Data Cleaning dan Query Konsumen Pembelian Tertinggi

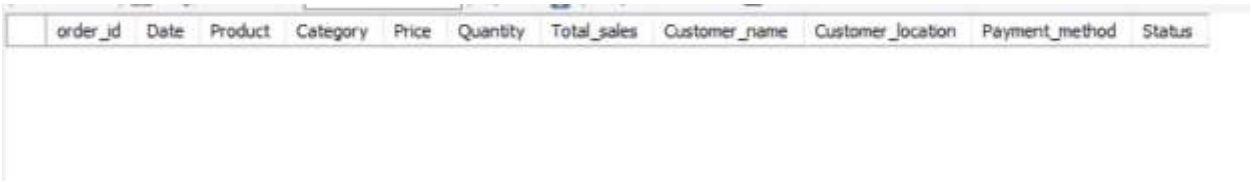
-- Mengganti Nama kolom agar mudah di query

```
ALTER TABLE amazon_sales.amazon_sales
RENAME column `Total Sales` TO Total_sales;

ALTER TABLE amazon_sales.amazon_sales
RENAME column `Customer Name` TO Customer_name;

ALTER TABLE amazon_sales.amazon_sales
RENAME column `Payment Method` TO Payment_method;

ALTER TABLE amazon_sales.amazon_sales
RENAME column `Customer Location` TO Customer_location;
```

-- Memastikan apakah ada kolom yang null

```
select * from amazon_sales.amazon_sales
where
order_id is null
or Date is null
or Product is null
or Category is null
or Price is null
or Quantity is null
or Total_sales is null
or Customer_name is null
or Customer_location is null
or Payment_method is null;
```

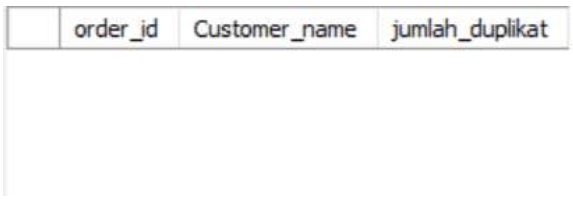
--Output:



order_id	Date	Product	Category	Price	Quantity	Total_sales	Customer_name	Customer_location	Payment_method	Status

-- memeriksa apakah ada kolom yang seharusnya unique terduplikasi/tidak

```
SELECT
   order_id,
   Customer_name,
   COUNT(*) AS jumlah_duplikat
FROM
   amazon_sales.amazon_sales
GROUP BY
   order_id,
    Customer_name
    -- Sertakan semua kolom yang ingin Anda periksa duplikasinya
HAVING
   COUNT(*) > 1;
```



order_id	Customer_name	jumlah_duplikat

-- data already clean

```
with revenue_per_customer as(
select
Customer_name,
SUM(Total_sales) as total_sales
from amazon_sales.amazon_sales
group by Customer_name
order by total_sales desc
)
```

-- memfilter konsumen tajir dengan min spend >=10000

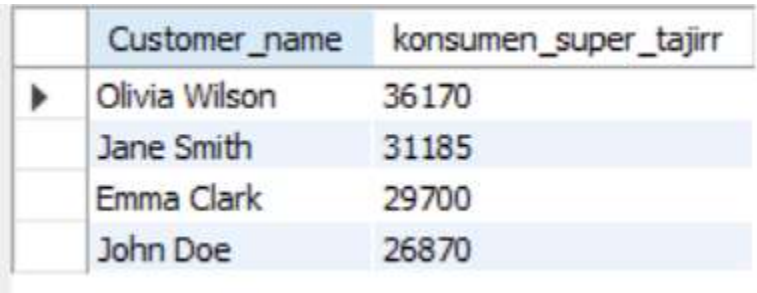
```
,konsumen_tajir as (
select
Customer_name,
SUM(total_sales) as total_sales_tajir
from revenue_per_customer
where total_sales >=10000
group by Customer_name
order by total_sales asc
)
```

-- memfilter customer super tajir yang spend nya di atas rata -rata

```
,avg_super_tajir as(
select
AVG(total_sales_tajir) as avg_sales
from konsumen_tajir

)
```

```
,konsumen_super_tajir as(

select
Customer_name,
total_sales_tajir
from konsumen_tajir
where total_sales_tajir > (select avg_sales from avg_super_tajir)
order by total_sales_tajir desc
)

select * from konsumen_super_tajir
```

Output:



	Customer_name	konsumen_super_tajirr
▶	Olivia Wilson	36170
	Jane Smith	31185
	Emma Clark	29700
	John Doe	26870

Tabel tersebut menampilkan empat pelanggan yang dianggap sebagai "konsumen super tajir" karena nilai pengeluaran mereka semua di atas 10.000, dan total spend nya lebih dari rata-rata konsumen_tajir