

databases final project

FIFA18 Video Game Database



December 16, 2017

Ye Chan Kim (315) | Jae Goan park (315)

ykim160@jhu.edu | jpark230@jhu.edu

Our database consists of various elements from the popular video game FIFA 18. The database includes soccer players’ personal information, their attributes, quantifiable skill level and associated clubs.

**3. Change from Phase I**

We have made some design changes to the FIFA 18 database. We decided to remove the continent relation as it did not directly relate to the goals of the project and instead added more parts to make it easier for a user to navigate and make queries to the database. In addition to the Offensive(Attack) and Defensive(Defense) skill relations we’ve also added the midfield skill relation to cover a broader range of players who may fall equidistantly between the three different categories.

Also we decided to add a new relation called “Country” that contains all the country FIFA codes as Key with the associated country names. This meant that we removed the country\_name column from the Player relation and instead added “NTC” which contains all the 3 letter FIFA country code abbreviations. Also in the Country relation, we have the country’s capital, Continent Abbreviation, sub-region, and its development status (whether or not it is a developed country).

We also made some changes to the soccer team relation. We removed the player’s associated football club from the Player relation, and instead created another relation called “Plays\_in” to link with the Player\_id from the player relation to the “Club” relation. The Club relation contains all the “Club\_id” numbers with the associated club name.

Finally, we’ve also added a attributes relation that contains a miscellaneous set of quantifiable attribute statistics to cover a broader range of skills for individual players.

**4. Loading the Database with values**

We obtained our dataset to populate out database with from Kaggle (<https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset)>, which is a renowned data science website that publishes many datasets for a variety of applications. The particular dataset contained all the data we needed to populate our data base in csv format. We used Pandas, which is a python library, to extract data from the different csv files contained in the dataset. To make the debugging process easier, we made a data extraction tool through python in several different files. Simply, each file that ends with “\_extraction.py” is responsible for extracting the data for each relation. The method of extraction was fairly simple due to the help of the pandas library. We used the pandas.read\_csv(filename, header etc.) command to read in the csv files and physically print the necessary sql commands to a sql file. We wrote to files the commands such as “DROP IF EXISTS Player” and “INSERT INTO Player” and filled out the necessary data parts by using a for loop to loop over the data extracted using the read\_csv method. The individual files that contained the extraction code outputted a sql file as a result and we ran the individual sql files on MySQL and was able to successfully populate the database.

**6. User’s guide**

1. Before proceeding with running the codes, please be sure to have either python2.7 or python3 installed each equipped with the pandas and numpy library written for the python programming language.
   1. To install pandas for python2.7: pip install pandas
   2. To Install pandas for python3: pip3 install pandas
2. Now to create the necessary sql files needed to populate the database we must run each individual \*\_extraction.py file on the command line.
   1. Please first run the file createSQL.py with command python createSQL.py to create another csv file necessary for country extraction with proper FIFA abbreviation mapping
   2. If the top comment in the extraction files ask you to run it in python2.7, run the command : python \*\_extraction.py
   3. Otherwise, if the comment asks you to run the file in python2, run the command : python3 \*\_extraction.py
   4. Python file list: Attribute\_extraction.py, club\_extraction.py, country\_extraction.py,
3. Once all the necessary sql files have been created, log into dbase and run the following command: \. ~/path\_to\_directory\_containing\_sql\_files/sql\_file
   1. Sql file list: attack.sql, attribute.sql, club.sql, country.sql, defense.sql, midfield.sql, player.sql, plays\_in.sql, Position.sql, Preferred.sql, Value.sql

Phase I

1. **List of Potential English Questions**
   1. Which player has the highest “Overall” Stat? (List all ties)
   2. List the nationality and ID of all players with the highest wages.
   3. List the club name and average wage of players by soccer club name.
   4. List all the teams that are located in Spain.
   5. List all players with “LB (Left Back)” stat higher than total LB stat average of all players.
   6. List the player who is a striker and has the highest potential to be a goal keeper.
   7. List players with more than one preferred position.
   8. List the top ten strongest players in the FIFA 18 database.
   9. List the 5 countries with the respective players that have the highest chance of scoring in a penalty kick
   10. List the most valuable club in the database.
   11. At what age do players have the highest overall average?
   12. List the fastest and the slowest player on the database.
   13. List the player name, age, and offensive stats for “Real Madrid” players
   14. Make the best team possible for a 4-3-3 formation where the players are from Spain.
   15. Who would most likely win between Real Madrid and Barcelona based on player stats.
2. Relational Model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Player** | Player ID | Name | Nationality | Age | Team | Overall | |  | 1179 | G. Buffon | Italy | 39 | Juventus | 89 | |  |  |  |  |  |  |  | | **Offensive Stat** | Player ID | Finishing | Dribbling | Agility | Acceleration |  | |  | 155862 | 80 | 90 | 76 | 89 |  | |  |  |  |  |  |  |  | | **Defensive Stat** | Player ID | Aggression | Interception | Strength |  |  | |  | 155462 | 40 | 78 | 95 |  |  | |  |  |  |  |  |  |  | | **Team** | Team Name | Country | Continent |  |  |  | |  | Real Madrid | Spain | Europe |  |  |  | |  |  |  |  |  |  |  | | **Continent** | Continent Name | Country Name | GDP (USD Tn) | Life Expectancy | Crime Rate Ranking |  | |  | Asia | Japan | 4.9 | 83 | 12 |  | |  |  |  |  |  |  |  | | **Finance** | Player ID | Wage (Mn) | Net Worth (mn) |  |  |  | |  | 1179 | 4 | 50 |  |  |  | |  |  |  |  |  |  |  | | **Position** | Player ID | Preferred Position |  |  |  |  | |  | 1179 | ST |  |  |  |  | |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

v) SQL Statements

Make the best team possible for a 4-3-3 (plus goal keeper) formation where the players are from Spain.

SELECT P.playerid

FROM Position as O,

(SELECT P.playerid

FROM Player as P

WHERE P.nationality = “Spain”) as S1

WHERE P.playerid = O.playerId AND O.preferredposition = “ST” AND

P.overall = (SELECT MAX(Overall)

FROM Player as P, Position as O

WHERE P.playerid = O.playerId AND O.preferredposition = “ST”);

\*NOTE: Repeat this procedure 11 times for all the other positions involved in the 4-3-3 formation (e.g. LW, ST, RW, LM, CM, RM, LB, CB, CB, RB, GK) and put together which a stored SQL procedure so output is integrated as one answer.

vi) Plan for loading database

We will download a comprehensive FIFA 18 player stat data from Kaggle.com from <https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>. We have not found appropriate database for country/continent stats, but are likely to collect from UN or CIA World Factbook website. The data collected will be in form of csv.

Some issues we have noticed are that some team names and player names display invalid broken text. This is probably because the original name or team name contains special characters that cannot be recognized in csv format. We will attempt to find a way to convert all of them to standard English so they are readable. Hopefully we do not have to convert them by hand one by one.

vii) Type of result we wish to generate

We wish to provide hardcore FIFA 18 gamers with the most optimal set of players to use for creating their team. Users of this database, will be able to input their general preferences about their dream team and be able to obtain several viable dream team options.

viii) Specialized/advanced topics

We plan on including advanced SQL topics during project implementation. To create the best team formation different combinations of player formations need to be formatted in a comprehensive way.