**DVD RENTAL SQL PROJECT**

This SQL project dataset consists of several tables that capture various aspects of a DVD rental system. The tables include film, category, film category, film actor, actor, language, customer, address, city, country, payment, rental, store, inventory, and staff. The dataset provides a comprehensive set of data for analysis and insights. The project involves performing several analysis tasks using SQL queries.

The SQL project dataset offers an opportunity to analyze customer rental behavior, identify high-spending customers, track actor movie counts, and explore films in specific genres. These queries provide valuable insights into the rental business, customer preferences, and the performance of actors and films in the dataset. Entity Relationship diagram is shown below:

A picture containing text, diagram, plan, parallel

Description automatically generated

**DATABASE USED FOR THIS ANALYSIS: POSTGRES 15**

The analysis has been performed using various basic queries involving comparison operators, logical operators, wildcard characters, clauses like WHERE, LIMIT, GROUP BY, ORDER BY and few complex queries using subquery, window (), CTE, CASE.

There is total 10 queries performed out of which two are listed and explained below: -

1.

A picture containing text, font, line, screenshot

Description automatically generated

The given SQL query retrieves the full name of the top three customers who have spent the most money in the payment table. Let's break down the query and understand its components:

The outer query starts with the SELECT statement, where `CONCAT(first\_name, ' ', last\_name) AS FULLNAME` is used to concatenate the first name and last name columns of the customer table, creating a new column named FULLNAME.

The FROM clause specifies the table name as "customer," indicating that the query will operate on the customer table.

The WHERE clause filters the rows based on the condition `customer\_id IN (...)`, which checks if the customer\_id is present in the subquery's result set.

The subquery within the parentheses is used to retrieve the top three customers with the highest total money spent. It begins with the SELECT statement, selecting the customer\_id and calculating the sum of the amount spent by each customer using the SUM() function. The result is aliased as "totalMoneySpentbyaCust."

The subquery also includes the GROUP BY clause to group the results by customer\_id and the ORDER BY clause to sort the results in descending order of the total money spent. The LIMIT clause ensures that only the top three rows are selected.

**RESULT**

A screenshot of a computer

Description automatically generated with medium confidence



This search result can also be downloaded in CSV format using download option (circled in red in above pic)

A picture containing text, line, screenshot, plot

Description automatically generated

2.

A picture containing text, font, screenshot

Description automatically generated

The given SQL query calculates the duration of rental for each customer and categorizes it based on a threshold of 7 days. Here's a summary of the query:

The query begins with a Common Table Expression (CTE) named "duration\_of\_Rental." It selects the rental\_id, customer\_id, and uses the `AGE()` function to calculate the number of days rented by subtracting the rental\_date from the return\_date.

The main query then selects the customer\_id and NO\_OF\_DAYS\_RENTED columns from the "duration\_of\_Rental" CTE.

The query includes a CASE statement to categorize the rental duration. If the NO\_OF\_DAYS\_RENTED is greater than or equal to 7 days, it returns the string 'SEND CUSTOMER A REMINDER MAIL TO RETURN DVD.' Otherwise, it returns 'NO NEED FOR REMINDER MAIL YET.' This CASE statement allows for conditional logic based on the duration of the rental.

The final result set includes the customer\_id, NO\_OF\_DAYS\_RENTED, and the categorization of whether a reminder mail should be sent to the customer or not.

In summary, the query calculates the rental duration for each customer and provides a categorization based on the duration threshold of 7 days. It determines whether a reminder mail should be sent to the customer based on the number of days rented and presents the results accordingly.

**RESULT**

A screenshot of a computer

Description automatically generated with low confidence

**Resources:**

[**https://www.postgresqltutorial.com/wp-content/uploads/2018/03/printable-postgresql-sample-database-diagram.pdf**](https://www.postgresqltutorial.com/wp-content/uploads/2018/03/printable-postgresql-sample-database-diagram.pdf)

*Dataset*

[**https://www.postgresqltutorial.com/postgresql-getting-started/postgresql-sample-database/**](https://www.postgresqltutorial.com/postgresql-getting-started/postgresql-sample-database/)

[**https://www.w3schools.com/sql/default.asp**](https://www.w3schools.com/sql/default.asp)