**Power BI**

**Lab 1 – ETL**

**Scopo del laboratorio:**

**▪ acquisizione delle tabelle utili alle analisi**

**▪ profilazione, preparazione, pulizia e ristrutturazione dati**

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.

**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!!

La sorgente dati è un DWH sul cloud di Azure.

*Cosa devi fare:*

1. *Creare una connessione alla sorgente dati*
2. *Acquisire i dati utili alle analisi (prodotti e vendite) scegliendo la corretta modalità di storage in base ai requisiti*
3. *Esplorare i dati*

**Task 1: Connessione Dati**

1. Crea una cartella dove salvare i tuoi report

2. Apri Power BI Desktop

3. Posizionati su **File** e poi **Salva con nome**.

A screenshot of a computer

Description automatically generated

4. Salva il report come Analisi Vendite

5. In file > Opzioni e impostazioni > Opzioni

A screenshot of a computer

Description automatically generated

Assicurati di impostare in **File Corrente > Caricamento dati** le opzioni come mostrato di seguito:

A screenshot of a computer

Description automatically generated

**Discuterete a lezione l’impatto delle impostazioni appena definite e quando conviene attivare la funzionalità di Business Intelligence per le gerarchie temporali.**

5. Nella barra multifunzione > tab **Home** di Power BI Desktop > clicca **Recupera Dati** e seleziona **SQL Server**

A screenshot of a computer

Description automatically generated

6. Inserisci il nome del server e il nome del database

7. Scegli la modalità di storage che ritieni opportuno e clicca OK

***Quale modalità di storage sceglieresti dati i requisiti indicati?***

8. Inserisci **Nome utente** e **Password** indicando come tipologia di connessione **Database,** poi clicca **Connetti**.

Nome utente e password sono indicati in *Individuazione delle sorgenti dati e utenza*

A screenshot of a computer

Description automatically generated

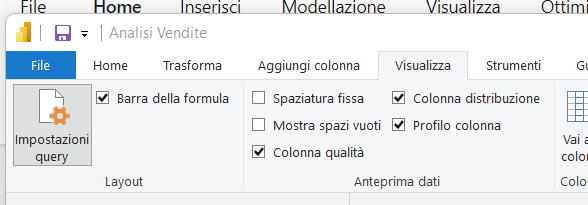
9. Seleziona le tabelle utili alle analisi, poi **Trasforma Dati**

A screenshot of a computer

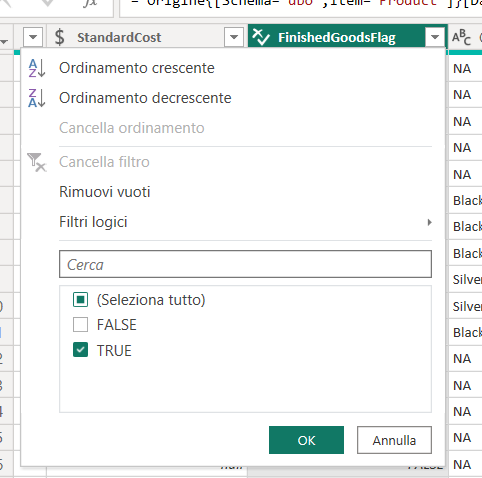
Description automatically generated

**Task 2: Esplorazione e ristrutturazione dati**

**Nel Power Query Editor abilita la profilazione dei campi**

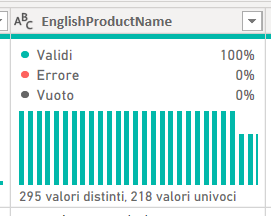


1. Esplora la tabella Product
   1. Rimuovi i campi non utili alle analisi (li devi dedurre dal contesto e dai requisiti!!)
   2. Filtra i soli prodotti finiti



Dopo l’applicazione del filtro la tabella ha 397 righe.

* 1. Analizza la colonna EnglishProductName (è necessario aver abilitato la profilazione!!)



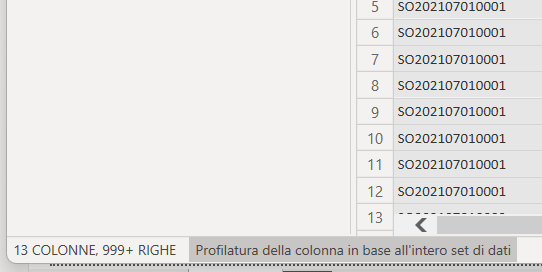
Cosa desumi dal conteggio dei valori distinti e univoci? Esiste un problema? Se sì, quale? *Ci sono nomi duplicati di prodotti*

* 1. Rinomina i campi se lo ritieni opportuno
  2. Soluzione plausibile per la tabella Product

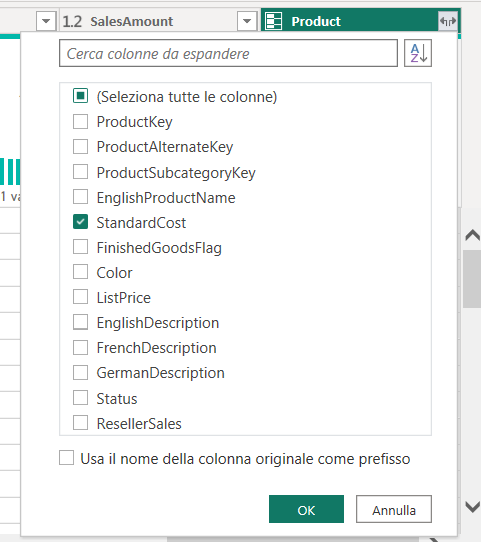
A screenshot of a computer

Description automatically generated

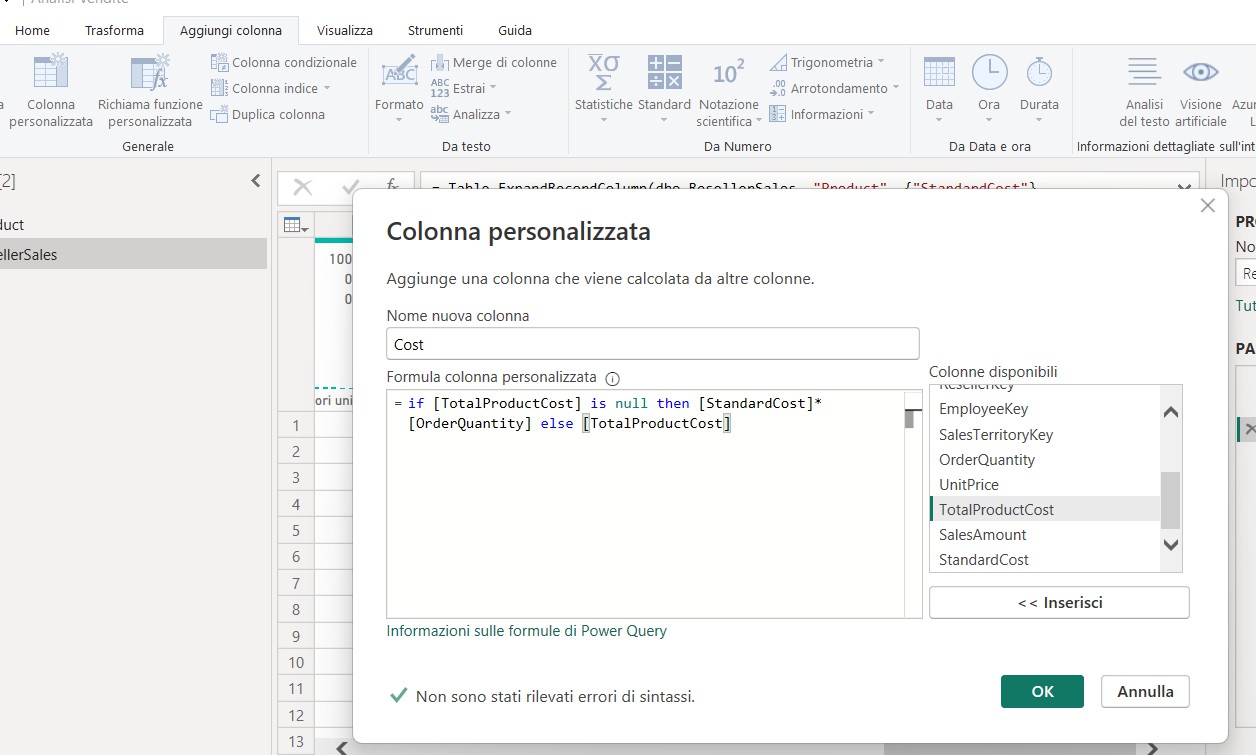
1. Esplora la tabella ResellerSales
   1. Controlla che il data type sia coerente alle informazioni contenute nei campi. In caso contrario modificalo opportunamente.
   2. Abilita la profilazione dell’intero dataset



* 1. Analizza la colonna **TotalProductCost**
  2. Aggiungi una nuova colonna calcolata per determinare il profitto. Il calcolo del profitto deve essere basato sul Costo Standard del prodotto (StandardCost in Product) quando il Costo del venduto (TotalProductCost in ResellerSales) è null.
     1. Recupera il campo StandardCost facendo una merge (in questo caso, puoi eseguire direttamente una merge oppure utilizzare il campo link). Utilizzando la colonna link



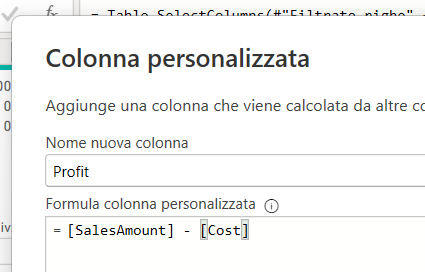
* + 1. Aggiungi una colonna personalizzata. Denomina la colonna **Cost**



La formula utilizzata è la seguente:

if [TotalProductCost] is null then [StandardCost]\*[OrderQuantity] else [TotalProductCost]

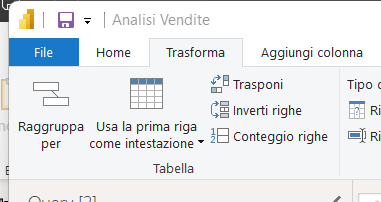
* + 1. Assegna il corretto data type alla colonna appena calcolata e rimuovi StandardCost e TotalProductCost
    2. Aggiungi una nuova colonna calcolata e calcola il profitto (assegna poi il corretto data type)



* 1. La preparazione della tabella Sales è conclusa.

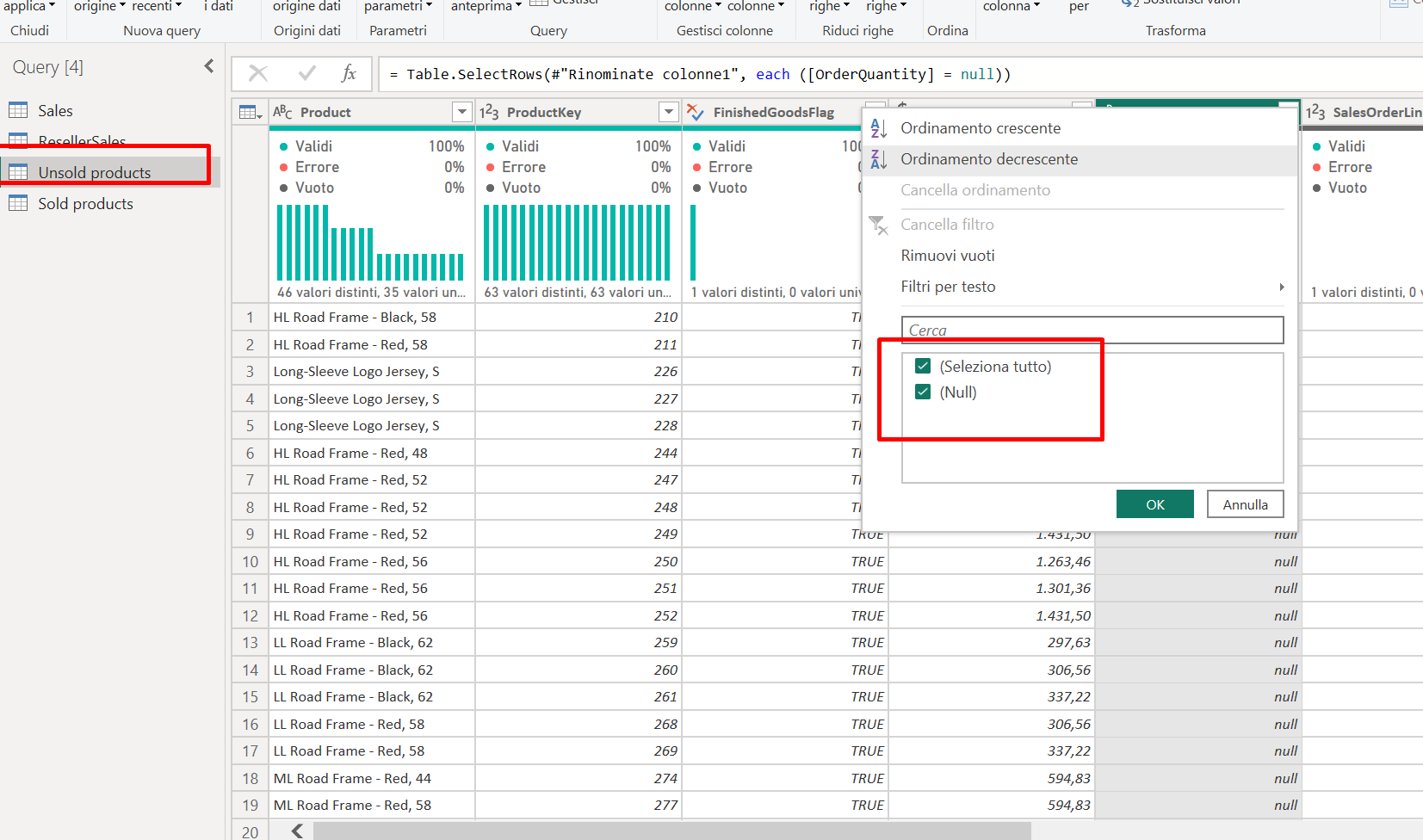
**Task 3: Esplorazione dati di vendita e considerazioni**

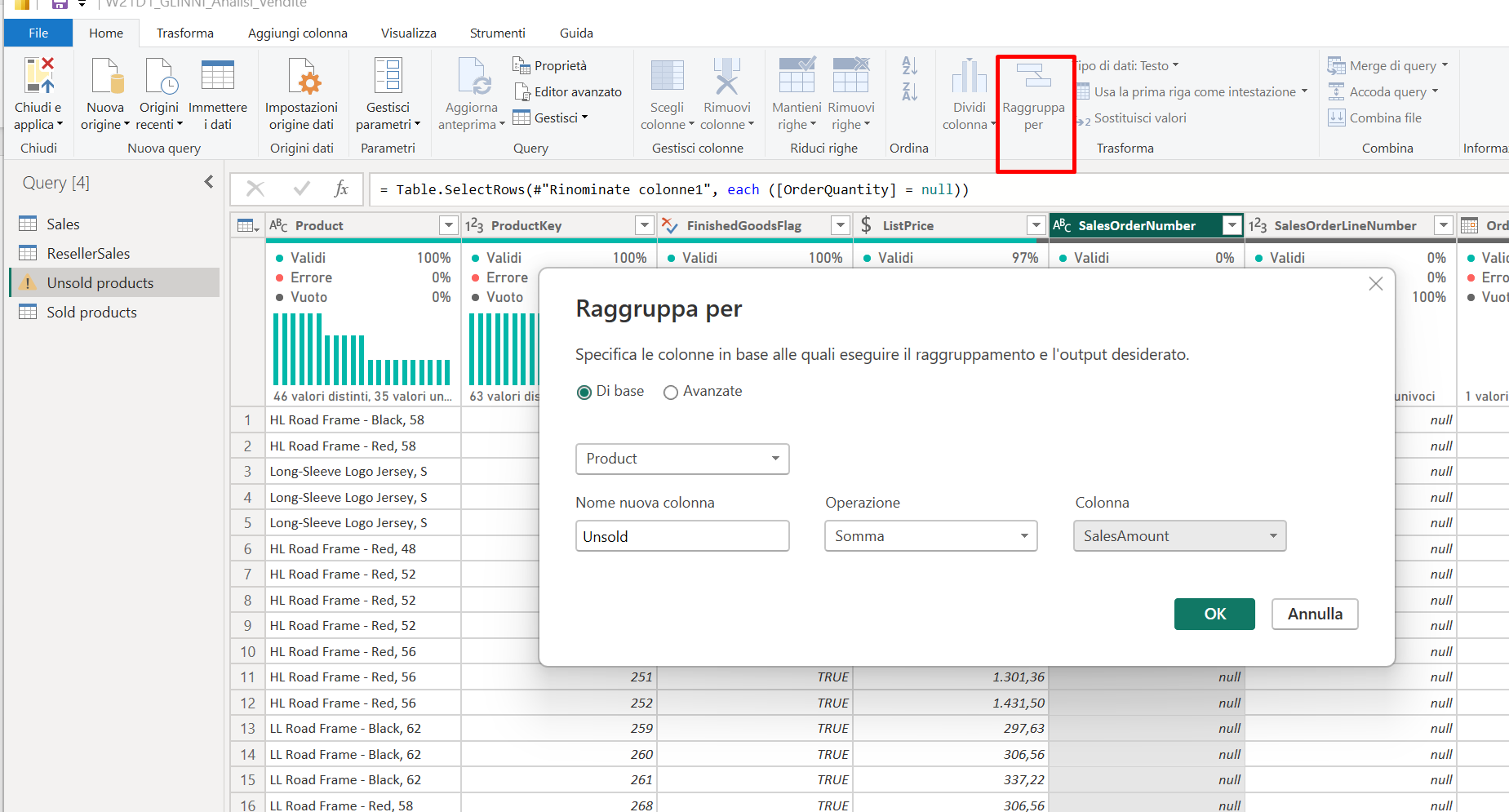
1. Ottieni una tabella contenente i soli prodotti venduti e l’importo totale per ciascuno di essi.
   1. Fai una Merge
   2. Esegui una Group by

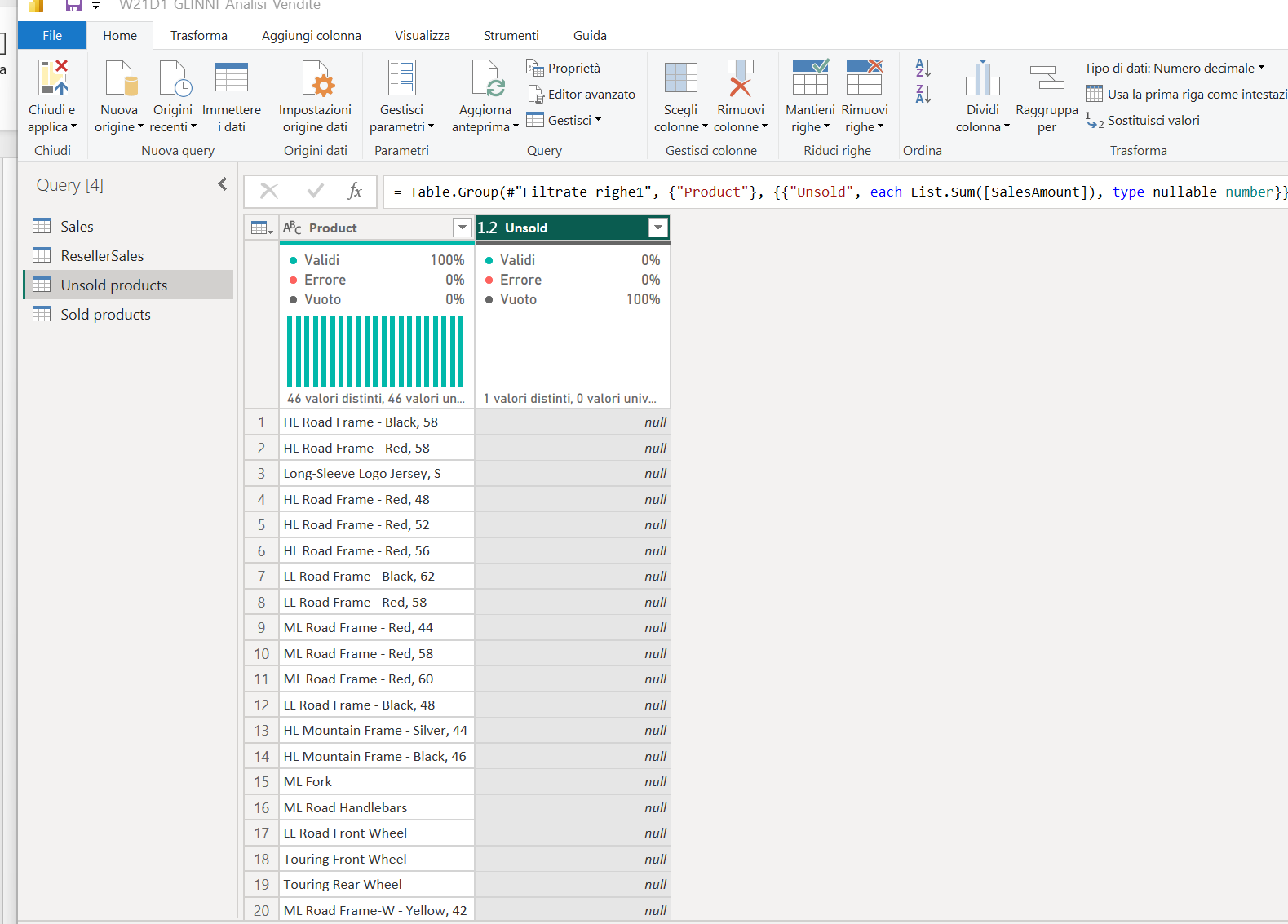


1. Ottieni una tabella contenente l’elenco dei prodotti non venduti se ci sono.

*Aggiunta GAB:*







*(In questo modo si evitano i nomi duplicati)*

1. Ottieni una tabella contenente l’elenco di tutti i prodotti e l’importo totale se venduti.

A screenshot of a computer

Description automatically generated

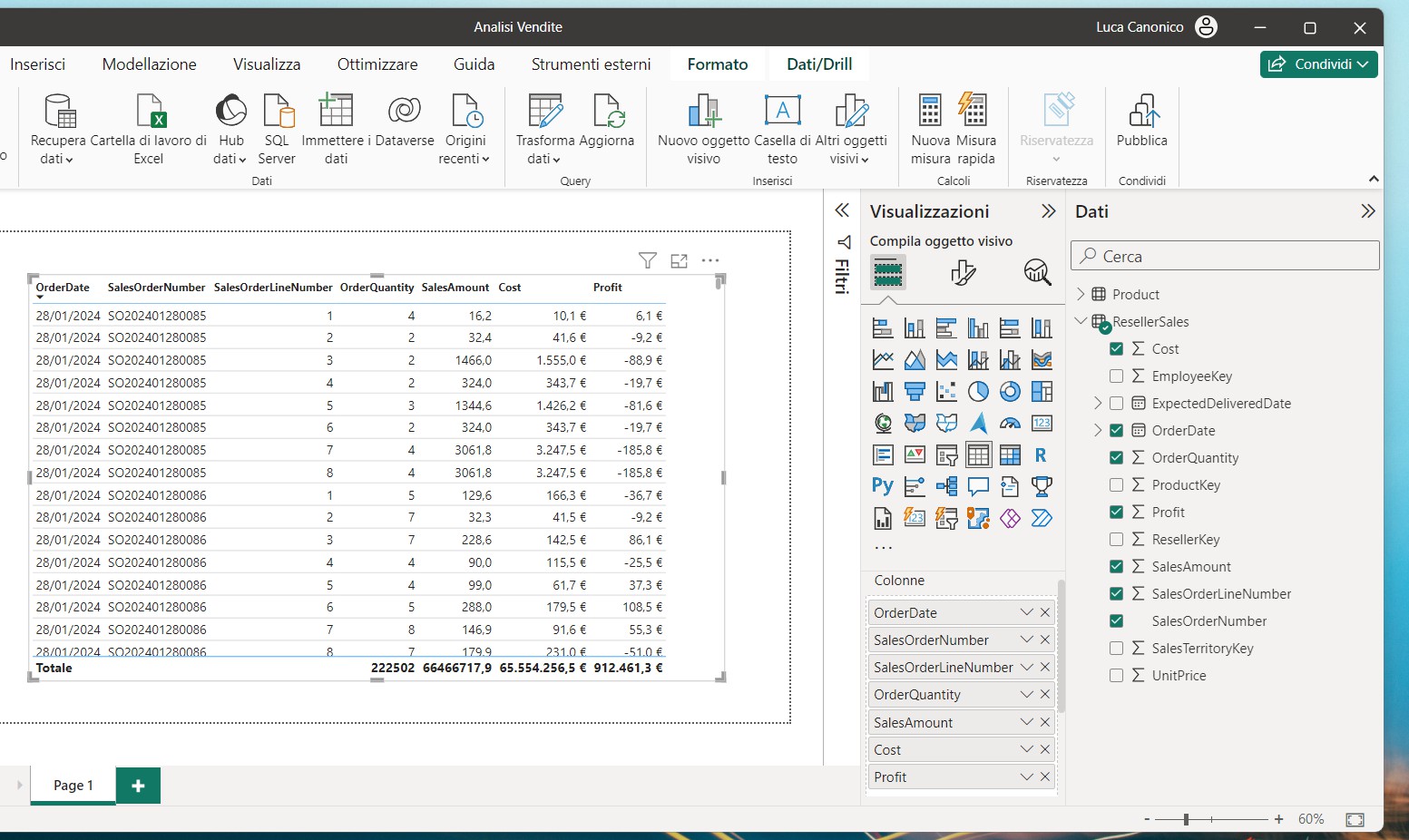
1. A screenshot of a computer

   Description automatically generatedChiudi e applica le modifiche

**Task 4: Visualizzazione**

Proponi una visualizzazione al fine esporre l’elenco delle transazioni con le metriche richieste.

Utilizza la visualizzazione tabella e compilala come indicato di seguito

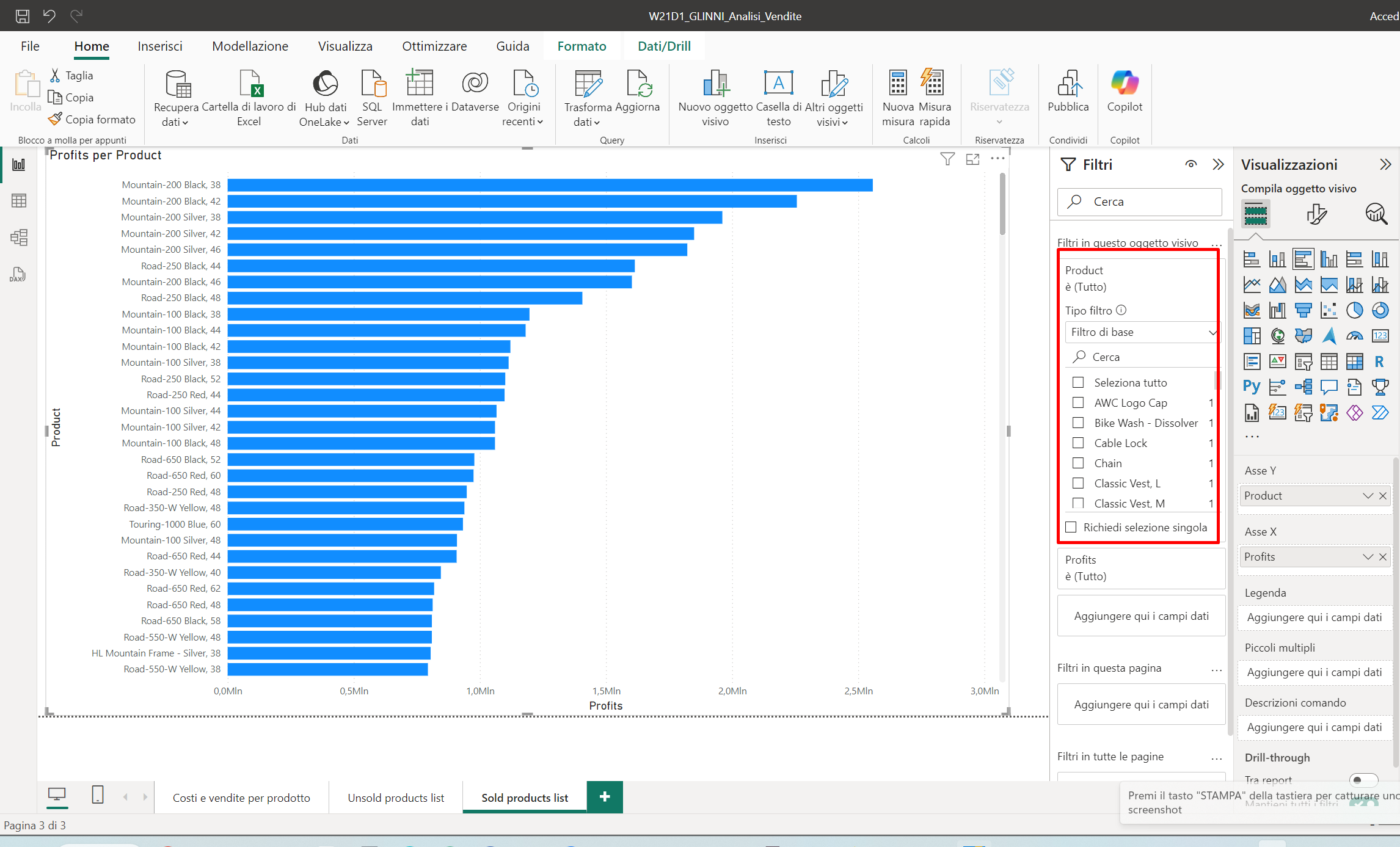


**Riesci a filtrare la visualizzazione per prodotto? Spiega il perché! (filtrare la tabella per prodotto significa arricchire la visualizza del campo Product della tabella Product)**

A screenshot of a computer

Description automatically generated

*Si può filtrare da qui:*



*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!*