Data-driven exploration of Shopping Behaviour in e-Grocery: Comparing Online and Offline Customers

Eleftheria Matta, George Stalidis[*] and Maria Pasiata

Department of Organisation Management, Marketing and Tourism, International Hellenic University, Thessaloniki, Greece

International Hellenic University – Thessaloniki campus

PO Box 141 GR

17 Km Thessaloniki-Sindos

57400 Thessaloniki, Greece

*email: stalidgi@ihu.gr

Data-driven exploration of Shopping Behaviour in e-Grocery: Comparing Online and Offline Customers

Online grocery customers are a new and rapidly growing market segment that calls for new marketing strategies and exploration of their profile. Understanding the differences in shopping behaviour between online and brick-and-mortar customers allows for targeted marketing and effective personalised promotions. This study examines the shopping behaviour of exclusive online and brick-and-mortar customers by analyzing their transaction history and loyalty program records. Factor analysis was performed to reveal the main dimensions that explain customers' behavior and clustering methods were used to identify the major customer profiles. Specifically, Multiple Correspondence Analysis (MCA) and Hierarchical Cluster Analysis on Principal Components (HCPC) were applied on a dataset spanning 60K customer records over a year. The research uncovered differences between online and brick-and-mortar customers in their purchase frequency, transaction amounts, and preferred product categories. The most important customer profiles of both categories were studied and valuable insights were derived about their needs and preferences.

Keywords: e-grocery, Multiple Correspondence Analysis, clustering, online shopping, brick-and-mortar supermarket, retail, shopping behaviour

Introduction

Online grocery shopping is experiencing significant increase in popularity, accelerated by the recent pandemic. Although in-store purchases remain popular due to perishable goods, allowing consumers to physically inspect and select fresh products (Galante et al., 2013), online customers are a new and rapidly growing market segment that calls for new marketing strategies and exploration. In the field of supermarkets and e-Grocery, businesses are exploiting a variety of channels, such as mobile apps, personalised ads and advanced loyalty programs, which allow for improved targeting and feedback collection. Intelligent data-driven techniques are widely used nowadays to capture the profiles of consumers and to perform targeted marketing actions, such as personalised

special offers, recommendations and push notifications (Stalidis, 2019) and marketing research is guided by the analysis of purchase history and customer data from loyalty programs. Although considerable research work has been reported in this field, a relatively unexplored problem is the profiling of online grocery customers and, in particular, their comparison with physical customers. This paper seeks to analyse the disparities in purchasing behaviours between online and offline consumers, offering valuable insights for a grocery company. These insights can aid in improving product offerings, pricing strategies, and targeted promotions, leading to enhanced customer satisfaction, loyalty, and revenue growth.

Existing research on the differences between online and offline grocery shoppers has shown that online shoppers tend to look at the pictures of products, rather than examine detailed product information, and 35% of them never look at information such as lists of ingredients or nutritional information (Benn et al, 2015). Online consumers demonstrate higher price sensitivity but exhibit reduced propensity for brand switching, while brick-and-mortar consumers frequently choose featured products even when there's minimal price reduction (Degeratu et al, 2000). Furthermore, Chu et al. (2010) suggested that light online shoppers demonstrate strong brand and size loyalties but minimal price sensitivity online, whereas heavy online shoppers exhibit minimal brand and size loyalties but strong price sensitivity. Recent research by Verstraeten et al. (2023) indicated that consumers tend to purchase a greater proportion of private label (PL) food products online, a trend supported also by Dawes & Nenycz-Thiel (2014). Furthermore, customers tend to select fewer unhealthy products when shopping online, as highlighted by Huyghe et al. (2017). Additionally, the buying process in online environments is typically shorter than in traditional supermarkets, as observed by Hanus (2016). Zatz et al., (2000) found that households that shop online are more likely to

have a female primary shopper compared to those exclusively shopping in-store, while, generally, online shoppers have higher incomes, tend to be younger, possess higher levels of education, and enjoy greater affluence compared to the general population.

In our own previous work (Matta & Stalidis, 2023), the various profiles of supermarket customers were analysed by applying multidimensional factor and clustering statistical methods to data from both physical stores and e-shop, integrated with demographic data available through the loyalty program of a large supermarket chain in Greece. In relation to the average supermarket customer, e-customers tended to be small spenders, not promo hunters, who mainly bought non-food products and did not show preference for private label products. In the current paper, we applied similar methods to a refined dataset, analysing online customers separately from the general supermarket customers. This study aims to compare the profiles of customers who are purely online with those who have made purchases exclusively from brick-and-mortar stores. The profiles were derived from the customers' purchase history, integrated with demographic information from loyalty cards. Sales data covered a rolling year (April 1, 2021 – March 31, 2022). The sample of online customers consisted of 617 individuals, and the sample of brick-and-mortar customers consisted of 59,906.

Methods and main findings

The primary technique employed in our research was Multiple Correspondence Analysis (MCA), a dimensionality-reduction method. Unlike approaches restricted to quantitative variables, MCA is well-suited for datasets containing numerous categorical variables (Greenacre, 2013). This method offers intuitive and graphical means of estimating and visualizing complex relationships among qualitative features, along with discovering trends in customer behaviour (Manca et al., 2018). Following MCA, Hierarchical Cluster Analysis on Principal Components (HCPC) was applied on the

dimensions generated by MCA, in order to identify customer groups with particular purchasing behaviour, product preferences and socio-demographic profile. The analysis process was applied separately for online and for brick-and-mortar customers, with subsequent comparison of the derived factors and clusters. The analysis was conducted using the FactoMineR R package.

In order to capture the customers' buying behaviour, the purchase data were aggregated at customer and product group levels. Across 35 different product categories, quantities were normalized per each customer's total purchases and per average quantities of the product category for all customers, and then discretised into three levels. The resulting variables expressed the preference level of each customer for particular product groups and were used as active qualitative variables. Furthermore, a series of 17 indicators were evaluated for each customer, e.g. their preference towards Private Label products and products on promotion. The 17 indicators acted as supplementary behavioural variables, meaning they didn't participate in the estimation of the factors but were projected onto the factorial planes to illustrate connections between product preferences and behavioural indicators.

Analysis of online customers

The MCA results for online customers are shown in Figure 1. It is noted that the points projected in red font on the factorial plane correspond to purchase intentions for particular product groups. Categories with suffix 1 indicate minimal or no purchasing of the specific product group, suffix 2 indicates average purchasing (compared to other purchases by the same customer and in relation to the average customer profile), and suffix 3 indicates high preference for the specific product group.

Figure 1 near here

On the left side of the factorial plane F1 X F2, we observed many low purchasing categories (i.e. level 1) of various product groups, whereas on the right side the corresponding average purchasing categories. It became apparent that the 1st factor (28.08% of inertia) delineated the distinction between limited purchases of **occasional** customers and the average purchases of **regular customers**. By analyzing the supplementary variables, we discovered that the occasional profile was associated with a lack of preference for private label and organic products, low monthly expenditures below 50€, a tendency towards premium products and seeking promotions, mostly receiving their orders during mornings or evenings, particularly on weekends and not having children. In contrast, the regular profile was linked to a strong preference for private labels and organic products, a normal behavior towards promotions, having elder family members and children. The level of spending ranged between 100-350€ per month. Along the vertical axis (2nd factor - 5.99% of inertia) high purchasing **food** product groups are positioned at the bottom side while **non-food** product groups are positioned at the top side. The 2nd dimension was thus the food vs non-food factor.

Continuing the interpretation of factorial planes, up to the 5th dimension, the 3rd factor differentiated the preference for stockpiling canned food & quick pasta products vs convenience, fresh & ready-made products. The 4th factor was characterized as quick bite vs regular meal. The fifth factor (2.91% of inertia) juxtaposed high-spending families with children who buy alcohol, packaged cold cuts, sugary snacks, and cooking ingredients such as frozen meat, with individuals or small households who purchase ready meals, fresh produce, bulk or frozen meat, cold cuts, and beverages. Therefore, the 5th factor was interpreted as families with large shopping baskets that cook, prefer home entertainment with alcohol and snacks, versus individuals, workers, students or young couples with smaller baskets who consume ready meals, meat, cold cuts, and

fresh greengroceries. The application of HCPC for online customers resulted in the 6 clusters shown in Table 1. In order to associate clusters with customer profiles, the clusters were projected on the factorial planes F1 x F2 and F3 x F5.

Table 1 near here

On the factorial plane F1XF2 (Occasional vs regular customers X Food vs non-food), we observed that: Clusters 1 and 2 are on the left side and correspond clearly to occasional customers with low spending. Since C1 is at the bottom side, it is characterized by preference to food products, while C2, being at the top side, to nonfood products. Regarding the 1st factor, clusters 3,4 and 5 are projected at a neutral position, indicating that they are neither typical occasional customers nor regular ones. By their position on the 2nd factor, it appears that cluster 3 is linked to more buys of food products (bottom side), while clusters 4 and 5 are linked to non-food products. Cluster 6 is projected to the right edge and therefore corresponds to regular high spending customers. Notice that clusters 4 and 5 overlap on the factorial plane F1 x F2 but are clearly distinguished on F3 x F5 (the stockpiling canned food & quick pasta vs convenience & ready-made X high spending families with children vs individuals or small households that consume ready meals). Cluster 4 is linked to stockpiling canned food and to small households, while Cluster 5 is linked to ready meals and to families that prefer home entertainment. The purchasing behavior of the six online customer clusters is summarized as follows:

Figure 2 near here

Cluster 1: Infrequent promo hunters & PL lovers (18% of the sample**).** This group purchases both food and non-food products and exhibits the highest percentage of products on discount as well as the highest rate of private label purchases.

Cluster 2: **Infrequent nonfood customers** (13%)**.** This group also shows no preference

for private label products.

Cluster 3: Infrequent promo hunters who dislike PL (29%)**.** They purchase primarily food but also nonfood product categories with a preference for dairy, eggs and cold cuts, canned food, pasta & pulses.

Cluster 4: Moderate spenders, nonfood customers (22%)**.** They purchase primarily nonfood products. The food categories they purchase are mostly canned food and some ingredients for cooking quick meals like pasta.

Cluster 5: Moderate spenders, balanced FNF customers (20%)**.** They maintain a balanced spending approximately 80% to food and 20% to non-food categories. They purchase alcohol and refreshments and ready meals and seem to be families with kids.

Cluster 6: High spenders, balanced FNF customers (7%)**.** They purchase all the product groups.

Analysis of Brick-and-Mortar Customers

Applying a similar process for brick-and-mortar customers, we found that the 1st factor (33.4% of inertia) expressed the contrast between limited **occasional** buys and the average purchases of **regular** customers and the 2nd dimension (4.7% of inertia) was the **food** versus **non-food** factor, as shown in Figure 3.

Figure 3 near here

The 3rd factor (3.9% of inertia) juxtaposed the purchasing of ingredients for cooking from preference for snacks and ready meals, and was labeled as **home cooking vs ready-made**. The 4th factor (2.6% of inertia) juxtaposed the preference for **bulk** cheese and cold cuts from the **packaged** ones, while the 5th factor (2%) juxtaposed

breakfast products from fresh food and alcohol. Applying HCPC, 6 physical customer clusters were identified, for which the descriptives are shown in Table 2.

Table 2 near here

By observing the projection of the clusters on the factorial planes, we extracted in more detail the associations among the clusters and the categories of both main and supplementary variables (Figure 4). Clusters 3,4,5 and 6 partially overlap on the factorial plane F1 x F2 and are better distinguished on F3xF4. The consolidated profiles of the six clusters are:

Cluster 1: **Occasional customers** (24% of the sample). They purchase all product categories. They tend to purchase one out of three products on promotion and appear to favour private label items.

Cluster 2: Low spenders, food customers (18%). They appear to actively seek promotions but show little interest in private label products.

Cluster 3: Moderate spenders, snackers & PL lovers (14%)**.** They prefer snacks over cooking and favor private label products.

Cluster 4: Moderate spenders, mostly nonfood customers (12%).

Cluster 5: Moderate spenders, balanced FNF customers (±80% of their amount is spent on food and ±20% on nonfood categories)**.** They prefer bulk over packaged products. Cluster 5 comprised 17% of the sample.

Cluster 6: High spenders, mostly food customers who seem to cook (15%).

Figure 4 near here

Comparing Online and Offline Customers

Online customers, even the most loyal ones, conduct significantly fewer transactions and purchase fewer products in total compared to customers of brick-and-mortar stores. Their monthly baskets are generally larger, but this could be attributed to the minimum

order amount required for making online purchases. Regarding private label products, we observed that proportionally, online consumers purchase slightly more private label products and fewer items on promotion. This conclusion is also supported by the research of Verstraeten et al. (2023). It was also found that they purchase more packaged products, paper goods, household cleaners, refreshments from the refrigerator, juices and water. Brick-and-mortar customers purchase proportionally more bulk products such as cheese, cold cuts, and butcher items, as well as more sugary and fresh bakery items.

Comparing the clusters resulting from the HCPC analysis for online and brick-and-mortar customers, we observed that in the first clusters of Infrequent/Occasional shoppers, the main distinguishing factor is their preference for Private Label products and items on promotion. Among the clusters of moderate spenders, we notice differentiation based on the preference for non-food items. Specifically, among brick-and-mortar customers, a cluster of snackers and PL lovers emerged. This behavior may possibly result from impulse and unplanned purchases. The fact that this is not observed among online customers is considered to confirm the findings of the study by Huyghe et al. (2017), which suggests that online consumers purchase fewer unhealthy products. Online high spenders purchase all product categories and are balanced between food and nonfood products, whereas offline high spenders, conduct purchases with higher frequency than online, and they are mostly food customers who seem to cook.

Conclusion

The above findings are valuable for developing targeted marketing strategies, conducting personalised promotional actions and optimizing the online shopping experience. Insights into customers' buying habits, preferred product categories, and shopping frequency can guide inventory management, pricing strategies, and

promotional campaigns. The company could apply attraction marketing strategies to infrequent customers, retention and growth marketing strategies to the moderate spenders and retention strategies to the high spenders. Regarding purely online customers, the results of the HCPC show that the three clusters of infrequent customers constitute 60% of the sample. They show much lower loyalty than physical store customers, therefore the company could aim at enhancing their visitation and loyalty. For example, we see that loyal customers purchase PL products at a rate of approximately 20%, so in cluster 3, that dislikes PL and constitutes almost 1/3 of the sample, offering free PL samples could enhance their trust. Clusters 1 and 3 are promo hunters, so strategies like engagement through gamification and reward challenges, abandoned cart discounts, and exclusive online offers could be applied. For clusters 2 and 4, which are non-food customers, cross-selling suggestions and recipe ideas could encourage them to explore and purchase food products, ultimately increasing their overall spending and engagement with the brand. For moderate spenders, cross-selling suggestions can be employed, recommending complementary products at checkout, reinforcing impulse buying by encouraging them to add more items to their basket. Finally for high spenders, VIP programs offering exclusive privileges and offers could be implemented. As a general conclusion, by delving deeper into this emerging market segment, grocery chains can capitalize on the growing trend of online shopping and stay competitive in the evolving retail landscape.

References

Benn, Y., Webb, T. L., Chang, B. P. I., & Reidy, J. (2015). What information do consumers consider, and how do they look for it, when shopping for groceries online? *Appetite*, *89*, 265–273. https://doi.org/10.1016/j.appet.2015.01.025

Chu, J., Arce-Urriza, M., Cebollada-Calvo, J.-J., & Chintagunta, P. K. (2010). An Empirical Analysis of Shopping Behavior Across Online and Offline Channels for Grocery Products: The Moderating Effects of Household and Product Characteristics. *Journal of Interactive Marketing*, 24(4), 251-268. https://doi.org/10.1016/j.intmar.2010.07.004

Dawes, J., & Nenycz-Thiel, M. (2014). Comparing retailer purchase patterns and brand metrics for in-store and online grocery purchasing. *Journal of Marketing Management*, *30*(3–4), 364–382. https://doi.org/10.1080/0267257X.2013.813576

Degeratu, A. M., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, *17*(1), 55–78. https://doi.org/10.1016/s0167-8116(00)00005-7

Galante, N., López, E. G., & Monroe, S. (2013). The Future of Online Grocery in Europe. *McKinsey & Company*, 22–31.

Greenacre, M. J. (2013). Correspondence analysis. The Oxford Handbook of Quantitative Methods. Statistical Analyses, 2, 142-153.

Hanus, G. (2016). Consumer Behaviour During Online Grocery Shopping. *CBU International Conference Proceedings*, *4*, 010–013. https://doi.org/10.12955/cbup.v4.737

Huyghe, E., Verstraeten, J., Geuens, M., & Van Kerckhove, A. (2017). Clicks as a healthy alternative to bricks: How online grocery shopping reduces vice purchases. *Journal of Marketing Research*, *54*(1), 61–74. https://doi.org/10.1509/jmr.14.0490

Manca, F., D'Uggento, A. M., & Convertini, N. (2018). Customer segmentation through multiple correspondence analysis. 2018 110th AEIT International Annual Conference, AEIT 2018, October 2020. https://doi.org/10.23919/AEIT.2018.8577279

Matta, E., Stalidis, G. (2023). Profiling online and physical supermarket customers using Factor and Clustering Methods. *International Conference Marketing and Technologies*, Prague 2023

Verstraeten, J., Heeremans, E., Geuens, M., & Vermeir, I. (2023). How online grocery shopping drives private label food purchases. *Journal of Business Research*, *167*(May), 114057. https://doi.org/10.1016/j.jbusres.2023.114057

Stalidis, G., Theodosios S., Kaplanoglou, P. I., Katsalis, A., Karaveli, I., Delianidi, M. and K. Diamantaras. "Multidimensional Factor and Cluster Analysis Versus

Embedding-Based Learning for Personalized Supermarket Offer Recommendations." In *Data Analysis and Rationality in a Complex World 16*, pp. 273-281. Springer International Publishing, 2021.

Zatz, L. Y., Moran, A. J., Franckle, R. L., Block, J. P., Hou, T., Blue, D., Greene, J. C., Gortmaker, S., Bleich, S. N., Polacsek, M., Thorndike, A. N., Mande, J. R., & Rimm, E. B. (2021). Comparing shopper characteristics by online grocery ordering use among households in low-income communities in Maine. *Public Health Nutrition*, *24*(15), 5127–5132. https://doi.org/10.1017/S1368980021002238
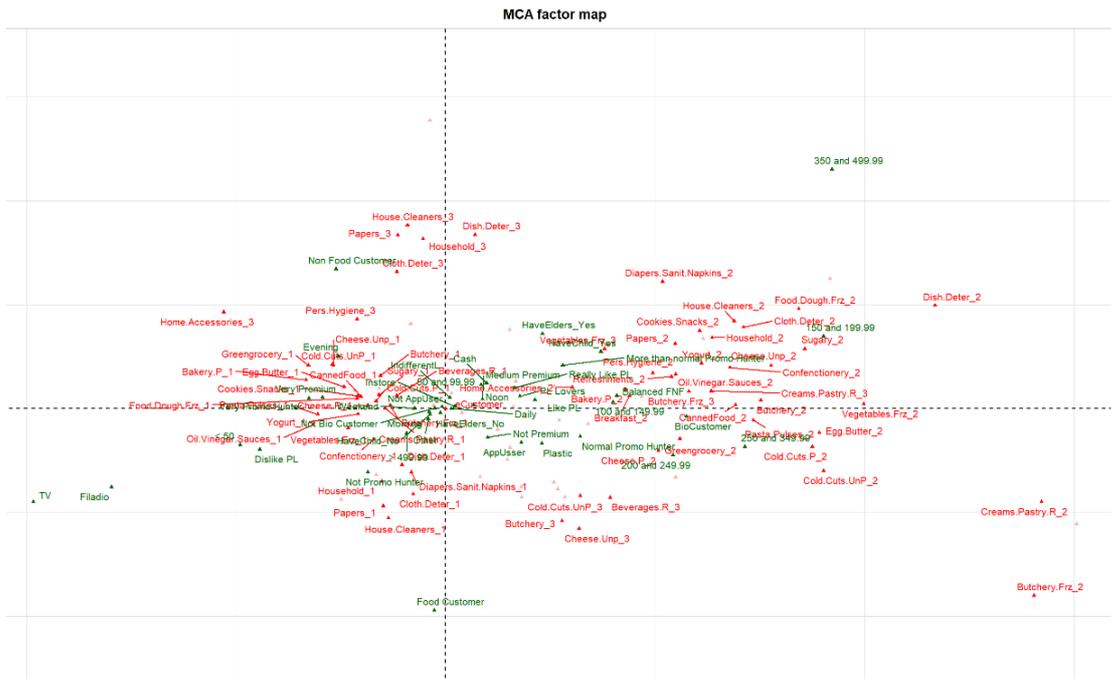
Figure 1. Representation of significant categories on the factorial plane F1 x F2. Preferences for product groups (active variables) appear in red and behavioural indicators (supplementary variables) appear in green.

Cluster	Cust Count	%Cust	Dist Items	Qty Total	Total Transactions	Monthly Basket	%Items on promo	%Private Label Products
1	113	18%	10.96	46.13	2.1	64.85	32%	28%
2	80	13%	23.53	76.19	2.35	65.04	23%	13%
3	180	29%	37.26	72.21	2.78	70.61	28%	16%
4	133	22%	67.53	129.16	3.98	100.37	22%	20%
5	65	11%	90.23	229.49	6.32	122.8	23%	18%
6	46	7%	188.67	501.18	16.2	146.04	23%	20%
Total	617	100%						

Table 1. Cluster descriptive statistics for online customers.
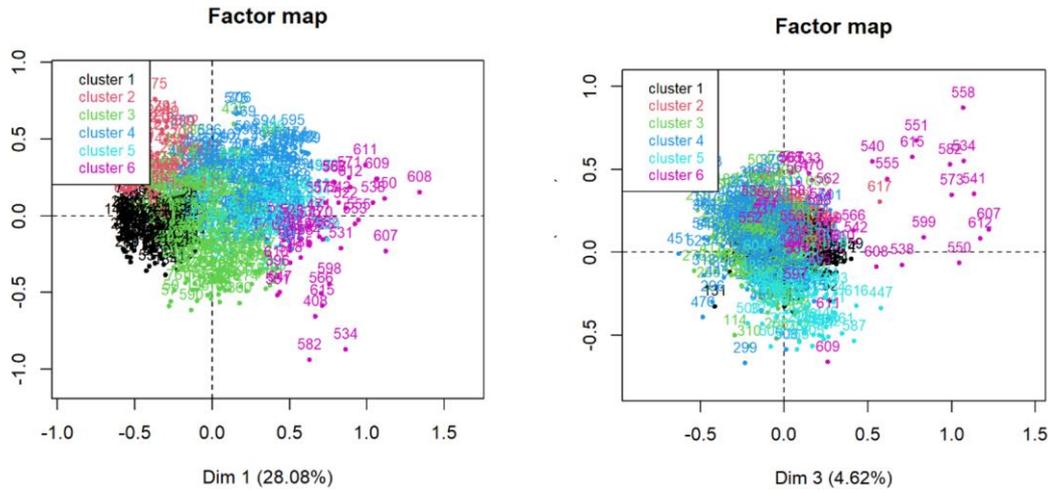
Figure 2. Projection of online clusters on the factorial planes.
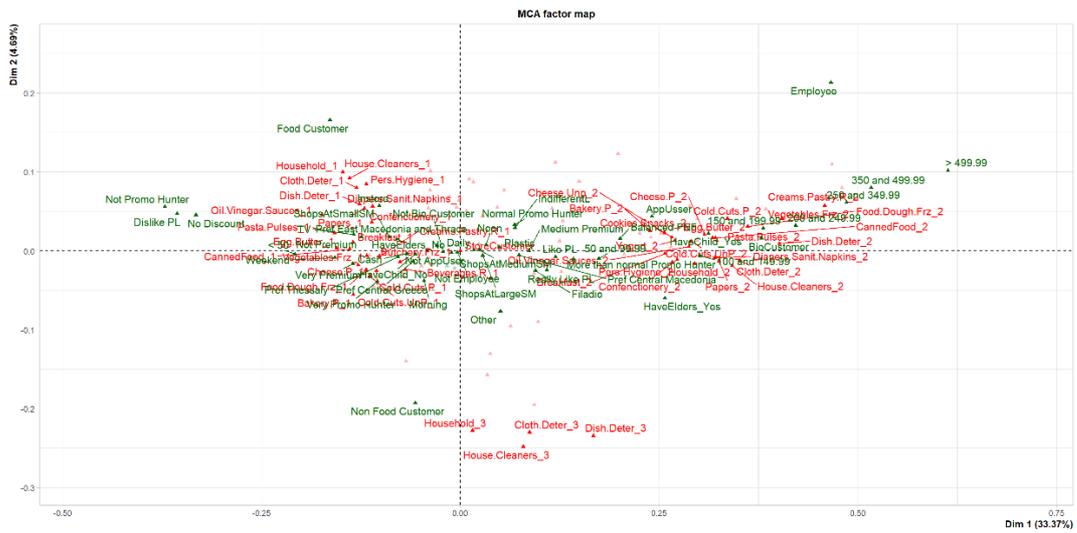
Figure 3. Representation of active and supplementary categories on the factorial plane F1 x F2 for brick-and-mortar customers.

Cluster	Cust Count	%Cust	Dist Items	Qty Total	Total Transactions	Monthly Basket	%Items on promo	%Private Label Products
1	14.353	24%	19.76	38.93	6	25.67	29%	18%
2	10.496	18%	55.67	123.94	15	43.68	27%	14%
3	8.483	14%	124.66	305.99	33	76.50	24%	21%
4	7.356	12%	126.02	288.35	30	69.12	26%	20%
5	10.040	17%	201.80	521.74	57	97.61	26%	18%
6	9.178	15%	301.04	750.08	70	138.69	26%	19%

Total	59.906	100%

Table 2. Cluster descriptive statistics for Brick-and-Mortar customers.
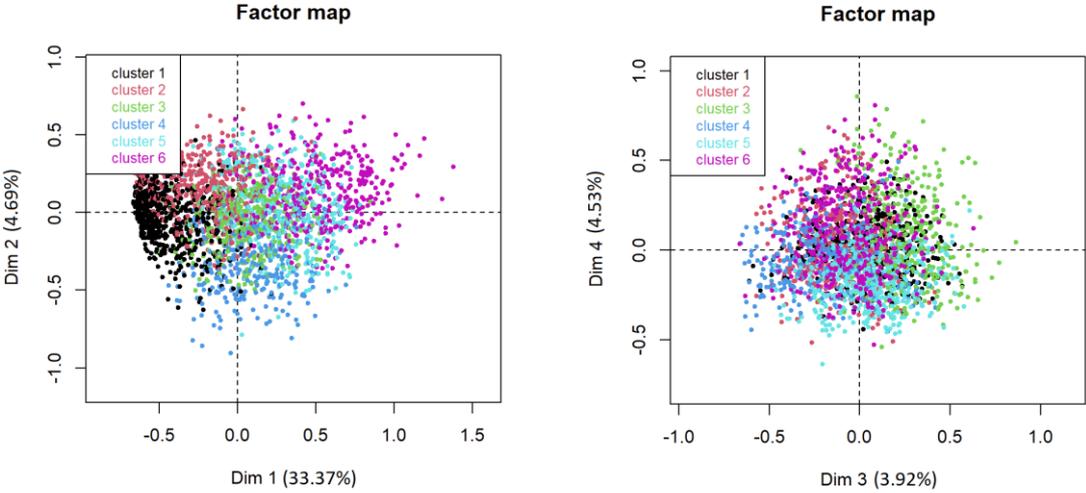

Figure 4. Projection of brick-and-mortar clusters on the factorial planes.