



## **UJIAN AKHIR SEMESTER (UAS)**

Untuk Memenuhi Salah Satu Syarat Penyelesaian

Mata Kuliah Data Warehouse

Program S-1 Teknik Informatika

Oleh

Toni Hari Wibowo

NIM. 20200801314

**PROGRAM STUDI S-1 TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS ESA UNGGUL**

**JANUARI 2023**

## SOAL UAS

Kompilasi kembali semua penugasan data warehouse yang sudah pernah dilakukan sehingga memenuhi kriteria sebagai berikut:

- **Gunakan data riil dari suatu bisnis (industri nya bebas) minimal 5 tahun**
- Buatlah semua disain yang memungkinkan dari data tersebut yang mencakup analisis kebutuhan (goal-driven, data-driven, user-driven dll), disain konseptual, disain logikal, disain fisik dan disain ETL
- Gunakan tools apapun (misal pentaho) untuk merancang ETL sehingga data riil yang dimiliki bisa masuk ke database DW/ DM yang sudah dibuat. Buat job untuk otomatisasi ETL tersebut. Jangan lupa sertakan penjelasan seperlunya
- Buat visualisasi dari hasil DW/ DM yang sudah didapatkan (misal dengan pivot excel)

Delivery yang harus dikumpulkan adalah:

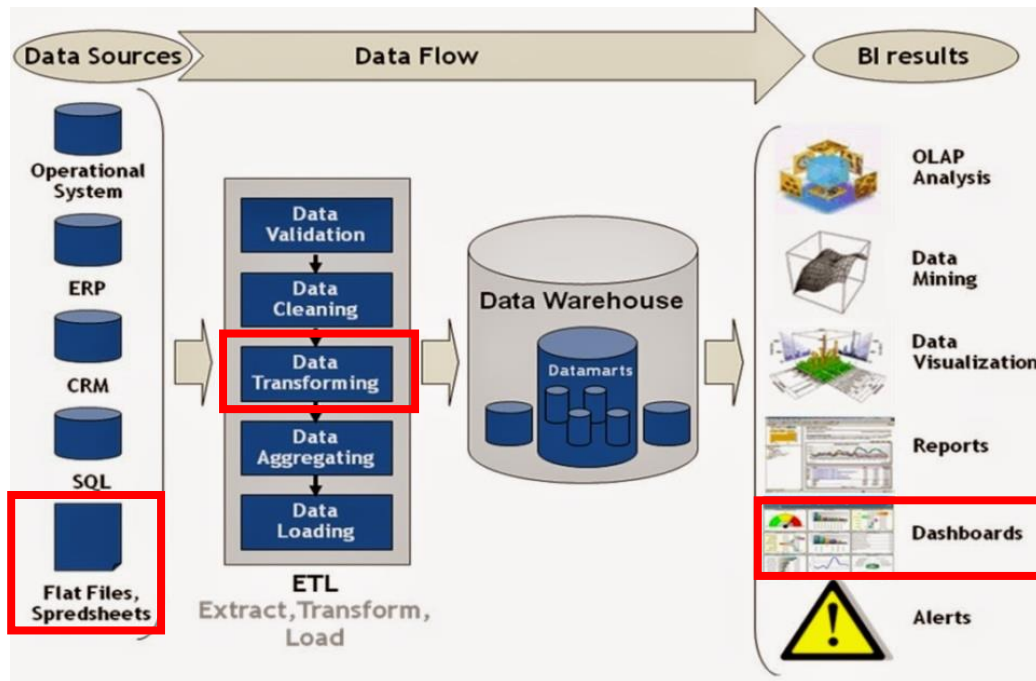
- Data riil minimal 5 tahun (Nilai 10)
- Disain lengkap DW/ DM (Nilai 35 )
- File .ktr dan job yang sudah dibuat. Jika pakai tools selain pentaho, sertakan juga instalasinya (Nilai 35)
- Visualisasi hasil DW/ DM (Nilai 10)
- Readme.txt yang berisi penjelasan rinci tentang apa yang dikumpulkan sehingga bisa terbaca DW/ DM apa yang di deliver serta bagaimana menjalankannya sehingga bisa muncul sesuai dengan yang diharapkan (Nilai 10)

### Jawab :

Dari soal diatas maka penulis akan menjawab soal tersebut sebagai berikut, dimana penulis menggunakan dataset yang di ambil dari situs kaggle dengan dataset publik E-commerce Brazilian by Olist. Dataset ini memiliki informasi 100 ribu pesanan dari tahun 2016 hingga 2018 yang dibuat di beberapa pasar di brasil. Fitur-fiturnya memungkinkan melihat pesanan dari berbagai dimensi: dari status pesanan, harga, kinerja pembayaran dan pengiriman hingga lokasi pelanggan, atribut produk, dan akhirnya ulasan yang ditulis oleh pelanggan. Berikut ini adalah link dataset yang digunakan :

[https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=olist\\_customers\\_dataset.csv](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=olist_customers_dataset.csv)

Dalam jawaban Ujian Akhir Semester (UAS) ini, penulis memproses dataset tersebut melalui alur sebagai berikut :



Gambar 1. Flow

## 1. Data Sources

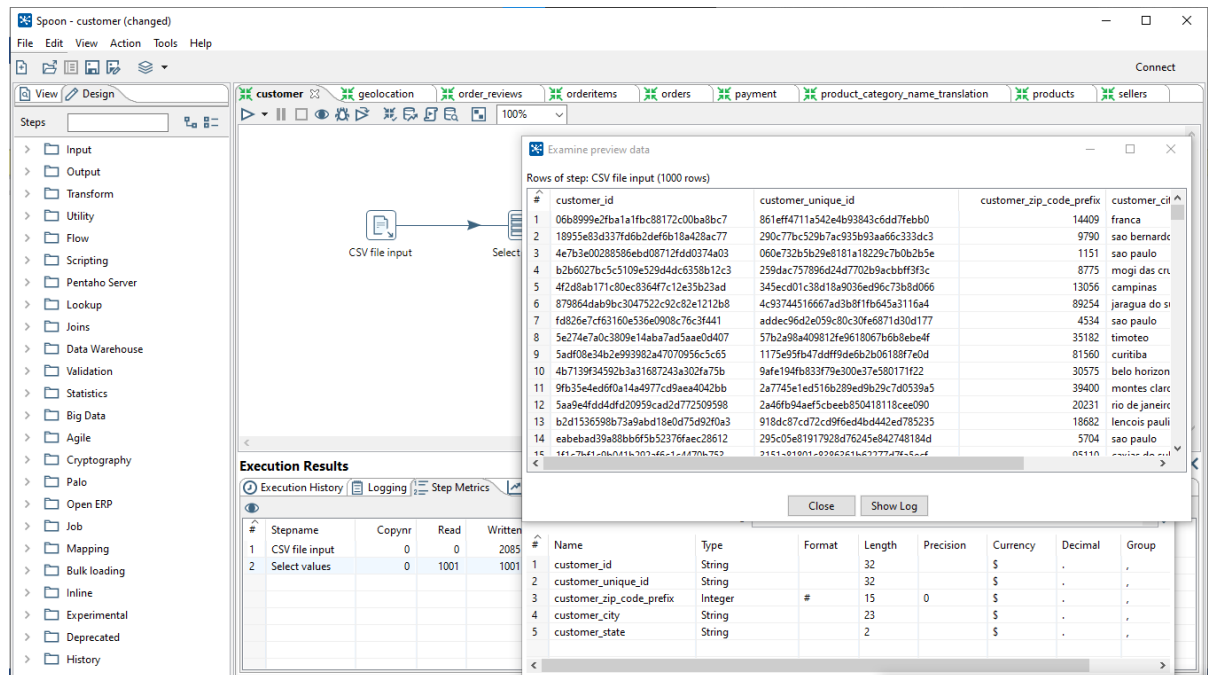
Data ini di ambil dapat dari berbagai sumber data, dengan contoh lain pada pengerjaan UAS ini data di ambil dari situs Kaggle dataset publik dalam bentuk Flat Files yang berformatkan .csv

## 2. Data Flow

Pada Tahap kedua ini dataset yang telah dimiliki akan dilakukan proses ETL (*Extract, Transform, Load*) menggunakan tools **Pentaho**. Hasil dari *extract* dan *transform* akan di *load* kedalam database yang ada pada server *local* didalam PostgreSQL.

- **Extract**

Pada layer ini merupakan tahapan *Extracting* data, sebelum data digunakan terlebih dahulu data yang berformatkan file .csv tersebut akan dibaca menggunakan tools Pentaho. Berikut adalah gambaran dari file .csv yang telah di *extract* kedalam tools Pentaho :



**Gambar 2.** Extract Customer Dataset

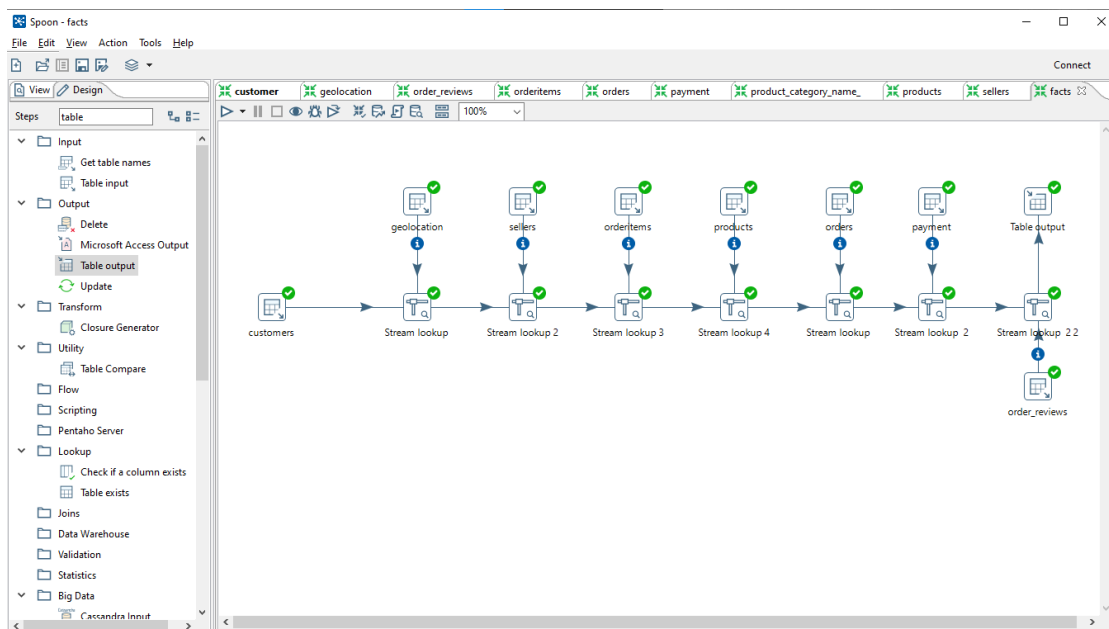
Pada gambar 2 diatas merupakan gambaran proses pada saat melakukan *extracting* disalah satu dataset menggunakan *tools* Pentaho. Dari dataset yang diperoleh melalui situs Kaggle, dataset ini memiliki 9 file *.csv*, diantaranya :

- 1.olist\_customers
- 2.olist\_geolocation
- 3.olist\_order\_items
- 4.olist\_order\_payments
- 5.olist\_order\_reviews
- 6.olist\_orders
- 7.olist\_products
- 8.olist\_sellers
- 9.product\_category\_name\_translation

## • Transform

Dari semua table yang telah di *extract* kedalam *tool* Pentaho, dilakukannya transformasi data, hal tersebut berguna untuk melakukan pembersihan dari baris yang kosong pada data yang akan dipakai untuk divisualisasi dan di analisa. Ketika setiap table telah di

transformasikan, maka dilakukannya proses integrasi table agar menjadi terstruktur. Berikut merupakan gambaran dari proses transform data menggunakan *tools* Pentaho :



**Gambar 3.** *Transform Dataset*

Dari gambar 3 diatas merupakan proses *transform* pada setiap dataset yang telah diubah kedalam tabel. Tabel - tabel tersebut selanjutnya dilakukan proses integrasi agar menjadi satu table yang memiliki kolom yang lengkap untuk dilakukan Visualisasikan.

- **Load**

Pada layer ini merupakan proses melakukan pengiriman data kedalam database menggunakan postgresQL. Data yang sudah dilakukan *transfrom* dan telah di integrasi menjadi satu kesatuan akan di *load* kedalam database yang ada di postgresQL. Berikut adalah gambaran pada saat melakukan *load* kedalam database yang ada di postgresQL

Execution Results													
Execution History   Logging   Step Metrics   Performance Graph   Metrics   Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	payment	0	0	103886	103886	0	0	0	0	Finished	2.2s	48,297	-
2	geolocation	0	0	1000163	1000163	0	0	0	0	Finished	12.0s	83,174	-
3	customers	0	0	99441	99441	0	0	0	0	Finished	16.3s	6,114	-
4	Stream lookup	0	1099604	99441	0	0	0	0	0	Finished	17.9s	61,441	-
5	products	0	0	32951	32951	0	0	0	0	Finished	2.5s	13,271	-
6	sellers	0	0	3095	3095	0	0	0	0	Finished	2.7s	1,145	-
7	Stream lookup 2	0	98568	95472	0	0	0	0	0	Running	18.8s	5,256	3968/9900
8	order_reviews	0	0	99224	99224	0	0	0	0	Finished	2.2s	44,217	-
9	orders	0	0	99441	99441	0	0	0	0	Finished	5.9s	16,769	-
10	orderitems	0	0	112650	112650	0	0	0	0	Finished	6.8s	16,542	-
11	Stream lookup 3	0	198222	85571	0	0	0	0	0	Running	18.8s	10,570	9900/9900
12	Stream lookup 4	0	108623	75671	0	0	0	0	0	Running	18.8s	5,792	9899/9901
13	Stream lookup	0	165211	65769	0	0	0	0	0	Running	18.8s	8,809	9901/9900
14	Stream lookup 2	0	159755	55868	0	0	0	0	0	Running	18.8s	8,518	9900/9900
15	Stream lookup 2.2	0	145192	45967	0	0	0	0	0	Running	18.8s	7,742	9900/9967
16	Table output	0	36000	35000	0	35000	0	0	0	Running	18.8s	1,920	9967/0

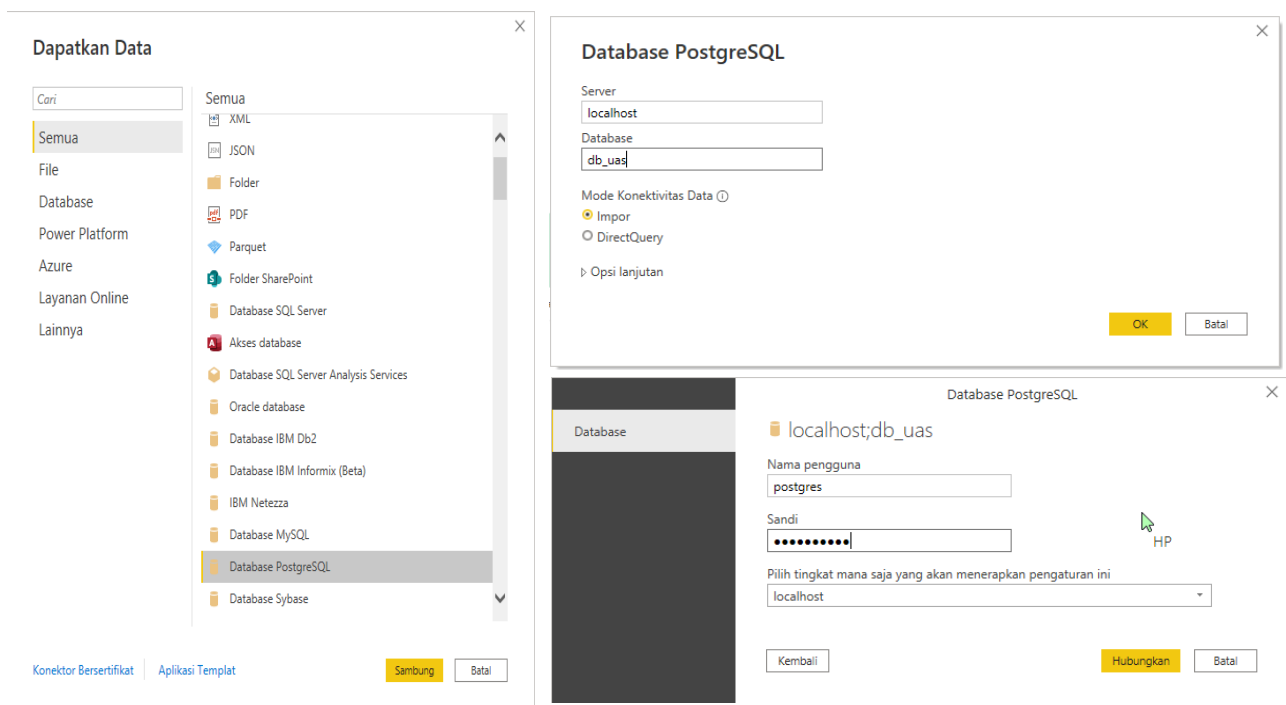
Gambar 4. Push Data into PostgreSQL

customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state	geoloca
06b8999e2fba1a1fbc88172c00ba8bc7	861ef4711a542e4b93843c6dd7febb0	14,409	franca	SP	franca
18955e83d37f46b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	9,790	sao bernardo do campo	SP	sao bernard
4e7b3e0028850ebd08712fd00374a03	060c732b5b29e8181a18229c7b0b2b5e	1,151	sao paulo	SP	sao paulo
b2b6027bc5c5109c52984dc6258b12c3	259d9c757896d24d7702b9acbbff9f3c	8,775	mogi das cruzeiros	SP	mogi das cr
4f2a8ab171c80ec8364f7c12c5b23ad	345ecd01c38d18a9036ced96c73b8d066	13,056	campinas	SP	campinas
879864dab9bc3047522c92e82e1212b8	4c93744516667ad3b8f1fb645a3116a4	89,254	jaragua do sul	SC	jaragua do s
fd826e7cf63160c536e908c76c3f441	addec96d2e059c80c30fe6871d30d177	4,534	sao paulo	SP	sao paulo
5e274e7ad0c3809e14aba7ad5aae0d407	57b2a98a409812fe9618067b6b8e4f	35,182	timoteo	MG	timoteo
5ad00e34b2e993982a47070956c5c65	1175e95fb47ddff99de6b2b061887e0d	81,560	curitiba	PR	curitiba
4b7139f34592b3a31687243a302fa75b	9afe194fb833f79c300e37c580171f22	30,575	belo horizonte	MG	belo horizon
9fb35e4ed6f0a14a4977cd9aa4042bb	2a7745e1ed516b289ed9b29c7d0539a5	39,400	montes claros	MG	montes clar
5aa9e4fd4d4d20959cad2d77250998	2a46fb94aef5cbeeb850418118cee090	20,231	rio de janeiro	RJ	rio de janeir
b2d1536598b73a9ab18e0d75d9210a3	918dc87cd72cd9f6e4b442ed785235	18,682	lencois paulista	SP	lencois paul
eaebad39a88b6ff5b5237f6aec28612	295cd5e81917928d76245e842748184d	5,704	sao paulo	SP	sao paulo
1f1c7bf1c9b041b292af6c1c470b753	3151a81801c8386361b6227d7fa3ecf	95,110	caxias do sul	RS	caxias do st
206f3129c0e4d7d0b9590426023f0a08	21f748a16f4e1c8e9014eb3ee6fa325	13,412	piracicaba	SP	piracicaba
a7c125a0a07b75146167b704a78e98	5c2991dbd08bb5c410713c4de5a0e5	22,750	rio de janeiro	RJ	rio de janeir
c5c61596a3b6bd0ce5766992c48a9a1	b6e99561fedf34a55b0b7da92f8ed775	7,124	guarulhos	SP	guarulhos
9b8ce80369b362d6aad4613ef426f	7f3a72e8f988c6e735ba118d54f47458	5,416	sao paulo	SP	sAo paulo
49d0ea0986edde72da777f15456a0ee0	3e6fd6b210d499a56a6a6820a40f2d79	68,485	pacaja	PA	pacaja
154c4ded6991bdfa3cd249d11abf4130	e607ede0e6343630866023f5a52da5e	88,034	florianopolis	SC	florianopoli
690172ab31962268d3bd4f42f676898	a96d5cfa0d3181817e2b946f921ea021	74,914	aparecida de goiania	GO	aparecida d
2938121a40a20953c43caa8c98787fb	482441ea6a06b1f72fe9784756c0e75	5,713	sao paulo	SP	sao paulo
237098a64674ae89babdc426746260fc	4390ddb6276a66ff1736a710205dca	82,820	curitiba	PR	curitiba
cb721d7b4f27f1d87011c4c83462c076	a5844ba4bfc8d0ccf1d13027c763bcc	8,225	sao paulo	SP	sao paulo

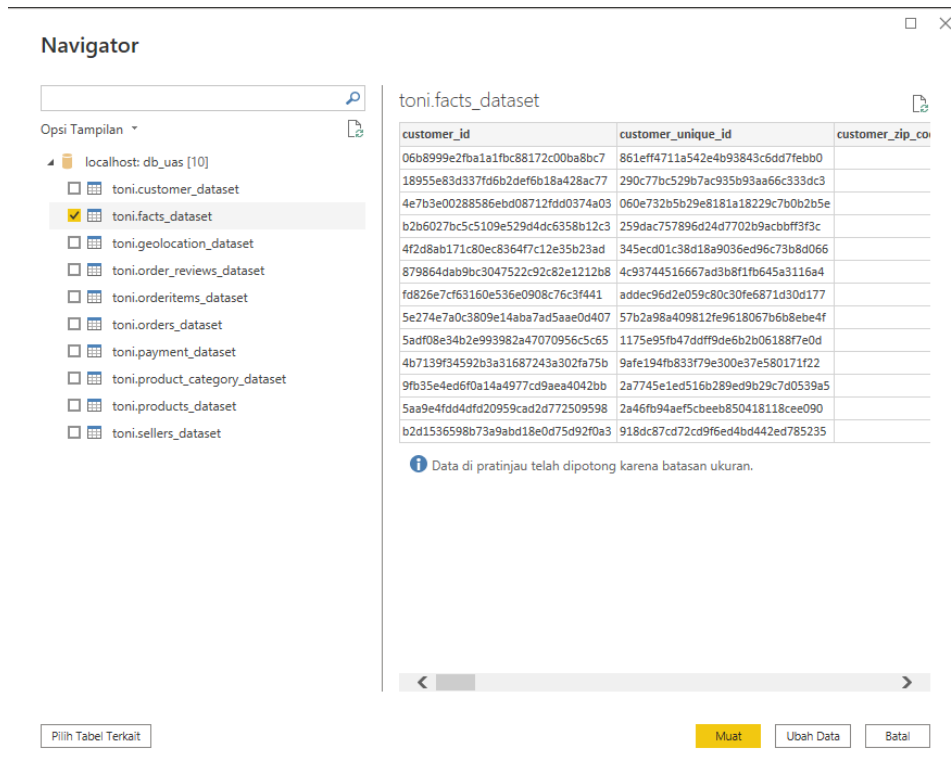
Gambar 5. Database PostgreSQL

### 3. Business intelligence (BI) Results

Pada tahap ini merupakan penerapan teknik transformasi data dari data mentah menjadi informasi yang berguna dan bermakna untuk tujuan analisis, salah satu teknik yang digunakan untuk melakukan pengerjaan Soal UAS ini yaitu membuat dashboard menggunakan PowerBI dengan cara menghubungkan database server PostgreSQL dengan PowerBI. Data yang telah disimpan di dalam PostgreSQL nantinya akan di panggil kembali kedalam PowerBI, sehingga perlu dihubungkan antara PostgreSQL dengan PowerBI yang nantinya di dalam PowerBI data tersebut akan di visualisasikan agar dapat dengan mudah di analisis. Berikut adalah gambaran proses hasil import data dari database PostgreSQL kedalam PowerBI :

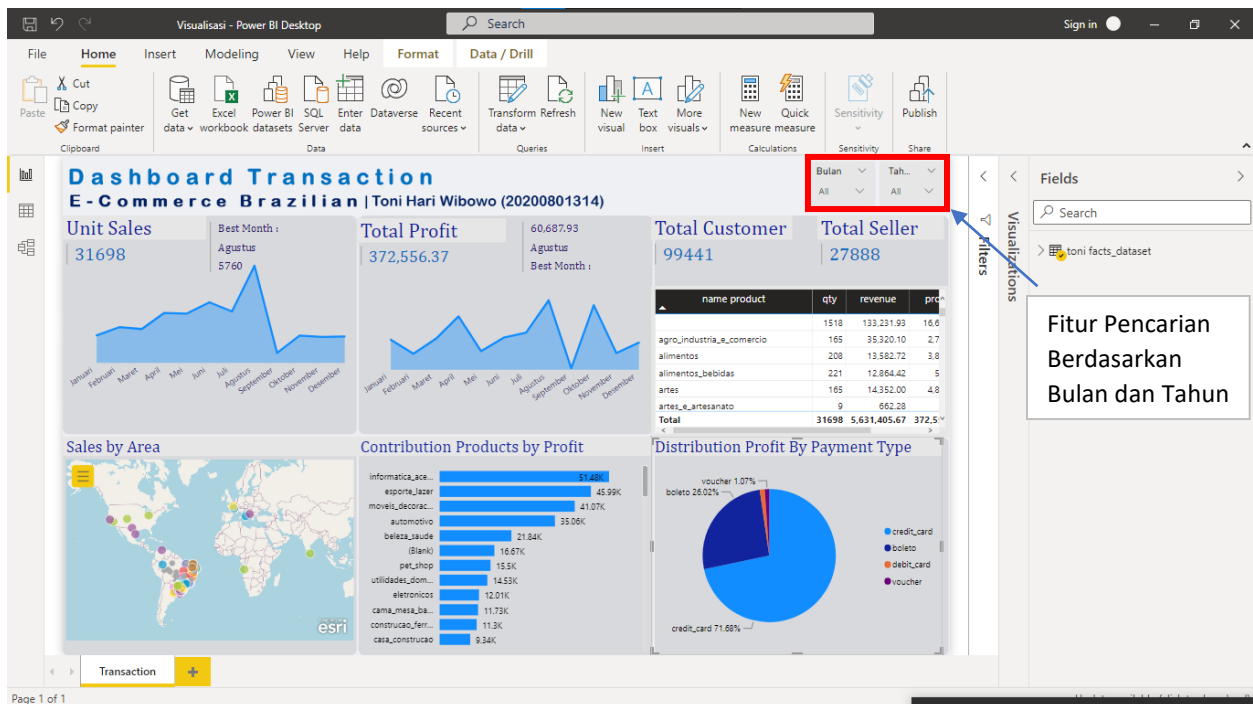


**Gambar 6.** Connecting From PowerBI Into PostgreSQL



**Gambar 7.** Import Data From PostgreSQL Into PowerBI

## Hasil



**Gambar 8.** Dashboard PowerBI, Halaman Transaction



Pada gambar 8 sebelumnya merupakan tampilan dari Dashboard yang telah dibuat dan pada gambar tersebut merupakan tampilan informasi dari Halaman Transaction, pada halaman ini penulis membuat beberapa fitur salah satunya untuk pencarian total transaksi berdasarkan bulan dan tahun yang ingin di tentukan, fitur ini penulis buat berdasarkan bulan dan tahun yang di ambil dari kolom **order\_delivered\_customer** yang ada pada database. Selain itu masih ada beberapa informasi yang penulis tampilkan pada halaman dashboard Transaksi seperti yang ada pada gambar diatas, yaitu berupa :

- **Unit Sales**

Informasi Unit Sales yang penulis buat menggunakan visualisasi multi\_row card dari visualisasi tersebut digunakan penjumlahan keseluruhan data dari variable quantity untuk menghasilkan informasi, selain itu dari visualisasi Area Chart penulis juga menggunakan quantity yang di bandingkan dengan variable order\_delivered\_date dan di filter berdasarkan top 1 untuk pencarian bulan terbaik pada Unit Sales.

- **Total Profit**

Informasi Total Profit yang penulis buat menggunakan visualisasi multi\_row card dan juga Area Chart, dari visualisasi multi\_row card menggunakan penjumlahan keseluruhan data dari variable profit, dimana variable profit dibuat setelah pembuatan variable revenue, variable tersebut penulis buat terlebih dahulu agar nantinya dapat digunakan untuk memberikan informasi berupa total profit, berikut adalah rumus yang penulis gunakan :

```
revenue = SUMX('toni facts_dataset', 'toni facts_dataset'[qty]*'toni facts_dataset'[price])
```

**Table 1.** Rumus penghitungan revenue

```
profit = SUMX('toni facts_dataset', 'toni facts_dataset'[revenue]-'toni facts_dataset'[price])
```

**Table 2.** Rumus penghitungan profit

- **Total Customer**

Informasi Total Customer yang penulis buat menggunakan visualisasi multi\_row card yang menggunakan penjumlahan keseluruhan data dari variable customer\_unique\_id, keputusan variable yang penulis gunakan untuk pembuatan informasi tersebut dikarenakan bahwa

setiap customer memiliki id yang unique, sehingga tidak ada duplikasi atau redundan pada penghitungan data.

- **Total Seller's**

Informasi Total Seller's yang penulis buat menggunakan visualisasi multi\_row card yang menggunakan penjumlahan keseluruhan data dari variable seller\_id, dari data yang ada pada kolom seller\_id penulis filter terlebih dahulu agar nantinya tidak ada id yang redundan untuk di jumlahkan.

- **Tabel**

Informasi pada Tabel yang penulis buat terisi dari product name, quantity, price, revenue, profit hingga gross profit margin(GPM). berikut penghitungan rumus dari GPM :

$\text{GPM (\%)} = \text{DIVIDE}([\text{profit}], [\text{revenue}])$
--

**Table 3.** Rumus penghitungan Gross Profit Margin(GPM)

- **Sales by Area**

Pada informasi Sales by Area penulis menampilkan persebaran peta dari product yang di order, Informasi tersebut penulis buat menggunakan visualisasi ArcGIS Maps yang berisikan value dari order\_item\_id dan di bandingkan dengan customer\_city lalu di filter berdasarkan customer\_state agar peta dari persebaran penjualan product dapat terlihat dengan jelas pada setiap daerah nya.

- **Contribution Products by Profit**

Informasi yang di tampilkan pada Contribution Products by Profit penulis buat menggunakan visualisasi Bar Chart, yang berisikan value berdasarkan perbandingan dari product\_category dengan profit.

- **Distribution Profit by Payment Type**

Pada Informasi Distributin Profit by Payment Type penulis buat menggunakan visualisasi pie chart yang berisikan value dari profit dan di bandingkan dengan payment\_type

## Kesimpulan

Dari penjelasan diatas dari proses *Extract, Transform, Load* data hingga pembuatan informasi data dalam bentuk Visual di dalam powerBI, penulis menyimpulkan, Dashboard Visual dari dataset E-Commerce Brazilian By Olist telah dibuat, dashboard tersebut dibuat guna untuk memenuhi syarat dalam pengerjaan SOAL UAS ini dan juga berguna untuk menampilkan informasi data dari data mentah menjadi dalam bentuk visualisasi yang berguna untuk tujuan pembelajaran analisis. Dari hasil analisis yang penulis lakukan, dimana pada dataset E-Commerce Brazilian By Olist, jumlah setiap tahun seperti customer, product, seller dan order barang terus meningkat, dari hasil peningkatannya tersebut banyak terjadi penjualan di bulan **Agustus** dari tahun **2016 s/d 2018**, selain itu category product yang paling laris di jual yaitu category product **informatica\_acessorios** dengan nama lain (*computers\_accessories*) yang berada di tingkatan pertama.

**D**engan jawaban pada soal UTS yang sudah di isi diatas maka, penulis mencantumkan link repository Github yang terdapat File **.ktr**, job, dataset yang digunakan hingga file Dashboard PowerBI yang telah dibuat sebagai bukti pengerjaan UAS ini dilakukan secara pribadi dan UAS ini merupakan hasil karya sendiri.

[https://github.com/ToniHariWibowo/UAS\\_DW\\_ETL\\_20200801314](https://github.com/ToniHariWibowo/UAS_DW_ETL_20200801314)