

# Project: Investigate a Soccer Database

## Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)
- [Conclusions](#)

## Introduction

### What is Soccer Database?

The ultimate Soccer Database is a collection of tables that contain information regarding players, teams, matches, and more. I shall first introduce you to the structure of the Tables :

### Country

Attribute	Type	Description
id	integer	table identifier
name	text	country name

### League

Attribute	Type	Description
id	integer	table identifier
country_id	integer	foreign key for country table
name	text	name of league

### Match

Attribute	Type	Description
id	integer	table identifier
country_id	integer	foreign key for country table
league_id	integer	foreign key for the league
season	text	season eg.2018/2019
stage	integer	stage number
date	text	the date of the match
home_team_api_id	integer	the foreign key for the team represents the team who plays in his stadium
away_team_api_id	integer	the foreign key for the team represents the team who doesn't play in his stadium
home_team_goal	integer	number of goals home team has scored in the match
away_team_goal	integer	number of goals away team has scored in the match
home_player_x1	integer	position of the home player x axis
.....	.....	.....
away_player_x1	integer	position of the away player x axis
.....	.....	.....
home_player_y1	integer	position of the home player y axis
.....	.....	.....
away_player_y1	integer	position of the away player y axis
.....	.....	.....
home_player_1	integer	the foreign key for the player table represents the first player of the home team
.....	.....	.....
away_player_1	integer	the foreign key for the player table represents the first player of the away team

shoton	text	XML text to represent the id of each player who made a shot on the target
shotoff	text	XML text to represent the id of each player who made a shot off the target
goal	text	XML text to represent the ids of the players who scored goals and how many
foulcommit	text	XML text to represent the ids of the players who commit fouls
card	text	XML text to represent the ids of the players who took a [yellow-red] card
corner	text	XML text to represent the ids of the players who played corners
possession	text	XML text to represent home and away possession
B365H	numeric	Bookmaker or betting site odds in decimal format
B365D	numeric	Bookmaker or betting site odds in decimal format

## Player

Attribute	Type	Description
id	integer	table identifier
player_api_id	integer	unique id
palyer_fifa_api_id	integer	unique id
palyer_name	text	name of the player
birthday	text	palyer birthday
weight	integer	weight of the player
height	integer	height of the player

## Player\_Attributes

Attribute	Type	Description
id	integer	table identifier
player_api_id	integer	foreign key for player table
palyer_fifa_api_id	integer	foreign key for player table
date	text	date of the update
overall_rating	integer	fifa rating
potential	integer	fifa rating
preferred_foot	text	fifa rating

## Team

Attribute	Type	Description
id	integer	table identifier
team_api_id	integer	unique id
team_fifa_api_id	integer	unique id
team_long_name	text	team long name
team_short_name	text	team short name

## Team\_Attributes

Attribute	Type	Description
id	integer	table identifier
team_api_id	integer	foreign key for player table
team_fifa_api_id	integer	foreign key for player table
date	text	date of season

buildupplayspeed	integer	rank from 1 to 100
buildupplayspeedclass	text	classification for build up speed
buildupplaydribbling	integer	rank from 1 to 100
buildupplaydribblingclass	text	classification for build up dribbling
.....	.....	.....

## Notes

- the dots ..... in the tables above mean that the next attributes will have the same sequence and i didn't mention them to save time
- there isn't a key between team table and player table
- I have used [tablesgenerator web \(https://www.tablesgenerator.com/markdown\\_tables#\)](https://www.tablesgenerator.com/markdown_tables#) site to write these tables in the markdown cell
- I have found a lot of information about betting websites [her \(http://www.football-data.co.uk/notes.txt\)](http://www.football-data.co.uk/notes.txt)

## Questions

- 1) which team has improved through all seasons in Spain LIGA BBVA?
- 2) How is the player's physical power related to weight and height?
- 3) How did the number of red cards change from 2008 to 2016?

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
from sqlalchemy import create_engine
from sqlalchemy.types import Integer
from bs4 import BeautifulSoup
%matplotlib inline
```

## Data Wrangling

### Sections:

- ▷ Loading the data
- ▷ General Properties
  - [Teams Dataset](#)
  - [Players Dataset](#)
  - [Matches Dataset](#)
- ▷ Cleaning the data
  - [Teams\\_Dataset](#)
  - [players\\_Dataset](#)
  - [Matches\\_Dataset](#)

### Loading Data

**Loading Data** is one of the most easiest thing with **sql** ,but reading multiple tables each time we want to answer a question lead to code redundancy so in the next cell i will make a two functions to do the following:

- [Build a connection to our database](#)
- [Load the data into a dataframe](#)
- [Convert all the columns names to lowercase and replace spaces with " \\_ "](#)
- [the second function will load a dataframe into our database](#)

In [3]:

```
def make_Query(query):
    """This function to make query from database.db
    inputs :
    query -----> string -----> the query to be done
    outputs:
    df -----> dataframe-----> dataframe contains the query result
    """
    # the path of my database
    path = r"C:\Users\Eslam\Desktop\Eslam\Summer\Data analysis Udacity\Professional\Introduction to data analysis
\Project 2\database.sqlite"
    #bulding a connection with create_engine object
    engine = create_engine('sqlite:///'+path)
    df = pd.read_sql(query, engine)
    df.rename(columns = lambda name:name.lower().replace(' ','_'),inplace = True)
    # types and look for instances of missing or possibly errant data.
    return df
```

In [4]:

```
def load_table( df, name, index = False, datatypes=None):
    """
    this function ask the user for dataframe to load it to our database
    inputs :
    df-----> DataFrame -----> dataframe to be converted into a table in database
    name-----> string -----> name of the table
    index-----> Boolean -----> specify if the table will come with index or not
    datatypes-----> dictionary -----> specify datatype for each column such that column name is the key
    outputs :
    None
    """
    # the path of my database
    path = r"C:\Users\Eslam\Desktop\Eslam\Summer\Data analysis Udacity\Professional\Introduction to data analysis
\Project 2\database.sqlite"
    #bulding a connection with create_engine object
    engine = create_engine('sqlite:///'+path)
    df.to_sql( name, engine, index = index, dtype = datatypes)
```

In [5]:

```
# first of all let's get the team information
query = """
select ta.* ,t.team_short_name,t.team_long_name from Team_Attributes ta
JOIN Team t on(ta.team_api_id = t.team_api_id AND ta.team_fifa_api_id = t.team_fifa_api_id) ;
"""
df_teams_info = make_Query(query)
```

In [6]:

```
# let's get all the players information
query = """
select p.player_name ,p.birthday,p.height,p.weight,pa.* from Player_Attributes pa
JOIN Player p on(pa.player_api_id = p.player_api_id AND pa.player_fifa_api_id = p.player_fifa_api_id ) ;
"""
df_players_info = make_Query(query)
```

In [7]:

```
query = """
SELECT t1.team_short_name as home_team_short_name,t1.team_long_name as home_team_long_name,
t2.team_short_name as away_team_short_name,t2.team_long_name as away_team_long_name,
l.name as league_name,
m.*
FROM Team t1
JOIN Match m on (t1.team_api_id = m.home_team_api_id)
JOIN Team t2 on (t2.team_api_id = m.away_team_api_id)
JOIN League l on(m.league_id = l.id) ;
"""
df_mathces_info = make_Query(query)
```

## General Properties

let's explore our data shape , columns names , datatypes , null values , duplicates and unique values

In [8]:

```
def explore_dataset(df):
    """
    this function to print the main properties of any data set
    inputs:
    df-----> dataframe -----> dataframe object
    outputs:
    None
    """
    num_rows = df.shape[0]
    num_columns = df.shape[1]
    column_nulls = df.isnull().sum()
    total_nulls = df.isnull().sum().sum()
    num_duplicates = df.duplicated().sum()
    print("we have {} rows and {} columns ".format(num_rows,num_columns))
    print("\n","_"*50)
    for column in df.columns :
        print("column name : ",column,"\n")
        print("column datatype : ",type(df[column][0]),"\n")
        print(" unique values : \n{}\nwwith total {}".format(df[column].unique(),df[column].nunique()))
        print("\n","_"*20)
    print("\n","_"*50)
    print("columns with null values count :\n {} \n with total {}".format(column_nulls,total_nulls))
    print("\n","_"*50)
    print("columns with duplicates count :\n{}".format(num_duplicates))
    print("\n","_"*50)
```

## Teams Dataset

In [9]:

```
# first 5 rows of data
df_teams_info.head()
```

Out[9]:

	id	team_fifa_api_id	team_api_id	date	buildupplayspeed	buildupplayspeedclass	buildupplaydribbling	buildupplaydribblingclass
0	1	434	9930	2010-02-22 00:00:00	60	Balanced	NaN	Little
1	2	434	9930	2014-09-19 00:00:00	52	Balanced	48.0	Normal
2	3	434	9930	2015-09-10 00:00:00	47	Balanced	41.0	Normal
3	4	77	8485	2010-02-22 00:00:00	70	Fast	NaN	Little
4	5	77	8485	2011-02-22 00:00:00	47	Balanced	NaN	Little

5 rows x 9 columns

In [10]:

```
# teams information dataset
explore_dataset(df_teams_info)
```

we have 1458 rows and 27 columns

column name : id

column datatype : <class 'numpy.int64'>

unique values :  
[ 1 2 3 ... 1456 1457 1458]  
with total 1458

column name : team\_fifa\_api\_id

column datatype : <class 'numpy.int64'>

unique values :  
[ 434 77 614 47 1901 650 245 1861 229 1530

111082	111989	159	112513	1	2	39	448	240	100409
57	1906	241	1848	896	32	21	675	1897	1889
1714	234	88	3	1926	1898	160	189	4	59
22	23	111376	1943	1896	190	378	1796	210	1842
647	1961	112409	110364	450	78	1750	110915	5	192
231	1867	110747	1799	110502	635	242	110569	180	181
182	467	110500	1824	468	1746	162	452	10020	2013
7	111271	1971	79	31	100626	1915	236	1903	286
894	246	110374	110636	25	111657	144	673	110556	674
1860	110316	1888	110744	100632	110565	111429	110832	322	1805
165	62	28	184	485	80	100634	100879	166	81
10029	1952	111239	44	620	110745	45	29	1832	82
111083	100081	110724	472	1862	46	1738	1739	347	873
111091	1871	95	10018	64	1853	239	65	9	1844
301	2007	217	897	66	169	573	453	10	11
1893	219	68	12	69	1747	70	1900	83	111560
1904	1891	1823	71	100737	1910	13	72	1792	112225
171	477	1564	111540	682	479	100087	1892	10030	1843
73	50	1754	1914	100741	200	111086	111087	110746	1570
1790	247	15	456	86	480	1793	449	243	457
571	203	379	74	744	1905	1838	1902	546	52
631	874	1819	1837	111974	58	1913	34	324	481
110770	111092	226	17	100646	670	237	459	48	898
100805	110329	100804	680	232	1806	106	1960	260	1715
54	18	1809	294	1908	55	1895	461	110456	462
206	36	483	1909	1887	665	100651	110913	1795	38
109	19	681	1917	1907	1873	175	110	435	1742
900	110749	244	112512	15005]					

with total 285

---

column name : team\_api\_id

column datatype : <class 'numpy.int64'>

unique values :

[	9930	8485	8576	8564	10215	10217	8593	9865	8635	8121
	8322	108893	9912	158085	9825	10252	8524	8315	9906	8406
	8583	10229	8634	9976	9931	8178	9823	9993	10211	9807
	6493	9772	8658	8655	8483	8613	9911	9857	8559	9827
	9789	9788	4170	8678	10264	9858	8521	8191	7819	8529
	7788	8344	208931	8530	9910	9925	9984	9880	8455	8533
	8342	7869	2186	9826	8262	8526	9783	9836	8284	9938
	8457	8372	9776	9810	10268	8534	8398	8558	7842	6351
	8668	4087	10218	8596	8722	6631	8674	9773	9908	9824
	10243	10235	8535	8194	8358	9891	9879	9987	10233	9991
	8305	6391	9764	8569	6433	8019	8020	7878	9956	9855
	8357	9747	9790	8429	9904	9860	9791	10278	8177	10251
	8226	8667	8234	8636	8066	1957	9885	8350	8295	8597
	8245	8571	8203	8306	7730	8543	9746	8682	9888	2182
	8030	8673	8197	6421	8588	8581	9989	8639	8650	8537
	8244	9994	8689	10199	9748	9905	9864	8661	8456	10260
	10212	8592	8550	8549	9829	9998	10249	8348	9927	274581
	9996	9761	10214	8481	9830	9809	8464	10261	9831	9850
	6269	8165	8388	8242	2033	8573	8371	1773	6403	8460
	8540	9847	10167	8426	6413	6547	9878	8028	8033	8023
	8031	2183	8462	8640	10172	8696	8548	8370	9798	8603
	8633	8560	8479	8690	9837	9851	7841	10219	8551	9803
	9999	8686	8649	1601	9853	9882	7943	7794	10228	10189
	9777	8302	10179	8025	9874	8466	8614	9986	9768	9869
	9875	10190	9800	8152	8467	9997	9985	10194	8472	10003
	9867	10191	9804	8586	9941	10242	8611	8600	9771	10267
	9873	10281	9876	10269	10205	8277	7844	10238	9839	8475
	9817	8697	8659	8654	10001	8024	8528	8525	10265	8721
	8602	7955	9868	10192	8021	8394	8027	10000]		

with total 288

---

column name : date

column datatype : <class 'str'>

unique values :

[	'2010-02-22 00:00:00'	'2014-09-19 00:00:00'	'2015-09-10 00:00:00'
	'2011-02-22 00:00:00'	'2012-02-22 00:00:00'	'2013-09-20 00:00:00']

with total 6

---

column name : buildupplayspeed

column datatype : <class 'numpy.int64'>

unique values :

[60 52 47 70 58 62 59 65 45 48 30 53 38 56 40 31 35 55 42 46 50 23 41 39  
69 66 75 25 67 63 64 57 68 43 24 36 61 73 37 51 44 49 71 74 76 54 32 80  
34 72 29 78 33 26 28 20 77]  
with total 57

---

column name : buildupplayspeedclass

column datatype : <class 'str'>

unique values :  
['Balanced' 'Fast' 'Slow']  
with total 3

---

column name : buildupplaydribbling

column datatype : <class 'numpy.float64'>

unique values :  
[nan 48. 41. 64. 57. 70. 53. 47. 40. 43. 46. 61. 49. 66. 51. 32. 37. 45.  
52. 50. 38. 55. 35. 63. 30. 29. 34. 24. 39. 31. 60. 44. 36. 56. 54. 33.  
59. 58. 42. 69. 62. 67. 65. 77. 28. 68. 71. 26. 27. 74.]  
with total 49

---

column name : buildupplaydribblingclass

column datatype : <class 'str'>

unique values :  
['Little' 'Normal' 'Lots']  
with total 3

---

column name : buildupplaypassing

column datatype : <class 'numpy.int64'>

unique values :  
[50 56 54 70 52 62 45 53 40 30 44 35 47 55 66 38 33 39 65 41 51 37 48 58  
32 29 26 59 72 36 69 49 31 46 34 25 57 22 28 60 79 75 64 73 61 43 27 74  
67 42 68 77 71 63 24 20 23 80]  
with total 58

---

column name : buildupplaypassingclass

column datatype : <class 'str'>

unique values :  
['Mixed' 'Long' 'Short']  
with total 3

---

column name : buildupplaypositioningclass

column datatype : <class 'str'>

unique values :  
['Organised' 'Free Form']  
with total 2

---

column name : chancecreationpassing

column datatype : <class 'numpy.int64'>

unique values :  
[60 54 70 53 45 40 56 51 65 48 55 50 66 30 67 39 58 57 68 35 52 33 49 41  
28 44 63 38 36 61 47 59 37 77 34 21 43 69 32 42 72 46 62 71 64 73 31 80  
76 29]  
with total 50

---

column name : chancecreationpassingclass

column datatype : <class 'str'>

unique values :  
['Normal' 'Risky' 'Safe']  
with total 3

---

column name : chancecreationcrossing

column datatype : <class 'numpy.int64'>

unique values :  
[65 63 70 48 50 68 72 35 34 38 45 60 20 53 36 57 51 67 55 52 64 54 62 40  
33 59 44 30 47 49 78 73 69 74 56 25 24 31 66 46 37 71 61 58 43 26 42 77  
39 41 76 27 80 23 75 32]  
with total 56

---

column name : chancecreationcrossingclass

column datatype : <class 'str'>

unique values :  
['Normal' 'Lots' 'Little']  
with total 3

---

column name : chancecreationshooting

column datatype : <class 'numpy.int64'>

unique values :  
[55 64 70 52 57 63 50 75 69 60 65 66 44 38 67 46 39 30 40 47 48 35 43 42  
56 34 51 59 37 36 79 54 49 72 53 68 61 22 41 58 45 76 73 62 80 78 33 32  
71 28 31 23 77 24 27 74 29]  
with total 57

---

column name : chancecreationshootingclass

column datatype : <class 'str'>

unique values :  
['Normal' 'Lots' 'Little']  
with total 3

---

column name : chancecreationpositioningclass

column datatype : <class 'str'>

unique values :  
['Organised' 'Free Form']  
with total 2

---

column name : defencepressure

column datatype : <class 'numpy.int64'>

unique values :  
[50 47 60 40 42 41 49 30 45 38 48 58 52 39 46 57 65 68 64 70 37 53 44 25  
36 35 51 43 55 54 59 33 34 67 66 61 72 32 28 27 23 24 63 31 26 29 56 62]  
with total 48

---

column name : defencepressureclass

column datatype : <class 'str'>

unique values :  
['Medium' 'Deep' 'High']  
with total 3

---

column name : defenceaggression

column datatype : <class 'numpy.int64'>

unique values :  
[55 44 70 47 40 42 45 35 50 49 57 30 38 62 58 65 53 68 43 48 32 54 37 39  
41 66 63 67 60 34 61 51 46 72 52 59 56 33 69 71 64 27 28 36 24 31 29]  
with total 47

---

column name : defenceaggressionclass

column datatype : <class 'str'>



unique values :  
['Press' 'Double' 'Contain']  
with total 3

column name : defenceteamwidth

column datatype : <class 'numpy.int64'>

unique values :  
[45 54 70 52 60 63 30 50 53 49 65 61 62 64 59 67 51 35 36 58 57 48 37 55  
68 42 56 41 40 66 47 38 46 43 39 44 33 69 32 34 29 31 73]  
with total 43

column name : defenceteamwidthclass

column datatype : <class 'str'>

unique values :  
['Normal' 'Wide' 'Narrow']  
with total 3

column name : defencedefenderlineclass

column datatype : <class 'str'>

unique values :  
['Cover' 'Offside Trap']  
with total 2

column name : team\_short\_name

column datatype : <class 'str'>

unique values :  
['AAR' 'ABE' 'AJA' 'ACM' 'ACA' 'HAA' 'ALM' 'AND' 'ANG' 'ARK' 'ARL' 'BIE'  
'ARO' 'ARS' 'AVL' 'ATA' 'BIL' 'AMA' 'AUG' 'AUX' 'ALK' 'BAR' 'BAS' 'LEV'  
'BMU' 'BAC' 'B-M' 'BEL' 'BEN' 'BIR' 'BLB' 'BLA' 'BOA' 'BOC' 'BOL' 'BOR'  
'DOR' 'GLA' 'BOU' 'BRA' 'BRE' 'BUR' 'CAE' 'CAG' 'CAM' 'CAR' 'CAP' 'CAT'  
'CEL' 'CEB' 'CES' 'CHE' 'CHI' 'CLB' 'COR' 'CKR' 'CRY' 'DAR' 'GRA' 'DIJ'  
'DUF' 'DUU' 'DUN' 'EIB' 'EFR' 'ELC' 'EMP' 'COT' 'ESP' 'EST' 'EUP' 'EVE'  
'ETG' 'EXC' 'FAL' 'FCK' 'GRO' 'POR' 'UTR' 'VAD' 'ZUR' 'FEY' 'FIO' 'FDU'  
'FRE' 'FRO' 'FUL' 'GEN' 'GET' 'GAJ' 'GV' 'GAE' 'LEC' 'GOR' 'GRE' 'GRF'  
'GUI' 'HAM' 'HAN' 'HEA' 'HER' 'HBE' 'HIB' 'HOF' 'HUL' 'ING' 'INT' 'INV'  
'BIA' 'JUV' 'KAI' 'KAR' 'KIL' 'KKI' 'KOR' 'MEC' 'LAS' 'LAU' 'LAZ' 'LEH'  
'LEM' 'POZ' 'LGD' 'LEG' 'LEI' 'LEN' 'LIE' 'LIL' 'LIV' 'LOD' 'LOK' 'LOR'  
'LUZ' 'LYO' 'MAI' 'MAL' 'MCI' 'MUN' 'MAR' 'MET' 'MID' 'MON' 'MOR' 'MOT'  
'MOP' 'MOU' 'NAC' 'NAN' 'NAV' 'NEC' 'NEW' 'NIC' 'NOR' 'NOV' 'NUR' 'NUM'  
'ODR' 'OLH' 'OOS' 'OSA' 'O-H' 'FER' 'PAD' 'PAL' 'PSG' 'PAR' 'ZWO' 'PEN'  
'PES' 'PIG' 'POD' 'POG' 'POB' 'PWA' 'PSV' 'QPR' 'SAN' 'RAN' 'RAY' 'REA'  
'BET' 'SOC' 'HUE' 'REG' 'REI' 'REN' 'RA' 'RKC' 'SIE' 'ROD' 'ROS' 'ROM'  
'CHO' 'ETI' 'SAM' 'SAS' 'HEE' 'S04' 'SER' 'SEV' 'SIO' 'SLA' 'SOU' 'SPA'  
'CHA' 'SCP' 'SPG' 'NAP' 'GAL' 'MIR' 'STP' 'JOH' 'STT' 'STL' 'STK' 'SUN'  
'SWA' 'TEN' 'THU' 'TOR' 'TOT' 'TOU' 'TRO' 'TWE' 'UDI' 'ULE' 'VAL' 'VER'  
'STU' 'VIL' 'VIT' 'SET' 'VEN' 'WAA' 'WAT' 'WBR' 'WBA' 'WHU' 'WES' 'WID'  
'WIG' 'WII' 'WIS' 'WOL' 'XAM' 'XER' 'YB' 'ZAG' 'ZAR' 'ZAW' 'ZUL']  
with total 251

column name : team\_long\_name

column datatype : <class 'str'>

unique values :  
['FC Aarau' 'Aberdeen' 'AC Ajaccio' 'Milan' 'Académica de Coimbra'  
'ADO Den Haag' 'Ajax' 'UD Almería' 'RSC Anderlecht' 'Angers SCO'  
'Arka Gdynia' 'AC Arles-Avignon' 'DSC Arminia Bielefeld' 'FC Arouca'  
'Arsenal' 'Aston Villa' 'Atalanta' 'Athletic Club de Bilbao'  
'Atlético Madrid' 'FC Augsburg' 'AJ Auxerre' 'AZ' 'FC Barcelona' 'Bari'  
'FC Basel' 'Bayer 04 Leverkusen' 'FC Bayern Munich' 'Beerschot AC'  
'SC Beira Mar' 'CF Os Belenenses' 'AC Bellinzona' 'SL Benfica'  
'Birmingham City' 'Blackburn Rovers' 'Blackpool' 'Boavista FC'  
'VfL Bochum' 'Bologna' 'Bolton Wanderers' 'Girondins de Bordeaux'  
'Borussia Dortmund' 'Borussia Mönchengladbach' 'US Boulogne Cote D'Opale'  
'Bournemouth' 'SC Braga' 'Brescia' 'Stade Brestois 29' 'Burnley'  
'SM Caen' 'Cagliari' 'SC Cambuur' 'Cardiff City' 'Carpi' 'Catania'  
'RC Celta de Vigo' 'Celtic' 'KSV Cercle Brugge' 'Cesena' 'Chelsea'  
'Chievo Verona' 'Club Brugge KV' 'Córdoba CF' 'Cracovia' 'Crystal Palace'  
'SV Darmstadt 98' 'De Graafschap' 'RC Deportivo de La Coruña' 'Dijon FC0'  
'Dundee FC' 'Dundee United' 'Dunfermline Athletic' 'SD Eibar'

'Eintracht Braunschweig' 'Eintracht Frankfurt' 'Elche CF' 'Empoli'  
'FC Energie Cottbus' 'RCD Espanyol' 'Estoril Praia' 'KAS Eupen' 'Everton'  
'Évian Thonon Gaillard FC' 'Excelsior' 'Falkirk' '1. FC Köln'  
'FC Dordrecht' 'FC Groningen' 'FC Porto' 'FC Utrecht' 'FC Vaduz'  
'FC Zürich' 'Feyenoord' 'Fiorentina' 'Fortuna Düsseldorf' 'SC Freiburg'  
'Frosinone' 'Fulham' 'KRC Genk' 'Genoa' 'KAA Gent' 'Getafe CF'  
'GFC Ajaccio' 'Gil Vicente FC' 'GKS Bełchatów' 'Go Ahead Eagles'  
'Górnik Łęczna' 'Polonia Bytom' 'Granada CF' 'Grasshopper Club Zürich'  
'Grenoble Foot 38' 'SpVgg Greuther Fürth' 'En Avant de Guingamp'  
'Hamburger SV' 'Hamilton Academical FC' 'Hannover 96'  
'Heart of Midlothian' 'Heracles Almelo' 'Hércules Club de Fútbol'  
'Hertha BSC Berlin' 'Hibernian' 'TSG 1899 Hoffenheim' 'Hull City'  
'FC Ingolstadt 04' 'Inter' 'Inverness Caledonian Thistle'  
'Jagiellonia Białystok' 'Juventus' '1. FC Kaiserslautern' 'Karlsruher SC'  
'Kilmarnock' 'Korona Kielce' 'KV Kortrijk' 'KV Mechelen' 'UD Las Palmas'  
'FC Lausanne-Sports' 'Lazio' 'Le Havre AC' 'Le Mans FC' 'Lecce'  
'Lech Poznań' 'Lechia Gdańsk' 'Legia Warszawa' 'Leicester City'  
'Leixões SC' 'RC Lens' 'Levante UD' 'Lierse SK' 'LOSC Lille' 'Liverpool'  
'Livorno' 'Widzew Łódź' 'Sporting Lokeren' 'FC Lorient' 'FC Luzern'  
'Olympique Lyonnais' '1. FSV Mainz 05' 'Málaga CF' 'RCD Mallorca'  
'Manchester City' 'Manchester United' 'CS Marítimo'  
'Olympique de Marseille' 'FC Metz' 'Middlesbrough' 'AS Monaco'  
'RAEC Mons' 'Montpellier Hérault SC' 'Moreirense FC' 'Motherwell'  
'Royal Excel Mouscron' 'NAC Breda' 'CD Nacional' 'AS Nancy-Lorraine'  
'FC Nantes' 'Naval 1º de Maio' 'N.E.C.' 'Newcastle United' 'OGC Nice'  
'Norwich City' 'Novara' '1. FC Nürnberg' 'CD Numancia' 'Odra Wodzisław'  
'S.C. Olhanense' 'KV Oostende' 'CA Osasuna' 'Oud-Heverlee Leuven'  
'FC Paços de Ferreira' 'SC Paderborn 07' 'Palermo' 'Paris Saint-Germain'  
'Parma' 'Partick Thistle F.C.' 'PEC Zwolle' 'FC Penafiel' 'Pescara'  
'Piast Gliwice' 'Podbeskidzie Bielsko-Biała' 'Pogoń Szczecin'  
'P. Warszawa' 'Portsmouth' 'PSV' 'Queens Park Rangers' 'Racing Santander'  
'Rangers' 'Rayo Vallecano' 'Reading' 'Real Betis Balompié'  
'Real Madrid CF' 'Real Sociedad' 'RC Recreativo' 'Reggio Calabria'  
'Stade de Reims' 'Stade Rennais FC' 'Rio Ave FC' 'RKC Waalwijk' 'Siena'  
'Roda JC Kerkrade' 'KSV Roeselare' 'Roma' 'Ross County FC' 'Ruch Chorzów'  
'AS Saint-Étienne' 'Sampdoria' 'Sassuolo' 'SC Bastia' 'SC Heerenveen'  
'FC Schalke 04' 'Servette FC' 'Sevilla FC' 'FC Sion' 'Śląsk Wrocław'  
'FC Sochaux-Montbéliard' 'Southampton' 'Sparta Rotterdam'  
'Sporting Charleroi' 'Sporting CP' 'Real Sporting de Gijón' 'Napoli'  
'FC St. Gallen' 'St. Mirren' 'FC St. Pauli' 'St. Johnstone FC'  
'Sint-Truidense VV' 'Standard de Liège' 'Stoke City' 'Sunderland'  
'Swansea City' 'CD Tenerife' 'FC Thun' 'Torino' 'Tottenham Hotspur'  
'Toulouse FC' 'ES Troyes AC' 'FC Twente' 'Udinese' 'União de Leiria, SAD'  
'Valencia CF' 'Valenciennes FC' 'Real Valladolid' 'Hellas Verona'  
'VfB Stuttgart' 'Villarreal CF' 'Vitesse' 'Vitória Guimarães'  
'Vitória Setúbal' 'VVV-Venlo' 'Waasland-Beveren' 'Watford'  
'SV Werder Bremen' 'West Bromwich Albion' 'West Ham United'  
'KVC Westerlo' 'Wigan Athletic' 'Willem II' 'Wisła Kraków'  
'VfL Wolfsburg' 'Wolverhampton Wanderers' 'Neuchâtel Xamax'  
'Xerez Club Deportivo' 'BSC Young Boys' 'Zagłębie Lubin' 'Real Zaragoza'  
'Zawisza Bydgoszcz' 'SV Zulte-Waregem']

with total 285

---

```

columns with null values count :
id                                0
team_fifa_api_id                 0
team_api_id                     0
date                             0
buildupplayspeed                 0
buildupplayspeedclass           0
buildupplaydribbling            969
buildupplaydribblingclass       0
buildupplaypassing              0
buildupplaypassingclass         0
buildupplaypositioningclass     0
chancecreationpassing           0
chancecreationpassingclass      0
chancecreationcrossing          0
chancecreationcrossingclass     0
chancecreationshooting          0
chancecreationshootingclass     0
chancecreationpositioningclass  0
defencepressure                 0
defencepressureclass            0
defenceaggression               0
defenceaggressionclass         0
defenceteamwidth                0
defenceteamwidthclass          0
defencedefenderlineclass       0
team_short_name                 0

```

team\_long\_name 0  
dtype: int64  
with total 969

columns with duplicates count :  
0

In [11]:

```
# print some summary statistics about this data
df_teams_info.describe()
```

Out[11]:

	id	team_fifa_api_id	team_api_id	buildupplayspeed	buildupplaydribbling	buildupplaypassing	chancecreationpassing
count	1458.000000	1458.000000	1458.000000	1458.000000	489.000000	1458.000000	1458.000000
mean	729.500000	17706.982167	9995.727023	52.462277	48.607362	48.490398	52.165295
std	421.032659	39179.857739	13264.869900	11.545869	9.678290	10.896101	10.360793
min	1.000000	1.000000	1601.000000	20.000000	24.000000	20.000000	21.000000
25%	365.250000	110.000000	8457.750000	45.000000	42.000000	40.000000	46.000000
50%	729.500000	485.000000	8674.000000	52.000000	49.000000	50.000000	52.000000
75%	1093.750000	1900.000000	9904.000000	62.000000	55.000000	55.000000	59.000000
max	1458.000000	112513.000000	274581.000000	80.000000	77.000000	80.000000	80.000000

- it seems that each team appears alot but with different date value
- id,team\_fifa\_api\_id and team\_api\_id are unnecessary
- date attribute need to be coverted to date datatype
- date attribute ranges from 2010 to 2015
- all the attributes that have class at the end are categorical
- the rest of the attributes are continous random variables
- buildupplaydribbling attribute has a lot of null values so we have to drop it

Players Dataset

In [12]:

```
# first 5 rows of dataset
df_players_info.head()
```

Out[12]:

	player_name	birthday	height	weight	id	player_fifa_api_id	player_api_id	date	overall_rating	potential	...	vision	penalties
0	Aaron Appindangoye	1992-02-29 00:00:00	182.88	187	1	218353	505942	2016-02-18 00:00:00	67.0	71.0	...	54.0	48.0
1	Aaron Appindangoye	1992-02-29 00:00:00	182.88	187	2	218353	505942	2015-11-19 00:00:00	67.0	71.0	...	54.0	48.0
2	Aaron Appindangoye	1992-02-29 00:00:00	182.88	187	3	218353	505942	2015-09-21 00:00:00	62.0	66.0	...	54.0	48.0
3	Aaron Appindangoye	1992-02-29 00:00:00	182.88	187	4	218353	505942	2015-03-20 00:00:00	61.0	65.0	...	53.0	47.0
4	Aaron Appindangoye	1992-02-29 00:00:00	182.88	187	5	218353	505942	2007-02-22 00:00:00	61.0	65.0	...	53.0	47.0

5 rows x 14 columns

In [13]:

```
# players information dataset
explore_dataset(df_players_info)
```

we have 183766 rows and 46 columns

column name : player\_name

```
column datatype :    <class 'str'>

unique values :
['Aaron Appindangoye' 'Aaron Cresswell' 'Aaron Doran' ... 'Zsolt Low'
 'Zurab Khizanishvili' 'Zvezdan Misimovic']
with total 10848

column name :    birthday

column datatype :    <class 'str'>

unique values :
['1992-02-29 00:00:00' '1989-12-15 00:00:00' '1991-05-13 00:00:00' ...
 '1996-01-10 00:00:00' '1986-12-18 00:00:00' '1979-04-29 00:00:00']
with total 5762

column name :    height

column datatype :    <class 'numpy.float64'>

unique values :
[182.88 170.18 172.72 165.1  190.5  175.26 187.96 177.8  185.42 180.34
 200.66 198.12 193.04 167.64 195.58 162.56 203.2  160.02 157.48 208.28]
with total 20

column name :    weight

column datatype :    <class 'numpy.int64'>

unique values :
[187 146 163 198 154 161 139 181 170 150 168 143 176 159 185 172 165 157
 179 174 212 218 190 148 152 141 132 201 183 205 194 137 203 196 126 192
 134 214 130 209 128 207 216 123 220 225 227 117 243 121]
with total 50

column name :    id

column datatype :    <class 'numpy.int64'>

unique values :
[      1      2      3 ... 183976 183977 183978]
with total 183766

column name :    player_fifa_api_id

column datatype :    <class 'numpy.int64'>

unique values :
[218353 189615 186170 ... 111191  47058 102359]
with total 11060

column name :    player_api_id

column datatype :    <class 'numpy.int64'>

unique values :
[505942 155782 162549 ...  36491  35506  39902]
with total 11060

column name :    date

column datatype :    <class 'str'>

unique values :
['2016-02-18 00:00:00' '2015-11-19 00:00:00' '2015-09-21 00:00:00'
 '2015-03-20 00:00:00' '2007-02-22 00:00:00' '2016-04-21 00:00:00'
 '2016-04-07 00:00:00' '2016-01-07 00:00:00' '2015-12-24 00:00:00'
 '2015-12-17 00:00:00' '2015-10-16 00:00:00' '2015-09-25 00:00:00'
 '2015-01-09 00:00:00' '2014-12-05 00:00:00' '2014-11-07 00:00:00'
 '2014-09-18 00:00:00' '2014-05-02 00:00:00' '2014-04-04 00:00:00'
 '2014-03-14 00:00:00' '2013-12-13 00:00:00' '2013-11-08 00:00:00'
 '2013-10-04 00:00:00' '2013-09-20 00:00:00' '2013-05-03 00:00:00'
 '2013-03-22 00:00:00' '2013-03-15 00:00:00' '2013-02-22 00:00:00'
 '2013-02-15 00:00:00' '2012-08-31 00:00:00' '2012-02-22 00:00:00'
 '2011-08-30 00:00:00' '2010-08-30 00:00:00' '2010-02-22 00:00:00'
 '2009-08-30 00:00:00' '2009-02-22 00:00:00' '2008-08-30 00:00:00']
```

'2015-10-09 00:00:00'	'2014-12-12 00:00:00'	'2014-04-18 00:00:00'
'2014-01-31 00:00:00'	'2013-11-29 00:00:00'	'2013-05-31 00:00:00'
'2013-04-26 00:00:00'	'2013-04-19 00:00:00'	'2013-04-05 00:00:00'
'2013-03-08 00:00:00'	'2011-02-22 00:00:00'	'2015-10-02 00:00:00'
'2015-07-03 00:00:00'	'2015-06-12 00:00:00'	'2015-01-16 00:00:00'
'2014-11-14 00:00:00'	'2014-06-06 00:00:00'	'2014-04-11 00:00:00'
'2013-05-10 00:00:00'	'2007-08-30 00:00:00'	'2015-05-08 00:00:00'
'2015-04-10 00:00:00'	'2014-01-17 00:00:00'	'2016-04-28 00:00:00'
'2016-02-25 00:00:00'	'2015-09-04 00:00:00'	'2014-02-28 00:00:00'
'2014-02-14 00:00:00'	'2013-10-18 00:00:00'	'2016-02-04 00:00:00'
'2014-10-02 00:00:00'	'2015-06-05 00:00:00'	'2015-02-06 00:00:00'
'2014-03-21 00:00:00'	'2014-02-07 00:00:00'	'2013-11-01 00:00:00'
'2013-05-24 00:00:00'	'2016-01-28 00:00:00'	'2014-12-27 00:00:00'
'2014-10-31 00:00:00'	'2014-10-10 00:00:00'	'2015-12-03 00:00:00'
'2015-11-26 00:00:00'	'2015-10-30 00:00:00'	'2014-01-03 00:00:00'
'2013-03-01 00:00:00'	'2016-01-21 00:00:00'	'2015-11-06 00:00:00'
'2015-05-22 00:00:00'	'2015-05-15 00:00:00'	'2015-02-13 00:00:00'
'2015-01-30 00:00:00'	'2014-09-19 00:00:00'	'2013-11-22 00:00:00'
'2014-04-25 00:00:00'	'2013-12-27 00:00:00'	'2016-04-14 00:00:00'
'2015-11-12 00:00:00'	'2015-03-06 00:00:00'	'2013-12-20 00:00:00'
'2016-03-10 00:00:00'	'2016-03-03 00:00:00'	'2014-12-19 00:00:00'
'2013-12-06 00:00:00'	'2013-10-25 00:00:00'	'2013-10-11 00:00:00'
'2013-09-27 00:00:00'	'2013-08-16 00:00:00'	'2015-02-27 00:00:00'
'2015-01-02 00:00:00'	'2014-01-10 00:00:00'	'2016-05-12 00:00:00'
'2016-01-14 00:00:00'	'2016-06-23 00:00:00'	'2013-03-28 00:00:00'
'2016-03-17 00:00:00'	'2014-07-18 00:00:00'	'2013-06-07 00:00:00'
'2015-01-23 00:00:00'	'2015-05-29 00:00:00'	'2014-05-16 00:00:00'
'2014-05-09 00:00:00'	'2013-04-12 00:00:00'	'2014-10-24 00:00:00'
'2014-11-28 00:00:00'	'2014-11-26 00:00:00'	'2013-05-17 00:00:00'
'2016-06-09 00:00:00'	'2016-03-24 00:00:00'	'2015-03-27 00:00:00'
'2014-03-28 00:00:00'	'2014-03-07 00:00:00'	'2014-02-21 00:00:00'
'2014-01-24 00:00:00'	'2015-12-10 00:00:00'	'2016-03-31 00:00:00'
'2015-04-17 00:00:00'	'2015-04-01 00:00:00'	'2008-02-22 00:00:00'
'2015-02-20 00:00:00'	'2016-05-05 00:00:00'	'2016-02-11 00:00:00'
'2014-11-21 00:00:00'	'2014-07-25 00:00:00'	'2013-07-12 00:00:00'
'2014-10-17 00:00:00'	'2013-07-05 00:00:00'	'2015-10-23 00:00:00'
'2015-03-13 00:00:00'	'2013-11-15 00:00:00'	'2016-02-19 00:00:00'
'2015-04-24 00:00:00'	'2015-08-14 00:00:00'	'2016-07-07 00:00:00'
'2014-05-30 00:00:00'	'2014-09-23 00:00:00'	'2016-06-02 00:00:00'
'2016-05-19 00:00:00'	'2015-05-01 00:00:00'	'2014-05-23 00:00:00'
'2016-06-16 00:00:00'	'2015-03-10 00:00:00'	'2013-09-06 00:00:00'
'2015-10-19 00:00:00'	'2013-06-14 00:00:00'	'2016-05-26 00:00:00'
'2015-08-27 00:00:00'	'2015-07-31 00:00:00'	'2013-08-23 00:00:00'
'2013-08-30 00:00:00'	'2013-08-09 00:00:00'	'2015-07-24 00:00:00'
'2014-08-22 00:00:00'	'2014-08-15 00:00:00'	'2014-08-08 00:00:00'
'2013-07-26 00:00:00'	'2015-01-26 00:00:00'	'2013-06-21 00:00:00'
'2013-06-28 00:00:00'	'2013-08-02 00:00:00'	'2013-07-19 00:00:00'
'2015-07-10 00:00:00'	'2014-09-12 00:00:00'	'2013-09-13 00:00:00'
'2015-12-30 00:00:00'	'2014-08-29 00:00:00'	'2015-08-21 00:00:00'
'2016-06-30 00:00:00'	'2015-07-16 00:00:00'	'2014-08-01 00:00:00'
'2015-08-07 00:00:00'	'2013-03-04 00:00:00'	'2015-01-28 00:00:00'
'2014-09-05 00:00:00'	'2015-06-19 00:00:00'	'2014-09-26 00:00:00'
'2015-09-10 00:00:00'	'2015-06-26 00:00:00'	'2015-09-01 00:00:00'
'2016-02-13 00:00:00'	'2014-07-20 00:00:00']	

with total 197

---

column name : overall\_rating

column datatype : <class 'numpy.float64'>

unique values :

```
[67. 62. 61. 74. 73. 71. 70. 69. 68. 65. 64. 54. 51. 52. 47. 53. 66. 59.
 75. 72. 76. 78. 77. 79. 60. 80. 81. 82. 84. 48. 63. 83. 55. 58. 50. 56.
 nan 49. 57. 42. 46. 45. 85. 44. 86. 89. 87. 88. 91. 40. 90. 41. 43. 38.
 93. 92. 39. 33. 36. 37. 35. 94.]
```

with total 61

---

column name : potential

column datatype : <class 'numpy.float64'>

unique values :

```
[71. 66. 65. 76. 75. 77. 78. 79. 80. 68. 64. 60. 67. 70. 72. 69. 82. 73.
 74. 81. 83. 86. 84. 85. 87. 90. 56. 57. 63. 62. 61. nan 59. 55. 58. 53.
 89. 54. 88. 52. 91. 92. 93. 51. 46. 44. 50. 47. 45. 95. 94. 48. 49. 42.
 97. 96. 39.]
```

with total 56

---

column name : preferred\_foot

column datatype : <class 'str'>

unique values :  
['right' 'left' None]  
with total 2

column name : attacking\_work\_rate

column datatype : <class 'str'>

unique values :  
['medium' 'high' None 'low' 'None' 'le' 'norm' 'stoc' 'y']  
with total 8

column name : defensive\_work\_rate

column datatype : <class 'str'>

unique values :  
['medium' 'high' 'low' '\_0' None '5' 'ean' 'o' '1' 'ormal' '7' '2' '8' '4'  
'tocky' '0' '3' '6' '9' 'es']  
with total 19

column name : crossing

column datatype : <class 'numpy.float64'>

unique values :  
[49. 48. 80. 79. 78. 77. 74. 58. 57. 22. 64. 65. 67. 69. 72. 73. 63. 56.  
59. 60. 46. 45. 75. 76. 47. 82. 71. 12. 62. 26. 70. 66. 40. 53. 55. 54.  
61. 39. 41. 42. 52. 68. nan 35. 34. 50. 38. 84. 81. 25. 11. 6. 23. 29.  
36. 44. 43. 51. 32. 33. 15. 14. 9. 21. 28. 20. 37. 19. 18. 83. 8. 31.  
24. 92. 10. 30. 27. 86. 90. 88. 13. 17. 85. 87. 16. 91. 93. 3. 89. 7.  
5. 4. 94. 95. 2. 1.]  
with total 95

column name : finishing

column datatype : <class 'numpy.float64'>

unique values :  
[44. 43. 53. 52. 51. 50. 40. 39. 48. 58. 56. 61. 69. 21. 20. 34. 33. 57.  
72. 73. 74. 71. 77. 79. 32. 31. 66. 65. 67. 68. 70. 62. 64. 15. 42. 41.  
45. 63. 54. 23. 75. 76. 26. 25. 55. 29. 17. 16. 22. 60. 37. nan 30. 49.  
47. 46. 38. 28. 27. 80. 78. 11. 10. 18. 59. 36. 82. 13. 9. 5. 8. 35.  
14. 24. 81. 88. 92. 83. 84. 86. 90. 89. 93. 85. 19. 6. 3. 12. 4. 87.  
91. 7. 2. 1. 95. 94. 97. 96.]  
with total 97

column name : heading\_accuracy

column datatype : <class 'numpy.float64'>

unique values :  
[71. 70. 58. 57. 56. 51. 50. 52. 60. 59. 62. 45. 68. 67. 73. 74. 79. 72.  
76. 64. 65. 66. 47. 46. 28. 27. 38. 49. 16. 61. 82. 42. 54. 53. 63. 55.  
43. 78. 77. 75. 48. 69. 39. 30. 80. nan 81. 84. 13. 25. 12. 23. 41. 44.  
37. 32. 90. 40. 11. 10. 21. 35. 24. 18. 17. 33. 31. 26. 36. 85. 86. 87.  
83. 22. 19. 34. 88. 89. 15. 8. 92. 20. 14. 4. 91. 9. 3. 93. 29. 6.  
7. 95. 5. 1. 94. 2. 98.]  
with total 96

column name : short\_passing

column datatype : <class 'numpy.float64'>

unique values :  
[61. 60. 71. 70. 69. 67. 64. 62. 59. 45. 44. 39. 66. 68. 63. 65. 53. 72.  
54. 79. 80. 75. 77. 78. 76. 23. 73. 84. 74. 58. 57. 52. 42. 83. 82. 85.  
55. 56. 40. 51. 49. nan 43. 47. 48. 34. 33. 41. 35. 81. 46. 25. 38. 16.  
15. 21. 26. 32. 28. 27. 50. 36. 37. 20. 30. 9. 13. 11. 10. 8. 29. 22.  
12. 86. 87. 89. 18. 19. 88. 31. 90. 92. 24. 3. 93. 91. 94. 95. 97. 14.  
17. 96. 7. 6. 5. 4.]  
with total 95

column name : volleys

column datatype : <class 'numpy.float64'>

unique values :  
[44. 43. 40. 32. 29. 28. 30. 52. 51. 56. 63. 48. 47. 33. 77. 80. 34. 68.  
67. 69. 14. 60. 59. 65. 64. 57. 25. 61. 79. 81. 78. 76. 74. 73. 70. 42.  
41. 24. 55. 62. 36. 23. 72. nan 58. 31. 66. 37. 46. 45. 49. 13. 39. 50.  
53. 54. 35. 75. 71. 10. 11. 5. 22. 21. 16. 38. 15. 27. 26. 12. 20. 6.  
18. 17. 9. 19. 82. 84. 83. 87. 86. 85. 4. 8. 7. 2. 88. 89. 3. 90.  
91. 1. 93. 92.]  
with total 93

---

column name : dribbling

column datatype : <class 'numpy.float64'>

unique values :  
[51. 50. 73. 71. 67. 66. 65. 64. 62. 44. 43. 26. 69. 70. 72. 74. 56. 55.  
59. 60. 52. 47. 46. 48. 53. 63. 78. 80. 79. 81. 82. 77. 54. 84. 83. 85.  
90. 88. 93. 94. 86. 15. 57. 75. 76. 68. 61. 49. 42. 45. 30. 41. 21. 58.  
nan 36. 35. 34. 31. 25. 14. 9. 23. 40. 28. 38. 37. 11. 10. 5. 20. 32.  
19. 18. 87. 33. 12. 89. 39. 16. 27. 22. 92. 13. 7. 8. 29. 1. 6. 91.  
17. 24. 95. 4. 97. 3. 96. 2.]  
with total 97

---

column name : curve

column datatype : <class 'numpy.float64'>

unique values :  
[45. 44. 70. 68. 67. 66. 65. 63. 64. 41. 71. 72. 43. 38. 37. 42. 52. 77.  
78. 79. 56. 55. 62. 14. 69. 73. 59. 57. 58. 21. 51. 35. 76. 75. 27. 23.  
22. nan 46. 40. 26. 54. 49. 36. 53. 15. 25. 28. 29. 30. 39. 31. 47. 48.  
50. 33. 32. 61. 60. 80. 83. 81. 12. 18. 17. 34. 74. 20. 84. 82. 86. 85.  
13. 9. 11. 8. 6. 87. 19. 10. 5. 2. 7. 90. 88. 16. 24. 92. 94. 91.  
89. 4. 93.]  
with total 92

---

column name : free\_kick\_accuracy

column datatype : <class 'numpy.float64'>

unique values :  
[39. 38. 69. 68. 43. 42. 26. 63. 62. 64. 41. 40. 25. 24. 81. 76. 74. 73.  
70. 57. 48. 53. 55. 54. 66. 18. 67. 65. 78. 71. 59. 61. 60. 58. 30. 50.  
77. 45. 23. 22. 47. 44. nan 32. 31. 49. 29. 52. 56. 28. 72. 14. 13. 9.  
75. 51. 33. 34. 79. 37. 83. 46. 80. 35. 12. 11. 27. 86. 36. 21. 20. 16.  
15. 82. 91. 87. 88. 84. 17. 19. 89. 6. 8. 85. 7. 10. 90. 93. 95. 96.  
92. 5. 94. 3. 4. 97. 2. 1.]  
with total 97

---

column name : long\_passing

column datatype : <class 'numpy.float64'>

unique values :  
[64. 63. 68. 67. 65. 46. 45. 49. 62. 61. 53. 57. 58. 54. 60. 55. 70. 48.  
59. 18. 69. 72. 73. 79. 77. 74. 71. 66. 56. 80. 81. 82. 83. 75. 52. 51.  
47. 39. 38. 33. 42. 27. 76. nan 35. 44. 43. 32. 41. 25. 24. 84. 78. 31.  
29. 40. 36. 12. 11. 30. 34. 37. 15. 26. 21. 50. 23. 20. 17. 28. 22. 8.  
14. 13. 85. 86. 87. 16. 19. 7. 90. 92. 94. 95. 96. 97. 89. 88. 10. 5.  
91. 6. 9. 93. 4. 3.]  
with total 95

---

column name : ball\_control

column datatype : <class 'numpy.float64'>

unique values :  
[49. 48. 71. 70. 68. 67. 66. 65. 64. 43. 42. 34. 73. 72. 74. 54. 62. 61.  
69. 58. 57. 59. 63. 79. 78. 76. 75. 77. 81. 80. 82. 84. 88. 83. 22. 60.  
52. 51. 44. 35. 46. 47. nan 53. 55. 50. 17. 28. 56. 85. 18. 25. 21. 19.  
24. 27. 26. 86. 87. 45. 30. 32. 89. 40. 36. 37. 92. 23. 31. 39. 29. 38.  
41. 13. 14. 10. 20. 15. 11. 91. 90. 94. 33. 16. 12. 93. 95. 6. 9. 96.  
7. 97. 8. 5.]  
with total 93

column name : acceleration

column datatype : <class 'numpy.float64'>

unique values :

[60. 79. 80. 84. 85. 69. 43. 42. 66. 77. 81. 78. 82. 74. 76. 68. 71. 65.  
33. 37. 50. 55. 75. 73. 49. 54. 52. 53. 70. 72. 89. 90. 92. 94. 96. 93.  
95. 15. 63. 67. 64. 51. 34. 61. 56. 83. 62. 57. nan 59. 86. 48. 40. 31.  
26. 36. 47. 87. 88. 44. 58. 91. 41. 35. 46. 39. 32. 25. 45. 29. 30. 18.  
21. 20. 28. 23. 27. 24. 38. 22. 13. 97. 14. 17. 10. 16. 19.]

with total 86

column name : sprint\_speed

column datatype : <class 'numpy.float64'>

unique values :

[64. 78. 82. 81. 72. 47. 46. 59. 74. 75. 80. 79. 77. 34. 37. 40. 48. 55.  
57. 61. 73. 31. 53. 54. 71. 69. 67. 89. 91. 95. 93. 94. 92. 90. 26. 65.  
66. 68. 76. 63. 60. 85. 58. 70. 84. 62. nan 49. 88. 83. 87. 86. 56. 52.  
32. 33. 38. 50. 42. 39. 45. 43. 51. 44. 41. 21. 30. 23. 29. 24. 35. 36.  
28. 27. 22. 20. 25. 19. 96. 97. 15. 17. 12. 14. 16. 13.]

with total 85

column name : agility

column datatype : <class 'numpy.float64'>

unique values :

[59. 78. 79. 81. 77. 74. 85. 76. 66. 62. 63. 41. 45. 49. 57. 60. 93. 94.  
88. 31. 69. 72. 67. 55. 71. 68. 82. 75. 70. 73. 44. 48. 33. 36. 80. 86.  
65. 50. 51. nan 56. 46. 47. 83. 34. 53. 87. 58. 32. 42. 52. 35. 91. 90.  
64. 84. 54. 61. 43. 40. 38. 39. 92. 37. 22. 89. 28. 25. 26. 95. 23. 30.  
21. 29. 16. 24. 27. 15. 19. 96. 11. 20.]

with total 81

column name : reactions

column datatype : <class 'numpy.float64'>

unique values :

[47. 46. 67. 71. 69. 70. 66. 62. 61. 65. 51. 50. 57. 56. 64. 68. 80. 77.  
79. 81. 78. 74. 76. 75. 73. 60. 84. 45. 72. 58. 52. 82. 53. 63. 55. 42.  
nan 59. 54. 41. 48. 49. 83. 39. 35. 33. 44. 43. 24. 37. 38. 90. 85. 88.  
32. 40. 22. 23. 36. 87. 86. 89. 34. 26. 31. 29. 92. 91. 27. 30. 96. 28.  
21. 93. 95. 94. 25. 20. 17.]

with total 78

column name : balance

column datatype : <class 'numpy.float64'>

unique values :

[65. 90. 87. 62. 92. 84. 56. 41. 44. 45. 52. 59. 66. 76. 72. 73. 67. 80.  
94. 24. 79. 77. 71. 64. 53. 82. 75. 74. 70. 68. 60. 61. 54. 49. 78. 57.  
63. nan 58. 69. 25. 31. 50. 43. 81. 55. 47. 51. 38. 83. 88. 86. 48. 85.  
40. 30. 37. 36. 28. 33. 42. 29. 32. 34. 39. 91. 35. 27. 95. 93. 89. 46.  
17. 23. 21. 22. 20. 26. 96. 15. 18. 12.]

with total 81

column name : shot\_power

column datatype : <class 'numpy.float64'>

unique values :

[55. 54. 71. 70. 65. 48. 47. 61. 60. 56. 62. 67. 74. 64. 59. 68. 76. 78.  
77. 73. 53. 52. 69. 72. 26. 66. 63. 37. 81. 75. 58. 50. 57. 42. 82. nan  
41. 40. 46. 49. 19. 34. 45. 51. 43. 33. 32. 39. 88. 85. 84. 83. 20. 80.  
79. 16. 35. 23. 44. 17. 25. 21. 36. 27. 86. 30. 24. 12. 31. 29. 87. 90.  
89. 93. 94. 96. 38. 28. 22. 14. 11. 18. 91. 13. 15. 6. 10. 8. 92. 9.  
95. 5. 4. 7. 2. 97. 3.]

with total 96

column name : jumping

column datatype : <class 'numpy.float64'>



```
unique values :
[58. 85. 84. 77. 73. 64. 48. 65. 71. 56. 70. 69. 67. 78. 83. 81. 52. 38.
 63. 75. 72. 62. 66. 61. 68. 79. 46. 60. 54. nan 74. 57. 43. 76. 51. 55.
 33. 50. 53. 80. 59. 90. 88. 87. 49. 82. 47. 86. 35. 37. 89. 34. 44. 91.
 42. 45. 39. 94. 41. 92. 93. 30. 40. 36. 32. 14. 31. 25. 96. 95. 24. 27.
 26. 29. 28. 19. 20. 21. 23. 22.]
with total 79
```

---

```
column name :   stamina
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[54. 79. 80. 77. 76. 74. 73. 63. 59. 55. 56. 66. 72. 81. 68. 69. 49. 39.
 48. 75. 60. 62. 78. 85. 84. 86. 87. 18. 67. 70. 83. 61. 53. 92. 93. 90.
 58. 64. 65. 50. 47. 71. 88. nan 57. 45. 52. 42. 43. 46. 82. 31. 37. 89.
 35. 36. 34. 51. 33. 40. 21. 27. 41. 44. 91. 28. 22. 38. 30. 32. 29. 25.
 19. 24. 14. 94. 26. 23. 95. 16. 20. 96. 15. 17. 10.]
with total 84
```

---

```
column name :   strength
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[76. 56. 50. 49. 48. 43. 37. 38. 63. 68. 71. 90. 87. 86. 81. 78. 77. 75.
 72. 70. 74. 69. 65. 66. 62. 59. 52. 57. 41. 33. 44. 73. 51. 30. 64. 61.
 82. 47. 60. 45. 67. 40. 79. 55. 58. 83. 53. nan 46. 95. 39. 42. 88. 84.
 85. 89. 91. 54. 80. 31. 92. 94. 22. 21. 35. 93. 32. 27. 34. 36. 29. 28.
 25. 26. 20. 23. 24. 96. 18. 16. 12. 13. 10.]
with total 82
```

---

```
column name :   long_shots
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[35. 34. 62. 60. 59. 58. 56. 55. 54. 53. 57. 51. 52. 61. 76. 67. 63. 29.
 28. 33. 26. 25. 20. 15. 75. 77. 78. 73. 72. 68. 38. 37. 12. 66. 24. 74.
 79. 70. 30. 23. 47. 46. 43. 48. 71. 39. 69. 50. nan 44. 27. 65. 64. 11.
 49. 41. 31. 32. 42. 45. 22. 40. 18. 14. 13.  8. 21. 19. 81. 80. 82. 16.
 84. 10. 83. 85. 88. 90.  6.  9. 86. 91.  5.  3.  4. 36. 87. 93. 17. 89.
  7. 92.  1. 94. 95.  2. 96.]
with total 96
```

---

```
column name :   aggression
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[71. 63. 62. 68. 67. 66. 64. 60. 59. 70. 69. 72. 65. 77. 75. 76. 57. 55.
 51. 49. 36. 74. 56. 37. 46. 24. 21. 73. 78. 30. 45. 44. 43. 53. 54. 41.
 58. 48. 39. 38. 85. 84. 61. 52. 40. nan 80. 79. 83. 81. 87. 91. 89. 90.
 93. 86. 50. 31. 18. 25. 17. 23. 82. 88. 47. 26. 33. 32. 34. 35. 20. 19.
 42. 22. 29. 10. 27. 28. 16. 94. 12. 14. 92. 15. 11. 13.  8. 97. 96.  6.
 95.  9.]
with total 91
```

---

```
column name :   interceptions
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[70. 41. 40. 67. 66. 65. 62. 59. 31. 32. 55. 54. 44. 35. 61. 68. 71. 77.
 80. 81. 83. 74. 56. 57. 37. 36. 60. 73. 19. 64. 63. 58. 45. 43. 39. 69.
 72. 82. 30. 29. 24. 23. 51. 49. 48. 21. 27. 20. 25. 33. 28. 47. 50. 79.
 nan 52. 26. 76. 75. 42. 46. 84. 53. 12. 78. 34. 13. 38. 22. 15. 16. 10.
 18. 85. 11. 14. 17. 87.  7. 88. 93. 86. 92. 90.  9. 91.  8. 89. 95. 96.
 94.  6.  2.  5.  1.  3.  4.]
with total 96
```

---

```
column name :   positioning
```

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
[45. 44. 60. 58. 57. 56. 19. 27. 28. 64. 67. 68. 66. 62. 50. 30. 26. 25.
```

78. 81. 29. 76. 77. 71. 74. 75. 73. 39. 79. 69. 14. 65. 63. 55. 21. 70.  
41. 83. 84. 82. 80. 61. 52. 51. 49. 48. 72. 38. 23. 47. 53. 36. 17. 46.  
31. 54. nan 24. 22. 20. 11. 10. 18. 42. 35. 32. 59. 40. 85. 37. 33. 43.  
13. 12. 16. 15. 34. 87. 88. 4. 3. 89. 91. 95. 86. 90. 93. 5. 8. 9.  
6. 94. 2. 92. 7. 96.]  
with total 95

---

column name : vision

column datatype : <class 'numpy.float64'>

unique values :  
[54. 53. 66. 65. 64. 61. 58. 63. 47. 38. 68. 71. 72. 73. 67. 41. 40. 60.  
76. 80. 82. 79. 83. 75. 77. 81. 15. 70. 62. 78. 31. 55. 50. 49. 44. 36.  
43. 74. 52. 51. 46. nan 33. 32. 69. 30. 57. 42. 56. 39. 35. 24. 25. 59.  
45. 34. 29. 48. 21. 18. 17. 20. 23. 16. 27. 26. 22. 28. 85. 87. 37. 14.  
10. 84. 12. 9. 13. 86. 2. 89. 90. 7. 5. 6. 19. 8. 94. 95. 96. 92.  
11. 93. 88. 91. 4. 1. 3. 97.]  
with total 97

---

column name : penalties

column datatype : <class 'numpy.float64'>

unique values :  
[48. 47. 59. 58. 35. 28. 29. 61. 60. 63. 64. 66. 36. 37. 45. 44. 81. 82.  
76. 72. 68. 67. 73. 53. 42. 41. 62. 84. 70. 74. 56. 55. 52. 34. 69. 65.  
75. 71. 79. 54. 40. 39. 50. 85. 80. 32. 31. 22. 46. 57. nan 49. 43. 51.  
77. 38. 20. 25. 33. 83. 21. 78. 18. 17. 27. 86. 30. 12. 88. 94. 7. 16.  
19. 26. 23. 15. 93. 24. 91. 90. 92. 14. 9. 11. 13. 6. 87. 95. 10. 89.  
8. 5. 3. 2. 96.]  
with total 94

---

column name : marking

column datatype : <class 'numpy.float64'>

unique values :  
[65. 62. 76. 73. 72. 66. 64. 53. 52. 23. 22. 21. 67. 78. 80. 77. 75. 74.  
81. 79. 82. 69. 28. 36. 30. 51. 29. 25. 18. 15. 63. 58. 60. 68. 56. 54.  
59. 57. 35. 34. 46. 43. 42. 61. 32. 26. 70. 13. 20. 47. 45. 83. nan 55.  
19. 44. 38. 17. 71. 48. 37. 39. 24. 14. 33. 49. 50. 12. 9. 31. 40. 27.  
16. 5. 84. 85. 10. 41. 11. 7. 6. 1. 88. 86. 87. 94. 93. 8. 4. 90.  
89. 91. 92. 3. 2. 96.]  
with total 95

---

column name : standing\_tackle

column datatype : <class 'numpy.float64'>

unique values :  
[69. 66. 63. 75. 72. 68. 64. 60. 59. 45. 22. 21. 71. 76. 78. 74. 70. 73.  
79. 80. 83. 34. 37. 30. 27. 25. 58. 56. 33. 23. 29. 15. 65. 62. 61. 54.  
20. 31. 24. 77. 67. 53. 47. 46. 19. 26. 42. 41. 32. 36. 51. 50. 57. 39.  
nan 55. 14. 13. 44. 43. 35. 81. 48. 52. 28. 17. 16. 84. 82. 11. 10. 40.  
18. 9. 12. 86. 85. 87. 88. 49. 8. 38. 2. 89. 90. 93. 95. 7. 91. 5.  
92. 6. 4. 3. 94. 1.]  
with total 95

---

column name : sliding\_tackle

column datatype : <class 'numpy.float64'>

unique values :  
[69. 66. 78. 72. 67. 63. 24. 21. 68. 65. 70. 71. 74. 75. 76. 77. 35. 32.  
29. 57. 56. 34. 25. 14. 12. 61. 60. 62. 73. 51. 44. 43. 17. 54. 52. 37.  
36. 45. 64. 59. 19. 18. 23. 33. 31. 38. 22. 30. nan 40. 53. 42. 55. 27.  
13. 82. 84. 79. 58. 46. 15. 11. 20. 41. 81. 47. 49. 26. 16. 50. 48. 39.  
80. 28. 10. 9. 83. 85. 6. 5. 7. 2. 4. 89. 90. 86. 87. 8. 88. 92.  
3. 91. 93. 94. 95.]  
with total 94

---

column name : gk\_diving

column datatype : <class 'numpy.float64'>

unique values :

```
[ 6.  5. 14. 13. 16. 15.  8.  7. 12.  9. 53. 11. 10. 24. 33. nan  2.  4.
 70. 66. 65. 58.  3. 60. 51. 46. 77. 73. 75. 72. 63. 55. 69. 68. 64. 67.
 74. 71.  1. 80. 79. 78. 56. 61. 62. 49. 76. 59. 57. 20. 81. 22. 21. 31.
 54. 88. 83. 82. 48. 84. 18. 35. 90. 85. 86. 45. 19. 87. 17. 43. 44. 28.
 23. 52. 39. 47. 91. 93. 50. 26. 34. 32. 92. 89. 25. 94. 41. 40. 42. 29.
 27. 30. 36. 37.]
with total 93
```

---

column name : gk\_handling

column datatype : <class 'numpy.float64'>

unique values :

```
[11. 10.  7.  6. 22. 21. 12.  5. 20.  9. 13. 25.  8. 41. 15. 14. 23. 30.
 24. nan 16. 63. 62. 59.  4.  3.  2. 61. 49. 44. 71. 66. 72. 67. 56. 46.
 60. 58. 65. 69. 68. 48.  1. 75. 76. 79. 82. 70. 74. 73. 64. 55. 57. 53.
 51. 78. 77. 52. 19. 45. 43. 26. 54. 50. 32. 81. 83. 42. 29. 80. 84. 17.
 47. 86. 88. 18. 28. 87. 85. 91. 35. 92. 93. 90. 37. 27. 33. 89. 39. 36.
 40.]
with total 90
```

---

column name : gk\_kicking

column datatype : <class 'numpy.float64'>

unique values :

```
[10.  9.  8. 45. 46. 49. 12. 11. 54. 13. 64. 65. 61. 58. 16. 15. 57. 55.
  7.  6. 63. 48. 14. 59. 39. 73.  5. 79. 75. 72. 67. 21. 38. 33. 47. 52.
 42. 62. 27. 71. 76. nan 35. 53. 43. 51. 69. 44. 60. 66.  3.  2. 56. 70.
 78. 68. 31. 29. 74. 20. 41. 82. 81. 84. 23.  4. 36. 32. 50. 28. 25. 83.
 86. 37. 77. 30. 80. 34. 24. 40. 26.  1. 95. 96. 97. 89. 90. 88. 18. 91.
 22. 19. 87. 85. 94. 92. 93. 17.]
with total 97
```

---

column name : gk\_positioning

column datatype : <class 'numpy.float64'>

unique values :

```
[ 8.  7.  9. 22. 21. 12. 11. 20. 10. 16. 15.  6. 14. 13. 25. 51. 31. 24.
 23. nan 70. 65. 64. 39.  5.  3.  2. 19. 18. 62. 58. 63. 71. 72. 68. 60.
 50. 57. 53. 54. 59. 69. 67. 66.  4. 75. 76. 74. 61. 55. 56. 79. 78. 77.
 82. 81. 73. 52. 30. 80. 47. 32. 49. 43. 48. 84. 42.  1. 44. 83. 17. 86.
 85. 87. 88. 45. 28. 46. 92. 38. 93. 91. 90. 94. 34. 41. 89. 96. 35. 27.
 37. 29. 33. 26. 36.]
with total 94
```

---

column name : gk\_reflexes

column datatype : <class 'numpy.float64'>

unique values :

```
[ 8.  7. 12. 11. 22. 13. 21. 10. 20.  6. 15. 14.  9. 25. 53. 30.  5. 24.
 23. nan  4. 16. 75. 74.  3.  2. 64. 60. 58. 55. 78. 76. 73. 72. 68. 54.
 69. 63. 62. 66. 56. 80. 86. 85. 84. 81. 79. 52. 70. 18. 67. 71. 65. 59.
 61. 57. 83. 77. 82. 38. 17.  1. 48. 33. 50. 49. 45. 89. 88. 51. 87. 43.
 47. 46. 27. 37. 92. 90. 41. 19. 29. 44. 42. 91. 35. 26. 40. 93. 94. 96.
 32. 34. 31.]
with total 92
```

---

columns with null values count :

|                     |      |
|---------------------|------|
| player_name         | 0    |
| birthday            | 0    |
| height              | 0    |
| weight              | 0    |
| id                  | 0    |
| player_fifa_api_id  | 0    |
| player_api_id       | 0    |
| date                | 0    |
| overall_rating      | 750  |
| potential           | 750  |
| preferred_foot      | 750  |
| attacking_work_rate | 3144 |
| defensive_work_rate | 750  |
| crossing            | 750  |
| finishing           | 750  |

heading\_accuracy 750  
short\_passing 750  
volleys 2627  
dribbling 750  
curve 2627  
free\_kick\_accuracy 750  
long\_passing 750  
ball\_control 750  
acceleration 750  
sprint\_speed 750  
agility 2627  
reactions 750  
balance 2627  
shot\_power 750  
jumping 2627  
stamina 750  
strength 750  
long\_shots 750  
aggression 750  
interceptions 750  
positioning 750  
vision 2627  
penalties 750  
marking 750  
standing\_tackle 750  
sliding\_tackle 2627  
gk\_diving 750  
gk\_handling 750  
gk\_kicking 750  
gk\_positioning 750  
gk\_reflexes 750  
dtype: int64  
with total 44033

columns with duplicates count :  
0

In [14]:

```
# summary statistics  
df_players_info.describe()
```

Out[14]:

|       | height        | weight        | id            | player_fifa_api_id | player_api_id | overall_rating | potential     | crossing      |   |
|-------|---------------|---------------|---------------|--------------------|---------------|----------------|---------------|---------------|---|
| count | 183766.000000 | 183766.000000 | 183766.000000 | 183766.000000      | 183766.000000 | 183016.000000  | 183016.000000 | 183016.000000 | 1 |
| mean  | 181.875022    | 168.769974    | 91973.292715  | 165645.222642      | 135957.515514 | 68.597522      | 73.457851     | 55.081687     |   |
| std   | 6.395603      | 15.092227     | 53104.857009  | 53870.761415       | 136979.374220 | 7.041099       | 6.591720      | 17.237884     |   |
| min   | 157.480000    | 117.000000    | 1.000000      | 2.000000           | 2625.000000   | 33.000000      | 39.000000     | 1.000000      |   |
| 25%   | 177.800000    | 159.000000    | 45986.250000  | 155715.750000      | 34763.000000  | 64.000000      | 69.000000     | 45.000000     |   |
| 50%   | 182.880000    | 168.000000    | 91946.500000  | 183480.000000      | 77767.000000  | 69.000000      | 74.000000     | 59.000000     |   |
| 75%   | 185.420000    | 179.000000    | 137950.750000 | 199848.000000      | 191081.000000 | 73.000000      | 78.000000     | 68.000000     |   |
| max   | 208.280000    | 243.000000    | 183978.000000 | 234141.000000      | 750584.000000 | 94.000000      | 97.000000     | 95.000000     |   |

8 rows x 10 columns

In [15]:

```
# inspect the null values with corresponding player name
df_players_info[df_players_info['heading_accuracy'].isnull()][['player_name']]
```

Out[15]:

| player_name |                    |
|-------------|--------------------|
| 483         | Abdeslam Ouaddou   |
| 983         | Abel Gomez         |
| 1752        | Adam Johnson       |
| 2089        | Adam Rooney        |
| 2482        | Adil Chihi         |
| ...         | ...                |
| 182375      | Youssef Hadji      |
| 183037      | Zbigniew Malkowski |
| 183097      | Zdenek Pospech     |
| 183149      | Zdravko Kuzmanovic |
| 183275      | Ze Roberto         |

750 rows × 1 columns

In [16]:

```
# we will use the player "Abdeslam Ouaddou" as example
# let's search for heading_accuracy values for that player
df_players_info.query('player_name == "Abdeslam Ouaddou"')[['heading_accuracy']]
```

Out[16]:

| heading_accuracy |      |
|------------------|------|
| 478              | 70.0 |
| 479              | 80.0 |
| 480              | 80.0 |
| 481              | 80.0 |
| 482              | 80.0 |
| 483              | NaN  |

Comment

one of the best ways to solve this problem is to fill null values with the mean value of the same calss  
let  $D$  be a data set with  $C_1, C_2, C_3, \dots, C_n$  columns.  
Suppose that  $C_i$  is the class column. Hence, we can devide each column into multiple categories  $G_1, G_2, G_3, \dots, G_j$   
Each category contain a list of elements  $e_1, e_2, e_3, \dots, e_k$  such that  $e_k$  is the last element before null value. Hence, we will fill the the null value as follows :

$$e_{k+1} = \sum_k^{i=1} \frac{e_i}{k}$$

In [17]:

```
# but what about categorical variables ?
# let's inspect the minimum number of occurrences for the players
# notice count ignore null values
print(df_players_info.groupby('player_name')['player_name'].count().min())
```

In [18]:

```
# we will use the player "Abdeslam Ouaddou" as example
# let's search for preferred_foot values for that player
df_players_info.query('player_name == "Abdeslam Ouaddou")[['preferred_foot']]
```

Out[18]:

|     | preferred_foot |
|-----|----------------|
| 478 | right          |
| 479 | right          |
| 480 | right          |
| 481 | right          |
| 482 | right          |
| 483 | None           |

#### Comment

*We can easily replace the null value with the previous value because we are 100% sure that the minimum number of occurrences of each player is 2*

In [19]:

```
# but if all the values is null then it would be impossible that the previous two methods work
# let's give an example
df_players_info.query('player_name == "Abdeslam Ouaddou")[['vision']]
```

Out[19]:

|     | vision |
|-----|--------|
| 478 | NaN    |
| 479 | NaN    |
| 480 | NaN    |
| 481 | NaN    |
| 482 | NaN    |
| 483 | NaN    |

#### Comment

*The only two ways to deal with this problem is :*

- 1. drop the records*
- 2. Replace them with the previous available value in the previous column*

- it seems that each player appears alot but with different date
- id , player\_fifa\_api\_id and player\_api\_id are unnecessary so we will drop them
- date , birthday attributes need to be coverted to date datatype
- attacking\_work\_rate , preferred\_foot and defensive\_work\_rate attributes are categorical
- the rest of the attributes are continous random variables
- The best solution to deal with null values in the numerical attributes here to use linear interpolation or the mean with the previuos player values in the same attribute .  
For Example : "Abdeslam Ouaddou" has only one null value so we can replace it depending on the previous values
- since all players have occurred at least twice then we can use thier previous categorical value to fill nulls
- the player "Abdeslam Ouaddou" all of his vision values are null so i will drop them since no way to fill them

## Matches Dataset

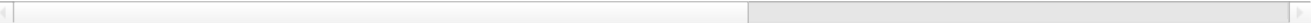
In [20]:

```
df_mathces_info.head()
```

Out[20]:

|   | home_team_short_name | home_team_long_name | away_team_short_name | away_team_long_name | league_name            | id | country_id | league_ |
|---|----------------------|---------------------|----------------------|---------------------|------------------------|----|------------|---------|
| 0 | GEN                  | KRC Genk            | BAC                  | Beerschot AC        | Belgium Jupiler League | 1  | 1          |         |
| 1 | ZUL                  | SV Zulte-Waregem    | LOK                  | Sporting Lokeren    | Belgium Jupiler League | 2  | 1          |         |
| 2 | CEB                  | KSV Cercle Brugge   | AND                  | RSC Anderlecht      | Belgium Jupiler League | 3  | 1          |         |
| 3 | GEN                  | KAA Gent            | MON                  | RAEC Mons           | Belgium Jupiler League | 4  | 1          |         |
| 4 | DEN                  | FCV Dender EH       | STL                  | Standard de Liège   | Belgium Jupiler League | 5  | 1          |         |

5 rows × 120 columns



In [21]:

```
explore_dataset(df_mathces_info)
```

we have 25979 rows and 120 columns

column name : home\_team\_short\_name

column datatype : <class 'str'>

unique values :

[ 'GEN' 'ZUL' 'CEB' 'DEN' 'MEC' 'ROS' 'TUB' 'WES' 'CLB' 'STL' 'LOK' 'MOU' 'BAC' 'AND' 'MON' 'CHA' 'KOR' 'STT' 'LIE' 'EUP' 'O-H' 'WAA' 'OOS' 'MOP' 'MUN' 'ARS' 'SUN' 'WHU' 'AVL' 'EVE' 'MID' 'BOL' 'HUL' 'CHE' 'STK' 'LIV' 'NEW' 'FUL' 'TOT' 'WBA' 'POR' 'MCI' 'BLB' 'WIG' 'WOL' 'BUR' 'BIR' 'BLA' 'QPR' 'NOR' 'SWA' 'REA' 'SOU' 'CRY' 'CAR' 'LEI' 'BOU' 'WAT' 'AUX' 'BOR' 'LEH' 'LEM' 'LYO' 'NAN' 'REN' 'SOC' 'VAL' 'LIL' 'LOR' 'MAR' 'NIC' 'ETI' 'TOU' 'CAE' 'GRE' 'PSG' 'LEN' 'ARL' 'BRE' 'AJA' 'DIJ' 'ETG' 'REI' 'TRO' 'BAS' 'GUI' 'MET' 'ANG' 'GAJ' 'BMU' 'LEV' 'S04' 'EFR' 'BIE' 'COT' 'GLA' 'KAR' 'HAM' 'WBR' 'HBE' 'FCK' 'BOC' 'DOR' 'STU' 'HAN' 'HOF' 'FRE' 'NUR' 'MAI' 'KAI' 'STP' 'AUG' 'GRF' 'FDU' 'BRA' 'PAD' 'DAR' 'ING' 'ATA' 'CAG' 'CAT' 'CHI' 'FIO' 'ACM' 'ROM' 'SAM' 'TOR' 'UDI' 'SIE' 'REG' 'PAL' 'LAZ' 'JUV' 'NAP' 'LEC' 'INT' 'PAR' 'BAR' 'CES' 'NOV' 'PES' 'VER' 'SAS' 'EMP' 'FRO' 'CAP' 'VIT' 'ROD' 'WII' 'NEC' 'UTR' 'HER' 'SPA' 'VOL' 'ALK' 'GRA' 'GRO' 'FEY' 'HAA' 'NAC' 'HEE' 'TWE' 'PSV' 'RKC' 'VEN' 'EXC' 'ZWO' 'CAM' 'GAE' 'WIS' 'GOR' 'LEG' 'SLA' 'LOD' 'POZ' 'ARK' 'PIG' 'BEL' 'BIA' 'CKR' 'CHO' 'PWA' 'ODR' 'LGD' 'POB' 'KKI' 'ZAG' 'WID' 'POD' 'POG' 'ZAW' 'TBN' 'SCP' 'FER' 'AMA' 'RA' 'NAV' 'BEN' 'SET' 'ACA' 'ULE' 'OLH' 'B-M' 'FEI' 'GV' 'MOR' 'EST' 'ARO' 'PEN' 'BOA' 'MAD' 'TON' 'FAL' 'HEA' 'KIL' 'ABE' 'CEL' 'DUU' 'INV' 'RAN' 'MOT' 'MIR' 'HIB' 'JOH' 'DUN' 'DUF' 'OSA' 'COR' 'NUM' 'SAN' 'SPG' 'BET' 'ESP' 'BIL' 'SEV' 'VIL' 'GET' 'MAL' 'ALM' 'HUE' 'ZAR' 'TEN' 'XER' 'RAY' 'ELC' 'EIB' 'LAS' 'YB' 'AAR' 'LUZ' 'XAM' 'SIO' 'VAD' 'ZUR' 'GAL' 'THU' 'SER' 'LAU' 'LUG' ]

with total 259

column name : home\_team\_long\_name

column datatype : <class 'str'>

unique values :

[ 'KRC Genk' 'SV Zulte-Waregem' 'KSV Cercle Brugge' 'KAA Gent' 'FCV Dender EH' 'KV Mechelen' 'KSV Roeselare' 'Tubize' 'KVC Westerlo' 'Club Brugge KV' 'Standard de Liège' 'Sporting Lokeren' 'Royal Excel Mouscron' 'Beerschot AC' 'RSC Anderlecht' 'RAEC Mons' 'Sporting Charleroi' 'KV Kortrijk' 'Sint-Truidense VV' 'Lierse SK' 'KAS Eupen' 'Oud-Heverlee Leuven' 'Waasland-Beveren' 'KV Oostende' 'Manchester United' 'Arsenal' 'Sunderland' 'West Ham United' 'Aston Villa' 'Everton' 'Middlesbrough' 'Bolton Wanderers' 'Hull City' 'Chelsea' 'Stoke City' 'Liverpool' 'Newcastle United' 'Fulham' 'Tottenham Hotspur' 'West Bromwich Albion' 'Portsmouth' 'Manchester City' 'Blackburn Rovers' 'Wigan Athletic' 'Wolverhampton Wanderers' 'Burnley' 'Birmingham City' 'Blackpool' 'Queens Park Rangers' 'Norwich City' 'Swansea City' 'Reading' 'Southampton' 'Crystal Palace' 'Cardiff City' 'Leicester City' 'Bournemouth' 'Watford' 'AJ Auxerre' ]

'Girondins de Bordeaux' 'Le Havre AC' 'Le Mans FC' 'Olympique Lyonnais'  
'AS Monaco' 'AS Nancy-Lorraine' 'Stade Rennais FC'  
'FC Sochaux-Montbéliard' 'Valenciennes FC' 'LOSC Lille' 'FC Lorient'  
'Olympique de Marseille' 'OGC Nice' 'AS Saint-Étienne' 'Toulouse FC'  
'SM Caen' 'Grenoble Foot 38' 'FC Nantes' 'Paris Saint-Germain'  
'Montpellier Hérault SC' "US Boulogne Cote D'Opale" 'RC Lens'  
'AC Arles-Avignon' 'Stade Brestois 29' 'AC Ajaccio' 'Dijon FCO'  
'Évian Thonon Gaillard FC' 'Stade de Reims' 'ES Troyes AC' 'SC Bastia'  
'En Avant de Guingamp' 'FC Metz' 'Angers SCO' 'GFC Ajaccio'  
'FC Bayern Munich' 'Bayer 04 Leverkusen' 'FC Schalke 04' 'VfL Wolfsburg'  
'Eintracht Frankfurt' 'DSC Arminia Bielefeld' 'FC Energie Cottbus'  
'Borussia Mönchengladbach' 'Karlsruher SC' 'Hamburger SV'  
'SV Werder Bremen' 'Hertha BSC Berlin' '1. FC Köln' 'VfL Bochum'  
'Borussia Dortmund' 'VfB Stuttgart' 'Hannover 96' 'TSG 1899 Hoffenheim'  
'SC Freiburg' '1. FC Nürnberg' '1. FSV Mainz 05' '1. FC Kaiserslautern'  
'FC St. Pauli' 'FC Augsburg' 'SpVgg Greuther Fürth' 'Fortuna Düsseldorf'  
'Eintracht Braunschweig' 'SC Paderborn 07' 'SV Darmstadt 98'  
'FC Ingolstadt 04' 'Atalanta' 'Cagliari' 'Catania' 'Chievo Verona'  
'Fiorentina' 'Milan' 'Roma' 'Sampdoria' 'Torino' 'Udinese' 'Siena'  
'Reggio Calabria' 'Palermo' 'Lazio' 'Juventus' 'Bologna' 'Napoli' 'Lecce'  
'Inter' 'Genoa' 'Livorno' 'Parma' 'Bari' 'Cesena' 'Brescia' 'Novara'  
'Pescara' 'Hellas Verona' 'Sassuolo' 'Empoli' 'Frosinone' 'Carpi'  
'Vitesse' 'Roda JC Kerkrade' 'Willem II' 'N.E.C.' 'FC Utrecht'  
'Heracles Almelo' 'Sparta Rotterdam' 'FC Volendam' 'AZ' 'De Graafschap'  
'Ajax' 'FC Groningen' 'Feyenoord' 'ADO Den Haag' 'NAC Breda'  
'SC Heerenveen' 'FC Twente' 'PSV' 'RKC Waalwijk' 'VVV-Venlo' 'Excelsior'  
'PEC Zwolle' 'SC Cambuur' 'Go Ahead Eagles' 'FC Dordrecht' 'Wisła Kraków'  
'Polonia Bytom' 'Legia Warszawa' 'Śląsk Wrocław' 'Widzew Łódź'  
'Lech Poznań' 'Arka Gdynia' 'Piast Gliwice' 'GKS Bełchatów'  
'Jagiellonia Białystok' 'Cracovia' 'Ruch Chorzów' 'P. Warszawa'  
'Odra Wodzisław' 'Lechia Gdańsk' 'Korona Kielce' 'Zagłębie Lubin'  
'Podbeskidzie Bielsko-Biała' 'Pogoń Szczecin' 'Zawisza Bydgoszcz'  
'Górnik Łęczna' 'Termalica Bruk-Bet Nieciecza' 'FC Porto' 'Sporting CP'  
'Vitória Guimarães' 'FC Paços de Ferreira' 'Amadora' 'Rio Ave FC'  
'Leixões SC' 'Naval 1º de Maio' 'SL Benfica' 'CF Os Belenenses'  
'SC Braga' 'Trofense' 'CS Marítimo' 'Vitória Setúbal' 'CD Nacional'  
'Académica de Coimbra' 'União de Leiria, SAD' 'S.C. Olhanense'  
'SC Beira Mar' 'Portimonense' 'Feirense' 'Gil Vicente FC' 'Moreirense FC'  
'Estoril Praia' 'FC Arouca' 'FC Penafiel' 'Boavista FC'  
'Uniao da Madeira' 'Tondela' 'Falkirk' 'Heart of Midlothian' 'Kilmarnock'  
'Aberdeen' 'Celtic' 'Hamilton Academical FC' 'Dundee United'  
'Inverness Caledonian Thistle' 'Rangers' 'Motherwell' 'St. Mirren'  
'Hibernian' 'St. Johnstone FC' 'Dunfermline Athletic' 'Ross County FC'  
'Dundee FC' 'Partick Thistle F.C.' 'Valencia CF' 'CA Osasuna'  
'RC Deportivo de La Coruña' 'CD Numancia' 'Racing Santander'  
'Real Sporting de Gijón' 'Real Betis Balompié' 'RCD Espanyol'  
'Athletic Club de Bilbao' 'Atlético Madrid' 'Sevilla FC' 'Villarreal CF'  
'Real Madrid CF' 'FC Barcelona' 'Getafe CF' 'RCD Mallorca' 'UD Almería'  
'Málaga CF' 'Real Valladolid' 'RC Recreativo' 'Real Zaragoza'  
'CD Tenerife' 'Xerez Club Deportivo' 'Hércules Club de Fútbol'  
'Levante UD' 'Real Sociedad' 'Rayo Vallecano' 'Granada CF'  
'RC Celta de Vigo' 'Elche CF' 'SD Eibar' 'Córdoba CF' 'UD Las Palmas'  
'Grasshopper Club Zürich' 'BSC Young Boys' 'FC Aarau' 'FC Luzern'  
'Neuchâtel Xamax' 'AC Bellinzona' 'FC Basel' 'FC Sion' 'FC Vaduz'  
'FC Zürich' 'FC St. Gallen' 'FC Thun' 'Servette FC' 'FC Lausanne-Sports'  
'Lugano']

with total 296

column name :    away\_team\_short\_name

column datatype :    <class 'str'>

unique values :  
[ 'BAC' 'LOK' 'AND' 'MON' 'STL' 'CLB' 'KOR' 'MOU' 'CHA' 'GEN' 'WES' 'ROS'  
'CEB' 'ZUL' 'DEN' 'TUB' 'MEC' 'STT' 'LIE' 'EUP' 'O-H' 'WAA' 'OOS' 'MOP'  
'NEW' 'WBA' 'LIV' 'WIG' 'MCI' 'BLB' 'TOT' 'STK' 'FUL' 'POR' 'WHU' 'EVE'  
'CHE' 'SUN' 'HUL' 'AVL' 'ARS' 'MUN' 'MID' 'BOL' 'BIR' 'BUR' 'WOL' 'BLA'  
'SWA' 'NOR' 'QPR' 'SOU' 'REA' 'CAR' 'CRY' 'LEI' 'WAT' 'BOU' 'NAN' 'CAE'  
'NIC' 'LOR' 'TOU' 'PSG' 'LIL' 'MAR' 'GRE' 'ETI' 'LYO' 'VAL' 'SOC' 'BOR'  
'LEM' 'LEH' 'AUX' 'REN' 'LEN' 'ARL' 'BRE' 'ETG' 'DIJ' 'AJA' 'BAS' 'TRO'  
'REI' 'GUI' 'MET' 'ANG' 'GAJ' 'HAM' 'DOR' 'HAN' 'FCK' 'HBE' 'WBR' 'HOF'  
'STU' 'BOC' 'LEV' 'GLA' 'BMU' 'COT' 'S04' 'BIE' 'EFR' 'KAR' 'FRE' 'NUR'  
'MAI' 'KAI' 'STP' 'AUG' 'FDU' 'GRF' 'BRA' 'PAD' 'ING' 'DAR' 'SIE' 'LAZ'  
'REG' 'JUV' 'NAP' 'INT' 'LEC' 'PAL' 'FIO' 'TOR' 'CHI' 'CAT' 'ROM' 'SAM'  
'ACM' 'UDI' 'ATA' 'CAG' 'BAR' 'PAR' 'CES' 'NOV' 'PES' 'SAS' 'VER' 'EMP'  
'CAP' 'FRO' 'GRO' 'TWE' 'GRA' 'PSV' 'FEY' 'HAA' 'HEE' 'NAC' 'VIT' 'ROD'  
'SPA' 'VOL' 'UTR' 'HER' 'WII' 'NEC' 'ALK' 'VEN' 'RKC' 'EXC' 'ZWO' 'GAE'  
'CAM' 'POB' 'CHO' 'PWA' 'LGD' 'ODR' 'BEL' 'BIA' 'CKR' 'LOD' 'WIS' 'SLA'  
'ARK' 'PIG' 'GOR' 'POZ' 'LEG' 'ZAG' 'KKI' 'WID' 'POD' 'POG' 'ZAW' 'TBN'  
'SET' 'ACA' 'BEN' 'AMA' 'NAV' 'RA' 'SCP' 'FER' 'OLH' 'ULE' 'B-M' 'FEI'  
'GV' 'MOR' 'EST' 'ARO' 'BOA' 'PEN' 'TON' 'MAD' 'RAN' 'MOT' 'HIB' 'INV'  
'MIR' 'DUU' 'ABE' 'FAL' 'CEL' 'KIL' 'HEA' 'JOH' 'DUN' 'DUF' 'MAL' 'VIL'



'SEV' 'GET' 'HUE' 'ALM' 'SPG' 'BET' 'ESP' 'BIL' 'SAN' 'NUM' 'COR' 'OSA'  
'XER' 'TEN' 'ZAR' 'RAY' 'ELC' 'EIB' 'LAS' 'SIO' 'VAD' 'ZUR' 'YB' 'AAR'  
'LUZ' 'XAM' 'GAL' 'THU' 'LAU' 'SER' 'LUG']  
with total 259

column name : away\_team\_long\_name

column datatype : <class 'str'>

unique values :

[ 'Beerschot AC' 'Sporting Lokeren' 'RSC Anderlecht' 'RAEC Mons'  
'Standard de Liège' 'Club Brugge KV' 'KV Kortrijk' 'Royal Excel Mouscron'  
'Sporting Charleroi' 'KAA Gent' 'KVC Westerlo' 'KSV Roeselare'  
'KSV Cercle Brugge' 'KRC Genk' 'SV Zulte-Waregem' 'FCV Dender EH'  
'Tubize' 'KV Mechelen' 'Sint-Truidense VV' 'Lierse SK' 'KAS Eupen'  
'Oud-Heverlee Leuven' 'Waasland-Beveren' 'KV Oostende' 'Newcastle United'  
'West Bromwich Albion' 'Liverpool' 'Wigan Athletic' 'Manchester City'  
'Blackburn Rovers' 'Tottenham Hotspur' 'Stoke City' 'Fulham' 'Portsmouth'  
'West Ham United' 'Everton' 'Chelsea' 'Sunderland' 'Hull City'  
'Aston Villa' 'Arsenal' 'Manchester United' 'Middlesbrough'  
'Bolton Wanderers' 'Birmingham City' 'Burnley' 'Wolverhampton Wanderers'  
'Blackpool' 'Swansea City' 'Norwich City' 'Queens Park Rangers'  
'Southampton' 'Reading' 'Cardiff City' 'Crystal Palace' 'Leicester City'  
'Watford' 'Bournemouth' 'FC Nantes' 'SM Caen' 'OGC Nice' 'FC Lorient'  
'Toulouse FC' 'Paris Saint-Germain' 'LOSC Lille' 'Olympique de Marseille'  
'Grenoble Foot 38' 'AS Saint-Étienne' 'Olympique Lyonnais'  
'Valenciennes FC' 'FC Sochaux-Montbéliard' 'Girondins de Bordeaux'  
'Le Mans FC' 'AS Monaco' 'Le Havre AC' 'AJ Auxerre' 'AS Nancy-Lorraine'  
'Stade Rennais FC' 'RC Lens' 'US Boulogne Cote D'Opale'  
'Montpellier Hérault SC' 'AC Arles-Avignon' 'Stade Brestois 29'  
'Évian Thonon Gaillard FC' 'Dijon FCO' 'AC Ajaccio' 'SC Bastia'  
'ES Troyes AC' 'Stade de Reims' 'En Avant de Guingamp' 'FC Metz'  
'Angers SCO' 'GFC Ajaccio' 'Hamburger SV' 'Borussia Dortmund'  
'Hannover 96' '1. FC Köln' 'Hertha BSC Berlin' 'SV Werder Bremen'  
'TSG 1899 Hoffenheim' 'VfB Stuttgart' 'VfL Bochum' 'Bayer 04 Leverkusen'  
'Borussia Mönchengladbach' 'FC Bayern Munich' 'FC Energie Cottbus'  
'FC Schalke 04' 'DSC Arminia Bielefeld' 'VfL Wolfsburg'  
'Eintracht Frankfurt' 'Karlsruher SC' 'SC Freiburg' '1. FC Nürnberg'  
'1. FSV Mainz 05' '1. FC Kaiserslautern' 'FC St. Pauli' 'FC Augsburg'  
'Fortuna Düsseldorf' 'SpVgg Greuther Fürth' 'Eintracht Braunschweig'  
'SC Paderborn 07' 'FC Ingolstadt 04' 'SV Darmstadt 98' 'Siena' 'Lazio'  
'Genoa' 'Reggio Calabria' 'Juventus' 'Bologna' 'Napoli' 'Inter' 'Lecce'  
'Palermo' 'Fiorentina' 'Torino' 'Chievo Verona' 'Catania' 'Roma'  
'Sampdoria' 'Milan' 'Udinese' 'Atalanta' 'Cagliari' 'Bari' 'Parma'  
'Livorno' 'Cesena' 'Brescia' 'Novara' 'Pescara' 'Sassuolo'  
'Hellas Verona' 'Empoli' 'Carpi' 'Frosinone' 'FC Groningen' 'FC Twente'  
'Ajax' 'De Graafschap' 'PSV' 'Feyenoord' 'ADO Den Haag' 'SC Heerenveen'  
'NAC Breda' 'Vitesse' 'Roda JC Kerkrade' 'Sparta Rotterdam' 'FC Volendam'  
'FC Utrecht' 'Heracles Almelo' 'Willem II' 'N.E.C.' 'AZ' 'VVV-Venlo'  
'RKC Waalwijk' 'Excelsior' 'PEC Zwolle' 'Go Ahead Eagles' 'SC Cambuur'  
'FC Dordrecht' 'Polonia Bytom' 'Ruch Chorzów' 'P. Warszawa'  
'Lechia Gdańsk' 'Odra Wodzisław' 'GKS Bełchatów' 'Jagiellonia Białystok'  
'Cracovia' 'Widzew Łódź' 'Wisła Kraków' 'Śląsk Wrocław' 'Arka Gdynia'  
'Piaśt Gliwice' 'Lech Poznań' 'Legia Warszawa' 'Zagłębie Lubin'  
'Korona Kielce' 'Podbeskidzie Bielsko-Biała' 'Pogoń Szczecin'  
'Zawisza Bydgoszcz' 'Górník Łęczna' 'Termalica Bruk-Bet Nieciecza'  
'CF Os Belenenses' 'Trofense' 'Vitória Setúbal' 'SC Braga'  
'Académica de Coimbra' 'SL Benfica' 'CD Nacional' 'CS Marítimo'  
'Vitória Guimarães' 'Amadora' 'Naval 1º de Maio' 'Rio Ave FC'  
'Leixões SC' 'Sporting CP' 'FC Porto' 'FC Paços de Ferreira'  
'S.C. Olhanense' 'União de Leiria, SAD' 'Portimonense' 'SC Beira Mar'  
'Feirense' 'Gil Vicente FC' 'Moreirense FC' 'Estoril Praia' 'FC Arouca'  
'Boavista FC' 'FC Penafiel' 'Tondela' 'Uniao da Madeira' 'Rangers'  
'Motherwell' 'Hibernian' 'Inverness Caledonian Thistle' 'St. Mirren'  
'Dundee United' 'Aberdeen' 'Falkirk' 'Hamilton Academical FC' 'Celtic'  
'Kilmarnock' 'Heart of Midlothian' 'St. Johnstone FC'  
'Dunfermline Athletic' 'Dundee FC' 'Ross County FC'  
'Partick Thistle F.C.' 'RCD Mallorca' 'Villarreal CF' 'Real Madrid CF'  
'FC Barcelona' 'Sevilla FC' 'Getafe CF' 'RC Recreativo' 'Real Valladolid'  
'UD Almería' 'Málaga CF' 'Real Sporting de Gijón' 'Real Betis Balompíe'  
'RCD Espanyol' 'Atlético Madrid' 'Valencia CF' 'Athletic Club de Bilbao'  
'Racing Santander' 'CD Numancia' 'RC Deportivo de La Coruña' 'CA Osasuna'  
'Xerez Club Deportivo' 'CD Tenerife' 'Real Zaragoza'  
'Hércules Club de Fútbol' 'Real Sociedad' 'Levante UD' 'Granada CF'  
'Rayo Vallecano' 'RC Celta de Vigo' 'Elche CF' 'Córdoba CF' 'SD Eibar'  
'UD Las Palmas' 'AC Bellinzona' 'FC Basel' 'FC Sion' 'FC Vaduz'  
'FC Zürich' 'Grasshopper Club Zürich' 'BSC Young Boys' 'FC Aarau'  
'FC Luzern' 'Neuchâtel Xamax' 'FC St. Gallen' 'FC Thun'  
'FC Lausanne-Sports' 'Servette FC' 'Lugano']

with total 296

column name : league\_name

column datatype : <class 'str'>

unique values :  
['Belgium Jupiler League' 'England Premier League' 'France Ligue 1'  
'Germany 1. Bundesliga' 'Italy Serie A' 'Netherlands Eredivisie'  
'Poland Ekstraklasa' 'Portugal Liga ZON Sagres' 'Scotland Premier League'  
'Spain LIGA BBVA' 'Switzerland Super League']  
with total 11

---

column name : id

column datatype : <class 'numpy.int64'>

unique values :  
[ 1 2 3 ... 25977 25978 25979]  
with total 25979

---

column name : country\_id

column datatype : <class 'numpy.int64'>

unique values :  
[ 1 1729 4769 7809 10257 13274 15722 17642 19694 21518 24558]  
with total 11

---

column name : league\_id

column datatype : <class 'numpy.int64'>

unique values :  
[ 1 1729 4769 7809 10257 13274 15722 17642 19694 21518 24558]  
with total 11

---

column name : season

column datatype : <class 'str'>

unique values :  
['2008/2009' '2009/2010' '2010/2011' '2011/2012' '2012/2013' '2013/2014'  
'2014/2015' '2015/2016']  
with total 8

---

column name : stage

column datatype : <class 'numpy.int64'>

unique values :  
[ 1 10 11 12 13 14 15 16 17 18 19 2 20 21 22 23 24 25 26 27 28 29 3 30  
31 32 33 34 4 5 6 7 8 9 35 36 37 38]  
with total 38

---

column name : date

column datatype : <class 'str'>

unique values :  
['2008-08-17 00:00:00' '2008-08-16 00:00:00' '2008-09-24 00:00:00' ...  
'2016-05-22 00:00:00' '2016-05-25 00:00:00' '2015-08-13 00:00:00']  
with total 1694

---

column name : match\_api\_id

column datatype : <class 'numpy.int64'>

unique values :  
[ 492473 492474 492475 ... 1992093 1992094 1992095]  
with total 25979

---

column name : home\_team\_api\_id

column datatype : <class 'numpy.int64'>

unique values :

```
[ 9987 10000 9984 9991 7947 8203 9999 4049 10001 8342
 9985 9994 9996 9993 8635 9998 9986 8571 9997 9989
 6351 1773 8475 8573 274581 10260 9825 8472 8654 10252
 8668 8549 8559 8667 8455 10194 8650 10261 9879 8586
 8659 8462 8456 8655 8528 8602 8191 8658 8483 10172
 9850 10003 9798 8466 9826 8344 8197 8678 9817 8583
 9827 9746 8682 9748 9829 8481 9851 9874 9873 8639
 8689 8592 9831 9853 9941 7819 9855 9830 9847 10249
 4170 8588 108893 8521 8576 9836 4087 9837 10242 7794
 9747 8550 8121 6391 9823 8178 10189 8721 9810 9912
 8398 9788 8295 9790 8697 8177 8722 9911 9789 10269
 9904 8226 8358 8165 9905 8350 8152 8406 8357 8194
 9776 8460 8262 8234 8524 8529 8530 8533 8535 8564
 8686 9882 9804 8600 8551 8690 8540 8543 9885 9857
 9875 9888 8636 10233 8537 10167 9976 9880 9858 6269
 9878 9876 7943 8534 9891 208931 8277 9803 8525 8464
 9908 9791 8614 6601 10229 8526 8593 8674 10235 10217
 9761 10228 8611 8640 10219 9839 10218 6413 7788 6433
 6631 10265 8020 8673 8025 8244 2182 8322 8028 8569
 1957 2186 1601 2183 8242 8030 8031 8245 8021 8024
 8033 8023 8027 8019 177361 9773 9768 7844 6403 10213
 7841 6421 9809 9772 9807 10264 7992 10212 10238 10214
 10215 9771 2033 10211 9765 4064 9764 8348 7842 158085
 6547 8613 6367 188163 8596 9860 8597 8485 9925 8429
 9938 8066 8548 9927 9800 10251 8467 8457 8649 8284
 8426 10267 8371 9783 8388 8696 9869 8603 8558 8315
 9906 8302 10205 8633 8634 8305 8661 9865 9864 10281
 8479 8394 9867 9868 10278 8581 8560 8370 7878 9910
 10268 8372 7869 8306 9956 10192 9930 10199 7955 6493
 9931 10179 9824 10243 10190 10191 9777 7730 7896]
with total 299
```

---

```
column name : away_team_api_id
```

```
column datatype : <class 'numpy.int64'>
```

```
unique values :
[ 9993 9994 8635 9998 9985 8342 8571 9996 9986 9991
 10001 9999 9984 9987 10000 7947 4049 8203 9997 9989
 6351 1773 8475 8573 274581 10261 8659 8650 8528 8456
 8655 8586 10194 9879 8462 8654 8668 8455 8472 8667
 10252 9825 10260 8549 8559 8658 8191 8602 8483 10003
 9850 10172 8466 9798 8344 9826 8197 9817 8678 9830
 7819 9831 8689 9941 9847 8639 8592 9855 9853 9748
 9873 9874 9827 8682 9829 9746 8583 8481 9851 8588
 4170 10249 108893 8521 4087 9836 8576 7794 10242 9837
 9747 8550 8121 6391 9790 9789 9904 8722 8177 8697
 8226 10269 9911 8178 9788 9823 8398 10189 9912 8721
 9810 8295 8358 8165 9905 8350 8152 8406 8194 8357
 9776 8460 8234 8262 8551 8543 10233 8690 9885 9857
 9875 8636 9888 8540 8535 9804 8533 8530 8686 9882
 8564 8600 8524 8529 9976 10167 8537 9880 9858 6269
 9878 7943 9876 8534 208931 9891 8674 8611 8593 8526
 8640 10235 10217 10228 9761 8277 9803 8614 6601 9908
 9791 8525 8464 10229 9839 10219 10218 6413 6433 7788
 6631 8031 1601 2183 8030 8242 8569 1957 2186 8244
 10265 8025 8322 8028 8020 2182 8673 8021 8245 8024
 8033 8023 8027 8019 177361 9807 7992 10238 10264 10215
 9772 10214 10212 7844 10213 9809 7841 6421 9768 9773
 6403 2033 9771 9765 10211 4064 9764 8348 7842 158085
 8613 6547 188163 6367 8548 9927 10251 8066 9800 9938
 8485 8596 8429 9925 8597 9860 8467 8457 8284 8649
 8426 8661 10205 8633 8634 8302 8305 8479 10281 9865
 9864 9869 8603 8558 9906 10267 8315 8696 8388 9783
 8371 9868 9867 8394 10278 8560 8581 7878 8370 9910
 10268 7869 8372 8306 6493 9931 10179 9824 10243 9956
 10192 9930 10199 7955 10190 10191 7730 9777 7896]
with total 299
```

---

```
column name : home_team_goal
```

```
column datatype : <class 'numpy.int64'>
```

```
unique values :
[ 1 0 5 2 4 3 7 6 9 8 10]
with total 11
```

---

```
column name : away_team_goal
```

```
column datatype : <class 'numpy.int64'>
```

unique values :  
[1 0 3 2 4 5 6 7 9 8]  
with total 10

---

column name : home\_player\_x1

column datatype : <class 'numpy.float64'>

unique values :  
[nan 1. 2. 0.]  
with total 3

---

column name : home\_player\_x2

column datatype : <class 'numpy.float64'>

unique values :  
[nan 2. 4. 3. 1. 5. 6. 8. 7. 0.]  
with total 9

---

column name : home\_player\_x3

column datatype : <class 'numpy.float64'>

unique values :  
[nan 4. 6. 8. 5. 3. 7. 2. 1.]  
with total 8

---

column name : home\_player\_x4

column datatype : <class 'numpy.float64'>

unique values :  
[nan 6. 8. 4. 7. 5. 3. 2.]  
with total 7

---

column name : home\_player\_x5

column datatype : <class 'numpy.float64'>

unique values :  
[nan 8. 6. 2. 7. 1. 3. 4. 9. 5.]  
with total 9

---

column name : home\_player\_x6

column datatype : <class 'numpy.float64'>

unique values :  
[nan 2. 6. 4. 1. 3. 9. 5. 7. 8.]  
with total 9

---

column name : home\_player\_x7

column datatype : <class 'numpy.float64'>

unique values :  
[nan 4. 8. 6. 5. 3. 7. 2. 9. 1.]  
with total 9

---

column name : home\_player\_x8

column datatype : <class 'numpy.float64'>

unique values :  
[nan 6. 2. 8. 3. 7. 5. 4. 9. 1.]  
with total 9

---

column name : home\_player\_x9

column datatype : <class 'numpy.float64'>

unique values :

[nan 8. 4. 2. 6. 7. 3. 9. 5. 1.]  
with total 9

---

column name : home\_player\_x10

column datatype : <class 'numpy.float64'>

unique values :  
[nan 4. 6. 9. 5. 7. 8. 3. 2. 1.]  
with total 9

---

column name : home\_player\_x11

column datatype : <class 'numpy.float64'>

unique values :  
[nan 6. 4. 5. 7. 3. 1.]  
with total 6

---

column name : away\_player\_x1

column datatype : <class 'numpy.float64'>

unique values :  
[nan 1. 2. 6.]  
with total 3

---

column name : away\_player\_x2

column datatype : <class 'numpy.float64'>

unique values :  
[nan 2. 4. 3. 1. 6. 8. 5. 7.]  
with total 8

---

column name : away\_player\_x3

column datatype : <class 'numpy.float64'>

unique values :  
[nan 4. 6. 5. 3. 8. 2. 7. 9.]  
with total 8

---

column name : away\_player\_x4

column datatype : <class 'numpy.float64'>

unique values :  
[nan 6. 8. 2. 7. 5. 4. 3. 1.]  
with total 8

---

column name : away\_player\_x5

column datatype : <class 'numpy.float64'>

unique values :  
[nan 8. 6. 4. 2. 7. 3. 1. 9. 5.]  
with total 9

---

column name : away\_player\_x6

column datatype : <class 'numpy.float64'>

unique values :  
[nan 2. 4. 3. 9. 1. 5. 6. 7. 8.]  
with total 9

---

column name : away\_player\_x7

column datatype : <class 'numpy.float64'>

unique values :  
[nan 4. 6. 5. 3. 7. 8. 2. 1. 9.]  
with total 9

---

column name :   away\_player\_x8

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 6. 8. 7. 5. 4. 3. 9. 1. 2.]  
with total 9

---

column name :   away\_player\_x9

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 8. 2. 6. 4. 3. 5. 7. 9. 1.]  
with total 9

---

column name :   away\_player\_x10

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 4. 6. 7. 5. 9. 8. 3. 1. 2.]  
with total 9

---

column name :   away\_player\_x11

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 6. 4. 3. 7. 5. 8.]  
with total 6

---

column name :   home\_player\_y1

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 1. 3. 0.]  
with total 3

---

column name :   home\_player\_y2

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 3. 0.]  
with total 2

---

column name :   home\_player\_y3

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 3. 5.]  
with total 2

---

column name :   home\_player\_y4

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 3. 5.]  
with total 2

---

column name :   home\_player\_y5

column datatype :   <class 'numpy.float64'>

unique values :  
[nan 3. 7. 6. 5. 8.]  
with total 5

---

column name : home\_player\_y6

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 3. 6. 5. 8. 9.]  
with total 6

---

column name : home\_player\_y7

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 6. 8. 5. 3. 9.]  
with total 6

---

column name : home\_player\_y8

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 8. 6. 5. 3. 9. 10.]  
with total 7

---

column name : home\_player\_y9

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 10. 8. 6. 9. 1.]  
with total 6

---

column name : home\_player\_y10

column datatype : <class 'numpy.float64'>

unique values :  
[nan 10. 7. 8. 9. 6. 11. 3.]  
with total 7

---

column name : home\_player\_y11

column datatype : <class 'numpy.float64'>

unique values :  
[nan 10. 11. 1. 3.]  
with total 4

---

column name : away\_player\_y1

column datatype : <class 'numpy.float64'>

unique values :  
[nan 1. 3.]  
with total 2

---

column name : away\_player\_y2

column datatype : <class 'numpy.float64'>

unique values :  
[nan 3.]  
with total 1

---

column name : away\_player\_y3

column datatype : <class 'numpy.float64'>

unique values :  
[nan 3. 7.]  
with total 2

---

column name : away\_player\_y4

column datatype : <class 'numpy.float64'>

unique values :  
[nan 3. 5. 7.]  
with total 3

---

column name : away\_player\_y5

column datatype : <class 'numpy.float64'>

unique values :  
[nan 3. 7. 6. 5. 9.]  
with total 5

---

column name : away\_player\_y6

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 3. 6. 8. 5. 9. 10.]  
with total 7

---

column name : away\_player\_y7

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 6. 8. 5. 3. 9. 10.]  
with total 7

---

column name : away\_player\_y8

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 8. 6. 5. 9. 3. 10.]  
with total 7

---

column name : away\_player\_y9

column datatype : <class 'numpy.float64'>

unique values :  
[nan 7. 10. 8. 9. 6. 11. 5.]  
with total 7

---

column name : away\_player\_y10

column datatype : <class 'numpy.float64'>

unique values :  
[nan 10. 7. 8. 9. 11. 6.]  
with total 6

---

column name : away\_player\_y11

column datatype : <class 'numpy.float64'>

unique values :  
[nan 10. 11. 8. 7.]  
with total 4

---

column name : home\_player\_1

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 39890. 38327. 95597. 30934. 37990. 38391. 39153. 37900.  
36835. 38252. 34480. 33676. 37937. 38341. 38318. 104378. 148308.  
32990. 39573. 38797. 170323. 148325. 148311. 131403. 148297. 40014.  
37971. 37868. 13131. 38289. 104388. 91929. 38434. 31226. 149260.  
208699. 38962. 20747. 131408. 67949. 26594. 111025. 208718. 69825.  
11336. 37839. 37854. 37993. 156553. 206592. 26589. 38793. 277845.  
148331. 214340. 242300. 181069. 346114. 13034. 166672. 270473. 107806.  
36872. 242234. 185644. 41301. 181572. 16199. 159883. 195782. 185618.]



|         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 614867. | 242243. | 104384. | 41818.  | 128083. | 305248. | 279874. | 487334. | 34003.  |
| 32657.  | 30726.  | 23686.  | 32562.  | 36374.  | 30380.  | 31465.  | 35442.  | 23932.  |
| 23021.  | 30859.  | 35906.  | 22978.  | 23794.  | 30660.  | 24224.  | 30633.  | 30455.  |
| 36373.  | 36286.  | 24527.  | 31432.  | 30622.  | 34421.  | 24886.  | 23984.  | 24227.  |
| 23027.  | 37149.  | 33948.  | 72810.  | 30973.  | 23950.  | 46518.  | 37233.  | 24147.  |
| 36283.  | 24788.  | 23813.  | 30669.  | 23307.  | 33439.  | 24368.  | 35477.  | 24229.  |
| 24694.  | 25156.  | 169718. | 24579.  | 23265.  | 40602.  | 37193.  | 26818.  | 24064.  |
| 182917. | 24634.  | 69650.  | 157303. | 155936. | 39351.  | 35495.  | 127712. | 38617.  |
| 30989.  | 33339.  | 278285. | 26295.  | 35496.  | 30974.  | 23979.  | 24341.  | 23793.  |
| 186054. | 119079. | 22998.  | 32446.  | 30841.  | 103428. | 32496.  | 169756. | 288880. |
| 24155.  | 37770.  | 42247.  | 95336.  | 268375. | 45571.  | 289159. | 28955.  | 50065.  |
| 190142. | 109060. | 163604. | 37780.  | 188971. | 73462.  | 40714.  | 30458.  | 41186.  |
| 145039. | 26233.  | 26523.  | 26359.  | 26150.  | 40712.  | 31293.  | 26350.  | 26168.  |
| 11321.  | 26190.  | 26126.  | 41097.  | 26252.  | 30742.  | 33617.  | 26117.  | 121488. |
| 26313.  | 40709.  | 26215.  | 26343.  | 33546.  | 33688.  | 26241.  | 78904.  | 94324.  |
| 26290.  | 103468. | 27445.  | 46705.  | 165528. | 39565.  | 38230.  | 188565. | 39989.  |
| 33129.  | 25852.  | 104432. | 94289.  | 26167.  | 103372. | 186806. | 30899.  | 93341.  |
| 23075.  | 46224.  | 11320.  | 26275.  | 40549.  | 33109.  | 180284. | 111193. | 190741. |
| 26367.  | 31047.  | 27659.  | 11319.  | 279175. | 46535.  | 38805.  | 26410.  | 189181. |
| 210114. | 41195.  | 94306.  | 30900.  | 140951. | 26335.  | 41886.  | 26164.  | 167035. |
| 46741.  | 156664. | 25540.  | 25329.  | 115184. | 215311. | 11316.  | 210682. | 352860. |
| 185386. | 263511. | 210164. | 187563. | 120598. | 46704.  | 563068. | 181619. | 215307. |
| 157955. | 206711. | 602493. | 311009. | 145550. | 103485. | 698273. | 607121. | 510569. |
| 291643. | 79248.  | 210387. | 529324. | 466439. | 27284.  | 36147.  | 37322.  | 30820.  |
| 27467.  | 36058.  | 36029.  | 37377.  | 25524.  | 35991.  | 24104.  | 26173.  | 31299.  |
| 27424.  | 27358.  | 30648.  | 27313.  | 27518.  | 29303.  | 27299.  | 38668.  | 27342.  |
| 27341.  | 30703.  | 36054.  | 42909.  | 27504.  | 32653.  | 27255.  | 49586.  | 36343.  |
| 27433.  | 25018.  | 114963. | 71524.  | 94841.  | 114737. | 27488.  | 128689. | 178778. |
| 36181.  | 42898.  | 178771. | 33473.  | 114211. | 37678.  | 27447.  | 36878.  | 41087.  |
| 35875.  | 184554. | 27503.  | 215168. | 210822. | 147733. | 38628.  | 278865. | 27376.  |
| 170322. | 40992.  | 39566.  | 69149.  | 5440.   | 177698. | 287894. | 94139.  | 97746.  |
| 36479.  | 41618.  | 212815. | 37335.  | 177953. | 112978. | 240338. | 408986. | 68813.  |
| 448805. | 66757.  | 40094.  | 304355. | 73409.  | 29529.  | 69259.  | 42390.  | 39599.  |
| 37503.  | 24454.  | 24503.  | 34530.  | 35644.  | 39323.  | 42422.  | 27691.  | 27697.  |
| 39510.  | 39425.  | 25735.  | 41243.  | 39204.  | 42042.  | 27558.  | 42702.  | 39477.  |
| 30713.  | 18558.  | 19344.  | 39354.  | 30717.  | 41322.  | 39313.  | 56829.  | 39170.  |
| 35626.  | 50669.  | 27560.  | 30720.  | 30281.  | 39163.  | 41864.  | 12209.  | 30850.  |
| 49541.  | 41762.  | 39522.  | 41671.  | 39415.  | 27667.  | 32746.  | 39401.  | 414788. |
| 27615.  | 42502.  | 39769.  | 41881.  | 111741. | 43283.  | 41643.  | 39725.  | 42692.  |
| 155246. | 74354.  | 181910. | 41045.  | 157284. | 39743.  | 42306.  | 39567.  | 41405.  |
| 39713.  | 42478.  | 281762. | 27658.  | 41489.  | 42009.  | 176186. | 27677.  | 30143.  |
| 197952. | 41478.  | 105828. | 181920. | 265273. | 172157. | 150747. | 212630. | 40648.  |
| 106294. | 423223. | 165952. | 176682. | 41948.  | 178732. | 33644.  | 74028.  | 213299. |
| 41466.  | 13033.  | 39294.  | 282706. | 163255. | 182843. | 103218. | 128803. | 162885. |
| 394824. | 553849. | 618878. | 181911. | 41458.  | 655544. | 39302.  | 150464. | 42429.  |
| 37216.  | 42710.  | 26510.  | 26663.  | 5044.   | 26002.  | 42829.  | 36617.  | 26536.  |
| 34489.  | 101584. | 26003.  | 45492.  | 36846.  | 26483.  | 110154. | 42573.  | 40916.  |
| 26484.  | 26585.  | 22822.  | 150210. | 37206.  | 26670.  | 26586.  | 26576.  | 45853.  |
| 184621. | 45836.  | 150217. | 39609.  | 36013.  | 112107. | 104033. | 26547.  | 212512. |
| 27329.  | 150242. | 33682.  | 110108. | 214984. | 27031.  | 243486. | 209422. | 110136. |
| 150236. | 26508.  | 5047.   | 68422.  | 13345.  | 177152. | 231674. | 21645.  | 206809. |
| 118848. | 156546. | 5710.   | 42592.  | 181377. | 278844. | 207058. | 33330.  | 150021. |
| 209785. | 189958. | 45182.  | 161014. | 110112. | 47532.  | 553476. | 352003. | 352365. |
| 212147. | 167059. | 143923. | 561819. | 46130.  | 243482. | 129241. | 467024. | 588337. |
| 141126. | 109650. | 243490. | 47249.  | 13436.  | 68835.  | 13320.  | 13390.  | 43313.  |
| 93450.  | 115268. | 68740.  | 136539. | 13454.  | 68768.  | 13471.  | 13494.  | 69004.  |
| 15361.  | 114822. | 138159. | 42929.  | 32633.  | 75579.  | 114817. | 13473.  | 206801. |
| 115006. | 40976.  | 206822. | 17869.  | 12340.  | 264578. | 249057. | 178434. | 309050. |
| 69382.  | 12318.  | 168295. | 432411. | 275751. | 181654. | 107001. | 69594.  | 277487. |
| 206805. | 114270. | 281844. | 275760. | 13472.  | 450742. | 169906. | 178584. | 38395.  |
| 72489.  | 463957. | 308842. | 361334. | 427161. | 206641. | 463806. | 232143. | 258206. |
| 543021. | 618737. | 97363.  | 40610.  | 150077. | 97498.  | 40604.  | 19515.  | 22125.  |
| 40963.  | 33550.  | 115171. | 97365.  | 30582.  | 163843. | 45245.  | 45306.  | 69341.  |
| 150275. | 30967.  | 96844.  | 19434.  | 11716.  | 40186.  | 16362.  | 19393.  | 22418.  |
| 40841.  | 25993.  | 40959.  | 32034.  | 96836.  | 150066. | 43006.  | 181198. | 11760.  |
| 45244.  | 164317. | 163981. | 37623.  | 171981. | 185908. | 33431.  | 32039.  | 121881. |
| 35453.  | 194632. | 121854. | 280969. | 124127. | 280348. | 277693. | 210046. | 280896. |
| 177126. | 30855.  | 181215. | 164430. | 120621. | 150044. | 363364. | 301804. | 96957.  |
| 301812. | 131117. | 202215. | 186705. | 416175. | 110382. | 25811.  | 25942.  | 163887. |
| 68482.  | 426359. | 261348. | 164316. | 361717. | 355347. | 202632. | 167285. | 500575. |
| 361013. | 104986. | 30657.  | 570468. | 52160.  | 380949. | 164354. | 181433. | 571298. |
| 268306. | 163893. | 38536.  | 522846. | 581726. | 450003. | 667964. | 43230.  | 23445.  |
| 32693.  | 32418.  | 11024.  | 9183.   | 32679.  | 24985.  | 35513.  | 40583.  | 32658.  |
| 25199.  | 34213.  | 43243.  | 32584.  | 95977.  | 43282.  | 34656.  | 148483. | 22924.  |
| 46365.  | 24210.  | 32713.  | 23141.  | 24984.  | 23772.  | 11498.  | 35492.  | 33914.  |
| 23918.  | 7334.   | 69793.  | 35954.  | 50051.  | 24000.  | 14100.  | 115675. | 75348.  |
| 23080.  | 354519. | 22962.  | 35336.  | 23114.  | 109622. | 24983.  | 407747. | 299380. |
| 165518. | 289118. | 41066.  | 47559.  | 186891. | 252672. | 241996. | 470003. | 620136. |
| 40980.  | 254472. | 237562. | 168673. | 160367. | 166768. | 172489. | 180197. | 289145. |
| 528212. | 19435.  | 24591.  | 33018.  | 75195.  | 37579.  | 38520.  | 46531.  | 25563.  |
| 33764.  | 33986.  | 74736.  | 34382.  | 32678.  | 30664.  | 96613.  | 33826.  | 96623.  |
| 33958.  | 33630.  | 33827.  | 30990.  | 41708.  | 37439.  | 75390.  | 71498.  | 33845.  |
| 24494.  | 33731.  | 30992.  | 33732.  | 33753.  | 2984.   | 38427.  | 113695. | 75600.  |

```
96651. 47566. 37511. 37421. 114762. 109222. 37656. 19403. 96618.
36622. 47421. 40672. 196126. 37456. 97865. 295212. 51949. 183348.
74689. 282680. 396526. 174030. 246201. 192064. 213504. 101983. 261787.
203527. 38484. 449484. 254241. 246422. 101590. 245555. 532942. 101920.
300532. 38559. 213684. 489240. 80265. 477498. 38467. 39584. 102394.
95231. 39420. 42276. 67311. 25947. 34485. 41601. 66902. 22700.
25760. 67374. 31470. 42231. 10637. 42259. 42113. 41722. 94678.
95224. 42141. 7621. 34270. 25779. 114018. 160901. 132814. 320187.
274787. 255251. 128239. 67357. 156175. 33272. 154261. 157856. 520798.
557342. 274776. 202616. 462944. 41975. 330458. 441302.]
with total 906
```

---

column name : home\_player\_2

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 67950. 39580. ... 92252. 197757. 214344.]  
with total 2414

---

column name : home\_player\_3

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38788. 67958. ... 134132. 678384. 705484.]  
with total 2375

---

column name : home\_player\_4

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38312. 67959. ... 115700. 155075. 39646.]  
with total 2606

---

column name : home\_player\_5

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 26235. 37112. ... 406283. 239959. 733787.]  
with total 2769

---

column name : home\_player\_6

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 36393. 46004. ... 119839. 185518. 186524.]  
with total 3798

---

column name : home\_player\_7

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 148286. 164732. ... 8893. 634310. 35831.]  
with total 3422

---

column name : home\_player\_8

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 67898. 39631. ... 598355. 614540. 36785.]  
with total 4076

---

column name : home\_player\_9

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 26916. 164352. ... 566785. 115862. 140490.]  
with total 4114

column name : home\_player\_10

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38801. 38423. ... 656668. 185518. 531309.]  
with total 3642

column name : home\_player\_11

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 94289. 26502. ... 209371. 42262. 195215.]  
with total 2890

column name : away\_player\_1

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 34480. 37937. 38252. 36835. 104378. 38318. 33676. 32990.  
38341. 37990. 37900. 30934. 39153. 38327. 39890. 38391. 95597.  
12381. 39573. 148308. 148297. 38797. 37854. 148325. 148311. 131403.  
131408. 37868. 13131. 37971. 38289. 40014. 91929. 104388. 170323.  
149260. 31226. 20747. 67949. 69825. 26594. 37839. 38962. 38434.  
208718. 111025. 208699. 11336. 40281. 36873. 26589. 206592. 25199.  
148331. 37993. 156553. 38793. 242300. 181069. 214340. 346114. 13034.  
166672. 38953. 107806. 270473. 242234. 36872. 185644. 195782. 185618.  
41301. 181572. 16199. 159883. 245599. 37117. 564854. 104384. 41818.  
487334. 128083. 305248. 242243. 279874. 34003. 32657. 24224. 36373.  
30660. 34421. 31432. 30622. 30455. 23794. 30633. 36286. 36374.  
31465. 30859. 24527. 23021. 30380. 23686. 30726. 22978. 23932.  
23984. 23307. 32562. 24886. 23027. 24227. 35442. 24147. 30973.  
72810. 23950. 35906. 46518. 33948. 24788. 37233. 36283. 24368.  
30669. 33439. 23813. 24399. 35477. 24694. 25156. 24229. 169718.  
104033. 24579. 23453. 30648. 157303. 37193. 182917. 40602. 23265.  
24064. 24634. 26818. 69650. 33339. 39351. 35495. 214748. 35496.  
119079. 278285. 30989. 26295. 38617. 30974. 23793. 23979. 186054.  
40581. 32496. 30841. 32446. 22998. 103428. 169756. 288880. 42247.  
23147. 95336. 24155. 37770. 268375. 289159. 45571. 50065. 28955.  
127712. 163604. 109060. 303919. 188971. 254472. 37780. 26252. 11321.  
33617. 40709. 26168. 30742. 26150. 31293. 26190. 26350. 26167.  
26313. 26359. 30458. 26126. 41097. 41186. 40714. 26233. 40712.  
26523. 26343. 41310. 26117. 26215. 26275. 26290. 26241. 33688.  
33129. 41178. 78904. 103468. 39565. 39989. 46705. 33546. 25852.  
165528. 38230. 94324. 189181. 104432. 11316. 33769. 115160. 180284.  
39233. 33599. 186806. 30899. 72623. 51409. 40549. 46224. 11320.  
46704. 190741. 111193. 26367. 26410. 11319. 279866. 27659. 279175.  
31047. 50651. 38805. 33109. 210114. 41195. 140951. 94306. 30900.  
26335. 41886. 215307. 26164. 167035. 210693. 210164. 115184. 25540.  
25329. 215311. 186811. 156664. 210682. 46689. 263511. 352860. 120598.  
187563. 563068. 181619. 185386. 513990. 145550. 103485. 79248. 510569.  
311009. 602493. 698273. 291643. 206711. 529324. 210387. 277348. 466439.  
25524. 27358. 27341. 26173. 24104. 27313. 42909. 31299. 27424.  
27518. 78279. 27284. 29303. 27299. 36058. 30820. 27467. 37377.  
36029. 36147. 37322. 35991. 27504. 30703. 32653. 36054. 27255.  
38668. 49586. 37376. 27433. 25018. 71524. 128689. 94841. 36343.  
114737. 27488. 114963. 178771. 42898. 36181. 37678. 27447. 178778.  
114211. 33473. 36878. 41087. 35875. 69149. 184554. 210822. 215168.  
27503. 147733. 278865. 212873. 38628. 27376. 170322. 40992. 39566.  
177698. 287894. 94139. 97746. 434259. 5440. 40094. 355347. 248437.  
36479. 112978. 177953. 41618. 37335. 212815. 245433. 408986. 66757.  
68813. 448805. 73409. 94658. 304355. 27691. 25735. 42702. 39510.  
30717. 41243. 42042. 27558. 30713. 24503. 39204. 39477. 37503.  
39425. 27697. 34530. 42422. 42390. 39599. 39323. 39354. 27615.  
19344. 41322. 24454. 35644. 56829. 18558. 39170. 35626. 50669.  
27560. 30720. 41864. 27666. 12209. 39313. 30281. 49541. 41129.  
41807. 32746. 39415. 41762. 41671. 39401. 27667. 42478. 42502.  
41881. 39567. 42692. 181920. 116243. 43283. 41643. 39769. 155246.  
213299. 39725. 190604. 39743. 42306. 74354. 157284. 41045. 41405.  
414788. 27658. 281762. 176186. 41489. 42009. 22305. 181910. 27677.  
39713. 30143. 197952. 212772. 190971. 106294. 105828. 265273. 40648.  
150747. 172157. 423223. 74028. 165952. 41948. 212630. 282774. 178732.  
176682. 128803. 163255. 282706. 13033. 41466. 182843. 162885. 103218.  
553849. 39294. 41534. 181911. 41458. 618878. 394824. 150464. 670927.  
39302. 435684. 282059. 299879. 36846. 40916. 36617. 34489. 45492.  
26483. 42829. 26663. 101584. 26002. 26510. 37216. 42710. 42573.  
110154. 26576. 5044. 26003. 26484. 37187. 150242. 26536. 36013.]

45853. 26586. 26670. 22822. 184621. 150210. 45836. 39609. 26585.  
42592. 112107. 37206. 33682. 26547. 212512. 27329. 209422. 214984.  
27031. 150236. 13345. 5047. 68422. 26508. 231674. 73462. 177152.  
209785. 181377. 21645. 206809. 118848. 5710. 156546. 110112. 278844.  
432470. 33330. 45182. 189958. 150021. 243548. 161014. 150217. 167059.  
352003. 47532. 352365. 561819. 212147. 143923. 46130. 553476. 467024.  
243482. 588337. 141126. 109650. 129241. 243490. 13436. 47249. 68835.  
136539. 13454. 115268. 68740. 43313. 24341. 13320. 68768. 93450.  
114822. 13390. 42929. 13471. 15361. 138159. 32633. 75579. 13494.  
69004. 114817. 13473. 115006. 40976. 17869. 264578. 249057. 12340.  
206822. 277487. 69341. 12318. 309050. 181654. 69382. 168295. 432411.  
206805. 275751. 178434. 107001. 69594. 281844. 114270. 38395. 13472.  
178584. 450742. 275760. 169906. 463957. 72489. 308842. 463806. 206641.  
427161. 232143. 258206. 206801. 618737. 543021. 32034. 30967. 45306.  
97363. 40610. 33550. 11746. 22125. 30582. 40963. 150077. 19515.  
40937. 97498. 163843. 45245. 40604. 96844. 19434. 11716. 43284.  
11760. 150275. 96836. 22418. 40841. 40959. 25993. 16362. 19393.  
40186. 43006. 163981. 181198. 96957. 37623. 121881. 185908. 32039.  
33431. 171981. 35453. 194632. 156714. 181215. 124127. 277693. 280969.  
280348. 210046. 177126. 150044. 280896. 30855. 121854. 164430. 164099.  
120621. 131117. 301804. 202215. 301812. 163887. 380949. 186705. 363364.  
25811. 110382. 416175. 150066. 25942. 361717. 426359. 361314. 68482.  
500575. 164316. 167285. 261348. 202632. 164354. 268306. 181433. 571298.  
570468. 30657. 38536. 522846. 581726. 163985. 450003. 24985. 43243.  
32679. 35513. 69259. 32418. 40583. 11024. 32693. 34213. 9183.  
32658. 95977. 43230. 32584. 43282. 34656. 22924. 35492. 148483.  
32713. 23141. 24210. 23772. 24984. 33693. 11498. 23445. 23918.  
33914. 7334. 69793. 35954. 50051. 24000. 14100. 35336. 315949.  
23080. 22962. 354519. 165518. 109622. 23114. 24983. 299380. 407747.  
470003. 41066. 186891. 47559. 115675. 212970. 241996. 237562. 172489.  
620136. 40980. 168673. 252672. 160367. 180197. 289145. 32845. 33845.  
33986. 34382. 33958. 96623. 96613. 33826. 33630. 46531. 25563.  
19435. 33764. 37579. 75195. 33018. 74736. 32678. 33827. 30664.  
24591. 30990. 38520. 37439. 33753. 109222. 75390. 113695. 24494.  
33731. 41708. 71498. 101590. 30992. 33732. 2984. 38427. 75600.  
96651. 37511. 37421. 114762. 37656. 19403. 33644. 151075. 97865.  
183701. 104986. 37456. 196126. 40672. 47566. 74689. 183348. 96618.  
51949. 193533. 246201. 101983. 203527. 192064. 213504. 2796. 261787.  
416409. 38484. 246422. 449484. 245555. 282680. 38467. 174030. 300532.  
101920. 38559. 396526. 477498. 489240. 532942. 80265. 39584. 102394.  
25947. 67311. 95231. 39420. 42276. 41601. 34485. 66902. 22700.  
25863. 67402. 25760. 42231. 10637. 27547. 42259. 31470. 42113.  
132814. 67374. 42098. 41722. 94678. 34270. 7621. 42141. 114018.  
33272. 25779. 255251. 274776. 160901. 274787. 156175. 128239. 154261.  
157856. 520798. 67357. 202616. 237645. 41975. 330458. 441302. 462944.]  
with total 926

column name : away\_player\_2

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38388. 38293. ... 27232. 458806. 92252.]  
with total 2504

column name : away\_player\_3

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 26458. 148313. ... 638593. 705484. 41415.]  
with total 2470

column name : away\_player\_4

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 13423. 104411. ... 115700. 173534. 656668.]  
with total 2657

column name : away\_player\_5

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38389. 148314. ... 733787. 214344. 231614.]  
with total 2884

column name : away\_player\_6

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38798. 37202. ... 186524. 358156. 143790.]  
with total 3930

column name : away\_player\_7

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 30949. 43158. ... 632356. 176298. 150084.]  
with total 3620

column name : away\_player\_8

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38253. 9307. ... 614540. 121115. 462608.]  
with total 4249

column name : away\_player\_9

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 106013. 42153. ... 538336. 176298. 140490.]  
with total 4319

column name : away\_player\_10

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 38383. 32690. ... 531309. 491794. 195215.]  
with total 3891

column name : away\_player\_11

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 46552. 38782. ... 527104. 209371. 119839.]  
with total 3040

column name : goal

column datatype : <class 'NoneType'>

unique values :  
[None  
'<goal><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>406</event\_incident\_typefk><elapsed>22</elapsed><player2>38807</player2><subtype>header</subtype><player1>37799</player1><sortorder>5</sortorder><team>10261</team><id>378998</id><n>295</n><type>goal</type><goal\_type>n</goal\_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>393</event\_incident\_typefk><elapsed>24</elapsed><player2>24154</player2><subtype>shot</subtype><player1>24148</player1><sortorder>4</sortorder><team>10260</team><id>379019</id><n>298</n><type>goal</type><goal\_type>n</goal\_type></value></goal>'  
'<goal><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>393</event\_incident\_typefk><elapsed>4</elapsed><player2>39297</player2><subtype>shot</subtype><player1>26181</player1><sortorder>2</sortorder><team>9825</team><id>375546</id><n>231</n><type>goal</type><goal\_type>n</goal\_type></value></goal>'  
...  
'<goal><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>71</event\_incident\_typefk><elapsed>35</elapsed><player1>38601</player1><sortorder>0</sortorder><team>10179</team><id>5639961</id><n>10</n><type>goal</type><goal\_type>n</goal\_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>71</event\_incident\_typefk><elapsed>60</elapsed><player1>38601</player1><sortorder>0</sortorder><team>10179</team><id>5639998</id><n>20</n><type>goal</type><goal\_type>n</goal\_type></value><value><comment>o</comment><stats><owngoals>1</owngoals></stats><event\_incident\_typefk>115</event\_incident\_typefk><elapsed>69</elapsed><player1>67349</player1><sortorder>0</sortorder><team>10179</team><id>5640012</id><n>28</n><type>goal</type><goal\_type>o</goal\_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event\_incident\_typefk>71</event\_incident\_typefk><elapsed>90</elapsed>

```
<player1>34082</player1><sortorder>0</sortorder><team>10199</team><id>5640044</id><n>33</n><type>goal</type><goal_type>n</goal_type></value></goal>'  
'<goal><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>19</elapsed><player1>25843</player1><sortorder>0</sortorder><team>10192</team><id>5639949</id><n>6</n><type>goal</type><goal_type>n</goal_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>58</elapsed><player1>245161</player1><sortorder>0</sortorder><team>10192</team><id>5640001</id><n>13</n><type>goal</type><goal_type>n</goal_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>71</elapsed><player1>37554</player1><sortorder>0</sortorder><team>10192</team><id>5640015</id><n>17</n><type>goal</type><goal_type>n</goal_type></value></goal>'  
'<goal><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>58</elapsed><player1>493418</player1><sortorder>0</sortorder><team>9824</team><id>5639993</id><n>16</n><type>goal</type><goal_type>n</goal_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>67</elapsed><player1>197757</player1><sortorder>0</sortorder><team>10243</team><id>5640008</id><n>20</n><type>goal</type><goal_type>n</goal_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>69</elapsed><player1>198082</player1><sortorder>0</sortorder><team>10243</team><id>5640010</id><n>21</n><type>goal</type><goal_type>n</goal_type></value><value><comment>n</comment><stats><goals>1</goals><shoton>1</shoton></stats><event_incident_typefk>71</event_incident_typefk><elapsed>76</elapsed><player1>121080</player1><sortorder>0</sortorder><team>10243</team><id>5640020</id><n>24</n><type>goal</type><goal_type>n</goal_type></value></goal>']  
with total 13225
```

50</elapsed><subtype>shot</subtype><player1>27267</player1><sortorder>2</sortorder><team>8659</team>  
<n>282</n><type>shoton</type><id>375680</id></value><value><stats><blocked>1</blocked></stats><event  
\_incident\_typefk>61</event\_incident\_typefk><elapsed>59</elapsed><subtype>blocked\_shot</subtype><play  
er1>36410</player1><sortorder>3</sortorder><team>9825</team><n>296</n><type>shoton</type><id>375717<  
/id></value><value><stats><blocked>1</blocked></stats><event\_incident\_typefk>61</event\_incident\_type  
fk><elapsed>66</elapsed><subtype>blocked\_shot</subtype><player1>31013</player1><sortorder>1</sorto  
er><team>9825</team><n>297</n><type>shoton</type><id>375740</id></value><value><stats><blocked>1</bl  
ocked></stats><event\_incident\_typefk>61</event\_incident\_typefk><elapsed>76</elapsed><subtype>blocked  
\_shot</subtype><player1>31088</player1><sortorder>0</sortorder><team>8659</team><n>305</n><type>shot  
on</type><id>375795</id></value><value><stats><blocked>1</blocked></stats><event\_incident\_typefk>61<  
/event\_incident\_typefk><elapsed>79</elapsed><subtype>blocked\_shot</subtype><player1>30935</player1><  
sortorder>0</sortorder><team>9825</team><n>310</n><type>shoton</type><id>375816</id></value><value><  
stats><shoton>1</shoton></stats><event\_incident\_typefk>136</event\_incident\_typefk><elapsed>82</elaps  
ed><subtype>header</subtype><player1>30843</player1><sortorder>1</sortorder><team>9825</team><n>314<  
/n><type>shoton</type><id>375831</id></value><value><stats><blocked>1</blocked></stats><event\_incid  
ent\_typefk>61</event\_incident\_typefk><elapsed>89</elapsed><subtype>blocked\_shot</subtype><player1>261  
81</player1><sortorder>1</sortorder><team>9825</team><n>321</n><type>shoton</type><id>375880</id></v  
alue></shoton>'

[illegible]



[illegible]

```
<sortorder><team>9869</team><n>354</n><type>shotoff</type><id>4960109</id></value><value><stats><br/><shotoff>1</shotoff></stats><event_incident_typefk>9</event_incident_typefk><coordinates><value>18</v  
alue><value>18</value></coordinates><elapsed>87</elapsed><subtype>distance</subtype><player1>213470<  
</player1><sortorder>1</sortorder><team>8315</team><n>398</n><type>shotoff</type><id>4960135</id></va  
lue><value><stats><shotoff>1</shotoff></stats><elapsed_plus>2</elapsed_plus><event_incident_typefk>7  
41</event_incident_typefk><coordinates><value>35</value><value>9</value></coordinates><elapsed>90</e  
lapsed><subtype>lob</subtype><player1>33871</player1><sortorder>4</sortorder><team>8315</team><n>416  
</n><type>shotoff</type><id>4960150</id></value></shotoff>'<br/><br/>'<shotoff><value><stats><shotoff>1</shotoff></stats><event_incident_typefk>46</event_incident_typef  
k><coordinates><value>35</value><value>12</value></coordinates><elapsed>12</elapsed><subtype>shot</s  
ubtype><player1>80492</player1><sortorder>7</sortorder><team>7878</team><n>93</n><type>shotoff</type  
<id>4946331</id></value><value><stats><shotoff>1</shotoff></stats><event_incident_typefk>9</event_i  
ncident_typefk><coordinates><value>28</value><value>16</value></coordinates><elapsed>43</elapsed><s  
ubtype>distance</subtype><player1>161291</player1><sortorder>2</sortorder><team>7878</team><n>332</n>  
<type>shotoff</type><id>4946748</id></value><value><stats><shotoff>1</shotoff></stats><event_inciden  
t_typefk>46</event_incident_typefk><coordinates><value>13</value><value>6</value></coordinates><elap  
sed>59</elapsed><subtype>shot</subtype><player1>534484</player1><sortorder>2</sortorder><team>7878<  
<team><n>419</n><type>shotoff</type><id>4947076</id></value><value><stats><shotoff>1</shotoff></stats  
><event_incident_typefk>46</event_incident_typefk><coordinates><value>14</value><value>9</value></co  
ordinates><elapsed>63</elapsed><subtype>shot</subtype><player1>534484</player1><sortorder>1</sortord  
er><team>7878</team><n>433</n><type>shotoff</type><id>4947095</id></value><value><stats><shotoff>1</  
shotoff></stats><event_incident_typefk>9</event_incident_typefk><coordinates><value>16</value><value  
>13</value></coordinates><elapsed>71</elapsed><subtype>distance</subtype><player1>2805</player1><so  
rtorder>0</sortorder><team>7878</team><n>500</n><type>shotoff</type><id>4947212</id></value><value><ev  
ent_incident_typefk>743</event_incident_typefk><coordinates><value>23</value><value>4</value></coord  
inates><elapsed>75</elapsed><subtype>miss_kick</subtype><player1>213711</player1><sortorder>2</sort  
order><team>7878</team><n>532</n><type>shotoff</type><id>4947299</id></value></shotoff>'<br/><br/>'<shotoff><value><stats><shotoff>1</shotoff></stats><event_incident_typefk>55</event_incident_typef  
k><coordinates><value>26</value><value>64</value></coordinates><elapsed>12</elapsed><subtype>crossba  
r</subtype><player1>52004</player1><sortorder>0</sortorder><team>8558</team><n>97</n><type>shotoff</  
type><id>4940296</id></value><value><stats><shotoff>1</shotoff></stats><event_incident_typefk>28</ev  
ent_incident_typefk><coordinates><value>18</value><value>5</value></coordinates><elapsed>19</elapsed  
><subtype>post</subtype><player1>36130</player1><sortorder>2</sortorder><team>8370</team><n>176</n>  
<type>shotoff</type><id>4940379</id></value><value><stats><shotoff>1</shotoff></stats><event_incident  
_typefk>53</event_incident_typefk><coordinates><value>23</value><value>15</value></coordinates><elap  
sed>44</elapsed><subtype>volley</subtype><player1>34104</player1><sortorder>4</sortorder><team>8370<  
<team><n>294</n><type>shotoff</type><id>4940624</id></value><value><stats><shotoff>1</shotoff></stat  
s><event_incident_typefk>317</event_incident_typefk><coordinates><value>21</value><value>55</value><  
</coordinates><elapsed>49</elapsed><subtype>deflected</subtype><player1>107930</player1><sortorder>3<  
</sortorder><team>8558</team><n>322</n><type>shotoff</type><id>4940738</id></value><value><stats><sho  
toff>1</shotoff></stats><event_incident_typefk>9</event_incident_typefk><coordinates><value>16</valu  
e><value>16</value></coordinates><elapsed>71</elapsed><subtype>distance</subtype><player1>210065</pl  
ayer1><sortorder>0</sortorder><team>8370</team><n>387</n><type>shotoff</type><id>4940963</id></value  
><value><stats><shotoff>1</shotoff></stats><event_incident_typefk>46</event_incident_typefk>< coordin  
ates><value>27</value><value>66</value></coordinates><elapsed>83</elapsed><subtype>shot</subtype><pl  
ayer1>629579</player1><sortorder>2</sortorder><team>8558</team><n>472</n><type>shotoff</type><id>494  
1072</id></value></shotoff>']<br/><br/>with total 8464
```

[illegible]



[illegible]

[illegible]









[illegible]

[illegible]

[illegible]

```

event_incident_typefk>749</event_incident_typefk><coordinates><value>42</value><value>15</value></co
ordinates><elapsed>74</elapsed><subtype>cross</subtype><player1>161291</player1><sortorder>1</sortord
er><team>7878</team><n>520</n><type>cross</type><id>4947275</id></value><value><stats><crosses>1</cr
osses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>42</value><value>6
0</value></coordinates><elapsed>80</elapsed><subtype>cross</subtype><player1>445907</player1><sortor
der>0</sortorder><team>8603</team><n>546</n><type>cross</type><id>4947387</id></value><value><stats>
<crosses>1</crosses></stats><event_incident_typefk>774</event_incident_typefk><coordinates><value>9<
/value><value>29</value></coordinates><elapsed>83</elapsed><subtype>cross</subtype><player1>254702</
player1><sortorder>1</sortorder><team>7878</team><n>570</n><type>cross</type><id>4947452</id></value
><value><stats><crosses>1</crosses></stats><event_incident_typefk>774</event_incident_typefk><coordi
nates><value>29</value><value>29</value></coordinates><elapsed>85</elapsed><subtype>cross</subtype><
player1>80492</player1><sortorder>0</sortorder><team>7878</team><n>577</n><type>cross</type><id>4947
478</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>770</event_incident
_typefk><coordinates><value>6</value><value>66</value></coordinates><elapsed>87</elapsed><subtype>cr
oss</subtype><player1>96652</player1><sortorder>1</sortorder><team>8603</team><n>589</n><type>cross<
/type><id>4947502</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>774</
event_incident_typefk><coordinates><value>5</value><value>4</value></coordinates><elapsed>88</elapse
d><subtype>cross</subtype><player1>362194</player1><sortorder>0</sortorder><team>7878</team><n>598</
n><type>cross</type><id>4947517</id></value></cross>'

```

```

/coordinates><elapsed>44/<elapsed><subtype>cross</subtype><player1>427893</player1><sortorder>2</sortorder><team>8370</team><n>293</n><type>cross</type><id>4940622</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>770</event_incident_typefk><coordinates><value>33</value><value>68</value></coordinates><elapsed>49/<elapsed><subtype>cross</subtype><player1>107930</player1><sortorder>0</sortorder><team>8558</team><n>318</n><type>cross</type><id>4940732</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>43</value><value>8</value></coordinates><elapsed>52/<elapsed><subtype>cross</subtype><player1>36130</player1><sortorder>0</sortorder><team>8370</team><n>329</n><type>cross</type><id>4940766</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>770</event_incident_typefk><coordinates><value>10</value><value>9</value></coordinates><elapsed>69/<elapsed><subtype>cross</subtype><player1>210065</player1><sortorder>0</sortorder><team>8370</team><n>382</n><type>cross</type><id>4940945</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>37</value><value>13</value></coordinates><elapsed>74/<elapsed><subtype>cross</subtype><player1>46808</player1><sortorder>3</sortorder><team>8370</team><n>411</n><type>cross</type><id>4940996</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>4</value><value>13</value></coordinates><elapsed>80/<elapsed><subtype>cross</subtype><player1>210065</player1><sortorder>0</sortorder><team>8370</team><n>458</n><type>cross</type><id>4941050</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>5</value><value>20</value></coordinates><elapsed>85/<elapsed><subtype>cross</subtype><player1>40220</player1><sortorder>0</sortorder><team>8370</team><n>478</n><type>cross</type><id>4941089</id></value><value><stats><crosses>1</crosses></stats><event_incident_typefk>7</event_incident_typefk><coordinates><value>6</value><value>14</value></coordinates><elapsed>87/<elapsed><subtype>cross</subtype><player1>40220</player1><sortorder>0</sortorder><team>8370</team><n>482</n><type>cross</type><id>4941100</id></value></cross>']
with total 8466

```

```

ubtype>cross</subtype><player1>26181</player1><sortorder>5</sortorder><team>9825</team><n>264</n><type>corner</type><id>375626</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>329</event_incident_typefk><elapsed>45</elapsed><subtype>cross</subtype><player1>31088</player1><sortorder>0</sortorder><team>8659</team><n>278</n><type>corner</type><id>375657</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>329</event_incident_typefk><elapsed>51</elapsed><subtype>cross</subtype><player1>23257</player1><sortorder>0</sortorder><team>8659</team><n>288</n><type>corner</type><id>375682</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>329</event_incident_typefk><elapsed>79</elapsed><subtype>cross</subtype><player1>30843</player1><sortorder>2</sortorder><team>9825</team><n>311</n><type>corner</type><id>375818</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>329</event_incident_typefk><elapsed>85</elapsed><subtype>cross</subtype><player1>24664</player1><sortorder>0</sortorder><team>8659</team><n>317</n><type>corner</type><id>375853</id></value></corner>'

```



```
<coordinates><elapsed>19</elapsed><subtype>short_left</subtype><player1>75004</player1><sortorder>6</sortorder><team>8370</team><n>184</n><type>corner</type><id>4940390</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>869</event_incident_typefk><coordinates><value>1</value><value>69</value></coordinates><elapsed>31</elapsed><subtype>cross_right</subtype><player1>498033</player1><sortorder>1</sortorder><team>8558</team><n>234</n><type>corner</type><id>4940494</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>867</event_incident_typefk><coordinates><value>45</value><value>69</value></coordinates><elapsed>32</elapsed><subtype>cross_left</subtype><player1>498033</player1><sortorder>1</sortorder><team>8558</team><n>240</n><type>corner</type><id>4940504</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>871</event_incident_typefk><coordinates><value>1</value><value>1</value></coordinates><elapsed>44</elapsed><subtype>short_left</subtype><player1>36130</player1><sortorder>0</sortorder><team>8370</team><n>286</n><type>corner</type><id>4940619</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>867</event_incident_typefk><coordinates><value>45</value><value>69</value></coordinates><elapsed>49</elapsed><subtype>cross_left</subtype><player1>498033</player1><sortorder>1</sortorder><team>8558</team><n>319</n><type>corner</type><id>4940734</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>873</event_incident_typefk><coordinates><value>1</value><value>69</value></coordinates><elapsed>50</elapsed><subtype>short_right</subtype><player1>498033</player1><sortorder>0</sortorder><team>8558</team><n>326</n><type>corner</type><id>4940742</id></value><value><stats><corners>1</corners></stats><event_incident_typefk>867</event_incident_typefk><coordinates><value>45</value><value>69</value></coordinates><elapsed>73</elapsed><subtype>cross_left</subtype><player1>498033</player1><sortorder>0</sortorder><team>8558</team><n>398</n><type>corner</type><id>4940990</id></value><value><stats><corners>1</corners></stats><elapsed_plus>2</elapsed_plus><event_incident_typefk>871</event_incident_typefk><coordinates><value>1</value><value>1</value></coordinates><elapsed>90</elapsed><subtype>short_left</subtype><player1>75004</player1><sortorder>4</sortorder><team>8370</team><n>499</n><type>corner</type><id>4941137</id></value></corner>']
```

with total 8465

```
'<possession><value><comment>59</comment><stats><homepos>59</homepos><awaypos>41</awaypos></stats><event_incident_typefk>352</event_incident_typefk><elapsed>22</elapsed><subtype>possession</subtype><sortorder>2</sortorder><awaypos>41</awaypos><homepos>59</homepos><n>204</n><type>special</type><id>4940420</id></value><value><comment>61</comment><stats><homepos>61</homepos><awaypos>39</awaypos></stats><event_incident_typefk>352</event_incident_typefk><elapsed>45</elapsed><subtype>possession</subtype><sortorder>0</sortorder><awaypos>39</awaypos><homepos>61</homepos><n>308</n><type>special</type><id>4940639</id></value><value><comment>60</comment><stats><homepos>60</homepos><awaypos>40</awaypos></stats><event_incident_typefk>352</event_incident_typefk><elapsed>68</elapsed><subtype>possession</subtype><sortorder>1</sortorder><awaypos>40</awaypos><homepos>60</homepos><n>381</n><type>special</type><id>4940941</id></value><value><comment>62</comment><stats><homepos>62</homepos><awaypos>38</awaypos></stats><elapsed_plus>1</elapsed_plus><event_incident_typefk>352</event_incident_typefk><elapsed>90</elapsed><subtype>possession</subtype><sortorder>2</sortorder><awaypos>38</awaypos><homepos>62</homepos><n>491</n><type>special</type><id>4941128</id></value></possession>']
```

with total 8420

---

column name : b365h

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |      |      |      |      |      |
|---|------|------|------|------|------|------|------|------|------|------|------|------|
| [ | 1.73 | 1.95 | 2.38 | 1.44 | 5.   | 4.75 | 2.1  | 3.2  | 2.25 | 1.3  | 2.6  | 1.91 |
|   | 2.9  | 1.7  | 2.35 | 6.   | 4.   | 1.36 | 2.4  | 1.75 | 1.4  | 1.83 | 2.   | 2.7  |
|   | 3.6  | 2.3  | 1.25 | 3.25 | 2.05 | 2.2  | 4.33 | 2.5  | 1.67 | 1.29 | 1.57 | 1.2  |
|   | 2.88 | 2.62 | 1.53 | 1.5  | 8.   | 2.8  | 1.61 | 3.   | 1.45 | 3.4  | 1.33 | 1.65 |
|   | 1.8  | 1.9  | 1.66 | 3.1  | 2.15 | 1.22 | 5.75 | 1.85 | 6.5  | 1.28 | 4.5  | nan  |
|   | 3.75 | 1.62 | 5.5  | 7.   | 1.14 | 1.18 | 11.  | 1.17 | 1.55 | 2.75 | 3.3  | 1.6  |
|   | 5.25 | 3.5  | 2.63 | 4.2  | 2.45 | 2.65 | 3.8  | 1.13 | 9.5  | 2.55 | 3.9  | 7.5  |
|   | 3.7  | 12.  | 2.37 | 1.72 | 4.8  | 8.5  | 2.87 | 10.  | 9.   | 13.  | 1.1  | 15.  |
|   | 1.11 | 1.38 | 4.1  | 5.8  | 1.48 | 4.9  | 2.95 | 6.25 | 4.3  | 4.6  | 1.08 | 1.06 |
|   | 1.09 | 17.  | 1.07 | 19.  | 21.  | 1.16 | 6.75 | 1.15 | 4.25 | 1.05 | 1.04 | 23.  |
|   | 26.  | 14.  | ]    |      |      |      |      |      |      |      |      |      |

with total 121

---

column name : b365d

column datatype : <class 'numpy.float64'>

unique values :

|   |     |      |      |      |      |      |      |      |      |      |     |      |
|---|-----|------|------|------|------|------|------|------|------|------|-----|------|
| [ | 3.4 | 3.2  | 3.3  | 3.75 | 3.5  | 3.25 | 5.25 | 4.75 | 4.33 | 4.5  | 4.2 | 3.6  |
|   | 6.  | 4.   | 3.8  | 6.5  | 5.5  | 5.   | 5.75 | 3.1  | nan  | 7.5  | 7.  | 3.9  |
|   | 3.7 | 4.1  | 8.   | 8.5  | 4.6  | 3.   | 10.  | 9.   | 4.4  | 6.75 | 9.5 | 4.25 |
|   | 11. | 6.25 | 2.88 | 2.9  | 2.87 | 2.8  | 2.4  | 2.38 | 2.75 | 4.8  | 15. | 13.  |
|   | 12. | 2.62 | 1.57 | 1.8  | 1.83 | 2.   | 1.62 | 1.5  | 2.1  | 1.91 | 2.3 | 2.63 |
|   | 1.7 | 2.05 | 1.4  | 1.75 | 1.53 | 2.5  | 2.7  | 2.15 | 2.25 | 1.73 | 5.2 | 7.25 |
|   | 17. | ]    |      |      |      |      |      |      |      |      |     |      |

with total 72

---

column name : b365a

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |       |      |      |      |      |
|---|------|------|------|------|------|------|------|-------|------|------|------|------|
| [ | 5.   | 3.6  | 2.75 | 7.5  | 1.65 | 1.67 | 3.3  | 2.2   | 2.88 | 9.5  | 2.5  | 4.   |
|   | 2.38 | 4.5  | 3.   | 1.57 | 1.83 | 8.5  | 2.9  | 2.25  | 4.33 | 8.   | 4.2  | 2.6  |
|   | 2.05 | 10.  | 3.5  | 2.1  | 2.8  | 3.4  | 1.75 | 7.    | 9.   | 1.85 | 5.5  | 3.2  |
|   | 6.   | 5.25 | 11.  | 3.25 | 2.3  | 2.62 | 6.5  | 1.36  | 2.4  | 13.  | 4.75 | 2.   |
|   | 1.95 | 1.66 | 2.7  | 1.72 | 2.15 | 3.8  | 3.1  | 2.37  | 5.75 | 2.45 | 2.87 | 1.55 |
|   | nan  | 1.8  | 1.6  | 1.53 | 1.73 | 15.  | 1.2  | 19.   | 1.4  | 12.  | 1.5  | 2.55 |
|   | 1.44 | 1.91 | 3.75 | 6.25 | 1.7  | 3.7  | 2.63 | 4.4   | 1.62 | 4.6  | 16.  | 17.  |
|   | 1.33 | 6.75 | 4.8  | 1.45 | 4.1  | 21.  | 23.  | 1.3   | 1.61 | 1.29 | 1.22 | 26.  |
|   | 10.5 | 14.  | 18.  | 3.9  | 29.  | 2.35 | 2.65 | 34.   | 1.25 | 1.9  | 51.  | 4.25 |
|   | 1.18 | 1.17 | 7.25 | 1.1  | 1.14 | 41.  | 1.08 | 1.13] |      |      |      |      |

with total 115

---

column name : bwh

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |      |      |      |      |      |
|---|------|------|------|------|------|------|------|------|------|------|------|------|
| [ | 1.75 | 1.8  | 2.4  | 1.4  | 5.   | 4.85 | 2.05 | 2.55 | 2.3  | 1.25 | 2.6  | 1.85 |
|   | 2.75 | 2.35 | 5.1  | 4.3  | 2.25 | 3.25 | 1.38 | 1.95 | 2.5  | 3.4  | 2.2  | 1.23 |
|   | 2.1  | 4.6  | 1.35 | 1.55 | 1.32 | 1.65 | 1.3  | 5.35 | 5.25 | 2.15 | 1.2  | 1.28 |
|   | 1.6  | 3.   | 2.95 | 3.35 | 1.45 | 7.15 | 1.5  | 1.7  | 2.9  | 1.16 | 2.8  | 2.85 |
|   | 4.4  | 2.   | 4.7  | 1.17 | 4.2  | 2.7  | 2.65 | 5.5  | 1.9  | 3.1  | 6.2  | nan  |
|   | 4.   | 3.75 | 4.8  | 1.71 | 4.35 | 1.67 | 2.45 | 4.05 | 1.15 | 3.9  | 12.  | 3.7  |
|   | 1.58 | 1.14 | 6.5  | 4.55 | 1.22 | 5.7  | 8.   | 1.48 | 3.6  | 1.18 | 1.26 | 1.66 |
|   | 1.46 | 3.3  | 5.95 | 3.2  | 3.5  | 4.9  | 1.36 | 3.45 | 3.85 | 4.5  | 6.   | 3.55 |
|   | 7.25 | 6.8  | 4.15 | 4.1  | 5.2  | 1.88 | 1.57 | 1.91 | 3.95 | 3.8  | 6.25 | 3.05 |
|   | 1.72 | 1.42 | 1.52 | 4.75 | 1.34 | 1.44 | 6.75 | 3.65 | 1.78 | 5.75 | 4.33 | 1.87 |



```

1.31 1.62 1.53 1.37 1.12 1.33 8.5 4.25 1.83 1.19 7.5 7.75
7. 10. 4.45 9. 4.65 1.73 6.15 1.27 1.41 5.4 2.28 5.3
3.15 8.25 5.85 6.35 6.95 11. 1.1 5.45 6.4 8.75 5.65 1.29
12.5 5.15 1.11 1.13 2.38 2.62 4.95 1.82 1.68 1.92 5.8 6.85
7.4 5.9 1.43 9.25 11.5 9.5 9.75 1.47 5.55 2.08 2.02 10.5
2.07 1.06 1.09 1.08 15. 14. 13. 13.5 1.24 16. 2.27 2.12
5.6 3.62 1.77 2.48 6.45 2.37 3.22 2.58 1.49 1.79 18. 8.1
6.1 1.63 1.93 21. 17. 5.05 6.65 7.05 6.55 2.17 1.98 9.65
2.22 8.35 7.35 1.07 7.7 7.95 1.56 1.05 6.3 8.2 15.5 14.5
1.54 1.96 2.68 2.18 1.03 6.9 16.5 34. 1.04 19. ]
with total 237

```

---

```
column name : bwd
```

```
column datatype : <class 'numpy.float64'>
```

```

unique values :
[ 3.35 3.3 4. 3.5 3.4 3.25 5. 3.2 3.45 3.6 3.15 3.1
 3.85 4.25 3. 4.5 4.7 5.1 3.55 3.65 3.05 3.7 3.75 6.
 5.6 5.5 4.55 nan 3.8 3.9 3.95 4.35 5.35 6.25 6.2 4.6
 5.25 4.2 5.8 6.5 4.65 5.2 4.3 4.4 4.1 4.85 4.75 4.9
 2.95 5.55 4.33 6.75 7.25 7. 5.75 4.95 3.56 4.15 8. 8.5
 8.25 8.75 7.75 4.05 6.05 5.65 7.5 2.9 9.25 2.75 2.85 2.8
 5.3 2.7 4.8 2.3 2.45 5.85 4.45 2.65 9. 9.75 10. 9.5
 6.8 5.15 6.55 6.7 5.7 6.3 5.4 6.85 6.35 12.5 10.5 11.
 6.4 2.68 2.55 3.37 3.38 8.35 3.27 2.15 1.9 2.35 1.8 2.
 1.65 2.4 2.5 1.95 1.67 2.2 5.05 3.73 6.6 7.4 3.58 5.45
 6.15 7.3 12. 5.9 3.12 15. 11.5 14. 13. 15.5 17. 16.
 13.5 19.5 ]
with total 133

```

---

```
column name : bwa
```

```
column datatype : <class 'numpy.float64'>
```

```

unique values :
[ 4.2 3.95 2.55 6.8 1.6 1.65 3.15 2.4 2.7 10. 3.8 2.25
 4. 1.55 1.68 9. 2.8 2.1 3.65 7.2 3.9 3.4 2.45 1.95
 3. 3.2 2. 2.95 3.6 3.1 1.7 2.75 7.75 8. 5.9 5.85
 2.9 8.25 4.9 8.5 3.3 3.05 6. 12. 2.15 5.65 7.15 2.35
 1.4 2.65 3.7 5.7 7. 5. 4.6 2.2 5.5 2.3 7.1 2.85
 5.75 4.5 7.5 5.15 4.3 5.2 2.6 4.1 13. 4.55 1.75 3.75
 10.5 4.8 4.05 6.5 2.5 1.5 3.85 nan 4.35 5.05 1.8 1.9
 5.25 5.8 10.25 7.3 3.55 9.5 4.7 1.52 11. 4.4 14. 6.25
 1.85 1.15 1.57 5.55 1.82 3.35 1.45 5.1 7.8 5.35 8.1 1.35
 5.3 4.25 7.25 4.65 6.85 6.3 4.15 9.1 4.75 9.55 9.75 6.75
 2.05 3.25 12.5 3.45 6.4 4.45 6.95 3.5 6.1 1.87 1.83 6.9
 6.6 11.5 1.62 6.65 6.15 8.75 1.78 1.42 6.2 4.33 15. 15.5
 16. 9.25 1.88 20. 1.67 1.91 13.5 1.72 1.48 14.5 17. 1.53
 1.25 5.4 1.3 7.4 7.85 6.05 8.9 1.58 18. 4.85 2.08 8.15
 7.65 1.22 1.47 18.5 1.63 17.5 1.33 21. 1.44 1.34 1.31 16.5
 1.37 1.36 7.6 2.62 1.73 8.85 3.62 5.45 2.42 4.95 9.35 8.3
 5.6 8.2 9.6 2.18 9.4 6.45 8.55 8.05 1.56 19.5 23. 26.
 1.26 6.7 6.35 7.7 8.65 7.9 7.45 10.75 1.2 1.28 1.24 1.23
 19. 1.18 1.38 29. 31. 34. 1.17 12.7 2.07 8.7 7.55 2.77
 8.6 7.35 6.55 2.17 7.05 2.72 1.77 4.77 5.95 1.1 9.2 9.15
 1.43 1.93 1.27 1.13 10.6 8.4 2.02 9.85 2.38 11.25 1.16 1.14
 8.35 1.19 4.22 7.95 1.66 1.12 36. 1.76 51. 41. ]
with total 261

```

---

```
column name : iwh
```

```
column datatype : <class 'numpy.float64'>
```

```

unique values :
[ 1.85 1.9 2.6 1.4 4. 3.7 2.4 2.1 1.3 1.8 2.5 1.7
 2.2 4.6 3.1 1.75 2. 2.45 3.9 2.8 3.4 2.3 1.35 3.8
 1.55 1.6 1.65 4.2 1.2 3.2 1.45 6. 1.5 1.37 1.33 2.9
 3.3 2.75 3. 1.17 4.4 1.15 4.8 5. 1.25 nan 1.27 12.
 1.22 3.5 2.85 3.6 5.4 2.7 2.55 4.3 2.25 2.65 4.1 4.5
 7. 4.7 6.5 5.5 2.15 7.3 6.1 5.1 5.6 1.95 8.5 7.5
 8. 2.35 1.43 5.9 6.7 6.3 2.05 4.9 5.3 7.6 5.2 3.35
 9.5 1.57 9. 1.12 10. 3.05 2.95 1.05 1.1 3.55 1.77 1.23
 7.2 1.42 7.8 5.7 1.32 1.53 1.28 1.47 1.18 1.67 7.1 3.85
 1.72 4.25 6.6 6.8 3.15 3.45 3.25 1.52 3.75 4.35 5.8 2.62
 4.85 2.83 7.9 4.65 8.3 1.62 11. 1.73 1.08 1.11 4.75 10.3
 13. 11.5 2.57 2.53 1.13 1.07 12.5 15. 14. 4.15 7.7 1.03
 10.5 6.2 6.4 20. ]
with total 147

```

---

column name : iwd

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |     |      |      |     |      |      |      |     |     |
|---|------|------|------|-----|------|------|-----|------|------|------|-----|-----|
| [ | 3.2  | 3.1  | 3.9  | 3.3 | 4.2  | 3.4  | 3.  | 4.   | 3.5  | 5.   | 3.8 | 3.7 |
|   | 5.5  | 6.   | 3.6  | 4.5 | nan  | 4.3  | 4.7 | 4.8  | 4.9  | 4.1  | 4.6 | 4.4 |
|   | 5.7  | 5.2  | 3.45 | 5.8 | 3.35 | 3.25 | 5.3 | 3.15 | 6.5  | 7.   | 6.2 | 10. |
|   | 7.3  | 5.4  | 3.75 | 8.  | 3.55 | 5.6  | 5.1 | 3.85 | 3.05 | 3.65 | 2.8 | 2.9 |
|   | 2.85 | 2.95 | 2.6  | 2.7 | 2.1  | 2.4  | 2.5 | 4.25 | 9.5  | 1.9  | 2.2 | 1.7 |
|   | 2.   | 7.5  | 1.55 | 2.3 | 1.85 | 1.8  | 1.5 | 1.6  | 9.   | 5.9  | 6.4 | 7.2 |
|   | 6.7  | 11.  | ]    |     |      |      |     |      |      |      |     |     |

with total 73

---

column name : iwa

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |      |      |      |      |      |
|---|------|------|------|------|------|------|------|------|------|------|------|------|
| [ | 3.5  | 2.3  | 6.   | 1.7  | 1.8  | 2.4  | 3.   | 8.   | 3.8  | 4.2  | 2.8  | 1.6  |
|   | 2.1  | 4.   | 3.1  | 2.45 | 1.75 | 3.2  | 2.2  | 1.85 | 2.6  | 7.   | 4.8  | 4.6  |
|   | 2.7  | 2.9  | 1.65 | 10.  | 4.4  | 2.   | 2.5  | 5.4  | 1.4  | 3.4  | 5.   | 3.7  |
|   | 1.9  | 2.75 | 11.  | 3.9  | 12.  | 5.8  | 1.55 | 1.5  | 9.   | nan  | 3.3  | 4.3  |
|   | 8.5  | 1.15 | 9.5  | 3.6  | 5.2  | 1.45 | 2.65 | 2.05 | 4.1  | 7.5  | 4.5  | 5.5  |
|   | 1.35 | 6.5  | 2.55 | 5.1  | 5.6  | 7.3  | 10.5 | 10.3 | 6.1  | 12.5 | 8.3  | 13.  |
|   | 4.9  | 4.7  | 7.6  | 1.3  | 1.37 | 6.7  | 2.95 | 2.85 | 5.3  | 3.35 | 1.43 | 2.15 |
|   | 3.25 | 14.  | 15.  | 1.42 | 1.33 | 22.  | 8.2  | 1.27 | 2.25 | 2.35 | 3.55 | 25.  |
|   | 5.7  | 4.45 | 3.45 | 1.47 | 1.28 | 7.8  | 7.2  | 1.67 | 1.52 | 1.57 | 6.6  | 1.72 |
|   | 1.87 | 1.95 | 16.  | 6.8  | 7.1  | 6.3  | 4.25 | 7.9  | 3.85 | 3.75 | 6.4  | 3.05 |
|   | 6.2  | 7.4  | 3.15 | 5.9  | 4.85 | 1.53 | 4.35 | 2.83 | 4.65 | 2.62 | 7.65 | 3.65 |
|   | 1.23 | 8.4  | 1.62 | 3.95 | 2.77 | 4.05 | 7.7  | 1.25 | 20.  | 1.22 | 1.2  | 2.57 |
|   | 2.53 | 1.1  | 1.77 | 9.2  | 1.63 | 1.17 | 1.12 | 4.15 | 2.73 | 18.  | 17.  | 11.5 |
|   | 1.32 | 21.  | 1.48 | 19.  | ]    |      |      |      |      |      |      |      |

with total 159

---

column name : lbh

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |      |      |      |      |      |
|---|------|------|------|------|------|------|------|------|------|------|------|------|
| [ | 1.8  | 1.9  | 2.5  | 1.44 | 4.   | 5.   | 1.83 | 2.25 | 1.25 | 1.29 | 2.6  | 1.67 |
|   | 2.2  | 5.5  | 1.36 | 2.75 | 1.73 | 1.33 | 1.91 | 3.75 | 2.   | 3.   | 2.1  | 2.4  |
|   | 1.5  | 1.57 | 1.72 | 4.5  | 4.33 | 2.3  | 1.17 | 1.62 | 2.38 | 2.88 | 2.37 | 6.   |
|   | 1.53 | 2.7  | 1.4  | 2.8  | 4.2  | 3.1  | 1.7  | 1.2  | 1.22 | 1.16 | 3.2  | 1.66 |
|   | nan  | 1.75 | 1.61 | 3.6  | 2.62 | 3.5  | 3.4  | 9.   | 1.14 | 7.   | 2.9  | 1.85 |
|   | 1.3  | 1.28 | 2.87 | 3.25 | 6.5  | 1.12 | 1.1  | 8.   | 1.6  | 2.85 | 2.05 | 2.55 |
|   | 1.18 | 3.8  | 4.75 | 2.15 | 2.45 | 4.8  | 7.5  | 1.55 | 1.95 | 4.1  | 1.48 | 3.3  |
|   | 1.35 | 1.65 | 5.75 | 11.  | 12.  | 10.  | 1.08 | 1.11 | 1.45 | 1.15 | 5.25 | 10.5 |
|   | 1.87 | 8.5  | 1.47 | 2.29 | 2.04 | 9.5  | 4.6  | 1.09 | 13.  | 17.  | 15.  | 1.13 |
|   | 1.07 | 1.77 | 1.26 | 1.86 | 3.9  | 1.06 | 1.05 | 2.35 | 4.25 | 2.65 | 2.63 | 5.2  |
|   | 6.75 | 14.  | 20.  | 1.21 | 21.  | 3.7  | 1.04 | 19.  | 26.  | 23.  | ]    |      |

with total 129

---

column name : lbd

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |      |      |      |      |      |      |      |     |      |
|---|------|------|------|------|------|------|------|------|------|------|-----|------|
| [ | 3.3  | 3.2  | 3.6  | 3.4  | 3.25 | 4.5  | 4.33 | 4.   | 3.1  | 3.5  | 3.  | 5.5  |
|   | 3.75 | 5.   | nan  | 6.   | 4.75 | 3.8  | 4.2  | 6.5  | 2.25 | 5.75 | 3.9 | 3.7  |
|   | 5.25 | 7.   | 5.2  | 4.6  | 4.1  | 2.88 | 8.   | 10.  | 7.5  | 9.   | 4.8 | 2.9  |
|   | 2.8  | 2.87 | 2.3  | 2.4  | 2.75 | 2.5  | 8.5  | 6.25 | 4.3  | 11.  | 12. | 2.6  |
|   | 1.8  | 1.9  | 2.1  | 1.91 | 1.65 | 2.7  | 2.2  | 2.38 | 1.73 | 2.   | 1.4 | 1.67 |
|   | 1.44 | 2.62 | 1.83 | 4.4  | 4.25 | 9.5  | 13.  | 2.85 | 4.59 | 10.5 | 15. | 17.  |
|   | 19.  | ]    |      |      |      |      |      |      |      |      |     |      |

with total 72

---

column name : lba

column datatype : <class 'numpy.float64'>

unique values :

|   |      |      |      |     |      |      |      |      |      |      |      |      |
|---|------|------|------|-----|------|------|------|------|------|------|------|------|
| [ | 3.75 | 3.5  | 2.5  | 6.5 | 1.72 | 1.62 | 3.6  | 2.75 | 10.  | 9.   | 2.38 | 4.5  |
|   | 2.8  | 1.5  | 1.73 | 7.  | 2.2  | 4.   | 8.   | 3.4  | 1.8  | 3.2  | 2.1  | 3.   |
|   | 2.6  | 6.   | 5.   | 2.7 | 1.61 | 1.66 | 2.62 | 12.  | 2.25 | 5.5  | 1.44 | 4.33 |
|   | 2.37 | 3.25 | 2.   | 11. | 1.67 | 2.4  | 2.88 | 1.57 | nan  | 1.83 | 2.3  | 1.53 |
|   | 1.9  | 2.87 | 1.91 | 1.2 | 1.4  | 1.85 | 8.5  | 4.2  | 13.  | 17.  | 1.29 | 3.8  |
|   | 7.5  | 1.6  | 3.1  | 9.5 | 1.95 | 2.9  | 14.  | 4.1  | 1.75 | 3.3  | 3.7  | 15.  |

```
4.75 2.15 2.05 5.75 5.25 1.48 1.36 1.33 1.3 21. 19. 26.
1.25 23. 1.7 4.8 1.22 10.5 2.29 1.45 4.6 2.55 1.55 1.65
3.9 16. 2.45 2.63 4.59 29. 1.35 34. 2.04 1.28 1.16 1.17
41. 51. 1.18 1.1 1.77 1.76 4.25 2.35 2.65 5.2 2.85 18.
20. 31. 36. 4.3 1.38 1.14 11.5 1.12 1.13]
with total 128
```

```
column name : psh
```

```
column datatype : <class 'numpy.float64'>
```

```
unique values :
[ nan 5.1 2.48 1.83 1.74 1.58 1.29 1.37 1.76 3.07
 3.15 2.19 1.79 1.65 2.09 2.63 1.24 1.53 1.55 1.99
 4.52 1.89 3.02 2.35 1.59 1.5 1.86 7.74 2.53 1.78
 3.13 1.6 1.92 1.7 2.36 1.54 5.04 2.16 2.25 1.71
 1.56 2.14 3.24 5.42 1.94 1.64 2.67 2.7 2.32 2.15
 3.28 1.33 2.93 2.44 2.18 3.61 4.85 3.34 1.43 1.96
 6.02 1.95 1.62 1.3 2.97 1.57 3.17 2.54 4.63 1.36
 1.38 1.46 4.36 2.55 2.52 5.13 6. 2.78 3.9 5.22
 3.42 1.48 2.3 3.35 2.65 3.49 1.41 2.39 1.77 1.52
 1.28 2.72 2.58 4.2 1.45 1.4 2.71 2.77 2.03 5.35
 2.46 4.57 2.69 2.4 1.67 2.22 1.47 3.18 1.81 1.44
 3.71 2.29 2.31 2.68 6.94 1.22 1.66 2.1 4.07 2.98
 1.93 2.13 4.31 2. 1.34 4.12 2.56 2.27 2.62 1.85
 2.95 1.68 3.97 4.06 2.75 1.39 4.35 2.02 3.76 5.01
 5.37 1.87 1.88 4.23 2.01 1.97 2.79 1.42 3.29 2.88
 1.61 2.45 3.82 3.92 2.61 1.63 1.98 3.04 2.59 4.45
 3.95 3.12 2.64 1.23 2.51 3.41 3. 1.84 2.84 3.69
 3.4 4.67 4.49 2.07 4.28 3.21 3.26 2.38 2.33 1.19
 1.27 3.77 4.43 2.8 4.26 2.66 4.04 2.92 4.91 2.11
 2.74 2.06 2.83 1.72 5.19 2.37 6.71 2.08 1.69 8.37
 1.21 2.04 4.64 2.57 8.66 1.2 5.38 3.19 1.49 3.53
 7.41 3.3 2.91 2.34 3.32 4.29 1.51 1.32 3.67 5.68
 4.1 3.66 2.41 2.23 1.9 7.81 2.89 4.81 4.3 2.86
 4.11 2.43 2.73 6.96 3.05 4.47 3.75 4.01 4.9 2.17
 4.34 5.67 2.42 1.8 2.99 3.88 5.09 2.9 4.89 5.45
 3.46 2.76 1.31 4.46 3.47 2.21 1.73 2.28 3.55 2.12
 3.79 3.2 4.71 2.82 5.66 3.51 4.54 2.96 1.25 3.44
 3.09 6.6 2.26 7.84 1.35 4.18 2.5 3.94 6.45 2.05
 4.62 3.59 4.22 3.03 3.39 3.85 6.26 6.83 1.26 1.91
 3.7 5.96 4.27 1.82 4.39 6.05 5.18 7.1 7.25 4.53
 3.89 5.06 1.18 3.57 7.75 4.21 2.24 5.05 3.23 5.36
 5.5 1.15 5.34 3.14 4.78 3.62 7.26 2.2 2.49 6.52
 3.31 5.24 6.13 4.4 1.75 3.93 4.65 7.48 7.52 5.73
 4. 3.56 5.65 4.59 3.78 6.66 3.65 7.62 4.48 4.87
 5.71 6.25 3.06 3.8 1.16 5.03 3.38 4.19 7.35 8.35
 4.68 5.9 4.6 6.4 4.5 5.3 3.52 3.01 3.98 6.2
 5.16 6.79 5.93 3.36 6.32 8.97 3.48 7.05 6.76 6.09
 6.42 2.85 5.31 6.59 2.47 3.33 9.03 6.58 2.94 1.13
 4.03 6.51 7.99 5.63 6.43 5.98 3.16 3.63 3.87 6.95
 5.89 9.51 8. 1.17 3.22 6.89 7.44 3.58 5.97 5.
 7.65 6.62 3.11 9.27 4.41 9.63 9.37 4.72 4.86 9.19
 8.91 4.95 4.25 8.38 2.6 3.27 5.79 3.45 6.87 5.62
 6.46 5.85 4.8 4.55 5.7 9.85 9.77 5.8 3.96 7.14
 5.54 3.72 3.08 10.2 2.87 7.23 6.88 5.55 6.77 6.21
 6.65 4.83 3.25 4.58 4.88 6.29 7.17 10.8 5.64 4.93
 5.58 4.14 4.24 6.11 8.52 6.15 3.5 5.6 9.2 4.15
 5.52 4.66 4.56 8.05 8.5 5.21 5.75 3.91 4.02 8.15
 4.79 6.68 5.25 4.75 6.24 3.37 3.6 5.51 5.14 2.81
 7.5 4.32 3.84 4.44 7.27 3.64 6.61 5.39 7.91 4.37
 4.51 6.55 4.08 4.7 8.34 7.42 4.05 3.1 5.46 5.82
 9. 5.02 3.73 8.36 10.27 5.33 9.68 3.81 16.92 9.82
 4.96 5.27 11.02 7.33 3.99 9.48 9.1 4.74 7.3 7.87
 6.18 4.97 10.47 5.84 10.88 11.07 7.98 7.94 6.01 4.38
 8.31 3.43 9.14 4.84 4.42 4.13 10.3 3.74 6.36 5.07
 11.25 4.98 7.32 3.68 1.14 7.28 3.83 7.29 5.41 6.64
 10.76 7.58 1.1 11.96 7.12 7.95 1.09 8.55 5.53 13.
 5.77 4.73 7.09 4.61 8.95 10.71 8.69 9.29 8.99 13.95
 7.15 8.4 6.17 8.61 14.24 7.83 4.09 6.37 5.2 7.68
 6.1 11. 8.25 1.11 11.9 1.12 1.07 5.88 13.51 4.17
 12.45 7.69 11.95 3.54 12.8 8.74 14.35 13.26 8.63 4.16
 10.78 9.34 15.25 8.84 13.6 9.7 5.81 5.15 14.15 18.1
 9.79 16.6 5.95 5.49 7.07 5.11 15.5 22.3 15.17 7.03
 10.55 4.82 5.56 5.4 11.8 9.92 8.82 19.24 12.9 9.55
 6.82 17.1 16.2 4.77 10.19 13.83 1.08 8.3 8.39 10.85
 6.84 8.71 7.06 16.77 18.7 13.31 15. 6.74 12.22 5.92
 9.74 6.56 6.7 6.27 9.78 1.909 5.94 23.6 18.5 20.
 4.33 9.76 4.69 8.64 10.33 8.42 8.68 9.46 7.04 6.5
 3.86 8.7 7.88 4.99 9.3 8.01 9.36 10.26 9.15 5.76
 16.43 5.28 5.99 11.93 10.37 5.12 7.46 8.75 4.76 8.03
 6.93 7.7 13.91 8.07 6.54 7.71 14.85 5.48 7.64 12.72
```

|       |       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 6.99  | 6.19  | 6.41  | 5.17  | 6.57  | 7.82  | 7.16  | 5.59  | 12.5  | 9.5   |
| 9.44  | 6.22  | 7.67  | 11.28 | 6.34  | 7.47  | 9.45  | 11.26 | 6.12  | 7.01  |
| 6.8   | 5.87  | 5.83  | 6.69  | 8.24  | 9.72  | 6.06  | 6.78  | 5.29  | 6.08  |
| 4.94  | 8.48  | 11.15 | 10.25 | 8.2   | 5.26  | 11.01 | 5.57  | 8.67  | 6.48  |
| 10.02 | 6.75  | 5.23  | 5.69  | 7.86  | 6.33  | 10.79 | 8.23  | 20.5  | 6.9   |
| 10.56 | 13.58 | 15.39 | 5.74  | 11.3  | 11.7  | 10.53 | 9.75  | 10.   | 13.5  |
| 10.82 | 12.   | 13.59 | 9.66  | 7.19  | 8.85  | 13.65 | 9.54  | 6.3   | 14.4  |
| 9.24  | 9.52  | 14.   | 8.93  | 7.8   | 6.14  | 5.08  | 12.1  | 5.91  | 12.48 |
| 7.89  | 8.89  | 15.77 | 8.8   | 11.69 | 5.43  | 7.93  | 13.43 | 14.49 | 12.74 |
| 26.   | 7.02  | 9.38  | 8.78  | 15.9  | 17.51 | 8.47  | 9.96  | 15.66 | 9.91  |
| 8.14  | 10.61 | 15.3  | 16.8  | 19.   | 14.79 | 11.11 | 6.49  | 8.29  | 9.64  |
| 11.29 | 12.47 | 25.5  | 6.81  | 7.96  | 7.72  | 11.6  | 8.08  | 13.81 | 10.65 |
| 10.08 | 10.7  | 11.55 | 9.07  | 15.1  | 11.21 | 9.53  | 10.93 | 9.98  | 14.3  |
| 11.62 | 9.62  | 17.52 | 11.82 | 14.9  | 7.66  | 6.63  | 9.11  | 7.08  | 8.32  |
| 11.5  | 7.34  | 8.92  | 7.13  | 9.59  | 10.09 | 1.188 | 1.794 | 1.862 | 7.76  |
| 9.43  | 11.33 | 6.73  | 10.39 | 12.36 | 14.76 | 12.2  | 6.44  | 9.28  | 6.92  |
| 11.77 | 10.69 | 12.77 | 9.08  | 10.96 | 13.48 | 13.4  | 7.59  | 10.24 | 12.18 |
| 11.45 | 11.24 | 13.7  | 15.37 | 10.81 | 11.04 | 9.41  | 7.78  | 8.1   | 11.19 |
| 10.92 | 11.1  | 11.39 | 11.22 | 12.65 | 11.42 | 13.78 | 8.59  | 5.47  | 8.81  |
| 19.86 | 9.25  | 13.07 | 8.22  | 1.06  | 16.19 | 12.62 | 1.05  | 9.61  | 8.56  |
| 12.43 | 14.06 | 7.2   | 16.81 | 14.2  | 16.85 | 18.09 | 10.03 | 14.45 | 21.08 |
| 6.04  | 17.05 | 6.07  | 16.4  | 21.2  | 10.21 | 17.   | 14.17 | 23.   | 16.   |
| 20.85 | 22.4  | 12.24 | 36.   | 11.58 | 18.43 | 5.44  | 10.9  | 18.07 | 12.98 |
| 11.4  | 6.97  | 10.05 | 21.1  | 6.85  | 15.2  | 11.08 | 18.95 | 7.45  | 18.   |
| 16.56 | 13.2  | 11.2  | 15.6  | 25.51 | 1.04  | 18.89 | 13.27 | 31.13 | ]     |

with total 948

---

```
column name :      psd

column datatype :   <class 'numpy.float64'>
```

```
unique values :
[ nan  3.82  3.52  3.79  4.27  5.74  5.16  3.93  4.   3.64  3.63  3.81
  4.01  3.51  3.56  6.71  4.6  4.44  3.76  3.88  3.59  3.57  4.4  4.79
  4.39  4.04  4.7  3.95  3.58  4.37  4.02  3.7  4.43  3.87  3.62  3.96
  4.46  3.71  4.08  3.68  3.97  4.26  4.38  4.3  3.47  3.61  5.65  3.49
  3.67  4.2  3.66  4.11  3.69  5.35  4.9  4.58  3.73  4.18  4.53  6.
  3.48  4.42  3.55  3.54  5.32  5.2  4.72  3.35  4.14  3.89  5.94  4.31
  3.84  3.65  4.65  3.3  4.95  3.5  3.53  6.03  4.35  4.85  5.09  3.45
  3.75  4.05  3.6  3.74  3.94  5.08  3.36  4.52  7.19  3.44  3.72  4.28
  5.41  4.15  3.34  5.45  3.4  5.58  3.9  6.2  4.03  4.24  3.85  5.13
  4.99  4.12  3.8  3.41  4.84  3.46  3.29  4.98  6.67  4.33  3.43  4.25
  4.22  3.86  4.09  4.54  4.1  3.31  6.23  4.47  5.28  7.63  6.09  3.77
  4.73  3.78  4.62  5.95  3.92  3.32  4.86  7.15  3.19  4.91  7.64  3.39
  4.06  4.82  7.75  4.67  3.38  5.43  5.25  4.5  5.52  3.42  4.74  7.42
  5.66  5.68  6.02  5.55  4.69  7.28  4.51  4.19  3.37  4.88  7.52  3.83
  6.82  4.57  5.04  4.63  3.91  5.44  6.75  4.29  4.45  4.93  6.05  6.87
  4.64  4.55  4.36  4.17  7.25  6.55  3.21  3.2  6.69  5.03  3.22  5.72
  4.68  7.41  4.34  6.06  4.07  3.98  4.48  6.42  5.76  4.76  4.16  7.5
  4.56  7.49  3.99  8.2  6.68  4.8  4.83  3.26  6.76  9.64  3.25  4.59
  6.8  3.24  5.6  4.94  5.83  7.2  5.05  5.77  5.4  5.75  6.66  7.02
  4.78  5.47  5.26  5.73  6.07  6.15  7.65  5.15  9.07  5.42  5.5  3.33
  6.54  6.38  4.61  6.16  4.96  5.62  5.9  3.12  6.77  5.69  6.22  5.01
  7.11  5.99  4.77  5.48  5.79  5.18  7.4  5.86  7.9  6.25  5.06  4.32
  4.21  3.27  4.13  7.  6.26  3.28  7.7  3.18  7.8  5.  5.21  4.49
  10.4  4.75  6.32  6.9  10.2  3.17  7.45  6.88  9.12  5.3  6.01  8.6
  5.29  8.65  11.03  5.89  8.04  5.23  3.16  3.23  9.47  5.92  5.8  4.71
  7.73  8.1  5.51  6.45  6.29  7.12  5.61  5.02  8.03  8.8  5.93  7.04
  5.87  4.23  4.81  5.1  6.21  6.64  5.36  5.12  4.41  7.92  5.7  6.4
  5.56  8.55  4.92  8.  6.14  5.98  3.14  6.63  5.85  5.46  5.33  6.33
  7.05  4.87  5.07  6.47  5.34  3.04  3.13  7.08  6.35  5.91  6.28  9.35
  6.12  3.08  3.11  3.15  3.1  4.66  8.5  7.61  3.09  7.69  2.93  2.96
  3.07  3.  4.97  8.75  3.06  6.61  6.89  5.63  3.02  5.19  9.2  6.3
  6.34  6.37  5.31  2.94  5.96  5.39  2.97  7.51  4.89  2.98  2.95  3.03
  6.46  9.  2.99  5.38  3.01  5.24  6.65  6.11  8.7  9.98  7.13  9.96
  6.08  3.05  5.64  2.9  11.5  10.  12.  7.83  5.22  7.09  5.11  6.81
  5.88  8.46  8.11  6.36  8.26  12.3  7.26  6.31  6.6  5.71  9.68  6.85
  7.31  5.78  7.79  5.14  8.91  5.57  6.59  10.08  10.87  15.76  13.85  11.07
  9.42  6.74  12.5  6.44  7.93  6.98  8.61  12.17  8.47  7.14  6.1  6.56
  10.1  8.45  12.14  9.04  6.78  13.5  7.38  9.8  6.92  12.55  5.59  8.4
  6.93  5.37  5.81  7.85  6.51  8.95  6.5  7.27  10.33  11.7  7.21  12.87
  8.05  8.52  12.4  5.97  5.27  13.  7.68  7.29  7.3  8.3  14.9  9.9
  10.3  12.6  7.06  5.53  6.86  14.3  9.75  6.13  9.5  6.79  5.67  7.01
  7.84  2.89  2.5  6.04  8.35  2.62  2.2  7.62  7.48  7.6  5.82  7.66
  6.91  9.21  6.24  7.33  5.17  8.25  5.84  8.82  7.32  9.67  6.7  6.52
  8.57  8.77  12.19  8.18  6.27  6.73  6.83  5.54  9.37  8.71  8.07  6.84
  8.41  6.19  8.66  7.78  5.49  6.18  7.94  7.46  8.02  6.62  6.39  7.07
  8.92  8.54  6.17  7.44  8.74  6.95  10.9  10.76  14.06  7.54  6.53  8.28
  8.86  7.1  6.43  7.47  8.06  7.99  7.81  12.83  7.35  8.69  11.79  8.15
  7.39  7.55  15.  8.76  8.81  7.97  8.32  12.2  9.09  9.27  10.8  8.67
  11.87  10.7  11.37  10.85  11.96  7.71  6.48  9.85  8.72  7.76  7.24  9.7
  7.36  9.53  10.01  6.58  8.31  13.19  13.3  6.57  8.08  12.1  9.08  13.34
```

```
7.95 10.47 10.52 9.06 16.75 10.51 14. 9.45 6.72 14.65 18.88 15.29
20.36 18.34 18.37 22.68 18.31 19. 15.99 8.48 11.39 8.23 13.15 16.85
10.39 17.49 7.34 13.52 14.32 19.8 13.4 22.5 11.6 19.75 16. 8.9
17.4 25. 13.25 19.08 10.84 13.49 8.19 15.34 15.5 14.99 9.15 11.1
18.8 11.45 20.5 11.4 15.75 12.22 8.68 20. 8.94 9.25 10.49 29.
13.64 16.5 12.8 9.55 9.84 11. ]
with total 665
```

---

column name : psa

column datatype : <class 'numpy.float64'>

unique values :  
[ nan 1.76 2.96 ... 13.21 1.13 31.5 ]  
with total 1475

---

column name : whh

column datatype : <class 'numpy.float64'>

unique values :  
[ 1.7 1.83 2.5 1.44 4.2 2.7 2.2 1.35 1.25 2.6 1.8 2.75  
5.25 4. 1.3 2.3 3. 1.4 1.9 2.05 3.25 2.15 3.2 2.1  
2.25 1.45 1.5 1.6 1.33 1.36 5. 1.22 1.65 1.57 6.5 2.4  
2. 1.73 1.67 1.17 2.9 2.8 1.62 3.8 1.12 4.33 4.5 1.91  
2.38 2.62 4.8 nan 5.5 1.29 3.75 3.6 1.53 1.95 1.2 3.5  
9. 7. 2.65 1.55 6. 1.14 1.85 2.87 1.75 3.3 3.1 1.61  
7.5 1.18 2.88 3.4 2.45 2.55 2.35 1.11 8. 2.37 4.6 1.15  
4.4 4.75 11. 8.5 9.5 10.5 3.9 10. 1.1 12. 1.28 3.7  
1.72 4.1 15. 1.08 4.35 4.25 13. 6.2 19. 17. 5.2 3.15  
1.07 21. 1.16 1.13 1.06 1.05 1.04 1.66 4.9 2.85 3.05 7.25  
3.45 5.1 1.02 5.7 23. 26. ]  
with total 125

---

column name : whd

column datatype : <class 'numpy.float64'>

unique values :  
[ 3.3 3.25 3.75 3.4 4.2 4.5 3.1 3.2 3.5 4. 4.8 4.3  
3.8 3.6 3.15 5. 6. 6.5 5.5 nan 4.33 5.3 3.9 5.1  
3. 4.6 7.5 4.75 2.9 5.25 5.75 7. 8. 4.1 9. 8.5  
9.5 4.4 2.8 2.88 2.87 2.95 2.85 2.75 2.5 3.7 2.7 1.02  
5.8 10. 2.6 1.8 1.83 2.05 1.6 1.95 1.91 1.67 2.62 4.25  
2.1 2.25 2.2 2.4 1.7 1.4 1.73 1.53 1.2 11. 3.45 3.05  
3.35 12. 6.2 13. 15. 2.38 17. ]  
with total 78

---

column name : wha

column datatype : <class 'numpy.float64'>

unique values :  
[ 4.33 3.6 2.4 6. 1.7 2.25 2.75 7. 9.5 3.75 4.2 2.8  
1.55 1.75 8.5 2.6 2.1 3.25 3. 1.95 9. 3.5 2. 2.7  
6.25 7.75 5.75 5. 3.2 2.62 2.9 7.5 8. 1.57 1.6 5.5  
4.5 1.44 2.5 4.8 2.3 10. 2.2 3.1 2.38 1.8 12. 1.67  
1.62 3.4 3.8 nan 1.53 1.83 1.73 3.3 4. 1.65 2.15 11.  
1.9 1.2 1.61 1.91 1.4 2.65 1.45 10.5 6.5 15. 2.37 13.  
2.05 2.88 1.5 1.36 4.6 1.85 2.45 17. 2.55 2.87 4.75 4.4  
19. 1.3 1.33 5.25 3.9 21. 1.25 1.28 23. 1.22 26. 3.15  
3.05 3.7 4.1 2.85 2.35 5.2 34. 1.29 3.55 2.95 1.35 5.4  
29. 1.15 6.8 5.7 6.2 1.12 4.3 5.8 1.14 1.17 1.18 1.72  
3.65 3.45 4.7 4.35 5.3 3.35 4.25 14. 5.1 4.05 6.9 6.6  
5.6 51. 41. 1.1 1.08]  
with total 136

---

column name : sjh

column datatype : <class 'numpy.float64'>

unique values :  
[ 1.9 1.95 2.63 1.44 4.5 5.5 1.91 2.6 2.2 1.27  
1.8 2.8 1.7 2.3 5.25 4.25 1.33 2.4 3.3 1.85  
1.4 1.83 2.05 3.6 2.15 2.1 3.5 2.25 4.33 1.36  
4.75 1.62 2.38 1.67 6. 1.2 3.2 3.4 1.5 6.5  
1.57 1.73 3. 2.5 1.18 3.25 2.75 1.25 5. 2.35  
1.22 2. 1.75 3.05 6.25 1.29 nan 4.2 4. 4.8

|      |      |      |       |       |       |       |       |      |      |
|------|------|------|-------|-------|-------|-------|-------|------|------|
| 1.17 | 3.1  | 1.14 | 1.53  | 3.75  | 11.   | 1.45  | 1.6   | 7.   | 4.65 |
| 1.3  | 2.88 | 8.   | 3.8   | 1.65  | 1.55  | 1.15  | 2.9   | 2.55 | 2.7  |
| 1.42 | 8.5  | 1.88 | 3.13  | 2.45  | 6.75  | 3.88  | 1.48  | 5.2  | 3.9  |
| 9.   | 7.5  | 3.12 | 1.571 | 1.909 | 1.727 | 2.375 | 1.533 | 2.62 | 5.8  |
| 4.1  | 9.5  | 15.  | 12.   | 1.13  | 10.   | 1.11  | 1.87  | 10.5 | 5.75 |
| 6.15 | 1.38 | 1.23 | 2.65  | 1.47  | 1.64  | 13.   | 1.35  | 1.08 | 1.86 |
| 1.41 | 17.  | 1.1  | 1.06  | 1.09  | 4.6   | 3.15  | 21.   | 14.  | 1.04 |
| 1.05 | 1.07 | 6.2  | 8.25  | 3.7   | 1.12  | 23.   | 19.   | ]    |      |

with total 137

---

```
column name :      sjd
```

---

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
```

|       |      |      |      |      |      |      |       |      |       |
|-------|------|------|------|------|------|------|-------|------|-------|
| [ 3.3 | 4.   | 3.5  | 3.75 | 3.4  | 5.   | 3.25 | 3.2   | 4.5  | 4.33  |
| 5.5   | 3.6  | 6.5  | 3.8  | 5.75 | 5.25 | 6.   | 3.15  | nan  | 3.1   |
| 3.35  | 4.2  | 3.65 | 3.45 | 7.5  | 3.7  | 4.75 | 4.8   | 6.75 | 7.    |
| 4.6   | 3.9  | 3.   | 5.8  | 6.25 | 3.13 | 4.1  | 5.2   | 8.   | 9.    |
| 8.5   | 3.33 | 2.88 | 2.8  | 2.75 | 2.3  | 2.4  | 2.9   | 3.05 | 3.12  |
| 2.5   | 4.25 | 10.  | 1.73 | 5.6  | 1.5  | 2.1  | 2.25  | 1.91 | 2.2   |
| 2.63  | 2.   | 1.4  | 1.83 | 1.62 | 2.7  | 1.8  | 2.375 | 4.4  | 3.85  |
| 13.   | 9.5  | 11.  | 2.6  | 6.2  | 5.4  | 12.  | 5.1   | 14.  | 15. ] |

with total 79

---

```
column name :      sjd
```

---

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
```

|       |       |        |      |      |      |      |      |       |       |
|-------|-------|--------|------|------|------|------|------|-------|-------|
| [ 4.  | 3.8   | 2.5    | 7.5  | 1.73 | 1.67 | 3.6  | 2.4  | 3.1   | 10.   |
| 4.25  | 2.38  | 4.75   | 2.88 | 1.62 | 1.8  | 9.   | 2.75 | 2.1   | 7.    |
| 3.5   | 2.6   | 1.95   | 3.3  | 3.4  | 3.2  | 1.83 | 2.8  | 8.5   | 5.5   |
| 3.    | 8.    | 5.     | 3.25 | 13.  | 5.25 | 2.25 | 2.15 | 6.5   | 1.5   |
| 6.    | 2.3   | 3.75   | 12.  | 2.05 | 2.2  | 4.5  | 11.  | 1.7   | 2.    |
| 2.95  | 4.1   | 1.53   | nan  | 4.33 | 2.7  | 9.5  | 1.75 | 2.63  | 1.57  |
| 14.   | 16.   | 1.18   | 1.45 | 1.44 | 1.85 | 1.9  | 15.  | 5.75  | 4.8   |
| 4.6   | 2.45  | 2.33   | 4.2  | 1.91 | 2.55 | 2.91 | 3.7  | 1.4   | 8.75  |
| 5.2   | 2.9   | 17.    | 1.65 | 3.9  | 1.33 | 2.62 | 5.8  | 2.875 | 1.727 |
| 3.13  | 3.12  | 6.25   | 19.  | 1.3  | 1.36 | 5.3  | 1.25 | 21.   | 1.29  |
| 29.   | 1.87  | 1.6    | 1.47 | 3.33 | 6.75 | 2.65 | 3.45 | 1.2   | 23.   |
| 1.86  | 1.88  | 26.    | 1.27 | 34.  | 1.22 | 1.1  | 1.55 | 1.64  | 1.571 |
| 1.167 | 1.615 | 18.    | 1.17 | 1.13 | 12.5 | 5.6  | 41.  | 1.38  | 25.   |
| 1.56  | 1.15  | 8.25 ] |      |      |      |      |      |       |       |

with total 132

---

```
column name :      vch
```

---

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
```

|        |       |       |       |       |       |       |       |       |       |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| [ 1.65 | 2.    | 2.35  | 1.45  | 4.5   | 4.35  | 2.1   | 2.8   | 2.25  | 1.3   |
| 1.28   | 2.65  | 2.75  | 1.7   | 2.2   | 5.5   | 3.6   | 3.25  | 1.75  | 1.35  |
| 1.8    | 2.4   | 3.75  | 1.25  | 3.2   | 4.    | 2.5   | 1.5   | 1.6   | 1.4   |
| 5.     | 1.55  | 1.2   | 2.3   | 2.6   | 3.    | 1.95  | 6.5   | 2.7   | 2.9   |
| 1.33   | 1.85  | 1.15  | 1.9   | 1.83  | 2.05  | nan   | 3.4   | 1.91  | 1.22  |
| 6.     | 1.18  | 10.   | 1.14  | 1.57  | 1.44  | 3.1   | 1.53  | 2.62  | 2.55  |
| 2.38   | 1.73  | 1.62  | 3.5   | 1.29  | 4.33  | 1.67  | 2.88  | 4.75  | 1.36  |
| 7.5    | 1.17  | 2.15  | 3.3   | 4.2   | 3.8   | 3.12  | 4.6   | 5.25  | 7.    |
| 1.12   | 5.4   | 1.364 | 1.444 | 1.667 | 1.615 | 4.333 | 9.5   | 4.3   | 4.4   |
| 1.87   | 5.75  | 2.45  | 4.8   | 4.1   | 3.9   | 3.7   | 1.92  | 6.25  | 8.    |
| 1.47   | 5.2   | 3.13  | 2.63  | 1.93  | 3.35  | 3.15  | 9.    | 1.72  | 11.   |
| 5.3    | 2.85  | 6.2   | 6.75  | 7.2   | 8.5   | 15.   | 1.11  | 1.09  | 1.1   |
| 2.875  | 1.571 | 1.909 | 1.13  | 10.5  | 13.   | 1.533 | 1.286 | 2.375 | 1.98  |
| 17.    | 12.   | 1.78  | 1.06  | 1.07  | 19.   | 1.833 | 1.952 | 2.625 | 1.083 |
| 11.5   | 18.   | 21.   | 1.08  | 20.   | 4.7   | 1.05  | 1.143 | 26.   | 23.   |
| 1.333  | 1.16  | 1.04  | 3.125 | 1.222 | 29.   | 1.925 | 22.   | 1.03  | 36.   |
| 31.    | ]     |       |       |       |       |       |       |       |       |

with total 160

---

```
column name :      vcd
```

---

```
column datatype :   <class 'numpy.float64'>
```

```
unique values :
```

|       |      |      |      |     |      |     |      |     |      |
|-------|------|------|------|-----|------|-----|------|-----|------|
| [ 3.4 | 3.25 | 3.75 | 4.35 | 4.5 | 3.3  | 3.2 | 3.5  | 4.  | 3.1  |
| 5.    | 3.6  | 5.5  | 3.8  | 4.2 | 3.7  | 6.  | nan  | 6.5 | 4.75 |
| 3.    | 4.33 | 3.12 | 4.4  | 4.6 | 5.25 | 7.5 | 7.   | 5.4 | 8.   |
| 8.5   | 5.75 | 3.9  | 5.2  | 4.3 | 4.1  | 4.8 | 6.25 | 6.1 | 3.35 |

```
3.85  3.45  6.7   5.8   4.7   4.9   10.    9.    9.5   2.5
3.13  2.9   2.8   2.88  2.75  2.25  2.38  2.4   11.5  11.
10.5  12.    13.    2.65  3.05  1.62  2.    1.73  2.62  1.7
3.125 2.2   2.3   2.45  2.15  15.   17.   19.    4.333 22.
18.   23.   26.   ]
with total 82
```

---

column name : vca

column datatype : <class 'numpy.float64'>

```
unique values :
[ 4.5  3.25  2.65  6.5  1.65  1.7  3.    2.25  2.8  8.5
 2.3  3.4  4.35  2.9  1.55  1.9  2.    7.5  4.    2.6
 1.85  9.    3.1  2.1  1.75  6.    2.75  7.    8.    1.6
 5.    11.   2.35  2.7  1.4  2.5  3.2  12.   4.75  2.2
 5.5  2.4  3.75  10.   3.5  3.6  9.5    nan  3.3  1.45
 1.5  1.2  1.8  13.   1.44  2.15  1.57  2.62  4.33  1.67
 1.62  1.91  2.88  2.38  1.95  2.05  1.73  1.83  4.6  3.8
 3.12  5.25  4.2  15.   19.   1.53  4.4  17.   5.4  26.
 1.833 6.25  23.   1.36  10.5  3.9  4.1  4.8  3.7  5.75
 2.45  2.55  4.3  11.5  1.98  5.2  1.87  3.13  1.92  2.63
 1.93  5.8  1.3  4.7  1.35  1.72  1.29  3.55  2.95  3.35
 16.   21.   5.9  3.45  4.9  1.33  1.25  1.22  29.   1.615
 4.333 2.875 20.   18.   22.   3.125 1.364 1.952 31.   36.
 41.   1.18 34.   51.   1.17 1.15  6.75  1.09  1.909 1.444
 14.   1.12 1.14  2.625 1.667 2.375 1.78  67.   1.533 46.
 1.13  1.08 ]
with total 151
```

---

column name : gbh

column datatype : <class 'numpy.float64'>

```
unique values :
[ 1.78  1.85  2.5  1.5  4.5  2.8  2.2  1.25  2.6  2.75  1.75  2.3
 5.25  4.1  1.3  3.1  1.8  1.4  1.95  3.6  2.1  1.27  2.    3.2
 4.25  2.45  1.45  1.37  4.6  1.53  1.62  2.25  1.33  1.7  2.15  4.75
 1.6  1.2  1.67  2.35  2.85  2.05  7.    1.55  1.65  3.    2.4  1.35
 2.9  1.18  3.4  2.65  1.57  2.7  2.55  1.73  1.83  6.25  nan  4.
 4.4  1.36  1.15  3.75  10.   1.47  1.9  6.5  1.22  5.5  5.    3.7
 6.    3.3  3.5  4.33  1.72  1.42  1.38  4.2  3.25  1.44  1.48  6.75
 3.15  1.28  1.14  1.87  8.5  2.95  1.91  1.34  5.75  1.26  1.88  3.9
 8.    7.25  1.16  9.    7.5  2.63  3.45  1.77  1.17  11.   1.13  3.8
 1.1  1.12  9.5  1.23  12.   7.75  2.38  3.65  3.55  1.32  3.05  1.19
 1.63  3.35  4.7  4.3  13.   9.25  1.11  10.5  2.07  1.31  17.   3.85
 1.58  1.43  5.4  1.93  8.75  5.3  21.   4.9  5.05  16.   1.82  1.52
 1.05  1.08  1.09  9.75  1.06  1.56  1.89  11.5  8.25  12.5  2.86  1.66
 1.29  1.07  15.  14.   ]
with total 159
```

---

column name : gbd

column datatype : <class 'numpy.float64'>

```
unique values :
[ 3.25  3.2  3.75  3.5  3.4  3.3  5.    3.6  4.75  4.2  3.8  4.25
 4.5  5.5  3.1  4.    6.    4.1  3.9  5.75  5.25  4.33  3.35  nan
 6.5  6.25  3.15  4.4  3.7  4.6  7.    2.95  3.    7.5  3.65  3.45
 3.33  7.25  3.85  8.    6.75  8.5  2.9  2.8  2.85  2.75  2.35  2.4
 2.65  9.25  9.    6.35  3.05  2.55  1.95  1.85  2.1  1.7  2.05  1.8
 2.6  2.25  2.3  2.2  1.75  1.45  2.5  2.7  3.55  5.35  3.95  4.3
 4.05  8.25  2.15  5.85  1.67  2.    5.05  7.75  9.75  11.   10.   8.75
 9.5 ]
with total 84
```

---

column name : gba

column datatype : <class 'numpy.float64'>

```
unique values :
[ 4.    3.75  2.5  5.5  1.65  1.7  2.25  2.8  10.   2.35  4.25  2.85
 1.6  1.8  7.5  2.75  2.15  4.2  7.    3.5  1.95  3.2  8.5  2.1
 3.    1.75  8.    1.73  6.    5.25  3.1  2.9  4.4  5.    11.   4.6
 2.3  2.2  3.3  2.6  1.4  5.75  4.75  2.7  2.45  12.   6.5  2.
 4.5  1.67  4.33  3.7  1.55  9.    nan  3.6  1.83  2.05  3.15  1.72
 4.8  2.4  7.75  3.8  3.4  4.3  13.   1.85  1.18  1.45  2.65  1.9
 6.25  3.9  1.5  1.77  3.25  2.55  9.5  4.1  1.62  1.57  15.   3.85
 1.53  16.   1.33  9.25  1.87  6.75  7.25  1.88  2.95  17.   2.38  1.48
```

```
1.3 14. 1.42 1.36 2.63 3.35 1.46 1.44 1.37 12.5 1.28 1.27
19. 1.35 21. 4.9 1.25 10.5 1.91 13.5 8.25 9.75 11.5 1.78
8.75 4.7 3.55 1.47 4.15 15.5 20. 17.5 18.5 14.5 5.6 1.63
5.2 1.2 1.34 18. 1.22 3.45 1.12 6.15 6.35 3.95 5.85 4.55
3.65 7.85 4.65 7.9 3.05 6.9 7.7 19.5 5.35 1.13 1.15 5.05
6.8 1.93 1.38 1.17 26. 1.23 25. 1.26 1.32 16.5 2.86 6.2
24. 34. 1.52 23. 29. ]
with total 172
```

---

column name : bsh

column datatype : <class 'numpy.float64'>

unique values :

```
[ 1.73  1.91  2.3   1.44  4.75   nan   2.1   2.88  2.25  1.29  1.25  2.62
 1.83  2.75  1.7   6.    3.8   1.3   3.1   1.36  1.8   2.    2.5   3.5
 2.2   4.33  4.2   1.5   1.6   1.33  4.5   5.    2.38  1.57  1.2   1.67
 3.3   2.7   7.    1.53  2.4   1.14  1.62  2.8   1.4   3.    1.75  1.17
 1.18  5.5   1.95  1.22  3.4   9.    3.6   2.15  2.6   6.5   2.05  3.25
 3.2   2.37  4.    3.75  5.4   1.28  1.12  8.    1.61  2.87  8.5   7.5
10.    1.85  1.65 11.    6.75  2.9  12.    5.25  1.9   6.25  1.1   1.11
 3.9  13.    9.5   2.45  2.17  2.63 17.    1.55 15.    5.75  1.08  1.06
 1.07 1.04 1.05 1.45 1.97 2.55]
```

with total 101

---

column name : bsd

column datatype : <class 'numpy.float64'>

unique values :

```
[ 3.4   3.25  3.2   3.75  3.3   nan   4.5   5.    3.6   4.2   3.1   3.5
 6.    4.    4.33  5.5   4.75  6.5   3.8   7.    5.4   5.25  3.7   3.
 3.9   6.75  7.5   5.75  8.    8.5   2.88  2.9   2.38  2.8   2.4   2.62
 2.75  1.91  2.1   1.67  2.    1.7   1.33  2.25  1.4   2.6   1.73  1.44
 2.5   2.7   2.15  2.2   9.    10.   13.   11.   12.   3.45  5.95  4.4 ]
```

with total 59

---

column name : bsa

column datatype : <class 'numpy.float64'>

unique values :

```
[ 4.2   3.6   2.75  6.5   1.67   nan   3.1   2.2   2.8   9.    2.38  3.8
 2.3   4.33  1.5   1.83  8.5   2.7   2.1   4.5   7.    4.    3.3   2.5
 1.91  2.9   3.    3.4   1.73  6.    5.    2.88  7.5   1.62  2.62  5.5
10.    4.75  2.    1.4   3.5  15.    2.25  2.15 11.    1.75 12.    2.6
 1.53  1.7   1.95  5.4   1.2   1.57  1.85 13.    2.4   1.44  3.2   8.
 3.75  2.05  2.37  1.61  1.8   3.25  2.87 17.    1.33  1.29 23.    19.
 1.36  9.5   5.75  5.25  3.7  21.    1.3   1.25  1.55 26.    1.22  3.9
 1.6   2.63 10.5   1.12  4.6   1.17 16.    29.   34.    1.14  1.18  2.55
 3.15]
```

with total 96

---

columns with null values count :

```
home_team_short_name    0
home_team_long_name     0
away_team_short_name    0
away_team_long_name     0
league_name             0
```

```
...
gbd                     11817
gba                     11817
bsh                     11818
bsd                     11818
bsa                     11818
```

Length: 120, dtype: int64  
with total 407395

---

columns with duplicates count :

0

---



In [22]:

```
df_mathces_info.describe()
```

Out[22]:

|       | id           | country_id   | league_id    | stage        | match_api_id | home_team_api_id | away_team_api_id | home_team_goal | away_team_goal |
|-------|--------------|--------------|--------------|--------------|--------------|------------------|------------------|----------------|----------------|
| count | 25979.000000 | 25979.000000 | 25979.000000 | 25979.000000 | 2.597900e+04 | 25979.000000     | 25979.000000     | 25979.000000   | 25979.000000   |
| mean  | 12990.000000 | 11738.630317 | 11738.630317 | 18.242773    | 1.195429e+06 | 9984.371993      | 9984.475115      | 1.544594       | 1.544594       |
| std   | 7499.635658  | 7553.936759  | 7553.936759  | 10.407354    | 4.946279e+05 | 14087.453758     | 14087.445135     | 1.297158       | 1.297158       |
| min   | 1.000000     | 1.000000     | 1.000000     | 1.000000     | 4.831290e+05 | 1601.000000      | 1601.000000      | 0.000000       | 0.000000       |
| 25%   | 6495.500000  | 4769.000000  | 4769.000000  | 9.000000     | 7.684365e+05 | 8475.000000      | 8475.000000      | 1.000000       | 1.000000       |
| 50%   | 12990.000000 | 10257.000000 | 10257.000000 | 18.000000    | 1.147511e+06 | 8697.000000      | 8697.000000      | 1.000000       | 1.000000       |
| 75%   | 19484.500000 | 17642.000000 | 17642.000000 | 27.000000    | 1.709852e+06 | 9925.000000      | 9925.000000      | 2.000000       | 2.000000       |
| max   | 25979.000000 | 24558.000000 | 24558.000000 | 38.000000    | 2.216672e+06 | 274581.000000    | 274581.000000    | 10.000000      | 10.000000      |

8 rows × 105 columns



Comments :

- The data from the matches seems missy
- Attributes like shoton , goal ,...etc are in xml format
- Attributes like sja, vch, vcd are numerical values which are odds from different websits in decimal formats
- Fortunately we don't need any of these attributes to answer our questions .so, we will drop them and just keep home\_team\_goal and away\_team\_goal ,home\_team\_name and away\_team\_name,season
- season attribute ranges from season 2008/2009 to 2015/2016
- Null values in card attribute means there is no red/yellow cards in this match or it's unknown

## Data Cleaning

Note :

This part of the section depend on the previous part. We will focus on the problems we have stated earlier and try to solve it.

### Teams\_Dataset

In [23]:

```
# first let's drop unneccessary columns
df_teams_info.drop(['id','team_fifa_api_id','team_api_id','buildupplaydribbling'],axis=1,inplace=True)
```

In [24]:

```
# change datatype for date
df_teams_info['date'] = pd.to_datetime(df_teams_info['date'])
# check for that
df_teams_info[['date']].dtypes
```

Out[24]:

```
date    datetime64[ns]
dtype: object
```

In [25]:

```
# we have to rearrange our dataset suchthat team name followed by the other attributes
# let's change short name attribute first
short_name = df_teams_info.pop('team_short_name')
df_teams_info.insert(0,'team_short_name',short_name)
# change long name attribute
long_name = df_teams_info.pop('team_long_name')
df_teams_info.insert(1,'team_long_name',long_name)
```

In [26]:

```
# check for changes
df_teams_info.head()
```

Out[26]:

|   | team_short_name | team_long_name | date       | buildupplayspeed | buildupplayspeedclass | buildupplaydribblingclass | buildupplaypassing | bui |
|---|-----------------|----------------|------------|------------------|-----------------------|---------------------------|--------------------|-----|
| 0 | AAR             | FC Aarau       | 2010-02-22 | 60               | Balanced              | Little                    | 50                 |     |
| 1 | AAR             | FC Aarau       | 2014-09-19 | 52               | Balanced              | Normal                    | 56                 |     |
| 2 | AAR             | FC Aarau       | 2015-09-10 | 47               | Balanced              | Normal                    | 54                 |     |
| 3 | ABE             | Aberdeen       | 2010-02-22 | 70               | Fast                  | Little                    | 70                 |     |
| 4 | ABE             | Aberdeen       | 2011-02-22 | 47               | Balanced              | Little                    | 52                 |     |

5 rows × 23 columns



Note :  
The dataset came from a database that's why it seems almost clean

Players\_Dataset

In [27]:

```
# first let's drop unnecessary columns
df_players_info.drop(['id', 'player_fifa_api_id'],axis=1,inplace=True)
```

In [28]:

```
# change datatype for date and birthday
df_players_info['date'] = pd.to_datetime(df_players_info['date'])
df_players_info['birthday'] = pd.to_datetime(df_players_info['birthday'])
# check for that
df_players_info[['date', 'birthday']].dtypes
```

Out[28]:

date datetime64[ns]  
birthday datetime64[ns]  
dtype: object

In [29]:

```
# change datatype from object to int or float
# since we have many attributes so using ""infer_objects" method will ease our mission
df_players_info = df_players_info.infer_objects()
# check for the datatypes
df_players_info.dtypes
```

Out[29]:

```
player_name      object
birthday      datetime64[ns]
height      float64
weight      int64
player_api_id    int64
date      datetime64[ns]
overall_rating    float64
potential      float64
preferred_foot    object
attacking_work_rate    object
defensive_work_rate    object
crossing      float64
finishing      float64
heading_accuracy    float64
short_passing    float64
volleys      float64
dribbling      float64
curve      float64
free_kick_accuracy    float64
long_passing      float64
ball_control      float64
acceleration      float64
sprint_speed      float64
agility      float64
reactions      float64
balance      float64
shot_power      float64
jumping      float64
stamina      float64
strength      float64
long_shots      float64
aggression      float64
interceptions      float64
positioning      float64
vision      float64
penalties      float64
marking      float64
standing_tackle    float64
sliding_tackle    float64
gk_diving      float64
gk_handling      float64
gk_kicking      float64
gk_positioning      float64
gk_reflexes      float64
dtype: object
```

In [30]:

```
df_players_info.head()
```

Out[30]:

|   | player_name        | birthday   | height | weight | player_api_id | date       | overall_rating | potential | preferred_foot | attacking_work_rate | ... | vision |
|---|--------------------|------------|--------|--------|---------------|------------|----------------|-----------|----------------|---------------------|-----|--------|
| 0 | Aaron Appindangoye | 1992-02-29 | 182.88 | 187    | 505942        | 2016-02-18 | 67.0           | 71.0      | right          | medium              | ... | 54.0   |
| 1 | Aaron Appindangoye | 1992-02-29 | 182.88 | 187    | 505942        | 2015-11-19 | 67.0           | 71.0      | right          | medium              | ... | 54.0   |
| 2 | Aaron Appindangoye | 1992-02-29 | 182.88 | 187    | 505942        | 2015-09-21 | 62.0           | 66.0      | right          | medium              | ... | 54.0   |
| 3 | Aaron Appindangoye | 1992-02-29 | 182.88 | 187    | 505942        | 2015-03-20 | 61.0           | 65.0      | right          | medium              | ... | 53.0   |
| 4 | Aaron Appindangoye | 1992-02-29 | 182.88 | 187    | 505942        | 2007-02-22 | 61.0           | 65.0      | right          | medium              | ... | 53.0   |

5 rows x 13 columns

In [31]:

```
# dealing with null values
# select columns that have null values by index then select columns with numeric datatype
null_column = df_players_info.iloc[:,5:].select_dtypes(include = np.number).columns
# loop through all the null column
for column in null_column :
    # replace each null value with the mean of "player" previous scores
    df_players_info[column] = df_players_info.groupby('player_name')[column].transform(lambda x:np.where(np.isnan(x),x.mean(),x))
```

In [32]:

```
# check for heading_accuracy value of Abdeslam Ouaddou player
# what we have done her is just calculate the mean of the previous score for Abdeslam Ouaddou (70+80+80+80+80)/5
=
df_players_info.query('player_name == "Abdeslam Ouaddou"')[['heading_accuracy']]
```

Out[32]:

|     | heading_accuracy |
|-----|------------------|
| 478 | 70.0             |
| 479 | 80.0             |
| 480 | 80.0             |
| 481 | 80.0             |
| 482 | 80.0             |
| 483 | 78.0             |

In [33]:

```
# select null columns which are categorical
null_column = ['preferred_foot','attacking_work_rate','defensive_work_rate']
# loop through all of these columns
for column in null_column :
    # replace each null value with the previous value
    df_players_info[column].fillna(method = 'ffill',inplace = True)
```

In [34]:

```
# check for preferred_foot value of Abdeslam Ouaddou player
df_players_info.query('player_name == "Abdeslam Ouaddou"')[['preferred_foot']]
```

Out[34]:

|     | preferred_foot |
|-----|----------------|
| 478 | right          |
| 479 | right          |
| 480 | right          |
| 481 | right          |
| 482 | right          |
| 483 | right          |

In [35]:

```
# let's check again for null values
df_players_info.isnull().sum()
```

Out[35]:

```
player_name      0
birthday         0
height           0
weight           0
player_api_id    0
date             0
overall_rating   0
potential        0
preferred_foot   0
attacking_work_rate 0
defensive_work_rate 0
crossing         0
finishing        0
heading_accuracy 0
short_passing    0
volleys         1848
dribbling        0
curve           1848
free_kick_accuracy 0
long_passing     0
ball_control     0
acceleration     0
sprint_speed     0
agility          1848
reactions        0
balance          1848
shot_power       0
jumping          1848
stamina          0
strength         0
long_shots       0
aggression       0
interceptions    0
positioning      0
vision           1848
penalties        0
marking          0
standing_tackle  0
sliding_tackle   1848
gk_diving        0
gk_handling      0
gk_kicking       0
gk_positioning   0
gk_reflexes      0
dtype: int64
```

In [36]:

```
# we will use foward filling method to fill this null values
df_players_info = df_players_info.fillna(method='ffill',axis = 1)
```

In [37]:

```
# check for vision value of Abdeslam Ouaddou player
df_players_info.query('player_name == "Abdeslam Ouaddou"')[['vision']]
```

Out[37]:

|     | vision |
|-----|--------|
| 478 | 70.0   |
| 479 | 71.0   |
| 480 | 71.0   |
| 481 | 71.0   |
| 482 | 71.0   |
| 483 | 70.8   |

In [38]:

```
# let's check again for null values
df_players_info.isnull().sum()
```

Out[38]:

```
player_name      0
birthday         0
height           0
weight           0
player_api_id    0
date             0
overall_rating   0
potential        0
preferred_foot   0
attacking_work_rate 0
defensive_work_rate 0
crossing         0
finishing        0
heading_accuracy 0
short_passing    0
volleys         0
dribbling        0
curve           0
free_kick_accuracy 0
long_passing     0
ball_control     0
acceleration     0
sprint_speed     0
agility          0
reactions        0
balance          0
shot_power       0
jumping          0
stamina          0
strength         0
long_shots       0
aggression       0
interceptions    0
positioning      0
vision           0
penalties        0
marking          0
standing_tackle  0
sliding_tackle   0
gk_diving        0
gk_handling      0
gk_kicking       0
gk_positioning   0
gk_reflexes      0
dtype: int64
```

Matches\_Dataset

In [39]:

```
# store card and date columns in separate dataframe to converted to tabular form
cards_info = df_mathces_info[['card','date']]
# we will just kepp the attributes that we want and drop the rest of them
df_mathces_info = df_mathces_info[['home_team_short_name', 'home_team_long_name', 'home_team_goal', 'away_team_sho
rt_name', 'away_team_long_name', 'away_team_goal', 'league_name', 'date', 'season', 'home_team_api_id', 'away_team_api
_id']]
```

In [40]:

```
# change datatype for date
df_mathces_info['date'] = pd.to_datetime(df_mathces_info['date'])
# check for that
df_mathces_info[['date']].dtypes
```

Out[40]:

```
date    datetime64[ns]
dtype: object
```

In [41]:

```
# check for the selected columns
df_mathces_info.columns
```

Out[41]:

```
Index(['home_team_short_name', 'home_team_long_name', 'home_team_goal',
       'away_team_short_name', 'away_team_long_name', 'away_team_goal',
       'league_name', 'date', 'season', 'home_team_api_id',
       'away_team_api_id'],
      dtype='object')
```

In [42]:

```
# in this section we will extract the players and the cards they have taken in specific date from xml files
# remove nulls from card column
cards_info = cards_info[cards_info.card.notnull()]
# let's extract players ids with card type red/yellow using BeautifulSoup
# define the following empty lists
player_id = []
card_type = []
dates = []
# extract the each xml file with the corresponding date
for xml, date in zip(cards_info.card, cards_info.date):
    soup = BeautifulSoup(xml, 'lxml') # parse the xml file to be converted into html
    players = [id.string for id in soup.find_all('player1')] # find all the players within player1 tag
    cards = [type.string for type in soup.find_all('card_type')] # find all the cards within card_type tag
    # sometimes card_type tag is not present so we replace it with comment tag
    comments = [comment.string for comment in soup.find_all('comment')]
    # we need to make the two lists cards and players consistent
    # if the we have players equal to cards that means our lists are consistent
    if len(players) == len(cards):
        player_id.extend(players) # append the current list with the bigger list
        card_type.extend(cards) # the something here
        dates.extend([date for e in range(len(players))]) # repeat the date along with the players list
    # if the we have players not equal cards then use comments instead
    elif len(players) == len(comments):
        player_id.extend(players)
        card_type.extend(comments)
        dates.extend([date for e in range(len(players))])
    # if the we have players not equal to cards or comments then there isn't enough information
    # that means we have palyers with unknown card types or card types with unknown players
    else:
        continue
# store the data into dataframe
cards_info = pd.DataFrame({"player_id":player_id,"card_type":card_type,"date":dates})
```

In [43]:

```
# explore the new dataset
explore_dataset(cards_info)
```

we have 61767 rows and 3 columns

---

column name :    player\_id

column datatype :    <class 'bs4.element.NavigableString'>

unique values :  
['24157' '30362' '30829' ... '25794' '95861' '214344']  
with total 5864

---

column name :    card\_type

column datatype :    <class 'bs4.element.NavigableString'>

unique values :  
['y' 'y2' 'r']  
with total 3

---

column name :    date

column datatype :    <class 'str'>

unique values :  
['2008-08-17 00:00:00' '2008-08-16 00:00:00' '2008-10-29 00:00:00' ...  
'2016-05-16 00:00:00' '2016-05-22 00:00:00' '2016-05-25 00:00:00']  
with total 1347

---

columns with null values count :

player\_id    0  
card\_type    0  
date    0  
dtype: int64  
with total 0

---

columns with duplicates count :

9

---

In [44]:

```
# change the datatype of date
cards_info.date = pd.to_datetime(cards_info.date)
```

In [45]:

```
# print the first 5 rows
cards_info.head()
```

Out[45]:

|   | player_id | card_type | date       |
|---|-----------|-----------|------------|
| 0 | 24157     | y         | 2008-08-17 |
| 1 | 30362     | y         | 2008-08-17 |
| 2 | 30829     | y         | 2008-08-17 |
| 3 | 37442     | y         | 2008-08-16 |
| 4 | 46621     | y         | 2008-08-16 |



# Exploratory Data Analysis

## Sections:

- ▷ [Research Question 1](#)
- ▷ [Research Question 2](#)
- ▷ [Research Question 3](#)

## Research Question 1 :

which teams has improved through all seasons in Spain LIGA BBVA ?

### Main Points :

1. *select the Spain LIGA BBVA league*
2. *define a function to plot the results to avoid code redundancy*
3. *call the function and comment on each season*

-----select the Spain LIGA BBVA league-----

In [46]:

```
df_matches_spain = df_mathces_info.query('league_name == "Spain LIGA BBVA"')
```

In [47]:

```
# unique seasons
df_matches_spain.season.unique()
```

Out[47]:

```
array(['2008/2009', '2009/2010', '2010/2011', '2011/2012', '2012/2013',
       '2013/2014', '2014/2015', '2015/2016'], dtype=object)
```

-----define a function to plot the results to avoid code redundancy-----

In [48]:

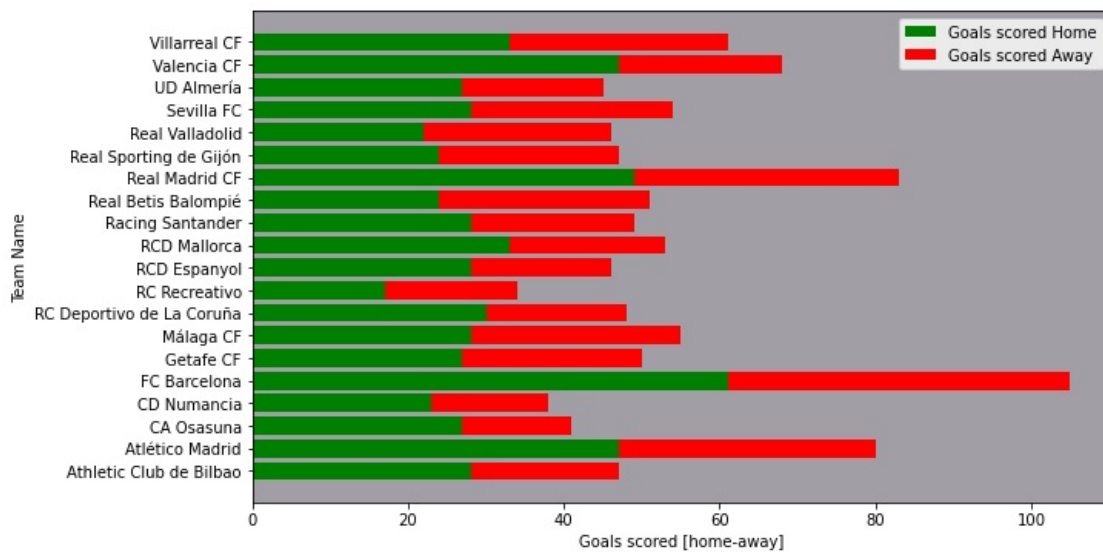
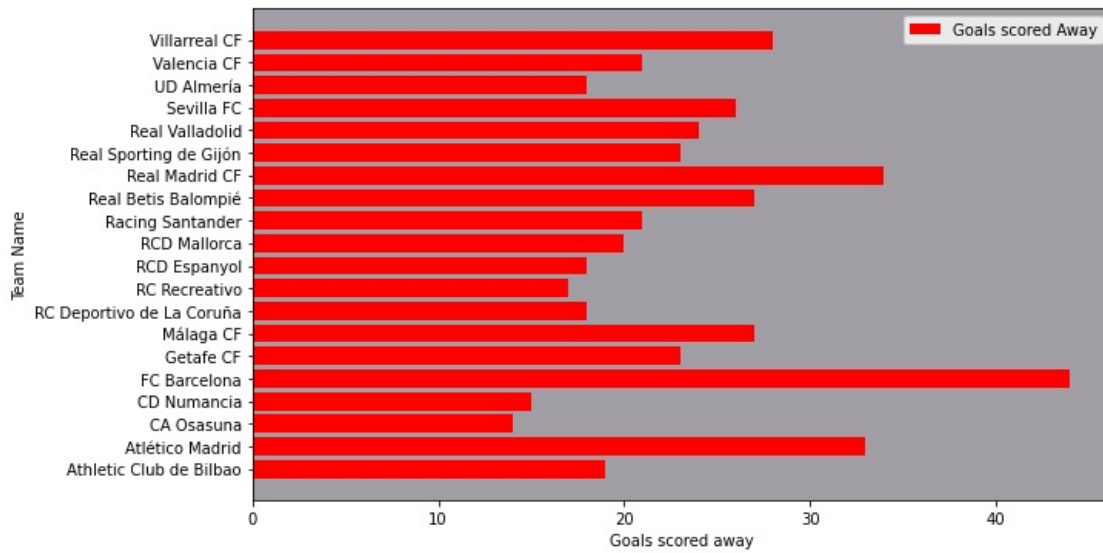
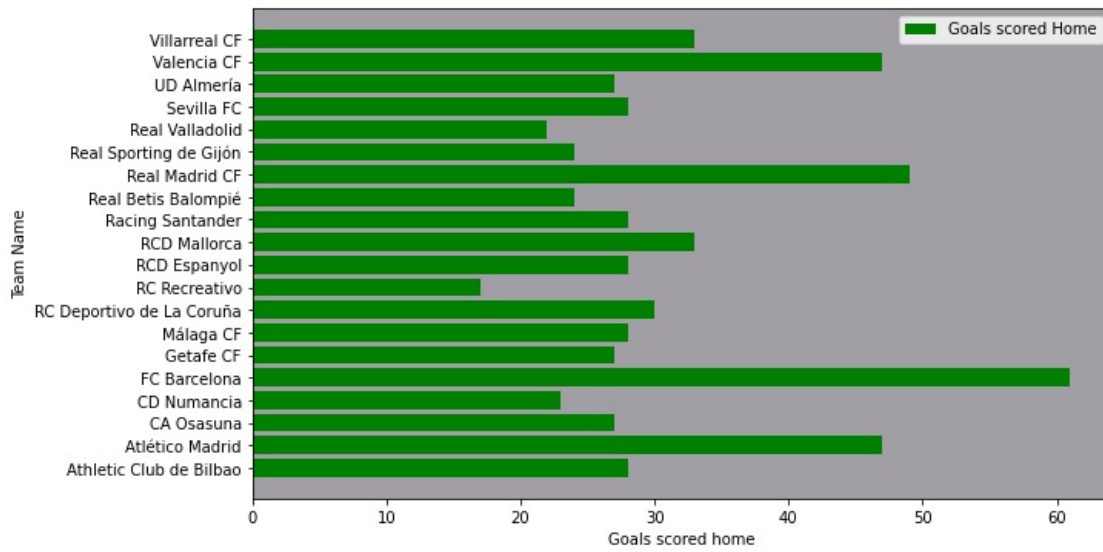
```
def compare_goals(season ,league,df):
    fig , (ax1 , ax2, ax3) = plt.subplots(3,figsize=(10,20))
    fig.suptitle("comparison Between goals scored Away and Home "+"\\n"+"season : "+season+"\\n"+league+" League")
    fig.patch.set_facecolor("#FFFFFF")
    # get the data related with season variable
    q = df.query("season == @season")
    # group by home goals
    home_goals = q.groupby(['season','home_team_long_name'])[['home_team_goal']].sum().reset_index()
    # select x and height for home teams
    x1 = home_goals['home_team_long_name']
    height1 = home_goals['home_team_goal']
    # group by away goals
    away_goals = q.groupby(['season','away_team_long_name'])[['away_team_goal']].sum().reset_index()
    # select x and height for away teams
    x2 = away_goals['away_team_long_name']
    height2 = away_goals['away_team_goal']
    # plot the home team goals
    ax1.set_xlabel("Goals scored home")
    ax1.set_ylabel("Team Name")
    ax1.set_facecolor('#A19FA5')
    ax1.barh( x1, height1, color = "g",label = "Goals scored Home")
    ax1.legend()
    # plot the away team goals
    ax2.set_xlabel("Goals scored away")
    ax2.set_ylabel("Team Name")
    ax2.set_facecolor('#A19FA5')
    ax2.barh(x2,height2,color='r', label = "Goals scored Away")
    ax2.legend()
    # stacked barchart to easily compare between them
    labels = list(home_goals['home_team_long_name'])
    x = np.arange(len(labels))
    width = 0.8
    ax3.set_facecolor('#A19FA5')
    ax3.barh( labels, height1, width ,label = 'Goals scored Home', color = 'g')
    ax3.barh( labels, height2, width,left = height1, label = 'Goals scored Away', color = 'r')
    ax3.set_xlabel("Goals scored [home-away]")
    ax3.set_ylabel("Team Name")
    ax3.legend()
    plt.show()
```

-----call the function and comment on each season-----

In [49]:

```
compare_goals('2008/2009','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2008/2009  
Spain LIGA BBVA League



season 2008/2009 :

*Barcelona* has the heighest score with more than 100 goals most of them scored Home

*Real Madrid* has the second heighest score with more than 80 goals most of them scored Home

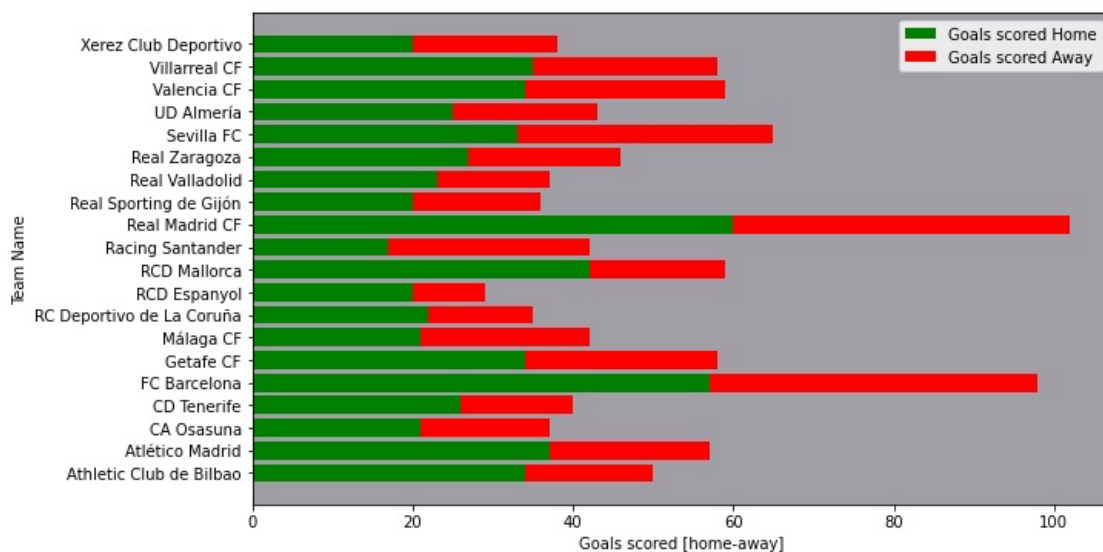
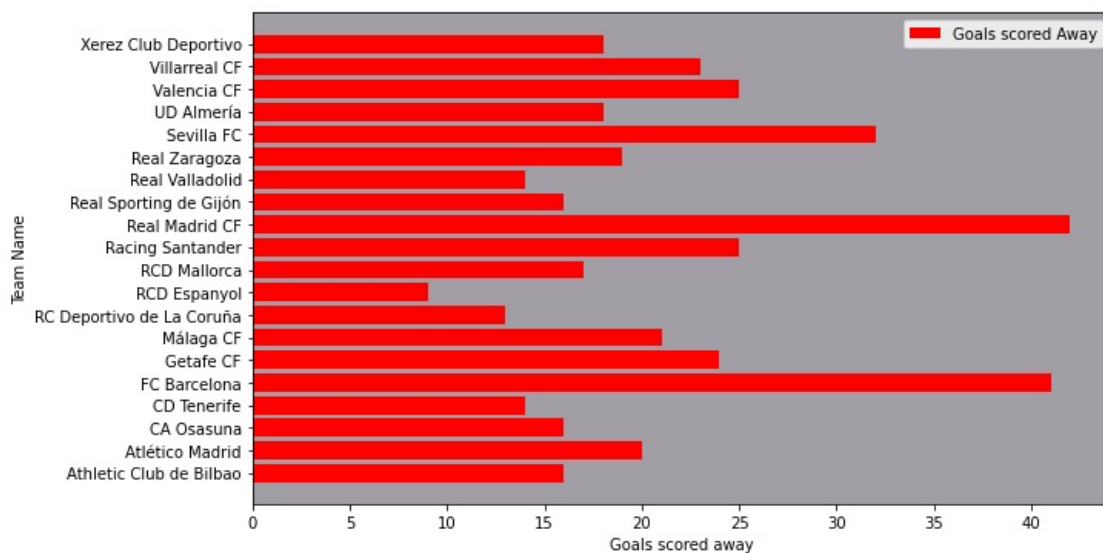
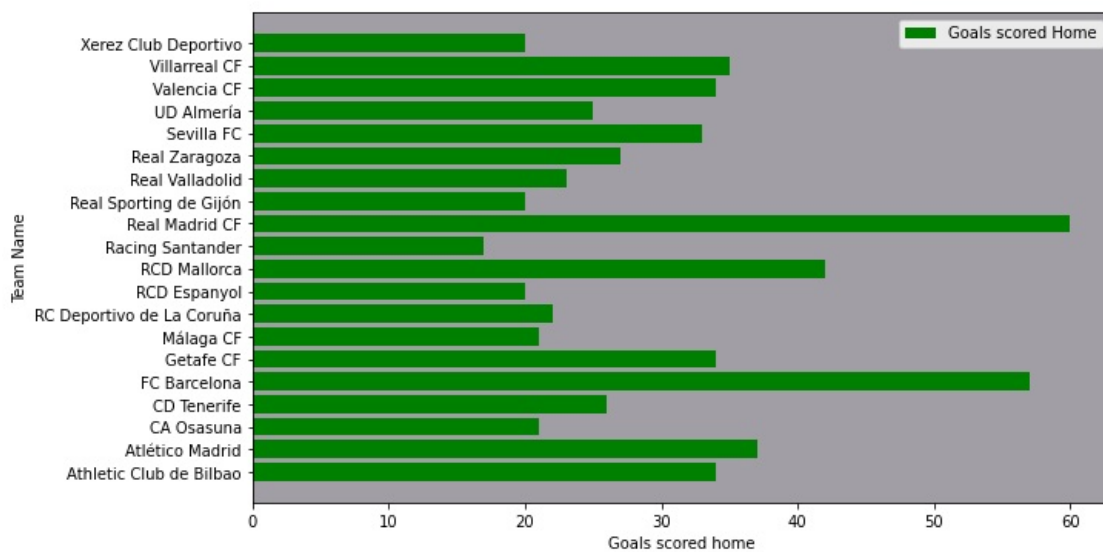
*Atlitco Madrid* has the third heighest score with 80 goalsmost of them scored Home

*Recreativo* has the lowest score with less than 40 goals

In [50]:

```
compare_goals('2009/2010','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2009/2010  
Spain LIGA BBVA League



season 2009/2010 :

*Real Madrid* has the heighest score with more than 100 goals most of them scored Home

*Real Madrid* has the second heighest score with 100 goals most of them scored Home

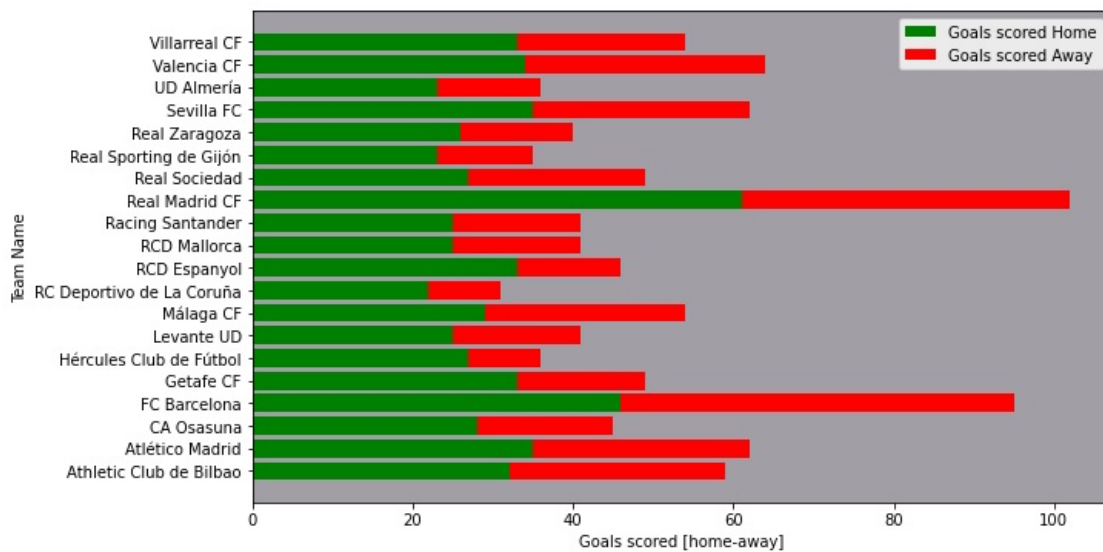
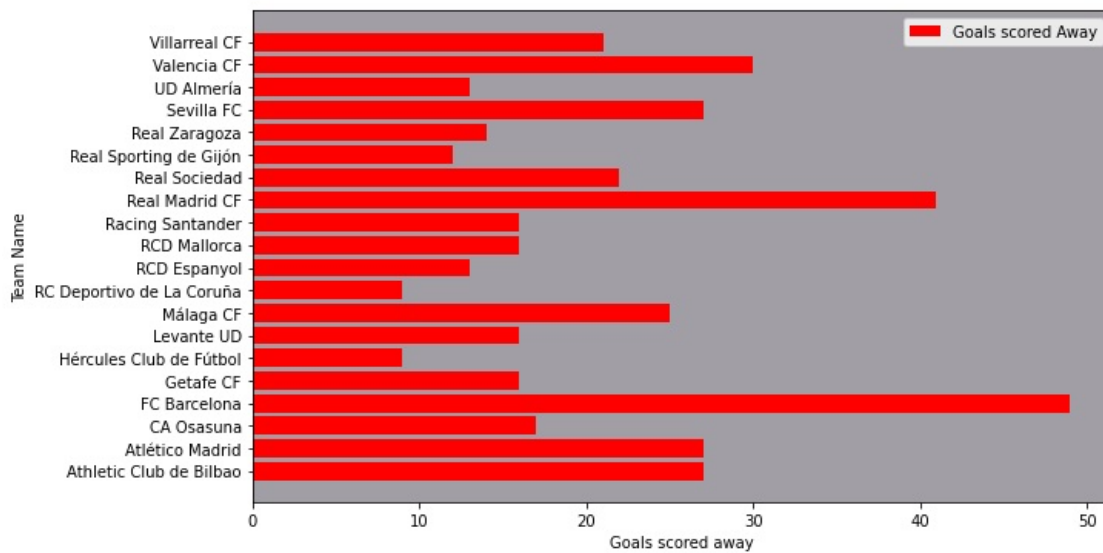
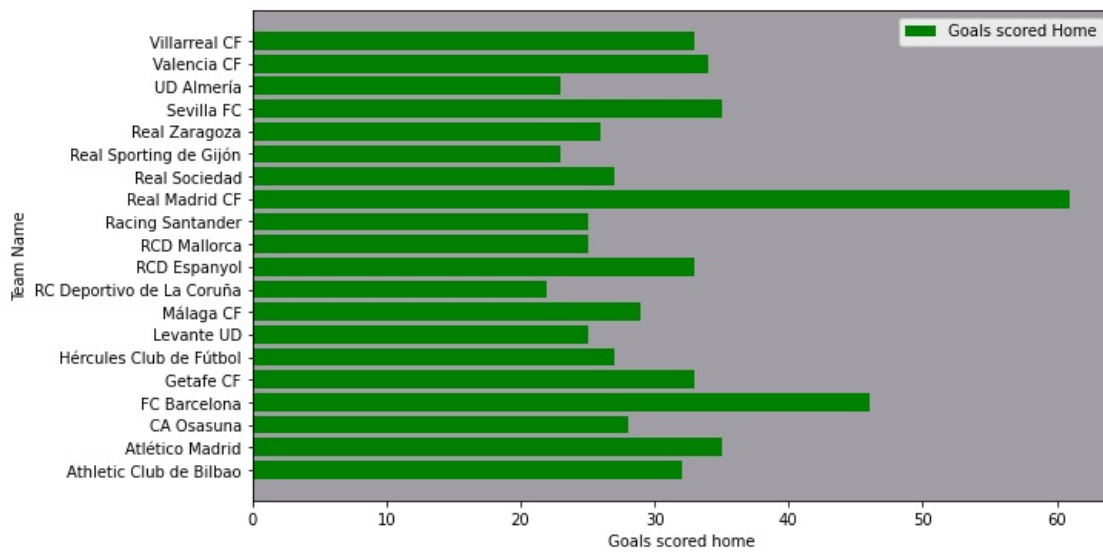
*Sevilla* has the third heighest score with 60 goals most of them scored Home

*Espanyol* has the lowest score with less than 40 goals

In [51]:

```
compare_goals('2010/2011','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2010/2011  
Spain LIGA BBVA League



season 2010/2011 :

*Real Madrid* has the heighest score with more than 100 goals most of them scored Home

*Barcelona* has the second heighest score with less than 100 goals most of them scored Away

*Valencia* has the third heighest score with 60 goals most of them scored Home

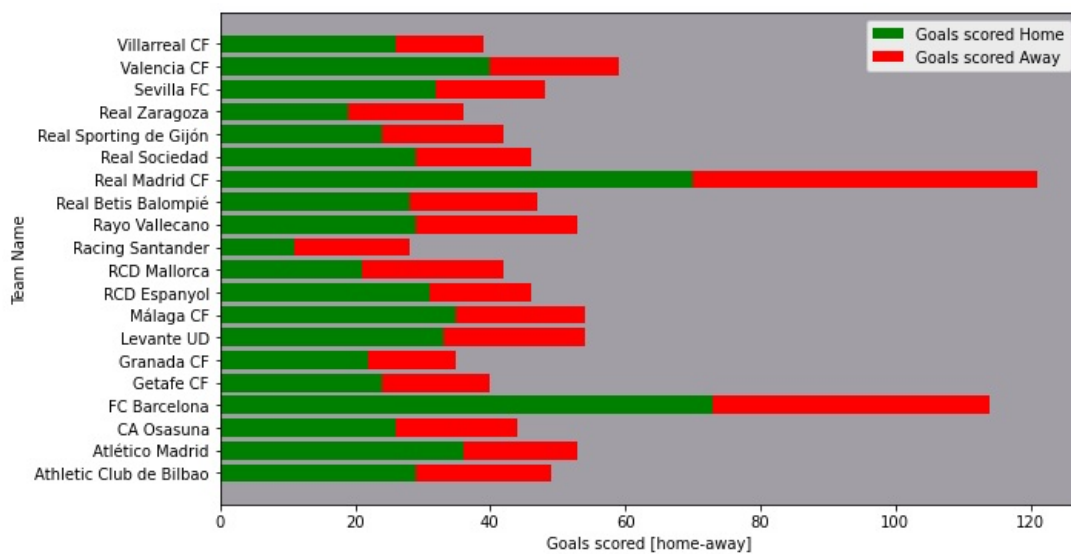
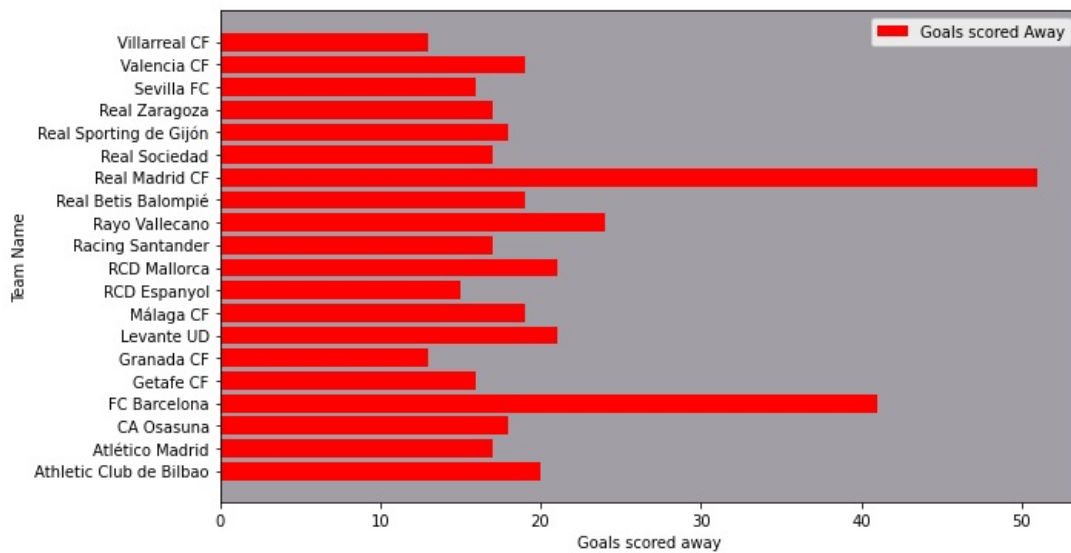
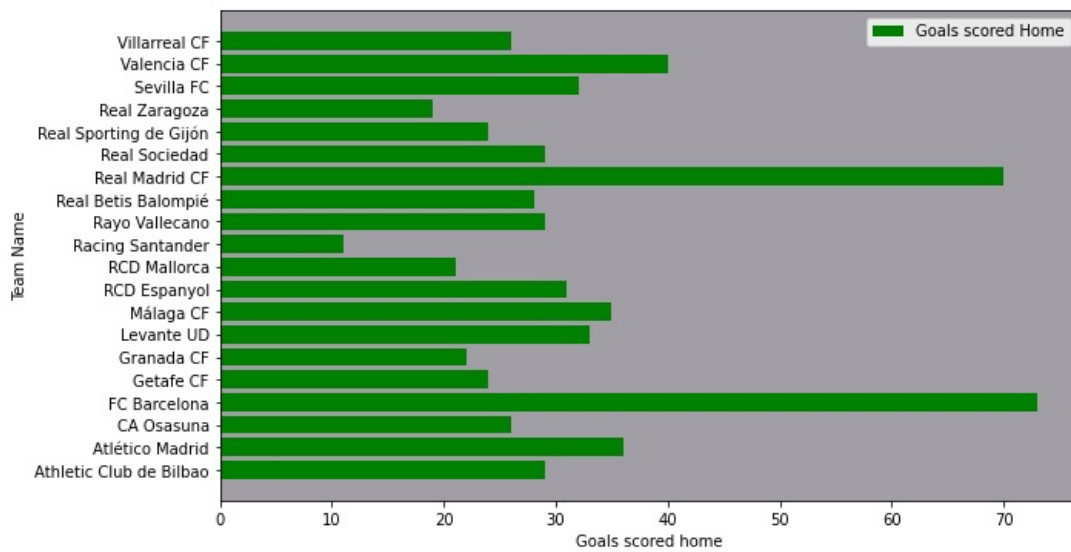
*Deportivo de la coruna* has the lowest score with less than 40 goals

In [52]:

```
compare_goals('2011/2012','Spain LIGA BBVA',df_matches_spain)
```



comparison Between goals scored Away and Home  
season : 2011/2012  
Spain LIGA BBVA League



season 2011/2012 :

*Real Madrid* has the heighest score with 120 goals most of them scored Home

*Barcelona* has the second heighest score with more than 100 goals most of them scored Home

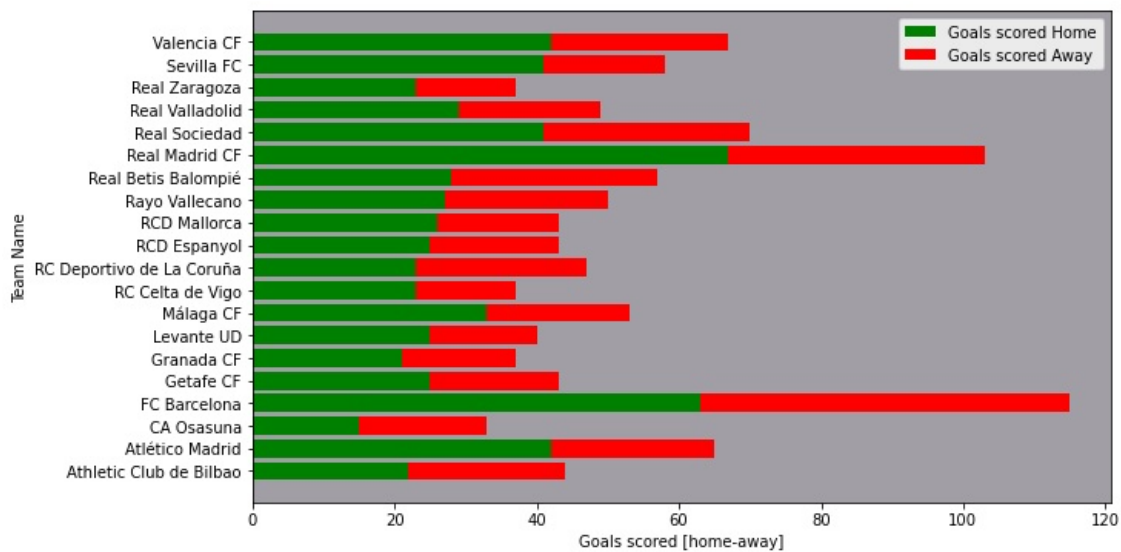
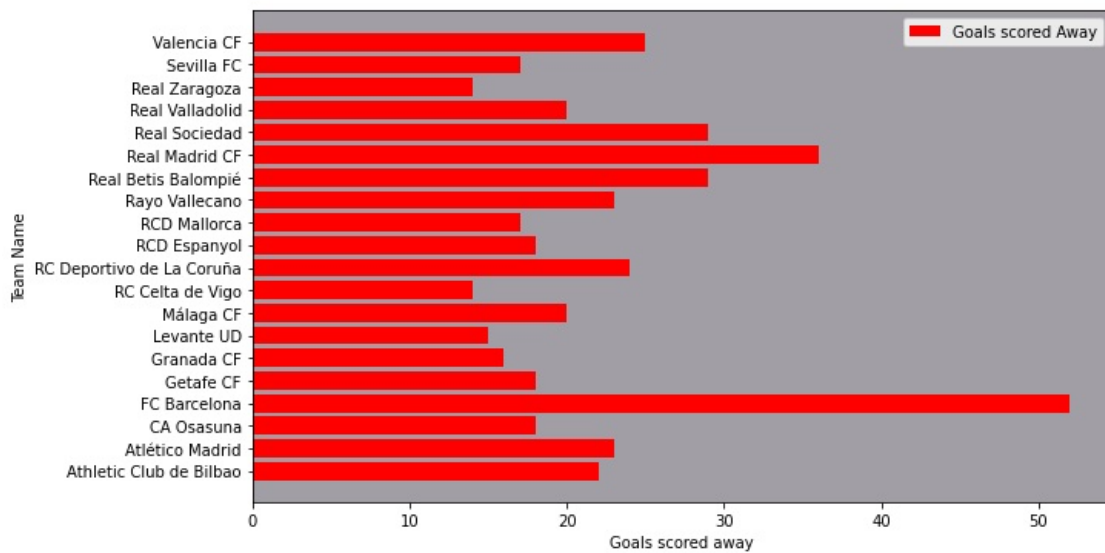
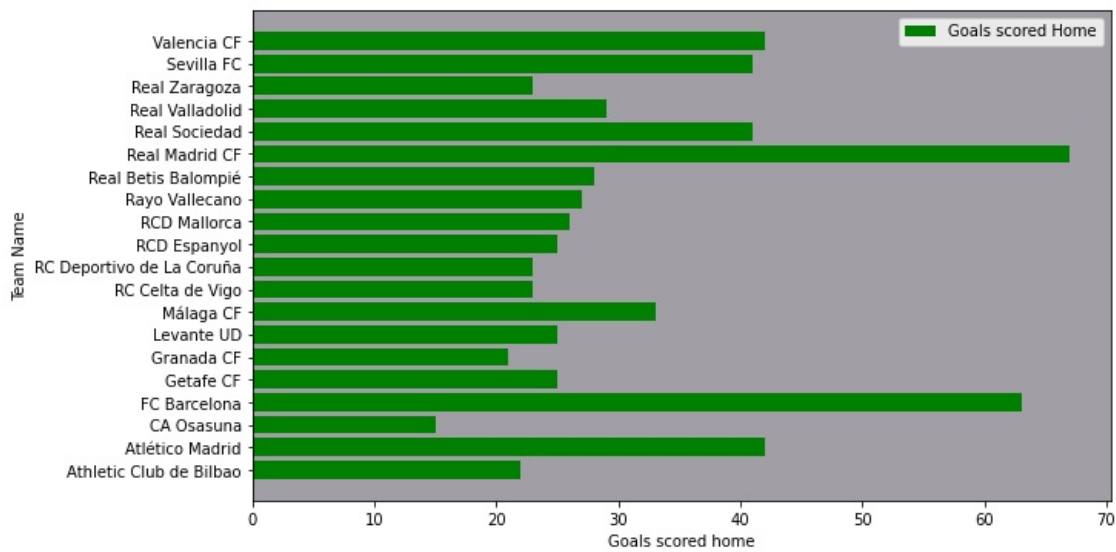
*Valencia* has the third heighest score with 60 goals most of them scored Home

*Racing santander* has the lowest score with less than 40 goals

In [53]:

```
compare_goals('2012/2013','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2012/2013  
Spain LIGA BBVA League



season 2012/2013 :

*Barcelona* has the heighest score with more than 100 goals most of them scored Home

*Real Madrid* has the second heighest score with more than 100 goals most of them scored Home

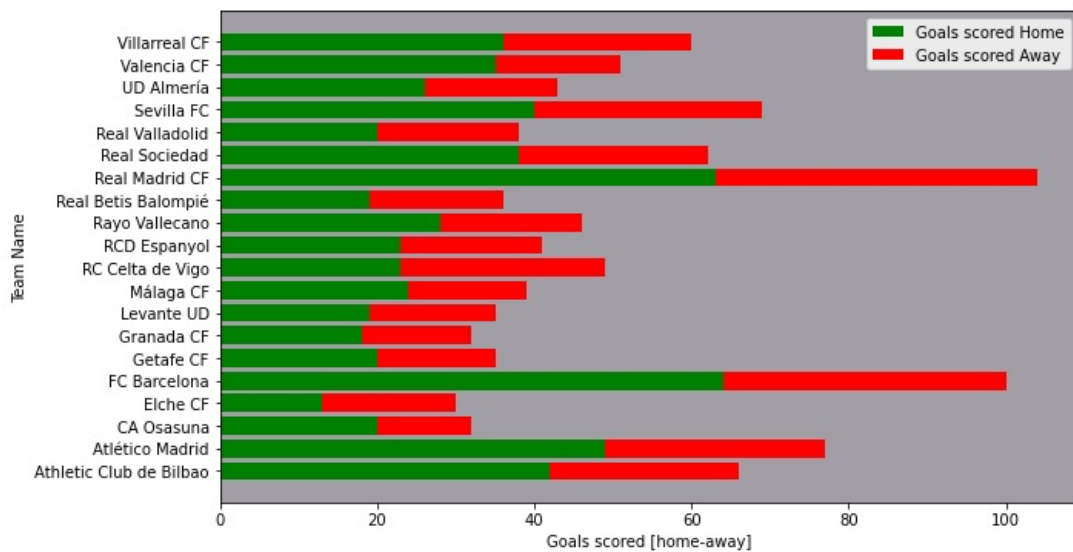
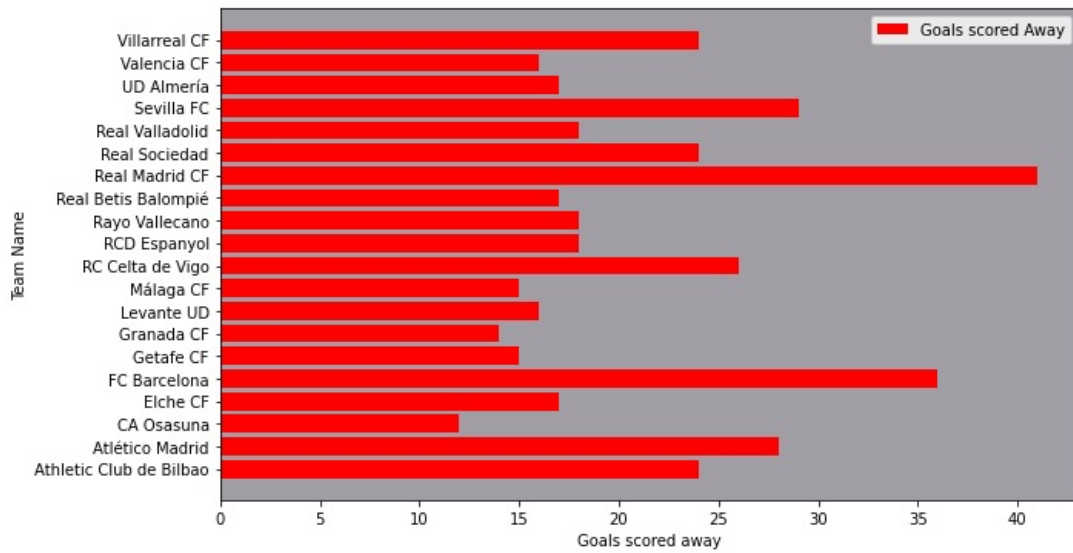
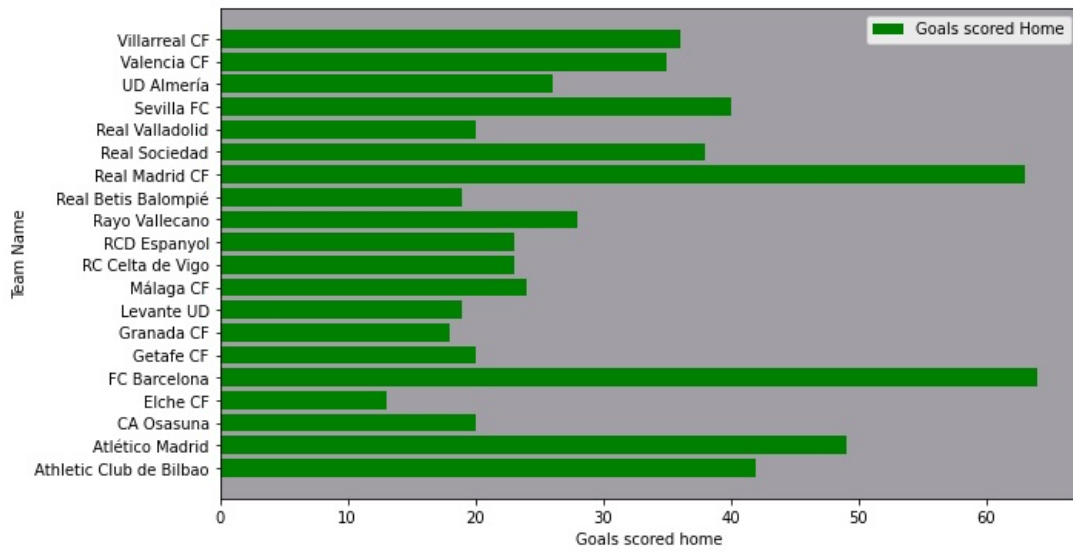
*Real sociedad* has the third heighest score with more than 60 goals most of them scored Home

*Osasuna* has the lowest score with less than 40 goals

In [54]:

```
compare_goals('2013/2014','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2013/2014  
Spain LIGA BBVA League



season 2013/2014 :

*Real Madrid* has the heighest score with more than 100 goals most of them scored Home

*Barcelona* has the second heighest score with 100 goals most of them scored Home

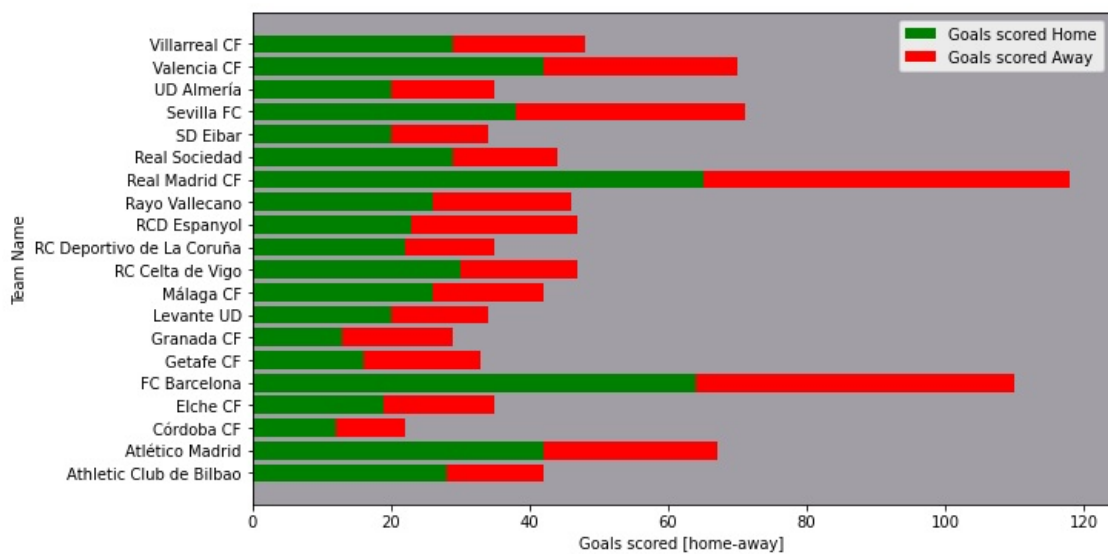
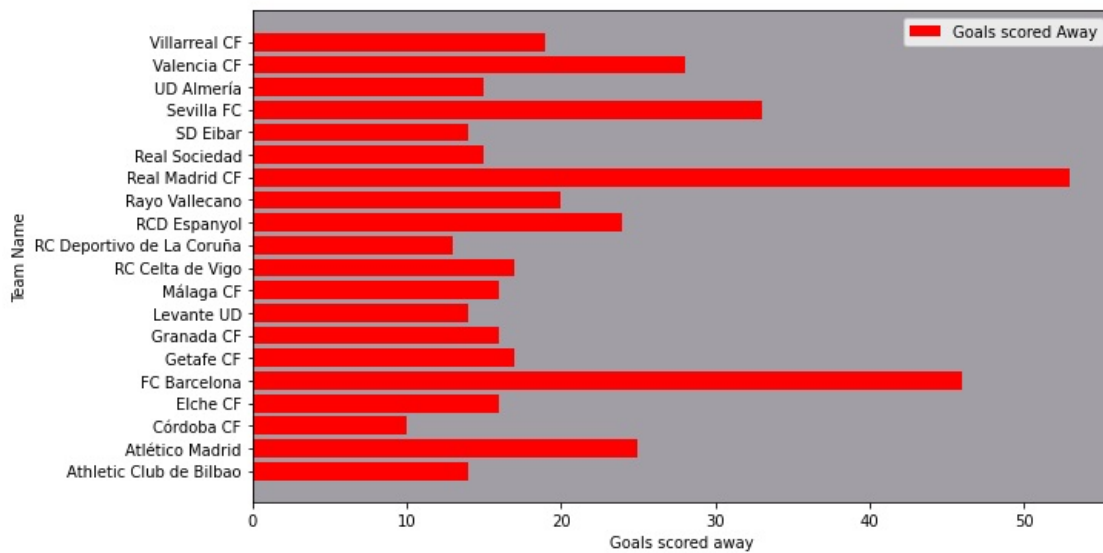
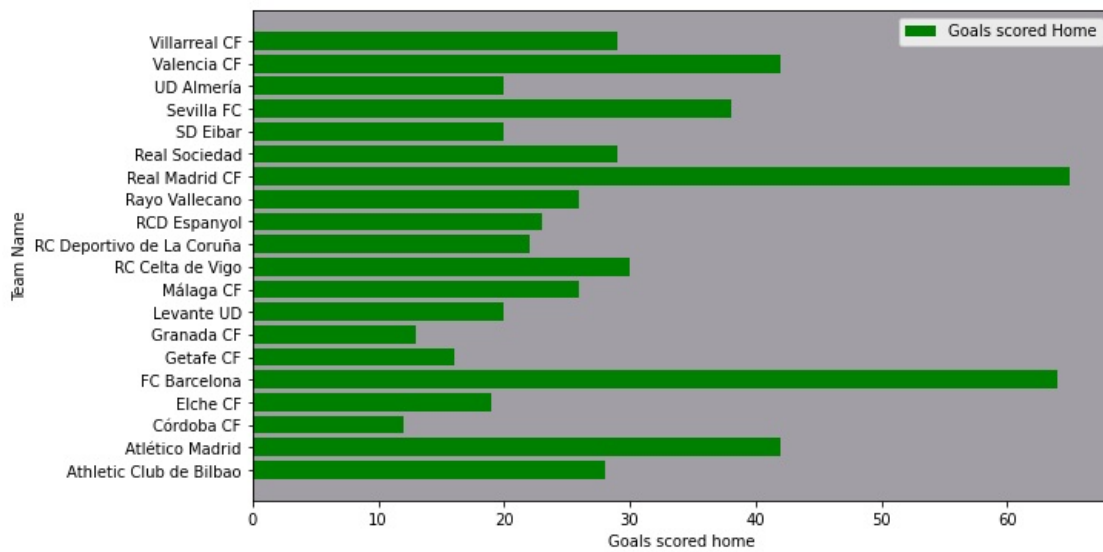
*Atletico Madrid* has the third heighest score with 60 goals most of them scored Home

*Elche* has the lowest score with less than 40 goals

In [55]:

```
compare_goals('2014/2015','Spain LIGA BBVA',df_matches_spain)
```

comparison Between goals scored Away and Home  
season : 2014/2015  
Spain LIGA BBVA League



season 2014/2015 :

*Real Madrid* has the heighest score with more than 100 goals most of them scored Home

*Barcelona* has the second heighest score with more than 100 goals most of them scored Home

*Sevilla* has the third heighest score with 60 goals most of them scored Home

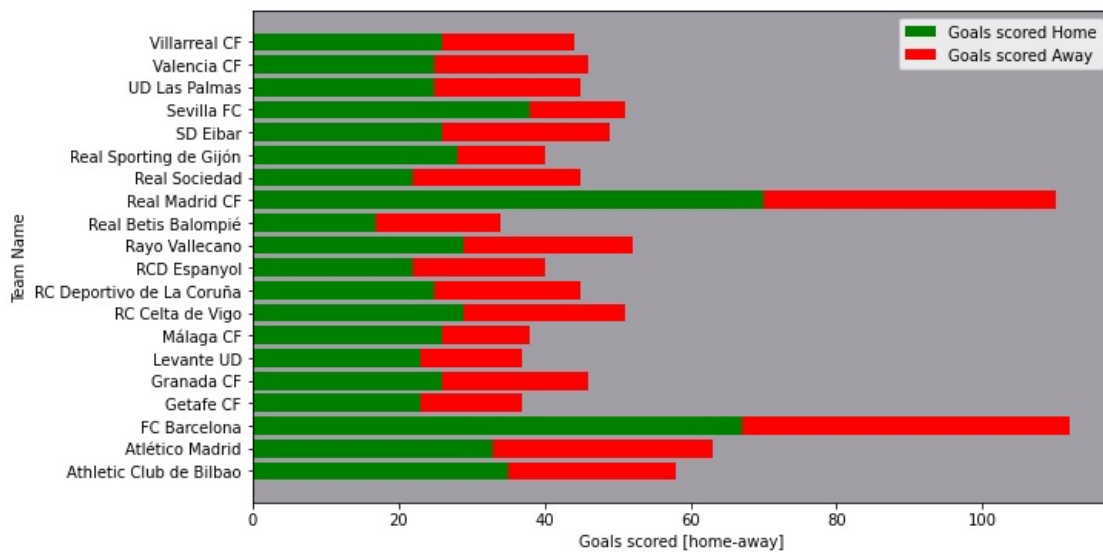
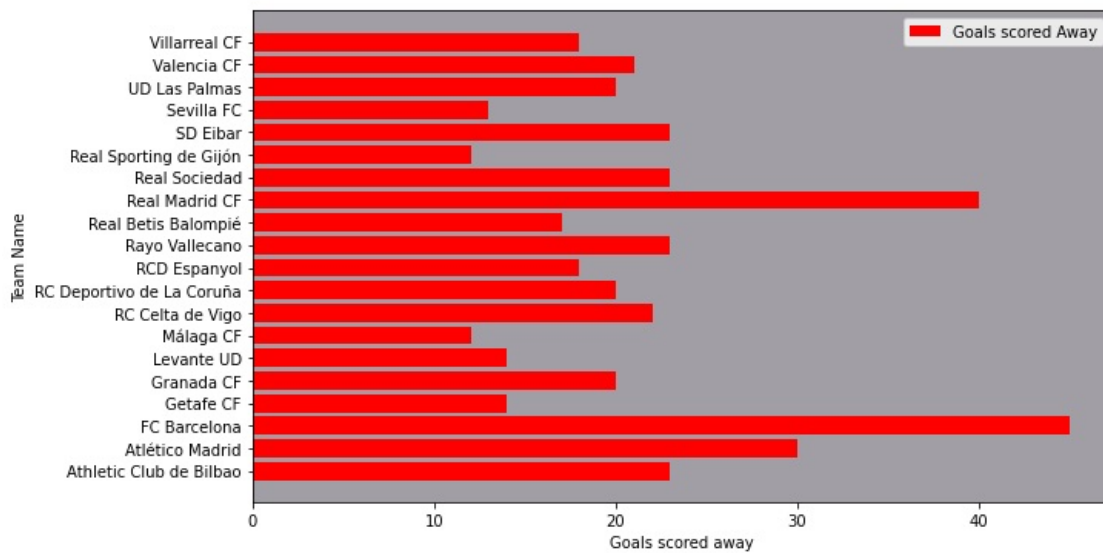
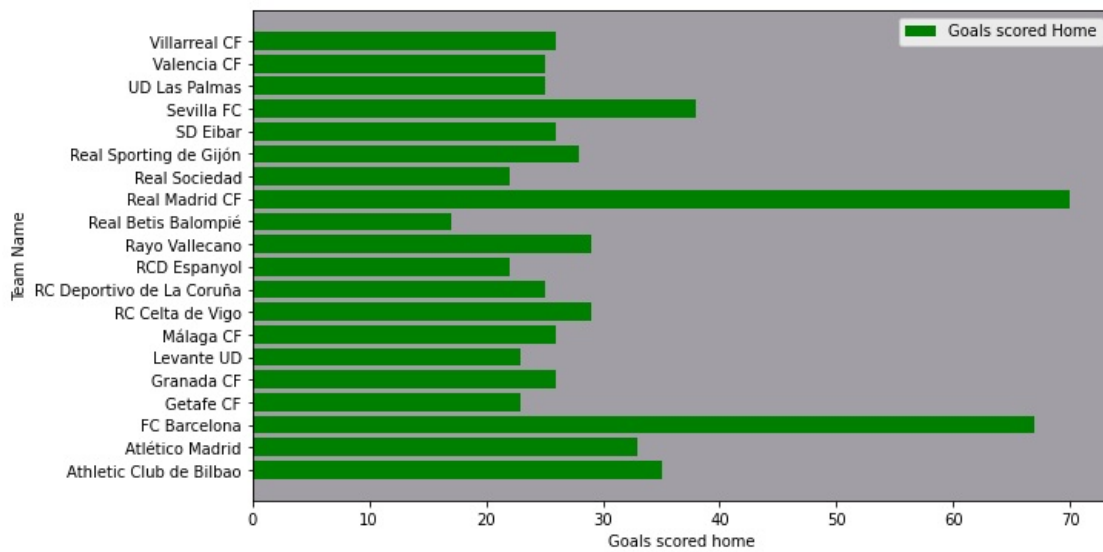
*Cordoba* has the lowest score with 20 goals

In [56]:

```
compare_goals('2015/2016','Spain LIGA BBVA',df_matches_spain)
```



comparison Between goals scored Away and Home  
season : 2015/2016  
Spain LIGA BBVA League



season 2015/2016 :

*Barcelona* has the highest score with more than 100 goals most of them scored Home

*Real Madrid* has the second highest score with more than 100 goals most of them scored Home

*Atletico Madrid* has the third highest score with 60 goals most of them scored Home

*Real betis balompie* has the lowest score with less than 40 goals

## Research Question 2 :

How is the player's physical power related to weight and height?

Main Points :

1. The physical power can be determined by some attributes like Acceleration, Agility, Balance, jumping, stamina and strength
2. select specific date to deal with
3. Find the distribution of weight
4. Find the distribution of height
5. Check if there any correlation between weight and height
6. Check for the correlation between weight and all the attributes
7. Check for the correlation between height and all the attributes

Note: use correlation value  $R^2$  such that :  $R^2 \approx \pm 0.4$  or higher

In [57]:

```
physical_attr = df_players_info[['player_name', 'date', 'acceleration', 'agility', 'balance', 'jumping', 'stamina', 'strength', 'reactions', 'weight', 'height']]
```

In [58]:

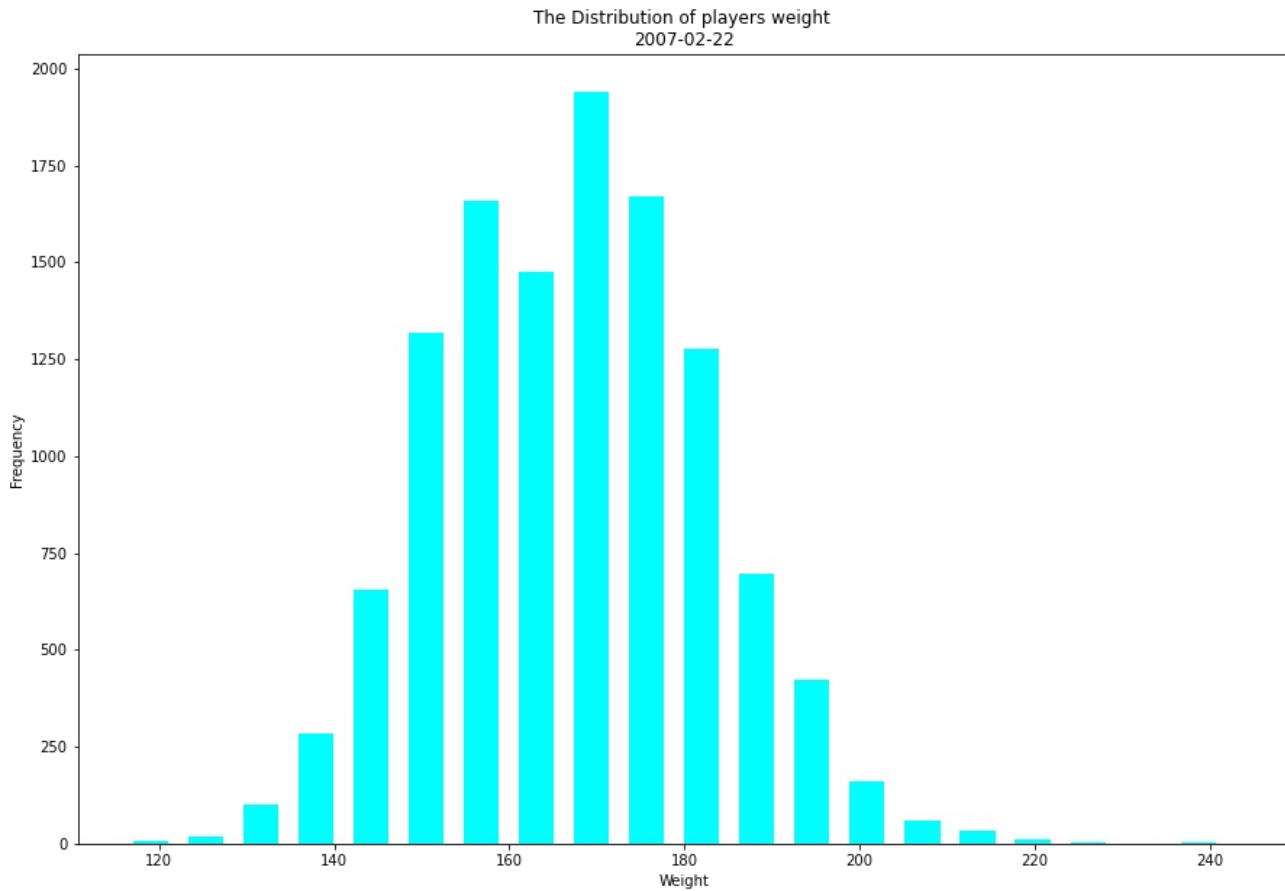
```
# select data within date 2007-02-22
# you can find unique dates in the previous phase Data Wrangling
physical_attr_227 = physical_attr.query('date == 20070222')
# make sure datatypes are consistent
physical_attr_227 = physical_attr_227.infer_objects()
```

.....View the distribution of Weight.....

In [59]:

```
# view the distribution of players weight in this date
weight = physical_attr_227['weight']
plt.figure(figsize=(15,10))
plt.title("The Distribution of players weight "+"\\n2007-02-22")
plt.ylabel("Frequency")
plt.xlabel("Weight")
plt.hist( weight,bins = 20,color='cyan',width=4);
print("The mean value of weight is {}".format(np.mean(weight)))
```

The mean value of weight is 168.44942294636795



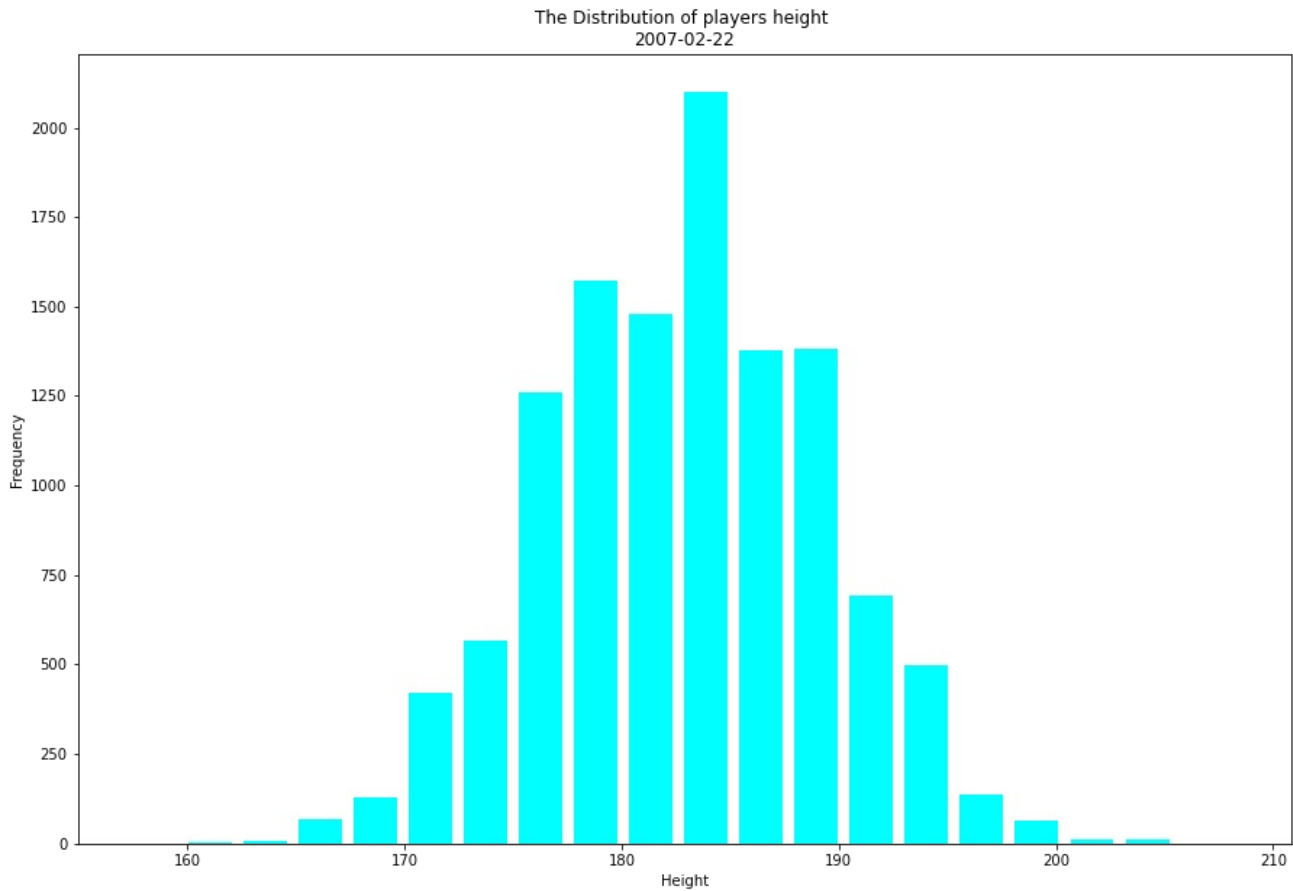
*The players weight are normally distributed*

.....View the distribution of Height.....

In [60]:

```
# view the distribution of players height in this date
height = physical_attr_227['height']
plt.figure(figsize=(15,10))
plt.title("The Distribution of players height "+"\\n2007-02-22")
plt.ylabel("Frequency")
plt.xlabel("Height")
plt.hist( height,bins =20,color='cyan',width=2);
print("The mean value of height is {}".format(np.mean(height)))
```

The mean value of height is 181.86908689747958

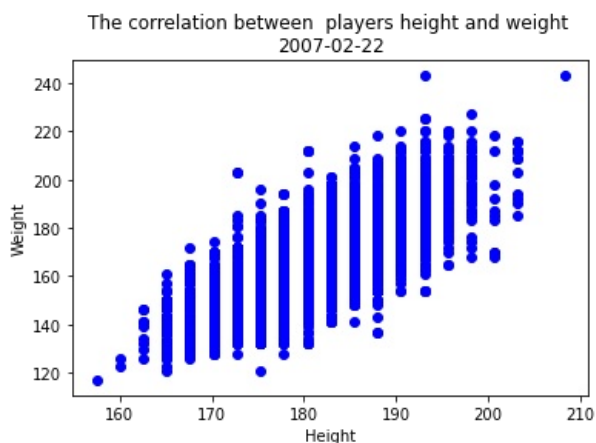


*The same thing for players height*

.....View the correlation between Height and Weight.....

In [61]:

```
# let's see if there any correlation between weight and height
height = physical_attr_227['height']
weight = physical_attr_227['weight']
plt.title("The correlation between players height and weight "+"\\n2007-02-22")
plt.ylabel("Weight")
plt.xlabel("Height")
plt.scatter( height, weight, color='blue');
```



There is a positive correlation between Hight and weight

.....View Which attribute has a correlation with Weight.....

In [62]:

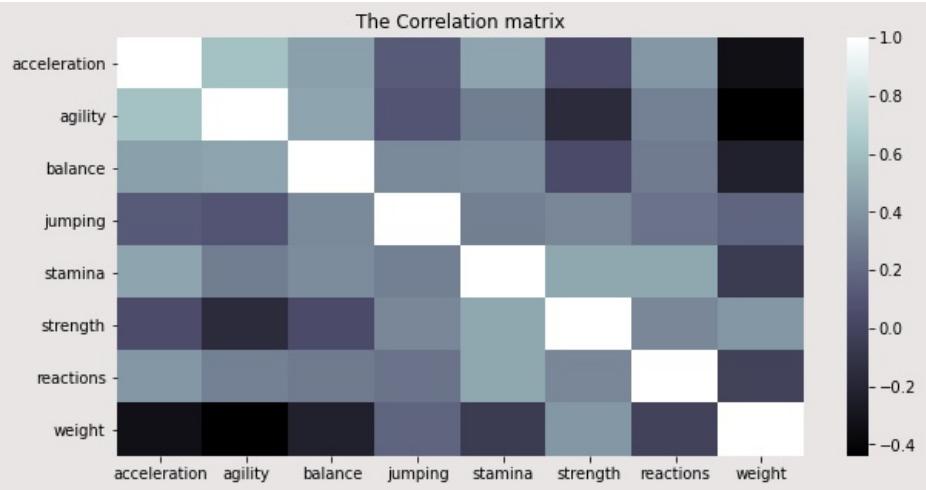
```
# build a correlation matrix between weight and all the other attributes
weight_correlation_matrix =physical_attr_227[['acceleration','agility','balance','jumping','stamina','strength','
reactions','weight']].corr()
weight_correlation_matrix
```

Out[62]:

|              | acceleration | agility   | balance   | jumping  | stamina   | strength  | reactions | weight    |
|--------------|--------------|-----------|-----------|----------|-----------|-----------|-----------|-----------|
| acceleration | 1.000000     | 0.613982  | 0.454475  | 0.130513 | 0.474922  | 0.052010  | 0.411562  | -0.337811 |
| agility      | 0.613982     | 1.000000  | 0.472631  | 0.099194 | 0.286786  | -0.163240 | 0.310720  | -0.439215 |
| balance      | 0.454475     | 0.472631  | 1.000000  | 0.351952 | 0.359711  | 0.048557  | 0.282136  | -0.220130 |
| jumping      | 0.130513     | 0.099194  | 0.351952  | 1.000000 | 0.297487  | 0.338856  | 0.242815  | 0.175596  |
| stamina      | 0.474922     | 0.286786  | 0.359711  | 0.297487 | 1.000000  | 0.483845  | 0.483094  | -0.061764 |
| strength     | 0.052010     | -0.163240 | 0.048557  | 0.338856 | 0.483845  | 1.000000  | 0.337293  | 0.410662  |
| reactions    | 0.411562     | 0.310720  | 0.282136  | 0.242815 | 0.483094  | 0.337293  | 1.000000  | -0.016343 |
| weight       | -0.337811    | -0.439215 | -0.220130 | 0.175596 | -0.061764 | 0.410662  | -0.016343 | 1.000000  |

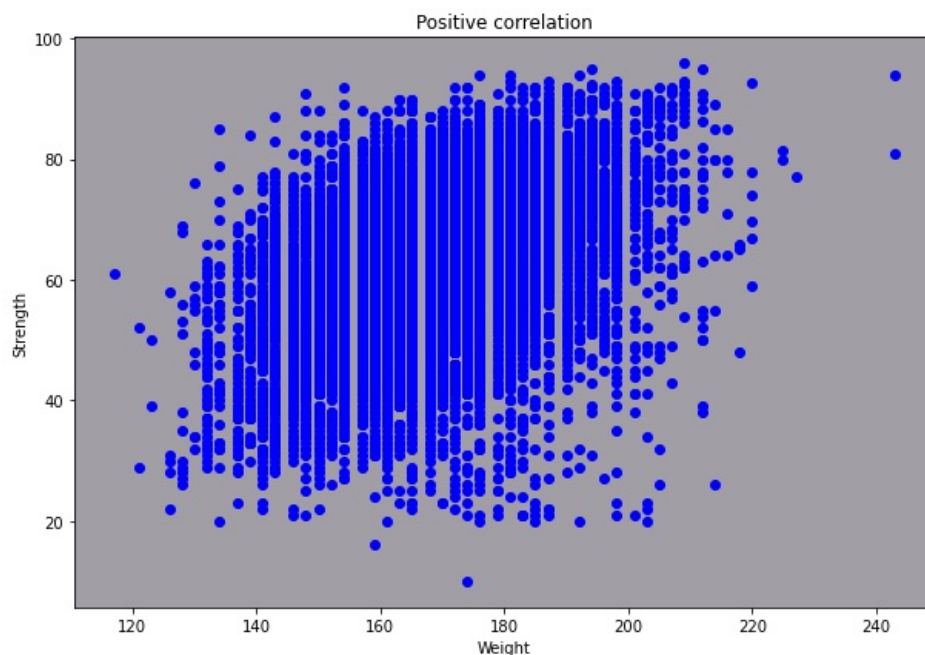
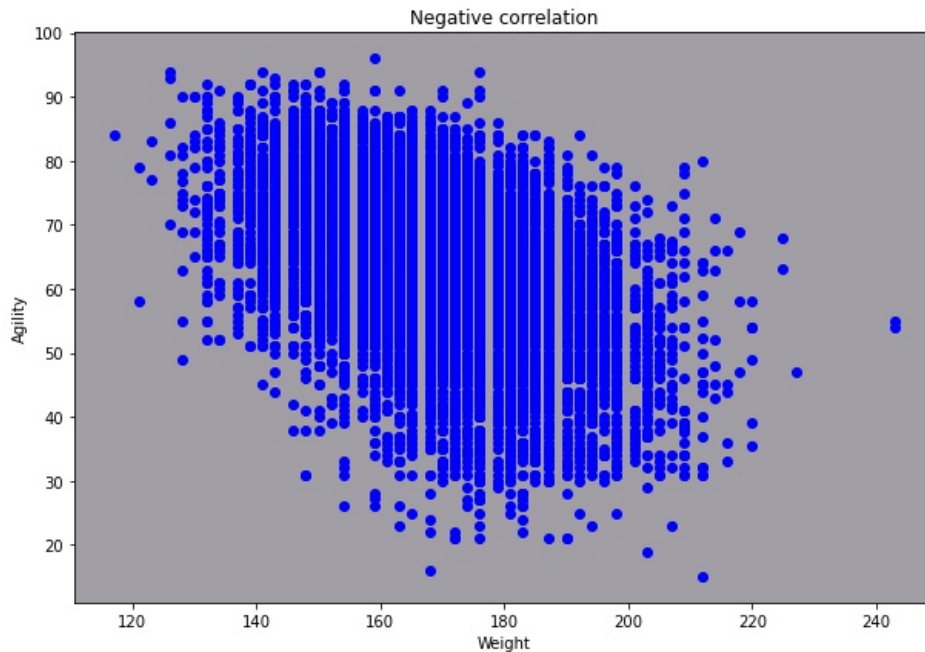
In [63]:

```
fig = plt.figure(figsize=(10,5))
plt.title("The Correlation matrix")
fig.patch.set_facecolor("#E9E5E5")
sns.heatmap(weight_correlation_matrix,cmap = 'bone');
```



In [64]:

```
# let's build a scatter plot for both correaltions
fig ,(ax1,ax2) = plt.subplots(2,figsize=(10,15))
fig.patch.set_facecolor("#FFFFFF")
weight = physical_attr_227['weight']
agility = physical_attr_227['agility']
strength = physical_attr_227['strength']
# scatter plot between weight and agility
ax1.set_facecolor('#A19FA5')
ax1.set_xlabel("Weight")
ax1.set_ylabel("Agility")
ax1.set_title("Negative correlation")
ax1.scatter(weight,agility,color='blue');
# scatter plot between weight and strenght
ax2.set_facecolor('#A19FA5')
ax2.set_xlabel("Weight")
ax2.set_ylabel("Strength")
ax2.set_title("Positive correlation")
ax2.scatter(weight,strength,color='blue');
```



- weight and agility has a negative correlation
- weight and strength has a positive correaltion

.....Repeat the same process to the Height.....

According to previous steps weight and Height has a positive correlation. Hence, we must find a correlation between height and agility and strength

In [65]:

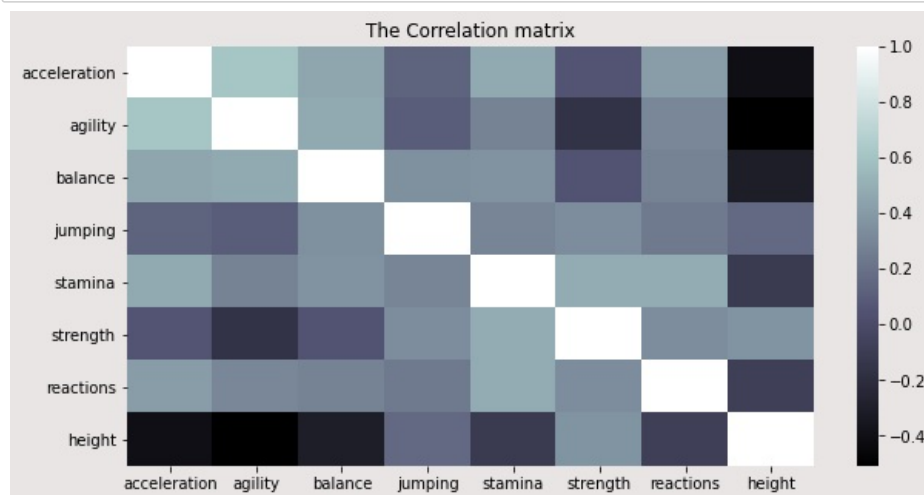
```
# build a correlation matrix between Height and all the other attributes
height_correlation_matrix = physical_attr_227[['acceleration', 'agility', 'balance', 'jumping', 'stamina', 'strength',
'reactions', 'height']].corr()
height_correlation_matrix
```

Out[65]:

|              | acceleration | agility   | balance   | jumping  | stamina   | strength  | reactions | height    |
|--------------|--------------|-----------|-----------|----------|-----------|-----------|-----------|-----------|
| acceleration | 1.000000     | 0.613982  | 0.454475  | 0.130513 | 0.474922  | 0.052010  | 0.411562  | -0.405290 |
| agility      | 0.613982     | 1.000000  | 0.472631  | 0.099194 | 0.286786  | -0.163240 | 0.310720  | -0.507899 |
| balance      | 0.454475     | 0.472631  | 1.000000  | 0.351952 | 0.359711  | 0.048557  | 0.282136  | -0.306263 |
| jumping      | 0.130513     | 0.099194  | 0.351952  | 1.000000 | 0.297487  | 0.338856  | 0.242815  | 0.161593  |
| stamina      | 0.474922     | 0.286786  | 0.359711  | 0.297487 | 1.000000  | 0.483845  | 0.483094  | -0.117280 |
| strength     | 0.052010     | -0.163240 | 0.048557  | 0.338856 | 0.483845  | 1.000000  | 0.337293  | 0.368340  |
| reactions    | 0.411562     | 0.310720  | 0.282136  | 0.242815 | 0.483094  | 0.337293  | 1.000000  | -0.083399 |
| height       | -0.405290    | -0.507899 | -0.306263 | 0.161593 | -0.117280 | 0.368340  | -0.083399 | 1.000000  |

In [66]:

```
# let's make the correlation matrix clear by using a heatmap
fig = plt.figure(figsize=(10,5))
plt.title("The Correlation matrix")
fig.patch.set_facecolor("#E9E5E5")
sns.heatmap(height_correlation_matrix, cmap = 'bone');
```

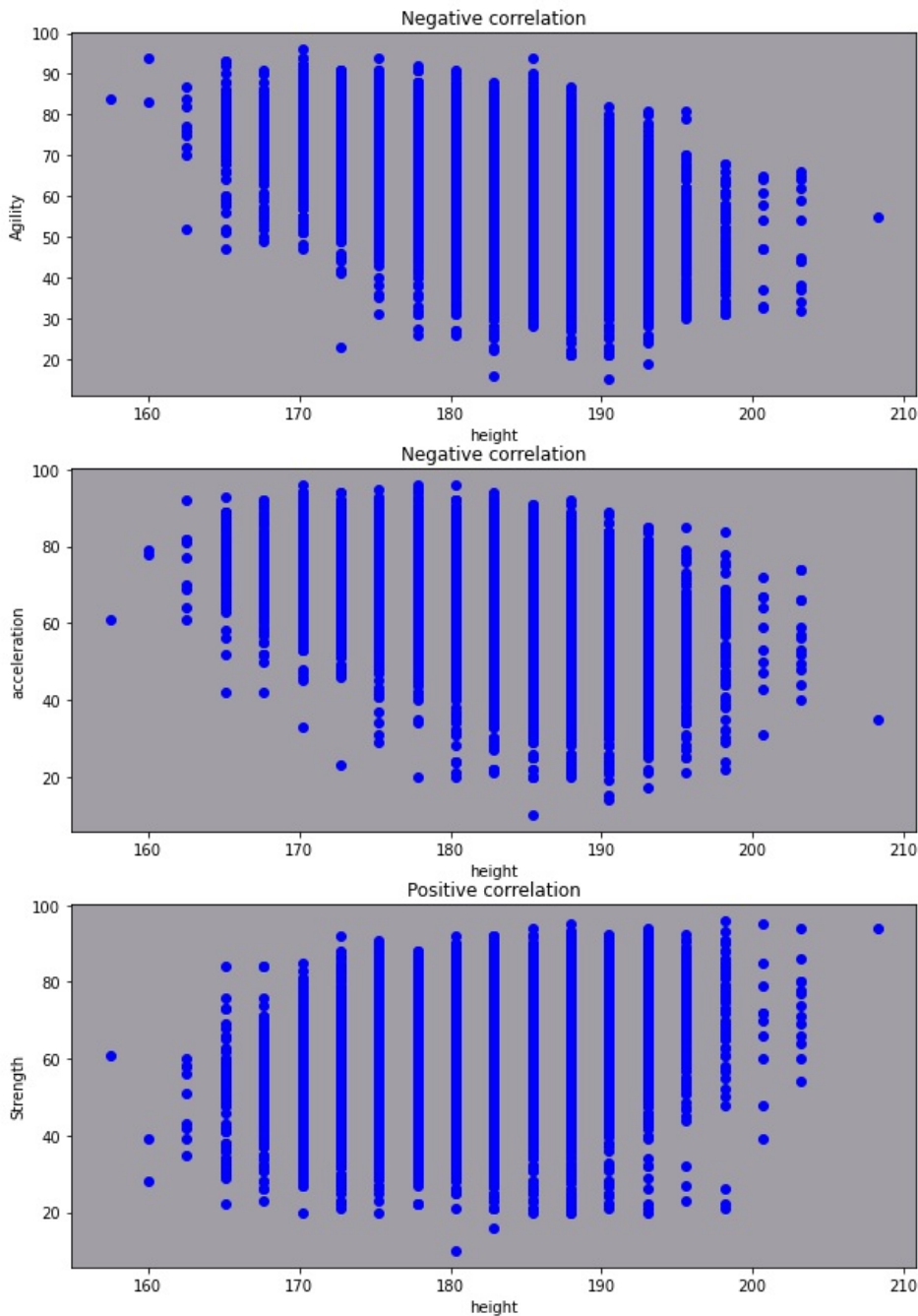


The Height has a negative correlation with agility and acceleration

The Height has a positive correlation with strength

In [67]:

```
# let's build a scatter plot for both correlations
fig, (ax1, ax2, ax3) = plt.subplots(3, figsize=(10, 15))
fig.patch.set_facecolor("#FFFFFF")
height = physical_attr_227['height']
agility = physical_attr_227['agility']
acceleration = physical_attr_227['acceleration']
strength = physical_attr_227['strength']
# scatter plot between weight and agility
ax1.set_facecolor('#A19FA5')
ax1.set_xlabel("height")
ax1.set_ylabel("Agility")
ax1.set_title("Negative correlation")
ax1.scatter(height, agility, color='blue');
# scatter plot between weight and acceleration
ax2.set_facecolor('#A19FA5')
ax2.set_xlabel("height")
ax2.set_ylabel("acceleration")
ax2.set_title("Negative correlation")
ax2.scatter(height, acceleration, color='blue');
# scatter plot between weight and strenght
ax3.set_facecolor('#A19FA5')
ax3.set_xlabel("height")
ax3.set_ylabel("Strength")
ax3.set_title("Positive correlation")
ax3.scatter(height, strength, color='blue');
```



Research Question 3 :



## How did the number of red cards change from 2008 to 2016?

### Main Points :

1. define a function to filter the cards\_info data by card\_tpe ,date and get the sum of red cards that a player got within this date
2. call the function using different dates
3. show how the number of red cards has changed from one date to another

-----define a function to filter the cards\_info data-----

In [68]:

```
# view the range of date
cards_info['date'].describe()
```

```
<ipython-input-68-8ef54e11ace5>:2: FutureWarning: Treating datetime data as categorical rather than
numeric in `.describe` is deprecated and will be removed in a future version of pandas. Specify `dat
etime_is_numeric=True` to silence this warning and adopt the future behavior now.
cards_info['date'].describe()
```

Out[68]:

```
count          61767
unique          1347
top    2015-04-04 00:00:00
freq           184
first    2008-08-15 00:00:00
last     2016-05-25 00:00:00
Name: date, dtype: object
```

In [69]:

```
# will use this function multiple times with different date from 2008 to 2016
def investigate_red_cards( df, table_name, start_date, end_date, card_type):
    """
    this function is to filter the cards_info data by date and sum over card_type then join it with players
    table using sql
    inputs :
    df -----> DataFrame-----> dataframe to store the result
    table_name -----> String -----> the name of the table
    start_date -----> String -----> the date to filter with in format yyyy0101
    end_date -----> String -----> the date to filter with in format yyyy0101
    ouputs :
    df -----> DataFrame-----> dataframe after filtering and joining with players table
    """
    if card_type == 'y':
        card_type = '!= '
    else :
        card_type = '=='
    # we have to filter players with red cards and within satrt_date to end_date
    query = 'card_type {} "r" and {}< date <{}'.format( card_type, start_date, end_date)
    df = cards_info.query(query)
    # we want to group the data by player id and count corresponding card type
    df = df.groupby(['player_id'])['card_type'].count().reset_index()
    # check if the table exist or not
    try :
        # save the dataframe into a table called red_card_2008 load it into the database
        load_table( df, table_name, False, {'player_id':Integer(),'card_type':Integer()})
    except ValueError:
        print("The table already exists")
    # query the result of joining the two tables players and red_card_2008
    query = """
    SELECT p.player_name,r.card_type
    FROM (SELECT DISTINCT player_api_id,player_name FROM Player) as p
    JOIN {} as r on(p.player_api_id = r.player_id);
    """.format(table_name)
    df = make_Query(query)
    return df
```

-----call the function using different dates-----

In [70]:

```
red_card_2008 = pd.DataFrame()
red_card_2008 = invistigate_red_cards( red_card_2008, 'red_card_2008', '20080101', '20090101','r')

yellow_card_2008 = pd.DataFrame()
yellow_card_2008 = invistigate_red_cards( yellow_card_2008, 'yellow_card_2008', '20080101', '20090101','y')
```

The table already exists  
The table already exists

In [71]:

```
red_card_2009 = pd.DataFrame()
red_card_2009 = invistigate_red_cards( red_card_2009, 'red_card_2009', '20090101', '20100101','r')

yellow_card_2009 = pd.DataFrame()
yellow_card_2009 = invistigate_red_cards( yellow_card_2009, 'yellow_card_2009', '20090101', '20100101','y')
```

The table already exists  
The table already exists

In [72]:

```
red_card_2010 = pd.DataFrame()
red_card_2010 = invistigate_red_cards( red_card_2010, 'red_card_2010', '20100101', '20110101','r')

yellow_card_2010 = pd.DataFrame()
yellow_card_2010 = invistigate_red_cards( yellow_card_2010, 'yellow_card_2010', '20100101', '20110101','y')
```

The table already exists  
The table already exists

In [73]:

```
red_card_2011 = pd.DataFrame()
red_card_2011 = invistigate_red_cards( red_card_2011, 'red_card_2011', '20110101', '20120101','r')

yellow_card_2011 = pd.DataFrame()
yellow_card_2011 = invistigate_red_cards( yellow_card_2011, 'yellow_card_2011', '20110101', '20120101','y')
```

The table already exists  
The table already exists

In [74]:

```
red_card_2012 = pd.DataFrame()
red_card_2012 = invistigate_red_cards( red_card_2012, 'red_card_2012', '20120101', '20130101','r')

yellow_card_2012 = pd.DataFrame()
yellow_card_2012 = invistigate_red_cards( yellow_card_2012, 'yellow_card_2012', '20120101', '20130101','y')
```

The table already exists  
The table already exists

In [75]:

```
red_card_2013 = pd.DataFrame()
red_card_2013 = invistigate_red_cards( red_card_2013, 'red_card_2013', '20130101', '20140101','r')

yellow_card_2013 = pd.DataFrame()
yellow_card_2013 = invistigate_red_cards( yellow_card_2013, 'yellow_card_2013', '20130101', '20140101','y')
```

The table already exists  
The table already exists

In [76]:

```
red_card_2014 = pd.DataFrame()
red_card_2014 = invistigate_red_cards( red_card_2014, 'red_card_2014', '20140101', '20150101','r')

yellow_card_2014 = pd.DataFrame()
yellow_card_2014 = invistigate_red_cards( yellow_card_2014, 'yellow_card_2014', '20140101', '20150101','y')
```

The table already exists  
The table already exists

In [77]:

```
red_card_2015 = pd.DataFrame()
red_card_2015 = invistigate_red_cards( red_card_2015, 'red_card_2015', '20150101', '20160101','r')

yellow_card_2015 = pd.DataFrame()
yellow_card_2015 = invistigate_red_cards( yellow_card_2015, 'yellow_card_2015', '20150101', '20160101','y')
```

The table already exists  
The table already exists

In [78]:

```
red_card_2016 = pd.DataFrame()
red_card_2016 = invistigate_red_cards( red_card_2016, 'red_card_2016', '20160101', '20170101','r')

yellow_card_2016 = pd.DataFrame()
yellow_card_2016 = invistigate_red_cards( yellow_card_2016, 'yellow_card_2016', '20160101', '20170101','y')
```

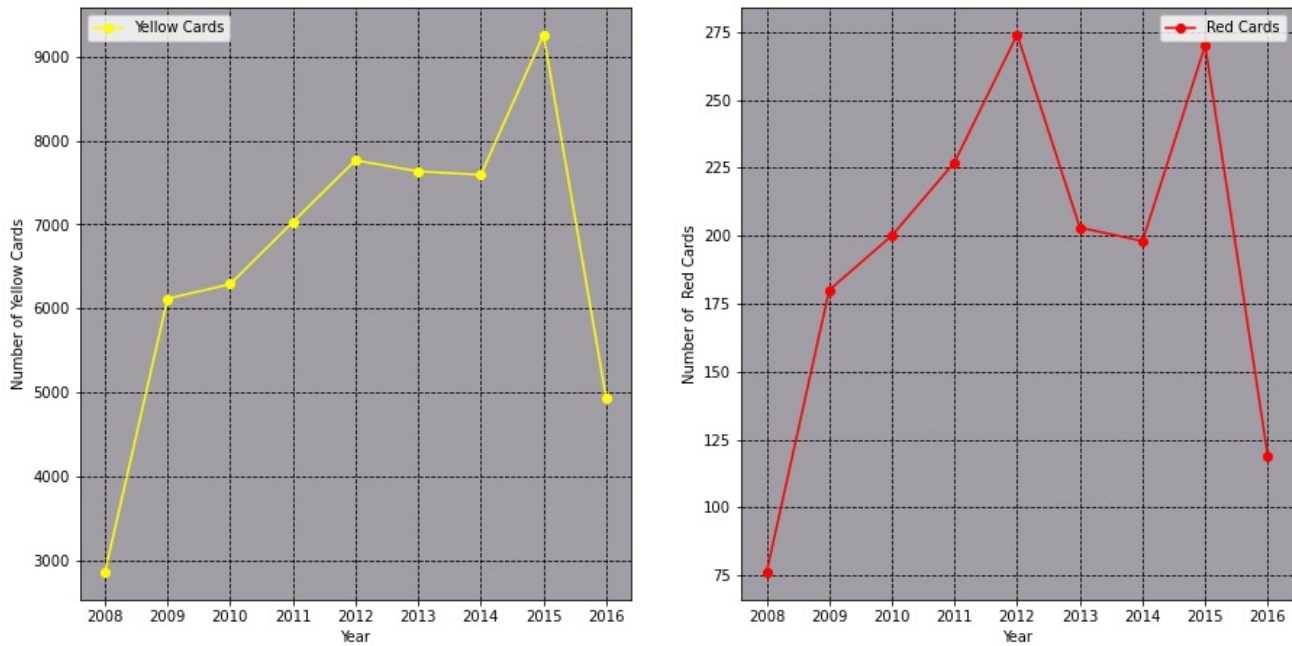
The table already exists  
The table already exists

-----show how the number of red cards has changed-----

In [79]:

```
# let's store the sum of all red/yellow cards in each season
red_cards_full = [ red_card_2008, red_card_2009, red_card_2010, red_card_2011, red_card_2012, red_card_2013, red_
card_2014, red_card_2015, red_card_2016]
yellow_cards_full = [ yellow_card_2008, yellow_card_2009, yellow_card_2010, yellow_card_2011, yellow_card_2012, y
ellow_card_2013, yellow_card_2014, yellow_card_2015, yellow_card_2016]
x = ['2008','2009','2010','2011','2012','2013','2014','2015','2016']
y1 = []
y2 = []
for element in red_cards_full :
    y1.append(element.card_type.sum())
for element in yellow_cards_full :
    y2.append(element.card_type.sum())
fig , (ax1 , ax2) = plt.subplots(1,2,figsize=(15,7.5)) # we will plot two figures
fig.suptitle(" Comparison between number of red cards and yellow cards"+"\\nfrom 2008 to 2016")
fig.patch.set_facecolor("#FFFFFF")
# plot first line chart for yellow cards
ax1.set_xlabel("Year")
ax1.set_ylabel("Number of Yellow Cards")
ax1.set_facecolor('#A19FA5')
ax1.grid(color='black',linestyle='dashed')
ax1.plot(x,y2,color= 'yellow',markersize = 12,marker='.',linewidth = 1.5,label="Yellow Cards");
ax1.legend();
# plot first line chart for yellow cards
ax2.set_xlabel("Year")
ax2.set_ylabel("Number of Red Cards")
ax2.set_facecolor('#A19FA5')
ax2.grid(color='black',linestyle='dashed')
ax2.plot(x,y1,color= 'red',markersize = 12,marker='.',linewidth = 1.5,label="Red Cards");
ax2.legend();
```

Comparison between number of red cards and yellow cards  
from 2008 to 2016



## Conclusions

- Most Teams in Spain LIGA BBVA league score in the home stadium more than in the away stadium
- Real Madrid and Barcelona score more than 100 goals most of the time each season
- Villarreal and Valencia are often close to each other when comparing the number of goals in each season
- The highest number of goals was made by Real Madrid in season 2011/2012
- The weight of the players are normally distributed around the mean which is 168 kg
- The height of the players are normally distributed around the mean which is 182 cm
- The height of the players are normally distributed around the mean which is 182 cm
- Once a player is tall and gains weight there is a high chance to have a low agility score and a high acceleration score
- In 2008 nearly 3000 yellow cards were recorded and increased by the double in 2009 and continued to increase to nearly 8000 cards till 2012 and decreased again to 7500 cards till 2014 and reached its peak with more than 9000 cards and decreased again to 5000
- The red cards follow the same pattern as the yellow cards the maximum number of red cards was recorded is 275 cards in 2012 and the minimum number of red cards was recorded is 75 cards in 2008

## Limitations

Lack of information is one of the limitations that I have encountered while dealing with this database *for example* there wasn't a foreign key `team_api_id` in the players' table that prevents us from relating the players' table with the teams' table. In addition to that, data inconsistency is another issue the database itself is consistent but some columns like `card`, `goal`, `shot off`, etc. written in XML format aren't consistent *for example* : in the `card` column, you might find a `player_id` but without specifying `card_type` or vice versa. Other issue like different date ranges were used *for example* : in the players' table, you might find some players' date ranges from 2008 to 2012 and others from 2012 to 2014 that make it impossible to deal with all the players on the same date.

## Resources

- This website helped me a lot to understand the structure of xml file visit the site [here \(https://www.anyjson.in/xml-to-table\)](https://www.anyjson.in/xml-to-table)
- Kaggle website (<https://www.kaggle.com/datasets/hugomathien/soccer/discussion?search=g>) for soccer database
- This website taught me [LaTeX mathematical symbols \(https://oeis.org/wiki/List\\_of\\_LaTeX\\_mathematical\\_symbols\)](https://oeis.org/wiki/List_of_LaTeX_mathematical_symbols)