

Business Intelligence

Data Warehouse per la società GiocoXtutti

Descrizione del problema

La società GiocoXTutti gestisce diversi negozi distribuiti sul territorio italiano che vendono giochi di ogni tipo (gioco di logica, di ruolo, di realtà interattiva, da tavolo, etc.) e per tutte le età. Ad ogni vendita è associata l'emissione di uno scontrino fiscale che registra la lista dei giochi acquistati, quantità e incasso parziale (per singolo gioco) e complessivo, la data in cui è stata realizzata la vendita, la modalità di pagamento e se il cliente è fidelizzato viene riportato il codice della tessera fedeltà. Le modalità di pagamento sono: contante, assegno (solo per importi superiori a 500 Euro), bancomat o carta di credito. Per i clienti fidelizzati sono note le seguenti informazioni: nome, cognome, data di nascita, indirizzo di residenza, indirizzo e-mail, numero di telefono, stato civile, ed eventualmente, per ogni figlio la data di nascita.

I negozi effettuano in alcuni periodi dell'anno delle campagne promozionali che consistono nell'applicazione di uno sconto percentuale (variabile) sul prezzo di listino dei diversi giochi. Per esempio a dicembre 2014 erano attive diverse campagne promozionali, una delle quali era denominata "Regalo Gioco Natale 2014" e permetteva di regalare un gioco ad un cliente, purchè fosse non fidelizzato, con percentuale di sconto variabile dal 30 al 50% in base al numero di figli del nuovo cliente.

Diversi negozi oltre a vendere giochi, organizzano diverse sessioni di attività ricreative di tipo competitivo e caratterizzate da obiettivi e regole rigorosamente definiti (e.g., giochi di società, giochi da tavolo) nei quartieri cittadini delle diverse città italiane. I ragazzi possono partecipare a più sessioni di gioco organizzati nello stesso quartiere o in quartieri diversi. Per alcuni giochi è necessario avere una o più conoscenze/abilità (e.g., lingua inglese, capacità di suonare uno strumento). Ogni sessione di gioco è gestita e organizzata da un numero variabile di moderatori in base al numero dei partecipanti e al livello di complessità del gioco.

Attualmente ogni negozio ha due base di dati indipendenti per gestire al meglio le vendite e organizzare le diverse sessioni di attività ricreative, ma la società è interessata a progettare e realizzare un data warehouse che offra una visione d'insieme. L'azienda, durante il processo di ETL vuole arricchire il proprio data warehouse con dei dati che caratterizzano il gioco e che sono disponibili in un'altra sorgente dati. Nello specifico, per ogni gioco è possibile conoscere il tipo di gioco e il livello di complessità (in una scala da 1 a 5), la fascia di età consigliata per l'utilizzo del gioco e il materiale principale utilizzato per realizzare il gioco (e.g., carta, plastica, etc). Il data warehouse realizzato dovrà consentire di analizzare:

1. l'efficienza e l'efficacia di vendita dei propri negozi di giochi;
2. la partecipazione dei ragazzi alle diverse attività ricreative organizzate durante l'anno nei quartieri delle città italiane.

Per il punto (1) la società è interessata ad analizzare l'incasso medio giornaliero e l'incasso medio per oggetto venduto in funzione di:

- data, mese, bimestre, trimestre, quadrimestre, semestre e anno in cui è stata effettuata la vendita;
- giorno della settimana, giorno festivo/feriale e settimana in cui è stata effettuata la vendita;
- modalità di pagamento utilizzato per l'acquisto;
- sconto (inteso come percentuale di sconto applicato ad un acquisto di un nuovo giocattolo nel contesto di una specifica campagna promozionale. Nel caso di vendita effettuata al di fuori di una campagna promozionale, la percentuale di sconto è 0% nel contesto di una specifica vendita tradizionale), percentuale di sconto (i valori ammessi sono 0%, 10%, 20%, 30%, 40%, e 50%), campagna promozionale (e.g., "Regalo Gioco Natale 2014", "Promo primavera 2014", "Vendita tradizionale 2013") e anno solare della campagna promozionale;
- negozio, quartiere, città, regione e dimensione del negozio (piccolo, medio, grande) che ha realizzato la vendita;
- informazione sulla fidelizzazione o meno del cliente,
- caratteristiche del cliente fidelizzato: numero figli (0, <3, >3), fascia di età (<25, >25 and <40, >40), stato civile;
- gioco, tipo di gioco, livello di complessità del gioco,
- fascia di età consigliata per il gioco e materiale principale con cui è stato realizzato il gioco.

Per il punto (2) la società GiocoXTutti è interessata ad analizzare il numero medio di partecipanti per sessione di gioco, l'età media dei partecipanti, il numero medio di partecipanti per moderatore in funzione di:

- tipo di gioco e livello di complessità del gioco;
- mese, bimestre, semestre e anno;
- quartiere, città e regione in cui è organizzata la sessione di gioco;
- abilità dei giocatori: conoscenza della lingua inglese, capacità di suonare uno strumento musicale, etc..

Sono di seguito riportate **alcune** delle interrogazioni frequenti di interesse:

- a) Considerando solo gli acquisti dei clienti fidelizzati con più di 3 figli, visualizzare per ogni negozio e tipo gioco l'incasso totale mensile, l'incasso medio giornaliero in ogni mese, l'incasso totale mensile cumulativo dall'inizio dell'anno. Calcolare inoltre per ogni negozio e tipo gioco la percentuale di incasso mensile rispetto al totale mensile realizzato da tutti i negozi nella stessa città, separatamente per ogni tipo gioco**
- b) Considerando solo i giochi di ruolo (tipo_gioco="gioco di ruolo") in cui i partecipanti devono conoscere la lingua inglese e saper suonare uno strumento musicale, visualizzare per ogni città e anno, i quartieri dove sono state organizzate il numero massimo di sessioni di gioco, e il numero massimo di sessioni organizzate. Ordinare e assegnare un rank in ordine decrescente di numero di sessioni totali organizzate, separatamente per ogni anno.**
- c) Considerando solo i giochi con un livello di complessità pari a 5 e realizzati prevalentemente con carta di riciclo, separatamente per i negozi grandi, piccoli e medi, calcolare la percentuale di incasso derivante da ogni campagna promozionale rispetto al totale annuo (l'incasso totale annuo è quello derivante da tutte le campagne promozionali attive in un determinato anno). Effettuare l'analisi separatamente per i clienti fidelizzati e quelli non fidelizzati.**
- d) Considerando solo gli incassi del 2013 e 2014, per ogni coppia (quartiere del negozio, tipo_gioco), visualizzare l'incasso totale trimestrale, l'incasso medio per oggetto venduto. Effettuare l'analisi separatamente per i clienti fidelizzati e quelli non fidelizzati.
- e) Per ogni coppia (città, semestre) visualizzare il numero medio di partecipanti per sessione di gioco, l'età media dei partecipanti, il numero totale di moderatori.
- f) Per ogni negozio e per ogni campagna promozionale visualizzare l'incasso totale annuo e l'incasso totale medio per oggetto venduto separatamente per ogni modalità di pagamento.
- g) Per ogni regione visualizzare il numero totale di sessioni di gioco organizzate, il numero medio di partecipanti per sessione di gioco, l'età media dei partecipanti. Effettuare l'analisi separatamente per ogni anno.
- h) Visualizzare per ogni città in cui è sito almeno un negozio l'incasso totale mensile. Effettuare l'analisi separatamente per ogni materiale principale in cui è realizzato il gioco, tipo gioco e livello di complessità.

Progettazione

1. Il data warehouse conterrà le informazioni relative agli anni 2009, 2010, 2011, 2012, 2013, 2014. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:
 - Valori distinti di materiale di gioco:~15
 - Numero massimo di abilità richieste ai giocatori delle attività ricreative :~5
 - Valori distinti di tipo di gioco:~20
 - Valori distinti di gioco:~1000
 - Valori distinti di negozio:~100
 - Valori distinti di sconto:~50
2. Esprimere le interrogazioni frequenti **(a)**, **(b)**, **(c)** delle specifiche del problema utilizzando il linguaggio SQL esteso.
3. Considerando le caratteristiche del data warehouse realizzato e la cardinalità dei dati memorizzati nel data warehouse, decidere se e quali viste materializzate potrebbe essere utile definire al fine di ottimizzare i tempi di risposta delle interrogazioni proposte nelle specifiche del problema. Considerare tutte le interrogazioni proposte e non solo quelle risolte in SQL al punto 2. Motivare le scelte fatte.
4. Decidere come gestire la dinamicità (variazioni) dei dati all'interno delle dimensioni. Motivare le scelte fatte.