**SQL Project Documentation**

Introduction: -

* Sakila Database represents a DVD rental store.
* This data mainly categorized into 3 parts. That are Customer Data, Customer Related Data and Inventory Data. Views also available in this database.
* In this data base total 16 tables are available. The important tables for analysis are Customer, Film, Category, Actor, Inventory, Rental, Payment.
* Film, Customer and Staff Tables are highly connected with other tables directly.
* Using functions and operations for data manipulation, aggregation, filtering, and joining, subqueries enabling you to extract meaningful insights from the Sakila database and make data-driven decisions for the movie rental store's growth and improvement.

Interesting Questions in my analysis: -

* 'Music of Queen' and 'Kris Kristofferson' have seen an unlikely resurgence. As an unintended consequence, films starting with the letters 'K' and 'Q' have also soared in popularity. Display the movies starting with the letter’s 'K' and 'Q'
* The film 'Agent Truman' has been a great success. Display the names of all actors who appeared in this film
* Sales have been lagging among young families, so the management wants to promote family movies. Identify all the movies categorized as family films.
* Display the movies in descending order of their rental frequencies, so the management can maintain more copies of those movies
* Display the list of records for the movies that are suitable for audience below 13 years of age.
* In how many film categories, the difference between the average film replacement cost and the average film rental rate is greater than $15? Display the list of film categories identified above, along with the corresponding average film replacement cost and average film rental rate.

Task wise Techniques: -

|  |  |
| --- | --- |
| TASKS | TECHNIQUES |
| * Task 1: - Display the full names of the actors available in the database | * SELECT * CONCAT * AS * FROM |
| * Task 2i :- Display the number of times each first name appears in the database. * Task 2ii :- What is the count of actors that have unique first names in the database? Display the first names of all these actors. | * SELECT * COUNT * AS * FROM * ORDER BY * GROUP BY * DESC   ----------------------------------------------   * SELECT * COUNT * AS * FROM * GROUP BY * HAVING |
| * Task 3i: - Display the number of times each last name appears in the database. * Task 3ii: - Display all unique last names in the database. | * SELECT * COUNT * AS * FROM * ORDER BY * GROUP BY * DESC   ----------------------------------------------------   * SELECT * COUNT * AS * FROM * GROUP BY * HAVING |
| * Task 4i: - Display the list of records for the movies with the rating 'R'. * Task 4ii: - Display the list of Records for the movies that are not rated 'R' * Task 4iii :- Display the list of records for the movies that are suitable for audience below 13 years of age. | * SELECT * FROM * WHERE |
| * Task 5i :- Display the list of records for the movies where the replacement cost is up to $11. * Task 5ii :- Display the list of records for the movies where the replacement cost is between $11 and $20. * Task 5iii :- Display the list of records for the all movies in descending order of their replacement costs. | * SELECT * FROM * WHERE   ------------------------------------   * SELECT * FROM * WHERE * BETWEEN – AND   -----------------------------------------   * SELECT * FROM * ORDER BY * DESC |
| * Task 6 :- Display the names of the top 3 movies with the greatest number of actors. | * SELECT * COUNT * AS * FROM * JOIN * ON * GROUP BY * ORDER BY * DESC * LIMIT |
| * Task 7 :- 'Music of Queen' and 'Kris Kristofferson' have seen an unlikely resurgence. As an unintended consequence, films starting with the letters 'K' and 'Q' have also soared in popularity. Display the movies starting with the letter’s 'K' and 'Q' | * SELECT * AS * FROM * WHERE * LIKE * OR * ORDER BY |
| * Task 8 :- The film 'Agent Truman' has been a great success. Display the names of all actors who appeared in this film | * SELECT * FROM * JOIN * ON * WHERE |
| * Task 9 :- Sales have been lagging among young families, so the management wants to promote family movies. Identify all the movies categorized as family films | * SELECT * FROM * JOIN * ON * WHERE   --------------------------- **Sub Query**   * SELECT * FROM * WHERE * IN |
| * Task 10i :- Display the maximum, minimum and average rental rates of movies based on their ratings. The output must be sorted in descending order of the average rental rates. * Task 10ii :- Display the movies in descending order of their rental frequencies, so the management can maintain more copies of those movies | * SELECT * MAX() * AS * MIN() * AVG() * FROM * GROUP BY * ORDER BY * DESC   ----------------------------------------   * SELECT * COUNT * AS * FROM * JOIN * ON * ORDER BY * GROUP BY * DESC |
| * Task 11 :- In how many film categories, the difference between the average film replacement cost and the average film rental rate is greater than $15? Display the list of film categories identified above, along with the corresponding average film replacement cost and average film rental rate. | * SELECT * AS * AVG() * FROM * JOIN * ON * GROUP BY * HAVING |
| * Task 12 :- Display the film categories in which the number of movies is greater than 70. | * SELECT * AS * COUNT * FROM * JOIN * ON * GROUP BY * HAVING |

Explanation of above techniques: -

|  |  |
| --- | --- |
| 1. SELECT | To retrieve specific data from one or more tables in a database. |
| 1. CONCAT | To concatenate (combine) two or more strings into a single string. |
| 1. AS | It is a keyword to assign an alias or a temporary name to a column or a table in the query. |
| 1. FROM | It is a keyword to specify the source table or tables from which data is being retrieved or referenced in a query |
| 1. COUNT | To calculate the number of rows or records that match a specified condition |
| 1. JOIN | To combine rows from two or more tables based on a related column between them |
| 1. ON | It is a keyword and is used in conjunction with the JOIN clause to specify the join condition between two tables |
| 1. IN | It is a keyword to specify multiple values in a subquery or a list for comparison in a WHERE clause |
| 1. GROUP BY | This clause divides the result set into groups based on the specified columns, and the aggregate functions operate on each group individually, returning a single result per group |
| 1. HAVING | To filter the result set of a query based on conditions that involve aggregate functions |
| 1. WHERE | To filter rows from a table based on specified conditions. |
| 1. ORDER BY | This clause is used to sort the result set of a query in ascending or descending order based on one or more columns. |
| 1. DESC | This keyword is used to specify the descending order for sorting in the "ORDER BY" clause. |
| 1. MIN() | This function is used to calculate the minimum value from a column or a set of values. |
| 1. MAX() | This function is used to calculate the maximum value from a column or a set of values. |
| 1. AVG() | This function is used to calculate the average value from a column or a set of values. |
| 1. LIKE | This operator is used to search for a specified pattern within a column. It is commonly used in the "WHERE" clause to filter rows based on a pattern match. Wild cards "%" (represents any sequence of characters) and "\_" (represents any single character) |
| 1. OR | This operator is a logical operator used to combine multiple conditions in a query's "WHERE" clause |
| 1. BETWEEN-AND | This operator is used to filter rows based on a range of values. It is used in the "WHERE" clause |
| 1. LIMIT | This clause is used to restrict the number of rows returned in a query result to a specified maximum number. |

Conclusion: -

* Most repeated first and last names of the actors are there. They are more popular in movies and people are easily identify and lovable actors so the management takes these repeated then audience more attract to those movies. (Task 2,3)
* G rating movies are lowest number of movies but these rating movies are appropriate for people of all ages if they make movies in different genres then the management get more profits. (From task 4iii)
* The replacement cost is good between $11and $20 so here I suggest that replacement cost with these costs have more movies 424 movies are there so management maintain this range of replacement cost. (task5ii)
* 41 movies have lowest replacement cost so these movies should increase their replacement cost these 41 movies are related to different categories. (task 5iii)
* 53 movies have highest replacement cost and 341 movies lowest rental rate so the management to increase the rental rate to lowest ones then they will get more profit. (task5iii)
* The movies starting with ‘k’ and ‘q’ are popular and the movies starting with ‘k’ are 12 and ‘q’ are 3. the management make more movies starting with ‘k’. (task 7)
* There are 69 family category movies and I suggested here these family category movies are promote in children category movies because most of the children movies are watching whole family. (task9)
* The average rental rate is lowest for G rating movies so management will increase the rental rate of G rating movies it will give profits. (task 10i)
* BUCKET BROTHERHOOD, ROCKETEER MOTHER are most frequently Rented movies so I suggest to make more copies of it then it will useful of management and customers also. (task 10ii)
* I performed that tasks and give information according to management requirements. I think my analysis will helpful to management.