

Salicyl Sales Dashboard

**Kimia Farma Big Data Analytics Virtual Internship
Program**

Presented by
Putri Kirey Eki Yogaswari



Putri Kirey Eki Yogaswari

About You

I am a fresh graduate from Public Health with a focus on Health Statistics. During college, I have knowledge of situation analysis, designing health programs, leadership skills, systems thinking and data analysis. I have a career interest in data management and proficient in using SPSS and Ms. Excel. Has managed data up to 2000+.

My Experiences

-  Project-Based Intern: Big Data Analytics
Virtual Internship
PT. KIMIA FARMA, TBK
-  Data Team Lead
Universitas Muhammadiyah Prof. Dr. Hamka Jakarta
-  Data Management Staf
Pusat Data dan Informasi, Kemenkes RI

Case Study

- 1. Pemahaman Query**
- 2. Menentukan Primary Key**
- 3. Design Datamart**
- 4. Data Visualization**

Pemahaman Query

Terdapat 2 query, mana yang bekerja lebih baik?

- (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:

Query yang kedua (b) akan bekerja lebih baik daripada query yang pertama (a).

- Query ini menggunakan operator LIKE untuk mencocokkan string yang dimulai dengan 'Mat'. Dalam kasus ini, % berfungsi sebagai wildcard, yang akan mencocokkan setiap karakter setelah 'Mat', termasuk kosong.
- Penggunaan LIKE dengan wildcard ini lebih tepat untuk mencari string yang dimulai dengan 'Mat'.
- Tidak efisien menggunakan fungsi pada column yang difilter. Jika kolom 'alamat' telah diindeks, penggunaan fungsi dalam filter query akan membuat indeks tersebut tidak efektif.

Pemahaman Query

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*

Pemahaman Query

Jawaban:

Query yang kedua (b) akan bekerja lebih baik daripada query yang pertama (a). Penggunaan pernyataan BETWEEN difungsikan untuk menangani rentang nilai. Penggunaan BETWEEN pada tanggal juga memungkinkan inklusi tanggal awal dan akhir dalam rentang tersebut, yang sesuai dengan kebutuhan kita untuk mendapatkan data dari tanggal 2000-01-01 hingga 2008-12-31. Lebih baik menghindari penggunaan operator logika (AND, OR, dan NOT) saat melakukan filter pada query, terutama jika kolom yang difilter merupakan bagian dari indeks yang ada.

Menentukan Primary Key

Menentukan primary key dari table penjualan.

Karena tidak ada kolom yang unique sehingga perlu mencari primary key nya sendiri dengan menggabungkan id_invoice dan id_barang dengan nama kolom id_penjualan.

id_penjualan	id_invoice	tanggal	id_customer	id_barang	jumlah_barang	unit	harga	mata_uang
IN5997_BRG0001	IN5997	2023-08-01	CUST55380	BRG0001		1	DUS	IDR
IN5997_BRG0002	IN5997	2023-08-01	CUST55380	BRG0002		4	DUS	IDR
IN5997_BRG0003	IN5997	2023-08-01	CUST55380	BRG0003		6	BOTOL	IDR
IN5997_BRG0004	IN5997	2023-08-01	CUST55380	BRG0004		11	TABLET	IDR
IN5997_BRG0005	IN5997	2023-08-01	CUST55380	BRG0005		40	TABLET	IDR
IN6297_BRG0002	IN6297	2023-08-01	CUST55381	BRG0002		5	DUS	IDR
IN6297_BRG0005	IN6297	2023-08-01	CUST55381	BRG0005		15	TABLET	IDR

Query selengkapnya dapat dilihat disini

Design Datamart

Terdiri dari tabel base, dan tabel aggregate

Tabel Base

Merupakan hasil penggabungan dari tabel penjualan, pelanggan dan barang.

tablebase_penjualan

Column	Data Type	Description	Transformation
id_penjualan	string	PK dari tabel penjualan	CONCAT(id_invoice, '_', id_barang)
id_invoice	string	id invoice	-
tanggal	date	tanggal transaksi dilakukan	-

Hasil tabel selengkapnya dapat dilihat disini

tablebase_penjualan

Column	Data Type	Description	Transformation
id_barang	string	id barang, key join ke tabel barang	-
nama_barang	string	nama barang	-
harga	numeric	harga barang	-
unit	string	kemasan untuk setiap penjualan	-
jumlah_barang	numeric	Jumlah barang yang dibeli	-
total_harga_per_barang	numeric	harga barang x jumlah barang	(jumlah_barang * pjl.harga)

Hasil tabel selengkapnya dapat dilihat disini

tablebase_penjualan

Column	Data Type	Description	Transformation
mata_uang	string	mata uang. default = IDR	-
kode_brand	string	kode brand dari barang	-
brand	string	nama brand	-
id_customer	string	id customer, key join ke tabel pelanggan	-
nama_customer	string	nama pelanggan	-
cabang_sales	string	cabang yang supply barang ke pelanggan	(jumlah_barang * pjl.harga)

Hasil tabel selengkapnya dapat dilihat disini

Tabel Aggregate

Merupakan turunan dari tabel base dimana data dikelompokkan berdasarkan tanggal, id_customer dan id_invoice

tableaggregate_penjualan harian

Column	Data Type	Description	Transformation
id_invoice	string	PK dari tabel penjualan harian	-
tanggal	date	tanggal transaksi dilakukan	-
id_customer	string	id pelanggan	-
nama_customer	string	nama pelanggan	-

Hasil tabel selengkapnya dapat dilihat disini

tableaggregate_penjualan harian

Column	Data Type	Description	Transformation
cabang_sales	string	nama cabang	-
id_distributor	numeric	distributor	-
group_category	string	kategori pelanggan, possible value: - Apotek - Klinik	-
total_barang	numeric	total barang yang dibeli untuk setiap invoice id	COUNT(DISTINCT id_barang) total_barang
total_pembelian	numeric	total harga yang dibayar untuk setiap invoice id	SUM(total_harga_per_ barang) total_pembelian

Hasil tabel selengkapnya dapat dilihat disini

Data Visualization

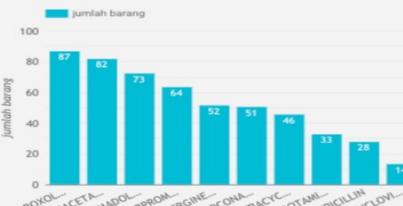
Report Penjualan PT Kimia Farma

Filter Data

Jan 1, 2022 - Jan 31, ... Pelanggan Cabang

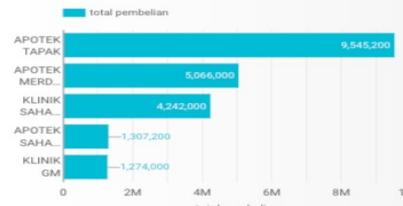
Total Invoice	Total Penjualan	Total Barang Terjual	Jumlah barang	Pelanggan Aktif
30	21,434,400	10	530	5

Jumlah barang terjual



total_barang

Trend Pembelian Pelanggan



total pembelian

APOTEK TAPAK
KLINIK SAHABAT
APOTEK MERDEKA
APOTEK SAHA...
KLINIK GM

10
7
5
4
4

1
1
1
1
1

Nama Barang	Tanggal	ID Invoice	Jumlah Bara...	Nama Customer	Harga	Total Harga	Jumlah Barang
1. AMBROXOL HC	Jan 2, 2022	IN6023	67	KLINIK SAHABAT	31000	2077000	67
2. PARACETAMOL	Jan 2, 2022	IN6063	67	APOTEK TAPAK	21000	1407000	67
3. ALERGINE TAB...	Jan 2, 2022	IN6063	25	APOTEK TAPAK	112000	2800000	25
4. ALERGINE TAB...	Jan 2, 2022	IN6259	25	APOTEK MERDEKA	112000	2800000	25
5. KLOPRIMAZI...	Jan 2, 2022	IN6063	24	APOTEK TAPAK	47000	1128000	24
6. TETRACYCLIN...	Jan 2, 2022	IN6301	24	APOTEK SAHA...	9800	235200	24
7. ERGOTAMINE ...	Jan 2, 2022	IN6063	23	APOTEK TAPAK	64700	1488100	23

1 - 30 / 30 < >

Additional Complementary Data

Data lain yang diperkirakan dapat melengkapi analisis, diantaranya:

1. Data Latitude Longitude untuk setiap lokasi pelanggan

- a. Berguna untuk mengamati sebaran geografis pelanggan
- b. mengidentifikasi lokasi potensial yang dapat ditargetkan untuk peningkatan pangsa pasar
- c. memungkinkan kita untuk memproyeksikan permintaan dari satu lokasi ke lokasi lain berdasarkan kebutuhan obat pada waktu dan kondisi tertentu, seperti saat musim hujan di suatu wilayah yang meningkatkan permintaan akan obat flu atau demam
- d. Dengan demikian, stok obat dapat dialokasikan dengan lebih efisien di distributor atau cabang tertentu sesuai dengan prediksi permintaan

2. Data Cuaca

Untuk memahami hubungan antara kebutuhan obat dan kondisi cuaca di suatu wilayah



**Link SQL, dan dataset
selengkapnya [disini](#)**



**Data visualization
selengkapnya [disini](#)**

Thank You

