Database Systems – ISYS2014/ISYS5008 2023 Semester 2

Final Assessment

Database Report

Dhilisha Flavio Ranatunga

20908391

# Introduction

This report describes the design and development of a database system in the context of the FIFA Women's World Cup. The project required an elaborate approach that included ER modeling, relational mapping, and the building of data description tables. Data was obtained from web platforms, CSV files were used, and queries were written to retrieve useful information. The system now includes advanced features like as stored procedures, views, and triggers. A link was also established between the database and a Python3 implementation. This report provides a detailed explanation of the project's techniques and outcomes.

# DATABASE DESIGN

The database is designed to hold essential FIFA Women's World Cup information. It contains the following entities:

**World Cup**: Key tournament information (year, host, teams, champions, and so forth).

**Location**: This table contains information about tournament venues, stadiums, and seat capacity.

**Group Stage**: Stores information on group stages that is relevant to the World Cup table and the performance of the team in the group stage.

**Team**: Individual team data, with links to the World Cup and Group Stage tables.

**Player**: Information on players associated with their respective teams.

**Matches**: Detailed match records linked with location and World Cup year.

**Plays In**: Manages many-to-many team participation in matches.

**Awards**: This table keeps track of player honors that are tied to the Player and World Cup year.

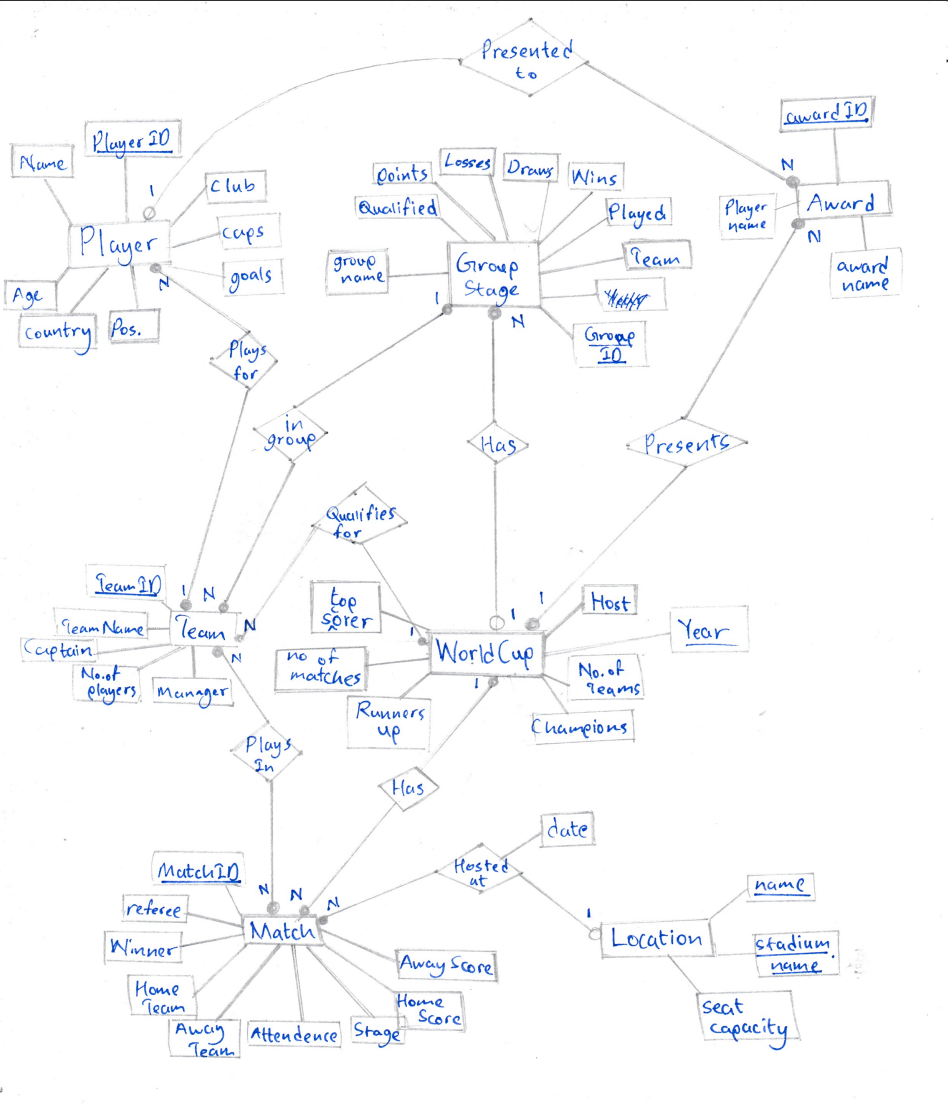
## ER Modelling

|  |  |  |
| --- | --- | --- |
| **Entity Sets** | **Keys** | **Other Attributes** |
| World Cup | Year | Host, Teams No., Champions, Runners Up, Top Scorer, Matches No. |
| Match | Match ID | Home team, Away team, Attendance, Stage, Home score, Away score, winner, referee |
| Team | Team ID | Team Name, No. of players, Manager, Captain |
| Player | Player ID | Name, Age, Country, Pos., Club, Caps, goals |
| Location | Name, Stadium name | Stadium Name, Seat capacity |
| Group Stage | Group ID | Group, Team, Played, Wins, Draws, Loses, Points, Qualified |
| Awards | AwardID | award name, player name |

|  |  |  |
| --- | --- | --- |
| **Relationship sets** | **Participation between sets** | **Attributes of relationship set** |
| Qualifies for | World Cup, Team |  |
| Has | World Cup, Match |  |
| Have | World Cup, Group stage |  |
| Plays in | Match, Team |  |
| Hosted at | Match, Location | Date |
| Plays for | Team, Player |  |
| In group | Team, Group Stage |  |
| Presented to | Award, Player |  |
| Presents | World Cup, Awards |  |

|  |  |  |
| --- | --- | --- |
| **Relationship sets** | **Cardinality Constraints** | **Participation/other constraints** |
| Qualifies for | One – many (World cup is played by many teams, Squad qualifies for one world cup) | World Cup – Total, Team – Total (World cup have at least one team, Teams cannot exist without world cup) |
| Has | One – many  (World cup has many matches, A match is played for one worldcup) | World Cup – Total, Match – Total (Worldcup cannot exist without a match, match cannot exist without a worldcup |
| Have | One - many | WorldCup – partial Group stage – Total (A world cup may exist without a groups stage format; groups may not exist without a world cup |
| Plays in | Many – many  (Match is played by two teams, Team plays many matches) | Match – Total, Team – Total  {Match must be played by teams; Team should play in the match) |
| Hosted at | One – many  (One location may host many matches, Match is held at one location) | Match – Total, Location – Partial (A location may not host any matches, match must be played at a location) |
| Plays for | One – many | Team – Total, Player – Total (Team should have players, Player should be in a team) |
| In group | One - many | Group – Total, Team – Total  (Group should consist of teams, Teams should have a group) |
| Presented to | One - many | Award – Total, Player – Partial (Award cannot exist without player, Player may not receive an award) |
| Presents | One - many | WorldCup – Total, Awards -Total (World cup typically presents awards to players and teams in general, Awards cannot exist without a world cup) |

## ER DIAGRAM



## Relational Schema

I've arranged the FIFA Women's World Cup data into a series of related tables in this schema. Foreign keys (FKs) establish relationships between entities, and each table represents a separate entity.

### Step 1: Mapping entities

**World Cup** (year, host, no. of teams, champions, runners up, top scorer, no. of matches)

**Match** (matchid, home team, away team, venue, attendance, stage, home score, away score, winner, referee)

**Team** (teamID, team name, no. of players, manager, captain)

**Player** (playerID., name, age, country, pos., club, caps, goals)

**Location** (name, stadium name, seat capacity)

**Group Stage** (groupID, group, team, played, wins, draws, loses, points, qualified)

**Awards** (awardID, award name, player name)

### Step 2: Mapping 1: N relationships

**World Cup** (year, host, no. of teams, champions, runners up, top scorer, no. of matches)

**Group Stage** (groupID, year, group, team, played, wins, draws, loses, points, qualified)

FK year REF WorldCup(year)

**Team** (teamID, year, groupID team name, no. of players, manager, captain)

FK year REF WorldCup(year),

FK groupID REF Group Stage(groupID)

**Match** (matchid, year, home team, away team, venue, attendance, stage, home score, away score, winner, player referee)

FK year REF WorldCup(year),

FK venue REF Location(name)

**Player** (playerID, teamID, name, age, country, pos., club, caps, goals)

FK teamID REF Team(teamID),

**Awards** (awardID, year, award name, playerID, player name)

FK playerID REF Player(playerID)

FK year REF WorldCup(year)

### Step 3: Each M:N relationship type become a separate relation.

**Match** (matchid, home team, away team, venue, attendance, stage, home score, away score, winner, referee, venue, date)

**Team** (teamID, year, groupID team name, no. of players, manager, captain)

PlaysIn (teamID, matchID)

FK teamID REF Team(teamID),

FK matchID REF Match(matchID)

## Data Descriptions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **World Cup** |  | | | | | |
| Description: | Consist of data about the world cup results | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| year | int |  | No | Yes | Year the world cup was held |  |
| host | varchar | 25 | No |  | Host country |  |
| no\_of\_team | int |  | No |  | Total number of teams playing |  |
| champions | varchar | 25 | No |  | Winner team of world cup |  |
| Runners\_up | varchar | 25 | No |  | Second of world cup |  |
| Top\_scorer | varchar | 25 |  |  | Top scorer of the world cup |  |
| No\_of\_matches | int |  |  |  | Total no. of matches |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Match** |  | | | | | |
| Description: | Has data about match statistics played between two teams | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| matchID | char | 8 | No | Yes | Match key |  |
| year | int |  | No |  | Year the world cup was held | FK WorldCup(year) |
| homeTeam | varchar | 25 | No |  | Team one |  |
| awayTeam | varchar | 25 | No |  | Team two |  |
| venue | varchar | 50 | No |  | Location the match is played |  |
| attendance | int |  |  |  | No of people attended for the match |  |
| stage | varchar | 25 | No |  | Match round (quarter final, semi final) |  |
| homeScore | int |  | No |  | Home team score |  |
| awayScore | int |  | No |  | Away team score |  |
| winner | varchar | 25 | No |  | Winner team of the match |  |
| referee | varchar | 50 |  |  | Name of referee |  |
| date | date |  |  |  | Date match was held |  |
| Venue (FK) | varchar | 50 | No |  | Foreign key from location entity | References Location(name) |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Location** |  | | | | | |
| Description: | Consist of information about the location and it’s stadium | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| name | varchar | 25 | No | Yes | Name of location |  |
| stadium\_name | varchar | 25 | No | Yes | Name of stadium |  |
| Seat\_capacity | int |  |  |  | Seating capacity of stadium |  |
| **Team** |  | | | | | |
| Description: | Keeps data about the team | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| teamID | char | 8 | No | Yes | Team key |  |
| year | int |  | No |  | World cup year | FK WorldCup(year) |
| groupID | int |  | No |  | Group key of team group | FK GroupStage(groupID) |
| team\_name | varchar | 25 | No |  | Name of team |  |
| no\_of\_players | int |  |  |  | Number of players in the team |  |
| manager | varchar | 50 |  |  | Name of the team manager |  |
| captain | varchar | 50 |  |  | Name of the team captain |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group stage** |  | | | | | |
| Description: | Contains data about the group | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| groupID | char | 4 | No | Yes | Group key |  |
| year | int |  | No |  | World cup year | FK WorldCup(year) |
| group | varchar |  | No |  | Group (etc. A, B , C..) |  |
| team | varchar | 25 | No |  | Name of team |  |
| played | int |  | No |  | Number of matches played |  |
| wins | int |  | No |  | Number of matches won |  |
| draws | int |  | No |  | Number of matches drawn |  |
| loses | int |  | No |  | Number of matches lost |  |
| points | int |  | No |  | Total points |  |
| qualified | varchar | 1 | No |  | Qualified or not |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Player** |  | | | | | |
| Description: | Contains data about the player | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| playerID | int |  | No | Yes | Player identifier |  |
| teamID | int |  | No |  | Foreign key referenced from the team entity | References Team(teamID) |
| playerNo | int |  | No |  | Player shirt number |  |
| name | varchar | 50 | No |  | Player name |  |
| age | int |  | No |  | Player age |  |
| country | varchar | 50 | No |  | Player’s nationality |  |
| position | varchar | 25 | No |  | Position played by the player |  |
| club | varchar | 50 |  |  | Players current club |  |
| caps | int |  |  |  | Matches played by the player |  |
| goals | int |  |  |  | Goals scored by the player |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Plays in** |  | | | | | |
| Description: | Contains data match played by two teams | | | | | |
| **Attribute** | **type** | **size** | **Null** | **Primary key** | **description** | **Other constraints** |
| matchID | char | 8 | No | Yes | Match identifier |  |
| homeTeamID | char | 8 | No | Yes | Team key | References Team (teamID) |

|  |  |
| --- | --- |
| **Business rules** | **Description** |
| BR1 | A team should not have more than 23 players per squad |
| BR2 | A player is not allowed to participate in more than one World Cup squad within a given year |
| BR3 | A specific location/stadium is allowed to host a maximum of one FIFA Women's World Cup match per day |

## Assumptions

* The database only considers teams which are only qualified for the specific world cup, players nor teams’ data who are not participating in the world cup are not included.
* The database only consists of data only relevant to the 2019 FIFA Women’s World Cup, this database solely focuses on the 2019 world cup.
* It specifically contains the data of the players as of the year 2019

# Designing and implementing queries

The MySQL is accessed using the terminal using the username and password,

mysql --local-infile=1 -u root -p

The feature which is used to import data from local csv files should be enabled.

The process begins with the creation of database "FIFAWWorldCup\_20908391." This database served as a container for the tables and information related to the 2019 Women's World Cup. This is done using the SOURCE worldcup\_tables.sql command.

A screen shot of a computer program

Description automatically generated

The 2019 FIFA Women's World Cup data were loaded to the relevant tables using local csv files. To record and arrange the tournament data, several tables were made, including "WorldCup," "Location," "GroupStage," "Team," "Matches," "Player," "PlaysIn," and "Awards." To make data analysis easier, stored methods like "GetPlayersByTeam," "AvgGoalsPerTeam," and "CalculateWinPercentage" were included in the “worldcup\_values.sql” script file. To ensure effective data management and analysis, a view called "MatchResultsView" was created to quicken the retrieval of match-related data from various tables.

LOAD DATA LOCAL INFILE '/home/flavio/DBS/FinalAssesment/Locations.csv'

INTO TABLE Location

FIELDS TERMINATED BY ','

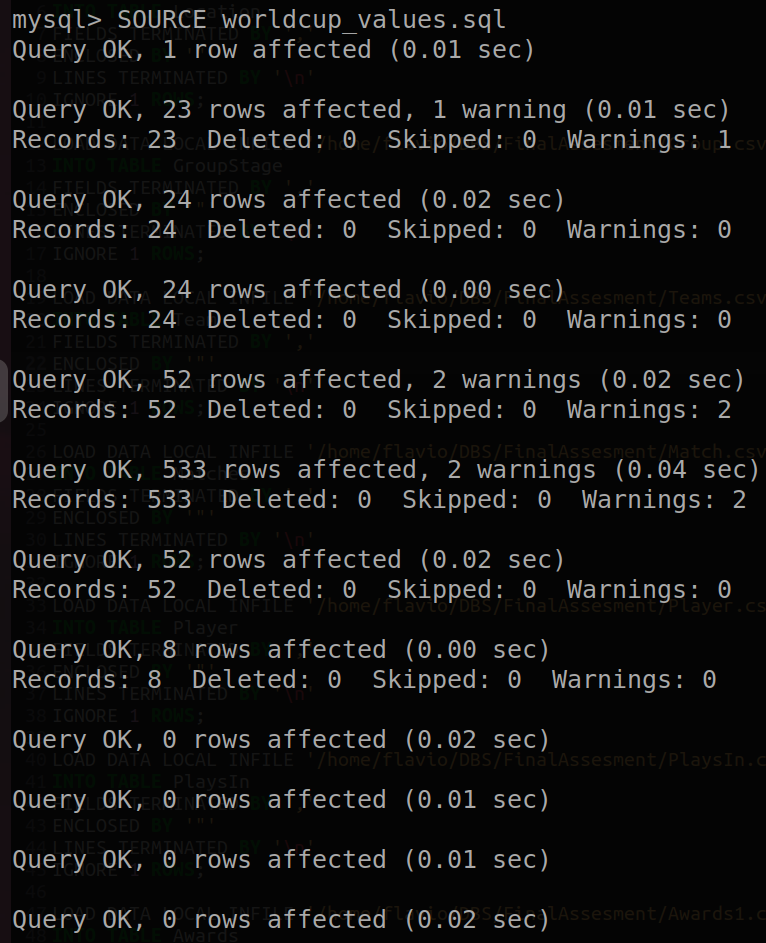
ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

The MySQL statement used to load sample location data from a local csv file is as follows. The structure of the data to be entered is specified using the INTO TABLE Location database. Additional requirements include comma separation, double quote marks, and a new line character at the end of each piece of information. The first row of the table, which contains the column names, is likewise skipped.

The database is compiled with information regarding the 2019 FIFA Women's World Cup from several trustworthy sources, such as Wikipedia, GitHub repositories, and Kaggle. Following that, I prepared a customized CSV file, manually choosing and structuring the relevant data to facilitate fast access and readability for my purpose.



# Application

## Query Design and Implementation

To retrieve vital data from a database, queries are essential. They assist in providing precise answers, resolving issues, and achieving particular objectives. They collect information that is useful for decision-making, statistics, and historical context. This section will demonstrate example results obtained from a series of deliberately run database queries that were created to extract relevant information regarding the 2019 FIFA Women's World Cup.

## Level 1

1. Get players who have scored more than 50 goals in their career, displays player’s information. This question helps in identifying the world cup's top performers to keep an eye out for. These athletes might be regarded as players who influence matches.

mysql> SELECT \* FROM Player WHERE goals > 50;

A screenshot of a computer screen

Description automatically generated

1. This query will return the names of teams that have qualified for the next stage ('Q') in their respective groups. Helps to identify the team’s progress and their ability by comparing with other world cup appearances.

A screen shot of a computer

Description automatically generatedmysql> SELECT group\_name, team FROM GroupStage WHERE qualified = 'Q';

1. The query shows a summary of each game, including the teams, results, and winner using string concatenation. For easier readability of needed data regarding the match.

A screen shot of a computer

Description automatically generatedmysql> SELECT CONCAT(homeTeam, ' vs ', awayTeam) AS Matchs, CONCAT(homeScore, ' - ', awayScore) AS Scores, winner AS Winner FROM Matches;

1. Calculate the difference in the number of days between the first match of the knockout stage (round of 16) and the final match. Indicates how many days the knockout stages of a world cup would generally last, helps people to plan out their stay and arrangements throughout the world cup.

mysql> SELECT DATEDIFF( (SELECT MAX(m\_date) FROM Matches), (SELECT MIN(m\_date) FROM Matches WHERE stage = 'Round of 16') ) AS days\_between\_first\_and\_final\_match;

A black screen with white text

Description automatically generated

## Level 2

1. This query retrieves the team names, team IDs, and their respective group names for teams that have not advanced to the next stage, as indicated by the 'N' qualification status in the "GroupStage" table. Specifies which teams will not be participating in the knockout bracket.

SELECT G.group\_name, T.team\_name FROM Team T LEFT JOIN GroupStage G ON T.team\_name = G.team WHERE G.qualified = 'N' ORDER BY G.group\_name;



1. Retrieving the data about award winners of the world cup with their respective country, club and goals, It could be a valuable insight to a football fan whose searching for standout players show casing their performance.

SELECT DISTINCT P.name, P.country, P.club, P.goals

FROM Player AS P

WHERE P.playerID IN (

SELECT A.playerID

FROM Awards AS A

)

ORDER BY P.goals DESC;

A screen shot of a computer

Description automatically generated

1. The query retrieves comprehensive information about the venues where the matches were hosted, including details on total attendance throughout the tournament, stadium names, and seating capacities, by seamlessly combining data from both the Match and Location tables. Also showcases the most visited stadiums throughout the tournament.

mysql> SELECT M.venue, L.stadium\_name, L.seat\_capacity, M.total\_attendance FROM (SELECT venue, SUM(attendance) AS total\_attendance FROM Matches GROUP BY venue) AS M JOIN Location AS L ON M.venue = L.name ORDER BY M.total\_attendance DESC;

A screen shot of a computer

Description automatically generated

1. Retrieving top scoring teams in the 2019 fifa world cup group stage

This query offers insightful information about how each team performed in the group stage.

SELECT GS.year, GS.team,

SUM(IF(GS.team = M.homeTeam, M.homeScore, IF(GS.team = M.awayTeam, M.awayScore, 0))) AS total\_goals\_scored

FROM GroupStage GS

JOIN Matches M ON GS.year = M.year AND (GS.team = M.homeTeam OR GS.team = M.awayTeam)

WHERE GS.year = 2019

GROUP BY GS.year, GS.team

ORDER BY total\_goals\_scored DESC;

A screen shot of a computer

Description automatically generated

## Design and Implementation of Advanced Features

## Stored Procedures

* **Retrieve Players and their details by National Team in the world cup.** Gives insightful information on player statistics and team rosters.

DELIMITER //

CREATE PROCEDURE GetPlayersByTeam(IN countryPar VARCHAR(50))

BEGIN

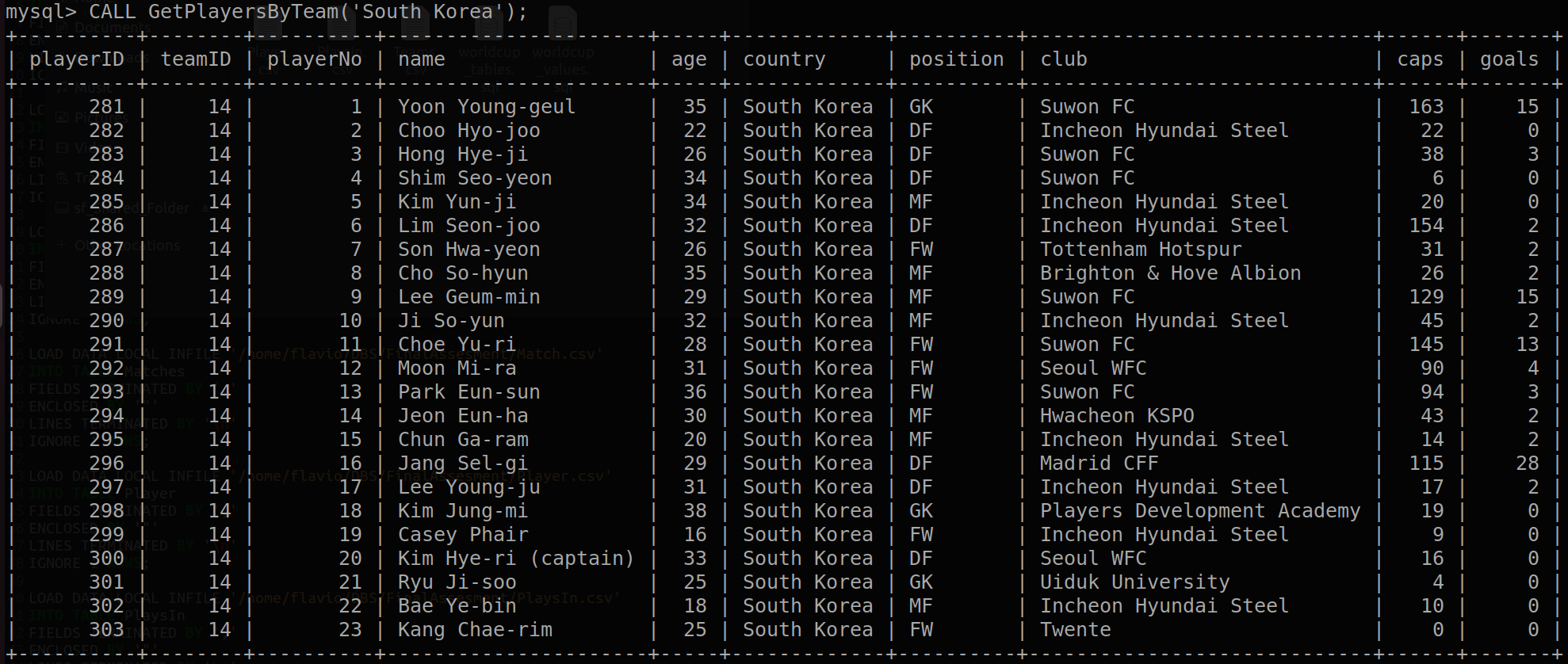
SELECT \* FROM Player

WHERE country = countryPar;

END //

DELIMITER ;

CALL GetPlayersByTeam('Brazil');



* **Average goals per match for each team.** It provides insight into which teams score more goals during a specific World Cup and can be utilized for tracking performance.

DELIMITER //

CREATE PROCEDURE AvgGoalsPerTeam(IN worldCupYear INT)

BEGIN

SELECT

T.team\_name AS TeamName,

SUM(IF(M.homeTeam = T.team\_name, M.homeScore, M.awayScore)) / COUNT(M.matchID) AS AverageGoalsPerMatch

FROM

Team AS T

LEFT JOIN Matches AS M ON T.team\_name = M.homeTeam OR T.team\_name = M.awayTeam

WHERE M.year = worldCupYear

GROUP BY T.team\_name;

END //

DELIMITER ;

A screen shot of a computer

Description automatically generated

* **Retrieving win percentage per team for a specific year,** shows the team's success rate at that particular World Cup.

DELIMITER //

CREATE PROCEDURE CalculateWinPercentage(IN teamName VARCHAR(255), IN worldCupYear INT)

BEGIN

DECLARE totalMatches INT;

DECLARE totalWins INT;

DECLARE winPercentage DECIMAL(5,2);

-- Calculate the total number of matches played by the team

SELECT COUNT(\*) INTO totalMatches

FROM Matches

WHERE (homeTeam = teamName OR awayTeam = teamName) AND year = worldCupYear;

-- Calculate the total number of matches won by the team

SELECT COUNT(\*) INTO totalWins

FROM Matches

WHERE ((homeTeam = teamName AND homeScore > awayScore) OR (awayTeam = teamName AND awayScore > homeScore)) AND year = worldCupYear;

-- Calculate the win percentage

IF totalMatches > 0 THEN

SET winPercentage = (totalWins / totalMatches) \* 100;

ELSE

SET winPercentage = 0;

END IF;

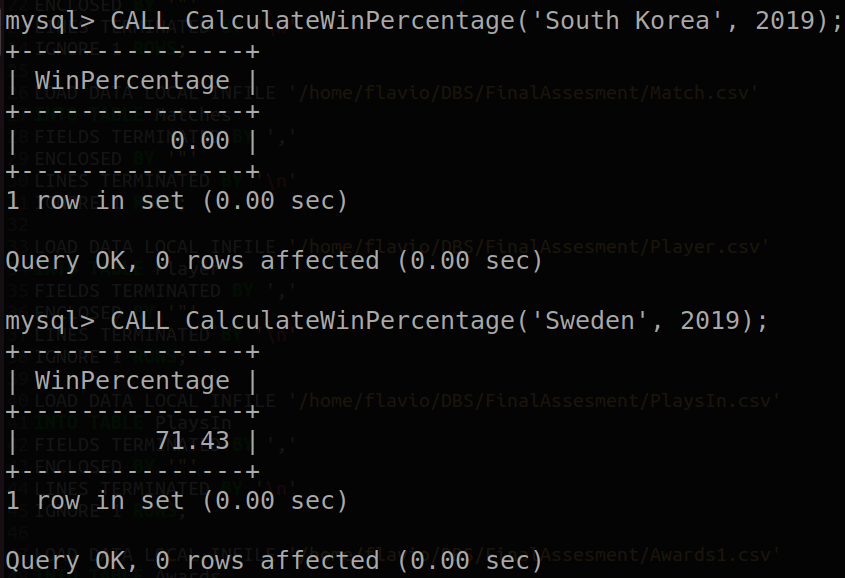
-- Return the win percentage

SELECT winPercentage AS WinPercentage;

END //

DELIMITER ;

CALL CalculateWinPercentage('Brazil', 2019)



## Views

* **Summarized match results view of the tournament**, provides a concise and convenient source of information regarding match outcomes, enabling comparisons, analysis, and the discovery of insights into the results of the competition.

CREATE VIEW MatchResultsView AS

SELECT

M.matchID,

T1.team\_name AS HomeTeam,

T2.team\_name AS AwayTeam,

M.homeScore AS HomeTeamScore,

M.awayScore AS AwayTeamScore,

L.stadium\_name AS Venue,

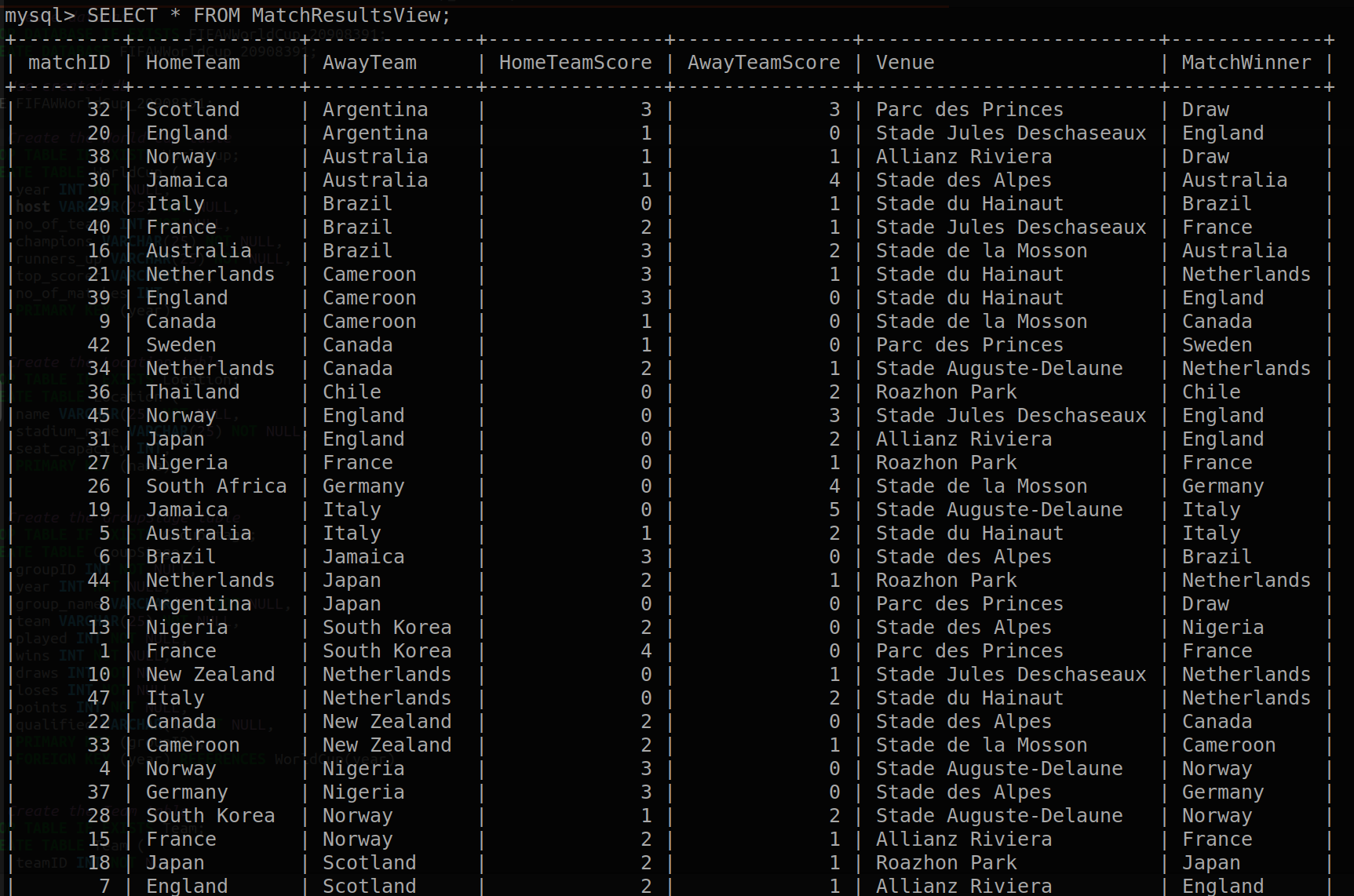
M.winner AS MatchWinner

FROM Matches AS M

JOIN Location AS L ON M.venue = L.name

JOIN Team AS T1 ON M.homeTeam = T1.team\_name

JOIN Team AS T2 ON M.awayTeam = T2.team\_name;



* **Team total goals view for specific worldcup**, helpful for evaluating team performance and locating the highest-scoring teams

CREATE VIEW TeamTotalGoals AS

SELECT t.year, t.team\_name, SUM(p.goals) AS total\_goals

FROM Team t

JOIN Player p ON t.teamID = p.teamID

GROUP BY t.year, t.team\_name;



A black background with white text

Description automatically generated

## Trigger

* **Trigger that would add the home team to the PlaysIn table automatically.**

DELIMITER //

CREATE TRIGGER AddHomeTeamToPlaysIn

AFTER INSERT ON Matches FOR EACH ROW

BEGIN

INSERT INTO PlaysIn (matchID, homeTeamID) VALUES (NEW.matchID, (SELECT teamID FROM Team WHERE team\_name = NEW.homeTeam AND year = NEW.year));

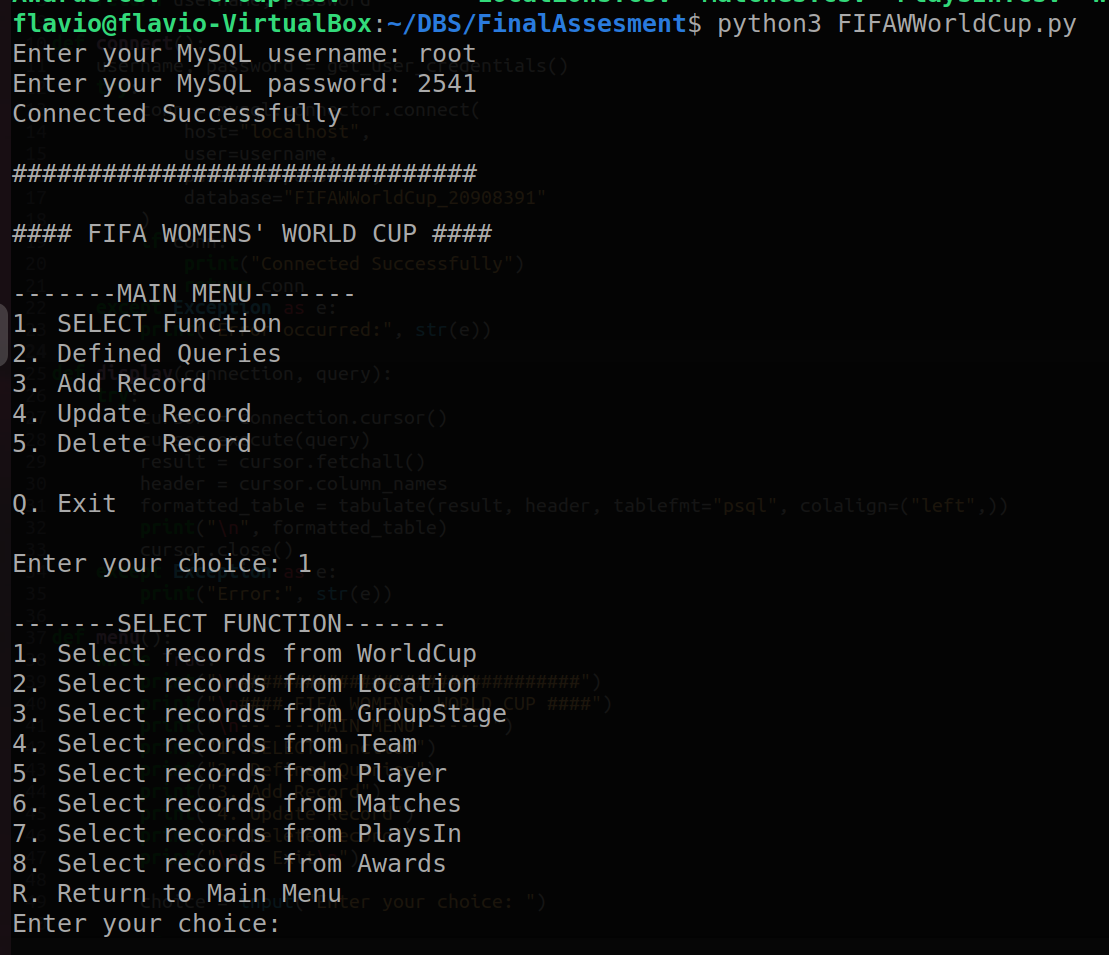
END;

//

DELIMITER ;

## DATABASE CONNECTIVITY AND PYTHON IMPLEMENTATION

By requesting user credentials, the Python software creates a connection to a MySQL database. It offers the ability to search for and show data in a tabular manner. Users can choose to run predefined queries or choose records from a variety of tables, and the script will gather the data and present it clearly in a tabular format.





# Reflection

In this project, I have achieved a deeper understanding of the importance of a well-structured database schema and the implementation process using local CSV files. I've gained valuable insights into the practical application of database systems, including the use of queries to extract meaningful information from real-world data. It took a lot of effort to collect data from various sources and make sure that it complied with the database schema. It was challenging to build the necessary links between entities since different datasets had varying tuples and data types. This work's present concentration on the 2019 FIFA Women's World Cup is a drawback. To facilitate more thorough historical analysis, it would be beneficial to expand the database to incorporate data from several years.

In conclusion, this project has offered a worthwhile learning opportunity in database design and implementation, highlighting the value of a well-structured schema and practical query programming for actual data analysis.

## References

* Wikipeida: 2019 FIFA Women's World Cup

<https://en.wikipedia.org/wiki/2019_FIFA_Women%27s_World_Cup>

* Kaggle: Football - FIFA Women's World Cup, 1991 – 2023

<https://www.kaggle.com/datasets/piterfm/football-fifa-womens-world-cup-1991-2023/data>

* The Fjelstul World Cup Database

https://github.com/jfjelstul/worldcup