POLITECNICO MILANO 1863 | DIPARTIMENTO DI INGEGNERIA GESTIONALE

Non contractual settings

Analytics for CRM

Marketing Analytics

Open challenges in CRM

- **Predictive Analytics**: forecasting customer behavior (e.g., churn, upselling or cross-selling) → regression, machine learning models, and time-series forecasting
- **Prescriptive Analytics**: suggesting actions potentially linked to a positive outcome (e.g., NBO/NBM, caring, suggested communication strategies, etc.) → Optimization and simulation algorithms, maybe inspired by MBA or similar
- **Customer Segmentation**: including demographic information, purchase behavior, engagement, etc. → K-means or hierarchical clustering, latent class analysis etc.
- **Sentiment Analysis**: understanding customer sentiments, opinions, and attitudes expressed (e.g., in social media or even in customer service channels) → NLP and text analysis

CRM Analytics:

NON CONTRACTUAL vs CONTRACTUAL

Churn is <u>not</u> explicit — Churn is explicit

P(0;1) — [0;1]

Contract: B2B, mobile bills, utility, subscriptions

Churn in explicit and predictable (either 0 or 1)

Non-Contract: grocery, stores, supermarket

It Is not explicit but has a probability of 0 to 1

Churn is <u>not</u> explicit — Churn is explicit

P(0;1) — [0;1]

↳ Clustering

↳ Market-Basket Analysis (MBA)

Open challenges in CRM

- **Social Network Analysis**: analyzing social relationships to identify hot spots in the network, communities, and its dynamics → network analysis
- **Web and E-commerce Analytics**: Continuous journey and conversion improvement moving from KPIs and analytics → clickstream analysis, conversion rate optimization, systematic A/B testing
- **Churn Analysis**: identifying levers and tools to explain/predict/prevent customer abandonment → machine learning, logistic regression, random forest, genetic models, survival analysis
- **Lifetime Value Analysis**: CLV modelling → bayesian models

RFM Analysis

Customer assessment: RFM Analysis

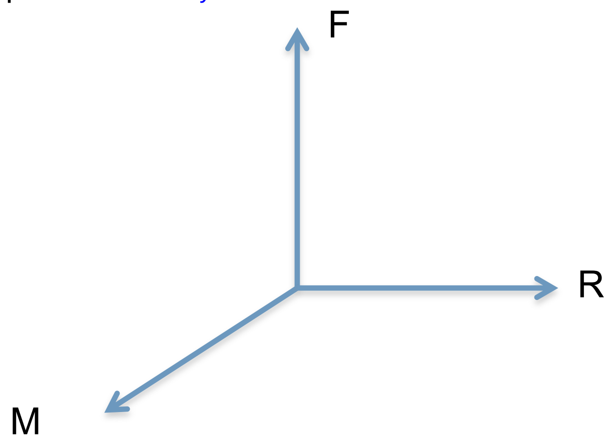
Recency – How recently has the customer bought? when the last purchase was?

Frequency – How often does the customer buy?

Monetary – How much does the customer spend? relatively to other customers

Thresholds:

- 20-50-80
- 5-20-50-80
- 33-33-33



A possible RFM-based segmentation

Customer Segment	Activity	Actionable Tip
Champions	Bought recently, buy often and spend the most	Reward them. Can be early adopters for new products. Will promote your brand.
Loyal Customers	Spend good money with us often. Responsive to promotions.	Upsell higher value products. Ask for reviews. Engage them.
Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	Offer membership / loyalty program, recommend other products.
New Customers	Bought most recently, but not often.	Provide on-boarding support, give them early success, start building relationship.
Promising	Recent shoppers, but haven't spent much.	Create brand awareness, offer free trials
Customers Needing Attention	Above average recency, frequency and monetary values. May not have bought very recently though.	Make limited time offers, Recommend based on past purchases. Reactivate them.
About To Sleep	Below average recency, frequency and monetary values. Will lose them if not reactivated.	Share valuable resources, recommend popular products / renewals at discount, reconnect with them.
At Risk	Spent big money and purchased often. But long time ago.	Send personalized emails to reconnect, offer renewals, provide helpful resources.
Can't Lose Them	Made biggest purchases, and often. But haven't returned for a long time.	Win them back via renewals or newer products, don't lose them to competition, talk to them.
Hibernating	Last purchase was long back, low spenders and low number of orders.	Offer other relevant products and special discounts. Recreate brand value.
Lost	Lowest recency, frequency and monetary scores.	Revive interest with reach out campaign, ignore otherwise.

Source: Putler, 2017

Reflections on RFM

1. Choose the right timeframe for your RFM analysis (it depends on the industry)
2. Choose the most adequate threshold values. Possibilities:

   Rules based (e.g., «who bought more than x€ is high in monetary)

   Algorithmic (e.g., above 66° percentile is high, below 33° is low, etc.)

   Mixed approach (e.g., let's calculate the percentiles, and then let's see when there is a decise drop)
3. Evaluate whether recency or regularity (or both) are better

   Recency is good when CRM actions are real-time

   Regularity is good when it depicts an actual buying behavior

time if a customer doesn't come back it's lost. for grocery it's like once a week or at least one a month. holiday exist

for frequency (tell alot about the regularity)

$$\text{interpurchase\_time} = \frac{\text{last purch.} - \text{first purchase}}{n-1}$$

$n = \#\ \text{tickets}$

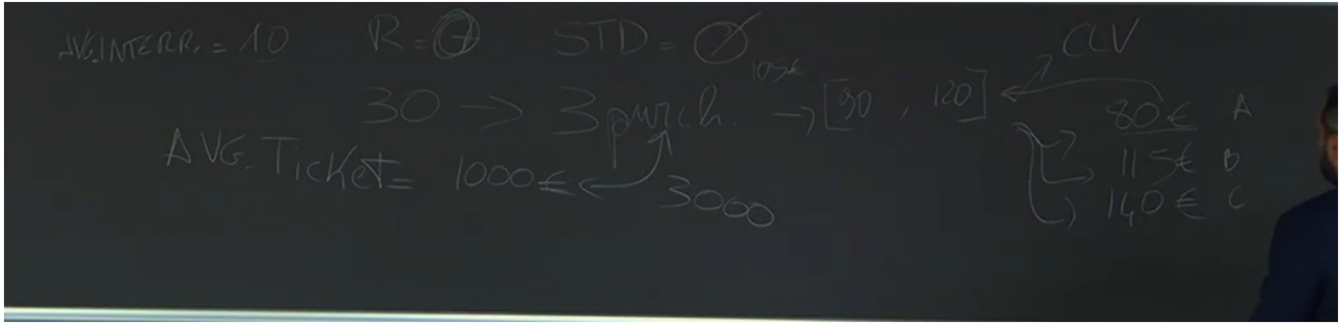
σ_{IPT}

$$VC = \frac{\sigma_{IPT}}{\mu_{IPT}} \begin{cases} \leq 0.5 \rightarrow \text{Regular} \\ (0.5-1] \rightarrow \text{Somewh. Reg.} \\ (1-1.5] \rightarrow \text{Somewh. Irr.} \\ > 1.5 \rightarrow \text{Highly Irregular} \end{cases}$$

Exercise

You have a dataset of tickets in a retailer; build the RFM determining:

- The timeframe of analysis
- The thresholds
- The resulting clustering
- The possible CRM strategies deriving from your analysis



Readings

Hansotia, Behram J., and Paul Wang. "Analytical challenges in customer acquisition." *Journal of Interactive Marketing* 11.2 (1997): 7-19.

Bansal, Harvir S., Shirley F. Taylor, and Yannik St. James. "“Migrating” to new service providers: Toward a unifying framework of consumers’ switching behaviors." Journal of the Academy of Marketing Science 33.1 (2005): 96-115.

Neslin, Scott A., et al. "Defection detection: Measuring and understanding the predictive accuracy of customer churn models." *Journal of marketing research* 43.2 (2006): 204-211.

Bijmolt, Tammo HA, et al. "Analytics for customer engagement." *Journal of Service Research* 13.3 (2010): 341-356.

Logistic Regression in SPSS: https://statistics.laerd.com/spss-tutorials/binomial-logistic-regression-using-spss-statistics.php#procedure

Decision Tree model in SPSS: http://ecapitaladvisors.com/blog/creating-decision-tree-analysis-using-spss-modeler/

Market basket analysis

AGENDA

- When and why is market basket analysis needed
- Affinity metrics
- Market basket analysis operationalization
- Marketing analytics lab troubleshooting

Frequently bought together

+

+


Total price: £304.21

Add all three to Basket

i These items are dispatched from and sold by different sellers. Show details

- ☑ **This item:** Canon EOS 4000D DSLR Camera and EF-S 18-55 mm f/3.5-5.6 III Lens - Black £255.00
- ☑ Canon ES100 Camera Bag - Black £38.63
- ☑ SanDisk Ultra SDXC Memory Card Up to 80 MB/s, Class 10, U1, 64 GB, Black/Grey £10.58

MATCH WITH

SIMILAR ITEMS

Market basket analysis is a data mining method focusing on discovering purchasing patterns by extracting associations or co-occurrences from transactional data.

Market basket analysis determines the products which are bought together and supports in:

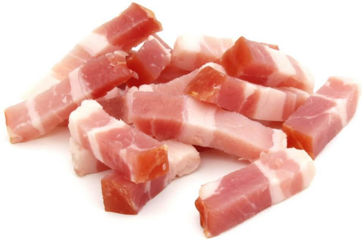
- Reorganising the supermarket layout or websites' contents;
- Designing promotional campaigns such that products' purchase can be improved.

Two basic assumptions of Market Basket Analysis:

1) Some items are bought together because of their natural affinity

Two basic assumptions of Market Basket Analysis:

2) Some items are bought together because of planning and consideration



Itemset: a collection of one or more items purchased by a customer

Antecedent: the itemset on left hand side (LHS) – eggs, cheese, and spaghetti

Consequent: the itemset on right hand side (RHS) – cheek lard

Support: frequency with which an antecedent–consequent pair appears together in the transactions of a dataset.

Confidence: proportion of transactions containing the consequent among those that include the antecedent, and therefore expresses the inferential reliability of the rule

Lift: measure of the performance of the model at predicting cases as having an enhanced response, measured against a random choice targeting model.

- Lift < 1: the presence of LHS has negative effect on RHS
- Lift = 1: LHS and RHS are independent
- Lift > 1: the presence of LHS has positive effect on RHS

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Each row represents one transaction (not one customer)

Each column is occupied by an item purchased in the given transaction.

Note that:

- It considers only the presence of an item in a transaction, not the quantity
- The sequence of purchase is not considered

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Itemset	*No. Transactions*	*Support (A&B)*
Eggs, Milk		
Fruit, Vegetable		
Fruit, Eggs		
Vegetable, Milk		

$$\text{Support (A\&B)} = \frac{\text{No. Transactions}}{\text{Total transactions}}$$

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Itemset	*No. Transactions*	*Support (A&B)*
Eggs, Milk		
Fruit, Vegetable		
Fruit, Eggs		
Vegetable, Milk		

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Itemset	*No. Transactions*	*Support (A&B)*
Eggs, Milk	3	60%
Fruit, Vegetable		
Fruit, Eggs		
Vegetable, Milk		

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

= 3 / 5

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Itemset	*No. Transactions*	*Support (A&B)*
Eggs, Milk	3	60%
Fruit, Vegetable	3	60%
Fruit, Eggs	3	60%
Vegetable, Milk	2	40%

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Confidence
Fruit => Vegetable	60%	
Vegetable => Fruit	60%	
Vegetable => Milk	40%	
Milk => Vegetable	40%	

Rule: *A* => *B*
A: Antecedent
B: Consequent

Meaning: if A is bought, then B is bought.

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Confidence
Fruit => Vegetable	60%	
Vegetable => Fruit	60%	
Vegetable => Milk	40%	
Milk => Vegetable	40%	

Rule: *A* => *B*
A: Antecedent
B: Consequent

Meaning: if A is bought, then B is bought.

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Confidence
Fruit => Vegetable	60%	75%
Vegetable => Fruit	60%	
Vegetable => Milk	40%	
Milk => Vegetable	40%	

= 3 / 4

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Confidence
Fruit => Vegetable	60%	75%
Vegetable => Fruit	60%	
Vegetable => Milk	40%	
Milk => Vegetable	40%	

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

= 3 / 4

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Confidence
Fruit => Vegetable	60%	80%	75%
Vegetable => Fruit	60%		
Vegetable => Milk	40%		
Milk => Vegetable	40%		

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Confidence
Fruit => Vegetable	60%	80%	75%
Vegetable => Fruit	60%	80%	75%
Vegetable => Milk	40%	80%	50%
Milk => Vegetable	40%	60%	67%

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Lift
Fruit => Vegetable	60%	80%		
Vegetable => Fruit	60%	80%		
Vegetable => Milk	40%	80%		
Milk => Vegetable	40%	60%		

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Lift
Fruit => Vegetable	60%	80%	80%	
Vegetable => Fruit	60%	80%		
Vegetable => Milk	40%	80%		
Milk => Vegetable	40%	60%		

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Lift
Fruit => Vegetable	60%	80%	80%	0.94
Vegetable => Fruit	60%	80%		
Vegetable => Milk	40%	80%		
Milk => Vegetable	40%	60%		

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

= 0.6 / (0.8 * 0.8)

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Lift
Fruit => Vegetable	60%	80%	80%	0.94
Vegetable => Fruit	60%	80%	80%	0.94
Vegetable => Milk	40%	80%	60%	0.83
Milk => Vegetable	40%	60%	80%	0.83

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

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Lift
Fruit => Vegetable	0.94
Vegetable => Fruit	0.94
Vegetable => Milk	0.83
Milk => Vegetable	0.83

- Lift < 1: the presence of A has negative effect on B
- Lift = 1: A and B are independent
- Lift > 1: the presence of A has positive effect on B

When milk is not purchased (transaction 1 & 5), vegetable is purchased 100% of the times. While vegetable is only purchase 67% of the times with milk.

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Confidence	Lift
Eggs => Milk	?			?	?

Transaction ID				
1	Fruit	Vegetable		
2	Fruit	Eggs	Milk	Meat
3	Vegetable	Eggs	Milk	Bread
4	Fruit	Vegetable	Eggs	Milk
5	Fruit	Vegetable	Eggs	

Rule	Support (A&B)	Support (A)	Support (B)	Confidence	Lift
Eggs => Milk	60%	80%	60%	75%	1,25

Support: high support identifies the items that are bought frequently.
These items potentially *drive traffic to the store*, and could be applied to a large number of customers

Confidence: high confidence item is most likely being bought together with some other items.
Recommendation or shelf placement, and bundling of such product could streamline customers' experience. Optimizing pricing of these items could enhance margin.

Lift: high, positive lift shows that antecedent itemset can predict the presence of the consequent itemset. Negative lift suggests that the absence of antecedent itemset better predicts the presence of consequent itemset.

Appendix: MBA on R

Reflections on MBA

Level (scope) of the MBA	Possible insights
Customer base	
RFM quadrants/CLV	
Individual customer	

Reflections on MBA

Level (scope) of the MBA	Possible insights
Customer base	Layout, promotions Using the rules to profile by tastes and preferences
RFM quadrants/CLV	Segmented insights (e.g., how to make customers stretch their ticket)
Individual customer	Recommendation

Further reflections on MBA

- How could the MBA insights be further used by a retailer or a marketplace?
- Who could be interested to MBA insights?
- Would the opposite of MBA be useful? Why?