

Kimia Farma Big Data Analytics Virtual Internship Program Rakamin Academy

Lingga Rizki Mandala

A short horizontal bar with a teal left half and an orange right half.

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 : Mencari huruf, angka, atau kata tertentu pada database dengan menggunakan query WHERE dan diikuti operator LIKE. Ketika ingin mencari value pada kolom alamat sesuai jawaban, selalu menggunakan kutip ' ' dan diawali dengan mencari nilai yang berhubungan dengan "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 : untuk mencari slicing suatu value dengan rentang tertentu seperti tanggal di atas, metode yang mudah dipahami dan praktis adalah menggunakan BETWEEN ketimbang penggunaan operator matematika

Soal 3: Menentukan Primary Key



- A. Tugas
Tentukan primary key dari table penjualan. jelaskan alasannya
- B. Jawaban & Penjelasan : PRIMARY KEY pada tabel penjualan terdapat pada kolom id_invoice karena mempunyai value unique dan tidak sama ketimbang kolom lainnya. Sedangkan id_customer mewakili tabel pelanggan, dan id_barang mewakili tabel barang.

Soal 4: Design Datamart

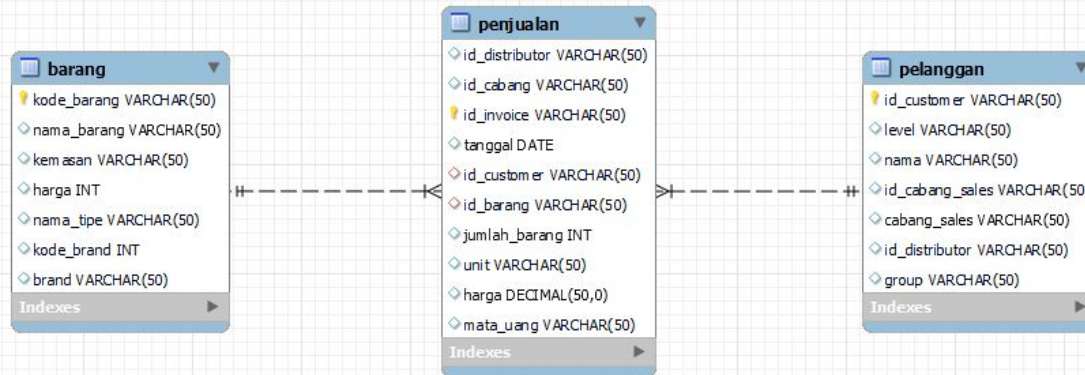


A. Tugas

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)

No	Nama File	Link
1	schema_db_kf.sql	https://drive.google.com/file/d/1EbKo6MMsA1n72PFGKmlQft0qfbglppCl/view?usp=sharing
2	base_tabel_kf.sql	https://drive.google.com/file/d/1ETZKKczF0udi6-pSqolrDS-sKYWMHqrY/view?usp=sharing
3	agg_tabel_kf.sql	https://drive.google.com/file/d/1EYkoGDeB58tsMd_wVt6xkiQd61pHbnsTN/view?usp=sharing

Tabel ERD



Membuat schema DB kimia_farma



```
-- membuat dan menggunakan schema database kimia_farma --
create schema kimia_farma;
use kimia_farma;

-- membuat tabel barang
create table kimia_farma.barang (
    `kode_barang` varchar(50) not null,
    `nama_barang` varchar(50) null,
    `kemasan` varchar(50) null,
    `harga` int null,
    `nama_tipe` varchar(50) null,
    `kode_brand` int null,
    `brand` varchar(50) null,
    primary key(kode_barang)
);

-- membuat tabel penjualan
create table kimia_farma.penjualan (
    `id_distributor` VARCHAR(50) null,
    `id_cabang` varchar(50) null,
    `id_invoice` varchar(50) not null,
    `tanggal` date null,
    `id_customer` varchar(50) null,
    `id_barang` varchar(50) null,
    `jumlah_barang` int null,
    `unit` varchar(50) null,
    `harga` float null,
    `mata_uang` varchar(50) null,
    primary key(id_invoice)
);

-- mengubah kolom harga menjadi decimal --
alter table kimia_farma.penjualan modify column harga decimal(50);
```


Membuat tabel base_tabel_kf



```
-- membuat datamart tabel penjualan penggabungan dataset barang, pelanggan, dan penjualan --
create table kimia_farma.base_tabel_kf(
select
concat(pnj.id_invoice, ' ',pnj.id_barang) as id_penjualan,
pnj.id_invoice,
pnj.tanggal,
pnj.id_customer,
plg.nama,
plg.level,
plg.id_cabang_sales,
plg.cabang_sales,
plg.id_distributor,
plg.group as grup,
pnj.id_barang,
brg.nama_barang,
brg.kemasan,
brg.nama_tipe,
brg.kode_brand,
brg.brand,
pnj.jumlah_barang,
pnj.harga,
pnj.mata_uang
from penjualan pnj
left join barang brg
on brg.kode_barang = pnj.id_barang
right join pelanggan plg
on plg.id_customer = pnj.id_customer
order by pnj.tanggal
);

-- menghapus null value pada row pertama --
delete from kimia_farma.base_tabel_kf
where id_penjualan is null and id_invoice is null and tanggal is null and id_customer is null and id_barang is null
and jumlah_barang is null and mata_uang is null and harga is null and nama_barang is null
and kemasan is null and nama_tipe is null and kode_brand is null and brand is null and `level`=''
and nama='' and id_cabang_sales='' and cabang_sales='' and id_distributor='' and grup='';

alter table kimia_farma.base_tabel_kf modify column harga decimal(50);

-- membuat id_penjualan menjadi primary key pada datamart tbl_penjualan --
alter table kimia_farma.base_tabel_kf add primary key(id_penjualan);
```

Membuat tabel agg_tabel_kf



```
-- membuat datamart tabel aggregate penjualan
create table kimia_farma.agg_tabel_kf (
select
id_penjualan,
id_invoice,
tanggal,
monthname(tanggal) as bulan,
id_customer,
nama,
cabang_sales,
grup,
id_barang,
nama_barang,
brand,
kemasan,
jumlah_barang,
harga as harga_satuan,
sum(harga * jumlah_barang) as total
from base_tabel_kf
group by id_penjualan
order by 3, 4, 5, 6, 8, 9, 10, 11, 12
);

select * from kimia_farma.agg_tabel_kf;
```

Table Base “base_tabel_kf”

column	data type	description	transformation
id_penjualan	varchar (PK)	id penjualan	gabungan id_invoice & id_barang
id_invoice	varchar	id invoice	
tanggal	date	tanggal transaksi	
id_customer	varchar	id customer	
nama	varchar	nama customer	
level	varchar	kategori perusahaan	
id_cabang_sales	varchar	id cabang customer	
cabang_sales	varchar	lokasi customer	
id_distributor	varchar	id distributor	
grup	varchar	kategori customer	
id_barang	varchar	id barang	
nama_barang	varchar	nama barang	
kemasan	varchar	jenis kemasan barang	
nama_tipe	varchar	jenis produk barang	
kode_brand	int	kode merek barang	
brand	varchar	merek barang	
jumlah_barang	int	jumlah barang	
harga	decimal	harga barang per satuan	
mata_uang	varchar	jenis mata uang	

Table Aggregate “agg_tabel_kf”

column	data type	description	transformation
id_penjualan	varchar	id penjualan	gabungan id_invoice & id_barang
id_invoice	varchar	id invoice	
tanggal	date	tanggal transaksi	
bulan	varchar	bulan transaksi	diambil dari tanggal transaksi
id_customer	varchar	id customer	
nama	varchar	nama customer	
cabang_sales	varchar	lokasi customer	
grup	varchar	kategori customer	
id_barang	varchar	id barang	
nama_barang	varchar	nama barang	
brand	varchar	merek barang	
kemasan	varchar	jenis kemasan barang	
jumlah_barang	int	jumlah barang	
harga_satuan	decimal	harga barang per satuan	
total	decimal	harga total	perkalian dari jumlah_barang dan harga_satuan

Soal 5 : Data Visualization

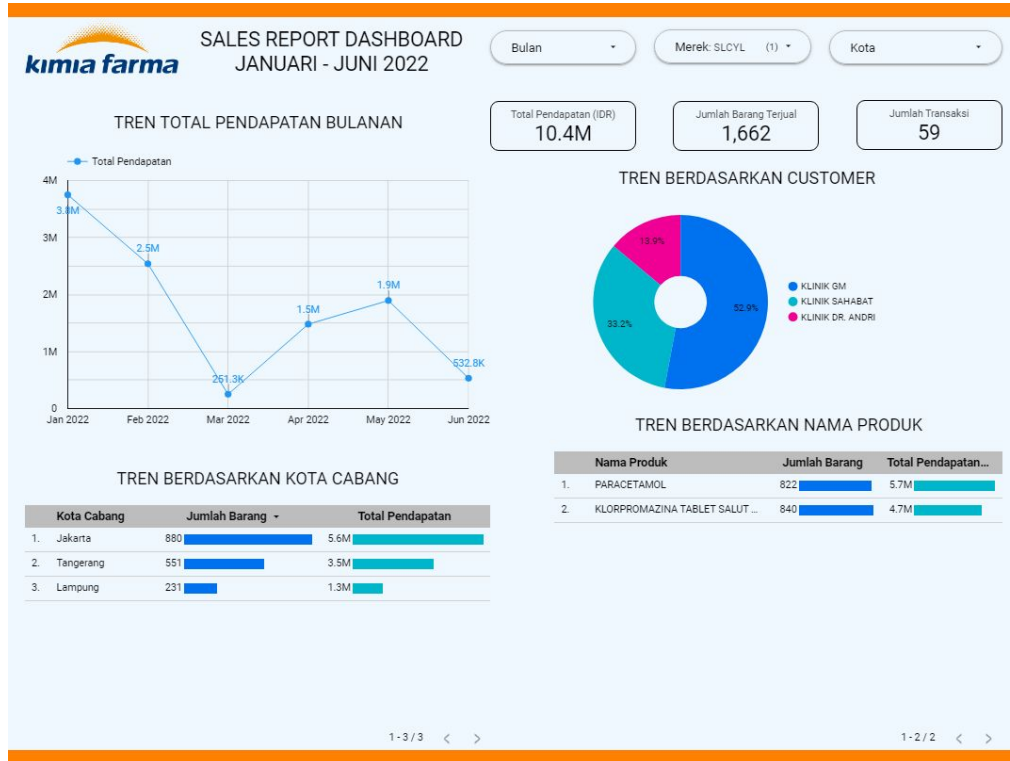


- A. Tugas
buatlah data visualiasasi nya, dan cantumkan linknya di bawah (pastikan bisa diakses publik). Lalu cantumkan juga screenshot visualisasinya

Link visualisasi : <https://intip.in/KFSalesDashboard>



SALICYL SALES REPORT DASHBOARD



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

Jawaban :

- informasi tambahan setiap dataset yang diberikan
- informasi tambahan mengenai distributor, cabang, dan lini brand karena tidak dijelaskan secara spesifik



Terima kasih