

Soal & Template Jawaban

Task 5

Nama : Ridho Aditya Rachman

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 : Alasan utama adalah penggunaan fungsi LIKE dengan pola 'Mat%' yang secara langsung menangkap semua entri yang dimulai dengan kata "Mat" dalam kolom alamat. Di sisi lain, query a menggunakan fungsi SUBSTR untuk memeriksa karakter pertama hingga ketiga dari kolom alamat. Namun, pendekatan ini tidak sepenuhnya tepat karena tidak memperhitungkan kemungkinan variasi format alamat yang lebih kompleks. Penggunaan LIKE dengan pola memberikan fleksibilitas yang lebih besar dalam menangkap dan mencocokkan pola tertentu dalam data.

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 : Alasan utama adalah bahwa menggunakan klausa BETWEEN yang secara langsung mengevaluasi rentang tanggal_lahir yang diinginkan. Ini membuat sintaksis lebih ringkas dan mudah dipahami daripada query a yang menggunakan dua kondisi terpisah dengan operator logika AND. Selain itu, pendekatan menggunakan BETWEEN umumnya dianggap lebih bersih dan ekspresif dalam konteks query yang melibatkan rentang nilai, seperti pada kasus tanggal_lahir ini.

Soal 3: Menentukan Primary Key

A. Tugas

Tentukan primary key dari table penjualan. jelaskan alasannya

B. Jawaban & Penjelasan :

Primary key yang dipilih untuk tabel penjualan, yaitu `id_invoice`, dipilih karena sifatnya yang unik untuk setiap transaksi penjualan. Keunikan ini memastikan identifikasi yang jelas dan optimal untuk operasi join serta query. Selain itu, `id_invoice` juga stabil dan cenderung tidak berubah, mendukung integritas referensial, dan memiliki indeks yang mendukung pencarian efisien, sehingga menjadikannya pilihan yang tepat untuk menjamin konsistensi dan performa dalam struktur data penjualan.

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

B. Jawaban :

No	Nama File	Link
1.	database	https://drive.google.com/file/d/1heEXevhalw2eTikzk9Da4mMlaahcz7kr/view?usp=sharing
2.	Datamart (Penjualan)	https://drive.google.com/file/d/1PEGgnaa58Yr-B7ErOedLYDwl1wWg6t2L/view?usp=sharing
3.	Agregat Datamart (Total Penjualan per Brand)	https://drive.google.com/file/d/1gCFRFN796NuotEn1SsBo37C6llsh4nI0/view?usp=sharing

Table Base “Datamart (Penjualan)”

Query	Query History
1	CREATE TABLE datamart_penjualan_base AS
2	SELECT
3	p.id_invoice,
4	p.tanggal,
5	p.id_customer,
6	p.id_barang,
7	p.jumlah_barang,
8	p.harga,
9	p.mata_uang,
10	b.sektor,
11	b.nama_barang,
12	b.tipe,
13	b.nama_tipe,
14	b.kemasan,
15	c.level AS customer_level,
16	c.nama AS customer_nama,
17	c.cabang_sales AS customer_cabang_sales
18	FROM
19	penjualan p
20	JOIN
21	barang b ON p.id_barang = b.kode_barang
22	JOIN
23	pelanggan c ON p.id_customer = c.id_customer;
24	

Table Base “Datamart (Penjualan)”

column	data type	description	transformation
id_invoice	VARCHAR	Nomor invoice yang unik untuk setiap transaksi.	
tanggal	DATE	Tanggal transaksi penjualan.	
id_customer	VARCHAR	ID pelanggan yang melakukan transaksi.	
id_barang	VARCHAR	Kode barang yang terjual.	
jumlah_barang	INT	Jumlah barang yang terjual.	
harga	INT	Harga barang dalam transaksi.	
mata_uang	VARCHAR	Mata uang dalam transaksi	
sektor	VARCHAR	Sektor barang	

column	data type	description	transformation
nama_barang	VARCHAR	Nama barang yang terjual.	
tipe	VARCHAR	Tipe barang.	
nama_tipe	VARCHAR	Nama tipe barang.	
kemasan	VARCHAR	Jenis kemasan barang.	
customer_level	VARCHAR	Level pelanggan (misalnya: Company, Retail, etc.).	c.level AS customer_level
customer_nama	VARCHAR	Nama pelanggan.	c.nama AS customer_nama
customer_cabang_sales	VARCHAR	Cabang sales pelanggan.	c.cabang_sales AS customer_cabang_sales

Table Aggregate “Agregat Datamart (Total Penjualan per Brand)”

Query	Query History
1	CREATE TABLE datamart_total_penjualan_brand AS
2	SELECT
3	b.brand,
4	SUM (p.jumlah_barang) AS total_barang_terjual,
5	SUM (p.harga) AS total_penjualan
6	FROM
7	penjualan p
8	JOIN
9	barang b ON p.id_barang = b.kode_barang
10	GROUP BY
11	b.brand;
12	

Table Aggregate “Agregat Datamart (Total Penjualan per Brand)”

column	data type	description	transformation
brand	VARCHAR	Nama brand yang terkait dengan penjualan.	
total_barang_terjual	INT	Jumlah total barang yang terjual untuk setiap brand.	SUM(p.jumlah_barang) AS total_barang_terjual
total_penjualan	INT	Total nilai penjualan untuk setiap brand.	SUM(p.harga) AS total_penjualan

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

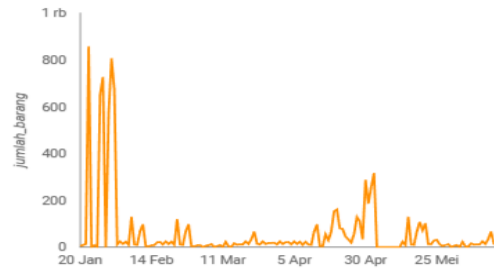
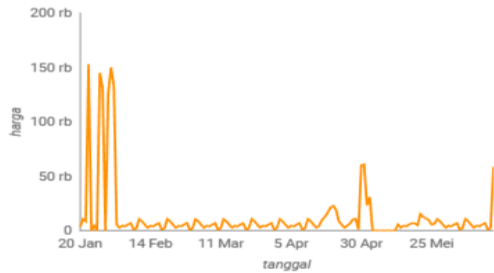
A. Jawaban :

Link visualisasi (ex link Google Data Studio) : <https://lookerstudio.google.com/reporting/1cf91be9-8afb-46d7-99ff-595fb2618fce>



Dashboard

Kimia Farma Sales Dashbiard(2023)



Pilih rentang tanggal ▾

harga
1.843.492,35

jumlah_barang
9.176

id_customer
350

Data Total Revenue dan Total Product

	tanggal	harga	jumlah_barang
1.	23 Jan 2022	153.037,36	858
2.	31 Jan 2022	150.098,65	807
3.	27 Jan 2022	145.188,75	651
4.	1 Feb 2022	132.648,7	672
5.	28 Jan 2022	130.888,3	727
6.	30 Jan 2022	126.285,5	571
7.	1 Mei 2022	61.045	186
8.	30 Apr 2022	60.263,1	288
9.	16 Jun 2022	58.730	152
10.	3 Mei 2022	31.029,2	317
11.	2 Mei 2022	23.487,23	257
12.	20 Apr 2022	22.898,71	161

1 - 100 / 136 < >

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 :

Data Stok Barang:

- Untuk memahami ketersediaan stok barang dan mengelola penjualan.

Data Pembelian Pelanggan Sebelumnya:

- Untuk memahami kebiasaan pembelian pelanggan dan memprediksi perilaku pembelian masa depan.