**Power BI**

**Lab 2 – ETL**

**Scopo del laboratorio:**

**▪ acquisizione delle tabelle utili alle analisi**

**▪ profilazione, preparazione, pulizia e ristrutturazione delle stesse**

La fase di ETL è fondamentale al fine di ridurre il volume dei dati, rinominare campi e tabelle, pulire valori errati ed inconsistenze dei dati, ristrutturare le tabelle coerentemente alle best practice di modellazione dati.

**Scenario**

BikesWorld è un’azienda che commercializza diversi modelli di biciclette, abbigliamento specifico e parti di ricambio in tutto il mondo.

Occorre sviluppare una soluzione di BI che consenta all’azienda di analizzare i dati storici e monitorare l’andamento corrente delle vendite.

L’azienda è interessata a quantificare le vendite in termini di Importo Totale, Numero Ordini, Numero di transazioni e Profitto.

**È necessario poter analizzare le metriche per prodotto, reseller e area geografica.**

**In particolare, è necessario poter esplorare le metriche rispetto alla gerarchia prodotto (Categoria, Sottocategoria e Prodotto)**

**Requisiti non funzionali**

Non è necessario un report real-time.

È necessario minimizzare il tempo di caricamento delle visualizzazioni (garantire le migliori prestazioni possibili).

**Individuazione delle sorgenti dati e utenza**

Sorgente dati: Azure SQL DB

▪ Server name: cnclcserver.database.windows.net

▪ Database: BikesWorld

▪ User: cn

▪ Password: lc\_pass1!!

Estrazione dati in .csv contenente l’anagrafica reseller e le rispettive informazioni geografiche

Le sorgenti dati sono un DWH sul cloud di Azure e un estrazione dati .csv

*Cosa devi fare:*

1. *Arricchire la tabella Product delle informazioni richieste (Sottocategoria, Categoria)*
2. *Acquisire ed esplorare i dati reseller*

**Task 1: Acquisizione dati da file .csv**

1. Importa i dati del file .csv inerente ai reseller nel report **Analisi Vendite.**
2. Esplora i dati.
3. Applica modifiche ai dati se necessarie.

**Task 2: Denormalizza l’entità Product**

Al fine di soddisfare le necessità di esplorazione del report, è necessario arricchire la tabella Product delle informazioni inerenti al prodotto.

Per ciascun prodotto bisogna “recuperare” la rispettiva sottocategoria e categoria.

1. Apri il Query Editor (se non lo hai già aperto!)

A screenshot of a computer

Description automatically generated

1. Individua nella tabella Product la colonna link ProductSubcategory

Se hai già eseguito delle operazioni di data preparation su questa tabella è plausibile che tu non veda la colonna ProductSubcategory in quanto rimossa.

In questo caso, modifica gli step applicati alla tabella Product. In particolare, controlla gli step di rimozione colonne.

A screenshot of a computer

Description automatically generated

***Potreste avere una sequenza di operazioni differente in base alla logica che avete applicato!***

*In ogni caso, gli step interessati alla modifica sono quelli medianti i quali avete rimosso colonne.*

*Occorre modificarli per recuperare il campo utile.*

*In questo esempio ci sono due operazioni di tipo* ***Rimosse altre colonne.*** *È necessario modificarne la logica per ciascuno di essi ‘a cascata’.*

*Ricorda che i passaggi applicati sono sequenziali!*

*Modifica step* ***Rimosse alter colonne***

A screenshot of a computer

Description automatically generated

*Modifica step* ***Rimosse alter colonne 1***

A screenshot of a computer

Description automatically generated

1. Individuata (o ottenuta) la colonna ProductSubcategory, puoi eseguire l’espansione della stessa
2. Esplora il risultato ottenibile.
3. A questo punto, continua nelle operazioni di ETL al fine di ottenere una tabella Product contenente le colonne **Subcategory** e **Category**.

**A screenshot of a computer

Description automatically generated**

1. Terminate le operazioni Chiudi e Applica
2. La ristrutturazione della tabella Product è conclusa

*Note*

*Al momento della compilazione della traccia i dati sono aggiornati al 29/02/2024. La visualizzazione è molto semplice, nei prossimi lab completeremo il report continuando con la fase di ETL, progettando un modello logico, arricchendo il dataset di nuove informazioni e facendo storytelling!*