

STUDY ABSTRACTION TOOL

This study titled “Analysis of supermarket grocery data for prediction of nutritional and health outcomes at the population level” aims to examine the purchasing patterns of consumers to understand their effects on the prevalence of non-communicable diseases (NCDs). It is crucial for informing policy interventions to promote health in Kenya, particularly in the context of rising NCDs and changes in dietary habits across age groups, sex, type of settlement (rural/urban) and location (counties).

The study is cross-sectional employing quantitative secondary data collection of purchase records of individuals in supermarkets across 11 counties in Kenya; Nairobi, Kiambu, Nakuru, Kajiado, Machakos, Nyeri, Embu, Murang'a, Kirinyaga, Meru and Garissa counties.

The data will be extracted using a standardized form (see table below for detailed variable information)

| Variable name | Description |
|-----------------------|---|
| Supermarket ID | Anonymized generated ID based on supermarket name |
| Supermarket branch ID | Anonymized generated ID based on supermarket sales location |
| County | Location of supermarket based on branches and town |
| Sub-county | Location of supermarket based on branches and town |
| Invoice ID | Computer generated Payment slip identification number |
| Customer ID | Anonymized individual customer identification number created with unique identifiers ¹ |
| Type of customer | Customer type recorded as loyalty card customers and non-loyalty customers |
| Age | Recorded as date of birth for customers ² |
| Gender | Gender of customer ³ |
| Item | Product name |
| Category | General item categorization groups – milk, fats & oils, sugar, beverages, health and beauty, home and lifestyle, sports and travel etc |
| Unit price | The unit price per item |
| Total | Total price including tax |
| Quantity | Number of items purchased by customer |
| Date | Date and Time of purchase |
| Transaction type | Payment method used by customer for purchase. Methods available – Cash (includes vouchers, redeeming of loyalty points), Credit card, Corporate and Mpesa |
| Recipe | Ingredients constituting an instore produced/deli item |
| Supplier | Vendor name |
| Barcode | Universal product lookup code |

Note

1. For loyalty card customers, a customer ID is created from the customer's name, national ID, phone number, gender, and date of birth, while non-loyalty customers are created from phone number, customer name, and gender (extracted from the names and phone number in M-Pesa payment transactions). Anonymization will be done at the supermarket and customer level before data is shared with APHRC.
2. Only customers with loyalty cards will have a date of birth recorded. Non-loyalty customers will have a blank date of birth field.
3. Only customers with loyalty cards will have their gender recorded. Missing gender records for loyalty and non-loyalty customers will be extracted from the names in M-Pesa payment transactions.