

sOURCEY

Analysis and Visualization for
Sourcing Request and Quotation Data

Gruop I

Morris
April
Joyce
Venus
Leah

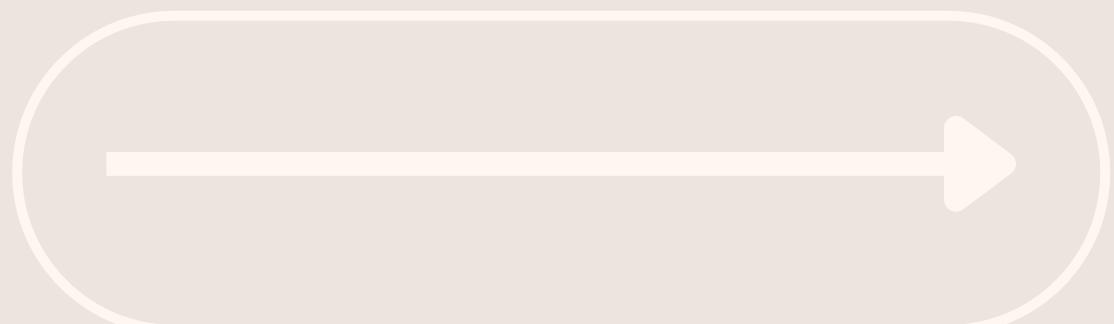
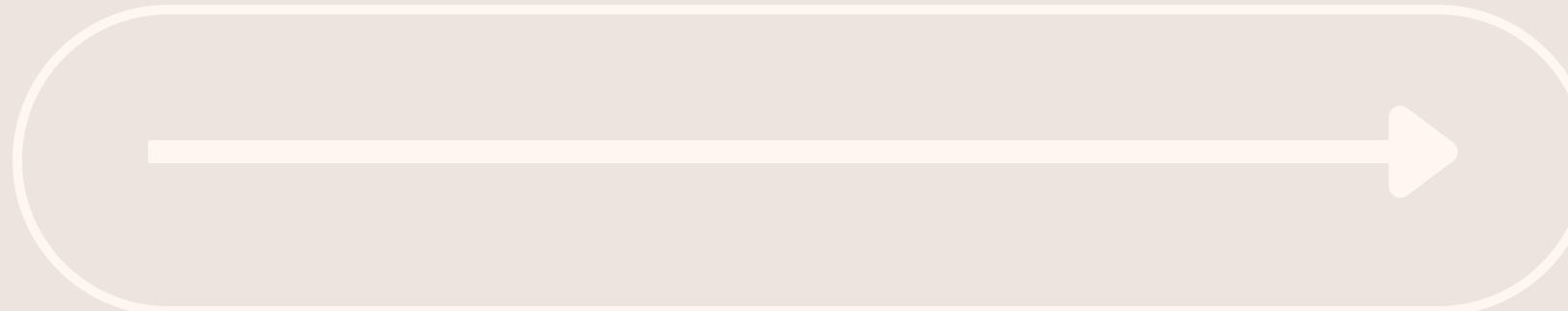




TABLE OF CONTENT

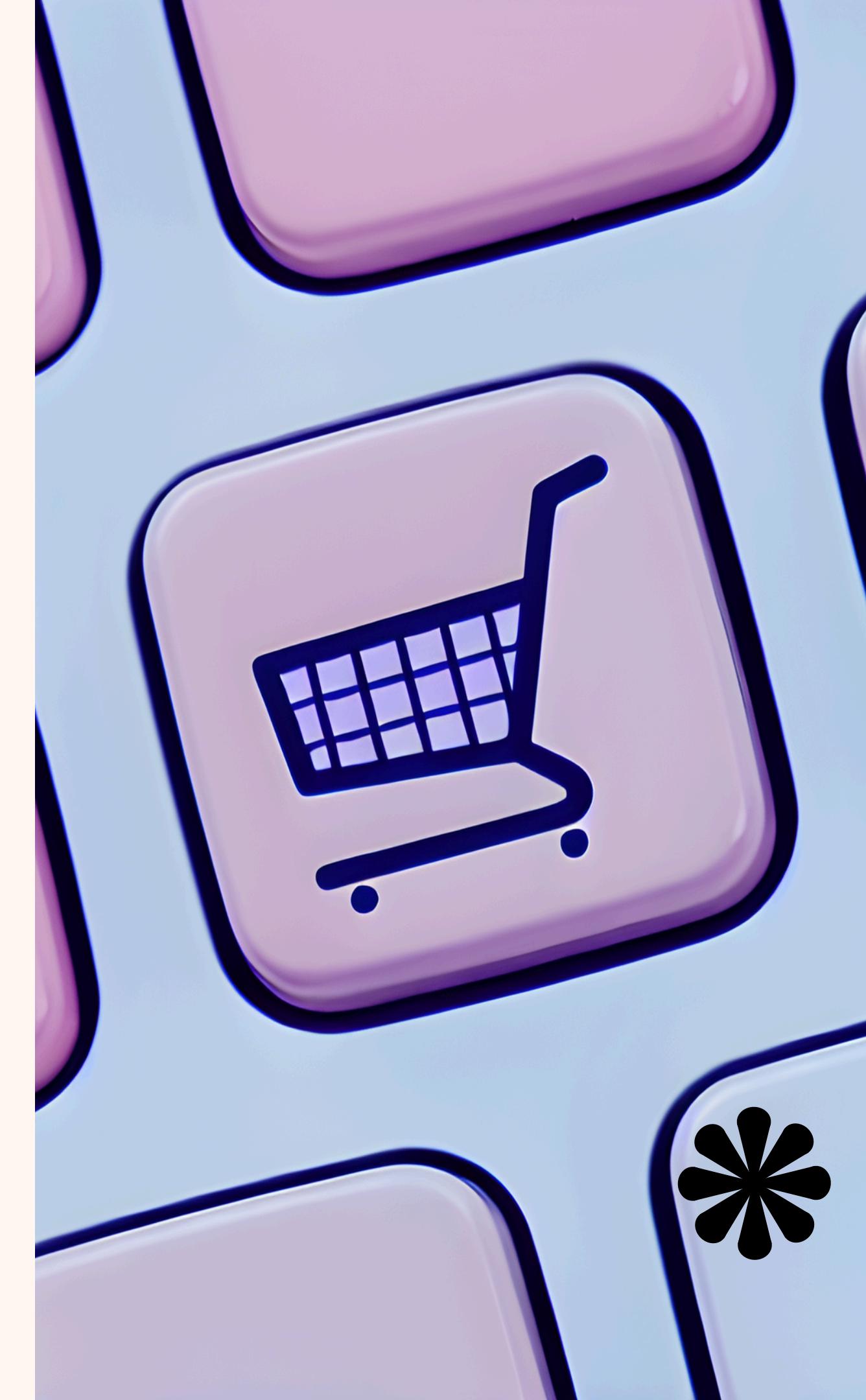
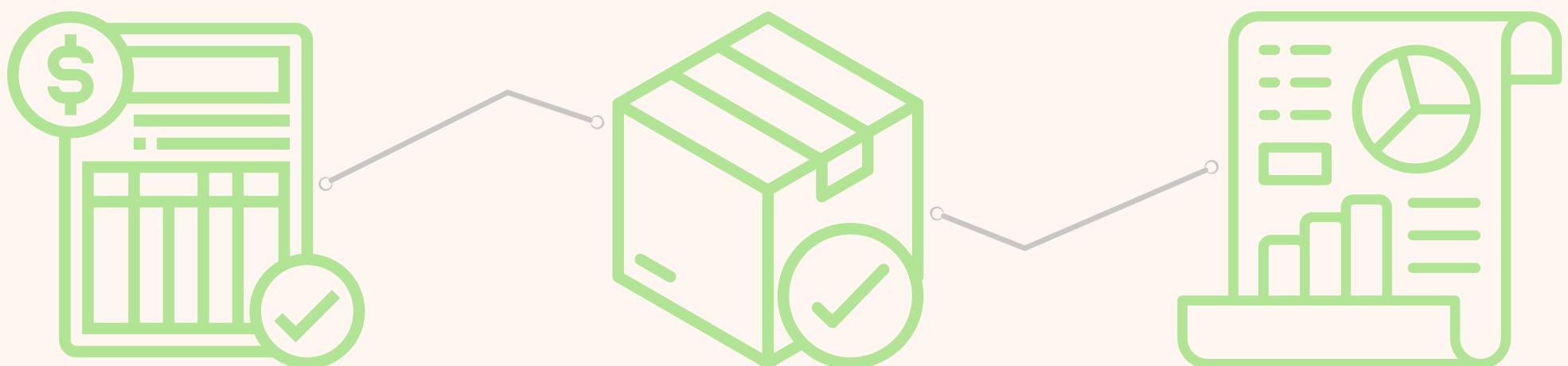


- COMPANY INTRODUCTION
- OBJECTIVES
- DATASET DESCRIPTION
- METHODOLOGY
- ANALYSIS AND VISUALIZATION
- CHALLENGES AND LIMITATIONS
- FUTURE WORK
- CONCLUSION

SOURCEY

Operator of an AI-based sourcing platform intended to digitize global B2B sourcing and international trade.

The company's platform includes features such as free sourcing quotations, data reports, product quality guarantees, customization, and private labeling options, enabling retailers to streamline their sourcing process, reduce costs, and manage their logistics.



OBJECTIVES

01

GMV performance for:

- Gained GMV
- Cost
- Gross Profit



02

Product Categories & SKUs for:

- Quoted Categories
- Purchased Categories



03

Sourcing Request (SR) for:

- No. of SR
- GMV per SR
- No. of average initial quotations per SR
- Status
- No of Product Categories & SKUs per SR



04

Data analytics (DA)

- No of DA request
- No of generated report per DA
- Category types per report



Dataset Description

Table Name	Object ID	Owner
SequelizeMeta	355,381	postgres
analytics_request	1,639,838	postgres
analytics_request_events	1,640,318	postgres
analytics_request_items	1,639,870	postgres
analytics_taxonomy	1,639,905	postgres
catalogues	1,639,479	postgres
cities	1,639,196	postgres
countries	1,639,186	postgres
currency_rates	1,639,690	postgres
customer_addresses	1,639,247	postgres
customer_analysis	1,639,309	postgres
customer_analysis_details	1,639,323	postgres
customer_code_mapping	327,186	postgres
customer_contacts	1,639,285	postgres
customers	1,639,220	postgres
discounts	1,640,076	postgres
employees	1,639,150	postgres
forwarders	1,639,700	postgres
gpt_categorization	1,605,942	postgres
gpt_output	1,640,342	postgres
labels	1,639,461	postgres
platforms	1,639,352	postgres
product_attributes	1,639,621	postgres
product_catalogues	1,639,581	postgres
product_deals	1,639,647	postgres
product_labels	1,639,561	postgres
product_promotions	1,639,597	postgres
product_rating	1,639,671	postgres
product_variants	1,639,535	postgres
products	1,639,498	postgres
quotation_events	1,640,286	postgres
quotation_items	1,640,186	postgres
quotation_shippings	1,661,296	postgres
quotations	1,640,120	postgres
refresh_tokens	1,639,164	postgres
request_item_shipping	326,965	postgres
request_shipping	1,640,041	postgres
shipping_methods	1,639,768	postgres
sourcing_request	1,639,948	postgres
sourcing_request_events	1,640,254	postgres
sourcing_request_items	1,639,992	postgres
supplier_ratings	1,639,421	postgres
supplier_tags	1,639,395	postgres
suppliers	1,639,368	postgres
taxonomies	1,639,444	postgres

The database under consideration comprises a total of 45 tables. The organization utilizes DBeaver as the primary platform for database management and development activities.

The timeframe of the data within this database spans from May 2024 onwards. The tables contain a diverse range of information, including customer details, product specifications, quotation items, sourcing requests, logistics costs, supplier data, and customer-requested data analysis reports.

We have employed SQL scripts within the 'test_dwh' schema to prevent the accidental modification of the original data. This approach ensures the integrity and preservation of the primary data sources while enabling us to safely conduct testing and development activities.

METHODOLOGY

GMV Performance: Gained GMV (Revenue) quotation_approved_final_2

quotations

sourcing_request_id,
delivery_mode,
grand_total_cost_outbound_logistics_air,
grand_total_cost_outbound_logistics_land,
grand_total_cost_outbound_logistics_lcl_ddp,
grand_total_cost_outbound_logistics_fcl_ddp,
grand_total_cost_outbound_logistics_fcl_cif
grand_total_air,
grand_total_land,
grand_total_lcl_ddp,
grand_total_fcl_ddp,
grand_total_fcl_cif

currency_rates

currency_rate_id
rate

	123 sourcing_request_id ▾	123 product_pric ▾	123 product_price_cny ▾	123 total_air ▾	123 total_land ▾	123 total_lcl_ddp ▾
1	42	4,015.582	28,985.002	0	0	7,826.985
2	74	446.9388	3,234	410.454	0	0
3	88	103,408.26	749,661.1	0	0	23,115.473
4	71	254,070.77	35,325.555	1,582,077	0	225,064.06

123 total_revenue ▾	123 total_revenue_cny ▾	123 total_logistics ▾	123 total_logistics_cny ▾
11,842.566	85,481.21	7,826.985	56,496.21
857.3928	6,204	410.454	2,970
126,523.734	917,237.44	23,115.473	167,576.28
2,061,211.8	285,514.34	1,807,141	250,188.81
22,382,660	161,923,328	17,082,126	123,577,552
76.567.3	10,594,616	63,598,125	8,800,072

The dataset comprises 45 records.

In visualization:

- product revenue
- Total revenue
- Logistic revenue

METHODOLOGY

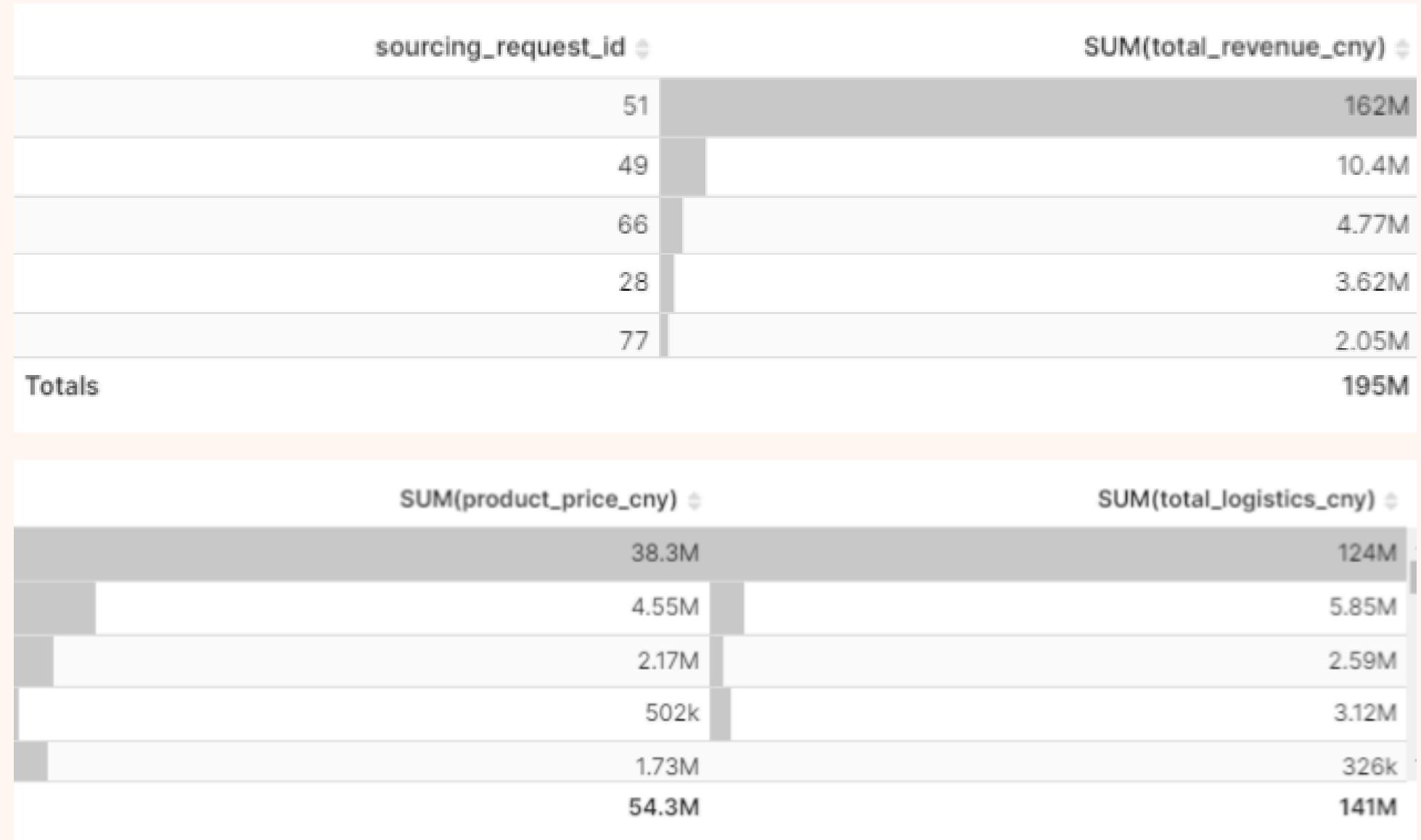
GMV Performance: Gained GMV (Revenue)

quotations

sourcing_request_id,
delivery_mode,
grand_total_cost_outbound_logistics_air,
grand_total_cost_outbound_logistics_land,
grand_total_cost_outbound_logistics_lcl_ddp,
grand_total_cost_outbound_logistics_fcl_ddp,
grand_total_cost_outbound_logistics_fcl_cif
grand_total_air,
grand_total_land,
grand_total_lcl_ddp,
grand_total_fcl_ddp,
grand_total_fcl_cif

currency_rates

currency_rate_id
rate



METHODOLOGY

GMV Performance: aggregated_sourcing_request_costs_1

quotations

sourcing_request_id,
delivery_mode,
status

sourcing_request_items

sourcing_request_id,
total_price_exfactory,
total_cost_outbound_logistics_air,
total_cost_outbound_logistics_lcl_ddp,
total_cost_outbound_logistics_fcl_ddp

123 sourcing_request_id	123 total_price_exfactory	123 total_cost_outbound_logistics_air	123 total_cost_outbound_logistics_lcl_ddp	123 total_cost_outbound_logistics_fcl_ddp	
56	110,768.9375	891,769.8125	135,132.859375	[NULL]	
11	961,875	442,680	74,373.609375	[NULL]	
70	28,700	1,669,150	273,477.5	[NULL]	
44	98,518	46,884.58203125	5,813.78222656	[NULL]	
49	3,047,271	3,485,448.25	551,205.875	35,973,448	
28	328,761.5	511,400.15625	81,406.8046875	[NULL]	
42	111,650	570,000	3,951,360.25	[NULL]	
45	8,343.75	5,032	2,850	[NULL]	

123 total_cost_air	123 total_cost_lcl_ddp	123 total_cost_fcl_ddp	RBC delivery_mode	123 total_costs
982,597.0625	225,960.140625	[NULL]	{AIR,LCL_DDP}	1,137,671.625
1,088,130	719,823.8125	[NULL]	{AIR,LCL_DDP}	1,478,928.625
1,698,205	302,532.5	[NULL]	{AIR,LCL_DDP}	1,971,327.5
105,234.5859375	73,549.78125	[NULL]	{AIR,LCL_DDP}	151,216.359375
6,540,010	3,605,767.75	39,028,000	{AIR,LCL_DDP}	7,083,925

The dataset comprises 45 records.
In visualization:

- Total logistics cost: \$141M
- Item costs: \$54.3M
- Overall total costs: \$195M

METHODOLOGY

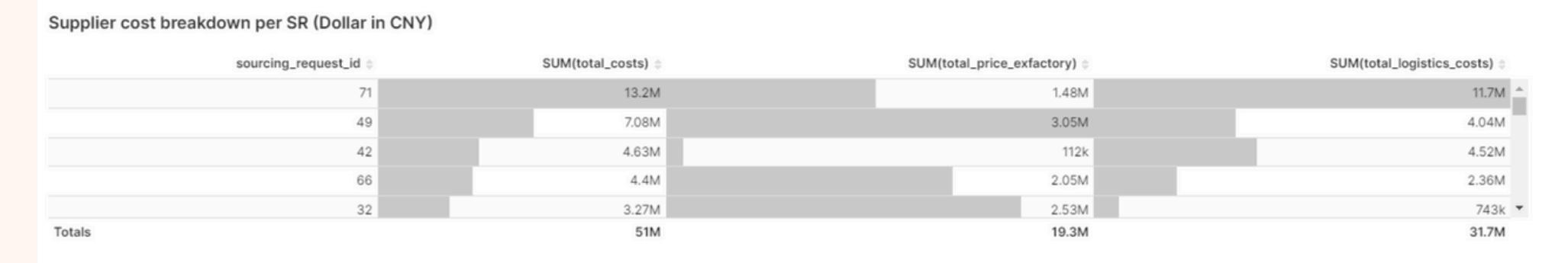
GMV Performance: aggregated_sourcing_request_costs_I

quotations

sourcing_request_id,
delivery_mode,
status

sourcing_request_items

sourcing_request_id,
total_price_exfactory,
total_cost_outbound_logistics_air,
total_cost_outbound_logistics_lcl_ddp,
total_cost_outbound_logistics_fcl_ddp



The dataset comprises 45 records.

In visualization:

- Total logistics cost: \$141M
- Item costs: \$54.3M
- Overall total costs: \$195M

METHODOLOGY

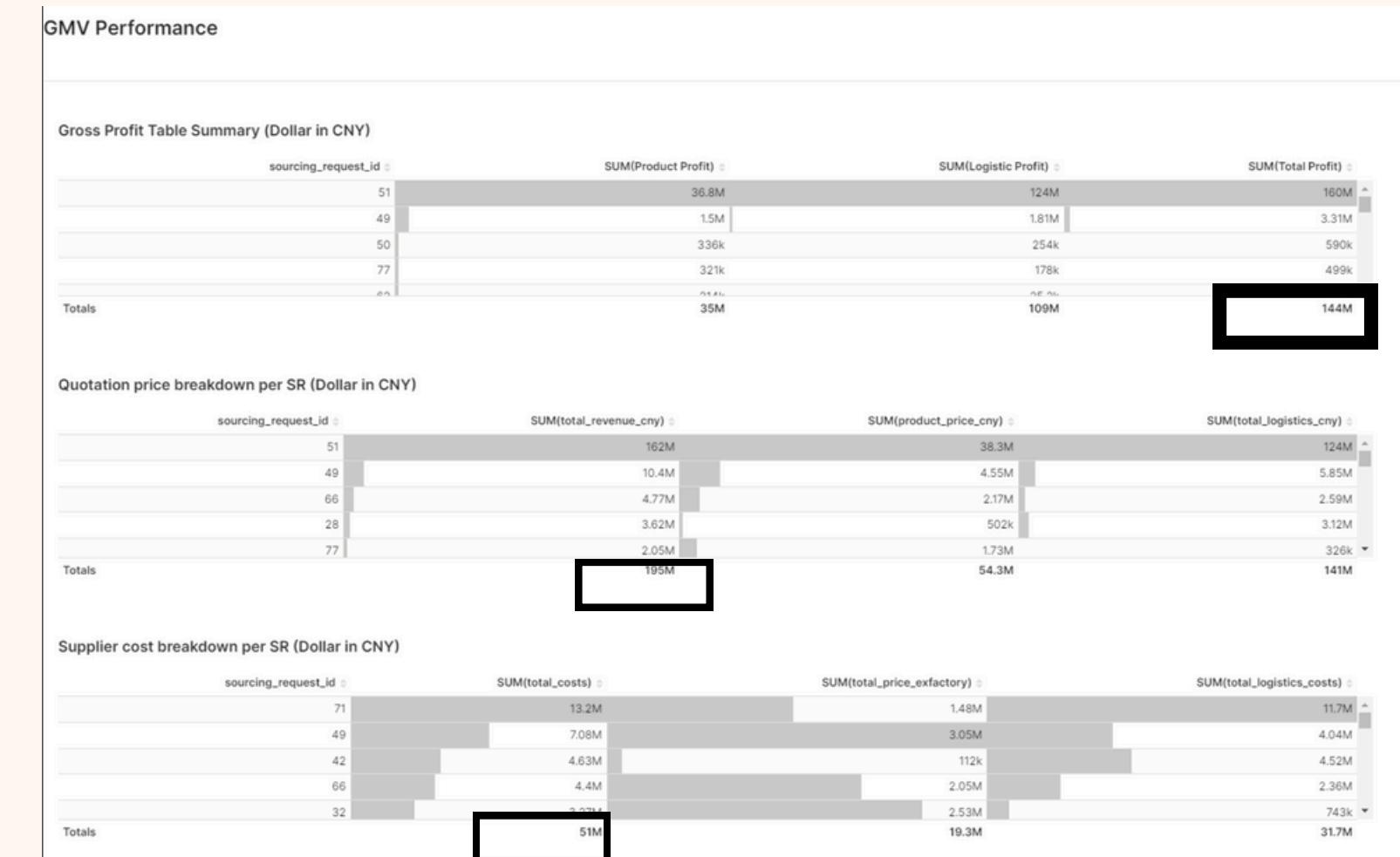
GMV Performance: Gross Profit (Revenue-Cost)

quotations

ID, date, type, status

sourcing_request_items

sourcing_request_id,
total_price_exfactory,
total_cost_outbound_logistics_air,
total_cost_outbound_logistics_lcl_ddp,
total_cost_outbound_logistics_fcl_ddp



The dataset comprises 45 records.

In visualization:

- Items profit: \$35M
- Logistics Profit: \$100M
- Total Profit: \$144M

METHODOLOGY

Product Categories & SKUs: test_skus1

quotations

ID, date, type, status

123 quotation_id	123 quotation_year	123 quotation_month	123 quotation_date	RBC quotation_type	RBC status	123 sourcing_request_item_id	123 quantity
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	363	1
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	364	2
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	365	250
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	366	1
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	367	1
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	368	20

quotation_items

ID

sourcing_request_items

quantity, customerID, productID

123 customer_id	RBC country	RBC subcategory	123 product_id	123 scrape_product_id	123 taxonomy_id	RBC category_level_1	RBC category_level_2	RBC category_level_3
116	PH	Hospitality	124	1,356	1,687	Hardware	Plumbing	Plumbing Fixtures
116	PH	Hospitality	125	1,357	1,687	Hardware	Plumbing	Plumbing Fixtures
116	PH	Hospitality	125	1,357	1,687	Hardware	Plumbing	Plumbing Fixtures
116	PH	Hospitality	126	1,358	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holders
116	PH	Hospitality	127	1,359	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holders
116	PH	Hospitality	128	1,360	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holders
116	PH	Hospitality	129	1,361	595	Home & Garden	Decor	Mirrors

products

scrape product ID

customers_READONLY

countries, subcategory

taxonomies

category level 1,2,3

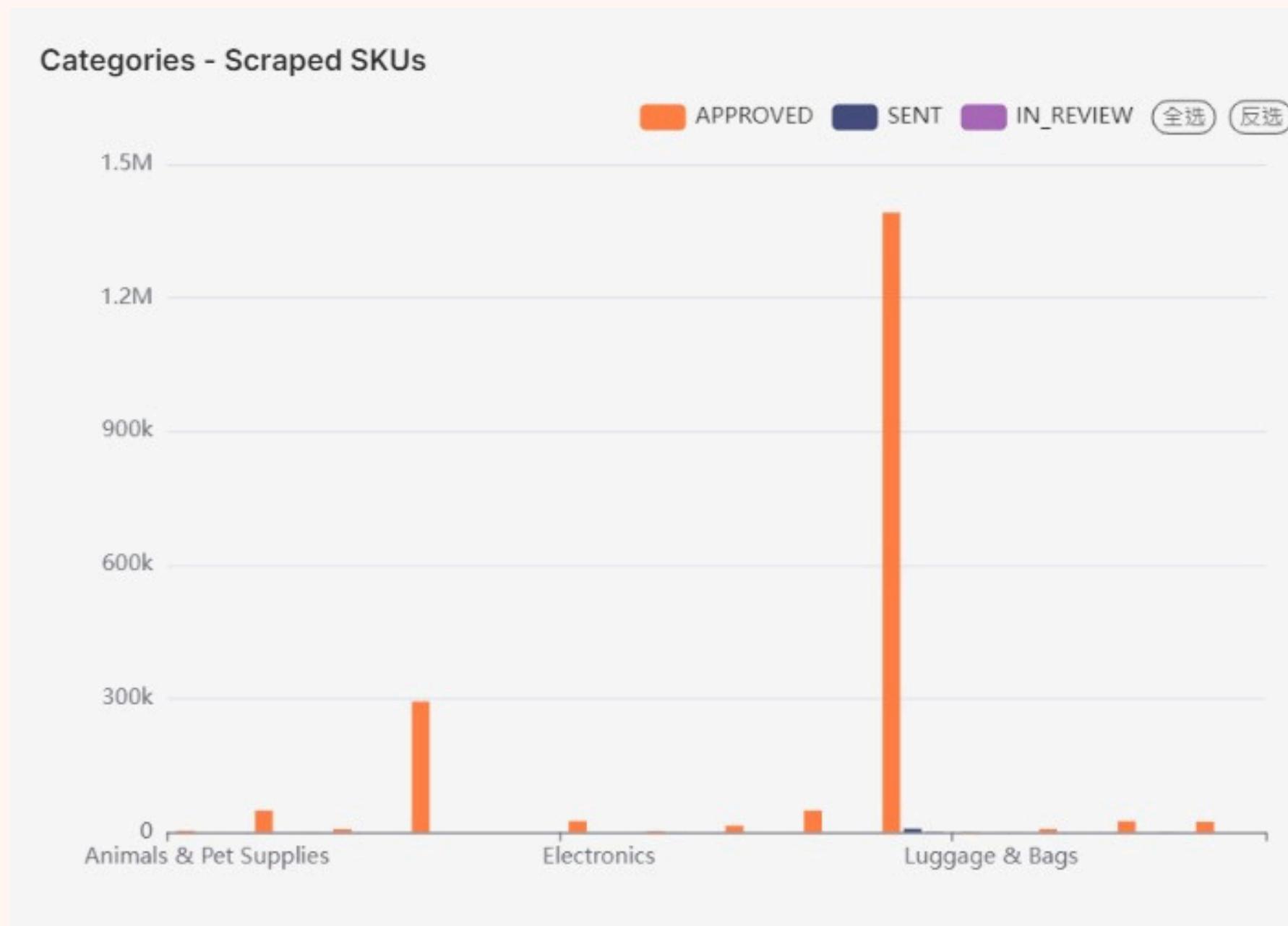
The dataset comprises 3504 records.

In visualization:

- auto scraping/manual scraping
- customer's business category
- Pinpoint the categories that demonstrate the highest popularity

PRODUCT CATEGORIES & SKUS

Quoted & Purchased Scraped Categories SKUs



Quoted Categories (SENT + IN REVIEW):

1. Home & Garden - 7K
2. Furniture - 50
3. Business & Industrial - 7

Purchased Categories (APPROVED):

1. Home & Garden - 1.39M
2. Business & Industrial - 0.29M
3. Apparel & Accessories, Health & Beauty - 48K

PRODUCT CATEGORIES & SKUS

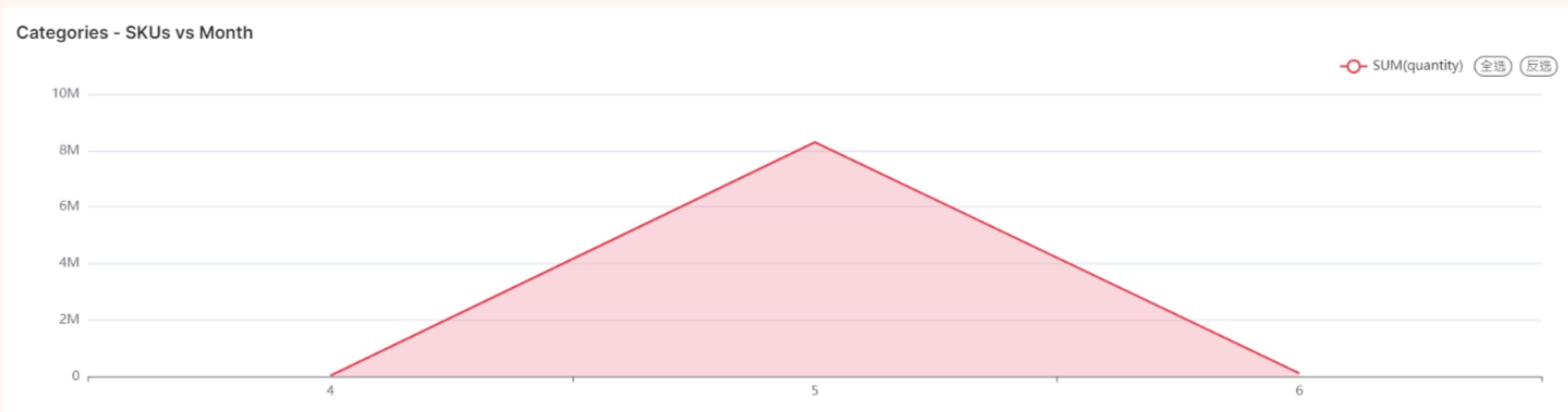
Quoted & Purchased Manual Categories SKUs



- Only **3 Apparel & Accessories** SKUs were quoted in Manual Categories

PRODUCT CATEGORIES & SKUS

Total Product Categories SKUs per Month in 2024



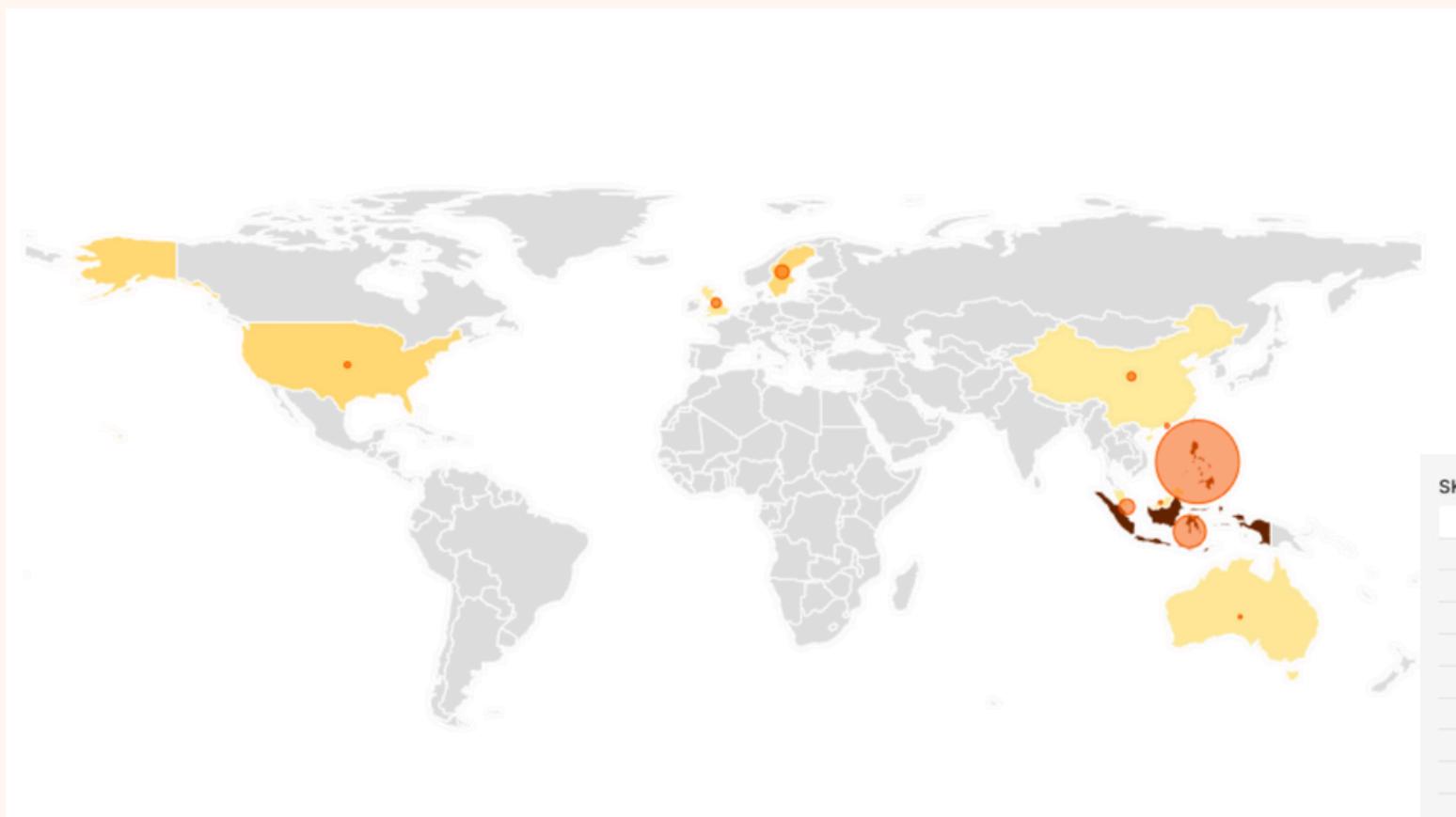
April - 25K SKUs (25,508)

May - 8.2M SKUs (8,293,778)

June - 108K SKUs (108,851)

PRODUCT CATEGORIES & SKUS

Categories vs SKUs by region



- Customer 117 (restaurant business) has the highest total quantity of 619k across multiple product categories, including Home & Garden, Business & Industrial, and Sporting Goods.

- The region with the highest total quantity is Asia, which accounts for the majority of the products across various categories such as Home & Garden, Business & Industrial, Health & Beauty, and Electronics.
- The top 3 product categories by total quantity are:
 - Home & Garden: 460k
 - Business & Industrial: 300k
 - Sporting Goods: 114k

SKU - Group by tables: Categories vs SKUs, customer, region		
customer_id	category_level_1	region
117	Home & Garden	Asia
116	Business & Industrial	Asia
125	Business & Industrial	Asia
128	Home & Garden	Asia
128	Sporting Goods	Asia
116	Home & Garden	Asia
122	Health & Beauty	Asia
66	Home & Garden	Asia
128	N/A	Asia
128	Electronics	Asia
128	Toys & Games	Asia
135	Apparel & Accessories	Asia
118	Home & Garden	Europe
111	Business & Industrial	Asia
65	N/A	Asia
123	Home & Garden	Europe
125	Home & Garden	Asia
67	Apparel & Accessories	Asia
42	Home & Garden	Asia
116	Hardware	Asia
116	Health & Beauty	Asia
115	N/A	Asia
Totals		
		8.43M
		6.19M
		460k
		300k
		136k
		114k
		88.3k
		76.3k
		70k
		67k
		66.6k
		65k
		60.8k
		51k
		50k
		43k
		41k
		37.1k
		36.4k
		35k
		32k
		30.5k

METHODOLOGY

Sourcing Request (SR): sku_n_sourcing_request

sourcing_request

sourcing_request_id, request_date,
request_type, status

sourcing_request_items

sourcing_request_item_id, total_cost_air,
total_cost_air, total_cost_lcl_ddp,
total_cost_fcl_ddp, total_cost_fcl_cif,
cost_currency

customers_READONLY

customer_id

countries

country_name, region

taxonomies

category_level_1

123 sourcing_request_id	① request_date	RBC request_type	RBC status	123 sourcing_request_item_id	123 total_cost_air	123 total_cost_lcl_ddp	123 total_cost_fcl_ddp
49	2024-05-14	INITIAL_HIGH_TOU	INITIAL_QUC	389	183.38000488	151.36746216	343.35140991
6	2024-04-22	SAMPLE	SAMPLE_QU	15	[NULL]	[NULL]	[NULL]
28	2024-04-29	SAMPLE	IN_PROGRES	90	45,010	21,312.625	[NULL]
94	2024-05-29	INITIAL_NON_TOU	INITIAL_QUC	1,131	12,086.66015625	3,337.59033203	[NULL]
6	2024-04-22	SAMPLE	SAMPLE_QU	8	1,702	765	[NULL]
1	2024-04-22	INITIAL_NON_TOU	IN_PROGRES	2	10,061.79980469	98.80000305	[NULL]
1	2024-04-22	INITIAL_NON_TOU	IN_PROGRES	3	10,061.79980469	98.80000305	[NULL]
1	2024-04-22	INITIAL_NON_TOU	IN_PROGRES	1	5,057.79980469	74.89600372	[NULL]
77	2024-05-24	INITIAL_NON_TOU	BACKLOG	828	[NULL]	[NULL]	[NULL]

RBC total_cost_fcl_cif	RBC cost_currency	123 customer_id	RBC country_name	RBC region	RBC category_level_1
	CNY	108	Philippines	Asia	Home & Garden
	CNY	116	Philippines	Asia	Home & Garden
	CNY	108	Philippines	Asia	Home & Garden
	CNY	117	Philippines	Asia	Home & Garden
	CNY	133	Indonesia	Asia	Home & Garden
	CNY	108	Philippines	Asia	Home & Garden
	CNY	67	Philippines	Asia	Apparel & Accessories
	CNY	67	Philippines	Asia	[NULL]
	CNY	67	Philippines	Asia	Apparel & Accessories

The dataset comprises 1161 records.

In visualization:

- No. of SR
- GMV per SR
- No. of Product Categories & SKUs per SR

SOURCING REQUEST

Sourcing Request (SR) - Total no. of SR

99

1. Total 99 sourcing request

Total No of Sourcing Request

SOURCING REQUEST

Sourcing Request (SR) - No. of SR per status in Summary

status	COUNT(sourcing_request_id)
INITIAL_QUOTATION_DONE	29
SAMPLE_QUOTATION_DONE	11
BULK_QUOTATION_DONE	2
Totals	42

Initial_Quotation_Done > Sample_Quotation_Done >
Bulk_Quotation_Done

SOURCING REQUEST

Sourcing Request (SR) - latest SR with status, timestamp

request_date	Metric	COUNT(sourcing_request_id)			Total (Sum)
	status	BULK_QUOTATION_DONE	INITIAL_QUOTATION_DONE	SAMPLE_QUOTATION_DONE	
2024-04-22				1	1
2024-04-24				1	1
2024-04-25				1	1
2024-04-26		1			1
2024-04-29					1
2024-05-02				2	2
2024-05-03				2	2
2024-05-07				1	1
2024-05-08				1	1
2024-05-13				3	3
2024-05-14				1	2
2024-05-15				2	1
2024-05-16				3	1
2024-05-17				1	1
2024-05-20				2	2
2024-05-22		1			1
2024-05-24				2	2
2024-05-27				1	1
2024-05-28				2	2
2024-05-29				3	1
Total (Sum)		2	29	11	42

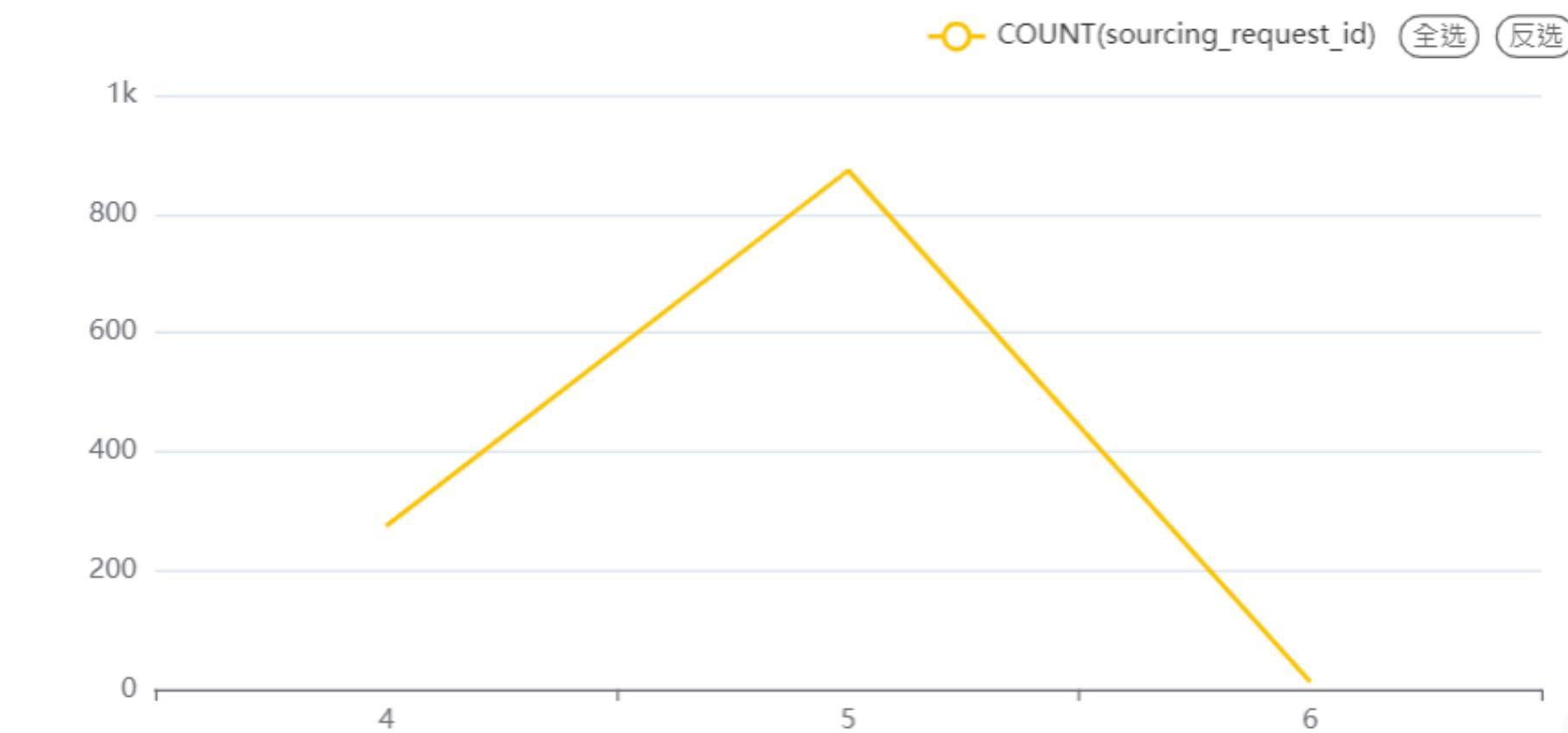
SOURCING REQUEST

No of Product Categories per Sourcing Request & Customer		
sourcing_request_id	customer_id	COUNT_DISTINCT(category_level_1)
66	128	11
88	115	8
30	118	4
3	104	4
21	111	4
57	125	4
56	125	3
39	116	3
71	128	3
49	116	3
Totals		16

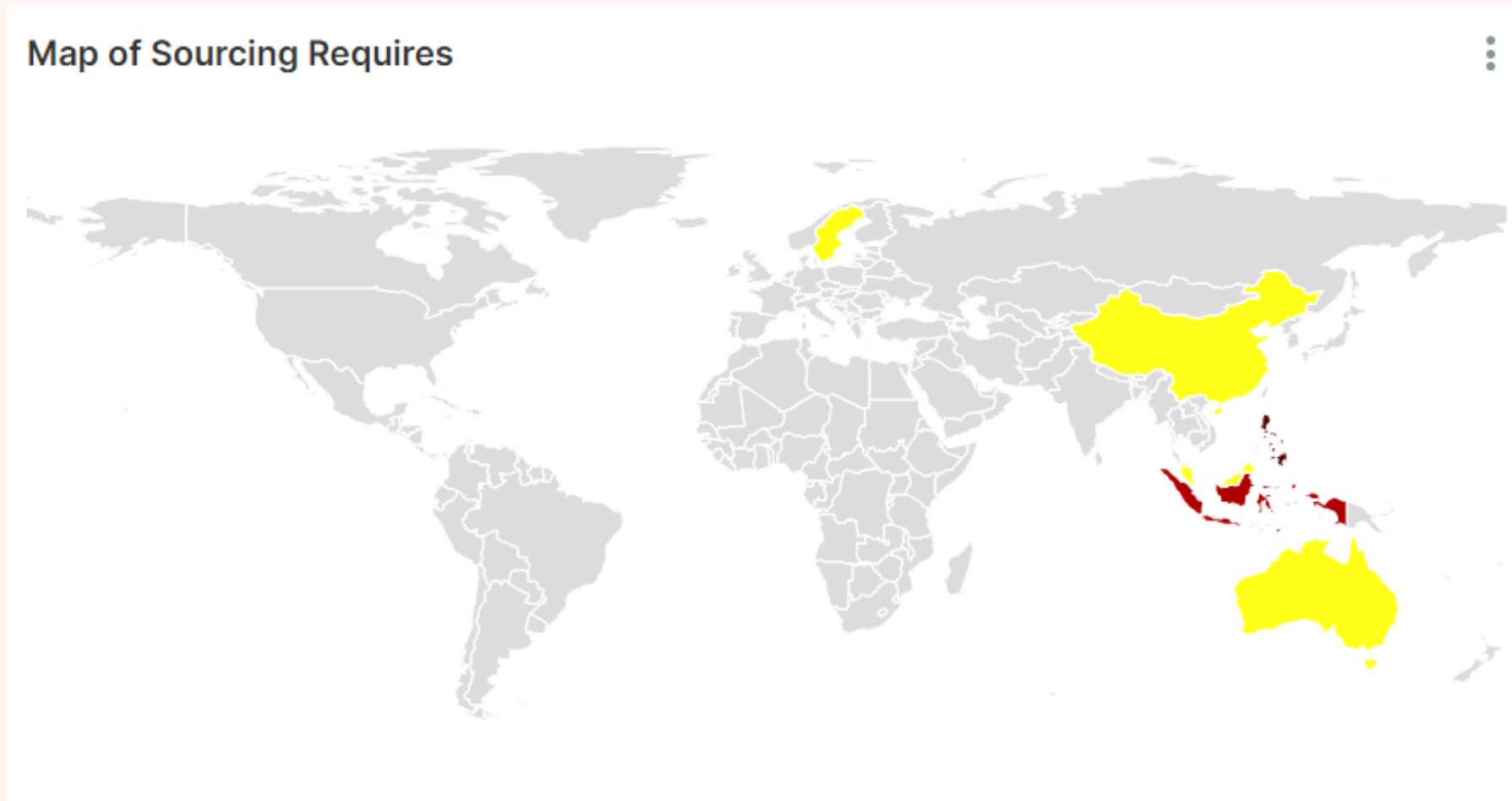
**NO.OF PRODUCT CATEGORIES PER
SOURCING REQUEST & CUSTOMER
TOTAL 16 CATEGORIES**

**NO.OF SR IN MONTH (APRIL -JUNE)
UPTREND
THE HIGHEST SR ORDER IN MAY**

No of Sourcing Request vs Month



SOURCING REQUEST



REGION DISTRIBUTION:
CHINA > SOUTHEAST ASIA >
AUSTRALIA > NORTH AMERICA

**GMV PER SR (INCLUDING
SHIPPING METHODS COST
AND PRODUCT COST
TOTAL 195 MILLION)**

GMV per SR

sourcing_request_id	SUM(total_revenue_cny)
51	162M
49	10.4M
66	4.77M
28	3.62M
77	2.05M
56	1.49M
50	1.35M
39	1.2M
62	1.2M
91	1.04M
Totals	195M

METHODOLOGY

Data analytics (DA): test_da

analytics_request

customer_id, analytics_request_id, request_type, status

analytics_request_events

analytics_request_event_id

analytics_request_items

category

123 quotation_id	123 quotation_year	123 quotation_month	123 quotation_date	RBC quotation_type	RBC status	123 sourcing_request_item_id	123 quantity	123 customer_id
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	364	2	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	365	250	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	366	1	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	367	1	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	368	20	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	369	1	116
184	2,024	5	24	INITIAL_HIGH_TOUCH	REJECTED	372	1	116

RBC country	RBC country_name	RBC region	RBC subcategory	123 product_id	123 scrape_product_id	123 taxonomy_id	RBC category_level_1	RBC category_level_2	RBC category_level_3
PH	Philippines	Asia	Hospitality	124	1,356	1,687	Hardware	Plumbing	Plumbing Fixtures
PH	Philippines	Asia	Hospitality	125	1,357	1,687	Hardware	Plumbing	Plumbing Fixtures
PH	Philippines	Asia	Hospitality	125	1,357	1,687	Hardware	Plumbing	Plumbing Fixtures
PH	Philippines	Asia	Hospitality	126	1,358	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holder
PH	Philippines	Asia	Hospitality	127	1,359	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holder
PH	Philippines	Asia	Hospitality	128	1,360	583	Home & Garden	Bathroom Accessories	Toilet Brushes & Holder
PH	Philippines	Asia	Hospitality	129	1,361	595	Home & Garden	Decor	Mirrors
PH	Philippines	Asia	Hospitality	132	1,365	595	Home & Garden	Decor	Mirrors

The dataset comprises 45 records.

In visualization:

- No. of DA requests
- with and without quotation marks
- identify the most popular categories represented within the data

DATA ANALYTICS

DA - no of report per customer

customer_id	COUNT(analytics_request_event_id)
130	28
115	15
35	2
Totals	45

- Number of Data Analytics requests: 3
- Total number of reports generated: 45
- Customer 130 made the most requests

DATA ANALYTICS



1st Data Analytics Request:

2 reports was cancelled.

2nd Data Analytics Request:

Brain Teaser Puzzle, Eyeshadow, Facial Toners & Astringents, Remote & App-Controlled Robots

3rd Data Analytics Request:

Brain Teaser Puzzle, Eyeshadow, Remote & App-Controlled Robots

DATA ANALYTICS

category	COUNT(analytics_request_event_id)
Lip Balms & Moisturizers	4
Facial Day Care	4
Electric Eyelash Curlers	4
Facial Masks	4
Facial Toners & Astringents	4
Eyeshadow	4
Wrinkle & Anti-Aging Devices	4
Remote & App-Controlled Robots	3
Jigsaw Puzzles	3
Toy Building Sets	3
Totals	45

Within these 45 reports, the category with the most reports per item is "**Lip Balms & Moisturizers**", "**Facial Day Care**", "**Electric Eyelash Curlers**", "**Facial Masks**", "**Facial Toners & Astringents**", "**Eyeshadow**", and "**Wrinkle & Anti-Aging Devices**", each with **4** reports per item.

GMV PERFORMANCE

Gross Profit Table Summary (Dollar in CNY)

sourcing_request_id	SUM(Product Profit)	SUM(Logistic Profit)	SUM(Total Profit)
51	36.8M	124M	160M
49	1.5M	1.81M	3.31M
50	336k	254k	590k
77	321k	178k	499k
69	214k	25.2k	239k
Totals	35M	109M	144M

Gross Profit Table Summary (Dollar in CNY)

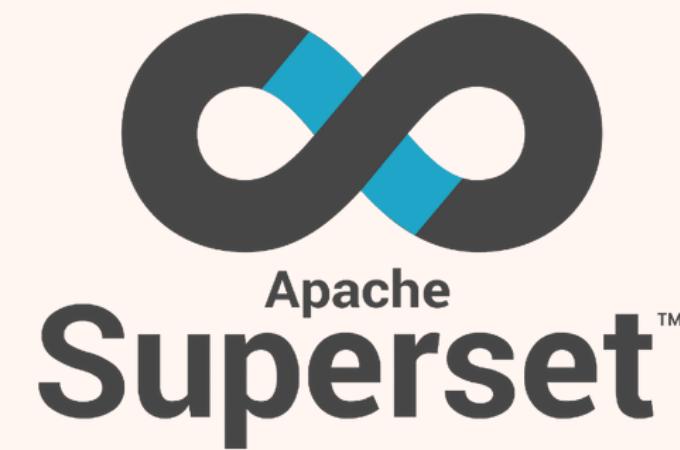
sourcing_request_id	SUM(Product Profit)	SUM(Logistic Profit)	SUM(Total Profit)
70	100k	20.0k	122k
11	-388k	-70.6k	-458k
78	-413k	-164k	-577k
71	-1.45M	-11.5M	-12.9M
32	-2.47M	-727k	-3.2M
Totals	35M	109M	144M

GMV PERFORMANCE

<u>Quotation price breakdown per SR (Dollar in CNY)</u>				
sourcing_request_id	SUM(total_revenue_cny)	SUM(product_price_cny)	SUM(total_logistics_cny)	
51	162M	38.3M	124M	
49	10.4M	4.55M	5.85M	
66	4.77M	2.17M	2.59M	
28	3.62M	502k	3.12M	
77	2.05M	1.73M	326k	
Totals	195M	54.3M	141M	

<u>Supplier cost breakdown per SR (Dollar in CNY)</u>				
sourcing_request_id	SUM(total_costs)	SUM(total_price_exfactory)	SUM(total_logistics_costs)	
71	13.2M	1.48M	11.7M	
49	7.08M	3.05M	4.04M	
42	4.63M	112k	4.52M	
66	4.4M	2.05M	2.36M	
32	3.27M	2.53M	743k	
Totals	51M	19.3M	31.7M	

INTERACTIVE DASHBOARD IN SUPERSET



“ Capstone - Group 1 “

<http://localhost:8088/superset/dashboard/p/v6gnwYlmrVN/>

CHALLENGES AND LIMITATIONS

Time scarcity

- limited time to study the database
- narrow scope for utilizing diverse technical capabilities

Platforms

- early phases of adopting and acclimating to DBeaver and Superset
- Limited Capability for Robust and Comprehensive Dashboards

Large-scale data import

- Missing value when moving database
- Time-consuming
- Data cross-checks

FUTURE WORK

GMV preference

- deeply analysis GMV preference

ETL automation

- create reuseable ETL scripts to automatic transform data

CONCLUSION & RECOMMENDATION

1. Understanding the business model
2. Fulfill all stakeholder's exceptions
3. Provide suitable tools
4. Setup a possible project timetable
5. Communication is important for all
6. Regular review and modify the dashboard

THANK YOU