Aidan Wall
CPSC 408-01
12/18/2021
Final Project Paper (Instructions at Bottom)

Europe's Top 5 Football Leagues

As a football player and fan, I watch a lot of football. My friends do as well but everyone has their own opinions and beliefs about which teams and players are better. Every fan has their own favorite player, and favorite team, and sometimes even a league they like the most. This leads to many arguments on which players or teams are better and it seems to be never ending. By creating a database with all player data, team data, and fixtures for all of Europe's top 5 leagues I have hoped to create a way in which players and teams can be compared across leagues, not just in their respective leagues. The end goal was to build a database that allows the user to compare players and teams across leagues, and also to compare 2 teams in different leagues and predict a winner. I built a database that looks at teams, players and leagues for Europe's top 5 different leagues in the 2020-2021 season. I wanted to also include data on managers for each team but I was not able to find usable data for any league but the Premier League, which still only had name and team.

I do not have much technical experience working specifically with football data but I have lots of knowledge on the game and what stats are interesting/useful. I really enjoyed this project and learning more about how to implement databases involving a topic that I like. I learned more about the intricacies of football and some specific stats that were interesting that I will definitely be sharing with my friends.

My data is divided up into leagues.. Each league has a standingsTable, which has each club and their final position in the league, points, wins, losses, and many other team stats. Each league also has a table of players, which has their information such as name, age, club they play for which is mapped to the teams in the standingsTable, nationality, and a bunch of other statistics from the 2020/2021 season. Each league also has a fixtures table, which is every game from every week of the season, and has attributes such as home team, away team, score, stadium, and other statistics. I also have a league table that has each league, their league ID and the country that the league is in. If I had manager data it would be its own table with team, league, league ID, salary and Win-Loss record. I make use of all functions required, creating a view of all player statistics so when the user wants to look at the top players by some statistic they are just selecting from the view. I make use of subqueries in many of my functions, such as … . I use aggregation in my query for showing most or least amount of nationalities represented for different squads. I enforce referential integrity by using primary keys, but also referencing squads across leagues and players. I also have the ability to create new players as well as delete players, and error check if they exist or not. I also implement a function to update a players Age, to utilize the update function. I error handle input so when creating a player, so that they cannot enter incorrect data for data such as age or year born, I do not allow the user to enter anything but numbers so it can be inserted into mySQL. I make use of transaction and rollback when updating a players age, to make sure that they are happy with their deletion or it will roll it back. If they are happy with the update then I will commit the transaction.  I formatted all printing output to make it very easily readable and understandable for the user when they are using the program. One of my functions allows a user to select a team, then download that team's schedule

as a CSV in the same directory that the user is currently in. The name of the CSV file is teamName.csv.

I found some interesting results from my project that led me to do some more research outside of the project. When building my function/query to look at which teams had the least and most different teams represented on their team, I found that some teams had only 2 or 3 nations represented, while some had up to 17. The teams that had very few nations were Athletic Bilbao, Real Sociedad and Osasuna, with Athletic Bilbao having only 2 different nationalities. When I wondered why so little, I asked my friends about this. When talking about my friend Tony who is Basque, he explained to me that Athletic Bilbao only allows players that are Basque in its team and club. The Basque Country is a subregion in the northern part of Spain and the southern part of France. Throughout the entire club's history they have only allowed players of this descent. Real Sociedad and Osasuna, two of the other teams that have the fewest number of nationalities represented are also Basque. They also had similar policies that only Basque players were allowed to play for their club but have since changed their rules.

The results of my interface and database system is a command line interface to access and use the database. It prompts the user with all different options that the user can perform, with one of them exiting the program. The user is then brought into that option and it either prompts them with more questions or displays the information desired. Here is what the prompt screen

looks like:

```
Welcome to your APP!
Select from the following menu options:
1 Find Top 6 of a league
2 Find top goal scorer on a team
3 Find the 5 players with the most yellow or red cards
4 Download a teams schedule as a csv
5 Find the team with the most distinct nationalities represented on their team
6 Update a players age
7 Add a new player to a league
8 Delete a player from a league
9 Show average match attendance per league
10 Show best offensive players by a stat
11 Exit
```

I also had some issues with inserting my data for Ligue1 fixtures. I do not understand why this is because the same code and processing worked for all the other leagues and all of their tables. I have tried replicating this process many times, and even ran the same code to put it in DataGrip just to show that it is there. I started getting an error in mySQL saying not all parameters are used even though I went line by line to see where the number of parameters expected does not equal the number of parameters given and they all match, so I don't understand where the error is coming from and cannot figure out a solution in mySQL.

SUMMARY:

I learned a lot from this project about how MySQL works and the struggles and challenges of building and implementing a good database. It was challenging at times doing all the work myself, while also balancing this class with studying for finals and projects for all my other classes. I wanted to implement a function that would allow me to compare teams and predict who would win but I did not finish one that I was happy with so I did not include it. That will probably be something I focus on over interterm. I overall enjoyed this project and

especially having the ability to choose what I am working on which is football data, which I am interested in and always want to learn more about.

INSTRUCTIONS:

To run, first create each table in the Create_Tables.sql file. Next, uncomment all of the pre_process functions in lines 566-585 in app.py and the add_leagues function in Line 585. Then you will be able to access and use my GUI. If you run the pre_process tables then exit the program, COMMENT OUT the pre_process functions in all the previous lines. Thank you!