Databases for data analytics

https://github.com/evidencebp/databases-course/

Assignments Feedback

Submission Requirements

You must submit **two files** as part of this assignment:

1. A **PDF file** containing the following for each question:

   ○ The SQL query you wrote.

   ○ The count of rows in the output.

   ○ A screenshot of the results of running the query.

   ○ A brief and clean explanation of your approach and reasoning in English

2. **A SQL file** containing exactly **X SELECT queries**, one for each question.

● The file names must be **identical** in both submissions and must include the student ID number (e.g., 123456789.pdf and 123456789.sql).

Select - Exercise #1 feedback

- **All those that already submitted did very well!**
- Pay attention to
    - Query structure (line per part)
    - Selected expressions (if you filter by length, it should be interesting)
    - Order by
- Minor mistakes
    - Lower case
    - Lower case on column and not search expression, no spaces
    - Condition order (length(name) < 80)
- Extra credit
    - Comments
    - Use imdb_ijs
    - Few options (yet that avoid in the future)
    - between

DML - Exercise #2 feedback

- Pay attention to
    - Assignment format
    - Use functional aliases when possible (e.g., monroe, costar)
    - test your results!
    - features not discussed in class - CTE, SET SQL_SAFE_UPDATES = 0;
    - consistency (like in alfred and Hitchcock)
- Minor mistakes
    - Better to join to the anchor tables and not related ones
    - unneeded tables (movie in the middle of Monroe)
    - Reflexiveness is a choice
    - prefer join over exists
    - name comparison - not with concat but first and next
    - space and newlines between parts
    - no cheating (Hitchcock's id), no redundancy (name and id)
    - COUNT(*) >= 1
    - remarks in Hebrew
- Extra credit
    - checks!
    - distinct in co-stars

Aggregations and Grouping - Exercise #4

● All those that already submitted did very well!

　　○ Most queries showed a good understanding of joins, grouping, and aggregation. Well done!

● Pay attention to

　　○ File name format - many lost 5 points for not following the required naming convention. ○ Use COUNT(DISTINCT actor_id) instead of COUNT(*) to avoid counting duplicates

　　○ Make sure the query answers the full question (e.g., distribution of actor counts, not just movies with zero actors)

　　○ Use meaningful and consistent aliases (actor_count, num_actors, etc.)

　　○ Don't forget ORDER BY when relevant

● Minor mistakes

　　○ Wrong join logic (e.g., joining movie ID to actor ID)

　　○ Some subqueries were more complex than needed

　　○ Missing or misused GROUP BY in a few submissions

　　● Chosen interesting question was very poor - **do not miss this opportunities next time**