DATA-DRIVEN MARKETING FOR RETAIL

Marketing Analytics

OUR METHODOLOGICAL APPROACH

The team has followed **a four-step methodology** with the overall **objective of creating value** for the retailer from data, starting by asking **"what is compromising performance?"**. From that, current deficiencies are identified, and business-oriented practical recommendations to improve performances are designed.

IDENTIFICATION → **EXPLORATION** → **RECOMMENDATION** → **IMPLICATION**

IDENTIFICATION

Identify through preliminary **descriptive analysis** the possible deficiencies, thus fronts of potential improvements

Descriptive analysis

- Sales
- Stores
- Products
- Customers

EXPLORATION

Explore the deficiencies identified through **data analytics methods** to figure out if and how to tackle them

Data analytics methods

- Kruskal-Wallis
- Market Basket Analysis (MBA)
- Hierarchical Clustering
- K-means

RECOMMENDATION

Develop recommendations to deal with the deficiencies **driven by data** and **marketing theory** to improve performance

Practical recommendations

- Operation optimization
- Top stores best practices
- Cross-selling
- New loyalty program

IMPLICATION

Analyse the **economic and business impacts** of the recommendations implementation and next steps

Implications

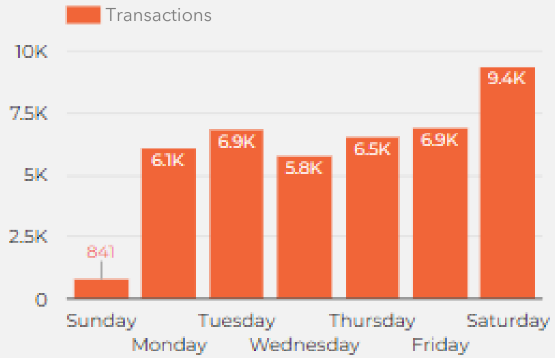
- Economic potential
- Advantages and challenges
- Rollout and next steps

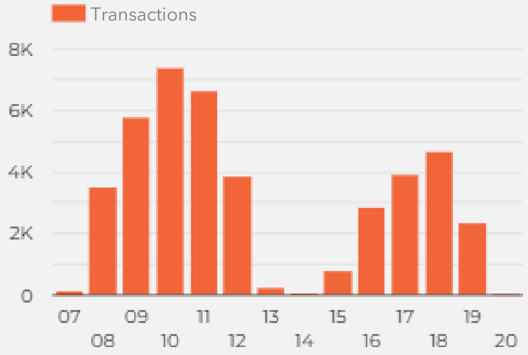
DIFFERENCE BETWEEN TIMEFRAMES AND STORES

Saturday and Sunday stood out in the amount of sales, as well as late morning and late afternoon. The store sizes have clear differences on purchase behaviour when comparing ticket and total sales.

SALES

In all locations and all store sizes, **most** of the transactions occurred **from Monday to Saturday**. The most popular time of the day is in the **morning**, from **9:00** to **11:00.**





STORES

There are **8 stores** of **3 different sizes** (900 [1], 600 [2] and 400 [5] sqm) and **3 different locations** (suburb main city, large town and main city/small town). The stores open from 7:30 to 20:30. Only the store of 900 sqm opens on Sunday.

The stores of **600 sqm** have the **highest** average and median **ticket**, whereas the total sales of the 900 sqm store alone is higher than the five 400 sqm and similar to the two 600 sqm.

Ticket

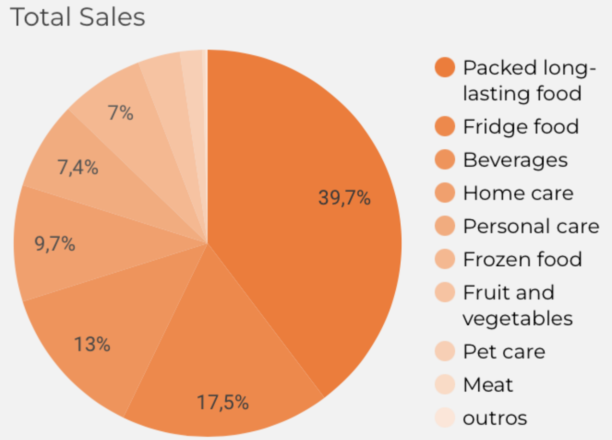
400 sqm	600 sqm	900 sqm
Average: €12.67	Average: €18.12	Average: €14.88
Median: €9.05	Median: €12.02	Median: €9.76
Total: 28.1%	Total: 36.4%	Total: 35.4%

A FEW OUTSTANDING PRODUCTS AND CUSTOMERS

Only three product's families* account for more than 70% of the sales, whereas the bottom five account for only 0,4%. Besides, there is a clear different levels of loyalty among customers.

PRODUCTS

Long-lasting food, fridge food and beverage are the main families of products in total sales. The families of meat, garden/vehicle, kitchen, outdoors and office supplies, besides a family non-identified, have extremely low sales.



PURCHASES

Purchases were made by 225 registered customers. Out of them, 35 **cashiers were identified**. The reasoning behind this and the effect on the loyalty program is explained later.

Considering only real customers (i.e., 190), 77 customers purchased on all 3 months, 36 customers on 2 months, and 77 only 1 month. The 3-month customers shop on average every week, the 2-months customers shop once or twice a month, and the 1-month customers went only once.

€ 15.12
Average ticket of all 42429 purchases
(median: €10,13)

1065 (2.5%)
Purchases of customers
5,6 average purchase frequency in 3 months

*Product's family description list related to its ID is in the Annex 1.

GOAL STATEMENT AND BREAKDOWN

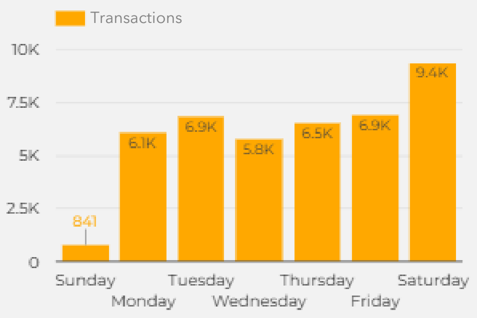
Our methodology brought us to breakdown the starting question «what is compromising performances?» into more **manageable pieces**. This tree summarizes the structure of the presentation, which was divided into four branches. The *exploration* and *recommendation* steps overlap, as the color scale indicates all along the presentation; thus, they will be presented together.

ASSESS SUNDAY OPENINGS AND DISTRIBUTE FLOWS

From the sales data analysis, it was clear that the store 576, the only one that opens on Sunday, should stay closed as well on that day.

Sunday has significantly lower sales

The total sales and number of transactions on Sunday's for the period of analysis were **significantly lower** (Kruskal-Wallis p-value < 0.01) compared with the other weekdays. If closing is not an option, the **optimal working hours** are from 9:00 to 12:00. We are aware that for competitive reasons* the supermarket shall be open on Sunday but data suggest that, even considering the narrower opening time, the total transactions are lower.



Distribute workload based on the peak hours

Regardless of the store or day, the customer flow during the day is the same. There is a **high flow in the morning** from 8:00 to 12:00 **and** the in the **evening** from 16:00 to 18:00.

Besides, **Saturday has the highest flow**, whereas there are no differences between the other weekdays (except Sunday).

From that, the **employees effort** of, e.g., cashier, assistants, customer service, stock clerk should be **managed according to the peaks**.

An alternative idea, even though it would also involve other functions than marketing, could be to incentivize customers to come during the **day's lowest peaks** to redistribute workload ad customer flows, for a higher efficiency and improved customer experience (e.g., less queues).

*In this case, an example of competitive reason is to maintain market share in a specific town or neighbourhood.

LEARN FROM THE BEST STORES PRACTICES

The stores 624 and 519 present superior performances regarding the total sales and average ticket of non-cashier customers. They should be analysed to understand if it is due to the location or execution.

Location

624 is a 400 m^2 store and 519 is a 600 m^2. From their description, they are **not in the same city**. There can be factors related to the location of the stores that can be **explored in the others** or even when thinking of **opening a new one**.

Those factors can be related location-related factors, such as the **purchasing power** of the local population, the store **neighbourhood characteristics**, the **proximity to public transport or main streets** of the city.

Execution

Both stores have different sizes, thus **different execution of retail practices**. This can be related to **store layout**, **product distribution** around the store, **product positioning** and availability on the shelves, **customer service** and attendance.

By understanding these factors, it is possible to **create a best practice standard** to be applied and followed in other stores of the similar size or location.

Store Size	Store ID	Revenue	Avg. Ticket
900 sqm	576	35.4%	14.88€
600 sqm	**519**	**21.2%**	**20.63€**
600 sqm	552	15.2%	15.50€
400 sqm	588	4.2%	10.37€
400 sqm	561	5.5%	13.19€
400 sqm	516	3.6%	17.94€
400 sqm	**624**	**10.4%**	**13.62€**
400 sqm	606	4.5%	10.23€

MBA INSIGHTS ARE LIMITED TO SPECIFIC TARGETS

There are enough transactions to identify promising rules that worth marketing investments but the majority (+50%) of the baskets have four or less products and most (80%) less than nine products.

Support is really low, but promising

As the size of the basket small, the support is at maximum 4.5% for the most sold products, which does not bring any insight. These rules exist just because the products are bought a lot.

The **most promising rules** have confidence from 20% to 50% and lift from 1.7 to 6, however the support is around 1%. This means that the support comes from around 450 transactions, which correspond to around 80 customers (using customer's average purchase frequency). Therefore, the rules are relevant for **targeted marketing actions based on customer's purchase history**, but **not** for **massive undifferentiated campaigns**. A sample of the rules obtained through MBA is in Annex 2.

The power of the rules

The table presents some examples of rules found. The support reinforce that those rules will be **triggered by specific customer segment**. However, from confidence and lift, **there is evidence** that these rules are not random.

Possible marketing strategies are reallocating products in-store, use visual merchandise next to the products to trigger the rule, or leverage on the loyalty program to recommend based on purchase history.

Left-hand side	=>	Right-hand side	Support	Confidence	Lift
{House cleaning detergents, Toilet paper}	=>	{Laundry soap}	0,0051	0,47	4,1
{UHT cream}	=>	{Ravioli/dumplings (fresh)}	0,0068	0,20	3,8
{Pasta, Sweet snacks}	=>	{Fruit juices}	0,0068	0,28	2,8
{Wiener/Wurstel}	=>	{Mozzarella}	0,014	0,24	2,0

CURRENT LOYALTY PROGRAM IS NOT EFFECTIVE

Cashier

There are 35 customer ID which were identified as cashiers. They have a high purchase volume and an **average daily frequency superior to twice a day**. This is considered unfeasible for both B2C or B2B customers. This happens for **three reasons**:

- There is not perceive benefit by the customers of owning a loyalty card, i.e., the **rewards aren't attractive**
- The **main benefit is the discount** on specific products that the customer would already buy anyways
- There is a motivation for the cashiers to offer the own loyalty card as they **win rewards without purchasing**, without restrictions

Customer

There is **no significant difference** (Kruskal-Wallis p-value<0.01) of the basket size, average ticket and unit price between transactions of customers registered in the loyalty programs and transactions made with the cashiers loyalty cards.

It is possible to say that the **loyalty program does not make the customers more profitable**. Therefore, the program has a lot of potential for improvements.

There is no information about their loyalty, as it is not possible to say if a cashier transaction is made by an previous customer. However, the **majority of the loyalty card customers** has **visited at least once a month a store** for the period of analysis.

Retailer

For retailers, loyalty programs can be useful ways to **drive sales, build frequency, and improve customer satisfaction**. Their potential is proven to become significantly higher if the company is able to design it in a **personalized** way; in fact, according to Bond's 2019 REDUX report, when companies do personalization well, it creates a **6.4x lift** in member satisfaction with the loyalty program[1].

Currently, the company's approach towards their program is completely undifferentiated; POS data risks turning from a source of customer value to overhead costs, since both gathering data and maintaining it has a cost.

[1] https://loyalty360.org/content-gallery/research-and-reports/the-loyalty-report-19-redux-the-new-story-of-loyal

RECOMMENDATION OF TAILORED LOYALTY PROGRAM

Designing **a stratified, data-driven loyalty program** allowing exploitation of customer base potential by simultaneously increasing loyalty of the devoted customers and inducing loyalty in the others.

Introduction of new program's layers

This objective can be achieved through a **layered game-like loyalty program,** where customers are split into **four levels**, based on their loyalty (defined through the characteristics that emerge from data). The higher the level of the customer, the better are their advantages and benefits.

Coal **Bronze** **Silver** **Gold**

Actions tailored to specific needs

With the proposed loyalty program, it is possible to pursue different objectives:

1) **Design actions towards consumers** in a more **targeted** way, to provide groups with different characteristics and behaviors (unveiled through data) with the most suitable experience
2) **Discover**, **understand**, and **fulfil** the retailer's **needs** and **deficiencies** and achieve its **goals**
3) **Monitor** and **influence** the **evolution of consumer patterns** and control the fulfilment of the customers' goal

GAMIFICATION INCREASES CUSTOMER ENGAGEMENT

Gamification is a renowned powerful approach to **incentivize certain behaviors** of the target of reference by designing game-like actions[1].

Incentivize behaviours → **Assign levels** → **Offer benefits**

Experiment is more engaging for customers

Customers are **clustered into different levels**, among which they can move as if they were in a **game**, by taking some actions that the retailer sets through data analysis. The more loyal a customer, the higher the level he/she reaches, and **the richer the benefits** he/she gets.

From the data, many metrics were created, such as **ticket, frequency, timeframe, product and store related metrics**. This allows a proper understanding of each customer behavior and characteristics, which in turn drive the clustering.

Positive economic impact for the retailer

The retailer can **design and update the program** to fulfil its needs and achieve its goals by "**moving the levers**" that the loyalty program allows (i.e., the marketing actions to target customers) by defining the actions that customers should do to reach the level above

For example, **the retailer can decide** to push certain product categories, make customers come in specific periods of the day or week, incentivize increase in frequency or average ticket value. These actions **depend on the goals** that the retailer wants to achieve.

[1] https://www.insidemarketing.it/gamification-come-strumento-di-marketing/

LAYERS OF THE LOYALTY PROGRAM

The goal of this program is to reward customers according to the different value they provide to the retailer and to create suitable incentives for each group of customers to strive for higher benefits.

Customer value is a combination of factors

A customer's value should not be defined mono-dimensionally, i.e., the through the **overall expenditure**. Instead, it should be related to other variables as well, namely the **purchases per day**, which better represents the frequency of purchase, and the **average ticket**.

The overall expenditure alone cannot capture different **shopping habits** or the **customer's behavior**; in fact, given the same total ticket, a customer that comes more frequently and makes a lower average ticket should be **engaged differently** compared to one coming less frequently but that makes higher tickets according to this program's philosophy. Loyalty, therefore, is defined by the parameters mentioned here, and customers have been clustered according to this definition.)

Internal and external validity check

To decide the optimal number of clusters, both <u>quantitative</u> and <u>qualitative</u> criteria have been followed.

- Quantitatively, through K-means (after scaling the data), the average silhouette (**silhouette method**) and the WSS within-cluster sum of squared errors (**elbow method**) were computed for each K (further details in the following slide)
- Qualitatively, it does not make sense to pick neither **too many** clusters – to avoid the program becoming too complex to be managed and discourage customers – nor **too few** (e.g., K = 2) – as it would not change from the current one. Moreover, the size of the cluster should be inversely proportional to the level they reach.

LAYERS OF THE LOYALTY PROGRAM

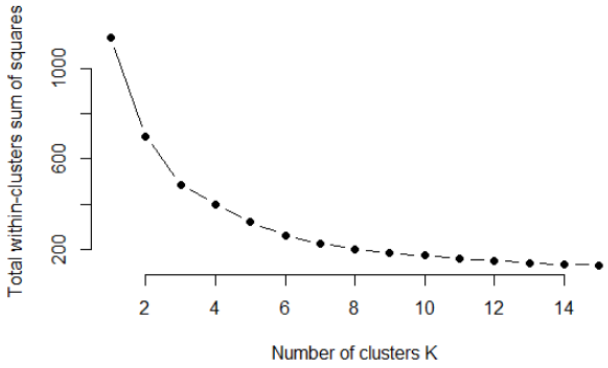
There is a clear difference between the four clusters of customers, which is supported qualitatively and quantitatively.

Four clusters is the perfect compromise

From the results, **K = 4** is the perfect compromise between these conditions: it presents the **highest average silhouette** after K = 2 (which was excluded by the qualitative conditions), and the **WSS presents a slight elbow** right in that point.

The resulting loyalty program rewards customers with a higher purchase frequency and total expenditure. A high average ticket is not enough.

K	Avg. silhouette
2	0.59
3	0.36
4	**0.53**
5	0.38
6	0.37
7	0.27



<u>COAL CUSTOMERS</u>

- N = **112**
- Avg. Ticket = 8.3€
- Total Exp. = 25€
- Monthly Exp. = 12.93€
- Purchases/day = 1

<u>BRONZE CUSTOMERS</u>

- N = **40**
- Avg. Ticket = 24€
- Total Exp. = 98€
- Monthly Exp. = 46.31€
- Purchases /day = 1

<u>SILVER CUSTOMERS</u>

- N = **30**
- Avg. Ticket = 13€
- Total Exp. = 174€
- Monthly Exp. = 58.36€
- Purchases /day = 1.1

<u>GOLD CUSTOMERS</u>

- N = **8**
- Avg. Ticket = 26€
- Total Exp. = 779€
- Monthly Exp. = 259.82€
- Purchases /day = 1.2

UNDERSTANDING WHO THE CUSTOMERS ARE

The **customers' characteristics** are not related to the layers of the program, but are required to understand what to suggest to a specific customer to buy, based on what they already buy

Four clusters were identified

Cluster 1 buys slightly more packed long-lasting food than the average but mostly stays on the average for the other product families. It will be named **Food stockers**.

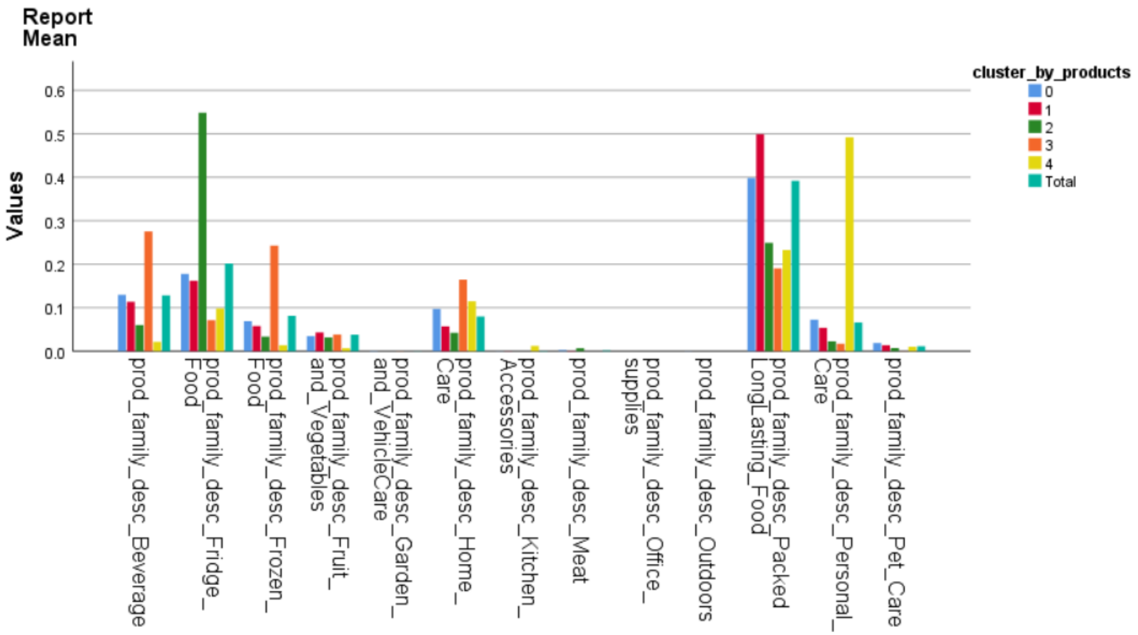
Cluster 2 buys lots of fridge food, while stays below the average on the other relevant categories. It will be named **Lazy cooks**.

Cluster 3 buys the most relevant categories in a balanced way. It has no particularities as the other clusters. It will be named **Ordinaries**.

Cluster 4 shows by far the highest personal care percentage, with all the other relevant ones being far from the average apart from home care. It will be named **Opportunists**.

Details on the cluster size assessment of this clustering method are on Annex 3. Compared to the behavioral ones, these clusters are less significant (no clear elbow and lower silhouette). More non-cashier observations are needed.

This time, the K-means method was performed on the ***percentage of product family* purchased*** *by each customer. Cluster 0 are the cashiers and total is the whole sample.*



*Product's family description list related to its ID is on the Annex 3.

COMBINING BOTH CLUSTERING IS POWERFUL

By merging the insights on customers' behavior and characteristics, we built a dashboard-like intuitive analysis tool to monitor their development over time, and to create more punctual and effective actions.

Behavior Clustering
(by total ticket, average ticket, purchase frequency)

Characteristics Clustering
(% product family purchased)

%	1	2	3	4	Tot
1	29	14	14	4	61
2	13	2.5	0.5	0	16
3	13	4	1	0	18
4	3.5	1	0.5	0	5
Tot	58.5	21.5	16	4	100

*% **number** of non-cashier customers*

%	1	2	3	4	Tot
1	10	17	25	34	86
2	3	2	1	0	6
3	2	2	2	0	6
4	1	0	1	0	2
Tot	16	21	29	34	100

*% on overall **expenditure** of non-cashier customers*

	1	2	3	4
1 2	58,5% cust. 32% sales		18,5% cust. 60% sales	
3 4	21,5% cust. 5% sales		1,5% cust. 3% sales	

PINK

This group of customers is made up of the ones used to purchase day-to-day products (fresh and packed food), but whose purchase frequency is low. Therefore, a driver to push them **horizontally** towards the green area is **frequency**: the retailer should incentivize them to come more often to the store, in addition to the other new loyalty program actions. Note that this cluster might also include the potential **new clients**, (low frequency and total ticket); there is potential.

GREEN

This group is made up of the most important customers for the retailer, because they come very frequently, buy day-to-day products, and account for **60% of sales**. The loyalty program should **reward them** for this, and the retailer must be able to **communicate** this the best way, both to engage this customers as much as possible, but also, for making this status appealing to others. The loyalty program should focus on both up- and cross-selling.

RED

This cluster is made up of the **worst customers** for the supermarket. They are the ones with shopping behaviors that are far from the day-to-day clients; plus, they show the lowest frequency and total and average ticket. As they seem disloyal to the supermarket, they should be the **least rewarded** by the program. To avoid losing profitability and "cannibalizing" other groups, the supermarket should make them move first **vertically** (cross-selling activities to push groceries), and afterwards work on building **frequency**.

BLUE

This group of customers is only made up of 1.5% of the total sample. They have an unusual behavior with respect to the others in terms of purchased products. As they do not have problems in terms of frequency and total ticket, **cross-selling actions** should push them to buy not only the current categories, but also more day-to-day ones, like **groceries**. Another good strategy could be to push **up-selling** for the categories that they currently buy - like home care for 3s - since their average ticket is not very high.

STRUCTURE OF THE LOYALTY PROGRAM

It is a gamified loyalty program aiming to engage the customer to **regular interaction** and **forming new shopping habits**. In order to make progress in the four customer levels, and hence unlock new benefits.

Advancing through levels

Just like in a game, customers are **encouraged to spend and engage more** to accomplish their missions and **move up through the ranks**, unlocking new privileges (once every 5-6 months, at goal completion). There are four levels available (coal, brown, silver, gold), each with **actions and rewards created based on the sample data**.

Communication is a key and the retailer must be in contact with the customer to **inform** them about the **structure of the program, required actions and available promotions**. Using subscription data, retailer can send messages to their customers.

Incentivized actions

The actions defined for each level are based on the **data analysis** performed. The **system incentivises buying the products** which customers from the next tier often buy, have sufficient lift and that encourage more frequent store visits. The points awarded for the given action are dependent on the revenue earned by the retailer.

Completed: **3/4**

Buy 3 packages of **Cookies** to unlock the next level!

Getting Rewards

In order to encourage the recurring behaviour, with advancing through the levels, customers gain access to specific rewards. **The higher the level, the better profits for the customer,** as the rewards incrementally increase in value.

Examples of rewards:

- Coal: fast track queue, self-scan device and self-checkout
- Bronze: recipes and video material related to products
- Silver: enhanced point collection, lotteries
- Gold: free delivery a certain times a month

Tracking the progress

To make customers feel the sense of accomplishment they have to be aware of their progress and how much they are lacking to reach the next level. They will be provided with a **dashboard** enabling them to **monitor their accomplishments, current level and requirements to reach the next one.**

Current level:

You are lacking **1500 points** to reach the silver level!

Practical **illustration** of how the retailer's information system can define the loyalty program **rewards** and **challenges** from the data sample for level 1 (**coal** customers) and level 3 (**silver** customers), with two customer IDs as example.

1 Coal

Frequency: **4th** Avg. Ticket: **4th** Monthly Exp: **4th**

The analyzed customer is: **899839452**. He came once in May and June, twice in July, and spent 2.93 € on avg. He is **pink**.

Rewards in Coal level:
Since it is the entering level, **visible** (but **cheap** for the retailer) benefits like the self-scan reader, or the checkout fast-track queue can be effective for both generating immediate gratification for novices, and for tempting others to join the program.

August challenges for 899839452:
Frequency & **Average Ticket**: make at least four different tickets > 8€ in four different days of the same month. Being pink, the main concern is to increase these two KPIs, to stimulate retention.
Products: (Cross-selling) as his last basket contained **UHT milk**, the system should recommend products that:

- Bronzes often buy, to push 899839452 to try new products that bronzes like, to push him to imitate them.
- Present a sufficient lift (> 1.5); higher values than 1.5 often correspond to too few transactions.
- Show the highest possible support, which can contribute to drive traffic to the store (frequency).
- Show a consistent confidence, but not a too high one (>= 0.25), to make sure that not too many people buy these products together and hence there is room for marketing actions.

Cookies respect these criteria (more detail in Annex 4). As a coal, a discount on an average cookies brand can be the action to pursue, communicating through an e-mail, and then by reminding it in the self-scan.

3 Silver

Frequency: **2nd** Avg. Ticket: **3rd** Monthly Exp: **2nd**

The analyzed customer is: **947804452**. He came 9 times in May, 4 in June, 5 in July, and spent 12.26€ on avg. He is **green**.

Rewards in Silver level:
Silver customers deserve **better benefits** than 1s and 2s: their point collection can be **enhanced** (e.g., 1 euro = 2 points, and points for each ticket they make), and exclusive initiatives, such as **lotteries**, and "**silver days**" in which they get unique, tailored rewards.

August challenges for 947804452:
Frequency & **Average Ticket**: customers in level 3 show a lower average ticket compared to both 2 and 4. 947804452 reflects this trend, therefore he should make at least one basket of 40 € in August (his all-time maximum ticket was 32€).
Products: being a silver customer, the supermarket can push on both cross-selling and up-selling:

- Cross-selling: he bough ready vegetables in his last basket, so the system should recommend **Canned tuna** (a typical "green" product), because it respects the association rules criteria set above (Annex 5), and because Gold customers like it a lot, and the retailer wants to incentivize them to behave like golds.
- Up-selling: initiatives on products that 947804452 often buys (such as **Parmesan**), either by offering point-related rewards on bigger packs, or by discounting premium brands for the whole month. This is done to incentivize him to get used to make higher avg. tickets, as his one is much lower compared to golds.

THE PROMOTION INITIATIVES CAN BE PROFITABLE

To estimate the impact of our program on increasing customer value, we estimate the contribution created from customers and moving through the different levels depending on the marketing strategy.

Current customers' perspective

Assume that the initiatives for marketing and promotion of products towards customers increase the transition of customers from one cluster to another by 20%, that means, the chances of a successful promotion and the movement of a customer is 20%.

Imagine we show each one of our customers the promotion. In this rough estimation, we derive the economic contribution of each customer as the increase in revenue they generate on a monthly basis to the retailer when they change the cluster (Annex 6).

Assume further that customer changes its behavior entirely to fit into the description of that new cluster, meaning that the customer either buys more often, spends more per visit, or buys more (expensive) products.

In the table on the right, the first column denoted the customer that changes behaviors from one cluster to another. On the second column, we estimate the benefits that that customer provides to the retailer if the promotion is successful.

Transition	Monthly Revenue Generated[1]
Red → Blue	Revenue generated through cross-selling ΔRevenue $_{\text{Red-Blue}}$ = 305% **190** * 21,5% * 20% * 49,40 = 403,60 €
Red → Pink	Revenue generated through cross-selling and increased frequency ΔRevenue $_{\text{Red-Pink}}$ = 47% **190** * 21,5% * 20% * 7,52 = 61,44 €
Pink → Green	Revenue generated aiming at increasing frequency ΔRevenue $_{\text{Pink-Green}}$ = 338% **190** * 58,5% * 20% * 80,10 = 1.780,62 €
Blue → Green	Revenue generated trough cross-selling and up-selling ΔRevenue $_{\text{Blue-Green}}$ = 58% **190** * 1,5% * 20% * 38,22 = 21,79 €

From these rough estimations, it is evident that the program we suggest can contribute with added revenue to the retailer. This analysis can help managers **understand** if and how to **invest in marketing** towards different clusters. As the transition of customers from one cluster to another has different economic impact, the managers can tune their budget on marketing initiatives.

[1]*Current customer base * % of number of customers in that cluster * Probability of successful promotion * Increase of monthly expenses in the new cluster compared to previous cluster*

LOYALTY PROGRAM CAN BOOST REVENUE SIGNIFICANTLY

In order to estimate the impact that new customers can have on the revenue if they decide to join the loyalty program, we can assume how many customers the retailer currently has.

Transition	Monthly Revenue Generated[1]
Red → Blue	Revenue generated through cross-selling: ΔRevenue $_{\text{Red-Blue}}$ = 305% **2216** * 21,5% * 20% * 49,40 = 4.707,23 €
Red → Pink	Revenue generated through cross-selling and increased frequency ΔRevenue $_{\text{Red-Pink}}$ = 47% **2216** * 21,5% * 20% * 7,52 = 716,57 €
Pink → Green	Revenue generated aiming at increasing frequency ΔRevenue $_{\text{Pink-Green}}$ = 338% **2216** * 58,5% * 20% * 80,10 = 20.767,69 €
Blue → Green	Revenue generated through cross-selling and up-selling ΔRevenue $_{\text{Blue-Green}}$ = 58% **2216** * 1,5% * 20% * 38,22 = 254,09 €

It is evident there are potential of **gaining revenue** from the suggested game-like loyalty program. These calculations are based on assumptions, the estimations are made cautiously through educated guesses. However, based on these rough estimation, we see an **increase in the revenue** from new customers. Only the return from the actions on pink to green has the potential to increase the monthly revenue of the retailer by 9,7%, and from all actions on current and new customers by 13,4%.

Future customers' perspective

We assume that the database is a good representation of all the customers that visit the retailer. Out of the **42,429 transactions** in the database, not all are assumed to be of cashiers. We assume that a proportion of the customers that purchase through cashiers enter the loyalty program of the retailer. There are **1065 registered customer transactions** and there are on average **5,6 transactions per registered customer** (1065/190). Furthermore, assume that this relationship holds for all the transactions in the database. This gives us a rough estimate of **7386 customers currently not registered** in the loyalty program, or just not using it. If 30% of these customers enter the loyalty program or start using it for its benefits, we get a customer base of **2216 customers**.

In this analysis, we assume a loyalty program that is successful and has been used by customers for a longer period, and therefore customers' need to change clusters.

In order to arrive at more precise results, a deeper analysis of customer willingness to change cluster and their response to rewards is needed as well as considerations about costs.

[1] *Current customer base * % of number of customers in that cluster * Probability of successful promotion * Increase of monthly expenses in the new cluster compared to previous cluster*

MAIN STRENGTHS AND WEAKNESSES OF THE PROGRAM

STRENGTHS

- Gamification is a way of using data that **might appeal the customers of the future**, i.e., Gen Z and Millennials
- Through technology and structured ideas, **the system can be very easy to understand and to use** also for middle-aged people and families
- A **wider and more loyal customer base** can be achieved
- The **system is extremely customizable**: if the supermarket has some needs (e.g., sell more of a product category, make customers come more often, …), it just needs to reset the parameters of the scale and customers will "execute" them
- The advantage against the previous loyalty program is that this one is **data-driven, more personalized** (different customers are pushed to do different things), based on more levers and rewards, so it allows the supermarket to influence customers in many more ways
- It is a more **targeted approach**: according to the type of customer, the customers will get different benefits, more tailored and meaningful to them and more advantageous for the company

WEAKNESSES

- Some customers **might feel stressed by the program**, as it has many more details than the current one
- Effective **communication is a challenge**
- **Frequent policies update** is required
- It is **not straightforward where to define the threshold** to go from one level to another one
- It is **not clear how to encourage each customer to move from one level to another one** as there are limited information about them
- The proposed program is based on the assumption that it can reach higher statistical significance by following the same procedure but with more data than just that of the non-cashier ones. Now, by looking at the cashier data (which represents 97% of the whole sample) we see that, taking the example of MBA, some unpredicted rules reach sufficient support, supporting our assumption. However, **the team cannot be sure of this until a sufficiently big data sample is analysed**.

CONCLUSIONS

In order to apply the data-driven recommendations into reality, the team proposes a tailored roll-out strategy for the loyalty program and suggests how to enrichen the analysis with additional data.

"Tried-and-true" roll-out strategy

Across many industries over the years, the so-called "**tried-and-true roll-out strategy**" has worked very well when deploying new loyalty programs (e.g., Target in the US) [3].

This strategy begins with **gaining share from existing consumers**, aiming to increase their recency, frequency, and avg. ticket, as we proposed in slide 17. Retailers do this by gaining one additional trip, one additional product per basket at a time. The process of converting the existing customer base to a more loyal, higher-frequency group can take time and should not be rushed. Once the existing customers show attachment to the program, the second step is to **market further out** from its existing trade area to begin gaining share from primary competitors.

According to practitioners [3], the first step can take up to 2 years after the roll-out, and besides, the team has no data about competitors. Therefore, both a **comprehensive economic evaluation** and a **location-wise competitor analysis** must be carried out carefully while designing the program.

With more data, much more can be achieved

The POS data provided to the team is a small sample of the data that the retailer gathers every day in all its stores, both in terms of timespan an in terms of number of attributes.

- With **more observations** (especially of non-cashier customers) the retailer can scale verbatim the presented procedure and reach the **statistical significance** that the sample of data provided does not allow, according to the analysis we chose to carry out (i.e., disregarding the cashier data in the loyalty program design, for both clustering and MBA).
- With **more detailed product data**, such as detailed description at the **SKU level**, the retailer can carry out a deeper **up-selling** analysis, to offer customers data-driven incentives to purchase higher quantities or more premium brands of the products they currently buy, to boost the average ticket.
- With **customers' demographics**, the retailer can organize the **communication campaign** to sustain the program and make deeper customer analysis.

[3] «Retail Analytics: The Secret Weapon», Emmet Cox (2012), Wiley

ANNEX 1: PRODUCT'S FAMILY DESCRIPTION

By looking at the components of the product family, we gave an interpretation to the family number, in order to analyze them in the most realistic and practical way.

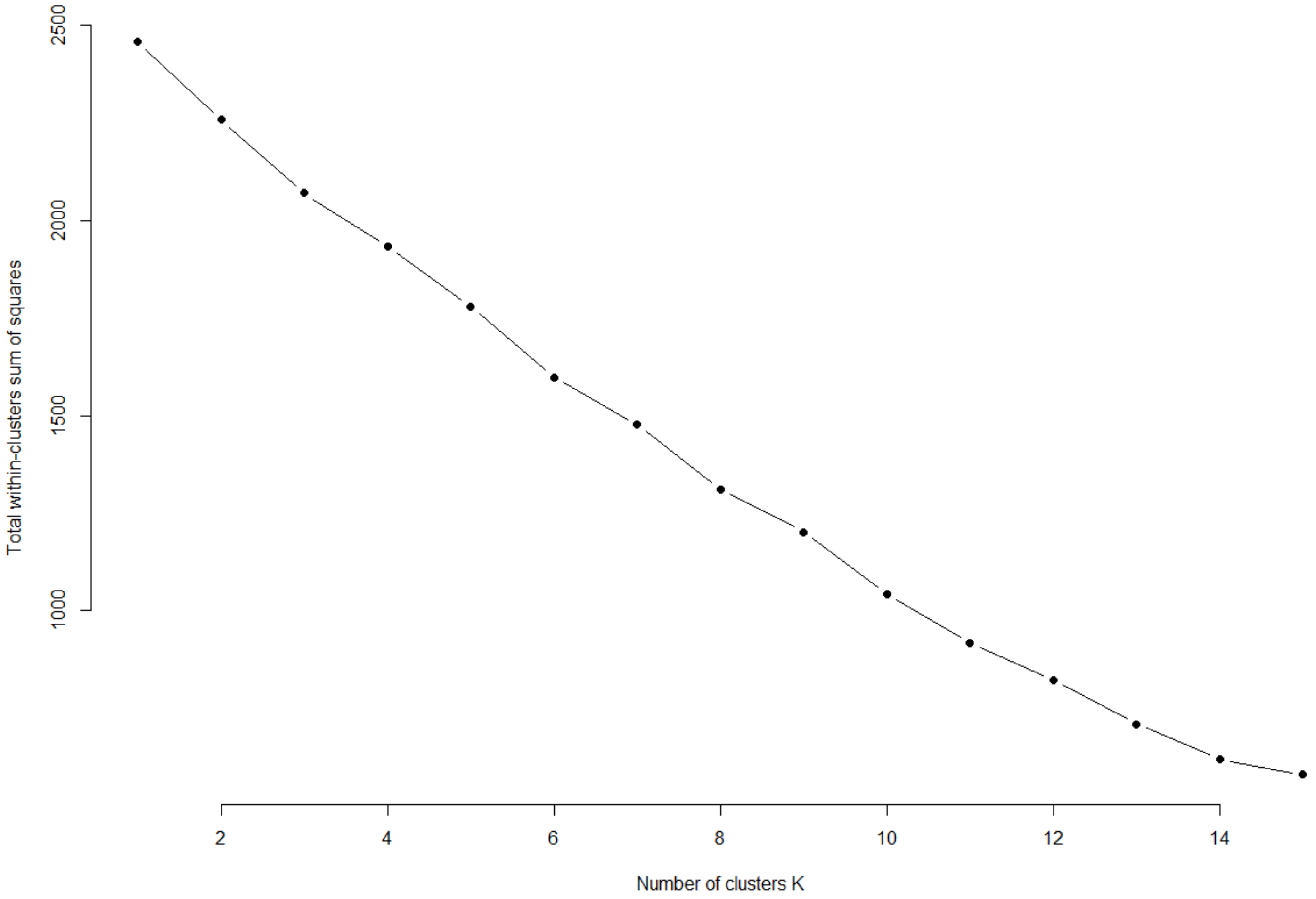
Family ID	Team's Interpretation	Family ID	Team's Interpretation
1	Packed long-lasting food	10	Fish
2	Beverages	50	Technology
3	Fridge food	51	Clothing
4	Frozen food	53	Garden and vehicle care
5	Home care	54	Kitchen accessories
6	Personal care	55	Office supplies
7	Pet care	56	Toys
8	Fruit and vegetables	57	Outdoors
9	Meat	58	Services

ANNEX 2: MARKET BASKET ANALYSIS RULES RESULTS

Left-hand side	=>	Right-hand side	Support	Confidence	Lift	Transactions
{Olives}	=>	{Pickles}	0,0074	0,22	6,2	314
{Vegetables in oil}	=>	{Pickles}	0,0067	0,21	5,7	284
{House cleaning detergents, Toilet paper}	=>	{Laundry soap}	0,0051	0,47	4,1	218
{Orange soft drinks}	=>	{Cola}	0,0083	0,29	3,9	352
{UHT cream}	=>	{Ravioli/dumplings (fresh)}	0,0068	0,20	3,8	288
{Pickles}	=>	{Wiener/Wurstel}	0,0077	0,21	3,4	328
{Olives}	=>	{Canned tuna (oil)}	0,0091	0,28	3,3	384
{Pickles}	=>	{Canned tuna (oil)}	0,0099	0,27	3,2	419
{Jam}	=>	{Rusks}	0,0075	0,22	3,1	320
{Mozzarella, Pasta}	=>	{Canned tuna (oil)}	0,0068	0,26	3,0	287
{Vegetables in oil}	=>	{Canned tuna (oil)}	0,0082	0,25	3,0	349
{Cookies, Fruit juices}	=>	{Rusks}	0,0053	0,21	3,0	224
{Pasta, Sweet snacks}	=>	{Fruit juices}	0,0068	0,28	2,8	289
{Fresh milk, Pasta}	=>	{Canned tuna (oil)}	0,0051	0,24	2,8	216
{Gloves and tools for cleaning}	=>	{Laundry soap}	0,011	0,28	2,4	477
{Tomato}	=>	{Mozzarella}	0,012	0,28	2,4	508
{Wiener/Wurstel}	=>	{Canned tuna (oil)}	0,013	0,20	2,4	533
{Tomato}	=>	{Ready vegetables}	0,012	0,29	2,3	527
{Cotton and handkerchiefs}	=>	{Laundry soap}	0,011	0,24	2,1	454
{Cola}	=>	{Beer}	0,016	0,22	2,1	686
{Wiener/Wurstel}	=>	{Mozzarella}	0,014	0,24	2,0	614

ANNEX 3: CHARACTERISTIC'S CLUSTER SIZE ASSESSMENT

Characteristics-based clustering is relevantly less significant than the behavioral one. There is no evident elbow, and the highest average silhouette was reached for K = 4.

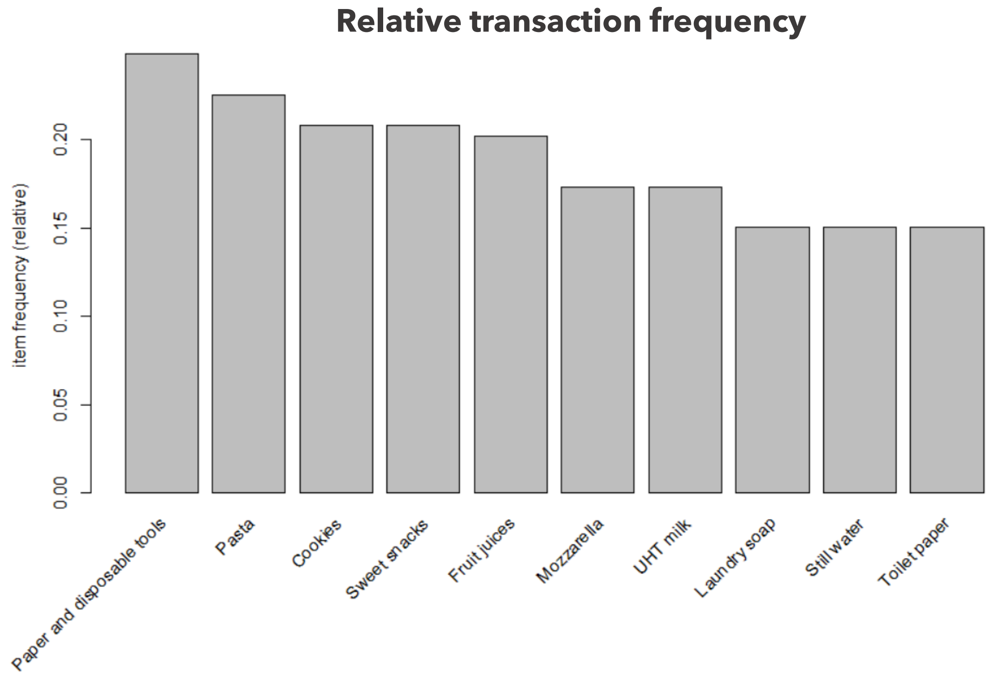


K	Average silhouette
2	0.11
3	0.1
4	**0.28**
5	0.18
6	0.09
7	0.06

ANNEX 4: CROSS-SELLING RULES RESULTS FOR COALS

Rules for cluster 1, sorted by support, that have *UHT milk* on the left side (antecedent). *Cookies* shows a good support, a borderline (but sufficient) confidence, and a good lift (1.76). The bar plot on the right shows the products that are most frequently purchased by cluster 2 (Bronze), and cookies are among them.

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{UHT milk}	=>	{Sweet snacks}	0.064	0.37	0.17	1.76	11
[2]	{UHT milk}	=>	{Cookies}	0.064	0.37	0.17	1.76	11
[3]	{UHT milk}	=>	{Paper and disposable tools}	0.058	0.33	0.17	1.34	10
[4]	{UHT milk}	=>	{Laundry soap}	0.052	0.30	0.17	2.00	9
[5]	{UHT milk}	=>	{Plain yogurt}	0.040	0.23	0.17	1.76	7
[6]	{UHT milk}	=>	{Pasta}	0.040	0.23	0.17	1.04	7
[7]	{UHT milk}	=>	{Sugar}	0.035	0.20	0.17	1.65	6
[8]	{UHT milk}	=>	{Patisserie}	0.035	0.20	0.17	1.44	6

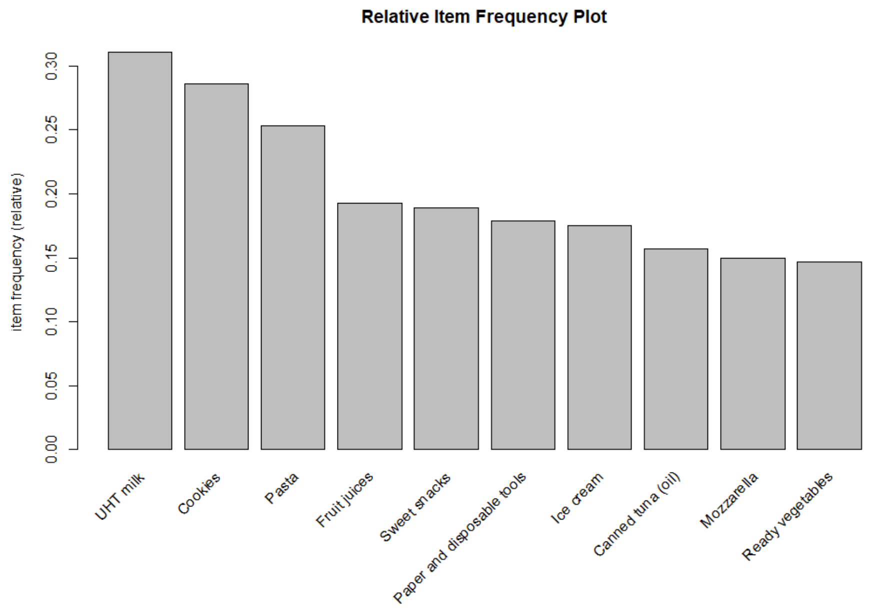


The team is aware that, by looking at rules within clusters, the total **count of transactions** to support these rules is very low. However, the team considers this data as a small sample of the data that the retailer gathers every day. Therefore, it is assumed that, with higher amounts of transactions, **the same principles can be applied**, and statistically significant results can be reached.

ANNEX 5: CROSS-SELLING RULES RESULTS FOR SILVERS

Rules for cluster 3, sorted by support, that have *ready vegetables* on the left side (antecedent). *Canned tuna* shows a good support, a confidence slightly above the threshold, and a good lift (2.3). The bar plot on the right shows the products that are most frequently purchased by cluster 4 (Gold), and *canned tuna* is among them.

```
    lhs                    rhs                             support confidence coverage lift count
[1] {Ready vegetables} => {Cookies}                        0.057   0.39       0.15     1.4  16
[2] {Ready vegetables} => {Canned tuna (oil)}              0.054   0.37       0.15     2.3  15
[3] {Ready vegetables} => {Pasta}                          0.054   0.37       0.15     1.4  15
[4] {Ready vegetables} => {UHT milk}                       0.050   0.34       0.15     1.1  14
[5] {Ready vegetables} => {Breakfast cereals}              0.046   0.32       0.15     3.0  13
[6] {Ready vegetables} => {Fresh milk}                     0.039   0.27       0.15     2.2  11
[7] {Ready vegetables} => {Paper and disposable tools}     0.039   0.27       0.15     1.5  11
```



ANNEX 6: KPIS BY COLORED CLUSTERS

Colored clusters	# Customers (% total)	Sales (% total)	Average monthly expenditure
Green	35 (18,5%)	€ 10.863,02 (59,8%)	€ 103,79
Blue	3 (1,5%)	€ 590,17 (3,2%)	€ 65,57
Pink	111 (58,5%)	€ 5.728,61 (31,5%)	€ 23,69
Red	41 (21,5%)	€ 994,20 (5,5%)	€ 16,17