PERANCANGAN DATA WAREHOUSE DAN BUSINESS INTELLIGENCE BERBASIS PERSPEKTIF FINANCIAL UNTUK ANALISIS PROFITABILITAS PADA CHINOOK MUSIC STORE

Disusun untuk memenuhi Tugas Besar
Mata Kuliah Data Warehouse Business Intelligence

Oleh:

Kelompok 6

Yudistira Sebastian Saftari	**102022300313**
Rajasyah Birra Munandar	**102022300284**
Devota Edra Athaloma	**102022300344**
Muhammad Rakha Alfaruq	**102022300368**

PROGRAM STUDI STRATA 1 SISTEM INFORMASI
FAKULTAS REKAYASA INDUSTRI
UNIVERSITAS TELKOM

2025

DAFTAR ISI

Daftar Isi 2
Daftar Gambar 4
Daftar Tabel 5
Bab I Pendahuluan 6
I.1 Latar Belakang 6
I.2 Identifikasi Masalah 6
I.3 Tujuan Penelitian 7
Bab II PERUMUSAN OBJECTIVE, BALANCED SCORECARD DAN KPI8
II.1 Tujuan Utama (Objective) 8
II.2 Perspektif Balanced Scorecard 8
II.3 Key Performance Indicators (KPI) 9
Bab III ANALISIS SUMBER DATA 11
III.1 Deskripsi Sumber Data 11
III.1.1 Sumber Data Utama (Primary Data) 11
III.2 Exploratory Data Analysis (EDA) 12
III.2.1 Pengecekan Kualitas Data (Data Quality Check) 12
III.2.2 Profiling Data dan Distribusi (Data Profiling) Analisis distribusi dilakukan untuk melihat pola awal data: 12
Bab IV PERANCANGAN STAR SCHEMA 14
IV.1 Definisi Star Schema 14
IV.2 Desain Star Schema 14
IV.2.1 Tabel Fakta 14
IV.2.2 Tabel Dimensi 15
IV.2.3 Diagram Star Schema 16
Bab V Implementasi star schema 17
V.1 Data Definition Language (DDL) 17
Bab VI IMPLEMENTASI PROSES ETL 20
VI.1 Alur Proses ETL 20
VI.2 Dokumentasi ETL 21
Bab VII Implementasi data mining 26
VII.1 Data Preparation 26
VII.2 Clustering: K-Means untuk Segmentasi Pelanggan 26
VII.2.1 Implementasi K-Means Clustering 26

VII.2.2 Hasil dan Analisis 28
VII.3 Linear Regression: Analisis Tren Penjualan.....29
VII.3.1 Implementasi Linear Regression 29
VII.3.2 Hasil dan Analisis 30
Bab VIII perancangan dashboard kpi.....31
VIII.1 ..... Tampilan Dashboard 31
VIII.2 ..... Analisis Ketercapaian KPI 32
Bab IX Pembagian tugas anggota tim.....34
LAMPIRAN 35
Link Akses Dashboard: 35

DAFTAR GAMBAR

Gambar 4.2.3 1 Diagram Star Schema 16
5.1 1 DDL dim_waktu 18
5.1 2 DDL dim_musik 18
5.1 3 DDL dim_pelanggan 19
5.1 4 DDL fact_penjualan 19
Gambar 6.2.1 1 Gambar dim_waktu sebelum transformasi 21
Gambar 6.2.1 2 Gambar dim_musik sebelum transformasi 22
Gambar 6.2.1 3 Gambar dim_pelanggan sebelum transformasi 22
Gambar 6.2.1 4 Gambar fact_penjualan sebelum transformasi 22

Gambar 6.2.2 1 Gambar ETL dim_waktu 23
Gambar 6.2.2 2 Gambar ETL dim_musik 23
Gambar 6.2.2 3 Gambar ETL dim_pelanggan 23
Gambar 6.2.2 4 Gambar ETL fact_penjualan 23

Gambar 6.2.3 1 Gambar dim_waktu setelah transformasi 24
Gambar 6.2.3 2 Gambar dim_musik setelah transformasi 24
Gambar 6.2.3 3 Gambar dim_pelanggan setelah transformasi 25
Gambar 6.2.3 4 Gambar fact_penjualan setelah transformasi 25
Gambar 7.2.1 1 Gambar potongan kode Clustering 27
Gambar 7.2.2 1 *Scatter Plot Clustering* 28

Gambar 7.3.1 1 Gambar potongan kode Linear Regression 29
Gambar 7.3.2 1 Grafik hasil Linear Regression 30
Gambar 8.1 1 Tampilan *dashboard* 31

DAFTAR TABEL

Tabel 2.3 1 Analisis KPI 10
Tabel 8.2 1 Tabel Analisis hasil KPI.....33

BAB I
PENDAHULUAN

I.1 Latar Belakang

Industri musik digital merupakan sektor yang sangat dinamis dan kompetitif, di mana profitabilitas perusahaan sangat bergantung pada volume penjualan dan efisiensi strategi pemasaran. *Chinook Music Store*, sebagai salah satu pengecer musik digital berskala global, menghasilkan data transaksi harian yang sangat besar, mencakup rincian *invoice*, informasi pelanggan, hingga performa penjualan per *track* dan *genre*.

Untuk memenangkan persaingan pasar dan meningkatkan kinerja perusahaan, diperlukan transformasi data menjadi informasi yang berharga melalui implementasi *Data Warehouse* dan *Business Intelligence* (BI). Struktur basis data operasional (OLTP) yang ada saat ini dirancang untuk kecepatan transaksi, namun tidak efisien untuk analisis strategis seperti melacak pertumbuhan pendapatan tahunan (*Year-over-Year Growth*), menganalisis kontribusi pendapatan per pelanggan (*Average Revenue per Customer*), atau mengidentifikasi tren penjualan lagu yang paling menguntungkan.

Untuk menjawab tantangan tersebut, diperlukan implementasi *Data Warehouse* dan *Business Intelligence* (BI) yang berfokus pada perspektif finansial. Dengan mentransformasi data mentah menjadi wawasan strategis, perusahaan dapat memantau indikator kinerja utama (KPI) finansial secara akurat. Penelitian ini bertujuan membangun solusi BI yang memungkinkan eksekutif Chinook untuk mengevaluasi pencapaian target pendapatan (*Total Revenue*) dan merumuskan strategi berbasis data guna meningkatkan profitabilitas perusahaan secara berkelanjutan.

I.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, permasalahan yang akan diselesaikan dalam laporan ini meliputi:

1. Bagaimana merancang arsitektur *Data Warehouse* dengan menggunakan *Star Schema* yang sesuai untuk kebutuhan analisis bisnis Chinook?

2. Bagaimana melakukan proses integrasi data (ETL) dari sumber data utama dan pendukung ke dalam data warehouse secara efektif?
3. Bagaimana menerapkan metode *Data Mining* (seperti klasifikasi, regresi, atau klastering) untuk menggali pola tersembunyi dari data Chinook guna menjawab KPI yang ditentukan?
4. Bagaimana menyajikan hasil analisis dalam bentuk *dashboard* interaktif yang informatif untuk mendukung pengambilan keputusan?

I.3 Tujuan Penelitian

Tujuan dari pengerjaan tugas besar ini adalah:

1. Membangun *Data Warehouse* dengan bentuk *Star Schema* dan mengimplementasikannya ke dalam RDBMS untuk mendukung kinerja perusahaan.
2. Melakukan proses ETL (*Extract, Transform, Load*) dari sumber data utama (Chinook) yang siap dianalisis.
3. Mengimplementasikan proses Data Mining untuk memberikan solusi prediktif atau deskriptif terhadap pendapatan perusahaan
4. Menampilkan hasil pengolahan data dan visualisasi KPI pada *dashboard* serta memberikan rekomendasi bisnis kepada tim eksekutif.

BAB II
PERUMUSAN OBJECTIVE, BALANCED SCORECARD DAN KPI

II.1 Tujuan Utama (Objective)

Tujuan utama dari penelitian ini adalah meningkatkan kinerja finansial Chinook Music Store melalui pemanfaatan data warehouse dan business intelligence sebagai alat pendukung pengambilan keputusan strategis. Seiring dengan berkembangnya industri musik digital, Chinook Music Store menghasilkan data transaksi dalam jumlah besar yang mencerminkan aktivitas penjualan, preferensi pelanggan, serta performa produk. Namun, data tersebut masih bersifat operasional sehingga belum dapat dimanfaatkan secara optimal untuk analisis strategis.

Melalui penerapan data warehouse, data transaksi akan diintegrasikan dan diorganisasikan dalam bentuk yang lebih terstruktur sehingga mendukung analisis historis dan analisis tren. Selanjutnya, business intelligence digunakan untuk mengolah data tersebut menjadi informasi yang relevan, khususnya terkait pendapatan, pertumbuhan penjualan, serta kontribusi produk dan pelanggan terhadap kinerja keuangan perusahaan. Dengan demikian, objective ini tidak hanya berfokus pada penyajian data, tetapi juga pada penyediaan insight yang dapat digunakan 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 digital yang bergerak di bidang penjualan musik, Chinook Music Store sangat bergantung pada pendapatan yang dihasilkan dari transaksi pelanggan. Oleh karena itu, pengukuran terhadap aspek finansial seperti total pendapatan, pertumbuhan penjualan, dan kontribusi produk menjadi sangat relevan untuk menggambarkan kondisi bisnis perusahaan secara menyeluruh.

Selain itu, perspektif finansial juga 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 Chinook Music Store 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 (Invoice, InvoiceLine)	Target: Total Revenue ≥ $500.00	Scorecard
Meningkatkan pertumbuhan penjualan dari waktu ke waktu	Monthly Sales Growth	Data penjualan bulanan (InvoiceDate)	Target: Growth YoY ≥ 10%	Time Series Chart
Meningkatkan nilai kontribusi pelanggan terhadap pendapatan	Average Revenue per Customer (ARPC)	Data pelanggan & transaksi (Customer, Invoice)	Target: semua customer ARPC menyentuh ≥ 40	Bar Chart, Pie Chart
Mengoptimalkan penjualan lagu dengan performa tertinggi	Top Track Revenue Contribution	Data lagu & transaksi (Track, InvoiceLine)	Target: Top 10 Track ≥ 30% dari total revenue	Table Chart

Mengetahui genre musik yang paling menguntungkan	Genre Revenue Distribution	Data genre & penjualan (Genre, Track)	Target: Top Genre ≥ 35% dari total revenue	Pie Chart

Tabel 2.3 1 Analisis KPI

BAB III
ANALISIS SUMBER DATA

III.1 Deskripsi Sumber Data

Dalam perancangan *Data Warehouse* ini, digunakan dua jenis sumber data untuk memenuhi kebutuhan analisis kinerja finansial perusahaan, yaitu sumber data utama dan sumber data pendukung.

III.1.1 Sumber Data Utama (Primary Data)

Sumber data utama yang digunakan berasal dari Chinook Database, sebuah dataset publik yang merepresentasikan data operasional toko musik digital. Dataset ini diperoleh dalam format CSV (Comma Separated Values) dengan pemisah berupa titik koma (;).

Data ini terdiri dari 11 tabel yang mencakup seluruh aktivitas bisnis perusahaan. Untuk keperluan analisis perspektif finansial, tabel-tabel yang menjadi fokus utama meliputi:

1. **Invoice.csv:** Menyimpan data transaksi penjualan level *header*, mencakup InvoiceId, InvoiceDate, CustomerId, dan Total pembayaran.
2. **InvoiceLine.csv:** Menyimpan rincian item transaksi (lagu) yang dibeli, mencakup UnitPrice, Quantity, dan TrackId. Tabel ini menjadi sumber utama perhitungan pendapatan (*revenue*).
3. **Customer.csv:** Berisi informasi demografis pelanggan seperti City, Country, dan Email yang digunakan untuk dimensi pelanggan.
4. **Track.csv:** Berisi data detail lagu, harga satuan, dan durasi. Tabel ini terhubung dengan tabel referensi lainnya.
5. **Tabel Referensi (Genre, Album, Artist, MediaType):** Tabel-tabel dimensi yang memberikan konteks terhadap produk yang dijual, digunakan untuk analisis kontribusi pendapatan berdasarkan kategori musik.

III.2 Exploratory Data Analysis (EDA)

Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk memahami karakteristik data, mengidentifikasi anomali, serta memvalidasi kualitas data sebelum proses ETL.

III.2.1 Pengecekan Kualitas Data (Data Quality Check)

Berdasarkan pemeriksaan awal terhadap file CSV, ditemukan beberapa karakteristik sebagai berikut:

- **Format Tanggal:** Kolom InvoiceDate pada tabel Invoice memiliki format YYYY-MM-DD HH:MM:SS (contoh: 2009-01-01 00:00:00). Hal ini memerlukan transformasi saat proses ETL untuk memisahkan komponen tanggal, bulan, dan tahun ke dalam Dimensi_Waktu.
- **Delimiter:** Format file menggunakan *delimiter* titik koma (;), bukan koma standar. Konfigurasi ini harus disesuaikan pada *tools* Pentaho agar pembacaan kolom tidak berantakan.
- **Konsistensi Numerik:** Dilakukan validasi pada kolom Total di tabel Invoice. Hasil pengecekan menunjukkan bahwa nilai Total di Invoice konsisten dengan penjumlahan (Quantity × UnitPrice) dari tabel InvoiceLine. Oleh karena itu, perhitungan *revenue* di tabel fakta akan menggunakan kalkulasi ulang dari InvoiceLine untuk menjaga granularitas data.

III.2.2 Profiling Data dan Distribusi (Data Profiling) Analisis distribusi dilakukan untuk melihat pola awal data:

1. **Missing Values (Nilai Kosong):**
    a. Tabel Track: Kolom Composer memiliki banyak nilai kosong (*null*). Penanganan dilakukan dengan mengganti nilai *null* menjadi "Unknown Composer" pada proses ETL.
    b. Tabel Customer: Kolom Company dan State memiliki banyak nilai kosong, namun hal ini wajar untuk data pelanggan perorangan (B2C).
2. **Distribusi Harga (Unit Price):**
    a. Mayoritas UnitPrice pada Track.csv adalah 0.99 (lagu standar) dan 1.99 (video/film). Variasi harga ini akan mempengaruhi total pendapatan.

3. **Rentang Waktu Data:**
    a. Data transaksi mencakup periode beberapa tahun (2009–2013). Analisis tren pertumbuhan (*Sales Growth*) akan difokuskan pada perbandingan *Year-over-Year* (YoY) berdasarkan rentang waktu tersebut

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 data serta melakukan analisis multidimensi berdasarkan berbagai sudut pandang seperti waktu, pelanggan, dan produk.

Dalam konteks analisis kinerja finansial Chinook Music Store, star schema digunakan untuk mendukung analisis pendapatan, tren penjualan, serta kontribusi produk dan pelanggan 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 dan beberapa tabel dimensi yang memberikan konteks terhadap data tersebut.

IV.2.1 Tabel Fakta

Tabel fakta yang digunakan adalah fact_sales, yang merepresentasikan transaksi penjualan lagu. Tabel ini menyimpan data numerik yang menjadi dasar perhitungan indikator kinerja utama (KPI) finansial.

Atribut utama dalam tabel fact_sales meliputi:

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

- CustomerID sebagai foreign key yang mengacu pada dimensi pelanggan,
- TrackID sebagai foreign key yang mengacu pada dimensi musik,
- Quanity sebagai jumlah item yang terjual,
- Amount sebagai total pendapatan hasil perkalian Quantity dan UnitPrice (dim_musik).

Granularitas tabel fakta berada pada level transaksi lagu per invoice, sehingga memungkinkan analisis yang detail maupun agregat.

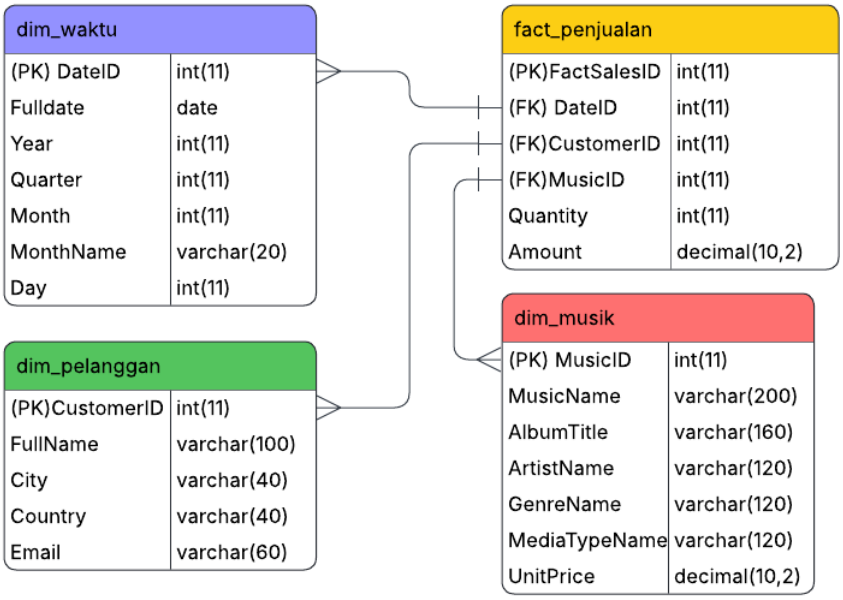
IV.2.2 Tabel Dimensi

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

1. **Dimensi Waktu (dim_waktu)**
   Digunakan untuk menganalisis data penjualan berdasarkan periode waktu tertentu, seperti tahun, bulan, dan tanggal. Dimensi ini berperan penting dalam analisis tren penjualan dan pertumbuhan pendapatan. berikut adalah kolom yang terdapat pada dim_waktu: DateID, FullDate, Year, Quarter, Month, MonthName, Day.
2. **Dimensi Pelanggan (dim_pelanggan)**
   Menyimpan informasi pelanggan seperti nama, kota, dan negara. Dimensi ini digunakan untuk menganalisis kontribusi pelanggan terhadap pendapatan serta menghitung rata-rata pendapatan per pelanggan. Berikut adalah kolom yang terdapat pada dim_pelanggan: CustomerID, FullName, City, Country, Email.
3. **Dimensi Musik (dim_musik)**
   Dimensi ini merupakan hasil penggabungan beberapa tabel referensi, yaitu track, album, artist, genre, dan media type. Penggabungan dilakukan untuk menyederhanakan struktur dimensi dan mempermudah analisis kontribusi pendapatan berdasarkan genre, artis, maupun lagu. Berikut adalah kolom yang terdapat pada dim_musik: TrackID, TrackName, Album Title, ArtistName, GenreName, MediaTypeName, dan UnitPrice.

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.

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 diterjemahkan ke dalam bentuk fisik melalui perintah Data Definition Language (DDL).

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

Setelah tabel dimensi berhasil dibuat, tabel fakta fact_penjualan dibangun dengan menambahkan foreign key yang mengacu pada primary key di masing-masing tabel dimensi. Penerapan foreign key constraint bertujuan untuk menjaga integritas referensial antar tabel.

Dengan implementasi DDL ini, struktur star schema telah siap digunakan sebagai fondasi untuk proses ETL dan analisis data lebih lanjut. Berikut adalah query DDL untuk pembuatan tabel dalam star schema:

1. dim_waktu

```sql
CREATE TABLE Dim_Waktu (
    DateID INT NOT NULL PRIMARY KEY,
    FullDate DATE NOT NULL,
    Year INT NOT NULL,
    Quarter INT NOT NULL,
    Month INT NOT NULL,
    MonthName VARCHAR(20) NOT NULL,
    Day INT NOT NULL
);
```

5.1 1 DDL dim_waktu

2. dim_musik

```sql
CREATE TABLE Dim_Pelanggan (
    CustomerID INT NOT NULL PRIMARY KEY,
    FullName VARCHAR(100) NOT NULL,
    City VARCHAR(40) NOT NULL,
    Country VARCHAR(40) NOT NULL,
    Email VARCHAR(60) NOT NULL
);
```

5.1 2 DDL dim_musik

3. dim_pelanggan

```sql
CREATE TABLE Dim_Musik (
MusicID INT NOT NULL PRIMARY KEY,
MusicName VARCHAR(200) NOT NULL,
AlbumTitle VARCHAR(160) NOT NULL DEFAULT 'Unknown Album',
ArtistName VARCHAR(120) NOT NULL DEFAULT 'Unknown Artist',
GenreName VARCHAR(120) NOT NULL,
MediaTypeName VARCHAR(120) NOT NULL,
UnitPrice DECIMAL(10,2) NOT NULL
);
```

5.1 3 DDL dim_pelanggan

4. fact_penjualan

```sql
CREATE TABLE Fact_Penjualan (
FactsalesID INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
DateID INT NOT NULL,
CustomerID INT NOT NULL,
MusicID INT NOT NULL,
Quantity INT NOT NULL,
Amount DECIMAL(10,2) NOT NULL,

CONSTRAINT fk_date FOREIGN KEY (DateID) REFERENCES Dim_Waktu(DateID),
CONSTRAINT fk_customer FOREIGN KEY (CustomerID) REFERENCES Dim_Pelanggan(CustomerID),
CONSTRAINT fk_track FOREIGN KEY (MusicID) REFERENCES Dim_Musik(MusicID)
);
```

5.1 4 DDL fact_penjualan

BAB VI
IMPLEMENTASI PROSES ETL

VI.1 Alur Proses ETL

A. Konsep ETL

Proses ETL (Extract, Transform, Load) merupakan tahapan penting dalam pembangunan data warehouse karena berfungsi memindahkan data dari sistem sumber ke dalam struktur star schema. Pada pengerjaan tugas ini, proses ETL dilakukan menggunakan Pentaho Data Integration (PDI). Tahapan ETL yang dilakukan meliputi:

1. **Extract**

   Tahap extract dilakukan dengan mengambil data dari sumber data Chinook yang tersedia dalam format CSV. Data yang diekstraksi mencakup tabel Invoice, InvoiceLine, Customer, Track, serta tabel referensi seperti Genre, Album, Artist, dan MediaType. Pada tahap ini dilakukan penyesuaian delimiter file agar data dapat dibaca dengan benar oleh tools ETL.

2. **Transform**

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

   - Pemisahan atribut tanggal transaksi menjadi tahun, bulan, dan tanggal untuk dimensi waktu.
   - Perhitungan ulang nilai pendapatan berdasarkan perkalian UnitPrice dan Quantity.
   - Penanganan nilai kosong (null) pada beberapa atribut.
   - Penggabungan data musik dari beberapa tabel referensi untuk membentuk dimensi musik.
   - Pemetaan surrogate key dari tabel dimensi ke tabel fakta menggunakan database lookup.

   Transformasi ini memastikan data yang dimuat ke dalam data warehouse bersih, konsisten, dan siap dianalisis.

3. **Load**

Tahap load merupakan proses memasukkan data hasil transformasi ke dalam tabel dimensi dan tabel fakta. Proses loading dilakukan secara bertahap, dimulai dari tabel dimensi dan diakhiri dengan tabel fakta untuk menjaga integritas foreign key.

B. Tools yang digunakan

Tools yang digunakan dalam pengerjaan tugas adalah sebagai berikut:

1. Database: MySQL, yang digunakan untuk menampung database sumber Pemesanan dan database tujuan *Chinook DataMart*
2. Tools ETL: Pentaho Data Integration (PDI) untuk merancang alur kerja ETL secara visual.

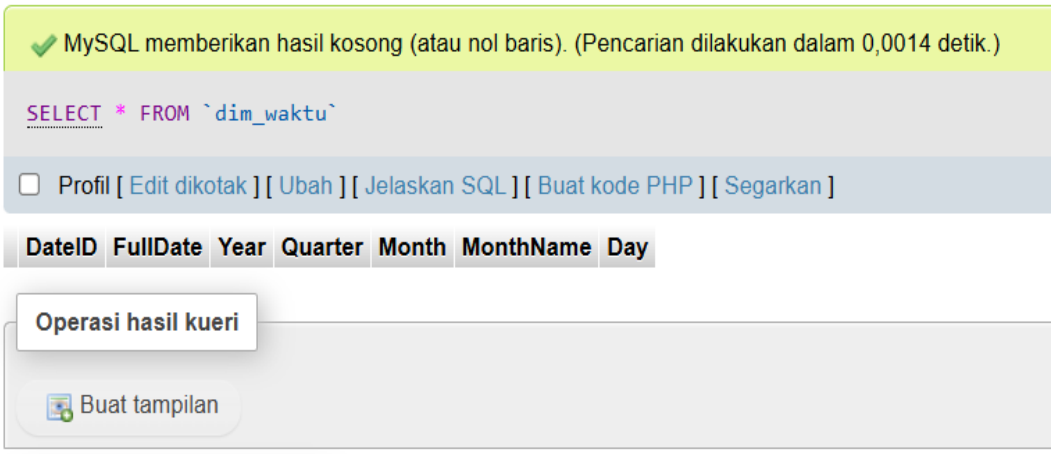
VI.2 Dokumentasi ETL

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_waktu



Gambar 6.2.1 1 Gambar dim_waktu sebelum transformasi

B. dim_musik

MySQL memberikan hasil kosong (atau nol baris). (Pencarian dilakukan dalam 0,0002 detik.)

```
SELECT * FROM `dim_musik`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

MusicID	MusicName	AlbumTitle	ArtistName	GenreName	MediaTypeName	UnitPrice

Operasi hasil kueri

Buat tampilan

Gambar 6.2.1 2 Gambar dim_musik sebelum transformasi

C. dim_pelanggan

MySQL memberikan hasil kosong (atau nol baris). (Pencarian dilakukan dalam 0,0007 detik.)

```
SELECT * FROM `dim_pelanggan`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

CustomerID	FullName	City	Country	Email

Operasi hasil kueri

Buat tampilan

Gambar 6.2.1 3 Gambar dim_pelanggan sebelum transformasi

D. fact_penjualan

MySQL memberikan hasil kosong (atau nol baris). (Pencarian dilakukan dalam 0,0004 detik.)

```
SELECT * FROM `fact_penjualan`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

FactSalesID	DateID	CustomerID	MusicID	Quantity	Amount

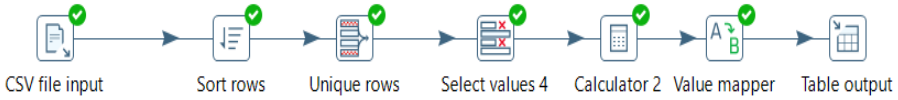
Operasi hasil kueri

Buat tampilan

Gambar 6.2.1 4 Gambar fact_penjualan sebelum transformasi

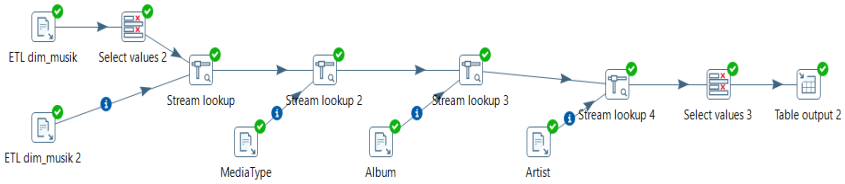
2. Proses Transformasi / ETL

A. dim_waktu



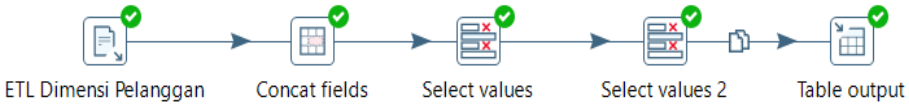
Gambar 6.2.2 1 Gambar ETL dim_waktu

B. dim_musik



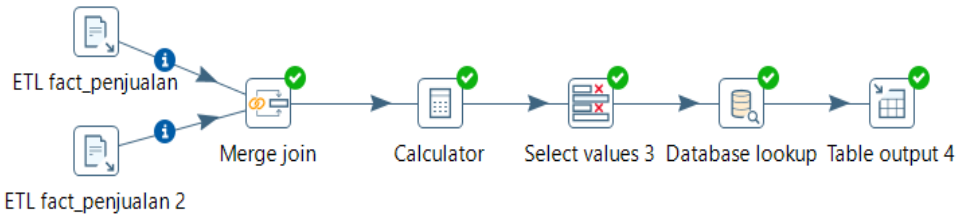
Gambar 6.2.2 2 Gambar ETL dim_musik

C. dim_pelanggan



Gambar 6.2.2 3 Gambar ETL dim_pelanggan

D. fact_penjualan



Gambar 6.2.2 4 Gambar ETL fact_penjualan

3. Setelah Transformasi

A. dim_waktu

Menampilkan baris 0 - 24 (total 412, Pencarian dilakukan dalam 0,0002 detik.)

```
SELECT * FROM `dim_waktu`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

1 > >> | Tampilkan semua | Jumlah baris: 25 Saring baris: Cari di tabel ini Sort by key: Tidak ada

Extra options

	DateID	FullDate	Year	Quarter	Month	MonthName	Day
Ubah Salin Hapus	1	2009-01-01	2009	1	1	January	1
Ubah Salin Hapus	2	2009-01-02	2009	1	1	January	2
Ubah Salin Hapus	3	2009-01-03	2009	1	1	January	3
Ubah Salin Hapus	4	2009-01-06	2009	1	1	January	6
Ubah Salin Hapus	5	2009-01-11	2009	1	1	January	11
Ubah Salin Hapus	6	2009-01-19	2009	1	1	January	19
Ubah Salin Hapus	7	2009-02-01	2009	1	2	February	1
Ubah Salin Hapus	8	2009-02-01	2009	1	2	February	1
Ubah Salin Hapus	9	2009-02-02	2009	1	2	February	2
Ubah Salin Hapus	10	2009-02-03	2009	1	2	February	3
Ubah Salin Hapus	11	2009-02-06	2009	1	2	February	6
Ubah Salin Hapus	12	2009-02-11	2009	1	2	February	11
Ubah Salin Hapus	13	2009-02-19	2009	1	2	February	19
Ubah Salin Hapus	14	2009-03-04	2009	1	3	March	4
Ubah Salin Hapus	15	2009-03-04	2009	1	3	March	4

Konsol

Gambar 6.2.3 1 Gambar dim_waktu setelah transformasi

B. dim_musik

Menampilkan baris 0 - 24 (total 3503, Pencarian dilakukan dalam 0,0002 detik.)

```
SELECT * FROM `dim_musik`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

1 > >> | Jumlah baris: 25 Saring baris: Cari di tabel ini Sort by key: Tidak ada

Extra options

	MusicID	MusicName	AlbumTitle	ArtistName	GenreName	MediaTypeName	UnitPrice
Ubah Salin Hapus	1	For Those About To Rock (We Salute You)	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	2	Balls to the Wall	Balls to the Wall	Accept	Rock	Protected AAC audio file	0.99
Ubah Salin Hapus	3	Fast As a Shark	Restless and Wild	Accept	Rock	Protected AAC audio file	0.99
Ubah Salin Hapus	4	Restless and Wild	Restless and Wild	Accept	Rock	Protected AAC audio file	0.99
Ubah Salin Hapus	5	Princess of the Dawn	Restless and Wild	Accept	Rock	Protected AAC audio file	0.99
Ubah Salin Hapus	6	Put The Finger On You	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	7	Let's Get It Up	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	8	Inject The Venom	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	9	Snowballed	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	10	Evil Walks	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	11	C.O.D.	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	12	Breaking The Rules	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	13	Night Of The Long Knives	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99
Ubah Salin Hapus	14	Spellbound	For Those About To Rock We Salute You	AC/DC	Rock	MPEG audio file	0.99

Gambar 6.2.3 2 Gambar dim_musik setelah transformasi

C. dim_pelanggan

Showing rows 0 - 58 (59 total, Query took 0.0007 seconds.)

```
SELECT * FROM `dim_pelanggan`
```

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

Show all | Number of rows: All Filter rows: Search this table Sort by key: None

Extra options

←T→			CustomerID	FullName	City	Country	Email
Edit	Copy	Delete	1	Luís Gonçalves	São José dos Campos	Brazil	luisg@embraer.com.br
Edit	Copy	Delete	2	Leonie Köhler	Stuttgart	Germany	leonekohler@surfeu.de
Edit	Copy	Delete	3	François Tremblay	Montréal	Canada	ftremblay@gmail.com
Edit	Copy	Delete	4	Bjørn Hansen	Oslo	Norway	bjorn.hansen@yahoo.no
Edit	Copy	Delete	5	František Wichterlová	Prague	Czech Republic	frantisekw@jetbrains.com
Edit	Copy	Delete	6	Helena Holý	Prague	Czech Republic	hholy@gmail.com
Edit	Copy	Delete	7	Astrid Gruber	Vienne	Austria	astrid.gruber@apple.at
Edit	Copy	Delete	8	Daan Peeters	Brussels	Belgium	daan_peeters@apple.be
Edit	Copy	Delete	9	Kara Nielsen	Copenhagen	Denmark	kara.nielsen@jubil.dk
Edit	Copy	Delete	10	Eduardo Martins	São Paulo	Brazil	eduardo@woodstock.com.br
Edit	Copy	Delete	11	Alexandre Rocha	São Paulo	Brazil	alero@uol.com.br
Edit	Copy	Delete	12	Roberto Almeida	Rio de Janeiro	Brazil	roberto.almeida@riotur.gov.br
Edit	Copy	Delete	13	Fernanda Ramos	Brasília	Brazil	fernadaramos4@uol.com.br
Edit	Copy	Delete	14	Mark Philips	Edmonton	Canada	mphilips12@shaw.ca

Gambar 6.2.3 3 Gambar dim_pelanggan setelah transformasi

D. fact_penjualan

Menampilkan baris 0 - 24 (total 2240, Pencarian dilakukan dalam 0,0002 detik.)

```
SELECT * FROM `fact_penjualan`
```

Profil [Edit dikotak] [Ubah] [Jelaskan SQL] [Buat kode PHP] [Segarkan]

1 > >> | Jumlah baris: 25 Saring baris: Cari di tabel ini Sort by key: Tidak ada

Extra options

←T→			FactSalesID	DateID	CustomerID	TrackID	Quantity	Amount
Ubah	Salin	Hapus	1	1	2	2	1	0.99
Ubah	Salin	Hapus	2	1	2	4	1	0.99
Ubah	Salin	Hapus	3	2	4	6	1	0.99
Ubah	Salin	Hapus	4	2	4	8	1	0.99
Ubah	Salin	Hapus	5	2	4	10	1	0.99
Ubah	Salin	Hapus	6	2	4	12	1	0.99
Ubah	Salin	Hapus	7	3	8	16	1	0.99
Ubah	Salin	Hapus	8	3	8	20	1	0.99
Ubah	Salin	Hapus	9	3	8	24	1	0.99
Ubah	Salin	Hapus	10	3	8	28	1	0.99
Ubah	Salin	Hapus	11	3	8	32	1	0.99
Ubah	Salin	Hapus	12	3	8	36	1	0.99
Ubah	Salin	Hapus	13	4	14	42	1	0.99
Ubah	Salin	Hapus	14	4	14	48	1	0.99
Ubah	Salin	Hapus	15	4	14	54	1	0.99

Konsol

Gambar 6.2.3 4 Gambar fact_penjualan setelah transformasi

BAB VII IMPLEMENTASI DATA MINING

VII.1 Data Preparation

Implementasi data mining dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka *Pandas* untuk manipulasi data dan *Scikit-Learn* untuk pemodelan algoritma. Sumber data yang digunakan berasal dari *Data Warehouse* yang telah dibangun sebelumnya, yaitu:

1. fact_penjualan.csv: Berisi data transaksi penjualan (Fakta).
2. dim_pelanggan.csv: Berisi data profil pelanggan (Dimensi).
3. dim_waktu.csv: Berisi informasi waktu transaksi (Dimensi).

Proses persiapan data (*preprocessing*) meliputi penggabungan tabel (*merging*), konversi tipe data tanggal, dan penanganan nilai kosong. Data transaksi kemudian diaggregasi ke dalam dua bentuk *Analytical View* yang berbeda sesuai kebutuhan algoritma:

- View Pelanggan (RFM): Agregasi per CustomerID untuk menghitung *Recency*, *Frequency*, dan *Monetary*.
- View Waktu (*Time Series*): Agregasi per Bulan (*MonthYear*) untuk menghitung total pendapatan bulanan.

VII.2 Clustering: K-Means untuk Segmentasi Pelanggan

Metode ini dilakukan guna mengelompokkan pelanggan berdasarkan perilaku pembelian mereka (Recency, Frequency, Monetary) untuk menjawab KPI Average Revenue per Customer dan strategi retensi. Atribut yang digunakan

VII.2.1 Implementasi K-Means Clustering

Atribut yang digunakan dalam proses clustering adalah:

1. **Recency (R):** Jumlah hari sejak transaksi terakhir pelanggan. Semakin kecil nilainya, semakin aktif pelanggan tersebut.
2. **Frequency (F):** Jumlah frekuensi transaksi yang dilakukan pelanggan.

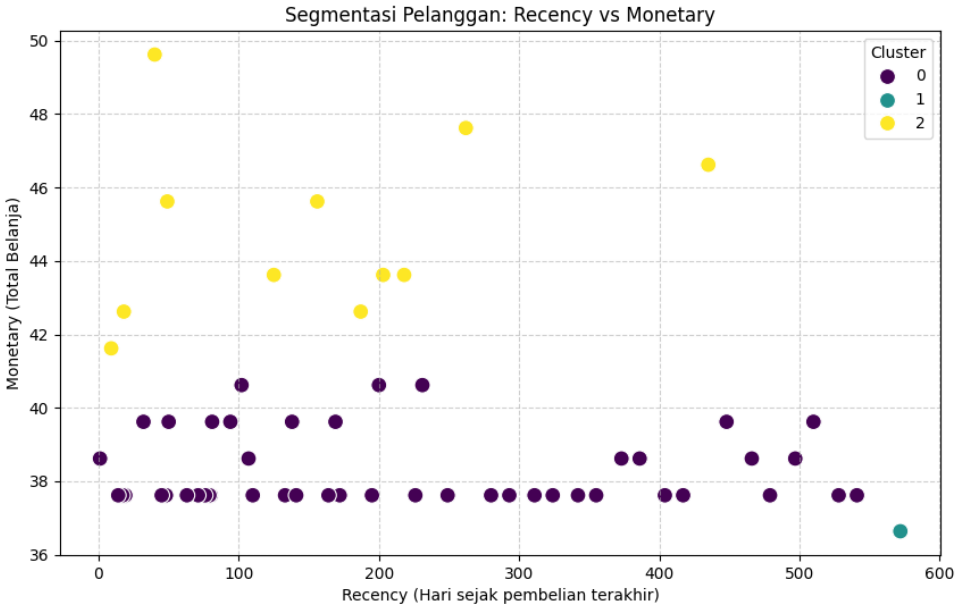
3. **Monetary (M):** Total nilai uang (*Revenue*) yang telah dikeluarkan oleh pelanggan.

Berikut adalah potongan kode Python yang digunakan untuk proses normalisasi dan clustering:

```python
rfm = sales_data.groupby('CustomerID').agg({
    'Date': lambda x: (snapshot_date - x.max()).days,
    'FactSalesID': 'count',
    'Amount': 'sum'
}).reset_index()

rfm.columns = ['CustomerID', 'Recency', 'Frequency', 'Monetary']

scaler = StandardScaler()
rfm_scaled = scaler.fit_transform(rfm[['Recency', 'Frequency', 'Monetary']])


kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
rfm['Cluster'] = kmeans.fit_predict(rfm_scaled)
```

Gambar 7.2.1 1 Gambar potongan kode Clustering

VII.2.2 Hasil dan Analisis



Gambar 7.2.2 1 *Scatter Plot Clustering*

Berdasarkan hasil eksekusi algoritma clustering menggunakan python, membentuk grafik scatter plot yang memetakan pelanggan. Terbentuk 3 kelompok pelanggan berdasarkan Recency dan Monetary sebagai berikut:

Cluster	Label Segmen	Karakteristik Utama	Rekomendasi Bisnis
0	Standar	Recency menengah, total belanja rata-rata	Pertahankan dengan layanan standar
1	Dormant	Recency tinggi (sudah lama tidak melakukan transaksi).	Perlu adanya strategi untuk membuat tertarik kembali
2	Premium	Recency rendah (belum lama melakukan transaksi),	Prioritas utama untuk retensi

		Monetary Tertinggi ($44.8).	

Tabel 7.2.2 1 Tabel hasil Clustering

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

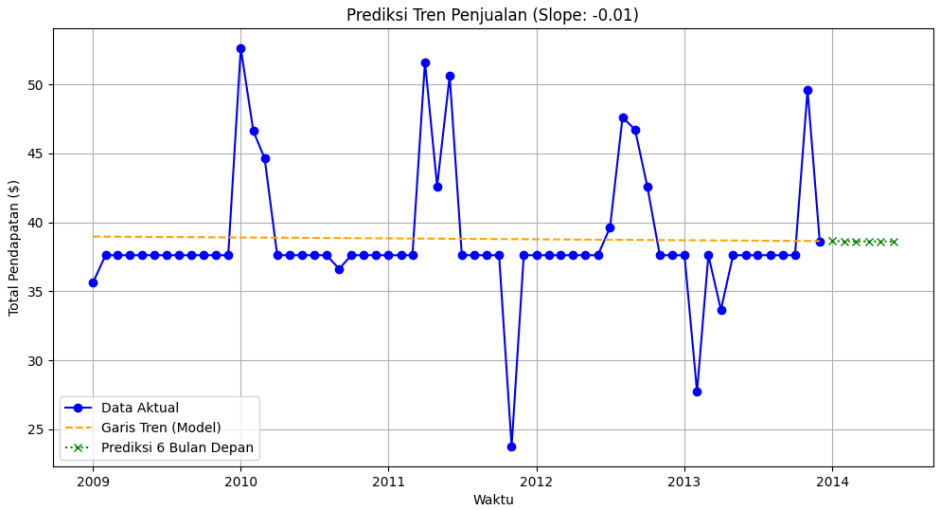
Linear Regression memodelkan hubungan antara variabel waktu (independen) dan total penjualan (dependen). Variabel yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen.

```
X = monthly_sales[['TimeIndex']]
y = monthly_sales['Amount']

reg = LinearRegression()
reg.fit(X, y)

y_pred = reg.predict(X)

last_index = monthly_sales['TimeIndex'].iloc[-1]
future_X = np.arange(last_index + 1, last_index + 7).reshape(-1, 1)
future_pred = reg.predict(future_X)
future_dates = [monthly_sales['DatePlot'].iloc[-1] + pd.DateOffset(months=i) for i in range(1, 7)]

r2 = r2_score(y, y_pred)
slope = reg.coef_[0]
```

Gambar 7.3.1 1 Gambar potongan kode Linear Regression

1. **Variabel X (Independent):** Indeks Waktu (*Time Index*) yang merepresentasikan urutan bulan (0, 1, 2, dst).
2. **Variabel Y (Dependent):** Total Pendapatan Bulanan (*Monthly Revenue*).

VII.3.2 Hasil dan Analisis



Gambar 7.3.2 1 Grafik hasil Linear Regression

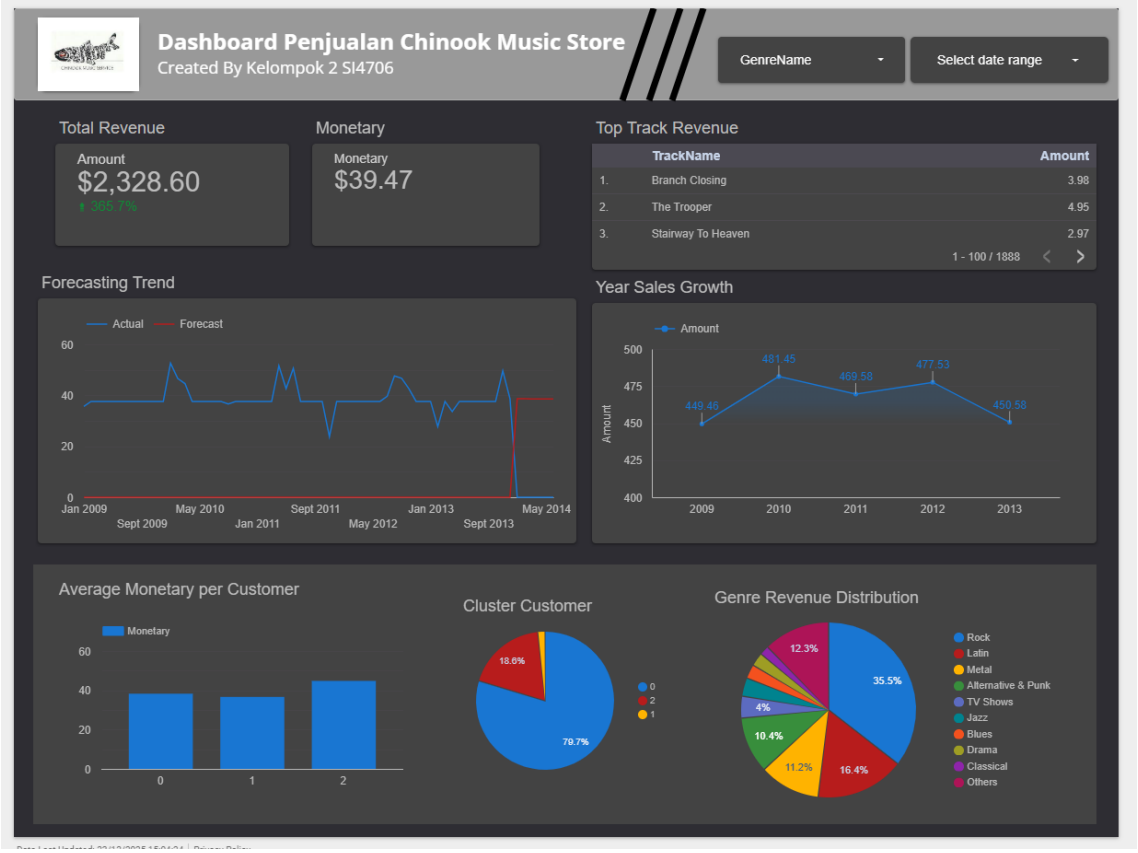
Berdasarkan hasil implementasi metode Linear Regression, model menghasilkan persamaan garis tren penjualan dengan karakteristik statistik. Model menampilkan nilai slope sebesar –0.01 yang mengindikasikan tidak adanya perubahan signifikan antara variabel independen dan variabel dependen.

Titik-titik data aktual menunjukkan tren penjualan data sangat fluktuaktif setiap bulannya. Secara statiktif, tidak ditemukan bukti adanya pertumbuhan penjualan organik yang signifikan pada periode penjualan. Prediksi untuk 6 bulan kedepan menunjukkan angka yang stabil di kisaran $37-$40 perbulan. Hal ini merupakan sinyal bagi manajemen untuk melakukan intervensi strategi penjualan.

BAB VIII
PERANCANGAN DASHBOARD KPI

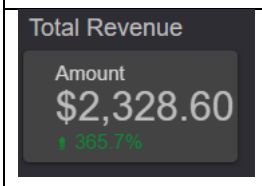
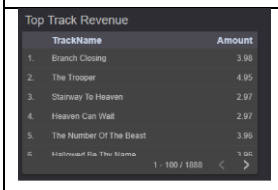
VIII.1 Tampilan Dashboard

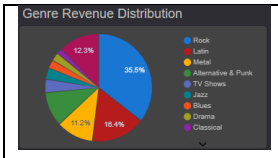
Dashboard Penjualan Chinook Music Store menyajikan ringkasan kinerja penjualan secara komprehensif berdasarkan revenue, pertumbuhan penjualan, perilaku pelanggan, serta distribusi genre musik. Dashboard ini menampilkan total revenue sebesar $2.328,60, daftar track dengan pendapatan tertinggi, tren penjualan aktual dan forecast, serta pertumbuhan penjualan tahunan (Year Sales Growth). Selain itu, dashboard juga menampilkan Average Revenue per Customer (ARPC) dan segmentasi pelanggan dalam bentuk cluster, serta kontribusi revenue berdasarkan genre musik, di mana genre Rock menjadi kontributor terbesar. Dashboard ini dirancang untuk membantu analisis performa bisnis, pemantauan KPI, dan pengambilan keputusan strategis.



Gambar 8.1 1 Tampilan *dashboard*

VIII.2 Analisis Ketercapaian KPI

Chart	Tercapai	Tidak Tercapai	Insight
	V		Total Revenue Mencapai Target Lebih dari $500.00
		V	KPI belum tercapai, karena kontribusi revenue 10 track teratas hanya 1,54%, masih di bawah target ≥30% dari total revenue.
 		V	KPI belum tercapai, karena pertumbuhan Year-on-Year (YoY) tidak mencapai target ≥10%.
		V	KPI tidak tercapai, Karena Average Revenue per Customer (ARPC) belum semua menyentuh 40

	V		KPI tercapai, karena Top Genre (Rock) berkontribusi sekitar 35,5% terhadap total revenue, melebihi target ≥35%.

Tabel 8.2 1 Tabel Analisis hasil KPI

BAB IX
PEMBAGIAN TUGAS ANGGOTA TIM

NIM	Nama	Pembagian Tugas
102022300368	Muhammad Rakha Alfaruq	Bab 1 Pendahuluan, Data Mining, Struktur laporan
102022300344	Devota Edra Athaloma	Implementasi Star Schema dan ETL
102022300313	Yudistira Sebastian Saftari	Bab 2 KPI, Implementasi Star Schema dan ETL
102022300284	Rajasyah Birra Munandar	Bab 3 Analisis Sumber Data, Bab 8 Dashboard

LAMPIRAN

Link Akses Dashboard:

https://lookerstudio.google.com/reporting/4123ca0f-b964-4568-9593-9d7634a63a9d

Link Dataset Kaggle:

https://www.kaggle.com/datasets/ranasabrii/chinook

Link Github:

https://github.com/rakhafaruq/Kelompok-2_Data-Warehouse-Business-Intellegence_Chinook-Music-Store