European Soccer Data Analysis With SQL

By Marvis Edorodion

Introduction

Welcome to the world of European soccer data analysis! In this project, you will have the opportunity to dive into the European Soccer Database, consisting of leagues, matches, teams, and player information. Get ready to explore and analyse this extensive collection of data to uncover valuable insights and trends in the world of soccer. Let's kick off this journey together!

Objectives

The goal of this SQL project is to analyse the European Soccer Database, specifically focusing on the information contained in the leagues, match, team, and player CSV files

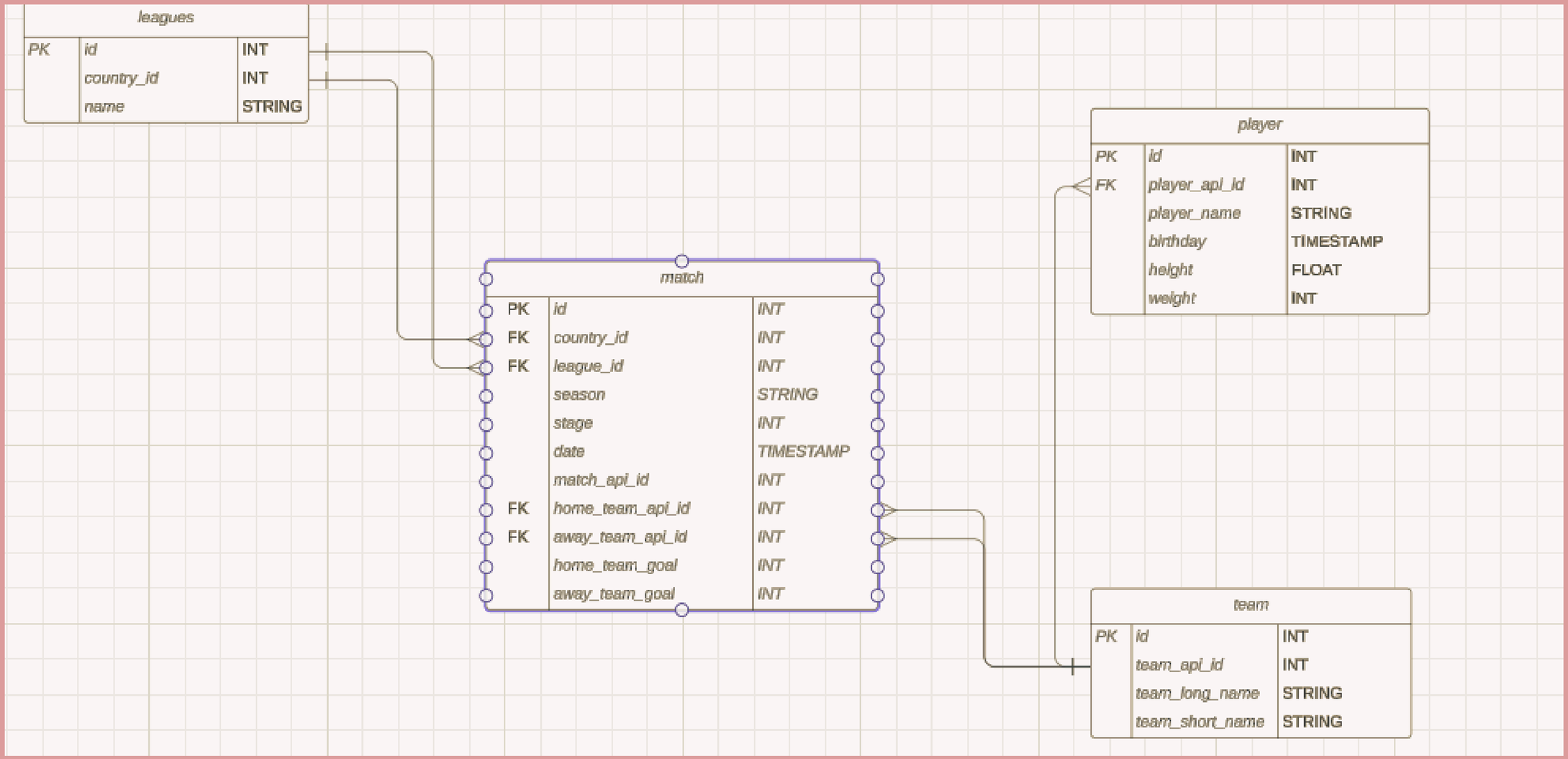
This analysis will involve querying the database to extract valuable insights such as team performance, player statistics, match outcomes, and trends within various leagues. By leveraging SQL queries and database manipulation techniques, we aim to gain a comprehensive understanding of the dynamics and patterns present in European soccer data.

Skills/Concepts Demonstrated

The following SQL features were incorporated in this analysis:

1. Data Exploration:
2. Data Filtering and Selection:
3. Data Aggregation and Analysis:
4. Data Manipulation and Joining:

PROJECT SCHEMA
Lucid Chat



days passed from the oldest Match to the most recent one

Query

```sql
SELECT
DATE_DIFF(MAX(date), MIN(date), DAY) AS num_of_days
FROM `sql-sandbox-422308.Final_Exercise.match`;
```

Result

JOB INFORMATION | RESULTS

Row	num_of_days
1	2868

Aggregate functions of home goals scored for each season and league name

Query

```sql
SELECT season, name, MIN(home_team_goal) AS min_home_goal, AVG(home_team_goal) AS avg_home_goal,
((MAX(home_team_goal) + MIN(home_team_goal))/2) AS mid_range_goal, MAX(home_team_goal) AS max_home_goal,
SUM(home_team_goal) AS sum_home_goal
FROM `sql-sandbox-422308.Final_Exercise.match` AS m
LEFT JOIN  `sql-sandbox-422308.Final_Exercise.leagues` AS l
ON m.country_id = l.country_id
GROUP BY season, name
ORDER BY sum_home_goal desc;
```

Result

	season	name	min_home_goal	avg_home_goal	mid_range_goal	max_home_goal	sum_home_goal
1	2009/2010	England Premier League	0	1.697368421052...	4.5	9	645
2	2012/2013	Spain LIGA BBVA	0	1.686842105263...	3.0	6	641
3	2011/2012	Spain LIGA BBVA	0	1.678947368421...	4.0	8	638
4	2008/2009	Spain LIGA BBVA	0	1.660526315789...	3.5	7	631
5	2010/2011	Spain LIGA BBVA	0	1.636842105263...	4.0	8	622
6	2013/2014	Spain LIGA BBVA	0	1.631578947368...	3.5	7	620
7	2010/2011	England Premier League	0	1.623684210526...	3.5	7	617
8	2015/2016	Spain LIGA BBVA	0	1.618421052631...	5.0	10	615
9	2009/2010	Spain LIGA BBVA	0	1.599999999999...	3.0	6	608

Results per page: 50 1 – 50 of

Unique seasons and number of matches played by each leagues

Query

```sql
SELECT season, name, count(m.id) AS count
FROM `sql-sandbox-422308.Final_Exercise.match` AS m
LEFT JOIN  `sql-sandbox-422308.Final_Exercise.leagues
ON m.country_id = l.country_id
GROUP BY season, name
ORDER BY count;
```

Result

1	2013/2014	Belgium Jupiler League	12
2	2011/2012	Switzerland Super League	162
3	2008/2009	Switzerland Super League	180
4	2009/2010	Switzerland Super League	180
5	2010/2011	Switzerland Super League	180
6	2012/2013	Switzerland Super League	180
7	2013/2014	Switzerland Super League	180
8	2014/2015	Switzerland Super League	180
9	2015/2016	Switzerland Super League	180
10	2009/2010	Belgium Jupiler League	210
11	2008/2009	Scotland Premier League	228
12	2009/2010	Scotland Premier League	228
13	2010/2011	Scotland Premier League	228

Players BMI table

```
CREATE TABLE `sql-sandbox-422308.Final_Exercise.PlayerBMI`
AS SELECT (weight/2.205) AS kg_weight, (height/100) AS m_height,
(weight/2.205)/POW((height/100), 2) AS BMI
FROM `sql-sandbox-422308.Final_Exercise.player`;
```

count of players with and without optimal BMI

```
SELECT COUNT(*) AS count_of_rows
FROM `sql-sandbox-422308.Final_Exercise.PlayerBMI`
WHERE BMI BETWEEN 18.5 AND 24.9;
```

Row	count_of_rows
1	10197

```
SELECT COUNT(*) AS count_of_rows
FROM `sql-sandbox-422308.Final_Exercise.PlayerBMI`
WHERE BMI NOT BETWEEN 18.5 AND 24.9;
```

Row	count_of_rows
1	863

Total goals scored(home and away) by a team during the most available season

Query

```
SELECT team_long_name, season, (SUM(home_team_goal) + SUM(away_team_goal)) AS total_goal
FROM `sql-sandbox-422308.Final_Exercise.match` AS m
LEFT JOIN  `sql-sandbox-422308.Final_Exercise.team` AS t
ON m.home_team_api_id = t.team_api_id
GROUP BY team_long_name, season
ORDER BY season DESC, total_goal DESC;
```

Result

Row	team_long_name	season	total_goal
1	Real Madrid CF	2015/2016	86
2	FC Barcelona	2015/2016	81
3	BSC Young Boys	2015/2016	74
4	Paris Saint-Germain	2015/2016	71
5	Manchester City	2015/2016	68
6	Celtic	2015/2016	67
7	Everton	2015/2016	65
8	SL Benfica	2015/2016	65
9	Borussia Dortmund	2015/2016	63
10	AZ	2015/2016	62
11	Chelsea	2015/2016	62
12	Ross County FC	2015/2016	62
13	PEC Zwolle	2015/2016	62

Highest goal scoring team for each season

Query

```
SELECT season, team_long_name,
FROM (SELECT season, team_long_name,
RANK() OVER (PARTITION BY season ORDER BY (SUM(home_team_goal) + SUM(away_team_goal)) DESC)
as position FROM `sql-sandbox-422308.Final_Exercise.match` AS m
 LEFT JOIN  `sql-sandbox-422308.Final_Exercise.team` AS t
 ON m.home_team_api_id = t.team_api_id
 GROUP BY season, team_long_name
 ORDER BY season)
WHERE position = 1;
```

Result

Row	season	team_long_name
1	2008/2009	Real Madrid CF
2	2009/2010	Chelsea
3	2010/2011	Willem II
4	2011/2012	Real Madrid CF
5	2012/2013	Real Madrid CF
6	2013/2014	Real Madrid CF
7	2014/2015	Real Madrid CF
8	2015/2016	Real Madrid CF

Top Team Scorer Table

```
CREATE TABLE `sql-sandbox-422308.Final_Exercise.TopScorer`
AS SELECT t.id AS team_id, team_long_name, (SUM(home_team_goal) + SUM(away_team_goal)) AS total_goal
FROM `sql-sandbox-422308.Final_Exercise.match` AS m
LEFT JOIN  `sql-sandbox-422308.Final_Exercise.team` AS t
ON m.home_team_api_id = t.team_api_id
GROUP BY t.id, team_long_name
ORDER BY total_goal DESC
```

```
SELECT * FROM `sql-sandbox-422308.Final_Exercise.TopScorer`
```

Result

Row	team_id	team_long_name	total_goal
1	43040	Real Madrid CF	652
2	43042	FC Barcelona	596
3	26556	PSV	496
4	3466	Manchester City	490
5	49117	BSC Young Boys	485
6	15617	FC Bayern Munich	478
7	39395	Celtic	477
8	49118	FC Basel	470
9	26562	SC Heerenveen	469
10	3475	Chelsea	462

Possible pair combination for the top ten team scorer

Query

```sql
SELECT t1.team_long_name AS team_1, t2.team_long_name AS team_2
FROM `sql-sandbox-422308.Final_Exercise.TopScorer` AS t1
JOIN `sql-sandbox-422308.Final_Exercise.TopScorer` AS t2
ON t1.team_id < t2.team_id;
```

Result

team_1	team_2
Real Madrid CF	FC Barcelona
Real Madrid CF	BSC Young Boys
Real Madrid CF	FC Basel
Manchester City	FC Bayern Munich
Manchester City	Real Madrid CF
Manchester City	Chelsea
Manchester City	PSV
Manchester City	Celtic
Manchester City	SC Heerenveen
Manchester City	FC Barcelona
Manchester City	BSC Young Boys
Manchester City	FC Basel