Maharaja Bir Bikram University
Statistics Honours [4th Semester]
Paper: STAT-SEC2 (Credit 2)

Database Management System

Assignments - MBB College – 2019 (Summer) - Total marks: 100

<u>Assignment – 2 (Total marks: 40)</u>

Details of Data

Data file format: csv / text
Database Name: olympic
Number of Tables: 2

Table 1: *athlete_events.csv*

No. of columns: 15
No. of records: 271116
Brief description: This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016.

It should be noted that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on. Globally speaking, the Summer Olympics are more popular. There are far bigger and more countries participate in them. The Winter Olympics are only popular in countries, where winter sports are a thing. This is mostly only the case on the northern hemisphere (North America, Europe+Russia).

Fields (columns):
ID: *Unique number for each athlete*
Name: *Athlete's name*
Sex: *M or F*
Age: *Integer*
Height: *In centimeters*
Weight: *In kilograms*
Team: *Team name*
NOC: *National Olympic Committee 3 letter code for Team/Country*
Games: *Year and season*
Year: *Year*
Season: *Summer or Winter*
City: *Host city*
Sport: *Sport*
Event: *Event*
Medal: *Gold, Silver, Bronze, or NA*

Table 2: *noc_regions.csv*

No. of columns: 3
No. of records: 230

Fields (columns):
      NOC: *National Olympic Committee 3 letter code for Team/Country*
      region: *Team/Country*
      notes: *Other information*

Required Queries and Analyses

1. Create a MySQL / SQLite / MS Access database named *olympic*. Insert all data of the database described above into two tables named *athlete_events* and *noc_regions*. Write down all necessary steps to complete this task. (10)
2. Select the ID, name and country of top 10 athletes in decreasing order of number of medals they awarded (10)
3. Select the ID, name and country of top 10 female athletes in decreasing order of number of Gold medals they awarded in summer Olympics. (10)
4. Obtain the correlation coefficient between height and weight of Male athletes whose age are not less than 25. (5)
5. Obtain ungrouped frequency distribution and summary statistics of age of athletes. Also draw the frequency column diagram of the distribution. (5)