Patient IQ SQL Questions

Write 4 SQL queries to answer the following questions. Assume the CSV files are loaded into the database or pandas dataframe with the same data structure.

1. How many movies have more than one spoken language?

SELECT COUNT(language)

FROM meta\_data

WHERE language LIKE ’%},%’;

1. What are the top 3 movie(s) in revenue in the 1990s and the 2000s, respectively?

SELECT TOP 3 title

FROM meta\_data

ORDER BY revenue DESC

WHERE realease\_date BETWEEN ‘1990-01-01 00:00:00' AND '2000-01-01 12:00:00'

UNION ALL

SELECT TOP 3 title

FROM meta\_data

ORDER BY revenue DESC

WHERE realease\_date BETWEEN ‘2000-01-01 00:00:00' AND '2010-01-01 12:00:00';

1. What are the annual profits for movies produced by `Walt Disney Pictures` or `Walt Disney Studios Motion Pictures`?

SELECT EXTRACT(year FROM release\_date) AS year,

(SUM(revenue) – SUM(budget)) AS profits

FROM meta\_data

WHERE production\_companies LIKE ‘%Walt Disney%’;

1. Which movie has the 3rd highest average rating in the 2000s? If there are ties, list the one with the most ratings left by users.

CREATE avg\_ratings AS

SELECT movieId, AVG(rating) AS avg\_rating, COUNT(userId) AS num\_ratings

FROM ratings

GROUP BY movieId;

SELECT meta\_data.title, avg\_ratings.avg\_rating

FROM meta\_data

JOIN links

ON meta\_data.imdb\_id = links.imdbId

JOIN avg\_ratings

ON avg\_rating.movieId = links.movieId

ORDER BY avg\_ratings.avg\_rating DESC

WHERE meta\_data.release\_date BETWEEN ‘2000-01-01 00:00:00' AND '2010-01-01 12:00:00'

OFFSET 2 ROWS

FETCH NEXT 1 ROWS ONLY;