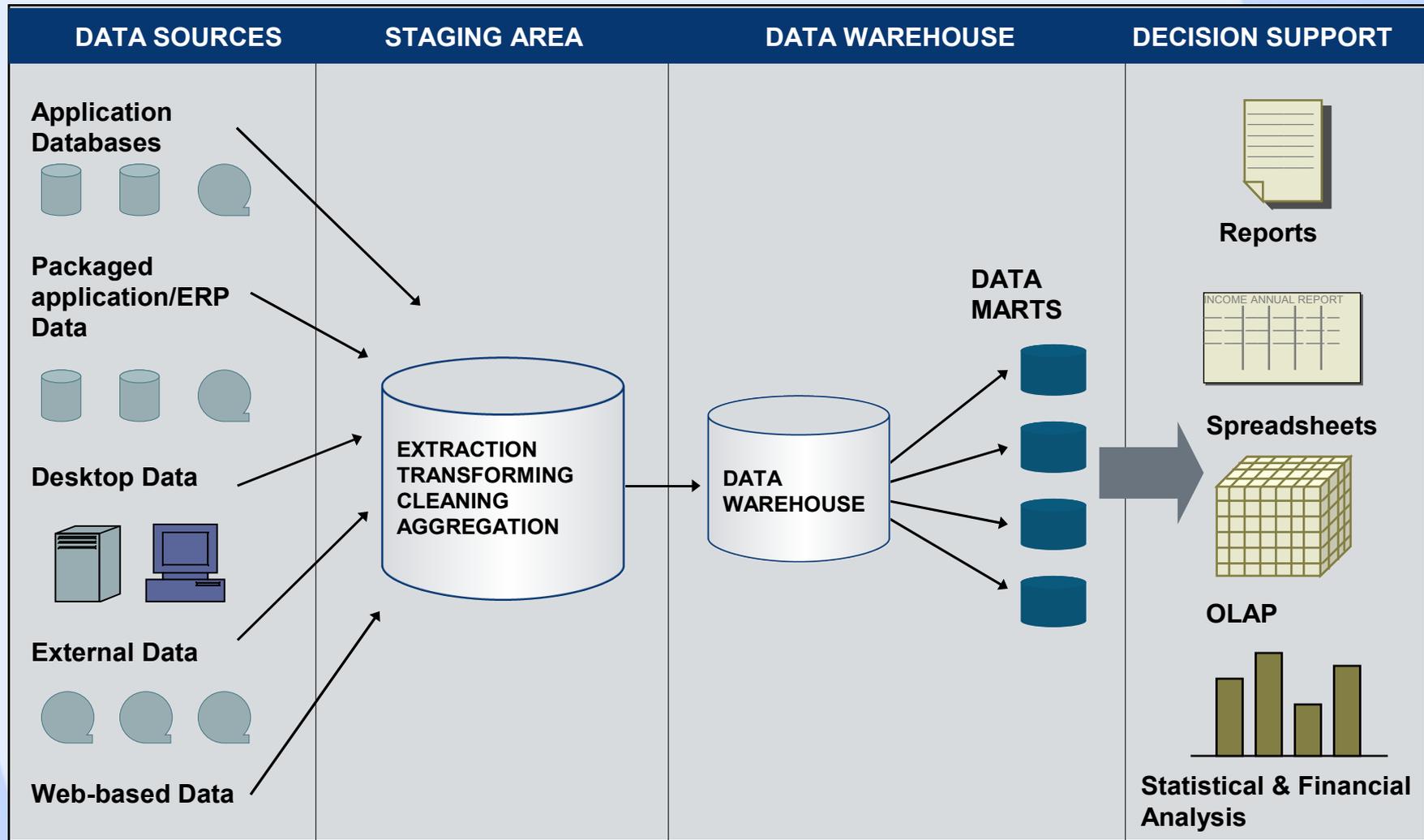
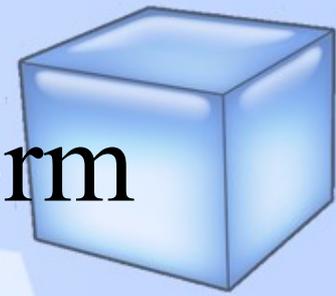


Business Intelligence & Data Mining In ambiente Retail

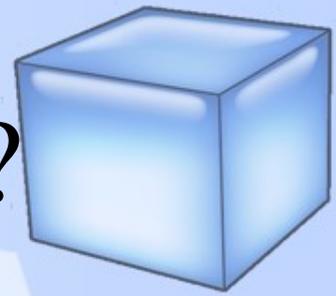


Business Intelligence Platform





Cos'è un data warehouse?

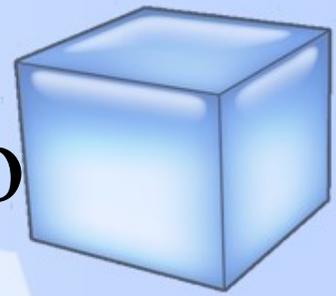


- Definito in molti modi
 - Un DB di supporto alle decisioni mantenuto separatamente dai DB operazionali dell'azienda.
 - Un processo di elaborazione dell'informazione che fornisce il basamento per dati consolidati e storici per l'analisi.
- “A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process.”

W.H. Inmon



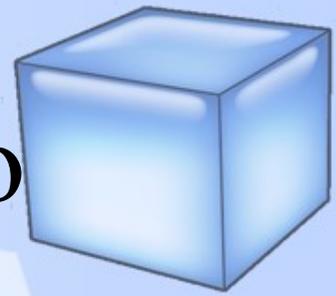
Il DW di Unicoop Tirreno



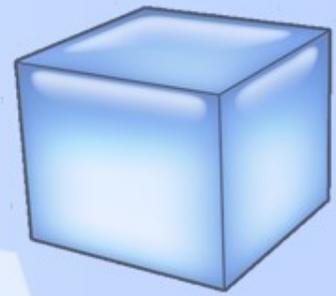
- 4 fatti:
 - Venduto
 - Pagamenti
 - Punti
 - Classificazione Clienti
- 18 dimensioni:
 - Data
 - Orario
 - Età
 - Prestito
 - Fasciaspesa
 - Tipologia
 - Percentuale
 - Sacchetto
 - Negozio
 - Articolo
 - Cliente
 - Promodettaglio
 - Svadettaglio
 - Pagamento
 - Salvatempo
 - Rilettura
 - Punto
 - lineaprezzo



Il DW di Unicoop Tirreno



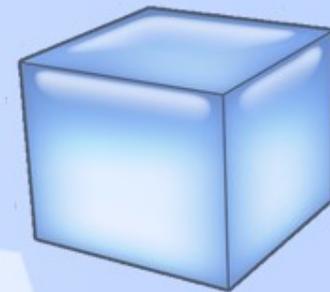
- È attualmente caricato con più di due anni di scontrini (Gennaio 2007 – Marzo 2009) di 110 negozi.
- 1Mld di righe totali per circa 150 GB di memoria
- L'aggiornamento avviene ogni settimana con un tempo di elaborazione di 3 ore macchina circa.
- Server usato: HP, 8 processori, 32GB memoria



Churn Analysis

Definizione del concetto di abbandono e
creazione di modelli previsionali

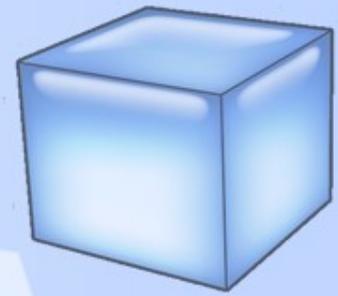
Sommario



- Introduzione
- Preparazione dei dati
- Modelli
- Valutazione dei risultati
- Scenario d'uso
- Conclusioni



Introduzione



Obiettivo:

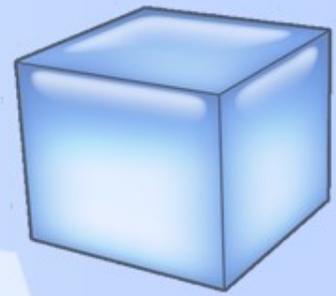
- definire opportunamente il concetto di fedeltà del cliente
- fornire strumenti previsionali in grado di stimare la fedeltà dei clienti nel futuro prossimo:
- individuare:
 - clienti che diminuiranno la loro attività
 - clienti che cesseranno definitivamente il rapporto

Finalità:

- arricchire la conoscenza dei propri clienti
- innescare meccanismi di *customer redemption* focalizzati.



Abbandono

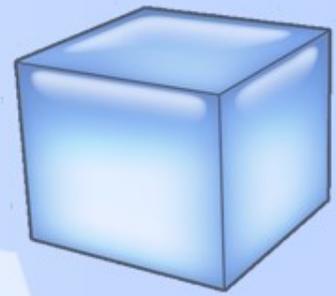


Definizione del problema:

- Stimare la probabilità di abbandono dei clienti a partire dalle informazioni presenti nel data warehouse:
 - dati di vendita
 - dati anagrafici

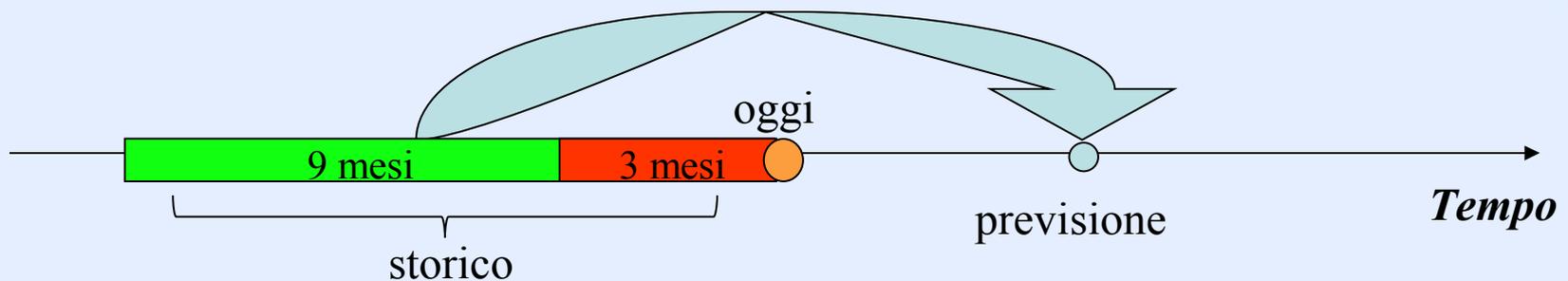
Definizione di abbandono:

- Un cliente è a rischio di abbandono quando evidenzia un netto calo nelle sue misure di spesa:
 - visite
 - volumi di spesa
 - battute di cassa

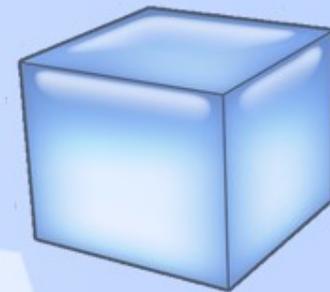


Analisi previsionale

- Raccolta dei dati storici per l'estrazione di:
 - Variabili di vendita e anagrafiche, i predittori (periodo verde)
 - Variabili obiettivo (periodo rosso)
- Costruzione di un modello predittivo
 - Addestrato in modo opportuno su dati storici
 - Utilizzabile per ottenere informazioni previsionali



Preparazione dati – osservazioni (periodo verde)



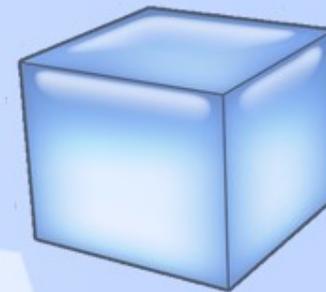
Si sono estratte dal data warehouse, per il periodo di 9 mesi (Dicembre 2006 – Agosto 2007) le seguenti informazioni:

- Dati anagrafici (sesso, età, professione etc.)
- Dati di spesa
 - Globale
 - Settori specifici: fresco, carni, pesce, ortofrutta
 - Pesata (abbattimento no-food)
- Trend di spesa:
 - Tipologia cliente (per ogni mese)
 - Regressione spese
 - Regressione spesa
 - Regressione battute



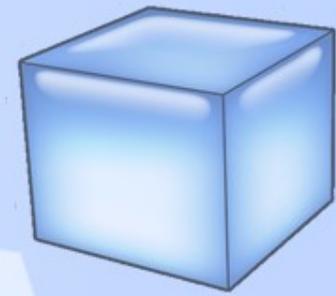


Preparazione dati – target (periodo rosso)



- Si sono estratte dal data warehouse, per il periodo di 3 mesi (Settembre 2007 – Novembre 2007) le seguenti informazioni:
 - Numero di spese
 - Variazione di spesa rispetto al periodo verde
 - Volume di spesa
 - Battute di cassa
 - Numero di visite

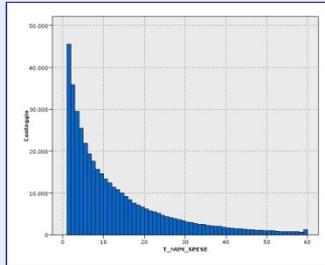
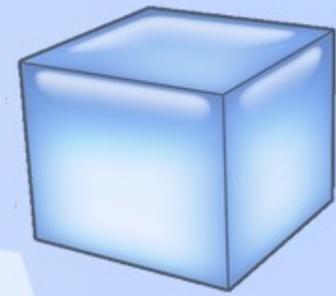
Dataset



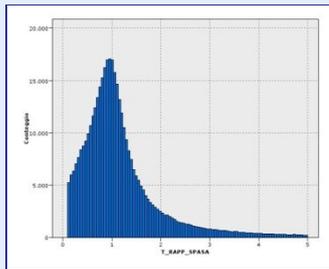
- Il dataset così ottenuto presenta una riga per ogni cliente che ha effettuato almeno una spesa nei nove mesi di osservazione. In tutto abbiamo ottenuto:
 - 517.000 righe
 - 47 attributi

Predittori Anagrafici	Predittori di spesa	Predittori di trend	Variabili target
CLIENTE_ID	DATA_ULTIMA_SPESA	TIPOLOGIA_01	T_NUM_SPESE
SESSO	NUM_SPESE	TIPOLOGIA_02	T_RAPP_SPESE
STATO_CIVILE	SPESA_TOT	TIPOLOGIA_03	T_RAPP_SPESA
PROFESSIONE	SPESA_TOT_PESATA	TIPOLOGIA_04	T_RAPP_BATTUTE
TITOLO_STUDIO	SPESA_MEDIA	TIPOLOGIA_05	
PROVINCIA	SPESA_MEDIA_PESATA	TIPOLOGIA_06	
REGIONE	BATTUTE	TIPOLOGIA_07	
ANNO_SOCIO	FRESCHI_TOT	TIPOLOGIA_08	
FASCIA_ANNO_SOCIO	FRESCHI_SPESE	TIPOLOGIA_09	
FL_INVIO_RIVISTA	CARNI_TOT	TIPOLOGIA_MEDIA	
COD_NEGOZIO	CARNI_SPESE	TIPOLOGIA_ZERI	
ETA	PESCE_TOT	REGR_NUM_SPESE	
ETA_FASCIA	PESCE_SPESE	REGR_SPESA	
	ORTOFRUTTA_TOT	REGR_SPESA_PESATA	
	ORTOFRUTTA_SPESE	REGR_BATTUTE	

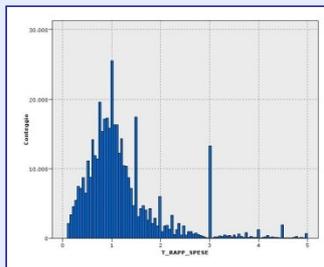
Funzioni Obiettivo



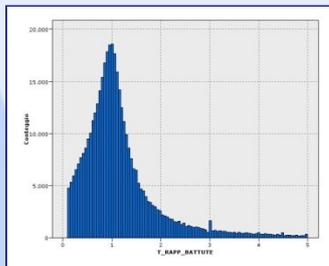
NUM_SPESE: spese del cliente nel periodo target



RAPP_SPESE: rapporto tra il numero delle spese del periodo target e quello del periodo d'osservazione

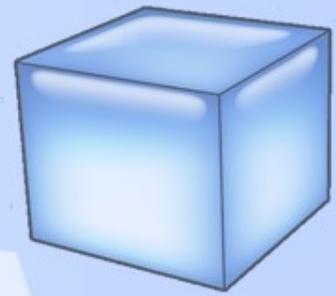


RAPP_SPESA: rapporto tra la spesa del periodo target e quella del periodo d'osservazione

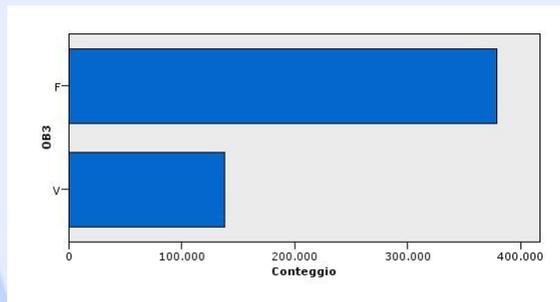


RAPP_BATTUTE: rapporto fra le battute di cassa del periodo target e quelle del periodo d'osservazione

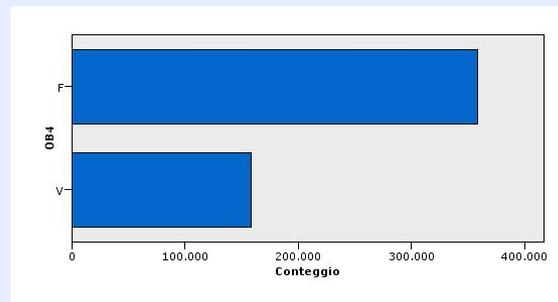
F. Obiettivo – Soglie



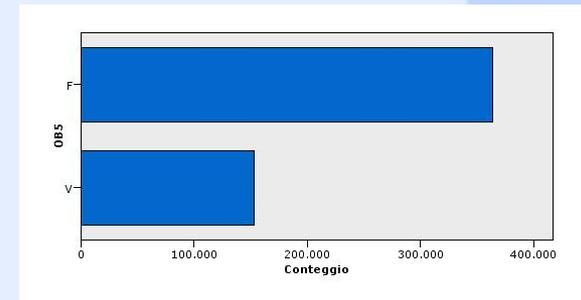
- Scelta una soglia di allarme per indicare un possibile cliente defezionario i rapporti si trasformano in tre indicatori di abbandono
- Abbiamo scelto come soglia una diminuzione sulle 3 misure del 50%
- Otteniamo le seguenti distribuzioni:



OB1: RAPP_SPESE

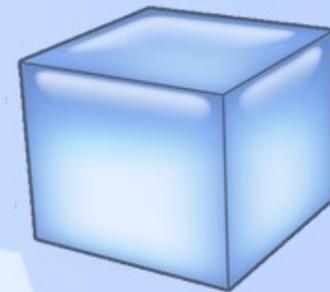


OB2: RAPP_SPESA



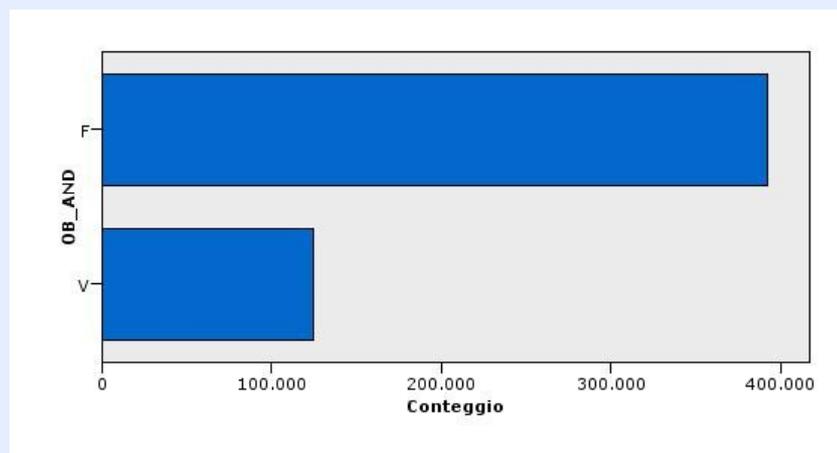
OB3: RAPP_BATTUTE

F. Obiettivo – Sintesi

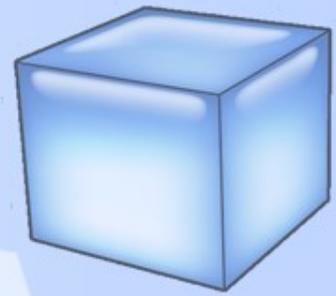


Per la funzione obiettivo finale si è deciso di considerare come potenziali defezionari tutti i clienti che superato la soglia di allarme, in ognuno dei tre indicatori OB1, OB2, OB3:

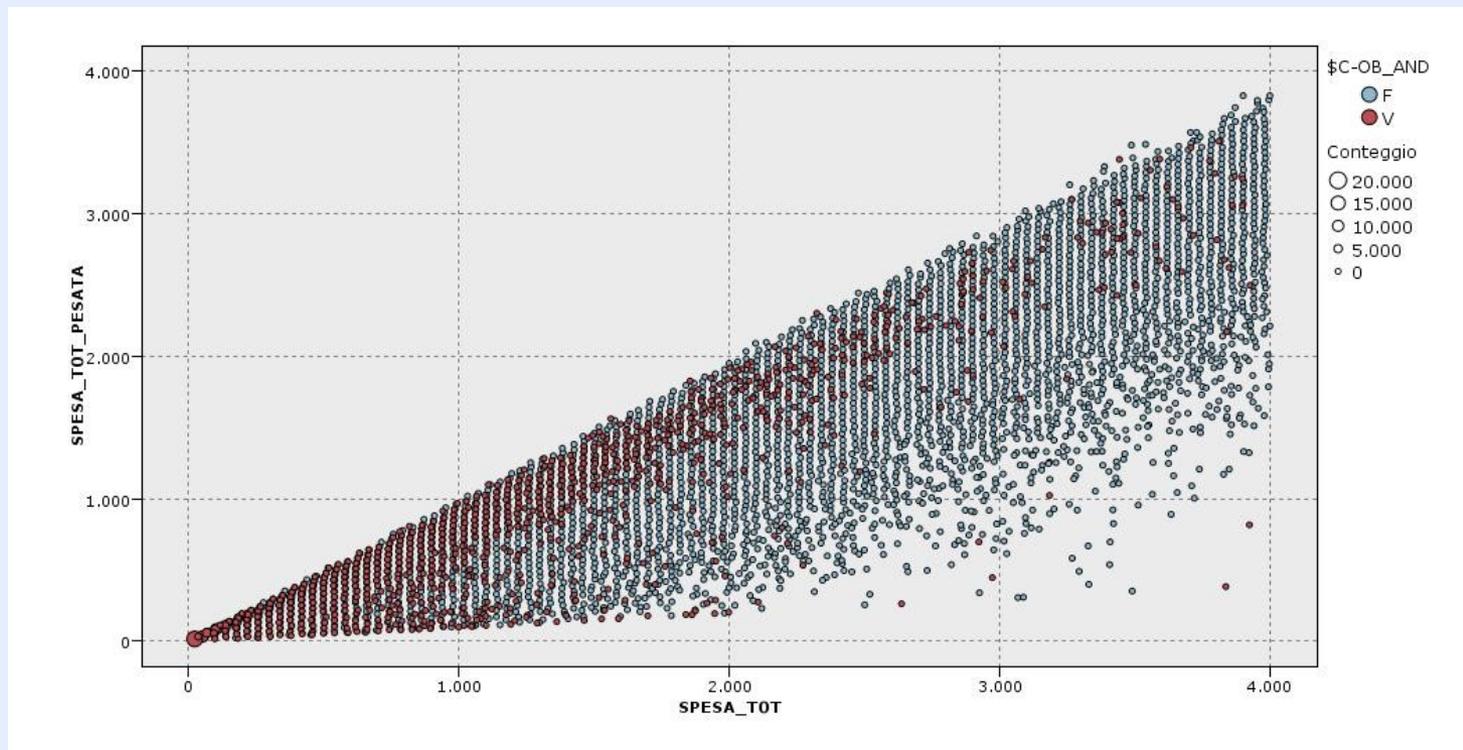
OB_AND: OB1 and OB2 and OB3



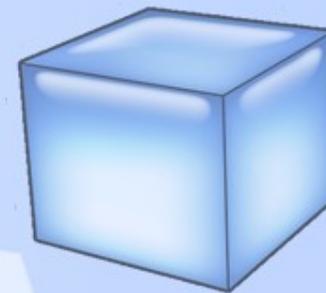
Modello previsionale e Risultati



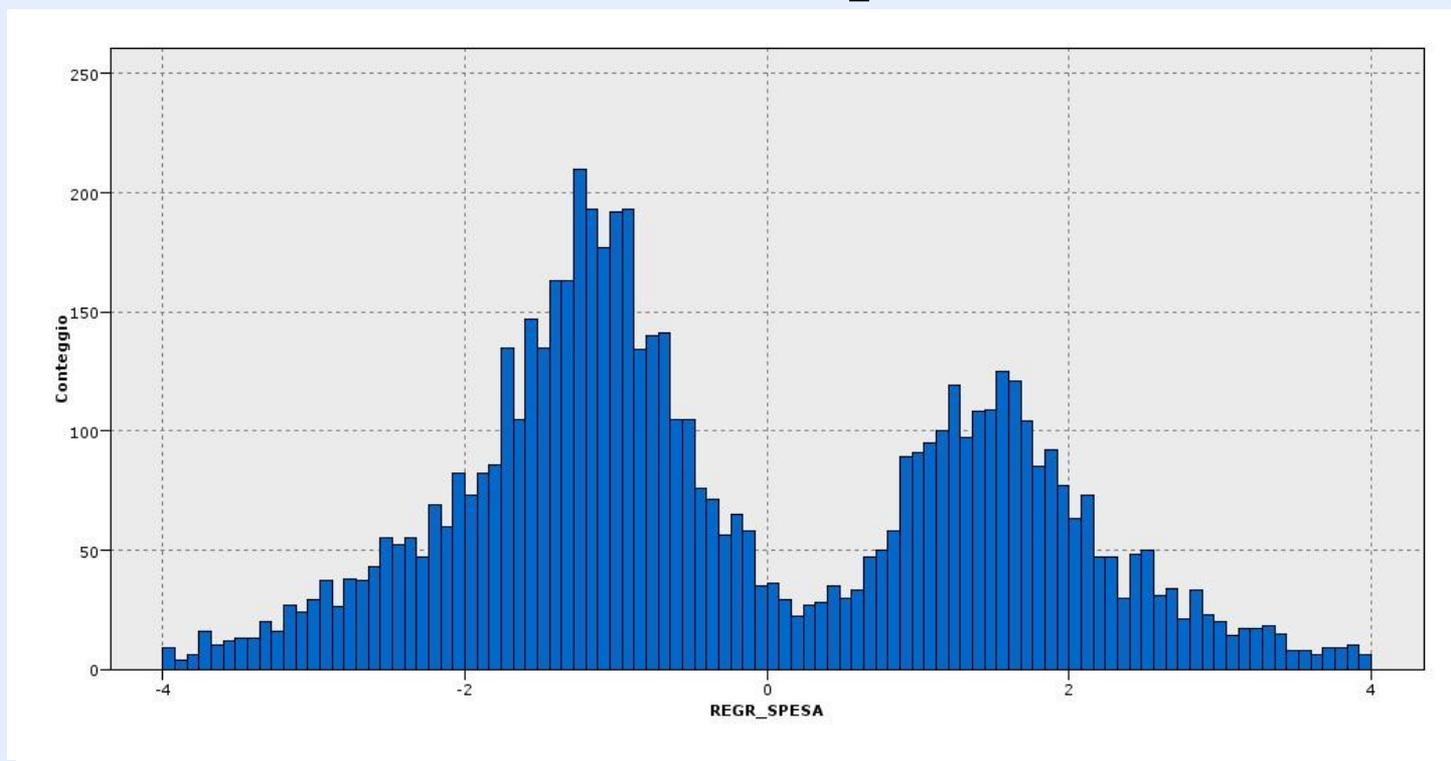
- Distribuzione spesa totale vs. spesa pesata



Modello previsionale e Risultati



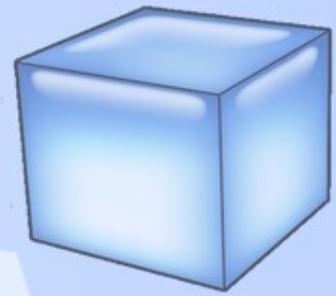
- Distribuzione trend di spesa



Trend dei clienti con spesa totale > 400€



Modello previsionale e Risultati



- Esempio di regole associative:

se REGIONE = TOSCANA

e NUM_SPESE \leq 128

e TIPOLOGIA_01 = 7

e TIPOLOGIA_09 = 0

e TIPOLOGIA_ZERI $>$ 2

e REGR_BATTUTE \leq -0,98

allora V (confidenza 82,8%)

se DATA_ULTIMA_SPESA $>$ 183

e NUM_SPESE \leq 21

e TIPOLOGIA_ZERI $>$ 1

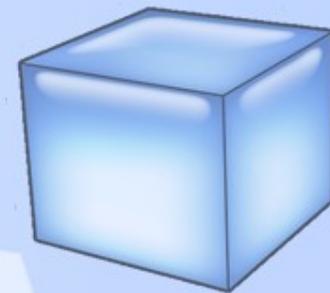
e REGR_NUM_SPESE \leq -0,02

e REGR_BATTUTE \leq -0,98

allora V (confidenza 92%)

Modello previsionale

Risultati Globali

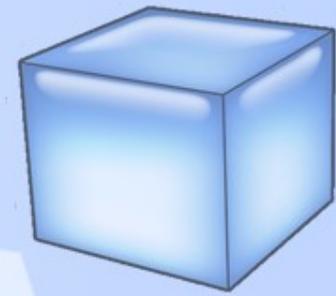


- Correttezza generale del modello:
 - 81.06% sul training set (70% del dataset, 360.000 righe)
 - 80.94% sul test set (30% del dataset, 155.000 righe)
- Matrici di confusione:

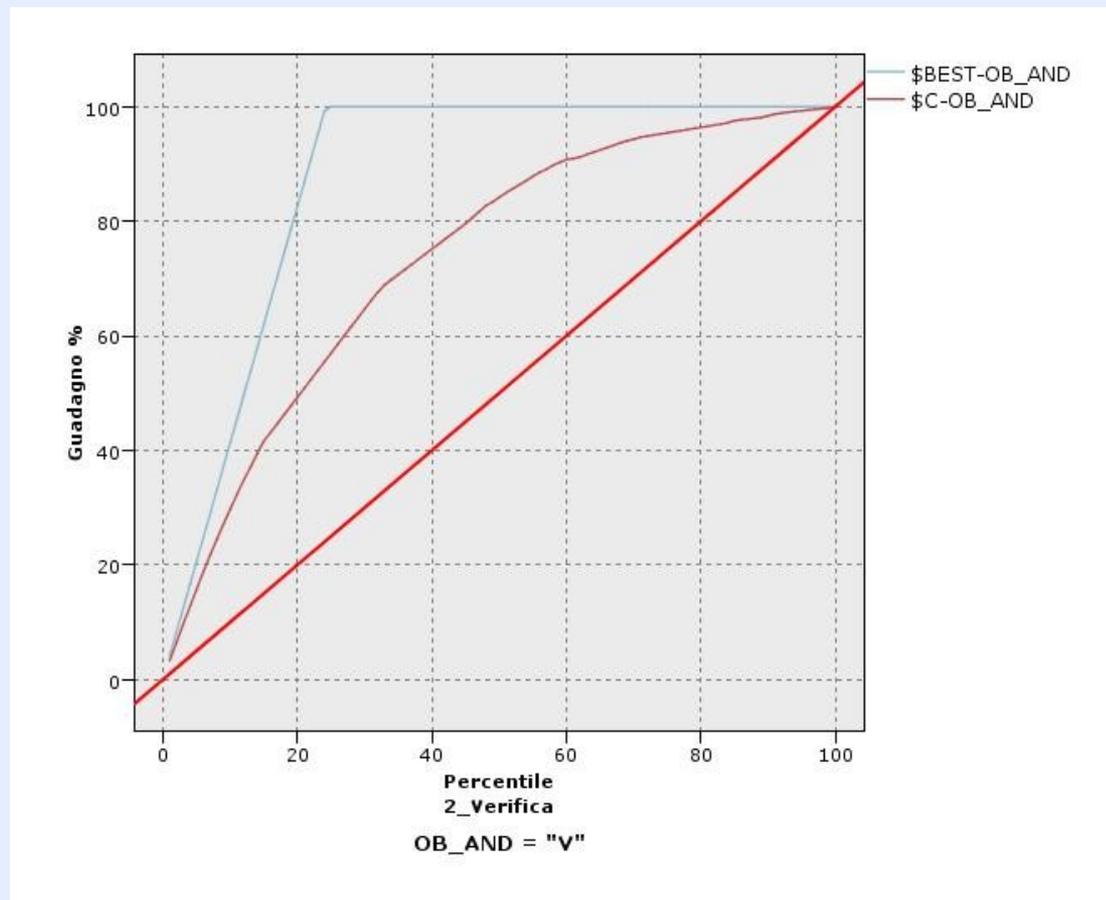
Valori Predetti

		Valori Predetti					
		Training Set	F	V	Test Set	F	V
Valori Reali	F		256.608	17.920		110.029	7.767
	V		50.540	36.466		21.855	15.734

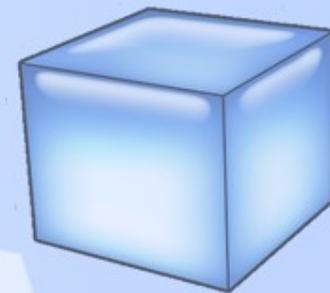
Modello previsionale Risultati Globali



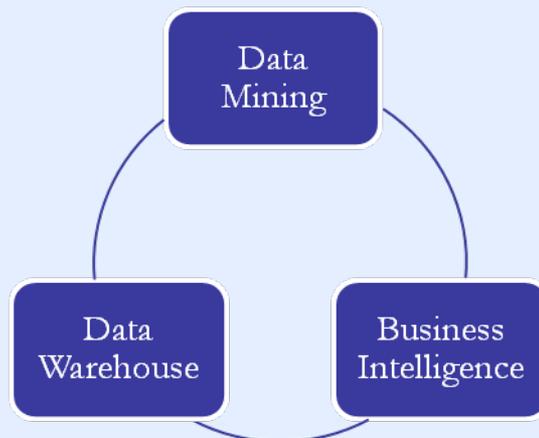
- Lift chart



Scenario d'uso – Esempio



- Creare un ambiente aperto e dinamico nel quale i dati forniti dal data warehouse vengono elaborati e trasformati in modelli di tipo previsionale.



- I modelli previsionali possono essere usati per arricchire il data warehouse, innestando un circolo virtuoso di informazioni utilizzabili direttamente in ambienti di Business Intelligence.