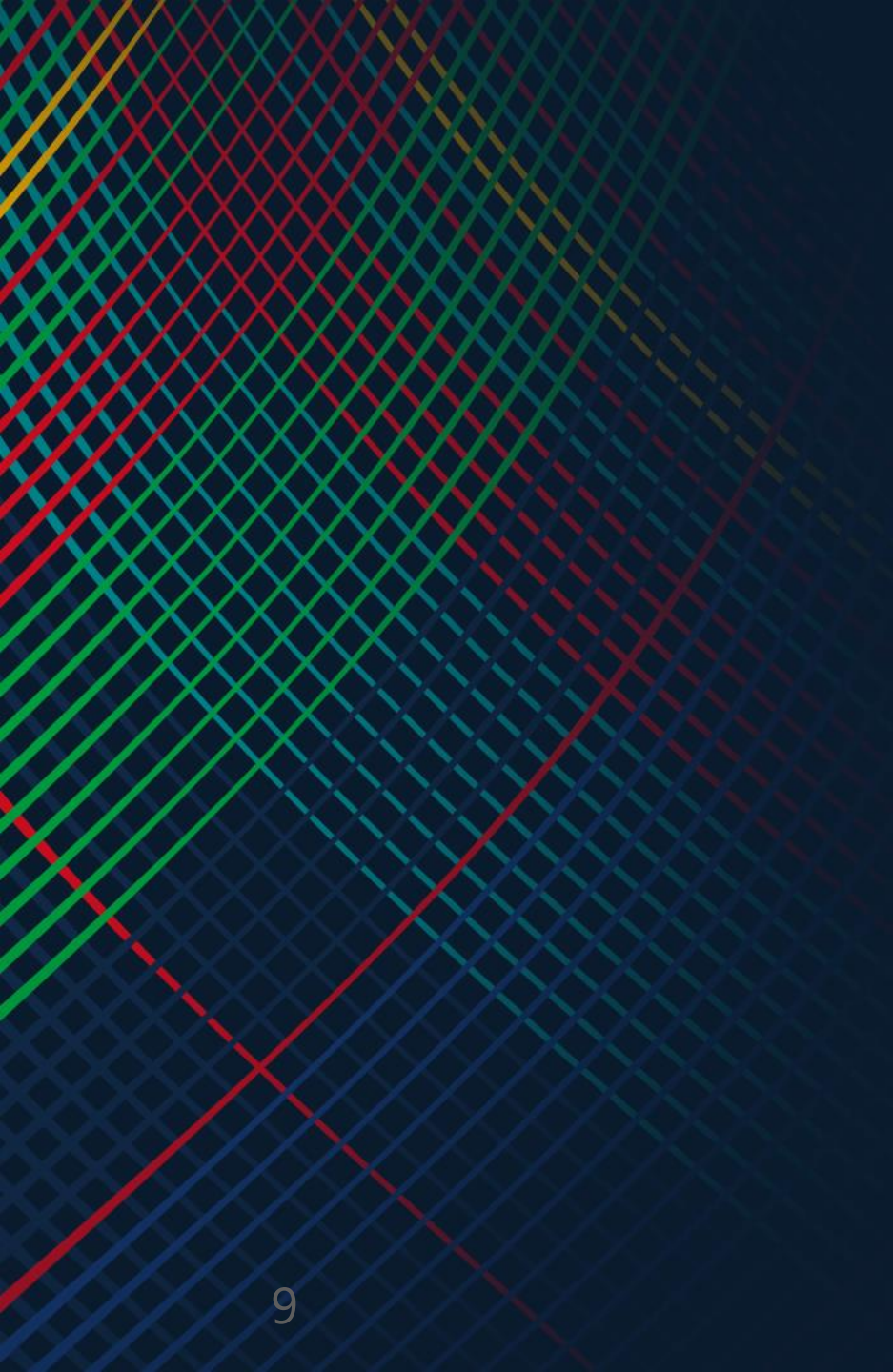
- Interdomain routing
- Network Security
- Internet Measurements
- Software Design

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 measure internet infrastructure resiliency

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

- 14* datasets used – one for each metric of interest
  - Most datasets obtained – 2 are still being downloaded
  - Open source and internal collections
- Preprocessing performed on some of datasets
  - Mainly to validate data quality and determine which countries have missing values

*Not including Geojson data needed to create dashboard

Category	Sub-category	Metric	Unit	Measurement
ISP Resilience	QoS	Uptime	%	Daily or weekly downtime/uptime ratio
		Throughput	mbps	Fluctuation in throughput
		Latency	ms	Latency to local services
	Link Resilience	Upstream	#	Number of upstream providers
		Congestion	?	Link congestion
		IPv6 capability	%	IPv6 capability of the ISP network
	Routing		%	% of prefixes covered by IRR object
		Security	%	% of prefixes covered by ROA
		AS hegemony	#	Compute the AS dependency of network
		MANRS Score	#	The consolidated MANRS score
Country-level resilience	Security	DDos Potential		Level of risks posed to other countries
	Spam	Infection	%	Composite Black List
	IXP	IXPs per 10M	#	Number of IXPs per 10M inhabitants
		IXP efficiency	%	% of ASes present at the IXP
	Cable ecosystem	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
	Market resilience	Affordability	index	How affordable is Internet services in this country

Data Description

- Ideally, want 54 entries per dataset (one per African country)
- Realistically ~48 entries per explored dataset
  - ~6 countries missing
  - Eritrea and South Sudan (eastern African countries) usually missing from data

Numerical Summaries - Affordability

- 37 countries represented
- Median cheapest prepaid mobile voice product in Africa
  (30 Calls / 100 SMS) is $5.47
- Maximum prepaid plan is $30.78

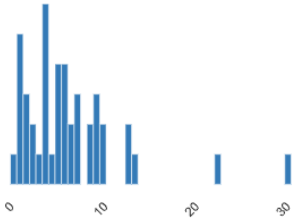
2020 Q2

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
MISSING

Distinct	44
Distinct (%)	100.0%
Missing	7
Missing (%)	13.7%
Infinite	0
Infinite (%)	0.0%
Mean	6.4825

Minimum	0.03
Maximum	30.78
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	536.0 B



Toggle details

Statistics | Histogram | Common values | Extreme values

Quantile statistics

Minimum	0.03
5-th percentile	1.1905
Q1	2.8925
median	5.47
Q3	8.6975
95-th percentile	13.2495
Maximum	30.78
Range	30.75
Interquartile range (IQR)	5.805

Descriptive statistics

Standard deviation	5.701391558
Coefficient of variation (CV)	0.879505061
Kurtosis	7.808477698
Mean	6.4825
Median Absolute Deviation (MAD)	2.92
Skewness	2.391350771
Sum	285.23
Variance	32.5058657
Monotocity	Not monotonic

Numerical Summaries - Nodes per Link

- Less nodes per link, the better
- 50 countries represented
- Median nodes per link is about 1
- Maximum nodes per link is about 5

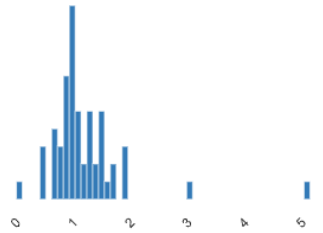
Q2_2020

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
ZEROS

Distinct	51
Distinct (%)	91.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.191016753

Minimum	0
Maximum	5.142857143
Zeros	1
Zeros (%)	1.8%
Negative	0
Negative (%)	0.0%
Memory size	576.0 B



Toggle details

Statistics | Histogram | Common values | Extreme values

Quantile statistics

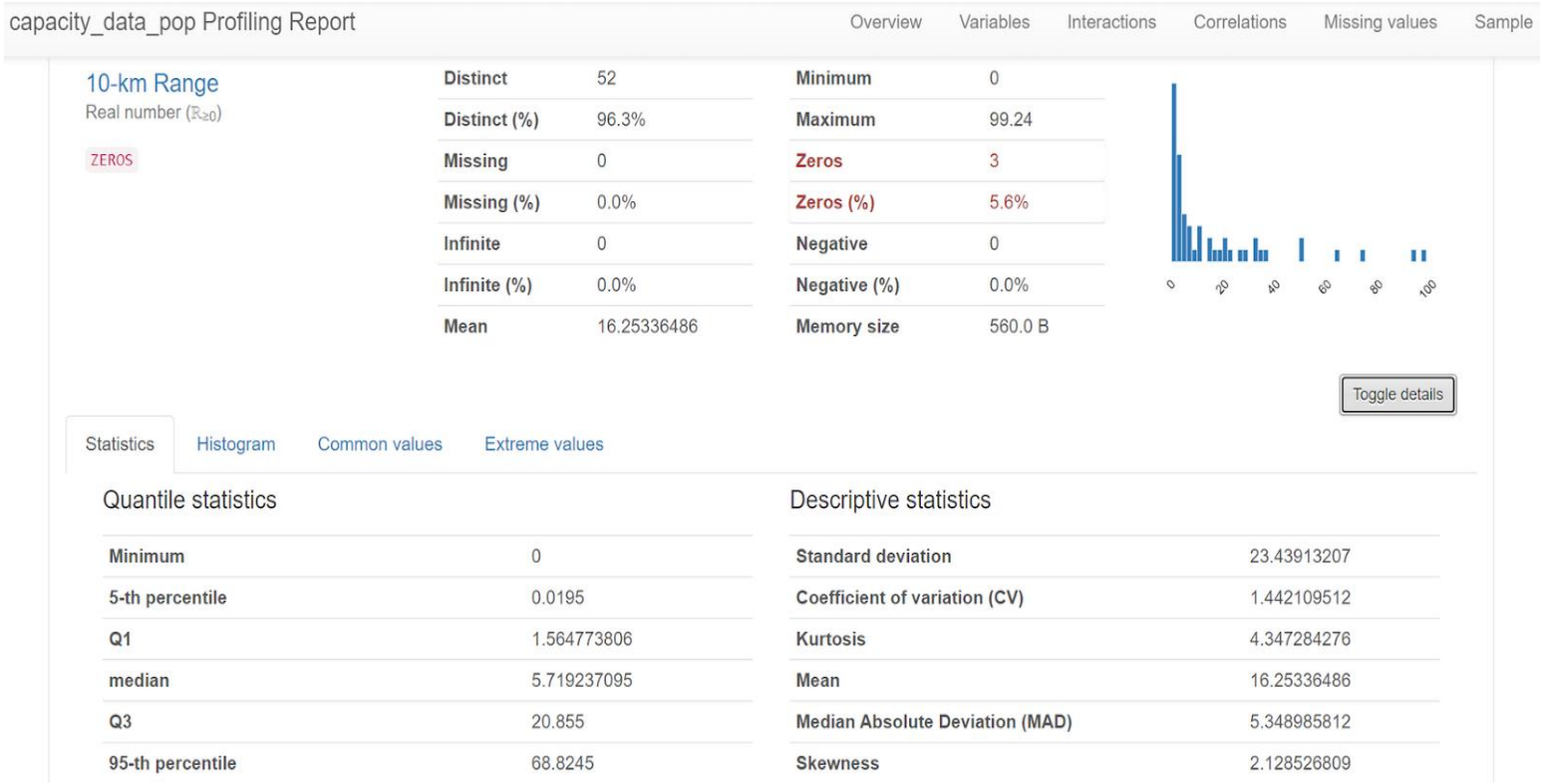
Minimum	0
5-th percentile	0.5
Q1	0.8623439667
median	1.026709402
Q3	1.402632827
95-th percentile	1.876309914
Maximum	5.142857143
Range	5.142857143
Interquartile range (IQR)	0.5402888606

Descriptive statistics

Standard deviation	0.7029627112
Coefficient of variation (CV)	0.5902206745
Kurtosis	18.30698551
Mean	1.191016753
Median Absolute Deviation (MAD)	0.2377761128
Skewness	3.54554522
Sum	66.69693816
Variance	0.4941565733
Monotocity	Not monotonic

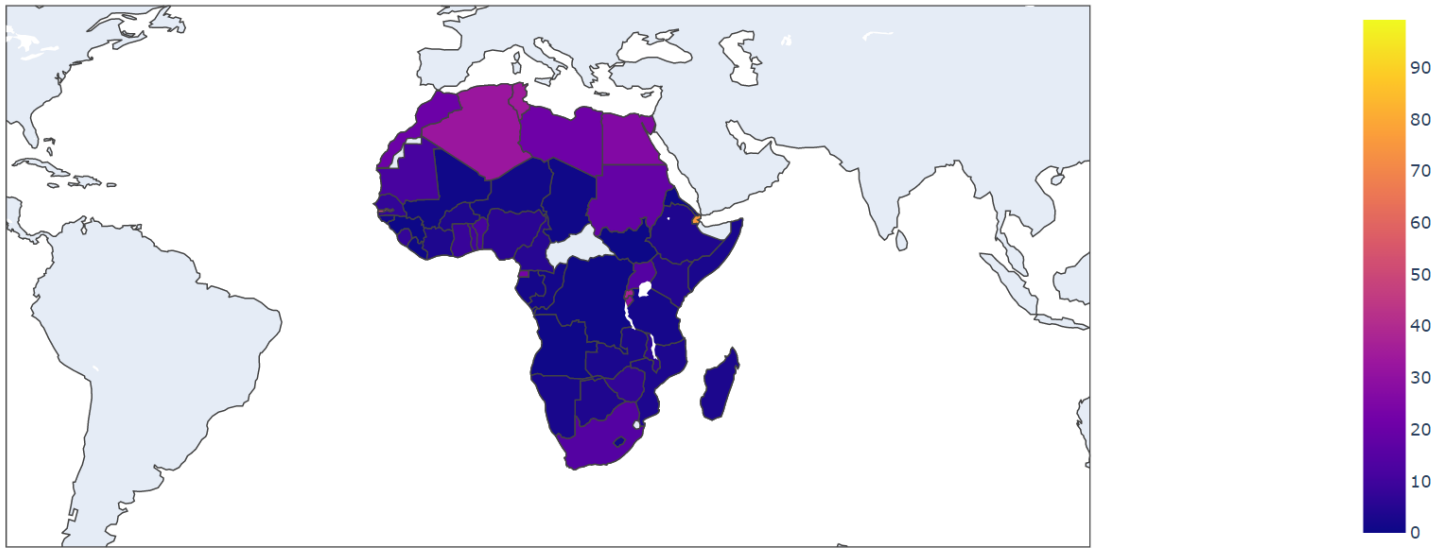
Numerical Summaries - 10 km Range

- 49 countries represented
- Median range is 5.71% coverage
- Maximum range is 99.24% coverage



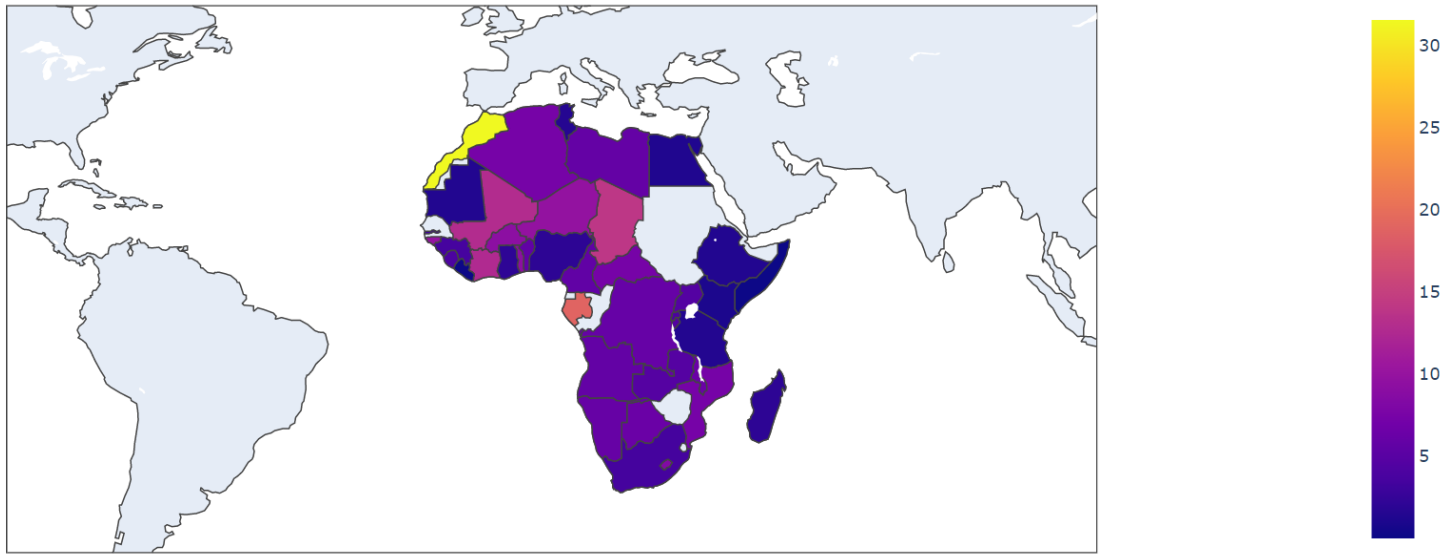
Exploratory Data Analysis – 10 Km % Reach

- Outer African countries usually have highest reach
- No country seems to have 100% coverage
- Population density may need to be considered for further analysis



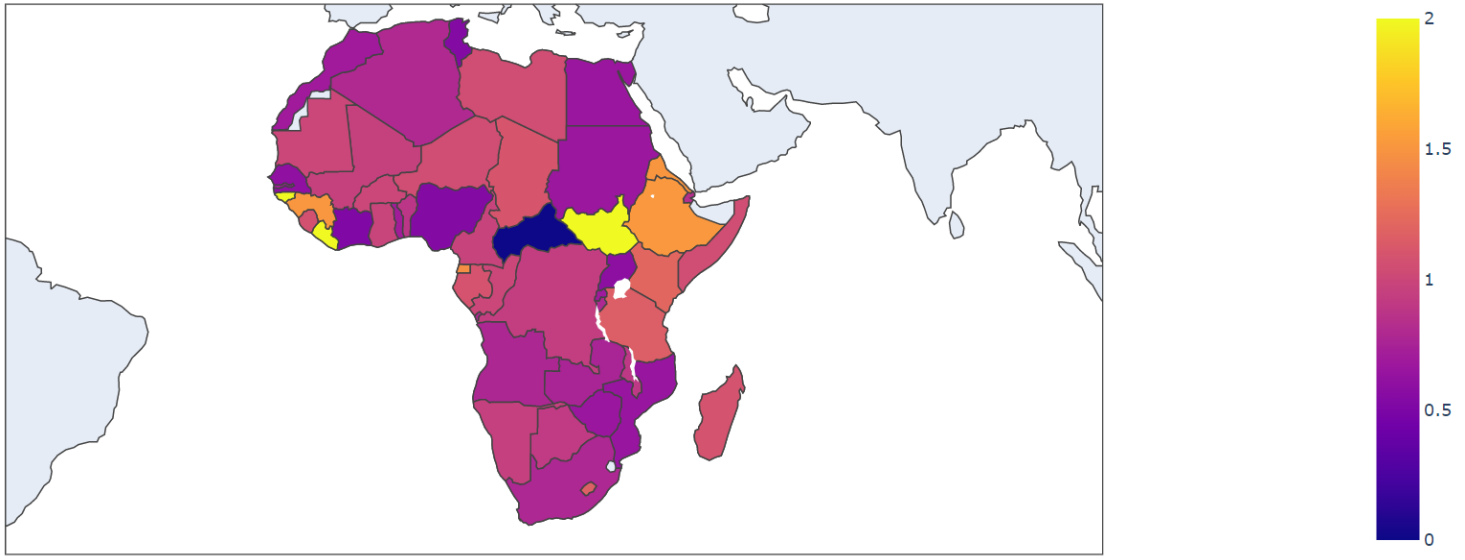
Exploratory Data Analysis – Affordability

- Cheapest prepaid mobile voice product in Africa (30 Calls / 100 SMS)
- Most expensive plans in western and north-western Africa
- Cheapest plans mostly along Eastern coastline



Exploratory Data Analysis – Nodes per Link

- Not as discernible of patterns here
- Countries near each other clustered in same color



Methods

Current Methodology

1. Select metrics based on relevance, independence, and data accessibility

   - Exploring internet measurements probe to have better understanding of data procurement process

2. Use a combination of research findings, intuition, subject matter opinions and various statistical methods for weighing/ranking to determine weights

   - Internet Lifecycle: Availability -> Affordability -> Relevance -> Readiness

3. Utilize web scraping to extract datasets

   - Document and automate extraction process

4. Possibly use Python dash library to create dashboard

5. Working in parallel for all above, potentially with Google Collab

Internet Probe

- Raspberry Pi
- Obtain various Internet measurements
- Uninterrupted measurements
- Low power hardware
- Easily obtain measurements in many parts of Rwanda
- Deployed at CMU Africa

Results

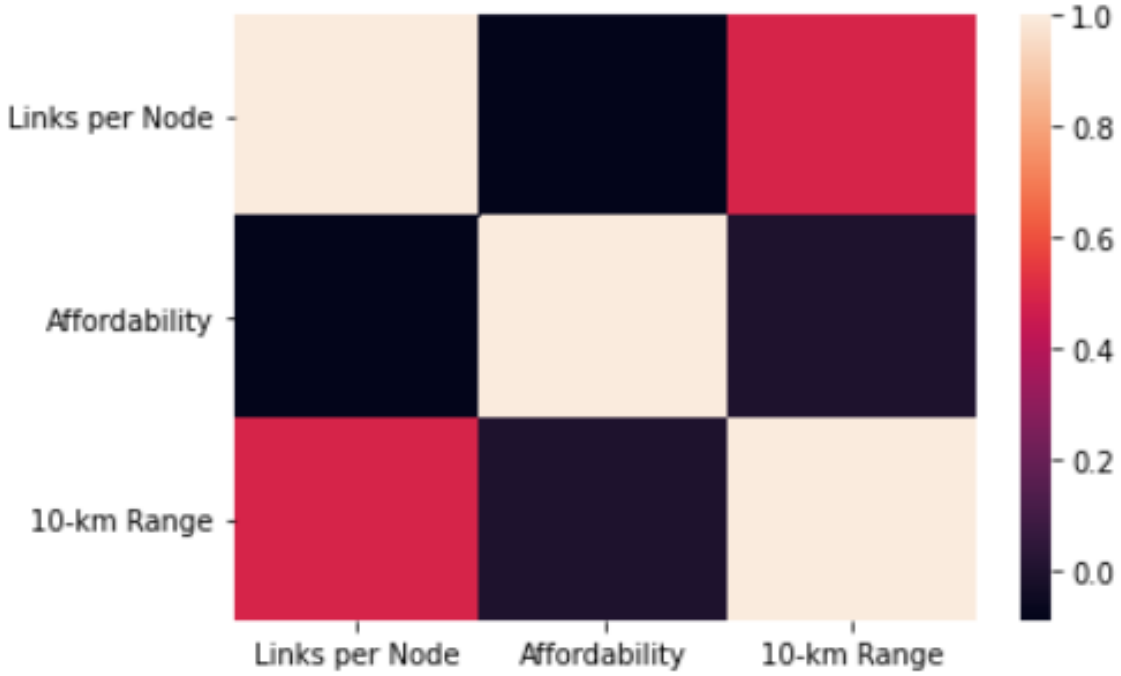
Current Results

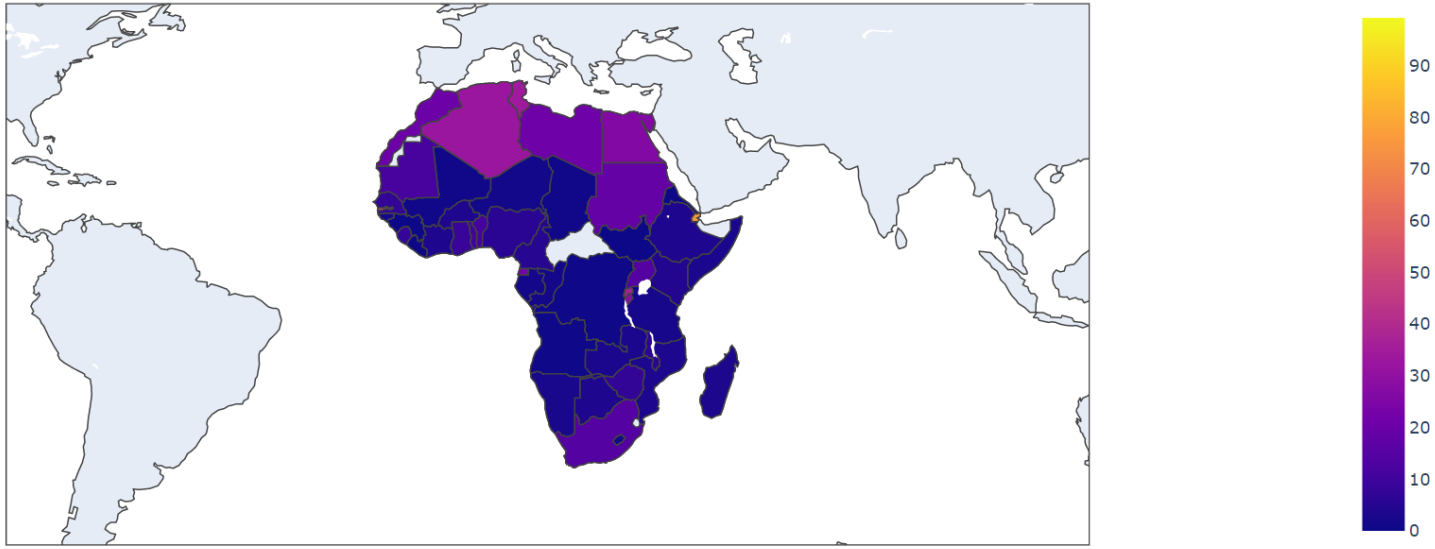
1. Mainly data exploration

2. Exploring ranks of various metrics per country

    - Large correlation between 10-km Range and Links per Node

1. Utilization of Python Plotly and Dash for dashboard





Next Steps & Roadblocks

Timeline

1. All deliverables are currently being worked on in parallel

2. We most importantly need to obtain and explore the last datasets to determine weights

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. Potential revision of subject matter harnessing tool to justify ranks per metric
5. Interactive dashboard that not only displays our score, but allows users to select weights per metric
6. Possible ad-hoc clustering analysis on countries based on our index
7. Research Paper/Final Presentation

Roadblocks

1. Clarification on measurements, frequency of measurements
2. Merging datasets
3. We are new to dash (mostly experience using R Shiny), so there is a learning curve to be expected
4. A lot of expected deliveries and how to best allocate tasks
5. Processing power, computing resources (certain metric data is GBs in size)

Questions

If any questions arise, please direct them to the following:

Blaise Viateur Niyigena: bniyigen@andrew.cmu.edu

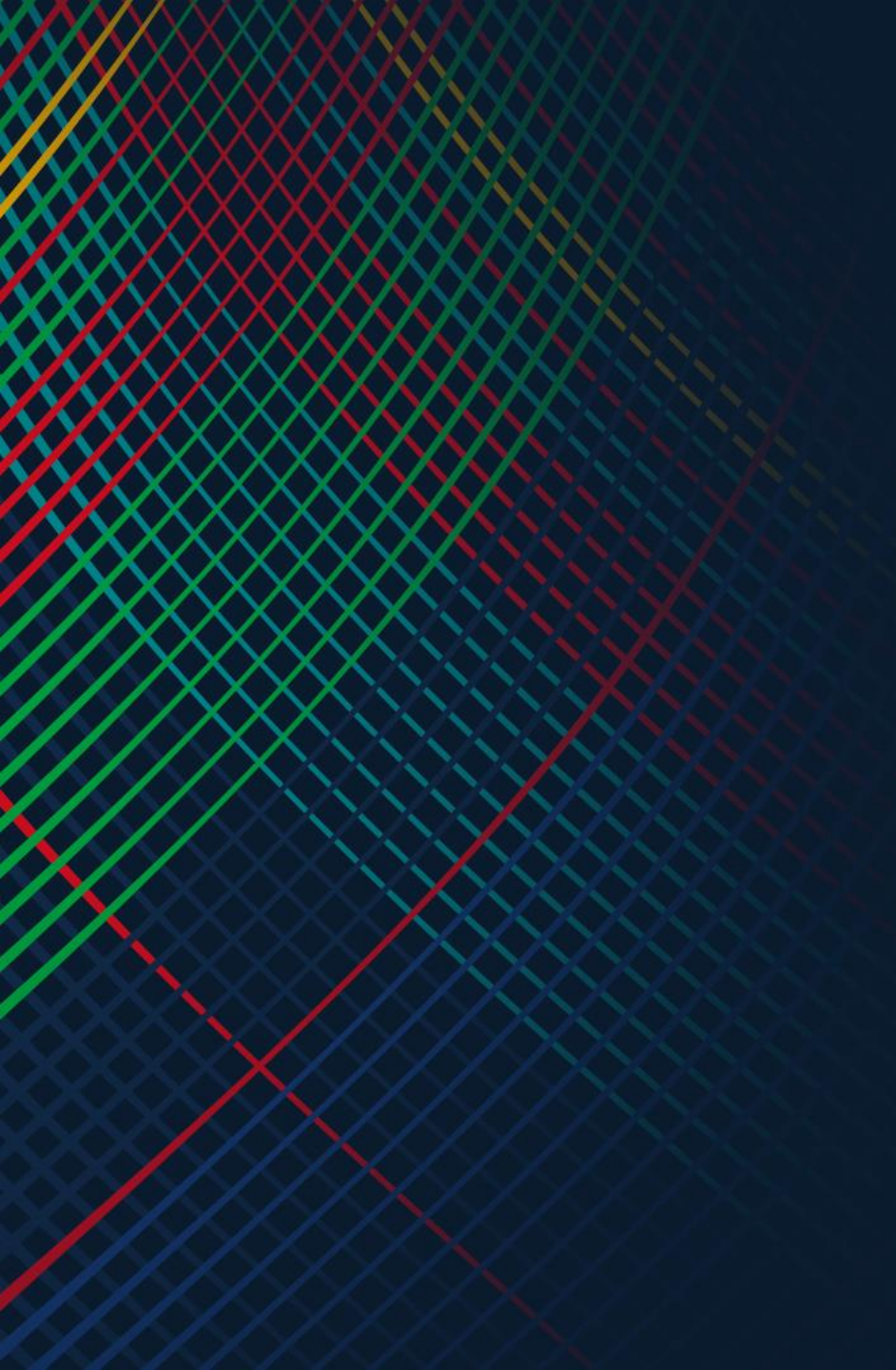
Esther Kamau: enkamau@andrew.cmu.edu

Isaac Manzi: isaacm@andrew.cmu.edu

Pasqua Ruggiero: pasquar@andrew.cmu.edu

Dr. Assane Gueye: assaneg@andrew.cmu.edu

Dr. Amreesh Phokeer: amreesh@afrinic.edu



Thank you!