

# Soal & Template Jawaban

Task 5

Nama : Zulhaji

# Petunjuk

**Silahkan merujuk pada Data Source Task 5 yang telah disediakan untuk mengerjakan soal soal di bawah ini**

Pada bagian data analytics, terdiri dari 4 soal dengan use case & tabel yang sama. Bayangkan kamu memiliki database erp yang terdiri dari 3 tabel: penjualan, pelanggan, barang. Tabel tersebut akan dibuat menjadi sebuah datamart yang nantinya digunakan untuk visualisasi.

# Query

## Soal 1\*:

Dari 2 query ini, mana yang bekerja lebih baik? Jelaskan mengapa.

(a) `SELECT * FROM pelanggan WHERE SUBSTR(alamat, 1, 3) = Mat;`

(b) `SELECT * FROM pelanggan WHERE alamat LIKE 'Mat%'`

*\*disclaimer: soal ini tidak terkait dengan data source*

Jawaban : b

Alasan : logika yang digunakan lebih efisien dalam mendapatkan data yang diinginkan dibandingkan dengan pilihan (a) karena untuk pilihan (a) akan memiliki 2 tahapan yakni mencari 3 huruf dari kolom tertentu kemudian tahapan selanjutnya mencari huruf yang berisi Mat

# Query

## Soal 2 \*:

Anggap kita memiliki tabel pelanggan dengan kolom: id, nama, tanggal\_lahir, alamat. Bagaimana cara yang lebih tepat dalam menulis query untuk mendapatkan data pelanggan yang tanggal\_lahir nya ada di antara 2000-01-01 sampai 2008-12-31? Pilihlah salah satu jawaban dan berikan alasannya.

- (a) `SELECT * FROM pelanggan WHERE tanggal_lahir >= '2000-01-01' AND tanggal_lahir <= '2008-12-31'`
- (b) `SELECT * FROM pelanggan WHERE tanggal_lahir BETWEEN '2000-01-01' AND '2008-12-31'`

*\*disclaimer: soal ini tidak terkait dengan data source*

Jawaban : (b)

Alasan : logika yang digunakan lebih efisien dalam mendapatkan data yang diinginkan dibandingkan dengan pilihan (a)

## Soal 3: Menentukan Primary Key

### A. Tugas

Tentukan primary key dari table penjualan. jelaskan alasannya

### B. Jawaban & Penjelasan : id customer & karena id customer merupakan value yang berisi nilai unik, tidak duplikat dan tidak memiliki nilai null, selain itu id customer berfungsi sebagai identitas untuk membedakan setiap record yang ada di dalam tabel

## Soal 4: Design Datamart

### A. Tugas

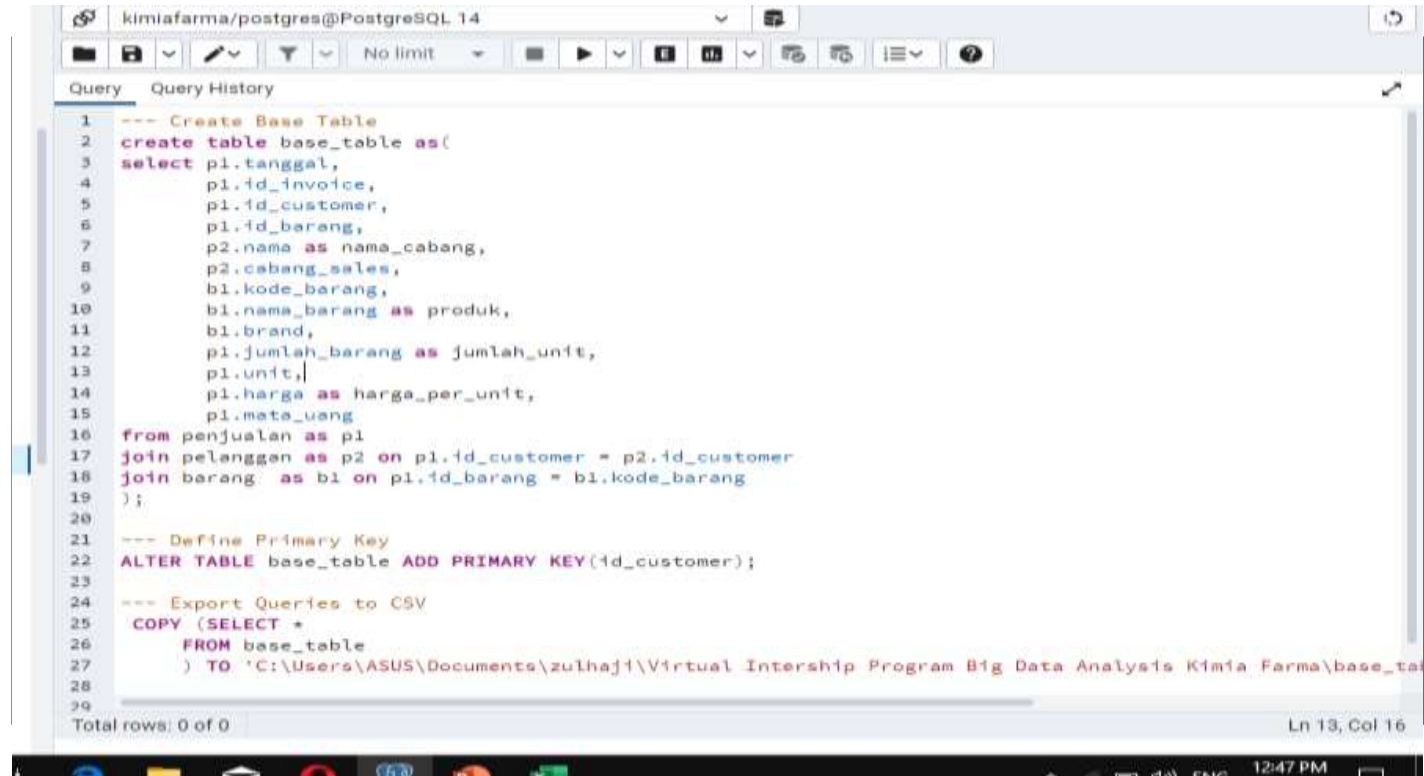
Buatlah design datamart (Terdiri dari tabel base, dan tabel aggregate). Upload file query dalam gdrive mu (pastikan dapat diakses public). Lalu masukkan linknya di tabel di bawah, dan cantumkan juga screenshoot query nya (jika lebih dari 1 file, maka masing masing file di-screenshoot)

Silahkan tambah halaman jika dibutuhkan

### A. Jawaban : .....

No	Nama File	Link
1	base_table	<a href="https://drive.google.com/file/d/1xYpHmcgE5cG48t8otqfcbxQJkMm0p-SB/view?usp=drive_open">https://drive.google.com/file/d/1xYpHmcgE5cG48t8otqfcbxQJkMm0p-SB/view?usp=drive_open</a>
2	anggregate_table	<a href="https://drive.google.com/file/d/1-nS2PVu3S11hz5ojdFk0TMjTF46bDjbS/view">https://drive.google.com/file/d/1-nS2PVu3S11hz5ojdFk0TMjTF46bDjbS/view</a>

# Table Base "<<table\_base>>"



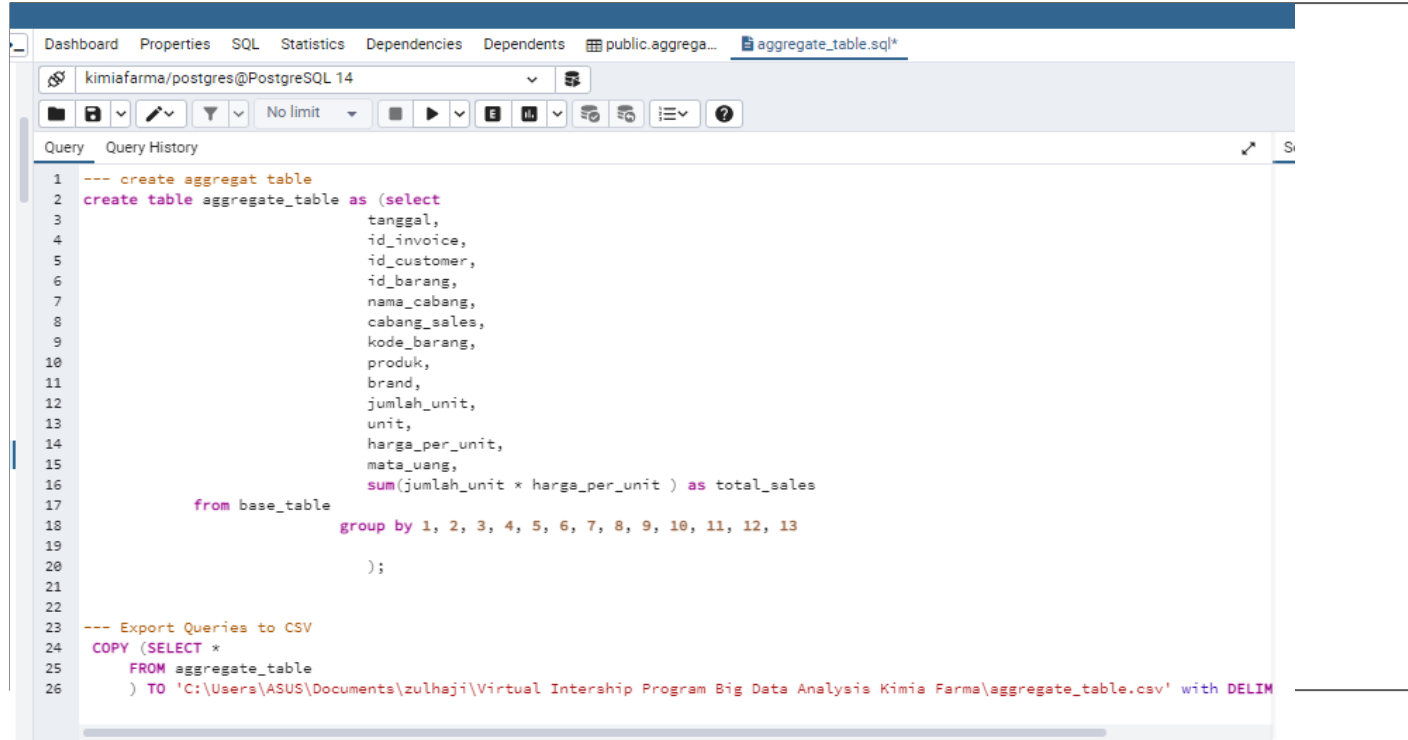
```
kimiafarma/postgres@PostgreSQL 14
Query History
1  --- Create Base Table
2  create table base_table as(
3  select
4      p1.tanggal,
5      p1.id_invoice,
6      p1.id_customer,
7      p1.id_barang,
8      p2.nama as nama_cabang,
9      p2.cabang_sales,
10     b1.kode_barang,
11     b1.nama_barang as produk,
12     b1.brand,
13     p1.jumlah_barang as jumlah_unit,
14     p1.unit,
15     p1.harga as harga_per_unit,
16     p1.mata_uang
17 from penjualan as p1
18 join pelanggan as p2 on p1.id_customer = p2.id_customer
19 join barang as b1 on p1.id_barang = b1.kode_barang
20 );
21
22 --- Define Primary Key
23 ALTER TABLE base_table ADD PRIMARY KEY(id_customer);
24
25 --- Export Queries to CSV
26 COPY (SELECT *
27      FROM base_table
28      ) TO 'C:\Users\ASUS\Documents\zulhajj\Virtual Internship Program Big Data Analysis Kimia Farma\base_table.csv'
29
30 Total rows: 0 of 0
Ln 13, Col 16
```

## Table Base “<<table\_base>>”

column	data type	description	transformation
tanggal	date		
id_invoice	varchar		
id_customer	varchar		
id_barang	varchar		
nama_cabang	varchar	perubahan nama	dari nama
cabang_sales	varchar		
kode_barang	varchar		
produk	varchar	perubahan nama	dari nama_barang
brand	varchar		
jumlah_unit	varchar		dari jumlah_barang
unit	varchar		
harga_per_unit	numeric		
mata_uang	varchar		



# Table Aggregate “<<aggregate\_table>>”



The screenshot shows a PostgreSQL query editor interface. The top menu bar includes Dashboard, Properties, SQL, Statistics, Dependencies, and Dependents. The current database is 'kimiafarma/postgres@PostgreSQL 14'. The query editor displays the following SQL code:

```
1 --- create aggregat table
2 create table aggregate_table as (select
3     tanggal,
4     id_invoice,
5     id_customer,
6     id_barang,
7     nama_cabang,
8     cabang_sales,
9     kode_barang,
10    produk,
11    brand,
12    jumlah_unit,
13    unit,
14    harga_per_unit,
15    mata_uang,
16    sum(jumlah_unit * harga_per_unit ) as total_sales
17 from base_table
18 group by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
19
20 );
21
22
23 --- Export Queries to CSV
24 COPY (SELECT *
25 FROM aggregate_table
26 ) TO 'C:\Users\ASUS\Documents\zulhaji\Virtual Intership Program Big Data Analysis Kimia Farma\aggregate_table.csv' with DELIM
```

# Table Aggregate “<<aggregate\_table>>”

column	data type	description	transformation
tanggal	date		
id_invoice	varchar		
id_customer	varchar		
id_barang	varchar		
nama_cabang	varchar		
cabang_sales	varchar		
kode_barang	varchar		
produk	varchar		
brand	varchar		
jumlah_unit	numeric	change datatype	from varchar to numeric
unit	varchar		
harga_per_unit	numeric		
mata_uang	varchar		
total_sales	numeric		

# Soal 5 : Data Visualization

## A. Tugas

buatlah data visualiasasi nya, dan cantumkan linknya di bawah (pastikan bisa diakses publik).  
Lalu cantumkan juga screenshot visualisasinya

Silahkan tambah halaman jika dibutuhkan

## A. Jawaban :

Link visualisasi (ex link Google Data Studio) :

<https://lookerstudio.google.com/u/0/reporting/190b44ce-9478-47de-bf96-b67cb4adf274/page/GKjDD/edit>

# Sales Report of PT Kimia Farma



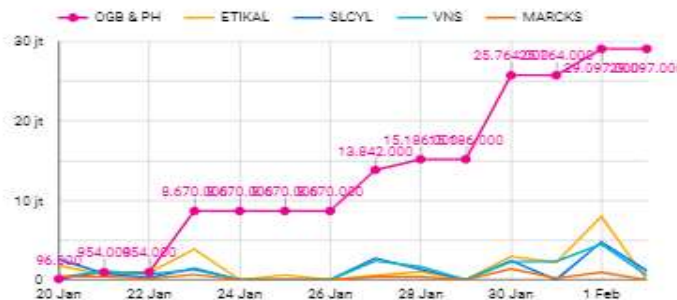
Pilih rentang -

produk -

jumlah... -

brand -

## Tren Total Penjualan



produk  
10,0

brand  
5

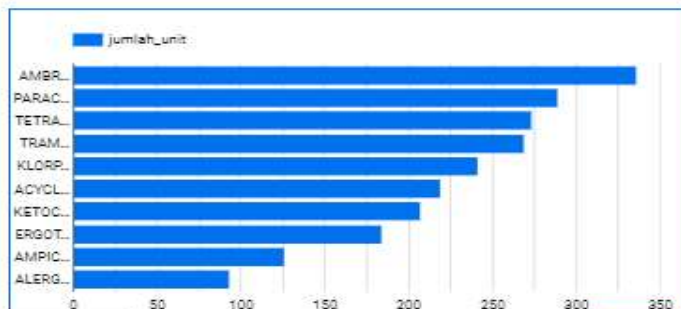
jumlah\_unit  
2,2 rb

total\_sales  
90.637.700

## Lokasi Cabang



## Jumlah Produk Terjual



produk	br...	jum...	harga...	brand	total_s...
1. ACYCLO...	OG...	98	96000	OGB & PH	9.408.000
2. AMBROX...	VNS	67	31000	VNS	6.221.000
3. ALERGIN...	ET...	25	112000	ETIKAL	5.600.000
4. ACYCLO...	OG...	45	96000	OGB & PH	4.320.000
5. ACYCLO...	OG...	40	96000	OGB & PH	3.840.000
6. ERGOT...	ET...	10	64700	ETIKAL	2.588.000
7. KLORPR...	SL...	24	47000	SLCYL	2.256.000
8. KETOCO...	OG...	25	39000	OGB & PH	1.950.000

## Soal 6 : Additional Complementary Data

A. Tugas :

Dari data yang tersedia, menurut kamu untuk melengkapi analisis nya apakah diperlukan data lain juga? jika iya, sebutkan data apa yang kamu maksud dan mengapa memerlukan data tersebut

A. Jawaban : manfaat atau kegunaan dari jenis-jenis obat yang mungkin perlu ditampilkan datanya agar bisa mengetahui obat yang punya frekuensi yang tinggi di setiap kota/provinsi agar dapat mengetahui kasus/penyakit yang di kota-kota tersebut