

# Penerapan Data Warehouse dan Business Intelligence Pada Perspektif Financial dan Customer Perusahaan Olist

- Muhammad Ayyub Ramli -



# Rundown



Part 1: Pendahuluan

---

Part 2: Objektif

---

Part 3: Sistem Data Warehouse

---

Part 4: Data Mining

---

Part 5: Dashboard

---

Part 6: Kesimpulan dan Saran

---



# Pendahuluan

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

# Profil Perusahaan



Olist adalah salah satu perusahaan jual beli online terbesar di Brazil. Sejak diluncurkan secara resmi, Olist telah menjadi salah satu perusahaan internet di Brazil dengan pertumbuhan yang pesat. Menggunakan model bisnis marketplace dan mall online, Olist memungkinkan setiap individu, toko kecil, dan merek untuk membuka dan mengelola toko online.

## Rumusan Masalah

- 1) Bagaimana hasil balanced scorecard pada perusahaan Olist jika dilihat menggunakan 4 perspektif?

---

- 2) Bagaimana hasil dari pengimplementasian desain multidimensional, hierarki, measure dan OLAP pada Perusahaan Olist?

---

- 3) Bagaimana hasil analisis dari data source dan desain data warehouse pada Perusahaan Olist?

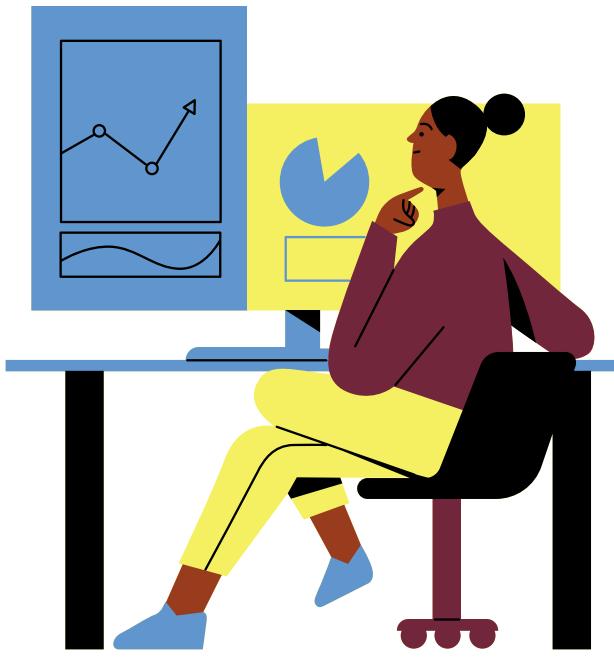
---

- 4) Bagaimana cara mengolah data menggunakan data mining berdasarkan objective Perusahaan Olist yang dipilih?

---

- 5) Bagaimana hasil dari pengolahan data sesuai dengan objective dan KPI pada Perusahaan Olist dalam bentuk dashboard?

# Batasan Masalah



- 1) Data yang digunakan pada pembahasan/penelitian kali ini hanya menggunakan database dari perusahaan Olist yang dikonversi dalam bentuk csv dan bersumber dari Kaggle
- 2) Data yang didapat dari rentang tahun 2016 sampai 2018
- 3) Tools yang digunakan dalam pembuatan Data Warehouse menggunakan Pentaho
- 4) Tools yang digunakan dalam penggerjaan Data Mining menggunakan bahasa pemrograman Python
- 5) Tools yang digunakan dalam pembuatan Dashboard menggunakan Google Data Studio



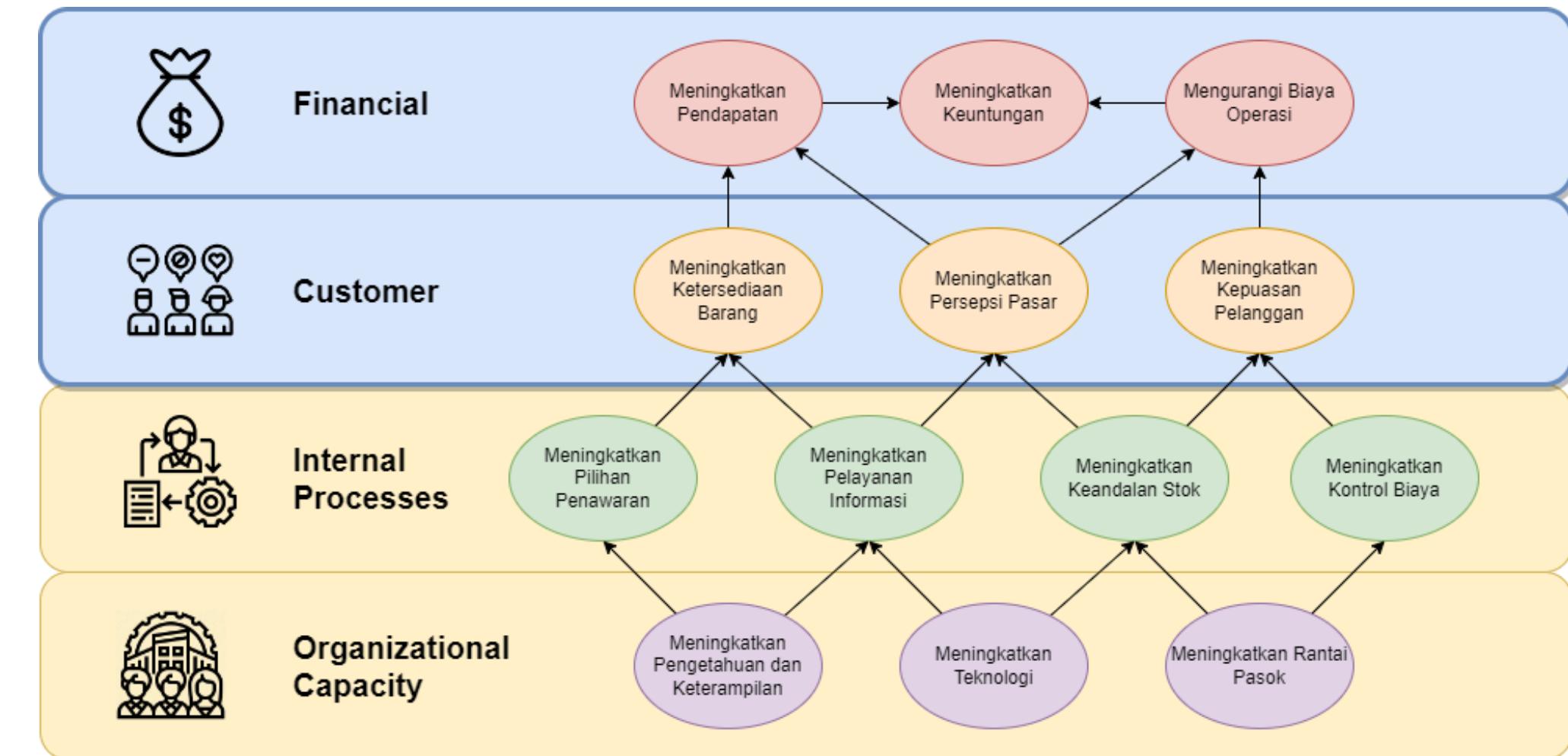
# Objektif

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

Group G SI-44-01

# Balance Scorecard

Balance scorecard adalah suatu sistem yang digunakan oleh perusahaan untuk mengukur dan mengikuti kinerja keuangan dan non-keuangan dengan menggunakan empat perspektif yang saling terkait.



# Key Performance Indicator (KPI)

KPI adalah singkatan dari Key Performance Indicator, yang berarti Indikator Kinerja Utama dalam bahasa Indonesia. KPI dapat digunakan untuk mengukur kinerja dalam berbagai bidang, seperti keuangan, operasional, pemasaran, dan sumber daya manusia.

Objective Strategis	KPI	Target Kinerja	Inisiatif
Financial	<ul style="list-style-type: none"><li>- Rata-rata pemasukan 3 tahun terakhir</li><li>- % Persentase pemasukan orderan cluster tertinggi</li><li>- Pemasukan dari 3 produk paling laku di 3 kota dengan penjualan tertinggi</li></ul>	<ul style="list-style-type: none"><li>- ↑ \$1000 per 3 tahun terakhir</li><li>- ↑ 5% per 3 tahun</li><li>- ↑ \$200 per 3 tahun</li></ul>	<ul style="list-style-type: none"><li>- Melakukan riset serta pemasaran sesuai dengan kategori barang</li><li>- Meningkatkan daya saing serta promosi barang dengan persentase pemasukan yang tinggi</li><li>- Memprioritaskan kategori barang berdasarkan skala penjualan dan pemasukan tertinggi serta lokasi seller</li></ul>

# Key Performance Indicator (KPI)

---

Customer	<ul style="list-style-type: none"><li>- Rata-rata pemrosesan barang sampai ke pelanggan</li><li>- Rata-rata barang telat sampai ke customer</li><li>- % Persentase ketepatan waktu pengiriman barang pelanggan 3 tahun terakhir</li><li>- % Persentase keterlambatan waktu pengiriman barang pada cluster terakhir (paling lambat)</li></ul>	<ul style="list-style-type: none"><li>- ↓ 3 Hari pengantaran</li><li>- ↓ 1 Hari</li><li>- ↑ 90% Pengiriman barang tepat waktu</li><li>- ↓ 10% Pengiriman barang terlambat</li></ul>	<ul style="list-style-type: none"><li>- Pemilihan jasa kurir dan rute pengiriman yang baik</li><li>- Pemilihan metode pengiriman sesuai dengan kebutuhan waktu pengiriman dan memonitor pengiriman secara berkala</li><li>- Pemilihan jasa kurir serta metode pengiriman yang sesuai dengan jenis barang dan jarak pengantaran</li><li>- Memprioritaskan pengiriman dengan estimasi keterlambatan yang tinggi</li></ul>
----------	--	---	---



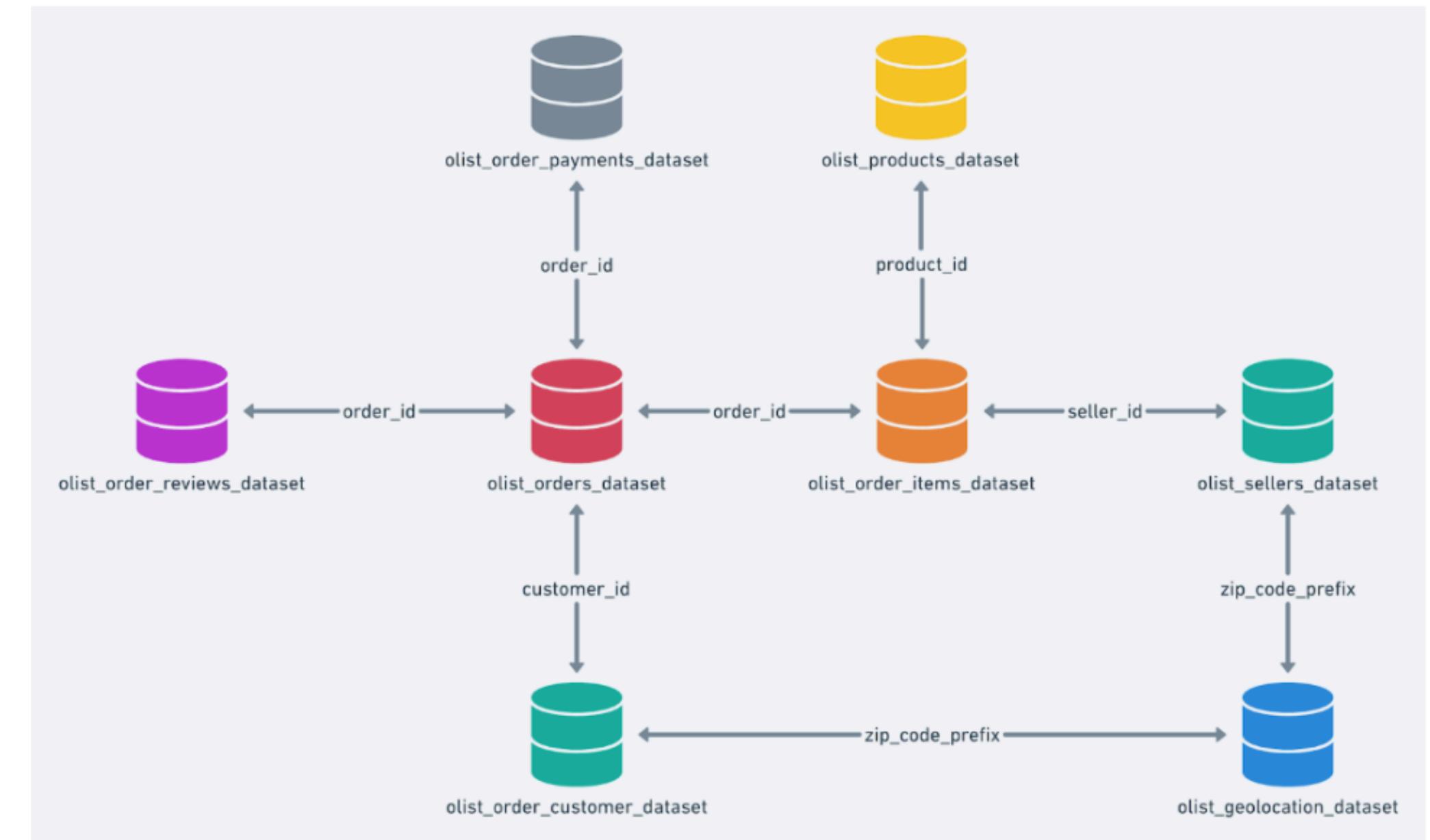
# Sistem Data Warehouse

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

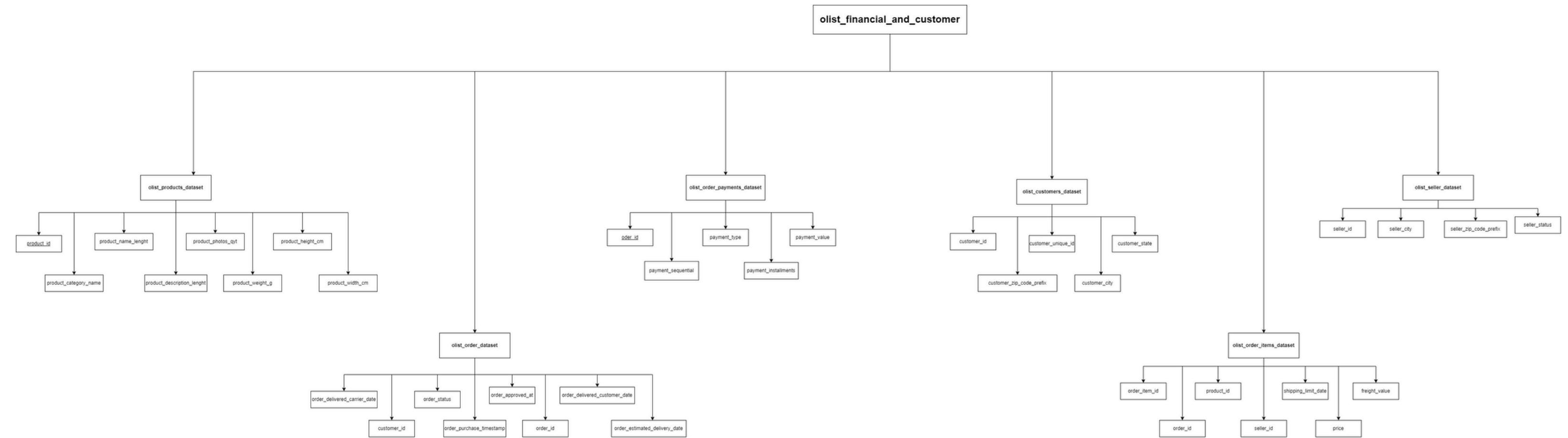
Group G SI-44-01

# Database Olist

Database adalah suatu sistem yang menyimpan dan mengelola data dengan cara yang terstruktur dan terorganisir. Database perusahaan Olist memiliki 8 jenis database yang terhubung satu sama lain berdasarkan relasinya masing - masing menggunakan primary key dan foreign key.



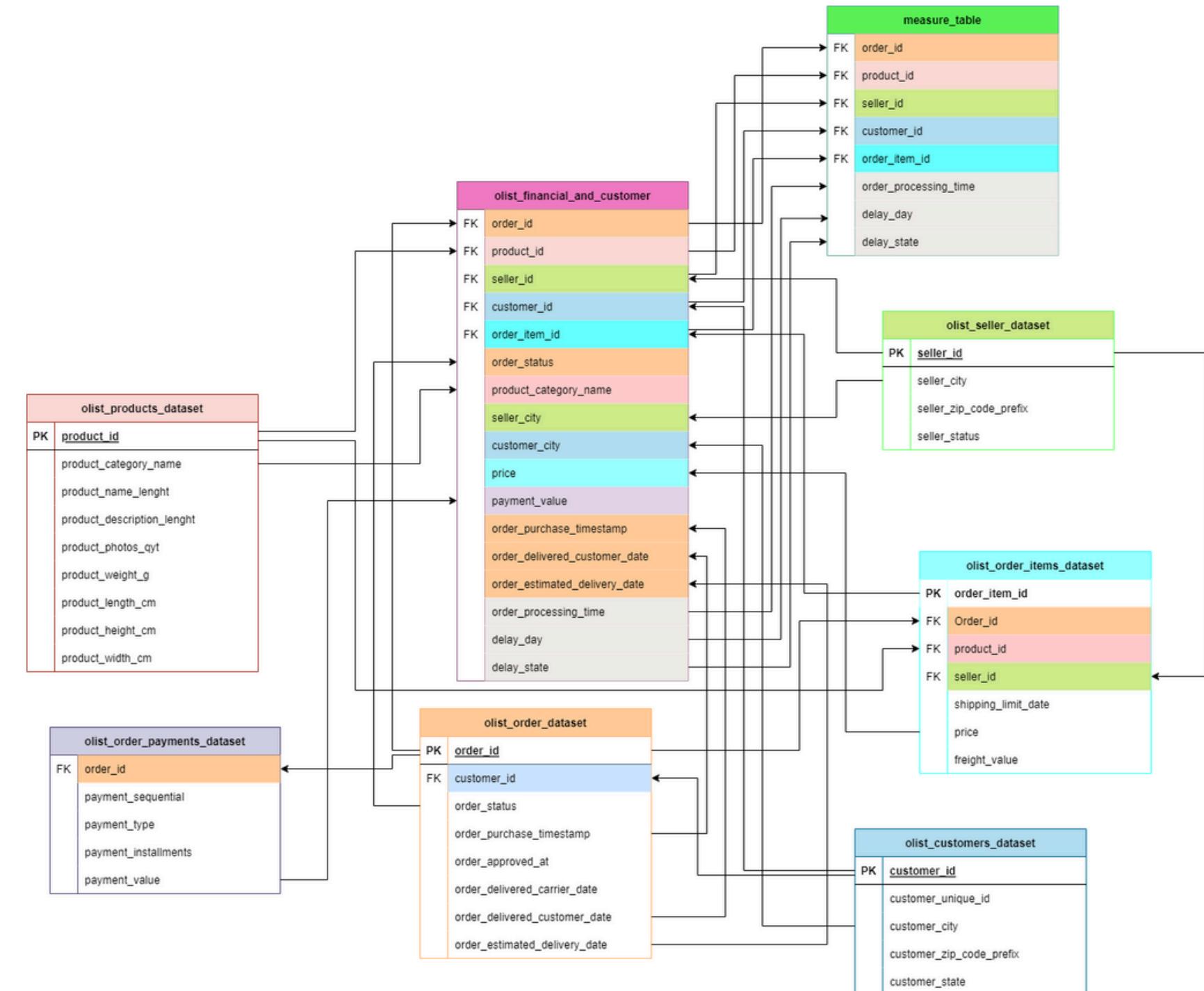
# Hierarki Basis Data



Model database Hirarki adalah model data yang menyusun data dalam bentuk struktur pohon. Hierarki basis data yang diimplementasikan pada database perusahaan Olist dapat dilihat pada gambar berikut.

# Star Schema

Star schema adalah salah satu model penyimpanan data yang sering digunakan dalam sistem informasi data warehouse. Berikut merupakan gambaran Star Schema dari perusahaan olist yang mana terdiri dari enam tabel dimensi serta satu tabel fakta dan measure.



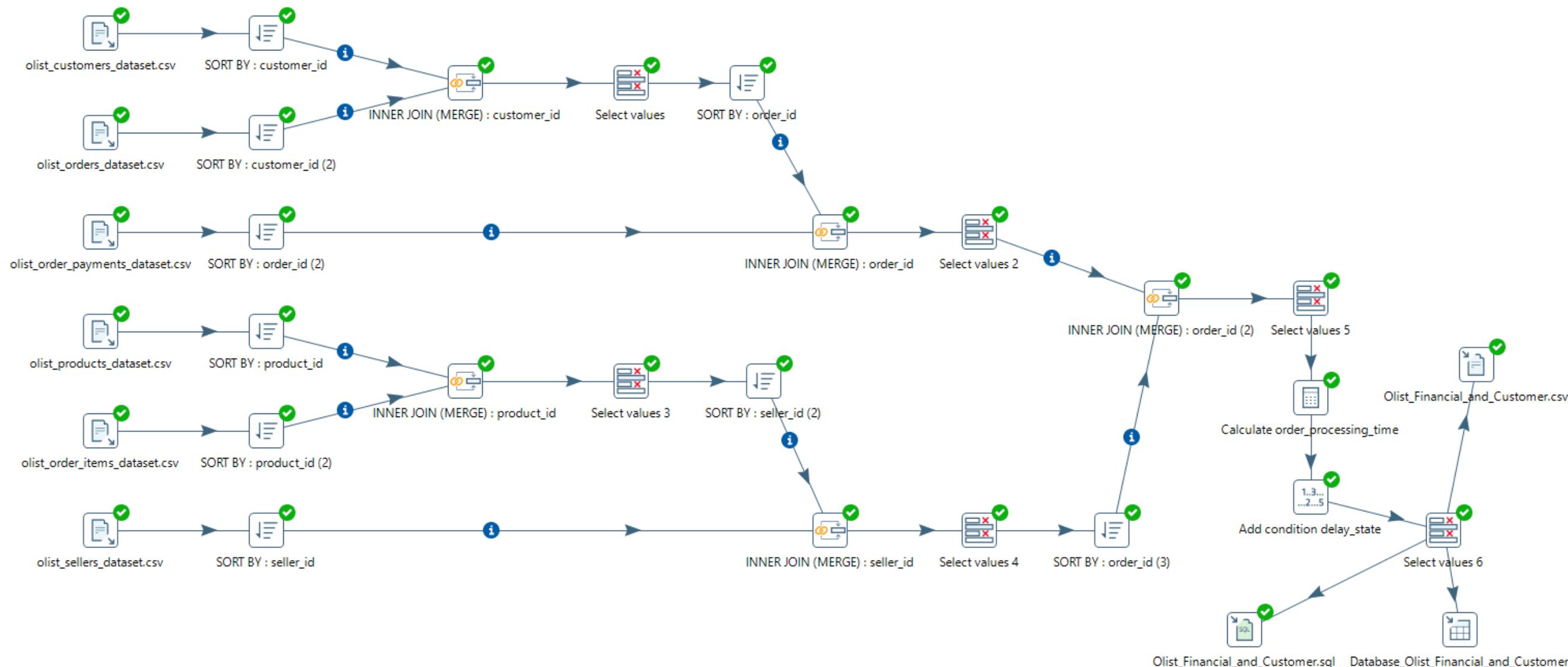
# Desain Data Warehouse

---



Rancangan desain Data Warehouse pada perusahaan Olist menggunakan tools Pentaho. Pentaho merupakan perangkat lunak yang dapat digunakan untuk mengekstrak, mentransformasikan, dan memuat data ke dalam sistem Data Warehouse. Desain Data Warehouse dari perusahaan Olist pada objektif Finance dan Customer dapat dilihat pada gambar berikut.

# Desain Data Warehouse



# Desain Data Warehouse

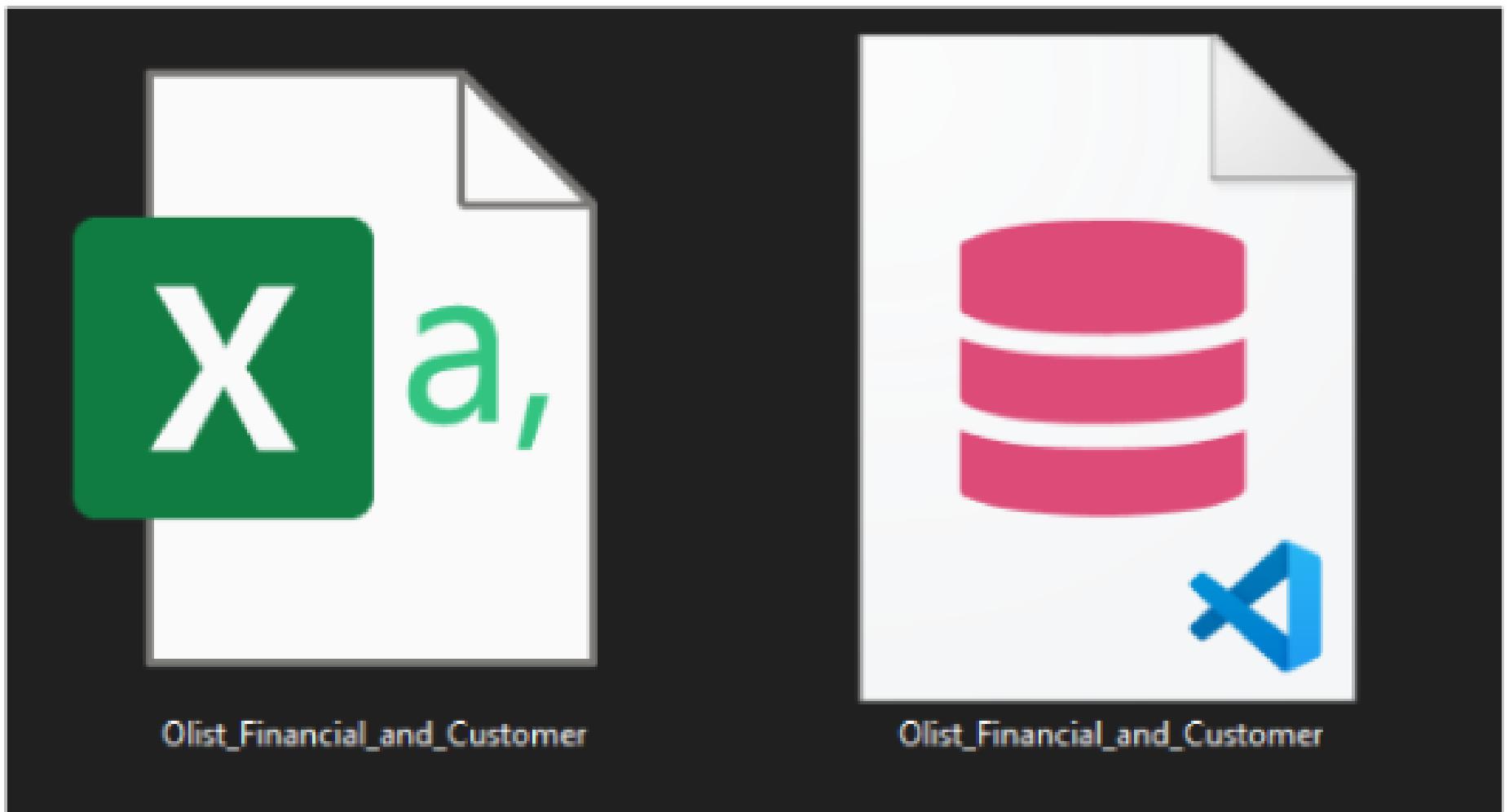
Hasil ekstraksi dataset tersebut kemudian dilakukan Transformasi dengan menggunakan metode sort, inner join, serta melakukan kalkulasi yang menghasilkan kolom pengukuran (measure), selanjutnya akan dilakukan proses Load untuk dilakukan penyimpanan dataset OLAP atau fact table hasil desain data warehouse. Hasil Load tersebut disimpan dalam database MySQL dengan nama database tugasbesar\_kelas01\_kelompokg dan pada table olist\_financial\_and\_customer yang dapat dilihat pada gambar berikut.

order_id	customer_id	customer_city	order_status	order_purchase_timestamp	order_delivered_customer_date	order_estimated_delivery_date	payment_value	product_id
000229ec398224e6ca0657da4fc703e	6489ae5e4333f3693df5ad4372dab6d3	para de minas	delivered	2018-01-14 14:33:31	2018-01-22 13:19:16	2018-02-05 00:00:00	216.67	c777355d18b72b67abee9df44fd0d
00048cc3ae77c65dbb7d2a0634bc1ea	816cbea969fe65b689b39cf97a506742	uberaba	delivered	2017-05-15 21:42:34	2017-05-22 13:44:35	2017-06-06 00:00:00	34.59	ef92defde845ab8450f9d70c526ef70f
0005a1a1728c9d785b8e2b08b904576c	16150771df4776261284213b89c304e	santos	delivered	2018-03-19 18:40:33	2018-03-29 18:17:31	2018-03-29 00:00:00	157.6	310ae3c140ff94b03219ad0adc3c778f
00063b381e2406b52ad429470734eb5d	6a899e55865de6549a58d2c6845e5604	monte alto	delivered	2018-07-27 17:21:27	2018-08-07 13:56:52	2018-08-07 00:00:00	57.98	f177554ea93259a5b282(24e33f65ab6
0008288aa423d2a3f0fb17cd7d8719	2355af7c75e7c9b843a87b2a7f210dc5	jandira	delivered	2018-02-13 22:10:21	2018-02-26 13:55:22	2018-03-06 00:00:00	126.54	368c6c730842d78016ad23897a372d
000aed2e25dbbad2f9db70584c5a2ded	ff5169e583d07fac9fec808962f189d	santa barbara d'este	delivered	2018-05-11 20:33:38	2018-05-18 16:46:31	2018-05-22 00:00:00	152.77	4fa33915031a8cde03dd03e0fb27f01
000e63d38ae8c00bbcb5a30573b9628	98884e672c5ba85f4394f2044e1a3eab	sao paulo	delivered	2018-03-23 19:48:26	2018-03-27 14:51:47	2018-04-05 00:00:00	56.78	553e0e7590d3116a072507a3635d287
001021efaa8f63c29475e7734483a457d	2dbf74659104ca1f00df3720a1d833d	ivaipora	delivered	2018-02-27 09:27:14	2018-03-08 20:06:33	2018-03-23 00:00:00	64.1	5d7c23057ed3fc8c5e699b373d5890b
001102c4b53e22e84023405fb467e57	2013d892495e1a101d742d533d2d1119	sao paulo	delivered	2018-01-16 21:43:23	2018-01-26 22:14:02	2018-02-15 00:00:00	315.33	c389f712c4b4510bc997ce93e8b1a28
0013503b13d1eac686219390b7d641b	f7277fce45fae3165379c15cad5ddbe	osasco	delivered	2017-12-05 23:40:59	2017-12-20 22:46:28	2017-12-28 00:00:00	137.22	38afdf723b956455b418a0f57d623c6b
00143d0f86d6fb9f9b38ab440ac16f5	2e45292979b9b2700ea09560eeb0f003	goiania	delivered	2017-10-16 15:29:43	2017-10-27 18:43:56	2017-11-06 00:00:00	109.29	e95ee6822b66ac6058e2e4aff656071a
0014ae671de39511f7575066200733b7	41065d9dcea52218c3943d2eed072b97	pinhais	delivered	2017-05-22 13:49:03	2017-06-07 13:52:52	2017-06-13 00:00:00	30.6	23365beed31653b4105bd800c46670
0016dfedd97fc2950e388d2971d718c7	2c8b917c5d7dd720ebe36a5ed3b501ec	canavieiras	delivered	2017-04-28 19:54:40	2017-05-22 18:32:21	2017-06-01 00:00:00	17.92	4089061a1bd4685da70bdd6b4f974f1
001862358bf850722e1e2ae000cfed8b	2cf869dd40c98f29686f36d83545cce	cunha	delivered	2018-02-06 19:11:57	2018-02-15 22:57:01	2018-03-02 00:00:00	113.71	c6dd917a0be2a7045b2055949915ab3
001ab0a7578dd66cd4b0715b6e1e41	8bb3bf4e75a959524235cc11a7331af	salvador	delivered	2017-12-27 00:38:47	2018-01-17 20:43:29	2018-01-29 00:00:00	127.56	0b0172eb0fd18479d29c3bc122c058c2
001b76dd48a5b1eee3e87778daa40d8	a62843b08086d1629415a203a7fd419d	santa salete	delivered	2018-03-26 17:42:53	2018-04-06 18:36:23	2018-04-27 00:00:00	95	dbb67791e405873b259e4656b9f71246
001d8f0e34a38e37f7dba237d4eba8b	d987da9fb4086ab7c20f83963cd6722	sao paulo	delivered	2017-05-14 17:19:44	2017-05-26 13:14:50	2017-05-24 00:00:00	53.54	e67307ff0f15ade43fcbe670be7a74c
001f1be1b19605ca0316e71309	b6a1445ee505f3a95c827e7a5188ad02	campo grande	delivered	2017-03-18 11:47:37	2017-03-28 14:23:33	2017-04-12 00:00:00	213.21	884la3cd42986ba480ea18ae425ff

# Desain Data Warehouse

---

Selain menghasilkan database MySQL, proses Load juga akan menghasilkan 2 ekstensi file yaitu .sql dan .csv yang akan digunakan sebagai backup data jika database MySQL bermasalah atau terhapus.





# Data Mining

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

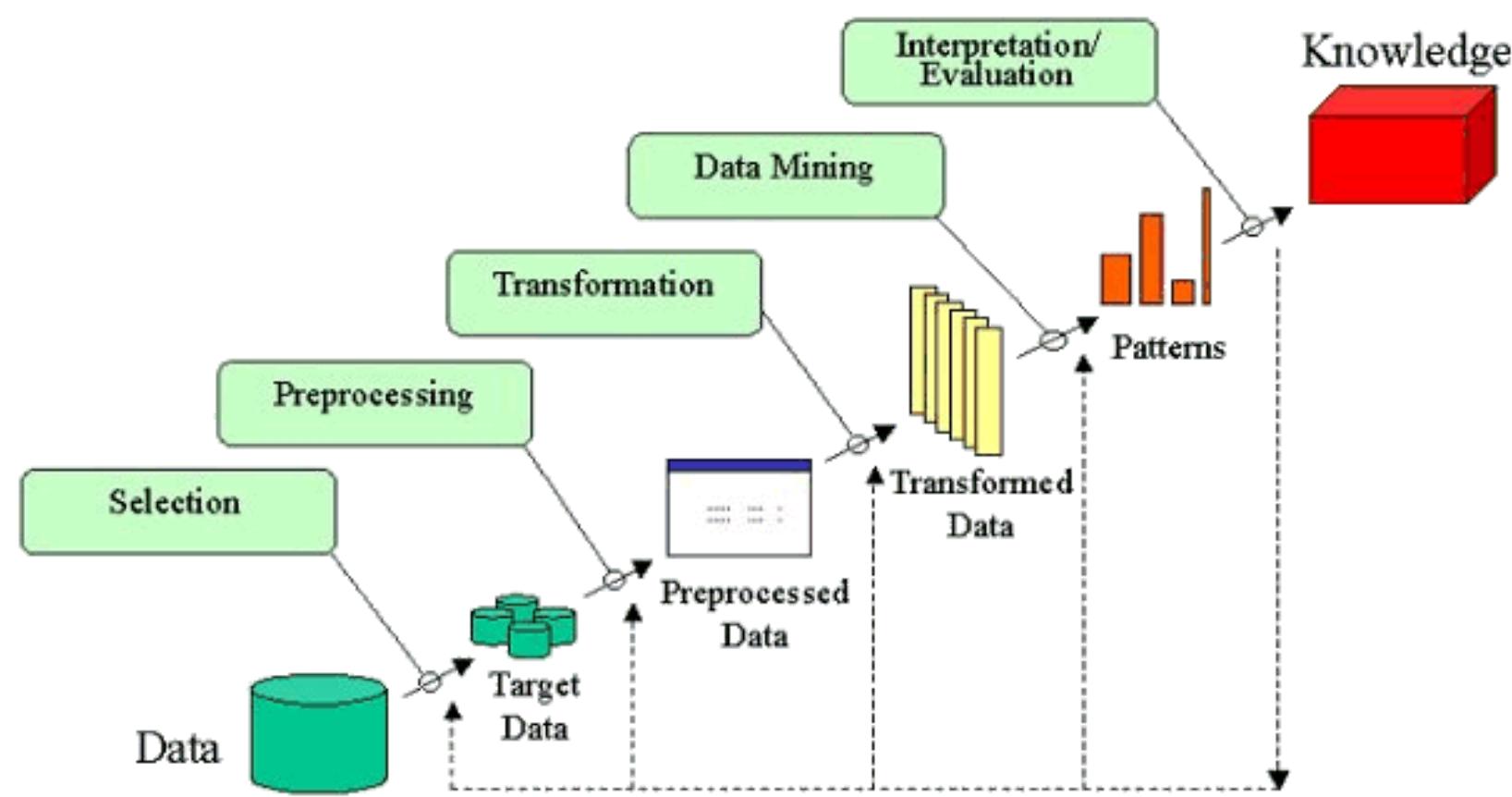
Group G SI-44-01

# Data Mining



Data mining adalah proses yang digunakan untuk mengekstrak informasi berguna dari data yang besar dan kompleks dengan cara menganalisisnya untuk menemukan pola-pola yang tidak terduga.

# Metode Data Mining



(Knowledge Discovery in Database)  
KDD merupakan proses mengekstrak informasi dari suatu set data yang besar dan kompleks untuk menemukan pola yang bermanfaat, atau pengetahuan baru yang tidak diketahui sebelumnya. Proses ini mencakup tiga tahap utama: pengumpulan data, preprocessing data, dan analisis data.

# Implementasi Data Mining



Rancangan Data Mining pada database perusahaan Olist menggunakan metode Knowledge Discovery in Database Process (KDD). Dataset yang digunakan yaitu dataset OLAP atau dataset yang berasal dari proses pembuatan Data Warehouse pada database perusahaan Olist. Pada proses pengimplementasianya menggunakan bahasa pemrograman Python sebagai tools utama dalam melakukan proses Data Mining.

# Implementasi Data Mining : 1) Data Cleansing

```
olist.isna().sum()
```

```
customer_id          0  
customer_city        0  
order_id             0  
order_status          0  
order_purchase_timestamp 0  
order_delivered_customer_date 2567  
order_estimated_delivery_date 0  
payment_value        0  
product_id           0  
product_category_name 1698  
order_item_id         0  
price                0  
seller_id             0  
seller_city           0  
order_processing_time 2567  
delay_day             2567  
delay_state            0  
dtype: int64
```

```
olist.dropna(inplace=True, axis=0)  
olist.isna().sum()
```

```
customer_id          0  
customer_city        0  
order_id             0  
order_status          0  
order_purchase_timestamp 0  
order_delivered_customer_date 0  
order_estimated_delivery_date 0  
payment_value        0  
product_id           0  
product_category_name 0  
order_item_id         0  
price                0  
seller_id             0  
seller_city           0  
order_processing_time 0  
delay_day             0  
delay_state            0  
dtype: int64
```

Menghapus baris data yang bernilai *null*

# Implementasi Data Mining : 1) Data Cleansing

```
olist.duplicated().sum()
```

774

```
olist = olist.drop_duplicates()  
olist.duplicated().sum()
```

0

Menghapus duplikasi data

## Implementasi Data Mining : 2) Data Integration

Tahap ini merupakan proses mengambil data dari berbagai sumber yang mungkin berbeda dan mengkombinasikannya ke dalam suatu format yang terstruktur dan dapat digunakan dengan mudah. Pada tahap ini tidak perlu dilakukan karena dataset yang digunakan yaitu dataset dalam bentuk OLAP atau dataset yang telah dilakukan Data Warehousing pada proses sebelumnya.

# Implementasi Data Mining : 3) Data Selection

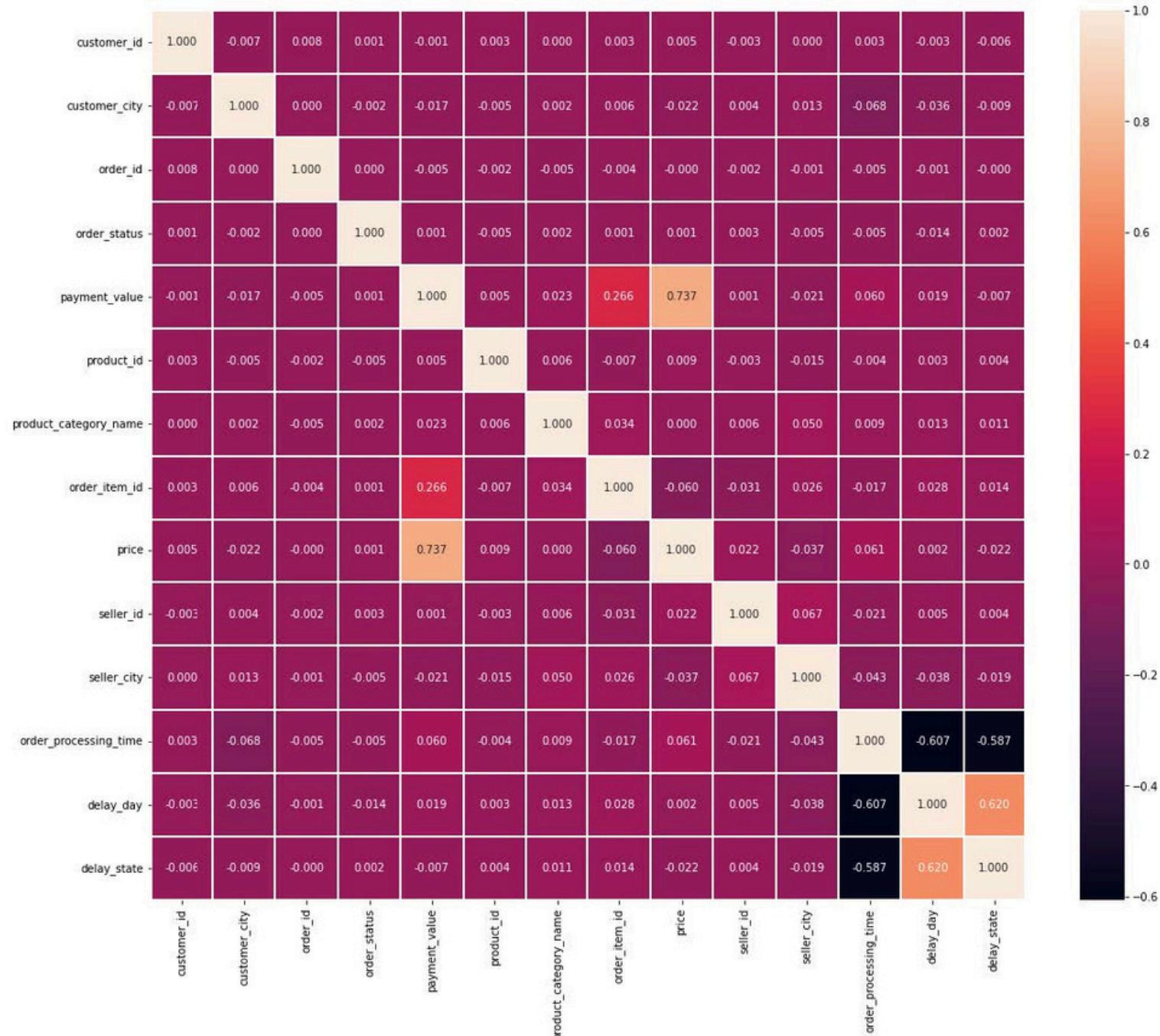


Diagram Heatmap  
Dataset Olist

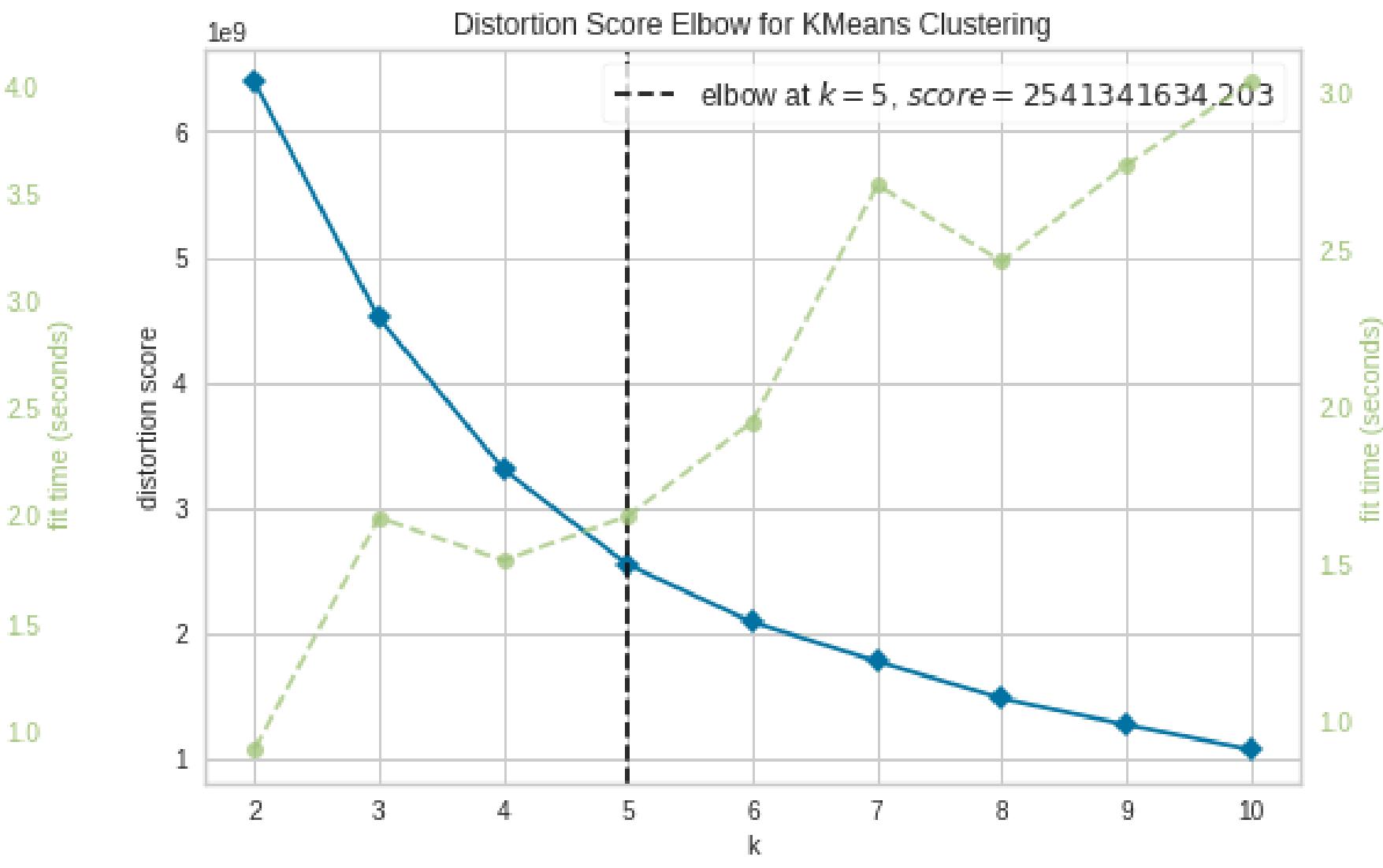
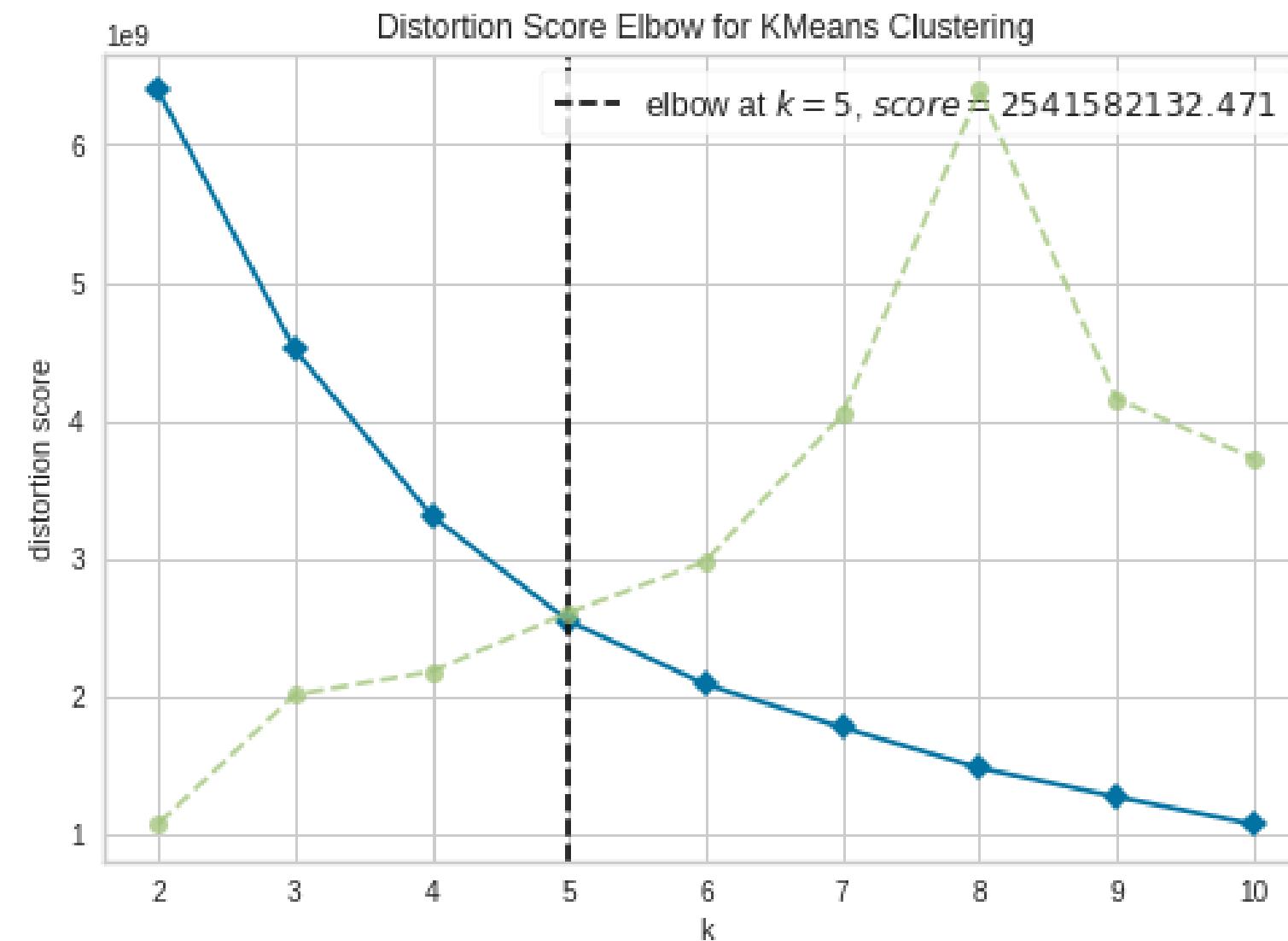
# Implementasi Data Mining : 4) Data Transformation

```
olist.dtypes
```

customer_id	object
customer_city	object
order_id	object
order_status	object
order_purchase_timestamp	datetime64[ns]
order_delivered_customer_date	datetime64[ns]
order_estimated_delivery_date	datetime64[ns]
payment_value	float64
product_id	object
product_category_name	object
order_item_id	int64
price	float64
seller_id	object
seller_city	object
order_processing_time	float64
delay_day	float64
delay_state	object
dtype: object	

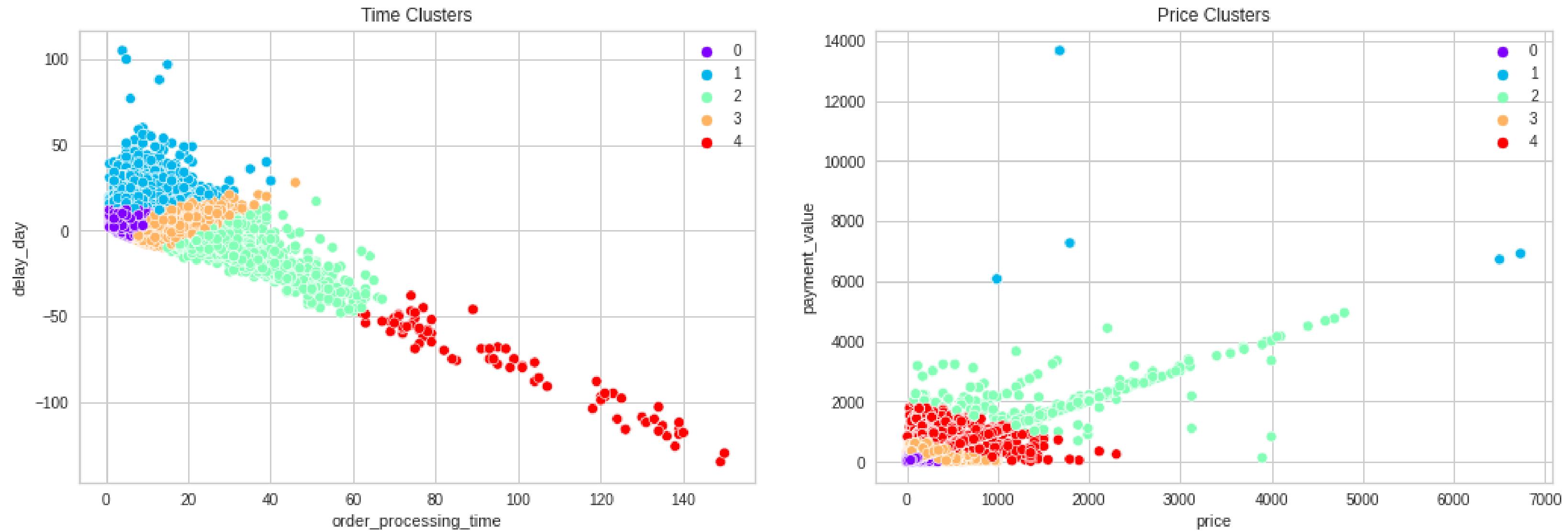
Tipe data dataset Olist

# Implementasi Data Mining : 5) Data Mining



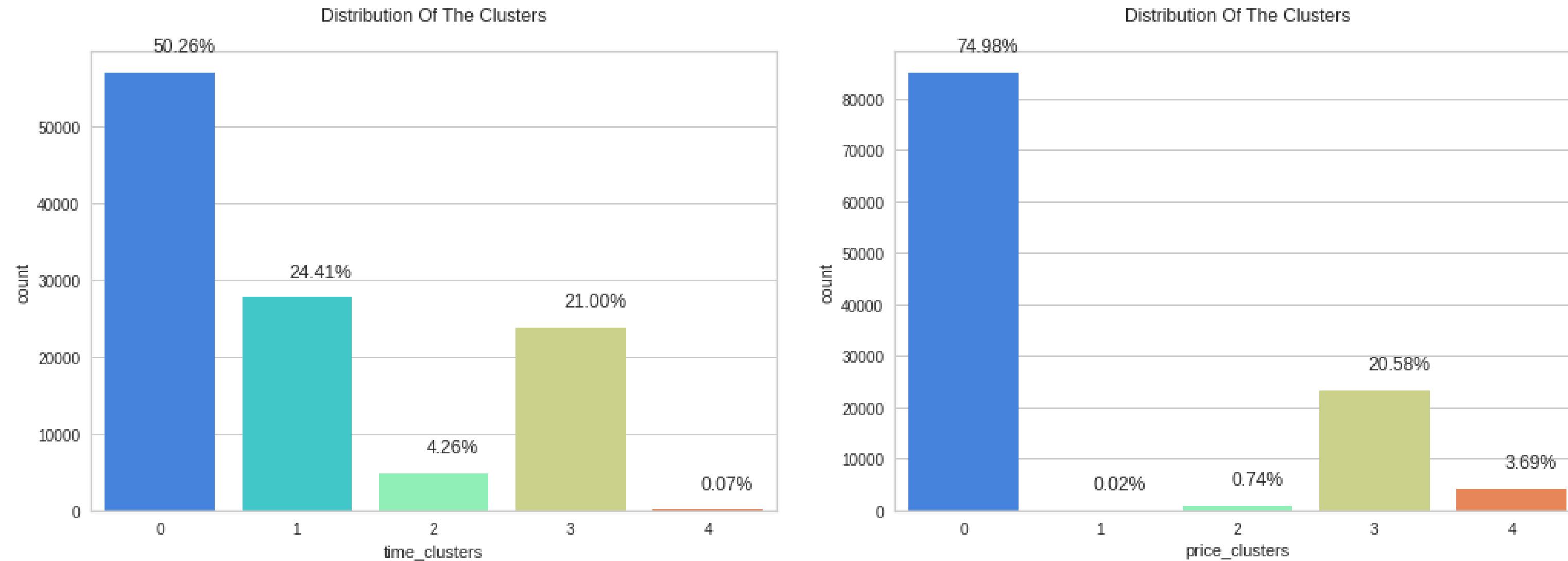
Elbow analysis objektif Financial dan Customer (K-Means Clustering)

# Implementasi Data Mining : 6) Pattern Evolution



Scatter Plot Clustering dataset Olist pada objektif Financial dan Customer  
(K-Means Clustering)

# Implementasi Data Mining : 7) Knowledge Presentation



Barplot perhitungan serta persentase segementasi dataset Olist pada objektif Financial dan Customer

**Link Notebook:**  
<https://datastudio.google.com/reporting/967357b4-b797-47d8-8dcc-3591ffff866f>



# Dashboard

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

Group G SI-44-01

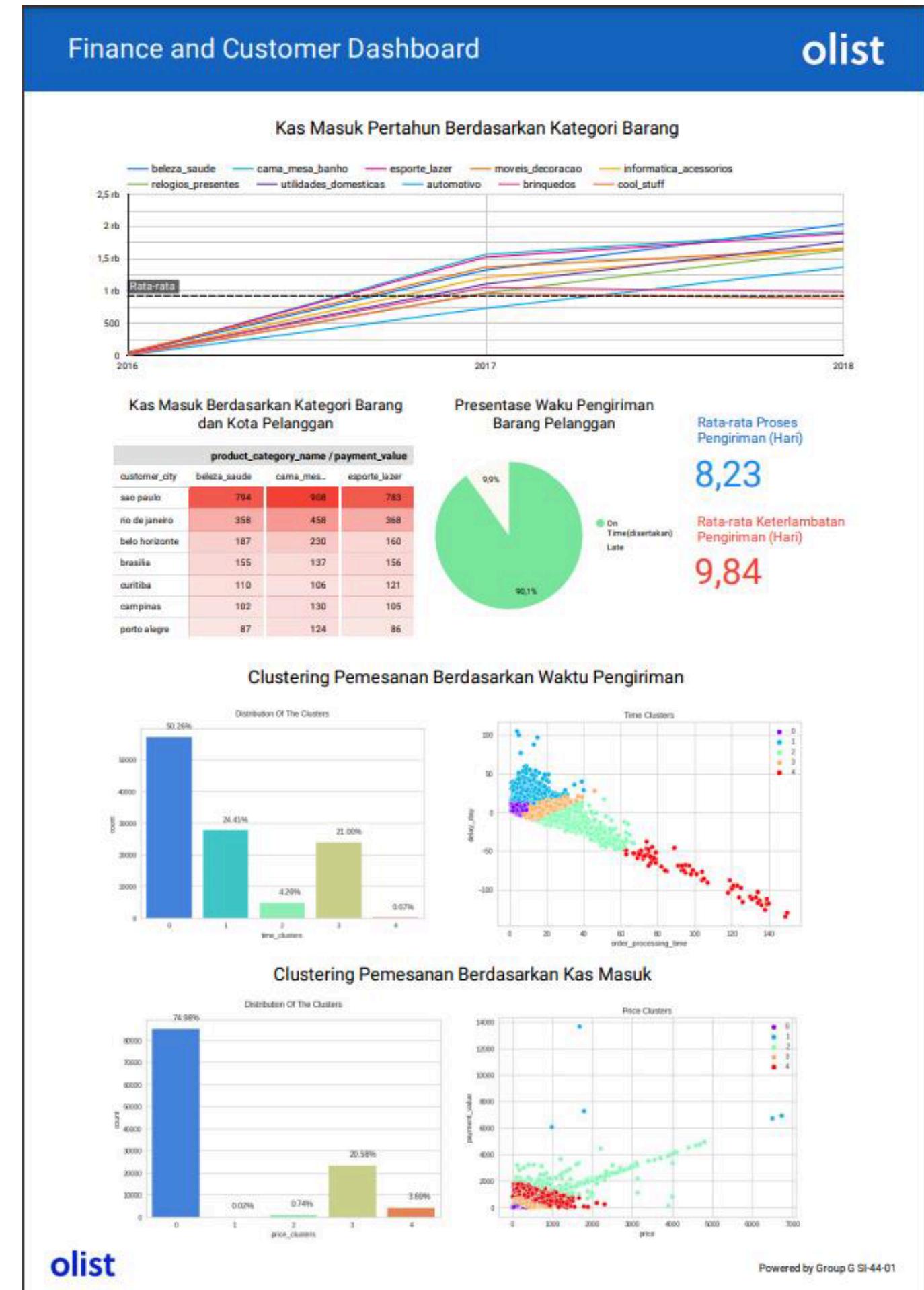
# Dashboard

Dashboard digunakan dalam data analytics untuk memberikan gambaran cepat tentang kinerja suatu sistem atau proses, atau untuk memberikan informasi yang relevan bagi para pemakai dalam membuat keputusan atau mengambil tindakan. Adapun tools yang digunakan dalam perancangan dashboard berdasarkan KPI perusahaan Olist yaitu menggunakan Google Data Studio.



# Dashboard

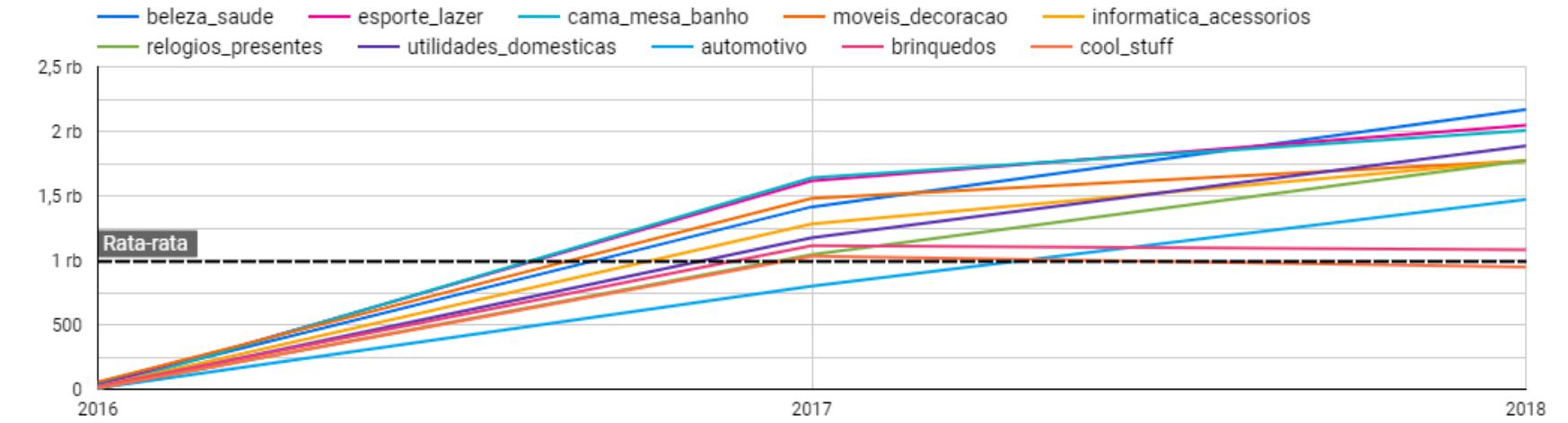
<https://datastudio.google.com/reporting/967357b4-b797-47d8-8dcc-3591ffff866f>



# Financial

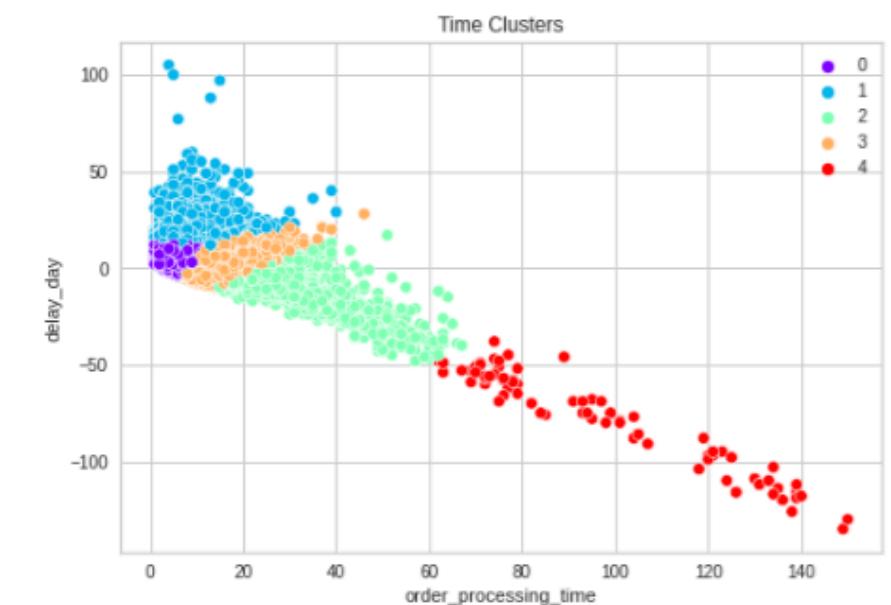
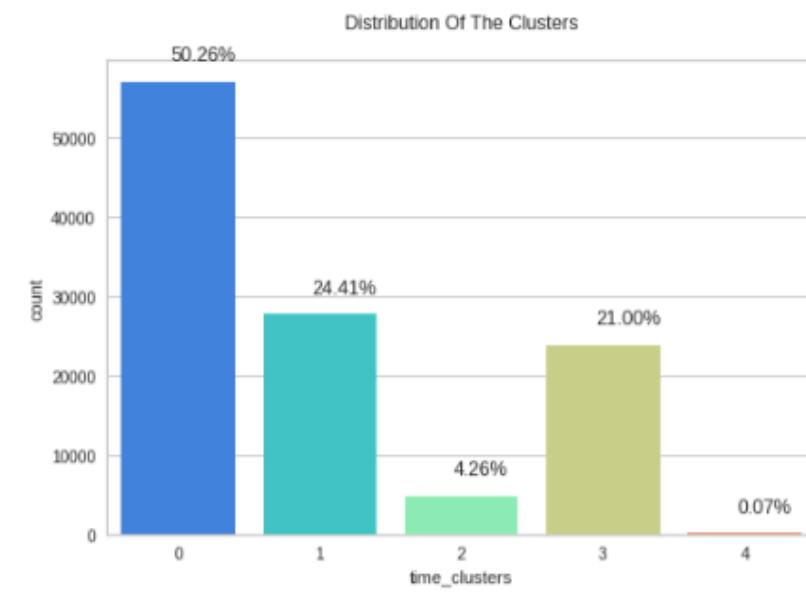
## KPI

- Rata-rata pemasukan 3 tahun terakhir
- % Persentase pemasukan orderan cluster tertinggi
- Pemasukan dari 3 produk paling laku di 3 kota dengan penjualan tertinggi



## Target Kinerja

- ↑ \$1000 per 3 tahun terakhir
- ↑ 5% per 3 tahun
- ↑ \$200 per 3 tahun



Kas Masuk Berdasarkan Kategori Barang dan Kota Pelanggan

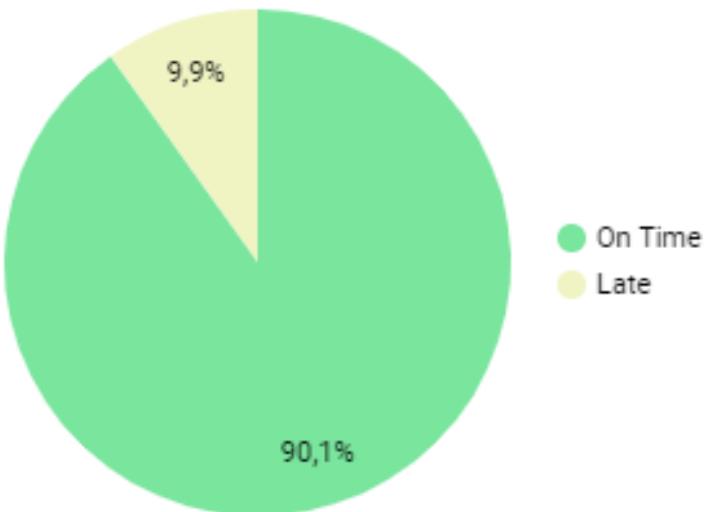
	product_category_name / payment_value			
customer_city	beleza_saude	esporte_lazer	cama_mes...	
sao paulo	847	836	967	
rio de janeiro	399	427	525	
belo horizonte	204	171	253	
brasilia	171	169	155	
curitiba	117	128	114	
campinas	116	114	142	
porto alegre	99	100	136	

# Customer

## KPI

- Rata-rata pemrosesan barang sampai ke pelanggan
- Rata-rata keterlambatan barang sampai ke customer
- % Persentase ketepatan waktu pengiriman barang pelanggan 3 tahun terakhir
- % Persentase keterlambatan waktu pengiriman barang pada cluster terakhir (paling lambat)

Presentase Waktu Pengiriman Barang Pelanggan



Rata-rata Proses Pengiriman (Hari)

9,38

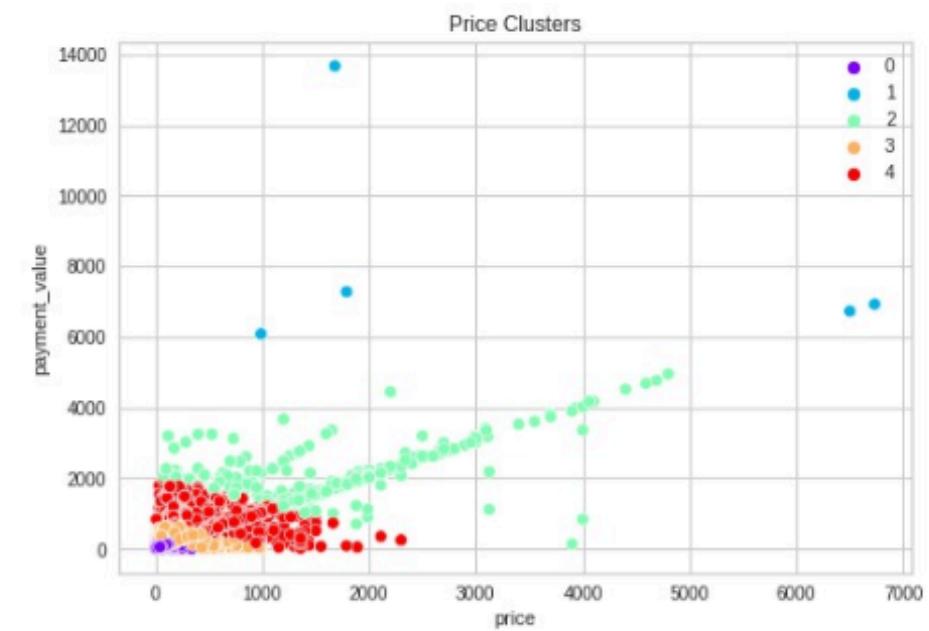
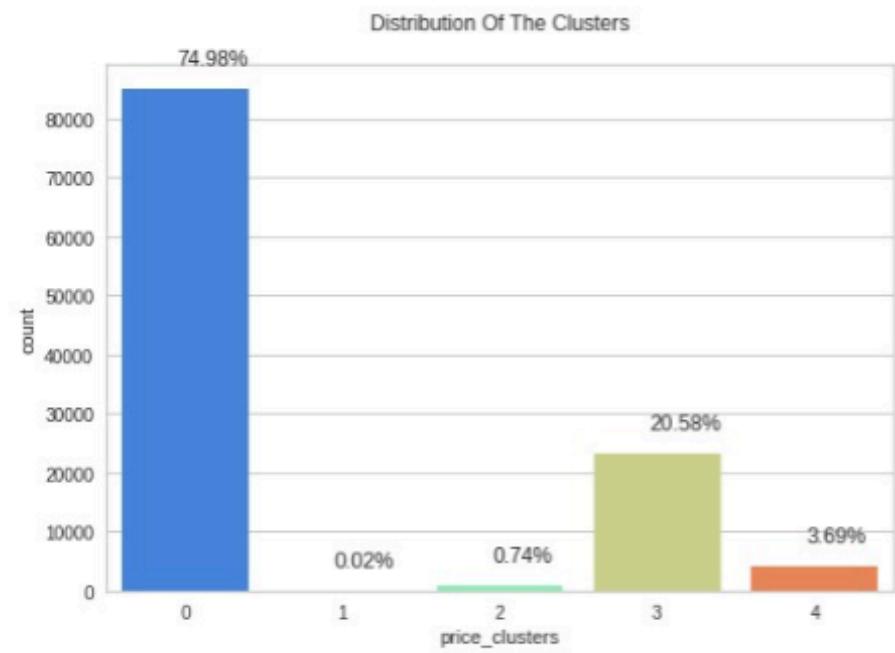
Rata-rata Keterlambatan Pengiriman (Hari)

8,5

## Target Kinerja

- ↓ 3 Hari pengantaran
- ↓ 1 Hari
- ↑ 90% Pengiriman barang tepat waktu
- ↓ 10% Pengiriman barang terlambat

Clustering Pemesanan Berdasarkan Kas Masuk





# Kesimpulan & Saran

DATA WAREHOUSE DAN  
BUSINESS INTELLIGENCE

Group G SI-44-01

# Kesimpulan

Berdasarkan hasil analisis dari pengimplementasian Data Warehouse, proses Data Mining, serta pembuatan Dashboard pada objektif Financial dan Customer perusahaan Olist dapat ditarik kesimpulan:

**01**

Perusahaan Olist cukup bagus dalam hal financial tiap tahunnya hanya saja untuk riset dan target tiap kategori produknya perlu ditingkatkan lagi.

**02**

Pemrosesan orderan customer dari Perusahaan Olist cukup melebih target dalam hal waktu pengantaran tetapi untuk waktu keterlambatan perlu diperhatikan lagi karena jauh melebihi target KPI.

## Saran

Dalam penyusunan tugas besar mata kuliah DWBI (Data Warehousing and Business Intelligence), kami menyarankan agar para mahasiswa menggunakan tools yang sama dalam proses pengembangan sistem (Data Warehouse, Data Mining, dan Dashboard). Penggunaan tools yang sama akan mempermudah dalam hal kolaborasi dan pemeliharaan kode.



**Terimakasih**

