



# E-commerce Analytics Project

End-to-End Data Warehousing & BI Solution

Ehab Farid

Data Engineer / BI Developer

# project overview

The project focuses on cleaning and preparing multiple datasets from a Brazilian e-commerce platform, and building a complete Data Warehouse to enable advanced analytics.

The final model supports analyzing sales performance, shipping and delivery efficiency, product performance, seller performance, and customer experience across the platform.

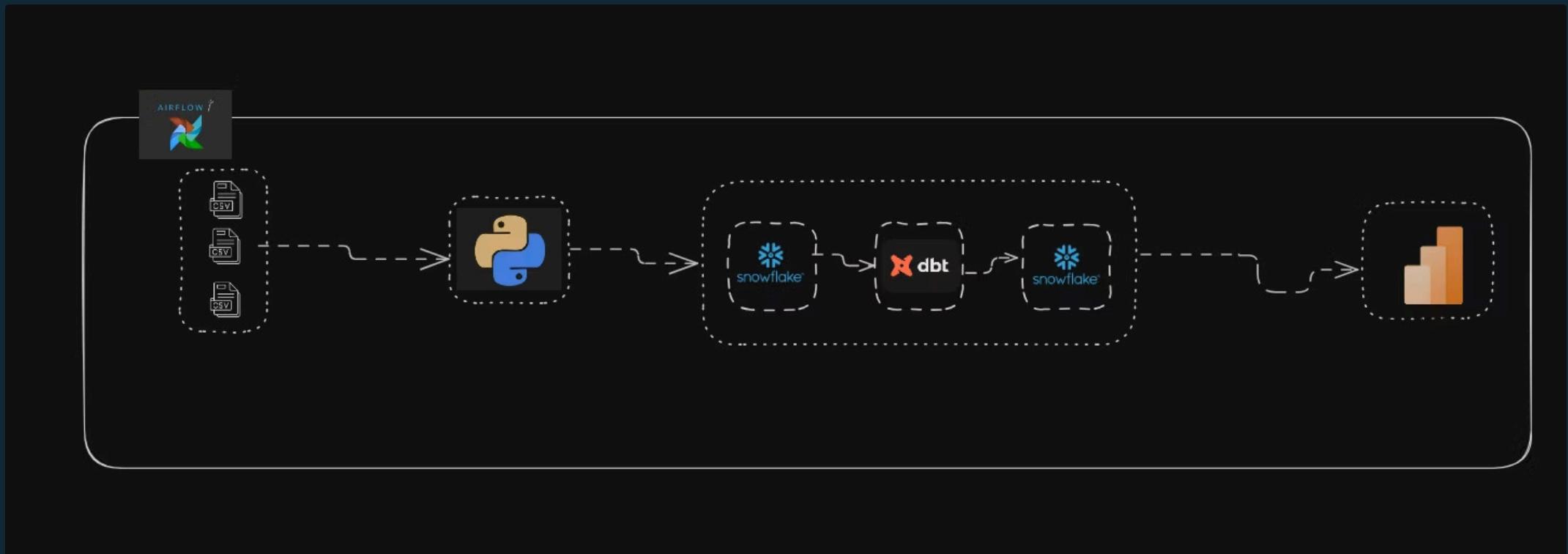




# Business Problem & Scope

- **The Context:** Brazilian marketplace connecting sellers to customers.
- **The Problem:** Data was scattered across 9 isolated CSV files (Silos), making analysis impossible.
- **The Goal:** Build a unified "Single Source of Truth" to analyze Sales, Logistics, and Seller Performance.
- **Target Audience:** Executives, Operations Managers, and Marketing Team.

# Modern Data Stack Architecture



# Raw Data

- 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
- product\_category\_name\_translation

# **Python (Pandas): Data Cleaning & Pre-processing**

# Data Cleansing (Python Layer):

- Remove Missing Values

Example: Dropped rows in products dataset where product\_category\_name or product\_photos\_qty was null.

## 1. Standardize Text Fields

Example: Converted customer\_city and geolocation\_city to lowercase + removed accents using unidecode.

(e.g., "sao paulo" vs "São Paulo").

## 1. Fix Data Types

Example: Converted order date columns (e.g., order\_purchase\_timestamp) to datetime using pd.to\_datetime.

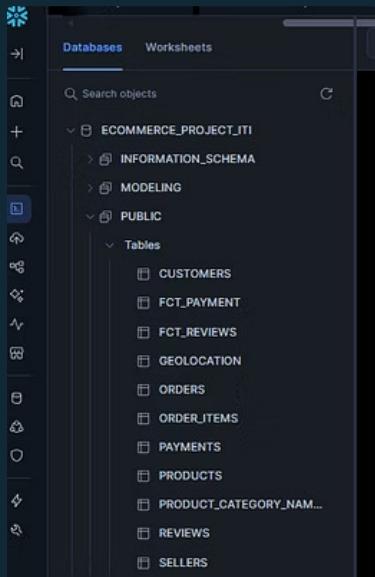
## 1. Remove Duplicates

Example: Removed duplicate rows in the geolocation dataset using .drop\_duplicates().

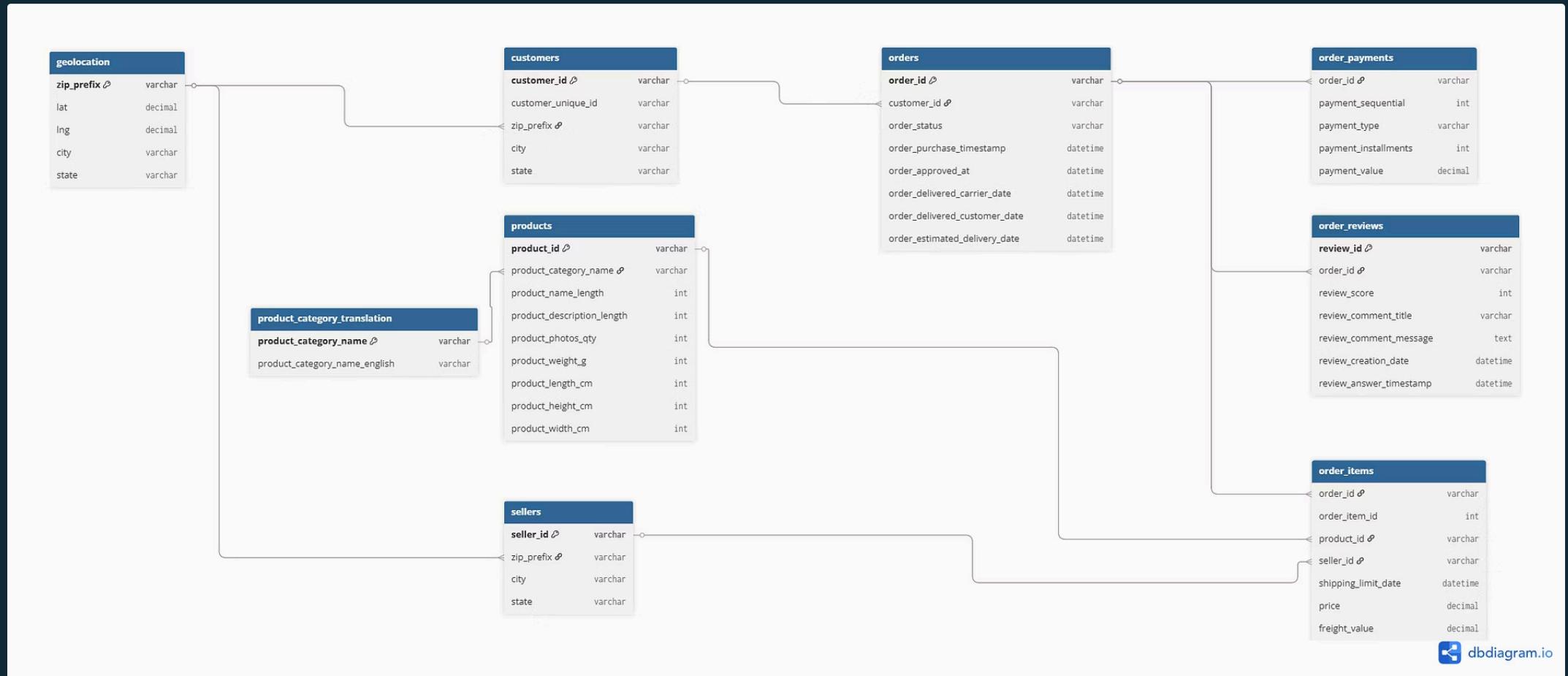
## 1. Export Clean Datasets

Example: Saved cleaned datasets into “clean csv files” folder (e.g., clean\_customers.csv).

# load data to snowflake



# ERD

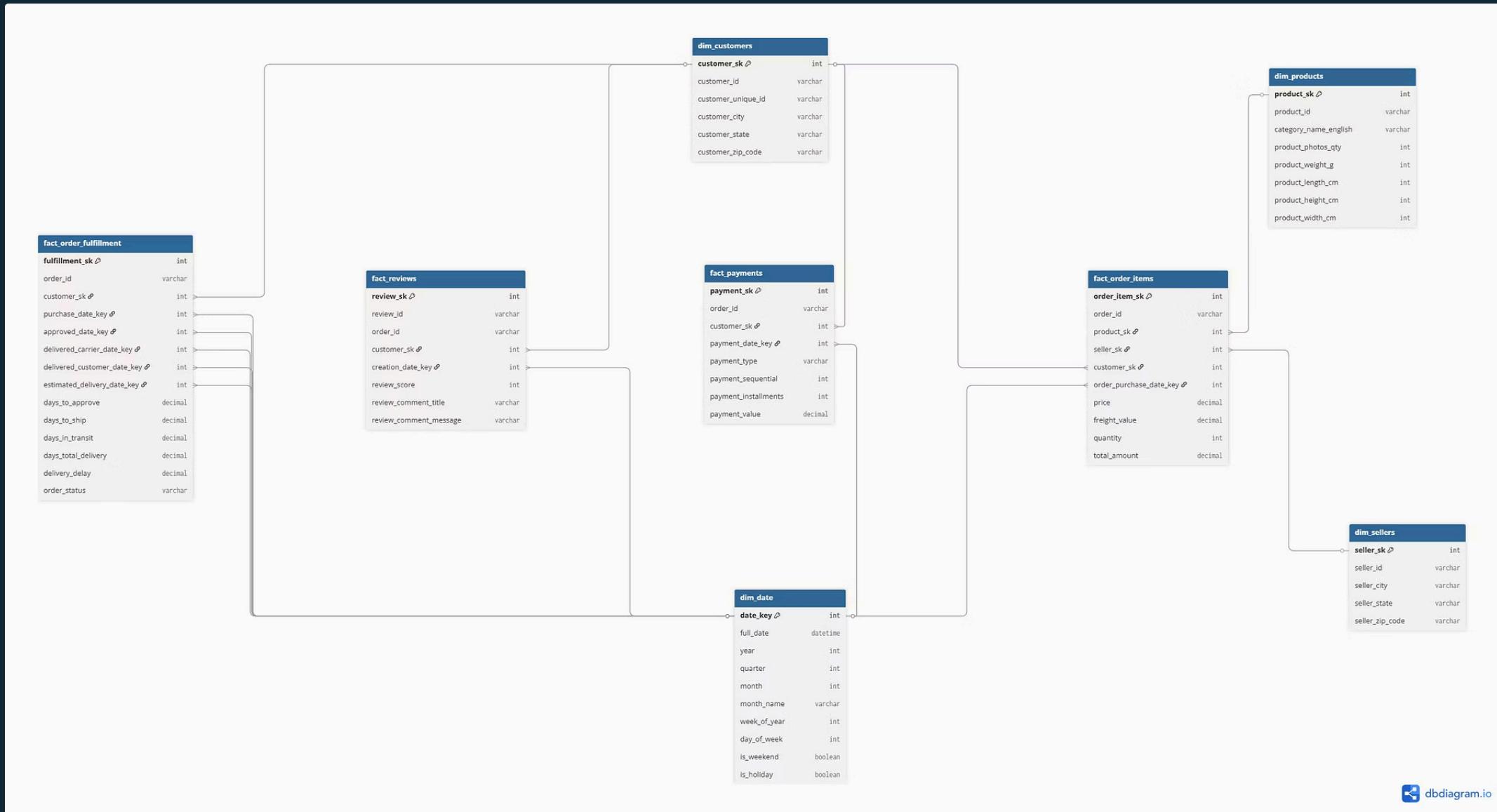


# Data Modeling (dbt Layer):

- Built the **Galaxy Schema** inside Snowflake.
- Created relationships between Facts and Dimensions.
- Calculated complex business logic (e.g., Delivery Delay, Shipping Time).

```
21:03:02 3 of 8 OK created sql table model PUBLIC_modeling.dim_products ..... [SUCCESS 1 in 4.94s]
21:03:02 1 of 8 OK created sql table model PUBLIC_modeling.dim_calender_date ..... [SUCCESS 1 in 5.00s]
21:03:03 4 of 8 OK created sql table model PUBLIC_modeling.dim_sellers ..... [SUCCESS 1 in 5.67s]
21:03:04 2 of 8 OK created sql table model PUBLIC_modeling.dim_customers ..... [SUCCESS 1 in 7.01s]
21:03:04 8 of 8 START sql table model PUBLIC_modeling.fct_payment ..... [RUN]
21:03:04 7 of 8 START sql table model PUBLIC_modeling.fct_order_transaction ..... [RUN]
21:03:04 5 of 8 START sql table model PUBLIC_modeling.fct_Reviews ..... [RUN]
21:03:04 6 of 8 START sql table model PUBLIC_modeling.fct_order_accumlate ..... [RUN]
21:03:07 5 of 8 OK created sql table model PUBLIC_modeling.fct_Reviews ..... [SUCCESS 1 in 2.41s]
21:03:07 7 of 8 OK created sql table model PUBLIC_modeling.fct_order_transaction ..... [SUCCESS 1 in 2.71s]
21:03:07 6 of 8 OK created sql table model PUBLIC_modeling.fct_order_accumlate ..... [SUCCESS 1 in 2.74s]
21:03:08 8 of 8 OK created sql table model PUBLIC_modeling.fct_payment ..... [SUCCESS 1 in 3.86s]
21:03:13
21:03:13 Finished running 8 table models in 0 hours 0 minutes and 21.95 seconds (21.95s).
21:03:13
21:03:13 Completed successfully
21:03:13
21:03:13 Done. PASS=8 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=8
(venv) PS C:\Users\ehabm\snowflake_data_project> |
```

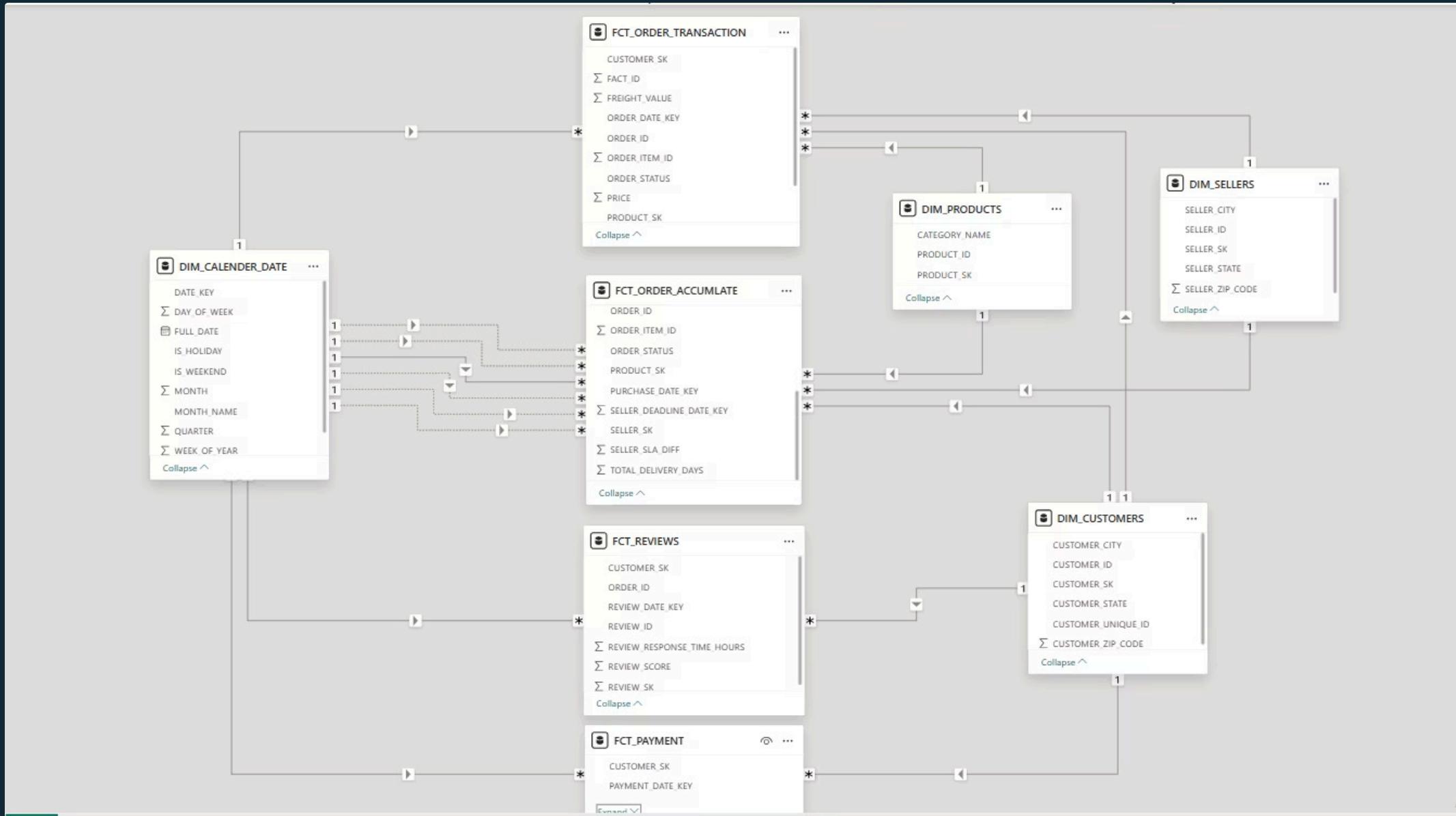
# DWH modelling



# why we use galaxy schema?

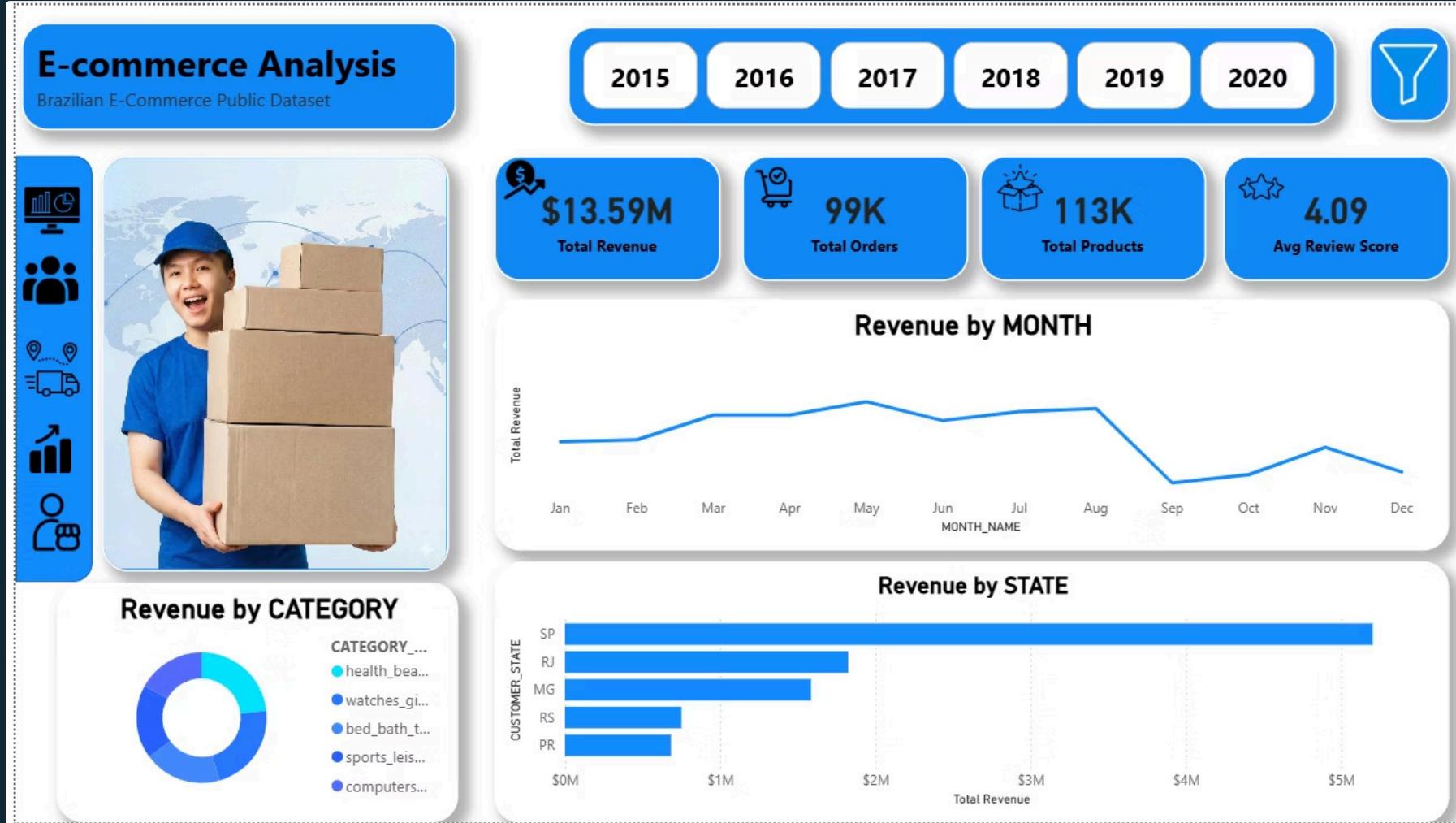
Because we wanted to analyze the entire product lifecycle from the very beginning, starting from when the customer places the order until the product is received and confirmed, including the time taken for preparation and processing by the sellers, as well as the time spent in transit until delivery to the customer.

# data modelling in power BI

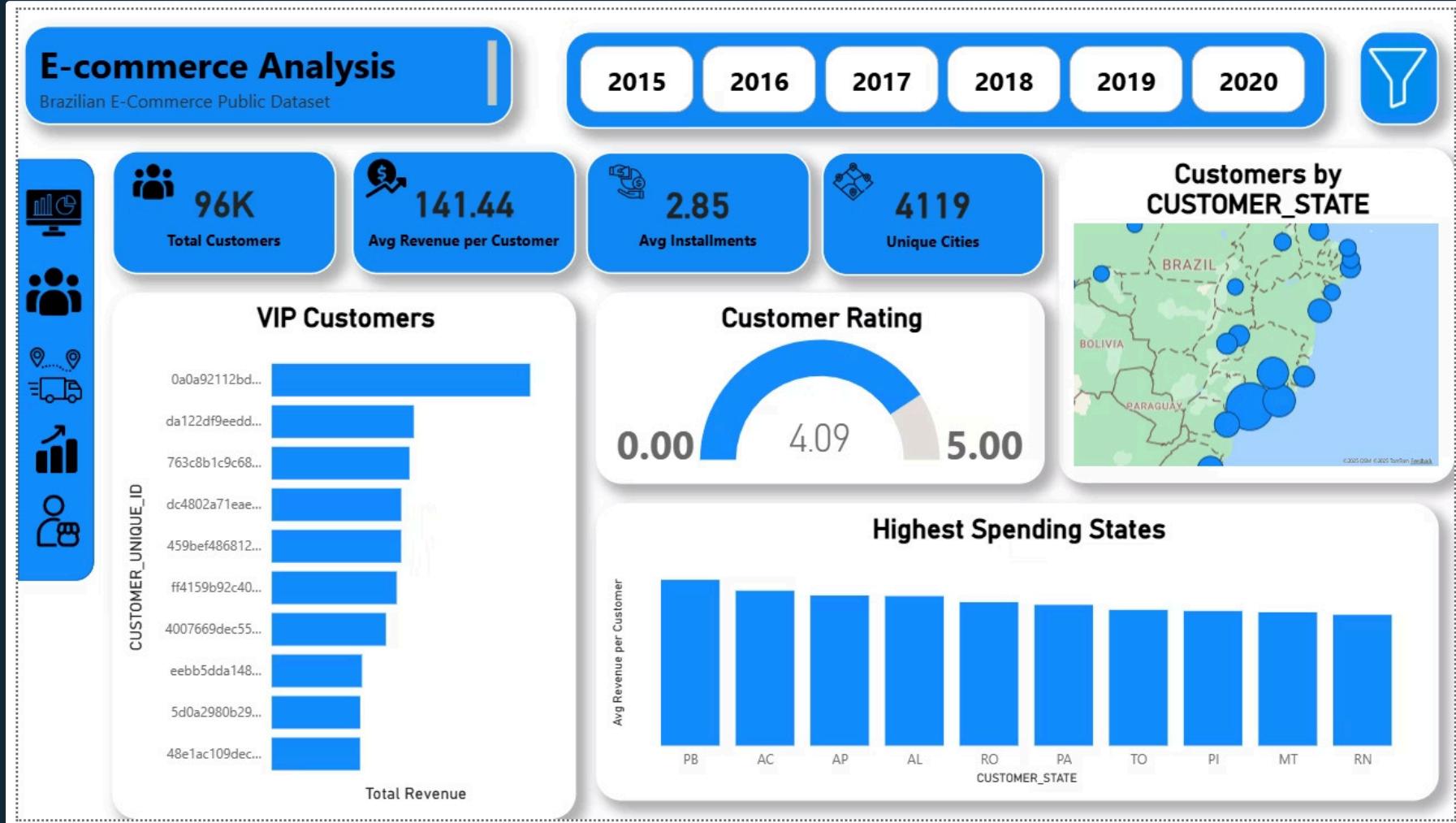


# Power BI Dashboard

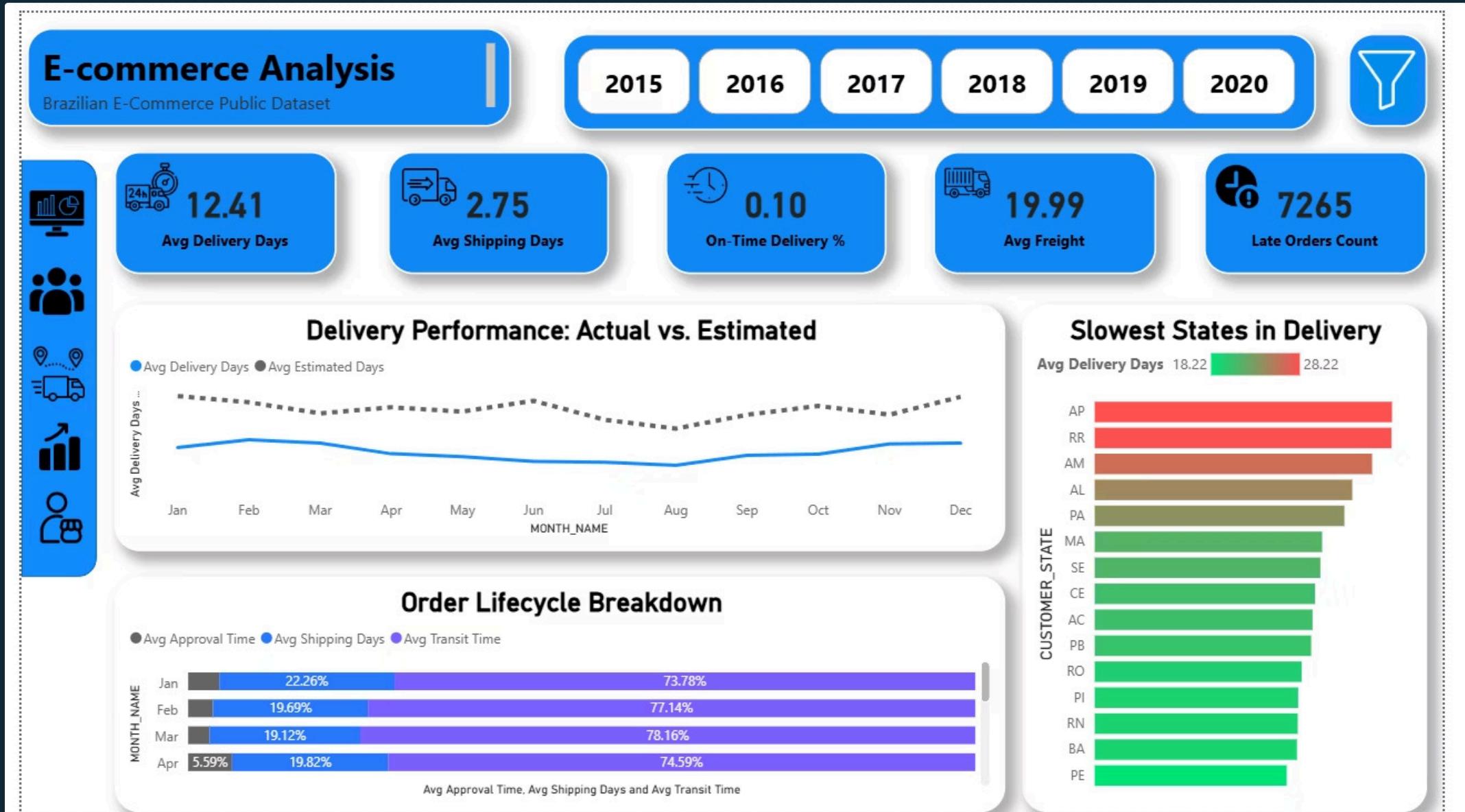
# overview page



# customers page



# Logistics page



# Sales & Products page

## E-commerce Analysis

Brazilian E-Commerce Public Dataset

2015 2016 2017 2018 2019 2020

113K Total Items Sold | \$120.65 Avg Item Price | 73 Categories

Price vs. Quantity Strategy

Total Revenue \$0.00M - \$1.26M

Avg Item Price

Revenue by CATEGORY

Total Revenue \$0.27M - \$1.26M

CATEGORY_NAME	Total Revenue	Total Items Sold	Avg
health_beauty	\$1,258,681.34	9670	
watches_gifts	\$1,205,005.68	5991	
bed_bath_table	\$1,036,988.68	11115	
sports_leisure	\$988,048.97	8641	
computers_accessories	\$911,954.32	7827	
furniture_decor	\$729,762.49	8334	
cool_stuff	\$635,290.85	3796	
housewares	\$632,248.66	6964	
auto	\$592,720.11	4235	
garden_tools	\$485,256.46	4347	
toys	\$483,946.60	4117	
baby	\$411,764.89	3065	
perfumery	\$399,124.87	3419	
telephony	\$323,667.53	4545	
office_furniture	\$273,960.70	1691	
stationery	\$230,943.23	2517	
computers	\$222,963.13	203	
pet_shop	\$214,315.41	1947	
musical_instruments	\$191,498.88	680	
small_appliances	\$190,648.58	679	
electronics	\$179,535.28	1603	
consoles_games	\$160,246.74	2767	
Total	\$13,591,643.70	112650	

Made with GAMMA

# sellers page

## E-commerce Analysis

Brazilian E-Commerce Public Dataset

2015 2016 2017 2018 2019 2020

Total Active Sellers: 3095

Avg Shipping Days: 2.75

Avg Review Score: 4.09

Seller Performance Detail

Total Revenue: \$0.00M - \$0.23M

Avg Seller Review Score: 1.0 - 5.0

Top Seller States by Revenue

Total Revenue: \$0.07M - \$8.75M

SELLER_ID	Total Revenue	Total Orders
4869f7a5dfa27	\$229,472.63	1132
7a7dca6462dcf3b52b2	\$222,776.05	358
53243585a1d6dc2643021fd1853d8905	\$200,472.92	1806
4a3ca9315b744ce9f8e9374361493884	\$194,042.03	585
fa1c13f2614d7b5c4749cbc52fecda94	\$187,923.89	982
7e93a43ef30c4f03f38b393420bc753a	\$176,431.87	336
da8622b14eb17ae2831f4ac5b9dab84a	\$160,236.57	1314
7a67c85e85bb2ce8582c35f2203ad736	\$141,745.53	1160
1025f0e2d44d7041d6cf58b6550e0bfa	\$138,968.55	915
955fee9216a6	\$135,171.70	1287
Total	\$13,591,643.70	98666

# Thank you