



**POLITECNICO**  
MILANO 1863

# MARKETING ANALYTICS

## GROUP 6

Landini Francesco

Lu Shiying

Ranganathan Narasimhan

Zhang Yasong

10803765

10781219

10750472

10754482



# AGENDA

- Objective
- Data pre-processing
- General analysis
- RFM analysis
- MBA analysis
- Marketing suggestions
- Appendix



# OBJECTIVE

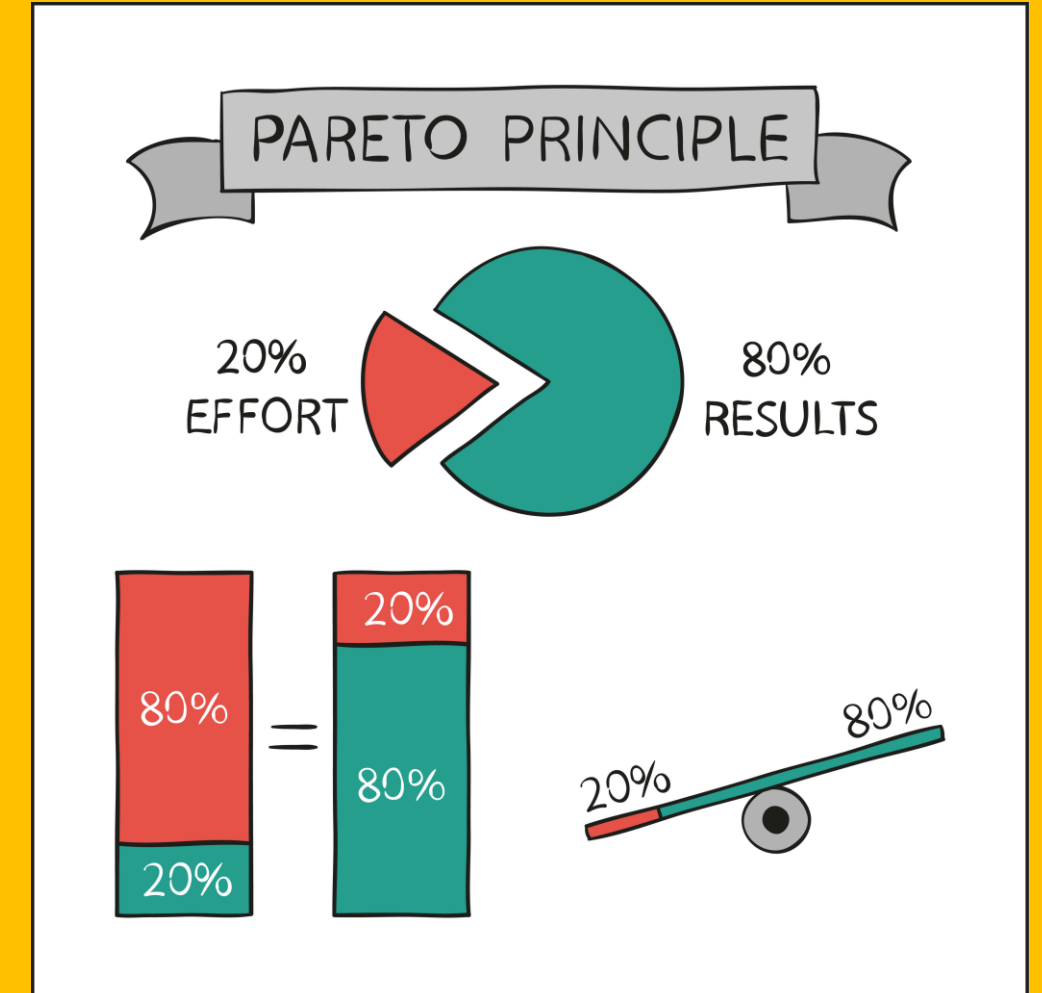
To analyze the dataset and find insights about who the valuable customers are, to carry out specific marketing actions and disregard the remaining customers.



*“Half the money I spend on advertising is wasted; the trouble is I don't know which half.”*

*- John Wanamaker,*

*father of modern advertising and a "pioneer in marketing."*



# DATA PRE-PROCESSING AND GENERAL INFORMATION



Check missing values

Outlier handling

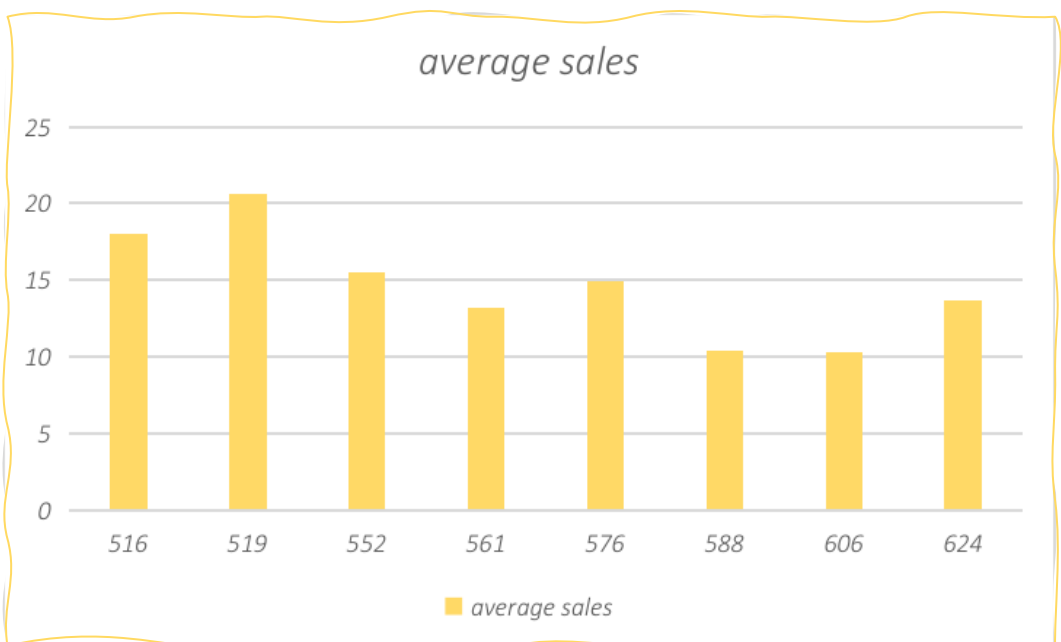
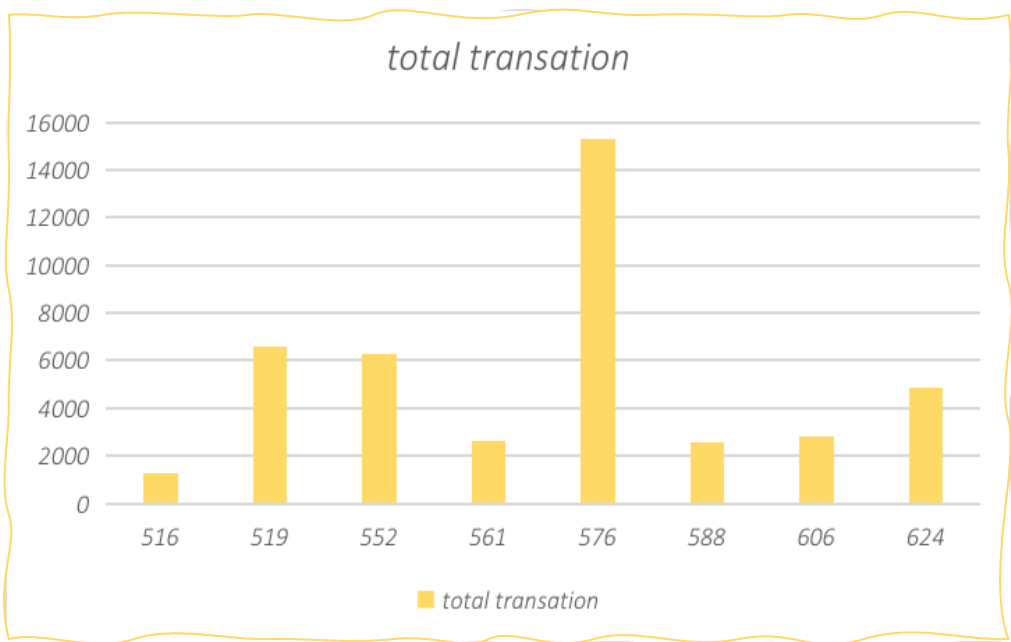
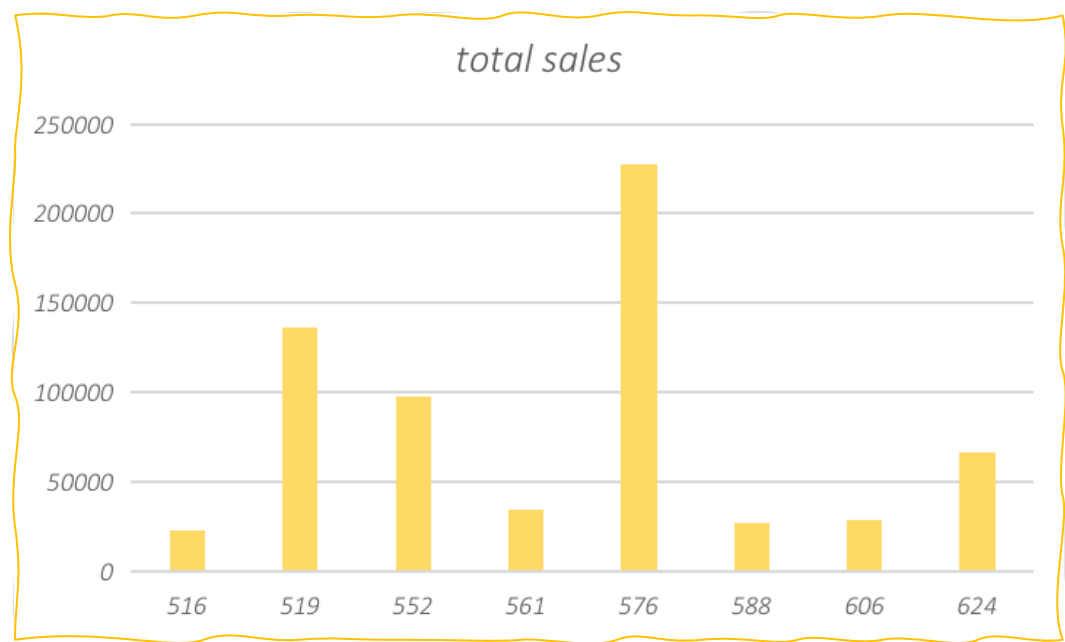
Such as frequent purchases from the same customer due to borrowing a loyalty card (tickets have similar monetary value, hence the term borrowers), they account for about 97% of transactions.

0 quantity (Consider as 1)

Consider as 1

Data Format Conversion

# GENERAL ANALYSIS- PERFORMANCE OF STORES



## SALES

### REVENUE PERFORMANCE

We see that Store #576 ranks first in terms of total sales and total transaction times because it is the largest store and maybe have more products and services compared to others.

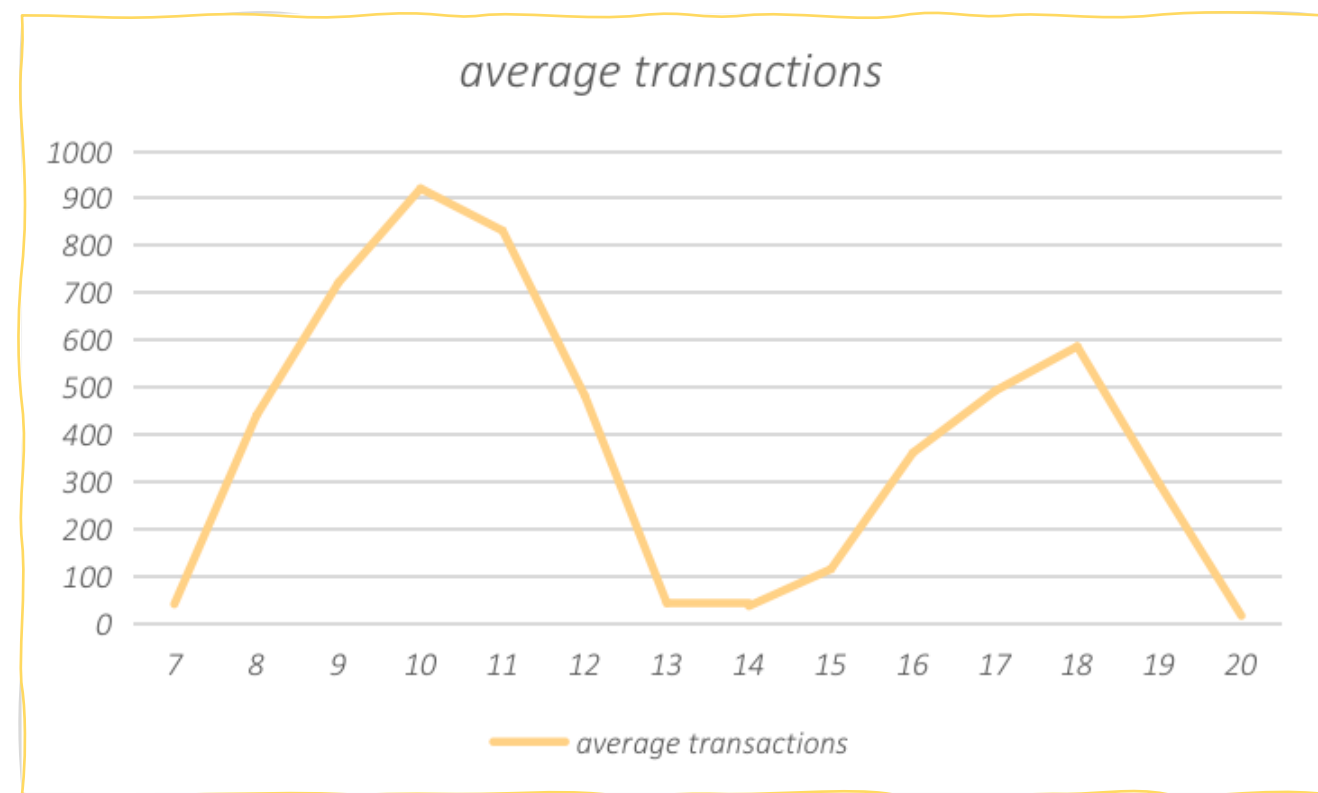
Overall, the total sales and the total transaction times have a similar distribution, and we can see that it is directly proportional to the size of the store. We can divide the stores into high, medium and low levels: High- 576, Medium- 519, 552 and Low- 516, 561, 588, 606, and 624. Considering the size of stores, 624 performs much better than others.

## AVERAGE SALES PER TICKET

### PURCHASING POWER

In order to evaluate the actual profitability of the store, we have to look at the average sales per ticket and we see that #576 no longer holds the top position, whereas top performer is #519 which is in the largest city. This indicates that the store has low customer traffic but the average purchase per visit is high. This reflects the purchasing power of the store's customers of the largest city. Therefore #519 is suitable for showcasing high quality and priced items. #576 performs worse than #516 which has the least total sales and total transactions. So, #576 can focus on low priced items.

# GENERAL ANALYSIS- PERFORMANCE OF STORES

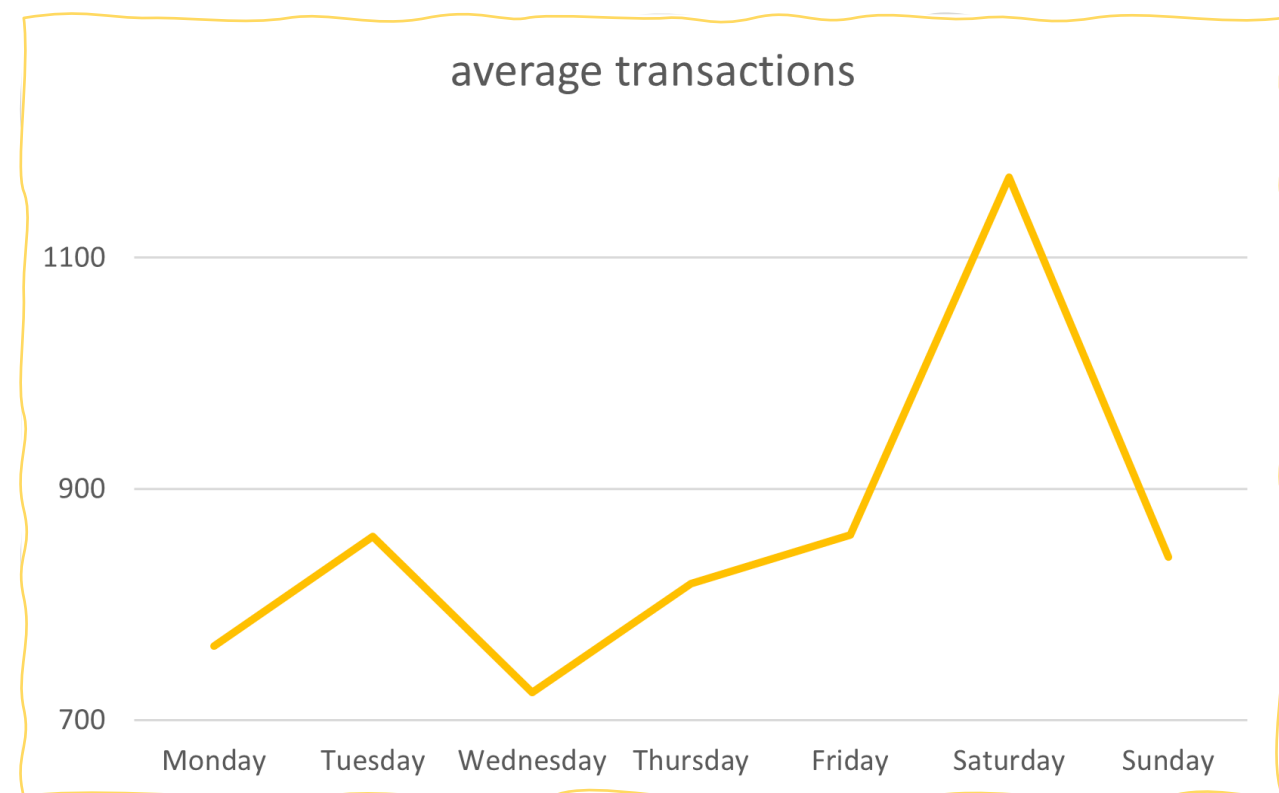


## DAILY TRANSACTION TREND

There are two peaks of customers in a day. One at 10:00 and another at 18:00. The first peak maybe due to visits by senior citizen and the second peak maybe due to customers after their office hours.

The trough is between 13:00 to 14:00, which may be because customers have lunch break and the stores being closed and we found that 2 stores #516, #576 are open.

Therefore, we can do some discounts to drive traffic during this period. During the peak flow of people, we should pay attention to increasing the employees at cashcounter and replenishing goods in time to improve consumer satisfaction.

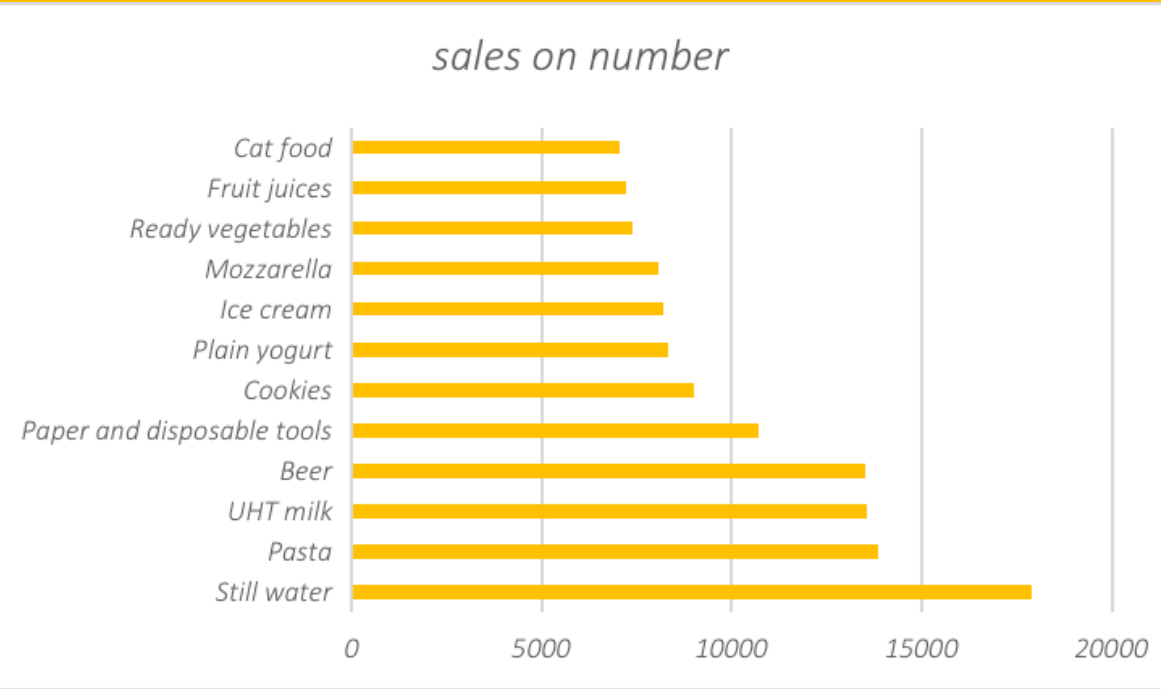
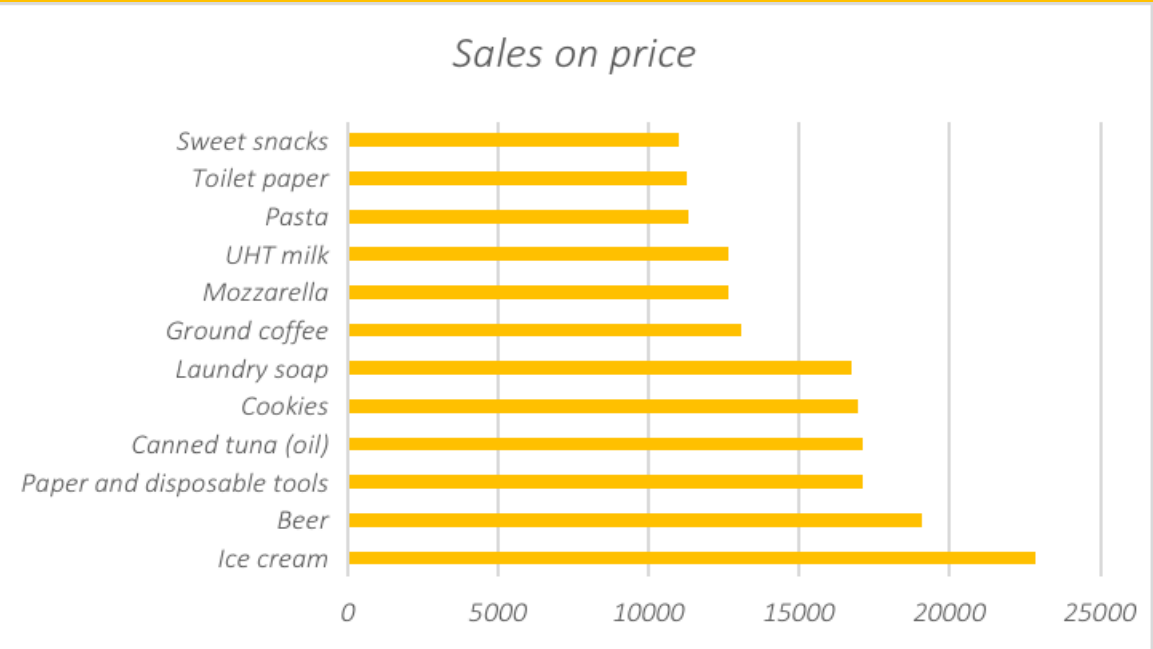


## WEEKLY TRANSACTION TREND

From the graph, we see a steep increase in transactions during the weekends with Friday, Saturday and Sunday accounting for major traffic. It is recommended to take advantage of this prime time to put high-profit and best-selling products in a prominent position for sale.

During the trough which is Wednesday, we can attract consumers with discounts to increase traffic.

# GENERAL ANALYSIS- PERFORMANCE OF PRODUCTS



## SALES REVENUE AND VOLUME

From the two graphs, we can see that 30% of the sales are from Beer, UHT milk, Pasta and Still water. Whereas, majority of sales include commonly used food items, drinking water and toilet paper. Supermarkets should pay more attention to these best-selling products for timely purchase and replenishment of goods to ensure a healthy supply chain.

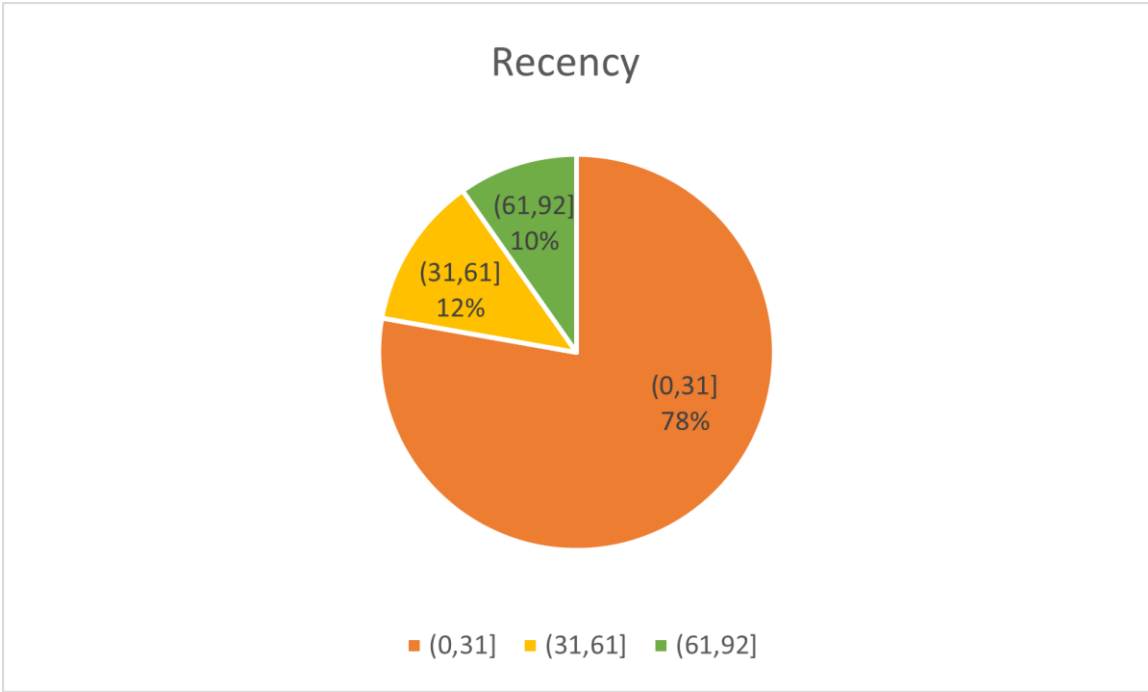
Then, we have products with low sales which needs attention as they will occupy more inventory, capital and product display space. We can get rid of these product categories to maximize the floor utilization.

We also see that of the 874 products in the catalog, only 376 were actually sold in 90 days. So, managers should focus whether the remaining 498 products were not stocked in the store or unsold. If it is unsold on the shelves, we need to make promotions to reduce the inventory and no longer stock these items.



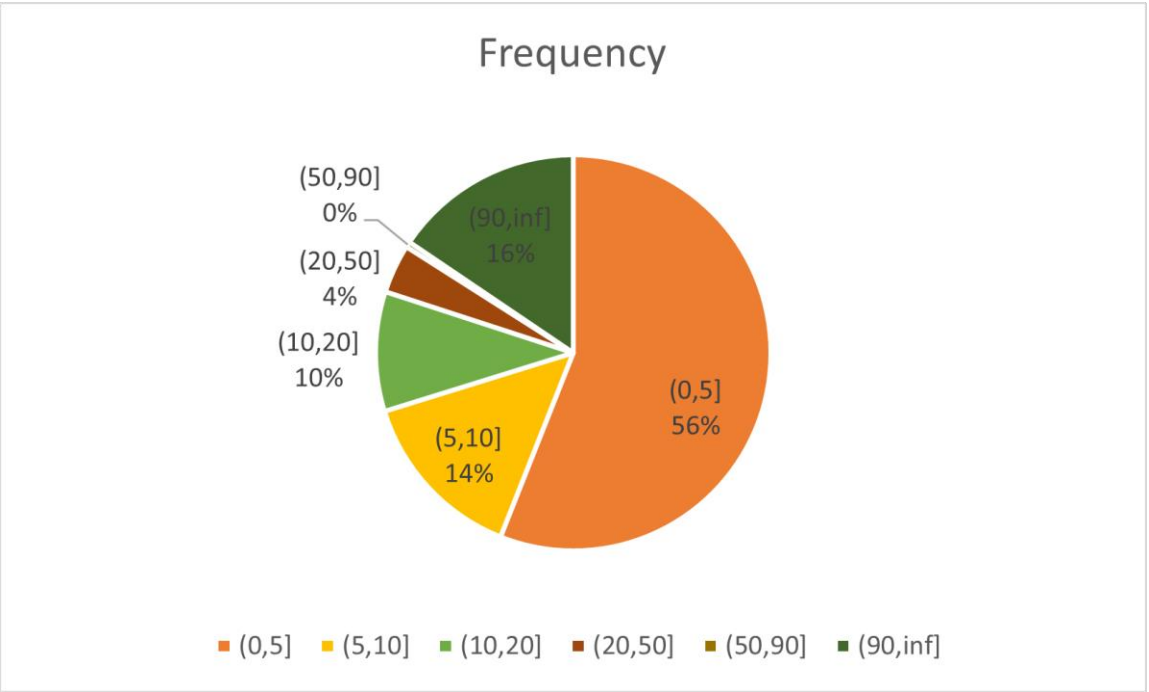
# GENERAL ANALYSIS- PERFORMANCE OF CUSTOMERS

The following refers to the transaction during the span of 3 months, borrowers of loyalty card are considered in each of the graph.



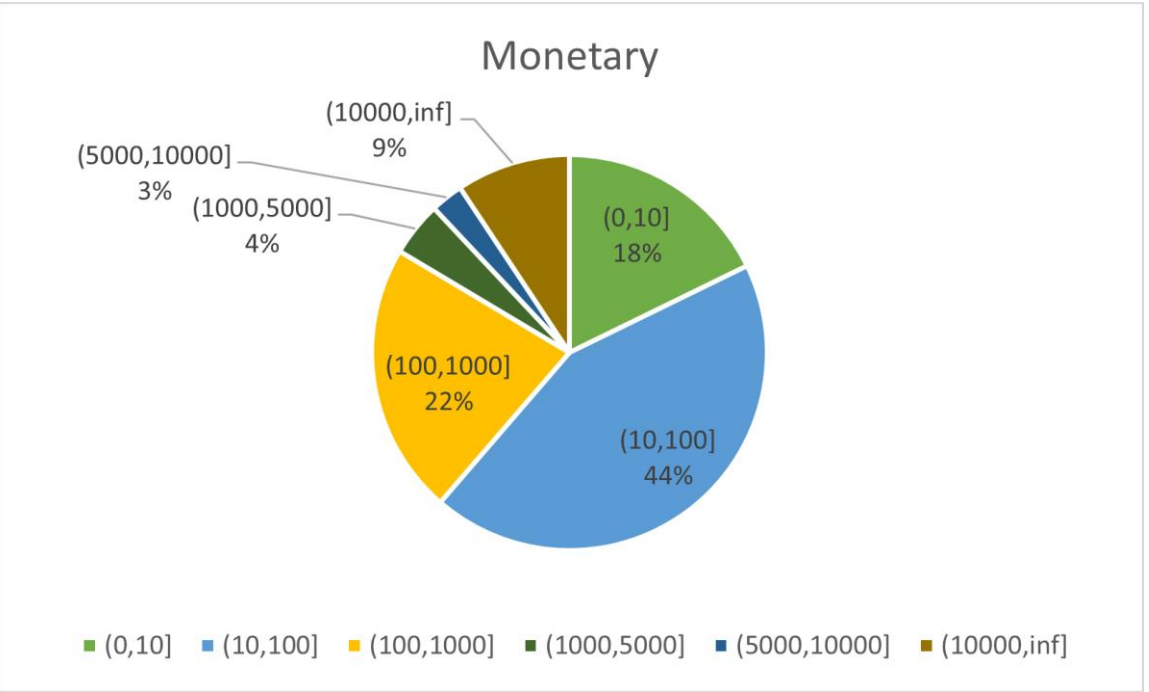
## RECENCY

Majority of customers around 78% shopped within the last 31 days. 12% within the last 31-61 days and 10% of the customers shopped within 61-92 days. This is good sign as a major portion are very recent customers and they still have the store awareness in the mind.



## FREQUENCY

Around 16% of the customers visit more than 90 times which we consider as borrowers and 14% of customers visiting between 5-10 times. A big portion of the customers around 56% have visited within 5 times hence more focus should be on converting these customers into more frequent ones.



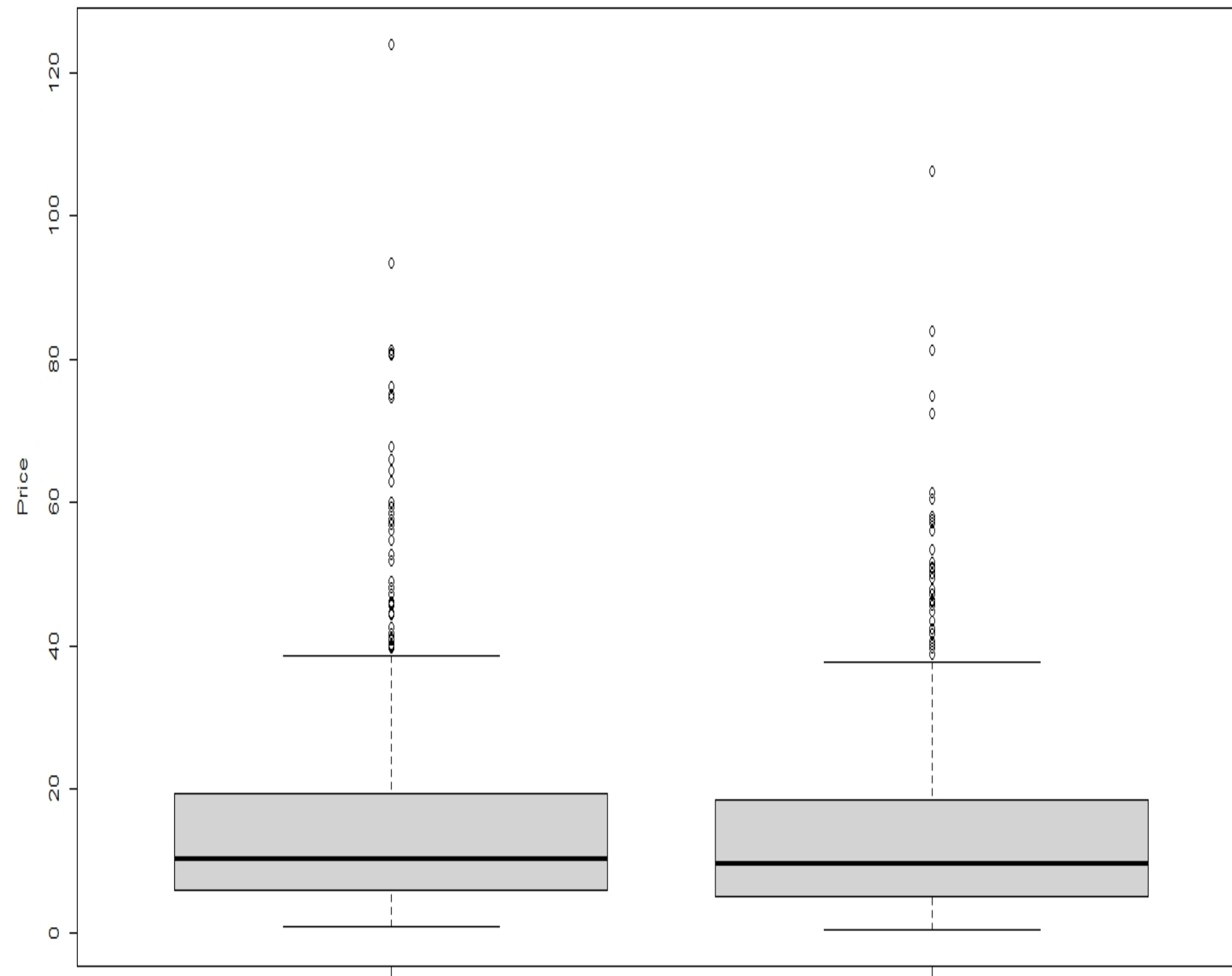
## MONETARY

Around 44% of the customers spent between 10-100 euros. 22% spent between 100-1000 euros, 18% within 10 euro. 9% of the customers spent more than 10000 euros which we consider to be borrowers. Majority of the customers spend less, and this reinforces our objective that a small number of customers contribute to most of the revenue and marketing focus should be on this segment.



# IDENTIFY THE REAL CUSTOMERS

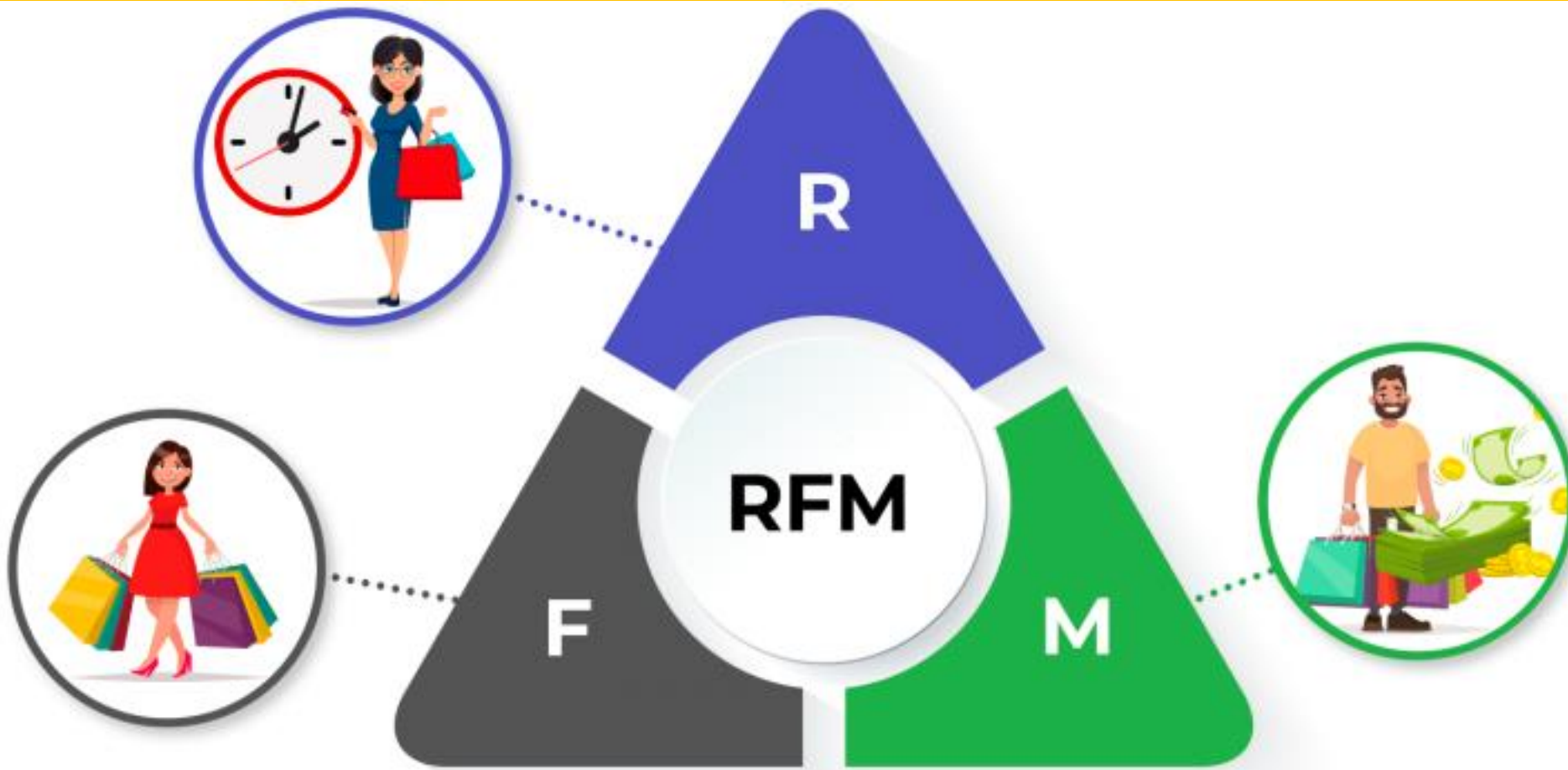
The average purchase per ticket for borrowers and for clients



## BORROWERS DETECTION

On the left, we see the boxplot of the price per ticket of customers vs. borrowers. Their purchases are similar whereas the number of transactions is different. This reinforces the fact that borrowers are also general customers but only borrow the loyalty card. Further analysis will be held only on real customers, but the validity of output could be more significant with a broader dataset or if it would be possible to distinguish customers and the borrowers.

# THE BIGGEST REVENUE COMES FROM MOST VALUABLE CUSTOMERS

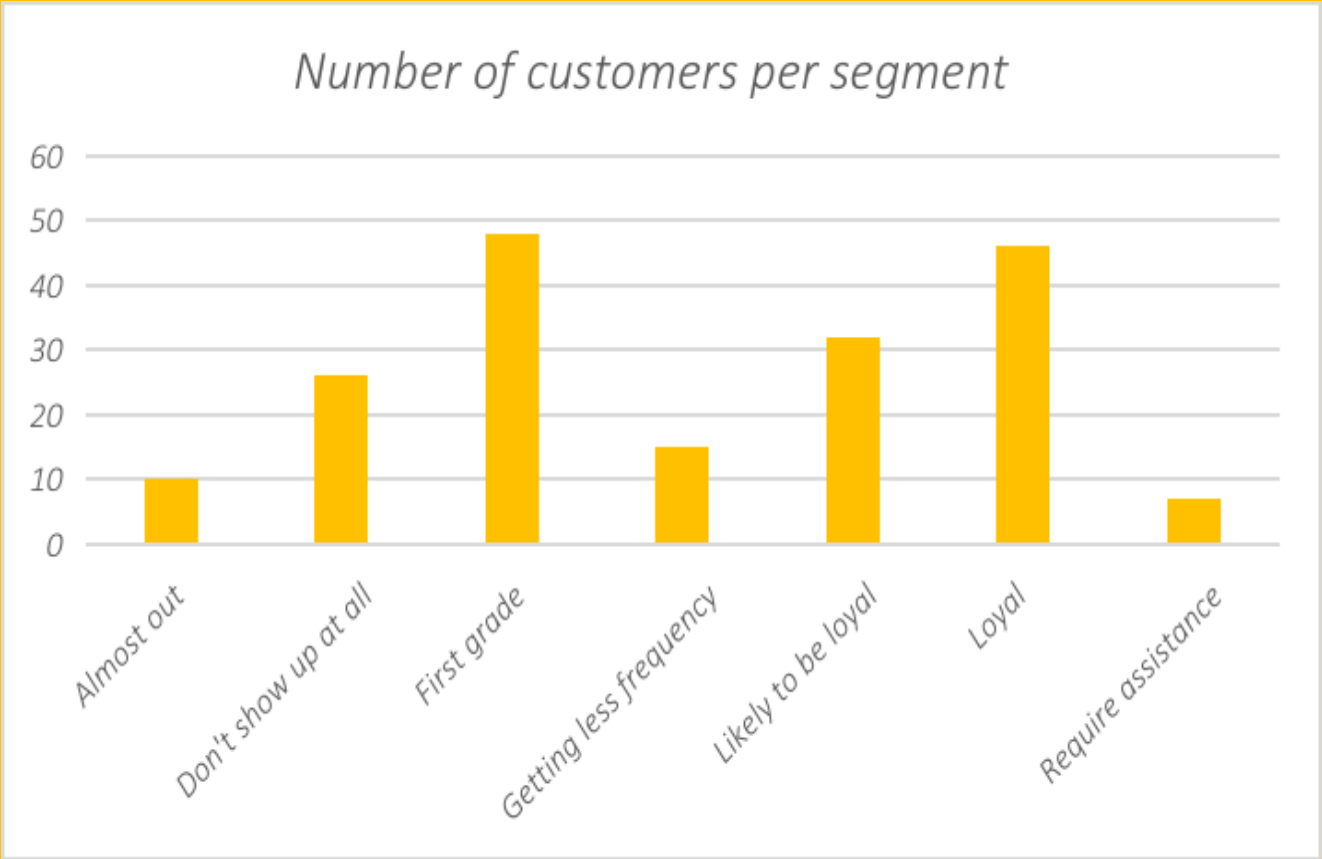


After analyzing the data from the descriptive analysis, we decided to do an RFM analysis, segmenting our customers and spending more on the ones that are likely to bring in revenue while ignoring others.

## Why RFM?

Not all customers are the same and each segment requires different marketing approach. Compared with the single- parameter evaluation model, the RFM model combines three different customer attributes to conduct a more comprehensive evaluation and take targeted marketing actions based on the segment.

# RFM ANALYSIS ON REAL CUSTOMERS (I/2)



	First grade	loyal	Likely to be loyal	New ones	Could be promising	Require assistance	Getting less frequency	Almost out	Can't lose them	Don't show up at all
R	4-5	2-5	3-5	4-5	3-4	2-3	2-3	1-2	1-1	1-2
F	4-5	3-5	1-3	1-1	1-1	2-3	1-2	2-5	4-5	1-2
M	4-5	3-5	1-3	1-1	1-1	2-3	1-2	2-2	4-5	1-2

- We eliminated the transactions done by borrowers in the dataset and performed the RFM analysis to have more accurate customer segmentation.
- We used a scale for Recency, Frequency and Monetary ranging from 1 to 5 and setting the limits for each segment as shown above.
- After performing it on the datasets, customers can be divided into 11 possible segments and there are 7 segments in our dataset, where most of the customers belong to the Loyal and First grade segments.



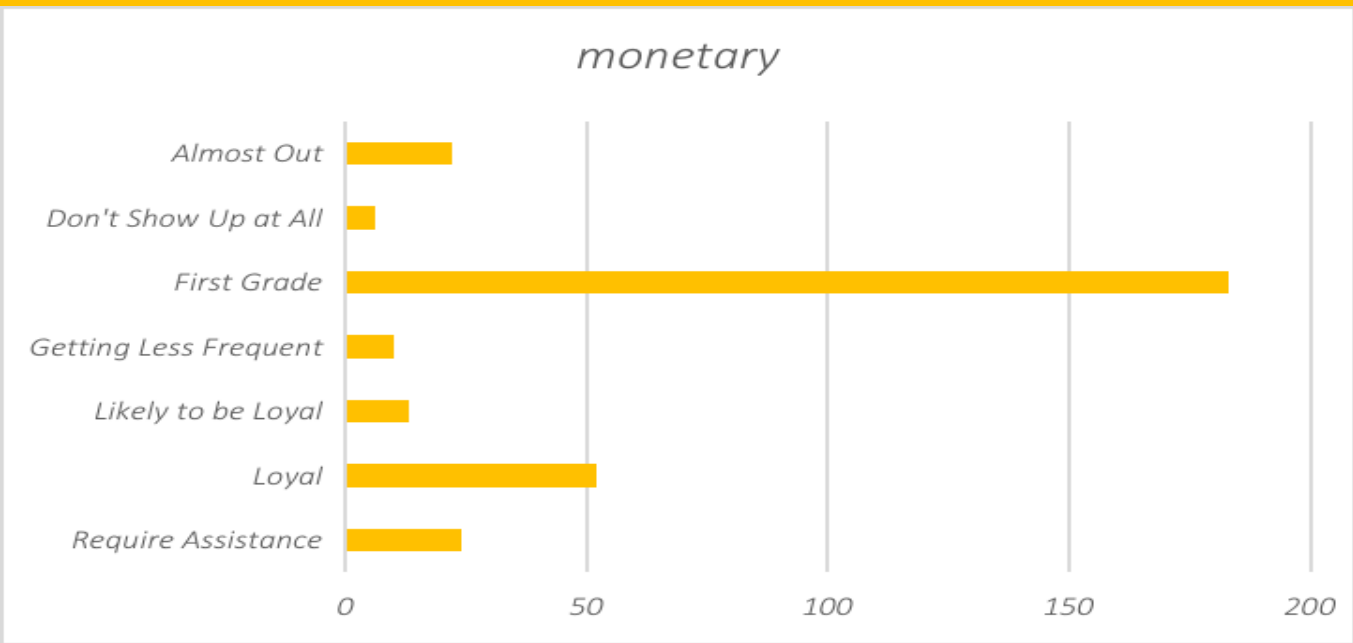
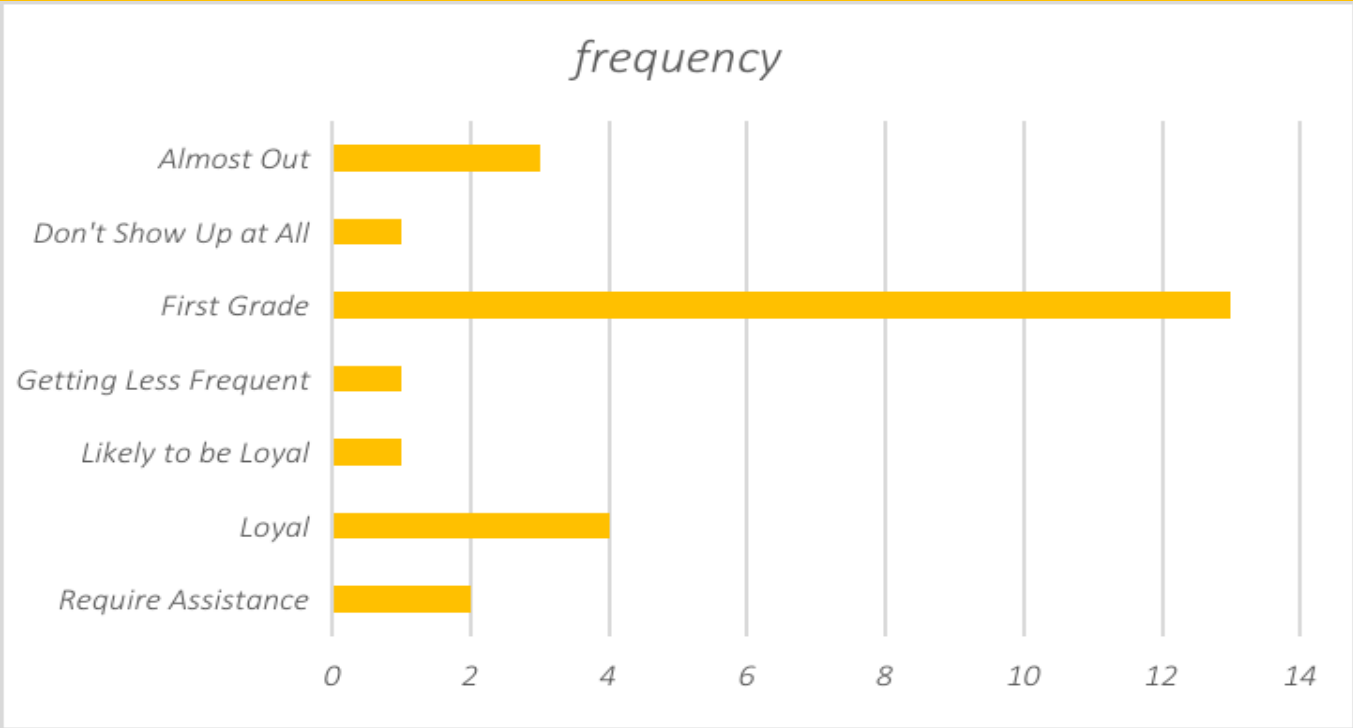
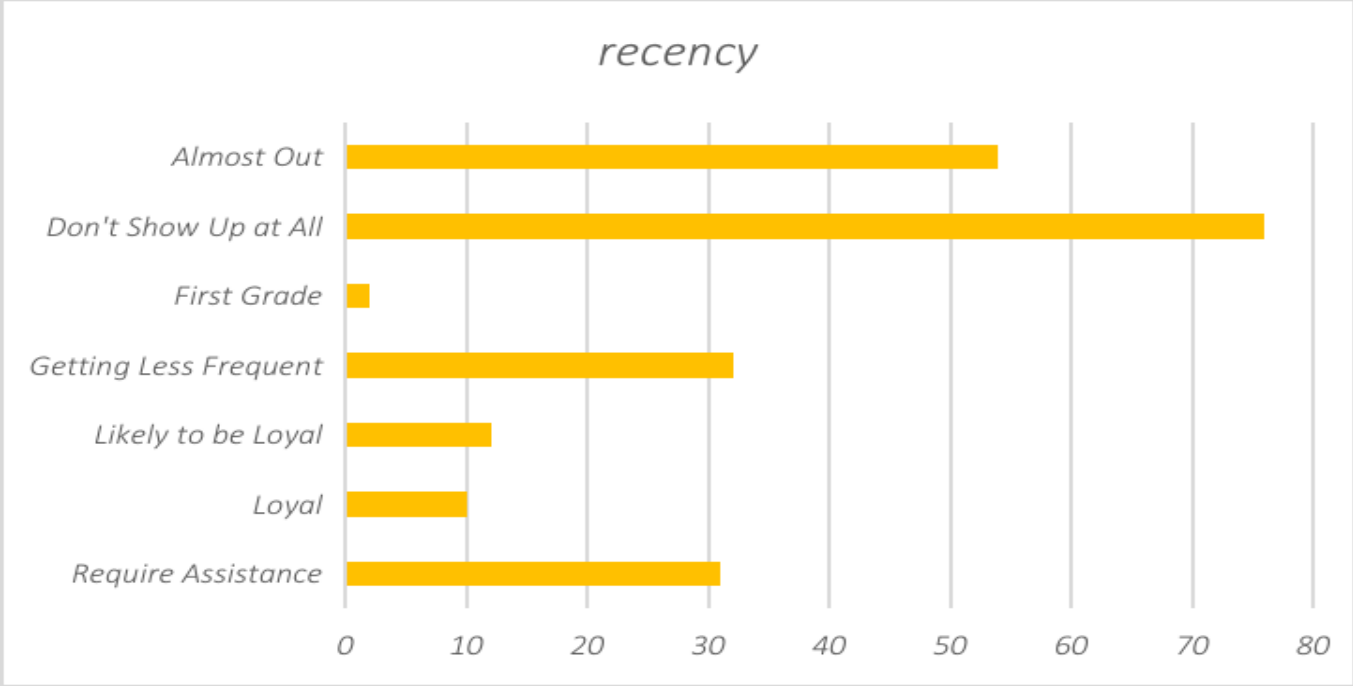
# RFM ANALYSIS ON REAL CUSTOMERS (2/2)

Furthermore, as the graphs show, the performance of different segments could be divided into following 3 levels.

level	Segments
Best customers	The First Grade,loyal
Medium customers	Likely To Be Loyal, Require Assistance
Worst customers	Almost Out, Don't Show Up At All, Getting Less Frequency

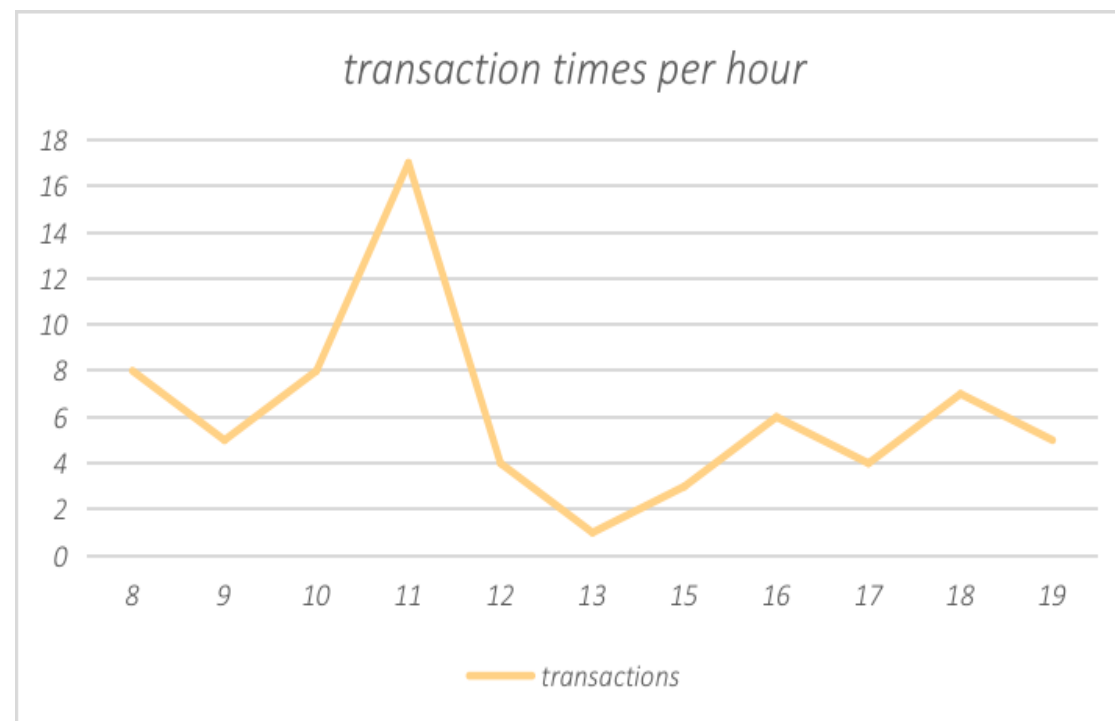
The monetary value and the frequency of the best customers are very high with respect to the other segments. They represent about 48% of the customers that account for 91% of the revenue.

Therefore, we believe that the focus should be more on these segments. Hence, we decided to do our further analysis on them.



# BEST CUSTOMER'S BEHAVIOR (1/2)

Further analysis was done on the customer to know when the customers visit, how much they spend, and which store they prefer.

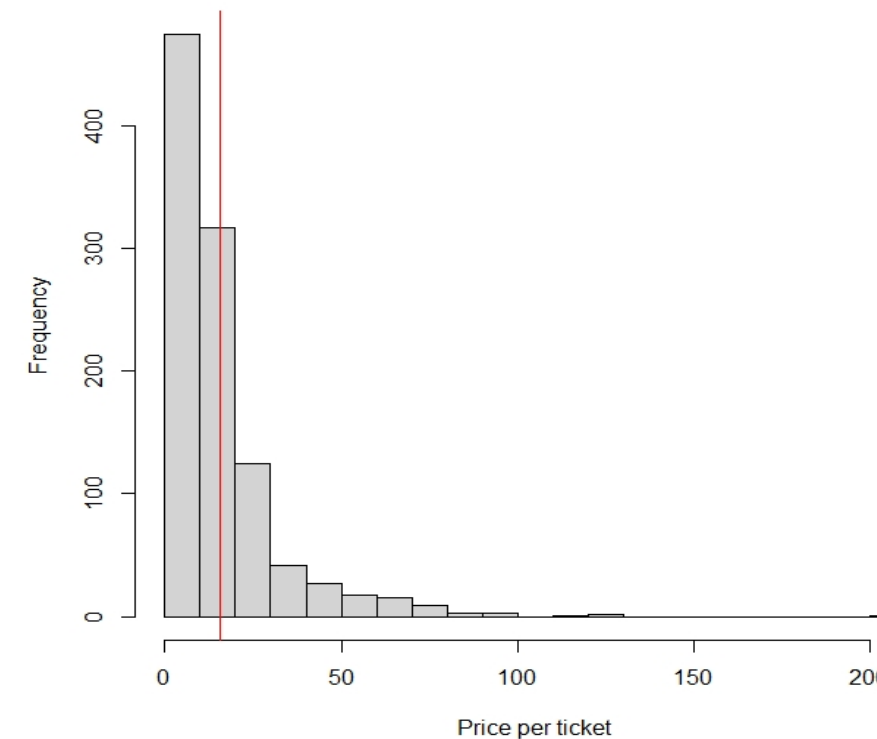


## WHEN?

---

The peak of the best customers is at 11:00 and the after-work peak is no longer noticeable maybe because a portion of our clients are old people who are more accustomed to going out in the morning.

monetary purchase per ticker for loyal customers (red line is the mean)

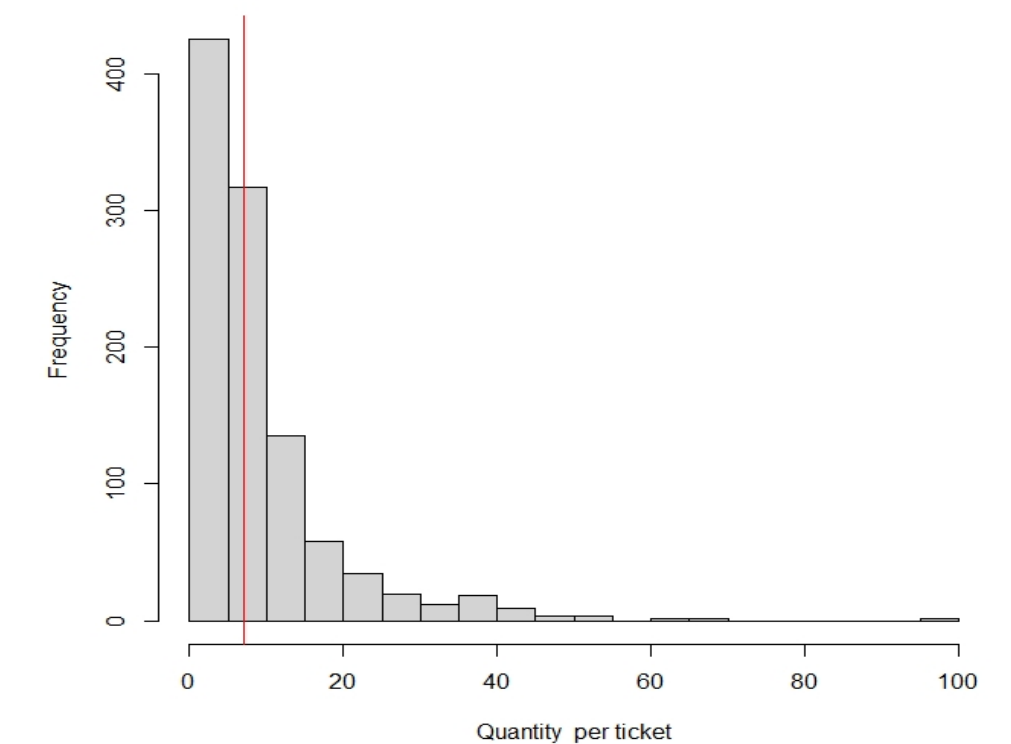


## HOW MUCH?

---

The mean value of best customer's spending is 15.96 euros, with the median at 10 euros. On average, customers buy about 10 items per visit, which shows that we have opportunity to try and sell more of what we want to sell to them, and at the same time, increase the overall consumption as much as possible

Quantity of items purchase per ticker, loyal customers (red line is the mean)



# BEST CUSTOMER'S BEHAVIOR (2/2)

## WHERE?

The histogram above shows the distribution of the best customers in each store. Even though store 624 is one of the smallest, has the highest number of loyal and tier one customers whereas the largest store 576 doesn't perform well considering its size.

The histogram below shows customers who visit more than one store. Clearly, these customers prefer to go to smaller stores for one of the two following possibilities:

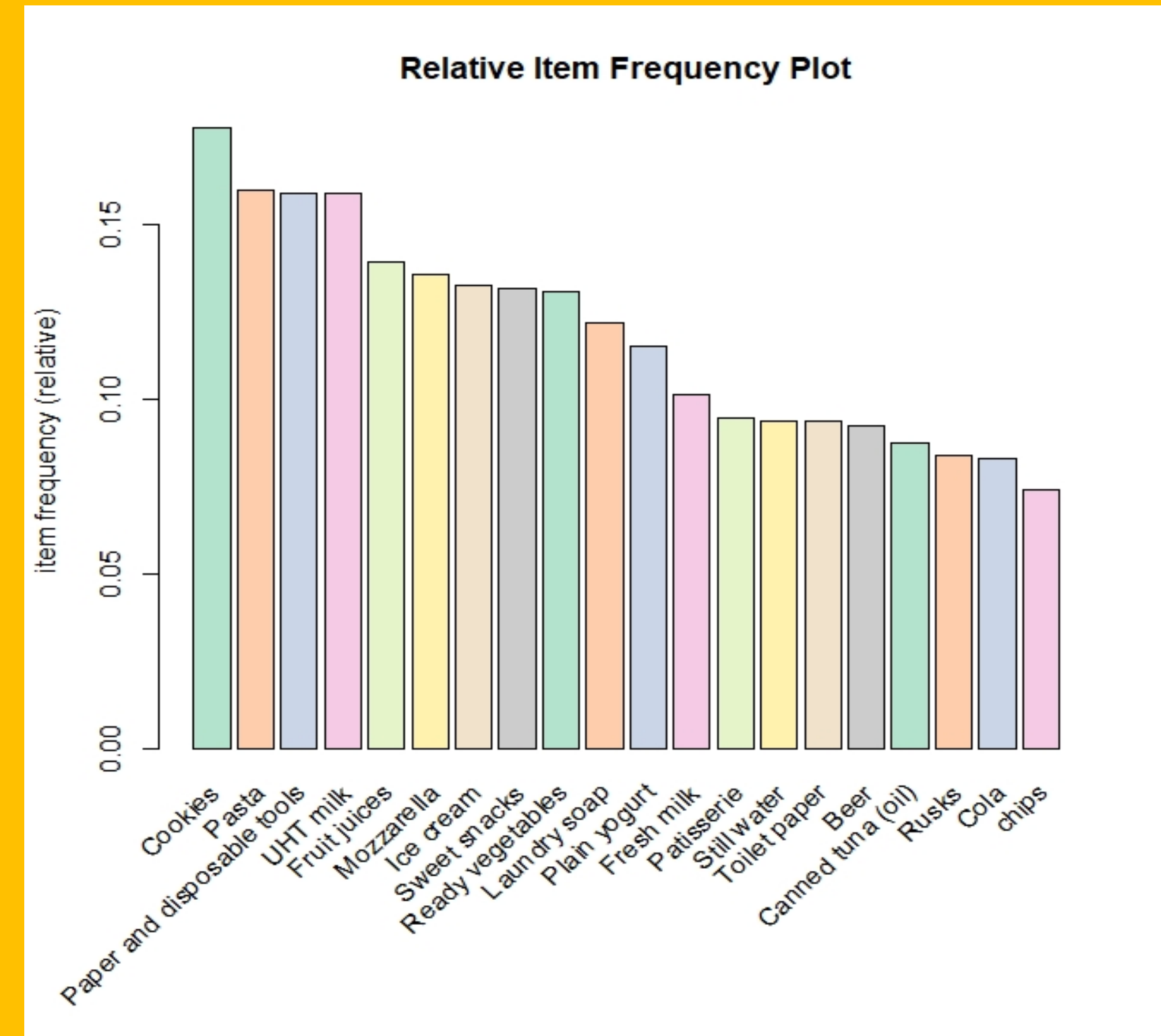
- Customers who visit smaller stores are also the more loyal to the brand because they don't care about the size but are in search of the value offered by the company. If so, it is important to understand this group as they account for company's success.
- Customers prefer to go to smaller stores as they are situated close to each other while medium and big sized stores are distant and so difficult to switch.





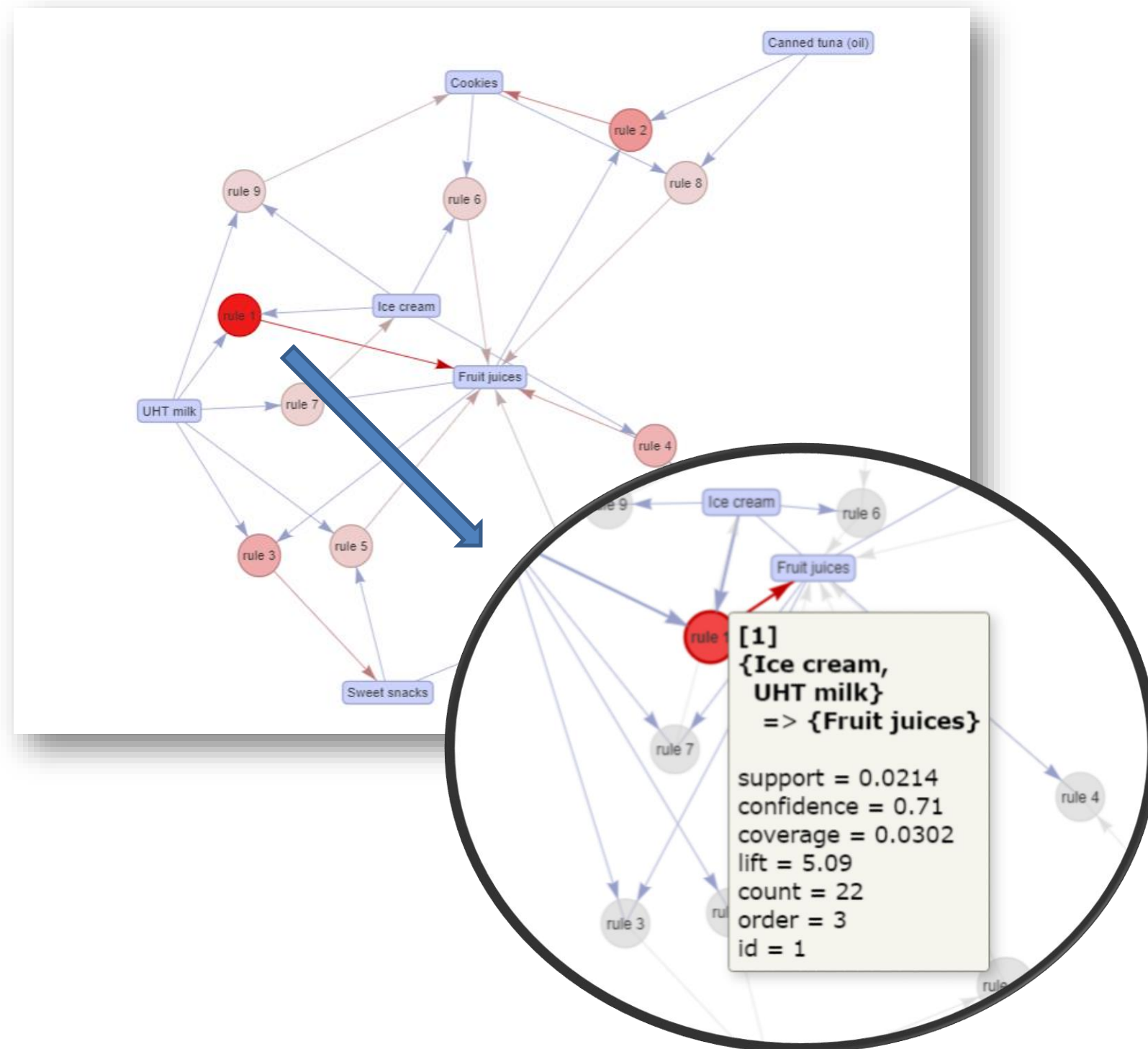
# BEST CUSTOMER'S PURCHASE

Relative item frequency plot was performed on the loyal and first grade customers and we can see that most of the items purchased by best customers are junk food and household items. The only fresh items that are bought being vegetables and mozzarella. So, it is clear that they buy a lot of ready, easy to eat and long shelf-life food.



# MARKET BASKET ANALYSIS

## Product package



- MBA was performed on the best customers to find their shopping pattern and then offer promotions to each segment based on the product subcategory that were frequently bought together.
- For MBA , Apriori algorithm was used with the following parameters: Minimum Support= 0.02, Minimum Confidence= 0.5 , and Max Length of the rules= 4.
- To limit the number of the rules, the subset of larger rules have been removed
- The rules obtained from Apriori algorithm were ordered by their lift.
- As shown in the picture on the left, according to Rule 1, customers who buy {Ice cream, UHT milk} will also prefer to buy {Fruit juices}.
- In order for cross-selling and up-selling to the target customers (the best customers), we could carry on:
  1. Product package sales
  2. Specific product promotions
  3. Discount redemption, etc

# OPERATIONAL SUGGESTIONS

We focus our marketing action to the different customer segments based on RFM analysis and Advertise them with appropriate promotions based on their market basket analysis. Email marketing, Promotions, ads inside and outside the stores are made to communicate with each segment.

## CROSS-SELLING & UP-SELLING- for best customers

- First grade customers have attitudinal loyalty but make purchases with low monetary value. Instead of addressing them through discounts, we recommend the newly launched products from categories that were bought together frequently. We offer them minimal discounts just to keep them engaged.
- Loyal customers respond well to promotions hence we can send regular promotions via Emails to them for Upselling product based on the product category they purchased.

## LIMITED-TIME PROMOTION- for medium customers

- Limited-time promotion is a tight promotion, its purpose is to stimulate consumers to have the mentality of "if you don't buy it, you will lose money"
- Limited time promotion is designed to activate the consumption of medium customers who are potential to be loyal. Medium customers are chosen for this promotions based on the fact that best customers still remain active without promotions.

## RECALL PROMOTION- for worst customers

- Recall promotion is a loose promotion which is to avoid churning. Churned customers are no more interested in buying and hence we can send Emails about Refer and Earn vouchers. This not only brings new customers to store but also invites the churned customers to shop again



# MARKETING ACTION BASED ON THE AGE CATEGORY

- **SENIORS**

- We see a peak in customers in the morning and working people cannot visit the store at that time. Although we do not have a dataset indicating age, based on our analysis, we assume that a portion of our customers are old people who prefer to go to small stores, as well as eat easy-to-chew food.
- Since, a majority of seniors are not that competent with technology, it is difficult to get in touch with them through online channels. Therefore, we must focus our promotions through offline methods.

- **STRATEGIES**

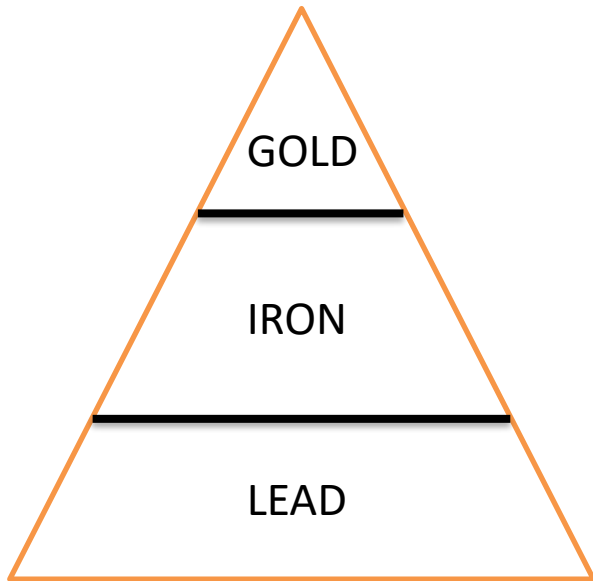
- Cash counter employees give vouchers to the customers
- Flyers kept near the counter indicating the upcoming promotions
- Promotions outside the store and in public places.





# LOYALTY PROGRAM IMPROVEMENT

Loyalty programs are designed to build long-term relationships with customers, foster their spending habits, and motivate them to shop more.



## RULES

- We divided the customers who have a loyal card into 3 grades basing on their monthly accumulated consumption, which will be allotted based on the last month’s performance.
- In general, you earn one point for every euro you spend, and these points can be accumulated to redeem for prizes.
- Ability to multiply points this week, based on the consumption performance of the last week. If you shopped at the store more than or equal to 2 days last week, you could get multiply points this week.

## INSIGHTS

- This system works only when customers got the loyalty card, so for the new customer, we need to lead them to join the loyal program as first by introducing the benefits.
- The higher the grade you are in, the higher rewarding points you can get.
- The system is dynamic to keep customers alive. Once the customer stops spending well, the multiply points end.
- In order to increase the appeal of the loyal program, we should set appropriate threshold to ensure it can be achieved with efforts and the rewarding prizes need to keep customers satisfied like Movie ticket, Spa, Medicine voucher, etc.

Grade	Monthly Accumulated consumption	Weekly consumption times	Multiply points
Gold	Above 300 €	2	1:4
Iron	100-300 €		1:3
Lead	Below 100 €		1:2

# APPENDIX

## 1. Open days and hours of each store

store	hour
516	7,8,9,10,11,12,13,14,15,16,17,18,19,20
519	7,8,9,10,11,12,13,15,16,17,18,19
552	7,8,9,10,11,12,13,15,16,17,18,19,20
561	8,9,10,11,12,15,16,17,18,19
576	7,8,9,10,11,12,13,14,15,16,17,18,19
588	8,9,10,11,12,13,16,17,18,19,20
606	8,9,10,11,12,13,15,16,17,18,19
624	8,9,10,11,12,15,16,17,18,19

store	week
516	6, 1, 2, 3, 4, 5
519	6, 1, 2, 3, 4, 5
552	6, 1, 2, 3, 4, 5
561	6, 1, 2, 3, 4, 5
588	6, 1, 2, 3, 4, 5
606	6, 1, 2, 3, 4, 5
624	6, 1, 2, 3, 4, 5
576	6, 7, 1, 2, 3, 4, 5



# APPENDIX

## 2. Products which are not sold but in the category

Product code	Description
10110	Healthy biscuits
10111	Wholegrain biscuits
10112	Biscuits with eggs
10204	Small chocolate eggs
12001	Panettone
12002	Pandoro
12003	Other Christmas sweets
12004	Easter eggs
12005	Easter colomba
12006	Other easter sweets
12007	Gift boxes
12008	Torrone

This is part of an unsold product

# APPENDIX

## 3. Top 10 rules

We denote association rules with  $X \rightarrow Y$   
Support =  $\frac{\text{\#orders contain } (X+Y)}{\text{\#total orders}}$   
Confidence =  $\frac{\text{\#orders contain } (X+Y)}{\text{\#orders contain } (X)}$   
Lift =  $\frac{\text{\#orders contain } (X+Y) * \text{\#total orders}}{[\text{\#orders contain } (X) * \text{\#orders contain } (Y)]}$

Lift > 1 means the more buy X, the more buy Y  
Lift = 1 means no connection between X and Y  
Lift < 1 means the more buy x, the less buy Y

Rules	Customers who buy	Also prefer to buy	Support	Confidence	Coverage	Lift	Count
1	Ice cream, UHT milk	Fruit juices	0.02144250	0.7096774	0.03021442	5.091811	22
2	Canned tuna (oil), Fruit juices	Cookies	0.02144250	0.7857143	0.02729045	4.429356	22
3	Fruit juices, UHT milk	Sweet snacks	0.02339181	0.5581395	0.04191033	4.241860	24
4	Ice cream, Pasta	Fruit juices	0.02144250	0.5789474	0.03703704	4.153846	22
5	Sweet snacks, UHT milk	Fruit juices	0.02339181	0.5454545	0.04288499	3.913541	24
6	Cookies, Ice cream	Fruit juices	0.02046784	0.5384615	0.03801170	3.863367	21
7	Fruit juices, UHT milk	Ice cream	0.02144250	0.5116279	0.04191033	3.859781	22
8	Canned tuna (oil), Cookies	Fruit juices	0.02144250	0.5365854	0.03996101	3.849906	22
9	Ice cream, UHT milk	Cookies	0.02046784	0.6774194	0.03021442	3.818859	21
10	Pasta, Sweet snacks	Fruit juices	0.02046784	0.5250000	0.03898635	3.766783	21