Python Programming

Assignment 1 – 20%

Deadline: 17ᵗʰ October 2024

I. (4 points)
Write a Python program to collect footballer player statistical data with the following requirements:

- Collect statistical data [*] for all players who have played more than 90 minutes in the 2023-2024 English Premier League season.
- Data source: https://fbref.com/en/
- Save the result to a file named 'results.csv', where the result table has the following structure:
  - Each column corresponds to a statistic.
  - Players are sorted alphabetically by their first name. If there is a name tie, sort them by age from oldest to youngest.
  - Any statistic that is unavailable or inapplicable should be marked as "N/a".
- [*] The required statistics are:
  - **Nation**
  - **Team**
  - **Position**
  - **Age**
  - **Playing Time**: matches played, starts, minutes
  - **Performance**: non-penalty goals, penalty goals, assists, yellow cards, red cards
  - **Expected**: xG, npxG, xAG
  - **Progression**: PrgC, PrgP, PrgR
  - **Per 90 minutes**: Gls, Ast, G+A, G-PK, G+A-PK, xG, xAG, xG + xAG, npxG, npxG + xAG
  - **Goalkeeping**:
    - Performance: GA, GA90, SoTA, Saves, Save%, W, D, L, CS, CS%
    - Penalty Kicks: PKatt, PKA, PKsv, PKm, Save%
  - **Shooting**:
    - Standard: Gls, Sh, SoT, SoT%, Sh/90, SoT/90, G/Sh, G/SoT, Dist, FK, PK, PKatt
    - Expected: xG, npxG, npxG/Sh, G-xG, np
  - **Passing**:
    - Total: Cmp, Att, Cmp%, TotDist, PrgDist
    - Short: Cmp, Att, Cmp%
    - Medium: Cmp, Att, Cmp%
    - Long: Cmp, Att, Cmp%
    - Expected: Ast, xAG, xA, A-xAG, KP, 1/3, PPA, CrsPA, PrgP
  - **Pass Types**:
    - Pass Types: Live, Dead, FK, TB, Sw, Crs, TI, CK
    - Corner Kicks: In, Out, Str

- - - Outcomes: Cmp, Off, Blocks
  - o **Goal and Shot Creation**:
    - SCA: SCA, SCA90
    - SCA Types: PassLive, PassDead, TO, Sh, Fld, Def
    - GCA: GCA, GCA90
    - GCA Types: PassLive, PassDead, TO, Sh, Fld, Def
  - o **Defensive Actions**:
    - Tackles: Tkl, TklW, Def 3rd, Mid 3rd, Att 3rd
    - Challenges: Tkl, Att, Tkl%, Lost
    - Blocks: Blocks, Sh, Pass, Int, Tkl + Int, Clr, Err
  - o **Possession**:
    - Touches: Touches, Def Pen, Def 3rd, Mid 3rd, Att 3rd, Att Pen, Live
    - Take-Ons: Att, Succ, Succ%, Tkld, Tkld%
    - Carries: Carries, TotDist, ProDist, ProgC, 1/3, CPA, Mis, Dis
    - Receiving: Rec, PrgR
  - o **Playing Time**:
    - Starts: Starts, Mn/Start, Compl
    - Subs: Subs, Mn/Sub, unSub
    - Team Success: PPM, onG, onGA
    - Team Success xG: onxG, onxGA
  - o **Miscellaneous Stats**:
    - Performance: Fls, Fld, Off, Crs, OG, Recov
    - Aerial Duels: Won, Lost, Won%
  - o Reference: https://fbref.com/en/squads/822bd0ba/2023-2024//en/squads/822bd0ba/2023-2024/Liverpool-Stats

II. **(2 points)**

- Identify the top 3 players with the highest and lowest scores for each statistic.
- Find the median for each statistic. Calculate the mean and standard deviation for each statistic across all players and for each team. Save the results to a file named 'results2.csv' with the following format:

		Median of Atttribute 1	Mean of Atttribute 1	Std of Atttribute 1	...	...
0	all					
1	Team 1					
...						
n	Team n					

- Plot a histogram showing the distribution of each statistic for all players and each team.
- Identify the team with the highest scores for each statistic. Based on your analysis, which team do you think is performing the best in the 2023-2024 Premier League season?
- Histogram Plot: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.hist.html

III. **(3 points)**

- Use the K-means algorithm to classify players into groups based on their similar statistics.
- How many groups should the players be classified into? Why? Provide your comments on the results.
- Use PCA to reduce the data dimensions to 2, then plot a 2D cluster of the data points.
- Write a Python program to draw a radar chart comparing two players with the following inputs:
  - `python radarChartPlot.py --p1 <player Name 1> --p2 <player Name 2> --Attribute <att1, att2, ..., att_n>`
  - `--p1`: The name of the first player
  - `--p2`: The name of the second player
  - `--Attribute`: The list of statistics to be compared
  - Radar chart: https://matplotlib.org/stable/gallery/specialty_plots/radar_chart.html

IV. **(1 point)**

- Collect player transfer values for the 2023-2024 season from https://www.footballtransfers.com.
- Propose a method for estimating player values.

Submission Instructions:

- The submission should include Python code.
- A report (.pdf).
- Submission details will be announced later.