

# Final Project

...

Kelompok D8

## Kelompok D8

Ahmad Rido Fairuji (SQL, Data Visualization)

Denny Darmawan (SQL, Data Visualization)

Desnita Rizki Irmawati (Python)

Galih Zaky Tristanaya (Python)

Jovan Aristito Arga (Python)

Modesta Berliansa Termatu Arsanta (SQL, Data Visualization)

Nadia Azahro Choirunisa (SQL, Python)

Rahmana Zahara Budi Putra (Python)

Rizky Huthama Arsyad (SQL, Data Visualization)

# SQL

**DATASET**

# Dataset

Dataset (set data) adalah kumpulan data terstruktur yang berisi informasi dari berbagai sumber, serta dapat dianalisis untuk memperoleh informasi baru.

Dataset yang digunakan pada project ini terdiri dari beberapa tabel sebagai berikut:

- order\_detail
- sku\_detail
- customer\_detail
- payment\_detail

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

# Dataset

## order\_detail

1. id → angka unik dari order / id\_order
2. customer\_id → angka unik dari pelanggan
3. order\_date → tanggal saat dilakukan transaksi
4. sku\_id → angka unik dari produk (sku adalah stock keeping unit)
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 ( $\text{price} * \text{qty\_ordered}$ )
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

# Dataset

## sku\_detail

1. id → angka unik dari produk (dapat digunakan untuk key saat join)
2. sku\_name → nama dari produk
3. base\_price → harga barang yang tertera pada tagging harga / price
4. cogs → cost of goods sold / total biaya untuk menjual 1 produk
5. category → kategori produk

## customer\_detail

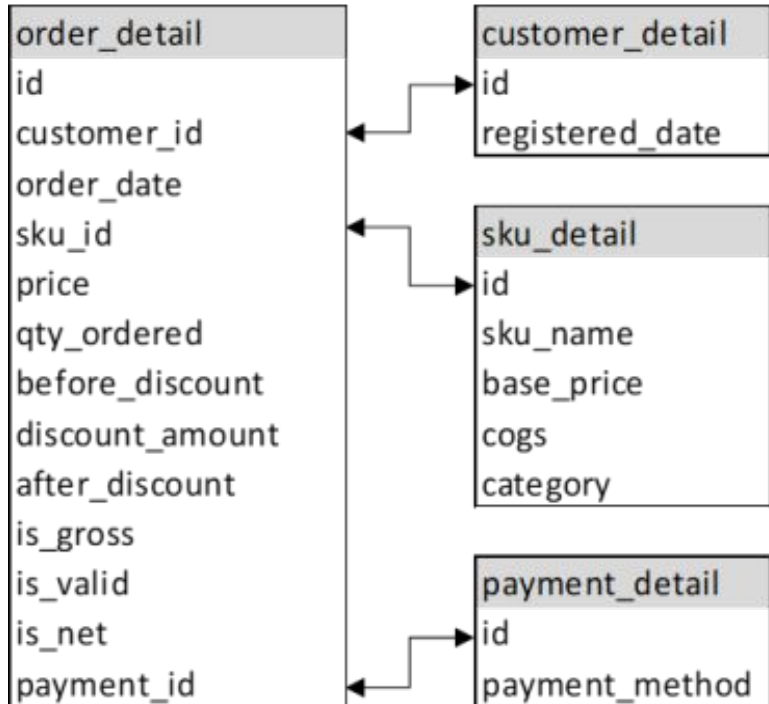
1. id → angka unik dari pelanggan
2. registered\_date → tanggal pelanggan mulai mendaftarkan diri sebagai anggota

## payment\_detail

1. id → angka unik dari metode pembayaran
2. payment\_method → metode pembayaran yang digunakan

# Dataset

Skema Tabel



Dari skema tabel di samping dapat dilihat bahwa tabel `order_detail` merupakan tabel utama pada dataset, yang terhubung dengan tabel lain yaitu tabel `customer_detail`, `sku_detail`, dan `payment_detail`.



# **PERTANYAAN DAN PEMBAHASAN**

# Pertanyaan

## No. 1

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**

# Jawaban

## Query

```
3 SELECT
4     EXTRACT(MONTH FROM order_date) AS MONTH,
5     SUM(after_discount) AS total_transaction_value
6 FROM
7     order_detail
8 WHERE
9     EXTRACT(YEAR FROM order_date) = 2021
10    AND is_valid = 1
11 GROUP BY
12     MONTH
13 ORDER BY
14     total_transaction_value DESC
15 LIMIT 1;
```

EXTRACT(MONTH FROM order\_date) AS MONTH: Ini akan mengekstrak bulan dari kolom order\_date. Misalnya, jika order\_date adalah 2021-08-15, maka hasilnya adalah bulan 8 (Agustus).

SUM(after\_discount) AS total\_transaction\_value: Ini akan menjumlahkan semua nilai transaksi setelah diskon (after\_discount) dalam satu bulan.

FROM order\_detail: Ini adalah tabel sumber data yang digunakan untuk mengambil transaksi.

EXTRACT(YEAR FROM order\_date) = 2021: Kondisi ini memastikan bahwa data yang diambil hanya dari tahun 2021.

AND is\_valid = 1: Ini adalah filter untuk hanya mengambil transaksi yang valid (berdasarkan kolom is\_valid yang bernilai 1).

GROUP BY MONTH: Mengelompokkan hasil berdasarkan bulan. Artinya, semua transaksi dalam bulan yang sama akan digabungkan (dijumlahkan).

ORDER BY total\_transaction\_value DESC: Hasilnya akan diurutkan berdasarkan total nilai transaksi dari yang terbesar ke yang terkecil.

LIMIT 1: Hanya mengembalikan 1 hasil, yaitu bulan dengan total transaksi tertinggi.

# Jawaban

## Hasil

month	total_transaction_value
8	227862744

Dalam hasil query, **bulan 8 (Agustus)** memiliki total nilai transaksi tertinggi sebesar **227,862,744** .

Jadi, bulan **Agustus** adalah bulan dengan total transaksi **terbesar di tahun 2021** .

# Pertanyaan

## No. 2

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**

# Jawaban

## Query

```
1 SELECT
2     sd.category AS kategori,
3     SUM(od.after_discount) AS total_penjualan
4 FROM order_detail od JOIN sku_detail sd ON od.sku_id = sd.id
5 WHERE
6     EXTRACT (YEAR FROM od.order_date) = 2022
7     AND od.is_valid = 1
8 GROUP BY kategori
9 ORDER BY total_penjualan DESC;
```

Select sd.category as kategori :

Memilih kategori produk dari tabel sku\_detail.

sum(od.after\_discount) as total\_penjualan :

Menghitung total nilai after\_discount dari tabel order\_detail.

From order\_detail od join sku\_detail sd on od.sku\_id = sd.id :  
Menggabungkan tabel order\_detail dan sku\_detail berdasarkan sku\_id.

where extract (year from od.order\_date) = 2022 :  
Memfilter transaksi yang terjadi pada tahun 2022.

and od.is\_valid = 1 :  
Memfilter transaksi yang valid.

group by kategori :  
Mengelompokkan data berdasarkan kategori produk.

order by total\_penjualan desc; :  
Mengurutkan hasil berdasarkan total nilai transaksi secara menurun.

# Jawaban

## Hasil

! kategori	total_penjualan
Mobiles & Tablets	918451576
Entertainment	365344148.9
Appliances	316358100
Computing	214028543.4
Men Fashion	135588253
Women Fashion	93014970.62
Home & Living	79483716.2
Health & Sports	54235579.6
Beauty & Grooming	46211019.18
Superstore	32643266.52
Kids & Baby	25931276.84
Others	21744646.02
Soghaat	17658332
School & Education	17362465.3
Books	6792519.2

Dari data yang sudah diperoleh, dapat disimpulkan bahwa kategori **"Mobiles & Tablets"** merupakan kontributor utama terhadap total penjualan pada tahun 2022. Hal ini mengindikasikan tingginya permintaan pasar terhadap produk-produk elektronik, khususnya perangkat mobile.

# Pertanyaan

## No. 3

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**



# Jawaban

## Query

```
74 WITH yearly_transaction AS (  
75     SELECT  
76         sd.category AS category,  
77         SUM(CASE WHEN EXTRACT(YEAR FROM od.order_date) = 2021 THEN after_discount END) AS transaction_2021,  
78         SUM(CASE WHEN EXTRACT(YEAR FROM od.order_date) = 2022 THEN after_discount END) AS transaction_2022  
79     FROM order_detail AS od  
80     LEFT JOIN sku_detail AS sd  
81         ON od.sku_id = sd.id  
82     WHERE is_valid = 1  
83     GROUP BY sd.category  
84 )
```

Common Table  
Expression

```
85  
86 SELECT  
87     category,  
88     transaction_2021,  
89     transaction_2022,  
90     CASE  
91         WHEN transaction_2021 < transaction_2022 THEN 'Increased'  
92         WHEN transaction_2021 > transaction_2022 THEN 'Decreased'  
93         ELSE 'No change'  
94     END AS Change  
95 FROM yearly_transaction  
96 ORDER BY Change DESC
```

Query utama

## Common Table Expression (CTE) yearly\_transaction

Menyimpan tabel sementara yang berisi total transaksi setiap kategori pada tahun 2021 dan 2022.

SELECT sd.category → mengambil kategori dari tabel sku\_detail.

SUM(CASE WHEN EXTRACT(YEAR from od.order\_date) = 2021 THEN after\_discount END) as transaction\_2021, SUM(CASE WHEN EXTRACT(YEAR from od.order\_date) = 2022 THEN after\_discount END) as transaction\_2022 → menghitung total nilai transaksi dari kolom after\_discount pada tahun 2021 dan 2022.

JOIN sku\_detail sd ON od.sku\_id = sd.id → menggabung tabel sku\_detail dan order\_detail berdasarkan sku\_id.

EXTRACT(YEAR FROM od.order\_date) = 2021 & EXTRACT(YEAR FROM od.order\_date) = 2022 → mengambil data pada tahun 2021 dan 2022 saja.

is\_valid = 1 → mengambil data yang valid saja (pelanggan sudah melakukan pembayaran).

GROUP BY sd.category → mengelompokkan data berdasarkan kategori produk

## Query utama

Membandingkan transaksi dari masing-masing kategori pada tahun 2017

SELECT y21.category → mengambil kategori dari CTE Year\_2021

transaction\_2021, transaction\_2022 → mengambil total nilai transaksi pada tahun 2021 dan 2022.

CASE WHEN transaction\_2021 < transaction\_2022 THEN 'Increased' WHEN transaction\_2021 > transaction\_2022 THEN 'Decreased' ELSE 'No Change' END as change → menentukan perubahan nilai transaksi dari masing-masing kategori dengan kondisi:

- Jika total transaksi di tahun 2021 lebih rendah dari total transaksi di tahun 2022 maka dihasilkan nilai 'Increased' (peningkatan).
- Jika total transaksi di tahun 2021 lebih tinggi dari total transaksi di tahun 2022 maka dihasilkan nilai 'Decreased' (penurunan).
- Selain kondisi tersebut maka dihasilkan nilai 'No change' (tidak ada perubahan)

ORDER BY change → mengurutkan data berdasarkan tabel change (menampilkan data yang 'Increased' terlebih dahulu).

# Jawaban

Hasil			
category	transaction_2021	transaction_2022	change
Women Fashion	84045961.4	93014970.62	Increased
Beauty & Grooming	46047360	46211019.18	Increased
Superstore	28828088	32643266.52	Increased
Appliances	218550177	316358100	Increased
Computing	172878860	214028543.4	Increased
Entertainment	162326357.4	365344148.9	Increased
Health & Sports	33837965.6	54235579.6	Increased
Home & Living	45797873	79483716.2	Increased
Kids & Baby	23971057.8	25931276.84	Increased
Men Fashion	58628198	135588253	Increased
Mobiles & Tablets	370606718	918451576	Increased
School & Education	11558982.4	17362465.3	Increased
Soghaat	15056202.6	17658332	Increased
Books	10124596	6792519.2	Decreased
Others	40468515.74	21744646.02	Decreased

Secara keseluruhan, nilai transaksi mengalami **kenaikan** dari tahun 2021 ke tahun 2022.

Beberapa kategori seperti **Men Fashion** dan **Mobiles & Tablets** mengalami peningkatan yang sangat pesat.

Namun pada kategori **Books** dan **Others** mengalami **penurunan** dari tahun sebelumnya, sehingga perlu dilakukan analisis lebih lanjut.

# Pertanyaan

## No. 4

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**

# Jawaban

## Query

```
70 SELECT
71     pd.payment_method,
72     COUNT (DISTINCT od.id) total_payment
73 FROM order_detail od LEFT JOIN payment_detail pd
74     ON od.payment_id = pd.id
75 WHERE
76     (order_date BETWEEN '2022-01-01' AND '2022-12-31')
77     AND is_valid = 1
78 GROUP BY 1
79 ORDER BY 2 DESC
80 LIMIT 5
```

select

pd.payment\_method,  
count (distinct od.id) total\_payment  
from order\_detail od left join payment\_detail pd  
on od.payment\_id = pd.id

Menggunakan left join antara table order\_detail dengan payment\_detail lalu membuat key dengan menggunakan payment\_id dengan id. Setelah itu kita ambil kolom yang dibutuhkan seperti payment\_method dan id (berdasarkan total unik order dan dilakukan fungsi agregat Count).

where (order\_date between '2022-01-01' and '2022-12-31') and is\_valid = 1

Melakukan filter dengan kolom order\_date dan is\_valid, yang menghasilkan order\_date selama tahun 2022 saja dan kolom is\_valid bernilai 1

group by 1

Pengelompokan data berdasarkan payment\_method

order by 2 desc

Mengurutkan data berdasarkan total\_payment secara descending

limit 5

Membatasi jumlah baris yang di panggil dalam query

# Jawaban

Hasil	
! payment_method	total_payment
cod	1809
Payaxis	181
customercredit	75
Easypay	69
jazzwallet	26

Berikut top 5 metode pembayaran yang paling populer digunakan selama 2022.

Dan **metode yang paling populer** atau posisi **puncak** yaitu menggunakan **metode COD** dengan angka **1809** , yang dimana terjadi perbedaan signifikan jauh dengan posisi kedua yaitu menggunakan **metode Payaxis** dengan angka **181** .

Hal ini membuktikan bahwa **metode pembayaran COD** paling banyak diminati oleh konsumen.

# Pertanyaan

## No. 5

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

## Query

```
1 SELECT *
2 FROM
3 (SELECT CASE
4     WHEN LOWER (sd.sku_name) LIKE '%samsung%' THEN 'Samsung'
5     WHEN LOWER (sd.sku_name) LIKE ANY (ARRAY['%apple%', '%macbook%', '%iphone%', '%mac%']) THEN 'Apple'
6     WHEN LOWER (sd.sku_name) LIKE '%sony%' THEN 'Sony'
7     WHEN LOWER (sd.sku_name) LIKE '%huawei%' THEN 'Huawei'
8     WHEN LOWER (sd.sku_name) LIKE '%lenovo%' THEN 'Lenovo'
9     ELSE 'Others'
10  END AS brand,
11   SUM (after_discount) AS total_trx
12 FROM order_detail od
13 LEFT JOIN sku_detail sd
14 ON od.sku_id = sd.id
15 WHERE is_valid = 1
16 GROUP BY 1
17 ORDER BY 2 DESC
18 ) AS proc
19 WHERE brand <> 'Others'
20 ORDER BY 2 DESC
```

Step pertama yang dilakukan yaitu membuat kolom baru dengan menggunakan CASE WHEN untuk pembagian kategori Brand dan menggunakan LIKE untuk mengambil data yang diinginkan. Serta diakhir CASE WHEN harus disertai END beserta nama aliasnya.

**SUM(after\_discount) as total\_trx :**

Menjumlahkan semua data di kolom after\_discount

**from order\_detail od**

**left join sku\_detail sd**

**on od.sku\_id = sd.id**

Mengambil data dari order\_detail dan melakukan left join dari sku\_detail untuk mengambil sku\_name ditabel sku\_detail dengan kata kunci sku\_id

**select \* from (.....) as proc**

**where brand <> 'Others'**

Untuk mengeluarkan nilai 'Others' maka dilakukan Subquery.

**order by 2 desc**

Melakukan pengurutan nilai dengan cara descending pada kolom total\_trx

# Jawaban

Hasil	
! brand	total_trx
Samsung	588764148
Apple	448464932
Sony	63960718
Huawei	63160260
Lenovo	62379800.4

Berikut hasil peringkat produk berdasarkan nilai transaksinya. Yang dimana **Brand Samsung** memiliki transaksi paling banyak yaitu **588,764,148** kali bertransaksi.

**Thank you**