

DESCRIPTIVE ANALYSIS PLAN

Kenya Household Survey: A Socioeconomic & Wellbeing Study (2022-2026)

1. Purpose

This plan outlines the descriptive analysis to be conducted on the cleaned Kenya Household Survey dataset. The survey covers 9420 households within 12 counties in Kenya, cutting across urban, peri-urban, and rural settings. All analysis is disaggregated by residence type and county. Gender breakdowns are included where relevant.

The findings are intended to inform programme targeting and resource allocation decisions for NGOs, county governments, and donors.

The central question driving the analysis is: **Which households are most vulnerable, and what do they have in common?**

2. Analytical Framework

The analysis is structured into five thematic modules. Each module produces summary statistics, cross-tabulations, and visualizations for Power BI.

Module 1: Demographic Profile

- Respondent count and percentage share by county
- Age distribution: mean, median, max, min by residence type, i.e (Rural, Urban and Peri-urban)
- Age group breakdown (15–24, 25–34, 35–49, 50–64, 65+) by county
- Gender composition by residence type and county
- Average household size by county, ranked largest to smallest
- Education level distribution by gender and residence type

Module 2: Economic Vulnerability

- Mean and median monthly income by residence type (urban vs rural income gap)
- Share of households in the bottom two income quintiles by county
- Top 5 livelihood sources in urban vs rural areas
- Employment status breakdown by county
- Remittance dependency rate by county
- Income by education level, disaggregated by residence type
- Dependency ratio distribution i.e. how does the ratio relate to income and food security?

Module 3: Health Access

- Health insurance coverage rate by residence type (% insured vs uninsured)
- Insurance type breakdown by county
- Average distance to nearest health facility by county, ranked worst to best
- % of households living more than 10km from a health facility by residence type
- % using unimproved water sources by county
- % practicing open defecation by residence type

Module 4: Food Security

- Food security status distribution by residence type
- Top 3 counties by rate of severe food insecurity
- Food security by income quintile (does severity worsen with lower income?)
- Livelihood sources most associated with food insecurity
- Relationship between household size and food security category
- Double deprivation e.g. severely food insecure households also using unimproved water

Module 5: Financial Inclusion

- Mobile money access rate by residence type and dominant platform
- Penetration by age group (is there a generational gap?)
- Gender gap in mobile money access
- Access by employment status (are casual workers less financially included?)
- Financially excluded core (no mobile money AND no health insurance, by residence type)

3. Cross-Cutting: Vulnerability Index

To bring all five modules together, each household is assigned a score from 0 to 5 based on the number of disadvantages they face at the same time. One point is assigned for each of the following:

- Income in the bottom quintile (Q1)
- Moderately or severely food insecure
- No health insurance of any kind
- No mobile money access
- Reliance on an unimproved water source

Households scoring 4 or 5 are flagged as the **highest-need group**. The index is mapped by county and residence type to show where multiple forms of vulnerability overlap and where support is most needed.

4. Deliverables

	Task	Duration	Output
1	Data analysis Plan	1 day	Analysis Plan
2	Pre-cleaning Audit findings	2 days	DQA Report
3	Data Cleaning (PostgreSQL)- Schema, views and cleaning steps	3 days	Cleaned dataset + cleaning log
4	PostgreSQL schema setup & data load	1 day	Database + views
5	Descriptive analysis (5 modules+ vulnerability index)	4 days	Analysis tables + charts
6	Report writing & visualisation	3 days	Final report + dashboard
7	Project overview, variable dictionary, sampling note, reproduction steps	2	README
Total		15 working days	Full project portfolio

5. Limitations

- This is a synthetic dataset and does not reflect real household conditions, county demographics, or survey findings in Kenya.
- The dataset is designed to be cross-sectional and does not support causal inference or trend analysis.
- Income data is self-reported and subject to recall and social desirability bias
- Food security classification is based on proxy indicators, not a validated HFIAS administration.