



MySkill



Final Project

Data Analysis using SQL in PostgreSQL

Kelompok B4 Batch 19





Amos Christ Kevin

Dani Harsalisman Febrian Putra

Irfan Rasyid

Muhamad Fatih Praga Ilhaq

Reza Muhamad Ricky

Adinda Aulia Hudianti

Dickna Nenden Woro Ramadityawati

Hanif Dwi Satria

Marchenko Svesda

Nando Rifki Utama Putra

Sharfina Zahra Akbar



Data Set

Overview

Dataset ini bermanfaat untuk memperoleh wawasan mendalam tentang pola perilaku pelanggan, efisiensi pelanggan, serta mengoptimalkan strategi penjualan produk.

Dataset yang digunakan:

1. order_detail
2. sku_detail
3. customer_detail
4. payment_detail

Data yang digunakan adalah data yang berasal dari Tokopedia (**bukan data sesungguhnya**).



Data Set Information

order_detail:

- | | |
|--------------------|--|
| 1. id | → angka unik dari order / <i>id_order</i> |
| 2. customer_id | → angka unik dari pelanggan |
| 3. order_date | → tanggal saat dilakukan transaksi |
| 4. sku_id | → angka unik dari produk (sku adalah <i>stock keeping unit</i>) |
| 5. price | → harga yang tertera pada tagging harga |
| 6. qty_ordered | → jumlah barang yang dibeli oleh pelanggan |
| 7. before_discount | → nilai harga total dari produk ($price * qty_ordered$) |
| 8. discount_amount | → nilai diskon product total |
| 9. after_discount | → nilai harga total produk ketika sudah dikurangi dengan diskon |
| 10. is_gross | → menunjukkan pelanggan belum membayar pesanan |
| 11. is_valid | → menunjukkan pelanggan sudah melakukan pembayaran |
| 12. is_net | → menunjukkan transaksi sudah selesai |
| 13. payment_id | → angka unik dari metode pembayaran |

Data Set Information

sku_detail:

- | | |
|---------------|---|
| 1. id | → angka unik dari produk (dapat digunakan untuk <i>key</i> saat join) |
| 2. sku_name | → nama dari produk |
| 3. base_price | → harga barang yang tertera pada tagging harga / <i>price</i> |
| 4. cogs | → <i>cost of goods sold</i> / total biaya untuk menjual 1 produk |
| 5. category | → kategori produk |

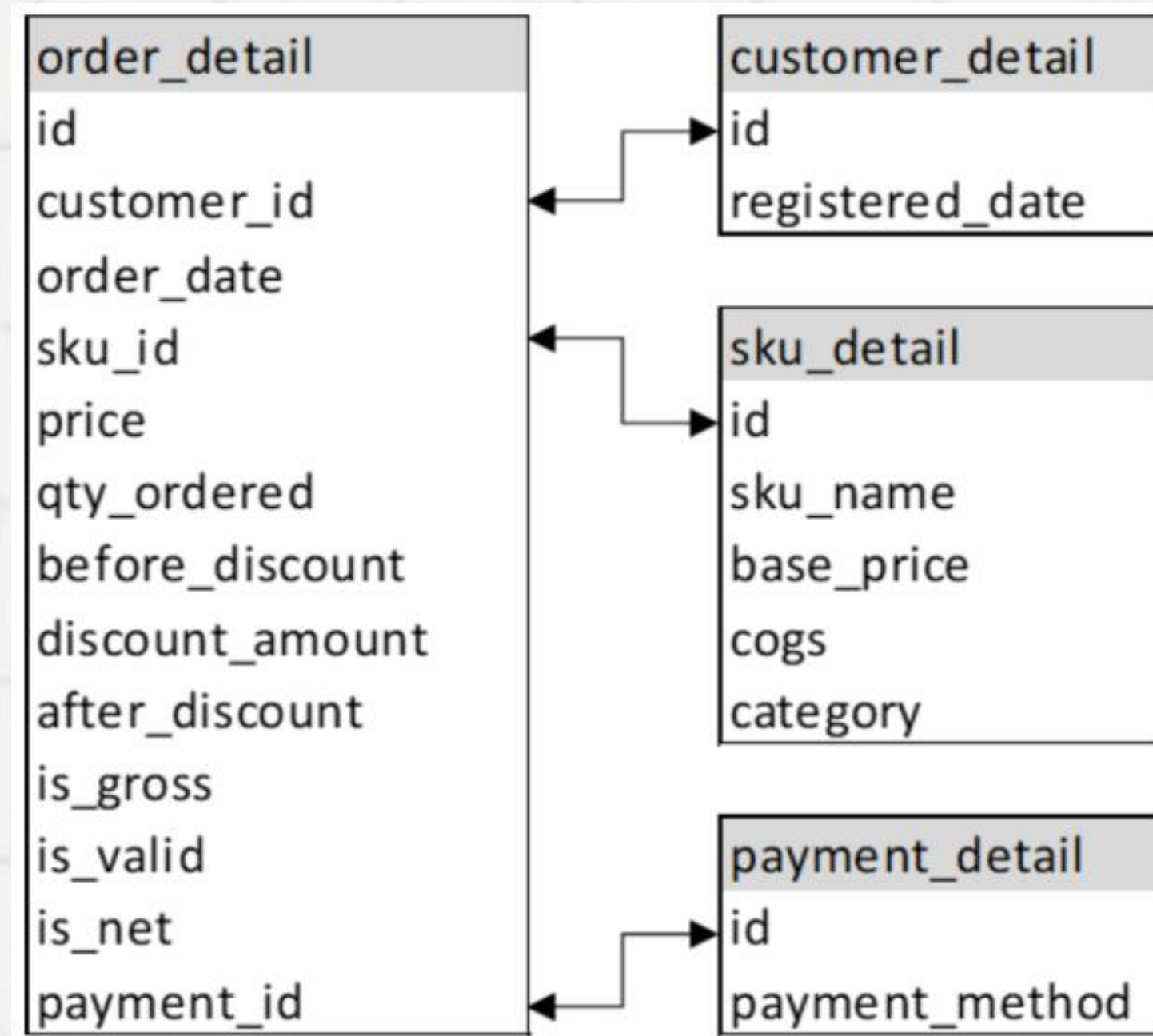
customer_detail:

- | | |
|--------------------|---|
| 1. id | → angka unik dari pelanggan |
| 2. registered_date | → tanggal pelanggan mulai mendaftarkan diri sebagai anggota |

Payment_detail:

- | | |
|-------------------|-------------------------------------|
| 1. id | → angka unik dari metode pembayaran |
| 2. payment_method | → metode pembayaran yang digunakan |

Skema Tabel



Beberapa data saling berhubungan, yang dijelaskan melalui skema tabel di samping.

Data dari order_detail berfungsi sebagai data utama (primary key) yang terhubung dengan data lainnya, seperti customer_detail, sku_detail, dan payment_detail.

Pertanyaan:

01.

Selama transaksi yang terjadi selama 2021, pada bulan apa total nilai transaksi (after_discount) paling besar? Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail

02.

Selama transaksi pada tahun 2022, kategori apa yang menghasilkan nilai transaksi paling besar? Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

03.

Bandingkan nilai transaksi dari masing-masing kategori pada tahun 2021 dengan 2022. Sebutkan kategori apa saja yang mengalami peningkatan dan kategori apa yang mengalami penurunan nilai transaksi dari tahun 2021 ke 2022. Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

04.

Tampilkan top 5 metode pembayaran yang paling populer digunakan selama 2022 (berdasarkan total unique order). Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, payment_method

05.

Urutkan dari ke-5 produk ini berdasarkan nilai transaksinya.

1. Samsung
2. Apple
3. Sony
4. Huawei
5. Lenovo

Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

Jawaban

01. Selama transaksi yang terjadi selama 2021, pada bulan apa total nilai transaksi (after_discount) paling besar? Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail

```
1  select
2      extract(month from order_date) as month,
3      sum(after_discount) as total_transaction
4  from
5      order_detail
6  where
7      extract(year from order_date) = 2021
8      and is_valid = 1
9  group by
10     1
11  order by
12     2 desc
13  limit 5; |
```


Jawaban

Penjelasan Query

Fungsi Select dan Agregat

- **select:** Memilih kolom atau nilai yang akan ditampilkan
- **extract(month from order_date) as month:** Mengambil nilai bulan dari kolom order_date dan memberi label "month"
- **sum(after_discount) as total_transaction:** Menjumlahkan nilai dari kolom after_discount untuk setiap bulan dan diberi label "total_transaction"

Sumber Data:

- **from order_detail:** Mengambil data dari tabel order-detail yang berisi rincian pesanan atau transaksi

Penjelasan Query

Klausula Where:

- **where extract (year from order_date) = 2021 and is_valid=1:** Membatasi hasil hanya pada transaksi yang terjadi di tahun 2021 dan telah divalidasi (dengan is_valid = 1 menunjukkan pembayaran telah dilakukan).

Fungsi Group By, Filter, dan Limit:

- **group by 1:** Mengelompokkan data berdasarkan bulan (month) yang diambil dari kolom order_date.
- **order by 2 desc:** Mengurutkan hasil berdasarkan total nilai transaksi terbesar
- **limit 5:** Menampilkan hanya 5 hasil teratas.

Jawaban

01.

	month numeric	total_transaction double precision
1	8	227862744
2	12	217309963
3	10	207603259.9
4	11	180396009.73999998
5	7	148007735

Insight

Berdasarkan data yang diperoleh, bulan ke-8 atau Agustus tahun 2021 memiliki total transaksi tertinggi sebesar 227.862.744

Analisis Lanjutan

Mengidentifikasi faktor penyebab meningkatnya nilai total transaksi pada Agustus 2021. Apakah hal ini berkaitan dengan promosi yang dilakukan oleh tim marketing selama bulan tersebut atau ada faktor lain? Jika berkaitan dengan tim marketing, apakah kampanye promosi tim marketing berhasil menarik lebih banyak pelanggan? Atau, apakah faktor eksternal seperti tren musiman, perayaan hari besar, atau peluncuran produk baru turut berperan?

Jawaban

02

Selama transaksi pada tahun 2022, kategori apa yang menghasilkan nilai transaksi paling besar? Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

```
1  select
2      s.category as category,
3      sum(after_discount) as total_transaction_value
4  from
5      order_detail o
6  left join sku_detail s
7  on o.sku_id = s.id
8  where extract (year from o.order_date) = 2022
9  and o.is_valid = 1
10 group by
11     1
12 order by
13     2 desc;
```

Jawaban

Penjelasan Query

Fungsi Select dan Agregat:

- **select s.category as category:** Memilih kolom category dari tabel sku_detail untuk ditampilkan dan diberi label sebagai "kategori"
- **sum(o.after_discount) as total_transaction_value:** Menjumlahkan nilai dari kolom after_discount untuk setiap bulan, diberi label "total_transaction_value"

Sumber Data dan Join:

- **from order_detail o:** Mengambil data dari tabel order_detail yang berisi rincian pesanan atau transaksi
- **left join sku_detail s on o.sku_id = s.id:** Menggabungkan tabel order_detail dan sku_detail menggunakan kolom sku_id dari order_detail dan id dari sku_detail sebagai penghubung

Penjelasan Query

Klausula Where:

- **where extract (year from o.order_date) = 2022 and o.is_valid = 1:** Membatasi data yang ditampilkan hanya pada transaksi yang terjadi di tahun 2022 dan yang sudah valid (ditandai dengan is_valid = 1)

Fungsi Group By dan Filter:

- **group by 1:** Mengelompokkan data berdasarkan kolom kategori produk
- **order by 2 desc:** Mengurutkan hasil berdasarkan total transaksi (after_discount), dari yang tertinggi ke yang terendah

Jawaban

02 Insight

Berdasarkan data yang dikumpulkan, kategori produk Mobile & Tablets mencatat nilai transaksi tertinggi sepanjang tahun 2022, dengan total transaksi mencapai 918.451.576.

	category text	total_transaction_value double precision
1	Mobiles & Tablets	918451576
2	Entertainment	365344148.9
3	Appliances	316358100
4	Computing	214028543.4
5	Men Fashion	135588253
6	Women Fashion	93014970.62
7	Home & Living	79483716.2
8	Health & Sports	54235579.6
9	Beauty & Grooming	46211019.18
10	Superstore	32643266.52
11	Kids & Baby	25931276.84
12	Others	21744646.02
13	Soghaat	17658332
14	School & Education	17362465.3
15	Books	6792519.2

Analisis Lanjutan

- Produk dalam kategori Mobile & Tablets memiliki potensi besar untuk meningkatkan pendapatan perusahaan di tahun mendatang. Oleh karena itu, perusahaan dapat memusatkan perhatian pada pengelolaan stok dan strategi pemasaran yang efektif untuk kategori ini.
- Dengan melakukan riset pasar mendalam dan mengidentifikasi tren teknologi terkini, perusahaan dapat merancang produk inovatif yang mampu meningkatkan loyalitas pelanggan. Di samping itu, promosi yang terarah, seperti potongan harga atau bundling dengan kategori lain, akan menarik lebih banyak pembeli potensial dan meningkatkan volume transaksi.

Jawaban

03

Bandingkan nilai transaksi dari masing-masing kategori pada tahun 2021 dengan 2022. Sebutkan kategori apa saja yang mengalami peningkatan dan kategori apa yang mengalami penurunan nilai transaksi dari tahun 2021 ke 2022. Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

```
1 with annual_transaction as (  
2     select  
3         s.category as category,  
4         sum(case when extract(year from o.order_date) = 2021 then after_discount end) as total_sales_2021,  
5         sum(case when extract(year from o.order_date) = 2022 then after_discount end) as total_sales_2022,  
6         sum(case when extract(year from o.order_date) = 2022 then after_discount end)  
7         -  
8         sum(case when extract(year from o.order_date) = 2021 then after_discount end) as different  
9     from  
10        order_detail as o  
11    left join sku_detail as s  
12        on o.sku_id = s.id  
13    where  
14        extract(year from o.order_date) in (2021, 2022)  
15        and o.is_valid = 1  
16    group by 1  
17    order by 4 desc  
18 )  
19 select  
20     category,  
21     total_sales_2021,  
22     total_sales_2022,  
23     different,  
24     case when total_sales_2022 < total_sales_2021 then 'Menurun'  
25     else 'Meningkat'  
26     end as Status  
27 from annual_transaction;
```

Jawaban

Penjelasan Query

Fungsi CTE (Common Table Expression):

- **with annual_transaction as (...):** CTE ini bertugas menghitung total penjualan tahunan (2021 dan 2022) berdasarkan kategori, serta selisih antara total penjualan tahun 2022 dengan tahun 2021

Fungsi Select dan Agregat:

- **select s.category as category:** Memilih kolom kategori yang diambil dari tabel sku_detail
- **sum(case when extract(year from o.order_date) = 2021 then after_discount end) as total_sales_2021:** Menghitung total penjualan pada tahun 2021
- **sum(case when extract(year from o.order_date) = 2022 then after_discount end) as total_sales_2022:** Menghitung total penjualan pada tahun 2022
- **sum(case when extract(year from o.order_date) = 2021 then after_discount end) as different:** Menghitung selisih antara total penjualan 2022 dan 2021

Penjelasan Query

Sumber Data dan Join:

- **from order_detail o:** Mengambil data dari tabel order_detail yang berisi rincian pesanan atau transaksi
- **left join sku_detail s on o.sku_id = s.id:** Menggabungkan tabel order_detail dan sku_detail menggunakan kolom sku_id dari order_detail dan id dari sku_detail sebagai penghubung

Klausula Where:

- **where extract(year from o.order_date) in (2021, 2022) and o.is_valid = 1:** Membatasi hasil hanya untuk tahun 2021 dan 2022, serta hanya untuk transaksi yang valid (is_valid = 1)

Fungsi Group By dan Filter

- **group by 1:** Mengelompokkan data berdasarkan kategori
- **order by 4 desc:** Mengurutkan hasil berdasarkan selisih penjualan, dari yang tertinggi ke yang terendah

Query

- **select * case when total_sales_2022 < total_sales_2021 then 'Meningkat' when total_sales_2022 > total_sales_2021 then 'Menurun' else 'Stabil' end as status:** Jika penjualan tahun 2021 lebih kecil dibandingkan dengan 2022 akan ditampilkan 'Meningkat', jika sebaliknya ditampilkan 'Menurun' di kolom status berdasarkan data dari CTE annual_transaction

Jawaban

03 Insight

Kategori produk yang mengalami penurunan nilai transaksi pada tahun 2022 adalah Books dan Others, sementara kategori lainnya mengalami peningkatan.

	category text	total_sales_2021 double precision	total_sales_2022 double precision	different double precision	status text
1	Mobiles & Tablets	370606718	918451576	547844858	Meningkat
2	Entertainment	162326357.4	365344148.9	203017791.49999997	Meningkat
3	Appliances	218550177	316358100	97807923	Meningkat
4	Men Fashion	58628198	135588253	76960055	Meningkat
5	Computing	172878860	214028543.4	41149683.400000006	Meningkat
6	Home & Living	45797873	79483716.2	33685843.2	Meningkat
7	Health & Sports	33837965.6	54235579.6	20397614	Meningkat
8	Women Fashion	84045961.4	93014970.62	8969009.219999999	Meningkat
9	School & Education	11558982.4	17362465.3	5803482.9	Meningkat
10	Superstore	28828088	32643266.519999996	3815178.519999996	Meningkat
11	Soghaat	15056202.6	17658332	2602129.4000000004	Meningkat
12	Kids & Baby	23971057.8	25931276.84	1960219.039999999	Meningkat
13	Beauty & Grooming	46047360	46211019.18	163659.1799999997	Meningkat
14	Books	10124596	6792519.2	-3332076.8	Menurun
15	Others	40468515.74	21744646.02	-18723869.720000003	Menurun

Analisis Lanjutan

- Kategori yang meningkat dapat dioptimalkan dengan perbaikan strategi promosi dan perluasan ragam penawaran produk sehingga perusahaan dapat mempertahankan atau bahkan meningkatkan pendapatan.
- Sebaliknya, penting untuk mengidentifikasi faktor-faktor yang menyebabkan penurunan transaksi pada kategori Books dan Others. Dengan memahami penyebab tersebut, perusahaan dapat merumuskan strategi pemasaran yang lebih efektif dan meningkatkan kualitas produk untuk mengatasi penurunan serta mendorong pertumbuhan pendapatan secara keseluruhan.

Jawaban

04

Tampilkan top 5 metode pembayaran yang paling populer digunakan selama 2022 (berdasarkan total unique order). Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, payment_method

```
1  ✓ select
2      p.payment_method,
3      count(distinct o.id) as total_transaction
4  from
5      order_detail o
6      left join payment_detail p on o.payment_id = p.id
7  where
8      extract (year from order_date) = 2022
9      and is_valid = 1
10 group by
11     1
12 order by
13     2 desc;
```

Jawaban

Penjelasan Query

Fungsi Select dan Agregat:

- **select p.payment_method:** Memilih kolom payment_method dari tabel payment_detail untuk ditampilkan
- **count(distinct o.id) as total_transaction:** Menghitung jumlah transaksi unik berdasarkan kolom id dari tabel order_detail. Fungsi DISTINCT digunakan untuk memastikan hanya transaksi unik yang dihitung, menghindari duplikasi. Diberi label "total_transaction"

Sumber Data dan Join:

- **from order_detail o:** Memilih tabel order_detail sebagai sumber data utama dan memberi label "o".
- **left join payment_detail p on o.payment_id = p.id:** Menggabungkan tabel order_detail dan payment_detail menggunakan kolom payment_id dari order_detail dan id dari payment_detail sebagai penghubung.

Penjelasan Query

Klausula Where

- **where extract(year from order_date) = 2022 and is_valid = 1:** Memastikan hanya data transaksi pada tahun 2022 yang diambil

Fungsi Group By dan Filter:

- **group by 1:** Mengelompokkan hasil berdasarkan bulan yang diambil dari kolom order_date
- **order by 2 desc:** Mengurutkan hasil berdasarkan nilai transaksi terbesar

Jawaban

04 Insight

Metode pembayaran yang paling sering digunakan oleh konsumen selama tahun 2022 adalah COD, Payaxis, Customer Credit, Easypay ,dan Jazzwallet.

Analisis Lanjutan

	payment_method text	total_transaction bigint
1	cod	1809
2	Payaxis	181
3	customercredit	75
4	Easypay	69
5	jazzwallet	26
6	jazzvoucher	9
7	cashatdoorstep	6
8	financesettlement	2
9	easypay_voucher	2

- Penting untuk mengevaluasi pengalaman pengguna pada metode pembayaran non-tunai seperti Payaxis, Customer Credit, Easypay, dan Jazzwallet. Dengan mengumpulkan umpan balik konsumen tentang kenyamanan, kecepatan transaksi, dan tingkat kepuasan, perusahaan bisa mendapatkan ide untuk meningkatkan penggunaan metode tersebut.
- Di sisi lain, perlu juga memahami mengapa banyak konsumen masih lebih memilih metode pembayaran COD. Apakah karena lebih praktis, lebih dipercaya, atau ada alasan lain? Mengetahui penyebab ini bisa membantu merancang strategi yang tepat untuk menjaga popularitas COD sekaligus mendorong lebih banyak orang menggunakan metode pembayaran non-tunai. Hal ini juga membuka peluang untuk memperbaiki promosi dan memberikan edukasi agar metode non-tunai lebih diminati.

Jawaban

05

Urutkan dari ke-5 produk ini berdasarkan nilai transaksinya.

1. Samsung
2. Apple
3. Sony
4. Huawei
5. Lenovo

Gunakan is_valid = 1 untuk memfilter data transaksi.

Source table: order_detail, sku_detail

```
--5
SELECT
  CASE
    WHEN s.sku_name LIKE '%Samsung%' THEN 'Samsung'
    WHEN s.sku_name LIKE '%Apple%'
      OR s.sku_name LIKE '%MACBOOK%'
      OR s.sku_name LIKE '%IMAC%' THEN 'Apple'
    WHEN s.sku_name LIKE '%Sony%' THEN 'Sony'
    WHEN s.sku_name LIKE '%Huawei%' THEN 'Huawei'
    WHEN s.sku_name LIKE '%Lenovo%' THEN 'Lenovo'
    ELSE 'Lainnya'
  END AS Produk,
  SUM(o.after_discount) AS Total_Transaksi
FROM
  order_detail AS o
JOIN
  sku_detail AS s
ON
  o.sku_id = s.id
WHERE
  o.is_valid = 1
  AND (
    s.sku_name LIKE '%Samsung%'
    OR s.sku_name LIKE '%Apple%'
    OR s.sku_name LIKE '%MACBOOK%'
    OR s.sku_name LIKE '%IMAC%'
    OR s.sku_name LIKE '%Sony%'
    OR s.sku_name LIKE '%Huawei%'
    OR s.sku_name LIKE '%Lenovo%'
  )
GROUP BY
  Produk
ORDER BY
  Total_Transaksi DESC;
```

Jawaban

Penjelasan Query

Fungsi Select dan Agregat:

- **select case when s.sku_name like '%Samsung%' then 'Samsung':** Mengelompokkan data berdasarkan merek dengan mencocokkan nama produk (sku_name) menggunakan pola LIKE. Contohnya, Jika sku_name mengandung kata Samsung, maka akan dikelompokkan ke "Samsung".
- **sum(o.after_discount) as total_transaksi:** Menghitung total nilai transaksi setelah diskon (after_discount) untuk setiap kelompok merek yang dihasilkan oleh pernyataan CASE.

Sumber Data dan Fungsi Join:

- **from order_detail as o:** Data utama diambil dari tabel order_detail, diidentifikasi dengan label o.
- **join sku_detail as s on o.sku_id = s.id:** Menggabungkan tabel order_detail dengan sku_detail menggunakan relasi antara sku_id di order_detail dan id di sku_detail.

Penjelasan Query

Klausula Where:

- **where o.is_valid = 1:** Hanya memilih data dari tabel order_detail di mana transaksi valid (is_valid = 1)
- **and s.sku_name like:** Memfilter data berdasarkan kata kunci di nama SKU. Hanya nama SKU yang mengandung salah satu kata kunci seperti "Samsung", "Apple", dll., yang akan diproses.

Fungsi Group By dan Filter:

- **group by 1:** Mengelompokkan hasil berdasarkan produk
- **order by 2 desc:** Mengurutkan hasil berdasarkan nilai transaksi terbesar

Jawaban

05 Insight

Samsung mencatat nilai transaksi tertinggi sebesar 380,8 juta, diikuti oleh Apple dengan total 232,3 juta, yang mencakup produk MacBook dan iPhone. Sementara itu, Sony mencatat nilai transaksi sebesar 43 juta, Huawei sebesar 31,1 juta, dan Lenovo sekitar 8,2 juta. Dominasi Samsung dan Apple ini mencerminkan popularitas produk mereka, sementara merek lain tertinggal.

Analisis Lanjutan

	produk text	total_transaksi double precision
1	Samsung	380818894
2	Apple	232392080
3	Huawei	43092260
4	Sony	31127440
5	Lenovo	8199402

- Untuk itu, perusahaan melalui strategi yang bisa ditingkatkan antara lain promosi pada produk Sony, Huawei, dan Lenovo untuk meningkatkan nilai transaksi. Selanjutnya, perusahaan juga melakukan peningkatan stock produk Samsung dan Apple yang memiliki popularitas tinggi untuk meraih maksimal keuntungan memenuhi permintaan konsumen yang meningkat. Dengan analisis ulasan dan umpan balik pelanggan, kekuatan dan kelemahan dari suatu produk dapat terdeteksi dan memberikan rekomendasi terkait pada tim pemasaran atau pengembangan produk mengenai peningkatan fitur atau penyelesaian masalah yang sering dikeluhkan.
- Strategi promosi juga perlu disesuaikan berdasarkan performa dan demografi konsumen, dengan menggunakan data penjualan sebelumnya untuk mengidentifikasi produk mana yang membutuhkan dorongan lebih dengan kampanye atau diskon khusus.

Kesimpulan

Proyek Tokopedia ini berhasil memberikan pemahaman mendalam terkait transaksi penjualan selama tahun 2021 dan 2022. Berdasarkan temuan proyek ini, Agustus 2021 tercatat sebagai bulan dengan transaksi tertinggi pada tahun tersebut, yang dapat dijadikan acuan untuk perencanaan promosi musiman. Selain itu, kategori Mobile & Tablets terus mendominasi nilai transaksi tertinggi sepanjang tahun 2022, mencerminkan meningkatnya minat terhadap produk teknologi.

Dari sisi metode pembayaran, COD tetap menjadi pilihan utama pelanggan, menunjukkan pentingnya mempertahankan opsi ini untuk mengoptimalkan tingkat konversi penjualan. Di sisi lain, evaluasi performa produk juga menjadi perhatian, di mana Samsung sebagai produk dengan nilai transaksi tertinggi perlu dijaga performanya, sedangkan Lenovo, yang memiliki nilai transaksi terendah, memerlukan strategi evaluasi untuk meningkatkan daya saing di pasar.

Hasil proyek ini memberikan arahan strategis yang jelas terkait pengelolaan produk, pemasaran, dan metode pembayaran ke depan. Selain itu, temuan ini membuka peluang untuk pengoptimalan lebih lanjut melalui data yang lebih kaya dan analisis lanjutan, meskipun diperlukan data yang lebih banyak dan beragam untuk hasil yang lebih komprehensif.





Thank you!

MySkill Data Analyst – Kelompok B4