



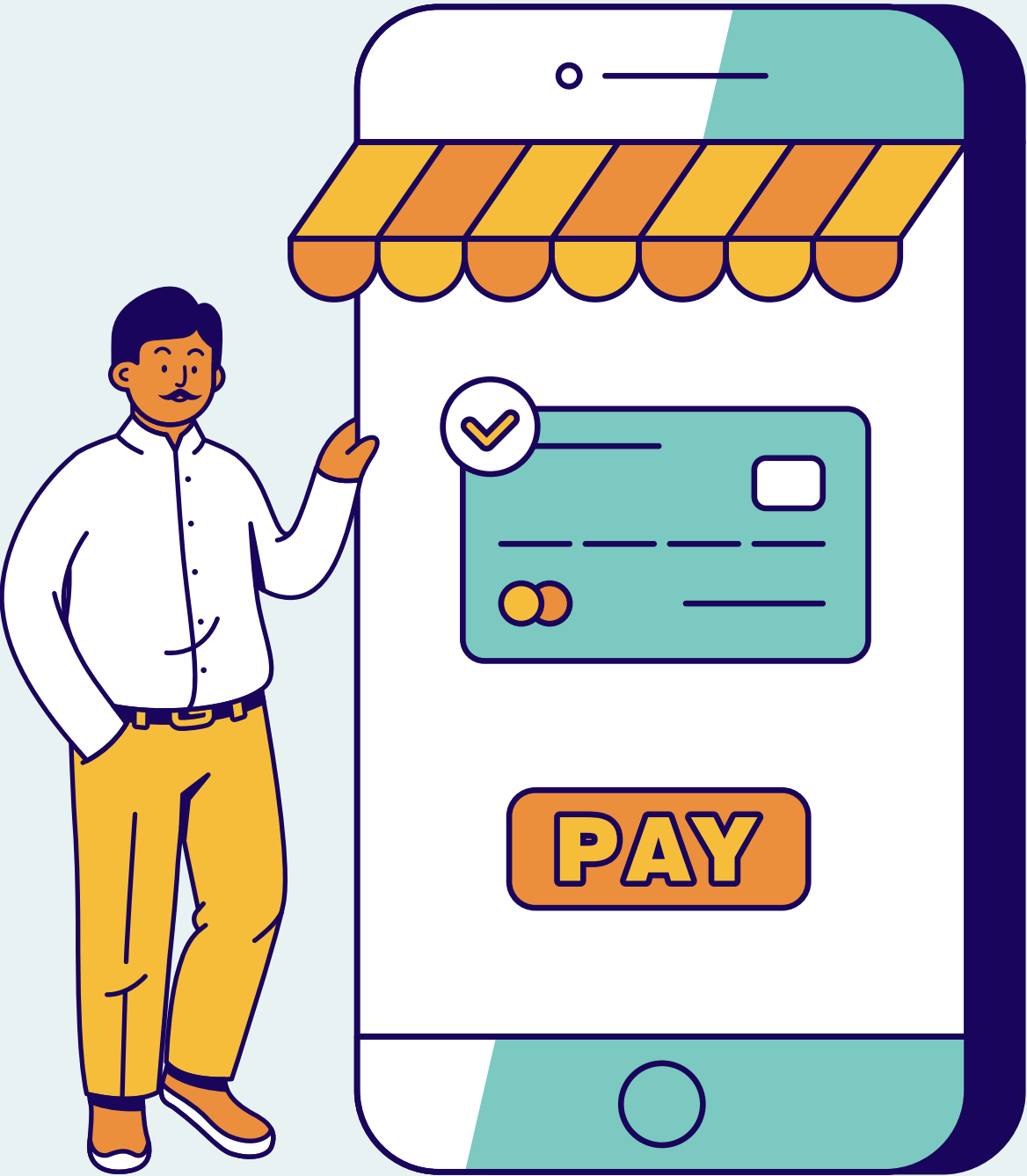
ECOMMERCE TRANSACTION MANAGEMENT APPLICATION

Information Management Project Fall'22

Group 4 - Aditya, Aishwarya, Kai, Prathmesh,
Rajshree, Yvonne

ABOUT THE DATA

A Brazilian e-commerce website (OLIST) has open sourced its dataset of 100,000 orders with product, customer and reviews information.



TABLES

Order Information	Order Status Details	Payment Information	Order Reviews
Customer Information	Seller Details	Product Categories	Geolocation

AGENDA

01

DATA STRATEGY

Current Data strategy

02

TRANSACTION MANAGEMENT APPLICATIONS

Setup of transactional databases, Data Models, ER Diagrams

03

DATA WAREHOUSE

Setup of data warehouse and different tables associated to it

04

NOSQL DATA LAKE

Setup of data lake, Elastic Search and Kibana, Denormalization, Piping Data

05

ANALYTICS AND INSIGHTS

Analyzing the data using a reporting layer on top of the data warehouse and NO SQL Datalake

06

REFLECTIONS

Learnings, key takeaways and conclusion



01

DATA STRATEGY

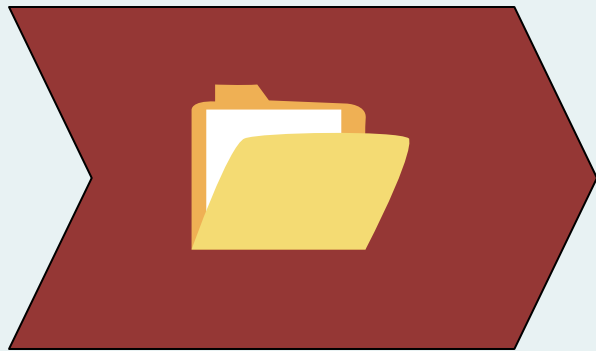
- Follows an offensive strategy
- Less regulated
- Limited sensitive personal data
- Need to react to competition and market changes at a faster pace



DATA PIPELINE

1.

Raw data flat files



The 9 raw data files are consolidated

2.

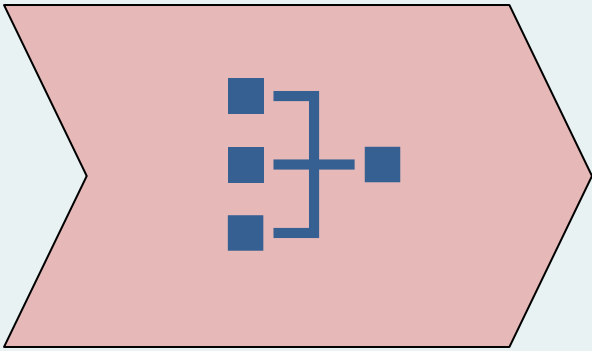
Transaction Management Applications



Development of the transaction management application for order processing, payment processing and buyer-seller information system

3.

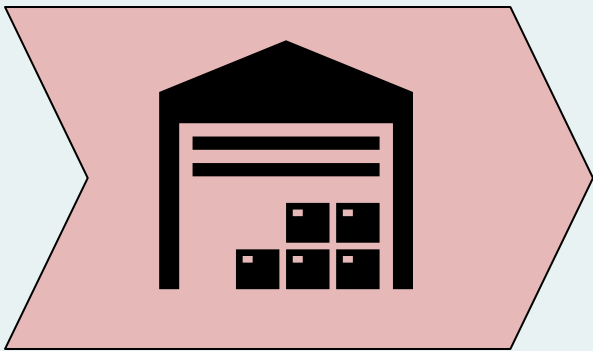
ETL



ETL process developed to consolidate data from multiple database systems into the warehouse

4.

Data Warehouse and analysis



Denormalized data is stored in the data warehouse

5.

Analysis on data lake



Kibana makes POST and GET requests to Elastic Search backend and populates the visualizations

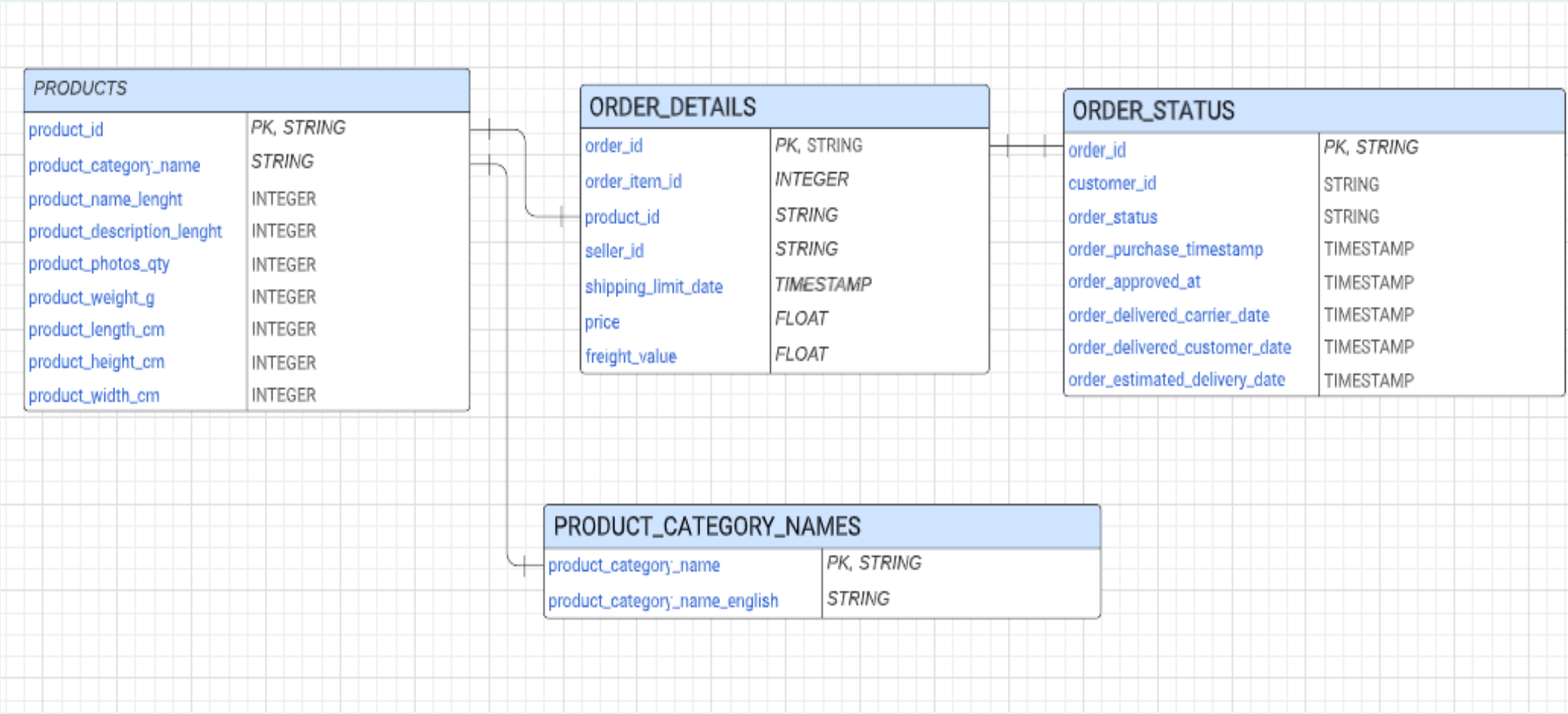
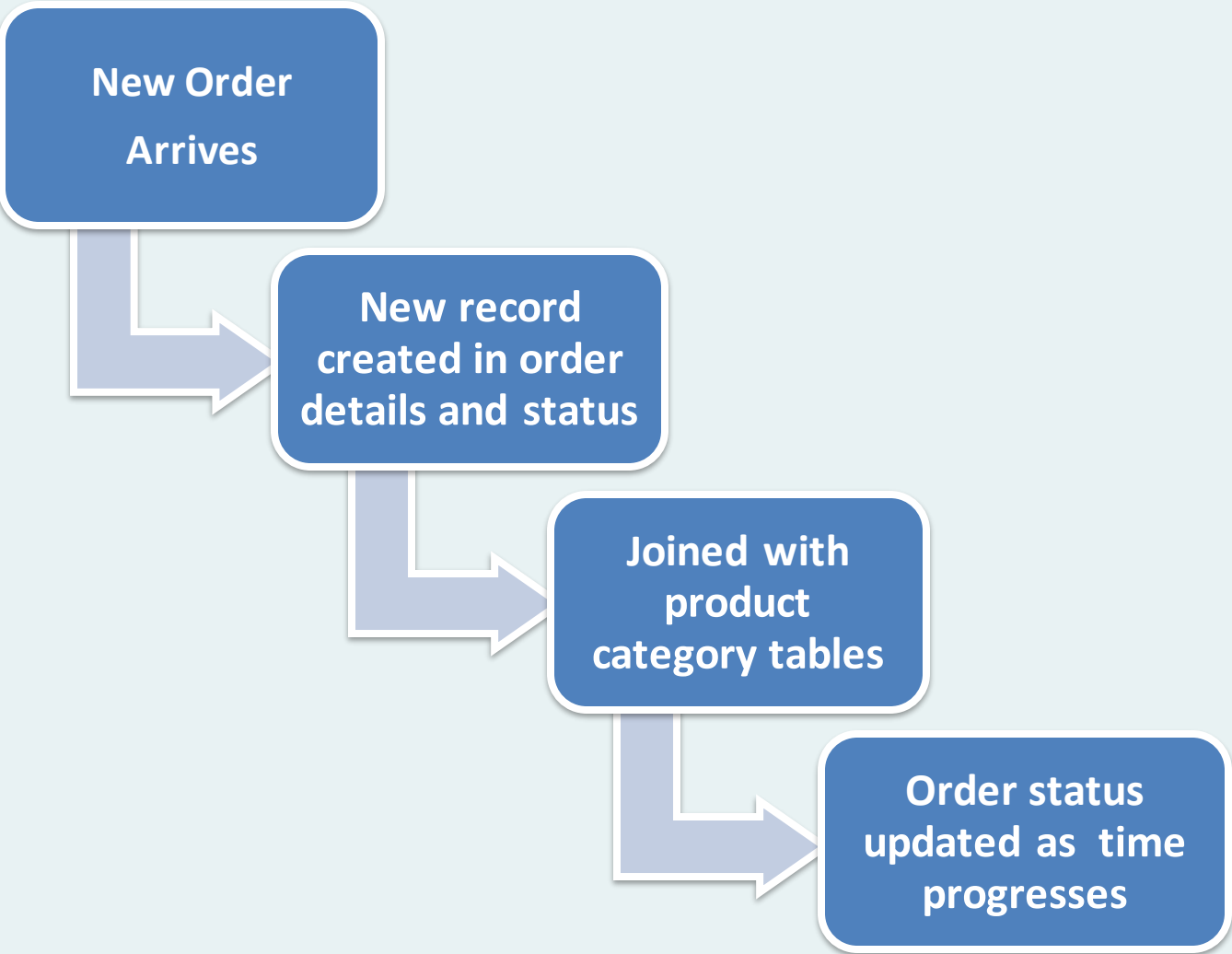
02

TRANSACTION MANAGEMENT APPLICATION



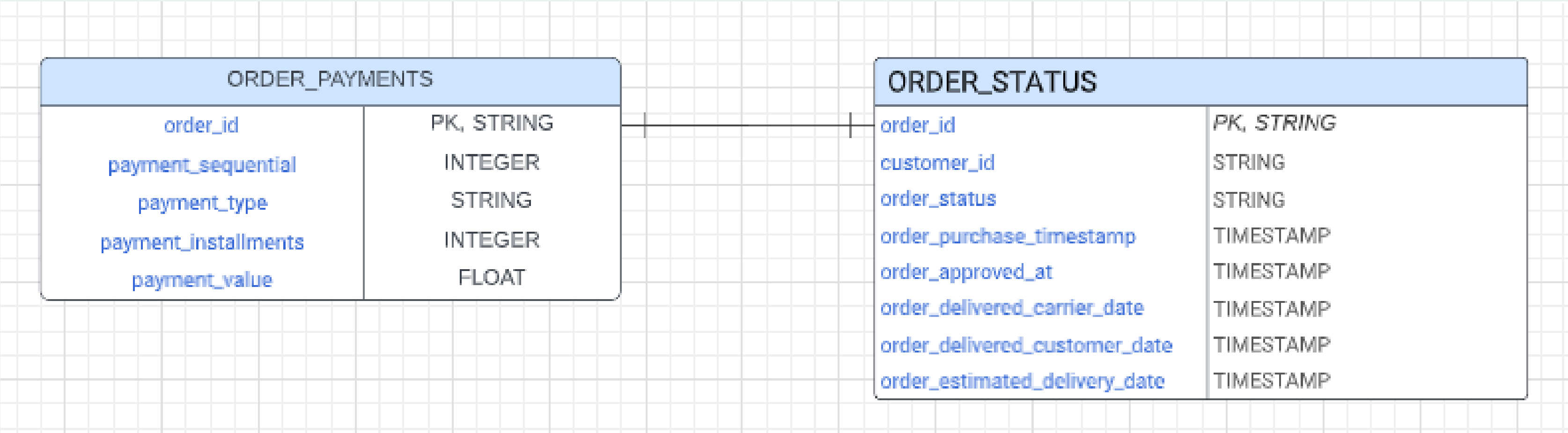
1. ORDER PROCESSING SYSTEM

- Maintains the integrity of order entries and updates across the database
- Check the number of orders placed against every product category for the purpose of optimal inventory management

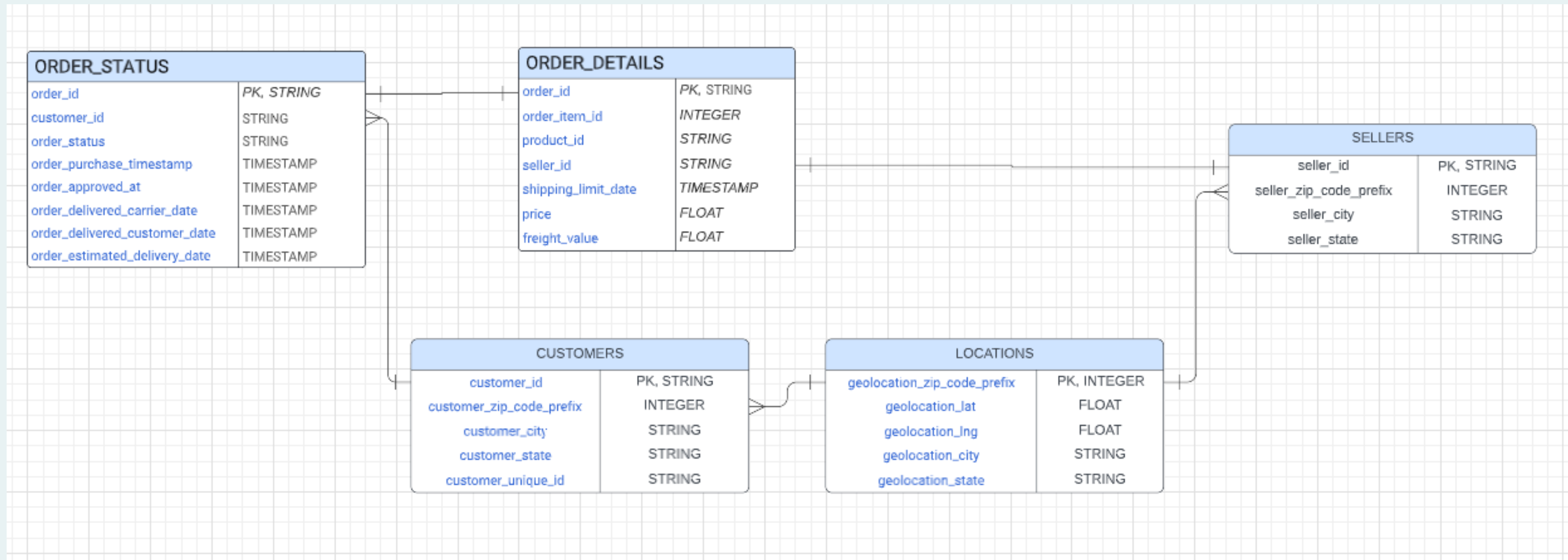


2. PAYMENT PROCESSING SYSTEM

- Order payments is updated with the payment details whenever a new order is placed and processed



3. Customer and Seller Information Management System

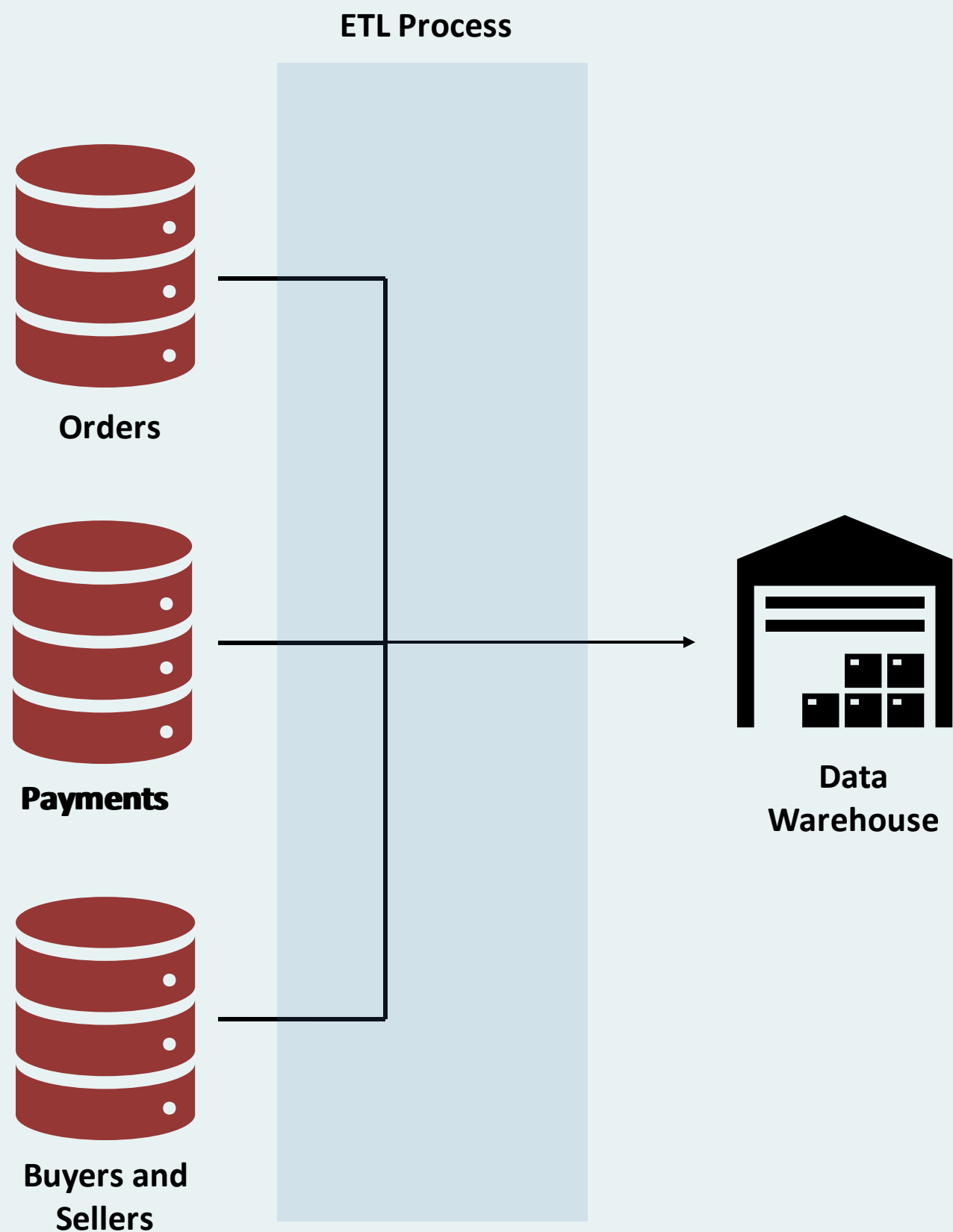


03

DATA WAREHOUSE



Building the Data Warehouse



ORDER_INFORMATION	
order_id	PK, STRING
customer_id	STRING
order_status	STRING
order_purchase_timestamp	TIMESTAMP
order_approved_at	TIMESTAMP
order_delivered_carrier_date	TIMESTAMP
order_delivered_customer_date	TIMESTAMP
order_estimated_delivery_date	TIMESTAMP
product_id	STRING
seller_id	STRING
customer_unique_id	STRING
price	FLOAT
freight_value	FLOAT
review_score	STRING
product_category	STRING
product_weight	INTEGER
product_length	INTEGER
product_width	INTEGER
payment_type	STRING
payment_value	FLOAT
payment_installments	INTEGER

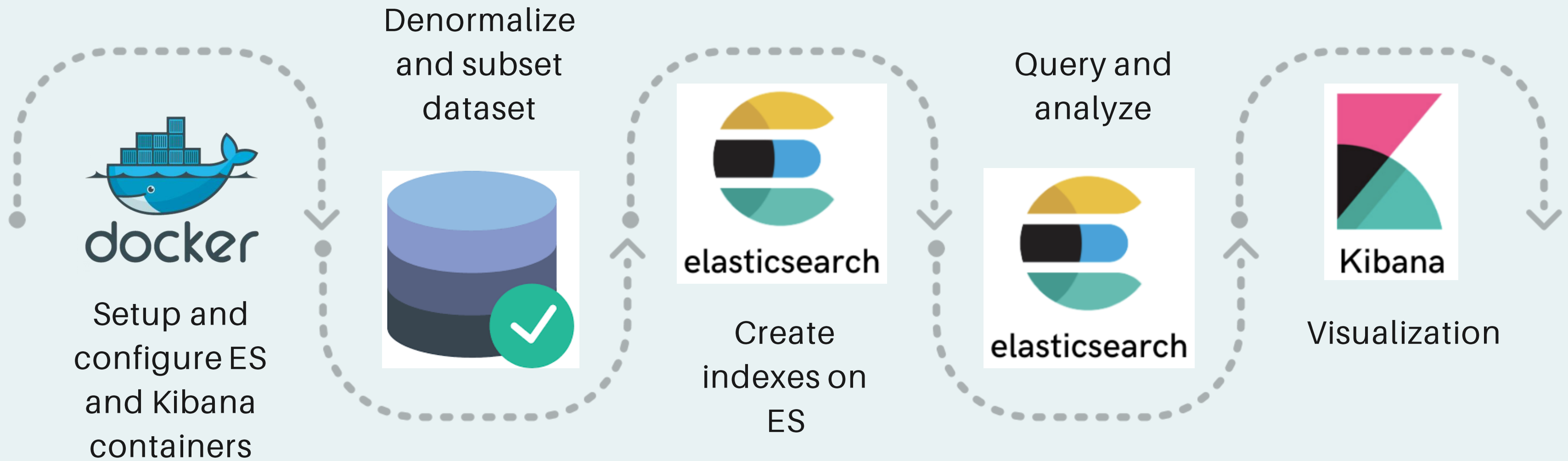
BUYER_SELLER_LOCATION	
unique_id	PK, STRING
city	INTEGER
zipCode	INTEGER
state	STRING
is_buyer_or_seller	STRING

04

NOSQL DATA LAKE



NO SQL DATA LAKE SETUP



05

ANALYTICS & INSIGHTS



PAYMENTS INSIGHTS

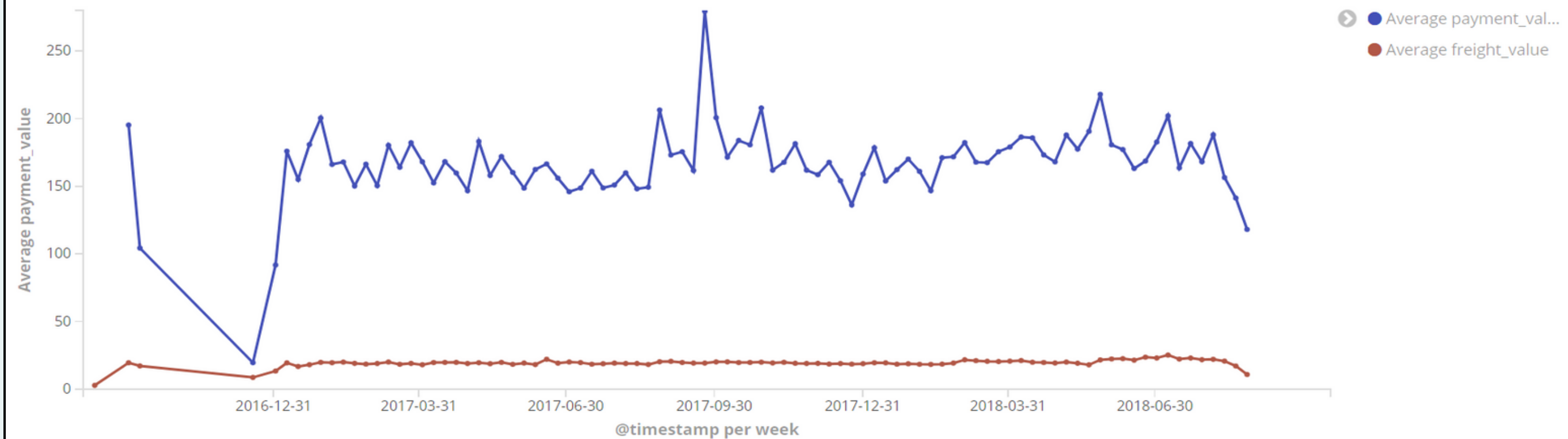
Overall Mean Payment Value

171.257
Average payment_value

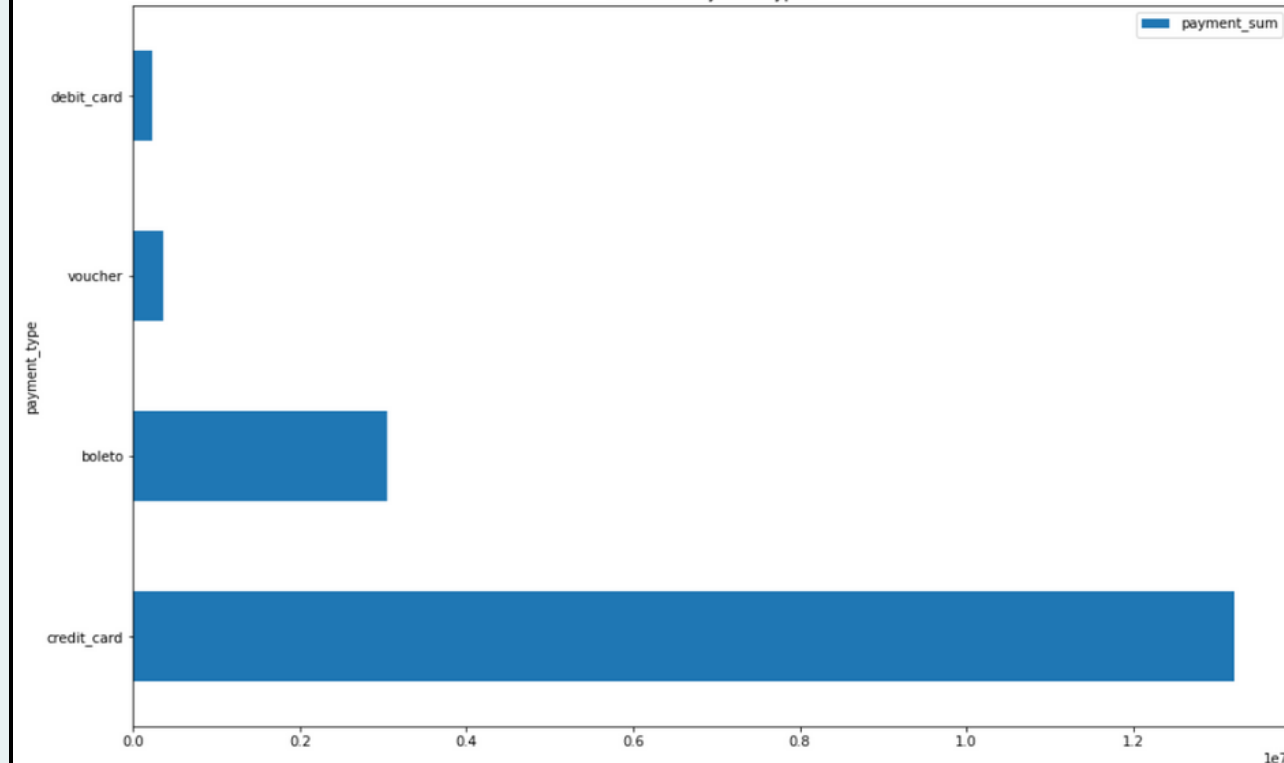
Overall Mean Freight Value

19.974
Average freight_value

Average Weekly Payment and Freight Value



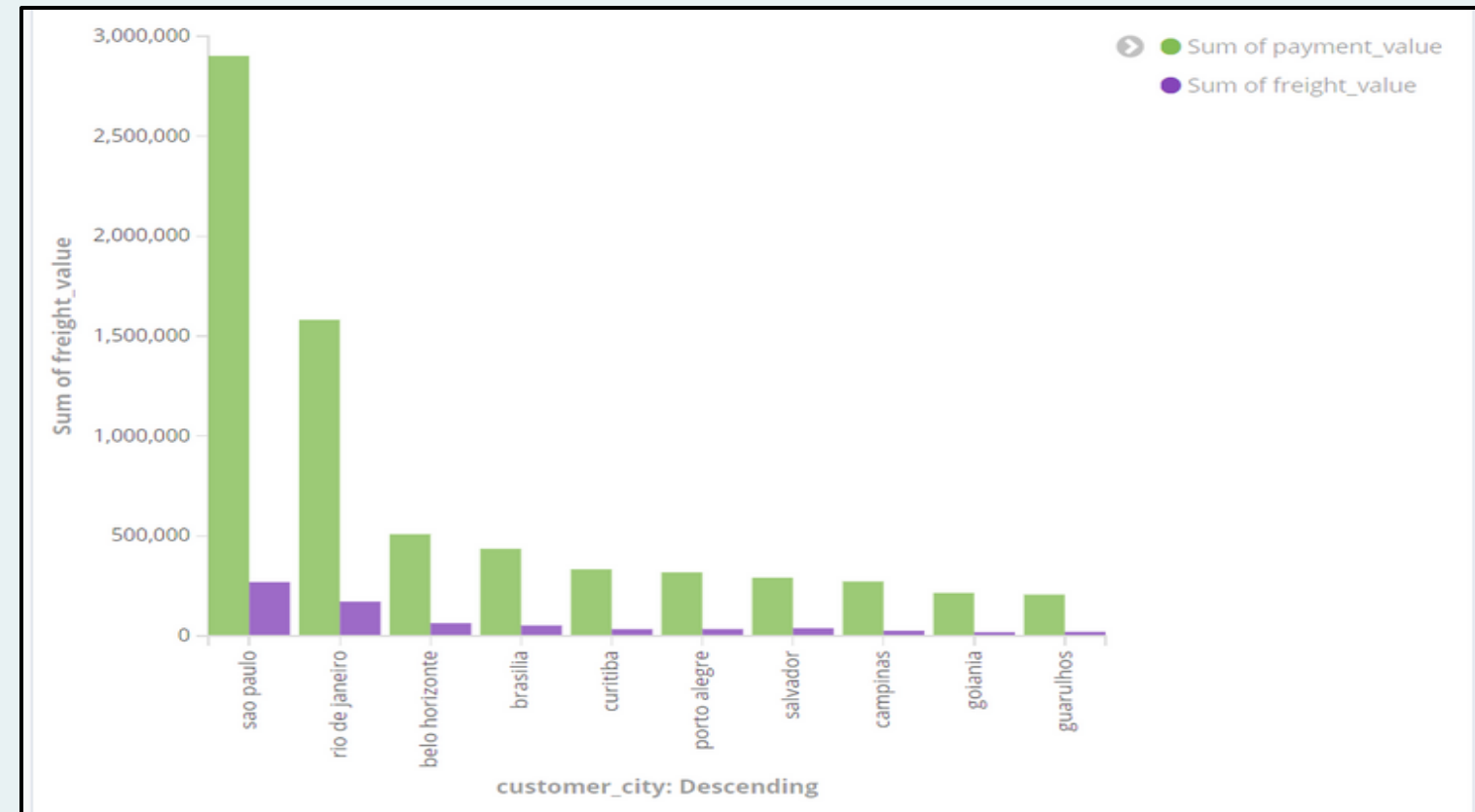
Different Payment Types



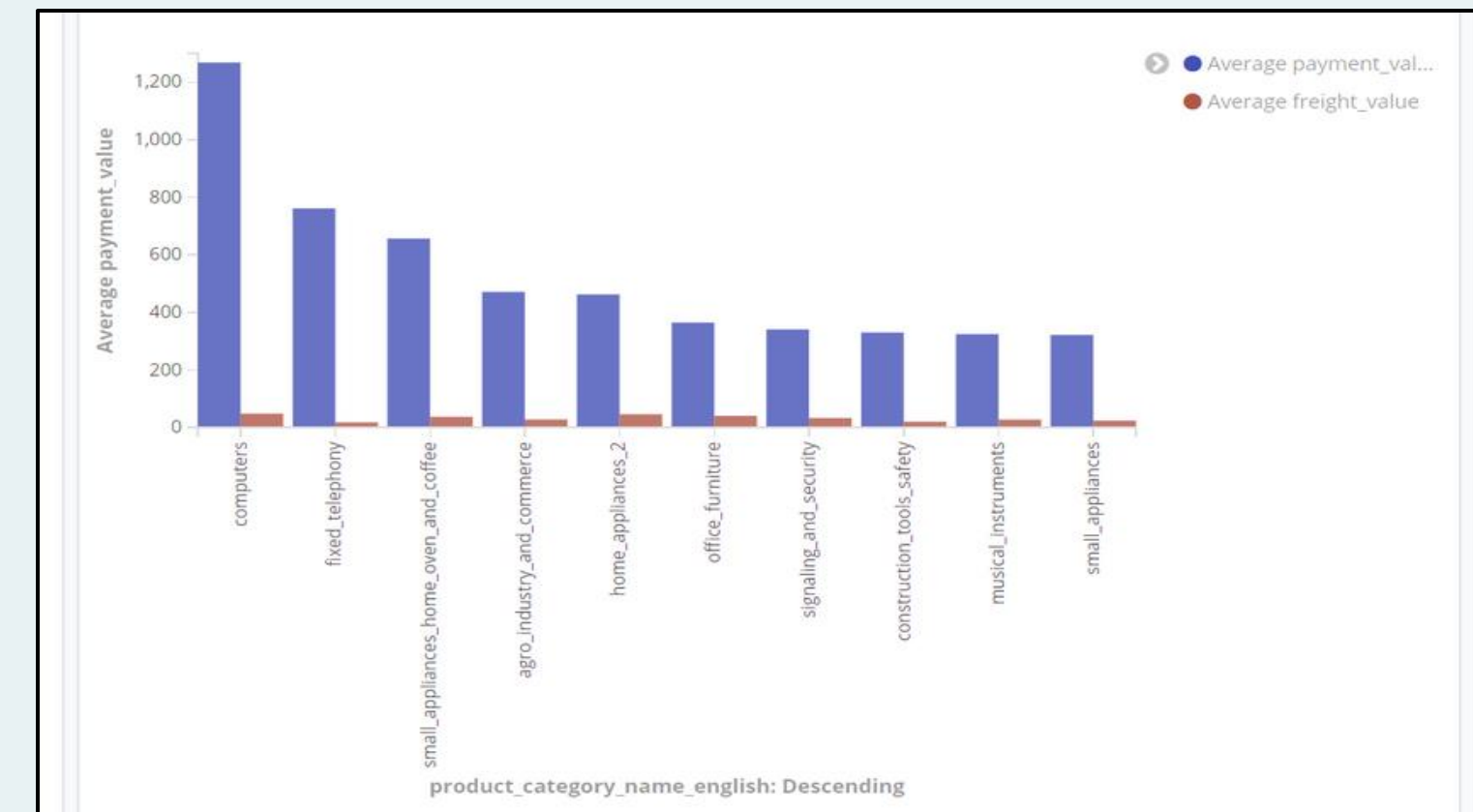
- The average payment value was 171.26 real, the mean freight value was about 20 real.
- The most popular payment method is credit card
- The payment and freight value trends are stable

PAYMENTS INSIGHTS

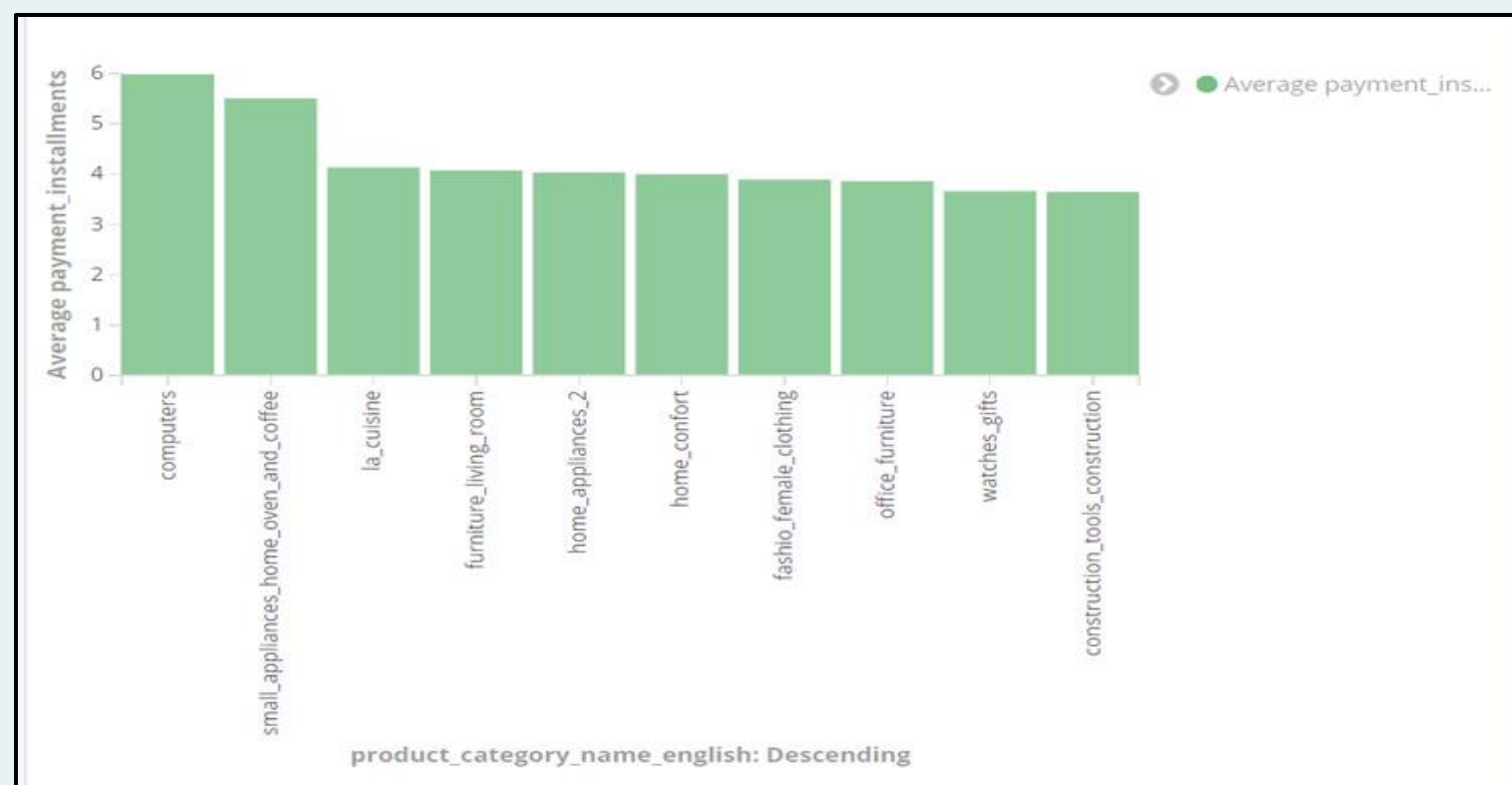
Sales and Freight Value across Cities



Average payment value across product categories



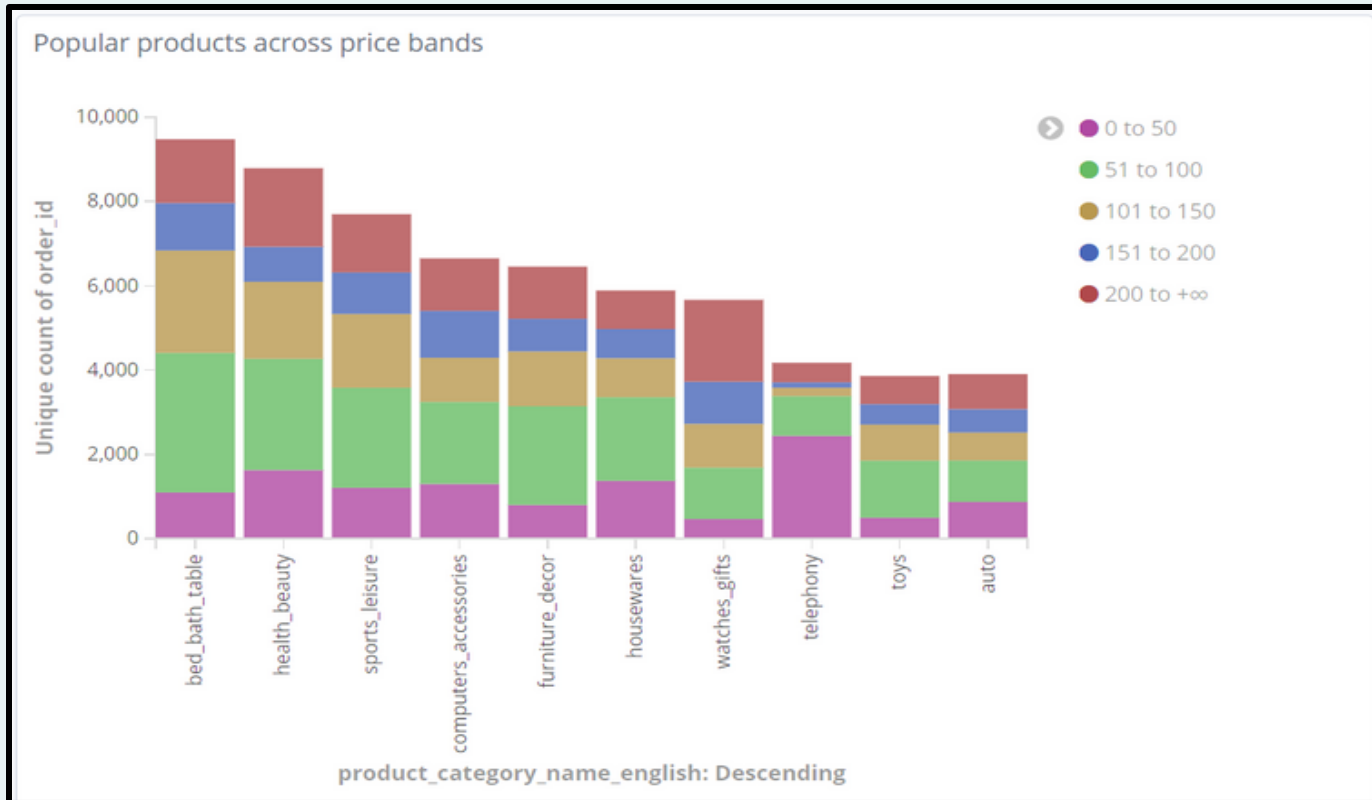
Average number of installments across product categories



- Sao Paulo and Rio De Janeiro are the business hotspots
- Freight value is proportional to the item type – bigger items (in terms of volume) have higher freight value
- Computers are the most expensive commodity sold on the website and its' payment has the highest average payment installment as well

PRODUCT CATEGORY AND ORDERS ANALYSIS

Popular products across price bands

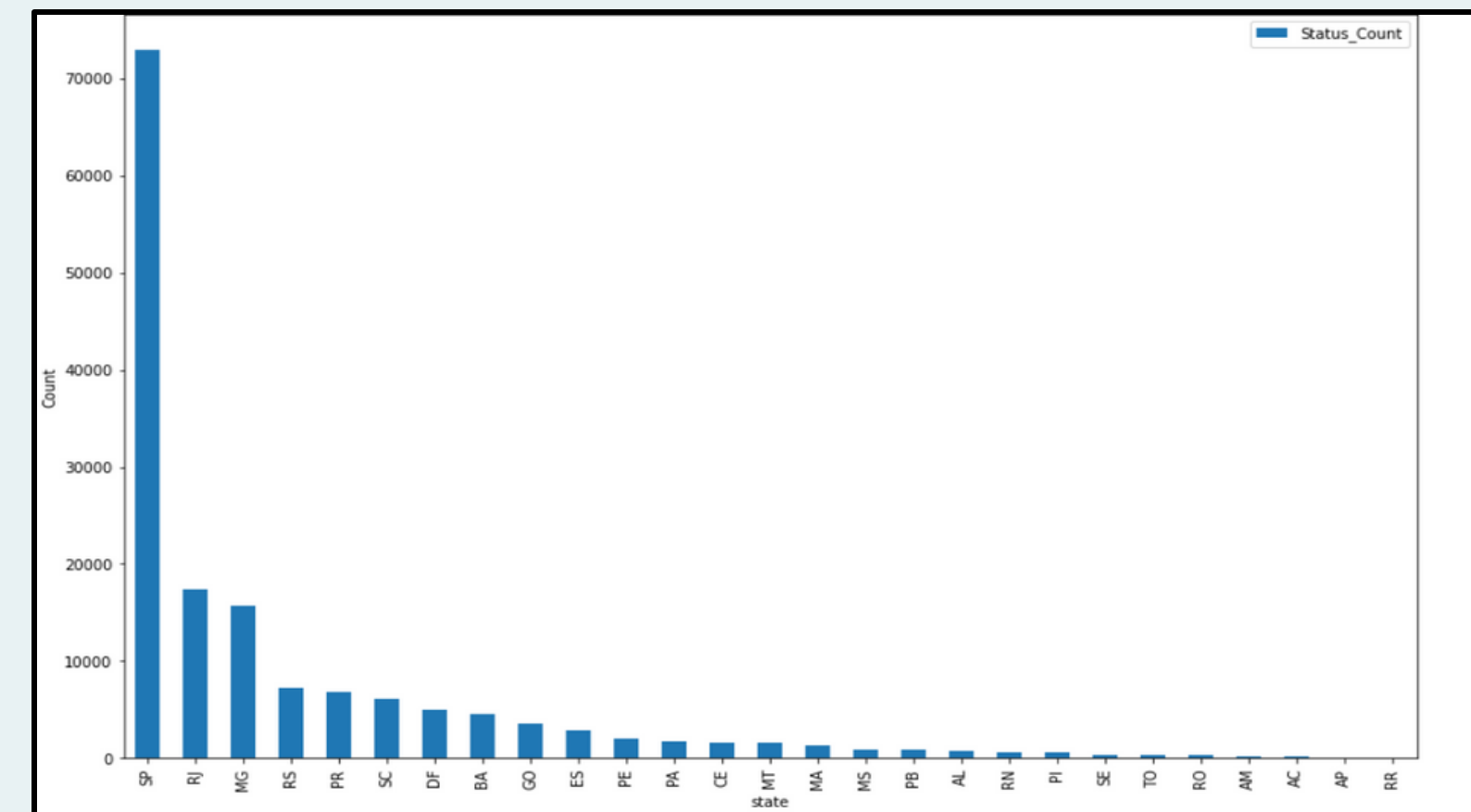


Low Sales Products across Price Bands

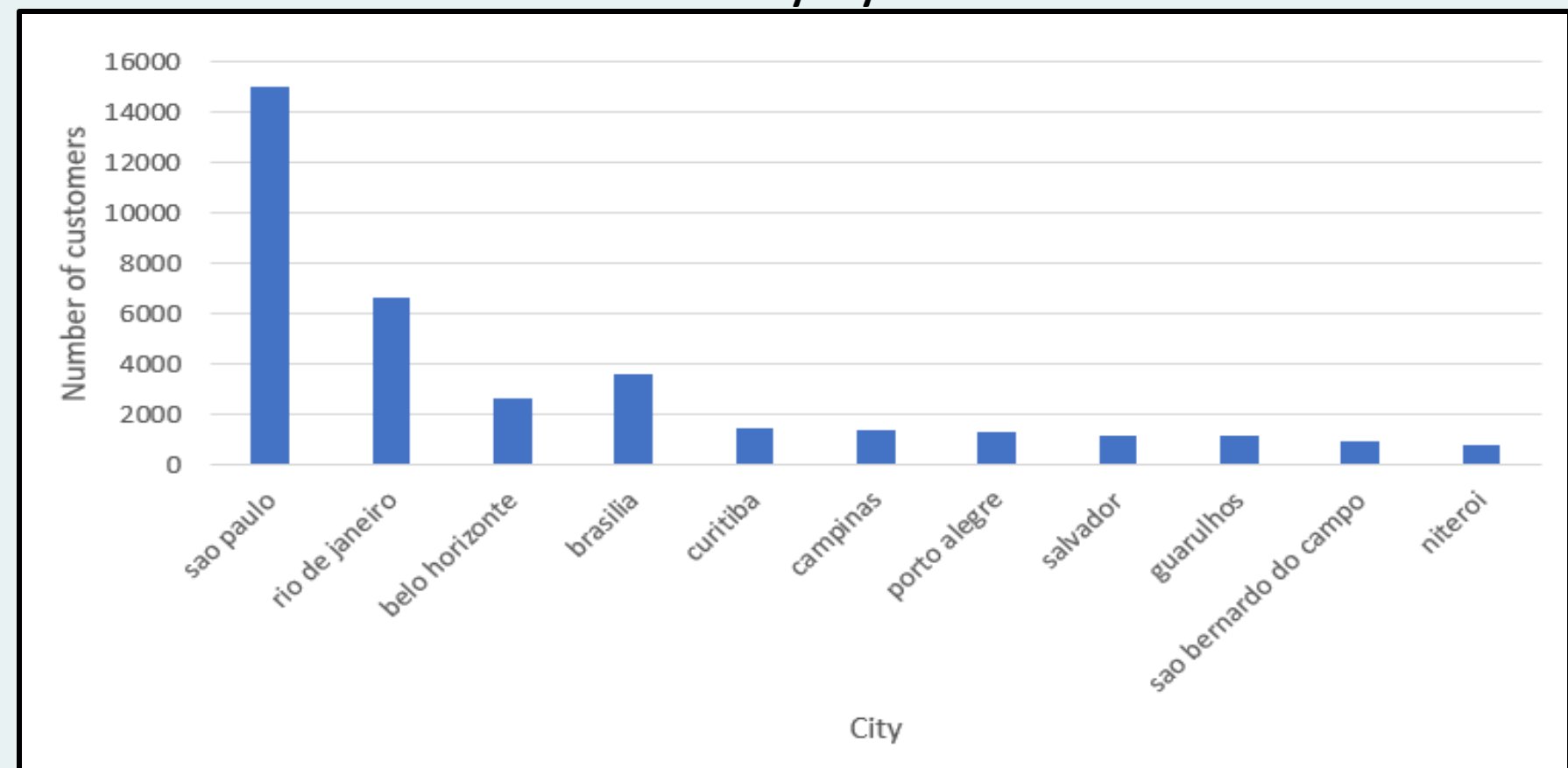


- Bed bath table, health and beauty are the most popular categories
- Watches and gifts have the higher price band products
- Children and women's clothes show lower sales volume with less price bands

Delivered Orders by State

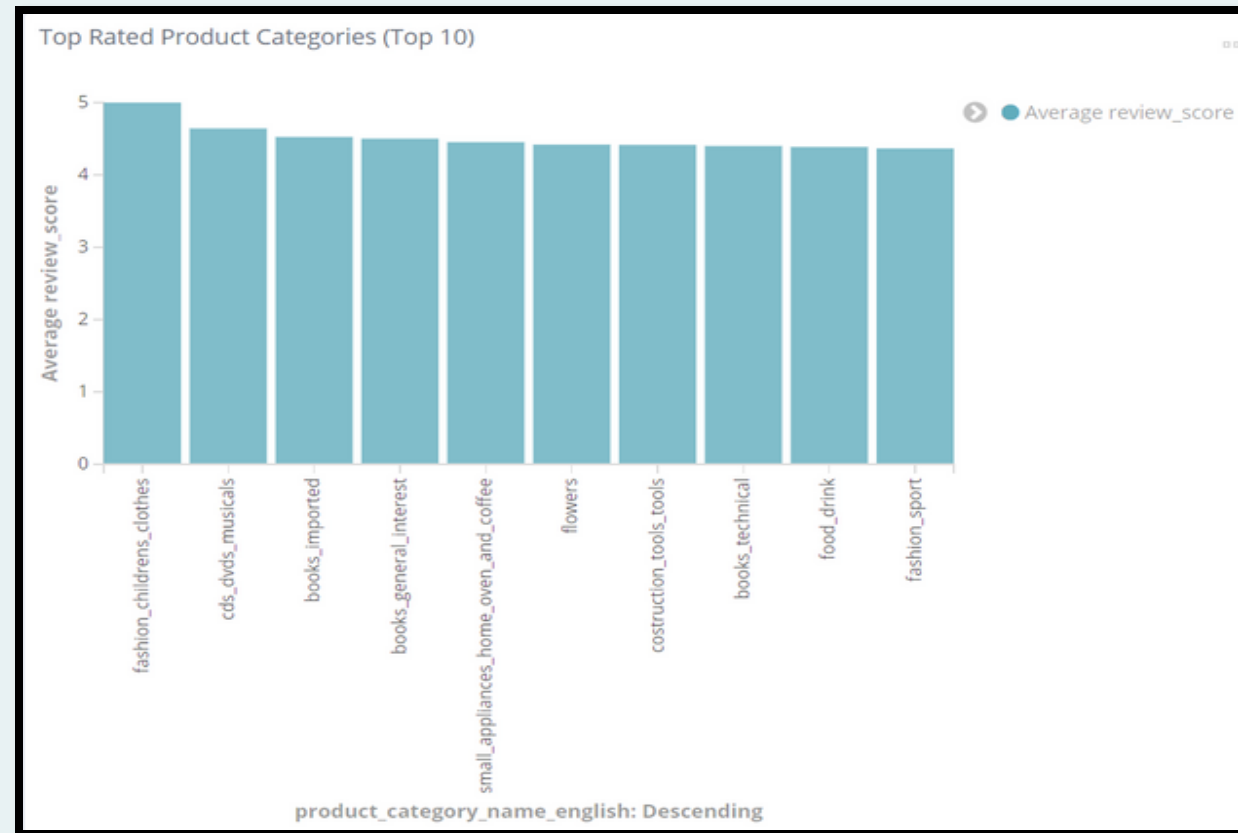


Number of Customers by city

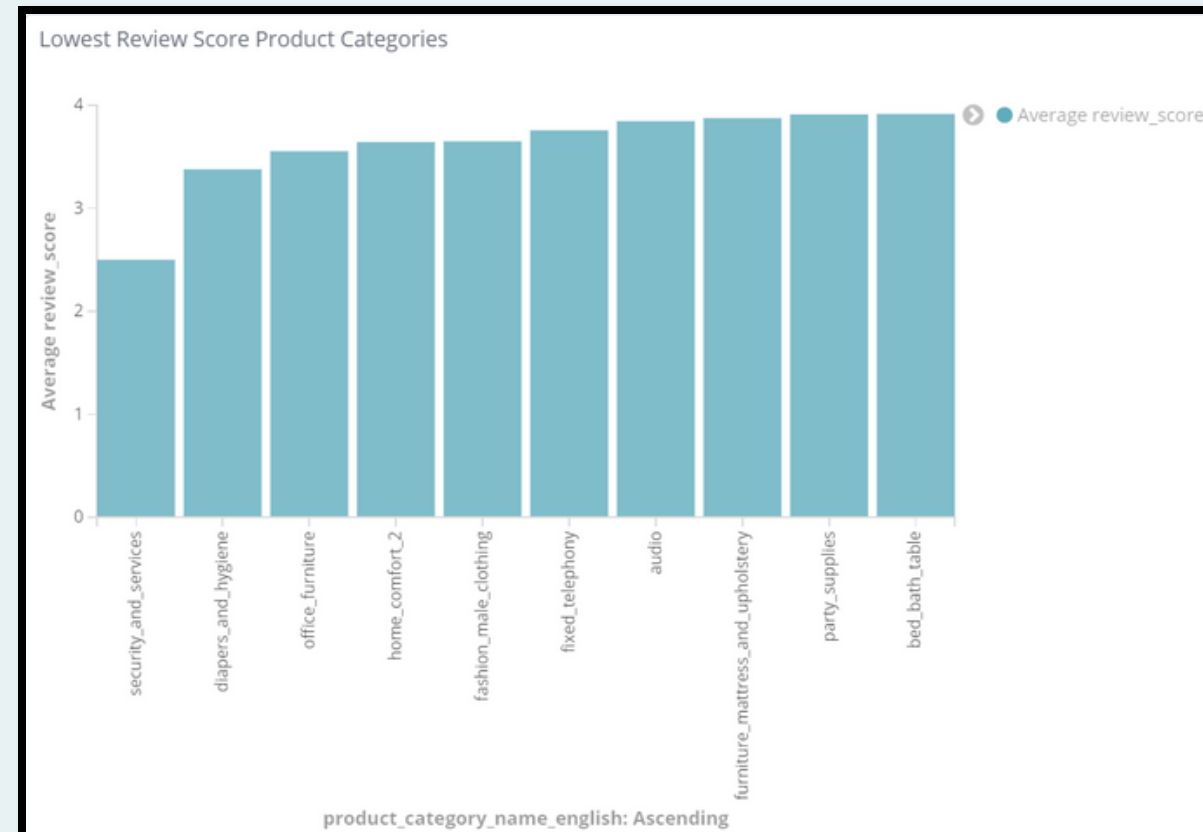


PRODUCT CATEGORY ANALYSIS

Top Rated Product Categories

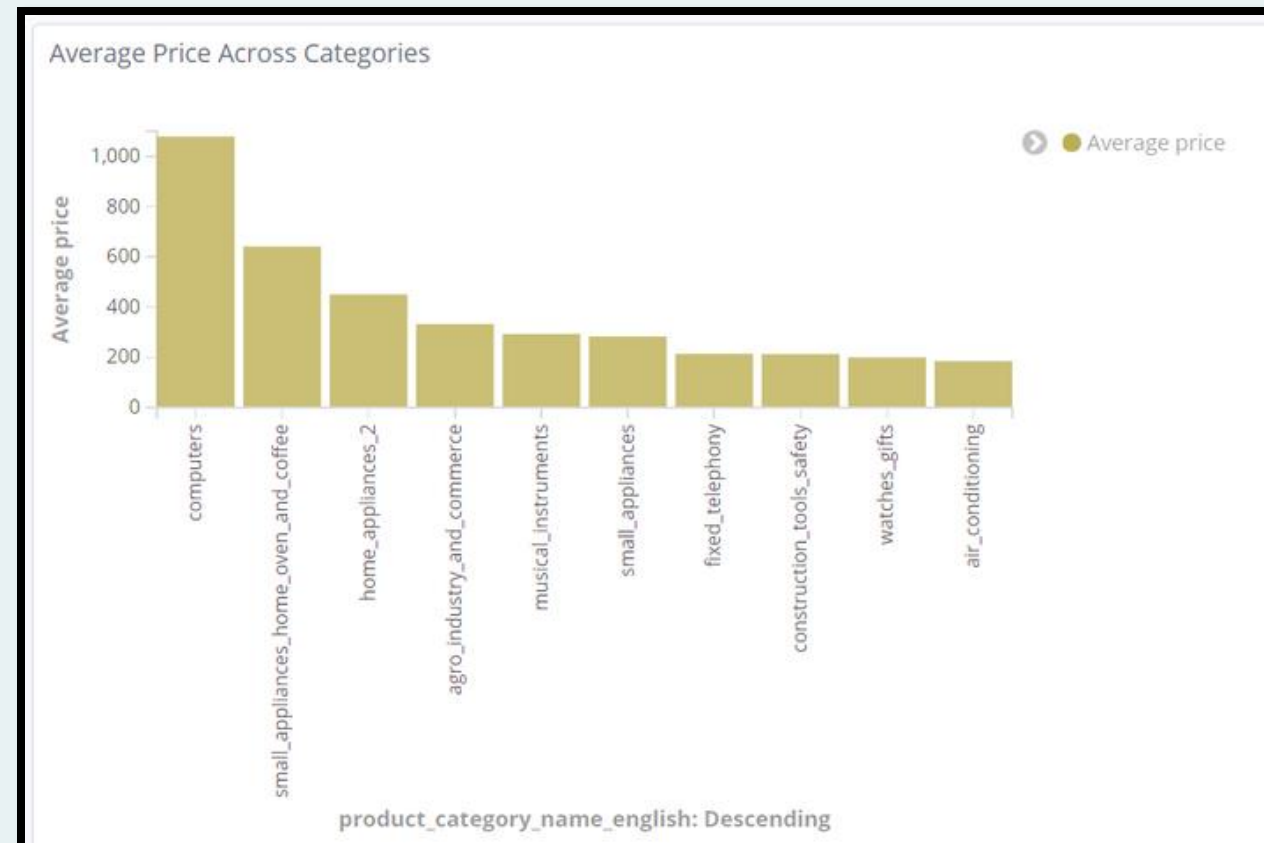


Lowest Review Score Product Categories

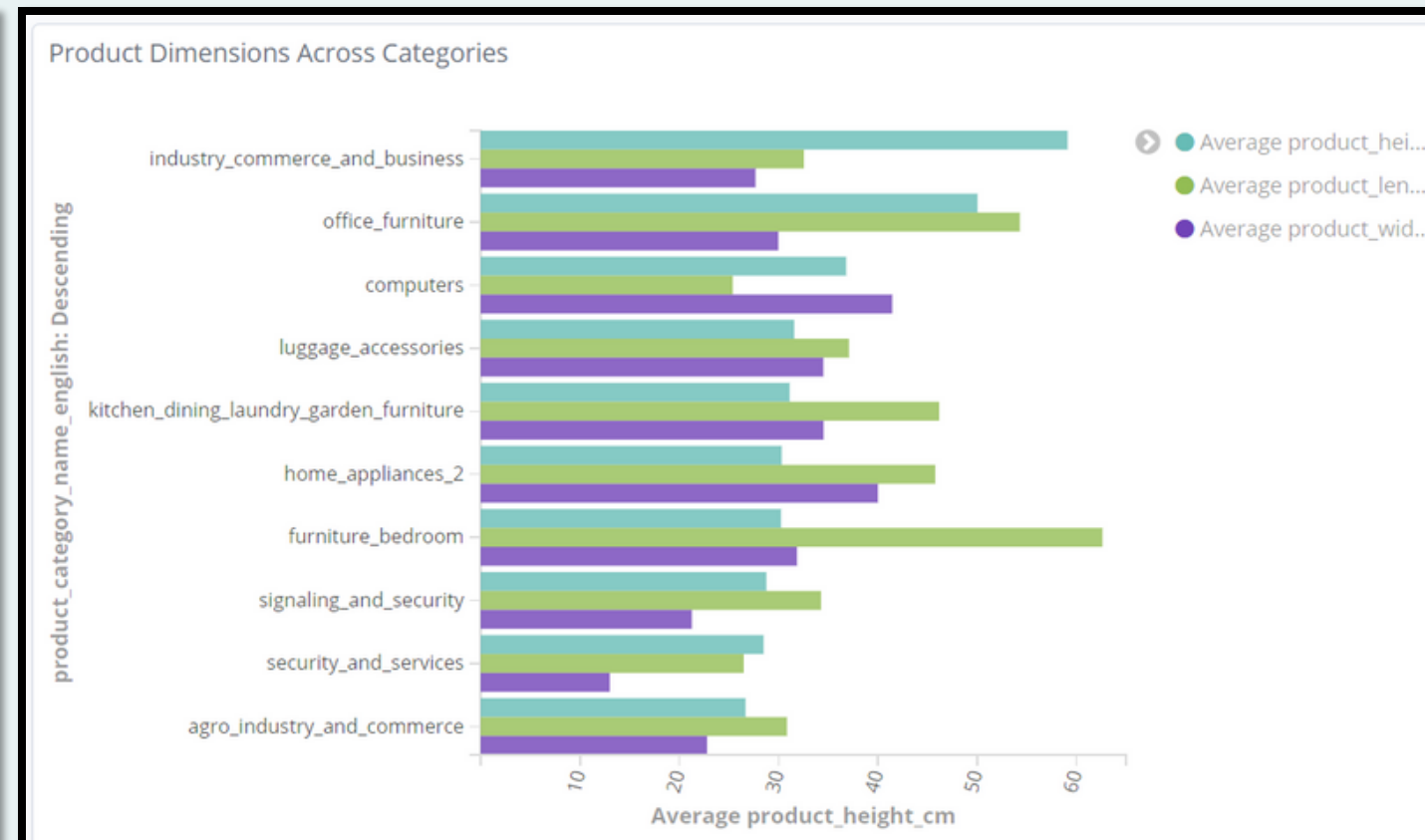


- Our top-rated product category is children clothing, whereas the lowest rated was security equipment
- Computers and small home appliances are the costliest categories

Average Price Across Categories



Product Dimensions Across Categories



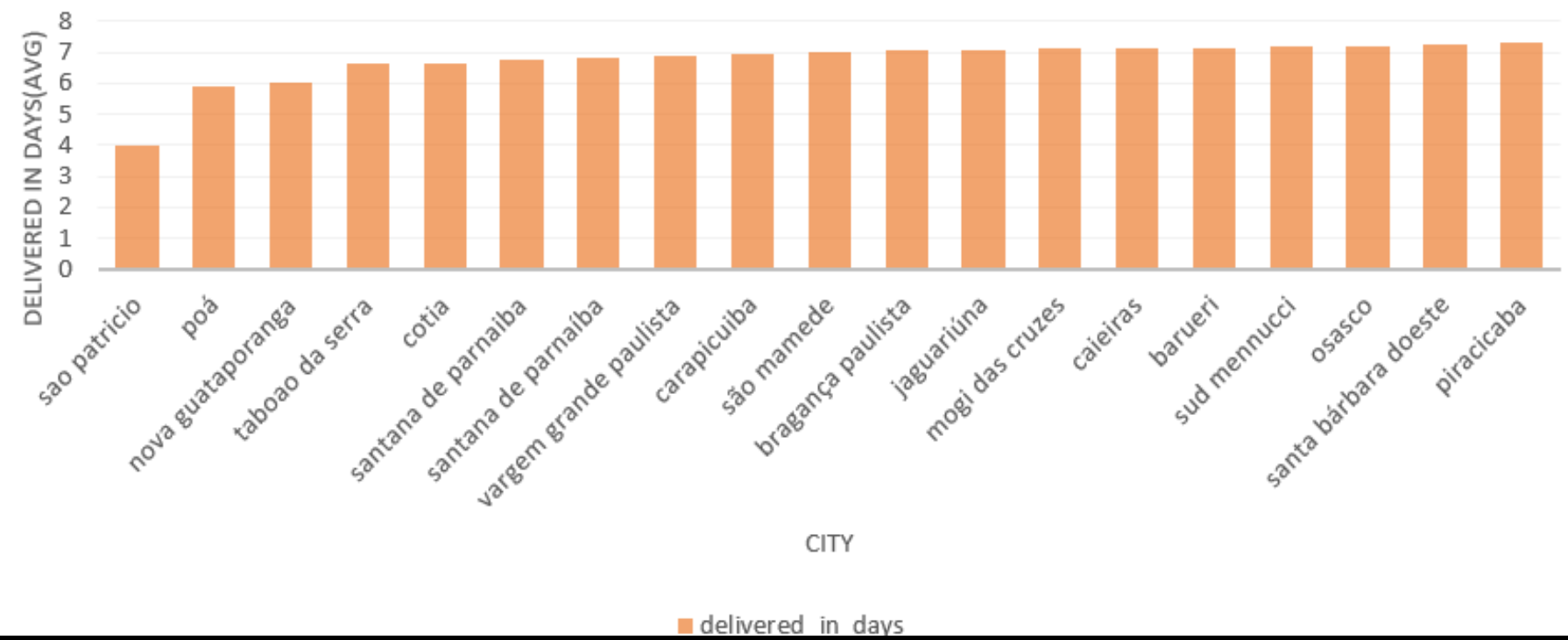
PRODUCT DELIVERY ANALYSIS

Delivery Time vs Review Score

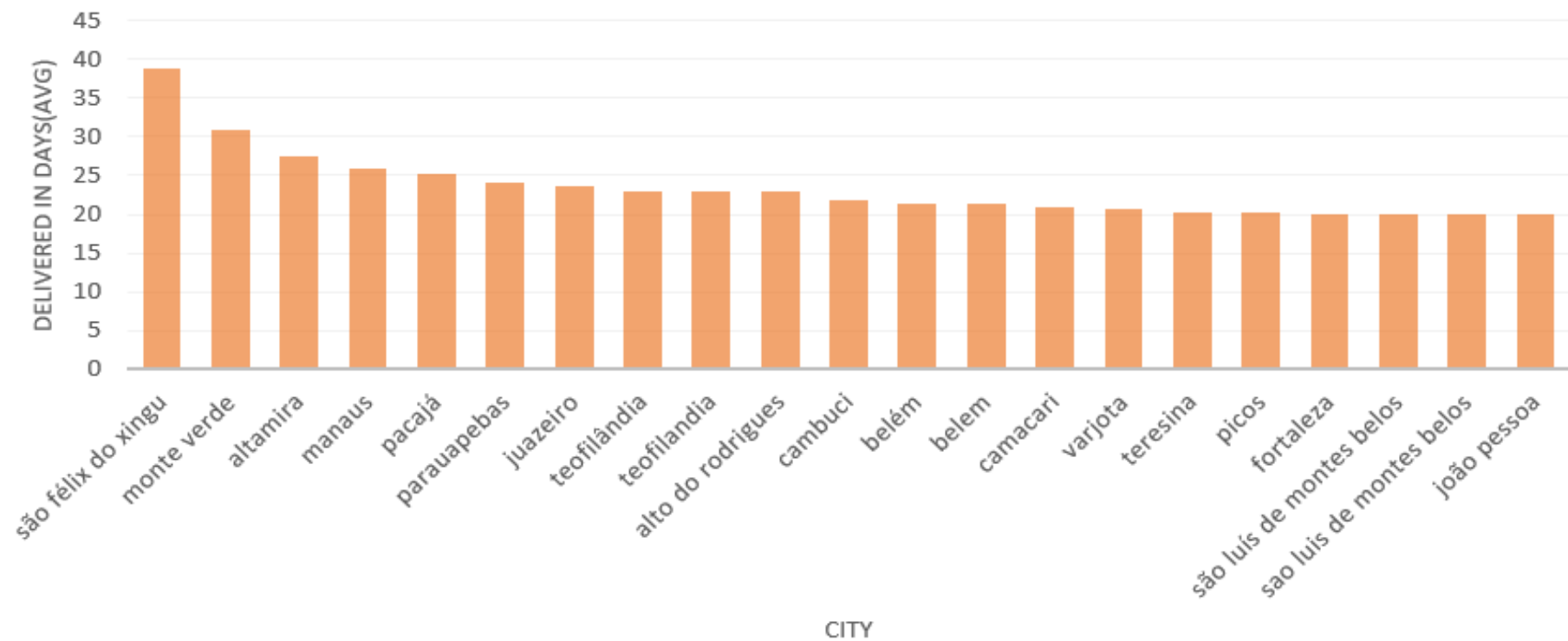


- Review score decreases with increased delivery time
- Quickest delivery for our website is in Sao Patricio
- Most delayed deliveries in Sao Felix

Quickest deliveries in Cities



Most delayed deliveries in Cities



06

REFLECTIONS



REFLECTIONS & FUTURE WORK

LEARNINGS



- Building a reporting layer on top of the data warehouse layer
- High Data integrity in large organizations

FUTURE WORK

