Optimizing TheLook eCommerce Business

By Mega Oceanna

Business and Problem Understanding

BUSINESS UNDERSTANDING

TheLook is a fictitious eCommerce clothing site developed by the Looker Team. Data Analysis Team here is to provide insights about product, business, and marketing to relevant stakeholders so that the business can be sustained (growing and profitable).

PROBLEM DISCOVERY

Due to potential crisis in 2023, the management team **decided to cut off resources in some categories** with lowest growth within 1 year.

From the problem discovery, we, as a Data Analyst need to give an insight to increase our business performance through:

Analyze and find the low performance product category and present with BCG Matrix

Understand customer behaviour with Cohort Analysis

SQL

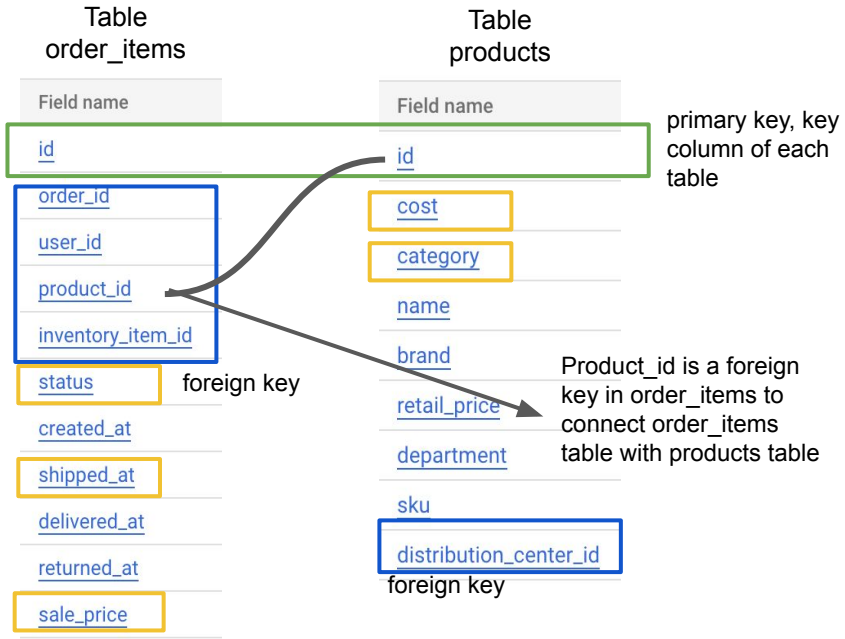
Analyze with SQL

With those following analysis, we need to give insight and recommendation to optimize our business in 2023.

01 Optimizing Category Resources

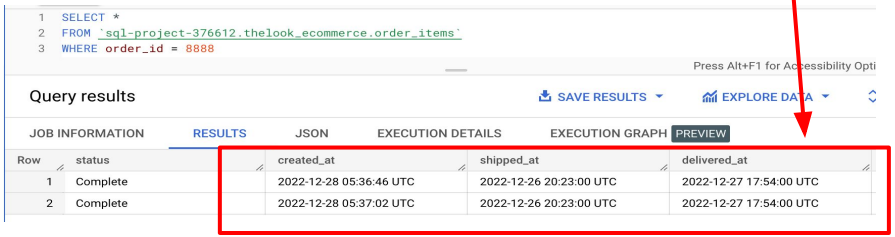
Find the category with lowest profit and revenue growth in 2022

Join 2 tables in SQL



The columns with yellow border are the column we'd like to focus on. Here are the reasons:

- **Sale_price**: to calculate the revenue
- **Cost**: cost of products to calculate profit (sale_price - cost)
- **Category**: to understand the category activity
- **Status**: to make sure the order is Complete (the revenue already processed to us)
- **Shipped_at**: due to the lack of our data, **created_at in order_items shows a delayed update.** To make it more accurate, we will use shipped_at that make sure our product already shipped to the users.



01 Optimizing Category Resources

Find the category with lowest profit and revenue growth in 2022

Table Schema and Table Result

Table Schema

category_growth QUERY SHA

SCHEMA DETAILS PREVIEW LINEAGE P

Filter Enter property name or value

Field name	Type	Mode
category	STRING	NULLABLE
month_num	INTEGER	NULLABLE
month_name	STRING	NULLABLE
revenue_per_month	FLOAT	NULLABLE
cost_per_month	FLOAT	NULLABLE

Table Result

category_growth QUERY SHARE COPY SNAPSHOT DELETE

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Row	category	month_num	month_name	revenue_per_mo	cost_per_month
1	Accessories	5	May	3810.67	1502.31
2	Active	5	May	3060.27	1298.8
3	Blazers & Jackets	5	May	2529.93	963.36
4	Clothing Sets	5	May	69.99	42.06
5	Dresses	5	May	2824.82	1273.75
6	Fashion Hoodies & Sweatshirts	5	May	7015.34	3616.91
7	Intimates	5	May	3370.56	1787.75
8	Jeans	5	May	13854.38	7473.46
9	Jumpsuits & Rompers	5	May	413.91	220.5
10	Leggings	5	May	570.8	342.24
11	Maternity	5	May	1662.99	745.46
12	Outerwear & Coats	5	May	10788.27	4797.08
13	Pants	5	May	3543.84	1621.88
14	Pants & Capris	5	May	2249.99	1178.78
15	Plus	5	May	2058.57	1033.9
16	Shorts	5	May	4300.25	2160.38
17	Skirts	5	May	712.18	281.79

From this table, we can analyze how each product contributes in our revenue and profit (checked by sales history). Therefore, we can put more efforts in significant category

01 Optimizing Category Resources

Find the category with lowest profit and revenue growth in 2022

SQL Syntax

What is/are the categories with the lowest revenue growth in the past 1 year?

1

```
WITH revenue AS (SELECT category, EXTRACT(MONTH
FROM shipped_at) as month_num, FORMAT_DATE('%B',
date(shipped_at)) as month_name, ROUND(SUM(sale_price),2)
as revenue_per_month
FROM `sql-project-376612.thelook_ecommerce.order_items` as
order_items
INNER JOIN `sql-project-376612.thelook_ecommerce.products`
as products
ON order_items.product_id = products.id
WHERE status = 'Complete'
AND date(shipped_at) >= DATE_SUB(DATE '2023-01-01',
interval 1 year) AND date(shipped_at) < '2023-01-01'
GROUP BY 1, 2, 3
ORDER BY 1, 2 DESC),
```

2

```
revenue_lag AS (
SELECT category, month_num, month_name,
revenue_per_month,
LAG(revenue_per_month, 1)
OVER(PARTITION BY category ORDER BY
month_num) as LAG_revenue,
((revenue_per_month -
(LAG(revenue_per_month, 1)
OVER(PARTITION BY category ORDER BY
month_num))) / LAG(revenue_per_month, 1)
OVER(PARTITION BY category ORDER BY
month_num)) * 100 as growth
FROM revenue
ORDER BY 1, 2 DESC)
```

3

```
SELECT revenue.category,
ROUND(AVG(revenue_lag.growth),2) as
average_growth_category
FROM revenue
JOIN revenue_lag
ON revenue.category = revenue_lag.category
AND revenue.month_num =
revenue_lag.month_num
AND revenue.month_name =
revenue_lag.month_name
GROUP BY 1
ORDER BY 2
```

01 Optimizing Category Resources

Find the category with lowest profit and revenue growth in 2022

SQL Syntax

What is/are the categories with the lowest profit growth in the past 1 year?

1

```
WITH profit AS (SELECT category, EXTRACT(MONTH FROM
shipped_at) as month_num, FORMAT_DATE('%B', shipped_at)
as month_name, ROUND(SUM(sale_price - cost),2) as
profit_per_month,
FROM `sql-project-376612.thelook_ecommerce.order_items` as
order_items
INNER JOIN `sql-project-376612.thelook_ecommerce.products`
as products
ON order_items.product_id = products.id
WHERE status = 'Complete'
AND date(shipped_at) >= DATE_SUB(DATE '2023-01-01',
interval 1 year) AND date(shipped_at) < '2023-01-01'
GROUP BY 1, 2, 3
ORDER BY 1, 2 DESC),
```

2

```
profit_lag AS (
SELECT category, month_num, month_name,
profit_per_month,
LAG(profit_per_month, 1) OVER(PARTITION
BY category ORDER BY month_num) as
LAG_profit, ((profit_per_month -
(LAG(profit_per_month, 1) OVER(PARTITION
BY category ORDER BY month_num))) /
(LAG(profit_per_month, 1) OVER(PARTITION
BY category ORDER BY month_num))) * 100
as profit_growth
FROM profit
ORDER BY 1, 2 DESC)
```

3

```
SELECT profit.category,
ROUND(AVG(profit_lag.profit_growth),2) as
average_growth_profit
FROM profit
JOIN profit_lag
ON profit.category = profit_lag.category
AND profit.month_num =
profit_lag.month_num
AND profit.month_name =
profit_lag.month_name
GROUP BY 1
ORDER BY 2
```

01 Optimizing Category Resources

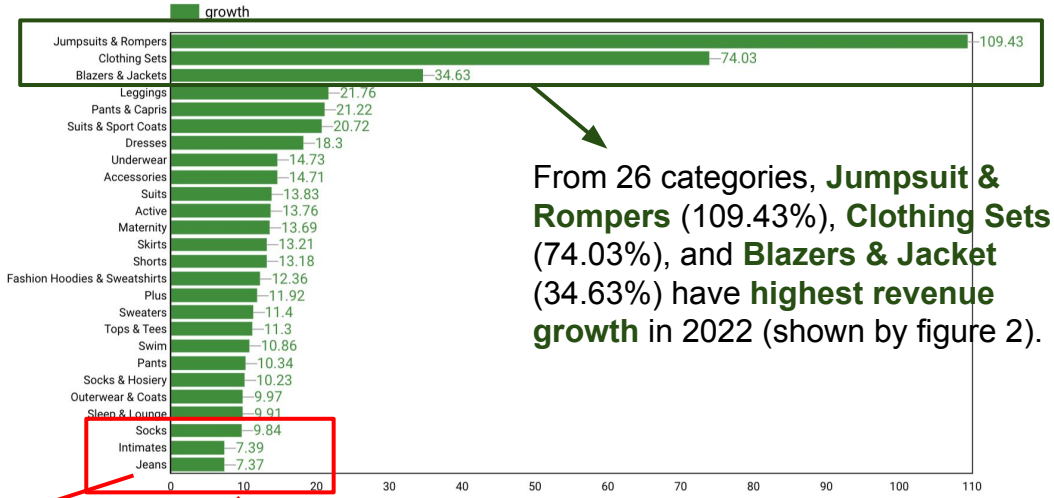
Find the category with lowest profit and revenue growth in 2022

Category with lowest revenue growth

	category	total_revenue
1.	Outerwear & Coats	173,964.51
2.	Jeans	170,914.7
3.	Sweaters	114,583.95
4.	Socks & Hosiery	8,071.25
5.	Jumpsuits & Rompers	3,865.69
6.	Clothing Sets	2,914.62
	Grand total	**474,314.72**

Figure 1. Top 3 and Bottom 3 Categories Based on Total Revenue ($)

Figure 1 shows us that **Outerwear & Coats** ($173.9K), **Jeans** ($170.9K), and **Sweater** ($114.5K) are the top 3 categories with the **highest revenue per month.**



From 26 categories, **Jumpsuit & Rompers** (109.43%), **Clothing Sets** (74.03%), and **Blazers & Jacket** (34.63%) have **highest revenue growth** in 2022 (shown by figure 2).

Figure 2. Average of Monthly Revenue Growth (%) per Category

Categories with the **lowest revenue growth are Socks** (9.54%), **Intimates** (7.39%), and **Jeans** (7.37%).

Jeans category is the lowest revenue growth (7.37%) category in the past 1 year (2022), but the total revenue is the 2nd highest revenue (up to $170.9K).

01 Optimizing Category Resources

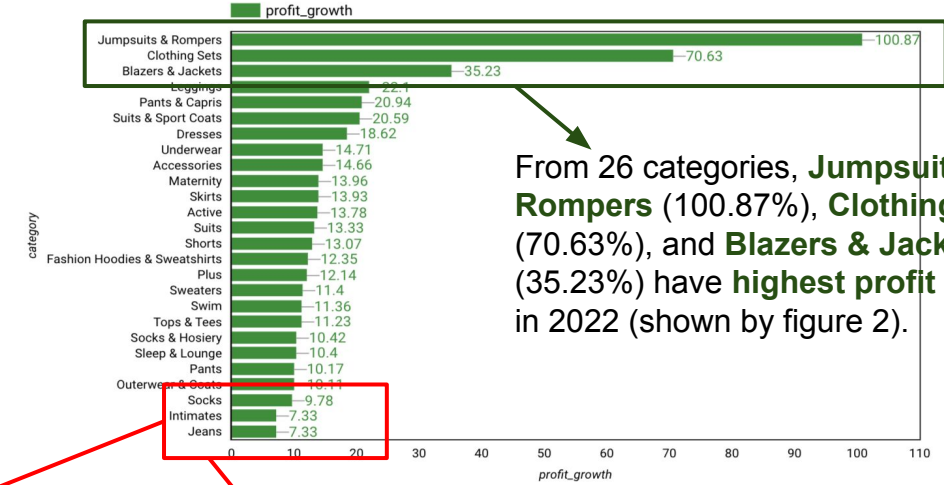
Find the category with lowest profit and revenue growth in 2022

Category with lowest profit growth

	category	total_profit
1.	Outerwear & Coats	96,732.84
2.	Jeans	79,364.29
3.	Sweaters	59,472.33
4.	Socks & Hosiery	4,817.29
5.	Jumpsuits & Rompers	1,783.21
6.	Clothing Sets	1,108.22
	Grand total	**243,278.18**

Figure 1. Top 3 and Bottom 3 Categories Based on Total Profit ($)

Figure 1 shows us that **Outerwear & Coats** ($96.7K), **Jeans** ($79.3K), and **Sweater** ($59.4K) are the top 3 categories with the **highest profit per month.**



From 26 categories, **Jumpsuit & Rompers** (100.87%), **Clothing Sets** (70.63%), and **Blazers & Jacket** (35.23%) have **highest profit growth** in 2022 (shown by figure 2).

Figure 2. Average of Monthly Profit Growth (%) per Category

Categories with the **lowest profit growth are Socks** (9.78%), **Intimates** (7.33%), and **Jeans** (7.33%).

Jeans and Intimates categories are the lowest profit growth (7.33%) categories in the past 1 year (2022), but the total profit of Jeans is the 2nd highest profit (up to $79.3K).

01 Optimizing Category Resources

Find the category with lowest profit and revenue growth in 2022

Revenue Contribution per Category to Total Revenue

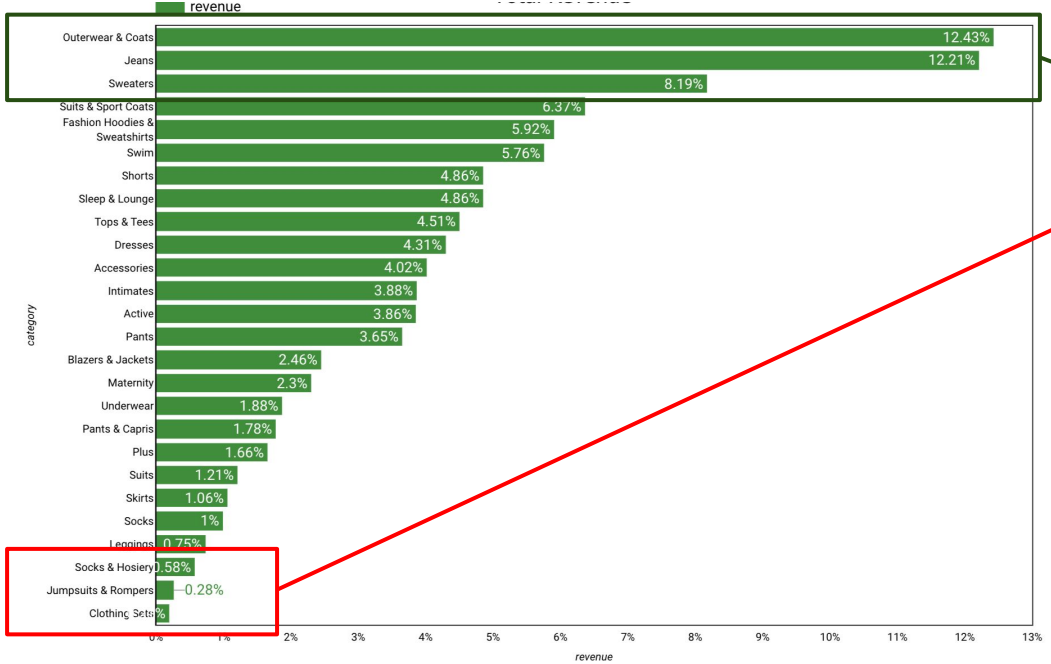


Figure 1. Revenue per Category Compared to Total Revenue in 2022 (Revenue Share)

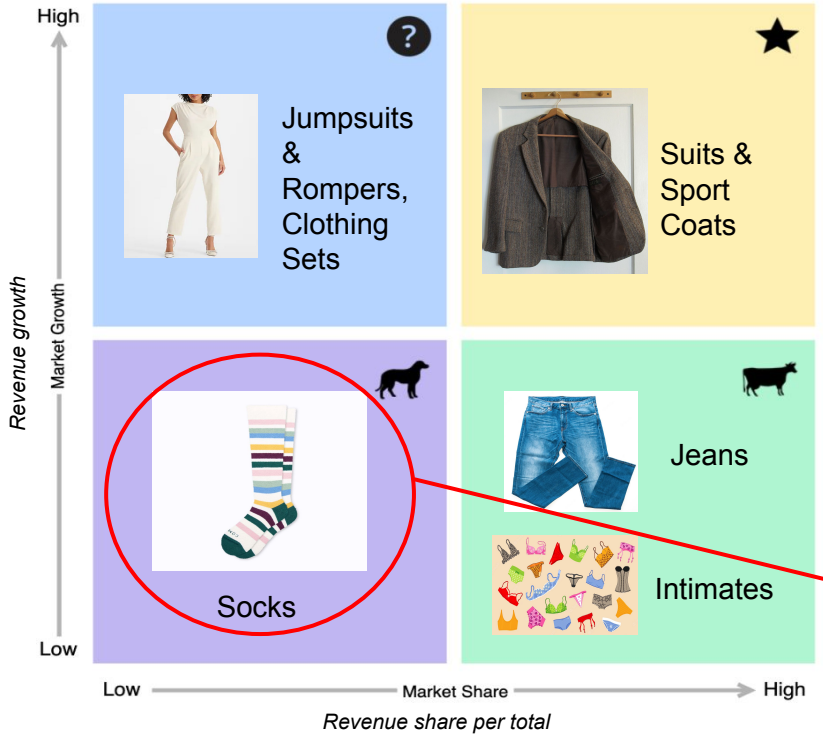
From revenue and profit analysis:

- Outerwear & Coats, **Jeans**, and Sweaters are the **top 3 categories that have the highest total profit and revenue.** Meanwhile, Socks & Hosiery, **Jumpsuits & Rompers, and Clothing Sets** are the **lowest 3 categories to total profit and revenue**.
- **Jumpsuits & Rompers (109.43%), Clothing Sets (74.03%)**, and Blazers & Jackets (34.63%) are the top 3 categories that give the highest profit growth and revenue growth. Meanwhile, **Jeans (7.37%)**, Intimates (7.39%), and Socks (9.54%) are 3 categories with the lowest profit growth and revenue growth.
- The **highlight words show the contrary between total profit & revenue and profit & revenue growth** with the following detail:
  - Jeans has the highest total profit and revenue, but the growth is lowest.
  - Jumpsuits & Rompers and Clothing Sets have the lowest total profit and revenue, but the growth is highest.

01 Optimizing Category Resources

Find the category with lowest profit and revenue growth in 2022

Recommendation with BCG Matrix



- **Star**: Suits & Sport Coats is the category that provides **high revenue** (6.37% of total revenue, the 4th highest) to the company and the **growth** (20.59%) itself **is high**. It brings a good impact to the company revenue.
- **Question Mark**: The **revenue** of Jumpsuits & Rompers (0.28% of total revenue) and Clothing Sets (0.10% of total revenue) are **lowest**, but the **growth** is **highest** (Jumpsuits & Rompers 100.87% and Clothing Sets 70.63%).
- **Cash Cow**: The **revenue** of Jeans (12.21% of total revenue) and Intimates (3.55% of total revenue) are **good**, but the **growth** (Jeans 7.37%, Intimates 7.39%) **is not great** or the demand is constant.
- **Dog**: Socks is the category that has a **low market growth** (9.84%, the 3rd lowest) and it's **not generating a good revenue** (1% of total revenue) to the company.

Based on the category analysis and BCG matrix, Socks is the category that should be deprioritized (low market growth and market share).

02 Understand Customer Retention

Understand the current retention performance with Cohort Analysis

Table Schema and Table Result

Table orders

Field name
order_id
user_id
status
gender
created_at
returned_at
shipped_at
delivered_at
num_of_item

primary key, key column of each table

foreign key

We will only use these 2 columns from one table, no need to join with other tables for **cohort analysis.** We will analyze how many users repurchase after the first purchase month.

- User_id: to calculate how many users in the first month and how many users repurchase in the next month
- Created_at: to analyze the users purchase based on month.

Table Schema

cohort_analys... QUERY SHARE

SCHEMA DETAILS PREVIEW LINEAGE PRE

Filter Enter property name or value

Field name	Type	Mode
cohort_month	DATE	NULLABLE
first_purchase_users	INTEGER	NULLABLE
different_time	INTEGER	NULLABLE
total_users	INTEGER	NULLABLE
percentage	FLOAT	NULLABLE

Table Result

cohort_analys... QUERY SHARE COPY SNAPS

SCHEMA DETAILS PREVIEW LINEAGE PREVIEW

Row	cohort_month	first_purchase_users	different_time	total_users	percentage
1	2022-08-01	4049	0	4049	1.0
2	2022-08-01	4049	1	287	0.07088169...
3	2022-08-01	4049	2	282	0.06964682...
4	2022-08-01	4049	3	283	0.06989380...
5	2022-08-01	4049	4	277	0.06841195...
6	2022-09-01	4265	0	4265	1.0
7	2022-09-01	4265	1	349	0.08182883...
8	2022-09-01	4265	2	341	0.07995310...
9	2022-09-01	4265	3	324	0.07596717...
10	2022-01-01	3444	0	3444	1.0
11	2022-01-01	3444	1	142	0.04123112...
12	2022-01-01	3444	2	166	0.04819976...
13	2022-01-01	3444	3	141	0.04094076...
14	2022-01-01	3444	4	142	0.04123112...
15	2022-01-01	3444	5	149	0.04326364...
16	2022-01-01	3444	6	148	0.04297328...
17	2022-01-01	3444	7	156	0.04529616...
18	2022-01-01	3444	8	159	0.04616724...
19	2022-01-01	3444	9	151	0.04384436...
20	2022-01-01	3444	10	136	0.03948896...

Results per page:

From this table, we can analyze our user behaviour since the first purchase time to the following months in 2022. We can also give insight on how to increase retention rate.

02 Understand Customer Retention

Understand the current retention performance with Cohort Analysis

SQL Syntax

Cohort Analysis

1

```
WITH cohort_item AS (
SELECT user_id, min(date_trunc(date(created_at), month)) as
cohort_month
FROM `sql-project-376612.thelook_ecommerce.orders`
WHERE date(created_at) >= DATE_SUB(DATE '2023-01-01', interval 1
year) AND date(created_at) < '2023-01-01'
GROUP BY 1),

user_activities AS (
SELECT orders.user_id, DATE_DIFF(date_trunc(date(created_at),
month), cohort_month, month) as different_time
FROM `sql-project-376612.thelook_ecommerce.orders` as orders
LEFT JOIN cohort_item as cohort
ON orders.user_id = cohort.user_id
WHERE date(created_at) >= DATE_SUB(DATE '2023-01-01', interval 1
year) AND date(created_at) < '2023-01-01'
group by 1,2),
```

2

```
cohort_size AS (
SELECT cohort_month, COUNT(DISTINCT user_id)
as first_purchase_users
FROM cohort_item
GROUP BY 1
ORDER BY 1),

retention AS (
SELECT cohort_month, different_time,
COUNT(DISTINCT user.user_id) as total_users
FROM user_activities as user
LEFT JOIN cohort_item as cohort
ON user.user_id = cohort.user_id
GROUP BY 1,2)
```

3

```
SELECT r.cohort_month, c.first_purchase_users,
r.different_time, r.total_users,
(r.total_users/c.first_purchase_users) as
percentage
FROM retention as r
LEFT JOIN cohort_size as c
ON r.cohort_month = c.cohort_month
WHERE r.cohort_month IS NOT NULL
ORDER BY 1, 3
```

02 Understand Customer Retention

Understand the current retention performance with Cohort Analysis

Cohort Analysis

different_time / percentage

cohort_month	first_purchase	0	1	2	3	4	5	6	7	8	9	10	11
Jan 1, 2022	3444	100%	4.12%	4.82%	4.09%	4.12%	4.33%	4.3%	4.53%	4.62%	4.38%	3.95%	4.33%
Feb 1, 2022	3118	100%	5.39%	3.72%	4.2%	4.71%	4.81%	4.55%	4.59%	4.94%	4.36%	4.75%	-
Mar 1, 2022	3498	100%	4.89%	4.95%	4.49%	5.03%	4.6%	4.72%	5%	5.29%	5.55%	-	-
Apr 1, 2022	3582	100%	4.83%	5.05%	5.11%	5.64%	5.58%	4.77%	4.89%	5.05%	-	-	-
May 1, 2022	3642	100%	5.11%	5.6%	6.04%	5.16%	5.19%	6.21%	5.08%	-	-	-	-
Jun 1, 2022	3669	100%	6.51%	5.4%	5.78%	5.81%	5.07%	5.42%	-	-	-	-	-
Jul 1, 2022	3946	100%	6.59%	6.44%	6.34%	6.39%	6.51%	-	-	-	-	-	-
Aug 1, 2022	4049	100%	7.09%	6.96%	6.99%	6.84%	-	-	-	-	-	-	-
Sep 1, 2022	4265	100%	8.18%	8%	7.6%	-	-	-	-	-	-	-	-
Oct 1, 2022	4490	100%	9.04%	9.13%	-	-	-	-	-	-	-	-	-
Nov 1, 2022	4818	100%	11.62%	-	-	-	-	-	-	-	-	-	-
Dec 1, 2022	5405	100%	-	-	-	-	-	-	-	-	-	-	-

Cohort_month increases, retention rate is also higher

Highest falling percentage from first purchase time (0) to the next month (1)

Percentage of customers do the repurchase in the following months after first purchase time (0)

- The retention rate for each cohort_month is different. As the **cohort_month increases, the retention rate is also higher**.
- From the total of users in each month, it is **only around 3% - 11% of users do the repurchase**, meaning that we have a higher churn rate (inversion of retention) instead of retention rate.
- The **highest falling percentage occurred in the first month after first purchase time** that can be happened because of low customer satisfaction rate from the first buying experience. It can be happened due to, but not limited to, our shipping time, product attraction, trouble in our source, and market competition.
- Therefore, it **needs collaboration from logistics team, product team, engineer team, and business team** to improve our product and service to maintain our users.

THANK YOU