MARKETING ANALYTICS
Project Work

GROUP 17

Alice Spreafico 10569951
Berk Ceyhan 10761821
Özge Tonbul 10703309
Yuliya Navrotska 10693496

INDEX

01
DESCRIPTIVE ANALYSIS
- Our Objective
- Store & Product Frequency
- Customer Purchase Frequency
- Seasonality of the Sales

02
CUSTOMER SEGMENTATION
- Data Preparation
- Factor Analysis
- Clustering
- Customized Promotions

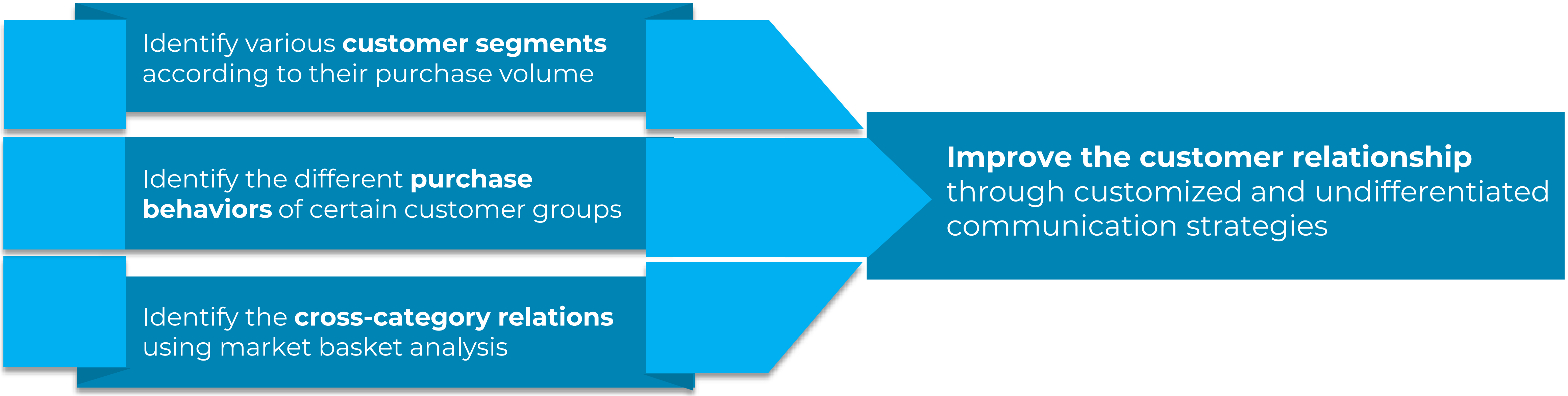
03
MARKET BASKET ANALYSIS
- Two Levels
- Procedure
- Store Promotion
- Additional Promotion
- Business Customers

04
FOR THE FUTURE
- Conclusions
- Future Suggestions

OUR OBJECTIVE

The establishment of loyalty relationships with customers is a main strategic goal for many retailers nowadays. This project work is related to a chain of supermarkets. These supermarkets (large and small stores) differ by the range of products offered, by the sales area and by the size of the city where they are located.

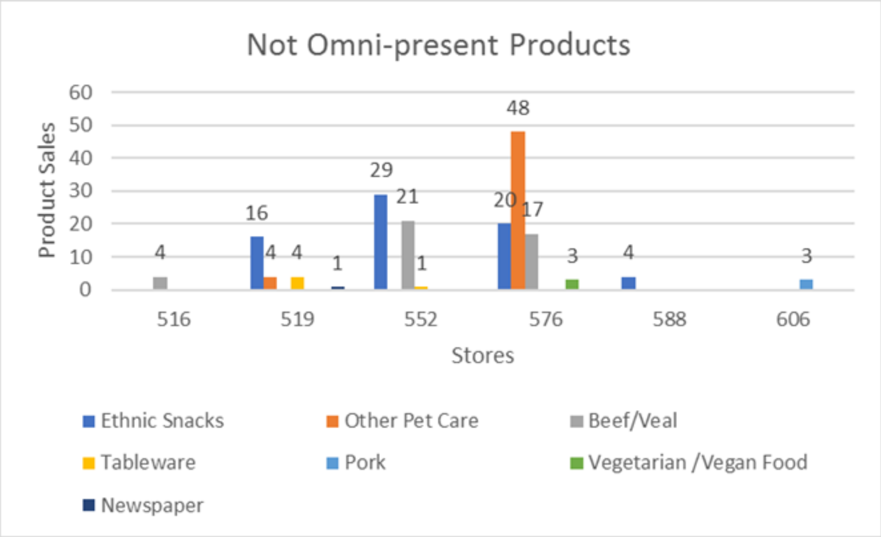
The company has provided us with data which is collected thanks to the implementation of a loyalty program. It basically contains customers (customer id,) and information related to their transactions in a specific time period (date, time, store, products and price). This project work is aimed to analyse this data and identify some strategical decisions to improve the customer relationship. The following graph represents the framework followed for this purpose.



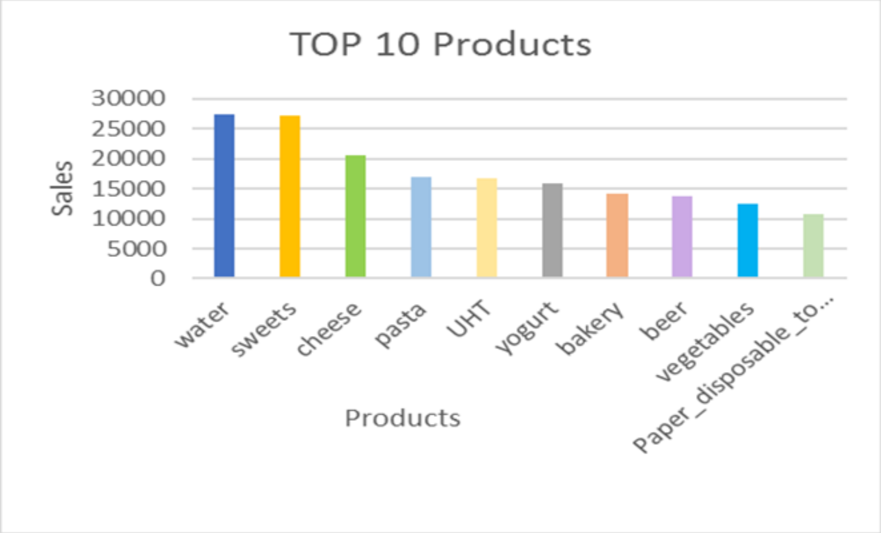
DESCRIPTIVE ANALYSIS

Store & Product Frequency

1.The graph on the right shows some of the products that are only sold in specific stores: ethnic snacks, tableware, newspaper, pork, beef/veal, vegetarian & vegan foods, other pet care.
Considering this, at the beginning of the analysis, we assume that these products can be sold in each store if the promotion we suggest requires it.



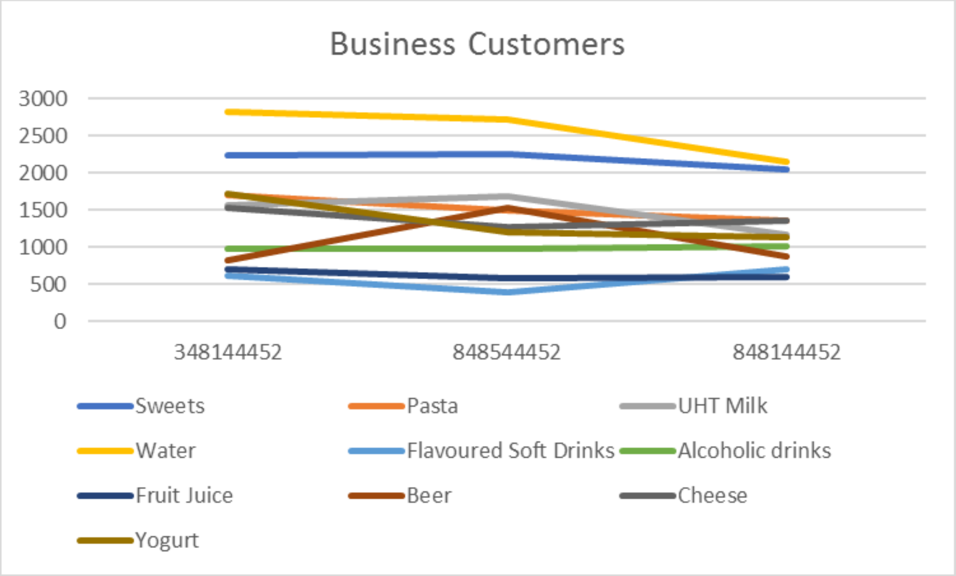
2. The graph shows the amount of sales of top 10 products that are sold in every store: water, sweet, cheese, pasta, UHT Milk, yogurt, bakery, beer, vegetables , paper disposable tools.
The number of sales of those products will be considered while designing the promotions.



DESCRIPTIVE ANALYSIS

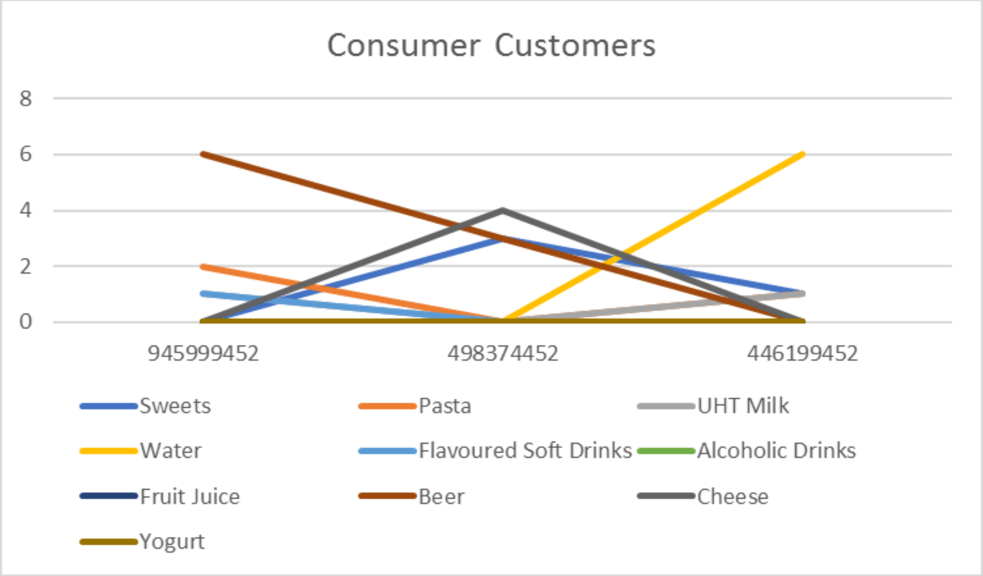
Customer Purchase Frequency: Differences Between Business and Consumer Customers

When the purchase amounts of customers is analyzed visually, it is seen that there are various shopping habits. It is evident that some customers buy **almost every product in high quantities**, and some others only buy **lower quantities of relatively few product categories**. To verify this hypothesis and to differentiate customers in terms of total quantity of purchases, Clustering Analysis is performed (it will be explained later). As a result of the analysis, a threshold for quantity of total purchases (1760) is determined to seperate two clusters of customers. Customers who purchased more than this threshold are defined as ***" business customers"*** while the ones that bought less are defined as ***"consumer customers"***.



Examples of Business Customers' Purchases

It is seen that they bought almost every product in high quantities.



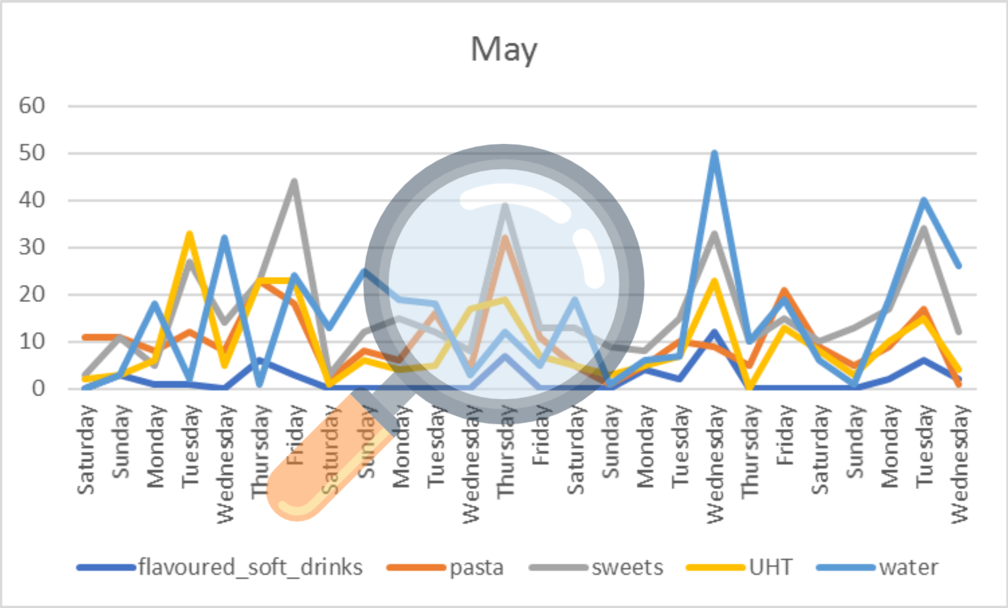
Examples of Consumer Customers' Purchases

It is seen that they didn't buy from all products and quantity is relatively low.

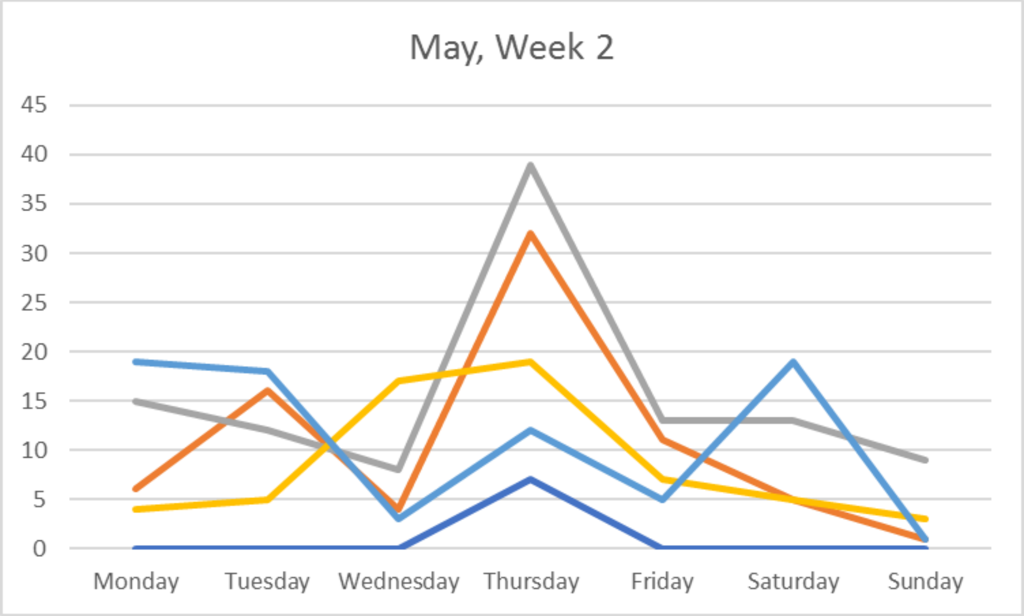
DESCRIPTIVE ANALYSIS

Seasonality of the Sales

- In all three months, it is observed that the sales followed the similar pattern in terms of daily total quantity sold.
- It is seen that, sales increase on Saturdays and usually the peaks are on Thursdays. On the other hand, the sales decrease on Wednesdays and Sundays. The reason might be that generally supermarkets are closed on Sundays or they work half-time. For this reason, it is recommended to choose weekdays for the promotions.



Daily Total Purchases of Store 516 in May 2021



Daily Total Purchases of Store 516 in the Second Week of May

DATA PREPARATION

- **Create Unique Transaction ID's**

For Market Basket Analysis, it is necessary to identify each unique transaction. For this purpose, a new column, Transaction ID, is created with the following formula.

Transaction ID = Store_Ticket Date_Ticket ID_Customer ID

transaction_id
576_20210501104800_853_848544452

- **Elimination of Returned and Free Products**

There are some products with prices smaller or equal to zero in the given dataset. It is assumed that these products are not true proxies for true purchase behavior of customers. That's why these products are excluded from further analysis.

```
select * from mba_no_return
where totprice>0;
```

- **Merging of Products Taking Place More than Once in Single Tticket_ID's**

When a single kind of product exists more than once in a basket, supermarket cashiers sometimes scan each of them and sometimes they just multiply one scan. When they scan these products seperately, they take more than one row in a receipt. This fact distorts true calculations. To solve the issue, these products are merged.

```
SELECT transaction_id, store, date, time,
id_cust, SUM(qty), sum(totprice), prod_id, prod_desc
FROM mba_no_return
GROUP BY transaction_id, prod_id;
```

store	ticketdate	ticket_	id_cust	qty	totprice	prod_id
576	20210501104800	853	848544452	3	5.25	11807
576	20210501104800	853	848544452	1	1.59	21003
576	20210501104800	853	848544452	2	2.78	30210
576	20210501104800	853	848544452	1	1.69	40211
576	20210501104800	853	848544452	1	1.55	40211
576	20210501104800	853	848544452	1	1.79	60105
576	20210501104800	853	848544452	1	0.69	12306
576	20210501104800	853	848544452	1	3.49	11401
576	20210501104800	853	848544452	2	2.9	21003
576	20210501104800	853	848544452	3	5.55	11807

Given Dataset

transaction_id	store	date	time	id_cust	qty	totprice	prod_id	prod_desc
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	6	10.8	11807	UHT vegetable milk
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	3	4.49	21003	Beer
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	2	2.78	30210	Greek yogurt
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	2	3.24	40211	Frozen vegetables
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	1	1.79	60105	Toilet paper
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	1	0.69	12306	Olives
576_20210501104800_853_848544452	576	01.05.21	10:48:00	8.49E+08	1	3.49	11401	Ground coffee
576_20210501104900_863_898544452	576	01.05.21	10:49:00	8.99E+08	2	1.38	11003	Breadsticks

Transformed Dataset

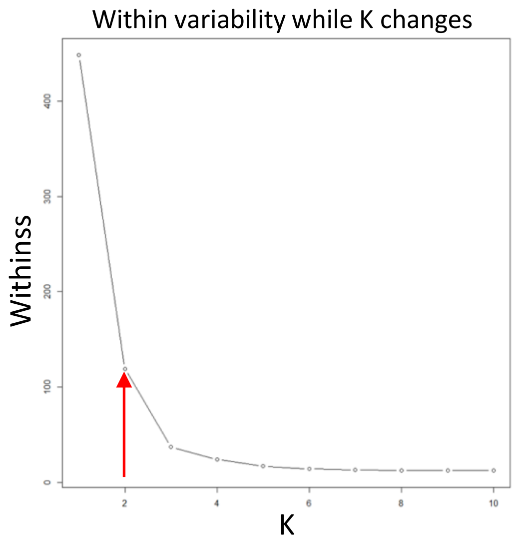
CUSTOMER SEGMENTATION

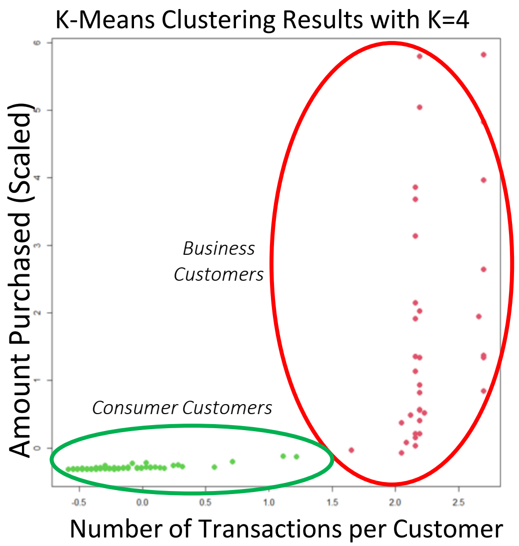
Clustering Based on Amount Spent and N. of Purchases / Visits

1. The first segmentation is performed with the aim of grouping customers based on their shopping habits, in particular according to:

- the **number of (daily) purchases** in the last 3 months (first variable)
- the **value of money spent** in the last 3 months

To perform this customer segmentation, **K- MEANS** algorithm was used. In order to support the choice of the number of clusters K, an elbow curve is plotted for different values of the number of clusters, as depicted in figure on the right. According to the Elbow curve, the most appropriate number of clusters would be two or three, as from those value of k, the marginal gain in the decrease of within variability starts to be much lower. By looking also at the plot of data, k=2 seemed to be more appropriate since k=3 will simply split the data of the green cloud). Therefore, we proceeded with customers' segmentation into 2 different groups.

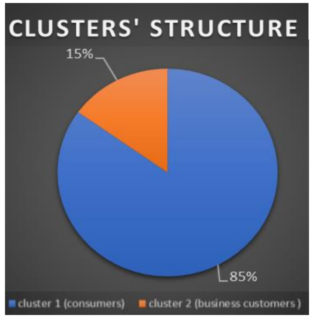




*Note: Data have been scaled

The **percentage of customers** in each cluster and their **profile** can be detailed as following:

- 85% are in Cl. 1 and it is composed by customers who have spent low amount per purchase and have visited the store less frequently . So, these customers were classified as ***consumer customers***.
- 15% are in Cl. 2 and it is made by customers who have spent huge amount and have visited the store frequently. These customers have been classified as ***business customers*** (they might be restaurants, caffes...)



Cluster	Average Amount per Purchase	#Visits in Three Months
#1	1,58€ - 56,7€: Low	1 – 51
#2	29,8€ - 693,6€: High	63 – 92

Strategical Application

This distinction among the 2 segments is an important factor to support the strategical decisions. Further analysis will be done separately to consumer customers and business customers. The customized promotions for both of the customer segments will be explained in the following slides.

CUSTOMER SEGMENTATION

The customer purchase behavior is analyzed with the help of second clustering method. This time a hierarchical clustering is performed.

2. CLUSTERING BASED ON CUSTOMERS' PREFERENCES (Factor Analysis + Hierarchical Clustering)

The other method of segmentation is based on customer **necessities** and **preferences**: in this case, customers are grouped into segments according to the **mix of categories and products** they purchase (each segment is distinguished by high relative weight of their purchases of certain products category compared with other segments).

a. FACTOR ANALYSIS is performed to reduce the dimensionality of large amount products and to better explain the result of the customer segmentation with meaningful factors. The following are the steps to factor analysis. They are performed separately for consumer customers and business customers to be able to understand purchase behaviors in more correct way.

1. Data is scaled in order to avoid strange results due to "difference in scales". In the scaled data, there were nine categories (columns) with zero variance which means that they do not capture any information. So, they have been removed from further analyses

2. Factor Analysis has been performed on all the remaining categories.

Factor analysis for business customers stated the fact that business customers don't have any particular purchase behavior. So, this segment is excluded from strategies taken by factor analysis.

For consumer customers, the scree plot showed that the marginal gain started to be lower after 6-7 factors but since their overall variability captured was low, it has been decided to take the first 9 factors that had a cumulative variance of about 58%. Moreover, the results in the Rotated Component Matrix showed the possibility to give meaningful interpretation to those factors:

Factor 1: **"Basic Daily Products"** since the variables with highest magnitude within this factor were categories like Pasta, Bakery, Dishwashing soap, House cleaning, Pork, Cotton handkerchiefs, Jam/Nutella/Honey…

Factor 2: **"Healthy Food"** due to high weight of Vegetables, Fruits, Rice, Canned vegetables.

Factor 3: **"Ready Food"** due to high loadings of categories like Ready made, Made from fruit, gastronomy, baby food…

Factor 4: **"Snacks"** because of the high weight of ethnic snacks, Dried fruits, Sweet snacks…

Factor 5: **"Long Drinks & Beer"** (Long drinks, Beer..)

Factor 6: **"Fast Food"** (Grease, Soft drinks…)

Factor 8: **"Perfumes & Shampoo"** (Home perfume, Body care..)

Factor 9: **"Breakfast"** (Fresh drinks, Fresh sweets, Cereals…)

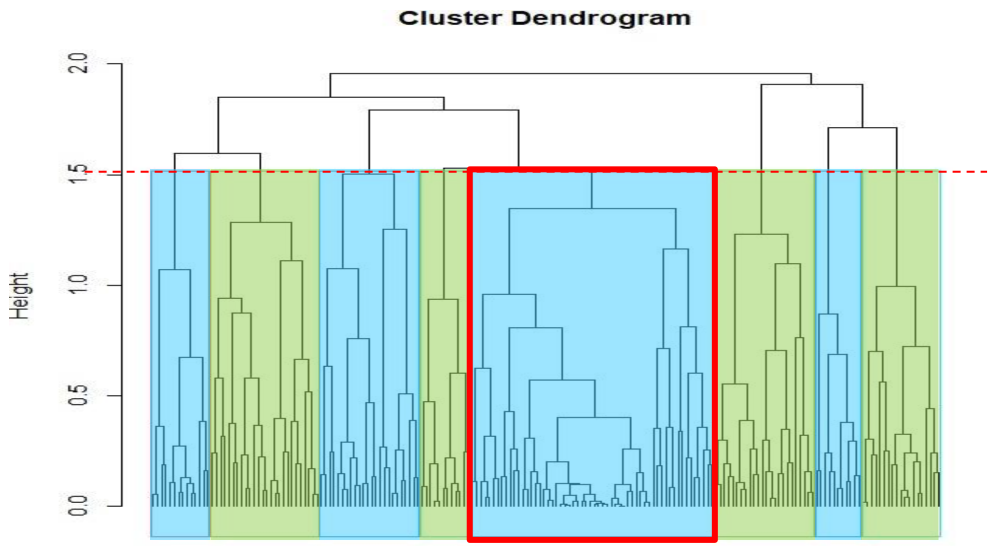
Rotated Component Matrix[a]

	Component								
	1	2	3	4	5	6	7	8	9
pork	,927								
Cotton_handkerchiefs	,743	,500							
pasta	,703								
bakery	,685								
bakery_ingredients	,669								
Garden_furniture	,648								,600
fresh_pasta	,610								
Paper_disposable_tools	,595	,446							
house_cleaning	,581								
canned_second_course	,576	,537							
fruit_juices	,556								
dressing	,550								
Laundry_soap	,542								
jam.nutella.honey	,539								
cosmetics	,523								
herbal.dressing	,497								
dishwahing_soap	,447								
vegetables		,798							
canned_vegetable.or.cereals.in.jar		,795							
milk_base		,753							
frozen_food		,643							
sweets	,519	,640							
from_pork	,442	,620							
rice		,559							
fuits		,543							
analcoholic_drinks		,536					,448		
ice_cream		,495							
cheese		,485		,419					
savory_snacks	,431	,472							
chicken		,408							
Personal_care									
baby_food			,927						
ready_made			,917						
made_from_fruit			,865						
DIY			,841						
gastronomy	,421		,719						
infused	,490		,540						
Baby_products			,527						
Other.pet.care									
First_aid				,668					
dried_fruits.vegetables				,543					
ethnic_snacks				,532					
cat_care	,424			,525					
other_laundry_soap				,502					
sweets_snacks				,471					
sauce				,435					
Gloves_tools_cleaning				,402					
eggs									
face_care									
fruit_syrups									
long_drinks					,875				
dog_care					,705	,407			
tomato				,445	,703				
bakery_food_preparation		,442			,577				
Mouth_care					,534				
beer		,425			,472				
grease						,796			
soft_drinks						,736			
strong_cleaning						,700			
coldcuts	,448				,429	,451			
Home_perfume							,614		
wine							,566		
body_care							,562		
water							,555		
Insecticides	,477						,501		
dietary_supplements									
perfumes								,900	
UHT			,428					,788	
other_services					,614			,655	
hair_care				,421				,522	
fresh_drinks									,800
fresh_sweets									,534
cereals									,484
falavoured_soft_drinks									,404

CUSTOMER SEGMENTATION

b. HIERARCHICAL CLUSTERING method is used to find the clusters of consumer customers in terms of their purchase behaviors. (K-Means is also used but results of hierarchical clustering were preferred due to easier interpretation and simply due to better results taken.)

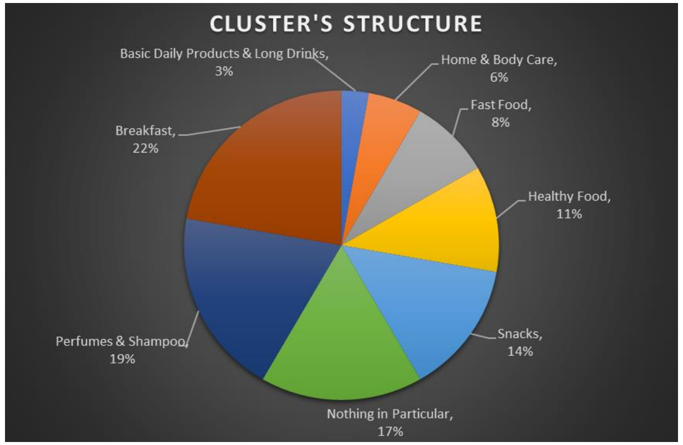
The model is applied on the 9 factors by using as dissimilarity criteria the correlation ("Centered Pearson") and as linkage "Complete linkage". These choices have been driven by interest in finding some similar patterns of behavior among customers, so correlation was the most appropriate type of "distance" for our scope. The results of this clustering can be shown through a Dendrogram in the figure below.



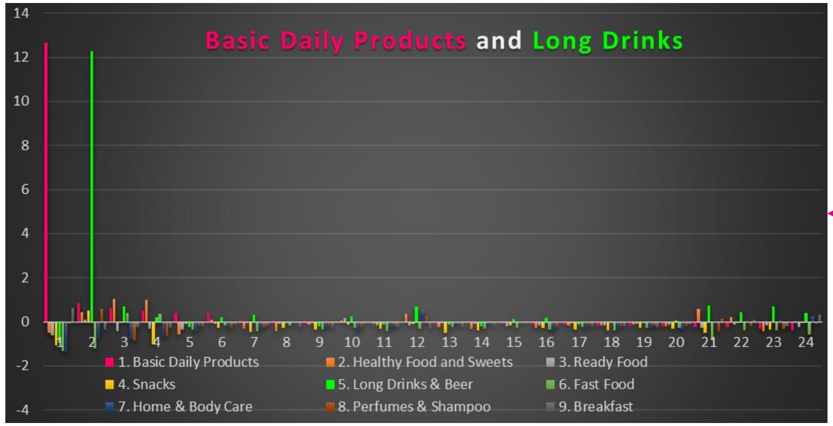
- The length of the branches are proportional to the marginal decrease of variability within each cluster (so the longer it is, the higher is the similarity in the purchase behavior of the customers within that group). It can be noticed that after certain point this marginal decrease is small, in our case for thresholds lower than 1.51 (y<1.51). By cutting the tree at that point we get 8 leaves that correspond to 8 clusters.
- After identifying 8 different clusters, an analyses of the purchasing trends within those groups could be done. In the following slide is showed that, except the cluster n.8 (in red rectangular) which does not present any particular trend, all the remaining ones seem to be characterized by high quantities bought of some factors. For instance, the first cluster is made, in the majority part, by customers that buy Basic Daily products and Long drinks more than the average.* The table above shows each cluster. Each cluster is named after the product categories that they tend to buy more. More detailed graphs about the purchase behavior of each cluster can be found in the following slide.

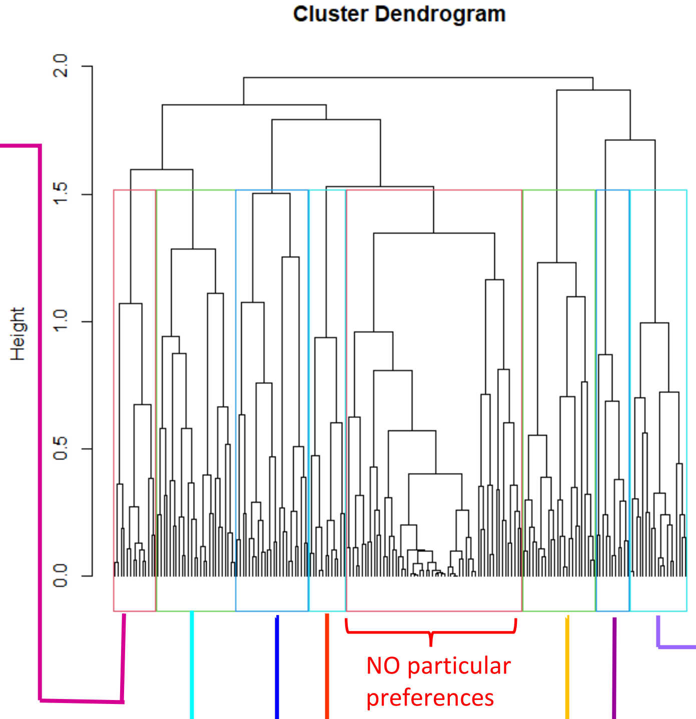
**Note: Data is scaled. The length of each rectangular in the histogram represents how much is purchased more/less with respect to the average quantity purchased of that factor.*

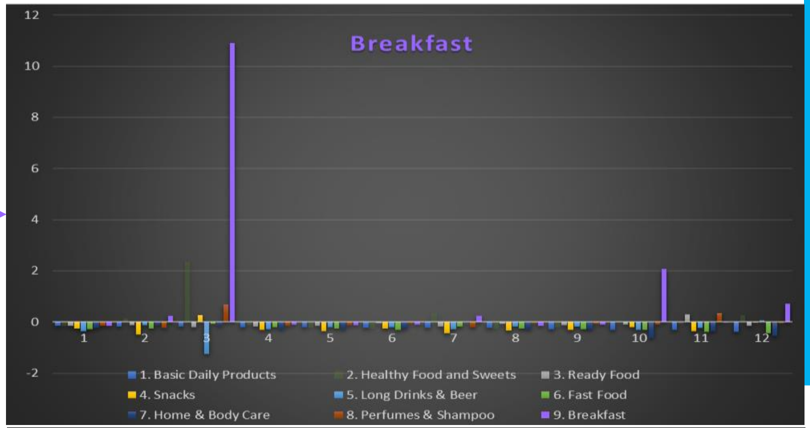
Cluster	Size
#1: Basic Daily Products and Long Drinks	24
#2: Home and Body Care	19
#3: Fast Food	24
#4: Healthy Food	11
#5: Snacks	14
#6: Perfumes & Shampoo	26
#7: Breakfast	12
#8: Nothing in Particular	58

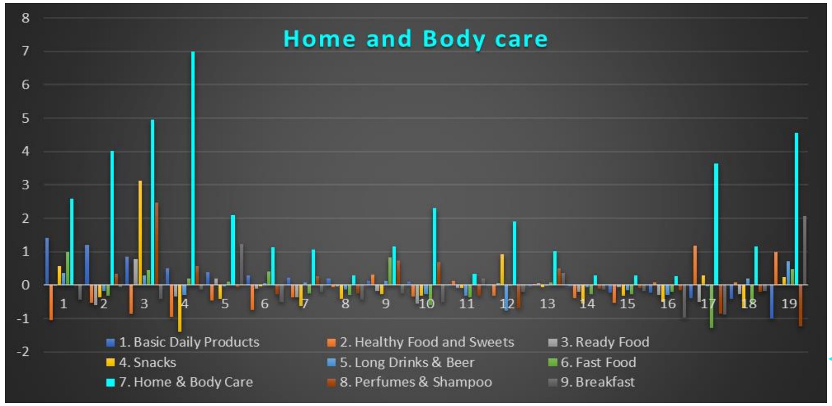


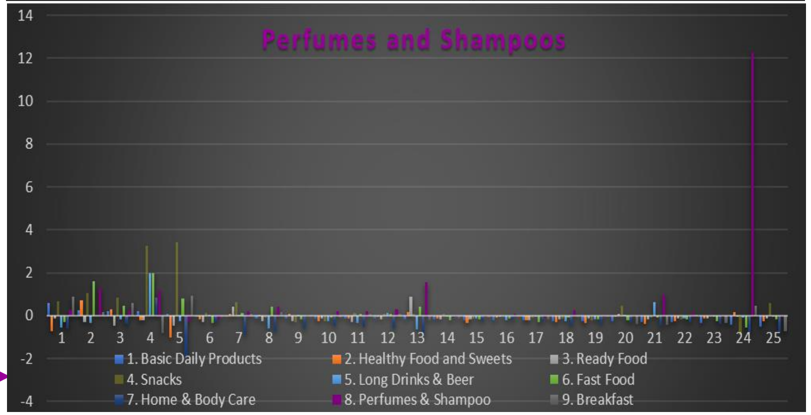
CUSTOMER SEGMENTATION

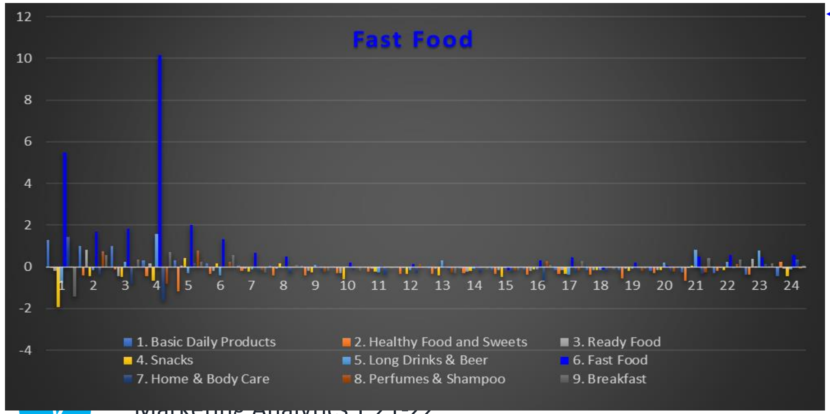


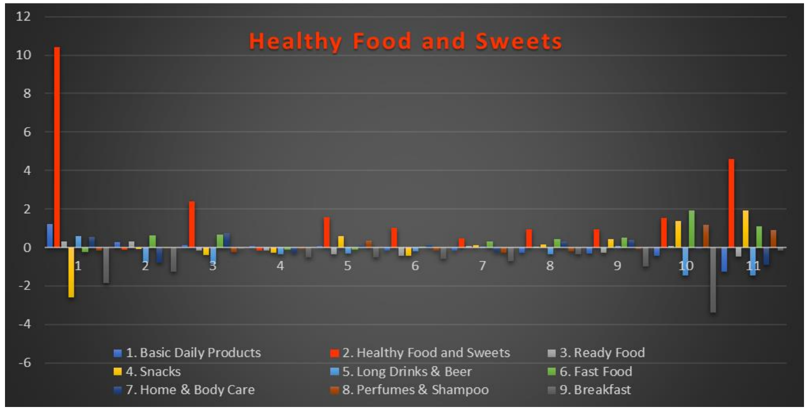


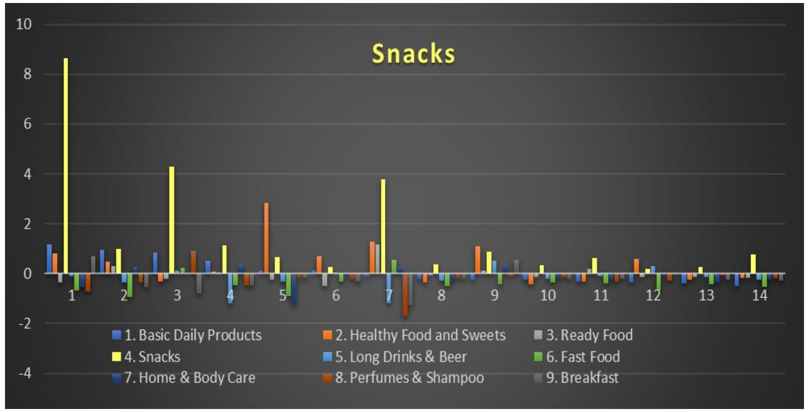












CUSTOMIZED PROMOTIONS

The segmentation allowed us to design several customized promotions that could be proposed to customers through ***email, SMS and application.*** The strategy developing process started with the definition of desired customer reactions. They are assigned as the "Aims". To reach each aim, several possible strategies are produced, and the best ones are filtered. The following graph shows the final table of Aims and Strategies.

Aim 1: Allow customers to discover more

St. 01

Offer discounts for ***various brands*** of same products that they tend to buy (these brands will be chosen from less-selling ones)

Example Strategy

One week send a discount in ***brand A*** pasta, next week send a discount in ***brand B*** pasta.

Expected reaction of customer

"I buy pasta all the time. I like brand A, but I am also open to try new tastes. Why don't I try also brand B now that there is a discount?"

St. 02

Offer discounts for ***relevant products*** with their purchases (by exploiting the results of MBA*)

Example Strategy

prod_desc_1	prod_desc_2	Support	confidence	lift
pasta	cheese	0.072313	0.376365	1.39376

If you buy ***cheese***, a package of ***pasta*** is 50% off.

Expected reaction of customer

"I already buy pasta, but I don't buy cheese. Anyways, I know that cheese is the best friend of pasta. Why don't I try it?"

Aim 2: Keep customers loyal

St. 03

Offer discounts for ***pairs of products*** that they tend to buy (these products will be chosen according to clustering)

Example Strategy

Pasta is for free if you buy 3 packages of it

Expected reaction of customer

"I already buy pasta all the time. This is a good deal for me. Why don't I go for it?"

These strategies will be followed to do marketing for each customer segments. The customization of each marketing action is required since the purchase behavior of each segment is different. So, the offered products will be shaped according to the preferences of every single segment. As can be seen in the example strategies, market basket analysis was used to understand the relations between various products. It will be explained in detail in the following sections.

* The new product will be chosen by Market Basket Analysis and it doesn't belong to the item list that this customer tends to buy a lot.

CUSTOMIZED PROMOTIONS

The following table shows each customer segments and example promotions of each strategy to each one of them:

Cluster	Size	St. 01 various brands	St. 02 relevant products	St. 03 pairs of products
#1: Basic Daily Products and Long Drinks	**24**	Try the new ***fruit-juice brand A*** with 50% discount (Change the offered brand each week)	Buy ***vegetables*** and get ***fresh-pasta*** with 50% discount.	One free ***bakery*** product if you buy 3 packages of it
#2: Home and Body Care	**19**	Try the new ***insecticides brand A*** with 50% discount (Change the offered brand each week)	Buy ***laundry-soap*** and get ***home-perfume*** with 50% discount.	One free ***body-care*** product if you buy 3 packages of it
#3: Fast Food	**24**	Try the new ***grease brand A*** with 50% discount (Change the offered brand each week)	Buy ***frozen-food*** and get ***cold-cuts*** with 50% discount.	One free ***soft-drink*** product if you buy 3 packages of it
#4: Healthy Food	**11**	Try the new ***rice brand A*** with 50% discount (Change the offered brand each week)	Buy ***cotton handkerchief*** and get ***vegetables*** with 50% discount.	One free ***cheese*** product if you buy 3 packages of it
#5: Snacks	**14**	Try the new ***sweet-snacks brand A*** with 50% discount (Change the offered brand each week)	Buy ***yogurt*** and get ***ethnic-snacks*** with 50% discount.	One free ***sauce*** product if you buy 3 packages of of it
#6: Perfumes & Shampoo	**26**	Try the new ***perfume brand A*** with 50% discount (Change the offered brand each week)	Buy ***vegetables*** and get ***hair-care*** with 50% discount.	One free ***hair-care*** product if you buy 3 packages of it
#7: Breakfast	**12**	Try the new ***flavored-soft-drink brand A*** with 50% discount (Change the offered brand each week)	Buy ***milk-base*** and get ***cereals*** with 50% discount.	One free ***cereals*** product if you buy 3 packages of it
#8: Nothing in Particular	**58**	Send random discount e-mails from time to time to lead them to buy some products. As they buy, they will be placed in one of the clusters and they will also be aimed with customized promotions		

MARKET BASKET ANALYSIS

Simultaneously with the clustering analysis, market basket analysis has also been done. The market basket analysis is intended to identify relevant product associations, which can support the design of general and customized promotions. Customized promotions can contribute to improve customer loyalty as they represent an effort from the company to communicate with the clients, intended to reward the customer relationship with the company and suggest the acquisition of products that are likely to be of interest to each customer. The main focus of this analysis is to identify baskets, which are sets of products highly correlated among them. To identify these baskets, firstly data preparation is done as explained before. From the given data, two datasets are prepared to discover the relationships in general and specific approaches. The following are the corresponding levels and the table on the right shows an example of transition from product level to category level.

- **PRODUCT LEVEL:** All the products belonging to the original dataset.
- **CATEGORY LEVEL:** Categories are composed of all the products with the same four initial digits in the product ID (if three digits are present in the category code, it means that the first one is 0).

Product code	Product	Category code	Category
10101	Patisserie	101	sweets
10102	Wafers	101	sweets
10103	Sweet snacks	101	sweets
10201	Bubble gum	102	sweets_snacks
10202	Bonbon	102	sweets_snacks
10203	Sweet snacks	102	sweets_snacks
10401	Ethnic snacks	104	ethnic_snacks
10501	Vinegar	105	dressing
10502	Olive Oil	105	dressing
10401	Ethnic snacks	104	ethnic_snacks
10501	Vinegar	105	dressing
10502	Olive Oil	105	dressing
100801	Frozen fish	1008	Frozen_fish
100802	Frozen fish	1008	Frozen_fish
500101	HARDWARE	5001	IT components
500102	SOFTWARE	5001	IT components
500303	Phone accessories	5003	phones
500401	Appliances	5004	Appliances

MARKET BASKET ANALYSIS

PROCEDURE

1. Application of the market basket analysis algorithm to two levels (product + category) through MySQL. After the data preparation, the analysis continued with the calculation of the **number of transactions of each product** and **number of transactions of each product couple.** Finally, the affinity matrix is obtained by calculating the following ratios which are important all together to understand if two products show some relevance or not.

Support: frequency with which the item set are bought together (the **coverage** of the itemset)

$$\frac{\text{No. Transactions}}{\text{Total transactions}}$$

```
cast(TRANS_COUNT as decimal)/42413 as Support,
# 42413 is equal to the number of unique transaction id's
```

TRANS_COUNT = # Transactions of each product couple

Confidence: % of transactions containing the consequent among those that include the antecedent (the **power** of the itemset);

$$\frac{\text{Support (A \& B)}}{\text{Support (A)}}$$

```
t_count_itemset/42413/support_product_1 as confidence,
```

Lift: measure the **causal relationship**, whether the antecedent has a positive (>1), negative(>1) or independent (=1) effect on the selling of the consequent;

$$\frac{\text{Support (A \& B)}}{\text{Support(A)} * \text{Support(B)}}$$

```
t_count_itemset/42413/(support_product_2*support_product_1) as lift
```

2. These ratios are considered in the following order: **Lift --> Support --> Confidence**

- Firstly, as a proxy of positive relationship between a product couple, the affinity matrix is filtered by **Lift > 1**
- Then, the ones having **relatively higher support** are left.
- Finally, the filtering is completed with the couples of products with **relatively higher confidence**.

The graph on the right shows an example section from the filtered affinity matrix. The full affinity matrix can be reached in the additional documents of the project work.

PRODUCT	prod_desc_1	prod_count_1	PRODUCT	prod_desc_2	prod_count_2	t_count_itemset	Support	confidenc	lift
10104	Cookies	6568	10106	Rusks	2933	942	0.02221	0.143423	2.073981
11802	UHT milk	6150	30501	Butter	2181	663	0.015632	0.107805	2.096437
10104	Cookies	6568	11003	Breadsticks	2043	658	0.015514	0.100183	2.079809
11502	Pasta	6762	12104	Tomato sauce	1471	577	0.013604	0.08533	2.460294
10104	Cookies	6568	10202	Bonbon	1978	509	0.012001	0.077497	1.661718

Example of MBA results (product level)

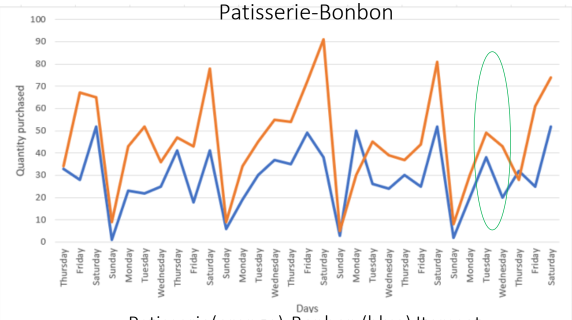
STORE PROMOTION

3. Based on the logic behind every couple, they were assigned to 3 different kind of strategies to communicate with the customer. The following are these strategies and some examples. Full results can be reached in the additional documents of the project.

- RECIPE: Exploit different recipes that include both the antecedent and the consequent* products to make people buy more of the consequent (like in the picture on the right, the image of the final products is shown near the antecedent and there is a QR code that contains information about the other ingredients and their position in the store). They last 1 week;
- GIFT-BOX: Some item-sets can be sold with a discount on the sum of their prices if they are bought together. They last 4-5 days;
- PROMOTION: Discount on the antecedent if a certain amount of consequent is bought. They last 3 days.

4. For creating these promotions, the historical data and affinity matrix are considered together.

5. For best timing of the promotions, decide the best possible period for applying these strategies. (when level of selling of consequent is expected to be low and the one of the antecedent enough high) Here is an example of July 2022:



Patisserie(orange)-Bonbon (blue) Itemset:
Historical data about the selling in July

Note: We refer to the couple antecedent (product bought first) + consequent (product bought second) as itemset

Example of Recipe Promotion Strategy

Week	Promotion propositions
Week 4-10 July	Tisamisù: mascarpone+ bakery ingredients+ cookies+ eggs Gift box: cookies+bonbonb Gift box: cookies+ breadstick Promotion: with 2 pairs of gloves to clean, 20% discount on paper disposable tools
Week 11-17 July	Fruit cake: bakery ingredients+dried fruit Promotion: with 2 pairs of gloves to clean, 20% discount on dishawashing soap Gift box: cookies+ chocolate bars Aperitivo: analcoholic+ sweet snacks+ coldcuts+cheese+fruitjuices
Week 18-24 July	Nutella cake: fresh pasta+ UHTmilk+ butter+ eggs+ sweet+ nutella+ sugar+flour Promotion: with 2 jars of jams, 30% discount on the rusks Gift box: Personal care product + hair care products Gift box: infused+ sweets
Week 25-31 July	Fruit cake:fresh pasta+ eggs+ UHT_milk+ butter+ cookies+ fruits Gift box: infused +sweets Promotion: with 2 pairs of gloves, 20% on house cleaning products Gift box: patisserie+bonbon Pasta recipe: pasta+mozzarella+tomato (or tomato sauce)

July Promotion Propositions

ADDITIONAL PROMOTION

The product IDs of products belonging to the same family and same category differ according to the subcategory. We assumed that in the information contained in the subcategory, there is also the brand of the product. Looking at the dataset, among the different brands of a product there is a significant difference in the selling. That's why we decide to propose an additional promotion to be placed in the stores.

In some item-sets resulting from MBA (product level), the less sold brand of a product is the consequent of another brand of another product. Among those couples we selected the most significant ones (in terms of support, confidence and lift) to create the following promotion: the antecedent can be sold with a discount, in case the consequent is purchased.

For Example:

Among the coldcuts brands, the less sold is the one with the prod ID 30603. In the MBA results (product level) there are the following item-sets:

PRODUCT_1	prod_desc_1	PRODUCT_2	prod_desc_2	Support	confidence	lift
21003	Beer	30603	Coldcuts	0.00061302	0.00587969	2.226566
10303	chips	30603	Coldcuts	0.00049513	0.00805523	3.050417

The store can offer a discount of 20% on Beer 21003 or on Chips 10303 in case 3 packages of coldcuts 30603 are purchased. (3 is chosen since it is above the average amount of that brand bought per transaction)

BUSINESS CUSTOMERS

As highlighted in the descriptive analysis and confirmed by clustering results, some of our customers are business customers. Since they are the ones responsible for most of the transactions and the ones that buy the highest amount per purchase, it is important to keep them loyal.

If the business customers reach a certain amount of purchase of product C in the first month (this quantity is chosen by the retailer coherently with the business customer purchasing habits) , in the second month that customer will have a significant discount on product A.

Product C and product A (where C is the consequent and A is the antecedent) are selected from the relevant item-sets resulting from MBA, according to the usual expenses of the client.

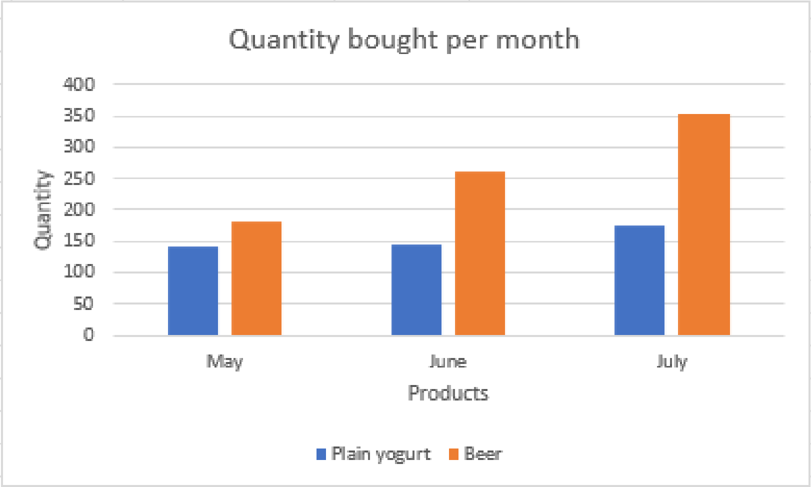
Example:

The business customer with the ID 348544452 shopping habits in the 3 months suggest that he tends to buy a huge quantity of beers (796 beers in total in the 3 months). In the MBA applied at product level, beer is in a significant itemset with plain yogurt, another product that the customer 348544452 buys a lot (460 plain yogurts in total in the 3 months).

PRODUCT_1	prod_desc_1	prod_count_1	PRODUCT_2	prod_desc_2	prod_count_2	t_count_itemset	Support	confidence	lift
21003	Beer	4422	30207	Plain yogurt	4230	516	0.01217	0.116689	1.17001

Doing the average of the quantity of plain yogurt bought, the retailers can create for him the following promotion: if the customer in the month 1 buys more than 153 plain yogurt, in month 2 he will have a 35% discount on beer.

	Average
Plain yogurt	153
Beer	265



The purchasing habits of customer 348544452 for beers and plain yogurt.

CONCLUSIONS & FUTURE SUGGESTIONS

Segmenting customers based on their shopping habits and identification of significant product associations enabled the design of tailored promotions to motivate customers to increase their purchases and keep being loyal to the company.

Some directions for future improvements can be given:

- It would be important to interview customers belonging to each cluster, in order to investigate the reasons behind their purchase preferences. For instance, a survey could be designed to discover why they do not buy or spend less for some product categories.

- It would be also interesting to monitor and compare the performances of the different types of promotions in terms of their impact on customers satisfaction and their ability to increase demand, profit and consumer surplus. This will allow to select the most effective promotions in the future.

- Finally, to implement some of the ideas proposed is important to invest (if the company has not done it yet) in an application that will improve the customer experience.

THANK YOU
FOR YOUR ATTENTION