

# **Olist E-Commerce Analysis Report**

Role: Data Analyst

Tools Used: Excel, Power BI

---

## **Executive Summary:**

This project presents an exploratory data analysis (EDA) conducted for Olist E-Commerce. The objective is to display the Key Point Indicators through Excel and Power BI by data cleaning, data validation.

The dataset contains 9 Excel files, each .CSV files with Datasets of Customers, Geolocation, Ordered Items, Order Payments, Order Reviews, Orders, Products, Sellers, Product Category Name.

## **About the Dataset**

1. olist\_customers\_dataset
  - a. customer\_id : A unique code for each customer transaction or interaction.
  - b. customer\_unique\_id : A unique code for each customer.
  - c. customer\_zip\_code\_prefix : The Postal code for each customer.
  - d. customer\_city : The city of the customer.
  - e. customer\_state : The state of the customer.
  
2. olist\_geolocation\_dataset
  - a. geolocation\_zip\_code\_prefix : Prefix of the Brazillian zip code (CEP) for a specific area.
  - b. geolocation\_lat : Latitude coordinate of the location.
  - c. geolocation\_lng : Longitude coordinate of the location.
  - d. geolocation\_city : Name of the city corresponding to the geolocation.
  - e. geolocation\_state : Name of the state corresponding to the geolocation.

### 3. olist\_order\_items\_dataset

- a. order\_id : Unique identifier for the order.
- b. order\_item\_id : A sequential number identifying each item within an order.
- c. product\_id : Unique identifier for the product.
- d. seller\_id : Unique identifier for the seller of the product.
- e. shipping\_limit\_date : The deadline date for the product to be shipped.
- f. price : The price of the product.
- g. freight\_value : The shipping cost for the product.

### 4. olist\_order\_payments\_dataset

- a. order\_id : Unique identifier for each order
- b. payment\_sequential : Sequence number if there are multiple payments for an order.
- c. payment\_type : The method of payment.
- d. payment\_installments : Number of payment with a credit card.
- e. payment\_value : Total value paid for the order.

### 5. olist\_order\_reviews\_dataset

- a. review\_id : Unique identifier for each review.
- b. order\_id : Unique identifier for the order being reviewed.
- c. review\_score : The review score given by the customer, scale from 1 to 5.
- d. review\_comment\_title : Title of the review comment, if any.
- e. review\_comment\_message : Detailed comment or message left by the customer as part of the review.
- f. review\_creation\_date : The date and time when the review was created.
- g. review\_answer\_timestamp : Timestamp when the review was responded to by the seller or the platform.

### 6. olist\_orders\_dataset

- a. order\_id : Unique identifier for each order.
- b. customer\_id : Unique identifier for the customer who placed the order.

- c. order\_status : Current status of the order.
- d. order\_purchase\_timestamp : Timestamp when the order was placed.
- e. order\_approved\_at : Timestamp when the order was approved.
- f. order\_delivered\_carrier\_date : Timestamp when the order was handed over to the delivery carrier.
- g. order\_delivered\_customer\_date : Timestamp when the order was delivered to the customer.
- h. order\_estimated\_delivery\_date : Estimated date for order delivery.

#### 7. olist\_products\_dataset

- a. product\_id : Unique identifier for each product.
- b. product\_category\_name : The category to which the product belongs.
- c. product\_name\_length : Length of the product name.
- d. product\_description\_length : Length of the product description.
- e. product\_photos\_qty : Quantity of photos available for the product.
- f. product\_weight\_g : Weight of the product in grams.
- g. product\_length\_cm : Length of the product in centimeters.
- h. product\_height\_cm : Height of the product in centimeters.
- i. product\_width\_cm : Width of the product in centimeters.

#### 8. olist\_sellers\_dataset

- a. seller\_id : Unique identifier for each seller.
- b. seller\_zip\_code\_prefix : The zip code prefix for the seller's address.
- c. seller\_city : The city where the seller is located.
- d. seller\_state : The state where the seller is located.

#### 9. product\_category\_name\_translation

- a. product\_category\_name : Original category name in Portuguese.
- b. product\_category\_name\_english : Translated category name in English.

## Steps

### **In Excel:**

1. Some .CSV files are opened using Excel.
2. **“olist\_order\_reviews\_dataset”** is opened, and Blank Cells are found in some columns ‘review\_comment\_title’ and ‘review\_comment\_message’
3. Using Home -> Find & Select -> Replace, Blank Cells were replaced with ‘NIL’.
4. **“olist\_orders\_dataset”** is opened, and Blank Cells are found in ‘order\_delivered\_carrier\_date’ and ‘order\_delivered\_customer\_date’.
5. Using Home -> Find & Select -> Go to Special -> Blanks are selected
6. The dataset having Blank Cells are selected.
7. Right click on any selected Blank cells are Select ‘Delete Option’
8. A Pop Window appears are ‘Entire Row’ is selected. The Rows containing Blank Cells are deleted from the dataset.
9. The files are saved after the necessary changes.

### **In Power BI**

1. All The .CSV files are Loaded into Power Query Editor.
2. In ‘olist\_order\_items\_dataset’ file, ‘shipping\_time\_limit’ column is used and the date is splitted into ‘Year’, ‘Month Name’, ‘Day Name’, ‘Time’.
3. Made sure that ‘shipping\_limit\_date’ is Date/Time format.
4. ‘shipping\_limit\_date’ column is selected and the following action is performed.
  - a. Add Column -> Date -> Year -> Year is selected for ‘Year’ Column.
  - b. Add Column -> Date -> Month -> Name of Month is selected for ‘Month Name’ Column.
  - c. Add Column -> Date -> Day -> Name of Day is selected for ‘Day Name’ Column.
  - d. Add Column -> Time -> Time Only is selected for ‘Time’ Column.
5. In ‘olist\_order\_reviews\_dataset’ file, ‘review\_creation\_date’ is used and the column is splitted into ‘review\_creation\_Year’, ‘review\_creation\_Month\_Name’, ‘review\_creation\_Day\_Name’, and ‘review\_creation\_Time’.

6. 'review\_creation\_date' column is selected and the following action is performed.
  - a. Add Column -> Date -> Year -> Year is selected for 'review\_creation\_Year' Column.
  - b. Add Column -> Date -> Month -> Name of Month is selected for 'review\_creation\_Month\_Name' Column.
  - c. Add Column -> Date -> Day -> Name of Day is selected for 'review\_creation\_Day\_Name' Column.
  - d. Add Column -> Time -> Time Only is selected for 'review\_creation\_Time' Column.
7. In 'olist\_orders\_dataset' file, 'order\_purchase\_timestamp' column is used and column is splitted into 'Day of Week'.
8. 'order\_purchase\_timestamp' column is selected and the following action is performed.
  - a. Add Column -> Date -> Day -> Day of Week is selected for 'Day of Week'.
9. In 'product\_category\_name\_translation' file, the First row is both Columns are Changed into Header by the following action
  - a. Home -> Use First Row as Headers, so the First Column Header will be 'product\_category\_name' and Second Column Header will be 'product\_category\_name\_english'.
10. After Applying the changes, In Data Model, the models are connected by the common columns by two files.
11. All the Cross-filter connections are from both sides of the files.

## **Dashboard Creation**

### **Weekday Vs Weekend Payment Statistics.**

- This visualization is used to show the Percentage and Count of the Orders, Categorizing Weekdays/Weekends.
- There is no Column for 'Weekdays/Weekends' in the dataset. Therefore, a Column Is created in the dataset using Power Query Editor.
- In 'olist\_orders\_dataset', A conditional column named 'Weekdays/Weekends' is created using 'Day of week' Column.

- If 'Day of Week' equals 0, then 'Weekend'.
- Else If 'Day of Week' equals 6, then 'Weekend'.
- Else 'Weekday'.
- All the Changes are Applied and Closed by Home -> Close & Apply.
- In Report View, Pie Chart Visualization is used, and following data fields are dragged into Visualizations Pane.
  - olist\_orders\_dataset -> Weekdays/Weekends in 'Legends'.
  - olist\_orders\_dataset -> Count of order\_status in 'Values'.
- The Visualization is titled as 'Order Status in Weekdays/Weekends'.

## **Payment Types in Review Score-5.**

- This visualization is used to Display the count of Orders in Credit Card Payment and Review Score 5.
- In Report View, Stacked Column Chart Visualization is used, and following data fields are dragged into Visualizations Pane.
  - 'olist\_orders\_payments\_dataset' -> payment\_type in 'X-axis'.
  - 'olist\_orders\_payments\_dataset' -> Count of order\_id in 'Y-axis'.
  - 'olist\_order\_reviews\_dataset' -> review\_score in 'Legend'.
- 'payment\_type' and 'review\_score' is dragged into filter pane.
  - 'payment\_type' -> Basic Filtering is used and 'not\_defined' option is unselected.
  - 'review\_score' -> 'Basic Filtering is used and '5' option is only selected.
- The Visualization is titled as 'Payment Types in Review Score-5'.

## **Average price and payment values from customers of sao paulo city.**

- This visualization is used to Show the Average price and Payments from customers of Sao Paulo City.
- In Report View, Clustered Column Chart Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_customers\_dataset' -> customer\_city in 'X-axis'.
  - 'olist\_order\_items\_dataset' -> Average of price in 'Y-axis'.

- 'olist\_order\_payments\_dataset' -> Average of payment\_value in 'Y-axis'.
- 'customer\_city' is dragged into filter pane.
  - 'customer\_city' -> Basic Filtering is used and 'sao paulo' is only selected.
- The Visualization is titled as 'Average of Payment Value & Price by Sao Paulo'.

## **Relationship between shipping days Vs review scores.**

- This visualization is used to Display the relationship between Shipping Days of the orders and Their Review Scores.
- There is no column for count of 'Shipping\_days' in 'olist\_orders\_dataset'. So, the Column is created using a Custom Column Formula.
  - 'order\_delivered\_customer\_date - order\_purchase\_timestamp' is the Formula used to create 'Shipping\_Days' Column.
- In Report View, Scatter Chart Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_order\_reviews\_dataset' -> review\_score in 'X Axis'.
  - 'olist\_orders\_dataset' -> Average of Shipping Days in 'Y Axis'.
- Trend Line is added to visualize the general direction and strength of a relationship between the two data.

## **Average number of days taken for order\_delivered\_customer\_date for pet\_shop.**

- This visualization is used to display the Average no of Days taken for Orders in pet shop.
- In Report View, KPI Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_orders\_dataset' -> Average of Shipping Days in 'Value'.
  - 'olist\_products\_dataset' -> product\_category\_name in 'Trend Axis'.
- The Visualization is titled as 'Average no of days taken for orders for Pet Shop'.

## **Number of orders with Review Score-5 and Credit Card.**

- This visualization is used to display the number of orders with Review Score-5 in Credit Card Payments.
- In Report View, Card Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_order\_payments\_dataset' -> Count of order\_id in Fields.
- 'payment\_type' and 'review\_score' is dragged in filter pane.
  - 'payment\_type' -> Basic Filtering is used and 'credit\_card' is only selected.
  - 'review\_score' -> Basic Filtering is used and '5' is only selected.
- This Visualization is titled as 'No of Orders with Review Score-5 and Credit Card'.

## **Total Customers**

- This Visualization is used to Display the Total Count of Customers in Olist Store.
- In Report View, Card Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_customers\_dataset' -> Count of customer\_unique\_id in Fields.
- This Visualization is titled as 'Total Customers'.

## **Total Orders**

- This Visualization is used to Display the Total Count of Orders in Olist Store.
- In Report View, Card Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_orders\_dataset' -> Count of order\_status in Fields.
  - order\_status is Renamed to Total Orders.
- This Visualization is titled as 'Total Customers'.



## Total Sellers

- This Visualization is used to Display the Total Count of Sellers in Olist Store.
- In Report View, Card Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_order\_items\_dataset' -> Count of seller\_id in Fields.
- This Visualization is titled as 'Total Sellers'.

## Total Sales

- This Visualization is used to Display the Total Sales in Olist Store.
- In Report View, Card Visualization is used and following data fields are dragged into Visualization Pane.
  - 'olist\_order\_payments\_dataset' -> Count of payment\_value in Fields.
- This Visualization is titled as 'Total Sales'.

## Total Profit

- This Visualization is used to Display the Total Profit in Olist Store.
- In Report View, Q&A Visualization is used and in 'Ask a Question about your data' Section, 'total payment value – total price' is entered.
- 'Turn this Q&A result into a Standard Visual' Option changed the Visual to Card Visualization.
- In Visualization Pane, Go to Visual -> Callout Value and Change the Display Units to Millions.
- This Visualization is titled as 'Total Sellers'.

## Month Wise Total Price and Sales

- This visualization is used to Compare the Total Price and Sales in Monthly Wise.
- In Report View, Line and Stacked Column Chart is used and following data fields are dragged into Visualization Pane.
  - 'olist\_orders\_dataset' -> order\_delivered\_customer\_date -> Date Hierarchy -> Year, Month in X-axis.

- 'olist\_order\_payments\_dataset' -> Sum of payment\_value in Column Y-axis.
- 'olist\_order\_items\_dataset' -> Sum of price in Line Y-axis.
- The Visualization is titled as 'Month Wise Total Price and Sales.'

## **Order Purchase Year**

- This is a Filter used for filtering the Dashboard based on Year.
- In Report View, Slicer is used for Displaying Year as options to Filter.
  - 'olist\_orders\_dataset' -> order\_purchase\_timestamp -> Date Hierarchy -> Year.
- This Slicer is titled as 'Order Purchase Year'.