ANALISIS PENYEWAAN DAN PERFORMA FILM BERDASARKAN GENRE, AKTOR, DAN TREN PENYEWAAN



Mentor:

Kurnia Andre Febrian

Kelompok 5:

1.	Tiara Andina Dalimunthe	NPM.222406057
2.	Marcellio Aurel Christian	NPM.22082010019
3.	Dwi Saumantri Oktarini	NIM.0110122174
4.	Yudi Waliyuddin	NIM.0110122167

CELERATES PT. MITRA TALENTA GROUP 2024

KATA PENGANTAR

Puji serta syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga kami dapat menyelesaikan laporan akhir ini yang merupakan Final Project Program Kampus Merdeka di bidang Data Analytics & Business Intelligence di PT Mitra Talenta Grup. Laporan ini disusun sebagai bentuk dokumentasi dari proses dan hasil yang telah kami capai selama mengikuti program ini.

Proyek ini bertujuan untuk mengasah keterampilan kami dalam mengolah data mentah menjadi informasi yang berguna, Membangun infrastruktur data yang lebih efisien, serta menyajikan hasil analisis melalui visualisasi interaktif yang mendukung dalam pengambilan keputusan bisnis. Proyek ini dapat memberikan pengalaman praktis dalam mengelola data menggunakan berbagai aplikasi seperti PostgreSQL, Pentaho, dan Tableau.

Kami mengucapkan terima kasih kepada PT Mitra Talenta Grup (Celerates) atas kesempatan yang diberikan kepada kami untuk berpartisipasi dalam program ini. Ucapan terima kasih juga kami sampaikan kepada trainer dan mentor atas bimbingan dan dukungannya selama pengerjaan proyek ini. Tak lupa juga, Kami berterima kasih kepada rekan-rekan kelompok yang telah bekerja sama dengan baik.

Kami menyadari proyek ini masih jauh dari sempurna. Oleh karena itu, kami mengharapkan kritik dan saran pada proyek ini dalam membangun perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan menjadi kontribusi positif dalam pengembangan kemampuan kami di bidang Data Analytics & Business Intelligence.

DAFTAR ISI

KATA PENGANTAR	1
DAFTAR ISI	2
DAFTAR GAMBAR	3
ABSTRAK	4
BAB 1 PENDAHULUAN	5
1.1 Latar Belakang	5
1.2 Tujuan Project	5
1.3 Manfaat Project	5
1.4 Lingkup Project	6
BAB 2 METODOLOGI PROJECT	7
2.1 Deskripsi Database	7
2.2 Tahapan Project	8
2.2.1 Input Database ke RDBMS (PostgreSQL)	8
2.2.2 Pembuatan Staging Area (menggunakan Pentaho)	9
2.2.3 Desain Data Warehouse (termasuk pembuatan tabel fakta dan dimensi)	10
2.2.4 Pembuatan Data Mart (topik analisis spesifik)	10
2.2.5 Pembuatan Dashboard (menggunakan Tableau atau Power BI)	11
2.3 Tools dan Teknologi.	12
BAB 3 IMPLEMENTASI DAN HASIL	13
3.1 Input Database	13
3.2 Pembuatan Staging Area.	13
3.3 Desain Data Warehouse	17
3.4 Pembuatan Datamart	24
3.5 Visualisasi dan Analisis Dashboard	26
BAB 4 PENUTUP	31
4.1 Kesimpulan	31
4.2 Saran	31
DAFTAR PHSTAKA	32

DAFTAR GAMBAR

Gambar 2.1 ERD Diagram DVD Rental	8
Gambar 2.2 Mockup Dashboard executive summary	11
Gambar 2.3 Mockup dashboard Aktor.	12
Gambar 3.1 Tabel Restore DVD Rental	13
Gambar 3.2 Perubahan Tipe Data.	14
Gambar 3.3 Table Address	14
Gambar 3.4 tf read data.	15
Gambar 3.5 tf log table masuk	15
Gambar 3.6 job execute data	15
Gambar 3.7 tf list table	16
Gambar 3.8 main job	16
Gambar 3.9 dim actor.	17
Gambar 3.10 dim category	18
Gambar 3.11 dim customer.	18
Gambar 3.12 dim date	19
Gambar 3.13 dim film	20
Gambar 3.14 dim staff	21
Gambar 3.15 dim store	22
Gambar 3.16 fact rental	23
Gambar 3.17 Data Mart.	25
Gambar 3.18 Select Values di Data Mart	25
Gambar 3.19 Calculated Field Total Penyewaan.	26
Gambar 3.20 Calculated Field Rank Film.	26
Gambar 3.21 Calculated Field Penyewaan Rata-Rata	27
Gambar 3.22 Calculated Field Zero.	27
Gambar 3.23 Dashboard 1	29
Gambar 3 24 Dashboard 2	29

ABSTRAK

Dalam era digital saat ini, Pengambilan keputusan yang berbasis data sudah menjadi kebutuhan utama bagi perusahaan dalam peningkatan efisiensi operasional dan daya saing. Final project ini dirancang untuk memberikan pengalaman praktis dalam mengimplementasikan *Data Analytics* dan *Business Intelligence* melalui tahapan pengolahan data yang mencakup pembuatan staging area, data warehouse, data mart, dan dashboard interaktif.

Tujuan dari proyek ini adalah untuk pengolahan data mentah menjadi informasi yang bernilai, serta menyajikan hasil analisis secara visual sehingga dapat mendukung dalam pengambilan keputusan strategis. Database yang digunakan dalam proyek ini dipilih dari sumber yang relevan dan diolah menggunakan aplikasi seperti PostgreSQL, Pentaho, dan Tableau.

Metodologi yang diterapkan ini mencakup proses *Extract, Transform, Load (ETL)*, Perancangan struktur data warehouse dengan tabel dimensi dan tabel fakta, Pembuatan data mart untuk analisis spesifik, hingga visualisasi data menggunakan dashboard interaktif. Setiap tahap dikerjakan dengan memperhatikan prinsip pengolahan data yang sistematis dan terstruktur.

Hasil proyek ini meliputi desain data warehouse yang optimal, data mart dengan analisis yang fokus, dan dashboard yang efektif dalam menyampaikan *insights*. Analisis yang dihasilkan dapat memberikan rekomendasi strategis yang digunakan untuk meningkatkan performa bisnis.

Kesimpulan dari proyek ini menunjukkan pentingnya infrastruktur data yang baik dalam mendukung analisis bisnis yang akurat dan relevan. Proyek ini juga dapat memberikan wawasan dan keterampilan praktis bagi peserta dalam bidang analisis data dan business intelligence, yang dapat diterapkan dalam lingkungan bisnis nyata.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam industri hiburan, khususnya penyewaan film dapat memahami tentang preferensi pelanggan dan performa produk yang sangat penting dalam peningkatan pendapatan dan kepuasan pelanggan. Genre film, popularitas aktor, dan tren penyewaan merupakan faktor utama yang mempengaruhi keputusan pelanggan dalam menyewa film. Namun, banyak perusahaan menghadapi tantangan dalam menganalisis data ini secara efektif karena kurangnya infrastruktur data yang memadai dan alat analisis yang mendukung pengambilan keputusan strategis.

Proyek ini berfokus pada analisis penyewaan dan performa film berdasarkan genre, aktor, dan tren penyewaan. Maka dari itu digunakan **DVD Rental Database**, sebuah database yang mencakup data transaksi penyewaan, informasi pelanggan, dan detail produk (film). Database ini relevan karena menyediakan informasi yang lengkap untuk mengidentifikasi pola penyewaan dan mengukur performa film secara kuantitatif.

1.2 Tujuan Project

Berikut adalah tujuan Proyek ini dilakukan.

- 1. Menganalisis data penyewaan film berdasarkan genre, aktor, dan tren penyewaan untuk mengidentifikasi preferensi pelanggan.
- 2. Membuat infrastruktur data yang mencakup staging area, data warehouse, dan data mart yang terstruktur dengan baik.
- 3. Menyediakan dashboard interaktif yang menyajikan *insights* penting dalam pengambilan keputusan yang strategis.

1.3 Manfaat Project

Proyek ini memberikan manfaat sebagai berikut.

- 1. Manfaat Teknis
 - a. Peningkatan keterampilan dalam pengelolaan data yang menggunakan aplikasi PostgreSQL, Pentaho, dan Tableau.

b. Pemahaman mengenai proses *Extract, Transform, Load (ETL)* dan Desain Data Warehouse.

2. Manfaat Bisnis.

- a. Dapat memberikan wawasan kepada perusahaan penyewaan film mengenai pola penyewaan dan preferensi pelanggan.
- b. Dapat membantu dalam pengambilan keputusan mengenai strategi pemasaran dan pemilihan koleksi film berdasarkan waktu atau periode tertentu.
- c. Dapat memilih aktor yang tepat berdasarkan genre yang dapat dijadikan dalam strategi promosi.

1.4 Lingkup Project

Proyek berfokus pada batasan pekerjaan yaitu:

- 1. Pembuatan staging area untuk menyimpan data mentah dari file DVD Rental Database.
- 2. Desain dan implementasi data warehouse dengan tabel dimensi dan fakta.
- 3. Pembuatan data mart untuk topik analisis spesifik, yaitu genre, aktor, dan tren penyewaan.
- 4. Pengembangan dashboard interaktif menggunakan Tableau untuk mengetahui hasil analisis kepada pemangku kepentingan.

BAB 2

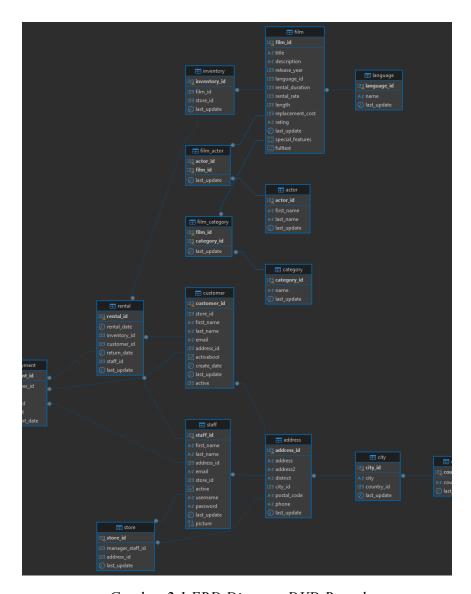
METODOLOGI PROJECT

2.1 Deskripsi Database

DVD Rental Database merupakan database proyek yang kami gunakan, database ini berisi terkait data penyewaan penyewaan film oleh pelanggan. Struktur database ini meliputi tabel utama yaitu tabel actor yang menyimpan data - data para aktor dalam film, tabel film berisi tentang detail film seperti judul, deskripsi, dan tahun rilis, tabel film_actor merupakan tabel yang terhubung dengan tabel actor dengan film melalui relasi many-to-many, tabel category berisi tentang informasi kategori film yang dihubungkan dengan tabel film melalui tabel film_category, tabel inventory berisi tentang stok DVD yang tersedia untuk disewa, dan setiap tabel dalam inventaris terkait dengan tabel rental yang menyimpan transaksi penyewaan oleh customer.

Customer yang tercatat dalam tabel customer dapat melakukan transaksi penyewaan yang tercatat dalam tabel rental. Tabel payment berisi tentang pembayaran yang dilakukan oleh pelanggan untuk setiap transaksi penyewaan, tabel staff berisi data staff yang mengelola toko penyewaan, dan tabel store berisi tentang informasi toko tempat DVD disewakan. Tabel address berisi tentang informasi alamat yang terkait dengan customer, staff, dan toko yang terhubung dengan tabel city dan country untuk mencatat lokasi geografis secara lebih spesifik. Tabel languange berisi tentang informasi bahasa pada film.

Hubungan antar tabel relasi one-to-many ada pada tabel languange dengan film dimana setiap film hanya memiliki satu bahasa, dan antara tabel store dengan inventory yang mencatat setiap toko memiliki stok DVD tertentu. Hubungan antar tabel relasi many-to-many ada pada tabel film_actor dan film_category yang menghubungkan film dengan aktor serta kategori film yang relevan.



Gambar 2.1 ERD Diagram DVD Rental

2.2 Tahapan Project

Project ini memiliki 5 tahapan yang dimulai dari Input Database ke PostgreSQL, Pembuatan Staging Area menggunakan Pentaho, Melakukan Desain Data Warehouse yang Termasuk dalam Pembuatan Tabel Dimensi dan Fakta, Melakukan Pembuatan Datamart sesuai dengan topik analisis, Pembuatan dashboard menggunakan Tableau. Berikut ini uraian dari tiap-tiap tahapan:

2.2.1 Input Database ke RDBMS (PostgreSQL)

- 1. Unduh file .tar dari DVD Rental Database.
- 2. Menggunakan PostgreSQL (pgAdmin) untuk melakukan restore database.

3. Verifikasi data di dalam database menggunakan DBeaver untuk memastikan semua tabel berhasil dimuat.

2.2.2 Pembuatan Staging Area (menggunakan Pentaho).

Staging Area adalah tempat penyimpanan sementara untuk data yang telah diekstraksi dari berbagai sumber data mentah (dengan struktur dan format berbeda). Di area ini, data sering kali dibersihkan, distandarisasi, diperkaya, dan disusun untuk memfasilitasi pemrosesan lebih lanjut.

1. Membuat Kerangka Staging

Melakukan pembuatan kerangka tabel dari database dvd rental dengan cara menyalin DDL dari database dvd rental untuk digunakan database staging.

2. Melakukan Transformasi pada Pentaho

Transformasi dilakukan untuk mengisi data dari kerangka stanging yang telah kita buat, ada pun caranya yaitu terdapat 3 Transformasi dan 2 Job untuk membaca, mencatat, dan menyalin data OLTP.

- a. tf_read_data, file ini berfungsi untuk membaca data dari OLTP yang akan dilakukan staging dan mencatat log dari data OLTP. File ini memiliki 2 Tahapan yang pertama membaca dan mengambil isi data dari OLTP lalu disalin kedalam database staging, yang kedua melakukan pencatatan log dari data OLTP yang akan disimpan kedalam schema log.
- b. tf_log_tabel_masuk, file ini berfungsi untuk melakukan pencatatan log tabel yang disalin dari data OLTP menuju staging sehingga menghasilkan data input date di dalam tabel log_tabel_masuk.
- c. Tf_list_tabel, file ini berfungsi untuk melakukan proses penyalinan dari data source_records dan staging_dvdrental berdasarkan nama tabel, serta mengatur sistem agar ketika terjadi kesalahan proses sistem tiak perlu melakukan penyalinan ulang tetapi hanya menyalin data yang kurang.
- d. main_job, file ini merupakan, file yang paling utama untuk menjalankan proses job_execute_data dan tf_list_tabel.
- e. Job_execute_data, file ini berfungsi untuk mengatur transformasi mulai dari tf_read_data dan tf_log_tabel_masuk, pada job ini juga terdapat beberapa step yang dimana untuk melakukan update pada tabel staging_log, dan menghitung jumlah baris dari variabel table_name, dan melakukan perbandingan antara source_targets dengan target_records.

2.2.3 Desain Data Warehouse (termasuk pembuatan tabel fakta dan dimensi).

Data warehouse sendiri merupakan suatu sistem yang dirancang untuk mengarsipkan hingga menganalisis riwayat data untuk bisa menunjang kepentingan informasi yang dimiliki oleh sebuah perusahaan ataupun bisnis tertentu sehingga keberadaannya pun sangat penting. Terdapat juga tabel fakta berisi kunci utama dari tabel dimensi yang dirujuk beserta beberapa metrik kuantitatif. Contoh tabel fakta meliputi pesanan pelanggan atau data keuangan deret waktu. Sedangkan tabel dimensi menyimpan informasi deskriptif untuk bidang terkait yang ada dalam rekaman tabel fakta. Biasanya tabel ini mewakili entitas fisik seperti "pelanggan" atau "produk".

Data warehouse yang kami rancang ini menggunakan model star schema, karena memiliki struktur yang sederhana serta pemahaman menggunakan star schema tidaklah sulit dan lebih mudah dibanding yang lainnya. Tabel fakta dan dimensi kami terdiri dari sebagai berikut:

- 1. Tabel Fakta: Total Penyewaan, yang berisi informasi transaksi seperti tanggal sewa, jumlah transaksi, dan total pendapatan.
- 2. Tabel Dimensi:
 - Dim_Film: Informasi genre, judul, dan rating (SCD Type 2).
 - Dim_Aktor: Daftar aktor yang berperan dalam film (SCD Type 2).
 - Dim_Waktu: Kalender yang mencakup data tanggal, bulan, dan tahun (SCD Type 1).
 - Dim Customer: Informasi demografis customer (SCD Type 2).
 - Dim Staff: Informasi staf terkait transaksi (SCD Type 2).
 - Dim_Category: Informasi kategori yang tersedia dalam film (SCD Type 1).
 - Dim Store: Informasi detail toko penyewaan film (SCD Type 2).

2.2.4 Pembuatan Data Mart (topik analisis spesifik)

Data mart adalah bentuk gudang data sederhana yang berfokus pada satu subjek atau lini bisnis. Dengan data mart, tim dapat mengakses data dan memperoleh wawasan lebih cepat, karena mereka tidak perlu menghabiskan waktu mencari dalam gudang data yang lebih kompleks atau menggabungkan data secara manual dari berbagai sumber. Data mart ini dibuat untuk analisis spesifik, Seperti:

- 1. Identifikasi genre film Terpopuler
- 2. Evaluasi aktor yang berkontribusi pada setiap genre

3. Analisis tren penyewaan berdasarkan waktu (bulan).

2.2.5 Pembuatan Dashboard (menggunakan Tableau atau Power BI).

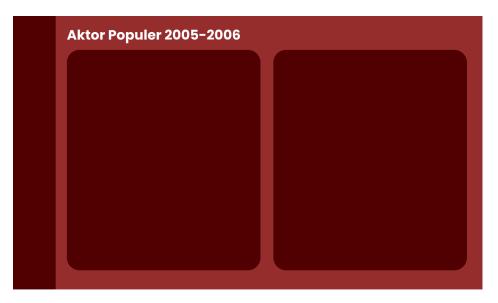
Visualisasi yang digunakan adalah Tableau untuk membuat dashboard interaktif. Berikut adalah langkah-langkah pembuatan:

- 1. Dapat menyambungkan Tableau ke data mart di PostgreSQL.
- 2. Membuat visualisasi genre favourite, fim terbaik, film terburuk dengan grafik batang dan tren penyewaan dengan grafik garis.
- 3. Dashboard yang diatur untuk menampilkan insights terkait tren dan waktu penyewaan, genre, serta aktor.



Gambar 2.2 Mockup Dashboard executive summary

Berikut adalah mockup dari dashboard executive summary. Mockup ini merupakan desain untuk dashboard yang akan menyajikan visualisasi data utama dari database DVD rental. Visualisasi yang ditampilkan sebagai berikut:



Gambar 2.3 Mockup dashboard Aktor

2.3 Tools dan Teknologi

- PostgreSQL: Digunakan untuk menyimpan database dan analisis data. Alasannya, Karena PostgreSQL memiliki kemampuan manajemen data, Mendukung untuk proses analisis lanjutan, serta menyediakan fitur fungsi yang sangat membantu dalam analisis.
- 2. Pentaho: Digunakan untuk proses ETL seperti ekstraksi, transformasi, dan Job. Alasannya, Memudahkan pengelolaan proses ETL tanpa memerlukan banyak kode, hanya dengan drag-and-drop.
- 3. Tableau: Digunakan untuk membuat visualisasi data interaktif. Alasannya, kemampuan menghasilkan dashboard yang intuitif dan mudah diakses oleh pengguna non-teknis.

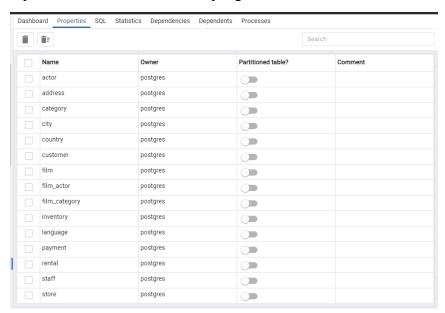
BAB 3

IMPLEMENTASI DAN HASIL

3.1 Input Database

Proses pengimporan database DVD Rental ke dalam PostgreSQL dilakukan melalui:

- Mempersiapkan Database DVD Rental dengan bentuk data .tar, lalu melakukan pengimporan menggunakan pgAdmin4 sebagai alat administrasi untuk PostgreSQL. Pada proses ini tidak membutuhkan banyak konfigurasi hanya perlu mengubah *custom files* menjadi *all files* lalu bisa memilih file yang akan di restore atau di impor.
- 2. Restore Database DVD Rental, setelah berhasil maka database dvdrental akan terisi dengan 15 tabel dengan data yang berbeda tiap tabelnya. Berikut ini tampilan dari database dvdrental yang telah terisi.



Gambar 3.1 Tabel Restore DVD Rental

Kendala Teknis: Pernah mengalami error pada saat restore database, Solusi yang dilakukan adalah melakukan pengecekan input file pada lokasi yang benar dan menyesuaikan versi PostgreSQL dan memerlukan waktu pemrosesan yang cukup lama saat restore file SQL database.

3.2 Pembuatan Staging Area

Pada tahap ini kita akan melakukan pembuatan kerangka untuk tabel staging terlebih dahulu, terdapat beberapa konfigurasi yang akan kita lakukan yang pertama

kita perlu mengubah tipe data dibagian release_year yang sebelumnya public."year" NULL menjadi int NULL, lalu dibagian rating yang sebelumnya public.mpaa_rating DEFAULT 'G' ::mpaa_rating NULL menjadi varchar DEFAULT 'G' NULL, seperti gambar dibawah ini.

```
● CREATE TABLE public.film (
film_id serial4 NOT NULL,
title varchar(255) NOT NULL,
description text NULL,
release year int NULL,
language_id int2 NOT NULL,
rental_duration int2 DEFAULT 3 NOT NULL,
rental_rate numeric(4, 2) DEFAULT 4.99 NOT NULL,
length int2 NULL,
replacement_cost numeric(5, 2) DEFAULT 19.99 NOT NULL,
rating varchar DEFAULT 'G' NULL,
last_update timestamp DEFAULT now() NOT NULL,
special_features_text NULL,
fulltext tsvector NOT NULL,
CONSTRAINT film_pkey PRIMARY KEY (film_id),
CONSTRAINT film_language_id_fkey FOREIGN KEY (language_id) REFERENCES public."language"(language_id) ON DELETE RESTRICT ON UPDATE CASCADE
);
```

Gambar 3.2 Perubahan Tipe Data

Setelah menjalankan query di atas selanjutnya kita membuat script baru yang berisi dibawah ini, lalu sebelum menjalankan query dibawah ini ada beberapa konfigurasi lagi yang akan kita lakukan yakni menghapus film_fulltext_idx di dalam folder indexes pada tabel film. Setelah menghapus data tersebut, query dibawah ini dapat kita jalankan tanpa error.

ALTER TABLE film

ALTER COLUMN special_features TYPE varchar

USING array_to_string(special_features, ',');

ALTER TABLE film

ALTER COLUMN fulltext TYPE varchar(255)

USING fulltext::text;

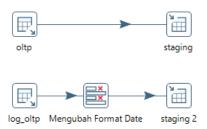
Setelah menjalankan query tersebut kita akan menonaktifkan Not Null pada tabel address dengan cara mengklik tanda v pada kolom Not Null, kolom yang akan kita nonaktifkan adalah district dan postal code, seperti gambar dibawah ini.

Column Name	# Data type	Identity	Collation	Not Null	Default	Comment
123 address_id	1 serial4			[v]	nextval('staging.address_address_id_seq'::regclass)	
A-Z address	2 varchar(50)		<u>default</u>	[v]		
A-Z address2	3 varchar(50)		<u>default</u>			
A-Z district	4 varchar(20)		<u>default</u>			
123 city_id	5 int2			[v]		
A-Z postal_code	6 varchar(10)		<u>default</u>			
A-Z phone	7 varchar(20)		<u>default</u>			
last_update	8 timestamp			[v]		

Gambar 3.3 Table Address

Pada tahap transformasi akan dibagi menjadi 5 step yaitu:

a. tf read data



Gambar 3.4 tf read data

Pada step ini terdapat dua transformasi yang akan dilakukan, transformasi ini berfungsi untuk mengambil seluruh tabel dari database production dimana pengambilan tersebut akan menggunakan parameter table_name, dan transformasi log berfungsi untuk mencatat log data dari data masuk.

b. tf log table masuk



Gambar 3.5 tf log table masuk

Transformasi ini berfungsi untuk mencatat log tabel yang telah masuk kedalam database staging.

c. job execute data

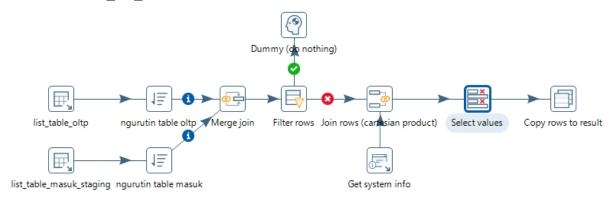


Gambar 3.6 job execute data

Job ini berfungsi untuk melakukan pemanggilan dari kedua transformasi diatas, lalu akan menambahkan dua query SQL, dimana target_records dan target_date berfungsi untuk memperbarui tabel staging_log pada database staging. Perintah tersebut akan melakukan pembaruan pada kolom target_records dan target_date dalam baris yang sesuai dengan variabel table_name. Pada kolom target_records nilai yang dimasukkan adalah hasil

dari jumlah perhitungan baris dengan *count(1)* dari tabel yang dinamai sesuai dengan nama variabel table_name dalam skema staging. Lalu pada kolom target_date akan di isi dengan waktu saat ini yang diambil menggunakan fungsi *now()*. Pada query SQL logical berfungsi untuk melakukan evaluasi apakah proses tersebut telah berhasil atau tidak.

d. tf list table



Gambar 3.7 tf list table

Transformasi ini berfungsi untuk melakukan perbandingan antara kedua database yaitu database production dan staging lalu menambahkan informasi tanggal saat proses staging dilakukan dan memilih field yang dibutuhkan saja.

e. main job



Gambar 3.8 main job

Job ini berfungsi sebagai tempat untuk menjalankan semua step yang telah dibuat diatas, sehingga urutan tersebut tertata dan tidak saling bertabrakan, jadi ketika saat dijalankan akan sesuai dengan job yang sudah diatur ini.

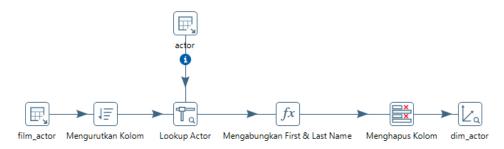
Pada saat dilakukan transformasi di pentaho ini ditemukan kendala saat proses pembuatan staging area dimana terdapat *foreign key constraint* sehingga memunculkan error. Masalah tersebut diselesaikan dengan cara menghapus seluruh

foreign key contraint yang berada di dalam folder Foreign Key dari tiap tablenya. Lalu terdapat kendala juga yang memunculkan pesan staging.actor.info does not exist. Masalah ini diselesaikan dengan cara saat melakukan pengambilan tabel base saja pada database production di transformasi list_tabel sehingga pesan erro tersebut tidak akan muncul. Setelah kendala tersebut terselesaikan maka jika pembuatan staging area ini berhasil seluruh tabel tabel akan terisi dengan data production yaitu database DVDRental.

3.3 Desain Data Warehouse

Desain data warehouse ini menggunakan model star schema yang terdiri dari tabel fakta dan tabel dimensi:

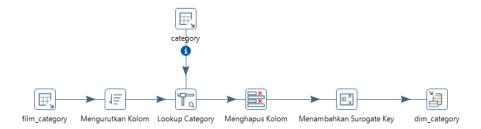
1. dim actor



Gambar 3.9 dim actor

Transfromasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel film actor lalu dilanjutkan dengan sort rows dengan mengurutkan kolom berdasarkan actor_id dari id 1, selanjutnya melakukan 'stream lookup' dengan tabel actor berdasarkan actor_id untuk menambahkan kolom first_name dan last_name. Setelah itu melakukan pengabungan kolom menjadi full_name dari kolom first_name dan last_name menggunakan 'formula', lalu dilanjut dengan menghapus kolom yang tidak digunakan seperti first_name, last_name & last_update. Terakhir melakukan 'dimension lookup/update' untuk memasukkan data hasil transformasi ke tabel dim_actor, dimensi ini menerapkan SCD Type 2, jadi jika ada perubahan data maka dibuat lah baris baru dengan kolom date_from dan date_to serta sk_actor akan dihasilkan untuk setiap baris data dimensi.

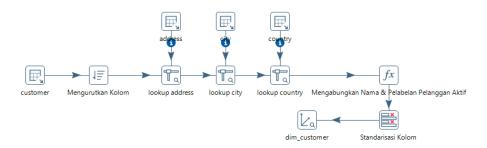
2. dim category



Gambar 3.10 dim category

Transfromasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel film category lalu dilanjutkan dengan sort rows dengan mengurutkan kolom berdasarkan category_id dari id 1, selanjutnya melakukan 'stream lookup' dengan tabel category berdasarkan category_id untuk menambahkan kolom name. Setelah itu menghapus kolom yang tidak digunakan last_update dan dilakukan standarisasi penamaan pada setiap kolom diawali dengan nama customer, lalu 'add sequence' untuk menambahkan sk_category pada tiap baris data category. Terakhir melakukan 'insert/update' untuk mengupdate data lama tanpa menambah baris baru pada data, maka dari itu data lama akan langsung diganti dengan data baru . Dimensi ini menerapkan SCD Type 1.

3. dim customer

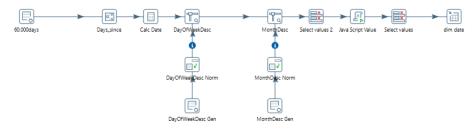


Gambar 3.11 dim customer

Transformasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel customer lalu dilanjutkan dengan sort rows dengan mengurutkan kolom berdasarkan customer_id dari id 1, selanjutnya melakukan 'stream lookup' dengan tabel address berdasarkan address_id untuk menambahkan kolom address, district, city_id, postal_code, phone. Kemudian melakukan 'stream lookup' dengan tabel city berdasarkan city_id untuk menambahkan kolom city, country id, lalu melakukan 'stream lookup'

lagi dengan tabel country berdasarkan country_id untuk menambahkan kolom country. Selanjutnya melakukan penggabungan nama dan pelabelan pelanggan aktif menggunakan 'formula' yang dimana penggabungan tersebut akan menggabungkan first_name dan last_name lalu untuk pelabelan pelanggan aktif akan memunculkan kolom customer_active yang diman bila kolom activebool bernilai "Y" Maka akan memunculkan label Aktif jika tidak maka sebaliknya. Lalu dilanjut dengan menstandarisasi kolom menggunakan 'select values' untuk menghapus kolom yang tidak digunakan seperti first_name, last_name dan last_update, lalu melakukan standarisasi penamaan kolom diawali dengan penamaan depan customer. Terakhir melakukan 'dimension lookup/update' untuk memasukkan data hasil transformasi ke tabel dim_customer, dimensi ini menerapkan SCD Type 2, jadi jika ada perubahan data maka dibuat lah baris baru dengan kolom date_from dan date_to serta sk customer akan dihasilkan untuk setiap baris data dimensi.

4. dim date

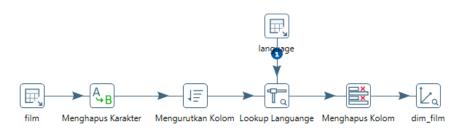


Gambar 3.12 dim date

Transformasi diatas dimulai dari 'generate rows' yang membuat tanggal sebanyak 150+ tahun ke depan, dilanjut dengan 'add sequence' yang digunakan untuk menghitung tanggal yang dihasilkan (hari sejak epoch time). Selanjutnya 'calculation' digunakan untuk menghitung tanggal spesifik berdasarkan jumlah hari yang diberikan sebelumnya. Lalu 'generate rows' yang digunakan untuk memasukkan deskripsi tambahan, selanjutnya 'rows normalizer' yang digunakan untuk menyimpan deskripsi hari normal. Lalu 'stream lookup' digunakan untuk mengambil data daydesc, daydescshort, weekendlnd, daynr, dilanjutkan 'generate rows' dengan memasukkan deskripsi bulan tambahan. Lalu 'rows normaliser' untuk menyimpan nama bulan, selanjutnya terdapat 'stream lookup' kepada monthdesc untuk mendapatkan

deskripsi bulan berdasarkan tanggal dengan lookup ke monthnr. Terdapat 'select values 2' untuk memilih kolom - kolom yang diperlukan saja, lalu step 'javascript value' digunakan untuk melakukan kalkulasi tambahan dan menghasilkan atribut waktu yang lebih rinci. Selanjutnya 'select values' digunakan untuk memilih kolom-kolom yang diperlukan serta melakukan standarisasi penamaan kolom. Lalu yang terakhir adalah 'table output' digunakan untuk menyimpan data hasil transformasi ke tabel dim date.

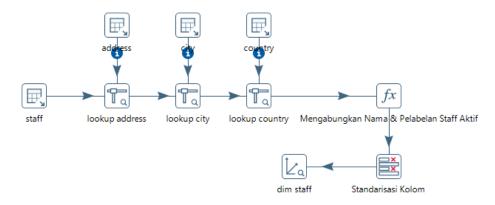
5. dim film



Gambar 3.13 dim film

Transfromasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel film lalu melakkan penghapusan karakter pada kolom special_features yang berisi ({}",), dilanjutkan dengan sort rows dengan mengurutkan kolom berdasarkan actor_id dari id 1, selanjutnya melakukan 'stream lookup' dengan tabel languange berdasarkan language_id untuk menambahkan kolom name. Setelah itu melakukan 'select values' penghapusan kolom yang tidak digunakan seperti last_update dan melakukan standarisasi penamaan kolom yang diawali dengan kata film. Terakhir melakukan 'dimension lookup/update' untuk memasukkan data hasil transformasi ke tabel dim_film, dimensi ini menerapkan SCD Type 2, jadi jika ada perubahan data maka dibuat lah baris baru dengan kolom date_from dan date to serta sk film akan dihasilkan untuk setiap baris data dimensi.

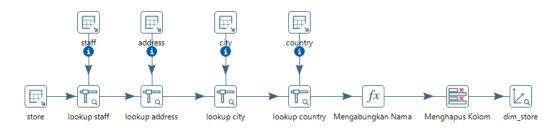
6. dim staff



Gambar 3.14 dim staff

Transformasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel staff lalu dilanjutkan dengan melakukan 'stream lookup' dengan tabel address berdasarkan address id untuk menambahkan kolom address, district, city id, postal code, phone. Kemudian melakukan 'stream lookup' dengan tabel city berdasarkan city id untuk menambahkan kolom city, country id, lalu melakukan 'stream lookup' lagi dengan tabel country berdasarkan country id untuk menambahkan kolom country. Selanjutnya melakukan penggabungan nama dan pelabelan staff aktif menggunakan 'formula' yang dimana penggabungan tersebut menggabungkan first name dan last name lalu untuk pelabelan pelanggan aktif akan memunculkan kolom staff active yang diman bila kolom active bernilai "Y" Maka akan memunculkan label Aktif jika tidak maka sebaliknya. Lalu dilanjut dengan menstandarisasi kolom menggunakan 'select values' untuk menghapus kolom yang tidak digunakan seperti first name, last name dan last update, lalu melakukan standarisasi penamaan kolom diawali dengan penamaan depan staf. Terakhir melakukan 'dimension lookup/update' untuk memasukkan data hasil transformasi ke tabel dim staff, dimensi ini menerapkan SCD Type 2, jadi jika ada perubahan data maka dibuat lah baris baru dengan kolom date from dan date to serta sk staff akan dihasilkan untuk setiap baris data dimensi.

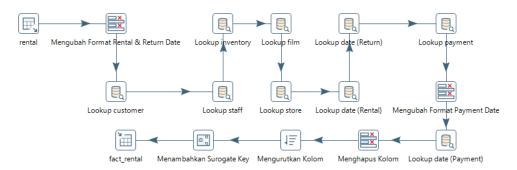
7. dim store



Gambar 3.15 dim store

Transformasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel store, setelah ini melakukan 'stream lookup' dengan tabel staff berdasarkan staff id untuk mengambil kolom first name, last name, & email, lalu dilanjutkan dengan melakukan 'stream lookup' dengan tabel address berdasarkan address id untuk menambahkan kolom address, district, city id, postal code, phone. Kemudian melakukan 'stream lookup' dengan tabel city berdasarkan city id untuk menambahkan kolom city, country id, lalu melakukan 'stream lookup' lagi dengan tabel country berdasarkan country id untuk menambahkan kolom country. Selanjutnya melakukan penggabungan nama menggunakan 'formula' yang dimana penggabungan tersebut akan menggabungkan first name dan last name. Lalu dilanjut dengan menstandarisasi kolom menggunakan 'select values' untuk menghapus kolom yang tidak digunakan seperti first name, last name dan last update, lalu melakukan standarisasi penamaan kolom diawali dengan penamaan depan store. Terakhir melakukan 'dimension lookup/update' untuk memasukkan data hasil transformasi ke tabel dim store, dimensi ini menerapkan SCD Type 2, jadi jika ada perubahan data maka dibuat lah baris baru dengan kolom date from dan date to serta sk store akan dihasilkan untuk setiap baris data dimensi.

8. fact rental



Gambar 3.16 fact rental

Transformasi diatas dimulai dari 'table input' yang mengambil data seluruh kolom yang ada pada tabel rental, lalu terdapat 'select values' yang mengubah format data rental dan return date menjadi string dengan format dd-MM-yy, lalu melakukan 'database lookup' dengan dim customer berdasarkan customer id untuk mengambil kolom sk customer. Selanjutnya melakukan 'database lookup' dengan dim staff berdasarkan staff id untuk mengambil kolom sk staff, lalu melakukan 'database lookup' dengan tabel staging inventory berdasarkan inventory id untuk mengambil kolom film id & store id. Step selanjutnya 'database lookup' dengan dim store berdasarkan store id untuk mengambil kolom sk store, lalu melakukan 'database lookup' dengan dim date berdasarkan deskripsi tanggal dengan rental date untuk mengambil kolom sk waktu. Step selanjutnya sama dengan step sebelumnya melakukan 'database lookup' dengan dim date berdasarkan yang deskripsi tanggal dengan return date untuk mengambil kolom sk waktu. Lalu dengan melakukan 'database lookup' staging payment berdasarkan customer id untuk mengambil kolom amoun & payment date, setelah itu melakukan pengubahan format pada kolom payment date menjadi string dengan format dd-MM-yy, lalu melakukan 'database lookup' dengan dim date berdasarkan deskripsi tanggal dengan payment_date untuk mengambil kolom sk waktu. Lalu melakukan 'select values' penghapusan kolom, jadi hanya mengambil kolom-kolom yang dibutuhkan saja, dilanjutkan dengan 'sort rows' untuk melakukan pengurutan kolom berdasarkan sk rental date mulai dari yang terkecil. Setelah itu melakukan 'add sequence' untuk menambahkan surogate key pada tiap data pada tabel fact rental. Lalu yang terakhir adalah

'table output' digunakan untuk menyimpan data hasil transformasi ke tabel fact rental.

3.4 Pembuatan Datamart

Berikut adalah langkah-langkah untuk membuat datamart:

1. Identifikasi Kebutuhan Bisnis

Pada tahap ini, diskusikan dengan tim untuk memahami kebutuhan analitik. Identifikasi fakta utama seperti transaksi pembayaran atau pengembalian film, serta dimensi terkait seperti pelanggan, film, staf, toko, dan waktu.

2. Desain Skema Data

Gunakan skema *Star Schema*. Mulai dengan membuat tabel fakta yang berisi metrik utama, misalnya jumlah pembayaran (*amount*). Selanjutnya, buat tabel dimensi yang mendeskripsikan fakta, seperti pelanggan, film, dan waktu.

3. Ekstraksi Data

Ambil data dari sumber operasional, seperti database transaksi, menggunakan query SQL untuk mengekstrak data yang relevan.

4. Transformasi Data

Pada tahap ini, data yang telah diekstraksi akan dibersihkan dan disesuaikan. Beberapa langkah yang dilakukan:

- Pembersihan Data: Menghapus duplikat dan mengisi nilai kosong.
- Penggabungan Tabel: Menghubungkan data menggunakan kunci primer dan kunci asing, seperti sk_customer untuk menghubungkan transaksi dengan pelanggan.
- Normalisasi atau Denormalisasi: Menyesuaikan data dengan desain skema yang telah dibuat.

5. Muat Data ke Data Warehouse

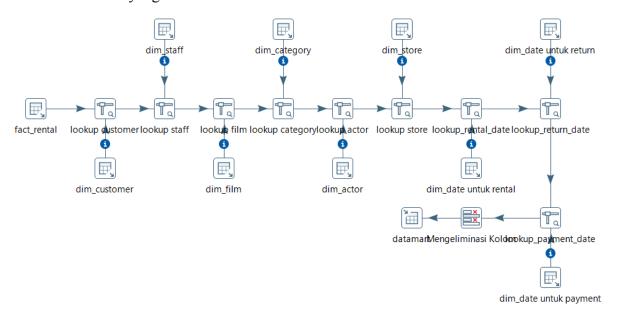
Gunakan alat ETL (*Extract, Transform, Load*) untuk memuat data yang telah diproses ke dalam data warehouse.

6. Pembuatan Datamart

Datamart sebagai subset data warehouse yang difokuskan pada kebutuhan spesifik, yaitu Analisis Penyewaan dan Performa Film Berdasarkan Genre, Aktor, dan Tren Penyewaan. Pada tahap ini, kombinasikan tabel fakta dan dimensi menggunakan query SQL untuk menghasilkan tampilan OLAP.

7. Pengujian dan Validasi

Lakukan validasi data untuk memastikan akurasi. Contohnya, periksa apakah total pembayaran sesuai dengan data sumber. Pastikan juga data mart memenuhi kebutuhan analitik yang telah ditentukan.



Gambar 3.17 Data Mart

#	Fieldname	Rename to
1	trx_id	
2	sk_customer	
3	cutomer_name	
4	customer_email	
5	customer_active	
6	customer_address	
7	customer_district	
8	customer_postal_code	
9	customer_phone	
1	customer_city	
1	customer_country	
1	sk_staff	
1	staff_name	
1	staff_email	
1	sk_film	
1	film_title	
1	film_description	
1	film_year	
1	film_languange_name	
2	film_rental_duration	
2	film_rental_rate	
2	film_length	
2	film_replacement_cost	
2	film_rating	
2	film_special_features	
2	sk_category	
2	category_name	
2	sk_actor	
2	actor_name	
3	sk_store	
3	store_address	
3	store district	

Gambar 3.18 Select Values di Data Mart

3.5 Visualisasi dan Analisis Dashboard

Dashboard dibuat menggunakan Tableau, dengan mengkoneksikan ke Data Mart dan menghubungkan Tableau ke PostgreSQL untuk mengambil data yang telah diolah untuk mendukung pengambilan keputusan dalam strategi bisnis.

1. Calculated Field

a. Total Penyewaan

Calculated field ini bertujuan untuk menghitung jumlah total transaksi penyewaan yang ada di dataset. Fungsi COUNT digunakan untuk menghitung banyaknya nilai dalam kolom [Trx Id], yang merepresentasikan setiap transaksi penyewaan.



Gambar 3.19 Calculated Field Total Penyewaan

b. Rank Film

Calculated field ini akan mengurutkan film-film dari yang tertinggi ke terendah, dan memberikan peringkat pada setiap film.



Gambar 3.20 Calculated Field Rank Film

c. Penyewaan rata-rata



Gambar 3.21 Calculated Field Penyewaan Rata-Rata

d. Zero



Gambar 3.22 Calculated Field Zero

2. Pembuatan Visualisasi:

Berikut adalah penjelasan tentang bagian-bagian dalam worksheet yang ditampilkan pada **Dashboard 1:**

- Total Penyewaan: Di bagian tengah atas, ditampilkan angka total penyewaan, yaitu 16.044. Ini adalah jumlah keseluruhan penyewaan film dalam periode yang dianalisis.
- 2. Genre: Diagram batang di bagian tengah kiri menunjukkan distribusi penyewaan berdasarkan genre film. Setiap batang mewakili jumlah penyewaan untuk genre tertentu.

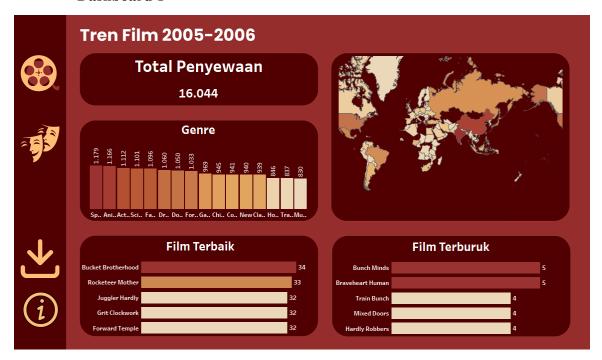
- 3. Grafik Map: Di kanan atas, terdapat peta dunia yang menunjukkan penyebaran tren penyewaan film secara geografis. Warna pada peta tampaknya menunjukkan intensitas penyewaan di berbagai wilayah (semakin gelap warnanya, semakin tinggi penyewaan).
- 4. Film Terbaik: Bagian kiri bawah mencantumkan film-film dengan performa terbaik, diukur berdasarkan jumlah penyewaan paling banyak: Bucket Brotherhood berada di posisi teratas dengan skor 34. Diikuti oleh Rocketeer Mother (33), Juggler Hardly (32), dan dua film lainnya.
- 5. Film Terburuk: Bagian kanan bawah mencantumkan film-film dengan performa terburuk, juga berdasarkan berdasarkan jumlah penyewaan paling sedikit: Film dengan skor terendah adalah Bunch Minds dan Braveheart Human, masing-masing mendapatkan skor 5. Film lainnya seperti Train Bunch, Mixed Doors, dan Hardly Robbers memiliki skor sedikit lebih tinggi (4).

Berikut adalah penjelasan tentang bagian-bagian dalam worksheet yang ditampilkan pada **Dashboard 2**:

- 1. Aktor Berdasarkan Genre (Aktor by Genre): Tabel heatmap di sisi kiri menunjukkan kontribusi masing-masing aktor dalam berbagai genre. Berikut detailnya:Angka di setiap sel menunjukkan jumlah penyewaan film yang melibatkan aktor tersebut dalam genre tertentu. Semakin gelap warna kotak, semakin tinggi jumlah penyewaan untuk kombinasi aktor dan genre tersebut.
- 2. Aktor Berdasarkan Total Penyewaan (Aktor by Total Penyewaan): Treemap di sisi kanan menunjukkan total penyewaan untuk setiap aktor tanpa memisahkan genre. Setiap kotak merepresentasikan seorang aktor, dengan ukuran kotak mencerminkan jumlah penyewaan total yang melibatkan aktor tersebut. Mary Keitel (memiliki kotak terbesar, menunjukkan ia adalah aktor dengan total penyewaan tertinggi). Aktor dengan kotak lebih kecil seperti Renee Ball memiliki penyewaan yang lebih rendah dibandingkan aktor dengan kotak lebih besar.
- 3. Filter Berdasarkan Bulan: Di bagian kanan atas terdapat filter "Berdasarkan Bulan" yang memungkinkan pengguna memilih data berdasarkan bulan tertentu. Dalam gambar ini, opsi yang dipilih adalah All (semua bulan), sehingga data mencakup seluruh periode 2005–2006.

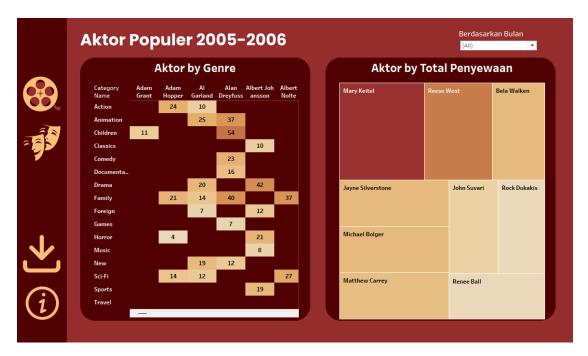
❖ Insight yang Dihasilkan:

Dashboard 1



Gambar 3.23 Dashboard 1

- Genre *Sports* dan *Animation* paling diminati pelanggan, sehingga perusahaan dapat menambah koleksi film dengan genre ini.
- Film Terbaik adalah Bucket Brotherhood dan Film Terburuk adalah Bunch Minds.



Gambar 3.24 Dashboard 2

- Aktor Mary Keitel memiliki pengaruh pada performa film yang dapat dimanfaatkan dalam strategi promosi.
- Tren Waktu: Penyewaan meningkat tajam di bulan Juli, sehingga promosi lebih intensif bisa dilakukan di periode tersebut.

* Kebutuhan pada Bisnis:

Dashboard interaktif ini dapat memberikan pemahaman kami tentang pola penyewaan film, Membantu pemilihan koleksi film berdasarkan genre, Melihat penyewaan aktor terbanyak dan optimalisasi penjualan di periode tertentu. Sehingga dapat menjadi kebutuhan bisnis dalam manajemen untuk menyusun strategi promosi.

BAB 4

PENUTUP

4.1 Kesimpulan

Proyek ini berhasil dalam menganalisis data penyewaan film berdasarkan genre, aktor dan tren penyewaan menggunakan database DVD Rental. Proses yang dilakukan dengan melakukan pembuatan staging area, Desain data warehouse, pembuatan data mart serta pengembangan interaktif dengan Tableau. Hasil Proyek:

- masii i toyek.
- Genre seperti *Sports* dan *Animation* memiliki tingkat penyewaan tertinggi, mengindikasikan preferensi pelanggan.
- Aktor tertentu berkontribusi signifikan terhadap popularitas film dan dapat memberikan wawasan dalam strategi pemasaran.
- Tren waktu menunjukkan banyaknya penyewaan pada akhir pekan dan periode liburan serta dapat membantu perusahaan dalam mengatur ketersediaan stok.

Proyek ini dapat memberikan manfaat teknis berupa peningkatan keterampilan kami dalam pengelolaan data serta dapat menganalisis menggunakan aplikasi seperti PostgreSQL, Pentaho dan Tableau. Selain itu proyek ini juga menghasilkan insight yang relevan dalam pengambilan keputusan bisnis yang strategis.

4.2 Saran

Proyek ini mungkin dapat kami tingkatkan dengan mengintegrasikan data tambahan, Seperti ulasan customer dan pendapatan per customer bertujuan agar mendapatkan analisis yang lebih mendalam. Kolaborasi dengan trainer dan mentor bagi kami juga sangat penting karena untuk memastikan metode yang digunakan dan hasil analisis yang telah sesuai dengan standar industri.

DAFTAR PUSTAKA

DATA SCIENCETEST.(2024). Staging-area. Etl.Procces. Retlieved From

https://datascientest.com/en/staging-area-what-does-this-stage-of-the-etl-process-involve

Repository.(2024). Undergraduate-Theses. Retlieved From

https://repository.its.ac.id/42834/1/5213100159-Undergraduate Theses.pdf

N-iX.(2023). Building-A-Data Warehouse. Retlieved From

https://www.n-ix.com/building-a-data-warehouse/

CAREERFOUNDRY.(2022). Data Analytics. Tableau-Visualizations. Retlieved From

https://careerfoundry.com/en/blog/data-analytics/tableau-visualizations/