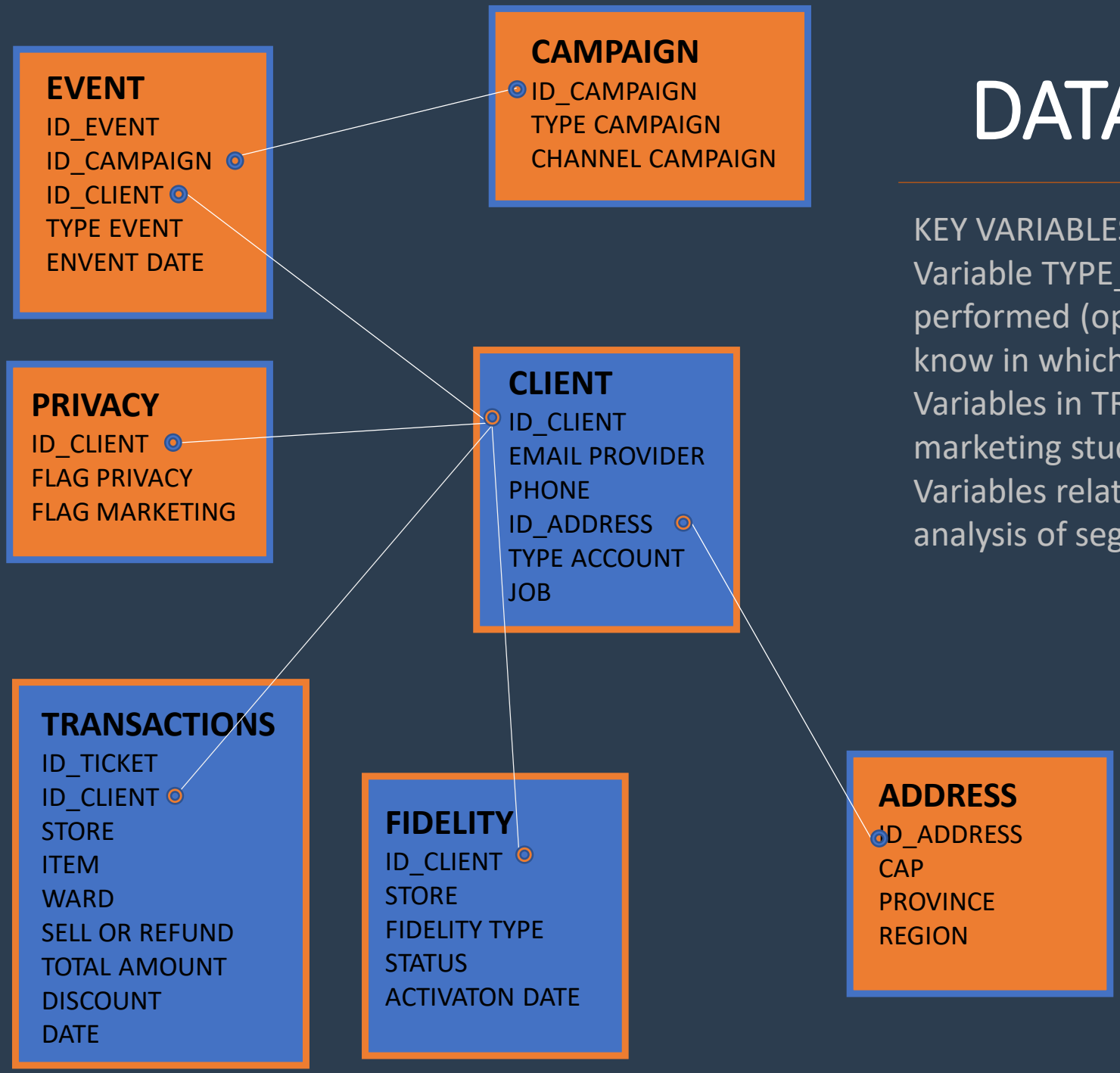
DIGITAL MARKETING

Sofia Davoli
Matricula: 813479
Email: s.davoli1@campus.unimib.it

AGENDA


MODELS DEVELOPEMENT
DATA PRESENTATION
BUSINESS QUESTIONS AND DATA EXPLORATION
RFM
MARKET BASKET ANALYSIS
CHURN MODEL
DATA DRIVEN CONCLUSION



DATA PRESENTATION

KEY VARIABLES:

Variable TYPE_EVENT allows to know the number of actions performed (open or click) with a campaign and allows to know in which date actions tooked place .

Variables in TRANSACTIONS dataset are at the center of marketing study.

Variables relative to client personal information allows analysis of segmentation (not in this project).

BUSINESS QUESTIONS

The study is based on a data-driven communication approach through the construction of Machine Learning models

RFM

Recency, Frequency and Monetary are Marketing analysis measures used for identify the most profitable customers. The RFM model is based on 3 main factors:

1. Recency of last purchases performed by customers
2. Frequency of purchases
3. Monetary amount of purchases

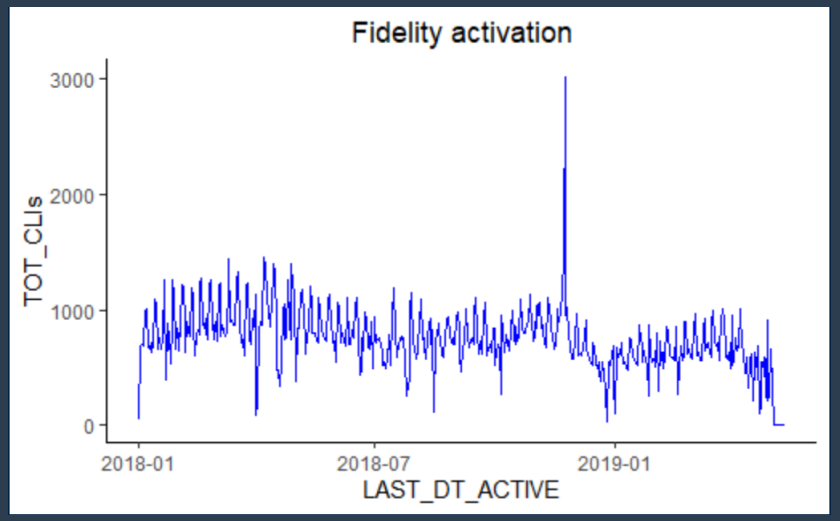
CHURN MODEL

The Churn model is a Marketing model fitted on a customer base which can identify if one of them no longer buys a product given a certain period. It is closely linked to the company's gain and reputation. Churn analysis is fundamental for CRM strategy plans.

MARKET BASKET ANALISYS

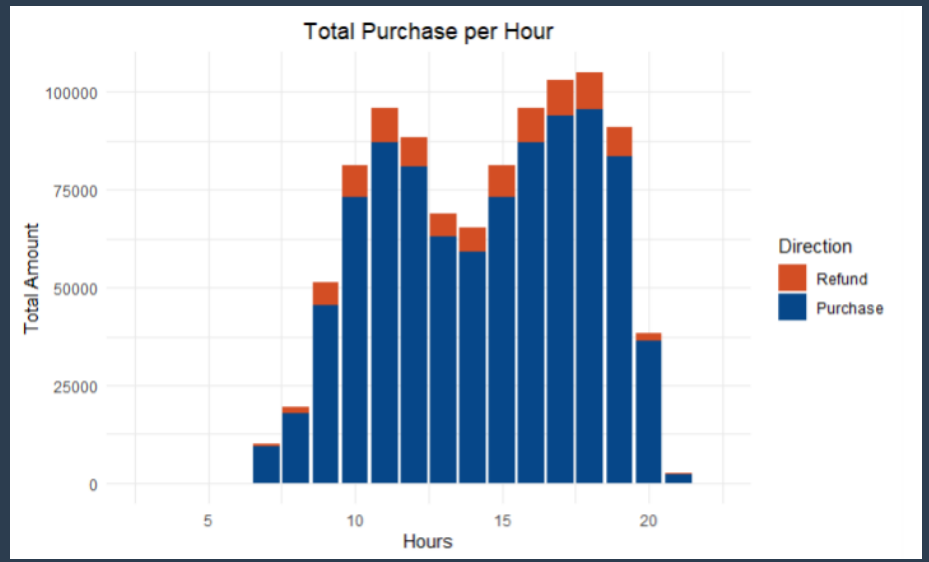
Market Basket Analysis is one of the main techniques used by companies to identify patterns in the purchase of products. The goal of the analysis is to identify the probability of identifying the association between 100 best selling products.

DATA EXPLORATION AND BUSINESS QUESTIONS



Fidelity subscriptions show remarkable increase during Christmas time. Probably during this period people are motivated to get a fidelity because of promotions for festivity. This information should be taken in account to set goals and to plan marketing campaign for next year Christmas period.
Store with more activation of fidelity is store 1, with a total of 15,7% of subscriptions, other stores contribute from 0,8%-3,4% each. Store number 1 is probably situated in a big city with many habitants with respect to other stores.

Highest **purchases** take place during 10-12 am and 4-6 pm. This information is useful to optimize the management of the working staff.
9% of total transactions are items returns. Most selled items are from department 10, followed by department 3 (in which it seems to be the highest rate of returned items). This information represent a start point for products improvement.

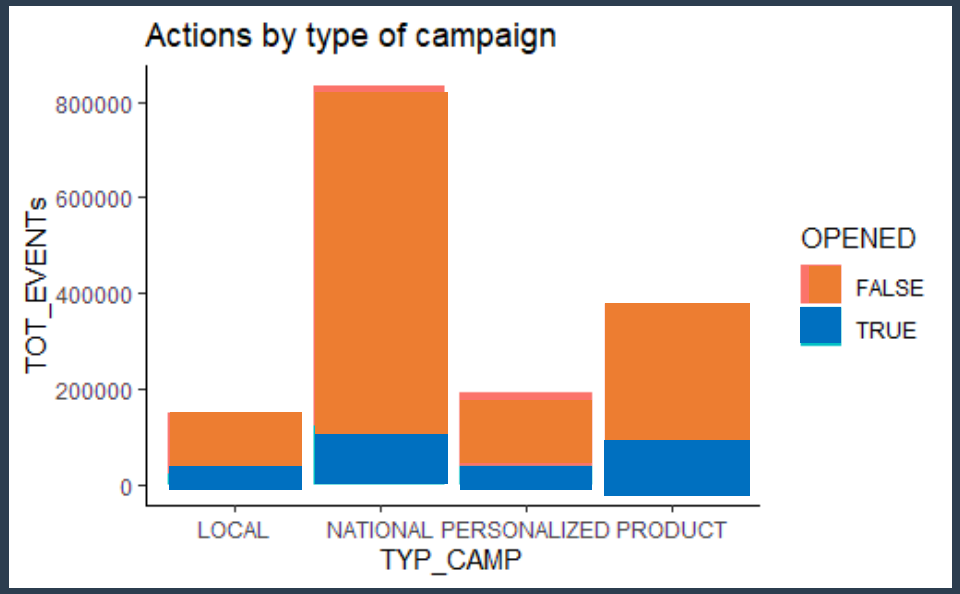


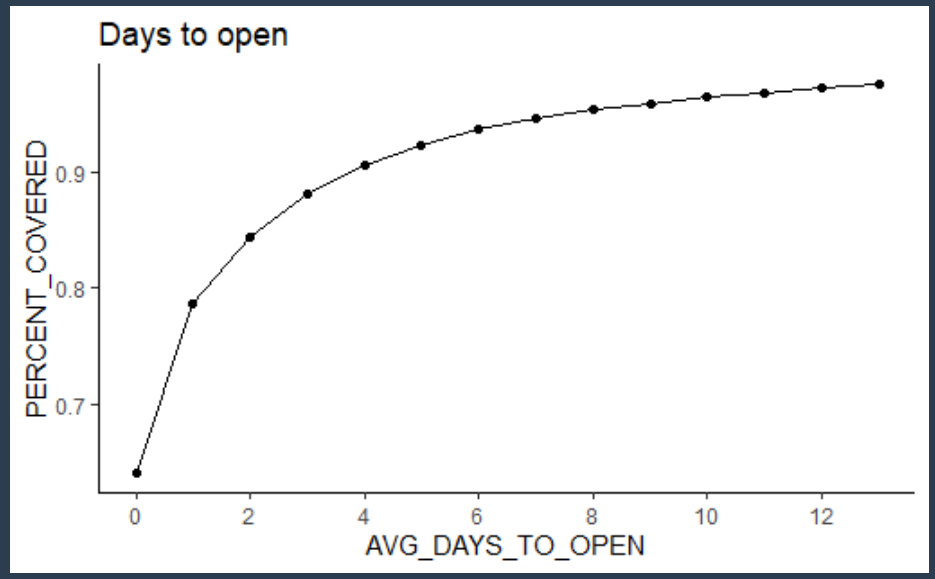
DATA EXPLORATION AND BUSINESS QUESTIONS

On a total of 369472 customers 11196 (33%) accepted **marketing analysis**. The distribution of **type of campaign** is: Product (43,8%), Personalized (19,9%), National (17,6%), Newsletter (12,8%) and Local (0,69%).

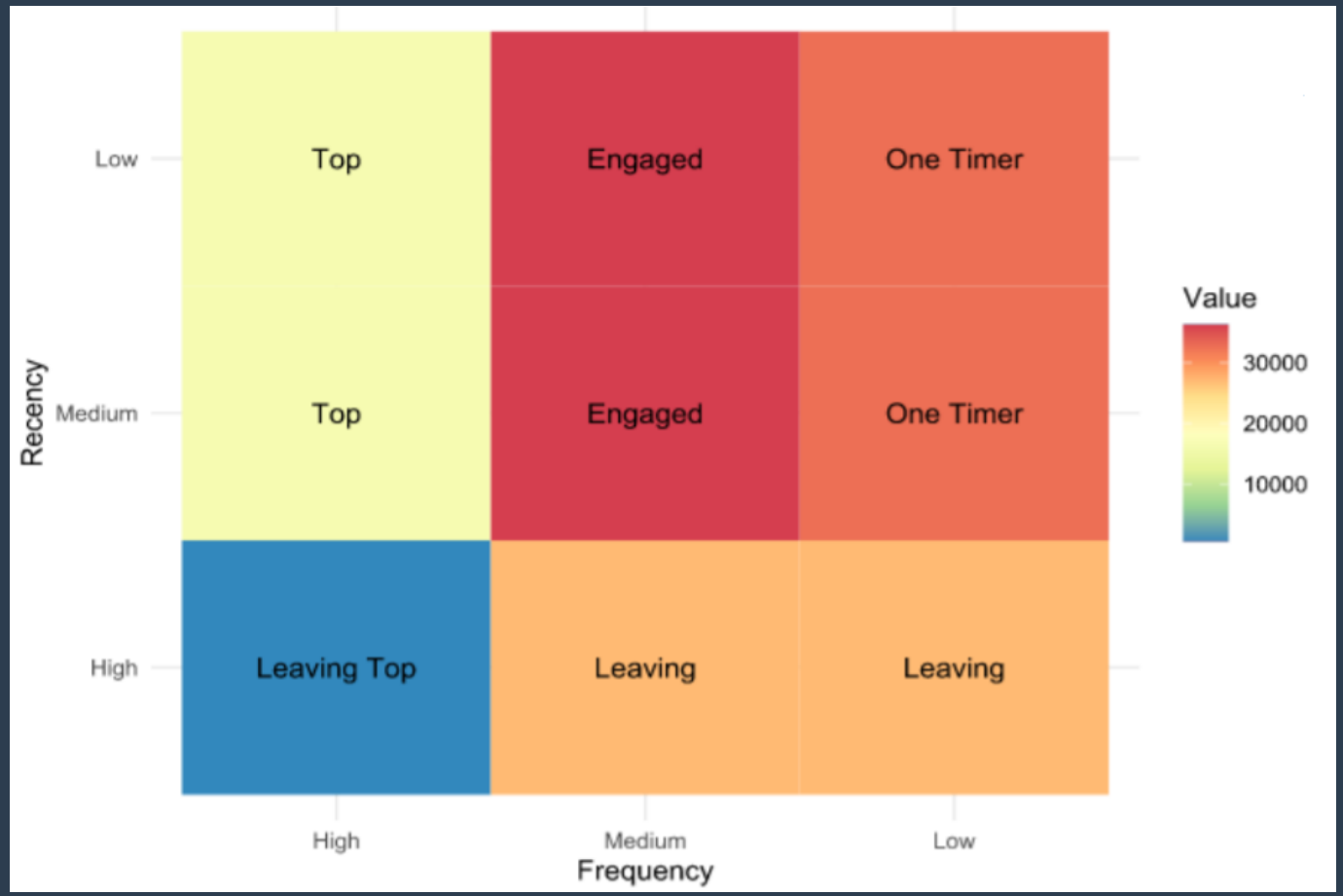
Result of campaign in **period from 3 January 2019 to 30 April 2019** show that product campaign get the highest success in term of opened action (more than 25% of opened), and national only get about 13% of open response.

Moreover, **opening action** are performed in the same day as sent email date time (about 65% of client open the received email the same day sent). In total 82,1% of people didn't open the campaign, 14% opened1 time. 97,7% didn't click the contents and only 1,7% clicked 1 time.





RFM MODEL



Customers (excluded those from shop 1) are analyzed based on how recently they bought a product, how often they buy and the total value spent. In this way it is possible to group customers into categories to evaluate the value of the same.
To analyze RFM:

1. Split customer base into active and non active
2. Split into groups based on the percentage distribution of clients to facilitate the interpretability of the data.
3. Recency-Frequency matrix show that mostly of customers belongs to Engaged class. Leaving Top class is the one with less customers.

Marketing actions should be performed to decrease number of Leaving and to increase loyalty of One Timer customer .

RFM MODEL

Finally Monetary information is added to the RF matrix.

RFM matrix allows to differentiate marketing action through customers base. For example many action should be taken to Diamond, Gold, Silver and Bronze customers, in particular actions of increasing loyality and supporting.

Customers belonging to the lowest right part of matrix are not valuables customer, best action to take are attracting action.

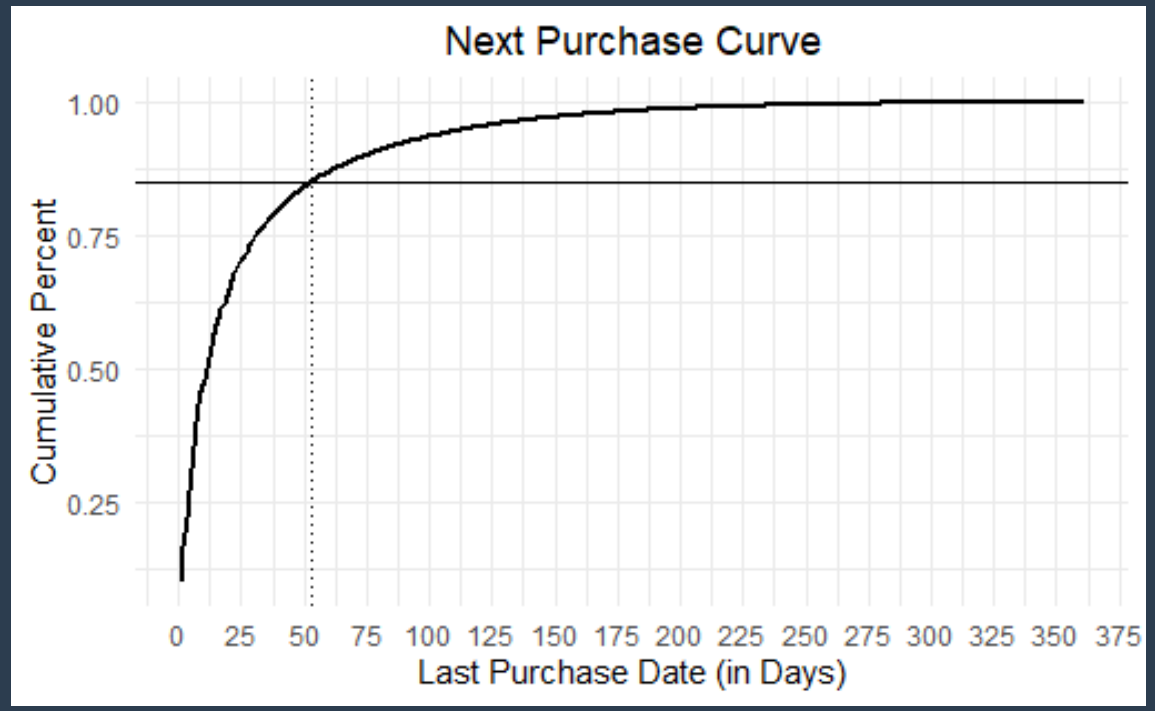
CHURN MODEL

The churn model provides a predictive estimate of the likelihood that a customer quit interacting with the company.

A customer is churned at a certain reference date if the customer does not purchase in the interval after the reference date corresponding to the Purchase Time Scale. In this case reference period is 53 days (length of a holdout period).

To each customer is assigned a target 0/1 variable such that 1 is assigned to customers who churned in the holdout period.

5 propensity supervised models are trained choosing as lookback date 2018-05-01 to 2019-1-1.



0 - non churn	2107266
1 - churn	3389527

CHURN MODEL

Training and testing partitions are created using 70%- 30% of data respectively.

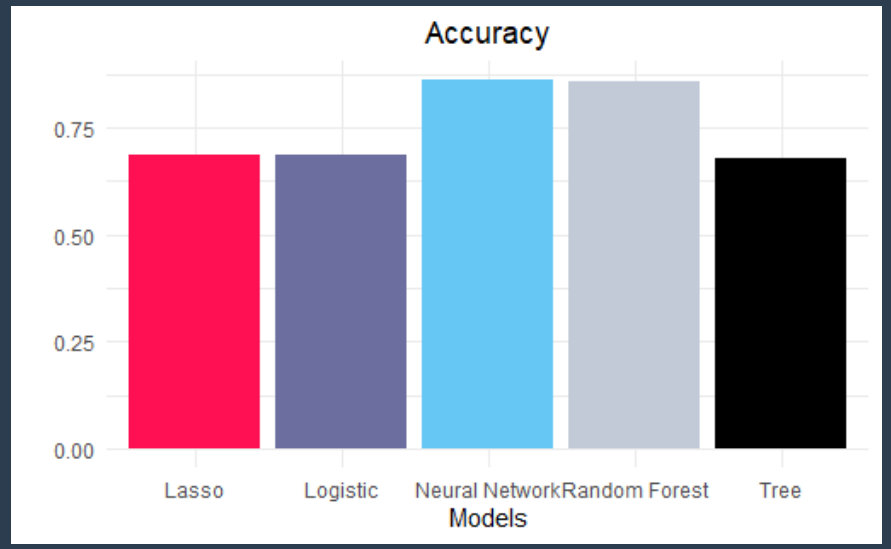
Relevant predictors variables are chosen based on:

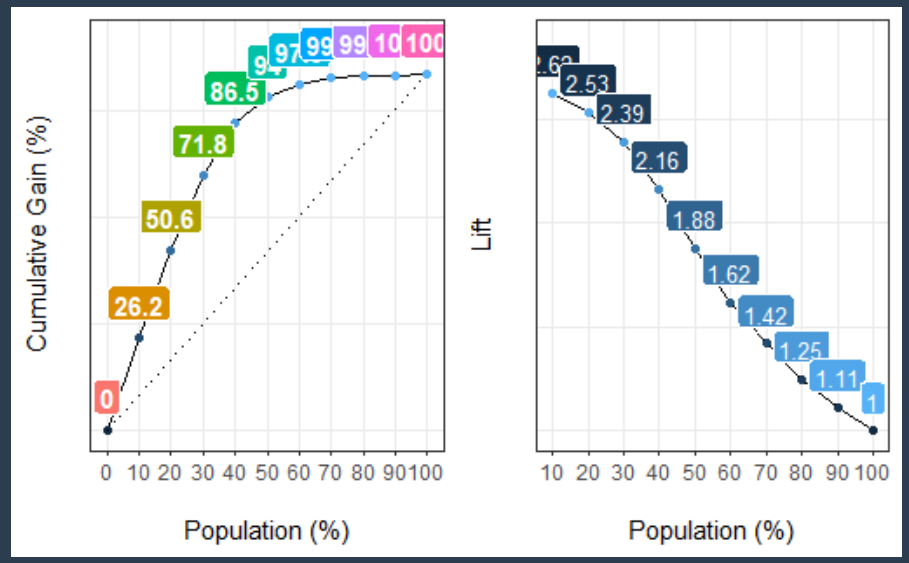
- Customer information (type of job, residence region and type of fidelity)
- Customer behavior (recency, frequency and monetary values)

Models trained are:

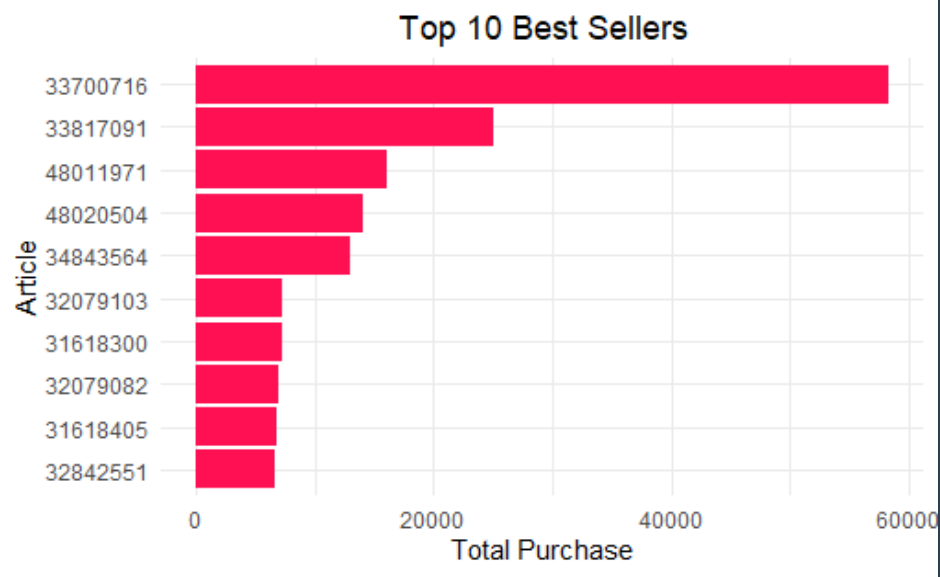
1. simple tree
2. random forest
3. logistic regression
4. lasso
5. neural network

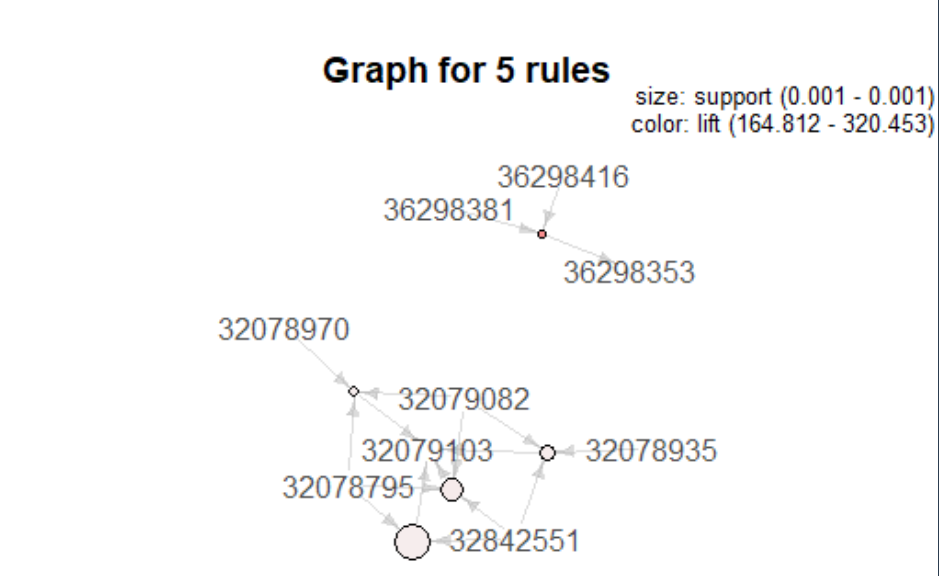
Best model result in being Neural Network model with accuracy metric equal to 86.2%. Lift curve show the percentage of people classified positively according to the size of the sample. In particular with only 50% of data we can predict with 86% probability of being correct, a churner.





MARKET BASKET ANALYSIS





The best-sold product is the 33700716, with approximatively 60,000 units sold followed by the product 33817091, with over 20,000 units purchased.

Most frequently observed patterns are from 3 to 7 products per transaction, with a mean of 32.65$ per transaction.

The Association Rule is used to analyze purchases of customers in transaction data, in order to identify strong item's relationships using measures of interest.

The Association Rule is built with support of at least 0.1 % and a confidence of 80%. It is observed that 95% of customers who purchase products {32078795, 32079082, 32842551} they also buy 32079103 product with a high relationship among the products.

Graph show top 5 associations rules.

CONCLUSIONS

RFM MODEL	RFM matrix allows to differentiate action to take across different types of customers. Principal actions should focus on attracting, gain loyalty and satisfying customers.
CHURN MODEL	Churn models highlights client to which fidelity actions should refer. Focus group for churner customer should be taken to improve services for customers.
MARKET BASKET ANALYSIS	Results from market basket analysis give an overview of general best-sold items set. Action suggested by this analysis includes re-allocation and promotion of items in stores
OTHER	A further develop of this study could be to conduct MB Analysis on single store data or on same region data to better allocate items in stores according to regional patterns. From explorative analysis company should take action to optimize e-mail campaign and should also stimulate customers through other channels.