

Project 02 – Given Week 07
Due Week 11 by Thursday 4PM

The second project is about designing and implementing an Olympic database. The main idea is to store information about the athletes, countries and the games in which they participated and won. We will get data from two databases available in Kaggle (see below). But the main problem is that the tables by Kaggle provided are very coarse (and very few) with lot of redundancies. I want to build a database that is in third normal forms by making some assumptions on functional dependencies. I will give a few ideas below, but I want you to think and design a good database.

An athlete has an Olympic ID, name, gender, country and other information. Every Olympic game is played in a particular year, has a season (winter or summer), city hosting it. Some assumptions need to be made and here are some suggestions. Each athlete participates in at least one game and can participate in several events in any one game. Obviously, an athlete can take part in several games over the year. An event has a name and a category and not all competitors get a medal. Also remember that some events are individual events and others are team events.

DATA (lot of redundancies):

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results#noc_regions.csv

<https://www.kaggle.com/the-guardian/olympic-games>

The queries that you perform would include the following. So, make sure you can answer these queries using your databases. Remember the main reason we need a database is to answer queries.

- For the year 1992 in Barcelona display the country name and the total number of competitors from that country, including those countries that have no competitors, in descending order of the number of competitors. Remember not all countries participate in every Olympics games.
- For the Vancouver (2010) games, list the competitor countries in the Curling competition.
- List all the competitors who have competed in more than 4 events in any Olympics since 1900.
- Find the number of competitors in each Olympics game who have competed in at least 3 events and group them by year, after 1940.
- Count the number of competitors who were from the India in every Olympics held since 1947.
- Display the results for swimming events (both men and women) in the 2004 Olympics.
- Find the medal counts for Michael Phelps across multiple years, show it by year and medal type.
- Which country has won the most Gold medals in the Men's marathon.
- Find names of competitors who have improved or maintained their standing across three games. Use years after 1940.

For additional information refer to this website. It has rich information about all Olympic games.

<https://www.olympic.org/olympic-games>

I want you to perform these queries, capture the queries and results in a text file and upload the results onto Canvas. I also want you to attach an ER Diagram for the database that you implemented. Remember to use the naming convention of LastName_FirstName_FileInfo.zzz for naming the files.
