



Available online at [www.alphanumericjournal.com](http://www.alphanumericjournal.com)

# alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 9, Issue 1, 2021



Received: June 13, 2020  
Accepted: January 26, 2021  
Published Online: June 30, 2021

AJ ID: 2021.09.01.MIS.01  
DOI: 10.17093/alphanumeric. 752505  
**Research Article**

## Market Basket Analysis of Basket Data with Demographics: A Case Study in E-Retailing

Ural Gökay Çiçekli, Ph.D. 

Assoc. Prof., Faculty of Economics and Administrative Sciences, Ege University, Izmir, Turkey, [gokay.cicekli@ege.edu.tr](mailto:gokay.cicekli@ege.edu.tr)

İnanç Kabasakal, Ph.D. \* 

Res. Assist., Faculty of Economics and Administrative Sciences, Ege University, Izmir, Turkey, [inanc.kabasakal@ege.edu.tr](mailto:inanc.kabasakal@ege.edu.tr)

\* Ege Üniversitesi İktisadi ve İdari Bilimler Fakültesi 1. Kat No:126 Bornova İzmir, Türkiye

### ABSTRACT

Businesses overcome with a high degree of competition that necessitates customer-focused strategies in most industries. In a digitalized business environment, the implementation of such strategies often requires the analysis of customer data. Market basket analysis is a well-known method in marketing that examines basket data to discover useful information about customers' purchase intentions. The analysis has been a playground for data mining researchers that aim to overcome with its practical challenges. Our study extends the conventional basket analysis by incorporating demographic variables along with purchase transactions. With such modification, we provide an example for the extraction of segment-specific rules that relate product-level purchase decisions with gender, location, and age group. For this purpose, we present a case study on monthly basket data obtained from an e-retailer in Turkey. Our findings demonstrate association rules that might guide marketing practitioners who need to discover segment-specific purchase patterns to designate personalized promotions.

### Keywords:

Association Rule Mining, Data Mining, Demographic Association Rules, Market Basket Analysis



## 1. Introduction

In a competitive environment, businesses aim to satisfy their customers while maintaining profitable relationships in the long run. Retaining existing customers with customer-centric practices is crucial for businesses due to the high cost of acquiring new customers (Bodapati, 2008). Customer Relationship Management (CRM) is often considered as a business strategy to achieve long-term customer satisfaction through a customer-oriented approach.

CRM strategy focuses on relationships established with customers and necessitates analyzing customer activities to understand their needs and behaviors (Winer, 2001). With this perspective, a business needs to keep track of its customers' actions, explore what they like, analyze their purchase patterns, and investigate their customers' needs, preferences, and behaviors (Tsiptsis & Chorianopoulos, 2009). In this regard, organizations have been utilizing information to establish long-lasting profitable relationships with their customers by offering better services (Sota et al., 2018).

Over the last decades, e-commerce has transformed how businesses and consumers interact with each other. As a side effect, marketing efforts for online shopping had become more and more critical for businesses (Aksoy, 2008). IT infrastructure facilitates capturing, storing, and rapidly accessing customer data, and provides a useful framework to implement data-driven marketing strategies. The need to extract useful information from large data stacks has led to the introduction of data mining techniques (Han et al., 2012).

Data mining is a semi-automatic process to discover interesting patterns and statistically significant structures within datasets (Vahaplar & İnceoğlu, 2001). This process involves the use of techniques for deriving meaningful and useful information within large and unstructured datasets (Dalal, 2012; Ching & Pong, 2002) and help to estimate future trends that affect decision-makers (Hudairy, 2004). In a broader context, data mining is often described as an essential step in the Knowledge Discovery Process (Bramer, 2016:2); which involves cleaning, integration, selection, and transformation of data before the application of data mining algorithm, followed by pattern evaluation and presentation (Han et al., 2012). Data mining techniques allow the discovery of previously unknown patterns and relationships within large data sets (Linoff & Berry, 2011; Marakas, 2003) and are widely employed in interdisciplinary studies across various domains (Savaş et al., 2012), including finance, e-commerce, medicine, business, and education. In this manner, data mining field stands at the intersection of computer science, machine learning, database management, applied mathematics, and statistics (Emel & Taşkın, 2005).

Data mining applications help to estimate future trends that affect decision-makers (Hudairy, 2004). The patterns revealed through data mining might help marketers to decide on the marketing mix, to determine new product opportunities, and to predict customer behaviors (Strauss & Frost, 2009). Moreover, segmentation, classification, and forecasting techniques in data mining are applied in various business problems to leverage data to implement customer-oriented strategies.

Among a variety of industries and sectors, retailing is one of the hotspots for data mining applications. In a study by Anderson et al. (2007), it was emphasized that retailers that aim to respond to the needs of their customers shift towards data-driven decisions. Data mining techniques have been widely adapted for retailing problems, including cross-selling, market basket analysis, risk management, fraud detection, customer acquisition, customer retention (Bala, 2008), shelf placement, and stock management (Hormozi & Giles, 2004). Basket data has been a valuable resource for data mining applications in retailing, and often analyzed for customer segmentation and extraction of purchase patterns (Griva et al., 2018). Notably, market basket analysis has been often addressed as a data mining problem with an objective to discover relationships, or association rules, which represent hypotheses on purchase intentions. Such findings have been used exploited in facility layout design (Halim et al., 2019) and online/mobile recommendations (Osadchiy et al., 2019) with the objective of increasing cross-sales.

As remarked in (Dippold & Hruschka, 2013a), prior research often focuses on cross-category purchase decisions extracted from basket data, where most exploratory models exclude demographics and marketing mix elements. However, it might be argued that the technique might be further applied to other product attributes in the retailing context, including brand (Kabasakal, 2020). Kooti et al. (2016) conducted a study to analyze the differences in customer purchase decisions and emphasized that an extensive count of attributes, including geo-location and demographics, might help to predict online shopping decisions. Moreover, Zhang and Pennacchiotti (2013) noticed that social media profile data, including gender and age, are useful in predicting purchase decisions in e-commerce. Due to their influence on purchase behaviors, customer profiles and demographics have been further utilized for product recommendations. With this motive, our study aims to analyze basket data and demographics together using the association rule mining technique. Along with a case study, our study presents association rules which help to relate purchase intentions with demographic variables. In the following sections, an introduction of market basket analysis and rule mining technique is provided. Subsequently, our case study is presented. Our findings involve prominent rules that were initially chosen by interestingness measures, then categorized with the inclusion of demographic attributes.

## 2. Market Basket Analysis

Market basket analysis is a popular technique for marketers that might be useful to designate customer-focused strategies for businesses (Özçakır & Çamurcu, 2007). The analysis extracts clues on customers' purchase intentions by discovering interrelated categories. The analysis depends on the assumption that the customers' purchase decisions across product categories might not be independent, thus follow similar patterns (Dippold & Hruschka, 2013b).

From a broader perspective, market basket analysis might be categorized into two types of models; exploratory models are designed to discover cross-category purchase patterns while explanatory models explore the effects of marketing mix variables right after the extraction of purchase patterns (Solnet et al., 2016).

However, the use of the analysis in e-retailing is often aimed at the discovery of cross-category purchase patterns for improved recommendations.

The findings in a market basket analysis typically imply which products could be sold together. Such results typically involve complementary products (Winer, 2001). Purchase patterns discovered by the analysis might be utilized to designate sales promotions (Chen et al., 2006). Furthermore, e-commerce web sites might provide relevant products for online users instantaneously.

The discovery of frequently purchased items has been commonly revisited as a frequent pattern mining problem in data mining studies (Aggarwal, 2015). Apriori algorithm (Agrawal et al., 1993) is a well-known algorithm proposed by R. Agrawal to discover the association rules that represent purchase patterns in a supermarket dataset. The primary advantage of the algorithm lies in the ability to scale for large input sizes efficiently (Kronberger & Affenzeller, 2011). Alternatively, Eclat (Equivalence Class Transformation) algorithm is widely employed in rule mining due to its high performance in smaller datasets (Şimşek-Gürsoy et al., 2019). As another alternative, the FP-Growth algorithm generates FP-trees to achieve better data compression in item-set discovery (Kotu & Deshpande, 2015:206).

The purchase behaviors are formulated as association rules. An association rule  $X \rightarrow Y$  for two sets of items  $X$  and  $Y$  represents a purchase pattern that indicates the purchase of  $Y$  along with  $X$ . In such representation, the antecedent ( $X$ ) and the consequent ( $Y$ ) denote discrete sets of items.

The significance of association rules and item-sets is assessed with several measures. The support criterion for a set of items indicates the fraction of all records that involve that set of items (Aggarwal, 2015). The confidence measure is used in rule mining to assess the importance of a rule. The confidence for the rule  $X \rightarrow Y$  is calculated by the ratio of transactions that involve  $X$  and  $Y$  together among all observations that involve  $X$ , as in the following (Aggarwal, 2015):

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

The support and confidence are among the criteria that indicate the usefulness of association rules (Bayardo Jr & Agrawal, 1999). Rules measured over a threshold for both measures are often described as 'strong' in most studies. On the other hand, high confidence score might be misleading in some cases; thus, the lift measure that signifies the correlation among the itemsets is useful to choose interesting rules (Han et al., 2012). The lift measure for the rule  $X \rightarrow Y$  can be formulated as follows:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)}$$

### 3. Methodology

In most studies on market basket analysis, rules are entirely discovered from the purchase history. Particularly, the rule mining technique (Agrawal et al., 1993) handles transactional data with binary attributes where each attribute signifies the purchase of a product. Our approach extends the basket data by integrating additional variables about the customers and the orders. However, our study sticks to the original Apriori algorithm for rule mining, after several steps of data preparation.

Market basket analysis typically formulates products within baskets in binary form regardless of the quantity, and such relations are occasionally represented in a table of binary relationships. In Table 1, each basket is represented as a column, and each row corresponds to an item (product or category). The cells at the intersection of columns and rows hold binary values. In this notation, a value of 1 corresponds to the presence of an item within a basket, and 0 corresponds to the opposite.

To examine demographic variables in rule mining, we split categorical values into binary attributes to form a binary table. For this purpose, subsequent rows were created for each value in demographic variables. Obviously, this requires matching baskets with a single value for each categorical variable.

		Code	Description	Baskets			
				B <sub>1</sub>	B <sub>2</sub>	....	B <sub>i</sub>
Product Category (340 categories)		7001	%100 Fruit Nectar	0	1		0
				1	0		1
		...	...	...	...		...
		7365	Olive Paste	0	1		0
Demographics	Gender	1	Female	1	0		1
		2	Male	0	1		0
	Age Group	11	<24	0	0		0
		12	25-34	1	0		0
		13	35-44	0	0		0
		14	45-54	0	1		1
	Location	15	55+	0	0		0
		41	(21 locations)	0	1		0
		42		1	0		1
		...		...		...	
61	0	0			0		

**Table 1.** Tabular representation of attributes analyzed

As listed in Table 1, the codes (1,2) represent gender. The numeric codes within the range [11-15] correspond to age groups, and the range [41-61] represents 21 delivery locations. Clearly, the numeric codes assigned do not overlap with product identifiers.

All items and transactions represented in the table above were stored in a relational database in SQLite. The transactional data initially involved purchased products for each basket, where each product was assigned with a category attribute. Accordingly, the products were reduced into product-categories to obtain generalized rules. Moreover, the basket-item relations were extended by including the demographic variables for analysis. For this purpose, a query was executed to prepare a combined transaction list that involves pairs of {Basket, Product Category}, {Basket, Gender}, {Basket, Age-Group} and {Basket, Location}. By this means, we assert that our transactional basket data represents the customers, even partially.

The subsequent step of our methodology involves the application of the Apriori algorithm to extract association rules. Rule mining was conducted by the 'RuleGenerator' software, which is a custom implementation of the Apriori algorithm introduced in Kabasakal (2020). The findings were separated into several groups based on the presence of demographics variables.

We suggest that the main advantage of our approach is the opportunity to identify behavioral purchase patterns by specific customer segments. Such rules can be exemplified as "women who purchase X also purchase Y", or "customers of age 25-

34 who purchase X also purchase Y". The resulting rules, including demographic variables on the left-hand side, might help to develop segment-specific offers. If a specific customer segment prefers a group of products more often than the others, our approach might demonstrate the difference with segment-specific association rules.

## 4. Case Study

In this study, we present a case study of a market basket analysis combined with demographic variables. The data was obtained from Adepo Sanal Market (adepo.com) for analysis. Since its foundation in 2001, adepo.com had been a substantial e-retailer in İzmir, Turkey. The company offered products of a variety of categories such as groceries, beverages, cleaning supplies, and household items; and had been in service until the end of December 2015. The dataset consists of 3163 purchase records, all of which were ordered in November 2013 by a total of 1717 online customers. Moreover, our dataset involves demographic variables that consist of gender and year of birth. The demographics had been provided by customers voluntarily in a membership form during their registration. Additionally, the delivery location was available in our dataset for each order.

Online customers were split into five segments according to their age. For such purpose, the years of birth selected in membership pages were used to calculate the customers' age by November 2013. Additionally, the location data of orders were available for each purchase in the dataset. With the additional variables included in the analysis, we argue that our study differentiates from most studies.

Table 2 presents the count of customers grouped by their gender and age, with subtotals. Among the customers who had ordered at least once in November 2013, 58.65% were females, while the ratio of males was 41.35%.

Age   Gender	Female	Male	Total
<24	46 (2.68%)	43 (2.50%)	89 (5.18%)
25-34	238 (13.86%)	164 (9.55%)	402 (23.41%)
35-44	369 (21.49%)	217 (12.64%)	586 (34.13%)
45-54	223 (12.99%)	170 (9.90%)	393 (22.89%)
55 +	131 (7.63%)	116 (6.76%)	247 (14.39%)
<b>Total</b>	<b>1007 (58.65%)</b>	<b>710 (41.35%)</b>	<b>1717 (100.00%)</b>

**Table 2.** Customers by Age and Gender

Table 3 demonstrates the distribution of customers grouped by their location. We should note that the numbers arise from the limited dataset analyzed in our study, which only involves the orders delivered in November 2013.

Location	Customers	Location	Customers	Location	Customers
54	294 (17.12%)	43	78 (4.54%)	47	28 (1.63%)
41	248 (14.44%)	45	75 (4.37%)	55	22 (1.28%)
44	232 (13.51%)	51	69 (4.02%)	50	18 (1.05%)
53	119 (6.93%)	57	69 (4.02%)	48	12 (0.7%)
46	89 (5.18%)	42	59 (3.44%)	58	11 (0.64%)
49	86 (5.01%)	60	56 (3.26%)	61	11 (0.64%)
56	81 (4.72%)	52	51 (2.97%)	59	9 (0.52%)

**Table 3.** Customers by delivery locations

Before analyzing our dataset for rule extraction, a preprocessing phase was required to eliminate products rewarded by the e-retailer. In particular, the company occasionally had a promotion for online customers whose purchase total exceeds a threshold. Accordingly, 2199 orders by 1291 customers had been rewarded with bottled water. The expert opinion by an IT specialist working for the e-retailer was to ignore such products. Moreover, a study by Häubl and Trifts (2000) describes online shopping as a two-step decision making process in which the customers initially screen a set of relevant products, then make a purchase through the examination of products based on their important attributes. From this perspective, it can be argued that free items offered by the e-retailers might be purchased by any customer, without proper consideration of the product attributes. Accordingly, our data preprocessing stage involved the elimination of gifts from the basket data.

As the Apriori algorithm requires, a support threshold should be set for the pruning of the infrequent item-sets. The minimum support threshold was set as 2% for our analysis. Additionally, the confidence threshold required to remove redundant rules was set as 25%. With such parameters, the analysis resulted in a total of 2956 association rules.

## 5. Findings

The rule mining technique initially discovers frequent item patterns that correspond to the most popular items in the basket data. As in the conventional market basket analysis, our findings involve cross-category purchase patterns. Due to the inclusion of demographic attributes in our analysis, our frequent patterns in Table 4 lists combinations of demographic attributes and purchased items.

Rank	Itemset	Frequency	Rank	Itemset	Frequency
1	[Gender=Female] & Vegetables	820	11	[Gender=Female] & Yogurt	440
2	[Gender=Female] & Milk	714	12	[Gender=Female] & [Age:25-34]	439
3	Vegetables & Fresh Fruits	697	13	[Gender=Male] & [Age:35-44]	438
4	[Gender=Female] & [Age:35-44]	659	14	[Age:35-44] & Milk	433
5	Vegetables & Milk	593	15	[Gender=Female] & [Age:45-54]	431
6	[Gender=Female] & Fresh Fruits	534	16	Yogurt & Vegetables	429
7	[Gender=Male] & Vegetables	495	17	Yogurt & Milk	409
8	[Age:35-44] & Vegetables	467	18	Vegetables & Eggs	393
9	[Gender=Male] & Milk	456	19	Fresh Fruits & Milk	391
10	[Gender=Female] & Vegetables & Fresh Fruits	442	20	[Gender=Female] & Paper Towels	389

**Table 4.** Product categories frequently purchased together

According to the table above, top-two itemsets suggest that females often purchase vegetables or milk. Moreover, the third itemset indicates the frequent purchase of vegetables and fresh fruits together, regardless of gender. On the other hand, a drawback from the inclusion of demographics was the discovery of irrelevant itemsets. Such findings ranked as 4, 12, 13, 15 in Table 4 identify the count of orders by particular groups of customers.

The inclusion of demographic variables in the analysis has also enabled the discovery of association rules, which occasionally point to a limited group of customers. Firstly, we present the top-10 rules in terms of the lift measure in Table 5.

Rule	Antecedent (X)	Consequent (Y)	Support	Confidence	Lift
1	Coffee Cream	Paper Towels & Instant Coffee	0.0651	0.3301	8.56
2	Paper Towels & Instant Coffee	Coffee Cream	0.0386	0.5574	8.56
3	Coffee Cream	Cube Sugar & Instant Coffee	0.0651	0.3689	8.40
4	Cube Sugar & Instant Coffee	Coffee Cream	0.0439	0.5468	8.40
5	Toilet Paper (3-Layered)	Handkerchief	0.0771	0.3197	7.90
6	Handkerchief	Toilet Paper (3-Layered)	0.0405	0.6094	7.90
7	Coffee Cream & Cube Sugar	Instant Coffee	0.0322	0.7451	6.64
8	Coffee Cream	[Gender=Male] & Instant Coffee	0.0651	0.3350	6.46
9	[Gender=Male] & Instant Coffee	Coffee Cream	0.0518	0.4207	6.46
10	Cracked Wheat	Lentil	0.0607	0.3490	6.27

**Table 5.** Rules on cross-category purchase patterns

Among the top-10 rules, two involved demographic variables, while the remaining signify cross-category purchase patterns. A cross-category relation, as in the last row, suggests an observation of “purchase of lentil is 6.27 times more frequent in baskets that involve cracked wheat”. Moreover, the confidence for the same rule signifies that 34.90% of the baskets that involve cracked wheat also involve lentil. The support measure in this rule signals that such a pattern relates to 6.07% of baskets that include cracked wheat. On the other hand, the rules ranked 8th, and 9th in Table 5 involves the gender variable, that signifies the presence of a particular purchase pattern within a specific customer group. To explore such findings further, top rules which involve the gender, age, and location variables have been filtered and presented separately in Tables 6-8.

Rule	Antecedent (X)	Consequent (Y)	Support	Confidence	Lift
1	Coffee Cream	[Gender: Male] & Instant Coffee	0.0651	0.3350	6.46
2	[Gender: Male] & Instant Coffee	Coffee Cream	0.0518	0.4207	6.46
3	[Gender: Male] & Coffee Cream	Instant Coffee	0.0345	0.6330	5.64
4	[Gender: Male] & Standard Pasta	Turkish Noodles	0.0790	0.2560	3.86
5	Turkish Noodles	[Gender: Male] & Standard Pasta	0.0664	0.3048	3.86
6	[Gender: Male] & Sugar	Tea	0.0654	0.4638	3.77
7	[Gender: Male] & Turkish Noodles	Standard Pasta	0.0304	0.6667	3.49
8	Washing Powder	[Gender: Female] & Fabric Softener	0.1028	0.3046	3.44
9	[Gender: Female] & Fabric Softener	Washing Powder	0.0885	0.3536	3.44
10	[Gender: Male] & Tea	Sugar	0.0651	0.4660	3.29

**Table 6.** Association rules that demonstrate relations among purchase decisions and gender

The rules that involve gender above help to discover interesting purchase patterns observed for males and females, separately. For instance, the 6th rule in Table 6 indicates that 65.4% of male customers who purchase sugar also purchase tea. Moreover, the lift for the rule suggests that male customers purchasing sugar are 3.77 times more likely to purchase tea, compared with others. Arguably, such a finding suggests a noticeable difference in tea consumption across females and males. On the other hand, one could argue that purchase decisions often originate from family members; thus, deriving such conclusions might be misleading due to the lack of a variable representing the family size. Nevertheless, such a rule might still be taken into consideration to recommend tea for males who have already added sugar into their basket when shopping.

In addition to the gender, age groups and delivery locations were also evident in the results. It could be argued that inclusion of the age variable might help to demonstrate how purchase behaviors come forward depending on age groups. Accordingly, interesting rules involving age groups are listed in Table 7.



Rule	Antecedent (X)	Consequent (Y)	Support	Confidence	Lift
1	[Age:35-44] & Yogurt	Milk & Eggs	0.0746	0.2881	2.95
2	[Age:35-44] & Yogurt & Vegetables	[Gender: Female] & Fresh Fruits	0.0481	0.4671	2.77
3	[Gender: Female] & [Age:35-44] & Yogurt & Vegetables	Fresh Fruits	0.0310	0.7245	2.74
4	[Gender: Female] & [Age:35-44] & Vegetables	Yogurt & Fresh Fruits	0.0888	0.2527	2.72
5	[Gender: Female] & [Age:35-44] & Fresh Fruits	Yogurt & Vegetables	0.0610	0.3679	2.71
6	[Gender: Female] & Yogurt & Vegetables	[Age:35-44] & Fresh Fruits	0.0873	0.2572	2.69
7	[Age:35-44] & Yogurt & Vegetables	Fresh Fruits	0.0481	0.6579	2.49
8	[Age:35-44] & Yogurt & Fresh Fruits	[Gender: Female] & Vegetables	0.0351	0.6396	2.47
9	[Age:35-44] & Milk & Eggs	Yogurt	0.0408	0.5271	2.44
10	[Age:35-44] & Fresh Fruits	Yogurt & Vegetables	0.0958	0.3300	2.43
11	[Gender: Female] & Yogurt & Fresh Fruits	[Age:35-44] & Vegetables	0.0626	0.3586	2.43
12	[Age:45-54] & Fresh Fruits	Yogurt & Vegetables	0.0654	0.3285	2.42
13	[Age:35-44] & Yogurt	Fresh Fruits & Milk	0.0746	0.2966	2.40
14	[Age:35-44] & Yogurt & Milk	Eggs	0.0474	0.4533	2.37
15	[Age:55+] & Vegetables & Milk	Fresh Fruits	0.0326	0.6214	2.35
16	[Age:45-54] & Vegetables & Milk	Yogurt	0.0443	0.5071	2.35
17	Yogurt & Fresh Fruits	[Age:35-44] & Vegetables	0.0929	0.3401	2.30
18	[Age:35-44] & Eggs	Yogurt & Milk	0.0730	0.2944	2.28
19	[Gender: Female] & [Age:35-44] & Vegetables & Milk	Fresh Fruits	0.0402	0.5984	2.26
20	[Age:45-54] & Yogurt	Vegetables & Milk	0.0534	0.4201	2.24

**Table 7.** Cross-category shopping patterns including age group

The most interesting rule at the top of Table 7 suggests that customers of an age between 35 and 44 who purchase yogurt are likely to purchase milk and eggs, too. Moreover, the customers described in the antecedent had purchased milk & eggs approximately three times more often than others. Furthermore, the third rule signifies a cross-category purchase pattern that is more common in females of age 35-44. Accordingly, 72.45% of females between the age of 35-44 who purchase yogurt and vegetables have purchased fresh fruits; moreover. Moreover, the purchase of fresh fruits was found 2.74 times more frequent in this group than other customers. In demonstrated in this rule, our approach might discover relationships among purchase behaviors and multiple demographic variables.

As the final set in our findings, Table 8 shows the associations between the delivery locations and the product categories.

Rule	Antecedent (X)	Consequent (Y)	Support	Confidence	Lift
1	[Location 54] & Fresh Fruits	[Gender: Female] & Vegetables	0.0541	0.5673	2.19
2	[Location 54] & Vegetables	[Gender: Female] & Fresh Fruits	0.0831	0.3688	2.18
3	[Gender: Female] & [Location 54] & Vegetables	Fresh Fruits	0.0547	0.5607	2.12
4	[Location 54] & Fresh Fruits	Vegetables & Milk	0.0541	0.3977	2.12
5	[Location 54] & Vegetables	Fresh Fruits	0.0831	0.5551	2.10
6	[Location 54] & Vegetables	Fresh Fruits & Milk	0.0831	0.2586	2.09
7	[Location 54] & Vegetables & Milk	Fresh Fruits	0.0395	0.5440	2.06
8	[Location 54] & Fresh Fruits	Vegetables	0.0541	0.8538	2.05
9	[Location 54] & Fresh Fruits & Milk	Vegetables	0.0256	0.8395	2.02
10	[Gender: Female] & [Location 54] & Fresh Fruits	Vegetables	0.0367	0.8362	2.01

**Table 8.** Cross-category shopping patterns including delivery locations

As mentioned before, locations were represented with identifier numbers to prevent revealing an overall location-based sales report. As an example, compared to the rule “Vegetables Fresh Fruits”, the fifth rule “[Location 54] and Vegetables Fresh Fruits” provides location-specific metrics about the purchase decisions. Moreover, the latter rule has a lift of 2.10, which is higher than the lift of 2.00 in the former.

An interesting detail in our findings was the dominance of Location-54 compared to other locations. Among the 2956 rules, 158 involved a location variable. Among those

158 rules, location-54 was present in 75, while the remaining 83 rules involved the other locations. Moreover, the average lift for rules with Location-54 was 1.33, whereas the average lift was found 1.18 in other segment-specific rules. Based on this difference, it might even be argued that the rules regarding location-54 indicate purchase characteristics, which differentiates 54 from other locations.

## 6. Conclusion

Customer data is an essential resource for analyses to implement customer-oriented strategies. For this purpose, purchase records have been extensively analyzed in prior research for a variety of problems. This study extends the conventional market basket analysis with the inclusion of demographic variables and presents findings that indicate segment-specific behaviors.

The data examined with market basket analysis consists of purchase records as well as delivery location, gender, and age group. The underlying motive to integrate those attributes was the opportunity to extract patterns that connect purchased products with demographic variables. The justification of this idea lies in the potential differences in purchase intentions across different customer segments. In this regard, our study aims to present a broadened use of conventional market basket analysis with demographics and discover segment-related purchase patterns.

Among the association rules discovered in our study, interesting results were chosen based on the lift and confidence measures and presented in Tables 5-8 separately. It could be argued that such rules might be useful for practitioners, especially when launching segment-specific campaigns. Moreover, the findings might be utilized to develop customized offers in e-retailing. We argue that our approach might result in more specific rules and lead to more-detailed purchase patterns in datasets with more demographic variables.

The consumer-oriented paradigm in the marketing context emphasizes an understanding of consumer behaviors and adopting more customer-focused practices. The model proposed in this study aims to contribute prior research with the inclusion of customer demographics in basket data for market basket analysis. Besides, the assessment of demographic cross-category association rules in recommender systems might be explored in further studies.

## Acknowledgement

The authors would like to kindly thank for the support of former Adepo.com executives and employees.

## References

- Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. (2007). "Customer relationship management in retailing: A content analysis of retail trade journals", *Journal of Retailing and Consumer Services*, 14(6), 394-399.
- Aggarwal, C. C. (2015). *Data mining: The Textbook*. Springer.
- Agrawal, R., Imieliński, T., Swami A. (1993). "Mining association rules between sets of items in large databases", In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Washington, DC, USA, 207-216.

- Aksoy, R. (2008). *İnternet Ortamında Pazarlama*, Seçkin Yayıncılık, Ankara.
- Bala, P. K. (2008). "Exploring Various Forms of Purchase Dependency in Retail Sale", In *Proceedings of the World Congress on Engineering and Computer Science 2008*, San Francisco, USA, 1101-1104.
- Bayardo Jr, R. J., Agrawal, R. (1999). "Mining the most interesting rules", In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, USA, 145-154.
- Bodapati, A. (2008). "Recommendation Systems with Purchase Data", *Journal of Marketing Research*, 45(1), 77-93.
- Bramer, M. (2016). *Principles of Data Mining*, Third Edition, Springer.
- Chen, Y. L., Chen, J. M., Tung, C. W. (2006). "A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-Space Adjacency on Sales", *Decisions Support Systems*, 3(42), 1503-1520.
- Ching, W. K., Pong, M. K. (2002). *Advances in Data Mining and Modeling*, World Scientific, 1st Edition, Hong Kong, China.
- Dalal, M. K. (2012). "Automatic Text Classification of Sports Blog Data", *Computing, Communications and Applications Conference*, Hong Kong, China, 219-222.
- Dippold, K., & Hruschka, H. (2013a). "Variable selection for market basket analysis", *Computational Statistics*, 28(2), 519-539.
- Dippold, K., Hruschka, H. (2013b). "A model of heterogeneous multicategory choice for market basket analysis", *Review of Marketing Science*, 11(1), 1-31.
- Emel, G. G., Taşkın, Ç. (2005). "Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması", *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6, 221-239.
- Griva, A., Bardaki, C., Pramatarı, K., & Papakiriakopoulos, D. (2018). "Retail business analytics: Customer visit segmentation using market basket data", *Expert Systems with Applications*, 100, 1-16.
- Halim, S., Octavia, T., & Alianto, C. (2019). "Designing Facility Layout of an Amusement Arcade using Market Basket Analysis", *Procedia Computer Science*, 161, 623-629.
- Han, J., Kamber, M., Pei, J. P. (2012). *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishing, Third Edition, USA.
- Häubl, G., Trifts, V. (2000). "Consumer Decision Making in Online Shopping Environments: The effects of interactive decision aids", *Marketing Science*, 19(1), 4-21.
- Hormozi, A. M., Giles, S. (2004). Data mining: A competitive weapon for banking and retail industries. *Information Systems Management*, 21(2), 62-71.
- Hudairy, H. (2004). *Data mining and decision making support in the governmental sector*, Master Thesis, Faculty of Graduate School of the University of Louisville, Kentucky, USA.
- Kabasakal, İ. (2020). Understanding shopping behaviors with category and brand-level market basket analysis. In *Tools and Techniques for Implementing International E-Trading Tactics for Competitive Advantage*, Editor: Yurdagül Meral, IGI Global, 242-267.
- Kooti, F., Lerman, K., Aiello, L. M., Grbovic, M., Djuric, N., & Radosavljevic, V. (2016). Portrait of an online shopper: Understanding and predicting consumer behavior. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, San Francisco, USA, 205-214.
- Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann, USA.
- Kronberger, G., Affenzeller, M. (2011). Market Basket Analysis of Retail Data: Supervised Learning Approach. In *Proceedings of the 13th International Conference on Computer Aided Systems Theory*, Las Palmas de Gran Canaria, Spain, 464-471.
- Linoff, G. S., Berry, M. J. A. (2011). *Data Mining Techniques for Marketing, Sales and Customer Relationship Management*, Wiley, Third Edition, Canada.
- Marakas, G. M. (2003). *Decision Support Systems in the 21st Century*, Prentice Hall, Second Edition, USA.
- Osadchiy, T., Poliakov, I., Olivier, P., Rowland, M., & Foster, E. (2019). "Recommender system based on pairwise association rules", *Expert Systems with Applications*, 115, 535-542.

- Özçakır, F. C., Çamurcu, A. Y. (2007). "Birliktelik Kuralı Yöntemi İçin Veri Madenciliği Yazılımı Tasarımı ve Uygulaması", İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 12, 21-37.
- Savaş, S., Topaloğlu, N., Yılmaz, M. (2012). "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri", İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 21, 1-23.
- Sota, S., Chaudhry, H., Chamaria, A., Chauhan, A. (2018). "Customer relationship management research from 2007 to 2016: An academic literature review", Journal of Relationship Marketing, 17(4), 277-291.
- Şimşek-Gürsoy, U. T., Kasapoğlu, Ö. A., Atalay, K. (2019). "R Programlama ile Birliktelik Kuralları Analizi: Tüketicilerin İnternet Üzerinden Yaptıkları Alışveriş Verisinin Apriori ve Eclat Algoritmalarıyla İncelenmesi", Alphanumeric Journal, 7(2), 357-368.
- Solnet, D., Boztug, Y., Dolnicar, S. (2016). "An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue", International Journal of Hospitality Management, 56, 119-125.
- Strauss, J., Frost, R. (2009). E-Marketing, Pearson Education, New Jersey.
- Tsiptsis, K., Chorianopoulos, A. (2009). Data Mining Techniques in CRM: Inside Customer Segmentation, John Wiley & Sons Ltd., Chichester, United Kingdom.
- Vahaplar, A., İnceoğlu, M. M. (2001). "Veri Madenciliği ve Elektronik Ticaret", VII. Türkiye'de İnternet Konferansı, İstanbul.
- Winer, R. S. (2001). "A framework for customer relationship management", California management review, 43(4), 89-105.
- Zhang, Y., Pennacchiotti, M. (2013). Predicting purchase behaviors from social media. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro Brazil, 1521-1532.