Table of Contents

Abstract ..... **ii**

Chapter 1: Introduction **1**

1.1 Description of the project idea 1
1.2 Problem statement 1
1.3 Technology and tools used 1

Chapter 2: Requirements and Analysis **2**

2.1 Project Life-Cycle 2
2.2 Project Functionalities and Capabilities 6

Chapter 3: Implementation **7**

3.1 Data Model 7
3.2 Project Architecture 8
3.3 Visualization Module 9
3.4 Implementation and Code 9
3.5 Excel Reports 12

Chapter 4: Project Output **13**

Abstract

Access to reliable and accurate information is essential for management of performance for a company. Administrative staff must have access to current and historical information to fulfill their administrative duties.

Accordingly, an analysis is proposed in this paper which help decision makers in Olist company to track sales performance of the company from a historical data during years. Concluding this analysis with reports will form a business intelligence suite that will improve the business value.

Key Words: Data Warehousing, Data base, Business Intelligence, Data Analysis, Sales, Olist, Data Science.

Chapter 1: Introduction

Olist is a Brazilian E-commerce which is the largest department store in Brazilian Marketplaces. Olist connects small businesses from all over Brazil to channels without hassle and with a single contract. Those merchants are able to sell their products through the Olist store.

1.1 Description of the project idea

In This project we are addressing sales performance of Olist company. We are making analysis about orders made on the website so we can have a valuable information about our orders selling performance during years.

so, we are getting some information about orders such as Order's date, ID of the customer bought that order, city the order delivered to and the rate given to that order.

1.2 Problem statement

Olist.com is one of the largest online stores in Brazil as Souq.com in Egypt so as any online store need to always improve their business and to know if this business is successful and on the right way or it is failing or starting to fail, we need to track the performance of the store in order to keep the business on the safe way.

1.3 Technology and tools used

Languages used:

- **S**tructured **Q**uery **L**anguage (**SQL**).
- **M**ulti-**D**imensional E**x**pressions **(MDX)**

Software used:

- Microsoft SQL server Management Studio 2019 (for SSIS)
- Visual Studio 2017 (for SSAS)
- Tableau (for Dynamic Visualizations)

Chapter 2: Requirements and Analysis

2.1 Project Life-Cycle

- ### Data Acquisition

  This is a Brazilian E-commerce public dataset of orders made at olist store. The dataset has information of 100k orders from 2016 to 2018 made at a multiple marketplace in Brazil. It allows viewing an order from multiple dimensions: from order status, order date, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. The dataset also has a geolocation that relates Brazilian zip codes to lat/lng coordinates.

  The Dataset has 9 tables (olist_customers_dataset, olist_geolocation_dataset, olist_order_items_dataset, olist_order_payments_dataset, olist_order_reviews_dataset, olist_orders_dataset, olist_products_dataset, olist_sellers_dataset)

  Important note:

  - An order might have multiple items (products) so the price of an order isn't determined with only one number but, it is the summation of the order items.

- ### Data Understanding and Filtering

  Olist connects small businesses from all over Brazil to channels without hassle and with a single contract. Those merchants are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners.

  After a customer purchases the product from Olist Store a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

Data Understanding:

- olist_customers_dataset: This table has data about the customer and its location.

- olist_geolocation_dataset: This table has data about the Brazilian zip codes and its lat/lng coordinates.

- olist_order_items_dataset: This table has data about the items purchased within each order.

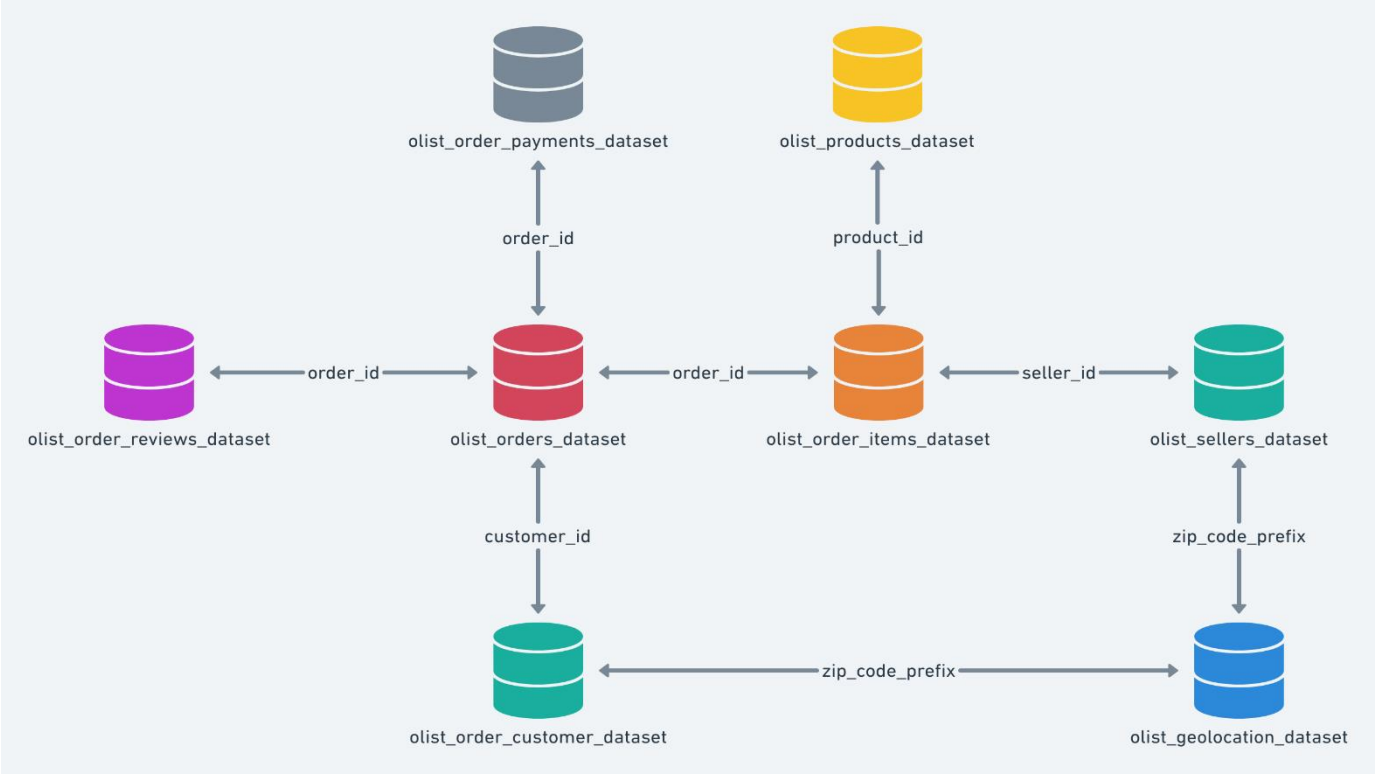
- olist_order_payments_dataset: This table has data about the orders payment options.

- olist_order_reviews_dataset: This table has data about the reviews made by the customers.

- olist_orders_dataset: This is the core table. From each order you might find all data you want as order status, delivery date, order price...

- olist_products_dataset: This table includes data about the products sold by Olist.

- olist_sellers_dataset: This table includes data about the sellers that fulfilled orders made at Olist.



Data Filtering:

In our project we need to work specifically on orders and analyze information about it so, main dimensions we will work on are:

- Customer who bought the order.
- Order Date.
- City that the order delivered to.

- Review about the order.

Therefore, the tables needed to work on with our Data Warehouse are:

- olist_customers_dataset.
- olist_orders_dataset.
- olist_order_items_dataset.
- olist_order_reviews_dataset.

• Building Star Schema

In the 4 tables we will use in our star schema we need these attributes: -

- olist_customers_dataset.
  - customer_id
  - customer_city
- olist_orders_dataset.
  - order_delivered_customer_date
- olist_order_reviews_dataset.
  - review_score

Motivation:

In our dataset as there are historical data for orders made in 3 years, we are creating an Order Sales star schema to analyze the sales profit statistics of our system during these years. So, our grain will be order date, customer who bought this order, city that the order is delivered to, review made on this order, number of products in this order and the total price of this order.

Therefore, our star schema will include four dimensions,

1) Time Dimension.
2) Customer Dimension.
3) Location Dimension.
4) Review Dimension

and 2 measures

1) Total Price of the order. (as each order could have more than one product)
2) Number of products in the order.

• OLAP

In this phase we are running in 3 subphases: -

First subphase:

The OLAP cube is a data structure optimized for very quick data analysis. It consists of numeric facts called measures which are categorized by dimensions. so, it allows us to use data in a multidimensional view.

Therefore, in this phase we will start building our cube with the star schema we build in the previous phase to represent our data for analysis.

So, hierarches that we need to make on that cube is: -

- Order Date will be in (Order year, Order Month, Order Day)

Second subphase:

MDX Queries allows us manipulate multidimensional information in our cube as it is optimized to query in a statistical way.

So, we build 4 MDX queries that outputs: -

1) Representation of all the orders made in 2018 only.

Order Year	Customer Id	Customer City	Review Score	Order Tprice	Products Qty
2018	000379cde...	sao paulo	4	93	1
2018	000419c54...	niteroi	1	34.3	1
2018	000598caf...	oliveira	5	1107	1
2018	0005aefbb...	sao paulo	4	134.9	1
2018	00066ccbe...	novo hambur...	1	199.6	4
2018	000fd45d6f...	piracaia	4	53.99	1
2018	001028b78...	sao paulo	5	29.9	1
2018	00104a47c...	uberlandia	4	83.99	1
2018	001051abf...	sao paulo	4	49.97	1
2018	0012a5c13...	rio de janeiro	5	349.9	1
2018	001328044...	sao paulo	5	149.9	1
2018	0013cd8e3...	sao paulo	5	79.9	1
2018	001450ebb...	ibiuna	2	37.64	2
2018	00146ad30...	belo horizonte	3	199.9	1
2018	0015bc9fd...	sao jose dos ...	5	122.99	1
2018	0017a0b4c...	curitiba	4	34.9	1
2018	0018c09f3...	santo andre	5	42.99	1
2018	0019c9aaa...	sao paulo	4	99.99	1

2) Total profit earned in month 2 in the 3 years 2016,2017 & 2018.

Order Month	Order Tprice
2	954836.77...

3)Total number of products sold in sao paulo city.

Customer City	Products Qty
sao paulo	17527

4) Total profit earned in each city in 2017 only.

Order Year	Customer City	Order Tprice
2017	abadia dos dourados	120
2017	abaete	1138.95
2017	abaetetuba	1785.51
2017	abaiara	93.9
2017	abare	149
2017	abatia	117.89
2017	abdon batista	125.4
2017	abelardo luz	447.8
2017	abre campo	281.68
2017	abreu e lima	312.88
2017	acaiaca	29.9
2017	acailandia	679.79
2017	acarau	604.59
2017	acopiara	177.89
2017	acreuna	43.8
2017	acu	349.6
2017	acucena	38

5) Customer satisfaction rate that gave rate 5 during years.

Order Year	Review Score	Sales Fact Count
2016	5	149
2017	5	24508
2018	5	32238

Third subphase:

Excel is the most used spreadsheet program in many business activities as it could help us in making static reports for our business needs.

So, we made one report and one chart that help us to represent sales performance of orders.

The Report: It represents the top 10 customers to olist as they are the top 10 customers in order's price measure.

The Chart: It represents the top 10 cities bought products as they are the top 10 cities in the number of products measure.

- **Business Intelligence**

Dynamic visualization is used to create different views to the same data. In addition to being a tool for communication, dynamic visualization serves as a tool for exploration: grouping and regrouping of variables, highlighting and filtering support decision making. The output chart could change its displayed results based on the required dimension attribute value, as we can also drill down or up in our cube and show how this affects our charts and graphs.

So, we used tableau software to make a dashboard to show the sales performance for olist company by making analysis on orders available in its dataset.

2.2 Project Functionalities and Capabilities

This project is able to answer questions about sales performance about olist company to help them making decisions and how the analysis predict the future of the company. As now from the available dataset about olist we made some analysis by filtering data, building Data Warehouse and building our cube.

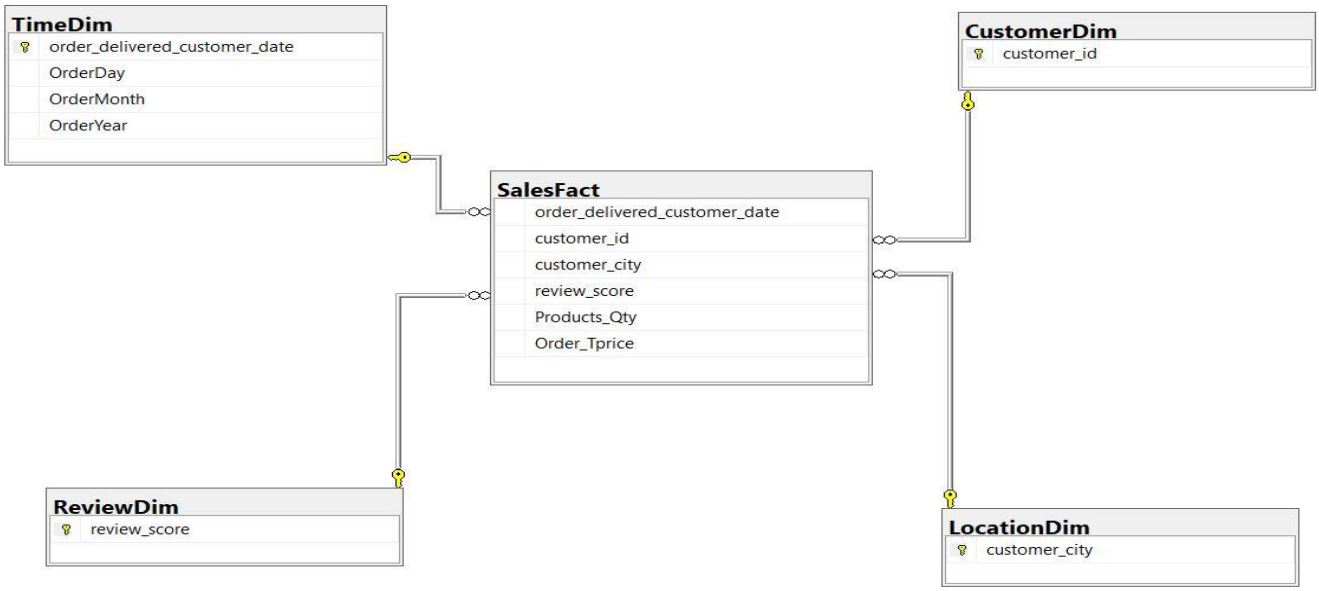
Now Olist company could see how their sales is going from this project as it answers some important questions as: -

- How much they sell products from all over the 3 years and months or in a specific year.
- How much they earn from all over the 3 years and months or in a specific year.

- Top customers with respect to sales for Olist.
- Top cities with respect to sales for Olist.
- Customer satisfaction rate for Olist orders.

Chapter 3: Implementation

3.1 Data Model



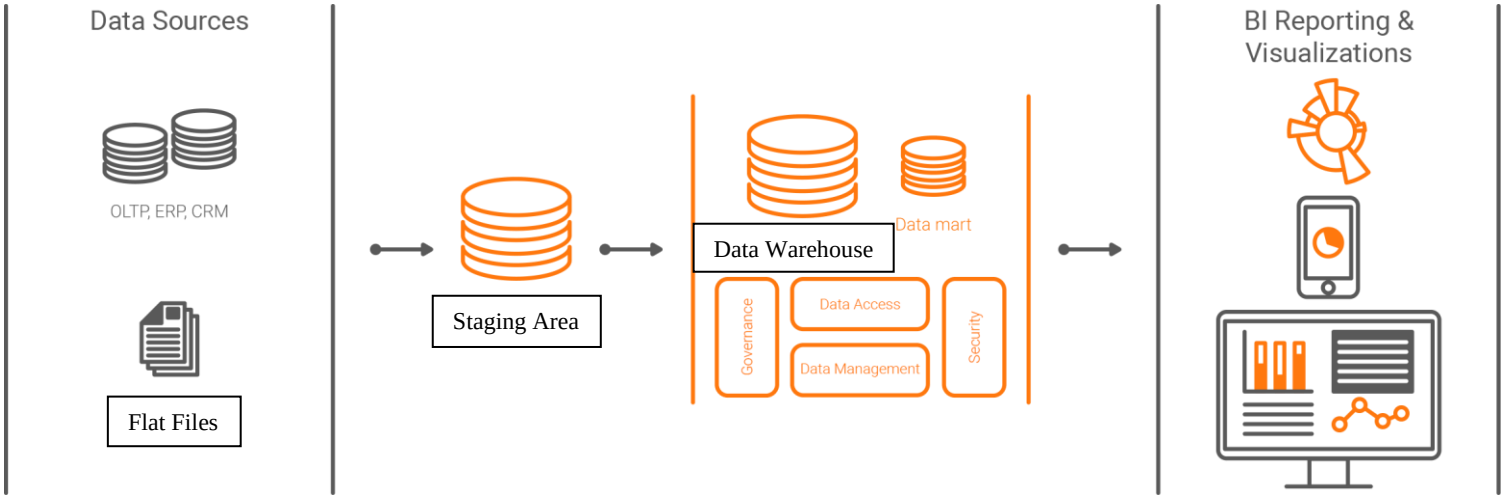
- TimeDim
  - Dimension shows data about dates for order. It contains 4 attributes.
    - Order_delivered_customer_date: - Date of delivering the order to the customer. **(Primary Key)**

    And this dimension has 3 hierarchies
    - OrderYear: - Year part of the delivered date.
    - OrderMonth:- Month part of the delivered date.
    - OrderDay: - Day part of the delivered date.
- CustomerDim
  - Dimension shows data about customers. It contains 1 attribute.
    - Customer_id for each customer who made an order. **(Primary Key)**
- ReviewDim
  - Dimension shows data about reviews of orders. It contains 1 attribute.
    - Review_score for each order made in olist from 1 (lowest) to 5 (highest). **(Primary Key)**
- LocationDim

- Dimension shows data about location of order's delivery. It contains 1 attribute.
  - Customer_city : - city of the customer that the order is delivered to. **(Primary Key)**

- SaleFact
  - Fact table that shows data about each order made in olist as each order has more than one item so data showed in the table is order date, the customer who bought this order, city that the order is delivered to, review of that order, total price of this order and number of products in this order. It contains 6 attributes.
    - Order_customer_dilevered_date: - Date of delivering the order to the customer. **(Foreign Key) from TimeDim**
    - Customer_id for each customer who made an order. **(Foreign Key) from CustomerDim**
    - Review_score for each order made in olist from 1 (lowest) to 5 (highest). **(Foreign Key) from reviewDim**
    - Customer_city: - city of the customer that the order is delivered to. **(Foreign Key) from LocationDim**
    - Products_Qty: - as each order may have more than one item; this measure shows the number of products in this order.
    - Order_Tprice: - as each order may have more than one item; this measure shows the total price of that order.

3.2 Project Architecture



First, data acquisition from any data sources in a flat files format that where you get your database that you will work on.

Second, go through staging area which is **E**xtract-**T**ransform-**L**oad (**ETL**) where we start extracting the data, data cleansing & Transformation and then loading the data.

Third, Data Warehouse where we start loading our filtered data to our data warehouse as a star schema with dimensions which is our descriptive data and a fact table which include our descriptive data and our needed measures.

Fourth, Reporting and visualizations that start with building our cube to make our star schema in a multidimensional format to use it in analysis then after than I could make needed MDX queries, build excel reports or build our dynamic visualizations with tableau.

3.3 Visualization Module

Tableau Software:

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format, used to drill down data and see the impact in a visual format that can be easily understood by any individual. Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets.

3.4 Implementation and Code

Star Schema Queries: -

- TimeDim (Getting the delivered date of each order)

```
Select distinct (order_delivered_customer_date) ,
DAY(order_delivered_customer_date) AS OrderDay
,MONTH(order_delivered_customer_date) AS OrderMonth
,YEAR(order_delivered_customer_date) AS OrderYear
From olist_orders_dataset
Where order_delivered_customer_date is not null
order by order_delivered_customer_date
```

- CustomerDim (Getting the customer id of the customer who bought that order)

```
Select customer_id
From olist_customers_dataset
```

- ReviewDim (Getting the review score the customer gave about the order)

```
Select DISTINCT review_score
From olist_order_reviews_dataset
```

- LocationDim (Getting the city of the customer where the order is delivered to)

```
Select DISTINCT customer_city
From olist_customers_dataset
```

- SalesFact

```
Select
        olist_orders_dataset.order_delivered_customer_date,
        olist_orders_dataset.customer_id,
        olist_customers_dataset.customer_city,
        olist_order_reviews_dataset.review_score,
        count(olist_order_items_dataset.order_id) AS Products_Qty,
        sum(olist_order_items_dataset.price) AS Order_Tprice
From
        olist_orders_dataset, olist_order_items_dataset,
        olist_customers_dataset,olist_order_reviews_dataset
Where
        olist_orders_dataset.order_id =
        olist_order_items_dataset.order_id
        and olist_orders_dataset.order_id =
        olist_order_reviews_dataset.order_id
        and olist_customers_dataset.customer_id =
        olist_orders_dataset.customer_id
Group by
        olist_orders_dataset.order_delivered_customer_date,olist_orde
        rs_dataset.customer_id,olist_customers_dataset.customer_city,
        olist_order_reviews_dataset.review_score
Order by Order_Tprice DESC
```

MDX Queries: -

1) All orders in 2018

```
SELECT NON EMPTY {[Measures].[Order Tprice],
[Measures].[Products Qty] } ON COLUMNS,
NON EMPTY { ([Time Dim].[Hierarchy].[Order Year].&[2018] *
[Customer Dim].[Hierarchy].[Customer Id].ALLMEMBERS *
[Location Dim].[Hierarchy].[Customer City].ALLMEMBERS *
[Review Dim].[Hierarchy].[Review Score].ALLMEMBERS ) }
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [OlistDWH]
```

2) Sales of Month 2 in all years

```
SELECT
[Measures].[Order Tprice] ON COLUMNS,
Descendants(
[Time Dim].[Order Month].&[2],
[Time Dim].[Order Year],self) ON ROWS
From [OlistDWH]
```

3) Total Order quantity for sao paulo city

```
SELECT {[Measures].[Products Qty]} ON Columns,
{Except ([Location Dim].[Customer City].&[sao paulo] ,[Location
Dim].[Customer City])} ON ROWS
FROM [OlistDWH]
```

4) Sales of cities in 2017

```
SELECT NON EMPTY { [Measures].[Order Tprice] } ON
COLUMNS,
NON EMPTY { ([Time Dim].[Order Year].[Order Year].&[2017] *
[Location Dim].[Hierarchy].[Customer City].ALLMEMBERS ) }
DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [OlistDWH]
```

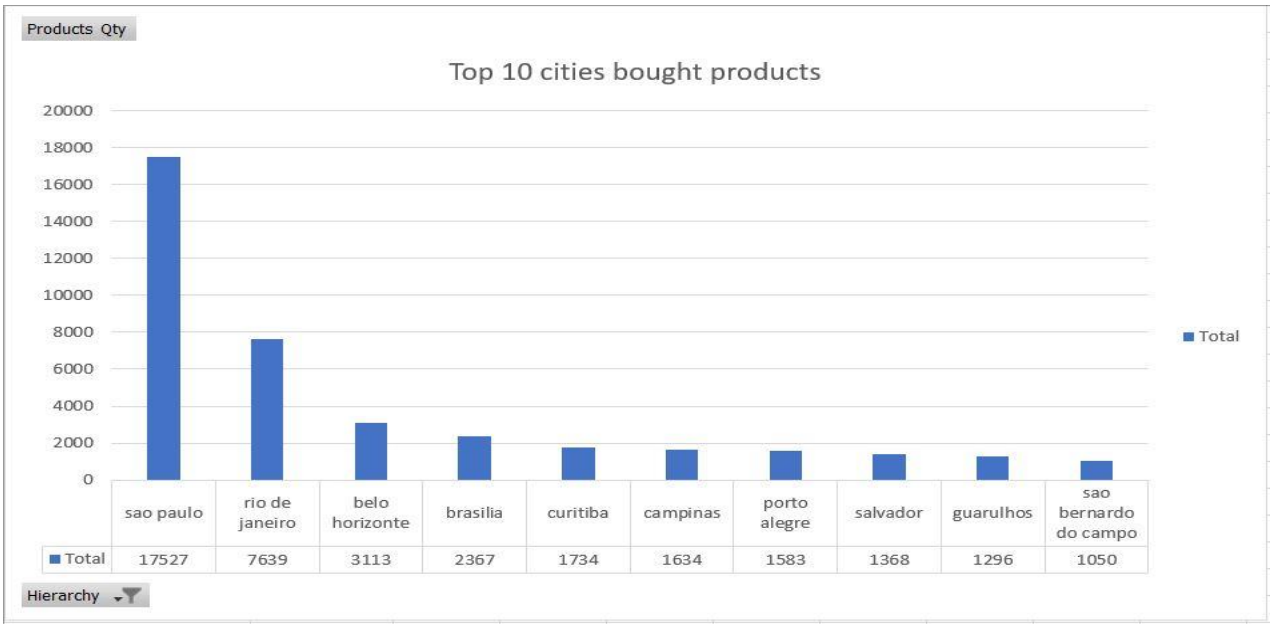
5) Number of customers gave rate 5 for orders during years

```
SELECT NON EMPTY { [Measures].[Sales Fact Count] } ON
COLUMNS,
NON EMPTY { ([Time Dim].[Hierarchy].[Order
Year].ALLMEMBERS * [Review Dim].[Hierarchy].[Review
Score].&[5] ) } DIMENSION PROPERTIES MEMBER_CAPTION,
MEMBER_UNIQUE_NAME ON ROWS
FROM [OlistDWH]
```

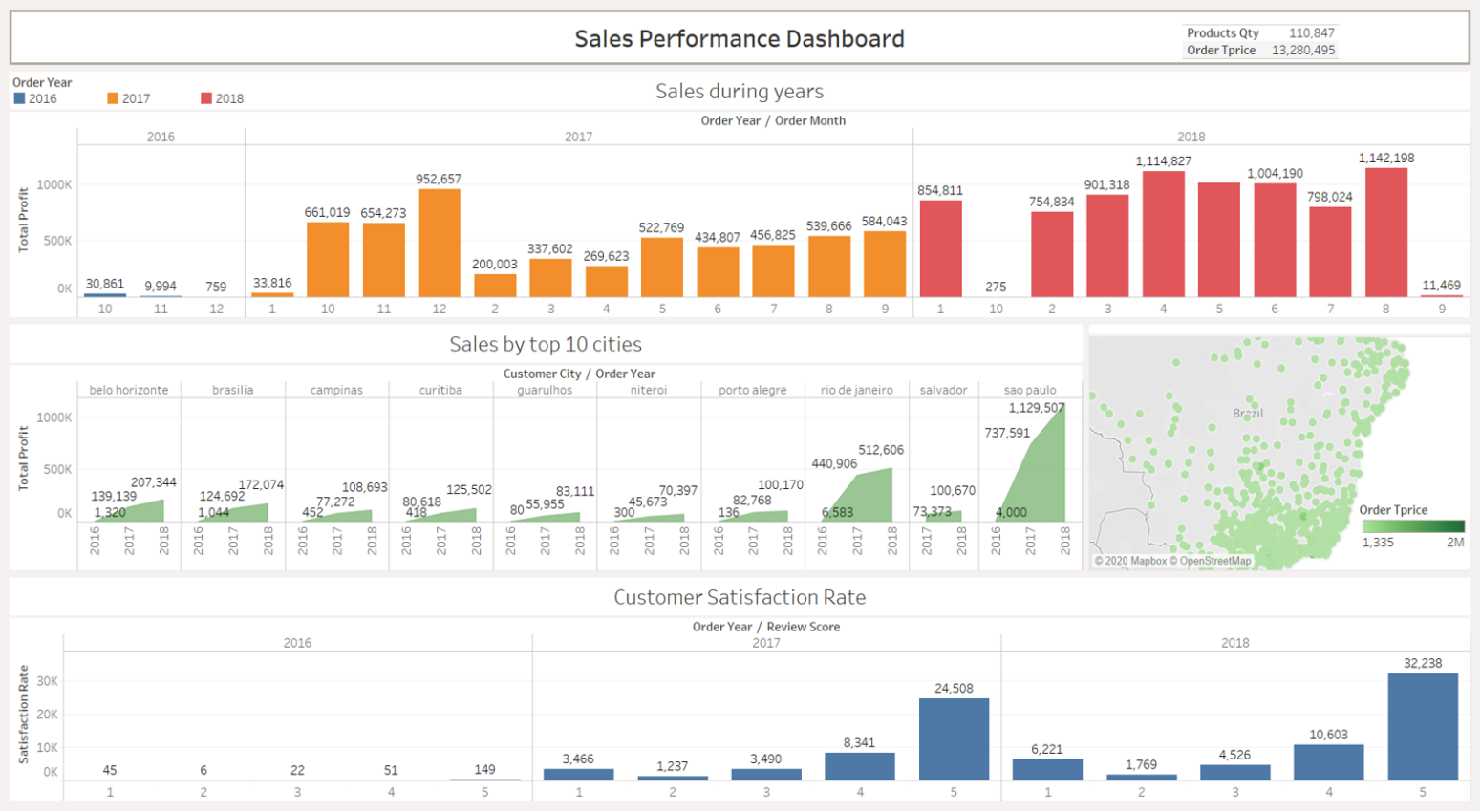
Excel Report: -

Top 10 Customers	
Customer_ID	Total price
1617b1357756262bfa56ab541c47bc16	13440
ec5b2ba62e574342386871631fafd3fc	7160
c6e2731c5b391845f6800c97401a43a9	6735
f48d464a0baaea338cb25f816991ab1f	6729
3fd6777bbce08a352fddd04e4a7cc8f6	6499
05455dfa7cd02f13d132aa7a6a9729c6	5934.6
df55c14d1476a9a3467f131269c2477f	4799
24bbf5fd2f2e1b359ee7de94defc4a15	4690
3d979689f636322c62418b6346b1c6d2	4590
cc803a2c412833101651d3f90ca7de24	4400
Grand Total	**64976.6**

Excel Chart:



Chapter 4: Project Output



Concluded report showed a great valuable information for decision makers some of them are: -

- Olist is making good sales during years with total profit 13,280,495 and sold 110,847 products.
- Sao Paulo city made a great income for olist.
- Customer satisfaction rate is good but not the best.
- Total profit is increasing as 2018 shows highest sales during the 3 years.

References

- https://www.kaggle.com/olistbr/brazilian-ecommerce
- https://www.tableau.com/
- Lectures