



Big Data Analytics - Kimia Farma

Final Project - Internship



Created By

Ammar Khaq Baasir

ammarbaasir19@gmail.com

<https://www.linkedin.com/in/ammar-baasir-973873234/>

Saya adalah mahasiswa S1 semester akhir jurusan teknik elektro bidang konsentrasi Telekomunikasi dan Informatika di Universitas Muhammadiyah Malang. Saya memiliki pengalaman dalam mempelajari dan mengimplementasikan Machine Learning dan Deep Learning dari Bangkit Academy by Google, GoTo, dan Traveloka. saya telah berhasil mengembangkan aplikasi berbasis Machine Learning dengan teknologi Computer Vision untuk mendeteksi objek produk - produk retail. Saya memiliki minat besar di bidang Data Scientist, Big Data, Machine Learning, dan Project Management. serta saya memiliki kemampuan dalam mengimplementasikan Machine Learning pada beberapa platform seperti Android, Website, dan Embedded System.



Daftar Isi

1	Our Goals
2	Problem 1 & 2 (Menentukan Penggunaan Query)
3	Problem 3 (Menentukan Primary Key)
4	Problem 4 (Design Datamart)
5	Problem 5 (Data Visualization/Dashboard)

Our Goals

Berikut beberapa tujuan yang harus dicapai dalam menyelesaikan final project ini



Goal # 1
Migrasi Data



Goal # 2
Make Primary Key



Goal # 3
Design Datamart

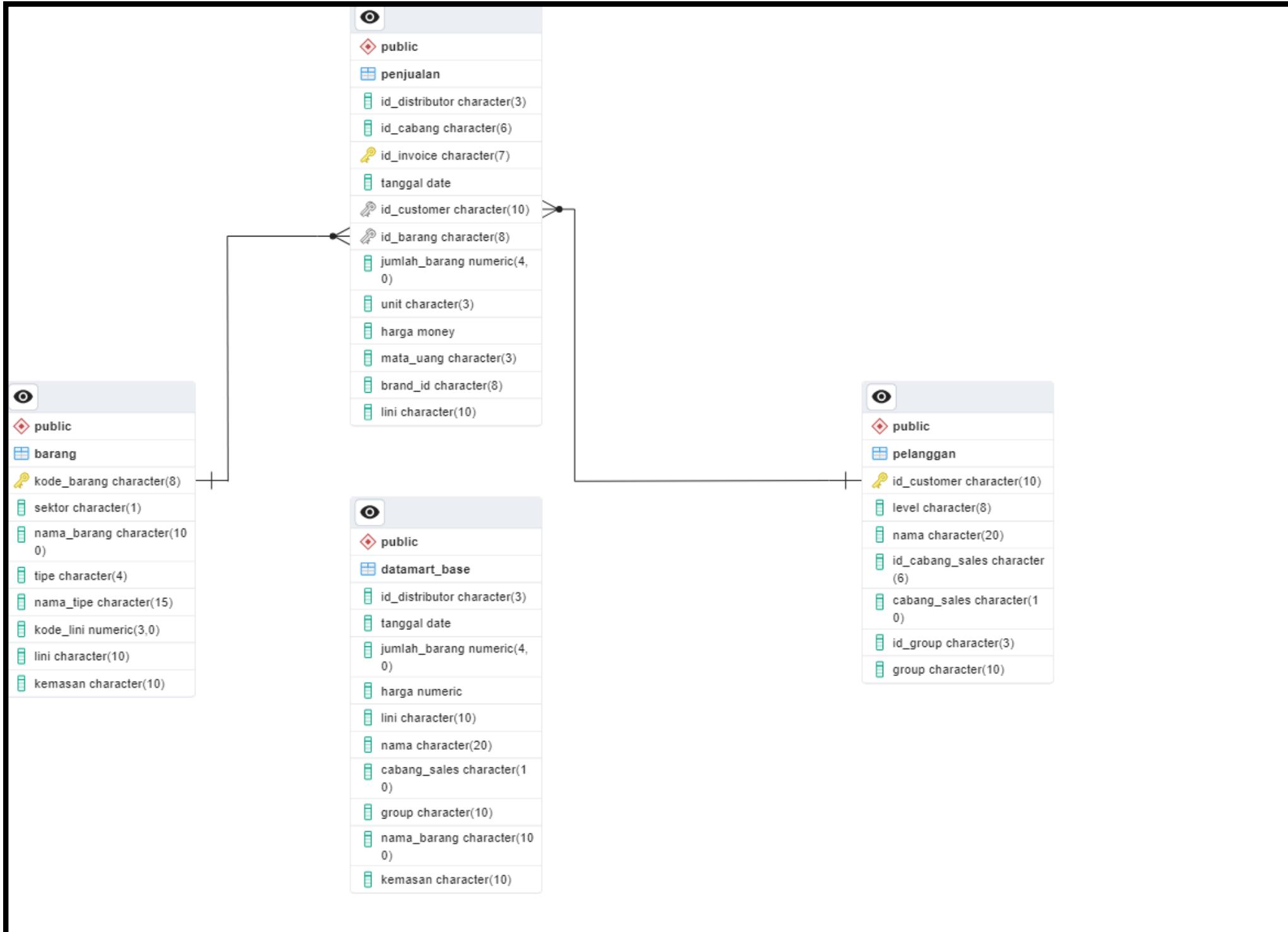


Goal # 4
Dashboard

MIGRASI DATA

Menentukan Design Database

DESIGN ERD DATABASE & DATAMART



Link Design ERD (Entity Relationship Diagram) : https://drive.google.com/file/d/1Ot-eGRoaBRk-N1H1JQoHergj-TFL_F4n/view?usp=sharing

MIGRASI DATA

Menentukan Design Database

QUERY DESIGN DATABASE

```
Query   Query History
5
6 CREATE TABLE IF NOT EXISTS public.barang
7 (
8     kode_barang character(8) COLLATE pg_catalog."default" NOT NULL,
9     sektor character(1) COLLATE pg_catalog."default",
10    nama_barang character(100) COLLATE pg_catalog."default",
11    tipe character(4) COLLATE pg_catalog."default",
12    nama_tipe character(15) COLLATE pg_catalog."default",
13    kode_lini numeric(3, 0),
14    lini character(10) COLLATE pg_catalog."default",
15    kemasan character(10) COLLATE pg_catalog."default",
16    CONSTRAINT barang_pkey PRIMARY KEY (kode_barang)
17 );
18
19 CREATE TABLE IF NOT EXISTS public.datamart_base
20 (
21     id_distributor character(3) COLLATE pg_catalog."default",
22     tanggal date,
23     jumlah_barang numeric(4, 0),
24     harga numeric,
```

Problem 1 & 2

Menentukan Penggunaan Query

Problem 1

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

- 1). SELECT * FROM pelanggan WHERE SUBSTR(alamat, 1, 3) = Mat;
- 2). SELECT * FROM pelanggan WHERE alamat LIKE 'Mat%'

Problem 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'

Problem 1 & 2

Menentukan Penggunaan Query

Problem 1

Jawaban :

Pilihan yang tepat adalah B. dari segi penulisan query kedua pilihan tersebut sebenarnya sudah tepat dan benar. Akan tetapi, dari segi kemudahan dalam membaca dan memahami query opsi B adalah pilihan yang tepat, hal ini dikarenakan BETWEEN secara langsung akan menyatakan rentang nilai yang akan diambil yakni '2000-01-01' hingga '2008-12-31'. Selain itu, menggunakan query opsi B akan membuat penulisan query lebih ringkas, dimana kita tidak perlu untuk menuliskan nama kolom yang akan dimasukkan kedalam suatu kondisi secara berulang – ulang.

Problem 2

Jawaban :

Query yang akan bekerja lebih baik adalah query 2. penggunaan query LIKE sangat dianjurkan ketika ingin mencari sebuah data dengan nilai atau kondisi tertentu, query ini akan memiliki efisiensi serta kecepatan proses yang lebih baik pada saat data yang dimiliki berukuran besar. LIKE akan memproses nilai yang memiliki awalan 'Mat%' tanpa harus melakukan proses pemotongan pada 3 karakter awal seperti pada penggunaan query SUBSTR.

Problem 3

Menentukan Primary Key

QUERY PEMBUATAN KOLOM “id_penjualan”

```
1 ALTER TABLE penjualan  
2 ADD COLUMN id_penjualan VARCHAR(20);  
3
```

OUTPUTAN TABEL

oice ter	tanggal date	id_customer character	id_barang character	jumlah_barang numeric (4)	unit character	harga money	mata_uang character	brand_id character	lini character	id_penjualan character varying (20)
1	7	2022-01-20	CUST55380	BRG0001	1 DUS	Rp116.991,00	IDR	BRND001	OGB & PH	[null]
2	7	2022-01-20	CUST55381	BRG0002	5 DUS	Rp23.375,00	IDR	BRND002	ETIKAL	[null]
3	5	2022-01-21	CUST55382	BRG0003	9 DUS	Rp106.906,00	IDR	BRND003	MARCKS	[null]
4	4	2022-01-22	CUST55383	BRG0004	13 DUS	Rp87.007,00	IDR	BRND004	VNS	[null]
5	0	2022-01-23	CUST55384	BRG0005	1 DUS	Rp56.483,00	IDR	BRND005	SLCYL	[null]

Dikarenakan pada tabel penjualan tidak terdapat satupun kolom yang memiliki nilai unique maka untuk penentuan primary key dilakukan dengan pembuatan kolom baru bernama id_penjualan seperti query diatas

Link Query Add Column : https://drive.google.com/file/d/1G_c4KuPj_gcvRkDExNzJslObwoBi2E/view?usp=sharing

Problem 3

Menentukan Primary Key

QUERY PEMBUATAN PRIMARY KEY “id_penjualan”

```
1 UPDATE penjualan
2 SET id_penjualan = CONCAT(RIGHT(id_cabang, 2), '_', id_invoice, '_', id_customer);
3
4 ALTER TABLE penjualan
5 ADD PRIMARY KEY (id_penjualan);
6
```

OUTPUTAN TABEL

	tanggal date	id_customer character	id_barang character	jumlah_barang numeric (4)	unit character	harga money	mata_uang character	brand_id character	lini character	id_penjualan [PK] character varying (20)
1	2022-01-28	CUST55401	BRG0002	25	DUS	Rp23.375,00	IDR	BRND002	ETIKAL	10_IN6141_CUST55401
2	2022-01-23	CUST55402	BRG0003	14	DUS	Rp106.906,00	IDR	BRND003	MARCKS	01_IN6125_CUST55402
3	2022-01-30	CUST55403	BRG0004	18	DUS	Rp87.007,00	IDR	BRND004	VNS	01_IN6176_CUST55403
4	2022-01-31	CUST55404	BRG0005	18	DUS	Rp56.483,00	IDR	BRND005	SLCYL	02_IN6175_CUST55404

Link Query Add Primary Key : <https://drive.google.com/file/d/10BR1hwbgjb8qkJ72b5T1yhA2lsg9ApR0/view?usp=sharing>

Problem 4

Design Datamart

MENENTUKAN DESIGN DATAMART

Dari 3 tabel yang tersedia (penjualan, pelanggan, dan barang) saya asumsikan 3 tabel tersebut ada di datalake.

Untuk membuat datamart, akan ada 2 langkah :

1. Tabel base

- Tabel base merupakan tabel dari hasil penggabungan 3 tabel dengan granularity id_penjualan sebagai Primary Key (yaitu penggabungan antara id_invoice dan id_barang)
- Total row data pada tabel base sama dengan total pada tabel penjualan
- Tabel base disimpan di dalam data warehouse

2. Tabel aggregate

- Tabel aggregate merupakan turunan dari tabel base dimana data dikelompokkan berdasarkan tanggal, id_customer dan id_invoice
- Tabel aggregate disimpan di dalam datamart

Problem 4

Design Datamart

LINK FILE DESIGN DATAMART

No	Nama File	Link
1	Datamart Base	https://drive.google.com/file/d/1IxS_0x0b5TJP HUDGNLkjRr-4q8wB3n/view?usp=sharing
2	Datamart Pendapatan	https://drive.google.com/file/d/1rdtT_YapM0ad_1vFwmUYwo94QInK8DD/view?usp=sharing
3	Datamart Transaksi	https://drive.google.com/file/d/11s9qO1DM-wfEMS6kQcA9PoP52o2y1znM/view?usp=sharing
4	Datamart Pelanggan	https://drive.google.com/file/d/1I_eZCLYMH4rshN1dsuuY2p3aqcl_NJy6/view?usp=sharing

Problem 4

Design Datamart

QUERY DESIGN DATAMART BASE (datamart_base)

```
CREATE TABLE datamart_base AS (
    SELECT
        sales.id_distributor, sales.tanggal, sales.jumlah_barang,
        sales.harga, sales.lini, customer.nama, customer.cabang_sales, customer.group,
        product.nama_barang, product.kemasan
    FROM
        penjualan AS sales
    JOIN
        pelanggan AS customer
    ON
        sales.id_customer = customer.id_customer
    JOIN
        barang AS product
    ON
        sales.id_barang = product.kode_barang
);
```

Problem 4

Design Datamart

DESKRIPSI KOLOM PADA TABEL DATAMART BASE

Column	Data Type	Description	Transformation
Id_distributor	Character Varying	Id atau kata kunci dari nama distributor	-
tanggal	Date	Tanggal terjadinya transaksi	-
jumlah_barang	Numeric	Jumlah barang yang terjual	-
harga	Money	Harga dari satu buah produk	-
lini	Character Varying	Nama brand dari produk	-
nama	Character Varying	Nama outlet dari mitra	-
cabang_sales	Character Varying	Tempat kota dari outlet	-
group	Character Varying	Jenis group dari outlet	-
nama_barang	Character Varying	Nama produk yang dijual pada outlet	-
kemasan	Character Varying	Kemasan yang digunakan pada produk	-

Problem 4

Design Datamart

TABEL DATAMART BASE “datamart_base”

	id_distributor character	tanggal date	jumlah_barang numeric	harga money	lini character	nama character	cabang_sales character	group character	nama_barang character	
1	TD	2022-01-20		1.0	Rp1.170,00	OGB & PH	APOTEK TAPAK	Aceh	Apotek	ACYCLOVIR DUS
2	TD	2022-01-20		5.0	Rp2.338,00	ETIKAL	APOTEK MAJA	Kuningan	Apotek	ALERGINE TABLET SALUT
3	TA	2022-01-21		9.0	Rp10.691,00	MARCKS	KLINIK GM	Jakarta	Klinik	AMPICILLIN
4	EPM	2022-01-22		13.0	Rp8.701,00	VNS	APOTEK MERDEKA ...	Bandung	Apotek	TRAMADOL KAPSUL 50 MG
5	TD	2022-01-23		1.0	Rp5.648,00	SLCYL	KLINIK SAHABAT	Tangerang	Klinik	KLORPROMAZINA TABLET SALUT
6	TA	2022-01-23		5.0	Rp2.819,00	OGB & PH	APOTEK SINAR JAY...	Bekasi	Apotek	KETOCONAZOLE TABLET 200 MG
7	EPM	2022-01-25		9.0	Rp4.592,00	ETIKAL	APOTEK SAHABAT ...	Padang	Apotek	ERGOTAMINE COFFEINE
8	TD	2022-01-23		1.0	Rp3.992,00	MARCKS	APOTEK TAPAK	Aceh	Apotek	TETRACYCLINE KAPSUL 250 MG
9	TA	2022-01-27		23.0	Rp5.781,00	VNS	APOTEK MAJA	Kuningan	Apotek	AMBROXOL HC
10	EPM	2022-01-28		4.0	Rp6.940,00	SLCYL	KLINIK GM	Jakarta	Klinik	PARACETAMOL
11	TD	2022-01-23		2.0	Rp1.170,00	OGB & PH	APOTEK MERDEKA ...	Bandung	Apotek	ACYCLOVIR DUS
12	TD	2022-01-30		17.0	Rp2.338,00	ETIKAL	KLINIK SAHABAT	Tangerang	Klinik	ALERGINE TABLET SALUT
13	TA	2022-01-31	11.3181818181818	Rp10.691,00	MARCKS	APOTEK SINAR JAY...	Bekasi	Apotek	AMPICILLIN	

Problem 4

Design Datamart

QUERY DATAMART AGGREGATE “datamart_pendapatan”

```
SELECT
    id_distributor AS distributor, cabang_sales AS kota,
    EXTRACT(MONTH FROM tanggal) AS bulan,
    ROUND(SUM(jumlah_barang)) AS jumlah_terjual,
    SUM(harga * jumlah_barang) AS pendapatan
FROM
    datamart_base
GROUP BY
    distributor, kota, bulan
ORDER BY
    bulan ASC, pendapatan DESC;
```

Problem 4

Design Datamart

DESKRIPSI KOLOM DATAMART AGGREGATE “datamart_penjualan”

Column	Data Type	Description	Transformation
distributor	Character Varying	Id atau kata kunci dari nama distributor	-
kota	Character Varying	Tempat kota dari outlet berada	-
bulan	Numeric	Bulan terjadinya transaksi	EXTRACT(MONTH FROM tanggal)
jumlah_terjual	Numeric	Jumlah barang yang terjual	-
pendapatan	Money	Total pendapatan dari transaksi yang dilakukan	jumlah_terjual * harga

Problem 4

Design Datamart

TABEL DATAMART AGGREGATE “datamart_pendapatan”

	distributor character 	kota character 	bulan numeric 	jumlah_terjual numeric 	pendapatan money 
1	TD	Jakarta	1	457	Rp3.277.352,62
2	TA	Jakarta	1	317	Rp2.392.649,76
3	TA	Bekasi	1	225	Rp1.798.811,23
4	TD	Padang	1	250	Rp1.161.466,95
5	EPM	Bandung	1	243	Rp1.080.450,08
6	EPM	Jakarta	1	170	Rp953.736,00
7	TA	Bandung	1	208	Rp940.279,01
8	TD	Aceh	1	159	Rp895.290,19
9	TA	Kuningan	1	158	Rp889.879,46
10	TD	Tangerang	1	184	Rp848.996,63
11	EPM	Kuningan	1	95	Rp710.525,04
12	EPM	Tangerang	1	158	Rp638.571,95
13	TD	Bekasi	1	122	Rp637.778,13
14	TD	Kuningan	1	124	Rp622.816,07

Problem 4

Design Datamart

QUERY TABEL DATAMART AGGREGATE “datamart_transaksi”

```
SELECT
    lini AS nama_brand, nama_barang,
    EXTRACT(MONTH FROM tanggal) AS bulan,
    ROUND(SUM(jumlah_barang)) AS jumlah_terjual,
    SUM(harga * jumlah_barang) AS pendapatan
FROM
    datamart_base
GROUP BY
    nama_brand, nama_barang, bulan
ORDER BY
    bulan ASC, jumlah_terjual DESC;
```

Problem 4

Design Datamart

DESKRIPSI KOLOM DATAMART AGGREGATE “datamart_transaksi”

Column	Data Type	Description	Transformation
nama_brand	Character Varying	Nama brand dari produk	-
nama_barang	Character Varying	Nama produk yang terjual pada outlet	-
bulan	Numeric	Bulan terjadinya transaksi	EXTRACT(MONTH FROM tanggal)
jumlah_terjual	Numeric	Jumlah barang yang terjual	-
pendapatan	Money	Total pendapatan dari transaksi yang dilakukan	jumlah_terjual * harga

Problem 4

Design Datamart

TABEL DATAMART AGGREGATE “datamart_transaksi”

	nama_brand character 	nama_barang character 	bulan numeric 	jumlah_terjual numeric 	pendapatan money 
1	MARCKS	AMPICILLIN	1	552	Rp5.902.889,87
2	OGB & PH	ACYCLOVIR DUS	...	1	Rp570.935,45
3	SLCYL	KLORPROMAZINA TABLET SAL...	1	402	Rp2.271.068,70
4	ETIKAL	ERGOTAMINE COFFEINE	...	1	Rp1.285.133,83
5	VNS	TRAMADOL KAPSUL 50 MG	...	1	Rp2.176.132,27
6	VNS	AMBROXOL HC	...	1	Rp1.430.757,07
7	MARCKS	TETRACYCLINE KAPSUL 250 ...	1	240	Rp958.080,00
8	SLCYL	PARACETAMOL	...	1	Rp1.482.854,76
9	ETIKAL	ALERGINE TABLET SALUT	...	1	Rp495.745,92
10	VNS	AMPICILLIN	1	198	Rp1.142.717,74
11	OGB & PH	KETOCONAZOLE TABLET 200 ...	1	198	Rp558.083,15
12	ETIKAL	ACYCLOVIR DUS	...	1	Rp753.842,63
13	MARCKS	ALERGINE TABLET SALUT	...	1	Rp544.880,08
14	SLCYL	TRAMADOL KAPSUL 50 MG	...	1	Rp506.425,88

Problem 4

Design Datamart

QUERY TABEL DATAMART AGGREGATE “datamart_pelanggan”

```
SELECT
    dm.cabang_sales AS kota, dm.nama AS nama_mitra,
    dm.group AS mitra_group,
    count(nama) AS jumlah_mitra
FROM
    datamart_base AS dm
GROUP BY
    kota, nama_mitra, mitra_group
ORDER BY
    jumlah_mitra DESC;
```

Problem 4

Design Datamart

DESKRIPSI KOLOM DATAMART AGGREGATE “datamart_pelanggan”

Column	Data Type	Description	Transformation
kota	Character Varying	Tempat kota dari outlet berada	-
nama_mitra	Character Varying	Nama outlet dari mitra	-
mitra_group	Character Varying	Jenis group dari outlet	-
jumlah_mitra	Bigint	Jumlah keseluruhan mitra berdasarkan kota dan mitra group	Count(nombre)

Problem 4

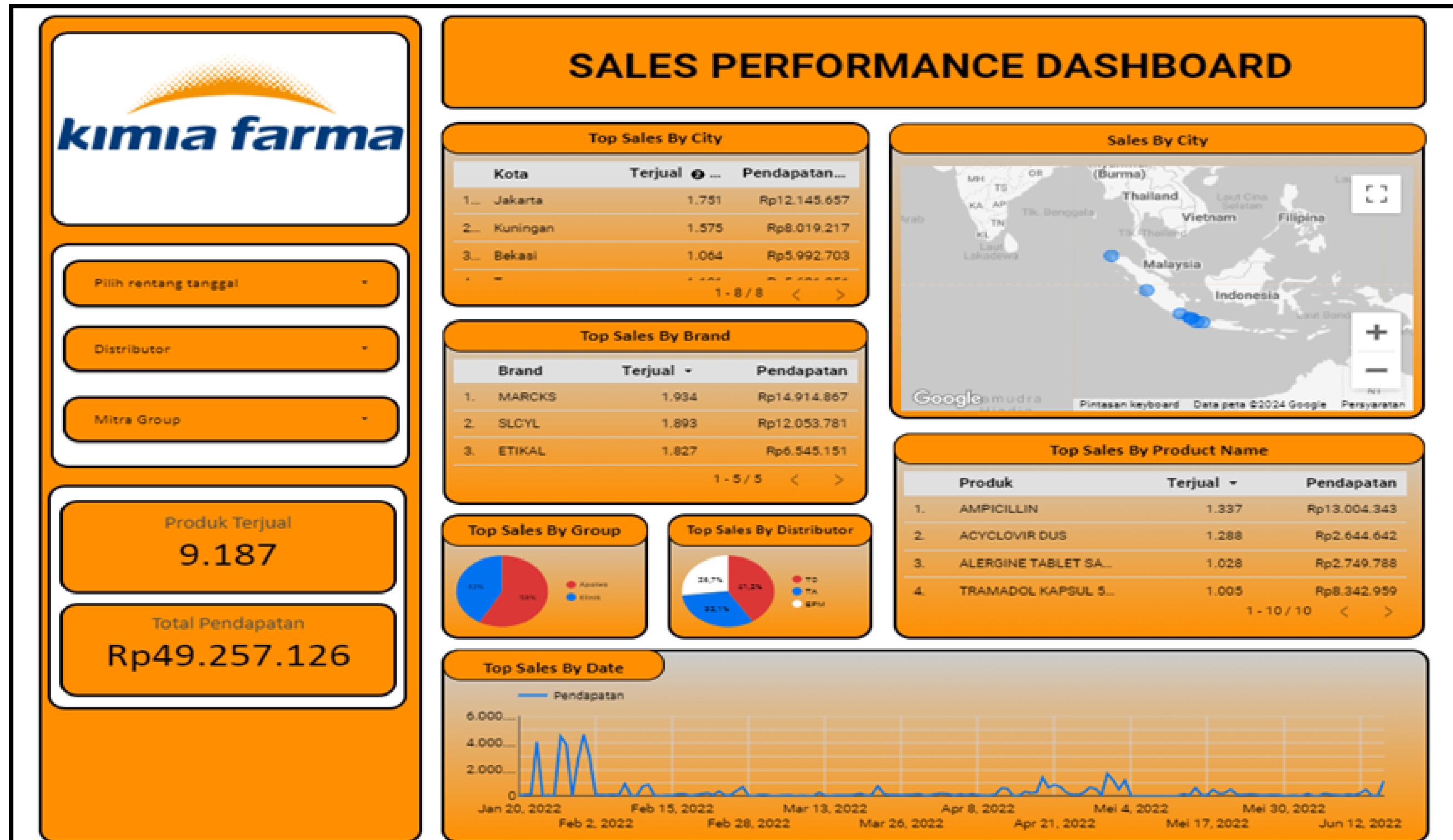
Design Datamart

TABEL DATAMART AGGREGATE
“datamart_pelanggan”

	kota character	nama_mitra character	mitra_group character	jumlah_mitra bigint
1	Bandung	APOTEK MERDEKA...	Apotek	47
2	Jakarta	KLINIK GM	Klinik	47
3	Tangerang	KLINIK SAHABAT	Klinik	47
4	Aceh	APOTEK TAPAK	Apotek	47
5	Kuningan	APOTEK MAJA	Apotek	47
6	Padang	APOTEK SAHABAT ...	Apotek	46
7	Bekasi	APOTEK SINAR JA...	Apotek	46
8	Lampung	KLINIK DR. ANDRI	Klinik	23

Problem 5

Data Visualization



Link Dashboard : <https://lookerstudio.google.com/reporting/5208159f-8027-4ba5-b105-3ec794073cd8>



TERIMA KASIH

Final Project - Internship



