

Analisi del carrello della spesa

Analizzare il comportamento dei clienti di un supermarket, prendendo in considerazione le diverse combinazioni dei prodotti in vendita (o di un gruppo di essi) negli scontrini emessi alle casse.

L'analisi può evidenziare quali siano le combinazioni più frequenti per effetto di sconti promozionali, campagne pubblicitarie, presenza di promoter tra le corsie del supermarket oppure disposizione di scaffali vicini con prodotti diversi (*layout* del supermarket).

In un contesto di commercio elettronico nel Web, queste analisi consentono di inserire nel sito Web forme di consigli collaborativi tra utenti (*social* o *collaborative recommendation*): durante la visualizzazione di un prodotto, il sito raccomanda altri prodotti sulla base di quello che altri clienti hanno acquistato insieme con il prodotto attualmente selezionato. Per esempio, con frasi del tipo "Il 70% dei clienti che hanno acquistato il prodotto X, hanno anche acquistato ...", oppure in uno store di app per dispositivi mobili: "Gli utenti che hanno visionato questa applicazione, hanno anche visto ...".

Supponiamo di avere un database *supermarket*. Le tabelle che forniscono i dati per l'analisi del carrello della spesa sono essenzialmente:

prodotti (CodProd, Descrizione, PrezzoUnitario)

scontrini (NumScontrino, Data)

riguescontrini (ID, NumScontrino, CodProd, Qta)

In alcune situazioni, basate sull'uso delle *Carte Fedeltà* dei clienti, potrebbe essere utile avere, nella tabella *scontrini*, anche un riferimento al codice cliente, che possa richiamare i dati del cliente (provenienza geografica, genere M/F, oppure età) per effettuare analisi aggiuntive sul comportamento di diverse categorie di clienti.

La tabella *riguescontrini* rappresenta le righe di dettaglio degli scontrini: ciascuna riga contiene, come chiavi esterne, i riferimenti al numero dello scontrino e al codice del prodotto. Il campo ID è il progressivo di registrazione delle righe di dettaglio. Ci possono essere più righe con lo stesso valore nel campo *NumScontrino*.

L'esempio seguente mostra l'analisi riferita, per semplicità, ad un numero limitato di prodotti (10) e ad un numero limitato di righe di scontrini (100).

La seguente query SQL consente di ottenere il numero di occorrenze per ognuna delle possibili combinazioni di prodotti nello stesso scontrino.

```
SELECT S.CodProd AS Prodotto1, T.CodProd AS Prodotto2, Count(*) AS Occorrenze
FROM rignescontrini AS S INNER JOIN rignescontrini AS T
ON S.NumScontrino = T.NumScontrino
WHERE S.CodProd < T.CodProd
GROUP BY S.CodProd, T.CodProd
```

Per realizzare questo calcolo vengono utilizzate due copie della stessa tabella *riguescontrini*, con un *join* per valori uguali di *NumScontrino*.

I dati ottenuti in MySQL possono essere esportati in formato CSV (per esempio, utilizzando il programma *phpMyAdmin*) per essere poi utilizzati con un foglio di *Excel*, che consente di fare ulteriori calcoli, ma anche rappresentazioni grafiche.

La figura seguente mostra la tabella dei risultati della query precedente, ordinati secondo i valori decrescenti delle occorrenze, per evidenziare le combinazioni più frequenti, insieme a un *grafico a bolle*, che illustra in modo efficace le informazioni ottenute dall'analisi. La bolla corrispondente al valore più alto è stata evidenziata con un colore più scuro.

