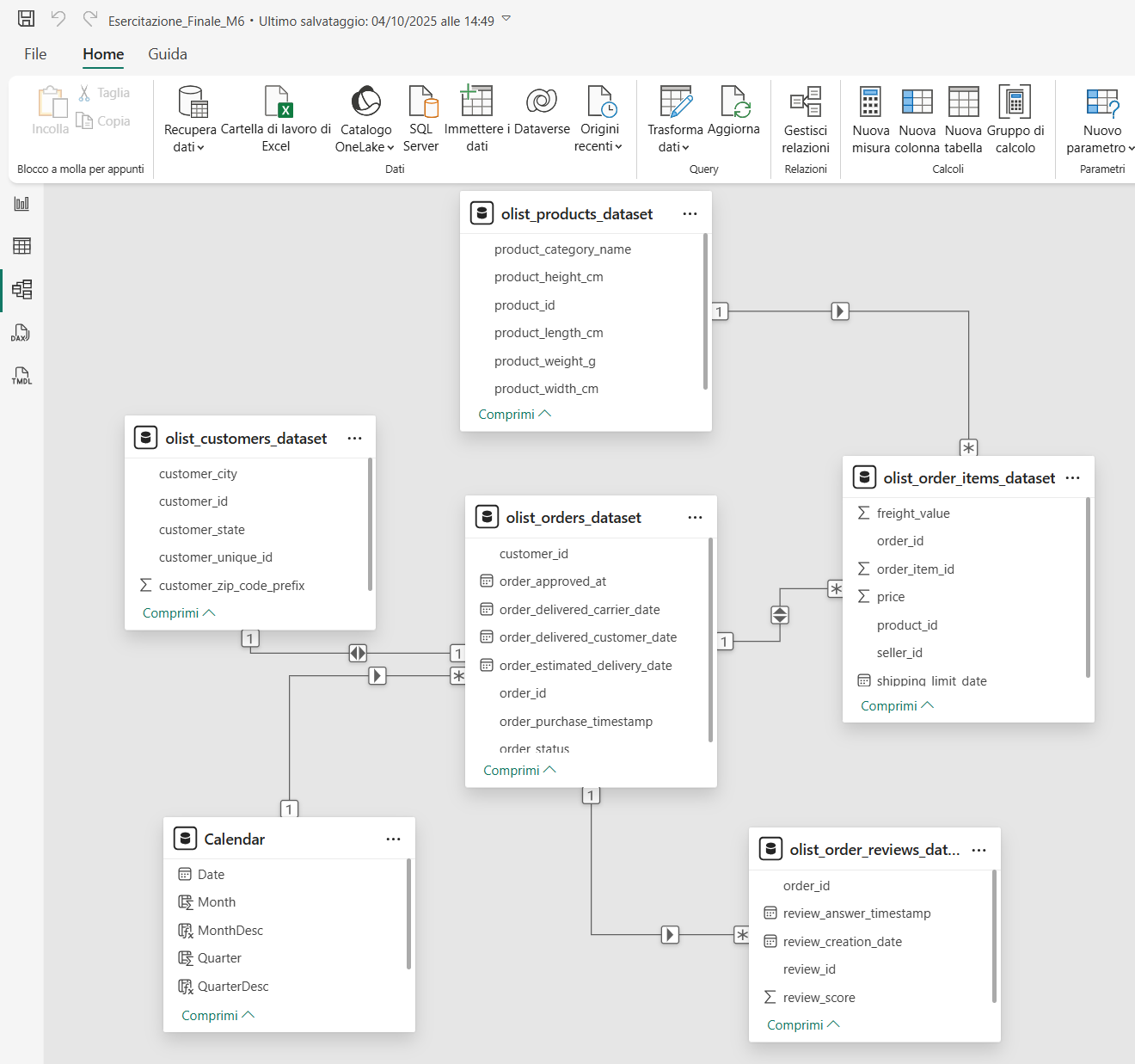
**Best Practices**

Ho iniziato pulendo su Power Query i 5 dataset strettamente necessari all’analisi, nello specifico ho:

* Eliminato delle colonne che ho ritenuto inutili ai fini dell’analisi;
* Verificato che i formati fossero corretti;
* Ho modificato il formato di “order\_purchase\_timestamp” in solo data in quanto, come visto insieme a lezione, avrebbe successivamente dato problemi;
* In “olist\_products\_dataset”, ho sostituito i valori vuoti (senza categoria) della colonna delle categoria prodotto con la parola “altro”;
* In “olist\_customers\_dataset” ho modificato il testo delle città in maiuscolo per ogni parola;
* Lasciato qualche colonna che, seppur inutile ai fini dell’analisi principale, potrebbe rivelarsi utile per un’analisi opzionale di approfondimento (es. differenza tra spedizione stimata e spedizione effettiva oppure sulla grandezza in cm dei prodotti).

Come step successivo ho creato il modello con le relazioni tra le tabelle.

Poi ho creato la tabella Calendar con le varie colonne Data, Anno, Mese, MeseDesc, Trimestre e TrimestreDesc per poi collegare anch’essa al modello relazionale:



**Esercizio 1**

Per il grafico a linee, ho ordinato MonthDesc per colonna Month in modo da avere i mesi in ordine cronologico e non alfabetico.

Ho creato una misura che mi contasse il numero di ordini:

NumeroOrdini = COUNTA(olist\_order\_items\_dataset[order\_id])

un’altra per il past year:

NumOrdiniPY = CALCULATE([NumeroOrdini], PARALLELPERIOD('Calendar'[Date], -12, MONTH))

e una terza per la variazione percentuale:

Ord% = DIVIDE([NumeroOrdini]-[NumOrdiniPY], [NumOrdiniPY])

A questo punto ho inserito i tre filtri, il grafico a linee e aggiunto le variazioni percentuali nella formattazione visiva.

nb: mi sono accorto che i dati sono abbastanza incompleti, infatti per il 2018 e soprattutto per il 2016 sono veramente pochi. Anche per questo il grafico a linee risulta essere un po’ “sballato” però mi sono concentrato sulla consegna e sui passaggi da fare ignorando questo aspetto.

**Esercizio 2**

Per il secondo esercizio, come per il precedente, ho creato le tre misure necessarie alla creazione di un grafico a linee che mostri i ricavi mese per mese, confrontandoli con quelli dell’anno precedente:

Ricavi = SUMX(olist\_order\_items\_dataset, olist\_order\_items\_dataset[price] + olist\_order\_items\_dataset[freight\_value])

RicaviPY = CALCULATE([Ricavi], PARALLELPERIOD('Calendar'[Date], -12, MONTH))

Ricavi% = DIVIDE([Ricavi] - [RicaviPY], [RicaviPY])

Infine mi sono concentrato sul layout.

**Esercizio 3**

Per l'esercizio sulla distribuzione del rating ho iniziato creando due misure:

NumeroRecensioni = DISTINCTCOUNT(olist\_order\_reviews\_dataset[review\_id])

MediaRecensioni = AVERAGE(olist\_order\_reviews\_dataset[review\_score])

In modo da iniziare ad avere sia due card utili alla visualizzazione sia un grafico a colonne che mostri quante recensioni ci sono per ogni punteggio che va da 1 a 5.

La struttura dei filtri sulla sinistra è la stessa rispetto ai precedenti esercizi, in modo da poter filtrare le recensioni per anno, stato dell’ordine e/o Stato del Brasile.

Come secondo grafico principale ho scelto un grafico a barre per mostrare la media recensione per categoria prodotto. Scorrendo verso il basso è possibile infatti vedere quali sono le categorie che hanno una maggiore media recensione, fino ad arrivare alle peggiori; il tutto ovviamente è filtrabile per anno, stato dell’ordine o Stato del Brasile.

Infine ho deciso di aggiungere altre card per mostrare recensioni positive e negative:

RecensioniPositive = CALCULATE(COUNTROWS(olist\_order\_reviews\_dataset),olist\_order\_reviews\_dataset[review\_score] >= 4)

RecensioniNegative = CALCULATE(COUNTROWS(olist\_order\_reviews\_dataset), olist\_order\_reviews\_dataset[review\_score] <=2)

RecPositive% = DIVIDE([RecensioniPositive], [NumeroRecensioni])

RecNegative% = DIVIDE([RecensioniNegative], [NumeroRecensioni])

Soltanto per queste ultime cards ho scelto di considerare 4 e 5 stelle per recensioni positive e 1 e 2 stelle per quelle negative, evitando quindi le 3 stelle.

**Analisi extra**

Come analisi extra ho creato una nuova pagina (“Dettaglio ordini e ricavi”) per l’analisi dettagliata di ordini e ricavi, stavolta però ho voluto usare dei pulsanti con azione segnalibro per differenziarli tra loro e averli a pagina intera.

Nello specifico ho inserito una tabella sopra e la mappa sotto in modo da avere informazioni dettagliate a seconda dei filtri applicati.

Successivamente ho creato una nuova pagina (“Analisi consegne e volume degli ordini“) per evidanziare se esiste correlazione tra ritardi di consegna e volume degli ordini.

Per prima cosa in “olist\_orders\_dataset” ho creato una nuova colonna che mi mostrasse la differenza tra consegna e consegna prevista:

GiorniDifferenzaConsegna =

IF(

NOT(ISBLANK(olist\_orders\_dataset[order\_estimated\_delivery\_date])) &&

NOT(ISBLANK(olist\_orders\_dataset[order\_delivered\_customer\_date])),

DATEDIFF(

olist\_orders\_dataset[order\_estimated\_delivery\_date],

olist\_orders\_dataset[order\_delivered\_customer\_date],

DAY

),

BLANK()

)

poi ho creato una nuova misura per calcolare il ritardo medio di consegna:

RitardoMedioConsegna = AVERAGE(olist\_orders\_dataset[GiorniDifferenzaConsegna])

purtroppo i dati non mi hanno aiutato molto in quanto olist store (o perlomeno questi dati) evidenziano consegne molto in anticipo rispetto alla data prevista, infatti i numeri sono per lo più negativi.

Dopo ho creato tre card per mostrare ritardo medio, percentuale di consegne anticipate e percentuale consegne in ritardo, queste le formule:

%ConsegneAnticipate =

DIVIDE(

COUNTROWS(FILTER(olist\_orders\_dataset, [GiorniDifferenzaConsegna] < 0)),

COUNTROWS(FILTER(olist\_orders\_dataset, NOT(ISBLANK([GiorniDifferenzaConsegna]))))

)

%ConsegneRitardo =

DIVIDE(

COUNTROWS(FILTER(olist\_orders\_dataset, [GiorniDifferenzaConsegna] > 0)),

COUNTROWS(FILTER(olist\_orders\_dataset, NOT(ISBLANK([GiorniDifferenzaConsegna]))))

)

Dopodiché ho creato un grafico a linee che mostrasse il ritardo medio di consegna per mese (filtrabile per anno, categoria prodotto e/o Stato).

Successivamente sono passato al dataset “olist\_products\_dataset” ed ho creato una nuova colonna che mi calcolasse il volume prodotto in centimetri cubi:

VolumeProdotto (cm³) = olist\_products\_dataset[product\_length\_cm] \* olist\_products\_dataset[product\_height\_cm] \* olist\_products\_dataset[product\_width\_cm]

Ho dunque creato due misure per vedere il volume medio del prodotto e degli ordini sotto forma di card:

VolumeMedioProdotto (cm³) = AVERAGE(olist\_products\_dataset[VolumeProdotto (cm³)])

VolumeMedioOrdine (cm³) = AVERAGEX(RELATEDTABLE(olist\_order\_items\_dataset), RELATED(olist\_products\_dataset[VolumeProdotto (cm³)]))

Infine ho creato un grafico a dispersione per vedere la correlazione tra volume medio ordine e differenza giorni di consegna. La distribuzione dei punti non evidenzia una relazione chiara tra la dimensione del pacco e la puntualità della consegna. La maggior parte degli ordini presenta un volume inferiore a 20.000 cm³ e viene consegnata in anticipo o in linea con la data stimata. Alcuni casi isolati di volumi molto elevati mostrano ritardi consistenti, ma non tali da determinare una tendenza generale.