

Modello Dati con Power BI

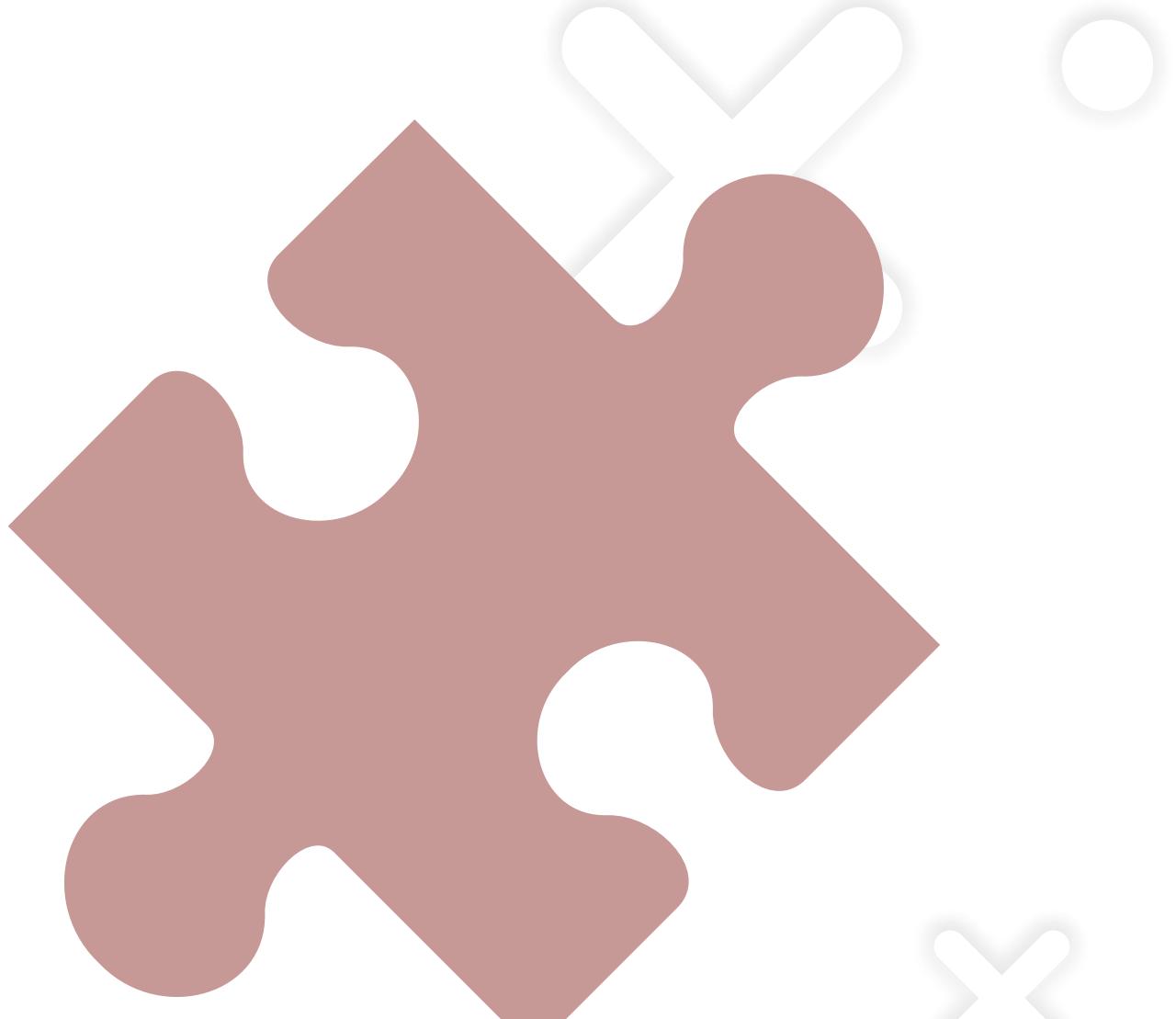


Obiettivi dell'analisi

Un questionario online ha rilevato che alcuni clienti sono scontenti perché hanno registrato un generico "aumento dei prezzi", non meglio precisato.

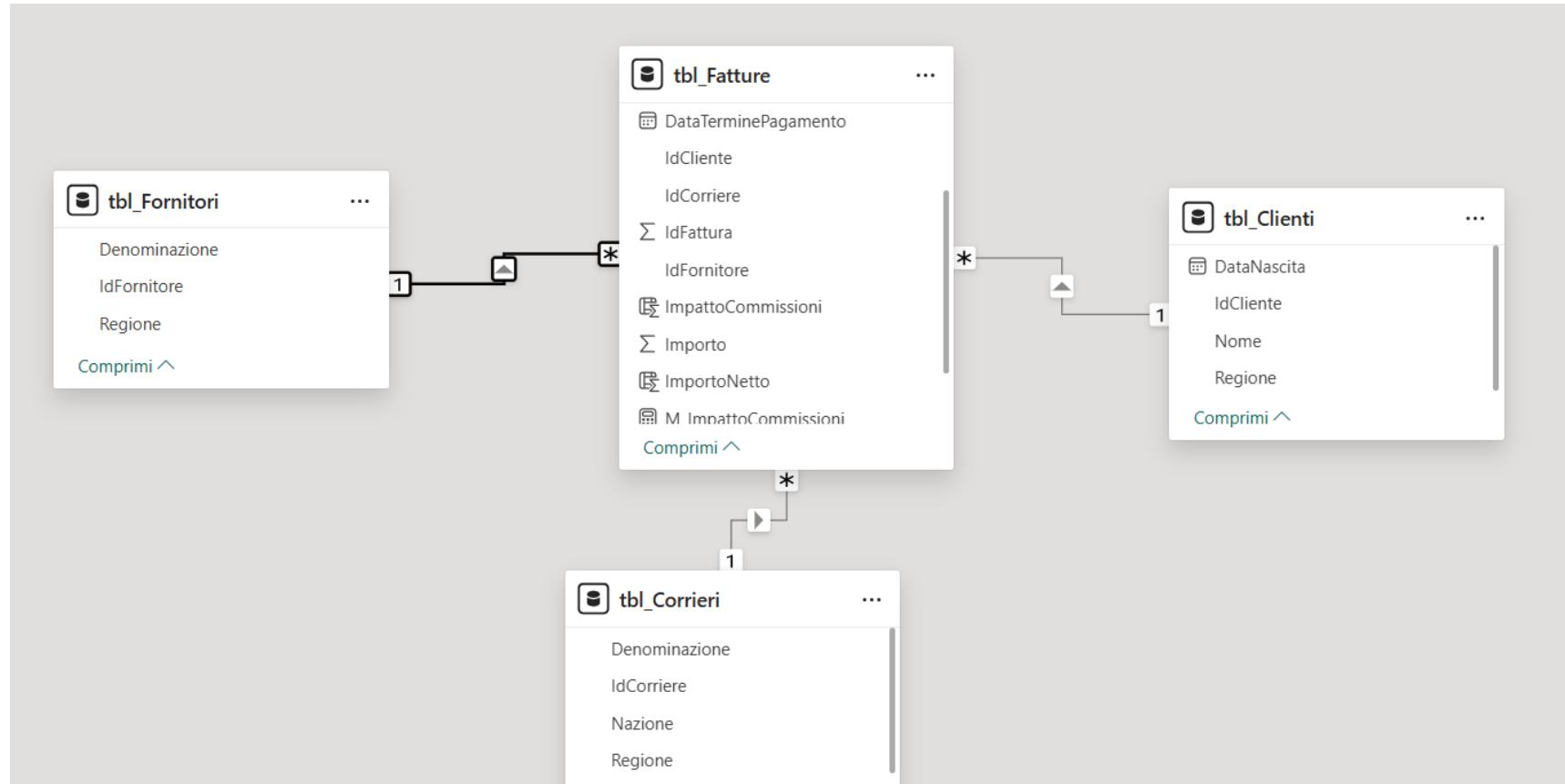
Vogliamo usare Power BI per andare in fondo alla questione...

Modello dati



Modellizzazione dei dati in Power BI

Il motore di calcolo di Power BI risulta particolarmente efficace se i dati sono organizzati secondo uno **schema a stella**, composto da una tabella dei fatti e varie tabelle delle dimensioni connesse.



Cosa sono i fatti

Un **fatto** è qualcosa che accade nella realtà che vogliamo analizzare.

Vediamo qualche esempio:

- all'interno di un progetto di BI di una banca, un fatto è un movimento di conto corrente.
- all'interno di un progetto di BI di un'azienda B2C, un fatto è una vendita.
- all'interno di un progetto di BI di una banca, un fatto è la particolare situazione a fine giornata di un conto corrente.

Misure associate ai fatti

Una misura è un valore numerico che rappresenta una particolare quantità associata a un fatto.

Vediamo qualche esempio:

- **l'importo** è una misura del movimento di un conto corrente.
- **la percentuale di sconto** è una misura di una vendita effettuata
- **il saldo** è una misura della situazione di un conto corrente a fine giornata

Dimensioni

Abbiamo visto come ai fatti siano associate delle misure quantitative.

Un fatto però è anche contestualizzato e descritto tramite **le dimensioni**.

Ad esempio il movimento di conto corrente è contestualizzato da:

- Il cliente che lo ha effettuato
- La data in cui il movimento è avvenuto
- Il luogo dove è avvenuto il movimento

Attributi delle dimensioni

Ogni *entità di business* rappresenterà una dimensione. Ad esempio in un'azienda che vende prodotti avrò come dimensioni:

- i prodotti
- i clienti
- i fornitori
- i negozi

Ogni dimensione avrà a sua volta una serie di attributi. Ad esempio la dimensione *prodotto* avrà come attributo il nome, il colore, la categoria, la sottocategoria.

Gerarchie degli attributi

Tra gli attributi delle dimensioni si possono creare delle gerarchie.

Ad esempio per la dimensione *clienti* c'è una gerarchia tra gli attributi:

- Città di residenza
- Regione di residenza
- Nazione di residenza

Invece per la dimensione *prodotto* c'è una gerarchia tra gli attributi:

- Nome prodotto
- Sottocategoria
- Categoria

Dimensioni degeneri

In alcuni casi un fatto è descritto anche da attributi che non sono né delle quantità (quindi non possono essere misure) né delle entità di business.

Vediamo qualche esempio:

- la valuta con cui è effettuata un' operazione sul conto corrente (euro, dollaro, ecc...)
- lo stato dell'ordine (ok, ko, sospeso)

Spesso ci si riferisce a questi attributi come dimensioni degeneri.

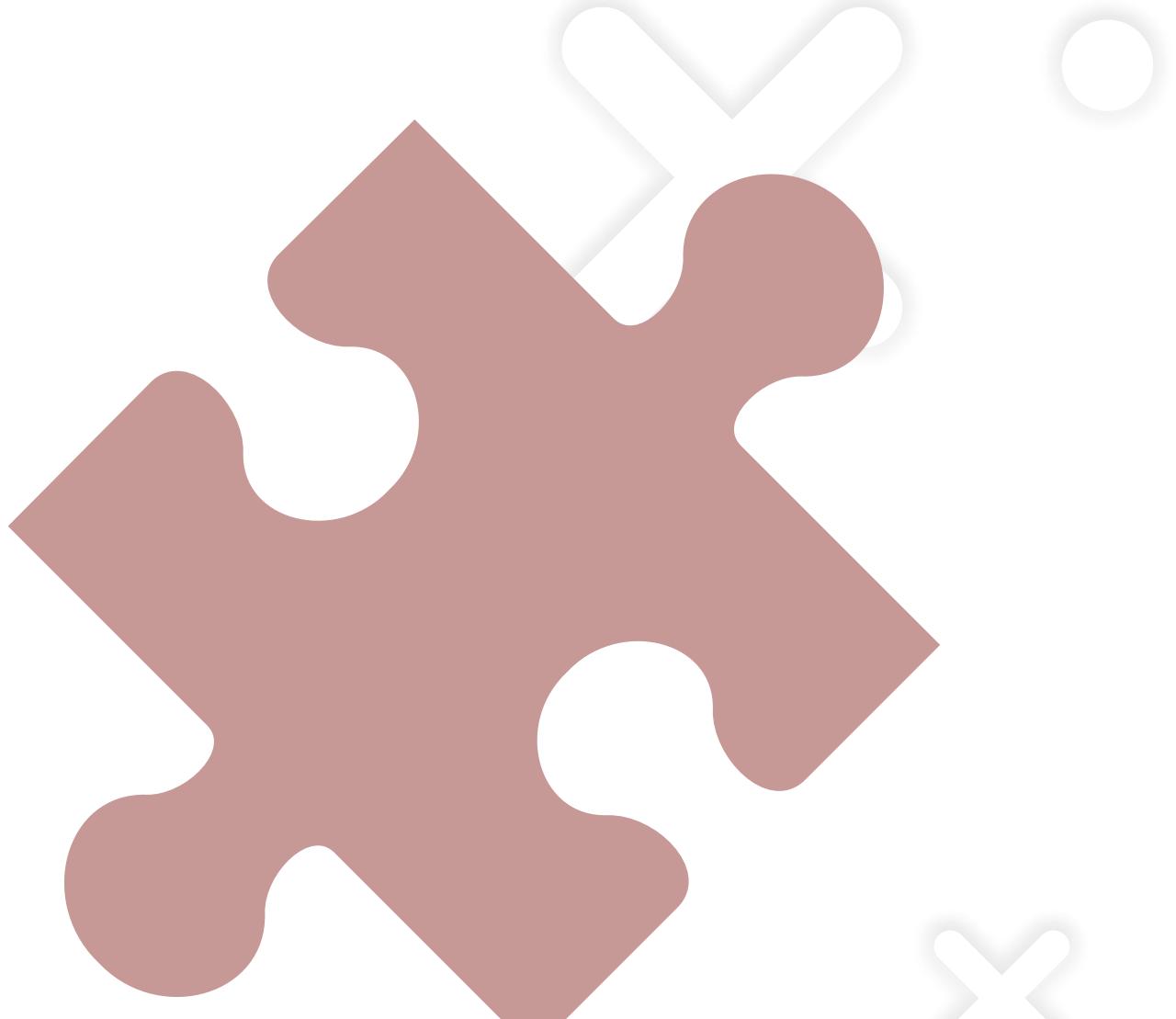
La distinzione tra dimensioni e dimensioni degenere può essere in alcuni casi no

Classificazione delle misure

Ora che abbiamo definito le dimensioni possiamo effettuare un'ulteriore classificazione per le misure dei fatti:

- **Misure additive:** possono essere sommate su tutte le dimensioni (ad esempio l'importo di un movimento del conto-corrente)
- **Misure semi-additive:** possono essere sommate solo su alcune dimensioni (ad esempio il saldo di fine giornata di un conto-corrente ha senso sommarlo al variare della dimensione cliente ma non al variare della dimensione tempo)
- **Misure non additive:** non possono essere sommate su nessuna dimensione (ad esempio la percentuale di sconto di una vendita)

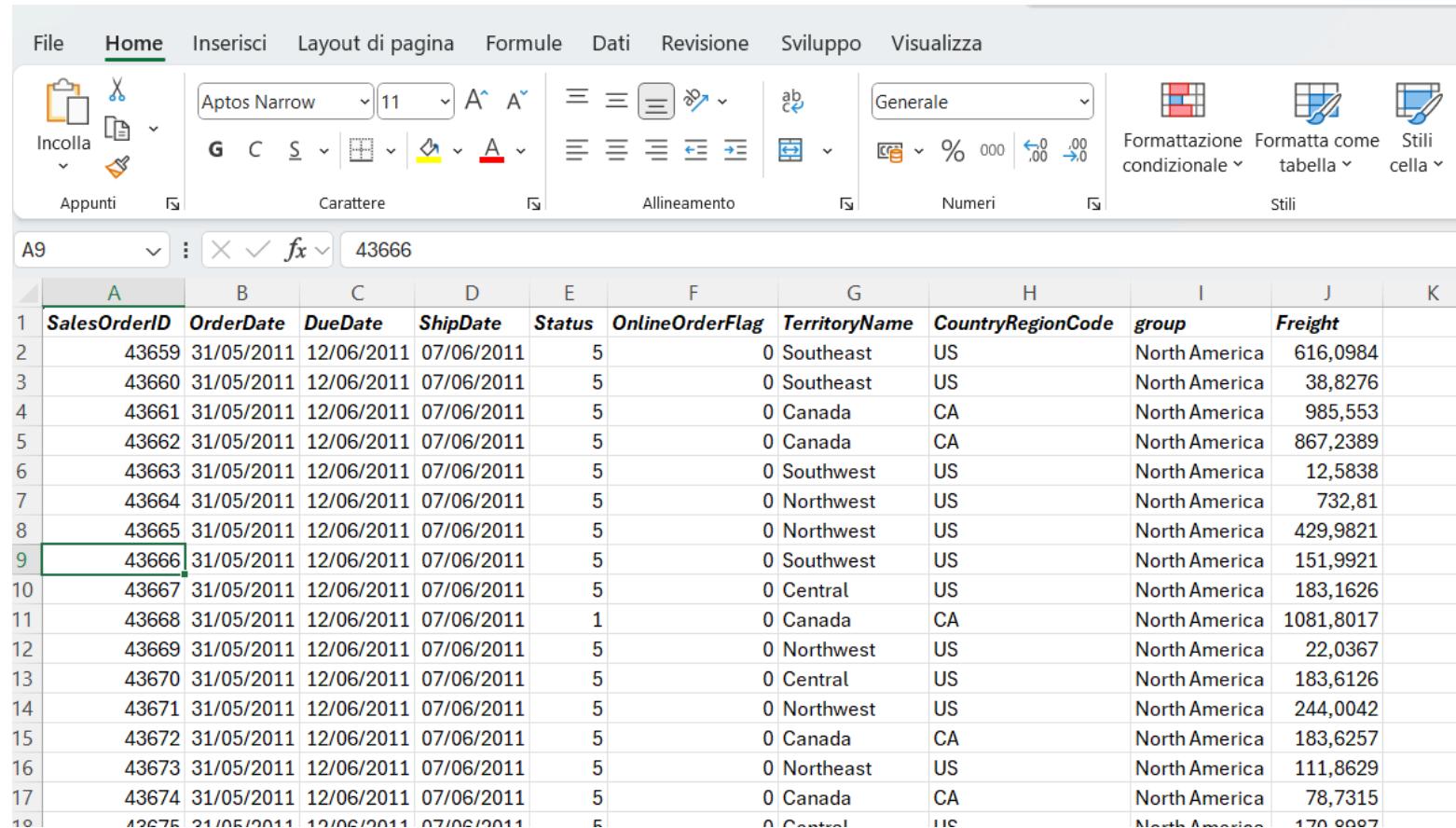
Import dei dati



Acquisiamo tutti i fogli del file Excel

I dati provengono dal database di esempio AdventureWorks, fornito da Microsoft per scopi didattici e dimostrativi.

<https://learn.microsoft.com/it-it/sql/samples/adventureworks-install-configure?view=sql-server-ver17&tabs=ssms>

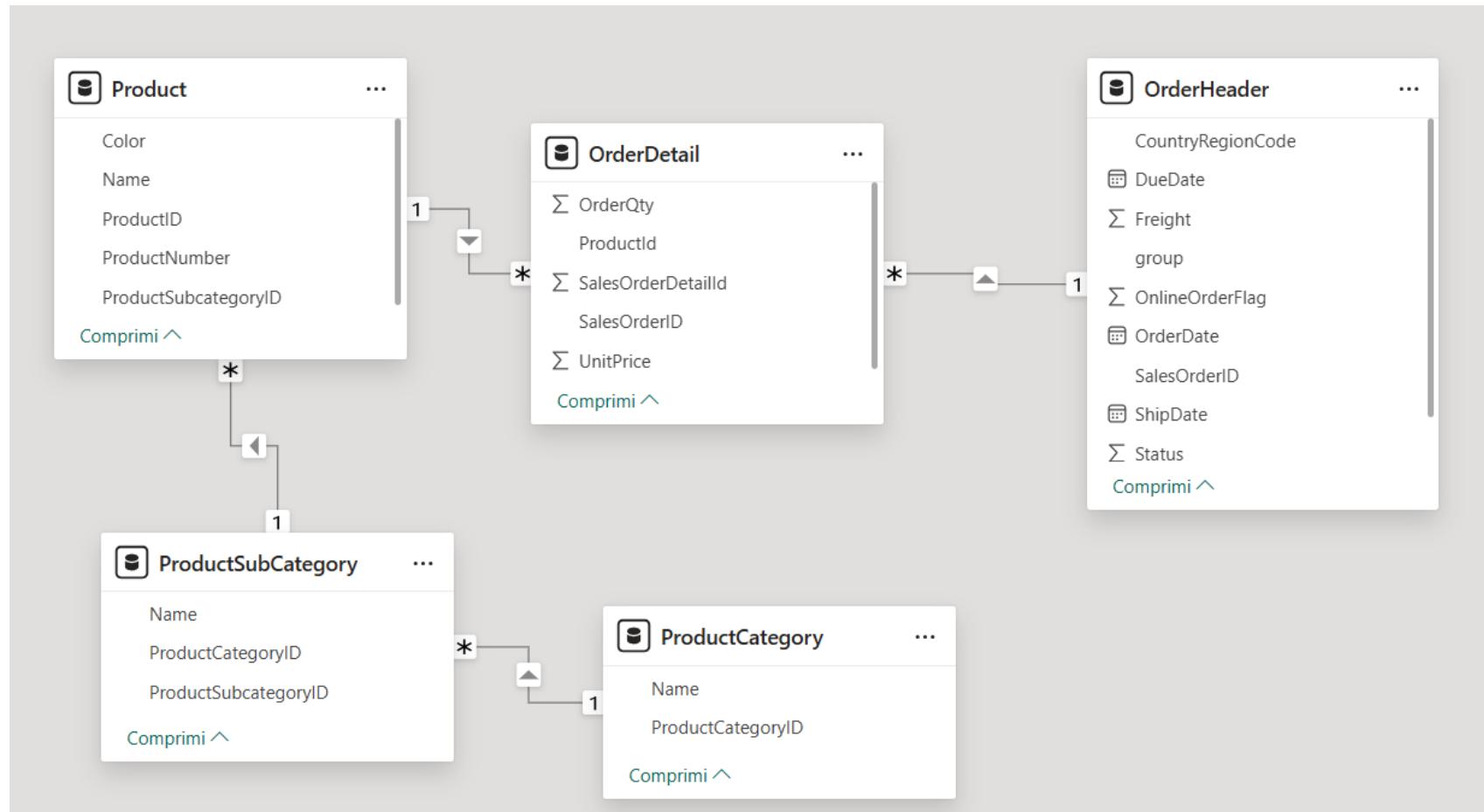


The screenshot shows a Microsoft Excel interface with the 'Home' tab selected in the ribbon. The ribbon also includes 'File', 'Inserisci', 'Layout di pagina', 'Formule', 'Dati', 'Revisione', 'Sviluppo', and 'Visualizza'. The 'Home' tab's ribbon bar contains icons for 'Incolla' (Paste), 'Appunti' (Notes), 'Carattere' (Font), 'Allineamento' (Orientation), 'Numeri' (Number Format), and 'Stili' (Styles). Below the ribbon, the formula bar shows 'A9' and '43666'. The main area displays a table with 17 rows of data from the 'Sales' table in the AdventureWorks database. The columns are labeled A through K, and the data includes SalesOrderID, OrderDate, DueDate, ShipDate, Status, OnlineOrderFlag, TerritoryName, CountryRegionCode, group, Freight, and a final column that appears to be a continuation of the group or territory information.

	A	B	C	D	E	F	G	H	I	J	K
1	SalesOrderID	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	TerritoryName	CountryRegionCode	group	Freight	
2	43659	31/05/2011	12/06/2011	07/06/2011	5	0	Southeast	US	North America	616,0984	
3	43660	31/05/2011	12/06/2011	07/06/2011	5	0	Southeast	US	North America	38,8276	
4	43661	31/05/2011	12/06/2011	07/06/2011	5	0	Canada	CA	North America	985,553	
5	43662	31/05/2011	12/06/2011	07/06/2011	5	0	Canada	CA	North America	867,2389	
6	43663	31/05/2011	12/06/2011	07/06/2011	5	0	Southwest	US	North America	12,5838	
7	43664	31/05/2011	12/06/2011	07/06/2011	5	0	Northwest	US	North America	732,81	
8	43665	31/05/2011	12/06/2011	07/06/2011	5	0	Northwest	US	North America	429,9821	
9	43666	31/05/2011	12/06/2011	07/06/2011	5	0	Southwest	US	North America	151,9921	
10	43667	31/05/2011	12/06/2011	07/06/2011	5	0	Central	US	North America	183,1626	
11	43668	31/05/2011	12/06/2011	07/06/2011	1	0	Canada	CA	North America	1081,8017	
12	43669	31/05/2011	12/06/2011	07/06/2011	5	0	Northwest	US	North America	22,0367	
13	43670	31/05/2011	12/06/2011	07/06/2011	5	0	Central	US	North America	183,6126	
14	43671	31/05/2011	12/06/2011	07/06/2011	5	0	Northwest	US	North America	244,0042	
15	43672	31/05/2011	12/06/2011	07/06/2011	5	0	Canada	CA	North America	183,6257	
16	43673	31/05/2011	12/06/2011	07/06/2011	5	0	Northeast	US	North America	111,8629	
17	43674	31/05/2011	12/06/2011	07/06/2011	5	0	Canada	CA	North America	78,7315	
18	43675	31/05/2011	12/06/2011	07/06/2011	5	0	Central	US	North America	170,9007	

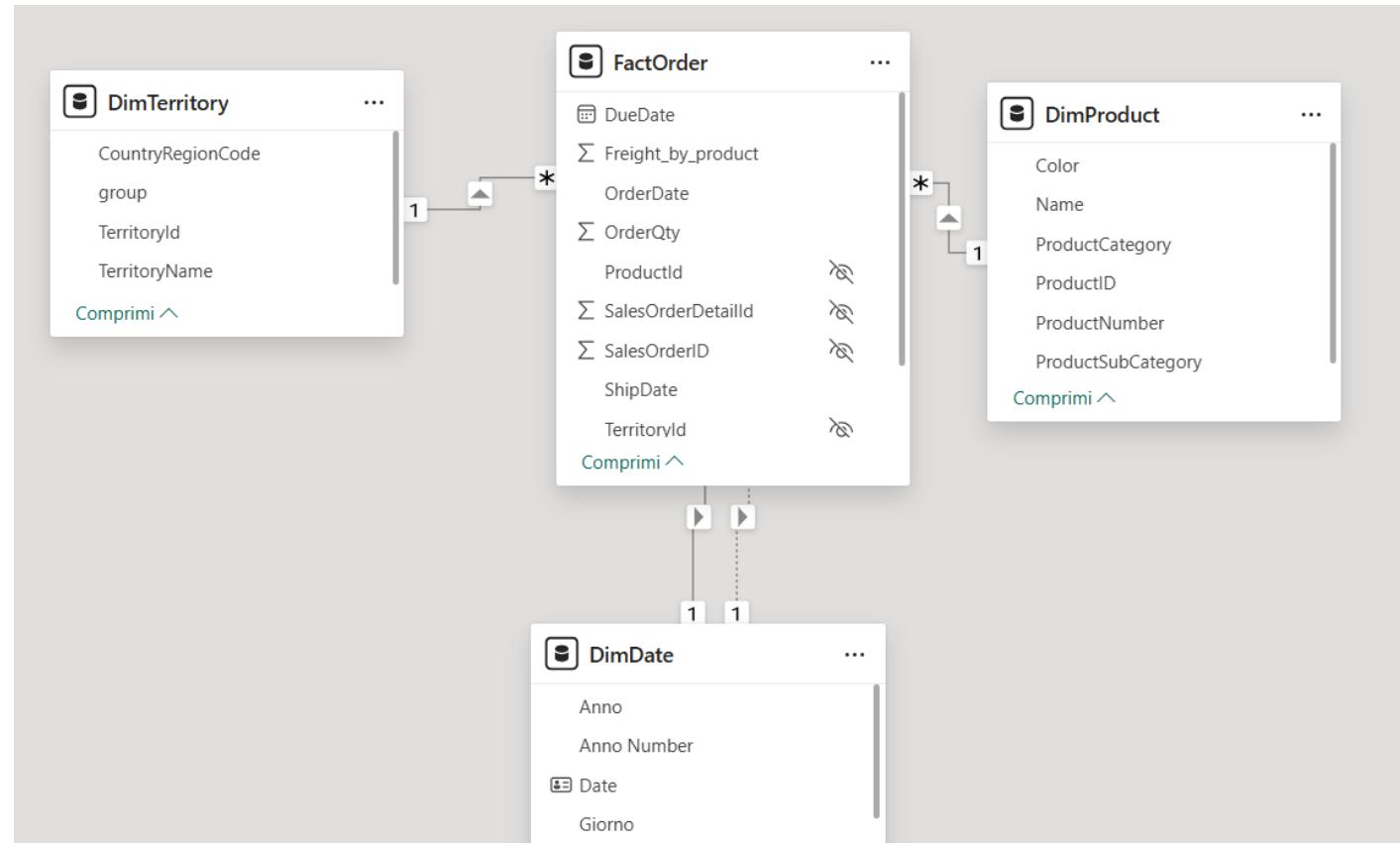
Il modello non è uno Star Schema

Se carichiamo direttamente i dati otterremo una tipologia di modello diverso dallo Star schema.

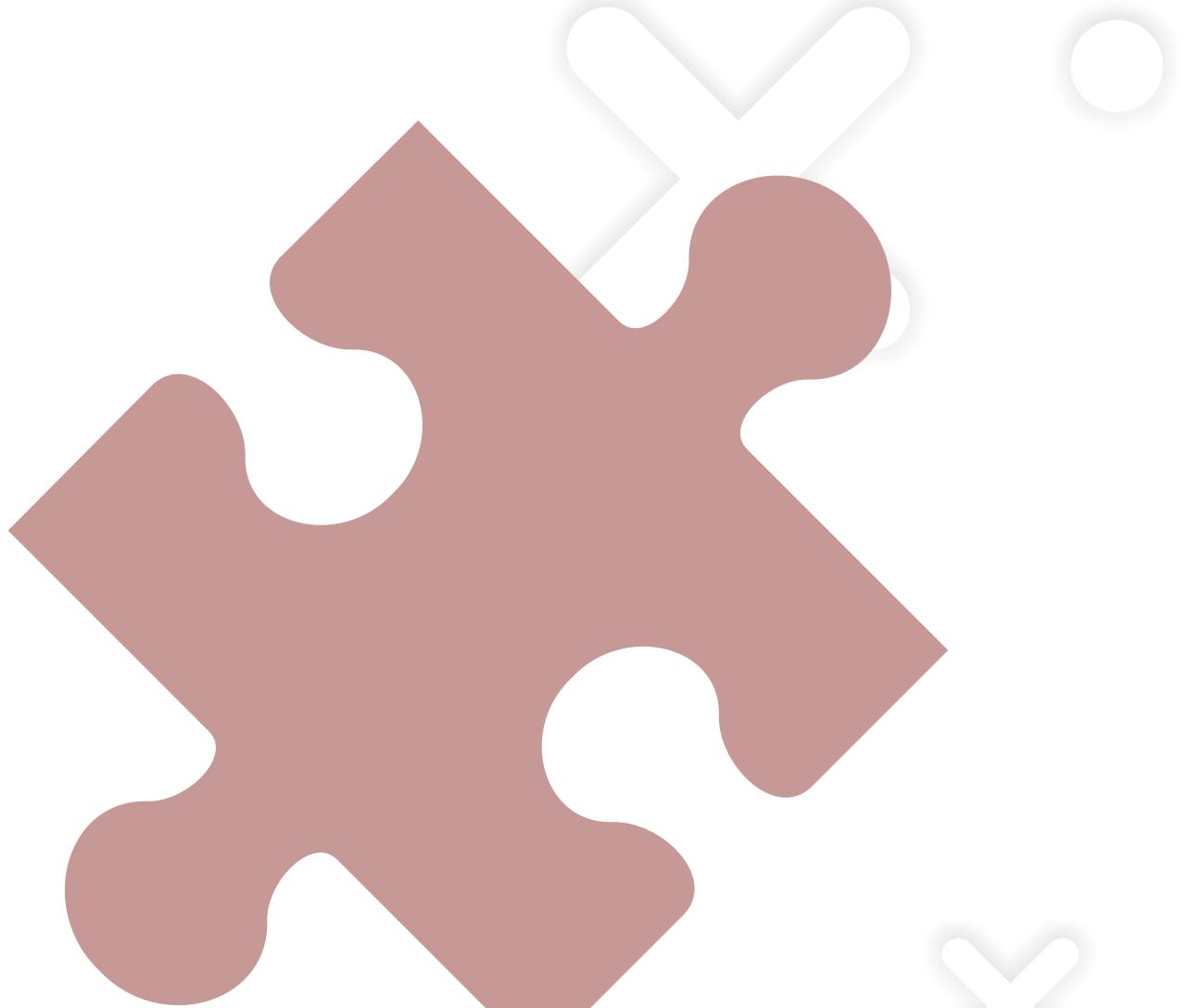


Trasformazione in uno Star Schema

Ottenerne uno Star Schema non è obbligatorio. Tuttavia per la valenza didattica del progetto e per efficientare performance e scrittura delle misure DAX, ci poniamo l'obiettivo di ristrutturare i dati per arrivare a questo modello.



Trasformazione dei dati con Power Query



Creazione della DimProduct - parte 1 di 2

1. Creiamo tre query per i fogli *Prodotti*, *ProductSubCategory* e *ProductCategory*
2. Posizioniamoci sul foglio Prodotti
3. Sostituiamo NULL con N/D nella colonna *Color*
4. Sostituiamo NULL con *null* nella colonna *ProductSubcategoryId*
5. Modifichiamo il tipo della colonna *ProductSubcategoryId* in INT
6. Eseguiamo una merge di tipo LEFT con il foglio *ProductSubCategory* utilizzando le colonne *ProductSubcategoryId*

Creazione della DimProduct – parte 2 di 2

7. Espandiamo la tabella e rimuoviamo le colonne inutili
8. Eseguiamo una merge di tipo LEFT con il foglio *ProductCategory* utilizzando le colonne *ProductCategoryId*
9. Espandiamo la tabella e rimuoviamo le colonne inutili
10. Effettuiamo le rinomine necessarie
11. Sostituiamo *null* con N/D nella colonna *ProductSubcategoryId* e *ProductCategory*
12. Rinominiamo la query in ***DimProduct***

Creazione della DimTerritory

1. Creiamo una query per il foglio *OrderHeader*
2. Selezioniamo le colonne *TerritoryName*, *CountryRegionCode* e *group*
3. Rimuoviamo i duplicati selezionando tutte le colonne. In generale facciamo attenzione alle **performance** per questa tipologia di operazione. In alcuni casi potrebbe essere più opportuno selezionare soltanto la colonna (o le colonne) che sono delle chiavi univoche. Questo però ci espone a rischi sulla presenza di valori duplicati inaspettati, che sarebbe meglio controllare e gestire tramite del codice M personalizzato. C'è il rischio di compromettere la correttezza dei dati nella costruzione della successiva tabella FactSales!
4. Aggiungiamo una colonna indice che parte da 1 di nome *TerritoryId*
5. Riordiniamo e rinominiamo le colonne
6. Rinominiamo la query in ***DimTerritory***

Creazione della FactSales - parte 1 di 3

1. Creiamo una seconda query per il foglio *OrderHeader*
2. Aggiungiamo una colonna condizionale per codificare la colonna *Type* (0->Offline e 1->Online)
3. Eseguiamo il merge con la *DimTerrrotory* sulla terna di colonne *TerritoryName*, *CountryRegionCode* e *group*
4. Rimuoviamo le colonne della *DimTerrrotory* (*TerritoryName*, *CountryRegionCode* e *group*)
5. Espandiamo la tabella e lasciamo la colonna *TerritoryId*
6. Riordiniamo e rinominiamo le colonne

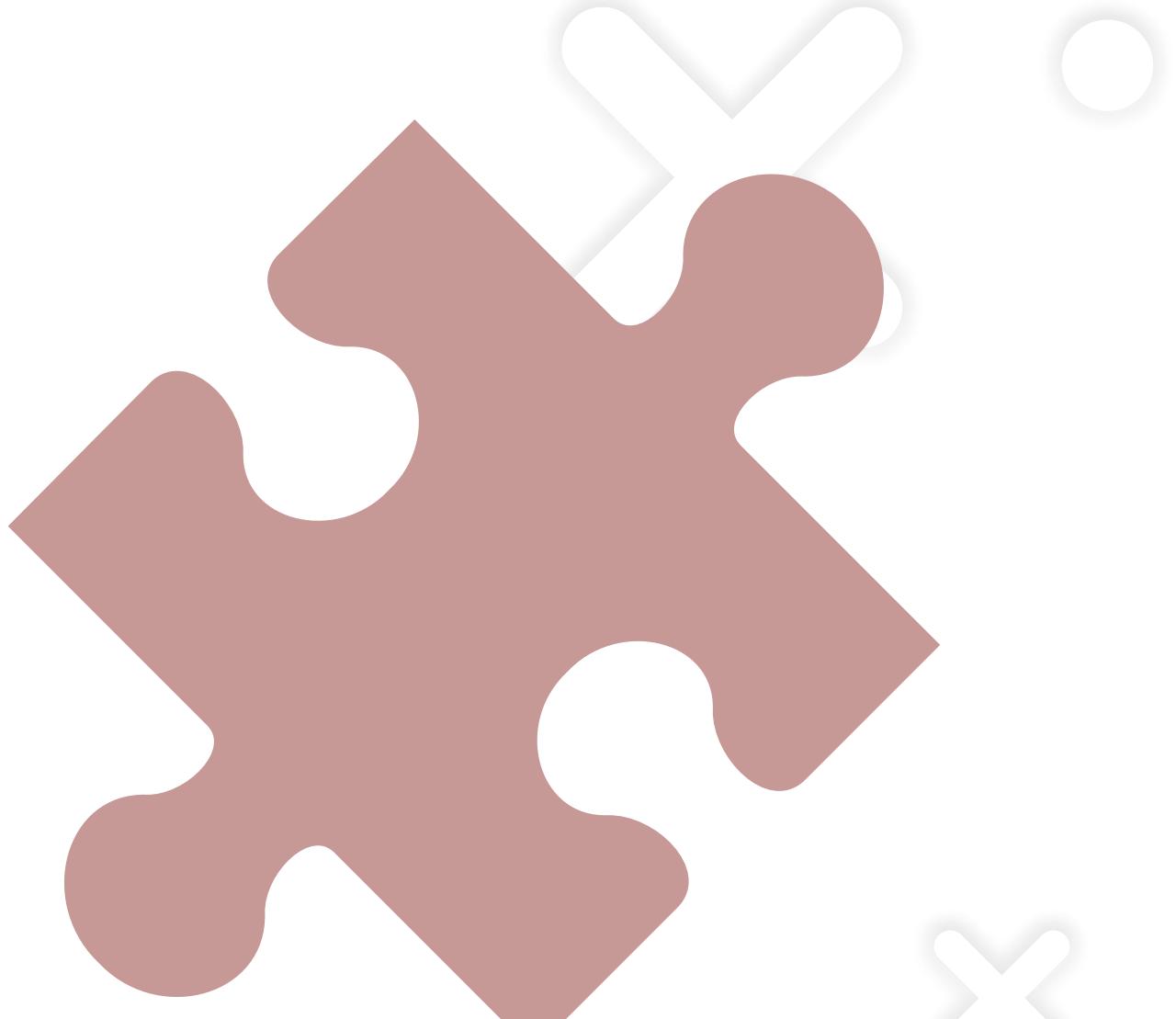
Creazione della FactSales - parte 2 di 3

7. Creiamo una query per il foglio *OrderDetails*
8. Raggruppiamo le righe di questa query per *SalesOrderId* e sommiamo la colonna *OrderQty*
9. Rinominiamo questa query *OrderQTY*
10. Creiamo una nuova query per il foglio *OrderDetails*
11. Eseguiamo il merge con la *OrderHeader* precedente sulle colonne *SalesOrderId*
12. Espandiamo la tabella e selezioniamo le colonne di interesse
13. Filtriamo le righe per visualizzare solo quelle con *status* = 5

Creazione della FactSales - parte 3 di 3

14. Eseguiamo il merge con la *OrderQTY* costruita prima sulle colonne *SalesOrderId*
15. Espandiamo la tabella
16. Creiamo una colonna personalizzata con la formula
 $[Freight]/[OrderQTY.1.QtyTotale]*[OrderQty]$
17. Arrotondiamo la colonna a 4 cifre decimali
18. Rimuoviamo le colonne inutile ed effettuiamo le rinomine necessarie
19. Chiamiamo la query *FactSales*

Finalizzazione del modello



Creazione dello StarSchema

- 1) Su PowerQuery disabilitiamo il caricamento delle query diverse da *DimTerritory*, *DimProduct* e *FactSales*
- 2) Creiamo la relazione uno a molti tra *DimTerritory* e *FactSales*
- 3) Creiamo la relazione uno a molti tra *DimProduct* e *FactSales*
- 4) Creiamo con il DAX una colonna *DimDate*

DimDate =

```
ADDCOLUMNS(CALENDARAUTO(),
"Anno", "Anno " & YEAR ([Date]),
"Anno Number", YEAR ([Date]),
"Trimestre","Trimestre " & QUARTER([Date]),
"Mese", FORMAT ([Date], "mmmm" ),
"Mese Number", MONTH ([Date]),
"Giorno",DAY ([Date]))
```

Creazione dello StarSchema

- 5) Impostiamo la tabella *DimDate* di tipo *Date*
- 6) Creiamo la relazione uno a molti tra *DimDate* e *FactSales*
- 7) Nascondiamo le colonne contenenti delle chiavi

