

**Università degli Studi di Milano-Bicocca**  
Dipartimento di Informatica, Sistemistica e Comunicazione  
Corso di Laurea Magistrale in Data Science

# Digital Marketing

RFM Customer Segmentation & Churn Model

Niccolò Bencini

# Project Pipeline



# Dataset

- **raw\_1\_cli\_fid**  
informazioni sugli abbonamenti fedeltà di ciascun cliente
- **raw\_2\_cli\_account**  
informazioni sugli account cliente
- **raw\_3\_cli\_address**  
informazioni sull'indirizzo corrispondente a ogni account cliente
- **raw\_4\_cli\_privacy**  
informazioni sulle politiche sulla privacy accettate da ciascun cliente
- **raw\_5\_camp\_cat**  
informazioni sulle e-mail di marketing
- **raw\_6\_camp\_event**  
eventi (inviati, aperti, clic, ...) relativi alle comunicazioni e-mail di marketing
- **raw\_7\_tic**  
contiene le transazioni di acquisto e rimborso di ciascun cliente

Per la realizzazione dei modelli RFM e Churn sono stati utilizzati principalmente i dati del file **raw\_7\_tic**.

# Business Questions



## RFM Customer Segmentation

Il modello RFM fornisce una descrizione deterministica del valore di ogni cliente in termini di comportamento d'acquisto.

Le metriche scelte per descrivere i comportamenti dei clienti sono:

- Recency
- Frequency
- Monetary Value



## Propensity to Churn

Il modello di churn fornisce una stima predittiva della probabilità che un cliente smetta di interagire con l'azienda.

# Data Overview



**907846** scontrini



**212124** clienti

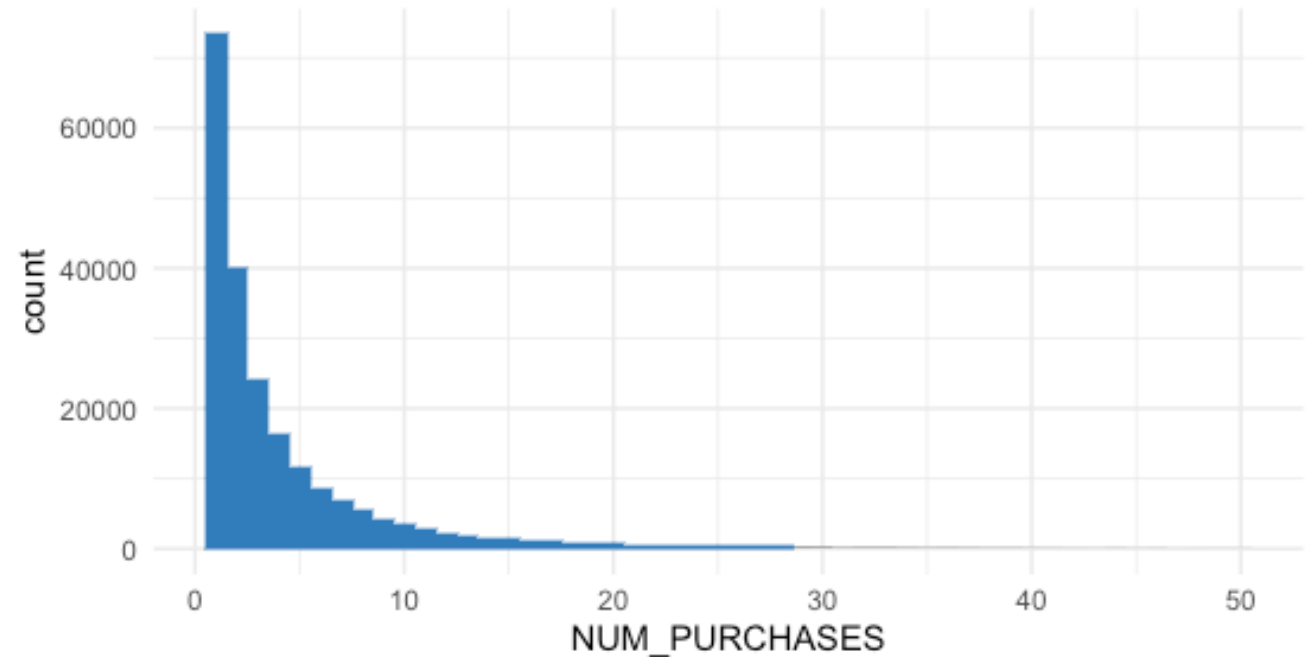
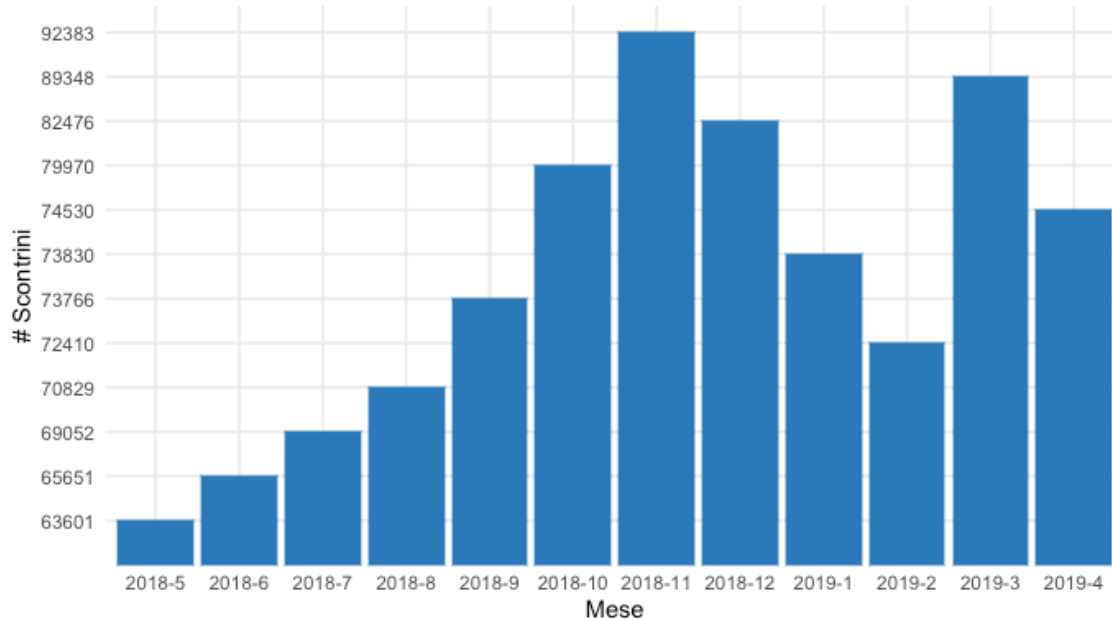


**01/05/2018 – 30/04/2019**



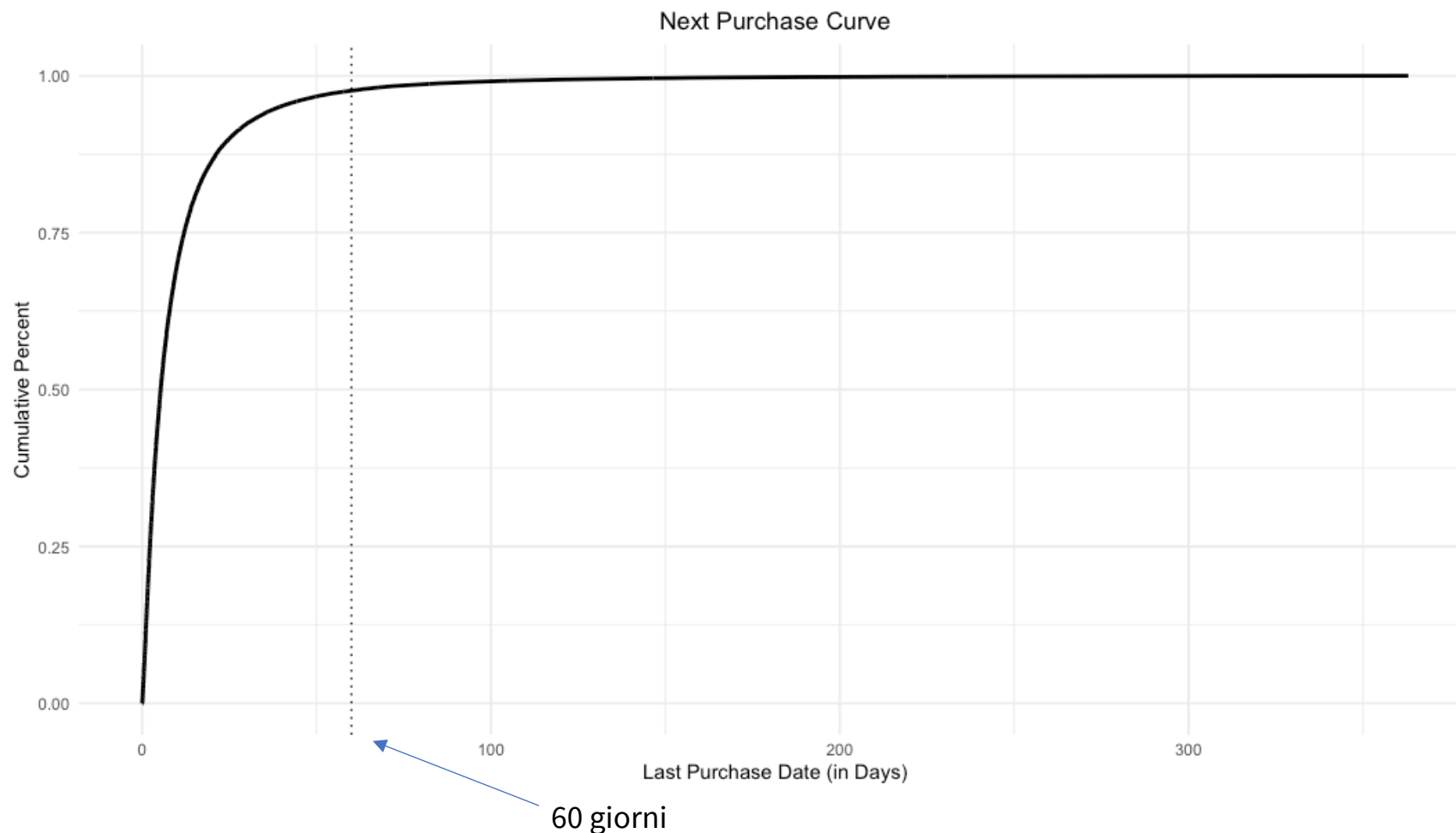
**32,65€** scontrino medio

Distribuzione Scontrini per Mese



# Data Overview

**Curva di Riacquisto:** distribuzione dei clienti in base alla differenza media in giorni tra un acquisto e il successivo



# RFM Customer Segmentation

I clienti sono segmentati in base a quanto recente hanno acquistato un prodotto (**recency**), quanto spesso acquistano (**frequency**) ed all'importo totale speso (**monetary**).

In questo modo è possibile raggruppare i clienti in categorie per valutarne il valore. Ad esempio, un cliente *Diamond* è molto più fidelizzato rispetto ad un cliente *Bronze*.

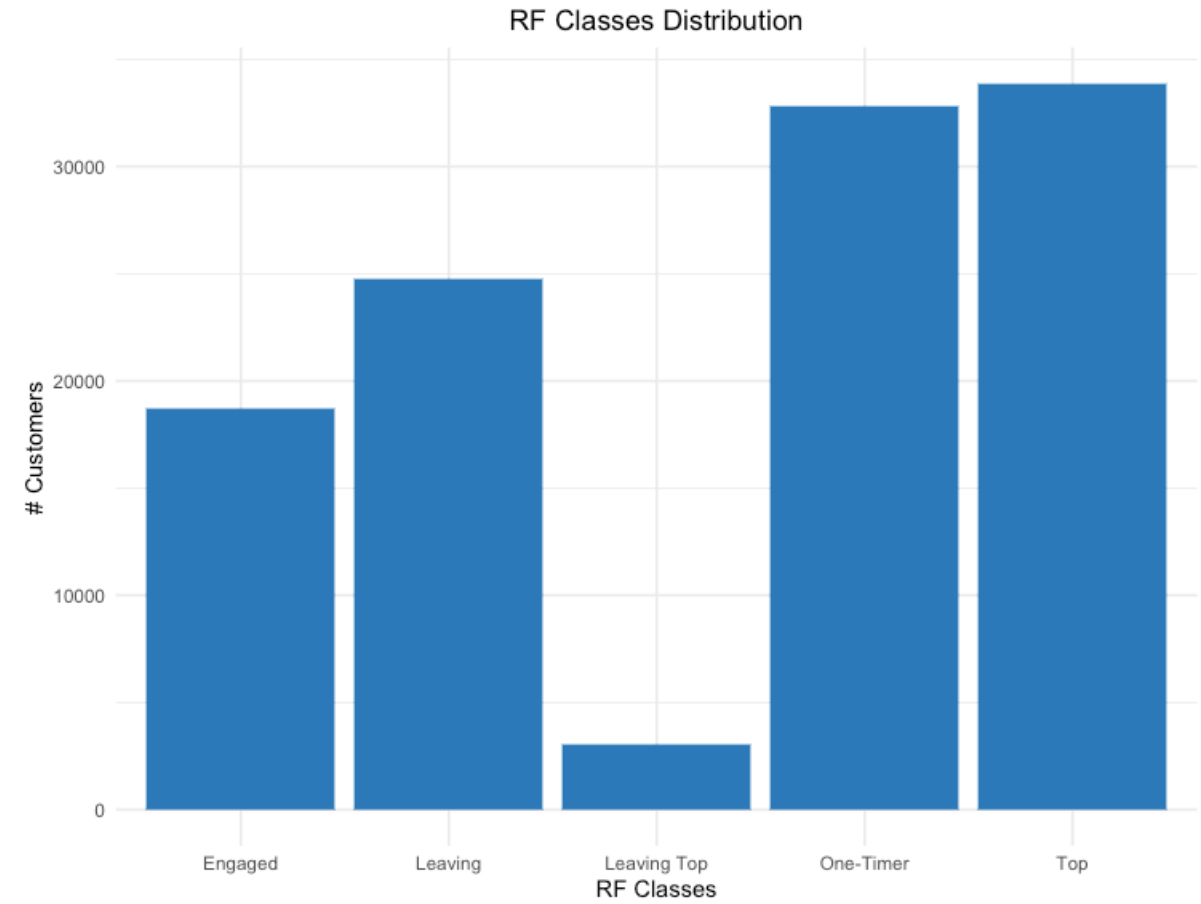
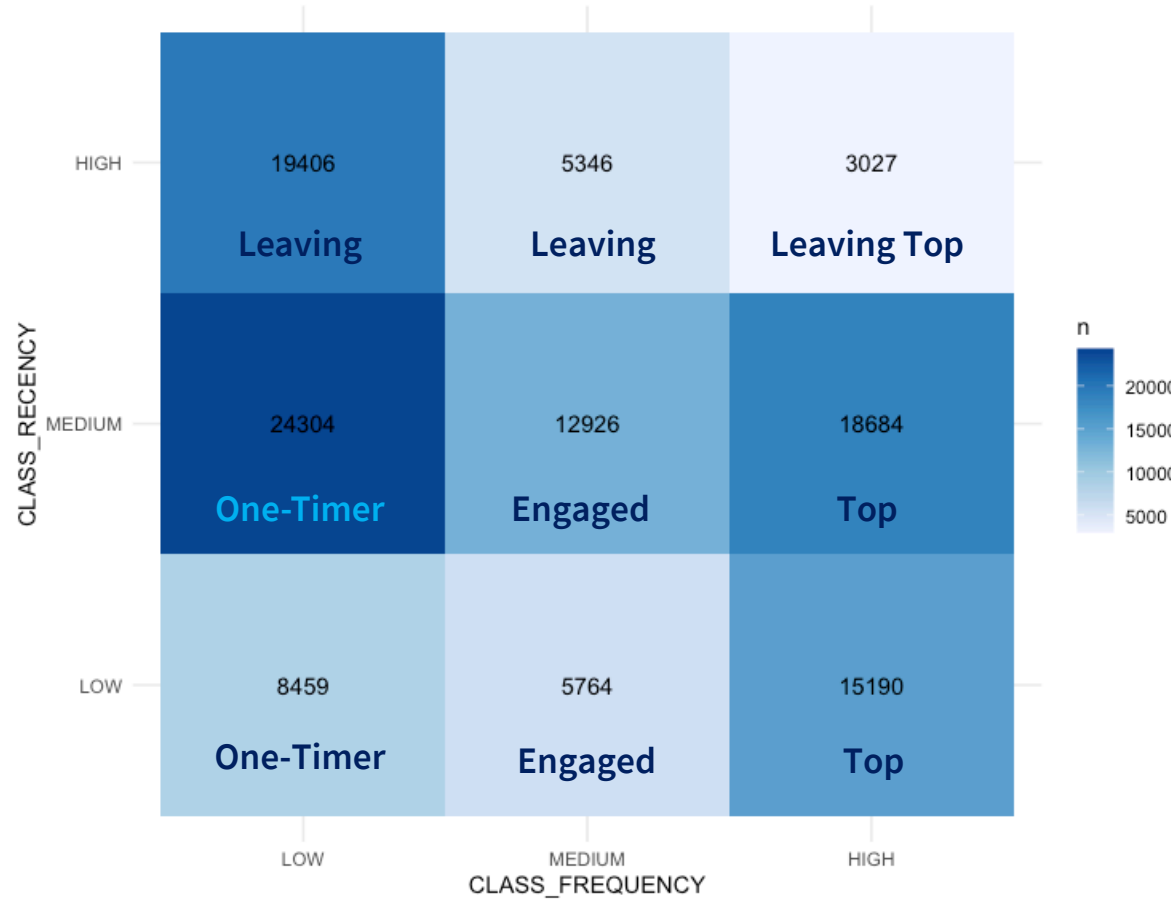
Ad ogni cliente viene associata una classe in base al percentile di appartenenza, secondo la seguente tabella.

Class	Recency	Frequency	Monetary
LOW	< 25%	< 50%	< 25%
MEDIUM	25% < x < 75%	50% < x < 75%	25% < x < 75%
HIGH	> 75%	> 75%	> 75%

# RFM Model

Inizialmente viene creata la matrice RF, calcolata sulla base dei valori di **recency** e **frequency** dei clienti.

Si ottengono i seguenti risultati. Come si vede la classe più numerosa è quella dei clienti **Top**.





# RFM Model

Successivamente, si uniscono i risultati della matrice RF con le classi di **monetary** per ottenere la **matrice RFM**.

In questo caso, la classe di clienti più frequente è **Tin** (RF Leaving e bassa monetary).



# RFM Model – Data Driven Actions



Investire di più in azioni di marketing per la cura dei clienti di classe Diamond / Gold



Identificare le azioni di marketing più efficienti in termini di costi per aumentare il valore dei clienti di classe Silver / Bronze



Risparmiare budget sulle azioni di marketing per i clienti a basso valore (Cheap, Copper, Tin)

# Propensity to Churn

Il modello di churn fornisce una stima predittiva della probabilità che un cliente smetta di interagire con l'azienda.



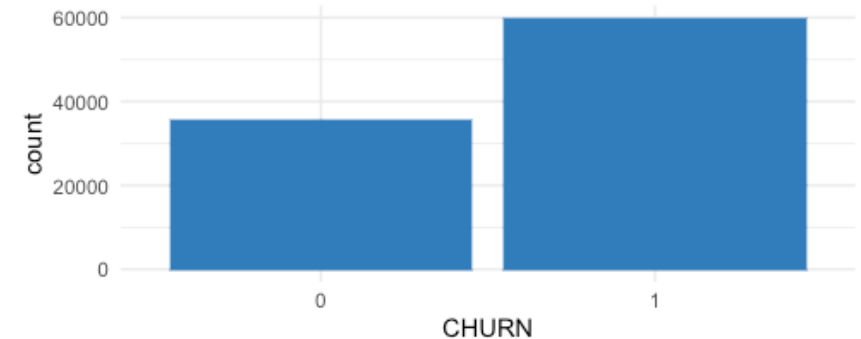
**Holdout Period:** 01/01/2019 – 28/02/2019



**Churner:** se un cliente non ha acquistato nel periodo di Holdout è considerato churner



**Train-Test split:** 70% Train, 30% Test



# Propensity to Churn

Features utilizzate per la costruzione dei modelli:

- Recency;
- Monetary;
- Importo Lordo;
- Tipo Carta Fedeltà;
- Tipo di Lavoro.



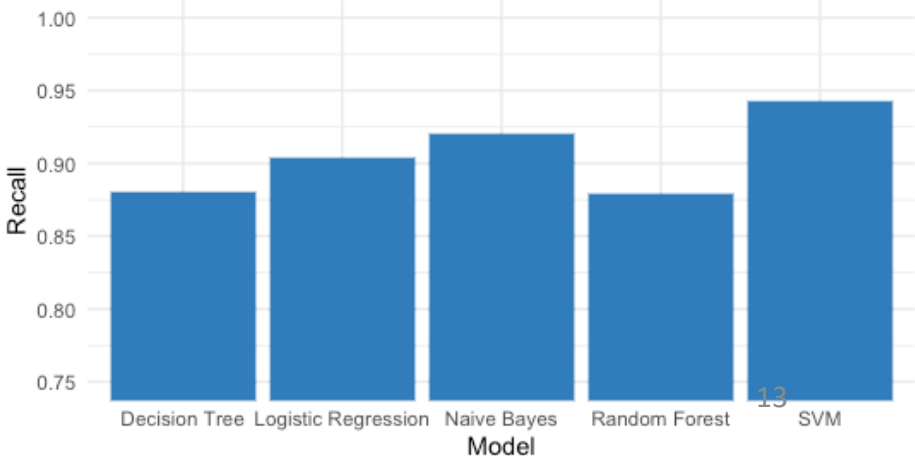
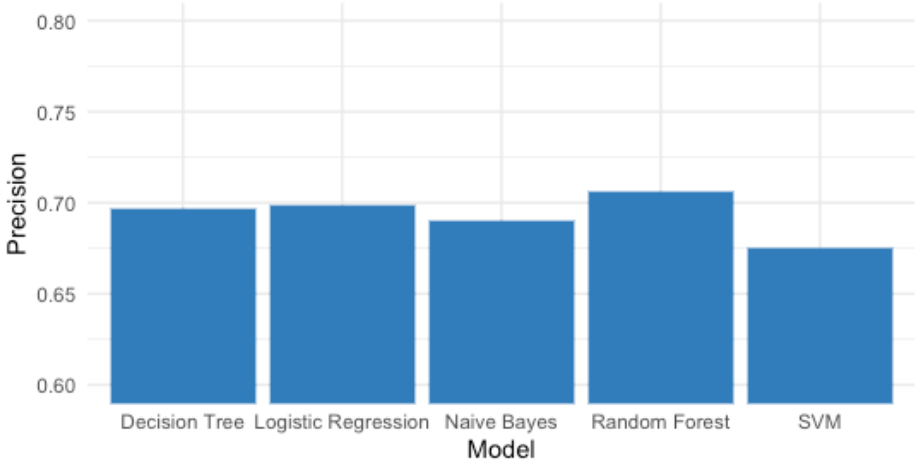
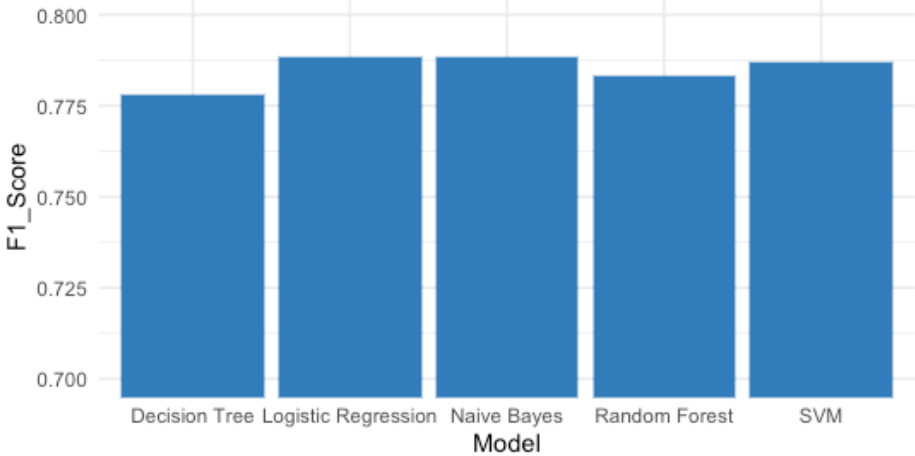
# Results

In termini di **F1-Score** le performance dei modelli sono molto simili, e i modelli che ottengono il valore più alto sono la **regressione logistica** e il **Naïve Bayes**.

Tutti i modelli ottengono elevati valori di **Recall** (la proporzione di churners correttamente classificati), e tra tutti spicca la **Support Vector Machine** che ottiene una recall di oltre il 94%.

I valori di **Precision** (il numero di veri churner tra quelli classificati dal modello) sono più bassi, ed il modello con precision maggiore è la **Random Forest**.

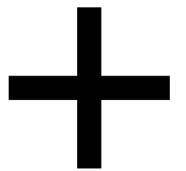
Model	Precision	Recall	F1-Score
Decision Tree	0.6964	0.8803	0.7776
Logistic Regression	0.6984	0.9039	<b>0.7880</b>
Naïve Bayes	0.6894	0.9195	<b>0.7880</b>
Random Forest	<b>0.7061</b>	0.8792	0.7832
Support Vector Machine	0.6749	<b>0.9428</b>	0.7867



# Churn Model – Data Driven Actions



Impostazione di una strategia di retention specifica per i clienti con maggiori probabilità di abbandono.



Impostazione di una strategia di retention con selezione dei target costruita combinando i risultati del modello di churn con i risultati di altri modelli analitici, come il modello RFM.