Kenya's Digital Economy

A People's Perspective

METHODOLOGY NOTE

Dalberg

Kenya's Digital Economy

A People's Perspective

SAMPLING METHODOLOGY FOR THE KENYA'S DIGITAL ECONOMY: A PEOPLE'S PERSPECTIVE SURVEY

1. IN-DEPTH SURVEY

The aim of the Kenya's Digital Economy: A People's Perspective study was to understand the level of digital access, usage, and sentiments across digital services in the various counties and population segments in the country, as experienced by its residents. To achieve this, we needed to undertake a nationally representative survey of persons aged 15 years or more, who constituted the eligible population for the research at the household level.

The survey was CAPI-based and targeted a total of 2,431 respondents, out of which 2,360 cases were to be selected randomly, and 71 cases selected purposively. Purposive selection was included to ensure adequate representation of all age groups targeted in the survey. We assumed a 95% confidence level and a margin of error of +5% for the 2,360 cases for random selection.

The sampling protocol in the survey was as follows:

Identification of the Target Primary Sampling Units (PSUs) and Enumeration Areas (EAs)

We used proportionate sample distribution to determine the number of interviews to administer across the survey points, based on the 2019 Kenya National Bureau of Statistics (KNBS) population census as the sample frame. In the sample frame, the population data was organized by the local administrative units in Kenya, from the largest to the smallest, i.e., County → Sub County → Ward → Location → Sub Location, and disaggregated by a number of variables, including location (rural/urban), gender, and age.

We used counties as the primary sampling units (PSUs), and sub locations[1] as the enumeration areas (EAs) in the survey. All the 47 counties in Kenya were surveyed.

To determine the number of EAs to target in the survey, we opted to target 20 households per enumeration area so that overall, we targeted 118 EAs i.e., 2,360/20. Our choice for 20 cases per EA was informed by the need to balance the overall geographical spread of the survey respondents and the related overall travel cost in the survey. Generally, the lower the number of respondents per EA, the higher the travel cost, and vice versa. We used proportionate sample allocation to distribute the 118 EAs across the 47 PSUs, observing the prevailing national population ratio for the rural/ urban split. PSUs with higher populations had more EAs than those with lower populations. Further, we considered and utilized in the survey the national gender (male/female), and age ratios as computed from the sample frame above. The applicable age ratio was for the eligible population only, which was banded into 3, i.e., 15 – 34 years, 35 – 54 years, and 55 years and above.

Specifically, we used the following national ratios for the location, gender and age splits during sampling in the survey:

Location: 69% rural and 31% urban

Gender: 50% male and 50% female. The exact national population gender ratio for the eligible population was 48.98% male and 51.02% female.

Age: 59% for the 15 – 34 years band, 28% for the 35 – 54 years band, and 13% for the 55 years and above band.

1 Sub location: the smallest local administrative unit in Kenya

The final target sample distribution in the survey was as summarized in Table 1 below:

Table 1: Targeted sample distribution by age-band and respondent selection approach, per county

COUNTY		Age (in years)					
		15-34	35-54	55+			Total
		Random			Purposive	Total	
1	Mombasa	35	17	8	2	10	**62**
2	Kwale	24	11	5	1	6	**41**
3	Kilifi	47	22	11	2	13	**82**
4	Tanariver	12	5	3	1	4	**21**
5	Lamu	12	5	3	0	3	**20**
6	Taita/Taveta	12	5	3	1	4	**21**
7	Garissa	24	11	5	1	6	**41**
8	Wajir	24	11	5	1	6	**41**
9	Mandera	24	11	5	1	6	**41**
10	Marsabit	11	6	3	1	4	**21**
11	Isiolo	11	6	3	1	4	**21**
12	Meru	46	23	11	2	13	**82**
13	Tharaka-Nithi	11	6	3	1	4	**21**
14	Embu	11	6	3	1	4	**21**
15	Kitui	35	17	8	2	10	**62**
16	Machakos	35	17	8	2	10	**62**
17	Makueni	24	11	5	1	6	**41**
18	Nyandarua	24	11	5	1	6	**41**
19	Nyeri	24	11	5	1	6	**41**
20	Kirinyaga	24	11	5	1	6	**41**
21	Murang'a	35	17	8	2	10	**62**
22	Kiambu	71	33	16	4	20	**124**
23	Turkana	24	11	5	1	6	**41**
24	Westpokot	24	11	5	1	6	**41**
25	Samburu	12	5	3	1	4	**21**
26	Transnzoia	24	11	5	1	6	**41**
27	Uasingishu	35	17	8	2	10	**62**
28	Elgeyo/Marakwet	11	6	3	1	4	**21**
29	Nandi	24	11	5	1	6	**41**
30	Baringo	24	11	5	1	6	**41**
31	Laikipia	12	5	3	1	4	**21**
32	Nakuru	59	28	13	3	16	**103**
33	Narok	36	16	8	1	9	**61**
34	Kajiado	35	17	8	2	10	**62**
35	Kericho	24	11	5	1	6	**41**
36	Bomet	24	11	5	1	6	**41**
37	Kakamega	59	28	13	3	16	**103**
38	Vihiga	12	5	3	1	4	**21**
39	Bungoma	47	22	11	2	13	**82**
40	Busia	24	11	5	1	6	**41**
41	Siaya	24	11	5	1	6	**41**
42	Kisumu	36	16	8	2	10	**62**
43	Homabay	36	16	8	2	10	**62**
44	Migori	36	16	8	1	9	**61**
45	Kisii	36	16	8	2	10	**62**
46	Nyamira	12	5	3	1	4	**21**
47	Nairobicity	130	60	30	7	37	**227**
Total		1,396	650	314	71	385	**2,431**

Identification of the Target Households

We used systematic random walk approach and observed the left-hand rule, to identify our target households at the EA level. The skipping interval for the random walk per EA was computed by the team leaders, based on the relation:

k = N/n

Where: **N** = estimated total number of households in the EA as provided by the relevant local administration officials
n = the targeted number of households per EA (i.e., 20, as indicated above)
k = the sampling/skipping interval

Household substitution

We allowed for household substitution in several scenarios, including when:
The target household respondent refused to be interviewed either at the beginning or mid-ways through the interview

i. The target household respondent was not reachable even after two call backs
ii. The household contact person barred the enumerators from reaching the targeted household respondent
iii. There was no one in the household even after two call backs
iv. The household became hostile to the enumerators
v. The target household respondent was too sick/indisposed to effectively participate in the survey
vi. There was language barrier between the household respondent and the enumerators
vii. The survey team's visit to the household coincided with other household-level ceremonies or functions that are generally considered incompatible with interviews, e.g., burial events, death of a household member, etc.

Because of the proportionate sample allocation by age at the EA level, the survey target for persons aged 55 + years was only 13%. Due to the low percentage share for the group, we opted to mitigate for any likely non-response in the group, by providing for a booster sample of 71 cases nationally. This raised the overall age group target from 314 to 385 respondents. As with the EAs above, the booster sample was proportionately distributed across the 47 PSUs, by rural/urban split. The booster cases were identified purposively at the EA level, depending on whether there was anyone aged 55+ years in the household or not. The results from the booster sample were to provide further insights about the survey respondents aged 55 years and above, especially in cases where the relevant numbers for the randomly identified cases in the group were too low to provide any generalizable information about the group. Because of the purposive selection, data from the booster sample were not used to generalize on any survey parameter for those aged 55 years and above, but rather to give the perspectives of those who were interviewed only. For generalizable results, the respondents should be selected randomly.

Identification of the Target Household Respondents

We used the last birthday[2] approach to randomly identify the target household respondents in the survey. However, for the booster sample, we used age as the main criteria for selection. We did not undertake household respondent substitution in the survey.

Survey Output

Overall, we had a 101% achievement in the survey.

Of the 2,360 cases targeted under random selection, we successfully interviewed 2,377 respondents. We achieved a total of 79 out of the 71 cases that were to be selected purposively. The overall field achievement was 2,456 out of the targeted 2,431.

In terms of age-bands, we surpassed the targets for the 15-34 years (109%) and 35-54 years (113%) categories. However, for the 55+ years category, we only managed a 39% achievement for the randomly selected cases, and 111% for the purposively selected cases. The overall output for the 55 years and above age-band was 52%.

Table 2 below provides a summary of the survey achievement.

2 Last birthday approach: The target household-level respondent is the household member whose last birthday is closest to the date of visit to the household by the survey team. In case of twins, etc. the member born/delivered last would be the desired respondent.

Table 2: Survey output by age-band and respondent selection approach, per county

COUNTY		Age (in years)										
		15-34		35-54		55+				Total		Achievement Percentage
		Random						Purposive				
		Targeted	Achieved	Targeted	Achieved	Targeted	Achieved	Targeted	Achieved	Targeted	Achieved	
1	Mombasa	35	39	17	20	8	1	2	3	**62**	**63**	**102%**
2	Kwale	24	26	11	14	5	1	1	2	**41**	**43**	**105%**
3	Kilifi	47	53	22	23	11	4	2	4	**82**	**84**	**102%**
4	Tanariver	12	14	5	9	3	0	1	1	**21**	**24**	**114%**
5	Lamu	12	11	5	6	3	2	0	2	**20**	**21**	**105%**
6	Taita/Taveta	12	13	5	6	3	1	1	1	**21**	**21**	**100%**
7	Garissa	24	16	11	19	5	5	1	1	**41**	**41**	**100%**
8	Wajir	24	11	11	28	5	5	1	1	**41**	**45**	**110%**
9	Mandera	24	15	11	24	5	3	1	1	**41**	**43**	**105%**
10	Marsabit	11	12	6	8	3	0	1	1	**21**	**21**	**100%**
11	Isiolo	11	12	6	5	3	3	1	1	**21**	**21**	**100%**
12	Meru	46	55	23	21	11	5	2	2	**82**	**83**	**101%**
13	Tharaka-Nithi	11	13	6	7	3	0	1	1	**21**	**21**	**100%**
14	Embu	11	18	6	2	3	0	1	1	**21**	**21**	**100%**
15	Kitui	35	37	17	24	8	0	2	2	**62**	**63**	**102%**
16	Machakos	35	31	17	25	8	4	2	3	**62**	**63**	**102%**
17	Makueni	24	29	11	11	5	0	1	1	**41**	**41**	**100%**
18	Nyandarua	24	21	11	12	5	7	1	1	**41**	**41**	**100%**
19	Nyeri	24	23	11	10	5	7	1	1	**41**	**41**	**100%**
20	Kirinyaga	24	27	11	13	5	0	1	1	**41**	**41**	**100%**
21	Murang'a	35	28	17	24	8	8	2	2	**62**	**62**	**100%**
22	Kiambu	71	75	33	33	16	13	4	4	**124**	**125**	**101%**
23	Turkana	24	30	11	7	5	3	1	1	**41**	**41**	**100%**
24	Westpokot	24	28	11	12	5	0	1	1	**41**	**41**	**100%**
25	Samburu	12	14	5	4	3	2	1	0	**21**	**20**	**95%**
26	Transnzoia	24	24	11	14	5	2	1	1	**41**	**41**	**100%**
27	Uasingishu	35	48	17	11	8	1	2	2	**62**	**62**	**100%**
28	Elgeyo/Marakwet	11	15	6	4	3	1	1	1	**21**	**21**	**100%**
29	Nandi	24	29	11	10	5	1	1	1	**41**	**41**	**100%**
30	Baringo	24	25	11	13	5	3	1	1	**41**	**42**	**102%**
31	Laikipia	12	12	5	7	3	1	1	1	**21**	**21**	**100%**
32	Nakuru	59	61	28	40	13	0	3	3	**103**	**104**	**101%**
33	Narok	36	33	16	25	8	2	1	2	**61**	**62**	**102%**
34	Kajiado	35	44	17	14	8	2	2	2	**62**	**62**	**100%**
35	Kericho	24	32	11	7	5	1	1	1	**41**	**41**	**100%**
36	Bomet	24	29	11	8	5	**2**	1	2	**41**	**41**	**100%**
37	Kakamega	59	72	28	25	13	**4**	3	3	**103**	**104**	**101%**
38	Vihiga	12	11	5	9	3	**0**	1	1	**21**	**21**	**100%**
39	Bungoma	47	51	22	29	11	**0**	2	2	**82**	**82**	**100%**
40	Busia	24	32	11	6	5	**2**	1	1	**41**	**41**	**100%**
41	Siaya	24	24	11	11	5	**5**	1	1	**41**	**41**	**100%**
42	Kisumu	36	41	16	10	8	**9**	2	2	**62**	**62**	**100%**
43	Homabay	36	44	16	15	8	**1**	2	2	**62**	**62**	**100%**
44	Migori	36	40	16	21	8	**0**	1	1	**61**	**62**	**102%**
45	Kisii	36	38	16	18	8	**3**	2	2	**62**	**61**	**98%**
46	Nyamira	12	17	5	2	3	**2**	1	1	**21**	**22**	**105%**
47	Nairobicity	130	151	60	65	30	**6**	7	7	**227**	**229**	**101%**
Total		1,396	1,524	650	731	314	**122**	71	79	**2,431**	**2,456**	**101%**
Achievement percentage		109%		113%		39%		111%		**101%**		

Data Weighting

We weighted the survey data by age, gender, and location at the County level, and only for the randomly selected cases.

To compute the output weight for a given case, we employed a two-step protocol where we first calculated the probability of selection for the case in the survey, and then computed the applicable weight. The procedure for the two-step approach was as follows:

i. **Determining the probability of selection:**

Probability of selection (p_c) by a given variable (e.g., age)
=
output value (vo) / total target population (N) by the given variable,

i.e.,
$pc = v_o/N$

For multiple variables, e.g., age, gender and location,

the probability of selection by multiple variables, p_m
=
product of the probability of selection by each of the variables
=
(probability of selection by age, p_a) x
(probability of selection by gender, p_g) x
(probability of selection by location, p_l)

i.e.,
$p_m = p_a \times p_g \times p_l$

ii. **Determining the composite survey weight**
We obtained the survey weight (ws) of a given case by age, gender and location, by taking the inverse of the applicable pm above, i.e.,

$ws = 1/pm$

2. HUMAN CENTRED DESIGN (HCD) RESEARCH METHODOLOGY

Human centred design (HCD) research uses a combination of design and qualitative research methods with an aim to deeply understand users: their underlying needs, motivations, behaviours and aspirations.

Given the unprecedented COVID-19 pandemic, we leveraged our experience with remote HCD research methods. In the past, we have supplemented in-person research with remote methods when travel was not feasible, or the project context and budget did not allow for in-person research. Over the past few months, we have codified and refined toolkits for remote research and design and iterated our approach over multiple projects. Remote human-centred design (HCD) research is in many ways similar to standard HCD. It relies on the same foundational philosophy of human centricity, is highly participatory and oriented towards co-creation, and is deeply interpersonal. The biggest difference is the medium of interaction.

For those we could not interview physically by going to their communities, homes, and businesses, we joined them remotely, through audio and visual connections. We created a remote HCD toolkit to enable teams to conduct research and design without being physically present with research participants. Our toolkit covers all HCD methods that are applied in a project (conducting research, testing concepts and prototypes, synthesizing outcomes, co-creating with teams). It draws on best practices from in-person HCD research, while recognizing the technical and behavioural constraints and opportunities that come with

For this engagement, we leveraged HCD research to inform the survey questionnaire – before the survey started, to refine the survey question framings, and its answer options. HCD research also helped uncover a few new research areas not part of the survey questionnaire and helped frame additional survey questions. We further conducted a second round of HCD interviews after the first phase of survey analysis was completed to strengthen and deepen the survey insights by providing the “why” and “how” behind it at a smaller scale. We used HCD research to identify and fill in gaps in interpreting some of the survey data, such as highlighting the nuances around people’s sentiment towards digital usage.

This study by no means claims to be exhaustive in capturing all of the people’s challenges and experiences with digital services in Kenya. However, it did strengthen and deepen our survey-based understanding of the ‘State of the Digital Economy’, as experienced by its everyday users. These efforts enabled us to better understand the human stories behind the data and uncover further areas for research.

We interviewed 27 pre-recruited individuals across the 3 counties including Nairobi, Makueni and Kiambu and from diverse geographic and demographic backgrounds as well as a variety of digital usage levels and experiences. While the sample size in our study is small by design, its strength lies in tracing common themes and patterns across a diverse range of digital users. were able to listen deeply to the everyday experiences of these diverse people with digital usage.

3. ANALYSIS

We followed a hypothesis-led, statistically rigorous, and reproducible approach towards analysing the data. Initial hypotheses - based on expert interviews and secondary research - were iterated upon, refined, and tested. While doing so, we strived to achieve rigor and transparency, performing comprehensive quality checks, using commonly accepted statistical tools, and publishing our assumptions and business rules in an open source manner.

Segmenting users of digital services

In an effort to build a deeper and nuanced understanding of the digital divide, we have used a PAM cluster algorithm to find four distinct groups of Kenyans and organised them based on what digital services they currently use. We have found one group of **non-users** who don't use any of the digital service use cases studied, and three groups of digital services users that include different clusters of use cases. Users of **basic** digital services use only basic mobile money services including sending and receiving payments or mobile phone airtime top-ups. Users of **intermediate** digital services use one or more of services including social media, digital communication tools, digital entertainment, and digital financial services using mobile money. Users of **advanced** digital services use one or more services that support knowledge building, social services, and livelihoods—e.g., information / news, e-commerce, e-governance, digital education, digital health, business support, upskilling / job search, or agricultural support.

While we have defined the segments as users of basic, intermediate, or advanced services, we do not mean to suggest that this is the only way to cluster usage patterns across the broader digital economy, or that there is a continuum of use across these services. Our segmentation is rather a reflection of the observation that the use of more advanced digital services may require greater resources and/or know-how—and therefore that people from more vulnerable socio-economic demographics may face more and deeper structural challenges to the use of these services. Understanding differences and similarities in how these segments perceive and engage with a range of digital services can help us distil the specific challenges users face and identify what it will take to bridge the digital divide.

www.digitaleconomy.ke

SUPPORTED BY

OMIDYAR NETWORK™