<u>Original Paper</u>

Tensorial Principal Component Analysis in Detecting Temporal Trajectories of Purchase Patterns in Loyalty Card Data: Retrospective Cohort Study

Reija Autio[1], PhD; Joni Virta[2], PhD; Klaus Nordhausen[3], PhD; Mikael Fogelholm[4], SCD; Maijaliisa Erkkola[4], PhD; Jaakko Nevalainen[1], PhD

[1]Faculty of Social Sciences (Health Sciences), Tampere University, Tampere, Finland
[2]Department of Mathematics and Statistics, University of Turku, Turku, Finland
[3]Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland
[4]Department of Food and Nutrition, University of Helsinki, Helsinki, Finland

Corresponding Author:
Reija Autio, PhD
Faculty of Social Sciences (Health Sciences)
Tampere University
P.O. Box 100
Tampere, FI-33014
Finland
Phone: 358 50 318 7364
Email: reija.autio@tuni.fi

Abstract

Background: Loyalty card data automatically collected by retailers provide an excellent source for evaluating health-related purchase behavior of customers. The data comprise information on every grocery purchase, including expenditures on product groups and the time of purchase for each customer. Such data where customers have an expenditure value for every product group for each time can be formulated as 3D tensorial data.

Objective: This study aimed to use the modern tensorial principal component analysis (PCA) method to uncover the characteristics of health-related purchase patterns from loyalty card data. Another aim was to identify card holders with distinct purchase patterns. We also considered the interpretation, advantages, and challenges of tensorial PCA compared with standard PCA.

Methods: Loyalty card program members from the largest retailer in Finland were invited to participate in this study. Our LoCard data consist of the purchases of 7251 card holders who consented to the use of their data from the year 2016. The purchases were reclassified into 55 product groups and aggregated across 52 weeks. The data were then analyzed using tensorial PCA, allowing us to effectively reduce the time and product group-wise dimensions simultaneously. The augmentation method was used for selecting the suitable number of principal components for the analysis.

Results: Using tensorial PCA, we were able to systematically search for typical food purchasing patterns across time and product groups as well as detect different purchasing behaviors across groups of card holders. For example, we identified customers who purchased large amounts of meat products and separated them further into groups based on time profiles, that is, customers whose purchases of meat remained stable, increased, or decreased throughout the year or varied between seasons of the year.

Conclusions: Using tensorial PCA, we can effectively examine customers' purchasing behavior in more detail than with traditional methods because it can handle time and product group dimensions simultaneously. When interpreting the results, both time and product dimensions must be considered. In further analyses, these time and product groups can be directly associated with additional consumer characteristics such as socioeconomic and demographic predictors of dietary patterns. In addition, they can be linked to external factors that impact grocery purchases such as inflation and unexpected pandemics. This enables us to identify what types of people have specific purchasing patterns, which can help in the development of ways in which consumers can be steered toward making healthier food choices.

(J Med Internet Res 2023;25:e44599) doi: 10.2196/44599

Introduction

Loyalty card data comprise a big data source [1], which is becoming increasingly important in research [2-5]. Loyalty cards are electronic customer cards used in grocery retailing that automatically register grocery expenditure per purchased item every time the customer swipes their card at the store. Card holder–specific loyalty card data should not be confused with aggregated data such as store-specific point of sales data; instead, the former provides information on both the product group and the time of each customer's purchases. The level of details in product groups (eg, amount and price) and time within loyalty card data make it an attractive source of information in, for instance, public health research [2]. Loyalty card data provide new insights into health-related purchase behavior [6-10]. Food, tobacco, and alcohol are well-known risk factors for many chronic diseases; however, their consumption is difficult to measure using traditional methods [11,12]. Using loyalty card data, the purchases of these items can be measured in an automated, detailed, and objective fashion.

Loyalty card data are structured in such a way that individual card holders, expenditures, and the volume of product groups purchased within a single period constitute the usual data matrix. However, when the data are arranged by time of purchase, they are longitudinal, and the third dimension is time, including information on, for example, time trends, seasonality, or change points. Therefore, loyalty card data are longitudinal data, having dimensions in both time and product groups, and can be seen as a tensor.

Tensorial data are multidimensional array data. Vectors are 1D or first-order tensors; matrices are 2D or second-order tensors; and more generally, tensors are generalization of matrices to the n-dimensional space. Tensorial data, images, and videos, for example, are becoming increasingly common. In addition, our loyalty card data resemble a video, which is set up as a dense sequence of images, each detailing the expenditures by product group for all participants during a fixed time. Simultaneously, methods for the analysis of tensor-valued data have developed substantially [13,14].

Although modern measurement systems have increased the dimensionality of data, these dimensions usually include a lot of redundant information within the data. Therefore, in data analysis, often the aim is to decrease the number of dimensions by removing the redundancy without losing the substantial information within the data. Dimension reduction has become a standard tool in exploratory statistics, with the dual goal of reducing the number of variables in a data set and extracting easy-to-interpret latent components for further analysis. Since the introduction of the standard principal component analysis (PCA), a search for uncorrelated subspaces containing a maximal amount of variation [15,16], a vast number of dimension reduction methods for different types of high-dimensional data have been proposed. Dimension reduction

for longitudinal data has been studied elsewhere [17] using multivariate functional PCA methodology.

Traditionally, research on the health effects of diet has primarily focused on examining the associations between individual nutrients, foods, or food groups and their impact on health outcomes. It is likely that this was related to the earlier focus on nutrient-specific deficiencies [18]. However, as individuals consume foods and beverages in various combinations, there has been a shift in epidemiological studies toward investigating dietary patterns [19-21]. A dietary pattern can be described as the typical quantities, proportions, variety, or combination of foods and drinks that an individual consumes. The dietary pattern analysis approach offers advantages, as nutrients present in food can confound or interact with each other. In addition, pattern analysis can enable the detection of associations between diet and health outcomes because the combined effect of an entire diet may be more powerful than the effects of its individual components [20]. However, identifying data-driven dietary patterns relies heavily on subjective and arbitrary decisions regarding the grouping of food items and the labeling of dietary patterns. Previous studies have mainly derived dietary patterns from self-reported data [22]. In this study, our hypothesis is that by using tensorial PCA, we will be able to explore food purchase patterns from loyalty card data more comprehensively than before. This presumption is based on the ability to simultaneously detect patterns in both the product group dimension and the temporal trajectory enabled by tensorial PCA, yielding a deeper understanding of product group preferences, purchasing behavior, and temporal patterns. Tensorial PCA has been used in various data domains such as functional magnetic resonance imaging (fMRI) data [23], image and facial recognition [24,25], videos [26], and financial time series prediction [27]; however, our study is the first to focus on retailer data.

Thus, we provide a case study on tensorial PCA that allows us to identify temporal characteristics of health-related purchase behavior from longitudinal loyalty card (LoCard) data. By implementing modern tensorial variants of PCA on high-resolution data, we systematically searched for typical food purchasing patterns across time and product groups, atypical behavior across product groups and time, and differing purchasing behaviors across groups of cardholders. We showcase and discuss the interpretation, benefits, and challenges of these methods compared with their standard counterparts.

This study contributes to the interdisciplinary fields of biostatistics, human nutrition, public health, and digital health data science. The overall purpose of this methodological study is to show how tensorial PCA enables a comprehensive analysis; to demonstrate how this analysis can be used for retailer data; and, subsequently, to interpret the findings of dietary pattern analysis.

Methods

Materials: LoCard Data

Setting and Participants

Loyalty card data by the S Group, a major Finnish retailer cooperative, disclosed the background characteristics of the consenting card holders (age, gender, and residential postal code) and grocery expenditure data from the year 2016. Consenting card holders were from Helsinki and 9 nearby municipalities. The full details of the data collection process are described by Nevalainen et al [2]. The initial data on expenditure consisted of the purchases of 143 product groups of 14,595 consenting loyalty card holders, who were the primary card holders of their households.

Exclusion Criteria

To analyze the purchase behavior of regular customers, we excluded personnel members (1962 card holders) and card holders who appeared to be frequently absent or to conduct a substantial amount of their grocery purchases elsewhere, defined as having: (1) >8 weeks, approximately 15% of all weeks, with no registered purchases at all (5919 card holders excluded) or (2) total expenditure of <€500 (US $547.3) per year (1331 card holders excluded). On average, a Finnish household annually used €4381 (US $4796) [28]; thus, €500 (US $547.3) is >10% of the average annual food expenditure and approximately 10 euros per week. Some of the card holders were excluded owing to more than one criterion, and after all exclusions, the final data set comprised 7251 card holders. The distributions of the demographic variables of those included and excluded were similar (Figure S1 in Multimedia Appendix 1 [23,24,29,30]).

Expenditure Variables

The initial grouping of foods provided by the retailer was based not only on ingredients but also on the package form and placement in the shelf system in stores. A nutrition researcher then regrouped the food, tobacco, and alcohol product groups into 76 groups relevant to nutrition and health research. Finally, product groups with minimal importance and expenditure, such as meal ingredients, were omitted, and the final data comprised 55 scientifically interpretable food, tobacco, and alcohol product groups. We used several criteria in our regrouping process: food price and status, connections to lifestyle or life stage, occasion of use, and wholesomeness. For instance, Tex-mex products were kept as a separate group because their consumption may be associated with special occasions such as parties or social evenings. The food group variables were comprehensive and nonoverlapping [10]. The variables were initially expressed in expenditure, that is, in euros spent on them. To give equal weighting to every card holder and to aid the interpretation of expenditure, we rescaled the expenditures in such a way that the annual expenditure for each card holder sums up to €1000. Thus, for each card holder, we observed standardized expenditure variables measured weekly, which can be interpreted as how many euros out of the €1000 were spent on product groups such as tobacco, vegetables, and meat. This scaling was necessary because the available data pertain to households of different sizes, structures, and shares of purchases

from S Group rather than individual food consumption. By rescaling the expenditures, we enabled meaningful comparisons and analysis of purchase composition across different card holders. The resulting dimension of the data tensor was 7251×55×52 (card holders, product groups, and weeks). The product groups and their labels are presented in Table S1 in Multimedia Appendix 1.

Statistical Methods

Principal Component Analysis

PCA has become a standard multivariate tool in dimension reduction [29,31]. With PCA, the aim is to reduce the original number of dimensions (ie, p) to a smaller number of derived variables (ie, k) that are linear combinations of the original variables such that "no information is lost" in terms of preserving as much of the variation as possible.

PCA for 3D Data

To provide a contrast to our proposed approach, we briefly expand on how PCA can be used to extract latent information from a horizontal cross-section of the LoCard data. Let the data matrix of $\mathbf{X} \in \mathbf{R}^{n \times p}$ contain the aggregated purchase data of n customers over p items during, for example, a single fixed week. The column-centered data matrix is denoted by $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{1}_n \bar{x}^T$, where $\mathbf{1}_n \in \mathbf{R}^n$ is a vector full of ones, and $\bar{x}$ contains the means of p items over n customers. Let the columns of $\mathbf{U} \in \mathbf{R}^{p \times k}$ contain the eigenvectors of the sample covariance matrix

$$ S = \frac{1}{n-1} \bar{\mathbf{X}}^T \bar{\mathbf{X}} $$

associated with its k (a user-chosen parameter) largest eigenvalues. The k columns of $\bar{\mathbf{X}}\mathbf{U}$ are known as the principal components, and they are usually expected to contain "hidden" information that is not discernible in the visualization of the original data. Moreover, the columns of $\mathbf{U}$ ("loadings") can be used to interpret the components in terms of the original p variables. This type of PCA analysis could be undertaken for the LoCard data for a time-aggregated version of the data or for a single time occasion.

However, the full data are expressed as a 3D object $\mathbf{X} \in \mathbf{R}^{n \times p \times t}$, where the additional dimension corresponds to the t time points during which the purchase history of each customer is recorded. The total data set (or data tensor) is most conveniently visualized as longitudinal data with a data matrix, $\mathbf{Y}_i = \mathbf{X}_{i,:,:} \in \mathbf{R}^{p \times t}$, $i=1,...,n$ of purchase history associated with each of the n customers. A standard way of applying PCA to data with such a structure is to vectorize the matrices $\mathbf{Y}_i$ into long vectors $vec(\mathbf{Y}_i) \in \mathbf{R}^{pt}$ by stacking their columns. PCA can then be applied to the vectorization data matrix $(vec(\mathbf{Y}_i),...,vec(\mathbf{Y}_n))^T$. However, this approach compromises the data structure by mixing the time and item dimensions, making the interpretation of the resulting principal components needlessly complicated. For example, in our data, having 55 product groups purchased at 52 weeks by customers would mean that the loading vectors have 2860

elements to interpret, making it very difficult to understand what they represent.

Structure-Preserving Dimension Reduction

A more preferable goal is to keep the 2 dimensions, product groups and time, separate and reduce their sizes individually. This approach is adopted in *tensorial dimension reduction*, a methodology aimed at reducing the dimensionalities of data sets consisting of observations of higher order than ordinary vectors such as matrices [32]. An extension of PCA in this case is higher-order singular value decomposition, known as higher-order singular value decomposition [30] and later rediscovered under the names 2-directional 2-dimensional PCA [24] and tensorial PCA [23], depending on the context, which also begins by centering the observed item-time matrices over the following sample:

$$\overline{\mathbf{Y}}_i = \mathbf{Y}_i - \frac{1}{n}\sum_{j=1}^{n} \mathbf{Y}_j , i = 1, \dots, n$$

The next step is to compute the *modal covariance matrices* of product groups and time separately, as follows:

$$\mathbf{S_1} = \frac{1}{n-1}\sum_{i=1}^{n} \overline{\mathbf{Y}}_i \overline{\mathbf{Y}}_i^{\mathbf{T}}, \mathbf{S_2} = \frac{1}{n-1}\sum_{i=1}^{n} \overline{\mathbf{Y}}_i^{\mathbf{T}}\overline{\mathbf{Y}}_i.$$

The interpretation of $\mathbf{S}_1$ is that it measures linear dependency among items while ignoring (or aggregating over) the time space, and vice versa for $\mathbf{S}_2$. To obtain the principal components, we consider the first p_0 eigenvectors (principal item directions) $\mathbf{U} \in \mathbf{R}^{p \times p_0}$ of $\mathbf{S}_1$ and the first t_0 eigenvectors (principal time directions) $\mathbf{V} \in \mathbf{R}^{t \times t_0}$ of $\mathbf{S}_2$, where p is the number of product groups, t is the number of weeks, and p_0 and t_0 are user-specified parameters for the number of principal components to be selected for further analysis. To aid in choosing p_0 and t_0, the eigenvalues of $\mathbf{S}_1$ and $\mathbf{S}_2$ can be plotted as in PCA to obtain scree plots. In such scree plots, one assumes that the last $p - p_0$ and $t - t_0$ eigenvalues are equal and one searches an "elbow." To aid the choice, information on eigenvector variation can be incorporated using resampling methods [33,34], where, for example, in the augmentation approach, the idea is to augment the data tensor mode wise. The criterion is then a weighted sum of the eigenvalues and eigenvector variation [35,36]. Here, we have used this augmentation method for selecting the suitable number of principal components for further analysis. The augmentation estimator is a nonheuristic method that guarantees the ability to estimate the dimension correctly under mild conditions [36].

The individual matrices of the principal components for each customer are then obtained as projections $\mathbf{z}_i = \mathbf{U}^{\mathbf{T}}\overline{\mathbf{Y}}_i\mathbf{V}$. We note that although the dimension reduction is performed separately for the item and time spaces, the method still produces a single matrix of principal components for each customer, in which each component is related to a time-space direction pair. An example of interpreting the tensorial PCA components and directions is provided in Multimedia Appendix 1 (Interpretation of the Tensorial PCA Components and Directions).

The main assumption of tensorial PCA is that the variation in data can be decomposed into row and column variations. Typically, this is expressed mathematically by requiring that the observed matrices $\mathbf{Y}_i$ admit representations $\mathbf{Y}_i = \mathbf{A}\mathbf{Z}_i\mathbf{B}^{\mathbf{T}}$, where the latent matrix $\mathbf{Z}_i$ has uncorrelated components, and $\mathbf{A}$ and $\mathbf{B}$ are orthogonal matrices of suitable sizes [36].

Interpretation of the Scores and Loadings Resulting From Tensorial PCA

Tensorial PCA, similar to standard PCA, transforms data from its original coordinates (variables) into a new coordinate system, where each coordinate, called principal component scores, corresponds to a linear combination of the original variables weighted by the estimated loadings. The scores are the new set of coordinates, where the first principal component captures most of the variability in the data, followed by the second principal component, and so on. Tensorial PCA returns these scores for each product group and week pair. In addition, tensorial PCA provides loadings representing the associations between each product group and the principal component (and each week and the principal component), reflecting the covariance between the component scores and the observations. When interpreting the loadings, the emphasis is on the magnitude of each loading. Variables with larger absolute loadings have a stronger influence on the principal component. The sign of the loading merely indicates the direction of the relationship (positive or negative correlation) between the variable and the component, both being equally interesting.

Illustration and Software Packages

To enhance the recognition of product groups and weeks exhibiting similar patterns, we used heat maps and hierarchical clustering of loadings (with correlation distance and average linkage) to illustrate purchase patterns. Hierarchical clustering is a standard clustering method that facilitates the visualization of a large number of loadings [37]. All analyses were performed with R (version 4.0.1; R Core Team) [38] and using packages *tensorBSS* [39], *gplots* [40], and *ggplot2* [41].

Ethical Considerations

This study was approved by the University of Helsinki Review Board in the Humanities and Social and Behavioral Sciences (statement 43/2016). Before inclusion in the study, all participants were invited to participate via email and provided informed consent electronically. They were asked to release their loyalty card data. To ensure privacy, the data were pseudonymized by S Group before the researchers obtained the data.

Results

Descriptive Analysis of the Data

We begin by reporting simple summaries and illustrations of the raw rata. When considering the product groups, the highest median expenditure was for cheese, which constituted together

7% of all food purchases. The customer-wise maximum expenditures were on beer (€882 [US $961.4]), cigarettes (€894 [US $974.5]), and wine and cider (€937 [US $1021.3]), and for all customers combined, of all purchases, these product groups constituted shares of 4.6%, 4.2%, and 1.5%, respectively (Figure 1; Table S1 in Multimedia Appendix 1). In addition, purchases of many product groups, such as cigarettes and alcohol, increased for some customers, whereas many others did not purchase these items at all (Figure 1; Table S1 in Multimedia Appendix 1).

There were also obvious differences in the purchase patterns based on time. If all purchases were distributed evenly across weeks, then each week would constitute 1.9% of the yearly purchases. However, exceptional weeks, such as week 51 (Christmas), constituted 2.4% of all purchases, which was 26% more than the average week. Similarly, weeks 25 (Midsummer)

and 12 (Easter) constituted 2.2% each, which was 16% more than purchases in an average week (Figure 2).

In addition, some of the product groups had a pattern that was related to the time of the year (Figure 2). For many product groups, the pattern was steady or there was only little weekly variation in the purchase behavior of customers. For some product groups, there were clear seasonal patterns during summer or winter and holiday times, that is, Christmas, Midsummer, and Easter, which can easily be distinguished from the weekly plot (Figure 2). For instance, the expenditure on beer tended to rise toward the summer season compared with other times of the year and to peak at Midsummer. Sweets and chocolate purchases clearly increased during Christmas and Easter, and pig and bovine meat had a similar trend, whereas mutton purchases increased clearly only during Easter.

Figure 1. Money spent (per €1000) by product group across customers and average purchase basket of participants (y-axis is the percentage of money spent on each product group). Only the product groups covering >1% of all purchases are illustrated. Thus, altogether 27 product groups, that is, 11.01% of all purchases were omitted.
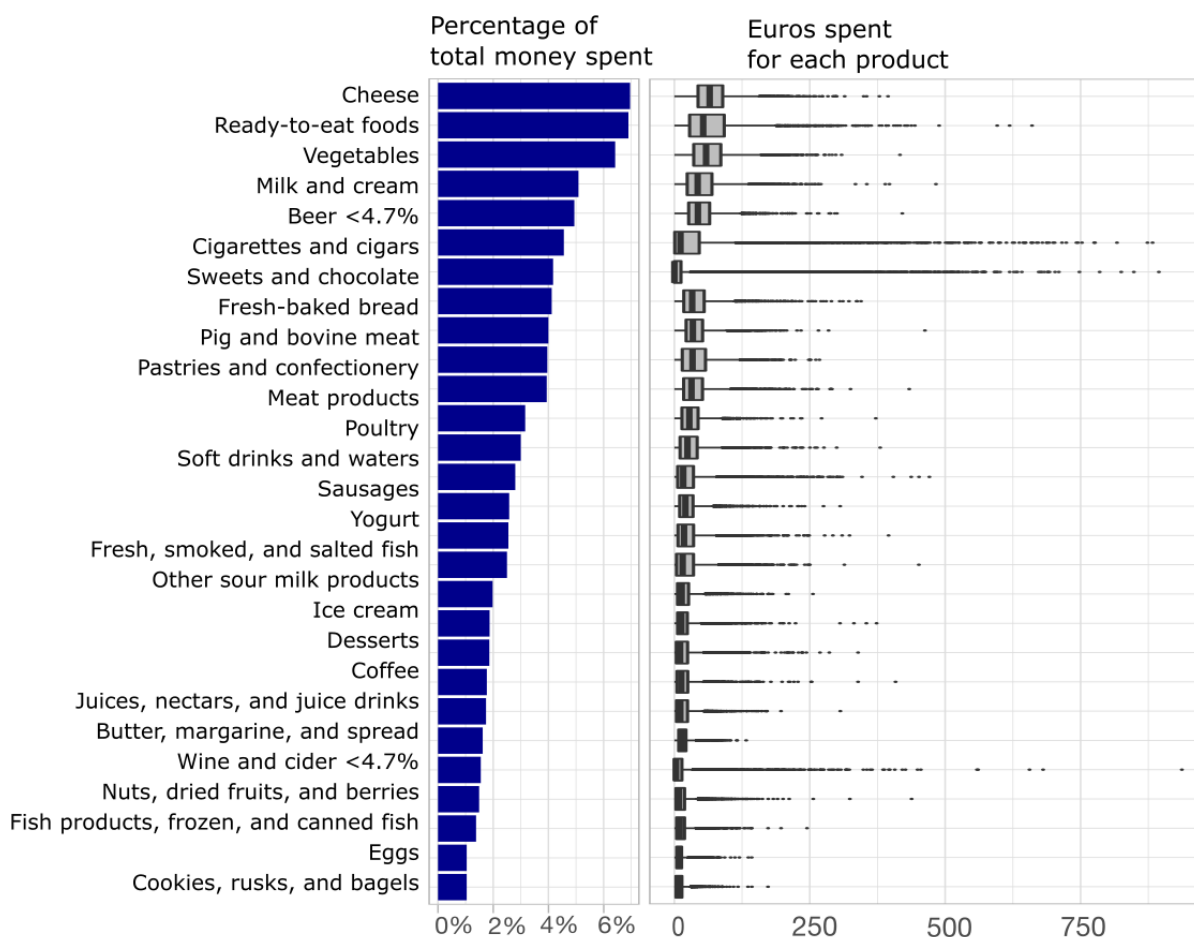
Figure 2. The upper line graph illustrates the percentage of purchases made across the weeks. Dashed red line represents 1.9%, which would be the weekly average if all purchases were distributed evenly across the year. Heat map shows weekly purchase pattern illustrated for total sum of the money spent on each product group (rows). The color indicates the row-wise z scores of each product group. Holiday weeks 12, 25, and 51 clearly stand out from the analysis. Simultaneously, the figure illustrates the sum patterns for the product groups, showing that some of the product groups are more often purchased during summer (eg, beer, wine, and cider), whereas others are purchased more during winter (eg, frozen fruits and frozen vegetables). Summed data are clustered with correlation distance and average linkage.

Joint Analysis of Time and Food Purchase Patterns

Using descriptive statistics and cluster analysis of product group expenditures over time, we were able to detect general patterns in the data. By using tensorial PCA, we could learn more from the data by taking the product group and time information into account simultaneously and could gain more insight into interindividual differences.

First, we analyzed the correlation structure of the product groups and time in Multimedia Appendix 1 (Correlation Structures of Products and Time). Especially weeks next to each other correlated heavily, which is a sign of serial correlation (Figure S2 in Multimedia Appendix 1). At the same time, the holiday weeks of Christmas (week 51), Easter (week 12), and Midsummer (week 25) stand out because of their understandably different purchase behavior; these may include different product groups and amounts compared with everyday life.

Tensorial PCA also allows outlier detection to be conducted simultaneously based on several dimensions [23,24,30]. Here, we used it for identifying atypicalities within the data for both time and product group dimensions, Multimedia Appendix 1 (Detecting Atypicalities). We revealed outstanding patterns

during different seasons and holiday times, that is, weeks 12, 25, and 51, as well as patterns indicating dominant product groups, that is, beer, cigarettes, and wine and cider (Figure S3 in Multimedia Appendix 1).

Longer-term health behavior is generally more relevant than occasional behavior during holidays or other special occasions. Therefore, we wanted to assess longer-term health behavior by purchase patterns, focusing on periods other than the holiday seasons. The health risks of beer and alcohol have been well documented [42-45], and we felt that here the focus on food purchasing behavior is more insightful without their inclusion. Therefore, to analyze the purchase (dietary) patterns of everyday life, we peeled outlying data from the data entity by removing the holiday weeks as well as the beer, cigarette, and wine and cider product groups. With these filtered data, we could identify different types of purchase behavior patterns and detect groups of individuals whose purchase patterns stand out with the specific combination of time and product group.

After filtering out the holiday weeks 12, 25, and 51 as well as the purchases of beer, cigarettes, and wine and cider of the scaled data, the resulting new tensor of data is $\mathbf{X} \in \mathbf{R}^{7251 \times 52 \times 49}$. These data are now free of the most obvious outliers in the time

and product group dimensions and were used to analyze time trends and food purchase patterns. On the basis of the reanalysis with tensorial PCA and using the augmentation method, we identified 18 product group principal components, explaining 81.8% of the product group variation, and 6 principal components for the time dimension, explaining 34.7% of the time variation (Figure 3; Figure S4 in Multimedia Appendix 1).

Overall, tensorial PCA of the filtered data was easier to read. The first week-based principal component shows the average purchase pattern over the year (Figure 3A). This illustrates the average food expenditure per week, meaning that the largest source of variation for the week dimension is the general level of expenditure. The second component illustrates cases in which the purchase pattern differs between the first and second half of the year. This can be the case, for example, if a card holder encounters major changes in household structure (eg, moving in with someone or new family), has moved to another address, implemented a lifestyle change, or just changed the grocery store where most purchases were made. The third principal component shows the differential pattern of purchases during winter and summer. We refer to these as the first 3 time components: PC1—weekly average, PC2—spring versus autumn, and PC3—summer versus winter. The remaining components seem to be more detailed variants of these, indicating, for instance, school or work holidays (PC4).

On the basis of the product group components, the first component gave high loadings for multiple product groups, with most of the product groups being fairly regular items in a Finnish food basket (Figure 3B). The other components were more product group specific. The second component loads highly on ready-to-eat food (including a variety of packaged and service counters selling ready meal portions such as pasta and pizza) and contrasts it with several fresh product groups such as pig and bovine meat, vegetables, and fruits. Therefore, with the second component, we can identify card holders who buy substantial amounts of ready-to-eat food but purchase very few fresh product groups and simultaneously card holders who buy only very little ready-to-eat food and a lot of fresh product groups. Thus, with each component, we can simultaneously detect the customers at both ends. The third component loads highly for pig and bovine meat, indicating a strong preference for eating inexpensive meats. We refer to these as the first 3 product group components: PC1—product average, PC2—ready-to-eat, and PC3—red meat.

Using these loadings, we simultaneously selected the time pattern and the product group pattern to illustrate individuals who exhibit such purchase patterns. We categorized our subjects into 3 categories: the 10% with the highest scores of the principal components formed the group "high" and the 10% with the lowest scores of the principal components formed the group "low," with 80% with scores between these 2 extremes representing the group "typical." This was done separately for the scores of each pair of the week and product group. In Figure

4, we illustrate the purchase pattern of the high and low groups defined by the scores of the principal components for the combination of week component 3, that is, PC3—summer versus winter and product group component 1 PC1–weekly average. As the first product group component takes the "average" and the third week component contrasts summer with winter, the seasonality of the highest and lowest deciles is clear; the figures reveal the juxtaposition between summer and winter observed in the loadings for the third time component. Moreover, this pattern persists across almost all product groups, which is in line with the first principal component of product groups representing an approximated average over a large number of product groups.

In addition, we selected the product groups with an absolute loading >0.3 in at least 1 of the selected 18 components and illustrated their purchase patterns for the groups high and low defined by all combinations of the first 3 principal components of weeks and the product groups (Figure S5 in Multimedia Appendix 1). The first product group based on the principal component shows the average expenditure of each group (PC1—product average), whereas the second focuses on pig and bovine meat (PC2—red meat) and the third on the tendency to buy ready-to-eat foods (PC3—ready-to-eat). These illustrations demonstrate how tensorial PCA identifies individuals with specific purchase patterns.

Furthermore, as ready-to-eat foods and pig and bovine meat (hereafter referred to as red meat) were clearly detected in principal components 2 and 3, we focused on them specifically. We divided the participants into deciles based on the selected component and week scores and illustrated them in Figures 5 and 6, in which we can clearly identify the different purchase patterns for the product groups throughout the weeks. For example, in Figures 5A and 6A, we can see the participants with steady differential purchase preferences of the product groups, whereas in Figures 5B and 6B, we see participants with consistently evolving changes in their purchase patterns. In addition, Figures 5C and 6C illustrate the groups having different seasonal purchases of the product groups in winter and summer. For the first deciles of red meat and PC2—spring versus autumn as well as PC3—summer versus winter (Figures 6B and 6C), the week before Christmas still spiked clearly: it is the time of year when Finnish people purchase ham for the Christmas table.

These values are the percentages of weekly purchases, reflecting the proportion of ready-to-eat foods or red meat out of the total purchases. Thus, the percentages indicate changes in the relative proportion of these categories and not changes in total expenditure. In addition, note that, for example, the highest decile in panel A consists of individuals different from those in panel B. This means that we could in principle correlate the scores of each panel with other variables, and potentially clarify the reasons for or the loyalty card holders behind the stable or changing purchase behavior.

Figure 3. Loadings of tensorial principal component analysis components (x-axis) for (A) weeks and (B) product groups. Red indicates a high positive loading, and green indicates a high negative loading, both equally interesting.
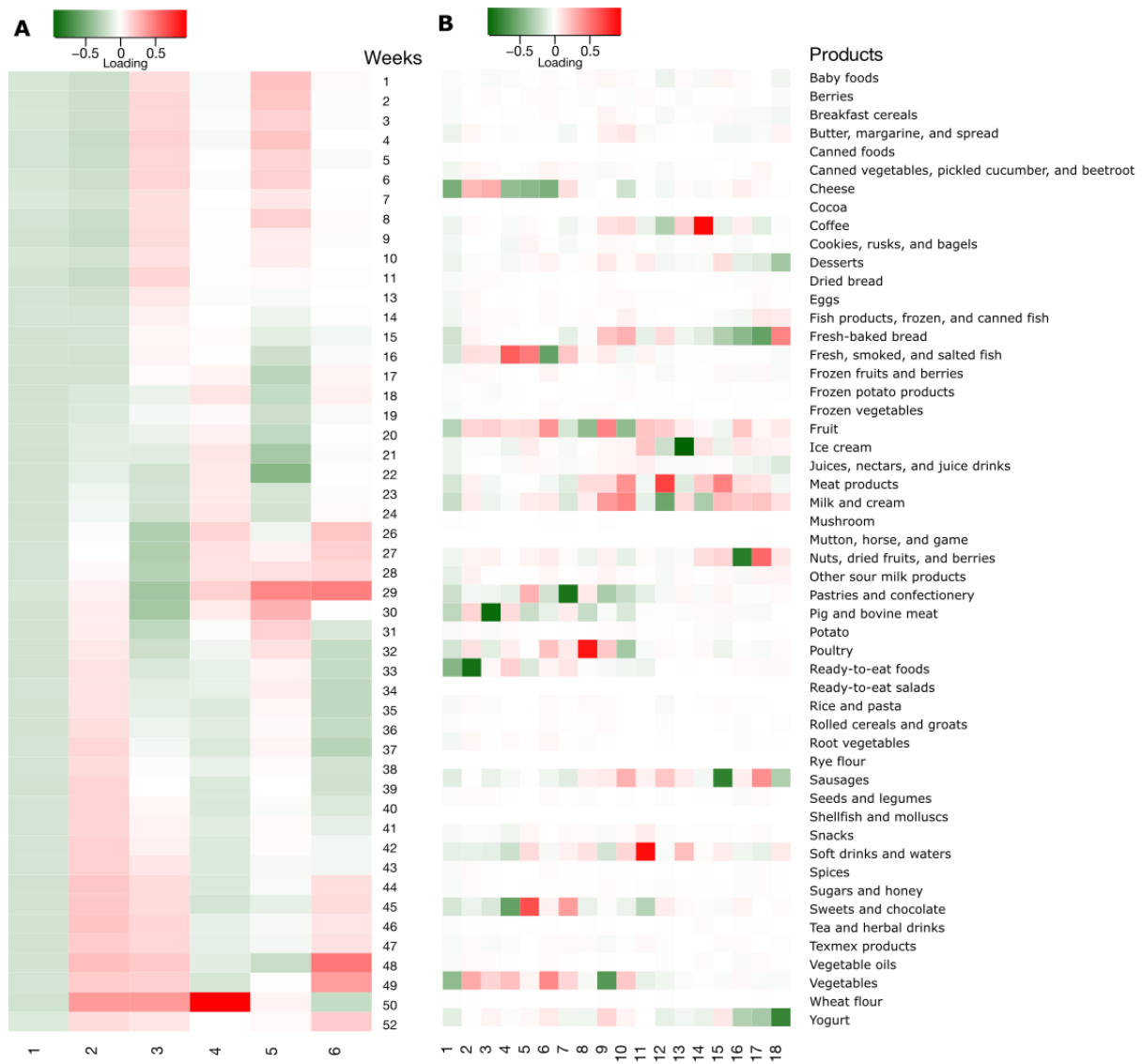
Figure 4. Illustration of the average purchase pattern of the card owners having the highest (A) and lowest (B) 10% scores of the third principal component of weeks (PC3—summer vs winter) and the first principal component of product groups (PC1—product average). The values have been standardized against all customers, after which the averages were computed, that is, a dark green value indicates that the average money spent within the illustrated group on each product group is 0.4 SDs lower than the average across all customers and a strong red value indicates that the average money spent within the illustrated group on each product group is 0.4 SDs higher than the average across all customers.
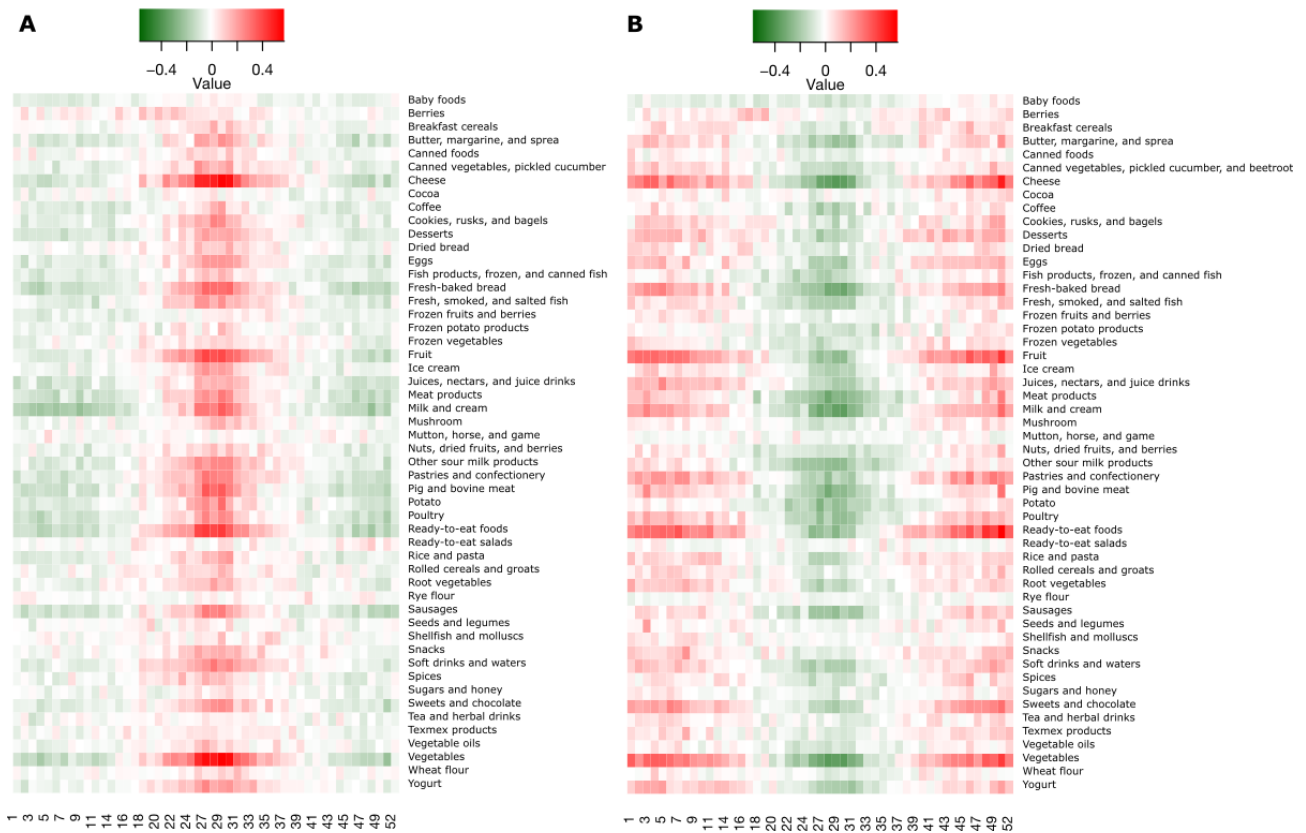
Figure 5. Average percentages of weekly expenditures on ready-to-eat foods, with the deciles divided based on the product group PC2-ready-to-eat and time PC1-weekly average (A), time PC2-spring versus autumn (B), and time PC3-summer versus winter (C). (A) The first time component finds the groups of participants with different levels but temporally stable purchase behavior for ready-to-eat food. (B) The second time component and especially its extreme deciles reveal the participants with increased or decreased use of ready-to-eat food. (C) With the third time component, we can identify participants with seasonal change in the ready-to-eat food purchase pattern.

Figure 6. Average percentages of weekly expenditures on red meat with the deciles divided based on the product group PC3-red meat and time PC1-weekly average (A), time PC2-spring versus autumn (B), and time PC3-summer versus winter (C). (A) The first time component finds the groups of participants with different levels but temporally stable purchase behavior for red meat. (B) The second time component reveals the participants with increased or decreased meat, and (C) detects the participants with summer versus winter difference in meat purchases.
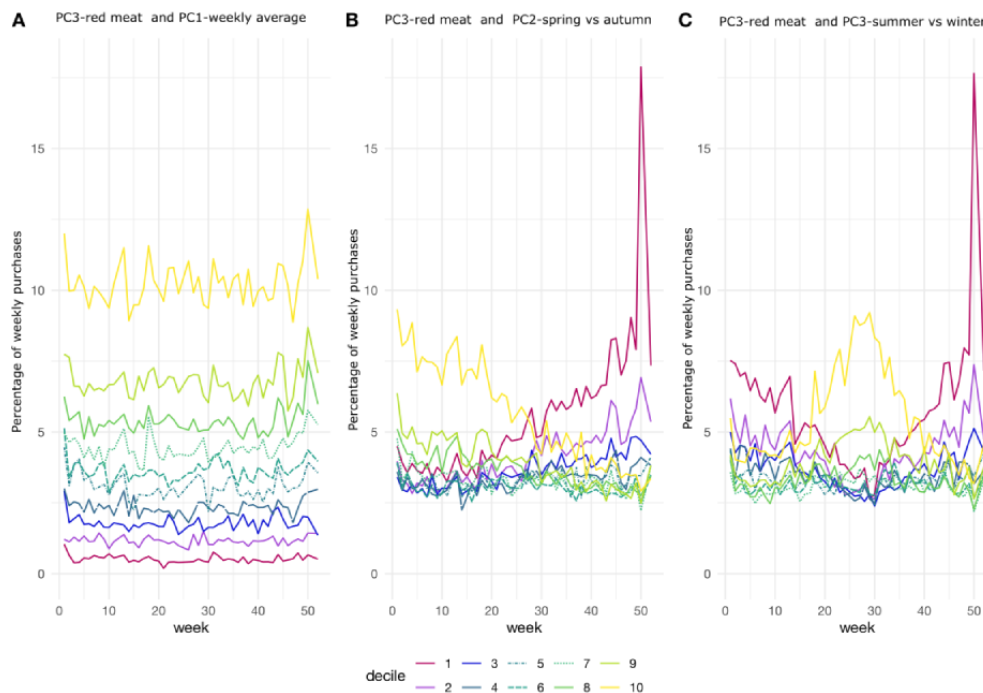
Comparison With the Results of Standard PCA

Table 1 summarizes the differences between the tensorial PCA and standard PCA results. The most evident difference in analyzing the data is that the format of the input data is different. Although tensorial PCA input data are multidimensional $\mathbf{X} \in \mathbf{R}^{7251 \times 52 \times 49}$, the data input to standard PCA is in a 2D matrix format.

To compare the actual results for purchases, we ran the standard PCA for the same filtered data and compared the results with those of tensorial PCA (Table 1). We ran the standard PCA with three data sets modified from the tensor data: (1) *product group data:* data were summed to the amount of money spent on each product group throughout the year $\mathbf{X} \in \mathbf{R}^{7251 \times 52}$; (2) *weekly data:* data were summed to the money spent in each week $\mathbf{X} \in \mathbf{R}^{7251 \times 49}$; and (3) *combination of product group and week:* combines all information in the tensor as such to a matrix $\mathbf{X} \in \mathbf{R}^{7251 \times (52*49)} = \mathbf{R}^{7251 \times 2548}$, where every week-product group combination has its own column.

Using the augmentation method for standard PCA, we selected 22 out of 52 product group components, 4 out of 49 week components, and 40 out of 2548 components for further analysis (Table 1). The analysis revealed that most of the variation between the participants came from the product groups that they buy, rather than from the purchase times, as the loadings behaved very similarly across the weeks within the product groups (Figures S6 and S7 in Multimedia Appendix 1). These results are also consistent with earlier results of tensorial PCA showing very similar patterns of PC loadings as well as weekly

loadings that are smaller than the loadings of the product groups (Figure 3; Figure S4 in Multimedia Appendix 1).

In addition, to compare the results of the standard PCA and tensorial PCA, we computed correlations between the scores of participants based on the first component and (1,1) component of tensorial PCA (PC1—weekly average and PC1—product average), yielding a significant correlation between the product group-wise scores, reaching r=0.988 over the first few components (Figure S8 in Multimedia Appendix 1). Therefore, the standard PCA for the aggregated data yielded approximately the same results as the tensorial PCA for the product group components combined with the first time PC. Recall that the first time component constructed approximately "an average over the year" (PC1—weekly average), which led, not surprisingly, to practically the same result. This finding was also observed for the PC2—ready-to-eat food and PC3—red meat component scores, with the first PC1—weekly average (Figure S8 in Multimedia Appendix 1). However, although the standard PCA could detect the time pattern of the summed purchases in week-based data, it could not easily do this for product groups. Tensorial PCA allowed us to delve deeper into the data and detect the most important changes in patterns over time for the most important product group combinations, a property that is potentially very useful for high time resolution. With standard PCA, we could also combine the product groups and weeks as separate variables; however, as the product group variation was much higher than the variation between weeks, it was not possible to detect timewise variation in the results (Figure S6 in Multimedia Appendix 1).

Table 1. Comparison of standard and tensorial principal component analysis (PCA) of the same but differently arranged data.

Characteristic	PCA for product group data	PCA for weekly data	PCA for product group+week combinations	Tensorial PCA
Interpretation of the results	"Money spent in total on each product group"	"Money spent in total in each week"	"Money spent on each product group in each week in 2D format"	"Money spent on each product group in each week in 3D format"
Data size	$\mathbf{X} \in \mathbf{R}^{7251 \times 52}$	$\mathbf{X} \in \mathbf{R}^{7251 \times 49}$	$\mathbf{X} \in \mathbf{R}^{7251 \times 2548}$	$\mathbf{X} \in \mathbf{R}^{7251 \times 52 \times 49}$
First dimension	Participants	Participants	Participants	Participants
Second dimension	Product groups	Weeks	Product groups×weeks	Product groups
Third dimension	N/A[a]	N/A	N/A	Weeks
Number of significant principal components based on the augmentation method	Product groups: 22	Weeks: 4	Combinations of product groups and weeks: 40	Product groups: 18 and weeks: 6
Cumulative percentage of explained variation	Product groups: 59.5%	Weeks: 14.5%	Combinations of product groups and weeks: 19.4%	Product groups: 81.8%; week: 34.7%
Can find purchase pattern in product groups	Yes	No	Yes	Yes
Can find purchase patterns across time	No	Yes	Yes, but patterns of product groups dominate patterns of times	Yes
Was successful in finding product group pattern–specific yearly trends	No	No	No	Yes
Notes in interpretation	Easy to interpret and limited insight owing to aggregation of time dimension	Easy to interpret and limited insight owing to product group aggregation	Difficult to interpret as loadings are for combinations of product groups and weeks	Both product group and week need to be taken into account in interpretation

[a]N/A: not applicable.

Discussion

Principal Findings

Loyalty card data, an automatic recording of all grocery purchases of the card owner, can be used to analyze the health behavior of customers [46]. Continuous data collection provides not only product group-wise information but also a time component showing when purchases were made. With tensorial PCA, we were able to analyze this multidimensional purchase data in greater detail than before by simultaneously focusing on both time and product group dimensions. The key advantage of tensorial PCA over standard PCA is its ability to effectively capture changes in patterns over time for specific product combinations. Although standard PCA detected the overall time pattern of the summed purchases in our week-based data, it faced challenges in detecting the temporal pattern when observations on different weeks and product groups were expressed as vectors (Figure S6C in Multimedia Appendix 1). Tensorial PCA, in contrast, allowed us to analyze not only at the aggregate level but also deeper; we were able to uncover the most significant changes in patterns over time and product groups. By leveraging the tensor structure of the data, tensorial PCA effectively captured the interplay between products and periods, enabling us to identify temporal variations in product-specific purchasing behaviors. It provided valuable insights into the dynamics of product combinations, allowing us to identify temporal shifts in card holders' purchasing patterns more accurately. This characteristic could be important in, for instance, assessing changing dietary (purchase) patterns following external alterations, such as price change following a price inflation or policy implementation (eg, sugar tax) or major disruptions in society (eg, lockdown during a pandemic).

Comparison With Prior Work

Traditional PCA has been widely used for nutrition data among different settings, cultures, and sociodemographic groups. Although the naming of the patterns and the foods loading to the components may vary slightly, typically at least 2 common patterns are identified in most countries, as in Finland: a prudent, healthy dietary pattern and an unhealthy "Western" pattern [47-51]. A third, almost equally typical pattern is often termed "traditional" [52], and this pattern is generally more context specific. A traditional Finnish diet is characterized by sausages, potatoes, milk, coffee, and butter [53]. A ready-to-eat pattern has also often been identified, characterized by a high consumption of ready-to-eat meals [51,54]. Similarly, a pattern indicating alcohol consumption has been identified previously in Finland [55]. As mentioned in our results, the standard PCA for the aggregated data yielded similar results to the tensorial PCA for the product group components combined with the first time PC. The added value of tensorial PCA was that it enabled us to identify patterns indicative of broader trends in purchasing behavior and to identify specific weeks for specific product groups.

The use of grocery purchase data for health research purposes is still relatively novel, and, to the best of our knowledge, only a few earlier studies have identified dietary patterns based on customer loyalty card data. One such study was conducted in the United Kingdom [3], whereas the other is an earlier study conducted by our group [51]. Both studies used the standard PCA method for analysis. Our previous study identified 8 patterns based on a more detailed food grouping within the same LoCard data used in this study [53]. Consistent with previous findings, we also identified patterns characterized by a high consumption of ready-to-eat products and red meat. Moreover, barcode scanning was used to examine purchase patterns, revealing a consistent finding of a pattern characterized by high consumption of ready-to-eat meals [54]. A direct comparison between our study and earlier studies on purchase patterns is not feasible because of several reasons. First, different countries have distinct food cultures that can significantly impact the identified patterns. In addition, the analyzed food product groups may consist of different items across studies, further complicating direct comparisons. Finally, the choice of analytic methods used can introduce differences in the results obtained. Therefore, it is important to acknowledge these issues and approach the comparison with caution.

Strengths and Limitations

The main strength of this study is the vast data set used, as S Group is a leading retailer with a market share as high as 47.2% in Finland [56]. Its shops cover the entire country, thus providing an excellent means of investigating the purchase patterns of Finnish people on a large scale. With grocery stores constantly collecting more detailed data and, thus, also loyalty card data having more dimensions than ever before, tensorial PCA provides an optimal means to analyze such data. Tensorial PCA simultaneously focuses on multiple dimensions and finds principal components in a dimension that otherwise would have been masked by another dimension with higher variability. Some limitations of this study also need to be addressed. First, there are other grocery stores in Finland; therefore, not all grocery purchases of the households are included. Second, only transactions made by customers using their loyalty card are captured. Although there are significant benefits for customers associated with using the card, not all customers actually use or even carry it with them, and the results thus perhaps represent the customers of S Group rather than the Finnish adult population. However, we have previously found that the age distribution of participants of the LoCard study is similar to that of the residents within the region, although the proportion of women is higher among loyalty customers (67.3%) than among residents (52.1%) [2]. Therefore, the interpretation of data is constrained by these limitations [57]. It should also be noted that the naming of PCA-derived purchase patterns is a highly subjective decision. Therefore, it is of great importance to publish information on food grouping and factor loadings along with a thoughtful naming of the components. This will enable the assessment of reproducibility and similarity.

Future Work

The purchase patterns disclosed here can be used in subsequent analysis. For example, we can identify the use of meat product groups and determine whether it is related to specific times of the year or whether it is stable throughout the year. In this way, the impact of timed social interventions, such as meat-free October or vegetarian months, on the purchase of meat can be evaluated. Several year trends in the consumption of red meat and its substitutes would also be important to monitor and understand [58,59].

We could determine the association between customer demographics (such as age and gender) and patterns to understand the behavior of different population subgroups, potentially important in sustainable food consumption, human health, and retailers' interests. For example, we can identify the type of individuals who buy more meat during summer or whose purchase pattern is focused more on vegetables overall or at a specific time of the year. Sociodemographic studies show that educated urban women are ahead of the curve in moving toward a more sustainable diet [59-62]. They are better able to follow the path of a larger sustainable dietary change, whereas for some other population groups, such as men and those less educated, making smaller dietary changes is more likely to be successful [59,61]. Tensorial PCA can be used for detailed monitoring of the nutrition and vulnerability of different population groups in a food system transformation and, for example, when significant changes occur in food prices.

In addition, the analysis can be used to detect the effects of external factors, including inflation or a global crisis, such as the COVID-19 pandemic or the Ukraine war, on the everyday purchase behavior of customer groups. The persistence of these changes can also be evaluated, yielding insights into which sociodemographic groups are most affected by these external factors. Tensorial PCA will also help to identify the influence of national steering instruments, such as taxation and updated nutrition recommendations, on purchase trends over time. In problems such as these, the ideal data would be multidimensional and of high resolution for each dimension. This makes tensorial dimension reduction methods, such as tensorial PCA, valuable in these contexts, as they can both reduce the data dimension in an interpretable manner and keep the subject, product group, and time dimensions separate.

Conclusions

This is the first study on the weekly purchase patterns of Finnish customers that simultaneously considers both the time and product group dimensions for each customer. With tensorial PCA, we could identify abnormalities in the purchasing behaviors during specific weeks or for specific product groups and were able to detect patterns of wider trends in customers' purchasing behavior. By using standard PCA, we found the principal components for either the weeks or the product groups, whereas with tensorial PCA, we identified purchase patterns simultaneously based on the dimensions in the data tensor. By selecting specific features identified based on the patterns within each dimension, we detected the participant groups with specific purchase patterns. In further analyses, these patterns based on time and product groups are likely to be directly linked to the socioeconomic and demographic predictors of dietary patterns of customers. These associations will enable us to identify what types of people have specific purchasing patterns, which in turn

can assist in the development of ways in which consumers can   be steered toward making healthier food choices.

Data Availability

The data used in this study are owned by a third party (S Group) and used under a research agreement. The data underlying this study cannot be shared publicly for the privacy of individuals who participated in the study.

Authors' Contributions

JN, ME, and MF conceptualized the project and curated the data. ME, JV, and JN performed data management. All authors participated in the study design. JV, KN, JN, and RA were responsible for the methodology development. JN, RA, JV, and KN planned the data analyses, which was conducted by RA. RA, JV, and JN wrote the original draft. All authors participated in writing the drafts and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary text, tables, and figures.
[PDF File (Adobe PDF File), 2146 KB-Multimedia Appendix 1]

References

1. Demchenko Y, Grosso P, De LC, Membrey P. Addressing big data issues in scientific data infrastructure. In: Proceedings of the 2013 International Conference on Collaboration Technologies and Systems. 2013 Presented at: CTS '13; May 20-24, 2013; San Diego, CA p. 48-55 URL: https://ieeexplore.ieee.org/document/6567203 [doi: 10.1109/cts.2013.6567203]
2. Nevalainen J, Erkkola M, Saarijärvi H, Näppilä T, Fogelholm M. Large-scale loyalty card data in health research. Digit Health. 2018 Nov 29;4:2055207618816898 [FREE Full text] [doi: 10.1177/2055207618816898] [Medline: 30546912]
3. Clark SD, Shute B, Jenneson V, Rains T, Birkin M, Morris MA. Dietary patterns derived from UK supermarket transaction data with nutrient and socioeconomic profiles. Nutrients. 2021 Apr 27;13(5):1481 [FREE Full text] [doi: 10.3390/nu13051481] [Medline: 33925712]
4. Rains T, Longley P. The provenance of loyalty card data for urban and retail analytics. J Retail Consum Serv. 2021 Nov;63:102650 [FREE Full text] [doi: 10.1016/j.jretconser.2021.102650]
5. Lintonen T, Uusitalo L, Erkkola M, Rahkonen O, Saarijärvi H, Fogelholm M, et al. Grocery purchase data in the study of alcohol use - a validity study. Drug Alcohol Depend. 2020 Sep 01;214:108145 [doi: 10.1016/j.drugalcdep.2020.108145] [Medline: 32663761]
6. Aiello LM, Schifanella R, Quercia D, Del Prete L. Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci. 2019 Apr 30;8(1) [FREE Full text] [doi: 10.1140/epjds/s13688-019-0191-y]
7. Hansel B, Roussel R, Diguet V, Deplaude A, Chapman MJ, Bruckert E. Relationships between consumption of alcoholic beverages and healthy foods: the French supermarket cohort of 196,000 subjects. Eur J Prev Cardiol. 2015 Feb;22(2):215-222 [doi: 10.1177/2047487313506829] [Medline: 24065742]
8. Moran AJ, Khandpur N, Polacsek M, Thorndike AN, Franckle RL, Boulos R, et al. Make it fresh, for less! A supermarket meal bundling and electronic reminder intervention to promote healthy purchases among families with children. J Nutr Educ Behav. 2019 Apr;51(4):400-408 [FREE Full text] [doi: 10.1016/j.jneb.2019.01.012] [Medline: 30765298]
9. Timberlake DS, Joensuu J, Kurko T, Rimpelä AH, Nevalainen J. Examining retail purchases of cigarettes and nicotine replacement therapy in Finland. Tob Induc Dis. 2019 May 3;17(May):39 [FREE Full text] [doi: 10.18332/tid/108537] [Medline: 31516482]
10. Uusitalo L, Erkkola M, Lintonen T, Rahkonen O, Nevalainen J. Alcohol expenditure in grocery stores and their associations with tobacco and food expenditures. BMC Public Health. 2019 Jun 20;19(1):787 [FREE Full text] [doi: 10.1186/s12889-019-7096-3] [Medline: 31221122]
11. Livingston M, Callinan S. Underreporting in alcohol surveys: whose drinking is underestimated? J Stud Alcohol Drugs. 2015 Jan;76(1):158-164 [Medline: 25486405]
12. Willett W. Nutritional Epidemiology. Oxford, United Kingdom. Oxford University Press; 2012.

13. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Rev. 2009 Aug 06;51(3):455-500 [doi: [10.1137/07070111X](#)]

14. Sun W, Hao B, Li L. Tensors in modern statistical learning. In: Balakrishnan N, Colton T, Everitt B, Piegorsch WW, Ruggeri F, Teugels JL, editors. Wiley StatsRef: Statistics Reference Online. Hoboken, NJ. John Wiley & Sons; 2021:1-25

15. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. Lond Edinb Dublin Philos Mag J Sci. 2010 Jun 08;2(11):559-572 [[FREE Full text](#)] [doi: [10.1080/14786440109462720](#)]

16. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933 Sep;24(6):417-441 [[FREE Full text](#)] [doi: [10.1037/h0071325](#)]

17. Kinson C, Tang X, Zuo Z, Qu A. Longitudinal principal component analysis with an application to marketing data. J Comput Graph Stat. 2019 Nov 05;29(2):335-350 [[FREE Full text](#)] [doi: [10.1080/10618600.2019.1677244](#)]

18. Tapsell LC, Neale EP, Satija A, Hu FB. Foods, nutrients, and dietary patterns: interconnections and implications for dietary guidelines. Adv Nutr. 2016 May;7(3):445-454 [[FREE Full text](#)] [doi: [10.3945/an.115.011718](#)] [Medline: [27184272](#)]

19. Newby P, Tucker K. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev. 2004 May;62(5):177-203 [[FREE Full text](#)] [doi: [10.1111/j.1753-4887.2004.tb00040.x](#)]

20. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002 Feb;13(1):3-9 [doi: [10.1097/00041433-200202000-00002](#)] [Medline: [11790957](#)]

21. Mozaffarian D, Rosenberg I, Uauy R. History of modern nutrition science-implications for current research, dietary guidelines, and food policy. BMJ. 2018 Jun 13;361:k2392 [[FREE Full text](#)] [doi: [10.1136/bmj.k2392](#)] [Medline: [29899124](#)]

22. Jenneson VL, Pontin F, Greenwood DC, Clarke GP, Morris MA. A systematic review of supermarket automated electronic sales data for population dietary surveillance. Nutr Rev. 2022 May 09;80(6):1711-1722 [[FREE Full text](#)] [doi: [10.1093/nutrit/nuab089](#)] [Medline: [34757399](#)]

23. Virta J, Taskinen S, Nordhausen K. Applying fully tensorial ICA to fMRI data. In: Proceedings of the 2016 Signal Processing in Medicine and Biology Symposium. 2016 Presented at: SPMB '16; August 16-20, 2016; Philadelphia, PA p. 1-6 URL: [https://ieeexplore.ieee.org/document/7846858](#) [doi: [10.1109/spmb.2016.7846858](#)]

24. Zhang D, Zhou ZH. (2D)2 PCA: two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing. 2005 Dec;69(1-3):224-231 [[FREE Full text](#)] [doi: [10.1016/j.neucom.2005.06.004](#)]

25. Hung H, Wu P, Tu I, Huang S. On multilinear principal component analysis of order-two tensors. Biometrika. 2012 May 22;99(3):569-583 [[FREE Full text](#)] [doi: [10.1093/biomet/ass019](#)]

26. Lu H, Plataniotis KN, Venetsanopoulos AN. MPCA: multilinear principal component analysis of tensor objects. IEEE Trans Neural Netw. 2008 Jan;19(1):18-39 [[FREE Full text](#)] [doi: [10.1109/tnn.2007.901277](#)]

27. Tingwei G, Xiu L, Yueting C, Youhua T. Deep learning with stock indicators and two-dimensional principal component analysis for closing price prediction system. In: Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science. 2016 Presented at: ICSESS '16; August 26-28, 2016; Beijing, China p. 166-169 URL: [https://ieeexplore.ieee.org/document/7883040](#) [doi: [10.1109/icsess.2016.7883040](#)]

28. Household consumption expenditure by type of household 1985-2016. Statistics Finland. URL: [https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__ktutk/statfin_ktutk_pxt_001.px/](#) [accessed 2023-09-14]

29. Jollife IT. Principal Component Analysis. New York, NY. Springer-Verlag; 2002.

30. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl. 2000 Aug;21(4):1253-1278 [doi: [10.1137/s0895479896305696](#)]

31. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016 Apr 13;374(2065):20150202 [[FREE Full text](#)] [doi: [10.1098/rsta.2015.0202](#)] [Medline: [26953178](#)]

32. Virta J, Li B, Nordhausen K, Oja H. Independent component analysis for tensor-valued data. J Multivar Anal. 2017 Nov;162:172-192 [[FREE Full text](#)] [doi: [10.1016/j.jmva.2017.09.008](#)]

33. Luo W, Li B. Combining eigenvalues and variation of eigenvectors for order determination. Biometrika. 2016 Dec 08;103(4):875-887 [[FREE Full text](#)] [doi: [10.1093/biomet/asw051](#)]

34. Luo W, Li B. On order determination by predictor augmentation. Biometrika. 2021 Aug;108(3):557-574 [[FREE Full text](#)] [doi: [10.1093/biomet/asaa077](#)]

35. Radojicic U, Lictzen N, Nordhausen K, Virta J. Dimension estimation in two-dimensional PCA. In: Proceedings of the 12th International Symposium on Image and Signal Processing and Analysis. 2021 Presented at: ISPA '21; September 13-15, 2021; Zagreb, CA p. 16-22 URL: [https://ieeexplore.ieee.org/document/9552114](#) [doi: [10.1109/ispa52656.2021.9552114](#)]

36. Radojicic U, Lietzen N, Nordhausen K, Virta J. Order determination for tensor-valued observations using data augmentation. arXiv. Preprint posted online July 21, 2022 [[FREE Full text](#)] [doi: [10.48550/arxiv.2207.10423](#)]

37. Bishop CM. Pattern Recognition and Machine Learning. New York, NY. Springer-Verlag; 2006.

38. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: [https://www.r-project.org/](#) [accessed 2023-11-18]

39. Virta J, Koesner CL, Li B, Nordhausen K, Oja H, Radojicic U. tensorBSS: blind source separation methods for tensor-valued observations R package version 0.3.8. 2021. Cran R Project. URL: [https://cran.r-project.org/package=tensorBSS](#) [accessed 2023-11-18]

XSL•FO

RenderX

40. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: various R programming tools for plotting data. Cran R. URL: https://cran.r-project.org/web/packages/gplots/index.html [accessed 2023-11-18]

41. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY. Springer-Verlag; 2016.

42. Iranpour A, Nakhaee N. A review of alcohol-related harms: a recent update. Addict Health. 2019 Apr;11(2):129-137 [FREE Full text] [doi: 10.22122/ahj.v11i2.225] [Medline: 31321010]

43. Alcohol. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/alcohol [accessed 2022-11-07]

44. GBD 2020 Alcohol Collaborators. Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020. Lancet. 2022 Jul 16;400(10347):185-235 [FREE Full text] [doi: 10.1016/S0140-6736(22)00847-9] [Medline: 35843246]

45. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. BMJ. 2011 Feb 22;342:d671 [FREE Full text] [doi: 10.1136/bmj.d671] [Medline: 21343207]

46. Vepsäläinen H, Nevalainen J, Kinnunen S, Itkonen ST, Meinilä J, Männistö S, et al. Do we eat what we buy? Relative validity of grocery purchase data as an indicator of food consumption in the LoCard study. Br J Nutr. 2021 Oct 18;128(9):1780-1788 [doi: 10.1017/s0007114521004177]

47. Kanerva N, Wachira LJ, Uusi-Ranta N, Anono EL, Walsh HM, Erkkola M, et al. Wealth and sedentary time are associated with dietary patterns among preadolescents in Nairobi City, Kenya. J Nutr Educ Behav. 2023 May;55(5):322-330 [FREE Full text] [doi: 10.1016/j.jneb.2023.02.001] [Medline: 36914443]

48. Mikkilä V, Vepsäläinen H, Saloheimo T, Gonzalez SA, Meisel JD, Hu G, et al. ISCOLE Research Group. An international comparison of dietary patterns in 9-11-year-old children. Int J Obes Suppl. 2015 Dec;5(Suppl 2):S17-S21 [FREE Full text] [doi: 10.1038/ijosup.2015.14] [Medline: 27152179]

49. Trudeau K, Rousseau MC, Csizmadi I, Parent MÉ. Dietary patterns among French-speaking men residing in Montreal, Canada. Prev Med Rep. 2019 Mar;13:205-213 [FREE Full text] [doi: 10.1016/j.pmedr.2018.12.017] [Medline: 30705807]

50. Peltonen H, Erkkola M, Abdollahi AM, Leppänen MH, Roos E, Sajaniemi N, et al. Associations of dietary patterns with common infections and antibiotic use among Finnish preschoolers. Food Nutr Res. 2023 Jun 14;67 [FREE Full text] [doi: 10.29219/fnr.v67.8997] [Medline: 37351020]

51. Meinilä J, Hartikainen H, Tuomisto HL, Uusitalo L, Vepsäläinen H, Saarinen M, et al. Food purchase behaviour in a Finnish population: patterns, carbon footprints and expenditures. Public Health Nutr. 2022 Nov;25(11):3265-3277 [FREE Full text] [doi: 10.1017/S1368980022001707] [Medline: 35979803]

52. Hinnig P, Monteiro J, de Assis M, Levy R, Peres M, Perazi F, et al. Dietary patterns of children and adolescents from high, medium and low human development countries and associated socioeconomic factors: a systematic review. Nutrients. 2018 Mar 30;10(4):436 [FREE Full text] [doi: 10.3390/nu10040436] [Medline: 29601553]

53. Mikkilä V, Räsänen L, Raitakari OT, Marniemi J, Pietinen P, Rönnemaa T, et al. Major dietary patterns and cardiovascular risk factors from childhood to adulthood. The Cardiovascular Risk in Young Finns study. Br J Nutr. 2007 Jul 01;98(1):218-225 [doi: 10.1017/s0007114507691831]

54. Piernas C, Mendez MA, Ng SW, Gordon-Larsen P, Popkin BM. Low-calorie- and calorie-sweetened beverages: diet quality, food intake, and purchase patterns of US household consumers. Am J Clin Nutr. 2014 Mar;99(3):567-577 [FREE Full text] [doi: 10.3945/ajcn.113.072132] [Medline: 24351878]

55. Uusitalo U, Arkkola T, Ovaskainen M, Kronberg-Kippilä C, Kenward MG, Veijola R, et al. Unhealthy dietary patterns are associated with weight gain during pregnancy among Finnish women. Public Health Nutr. 2009 Dec;12(12):2392-2399 [doi: 10.1017/S13689800900528X] [Medline: 19323867]

56. Finnish grocery trade 2016. Finnish Grocery Trade Association. 2016. URL: https://www.pty.fi/paeivittaeistavarakaupan-myynti-ja-kaupan-ryhmittymien-markkinaosuudet-2016-julkistettiin/ [accessed 2023-12-04]

57. Vuorinen AL, Erkkola M, Fogelholm M, Kinnunen S, Saarijärvi H, Uusitalo L, et al. Characterization and correction of bias due to nonparticipation and the degree of loyalty in large-scale Finnish loyalty card data on grocery purchases: cohort study. J Med Internet Res. 2020 Jul 15;22(7):e18059 [FREE Full text] [doi: 10.2196/18059] [Medline: 32459633]

58. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet. 2019 Feb 02;393(10170):447-492 [doi: 10.1016/S0140-6736(18)31788-4] [Medline: 30660336]

59. Erkkola M, Kinnunen SM, Vepsäläinen HR, Meinilä JM, Uusitalo L, Konttinen H, et al. A slow road from meat dominance to more sustainable diets: an analysis of purchase preferences among Finnish loyalty-card holders. PLOS Sustain Transform. 2022 Jun 16;1(6):e0000015 [doi: 10.1371/journal.pstr.0000015]

60. Hjorth T, Huseinovic E, Hallström E, Strid A, Johansson I, Lindahl B, et al. Changes in dietary carbon footprint over ten years relative to individual characteristics and food intake in the Västerbotten Intervention Programme. Sci Rep. 2020 Jan 08;10(1):20 [FREE Full text] [doi: 10.1038/s41598-019-56924-8] [Medline: 31913331]

61. Kaljonen M, Karttunen K, Kortetmäki T. A just food system transformation. Pathways to a Sustainable and Fair Food System. URL: https://helda.helsinki.fi/handle/10138/349713 [accessed 2022-10-28]

62.   Tschanz L, Kaelin I, Wróbel A, Rohrmann S, Sych J. Characterisation of meat consumption across socio-demographic, lifestyle and anthropometric groups in Switzerland: results from the National Nutrition Survey menuCH. Public Health Nutr. 2022 Nov;25(11):3096-3106 [FREE Full text] [doi: 10.1017/S136898002200101X] [Medline: 35466905]

Abbreviations

fMRI: functional magnetic resonance imaging
PCA: principal component analysis