NOVA Information Management School, Universidade Nova de Lisboa
Data Science and Advanced Analytics with a major in Business Analytics
Business Intelligence, 2020-2021

Project Report

Business Intelligence Solution:

Olist Business

Ana Luísa Mestre, 20200599
Ana Marta Silva, m20200971
Débora A. Santos, m20200748
Mariana Domingues, m20201040
Pedro Henrique Medeiros, m20200742

INDEX

1 – INTRODUCTION

The world is experiencing a great technological and digital revolution that is changing the consumer shopping habits. Taking advantage of that, e-commerce is growing daily and now it is a fundamental strategy for any business to gain strength, visibility and to stay relevant in the market.

Consequently, the competition between sellers leads to a constant strive for improving their business models and decisions.

In this context, the Brazilian company Olist challenged our consulting group Inter BI to design a Self-Service BI solution to support its strategic, tactical and operational business decisions.

The present report intends to provide the details of our solution, such as the dimensional model and the perspectives of analysis in order to make sure that it matches the Olist's expectations.

In the following sections we will start by describing the Olist Business, followed by the construction of the Dimensional Model and the Power BI Report.

We are excited to take part of this challenge.

Link for the workspace of the group (that the Professors have access to): https://app.powerbi.com/groups/2715a87e-62ac-427f-9bb4-10f73c526d9c/list

2 – BUSINESS PRESENTATION

2.1 – OLIST'S BACKGROUND

Olist is a Brazilian start-up that operates in the e-commerce[1] segment, but it is not a true e-commerce. Instead, the company helps small retailers to sell in several marketplaces[1], such as Mercado Livre, Americanas, and Magazine Luiza. On the other hand, it concentrates the products from all sellers in a single store that is visible to the end consumer.

The company was founded in February 2015 by Tiago Dalvi, but its history goes back to times before its foundation.

Tiago started by creating a simple store that bridged the gap between artisans who wanted to sell their pieces, and the people who wanted to buy them, founding Solidarium. This store had as its main goal to find the path of the products that were being sought but that were not so visible since they were and are manufactured by small and medium companies. Solidarium started by having stores in malls and making this connection in a more traditional way than the current Olist. They had more than 15000 shopkeepers and 380000 products, and it was there that they realized that it was very costly, both in terms of effort

[1] E-commerce – online store where companies sell their own products.
Marketplace – platform where several companies and / or brands can register and sell their products. It can be said that a marketplace assumes itself as a virtual mall.

and money, to find a final consumer, as well as campaigning to promote the products – the costs often did not outweigh the gains.

However, a large contribution from one of the companies that most drives start-ups made it possible for the company to enter the digital world; hence, Olist was born. It was also not restricted to just artisans, opening doors to all kinds of products that can be commercialized. They also made it easier for small and medium-sized retailers to enter this new e-commerce market that was increasingly in use.

Currently, the company is headquartered in Curitiba, Paraná, and São Paulo. With 300 employees, more than 90 thousand registered shopkeepers, and 2 million unique consumers, Olist is already one of the largest virtual stores within the main marketplaces in the country. It is present on the websites of Mercado Livre, Walmart, B2W (Submarino, Americanas and Shoptime), Via Varejo (Casas Bahia, Ponto Frio and Extra.com), and Amazon, selling products from micro, small and medium retailers from various parts of the country and several segments. The business model is guaranteed to the company growth rates of up to ten times a year and bringing it closer to the objective of being the largest virtual store of the main marketplaces in operation in Brazil.

“Balancing quality and quantity is a challenge for any marketplace that wants to bring a large volume of tenants (to its website). What we managed to do was to automate this entry process, by making available (to the marketplace) a very high-quality catalog, with a very large volume of products already cured, moderated and approved”, Tiago Dalvi, CEO and founder of Olist.

The whole concept of Olist comes down to making the competition between the small shopkeepers, who are mostly offline sellers that found it difficult to create their own e-commerce, and the big sellers, fairer.

Olist’s shopkeepers pay a monthly fee that allows them to use the platform, to make their inventory available, and to receive a suggestion regarding the sale value of their products. When products are sold, Olist takes care of the product's path from the shopkeeper to the final consumer, thus receiving a certain percentage of the sale.

In 2017 Olist continued to grow, receiving around 2500 new products per day and with them 300/350 new shopkeepers per month. It was also in negotiations to integrate more marketplaces.

In the end of the years 2019 and 2020, Olist received two investments from the Japanese conglomerate SoftBank of 190 million Reais and 310 million Reais, respectively. Paulo Passoni, a partner at SoftBank, stated that the second investment shows confidence in Olist's work. Additionally, the investment took place in a year of atypical growth for the Brazilian e-commerce. In 2020, sales were expected to grow 30% compared to the previous year, according to the Brazilian Electronic Commerce Association (ABComm). From the 117 billion Reais that the e-commerce is expected to bill in 2020, about 45% will come from marketplaces, while three years ago, the share was 24%.

This growth in marketplaces benefited Olist, which will use the capital received to accelerate its growth, investing in technology, new products and in mergers and acquisitions.

2.2 – STATISTICAL DATA AND DESCRIPTIVE ASPECTS

Billing and Order Volume

- The Brazilian e-commerce registered 123 million orders and received revenues of R$ 53.2 billion in 2018.
- 58 million customers made at least one purchase over the internet in 2018.
- For the customers, the main advantages of buying online are cheaper prices than in physical stores (70.9%), ease of comparing prices and products (69.8%), and convenience (67.1%).
- In first semester of 2020 the Brazilian regions that recorded the greatest growth in e-commerce were the Northeast (107%), the North (93%) and the Midwest (45%) against the first semester of 2019. (Figs. 2.1 and 2.2)

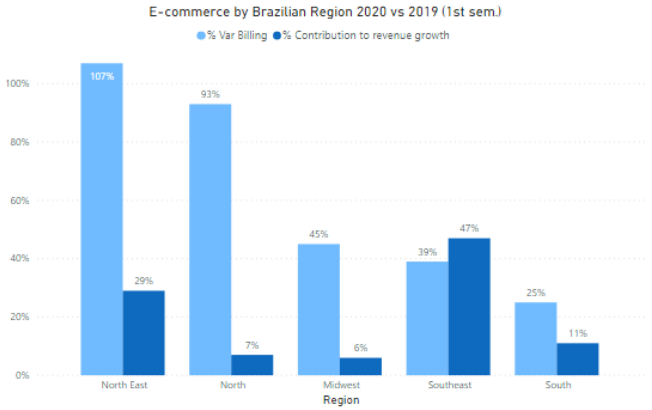

Fig. 2.1 – E-commerce Brazilian Region

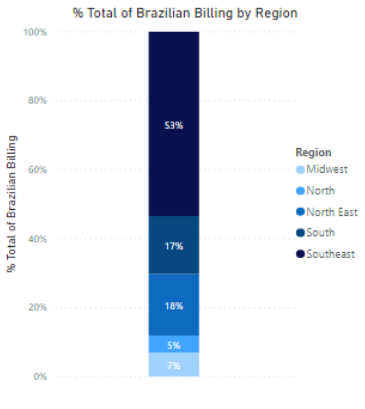

Fig. 2.2 - % Brazilian Billing by Region

Marketplaces

- The sales in marketplaces represented around 18.5% of total Brazilian e-commerce revenue in 2017.
- The main reasons that motivate sales in marketplaces are to increase sales (81.5%), to reach more customers (67.9), and to strengthen the brand (42%).
- 18.3% of the stores present in marketplaces sell on more than 5 websites of the same type, while 14.8% of them advertise on only one channel.
- The number of sellers on marketplaces increase 90,7% from September 2017 to September 2018. [10] (Figure 2.3)

Fig. 2.3 – Number of Sellers on Marketplaces

Best-Selling Categories in E-Commerce

- The categories that grew the most in the number of orders in e-commerce in 2018 were Perfumery / Cosmetics / Health (51%), IT (27%), Food and Beverages (23%), and Home and Decoration (16%).
- Some of the segments that registered the highest average ticket in 2018 were Electronics (R$ 1,000), Appliances (R$ 677), Computers (R$ 481), and Fashion (R$ 319).
- The average ticket for the consumer electronics category was R$558.2 in 2017, while the average for other segments was R$314.8.

Fig. 2.4 – Categories with the Highest Growth Demand Ticket

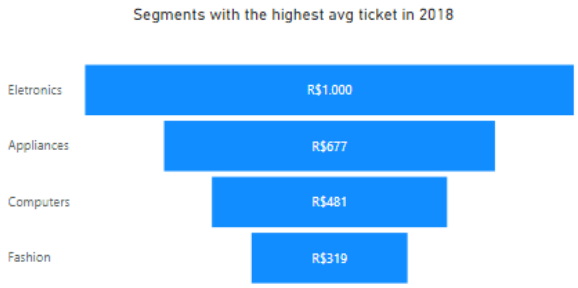

Fig. 2.5 – Segments with the Highest Average

Product Registration, Categorization, and Reviews

- 95% of the public considers it is very important to present in-depth information for making purchases over the internet.
- Optimization in product registration increases the conversion rate of ads by up to 30%.
- More than 70% of the goods evaluated by customers in Brazil have a score between 4.5 and 5.
- The chance of buying an item with 5 comments is 270% higher than a product without comments.

3 – DATA SOURCES

The data used in this project was made available by Olist on Kaggle [1]. Kaggle is a website with lots of datasets and discussions regarding data analysis.

The different files contain transactional data about the sales orders at Olist Store from September 2016 to October 2018. It has many details about orders, such as the customers and their locations, the sellers and their locations, the products and their categories, the reviews' comments and scores per order, price and freight value of each order item, and many date dimensions (order purchase time, delivery date estimated, effective delivery date, review date, etc.).

However, the company did not shared all the data, taking some information from the database, such as the customers' names, the sellers' names and the products' names. Thus, some features like customer_id, seller_id and product_id, are anonymized. And any text identifying stores and partners were replaced by the names of Game of Thrones great houses [1].

The nine files available are listed below:

- olist_customers_dataset.csv
- olist_geolocation_dataset.csv
- olist_order_items_dataset.csv
- olist_order_payments_dataset.csv
- olist_order_reviews_dataset.csv
- olist_orders_dataset.csv
- olist_products_dataset.csv
- olist_sellers_dataset.csv
- product_category_name_translation.csv

The figure 3.1[1] represents the schema showing how the files are connected.

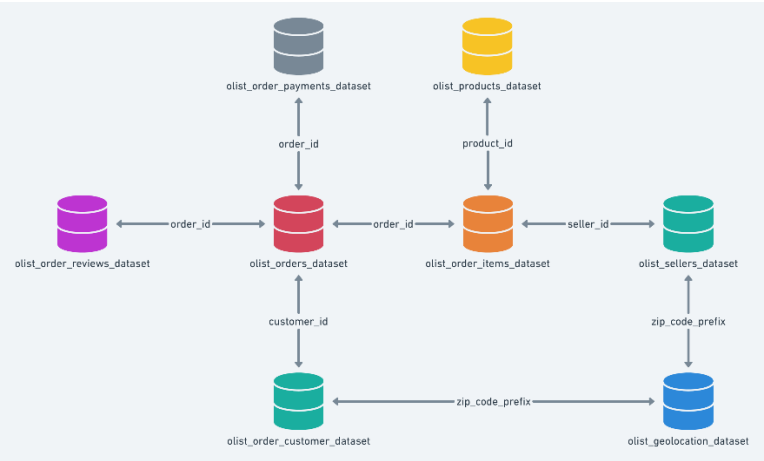

Fig. 3.1 – Tables' Schema

In the project, we decided not to use the olist_order_payments_dataset.csv and product_category_name_translation.csv.

We transformed to an excel document the following datasets: olist_customers_dataset.csv, olist_sellers_dataset.csv, olist_products_dataset.csv, and we inserted mannually the customers' names and sellers' names.

The data has sales information corresponding of 98666 orders, with 32951 different products from 3095 sellers to 98666 customers. The orders cancelled or unavailable were out of the scope of this work.

All features' descriptions are in the appendix.

4 – BUSINESS PROBLEM

In the first semester of 2020, the Brazilian e-commerce had a growth of 47% [2]. Circulation restrictions imposed due to the pandemic boosted e-commerce sales.

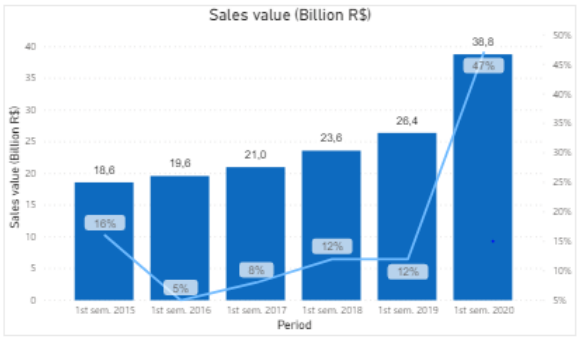

Fig. 4.1 – Brazillian E-Commerce Growth [3]

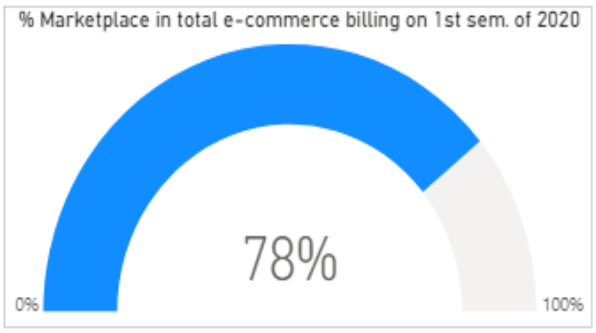

Fig. 4.2 - % Marketplace on the Total E-Commerce [3]

Olist also had a big growth in the past years.

Following this, the executive team would like to implement a Business Intelligence solution to better understand the business, which lead it to reach Inter BI to develop it.

The solution should be developed in Power BI and it would provide reports and statistics regarding three different perspectives: Sales Performance, Logistics Performance and Customer Satisfaction Analysis.

4.1 – SALES PERFORMANCE

In this first perspective, the report will show the sales performance from three dimensions:

- Product dimension
- Customer dimension
- Sellers dimension

For the product dimension, the report will provide the following information:

1) How is the distribution of the total sales amount per customer's city and per product category in each city?
2) What is the total quantity, total sales amount, and average price per product category?
3) What is the percentage of delivered and not delivered product(s)?
4) Provide the forecasting of the total sales' quantity for 6 months.
5) A comparison of number of products delivered in the current selected year versus the same period of the previous one.

Regarding the customer dimension, the report will provide answers to questions:

1) How are customers geographically distributed?
2) What are the number of orders and total sales amount related to the location of customers through time?
3) How are the number of orders and total sales values distributed over time?

4) Which regions were responsible for the largest sales amount?

For the sellers dimension, the report will have two pages:

- The first one is named Sellers' General. As the sellers are customers from the Olist perspective, we create this view with the objective for the company to visualize the performance of its sellers. The information provided by this page is related to the following questions:

    1) How many sellers are in the database?
    2) Who are the top sellers in terms of sales volumes and sales amount?
    3) What is the sales amount and quantity sold by Seller's Region?
    4) What is Year to Date sales amount by month and by year?
    5) What is the review average score of each seller and how many times they were reviewed with each score?

- The second one is named Sellers' Single. This one was built with the objective of the company make it available so that each seller can see your own performance. It is an individual seller report. The information provided by this page answers to the problems:

    1) What is the average score of the selected seller?
    2) How many orders has the selected seller?
    3) When was the first and the last sale date of the selected seller?
    4) How is the distribution of quantity sold and sales amount by customer's region of the selected seller?
    5) What is the sales amount and quantity sold by product category and by year of the selected seller?
    6) A comparison of the average review score of the selected seller with the overall average review score.
    7) A comparison of the total sales amount of the selected seller in the current selected year with the same period of the previous one.

4.2 – LOGISTICS PERFOMANCE

One of the main problems in the e-commerce is the shipment deadline. In the second quarter of 2020, the percentage of orders delivered out of the deadline in the Brazilian e-commerce increased to 15% [3] (Fig. 4.3). This is also a problem to Olist. In the Brazilian website "Reclame aqui", a platform for customers to make complaints about companies, Olist has more than 1800 complaints [4]. Going into more details, most of these complaints are about problems with the shipment.

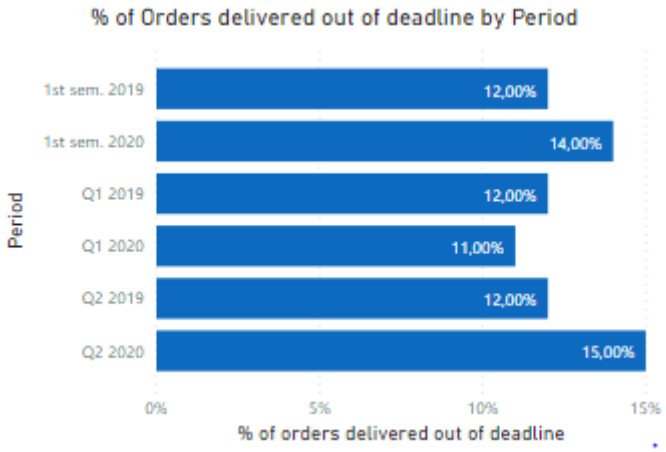

Fig. 4.3 – % of Orders Delivered Out of Deadline

Aware of this, Olist asked InterBI the construction of useful reports about logistics' performance, in order to have a solution for this.

Thus, the report will also respond to the following questions:

1) What is the total value of freight that the customers pay based on their location?
2) What is the average freight value for each category of products? And what is the total freight value paid by the customers for each category?
3) What s the percentage of orders hat were not delivery on time through time?
4) What is the amount of order that were not delivery on time? It was carrier's or seller's fault?

4.3 – CUSTOMER SATISFACTION ANALYSIS

Finally, Olist would like to have some insights regarding the customers' opinion regarding the service provided. Considering this, the report should present visualizations regarding customer satisfaction in different dimensions: sellers, products and logistics.

Finally, the report will present information about:

1) What is the impact through time of late deliveries on customers' review scores?
2) What are the top 10 product categories in terms of review score? And the worst?
3) What are the words that customers most apply in the reviews?
4) How did the overall satisfaction of the customer evolved compared to the same period of the previous year?

5 – DIMENSIONAL MODELLING

A Dimensional Model is designed to read, summarize, and analyze numeric information like values, balances, counts, weights, etc. in a data warehouse.
The Olist data warehouse was built with a dimensional data model, which was created following the Kimball methodology. The concept stated by this methodoly is that the dimensional model is comprised of dimension tables and a fact table.

The fact table represents the facts/events and it has the measures of the events that can be aggregated. In this project, the facts are the sales done by the sellers through Olist.
The dimension tables are responsible for the description of the facts. In this project, the model contains the following dimensions:

- Dim_Customer (Customer Dimension)
- Dim_Seller (Seller Dimension)
- Dim_Date (Date Dimension)
- Dim_Product (Product Dimension)
- Dim_Location (Location Dimension)

The dimension Location was denormalized and embedded on the Customers and Sellers' dimensions. In this way, the report is able to show visualizations using both the customers' location and the sellers' location.
Additionally, the model follows a star-schema structure with one-to-many relationships between each dimension table and the fact table. The four dimension tables are connected directly to the fact table. This style of dimension model allows for performance increase and easy deriving of insights. A representation of the model can be seen below:

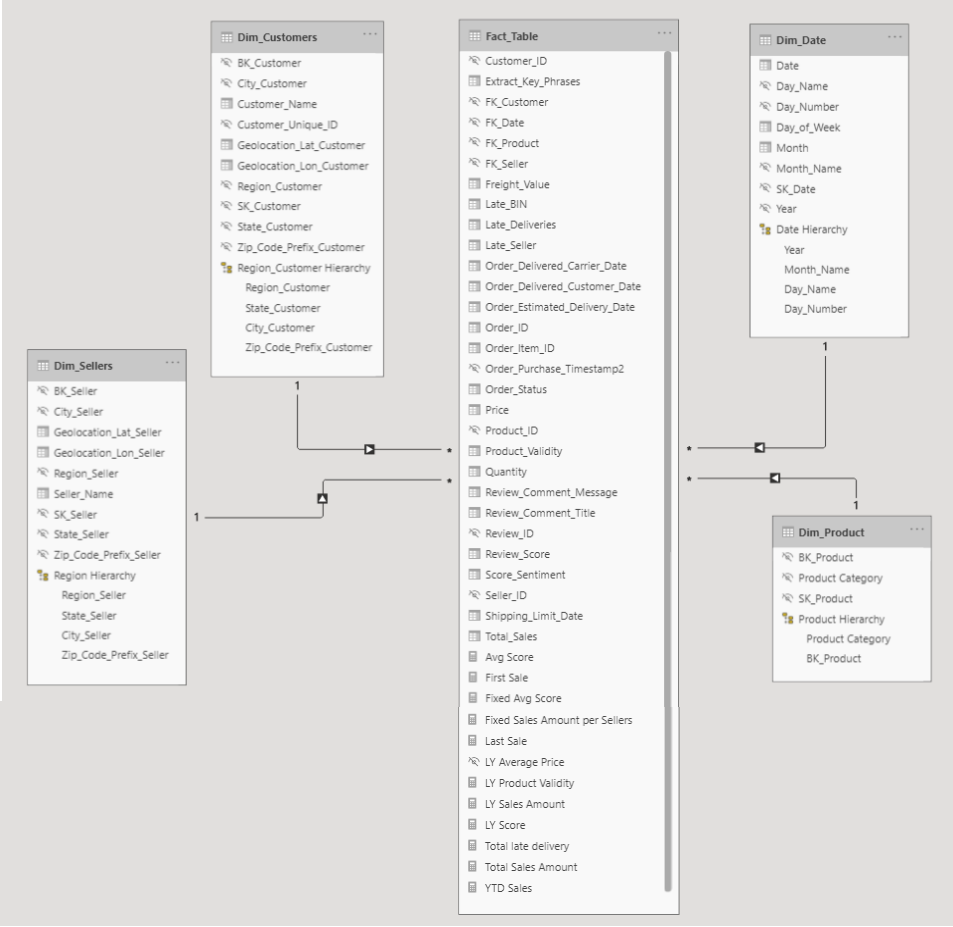

Fig. 5.1 – Dimensional Model

The dimension tables, as well as their attributes, metrics and granularity level, were chosen based on the different aspects of the business processes. Additionally, the necessity of

performing different analysis for the sales, logistics and marketing activities of the Olist business supports the chosen design.

Morevover, the model was designed having in mind the end user to provide reliable, easily accessible, and consistent information. The data model is adaptable, receptive to change, and delivers information in a timely manner for better analysis and decision making.

In this context, the dimension table **Dim_Date** provides information about the timing when the orders were placed in the plataform Olist. In other words, the business key used to build this dimension is the field Order_purchase_timestamp. We opted by using the level of granularity until the day of the purchase, since Olist only required that level. This dimension includes an hierarchy of 4 levels, which are the attributes Year, Month_Name, Day_Name and Day_Number. In summary, this dimension table contains:

1) SK_Date (surrogate key)
2) Date
3) Month
4) Day_of_Week
5) Hierarchy of 4 levels:
    a) Year
    b) Month_Name
    c) Day_Name (name of the day of the week)
    d) Day_Number

The dimension table **Dim_Product** provides information that allows to answer queries about the products sold and it includes 3 fields: a surrogate key named SK_Product, a business key named BK_Product, which consists of the unique identifier of each product, and Product_Category, both constituting a hierarchy of 2 levels. In summary, this dimension table contains:

1) SK_Product (surrogate key)
2) Hierarchy of 2 levels:
    a) BK_Product (business key reffering to the product id)
    b) Product_Category

The dimension table **Dim_Customer** provides information about the buyers in the platform Olist. It includes the index of the table (surrogate key), the business key and more 7 attributes. Specifically, it contains:

1) SK_Customer (surrogate key)
2) BK_Customer (business key referring to the customer id)
3) Customer_Name
4) Geolocation_Lat (latitude of the location of the customer)
5) Geolocation_Long (longitude of the location of the customer)
6) Hierarchy of 4 levels:
    a) Customer_Region
    b) Customer_City
    c) Customer_State
    d) Zip_Code_Prefix

The dimension table **Dim_Seller** contains descriptive information regarding the sellers. It includes the index of the table (surrogate key), the business key and more 7 attributes. Specifically, it contains:

1. SK_Seller (surrogate key)
2. BK_Seller (business key referring to the id of the seller)
3. Seller_Name
4. Geolocation_Lat
5. Geolocation_Long
6. Hierarchy of 4 levels:
   a) Seller_Region
   b) Seller_City
   c) Seller_State
   d) Zip_Code_Prefix

These last two dimensions (Dim_Customer and Dim_Seller) have embedded the Dim_Location dimension. Thus, some atributes related to the location of both are represented by an hierarchy of 4 levels: Region > State > City > Zip_Code_Prefix.

Finally, the fact table (Fact_Table) consists of the orders transactional data with measures of each event. Its rows include the main order attributes and the combination of the foreign keys of the dimension tables Dim_Customer, Dim_Seller, Dim_Date, and Dim_Product. Additionally, we also created some columns in order to build the Report. In detail, the fact table contains the following fields:

Original Fields	Columns created by merging tables	Collumns created Customm Column on Power Query	Measures created using DAX	Calculated Collumns created using DAX	Collumns created using the option AI Insights
Order Delivered Carrier date	FK_Product	Quantity	Avg Score	Total Sales	Extract Key phrases
Order Delivered Customer date	FK_Customer		First Sale	Product Validity	
Order Estimated Delivery date	FK_Seller		Fixed Average Score	Late Deliveries	
Order_id	FK_Date		Fixed Sales Amount	Late Seller	
Order Status			Last Sale	Late BIN	
Price			LY Product Validity		
Review_comment_message			LY Sales Amount		
Review_comment_title			LY Score		
Review_score			Total_Late_Delivery		
Shipping_limit_date			Total Sales Amount		
			YTD Sales		

Fig. 5.2 – Fields of the Fact Table

To sum up, the final dimension model has 4 dimension tables (Dim_Customer, Dim_Seller, Dim_Date and Dim_Product) with the Dim_Location dimension denormalized inside the dimensions Dim_Customer and Dim_Seller. It also has the fact table cointaining the transactional data and the measures. As for the hierarchies, the model has 3:

1) Product hierarchy in the dimension table Dim_Product with 2 levels: Product_Category > BK_Product
2) Date hierarchy in the dimension table Dim_Date with 4 levels: Year > Month_Name > Day_Name > Day_Number
3) Location hierarchy in the dimension tables Dim_Customer and Dim_Seller with 4 levels: Customer/Seller_Region > Customer/Seller_State > Customer/Seller_City > Customer/Seller_Zip_Code_Prefix.

6 – DATA INTEGRATION, TRANSFORMATION AND MODELLING

In this section, it will be presented how the model was built in the Power BI. This process had three main steps: Data Integration, Data Pre-Processing and Modelling.

In the first step, and as already mentioned in the section 3, we integrated seven files with two different types of source:

1) 3 excel files:
   a) olist_customers_dataset.xlsx
   b) olist_sellers_dataset.xlsx
   c) olist_products_dataset.xlsx
2) 4 csv files:
   a) olist_geolocation_dataset.csv
   b) olist_order_items_dataset.csv
   c) olist_order_reviews_dataset.csv
   d) olist_orders_dataset.csv

The tool used to make this integration was the Power Query Editor. After this, the next step was to pre-process the data.

The pre processing phase was also made using the Power Query Editor. The first two main transformations on this phase were regarding the merge of some tables. The first merge was done in order to build the fact table. For that, the tables olist_order_items_dataset, olist_order_reviews_dataset, and olist_orders_dataset were merged in a new table, posteriorly named Fact_Table. The key used to make this merge was the Order_id. In the second merge, the columns latitude and longitude of the olist_geolocation_dataset were included in the costumers and sellers' datasets. In this case, the key used for the merge was the Zip_Code_Prefix. After this, the load of the four tables (geolocation, order_itens, orders, reviews) was hidden so that they would not appear in the model.

The next main transformation was to create the dimension table Dim_Date. The business key used to create this table was the Order_Purchase_Timestamp. The column Date was created using the range from 01/09/2016 to 05/09/2018. On this moment, we stated the granularity of the dimension date and we decided to keep the lower level as the date of the sale. Starting from the date column, we created six more columns: Year, Month, Month_Name, Day_of_Week, Day_Number and Day_Name.

Besides this, other tranformations were also applied, such as: rename the queries, remove some columns, change the data type of the columns, correct some wrong values, remove duplicate rows in the tables, reorder columns and/or rename some columns.

The next tranformations are more related with the modeling, although they were still done in Power Query Editor.

Every dimension table should have one column designated as the Primary Key. This key can consist of one or more fields that uniquely identify a row in a table, and cannot be null (blank). In this project, we already have in each dimension a Business Key, however, we still need to create a Surrogate Key (SK). This key, on the other hand, is the column that connects the dimensional tables with the fact table. It has to be an integer continuos starting from 0.

Therefore, the next step was to create the four SKs in each dimension table: SK_Customer, SK_Seller, SK_Date and SK_Product.

After we creat the SK, it was time to create the Foreign Keys (FK) in the fact table, which always contains foreign keys for each of its associated dimensions. For this, we merged the four dimension tables with the fact table by expanding the columns corresponding to the SKs of the dimension tables into the fact one. Then, we renamed the four new columns to: FK_Customer, FK_Seller, FK_Date and FK_Product.

The main transformations done in Power Query BI were explained and can also be checked in the table 10.4 in the Appendix section.

Now, in the Report View, we created the three hierarchies: Product hierarchy, in the Dim_Product, Date hierarchy, in the Dim_Date, and Location hierarchy, in the Dim_Seller and Dim_Customer.

Additionally, in the Model View, we hided the columns that we did not want to see in the report. Here, we also checked the relationships between the tables and the columns used to link them to verify if the model was correctly built and if everything was working fine.

The last step was to create the following column:

- Quantity: created in the Power Query using the option 'Add a Custom Column'. Each row received the value equal to 1.

The other columns and measures created are going to be explained in the next section where the details of each page of the report will be presented.

With everything working perfectly well, it was time to make the analysis and the report.

7 – POWER BI REPORT

Regarding the Power BI Report, we organized the three perpectives of analysis into six different pages – products, customers, sellers (general and single), logistics, and reviews.

7.1 – TECHNICAL ASPECTS

To be able to build the pages of the report, some technical work behind them had to be done.

Thus, in the Power BI Desktop, and using the DAX functions, we started by creating a calculated column inside the Fact Table named Total Sales which, as already mentioned above, is equal to the multiplication of the quantity per product by the sum of the unit price and freight value for each row (Quantity * (Price + Freight)). Then, from this column, with the SUMX function, we created a new measure named Total Sales Amount.

Starting by the **Products' page**, the column Total Sales was used in the matrix to visualize the sales amount per product category and the measure Total Sales Amount was used in the map to plot the amount of the overall sales for all the products in each city. The pie chart and the KPI implementation required the creation of a binary variable that we created using the option 'New Column' in the Data View. This column was named Product Validity, and it has value 1 if a product was delivered and 0 if not. Besides this, it was created from the already

existed order_status column and with the IF function. From this column, we also created one DAX measure – LY Product Validity – with the functions CALCULATE, COUNT, and SAMEPERIODLASTYEAR, posteriorly used in the KPI.

To diversify the analysis and present more information to the user, when filling the fields to build some of the visualizations, instead of creating new measures, we used more practical tools of the Power BI. For example, in the matrix, we selected the average option to get the average price per product category. Also, both in the card and KPI, we selected the count option of the BK_Product and Product Validity to get the total number of different products and products that were delivered. In the end, we filtered the graphics by eliminating the blank rows.

In this first page of the report, there is a map with the distribution of the total sales amount in the geographical space, followed by a tooltip with the top product categories in terms of total sales amount in each city; a matrix with the total quantity sold, the total sales amount, and the average price per product category, providing to the user the possibility of sorting the values of the columns by ascending or descending order; a pie chart with the percentages of products that were and were not delivered, as well as a card with the number of different products that exist of one or all the categories in a certain period of time; a line chart showing the forecast of the quantities for the next 6 months and, finally, a KPI, which compares the total number of delivered products in the current selected year with the previous one.

On the **Customers' page**, we have four main visualizations: two combined graphs, a pie chart and a map. The first graph, composed of line and clustered columns, shows the total number of orders placed by customers and their respective total values, distributed according to the year and location of the customers. For its construction, the Region_Customer Hierarchy was used in the shared axis field, the Year column in the column series, the sum of the Total Sales column for the column values, and the distinct count in the Order_id column for the row values.

The second graph, also composed of line and clustered columns, shows the quantities of orders placed by customers and their respective total values, distributed over time. For its creation, the Date Hierarchy was used, the sum of the Total Sales column for the column values, and the distinct count of the Order_id column for the row values.

The pie chart shows the values of the total quantities of orders placed in each of the five regions of Brazil, as well as the respective percentages of the quantities of customers by region. For its construction, the Region_Customer column was used for the legend, the sum of the Total Sales column for the values, and, as a Tooltip, the distinct count of the SK_Customer column was used.

The map visualization displays the customers' locations across the Brazilian territory. For the construction of this graph, the State_Customer column was used as the legend, the average of the Geolocation_Lat column as latitude, the average of the Geolocation_Long column as logintude, and the distinct count of the SK_Customer column as the size.

In the **Sellers' page**, we create two report pages.

The first one is named Sellers General and its objective is to visualize the overall performance of the Olist' sellers.
The first visualization of this page is a card with the count of BK_Seller, so that the company can know how many sellers are in the data base.

The next two visualizations show the top sellers. In terms of sales volumes, the top sellers are shown with a funnel visualization. In this plot, we also have a tooltip with the number of orders of each seller. Additionally, the top sellers in terms of sales amount are presented in a treemap visualization.

A pie chart was used to represent the distribution of sales amount and volumes through Brazilian regions. Here, we used the sellers' regions, so that the company can see where the sellers are concentrated.

On the next visualization, YTD Sales amount by month is represented by a line plot. To produce this visualization, we created a new DAX Time Intelligence Measure named YTD Sales. This was created using the option 'Create New Measure' in the report view. It was inserted into the fact table. The function used to calculate it is called TOTALYTD.

Finally, this page has a table containing details of the scores of each seller. In this table, for each seller is presented the average score and how many times the seller was reviewed. We also have the number of reviews for each score for each seller. To build this table, we created a new measure named Avg Score by using the option 'Create New Measure' in the report view. It was inserted into the fact table. The function used to calculate it is called AVERAGE.

This page also has a filter by year, so that all the visualizations can be seen considering all years or year by year.

The second page is named Sellers Single, whose objective is to, if Olist wants, make it available for each seller so that he can see his own performance.

The first thing to do in this page it is to select a seller. After that, the report will present only information regarding the selected seller.

The left side of the page presents some information about the seller in a multi row card. Here it is possible to visualize when was the first and the last sale date of the seller. To be possible to present this information, two measures were created using the option 'Create New Measure' in the report view. The measures First Sale and Last Sale were inserted into the fact table. Also, in the multirow card, the number of orders and the average score of the seller are presented. The average score was calculated using the same metric created on the sellers' general view.

The next visualization is a bar plot which shows the total sales amount and quantity sold by product category by year.
Additionally, the pie chart presents the distribution of the sales amount and quantity by region of customer, providing the seller the possibility of seeing where his customers are concentrated.

Finally, on the right side of the page, two KPI's are shown. The first one is a comparison of the average score of the selected seller with the overall score. For this KPI, we created another column using the option 'Create New Measure' in the Report View with a combination of the functions CALCULATE AND ALL, in order to determine the average score per seller. This measure was then named Fixed Avg Score. The second KPI is a sales amount comparison. It compares the sales amount of the selected period with the same period of the previous year. The measure Same Period of Last Year was created using the option 'Create New Measure' in the report view. It was inserted into the fact table. The functions used to calculate it are called SAMEPERIODLASTYEAR and CALCULATE.

For the **Logistics' page**, three columns were created using the option 'New Column' in the Data View: Late Deliveries, Late_BIN and Late Seller.

The first one indicates whether the order has arrived after the customer is due, the second one is the binary value of the first, which presents the number 1 when the delivery was delayed and 0 when it was not, while the third column indicates whether the order was delivered to the carrier after shipping limit time, that is, whether the seller delivered or not to the carrier after the announced and stipulated date. These columns were important for us to be able to make the visualizations we wanted. Regarding measures, we created a Total Late Delivery that would sum the quantity of not delivered orders in time and then divide it by the total of orders considered.

On the page of the report regarding logistics, we can see a map with the total value of freight paid by customers depending on their region, as well as a tooltip in which we can see the total values of freight paid in that state by the customers, depending on the product supplier. In the barplot on the right, we can see the average freight value by product type, that is, its category. In the tooltip associated with it, we were able to have a perception of the total amount paid by all customers when they bought this type of product.

Below, we have two cards relating to the quantities of orders that were or were not delivered late and the quantities of orders that were or were not delivered to the carrier late, two pie-charts indicating the percentage of orders that were delivered late to customers and also the percentage of orders that were delivered late to carriers, that is, the seller was late in delivering the order to the person who would be transporting it. Besides this, we can see if we select the part of the percentage of orders delivered after the deadline, which of these were already late in delivery to the carrier or, even, on the contrary, select which deliveries were delivered to the carrier in time and arrived late to the customer, thus suggesting that the carrier was at fault in these cases.

Finally, in the lower-left corner, we can see the percentage of orders delivered after what was stipulated in the various hierarchical levels of the date, that is, we can see by year as by month, and even months for a year. When selecting the model to see per year, we were able to get a sense of the development of the percentage of orders overdue over the 3 years, and so we can see that it is useful as a Key Performance Indicator. For this visual we used the measure created to this perspective, the Total late delivery.

In the **Reviews' page**, a DAX Time Intelligence Measure, named LY Score, was created in order to calculate the average score corresponding to the previous year to the one in analysis, for it were used the fuctions CALCULATE and SAMEPERIODLASTYEAR .

This measure and the Avg Score, already created before, allowed to create 6 graphs: 2 clustered bar charts, each representing the 10 top and worst product categories by review score. 1 clustered column chart in order to asses the impact of late and on time deliveries on the review scores over the years, 1 KPI to show the evolution of the average review score compared to the previous year, 1 card to show the total number of reviews according to the filters applied and one advanced visual, a wordcloud to show the most frequent words in the reviews, which was obtained using Machine learning competencies of Power BI (AI insights), particularly Text Analytics in order to extract key phrases.

7.2– DATA INSIGHTS

As already mentioned above, the first page of the report contains the **Products' analysis** with a set of different interactive visualizations from where it is possible to extract many useful insights.

Starting by the map, we notice that the dispersion of the total sales amount was much higher in 2016 comparing to the following two years, 2017 and 2018, where they were more concentrated.

In the matrix, the data tell us that, for example, the company sold a total of 111047 products in the 3 years considered with a total sales amount of around R$ 15 million and an average price per product category of R$ 121. Additionally, 2016 was the year with the highest average price per product.

In the pie chart, considering all the three years, 5,13% of the products of the "fraldas_higiene" category was not delivered. Also, considering all the categories, 2016 was the year with the highest percentage of products that were not delivered (14,4%).

By observing the forecast graph, and considering all the categories, we can see that the quantities of the products will not vary much in the future.

Finally, the KPI tells us that the number of products delivered in 2017 exceeded the expectations in around 13473%, however, this is probably because the previous year (2016) is the one that we have less data.

Regarding the interactive visualizations of the **Customers' page**, of all the 99440 different consumer records, spread across the 5 Brazilian regions, the Southeast has the largest number of consumers, comprising 68221 records, which corresponds to 64.73% of the total. The second region with the largest number of consumers in the South with 14133 (14.53%), followed by the Northeast with 9364 (11.86%), the Midwest with 5600 (6.29%), and the North with only 1845 (2.59%).

Following the same order, and considering the records of all years (2016, 2017, and 2018) found in the database, the Southeast region was the one that generated the highest total sales value, with 10.15 million Reais. Then the South, with 2.28 million, the Northeast with 1.87 million, the Midwest with 98494 thousand, and the North with R$ 40864.

The year 2018 was the one with the highest purchase record, counting 53530 orders and generating 8.59 million Reais. In 2017, 44380 orders were placed, which totaled R$ 7.09 million. 2016 presented only 293 orders, generating R$ 51.65 thousand.

Considering the orders placed in all years, the month with the highest number of purchases was August, with a total of 10745 orders, which generated R$ 1.671 million. With 10513, the month of May was the one with the second-highest number of orders, being, however, the month that provided the highest monetary value, with R$ 1.735 million.

Regarding the days of the month, the 24th was the day with the highest number of purchases (3850), generating a total of R$ 600,095.11. The 31st was the one with the lowest number of purchases, 1664, 114.90% less than the 24th, and providing R$ 279,249.85.

Regarding the days of the week, Monday presented the largest number of orders, representing 16.41% of the total, and generating R$ 2,600,533.82. Saturday was the day with the lowest number of sales, with R$ 1,752,075.49, 48.43% less than Monday.

The **Sellers' General analysis** shows that the top 10 sellers in terms of volumes are different of the top 10 sellers in terms of amount. The 10th seller in terms of volumes, sells 57,8% of volumes when compared with the top one seller. In terms of sales values, the 10th seller has a sale amount of 64% when compared with the top one.

More than 78% of sellers are concentrated on southeast region, followed by south region with approximately 15% of the total sales amount.

Regarding the score analysis, the most reviewed sellers are the same with the highest sales volumes. It shows that the customers are really answering the survey about satisfaction. In this way, Olist can try to improve the average satisfaction of the customers by analyzing the quality of service of each seller and try to help them to improve that.

The next page, **Sellers' Single analysis**, tells us that the customers are also concentrated on Southeast region.

The average score compared with other sellers is a good benchmark to help each seller to improve their business. Also, the comparison of the sales with the previous period can help the sellers to better understand their performance.

Choosing a seller to analyze, the most sold category is 'cama mesa e banho', followed by 'casa e conforto', which means that this seller sells products for houses and families. A decrease of 10% of sales amount against last year is observed, but the score is aligned with the overall one.

In the **Logistics' page**, starting by the map, the data says that the state with the highest value collected by freight is São Paulo, however, this is also because this is the state with the most orders. We can also see, for example, that in the state of Alagoas, the seller that generated the most value in terms of freight was Leon Ltda (information taken from the tooltip) with R $ 419.26.

The barplot tells us that, considering all the years and all the locations of our customers, the category with the highest average value of freight paid is 'pcs' with an average value greater than R $ 40 and a total value of R $ 9.67K. However, when we select the year 2017 we see that the category with the lowest average value in freight is 'artes_e_artesanato' with a little more than R $ 10 and a total value of R $ 23.33.

In the pie-charts, we see that, considering all the years and all the locations of the customers, we have a percentage of about 93% of orders delivered on time and that 94% of the orders were delivered within the agreed time to the carrier. If we click on the situations in which the order was delivered on time to the seller, we see that about 5% of them arrived late to the customer, corresponding to 5562 orders.

In the bar graph in the lower-left corner, we can see an increase in the percentage of orders delivered late, which is not a good sign.

The **Reviews' page** provides insights about the impact throughout the years of late deliveries on customer satisfaction. It also allows to analyse which are the top 10 product categories with the highest and lowest review scores and which are the most used words in reviews per category, type of delivery and year. Besides this, it provides information about the evolution of customer satisfaction compared with the previous year and how many reviews were provided by the customers in total, per year and category.

More particularly, the data tells us that, on average, per year, orders with late deliveries produce a much lower review score (approximately 2) compared to the orders with on time deliveries (around 4). This behaviour is similar along the years.

Additionally, while utilizing the AI Insight competencies of Power BI, specifically the text analytics one, we were able to identity the words most used in the reviews. We verified that 57.3% of the reviews provided does not include a comment, just a score. Also, the most used words are 'produto' and 'prazo', this is valid across the years too.

Moreover, we concluded that the total number of reviews in the database of Olist is 112.65K.

370 of them belong to the year of 2016 and the average score is of 3.51. In 2017, the number of reviews was equal to 50.86K and the average score was 4.03. Finally, 2018 accounted for 61.42K reviews with an average score of 4.01. In other words, the overall score in 2018 increased 0.13% compared with the year of 2017. Besides, the satisfaction of the customer in 2017 increased 0.84%.

Regarding the 10 product categories with highest review scores, they are in descending order: cds_dvds_musicais with a score of 4.64, fashion_roupa_infanto_juvenil with a score of 4.50, livros_interesse_geral with a score of 4.44, livros importados with a score of 4.40, construção_ferramentas_ferramentas with a score of 4.36, livros técnicos with a score of 4.33, malas_acessorios with a score of 4.31, portateis_casa_forno_e_cafe with a score of 4.30, alimentos_bebidas with a score of 4.30, and fashion_esporte with a score of 4.23.

In 2018, the category of cds_dvds_musicais had an average score of 5, which signified an increased of 25% compared with 2017. All the deliveries of this category were delivered on time.

On the other hand, the categories with the lowest score in ascending order are: seguros_e_servicos with a score of 2.50, fraldas_higiene with a score of 3.26, portáteis_cozinha_e_preparadores_de_alimentos with a score of 3.27, pc_gamer with a score of 3.33, casa_conforto_2 with a score of 3.37, moves_escritorio with a score of 3.48, roupa_fashion_masculina with a score of 3.62, telefonia_fixa with a score of 3.67,

fashion_roupa_feminina with a score of 3.75 and, lastly, artigos_de_festas with a score of 3.77.

The two lowest deliveries correspond to the orders that were delivered on time and the words most used were 'produto' which means that there is a problem of quality with the products inside these categories.

The place of this categories in the rankings was similar across the years.

From these insights we can conclude that the time of delivery and the quality of the product are very significant points for the clients of Olist, therefore they should prioritize partnerships with suppliers who have an on time delivery track record and use the reviews scores to identify categories where the quality of the products is poor and, consequently, change the suppliers. By doing this, Olist will assure that the products sold and the suppliers existent in their platform match the high quality for which they want to be recognized.

8 – CRITICAL ASSESMENT

In this project, it was possible to acquire a deep knowledge and understanding of the Power BI, as well as the relationships between the data and how to build a dimensional modeling.

Using the techniques presented by Kimball and Star-Schema, the dimensional modeling produced, including the dimensional tables, the fact table and its measures, attends all the business perspectives and the business needs. The dimensional model built allowed us to create informative, interactive, and easy-to-interpret reports that help to tell the narrative of the data. In addition, we developed the ability to discern which view is most suitable for demonstrating the analyzes performed.

The reports can be easily updated in order to keep up with the dynamism of the business. Moreover, Power BI offered several tools among which we highlight machine learning tools, AI insights, as well as the integration in the Power BI Online, which increases the accessibility of reports by third parties and in multiple devices.

On the other hand, the lack of descriptive data, such as customers and sellers' demographics (gender, date of birth, profession, marital status) prevented the customer segmentation analysis, for example. Besides this, the fact that some of the data was hidden by Olist also ended up limiting the analysis. Additionally, it would be interesting to have more information on who the carrier is, allowing the analysis of deliveries by carrier, understanding those that most impact delays and, consequently, customer satisfaction. However, we understood that the confidentiality of the information had to be preserved.

For future work, some analysis could be added to the report, such as understanding which would be the most frequent routes, the most expensive ones and if they are related to the weight of the products. Additionally, we could have detailed the analysis at time level to consider which hours have more traffic on the website to direct the promotion of the platform at more effective times. Another possible analysis would have been to understand the relationships between the products (substitutes, complementary) by integrating the code built in Python into Power BI and, in that way, perform a market basket analysis.

9 – CONCLUSION

The main objective of this project was to develop a Business Intelligence solution to Olist.

As a first step, Inter BI presented a draft of the dimensional modelling and which perspectives of analysis would be under the scope of the solution.

As a second step, Inter BI built the dimensional modelling, created the report pages and the dashboard in Power BI with the objective of attending the business needs of Olist.

The construction of the model required some data integration, data treatment and feature enginering to estabilish the relationship between the tables. On top of this, columns and measures were created to enable the construction of the report pages.

The final dimensional model achieved the requirements suggested of having 5 dimension tables from which location is embedded inside the customer and seller, 1 fact table and 3 hierarchies with an average depth of 3 levels. Also the requirements of creation of DAX measures for each perpective, calculated columns and DAX Time Intelligence measures were achieved.

The report pages were developed having in mind the business needs that we would like to address and availability of the data. Following the correct visualization were chosen to depict the analysis performed. To facilitate interpretation, we opted for a coherent design on all pages, we used several storytelling tools and several machine learning and predictive analytics techniques.

Finally, the BI solution was uploaded in Power BI online environment, so it can be accessed with a link . It also allowed us to create a dashboard with the most useful and easily to interpret visuals for Olist. Thus, the user of the solution can acess the report and the dashboard anywhere and any time.

This final solution enables the end users to effectively conduct analysis from three different perspectives (sales, logistics and customer satisfaction) in order to help decision-makers to make betterand fast informed decisions.

We hope Olist is satisfied with Inter BI work and we continue working together.

10 – APPENDIX

10.1 - Features' Descriptions Table

BK_Seller	Seller unique identifier
Zip_Code_Prefix_ Seller	The first 5 numbers of seller's zip code
Seller_Name	Name of the seller
City_ Seller	Seller city name
State_ Seller	Seller state abbreviation
Region_ Seller	Seller Brazilian Region Name
BK_Product	Product unique identifier
Product_Category_Name	Category name of the product

Order_ID	Order unique identifier
Order_Item_ID	Number itens in the invoice (sequential)
Shipping_Limit_Date	Deadline of the seller to deliver the order to the logistic partner
Price	Item price
Freight_Value	Item freight value. (if an order has more than one item, the total freight is splitted betwwen itens)
Order_Purchase_Timestamp	Shows the timestamp which the purchase was made
Order_Delivered_Carrier_Date	Show the timestamp which the seller handled the order to the logistic partner
Order_Delivered_Customer_Date	Shows the date that the order was delivered to the customer
Order_Estimated_Delivery_Date	Shows the estimated delivery date that was informed to the customer at the purchase moment
Order_Approved_At	Show the timestamp when the payment was approved
Review_ID	Review unique identifier
Review_Score	Score given by the customer on a satisfaction survey (range 1 to 5)
Review_Comment_Title	Comment title of the review given by customer
Review_Comment_Message	Comment of the review given by customer
BK_Customer	Key to identify the customer. Each order has a unique customer_id.
Zip_Code_Prefix_Customer	The first 5 numbers of customer's zip code
Customer_Name	Name of customer
City_Customer	Customer city name
State_Customer	Customer state abbreviation
Region_Customer	Customer Brazilian Region Name

10.2 – Names and Formulas of the Measures Created

Avg Score	Avg Score = ROUND(AVERAGE(Fact_Table[review_score]),0)
First Sale	First Sale = MIN(Fact_Table[order_purchase_timestamp2])
Fixed Avg Score	Fixed Avg Score = CALCULATE([Avg Score],ALL(Dim_Sellers[Seller_Name]))
Fixed Saled Amount per Sellers	Fixed Sales Amount per Sellers = CALCULATE(Fact_Table[Total Sales Amount],ALL(Dim_Sellers[Seller_Name]))
Last Sale	Last Sale = MAX(Fact_Table[order_purchase_timestamp2])

LY Product Validity	LY Product Validity = CALCULATE(COUNT(Fact_Table[Product_Validity]), SAMEPERIODLASTYEAR(Dim_Date[Date]))
LY Sales Amount	LY Sales Amount = CALCULATE(Fact_Table[Total Sales Amount], SAMEPERIODLASTYEAR(Dim_Date[Date]))
LY Score	LY Score = CALCULATE(Fact_Table[Avg Score], SAMEPERIODLASTYEAR(Dim_Date[Date]))
Total Late Delivery	Total late delivery = DIVIDE(SUM(Fact_Table[Late_BIN]), COUNT(Fact_Table[Late_BIN]))
Total Sales Amount	Total Sales Amount = SUMX(Fact_Table, Fact_Table[Total Sales])
YTD Sales	YTD Sales = TOTALYTD(SUM(Fact_Table[Total Sales]), Dim_Date[Date])

10.3 – Names and Formulas of the Calculated Columns Created

Late_Deliveries	Late Deliveries = IF(Fact_Table[order_estimated_delivery_date].[Date]<Fact_Table[order_delivered_customer_date].[Date], "Late", "On Time")
Late_Seller	Late Seller = IF(Fact_Table[shipping_limit_date].[Date]<Fact_Table[order_delivered_carrier_date].[Date],"Late Seller", "On Time Seller")
Late_BIN	Late_BIN = IF(Fact_Table[order_estimated_delivery_date].[Date]<Fact_Table[order_delivered_customer_date].[Date], 1, 0)
Product_Validity	Product_Validity = IF(Fact_Table[order_status]="delivered", "delivered", "not delivered")
Total_Sales	Total Sales = (Fact_Table[price] + Fact_Table[freight_value])*Fact_Table[Quantity]

10.4 – Summary with all the information of the dimension tables

Dimension	Transformation	Column/Table	Power BI Tools	Comment
Fact Table	Marge queries/Expanded columns	Olist_orders_datase Olist_order_reviews_dataset Olist_order_itens_dataset	Power Query Editor	Create the fact table merging the 3 datasets mentioned, using the column order_id as a key. Starting by the order_itens_dataset because it has more details regarding orders transactions
	Add column	Quantity	Power Query Editor	Insert custom column, Quantity
	Removed Columns		Power Query Editor	Removed all columns unnecessary to our model
	Change type of the columns		Power Query Editor	Change type of all columns
	Marge queries/Expanded columns	Create the FK Columns (Date, Customer, Sellers and Product	Power Query Editor	Create the Foreign Columns of Dim Customers, Dim Sellers, Dim Date and Dim Product using the Business Key for the merge: Customer id, Seller id, Time stamp and product id respectively
	Rename the columns add above	FK_Customer, FK_Seller, FK_Date, FK_Product	Power Query Editor	Rename SK_Customer, SK_Seller,SK_Date, SK_Product to FK_Customer, FK_Seller, FK_Date, FK_Product
Product	Add Index Columns	SK_Product	Power Query Editor	Create the SK_Customer
	Marge queries/Expanded columns	SK_Product/Fact_table	Power Query Editor	Merge SK_Customer into the Fact_Table
	Removed Columns		Power Query Editor	Removed all columns unnecessary to our model
	Change type of the columns		Power Query Editor	Change type of all columns
	Rename columns		Power Query Editor	Rename all columns that were necessary
Customer	Add Index Columns	SK_Customer	Power Query Editor	Create the SK_Customer
	Marge queries/Expanded columns	SK_Customer/Fact Table	Power Query Editor	Merge SK_Customer into the Fact_Table
	Marge queries/Expanded columns	Olist_geolocation_dataset	Power Query Editor	Expanded the columns latitude and longitude into Customer Dimension using the Zip_Code_Prefix as key from geolocation dataset. The last one was stopped load.
	Removed Columns		Power Query Editor	Removed all columns unnecessary to our model
	Change type of the columns		Power Query Editor	Change type of all columns
	Rename columns		Power Query Editor	Rename all columns that were necessary
Seller	Add Index Columns	SK_Seller	Power Query Editor	Create the SK_Seller
	Marge queries/Expanded columns	SK_Seller/Fact Table	Power Query Editor	Merge SK_Seller into the Fact_Table
	Marge queries/Expanded columns	Olist_geolocation_dataset	Power Query Editor	Expanded the columns latitude and longitude into Seller Dimension using the Zip_Code_Prefix as key from geolocation dataset. The last one was stopped load.
	Removed Columns		Power Query Editor	Removed all columns unnecessary to our model
	Change type of the columns		Power Query Editor	Change type of all columns
	Rename columns		Power Query Editor	Rename all columns that were necessary
Date	Create a dimension table with the dates	Dimension Table	Power Query Editor	Using the option new query, blank query. It was inserted the list of dates using this formula: List.Dates(#date(2016,09,1),Number.From(#date(2018,9,5)) - Number.From(#date(2016,09,1)), #duration(1,0,0,0)). We based on the order_purchase_timestamp to chose the list of dates.
	Create collumns related to Date		Power Query Editor	Create the columns referent to dates: Year, Month, Month_Name, Day_of_Week, Day Name,Day.
	Add Index Columns	SK_Date	Power Query Editor	Create the SK_Date
	Marge queries/Expanded columns	SK_Date/Fact Table	Power Query Editor	Merge SK_Date into the Fact_Table
	Change type of the columns		Power Query Editor	Change type of all columns
	Rename columns		Power Query Editor	Rename all columns that were necessary

11 – REFERENCES

[1] Olist and André Sionek, “Brazilian E-Commerce Public Dataset by Olist.” Kaggle, 2018, doi: 10.34740/KAGGLE/DSV/195341. Viewed on 05 March 2021

<https://www.kaggle.com/olistbr/brazilian-ecommerce>

[2] Ingizza, C., 2020, Vendas online no Brasil crescem 47% no 1º semestre, maior alta em 20 anos. Viewed on 20 March 2021

<https://exame.com/pme/e-commerce-brasil-cresce-47-primeiro-semestre-alta-20-anos/>

[3] Ebit | Nielsen and Elo, 2020. 42º Webshoppers. Viewed on 18 March 2021.

<https://company.ebit.com.br/webshoppers>

[4] Reclame aqui, “Olist” Viewed on 20 March 2021.

< https://www.reclameaqui.com.br/empresa/olist/>

[5] Carolina Ingizza, Exame,19th November 2020, “Olist, marketplace de marketplaces, recebe novos R$ 310 mi do SoftBank” Viewed on 20th March 2021.

<https://exame.com/pme/olist-marketplace-de-marketplaces-recebe-novos-r-310-mi-do-softbank/>

[6] Camilla de Oliveira, Olist, 2nd October 2019 “Números do e-commerce brasileiro: 34 dados que você PRECISA conhecer!” Viewed on 20th March 2021.

<https://blog.olist.com/1-numeros-e-commerce-brasileiro-e-dados-vendas-online/>

[6] Tiago Dalvi, Olist, 30th June 2016 “A história do Olist, por Tiago Dalvi” Viewed on 20th March 2021

<https://www.youtube.com/watch?v=z01QmZyvhes>

[7] Wikipedia, 6th December 2019 “Olist” Viewed on 20th March 2021.

<https://pt.wikipedia.org/wiki/Olist>

[8] Jéssica Sant’Ana, Gazeta do Povo, 14th November 2017 “Olist quer ser a maior loja virtual dentro dos principais marketplaces do país” Viewed on 20th March 2021.

<https://www.gazetadopovo.com.br/economia/nova-economia/olist-quer-ser-a-maior-loja-virtual-dentro-dos-principais-marketplaces-do-pais-3xbe9k4mn13uzs1pm60ym6znz/>

[9] Redator Rock Content, 24th August 2018, “Marketplace x E-commerce: entenda as diferenças entre esses modelos de negócios online” Viewed on 20th March 2021

<https://rockcontent.com/br/blog/marketplace-e-ecommerce/>

[10] Precifica. Setembro 2018. “Panoramo dos Marketplaces no Brasil”. Viewed on 20 March 2021.

<https://www.precifica.com.br/wp-content/uploads/2018/10/panorama-dos-marketplaces-no-brasil-edicao-setembro-2018.pdf>