SQL & ETL

Kimia Farma

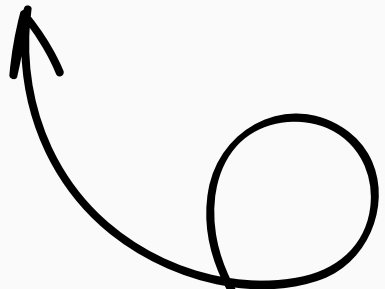By: Raden Bimo

Outline

1. Introduction

- Background Data
- Tools
- Objective

2. Create Database

- Star Schema with SQL using MYSQL
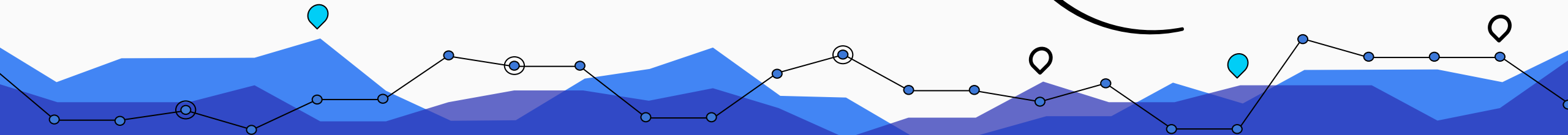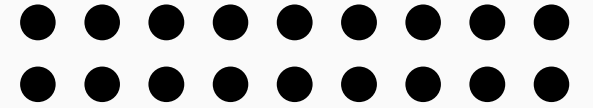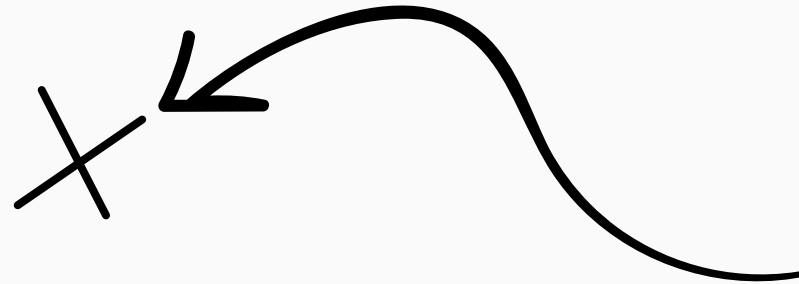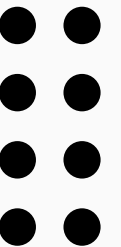- Snow Flake Schema with ETL process using Talend

3. Data Visualization

- Dashboard
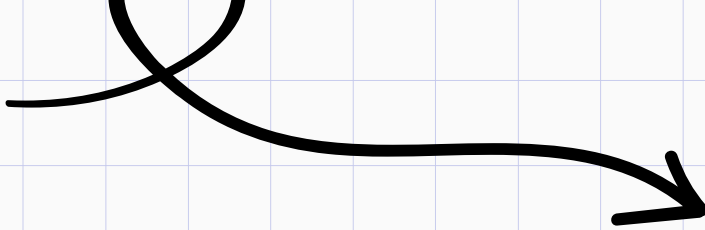- Component Visualization purpose

01 Introduction

- Data Background
- Tools
- Objective

Data Background

This data set is a dummy data set from Kimia Farma, one of the Indonesia pharmaceutical producers and distributors. One file contains three sheets, with each sheet having one table. However, before processing it, I split every sheet into a different file.
The sheets' content:
1. Pelanggan: Contain customer information
2. Produk: Contain product information
3. Penjualan: Contain sales information

Example 5 rows on each sheet

Product

kode_barang	sektor	nama_barang	tipe	nama_tipe	kode_lini	lini	kemasan
BRG0001	P	ACYCLOVIR DUS	ZPJ1	Produk jadi	206	OGB & PH	DUS
BRG0002	P	ALERGINE TABLET SALUT	ZPJ1	Produk jadi	203	ETIKAL	DUS
BRG0003	P	AMPICILLIN	ZPJ1	Produk jadi	210	MARCKS	BOTOL
BRG0004	P	TRAMADOL KAPSUL 50 MG	ZPJ1	Produk jadi	201	VNS	KAPSUL

Pelanggan

id_customer	level	nama	id_cabang_sales	cabang_sales	id_group	group
CUST55380	Company	APOTEK TAPAK	CAB01	Aceh	Z32	Apotek
CUST55381	Company	APOTEK MAJA	CAB02	Kuningan	Z32	Apotek
CUST55382	Company	KLINIK GM	CAB03	Jakarta	Z31	Klinik
CUST55383	Company	APOTEK MERDEKA	CAB04	Bandung	Z32	Apotek

Penjualan

id_distributor	id_cabang	id_invoice	tanggal	id_customer	id_barang	jumlah_barang	unit	harga	mata_uang	brand_id	lini
TD	CAB01	IN5997	20/01/22	CUST55380	BRG0001	1	DUS	1169,91	IDR	BRND001	OGB & PH
TD	CAB01	IN6297	20/01/22	CUST55381	BRG0002	5	DUS	2337,5	IDR	BRND002	ETIKAL
TA	CAB02	IN6155	21/01/22	CUST55382	BRG0003	9	DUS	10690,6	IDR	BRND003	MARCKS
EPM	CAB03	IN6144	22/01/22	CUST55383	BRG0004	13	DUS	8700,7	IDR	BRND004	VNS
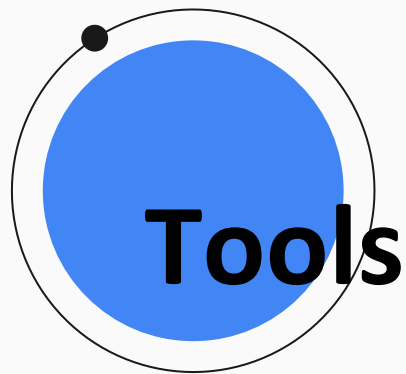
Tools

Tableau
- Data visualization tools

MySQL
- Data Base Management tools
- Create local database
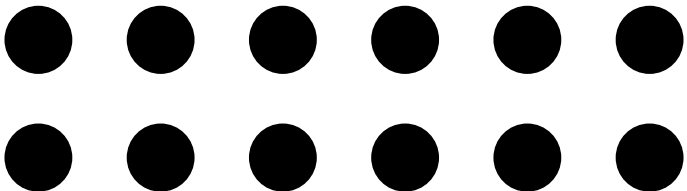- Create Entity Relationship Diagram (ERD)

Talend
- Extract, Tranform, and Load (ETL) tools
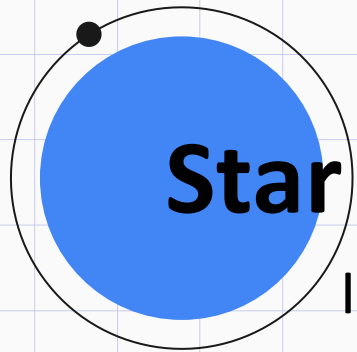
Objective

Project have 2 main objective:

1. **Create Database**
   a. Star Schema with SQL using MYSQL
   b. Snow Flake Schema with ETL process using Talend
2. **Visualize the data using Tableau**

02 Create Database

- Star Schema with SQL using MYSQL
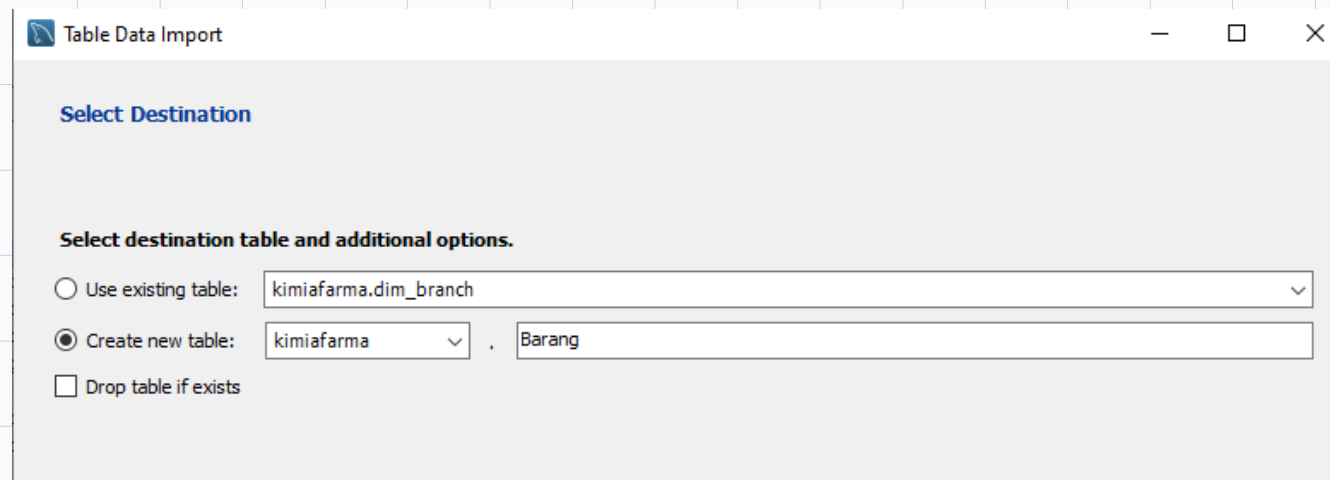- Snow Flake Schema with ETL process using Talend
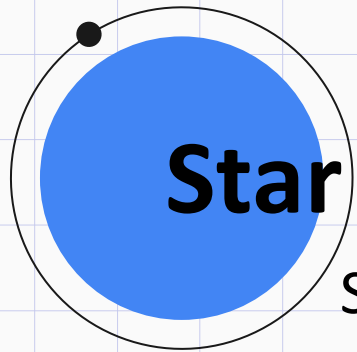
Star Schema with SQL using MYSQL

Importing Data

There are several ways for import data into database using MYSQL, such as using:
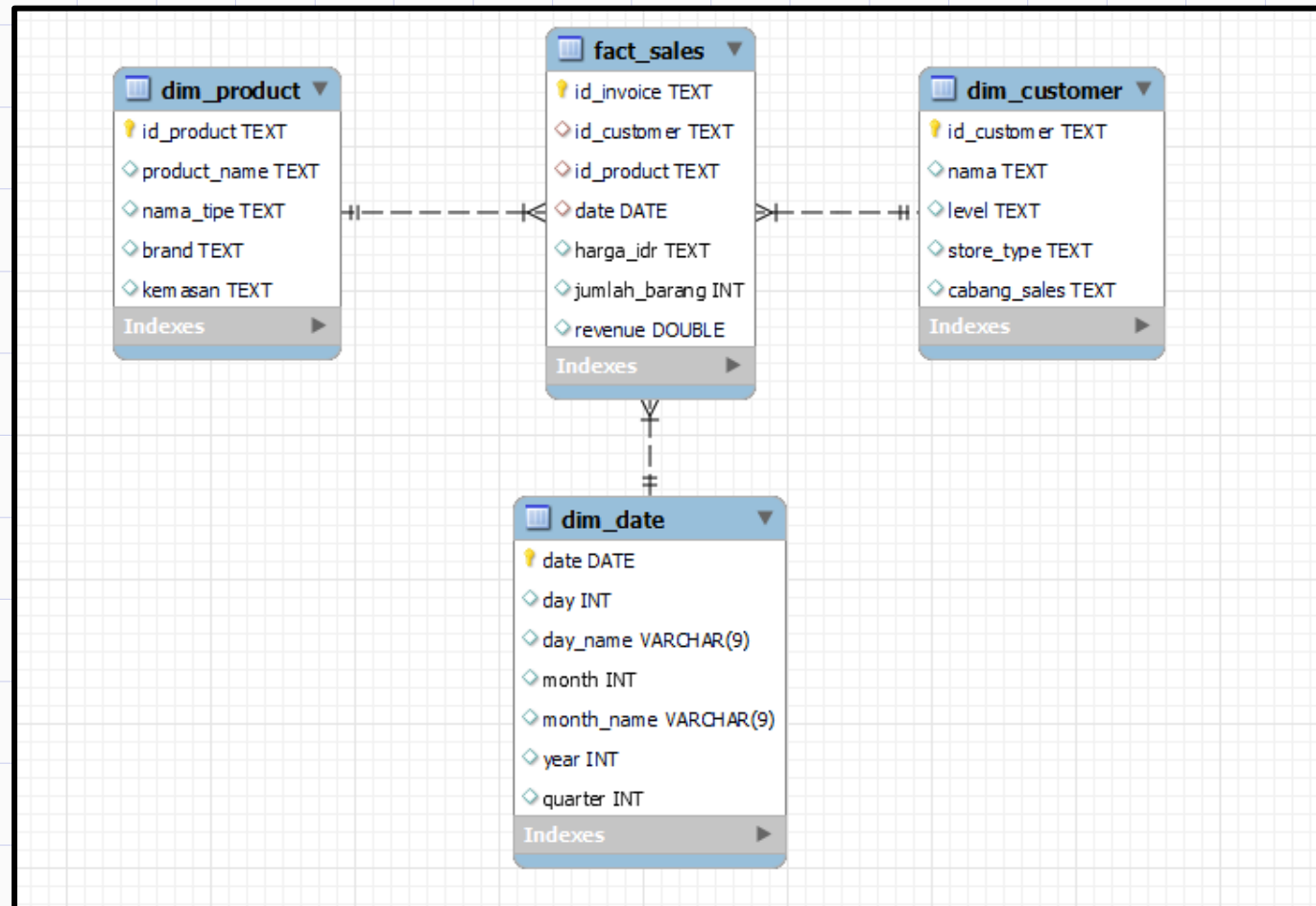
- Command Line
- MySQL GUI
- Etc.

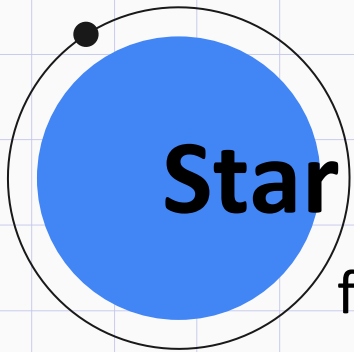However this project we will using MySQL GUI Workbench

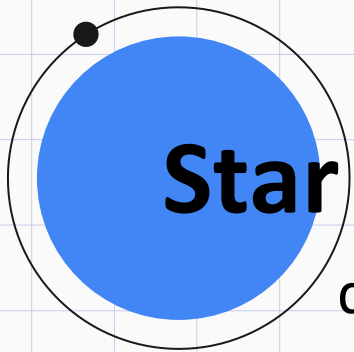# Star Schema with SQL using MYSQL

Star Schema

Star Schema with SQL using MYSQL

fact_sales

Query

```sql
1   CREATE TABLE fact_sales
2   SELECT
3       id_invoice,
4       id_customer,
5       id_barang AS id_product,
6       str_to_date(tanggal,'%d/%m/%y') AS date,
7       harga AS harga_idr,
8       jumlah_barang,
9       (harga*jumlah_barang) AS revenue
10  FROM kimia_farma.penjualan
11  WHERE
12      IFNULL(harga,'')!= ''
13      AND IFNULL(jumlah_barang,'')!= '';
```

Result

id_invoice	id_customer	id_product	date	harga_idr	jumlah_barang	revenue
IN5997	CUST55380	BRG0001	2022-01-20	1169,91	1	1169
IN6297	CUST55381	BRG0002	2022-01-20	2337,5	5	11685
IN6155	CUST55382	BRG0003	2022-01-21	10690,6	9	96210
IN6144	CUST55383	BRG0004	2022-01-22	8700,7	13	113100
IN6280	CUST55384	BRG0005	2022-01-23	5648,3	1	5648

Star Schema with SQL using MYSQL

dim_product

Query

```sql
1   CREATE TABLE dim_product
2   SELECT
3       kode_barang AS id_product,
4       nama_barang AS product_name,
5       nama_tipe,
6       lini AS brand,
7       kemasan
8   FROM kimia_farma.barang
9   WHERE
10      IFNULL(kode_barang,'')!= ''
11      AND IFNULL(nama_barang,'')!= '';
```

Result

kode_barang	sektor	nama_barang	tipe	nama_tipe	kode_lini	lini	kemasan
BRG0001	P	ACYCLOVIR DUS	ZPJ1	Produk jadi	206	OGB & PH	DUS
BRG0002	P	ALERGINE TABLET SALUT	ZPJ1	Produk jadi	203	ETIKAL	DUS
BRG0003	P	AMPICILLIN	ZPJ1	Produk jadi	210	MARCKS	BOTOL
BRG0004	P	TRAMADOL KAPSUL 50 MG	ZPJ1	Produk jadi	201	VNS	KAPSUL
BRG0005	P	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	ZPJ1	Produk jadi	203	SLCYL	TABLET
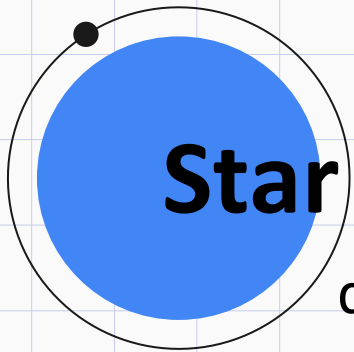
Star Schema with SQL using MYSQL

dim_customer

Query

```sql
1 ●  CREATE TABLE dim_customer
2    SELECT
3        id_customer,
4        nama,
5        level,
6        p.group AS store_type,
7        cabang_sales
8    FROM pelanggan p
9    WHERE
10       IFNULL(id_customer,'')!= ''
11       AND IFNULL(nama,'')!= ''
```

Result

id_customer	nama	level	store_type	cabang_sales
CUST55380	APOTEK TAPAK	Company	Apotek	Aceh
CUST55381	APOTEK MAJA	Company	Apotek	Kuningan
CUST55382	KLINIK GM	Company	Klinik	Jakarta
CUST55383	APOTEK MERDEKA	Company	Apotek	Bandung
CUST55384	KLINIK SAHABAT	Company	Klinik	Tangerang
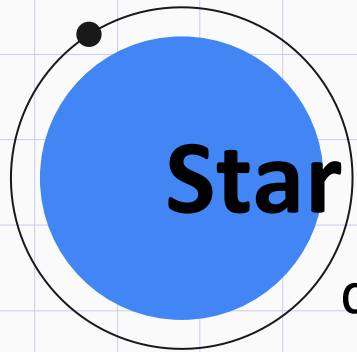
# Star Schema with SQL using MYSQL

dim_date

Query

```sql
1   CREATE TABLE dim_date
2   WITH waktu AS (SELECT distinct str_to_date(tanggal,'%d/%m/%y') as tanggal
3                  FROM penjualan
4                  WHERE IFNULL(tanggal,'')!= '')
5   SELECT
6       tanggal AS date,
7       day(tanggal) AS day,
8       dayname(tanggal) AS day_name,
9       month(tanggal) AS month,
10      monthname(tanggal) AS month_name,
11      year(tanggal) AS year,
12      quarter(tanggal) AS quarter
13  FROM waktu
14  ORDER BY date;
```
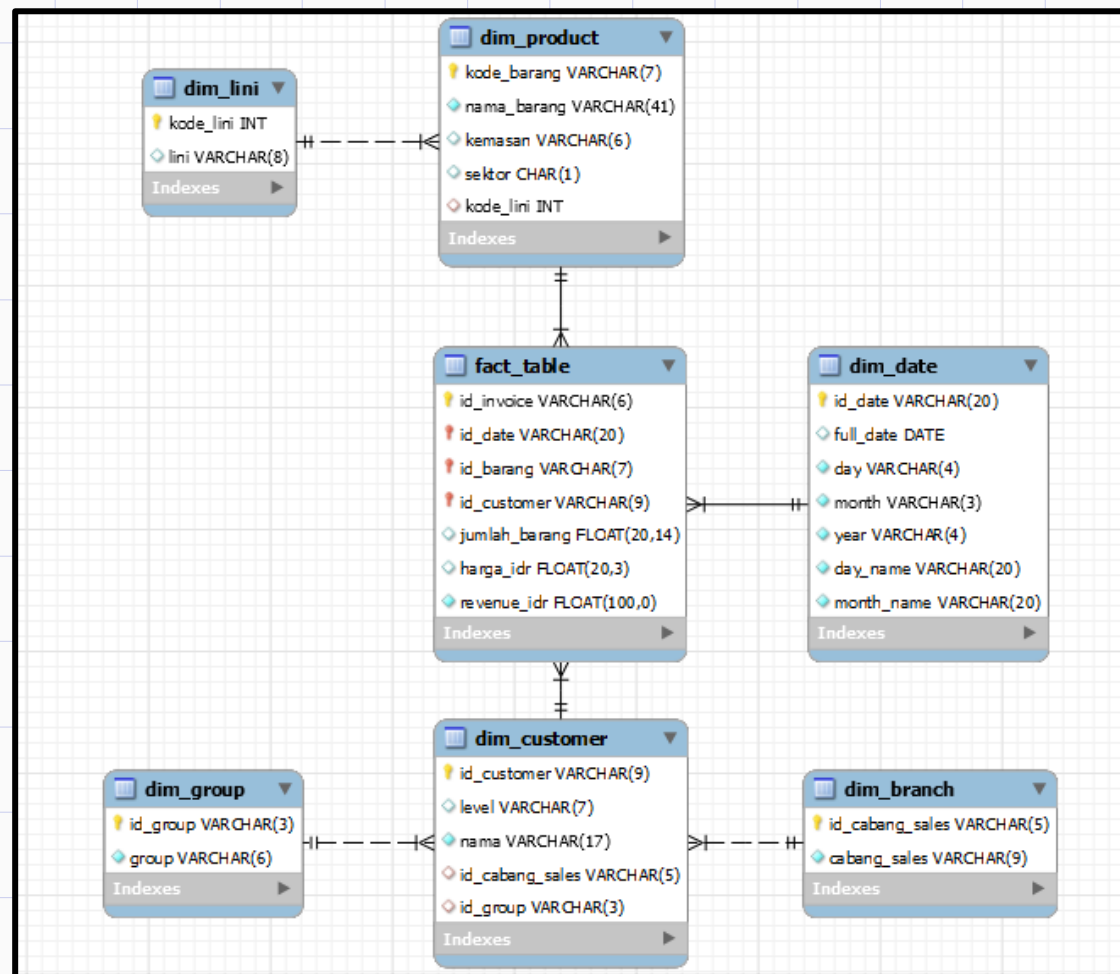
Result

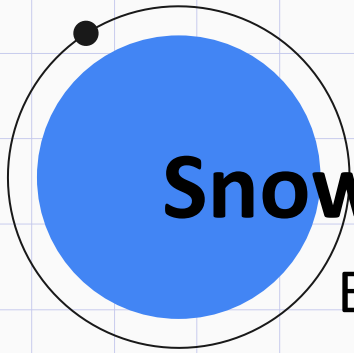| date	day	day_name	month	month_name	year	quarter
2022-01-20	20	Thursday	1	January	2022	1
2022-01-21	21	Friday	1	January	2022	1
2022-01-22	22	Saturday	1	January	2022	1
2022-01-23	23	Sunday	1	January	2022	1
2022-01-25	25	Tuesday	1	January	2022	1
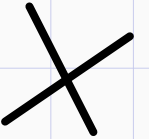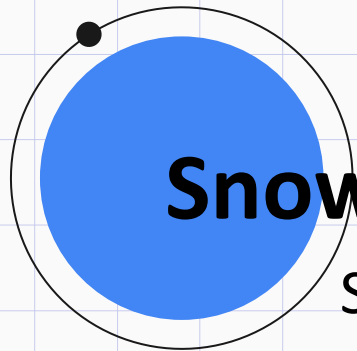
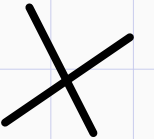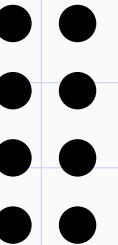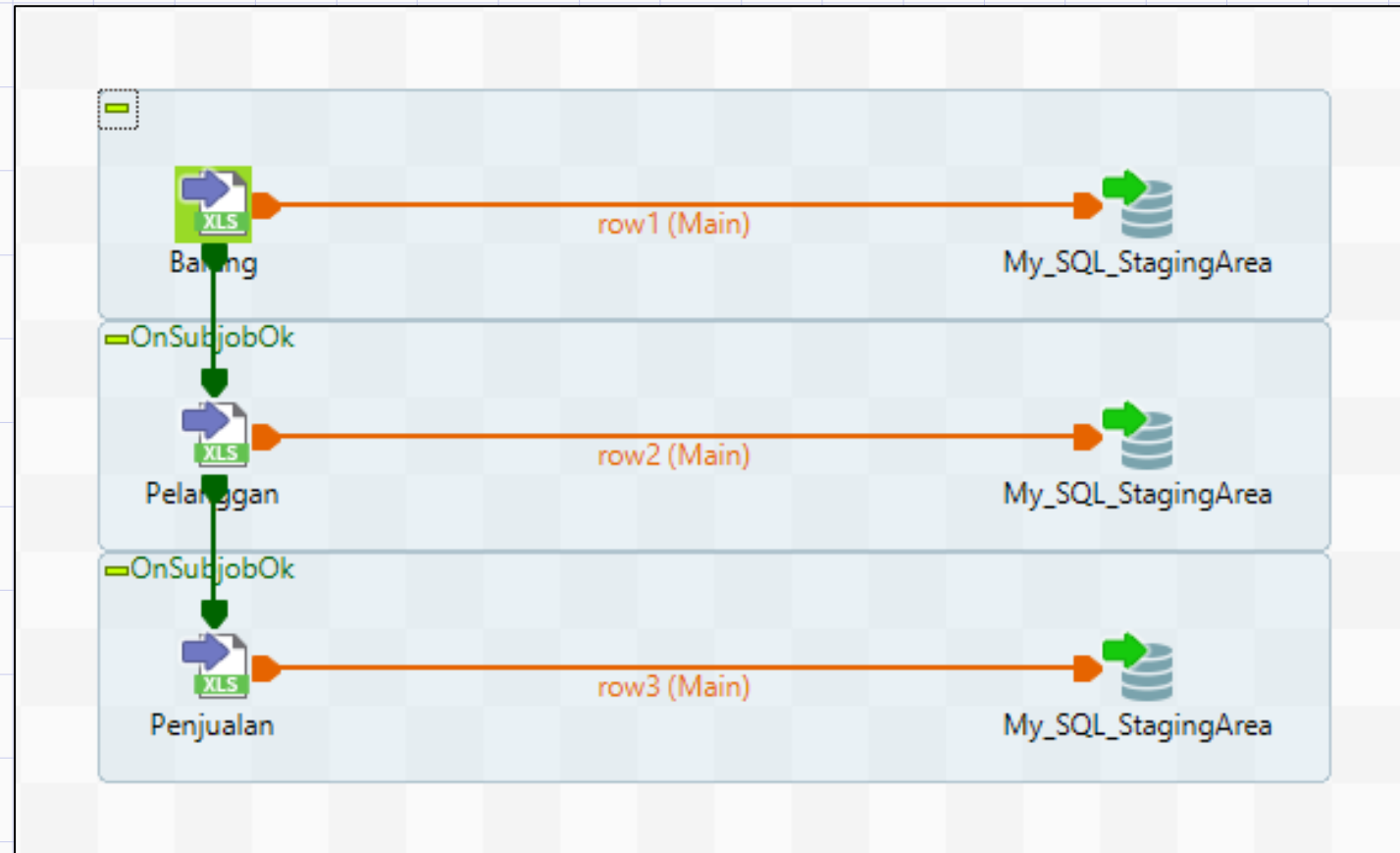# Snow Flake Schema with ETL process using Talend

Snow Flake Schema
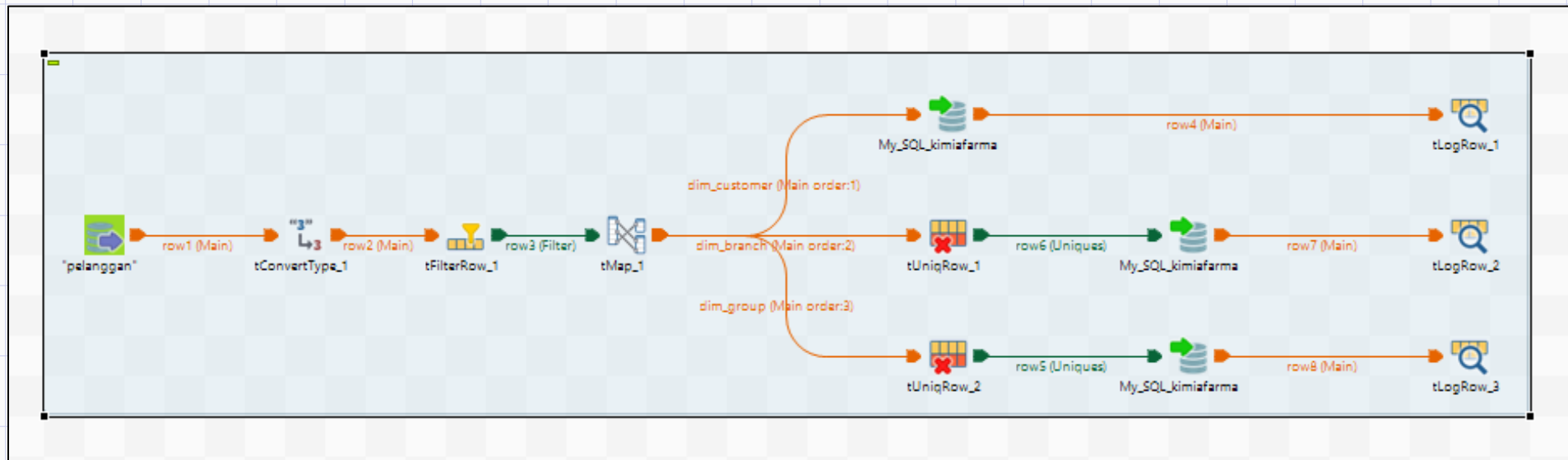
Snow Flake Schema with ETL process using Talend

Extract into Staging Area

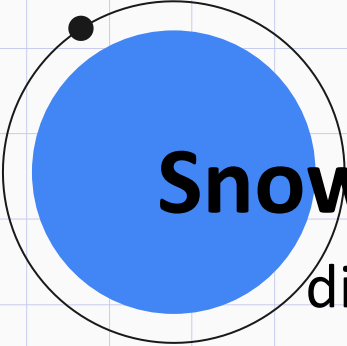# Snow Flake Schema with ETL process using Talend

dim_customer & dim_branch

Job

Result

dim_customer

	id_customer	level	nama	id_cabang_sales	id_group
▶	CUST55380	Company	APOTEK TAPAK	CAB01	Z32
	CUST55381	Company	APOTEK MAJA	CAB02	Z32
	CUST55382	Company	KLINIK GM	CAB03	Z31
	CUST55383	Company	APOTEK MERDEKA	CAB04	Z32
	CUST55384	Company	KLINIK SAHABAT	CAB05	Z31

dim_branch

	id_cabang_sales	cabang_sales
▶	CAB01	Aceh
	CAB02	Kuningan
	CAB03	Jakarta
	CAB04	Bandung
	CAB05	Tangerang

Snow Flake Schema with ETL process using Talend

dim_product & dim_lini

Job
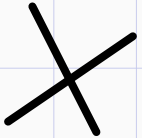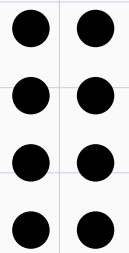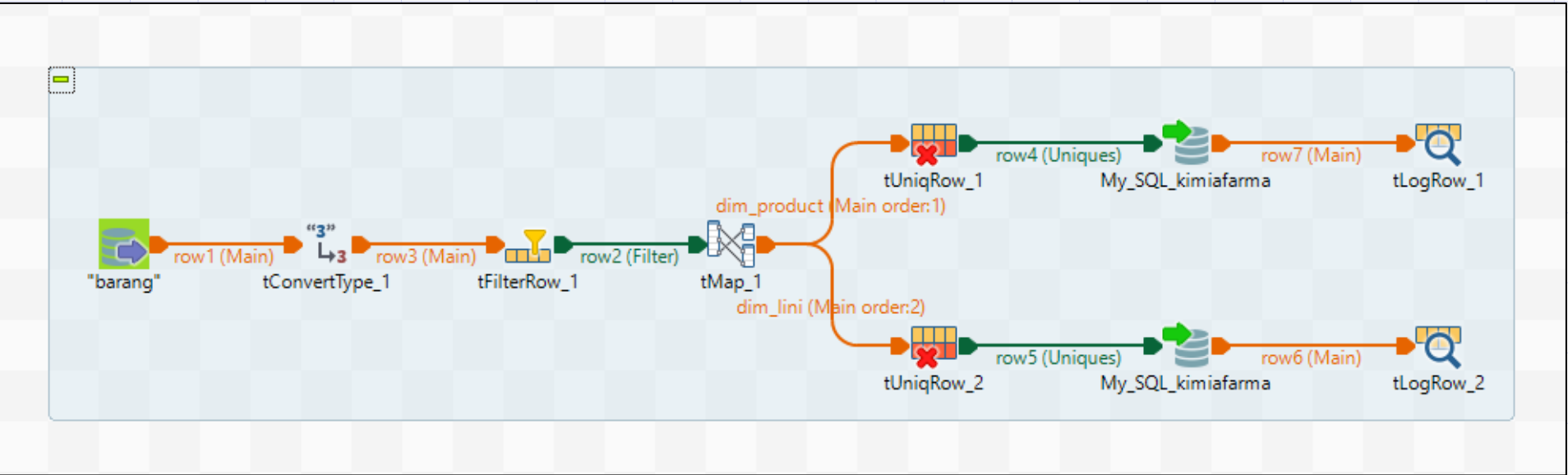
Result

dim_product

	kode_barang	nama_barang	kemasan	sektor	kode_lini
▶	BRG0001	ACYCLOVIR DUS	DUS	P	206
	BRG0002	ALERGINE TABLET SALUT	DUS	P	203
	BRG0003	AMPICILLIN	BOTOL	P	210
	BRG0004	TRAMADOL KAPSUL 50 MG	KAPSUL	P	201
	BRG0005	KLORPROMAZINA TABLET SALUT SELAPUT 100 ...	TABLET	P	203

dim_lini

	kode_lini	lini
▶	201	VNS
	203	ETIKAL
	204	MARCKS
	205	OGB & PH
	206	OGB & PH
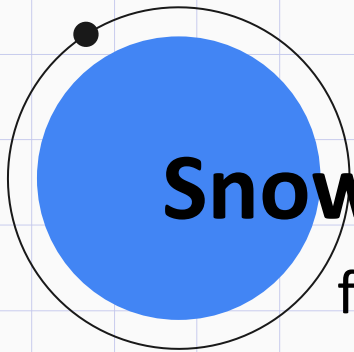
# Snow Flake Schema with ETL process using Talend

fact sales

Job
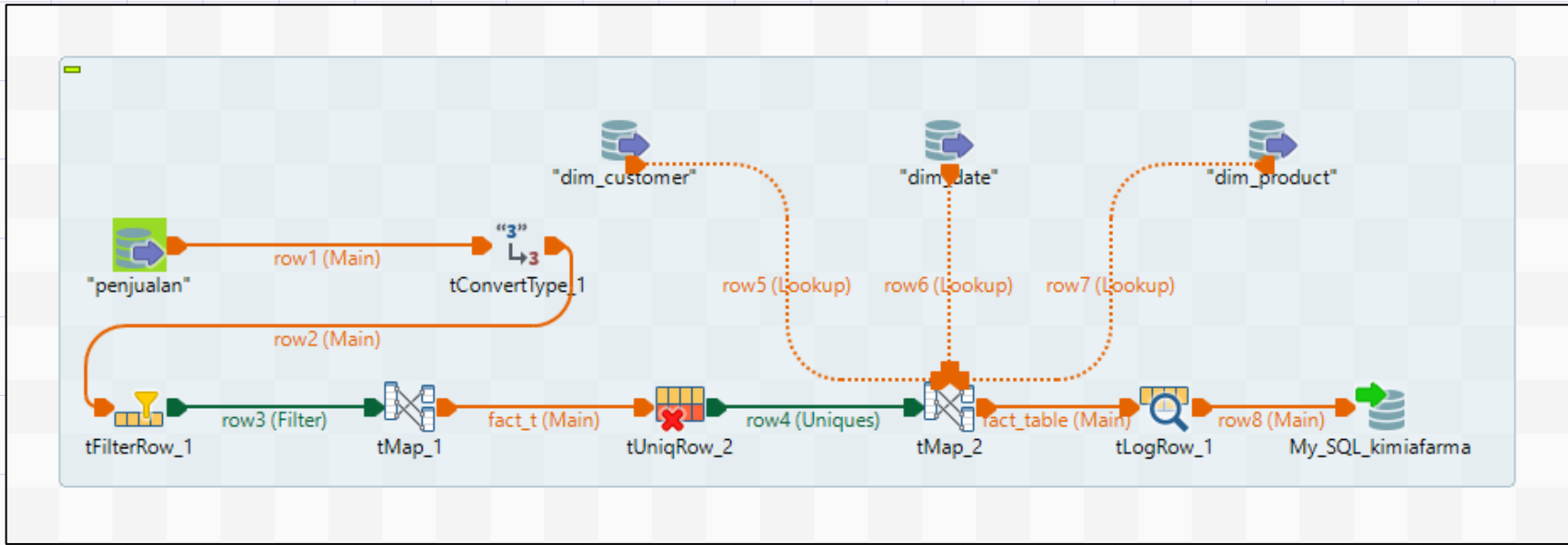
Result

id_invoice	id_date	id_customer	kode_barang	jumlah_barang	harga_idr	revenue_idr
IN5978	t114	CUST55670	BRG0001	9.00000000000000	1169.910	10529.2
IN5979	t65	CUST55598	BRG0009	18.36713218688965	5780.700	106175
IN5980	t4	CUST55432	BRG0003	20.00000000000000	10690.600	213812
IN5981	t7	CUST55443	BRG0004	76.00000000000000	8700.700	661253
IN5982	t8	CUST55535	BRG0006	5.00000000000000	2819.200	14096
IN5983	t89	CUST55622	BRG0003	67.00000000000000	10690.600	716270
IN5984	t99	CUST55714	BRG0005	8.00000000000000	5648.300	45186.4
IN5985	t86	CUST55640	BRG0001	21.89160919189453	1169.910	25611.2

03 Data Visualization

- Dashboard
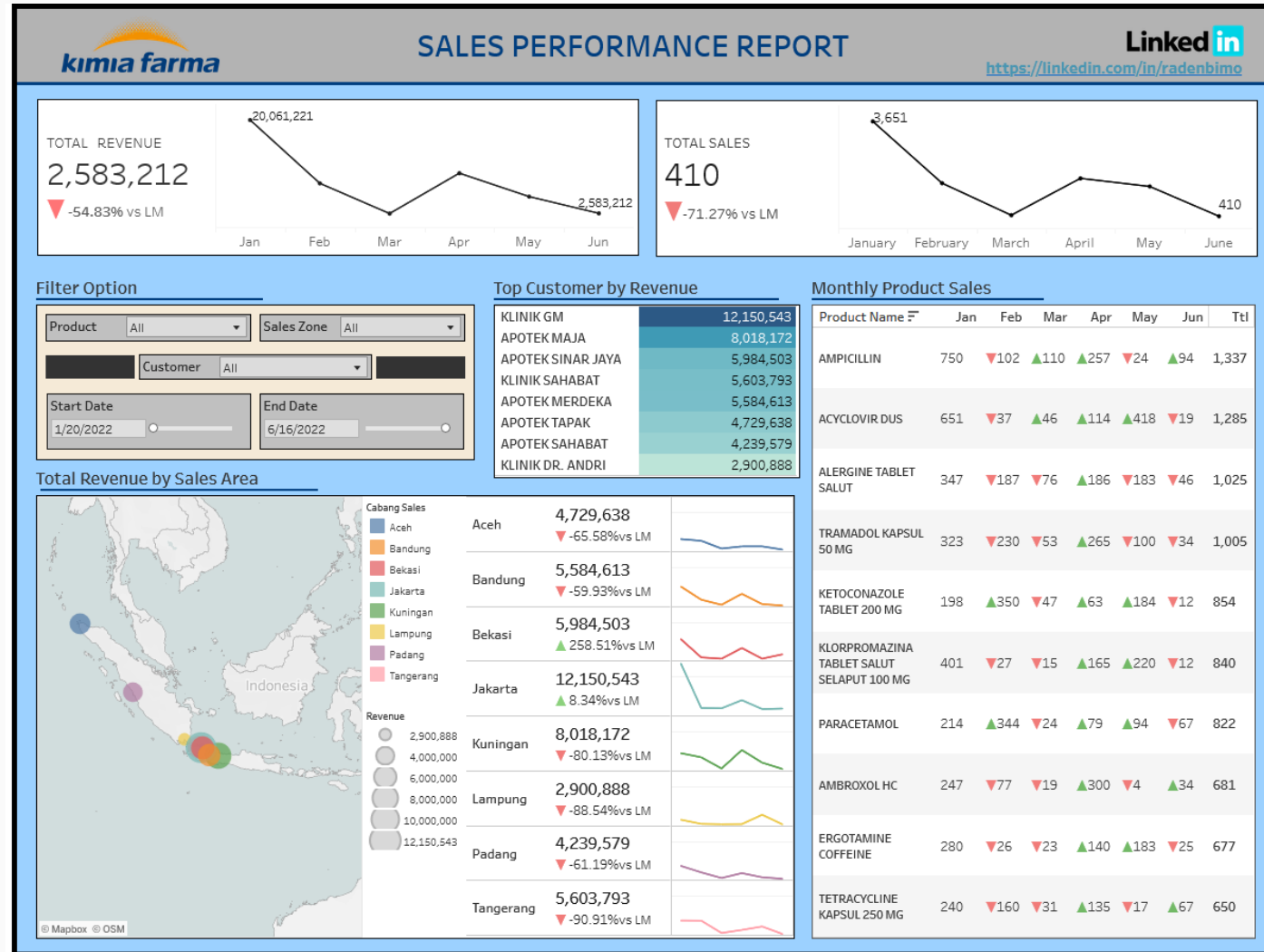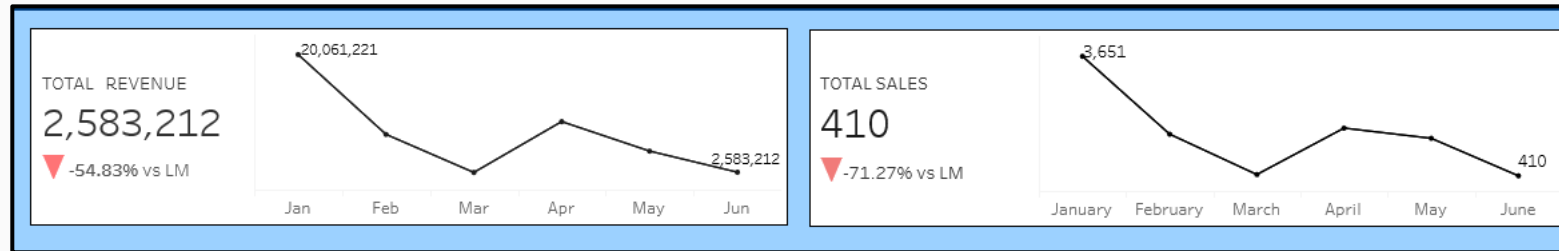- Component Visualization purpose
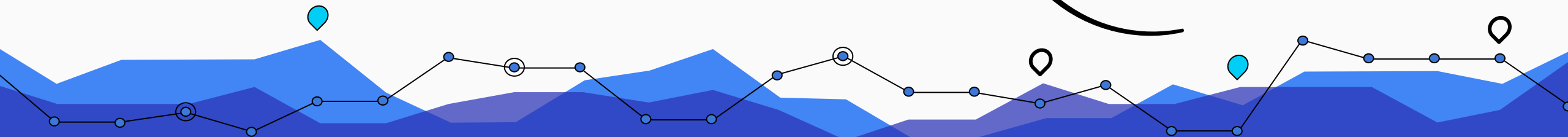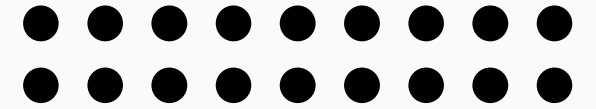
Dashboard

Tableu Public Link

Component Visualization Purpose

- 2 Line charts at top dashboard show the total revenue and sales. the purpose of placement is when someone see the dashboard the first thing they see is sales and performance this month.
- Build with line chart to make it easy to compare performance with previous months.
- Add an automatic feature when the performance is smaller than the previous month it will show a red reverse-triangle (▼), and if it increase it will show green triangle(▲).
- Add an automatic feature that calculates the percentage difference from the current month's performance compared to last month.

Tableu Public Link

Component Visualization Purpose

Filter Option

| Product | All | Sales Zone | All |

| Customer | All |

| Start Date | End Date |
| 1/20/2022 | 6/16/2022 |

Top Customer by Revenue

| KLINIK GM | 12,150,543 |
APOTEK MAJA	8,018,172
APOTEK SINAR JAYA	5,984,503
KLINIK SAHABAT	5,603,793
APOTEK MERDEKA	5,584,613
APOTEK TAPAK	4,729,638
APOTEK SAHABAT	4,239,579
KLINIK DR. ANDRI	2,900,888

- Add a filter option that allows the user to filter the time range to look at, filter customers, choose a sales zone, or look at a specific product.

- Small table to show the top 10 customer who contribute the highest revenue to company

Tableu Public Link

Component Visualization Purpose
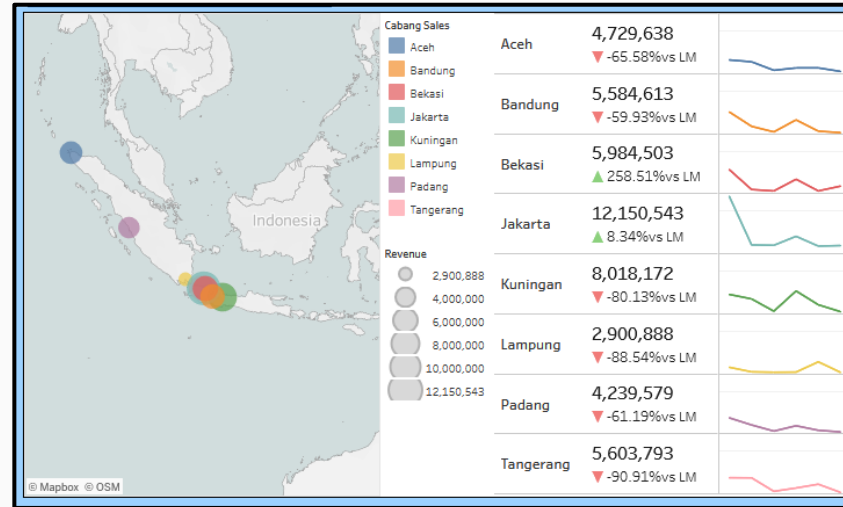
- Made with a map chart so that users can easily find out where the sales located and the amount of revenue generated compare with other location.
- Add an automatic feature when the performance is smaller than the previous month it will show a red reverse-triangle (▼), and if it increase it will show green triangle(▲).
- Every location has detailed current month revenue as well as an automatic percentage revenue compare feature against the previous month, giving users a better understanding of sales performance in specific locations.
- Add Line chart at right side so User have rough understanding what's happening at previous months.

Tableu Public Link

Component Visualization Purpose

Monthly Product Sales

Product Name	Jan	Feb	Mar	Apr	May	Jun	Ttl
AMPICILLIN	750	▼102	▲110	▲257	▼24	▲94	1,337
ACYCLOVIR DUS	651	▼37	▲46	▲114	▲418	▼19	1,285
ALERGINE TABLET SALUT	347	▼187	▼76	▲186	▼183	▼46	1,025

- This table is dedicated to displaying detailed information on the total number of products sold each month. so that users can better understand which product affect company revenue.
- Add an automatic feature when the performance is smaller than the previous month it will show a red reverse-triangle (▼), and if it increase it will show green triangle(▲).

Tableu Public Link
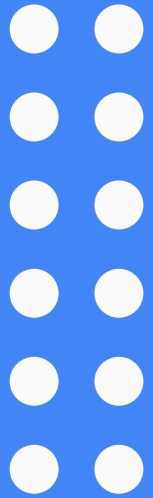
Thank You!

Email
rbimomandalaputra@gmail.com

Linkedin
www.linkedin.com/in/radenbimo

Github
https://github.com/RadenBimo