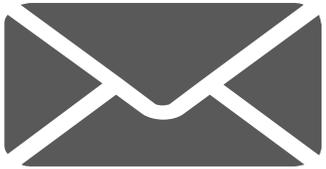




# Background

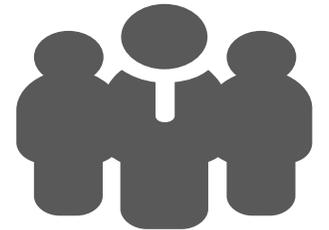
Informazioni  
sulle **campagne**  
**email marketing**



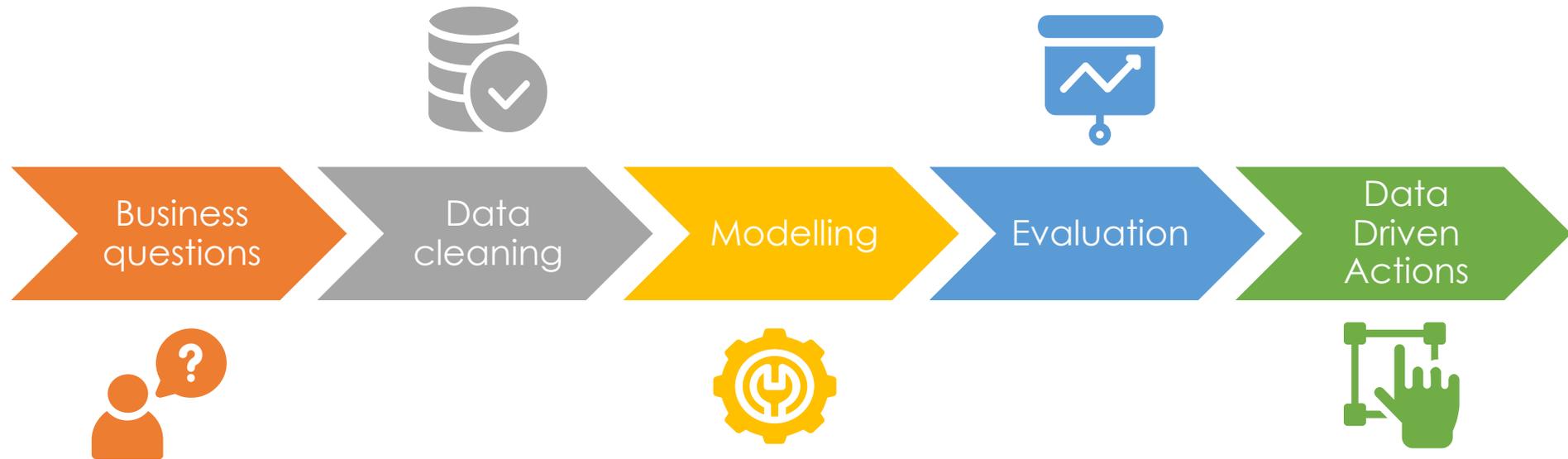
**Retail company nota**



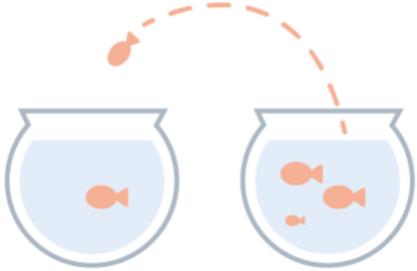
Informazioni sui  
**clienti**



# Workflow



## Propensity to **churn**



L'obiettivo è quello di prevenire l'abbandono dei propri clienti cercando di aumentare il tasso di fidelizzazione e retention. Diventata fondamentale comprendere quali azioni svolgere per trattenere i clienti

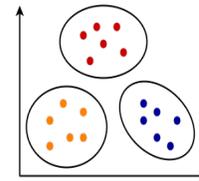
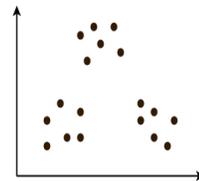
## RFM

L'intento è quello di segmentare i propri clienti assegnando a ciascuno di essi un valore in modo tale da pianificare differenti strategie di marketing e comunicazione

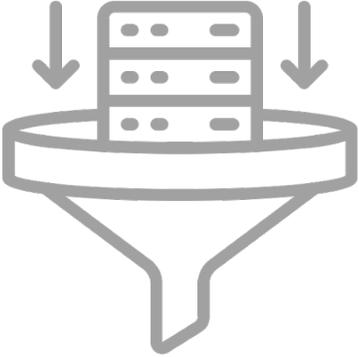


## Clustering

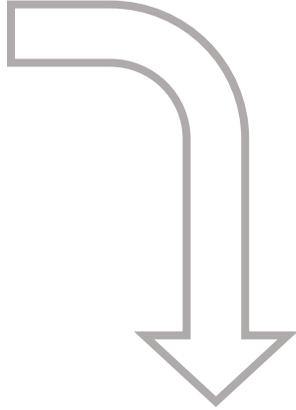
Si cerca di individuare gruppi di clienti con caratteristiche simili in modo tale da avere una visione strategica della customer base



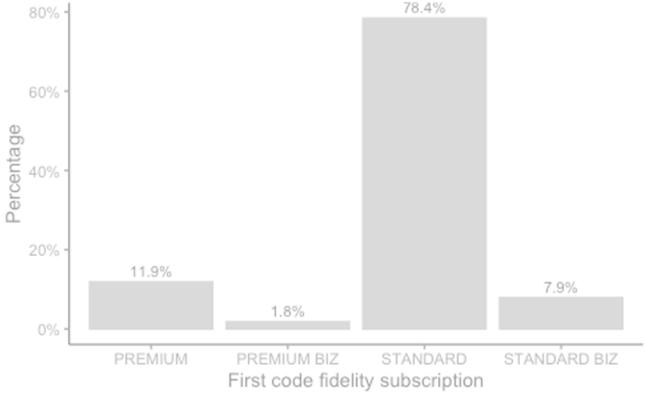
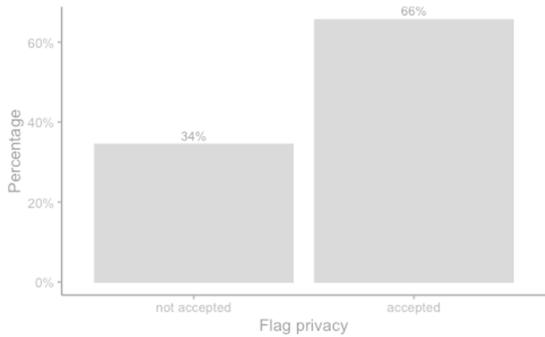
# Data preprocessing



- ⚙️ Eliminazione duplicati
- ⚙️ Eliminazione valori NULL
- ⚙️ Join tra dataframe
- ⚙️ ...

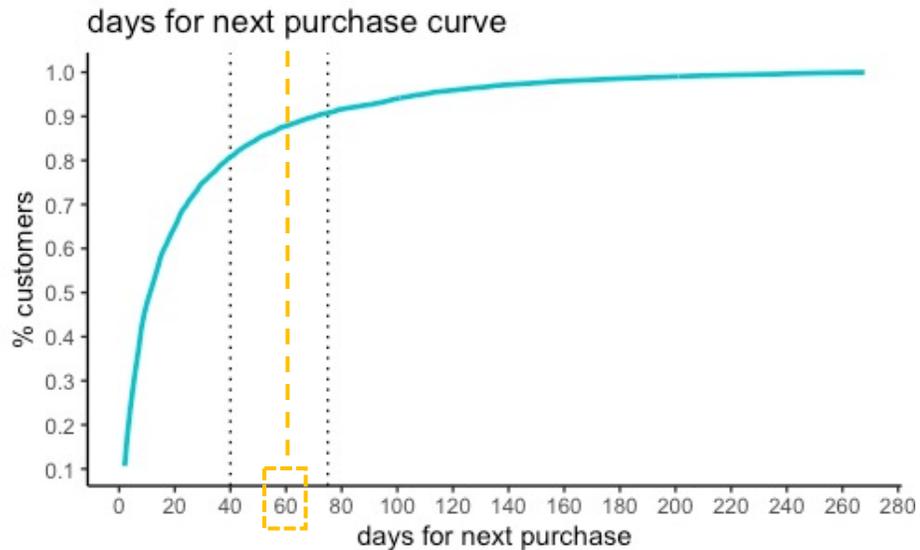


# Esplorazione dei dati



# Churn

## Soglia temporale



Se un cliente non effettua alcun acquisto nell'arco di 60 giorni allora viene definito come **churner**

## Variabili considerate

- spesa totale
- numero di acquisti
- recency
- regione di provenienza
- tipo di programma fedeltà a cui è iscritto il cliente
- se l'account del cliente è primario o meno
- identificativo del reference store

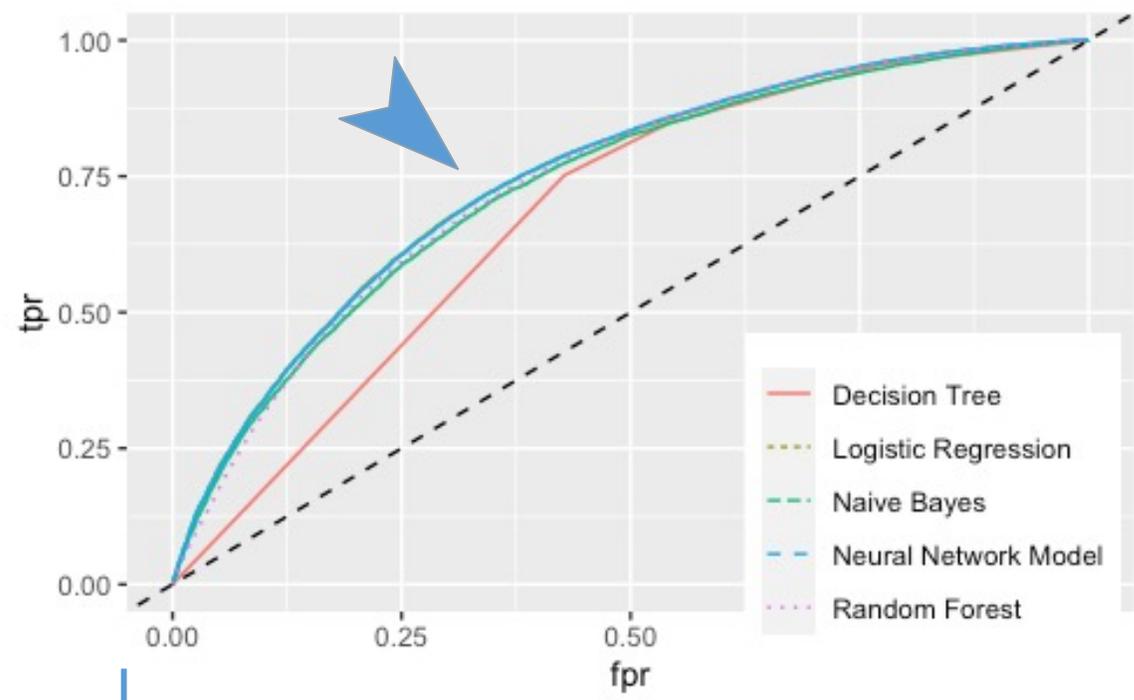
## Algoritmi utilizzati

- Decision Tree
- Random Forest
- Logistic Regression
- Neural Network Model
- Naive Bayes



# Churn

Confronto curve **ROC**



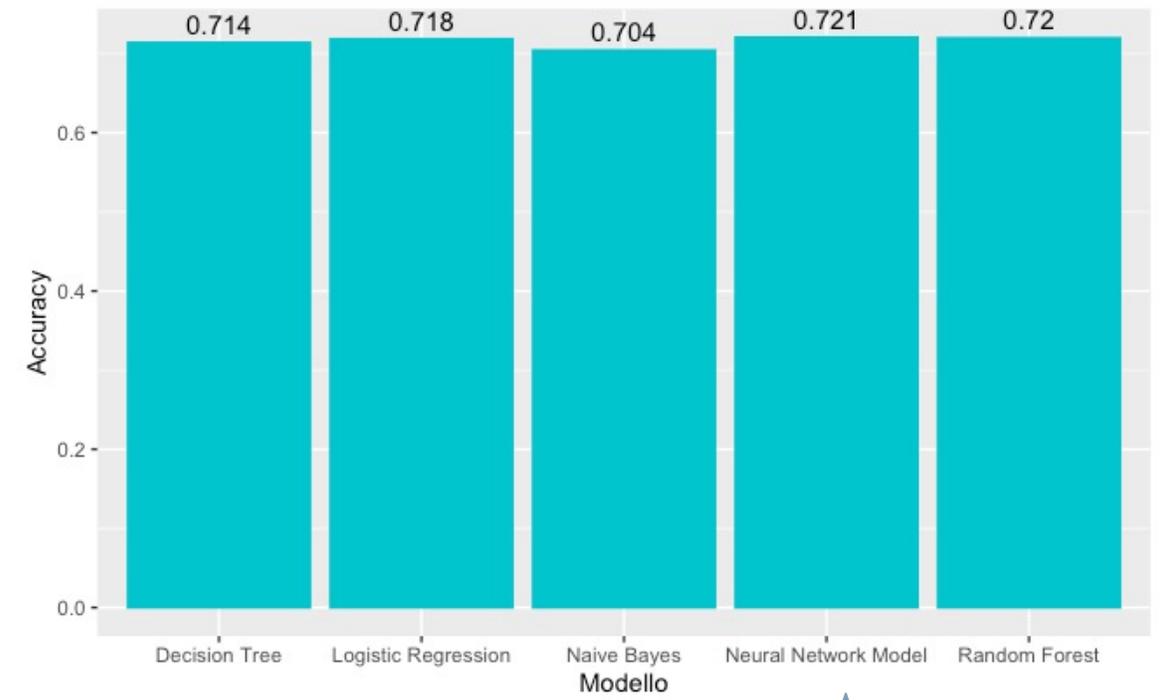
**Modello**

- Decision Tree
- Random Forest
- Logistic Regression
- Neural Network Model
- Naive Bayes

**Area Under Curve**

- 0.6824019
- 0.7386719
- 0.7461111
- 0.7462316
- 0.7339499

Confronto **Accuracy**



Il modello migliore prendendo in considerazione la metrica Accuracy risulta essere il **Neural Network Model**



# RFM

Le **metriche** utilizzate sono:

Recency



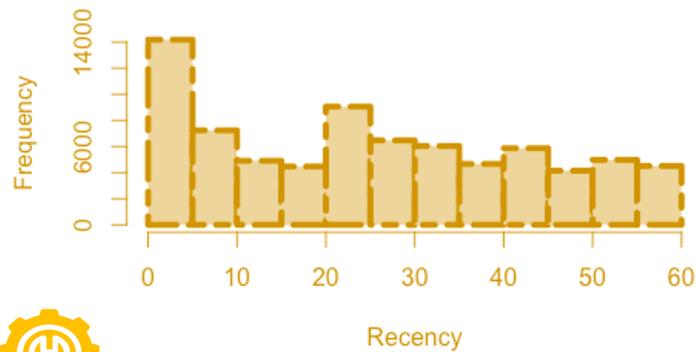
Frequency



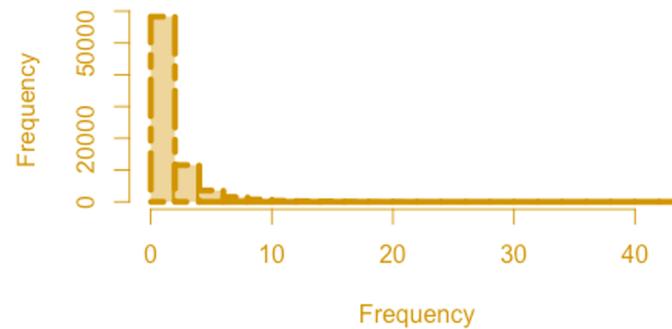
Monetary



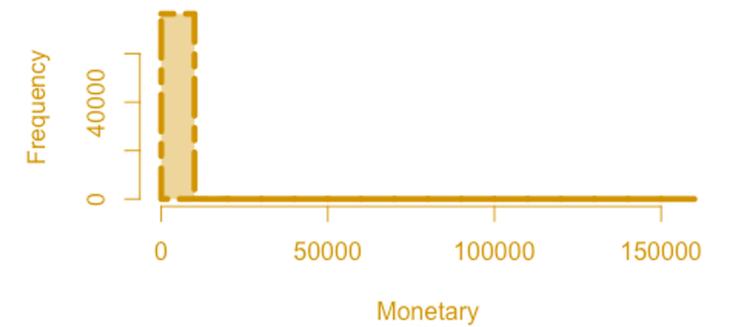
Distribution RECENCY



Distribution FREQUENCY

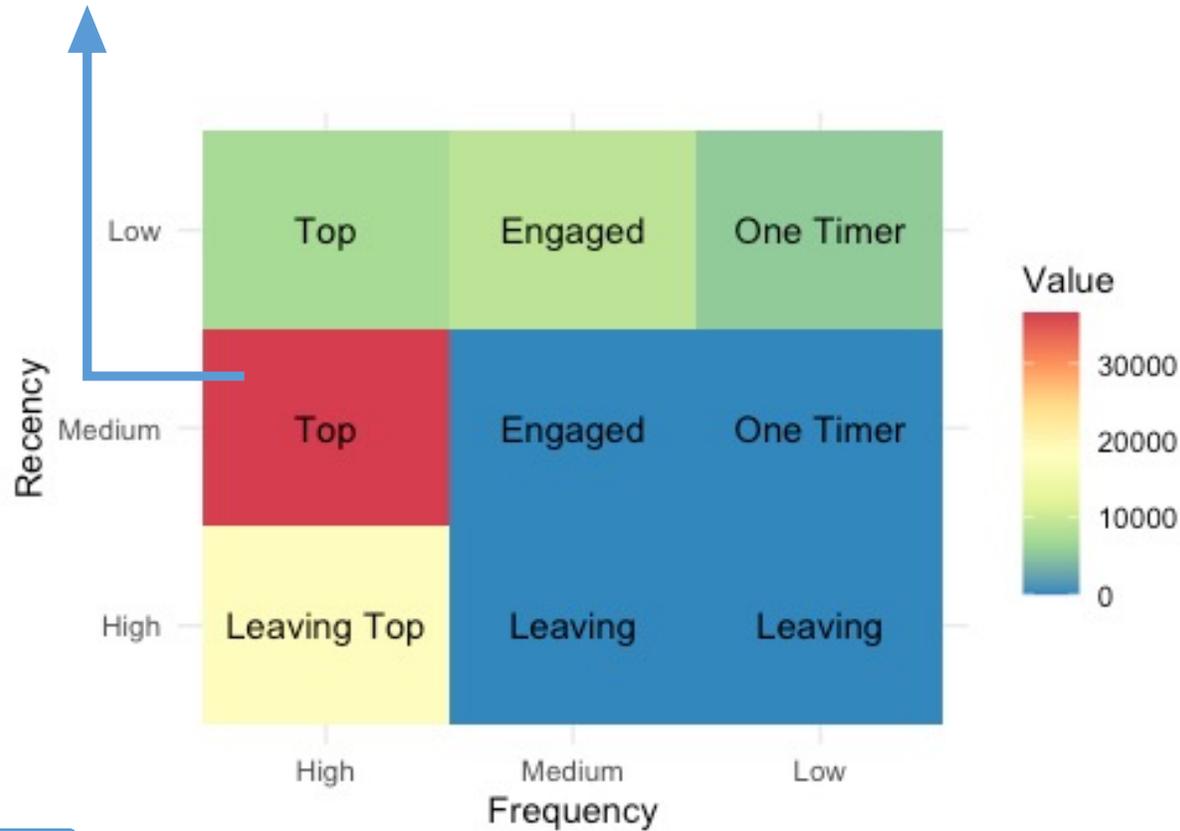


Distribution MONETARY



# RFM

**Top:** clienti che hanno acquistato recentemente o relativamente recentemente e con alta frequenza



**Gold:** clienti con valori di Monetary e RF\* rispettivamente "Top" e "Medium" oppure "Leaving Top" e "High"

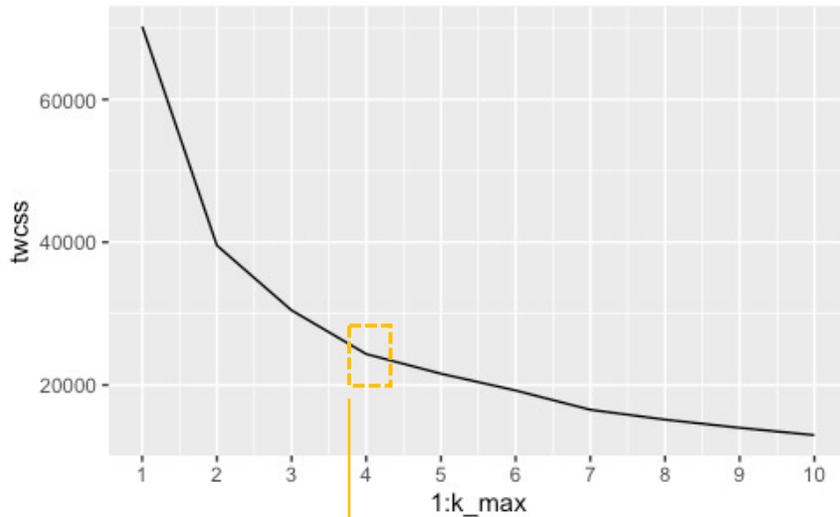


\* RF: combinazione tra Recency e Frequency



# Clustering

## Numero ottimale di cluster



Osservando il grafico si sceglie come numero ottimale di cluster **4**

## Variabili considerate

- numero acquisti
- numero articoli
- spesa (importo lordo - sconto)

## Algoritmi utilizzati

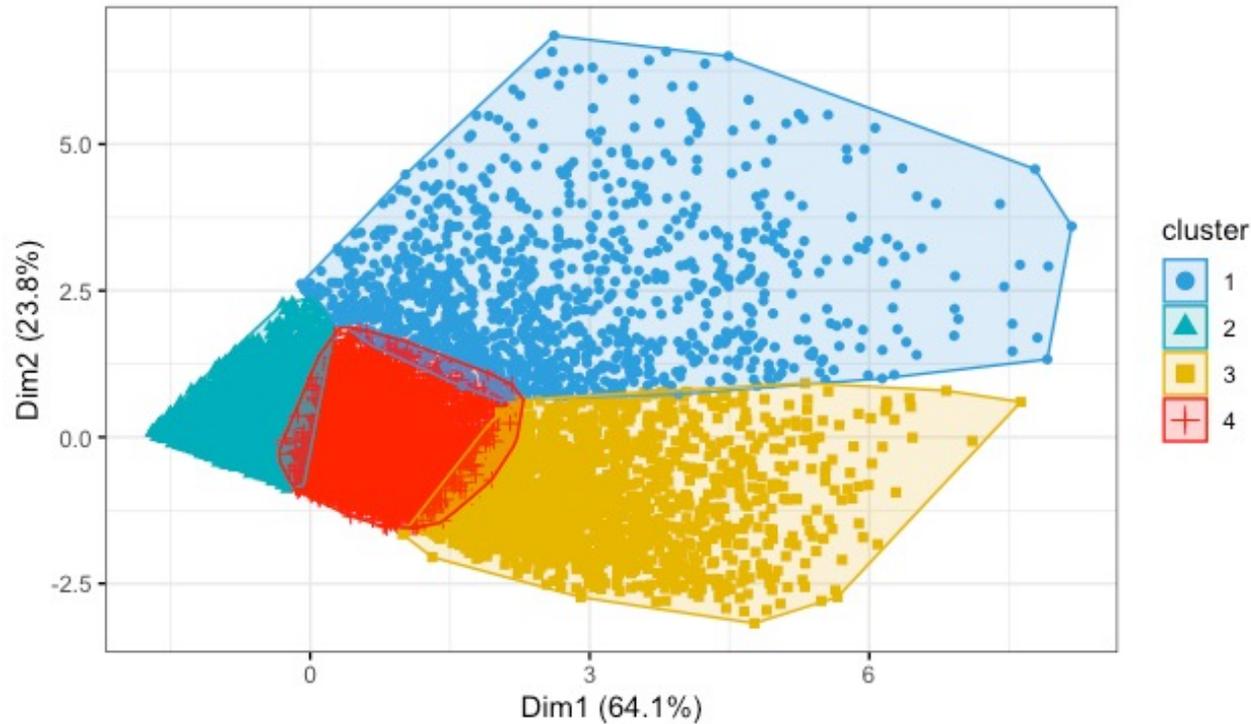
- K-means
- K-medoids
- DBSCAN



# Clustering

Output dell'algoritmo **K-medoids**

Cluster plot



**Valori medi** dei 4 cluster

Numero acquisti	Numero articoli	Spesa
6	24	446.91
11	47	939.29
4	11	232.03
7	29	2575.03



### Real Time Marketing

Reagire in modo tempestivo agli stimoli dell'ambiente esterno in modo tale da ottenere visibilità e conquistare nuovi clienti



### Multi-channel strategy

Utilizzare molteplici canali di comunicazione per raggiungere i clienti



### Buyer personas

Creare una rappresentazione del profilo del cliente ideale in modo tale comprendere meglio i suoi bisogni



### CRM

Gestione delle relazioni dei clienti al fine di rendere durature le relazioni con essi



Grazie per l'attenzione