

Olist Project Report

Dashboard link: <https://app.powerbi.com/groups/me/reports/2c142bbb-6a8d-4142-a384-6816f4f2650d/134945030290ba967076?experience=power-bi>

Data Understanding:

- Olist Ecommerce Dataset describes the sales data of the Brazilian Ecommerce website, It contains a vast amount of data collected for the years (2016 – 2017 – 2018)
- Data source: Kaggle (9 tables – CSV Files)
 - Customers – Customer ID
 - Geolocation – Geolocation zip code prefix
 - Orders – Order ID
 - Order Items – Order ID – Order Item ID
 - Order Payment – Order ID
 - Order Reviews – Review ID
 - Products – Product ID
 - Sellers – Seller ID
 - Product Category Name Translation In English – Product Category Name

Business Questions & Goals & KPIs:

- Our analysis is targeting all the departments and managerial levels who seeks growth , performance and logistics insights
 - Sales
 - Ceo
 - Logistics
 - Customer service and satisfaction
- Analysis Goal:
 - Increase sales and revenue
 - Improve customer satisfaction
 - Enhance product and category performance
 - Boast seller performance
 - Optimize delivery durations
- KPIs
 - Sales and revenue
 - What are the top performing products ?

- Which region generates the most sales ?
- Which product categories generate the most sales ?
- Customer insights & Delivery Performance
 - What is customer density for every state ?
 - Which customers generate the largest amount of revenue ?
 - What is the number of customers over time ?
- Seller Performance
 - Top Sellers by revenue ?
 - Sales contribution by sellers ?
 - Sales trend by sellers ?

Cleaning & Transformation:

We handled inconsistencies in data, implemented data filtering, Addressed missing values strategically through deletion, imputation, or averaging and used the appropriate transformation functions

- Geolocation;
 - Standardized State column values (Sao Paulo)
 - Removed duplicates from zip code column
 - Filtered longitude and latitude columns from outliers
- Order Payments;
 - Created a new payment ID
- Order reviews;
 - Replaced empty values with N/A
 - Created a new review ID
- Products;
 - Replaced the blank values
 - Removed duplicates from Product ID
 - Merged with Product category name translation table
 - Expanded The Product category name in English column
 - Stopped the loading of product category name translation table
 - Filtered rows from null values
- Seller;
 - Removed Duplicates from seller ID

- Orders;
 - Created a custom column to calculate the delivery duration by subtracting customers purchase time stamp from customer delivery date
 - Merge with Customers
 - Expanded customer zip code prefix
 - Merge with order items
 - Expanded the whole table
 - Stopped the loading of order items table
 - Merge with order payments
 - Expanded payment ID
 - Merge with sellers
 - Expanded seller zip code prefix
 - Merge with order reviews
 - Expanded review ID
 - Merge with Geolocation
 - Expanded geolocation state
 - Removed null values

Data modeling & relationships:

- Created a star schema – fact table; orders
- Created a date table, marked the table as a date table and created a many to one relationship between the table and my fact table (orders)
- Created a state data table and created a many to one relationship with the fact table
- Created a many to one relationship between fact table and the other 8 Dims (Geolocation , order reviews, order payments ,sellers ,products ,customers ,state data , Date table)

Data Visuals and DAX:

We created three dashboards , every one of them represents a view of a specific KPI

- **Sales Dashboard**
 1. Total Sales = `SUM(Orders[Price])`
 2. Num. Of Orders = `DISTINCTCOUNT(Orders[Order ID])`
 3. Average Order Value = `DIVIDE([Total Sales],[Num. Of Orders])`
 4. Current Year sales = `TOTALYTD([Total Sales],DATESYTD('Date Table'[Date]))`

5. Total sales over time, line chart to visualize the sales trend over time
6. Top 10 categories by sales, stacked column chart , sales by product category to show the distribution of sales across different product categories
7. Number of orders and total sales by state, stacked area chart to show the performance of every state in the terms of orders and sales
8. Top 10 products by sales, stacked bar chart to visualize the top performing products in terms of sales
9. Total sales by payment method ,pie chart to visualize the distribution of sales for every payment method
10. Sales distribution by state, a map visual

- **Customers Dashboard**

1. Num. Of Customers = `DISTINCTCOUNT(Customers[Customer ID])`
2. Average Review Score = `AVERAGE('Order Reviews'[Review Score])`
3. Average Delivery Time = `AVERAGE(Orders[Delivery Duration])`
4. Num. Of Products = `COUNT(Products[Product ID])`
5. Top 10 Customers by revenue, stacked bar chart to visualize the top 10 contributing customers to the revenue
6. Customers by time , line chart to visualize customers trend and numbers over time
7. Customers order status, stacked bar chart to visualize the delivery status
8. Customer reviews, pie chart to visualize customer review scores to indicate customer satisfaction
9. Customer payment type, pie chart to visualize the percentage of usage of every payment method for customers
10. Customer density map for every state

- **Sellers dashboard**

1. Number Of Sellers = `DISTINCTCOUNT(Sellers[Seller ID])`
2. Num. Of Items Sold = `COUNT(Orders[Order Item ID])`
3. Average Order Value = `DIVIDE([Total Sales],[Num. Of Orders])`
4. Total Sales = `SUM(Orders[Price])`
5. Top 10 sellers by sales, stacked column chart to visualize top 10 sellers generating sales
6. Sales by state and seller count, stacked bar chart to visualize states performance in sales and number of sellers contributing in each state
7. Sales by sellers city , bar chart to visualize the best performing seller cities
8. Top 5 sellers contribution in sales, pie chart to visualize the generated sales amount for each of the top 5 sellers
9. Top 5 sellers trend over time, line chart to track sales trend over time and performance fluctuations for each one of the top 5 sellers

10. Density map for sellers by state

- **DAX**

1. **We created a date table to ease our analysis and visualizations related to dates and over time trends and marked our table as a date table**

DAX Formula:

Date Table =

```
ADDCOLUMNS(  
    CALENDAR(DATE(2016, 1, 1), DATE(2018, 12, 31)),  
    "Year", YEAR([Date]),  
    "Month Name", FORMAT([Date], "MMMM"),  
    "Month Number", MONTH([Date]),  
    "Quarter", "Q" & FORMAT([Date], "Q"),  
    "Day", DAY([Date]),  
    "Weekday", FORMAT([Date], "dddd")  
)
```

2. **Also we created a state data table to ease using map visuals that give specific insights for every state.**
3. **We merged geolocation with orders , expanded prices , removed all columns except geolocation state and price and grouped by geolocation, aggregating the sum of prices**
4. **Then in the power view we created four columns for the state data table**

Customer Count =

```
CALCULATE(  
    COUNT(Orders[Customer ID]),  
    FILTER(  
        Geolocation,  
        Geolocation[Geolocation State] = 'State Data'[Geolocation State]  
    ))
```

Seller Count =

```
CALCULATE(  
    DISTINCTCOUNT(Orders[Seller ID]),  
    FILTER(  
        Geolocation,  
        Geolocation[Geolocation State] = 'State Data'[Geolocation State]  
    ))
```

))

```
Latitude =  
CALCULATE(  
    FIRSTNONBLANK(Geolocation[Geolocation Latitude], 0),  
    FILTER(  
        Orders,  
        Orders[Geolocation State] = 'State Data'[Geolocation State]  
    )  
)
```

```
Longitude =  
CALCULATE(  
    FIRSTNONBLANK(Geolocation[Geolocation Longitude], 0),  
    FILTER(  
        Orders,  
        Orders[Geolocation State] = 'State Data'[Geolocation State]  
    )  
)
```

Conclusion of the analysis:

- **Key insights and findings:**
 1. the analysis showed a consistent growth of sales over time in the analyzed period (2016 – 2017 – 2018) in certain months specially at the end of 2017 and the whole of 2018 showed a noticeable growth, indicating potential seasonal trends of events
 2. the analysis showed a diverse range for top selling products and categories, for example health beauty and watches gifts , both generate the highest revenue
 3. with the analysis of customers, we found out that sao paulo and rio de janero has the highest customer density and shipped orders also the state of parana and the city of ibitinga, all of them generate the most amount of sales
 4. the analysis of sellers helped us to see the top performing sellers and performance fluctuations and also the sellers density in each state

Recommendations:

- give more sellers training and support to spread the sales in more states, as we noticed that there is a huge gap between the top 10 performing sellers and the rest of them
- invest more on the top performing product categories
- create a marketing campaign to reach more states and cities as we noticed that the highest amount of sales only generated by only 3 states and cities