



Digital Marketing

Andrea Maver - 828725

Anna Nava - 827719

Indice

Domande di business

Database

Esplorazione dei dati

RFM Model

Churn Model

Market Basket Analysis




Domande di Business



VALORE DEI CLIENTI

- Quali sono i clienti con maggior valore?

Per rispondere è stato utilizzato il modello **RFM**, basato su:

- **Recency**: il cliente quanto recentemente ha acquistato?
 - **Frequency**: quanto frequentemente acquista?
 - **Monetary**: quanto spende?
- 

FEDELTÀ DEI CLIENTI

- Qual è la probabilità di churn dei clienti?
- Quali, fra i probabili churning, hanno maggior valore?
- Quali strategie adottare per non perderli?

Si è ricorsi ai modelli di Churn, studiando il modello migliore.

MIGLIORAMENTO PUNTI VENDITA

- Quali sono i prodotti maggiormente venduti assieme?
- Come distribuire i prodotti negli scaffali per ottimizzare le vendite?

È stata svolta una Market Basket Analysis.

Database

Sette dataset (due relativi alle campagne marketing).

Programmi Fedeltà

- Negozio attivazione
- Date attivazione
- Programma fedeltà
- Numero sottoscrizioni

Account

- Provider account email
- Aggiunta cellulare
- Tipo account
- Lavoro cliente

Indirizzi

- CAP
- Provincia
- Regione

Privacy

- Flag privacy
- Flag profiling
- Flag marketing diretto

Campagne marketing

- Tipo campagna
- Numero click/open/inviati falliti

Transazioni

- Cliente e negozio di riferimento;
- Acquisto o rimborso;
- Costo netto.

Variabili aggiunte

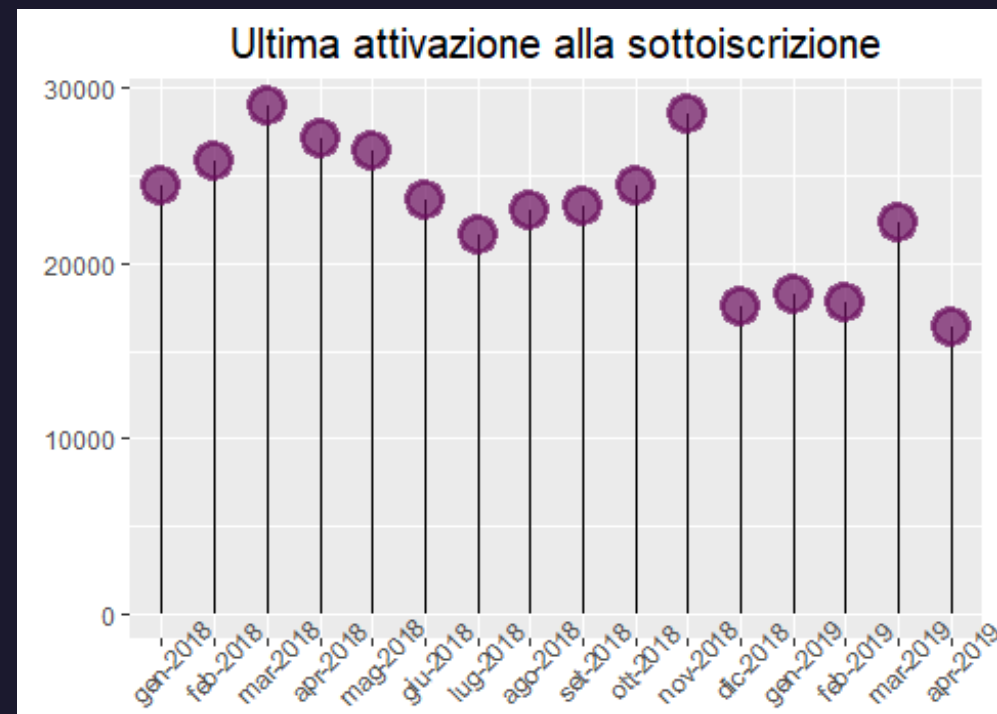
- Spesa totale per ciascun cliente;
- Numero acquisti;
- Ricezione mail da campagne marketing.

Esplorazione dati: clienti e sottoscrizioni

Per orientarsi nella grande quantità di dati a disposizione, è opportuno svolgere delle analisi esplorative preliminari; in seguito vengono mostrate alcune delle più interessanti.



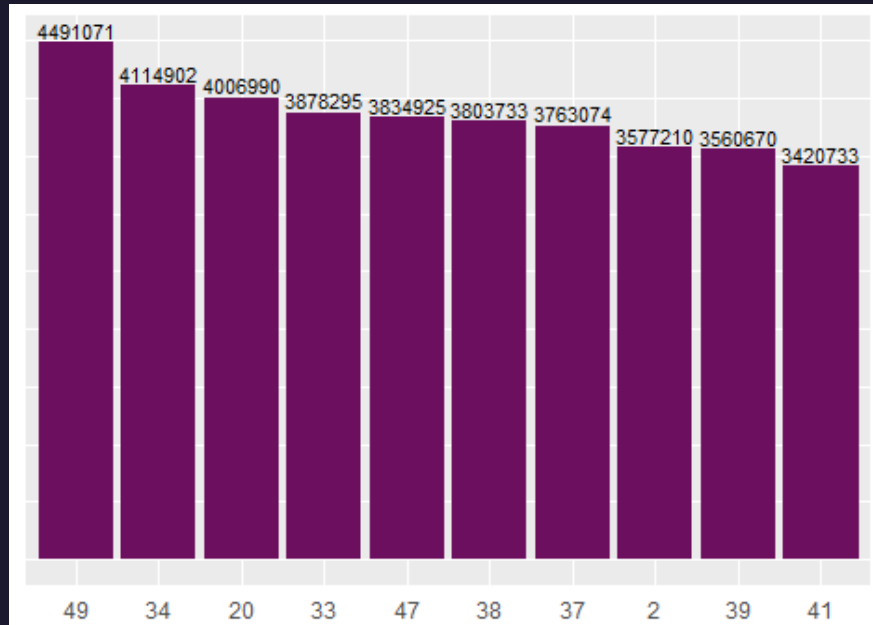
Utilizzando i dati sugli indirizzi dei clienti se ne può visualizzare facilmente la provenienza: una netta maggioranza risiede in Lombardia, seguita da Lazio e Campania.



È raffigurato il numero di sottoscrizioni mensili, considerando solo l'ultima attivazione per ciascun cliente. Si osserva che negli ultimi mesi, le sottoscrizioni sono diminuite significativamente.

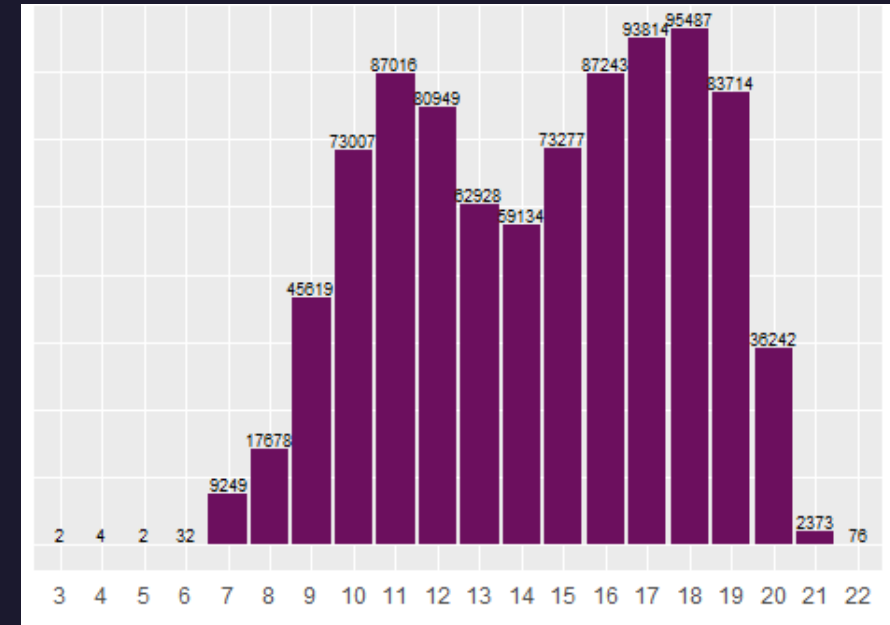
Esplorazione dati: transazioni

GUADAGNI NETTI PER NEGOZIO



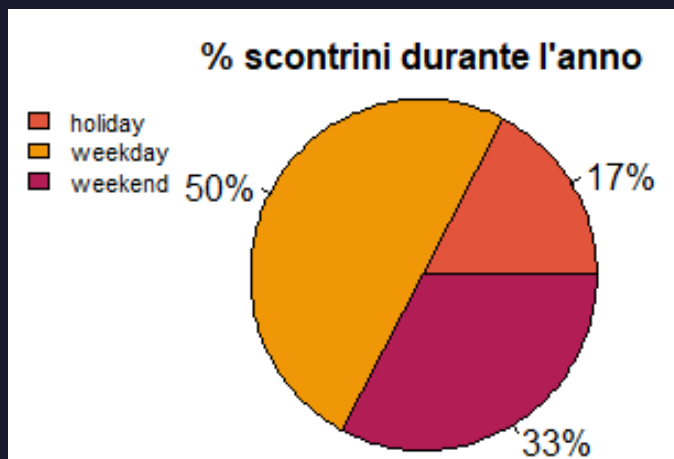
In ordine decrescente, sono raffigurati i guadagni netti per i 10 negozi con i guadagni maggiori.

SCONTRINI DURANTE LA GIORNATA



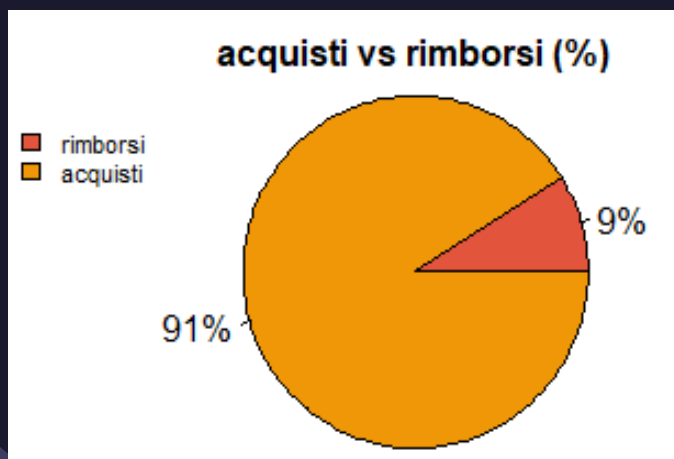
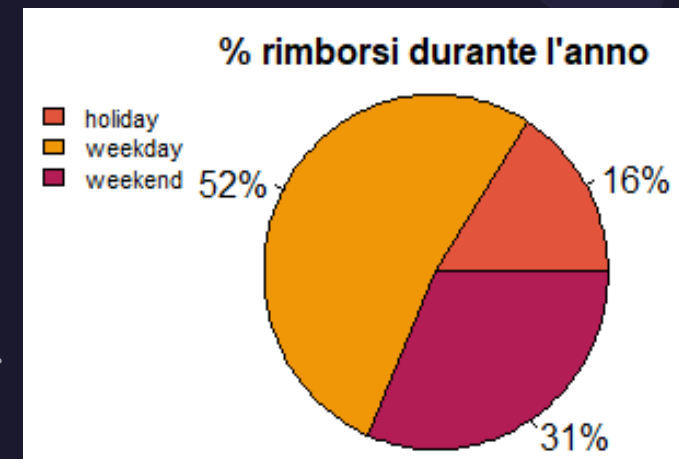
Il numero maggiore di acquisti si concentra in tarda serata (17-18) e in tarda mattinata (11-12). Di primo pomeriggio e passate le 21 (orari di apertura e chiusura) gli acquisti calano nettamente.

Esplorazione dati: acquisti e rimborsi



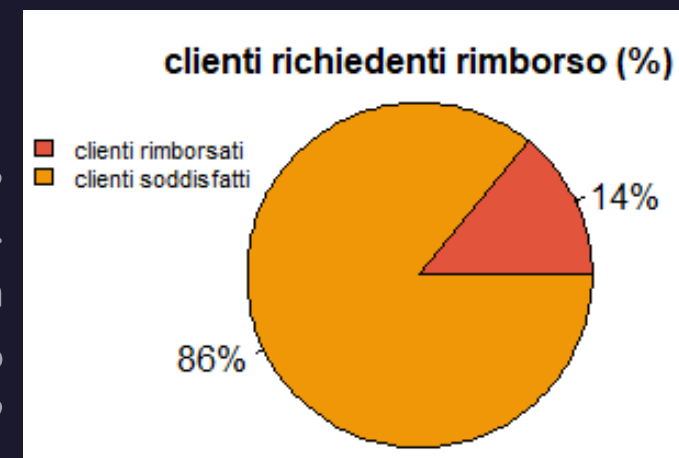
Ben il 50% degli acquisti si ha nelle festività e nei weekend.

I rimborsi si distribuiscono in maniera analoga.



Solo il 9% degli scontrini sono dei rimborsi.

Nonostante ciò,
ben il 14% dei clienti ha ottenuto dei rimborsi.
Solo errori di cassa o ci sono articoli in
particolare di cui viene frequentemente chiesto
il rimborso?



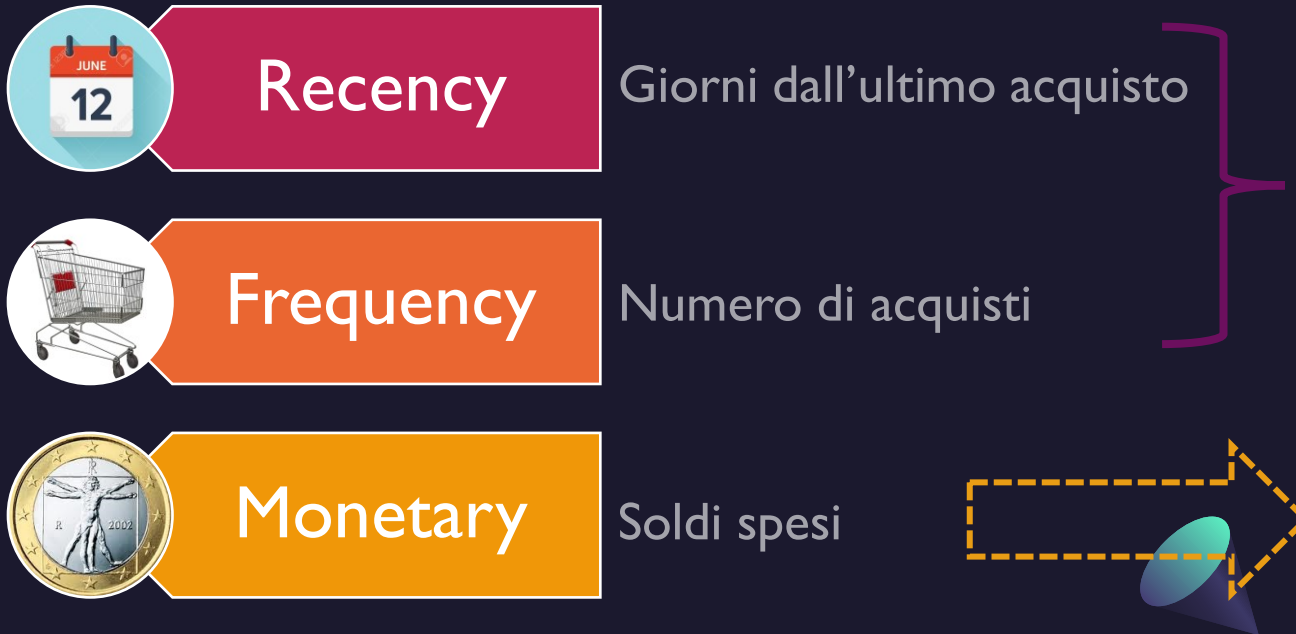
Modello RFM

Questa analisi permette di individuare i clienti con maggior valore.

Prima di tutto, si escludono i clienti *inattivi* – i clienti che non hanno acquistato negli ultimi 160 giorni.

I 160 giorni corrispondono al numero di giorni dopo cui riacquista il 90% dei clienti non casuali – i clienti che hanno fatto acquisti in più di un giorno.

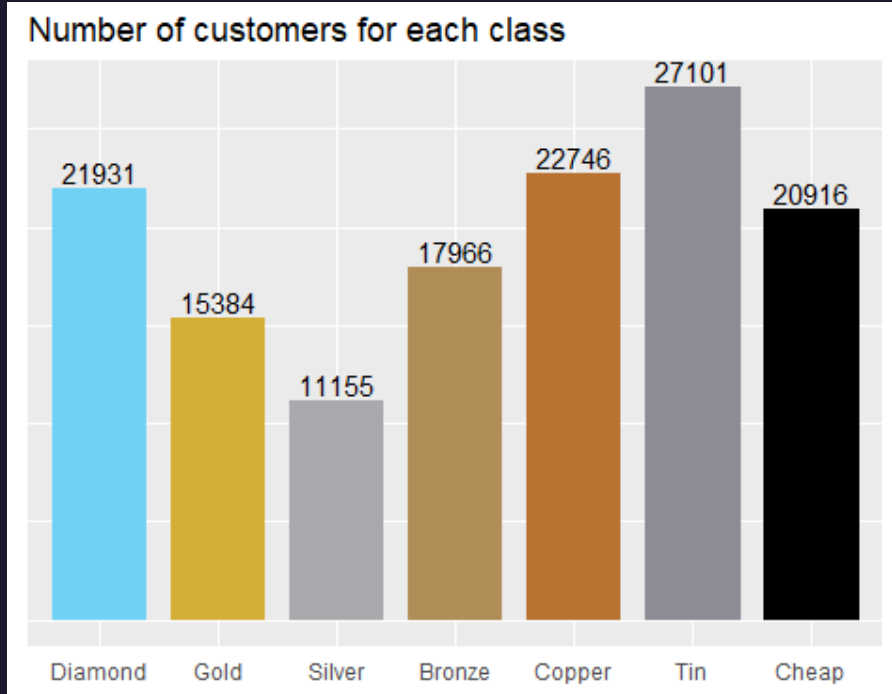
Segmentiamo i clienti attivi in base a:



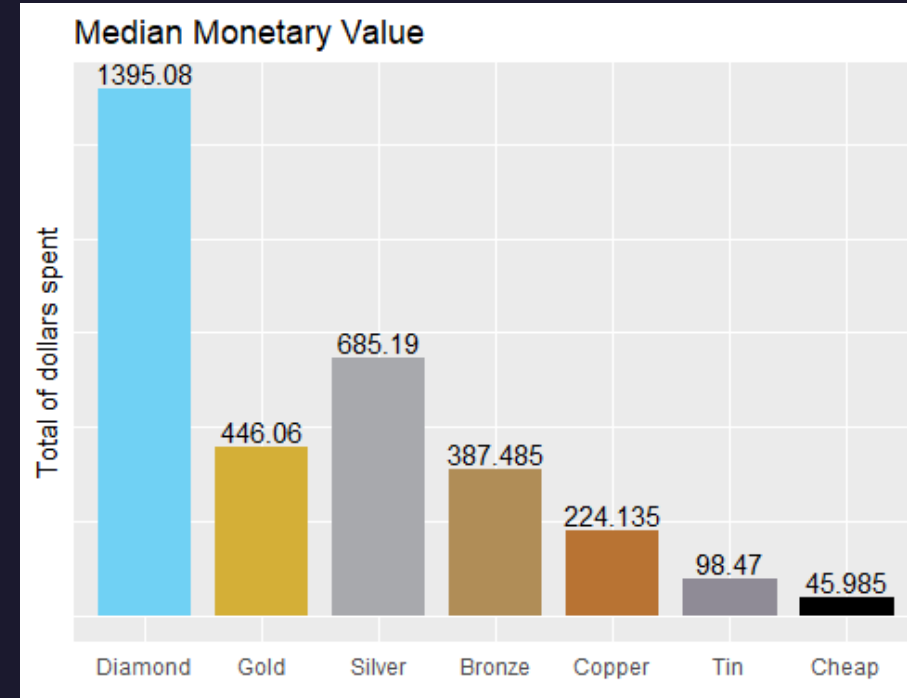
In base a **Recency** e **Frequency**, suddividiamo in cluster che corrispondono alla *fedeltà* dei clienti.
In ordine decrescente:
Top, Leaving Top, Engaged, Leaving, One-Timer.

Infine, in base a **Monetary**, suddividiamo i cluster in sette categorie di clienti:
Diamond, Gold, Silver, Bronze, Copper, Tin e Cheap.

Risultati RFM



Il 15,98% dei clienti è Diamond, percentuale da non sottovalutare perché piuttosto consistente. I clienti con più valore (Diamond, Gold, Silver) corrispondono al 35,33% del totale.



Si nota una particolarità: il valore monetario dei clienti Gold è piuttosto basso; significa che acquistano frequente ma con spese contenute. Per quanto riguarda i clienti Tin e Cheap è evidente che non convenga investire nei loro confronti.

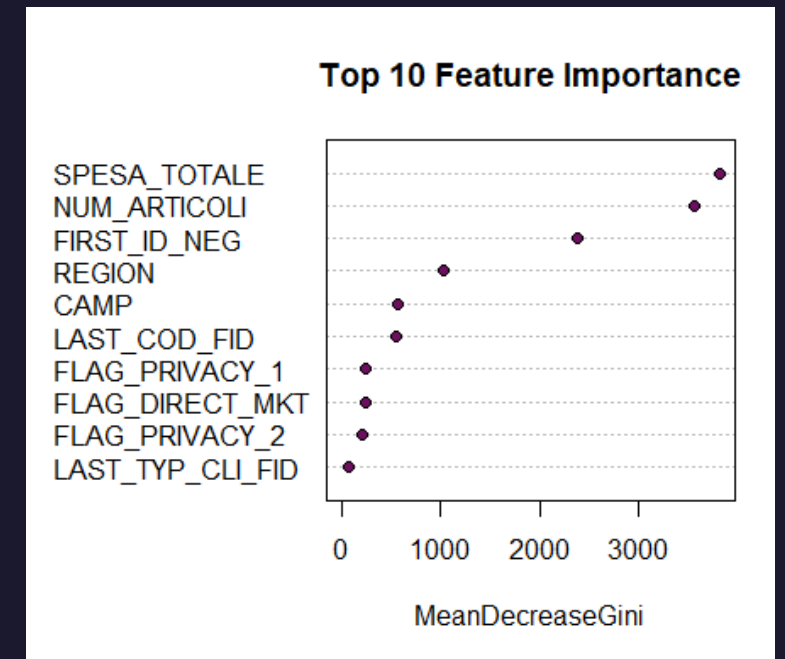
Modello di churn 1/3

Questa analisi permette di individuare i clienti con la maggiore probabilità di abbandonare l'azienda, studiandone alcuni parametri. Identificare i churmer permette di attuare delle campagne marketing personalizzate.

Il primo passo consiste nel determinare il periodo di tempo dopo il quale un cliente viene considerato churmer; dopo aver scelto una data di riferimento, si individua il numero di giorni a ritroso per cui il 90% dei clienti effettua una nuova transazione. Il numero di giorni individuato è 160, per cui i clienti che 160 giorni prima della data di riferimento non hanno effettuato acquisti verranno considerati churmer.

È ora necessario scegliere le variabili da utilizzare nei modelli. Dopo diversi test, vengono scelte le tre più significative, cioè:

- SPESA_TOTALE: somma di tutti gli scontrini di un cliente
- NUM_ARTICOLI: numero di prodotti acquistati per cliente
- FIRST_ID_NEG: primo punto vendita di riferimento



Modello di churn

2/3

I classificatori utilizzati per l'individuazione dei churner sono Logistic Regression, Decision Tree, Random Forest e Multilayer Perceptron.

Il dataset originale viene diviso in training set (80%) per fare apprendere i modelli e test set (20%) per testarne i risultati.

A causa degli scarsi risultati iniziali dovuti allo sbilanciamento del dataset, situazione in cui una classe è molto più frequente di quella che interessa individuare, viene applicata la tecnica dell'undersampling, che consiste nel campionare dal dataset originale meno osservazioni con valore 0, in modo da ottenere un nuovo dataset che presenti uguale distribuzione tra churner e non.

Distribuzione churner originale	
Non churner - 0	Churner - 1
74.0 %	25.9 %

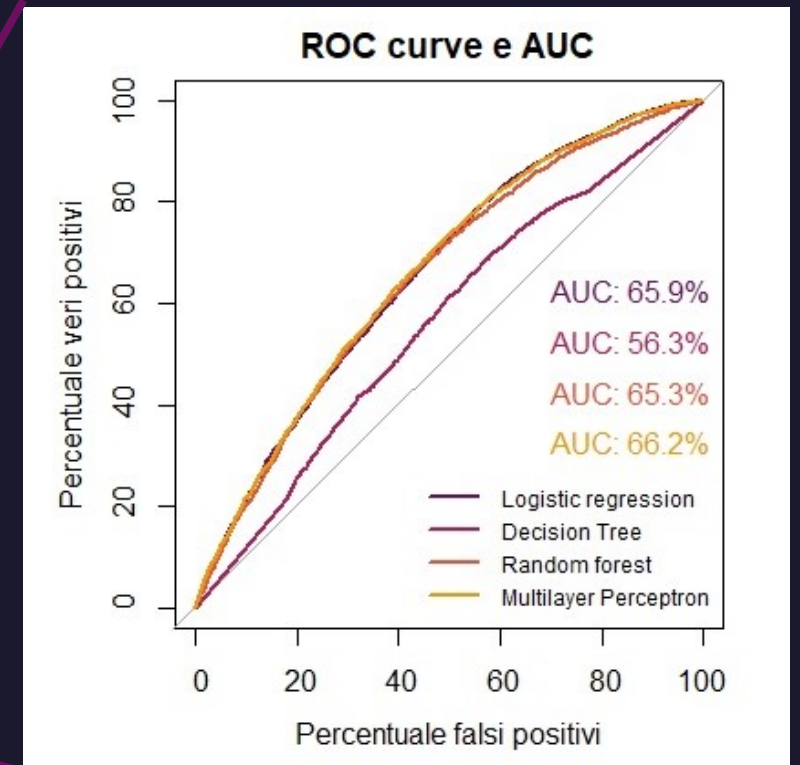
Distribuzione churner post undersampling	
Non churner - 0	Churner - 1
50.0 %	50.0 %

Modello di churn 3/3

I modelli forniscono i risultati in tabella. È importante osservare il valore di Recall: in questo studio assume un significato più importante di quello della Precision perché perdere un cliente di grande valore sarebbe più costoso che acquisirne uno poco profittevole.

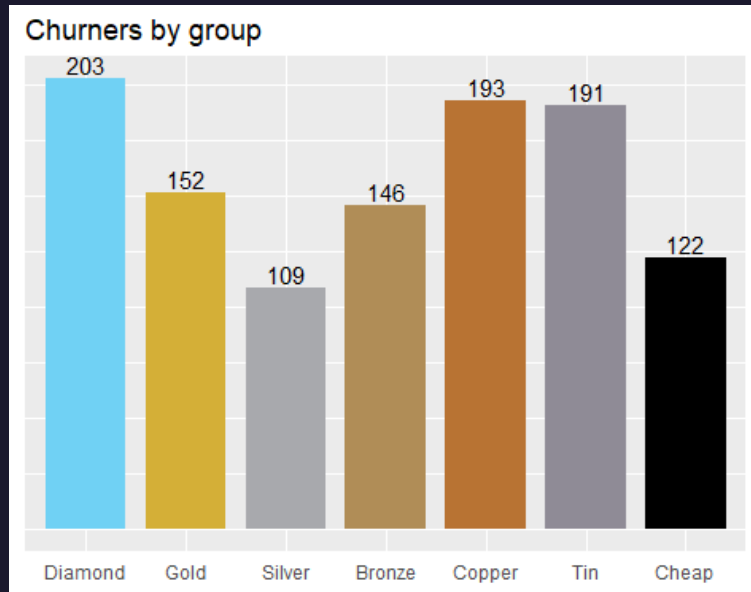
Il classificatore migliore risulta essere il **Multilayer Perceptron**, che, come indicato dal valore AUC, individua correttamente il 66.2% dei churner

	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.616	0.596	0.723	0.653
Decision Tree	0.552	0.551	0.563	0.557
Random Forest	0.613	0.600	0.680	0.637
Multilayer Perceptron	0.620	0.598	0.731	0.658



Churn e RFM

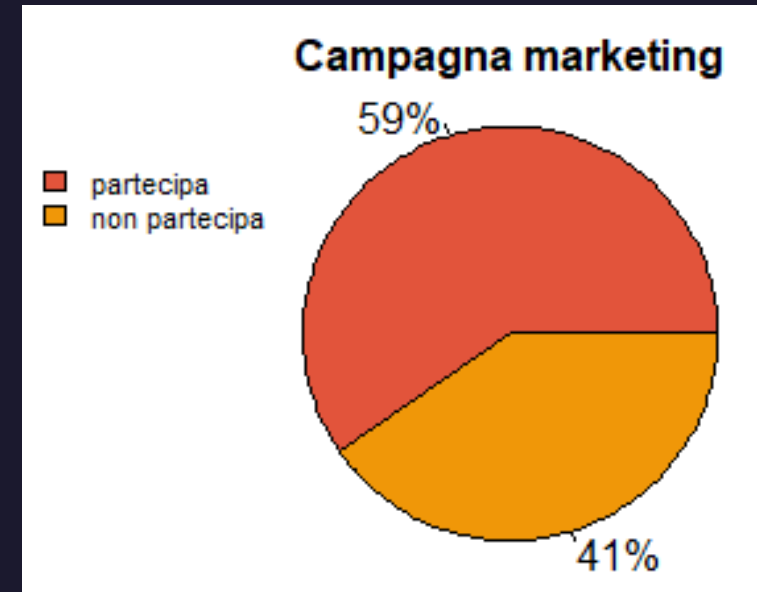
A questo punto, in particolare, si vogliono individuare i clienti con alta probabilità di churn e con maggior valore. Per questo, si confrontano i risultati ottenuti con il Multilayer Perceptron con i precedenti del modello RFM.



Un grande numero di clienti ad alto valore sono probabili churner.

In particolare, la maggioranza dei churner predetti rientra nella categoria Diamond.

I risultati sono, pertanto, significativamente negativi.

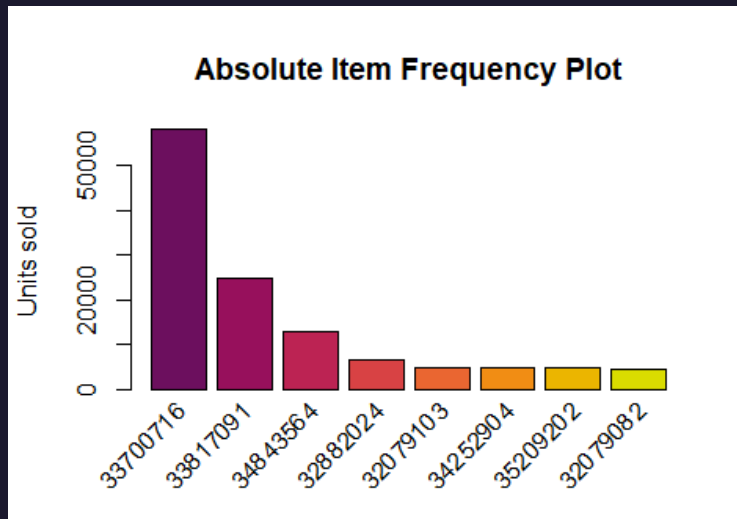


Fra i clienti con più valore (Diamond, Gold, Silver), quanti rientrano in almeno una campagna marketing? Ben il 59%. Sembra essere pertanto opportuno, prima ancora che investire in nuove campagne marketing, studiare gli errori di quelle passate.

Market Basket analysis

La market basket analysis permette di scoprire associazioni tra prodotti che appaiono nelle stesse transazioni e che vengono acquistati insieme. Lo scopo dell'analisi è quello di scoprire legami tra prodotti per poterne sfruttare la profittabilità e creare particolari disposizioni sugli scaffali o bundle di offerte.

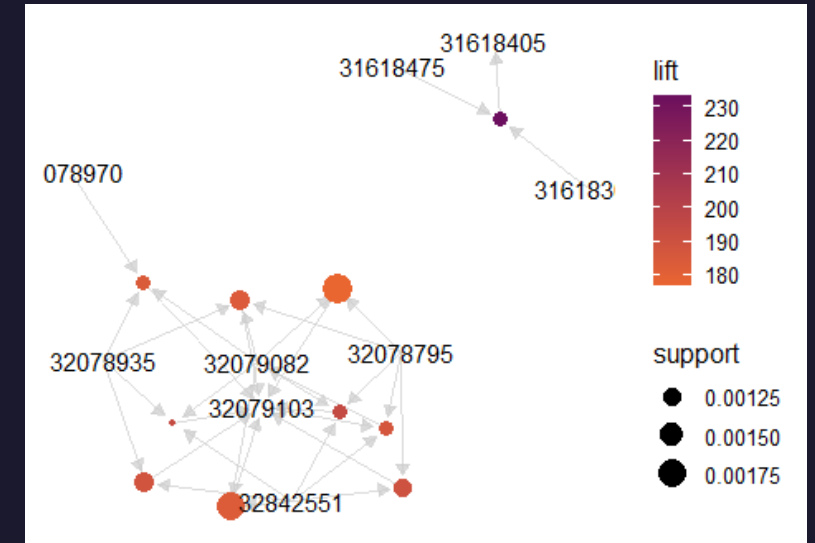
Applicando la MBA si osserva che di quasi un milione di transazioni analizzate, circa il 30% sono costituite da un solo item e del prodotto più venduto, con codice 33700716, ne sono state acquistate poco meno di sessantamila unità.



Gli otto prodotti più acquistati

Association rules: utilizzando support di 0.001 e confidenza di 0.8 vengono generate 26 regole.

Ad esempio, il 92% dei clienti che acquista i prodotti 32078935 e 32842551 acquista anche 32079103



Network delle regole di associazione



Grazie per l'attenzione

Andrea Maver – a.maver@campus.unimib.it

Anna Nava – a.nava38@campus.unimib.it