Marketing Analytics

Paolo Caggiano 902877

What will we talk about?

1. Customer Focus (RFM + CHURN)
2. Product Focus(MBA)
3. Feedback Focus(Sentiment Analysis)

Business questions 1/2

- How is divided the customer base?
- What is the value of the various segments?

RFM MODEL

- Who are the customers with the highest probability to churn?
- Which are the possible churners characteristics?

CHURN MODEL

Business questions 2/2

- Which are the most selled products?
- Is there association between products purchased together?

MARKET BASKET ANALYSIS

- How is the feedback of the users?
- How to act with the different customers?

SENTIMENT ANALYSIS

The Data

tbl_customer

customer_id
address_id
birthdate
gender
job_type
email_provider
flag_phone_provided
flag_privacy

tbl_addresses

address_id
postal_code
district
region

tbl_customer_accounts

customer_id
account_id
favorite_store
Loyalty_type
loyalttly_status
activation_date

tbl_orders

order_id
customer_id
store_id
product_id
direction
gross_price
price_reduction
purchase_datetime

tbl_product

product_id
product_class

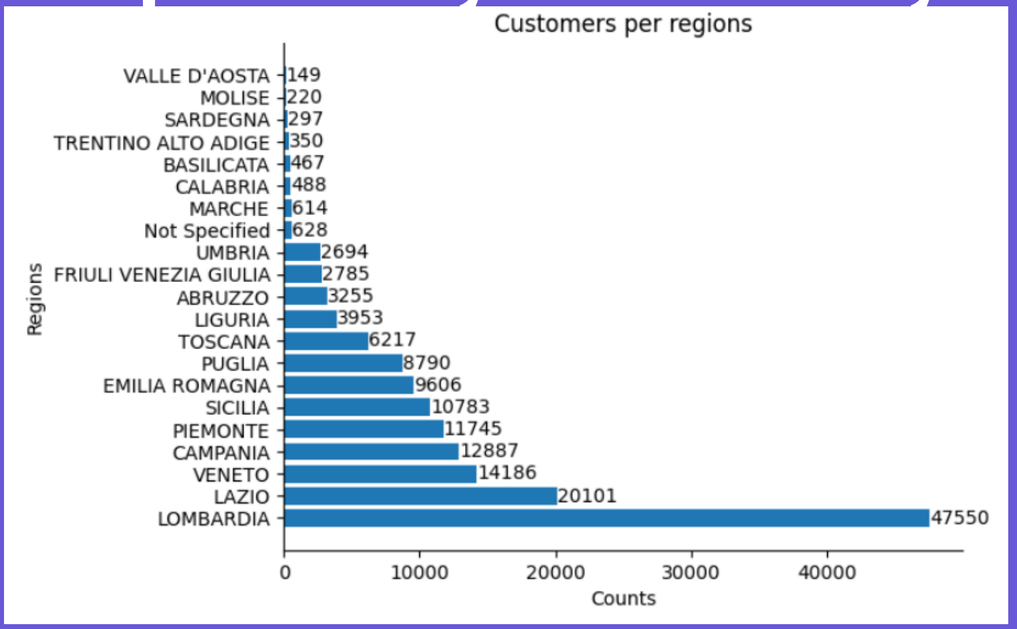
tbl_customer_review

review_id
customer_id
review_text

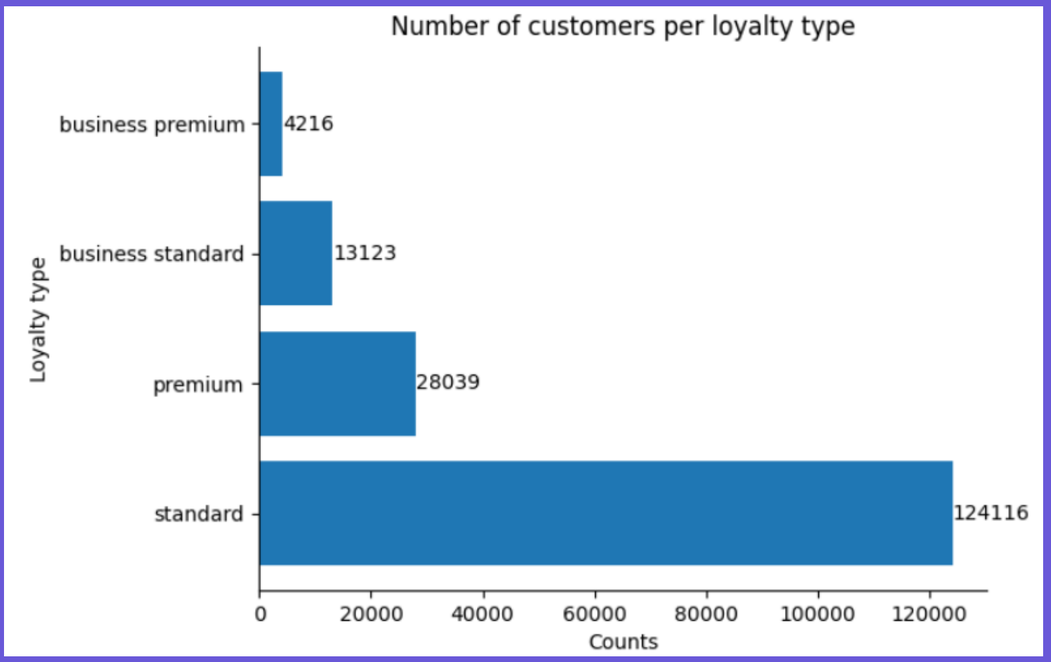
tbl_labelled_reviews

labelled_review_id
review_text
sentiment_label

Exploratory Data Analysis



- Most of the customers are from Lombardia
- Valle D'Aosta and Molise are the less represented regions
- For 628 costumers there is no information about their region



- The most common loyalty type is standard
- Business premium is the less usual loyalty type

The 10 most frequent customers year of birth

Year of birth	Number of customers
1980	9297
1981	9190
1983	9064
1982	9033
1979	8987
1984	8936
1978	8886
1977	8845
1976	8625
1975	6113

- The majority of the customers are in the middle age
- They can be used both the "old" metodologies like the use of telemarketing and "new" techniques
- Digital marketing: the customer base has growed with the digital revolution
- Use of social media in order to attract and acquire new clients

RFM MODEL

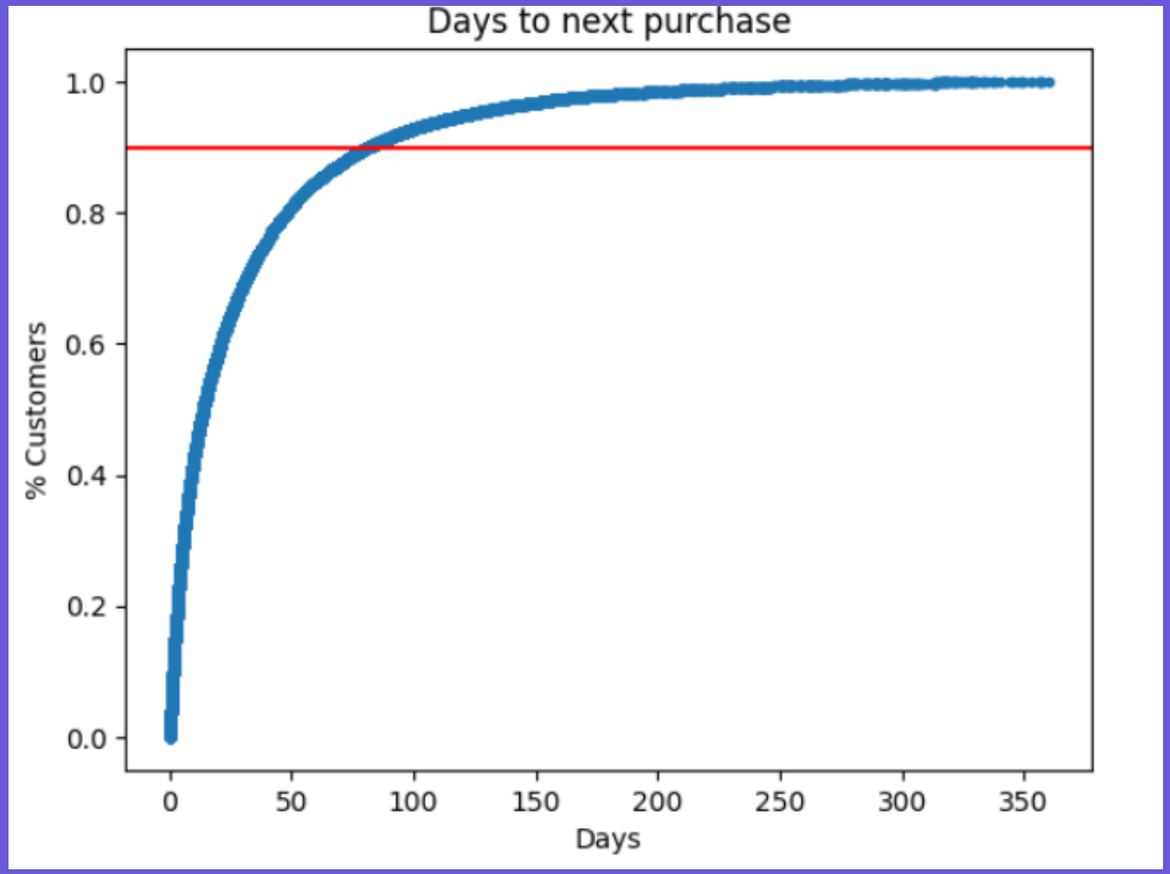
Recency and **Frequency** indicators allow the definition of the classes:

- One-Timer
- Leaving
- Engaged
- Leaving Top
- Top

Monetary metric permits another division in the classes:

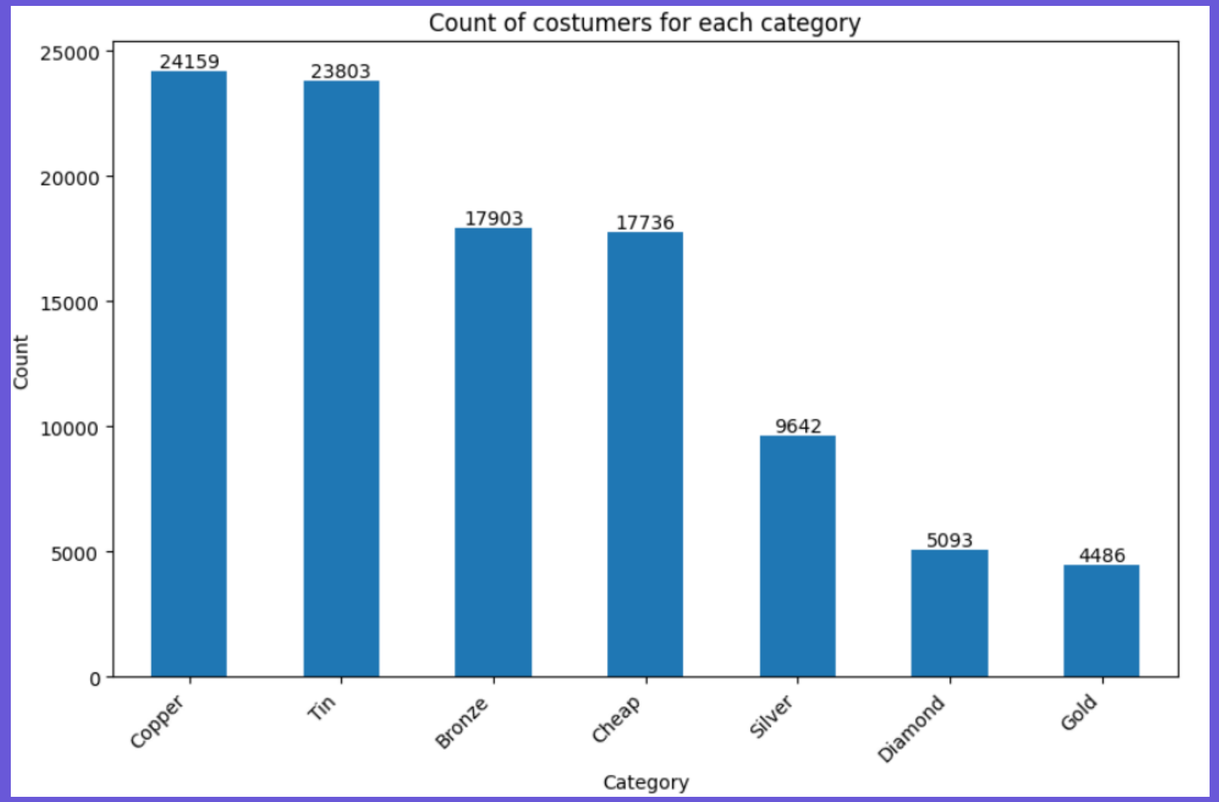
- Copper
- Tin
- Bronze
- Cheap
- Silver
- Gold
- Diamond

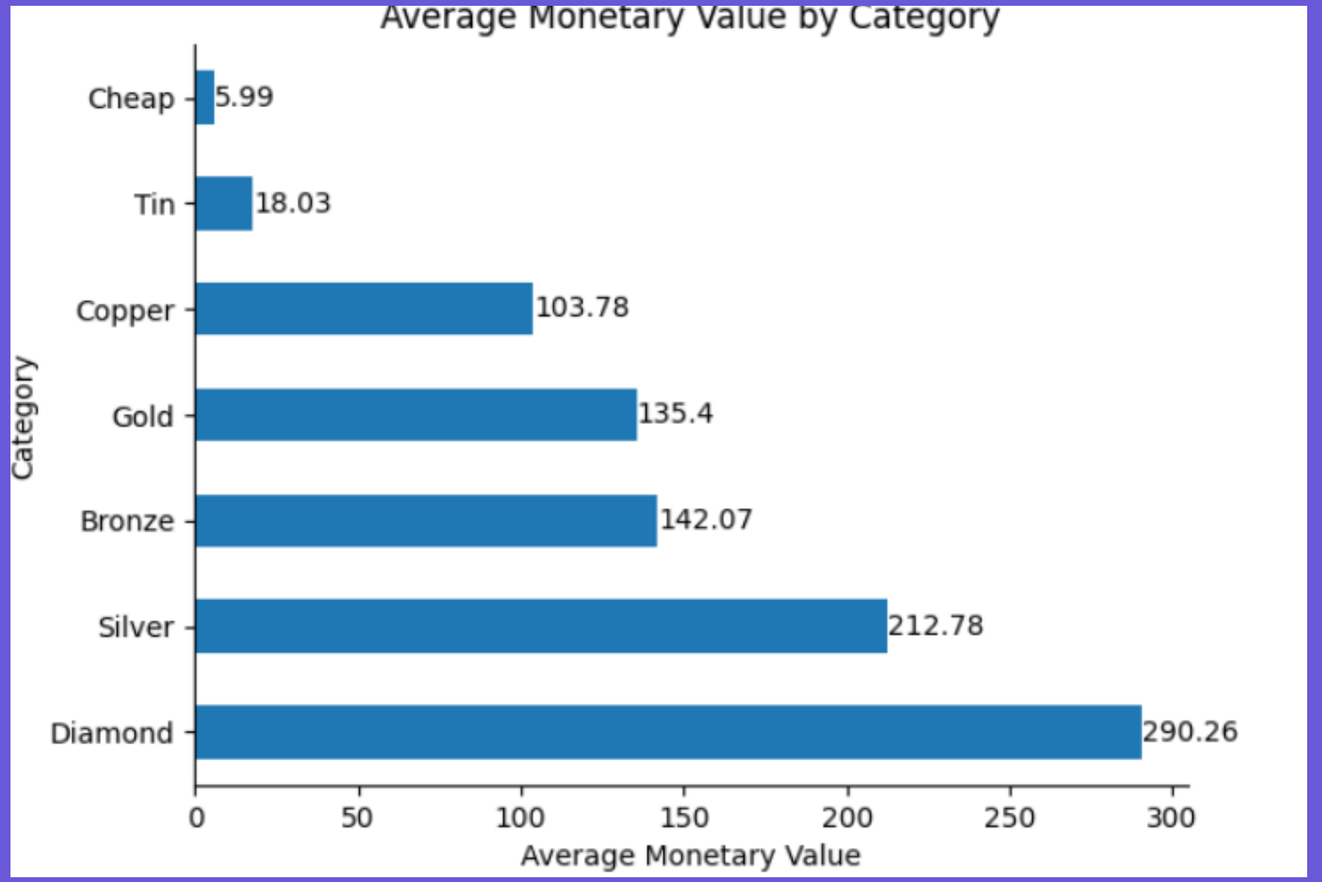
Repurchase Curve



- Approximately 90% of customers make a repeated purchase after 82 days.
- Since the first date in the orders is 01-05-2022, I make a distinction between **active and inactive customers**. Those clients that did not purchase in the range : (01-05-2022, 30-04-2023) have been considered as non-active.

RFM MODEL





As we can see most of the customers belong to the class Copper and Tin, while the classes: **Diamond** and **Gold** represent the 13% of the total.
Idea: invite them to join loyalty programs or subscribe to newsletters for future promotions.

Diamond customers have the highest average monetary value. Cheap customers have the lowest monetary values , probably it is not profitable to invest on them.
Idea: provide exclusive access to new product launches or pre-sales for loyalty program members

CHURN MODEL

Churn refers to the behavior of the customer intent on terminating his or her relationship with the company.

I use the repurchase curve to determine after how many days from the last purchase a customer can be define churner.

In fact, from the repurchase curve I discovered that above **the 90%** of the customer repurchase after **82 days.**

I choose as **reference date** : 07-02-2023. The customers that did not purchase in the 82 days before this date are classified as churners.

The classes are not balanced

CHURN(1)	**65 %**
NO-CHURN(0)	**35%**

Oversampling procedure →

Rebalanced dataset

CHURN(1)	**50%**
NO-CHURN(0)	**50%**

PERFORMANCE MEASURES

It is important to consider different metrics in order to have a better overview on models' performance.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

It refers to the fraction of instances correctly classified

$$Precision = \frac{T_p}{T_p + F_p}$$

it represents the fraction of instances which are classified as positive and that result to be effectively positive

$$Recall = \frac{T_p}{T_p + T_n}$$

it indicates the fraction of instances of positive class that are correctly identified

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

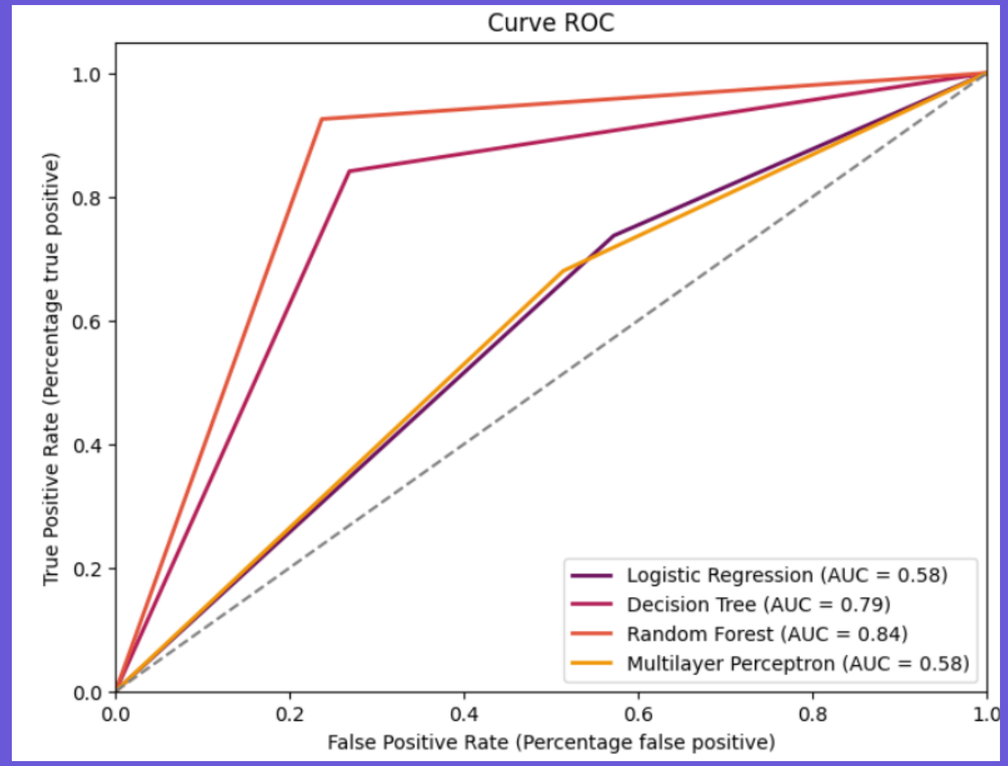
it is an harmonic mean between Precision and Recall

CHURN MODELS PERFORMANCES

ALGORITHM	ACCURACY	PRECISION	RECALL	F_SCORE
Random Forest	0.84	0.79	0.92	0.85
Decision Tree	0.78	0.75	0.84	0.79
Logistic Regression	0.58	0.55	0.74	0.63
Multilayer Perceptron	0.58	0.56	0.67	0.61

RELEVANT VARIABLES USED:

- total_expenditure
- favorite_store
- number of articles
- flag_privacy



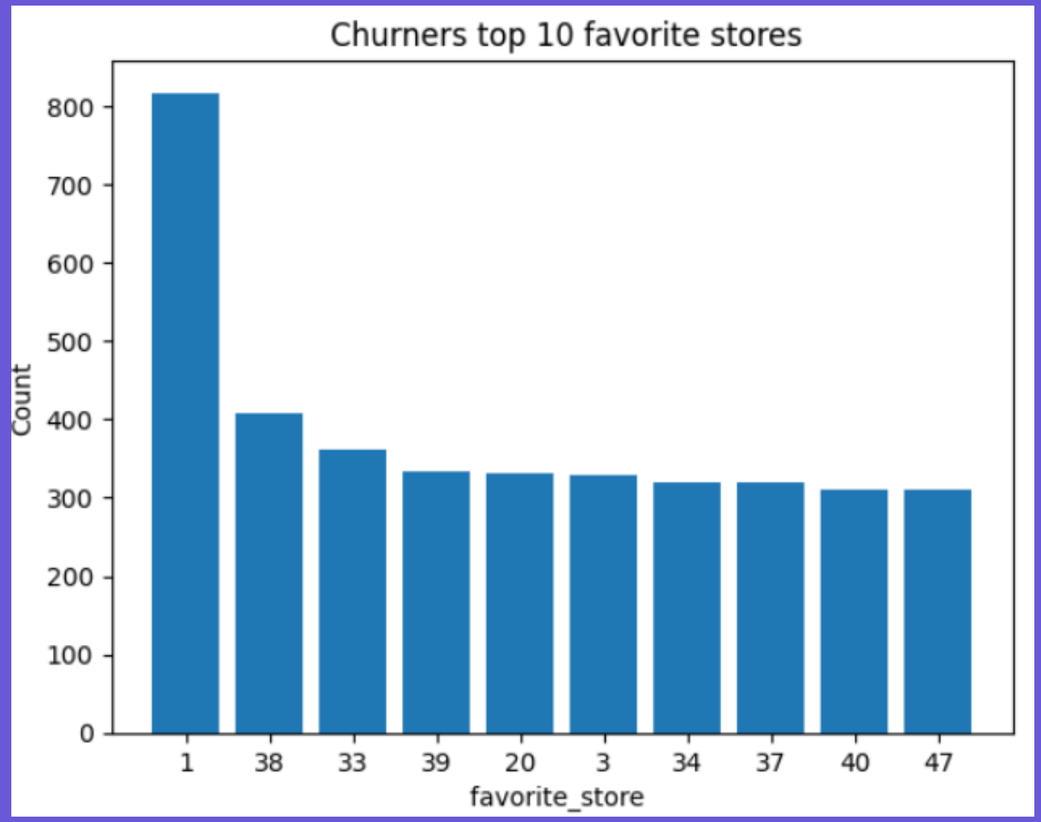
The Random Forest and Decision Tree models have the best performances.

Random Forest identifies the 92% of the churners

CHURNERS BEHAVIOURS

CHURN	AVERAGE EXPENDITURE
YES(1)	184.5
NO(0)	239.6

CHURN	AVERAGE NUMBER OF PURCHASED ARTICLES
YES(1)	7.94
NO(0)	13.3



- The store 1 is the most preferred by churners customers. It can be used to understand churners behaviours
- Tailor special offers or discounts specifically for churners based on their preferred store

Product analysis

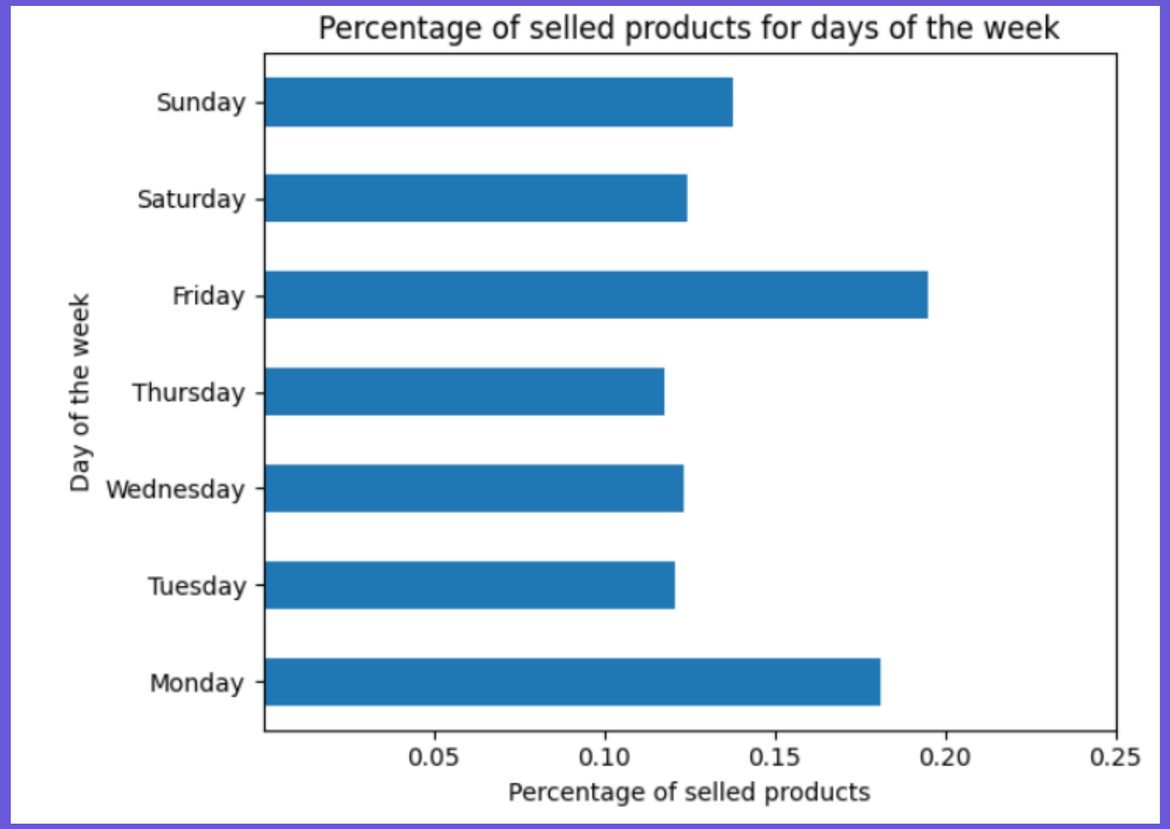
ORDERS
DIRECTION:
PURCHASES(97%)
REFUNDS(3%)

MOST PURCHASED
PRODUCT:
- 256686 (4%)

MOST PURCHASED
CATEGORIES:
- 3 (28%)
- 10 (14%)
- 11 (12%)

Use of these categories to promote new or less selled products

Products date -time analysis



Selled products time slot

Range	Proportion
(4-8)	0.01
(8-12)	0.24
(12-16)	0.31
(16-20)	0.39
(20-24)	0.05

The majority of the customer purchases' is made on Friday.
The most frequent time slot is 16-20.

MARKETING ACTION

DISCOUNTS ON THE PRODUCTS PURCHASED AT THE BEGINNING OF THE WEEK

PRODUCT ANALYSIS-MBA

The Market Basket Analysis model identifies **relationship** among products

- I compute frequent itemsets with the Apriori Algorithm
- Generation of association rules from the frequent itemsets(lift-metric)
- Other measures: antecedent , consequent , support, confidence, lift, leverage, and conviction evaluation
- I fix 0.01 as the support threshold
- I remove redundant rules

MBA MEASURES

$$\text{supp}(X) = P(X) = \frac{\#\text{ purchases with products } x_1, x_2, \ldots}{\#\text{ all purchases}}$$

$$\text{conf}(X \Rightarrow Y) = P(Y|X) = \frac{P(X \cap Y)}{P(X)}$$

$$\text{lift}(X \Rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{P(X \cap Y)}{P(X) * P(Y)}$$

What they indicate?

How common is the purchase of the product among all the products

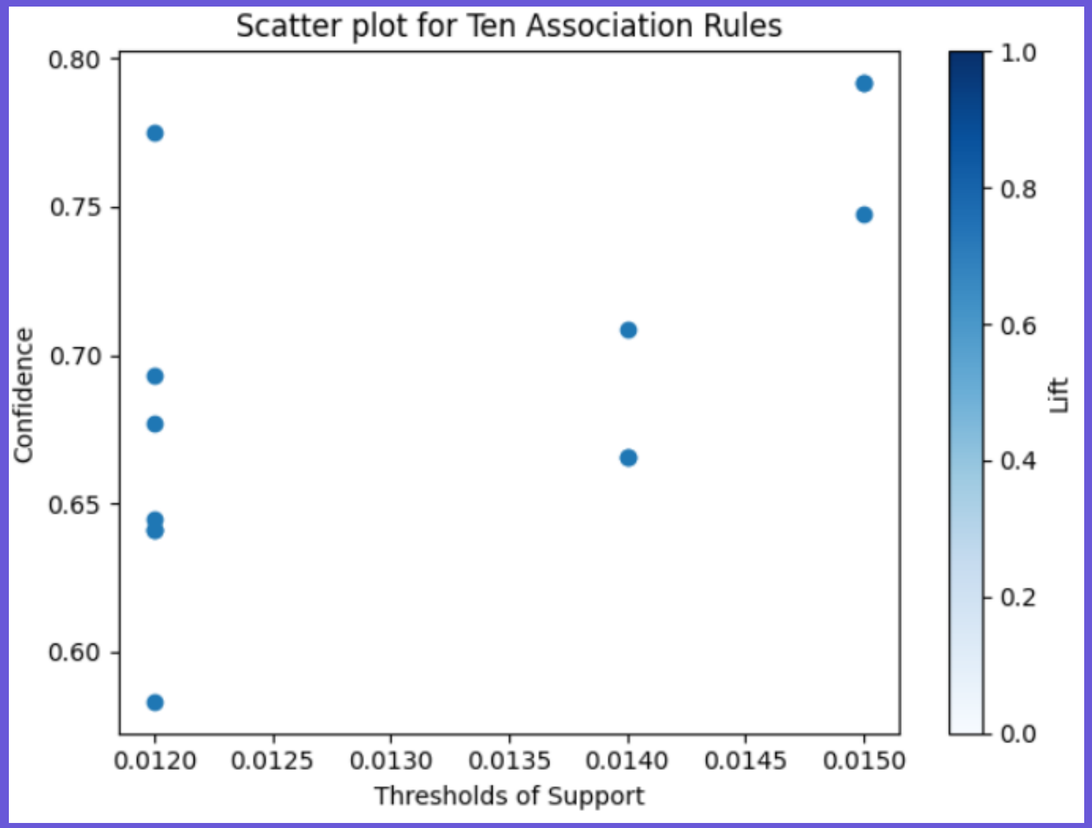
it refers to the frequency at which products Y and products X are purchased together compared to the total number of purchases that include products X.

the lift quantifies the difference in probabilities between buying products Y with the presence of products X and buying products Y without considering products X.

MBA-INSIGHTS

Taking into consideration the highest lift values, interesting associations emerge. For example those who frequently purchase product 32079082 often also buy products 32078795 and 32079103

antecedents	consequents	confidence	Lift
32079082	(32078795,32079103)	0.55	46.14

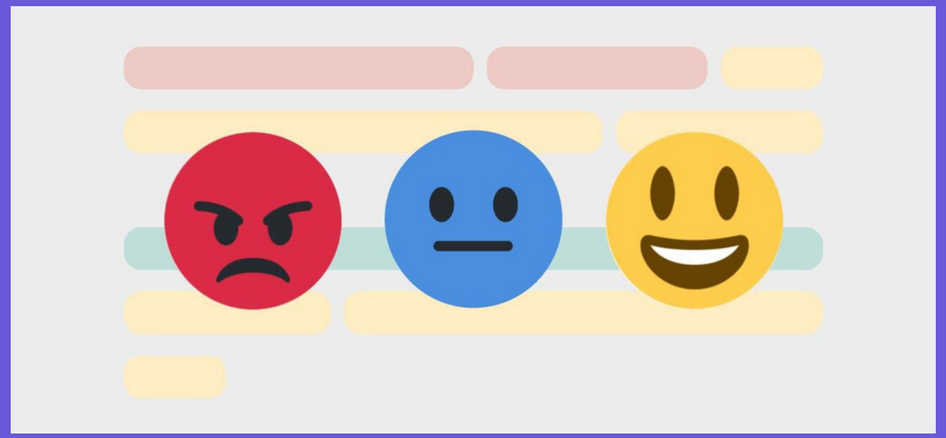


- Place these products in the same aisle or section of the supermarket.
- Offer the two products as a package or bundle at a discounted price compared to purchasing the individual products separately

FEEDBACK FOCUS

Sentiment Analysis is useful to analyse the feedback of the users of a product.
The Sentiment can be:

- positive
- negative
- neutral



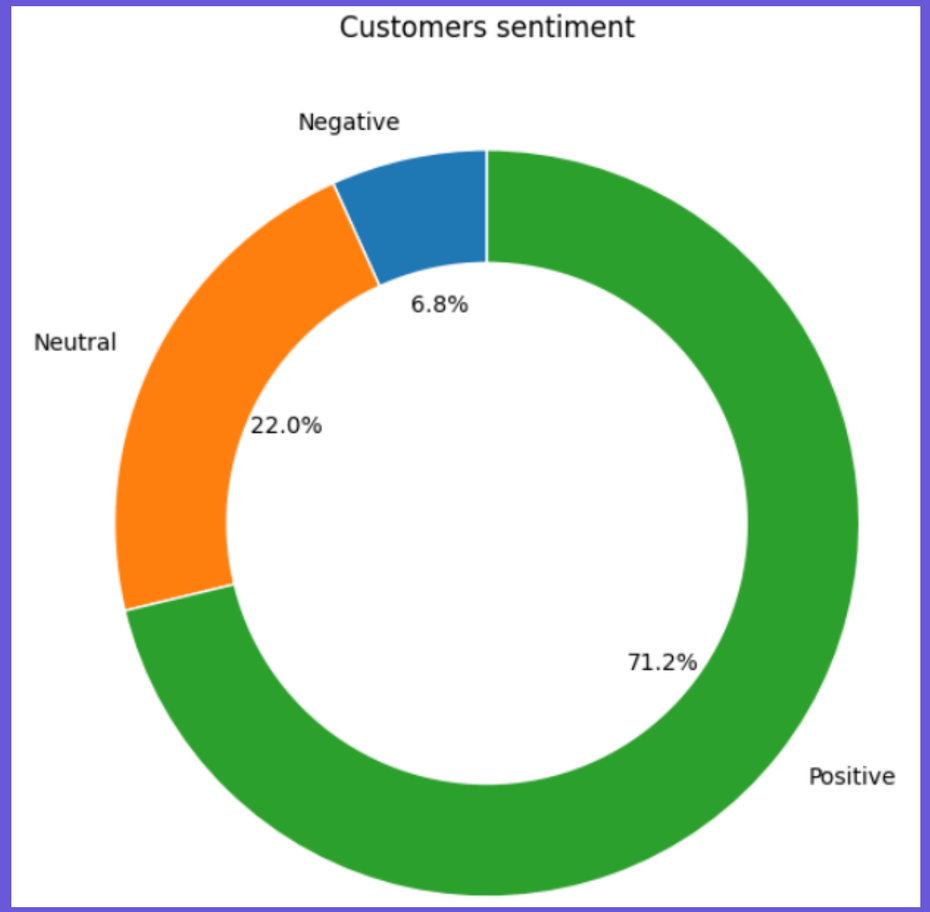
Before the use of the model,
it is necessary a preprocessing phase:

- elimination of punctuation → comma, dot,..
- elimination of stop words → a , the,..
- construction of usable features → Tf_Idf procedure

How is the sentiment of the costumers?

I train the Logistic Regression model on the labelled reviews and after I applied it to the customer reviews dataset

About the 7% of the reviews express a negative sentiment.
The 71% of the customers made a positive reviews



Gather reviews and testimonials from satisfied customers who have purchased both products together. This can provide social proof and increase the trust of potential buyers in purchasing both products.

HOW TO ACT

The 76% of the customers made a review

IDEA:
Investigate the opinion of the remaining customers through survey

Identify the detractor customers, understand their opinions

IDEA:
Propose a discount to the detractor customers

Identify the promoter customers, they should become the brand ambassador

IDEA:
Promotions: Bring a friend and get a discount!

Thank You