MEF UNIVERSITY

MARKET BASKET ANALYSIS - AYDINLI GROUP

Capstone Project

Çağlayan Özgür Öney

İSTANBUL, 2018

MEF UNIVERSITY

MARKET BASKET ANALYSIS - AYDINLI GROUP

Capstone Project

Çağlayan Özgür Öney

Advisor: Dr. Tuna Çakar

İSTANBUL, 2018

EXECUTIVE SUMMARY

MARKET BASKET ANALYSIS - AYDINLI GROUP

Çağlayan Özgür Öney

Advisor: Dr. Tuna Çakar

JANUARY, 2018, 28 pages

In this paper, we have analyzed the purchase transaction data of Aydınlı Group. Aydınlı offers their customers diverse set of products by providing Polo, Cacharel and Pierre Cardin brands on both retail and online store. The million dollar question that we seek an answer in our research is "can we determine the purchase pattern of customers?". In order to reveal patterns among customers purchase decisions, we got help from the Market Basket Analysis. But before getting into the process where we create association rules, first with the help of exploratory data analysis we examined our data. By looking the data's general attributes, brand and product distribution, distribution of products sold, number of items in the basket, products gender distribution in the basket, city distribution and much more we better understood the data set.

After getting familiar with the features of the data set we have started to analyze rules between transactions by using apriori algorithm on R. Before getting into outcomes of the association rules we have explained the support, confidence and lift factors and how they can be interpreted. After giving information about terminology first we focused on creating rules that generally applicable to the data. We created rules for both baskets that contains 2 items and 3 items. Afterwards, since we believe that one of the crucial goals of Aydınlı Group is to increase the sales of the items that doesn't preferred that much. So, we have created association rules by targeting those products that customers don't aim to buy in the first place. By separating man and woman products we have come up with different

association rules which will help Aydınlı Group to understand and manipulate customer's purchase pattern.

Moreover, since Aydınlı Group has different products to offer in warm and cold days we have diversified our analysis and come up with different rules for different weather conditions. Even though there are no surprising results we have clearly described the relationship between purchase patterns and strengthened our general rules.

As we continue, we are curious about whether are there any products that don't persist in the association rule but have similar sales rate in the same months. So, in order understand the similarity we have used cosine similarity method. After giving terminological information we have created rules to be used to create recommendations for products that sold in the same period of the year but doesn't exist in the same basket.

By using the rules that we have revealed we make sure that each rule indicates a business incentive where Aydınlı can definitely apply to their retail and online stores. Such as; redesigning shelf structure, created bundle packages, create discounted products, enhance recommendation engine on online shopping, offer particular products just before checkout, etc. We believe that by using these association rules and applying corresponding business incentives Aydınlı Group will increase their sales rate and customer loyalty.

ÖZET

PAZAR SEPET ANALİZİ – AYDINLI GRUP

Çağlayan Özgür Öney

Advisor: Dr. Tuna Çakar

OCAK, 2018, 28 sayfa

Araştırmamızda Aydınlı Gruba ait alışveriş işlemi verisini analiz ettik. Aydınlı, Polo, Cacharel ve Pierre Cardin gibi markalarının ürünleri hem fiziksel hem sanal mağazalarında müşterilerine sunuyor. Araştırmamızda "müşterilerin satın alma davranışlarını belirleyebilir miyiz?" sorusuna yanıt arıyoruz. Müşterilerin satın almaları arasındaki örüntüyü bulabilmek için Sepet Analizinden yardım alıyoruz. Fakat, sonucunda kurallar çıkartacağımız analizimize başlamadan önceden keşifsel veri analizi tekniği ile verimizi inceledik. Verinin genel özelliklerine, marka ve ürün dağılım, satılan ürünlerin dağılımı, sepetteki ürün sayısı, satın alınan ürün ve cinsiyet dağılımı, şehir dağılımı ve daha fazlasını veri setindeki dağılımlarını inceledik.

Veri setinin özelliklerini anladıktan sonra R programın üzerinden apropriori algorithması çalıştırarak alışveriş işlemlerini analiz ettik. Sonuçlara girmeden önce analiz çıktılarımız olan support, condifdence ve liftten bahsettik. Terminolojiden bahsettikten sonra öncelikle genel olarak uygulanabilir kurallar üzerinde çalıştık. Bunun sonucunda 2 ve 3 ürüne sahip sepetler için kurallar ürettik. Analizimize devam ederken Aydınlı Grup'un hedeflerinden birinin satılmayan ürünlerdeki satış oranlarını arttırmak olduğuna inandığımızdan ilk aşamada tercih edilmeyen ürünlerin satılması üzerine kurallar çıkarttık. Veriyi kadın ve erkek ürünleri olarak da ayırarak yaptığımız analizimizde çıkarttığımız

kurallar ile Aydınlı Grup'un müşterilerinin satın alma davranışlarını daha iyi manipüle edeceklerini düşünüyoruz.

Aydınlı Grup müşterilerine soğuk ve sıcak sezonlarda farklı ürün sunduğu için biz de analizimizi yaparken havanın sıcak veya soğuk olmasının etkisini görmek istedik. Soğuk ve sıcak aylar olarak ikiye böldüğümüz verimizin analizi sonucunda her ne kadar çok sürpriz bir sonuç çıkmamış olsa da öncesinde çıkarttığımız sonuçları güçlendirir nitelikte sonuçlar elde ettik.

Analizimize devam ederken, çıkarttığımız kurallarda izleyemediğimiz aynı sepette olmamasına rağmen aynı aylarda satılan ürünler olup olmadığını merak ediyorduk. Bunun gibi bir benzerliğin olup olmadığını anlamak adına cosine similarity metodunu kullandık. Biraz terminolojik bilgi verdikten sonra senenin aynı zamanında satılan ürünleri daha çok sattırmak adına öneriler çıkarttık.

Analizimiz sonrasında ortaya çıkarttığımız her kural için Aydınlı Grup'un hem fiziksel hem sanal mağazalarında uygulayabileceği iş geliştirme alanlarından bahsettik. Bunlardan bir kaçı; mağazadaki rafların yerini yeniden tasarlamak, birden fazla ürünün yer aldığı paketler oluşturmak, indirimli ürünler oluşturmak, online mağazadaki öneri sistemini geliştirmek, satın alma aşamasından hemen önce bazı ürünleri önermek olduğunu söyleyebiliriz. Aydınlı Grup'un ortaya çıkarttığımız kurallar ve bunlara bağlı iş geliştirme çözümlerini hayata geçirerek satış oranlarını ve müşteri sadakatini arttıracağına inanıyoruz.

Anahtar Kelimeler: Pazar sepet analizi, müşteri satınalma davranışı, apriori algoritması

TABLE OF CONTENTS

INTRODUCTION

The clothing and accessories market has always been a dynamic one. We live as a consumer society and we tend to buy frequently. In order to grow the sales volume the companies always wanted to know their consumers' purchase behavior. Since, by knowing that they can offer right products which that customer is likely to buy, create promotions to increase cross-sell, aim their discounts and even their marketing campaigns to create a higher impact.

So how can we determine our customers purchase patterns? With the help of basket analysis, we can determine which products are likely to be bought with which products. In terminologically speaking, by creating association rules we can see that if there are combinations of products that frequently co-occur in transactions.

In our case, we examined purchase transactions of Aydınlı Group's customers. Aydınlı, owns Polo, Pierre Cardin, Cacharel stores for both retail and online sail (Table 1). In order to understand the data set of the Aydınlı Group's customers' transactions, we shall move on to exploratory data analysis section.

Table 1

Brand and Product Distribution

Cacharel
2,30%

Pierre Cardin
14,97%

Polo
82,73%
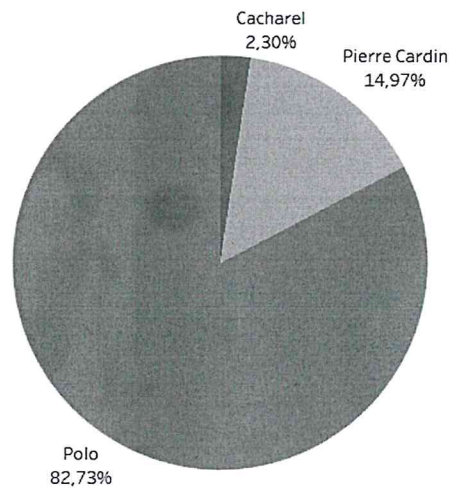
ABOUT THE DATA

Exploratory Data Analysis

The data set has dimensions of 24 attributes which consists 22 nominal fields, 2 continuous fields 1 for date and 1 for age. It consists of 300,000 total items purchased which belong to 263,400 transactions. There are 164 distinct items to buy which are categorized into 5 different classes; for women, men, boys and girls. As you can observe in Table 2 products for man preferred much more than others with %61.72. And there are woman products which exist in %23.36 of the transactions. Since the %85 of the transactions are about man and woman products we are going to create rules for these two types. So, we continuing our analysis of 228.235 transactions which includes 255.218 items and 102 distinct items.

Table 2

Distribution of Products Sold

Boy
10,53%

Girl
4,37%

Woman
23,36%

Unknown
0,03%

Man
61,72%
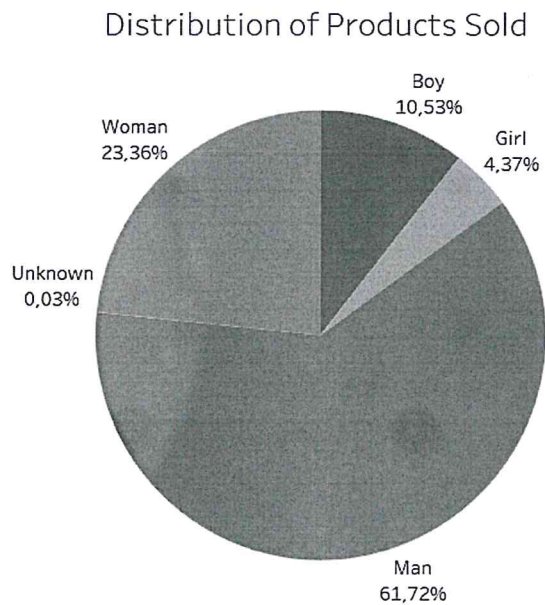
When we get frequency plot of the items for man and woman we can see that there are 2 products which are popular with others for both genders. Please find the distribution on Table 3 and 4. So, rather than focusing top 2 products we picked 2nd, 3rd, 4th and 5th products as targets to sell. We will get in details in Targeting section.
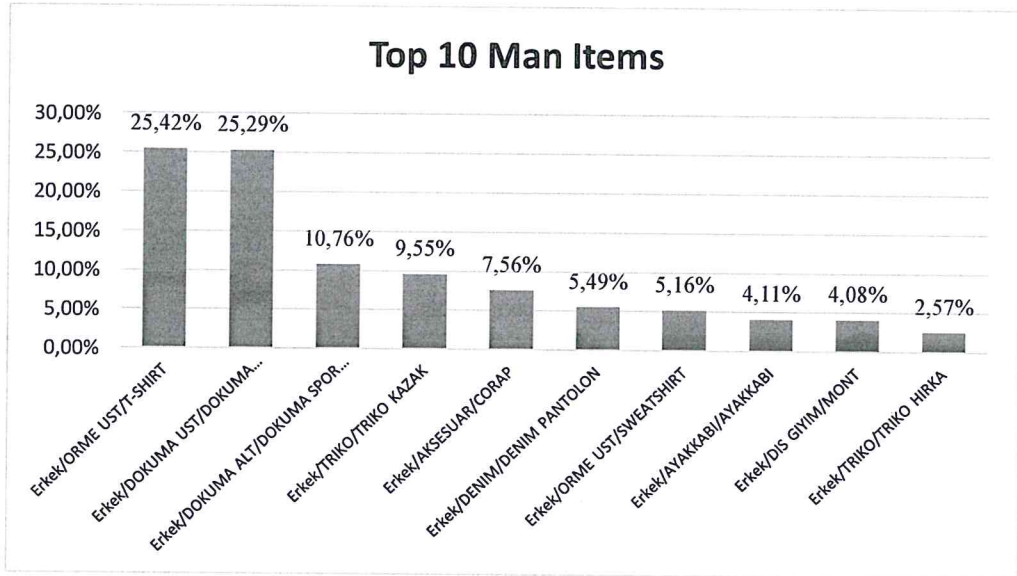
Table 3

Top 10 Man Items

Item	%
Erkek/ORME UST/T-SHIRT	25,42%
Erkek/DOKUMA UST/DOKUMA...	25,29%
Erkek/DOKUMA ALT/DOKUMA SPOR...	10,76%
Erkek/TRIKO/TRIKO KAZAK	9,55%
Erkek/AKSESUAR/CORAP	7,56%
Erkek/DENIM/DENIM PANTOLON	5,49%
Erkek/ORME UST/SWEATSHIRT	5,16%
Erkek/AYAKKABI/AYAKKABI	4,11%
Erkek/DIS GIYIM/MONT	4,08%
Erkek/TRIKO/TRIKO HIRKA	2,57%

Table 4

Top 10 Woman Items

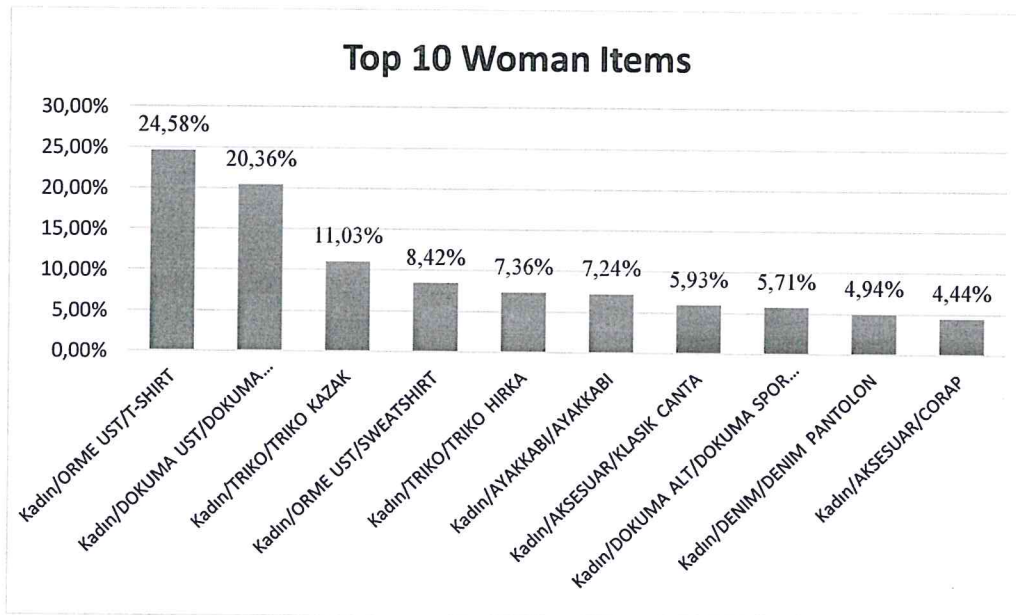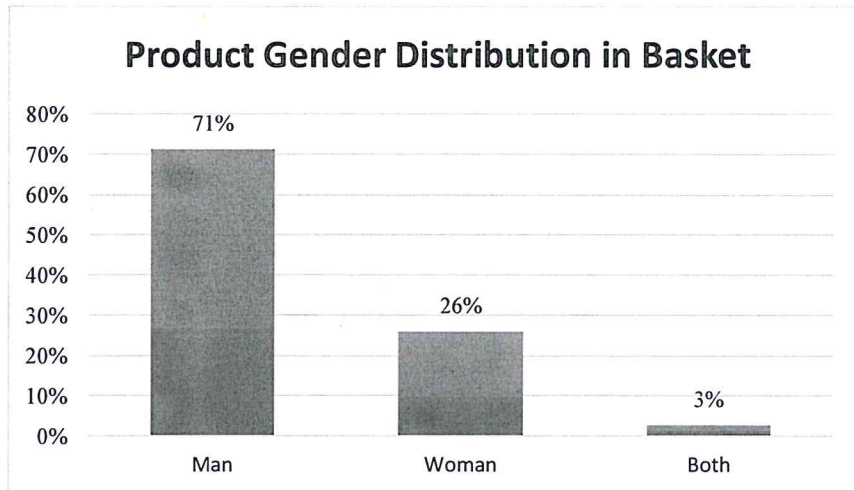| Item	%
Kadin/ORME UST/T-SHIRT	24,58%
Kadin/DOKUMA UST/DOKUMA...	20,36%
Kadin/TRIKO/TRIKO KAZAK	11,03%
Kadin/ORME UST/SWEATSHIRT	8,42%
Kadin/TRIKO/TRIKO HIRKA	7,36%
Kadin/AYAKKABI/AYAKKABI	7,24%
Kadin/AKSESUAR/KLASIK CANTA	5,93%
Kadin/DOKUMA ALT/DOKUMA SPOR...	5,71%
Kadin/DENIM/DENIM PANTOLON	4,94%
Kadin/AKSESUAR/CORAP	4,44%

Since if the customer did not use loyalty card on their purchase we cannot determine whether two separate transactions with different slip number made by the same customer or not and the company offers diverse products we find out that with a ratio of %7 percent of the customers buys 2 or more product (Table 5) the density of 228,235 X 102 matrix is %1.078275.

Table 5

Number of Products in Basket
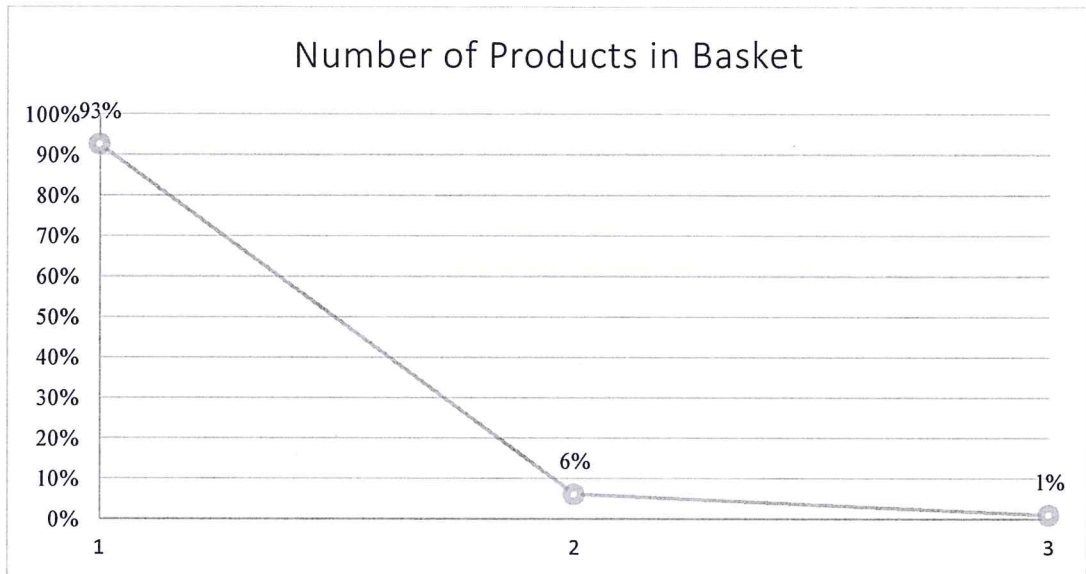
Additionally, when we look at the distributions of products' gender in baskets we see that only %3 of distinct transactions have both man and woman product together. So, in order to create effective rules, we will be focusing on rules only consists of one gender's products. Please find the distribution on Table 6 below.

Table 6

Product Gender Distribution in Basket

Moreover, when we look at the city distribution of transactions' store we can see that the items sold are mostly Polo branded and sold from biggest cities in Turkey. The top cities are İstanbul, Ankara, Antalya, İzmir, Bursa. Please find the distribution in Table 7.

Table 7

Products Sold by Store City

Data Preparation

Since we wanted to analyze the transaction of customers and we have dummy customer numbers for those who performed transactions with the loyalty card, we transformed transaction data into customer based one. Even though there is not a customer information

for transactions made without loyalty card we have merged transactions by looking a particular transaction's bill id.

MARKET BASKET ANALYSIS AND BUSINESS OPPORTUNITIES

Project Definition

As we mentioned in introduction part our project's aim is to understand customers' purchase patterns. In order to reveal hidden associations, we should have analyzed the data to detect purchase patterns with the help of market basket analysis. We are willing to create rules that to be used in online recommendation systems and any other in-store strategical decision that Aydınlı Group makes. To put in a nutshell, with the help of association rules Aydınlı Group's aim is to increase their sales rate and enhance the customer loyalty like any other financial organization.

Terminology

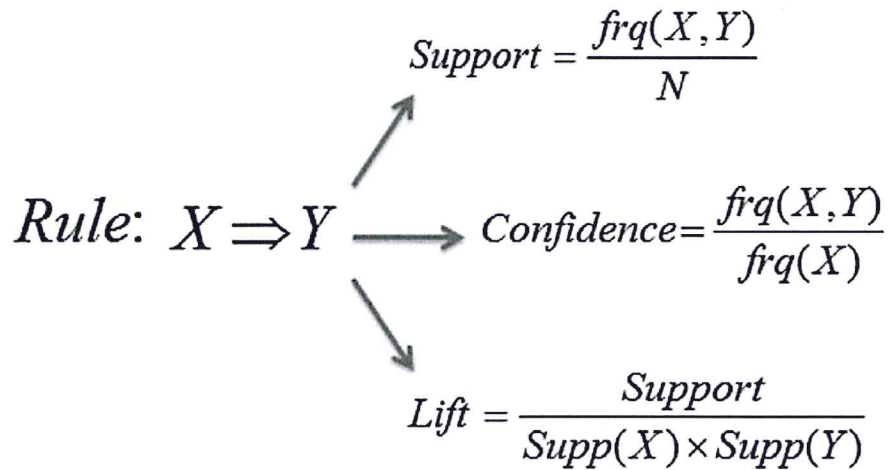In terminology association rules can be illustrated like:

Item set (In our case the n is 102): $I = \{i_1, i_2, \ldots, i_n\}$

Transactions: $T = \{i_i, i_j, \ldots, i_k\}$

Then by looking purchase patterns we create associations such as: $\{i_1, i_2, \ldots\} \rightarrow \{i_k\}$

In our example, this would be: $\{Shirt, Pants\} \rightarrow \{Sweaters\}$

After creating association rules in order to assess it we must analyze the attributes. These attributes are support, confidence and lift factors. Please find the terminological illustration of the support, confidence and lift features below.

$$Support = \frac{frq(X,Y)}{N}$$

$$Rule: X \Rightarrow Y \longrightarrow Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Support has the value between 0 and 1 and shows us whether the corresponding rule occurs a large number of cases. In other words, it is the proportion of transactions in our data set that contains that product.

The confidence shows us that at what percent rate the rule correct should be correct. In other words, it shows how likely the customer will purchase that item on the right-hand side with the items on the left-hand side. The formula is Support(i_1,i_2, i_3)/ Support(i_1,i_2) In an example, imagine that there are total 1.000 transactions and there are 120 transactions contains t-shirt and 50 contains pants. The support would be 120/1,000 = %12 for t-shirt, 50/1,000 = %5 for pants and there are 30 (30/1.000=%3) transactions that contains both product together. Then the confidence for {T-Shirt} => {Pants} is (%3)/%12 = %25.

The lift shows the power of this rule over the population. It is calculated by Support/Support (i_1)* Support (i_2). In our example the lift is %3/(%12*%5) = 5. In other words, we can say that customers who have shirt and pants together in their basket are 5 times more likely to buy a sweater than population. So, basically lift greater than 1 shows that corresponding rule shows there is a pattern between products compared to the population.

Methodology

In order to make market basket analysis and create association rules we have used "arulez" library in R. In this package, we used apriori algorithm. This algorithm uses a step known as candidate generation. By using bottom-up approach it searches frequent subsets and extends them one item at a time. Then candidates are tested against the data. Moreover, it uses tree structure and breadth-first search in order to count candidate item sets' efficiency. The algorithm generates candidate item sets of length with x item sets from item sets with x – 1 length. Then it prunes. Then it decides whether the candidates have enough predefined confidence and support levels or not. Eventually, it scans the transaction database in order to determine item sets that have high confidence and support levels.

Results

Market Basket Analysis

First, we must indicate what are the lowest confidence and support that should be used to perform analysis and create rules. Since the density of the sparse matrix is %1.078275 and there is only %7 percent of the transactions contains more than 2 products our support levels will be low. In order to observe rules even not powerful we have set minimum support level to %0.01, minimum confidence to %1 and minimum lift to 1. By this constraints, we have calculated 11 rules for 1 lhs – 1 rhs and 696 rules for 2 lhs – 1 rhs. After setting these boundaries and getting results we have evaluated rules and determined the crucial ones. Please find the rules which we believe that can be transformed into business opportunities.

For Man Products

1 Item on the LHS

> {Erkek/KLASIK/KLASIK PANTOLON} => {Erkek/KLASIK/CEKET}

Support	Confidence	Lift
0,01%	2%	2.9

Here we can say that a customer who buys man pants are willing to buy jacket 2.9 times more than a population. So, the jacket can be recommended at a discounted price to who buys pants or there can be a bundle package for pants and jackets together (not suits).

{Erkek/KLASIK/CEKET} => {Erkek/AKSESUAR/KRAVAT MENDIL}

Support	Confidence	Lift
0,01%	2%	2.3

People who buy jacket willing to buy handkerchief more than 2,3 than population. This wasn't a surprise since the handkerchief is used with jackets but we can see that people who buy jacket from Aydınlı group's stores also buy handkerchief from that store. So there is an opportunity for cross sale. There can be a handkerchief bundle which contains 3-4 items with a discounted price for just to offer who buys a jacket.

2 Item on the LHS

> {Erkek/DOKUMA UST/DOKUMA GOMLEK,Erkek/ORME UST/T-SHIRT} => {Erkek/AKSESUAR/CORAP}

Support	Confidence	Lift
0.06%	10%	2

Since shirt and t-shirt for the man are the most preferred products in the man category setting a rule for this products will explain the macro behavior of the customers. Here is a 2 lift which shows that men who bought both shirt and t-shirt are willing to buy a sock more than 2 times than the population. So, as a business opportunity socks can be offered to the e-commerce customers while they are at checkout page or just before. Since men's first initiative about Aydınlı Group's products isn't to buy a sock (frequency is %4) it can be hard to recommend and sell the product on the first shopping stage. But if the business offers it on the checkout page or just before it will be like offering a sauce on top of a cone of ice cream. Since it has fairly high confidence level sock can be easily sold in this way.

➢ {Erkek/ORME UST/T-SHIRT,Erkek/TRIKO/TRIKO KAZAK} => {Erkek/ORME ALT/ORME PANTOLON}

Support	Confidence	Lift
0.01%	7%	
		5.6

Here we can observe that men who buy t-shirt and sweater together are willing to buy pants more than 5.6 compared to the population. So, locating shelf of the products near each other in stores can increase customer experience.

➢ {Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON,Erkek/ORME UST/SWEATSHIRT} => {Erkek/TRIKO/TRIKO KAZAK}

Support	Confidence	Lift
0,02%	23%	3.6

The rule implies that customers 3.6 times likely to buy pullovers while sports pants and sweaters at the basket. With the high confidence levels, this rule is often valid. Since pullover and sweatshirt are similar products but can be combined with different things and can be used at different temperatures they can be both offered whereas the other one is in the basket.

For Woman Products

1 Item on the LHS

➢ {Kadın/ORME ALT/ORME PANTOLON} => {Kadın/ORME UST/SWEATSHIRT}

Support	Confidence	Lift
0,02%	3%	1.26

The similar story persists in the woman products; customers slightly more likely to buy a sweatshirt with pants together. So, by knowing this pattern pants and sweatshirts can be served at a discounted price at the same time. We believe that by performing such kind of campaign would increase the lift values between these products.

2 Item on the LHS

> {Kadın/DOKUMA UST/DOKUMA GOMLEK,Kadın/ORME UST/T-SHIRT} => {Kadın/TRIKO/TRIKO KAZAK}

Support	Confidence	Lift
0,02%	12%	4.5

With the 4.5 lift and %12 confidence, this rule has strong attributes to rely on. Since women like to combine different styles much more than men shirt and t-shirt can be increases their probability to buy also a pullover 4.5 more than the population. So, by knowing this rule exists Aydınlı Group can tune their online recommendation system and increase the sales of pullovers.

Targeting Products

All the business establishments have one duty; make money. In this part, we are going to target the products that have low frequency on transactions and bring out rules to clarify the answer for the question "how can we make them sell". Please find below the targeted items in order to create business incentives and create a strategy.

For Man Products

How to sell Sport Pant

{Erkek/AYAKKABI/AYAKKABI,Erkek/DENIM/DENIM PANTOLON} => {Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON}

{Erkek/AYAKKABI/AYAKKABI,Erkek/TRIKO/TRIKO KAZAK} => {Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON}

Both of these rules have the same support, confidence and lift level.

Support	Confidence	Lift
0,01%	28%	3.9

Also, as you can see there are shoes on the lhs for rules. So we move on to interpret the rules shoe and pants. Please find the relevant attributes below.

{Erkek/AYAKKABI/AYAKKABI} => {Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON}

Support	Confidence	Lift
0,07%	3%	0.36

Since there is a lift lower than 1, we cannot say that buying a shoe will increase the chance to buy pants. But with the effect of jean pants and pullovers, the lift levels almost increase by 10 times. As a business incentive, on the website if a customer buys jean pants or pullovers recommend them a shoe or in the store bring the shelf together and make shoes appealing.

How to sell Pullover

{Erkek/DOKUMA UST/DOKUMA GOMLEK,Erkek/ORME UST/T-SHIRT} => {Erkek/TRIKO/TRIKO KAZAK}

Support	Confidence	Lift
0,06%	12%	1.81

Since shirt and t-shirt are the most frequently sold products the rule's support is relatively higher among rules that have 2 lhs item. As the rule shows, people who have shirt and t-shirt in their basket will likely to buy a pullover 1.8 times more than a population. So, in the winter and autumn seasons pullovers can be recommended an e-commerce user when they have both t-shirt and shirt in the basket.

How to sell Sock

{Erkek/AKSESUAR/KASKOL BERE ELDIVEN,Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON} => {Erkek/AKSESUAR/CORAP}

Support	Confidence	Lift
0,004%	39%	7.8

Socks can be recommended who buys scarf, hat, gloves and sports pants together. As we mention on the "2 Item on the LHS" part socks are not the primary objective to buy

for Aydınlı Group's customers and they can be offered to the customer at the e-commerce web page at check out or just before it.

How to sell Jean Pant

{Erkek/DOKUMA ALT/DOKUMA SPOR PANTOLON,Erkek/ORME UST/T-SHIRT} => {Erkek/DENIM/DENIM PANTOLON}

Support	Confidence	Lift
0,03%	10%	2.8

We believe that jean pants can be purchased without having an underlying pattern with other items. However, customers who buy sports pants and t-shirt together tends to buy jean pants 2.8 times more than the population. As a business incentive, there can be promotion placed where offers discount on total price when sports pants and jeans purchased together.

For Woman Products

How to sell Pullover

{Kadın/DOKUMA UST/DOKUMA GOMLEK,Kadın/ORME UST/T-SHIRT} => {Kadın/TRIKO/TRIKO KAZAK}

Support	Confidence	Lift
0,02%	12%	4.5

As we mentioned in "2 Item on the LHS" section for Woman customers who has t-shirt and shirt together at their basket will likely to also buy pullovers 4.5 times more than the population. So, there can be a bundle of these products to increase cross-sell rates and make customers experience the woman pullovers of Aydınlı Group's.

How to sell Sweatshirt

{Kadın/ORME ALT/ORME PANTOLON} => {Kadın/ORME UST/SWEATSHIRT}

Support	Confidence	Lift
0,02%	3%	1.26

Since woman likes to shop pants and sweatshirts together instore shelves of these products can be located near each other.

How to sell Cardigan

{Kadın/DOKUMA UST/DOKUMA GOMLEK,Kadın/ORME UST/T-SHIRT} => {Kadın/TRIKO/TRIKO HIRKA}

Support	Confidence	Lift
0,01%	9%	5

Shirts and t-shirts are the most popular products among woman's products. They almost persist 1 of every 2 transactions. The rule shows that customers who have t-shirt and shirt together at their basket will 5 likely to buy a cardigan 5 times more than the population.

How to sell Shoe

{Kadın/ORME ALT/ORME ETEK} => {Kadın/AYAKKABI/AYAKKABI}

Support	Confidence	Lift
0,002%	4%	2.2

{Kadın/ORME ALT/ORME PANTOLON,Kadın/ORME UST/T-SHIRT} => {Kadın/AYAKKABI/AYAKKABI}

Support	Confidence	Lift
0,004%	16%	9

A Customer with pants and t-shirt in the basket will 9 times likely to buy a shoe. Also, skirts increase the chance to sell the shoes 2.2 times. So by placing shoes appealing to the eye, the sales rate can be increased.

Analysis by Season

While conducting basket analysis and creating association rules we are also curious about whether there are different rules according to the season of the year. So, to compare and contrast this effect we grouped 10th, 11th,12th,1st,2nd,3rd months as cold months and 4th,

5th, 6th, 7th, 8th, 9th months as warm months. Even there are not huge differences among rules we have come up with few rules which can help to increase Aydınlı Group's sales by knowing the season's affect over customer purchase pattern.

Cold Season

{Erkek/AYAKKABI/AYAKKABI,Erkek/TRIKO/TRIKO KAZAK} => {Erkek/DENIM/DENIM PANTOLON}

Support	Confidence	Lift
0,01%	22%	6.1

We have examined this rule in the previous parts. However, when we analyze the effect of season attributes of the association rule strengthens. We had 3.9 lift with the same amount of support and %28 confidence but when we only focus on the products that sold on winter days even the confidence lowers a bit the lift increases up to 6.1. So we can clearly say that this rule has a higher power on explaining purchase pattern of the customers in the cold season.

Warm Season

{Erkek/AYAKKABI/AYAKKABI,Erkek/ORME UST/T-SHIRT} => {Erkek/TRIKO/TRIKO KAZAK}

| Support	Confidence	Lift
0,01%	7%	7.6

During the warm season, customers who have man t-shirt and man sweaters in their basket will likely to buy man shoes more than 7.6 compared to the population. The lift is fairly high. By following this rule Aydınlı Group can increase their sales rates by making the shelf of the sweaters close to t-shirt and shoes and making it appealing to the eye. Even more, since sweaters are not the first choice to buy in the warm weathers in order to create loyal customers there can be a discount on sweater prices.

Cosine Similarity

In order to conduct a further analysis, we wanted to examine the products sold in the same month. They are not supposed to be in the same basket. Since we believe that there can be a similar sales rate between products that don't exist in the same basket. By creating this kind of rule, Aydınlı Group can offer products that have similar sales rate with the current basket of the customer.

In order to measure the similarity between products' sales rate, we have used cosine similarity method. The cosine similarity examines similarity between two vectors by looking the cos angle formed by their intersection. Since our data consists of 2.5 year's transactions we conducted an analysis whereas creating vectors of a product by aggregating their existence in a transaction per month. Then, we examined the cosine similarity matrix between product's monthly sales (formed by numbers 0 to 1) and comprehend the products that sold during the same period of the month. Please find the cosine similarity function below.

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$
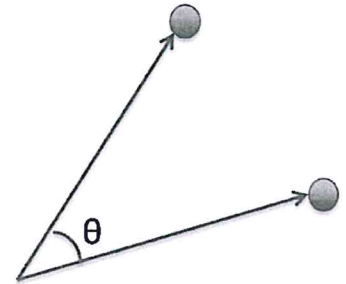
Please find the rules below which have revealed in the light of cosine similarity analysis and couldn't be revealed through basket analysis.

For Man Products

➤ Erkek/DENIM/DENIM CAPRI BERMUDA, Erkek/YUZME GIYIM/YUZME GIYIM

Denim shorts and swimwear has 0.92 similarity which shows that men tend to buy jean short with swimwear at the same period of the year.

➢ Erkek/DIS GIYIM/DERI MONT, Erkek/IC GIYIM/PIJAMA

The leather coat and pajamas have slightly high similarity, 0.82. This similarity tells that men tend to buy pajamas at the same period where they also buy leather coats. So, we believe that using this rule in the e-commerce recommendation would be effective.

➢ Erkek/DOKUMA ALT/KLASIK PANTOLON, Erkek/AKSESUAR/PAPYON KUSAK

The pants and bow tie has 0.88 similarity which is also high like other rules. We believe that placing bow tie shelf near to the pants' can increase bow tie sales.

For Woman Products

Kadın/DENIM/DENIM CAPRI BERMUDA, Kadın/AKSESUAR/SAPKA

➢ Woman are likely to buy jean shorts and hat together at the same period of the year. It has really high similarity with 0.97. As an inference, there can be dummy mannequin dressed up with the jean and shorts in the stores. Even more, there can be a discounted bundle that includes this two products to increase the sales rate of both items.

Kadın/EV TEKSTIL/PIJAMA - Kadın/DIS GIYIM/PALTO

➢ The rule which shows the relationship between coat and pajamas persists also in woman clothes. With the 0.86 similarities, we can clearly say that on e-commerce store the pajamas can be recommended to the customers who have also coat in their basket.

Kadın/KLASIK/KLASIK PANTOLON - Kadın/TRIKO/TRIKO YELEK 0.77

➢ We can conclude that with 0.77 similarity pants and sweater vests are likely to be bought in the same period of the year. Placing shelf of these products near to each other will most probably increase the sales rate of both items.

SOCIAL AND ETHICAL ASPECTS

The study is irrelevant in terms of social and ethical aspects.

VALUE DELIVERED

As we discussed throughout the paper Aydınlı Group offers diversified product catalog for their customers. Even the customer's purchase behavior seems to be different than each other from the outside, there are powerful association rules lies among them. In order to create solutions t increase customer loyalty and sales rates, enhance customer experience, reduce costs and have powerful prediction we have come up with lots of different rules.

With the help of these solutions and by understanding the pattern of customer's behavior, Aydınlı Group can build recommendation systems for e-commerce customers, reorder their shelves, create bundles of products, pursue discount campaigns and furthermore manage their customer's whole interaction with the brand much better.

We believe that with the help of further research, in-depth analysis and improvements can be conducted by Aydınlı Group with more diversified data, creating customer segments and performing control grouped experiments.

REFERENCES

➢ Srikant, R., Vu, Q., & Agrawal, R. (1997, August). Mining association rules with item constraints. In KDD (Vol. 97, pp. 67-73).
➢ RDataMining Reference Card, Yanchang Zhao, 2015

➤ Apriori Algorithm - http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab8-Apriori.pdf

➤ Zhao, Y., & Bhowmick, S. S. (2016). Association Rule Mining with R. A Survey Nanyang Technological University, Singapore.

➤ Gupta, S., & Mamtora, R. (2014). A Survey on Association Rule Mining in Market Basket Analysis. Int. Journal of Information and Computation Technology, 4(4), 0974-2239.

➤ Srikant, R. (1996). Fast algorithms for mining association rules and sequential patterns (Doctoral dissertation, University of Wisconsin, Madison).

➤ Loraine Charlet Annie M.C. and Ashok Kumar D. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining

➤ http://infocenter.informationbuilders.com/wf80/index.jsp?topic=%2Fpubdocs%2FRStat16%2Fsource%2Ftopic49.htm