



Big Data Analytics Virtual Internship Experience

Final Project

Query

Soal 1: Query kedua lebih baik daripada Query pertama

Karena tidak efisien menggunakan fungsi pada column yang difilter. Apabila column alamat tersebut di index, index yang tujuannya untuk mengoptimalkan query jadi tidak berguna karena ada fungsi di dalam column nya

Soal 2: Query kedua lebih baik daripada Query pertama

Karena lebih baik kita menghindari operator AND, OR, & NOT. Apabila column tanggal_lahir tersebut di index, index yang tujuannya untuk mengoptimalkan query jadi tidak berguna karena ada operator tersebut. Sehingga lebih baik kita menggunakan cara lain sebagai penggantinya. Dalam case ini, AND digantikan dengan BETWEEN

Menentukan Primary Key

Pada table penjualan, tidak ada column yang unique sehingga primary key nya perlu di define sendiri.

Caranya dengan menggabungkan id_invoice dan id_barang

```
1 SELECT
2   CONCAT(id_invoice, '-', id_barang) id_penjualan, *
3 FROM   composite-silo-l30101.rakamin.penjualan
```

Run Save query Save view Schedule query More

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.5 sec elapsed, 24 KB processed)

Job information Results JSON Execution details

Row	id_penjualan	id_invoice	tanggal	id_customer	id_barang	jumlah_barang	unit	harga	mata_uang
1	IN5997_BRG0001	IN5997	2022-01-20	CUST55380	BRG0001	1	DUS	96000	IDR
2	IN5997_BRG0002	IN5997	2022-01-20	CUST55380	BRG0002	4	DUS	112000	IDR
3	IN5997_BRG0003	IN5997	2022-01-20	CUST55380	BRG0003	6	BOTOL	17000	IDR
4	IN5997_BRG0004	IN5997	2022-01-20	CUST55380	BRG0004	11	TABLET	24500	IDR
5	IN5997_BRG0005	IN5997	2022-01-20	CUST55380	BRG0005	40	TABLET	47000	IDR
6	IN6297_BRG0002	IN6297	2022-01-20	CUST55381	BRG0002	5	DUS	112000	IDR
7	IN6297_BRG0005	IN6297	2022-01-20	CUST55381	BRG0005	15	TABLET	47000	IDR
8	IN6297_BRG0007	IN6297	2022-01-20	CUST55381	BRG0007	12	BOTOL	64700	IDR
9	IN6297_BRG0008	IN6297	2022-01-20	CUST55381	BRG0008	30	TABLET	9800	IDR

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
 - a. Tabel base merupakan tabel dari hasil penggabungan 3 tabel dengan granularity id_penjualan sebagai Primary Key (yaitu penggabungan antara id_invoice dan id_barang)
 - b. Total row data pada tabel base sama dengan total pada tabel penjualan
 - c. Tabel base disimpan di dalam data warehouse
2. Tabel aggregate
 - a. Tabel aggregate merupakan turunan dari tabel base dimana data dikelompokkan berdasarkan tanggal, id_customer dan id_invoice
 - b. Tabel aggregate disimpan di dalam datamart

Table Base “Penjualan”

```
SELECT
  CONCAT(id_invoice, '_', id_barang) id_penjualan,
  pjl.id_invoice,
  pjl.tanggal,
  pjl.id_barang,
  brg.nama_barang,
  pjl.harga,
  pjl.unit,
  pjl.jumlah_barang,
  (jumlah_barang * pjl.harga) AS total_harga_per_barang,
  pjl.mata_uang,
  brg.kode_brand,
  brg.brand,
  pjl.id_customer,
  plg.nama nama_customer,
  plg.cabang_sales,
  plg.id_distributor,
  plg.group AS group_category
FROM `composite-silo-130101.rakamin.penjualan` pjl
LEFT JOIN `composite-silo-130101.rakamin.barang` brg
ON (pjl.id_barang = brg.kode_barang)
LEFT JOIN `composite-silo-130101.rakamin.pelanggan` plg
ON (pjl.id_customer = plg.id_customer)
```

Table Base “Penjualan”

column	data type	description	transformation
id_penjualan	string	PK dari tabel penjualan	CONCAT(id_invoice, '_', id_barang)
id_invoice	string	id invoice	-
tanggal	date	tanggal transaksi dilakukan	-
id_barang	string	id barang, key join ke tabel barang	-
nama_barang	string	nama barang	-
harga	numeric	harga barang	-
unit	string	kemasan untuk setiap penjualan	-
jumlah_barang	numeric	jumlah barang yang dibeli	-
total_harga_per_barang	numeric	harga barang x jumlah barang	(jumlah_barang * pjl.harga)
mata_uang	string	mata uang. default = IDR	-
kode_brand	string	kode brand dari barang	-

Table Base “Penjualan”

column	data type	description	transformation
brand	string	nama brand	-
id_customer	string	id customer, key join ke tabel pelanggan	-
nama_customer	string	nama pelanggan	-
cabang_sales	string	cabang yang supply barang ke pelanggan	-
id_distributor	string	distributor yang supply barang ke cabang	-
group_category	string	kategori pelanggan, possible value: <ul style="list-style-type: none">- Apotek- Klinik	-

Table Aggregate “Penjualan harian”

```
SELECT
  id_invoice, tanggal,
  id_customer, nama_customer, cabang_sales,
  id_distributor, group_category,
  COUNT(DISTINCT id_barang) total_barang, SUM(total_harga_per_barang) total_pembelian
FROM `composite-silo-130101.datamart_rakamin.penjualan`
GROUP BY 1,2,3,4,5,6,7
ORDER BY 1
```


Table Aggregate “Penjualan Harian”

column	data type	description	transformation
id_invoice	string	PK dari tabel penjualan harian	-
tanggal	date	tanggal transaksi dilakukan	-
id_customer	string	id pelanggan	-
nama_customer	string	nama pelanggan	-
cabang_sales	string	nama cabang	-
id_distributor	numeric	distributor	-
group_category	string	kategori pelanggan, possible value: <ul style="list-style-type: none">- Apotek- Klinik	-
total_barang	numeric	total barang yang dibeli untuk setiap invoice id	COUNT(DISTINCT id_barang) total_barang
total_pembelian	numeric	total harga yang dibayar untuk setiap invoice id	SUM(total_harga_per_barang) total_pembelian

Data Visualization

link: [data viz](#)

Filter Data

1 Jan 2022 - 28 Feb 2022

pelanggan

cabang

total invoice

132

total penjualan

90.637.700

total barang terjual

10

jumlah barang terjual

2.237

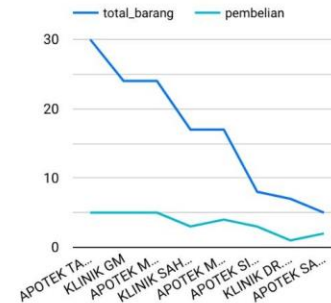
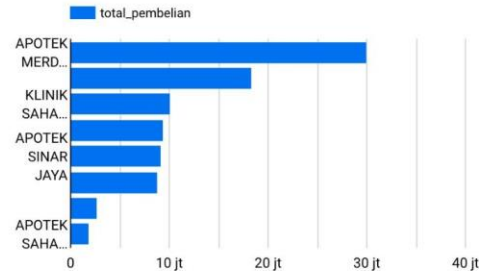
pelanggan aktif

25

Jumlah barang terjual



Trend Pembelian Pelanggan



	tanggal ▾	id_invoice	nama_customer	nama_barang	jumlah_bar...	harga	total harga
1.	2 Feb 2022	IN6064	APOTEK TAPAK	KLORPROMAZINA TABLET S...	24	47000	1128000
2.	2 Feb 2022	IN6064	APOTEK TAPAK	TRAMADOL KAPSUL 50 MG	24	24500	588000
3.	1 Feb 2022	IN6113	KLINIK GM	KLORPROMAZINA TABLET S...	12	47000	564000
4.	1 Feb 2022	IN6301	APOTEK SAHABAT	TETRACYCLINE KAPSUL 250 ...	24	9800	235200
5.	1 Feb 2022	IN6113	KLINIK GM	TRAMADOL KAPSUL 50 MG	12	24500	294000
6.	1 Feb 2022	IN6259	APOTEK MERDEKA	ALERGINE TABLET SALUT	25	112000	2800000
7.	1 Feb 2022	IN6259	APOTEK MERDEKA	KETOCONAZOLE TABLET 20...	19	39000	741000

Additional Complementary Data

1. data latitude longitude untuk setiap lokasi pelanggan
 - a. tujuannya untuk melihat persebaran lokasi pelanggan
 - b. dari data tersebut, kita bisa melihat potential lokasi yang bisa kita approach untuk menjadi pelanggan
 - c. selain itu, kita juga bisa memprediksi demand lokasi 1 ke lokasi sekitar dari kebutuhan obat di lokasi tertentu pada waktu tertentu. misalkan sedang musim pengujian di daerah A sehingga kebutuhan obat di sekitar daerah tersebut terkait dengan obat flu atau demam. dan kita bisa mengalokasikan stok obat pada distributor / cabang tertentu terhadap demand prediction tersebut
2. data cuaca
 - a. untuk melihat korelasi antara kebutuhan obat dengan cuaca di daerah tersebut