

# DIGITAL MARKETING ANALYSIS

Digital Marketing Project  
A.A. 2018/2019

# Strategy: a model for data-driven actions



## E-mail engagement

I dati demografici e comportamentali del consumatore possono essere utilizzati per inviare campagne mirate



## Cluster analysis

I risultati relativi alla cluster analysis sono fondamentali per individuare il segmento di clientela al quale rivolgersi, limitando la dispersione delle proprie risorse

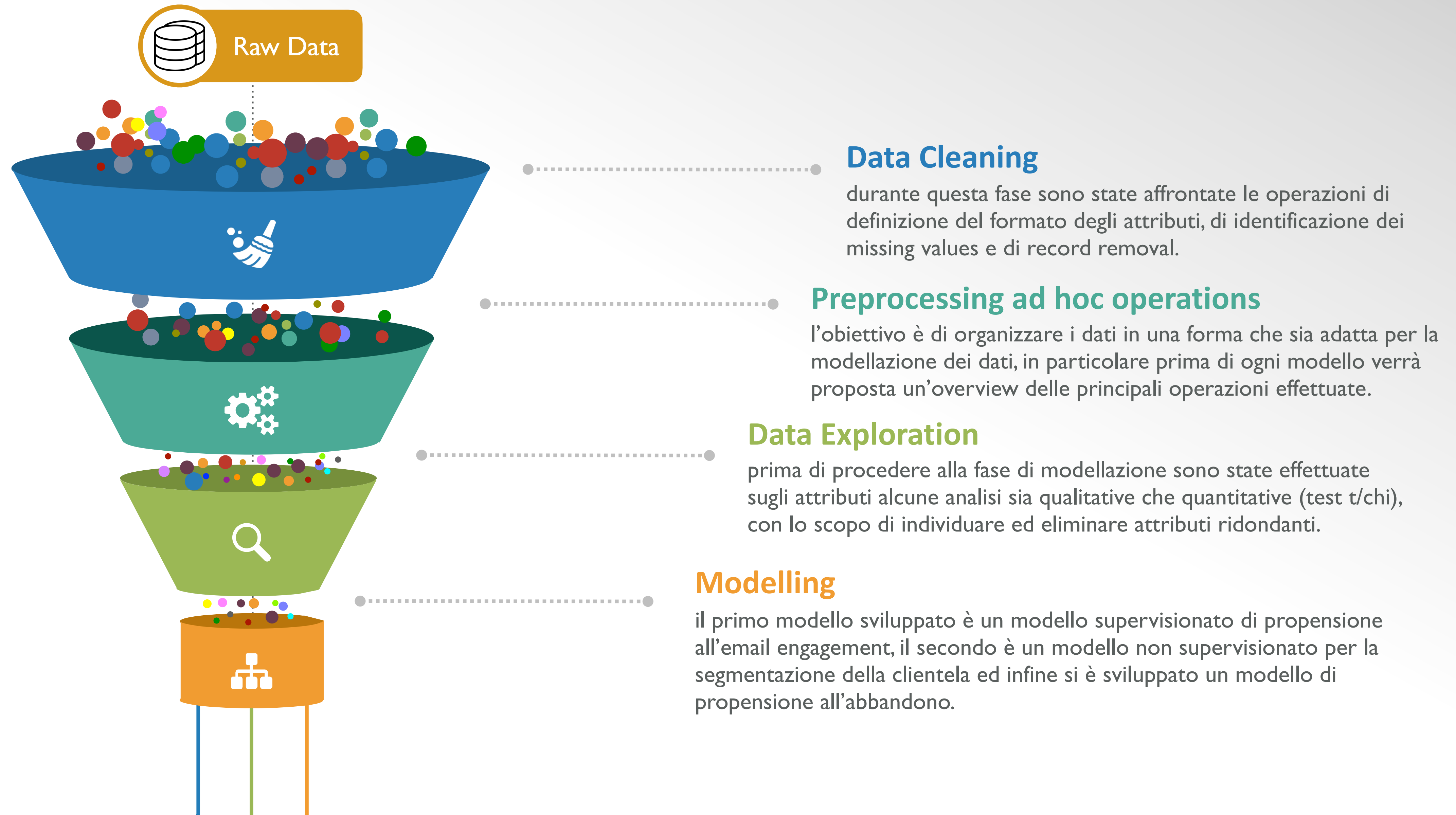


## Churn prediction

I dati comportamentali dei clienti possono aiutare ad individuare e prevenire l'abbandono con strategie di marketing ad hoc



# Analysis workflow: a data-driven communication strategy





# Propensity to E-mail engagement model - Overview

## Variabili esplicative:

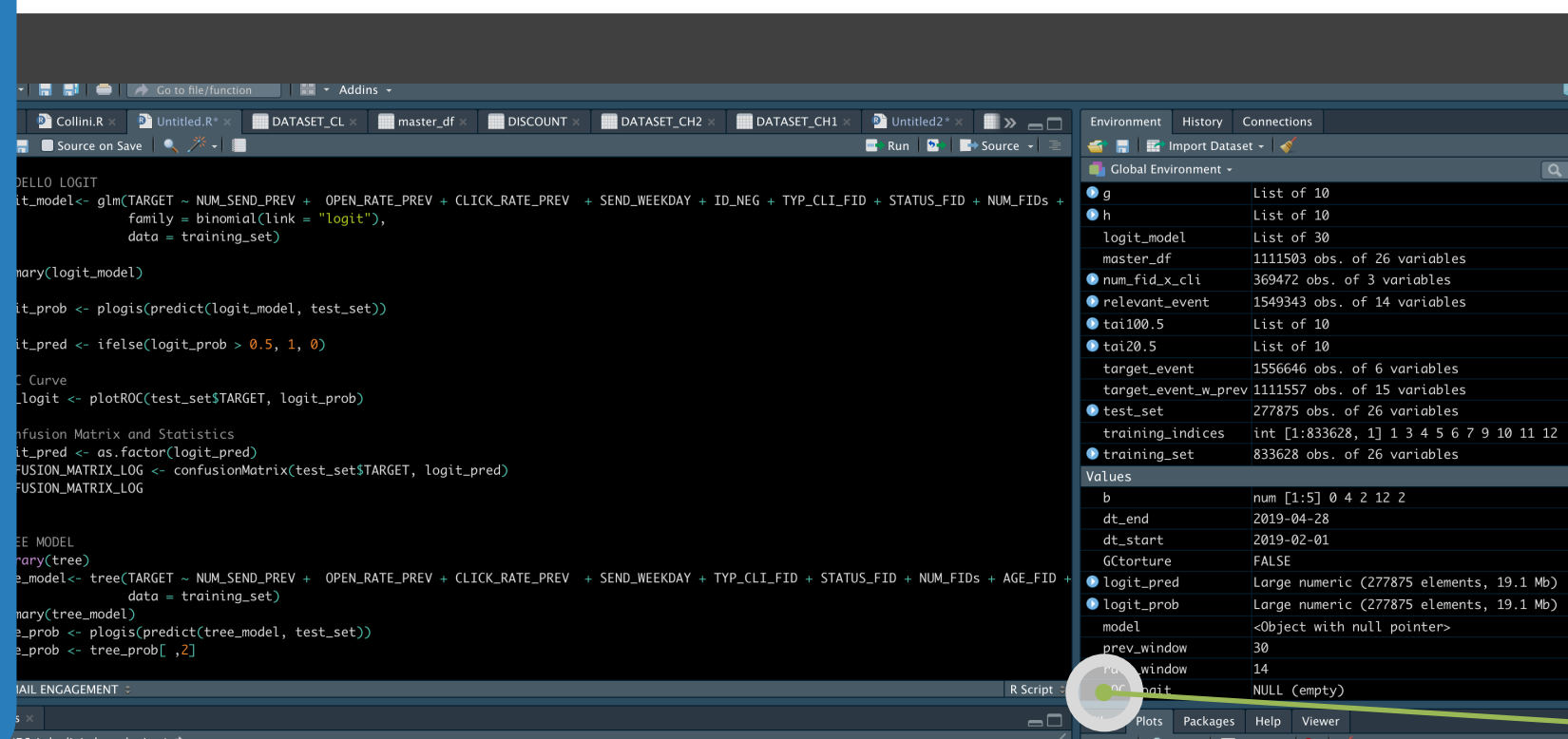
- n. di campagne inviate
- tasso di apertura
- tasso di click
- giorno della settimana
- tipologia di account
- stato della fidelity
- n. di fidelity
- tipologia di fidelity
- tempo di fidelity
- numero
- email provider
- regione
- ID negozio

## Creazione del datamart di analisi

- ✓ Controllo e **selezione** delle potenziali variabili da inserire del modello
- ✓ Creazione di alcune nuove variabili tra cui la variabile **target**
- ✓ Definizione di una **finestra temporale** di due giorni
- ✓ Creazione del **datamart** di analisi

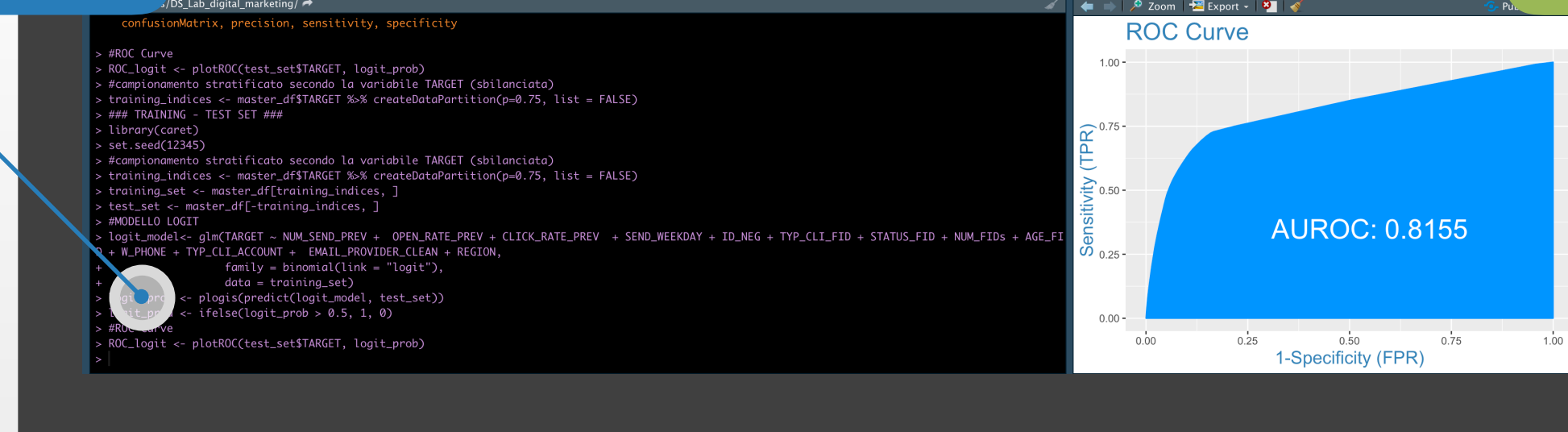
N. record: 1111503

Analisi relativa al periodo:  
01/02/2019 - 30/04/2019



## Metodologia

- ✓ Test per la verifica della significatività delle variabili (t-test e chi-squared test)
- ✓ Partizione dei dati in training e test set, con campionamento stratificato secondo la variabile target
- ✓ Implementazione di un modello logistico e di un albero decisionale



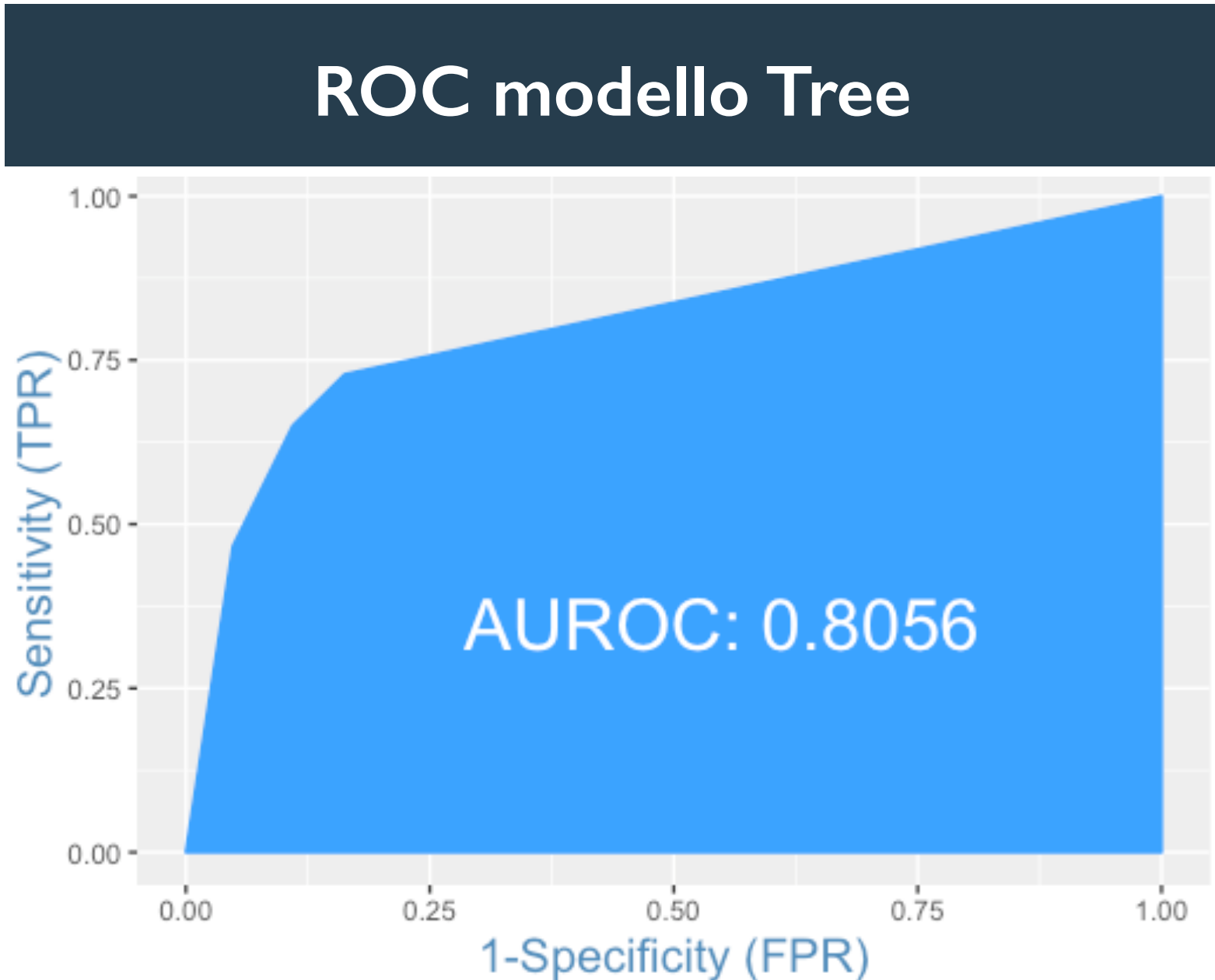
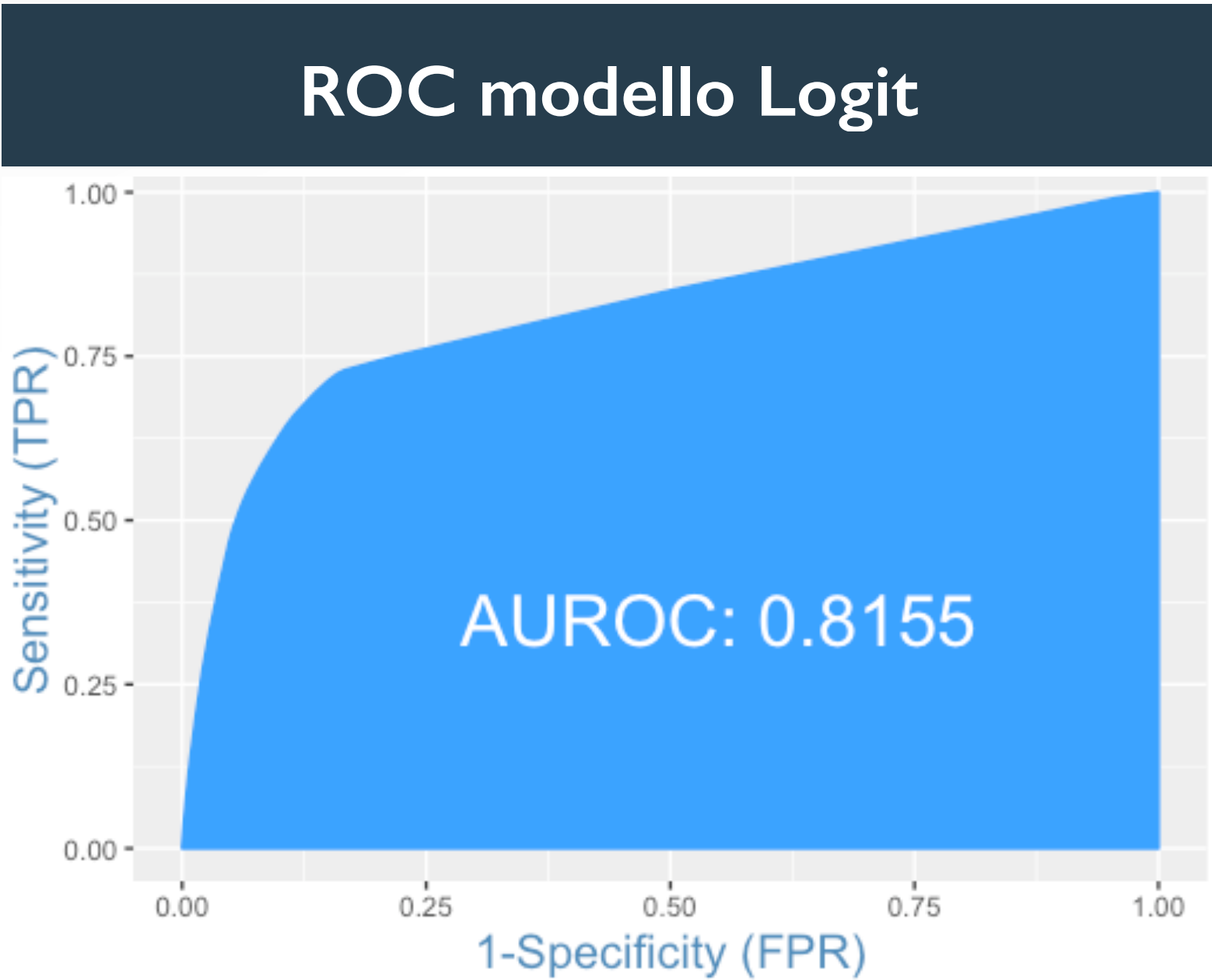
Variabile target: assume valore 1 se una mail viene aperta nella finestra temporale definita, diversamente assume valore 0

# Propensity to E-mail engagement model - Results

## Variabili esplicative:

- n. di campagne inviate
- tasso di apertura
- tasso di click
- giorno della settimana
- tipologia di account
- stato della fidelity
- n. di fidelity
- tipologia di fidelity
- tempo di fidelity
- numero
- email provider
- regione
- ID negozio

	Accuracy	Precision	Recall	Specificity
Modello Logit	0.8784	0.4666	0.6326	0.9092
Modello Tree	0.8782	0.4562	0.6356	0.9077





# Cluster analysis - Overview



## Variabili esplicative:

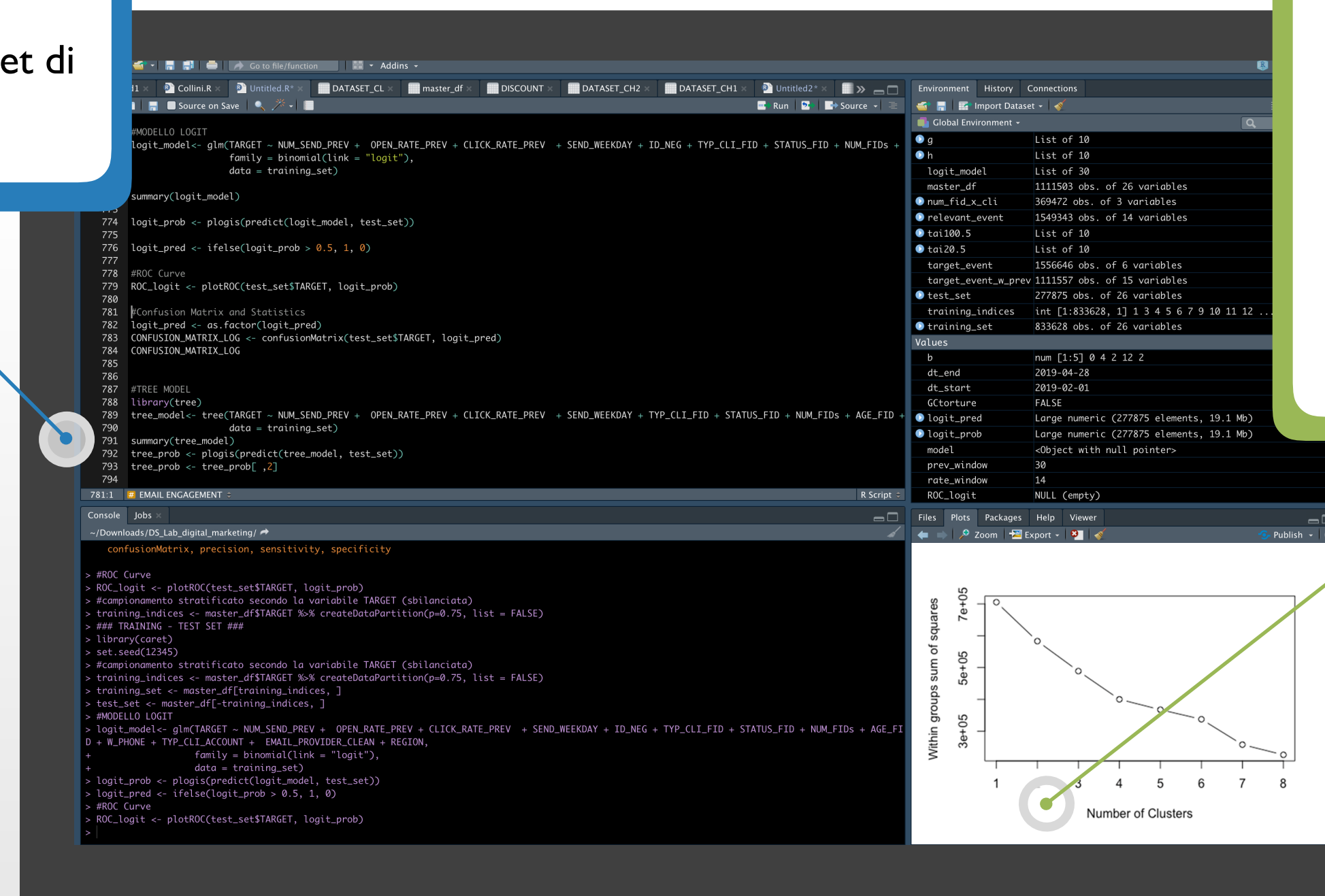
- n. di scontrini per cliente
- valore scontrino medio
- tasso di sconto complessivo applicato al cliente
- n. medio di articoli per scontrino

## Creazione del dataset di analisi

- ✓ **Selezione** dei soli attributi numerici da inserire del modello
- ✓ **Standardizzazione** di tutte le variabili e creazione del dataset di analisi

N. record: 176396

Analisi relativa al periodo:  
28/02/2019 - 30/04/2019



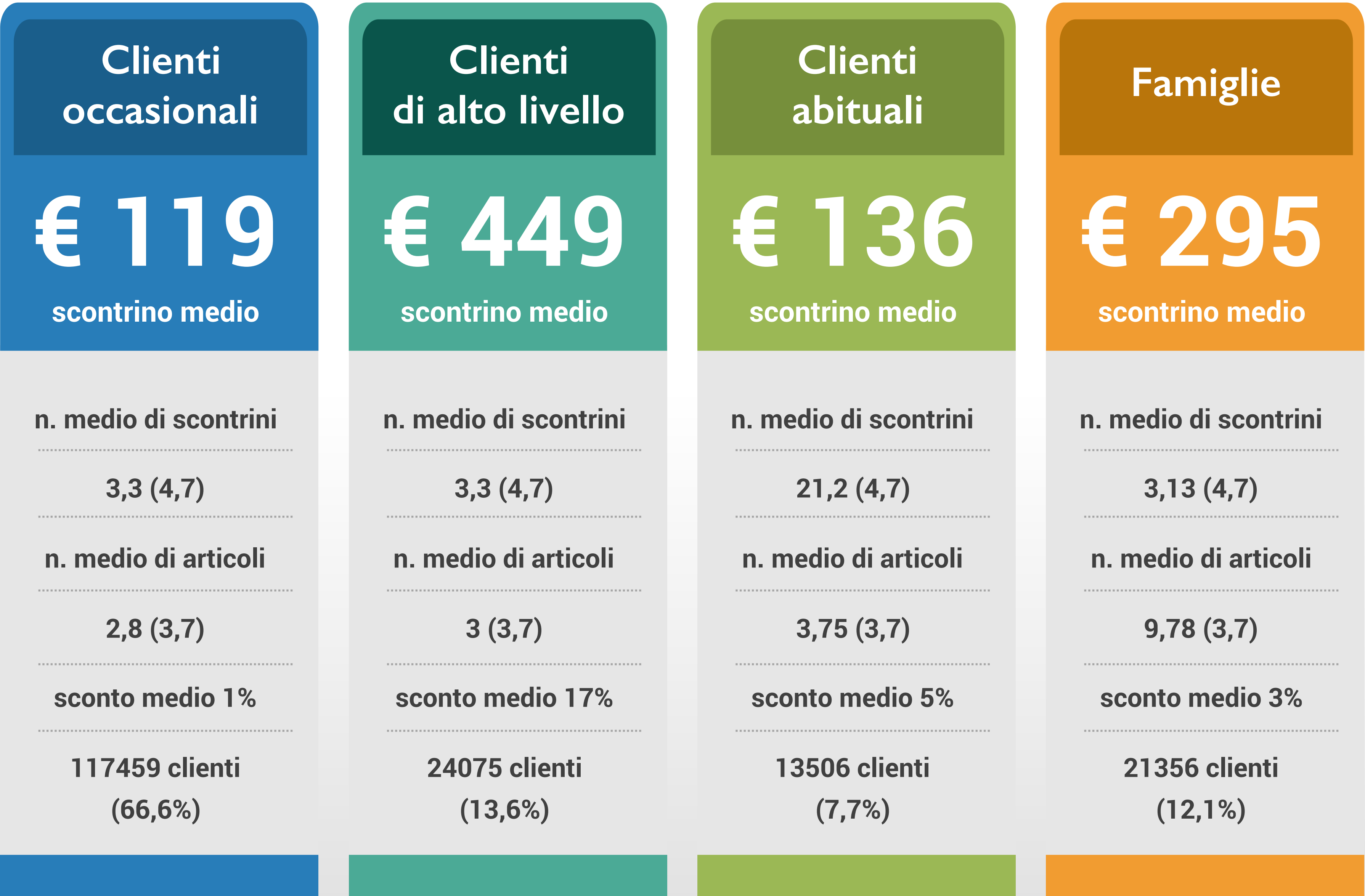
## Metodologia

- ✓ Definizione del **numero ottimale di cluster** tramite la valutazione della somma delle distanze dentro i gruppi
- ✓ Analisi della correlazione per la selezione delle variabili
- ✓ Implementazione dell'algoritmo di clustering **K-means** e **segmentazione deterministica**.

# Cluster analysis - Results

Variabili esplicative:

- n. di scontrini per cliente
- valore scontrino medio
- tasso di sconto complessivo applicato al cliente
- n. medio di articoli per scontrino



# Propensity to churn model - Overview

## Variabili esplicative:

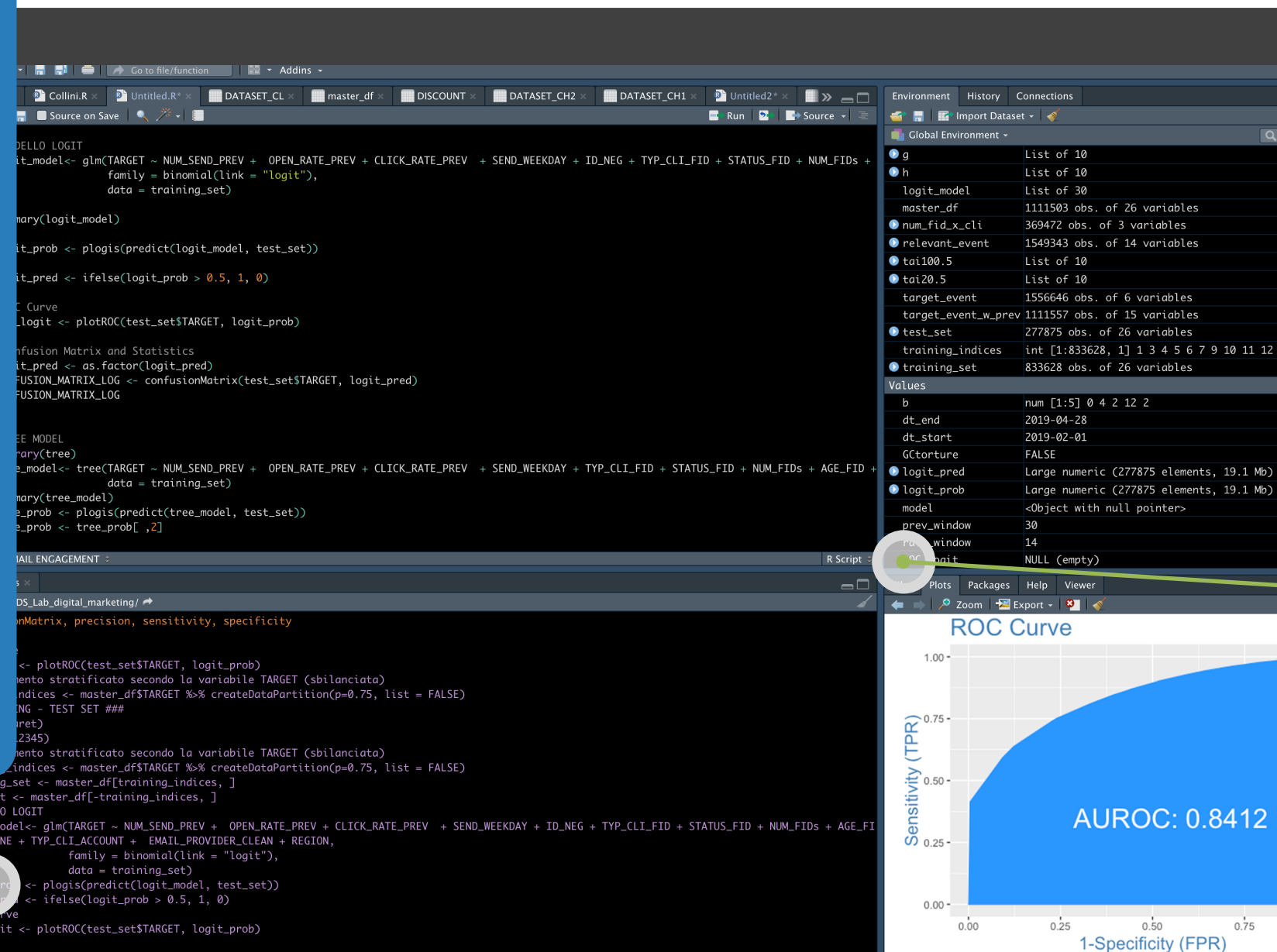
- n. scontrini per cliente
- valore scontrino medio
- tasso di sconto complessivo applicato al cliente
- n. medio di articoli per scontrino
- stato della fidelity
- tipologia di fidelity
- macroarea
- ID negozio



## Creazione del datamart di analisi

- ✓ **Selezione** delle variabili necessarie per la creazione di nuovi attributi
- ✓ Definizione di una **soglia temporale** di riacquisto di 60 giorni
- ✓ **Esclusione** dei clienti che hanno effettuato il primo acquisto dopo il 28/02
- ✓ **Standardizzazione** delle variabili numeriche e creazione del **dataset** per l'analisi

N. record: 176396  
Analisi relativa al periodo:  
28/02/2019 - 30/04/2019



## Metodologia

- ✓ Partizione dei dati in training e test set, con campionamento stratificato secondo la variabile target
- ✓ definizione dell'**optimal cutoff** per la variabile target nel rispetto della missclassification.
- ✓ Implementazione di un modello **logistico**, di un **albero decisionale** ed un modello **logit boost**.

Variabile target: assume valore 0 se il consumatore non riacquista nella finestra temporale definita, assume valore 1 e il cliente verrà classificato come churmer

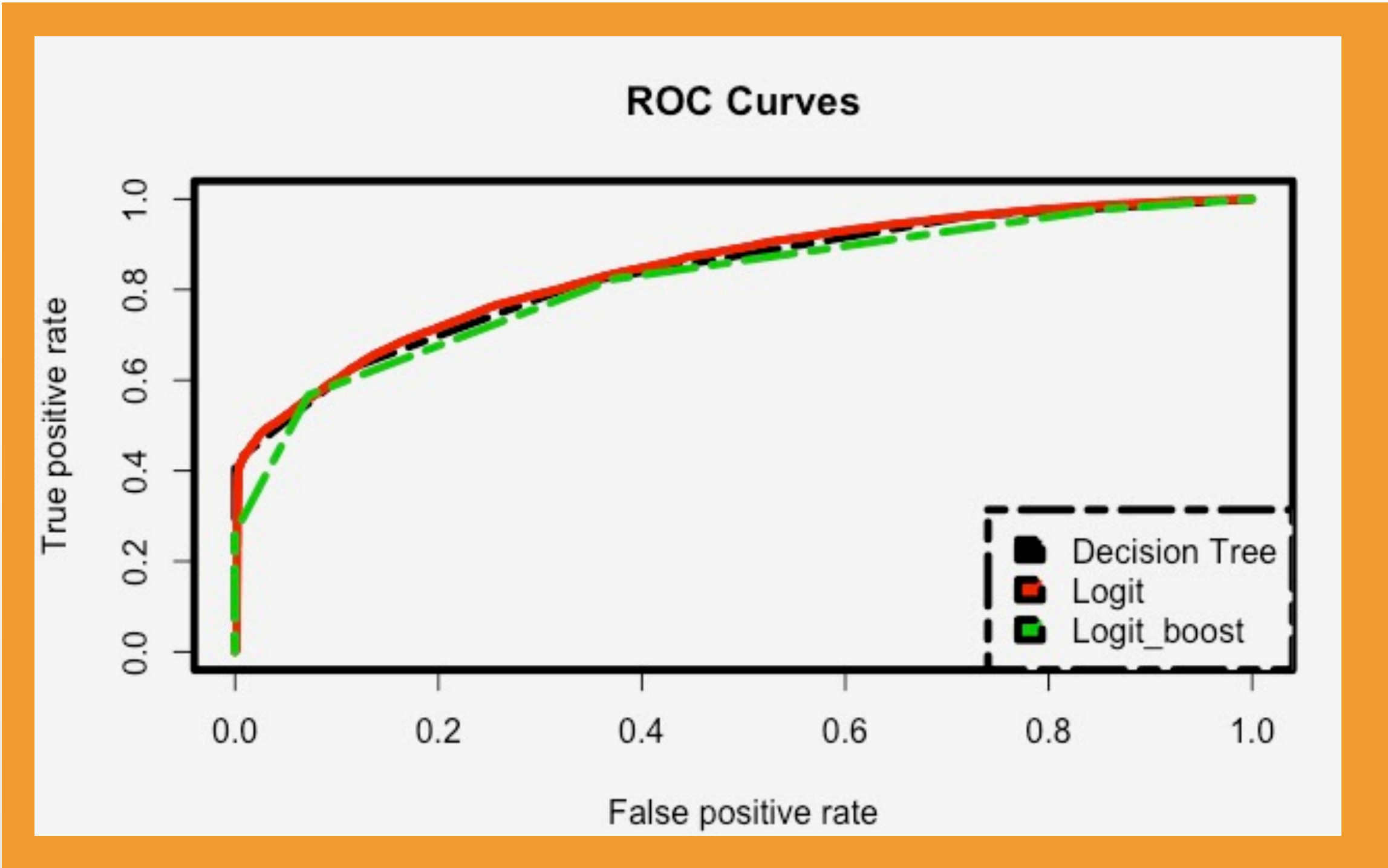


# Propensity to churn model - Results

Variabili esplicative:

- n. scontrini per cliente
- valore scontrino medio
- tasso di sconto complessivo applicato al cliente
- n. medio di articoli per scontrino
- stato della fidelity
- tipologia di fidelity
- macroarea
- ID negozio

	Accuracy	Precision	Recall	Specificity
Modello Tree	0.7766	0.9623	0.7815	0.7347
Modello Logit	0.7842	0.9423	0.7979	0.7001
Modello Logit Boost	0.77	0.8228	0.8557	0.5698

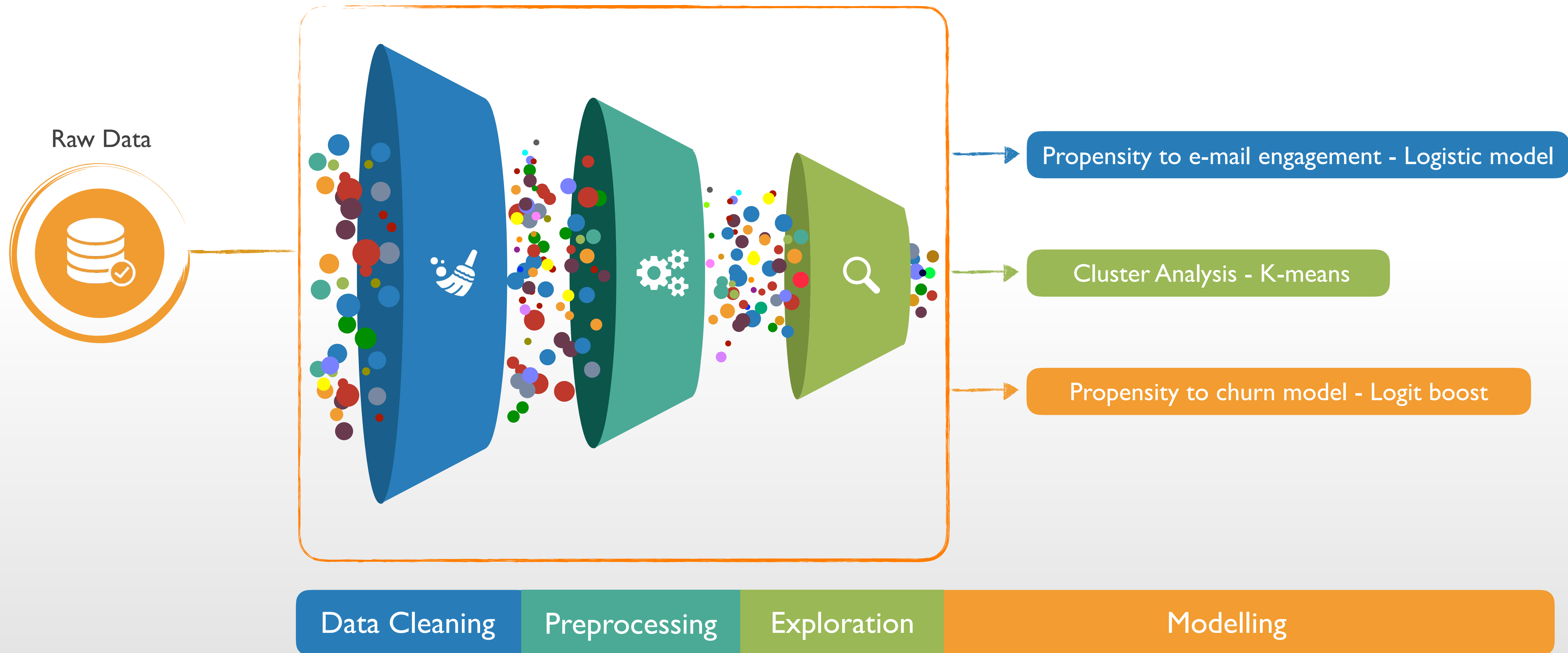


Area Under the curve

- ✓ Decision Tree  
AUC = 0.8348
- ✓ Logit model  
AUC = 0.8412
- ✓ Logit Boost  
AUC = 0.8167

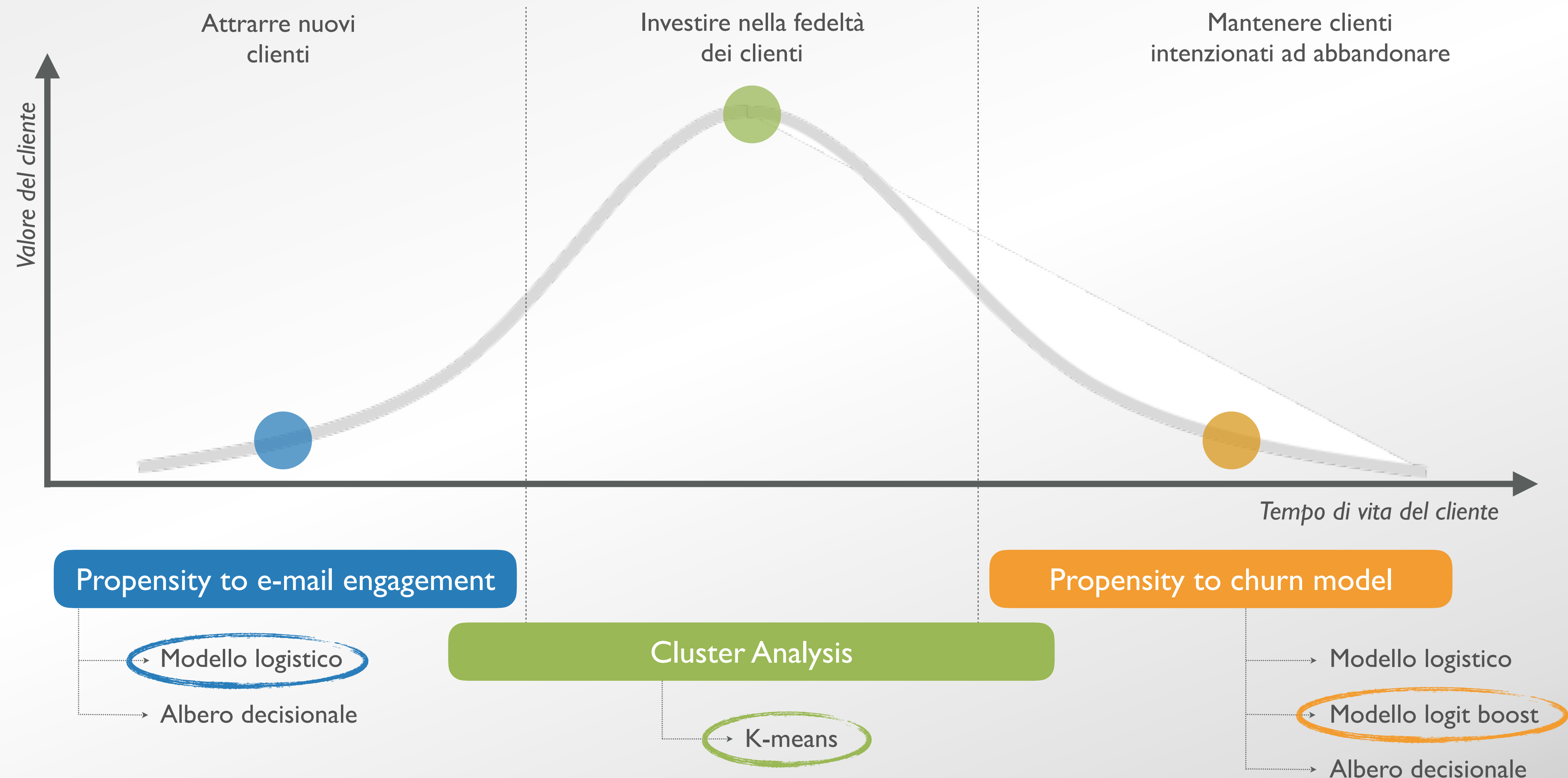
# Strategy: Our data-driven model

Di seguito si presenta la struttura della data-driven strategy definita dalla nostra analisi selezionando i modelli più performanti



# Analytics Goal: a customer life-cycle perspective analysis

- L'obiettivo di questa analisi è di **generare conoscenza** sui clienti descritti nei dati e di raggiungere una **visione strategica della Consumer Base**.
- In particolare si vuole sviluppare una **strategia di comunicazione data-driven** che sia **basata su algoritmi di machine learning**.



# Grazie per l'attenzione

---

Lorenzo Pastore

847212

Raffaele Anselmo

846842