

# ESAME POWER BI di Turchi Giulia

## Scenario

[Olist store](#) è un sito di e-commerce brasiliano per seller (venditori).

La piattaforma consente ai seller di proporre i propri prodotti ai diversi mercati brasiliani.

Il dataset pubblico (e anonimizzato) espone gli ordini di vendita dal 2016 al 2018 e consente di quantificarli, misurarli, analizzarsi rispetto dimensioni di analisi diverse: cliente, prodotto, metodi di pagamento, status dell'ordine, ...

Obiettivi di analisi

È necessario sviluppare un report di BI che consenta di analizzare:

1. L'andamento degli ordini nel tempo per stato\*
2. L'andamento dei ricavi nel tempo per stato\*\*
3. Distribuzione del rating\*\*\*

### Tip

\*l'utente, dato un particolare anno selezionato, vorrebbe poter visualizzare il conteggio degli ordini mese per mese. Inoltre, vorrebbe poter confrontare i risultati con quelli dell'anno precedente e, rispetto a questi, visualizzare la variazione percentuale mese per mese. È fondamentale poter filtrare rispetto allo status dell'ordine.

Il campo da considerare per la metrica è **olist\_order\_items\_dataset | order\_item\_id**.

Per i dati geografici bisogna fare riferimento al campo **olist\_customers\_dataset | customer\_state**.

Per applicare il filtro sullo status degli ordini utilizza il campo **olist\_orders\_dataset | order\_status**.

\*\* l'utente, dato un particolare anno selezionato, vorrebbe poter visualizzare i ricavi totali mese per mese. Inoltre, vorrebbe poter confrontare i risultati con quelli dell'anno precedente e, rispetto a questi, visualizzare la variazione percentuale mese per mese. È fondamentale poter filtrare rispetto allo status dell'ordine.

Per calcolare il ricavo della singola transazione è necessario considerare la somma di **olist\_order\_items\_dataset | price** e **olist\_order\_items\_dataset | freight\_value** cioè la somma del prezzo di vendita e del costo di spedizione/consegna.

\*\*\*utilizza il campo **olist\_order\_reviews\_dataset | review\_score**.

Applica tutte le best practice condivise:

1. Riduci il volume del dataset
2. Ristruttura i dati grezzi in modo tale da ottenere uno star schema
3. Utilizza una dimensione calendario
4. Progetta un buon layout
5. Progetta una buona UX (bottoni, drill-through per esporre maggiori livelli di dettaglio rispetto a certe aggregazioni...)
6. Arricchisci il report anche di altre analisi: distribuzione del rating (**utilizza la tabella olist\_order\_reviews\_dataset**, analisi per prodotto e/o per area geografica).

Le tabelle strettamente necessarie alle analisi sono:

- olist\_orders\_dataset
- olist\_order\_items\_dataset
- olist\_products\_dataset
- olist\_order\_reviews\_dataset
- olist\_customers\_dataset

Nota bene: ti viene fornita l'estrazione dati completa per tua comodità qualora tu volessi approfondire per tuo esercizio lo scenario proposto.

---

## Svolgimento:

### OBIETTIVO DEL REPORT

Analizzare l'andamento:

- degli ordini e dei ricavi per mese/anno per stato (con variazione percentuale);
- delle recensioni dei clienti (review score);
- con confronto rispetto all'anno precedente;
- e approfondimenti per categoria prodotto .

---

## 1. DATI & MODELLAZIONE

Eseguita una pulizia dei dati in Power Query (Excel):

Tabelle coinvolte:

- **f\_orders** (è il set di dati principale, da ogni ordine si possono trovare tutte le altre informazioni. Include dati sullo status spedizione di ciascun ordine.) → **fact table principale**. Contiene 99.441 righe, che sono anche valori distinti ed univoci ('order\_id');
- **f\_order\_items** (include dati sugli articoli acquistati in ciascun ordine). Contiene colonne con costo dell'ordine e costo di spedizione ed abbiamo nella colonna 'order\_id', 88.863 valori univoci, 98.666 ordini distinti, di cui 9.803 ordini contengono più di un articolo (totale di 112.650 righe complessive, di cui 13.984 righe sono ordini con più articoli, dal secondo articolo in poi).

Pulizia eseguita per le due tabelle 'orders':

- ✓ constatato il fatto di una discrepanza di 775 ordini tra le due tabelle, ho controllato quali fossero le informazioni presenti nel dataset di tali ordini e quali mancassero → questi ordini hanno solo status spedizione, clienti e pagamenti effettuati (anche se in alcuni casi trattasi di cifre sotto 1€), mancano prezzi e dettagli articoli (quantità e prodotto);
- ✓ stabilito di prendere in considerazione solo ordini completi di tutte le informazioni (per effettuare un'analisi più veritiera possibile), ho effettuato un merge tra le due tabelle, per tenere nella tabella 'f\_orders' solo i campi interessati (98.666 righe, che sono anche valori distinti ed univoci in 'order\_id') → stesso numero di valori distinti

- nella tabella 'f\_order\_items', dove i campi sono rimasti invariati, in quanto contengo le info sui prezzi\* degli ordini (creata colonna 'sales\_amount', per ogni singola riga);
- ✓ eliminato colonne non necessarie e rinominate quelle restanti;
  - ✓ verificate in 'f\_orders', le colonne con le date relative ai vari status spedizione → alcuni stati non coincidono con ultima data disponibile. Le colonne che rispettano le date per intero sono 'order\_date' e 'delivered\_customer\_date' (ci sono alcune date di consegna al cliente che differenziano più di 30 giorni con data di spedizione e date dove lo stato non è 'Delivered', che sono mancanti);
  - ✓ colonne finali:  
 tabella 'f\_orders' → 'order\_id', 'customer\_id', 'order\_status', 'order\_date', 'delivered\_customer\_date' e 'estimated\_delivery\_date';  
 tabella 'f\_order\_items' → 'order\_id', 'order\_item\_id', 'product\_id', 'price' e 'freight\_value';
  - ✓ in PowerBI ho creato le colonne 'days\_passed' (per misurare i tempi effettivi di consegna per ogni ordine), 'delivery\_range' (per raggruppare le consegne in 4 fasce di durata\*\*) e 'sales\_status\_group' (per semplificare e raggruppare gli stati ordine in 'Delivered' e 'Not Delivered').

**\* Esempio calcolo ricavo (colonna 'sales\_amount' → [price] + [freight\_value]) nella tabella 'f\_order\_items':**

L'order\_id = 00143d0f86d6fbd9f9b38ab440ac16f5 contiene 3 articoli (stesso prodotto). Per ogni articolo, il costo di spedizione è calcolato in base alle sue dimensioni e al suo peso. Per ottenere il valore totale della spedizione per ogni ordine, è sufficiente sommare.

- Il valore totale dell'ordine\_articolo è: **21,33 \* 3 = 63,99**
- Il valore totale del trasporto è: **15,10 \* 3 = 45,30**
- Il valore totale dell'ordine (prodotto + trasporto) è: **45,30 + 63,99 = 109,29**

**\*\* delivery\_range = SWITCH( TRUE(),**  
 ISBLANK(f\_orders[days\_passed]), "Data Missing", # Dati mancanti  
 f\_orders[days\_passed] <= 19, "On Time", # Consegnato puntualmente  
 f\_orders[days\_passed] <= 38, "Late but Tolerable", # Consegna accettabile  
 f\_orders[days\_passed] > 38, "Significant Delay") # Ritardo significativo

Range massimo per la consegna di 38 giorni ricavato da vari commenti analizzati.

- **f\_reviews** (include dati sulle recensioni lasciate dai clienti);  
Pulizia eseguita:
  - ✓ da un primo controllo risultano duplicati sia nella colonna 'review\_id', sia nella colonna 'order\_id';
  - ✓ dopo vari controlli incrociati, ho deciso di eliminare i duplicati nella colonna 'order\_id': la maggior parte avevano una 'review\_id' nella tabella ordini completi e l'altra nella tabella ordini senza prezzo (che si eliminano in automatico con il merge); altri erano duplicati tra loro in entrambe le colonne ed ho effettuato una verifica tra

data di consegna e data della recensione, e tenuto solo il valore che rispettava le tempistiche ed eliminato quelle con date superiori alla data di consegna);

- ✓ ho mantenuto i duplicati nella colonna 'review\_id', in quanto un ordine può avere più di una recensione;
- ✓ eseguito merge per eliminare del tutto i commenti riferiti ai 775 ordini incompleti;

Risultano quindi 97.705 righe, che sono anche valori distinti ed univoci nella colonna 'review\_id'; Invece per la colonna 'order\_id', risultano 97.705 righe, con 97.359 valori distinti e 97.014 valori univoci;

- **f\_payments** (include dati sulle opzioni di pagamento degli ordini).

Pulizia eseguita:

- ✓ eseguito merge con la tabella 'f\_orders', per prendere solo i pagamenti degli ordini completi di prezzo (98.666) → risultato 100.892 pagamenti (righe totali, alcuni ordini anno più pagamenti con diversa tipologia, tipo 'Credit\_card' e 'Voucher'), con 98.665 valori distinti (per l'ordine bfb0f9bdef84302105ad712db648a6c non risultano dati di pagamento);
- ✓ eliminato colonne non necessarie;
- ✓ rinominate le colonne e tenute solo 'order\_id', 'number', 'type' e 'value';

- **d\_customers** (contiene informazioni sul cliente e sulla sua posizione. Identifica i clienti univoci nel set di dati degli ordini e trova luogo di consegna degli ordini). Qui abbiamo due codici ID → 'customer\_unique\_id', identificatore univoco di un cliente (95.420 valori distinti) e 'customer\_id', che è la chiave per il dataset degli ordini (98.666 valori distinti).

Pulizia eseguita:

- ✓ rimosso colonna 'customer\_zip\_code\_prefix';
- ✓ duplicato la colonna 'customer\_state', per avere nome dei stati completi;
- ✓ rinominato le colonne;
- ✓ aggiunto colonna personalizzata 'location' (concatenazione tra 'city', 'state\_name' & Brazil);
- ✓ fatto merge con la tabella 'f\_orders', per eliminare clienti dei 775 ordini senza prezzo e rinominato le tabelle;

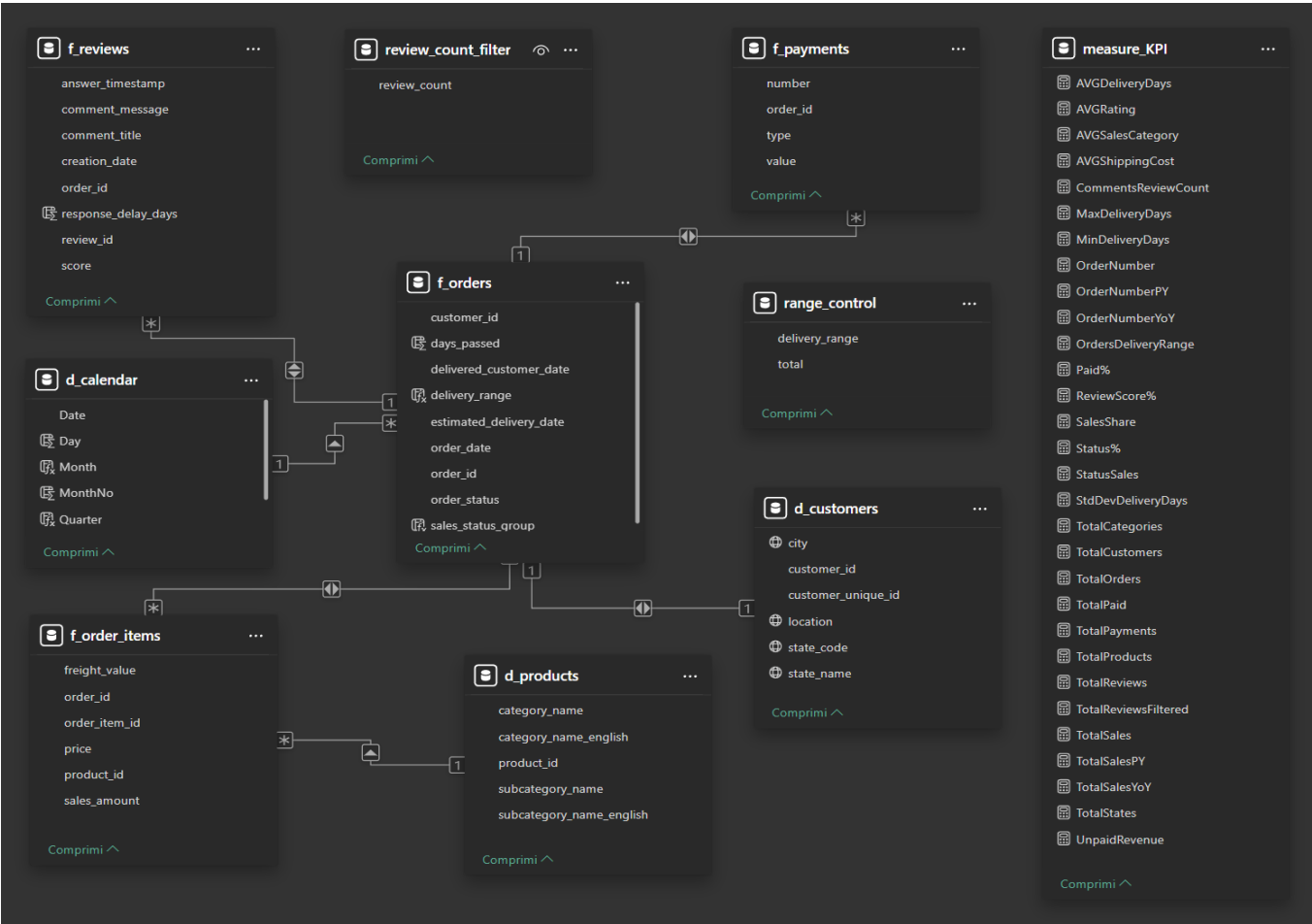
- **d\_products** (include dati sui prodotti venduti). Abbiamo 32.951 righe, che sono anche valori distinti ed univoci ('product\_id').

Pulizia eseguita:

- ✓ sostituito i valori 'null', nella colonna 'product\_category\_name', con 'Nao\_codificado';
- ✓ eliminato le colonne che non servivano, tranne 'product\_id' e 'product\_category\_name';
- ✓ eseguito merge con la tabella **product\_category\_name\_translation** (che traduce product\_category\_name in inglese);
- ✓ esaminando le 74 categorie presenti nel dataset, ho deciso di classificarle come 'subcategory\_name' (con relativa 'subcategory\_name\_english'), e creato nuove colonne personalizzate ('category\_name' e 'category\_name\_english'), con un totale di 14 categorie;

Tabelle create:

- **d\_calendar** (contiene informazioni sulle date, come giorni, mesi e anni, utile per analisi temporali), in Power BI Desktop;
- **measure\_KPI** per raggruppare le misure create;
- **range\_control** per controllo valori totali del 'delivery\_range';
- **review\_count\_filter** per distinguere ordini che hanno più di una recensione e poter creare misura per applicare filtro.



+ Nuova relazione		⚡ Rilevamento automatico		Modifica	Elimina	Filtro
<input type="checkbox"/> Da: tabella (colonna) ↑	Relazione	A: tabella (colonna)	Stato			
<input type="checkbox"/> f_order_items (order_id)	* —> 1	f_orders (order_id)	Attivo			
<input type="checkbox"/> f_order_items (product_id)	* —> 1	d_products (product_id)	Attivo			
<input type="checkbox"/> f_orders (customer_id)	1 —> 1	d_customers (customer_id)	Attivo			
<input type="checkbox"/> f_orders (order_date)	* —> 1	d_calendar (Date)	Attivo			
<input type="checkbox"/> f_payments (order_id)	* —> 1	f_orders (order_id)	Attivo			
<input type="checkbox"/> f_reviews (order_id)	* —> 1	f_orders (order_id)	Attivo			

## 2. MISURE CALCOLATE (DAX)

Misure principali richieste	Commento
<b>OrderNumber</b>	Numero totale di ordini
<b>OrderNumberPY</b>	Numero totale di ordini dell'anno precedente
<b>OrderNumberYoY</b>	Variazione % degli ordini rispetto all'anno scorso
<b>TotalSales</b>	Ricavi totali generati dagli ordini
<b>TotalSalesPY</b>	Ricavi totali dell'anno precedente
<b>TotalSalesYoY</b>	Variazione % dei ricavi anno su anno
<b>AVGRating</b>	Valutazione media delle recensioni
<b>ReviewScore%</b>	Percentuale di ogni punteggio recensione
<b>TotalReviews</b>	Numero totale di recensioni

---

### 1) RICHIESTE PER ANDAMENTO ORDINI

**OrderNumber = COUNT(f\_orders[order\_id])**

✓ Uso nel report:

- KPI card;
- Grafico linee per "Andamento Ordini nel tempo";
- Grafico barre per "Totale Ordini per Stato" e mappa geografica;

**OrderNumberPY = CALCULATE([OrderNumber], SAMEPERIODLASTYEAR(d\_calendar[Date]))**

✓ Uso nel report:

- Confronto anno corrente vs anno precedente (KPI card e Tooltip);

**OrderNumberYoY = DIVIDE([OrderNumber] - [OrderNumberPY], [OrderNumberPY])**

✓ Uso nel report:

- KPI card e Tooltip.

### 2) RICHIESTE PER ANDAMENTO RICAVI

**TotalSales = SUM(f\_order\_items[sales\_amount])**

✓ Uso nel report:

- KPI card;
- Grafico linee per "Andamento Ricavi nel tempo";
- Grafico barre per "Totale Ricavi per Stato" e mappa geografica;

**TotalSalesPY = CALCULATE([TotalSales], SAMEPERIODLASTYEAR(d\_Calendar[date]))**

✓ Uso nel report:

- Confronto anno corrente vs anno precedente (KPI card e Tooltip);

**TotalSalesYoY = DIVIDE([TotalSales] - [TotalSalesPY], [TotalSalesPY])**

✓ Uso nel report:

- KPI card e Tooltip.

### 3) RICHIESTE PER DISTRIBUZIONE RATING

**AVGRating = AVERAGE(f\_reviews[score])**

✓ Uso nel report:

- KPI card;

**ReviewScore% =**

**DIVIDE(COUNTROWS(VALUES(f\_reviews[review\_id])),CALCULATE(COUNTROWS(f\_reviews),ALL(f\_reviews[score]))))**

✓ Uso nel report:

- Grafico a torta: "Distribuzione % review\_score";

**TotalReviews = COUNT(f\_reviews[review\_id])**

✓ Uso nel report:

- Grafico linee per "Andamento recensioni nel tempo" e Tooltip;
- Grafico a barre: "Conteggio review per Categoria" e Tooltip.

### 4) MISURE PER ULTERIORI ANALISI

✓ Misure create per le KPI card, nella pagina Overview:

- Numero totale pagamenti

**TotalPayments = DISTINCTCOUNT(f\_payments[order\_id])**

- Numero totale prodotti univoci

**TotalProducts = CALCULATE(DISTINCTCOUNT(f\_order\_items[product\_id]))**

- Numero totale clienti univoci

**TotalCustomers = DISTINCTCOUNT(d\_customers[customer\_unique\_id])**

- Numero totale categorie univoche

**TotalCategories = CALCULATE(DISTINCTCOUNT(d\_products[category\_name]),  
CROSSFILTER(f\_order\_items[product\_id], d\_products[product\_id], BOTH))**

- Numero totale stati univoci

**TotalStates = DISTINCTCOUNT(d\_customers[state\_name])**

✓ Misura creata solo per grafico 'Top 5 Categories by Orders', che restituisce totale ordini per le prime 5 categorie, rispetto al contesto, nella pagina Overview:

- Numero totale ordini per categoria prodotto

**TotalOrders = DISTINCTCOUNT(f\_order\_items[order\_id])**

- ✓ Misura creata solo per tabella 'Top 5 Categories by Sales', che restituisce vendite per le prime 5 categorie, rispetto al contesto, nella pagina Overview:
  - Incidenza percentuale della categoria sul totale vendite  
**SalesShare = DIVIDE([TotalSales], CALCULATE([TotalSales], ALL(d\_products)))**
- ✓ Misure create solo per grafico 'How Fast Are We Improving?' e 'Who Takes the Longest?' e relativo tooltip, nella pagina Orders:
  - Media dei giorni trascorsi tra acquisto e consegna (days\_passed)  
**AVGDeliveryDays = AVERAGE(f\_orders[days\_passed])**
  - Min / Max / Deviazione Standard dei giorni di consegna (popolazione intera)  
**MinDeliveryDays = MIN(f\_orders[days\_passed])**  
**MaxDeliveryDays = MAX(f\_orders[days\_passed])**  
**StdDevDeliveryDays = STDEV.P(f\_orders[days\_passed])**
  - Numero di ordini per fascia di consegna ('delivery\_range'), utile per analizzare la distribuzione dei tempi di spedizione  
**OrdersDeliveryRange = CALCULATE(COUNTROWS(f\_orders), ALLEXCEPT(f\_orders, f\_orders[delivery\_range]))**
- ✓ Misura creata per analizzare la distribuzione delle vendite in base allo stato dell'ordine, nella pagina Sales:
  - Calcolare il totale delle vendite per ciascun gruppo di stato d'ordine, mantenendo il contesto 'sales\_status\_group'  
**StatusSales = CALCULATE([TotalSales], ALLEXCEPT(f\_orders, f\_orders[sales\_status\_group]))**
  - Percentuale delle vendite per ciascun gruppo, rispetto al totale vendite  
**Status% = DIVIDE([StatusSales], CALCULATE([TotalSales], ALL(f\_orders)))**
- ✓ Misure create per avere percentuale di incasso e usate nel tooltip, nella pagina Sales:
  - Totale di tutti i pagamenti ricevuti dagli ordini  
**TotalPaid = SUM(f\_payments[value])**
  - Revenue non ancora incassato (differenza tra vendite totali e importo incassato)  
**UnpaidRevenue = [TotalSales] - [TotalPaid]**
  - Percentuale di incasso rispetto alle vendite totali  
**Paid% = DIVIDE([TotalPaid], [TotalSales])**
- ✓ Misure create per medie costi, nella pagina Sales (sezione Category):
  - Valore medio delle vendite per categoria prodotto  
**AVGSalesCategory = AVERAGEX(VALUES(d\_products[category\_name]),[TotalSales])**
  - Valore medio del costo di spedizione per ordine **AVGShippingCost = AVERAGE(f\_order\_items[freight\_value])**



✓ Misure create per KPI card 'Review' per collegamento filtri, nella pagina Reviews (sezione Comments):

- Misura che conta le recensioni filtrate, in base al numero selezionato da 'review\_count\_filter'

**TotalReviewsFiltered = VAR SelectedCount =**

**SELECTEDVALUE(review\_count\_filter[review\_count]) RETURN**

**CALCULATE(COUNTROWS(f\_reviews), FILTER(f\_reviews, VAR ReviewPerOrder =**

**CALCULATE(COUNTROWS(f\_reviews), ALLEXCEPT(f\_reviews, f\_reviews[order\_id]))**

**RETURN ISBLANK(SelectedCount) || ReviewPerOrder = SelectedCount))**

- Misura che restituisce 1, se il numero di recensioni per ordine corrisponde al filtro selezionato, altrimenti 0 (per logica condizionale nei visual)

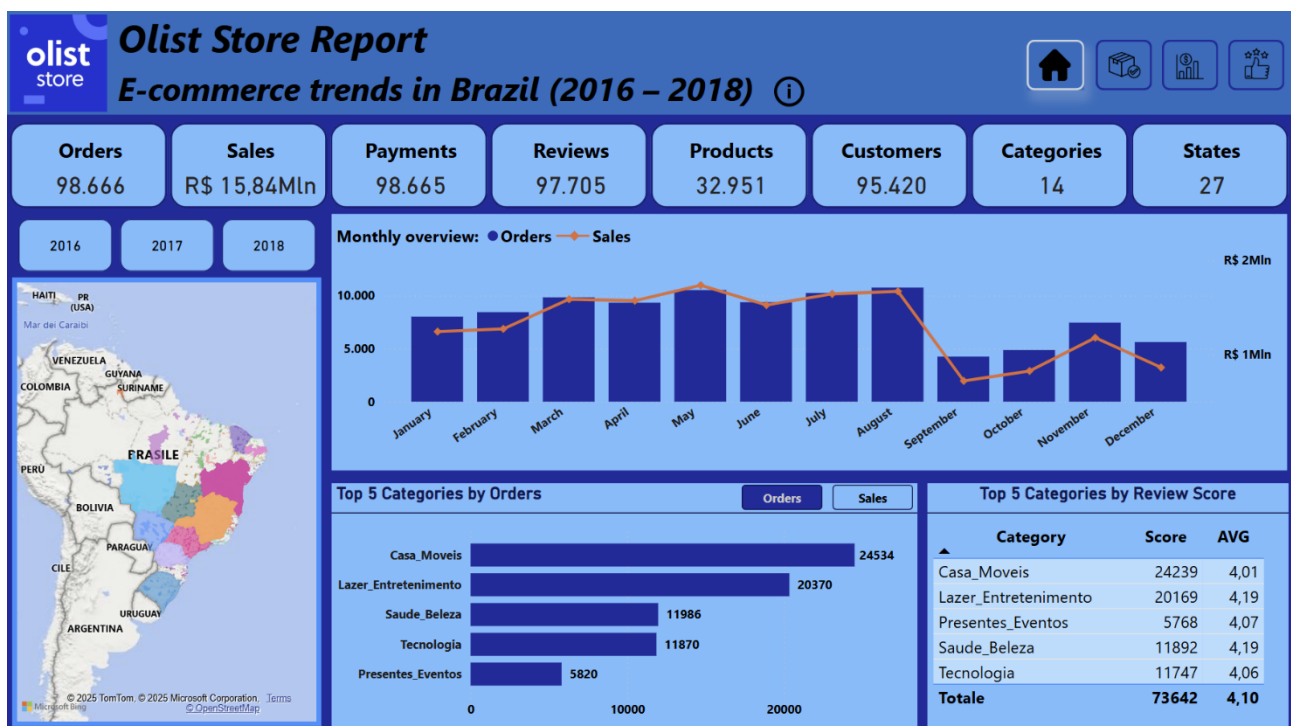
**CommentsReviewCount = VAR SelectedCount =**

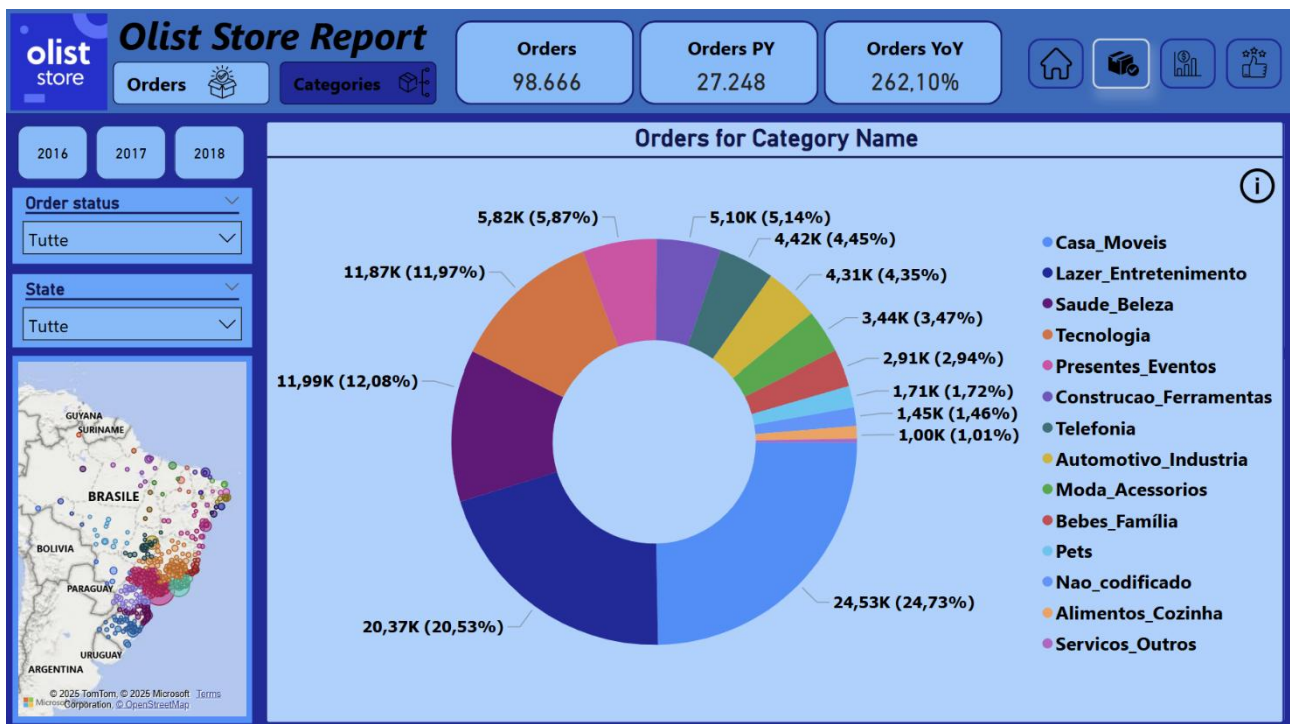
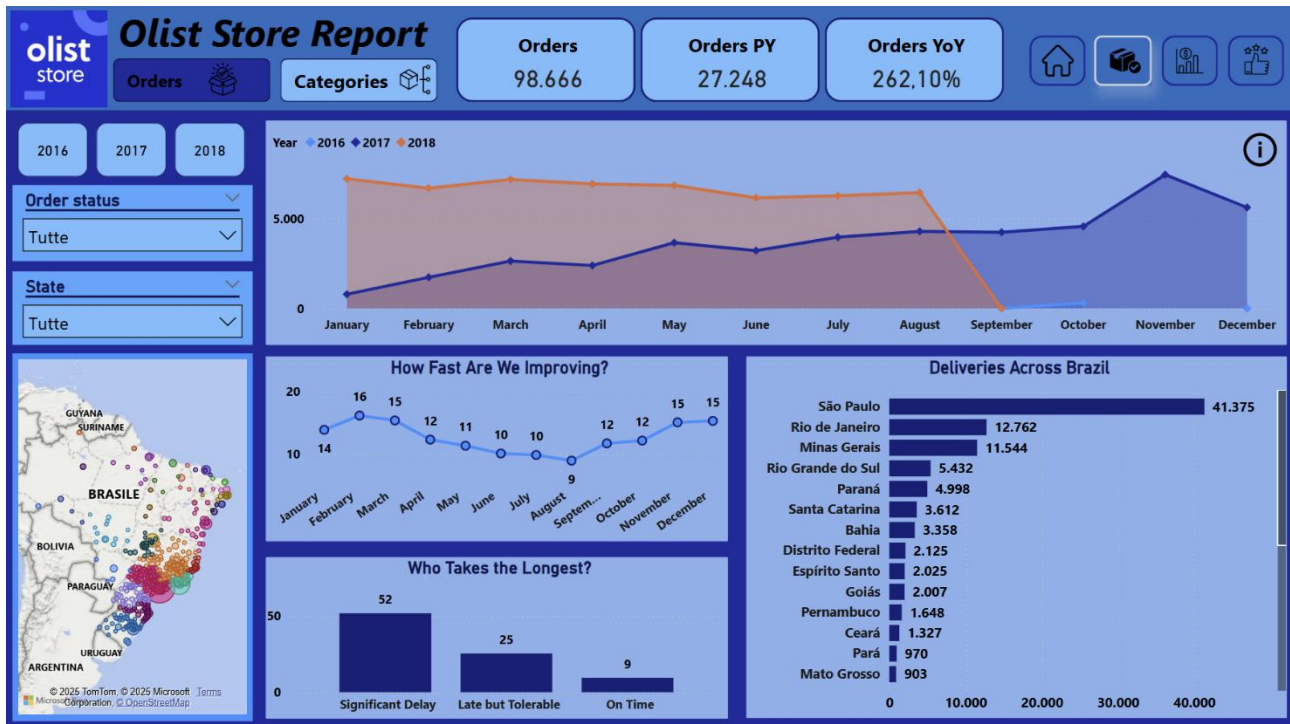
**SELECTEDVALUE(review\_count\_filter[review\_count]) VAR ReviewsPerOrder =**

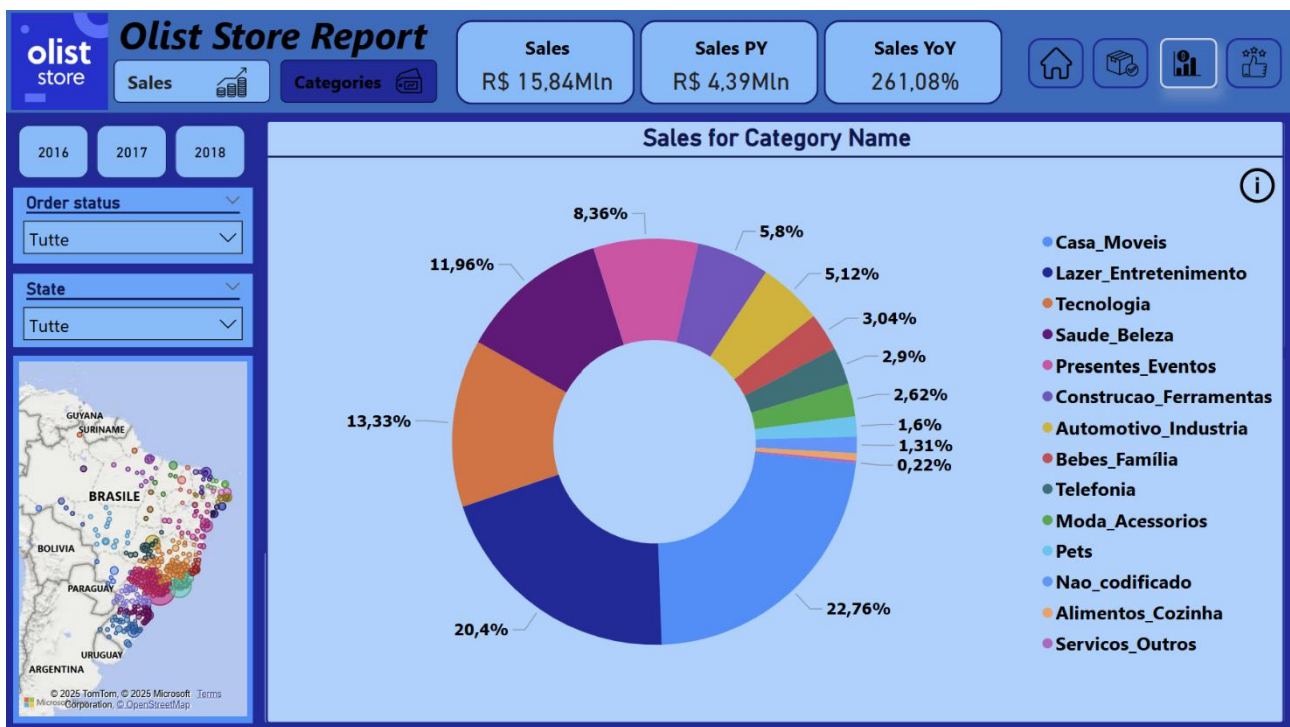
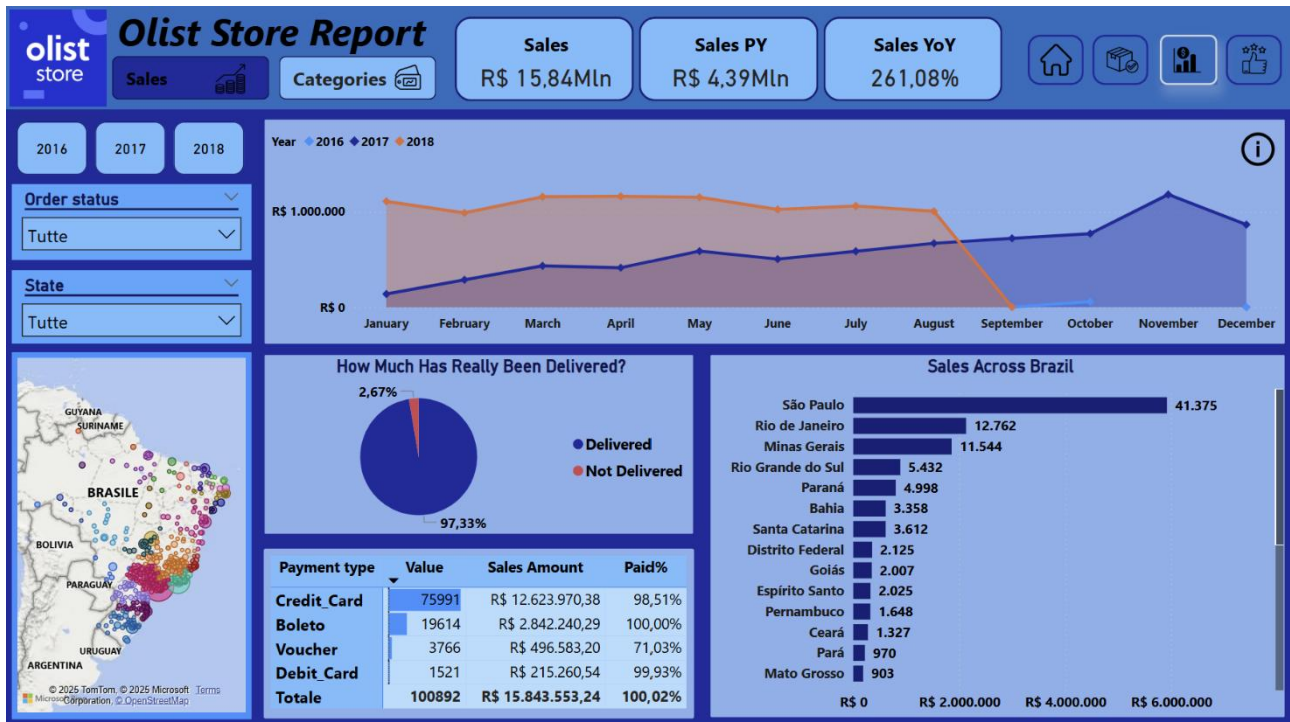
**CALCULATE(COUNTROWS(f\_reviews), ALLEXCEPT(f\_reviews, f\_reviews[order\_id]))**

**RETURN IF(ISBLANK(SelectedCount) || ReviewsPerOrder = SelectedCount, 1, 0)**

### 3. VISUALIZZAZIONE REPORT







2016

2017

2018

Order status

Tutte

State

Tutte

Year

2016

2017

2018

Sales for Category Name

11,96%

8,36%

5,8%

5,12%

3,04%

2,9%

2,62%

1,6%

1,31%

0,22%

22,76%

20,4%

13,33%

Casa\_Moveis

Lazer\_Entretenimento

Tecnologia

Saude\_Beleza

Presentes\_Eventos

Construcao\_Ferramentas

Automotivo\_Industria

Bebes\_Familia

Telefonia

Moda\_Acessorios

Pets

Nao\_codificado

Alimentos\_Cozinha

Servicos\_Outros

Map

BRASILE

GUYANA

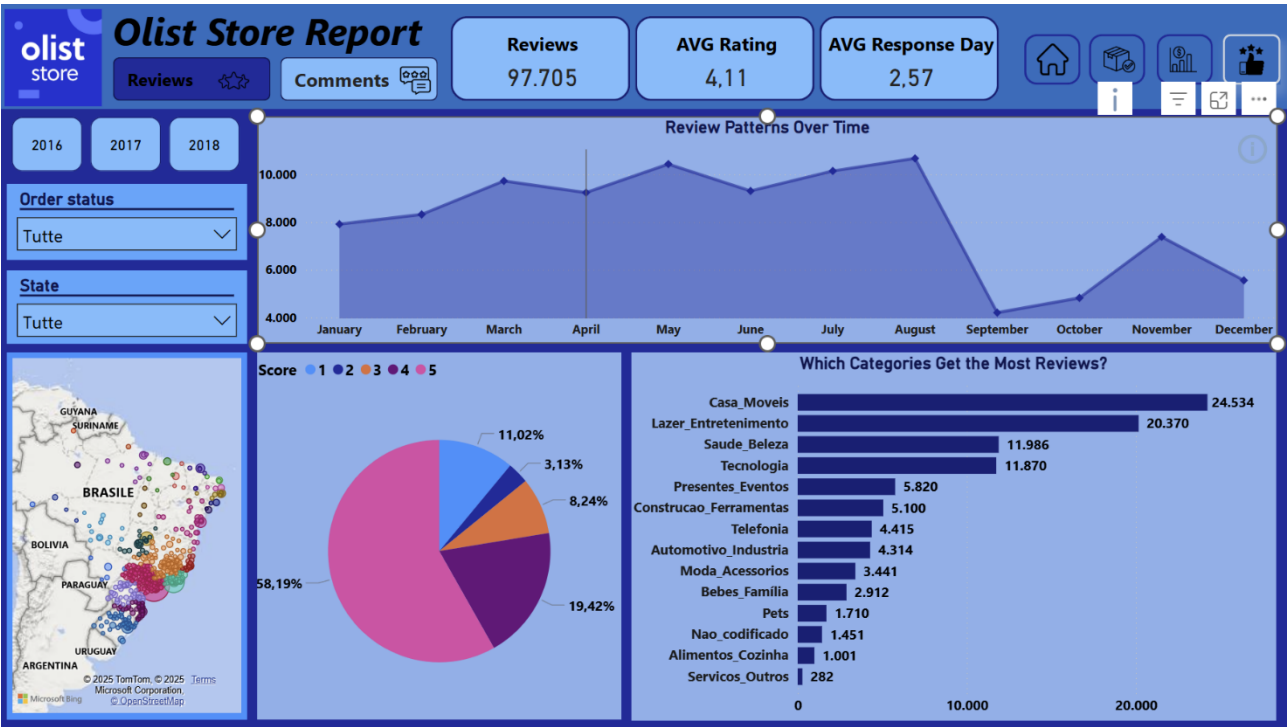
SURINAME

BOLIVIA

PARAGUAY

URUGUAY

ARGENTINA



olist  
store

Reviews

Comments

Olist Store Report

Reviews for Order

Tutte

Reviews

97705

Score

Tutte

Home

Box

Bar Chart

Star

Info

Menu

Share

More

Order_id	Date	Score	Comment
00010242fe8c5a6d1ba2dd792cb16214	21/09/2017	5	Perfeito, produto entregue antes do combinado.
00018f77f2f0320c557190d7a144bdd3	13/05/2017	4	
000229ec398224efeca0657da4fc703e	23/01/2018	5	Chegou antes do prazo previsto e o produto surpreendeu pela qualidade. Muito satisfatório.
00024acbcd0a6daa1e931b038114c75	15/08/2018	4	
00042b26cf59d7ce69dfabb4e55b4fd9	02/03/2017	5	Gostei pois veio no prazo determinado .
00048cc3ae777c65dbb7d2a0634bc1ea	23/05/2017	4	
00054e8431b9d7675808bcb819fb4a32	19/12/2017	4	
000576fe39319847cbb9d288c5617fa6	10/07/2018	5	
0005a1a1728c9d785b8e2b08b904576c	30/03/2018	1	Na descrição do produto quando fui efetuar a compra, estava detalhado os pentes de disfarce. Quando recebo o produto verifiquei e notei que não havia os pentes 0.5 e 1.5
0005f50442cb953dcd1d21e1fb923495	05/07/2018	4	
00061f2a7bc09da83e415a52dc8a4af1	30/03/2018	5	A caixa do produto veio com uma pequena avariação, mas de resto veio perfeito...
00063b381e2406b52ad429470734ebd5	08/08/2018	5	Fiquei um pouco triste, achei que a cor do coração seria verde, conforme a foto..... Mas depois que recebi o coração cor de rosa que fui ver que são imagens ilustrativas...
0006ec9db01a64e59a68b2c340bf65a7	31/07/2018	5	Excelente serviço de compras on line.
0008288aa423d2a3f00fcb17cd7d8719	27/02/2018	5	
0009792311464db532ff765bf7b182ae	23/08/2018	5	
0009c9a17f916a706d71784483a5d643	01/05/2018	4	
000aed2e5dbad2f9ddb70584c5a2ded	19/05/2018	1	Mudo minha opinião assim que receber o produto. Consta no site que o produto foi entregue, mas não recebi.
000c3e6612759851cc3cbb4b83257986	20/08/2017	5	Recebi td certinho antes do prazo!
000e562887b1f2006d75e0be9558292e	13/03/2018	3	Bom produto
000e6d38ae8c00bbc5a30573b99628	28/03/2018	3	5
000e906b789b55f64edcb1f84030f90d	09/12/2017	3	
000f25f4d72195062c040b12dce9a18a	23/03/2018	4	
001021efaa8636c29475e7734483457d	09/03/2018	4	super recomendo
0010b2e5201cc5f1ae7e9c6cc8f5bd00	24/09/2017	5	
00119f934e539cf26f92b9ef0cdfed8	17/08/2017	5	Recebi o long 10 dias após a compra, e chegou em perfeito estado. A ilustração é idêntica à imagem do site. O material é de ótima qualidade e o tamanho é perfeito. Rec
0011d82c4b53e22e84023405fb467e57	27/01/2018	4	
00125cb692d04887809806618a2a145f	08/04/2017	5	produto como descrito. satisfetíssima com a compra.