D191 Advanced Data Management

VDM2 Task: Data Analysis

Alina Hendrix

4037.4.1 : **Writing Complex SQL Statements**
The graduate writes complex SQL statements in order to implement functions, stored procedures, and triggers to prepare data sets for data analysis and manipulation.

4037.4.2 : **Configuring Extraction, Transformation, and Loading (ETL) Workflows**
The graduate configures data extraction, transformation, and loading tasks to automate data integration.

My project reflects sales data for film rating categories at a movie rental store. My business question is: "What are total profits from films based upon their rating?"

The DVD business will find this report beneficial because it will allow the business analyze sales trends of films, stock films with ratings that are popular and that appeal to audiences in this geographic store location.

Part A

SUMMARIZE ONE REAL-WORLD WRITTEN BUSINESS REPORT THAT CAN BE CREATED FROM THE DVD DATASET FROM THE "LABS ON DEMAND ASSESSMENT ENVIRONMENT AND DVD DATABASE" ATTACHMENT.

1. Identify the specific fields that will be included in the detailed table and the summary table of the report.

2. Describe the types of data fields used for the report.

3. Identify *at least* **two** specific tables from the given dataset that will provide the data necessary for the detailed table section and the summary table section of the report.

Detailed Table

Variable Name	Database Table Location	Datatype	Definition
Rating	Film	Mpaa Rating	Rating of film for audience depending on content
Total Sales	Payment	Numeric	Sum of sales from each rating category
Average Price Per Sale	Payment	Numeric	Averaged price of each film from each rating category

Summary Table

Variable Name	Database Table Location	Datatype	Definition
Rating	Film	Mpaa Rating	Rating of film for audience depending on content
Total Sales	Payment	Numeric	Sum of sales from each rating category
Description		Varchar	Description of rating category as suitability for children and adult audiences

4. Identify *at least* **one** field in the detailed table section that will require a custom transformation with a user-defined function and explain why it should be transformed (e.g., you might translate a field with a value of *N* to *No* and *Y* to *Yes*).

The "Rating" field in my detailed table will require a custom transformation that will be outputted in the summary table. This field needs to be transformed to output a description based upon the rating. Ex 'PG' becomes 'Some Material May Not Be Suitable For Children'.

5. Explain the different business uses of the detailed table section and the summary table section of the report.

A stakeholder can utilize the data in the detailed table to see which rating categories have been the most profitable and the average for each film sold in its category.

The data in the summary table can be utilized to see total sales for rating categories and what audiences they are most suitable for. Stakeholders may adjust inventory to appeal to audiences in the geographical location according to sales and age of customers.

6. Explain how frequently your report should be refreshed to remain relevant to stakeholders.

This report should be refreshed quarterly to continuously appeal to customers in the area and to ensure inventory consists of the most profitable films. Refreshing at this rate allows enough sales to occur before analyzing trends.

Part B

PROVIDE ORIGINAL CODE FOR FUNCTION(S) IN TEXT FORMAT THAT PERFORM THE TRANSFORMATION(S) YOU IDENTIFIED IN PART A4.

```
create or replace function rating_description(rating mpaa_rating)
returns varchar
language plpgsql
as
$$
declare
description varchar;
        begin
        if rating = 'G' then
                description := 'All Ages Admitted';
        elseif rating = 'PG' then
                description := 'Some Material May Not Be Suitable For Children';
        elseif rating = 'PG-13' then
```

```
            description := 'May Be Inappropiate For Children Under 13';
        elseif rating = 'R' then
            description := 'Under 17 Requires Guardian Or Parent';
        elseif rating = 'NC-17' then
            description := 'Under 17 NOT Permitted';
        end if;

        return description;
end;
$$;
```

Part C

PROVIDE ORIGINAL SQL CODE IN A TEXT FORMAT THAT CREATES THE DETAILED AND SUMMARY TABLES TO HOLD YOUR REPORT TABLE SECTIONS.

```
create table detailed(
        rating mpaa_rating,
        total_sales numeric,
        average_price_per_sale numeric
);
drop table if exists detailed;
create table summary(
        rating mpaa_rating,
        total_sales numeric
);
```

Part D

PROVIDE AN ORIGINAL SQL QUERY IN A TEXT FORMAT THAT WILL EXTRACT THE RAW DATA NEEDED FOR THE DETAILED SECTION OF YOUR REPORT FROM THE SOURCE DATABASE.

```
insert into detailed
select
```

```
        rating,
        sum(amount),
        avg(rental_rate)
from payment, film
group by rating
order by sum(amount) desc;
select * from detailed;
```

Part E

PROVIDE ORIGINAL SQL CODE IN A TEXT FORMAT THAT CREATES A TRIGGER ON THE DETAILED TABLE OF THE REPORT THAT WILL CONTINUALLY UPDATE THE SUMMARY TABLE AS DATA IS ADDED TO THE DETAILED TABLE.

```
create or replace function update_summary()
returns trigger
language plpgsql
as
$$
begin
        delete from summary;
        insert into summary

        select rating, total_sales, rating_description(rating)
        from detailed;
        return new;
        end;
        $$;

create  trigger summary_changes
after update
on detailed
for each row
```

```
execute procedure update_summary();


        update detailed
        set total_sales = '12054677.54'
        where rating = 'PG';
select * from summary;
```

Part F

PROVIDE AN ORIGINAL STORED PROCEDURE IN A TEXT FORMAT THAT CAN BE USED TO REFRESH THE DATA IN BOTH THE DETAILED TABLE AND SUMMARY TABLE. THE PROCEDURE SHOULD CLEAR THE CONTENTS OF THE DETAILED TABLE AND SUMMARY TABLE AND PERFORM THE RAW DATA EXTRACTION FROM PART D.

```
create procedure refresh_data()
language plpgsql
as $$
begin
delete from detailed;
delete from summary;

insert into detailed
select
        rating,
        sum(amount),
        avg(rental_rate)
from payment, film
group by rating
order by sum(amount) desc;

insert into summary
select
        rating,
```

```
        total_sales,
        rating_description(rating)
from detailed;
return;
end;
$$;
call refresh_data();
select * from detailed;
```

Part G

PROVIDE A PANOPTO VIDEO RECORDING THAT INCLUDES THE PRESENTER AND A VOCALIZED DEMONSTRATION OF THE FUNCTIONALITY OF THE CODE USED FOR THE ANALYSIS.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=309e4be2-0aae-435f-bb53-b165014269dd

Part H

ACKNOWLEDGE ALL UTILIZED SOURCES, INCLUDING ANY SOURCES OF THIRD-PARTY CODE, USING IN-TEXT CITATIONS AND REFERENCES. IF NO SOURCES ARE USED, CLEARLY DECLARE THAT NO SOURCES WERE USED TO SUPPORT YOUR SUBMISSION.

PostgreSQL Tutorial. (2024b, April 21). *PostgreSQL tutorial*.

https://www.postgresqltutorial.com/