



ASSIGNMENT DAY 14 : MENGANALISIS DESIGN DATA WAREHOUSE DAN IMPLEMENTASINYA

TABLE OF CONTENT

- Amazon Sale Report
- Data Cleaning
- Data Normalization
- ERD
- Data Visualization
- Data Insight
- Recomendation

AMAZON SALE REPORT

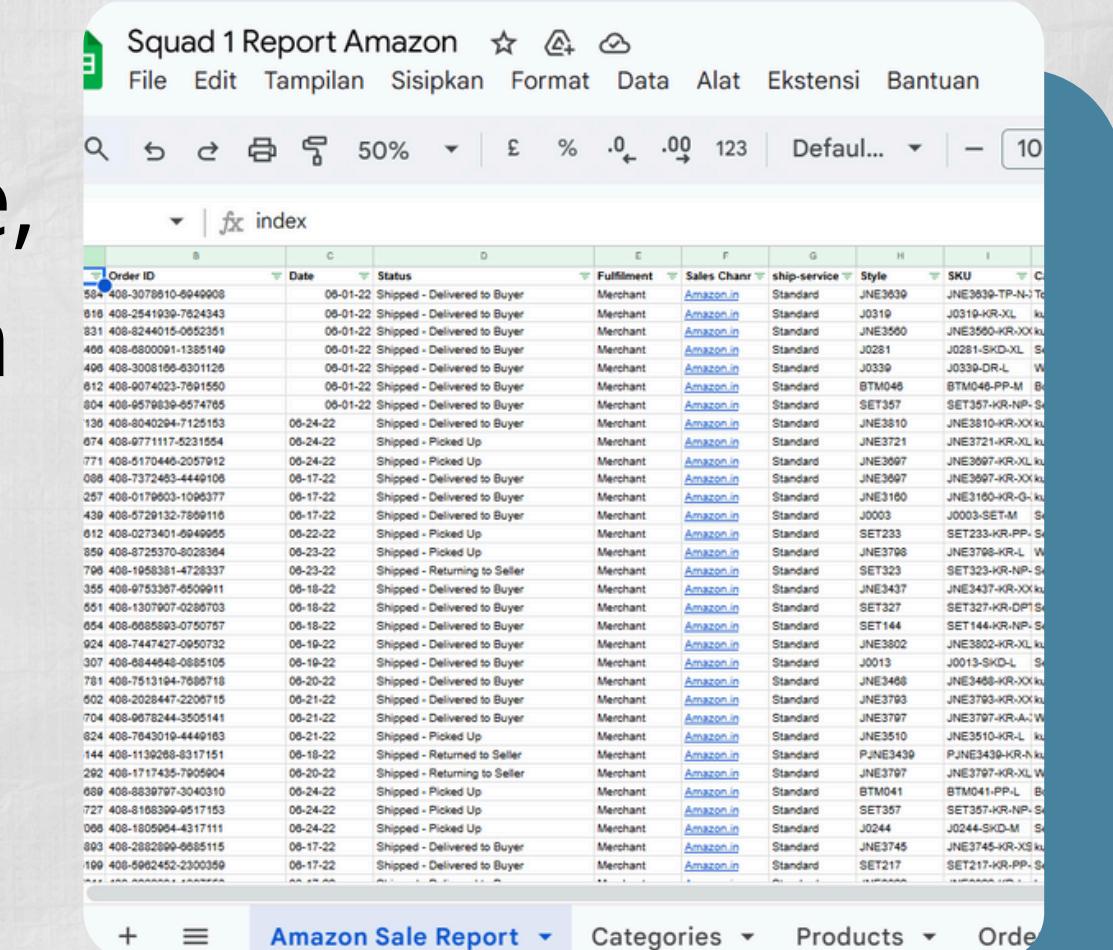
Amazon Sale Report Contains raw sales data from Amazon India. Period from April 1, 2022, to June 29, 2022. This raw data needs to be managed and processed to extract insights on Product Sales, Order Status (Shipped vs Cancelled), Sales Over Time, State Sales, and ...

The screenshot shows a Microsoft Excel spreadsheet titled "Amazon Sale Report". The ribbon menu includes File, Edit, Tampilan, Sisipkan, Format, Data, Alat, Ekstensi, and Bantuan. The toolbar includes icons for search, print, insert, filter, zoom (50%), and comment (Hanya komentar). The main content area displays a large table of sales data starting with row A1. The columns include: index, Order ID, Date, Status, Fulfilment, Sales Channel, ship-service-level, Style, SKU, Category, Size, ASIN, Courier Status, Qty, currency, Amount, ship-city, ship-state, ship-postal-code, ship-country, promotion-ids, B2B, and fulfilled-by. The data consists of approximately 25 rows of sales records.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Style	SKU	Category	Size	ASIN	Courier Status	Qty	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	promotion-ids	B2B	fulfilled-by
0	405-8078784-57	04-30-22	Cancelled	Merchant	Amazon.in	Standard	SET389	SET389-KR-NP- Set	S	B09KXVBD7Z		0	INR	647.62	MUMBAI	MAHARASHTR	400081	IN		FALSE	Easy Ship	
1	171-9198151-11	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	JNE3781	JNE3781-KR-XX kurta	3XL	B09K3WFS32	Shipped	1	INR	406	BENGALURU	KARNATAKA	580085	IN	Amazon PLCC F	FALSE	Easy Ship	
2	404-0687678-72	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3371	JNE3371-KR-XL kurta	XL	B07WV4JV4D	Shipped	1	INR	329	NAVI MUMBAI	MAHARASHTR	410210	IN	IN Core Free Sh	TRUE		
3	403-9615377-81	04-30-22	Cancelled	Merchant	Amazon.in	Standard	J0341	J0341-DR-L Western Dress	L	B099NRCT7B		0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008	IN		FALSE	Easy Ship	
4	407-1069790-72	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3671	JNE3671-TU-XX Top	3XL	B098T14BZP	Shipped	1	INR	574	CHENNAI	TAMIL NADU	600073	IN		FALSE		
5	404-1490984-45	04-30-22	Shipped	Amazon	Amazon.in	Expedited	SET264	SET264-KR-NP- Set	XL	B08YNTXDSG	Shipped	1	INR	824	GAZIABAD	UTTAR PRADES	201102	IN	IN Core Free Sh	FALSE		
6	408-5748499-68	04-30-22	Shipped	Amazon	Amazon.in	Expedited	J0095	J0095-SET-L Set	L	B08CMHNWB	Shipped	1	INR	653	CHANDIGARH	CHANDIGARH	160036	IN	IN Core Free Sh	FALSE		
7	406-7807733-37	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	JNE3405	JNE3405-KR-S kurta	S	B081WX4G4Q	Shipped	1	INR	399	HYDERABAD	TELANGANA	500032	IN	Amazon PLCC F	FALSE	Easy Ship	
8	407-5443024-52	04-30-22	Cancelled	Amazon	Amazon.in	Expedited	SET200	SET200-KR-NP- Set	3XL	B08L91ZZXN	Cancelled	0			HYDERABAD	TELANGANA	500008	IN	IN Core Free Sh	FALSE		
9	402-4393761-03	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3461	JNE3461-KR-XX kurta	XXL	B08B3XF5MH	Shipped	1	INR	363	Chennai	TAMIL NADU	600041	IN		FALSE		
10	407-5633625-69	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3160	JNE3160-KR-G- kurta	S	B07K3YQLF1	Shipped	1	INR	685	CHENNAI	TAMIL NADU	600073	IN		FALSE		
11	171-4638481-63	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3500	JNE3500-KR-XS kurta	XS	B098117DJ3	Shipped	1	INR	364	NOIDA	UTTAR PRADES	201303	IN		FALSE		
12	405-5513694-81	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	JNE3405	JNE3405-KR-XS kurta	XS	B081XCMYXJ	Shipped	1	INR	399	Amravati	MAHARASHTR	444606	IN	Amazon PLCC F	FALSE	Easy Ship	
13	408-7955685-30	04-30-22	Shipped	Amazon	Amazon.in	Expedited	SET182	SET182-KR-DH- Set	XS	B085HS947T	Shipped	1	INR	657	MUMBAI	MAHARASHTR	400053	IN		FALSE		
14	408-1298370-19	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	J0351	J0351-SET-L Set	L	B09CSSQY4F	Shipped	1	INR	771	MUMBAI	MAHARASHTR	400053	IN	Amazon PLCC F	FALSE	Easy Ship	
15	403-4965581-95	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	PJNE3368	PJNE3368-KR-6 kurta	6XL	B08PY99SVJ	Shipped	1	INR	544	GUNTAKAL	ANDHRA PRAD	515801	IN	Amazon PLCC F	FALSE	Easy Ship	
16	406-9379318-65	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3721	JNE3721-KR-XX kurta	XXL	B099FCT65D	Shipped	1	INR	329	JAIPUR	RAJASTHAN	302020	IN	IN Core Free Sh	FALSE		
17	405-0013803-80	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3405	JNE3405-KR-XL kurta	XL	B081WT6GG7	Shipped	1	INR	399	NEW DELHI	DELHI	110074	IN		FALSE		
18	402-4030358-58	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	JNE3697	JNE3697-KR-XX kurta	XXL	B098133PV5	Shipped	1	INR	458	Gurgaon	HARYANA	122004	IN	Amazon PLCC F	FALSE	Easy	
19	405-5957858-10	04-30-22	Shipped	Amazon	Amazon.in	Expedited	SET254	SET254-KR-NP- Set	XS	B0983DDPL6	Shipped	1	INR	886	BENGALURU	KARNATAKA	580017	IN		FALSE		
20	405-0607769-07	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3795	JNE3795-KR-XX kurta	3XL	B09HMXJVFS	Shipped	1	INR	517	TIRUCHIRAPP	TAMIL NADU	620018	IN	IN Core Free Sh	FALSE		
21	404-8494550-58	04-30-22	Shipped	Amazon	Amazon.in	Expedited	SET345	SET345-KR-NP- Set	M	B09KXV4BN8	Shipped	1	INR	666	BENGALURU	KARNATAKA	580040	IN	IN Core Free Sh	FALSE		
22	171-1305077-28	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3373	JNE3373-KR-L kurta	L	B082W7GVH7	Shipped	1	INR	376	HYDERABAD	TELANGANA	500072	IN		FALSE		
23	404-6019946-29	04-30-22	Cancelled	Merchant	Amazon.in	Standard	SET291	SET291-KR-PP- Set	M	B099NK55YG		0	INR	570.48	pune	MAHARASHTR	411044	IN				
24	402-3384087-40	04-30-22	Shipped	Amazon	Amazon.in	Expedited	MEN5002	MEN5002-KR-L kurta	L	B08YYYDN9R	Shipped	1	INR	499	TEZPUR	ASSAM	784001	IN	IN Core Free Sh			
25	405-8191138-51	04-30-22	Shipped - Delive	Merchant	Amazon.in	Standard	NW030	NW030-TP-PJ-X Set	XS	B09G2RQSRZ	Shipped	1	INR	582	RANCHI	JHARKHAND	834002	IN	Amazon PLCC F	FALSE		

DATA CLEANING

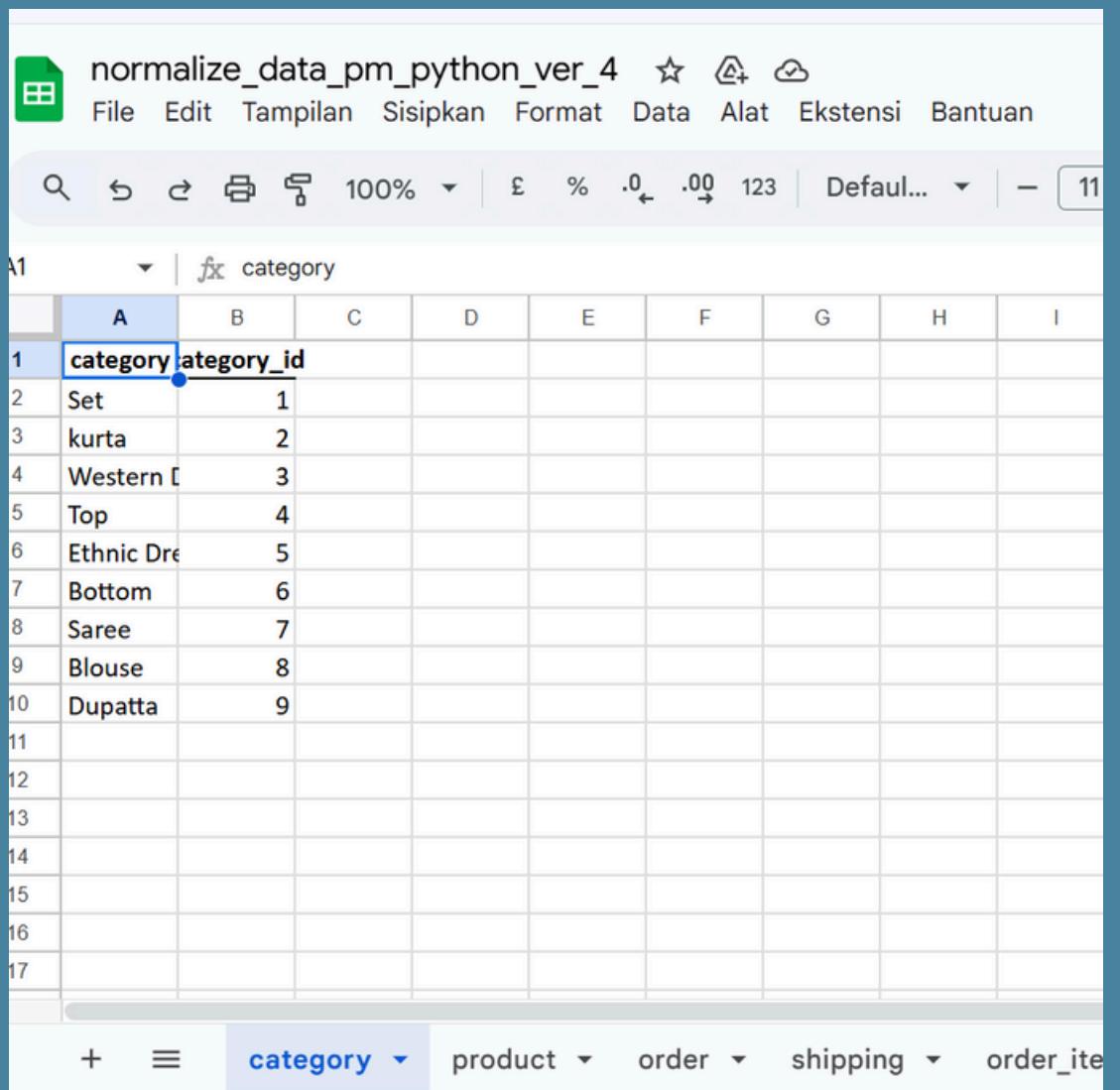
Data Cleaning transforms raw, messy data into a reliable, usable, and valuable asset. That data can drive accurate reports, better decisions, powerful models, and efficient operations. We perform data cleaning using both Google Spreadsheets and Phyton.



The screenshot shows a Google Sheets document with the title "Squad 1 Report Amazon". The spreadsheet contains a single sheet named "index" with data from an "Amazon Sale Report". The columns include Order ID, Date, Status, Fulfillment, Sales Chann, ship-service, Style, SKU, and C. The data consists of numerous rows of Amazon sales information, such as order numbers, dates, and fulfillment details. The interface includes standard Google Sheets tools like search, filters, and a toolbar at the top.

B	C	D	E	F	G	H	I
504 408-3078610-6949908	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3639	JNE3639-TP-N-JT
619 408-2541939-7624343	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	J0319	J0319-KR-XL
831 408-6244015-0523251	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3560	JNE3560-KR-XX
408 408-6800001-1385149	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	J0281	J0281-SKD-XL
406 408-3008106-6301126	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	J0339	J0339-DR-L
612 408-9074023-7691550	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	BTM046	BTM046-PP-J
804 408-9579839-6574785	06-01-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	SET357	SET357-KR-NP-S
136 408-6040294-7125153	06-24-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3810	JNE3810-KR-XX
674 408-9771117-6231554	06-24-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	JNE3721	JNE3721-KR-XL
771 408-5170446-2057912	06-24-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	JNE3597	JNE3597-KR-XX
098 408-7372453-4449108	06-17-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3597	JNE3597-KR-XX
257 408-0179603-1095377	06-17-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3160	JNE3160-KR-G
438 408-5729132-7869116	06-17-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	J0003	J0003-SET-M
612 408-0273401-6949955	06-22-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	SET233	SET233-KR-PP-S
859 408-6725370-8028354	06-23-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	JNE3798	JNE3798-KR-L
798 408-1953831-4723337	06-23-22	Shipped - Returning to Seller	Merchant	Amazon.in	Standard	JNE323	JNE323-KR-NP-S
355 408-9733557-6559911	06-18-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3437	JNE3437-KR-XX
551 408-1307907-0286703	06-18-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	SET327	SET327-KR-DPI-S
054 408-6655893-0750757	06-18-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	SET144	SET144-KR-NP-S
924 408-7447427-0650732	06-19-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3802	JNE3802-KR-XL
307 408-6844446-0685105	06-19-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	J0013	J0013-SKD-L
781 408-7513194-7696718	06-20-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3468	JNE3468-KR-XX
602 408-2028447-2206715	06-21-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3793	JNE3793-KR-XX
704 408-9078244-3505141	06-21-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3797	JNE3797-KR-A
624 408-7643019-4449103	06-21-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	JNE3510	JNE3510-KR-L
144 408-1139288-8317151	06-18-22	Shipped - Returned to Seller	Merchant	Amazon.in	Standard	PJNE3439	PJNE3439-KR-N
292 408-1717435-7905604	06-20-22	Shipped - Returning to Seller	Merchant	Amazon.in	Standard	JNE3797	JNE3797-KR-XL
659 408-8839797-3040310	06-24-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	BTM041	BTM041-PP-L
727 408-6168399-0517153	06-24-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	SET357	SET357-KR-NP-S
066 408-1805904-4317111	06-24-22	Shipped - Picked Up	Merchant	Amazon.in	Standard	J0244	J0244-SKD-M
693 408-2882899-6685115	06-17-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3745	JNE3745-KR-XS
199 408-5982452-2300359	06-17-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	SET217	SET217-KR-PP-S

DATA NORMALIZATION



	A	B	C	D	E	F	G	H	I
1	category	category_id							
2	Set	1							
3	kurta	2							
4	Western D	3							
5	Top	4							
6	Ethnic Dre	5							
7	Bottom	6							
8	Saree	7							
9	Blouse	8							
10	Dupatta	9							
11									
12									
13									
14									
15									
16									
17									

Data Normalization is performed to :

- Eliminate redundancy and minimize the risk of data errors.
- Separate each entity (Product, Shipping, Order, Category, Order item) into appropriate tables.
- Facilitate cross-entity data analysis and visualization.
- Remove incomplete data rows (missing values) to maintain analytical integrity.
- Standardize text formats, such as city names, to be uniform (all converted to uppercase)
- Adjust data formats to be understood by visualization applications (Looker Studio), specifically date formats.

Assets : [Normalize Data](#)

STRUCTURE OF NORMALIZED TABLE RESULTS

Table	Column	Data Type	Description
category	category_id	Integer	Unique category ID
category	category	Text	Product category name
product	sku	Text	Unique product ID
product	style	Text	Product name/type
product	category	Text	Product Category
product	category_id	Integer	Reference to category table
order	order_id	Text	Unique order ID
order	date	Date	Transaction date (will be cleaned)
order	status	Text	Order status
order	sales_channel	Text	Sales channel (Amazon.in, etc.)
order	fulfilment	Text	Order fulfillment method
order	ship_service_level	Text	Selected shipping service type
order	promotion_ids	Text	Promo name (if any)
order	b2b	Boolean	Marks whether the transaction is B2B or B2C
order	total_amount	Decimal	Total transaction value

shipping	order_id	Text	Order ID (foreign key)
shipping	ship_city	Text	Destination city (already standardized)
shipping	ship_state	Text	Destination province
shipping	ship_postal_code	Text	Destination zip code
shipping	ship_country	Text	Destination country
order_item	order_item_id	Integer	Unique item ID in order
order_item	order_id	Text	Order ID (foreign key)
order_item	sku	Text	Product SKU (foreign key)
order_item	qty	Integer	Quantity of units ordered
order_item	amount	Decimal	Product price x quantity value
order_item	currency	Text	Transaction currency (e.g. INR)
order_item	suspicious	Boolean	Indicator if qty=0 but there is an average

Incomplete Data

- Rows with empty values in critical columns such as date, amount, and ship_city have been removed from the dataset.
- Text formats such as city names are standardized using UPPER() in Google Sheets or .str.upper() in Python.
- Date formats not recognized by the system are converted using formulas or processing logic to match the Date type.
- Added Suspicious Flag in the Order_item table to identify missed in server logic

Rule Applied	Action Taken
qty = 0 but amount > 0	Marked as suspicious = TRUE in order_item table
Purpose	To flag potential data entry errors or anomalies in transaction
How It Helps	Enables better monitoring of unusual orders and improves data quality

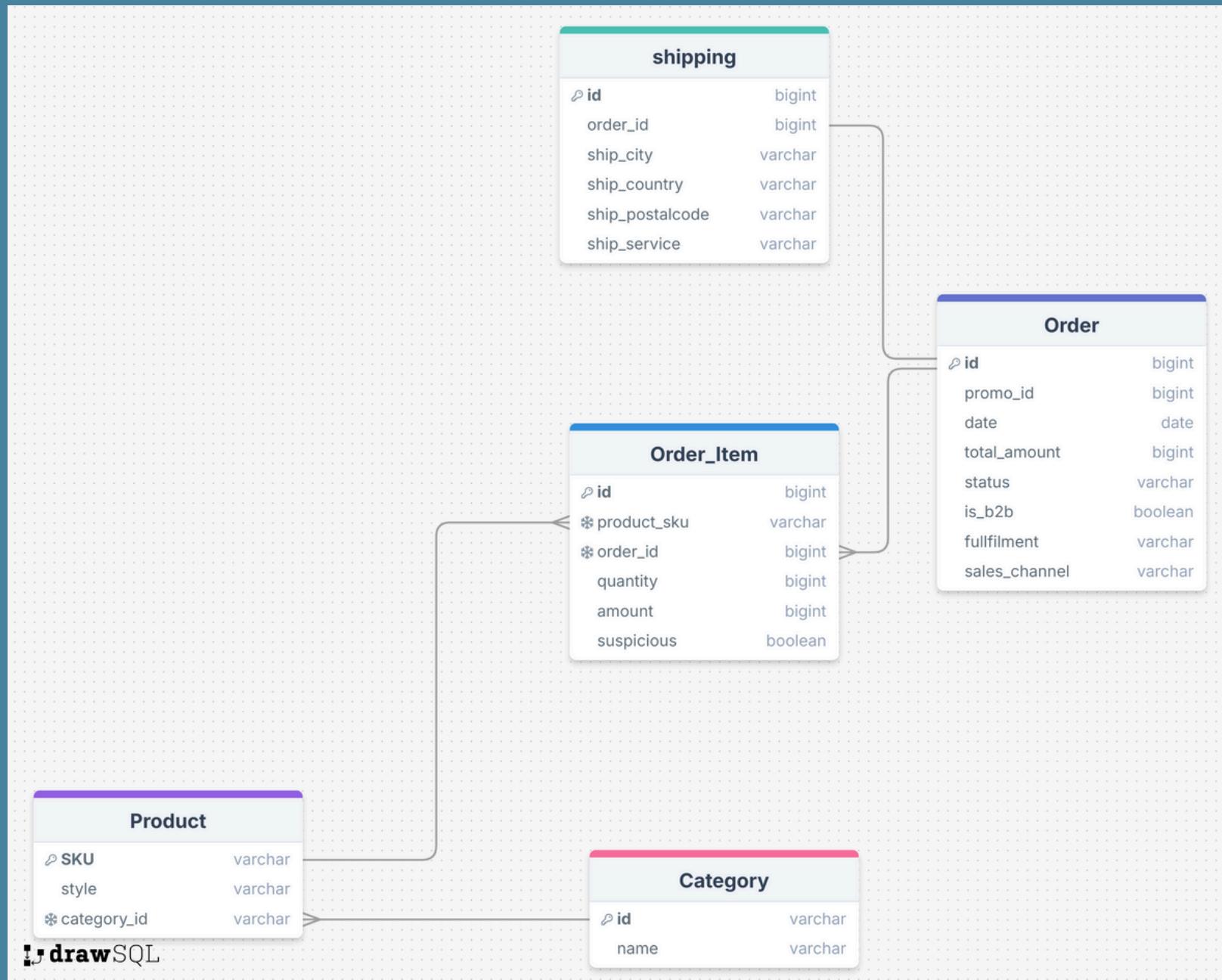


Normalization Workflow

- Loading raw data from an Excel file
- Cleaning empty or invalid columns
- Creating a category table from a list of unique categories
- Forming a product table based on unique SKUs
- Compiling an order table with total transaction calculations
- Separating shipping information into a dedicated table
- Creating an order_item table based on details per SKU per order
- Flagging anomalous data such as qty=0 but having an amount value
- Standardizing text format (UPPER) and date format (DATE)
- Exporting the final results to Google Sheets

ERD

We Separate the data into some tables for efficiency.
These tables will be used for data visualization.



Assets : [ERD](#)

1.Order ↔ Order_Item

Relation: One-to-Many

Explanation: One Order can contain multiple different items (Order_Item). This relation is established through the Order_ID

2.Order_Item ↔ Product

Relation: Many-to-One

Explanation: Each order item (Order_Item) refers to one specific product (Product) based on the Product_SKU. The same product can appear in many order items.

3.Product ↔ Category

Relation: Many-to-One

Explanation: Each Product (product) is assigned to one specific category (Category) based on category_id. One category can have many Products.

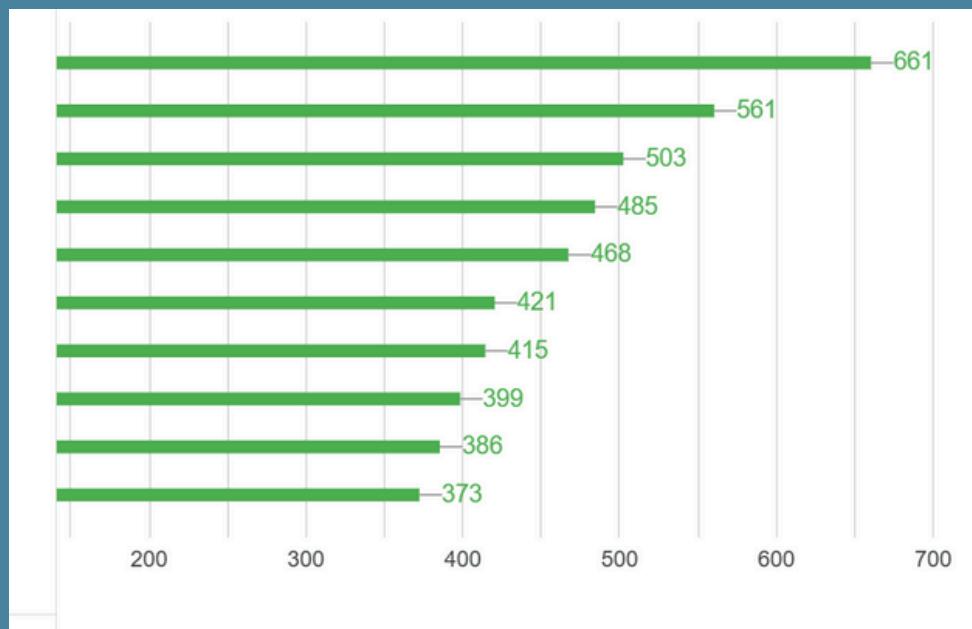
4.Order ↔ Shipping

Relation: One-to-One

Explanation: Each Order (Order) has exactly one set of shipping information (Shipping). This Relation is identified by the order_id.

ERD

DATA VISUALIZATION



Product Sales Data, Showing the product sold during the specified period in India. The top 3 best selling product :

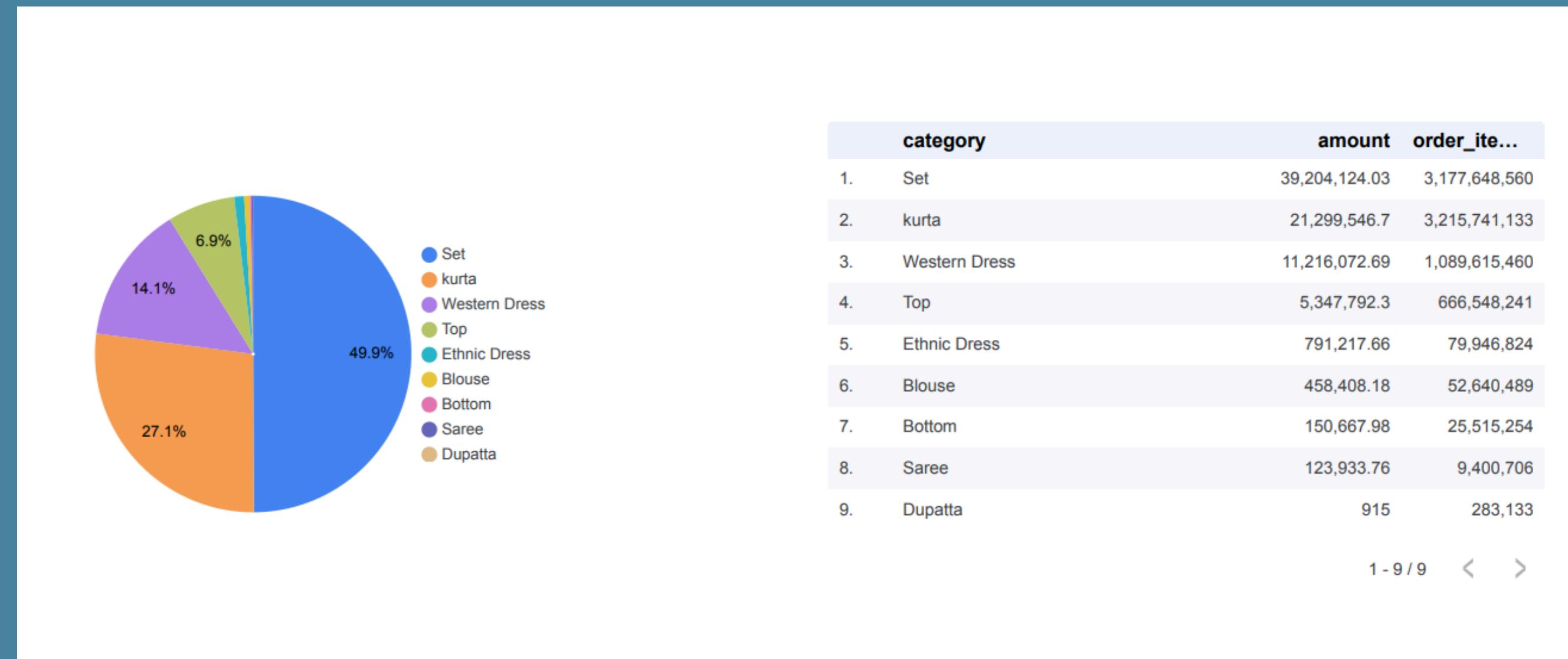
- JNE3797-KR-L (661 item sold)
- JNE3797-KR-M (561 item sold)
- JNE3797-KR-S (503 item sold)

The lowest sales among the top 10 best selling products are :

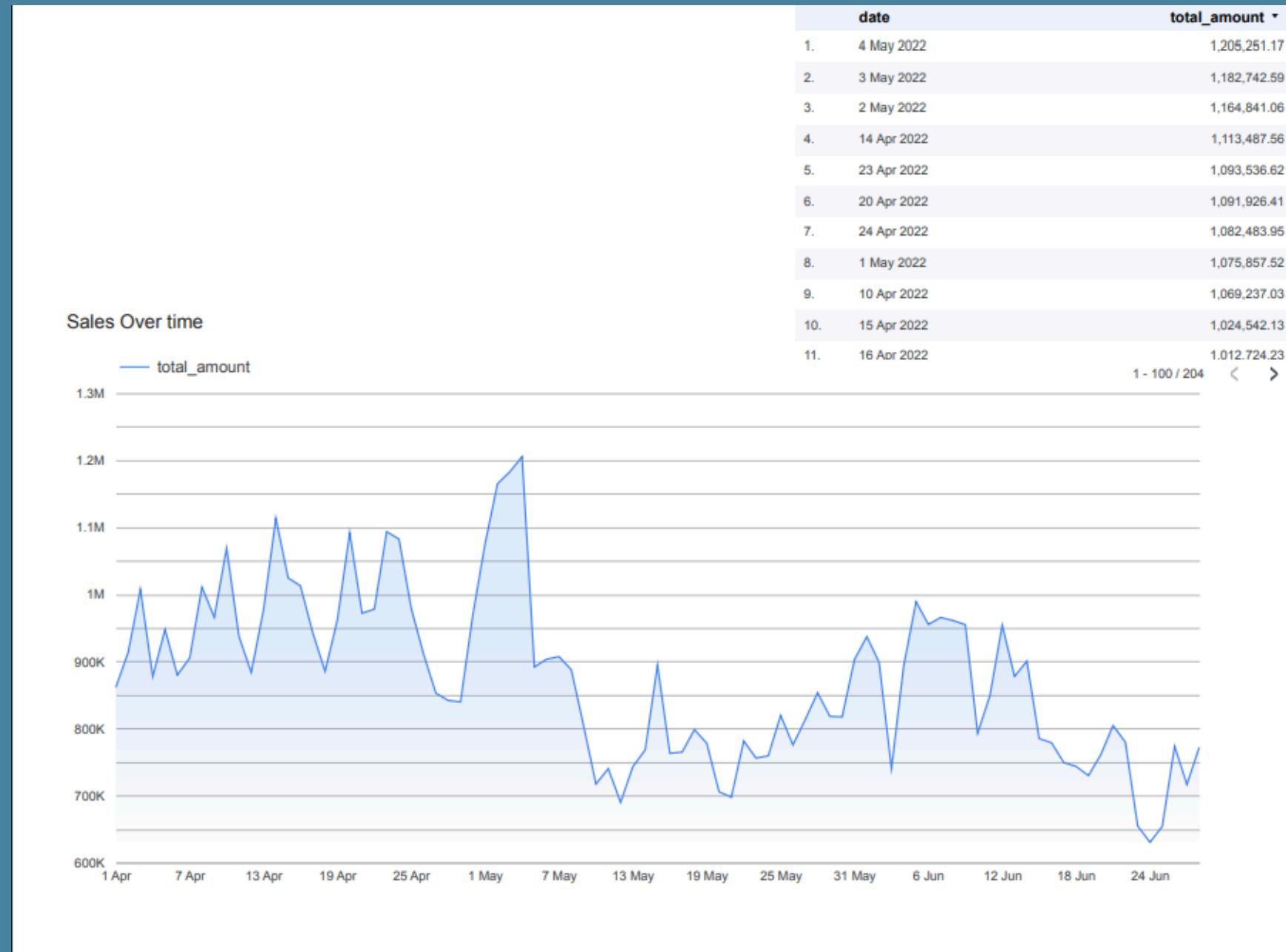
- JNE3405-KR-S (399 item sold)
- JNE3797-KR-XS (386 item sold)
- SET268-KR-NP-XL (373 item sold)

sku	qty
1. JNE3797-KR-L	661
2. JNE3797-KR-M	561
3. JNE3797-KR-S	503
4. JNE3405-KR-L	485
5. J0230-SKD-M	468
6. J0230-SKD-S	421
7. JNE3797-KR-XL	415
8. JNE3405-KR-S	399
9. JNE3797-KR-XS	386

1 - 100 / 7195 < >

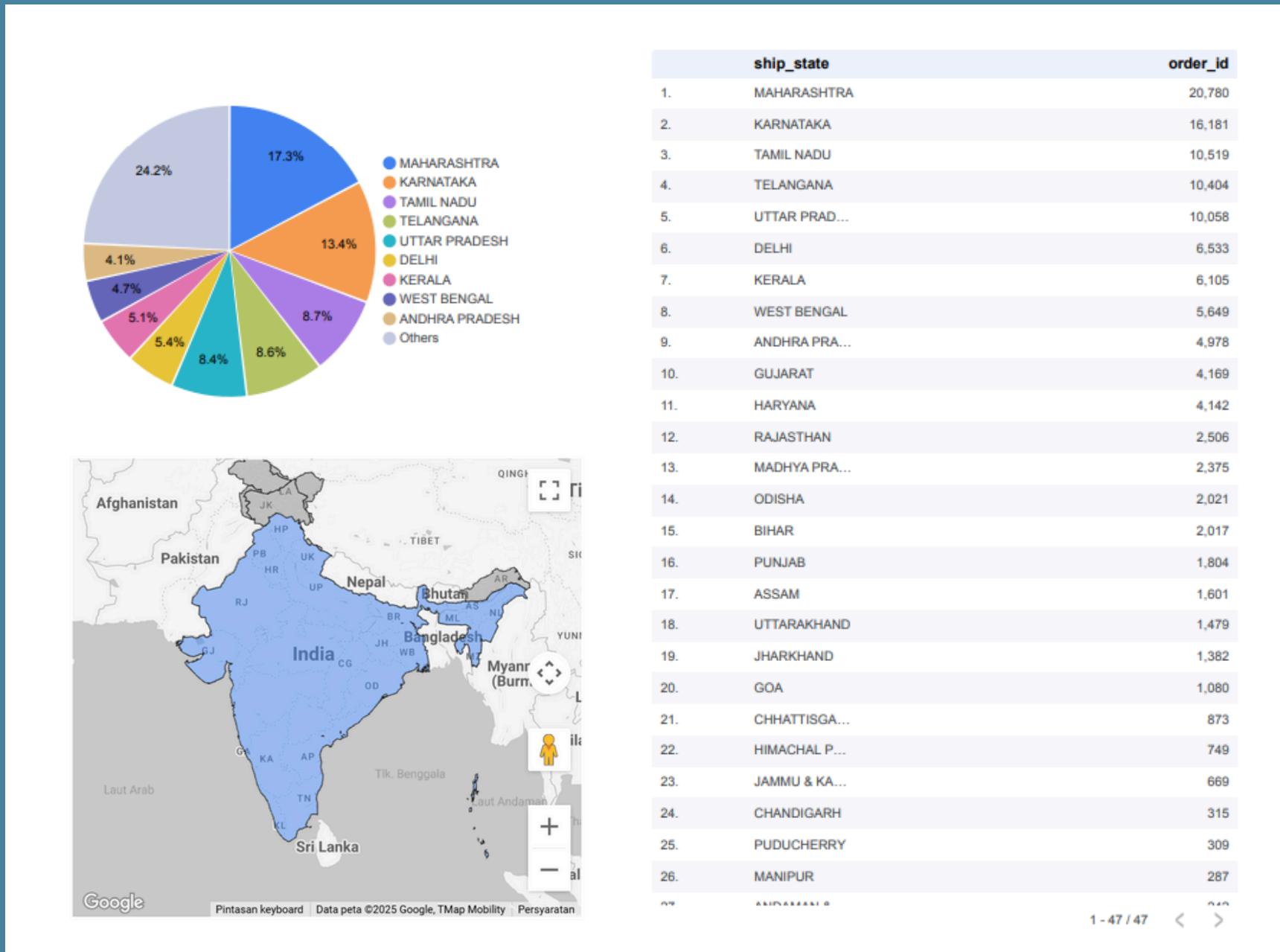


- The Set category dominates the sales, contributing 49.9% of total revenue (₹39.2M), followed by Kurta at 27.1% (₹21.3M).
- The top 3 categories (Set, Kurta, Western Dress) together account for over 91% of total sales — a clear indication of customer preference.



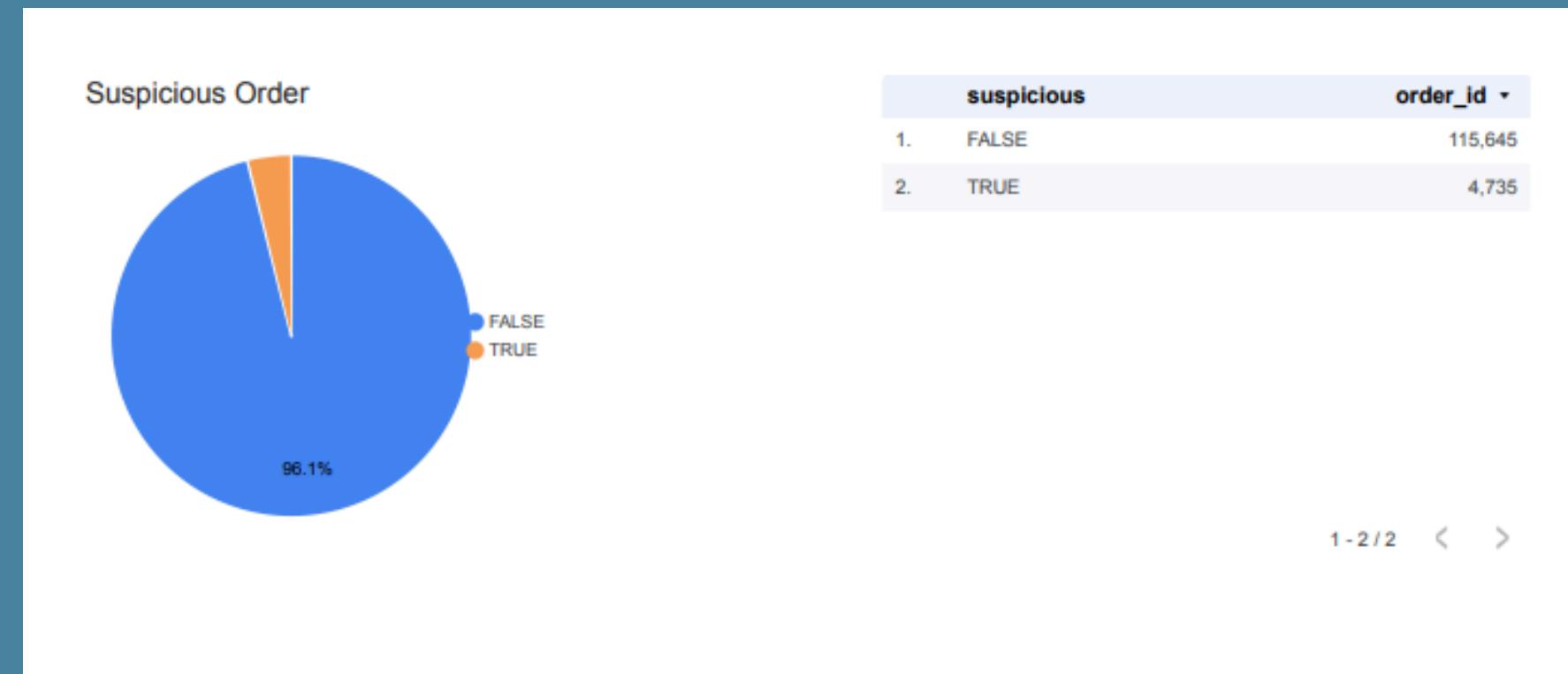
Sales Over Time

- The highest daily sales occurred on 4 May 2022, reaching over 1.2 million INR.
- There's a noticeable peak between late April and early May, followed by a gradual decline throughout June.
- This pattern may indicate the impact of seasonal promotions, payday periods, or product launches.



Top Sales By State

- The majority of orders are concentrated in a few states:
- Maharashtra (17.3%)
- Karnataka (13.4%)
- Tamil Nadu (8.7%)
- Together, the top 5 states account for nearly 60% of total orders.

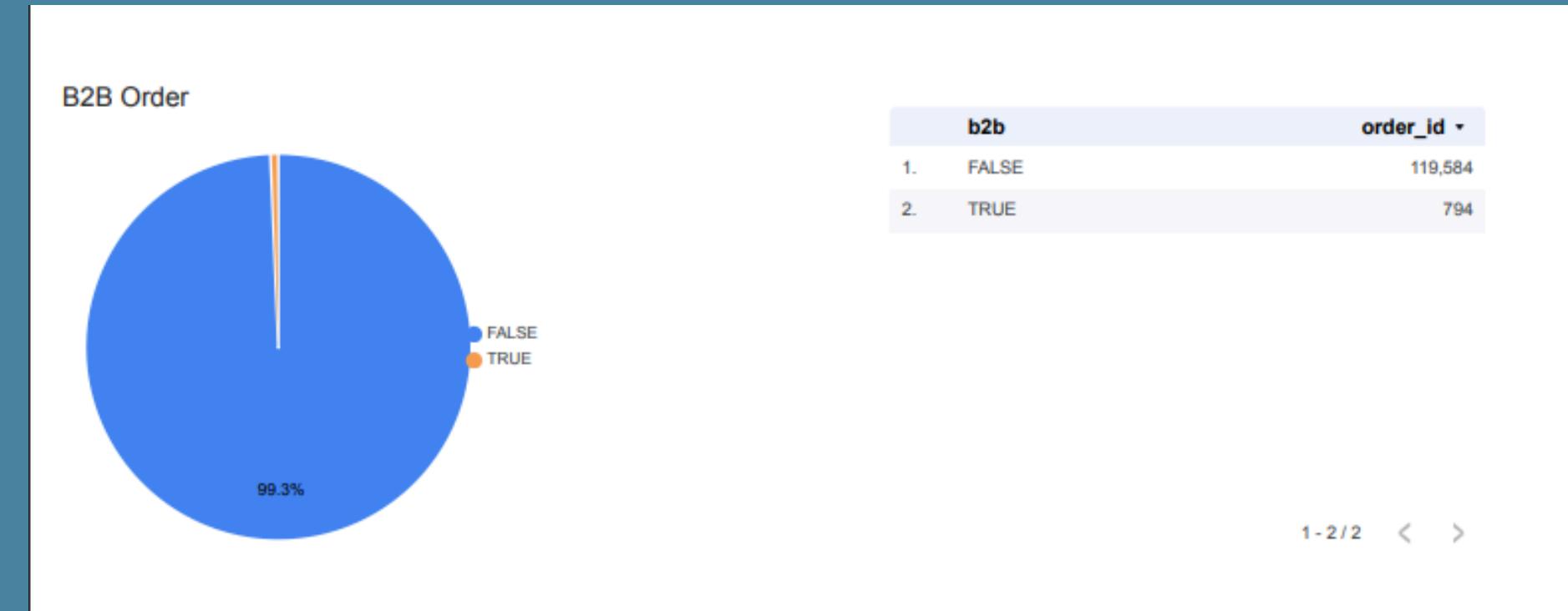


Suspicious Order

- A total of 4,735 orders (3.9%) were flagged as suspicious:
quantity = 0 but amount > 0
- These anomalies were automatically identified and labeled using a custom suspicious flag during data preprocessing.

B2B ORDER

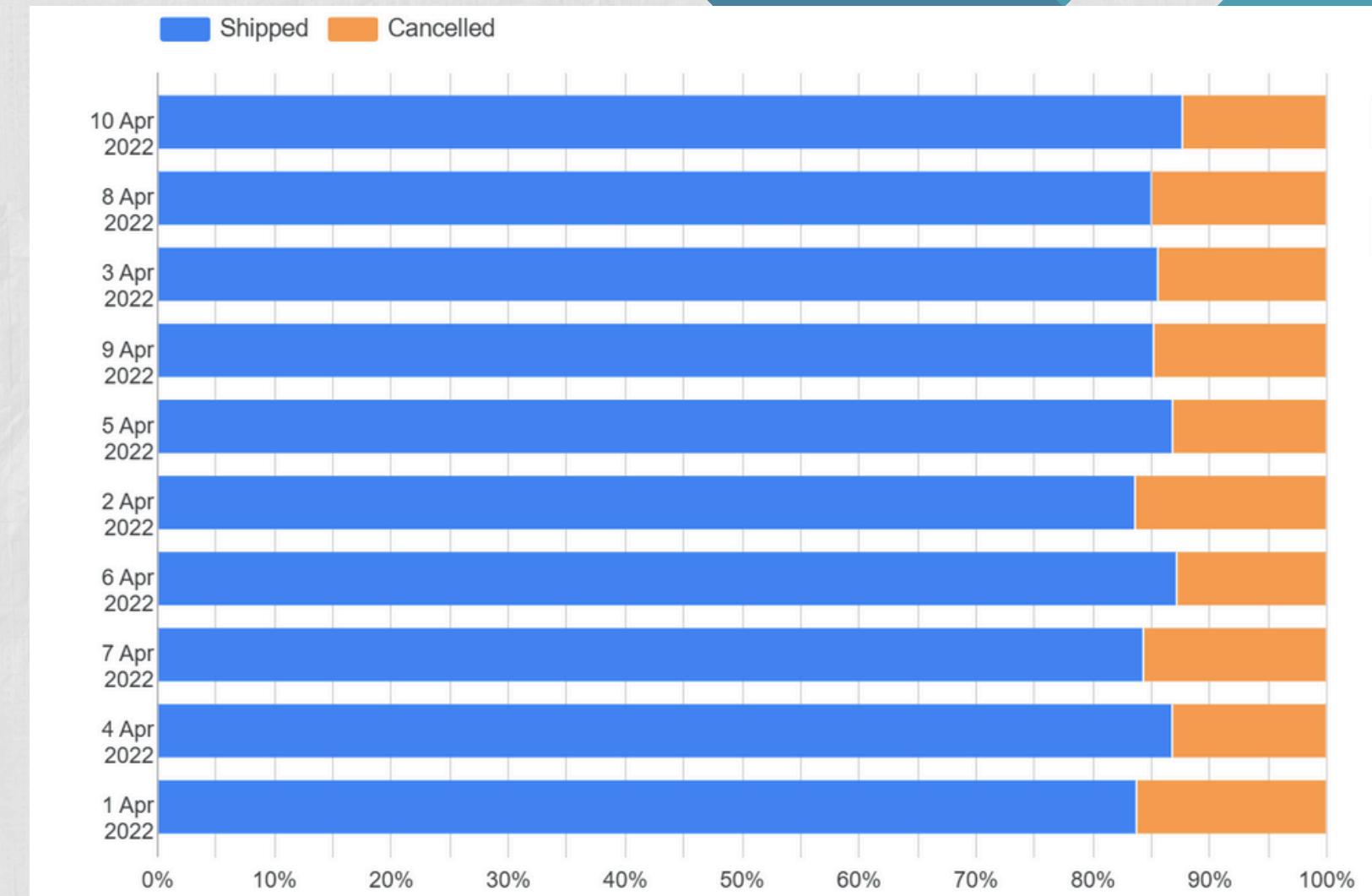
- The vast majority of orders (99.3%) are B2C (retail).
- Only 794 orders (0.7%) are categorized as B2B.





Total Shipped and Cancelled Orders during April Sales period.

From this data, we can see that the highest shipments on April 5th, with a total of 1.303 (86,81%) item shipped and 198 (13,19%) item cancelled. The lowest for April sales period were on April 2nd, with 1.213 (83,6%) item Shipped and 238 (16,4%) item cancelled.



status	Record Count
1. Shipped	101.288
2. Cancelled	17.185

The data shows a total of shipment and Cancelled order. That 101.288 total of shipped order and 17.185 cancelled orders for April - June 2022 Sales Period.



DATA INSIGHT

Sales Performance

- Top Products:
- SKUs JNE3797-KR-L and JNE3405-KR-L contribute the highest order volume and revenue.
- Categories like Set, Kurta, and Dress are the most frequently ordered.

Geographical Insights

- Top Shipping States:
- Maharashtra, Karnataka, and Tamil Nadu lead in total number of orders.
- Recommendation: Focus logistics and marketing in high-performing regions.

Order Fulfillment

- Order Status Breakdown:
- 85.5% of orders are Shipped, while 14.5% are Canceled.
- Suggests solid fulfillment performance, though reasons behind cancellations should be investigated.

Data Anomalies

- 4,735 suspicious orders flagged where qty = 0 but amount > 0.
- Indicates potential data errors or fraud — should be validated or excluded from analysis.

Order Segmentation

- B2C dominates with 99.3% of orders; only 0.7% are B2B.
- Opportunity to consider B2B outreach or tailor strategies for retail-heavy operations.



RECOMENDATION

1. Optimize Inventory for High-Demand SKUs
 - Focus on restocking and forecasting for products like JNE3797-KR-L and JNE3405-KR-L.
 - Bundle top-performing categories (Kurta, Set) in future promotions.
2. Strengthen Logistics in Key Regions
 - Allocate more shipping resources to top-performing states: Maharashtra, Karnataka, Tamil Nadu.
 - Consider warehouse or delivery optimization in high-volume areas.
3. Investigate Suspicious Orders
 - 4,735 orders have qty = 0 but still show positive amount.
 - Set up validation logic during order intake or flag for manual review.
4. Reduce Cancellations
 - Analyze trends behind the 14.5% canceled orders (e.g., stockouts, delivery delays).
 - Improve order tracking communication to reduce buyer drop-off.
5. Explore B2B Potential
 - Although only 0.7% of orders are B2B, this segment can be expanded.
 - Consider launching a targeted campaign or partnership to grow B2B share.
6. Implement Suspicious Order Validation
 - Build automated checks in the data pipeline to flag future orders with:
 - quantity = 0 but amount > 0
 - Review flagged orders periodically to identify root causes (manual entry error, system bug, or fraud).
 - Exclude suspicious orders from revenue metrics in dashboards unless validated.



TERIMAKASIH SQUAD 1

DISUSUN OLEH :
RIFQI RAFIALDY DWIJAYA
WIDYA CAHYANINGRUM
BERRYL CARLOS MANUEL
NURUL HAMIDAH