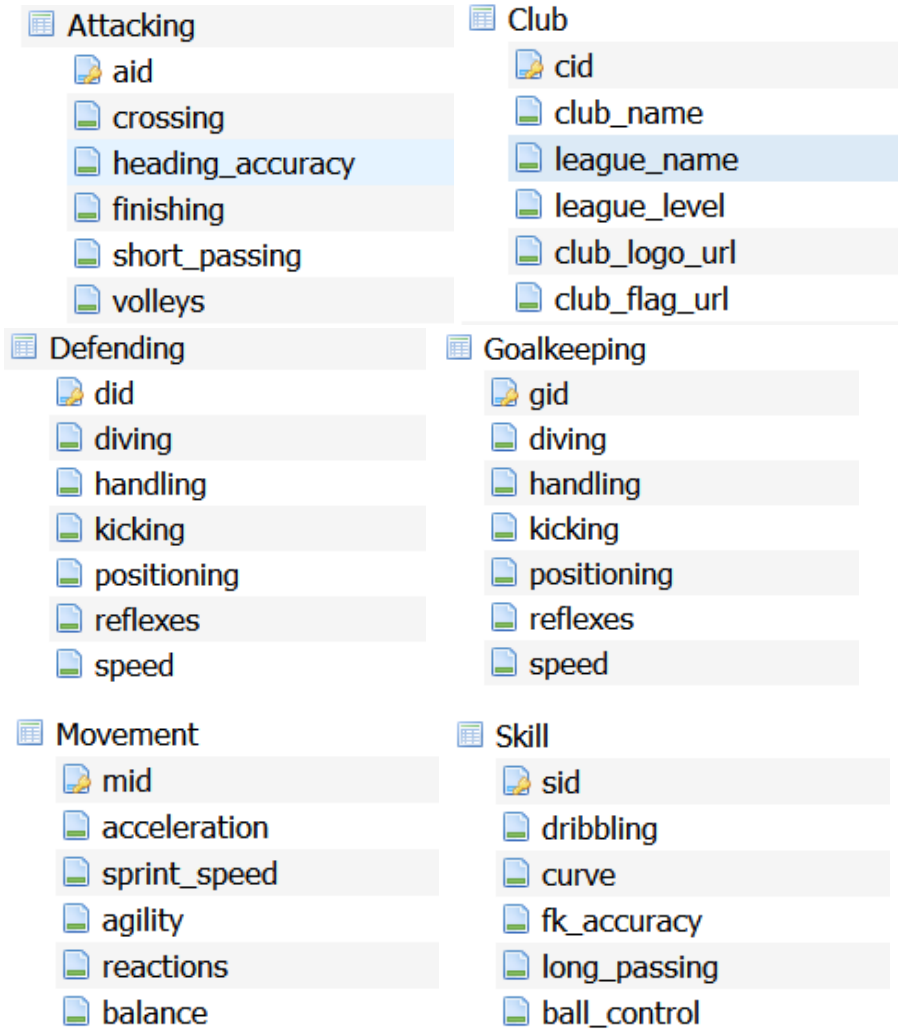
Data Management

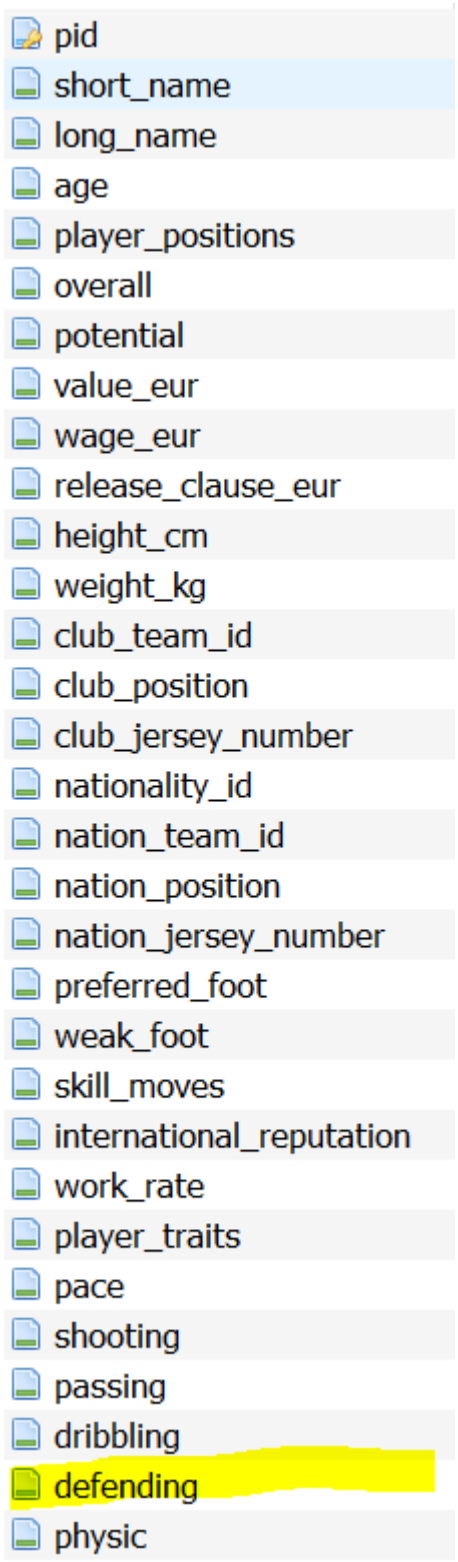
Team project

Note: Please answer all questions and hand in source codes and the results

(9 pts) Q1. Work on the FIFA dataset: FIFA.db

The FIFA dataset contains seven tables: Attacking, Club, Defending, Goalkeeping, Movement, Player and Skill.





The relationship between these seven tables is summarised below:

- pid in Player is also the foreign key to Attacking, Defending, Movement, Goalkeeping, and Skill.

- club_team_id in Player is the foreign key to Club

(3 pts) A. Find the average defending rating of players in each club and display the result based on the average defending rating, ranked from highest to lowest.

(3 pts) B. Find the names of players, their clubs, and their attacking ratings for players who have an attacking rating greater than 80. Note: attacking rating = (crossing + finishing + heading_accuracy + short_passing + volleys)/5

(3 pts) C. Find clubs (i.e., the id and the name of the club) where the average shooting rating is higher than or equal to the average shooting rating of 'AC Milan'.

(11 pts) Q2. Work on the red wine dataset: Wine.db

The data of red wine contains four tables: Wine, Winery, Grape, and Rater.

- Features in Wine:
  - id: primary key
  - points: points of the wine (from 0 to 100)
  - title: title of the wine
  - description: comment on the wine
  - price: the price of the wine in USD
  - designation: name of the wine
  - year: the year of the wine
  - gid: foreign key to the table Grape
  - wid: foreign key to the table Winery
  - rid: foreign key to the table Rater
- Features in Winery:
  - id: primary key
  - country
  - province
  - winery: the name of the winery
- Features in Grape:
  - id: primary key
  - variety
- Features in Rater:
  - id: primary key
  - name: the name of the Rater
  - twitter_handle: the twitter account of the Rater

(3 pts) A. Grapes are the raw material for making red wine. Which country has the most varieties of grapes, please report the name of the country and the number of types of grapes.

(3 pts) B. According to the guide published by wine enthusiasts, red wine can be divided into four classes based on its price: everyday wine, mid-range wine, fine wine, and super fine wine. If the price of red wine is higher than or equal to $340, it will be classified as super fine wine. Please report the average price and points of the super fine wines.

(5 pts) C. A rater evaluates the score and price of the wine fairly and unbiasedly based on the quality of the wine. Please write SQL code and python code to analyse the data statistics of points and prices of wine evaluated by each taster and use the format rater_name.xlsx to export the data statistics.

(Note: data statistics includes the min, max, median, 25-quantile and 75-quantile information of the points and prices.)