



olist

How Olist can recover from its losses and manage its inventory efficiently...

a Marketing and Retail Analytics Project by Abhinav Kumar



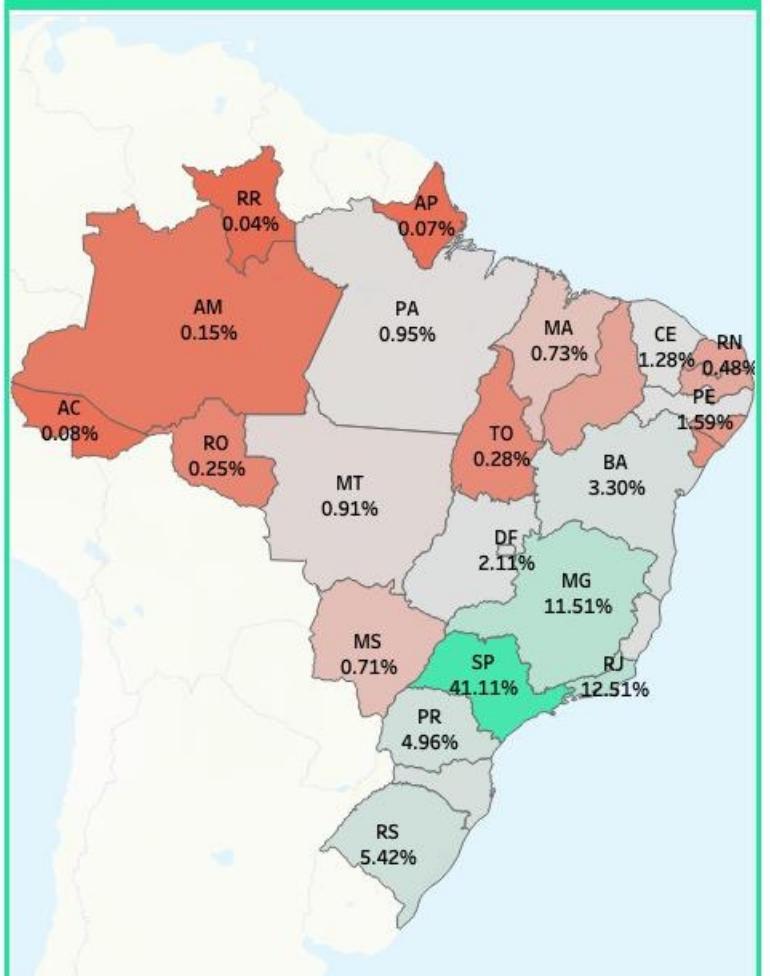
Brazilians
Love
Toys!

Share of Orders across Brazil

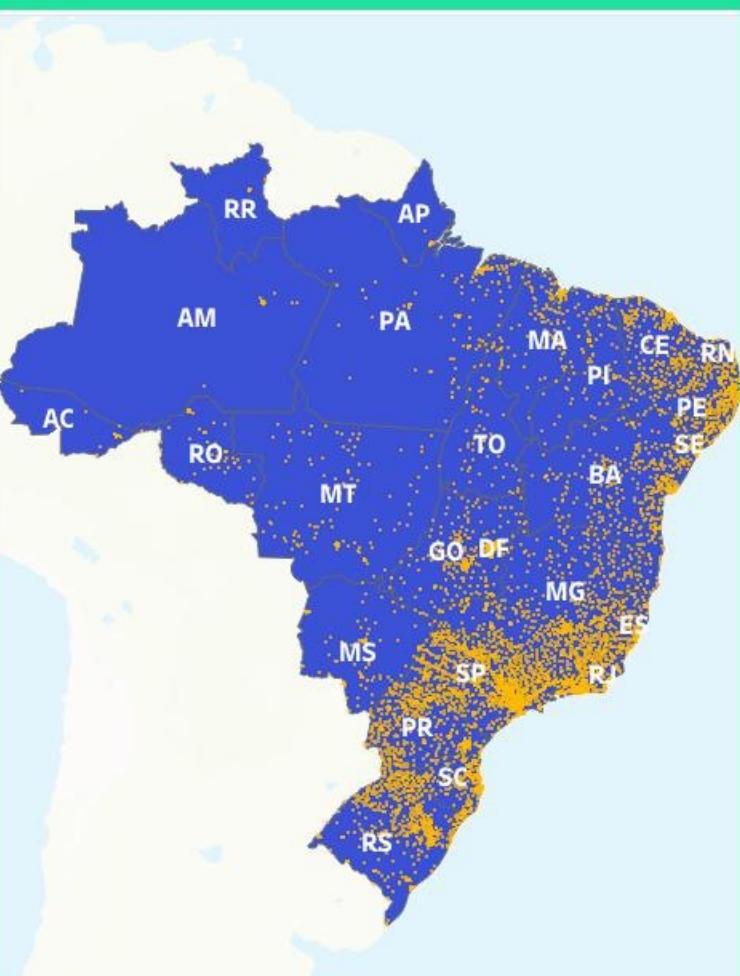


olist

olist Share of Orders across Brazil



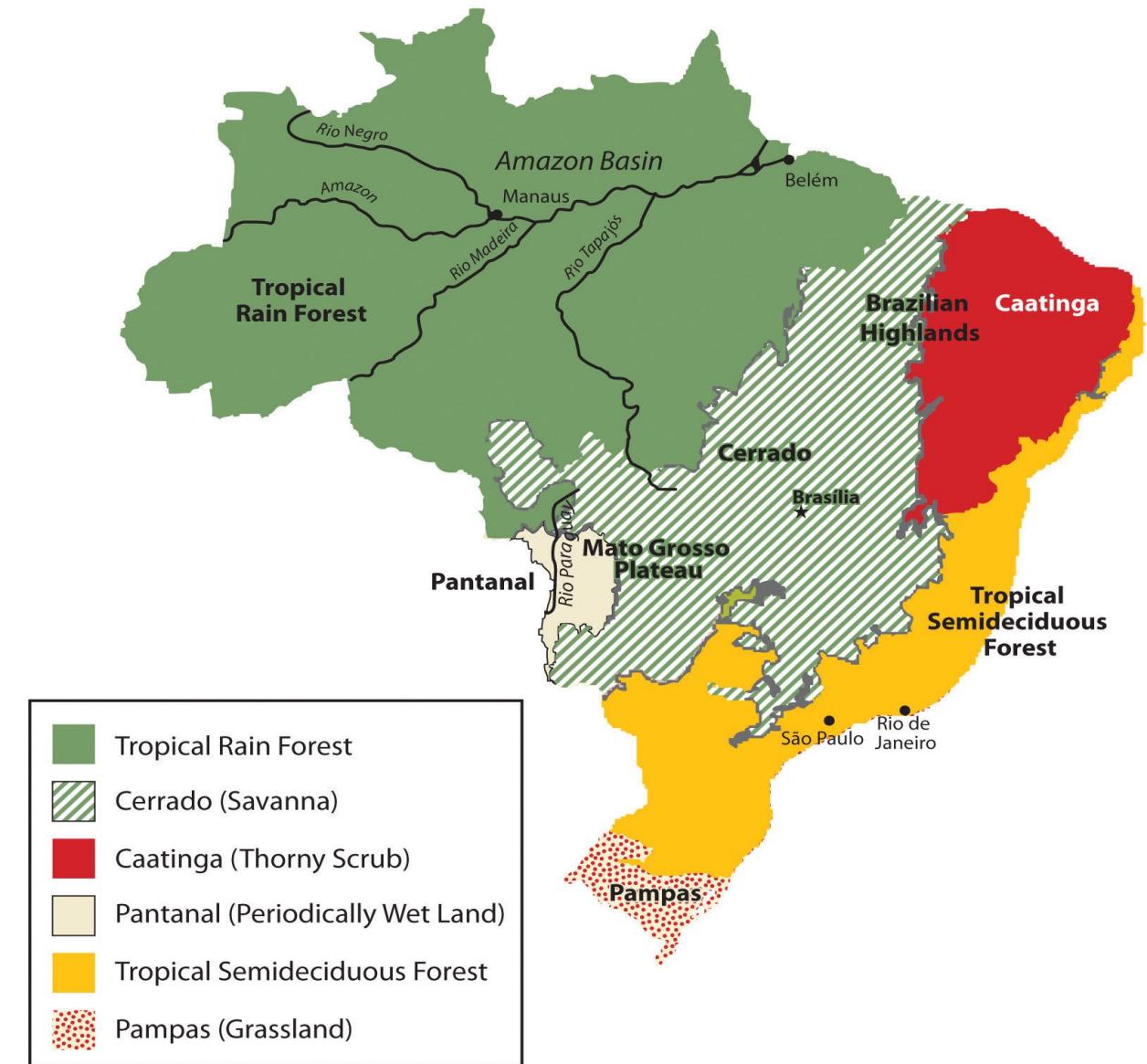
olist Share of Orders across Brazil



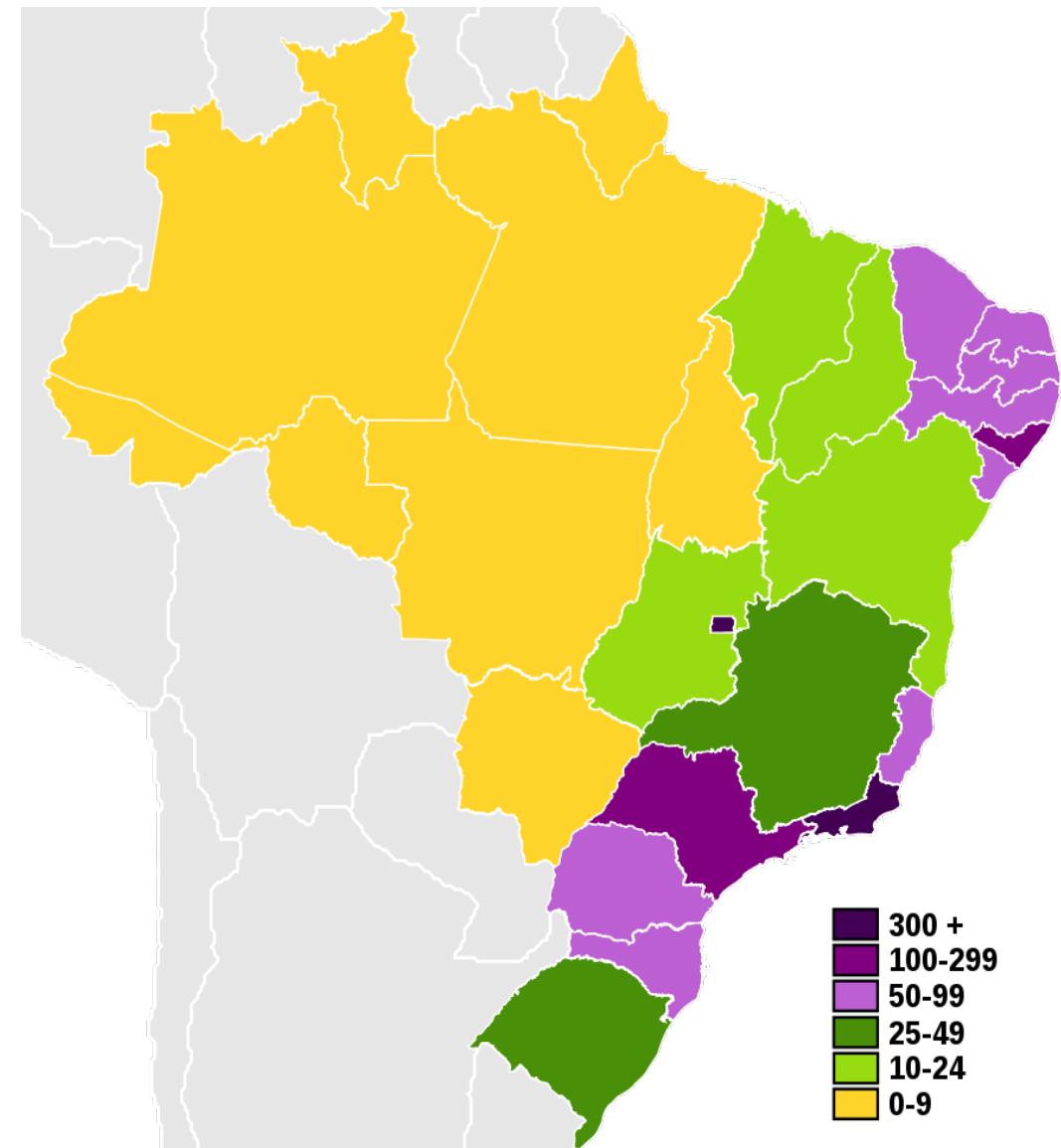
Brazilians love Toys!



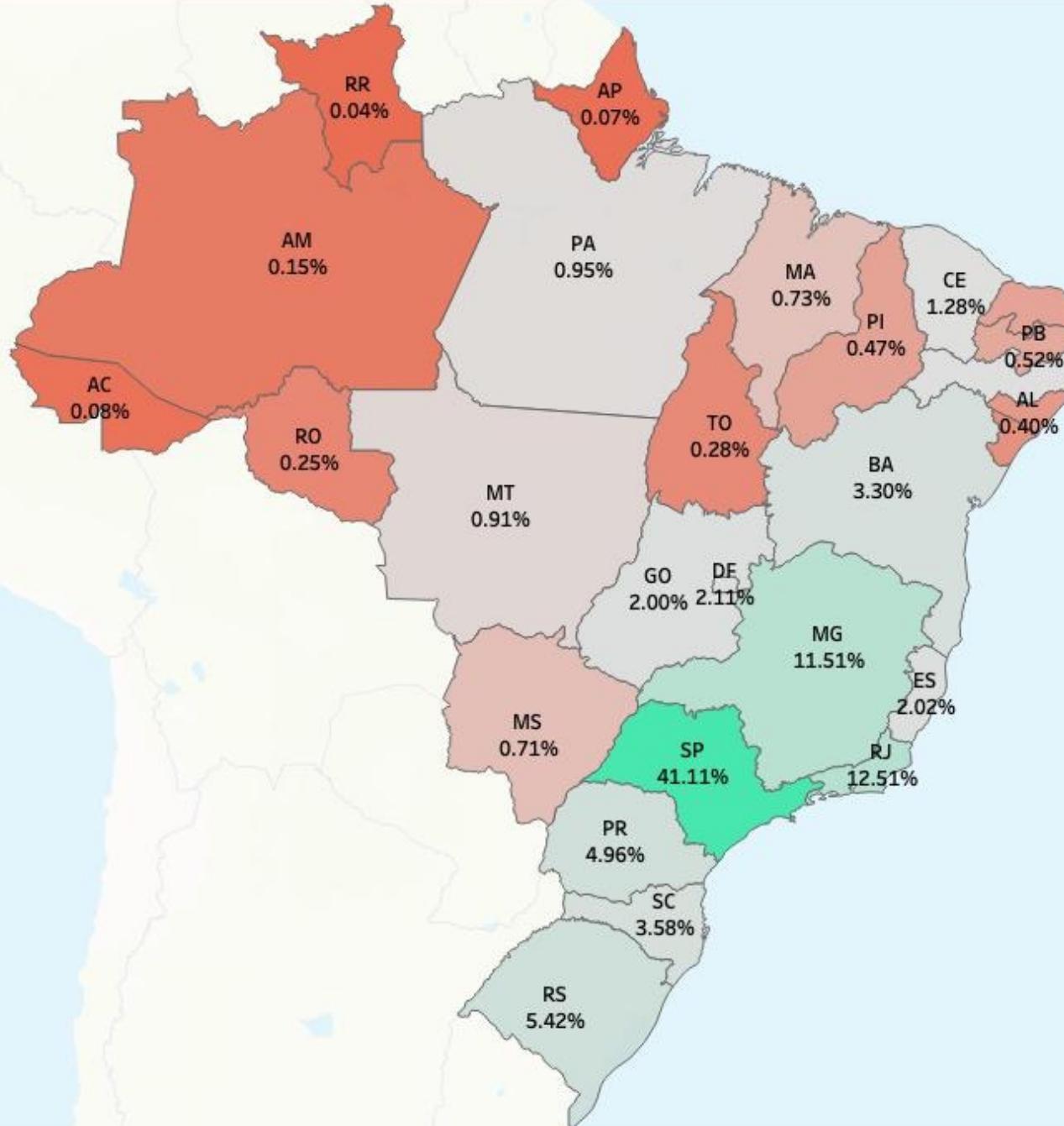
Regions in Brazil



Population Distribution in Brazil



olist Share of Orders across Brazil



Agenda

•

- 1 Objective
- 2 Background
- 3 Inferences
- 4 Suggestions
- 5 Appendix



Objective::



Background



To recover from losses, Olist must reduce any unnecessary costs it might be bearing.

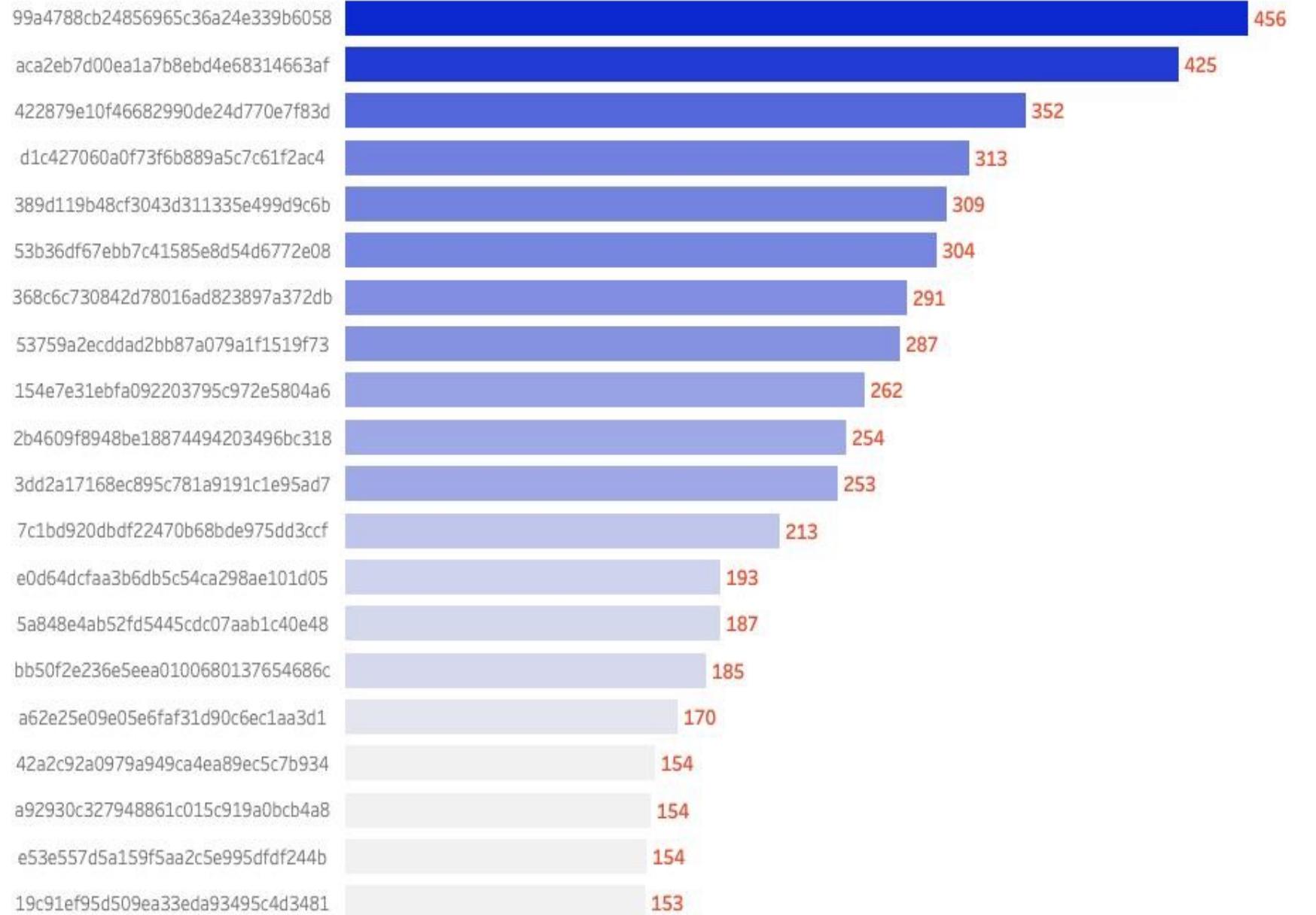
Olist has faced some losses recently and want to manage their inventory very well.

Inferences:



Top 20 Ordered Products by Quantity

Product Id



Number of Orders ↗



Top 20 Products by Revenue

Product Id



Revenue ₡



Top 20 Products Category by Orders Quantity

toys
77,154

health_beauty
2,888

sports_leisure
2,094

telephony
1,104 auto
1,022

garden_tools
842

perfumery
775

baby pet_shop





Market Basket Analysis

Product Categories which are Ordered more than 5 Times

toys
74,929

health_beauty
2,868

housewares
1,477

auto
1,015

garden_tools
836

perfumery
770

baby
727

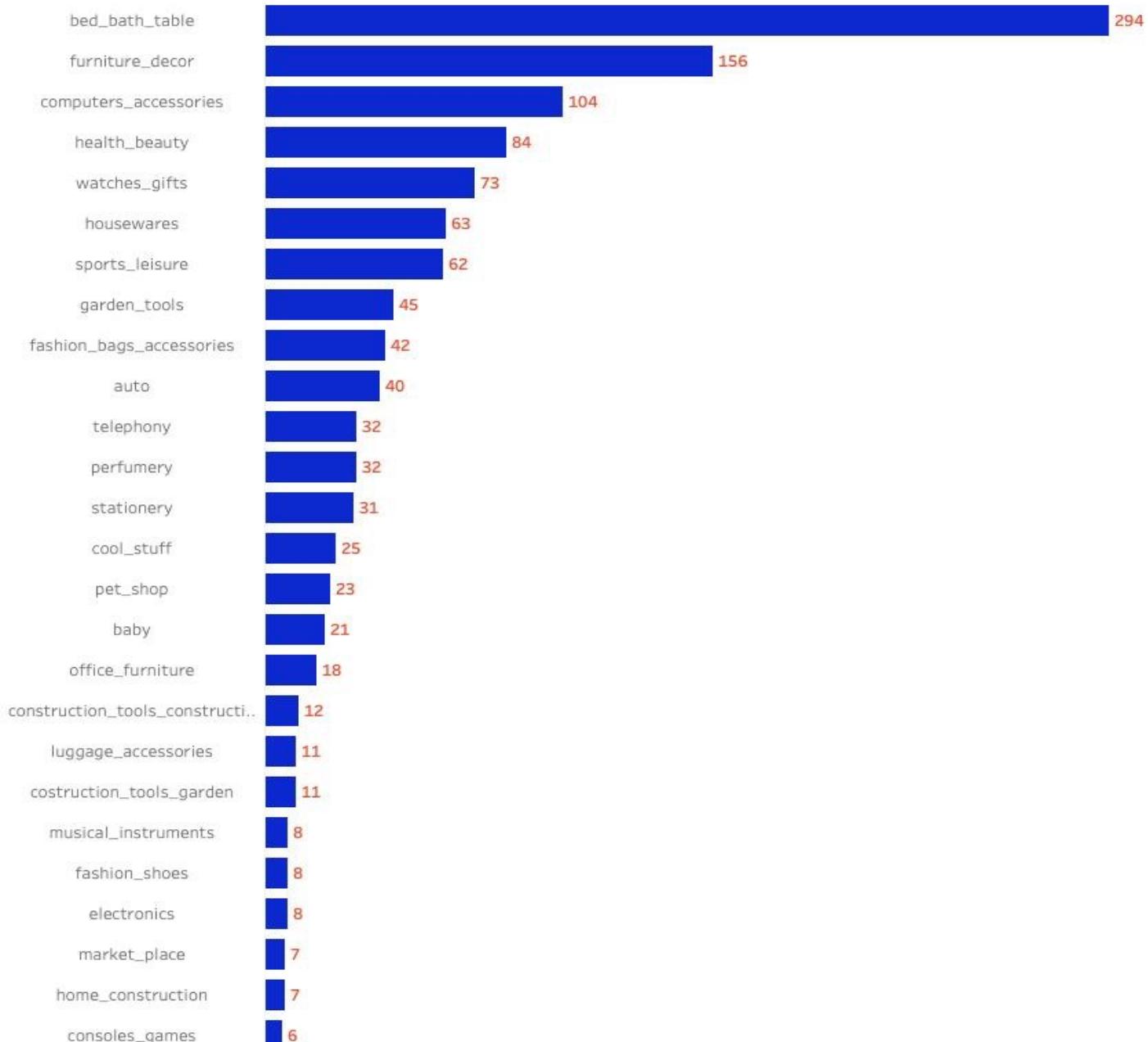
electronics





Combinations of Product Categories which are Frequently Ordered together

Basket also Contains

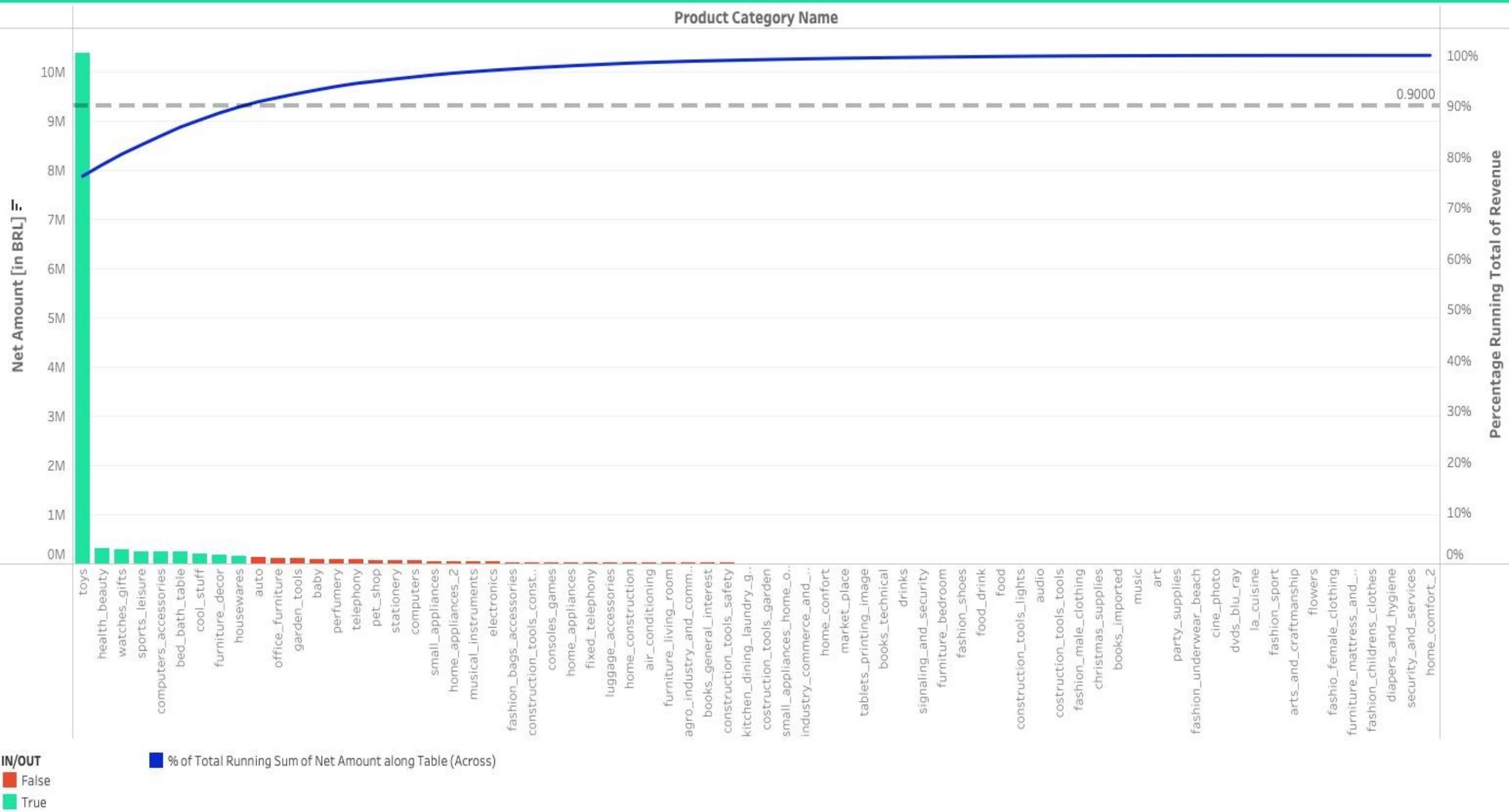


- An example, these are the product categories which are frequently ordered together with 'toys'.



Pareto Analysis

Pareto Analysis on Contribution of Distinct Products Categories by their Revenue in olist business



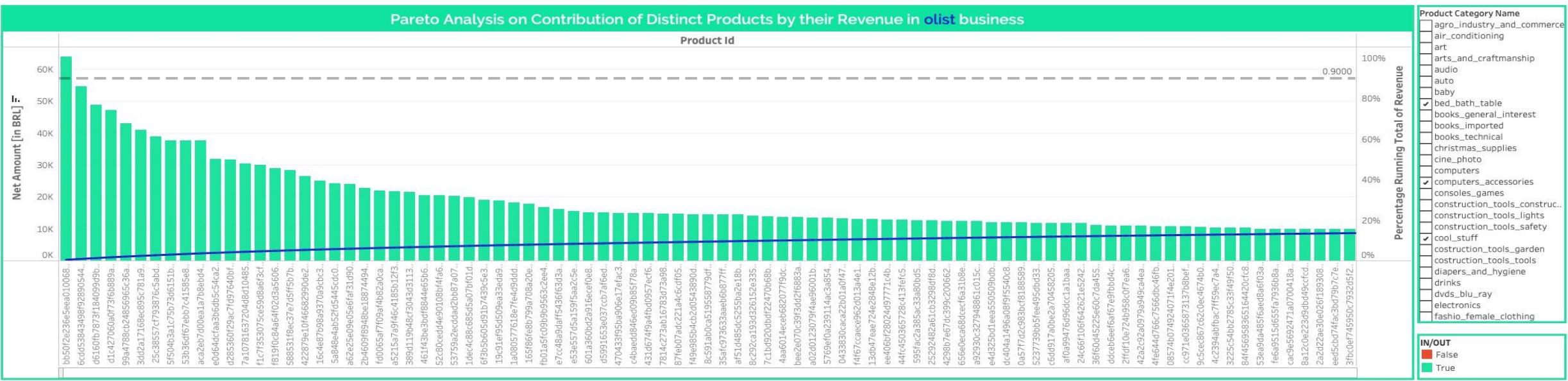
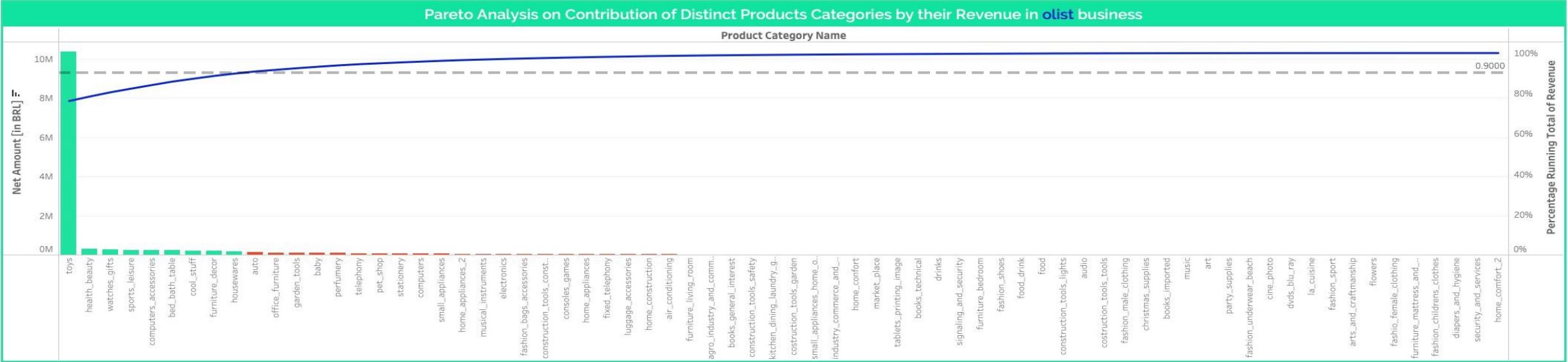
**90%
Revenue
result from
10% Product
Categories!**

- toys
- health_beauty
- watches_gifts
- sports_leisure
- computers_accessories
- bed_bath_table
- cool_stuff
- furniture_decor
- housewares

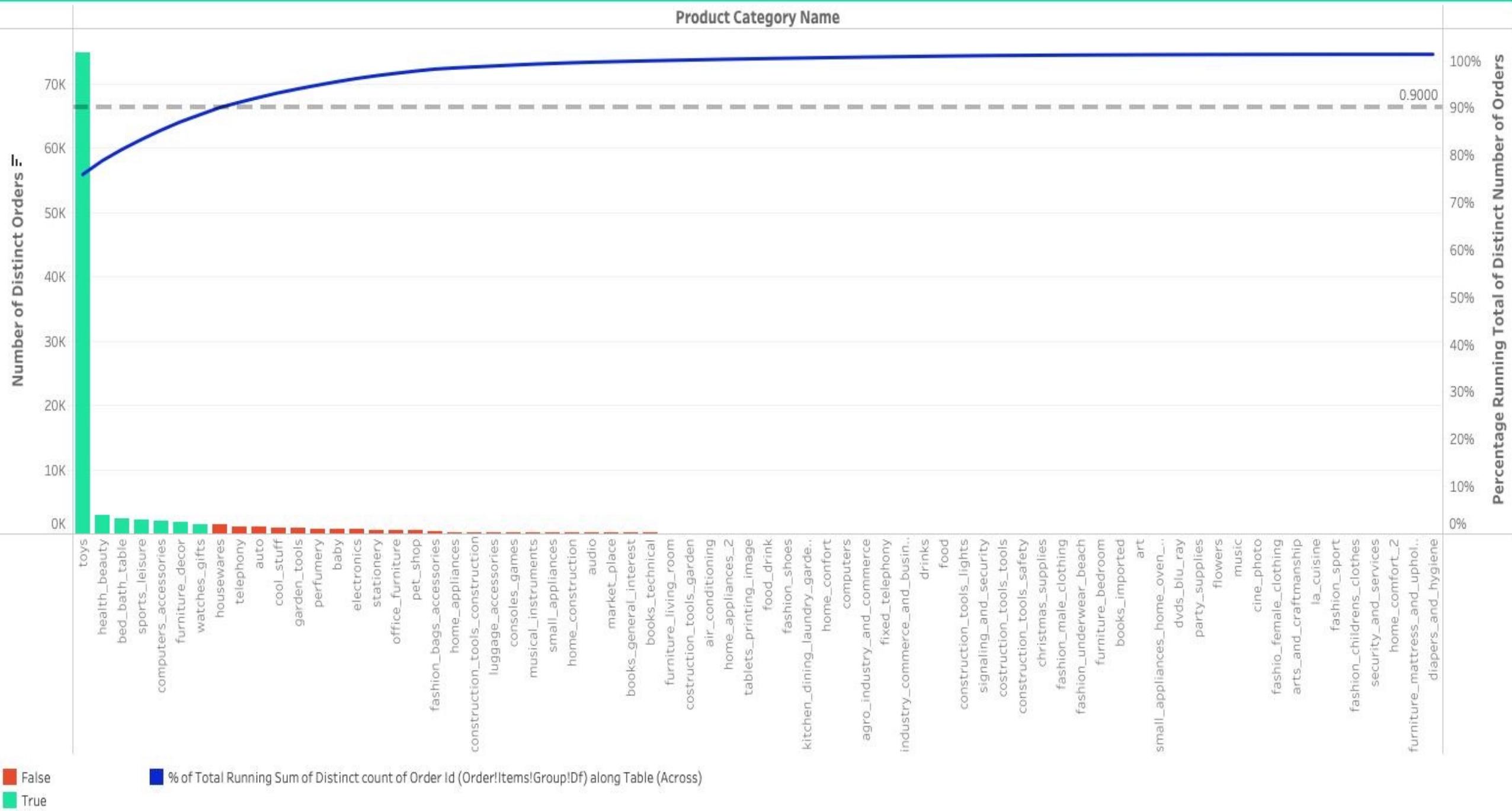
Pareto Analysis on Product Categories & Products by Revenue



olist



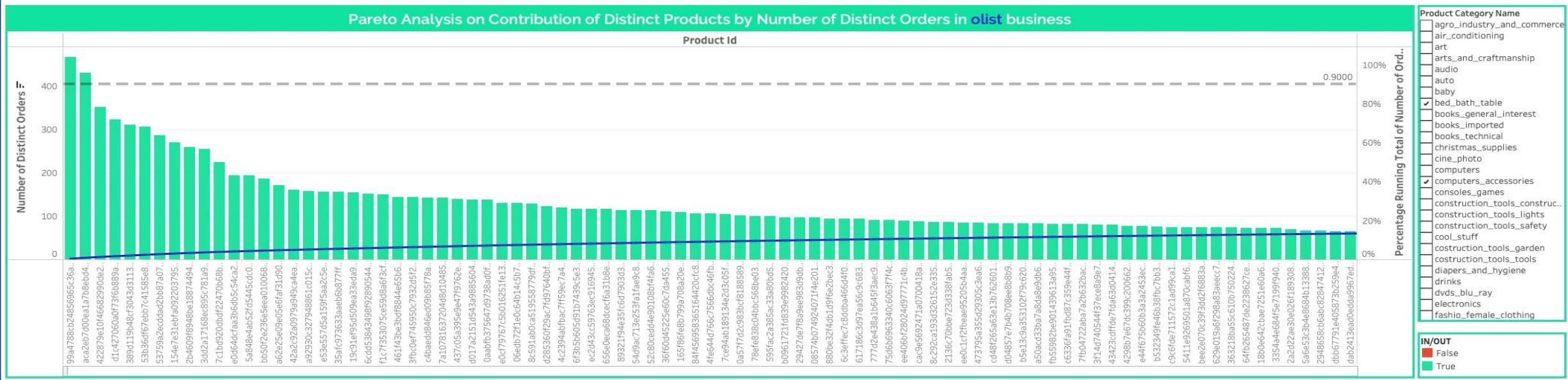
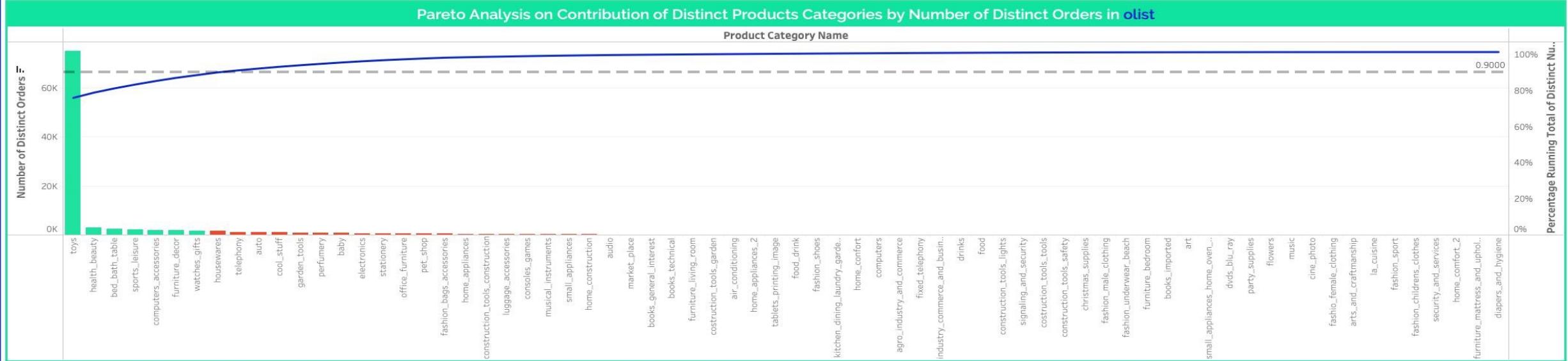
Pareto Analysis on Contribution of Distinct Products Categories by Number of Distinct Orders in olist



**90%
Orders
result from
10% Product
Categories!**

- toys
- health_beauty
- bed_bath_table
- sports_leisure
- computers_accessories
- furniture_decor
- watches_gifts

Pareto Analysis on Product Categories & Products by Distinct Orders Count



Suggestions:





Concentrate on the top 20 product categories based on the number of orders and those that have been ordered more than five times.



Customers who purchase toys also purchase other items. Toys should be offered in conjunction with new offers.

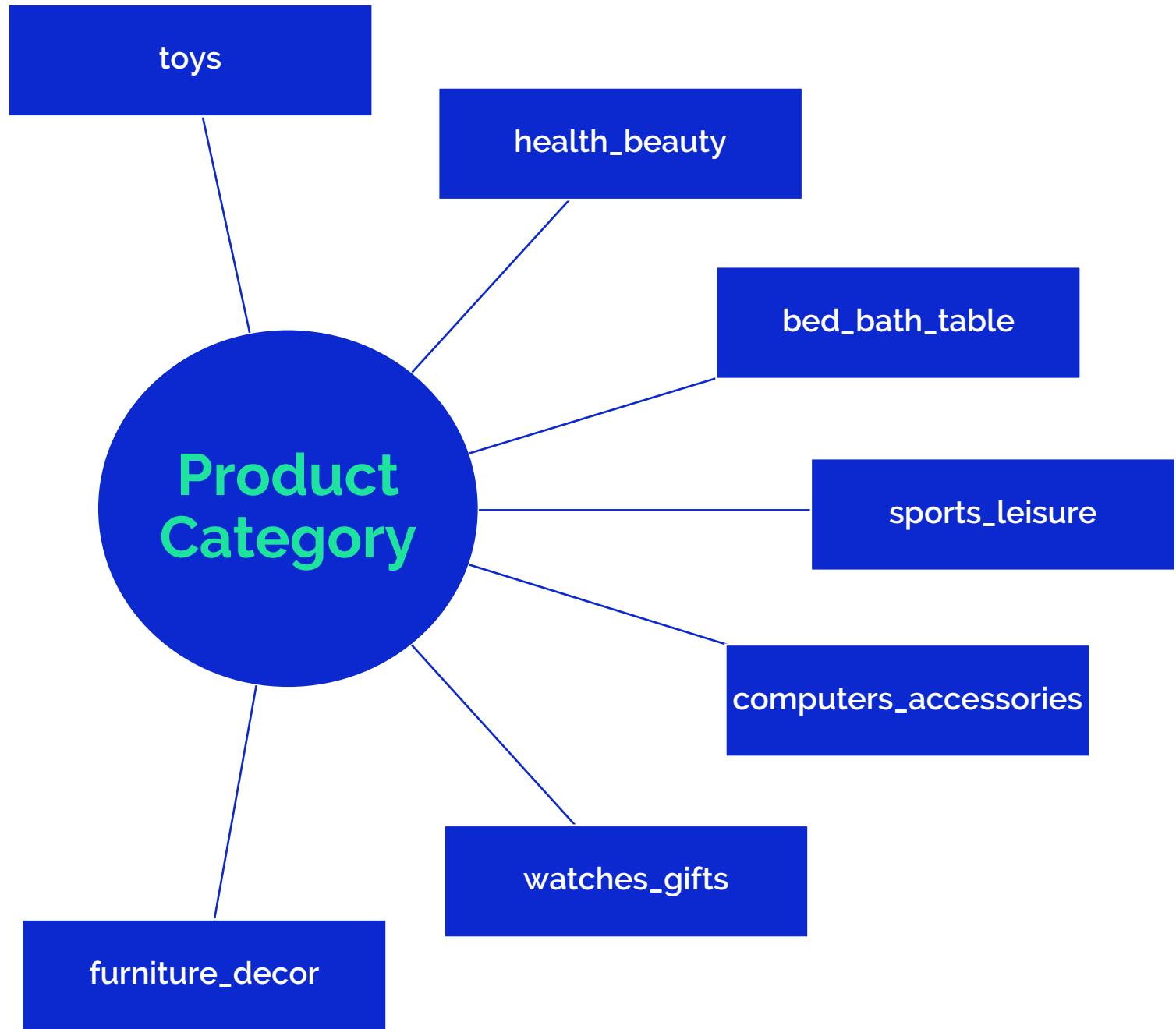
Customers who purchase lower-cost toys should be offered relevant or complementary things in addition to their purchase.



Customers who purchase higher-end toys should be offered a comparable higher-end product in addition to the one they purchased.

Products that are frequently ordered and generate a lot of revenue must always be available in stock.

Focus on the following product categories to manage inventory, as they generate large revenue and continue to be in high demand among customers





Check the Tableau workbook for 10% of items responsible for 90% of Olist's income or sale with their product ids, which sells as well as creates revenue for the company, to retain depth in the aforementioned categories.

Appendix...





Data Sources

Data Dictionary

Dataset name	Column Name	Description
orders	order_id	Unique identifier for an order, acts as the primary key of this table
orders	customer_id	Unique identifier for a customer, however, this table wont be unique at this level
orders	order_status	Indicates the status of an order, for example: delivered, cancelled, processing etc.
orders	order_purchase_timestamp	Timestamp when the order was made from the customer
orders	order_approved_at	Timestamp when the order was approved from the sellers' side
orders	order_delivered_timestamp	Timestamp when the order was delivered at customer's location
orders	order_estimated_delivery_date	Estimated date of delivery shared with the customer while placing the order
order_items	order_id	Unique identifier for an order
order_items		Item number in each order.
order_items	order_item_id	Order_id along with this column acts as the primary key of this table
order_items	product_id	Unique identifier for a product
order_items	seller_id	Unique identifier for the seller
order_items	price	Selling price of the product
order_items	shipping_charges	Charges associated with the shipping of the product
customers	customer_id	Unique identifier for a customer, acts as the primary key of this table
customers	customer_zip_code_prefix	Customer's Zip code
customers	customer_city	Customer's Zip city
customers	customer_state	Customer's Zip state
payments	order_id	Unique identifier for an order, this table can have duplicates in this column
payments	payment_sequential	Provides the info of the sequence of payments for the given order
payments	payment_type	Type of payment like credit_card, debit_card etc.
payments	payment_installments	Payment installement number in case of credit cards
payments	payment_value	Transaction value
products	product_id	Unique identifier for each product, acts as the primary key of this table
products	product_category_name	Name of the category the product belongs to
products	product_weight_g	Product weight in grams
products	product_length_cm	Product length in centimeters
products	product_height_cm	Product height in centimeters
products	product_width_cm	Product width in centimeters



Solution Methodology

A photograph of a woman and two children shopping in a grocery store. The woman, wearing a light blue long-sleeved shirt and jeans, stands on the left, looking at a bottle of juice. Two young boys, one in a light green shirt and jeans, and another in a white and yellow checkered shirt and khaki pants, stand next to her, also looking at bottles. A shopping cart is visible on the far left. The background shows well-stocked grocery shelves filled with various brands of juice and soda.

1. Data Collection

Retail_dataset
is provided by
Olist.

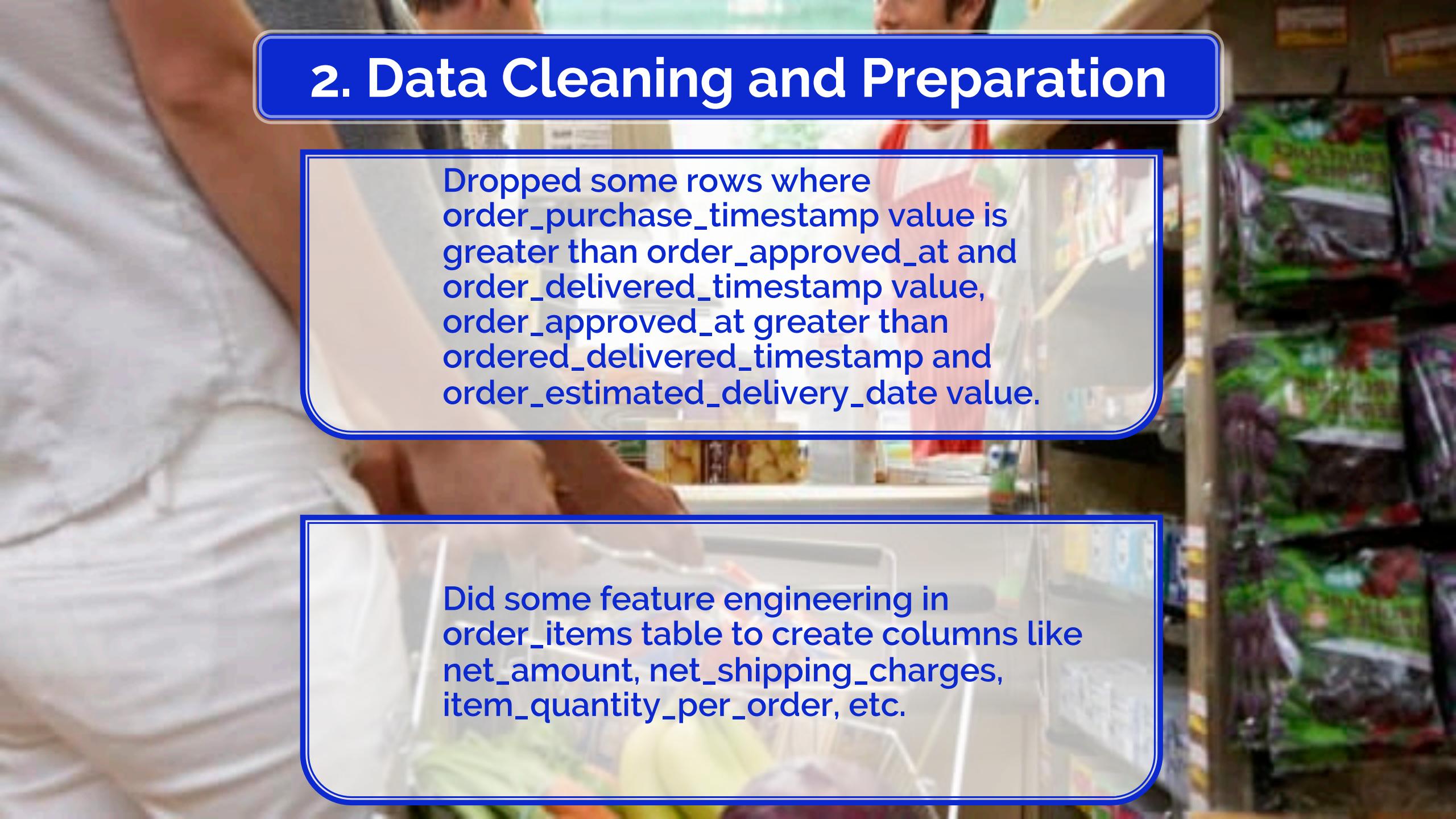
2. Data Cleaning and Preparation

Did the preliminary inspection of data by figuring the percentage of null values both in rows and columns, shape and number of categorical and numerical columns.

Updated the name of a few columns to avoid any confusion.

Dropped 'order_item_id' and 'payment_sequential' column as they are irrelevant for our data analysis from order_items and payments tables respectively.

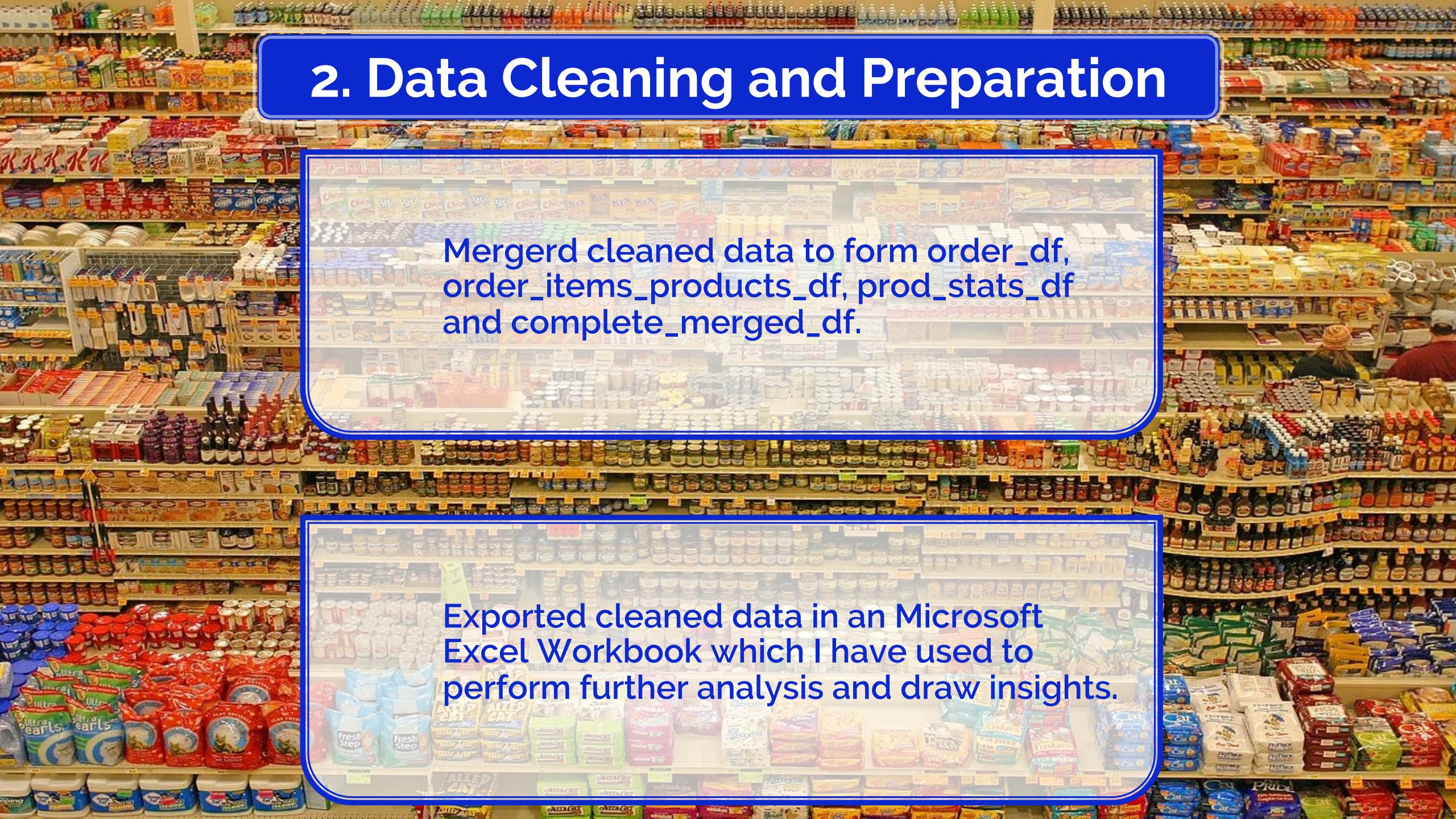
2. Data Cleaning and Preparation

A blurred background image of a person shopping in a grocery store aisle, with shelves of food items like lettuce and other packaged goods visible.

Dropped some rows where `order_purchase_timestamp` value is greater than `order_approved_at` and `order_delivered_timestamp` value, `order_approved_at` greater than `ordered_delivered_timestamp` and `order_estimated_delivery_date` value.

Did some feature engineering in `order_items` table to create columns like `net_amount`, `net_shipping_charges`, `item_quantity_per_order`, etc.

2. Data Cleaning and Preparation



Merged cleaned data to form order_df, order_items_products_df, prod_stats_df and complete_merged_df.

Exported cleaned data in an Microsoft Excel Workbook which I have used to perform further analysis and draw insights.



Tableau Desktop was used to visualise attributes in order to get relevant insights, performing market basket analysis and pareto analysis while python was used to perform exploratory data analysis.

4. Data Insights Generation

Insight 1

- Over a two-year period, the maximum number of orders per month occurs between the months of March and August.

Insight 2

- Saturday and Sunday have a lower percentage of Olist customers ordering, but this is not the case with Monday, Tuesday, and Wednesday, which have the highest number of orders during the last two years.

Insight 3

- Approximately '56 percent' of consumers prefer to order in the morning, between 7 a.m. and 1 p.m., or at night, between 7 p.m. and 12 a.m.

Insight 4

- The majority of customers like to order in the afternoon, between 1pm and 7pm, accounting for roughly 38% of all orders.

4. Data Insights Generation

Insight 5

- Toys account for more than 75% of all products sold, with the remaining product categories accounting for 25% of total revenue.

Insight 6

- Over the course of two years, 78974 payments (about 75%) were made by credit cards, followed by 20288 payments (around 20%) made via wallet.

Insight 7

- Over the course of two years, Olist received 15022 (15.63 percent) of orders from the city of Sao Paulo, followed by 6593 (6.92 percent) from the city of Rio de Janeiro.

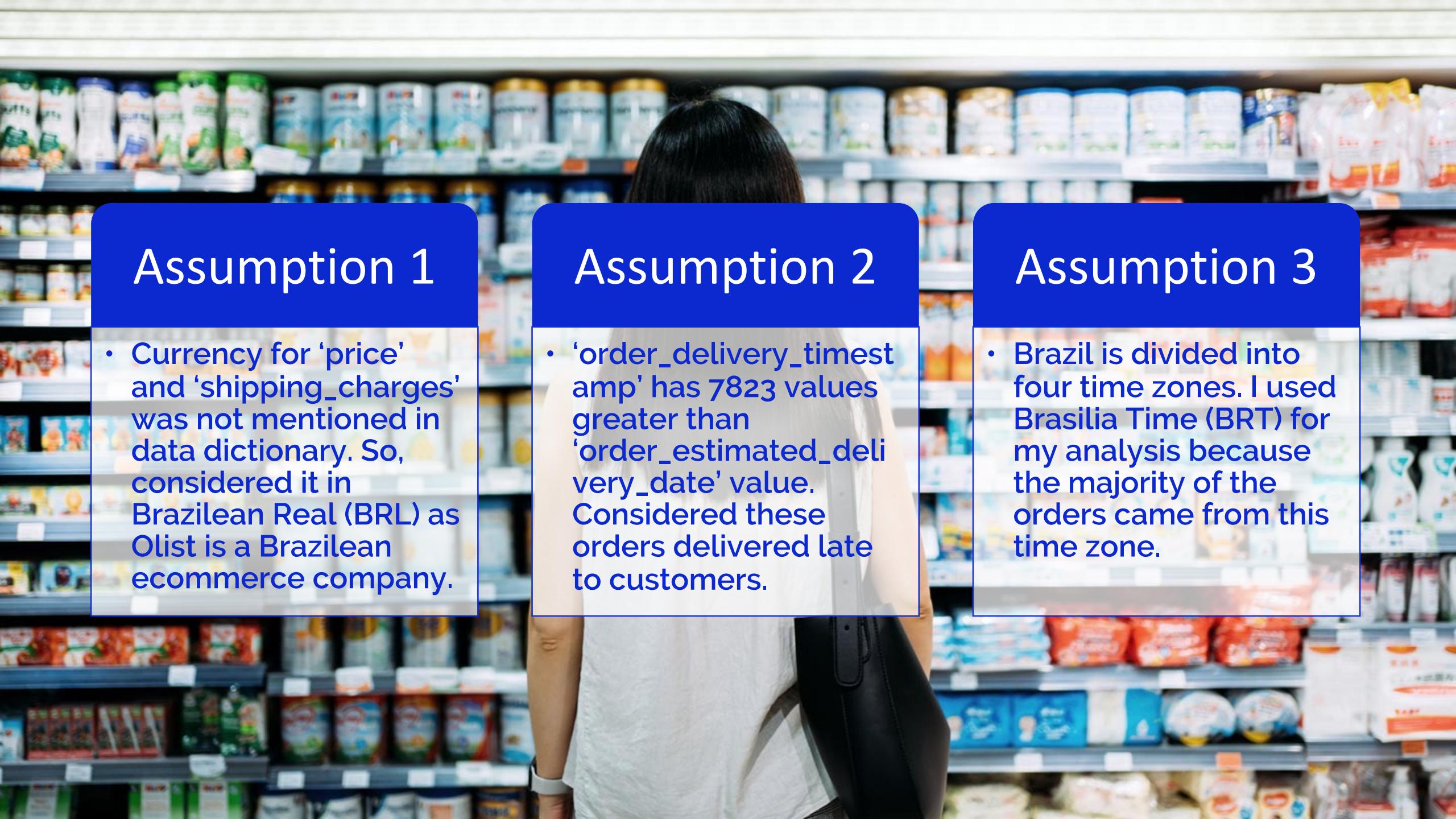
Insight 8

- Over the course of two years, Olist received the most orders in the state of Sao Paulo (SP) with 41.98 percent orders, followed by the state of Rio de Janeiro (RJ) with around 12.92 percent customers.



Find "Executive Summary" document attached with this presentation.





Assumption 1

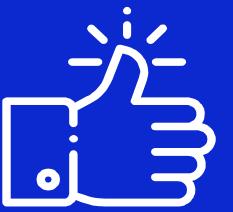
- Currency for 'price' and 'shipping_charges' was not mentioned in data dictionary. So, considered it in Brazilian Real (BRL) as Olist is a Brazilian ecommerce company.

Assumption 2

- 'order_delivery_timestamp' has 7823 values greater than 'order_estimated_delivery_date' value. Considered these orders delivered late to customers.

Assumption 3

- Brazil is divided into four time zones. I used Brasilia Time (BRT) for my analysis because the majority of the orders came from this time zone.



Thank You

