

Data Engineering Career Track

Advanced SQL Mini-Project

Euro Cup Data Analysis

Estimated Time: 3-5 hours



In this project, you'll look at Euro Cup 2016 data and write queries to answer our questions.

The data can be downloaded [here](#). Please extract the ZIP archive downloaded to a folder and import data contained in the CSV files to a database using the following steps:

1. Create a new database named euro_cup_2016 :

```
mysql> CREATE SCHEMA euro_cup_2016;
```

2. Follow the instructions as given in the [documentation](#) to import the CSV files from the ZIP archive into the database.

Please use the database name that you created above (when in step as shown in figure Figure 6.16 in the documentation).

Please configure the table names same as the CSV file names.

The data model for the data can be found [here](#). In order to understand the content of the tables and what data each column contains, please refer to [this](#) document.

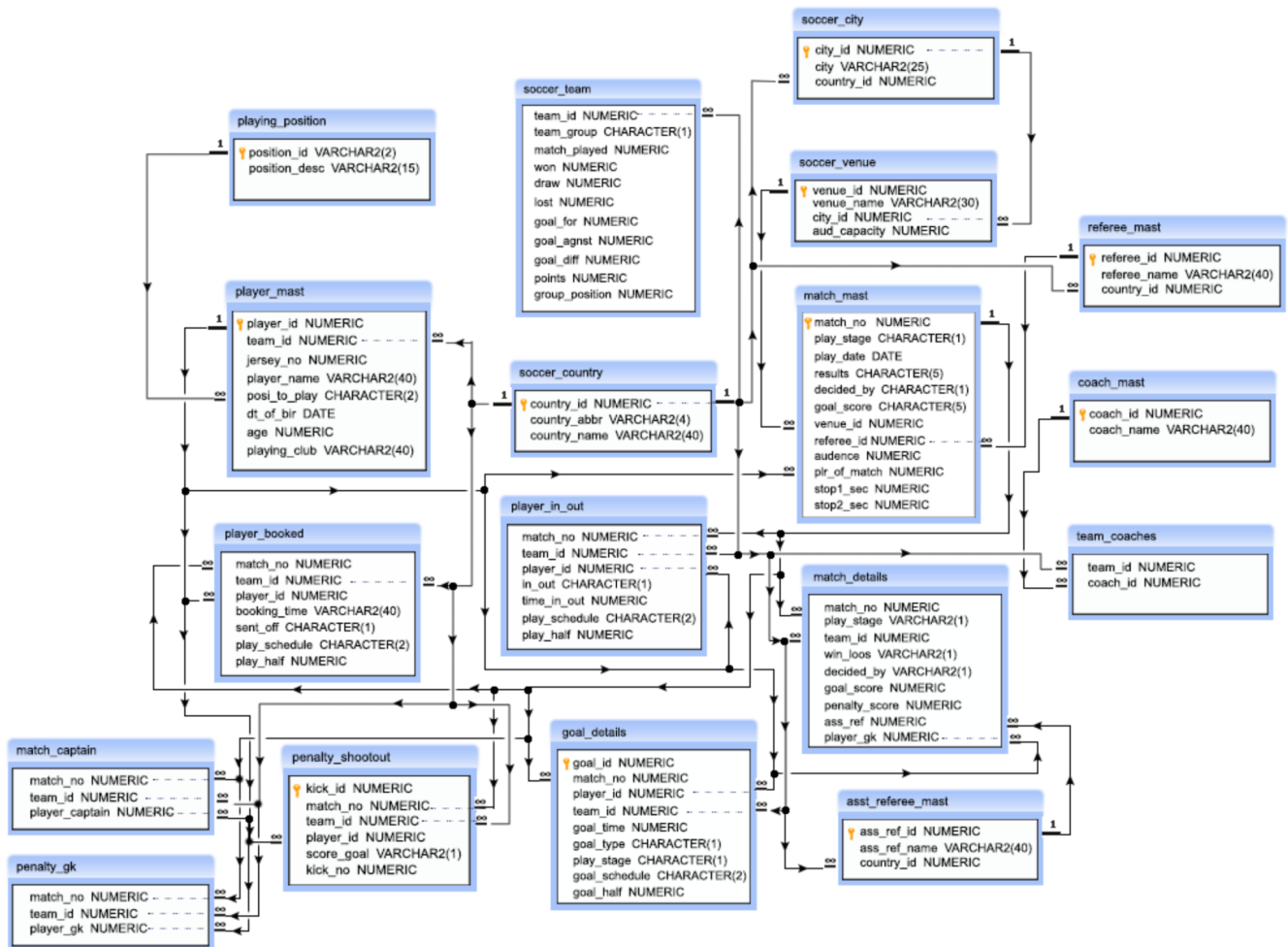
You'll find the queries to write on the next page.

Please answer write the following queries:

1. Write a SQL query to find the date EURO Cup 2016 started on.
2. Write a SQL query to find the number of matches that were won by penalty shootout.
3. Write a SQL query to find the match number, date, and score for matches in which no stoppage time was added in the 1st half.
4. Write a SQL query to compute a list showing the number of substitutions that happened in various stages of play for the entire tournament.
5. Write a SQL query to find the number of bookings that happened in stoppage time.
6. Write a SQL query to find the number of matches that were won by a single point, but do not include matches decided by penalty shootout.
7. Write a SQL query to find all the venues where matches with penalty shootouts were played.
8. Write a SQL query to find the match number for the game with the highest number of penalty shots, and which countries played that match.
9. Write a SQL query to find the goalkeeper's name and jersey number, playing for Germany, who played in Germany's group stage matches.
10. Write a SQL query to find all available information about the players under contract to Liverpool F.C. playing for England in EURO Cup 2016.
11. Write a SQL query to find the players, their jersey number, and playing club who were the goalkeepers for England in EURO Cup 2016.
12. Write a SQL query that returns the total number of goals scored by each position on each country's team. Do not include positions which scored no goals.
13. Write a SQL query to find all the defenders who scored a goal for their teams.
14. Write a SQL query to find referees and the number of bookings they made for the entire tournament. Sort your answer by the number of bookings in descending order.
15. Write a SQL query to find the referees who booked the most number of players.
16. Write a SQL query to find referees and the number of matches they worked in each venue.
17. Write a SQL query to find the country where the most assistant referees come from, and the count of the assistant referees.
18. Write a SQL query to find the highest number of foul cards given in one match.
19. Write a SQL query to find the number of captains who were also goalkeepers.
20. Write a SQL query to find the substitute players who came into the field in the first half of play, within a normal play schedule.

Deliverables:

1. Individual SQL script for solution to each question (SQL file naming convention is sql_q<question number>_sol.sql e.g.- sql_q1_sol.sql, sql_q2_sol.sql, etc.)



Description of tables:

soccer_country:

- country_id – this is a unique ID for each country
- country_abbr – this is the sort name of each country
- country_name – this is the name of each country

soccer_city:

- city_id – this is a unique ID for each city
- city – this is the name of the city
- country_id – this is the ID of the country where the cities are located and only those countries will be available which are in soccer_country table

soccer_venue:

- venue_id – this is a unique ID for each venue
- venue_name – this is the name of the venue
- city_id – this is the ID of the city where the venue is located and only those cities will be available which are in the soccer_city table
- aud_capacity – this is the capacity of audience for each venue

soccer_team:

- team_id – this is the ID for each team. Each teams are representing to a country which are referencing the country_id column of soccer_country table
- team_group – the name of the group in which the team belongs
- match_played – how many matches a team played in group stage
- won – how many matches a team won
- draw – how many matches a team draws
- lost – how many matches a team lose
- goal_for – how many goals a team conceded
- goal_agnst – how many goals a team scored

- goal_diff – the difference of goal scored and goal conceded
- points – how many points a team achieved from their group stage matches
- group_position – in which position a team finished their group stage matches

playing_position:

- position_id – this is a unique ID for each position where a player played
- position_desc – this is the name of the position where a player played

player_mast:

- player_id – this is a unique ID for each player
- team_id – this is the team where a player played, and only those teams which referencing the country_id column of the table soccer_country
- jersey_no – the number which labeled on the jersey for each player
- player_name – name of the player
- posi_to_play – the position where a player played, and the positions are referencing the position_id column of playing_position table
- dt_of_bir – date of birth of each player
- age – approximate age at the time of playing the tournament
- playing_club – the name of the club for which a player was playing at the time of the tournament

referee_mast:

- referee_id – this is the unique ID for each referee
- referee_name – name of the referee
- country_id – the country, where a referee belongs and the countries are those which referencing the country_id column of soccer_country table

match_mast:

- match_no – this if the unique ID for a match

- play_stage – this indicates that in which stage a match is going on, i.e. G for Group stage, R for Round of 16 stage, Q for Quarter final stage, S for Semi Final stage, and F for Final
- play_date – date of the match played
- results – the result of the match, either win or draw
- decided_by – how the result of the match has been decided, either N for by normally or P for by penalty shootout
- goal_score – score for a match
- venue_id – the venue where the match played and the venue will be one of the venue referencing the venue_id column of soccer_venue table
- referee_id – ID of the referee who is selected for the match which referencing the referee_id column of referee_mast table
- audience – number of audience appears to watch the match
- plr_of_match – this is the player who awarded the player of a particular match and who is selected a 23 men playing squad for a team which referencing the player_id column of player_mast table
- stop1_sec – how many stoppage time (in second) have been added for the 1st half of play
- stop2_sec – how many stoppage time (in second) have been added for the 2nd half of play

coach_mast:

- coach_id – this is the unique ID for a coach
- coach_name – this is the name of the coach

asst_referee_mast:

- ass_ref_id – this is the unique ID for each referee assists the main referee
- ass_ref_name – name of the assistant referee
- country_id – the country where an assistant referee belongs and the countries are those which are referencing the country_id column of soccer_country table

match_details:

- match_no – number of the match which is referencing the match_no column of match_mast table
- play_stage - stage of the match, i.e. G for group stage, R for Round of 16, Q for Quarter Final, S for Semi final and F for final
- team_id – the team which is one of the playing team and it is referencing the country_id column of soccer_country table
- win_lose – team either win or lose or drawn indicated by the character W, L, or D
- decided_by - how the result achieved by the team, indicated N for normal score or P for penalty shootout
- goal_score – how many goal scored by the team
- penalty_score – how many goal scored by the team in penalty shootout
- ass_ref – the assistant referee assist the referee which are referencing the ass_ref_id column of asst_referee_mast table
- player_gk - the player who is keeping the goal for the team, is referencing the player_id column of player_mast table

goal_details:

- goal_id – this is the unique ID for each goal
- match_no – this is match_no which is referencing the match_no column of match_mast table
- player_id - this is the ID of a player who is selected for the 23 men squad of a team for the tournament and which is referencing the player_id column of player_mast table
- team_id – this is the ID of each team who are playing in the tournament and referencing the country_id column of soccer_country table
- goal_time – this is the time when the goal scored
- goal_type – this is the type of goal which came in normally indicated by N or own goal indicating by O and goal came from penalty indicated by P
- play_stage – this is the play stage in which goal scored, indicated by G for group stage, R for round of 16 stage, Q for quarter final stage, S for semifinal stage and F for final match

- goal_schedule – when the goal came, is it normal play session indicated by NT or in stoppage time indicated by ST or in extra time indicated by ET
- goal_half – in which half of match goal came

penalty_shootout:

- kick_id – this is unique ID for each penalty kick
- match_no - this is the match_no which is referencing the match_no column of match_mast table
- team_id – this is the ID of each team who is playing in the tournament and referencing the country_id column of soccer_country table
- player_id - this is the ID of a player who is selected for the 23 men squad of a team for the tournament and which is referencing the player_id column of player_mast table
- score_goal – this is the flag Y if able to score the goal or N when not
- kick_no – this is the kick number for the kick of an individual match

player_booked:

- match_no - this is the match_no which is referencing the match_no column of match_mast table
- team_id – this is the ID of each team who are playing in the tournament and referencing the country_id column of soccer_country table
- player_id - this is the ID of a player who is selected for the 23 men squad of a team for the tournament and which is referencing the player_id column of player_mast table
- booking_time – this is the time when a player booked
- sent_off – this is the flag Y when a player sent off
- play_schedule – when a player booked, is it in normal play session indicated by NT or in stoppage time indicated by ST or in extra time indicated by ET
- play_half – in which half a player booked

player_in_out:

- match_no - this is the match_no which is referencing the match_no column of match_mast table

- team_id – this is the ID of each team who are playing in the tournament and referencing the country_id column of soccer_country table
- player_id - this is the ID of a player who is selected for the 23 men squad of a team for the tournament and which is referencing the player_id column of player_mast table
- in_out – this is the flag I when a player came into the field or O when go out from the field
- time_in_out – when a player come into the field or go out from the field
- play_schedule – when a player come in or go out of the field, is it in normal play session indicated by NT or in stoppage time indicated by ST or in extra time indicated by ET
- play_half - in which half a player come in or go out

match_captain:

- match_no - this is the match_no which is referencing the match_no column of match_mast table
- team_id – this is the ID of each team who are playing in the tournament and referencing the country_id column of soccer_country table
- player_captain - the player who represents as a captain for a team, is referencing the player_id column of player_mast table

team_coaches:

- team_id – this is the ID of a team who is playing in the tournament and referencing the country_id column of soccer_country table
- coach_id – a team may be one or more coaches, this indicates the coach(s) who is/are coaching the team is referencing the coach_id column of coach_mast table

penalty_gk:

- match_no - this is the match_no which is referencing the match_no column of match_mast table
- team_id – this is the ID of each team who are playing in the tournament and referencing the country_id column of soccer_country table
- player_gk - the player who kept goal at the time of penalty shootout, is referencing the player_id column of player_mast table