**DATA EXPLORATION - EUROPEAN SOCCER DATABASE**

1. How many days have passed from the oldest **Match** to the most recent one (dataset time interval)?

**ANSWER**: 2868 days

**QUERY:**

SELECT

MAX(date) AS Newest\_date,

MIN(date) AS Oldest\_date,

TIMESTAMP\_DIFF(MAX(date),MIN(date), DAY) AS Days\_Dataset\_inteval

FROM `progetto-di-prova-365418.Final\_Exercise.Match`

1. Produce a table which, for each Season and **League** Name, shows some statistics (min, avg, mid-range, max, sum) about the home goals scored.Which combination of Season-League has the highest number of home goals?

**ANSWER:** England Premier League - Season 2009/2010 with 645 home goals.

**QUERY:**

SELECT match.season,

leagues.name AS league\_name,

MIN(match.home\_team\_goal) AS min\_homegoals,

MAX(match.home\_team\_goal) AS max\_homegoals,

ROUND(AVG(match.home\_team\_goal),2) AS avg\_homegoals,

(MIN(match.home\_team\_goal) + MAX(match.home\_team\_goal))/2 AS midrange\_homegoals,

SUM(match.home\_team\_goal) AS sum\_homegoals,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

GROUP BY match.season, league\_name

ORDER BY sum\_homegoals DESC

LIMIT 1;

1. Find out how many unique seasons there are in the **Match** table.   
   Then write a query that shows, for each Season, the number of matches played by each League. Do you notice anything out of the ordinary?

**ANSWER:** The number of seasons in the "match" table is 8. It should be noted that in the 2013/2014 season only 12 matches were played in the Belgium Jupiler League.

**QUERY:**

SELECT COUNT(DISTINCT season) AS distinct\_season

FROM `progetto-di-prova-365418.Final\_Exercise.Match`;

SELECT

match.season,

leagues.name AS league\_name,

COUNT(match.id) AS number\_matches,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

GROUP BY match.season, league\_name

ORDER BY number\_matches;

1. Using Players as the starting point, create a new table (PlayerBMI) and add:
   1. a new variable that represents the players’ weight in kg (divide the mass value by 2.205) and call it kg\_weight;
   2. a variable that represents the height in metres (divide the cm value by 100) and call it m\_height;
   3. a variable that shows the body mass index (BMI) of the player;
   4. Filter the table to show only the players with an optimal BMI (from 18.5 to 24.9).

How many rows does this table have? How many players do not have an optimal BMI?

**ANSWER:** Players with an excellent BMI are 10197. Players who do not have an excellent BMI are 863.

**QUERY:**

CREATE TABLE `progetto-di-prova-365418.Final\_Exercise.PlayerBMI` AS

SELECT \*,

weight/2.205 AS kg\_weight,

height/100 AS m\_height,

(weight/2.205) / ((height/100)\*(height/100)) AS BMI

FROM `progetto-di-prova-365418.Final\_Exercise.Player`;

SELECT COUNT(\*) AS N\_player\_with\_optimal\_BMI,

FROM `progetto-di-prova-365418.Final\_Exercise.PlayerBMI`

WHERE BMI >= 18.5 AND BMI <= 24.9;

SELECT COUNT(\*) AS N\_player\_with\_not\_optimal\_BMI,

FROM `progetto-di-prova-365418.Final\_Exercise.PlayerBMI`

WHERE BMI < 18.5 OR BMI > 24.9;

1. Which **Team** has scored the highest total number of goals (home + away) during the most recent available season? How many goals has it scored?

**ANSWER:** The team that has scored the highest goals number goals in the most recent season (2015/16) is Barcelona with 112 goals.

**QUERY:**

SELECT MAX(DISTINCT season) AS season\_most\_recent

FROM `progetto-di-prova-365418.Final\_Exercise.Match`;

SELECT league\_name, team\_name, SUM(team\_goal) AS tot\_team\_goal

FROM(

SELECT \* FROM

(SELECT

leagues.name AS league\_name,

team\_home.team\_long\_name AS team\_name,

SUM(match.home\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_home

on match.home\_team\_api\_id = team\_home.team\_api\_id

WHERE season = "2015/2016"

GROUP BY league\_name, team\_name

) AS table\_home\_goals

UNION ALL

SELECT \* FROM

(SELECT

leagues.name AS league\_name,

team\_away.team\_long\_name AS team\_name,

SUM(match.away\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_away

on match.away\_team\_api\_id = team\_away.team\_api\_id

WHERE season = "2015/2016"

GROUP BY league\_name, team\_name

) AS table\_away\_goals

)

GROUP BY league\_name, team\_name

ORDER BY tot\_team\_goal DESC

LIMIT 1;

1. Create a query that, for each season, shows the name of the team that ranks first in terms of total goals scored.   
   Which team was the one that ranked first in most of the seasons?

**ANSWER:** The team that comes first most times for goals scored in the seasons played (among those available in the Dataset) is "Real Madrid CF" (4 times out of 8 seasons), followed by "FC Barcelona" (3 times out of 8 seasons) and “Ajax” (1 time in 8 seasons).

**QUERY:**

SELECT team\_name, count(team\_name) as num\_seasons\_as\_1stForGoalScored,

FROM(

SELECT \*,

RANK() over(partition by season order by tot\_team\_goal DESC) AS Frist\_team\_per\_goal\_each\_season

FROM(

SELECT season, league\_name, team\_name, SUM(team\_goal) AS tot\_team\_goal,

FROM(

SELECT \* FROM

(SELECT

match.season,

leagues.name AS league\_name,

team\_home.team\_long\_name AS team\_name,

SUM(match.home\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_home

on match.home\_team\_api\_id = team\_home.team\_api\_id

GROUP BY match.season, league\_name, team\_name

) AS table\_home\_goals

UNION ALL

SELECT \* FROM

(SELECT

match.season,

leagues.name AS league\_name,

team\_away.team\_long\_name AS team\_name,

SUM(match.away\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_away

on match.away\_team\_api\_id = team\_away.team\_api\_id

GROUP BY match.season, league\_name, team\_name

) AS table\_away\_goals

)

GROUP BY season, league\_name, team\_name

)

)

WHERE Frist\_team\_per\_goal\_each\_season = 1

GROUP BY team\_name;

1. From the query above create a new table (TopScorer) containing the top 10 teams in terms of total goals scored.   
   Then write a query that shows all the possible “pair combinations” between those 10 teams. How many “pair combinations” did it generate?

**ANSWER:** The top 10 teams for goals scored in the "TopScorer" table are: FC Barcelona, Real Madrid CF, Celtic, FC Bayern Munich, PSV, Ajax, FC Basel, Manchester City, Chelsea, Manchester United.

Considering all matches in which the "TopScorer" teams face each other in each championship, there are a total of 80 matches (10 for each season among those available in the dataset).

**QUERY:**

CREATE TABLE `progetto-di-prova-365418.Final\_Exercise.TopScorer` AS

SELECT team\_name, SUM(team\_goal) AS tot\_team\_goal,

FROM(

SELECT \* FROM

(SELECT

match.season,

leagues.name AS league\_name,

team\_home.team\_long\_name AS team\_name,

SUM(match.home\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_home

on match.home\_team\_api\_id = team\_home.team\_api\_id

GROUP BY match.season, league\_name, team\_name

) AS table\_home\_goals

UNION ALL

SELECT \* FROM

(SELECT

match.season,

leagues.name AS league\_name,

team\_away.team\_long\_name AS team\_name,

SUM(match.away\_team\_goal) AS team\_goal,

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Leagues` AS leagues

on match.league\_id = leagues.id

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team\_away

on match.away\_team\_api\_id = team\_away.team\_api\_id

GROUP BY match.season, league\_name, team\_name

) AS table\_away\_goals

)

GROUP BY team\_name

ORDER BY tot\_team\_goal DESC

LIMIT 10;

SELECT season, COUNT(match\_id) AS NumTopMatches

FROM (

SELECT id as match\_id,

season,

home\_team\_api\_id,

away\_team\_api\_id,

team\_long\_name\_home,

team\_long\_name\_away,

CASE WHEN team\_long\_name\_home IN (SELECT team\_name FROM `progetto-di-prova-365418.Final\_Exercise.TopScorer`) THEN "TopScorer"

ELSE "NormalScorer"

END AS ScorerType\_home,

CASE WHEN team\_long\_name\_away IN (SELECT team\_name FROM `progetto-di-prova-365418.Final\_Exercise.TopScorer`) THEN "TopScorer"

ELSE "NormalScorer"

END AS ScorerType\_away

FROM (

SELECT match.\*, team.team\_long\_name as team\_long\_name\_away

FROM(

SELECT match.\*, team.team\_long\_name as team\_long\_name\_home

FROM `progetto-di-prova-365418.Final\_Exercise.Match` AS match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team

on match.home\_team\_api\_id = team.team\_api\_id

) as match

LEFT JOIN `progetto-di-prova-365418.Final\_Exercise.Team` AS team

on match.away\_team\_api\_id = team.team\_api\_id

)

)

WHERE ScorerType\_home = "TopScorer" AND ScorerType\_away = "TopScorer"

GROUP BY season