CS 5200 Homework #6

Creating and manipulating a MySQL schema for soccer

Domain Description
The attached files contain information on the soccer (European football) players for the 2017/2018 Premier League. The Premier league is an English professional soccer league. The Premier League is the most watched sports league in the world, broadcasted in 212 territories to 643 million homes and a potential TV audience of 4.7 billion people.

A game of soccer is played with 2 teams. There are 20 teams in the league. Each team plays 2 matches with each of the other 19 opponents, one in the team's home stadium and the other in an away stadium. Thus, in a season there are 380 matches in total and 38 matchdays. Each match is played in two halves, each half is 45 minutes. The team which scores more goals at the end of the second half wins, if both teams score the same number of goals, the match ends in a draw.

Data File Description
There are three different csv files. Each row within the **matches file** is data specific to one match. Each row within the stadium file is specific to each team's **home stadium**. Each team has only one home stadium. Each row within the **manager file** is data specific to one manager. A team may have one to many managers. Let's review the data fields stored in the files.

The data fields in the matches file (EPL_matches.csv) are the following:
Match number: Match number of the season
Match day: Match day (1 to 38)
Date: Date, in which match was played (In YYYY-MM-DD) format
(Note: In excel, if you open the csv, you might see a different format, but do not worry. After importing in MySQL you will see the above format)
Team 1: Home team
Team 2 : Away team
Half time score team 1: Home team number of goals at the end of first half
Half time score team 2: Away team number of goals at the end of first half
Full time score team 1: Home team number of goals at the end of second half
Full time score team 2: Away team number of goals at the end of second half

The data fields in the stadium's file (EPL_stadiums.csv) are the following:

Team: the team or the name of the Premier League team
Venue: the name of the team's home stadium
Each team has one home stadium.

The data fields within the managers file (EPL_managers.csv) are the following:
Manager: the name of the manager for the team.
Team: the name of the Premier league team
Nationality: the nationality of the manager
Status: Status of the manager (Active or Sacked)

Query creation: When creating the SQL SELECT statements, make sure you do not design queries that are dependent on the data stored in the database or that only work for this specific instance of the schema. **Also, please provide the question number in comments before the solution.**
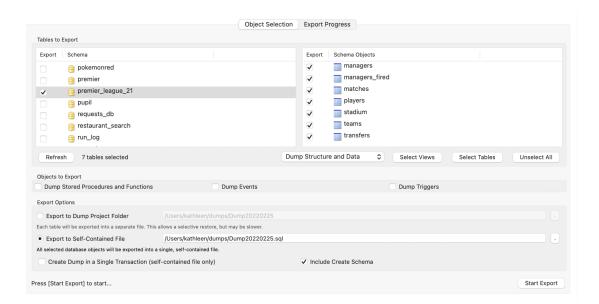

Assignment Description

1. Create a database for your schema named *premierlastnamefirstnameinitial*. You may create the tables' schemas using the SQL CREATE command or using the EER modeling tool. (15 points)
2. Create a diagram for your Premier Football League database (*premierlastnamefirstnameinitial)* by using the "Reverse Engineer" command. This can be done in the modeling tool found in the MySQL workbench. Create a screenshot of the diagram or create a pdf from the EER model,. (10 points)
3. Import the .csv files into your tables using the Import table wizard or by writing INSERT commands for the data.(5 points)

Compose queries to answer the following questions

4. Generate a list of matches for match day 1, in which the home team won. The result should contain the match number, home team and the away team name. (5 points)
5. Which teams had more than one manager in the season? The result should contain the team name and the number of managers. (5 points)
6. Which manager/managers worked for more than one team? The result should contain the manager name and number of teams. (5 points)
7. Generate a result that contains managers, teams and the number of goals scored by the team in the home stadium for each team for this season. Consider only the active managers. The list should be in descending order of number of goals. The result should contain the manager's name, the team name, and the number of goals. (5 points)
8. Generate a list consisting of a manager's name, total number of matches won by the manager in the season. The list should be in descending order of number of matches. Consider only the active managers. (5 points)
9. Determine the stadium, where the most number of goals were scored. The result should only contain the stadium name. (5 points)
10. Determine the number of matches ended as a draw per team. The result should contain the team name and the number of matches. (5 points)

11. Clean sheets means that the team did not allow an opponent to score a goal in the match. Determine the top 5 teams ranked by number of clean sheets in the season. The result should contain the team's name and the count of the clean sheets. The result should be ordered in descending order by the count of clean sheets. (5 points)

12. Generate a list of matches played between Christmas and 3rd January where the home team scored 3 or more goals. Consider the date range 25th December to 3rd January(Including). Display all the fields for the match. (5 points)

13. Generate the list of all the matches, where a team came back from losing the game at the end of the first half of the game , to winning at the end of the second half. The result should contain all the columns in the match tuple. (5 points)

14. Determine the top 5 teams by the number of matches won by the teams. The result should just contain team names. (5 points)

15. Write a query that computes for each team the average number of goals conceded at home, the average number of goals scored at home, the average number of goals scored away from home and the average number of goals conceded away from home. (5 points)

16. Generate a self-contained extract of your database to canvas using the 'data export' tool from the 'Server' menu. Make sure you include the create schema as well as other objects in the database. The file should be a self-contained file. We must be able to import your schema so please ensure the extracted file works with import. Below is the panel for export. (10)

Assignment submission

Please submit the following 3 files :

- hwk6EERlastnamefi.pdf (picture of the model)
- hwk6footballdumplastnamefi.sql (export of the created schema)
- Hwk6footballquerieslastnamefi.sql (SQL SELECT commands for the queries)

Submit the multiple files to canvas.