Original Research Article

Analyzing Demographic Grocery Purchase Patterns in Kenyan Supermarkets Through Unsupervised Learning Techniques

INQUIRY: The Journal of Health Care Organization, Provision, and Financing
Volume 62: 1–17

DOI: 10.1177/00469580251319905
journals.sagepub.com/home/inq
Sage

Reinpeter Momanyi, MSc[1], Steve Bicko Cygu, PhD[1], Agnes Kiragga, PhD[1,2], Henry Owoko Odero, BA[1], Maureen Ng'etich, MSc[1], Gershim Asiki, PhD[1], and Tatenda Duncan Kavu, PhD[1]

Abstract

Kenya is experiencing a significant increase in the prevalence of non-communicable diseases (NCDs) such as cardiovascular diseases, hypertension, Type 2 diabetes, and certain cancers (bowel, lung, prostate, and uterine). This case is not unique to Kenya but is common in many Low and Middle-Income Countries (LMICs) in Africa. Many NCDs, are linked to diets high in added sugars, sodium, saturated fat, and low in fiber. There is a notable lack of information regarding the demographic differences among supermarket customers and their purchasing habits of healthy versus unhealthy foods in some parts of Africa. This gap in knowledge hinders the ability to connect grocery purchase patterns to NCDs, including obesity. Supermarkets in LMICs offer valuable demographic insights through grocery data. This research utilizes NOVA classification tool, data mining and unsupervised machine learning techniques to analyze grocery purchase patterns in 10 supermarkets across 5 counties in Kenya between 2022 and 2023. The apriori algorithm was used to create association rules and an analysis was done on the association rules to find out the relationship between demography (location, gender, and age) with purchase patterns. Individual data was collected along with transaction data, since the supermarkets logged transactions done by loyalty card customers. The main aim is to provide guidance to policymakers in public health. We collected 3 934 122 unique transactions and each transaction was associated with a customer who was identified with a unique customer ID. Findings from this research demonstrate that 53% of food purchases from these transactions were mainly industrially processed food items and males above the age of 50 years were the main consumers of these food items. The findings lead to the conclusion that this purchase trend has a chance of rising NCDs in older people. Therefore we recommend that policymakers adopt our recommendations to safeguard public health.

Keywords

low and middle-income countries (LMICs), association rule mining, NOVA, non-communicable diseases (NCDs), demography

Why was this study done?

- Low- and middle-income countries (LMICs) in Africa are experiencing increasing cases of non-communicable diseases (NCDs) such as diabetes, cardiovascular conditions, and certain cancers.
- NCDs are linked to overweight and obesity associated with high intake of added sugars, sodium, saturated fats, and low fiber.
- Grocery transaction data can provide insights into consumer demographics and health trends.

What did this study find?

- This study utilized supermarket grocery data to analyze customer purchase patterns using NOVA food classification and unsupervised machine learning.
- Results revealed that 53% of food purchases were industrially processed foods with men over 50 as the main consumers.
- These food purchase trends may increase NCD risks in older adults.

What do these findings mean?

- The results indicate that policymakers need to take action to promote healthier food consumption.
- The study was limited to shoppers in a tier 2 supermarket drawn from 5 counties out of the 47 counties in Kenya, therefore the data might not represent other regions in Kenya. It did not consider alternative traditional food purchasing outlets such as kiosks, open-air markets, and street vendors.

Introduction

Non-communicable diseases (NCDs) which include nutrition related non-communicable diseases (NR-NCDs) account for 74% of deaths globally. Of these deaths, 77% are in low-and-middle income countries.[1] Nearly all NCDs, can be linked with consumption of diets high in added sugars, sodium, saturated fat, and low in fiber.[2-4] These associations have been observed on both an individual and population level. The evidence of the global nutrition transition toward this diet pattern often referred to as the Western Diet has shown population level effects on NCDs prevalence and mortality.[5,6] East Africa has not been immune to the effects of global nutrition transition. In both urban and rural regions, overweight and obesity prevalence has increased dramatically along with the increases in NCDs. Overall in 2022, overweight (Body Mass Index(BMI) between 25.0 and 29.9) and obesity (BMI greater than or equal to 30.0) prevalence of age 20 to 49 in Kenya was 19% among men and 45% among women while diabetes prevalence of age 20 to 49 in Kenya was 1% among men and 1% among women[7]; these numbers have undoubtedly increased since then.

To investigate the increasing over-reliance on industrially processed products, the NOVA food classification system[8] has been prominently used to provide guidance and information on processed foods. NOVA has been applied worldwide in scientific literature. It has been used in Brazil, Canada, Chile, Mexico, Norway, Sweden, Spain, UK, and USA to assess the association, determinants and impact of consumption of ultra-processed products.[9-21] Little is known about studies in Africa that have utilized the use of NOVA food classification.

Little information is known about demographic differences in supermarket purchasing habits of healthy and unhealthy foods in Africa. Available studies mostly refer to demographic differences in food consumption habits in Canada, Hungary, Lebanon, Norway, and USA have shown that men consume more unhealthy foods (carbonated beverages, fast foods) while women consume more healthy foods (fruits, vegetables).[22-26]

Use of innovative ways with big data is on the rise due to advances in software development, storage capacity, and computational power which have made it possible to analyze large datasets.[27] One such industry that utilizes big data is the retail industry, specifically supermarkets. Supermarkets on a daily basis collect sales data at the store level and grocery data at the household level through loyalty cards. Grocery data can provide an accurate measure of household or individual level dietary-related behaviors in a real-life setting.[28] Despite the rapid growth of supermarkets in LMIC's termed "supermarket revolution,"[29] the full potential of grocery data has been underutilized.

With most households in large towns in LMIC's relying heavily on supermarkets for their food variety due to lower prices, there has been a rapid change of diets toward more energy-dense foods and beverages that tend to be rich in unhealthy types of fat, refined starches, caloric sweeteners, and salt.[30] These diet changes contribute to surging rates of obesity, which are risk factors for NR-NCDs.[6]

Insights From Related Work

Toiba et al[31] investigated the relationship between the increased use of modern food retail outlets in Indonesia and the emergence of unhealthy dietary patterns. They found it quite interesting that modern retail food was expensive and not very healthy. Specifically, a one-percentage-point increase in the share of food expenditures at supermarkets was associated with a 0.15% to 0.17%-point decrease in the household's share of food expenditures on healthy food. This shows that individuals would buy expensive, unhealthy foods at the expense of relatively low-cost healthy food, what an irony. Notably, they found that food consumed away from home tends to be less healthy. This study in Indonesia also provided demographic insights relevant to the context of Kenya. Factors such as the location and age of consumers can influence where they purchase their food items and at what cost.

De Roos et al[32] conducted an investigation to quantitatively analyze expenditure on all fresh foods, fruits and vegetables, and fish in Scotland, since fresh foods are perceived to contribute more to health than processed foods. They used ANOVA to test for differences in demographic characteristics and food purchases across urban-rural classifications. They found that rural households bought more in kilogram terms of fresh foods and fruits and vegetables than urban households. This study was interesting to this particular study since we also wanted to find out the purchase patterns of people in rural counties in Kenya and compare with other countries.

Odunitan-Wayas et al[33] conducted a study as part of the larger STOP Obesity in the People of South Africa. They found out that high-income shoppers purchased more fruits and vegetables, while low- and middle-income shoppers spent more on snacks, sugar-sweetened beverages, grains,

[1]African Population and Health Research Center, Nairobi, Kenya
[2]Makerere University, Kampala, Uganda

Received 24 September 2024; revised 7 January 2025; revised manuscript accepted 27 January 2025

Corresponding Author:
Reinpeter Momanyi, African Population and Health Research Center, Manga Close, Off Kirawa Road, 2nd Floor, APHRC Campus, Nairobi P.O. Box 10787-00100, Kenya.
Emails: rmomanyi@aphrc.org; reinp.mom@gmail.com

and potatoes. They also found that high Socio-Economic Area (SEA) shoppers (with monthly income ≈US$857) purchased more fruits and vegetables, while low (monthly income ≈US$114) and middle-SEA shoppers (monthly income ≈US$373) spent more on snacks, sugar-sweetened beverages (SSB), grains, and potatoes. Another unfortunate situation was that fresh fruits and vegetables in high SEA supermarkets were of better quality compared to those in middle- and low-SEA supermarkets. Since this study was conducted in Africa, we also wanted to explore how it compares to Kenya. Specifically, we aimed to investigate what perceived high-income shoppers, identified based on location and age, tend to purchase in Kenya.

Another study was conducted by Koroknay et al[23] in Hungary. The survey was conducted among university students to evaluate gender differences in consumption habits and spending behavior regarding various food groups. The study highlighted that consumption of less healthy food was more prevalent among young men, greatly contributing to obesity. They also discovered that unhealthy eating patterns established in young adulthood persist and contribute to the development of NCDs. Some key insights were that women consumed more coffee, dairy products, vegetables, and fruits compared to men. Men consumed more trendy and less healthy products like snacks, sweets, fast food, and energy drinks. This study was particularly interesting because purchasing patterns related to gender were observed in Hungary. These findings provided valuable insights that guided us in conducting this experiment in Kenya using the available data.

An interesting study was conducted by Moodley et al.[34] Data was collected using a digital camera and GPS navigation system to record the location of sugar-sweetened beverage (SSB) advertisements and food vendors in Soweto, South Africa. The research highlights that the high sugar intake from SSBs significantly exceeds the World Health Organization's daily sugar intake recommendations and contributes to obesity. The study also suggests that the strategic placement of SSB advertisements in schoolyards or close to schools targets children and promotes unhealthy dietary practices. The findings have significant implications for policies that regulate SSB advertising, particularly around schools, to prevent childhood obesity and promote healthier dietary habits. This study led to the development of a new policy in South Africa, and it was intriguing to consider how our study might similarly influence the creation of a policy promoting healthy food purchases in Kenyan supermarkets.

Machado et al[35] conducted research in Brazil involving a survey on a probabilistic sample of 55 970 Brazilian households between 2008 and 2009. The results of the study suggested that the convenience and lower prices of ultra-processed foods at supermarkets contributed to their higher consumption, which is linked to poorer dietary quality and higher incidences of NCDs. The results also showed that a 1% increase in the number of food items purchased resulted in a 1.83% increase in calorie acquisition from ultra-processed foods. Another notable finding was that supermarkets in Brazil accounted for a higher percentage of calories from ultra-processed foods and beverages (25% higher) than other food stores, driven by lower prices and convenience. The authors suggested that public policies should address the role of supermarkets in terms of the consumption of ultra-processed foods through pricing and convenience mechanisms to promote healthier diets and reduce NCDs. This study similar to Moodley et al[34] showed how supermarkets might be the major driving force of purchasing of unhealthy foods, we used this case study to find out if it is the same case in Kenya. We classified the food items purchased in the 3 million transactions as a means to find out the percentage of ultra-processed food items in Kenyan supermarkets.

Debela et al[36] investigated the relationship between supermarket purchases and child nutrition by analyzing data from households in 3 towns in Kenya, with varying access to supermarkets. The results showed that households with supermarket purchases had significantly higher food variety and dietary diversity scores. These households were more likely to consume nutrient-rich foods such as meat, fish, eggs, and fruits. This study was important as it provided insights into what supermarkets offer to the Kenyan population that other food stores do not. Understanding the Kenyan supermarket landscape significantly enriched the scope of our study.

Rischke et al,[37] like Debela et al,[36] investigated the relationship between supermarket purchases and child nutrition by analyzing data from 453 households in 3 small towns in Central Province, Kenya. The analysis focused on the effect of supermarket purchases on food expenditure shares, calorie shares, and overall calorie availability per adult equivalent per day. The study showed that supermarket purchases increased the consumption of processed and highly processed foods at the expense of unprocessed foods. Similar to findings in Brazil, households that purchased food from supermarkets had higher overall calorie availability per adult equivalent per day. This increase in calorie consumption was driven by lower prices per calorie for foods purchased from supermarkets, particularly processed foods.

Table 1 summarizes the relevant related work explored in this study. Investigations from published work in different countries show that there is a limited amount of research using quantitative data or supermarket transactional data and NOVA classification. Related work did not focus extensively on actual transactional data, especially those conducted in Kenya, to capture the minute details of how populations in their samples made purchases and how these were linked to their demographics and subsequent health effects. This study aims to examine individuals' grocery purchasing patterns, their demographics, and the combined effects of these factors on NCDs and other health-related issues.

Aim

The main aim of this study is to examine the food purchasing patterns of supermarket shoppers and identify associations with demographics in Kenya. Understanding this connection

Table 1. Summary of Related Work.

Author(s)	Sample data size	Method/Approach	Study area	Notable findings
Toiba et al[31]	1180 urban households	Ordinary Least Squares (OLS) and Instrumental Variables (IV) regression	Indonesia	Modern food is expensive and unhealthy
De Roos et al[32]	577 382 entries from the Kantar Worldpanel (KWP) database	ANOVA, Kruskal-Wallis tests, simple linear regression	Scotland	Rural households buy more fruits & veggies
Odunitan-Wayas et al[33]	395 shoppers 11 supermarkets in Cape Town	ANOVA, and ANCOVA	South Africa	High (sea) shoppers purchased healthier food than low (sea) shoppers, fruits and veggies in high sea supermarkets were of better quality than those in low sea supermarkets
Koroknay et al[23]	701 university students, 38.5% male, and 61.5% female	PCA, PLS2B	Hungary	Women eat healthy foods than men, unhealthy eating patterns leads to NCD's
Moodley et al[34]	145 advertisements for SSBs and 180 vendors in Soweto	Spatial point analyzes	South Africa	High consumption of SSBs lead to obesity and type 2 diabetes
Machado et al[35]	55 970 Brazilian households	Log linear regression	Brazil	Supermarkets accounted 25% intake of calories from ultra-processed foods than other food stores
Debela et al[36]	Unknown households	Regression models	Kenya	Purchasing food from supermarkets had a positive effect on child height
Rischke et al[37]	453 households	OLS and instrumental Variables (IV) regression models	Kenya	Supermarkets increase the intake of calorie

will help the general public and policymakers implement demographic-driven measures to reduce the rate of increase of NCDs to manageable levels. Since published work[2-4] has shown a strong connection between purchased food items and NCDs, the objective of this work is to reveal hidden and currently unknown insights from Kenya's purchasing behavioral patterns in their demographic differences and how these can be related to the prevalence of NCDs. The rest of the manuscript is organized as follows: Materials and Methods, Results, Discussion, and Conclusion.

Materials and Methods

The study is a cross-sectional exploratory study with a phased approach employing quantitative secondary data collection from electronic transactional records of supermarkets (departmentalized self-service stores offering wide range of goods from groceries to household goods, electrical appliances to furniture, toys to clothing, stationery to cosmetics all under 1 roof) in 5 counties namely (Nakuru, Kajiado, Nairobi, Kirinyaga, and Machakos) in Kenya.

The county profiles are as follows:

(i) Nakuru is county number 32 out of the 47 Kenyan counties. With an estimated population of over 2.2 million it is the third most populous county after Nairobi and Kiambu. 72% of the population is less than 30 years of age. The population of the county is 52% rural and 48% urban and comprises all major tribes of Kenya. The county is strategically located at the nerve center of the entire nation and opens Kenya to East and Central Africa. Nakuru is an agriculturally-rich county with various tourist attractions such as craters, lakes and national parks and other beneficial economic activities such as production of geothermal energy. The main economic activity in Nakuru County is food and cash crop farming as 70% of the land is arable and highly productive. The main crops produced in the county include: maize, beans, Irish potato, wheat, fruits, vegetables, and flowers. Nakuru County is also endowed with a high capacity for livestock and dairy production.

(ii) Kajiado County is adjacent to Nairobi County and to its south, borders the Tanzanian regions of Arusha and Kilimanjaro. It is county number 34 out of the 47 Kenyan counties and has an estimated total population of over 1.2 million with 59% living in rural areas. The county has several towns that serve as a residential area for many people from different parts of the country who work in the urban centers, especially Nairobi City. Rural-urban migration has contributed to the rise in multiethnic

representation in the major urban centers, with the Maasai being the dominant tribe in the rural areas. Kajiado County is mainly water-stressed with a poverty rate of around 60%. Economic growth and development depend on the main strengths of agriculture, horticulture (tomato, cabbage, kales, and banana), food crop farming (maize, sorghum millet, beans, cowpeas, greengrams, sweet potatoes, cassava, irish potatoes), livestock production (cattle, sheep, goats, and they are kept for milk, meat, offal, raw fats, fresh hides and skins.

(iii) Nairobi County serves as the capital city of Kenya. It is county number 47 out of the 47 Kenyan counties. With an estimated population of over 4.5 million, it is the most populous. Nairobi County which is an established hub for business, manufacturing industries, transport, shopping malls, tourism, and culture. As a cosmopolitan urban African city, it hosts all the 42 diverse tribes of Kenya, and a large immigrant population from other countries in the world. More than 90% of Nairobi residents work within the Nairobi metropolitan area, in the formal and informal sectors. 10% of Nairobi population is located in the high-income upmarket suburbs, 30% in lower-middle and upper-middle-income settlement areas whereas 60% of low and lower-income settlements are located mainly in high-density slum areas.

(iv) Kirinyaga County is located south of Mt. Kenya and is county number 20 out of the 47 Kenyan counties. With an estimated total population of over 0.7 million, 78% live in rural areas. The Kikuyu (largest ethnic group in Kenya) are the dominant tribe in the county. The poverty rate for Kirinyaga is 20%. Agriculture is the main economic activity. The county is best known for rice production from the Mwea irrigation scheme, coffee, and tea. Due to the scarcity of land and high population, most agriculture is done on a small scale. The county also boasts of dairy production and agricultural exports such as horticultural crops (maize, beans, tomatoes, French beans) and macadamia. Kirinyaga County is endowed with a forest cover of around 23% and 6 rivers and is inhabited by a variety of wildlife including elephants, buffaloes, monkeys, bushbucks, and colorful birds while the lower parts of the forest zone provide grazing land for livestock.

(v) Machakos County borders Nairobi to the west, Kajiado to the south west, and Kirinyaga to the north west. It is county number 16 out of the 47 Kenyan counties and has an estimated population of 1.5 million. The Kamba (fifth largest ethnic group in Kenya) are the dominant tribe in the county. Around 30% of the population living in the urban areas are middle-income and mostly work in Nairobi's city due to the close proximity. The county has a monetary poverty rate of 23.1% and a multidimensional poverty rate of 39.4%. After the manufacturing sector (cement factories, steel mills, export processing zone companies), agriculture is the second most important sector. Over 70% of the County population live in rural areas and derive their principal livelihoods, directly or indirectly from agriculture. Subsistence agriculture is mostly practiced with maize and drought-resistant crops such as sorghum and millet being grown due to the area's semi-arid state. Machakos County is faced with an accelerating crisis of drought, heat stress, diminishing water resources and increased temperatures leading to food shortage, increased food prices, and scarcity of water.

The electronic records contain transactional data from January 2022 to December 2023. These counties were purposely sampled since grocery data was received from supermarket branches located in those counties. These counties are represented by different ethnic and cultural differences in the populations.

The study population includes records of individuals who have made purchases in a tier 2 supermarket with 10 branches spread across 5 Kenyan counties. Currently in Kenya, there are over 20 supermarket chains with over 300 stores where food (dry groceries and fresh produce) represents around 60% of total sales. They are classified in terms of their market share, number of stores, and amount of space occupied as first-tier, second-tier, and third-tier. The first tier comprises International and hybrid (International and local) owned supermarket chains that are the clear market leaders with over 25 stores located in major towns and strategic shopping malls, the second comprises mostly local owned large supermarket chains with between 5 and 25 stores and third comprises of independent and mini supermarket chains with between 1 and 4 stores.

Ethical approval was granted by Amref Ethics and Review Committee (ESRC) (approval no. AMREF-ESRC P1526/2023) on November 15, 2023.

The reporting of this study conforms to STROBE-nutreporting guidelines.[38]

Description of Data

Supermarket variables: An abstraction tool was developed to guide in secondary data collection of the main variables of interest. Data was not identifiable as it was anonymized at the supermarket and client level. Variables collected are presented in Table 2. This table shows that there were on average 47% of males and 53% of females who were recorded in the transactions. The average age of the shoppers was 40.2 years of age, showing that mature people were the main customers not children. This is important because the results

Table 2. Main Data Values.

Attribute	Data value
Sample size shoppers	103 931
Gender	Females = 53.0% Males = 47.0%
Age (Years)	Mean (SD) = 40.2 (11.7) Median (IQR) = 38.7 (31.1, 47.7)
Shoppers per county	Kajiado (30.4%), Kirinyaga (20.6%), Machakos (7.8%), Nairobi (33.4%), Nakuru (7.8%)
Unique transactions	3 934 122
Transaction dates	1st January 2022-31st December 2023
Transactions per county	Kajiado (1 495 392), Kirinyaga (503 877), Machakos (395 188), Nairobi (1 245 453), Nakuru (294 212)
Supermarket type	Tier 2
Supermarket branches	10
Supermarket branches per county	Kajiado (2), Kirinyaga (2), Machakos (1), Nairobi (4), Nakuru (1)
Food categories	103
Food category items per transactions	Range: 1-96 1 item (18.0%), 2 items (22.4%), 3 items (17.9%), 4 items (12.8%), 5 items (8.7%), 6 items (6.0%), 7 items (4.0%), 8 items (2.8%), 9 items (2.0%), 10 items (1.4%), 11 + items (4%)
Nova classes	Processed foods, Processed culinary Ingredients, Unprocessed/Minimally processed foods, Ultra-processed foods
Popular nova class 1	Ultra-processed foods (53.4%)
Popular nova class 2	Unprocessed/minimally processed foods (35.2%)
Popular nova class 3	Processed culinary ingredients (10.5%)
Popular nova class 4	Processed foods (0.9%)

will speak to the purchasing patterns of the adult population of Kenya. There were a total of 103 931 unique supermarket shoppers. The majority of the shoppers were from Nairobi (33.4%) and the biggest proportion aged 36 to 49 (40.2%).

NOVA food classification is a system developed by researchers at the University of Sao Paulo, Brazil[39,40] and categorizes all foods into 4 groups according to the nature, extent, and purposes of the industrial processes they undergo.

Unprocessed foods: natural foods which are edible parts of plants such as fruit, leaves, stems, seeds, or animals such as muscle, offal, eggs, milk.

Processed culinary ingredients: ingredients include oils, butter, lard, sugar, and salt and these are substances derived from nature by processes such as pressing, refining, grinding, milling, and drying.

Processed foods: include canned or bottled vegetables or legumes (pulses) preserved in brine, they are usually made by adding salt, oil, sugar, or other substances from processed culinary foods.

Ultra-processed foods: the foods are formulations of ingredients, mostly of exclusive industrial use, typically created by a series of industrial techniques and processes. Some common ultra-processed products are carbonated soft drinks; sweet, fatty, or salty packaged snacks.

Process Workflow

Table 2 was created from descriptive statistics in the form of frequency, percentage, sum, mean, and standard deviation to examine demographic characteristics of the customers and food classification. The extracted grocery data in the form of csv files was saved into a local database using PostgreSQL version 15.2[41] and imported into R version 4.3.3[42] for cleaning and analysis as highlighted in Figure 1. The data had 47 748 603 transactional records. Since the data came from 10 supermarket branches in the 5 counties, the data was augmented to have a holistic view of the transactions. The alignment included transactions and demographics alignment. Some data preprocessing techniques were applied (which include: parsing, re-formatting, dealing with missing values etc) to check for data consistency and quality assurance checks, filtering non-food items and non-loyalty shoppers. The final dataset for analysis had 15 210 101 transactional records.

After the data preprocessing step, the data was restructured to fit both the alignment and to prepare it for some basic data analysis which will give us some more insights in the transactions. The basic analytics included a chi-square test and Fisher's exact test (when more than 20% of cells have expected frequencies <5) where appropriate to analyze the differences in NOVA food classification groups with gender, age, location, and socio-demographic analysis. Statistical significance was considered at P-value $< .05$. The reason for this analysis was to have a detailed understanding of the social and demographic structure of the customers, as well as to understand the relationship between these 2 with the transactions before association rule mining. Association rule mining was implemented

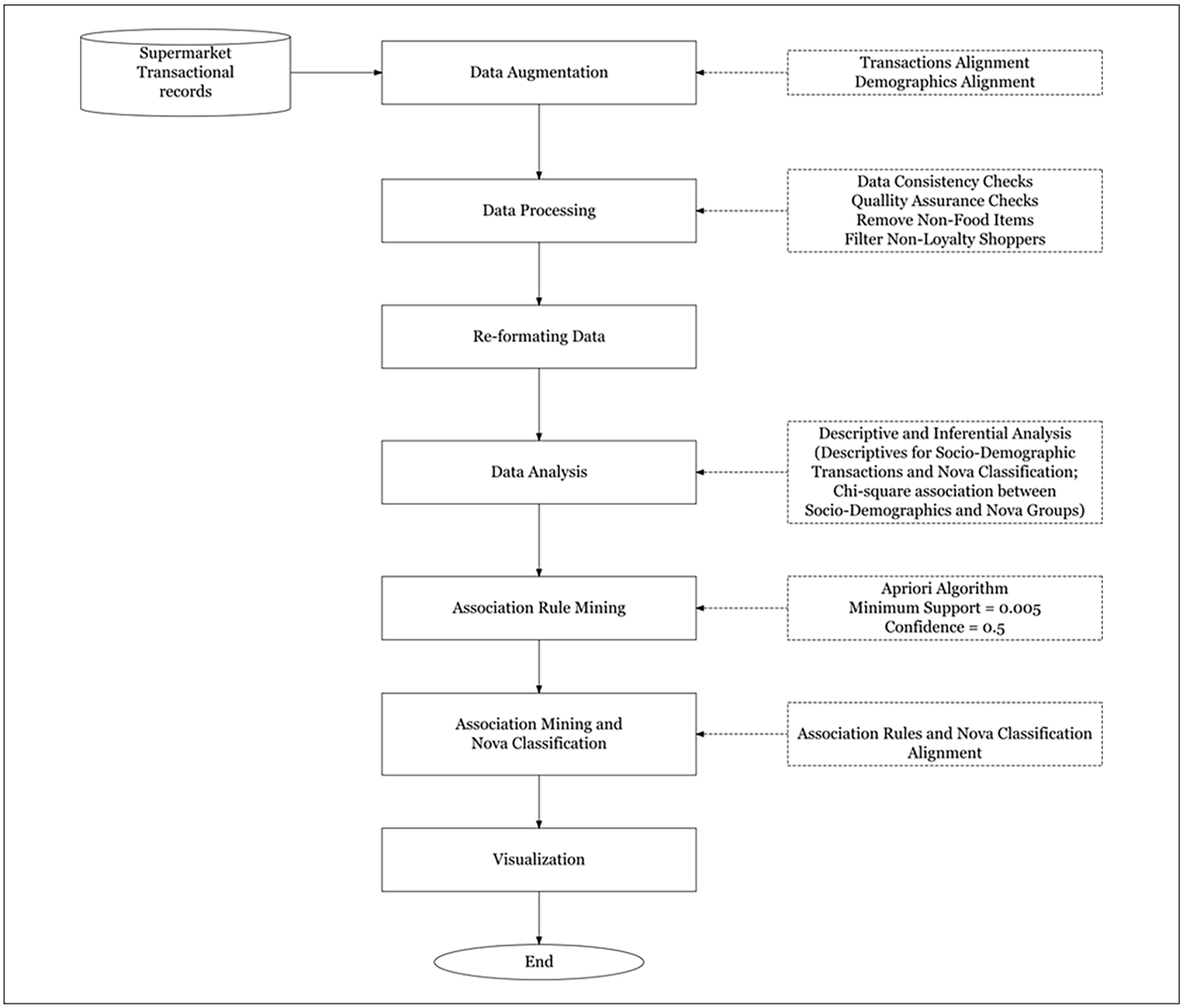

Figure 1. Process workflow.

using the apriori algorithm to understand items which were bought together frequently and with which demographic groups. After that, a linkage was built between the association rules and the NOVA classification, the main reason for that was to understand the frequent transaction patterns and the nature of the food items in those frequent transactions. Finally, visualization was implemented to generate the graphic view of results and their implications. The graphical representations create a holistic understanding of everything to the audience, especially the policy makers.

Results

Initially, the results reveal insights into how different demographic groups behave in terms of their purchasing habits. This is followed by an exploration of the variations in purchase patterns across different demographic groups concerning the NOVA classification of food items. Additionally, the results identify the main food items prevalent in the frequently purchased NOVA categories for each demographic group. By linking these purchasing patterns to demographic data, the study aims to identify potential non-communicable diseases (NCDs) that may disproportionately affect certain demographic groups.

Purchase Patterns by Demographics

Shoppers aged 26 to 35, 36 to 49, 50, and above were proportionally more in Nakuru, Nairobi, and Machakos counties respectively. Figure 2 shows that there were proportionally more female shoppers than males across age groups of 18 to 25 and 36 to 49 years whereas there were proportionally

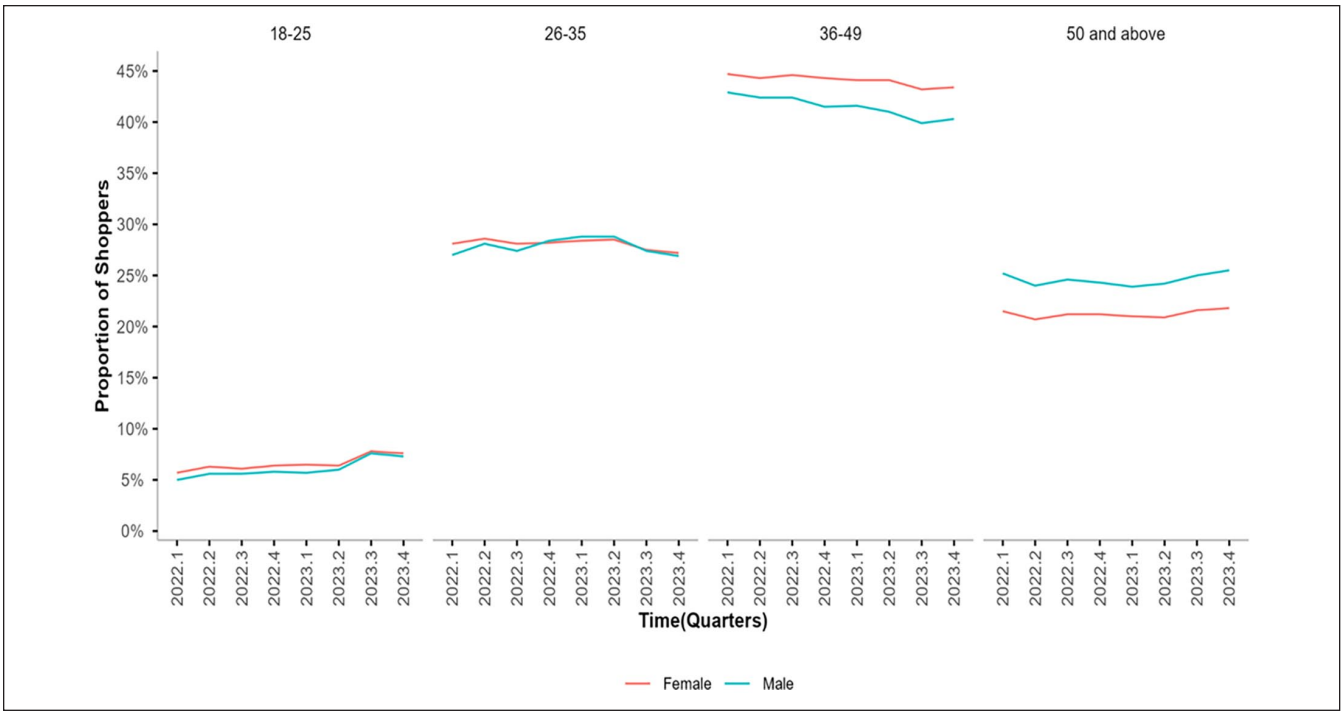

Figure 2. Age-gender trend of shoppers across the study period.

more male shoppers than females in the age group of 50 and above years.

Table 3 summarizes the proportion of different food category purchases and the association with gender, age, and location. Over three-fifths (60.87%) of the food purchases are accounted for by 12 out of 103 unique food categories, that is, breads (13.58%); whole milk (12.25%); sugar (5.95%); maize/corn flour (4.82%); flavored/fruit yogurt (4.65%); wheat flour (3.94%); sweet and savory snacks (3.52%); cookies, biscuits, and wafers (2.82%); sodas and carbonated drinks (2.67%); tea (2.33%); fruit drinks/sweetened juice and bottled teas (2.19%); pastries (2.16%).

Gender and age. Figure 3 shows that males in the age group between 26 and 35 were frequently purchasing bread more than females. However, more males in the age groups of 36 to 49 and above 50 added sugar and whole milk to their purchases. We can also observe a slight difference in the purchasing patterns of males and females from Figure 3. Males in the age groups of 36 to 49 and above 50 were also frequently purchasing flavored yogurts, pastries, carbonated drinks, and candies as shown by higher values of support (support for an association rule flavored yogurt $\Rightarrow$ pastries is the percentage of transactions in the database that contain flavored yogurt $\cup$ pastries). Additionally, more males above 50 were frequently purchasing sweetened juices and bottled teas. On the other hand, females, especially those in the age groups of 36 to 49 and above 50, were frequently purchasing whole milk, salt, whole flour, and maize flour.

Age and location. Figure 4 shows that people in the age group of 18 to 25, especially those in Machakos, were frequently purchasing bread and whole milk more than in any other counties. People in Kajiado, Nairobi, and Machakos in the age groups of 26 to 35, 36 to 49, and above 50 were frequently purchasing bread, whole milk, and sugar. We can also observe a trend of purchasing flavored yogurts, sweet and savory snacks in Kajiado by the age group of 26 to 35. Those in the ages of 36 to 49 and above 50 added corn flour to their purchases in the same county. Similarly, those in Machakos aged 36 to 49 and above 50 were buying flavored yogurts, sweet and savory snacks. Nairobi comes in third place in this trend.

Gender and location. Figure 5 shows that both males and females in Kajiado, Machakos, and slightly Nairobi were very interested in bread and whole milk. In Machakos and Kirinyaga, they also added maize or corn flour. We can also observe that both males and females in Kajiado were much more interested in purchasing flavored yogurt, candies, sweet and savory snacks, sodas and carbonated drinks, pastries, cupcakes, and many more items compared to any other county.

Frequent Purchase Patterns and NOVA Classes

Table 3 summarizes the proportion of food purchases found in this study according to NOVA classification and the association with gender, age, and location. Ultra-processed foods accounted for the highest purchases (53.440%), closely

Table 3. Food Category Purchases Stratified by Gender, Age and Location.

			Gender*			Age*					Location*					
		Overall N = 15210101	Female N = 8034182	Male N = 7150966		18-25 N = 636358	26-35 N = 4120386	36-49 N = 6964184	50 and above N = 3401311		Kajiado N = 6563764	Kirinyaga N = 1928522	Machakos N = 1543953	Nairobi N = 4065242	Nakuru N = 1108620	
Nova classification	Food category	%	%	%	*P*-value	%	%	%	%	*P*-value	%	%	%	%	%	*P*-value
Processed foods	Total	0.864	0.851	0.879	**<.001**	0.944[a]	0.905[b]	0.837[c]	0.861[d]	**<.001**	0.759[a]	1.196[b]	0.918[c]	0.839[d]	0.918[c]	**<.001**
	Salted/sugared nuts	0.445	0.387	0.510	**<.001**	0.333[a]	0.406[b]	0.453[c]	0.503[d]	**<.001**	0.406[a]	0.835[b]	0.643[c]	0.244[d]	0.461[e]	**<.001**
	Tomato paste	0.315	0.375	0.248	**<.001**	0.481[a]	0.395[b]	0.288[c]	0.244[d]	**<.001**	0.253[a]	0.222[b]	0.185[c]	0.515[d]	0.287[e]	**<.001**
	Malt drinks	0.049	0.027	0.075	**<.001**	0.077[a]	0.046[b]	0.049[b]	0.047[b]	**<.001**	0.037[a]	0.076[b]	0.024[c]	0.056[d]	0.085[e]	**<.001**
	Bacon/ham	0.030	0.036	0.024	**<.001**	0.034[a,b]	0.032[b]	0.026[c]	0.038[a]	**<.001**	0.033[a]	0.036[b]	0.045[c]	0.014[d]	0.045[c]	**<.001**
	Canned pulses - peas, lentils, beans	0.009	0.009	0.009	.428	0.007[a]	0.010[b]	0.008[a]	0.011[b]	**<.001**	0.012[a]	0.009[b]	0.008[b]	0.004[c]	0.013[a]	**<.001**
	Cheese	0.009	0.010	0.007	**<.001**	0.008[a,b]	0.010[c]	0.007[b]	0.009[a,c]	**<.001**	0.011[a,b]	0.010[b]	0.006[c]	0.004[d]	0.013[a]	**<.001**
	Processed fish	0.004	0.004	0.004	.553	0.003[a]	0.003[a]	0.003[a]	0.007[b]	**<.001**	0.004[a]	0.004[a]	0.005[a,b]	0.002[c]	0.006[b]	**<.001**
	Canned corn kernel	0.002	0.002	0.001	.601	0.001[a,b]	0.002[b]	0.001[a]	0.002[a,b]	**.018**	0.001[a]	0.002[b]	0.001[a]	0.001[a]	0.004[c]	**<.001**
	Others	0.001	0.001	0.001	**.001**	0.001[a,b]	0.002[b]	0.001[a]	0.001[a]	**<.001**	0.001[a]	0.002[b]	0.001[a,c]	0.000[c]	0.003[b]	**<.001**
Processed culinary ingredients	Total	10.468	10.061	10.931	**<.001**	10.973[a]	10.188[b]	10.112[c]	11.420[d]	**<.001**	8.640[a]	13.016[b]	12.557[c]	11.813[d]	9.023[e]	**<.001**
	Sugar	5.947	5.539	6.410	**<.001**	6.207[a]	5.809[b]	5.827[b]	6.292[c]	**<.001**	4.688[a]	7.329[b]	7.075[c]	7.191[d]	4.861[e]	**<.001**
	Cooking oil	2.144	2.016	2.290	**<.001**	2.199[a]	2.034[b]	2.000[c]	2.559[d]	**<.001**	1.856[a]	2.814[b]	3.016[c]	2.026[d]	1.903[e]	**<.001**
	Salt	1.538	1.683	1.376	**<.001**	1.682[a]	1.540[b]	1.470[c]	1.652[a]	**<.001**	1.346[a]	1.839[b]	1.693[c]	1.672[c]	1.444[d]	**<.001**
	Cooking fat	0.514	0.502	0.526	**<.001**	0.561[a]	0.470[b]	0.514[c]	0.553[a]	**<.001**	0.412[a]	0.651[b]	0.434[c]	0.645[b]	0.504[d]	**<.001**
	Sugar - powdered	0.154	0.154	0.153	.928	0.177[a]	0.179[a]	0.147[b]	0.132[c]	**<.001**	0.139[a]	0.187[b]	0.189[b]	0.152[c]	0.136[a]	**<.001**
	Honey	0.124	0.118	0.131	**<.001**	0.118[a]	0.114[a]	0.113[a]	0.162[b]	**<.001**	0.124[a]	0.179[b]	0.117[c]	0.102[d]	0.121[a,c]	**<.001**
	Butter	0.036	0.037	0.034	**<.001**	0.023[a]	0.030[b]	0.032[b]	0.056[c]	**<.001**	0.055[a]	0.013[b]	0.027[c]	0.020[d]	0.035[e]	**<.001**
	Ghee	0.010	0.010	0.010	**.319**	0.005[a]	0.011[b]	0.009[c]	0.012[b]	**<.001**	0.017[a]	0.002[b]	0.005[c]	0.003[b]	0.012[d]	**<.001**
	Others	0.002	0.003	0.001	**<.001**	0.001[a]	0.002[a]	0.002[a]	0.003[b]	**<.001**	0.002[a]	0.002[a]	0.001[b]	0.002[a]	0.006[c]	**<.001**
Unprocessed/minimally processed foods	Total	35.228	34.632	35.912	**<.001**	33.977[a]	34.589[b]	34.660[c]	37.364[d]	**<.001**	36.237[a]	25.881[b]	39.396[c]	37.822[d]	30.203[e]	**<.001**
	Whole milk	12.252	11.445	13.165	**<.001**	10.000[a]	11.231[b]	12.520[c]	13.295[d]	**<.001**	12.389[a]	6.778[a]	13.983[b]	15.392[a]	7.037[c]	**<.001**
	Maize/corn flour	4.815	4.404	5.283	**<.001**	5.485[a]	5.032[b]	4.618[c]	4.812[d]	**<.001**	4.124[a]	4.033[b]	4.996[c]	6.611[d]	3.432[e]	**<.001**
	Wheat flour	3.942	4.128	3.735	**<.001**	4.526[a]	4.010[b]	3.848[c]	3.942[d]	**<.001**	3.078[a]	4.518[b]	4.149[c]	4.983[d]	3.956[e]	**<.001**
	Tea	2.330	2.290	2.377	**<.001**	2.239[a]	2.300[b]	2.255[a]	2.536[c]	**<.001**	2.230[a]	1.828[b]	3.722[c]	2.391[d]	1.628[e]	**<.001**
	Natural/plain fermented milk	1.663	1.588	1.748	**<.001**	1.106[a]	1.458[b]	1.693[c]	1.941[d]	**<.001**	1.814[a]	0.613[b]	1.994[c]	1.886[d]	1.314[e]	**<.001**
	Cereal grains	1.256	1.120	1.411	**<.001**	1.471[a]	1.234[b]	1.182[c]	1.393[d]	**<.001**	1.321[a]	0.113[b]	2.518[c]	1.189[d]	1.352[e]	**<.001**
	Herbs and spices	1.157	1.403	0.880	**<.001**	1.259[a]	1.162[b]	1.053[c]	1.358[d]	**<.001**	1.594[a]	0.878[b]	0.811[c]	0.561[d]	1.723[e]	**<.001**
	Water	1.076	0.945	1.224	**<.001**	0.897[a]	0.892[a]	1.103[b]	1.254[c]	**<.001**	1.027[a]	1.005[b]	1.161[c]	1.229[d]	0.813[e]	**<.001**
	Coffee	0.867	0.859	0.876	**<.001**	1.036[a]	0.873[b]	0.818[c]	0.934[d]	**<.001**	0.827[a]	0.968[b]	0.791[c]	0.900[d]	0.909[d]	**<.001**
	Fruits	0.799	0.832	0.763	**<.001**	0.616[a]	0.862[b]	0.732[c]	0.912[d]	**<.001**	1.186[a]	0.377[b]	1.179[a]	0.096[c]	1.297[d]	**<.001**
	Pastas	0.766	0.863	0.658	**<.001**	1.022[a]	0.950[b]	0.744[c]	0.548[d]	**<.001**	0.909[a]	0.798[b]	0.715[c]	0.543[d]	0.756[e]	**<.001**
	Fresh/fruit juices - no added sugar	0.727	0.736	0.718	**<.001**	0.769[a]	0.828[b]	0.718[c]	0.620[d]	**<.001**	0.825[a]	0.824[a]	0.717[b]	0.466[c]	0.951[d]	**<.001**

(continued)

Table 3. (continued)

			Gender*			Age*					Location*					
		Overall N = 15210101	Female N = 8034182	Male N = 7150966		18-25 N = 636358	26-35 N = 4120386	36-49 N = 6964184	50 and above N = 3401311		Kajiado N = 6563764	Kirinyaga N = 1928522	Machakos N = 1543953	Nairobi N = 4065242	Nakuru N = 1108620	
Nova classification	Food category	%	%	%	*P*-value	%	%	%	%	*P*-value	%	%	%	%	%	*P*-value
	Vegetables	0.657	0.796	0.501	**<.001**	0.490[a]	0.545[b]	0.640[c]	0.872[d]	**<.001**	1.355[a]	0.000[b]	0.000[b]	0.000[b]	0.995[c]	**<.001**
	Red millet flour	0.580	0.599	0.559	**<.001**	0.565[a]	0.652[b]	0.530[c]	0.598[d]	**<.001**	0.568[a]	0.609[b]	0.971[c]	0.463[d]	0.482[e]	**<.001**
	Red meat	0.568	0.682	0.437	**<.001**	0.623[a]	0.581[b]	0.538[c]	0.613[a]	**<.001**	0.756[a]	0.969[b]	0.000[c]	0.024[d]	1.538[e]	**<.001**
	Eggs	0.381	0.408	0.352	**<.001**	0.488[a]	0.446[b]	0.346[c]	0.360[d]	**<.001**	0.312[a]	0.304[a]	0.651[b]	0.386[c]	0.534[d]	**<.001**
	Pulses - peas, lentils, beans	0.253	0.271	0.234	**<.001**	0.234[a]	0.254[b]	0.247[c]	0.272[d]	**<.001**	0.332[a]	0.070[b]	0.307[c]	0.197[d]	0.241[e]	**<.001**
	Poultry	0.199	0.229	0.164	**<.001**	0.261[a]	0.221[b]	0.182[c]	0.197[d]	**<.001**	0.287[a]	0.264[b]	0.056[c]	0.043[d]	0.336[e]	**<.001**
	Baby foods - porridge flour	0.186	0.172	0.203	**<.001**	0.198[a]	0.297[b]	0.173[c]	0.078[d]	**<.001**	0.189[a]	0.207[b]	0.261[c]	0.139[d]	0.208[b]	**<.001**
	Natural/plain yogurt	0.138	0.151	0.124	**<.001**	0.114[a]	0.138[b]	0.127[c]	0.165[d]	**<.001**	0.143[a]	0.184[b]	0.082[c]	0.126[d]	0.155[e]	**<.001**
	Root vegetables	0.137	0.163	0.107	**<.001**	0.123[a]	0.134[b]	0.122[a]	0.176[c]	**<.001**	0.308[a]	0.000[b]	0.000[b]	0.000[b]	0.054[c]	**<.001**
	Popcorn kernels	0.078	0.094	0.059	**<.001**	0.103[a]	0.094[b]	0.078[c]	0.053[d]	**<.001**	0.088[a]	0.100[b]	0.062[c]	0.043[d]	0.129[e]	**<.001**
	Dried fruits	0.060	0.069	0.049	**<.001**	0.040[a]	0.053[b]	0.058[c]	0.077[d]	**<.001**	0.124[a]	0.009[b]	0.013[c]	0.009[b]	0.020[d]	**<.001**
	Fish	0.058	0.064	0.051	**<.001**	0.059[a]	0.060[a]	0.052[b]	0.069[c]	**<.001**	0.068[a]	0.100[b]	0.121[c]	0.006[d]	0.030[e]	**<.001**
	Porridge flour	0.046	0.046	0.046	.634	0.052[a]	0.054[a]	0.044[b]	0.040[c]	**<.001**	0.047[a]	0.062[b]	0.036[c]	0.032[d]	0.077[e]	**<.001**
	Organ meat	0.045	0.057	0.032	**<.001**	0.029[a]	0.046[b,c]	0.045[c]	0.049[b]	**<.001**	0.085[a]	0.066[b]	0.000[c]	0.000[c]	0.005[d]	<.001
	Edible clay	0.042	0.060	0.021	**<.001**	0.034[a]	0.040[b]	0.052[c]	0.024[d]	**<.001**	0.048[a]	0.069[b]	0.021[c]	0.033[d]	0.021[c]	**<.001**
	Milk powder	0.041	0.040	0.042	.066	0.052[a]	0.042[b]	0.038[c]	0.043[b]	**<.001**	0.055[a]	0.064[b]	0.019[c]	0.016[d]	0.035[e]	**<.001**
	Herbal infusions	0.033	0.041	0.025	**<.001**	0.029[a,b]	0.029[b]	0.033[a]	0.040[c]	**<.001**	0.034[a]	0.049[b]	0.029[c]	0.027[c]	0.035[a]	**<.001**
	Poultry giblets	0.033	0.038	0.028	**<.001**	0.025[a]	0.037[b]	0.029[a]	0.039[b]	**<.001**	0.051[a]	0.015[b]	0.009[c]	0.003[d]	0.106[e]	**<.001**
	Dried fruits and nuts	0.013	0.011	0.016	**<.001**	0.005[a]	0.010[b]	0.014[c]	0.017[d]	**<.001**	0.015[a]	0.006[b]	0.019[c]	0.010[d]	0.024[e]	**<.001**
	Others	0.010	0.010	0.010	.55	0.020[a]	0.011[b]	0.009[c]	0.008[d]	**<.001**	0.010[a]	0.002[b]	0.004[c]	0.017[d]	0.006[c]	**<.001**
	Raw nuts	0.010	0.010	0.009	.148	0.005[a]	0.005[a]	0.009[b]	0.018[c]	**<.001**	0.022[a]	0.000[b,c]	0.000[c]	0.001[b]	0.002[d]	**<.001**
	Edible seeds	0.008	0.009	0.006	**<.001**	0.003[a]	0.006[b]	0.007[c]	0.012[d]	**<.001**	0.018[a]	0.000[b]	0.000[b]	0.000[b]	0.003[c]	**<.001**
Ultra-processed foods	Total	53.440	54.457	52.278	**<.001**	54.105[a]	54.317[b]	54.392[c]	50.355[d]	**<.001**	54.365[a]	59.906[b]	47.129[c]	49.520[d]	59.857[b]	**<.001**
	Breads	13.580	12.275	15.046	**<.001**	10.645[a]	11.395[b]	14.011[c]	15.840[d]	**<.001**	13.628[a]	10.404[b]	14.231[c]	15.159[d]	12.121[e]	**<.001**
	Flavored/fruit yogurt	4.648	4.689	4.602	**<.001**	4.650[a]	5.412[b]	4.678[a]	3.675[c]	**<.001**	4.583[a]	6.302[b]	3.724[c]	4.390[d]	4.379[d]	**<.001**
	Sweet and savory snacks	3.523	3.947	3.045	**<.001**	3.379[a]	3.939[b]	3.861[c]	2.364[d]	**<.001**	3.709[a]	4.387[b]	2.750[c]	3.024[d]	3.832[e]	**<.001**
	Cookies, biscuits and wafers	2.818	3.073	2.533	**<.001**	2.723[a]	3.100[b]	2.977[c]	2.173[d]	**<.001**	2.879[a]	3.072[b]	2.858[a]	2.461[c]	3.272[d]	**<.001**
	Sodas and carbonated drinks	2.665	2.588	2.751	**<.001**	3.144[a]	2.592[b]	2.639[c]	2.701[d]	**<.001**	2.663[a]	3.329[b]	2.613[c]	2.250[d]	3.113[e]	**<.001**
	Fruit drinks/sweetened juices & bottled teas	2.188	2.159	2.221	**<.001**	2.420[a]	2.601[b]	2.122[c]	1.774[d]	**<.001**	2.125[a]	2.427[b]	2.350[c]	1.998[d]	2.611[e]	**<.001**
	Pastries	2.163	2.062	2.277	**<.001**	1.944[a]	2.241[b]	2.264[c]	1.926[a]	**<.001**	2.375[a]	2.827[b]	1.540[c]	1.594[d]	2.717[e]	**<.001**
	Cupcakes/muffins/ queencakes	1.996	2.080	1.902	**<.001**	1.984[a]	2.015[a]	2.113[b]	1.746[c]	**<.001**	1.953[a]	2.236[b]	2.249[b]	1.945[a]	1.670[c]	**<.001**
	Fast foods	1.965	2.183	1.716	**<.001**	3.095[a]	2.205[b]	1.813[c]	1.773[d]	**<.001**	1.669[a]	2.356[b]	0.000[c]	1.967[d]	5.770[e]	**<.001**
	Candy and confectionery	1.912	2.215	1.566	**<.001**	1.897[a]	2.063[b]	2.053[b]	1.452[c]	**<.001**	2.400[a]	2.153[b]	1.529[c]	0.948[d]	2.669[e]	**<.001**
	Margarine	1.911	1.980	1.834	**<.001**	1.647[a]	1.806[b]	1.963[c]	1.981[d]	**<.001**	1.829[a]	2.243[b]	2.094[c]	1.901[d]	1.604[e]	**<.001**
	Buns/rolls	1.808	1.916	1.685	**<.001**	1.589[a]	1.526[b]	1.764[c]	2.301[d]	**<.001**	2.010[a]	2.130[b]	0.429[c]	1.658[d]	2.520[e]	**<.001**
	Cocoa	1.221	1.339	1.086	**<.001**	1.015[a]	1.114[b]	1.241[c]	1.353[d]	**<.001**	1.148[a]	1.774[b]	1.239[c]	1.118[d]	1.040[e]	**<.001**
	Processed meat	1.082	1.266	0.875	**<.001**	1.128[a]	1.108[a]	1.091[b]	1.022[c]	**<.001**	1.061[a]	1.476[b]	0.973[c]	1.004[d]	0.956[c]	**<.001**

(continued)

Table 3. (continued)

			Gender*			Age*					Location*					
		Overall N = 15210101	Female N = 8034182	Male N = 7150966		18-25 N = 636358	26-35 N = 4120386	36-49 N = 6964184	50 and above N = 3401311		Kajiado N = 6563764	Kirinyaga N = 1928522	Machakos N = 1543953	Nairobi N = 4065242	Nakuru N = 1108620	
Nova classification	Food category	%	%	%	*P*-value	%	%	%	%	*P*-value	%	%	%	%	%	*P*-value
	Chocolate	1.017	1.114	0.907	**<.001**	1.320[a]	1.168[b]	1.014[c]	0.787[d]	**<.001**	1.100[a]	1.165[b]	1.103[a]	0.680[c]	1.385[d]	**<.001**
	Cakes	0.929	0.976	0.875	**<.001**	1.271[a]	1.024[b]	0.926[c]	0.767[d]	**<.001**	1.124[a]	1.400[b]	0.143[c]	0.499[d]	1.625[e]	**<.001**
	Milk drinks	0.804	0.809	0.796	**.004**	0.814[a]	0.937[b]	0.855[c]	0.541[d]	**<.001**	0.932[a]	0.918[a]	0.585[b]	0.633[c]	0.777[d]	**<.001**
	Table sauces	0.753	0.912	0.573	**<.001**	0.961[a]	0.832[b]	0.724[c]	0.683[d]	**<.001**	0.831[a]	0.814[b]	0.544[c]	0.654[d]	0.842[a]	**<.001**
	Noodles	0.752	0.875	0.613	**<.001**	1.096[a]	0.849[b]	0.767[c]	0.539[d]	**<.001**	0.783[a]	0.971[b]	0.642[c]	0.647[c]	0.724[d]	**<.001**
	Diluting juice	0.713	0.748	0.674	**<.001**	1.201[a]	0.882[b]	0.652[c]	0.549[d]	**<.001**	0.694[a]	1.194[b]	0.754[c]	0.509[d]	0.686[a]	**<.001**
	Marble cakes	0.705	0.701	0.709	.083	0.928[a]	0.797[b]	0.711[c]	0.538[d]	**<.001**	0.523[a]	1.341[b]	0.833[c]	0.604[d]	0.865[e]	**<.001**
	Energy drinks	0.598	0.384	0.838	**<.001**	0.604[a]	0.561[b]	0.598[a]	0.634[c]	**<.001**	0.529[a]	0.691[b]	0.583[c]	0.674[d]	0.586[c]	**<.001**
	Breakfast cereals	0.585	0.663	0.496	**<.001**	0.500[a]	0.672[b]	0.605[c]	0.453[d]	**<.001**	0.669[a]	0.565[b]	0.570[b]	0.455[c]	0.615[d]	**<.001**
	Ice cream	0.530	0.633	0.413	**<.001**	1.148[a]	0.677[b]	0.467[c]	0.365[d]	**<.001**	0.538[a]	0.907[b]	0.540[a]	0.297[c]	0.659[d]	**<.001**
	Soup flavor cubes & mixes	0.496	0.599	0.380	**<.001**	0.603[a]	0.529[b]	0.461[c]	0.509[d]	**<.001**	0.489[a]	0.518[b]	0.440[c]	0.492[a]	0.583[d]	**<.001**
	Jam/other spreads	0.426	0.473	0.374	**<.001**	0.376[a]	0.418[b]	0.434[c]	0.429[c]	**<.001**	0.394[a]	0.517[b]	0.419[c]	0.435[d]	0.434[c,d]	**<.001**
	Flavored fermented milk	0.351	0.333	0.371	**<.001**	0.294[a]	0.382[b]	0.351[c]	0.325[d]	**<.001**	0.368[a]	0.233[b]	0.513[c]	0.297[d]	0.427[e]	**<.001**
	Baking additives	0.259	0.294	0.220	**<.001**	0.377[a]	0.267[b]	0.239[c]	0.271[b]	**<.001**	0.219[a]	0.190[b]	0.169[c]	0.418[d]	0.161[c]	**<.001**
	Peanut butter	0.214	0.235	0.189	**<.001**	0.227[a]	0.226[a]	0.209[b]	0.206[b]	**<.001**	0.208[a]	0.296[b]	0.225[c]	0.165[d]	0.266[e]	**<.001**
	Processed poultry	0.189	0.215	0.159	**<.001**	0.214[a]	0.209[a]	0.187[b]	0.165[c]	**<.001**	0.243[a]	0.213[b]	0.010[c]	0.140[d]	0.248[a]	**<.001**
	Cooking sauces	0.151	0.175	0.123	**<.001**	0.252[a]	0.175[b]	0.136[c]	0.134[c]	**<.001**	0.141[a]	0.278[b]	0.127[c]	0.106[d]	0.187[e]	**<.001**
	Baby foods - wheat cereals	0.134	0.138	0.129	**<.001**	0.126[a]	0.202[b]	0.136[c]	0.049[d]	**<.001**	0.163[a]	0.136[b]	0.101[c]	0.091[d]	0.162[a]	**<.001**
	Soya chunks/mince	0.098	0.116	0.079	**<.001**	0.205[a]	0.109[b]	0.091[c]	0.081[d]	**<.001**	0.121[a]	0.079[b]	0.051[c]	0.089[d]	0.094[d]	**<.001**
	Food color additives	0.054	0.062	0.045	**<.001**	0.094[a]	0.071[b]	0.047[c]	0.043[d]	**<.001**	0.044[a]	0.068[b]	0.024[c]	0.079[d]	0.044[a]	**<.001**
	Soya drink	0.051	0.050	0.053	**.003**	0.039[a]	0.038[a]	0.048[b]	0.078[c]	**<.001**	0.048[a]	0.104[b]	0.038[c]	0.040[c]	0.037[c]	**<.001**
	Food flavor additives	0.039	0.048	0.030	**<.001**	0.068[a]	0.047[b]	0.038[c]	0.029[d]	**<.001**	0.039[a]	0.067[b]	0.027[c]	0.028[c]	0.050[d]	**<.001**
	Chocolate spreads	0.025	0.030	0.020	**<.001**	0.025[a,b]	0.025[b]	0.029[a]	0.018[c]	**<.001**	0.035[a]	0.026[b]	0.029[b]	0.010[c]	0.012[c]	**<.001**
	Infant formula	0.019	0.019	0.019	.946	0.022[a]	0.029[b]	0.016[c]	0.013[d]	**<.001**	0.021[a]	0.032[b]	0.025[c]	0.008[d]	0.019[a]	**<.001**
	Plant milk	0.017	0.021	0.013	**<.001**	0.025[a]	0.017[b]	0.016[c]	0.018[b]	**<.001**	0.023[a]	0.011[b,c]	0.013[c]	0.010[b]	0.024[a]	**<.001**
	Bread crumbs	0.014	0.016	0.011	**<.001**	0.019[a]	0.019[a]	0.012[b]	0.010[c]	**<.001**	0.011[a]	0.008[b]	0.012[a,c]	0.023[d]	0.015[c]	**<.001**
	Isotonic drinks	0.013	0.013	0.012	**.043**	0.012[a,b]	0.014[b]	0.012[a]	0.013[a,b]	.057	0.012[a]	0.020[b]	0.012[a]	0.008[c]	0.022[b]	**<.001**
	Peanut powder	0.007	0.009	0.004	**<.001**	0.007[a,b]	0.009[b]	0.006[a]	0.005[c]	**<.001**	0.007[a]	0.010[b]	0.004[c]	0.004[c]	0.011[b]	**<.001**
	Cake mixes	0.005	0.008	0.003	**<.001**	0.004[a]	0.005[a]	0.005[a]	0.007[b]	**<.001**	0.007[a]	0.007[a]	0.006[a]	0.003[b]	0.002[b]	**<.001**
	Instant soups	0.005	0.007	0.004	**<.001**	0.006[a,b]	0.004[c]	0.005[b]	0.007[a]	**<.001**	0.006[a]	0.004[b]	0.004[b]	0.004[b]	0.011[c]	**<.001**
	Dessert mixes	0.003	0.003	0.002	**<.001**	0.004[a]	0.003[a,b]	0.002[b]	0.003[a,b]	**.038**	0.003[a]	0.001[b]	0.002[b,c]	0.002[b,c]	0.003[a,c]	**<.001**
	Dessert sauces	0.003	0.003	0.003	.281	0.003[a,b,c]	0.003[c]	0.002[b]	0.003[a,c]	**.002**	0.004[a]	0.002[b]	0.004[a,c]	0.001[d]	0.005[c]	**<.001**
	Ready to heat dishes	0.002	0.003	0.002	**.036**	0.002[a]	0.002[a]	0.002[a]	0.003[a]	.391	0.004[a]	0.000[b]	0.000[b]	0.002[c]	0.000[d]	**<.001**

Note. *= column percentages; *P*-value = chi-square test/fisher's exact test significance level; Bold *P*-values indicate significance at 0.05 level; Others (Processed foods) = Canned Vegetables/Canned Fruits/Canned Spices; Others (Processed Culinary Ingredients) = Maple Syrup/Sugar-Molasses; Others (Unprocessed/Minimally processed foods) = Gram Flour/Carrot Flour; Each superscript letter ([a,b,c,d,e]) compares the proportion of food categories between groups with more than two categories and denotes a subset of age and location categories whose column proportions do not differ significantly from each other at the 0.05 level.

Figure 3. Purchase patterns associated with age and gender; support ≥0.02.

followed by unprocessed/minimally processed foods (35.228%) and processed culinary ingredients (10.468%) while processed foods accounted for the least purchases (0.864%).

Processed foods. Males in the age groups of 36 to 49 and above 50, especially those located in Kirinyaga, purchased a higher proportion of salted/sugared nuts and malt drinks compared to females. Meanwhile, females purchased a higher proportion of tomato paste, bacon/ham, and cheese compared to males. There was no significant gender difference in the purchases of canned pulses (dry peas, lentils, beans), processed fish, and canned corn kernels.

Processed culinary ingredients. Females in the age groups of 18 to 25 and above 50, mainly in Kirinyaga and Kajiado, purchased a higher proportion of salt and butter compared to males. Males in the age groups of 18 to 25 and above 50, mainly in Kirinyaga, purchased a higher proportion of sugar, cooking oil, cooking fat, and honey than females. There was no significant gender difference in the purchases of powdered sugar and ghee.

Unprocessed/minimally processed foods. Females above 50 years old, mainly in Nakuru, purchased a higher proportion of wheat flour, herbs and spices, fruits, pasta, fresh/fruit juices with no added sugar, vegetables, red millet flour, red meat, eggs, pulses (dry peas, lentils, beans), poultry, natural/plain yogurt, root vegetables, dried fruits, fish, and organ meat compared to males. Males in the age groups of 18 to 25 and above 50, mainly in Nairobi and Machakos, purchased a significantly higher proportion of whole milk, tea, and coffee compared to females. There was no significant gender difference in the purchases of porridge flour, milk powder, and raw nuts.

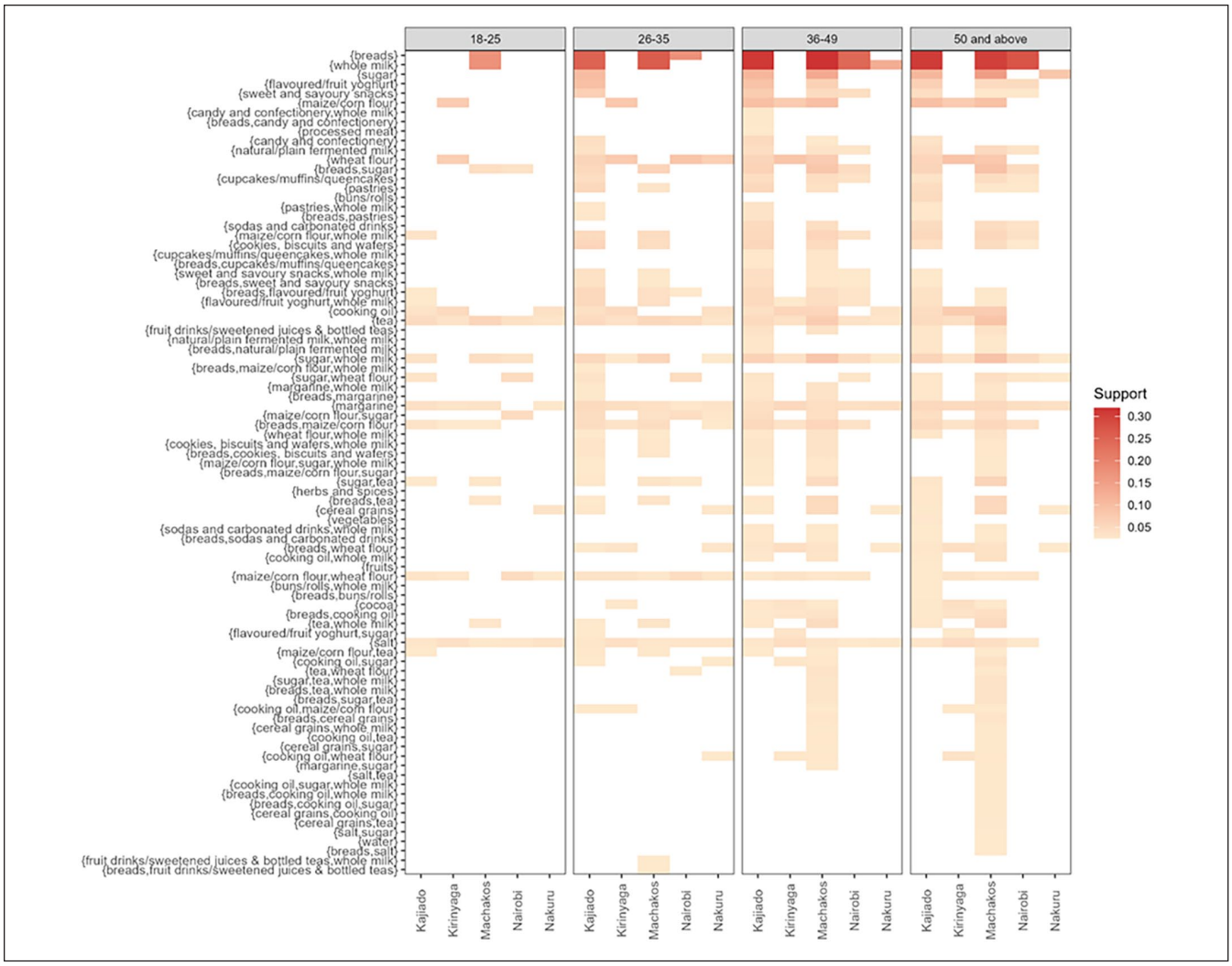

Figure 4. Purchase patterns associated with age and location; support ≥0.025.

Ultra-processed foods. Males in the age groups of 26 to 35 and above 50, mainly in Nairobi and Kajiado, purchased a higher proportion of bread, sodas, fruit drinks/sweetened juice, pastries, and energy drinks compared to females. Females in the age groups of 26 to 35 and 36 to 49, mainly in Kirinyaga, purchased a higher proportion of flavored/fruit yogurt, sweet and savory snacks, cookies, biscuits and wafers, pre-prepared food, margarine, processed meat, marble cakes, cakes, chocolate, diluting juice, noodles, table sauces, ice cream, breakfast cereals, jam, sweet rolls/buns, peanut butter, cooking sauces, processed poultry, soya chunks/mince, and chocolate spreads compared to males. There was no significant gender difference in the purchases of cocoa and malt drinks, milk drinks, flavored fermented milk, soya drinks, and dessert sauces.

Discussion

To the best of our knowledge, this is the first study in Kenya to use grocery data to examine supermarket purchasing patterns at individual level in Kenya. This is also the first study in Kenya that utilizes the NOVA food classification to examine individual food purchasing patterns.

Our results show that, generally, females bought more healthy food items (unprocessed according to NOVA classification) than males, especially females above 50 years of age. This was a trend among elderly women in Nakuru. These findings concur with those of,[23] who also found that females in Hungary tend to prefer healthy and less processed foods than males. This might be because Nakuru County is well known for farming, according to the Nakuru County Government (2021). Males who had a trend of purchasing slightly healthier foods or relatively unprocessed food items were in the age groups of 18 to 25 and above 50, mainly in Nairobi and Nakuru. The reason for this trend might be that males between 18 and 25 are not yet in the working class, and those above 50 are out of the working class; therefore, they will be at home to eat farm produce, especially those who live in Nakuru, which is a farming community. This trend concurs with the findings of,[32] who found that people

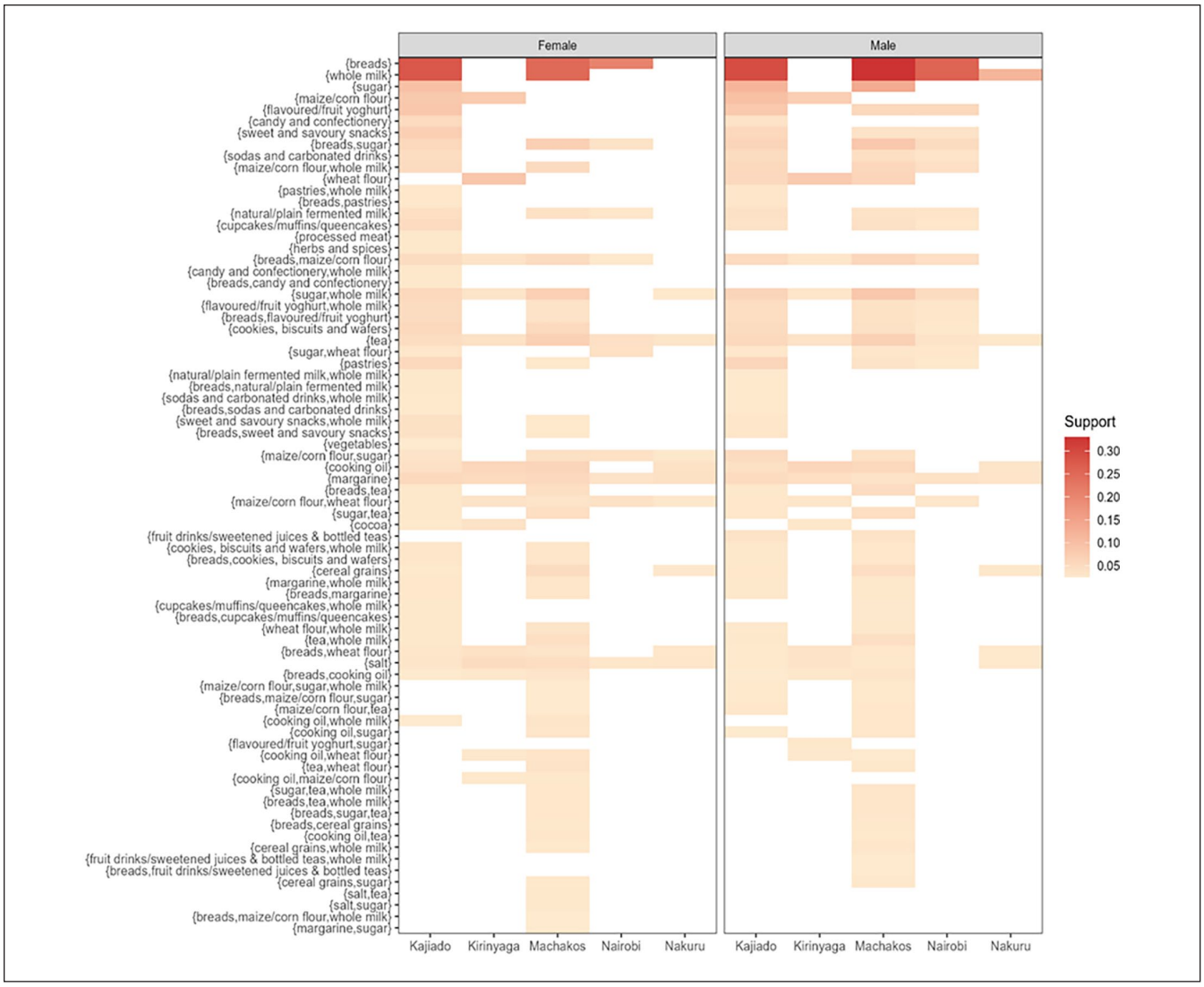

Figure 5. Purchase patterns associated with location and gender; support ≥0.025.

in the rural or peri-urban areas of Scotland tend to buy healthy foods.

In terms of industrially processed foods, males in the age groups of 26 to 35 and above 50, mainly in Nairobi and Kajiado, purchased a higher proportion of unhealthy, highly processed foods compared to females. These results concur with the findings of,[23] where they also found that males in Hungary were the main consumers of unhealthy foods. This might be because Nairobi and some parts of Kajiado, like Ngong, are urban and known for tourism.[31] found that food consumed away from home tends to be less healthy, and[32] found that households in urban settings tend to buy less healthy foods. In another investigation in South Africa,[33] found that middle-SEA shoppers in urban settings spent more on unhealthy foods. On the same note, females who were more interested in unhealthy foods were in the age groups of 26 to 35 and 36 to 49, mainly in Kirinyaga. This age group is mainly in the working class and tends to purchase food away from home, which, according to,[31] is often less healthy.

Overall, we found that consumers were purchasing more ultra-processed foods, as 53% of the food transactions were ultra-processed foods. This might be due to the fact that foods in supermarkets are, to a large extent, unhealthy. This was also found in Indonesia in an investigation conducted by.[31] Another research conducted by[35] in Brazil found that lower prices of ultra-processed foods in supermarkets contributed to their higher consumption. Research conducted in Kenya by[37] showed that supermarket purchases increased the consumption of highly processed foods at the expense of unprocessed foods.

From the results of this study, we observed that consumers in these 5 counties are generally purchasing ultra-processed foods. These foods are unhealthy, and the trend of purchasing them was especially observed among males above 50 years of age across all counties. There might be a

need to investigate why these males had these purchase trends. Research conducted by[6] showed that these foods contribute to surging rates of obesity, which are risk factors for NCDs. With this understanding we recommend that policymakers can create policies that: (1) discourage consumers from buying unhealthy foods by using technology to raise awareness about unhealthy foods; (2) encourage consumers to purchase healthy foods by reducing taxes on farm produce and all materials needed for fresh agricultural produce; (3) create affordable medical insurance for older people who are mainly affected by NCDs; and (4) promote the consumption of healthy foods in the fast food and tourism industries.

Our study had some limitations. First, the study focused on shoppers in a tier 2 supermarket with 10 branches and did not consider alternative traditional food purchasing outlets such as kiosks, open-air markets, and street vendors. Second, the grocery data was drawn from 5 counties out of the 47 counties in Kenya; therefore, the data might not be representative of other regions in Kenya.

Conclusion

This study focused on gaining valuable, hidden and previously unknown insights from transactional records of 10 supermarkets in 5 counties in Kenya. The main aim was to determine the relationship between demographic differences in supermarket purchasing patterns in LMICs like Kenya and the rise of NCDs. While there have been some studies across the world looking into similar topics, to our knowledge, this is the first study in Kenya to examine real supermarket transactions using individual purchase records.

General results showed that males aged above 50 in all counties were purchasing unhealthy foods more than females of the same age, which might lead to a higher prevalence of NCDs among this demographic. Additionally, the results indicated that 53% of the food purchase transactions were for ultra-processed (industrially processed) foods, which other research has linked to NCDs.

In light of these findings, we recommend that policymakers implement measures to: (1) discourage consumers from buying unhealthy foods by raising taxes on these items in urban areas or by using technology to raise awareness about unhealthy foods; (2) encourage consumers to purchase healthy foods by reducing taxes on farm produce and materials needed for agricultural production; (3) create affordable medical insurance for older people who are primarily affected by NCDs; and (4) promote the consumption of healthy foods in the fast food and tourism industries. Lastly, there is a need for targeted policies such as sugar-sweetened beverage taxation to curb consumption of unhealthy processed foods.

Acknowledgments

The authors would like to thank the African Population and Health Research Center for financial support and the supermarkets for providing the data.

Authors' contributions

RM - Data management, data analysis, and drafting of the paper; SC - conceptualization and oversight of the study; AK - conceptualization and oversight of the study; HOO - Literature review; MN - ethical issues and critique of the manuscript; GA - critique of the manuscript; TDK - drafting of the paper.

Availability of data and material

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Source code available from github.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical considerations

This study was approved by the Amref Ethics and Review Committee (ESRC; approval no. AMREF-ESRC P1526/2023) on November 15, 2023.

Consent to participate

Not applicable

Consent for publication

Not applicable

ORCID iDs

Reinpeter Momanyi https://orcid.org/0009-0002-9059-393X
Steve Bicko Cygu https://orcid.org/0000-0002-9284-8863
Agnes Kiragga https://orcid.org/0000-0002-0969-0699
Henry Owoko Odero https://orcid.org/0000-0002-3678-4332
Maureen Ng'etich https://orcid.org/0009-0007-4945-4635
Gershim Asiki https://orcid.org/0000-0002-9966-1153
Tatenda Duncan Kavu https://orcid.org/0000-0002-9479-1143

References

1. World Health Organization. *Noncommunicable diseases progress monitor 2022*. World Health Organization 2022. Licence: CC BY-NC-SA 3.0 IGO; 2022.
2. Oikonomou E, Psaltopoulou T, Georgiopoulos G, et al. Western dietary pattern is associated with severe coronary artery disease. *Angiology*. 2018;69(4):339-346. doi:10.1177/0003319717721603
3. Shu L, Shen XM, Li C, Zhang XY, Zheng PF. Dietary patterns are associated with type 2 diabetes mellitus among middle-aged adults in Zhejiang Province, China. *Nutr J*. 2017;16(1):81-89. doi:10.1186/s12937-017-0303-0
4. Sun J, Buys NJ, Hills AP. Dietary pattern and its association with the prevalence of obesity, hypertension and other cardio-

vascular risk factors among Chinese older adults. *Int J Environ Res Public Health*. 2014;11(4):3956-3971. doi:10.3390/ijerph110403956
5. Bodirsky BL, Dietrich JP, Martinelli E, et al. The ongoing nutrition transition thwarts long-term targets for food security, public health and environmental protection. *Sci Rep*. 2020;10(1):19778. doi:10.1038/s41598-020-75213-3
6. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev*. 2012;70(1):3-21. doi:10.1111/j.1753-4887.2011.00456.x
7. KNBS and ICF. *Kenya demographic and health survey 2022*. Kenya National Bureau of Statistics (KNBS) and ICF; 2023.
8. Monteiro CA, Levy RB, Claro RM, de Castro IR, Cannon G. Increasing consumption of ultra-processed foods and likely impact on human health: evidence from Brazil. *Public Health Nutr*. 2010;14(1):5-13. doi:10.1017/S1368980010003241
9. Canella DS, Levy RB, Martins AP, et al. Ultra-processed food products and obesity in Brazilian households (2008-2009). *PLoS O*ne. 2014;9(3):e92752. doi:10.1371/journal.pone.0092752
10. Cediel G, Reyes M, da Costa Louzada ML, et al. Ultra-processed foods and added sugars in the Chilean diet (2010). *Public Health Nutr*. 2018;21(1):125-133. doi:10.1017/S1368980017001161
11. Crovetto MM, Uauy R, Martins AP, Moubarac JC, Monteiro C. Consume food and drink products in Chile: impact on nutritional quality of the diet. *Rev Med Chil*. 2014;142(7):850-858.
12. Louzada ML, Baraldi LG, Steele EM, et al. Consumption of ultra-processed foods and obesity in Brazilian adolescents and adults. *Prev Med*. 2015;81:9-15. doi:10.1016/j.ypmed.2015.07.018
13. Bhurosy T, Kaschalk E, Smiley A, He K. Comment on "ultra-processed food consumption and risk of overweight and obesity: the University of Navarra Follow-Up (SUN) cohort study". *Am J Clin Nutr*. 2017;105(4):1012. doi:10.3945/ajcn.116.149302
14. Juul F, Hemmingsson E. Trends in consumption of ultra-processed foods and obesity in Sweden between 1960 and 2010. *Public Health Nutr*. 2015;18(17):3096-3107. doi:10.1017/S1368980015000506
15. Marrón-Ponce JA, Sánchez-Pimienta TG, Louzada MLDC, Batis C. Energy contribution of NOVA food groups and sociodemographic determinants of ultra-processed food consumption in the Mexican population. *Public Health Nutr*. 2018;21(1):87-93. doi:10.1017/S1368980017002129
16. Mendonça RDD, Lopes ACS, Pimenta AM, Gea A, Martinez-Gonzalez MA, Bes-Rastrollo M. Ultra-processed food consumption and the incidence of hypertension in a mediterranean cohort: the seguimiento universidad de navarra project. *Am J Hypertens*. 2017;30(4):358-366. doi:10.1093/ajh/hpw137
17. Moreira PV, Baraldi LG, Moubarac JC, et al. Comparing different policy scenarios to reduce the consumption of ultra-processed foods in UK: impact on cardiovascular disease mortality using a modelling approach. *PLoS One*. 2015;10(2):e0118353. doi:10.1371/journal.pone.0118353
18. Moubarac JC, Batal M, Louzada ML, Martinez Steele E, Monteiro CA. Consumption of ultra-processed foods predicts diet quality in Canada. *Appetite*. 2017;108:512-520. doi:10.1016/j.appet.2016.11.006
19. Moubarac JC, Martins AP, Claro RM, Levy RB, Cannon G, Monteiro CA. Consumption of ultra-processed foods and likely impact on human health. Evidence from Canada. *Public Health Nutr*. 2013;16(12):2240-2248. doi:10.1017/S1368980012005009
20. Solberg SL, Terragni L, Granheim SI. Ultra-processed food purchases in Norway: a quantitative study on a representative sample of food retailers. *Public Health Nutr*. 2016;19(11):1990-2001. doi:10.1017/S1368980015003523
21. Martínez Steele E, Baraldi LG, Louzada ML, Moubarac JC, Mozaffarian D, Monteiro CA. Ultra-processed foods and added sugars in the US diet: evidence from a nationally representative cross-sectional study. *BMJ Open*. 2016;6(3):e009892. doi:10.1136/bmjopen-2015-009892
22. Fagerli RA, Wandel M. Gender differences in opinions and practices with regard to a "Healthy Diet". *Appetite*. 1999;32(2):171-190. doi:10.1006/appe.1998.0188
23. Koroknay Z, Kovács S, Pfau C. Gender differences in consumption habits and spending behaviour regarding food groups in one of the most obese countries in Europe. *Phys Educ Sport*. 2021;21(2):791-796. doi:10.7752/jpes.2021.02098
24. Leblanc V, Bégin C, Corneau L, Dodin S, Lemieux S. Gender differences in dietary intakes: WHAT is the contribution of motivational variables? *J Hum Nutr Diet*. 2015;28(1):37-46. doi:10.1111/jhn.12213
25. Li KK, Concepcion RY, Lee H, et al. An examination of sex differences in relation to the eating habits and nutrient intakes of university students. *J Nutr Educ Behav*. 2012;44(3):246-250. doi:10.1016/j.jneb.2010.10.002
26. Salameh P, Jomaa L, Issa C, et al.; Lebanese national conference for health in university research group. Assessment of dietary intake patterns and their correlates among university students in Lebanon. *Front Public Health*. 2014;2:185. doi:10.3389/fpubh.2014.00185.
27. Emrouznejad A, Marra M. Big data: who, what and where? Social, cognitive and journals map of big data publications with focus on optimization. *Stud Big Data*. 2016;18(3):1-16. doi:10.1007/978-3-319-30265-2_1
28. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? *Int J Behav Nutr Phys Act*. 2017;14(1):46-10. doi:10.1186/s12966-017-0502-2
29. Reardon T, Timmer CP. The economics of the food system revolution. *Annu Rev Resour Econ*. 2012;4(1):225-264. doi:10.1146/annurev.resource.050708.144147
30. Popkin BM. The nutrition transition: an overview of world patterns of change. *Nutr Rev*. 2004;62(7 II):S140-143. doi:10.1111/j.1753-4887.2004.tb00084.x
31. Toiba H, Umberger WJ, Minot N. Diet transition and supermarket shopping behaviour: is there a link? *Bull Indones Econ Stud*. 2015;51(3):389-403. doi:10.1080/00074918.2015.1111997
32. De Roos B, Binacchi F, Whybrow S, Sneddon AA. Differences in expenditure and amounts of fresh foods, fruits and vegetables, and fish purchased in urban and rural Scotland. *Public Health Nutr*. 2017;20(3):524-533. doi:10.1017/S1368980016002688

33. Odunitan-Wayas FA, Okop KJ, Dover RV, et al. Food purchasing behaviour of shoppers from different South African socio-economic communities: results from grocery receipts, intercept surveys and in-supermarkets audits. *Public Health Nutr*. 2020;24(4):1-676. doi:10.1017/S1368980020001275
34. Moodley G, Christofides N, Norris SA, Achia T, Hofman KJ. Obesogenic environments: access to and advertising of sugar-sweetened beverages in Soweto, South Africa, 2013. *Prev Chronic Dis*. 2015;12(10):E186. doi:10.5888/pcd12.140559
35. Machado PP, Claro RM, Canella DS, Sarti FM, Levy RB. Price and convenience: the influence of supermarkets on consumption of ultra-processed foods and beverages in Brazil. *Appetite*. 2017;116:381-388. doi:10.1016/j.appet.2017.05.027
36. Debela BL, Demmler KM, Klasen S, Qaim M. Supermarket food purchases and child nutrition in Kenya. *Glob Food Secur*. 2020;25:100341. doi:10.1016/j.gfs.2019.100341
37. Rischke R, Kimenju SC, Klasen S, Qaim M. Supermarkets and food consumption patterns: the case of small towns in Kenya. *Food Policy*. 2015;52:9-21. doi:10.1016/j.foodpol.2015.02.001
38. Lachat C, Hawwash D, Ocké MC, et al. Strengthening the eporting of observational Studies in pidemiology – nutritional epidemiology (STROBE -nut): an extension of the STROBE statement. *Nutr Bull*. 2016;41(3):240-251. doi:10.1111/nbu.12217
39. Monteiro CA, Cannon G, Moubarac JC, Levy RB, Louzada MLC, Jaime PC. The un Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutr*. 2018;21(1):5-17. doi:10.1017/S1368980017000234
40. Monteiro CA, Levy RB, Claro RM, Castro IR, Cannon G. A new classification of foods based on the extent and purpose of their processing. *Cad Saúde Pública*. 2010;26(11):2039-2049. doi:10.1590/s0102-311x2010001100005
41. The PostgreSQL Global Development Group. PostgreSQL: the world's most advanced open source relational database. Published online February, 2023.
42. R Core Team. R: A language and environment for statistical computing. Published online March, 2024.