

The Look E-Commerce Data Analysis Using SQL

Pendahuluan

Dalam proyek ini, saya menggunakan SQL untuk menganalisis data e-commerce dari The Look. The Look adalah toko pakaian e-commerce yang dikembangkan oleh tim Google Looker. Data e-commerce ini dihosting di Google Big Query dan berisi informasi tentang pelanggan, produk, pesanan, logistik, acara web, dan kampanye pemasaran digital. Konten dataset ini disediakan untuk praktisi industri untuk tujuan penemuan produk, pengujian, dan evaluasi.

Pada saat analisis ini dilakukan, toko e-commerce mencatat data transaksi penjualan dari Agustus 2019 hingga September 2023.

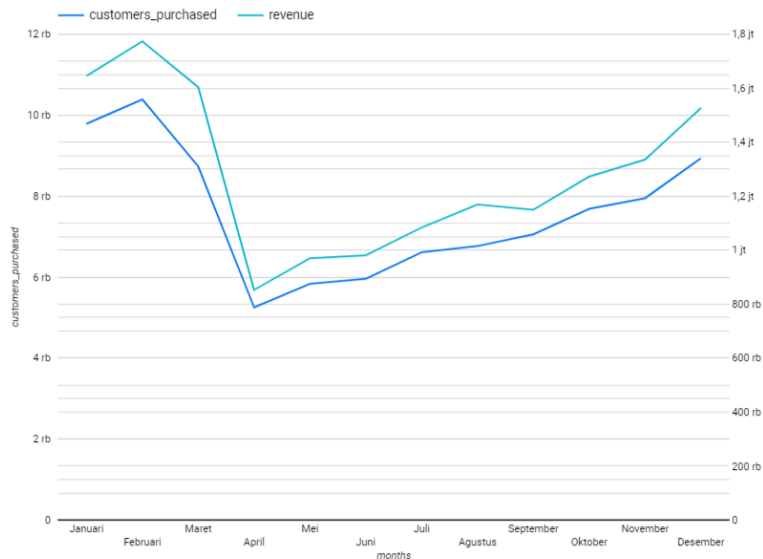
Tujuan

Tujuan dari analisis ini adalah untuk memahami dan menjawab beberapa pertanyaan bisnis terkait performa marketplace Look e-commerce, mendapatkan wawasan, dan memberikan beberapa rekomendasi untuk meningkatkan pendapatan.

Kita akan fokus pada pesanan yang tidak dibatalkan atau dikembalikan karena kita akan berurusan dengan metrik yang terkait dengan pendapatan.

Menjawab Pertanyaan Bisnis

- 1. Bagaimana trend pendapatan pada e-commerce setiap bulannya ? Apakah tinggi atau rendah dibandingkan dengan bulan lalu ?**
 - a. Bagaimana proyeksi pendapatan tiap bulannya ?
 - b. Adakah bulan tertentu yang mendominasi penjualan tiap tahunnya? Mengapa hal tersebut dapat terjadi?



Insight:

- Dari hasil Analisa diperoleh fakta bahwa pendapatan pada perusahaan, mengalami trend naik-turun dari bulan ke bulan.
- Bulan Februari menjadi puncak keuntungan dan jumlah pembelian customer. Hal ini bisa terjadi dikarenakan perayaan valentine bersamaan dengan suksesnya event yang dilakukan.
- Bulan Januari dan Desember juga menjadi penyumbang keuntungan dan jumlah pembelian customer yang masif. Hal ini bisa terjadi dikarenakan adanya perayaan natal, tahun baru, dan musim liburan sekolah.
- Bulan April menjadi bulan yang mencatatkan keuntungan dan jumlah pembelian customer paling rendah dibandingkan dengan bulan lainnya.

Source code :

```
SELECT
EXTRACT(MONTH FROM oi.Created_at) AS months,
ROUND(SUM(oi.sale_price * o.num_of_item), 2) AS revenue,
COUNT(DISTINCT oi.order_id) AS order_count,
COUNT(DISTINCT oi.user_id) AS customers_purchased
FROM bigquery-public-data.thelook_ecommerce.order_items oi
INNER JOIN `bigquery-public-data.thelook_ecommerce.orders` AS o
ON oi.order_id = o.order_id
WHERE oi.status NOT IN ('Cancelled', 'Returned')
GROUP BY months
ORDER BY months ASC;
```

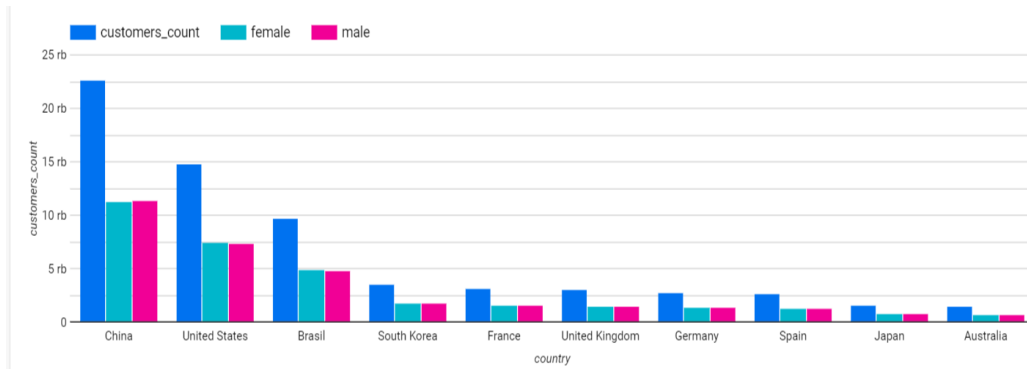
Rekomendasi :

- Dengan demikian perusahaan perlu mempertahankan atau meningkatkan keuntungan dan jumlah pembelian customer pada bulan Januari, Februari, dan Desember dengan mengadakan berbagai event-event yang menarik.
- Perusahaan juga dapat mengadakan berbagai event lain untuk meningkatkan keuntungan dan jumlah pembelian customer pada bulan-bulan lainnya.

2. Siapa customer kita ? Dari negara mana, kelompok jenis kelaminnya, dan usia customer ?

- Darimana saja pelanggan kita berasal ?
- Kelompok jenis kelamin mana yang memberikan ?
- Berapa rentang usia customer kita ?

A. Customer berdasarkan negara



Insight :

- Mayoritas customer kita berasal dari Cina, Amerika Serikat, dan Brasil.
- Proporsi pendapatan yang dihasilkan oleh masing-masing gender hampir sama di setiap negara.

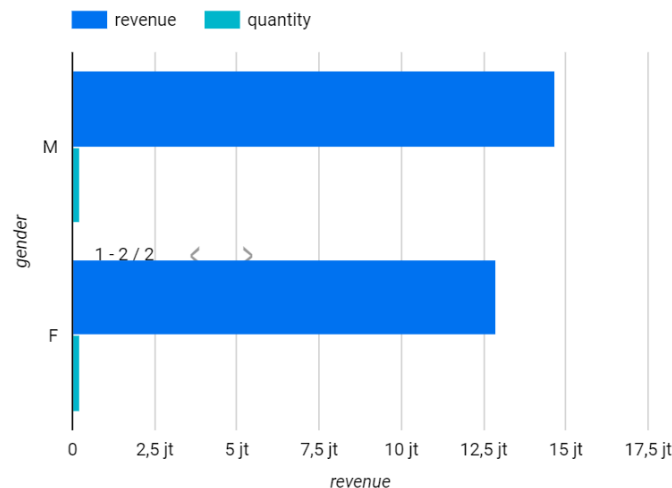
Source Code :

```
WITH customers AS (  
  SELECT  
    DISTINCT oi.user_id,  
    SUM(CASE WHEN u.gender = 'M' THEN 1 ELSE null END) AS male,  
    SUM(CASE WHEN u.gender = 'F' THEN 1 ELSE null END) AS female,  
    u.country AS country  
  FROM bigquery-public-data.thelook_ecommerce.order_items AS oi  
  INNER JOIN bigquery-public-data.thelook_ecommerce.users AS u  
  ON oi.user_id = u.id  
  WHERE oi.status not in ('Cancelled', 'Returned')  
  GROUP BY user_id, country  
)  
  
SELECT  
  country,  
  COUNT(DISTINCT user_id) AS customers_count,  
  COUNT(female) AS female,  
  COUNT(male) AS male  
FROM customers  
GROUP BY country  
ORDER BY customers_count DESC;
```

Rekomendasi :

- a. Perusahaan dapat mengadakan berbagai event khusus pada setiap negara sesuai karakteristik demografinya sebagai cara untuk memperkuat branding dan menarik customer baru. Negara dengan jumlah customer lebih banyak, lebih baik diprioritaskan untuk menjaga loyalitas customer, namun hal ini tidak berarti mengabaikan negara dengan jumlah customer yang rendah.

B. Customer berdasarkan jenis kelamin



Insight :

- a. Customer laki-laki memberikan lebih banyak pendapatan dan jumlah pembelian dibandingkan dengan customer perempuan. Namun, jumlah pembelian antara laki-laki dan perempuan memiliki jumlah yang hampir sama.

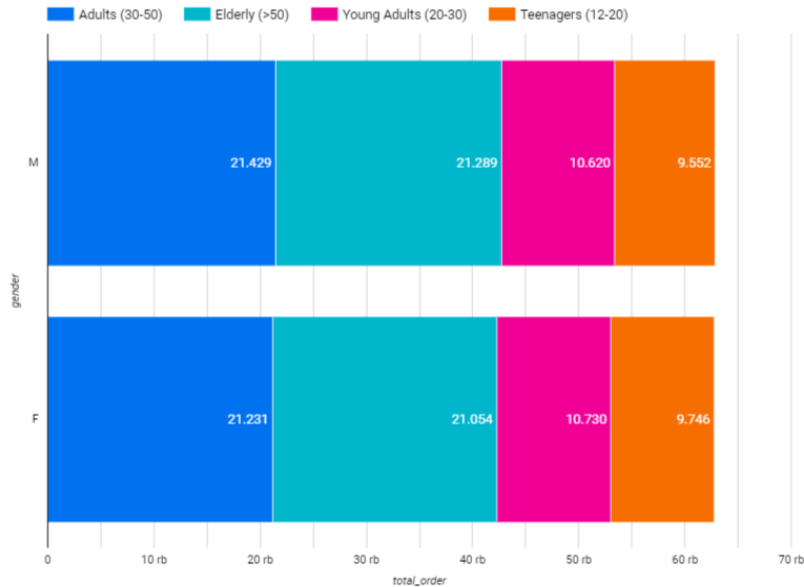
Rekomendasi :

- a. Basis pelanggan yang terbagi rata memberikan keuntungan untuk dapat melakukan berbagai strategi pemasaran yang fokus kepada laki-laki atau perempuan atau bahkan keduanya, tanpa khawatir nilai resiko yang terlalu besar.

Source Code :

```
SELECT
  o.gender,
  SUM(oi.sale_price*o.num_of_item) revenue,
  SUM(o.num_of_item) quantity
FROM `bigquery-public-data.thelook_ecommerce.order_items` oi
LEFT JOIN `bigquery-public-data.thelook_ecommerce.orders` o
ON oi.order_id = o.order_id
WHERE oi.status NOT IN ('Cancelled','Returned')
GROUP BY 1
ORDER BY 2
```

C. Customer berdasarkan grup usia



Insight :

- Mayoritas customer kita adalah lansia (>50) dan orang dewasa (30-50). Sisanya adalah orang dewasa muda (20-30) dan remaja (12-20).
- Jumlah customer laki-laki dan perempuan menunjukkan jumlah yang hampir sama di setiap grup usia.

Source Code:

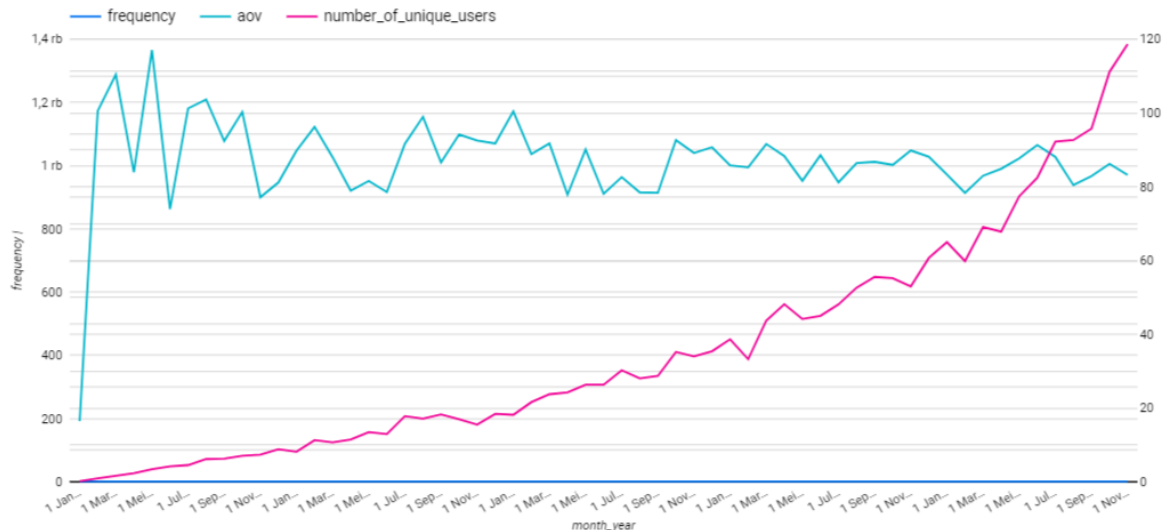
```
SELECT
CASE
  WHEN u.age BETWEEN 12 AND 20 THEN 'Teenagers (12-20)'
  WHEN u.age BETWEEN 20 AND 30 THEN 'Young Adults (20-30)'
  WHEN u.age BETWEEN 30 AND 50 THEN 'Adults (30-50)'
  WHEN u.age > 50 THEN 'Elderly (>50)'
END AS age_segment,
U.gender,
COUNT(O.num_of_item) AS total_order
FROM
bigquery-public-data.thelook_ecommerce.users U
LEFT JOIN
bigquery-public-data.thelook_ecommerce.orders O
ON U.id = O.user_id
GROUP BY age_segment, U.gender
ORDER BY age_segment, U.gender;
```

Rekomendasi :

- Tetap mempertahankan dan meningkatkan strategi pemasaran yang dilakukan untuk segmen orang dewasa untuk menjaga loyalitas customer.
- Mencoba menarik anak-anak muda dengan cara menambahkan produk-produk yang sedang trend di tingkat global atau local dan memberikan berbagai penawaran khusus pada produk-produk tersebut.

3. Bagaimana trend nilai frekuensi, AOV, dan jumlah pengguna pada e-commerce setiap bulannya pada rentang 1 Januari 2019 - 30 November 2023 ?

- a. Bagaimana proyeksi nilai frekuensi, AOV, dan jumlah pengguna e-commerce setiap bulannya pada rentang 1 Januari 2019 – 30 November 2023 ?



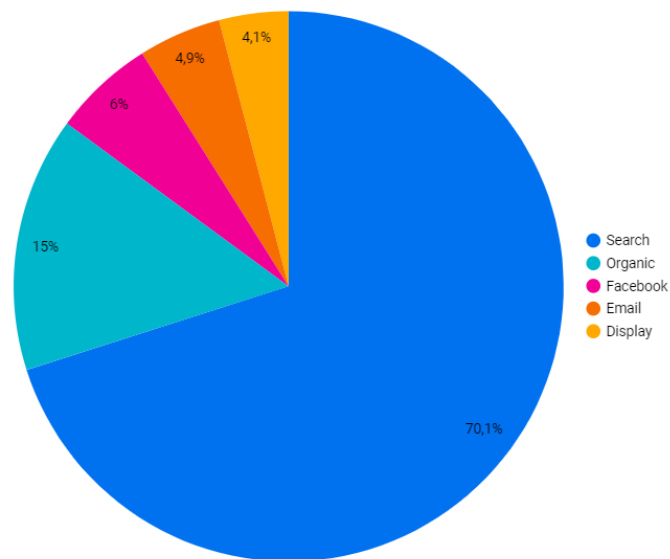
Insight :

- a. Dari Januari 2019 – November 2023, kita dapat melihat jumlah customer yang selalu meningkat dari tahun ke tahun menunjukkan perkembangan yang baik, untuk nilai Average Order Value (AOV) menunjukkan nilai yang naik turun (masih dalam kondisi stabil) dengan puncaknya pada 1 Maret 2019, sedangkan untuk frekuensi menunjukkan kondisi yang stagnan.

Source Code :

```
SELECT DATE_TRUNC(DATE(shipped_at), MONTH) AS month_year,
       ROUND(COUNT(DISTINCT order_id)/COUNT(DISTINCT user_id), 2) AS frequency,
       ROUND(SUM(sale_price)/COUNT(DISTINCT order_id), 2) AS aov,
       COUNT(DISTINCT(user_id)) AS number_of_unique_users
FROM `bigquery-public-data.thelook_ecommerce.order_items`
WHERE shipped_at BETWEEN '2019-01-01' AND '2023-11-30'
      AND status = 'Complete'
GROUP BY 1
ORDER BY 1
```

4. Apa media marketing paling baik untuk mempromosikan e-commerce kita ?



Insight:

- Sebagian besar pelanggan menemukan tentang e-commerce kita melalui pencarian internet (70,1%). Sisanya melalui organik (15%), facebook (6%), email (4,9%), dan display (4,1%).

Source Code :

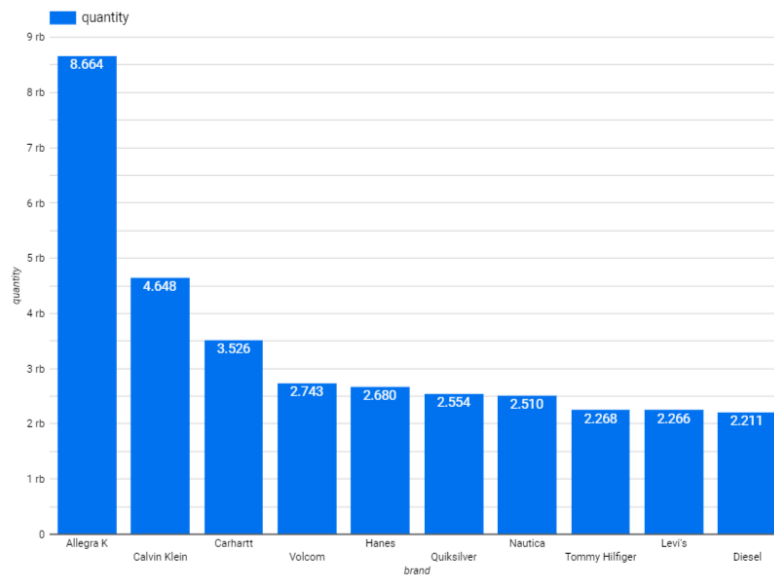
```
SELECT
  u.traffic_source,
  COUNT(DISTINCT oi.user_id) total_customer
FROM `bigquery-public-data.thelook_ecommerce.order_items` oi
LEFT JOIN `bigquery-public-data.thelook_ecommerce.users` u
ON oi.user_id = u.id
WHERE oi.status NOT IN ('Cancelled','Returned')
GROUP BY 1
ORDER BY 2 DESC
```

Rekomendasi :

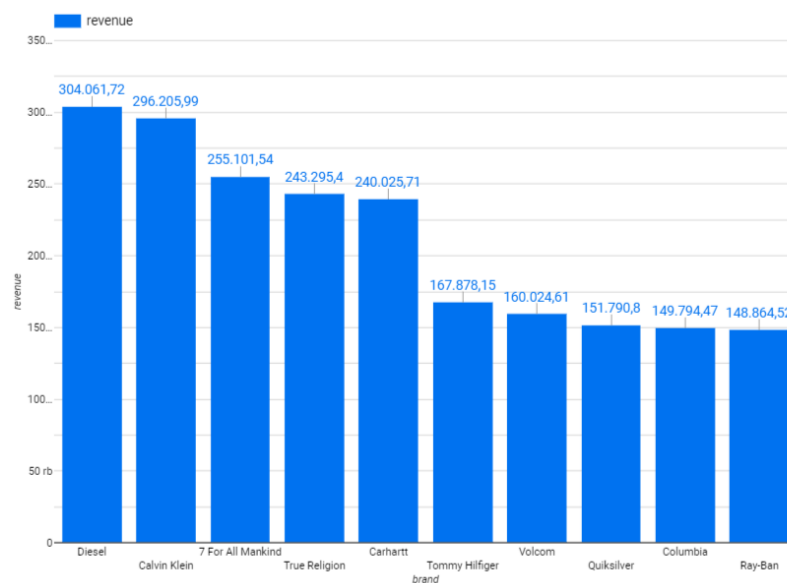
- Dana marketing yang digunakan untuk iklan di platform pencarian internet lebih baik tidak dikurangi dan akan lebih baik diberikan penambahan.
- Memperluas jangkauan iklan di platform lain seperti Tiktok, Instagram, dan Twitter dapat dipertimbangkan untuk menjangkau target customer yang lebih luas dan beragam. Hal ini tidak lain dikarenakan pengguna Tiktok, Instagram, dan Twitter yang terus menerus mengalami kenaikan.

5. Apa kategori produk dan brands yang paling banyak dan paling sedikit memberikan pendapatan dan kuantitas penjualan ?

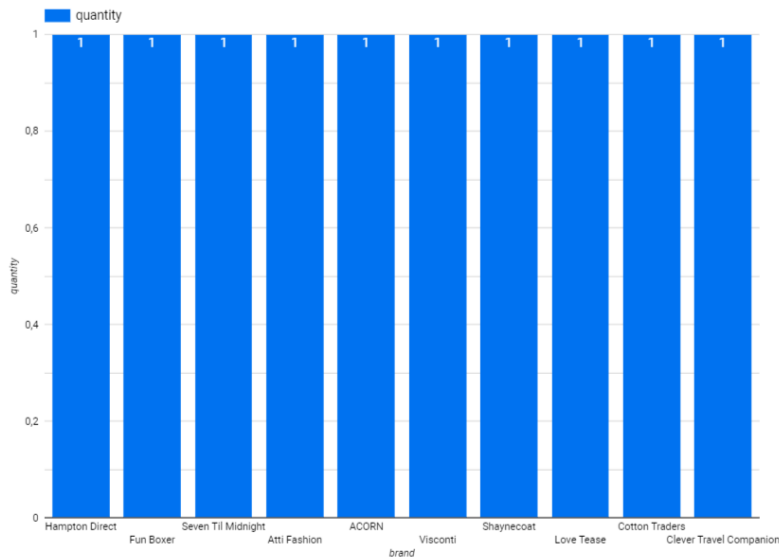
A. Brand Sales



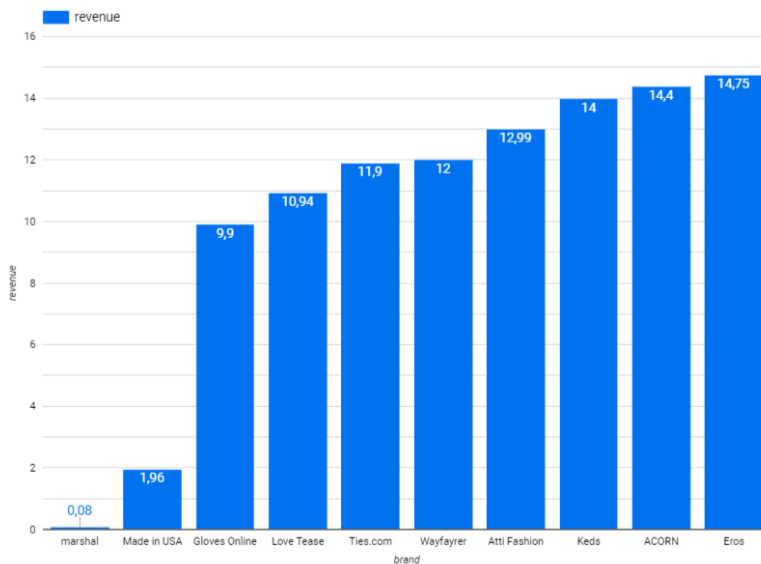
Brand Produk Paling Banyak Jumlah Penjualannya



Brand Produk Paling Banyak Menguntungkan



Brand Produk Paling Sedikit Jumlah Penjualannya



Brand Produk Paling Sedikit Keuntungannya

Insight :

- Brand Diesel memberikan pendapatan paling tinggi di antara brand lain.
- Brand Allegra K memiliki kuantitas penjualan terbanyak, hampir 2 kali lipat dari jumlah penjualan Calvin Klein.
- Brand Marshal memberikan pendapatan paling rendah di antara brand lain.
- Perusahaan kita memiliki banyak brand yang memiliki kuantitas penjualan 1,2, dan 3 produk.

Source Code :

```
SELECT
  p.brand AS brand,
  ROUND(SUM(sale_price * num_of_item), 2) AS revenue,
  SUM(num_of_item) AS quantity
FROM bigquery-public-data.thelook_ecommerce.order_items oi
```

```

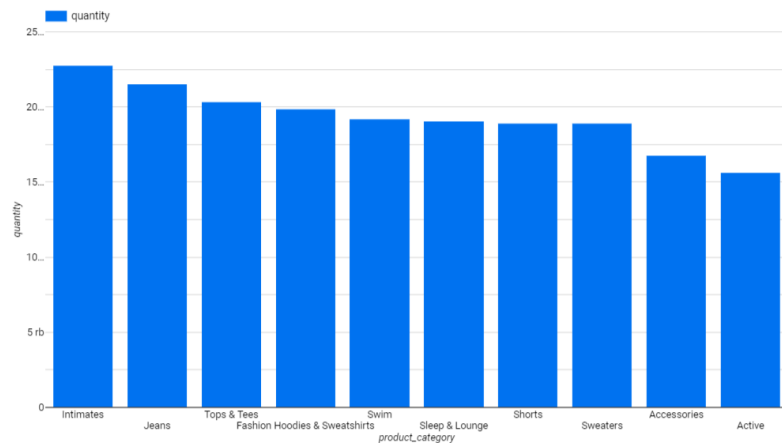
INNER JOIN bigquery-public-data.thelook_ecommerce.orders o
ON oi.order_id = o.order_id
INNER JOIN bigquery-public-data.thelook_ecommerce.products p
ON oi.product_id = p.id
WHERE oi.status NOT IN ('Cancelled', 'Returned')
GROUP BY brand
ORDER BY revenue DESC;

```

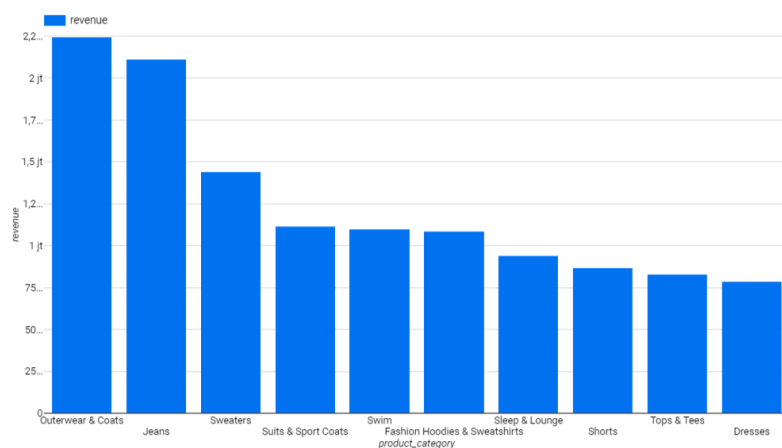
Rekomendasi :

- Perusahaan dapat melakukan kerja sama kepada brand-brand yang menunjukkan penjualan dan pendapatan yang luar biasa. Dengan melakukan kerja sama, perusahaan dapat mengadakan event-event khusus dengan brand tersebut untuk membuat produk exclusive atau harga khusus.

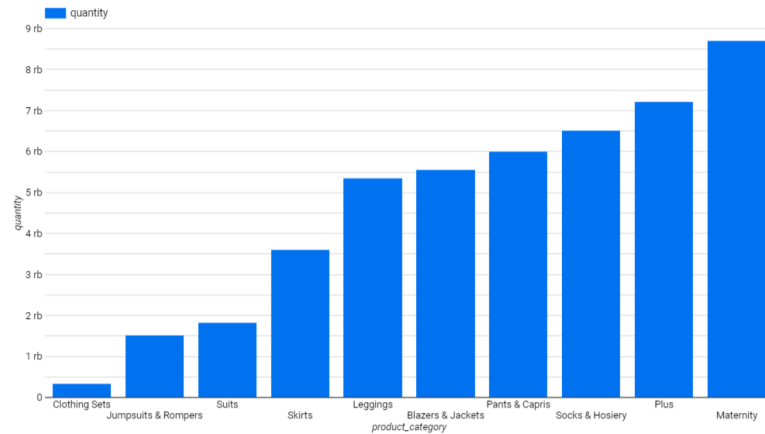
B. Product Category Sales



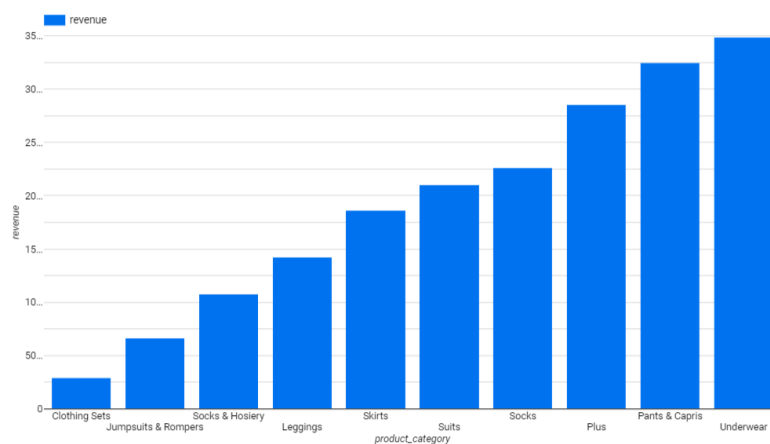
Kategori Produk Paling Banyak Jumlah Penjualannya



Kategori Produk Paling Menguntungkan



Kategori Produk Paling Sedikit Jumlah Penjualannya



Kategori Produk Paling Sedikit Keuntungannya

Insight :

- Jenis produk outerwear dan coats memberikan pendapatan paling tinggi di antara produk lainnya.
- Jenis produk intimates memiliki kuantitas penjualan terbanyak di antara produk lainnya.
- Jenis produk clothing sets mencatatkan pendapatan dan kuantitas penjualan paling rendah di antara produk lainnya.

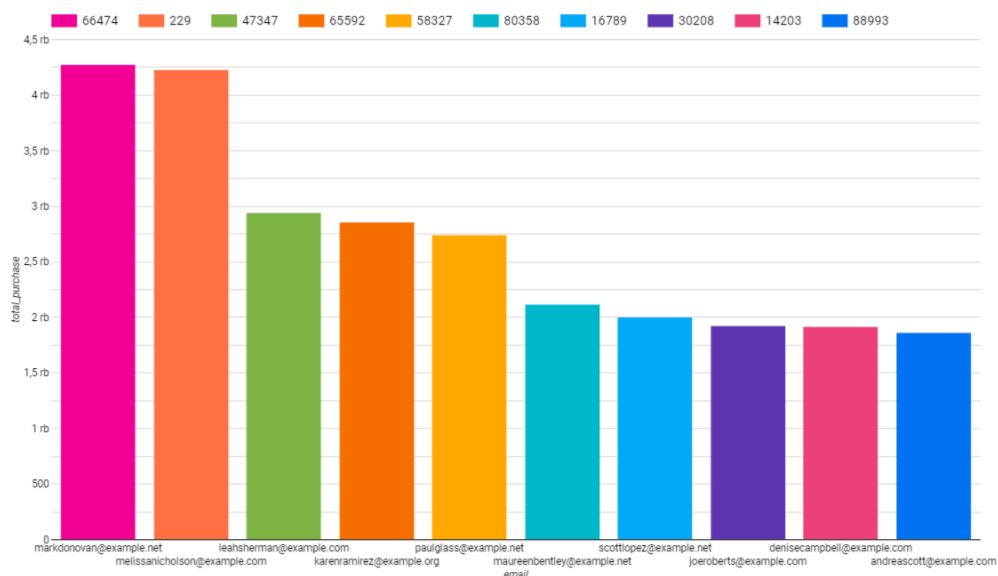
Source Code :

```
SELECT
category AS product_category,
ROUND(SUM(sale_price * num_of_item), 2) AS revenue,
SUM(num_of_item) AS quantity
FROM bigquery-public-data.thelook_ecommerce.order_items oi
INNER JOIN bigquery-public-data.thelook_ecommerce.orders o
ON oi.order_id = o.order_id
INNER JOIN bigquery-public-data.thelook_ecommerce.products p
ON oi.product_id = p.id
WHERE oi.status NOT IN ('cancelled', 'Returned')
GROUP BY category
ORDER BY revenue DESC;
```

Rekomendasi :

- Perusahaan dapat melakukan berbagai event khusus yang melibatkan jenis produk dengan kinerja tinggi untuk menghasilkan lebih banyak keuntungan.
- Dalam hal ini, Perusahaan juga dapat melakukan kerja sama khusus dengan brand ternama pada jenis produk brand yang menunjukkan penjualan dan pendapatan yang luar biasa.
- Perusahaan juga dapat melakukan event khusus yang berisi jenis produk dengan keuntungan dan jumlah penjualan yang sedikit untuk mengenalkan produk tersebut ke customer.

6. Mengidentifikasi 10 customer dengan total pembelian terbanyak di negara Australia untuk diskon terbesar event perayaan hari Kemerdekaan Australia ?



Insight :

- Pengguna dengan id 66474 menjadi penerima diskon terbesar dengan jumlah pembelian sebanyak 4276,3.
- Pengguna dengan id 88993 menjadi penerima diskon terbesar terakhir dengan jumlah pembelian sebanyak 1.865,2.
- Rata-rata jumlah pembelian penerima diskon terbesar adalah 2.689,573.

Source Code :

```
SELECT
  u.id AS customer_id,
  u.email AS email,
  ROUND(SUM(oi.sale_price * o.num_of_item), 2) AS total_purchase
FROM bigquery-public-data.thelook_ecommerce.order_items AS oi
INNER JOIN bigquery-public-data.thelook_ecommerce.orders AS o
ON oi.order_id = o.order_id
INNER JOIN bigquery-public-data.thelook_ecommerce.users AS u
ON o.user_id = u.id
WHERE country = 'Australia'
GROUP BY customer_id, email
```

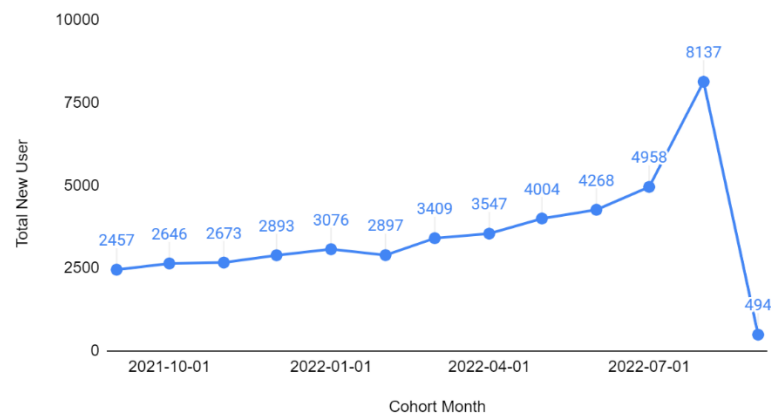
ORDER BY total_purchase DESC
LIMIT 10;

7. Membuat cohort retensi bulanan (Mulai dari tanggal 01 September 2021 sampai 1 September 2022)

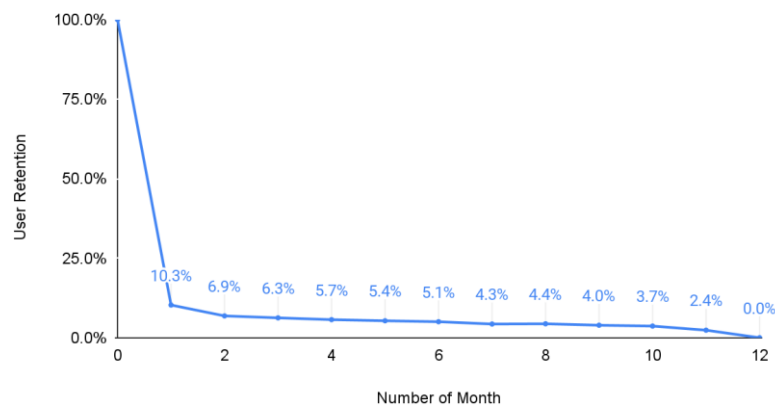
AVERAGE of pe num_month														
cohort_month	0	1	2	3	4	5	6	7	8	9	10	11	12 Grand Total	
2021-09-01	100.00%	7.37%	4.80%	5.58%	4.76%	4.03%	5.66%	4.52%	5.21%	5.25%	5.82%	4.80%	0.04%	12.13%
2021-10-01	100.00%	6.88%	5.56%	5.14%	4.80%	5.48%	4.57%	4.84%	5.48%	4.88%	5.22%	0.23%		12.76%
2021-11-01	100.00%	7.78%	5.95%	5.01%	5.65%	5.61%	6.17%	5.20%	5.57%	5.69%	0.07%			13.88%
2021-12-01	100.00%	8.84%	5.05%	6.01%	6.05%	7.09%	5.81%	5.25%	5.74%	0.07%				14.97%
2022-01-01	100.00%	8.03%	6.21%	6.47%	6.63%	5.49%	6.79%	6.01%	0.10%					16.19%
2022-02-01	100.00%	8.86%	7.01%	6.73%	6.45%	6.97%	6.32%	0.24%						17.80%
2022-03-01	100.00%	9.65%	7.36%	7.83%	8.15%	8.10%	0.18%							20.18%
2022-04-01	100.00%	11.53%	9.05%	9.36%	8.77%	0.17%								23.15%
2022-05-01	100.00%	10.91%	11.84%	10.21%	0.22%									26.64%
2022-06-01	100.00%	16.05%	12.46%	0.52%										32.26%
2022-07-01	100.00%	19.69%	0.61%											40.10%
2022-08-01	100.00%	8.61%												54.31%
2022-09-01	100.00%													100.00%
Grand Total	100.00%	10.32%	6.90%	6.29%	5.72%	5.37%	5.07%	4.34%	4.42%	3.97%	3.70%	2.41%	0.04%	19.48%

SUM of total_us num_month														
cohort_month	0	1	2	3	4	5	6	7	8	9	10	11	12 Grand Total	
2021-09-01	2457	181	118	137	117	99	139	111	128	129	143	113	1	3873
2021-10-01	2646	182	147	136	127	145	121	128	145	129	138	6		4050
2021-11-01	2673	208	159	134	151	150	165	139	149	152	2			4082
2021-12-01	2893	250	146	174	175	205	168	152	166	2				4331
2022-01-01	3076	247	191	199	204	169	209	185	3					4483
2022-02-01	2897	251	203	195	187	202	183	7						4125
2022-03-01	3409	329	251	267	278	276	8							4816
2022-04-01	3547	409	321	332	311	8								4926
2022-05-01	4004	437	474	409	9									5333
2022-06-01	4268	685	532	22										5507
2022-07-01	4958	976	30											5964
2022-08-01	8137	791												8838
2022-09-01	494													494
Grand Total	45459	4856	2572	2005	1559	1252	991	722	591	412	283	119	1	60822

Total New User Each Month



User Retention Ratio



Insight:

- a. Total pengguna baru menunjukkan tren naik dalam setahun terakhir. Namun, terdapat penurunan besar pada bulan September 2022 yang perlu dianalisis lebih lanjut.
- b. Sementara itu, rasio retensi mengalami penurunan besar pada bulan pertama setelah penggunaan pertama mereka. Sepanjang tahun, pengguna yang kembali masih rendah dengan rata-rata kurang dari 10%.

Source Code:

```
WITH monthly_recap AS
(
SELECT
  DATE_TRUNC(DATE(order_items.created_at),month) AS date,
  inventory_items.product_category AS product_categories,
  COUNT(order_items.id) AS order_qty,
  ROUND(SUM(order_items.sale_price),2) AS revenue
FROM `bigquery-public-data.thelook_ecommerce.order_items` order_items
INNER JOIN
  `bigquery-public-data.thelook_ecommerce.inventory_items` inventory_items
ON order_items.inventory_item_id = inventory_items.id
WHERE DATE(order_items.created_at) BETWEEN DATE(2019,01,01) AND DATE(2022,04,30)
AND order_items.status = 'Complete'
GROUP BY
  1,2
ORDER BY
  1 DESC,2
),

growth AS
(
SELECT
  *,
  LAG(order_qty)OVER(PARTITION BY product_categories ORDER BY date) previous_order_qty,
  LAG(revenue)OVER(PARTITION BY product_categories ORDER BY date) previous_revenue,
FROM monthly_recap
)

SELECT
  date,
  product_categories,
  order_qty AS TPO,
  revenue AS TPV,
  CONCAT(COALESCE(ROUND((order_qty-previous_order_qty)/previous_order_qty*100),0), '%')
order_growth_percentage,
  CONCAT(COALESCE(ROUND((revenue-previous_revenue)/previous_revenue*100),0), '%')
revenue_growth_percentage
FROM growth
ORDER BY
  product_categories, date DESC;
-- Determine cohort item: 1. unique identifier (user_id) and 2. Initial Start Date
WITH cohort_item AS
(
SELECT
  user_id,
  MIN(DATE(DATE_TRUNC(created_at,MONTH))) AS cohort_month
```

```

FROM `bigquery-public-data.thelook_ecommerce.order_items`
GROUP BY
    user_id
),

-- Determine size of each cohort by counting user_id that purchase for the first time in cohort size
cohort_size AS
(
SELECT
    cohort_month,
    COUNT(user_id) AS num_users,
FROM cohort_item
GROUP BY
    cohort_month
),

-- Measure Activity After Cohort Month/initial start date
user_activities AS
(
SELECT
    oi.user_id AS user_id,
    DATE_DIFF(DATE(DATE_TRUNC(oi.created_at,MONTH)), cohort_item.cohort_month, MONTH)
AS num_month
FROM `bigquery-public-data.thelook_ecommerce.order_items` AS oi
LEFT JOIN
    cohort_item
ON cohort_item.user_id = oi.user_id
WHERE EXTRACT(YEAR FROM cohort_item.cohort_month) IN (2019,2020,2021,2022)
GROUP BY 2, 1
),

retention_table AS
(
SELECT
    c.cohort_month,
    a.num_month,
    COUNT(1) AS num_users
FROM user_activities AS a
LEFT JOIN
    cohort_item AS c
ON a.user_id = c.user_id
GROUP BY 1, 2
)

SELECT
    r.cohort_month,
    s.num_users AS cohort_size,
    r.num_month,
    r.num_users AS total_users,
    r.num_users/s.num_users AS percentage
FROM retention_table AS r
LEFT JOIN
    cohort_size AS s
ON r.cohort_month = s.cohort_month
WHERE r.cohort_month IS NOT NULL
ORDER BY 1, 3;

```

