

Final Project

By Eni Sawitri Setianingsih

Konteks:

Kumpulan Data Perilaku Konsumen dan Kebiasaan Belanja memberikan wawasan komprehensif tentang preferensi, kecenderungan, dan pola konsumen selama pengalaman berbelanja mereka. Kumpulan data ini mencakup beragam variabel, termasuk informasi demografis, riwayat pembelian, preferensi produk, frekuensi belanja, dan perilaku belanja online/offline. Dengan kumpulan data yang kaya ini, analis dan peneliti dapat menyelidiki seluk-beluk proses pengambilan keputusan konsumen, membantu bisnis dalam menyusun strategi pemasaran yang ditargetkan, mengoptimalkan penawaran produk, dan meningkatkan kepuasan pelanggan secara keseluruhan.

Isi:

Kumpulan Data Perilaku Konsumen dan Kebiasaan Belanja memberikan gambaran rinci tentang preferensi konsumen dan perilaku pembelian. Ini mencakup informasi demografis, riwayat pembelian, preferensi produk, dan saluran belanja pilihan (online atau offline). Kumpulan data ini penting bagi bisnis yang ingin menyesuaikan strategi mereka untuk memenuhi kebutuhan pelanggan dan meningkatkan pengalaman berbelanja, yang pada akhirnya mendorong penjualan dan loyalitas.

Glosarium Kumpulan Data (berdasarkan Kolom)

Customer ID : Pengidentifikasi unik yang diberikan kepada masing-masing pelanggan, memfasilitasi pelacakan dan analisis perilaku belanja mereka dari waktu ke waktu.

Order Date: Menunjukkan tanggal pembelian oleh pelanggan

Order Time: Menunjukkan waktu pembelian oleh pelanggan

Age: Usia pelanggan, memberikan informasi demografis untuk segmentasi dan strategi pemasaran yang ditargetkan.

Gender: Identifikasi gender pelanggan, variabel demografis utama yang mempengaruhi preferensi produk dan pola pembelian.

Item Purchased: Produk atau barang tertentu yang dipilih oleh pelanggan selama transaksi.

Category: Klasifikasi luas atau kelompok barang yang dibeli (misalnya pakaian, elektronik, bahan makanan).

Purchase Amount (USD): Nilai moneter transaksi, dinotasikan dalam Dolar Amerika Serikat (USD), menunjukkan harga barang yang dibeli.

Discount: Menunjukkan persentase diskon untuk barang tertentu

Location: Lokasi geografis tempat pembelian dilakukan, memberikan wawasan mengenai preferensi regional dan tren pasar.

Size: Spesifikasi ukuran (jika ada) barang yang dibeli, relevan untuk pakaian jadi, alas kaki, dan barang konsumsi tertentu.

Color: Varian atau pilihan warna yang terkait dengan barang yang dibeli, mempengaruhi preferensi pelanggan dan ketersediaan produk.

Season: Relevansi musiman dari barang yang dibeli (misalnya, musim semi, musim panas, musim gugur, musim dingin), yang mempengaruhi manajemen inventaris dan strategi pemasaran.

Rating Review: Penilaian numerik atau kualitatif yang diberikan oleh pelanggan mengenai kepuasan mereka terhadap barang yang dibeli.

Subscription: Menunjukkan apakah pelanggan telah memilih layanan berlangganan, yang menawarkan wawasan tentang tingkat loyalitas dan potensi pendapatan berulang mereka.

Shipping Type: Menentukan metode yang digunakan untuk mengirimkan barang yang dibeli (misalnya, pengiriman standar, pengiriman ekspres), yang mempengaruhi waktu dan biaya pengiriman.

Discount Applied: Menunjukkan discount yang diterapkan oleh pelanggan pada aplikasi, berisikan Yes dan No

Kode Promo yang Digunakan: Mencatat apakah kode promosi atau kupon digunakan selama transaksi, membantu dalam evaluasi keberhasilan kampanye pemasaran.

Payment Method: Menentukan cara pembayaran yang digunakan oleh pelanggan (misalnya, e-wallet, uang tunai), menawarkan wawasan tentang opsi pembayaran pilihan.

Descriptive Analysis

Objektif: Menyediakan gambaran umum tentang dataset, termasuk sumber data, jumlah observasi, variabel Sebelum memulai analisis lebih lanjut, perlu memahami karakteristik dari sebuah dataset agar memahami gambaran umum dari data transaksi di tahun 2023 ini.

1. Berapa jumlah transaksi yang terjadi di tahun 2023?
2. Berapa jumlah customer unqiue di tahun 2023?
3. Berapa total revenue yang didapatkan di tahun 2023?
4. Ada berapa kategori product yang dijual di tahun 2023?
5. Ada Berapa product yang dijual di tahun 2023?
6. Berapa rata-rata umur customer yang bertransaksi di tahun 2023?

Query dan Output

```
SELECT COUNT(*) AS jumlah_transaksi,  
       COUNT(DISTINCT customer_id) AS jumlah_customer,  
       ROUND(SUM(CASE WHEN discount_applied = 'Yes'  
                     THEN (100-discount)/100*purchase_amount_usd  
                     ELSE purchase_amount_usd END),2) AS total_revenue,  
       COUNT(DISTINCT category) AS jumlah_kategori_produk,  
       COUNT(DISTINCT item_purchased) AS jumlah_produk_terjual,  
       ROUND(AVG(age),0) AS rata_rata_umur  
FROM customer_transaction
```

	jumlah_transaksi bigint	jumlah_customer bigint	total_revenue numeric	jumlah_kategori_produk bigint	jumlah_produk_terjual bigint	rata_rata_umur numeric
1	3900	551	129119.92	4	25	24

Selama tahun 2023, jumlah transaksi yang terjadi sebanyak 3900 transaksi. Dari sebanyak 3900 transaksi jumlah customer yang melakukan transaksi ada sebanyak 551 customer. Dimana dalam hal ini terbukti bahwa satu orang customer bisa melakukan transaksi lebih dari 1x. Bisa dihitung rata-rata satu customer melakukan transaksi 7x. Untuk pendapatan yang didapatkan dari GenggamData Store, dimana dalam hal ini sudah dikurangi diskon yaitu sejumlah \$129119,92. Produk yang terjual selama tahun 2023 ada sebanyak 4 kategori diantaranya Accessories, Footwear, Clothing, dan Outwear. Sedangkan untuk jumlah item produk yang terjual ada sebanyak 25 jenis produk. Dari 551 customer di GenggamData Store ini rata-rata umur customer selama tahun 2023 ini adalah Gen Z yaitu umur 24 tahun.

Performance Sales & Growth Analysis

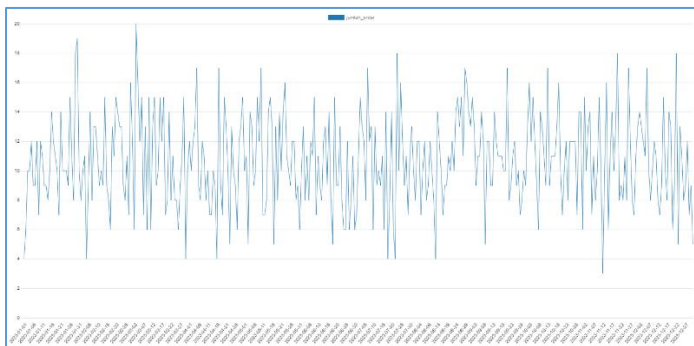
Menganalisis kinerja penjualan dan pertumbuhan bisnis dari waktu ke waktu. Melibatkan pemahaman tentang tren penjualan, faktor yang memengaruhi pertumbuhan, dan identifikasi peluang atau tantangan yang mungkin timbul.

1. Bagaimana overall performance GenggamData Store ditahun 2023 untuk jumlah order dan total sales ?

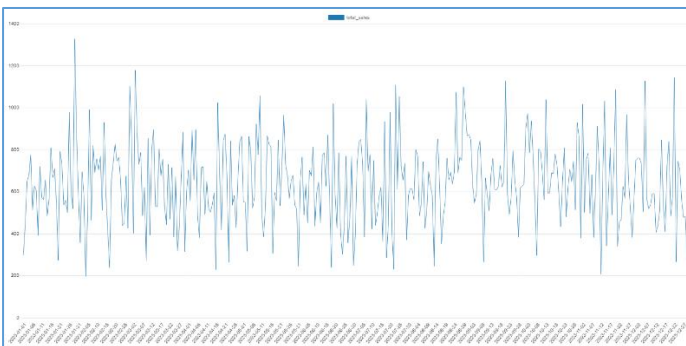
```
SELECT order_date,
       COUNT(*) AS jumlah_order,
       SUM(purchase_amount_usd) AS total_sales
FROM customer_transaction
GROUP BY order_date
ORDER BY order_date
```

Output

	order_date date	jumlah_order bigint	total_sales numeric
1	2023-01-01	4	296
2	2023-01-02	6	444
3	2023-01-03	10	649
4	2023-01-04	10	679
5	2023-01-05	12	777
6	2023-01-06	9	512
7	2023-01-07	9	627
8	2023-01-08	12	607
9	2023-01-09	7	391
10	2023-01-10	12	720



Grafik Jumlah Order (Daily)



Grafik Total Sales (Daily)

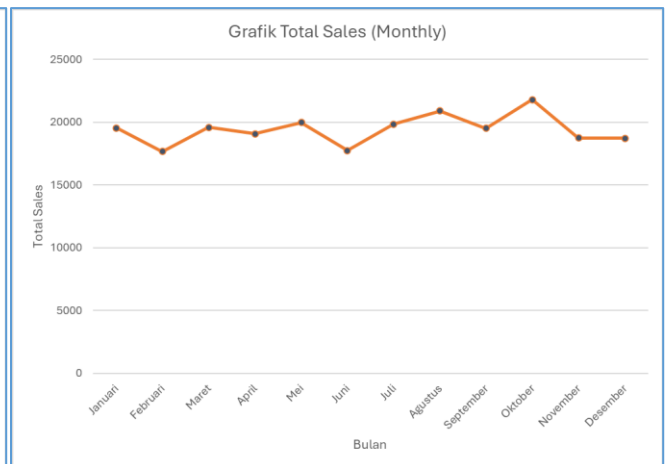
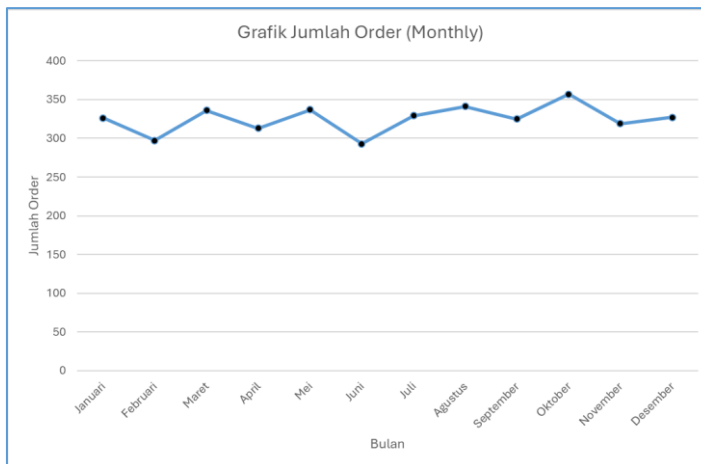
Dari kedua grafik baik jumlah order dan total sales memiliki pola yang hampir samayaitu konstan dan tidak ada kenaikan atau penurunan yang signifikan. Terlihat dalam grafik bahwa terjadi puncak jumlah order di tahun 2023 pada tanggal 3 Maret 2023 yaitu sebanyak 20 pesanan, dimana dalam hal ini juga berbanding lurus dengan total sales yang didapatkan. Selanjutnya akan dilihat juga trend untuk setiap bulan dari jumlah order dan total sales.

Query dan Output

```
SELECT
  CASE
    WHEN EXTRACT(MONTH FROM order_date) = 1 THEN 'januari'
    WHEN EXTRACT(MONTH FROM order_date) = 2 THEN 'februari'
    WHEN EXTRACT(MONTH FROM order_date) = 3 THEN 'maret'
    WHEN EXTRACT(MONTH FROM order_date) = 4 THEN 'april'
    WHEN EXTRACT(MONTH FROM order_date) = 5 THEN 'mei'
    WHEN EXTRACT(MONTH FROM order_date) = 6 THEN 'juni'
    WHEN EXTRACT(MONTH FROM order_date) = 7 THEN 'juli'
    WHEN EXTRACT(MONTH FROM order_date) = 8 THEN 'agustus'
    WHEN EXTRACT(MONTH FROM order_date) = 9 THEN 'september'
    WHEN EXTRACT(MONTH FROM order_date) = 10 THEN 'oktober'
    WHEN EXTRACT(MONTH FROM order_date) = 11 THEN 'november'
    ELSE 'desember'
  END AS bulan,
  COUNT(*) AS jumlah_order,
  SUM(purchase_amount_usd) AS total_sales
FROM
  customer_transaction
GROUP BY
  EXTRACT(MONTH FROM order_date)
ORDER BY
  CASE
    WHEN EXTRACT(MONTH FROM order_date) = 1 THEN 1
    WHEN EXTRACT(MONTH FROM order_date) = 2 THEN 2
    WHEN EXTRACT(MONTH FROM order_date) = 3 THEN 3
    WHEN EXTRACT(MONTH FROM order_date) = 4 THEN 4
    WHEN EXTRACT(MONTH FROM order_date) = 5 THEN 5
    WHEN EXTRACT(MONTH FROM order_date) = 6 THEN 6
    WHEN EXTRACT(MONTH FROM order_date) = 7 THEN 7
    WHEN EXTRACT(MONTH FROM order_date) = 8 THEN 8
    WHEN EXTRACT(MONTH FROM order_date) = 9 THEN 9
    WHEN EXTRACT(MONTH FROM order_date) = 10 THEN 10
    WHEN EXTRACT(MONTH FROM order_date) = 11 THEN 11
    ELSE 12
  END
END
```

	bulan text	jumlah_order bigint	total_sales numeric
1	januari	326	19547
2	februari	297	17673
3	maret	336	19587
4	april	313	19076
5	mei	337	19972
6	juni	293	17740
7	juli	329	19839
8	agustus	341	20890
9	september	325	19502
10	oktober	357	21792
11	november	319	18738
12	desember	327	18725

Grafik Trend Analyst



Dari grafik bulanan, jumlah order dan total sales di tahun 2023 cenderung stabil dan tidak memiliki trend yang signifikan baik naik atau turun. Dari hasil grafik ini bisa diketahui bahwa performance dari GenggamData Store kurang. Maka bisa lebih ditingkatkan dalam hal marketing lagi, mulai dari memperbanyak diskon, promosi besar-besaran, atau iklan produk yang lebih massif. Bisa dilihat bahwa jumlah order yang paling tinggi terjadi pada bulan Oktober yaitu sebanyak 357 pesanan dan total sales yang didapatkan sebesar \$21792. Sedangkan jumlah order paling sedikit pada bulan Juni yaitu sebanyak 293 pesanan dan total sales sebesar \$17740. Hal ini bisa dilihat bahwa jumlah order dan total sales polanya sama berbanding lurus, Ketika jumlah order semakin banyak maka total sales yang didapatkan semakin besar.

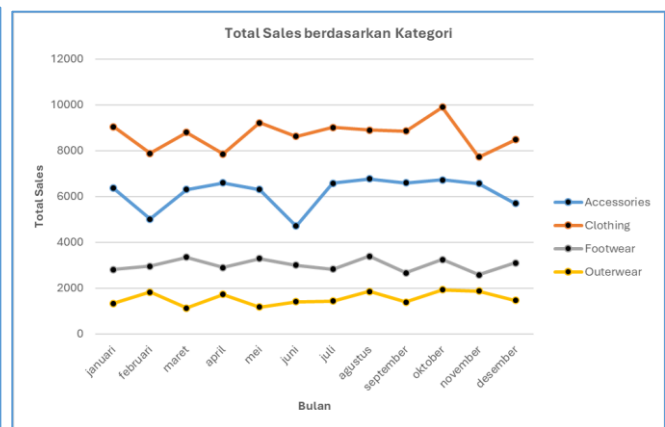
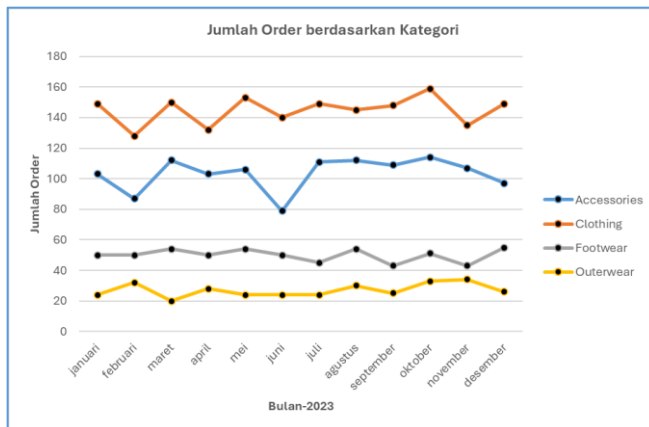
2. Bagaimana overall performance berdasarkan kategori produk GenggamData Store ditahun 2023 untuk jumlah order dan total sales ?

Query dan Output

```
SELECT
CASE
WHEN EXTRACT(MONTH FROM order_date) = 1 THEN 'januari'
WHEN EXTRACT(MONTH FROM order_date) = 2 THEN 'februari'
WHEN EXTRACT(MONTH FROM order_date) = 3 THEN 'maret'
WHEN EXTRACT(MONTH FROM order_date) = 4 THEN 'april'
WHEN EXTRACT(MONTH FROM order_date) = 5 THEN 'mei'
WHEN EXTRACT(MONTH FROM order_date) = 6 THEN 'juni'
WHEN EXTRACT(MONTH FROM order_date) = 7 THEN 'juli'
WHEN EXTRACT(MONTH FROM order_date) = 8 THEN 'agustus'
WHEN EXTRACT(MONTH FROM order_date) = 9 THEN 'september'
WHEN EXTRACT(MONTH FROM order_date) = 10 THEN 'oktober'
WHEN EXTRACT(MONTH FROM order_date) = 11 THEN 'november'
ELSE 'desember'
END AS bulan,
category,
COUNT(*) AS jumlah_order,
SUM(purchase_amount_usd) AS total_sales
FROM
customer_transaction
GROUP BY
EXTRACT(MONTH FROM order_date),category
ORDER BY
CASE
WHEN EXTRACT(MONTH FROM order_date) = 1 THEN 1
WHEN EXTRACT(MONTH FROM order_date) = 2 THEN 2
WHEN EXTRACT(MONTH FROM order_date) = 3 THEN 3
WHEN EXTRACT(MONTH FROM order_date) = 4 THEN 4
WHEN EXTRACT(MONTH FROM order_date) = 5 THEN 5
WHEN EXTRACT(MONTH FROM order_date) = 6 THEN 6
WHEN EXTRACT(MONTH FROM order_date) = 7 THEN 7
WHEN EXTRACT(MONTH FROM order_date) = 8 THEN 8
WHEN EXTRACT(MONTH FROM order_date) = 9 THEN 9
WHEN EXTRACT(MONTH FROM order_date) = 10 THEN 10
WHEN EXTRACT(MONTH FROM order_date) = 11 THEN 11
ELSE 12
END
```

	bulan text	category character varying (512)	jumlah_order bigint	total_sales numeric
1	januari	Footwear	50	2816
2	januari	Accessories	103	6365
3	januari	Outerwear	24	1329
4	januari	Clothing	149	9037
5	februari	Outerwear	32	1819
6	februari	Footwear	50	2954
7	februari	Clothing	128	7888
8	februari	Accessories	87	5012
9	maret	Outerwear	20	1134
10	maret	Footwear	54	3350
11	maret	Accessories	112	6310
12	maret	Clothing	150	8793
13	april	Footwear	50	2901
14	april	Clothing	132	7854
15	april	Accessories	103	6598

Grafik Trend Analyst by Category



Dilihat dalam grafik keduanya, pola jumlah order dan total sales sama. Kategori produk yang paling diminati dan mendapatkan total sales (penjualan) paling tinggi yaitu *Clothing*. Sedangkan untuk urutan kategori yang diminati berikutnya berturut-turut *Accessories*, *Footwear*, dan yang sepi peminat ada *Outerwear*. Namun untuk setiap kategori produk masing-masing penjualan laku terjual berada di bulan yang berbeda. Misalnya *Accessories* dan *Clothing* mencapai puncak penjualan tertinggi di bulan Oktober yaitu sebanyak 114 dan 159, namun total penjualan tertinggi masing-masing di bulan Agustus dan Oktober sebesar \$6760 dan \$9903. Begitu pula dengan kategori yang lain bisa kita lihat pola pembelian dan jumlah pendapatan yang tinggi, sehingga bisa digunakan untuk asumsi penyediaan barang untuk bulan-bulan berikutnya.

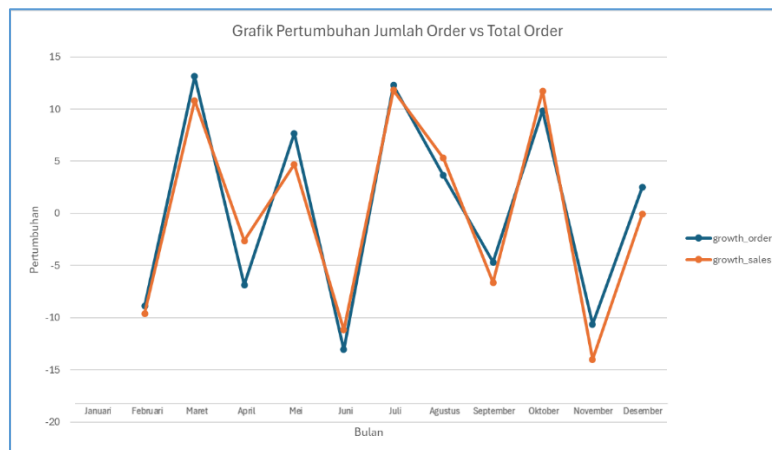
3. Bagaimana growth order dan sales GenggamData Store ditahun 2023 ?

Query dan Output

```
SELECT
    bulan,
    ROUND(((jumlah_order/LAG(jumlah_order) OVER(ORDER BY bulan))-1)*100, 2) as growth_order,
    ROUND(((total_sales/LAG(total_sales) OVER(ORDER BY bulan))-1)*100,2) as growth_sales
FROM
(
    SELECT
        EXTRACT (MONTH FROM order_date) AS bulan,
        CAST(COUNT(*) AS numeric) AS jumlah_order,
        SUM(purchase_amount_usd) AS total_sales
    FROM customer_transaction
    GROUP BY bulan
    ORDER BY bulan
)
```

	bulan numeric	growth_order numeric	growth_sales numeric
1	1	[null]	[null]
2	2	-8.90	-9.59
3	3	13.13	10.83
4	4	-6.85	-2.61
5	5	7.67	4.70
6	6	-13.06	-11.18
7	7	12.29	11.83
8	8	3.65	5.30
9	9	-4.09	-6.64
10	10	9.85	11.74
11	11	-10.64	-14.01
12	12	2.51	-0.07

Grafik Trend Growth Jumlah Order vs Total Sales



Pertumbuhan (*growth*) adalah tingkat dari sesuatu yang diamati (jumlah order dan total sales) dari waktu ke waktu dimana dapat mengukur seberapa baik dari usaha yang dilakukan oleh sebuah perusahaan (GenggamData Store). Dilihat dalam grafik baik jumlah order maupun total sales nilai pertumbuhan selalu naik turun dari bulan ke bulan (cenderung tidak stabil mengikuti trend naik). Dalam hal ini bisa diketahui bahwa Perusahaan GenggamData Store masih perlu meningkatkan kinerja bisnisnya dalam hal untuk meningkatkan jumlah order maupun total penjualannya (sales). Pertumbuhan yang positif bisa diartikan bahwa nilai penjualan meningkat dari periode sebelumnya, begitu sebaliknya jika pertumbuhan negative (<0) maka nilai penjualan sedang turun daripada periode sebelumnya. Bisa dilihat bahwa masih banyak pada bulan Juni dan November jumlah order dan total sales angka pertumbuhannya -10%. Sehingga dalam hal ini GenggamData store perlu membuat banyak inovasi-inovasi untuk langkah ke depannya agar pertumbuhan di setiap bulannya tidak ada yang di bawah 0 (negatif).

4. Bagaimana growth order dan sales GenggamData Store ditahun 2023 berdasarkan kategori produk

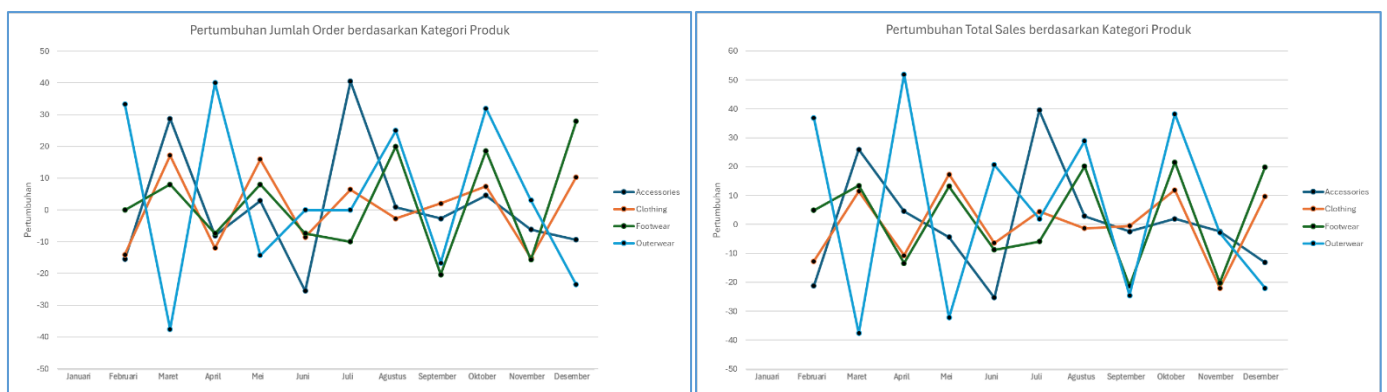
Query

```
SELECT
    bulan,
    category,
    ROUND(((jumlah_order/LAG(jumlah_order) OVER(PARTITION BY category ORDER BY bulan))-1)*100, 2) as growth_order,
    ROUND(((total_sales/LAG(total_sales) OVER(PARTITION BY category ORDER BY bulan))-1)*100,2) as growth_sales
FROM
(
    SELECT
        EXTRACT (MONTH FROM order_date) AS bulan,
        category,
        CAST(COUNT(*) AS numeric) AS jumlah_order,
        SUM(purchase_amount_usd) AS total_sales
    FROM customer_transaction
    GROUP BY bulan, category
    ORDER BY bulan
)
```

Output

	bulan numeric	category character varying (512)	growth_order numeric	growth_sales numeric
1	1	Accessories	[null]	[null]
2	2	Accessories	-15.53	-21.26
3	3	Accessories	28.74	25.90
4	4	Accessories	-8.04	4.56
5	5	Accessories	2.91	-4.41
6	6	Accessories	-25.47	-25.31
7	7	Accessories	40.51	39.50
8	8	Accessories	0.90	2.86
9	9	Accessories	-2.68	-2.47
10	10	Accessories	4.59	1.96
11	11	Accessories	-6.14	-2.45
12	12	Accessories	-9.35	-13.18
13	1	Clothing	[null]	[null]
14	2	Clothing	-14.09	-12.71
15	3	Clothing	17.19	11.47
16	4	Clothing	-12.00	-10.68

Grafik Trend Growth Jumlah Order vs Total Sales by Category



Peningkatan jumlah order dan total sales yang positif menandakan bahwa terjadi peningkatan dari bulan ke bulan daripada periode bulan sebelumnya. Namun bisnis GenggamData Store ini baik secara overall maupun berdasarkan kategori, masih dinilai kinerjanya masing belum baik. Karena masih banyak kategori yang angka pertumbuhannya di bawah 0%. Oleh karena itu, perlu menggiatkan iklan, promosi, diskon atau strategi marketing lainnya agar penjualan bisa meningkat dari bulan ke bulan.

Promotional Cost Efficiency Analysis

Mengukur efisiensi biaya promosi dan menganalisis hubungan antara biaya promosi dengan hasil penjualan. Tujuannya adalah menilai sejauh mana pengeluaran promosi dapat memberikan nilai tambah yang sebanding dengan hasil yang dicapai.

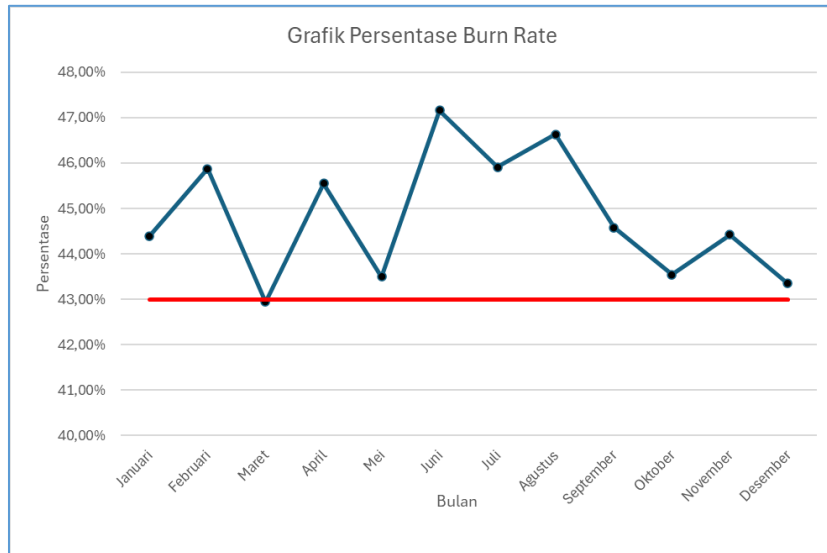
Dengan objektif diatas tim marketing memintamu untuk melakukan analisis dan memberikan insight dari:

1. Efektifitas dan efisiensi promosi yang dilakukan selama tahun 2023, dengan menghitung *burn rate* dari promosi yang dilakukan overall berdasarkan bulan.

Query dan Output

```
SELECT
    bulan,
    total_diskon,
    total_sales,
    ROUND(total_diskon/total_sales*100,2) as burn_rate
FROM
(
    SELECT
        EXTRACT (MONTH FROM order_date) AS bulan,
        SUM(purchase_amount_usd) AS total_sales,
        ROUND(SUM(purchase_amount_usd*discount/100),2) as total_diskon
    FROM customer_transaction
    GROUP BY bulan
    ORDER BY bulan
)
```

	bulan numeric	total_diskon numeric	total_sales numeric	burn_rate numeric
1	1	8677.53	19547	44.39
2	2	8106.39	17673	45.87
3	3	8410.95	19587	42.94
4	4	8689.96	19076	45.55
5	5	8687.12	19972	43.50
6	6	8366.08	17740	47.16
7	7	9107.35	19839	45.91
8	8	9740.11	20890	46.63
9	9	8693.20	19502	44.58
10	10	9488.02	21792	43.54
11	11	8323.74	18738	44.42
12	12	8116.82	18725	43.35



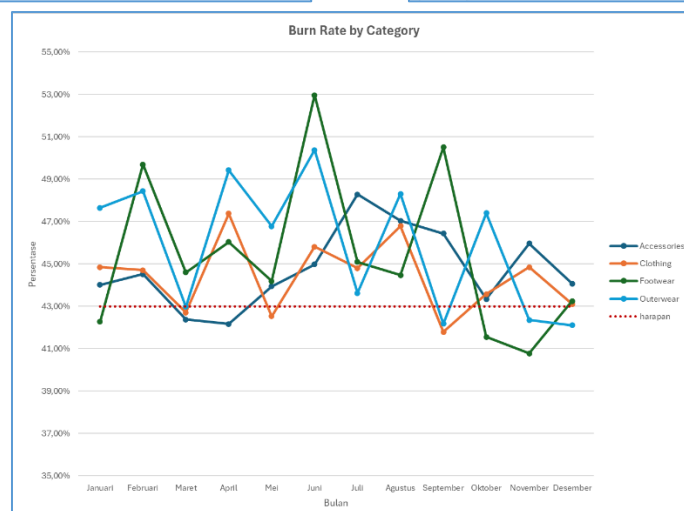
Burn rate adalah nilai untuk mengukur tingkat penggunaan dana dan mengevaluasi seberapa efisien perusahaan dalam mengelola modalnya. Dalam hal ini melihat seberapa efisien total penggunaan dana untuk promosi terhadap hasil penjualan dari GenggamData Store. Persentase *burn rate* semakin besar maka bisa diindikasikan bahwa Perusahaan menggunakan modalnya tinggi tetapi total pendapatan yang dihasilkan masih rendah. Dalam konteks ini, manajemen Perusahaan berharap bahwa *burn rate* Perusahaan tetap berada di angka 43%. Namun bisa dilihat dari hasil *burn rate* tiap bulannya masih >43%. Hanya di bulan Maret saja yang nilai *burn rate* nya di bawah 43%. Dari sini bisa didapatkan insight bahwa biaya promosi yang digunakan oleh Perusahaan kurang efisien sehingga perlu ada evaluasi lebih lanjut mengenai biaya promosi yang berdampak pada tingkat penjualan yang rendah.

2. Efektifitas dan efisiensi promosi yang dilakukan selama tahun 2023, dengan menghitung *burn rate* dari promosi yang dilakukan overall berdasarkan product category dan bulan

Query dan Output

```
SELECT
  bulan,
  category,
  total_diskon,
  total_sales,
  ROUND(total_diskon/total_sales*100,2) as burn_rate
FROM
(
  SELECT
    EXTRACT (MONTH FROM order_date) AS bulan,
    category,
    SUM(purchase_amount_usd) AS total_sales,
    ROUND(SUM(purchase_amount_usd*discount/100),2) as total_diskon
  FROM customer_transaction
  GROUP BY bulan, category
  ORDER BY bulan
)
```

	bulan numeric	category character varying (512)	total_diskon numeric	total_sales numeric	burn_rate numeric
1	1	Footwear	1190.33	2816	42.27
2	1	Accessories	2801.40	6365	44.01
3	1	Outerwear	633.33	1329	47.65
4	1	Clothing	4052.47	9037	44.84
5	2	Outerwear	881.15	1819	48.44
6	2	Footwear	1467.73	2954	49.69
7	2	Clothing	3526.09	7888	44.70
8	2	Accessories	2231.42	5012	44.52
9	3	Outerwear	487.40	1134	42.98
10	3	Footwear	1494.19	3350	44.60
11	3	Accessories	2673.98	6310	42.38
12	3	Clothing	3755.38	8793	42.71
13	4	Footwear	1335.48	2901	46.04



Dilihat dari masing-masing kategori pun, nilai *burn rate* rata-rata masih di atas 43%. Sehingga perlu evaluasi dari masing-masing kaegori terhadap pemberian promosi di setiap bulan terlebih untuk bulan-bulan yang nilai *burn rate* nya masih >43%.

Customer Retention Analysis

Menilai seberapa baik bisnis dapat mempertahankan pelanggan yang sudah ada. Analisis ini melibatkan retensi pelanggan, faktor yang mempengaruhi retensi, dan pengidentifikasian strategi untuk meningkatkan retensi pelanggan.

Dengan objektif diatas tim marketing dan business development memintamu untuk menganalisis dan memberikan insight dari :

1. Bagaimana retensi pelanggan dalam membeli barang di GenggamData Store pada tahun 2023 berdasarkan bulan (tampilkan nilai dalam persentase)

Query

```
WITH
periode_pertama_transaksi AS
(
    SELECT
        customer_id,
        MIN(EXTRACT(MONTH FROM order_date)) as bulan_pertama_transaksi
    FROM customer_transaction
    GROUP BY customer_id
    ORDER BY customer_id
),
bulan_transaksi AS
(
    SELECT DISTINCT customer_id,
        EXTRACT(MONTH FROM order_date) AS bulan
    FROM customer_transaction
    ORDER BY customer_id
),
tabel_gab AS
(
    SELECT ppt.customer_id,
        bt.bulan,
        ppt.bulan_pertama_transaksi,
        bt.bulan - ppt.bulan_pertama_transaksi AS periode_ke
    FROM periode_pertama_transaksi ppt
    JOIN bulan_transaksi bt
    ON ppt.customer_id = bt.customer_id
),

tabel_cohort AS
(
    SELECT bulan_pertama_transaksi,
        SUM(CASE WHEN periode_ke = 0 THEN 1 ELSE 0 END) AS m0,
        SUM(CASE WHEN periode_ke = 1 THEN 1 ELSE 0 END) AS m1,
        SUM(CASE WHEN periode_ke = 2 THEN 1 ELSE 0 END) AS m2,
        SUM(CASE WHEN periode_ke = 3 THEN 1 ELSE 0 END) AS m3,
        SUM(CASE WHEN periode_ke = 4 THEN 1 ELSE 0 END) AS m4,
        SUM(CASE WHEN periode_ke = 5 THEN 1 ELSE 0 END) AS m5,
        SUM(CASE WHEN periode_ke = 6 THEN 1 ELSE 0 END) AS m6,
        SUM(CASE WHEN periode_ke = 7 THEN 1 ELSE 0 END) AS m7,
        SUM(CASE WHEN periode_ke = 8 THEN 1 ELSE 0 END) AS m8,
        SUM(CASE WHEN periode_ke = 9 THEN 1 ELSE 0 END) AS m9,
        SUM(CASE WHEN periode_ke = 10 THEN 1 ELSE 0 END) AS m10,
        SUM(CASE WHEN periode_ke = 11 THEN 1 ELSE 0 END) AS m11
    FROM tabel_gab
    GROUP BY bulan_pertama_transaksi
    ORDER BY bulan_pertama_transaksi
)

SELECT bulan_pertama_transaksi as cohort_month,
    ROUND((m0/CAST(m0 AS numeric))*100, 0) AS m0,
    ROUND((m1/CAST(m0 AS numeric))*100, 0) AS m1,
    ROUND((m2/CAST(m0 AS numeric))*100, 0) AS m2,
    ROUND((m3/CAST(m0 AS numeric))*100, 0) AS m3,
    ROUND((m4/CAST(m0 AS numeric))*100, 0) AS m4,
    ROUND((m5/CAST(m0 AS numeric))*100, 0) AS m5,
    ROUND((m6/CAST(m0 AS numeric))*100, 0) AS m6,
    ROUND((m7/CAST(m0 AS numeric))*100, 0) AS m7,
    ROUND((m8/CAST(m0 AS numeric))*100, 0) AS m8,
    ROUND((m9/CAST(m0 AS numeric))*100, 0) AS m9,
    ROUND((m10/CAST(m0 AS numeric))*100, 0) AS m10,
    ROUND((m11/CAST(m0 AS numeric))*100, 0) AS m11
FROM tabel_cohort
ORDER BY bulan_pertama_transaksi
```

Output

	cohort_month numeric	m0 numeric	m1 numeric	m2 numeric	m3 numeric	m4 numeric	m5 numeric	m6 numeric	m7 numeric	m8 numeric	m9 numeric	m10 numeric	m11 numeric
1	1	100	43	46	44	51	44	47	48	45	54	51	45
2	2	100	57	43	46	48	44	48	45	53	50	42	0
3	3	100	39	36	26	40	36	34	29	33	39	0	0
4	4	100	47	35	27	57	39	31	35	47	0	0	0
5	5	100	32	26	32	37	42	42	32	0	0	0	0
6	6	100	29	38	43	33	24	19	0	0	0	0	0
7	7	100	10	50	20	40	20	0	0	0	0	0	0
8	8	100	13	13	25	38	0	0	0	0	0	0	0
9	9	100	17	33	17	0	0	0	0	0	0	0	0
10	10	100	0	0	0	0	0	0	0	0	0	0	0
11	11	100	0	0	0	0	0	0	0	0	0	0	0
12	12	100	0	0	0	0	0	0	0	0	0	0	0

Retention rate mengacu pada persentase pelanggan atau pengguna yang tetap menggunakan produk atau layanan dari sebuah perusahaan dalam periode waktu tertentu. Ini adalah pengukuran yang penting dalam menilai seberapa efektif sebuah perusahaan dalam mempertahankan pelanggan dan mendorong loyalitas pelanggan.

Secara vertical, tabel cohort digunakan untuk meembandingkan kelompok pelanggan dari waktu ke waktu. Misalnya dilihat dari periode bulan m1, pada bulan 1 (Januari) sebanyak 43% pelanggan melakukan transaksi kemudian di bulan 2 (Februari) terjadi peningkatan jumlah pelanggan yang melakukan transaksi yaitu sebanyak 57%. Namun, seiring berjalannya waktu di 3 bulan terakhir tidak ada pelanggan yang melakukan transaksi Kembali setelah transaksi yang dilakukan pertama kali. Secara horizontal digunakan untuk menganalisis pola setiap perilaku kelompok pelanggan secara spesifik. Rata-rata tingkat retention customer dari waktu ke waktu berada di bawah 50%. Hal ini menunjukkan bahwa tingkat perilaku customer misalnya di bulan Maret 2023, bahwa ada sebanyak 39% dari pelanggan yang melakukan Kembali di bulan selanjutnya, namun pada bulan selanjutnya mengalami penurunan yaitu hanya sebesar 36% yang melakukan transaksi Kembali dari awal dia melakukan transaksi pertama kali

Secara diagonal, melihat peristiwa yang memicu perubahan dalam perilaku kelompok secara keseluruhan. Dapat dilihat bahwa 3 bulan terakhir (Oktober-Desember) terlihat bahwa tidak ada customer yang melakukan pembelian lagi pada transaksi periode setelahnya. Hal ini berarti di bulan tersebut promosi yang dilakukan gagal, sehingga perlu Analisa lebih lanjut karena secara keseluruhan juga jumlah customer yang melakukan transaksi lagi semakin sedikit.

2. Bagaimana persentase churn rate nya?

	cohort_month numeric	churn_m0 numeric	churn_m1 numeric	churn_m2 numeric	churn_m3 numeric	churn_m4 numeric	churn_m5 numeric	churn_m6 numeric	churn_m7 numeric	churn_m8 numeric	churn_m9 numeric	churn_m10 numeric	churn_m11 numeric
1	1	0	57	54	56	49	56	53	52	55	46	49	55
2	2	0	43	57	54	52	56	52	55	47	50	58	100
3	3	0	61	64	74	60	64	66	71	67	61	100	100
4	4	0	53	65	73	43	61	69	65	53	100	100	100
5	5	0	68	74	68	63	58	58	68	100	100	100	100
6	6	0	71	62	57	67	76	81	100	100	100	100	100
7	7	0	90	50	80	60	80	100	100	100	100	100	100
8	8	0	87	87	75	62	100	100	100	100	100	100	100
9	9	0	83	67	83	100	100	100	100	100	100	100	100
10	10	0	100	100	100	100	100	100	100	100	100	100	100
11	11	0	100	100	100	100	100	100	100	100	100	100	100
12	12	0	100	100	100	100	100	100	100	100	100	100	100

Churn rate diartikan sebagai tingkat kehilangan pelanggan atau pengguna dari produk atau layanan perusahaan dalam periode waktu tertentu. Semakin tinggi tingkat persentase churn rate maka, Perusahaan akan kehilangan pelanggan begitu sebaliknya jika persentase churn rate rendah berarti pelanggan loyal dengan Perusahaan. Misalkan di bulan Maret 2023 rata-rata tingkat churn rate sebesar 60%. Berarti bisa disimpulkan bahwa sebesar 605 pelanggan tidak akan melakukan transaksi lagi pada bulan-bulan berikutnya.

Customer Behavior Analysis

Menganalisis perilaku pelanggan untuk memahami kebiasaan pembelian, preferensi produk, dan interaksi dengan merek. Tujuannya adalah untuk memberikan wawasan tentang cara pelanggan berinteraksi dengan bisnis dan untuk mendukung pengambilan keputusan pemasaran yang lebih baik.

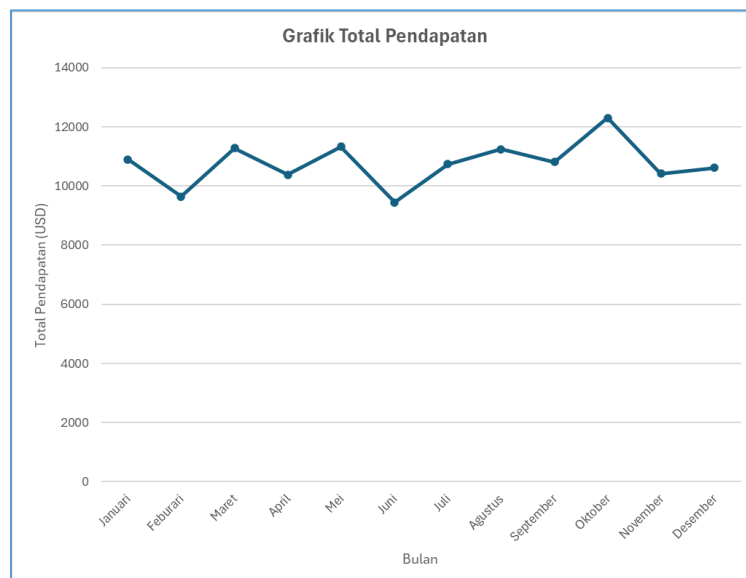
Dengan objektif diatas tim marketing memintamu untuk menganalisis dan memberikan insight dari :

1. Trend Pembelian

a. Bagaimana Trend Pembelian dari bulan ke bulan ?

```
SELECT
CASE
WHEN bulan = '1' THEN 'Januari'
WHEN bulan = '2' THEN 'Februrari'
WHEN bulan = '3' THEN 'Maret'
WHEN bulan = '4' THEN 'April'
WHEN bulan = '5' THEN 'Mei'
WHEN bulan = '6' THEN 'Juni'
WHEN bulan = '7' THEN 'Juli'
WHEN bulan = '8' THEN 'Agustus'
WHEN bulan = '9' THEN 'September'
WHEN bulan = '10' THEN 'Oktober'
WHEN bulan = '11' THEN 'November'
ELSE 'Desember' END as bulan,
total_revenue
FROM
(
SELECT
EXTRACT(MONTH FROM order_date) as bulan,
ROUND(SUM(CASE WHEN discount_applied = 'Yes' THEN (100-discount)/100*purchase_amount_usd
ELSE purchase_amount_usd END),2) AS total_revenue
FROM customer_transaction
GROUP BY bulan
ORDER BY bulan
)
```

	bulan text	total_revenue numeric
1	Januari	10900.97
2	Februrari	9639.66
3	Maret	11277.69
4	April	10386.04
5	Mei	11332.40
6	Juni	9451.20
7	Juli	10739.13
8	Agustus	11238.15
9	September	10808.80
10	Oktober	12303.98
11	November	10421.12
12	Desember	10620.78



Di tahun 2023, trend pembelian dari bulan ke bulan cenderung stabil. Pembelian mencapai puncak tertinggi pendapatan pada bulan Oktober dengan total pendapatan sebesar \$12303,98. Di sisi lain para pelanggan cenderung melakukan sedikit transaksi pada bulan Februari dan Juni dengan masing-masing total pendapatan yang didapatkan dari GenggamData Store sebesar \$9639,66 dan \$9451,20. Total pendapatan yang tinggi ataupun rendah bisa dipengaruhi oleh barang yang diminati oleh pelanggan, sehingga bisa menjadi bahan evaluasi di bulan-bulan selanjutnya.

2. Preferensi Produk

- a. Produk apa yang paling sering dibeli oleh pelanggan?

Query dan Output

```
SELECT
  DENSE_RANK () OVER(ORDER BY COUNT(item_purchased) DESC) as denserank_num,
  item_purchased,
  COUNT(item_purchased) as jumlah_produk
FROM customer_transaction
GROUP BY item_purchased
ORDER BY denserank_num
```

	denserank_num bigint	item_purchased character varying (512)	jumlah_produk bigint
1	1	Jewelry	171
2	1	Pants	171
3	1	Blouse	171
4	2	Shirt	169
5	3	Dress	166
6	4	Sweater	164
7	5	Jacket	163
8	6	Sunglasses	161
9	6	Belt	161
10	6	Coat	161

Sebelumnya bisa dilihat trend pendapatan di setiap bulannya, dari sana bisa dihubungkan untuk mengetahui barang apa yang sering dibeli sehingga dapat meningkatkan pendapatan. Dari hasil query bisa dilihat item produk apa yang paling diminati customer atau sebaliknya. Top 3, item yang sering dibeli customer yaitu (Jewelry, Pants, Blouse) dengan total order sama, selanjutnya Shirt dan Dress. Sedangkan untuk barang yang kurang diminati yaitu Jeans yang berada di urutan terakhir. Untuk selanjutnya bisa dievaluasi kira-kira produk apa saja yang harus memiliki stok berlebih untuk persiapan bulan-bulan berikutnya atau sebaliknya bisa meningkatkan penjualan dari produk-produk yang berada di rank bawah agar bisa meningkatkan pendapatan.

- b. Apakah ada produk yang lebih disukai oleh pelanggan tertentu? (Misal marketing ingin membuat iklan yang akan disebar berdasarkan gender)

Query dan Output

```
SELECT
  DENSE_RANK() OVER(ORDER BY female DESC) AS denserank_num,
  item_purchased,
  female
FROM (
  SELECT
    item_purchased,
    SUM(CASE WHEN gender = 'Male' THEN 1 ELSE 0 END) AS male,
    SUM(CASE WHEN gender = 'Female' THEN 1 ELSE 0 END) AS female
  FROM customer_transaction
  GROUP BY item_purchased
) AS ringkasan_pembelian
ORDER BY denserank_num
```

```
--male
SELECT
  DENSE_RANK() OVER(ORDER BY male DESC) AS denserank_num,
  item_purchased,
  male
FROM (
  SELECT
    item_purchased,
    SUM(CASE WHEN gender = 'Male' THEN 1 ELSE 0 END) AS male,
    SUM(CASE WHEN gender = 'Female' THEN 1 ELSE 0 END) AS female
  FROM customer_transaction
  GROUP BY item_purchased
) AS ringkasan_pembelian
ORDER BY denserank_num
```

	denserank_num bigint	item_purchased character varying (512)	female bigint
1	1	Blouse	66
2	2	Shirt	59
3	2	Sandals	59
4	3	Socks	58
5	3	Handbag	58
6	4	Sunglasses	56
7	5	Belt	55
8	6	Jacket	54
9	7	Dress	52
10	7	Jewelry	52

	denserank_num bigint	item_purchased character varying (512)	male bigint
1	1	Pants	123
2	2	Jewelry	119
3	3	Dress	114
4	3	Coat	114
5	3	Sweater	114
6	4	Scarf	112
7	5	Shirt	110
8	6	Skirt	109
9	6	Shorts	109
10	6	Jacket	109

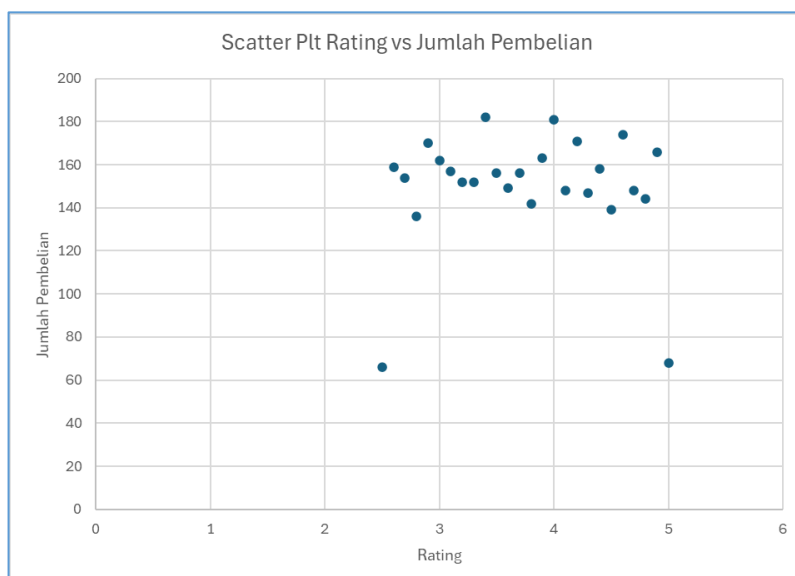
Dari hasil query bisa dilihat bahwa produk-produk yang diminati oleh wanita top 3 rank di tahun 2023 diantaranya Blouse, Shirt, Sandals, Sooks, dan Handbag. Sebagian besar produk wanita yang sering dibeli adalah produk-produk untuk dirinya sendiri. Sedangkan untuk produk pria ada sedikit produk yang mungkin bisa saja menjadi analisis lebih lanjut yaitu rank 3 ditempati oleh posisi produk Dress. Berbeda di gender wanita, Dress masuk di posisi rank 7. Walaupun termasuk produk wanita, hal ini bisa saja terjadi. Diasumsikan bahwa pria banyak membeli dress untuk hadiah orang tersayang, seperti pacar, ibu, atau lainnya. Hal ini bisa menjadi bahan pertimbangan bagi produk-produk yang sering dibeli untuk bisa meningkatkan stok produknya. Tim marketing juga bisa mendapatkan gambaran bahwa iklan produk wanita tidak hanya diberikan oleh gender wanita saja, bisa menjadi referensi untuk gender pria untuk ditampilkan di iklan-iklan media sosial.

3. Korelasi Rating dan Review

- Apakah rating berpengaruh kepada jumlah pembelian? Gunakan Pearson Correlation & buat scatter plot nya

```
SELECT
  CORR(review_rating, jumlah_pembelian) AS pearson_correlation
FROM
(
  SELECT
    review_rating,
    COUNT(customer_id) as jumlah_pembelian
  FROM customer_transaction
  GROUP BY review_rating
)
```

	pearson_correlation double precision
1	-0.00038474951539880977



Korelasi Pearson adalah sebuah ukuran statistik yang menggambarkan seberapa erat hubungan linear antara dua variabel. Nilai korelasi Pearson berkisar antara -1 hingga 1, nilai korelasi mendekati 1 berarti bahwa ada hubungan yang sangat kuat antara 2 variabel yang diamati. Korelasi Pearson antara pemberian rating terhadap jumlah pembelian yaitu sebesar -0,0003847. Dapat disimpulkan bahwa nilai ini sangat menjauhi nilai 1 dan bertanda negatif, sehingga dapat diartikan bahwa memiliki korelasi negatif sebesar 0,0003847 atau bisa disimpulkan bahwa **tidak memiliki pengaruh yang signifikan antara pemberian rating terhadap jumlah pembelian**. Korelasi negatif juga bisa diartikan bahwa ketika satu variabel tinggi, maka variabel yang lain sebaliknya. Hal ini juga bisa diamati dari pemberian rating 5 (pada titik scatter plot di bagian bawah), bahwa pelanggan memberikan rating 5 namun jumlah pembeliannya sedikit.

4. Pola Diskon

- a. Berapa sering pelanggan menggunakan diskon?

Overall

```
SELECT
    pengguna_diskon,
    jumlah_pengguna,
    ROUND(CAST (pengguna_diskon AS numeric)/CAST (jumlah_pengguna AS numeric)*100,2) AS persentase_penggunaan_diskon
FROM
(
    SELECT
        SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) AS penggunaan_diskon,
        COUNT(discount_applied) AS jumlah_pengguna
    FROM customer_transaction
) AS tabel_persentase
```

	penggunaan_diskon	jumlah_pengguna	persentase_penggunaan_diskon
	bigint	bigint	numeric
1	3883	3900	99.56

Seberapa sering customer atau pelanggan menggunakan diskon bisa dilihat dari banyaknya pelanggan yang menggunakan diskon dibagi dengan total keseluruhan dari transaksi. Bisa dilihat bahwa selama tahun 2023, secara keseluruhan bisa dikatakan mayoritas seluruh pelanggan menggunakan diskon untuk setiap transaksi yang dilakukan. Hanya 0,44% atau sebanyak 17 transaksi saja customer tidak menggunakan diskon. Sehingga adanya diskon ini bisa berpengaruh besar bagi Keputusan customer dalam membeli sebuah barang. Untuk selanjutnya bisa menjadi bahan untuk memperbanyak lagi pemberian diskon dan promo lainnya agar customer tetap sering melakukan pembelian di GenggamData Store.

By customer

```
SELECT
    customer_id,
    transaksi_diskon,
    jumlah_transaksi,
    ROUND(CAST (transaksi_diskon AS numeric)/CAST (jumlah_transaksi AS numeric)*100,2) AS persentase_penggunaan_diskon
FROM
(
    SELECT
        customer_id,
        SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) AS transaksi_diskon,
        COUNT(customer_id) AS jumlah_transaksi
    FROM customer_transaction
    GROUP BY customer_id
    ORDER BY customer_id
) AS tabel_persentase

ORDER BY persentase_penggunaan_diskon
```

	customer_id integer	transaksi_diskon bigint	jumlah_transaksi bigint	persentase_penggunaan_diskon numeric
1	538	1	2	50.00
2	360	2	3	66.67
3	550	2	3	66.67
4	464	3	4	75.00
5	82	3	4	75.00
6	353	4	5	80.00
7	420	4	5	80.00
8	502	5	6	83.33
9	26	7	8	87.50
10	299	9	10	90.00

Selain bisa dilihat berdasarkan keseluruhan dari transaksi yang dilakukan customer, bisa dilihat juga customer mana saja yang sering menggunakan diskon untuk melakukan transaksi atau sebaliknya. Yang menjadi perhatian disini mungkin bisa dilihat dari customer mana yang jarang menggunakan diskonnya. Hal ini bisa diasumsikan bahwa customer tersebut jarang mendapatkan diskon atau bisa saja lebih menyukai promo selain diskon. Mungkin evaluasi selanjutnya, bisa dipertimbangkan dengan pemberian diskon bagi pelanggan yang minim mendapatkan diskon agar tetap bisa menikmati promo-promo lainnya agar meningkatkan total pendapatan.

Link Query Sql : <https://github.com/ennisawitri/fp-sql>