

A.A. 2020-2021

Digital Marketing

Beatrice Barzaghi - 831829

Guglielmo Muoio - 826029



Business Questions

RFM

L'analisi RFM è una tecnica di segmentazione deterministica utilizzata per classificare e raggruppare quantitativamente i clienti in base a 3 fattori: recency, frequency e monetary value.

OBIETTIVO: identificare gruppi di clienti diversi e adottare strategie personalizzate.



CHURN

L'espressione "churner" indica i clienti che abbandonano il servizio e/o l'acquisto di prodotti dell'azienda.

OBIETTIVO: creare un buon modello per assegnare una probabilità di abbandono a ciascun cliente in modo tale da identificare potenziali churner e orientare le campagne di marketing



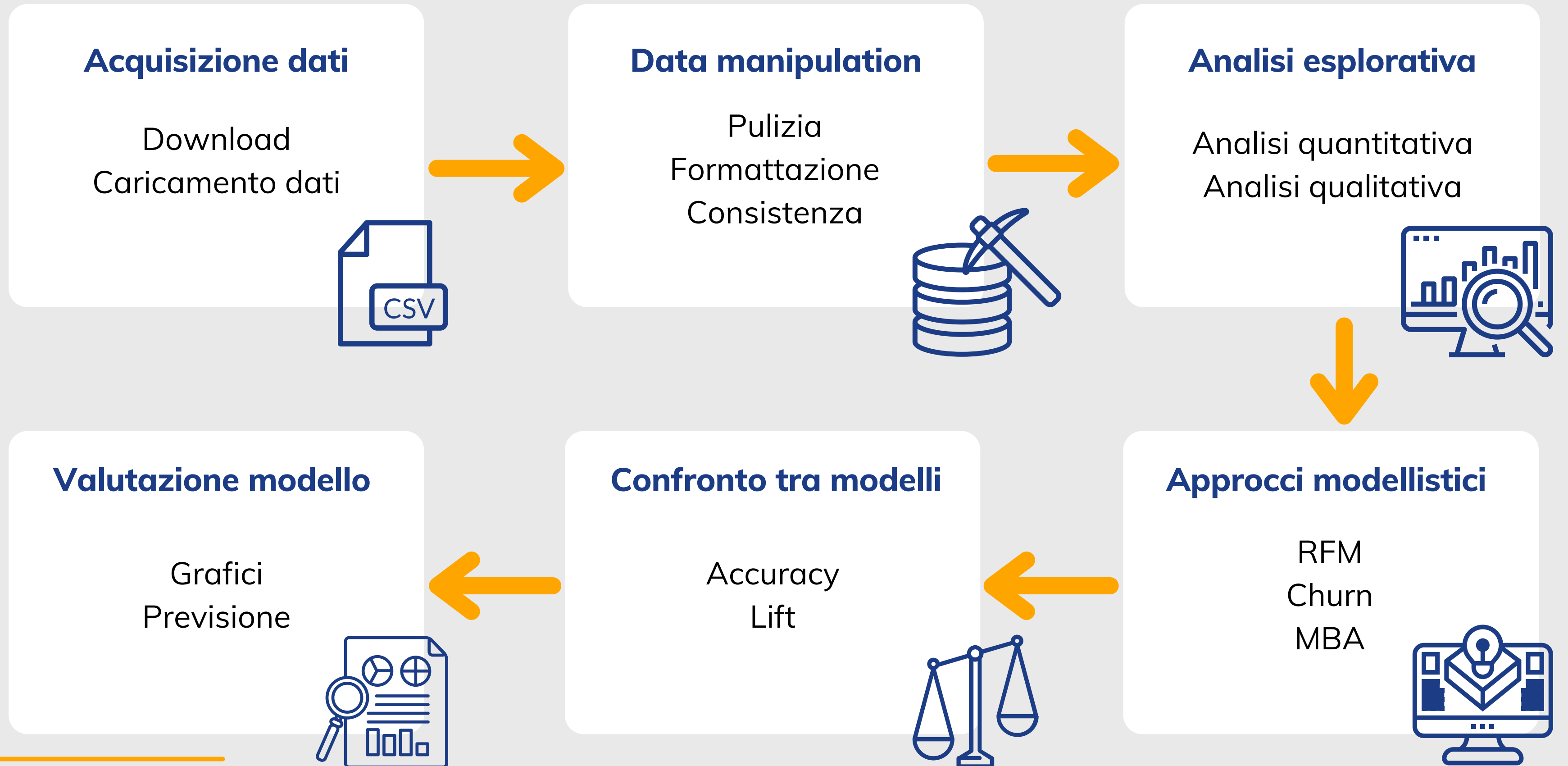
MBA

La market basket analysis (MBA) è un processo che analizza le abitudini di acquisto dei clienti, trovando associazioni tra i diversi prodotti comprati.

OBIETTIVO: identificare l'associazione tra i 100 prodotti più venduti per l'adozione di strategie di marketing ad hoc.

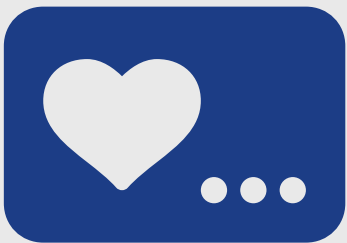


Workflow



Dataset

Fidelity



Cliente
Fedeltà
Negozio
Main Account
Tipo Fedeltà
Status
Data Attivazione

Account



Cliente
Email
Telefono
Indirizzo
Tipo Account
Professione

Address



Indirizzo
CAP
Provincia
Regione

Privacy



Cliente
Privacy
Profiling
Direct Marketing

Campaign



Campagna
Tipo Campagna

Event



Evento
Cliente
Campagna
Consegna
Tipo Evento
Data

Tickets



Scontrino
Cliente
Negozio
Articolo
Reporto
Direzio
Importo Lordo
Sconto
Data

Analisi Esplorativa

L'analisi esplorativa permette di avere una visione sulla distribuzione delle variabili e l'individuazione di eventuali relazioni presenti tra di esse, portando alla scoperta di nuove informazioni. Aiuta anche ad individuare eventuali outliers ed errori.

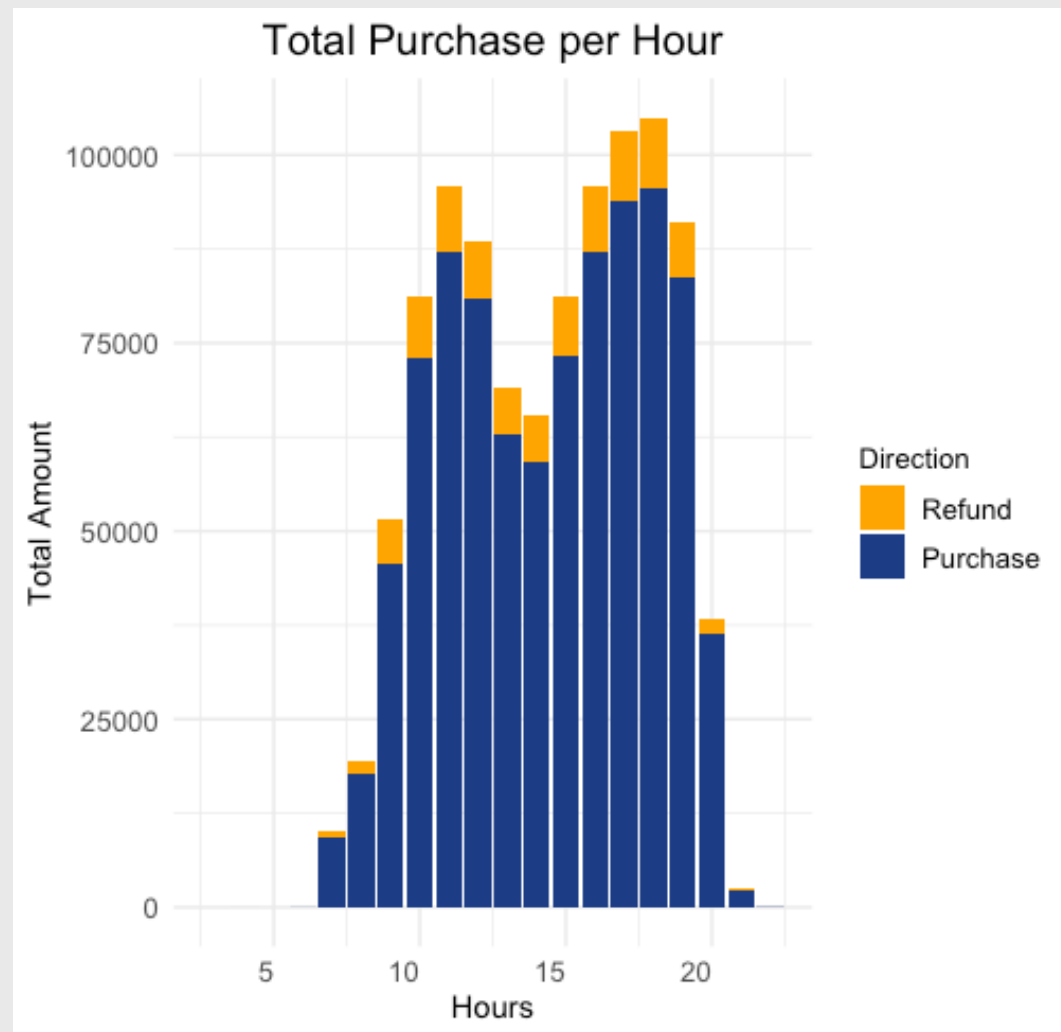


Figura 1. Distribuzione degli acquisti per ora

Prendendo in considerazione solo gli **acquisti** si può notare che le fasce orarie meno proficue sono quelle di apertura e chiusura. Il maggiore numero di acquisti avviene dalle 11.00 alle 12.00 e dalle 16.00 alle 18.00, ovvero subito prima della pausa pranzo e verso la fine della giornata lavorativa.

Considerando i negozi dove è avvenuta la prima sottoscrizione al programma fedeltà è emerso come gran parte dei clienti, circa il 16%, si sia appoggiata al negozio con ID 1. Per tutti gli altri negozi non sono presenti valori elevati di sottoscrizione da parte dei clienti.

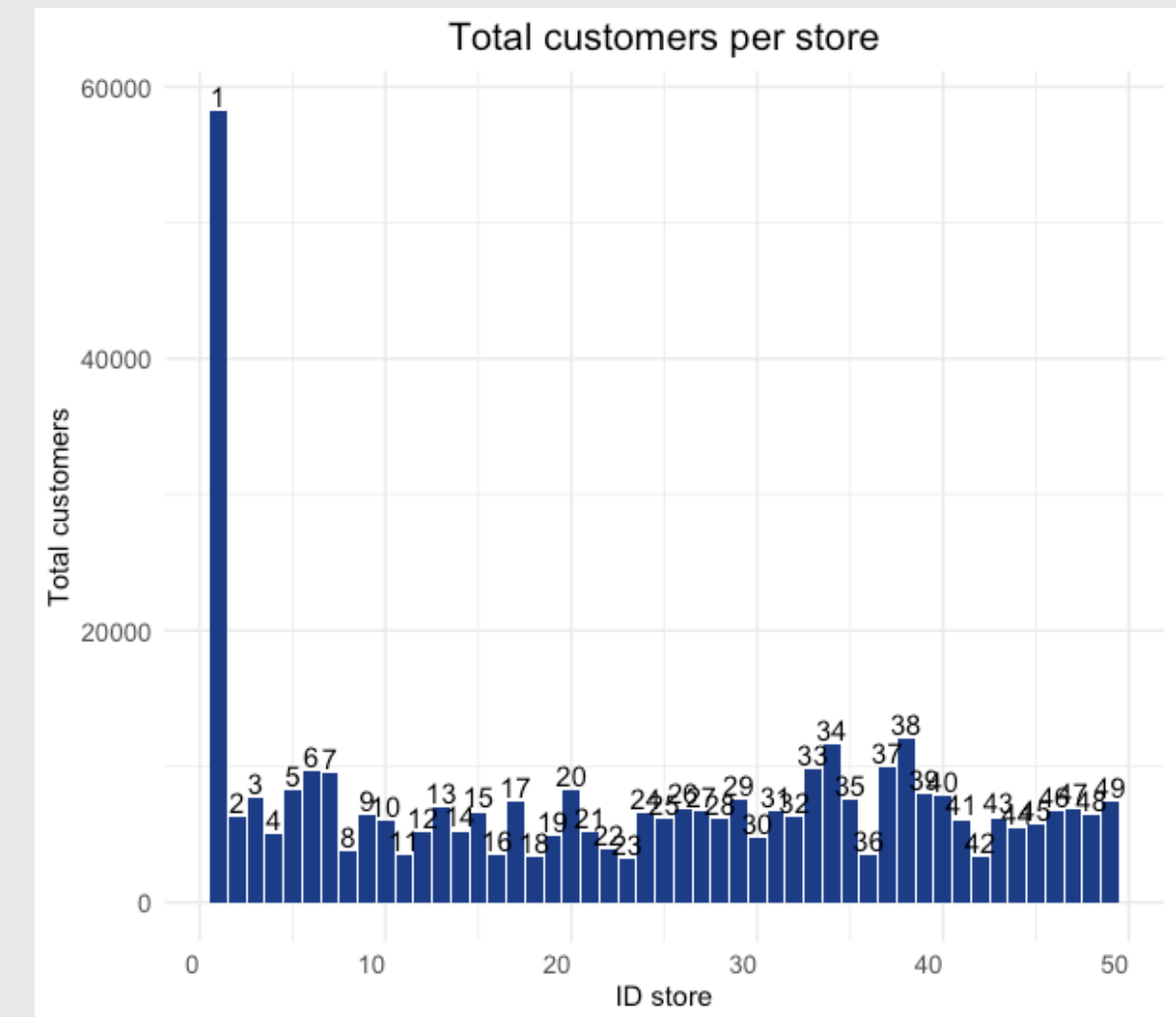


Figura 2. Distribuzione dei clienti per negozio

Modello RFM

Ottenere una descrizione del valore del cliente, creare una classificazione e una gerarchizzazione dei loro comportamenti per poter differenziare le azioni di marketing. Le metriche utilizzate sono:

- Recency ---> quando il cliente ha effettuato l'ultimo acquisto?
- Frequency ---> quanti acquisti ha effettuato il cliente in un certo periodo?
- Monetary Value ---> quanto ha acquistato il cliente in un certo periodo?

Come fare questo? Utilizzando lo storico delle transazioni: la customer base viene prima suddivisa in Attivi e Non Attivi (clienti che non hanno acquistato da 4 mesi) e successivamente i primi vengono ulteriormente divisi in base ai quantili di ognuna delle 3 metriche.

| Classe | Recency | Frequency | Monetary |
|--------|---------------|------------------|---------------|
| Low | < 25% | < 2 purchase | < 25% |
| Medium | 25% < x < 75% | 2 < purchase < 5 | 25% < x < 75% |
| High | > 75% | purchase > 5 | > 75% |

Tabella 1. Divisione per quantili delle tre metriche

Per ottenere la suddivisione finale si procede prima attraverso l'utilizzo dei dati di Recency e di Frequency, solo in seguito verrà aggiunta la misura Monetary. È possibile osservare che la classe con meno clienti è Leaving Top (più di 5 acquisti ma nessuno recente), mentre quella con un numero totale di clienti molto elevato è Engaged (tra 2 e 4 acquisti effettuati abbastanza recentemente).

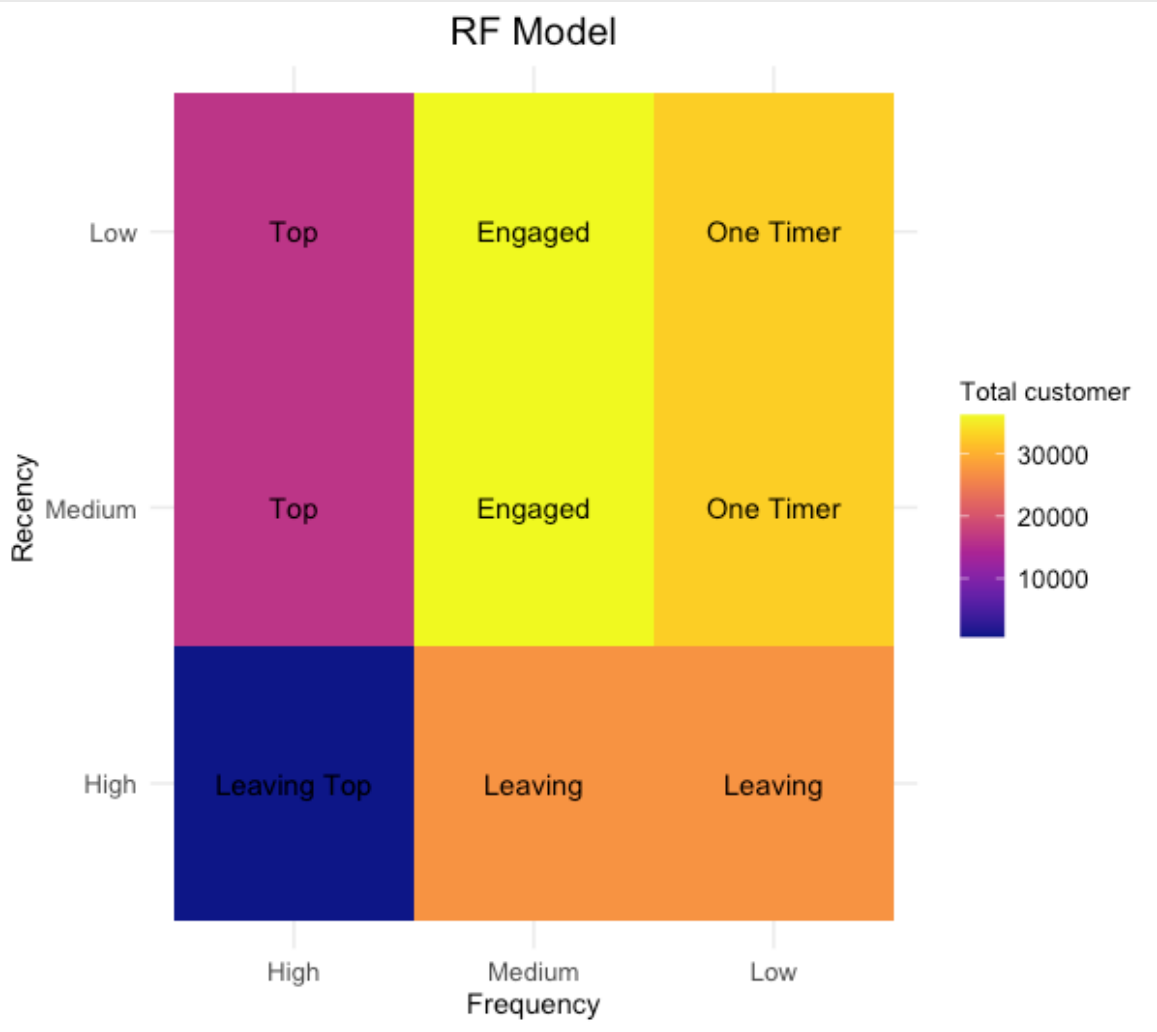


Figura 3. Matrice RF

Modello RFM

Le classi ottenute al passo precedente vengono combinate con i differenti gruppi di *Monetary* per ottenere le 7 classi definitive che vanno a caratterizzare lo status di un cliente. Grazie a queste categorie è possibile proporre delle azioni di marketing *ad hoc* per ciascun cliente e valutare l'eventuale migrazione di questi tra i diversi segmenti (sfruttando periodicamente la valutazione RFM). Tutte queste informazioni permettono l'ottimizzazione degli investimenti marketing.

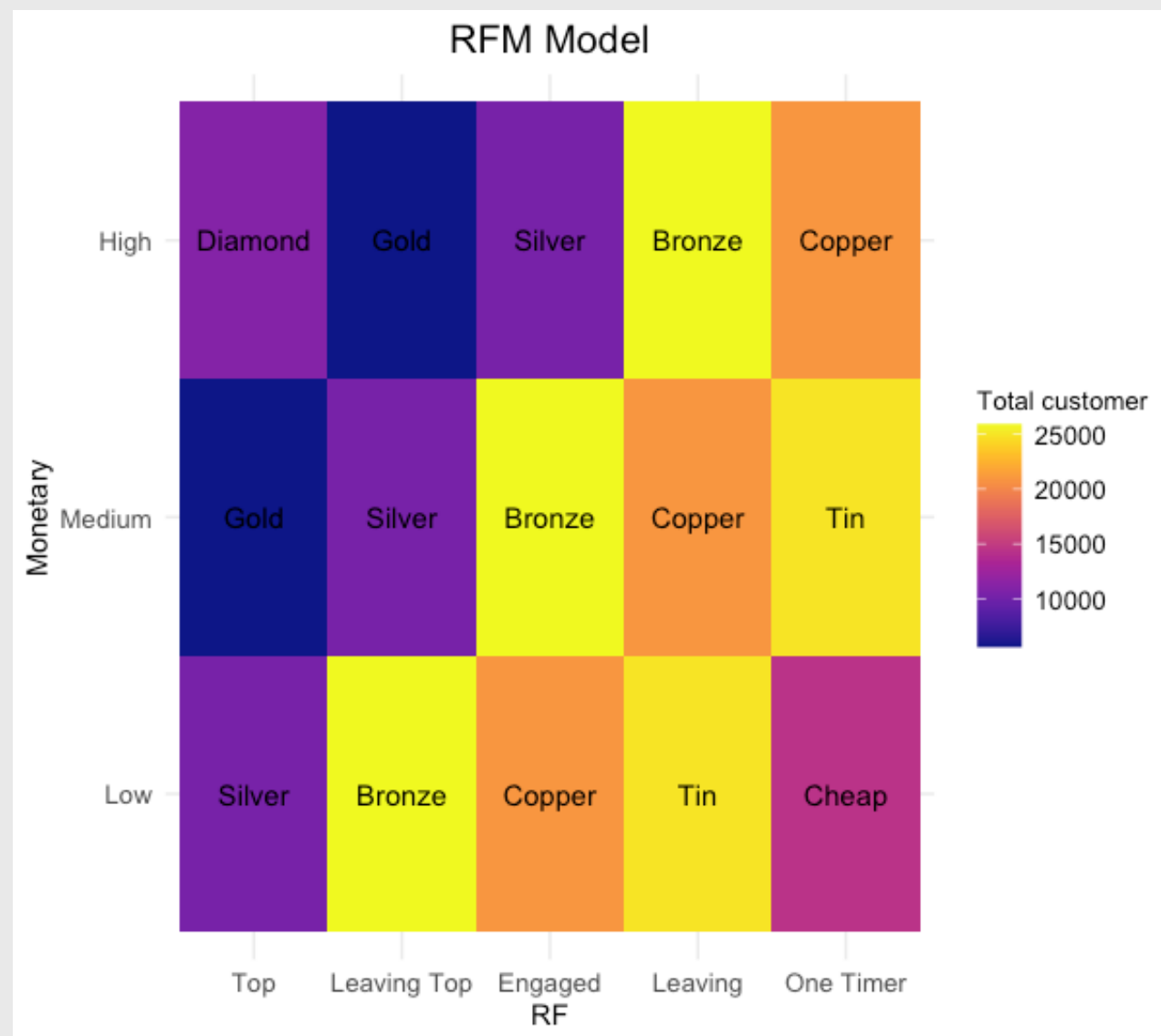


Figura 4. Matrice RFM

Si nota dalla matrice RFM come i clienti *Gold* siano i meno frequenti al contrario di quelli di classe *Bronze*.

L'RFM permette di incrementare la customer retention, il tasso di risposta o il tasso di conversione e, a seconda dei propri obiettivi, si può delineare un piano opportuno.

Potenziali strategie:

- Investire di più in campagne marketing volte al mantenimento di clienti *Diamond* e *Gold*
- Identificare possibili azioni marketing per incrementare il valore dei clienti *Silver* e *Bronze*
- Risparmiare budget per le classi di clienti a basso valore: *Cheap* e *Tin*

Modello Churn

L'obiettivo del modello di churn consiste nell'assegnare a ciascun cliente la sua probabilità di abbandono. Sarà così possibile implementare specifiche azioni di marketing correttive al fine di trattenere i clienti a più alto valore.

Come fare questo? Attraverso la costruzione di un *propensity supervised model* in cui viene sfruttata principalmente la frequenza relativa all'acquisto da parte del cliente.

Step necessari alla costruzione del modello:

1. scelta di una reference date (01/01/2019)
2. imporre la lunghezza del periodo di holdout (01/01/2019-28/02/2019)
3. scelta del periodo di lookback (01/10/2018 - reference date)
4. assegnamento della variabile target (0=no churn, 1=churn) ad ogni cliente che si è rilevato *churner* nel periodo di holdout
5. definizione dei predittori da utilizzare

Si ottiene che tra il periodo di lookback ed il periodo di holdout i valori di *churn* e di *no churn* sono i seguenti

| No Churn - 0 | Churn - 1 |
|--------------|-----------|
| 35 373 | 59 752 |

Tabella 2. Divisione in No Churn e Churn

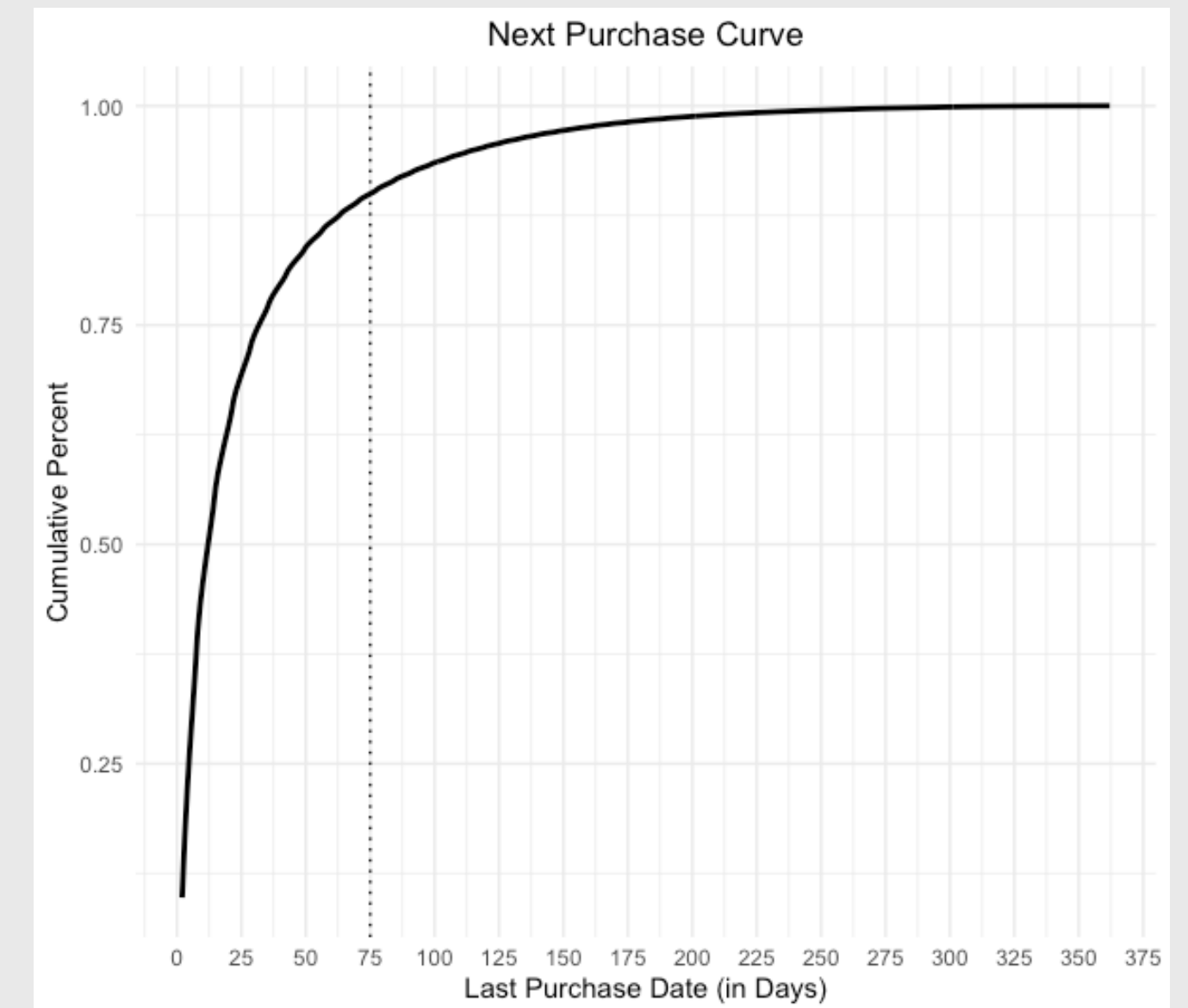


Figura 5. Next Purchase Curve

La *Next Purchase Curve* evidenzia come il 90% dei clienti abbia effettuato almeno un acquisto negli ultimi 75 giorni.

Modello Churn

Per approcciarsi ai modelli in maniera corretta è necessario dividere il dataset in *Train* (70%) e *Test* (30%). Il primo viene utilizzato per imparare il modello mentre il secondo viene utilizzato per misurare i livelli di accuratezza attraverso procedura *holdout*.

Variabili esplicative utilizzate:

- *Recency* - giorni dall'ultimo acquisto
- *Spesa* - ammontare totale degli acquisti effettuati in euro
- *Tot purchase* - numero di acquisti effettuati
- *Region* - regione di residenza del cliente
- *Type job* - tipo di lavoro del cliente

Modelli di *machine learning* implementati:

- Recursive Partitioning And Regression Trees
- Random Forest
- Logistic Regression
- Lasso

Il modello migliore è il *Logistic Regression* con un livello di accuracy pari a 69.2%. In particolare questo garantisce un livello di *specificity* del 90.1%, dovuto alla corretta classificazione di 15411 *churner* su un totale di 17093.

Analisi *lift curve*: evidenza di quanto nel sottogruppo selezionato si prevede meglio il fenomeno di *churn* rispetto a quello che si otterrebbe dalla popolazione totale. Sfruttando il modello migliore e considerando il 70% dell'intero dataset, si è in grado di prevedere correttamente la variabile *churn* all' 83.4%.

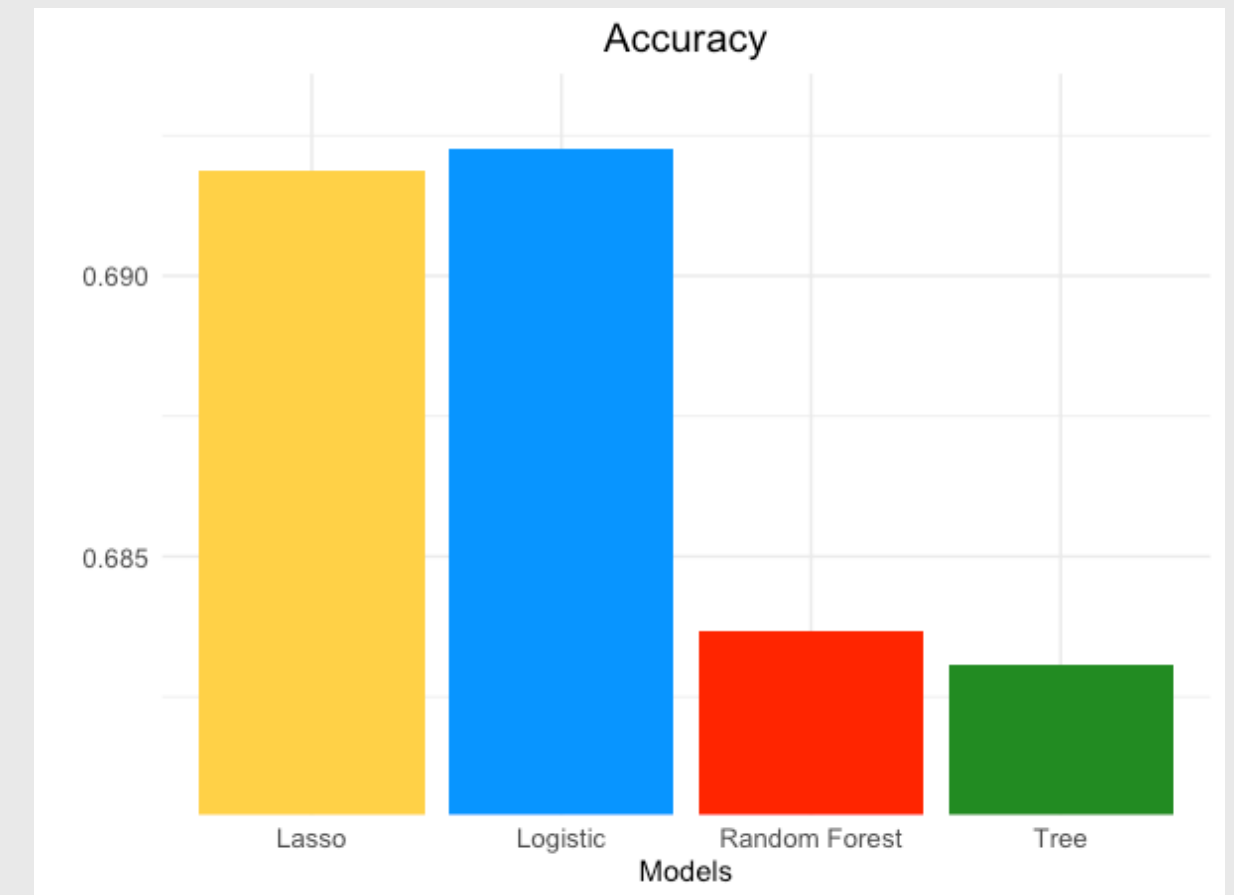


Figura 6. Accuracy dei modelli

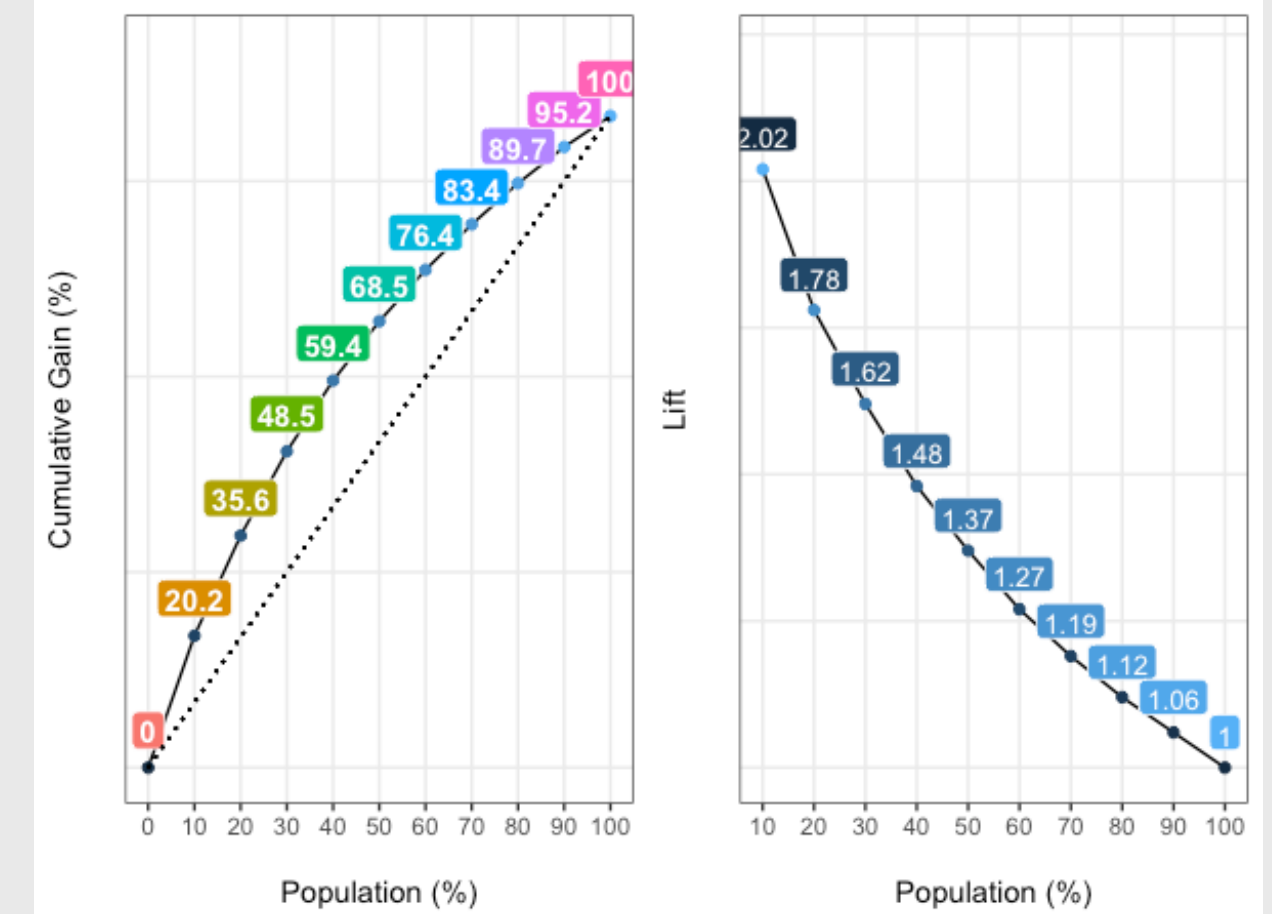


Figura 7. Gain e Lift del modello Logistico

Market Basket Analysis

La market basket analysis (MBA) ha l'obiettivo di individuare regole di associazione tra prodotti che appaiono congiuntamente all'interno delle transazioni, identificando pattern di prodotti che vengono acquistati con maggiore probabilità. Consente di identificare prodotti "trainanti", ovvero item che portano all'acquisto di altri prodotti.

Applicando la MBA è possibile:

- creare dei bundle promozionali di prodotti,
- arricchire la contact strategy attraverso il "next best product to buy"
- ottimizzare la collocazione dei prodotti

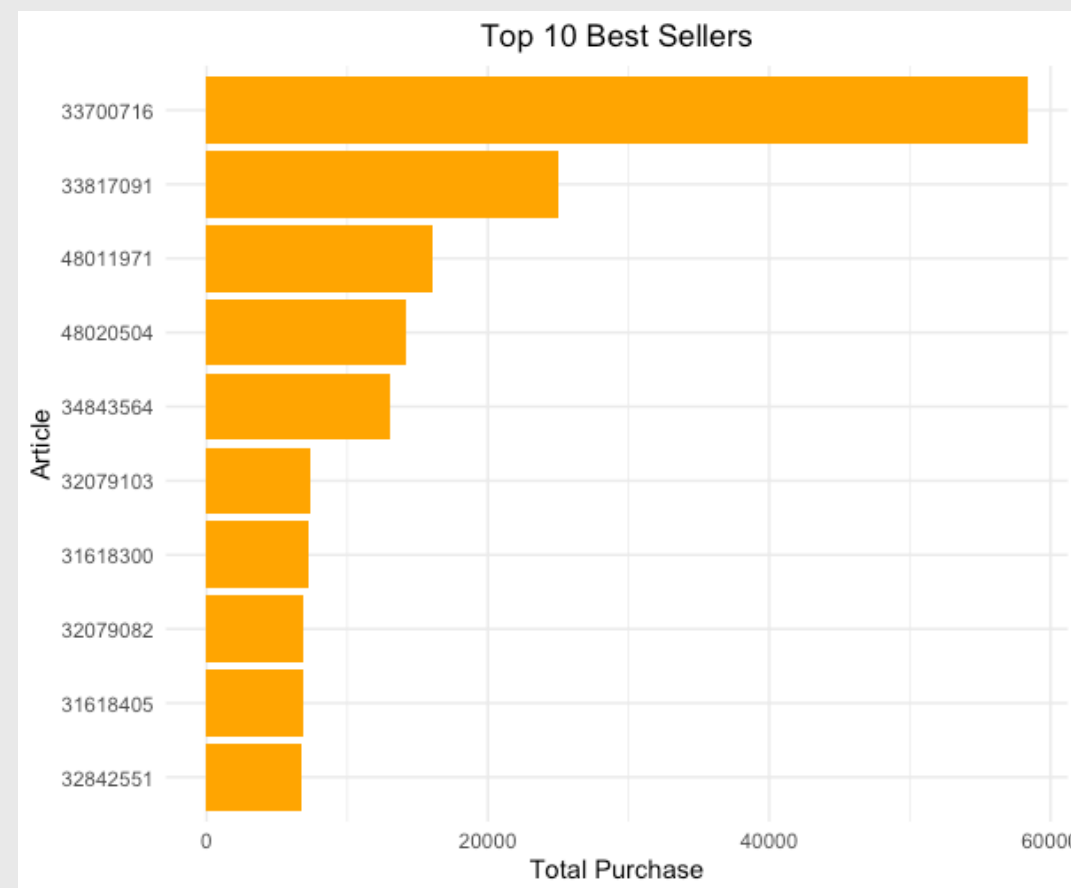


Figura 8. Prodotti più venduti

Il prodotto più venduto ha codice 33700716, con un totale di circa 60 000 unità vendute.

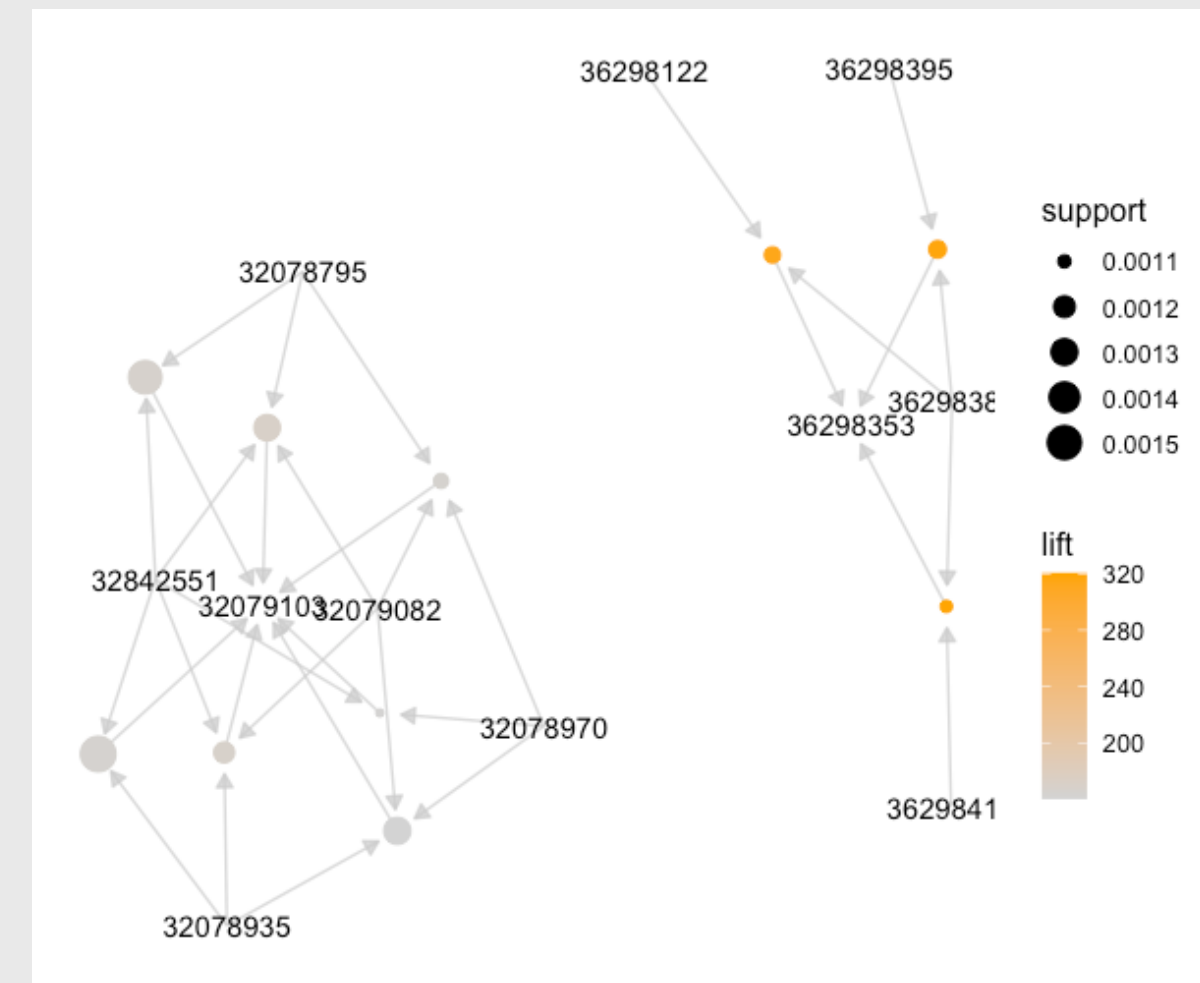


Figura 9. Associazione tra prodotti

Association Rules con una confidenza dell'80% e con supporto almeno dello 0.1% . È possibile osservare che il 95% dei clienti che acquista i prodotti {32078795, 32079082, 32842551} acquistano anche il prodotto 32079103.

Azioni Data-driven

MODELLO RFM

Interagire al meglio con la clientela al fine di trattenere i clienti ad alto valore, non investire eccessivamente in clienti del tipo *Cheap* o *Tin*

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

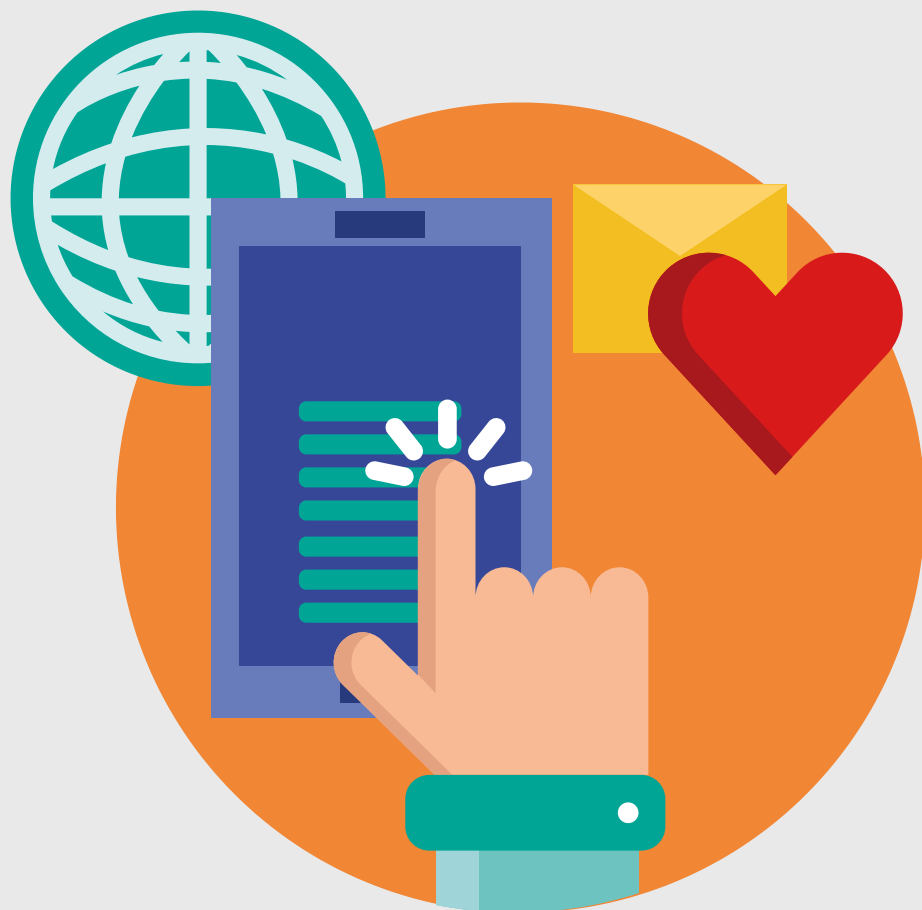
- campagne fidelity
- migliorare interazioni tramite social network
- proposta di promozioni ai clienti ad "alto rischio"



RIORGANIZZAZIONE PRODOTTI

- riorganizzazione della disposizione dei prodotti
- promozioni tra categorie di prodotti ad alto livello di associazione
- proposte di next best purchase





**Grazie per
l'attenzione**

