

DATA 601 - Homework Assignment 2

Deadline: 23:59 pm, October 10, 2021

This week, you will work on four datasets. You can either download these csv files from our GitHub repository or the Blackboard page to work with files locally or you can use the "requests" module and access these files online (as explained in DATA601Lecture02_p2_Python_IO.pdf, slide 47).

Here's a link to the repo: [Datasets](#)

Note that in this assignment,

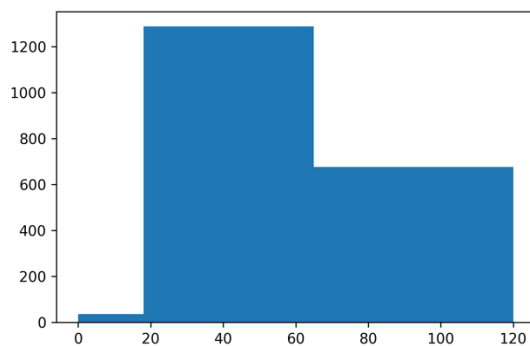
- You are not allowed to use pandas, because we haven't learned it yet. So you need to use the "csv" module to read these csv files.
- All the csv files have a header. You can skip the header with the next function, e.g. assuming you are reading local files

```
readCSV = csv.reader(csvfile, delimiter=',')
next(readCSV)
```

- Do not forget that the "csv.reader" reads everything as text. You will need to convert those readings to integers and floats, when necessary.
- Do not forget labeling x and y-axes, using grids and legends (if there are more than one line).

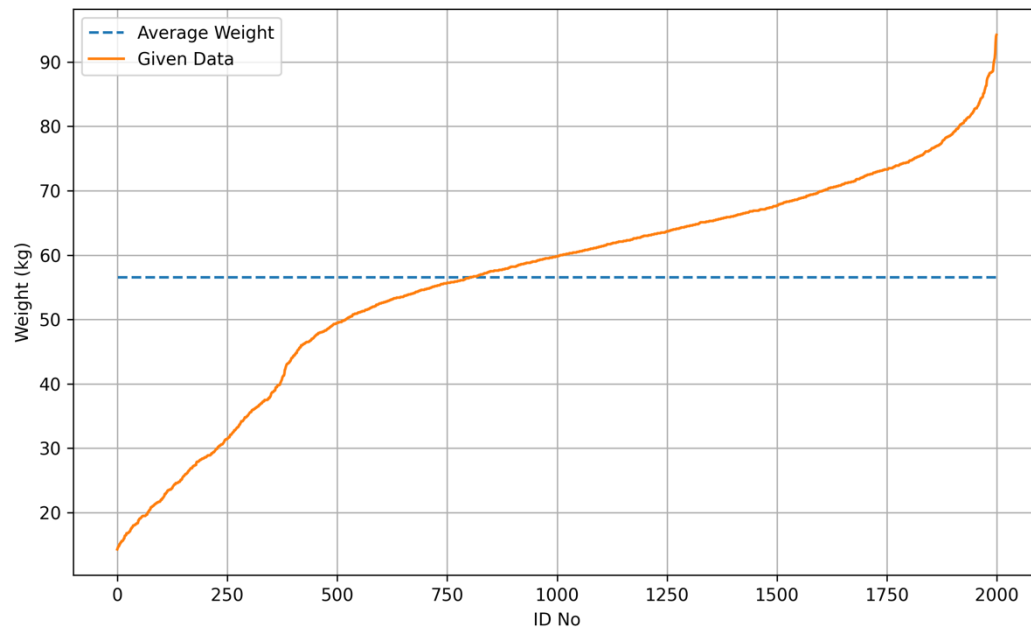
[9 points] Question 1

- Read the "HW2_People.csv" with the csv module's reader function and create lists of firstnames, lastnames, birthyears, birthmonths, birthdays, weights. Note that first two lists would be strings, the following three are integers, and the last one is a list floats.
- Determine the length of the one of the lists.
- Plot the histogram of "weights" using 16 bins.
- Define "edges" and determine the number of people younger than 18, older than 65, and the rest. Plot the histogram showing the number of people in these three age groups. It should look like this (with labels)



- Convert the list "weight" to a numpy array (e.g. `w = np.asarray(weights)`), sort "w", and plot it, where x axis

will be the index (i.e. changing from 1 to 2000). On the same figure, plot the average weight with a horizontal dashed line, as shown below.



- f. Plot the histograms of birth-years with 7 bins, birth-months with 12 bins, days of the births with 16 bins.
- g. Determine mean, max, and min weights.
- h. Determine mean, max, and min ages (with respect to current year, 2021).

[7 points] Question 2

For this question, you will work with Messi.csv and Ronaldo.csv files. For those who are not familiar with Messi and Ronaldo, they are two of the best soccer players of all times.

In both data sets, there are 6 columns as follows:

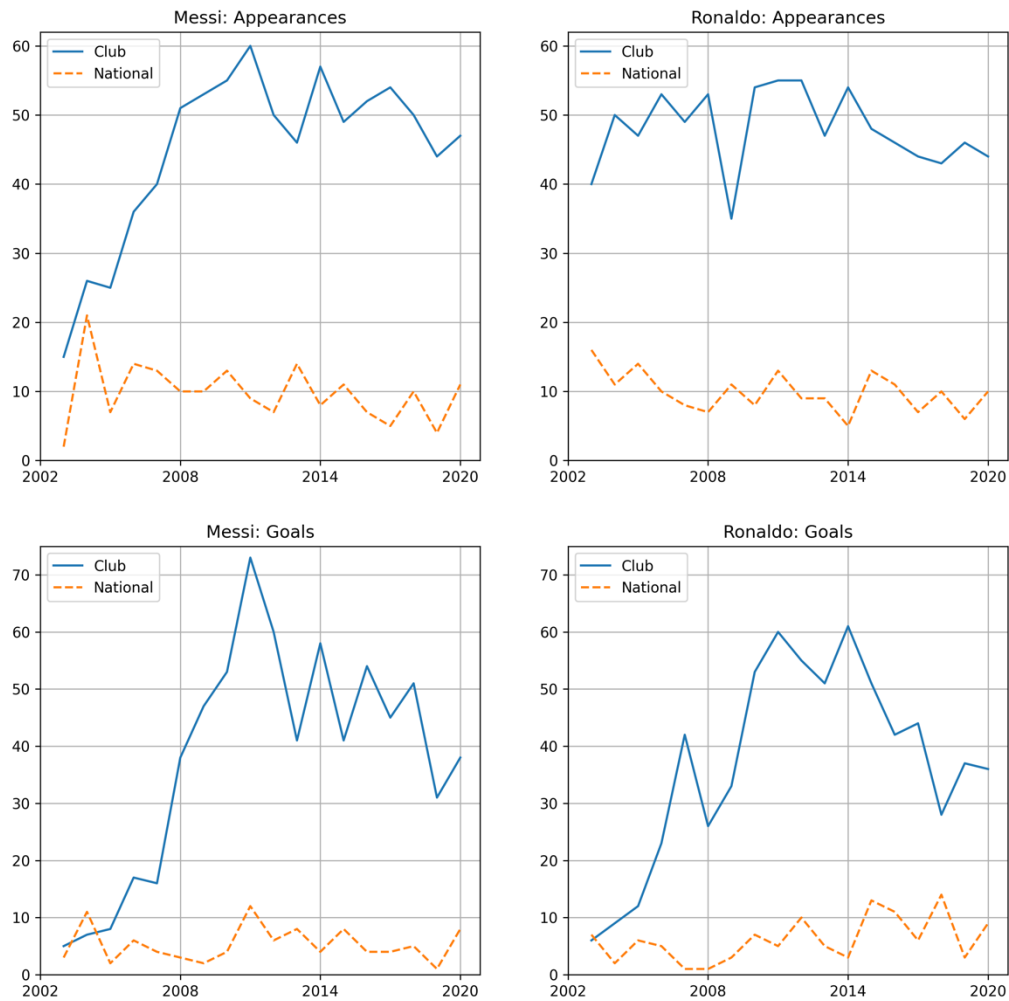
1. Start Year of the Season,
2. End Year of the Season,
3. Number of games played for the professional club
4. Number of games goals made for professional club
5. Number of games played for the national team
6. Number of games goals made for national team

a. Read these two files and create lists of "seasonstart, seasonend, clubapp, clubgoal, natapp, natgoal" for both players.

b. Generate the figure below in which we have 4 subplots.

- Start Year of the Season vs Number of games for Messi (club vs. national)
- Start Year of the Season vs Number of games for Ronaldo (club vs. national)

- Start Year of the Season vs Number of goals for Messi (club vs. national)
- Start Year of the Season vs Number of goals for Ronaldo (club vs. national)



Note that there are so many small details in the figure above. Pay attention to details

- Calculate total number of goals made by Messi (club+national).
- Calculate total number of goals made by Ronaldo (club+national).
- Calculate goals per game for Messi (club+national).
- Calculate goals per game for Ronaldo (club+national).
- Which one is better: Messi or Ronaldo? Be careful about your answer if you don't want to fail this class!

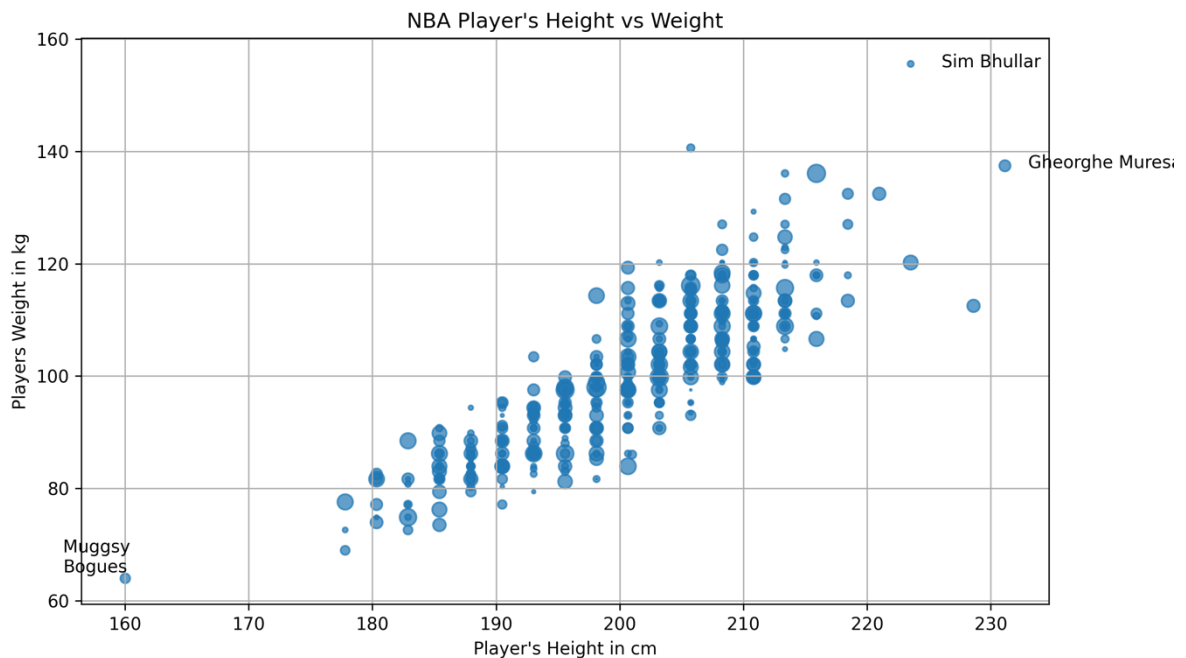
[4 points] Question 3

Read the "season96_97.csv", where we have NBA player statistics from 1996-97 season. Use players' heights, weights and pts to recreate the visualization below, using matplotlib.pyplot's scatter function.

Note that

- `scatter(x,y,z)` creates small circles with a radius proportional to `z` at `(x,y)`. Here, the size of these circles should change according to player's average points.
- Set the transparency value to 0.7.
- Add text at `x = 155, y = 65` as 'Muggsy Bogues'
- Add text at `x = 226, y = 155`, as 'Sim Bhullar'
- Add text at `x= 233, y = 137`, as 'Gheorghe Muresan'

Final output should look this.



Save the output as a png file, whose name includes yourname, e.g. 'HW2Q3_Ergun.png'

BONUS

You will get 2 extra points if you read the csv files from the GitHub repository.