Data Science & Business Analytics

Module 6 - Analyzing and Visualizing Data

Ana Rita Ribeiro

Brazilian e-commerce public dataset by Olist

1. **Base Dashboard:** Sales and Revenue (Sales Department)

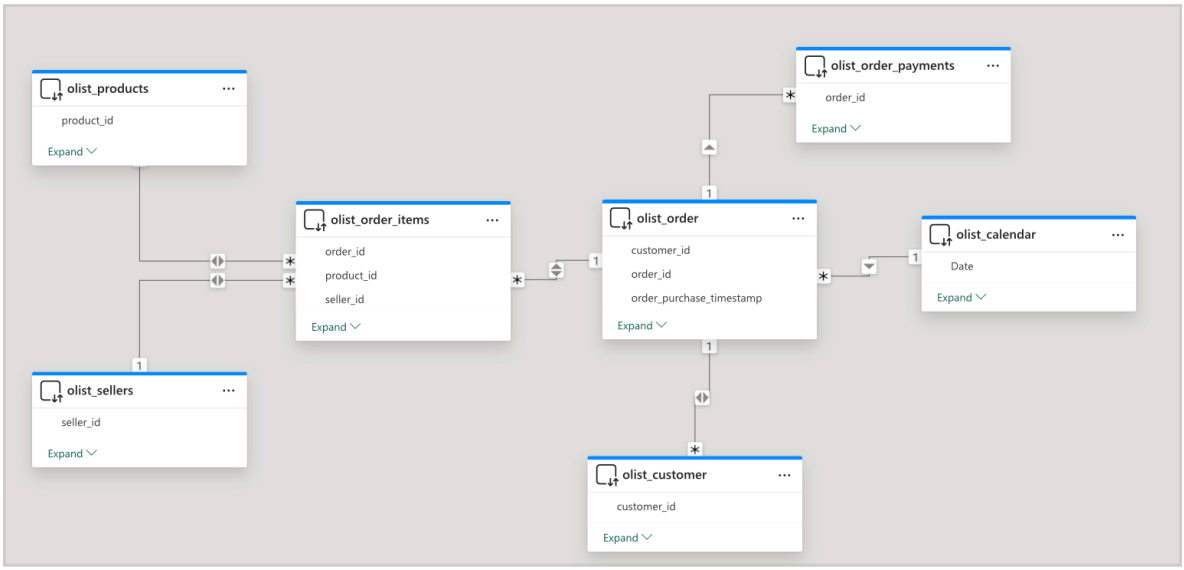
`Key Metrics: Total Sales, Revenue by Product Category, Sales by State, Monthly Sales Trends, Average Order Value.`

`Charts: Sales trend over time, top-selling products and categories, revenue by region, sales by payment type.`

`Insights: Identify best-selling products, track revenue growth, and analyze sales patterns across different regions.`

2. Tables used (downloaded from Kaggle): *olist_customer; olist_order; olist_order_items, olist_order_payments; olist_products; olist_sellers*.
3. Table created in Power Query: *olist_calendar*. Created to have Date, Year, Month Number, Month Name, Year-Month.

Connections created between fact tables and dimension tables:



- Data transformation (Power Query)

In the *olist_order* table, I chose to convert the data type from date/time to date,

as I believed the time component wouldn't significantly impact the dashboard's insights and simplifying it would make measure creation easier.

Column Creation: In the olist_order table, I created a new column called Delivery_Days to calculate the difference between the delivery date and the purchase date, allowing for an evaluation of delivery duration.

Measures Created (DAX)

- olist_customer:

# of Customers = DISTINCTCOUNT(olist_customer[customer_id])
Orders_by_State = CALCULATE(COUNTROWS('olist_order'), CROSSFILTER('olist_order'[customer_id], 'olist_customer'[customer_id], BOTH))
AOV_by_State = DIVIDE([Revenue_by_State], [Orders_by_State])
Top_State_By_Revenue =

```
CALCULATE (
    MAX('olist_customer'[customer_state]),
    TOPN (
        1,
        SUMMARIZE (
            'olist_customer',
            'olist_customer'[customer_state],
            "TotalRevenue",
            CALCULATE(SUM('olist_order_items'[price]))
        ),
        [TotalRevenue], DESC
    )
)
```

Top_City_By_Orders =

```
CALCULATE (
    MAX('olist_customer'[customer_city]),
    TOPN (
        1,
        SUMMARIZE (
            'olist_customer',
            'olist_customer'[customer_city],
            "OrderCount",
            CALCULATE(COUNTROWS('olist_order'))
        ),
        [OrderCount], DESC
    )
)
```

🔷 olist_customer

Measure	Description
`# of Customers`	Count of unique customers.
`Orders_by_State`	Number of orders per state using bidirectional filtering.
`AOV_by_State`	Average order value by state, calculated from revenue and orders.
`Top_State_By_Revenue`	State with the highest total revenue.
`Top_City_By_Orders`	City with the highest number of orders.

- olist_order:

Orders_by_City =
CALCULATE(
    COUNTROWS('olist_order'),
    VALUES('olist_customer'[customer_city])
)
Dashboard_Title = "Brazilian E-Commerce Public Dataset by Olist. EDIT Module 6"
Delivered_Orders =
CALCULATE(
    COUNTROWS('olist_order'),
    'olist_order'[order_status] = "delivered"
)
Canceled_Orders =
CALCULATE(
    COUNTROWS('olist_order'),
    'olist_order'[order_status] = "canceled"
)
On_Time_Percentage = DIVIDE([Delivered_Orders], [Total_Orders])
Gauge_Min = 0
Gauge_Max = 1
Gauge_Target = 0.95

🔷 olist_order

Measure	Description
`Orders_by_City`	Number of orders by city.
`Dashboard_Title`	Static text for use as a title.
`Delivered_Orders`	Count of orders with status “delivered”.
`Canceled_Orders`	Count of orders with status “canceled”.
`On_Time_Percentage`	Share of delivered orders over total orders.
`Gauge_Min/Max/Target`	Constants for use in gauge visualizations.

- olist_order_items

Total_Revenue_With_Shipping = SUMX(olist_order_items, olist_order_items[price] + olist_order_items[freight_value])
Average_Order_Value = DIVIDE([Total_Revenue_With_Shipping], [Total_Orders])
Monthly_Revenue =
CALCULATE(
    [Total_Revenue_With_Shipping],
    FILTER(
        olist_order_items,
        EOMONTH(olist_order_items[shipping_limit_date], 0) = EOMONTH(TODAY(), 0)
)
)
Revenue_by_Product = CALCULATE(SUM(olist_order_items[price]))
# of Orders = COUNTROWS(olist_order_items)
Total_Orders = DISTINCTCOUNT(olist_order[order_id])
Total_Sales =
CALCULATE(
    SUM('olist_order_items'[price]),
    USERELATIONSHIP('olist_calendar'[Date], 'olist_order'[order_purchase_timestamp])
)
Top_Category =
CALCULATE (

```
    MAX('olist_products'[product_category_name]),
    TOPN(1, SUMMARIZE('olist_order_items',
'olist_products'[product_category_name],
        "Revenue", SUM('olist_order_items'[price])), [Revenue], DESC)
)
```

Revenue_by_State =

```
CALCULATE(
    SUM('olist_order_items'[price]),
    USERELATIONSHIP('olist_order'[order_id], 'olist_order_items'[order_id])
)
```

Profit_Ratio = DIVIDE(SUM('olist_order_items'[price]) * 0.2, SUM('olist_order_items'[price]))

🔷 **olist_order_items**

Measure	Description
`Total_Revenue_With_Shipping`	Total revenue including shipping costs.
`Average_Order_Value`	Revenue per order.
`Monthly_Revenue`	Revenue for the current month.
`Revenue_by_Product`	Revenue by product category.
`# of Orders`	Total order lines.
`Total_Orders`	Distinct orders.
`Total_Sales`	Total item sales using calendar relationship.
`Top_Category`	Best-selling product category.
`Revenue_by_State`	Revenue by customer state (via relationship).
`Profit_Ratio`	Simplified profit margin (assumed 20% profit on sales).

- olist_order_payments

Top_Payment_Type =

```
CALCULATE (
    MAX('olist_order_payments'[payment_type]),
    TOPN(1,
        SUMMARIZE('olist_order_payments',
```

```
            'olist_order_payments'[payment_type],
            "PayVal", SUM('olist_order_payments'[payment_value])),
        [PayVal], DESC
    )
)
```

🔷 **olist_order_payments**

Measure	Description
`Top_Payment_Type`	Most common payment method by value.

- olist_products

Top Categories by Revenue =

```
TOPN(
    5,
    SUMMARIZE(
        olist_order_items,
        olist_products[product_category_name],
        "Revenue", SUM(olist_order_items[price])
    ),
    [Revenue],
    DESC
)
```

🔷 **olist_products**

Measure	Description
`Top Categories by Revenue`	Top 5 product categories by revenue.

DASHBOARD:

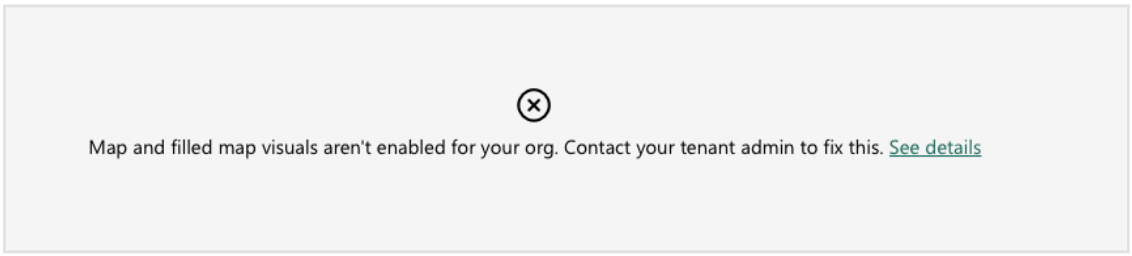
4 pages:

1. Overview: summary of key business metrics, including total revenue, total sales, number of orders, and customer count. Visuals: trend lines and word clouds.
2. Sales Insights: Focuses on analyzing sales performance across

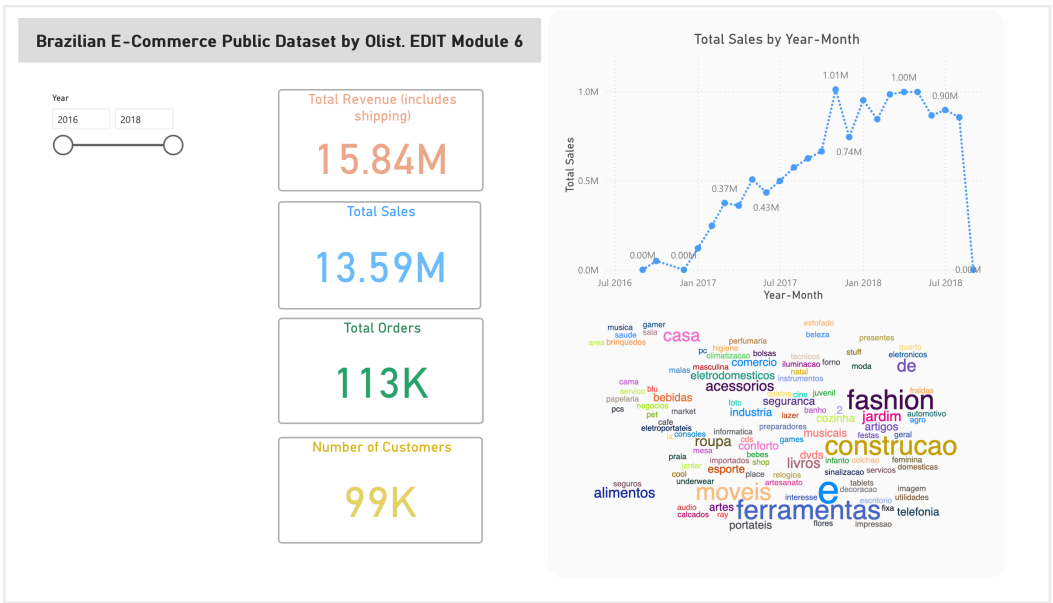
product categories and payment methods. This page identifies top-selling categories, the most used payment types, and highlights average order values to better understand customer purchasing behavior.

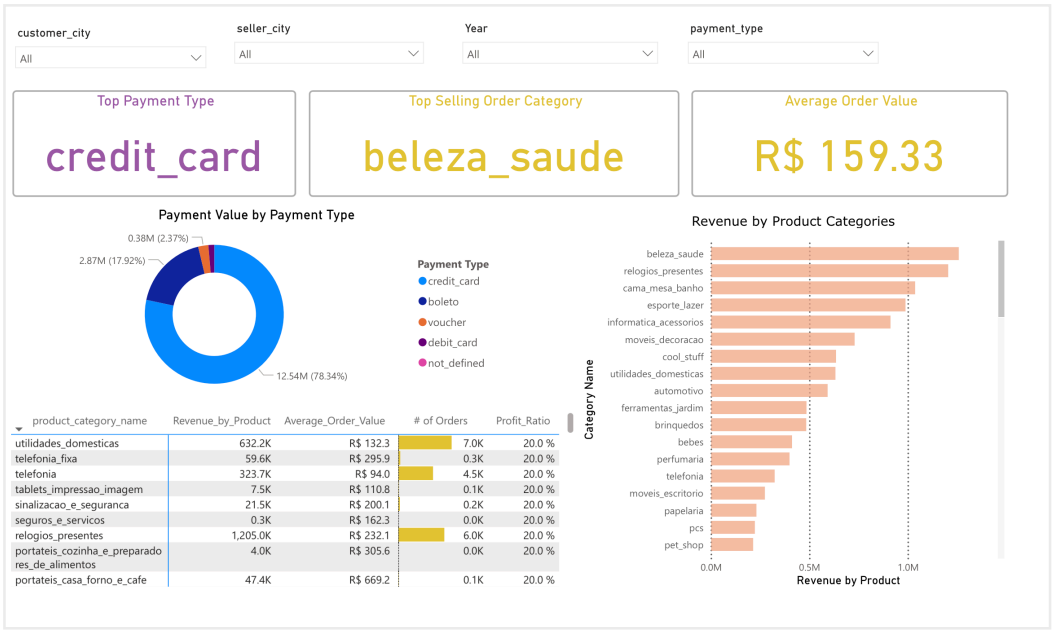
3. Sales Distribution: Breaks down sales by geography and customer demographics. This includes revenue and order distribution across states and cities, helping identify regional trends and high-performing areas.

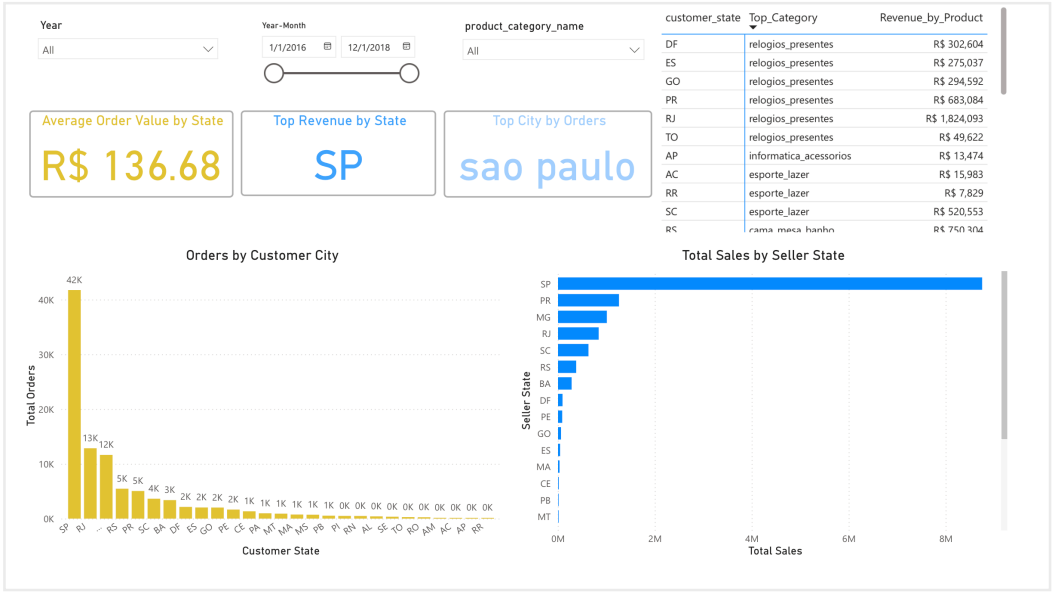
Tried to do some visualizations using map ou filled map but always encountered this warning:

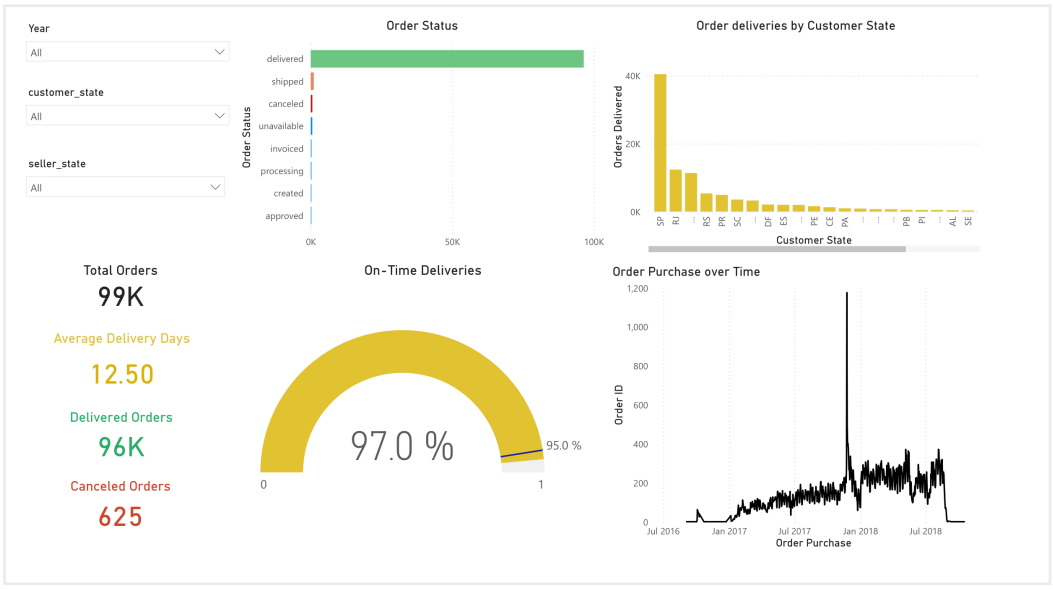


4. Delivery Performance: Evaluates the efficiency and reliability of order deliveries. Displays on-time delivery rates, average delivery time, and trends over time, offering insight into customer satisfaction and operational effectiveness.




customer_city
All
seller_city
All
Year
All
payment_type
All
Top Payment Type
credit_card
Top Selling Order Category
beleza_saude
Average Order Value
R$ 159.33
Payment Value by Payment Type
0.38M (2.37%)
2.87M (17.92%)
12.54M (78.34%)
Payment Type
credit_card
boleto
voucher
debit_card
not_defined
Revenue by Product Categories
beleza_saude
relogios_presentes
cama_mesa_banho
esporte_lazer
informatica_acessorios
moveis_decoracao
cool_stuff
utilidades_domesticas
automotivo
ferramentas_jardim
brinquedos
bebes
perfumaria
telefonia
moveis_escritorio
papelaria
pcs
pet_shop
Category Name
0.0M
0.5M
1.0M
Revenue by Product
product_category_name Revenue_by_Product Average_Order_Value # of Orders Profit_Ratio
utilidades_domesticas 632.2K R$ 132.3 7.0K 20.0 %
telefonia_fixa 59.6K R$ 295.9 0.3K 20.0 %
telefonia 323.7K R$ 94.0 4.5K 20.0 %
tablets_impressao_imagem 7.5K R$ 110.8 0.1K 20.0 %
sinalizacao_e_seguranca 21.5K R$ 200.1 0.2K 20.0 %
seguros_e_servicos 0.3K R$ 162.3 0.0K 20.0 %
relogios_presentes 1,205.0K R$ 232.1 6.0K 20.0 %
portateis_cozinha_e_preparadores_de_alimentos 4.0K R$ 305.6 0.0K 20.0 %
portateis_casa_forno_e_cafe 47.4K R$ 669.2 0.1K 20.0 %


Year
All
Year-Month
1/1/2016
12/1/2018
product_category_name
All
Average Order Value by State
R$ 136.68
Top Revenue by State
SP
Top City by Orders
sao paulo
customer_state Top_Category Revenue_by_Product
DF relogios_presentes R$ 302,604
ES relogios_presentes R$ 275,037
GO relogios_presentes R$ 294,592
PR relogios_presentes R$ 683,084
RJ relogios_presentes R$ 1,824,093
TO relogios_presentes R$ 49,622
AP informatica_acessorios R$ 13,474
AC esporte_lazer R$ 15,983
RR esporte_lazer R$ 7,829
SC esporte_lazer R$ 520,553
Orders by Customer City
Total Orders
Customer State
Total Sales by Seller State
Seller State
Total Sales


Year
All
customer_state
All
seller_state
All
Order Status
delivered
shipped
canceled
unavailable
invoiced
processing
created
approved
Order Status
0K
50K
100K
Order deliveries by Customer State
40K
20K
0K
Orders Delivered
SP
RJ
BS
PR
SC
DF
ES
PE
CE
PA
PB
PI
AL
SE
Customer State
Total Orders
99K
Average Delivery Days
12.50
Delivered Orders
96K
Canceled Orders
625
On-Time Deliveries
97.0 %
95.0 %
0
1
Order Purchase over Time
1,200
1,000
800
600
400
200
0
Order ID
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Order Purchase