Rakamin Academy

KIMIA FARMA

kimia farma

Big Data Analytics: Salicyl Sales Dashboard

Anually recap by Bashara Aina

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

List of Content

Background - Determine the background data & analysis

Query - Q&A about Query Questions

Primary Key (PK) - Decide the PK of the database

Datamart Design - From raw 6 tables to Datamart

Base Table - Detail and raw data

Aggregate Table - Summary of Base Table

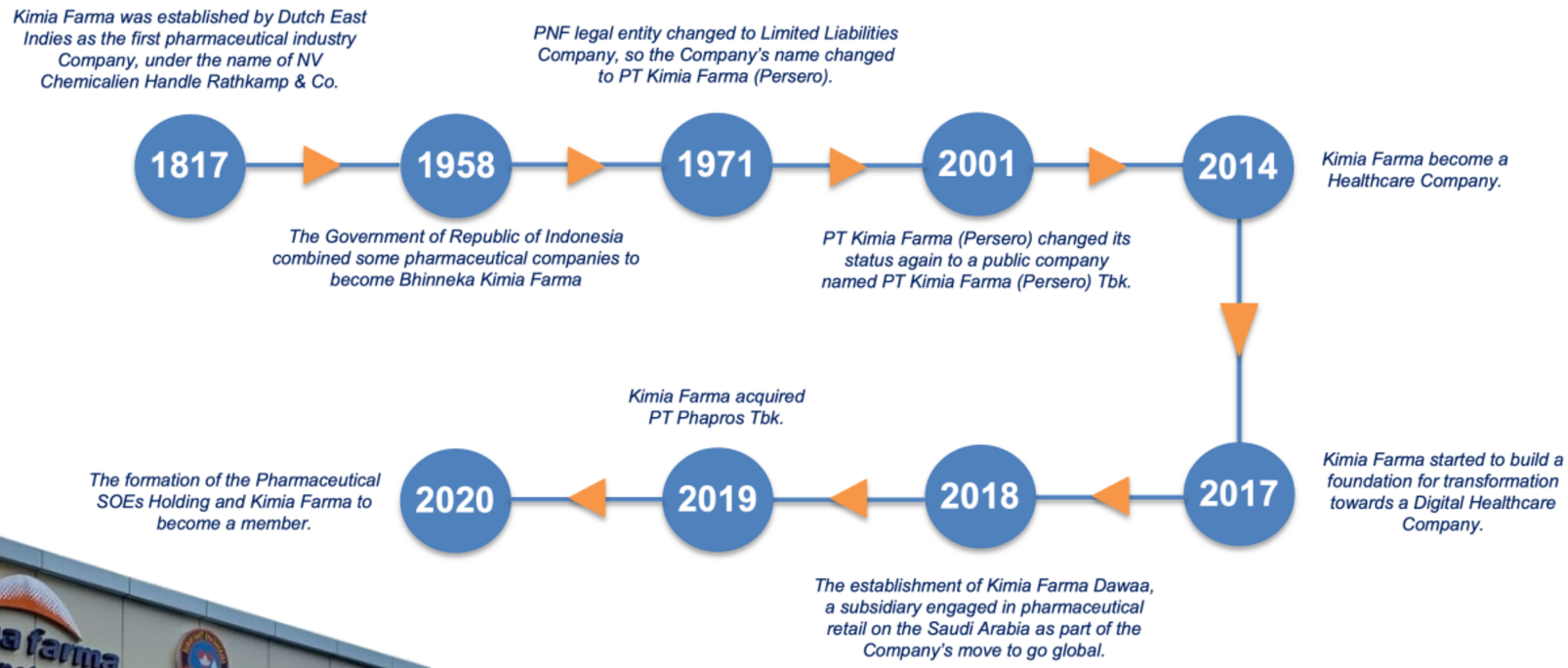
Data Visualization - A step of data process until it turn out visualize and the elaborate of insights

Additional Complementary - Advice and recommendations for company

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Background

Kimia Farma, established in 1817, is the pioneer of the pharmaceutical industry in Indonesia, founded by the Dutch East Indies Government. As a Big Data Analytics professional at PT. Kimia Farma, Tbk, I was tasked with generating reports and dashboards related to the sales data of Brand Salycil. The data spans from January to June 2022 and after conducting a brief overview, I have identified key insights to help inform business decisions.



Rakamin Academy

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

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%'

Alasan : Both of these queries will retrieve data from the pelanggan table that has an address starting with the word "Mat". However, the second query, SELECT * FROM pelanggan WHERE alamat LIKE 'Mat%', will work better because it uses the LIKE operator. This operator will read data that starts with the word "Mat" in the database. The first query will not read data that starts with "Mat" if the address name is not exactly 3 characters. Therefore, the second query will work better in this case.

Advantages:

- LIKE operator allows for more flexible search patterns by using wildcards such as % and _ to match any number of characters or single character respectively. This makes it easier to retrieve data that meets specific criteria without having to specify exact string matches like in the first query.
- The SUBSTR function in the first query is more restrictive and only retrieves data that matches a specific substring of a certain length. This makes it more difficult to retrieve data that does not exactly match the criteria specified in the query, leading to a potentially incomplete or inaccurate result.

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

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'

Alasan : Both queries will give the same result, but using the BETWEEN operator is more concise and easier to read. It specifies a range of dates that we want to retrieve, and the BETWEEN operator includes both endpoints of the range. On the other hand, the first query uses the comparison operators >= and <= to check if the date of birth is within the specified range. Although it achieves the same result, it requires more characters to write and may be harder to understand for beginners.

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Primary Key

Soal 3:

Tentukan primary key dari table penjualan. jelaskan alasannya

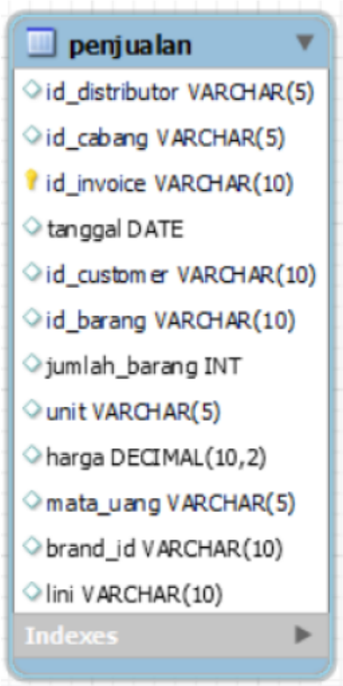
Alasan : Based on the given attributes, the most suitable primary key for the tabel penjualan is the id_invoice. The reason is that id_invoice has a unique value for each row or record in the table. This fulfills the basic criteria of a primary key, which is a unique and non-null value.

On the other hand, other attributes such as id_distributor, id_cabang, and id_barang have some duplicate values, making them unsuitable as a single primary key. Meanwhile, attributes such as tanggal, unit, brand_id, and lini cannot be used as a primary key because their values are not unique. However, customer_id has unique values, but it is already the primary key of the tabel pelanggan. Therefore, customer_id becomes a foreign key in the tabel penjualan.

By using id_invoice as the primary key, each record in the tabel penjualan can be identified uniquely and efficiently. Additionally, id_invoice is often used as a reference in the process of recording and tracking sales, making it a good candidate for the primary key in the tabel penjualan

Unik

id_distributor	id_cabang	id_invoice	tanggal	id_customer	id_barang
TD	CAB09	IN5978	2022-05-25	CUST55670	BRG0001
TA	CAB01	IN5979	2022-03-28	CUST55598	BRG0009
TA	CAB01	IN5980	2022-01-23	CUST55432	BRG0003
EPM	CAB07	IN5981	2022-01-28	CUST55443	BRG0004
TD	CAB09	IN5982	2022-01-30	CUST55535	BRG0006
EPM	CAB06	IN5983	2022-04-21	CUST55622	BRG0003
TD	CAB08	IN5984	2022-05-01	CUST55714	BRG0005
TD	CAB09	IN5985	2022-04-18	CUST55640	BRG0001
TA	CAB01	IN5986	2022-02-15	CUST55557	BRG0002
EPM	CAB02	IN5987	2022-04-19	CUST55659	BRG0010
EPM	CAB10	IN5988	2022-02-09	CUST55551	BRG0002
TA	CAB02	IN5989	2022-05-01	CUST55719	BRG0010
TA	CAB01	IN5990	2022-05-03	CUST55706	BRG0001
TD	CAB08	IN5991	2022-01-23	CUST55414	BRG0005



List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Datamart Design

Soal 4:

Buatlah design datamart (Terdiri dari tabel base, dan tabel aggregate).

Alasan : Kimia Farma's database consists of 6 tables which are Table Penjualan (Sales), Penjualan_ds (details), Table Pelanggan (Customer), pelanggan_ds (details), Table Barang (Product), and barang_ds. These tables contain data related to sales, customers, and products. The database has been converted from Kimia Farma's data using MySQL.

The Table Penjualan (Sales) contains information about the sales transactions, such as the invoice number, date, and total sales amount. The Penjualan_ds (details) table contains more detailed information about each transaction, such as the product sold, the quantity, and the price. These tables are related through the invoice number.

The Table Pelanggan (Customer) contains information about the customers, such as their name, address, and phone number. The pelanggan_ds (details) table contains more detailed information about each customer, such as their transaction history. These tables are related through the customer ID.

The Table Barang (Product) contains information about the products sold, such as the name, brand, and unit price. The barang_ds table contains more detailed information about each product, such as the supplier and the manufacturing date. These tables are related through the product ID.

Table	Action	Rows	Type	Collation	Size	Overhead
barang	Browse Structure Search Insert Empty Drop	10	InnoDB	utf8mb4_general_ci	16.0 KiB	-
barang_ds	Browse Structure Search Insert Empty Drop	10	InnoDB	utf8mb4_general_ci	16.0 KiB	-
pelanggan	Browse Structure Search Insert Empty Drop	350	InnoDB	utf8mb4_general_ci	64.0 KiB	-
pelanggan_ds	Browse Structure Search Insert Empty Drop	350	InnoDB	utf8mb4_general_ci	64.0 KiB	-
penjualan	Browse Structure Search Insert Empty Drop	350	InnoDB	utf8mb4_general_ci	64.0 KiB	-
penjualan_ds	Browse Structure Search Insert Empty Drop	132	InnoDB	utf8mb4_general_ci	16.0 KiB	-
6 tables	Sum	1,202	InnoDB	utf8mb4_general_ci	240.0 KiB	0 B

link to dataset: https://intip.in/KimiaFarmaFP

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Datamart Design

Soal 4:

Buatlah design datamart (Terdiri dari tabel base, dan tabel aggregate), Cont.

Nama File	Nama Table		Link
Data_Mart.sql	BASE TABLE penjualan_all salicyl_database	AGGREGATE TABLE sales_date total_customer total_sales_per_cbranch total_sales_per_customer	https://intip.in/DataMartKF

Table	Action	Rows	Type	Collation	Size	Overhead
penjualan_all	Browse Structure Search Insert Empty Drop	350	InnoDB	utf8mb4_general_ci	48 KiB	-
sales_date	Browse Structure Search Insert Empty Drop	31	InnoDB	utf8mb4_general_ci	16 KiB	-
salicyl_database	Browse Structure Search Insert Empty Drop	59	InnoDB	utf8mb4_general_ci	16 KiB	-
total_customer	Browse Structure Search Insert Empty Drop	59	InnoDB	utf8mb4_general_ci	16 KiB	-
total_sales_per_branch	Browse Structure Search Insert Empty Drop	3	InnoDB	utf8mb4_general_ci	16 KiB	-
total_sales_per_customer	Browse Structure Search Insert Empty Drop	2	InnoDB	utf8mb4_general_ci	16 KiB	-
6 tables	Sum	504	InnoDB	latin1_swedish_ci	128 KiB	0 B

link to dataset: https://intip.in/KimiaFarmaFP

Base Table - penjualan_all

- List of Content
- Background
- Query
- Primary Key
- Datamart Design
- Base Table
- Aggregate Table
- Data Visualization
- Complementary


```
-- Membuat tabel sementara untuk menampung data penjualan setelah penyesuaian harga
CREATE TABLE penjualan_all AS
SELECT id_invoice, tanggal, id_customer, id_barang, jumlah_barang, unit, harga
FROM (
  -- Query untuk mengupdate harga pada tabel finalproject.penjualan
  SELECT
    id_invoice, tanggal, id_customer, id_barang, jumlah_barang, unit,
    CASE
      WHEN id_barang = 'BRG0005' THEN 47000 -- Update harga untuk barang BRG0005
      WHEN id_barang = 'BRG0010' THEN 21000 -- Update harga untuk barang BRG0010
      ELSE harga
    END AS harga
  FROM finalproject.penjualan
) penjualan_harga;
```

Column Name	Datatype
id_invoice	VARCHAR(10)
tanggal	DATE
id_customer	VARCHAR(10)
id_barang	VARCHAR(10)
jumlah_barang	INT
unit	VARCHAR(5)
harga	INT

id_invoice	tanggal	id_customer	id_barang	jumlah_barang	unit	harga
IN5978	2022-05-25	CUST55670	BRG0001	9	DUS	1169
IN5979	2022-03-28	CUST55598	BRG0009	18	DUS	5780
IN5980	2022-01-23	CUST55432	BRG0003	20	DUS	10690
IN5981	2022-01-28	CUST55443	BRG0004	76	DUS	8700
IN5982	2022-01-30	CUST55535	BRG0006	5	DUS	2819
IN5983	2022-04-21	CUST55622	BRG0003	67	DUS	10690
IN5984	2022-05-01	CUST55714	BRG0005	8	DUS	47000
IN5985	2022-04-18	CUST55640	BRG0001	22	DUS	1169
IN5986	2022-02-15	CUST55557	BRG0002	9	DUS	3991
IN5987	2022-04-19	CUST55659	BRG0010	14	DUS	21000
IN5988	2022-02-09	CUST55551	BRG0002	10	DUS	2337
IN5989	2022-05-01	CUST55719	BRG0010	67	DUS	21000
IN5990	2022-05-03	CUST55706	BRG0001	10	DUS	4592
IN5991	2022-01-23	CUST55414	BRG0005	24	DUS	47000
IN5992	2022-06-16	CUST55707	BRG0002	23	DUS	3991
IN5993	2022-05-26	CUST55671	BRG0002	1	DUS	2337

Base Table - salicyl_database

- List of Content
- Background
- Query
- Primary Key
- Datamart Design
- Base Table
- Aggregate Table
- Data Visualization
- Complementary


```sql
-- Membuat tabel salicyl_database dari tabel penjualan_all
CREATE TABLE salicyl_database AS (
  SELECT
    pa.tanggal,
    pa.id_invoice,
    pa.id_customer,
    pl.nama AS nama_pelanggan,
    pl.cabang_sales,
    pl.group,
    pl.id_distributor,
    pa.id_barang,
    ba.nama_barang,
    pa.unit,
    pa.harga,
    pa.jumlah_barang,
    pa.harga * pa.jumlah_barang AS harga_total
  FROM penjualan_all pa
  JOIN barang_ds ba ON pa.id_barang  = ba.kode_barang
  JOIN pelanggan_ds pl ON pa.id_customer = pl.id_customer
  WHERE pa.id_barang IN ('BRG0005', 'BRG0010')
);
```

Column Name	Datatype
tanggal	DATE
id_invoice	VARCHAR(10)
id_customer	VARCHAR(10)
nama_pelanggan	VARCHAR(50)
cabang_sales	VARCHAR(10)
group	VARCHAR(10)
id_distributor	VARCHAR(5)
id_barang	VARCHAR(10)
nama_barang	VARCHAR(50)
unit	VARCHAR(5)
harga	INT
jumlah_barang	INT
harga_total	INT

tanggal	id_invoice	id_customer	nama_pelanggan	cabang_sales	group	id_distributor	id_barang	nama_barang	unit	harga	jumlah_barang	harga_total
2022-05-01	IN5984	CUST55714	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	8	376000
2022-04-19	IN5987	CUST55659	KLINIK GM	Jakarta	Klinik	TA	BRG0010	PARACETAMOL	DUS	21000	14	294000
2022-05-01	IN5989	CUST55719	KLINIK GM	Jakarta	Klinik	TA	BRG0010	PARACETAMOL	DUS	21000	67	1407000
2022-01-23	IN5991	CUST55414	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	24	1128000
2022-01-27	IN6003	CUST55424	KLINIK DR. ANDRI	Lampung	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	9	423000
2022-05-01	IN6004	CUST55729	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0010	PARACETAMOL	DUS	21000	5	105000
2022-02-12	IN6006	CUST55554	KLINIK GM	Jakarta	Klinik	TA	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	5	235000
2022-03-09	IN6016	CUST55579	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0010	PARACETAMOL	DUS	21000	5	105000
2022-05-23	IN6022	CUST55654	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	9	423000
2022-02-01	IN6023	CUST55399	KLINIK SAHABAT	Tangerang	Klinik	TD	BRG0010	PARACETAMOL	DUS	21000	15	315000
2022-01-27	IN6024	CUST55394	KLINIK DR. ANDRI	Lampung	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	12	564000
2022-01-28	IN6029	CUST55389	KLINIK GM	Jakarta	Klinik	TA	BRG0010	PARACETAMOL	DUS	21000	4	84000
2022-05-01	IN6044	CUST55694	KLINIK DR. ANDRI	Lampung	Klinik	TD	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	18	846000
2022-01-31	IN6048	CUST55464	KLINIK GM	Jakarta	Klinik	TA	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	25	1175000
2022-01-28	IN6049	CUST55479	KLINIK GM	Jakarta	Klinik	TA	BRG0010	PARACETAMOL	DUS	21000	9	189000
2022-01-31	IN6050	CUST55434	KLINIK GM	Jakarta	Klinik	TA	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	DUS	47000	34	1598000

Aggregate Table

- List of Content
- Background
- Query
- Primary Key
- Datamart Design
- Base Table
- Aggregate Table
- Data Visualization
- Complementary


```sql
CREATE TABLE sales_date AS( -- Menghitung total penjualan per tanggal
SELECT tanggal, SUM(harga_total) AS total_penjualan, SUM(jumlah_barang) AS total_barang
FROM salicyl_database
GROUP BY tanggal);
```

Column Name	Datatype
tanggal	DATE
total_penjualan	INT

tanggal	total_penjualan	total_barang
2022-05-01	5801000	171
2022-04-19	294000	14
2022-01-23	6862000	146
2022-01-27	2068000	44
2022-02-12	235000	5
2022-03-09	105000	5
2022-05-23	423000	9
2022-02-01	2436000	116
2022-01-28	4494000	214
2022-01-31	9917000	211
2022-03-24	564000	12
2022-03-14	94000	2
2022-05-29	611000	13
2022-03-29	399000	19
2022-02-27	2058000	98
2022-04-23	2115000	45

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Aggregate Table

```sql
CREATE TABLE total_customer AS( -- Menghitung total penjualan per pelanggan
SELECT id_customer, nama_pelanggan, SUM(harga_total) AS total_penjualan
FROM salicyl_database
GROUP BY id_customer, nama_pelanggan);
```

Column Name	Datatype
id_customer	VARCHAR(10)
nama_pelanggan	VARCHAR(50)
total_penjualan	INT

id_customer	nama_pelanggan	total_penjualan
CUST55714	KLINIK SAHABAT	376000
CUST55659	KLINIK GM	294000
CUST55719	KLINIK GM	1407000
CUST55414	KLINIK SAHABAT	1128000
CUST55424	KLINIK DR. ANDRI	423000
CUST55729	KLINIK SAHABAT	105000
CUST55554	KLINIK GM	235000
CUST55579	KLINIK SAHABAT	105000
CUST55654	KLINIK SAHABAT	423000
CUST55399	KLINIK SAHABAT	315000
CUST55394	KLINIK DR. ANDRI	564000
CUST55389	KLINIK GM	84000
CUST55694	KLINIK DR. ANDRI	846000
CUST55464	KLINIK GM	1175000
CUST55479	KLINIK GM	189000
CUST55434	KLINIK GM	1598000

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Aggregate Table

```sql
CREATE TABLE total_sales_per_customer AS( -- Menghitung total penjualan per produk
SELECT id_barang, nama_barang, SUM(jumlah_barang) AS total_barang_terjual, SUM(harga_total) AS total_penjualan
FROM salicyl_database
GROUP BY id_barang, nama_barang);
```

id_barang	nama_barang	total_barang_terjual	total_penjualan
BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	840	39480000
BRG0010	PARACETAMOL	822	17262000

Column Name	Datatype
id_barang	VARCHAR(10)
nama_barang	VARCHAR(50)
total_barang_terjual	INT
total_penjualan	INT

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Aggregate Table

```sql
CREATE TABLE total_sales_per_branch AS( -- Menghitung total penjualan per cabang sales
SELECT cabang_sales, SUM(harga_total) AS total_penjualan
FROM salicyl_database
GROUP BY cabang_sales);
```

cabang_sales	total_penjualan
Tangerang	17863000
Jakarta	28022000
Lampung	10857000

Column Name	Datatype
cabang_sales	VARCHAR(10)
total_penjualan	INT

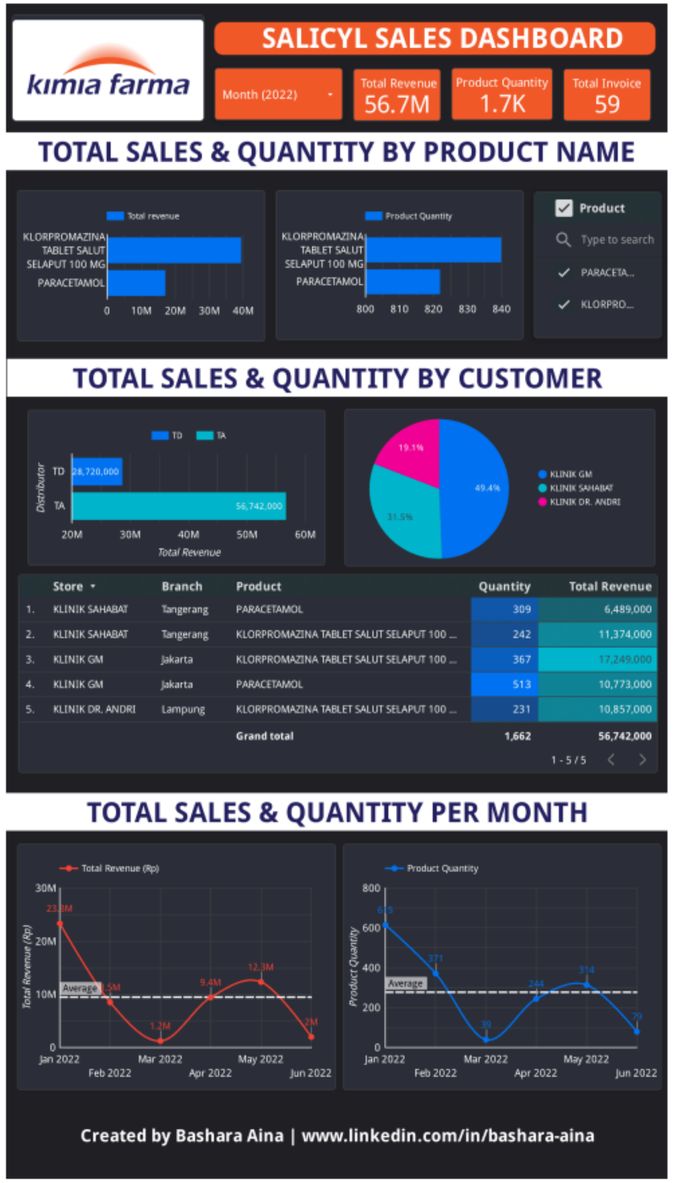
Rakamin Academy

List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Data Visualization (Dashboard)

Overview Visualization

This dashboard presents a comprehensive overview of the sales performance of Kimia Farma's brand Salicyl for the period of January to June 2022. The dashboard provides various summary metrics, including Total Revenue, Product Quantity, and Total Invoice, which can be filtered by each month or product to gain deeper insights. The dashboard is designed to cover several important aspects, including Total Sales & Quantity by Product Name, Total Sales & Quantity by Customer, and Total Sales & Quantity per month. Each aspect is presented with different data visualizations, such as bar chart, pie chart, line chart, table, etc., to provide users with a more holistic understanding of the sales performance of the brand Salicyl. This dashboard serves as a valuable tool for decision-makers to make informed decisions and optimize their sales strategies.



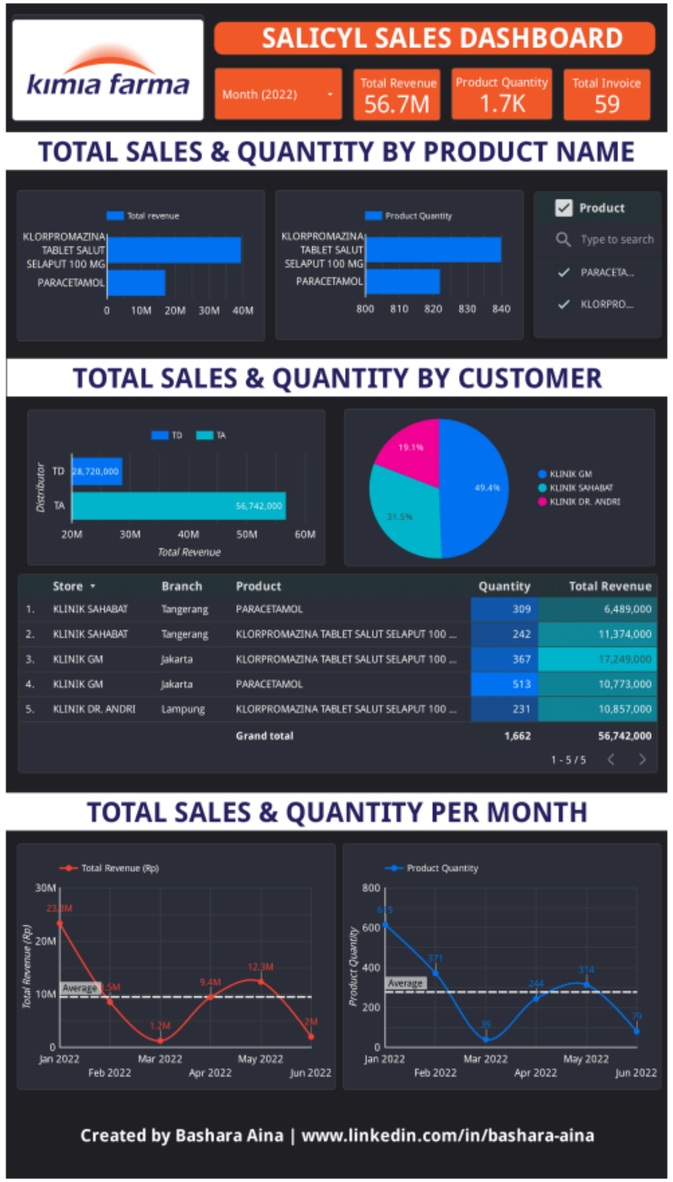
Visualization link: https://intip.in/VisualizationFPKF

Data Visualization (Dashboard)

- List of Content
- Background
- Query
- Primary Key
- Datamart Design
- Base Table
- Aggregate Table
- Data Visualization
- Complementary

Insight

Based on the analysis, it is observed that Klorpromazina has twice the sales of Paracetamol with a total revenue of Rp39.480.000 compared to Rp17.262.000. The Salicyl brand distributor TA has contributed significantly to the brand's sales, surpassing distributor TD by half of the sales. Klinik GM dominates the sales of Brand Salicyl with a percentage of 49.4%, followed by Klinik Sahabat with 31.5%, and Klinik DR. Andri with 19.1%. In the line chart representing sales and quantity per month, it can be seen that the sales of the Salicyl brand are fluctuating. There is a significant drop in revenue from 23M and 615 product quantity sold to 2M revenue and 79 products sold.



Visualization link: https://intip.in/VisualizationFPKF

kimia farma
List of Content
Background
Query
Primary Key
Datamart Design
Base Table
Aggregate Table
Data Visualization
Complementary

Additional Complementary

1. The dataset for customer data would benefit from additional information such as address, email, and phone number. This would enable the company to investigate any problems in the future and offer promotions.
2. Product quantity (stocks) can be adjusted based on weather conditions or health data in certain areas. It's important to focus on areas where diseases are prevalent and ensure that the right product needs are available and that no products accumulate and expire.
3. To enhance the analysis results, it's important to consider the distributor with the most complete data. Additional data on types of sales such as profits, losses, discounts/promotions, price increases and decreases would also be beneficial.
4. For more competitive analysis, data from other data providers who are competitors can be examined.
5. If there is user/buyer data available, a segmentation process can be carried out to further refine the analysis.

THANK YOU

LinkedIn linkedin.com/in/bashara-aina/