Mining Segmentation Patterns Using e-Commerce Retail Data: An Experience Report

Anastasia Griva, Denis Dennehy

Mining segmentation patterns using e-commerce retail data: An experience report

Anastasia Griva[1[0000-0001-8346-9389]] and Denis Dennehy[1[0000-0001-9931-762X]]

[1] Lero - The Science Foundation Ireland Research Centre for Software, NUI Galway
Galway, Ireland
anastasia.griva@nuigalway.ie

Abstract.
The goal of this experience report is to study and advance our understanding on visit and shopper segmentation in retail. Current segmentation studies mainly use customers as unit of analysis to identify shopper segments and mine patterns in their behaviors. Another stream of studies utilizes shopper baskets or visits to perform basket/visit segmentation and elicit shopping patterns. However, given the fact that we live in the era of personalization focusing solely on visits leads on neglecting the shopper. On the other hand, focusing solely on shoppers and their behavior over time, leads on losing his/her daily purchasing behavior. In this study we combine shopper and visit segmentation and we apply data mining to identify purchase patterns using data from an e-commerce grocery retailer.

Keywords: retail analytics, business analytics, data mining, e-commerce.

1 Introduction

In contemporary retail, both practitioners and researchers agree that old-school shopper segmentation is not enough and cannot describe the new shopper habits and preferences. They suggest that retail nowadays, demands a transformation of shopper segmentation approaches. This happens since the modern shopper has changed. Nowadays, the shopper performs a complex shopper journey with the purpose to satisfy his/her increasing demands [1]. Shopper behavior is no longer predictable; it is changing through time and, even, between shopping visits in the same store [2]. Thus, there is a need to focus on shoppers' visits and perform visit segmentation to cope with shoppers' changing behavior and perform data mining to identify segmentation patterns.

Practitioners have coined the term "shopping mission" to refer to the intention behind a shopper's visit [3][4]. Similarly, researchers [5], [6] agree with practitioners and suggest that we should pay attention to each shopper visit which reflects the actual shopper needs. Looking at each specific shopper visit, instead of a shopper's total buying behavior over many visits, provides a better view of the shopper desires that change frequently due to unpredicted events e.g., COVID-19 pandemic. Approaches and studies that respects the different views of shoppers and considers their heterogenous behaviors, are becoming crucial [7].

Looking into the segmentation literature we can classify the pertinent studies in two categories: those focusing on shopper segmentation, and those that focus on visit segmentation. The first group of studies usually perform clustering as the data mining approach to identify patterns on everything a shopper has purchased in bulk, and overlook the shopping intentions, and missions of each shopper, which by nature are not the same in every store visit. Regarding the second group of studies, they examine shoppers' visits, usually by using association rule mining to perform visit segmentation [1], [8]–[11] e.g. a form of visit segmentation is market basket analysis.

In this study we combine visit and shopper segmentation, as we utilize visit segments (or else shopping missions) to then perform customer segmentation and identify the characteristics of the derived customer segments. To the best of our knowledge, there is paucity of studies that perform visit segmentation and use the derived segments to further extract and identify shopper segment. Current studies either perform visit segmentation, or shopper segmentation.

Below we describe the data used to perform visit and shopper segmentation, and the respective segmentation results. Then, we conclude our practitioners' paper via presenting the practical implications of our study

2 Background

3 Data analysis

3.1 The dataset

We used point-of-sales (POS) data from the e-commerce store of a small retail chain named "XYZ" having 222 stores in the urban areas of Greece country. Retailer provided us with all the transactions that had been performed using their website within a year by retailer's shoppers using loyalty card. Thus, we received more than 120 billion product records that correspond to 15 billion transactions/baskets performed by 1.120.021 shoppers. A sample of the given dataset is provided in Table 1. Here, we should admit that almost the 96% of the total retail chain transactions happen using loyalty card; thus, sample is representative. The average basket costed 16,7€ and contained 8,1 products from 4,9 different product categories.

Table 1. Data sample

Basket_ID	Date	Barcode	Sum_Units	Sum_Value	Card_ID
1103084867	1/3/2021	800220505783	2	1.96	9160003751260
1103084867	1/3/2021	520139501183	1	5.349993	9160003751260
1092750793	1/3/2021	520423907421	6	1.740015	9164012915385
1106160983	1/3/2021	211069400000	1	0.749817	9162005811409
1108695491	1/3/2021	520286400380	-2	-0.6	NULL

3.2 The analysis

In this section first we perform visit segmentation to identify customers' shopping missions or else the visit segments, and then we utilize these visit segments to perform shopper segmentation. In both the segmentation phases we use clustering as a data mining technique to derive the shopping patterns.

Visit segmentation/shopping missions

Drawing on the method suggested by (Griva, Bardaki, Pramatari, & Papakiriakopoulos, 2018; Griva, Bardaki, Pramatari, & Doukidis, 2021), we identified customers shopping missions or else the visit segments. In a nutshell, we conducted clustering with expectation maximization algorithm on the product categories the shoppers purchased on each visit. This way we identified 7 distinct shopping missions, as shown in Fig. 1.

Main course		**Snacks and beverages**		**Pastry making**		**Health and beauty**		**Breakfast**		**House cleaning and maintenance**		**Sandwich**	
fresh vegetables	62 %	**beverages**	**35 %**	**kit desserts**	**27 %**	**shower gel**	**63 %**	**coffee**	50 %	**house cleaning**	**65 %**	**counter cheese**	63 %
counter red meat	**54 %**	**salty snacks**	24 %	**sugar**	**25 %**	**shampoo**	**50 %**	cereals	31 %	**paper tissue and rolls**	**61 %**	**deli counter**	61 %
counter poultry	**50 %**	**chips**	22 %	fresh milk	25 %	oral care	45 %	marmalade	27 %	laundry	59 %	**bread slices**	47 %
counter- fishmonger	**47 %**	**beers**	**21 %**	**flour**	**24 %**	women haircare	41 %	toast bread	23 %	dishwashers	56 %	pies	19 %
bread slices	44 %	biscuits	20 %	confectionary	24 %	facecare	29 %	packed sliced cheese	22 %	food storage	55 %	packed sliced cheese	16 %
pasta	27 %	spirits	20 %	cocoa	23 %	deodorants	25 %	yogurt - desserts	20 %	house and garden	14 %	packed ham slices	15 %
packed salad	26 %	sweets	19 %	coffee	23 %	sanitary protection	13 %	cookies	19 %	linen	13 %	bread	10 %
bread	20 %	deserts	18 %	culinary aids	21 %	bodycare	13 %	fresh milk	17 %	paper- schools	13 %	packed salads	5 %
counter cheese	18 %	watter	17 %	margarine	20 %	baby care	12 %	long-life milk	16 %	DIY and car	12 %		
fresh fruits	18 %	juices & smoothies	16 %	eggs	20 %	conditioner	9 %	bakery sweets	15 %	insecticides	11 %		
deli counter	15 %	wines	16 %	butter	20 %	clothing	8 %	juices & smoothies	14 %				
tinned tomatoes	14 %	fresh milk	15 %	spices and herbs	17 %	make up	7 %	honey	14 %				
biological vegetables	12 %	fresh vegetables	13 %	milk cream - sandy	17 %	men haircare	7 %						
table sauces	10 %	beverages	11 %	sweeteners	16 %	perfumes	5 %						
frozen vegetables	10 %	processed fruits	8 %	chocolates	16 %	accessories	4 %						
butter	9 %	empty cans	6 %	powder milk	7 %	vitamins	3 %						
oils and fats	9 %	party equipment	5 %										
dressing	9 %												
rice	8 %												
seafood	7 %												
ethnic food	5 %												
white cheese	4 %												
flour	3 %												

Fig. 1. Resulting visit segments/ shopping missions

Shopper segmentation based on the visit segments

Afterwards, we exploited the loyalty cards data, to detect the visit segments or else the shopping missions a shopper performed during his/her purchase history. We rated each shopper (Table 2) with a value ranging from 1 (low) to 5 (high) according to the s/he visited the store for each mission during the whole year weeks (weeks of presence).

In more detail, to calculate this value we used as benchmark the weeks of presence of all the shoppers per mission. Thus, this value is different for the various shopping missions. For example, a value equals 5 at the breakfast mission is not the same with a 5 into house cleansing and maintenance, as a shopper purchases more often breakfast than house cleansing products.

We used text mining to extract more meta-data from the available product descriptions. Thus, we identified whether a product is premium, private label (PL), and/or biological. Also, we classified the products based on their descriptions into children or elder usage. In addition, we added a binary flag in the cases that the product sold was in promo. Hence, we enriched our data mining table (i.e., fact table) by adding more meta-data per customer. Then, we utilized shoppers' fact table as input in the data mining model,

we executed clustering with k-means algorithm, and we segmented shoppers based on the missions they had performed during their yearly visits.
We shaped ten shopper clusters i.e., ten shopper segments. Each shopper is assigned to one shopper segment. Segment 1 contains the 4,6% of shoppers. These shoppers visited the store for all the identified shopping missions as they are rated with five in all of them. Thus, here we have the more loyal shoppers that spend more than 281,83€ per month. These shoppers purchase many private label (PL), baby and biological products. Similarly, segment 2 contains loyal shoppers, as they are rated with five in almost all the missions. This leads them spend almost 100€ less (197,58€) than the previous segment. In addition, shoppers in this segment purchase more premium products. Closing, segments 1 and 2 could be declared as "the most loyal retailer's shoppers" that seem to have this retail chain as primary for their purchases. Thus, customer retention strategies could be more appropriate for these two segments.
Similarly, we identified the rest shopper segments based on the identified shopping missions. Table 2 presents the resulting shopper segments and their characteristics. Based on the descriptive statistics derive for each segment and the patterns identified regarding the purchased shopping missions we inferred shoppers' loyalty per segment.

Table 2. Segment characteristics

Segment No	Size	Average monthly value	Shoppers' loyalty classification	Segment characteristics
1	4,6%	281,83	Very high loyalty	Purchase all the shopping missions. Preference in bio, baby and PL products.
2	4,9%	197,58	Very high loyalty	Purchase all the shopping missions. Preference in premium products.
3	5,4%	121	High loyalty	Purchase all shopping missions but seem to visit a second retail chain for non-food products. Preference in PL products, elderly products,
4	5,1%	128,2	High loyalty	Purchase all shopping missions but seem to visit a second retail chain for food products.
5	9,1%	85,6	Medium loyalty	This retailer is not their primary choice. Purchase occasionally all shopping missions. Prefer baby and bio products.
6	6,9%	59,48	Medium loyalty	This retailer is not their primary choice. Purchase only food shopping missions.
7	10,9%	61,65	Medium loyalty	This retailer is not their primary choice.

				Purchase only non-food shopping missions. High preference in promo products.
8	14,0%	53,95	Low loyalty	Purchase missions such as breakfast, sandwich and personal care and hygiene. Visit stores close to student campuses.
9	18,1%	24,33	No loyalty	Purchase only promotional products from various shopping missions.
10	21,0%	26,05	No loyalty	Have visited retailer's store only 2 times per year and purchase premium products across all missions that were in promo

4 Conclusions

Business analytics can be powerful for companies in various domains ranging from software development [12] to retail [1]. In this study, we focused on retailing and we analyzed data from the e-commerce store of a grocery retailer. Our goal was to combine shopper and visit segmentation to showcase the potential of retail analytics. First, in the first phase of our analysis, we used a method to identify visit segments and patterns in shopper visit behaviors from the available data. Then, in our second phase we used these segments combined with loyalty and other data, and we applied data mining through clustering to identify loyalty-based shopper segments. From a research perspective these is lack of studies that combine shopper and visit segmentation. Researchers mainly focus on the first, and more recently there are some examples related to the second.

As an experience report, below we focus on the practical implications this study offers. The practical value of this work is stressed when considering the business decisions, it can support. This analysis can be used and evolves into a tool for designing bundled promotions for product categories belonging to the same segment. For example, retailers may plan cross-coupon programs for addressing the needs of customers visiting the store with a specific purpose in mind. Alternatively, it can be used as the basis of a recommendation system for real-time purchases in retail stores. It can suggest to the customers products that they might have forgotten to buy, considering their prior or current visit(s).

Similarly, we can create offline and online product catalogues for specific visit segments. The extracted knowledge could also be valuable for advertising purposes; for instance, instead of making advertisements of specific product categories, retailers could advertise bundled product categories that correspond to a shopping mission, e.g., breakfast products advertisements.

On the other hand, the customer visit segments can dictate a new redesigned retail store physical or web store layout. For example, the product categories in the same visit segment could be positioned in nearby store aisles and shelves. Considering the bigger picture, we can move from a category-based layout to a mission-based layout that can help customers locate products in the store more easily and buy more in less time [8], [13]. Alternatively, second in-store placement spots can be detected.

Further, the value of such a system could be further enhanced when we use the resulting visit segments to perform shopper segmentation. By looking into the shopping missions that each shopper performs in all the stores of a retail chain, we can boost shopper marketing activities. Shopper segmentation based on the identified shopping missions can aid retailers identify selling gaps and opportunities and enhance personalization.

The results of Table 2 can be of significant value to the marketers. Based on these they can build retention strategies for the very high loyal segments, customer development strategies for the high and medium loyal segments, and customer attraction strategies for the less loyal segments. This way they can approach and treat the various loyalty segments they have differently, based on their actual behavior and needs.

Additionally, the store manager could reengineer store operations management and replenishment strategies by ordering groups of products based on the identified visit segments. Additionally, this approach could be even utilized to rearrange and modify a retailer's warehouse, by placing in nearby aisles products matching online orders to decrease order-picking time.

As it is obvious our study includes several limitations, which can all be considered and resolved in future studies. For instance, in this study we extract patterns using data from the e-commerce website of a grocery retailer, however, we neglect that some shopping gaps might be fulfilled in retailer's physical stores etc. Thus, from a marketing perspective future research may examine the omni-channel shopper behaviors. Moreover, from a technical perspective, future research may focus on mining behavioral shopping patterns using other data mining techniques such as graph mining etc. Closing, from an IS perspective future studies may focus on how retailers can make sense of business analytics and segmentation [14].

Acknowledgements

This research was funded by the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant 754489 and from the Science Foundation Ireland grant 13/RC/2094_2.

References

[1] A. Griva, C. Bardaki, K. Pramatari, and G. Doukidis, "Factors Affecting Customer Analytics: Evidence from Three Retail Cases," *Inf. Syst. Front.*, 2021.

[2] H. Sorensen *et al.*, "Fundamental patterns of in-store shopper behavior," *J. Retail. Consum. Serv.*, vol. 37, no. 2017, pp. 182–194, 2017.

[3] A. Griva, "Data-driven innovation in shopper marketing: a business analytics approach for visit segmentation in the retail industry," Department of Management Science & Technology, Athens University of Economics and Business, 2019.

[4] A. Griva, C. Bardaki, P. Sarantopoulos, and D. Papakiriakopoulos, "A Data Mining-Based Framework To Identify Shopping Missions," in *MCIS 2014*

Proceedings, 2014.
[5] D. R. Bell, D. Corsten, and G. Knox, “From Point of Purchase to Path to Purchase: How Preshopping Factors Drive Unplanned Buying,” *J. Mark.*, vol. 75, no. 1, pp. 31–45, 2011.
[6] R. G. Walters and M. Jamil, “Exploring the relationships between shopping trip type, purchases of products on promotion, and shopping basket profit,” *J. Bus. Res.*, vol. 56, no. 1, pp. 17–29, 2003.
[7] R. T. Rust and M.-H. Huang, “The Service Revolution and the Transformation of Marketing Science.,” *Mark. Sci.*, vol. 33, no. 2, pp. 206–221, Mar. 2014.
[8] I. Cil, “Consumption universes based supermarket layout through association rule mining and multidimensional scaling,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 8611–8625, 2012.
[9] N. Beck and D. Rygl, “Categorization of multiple channel retailing in Multi-, Cross-, and Omni-Channel Retailing for retailers and retailing,” *J. Retail. Consum. Serv.*, vol. 27, pp. 170–178, 2015.
[10] Y. Boztuğ and T. Reutterer, “A combined approach for segment-specific market basket analysis,” *Eur. J. Oper. Res.*, vol. 187, no. 1, pp. 294–312, 2008.
[11] A. Griva, C. Bardaki, K. Pramatari, and D. Papakiriakopoulos, “Retail business analytics: Customer visit segmentation using market basket data,” *Expert Syst. Appl.*, vol. 100, no. 2018, pp. 1–16, 2018.
[12] K. Conboy, P. Mikalef, D. Dennehy, and J. Krogstie, “Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda,” *Eur. J. Oper. Res.*, vol. 281, no. 2020, pp. 656–672, 2020.
[13] P. Sarantopoulos, A. Theotokis, K. Pramatari, and A. L. Roggeveen, “The Impact of a Complement-Based Assortment Organization on Purchases,” *J. Mark. Res.*, pp. 1–20, 2019.
[14] E. D. Zamani, A. Griva, K. Spanaki, P. O’Raghallaigh, and D. Sammon, “Making sense of business analytics in project selection and prioritisation: insights from the start-up trenches,” *Inf. Technol. People*, Apr. 2021.