LAPORAN UJIAN AKHIR SEMESTER

BUSINESS INTELLIGENCE

Ditulis oleh :

Achmad Anfasa Rabbany (2341720105)

Ericha Rizki Wardani (2341720202)

Hizkia Elsadanta (2341720253)

Nakita Gayuh C (2341720181)

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perusahaan retail berskala global umumnya memiliki data transaksi dalam jumlah besar yang berasal dari berbagai wilayah, kategori produk, dan segmen pelanggan. Pada studi kasus Retail dataset of a global superstore for 4 years, data transaksi penjualan tersimpan dalam bentuk data mentah (raw) yang berisi detail order, pelanggan, produk, waktu transaksi, serta nilai penjualan.

Agar data tersebut dapat digunakan untuk analisis bisnis, diperlukan proses ETL (Extract, Transform, Load) untuk mengintegrasikan data dari sumber mentah, membersihkan data (misalnya perbaikan tipe tanggal/angka dan pengurangan duplikasi), serta menstandarisasi format agar konsisten. Setelah itu, data perlu disimpan dalam Data Warehouse dengan desain yang sesuai untuk analisis, seperti Star Schema, sehingga perhitungan KPI dapat dilakukan lebih cepat, terstruktur, dan akurat. Dengan data warehouse, proses analitik seperti melihat tren total penjualan per periode dan produk terlaris dapat dilakukan secara efisien dan divisualisasikan dalam dashboard untuk mendukung pengambilan keputusan.

1.2 Tujuan Proyek

Tujuan dari proyek ini adalah:

1. Membangun pipeline ETL menggunakan Pentaho untuk mengekstrak data retail global superstore, melakukan transformasi (pembersihan, konversi tipe data, deduplikasi, dan pembentukan struktur dimensi-fakta), lalu memuatnya ke database.
2. Membangun data warehouse menggunakan MySQL dengan desain Star Schema yang terdiri dari tabel fakta dan tabel dimensi untuk mendukung analisis penjualan.
3. Membuat dashboard KPI menggunakan Looker Studio guna menampilkan hasil analisis penjualan dalam bentuk visual yang interaktif dan mudah dipahami.

1.3 Ruang Lingkup

Ruang lingkup proyek ini meliputi:

- Data: Dataset retail global superstore selama 4 tahun (2015–2018) yang memuat informasi transaksi penjualan seperti order ID, tanggal order/pengiriman, customer, produk, kategori, segmen, wilayah, serta nilai penjualan (sales).
- Proses: Implementasi ETL dari data mentah (CSV) menuju data warehouse di MySQL, serta perancangan struktur Star Schema (tabel dimensi dan tabel fakta).

- Output Analisis (KPI):
    1. Total penjualan per periode tertentu (misalnya per bulan/per tahun) untuk melihat tren penjualan.
    2. Top 5 produk terlaris berdasarkan total nilai penjualan (sales).
- Tools yang digunakan:
    1. Pentaho Data Integration (PDI) untuk proses ETL,
    2. MySQL Database untuk penyimpanan data warehouse (Star Schema),
    3. Looker Studio untuk visualisasi dashboard dan analisis KPI.

BAB II
STUDI KASUS & DATASET

2.1 Studi Kasus Pengelolaan Data

Studi kasus pada proyek ini adalah retail dataset of a global superstore selama 4 tahun (2015–2018). Perusahaan global superstore menjual berbagai jenis produk (misalnya technology, office supplies, dan furniture) kepada beberapa segmen pelanggan (consumer, corporate, home office) dan melayani pengiriman ke berbagai wilayah. Kebutuhan analisis pada studi kasus ini adalah memantau performa penjualan dan memperoleh insight melalui indikator bisnis, seperti:

- Bagaimana tren total penjualan pada periode tertentu (per bulan/per tahun),
- Produk apa yang menjadi penyumbang penjualan terbesar (top product),
- Distribusi kontribusi penjualan berdasarkan kategori/segmen/wilayah (untuk evaluasi strategi penjualan).

Karena data transaksi bersifat detail dan jumlahnya besar, proses analisis membutuhkan pengelolaan data yang terstruktur menggunakan ETL dan data warehouse agar data siap dipakai untuk perhitungan KPI dan visualisasi dashboard.

2.2 Sumber Data Dummy

- Sumber dataset: Dataset “Global Superstore / Superstore Sales” yang tersedia di Kaggle
- Format data: CSV
- Cara memperoleh data: Dataset diperoleh dengan cara mengunduh (download) file CSV dari sumber publik, lalu digunakan sebagai data mentah (raw) untuk proses ETL.

2.3 Struktur Data (Data Understanding)

2.3.1 Daftar Tabel / Dataset yang Digunakan

Dataset awal hanya berupa 1 tabel transaksi (CSV). Namun, untuk memenuhi kebutuhan data warehouse (Star Schema), tabel transaksi tersebut dipecah saat proses ETL/DWH menjadi beberapa tabel, yaitu:

1. Dim_customer (dimensi pelanggan)
2. Dim_product (dimensi produk)
3. Dim_time (dimensi waktu)
4. Fact_sales (table fakta pelanggan)

Table	Action	Rows	Type	Collation	Size	Overhead
dim_customer	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_general_ci	16.0 KiB	-
dim_product	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_general_ci	16.0 KiB	-
dim_time	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_general_ci	16.0 KiB	-
fact_sales	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_general_ci	48.0 KiB	-
4 tables	Sum	0	InnoDB	utf8mb4_0900_ai_ci	96.0 KiB	0 B

Rasional pemecahan ini adalah agar data siap dianalisis secara cepat dan terstruktur: tabel fakta menyimpan nilai numerik (sales), sedangkan tabel dimensi menyimpan atribut deskriptif (customer, product, time).

2.3.2 Data Dictionary Ringkas (kolom penting)

Berikut ringkasan kolom-kolom utama yang akan digunakan dari dataset transaksi:

- Order information:
  - Order ID: kode unik order/transaksi
  - Order Date: tanggal pemesanan
  - Ship Date: tanggal pengiriman
  - Ship Mode: metode pengiriman
- Customer Information:
  - Customer ID: kode unik pelanggan
  - Customer Name: nama pelanggan
  - Segment: segmen pelanggan (Consumer/Corporate/Home Office)
- Product Information:
  - Product ID: kode unik produk
  - Product Name: nama produk
  - Category: kategori produk
  - Sub-Category: sub kategori produk
- Location Information:
  - Country, Region, State, City, Postal Code
- Measure:
  - Sales: nilai penjualan (digunakan sebagai metrik utama untuk KPI)

BAB III
PERANCANGAN DATA WAREHOUSE

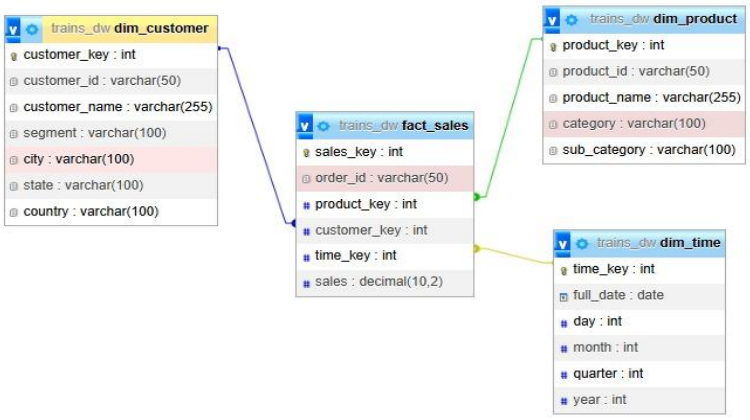
3.1 Konsep & Desain Star Schema

Data warehouse pada proyek ini dirancang menggunakan Star Schema, yaitu model data analitik yang terdiri dari tabel fakta (fact table) di pusat dan beberapa tabel dimensi (dimension tables) di sekelilingnya. Star schema dipilih karena struktur ini memudahkan proses analisis dan pelaporan KPI, serta mendukung query yang lebih sederhana dan cepat untuk kebutuhan dashboard. Pada studi kasus global superstore retail, tabel fakta merepresentasikan transaksi penjualan, sedangkan tabel dimensi menyimpan atribut deskriptif seperti informasi pelanggan, produk, dan waktu. Dengan pemisahan ini, proses agregasi (misalnya total sales per bulan atau top product) dapat dilakukan dengan efisien.

Grain (tingkat detail) tabel fakta, satu baris pada fact_sales merepresentasikan satu transaksi penjualan pada tingkat order line (per produk dalam suatu order). Dengan grain ini, analisis dapat dilakukan secara fleksibel, baik pada level harian/bulanan/tahunan, maupun per produk atau per pelanggan.

3.2 Skema Star Schema

Skema star schema yang digunakan terdiri dari 4 tabel utama:



Dengan relasi tersebut, fact_sales menjadi pusat analisis, sedangkan tabel dimensi menyediakan konteks untuk slicing/dicing data (misalnya filter berdasarkan tahun, kategori produk, atau segment customer).

3.3 Data Mart / Tabel yang Dibangun

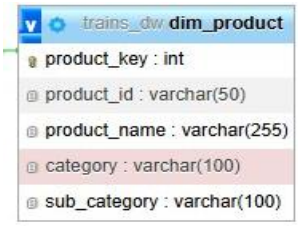
3.3.1 Dim_customer (Dimensi Customer)

Fungsi: menyimpan informasi pelanggan untuk analisis berdasarkan customer dan segment.

Aturan: customer_key sebagai primary key.

3.3.2 Dim_product (Dimensi Product)

Fungsi: menyimpan informasi produk untuk analisis berdasarkan kategori dan produk.



Aturan: product_key sebagai primary key.

3.3.3 Dim_time (Dimensi Time)

Fungsi: menyimpan atribut waktu agar analisis dapat dilakukan per hari/bulan/kuartal/tahun.

Aturan: time_key sebagai primary key.

3.3.4 Fact_sales (Fakta Penjualan)

Fungsi: menyimpan transaksi penjualan (measure) serta penghubung ke dimensi.

Aturan relasi (Foreign Key):

- fact_sales.customer_key REFERENCES dim_customer(customer_key)
- fact_sales.product_key REFERENCES dim_product(product_key)
- fact_sales.time_key REFERENCES dim_time(time_key)

Measure sales digunakan sebagai metrik utama untuk perhitungan KPI:

- Total penjualan per periode: SUM(sales)
- Top 5 produk terlaris: SUM(sales) dikelompokkan per produk

BAB IV
IMPLEMENTASI HASIL ANALISIS

4.1 Implementasi Database (MySQL)

4.1.1 Pembuatan Database

- Nama database: traind_dw
- Tujuan database: menampung tabel dimensi dan fakta untuk kebutuhan analitik.

4.1.2 Pembuatan Tabel Data Warehouse

- Dim_customer

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	customer_key	int			No	None		AUTO_INCREMENT	Change Drop More
2	customer_id	varchar(50)	utf8mb4_general_ci		Yes	NULL			Change Drop More
3	customer_name	varchar(255)	utf8mb4_general_ci		Yes	NULL			Change Drop More
4	segment	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More
5	city	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More
6	state	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More
7	country	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More

- Dim_product

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	product_key	int			No	None		AUTO_INCREMENT	Change Drop More
2	product_id	varchar(50)	utf8mb4_general_ci		Yes	NULL			Change Drop More
3	product_name	varchar(255)	utf8mb4_general_ci		Yes	NULL			Change Drop More
4	category	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More
5	sub_category	varchar(100)	utf8mb4_general_ci		Yes	NULL			Change Drop More

- Dim_time

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	time_key	int			No	None		AUTO_INCREMENT	Change Drop More
2	full_date	date			Yes	NULL			Change Drop More
3	day	int			Yes	NULL			Change Drop More
4	month	int			Yes	NULL			Change Drop More
5	quarter	int			Yes	NULL			Change Drop More
6	year	int			Yes	NULL			Change Drop More

- Fact_sales

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	sales_key	int			No	None		AUTO_INCREMENT	Change Drop More
2	order_id	varchar(50)	utf8mb4_general_ci		Yes	NULL			Change Drop More
3	product_key	int			Yes	NULL			Change Drop More
4	customer_key	int			Yes	NULL			Change Drop More
5	time_key	int			Yes	NULL			Change Drop More
6	sales	decimal(10,2)			Yes	NULL			Change Drop More

4.1.3 Validasi Struktur

- Relasi Foreign Key

Foreign key constraints

Actions	Constraint properties	Column	Foreign key constraint (INNODB)		
			Database	Table	Column
Drop	fact_sales_ibfk_1 ON DELETE RESTRICT ON UPDATE RESTRICT	product_key + Add column	trains_dw	dim_product	product_key
Drop	fact_sales_ibfk_2 ON DELETE RESTRICT ON UPDATE RESTRICT	customer_key + Add column	trains_dw	dim_customer	customer_key
Drop	fact_sales_ibfk_3 ON DELETE RESTRICT ON UPDATE RESTRICT	time_key + Add column	trains_dw	dim_time	time_key

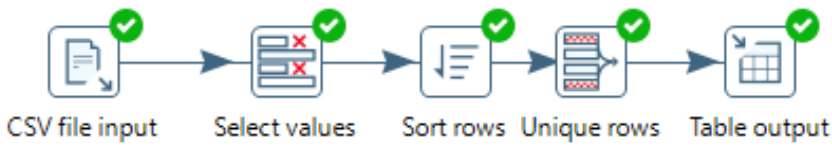
4.2 Perancangan & Implementasi ETL (Pentaho)

Implementasi ETL dilakukan menggunakan *tools* Pentaho Data Integration (Spoon). Proses ini bertujuan untuk memindahkan data mentah dari file CSV ke dalam Data Warehouse dengan struktur *Star Schema*.

4.2.1 Implementasi ETL Tabel Dimensi

Sebelum mengisi tabel fakta, proses ETL dilakukan terlebih dahulu pada tabel dimensi (dim_customer, dim_product, dan dim_time) untuk membentuk data referensi dan menghasilkan *Primary Key* (SK).

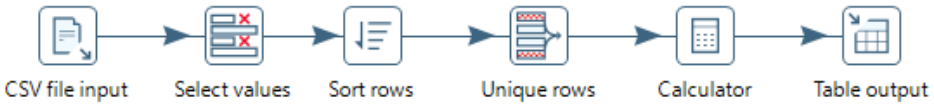
a. Alur Transformasi Dimensi (Contoh: dim_customer & dim_product) Proses pengisian tabel dimensi pelanggan dan produk menggunakan alur sebagai berikut:



Langkah-langkah proses:

1. Extract (CSV File Input): Mengambil data mentah dari file train.csv.
2. Transform (Select Values): Memilih kolom yang diperlukan dan menyesuaikan tipe data.
3. Transform (Sort & Unique Rows): Mengurutkan data dan menghapus duplikasi agar setiapdata hanya muncul satu kali .
4. Load (Table Output): Memasukkan data bersih ke tabel dim_customer dan dim_product di database.

b. Alur Transformasi Dimensi Waktu (dim_time) Khusus untuk dimensi waktu, dilakukan ekstraksi komponen tanggal.



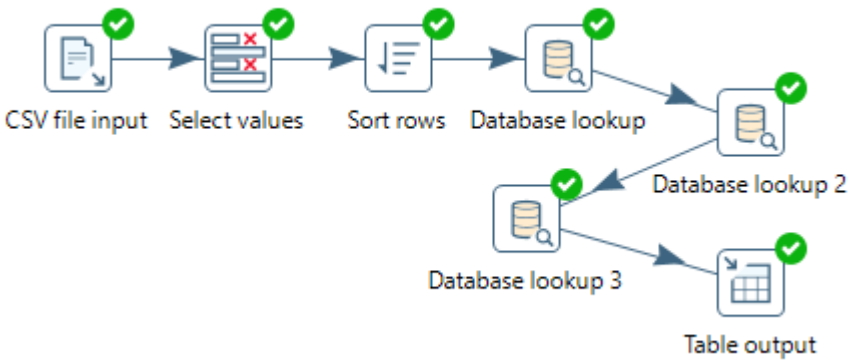
Langkah-langkah proses:

1. Extract: Menggunakan CSV File Input untuk mengambil data transaksi, khususnya kolom Order Date.
2. Select Values (Data Formatting): Mengubah tipe data kolom Order Date dari String menjadi Date dengan format dd/MM/yyyy. Langkah ini krusial agar fungsi kalkulasi tanggal dapat berjalan.
3. Sort & Unique Rows: Mengurutkan data berdasarkan Order Date dan menghapus duplikasi. Tujuannya adalah membuat daftar tanggal referensi unik (kalender) di mana satu tanggal hanya diwakili oleh satu baris data.
4. Calculator (Date Extraction): Menggunakan step Calculator untuk memecah Order Date menjadi atribut terpisah:
   - Day: Mengambil tanggal (hari ke-n).
   - Month: Mengambil bulan (angka 1-12).
   - Year: Mengambil tahun (4 digit).
   - Quarter: Mengambil kuartal (1-4).
5. Load: Memasukkan hasil ekstraksi tersebut ke dalam tabel dim_time. Kolom time_key dibiarkan kosong karena akan diisi otomatis (Auto Increment) oleh database.

4.2.2 Implementasi ETL Tabel Fact Sales

Setelah tabel dimensi terisi, dilakukan proses pengisian tabel fakta fact_sales. Tabel ini menyimpan data transaksi penjualan dan kunci asing (*Foreign Keys*) yang menghubungkan ke tabel dimensi.

Alur Transformasi Fact Sales:



Penjelasan langkah-langkah transformasi:

1. Extract (CSV File Input): Membaca data transaksi dari sumber train.csv. Kolom yang diambil disederhanakan menjadi: Order ID, Order Date, Customer ID, Product ID, dan Sales.

2. Transform (Data Type Formatting): Menggunakan step Select Values untuk memastikan kolom Order Date memiliki format Date yang sesuai agar bisa dicocokkan dengan database.

3. Sort & Unique Rows: Mengurutkan data berdasarkan Order ID

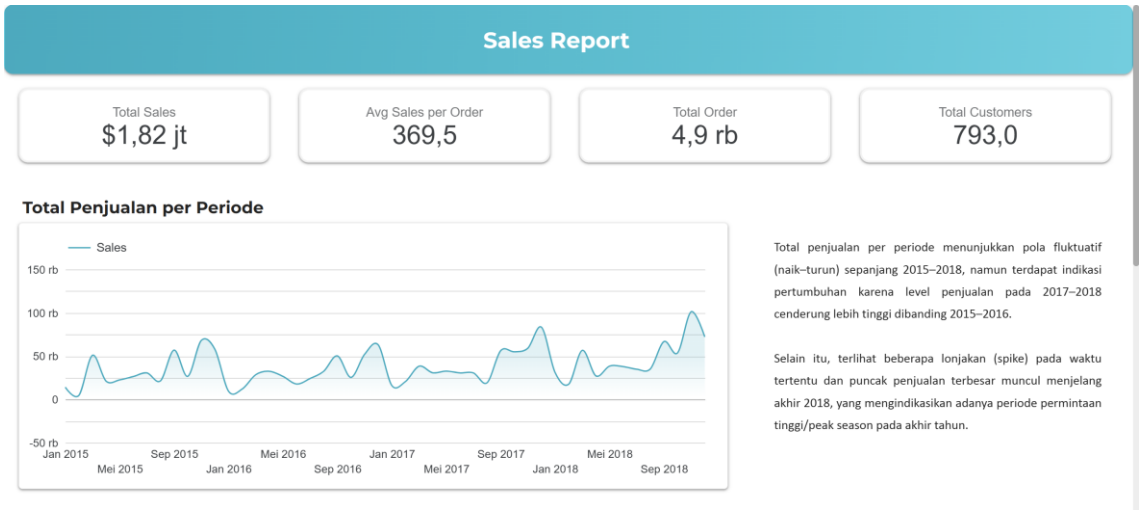
4. Lookup Process: Dilakukan proses Database Lookup untuk menukar ID operasional menjadi Surrogate Key dari Data Warehouse:

    - Lookup 1: Mencocokkan Customer ID (CSV) dengan dim_customer untuk mendapatkan customer_key.
    - Lookup 2: Mencocokkan Product ID (CSV) dengan dim_product untuk mendapatkan product_key.
    - Lookup 3: Mencocokkan Order Date (CSV) dengan full_date di dim_time untuk mendapatkan time_key.

5. Load (Table Output):

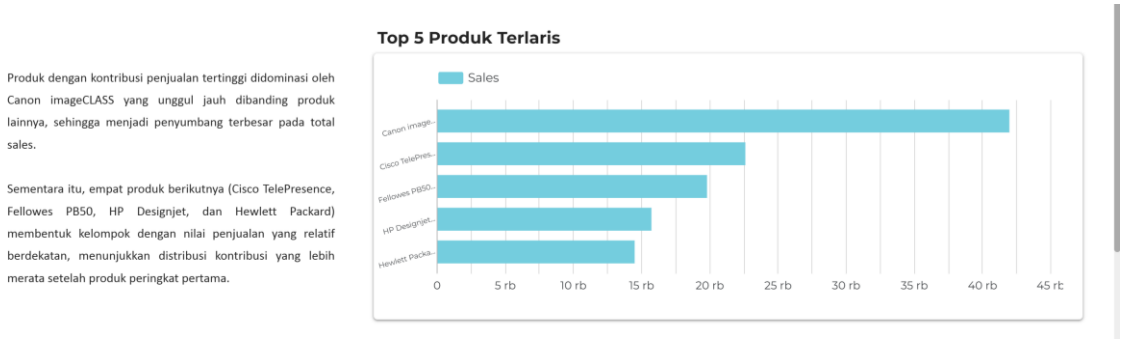
Menyimpan hasil akhir ke tabel fact_sales. Mapping kolom dilakukan

4.3 Visualisasi Dashboard



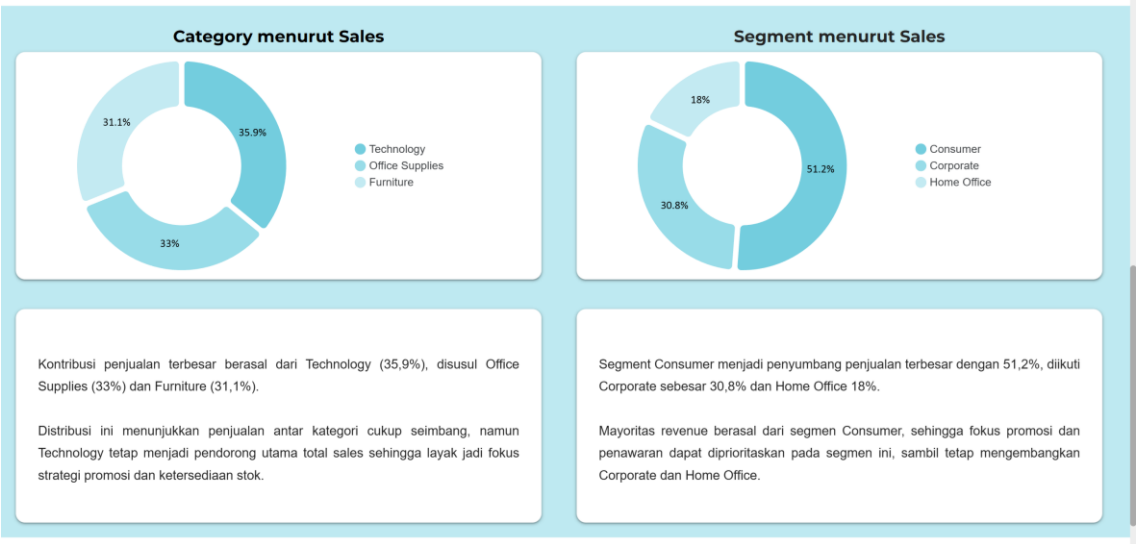
Gambar tersebut menunjukkan tampilan dashboard “Sales Report” pada Looker Studio yang merangkum performa penjualan dataset Global Superstore periode 2015–2018. Pada bagian atas terdapat 4 KPI utama dalam bentuk scorecard, yaitu Total Sales sebesar $1,82 jt, Average Sales per Order sebesar 369,5, Total Order sebesar 4,9 rb, dan Total Customers sebanyak 793. KPI ini memberikan ringkasan cepat mengenai total pendapatan, rata-rata nilai transaksi, jumlah transaksi yang terjadi, serta jumlah pelanggan unik.

Di bagian bawah, terdapat grafik “Total Penjualan per Periode” yang menampilkan tren penjualan dari waktu ke waktu. Grafik menunjukkan pola penjualan yang fluktuatif (naik–turun) namun memiliki kecenderungan meningkat pada periode 2017–2018 dibandingkan 2015–2016, serta terlihat beberapa lonjakan (spike) pada waktu tertentu. Di sisi kanan grafik, disertakan narasi insight yang menjelaskan interpretasi tren tersebut, termasuk indikasi adanya periode permintaan tinggi (peak season) menjelang akhir tahun 2018.



Gambar tersebut menampilkan visualisasi “Top 5 Produk Terlaris” dalam bentuk bar chart horizontal pada dashboard Looker Studio. Grafik ini membandingkan total penjualan (Sales) dari lima produk dengan nilai penjualan tertinggi pada periode analisis.

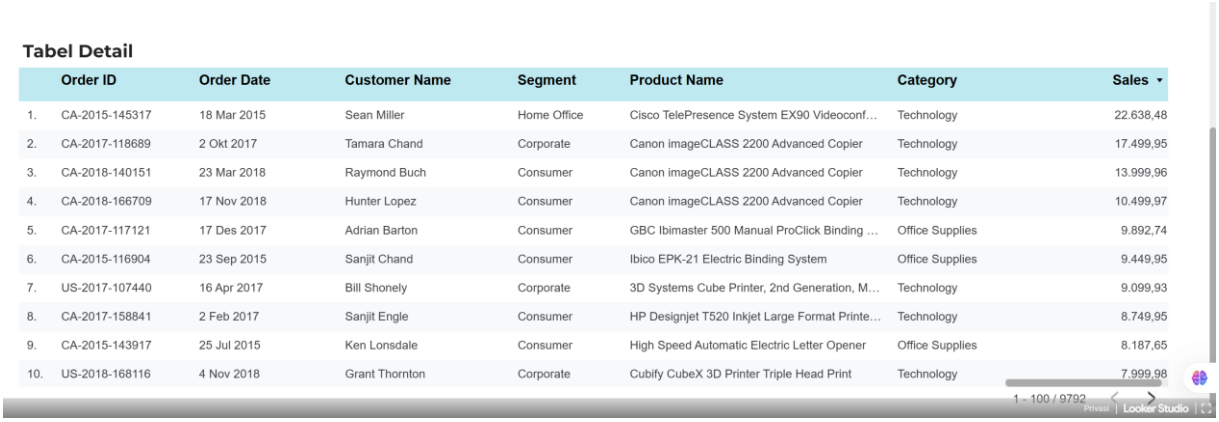
Dari hasil visualisasi terlihat bahwa Canon imageCLASS memiliki total penjualan paling tinggi dan jaraknya cukup signifikan dibanding empat produk lain, sehingga menjadi kontributor terbesar terhadap penjualan. Sementara itu, empat produk berikutnya (Cisco TelePresence, Fellowes PB50, HP Designjet, dan Hewlett Packard) memiliki nilai penjualan yang relatif lebih berdekatan, yang menunjukkan kontribusi penjualan yang lebih merata setelah produk peringkat pertama.



Gambar tersebut menampilkan dua visualisasi distribusi penjualan dalam bentuk donut chart, yaitu Category menurut Sales (kiri) dan Segment menurut Sales (kanan), lengkap dengan narasi insight di bawah masing-masing grafik.

Pada grafik Category menurut Sales, kontribusi penjualan terbesar berasal dari Technology (35,9%), diikuti Office Supplies (33%) dan Furniture (31,1%). Komposisi ini menunjukkan distribusi penjualan antar kategori relatif seimbang, namun kategori Technology tetap menjadi penyumbang utama sehingga dapat dijadikan fokus strategi promosi dan pengelolaan stok.

Pada grafik Segment menurut Sales, segmen Consumer menjadi kontributor terbesar dengan 51,2%, diikuti Corporate (30,8%) dan Home Office (18%). Hal ini mengindikasikan bahwa penjualan paling banyak didorong oleh pelanggan individu (Consumer), sehingga strategi pemasaran dan penawaran dapat diprioritaskan ke segmen ini sambil tetap mengembangkan potensi penjualan pada segmen Corporate dan Home Office.



Tabel Detail

	Order ID	Order Date	Customer Name	Segment	Product Name	Category	Sales
1.	CA-2015-145317	18 Mar 2015	Sean Miller	Home Office	Cisco TelePresence System EX90 Videoconf…	Technology	22.638,48
2.	CA-2017-118689	2 Okt 2017	Tamara Chand	Corporate	Canon imageCLASS 2200 Advanced Copier	Technology	17.499,95
3.	CA-2018-140151	23 Mar 2018	Raymond Buch	Consumer	Canon imageCLASS 2200 Advanced Copier	Technology	13.999,96
4.	CA-2018-166709	17 Nov 2018	Hunter Lopez	Consumer	Canon imageCLASS 2200 Advanced Copier	Technology	10.499,97
5.	CA-2017-117121	17 Des 2017	Adrian Barton	Consumer	GBC Ibimaster 500 Manual ProClick Binding …	Office Supplies	9.892,74
6.	CA-2015-116904	23 Sep 2015	Sanjit Chand	Consumer	Ibico EPK-21 Electric Binding System	Office Supplies	9.449,95
7.	US-2017-107440	16 Apr 2017	Bill Shonely	Corporate	3D Systems Cube Printer, 2nd Generation, M…	Technology	9.099,93
8.	CA-2017-158841	2 Feb 2017	Sanjit Engle	Consumer	HP Designjet T520 Inkjet Large Format Printe…	Technology	8.749,95
9.	CA-2015-143917	25 Jul 2015	Ken Lonsdale	Consumer	High Speed Automatic Electric Letter Opener	Office Supplies	8.187,65
10.	US-2018-168116	4 Nov 2018	Grant Thornton	Corporate	Cubify CubeX 3D Printer Triple Head Print	Technology	7.999,98

1 – 100 / 9792

Gambar tersebut menampilkan Tabel Detail pada dashboard Looker Studio yang berfungsi sebagai tampilan data transaksi penjualan secara rinci (drill-down) untuk mendukung grafik dan KPI pada dashboard.

Tabel memuat atribut utama transaksi, yaitu Order ID, Order Date, Customer Name, Segment, Product Name, Category, dan Sales. Melalui tabel ini, pengguna dapat memverifikasi nilai yang membentuk agregasi pada visualisasi (misalnya Total Sales atau Top 5 Produk), melihat

transaksi dengan nilai penjualan tinggi, serta melakukan analisis lebih detail berdasarkan pelanggan, segmen, kategori, maupun produk tertentu. Selain itu, tabel mendukung navigasi data (pagination) sehingga pengguna dapat menelusuri banyaknya record transaksi yang tersedia.

BAB V
KESIMPULAN

5.1 Kesimpulan

Berdasarkan proyek yang telah dikerjakan, proses ETL menggunakan Pentaho berhasil dilakukan untuk mengekstrak data retail global superstore (CSV), melakukan transformasi (konversi tipe data, deduplikasi, dan pembentukan key), serta memuatnya ke database. Selanjutnya, data warehouse di MySQL berhasil dibangun menggunakan star schema yang terdiri dari tabel dimensi (dim_customer, dim_product, dim_time) dan tabel fakta (fact_sales) dengan relasi melalui foreign key. Hasil data warehouse kemudian berhasil diintegrasikan ke Looker Studio untuk membangun dashboard yang menampilkan KPI utama, seperti total penjualan per periode dan top 5 produk terlaris, serta visualisasi pendukung (kategori dan segmen). Dengan demikian, dashboard mampu memberikan ringkasan performa penjualan dan mempermudah analisis data secara cepat dan interaktif.

Link dashboard:
https://lookerstudio.google.com/embed/reporting/de99cd2d-5b95-4f58-a091-e2756ca6ef4f/page/vKqhF

link dataset:
https://docs.google.com/spreadsheets/d/1D4gwDAZoBI3GHmeFxRgqFaIoaL8f3Eem8UZYGXbOzAI/edit?usp=sharing

link github:
https://github.com/Ericharw/PBL_Kelompok_6/tree/main/Business_Intelligence

link file ktr:
https://drive.google.com/drive/folders/18pPve4MhhAoHmEv0G3mOq2mhiS3FPt1d?usp=sharing