Sample Exercise Instructions

1. Locate and download data sets you have been given access to
   a. athlete_info.csv
   b. event_info.csv
   c. NOC_info.csv

2. Open the data sets and review the structure of the data. (column and row headings).

3. Merge all 3 data sets together to create 1 dataset combining all the columns

4. Create a Year and a Season column from the Games column information

5. For the Age column replace any non numeric values with the most popular age

6. For the Height column replace any non numeric values with the average height

7. For the Weight column replace any non numeric values with the middle value for weight

8. Ensure your merged file has only unique columns across the two datasets.

9. Save this file to create an appropriately named separate file or data source of the merged data

10. Create a dataset of only Gold Medal winners and save this as an appropriately named separate file or data source of the gold medal winners

11. Create a chart of the number of gold medals with respect to age. Choose a chart you think best summarises this information.

12. Show the average age for each sport and each event for gold medalists.

13. Display a count of all women in athletics over the years and use a chart to visualise

14. Create a second chart showing the male v female numbers over the years

15. Display the top 5 countries who have won the most medals.

16. Display the weight range for males over years for each of the sports