Construct database with real data:

1. Preprocessed the data:

Our project will use real data source from FIFA official dataset[[1]](#footnote-1) which contains all the players’ information from season 14/15 to 20/21. The dataset is given in the form of 7 csv files(which is put in the /FIFADB/Data/), each of them contains the whole information of each player in each season.

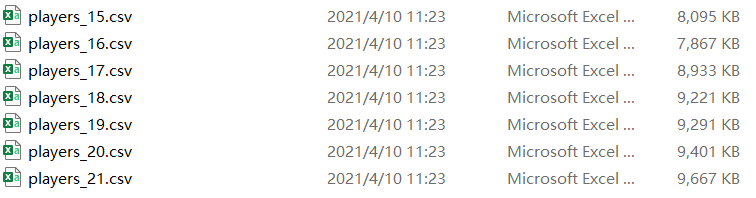


Figure 1: data source file

To make it more convenient to import data to the database, we first need to retrieve all the information needed for each table in the normalized schema. Thus we write a python program (/FIFADB/Codes/ step\_0\_process\_data.py) to scan through all the csv files and put data for each table in a corresponding csv file (put in the /FIFADB/Data/Data\_processed). Notice that we further clean and modify these data by modifying these files generated on python that is why the file generated from python program will be different from these files in the folder.

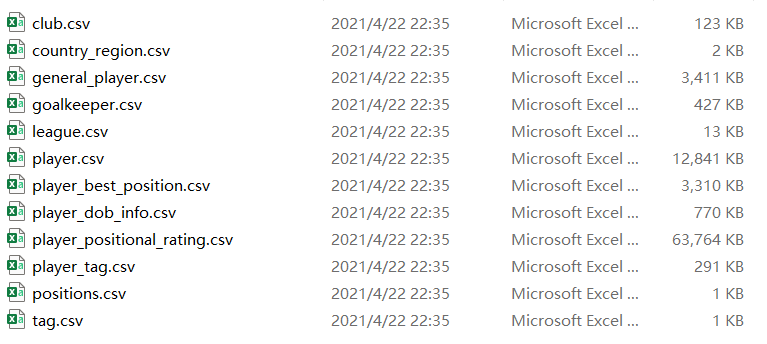


Figure 2: csv file for each table

1. Construct database tables and import those prepared data:

After preparing data, we then need to construct a database contains all the tables in normalized schema. Notice that to avoid any importing data abnormal, we construct the original table with all the attributes in type varchar, which we will set back to appropriate type in the later step. All the attributes are set in a proper size represented in length of varchar and we have set some attribute to be not null to follow our database rules. (All the codes of constructing tables are in

/FIFADB/Codes/ step\_1\_construct\_tables.sql).

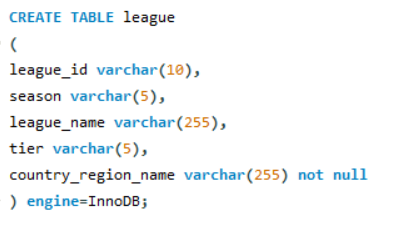


Figure 3: example of constructing tables

With all tables constructed, we then need to import all the csv file data into their tables. Codes used to import data are put in /FIFADB/Codes/step\_1\_construct\_tables.sql. Notice that if you want to execute this sql file, you need to change the csv files path to the corresponding data path in your computer.



Figure 4: example of importing data

1. Set keys and constraints of tables

In the next step, we need to set proper constraints to those tables. Firstly we need to clean some abnormal data which will cause failure of setting further constraints.



Figure 5: clean abnormal data

Since we initially constructing all the table using attribute type varchar, at this stage we need to convert them back to normal type.

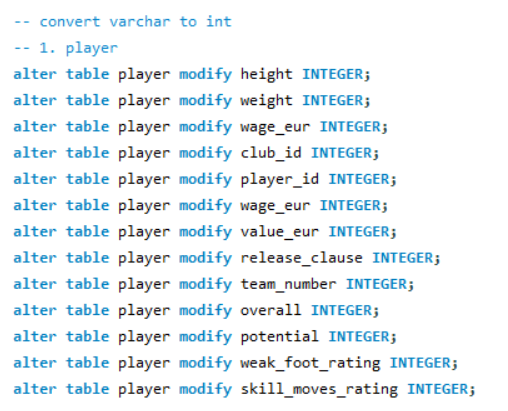


Figure 6: set attribute type(partial)

With all the attributes in proper types, we can then set primary keys and foreign keys for each table.

Notice that we need to first set up all the primary keys for all tables and set their foreign keys afterwards otherwise it will cause error.



Figure 7: set primary and foreign keys(partial)

To make sure the attributes in the database will follow the constraints we set, we also need to set constraints for some attributes in certain table.

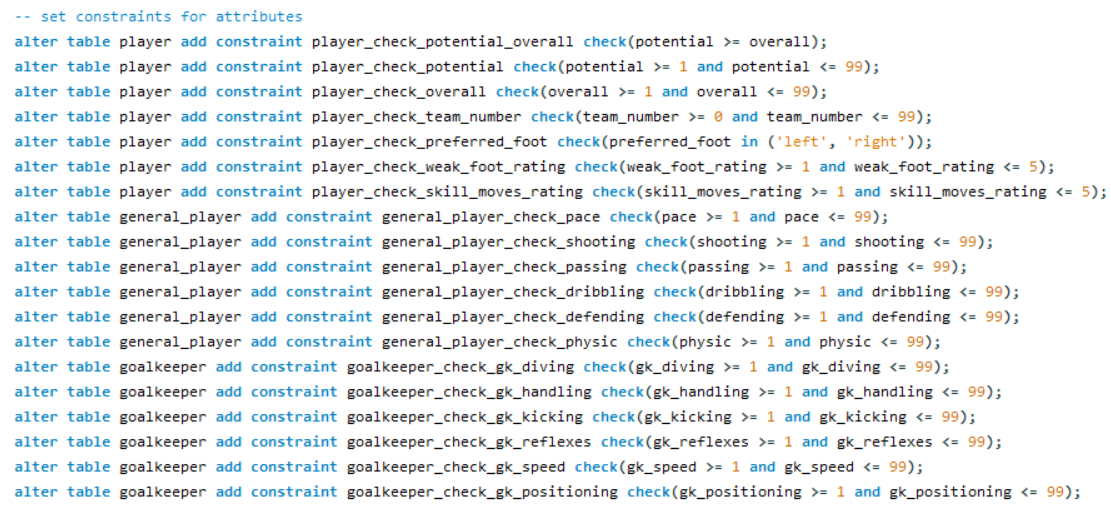


Figure 8: set constraints for some attributes

Notice the codes mentioned in this section are put in /FIFADB/Codes/step\_3\_set\_constraints.sql.

1. Dataset is retrieved from https://www.kaggle.com/bryanb/fifa-player-stats-database. [↑](#footnote-ref-1)