

**Laporan Studi Kasus Proyek
Business Intelegence**



Ditulis oleh :

Fahmi Yahya	(2341720089)
Fiera Ziadattun Nisa'	(2341720220)
Nabeel Nizam	(2341720155)
Rafiqo Adib D.	(2341720214)

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2025**

I. Studi Kasus

Dalam industri e-commerce, kualitas layanan menjadi faktor yang sangat menentukan tingkat kepuasan pelanggan. Salah satu indikator utama yang sering dievaluasi adalah ketepatan waktu pengiriman, karena proses pengiriman merupakan fase akhir yang sangat memengaruhi persepsi pelanggan terhadap keseluruhan layanan. Pada proyek Business Intelligence ini, dilakukan analisis mendalam untuk mengetahui bagaimana performansi pengiriman memengaruhi tingkat kepuasan pelanggan, yang direpresentasikan melalui review score yang diberikan pelanggan setelah transaksi selesai.

Studi kasus ini menggunakan dataset [*Brazilian E-Commerce Public Dataset*](#) yang berisi data transaksi, data pengiriman, data ulasan pelanggan, dan data detail produk, dan lain – lain. Dari sekian banyak table, dipilih ada empat tabel utama, diantaranya:

- *orders* : menyimpan informasi waktu pemesanan, waktu pengiriman, dan waktu penerimaan barang
- *order_items* : menyimpan detail barang per transaksi, termasuk estimasi waktu pengiriman kurir
- *order_reviews* : menyimpan skor review (1–5), komentar, dan waktu evaluasi
- *products* : menyimpan kategori dan karakteristik produk

II. Rumusan Masalah

Berdasarkan studi kasus di atas, maka rumusan masalah yang dapat dikaji adalah sebagai berikut:

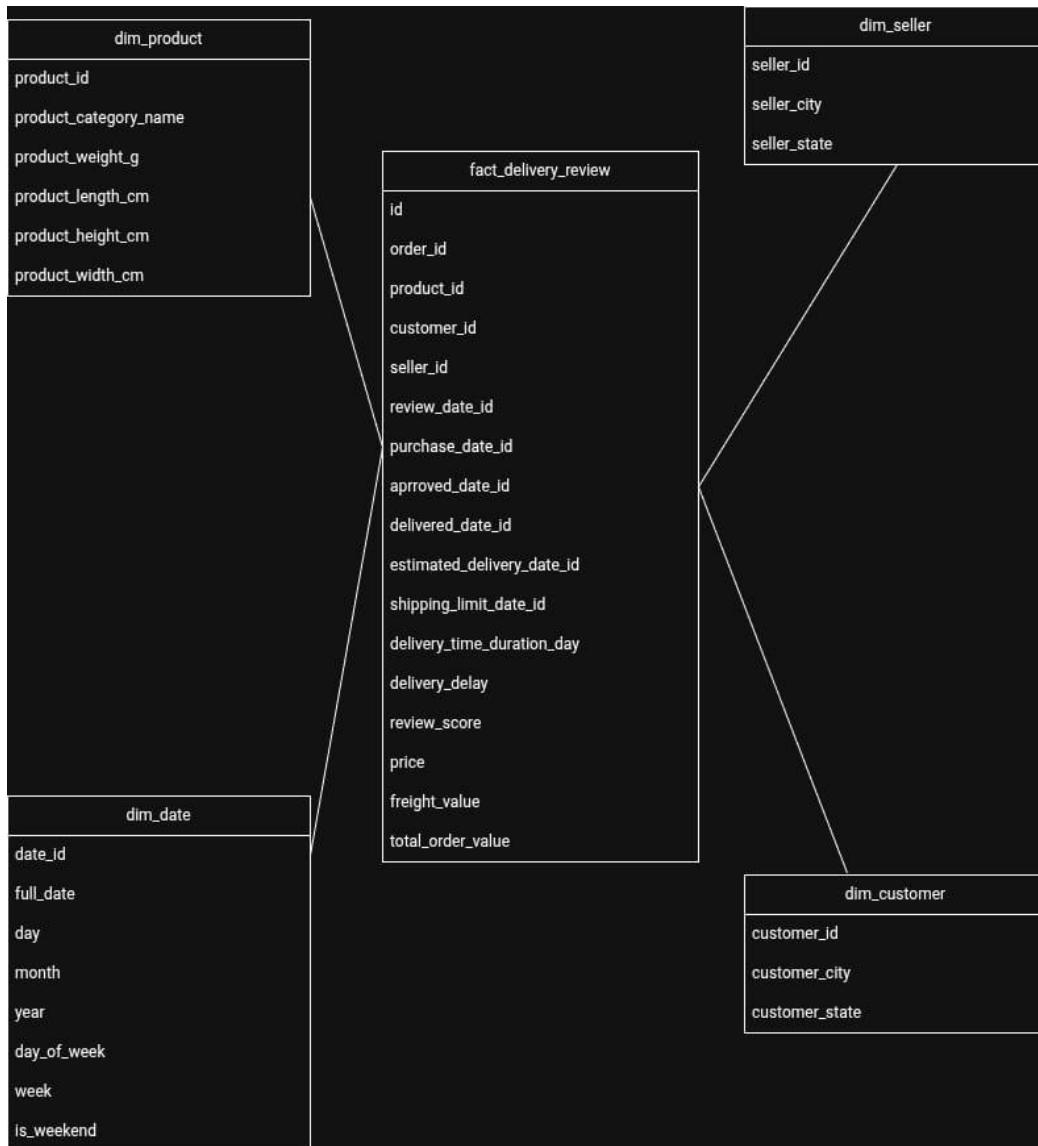
1. Apakah keterlambatan pengiriman berpengaruh terhadap rating yang diberikan pelanggan?
2. Kategori produk apa yang paling sering mendapatkan ulasan (review) dengan rating rendah?
3. Berapa rata-rata keterlambatan pengiriman pada setiap kategori produk?

III. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memberikan jawaban dari setiap rumusan masalah, yaitu:

1. Menilai pengaruh keterlambatan pengiriman terhadap rating yang diberikan pelanggan, sehingga dapat diketahui apakah waktu pengiriman merupakan faktor dominan dalam kepuasan pelanggan.
2. Mengidentifikasi kategori produk yang paling sering memperoleh rating rendah, sehingga bisnis dapat melakukan evaluasi kualitas produk maupun mitra logistik.
3. Menghitung rata-rata keterlambatan pengiriman berdasarkan kategori produk, untuk mengetahui pola lama pengiriman pada jenis barang tertentu dan menentukan area perbaikan yang lebih tepat.

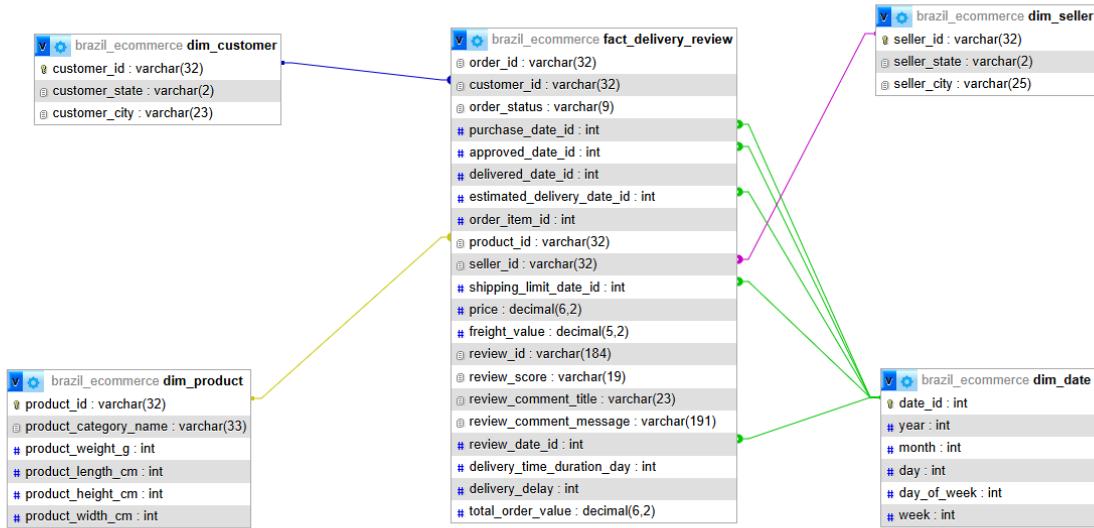
IV. Rancangan Skema Bintang (Star Schema) untuk Data Warehouse



Berdasarkan gambar skema diatas, tabel fakta yang digunakan adalah *fact_delivery_review*, yang berisi catatan lengkap terkait proses pemesanan, pengiriman, dan ulasan pelanggan. Tabel ini memuat berbagai atribut penting seperti waktu pembelian, waktu pengiriman, penilaian pelanggan, harga produk, nilai ongkos kirim, serta informasi kuantitatif lain yang berfungsi sebagai measures untuk keperluan analisis performa pengiriman dan kepuasan pelanggan.

Untuk memberikan konteks deskriptif terhadap data pada tabel fakta, digunakan empat tabel *dimensi*, yaitu *dim_product*, *dim_customer*, *dim_seller*, dan *dim_date*. Dimensi produk berisi informasi detail mengenai produk seperti kategori, berat, serta ukuran fisik. Dimensi pelanggan mencakup data wilayah pelanggan, sedangkan dimensi penjual menyimpan informasi lokasi

penjual. Sementara itu, dimensi waktu memuat atribut tanggal secara terstruktur mencakup hari, bulan, tahun, minggu, dan indikator hari kerja atau akhir pekan. Keempat dimensi ini berfungsi untuk memperkaya analisis sehingga data transaksi dapat dipecah dan dianalisis berdasarkan produk, pelanggan, penjual, maupun waktu.

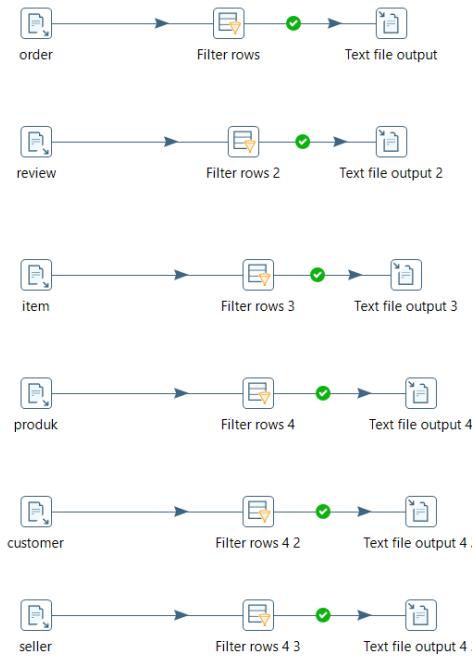


Relasi antar tabel dibangun melalui foreign key yang menghubungkan tabel fakta dengan masing-masing tabel dimensi. Atribut seperti `product_id`, `customer_id`, dan `seller_id` pada tabel fakta mengarah ke tabel dimensi terkait, sedangkan berbagai atribut tanggal seperti `purchase_date_id`, `delivered_date_id`, dan `review_date_id` merujuk ke `dim_date`. Pola hubungan ini menghasilkan struktur star schema yang efisien dan mudah digunakan untuk analisis multidimensional, karena setiap transaksi dapat ditinjau dari berbagai perspektif berdasarkan dimensi yang terhubung. Dengan demikian, rancangan ini mendukung kegiatan analisis data yang komprehensif, cepat, dan terorganisasi sesuai kebutuhan sistem pelaporan maupun business intelligence.

V. Rancangan dan Penjelasan Proyek ETL (*pipeline*)

Proyek ETL (Extract, Transform, Load) ini dirancang untuk mengolah data dari beberapa sumber operasional agar siap digunakan dalam proses analisis dan pelaporan. Pipeline ETL dibagi menjadi tiga tahap utama, yaitu Extract, Transform, dan Load. Pada bagian ini dijelaskan rancangan serta alur kerja pipeline ETL yang digunakan, diantaranya sebagai berikut:

1. Extract



Tahap Extract merupakan proses pengambilan data dari sumber data awal. Pada proyek ini, data diekstraksi dari beberapa tabel utama, yaitu order, review, item, produk, customer, dan seller. Masing-masing tabel diproses secara terpisah untuk menjaga konsistensi dan kemudahan pengolahan data.

Setiap alur ekstraksi diawali dengan pengambilan data dari sumber, kemudian dilanjutkan dengan proses penyaringan menggunakan komponen Filter Rows. Proses ini bertujuan untuk menghilangkan data yang tidak valid, seperti nilai null, data kosong, atau data yang tidak memenuhi kriteria tertentu. Setelah data difilter, hasil ekstraksi disimpan sementara dalam bentuk file teks (Text File Output). Penyimpanan ini berfungsi sebagai data antara (staging) sebelum memasuki tahap transformasi selanjutnya.

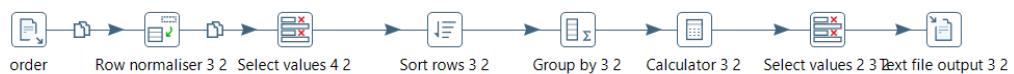
- Secara rinci, alur ekstraksi meliputi:
 - Data order diekstraksi, difilter, dan disimpan ke dalam file teks.
 - Data review diproses untuk memastikan hanya data valid yang disertakan.
 - Data item difilter untuk menghindari inkonsistensi nilai.
 - Data produk diseleksi agar hanya informasi produk yang relevan yang digunakan.
 - Data customer diproses untuk memastikan kelengkapan dan keakuratan data pelanggan.
 - Data seller diekstraksi dengan proses penyaringan serupa untuk menjaga kualitas data penjual.

Dengan penerapan tahap Extract ini, data yang masuk ke proses berikutnya telah melalui proses validasi awal sehingga lebih bersih, terstruktur, dan siap untuk ditransformasikan.

2. Transform

Tahap Transform bertujuan untuk meningkatkan kualitas dan konsistensi data sebelum dimuat ke dalam data warehouse. Proses transformasi meliputi penghapusan data duplikat serta pembersihan kolom yang memiliki nilai kosong agar tidak menimbulkan kesalahan pada proses analisis.

Selain itu, dilakukan normalisasi teks pada atribut brand dan kategori untuk menyeragamkan penulisan, serta konversi tipe data pada kolom harga dan rating agar sesuai untuk perhitungan numerik. Pada tahap ini juga dibentuk tabel dimensi, yaitu product, brand, category, marketing, dan time. Selanjutnya, tabel fakta dibangun melalui proses join antara data transaksi dan seluruh tabel dimensi, sehingga data siap digunakan untuk analisis dan pelaporan.



3. Load

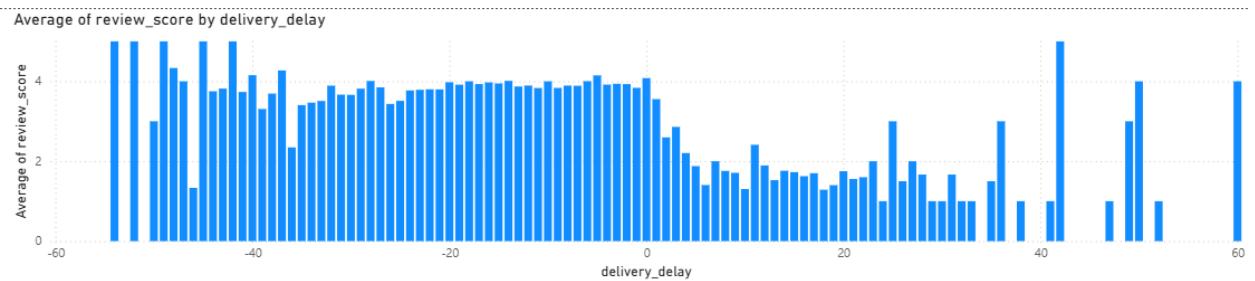
Tahap Load merupakan proses pemuatan data hasil transformasi ke dalam sistem tujuan. Pada proyek ini, hasil ETL disimpan dalam bentuk file CSV sebagai output dari Pentaho. Oleh karena itu, proses Load difokuskan pada penyiapan struktur tabel di data warehouse yang sesuai dengan skema tabel dimensi dan tabel fakta.

File CSV yang dihasilkan selanjutnya dapat dimuat ke dalam tabel yang telah disiapkan tanpa memerlukan proses pemrosesan tambahan, karena data sudah bersih dan terstruktur pada tahap sebelumnya.

VI. Hasil Visualisasi Data

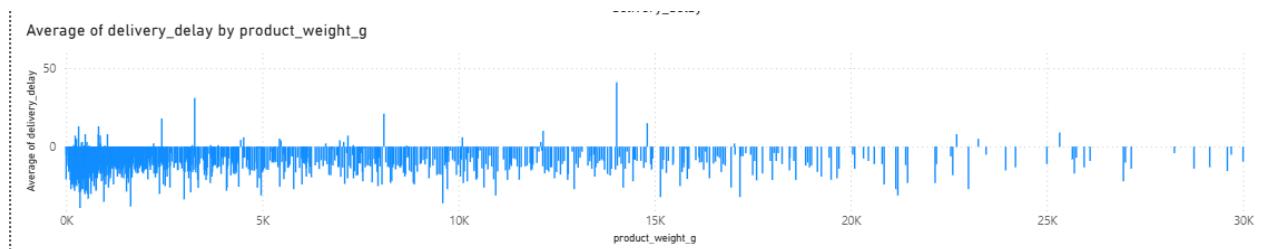
Grafik dan diagram yang ditampilkan merepresentasikan hasil pengolahan data berdasarkan tabel fakta dan tabel dimensi yang telah dibentuk. Visualisasi ini menjadi dasar dalam melakukan analisis lebih lanjut terhadap performa penjualan, karakteristik produk, serta perilaku pengguna, yang akan dijelaskan pada bagian analisis grafik selanjutnya.

1. Pengaruh keterlambatan terhadap review



Grafik pertama menunjukkan hubungan antara delivery delay dan rata-rata skor ulasan (review score). Terlihat bahwa nilai delay negatif, yang menandakan pengiriman lebih cepat dari estimasi, cenderung menghasilkan skor ulasan yang lebih tinggi. Sebaliknya, semakin besar nilai delay positif (pengiriman terlambat), rata-rata skor ulasan cenderung menurun. Hal ini menunjukkan bahwa ketepatan waktu pengiriman memiliki pengaruh signifikan terhadap tingkat kepuasan pelanggan.

2. Pengaruh Berat terhadap keterlambatan



Grafik kedua menggambarkan hubungan antara berat produk dan rata-rata keterlambatan pengiriman. Secara umum, produk dengan berat yang lebih besar cenderung memiliki nilai delay yang lebih tinggi, meskipun terdapat fluktuasi pada beberapa titik. Pola ini mengindikasikan bahwa berat produk dapat memengaruhi proses distribusi, terutama dalam hal waktu pengiriman.

3. Pendapat dari tahun ke tahun



Grafik ketiga memperlihatkan rata-rata nilai pendapatan (total order value) per tahun. Terlihat adanya peningkatan pendapatan dari tahun 2017 ke 2018. Tren ini menunjukkan pertumbuhan kinerja penjualan yang positif dari waktu ke waktu, yang dapat dikaitkan dengan meningkatnya jumlah transaksi atau nilai pembelian pelanggan.