

# DIGITAL MARKETING

Franческа De Cola 819343

Laura Nembrini 819059

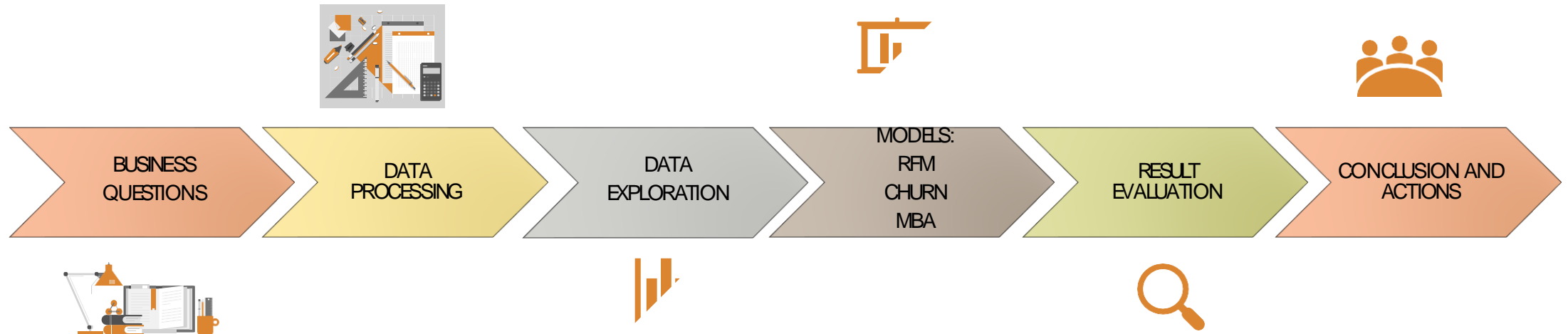
A.A. 2019-2020 DISCO



Questa foto di Autore sconosciuto è concesso in licenza da [CC BY-SA](#)

# PIPELINE FOR THE QUANTITATIVE APPROACH

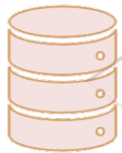
Il punto di partenza per un approccio quantitativo al marketing è porsi le giuste **business questions**.



# DATA

## INFORMATION:

- Informazioni sui clienti,
- Informazioni su campagne di marketing via mail



### FIDELITY:

informazioni sulle fidelity subscriptions di ogni customer account.



### ACCOUNT:

informazioni su ogni customer account.



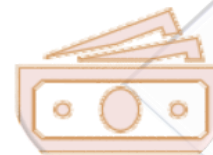
### ADDRESS:

informazioni sugli indirizzi corrispondenti a un customer account.



### CAMPAIGN EVENT:

contiene gli eventi (sents, opens, clicks.) relativi alle comunicazioni email



### TIC:

contiene i purchase e i refund delle transazioni di ogni customer.



### CAMPAIGN:

categorizzazione delle comunicazioni di marketing email.



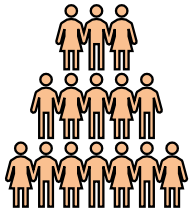
### PRIVACY:

informazioni sulle privacy policies accettate da ogni customer.

# BACKGROUND & BUSINESS QUESTIONS



**CHI?**  
DIGITAL RETAIL  
COMPANY



**OBIETTIVO:**  
CONSERVARE GLI '*HIGH-  
VALUE CUSTOMERS*',  
*AUMENTARE PROFITTI*



Come prevedere la probabilità di abbandono di un cliente?



Come ottenere una visione generale del comportamento della clientela al fine di segmentare i vari clienti per svolgere azioni ad hoc?



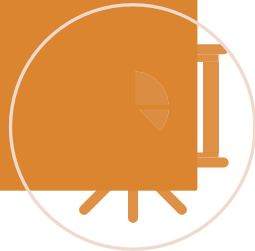
Quali sono le tendenze di acquisto?

# MODELLING APPROACHES

Il modello **RFM** riesce a differenziare le azioni di marketing segmentando la customer base sulla base della distribuzione di tre variabili chiave:

1. **Recency**: quando il cliente ha effettuato l'ultimo acquisto?
2. **Frequency**: quanti acquisti ha effettuato il cliente in un certo periodo?
3. **Monetary**: quanto ha acquistato il cliente in un certo periodo?

## RFM ANALYSIS



La **MBA** è l'analisi delle associazioni tra prodotti che ha come obiettivo l'individuazione delle **regole di associazione** tra item che appaiono congiuntamente in un evento, identificando pattern di prodotti che vengono acquistati con maggiore probabilità.

## MARKET BASKET ANALYSIS



Lo scopo del modello di **CHURN** è l'assegnazione per ciascun cliente di una probabilità di **abbandono** al fine di implementare azioni specifiche finalizzate al **trattenimento** dei clienti con più alto valore. Risulta fondamentale per il profitto dell'azienda riuscendo a identificare le aree dove i servizi sono carenti.

## PROPENSITY TO CHURN



# ANALISI ESPLORATIVA

Prima dell'implementazione dei modelli è necessario svolgere un'analisi esplorativa dei dati mediante l'uso di grafici e misure statistiche al fine di scoprire relazioni importanti e generare i KPI.

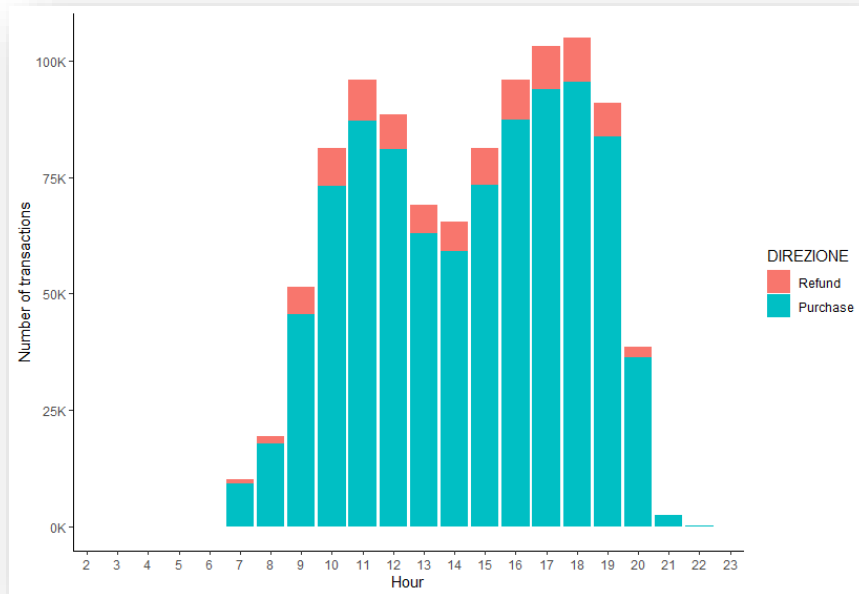


Figura 1. Distribuzione delle transazioni per ora.

Distribuzione degli acquisti (blu) dei prodotti per ora. Gli orari di maggior frequenza sono dalle 10 alle 12 e dalle 17 alle 19 in linea a ciò che ci si aspetta. Gli orari sono forse corrispondenti alle pause da lavoro nel mattino e alla fine della giornata.

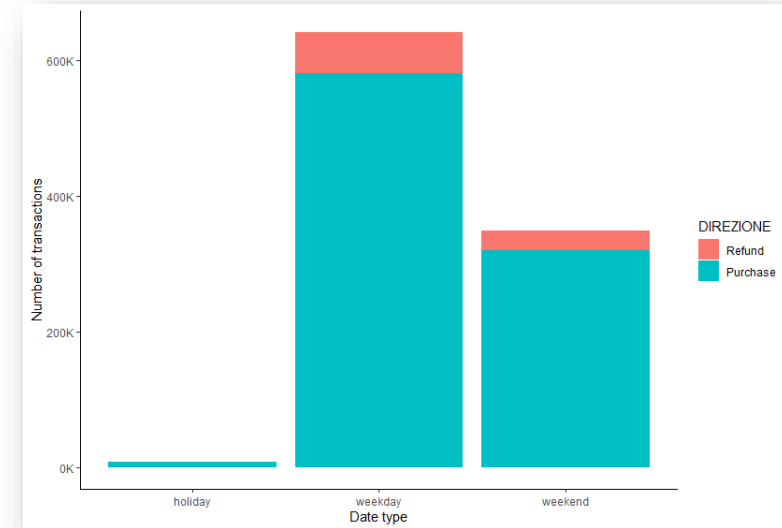


Figura 2. Numero di transazioni (acquisto e refund) per vacanza, weekend, e giorno della settimana.

Dai dati sugli scontrini si evince che gli acquisti sono pochi nei giorni di vacanza contrariamente a come ci si aspetterebbe. Questa informazione potrebbe essere utile per condurre campagne ad hoc su specifici prodotti nei periodi in cui si acquista di più

# RFM

Per studiare la customer base è necessario raggruppare i clienti in fasce secondo la distribuzione percentuale per i valori di **Recency**, **Frequency** e **Monetary** dopo averli divisi in **Attivi** e **Non Attivi**.

Soglia oltre la quale si classificano i clienti come **ATTIVI**:

- si hanno dati dal 01/05/2018 al 30/04/2019
- se l'ultimo acquisto è dopo **01/01/2019** il cliente è attivo



Figura 3. Matrice RF

	RECENCY	FREQUENCY	MONETARY
LOW	<25%	<2 Acquisti	<25%
MEDIUM	25% < X < 75%	2< Acquisti <5	25% < X < 75%
HIGH	>75%	>5 Acquisti	>75%

Tabella 1. Distribuzione percentuale per costruzione matrice RFM

Nella matrice RF si osserva che la maggior parte dei clienti è **Engaged**, pochi **Leaving Top**. Si potrebbe mirare con azioni di marketing la clientela One Timer al fine di renderla Engaged. I **Leaving** potrebbero essere incentivati a non lasciare l'azienda con delle sconti personalizzate.

# RFM

Combinando i valori della matrice RF con la **Monetary** si ottiene la matrice **RFM**.

Posso scegliere di condurre azioni per i clienti più solidi come i **'Diamond'** o risparmiare budget non investendo in quelli più deboli come i **'Cheap'** che spendono e acquistano poco.

Il gruppo con più clienti è il **'Bronze'** con circa **25k** persone, pochi sono i clienti **'Gold'**.

Azioni marketing mirate ad accrescere il valore della categoria più numerosa dei **'Bronze'** porterebbe ad un **incremento** delle vendite.



Figura 4. Matrice RFM



# CHURN

Il modello di CHURN provvede una stima di verosimiglianza che un cliente smetta di interagire con l'azienda (es: smette di acquistare) attraverso un propensity supervised model. Gli step dell'analisi sono :

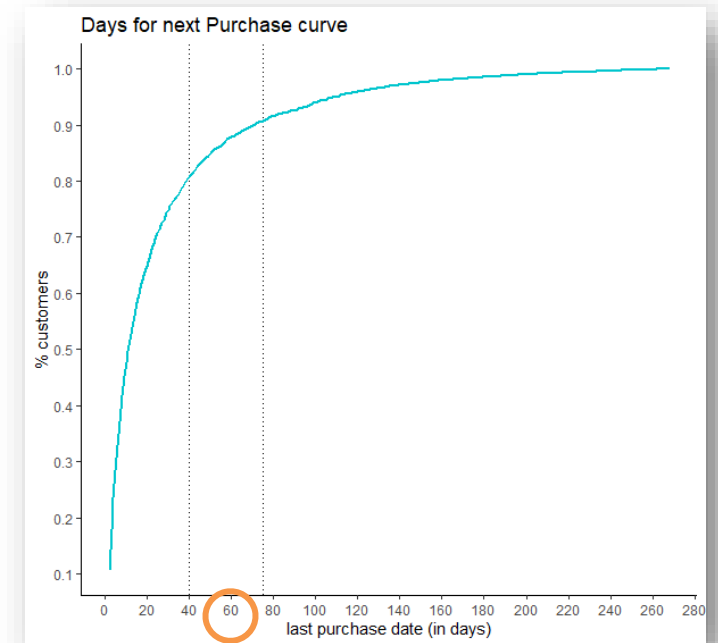
- Scegliere una reference date nel passato:
  - **28 Febbraio 2019**
- Impostare lunghezza del periodo di holdout dopo la reference date:
  - **28 Febbraio 2019 - 30 aprile 2019**
  - **60 giorni** (Days for next Purchase Curve)
  - Tutti i clienti nell'holdout hanno effettuato un acquisto
- Scegliere la lunghezza del lookback period prima della reference date:
  - **01 maggio 2018 - 28 febbraio 2019**
- Assegnare ad ogni cliente una variabile churn target 0/1, dove 1 è per coloro che hanno 'churned' nel periodo di holdout;
- Definire un set di potenziali predittori da usare nel lookback period scelti calcolando la correlazione tra ognuno e il dataset:

- TIPO DI PROGRAMMA FEDELTA' DEL CLIENTE
- SPESA TOTALE
- NUMERO DI ACQUISTI
- RECENCY
- REGIONE DI PROVENIENZA
- SE ACCOUNT è PRIMARIO O MENO
- ID DELLO STORE DI REFERENZA.

CHURN (1)	NO CHURN (0)
39398	74940

Tabella 2. Divisione Churn nel dataset

Figura 5. Days for next purchase curve.



Per circa il **90%** dei clienti passano **60/70 giorni** dall'ultimo acquisto. Indagando i comportamenti di acquisto è possibile individuare le aree che richiedono un intervento e aumentare la **profitability**.

# CHURN

Per approcciarsi ai modelli in maniera corretta, è necessario dividere il dataset in **Train (70%)** e **Test (30%)**. Il primo viene utilizzato per allenare il modello mentre il secondo per valutare le performance del modello.

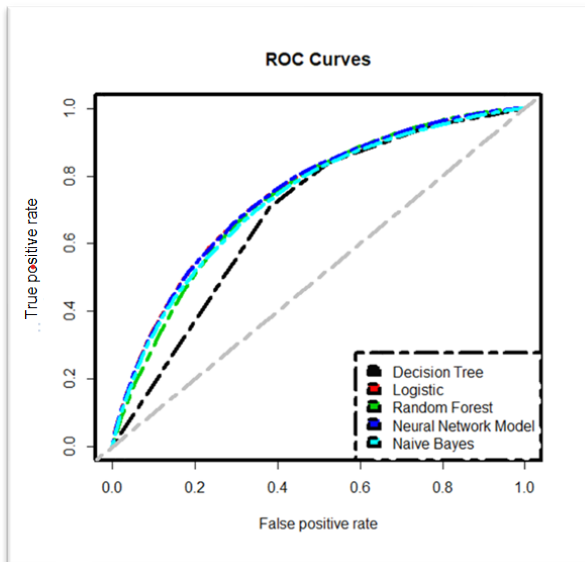
Modelli implementati:

- RANDOM FOREST
- DECISION TREE
- NEURAL NETWORK
- LOGISTIC REGRESSION
- NAIVE BAYES

	Decision Tree	Logistic Regression	Naive Bayes	Neural Network	Random Forest
Accuracy	0.715	0.716	0.704	0.719	0.719

Tabella 3. Accuracy modelli.

Confronto curve di ROC



## AUC

- NN 0.7456155
- Naive Bayes 0.7339023
- Random Forest 0.736105
- Decision Tree 0.6905359
- Logistic Regression 0.7452261

Dalla tabella, è possibile notare che i modelli più performanti in termini di accuracy sono **Neural Network** e **Random Forest**. Valutando la curva ROC di questi modelli è possibile notare come il valore di AUC maggiore corrisponde al modello **Neural Network**, che quindi risulta essere il più performante.

# MARKET BASKET ANALYSIS

La MBA individua **regole associative** al fine di studiare le abitudini di acquisto dei consumatori, per evidenziare eventuali **affinità tra i prodotti venduti**.  
I risultati mostrano i prodotti che più spesso vengono venduti insieme.

Risulta importante per l'azienda tenere conto di questi dati e sfruttare le associazioni vincenti in eventuali **campagne marketing**.

Prime **quattro associazioni** molto rilevanti, 1132 volte i clienti hanno acquistato il prodotto 36298416 insieme al 36298353

lhs	rhs	support	confidence	coverage	lift	count
{36298416}	=> {36298353}	0.001246905	0.8608365	0.001448480	320.1592	1132
{36298395}	=> {36298353}	0.001309691	0.8591040	0.001524484	319.5149	1189
{36298206}	=> {36298122}	0.001483729	0.8173544	0.001815282	312.0410	1347
{36298381}	=> {36298353}	0.001981609	0.8513961	0.002327482	316.6482	1799
{36298381,36298416}	=> {36298353}	0.001008979	0.9318413	0.001082780	346.5671	916
{36298353,36298416}	=> {36298381}	0.001008979	0.8091873	0.001246905	347.6664	916
{36298381,36298395}	=> {36298353}	0.001047532	0.9250973	0.001132348	344.0589	951
{32078802,32079082}	=> {32079103}	0.001016690	0.8538390	0.001190728	167.7101	923
{32078802,32079103}	=> {32079082}	0.001016690	0.8315315	0.001222672	171.2966	923
{36298122,36298381}	=> {36298353}	0.001027705	0.9165029	0.001121333	340.8625	933

# AZIONI DATA-DRIVEN

## MODELLO RFM

Interagire al meglio con la clientela al fine di mantenere i clienti ad alto valore, risparmiare budget sui clienti di tipo *Cheap* o *Tin*. Specifiche azioni anche sulla clientela Bronze che è risultata essere la categoria più numerosa

## MODELLO CHURN

Diminuire il tasso di abbandono, salvaguardare il budget da campagne troppo dispendiose, ad esempio estese a tutta la clientela, prediligendo un approccio mirato ed efficace

## MODELLO MBA

Identificazione di pattern di prodotti che vengono acquistati congiuntamente e con maggiore frequenza per poter offrire una migliore e facilitata esperienza di acquisto

## CAMPAGNE PROMOZIONALI AD HOC

- Azioni di e-mail marketing a seconda della classe di appartenenza del cliente
- Creazione di offerte pertinenti e personalizzate

## STRATEGIE MARKETING

- Campagne fidelity
- Migliorare interazioni tramite social network
- Proposta di promozioni ai clienti con 'alto rischio' di abbandono (es: voucher, coupon)

## RIORGANIZZAZIONE DEI PRODOTTI

- Riorganizzazione della disposizione dei prodotti nel sito per facilitare la ricerca di quelli acquistati insieme
- Promozioni tra categorie di prodotti ad alto livello di associazione
  - Proposte di next best purchase



**GRAZIE PER L'ATTENZIONE**