



# Project-Based Intern: Big Data Analytics

## PT. Kimia Farma Tbk. X Rakamin Academy

By : Moh. Harwin Prayoga



# Moh. Harwin Prayoga

## About Me

---

Yoga is an Engineering Physics graduate at Institut Teknologi Sepuluh Nopember that passionate about Data Science and Data Analysis. He joined a data science bootcamp to forge his programming skills in Python and RDBMS querying with SQL and gratefully his final project team got awarded as The Winner of Best Final Project Team. With Data Science skills and experience, he believes that he is able to deliver strategic insights and recommendations through data to achieve company goals.



**Reach me at:**

[LinkedIn](#) | [GitHub](#) | [Medium](#)

# Outline

- About This Project
- Query\_1
- Query\_2
- Soal 3 : Menentukan Primary Key
- Soal 4 : Desain Datamart
- Base Table
- Aggregate Table
- Data Visualization (Dashboard)
- Additional Complementary Data



# About This Project



Proyek ini dibuat dalam rangka memenuhi tugas akhir dari ***Project-Based Intern: Big Data Analytics Kimia Farma Tbk. X Rakamin Academy*** yang dimana saya berperan sebagai ***Big Data Analyst*** yang ditugaskan untuk membuat visualisasi berupa Dashboard dengan menggunakan dataset yang telah disediakan oleh perusahaan.

# Query 1

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) SELECT \* FROM pelanggan WHERE alamat LIKE 'Mat%'

Alasan :

Performa yang lebih baik, karena klausa LIKE bekerja dengan mencari index tertentu yang kita cari secara langsung yang dimana pada kasus ini kolom alamat yang diawali dengan "Mat".

# Query 2

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) `SELECT * FROM pelanggan WHERE tanggal_lahir BETWEEN '2000-01-01' AND '2008-12-31'`

Alasan :

Klausula BETWEEN sudah dikenal lebih luas oleh pengguna sehingga query akan lebih mudah dipahami, selain itu klausula BETWEEN juga memanfaatkan index pada kolom tanggal\_lahir sehingga optimasi query lebih optimal.

# Soal 3: Menentukan Primary Key

1. Tentukan primary key dari table penjualan. jelaskan alasannya
2. Jawaban & Penjelasan :

“**id\_invoice**”

**Penjelasan :** kolom “**id\_invoice**” disini menjelaskan kode faktur yang dimana masing-masing kode memiliki nilai unik pada tiap transaksinya serta tidak dipengaruhi oleh faktor dari kolom / tabel lain. Sehingga, kolom “**id\_invoice**” dapat menjadi kolom *primary key* yang stabil.

# Soal 4: Design Datamart

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 screenshot query nya (jika lebih dari 1 file, maka masing masing file di-screenshot)

Jawaban : .....

No.	Nama File	Link
1	Base_Table	<a href="https://github.com/mpythree/vix_kimiafarm_a_bda/blob/main/Base%20Table.sql">https://github.com/mpythree/vix_kimiafarm_a_bda/blob/main/Base%20Table.sql</a>
2	Aggregate_Table	<a href="https://github.com/mpythree/vix_kimiafarm_a_bda/blob/main/Aggregate%20Table.sql">https://github.com/mpythree/vix_kimiafarm_a_bda/blob/main/Aggregate%20Table.sql</a>

# Base Table

```
1   -- Membuat base table--  
2  
3   -----  
4 • CREATE TABLE base_table (  
5       SELECT  
6           pj.id_distributor, pj.id_cabang, pj.id_invoice,  
7           pj.tanggal, pj.id_customer, pj.id_barang,  
8           pj.jumlah_barang, pj.unit, pj.harga,  
9           pj.mata_uang, pj.brand_id, pj.lini,  
10          pl.level, pl.nama, pl.id_cabang_sales,  
11          pl.cabang_sales, pl.id_group, pl.group,  
12          b.sektor, b.nama_barang, b.tipe,  
13          b.nama_tipe, b.kode_lini, b.kemasan  
14      FROM penjualan pj  
15      LEFT JOIN barang b  
16          ON pj.id_barang = b.kode_barang  
17      LEFT JOIN pelanggan pl  
18          ON pj.id_customer = pl.id_customer  
19      );  
20      -- Menentukan primary key  
21 •     ALTER TABLE base_table ADD PRIMARY KEY (id_invoice);
```

# <Base Table>

Column	Data type	Description	Transformation
id_distributor	text	id unik distributor	
id_cabang	text	id unik cabang	
id_invoice	text	id unik faktur	
tanggal	text	tanggal transaksi	konversi format menjadi datetime
id_customer	text	id unik pelanggan	
id_barang	text	id unik barang	
jumlah_barang	int	Jumlah produk terjual	
unit	text	Unit ukuran produk terjual	

# <Base Table>

Column	Data type	Description	Transformation
harga	double	harga per unit item terjual	
mata_uang	text	mata uang transaksi	
brand_id	text	id unik dari merek produk terjual	
lini	text	nama merek produk terjual	
level	text	tingkat pelanggan	
nama	text	nama pelanggan	
id_cabang_sales	text	id unik cabang sales	
cabang_sales	text	lokasi cabang sales	

# <Base Table>

Column	Data type	Description	Transformation
id_group	text	id unik grup pelanggan	
group	text	grup pelanggan	
sektor	text	sektor barang	
nama_barang	text	nama produk terjual	
tipe	text	tipe produk terjual	
nama_tipe	text	nama tipe produk terjual	
kode_lini	text	kode nama barang	
kemasan	text	kemasan produk terjual	

# Aggregate Table

```
1 •-----  
2      SELECT  
3          EXTRACT(MONTH FROM tanggal) bulan_ke,  
4          MONTHNAME(tanggal) bulan,  
5          cabang_sales,  
6          bt.group,  
7          nama nama_customer,  
8          lini,  
9          SUM(jumlah_barang * harga) total_harga  
10         FROM base_table bt  
11         GROUP BY 1,2,3,4,5,6  
12         ORDER BY 1
```

# <Aggregate Table>

Column	Data type	Description	Transformation
bulan_ke	int	Urutan bulan	mengambil nomor bulan dari tanggal
bulan	varchar(9)	nama bulan	mengambil nama bulan dari tanggal
cabang sales	text	lokasi cabang sales	
group	text	nama grup pelanggan	
nama_customer	text	nama pelanggan	
lini	text	merek produk terjual	
total_harga	double	total pendapatan dari transaksi	menjumlahkan harga dan jumlah barang terjual

# Data Visualization (Dashboard)

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

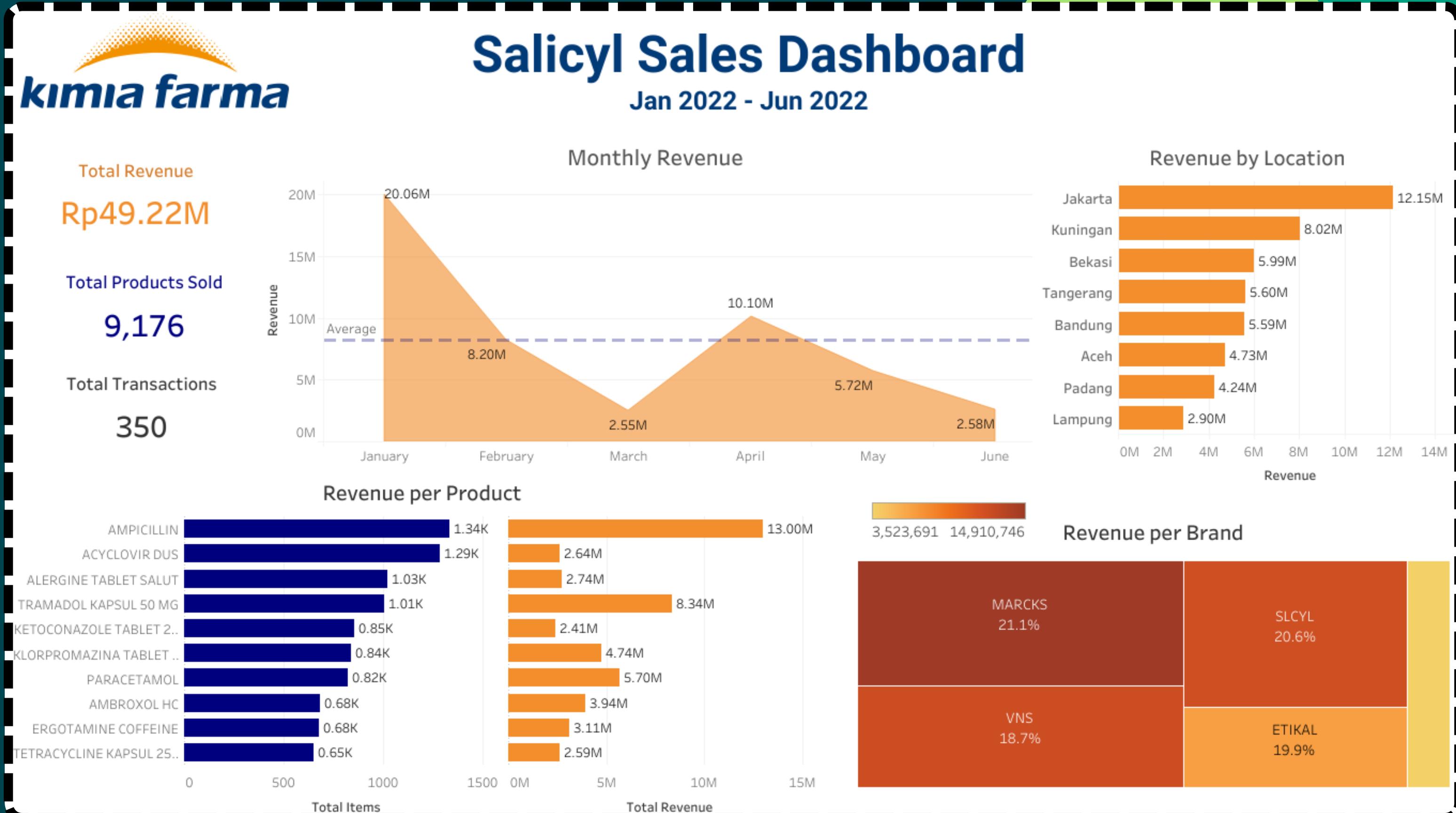
Silahkan tambah halaman jika dibutuhkan

Jawaban :

Link visualisasi (Tableau) :

<https://public.tableau.com/app/profile/moh.harwin.prayoga/viz/PBIBDAKimiaFarmaDashboard/SalesDashboard?publish=yes>

# Dashboard Visualization



# Additional Complementary Data

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 :

Iya, karena dibutuhkan data lain untuk mencari informasi dan pola yang lebih spesifik.

- Data promosi : untuk mengetahui apakah pelanggan tersebut membeli melalui promosi yang diterima atau tidak untuk mengetahui efektifitas berbagai promosi yang dilakukan oleh perusahaan (PT Kimia Farma Tbk.)
- Data demografi pelanggan : untuk mengetahui karakteristik (usia, jenis kelamin, pendapatan, dsb) serta alasan pembelian dari pelanggan yang membeli produk spesifik yang telah disalurkan oleh sales melalui gerai pelanggan.
- Data kepuasan pelanggan : untuk mengetahui seberapa puas pelanggan terhadap produk yang dijual. Dengan mengetahui kepuasan pelanggan, perusahaan dapat mengevaluasi serta merancang strategi perbaikan atau inovasi untuk kedepannya.

# Thank You!

Please reach me at:

- [LinkedIn](#)
- [GitHub](#)
- [Medium](#)

This portfolio  
[GitHub \(portfolio\)](#)

