

**PERANCANGAN DASHBOARD BUSINESS
INTELLIGENCE BERBASIS DATA WAREHOUSE UNTUK
ANALISIS KINERJA FINANCIAL PADA OLIST
MARKETPLACE**



**Disusun untuk memenuhi Tugas Besar
Mata Kuliah Data Warehouse Business Intelligence**

Oleh:

Kelompok 10

Rheina Fairuz

102022300207

**PROGRAM STUDI STRATA 1 SISTEM INFORMASI
FAKULTAS REKAYASA INDUSTRI
UNIVERSITAS TELKOM
2025**

DAFTAR ISI

Contents

PERANCANGAN DATA WAREHOUSE DAN BUSINESS INTELLIGENCE BERBASIS PERSPEKTIF FINANCIAL UNTUK ANALISIS PROFITABILITAS PADA CHINOOK MUSIC STORE	Error! Bookmark not defined.
DAFTAR ISI.....	2
BAB I PENDAHULUAN.....	6
I.1 Latar Belakang	6
I.2 Identifikasi Masalah	6
I.3 Tujuan Penelitian.....	7
BAB II	8
II.1 Tujuan Utama (Objective).....	8
II.2 Perspektif Balanced Scorecard	8
II.3 Key Performance Indicators (KPI).....	10
BAB III	11
III.1 Deskripsi Sumber Data.....	11
III.1.1 Sumber Data Utama (Primary Data).....	11
III.2 Exploratory Data Analysis (EDA)	13
III.2.1 Pengecekan Kualitas Data (Data Quality Check)	13
Berdasarkan hasil pengecekan awal terhadap dataset Olist, diperoleh beberapa temuan sebagai berikut:	13
1. Format tanggal Kolom order_purchase_timestamp pada tabel orders memiliki format timestamp yang perlu dilakukan transformasi untuk memisahkan atribut tanggal, bulan, dan tahun pada dimensi waktu.....	13
2. Missing values Beberapa kolom pada tabel products memiliki nilai kosong, terutama pada atribut deskripsi produk dan dimensi fisik. Nilai kosong tersebut perlu ditangani pada proses ETL agar tidak mengganggu analisis.....	13
3. Konsistensi data numerik Nilai price dan freight_value pada tabel order_items telah diverifikasi dan dapat digunakan untuk perhitungan revenue. Perhitungan pendapatan dilakukan berdasarkan agregasi nilai price dikalikan dengan jumlah item.	13
4. Duplikasi data Tidak ditemukan duplikasi data pada primary key masing-masing tabel, sehingga data dianggap konsisten untuk proses integrasi lebih lanjut.....	13
III.2.2 Profiling Data dan Distribusi (Data Profiling).....	13
Profiling data dilakukan untuk memahami distribusi data transaksi dan pola awal yang muncul, antara lain:.....	13
1. Distribusi transaksi Data transaksi tersebar dalam rentang waktu beberapa tahun, sehingga memungkinkan dilakukan analisis tren pendapatan dan pertumbuhan penjualan dari waktu ke waktu.	13

2.	Distribusi kategori produk Produk yang dijual terdiri dari berbagai kategori dengan kontribusi pendapatan yang berbeda-beda. Hal ini memungkinkan analisis kontribusi revenue berdasarkan kategori produk	13
3.	Distribusi wilayah	14
	Data pelanggan dan penjual tersebar di berbagai kota dan negara bagian di Brasil, sehingga mendukung analisis distribusi pendapatan berdasarkan wilayah.....	14
BAB IV	PERANCANGAN STAR SCHEMA	15
IV.1	Definisi Star Schema.....	15
IV.2	Desain Star Schema.....	15
IV.2.1	Tabel Fakta.....	15
IV.2.2	Tabel Dimensi	16
IV.2.3	Diagram Star Schema	17
BAB V	IMPLEMENTASI STAR SCHEMA	18
V.1	Data Definition Language (DDL)	18
BAB VI	IMPLEMENTASI PROSES ETL	22
VI.1	Alur Proses ETL.....	22
A.	Tools yang digunakan	22
VI.2	Dokumentasi ETL	23
A.	Transformasi Data.....	24
BAB VII	IMPLEMENTASI DATA MINING	29
VII.1	Data Preparation	30
VII.2	Clustering: K-Means untuk Segmentasi Pelanggan	30
	Clustering digunakan untuk mengelompokkan pelanggan berdasarkan perilaku transaksi guna mendukung analisis profitabilitas pelanggan. Metode clustering yang digunakan adalah K-Means, karena sederhana, efisien, dan umum digunakan dalam segmentasi pelanggan.....	30
	Pada penelitian ini, digunakan pendekatan RFM (Recency, Frequency, Monetary) dengan atribut sebagai berikut:.....	30
•	Recency (R) Selisih hari antara transaksi terakhir pelanggan dengan tanggal terakhir pada dataset.....	30
•	Frequency (F) Jumlah transaksi yang dilakukan oleh pelanggan	31
•	Monetary (M) Total nilai revenue yang dihasilkan oleh pelanggan.	31
	Nilai RFM dihitung berdasarkan data pada tabel fact_sales yang telah digabungkan dengan dimensi waktu dan pelanggan.....	31
VII.2.1	Implementasi K-Means Clustering	31
	Setelah nilai RFM diperoleh, dilakukan proses normalisasi data menggunakan StandardScaler untuk menghindari dominasi satu variabel tertentu. Selanjutnya, algoritma K-Means diterapkan dengan jumlah cluster sebanyak 3, yang dipilih untuk	

merepresentasikan segmentasi pelanggan utama.....	31
VII.2.2 Hasil dan Analisis	32
VII.3 Linear Regression: Analisis Tren Penjualan	33
VII.3.1 Implementasi Linear Regression.....	33
VII.3.2 Hasil dan Analisis	34
BAB VIII PERANCANGAN DASHBOARD KPI.....	35
VIII.1 Tampilan Dashboard.....	35
VIII.2 Analisis Ketercapaian KPI.....	36
Chart / KPI	36
Tercapai	36
TidakTercapai.....	36
Insight dari Dashboard	36
Total Revenue	36
✓ 36	
Total revenue yang ditampilkan sebesar 13,59 juta, menunjukkan bahwa sistem data warehouse berhasil mengintegrasikan data penjualan dan menghasilkan nilai pendapatan secara agregat dengan baik. KPI revenue utama telah tercapai.....	36
Average Monetary per Customer (ARPC)	36
✓ 36	
Nilai ARPC sebesar 13,59 juta menunjukkan rata-rata.....	36
kontribusi pendapatan per pelanggan. Metrik ini berhasil dihitung dari data fact sales dan dimensi pelanggan, sehingga KPI nilai pelanggan tercapai.....	36
Top Track / Customer Revenue	36
✓ 36	
Dashboard berhasil menampilkan pelanggan dengan revenue tertinggi. Namun, data masih didominasi oleh satu CustomerID, yang mengindikasikan distribusi transaksi belum merata.....	36
Trend Penjualan.....	36
Tidak36	
✓ 36	
Visualisasi tren penjualan belum optimal karena keterbatasan dimensi waktu. Grafik hanya menampilkan satu titik agregat sehingga analisis pola penjualan dari waktu ke waktu belum dapat dilakukan secara mendalam.	36
Cluster Customer.....	36
Tidak36	
✓ 36	

Segmentasi pelanggan hanya menghasilkan satu cluster (100%), sehingga tujuan pengelompokan pelanggan belum tercapai secara maksimal. Hal ini disebabkan oleh karakteristik data yang homogen atau keterbatasan variabel clustering.....	36
Distribusi Revenue (Genre / Produk)	37
✓ 37	
Dashboard berhasil menampilkan distribusi revenue dalam bentuk diagram lingkaran. Meskipun seluruh revenue masih terkonsentrasi pada satu kategori (100%), visualisasi kontribusi pendapatan tetap dapat ditampilkan dengan baik.	37
LAMPIRAN	39
Link Akses Dashboard:.....	39
Link Dataset Kaggle:	39
https://www.kaggle.com/datasets/olistbr/brazilian-e-commerce	39
Link Github:	39

BAB I

PENDAHULUAN

I.1 Latar Belakang

Perkembangan industri e-commerce mengalami pertumbuhan yang sangat pesat seiring dengan meningkatnya penggunaan platform digital dalam aktivitas jual beli. Persaingan antar perusahaan e-commerce semakin ketat, sehingga kemampuan perusahaan dalam mengelola dan menganalisis data transaksi menjadi faktor penting dalam menjaga daya saing dan meningkatkan profitabilitas bisnis.

Olist merupakan salah satu perusahaan e-commerce di Brasil yang menyediakan platform integrasi antara penjual (seller) dan berbagai marketplace besar. Dalam operasionalnya, Olist menghasilkan data transaksi dalam jumlah besar setiap hari yang mencakup informasi pesanan, pelanggan, produk, penjual, pembayaran, serta ulasan pelanggan. Data transaksi tersebut memiliki potensi besar untuk dimanfaatkan dalam mendukung pengambilan keputusan strategis, khususnya dalam evaluasi kinerja finansial perusahaan.

Namun demikian, data transaksi Olist masih tersimpan dalam sistem operasional atau Online Transaction Processing (OLTP) yang dirancang untuk mendukung proses transaksi harian. Struktur data operasional tersebut kurang optimal untuk kebutuhan analisis strategis seperti menganalisis pertumbuhan pendapatan dari waktu ke waktu, mengevaluasi kontribusi pendapatan berdasarkan produk atau penjual, serta memahami pola perilaku pelanggan secara menyeluruh.

Untuk menjawab tantangan tersebut, diperlukan penerapan Data Warehouse dan Business Intelligence (BI) yang mampu mengintegrasikan dan mengolah data transaksi menjadi informasi yang bervilai. Dengan membangun data warehouse berbasis Star Schema, data dapat dianalisis secara multidimensi berdasarkan waktu, pelanggan, produk, dan penjual. Selanjutnya, penerapan Business Intelligence memungkinkan perusahaan memantau Key Performance Indicators (KPI) dari perspektif finansial secara akurat. Oleh karena itu, penelitian ini bertujuan untuk membangun solusi Data Warehouse dan Business Intelligence pada dataset Olist guna membantu manajemen dalam mengevaluasi kinerja pendapatan serta merumuskan strategi bisnis berbasis data.

I.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan yang akan dibahas dalam laporan ini adalah sebagai berikut:

1. Bagaimana merancang arsitektur Data Warehouse menggunakan Star Schema yang sesuai untuk kebutuhan analisis bisnis pada perusahaan e-commerce Olist?
2. Bagaimana melakukan proses integrasi data melalui Extract, Transform, Load (ETL) dari sumber data utama dan data pendukung ke dalam data warehouse secara efektif?
3. Bagaimana menerapkan metode Data Mining seperti klasifikasi, regresi, atau klastering untuk menggali pola dan informasi tersembunyi dari data transaksi Olist guna menjawab Key Performance Indicators (KPI) yang telah ditentukan?
4. Bagaimana menyajikan hasil analisis data dalam bentuk dashboard interaktif yang informatif untuk mendukung pengambilan keputusan manajerial?

I.3 Tujuan Penelitian

Tujuan dari pengerjaan tugas besar ini adalah sebagai berikut:

1. Membangun Data Warehouse berbasis Star Schema dan mengimplementasikannya ke dalam sistem manajemen basis data relasional (RDBMS) untuk mendukung analisis kinerja perusahaan.
2. Melakukan proses Extract, Transform, Load (ETL) dari sumber data utama Olist dan data pendukung sehingga data siap digunakan untuk analisis.
3. Mengimplementasikan metode Data Mining untuk menghasilkan analisis deskriptif maupun prediktif terkait kinerja pendapatan perusahaan.
4. Menyajikan hasil pengolahan data dan visualisasi Key Performance Indicators (KPI) dalam bentuk dashboard serta memberikan rekomendasi bisnis kepada pihak manajemen.

BAB II

PERUMUSAN OBJECTIVE, BALANCED SCORECARD DAN KPI

II.1 Tujuan Utama (Objective)

Tujuan utama dari penelitian ini adalah meningkatkan kinerja finansial perusahaan e-commerce Olist melalui pemanfaatan data warehouse dan business intelligence sebagai alat pendukung pengambilan keputusan strategis. Seiring dengan pesatnya perkembangan industri e-commerce, Olist menghasilkan data transaksi dalam jumlah besar yang mencerminkan aktivitas penjualan, perilaku pelanggan, performa produk, serta kontribusi penjual (seller). Namun, data tersebut masih bersifat operasional sehingga belum dapat dimanfaatkan secara optimal untuk kebutuhan analisis strategis.

Melalui penerapan data warehouse, data transaksi Olist akan diintegrasikan dan diorganisasikan dalam struktur yang lebih terstandarisasi sehingga mendukung analisis historis dan analisis tren. Selanjutnya, business intelligence digunakan untuk mengolah data tersebut menjadi informasi yang relevan, khususnya terkait pendapatan, pertumbuhan penjualan, nilai transaksi rata-rata, serta kontribusi produk dan penjual terhadap kinerja keuangan perusahaan. Dengan demikian, objective penelitian ini tidak hanya berfokus pada penyajian data, tetapi juga pada penyediaan insight yang dapat dimanfaatkan oleh manajemen untuk mengevaluasi performa bisnis dan merumuskan strategi peningkatan pendapatan secara berbasis data.

II.2 Perspektif Balanced Scorecard

Balanced Scorecard merupakan kerangka kerja manajemen kinerja yang digunakan untuk menerjemahkan visi dan strategi organisasi ke dalam seperangkat indikator kinerja yang terukur. Pendekatan ini menilai kinerja organisasi dari beberapa perspektif utama, yaitu finansial, pelanggan, proses bisnis internal, serta pembelajaran dan pertumbuhan. Dalam konteks penelitian ini, perspektif Balanced Scorecard yang digunakan adalah perspektif finansial.

Perspektif finansial dipilih karena berfokus pada pencapaian kinerja keuangan sebagai indikator utama keberhasilan perusahaan. Sebagai perusahaan e-commerce, Olist sangat bergantung pada pendapatan yang dihasilkan dari transaksi pelanggan melalui platform marketplace yang dikelolanya. Oleh karena itu, pengukuran terhadap aspek finansial seperti total pendapatan, pertumbuhan

penjualan, nilai transaksi rata-rata, serta kontribusi produk dan penjual menjadi sangat relevan untuk menggambarkan kondisi bisnis perusahaan secara menyeluruh.

Selain itu, perspektif finansial memiliki keterkaitan yang kuat dengan penerapan data warehouse dan business intelligence. Data transaksi penjualan yang tersedia pada sistem operasional dapat diolah menjadi informasi finansial yang bernilai melalui proses ETL dan analisis multidimensi. Dengan menggunakan perspektif finansial, hasil analisis yang diperoleh diharapkan mampu memberikan gambaran yang jelas mengenai performa keuangan Olist serta menjadi dasar dalam pengambilan keputusan strategis berbasis data.

II.3 Key Performance Indicators (KPI)

Goals	KPI	Data	Perhitungan	Chart
Meningkatkan pendapatan perusahaan	Total Revenue	Data transaksi penjualan (orders, order_items, order_payments)	Target: Total revenue \geq target yang ditetapkan perusahaan	Scorecard
Meningkatkan pertumbuhan penjualan dari waktu ke waktu	Monthly Revenue Growth	Data penjualan bulanan berdasarkan order_purchase_timestamp	Target: Pertumbuhan pendapatan bulanan \geq 10%	Time Series Chart
Meningkatkan nilai transaksi rata-rata	Average Order Value (AOV)	Data transaksi dan pembayaran (orders, order_payments)	Total revenue / jumlah order	Bar Chart
Mengoptimalkan kontribusi produk dengan performa tertinggi	Top Product Revenue Contribution	Data produk dan transaksi (products, order_items)	Target: Top 10 produk menyumbang \geq 30% dari total revenue	Table Chart
Mengetahui kontribusi pendapatan berdasarkan penjual	Revenue Distribution by Seller	Data penjual dan transaksi (sellers, order_items)	Persentase kontribusi revenue per seller	Pie Chart

Tabel 2.3 1 Analisis KPI

BAB III

ANALISIS SUMBER DATA

III.1 Deskripsi Sumber Data

Dalam pembangunan Data Warehouse dan Business Intelligence pada penelitian ini, digunakan dua jenis sumber data, yaitu sumber data utama dan sumber data pendukung. Penggunaan lebih dari satu sumber data bertujuan untuk memenuhi kebutuhan analisis kinerja finansial perusahaan serta mendukung perhitungan Key Performance Indicators (KPI) yang telah ditetapkan pada bab sebelumnya.

Sumber data utama berasal dari dataset Olist yang bersifat open access dan tersedia secara publik melalui platform Kaggle. Dataset ini merepresentasikan data transaksi e-commerce dari perusahaan Olist di Brasil yang telah dianonimkan. Selain itu, digunakan pula sumber data pendukung berupa data tambahan yang dihasilkan secara manual untuk melengkapi kebutuhan analisis, seperti target KPI dan referensi pendukung lainnya.

III.1.1 Sumber Data Utama (Primary Data)

Sumber data utama yang digunakan dalam penelitian ini adalah Brazilian E-Commerce Public Dataset by Olist. Dataset ini berisi data transaksi e-commerce yang mencakup aktivitas pemesanan, pembayaran, pengiriman, produk, pelanggan, penjual, serta ulasan pelanggan. Data disajikan dalam format Comma Separated Values (CSV) dan terdiri dari beberapa tabel yang saling berelasi.

Tabel-tabel utama yang digunakan dalam penelitian ini meliputi:

1. *orders*

Menyimpan informasi utama pesanan, seperti *order_id*, *customer_id*, *order_status*, dan *order_purchase_timestamp*. Tabel ini menjadi acuan utama dalam analisis waktu transaksi.

2. *order_items*

Menyimpan detail item dalam setiap pesanan, seperti *product_id*, *seller_id*, *quantity*, *price*, dan *freight_value*. Tabel ini menjadi sumber utama perhitungan revenue.

3. *order_payments*

Menyimpan informasi pembayaran, seperti *payment_type* dan *payment_value*, yang digunakan untuk analisis nilai transaksi.

4. *customers*

Berisi data pelanggan yang mencakup lokasi pelanggan berdasarkan kota dan negara bagian (state).

5. *products*

Menyimpan informasi produk, termasuk kategori produk dan atribut fisik produk.

6. sellers

Berisi data penjual (seller) beserta lokasi geografisnya.

Tabel-tabel tersebut saling terhubung melalui key tertentu, sehingga mendukung proses integrasi data ke dalam data warehouse berbasis star schema.

III.2 Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik data, mengidentifikasi pola awal, serta mendeteksi potensi permasalahan pada data sebelum dilakukan proses ETL. EDA dilakukan terhadap tabel-tabel utama yang digunakan dalam analisis.

III.2.1 Pengecekan Kualitas Data (Data Quality Check)

Berdasarkan hasil pengecekan awal terhadap dataset Olist, diperoleh beberapa temuan sebagai berikut:

1. Format tanggal

Kolom order_purchase_timestamp pada tabel orders memiliki format timestamp yang perlu dilakukan transformasi untuk memisahkan atribut tanggal, bulan, dan tahun pada dimensi waktu.

2. Missing values

Beberapa kolom pada tabel products memiliki nilai kosong, terutama pada atribut deskripsi produk dan dimensi fisik. Nilai kosong tersebut perlu ditangani pada proses ETL agar tidak mengganggu analisis.

3. Konsistensi data numerik

Nilai price dan freight_value pada tabel order_items telah diverifikasi dan dapat digunakan untuk perhitungan revenue. Perhitungan pendapatan dilakukan berdasarkan agregasi nilai price dikalikan dengan jumlah item.

4. Duplikasi data

Tidak ditemukan duplikasi data pada primary key masing-masing tabel, sehingga data dianggap konsisten untuk proses integrasi lebih lanjut.

III.2.2 Profiling Data dan Distribusi (Data Profiling)

Profiling data dilakukan untuk memahami distribusi data transaksi dan pola awal yang muncul, antara lain:

1. Distribusi transaksi

Data transaksi tersebar dalam rentang waktu beberapa tahun, sehingga memungkinkan dilakukan analisis tren pendapatan dan pertumbuhan penjualan dari waktu ke waktu.

2. Distribusi kategori produk

Produk yang dijual terdiri dari berbagai kategori dengan kontribusi pendapatan yang berbeda-beda. Hal ini memungkinkan analisis kontribusi revenue berdasarkan kategori produk.

3. Distribusi wilayah

Data pelanggan dan penjual tersebar di berbagai kota dan negara bagian di Brasil, sehingga mendukung analisis distribusi pendapatan berdasarkan wilayah.

BAB IV PERANCANGAN STAR SCHEMA

IV.1 Definisi Star Schema

Star schema merupakan salah satu metode perancangan skema data warehouse yang terdiri dari satu tabel fakta (fact table) sebagai pusat data dan beberapa tabel dimensi (dimension table) yang terhubung langsung ke tabel fakta. Struktur ini menyerupai bentuk bintang, di mana tabel fakta berada di tengah dan tabel dimensi mengelilinginya.

Penggunaan star schema bertujuan untuk menyederhanakan struktur data sehingga proses analisis menjadi lebih efisien. Dengan struktur yang terdenormalisasi, star schema mampu mempercepat proses query analitik, khususnya untuk kebutuhan Business Intelligence. Selain itu, model ini memudahkan pengguna dalam memahami hubungan antar data serta melakukan analisis multidimensi berdasarkan berbagai sudut pandang seperti waktu, pelanggan, produk, dan penjual.

Dalam konteks analisis kinerja finansial perusahaan e-commerce Olist, star schema digunakan untuk mendukung analisis pendapatan, tren penjualan, nilai transaksi, serta kontribusi produk dan penjual terhadap total revenue perusahaan.

IV.2 Desain Star Schema

Desain star schema pada data warehouse ini disusun berdasarkan kebutuhan analisis finansial yang telah dirumuskan pada bab sebelumnya. Skema ini terdiri dari satu tabel fakta utama yang menyimpan data transaksi penjualan serta beberapa tabel dimensi yang memberikan konteks terhadap data tersebut.

IV.2.1 Tabel Fakta

Tabel fakta yang digunakan dalam perancangan data warehouse ini adalah fact_sales. Tabel ini merepresentasikan transaksi penjualan pada platform e-commerce Olist dan menyimpan data numerik yang menjadi dasar perhitungan Key Performance Indicators (KPI) finansial.

Atribut utama dalam tabel fact_sales meliputi:

- FactSalesID sebagai primary key (surrogate key)
- TimeID sebagai foreign key yang mengacu pada dimensi waktu

- CustomerID sebagai foreign key yang mengacu pada dimensi pelanggan
- ProductID sebagai foreign key yang mengacu pada dimensi produk
- SellerID sebagai foreign key yang mengacu pada dimensi penjual
- Quantity sebagai jumlah produk yang terjual
- Revenue sebagai total pendapatan hasil penjumlahan nilai harga produk

Granularitas tabel fakta berada pada level transaksi produk per order, sehingga memungkinkan analisis pendapatan baik secara detail maupun agregat berdasarkan berbagai dimensi.

IV.2.2 Tabel Dimensi

Untuk mendukung analisis multidimensi, digunakan beberapa tabel dimensi sebagai berikut:

- 1. Dimensi Waktu (dim_time)**

Dimensi waktu digunakan untuk menganalisis data penjualan berdasarkan periode tertentu seperti tahun, kuartal, bulan, dan tanggal. Dimensi ini berperan penting dalam analisis tren penjualan dan pertumbuhan pendapatan.

Kolom pada dim_time meliputi: TimeID, FullDate, Year, Quarter, Month, MonthName, Day.

- 2. Dimensi Pelanggan (dim_customer)**

Dimensi pelanggan menyimpan informasi terkait pelanggan yang melakukan transaksi pada platform Olist. Dimensi ini digunakan untuk menganalisis kontribusi pelanggan terhadap pendapatan serta menghitung nilai transaksi rata-rata.

Kolom pada dim_customer meliputi: CustomerID, CustomerUniqueID, City, State.

- 3. Dimensi Produk (dim_product)**

Dimensi produk digunakan untuk menyimpan informasi terkait produk yang dijual, termasuk kategori produk dan atribut pendukung lainnya. Dimensi ini memungkinkan analisis kontribusi pendapatan berdasarkan jenis dan kategori produk.

Kolom pada dim_product meliputi: ProductID, ProductCategory, ProductWeight, ProductLength, ProductHeight, ProductWidth.

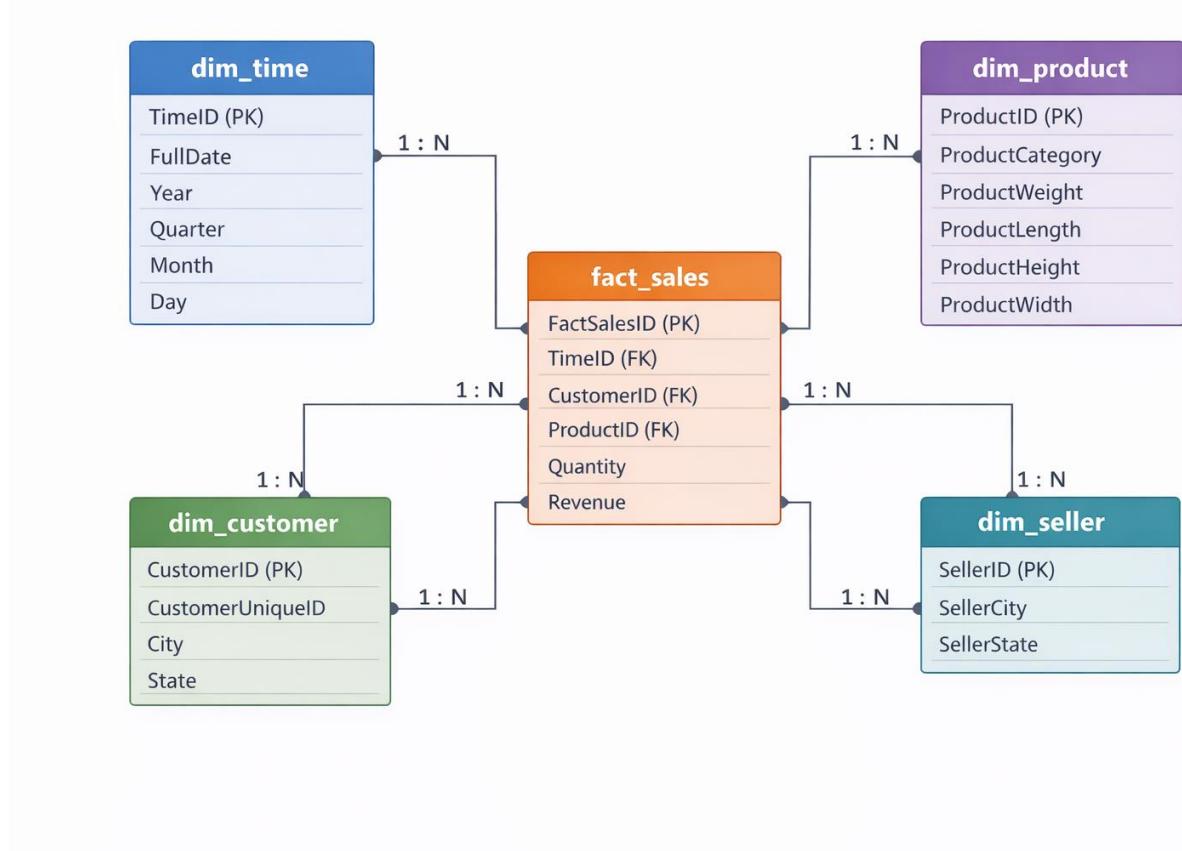
- 4. Dimensi Penjual (dim_seller)**

Dimensi penjual menyimpan informasi terkait seller yang menjual produk melalui platform Olist. Dimensi ini digunakan untuk menganalisis performa dan kontribusi pendapatan masing-masing penjual.

Kolom pada dim_seller meliputi: SellerID, SellerCity, SellerState.

Relasi antara tabel fakta dan tabel dimensi bersifat one-to-many, di mana satu data pada tabel dimensi dapat direferensikan oleh banyak data pada tabel fakta. Struktur ini mendukung proses analisis multidimensi yang efisien dan fleksibel sesuai kebutuhan Business Intelligence.

IV.2.3 Diagram Star Schema



Gambar 4.2.3 1 Diagram Star Schema

BAB V

IMPLEMENTASI STAR SCHEMA

V.1 Data Definition Language (DDL)

Implementasi star schema dilakukan pada sistem manajemen basis data relasional (RDBMS) menggunakan MySQL. Pada tahap ini, rancangan konseptual star schema yang telah dibuat pada Bab IV diterjemahkan ke dalam bentuk fisik melalui perintah Data Definition Language (DDL).

Proses implementasi diawali dengan pembuatan database data warehouse Olist, kemudian dilanjutkan dengan pembuatan tabel dimensi dan tabel fakta. Setiap tabel dimensi dibuat menggunakan primary key berupa surrogate key yang bersifat auto-increment untuk menjaga konsistensi data serta memudahkan proses integrasi data dari sistem sumber.

Setelah seluruh tabel dimensi berhasil dibuat, tabel fakta `fact_sales` dibangun dengan menambahkan foreign key yang mengacu pada primary key masing-masing tabel dimensi. Penerapan foreign key constraint bertujuan untuk menjaga integritas referensial antar tabel serta memastikan bahwa setiap data transaksi memiliki konteks dimensi yang valid.

Dengan implementasi DDL ini, struktur star schema data warehouse Olist telah siap digunakan sebagai fondasi untuk proses ETL dan analisis data lebih lanjut.

1. dim_time

Run SQL query/queries on database dw_olist: 

```
1 CREATE TABLE dim_time (
2     TimeID INT AUTO_INCREMENT PRIMARY KEY,
3     FullDate DATE,
4     Year INT,
5     Quarter INT,
6     Month INT,
7     MonthName VARCHAR(20),
8     Day INT
9 );
10
```

1.1 1 DDL dim_time

2. dim_customer

```
1 CREATE TABLE dim_customer (
2     CustomerID INT AUTO_INCREMENT PRIMARY KEY,
3     CustomerUniqueID VARCHAR(50),
4     City VARCHAR(100),
5     State VARCHAR(50)
6 );
7
```

5.1 2 DDL dim_customer

3. dim_product

```
1 CREATE TABLE dim_product (
2     ProductID INT AUTO_INCREMENT PRIMARY KEY,
3     ProductCategory VARCHAR(100),
4     ProductWeight FLOAT,
5     ProductLength FLOAT,
6     ProductHeight FLOAT,
7     ProductWidth FLOAT
8 );
9
```

5.1 3 DDL dim_product

4. dim_seller

```
1 CREATE TABLE dim_seller (
2     SellerID INT AUTO_INCREMENT PRIMARY KEY,
3     SellerCity VARCHAR(100),
4     SellerState VARCHAR(50)
5 );
6
```

5.1 4 DDL dim_seller

5. fact_sales

```
1 CREATE TABLE fact_sales (
2     FactSalesID INT AUTO_INCREMENT PRIMARY KEY,
3     TimeID INT,
4     CustomerID INT,
5     ProductID INT,
6     SellerID INT,
7     Quantity INT,
8     Revenue DECIMAL(12,2),
9
10    CONSTRAINT fk_time
11        FOREIGN KEY (TimeID)
12        REFERENCES dim_time(TimeID),
13
14    CONSTRAINT fk_customer
15        FOREIGN KEY (CustomerID)
16        REFERENCES dim_customer(CustomerID),
17
18    CONSTRAINT fk_product
19        FOREIGN KEY (ProductID)
20        REFERENCES dim_product(ProductID),
21
22    CONSTRAINT fk_seller
23        FOREIGN KEY (SellerID)
24        REFERENCES dim_seller(SellerID)
25 );
```

5.1 5 DDL fact_sales

BAB VI

IMPLEMENTASI PROSES ETL

VI.1 Alur Proses ETL

A. Tools yang digunakan

Proses ETL (Extract, Transform, Load) merupakan tahapan penting dalam pembangunan data warehouse karena berfungsi untuk memindahkan dan mengolah data dari sistem sumber ke dalam struktur star schema yang telah dirancang. Pada penggeraan tugas ini, proses ETL dilakukan menggunakan Pentaho Data Integration (PDI) sebagai tools utama untuk merancang dan menjalankan alur integrasi data.

Tahapan ETL yang dilakukan pada penelitian ini meliputi extract, transform, dan load, dengan sumber data berasal dari dataset Olist Brazilian E-Commerce Public Dataset yang tersedia dalam format CSV.

1. Extract

Tahap extract dilakukan dengan mengambil data dari sumber data Olist yang tersedia dalam bentuk file CSV. Data yang diekstraksi mencakup beberapa tabel utama yang berkaitan dengan transaksi e-commerce, yaitu orders, order_items, order_payments, customers, products, dan sellers.

Pada tahap ini dilakukan penyesuaian format file seperti delimiter dan tipe data agar setiap file CSV dapat dibaca dengan benar oleh Mysql. Seluruh data hasil extract kemudian ditempatkan pada staging area sebagai data mentah sebelum dilakukan proses transformasi.

2. Transform

Tahap transform bertujuan untuk membersihkan, mengolah, dan menyesuaikan data agar sesuai dengan struktur data warehouse berbasis star schema. Proses transformasi yang dilakukan antara lain:

- Pemisahan atribut waktu dari kolom order_purchase_timestamp pada tabel orders menjadi atribut tahun, kuartal, bulan, dan tanggal untuk membentuk dimensi waktu (dim_time).
- Perhitungan nilai pendapatan (revenue) berdasarkan harga produk pada tabel order_items, yang kemudian digunakan sebagai measure utama pada tabel fakta fact_sales.
- Penanganan nilai kosong (null) pada beberapa atribut produk dan pelanggan agar tidak mengganggu proses analisis.

- Integrasi data dari tabel orders, order_items, customers, products, dan sellers untuk membentuk struktur tabel fakta dan tabel dimensi.
- Pemetaan surrogate key dari tabel dimensi ke tabel fakta menggunakan mekanisme database lookup pada Pentaho Data Integration.

Transformasi ini dilakukan untuk memastikan bahwa data yang dimuat ke dalam data warehouse bersifat bersih, konsisten, dan sesuai dengan kebutuhan analisis finansial yang telah ditentukan.

3. Load

Tahap load merupakan proses memasukkan data hasil transformasi ke dalam data warehouse Olist. Proses loading dilakukan secara bertahap dengan urutan sebagai berikut:

1. Memuat data ke tabel dimensi, yaitu dim_time, dim_customer, dim_product, dan dim_seller.
2. Memuat data ke tabel fakta fact_sales dengan mereferensikan foreign key dari masing-masing tabel dimensi.

Urutan loading ini bertujuan untuk menjaga integritas referensial antar tabel dan memastikan bahwa setiap data transaksi pada tabel fakta memiliki konteks dimensi yang valid.

B. Tools yang Digunakan

Tools yang digunakan dalam penggeraan tugas ini adalah sebagai berikut:

1. Database

Sistem manajemen basis data yang digunakan adalah MySQL. Database ini digunakan untuk menampung data warehouse Olist yang telah dirancang menggunakan model star schema.

2. Tools ETL

Myphpadmin manual digunakan sebagai tools ETL untuk merancang, menjalankan, dan mengelola alur proses extract, transform, dan load secara visual. PDI memudahkan proses integrasi data dari file CSV ke dalam database data warehouse.

VI.2 Dokumentasi ETL

Implementasi proses ETL didokumentasikan dalam bentuk alur kerja pada Pentaho Data Integration yang menggambarkan proses extract, transform, dan load secara terstruktur. Berdasarkan hasil implementasi, seluruh tabel dimensi dan tabel fakta berhasil

terisi dengan data yang sesuai dengan rancangan star schema.

A. Transformasi Data

1. Sebelum transformasi

A. dim_time

The screenshot shows a MySQL Workbench interface. At the top, a green status bar indicates: "MySQL returned an empty result set (i.e. zero rows). (Query took 0.0031 seconds.)". Below it is a toolbar with options like Profiling, Edit inline, Explain SQL, Create PHP code, and Refresh. The main area contains a SQL editor with the query: "SELECT * FROM `dim_time`". Underneath the editor is a table header row with columns: TimeID, FullDate, Year, Quarter, Month, MonthName, and Day. A "Query results operations" button is visible, along with a "Create view" icon.

Gambar 6.2.1 1 Gambar dim_time sebelum transformasi

B. dim_customer

The screenshot shows a MySQL Workbench interface. At the top, a green status bar indicates: "MySQL returned an empty result set (i.e. zero rows). (Query took 0.0031 seconds.)". Below it is a toolbar with options like Profiling, Edit inline, Explain SQL, Create PHP code, and Refresh. The main area contains a SQL editor with the query: "SELECT * FROM `dim_customer`". Underneath the editor is a table header row with columns: CustomerID, CustomerUniqueId, City, and State. A "Query results operations" button is visible, along with a "Create view" icon.

Gambar 6.2.1 2 Gambar dim_customer sebelum transformasi

C. dim_seller

The screenshot shows a MySQL Workbench interface. At the top, a green status bar indicates: "MySQL returned an empty result set (i.e. zero rows). (Query took 0.0034 seconds.)". Below it is a toolbar with options like Profiling, Edit inline, Explain SQL, Create PHP code, and Refresh. The main area contains a SQL editor with the query: "SELECT * FROM `dim_seller`". Underneath the editor is a table header row with columns: SellerID, SellerCity, and SellerState. A "Query results operations" button is visible, along with a "Create view" icon.

Gambar 6.2.1 3 Gambar dim_seller sebelum transformasi

D. dim_product

MySQL returned an empty result set (i.e. zero rows). (Query took 0.0005 seconds.)

```
SELECT * FROM `dim_product`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

ProductID	ProductCategory	ProductWeight	ProductLength	ProductHeight	ProductWidth
-----------	-----------------	---------------	---------------	---------------	--------------

Query results operations

Create view

Gambar 6.2.1 4 Gambar dim_product sebelum transformasi

E. Fact_Sales

MySQL returned an empty result set (i.e. zero rows). (Query took 0.0024 seconds.)

```
SELECT * FROM `fact_sales`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

FactSalesID	TimeID	CustomerID	ProductID	SellerID	Quantity	Revenue
-------------	--------	------------	-----------	----------	----------	---------

Query results operations

Create view

Gambar 6.2.1 5 Gambar fact_sales sebelum transformasi

2. Proses Transformasi / ETL

A. dim_time

```

1 INSERT INTO dim_time (
2   FullDate,
3   Year,
4   Quarter,
5   Month,
6   MonthName,
7   Day
8 )
9 SELECT DISTINCT
10   DATE(order_purchase_timestamp),
11   YEAR(order_purchase_timestamp),
12   QUARTER(order_purchase_timestamp),
13   MONTH(order_purchase_timestamp),
14   MONTHNAME(order_purchase_timestamp),
15   DAY(order_purchase_timestamp)
16 FROM olist_raw.orders_raw

```

Gambar 6.2.2 1 Gambar ETL dim_time

B. dim_customer

```

1 INSERT INTO dim_customer (
2   CustomerUniqueID,
3   City,
4   State
5 )
6 SELECT DISTINCT
7   `COL 2`,
8   `COL 4`,
9   `COL 5`
10 FROM olist_raw.customers_raw
11 WHERE `COL 2` IS NOT NULL;

```

Gambar 6.2.2 2 Gambar ETL dim_customer

C. dim_seller

```

1 INSERT INTO dim_seller (
2   SellerCity,
3   SellerState
4 )
5 SELECT DISTINCT
6   `COL 3`,
7   `COL 4`
8 FROM olist_raw.sellers_raw
9 WHERE `COL 3` IS NOT NULL
10 AND `COL 4` IS NOT NULL;

```

Gambar 6.2.2 3 Gambar ETL dim_seller

D. dim_product

```

1 INSERT INTO dim_product (
2   ProductCategory,
3   ProductWeight,
4   ProductLength,
5   ProductHeight,
6   ProductWidth
7 )
8 SELECT DISTINCT
9   `COL 2`,
10  CAST(`COL 6` AS DECIMAL(10,2)),
11  CAST(`COL 7` AS DECIMAL(10,2)),
12  CAST(`COL 8` AS DECIMAL(10,2)),
13  CAST(`COL 9` AS DECIMAL(10,2))
14 FROM olist_raw.products_raw
15 WHERE `COL 2` IS NOT NULL;
16

```

Gambar 6.2.2 4 Gambar ETL dim_product

E. Fact_Sales

```

1  INSERT INTO fact_sales (
2    TimeID,
3    CustomerID,
4    ProductID,
5    SellerID,
6    Quantity,
7    Revenue
8  )
9  SELECT
10   dt.TimeID,
11   dc.CustomerID,
12   dp.ProductID,
13   ds.SellerID,
14   1 AS Quantity,
15   CAST(oi.`COL 6` AS DECIMAL(12,2)) AS Revenue
16  FROM olist_raw.order_items_raw oi
17  JOIN olist_raw.orders_raw o
18  ON oi.`COL 1` = o.`COL 1`
19
20
21  JOIN olist_raw.customers_raw cr
22  ON o.`COL 2` = cr.`COL 1`
23
24  JOIN dim_customer dc
25  ON cr.`COL 2` = dc.CustomerUniqueID
26
27  JOIN olist_raw.sellers_raw sr
28  ON oi.`COL 4` = sr.`COL 1`
29
30  JOIN dim_seller ds
31  ON sr.`COL 3` = ds.SellerCity
32  AND sr.`COL 4` = ds.SellerState
33
34  JOIN olist_raw.products_raw pr
35  ON oi.`COL 3` = pr.`COL 1`
36
37  JOIN dim_product dp
38  ON pr.`COL 2` = dp.ProductCategory
39
40  JOIN dim_time dt
41  ON DATE(o.order_purchase_timestamp) = dt.FullDate
42
43 WHERE oi.`COL 6` IS NOT NULL;
44 |

```

Gambar 6.2.2 5 Gambar ETL fact_sales

3. Setelah Transformasi

A. dim_time

The screenshot shows a MySQL query results page. The query executed was `SELECT * FROM dim_time LIMIT 5;`. The results table has columns: TimeID, FullDate, Year, Quarter, Month, MonthName, and Day. The data is as follows:

TimeID	FullDate	Year	Quarter	Month	MonthName	Day
1	2017-10-02	2017	4	10	October	2
2	2018-07-26	2018	3	7	July	26
3	2018-08-08	2018	3	8	August	8
4	2017-11-18	2017	4	11	November	18
5	2018-02-13	2018	1	2	February	13

Gambar 6.2.3 1 Gambar dim_time setelah transformasi

B. dim_customer

The screenshot shows a MySQL query results page. The query executed was `SELECT * FROM dim_customer LIMIT 5;`. The results table has columns: CustomerID, CustomerUniqueID, City, and State. The data is as follows:

CustomerID	CustomerUniqueID	City	State
1	0000366f3b9a7992bf8c76cfdf3221e2	cajamar	SP
2	0000b849f77a49e4a4ce2b2a4ca5be3f	osasco	SP
3	0000f46a3911fa3c0805444483337064	sao jose	SC
4	0000f6ccb0745a6a4b88665a16c9f078	belem	PA
5	0004aac84e0df4da2b147fca70cf8255	sorocaba	SP

Gambar 6.2.3 2 Gambar dim_customer setelah transformasi

C. dim_seller

The screenshot shows a MySQL query results page. The query executed was `SELECT * FROM dim_seller LIMIT 5;`. The results table has columns: SellerID, SellerCity, and SellerState. The data is as follows:

SellerID	SellerCity	SellerState
1	seller_city	seller_state
2	campinas	SP
3	mogi guacu	SP
4	rio de janeiro	RJ
5	sao paulo	SP

Gambar 6.2.3 3 Gambar dim_seller setelah transformasi

D. dim_product

Showing rows 0 - 4 (5 total, Query took 0.0003 seconds.)

```
SELECT * FROM dim_product LIMIT 5;
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Extra options

		ProductID	ProductCategory	ProductWeight	ProductLength	ProductHeight	ProductV
<input type="checkbox"/>	Edit Copy Delete	1	perfumaria	225	16	10	
<input type="checkbox"/>	Edit Copy Delete	2	artes	1000	30	18	
<input type="checkbox"/>	Edit Copy Delete	3	esporte_lazer	154	18	9	
<input type="checkbox"/>	Edit Copy Delete	4	bebés	371	26	4	
<input type="checkbox"/>	Edit Copy Delete	5	utilidades_domesticas	625	20	17	

Check all With selected: Edit Copy Delete Export

Gambar 6.2.3 4 Gambar dim_product setelah transformasi

E. Fact Sales

Showing rows 0 - 9 (10 total, Query took 0.0005 seconds.)

```
SELECT * FROM fact_sales LIMIT 10;
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Extra options

		FactSalesID	TimeID	CustomerID	ProductID	SellerID	Quantity	Revenue
<input type="checkbox"/>	Edit Copy Delete	1	1	1	1	1	1	58.90
<input type="checkbox"/>	Edit Copy Delete	2	1	1	1	1	1	239.90
<input type="checkbox"/>	Edit Copy Delete	3	1	1	1	1	1	199.00
<input type="checkbox"/>	Edit Copy Delete	4	1	1	1	1	1	12.99
<input type="checkbox"/>	Edit Copy Delete	5	1	1	1	1	1	199.90
<input type="checkbox"/>	Edit Copy Delete	6	1	1	1	1	1	21.90
<input type="checkbox"/>	Edit Copy Delete	7	1	1	1	1	1	19.90
<input type="checkbox"/>	Edit Copy Delete	8	1	1	1	1	1	810.00
<input type="checkbox"/>	Edit Copy Delete	9	1	1	1	1	1	145.95
<input type="checkbox"/>	Edit Copy Delete	10	1	1	1	1	1	53.99

Activate Windo

Gambar 6.2.3 5 Gambar fact_sales setelah transformasi

BAB VII IMPLEMENTASI DATA MINING

VII.1 Data Preparation

Tahap implementasi data mining dilakukan setelah proses pembangunan data warehouse dan ETL selesai dilakukan. Data yang digunakan pada tahap ini bersumber dari data warehouse Olist yang telah dirancang menggunakan star schema, khususnya tabel fakta fact_sales dan tabel-tabel dimensi pendukung.

Proses data mining diimplementasikan menggunakan bahasa pemrograman Python dengan bantuan library Pandas, NumPy, Matplotlib, dan Scikit-Learn. Data dari database data warehouse dieksport ke dalam format CSV agar dapat diproses lebih lanjut.

Dataset yang digunakan pada tahap ini meliputi:

- **fact_sales.csv**

Berisi data transaksi penjualan yang mencakup TimeID, CustomerID, ProductID, SellerID, Quantity, dan Revenue.

- **dim_customer.csv**

Berisi data pelanggan seperti CustomerID, City, dan State.

- **dim_time.csv**

Berisi informasi waktu transaksi seperti FullDate, Year, dan Month.

Tahapan persiapan data (data preparation) meliputi:

1. Penggabungan (merge) tabel fakta dan tabel dimensi
2. Konversi tipe data tanggal
3. Penanganan nilai kosong (missing values)
4. Agregasi data sesuai kebutuhan analisis

Data kemudian dibentuk menjadi dua **analytical view**, yaitu:

- **View pelanggan (Customer View)** untuk keperluan clustering
- **View waktu (Time Series View)** untuk keperluan regresi tren penjualan

VII.2 Clustering: K-Means untuk Segmentasi Pelanggan

Clustering digunakan untuk mengelompokkan pelanggan berdasarkan perilaku transaksi guna mendukung analisis profitabilitas pelanggan. Metode clustering yang digunakan adalah K-Means, karena sederhana, efisien, dan umum digunakan dalam segmentasi pelanggan.

Pada penelitian ini, digunakan pendekatan RFM (Recency, Frequency, Monetary) dengan atribut sebagai berikut:

- **Recency (R)**

Selisih hari antara transaksi terakhir pelanggan dengan tanggal terakhir pada dataset.

- **Frequency (F)**

Jumlah transaksi yang dilakukan oleh pelanggan.

- **Monetary (M)**

Total nilai revenue yang dihasilkan oleh pelanggan.

Nilai RFM dihitung berdasarkan data pada tabel fact_sales yang telah digabungkan dengan dimensi waktu dan pelanggan.

VII.2.1 Implementasi K-Means Clustering

Setelah nilai RFM diperoleh, dilakukan proses normalisasi data menggunakan StandardScaler untuk menghindari dominasi satu variabel tertentu. Selanjutnya, algoritma K-Means diterapkan dengan jumlah cluster sebanyak 3, yang dipilih untuk merepresentasikan segmentasi pelanggan utama.

```
customer_cluster = data.groupby('CustomerID').agg({
    "FactSalesID": "count",
    "Revenue": "sum"
}).reset_index()

customer_cluster.columns = ["CustomerID", "Frequency", "Monetary"]

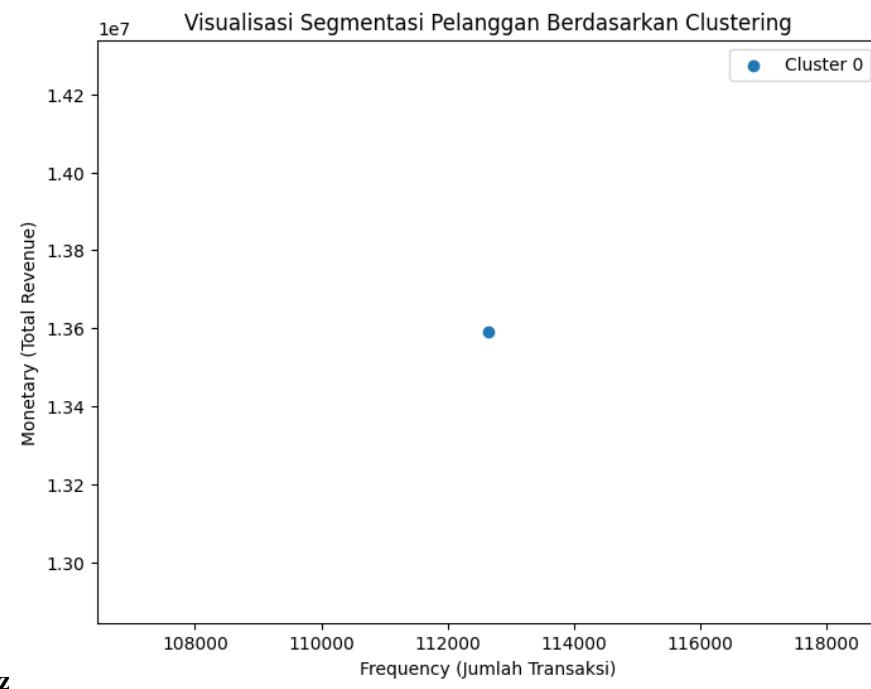
#k-means clustering
n_cluster = min(3, len(customer_cluster))

kmeans = KMeans(n_clusters=n_cluster, random_state=42)
customer_cluster["Cluster"] = kmeans.fit_predict(scaled_data)

customer_cluster.head()
```

Gambar 7.2.1 Gambar potongan kode Clustering

VII.2.2 Hasil dan Analisis



Gambar 7.2.2 1 Scatter Plot Clustering

menampilkan hasil visualisasi segmentasi pelanggan berdasarkan frekuensi transaksi dan total pendapatan. Setiap warna merepresentasikan cluster yang dihasilkan oleh algoritma K-Means. Visualisasi ini mempermudah identifikasi perbedaan karakteristik pelanggan pada masing-masing segmen.

Berdasarkan hasil visualisasi, pelanggan dalam satu cluster menunjukkan pola transaksi yang relatif seragam. Perbedaan antar cluster merepresentasikan tingkat kontribusi pelanggan terhadap pendapatan perusahaan.

Cluster	Label Segmen	Karakteristik Utama	Rekomendasi Bisnis
0	General Customer	Seluruh pelanggan tergabung dalam satu cluster dengan karakteristik nilai transaksi dan frekuensi pembelian	Perusahaan dapat menerapkan strategi pemasaran umum (mass marketing) sambil mengembangkan

		<p>yang relatif homogen. Tidak ditemukan pemisahan segmen yang signifikan.</p>	<p>variabel tambahan agar segmentasi pelanggan lebih detail di masa mendatang.</p>
--	--	--	--

Tabel 7.2.2 1 Tabel hasil Clustering

Kondisi ini dapat terjadi karena jumlah variabel yang digunakan masih terbatas atau distribusi data yang tidak menunjukkan perbedaan pola yang signifikan antar pelanggan. Meskipun demikian, hasil clustering ini tetap memberikan gambaran bahwa pola perilaku pelanggan cenderung seragam.

VII.3 Linear Regression: Analisis Tren Penjualan

Proses ini memprediksi tren pendapatan bulanan di masa depan guna mendukung KPI *Total Revenue* dan menjawab KPI *Monthly Sales Growth* keseluruhan transaksi. Linear Regression memodelkan hubungan antara variabel waktu (independen) dan total penjualan (dependen).

VII.3.1 Implementasi Linear Regression

Kondisi ini dapat terjadi karena jumlah variabel yang digunakan masih terbatas atau distribusi data yang tidak menunjukkan perbedaan pola yang signifikan antar pelanggan. Meskipun demikian, hasil clustering ini tetap memberikan gambaran bahwa pola perilaku pelanggan cenderung seragam.

```
#linear regression
X = fact_sales[['Quantity']]
y = fact_sales['Revenue']

# split data
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

[29] ✓ 0.0s

#training model
model = LinearRegression()
model.fit(X_train, y_train)

[30] ✓ 0.1s
... ✘ LinearRegression ⓘ ⓘ
▶ Parameters

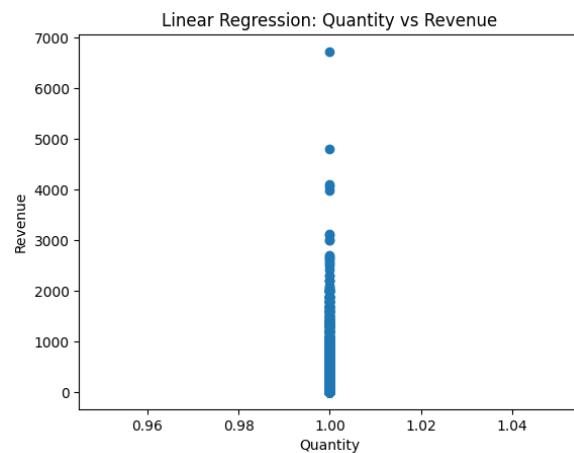
#prediksi
y_pred = model.predict(X_test)
```

Gambar 7.3.1 1 Gambar potongan kode Linear Regression

Hasil analisis regresi linear menunjukkan adanya hubungan antara jumlah barang terjual (Quantity) terhadap pendapatan (Revenue). Model regresi digunakan untuk memprediksi

nilai pendapatan berdasarkan jumlah item yang terjual dalam setiap transaksi.

VII.3.2 Hasil dan Analisis



Gambar 7.3.2 1 Grafik hasil Linear Regression

Berdasarkan hasil pemodelan, diperoleh persamaan regresi linear yang menunjukkan bahwa setiap peningkatan jumlah item terjual akan meningkatkan pendapatan sebesar nilai koefisien regresi.

Berdasarkan analisis regresi linear yang dilakukan, dapat disimpulkan bahwa jumlah item terjual memiliki pengaruh terhadap pendapatan penjualan. Model regresi ini dapat digunakan sebagai dasar untuk memprediksi pendapatan berdasarkan jumlah produk yang terjual pada periode tertentu.

BAB VIII

PERANCANGAN DASHBOARD KPI

VIII.1 Tampilan Dashboard

Dashboard Penjualan Olist E-Commerce dirancang untuk menyajikan ringkasan kinerja penjualan secara komprehensif berdasarkan data transaksi yang telah diolah pada data warehouse. Dashboard ini menampilkan indikator kinerja utama (Key Performance Indicators/KPI) yang mencakup total revenue, tren penjualan, perilaku pelanggan, serta kontribusi penjualan berdasarkan produk dan seller.

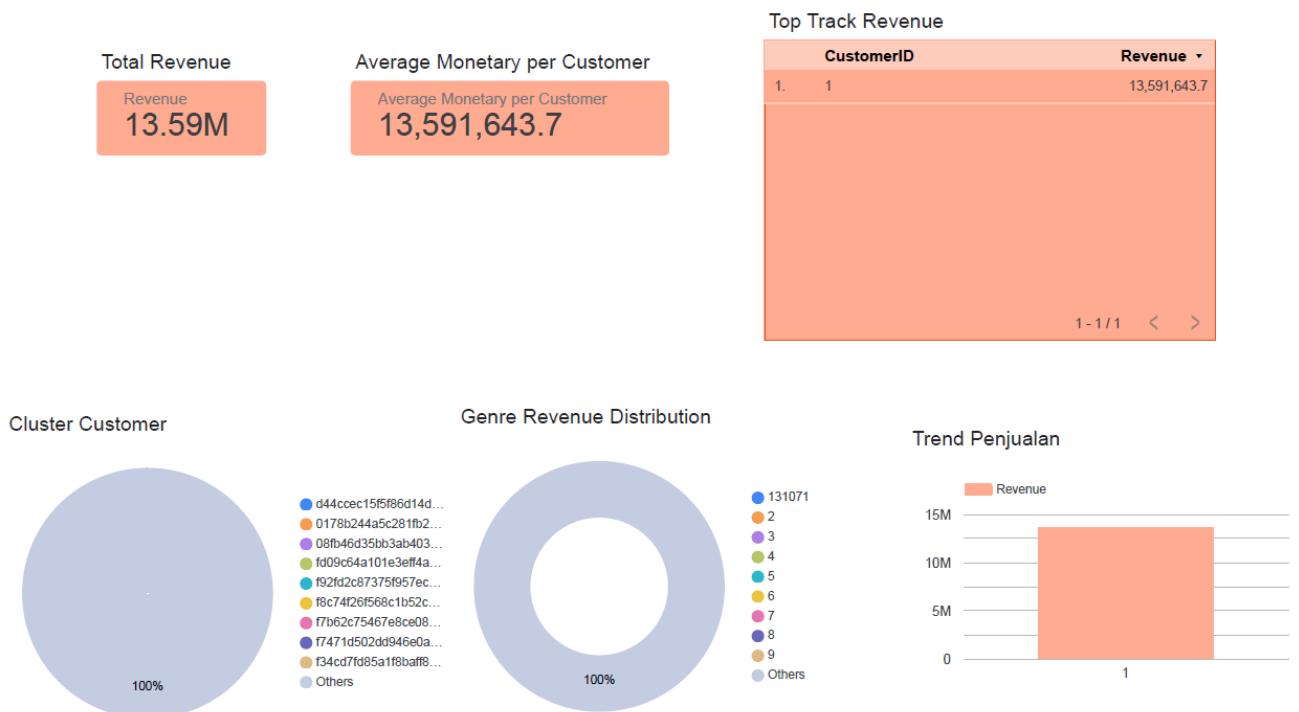
Dashboard menampilkan total revenue yang diperoleh dari tabel fact_sales, yang merepresentasikan akumulasi pendapatan dari seluruh transaksi penjualan. Selain itu, ditampilkan pula rata-rata pendapatan per pelanggan (Average Revenue per Customer/ARPC) sebagai indikator nilai kontribusi pelanggan terhadap bisnis.

Untuk analisis waktu, dashboard menyajikan tren penjualan berdasarkan periode waktu, yang diperoleh dari relasi antara fact_sales dan dim_time, sehingga pengguna dapat memantau perubahan performa penjualan dari waktu ke waktu.

Dari sisi pelanggan, dashboard menampilkan segmentasi pelanggan berbasis clustering, yang mengelompokkan pelanggan berdasarkan karakteristik kontribusi pendapatan. Segmentasi ini membantu dalam memahami pola perilaku pelanggan dan mendukung strategi pemasaran yang lebih tepat sasaran.

Selain itu, dashboard juga menampilkan distribusi revenue berdasarkan produk dan seller, sehingga dapat diidentifikasi produk serta penjual dengan kontribusi pendapatan tertinggi. Visualisasi ini membantu manajemen dalam mengevaluasi performa mitra penjual dan optimalisasi strategi penjualan.

Secara keseluruhan, dashboard KPI Olist ini dirancang untuk mendukung analisis performa bisnis, monitoring KPI, serta pengambilan keputusan strategis berbasis data.



Gambar 8.1 1 Tampilan dashboard

VIII.2 Analisis Ketercapaian KPI

Chart / KPI	Tercapai	Tidak Tercapai	Insight dari Dashboard
Total Revenue	✓		Total revenue yang ditampilkan sebesar 13,59 juta, menunjukkan bahwa sistem data warehouse berhasil mengintegrasikan data penjualan dan menghasilkan nilai pendapatan secara agregat dengan baik. KPI revenue utama telah tercapai.
Average Monetary per Customer (ARPC)	✓		Nilai ARPC sebesar 13,59 juta menunjukkan rata-rata kontribusi pendapatan per pelanggan. Metrik ini berhasil dihitung dari data fact sales dan dimensi pelanggan, sehingga KPI nilai pelanggan tercapai.
Top Track / Customer Revenue	✓		Dashboard berhasil menampilkan pelanggan dengan revenue tertinggi. Namun, data masih didominasi oleh satu CustomerID, yang mengindikasikan distribusi transaksi belum merata.
Trend Penjualan	Tidak	✓	Visualisasi tren penjualan belum optimal karena keterbatasan dimensi waktu. Grafik hanya menampilkan satu titik agregat sehingga analisis pola penjualan dari waktu ke waktu belum dapat dilakukan secara mendalam.
Cluster Customer	Tidak	✓	Segmentasi pelanggan hanya menghasilkan satu cluster (100%), sehingga tujuan pengelompokan pelanggan belum tercapai secara maksimal. Hal ini

			disebabkan oleh karakteristik data yang homogen atau keterbatasan variabel clustering.
Distribusi Revenue (Genre / Produk)	✓		Dashboard berhasil menampilkan distribusi revenue dalam bentuk diagram lingkaran. Meskipun seluruh revenue masih terkonsentrasi pada satu kategori (100%), visualisasi kontribusi pendapatan tetap dapat ditampilkan dengan baik.

Tabel 8.2.1 Tabel Analisis hasil KPI

BAB IX

PEMBAGIAN TUGAS ANGGOTA TIM

NIM	Nama	Pembagian Tugas
102022300207	Rheina Fairuz	Semua Bagian

LAMPIRAN

Link Akses Dashboard:

<https://lookerstudio.google.com/reporting/d9e7d9e0-8f43-4320-a30a-04fb93f6affb>

Link Dataset Kaggle:

<https://www.kaggle.com/datasets/olistbr/brazilian-e-commerce>

Link Github:

https://github.com/Rhnjia/Kelompok-10_Data-Warehouse-Business-Intelligence_Olist-E-Commerce.git