

Athens University of Economics and Business
Department: Computer Science
Course: Databases
Academic Year: 2019-2020
Instructor: Vasilis Vassalos

Final Project

Goal:

In this project we will deal with setting up a database that consists of movies data as well as with a small analysis and visualization on the data.

Part A

Data:

Dataset can be found [here](#). According to these data, you should create the database, the relationships between the tables and also insert the data into the tables by defining the correct data types in all attributes.

Data pre-processing:

For the needs of this assignment you should delete all duplicates from the tables (except “ratings”) as well as delete movies data that do not exist in the “movies_metadata” table, but exist in one of the other tables.

Part A Output:

You must deliver a sql file in a folder named “partA” which will contain table creation commands, key generation commands. You must also write a short report (.pdf file) describing the steps that you followed when processing your data in order they get stored in the database. Finally an ER Diagram must be delivered in a .png file.

Part B

In this section you must calculate the following statistics. Whoever visualizes them in charts will have a 0.5 bonus.

- Number of movies per year

- Number of movies per genre
- Number of movies per genre and per year
- Average rating per genre
- Number of ratings per user
- Average rating per user

Finally create a view table and save the number of ratings and the average score of each user.
Do we get some insight from this relationship?

Part B Output:

You must deliver a sql file in a folder named “partB” which will contain the commands you used to calculate all of the above as well as the view_table creation command. Those of you who visualize the data, place image files in the folder as well.