

Kenya - Analysis of Supermarket Grocery Data for Prediction of Nutritional and Health Outcomes at the Population Level - Supermarket F

Agnes Kiragga

Report generated on: July 16, 2025

Visit our data catalog at: <https://microdataportal.aphrc.org/index.php>

Overview

Identification

ID NUMBER

DDI-KEN-APHRC-SUPERMARKET-F-2023-V1.0

Version

VERSION DESCRIPTION

PRODUCTION DATE

2025-07-08

NOTES

N/A

Overview

ABSTRACT

Rates of overweight, obesity, and chronic diseases such as cardiovascular diseases, hypertension, type 2 diabetes and certain cancers (bowel, lung, prostate and uterine) are on the rise in most sub-saharan Africa (SSA) countries like Kenya. These increases can be largely attributed to the shift toward unhealthy diet patterns and increased access to processed foods that are high in fat, sugar, and sodium. The influx of supermarkets in East Africa and the replacement of traditional foods for processed foods places this region in a vulnerable position for greater increases in chronic disease rates. Consumer purchasing history from supermarkets can provide valuable insight to food intake over time and the present and future effects on chronic diseases. Purchasing data from supermarkets is available yet underutilized in SSA.

The study aimed to harmonize and increase accessibility to grocery data, use statistical methods to explore purchasing patterns and predict the effects of nutrition on chronic diseases, and inform policy on the various influences on consumer purchases.

UNITS OF ANALYSIS

Individuals and supermarket transaction records.

Scope

NOTES

- Transaction Level Data: Food Item Details (Specific products purchased), Quantity, Price, Date of Purchase, Location of Purchase, Customer Demographics (age, gender when collected via loyalty programs), Payment Method(Cash, credit card, digital payment, etc.), Basket Composition

KEYWORDS

Supermarkets, Ultra-Processed Foods, Processed Foods, Non-communicable Chronic Diseases

Coverage

GEOGRAPHIC COVERAGE

County coverage: Nairobi and Kiambu.

UNIVERSE

The survey covers transaction records of individuals who made purchases in supermarkets.

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Agnes Kiragga	African Population and Health Research Centre (APHRC)

OTHER PRODUCER(S)

Name	Affiliation	Role
Steve Cygu	African Population and Health Research Center (APHRC)	Co-Investigator - Study coordination and Co-lead
Maureen Ng'etich	African Population and Health Research Center (APHRC)	Co-Investigator - Study coordination and Co-lead
Lindsey English	African Population and Health Research Center (APHRC)	Co-Investigator - Supporting methods of data mapping, analysis, and nutrition policy
Reinpeter Momanyi	African Population and Health Research Center (APHRC)	Co-Investigator - Supporting methods of data mapping, analysis and documentation of study
Elizabeth Kimani	African Population and Health Research Center (APHRC)	Co-Investigator - Support methods of data analysis and co-lead in policy analysis
Gershim Asiki	African Population and Health Research Center (APHRC)	Co-Investigator - Support methods of data analysis and co-lead in policy analysis

FUNDING

Name	Abbreviation	Role
African Populattion and Health Research Center	APHRC	Funder (Big Idea)

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Bonface Ingumba	African Population and Health Research Center (APHRC)	Data Governance Officer
Mambe Shem	African Population and Health Research Center (APHRC)	Data Documentation Officer

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
African Population and Health Research Center	APHRC		Documentation of the DDI

DATE OF METADATA PRODUCTION

2025-07-08

DDI DOCUMENT VERSION

Version 1.0 (July 2025)

DDI DOCUMENT ID

DDI-KEN-APHRC-SUPERMARKET-F-2023-V1.0

Sampling

Sampling Procedure

The study is a cross-sectional exploratory study with a phased approach employing quantitative secondary data collection from a third-party information management solution provider. The third party provider employs an open integrated point of sale and store information retail system that connects retail touch points and sales channels in several counties in Kenya.

Sampling was conducted after a census of all supermarkets subscribed to the third party system was done. Only those counties with supermarkets subscribed to the platform were sampled. A sample of large, medium sized and small supermarkets were selected to participate in the study. The supermarket sizes were determined as follows; large supermarkets (supermarkets with a cumulative total of more than 8 branch networks). Medium size supermarkets will be those with 3-8 branch networks in the counties and smaller supermarkets are those with 1-2 branch networks.

Grocery data was received from a supermarket chain with 3 branches.

Deviations from Sample Design

Not Applicable

Response Rate

Not Applicable

Weighting

Not Applicable

Questionnaires

Overview

A standardized form was developed to guide in extraction of information from 3rd party information provider for supermarket purchase data. Variables of interest includes supermarket name, supermarket branch, location of supermarket, invoice id, customer id, customer demographics (gender, age), date and time of purchase, product name purchased, unit price per item, number of items purchased, payment method used by customer for purchase etc.

Secondary data collected will not be identifiable as it will be anonymized at the supermarket and client level.

The standardized form is provided as external resources data.
V1-V18 the questions are found in the "Study abstraction tool"

Data Collection

Data Collection Dates

Start	End	Cycle
2022-06-09	2023-12-31	Supermarket F

Data Collection Mode

Other [oth]

Questionnaires

A standardized form was developed to guide in extration of information from 3rd party information provider for supermarket purchase data. Variables of interest includes supermarket name, supermarket branch, location of supermarket, invoice id, customer id, customer demographics (gender, age), date and time of purchase, product name purchased, unit price per item, number of items purchased, payment method used by customer for purchase etc.

Secondary data collected will not be identifiable as it will be anonymized at the supermarket and client level.

The standardized form is provided as external resources data.
V1-V18 the questions are found in the "Study abstraction tool"

Supervision

Not Applicable

Data Processing

Data Editing

Not Applicable

Other Processing

The extracted grocery data was in the form of csv files and was saved into a local database using PostgreSQL version 15.2 and imported into R version 4.3.3 for cleaning and pre-processing.

Data preprocessing techniques applied included: Transactions and demographics alignment, dealing with missing values, checking for data consistency, quality assurance checks and filtering non-food items.

Data Appraisal

Estimates of Sampling Error

Not Applicable

File Description

Variable List

clean_supermarket_f_sample

Content

Cases 279307

Variable(s) 18

Structure Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V1	id	Shopper id	discrete	character	Shopper id
V2	county	County acronym	discrete	character	County acronym
V3	gender	Gender	discrete	character	Gender
V4	description	Product name description	discrete	character	Product name description
V5	price	Unit price of product	contin	numeric	Unit price of product
V6	quantity	Quantity of product purchased	contin	numeric	Quantity of product purchased
V7	total	Sales invoice total price	contin	numeric	Sales invoice total price
V8	trnref	Transaction id	contin	numeric	Transaction id
V9	sdatetime	Transaction date	contin	numeric	Transaction date
V10	paymentmode	Payment mode	discrete	character	Payment mode
V11	branch	Branch id of supermarket	discrete	character	Branch id of supermarket
V12	transaction_id	Supermarket branch transaction id	discrete	character	Supermarket branch transaction id
V13	dob_new	Date of birth of shopper	discrete	character	Date of birth of shopper
V14	supermarket_name	Supermarket name	discrete	character	Supermarket name
V15	branch_name	Branch name of supermarket	discrete	character	Branch name of supermarket
V16	county_name	County	discrete	character	County
V17	sub_county_name	Sub-county	discrete	character	Sub-county
V18	age	Age of shopper	contin	numeric	Age of shopper

Shopper id (id)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 80171
Format: character	Invalid: 0
Width: 5	

Description

This question seeks to get generated ID for the shopper

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Shopper id

Post question

N/A

Interviewer instructions

N/A

County acronym (county)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 279307
Format: character	Invalid: 0
Width: 7	

Description

This question seeks to get generated County acronym

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

County acronym

Post question

N/A

Interviewer instructions

N/A

Gender (gender)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 80171
Format: character	Invalid: 0
Width: 6	

Description

This question seeks to get the Gender of the shopper

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Gender

Post question

N/A

Interviewer instructions

N/A

Product name description (description)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Format: character

Width: 50

Valid cases: 279307

Invalid: 0

Description

This question seeks to get the Product name description of the item the shopper bought.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Product name description

Post question

N/A

Interviewer instructions

N/A

Unit price of product (price)

File: clean_supermarket_f_sample

Overview

Type: Continuous

Format: numeric

Width: 10

Decimals: 0

Range: 1-4755

Valid cases: 279307

Invalid: 0

Minimum: 1

Maximum: 4755

Mean: 115.6

Standard deviation: 133

Description

This question seeks to get the Unit price of product the shopper bought.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Unit price of product

Post question

N/A

Interviewer instructions

N/A

Quantity of product purchased (quantity)

File: clean_supermarket_f_sample

Overview

Type: Continuous

Format: numeric

Width: 10

Decimals: 0

Range: -100-182

Valid cases: 279307

Invalid: 0

Minimum: -100

Maximum: 182

Mean: 1.7

Standard deviation: 3.1

Description

This question seeks to get the Quantity of product purchased by the shopper .

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Quantity of product purchased

Post question

N/A

Interviewer instructions

N/A

Sales invoice total price (total)

File: clean_supermarket_f_sample

Overview

Type: Continuous

Format: numeric

Width: 10

Decimals: 0

Range: -3559-66813

Valid cases: 279307

Invalid: 0

Minimum: -3559

Maximum: 66813

Mean: 834.9

Standard deviation: 1249.2

Description

This question seeks to get the Sales invoice total price.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Sales invoice total price

Post question

N/A

Interviewer instructions

N/A

Transaction id (trnref)

File: clean_supermarket_f_sample

Overview

Type: Continuous	Valid cases: 279307
Format: numeric	Invalid: 0
Width: 12	Minimum: 3
Decimals: 0	Maximum: 1219629
Range: 3-1219629	Mean: 494411.9
	Standard deviation: 313806.3

Description

This question seeks to get the Transaction id.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Transaction id

Post question

N/A

Interviewer instructions

N/A

Transaction date (sdatetime)

File: clean_supermarket_f_sample

Overview

Type: Continuous	Valid cases: 279307
Format: numeric	Invalid: 0
Width: 11	Minimum: 1970375202000
Decimals: 0	Maximum: 2019683232000
Range: 1970375202000-2019683232000	Mean: 1996190306040.5
	Standard deviation: 13975172732.4

Description

This question seeks to get the Transaction date.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Transaction date

Post question

N/A

Interviewer instructions

N/A

Payment mode (paymentmode)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Format: character

Width: 18

Valid cases: 279307

Invalid: 0

Description

This question seeks to get the Payment mode.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Payment mode

Post question

N/A

Interviewer instructions

N/A

Branch id of supermarket (branch)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Format: character

Width: 7

Valid cases: 279307

Invalid: 0

Description

This question seeks to get the generated Branch id of the supermarket.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Branch id of supermarket

Post question

N/A

Interviewer instructions

N/A

Supermarket branch transaction id (transaction_id)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 279307
Format: character	Invalid: 0
Width: 15	

Description

This question seeks to get the generated Supermarket branch transaction id.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Supermarket branch transaction id

Post question

N/A

Interviewer instructions

N/A

Date of birth of shopper (dob_new)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 9048
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

Description

This question seeks to get the Date of birth of the shopper.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Date of birth of shopper

Post question

N/A

Interviewer instructions

N/A

Supermarket name (supermarket_name)

File: clean_supermarket_f_sample

Overview

Type: Discrete	Valid cases: 279307
Format: character	Invalid: 0
Width: 1	

Description

This question seeks to get the Supermarket name.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Supermarket name

Post question

N/A

Interviewer instructions

N/A

Branch name of supermarket (branch_name)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Valid cases: 279307

Format: character

Invalid: 0

Width: 2

Description

This question seeks to get the Branch name of supermarket

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Branch name of supermarket

Post question

N/A

Interviewer instructions

N/A

County (county_name)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Valid cases: 279307

Format: character

Invalid: 0

Width: 7

Description

This question seeks to get the County where the supermarket is located.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

County

Post question

N/A

Interviewer instructions

N/A

Sub-county (sub_county_name)

File: clean_supermarket_f_sample

Overview

Type: Discrete

Valid cases: 279307

Format: character

Invalid: 0

Width: 10

Description

This question seeks to get the Sub-county where the supermarket is located.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Sub-county

Post question

N/A

Interviewer instructions

N/A

Age of shopper (age)

File: clean_supermarket_f_sample

Overview

Type: Continuous

Valid cases: 9048

Format: numeric

Invalid: 270259

Width: 10

Minimum: 18

Decimals: 0

Maximum: 71.9

Range: 18-71.9

Mean: 19.9

Standard deviation: 6.5

Description

This question seeks to get the Age of shopper.

Universe

Individuals and supermarket transaction records.

Source of information

Individuals and supermarket transaction records.

Pre question

N/A

Literal question

Age of shopper

Post question

N/A

Interviewer instructions

N/A

Documentation

Questionnaires

Study abstraction tool.pdf

Title Study abstraction tool.pdf
Author(s) African Population and Health Research Center
Date 16/07/2025
Country KENYA
Language ENGLISH
Contributor(s) Dr. Agnes Kiragga
Publisher(s) APHRC
Filename Study abstraction tool.pdf

Technical documents

Metadata of supermarkets.pdf

Title Metadata of supermarkets.pdf
Author(s) African Population and Health Research Center
Date 16/07/2025
Country KENYA
Language ENGLISH
Contributor(s) Dr. Agnes Kiragga
Publisher(s) APHRC
Filename Metadata of supermarkets.pdf
