#First step: I created the 'European Soccer Database' on Google BigQuery

#Second step: I loaded the 4 tables into it in csv (match, leagues, team and player)

#Below is the list of queries used for data analysis

#Remember to replace ‘`European\_Soccer\_Database’ with your Dataset ID

#How many matches in total are we analyzing?

SELECT COUNT(id) Tot\_match FROM `European\_Soccer\_Database.match`

#How many leagues in total are we analyzing?

SELECT DISTINCT(name) FROM `European\_Soccer\_Database.leagues`

#Calculate the difference between the most recent date and the most distant one and the two dates at the extremes

SELECT DATE\_DIFF(max(date), min(date), day) AS Total\_Range,

FORMAT\_TIMESTAMP('%m-%d-%Y', TIMESTAMP(max(date))) AS Most\_recent,

FORMAT\_TIMESTAMP('%m-%d-%Y', TIMESTAMP(min(date))) AS Less\_recent

FROM `European\_Soccer\_Database.match`;

#Count how many seasons are there in total

SELECT DISTINCT season Season

FROM `European\_Soccer\_Database.match`

#How many matches for each season?

SELECT COUNT(id) TotMatch, season Season

FROM `European\_Soccer\_Database.match`

GROUP BY Season

ORDER BY Season

#Count how many matches have been played for each League

SELECT DISTINCT (m.season) Season, l.name LeagueName, COUNT(match\_api\_id) TotMatch

FROM `European\_Soccer\_Database.match` m

LEFT JOIN `European\_Soccer\_Database.leagues` l

on m.league\_id = l.id

GROUP BY m.season, l.name

ORDER BY TotMatch DESC

#Do we notice anything out of the ordinary?

SELECT DISTINCT (m.season) Season, l.name LeagueName, COUNT(match\_api\_id) TotMatch

FROM `European\_Soccer\_Database.match` m

LEFT JOIN `European\_Soccer\_Database.leagues` l

on m.league\_id = l.id

GROUP BY m.season, l.name

ORDER BY TotMatch ASC

LIMIT 1

#How many matches were there for each month of the year? I used this query to download a dataset match\_per\_month

SELECT COUNT(id) Tot\_match, EXTRACT(MONTH FROM date) Month FROM `pure-respect-400108.European\_Soccer\_Database.match`

GROUP BY Month

#Produce a table that shows for each Season and League Name, the following statistics about the home goals scored: min, average, mid-range, max and sum

SELECT m.season Season, l.name LeagueName,

MIN(m.home\_team\_goal) minHTG,

ROUND(AVG(m.home\_team\_goal),2) avgHTG,

CAST(((MIN(m.home\_team\_goal) + MAX(m.home\_team\_goal))/2) AS INT64) midrangeHTG,

MAX(m.home\_team\_goal) maxHTG,

SUM(m.home\_team\_goal) sumHTG

FROM `European\_Soccer\_Database.match` m

LEFT JOIN `European\_Soccer\_Database.leagues` l

ON m.league\_id = l.id

GROUP BY m.season, l.name

ORDER BY sumHTG desc

#Create a new 'PlayerBMI' table in which we insert:

# - the weight in kg (kg\_weight);

# - the height in meters (m\_height);

# - the player's body mass index (BMI).

#We filter the table to only show players with an optimal BMI (18.5 to 24.9)

CREATE TABLE `European\_Soccer\_Database.PlayerBMI` AS SELECT \*,

ROUND ((weight / 2.205),2) AS kg\_weight,

ROUND ((height / 100),2) AS m\_heigth,

ROUND ((weight / 2.205) / power(height / 100,2),2) AS BMI,

FROM `European\_Soccer\_Database.player`

WHERE (weight/2.205)/power(height/100, 2) between 18.5 and 24.9

#How many players do not have an optimal BMI?

SELECT

(SELECT count(id)

FROM `European\_Soccer\_Database.player`) -

(SELECT count(id)

FROM `European\_Soccer\_Database.PlayerBMI`) as PlayerNoBMI

#Which Team has scored the highest total number of goals during the most recent available season?

SELECT h.team\_long\_name, h.SumOfGoalHome, a.SumOfGoalAway,

h.SumOfGoalHome + a.SumOfGoalAway AS TotalGoal

FROM ( SELECT t.team\_long\_name, SUM(m.home\_team\_goal) AS SumOfGoalHome

FROM `European\_Soccer\_Database.match` m

INNER JOIN `European\_Soccer\_Database.team` t

ON m.home\_team\_api\_id = t.team\_api\_id

WHERE m.season = (SELECT MAX(season) FROM `European\_Soccer\_Database.match`)

GROUP BY t.team\_long\_name ORDER BY SumOfGoalHome) h

INNER JOIN

(SELECT t.team\_long\_name, SUM(m.away\_team\_goal) AS SumOfGoalAway

FROM `European\_Soccer\_Database.match` m

INNER JOIN `European\_Soccer\_Database.team` t

ON m.away\_team\_api\_id = t.team\_api\_id

WHERE m.season = (SELECT MAX(season) FROM `European\_Soccer\_Database.match`)

GROUP BY t.team\_long\_name ORDER BY SumOfGoalAway) a

ON h.team\_long\_name = a.team\_long\_name

ORDER BY TotalGoal DESC

LIMIT 1

#For each season, which team ranks first in terms of total goals scored?

SELECT \* FROM

(SELECT h.season, h.team\_long\_name, h.SumOfGoalHome, a.SumOfGoalAway, h.SumOfGoalHome + a.SumOfGoalAway AS TotalGoal,

RANK() OVER (PARTITION BY a.season ORDER BY h.SumOfGoalHome + a.SumOfGoalAway DESC) AS rank\_season FROM

(SELECT m.season, t.team\_long\_name, SUM(m.home\_team\_goal) AS SumOfGoalHome

FROM `European\_Soccer\_Database.match` m INNER JOIN

`European\_Soccer\_Database.team` t ON m.home\_team\_api\_id = t.team\_api\_id

GROUP BY m.season, t.team\_long\_name ORDER BY SumOfGoalHome) h

INNER JOIN

(SELECT m.season,t.team\_long\_name, sum(m.away\_team\_goal) AS SumOfGoalAway

FROM `European\_Soccer\_Database.match` m INNER JOIN

`European\_Soccer\_Database.team` t ON m.away\_team\_api\_id = t.team\_api\_id

GROUP BY m.season, t.team\_long\_name ORDER BY SumOfGoalAway) a

ON h.team\_long\_name = a.team\_long\_name AND h.season=a.season)

WHERE rank\_season = 1

ORDER BY season DESC

#Create a new table 'TopScorer' containing the top 10 teams in terms of total goals scored

CREATE TABLE `European\_Soccer\_Database.TopScorer` AS

(SELECT h.team\_api\_id ,h.team\_long\_name, h.SumOfGoalHome, a.SumOfGoalAway,

h.SumOfGoalHome + a.SumOfGoalAway AS TotalGoal FROM

(SELECT t.team\_api\_id ,t.team\_long\_name, SUM(m.home\_team\_goal) AS SumOfGoalHome

FROM `European\_Soccer\_Database.match` m INNER JOIN

`European\_Soccer\_Database.team` t ON m.home\_team\_api\_id = t.team\_api\_id

where m.season = (select MAX(season) FROM `European\_Soccer\_Database.match`)

GROUP BY t.team\_api\_id, t.team\_long\_name ORDER BY SumOfGoalHome) h INNER JOIN

(SELECT t.team\_long\_name, SUM(m.away\_team\_goal) AS SumOfGoalAway

FROM `European\_Soccer\_Database.match` m INNER JOIN

`European\_Soccer\_Database.team` t ON m.away\_team\_api\_id = t.team\_api\_id

WHERE m.season = "2015/2016"

GROUP BY t.team\_long\_name ORDER BY SumOfGoalAway) a ON h.team\_long\_name = a.team\_long\_name

ORDER BY TotalGoal DESC

LIMIT 10)