Programming for Data Analytics

SOFT 8032

2023 Project Specification

Task 1

Read the dataset and import it into an appropriate format. Select the 20 oldest athletes and utilize

an appropriate visualization technique to display the percentage of athletes from each country

(Team) involved. See Figure 1.

Explain your implementation strategy and the detail of the algorithm in the recorded video.

Task 2

Use an appropriate visualization technique and discuss the Height distribution among athletes.

Visually show what is median, outliers, max and min.

Explain your implementation strategy and the detail of the algorithm in the recorded video.

Task 3

Use an appropriate visualization technique and display the top 10 countries (Teams) that gained

the maximum number of Gold medals.

Explain your implementation strategy and the detail of the algorithm in the recorded video.

Task4

Among the following features, what group of features with size 4 can predict the Medal better?

E.g., (Sex, Height, Team, Games) or (Sex, Sport, Team, Games) or ...

1. Sex

2. Height

3. Team

4. Games

5. Weight

6. Sport

Your solution should be based on utilizing supervised learning algorithms. Use a visualization technique and display what group of features (with size 4) can predict Medal better.

Explain your implementation strategy and the detail of the algorithm in the recorded video.

Task5

In this task, you are required to find out in which sports, women received more Gold medals than men. Print the name of those sports in the consul.

Explain your implementation strategy and the detail of the algorithm in the recorded video.

Task6

In this task, you are required to find the followings:

1. The year with the maximum number of Gold medals.

2. The year with the maximum number of Silver medals.

3. The year with the maximum number of Bronze medals.

Explain your implementation strategy and the detail of the algorithm in the recorded video.
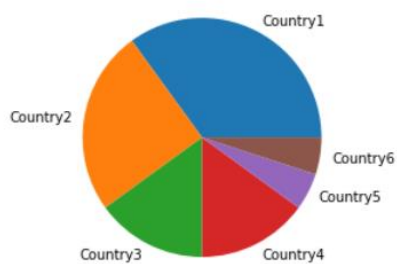

Figure 1: Countries.