movies_ratings

AUTHOR
Radzhana Rabdanova

1. Load necessary libraries

library(DBI) library(RSQLite) library(dplyr) library(ggplot2)

2. Connect to the SQLite database

Using the path to your file in C:/Csqlite

con <- dbConnect(RSQLite::SQLite(), "C:/Csqlite/sqlite-tools-win-x64-3510200/movie_ratings.db")

3. Fetch data using SQL JOIN to combine tables

query <- " SELECT u.name AS user_name, m.title AS movie_title, r.rating FROM ratings r JOIN users u ON r.user_id = u.user_id JOIN movies m ON r.movie_id = m.movie_id"

df <- dbGetQuery(con, query)

4. Close the connection

dbDisconnect(con)

5. Data Analysis: Calculate average ratings

We use na.rm = TRUE to handle the NULL values from our survey correctly

average_ratings <- df %>% group_by(movie_title) %>% summarize(avg_score = mean(rating, na.rm = TRUE)) %>% arrange(desc(avg_score))

Display the resulting table

print(average_ratings)

6. Visualization: Create a bar chart for the report

ggplot(average_ratings, aes(x = reorder(movie_title, avg_score), y = avg_score)) + geom_col(fill = "steelblue") + coord_flip() + labs(title = "Average Movie Ratings from Survey Participants", subtitle =

“Analysis of popular movies (Missing values handled as NULL)”, x = “Movie Title”, y = “Average Score (1-5)”) + theme_minimal()