THE LOOK

E-COMMERCE

STRUCTURED QUERY LANGUAGE

BY ISMATUL MAULA

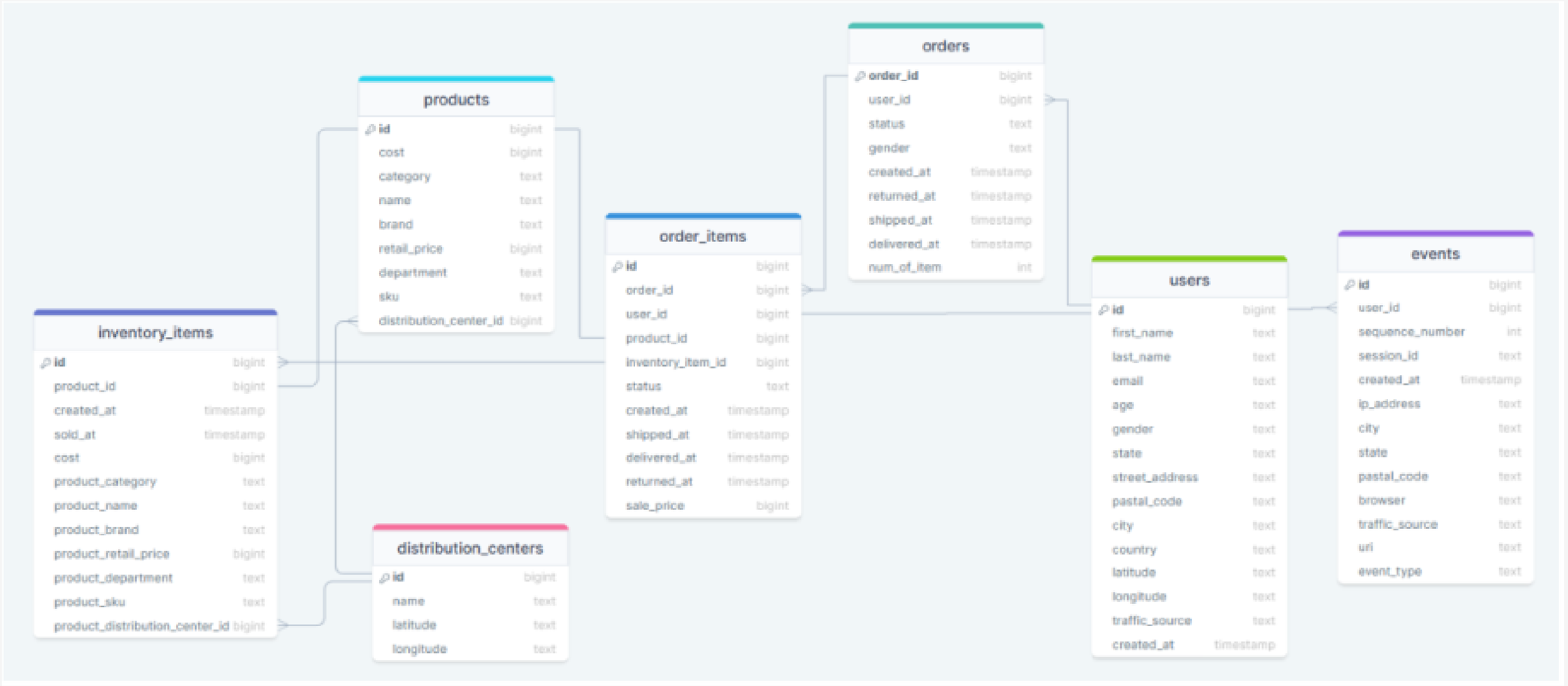
BUSINESS BACKGROUND

A Data Analyst (in e-Commerce clothing named TheLook) asked to provide insights about product, business, and marketing by understanding about the business current performance (whether the business is growing, steady, or dropping) and knowing key drivers breakdown per variables causing the drop/growth of the business in order to adjust the new strategy that can boost the business performance in the future

The Look is a fictitious eCommerce clothing site developed by the Looker team. The dataset contains information about customers, products, orders, logistics, web events and digital marketing campaigns. The contents of this dataset are synthetic, and are provided to industry practitioners for the purpose of product discovery, testing, and evaluation.

ABOUT DATASET

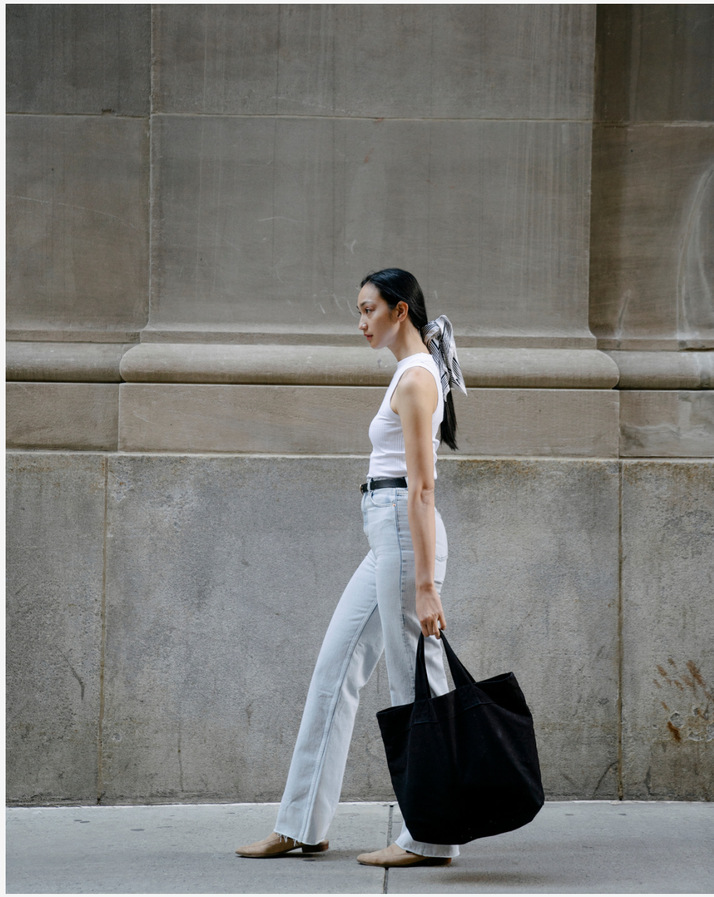
ERD


products
id bigint
cost bigint
category text
name text
brand text
retail_price bigint
department text
sku text
distribution_center_id bigint
orders
order_id bigint
user_id bigint
status text
gender text
created_at timestamp
returned_at timestamp
shipped_at timestamp
delivered_at timestamp
num_of_item int
order_items
id bigint
order_id bigint
user_id bigint
product_id bigint
inventory_item_id bigint
status text
created_at timestamp
shipped_at timestamp
delivered_at timestamp
returned_at timestamp
sale_price bigint
users
id bigint
first_name text
last_name text
email text
age text
gender text
state text
street_address text
postal_code text
city text
country text
latitude text
longitude text
traffic_source text
created_at timestamp
events
id bigint
user_id bigint
sequence_number int
session_id text
created_at timestamp
ip_address text
city text
state text
postal_code text
browser text
traffic_source text
uri text
event_type text
inventory_items
id bigint
product_id bigint
created_at timestamp
sold_at timestamp
cost bigint
product_category text
product_name text
product_brand text
product_retail_price bigint
product_department text
product_sku text
product_distribution_center_id bigint
distribution_centers
id bigint
name text
latitude text
longitude text

TOOLS

The dataset is restored in Google Big Query. In order to answer the business problem we conduct the analysis using query. Therefore we use Google Big Query as the tools.

INTERMEDIATE LEVEL



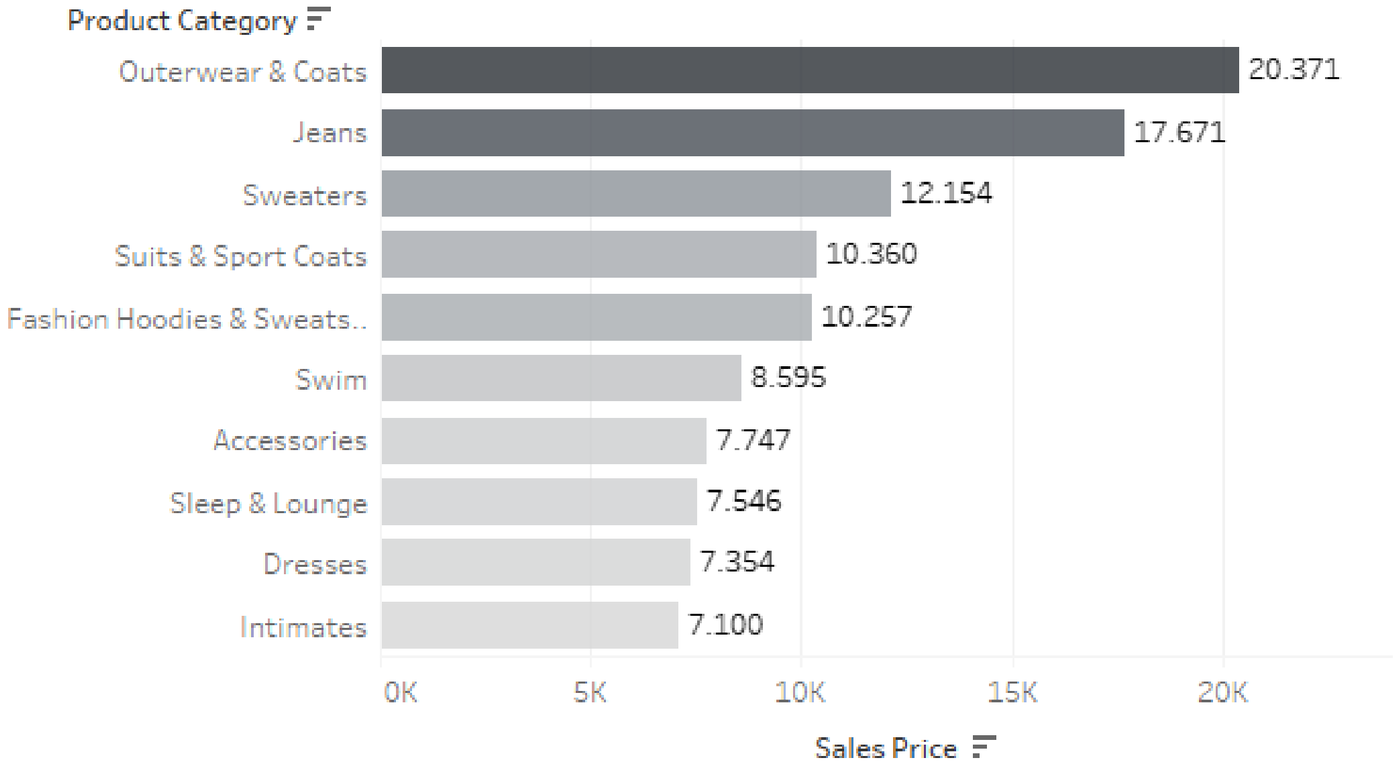
BUSINESS CASE

QUESTIONS

- Which categories that generate the most sales price with status shipped on dec 2022?
- Which category faced lowest % (most shrinking) in term of order number with status shipped on dec 2022 compared to previous month?
- What is the best traffic source to get a session on the look ecommerce on dec 2022?
- What top 3 users first traffic sources that provide the most revenue in the past 3 months?

BEST SELLER CATEGORY

>> Link to Query<<

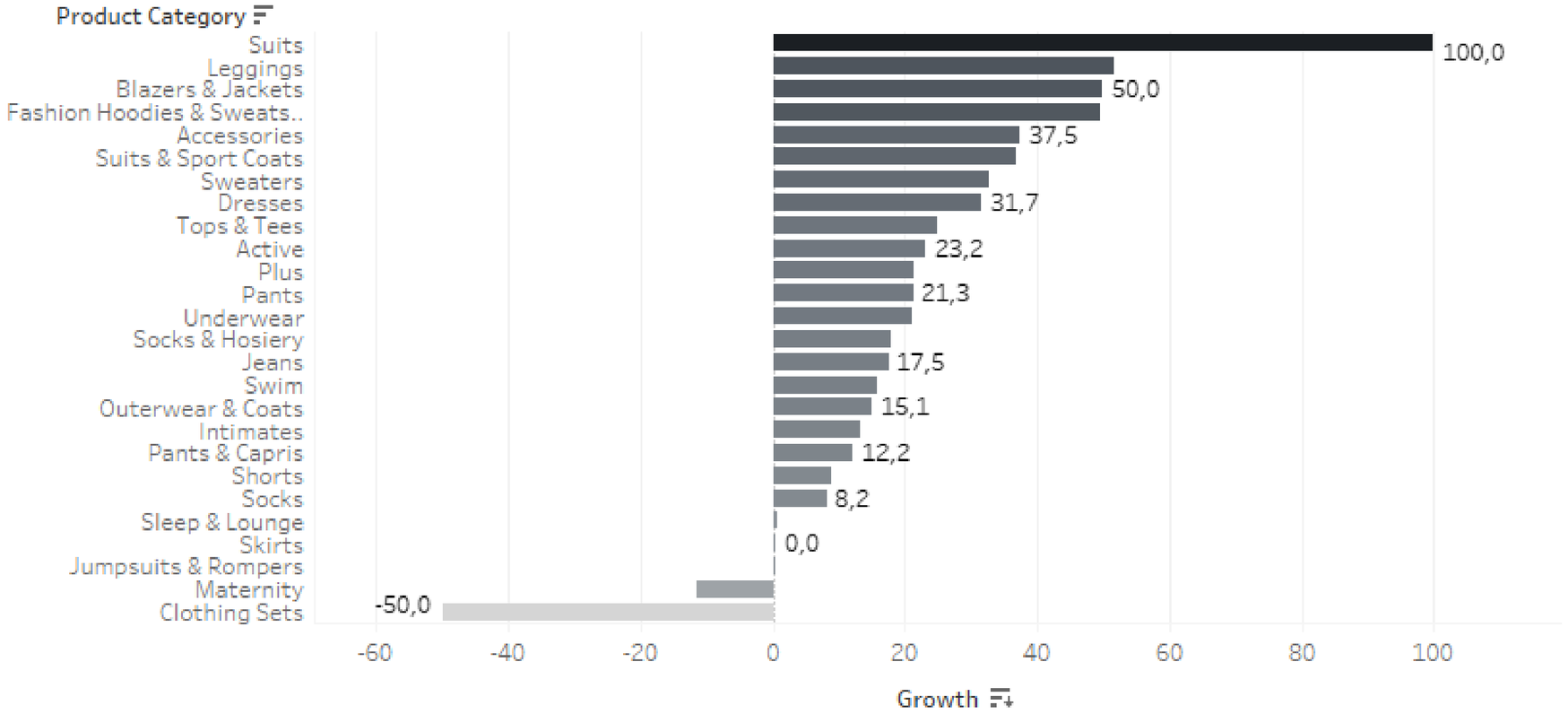


The are Top 10 product categories that being a best seller on December 2022.

Outerwear & Coats is the most favorite one that generate the most sales prices with total sales 20.371

THE MOST SHRINKING CATEGORY

>> Link to Query<<

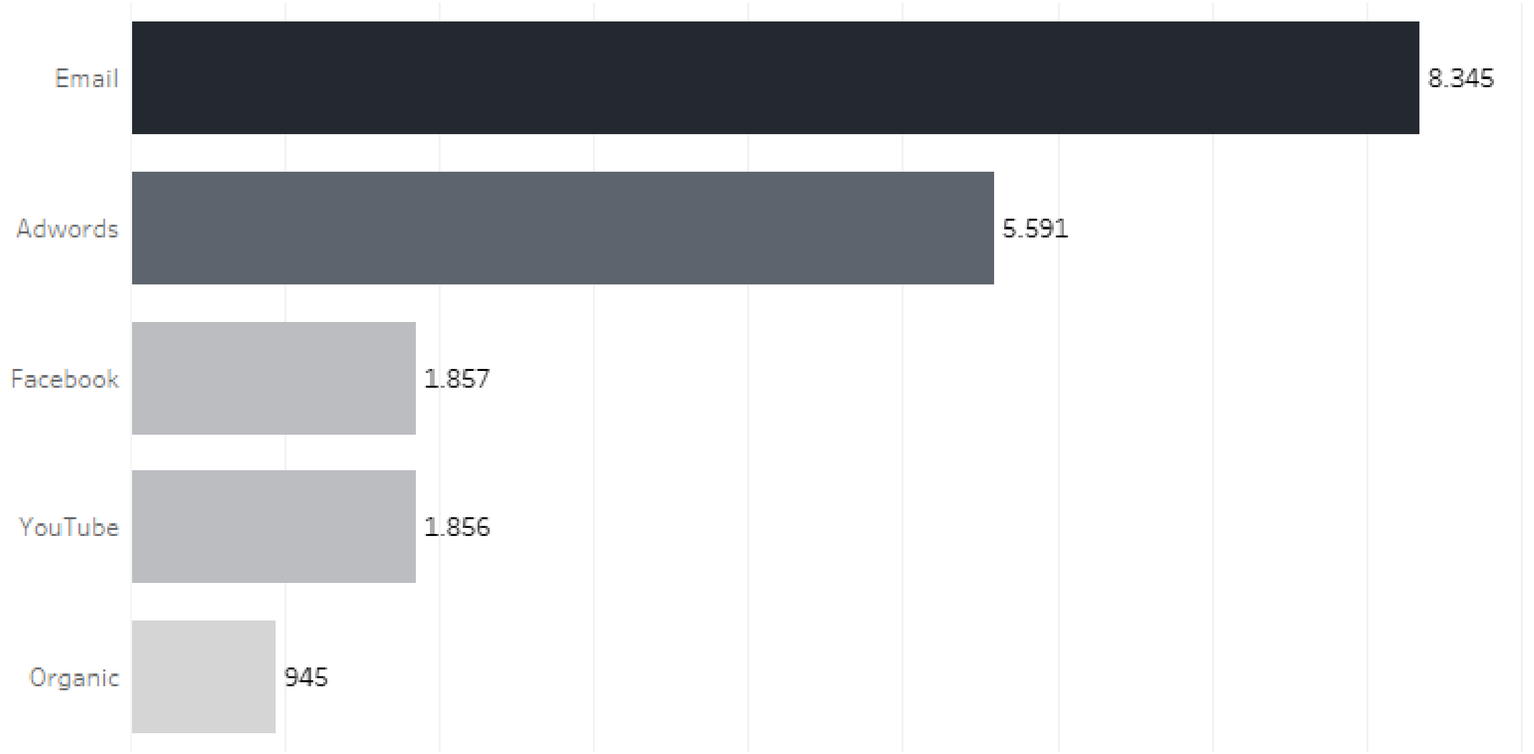


Compared to the previous month, As we look at the chart there are 2 product categories are the most shringking.

Clothing Sets is product category that faces the lowest order on December 2022

BEST TRAFFIC SOURCE

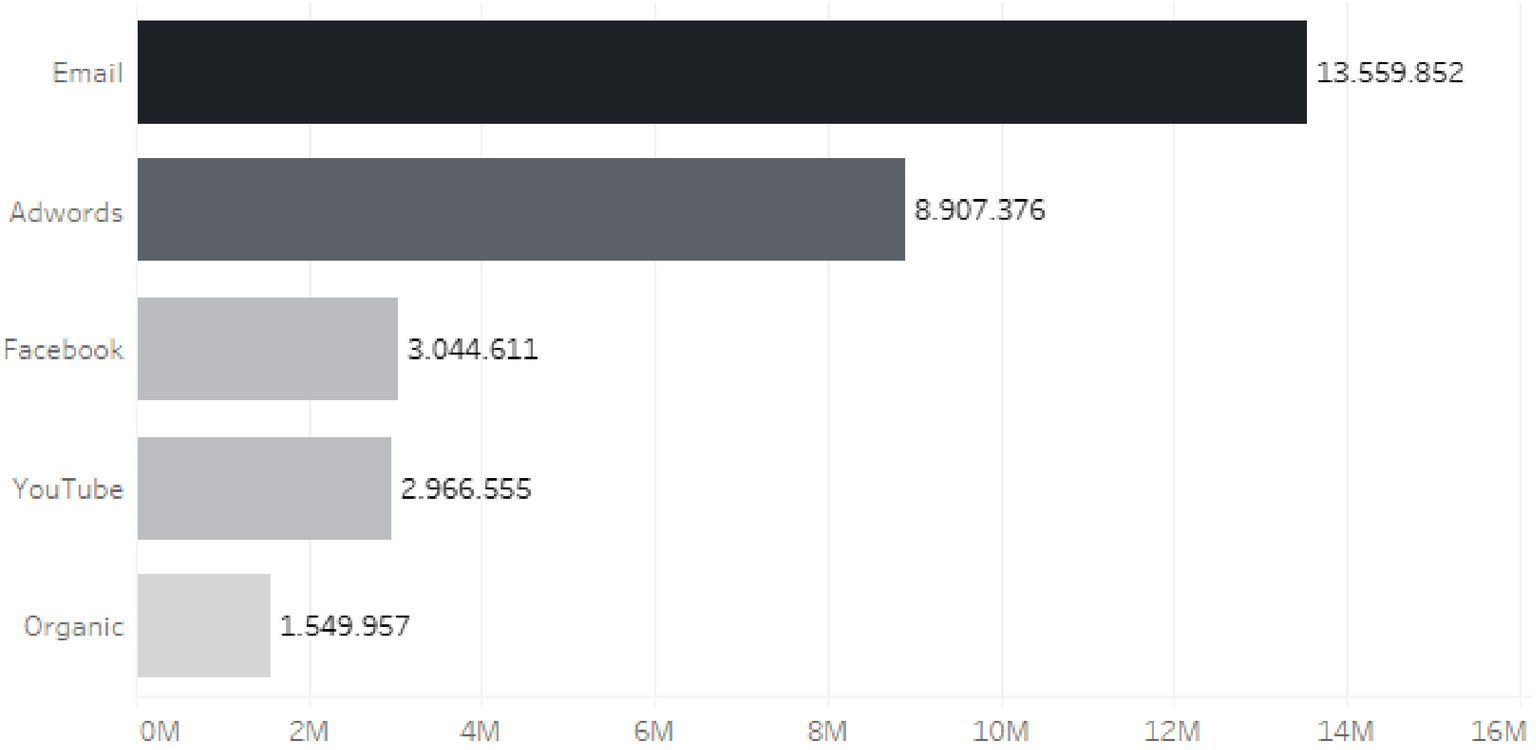
>> Link to Query<<



From the chart we can see that there are 5 best traffic sources which are Email that provide the highest contribution with total 8.345 followed by Adwords, Facebook Youtube and Organic

TRAFFIC SOURCE WITH THE MOST REVENUE

>> Link to Query<<



There are 5 traffic sources that give most revenue to the company in the past 3 months. The first source is E-mail that provide revenue until 13.559.852

LEVEL ADVANCED

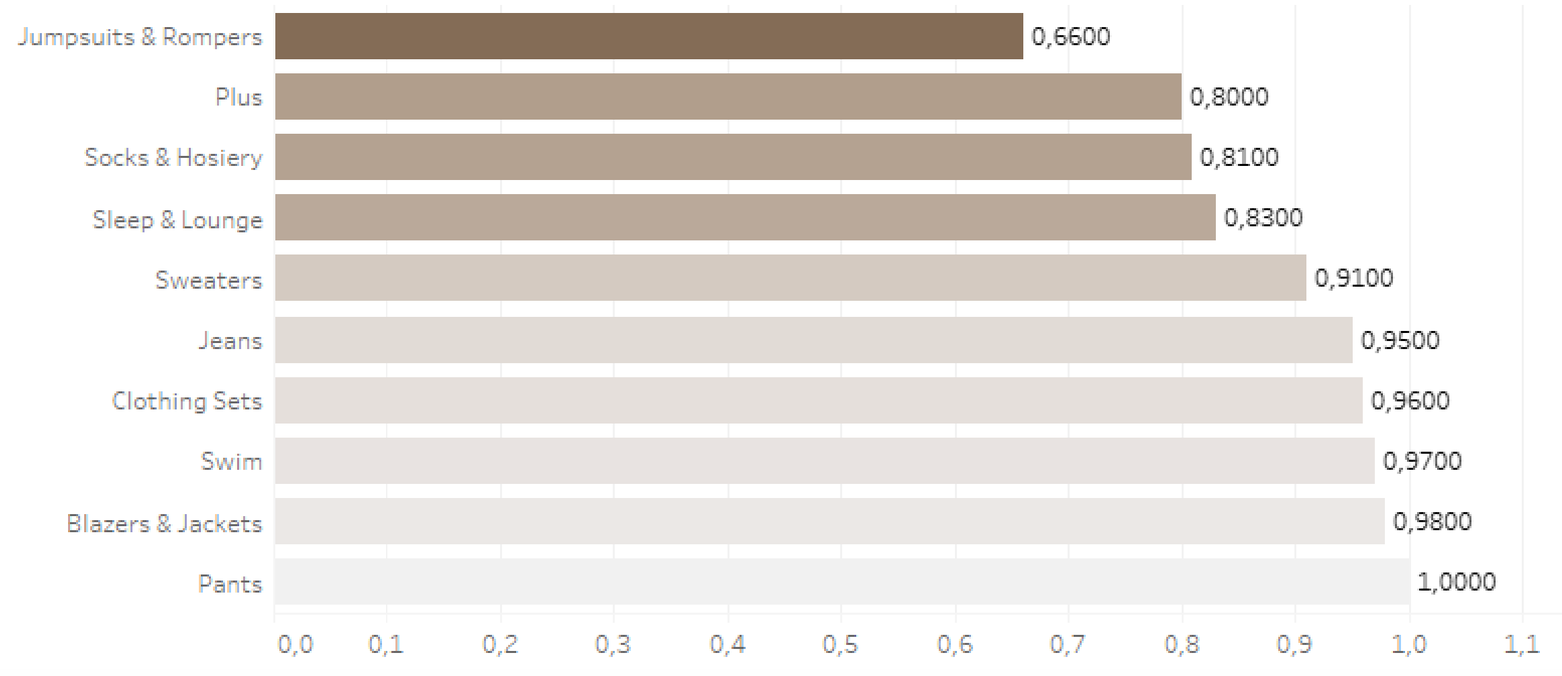
BUSINESS BACKGROUND

The company is in the optimization mode caused by the potential crisis in 2023. The management has decided to cut off resources in some categories with the lowest growth in the past 1 year. On another side, they want to continue the analysis by understanding the retention behaviors of the users and how to increase the retention rate

QUESTIONS

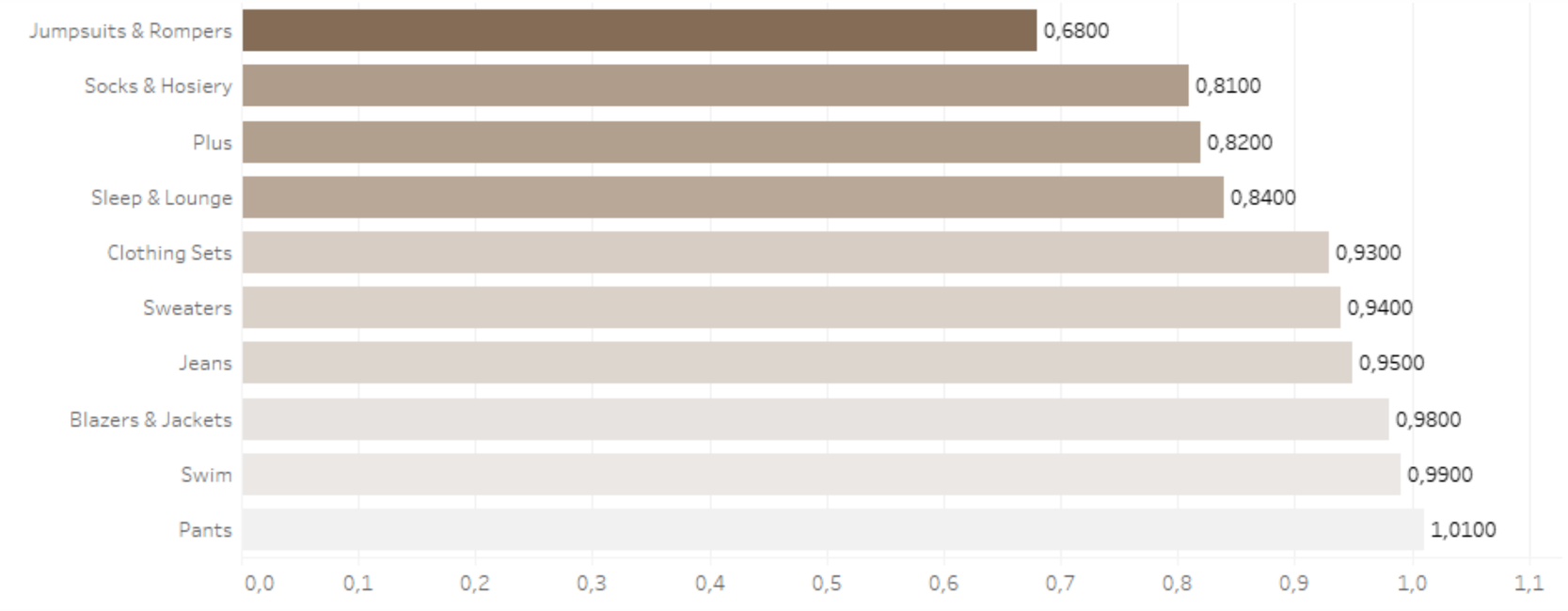
- What is the categories with the lowest revenue and profit growth in the past 1 year?
- What is the categories need to be deprioritized?
- How is the retention rate of the business?
- Is there any new initiatives to boost overall retention rate?

>> Link to Query<<



Based on the chart above the category with the lowest profit growth is Jumpsuits and Rompers that just increase 66% from 2021-2022.

LOWEST PROFIT GROWTH



Based on the chart above the category with the lowest revenue growth is Jumpsuits and Rompers that just increase 68% from 2021-2022.

LOWEST REVENUE GROWTH

DEPRIORITIZED CATEGORY

Profit Growth	High	Low
High	Active Fashion Hoodies & Sweatshirts Shorts Suits & Sport Coats Outwear & Coats	Skirts Accessories Intimates Underwear Socks Leggings Pants & Capris
Low	Sleep & Lounge Sweaters Jeans Swim Pants Dresses	Jumpsuits & Rompers Plus Socks & Hosiery Clothing Sets Blazers & Jackets

Market Share

The use of BCG Matrix

Based on the BCG matrix The category need to be prioritized are

- Jumpsuits & Rompers
- Plus
- Socks & Hosiery
- Clothing Sets
- Blazers & Jackets

Since they are in th edog are which means they are categories with low market share and low growth market.

RETENTION RATE

>>Link To Query<<

SUM of cohort_percentage		datediff											
first_order_mont	first_users	0	1	2	3	4	5	6	7	8	9	10	11
2022-01-01	733	100	1.09	0.95	0.95	0.55	0.82	0.68	0.55	0.82	0.55	1.09	1.64
2022-02-01	706	100	0.99	1.56	1.27	0.57	1.27	0.99	1.13	0.85	1.42	1.13	
2022-03-01	755	100	0.53	0.53	1.32	0.79	1.19	1.06	1.06	1.19	1.06		
2022-04-01	750	100	1.33	1.33	0.80	1.47	1.60	0.80	1.87	1.47			
2022-05-01	895	100	1.68	1.01	1.23	1.45	1.45	0.78	1.12				
2022-06-01	868	100	1.04	1.27	1.27	0.92	1.27	1.50					
2022-07-01	994	100	1.31	0.91	1.11	1.91	1.61						
2022-08-01	1019	100	1.18	1.08	1.08	1.18							
2022-09-01	1040	100	1.15	1.54	1.63								
2022-10-01	1083	100	2.31	1.75									
2022-11-01	1268	100	1.89										
2022-12-01	1334	100											

- **The number of new customer in our site is increasing**
- **After the first month, user typically retain some stickiness without a sharp change in trend**
- **We can see that October 2022 cohort had the highest percentage of customer making purchase in the first month compared to the earlier month.**

RECOMMENDATION

- **Improve Customer satisfaction to evaluate the effectiveness of our service. For instance offering voucher discount, faster shipping, simplifying the check-out process, improving website's user interface**

- **In order to increase the customer retention rate we can provide the loyalty program or referral program where the customer can get any reward or discount by being loyal customer or by referring their friends or family.**

- **Conducting further analysis regarding the fashion trend interest by customer such as the colours, style, material.**

APPENDIX

Query

```
SELECT
  a.category,
  SUM(b.sale_price) most_sale_price
FROM
  `sql-project-376612.thelook_ecommerce.products` a
INNER JOIN
  `sql-project-376612.thelook_ecommerce.order_items` b
ON
  a.id = b.product_id
WHERE
  b.status = "Shipped"
  AND b.shipped_at BETWEEN "2022-12-01"
  AND "2022-12-31"
GROUP BY
  1
ORDER BY
  2 DESC
```

>> Link to Google Big Query<<

Query Result

Row	category	most_sale_price
1	Outerwear & Coats	20370.5399...
2	Jeans	17670.5200...
3	Sweaters	12154.4500...
4	Suits & Sport Coats	10360.3699...
5	Fashion Hoodies & Sweatshirts	10257.1200...
6	Swim	8595.10001...
7	Accessories	7746.55001...
8	Sleep & Lounge	7546.24999...

Query

```sql
WITH
  dec AS(
  SELECT
    a.category,
    COUNT(b.order_id) order_dec
  FROM
    `sql-project-376612.thelook_ecommerce.products` a
  JOIN
    `sql-project-376612.thelook_ecommerce.order_items` b
  ON
    a.id = b.product_id
  WHERE
    status = "Shipped"
    AND shipped_at BETWEEN "2022-12-01"
    AND "2022-12-31"
  GROUP BY
    1),
  nov AS(
  SELECT
    c.category,
    COUNT(d.order_id) order_nov
  FROM
    `sql-project-376612.thelook_ecommerce.products` c
  JOIN
    `sql-project-376612.thelook_ecommerce.order_items` d
  ON
    c.id = d.product_id
  WHERE
    status = "Shipped"
    AND shipped_at BETWEEN "2022-11-01"
    AND "2022-11-30"
  GROUP BY
    1)
SELECT
  order_dec,
  order_nov,
  ((order_dec-order_nov)/order_nov)*100 AS growth,
  e.category
FROM
  dec e
JOIN
  nov f
ON
  e.category = f.category
ORDER BY
  3 ASC
```

>> Link to Google Big Query<<

Query Result

Row	order_dec	order_nov	growth	category
1	2	4	-50.0	Clothing Sets
2	62	70	-11.428571...	Maternity
3	16	16	0.0	Skirts
4	13	13	0.0	Jumpsuits & Rompers
5	149	148	0.67567567...	Sleep & Lounge
6	79	73	8.21917808...	Socks
7	148	136	8.82352941...	Shorts

Query

[>> Link to Google Big Query<<]

```sql
SELECT
  traffic_source AS source_traffic,
  COUNT(DISTINCT session_id) AS most_traffic_source
FROM
  `sql-project-376612.thelook_ecommerce.events`
WHERE
  created_at BETWEEN "2022-12-01"
  AND "2022-12-31"
GROUP BY
  1
ORDER BY
  2 DESC
```

Query Result

Row	source_traffic	most_traffic_sou
1	Email	8345
2	Adwords	5591
3	Facebook	1857
4	YouTube	1856
5	Organic	945

Query

```sql
WITH
  sources AS(
  SELECT
    user_id,
    traffic_source,
    created_at,
    ROW_NUMBER() OVER(ORDER BY created_at ASC) User_rank
  FROM
    `sql-project-376612.thelook_ecommerce.events`
  WHERE
    DATE(created_at)>= DATE_SUB (date "2023-01-01",INTERVAL 3 MONTH)),
  revenue AS(
  SELECT
    o.user_id,
    SUM(sale_price) total_price
  FROM
    `sql-project-376612.thelook_ecommerce.order_items` o
  WHERE
    status = "Complete"
  GROUP BY
    1 )
SELECT
  a.traffic_source,
  SUM(b.total_price) total_revenue,
  RANK()OVER (ORDER BY SUM(b.total_price)DESC) AS Rank_revenue
FROM
  sources a
JOIN
  revenue b
ON
  a.user_id = b.user_id
GROUP BY
  1
ORDER BY
  2 DESC
```

>> Link to Google Big Query<<

Query Result

Row	traffic_source	total_revenue	Rank_revenue
1	Email	13559852.7...	1
2	Adwords	8907376.28...	2
3	Facebook	3044611.09...	3
4	YouTube	2966555.62...	4
5	Organic	1549957.46...	5

Query

```sql
WITH data_21 AS
(
SELECT
  EXTRACT(YEAR FROM oi.created_at) IN (2021) as year21,
  category as Kategori_Produk,
  SUM(sale_price) as revenue21,
  SUM(sale_price-cost) profit21
  FROM `sql-project-376612.thelook_ecommerce.order_items` as oi
  JOIN `sql-project-376612.thelook_ecommerce.products` as p
  ON oi.product_id = p.id
  WHERE status = 'Complete'
  AND EXTRACT(YEAR FROM oi.created_at) = 2021
  GROUP BY 1,2
  ORDER BY 2
)
, data_22 AS (
  SELECT
  EXTRACT(YEAR FROM oi.created_at) IN (2022) as year22,
  category as Kategori_Produk,
  SUM(sale_price) as revenue22,
  SUM(sale_price-cost) as profit22
  FROM `sql-project-376612.thelook_ecommerce.order_items` as oi
  JOIN `sql-project-376612.thelook_ecommerce.products` as p
  ON oi.product_id = p.id
  WHERE status = 'Complete'
  AND EXTRACT(YEAR FROM oi.created_at) = 2022
  GROUP BY 1,2
  ORDER BY 2
),
  GROWTH AS (
  SELECT
  data_21.Kategori_Produk as Category,
  year21,
  revenue21,
  profit21,
  ROUND(((revenue22-revenue21)/revenue21),2) as growth_rev,
  ROUND(((profit22-profit21)/profit21),2) as growth_profit
FROM GROWTH
ORDER BY 2 ASC
)

SELECT *
  ,CASE
    WHEN revenue_market_share >= AVG(revenue_market_share) OVER() AND growth_rev >= AVG(growth_rev) OVER()
      THEN 'Star'
    WHEN revenue_market_share >= AVG(revenue_market_share) OVER() AND growth_rev < AVG(growth_rev) OVER()
      THEN 'Cash Cow'
    WHEN revenue_market_share < AVG(revenue_market_share) OVER() AND growth_rev >= AVG(growth_rev) OVER()
      THEN 'Question Mark'
    ELSE 'Dog'
    END Quadrant
FROM market_share
ORDER BY growth_profit
```

>> Link to Google Big Query<<

Query Result

Row	Category	growth_profit	growth_rev	revenue_market	profit_market_sh	Quadrant
1	Jumpsuits & Rompers	0.66	0.68	0.004	0.0	Dog
2	Plus	0.8	0.82	0.018	0.02	Dog
3	Socks & Hosiery	0.81	0.81	0.007	0.01	Dog
4	Sleep & Lounge	0.83	0.84	0.056	0.06	Cash Cow
5	Sweaters	0.91	0.94	0.079	0.08	Cash Cow
6	Jeans	0.95	0.95	0.121	0.11	Cash Cow
7	Clothing Sets	0.96	0.93	0.001	0.0	Dog

Query

```sql
with first_order_ref as (
  SELECT
    user_id
    , min(date_trunc(date(created_at), month)) as first_order_month
  FROM bigquery-public-data.thelook_ecommerce.orders
  where date_trunc(date(created_at), month) BETWEEN '2022-01-01' AND '2022-12-01'
    and status = 'Complete'
    and user_id is not null
  group by 1
)
, act_ref as (
  select
    distinct o.user_id
    , date_trunc(date(o.created_at), month) as period
    , first_order_month
  FROM bigquery-public-data.thelook_ecommerce.orders as o
  left join first_order_ref as fst
    on o.user_id = fst.user_id
  where o.user_id is not null and date_trunc(date(o.created_at), month) BETWEEN '2022-01-01' AND '2022-12-01'
    and status = 'Complete'
)
, raw as (
  select
    *
    , date_diff(period, first_order_month, month) datediff
  from act_ref
)
, cohort_size as (
  select
    first_order_month
    , count(distinct user_id) first_users
  from first_order_ref
  group by 1
)
, cohort as (
  select
    first_order_month
    , datediff
    , count(distinct user_id) users
  from raw
  group by 1,2
)
select
  c.*
  , first_users
  , (users/first_users)*100 cohort_percentage
from cohort as c
left join cohort_size as cs
  on c.first_order_month = cs.first_order_month
order by 1,2;
```

>> Link to Google Big Query<<

Query Result

Row	first_order_mont	datediff	users	first_users	cohort_percenta
1	2022-01-01	0	733	733	100.0
2	2022-01-01	1	8	733	1.09140518...
3	2022-01-01	2	7	733	0.95497953...
4	2022-01-01	3	7	733	0.95497953...
5	2022-01-01	4	4	733	0.54570259...
6	2022-01-01	5	6	733	0.81855388...
7	2022-01-01	6	5	733	0.68212824...

FASHION COLLECTION

THANKS

Click Here

Click Here

maismaula01@gmail.com