

IE6700: Data Management for Analytics

Soccer Management System

Group 15

Ankita Nitin Bandal

Shaikh Zeeshan Ali

Mobile No.1: 508-762-7918

Mobile No.2: 857-693-9325

bandal.a@northeastern.edu

shaikh.zee@northeastern.edu

Percentage of Effort contributed by Student 1: 50%

Percentage of Effort contributed by Student 2: 50%

Signature of Student 1: Ankita Nitin Bandal

Signature of Student 2: Shaikh Zeeshan Ali

Submission Date: 02/12/2023

I. Introduction

A huge sports organization may find it difficult to manage a soccer sports database. The database may become out of date and contain duplicate or incorrect data. As a result, examining player data and team performance can be challenging, making informed judgments about future investments in players and teams difficult. To make informed decisions based on reliable information, the organization requires a solution that can accurately store and manage real-time data.

Business Problem

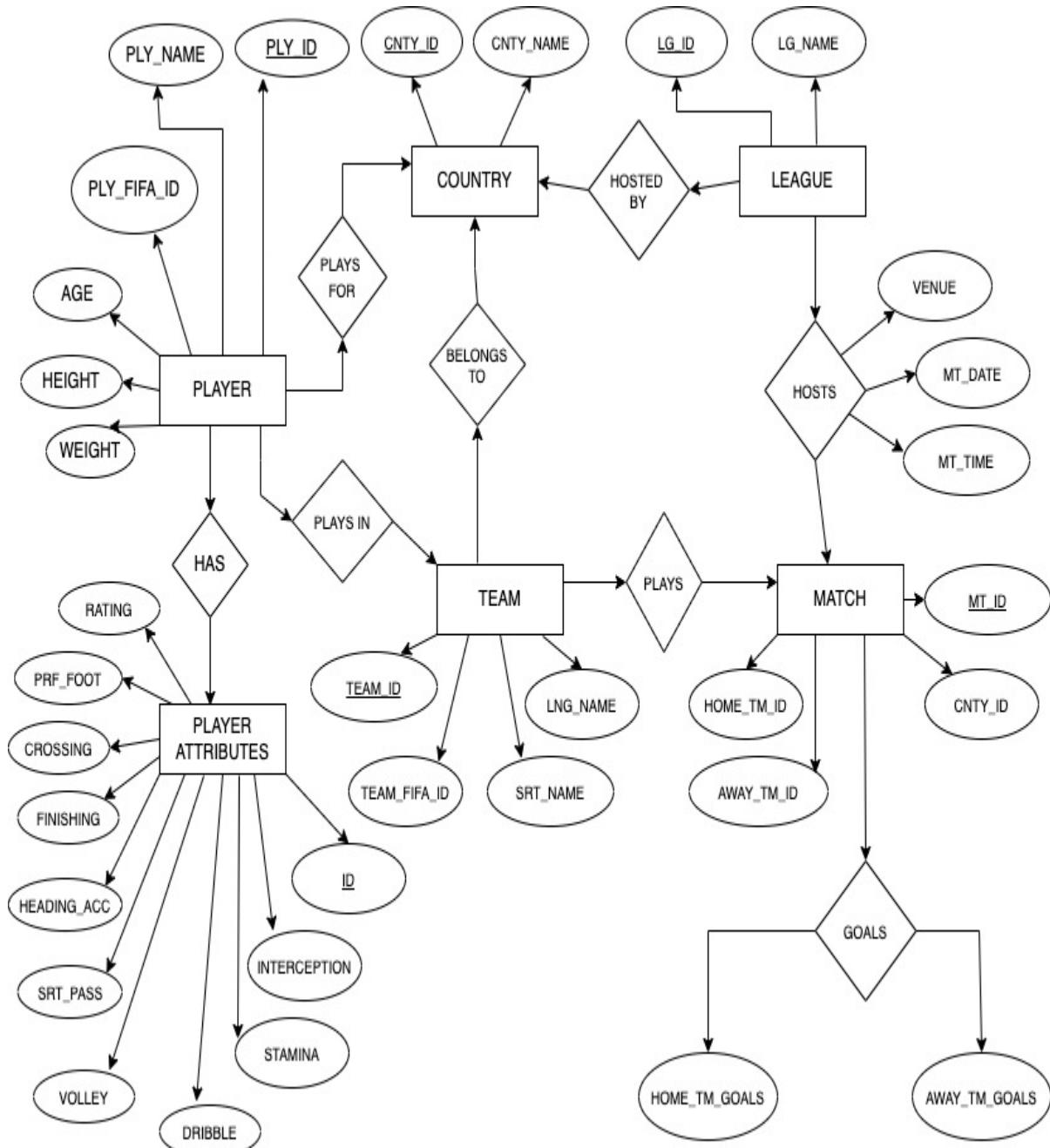
The present soccer sports database system lacks data analytic skills and is difficult to use. Because the firm is unable to generate practical insights from the data, its competitiveness suffers. To overcome this issue, the company needs a flexible database solution that authorized personnel can access from multiple locations. To eliminate lag time and mistakes in player statistics and team performance analysis, the system should give real-time updates. Integration with other systems is also required to eliminate manual data entry and reduce duplicative labor. The company seeks a user-friendly database solution that would allow for better decision-making and increased productivity. The new system should enable the business to analyze data more efficiently, give insights, and implement remedial actions more quickly.

Business Requirements

- **Scalable and accurate storage:** The database must store large amounts of data accurately, efficiently, and be adaptable to accommodate future growth.
- **User-friendly interface:** The database should provide simple tools for data analysis and reporting to enable informed decision-making.
- **Accessible and collaborative:** Authorized users must be able to access the database from different locations to ensure collaboration and efficiency.
- **Real-time updates:** The database should handle real-time updates and data entry to provide accurate and timely information for decision-making.
- **Seamless integration:** The database must integrate seamlessly with other systems to minimize manual data entry and reduce the possibility of errors.
- **Data security and consistency:** The database must ensure data accuracy, consistency, and security through robust data quality checks, backup and recovery processes, and user authentication and access controls.

II. Conceptual Data Modeling

Enhanced Entity Relationship Model.



UML Model



III. Mapping Conceptual Model to Relational Model

Primary Key: Bold Underlined Foreign Key: **Bold Italicized**

COUNTRY (**CNTY_ID**, CNTY_NAME, ***PLY_ID***)

TEAM (**TEAM_ID**, TEAM_FIFA_ID, SRT_NAME, LNG_NAME, **CNTY_ID**)

PLAYER (**PLY_ID**, PLY_NAME, PLY_FIFA_ID, AGE, HEIGHT, WEIGHT, **TEAM_ID**)

PLAYERATTRIBUTES (**ID**, INTERCEPTION, STAMINA, DRIBBLE, VOLLEY, SRT_PASS, HEADING_ACC, FINISHING, CROSSING, PRF_FOOT, RATING, **PLY_ID**)

MATCH (**MT_ID**, CNTY_ID, HOME_TM, AWAY_TM, **HOME_TM**, **AWAY_TM**)

LEAGUE (**LG_ID**, LG_NAME)

GOAL (HOME_TM_GOALS, AWAY_TM_GOALS, **MT_ID**)

IV. Implementation of Relational Model via MySQL & NoSQL

MySQL

Constructed the Soccer Management database and added tables to it by creating tables that included primary keys that uniquely identified the table, and foreign keys used to refer to the data in the referenced table.

```
USE `Soccer Management`;
CREATE TABLE country (
    country_id INT PRIMARY KEY,
    country_name VARCHAR(255) NOT NULL UNIQUE
);
CREATE TABLE team (
    team_id INT PRIMARY KEY,
    team_fifa_id INT NOT NULL UNIQUE,
    team_long_name VARCHAR(255) NOT NULL,
    team_short_name VARCHAR(255) NOT NULL,
    country_id INT NOT NULL,
    FOREIGN KEY (country_id) REFERENCES country(country_id)
);
CREATE TABLE player (
    player_id INT PRIMARY KEY,
    player_name VARCHAR(255) NOT NULL,
    player_fifa_id INT NOT NULL UNIQUE,
    Birthdate DATE NOT NULL,
    height INT NOT NULL,
    weight INT NOT NULL
);
CREATE TABLE player_attribute (
    attribute_id INT PRIMARY KEY,
    player_fifa_id INT NOT NULL,
    player_id INT NOT NULL,
    date DATE NOT NULL,
    overall_rating INT NOT NULL,
    preferred_foot VARCHAR(10) NOT NULL,
    crossing INT NOT NULL,
    finishing INT NOT NULL,
    heading_accuracy INT NOT NULL,
    short_passing INT NOT NULL,
    volleys INT NOT NULL,
    dribbling INT NOT NULL,
    stamina INT NOT NULL,
    strength INT NOT NULL,
    interceptions INT NOT NULL,
    FOREIGN KEY (player_fifa_id) REFERENCES player(player_fifa_id),
    FOREIGN KEY (player_id) REFERENCES player(player_id)
);
CREATE TABLE league (
    league_id INT PRIMARY KEY,
    country_id INT NOT NULL,
    league_name VARCHAR(255) NOT NULL,
    FOREIGN KEY (country_id) REFERENCES country(country_id)
);
CREATE TABLE Matchs (
    match_id INT PRIMARY KEY,
    match_date DATE NOT NULL,
    home_team_id INT NOT NULL,
    away_team_id INT NOT NULL,
    home_team_goals INT NOT NULL,
    away_team_goals INT NOT NULL,
    league_id INT NOT NULL,
    stage INT NOT NULL,
    FOREIGN KEY (home_team_id) REFERENCES team(team_id),
    FOREIGN KEY (away_team_id) REFERENCES team(team_id),
    FOREIGN KEY (league_id) REFERENCES league(league_id)
);
```

IMPORTING DATA INTO TABLES

Select * from country;
 Select * from team;
 Select * from Player
 Select * from player_attribute
 Select * from league
 Select * from Matchs

Queries

As report should not exceed more than 10 pages SQL queries screenshots are small, A file of all SQL outputs are submitted separately.

team_long_name	stage	total_goals
Melaga Cf	38	1
Rcd Espanyol	38	1
Re Recreativo	38	1
Real Sporting De Gij	38	1
Getafe Cf	38	1
Racing Santander	38	1
Sevilla Fc	38	1
Cd Numancia	38	1
Fc Barcelona	38	1
Rc Deportivo De La Coruna	38	1
Real Madrid Cf	38	1
Cd Osasuna	38	1
Villena Cf	38	1
Rcd Mallorca	38	1
Real Valladolid	38	1
Real Betis Balompi	38	1
Rc Deportivo De La Coruna	4	1
Cd Osasuna	4	1
Cd Numancia	4	1
Rcd Mallorca	4	1
Racing Santander	4	1
Villena Cf	4	1
Real Sporting De Gij	4	1
Real Madrid Cf	4	1
Real Betis Balompi	4	1
Fc Barcelona	4	1
Rcd Espanyol	4	1
Sevilla Fc	4	1
Atletico Madrid	4	1
Getafe Cf	4	1

Result Grid | Filter Rows: Q Search Export: Fetch rows: ⏪

country	average_height	average_weight
Belgium	182.88	187.0000
England	187.96	190.0000
France	185.42	183.0000

Result Grid | Filter Rows: Q Search Export: ⏪

league_name	team_long_name	num_matches
Belgium Jupiler League	Beerschot Ac	1
Belgium Jupiler League	Krc Genk	1
England Premier League	Manchester United	1
England Premier League	Manchester United	1
France Ligue 1	Fc Nantes	1
France Ligue 1	Aj Auxerre	1
Germany 1. Bundesliga	Hannover 96	1
Germany 1. Bundesliga	Fc Bayern Munich	1
Italy Serie A	Siena	1
Italy Serie A	Atalanta	1
Netherlands Eredivisie	Pg Groningen	1
Netherlands Eredivisie	Utrecht	1
Poland Ekstraklasa	Polonia Bytom	1

Result Grid | Filter Rows: Q Search Export: ⏪

```

36  # 5.Get the player_name, date, and overall rating, along with the average overall rating for the previous 5 days.
37  ● SELECT
38    pa.player_fifa_id,
39    pa.date,
40    pa.overall_rating,
41    AVG(pa.overall_rating) OVER (PARTITION BY pa.player_fifa_id ORDER BY pa.date ROWS BETWEEN 4 PRECEDING AND CURRENT ROW) AS avg_rating
42  FROM
43    player_attributes AS pa
44  ORDER BY
45    pa.date
46    pa.player_fifa_id,
47    pa.date]
48
49
50  ◇ 19:47
51
52  Result Grid Filter Rows: Search Export:
53
54
55  # 6.Get the player name and their overall rating for the latest date in the player_attributes table,
56  # for players who have a potential rating greater than or equal to 80.
57  ● SELECT
58    p.player_name,
59    pa.overall_rating
60  FROM
61    player_attributes AS pa
62  JOIN
63    players AS p ON pa.player_id = p.player_api_id
64  WHERE
65    pa.date = (SELECT MAX(date) FROM player_attributes)
66    AND pa.potential >= 80
67  ORDER BY
68    pa.overall_rating DESC;
69
70
71  ◇ 28:08
72
73  Result Grid Filter Rows: Search Export:
74
75
76  player_name overall_rating
77
78  ▶ Fred 81
79
80  ▶ Adriyan 70
81
82
83  Result 39
84
85  # 7.Get the player name and their preferred foot for players who have an overall rating
86  # greater than or equal to 85 and a potential rating greater than or equal to 90.
87  ● SELECT
88    DISTINCT(p.player_name),
89    pa.preferred_foot
90  FROM
91    player_attributes AS pa
92  JOIN
93    players AS p ON pa.player_id = p.player_api_id
94  WHERE
95    pa.overall_rating >= 85
96    AND pa.potential >= 90
97  ORDER BY
98    p.player_name ASC;
99
100
101  ◇ 23:74
102
103  Result Grid Filter Rows: Search Export:
104
105
106  player_name preferred_foot
107
108  ▶ Adriano left
109  ▶ Alberto Gilardino right
110  ▶ Alessandro Del Piero right
111  ▶ Alessandro Del Piera right
112  ▶ Alexandre Pato right
113  ▶ Andreia Pinto right
114  ▶ Andres Iniesta right
115  ▶ Andres Palop right
116  ▶ Andrey Shevchenko right
117  ▶ Andriy Shevchenko right
118  ▶ Antonio Cassano right
119  ▶ Arjen Robben left
120  ▶ Carles Puyol right
121  ▶ Cesinha right
122  ▶ Cesc Fabregas right
123
124
125  Result 42
126
127
128  # 8.Get the number of players in each age group.
129  ● SELECT
130    CASE
131      WHEN age < 20 THEN 'Under 20'
132      WHEN age >= 20 AND age < 30 THEN '20-29'
133      WHEN age >= 30 AND age < 40 THEN '30-39'
134      ELSE '40 and Over'
135    END as age_group,
136    COUNT(*) as total_players
137  FROM
138    players
139  GROUP BY
140    age_group;
141
142
143  ◇ 22:01
144
145  Result Grid Filter Rows: Search Export:
146
147
148  age_group total_players
149
150  ▶ 30-39 3048
151
152  ▶ 40 and Over 1268
153
154  ▶ 20-29 679
155
156
157  Result 42

```

```

158  # 9. Get the total number of goals scored by each team in each match.
159  ● SELECT
160    match_id,
161    home_team_id,
162    away_team_id,
163    home_team_goal,
164    away_team_goal,
165    SUM(home_team_goal + away_team_goal) OVER (PARTITION BY match_id) as total_goals
166  FROM
167    `match`;
168
169
170  ◇ 13:11
171
172  Result Grid Filter Rows: Search Export:
173
174
175  match_id home_team_id away_team_id home_team_goal away_team_goal total_goals
176
177  ▶ 9984 9835 0 3 3
178  ▶ 9991 9998 5 0 5
179  ▶ 7947 9985 1 3 4
180  ▶ 8203 8342 1 1 2
181  ▶ 9999 9571 2 2 4
182  ▶ 4649 9995 1 2 3
183  ▶ 10001 9995 1 0 1
184  ▶ 8342 8571 4 1 5
185  ▶ 9985 9986 1 2 3
186  ▶ 10000 9991 0 2 2
187  ▶ 9994 9998 0 0 0
188  ▶ 7947 10001 2 2 4
189  ▶ 8203 9999 1 2 3
190  ▶ 9996 9984 0 1 1
191  ▶ 4049 9987 1 3 4
192  ▶ 9993 9835 1 3 4
193  ▶ 8635 9994 2 3 5
194  ▶ 9998 9996 0 0 0
195  ▶ 8342 8342 2 2 4
196  ▶ 10000 9994 2 0 2
197
198
199  Result 45

```

Local instance 2008

Schemas

- 1 Use soccermanagementsystem;
- 2 SELECT team_long_name,
- 3 (SELECT COUNT(*)
- 4 FROM Match
- 5 WHERE home_team_id = teams.team_id OR away_team_id = teams.team_id) AS num_matches
- 6 FROM teams;

Result Grid Filter Rows: Search Export:

team_long_name	num_matches
Kic Gerki	212
Beşiktaş JK	213
FC Schalke 04	213
Spalding Lawyer	212
Kic Gerke Brugge	182
RSC Anderlecht	212
FC Twente	219
Race Monz	64
Standard Liège	212
FC Twente Enschede	212
Club Brugge K	212
Kic Rosenda	212
Kic Kerji	212
FC Twente Alkmaar	212
Kic Venlo	182
Spalding Deinze	182
FC Twente V	18
Lierse SK	128
Kic Expert	30
Oud Heverlee Le	80
Kic Onderda	36
Royal Excel Mous	60
Monarcida United	30
FC Twente Venlo	28
Res 2	2

Action Output

Time	Action	Response
12:53:54	SELECT p.player_name,pa.date,pa.overall_rating FROM player_pa...	Error Code: 1054 Unknown column 'p.player_fifa_id' in 'on clause'
12:54:19	SELECT t.team_long_name,p.player_name,pa.overall_rating FROM Ma...	Error Code: 1066 Table 'soccermanagementsystem.player_attributes' doesn't exist
12:54:42	SELECT t.team_long_name,p.player_name,pa.overall_rating FROM Ma...	Error Code: 1066 Table 'soccermanagementsystem.player_attributes' doesn't exist
12:54:48	SELECT t.team_long_name,c.player_name,ca.avail_ratio FROM Ca...	Error Code: 1066 Table 'soccermanagementsystem.ca_avail_ratio' doesn't exist

Query Complied

NoSQL

- What are all the countries currently represented in the database and how many teams from each country are included?

```
1 > db.country.find();
2
3
4
country > id
_id      id          name
643c9af6c4c4 1       Belgium
643c9af6c4c4 1729    England
643c9af6c4c4 4769    France
643c9af6c4c4 7809    Germany
643c9af6c4c4 10257   Italy
643c9af6c4c4 13274   Netherlands
643c9af6c4c4 15722   Poland
643c9af6c4c4 17642   Portugal
643c9af6c4c4 19694   Scotland
```

- What are the players in the database with an overall rating higher than 50, and how do their attributes compare to players with a lower overall rating?

```
0 > db.player_attributes.find({overall_rating: {$gt: 50}});
1
2
player_attributes > player_fifa_id
_id      player_fifa_id  player_id    date        overall_rating  potential  preferred_foot  crossing  finishi
643c9b006c4c 218353    505942     2/18/16    67           71         right        49        4
643c9b006c4c 218353    505942     11/19/15   67           71         right        49        4
643c9b006c4c 218353    505942     9/21/15    62           66         right        49        4
643c9b006c4c 218353    505942     3/20/15    61           65         right        48        4
643c9b006c4c 218353    505942     2/22/07    61           65         right        48        4
643c9b006c4c 189615    155782     4/21/16    74           76         left         80        5
643c9b006c4c 189615    155782     4/7/16     74           76         left         80        5
643c9b006c4c 189615    155782     1/7/16     73           75         left         79        5
643c9b006c4c 189615    155782     12/24/15   73           75         left         79        5
```

- What is the average height of players in the database, and how does it vary by country or league?

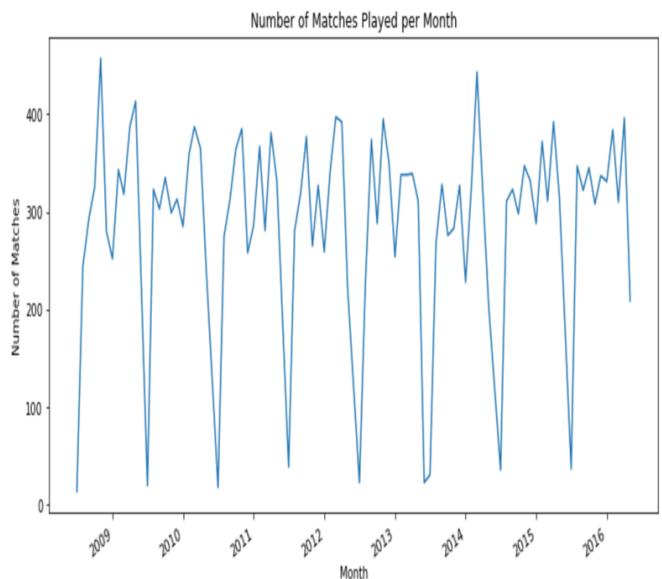
```
31
32 > db.players.aggregate([{$group: {_id: null, avg_height: {$avg: '$height'}}}]);
33
players > avg_height
_id      avg_height
null     181.44744
```

V. Database Access via R or Python

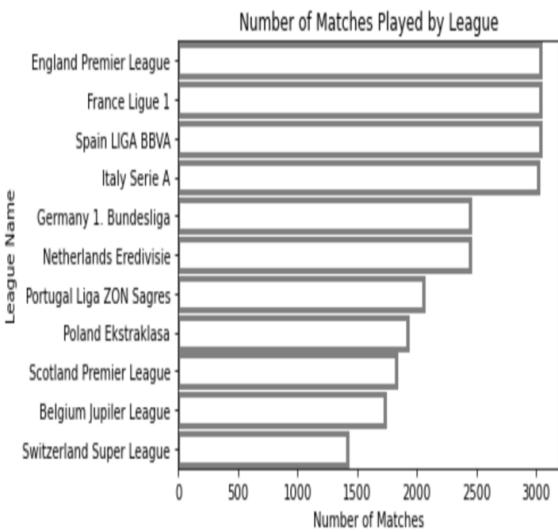
Connecting MySQL Database with Python Notebook

```
mydb = mysql.connector.connect(host='localhost', port='3306', user='root',
passwd='Samsungstar@5', db="Football", auth_plugin='mysql_native_password',
buffered=True)
```

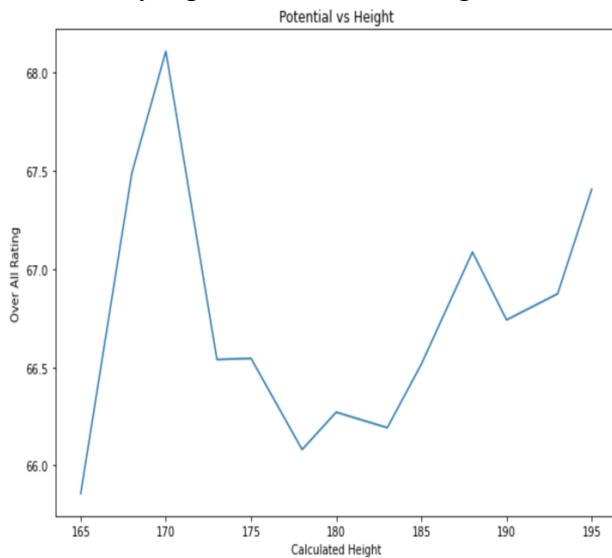
1. Number of Matches played per Month.



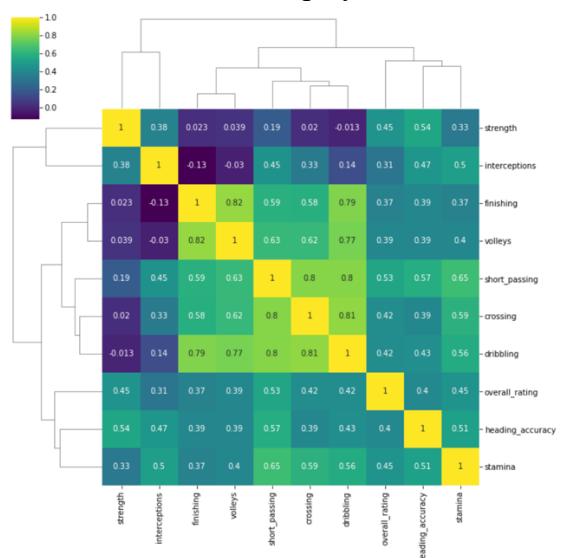
2. Number of matches played in each League.



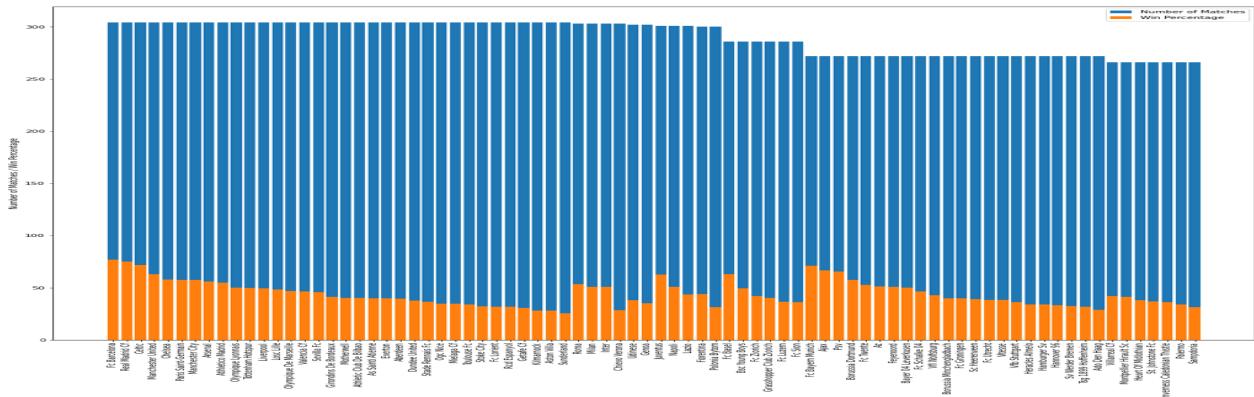
3. Player potential vs their Height



4. Correlation between player Attributes



5. Number of matches with winning and losing percent



VI. Summary & Recommendation

A soccer management system implemented within a database management system (DBMS) is a software tool designed to assist with the daily management of a soccer club or league. We started by drawing EER and UML diagrams based on this business challenge, taking into account all of the available entities and their attributes. We allocated cardinalities to each of the entities based on the business challenge. We created a Relational model using the EER model and its cardinalities to establish new relations and bring them to the highest normal form in order to reduce data redundancy. We used the fillDB site to generate dummy data for each entity and populate our database. We created some SQL queries to demonstrate how it works. Because the dataset was connected to the main server, we used Jupiter notebook in Python to access to the database and conduct some basic visualizations. We filled the database in MongoDB by exporting the constructed tables and writing various queries to aggregate tables.

A soccer management system integrated into a database management system can increase the efficiency and productivity of a soccer club or league by streamlining day-to-day operations, freeing up time for coaching and player development. Furthermore, greater communication among staff, players, and parents might result in a more cohesive squad. The system's valuable insights can aid in making better decisions concerning player selection, training, and tactics. It can also boost fan participation by making game, player, and team information available, resulting in a more supportive fan community.