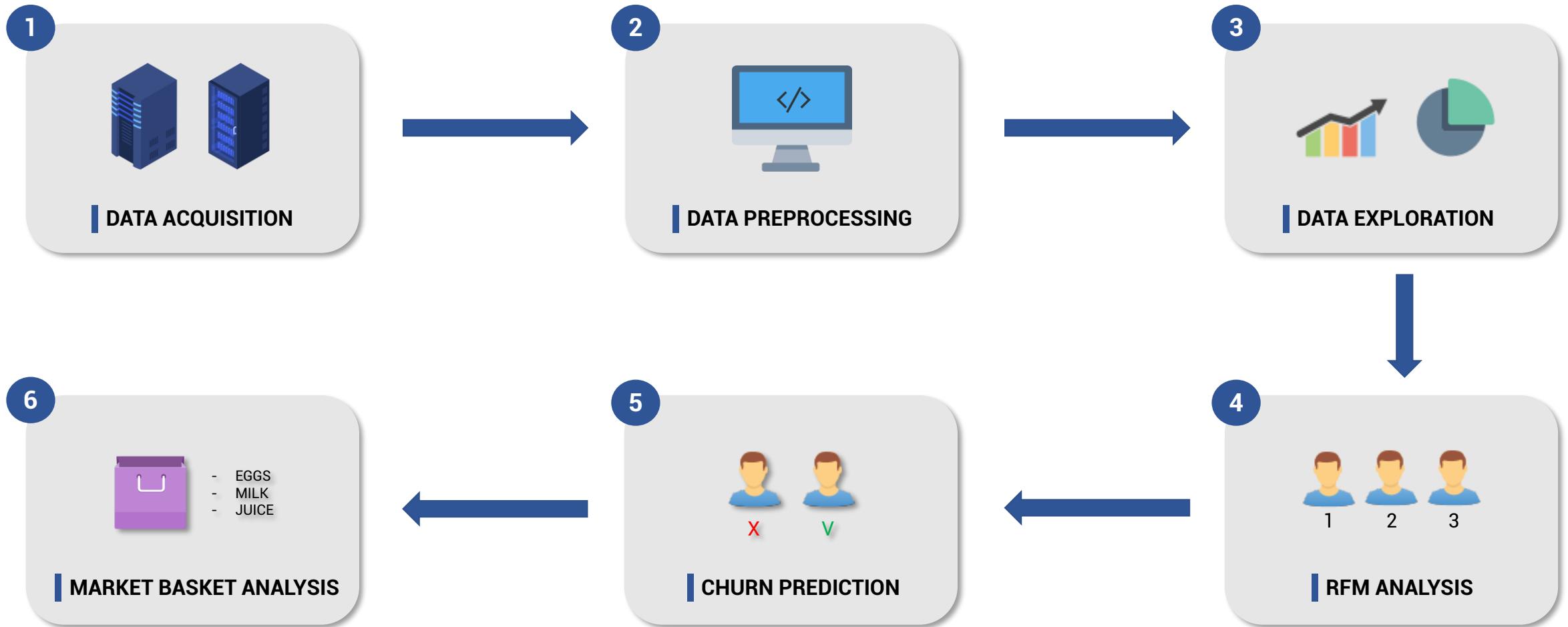


Laurea Magistrale in DATASCIENCE



Digital Marketing
Federico De Servi – 812166
AA 2019-2020

PIPELINE



OBIETTIVI

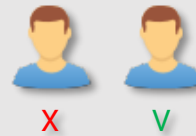


RFM ANALYSIS

L'analisi **RFM** (Recency, Frequency, Monetary) è uno dei metodi più utilizzati nella segmentazione dei propri clienti. L'esigenza a cui risponde è quella di creare liste di clienti differenti ed assegnarne un diverso.

Essa si basa su 3 principali fattori:

- **Recency**: la frequenza di acquisto di un cliente in un periodo determinato di tempo.
- **Frequency**: il numero totale di acquisti
- **Monetary**: l'ammontare monetario di tutti gli acquisti che ha generato il cliente in un determinato periodo di tempo



CHURN PREDICTION

Per tasso di **churn** (anche detto churn rate) si intende la **misura percentuale di perdita di clienti** (o altri soggetti di analisi) in un determinato **arco temporale**.

Quello che un'azienda vuole fare è costruire un **modello** per cercare di predire la probabilità con cui un cliente effettuerà «churn» o meno, in un determinato arco temporale.

Questo è reso possibile grazie a tecniche di machine learning di vario genere.



MARKET BASKET ANALYSIS

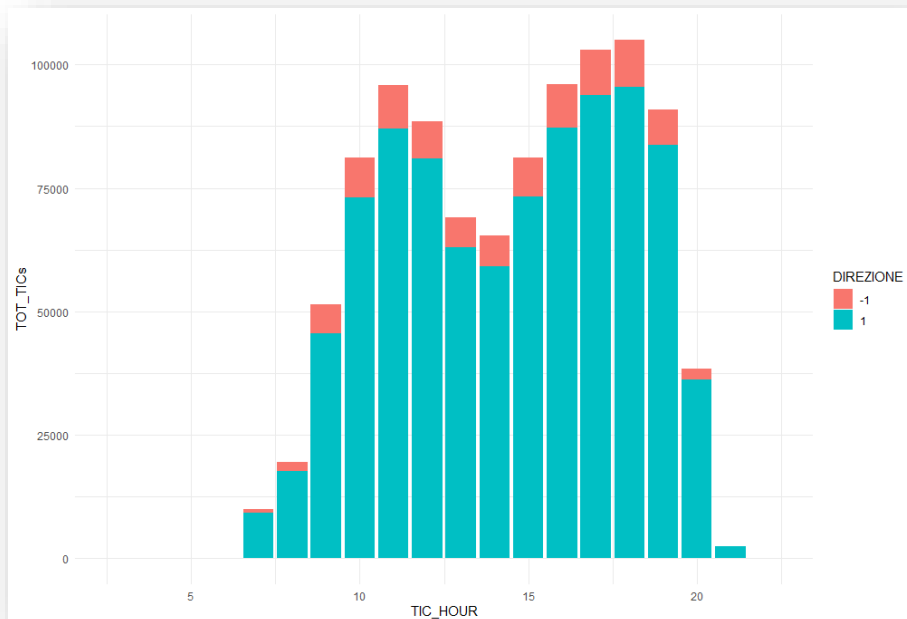
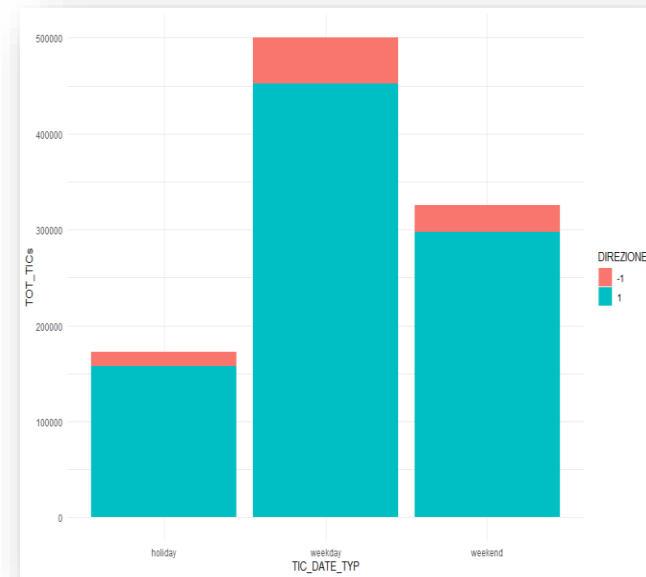
La **Market Basket Analysis** è uno strumento di data mining basato sull'individuazione di **regole associative**, che permette di studiare le abitudini di acquisto dei consumatori per evidenziare eventuali affinità esistenti tra i prodotti venduti.

Lo scopo è quello di trovare **pattern** ricorrenti nei comportamenti di acquisto dei consumatori. Questi pattern risultano di **interesse fondamentale** per l'azienda. Basti pensare, per esempio, quanto sia importante conoscere quali prodotti vengono più spesso acquistati insieme.

ANALISI ESPLORATIVA

Durante l'**analisi esplorativa** si possono scoprire molti **insights** interessanti, specialmente in ottica business.; ne riportiamo alcuni.

Nel grafico a destra possiamo vedere come la maggior parte degli acquisti si concentri nei giorni feriali, seguiti a debita distanza dai giorni del weekend. Pochi acquisti vengono effettuati durante i giorni di vacanza.



Nel grafico di sinistra invece vediamo come si distribuiscano gli acquisti durante le varie ore del giorno. Vediamo quindi che abbiamo un picco alle ore 11 e uno alle ore 17-18.

RFM ANALYSIS

Per un'azienda è importante **suddividere i clienti** in base alla loro importanza. Per questo motivo si calcola un punteggio da 1 a 5 (calcolati in base alla distribuzione totale dei valori: vedi sotto) per ciascuna delle 3 misure fondamentali: **recency**, **frequency** e **monetary**. In base ai valori ottenuti, ciascun cliente viene fatto rientrare in uno dei seguenti gruppi.

Segment	Description	R	F	M
Champions	Bought recently, buy often and spend the most	4 - 5	4 - 5	4 - 5
Loyal Customers	Spend good money. Responsive to promotions	2 - 5	3 - 5	3 - 5
Potential Loyalist	Recent customers, spent good amount, bought more than once	3 - 5	1 - 3	1 - 3
New Customers	Bought more recently, but not often	4 - 5	<= 1	<= 1
Promising	Recent shoppers, but haven't spent much	3 - 4	<= 1	<= 1
Need Attention	Above average recency, frequency & monetary values	2 - 3	2 - 3	2 - 3
About To Sleep	Below average recency, frequency & monetary values	2 - 3	<= 2	<= 2
At Risk	Spent big money, purchased often but long time ago	<= 2	2 - 5	2 - 5
Can't Lose Them	Made big purchases and often, but long time ago	<= 1	4 - 5	4 - 5
Hibernating	Low spenders, low frequency, purchased long time ago	1 - 2	1 - 2	1 - 2
Lost	Lowest recency, frequency & monetary scores	<= 2	<= 2	<= 2

Fonte: <https://cran.r-project.org/web/packages/rfm/vignettes/rfm-customer-level-data.html#segments>

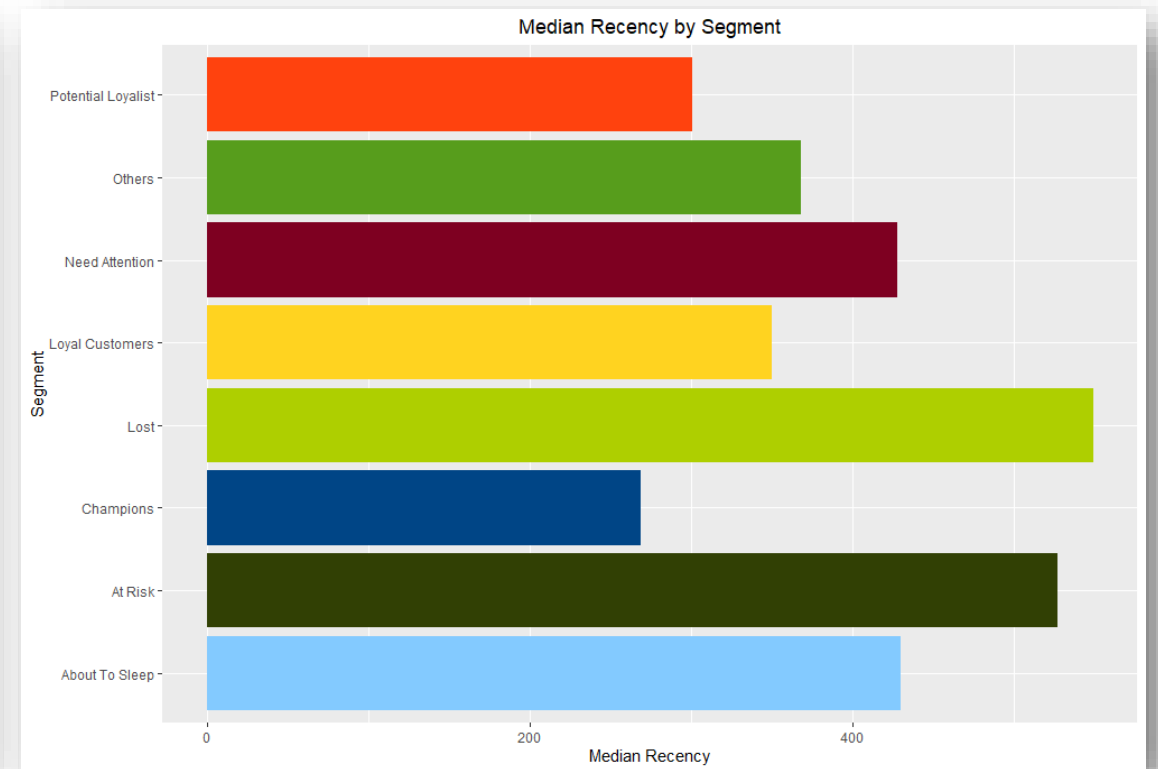
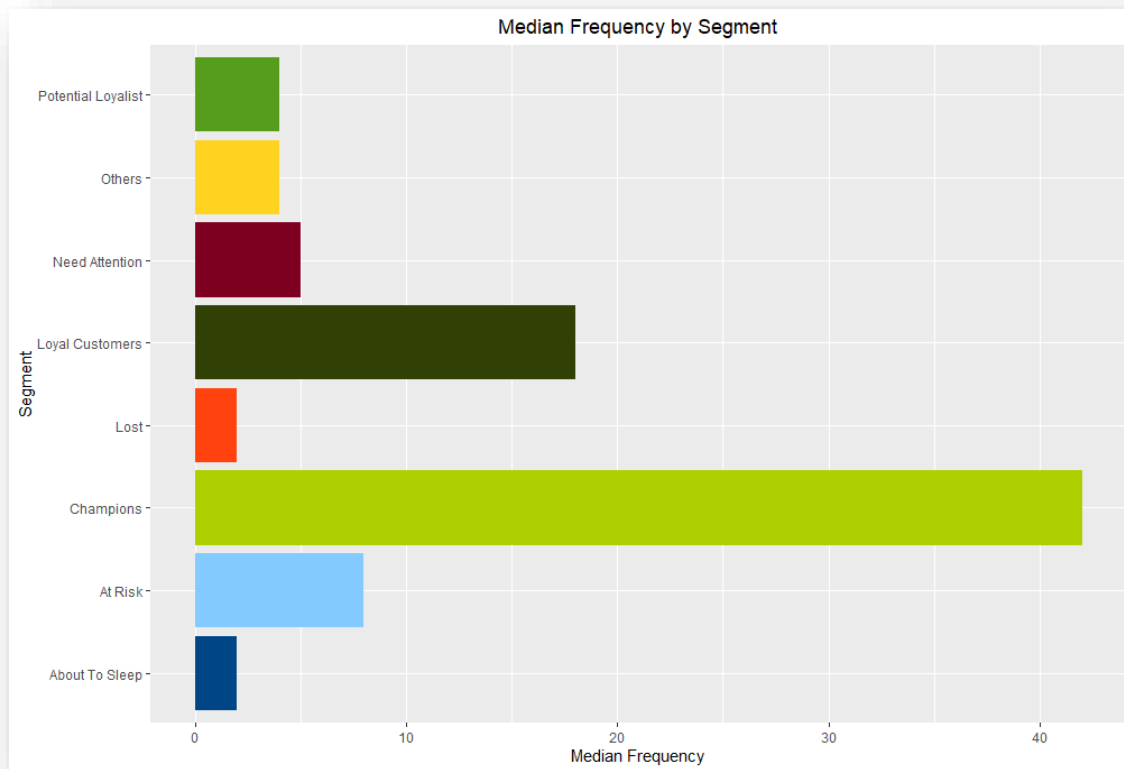
Bin	Recency	Frequency	Monetary
1	< 20%	< 20%	< 20%
2	20% < X < 40%	20% < X < 40%	20% < X < 40%
3	40% < X < 60%	40% < X < 60%	40% < X < 60%
4	60% < X < 80%	60% < X < 80%	60% < X < 80%
5	> 80%	> 80%	> 80%

	Segment	Count
1	Loyal Customers	52509
2	Potential Loyalist	45277
3	Champions	33130
4	At Risk	26033
5	Others	21569
6	Lost	13869
7	About To Sleep	11126
8	Need Attention	7423

In questa tabella si riportano tutti i gruppi trovati nella clientela in analisi ed il numero di clienti appartenenti a ciascun gruppo. Come vediamo, i **3 gruppi più numerosi** sono «loyal customers», «potential loyalist» e «champions». Tuttavia si nota anche che abbiamo un considerevole numero di clienti a rischio di essere persi («at risk»).

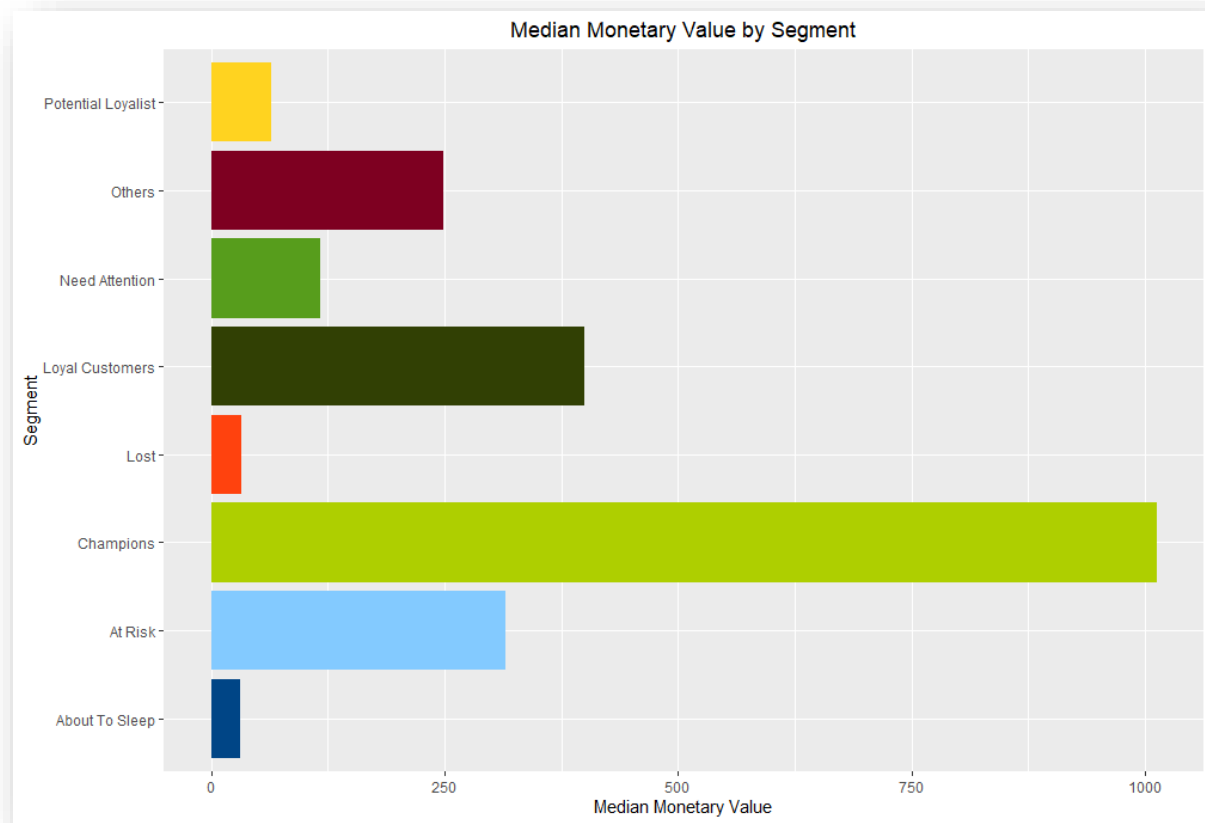
RFM ANALYSIS

Dopo aver effettuato l'analisi e la divisione dei vari clienti nelle rispettive ed adeguate tipologie, possiamo calcolare alcune caratteristiche di tali gruppi, quali la frequency mediana, la recency mediana e la monetary mediana per ogni gruppo.

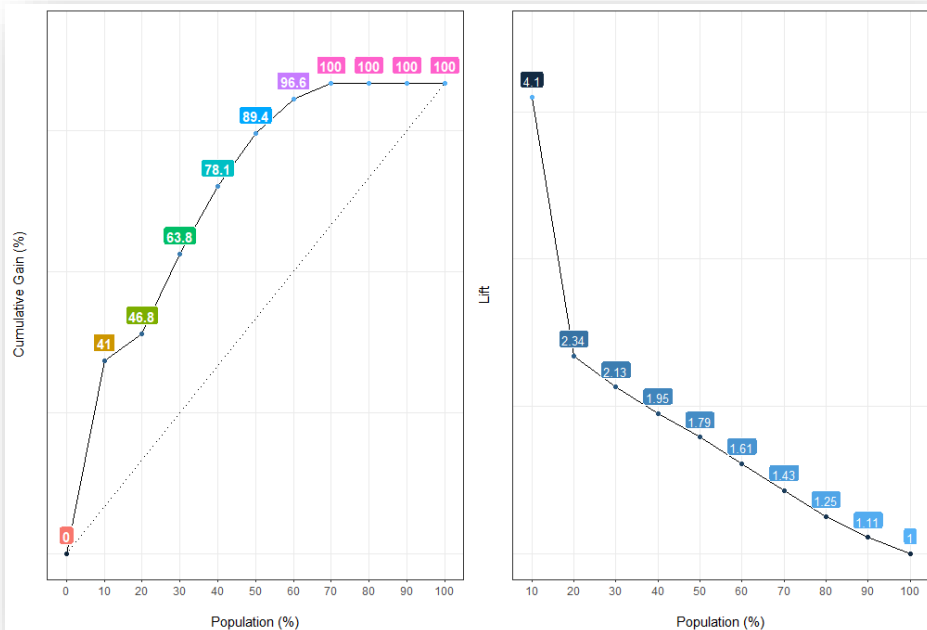


RFM ANALYSIS

In questo grafico vediamo quali siano i valori mediani di monetary per ciascuno dei gruppi ottenuti con l'analisi RFM. Come ci aspettiamo, più si sale in importanza tra i vari gruppi, più la monetary mediana aumenta.



CHURN MODELING



Models	Accuracy
Random Tree	0.8552307
Random Forest	0.8541877
Logistic	0.7437737

Per tasso di **churn** (anche detto churn rate) si intende la **misura percentuale di perdita di clienti** (o altri soggetti di analisi) in un determinato **arco temporale**.

Si vuole quindi costruire un **modello** per cercare di predire la probabilità con cui un cliente effettuerà «**churn**» o meno, in un determinato arco temporale.

Per fare questo, dopo aver applicato il preprocessing, suddividiamo il dataset in **Train** (70%), ovvero il dataset usato per allenare l'algoritmo, e **Test** (30%), usato per testare l'accuratezza del modello ottenuto.

Le **variabili esplicative** usate per la costruzioni dei modelli sono state:

- Il segmento di appartenenza
- Il numero di transazioni totale
- Il numero di giorni dall'ultima transazione
- Il totale speso dal cliente

I **modelli** usati sono stati:

- Random forest
- Logistic Regression
- Random Tree

Il modello migliore è risultato essere la **Random Tree**, con una **accuracy pari a 85.52%**. A sinistra riportiamo i grafici che mostrano il Lift e il Gain relativo al Random Tree.


MARKET BASKET ANALYSIS

Come abbiamo anticipato, la **Market Basket Analysis** è uno strumento di data mining basato sull'individuazione di **regole associative**, che permette di studiare le abitudini di acquisto dei consumatori per evidenziare eventuali affinità esistenti tra i prodotti venduti.

Tramite questa analisi possiamo vedere quali sono i prodotti che vengono più spesso venduti insieme. Questo risulta di particolare importanza per l'azienda, in quanto potrà studiare campagne di marketing mirate a sfruttare queste associazioni.

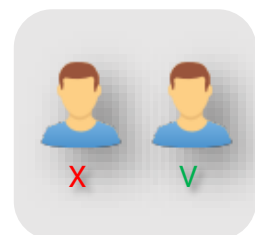
In particolare, le **prime quattro associazioni** sono particolarmente **rilevanti** visto il numero di volte che sono state ritrovate nei dati in analisi.

Per esempio, è accaduto ben 1799 volte che un cliente che acquistasse il prodotto 3629831 acquistasse il prodotto 3629853.



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

COME SFRUTTARE I RISULTATI DI QUESTA ANALISI?



CHURN E RFM

L'azienda dovrebbe cercare di intercettare i clienti «a rischio» e proporre campagne mirate a fidelizzarli.



MARKET BASKET ANALYSIS

L'azienda dovrebbe usare al meglio le associazioni tra prodotti ottenuti dall'analisi e sfruttarli a proprio favore con promozioni mirate



COMPrensione DEI PROPRI CLIENTI

L'azienda dovrebbe cercare di comprendere al meglio quali sono i comportamenti e i sentimenti dei propri clienti verso l'azienda



GRAZIE

Nome e cognome: Federico De Servi

Matricola: 812166

Email: f.deservi1@campus.unimib.it