



## Salicyl Sales Analysis

Kimia Farma Big Data Analyst Project Based

Presented by

Arya Vito Fauzal

### Petunjuk

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 data mart 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 : Karena jawaban A seharusnya SUBSTR(alamat, 1,3) = "Ala" apabila ingin mencari alamat dengan awalan ala atau apabila ingin mat maka diganti SUBSTR(alamat, 3,6). Pilihan B tepat karena 'Mat%' akan mencari seluruh alamat yang berawalan dengan kata Mat dan ditambah apapun kata, simbolik, atau numerik yang ada setelahnya seperti Matoga 5, Mataram, Mato Kopi 3A, dsb

### 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: Diantara, berarti 2000-01-01 dan 2008-12-31 tidak termasuk kriteria sehingga opsi A menjadi kurang tepat,pilihan B menjadi tepat karena menggunakan kata between yang menunjukkan rentang waktunya 2000-01-01 dan 2008-12-31 terekslusi

## Soal 3: Menentukan Primary Key

#### **Tugas**

Tentukan primary key dari table penjualan. jelaskan alasannya

#### Jawaban & Penjelasan:

id\_customer karena pada table 'penjualan' ke 3 field ini terhubung dengan table lain seperti 'pelanggan' dan 'barang' dan mempunyai unique identifier dimana tidak mungkin ada kesamaan terhadap ID yang ada sehingga tidak memungkinkan adanya duplikasi record apabila disambungkan antar table dengan menggunakan inner join, outer join, join lainnya.



## Soal 4: Design Datamart

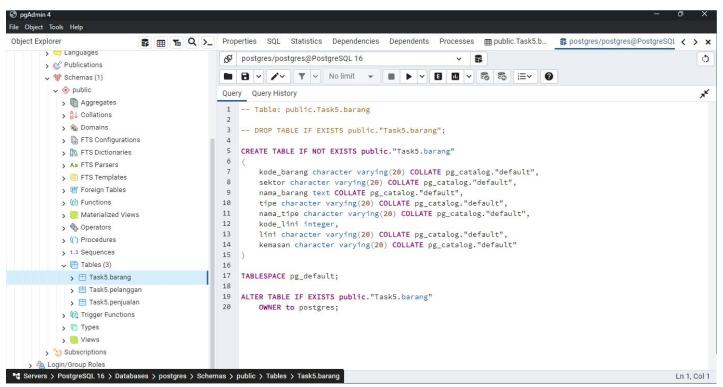
#### A. Tugas

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

#### B. Jawaban: Terlampir

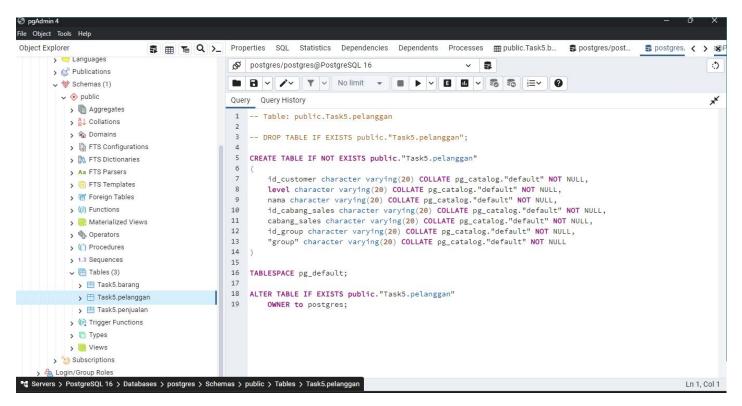
No	Nama File	Link
1	Task 5 Rakamin	Dataset dan Mekanisme ETL Data

## Table Base "<<Task5.barang>>"



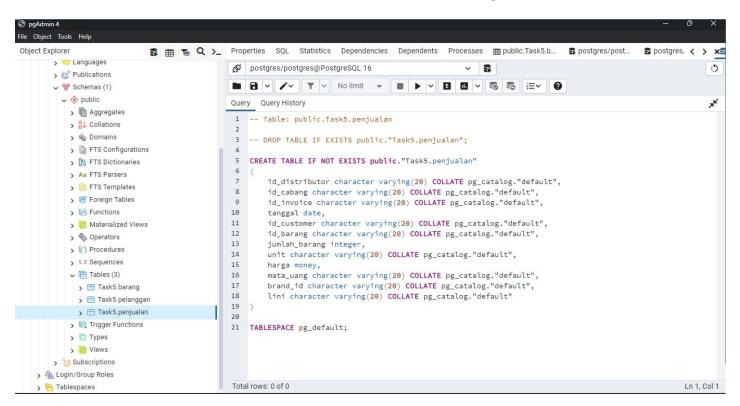
## Table Base "<<Task5.barang>>"

column	data type	description	transformation
kode_barang	varchar	kode dari obat	-
sektor	varchar	sektor terhadap obat tersebut	-
nama_barang	text	nama obat	-
tipe	varchar	kode dari tipe obat	-
nama_tipe	varchar	jenis produk obat	-
kode_lini	int	kode lini obat	-
lini	varchar	lini obat-obatan	-



column	data type	description	transformation
id_customer	varchar	ID customer pembeli obat	-
level	varchar	perusahaan atau bukan	-
nama	varchar	nama apotek/klinik pembeli obat	-
id_cabang_sales	varchar	ID cabang apotek/klinik	-
cabang_sales	varchar	cabang apotek/klinik	-
id_group	varchar	ID penentu apotek/klinik	-
group	varchar	apotek/klinik yang memesan obat	-

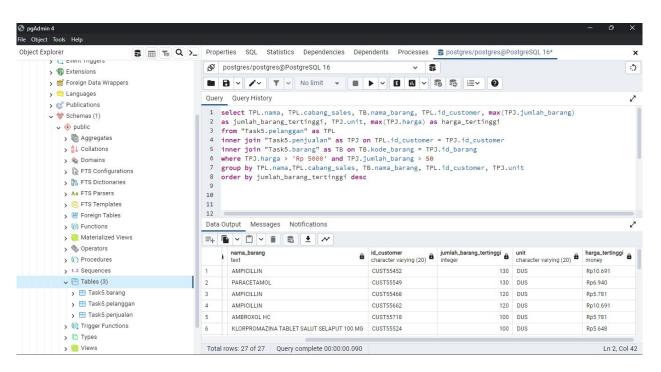
### Table Base "<<Task5.penjualan>>"



column	data type	description	transformation
id_distributor	varchar	ID distributor obat	-
id_cabang	varchar	ID cabang pembeli	-
id_invoice	varchar	ID resi pembelian obat	-
tanggal	date	tanggal pembelian obat	-
id_customer	varchar	ID customer pembeli	-
id_barang	varchar	ID barang yang dibeli	-
jumlah_barang	int	jumlah barang yang dibeli	-

column	data type	description	transformation
unit	varchar	berapa unit yang dibeli (Dus/Kardus)	-
harga	money	Harga obat/unit	-
mata_uang	varchar	mata uang yang digunakan untuk pembelian	_
brand_id	varchar	ID brand yang dibeli	-
lini	varchar	lini dari obat	-

# Table Aggregate "<<Data Analysis Rakamin>>"



# Table Aggregate "<<Data Analysis Rakamin>>"

column	data type	description	transformation
Nama	varchar	nama dari apotek dan klinik	Nama Apotek
cabang_sales	varchar	cabang sales apotek	Cabang Sales
nama_barang	text	nama dari obat-obatan yang dipesan	Nama Obat
id_customer	varchar	ID dari customer yang memesan obat	Customer ID
jumlah_barang_terting gi	int	Jumlah pemesanan obat tertinggi dari tiap cabang apotek	Jumlah Obat (Dus)
unit	varchar	satuan Dus (Kardus) tiap obat	Unit
harga_tertinggi	money	harga per 1 dus tiap pemesanan obat di cabang tertentu	Harga Tertinggi

#### Soal 5 : Data Visualization

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

#### **B.** Jawaban:

Link visualisasi (ex link Google Data Studio): Salicyl Sales 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

B. Jawaban: Menurut saya sudah cukup lengkap bagi saya yang pemula dan masih belajar. Untuk memperlengkap analisa, dapat ditambahkan di beberapa wilayah yang tersebar agar saat menambahkan geo chart tidak terkesan sepi dan untuk di tabel penjualan dapat diberikan keuntungan agar dapat dilihat berapa keuntungan yang didapatkan pada apotek X di cabang X. Dapat ditambahkan juga waktu (seperti tahun atau bulan atau batch) agar dapat memasukkan chart berupa tren yang dapat dianalisa dari berbagai aspek nantinya