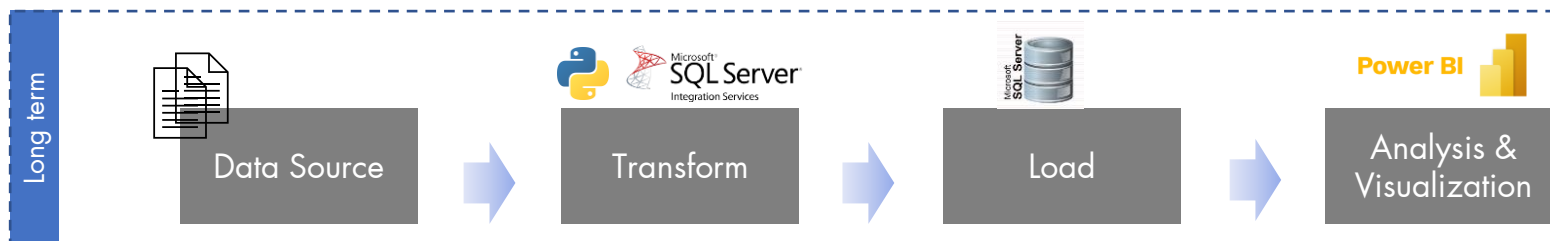
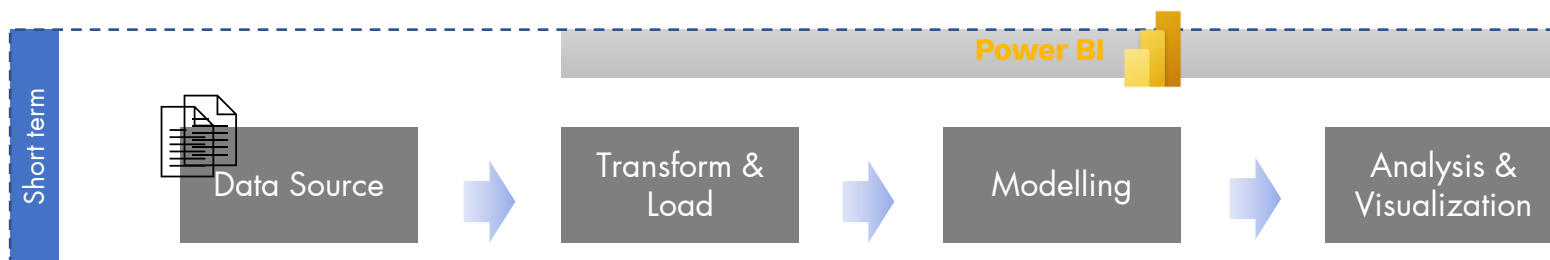


Brazilian E-Commerce

Technical design

Thiết kế



Define datasource

Overview

- 9 flat csv files : Customers, Products, Product Category, Sellers, Geolocation, Orders, Order items, Order payments, Order reviews
- Total size : 120M
- Biggest file : Geolocation, > 1mil rows

Name	Size	
olist_customers_dataset.csv	8.823 KB	olist_customers_dataset 99.441 rows loaded.
olist_geolocation_dataset.csv	59.838 KB	olist_geolocation_dataset 1.000.163 rows loaded.
olist_order_items_dataset.csv	15.077 KB	olist_order_items_dataset 112.650 rows loaded.
olist_order_payments_dataset.csv	5.642 KB	olist_order_payments_dataset 103.886 rows loaded.
olist_order_reviews_dataset.csv	14.072 KB	olist_order_reviews_dataset 100.000 rows loaded.
olist_orders_dataset.csv	17.242 KB	olist_orders_dataset 99.441 rows loaded.
olist_products_dataset.csv	2.324 KB	olist_products_dataset 32.951 rows loaded.
olist_sellers_dataset.csv	171 KB	olist_sellers_dataset 3.095 rows loaded.
product_category_name_translation.csv	3 KB	product_category_name_translation 71 rows loaded.

Quality

- Repetitive customer city and State

Geolocation

A ^B _C geolocation_city	A ^B _C geolocation_state
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
sao paulo	SP
sao paulo	SP
sao paulo	SP

Customer

A ^B _C customer_city	A ^B _C customer_state
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
franca	SP
sao bernardo do campo	SP
sao paulo	SP
mogi das cruzeiras	SP

- Multiple geolocation city versions for 1 zipcode (Geolocation file and Customer file)

Geolocation

geolocation_zip_code_prefix	geolocation_city
01307	sao bernardo do campo
	sao paulo
	são paulo
01308	sao paulo
	são paulo
01309	sao paulo
	são paulo

Customer

customer_id	customer_unique_id	customer_zip_code_prefix	customer_city
502de0f9e9835ab231c2851acd163376	6d004368588f6fee399cf...	06806	embu das artes
ad587a1071e2eb39139bdf6c7e49d90b	83d0c398734ab60891d3...	06806	embu das artes
8b8e02be8fef10b7c6e2d210dbcfba7c	ed05a1bbf9f8816af23fc...	06806	embu
2ed1b5c01561bcf39cc2f1566a8ac9f1	ed05a1bbf9f8816af23fc...	06806	embu
olist_customers_dataset[customer_zip_code_prefix]		olist_customers_dataset[customer_city]	
89294		fragosos	
89294		campo alegre	

- Duplicated data (Geolocation)

Geolocation

geolocation_zip_code_prefix	geolocation_city	geolocation_lat	geolocation_lng	Count of geolocation_state
01001	sao paulo	-23.549291999999999	-46.633559478233785	6
		-23.54969829946912	-46.63390859285005	1
		-23.549779299469115	-46.6339571183853	1
		-23.5498252739339	-46.63396956010149	2
		-23.550497706907514	-46.63433817805407	8
		-23.55064182209015	-46.634409790322515	1
		-23.551336655288804	-46.63402699777831	3

Transformation data

ETL Process

Bước 1	Load data vào Power Query, sắp xếp thứ tự cột
Bước 2	Xác định dữ liệu liên quan, xác định các primary key của từng bảng
Bước 3	Kiểm tra chất lượng dữ liệu : khóa chính là giá trị duy nhất, tỷ lệ dữ liệu lỗi
Bước 4	Xóa các cột dữ liệu thừa, hoặc trùng lặp với bảng khác. Xóa dòng dữ liệu trùng lặp
Bước 5	Chuyển đổi dữ liệu, kiểu dữ liệu

Geolocation					
	1^2_3 geolocation_zip_code_prefix	1.2 geolocation_lat	1.2 geolocation_lng	A^B_C geolocation_city	A^B_C geolocation_state
1	1037	-23,54562128	-46,63929205	sao paulo	SP
2	1046	-23,54608113	-46,6448203	sao paulo	SP
3	1046	-23,54612897	-46,64295148	sao paulo	SP
4	1041	-23,54439216	-46,63949931	sao paulo	SP

	1^2_3 geolocation_zip_code_prefix	1.2 geolocation_lat	1.2 geolocation_lng	A^B_C geolocation_city	A^B_C geolocation_state
1	1037	-23.54672337	-46.64033633	sao paulo	SP
2	1046	-23.54719064	-46.64577103	sao paulo	SP
3	1041	-23.54463817	-46.64090618	sao paulo	SP
4	1035	-23.54394733	-46.64414706	sao paulo	SP

- Step 1 : Remove duplicated
- Step 2 : Mỗi Zip code xác định Min lat và Max long
- Step 3 : Thay thế các ký tự đặc biệt để mỗi city chỉ 1 tên duy nhất

Seller			
	A^B_C seller_id	1^2_3 seller_zip_code_prefix	A^B_C seller_city
1	3442f8959a84dea7ee197c632cb2df15	13023	campinas
2	d1b65fc7debc3361ea86b5f14c68d2e2	13844	mogi guacu
3	ce3ad9de960102d0677a81f5d0bb7b...	20031	rio de janeiro

	A^B_C seller_id	1^2_3 seller_zip_code_prefix
1	3442f8959a84dea7ee197c632cb2df15	13023
2	d1b65fc7debc3361ea86b5f14c68d2e2	13844
3	ce3ad9de960102d0677a81f5d0bb7b...	20031

Customer					
	A^B_C customer_id	A^B_C customer_unique_id	1^2_3 customer_zip_code_prefix	A^B_C customer_city	A^B_C customer_state
1	06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	14409	franca	SP
2	18955e83d337fd6b2def6b18a428ac...	290c77bc529b7ac935b93aa66c333d...	9790	sao bernardo do campo	SP
3	4e7b3e00288586ebd08712fdd0374...	060e732b5b29e8181a18229c7b0b2...	1151	sao paulo	SP

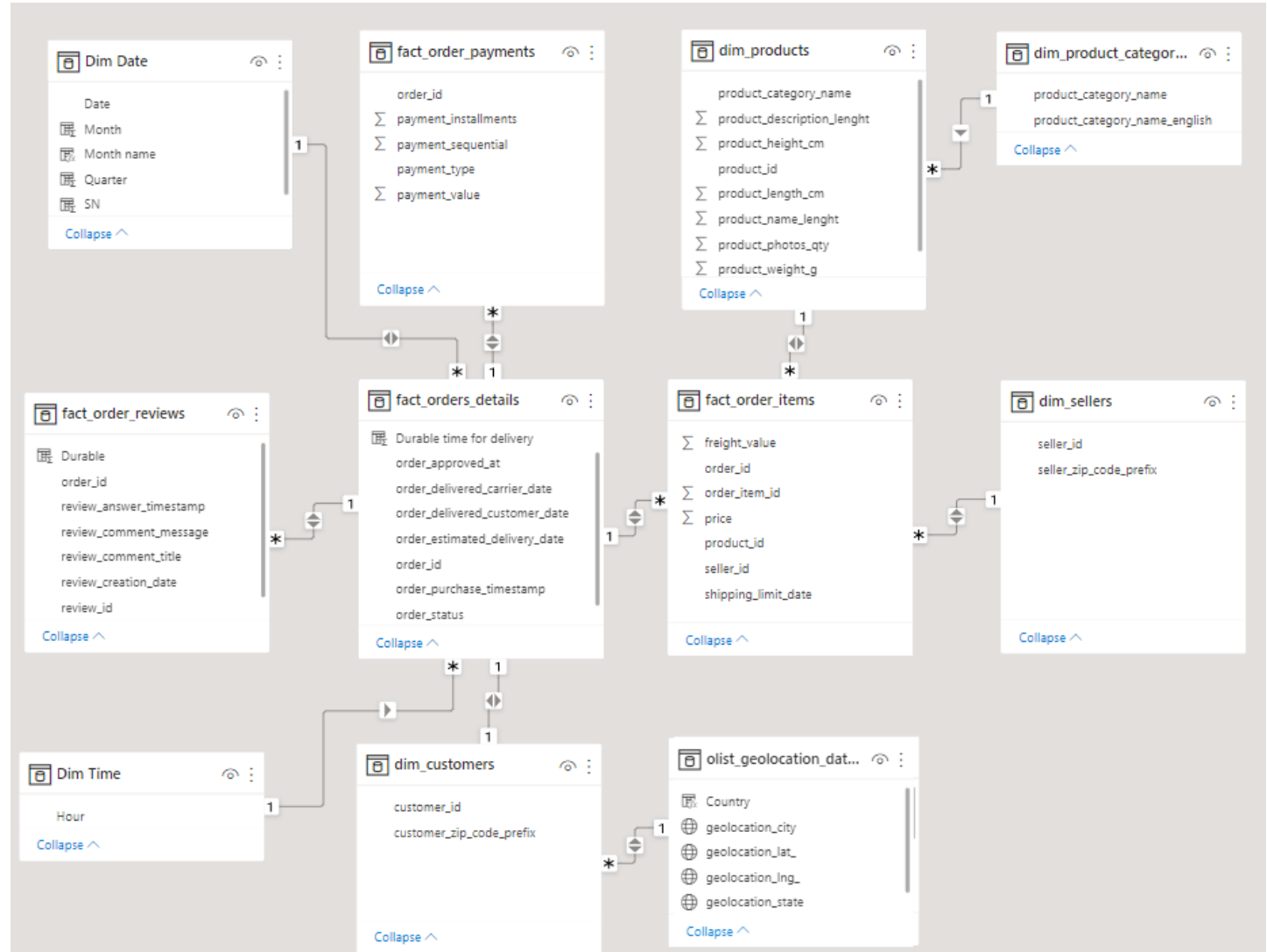
	A^B_C customer_id	1^2_3 customer_zip_code_prefix
1	06b8999e2fba1a1fbc88172c00ba8bc7	14409
2	18955e83d337fd6b2def6b18a428ac...	9790
3	4e7b3e00288586ebd08712fdd0374...	1151

Data Modelling

- Dax Measure
- Data Relationship

✓ Dax

📊	_Count order no by cancel status
📊	_Count order no by delivery status
📊	_Rate of Freight cost on Sales
📊	_Total Freight
📊	_Total Order
📊	_Total Sales
📊	_Total Sales by Country
📊	% freight growth with lastyear
📊	% order growth with lastyear
📊	% sales growth with lastyear



Dax Measure

- Tạo các measure để đo lường tính toán

```
1 % sales growth with lastyear =
2 VAR _sales_thisyear =
3     TOTALYTD ( SUM ( fact_order_payments[payment_value] ), 'Dim Date'[Date] )
4 VAR _sales_LY =
5     CALCULATE (
6         SUM ( fact_order_payments[payment_value] ),
7         REMOVEFILTERS ( 'Dim Date'[Year] ),
8         YEAR ( 'Dim Date'[Date] )
9         = SELECTEDVALUE ( 'Dim Date'[Year] ) - 1
10    )
11 VAR _result = ( ( _sales_thisyear - _sales_LY ) / _sales_LY * 100 )
12 RETURN
13     SWITCH (
14         TRUE (),
15         _result > 0,
16         UNICHAR ( 9650 ) & FORMAT ( _result, "###.#" ) & "%",
17         _result = 0, UNICHAR ( 9654 ),
18         _result < 0,
19         UNICHAR ( 9660 ) & FORMAT ( _result, "###.#" ) & "%"
20     )
21
```

Fields

Search

Dax

- _Count order no by cancel status
- _Count order no by delivery status
- _Rate of Freight cost on Sales
- _Total Freight
- _Total Order
- _Total Sales
- % freight growth with lastyear
- % order growth with lastyear
- % sales growth with lastyear

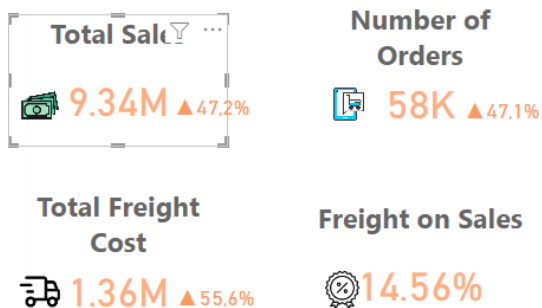
Column

- Dim Date
- Dim Time

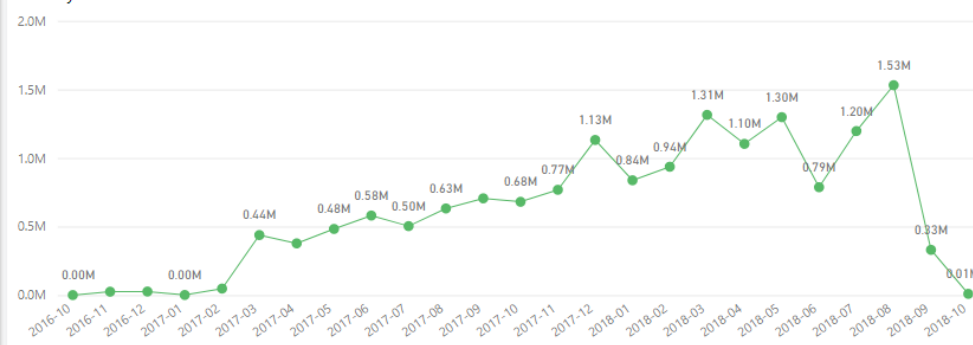
Thiết kế và tạo báo cáo

- Tạo các measure để đo lường tính toán

SALES ACTIVITY DASH BOARD



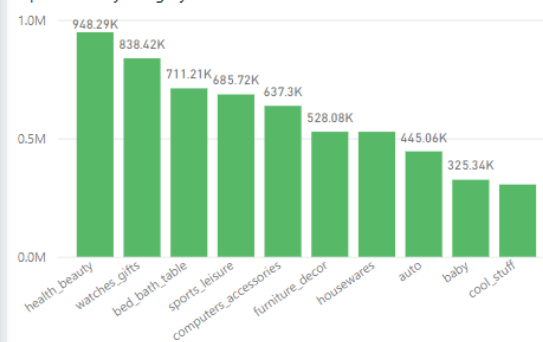
Sales by timeseries



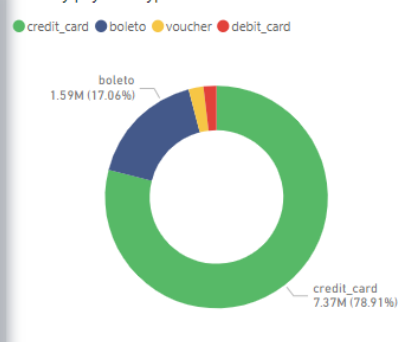
Sales by State



Top 10 sales by category



Sales by payment type



Visualizations

Build visual

Filters

Fields

Search

Dax

- ☐ _Count order no by cancel status
- ☐ _Count order no by delivery status
- ☐ _Rate of Freight cost on Sales
- ☐ _Total Freight
- ☐ _Total Order
- ☒ _Total Sales
- ☐ % freight growth with lastyear
- ☐ % order growth with lastyear
- ☐ % sales growth with lastyear

Dim Date

- ☐ Date
- ☐ Month
- ☐ Month name
- ☐ Quarter
- ☐ SN
- ☐ Year

Dim Time

dim_customers

dim_product_category_name_translation

- ☐ product_category_name
- ☐ product_category_name_english

dim_products

- ☐ product_category_name
- ☐ product_description_lenght
- ☐ product_height_cm
- ☐ product_id
- ☐ product_length_cm

Fields

_Total Sales

Drill through

Cross-report

Keep all filters

Add drill-through fields here