University of Dhaka
Department of Computer Science and Engineering

STAT – 2122
Survey on Spanish Premier League

Team Members:
Md. Shahriar Hossain Sajib: 06
Md. Redwan Rahman Rizvee: 09
Md. Zahidul Hasan: 11
Amar Debnath: 23
Sherajus Salehin: 37
Mesbah Uddin Tanvir: 57
Sadman Sakib: 58.

Submitted To:
Dr. Moinul Islam Zaber
Dept. of CSE
University of Dhaka

Appendix

page

SPANISH LA LIGA

La Liga is one of the most popular professional sports leagues in the world, with an average attendance of 26,741 for league matches in the 2014–15 seasons. This is the sixth-highest of any domestic professional sports league in the world and the third-highest of any professional association football league in the world, behind Germany's Bundesliga and England's Premier League.

This project is about a survey on various information about players of various clubs of Spanish League and finally coming to a final decision about prediction about future matches.For surveying we have considered eight most renowned teams of La Liga. Our respictive teams were **Real Madrid, Barcelona, Athletico Madrid, Real Sociedad, Athletico Bilbao, Celeta,Malaga and Espanyol.**

GOAL OF OUR PROJECT:

The major concerns of our projects are:
1) Based on our collected data over previous seasons, predicting the best club for the upcoming season.
2) Determining the strength and weakness of the teams from Attacking, Defense and Possession Units.
3) Giving a hypothesis between player's performance and with his age.

Scope of the Project:

Majority of our time and effort was given to collect the dataset of each player carefully. So, we have come up with a rich Database of information about each player from the current lineup including last 4 seasons.

The Categories in which we gathered data are:

- Player Name
- CountryAge
- Position
- Height(m)
- Weight(kg)
- Date of birth
- Place of birth
- Previous Clubs
- Season
- Possession
- Performance Score
- Attack

- Defense
- Games
- Shot Accuracy%
- Total shots
- Total Goals
- Assists
- Key Passes
- Passing Accuracy%
- Successful
- Headed
- A V passes length
- Duels Won%
- Tackled
- Suffered %
- Commited %
- Take Ons
- Avg Defensive Actions
- Blocks%
- Interceptions%
- Clearance%
- Defensive Error
- Yellow Card
- Red Card

STATISTICAL ANALYSIS:

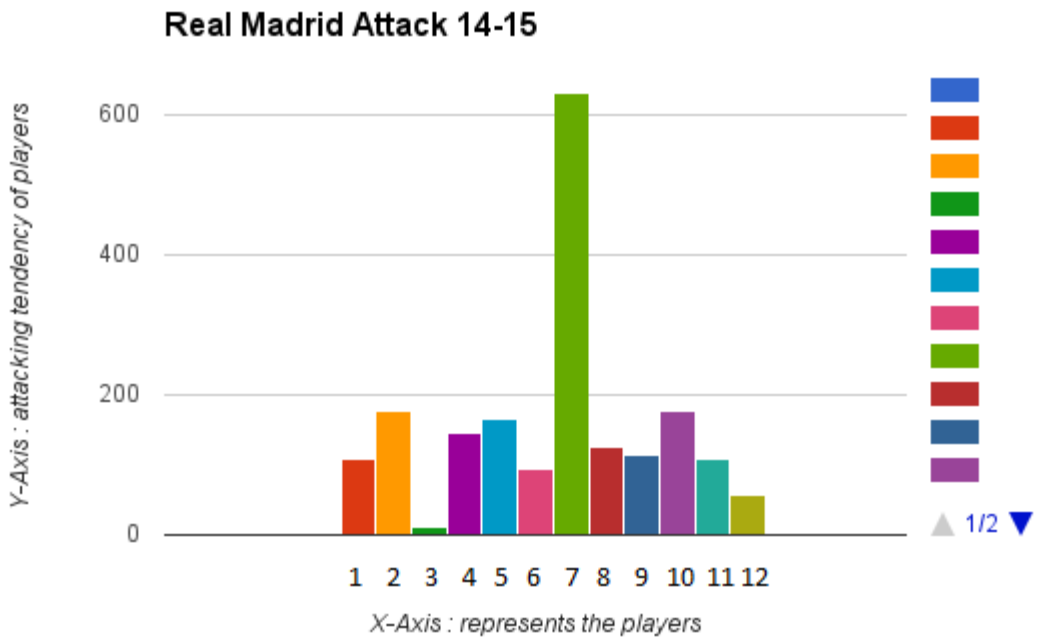
In the following pages we have given some of the attributes, we collected through our analysis.We had huge graphical formations on possession,performance,attack,deffense etc. Here we are giving some of them and discussing their properties.

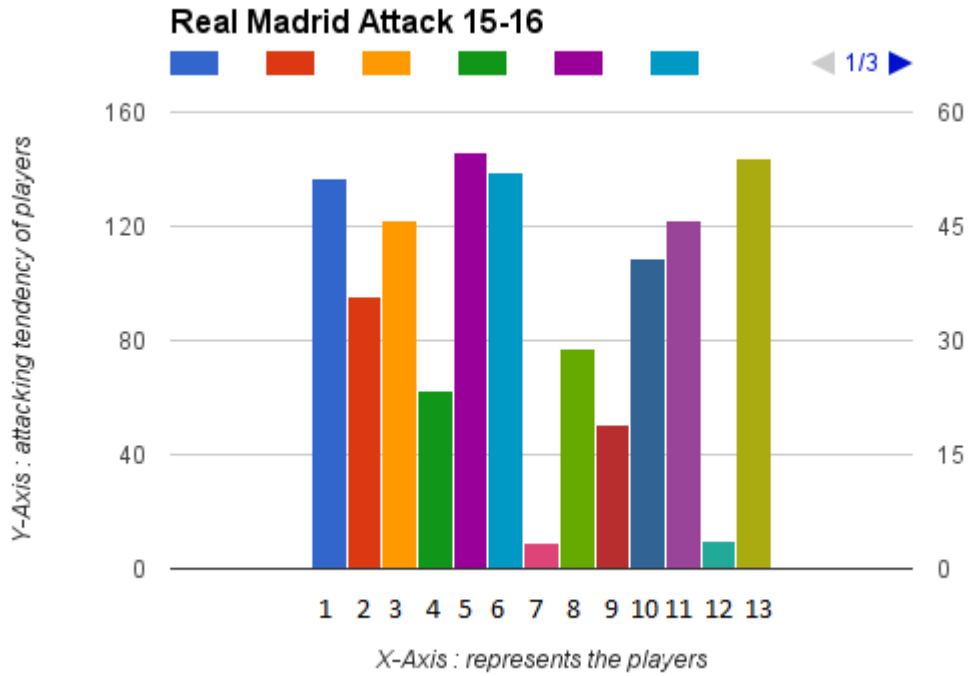
Attack: The attacking power parameter of each player were collected to compare between the players and determine who is the best attacking player of a team. Among them, the best histograms where chosen..

We made the graph by comparing the “attacking tendency” of all players. X axis contains Players and Y axis contains the attacking unit.

Real Madrid:

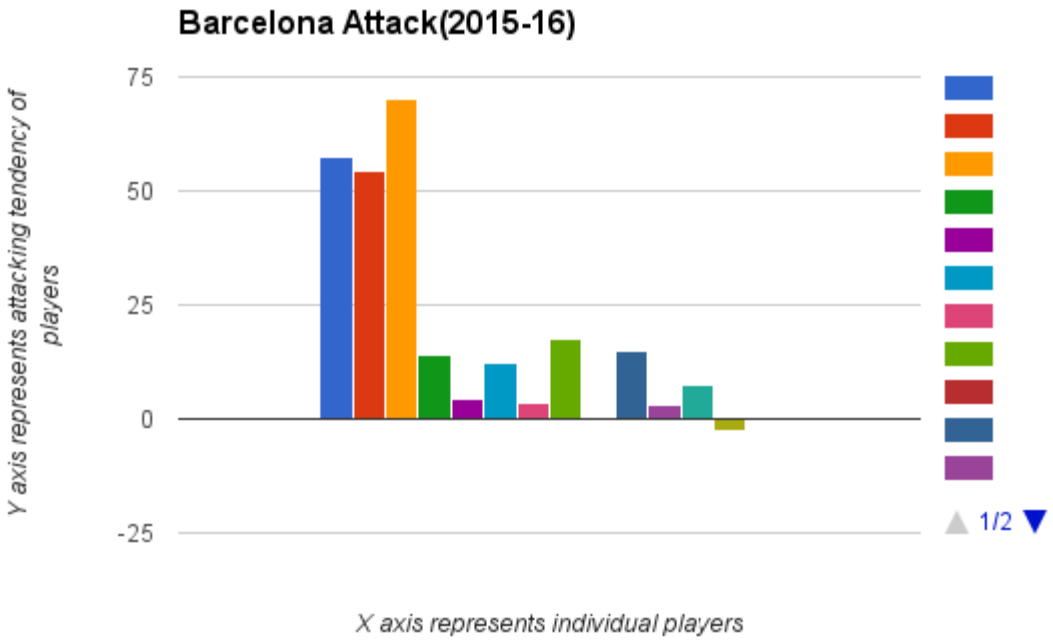
The following graphs show the attacking tendency of each player of Real Madrid who played in season 2015-16. Note that both the graphs for the season 2014-15 and 2015-16 does not have same number of players because many of the players from last season didn’t appear in previous seasons.

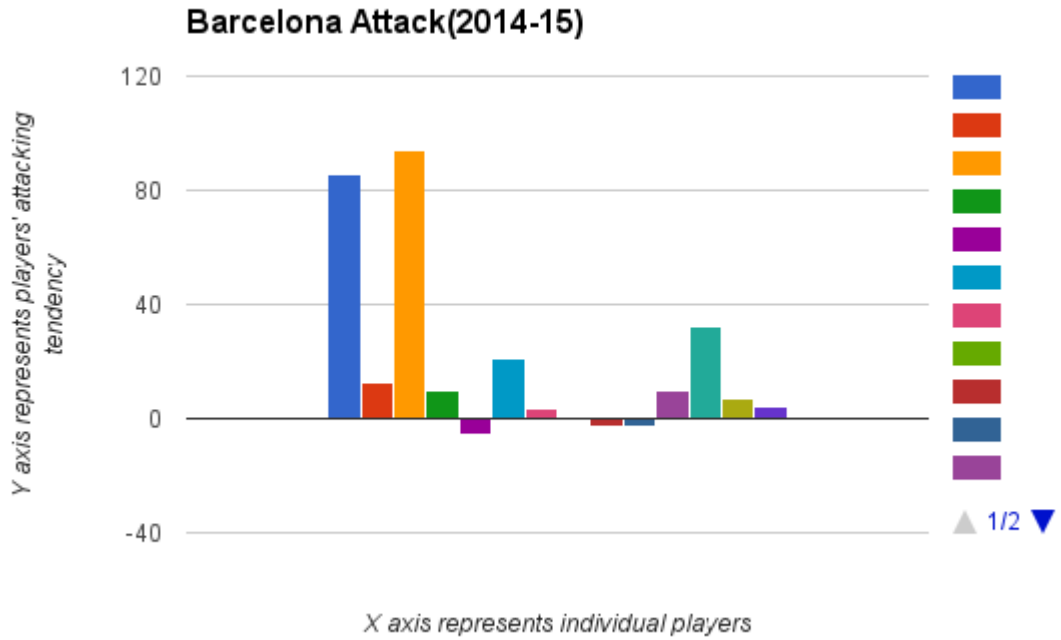




Barcelona:

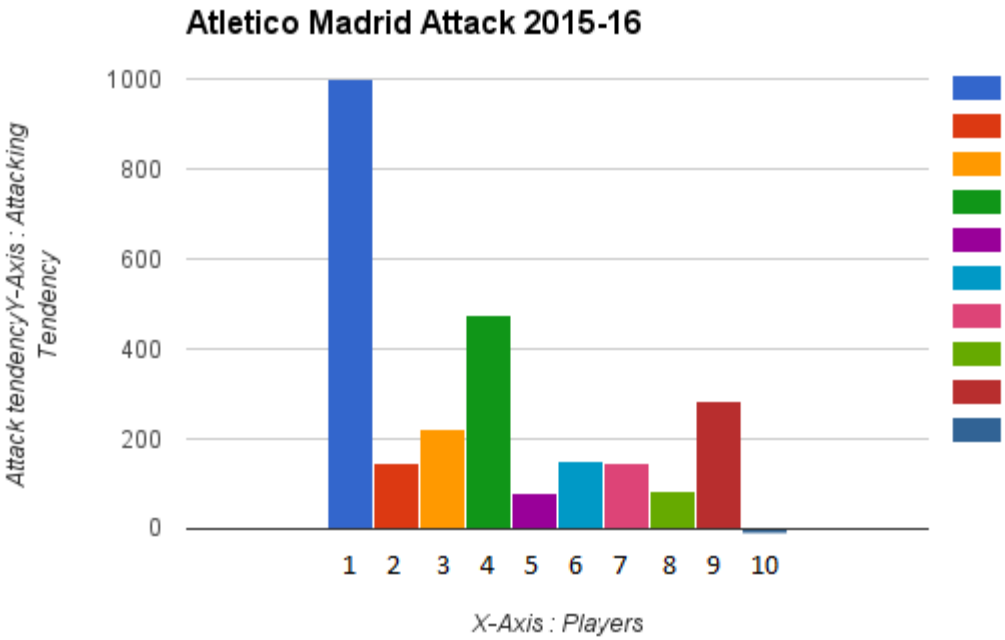
The following graphs show the attacking tendency of each player of Club Barcelona who played in season 2015-16. Note that both the graphs for the season 2014-15 and 2015-16 does not have same number of players because many of the players from last season didn't appear in previous seasons.





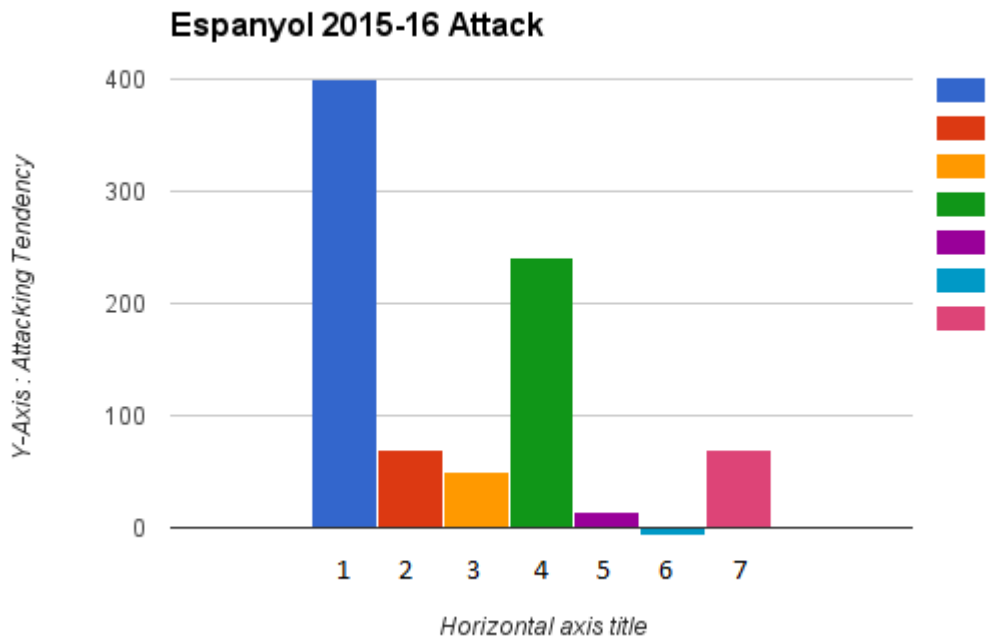
Athletico Madrid:

The following graphs show the attacking tendency of each player from Athletico Madrid who played in season 2015-16.



Espanyol:

The following graphs show the attacking tendency of each player who played in season 2015-16.



Attacking graph actually indicates a team's normal strategy, how they carry their performance in a match. Some of the teams show high attack version, which basically indicates how they create pressure on their opponent in the match. Generally strong team shows high attacking mode to pressurize and to bring out the match. We can generally see, Real Madrid, Barcelona show strong attacking graph and we all know they are two of the best teams in Spanish premiere league.

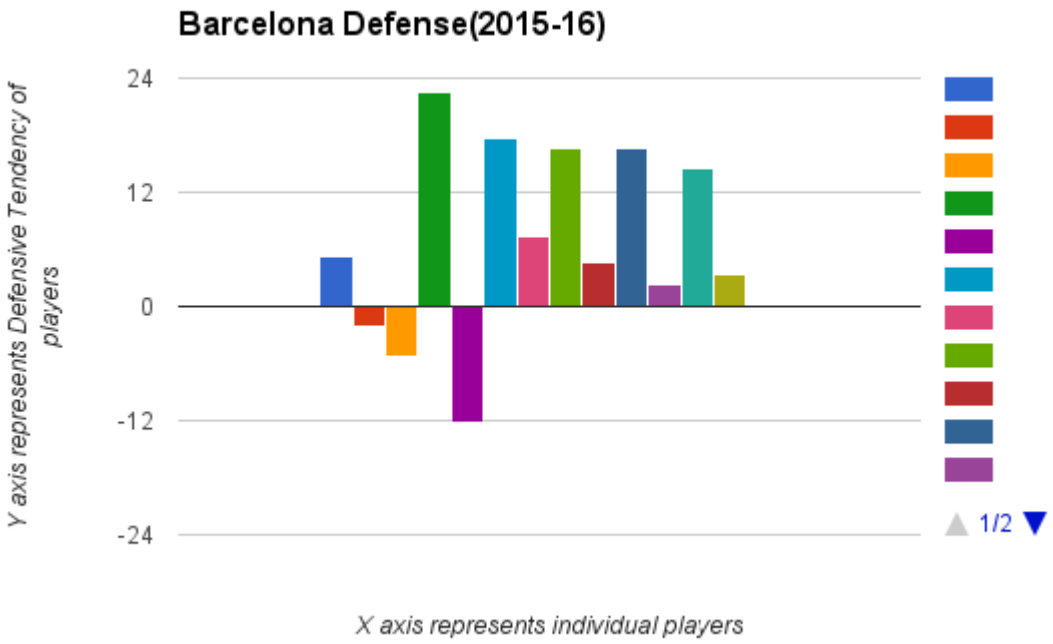
Though, it depends on the match situation and strategy how a team will play. Generally we have shown a normal statistics how a team perform, where from the previous data collection we can actually say Real Madrid and Athletico Madrid show a good scoring performance here. Where other teams we have found not so much attacking team in those years on our respective seasons. However only attacking doesn't show a team's total performance.

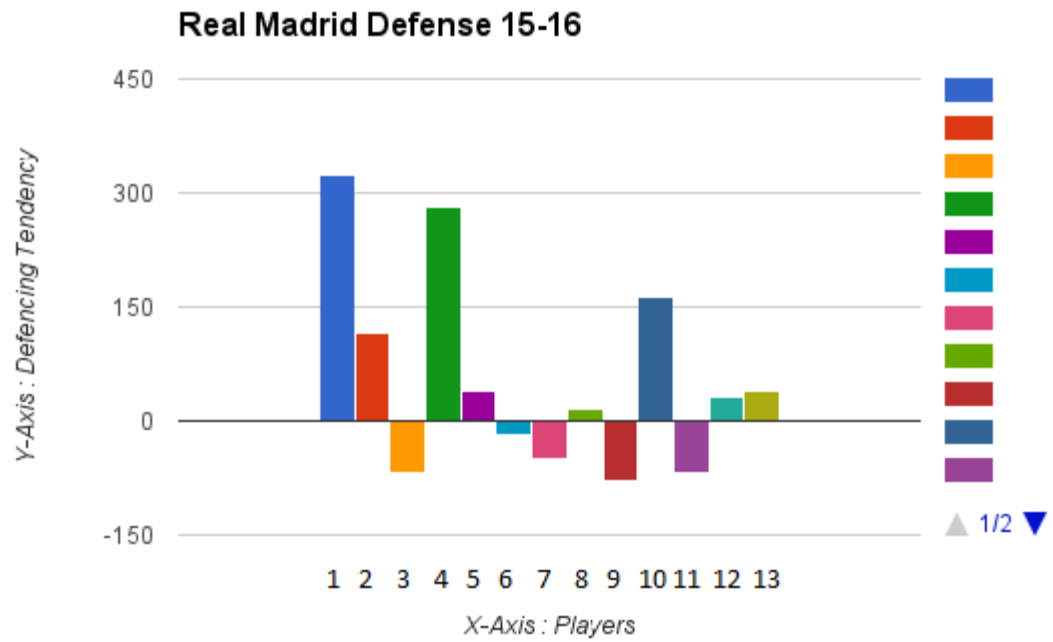
So we have also considered other attributes too. Like defense, possession etc. to find the real performance of a team.

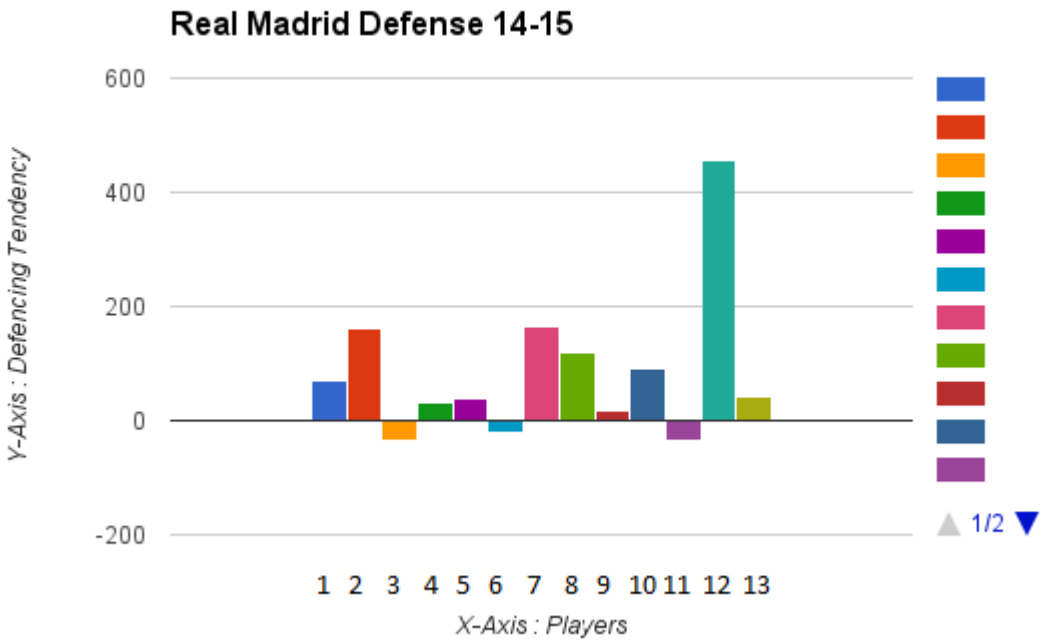
Defense : The defending power parameter of each player were collected to compare between the players and determine who is the best defending player of a team. Among them, the best histograms where chosen..

We made the graph by comparing the “defensing tendency” of all players. X axis contains Players and Y axis contains the defensing unit.

For space and time conviniene we have added one graph from each of the best defencing teams here from last 4 seasons.





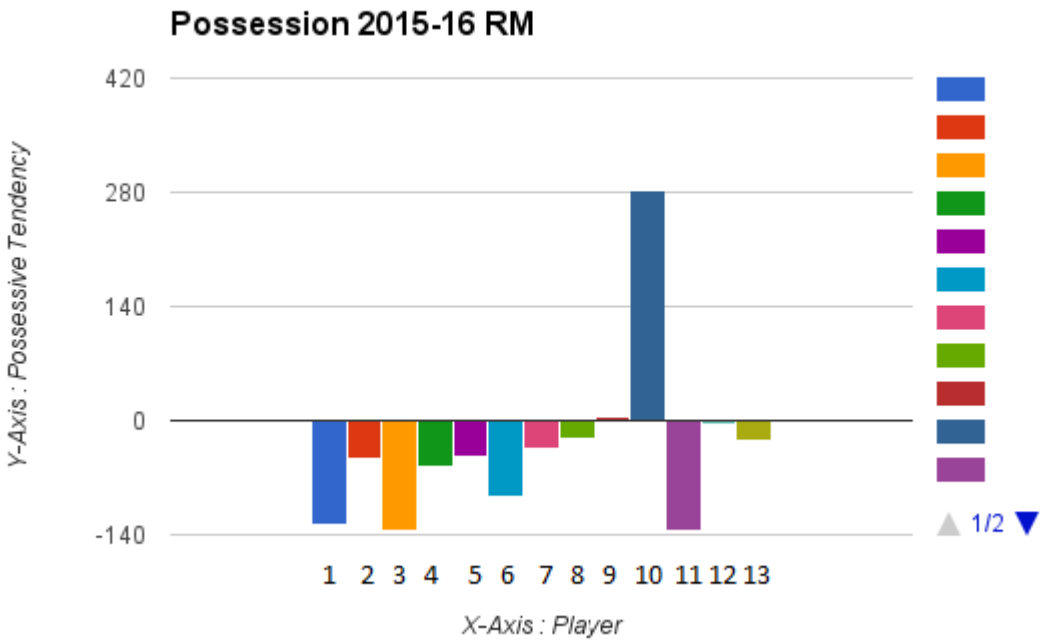


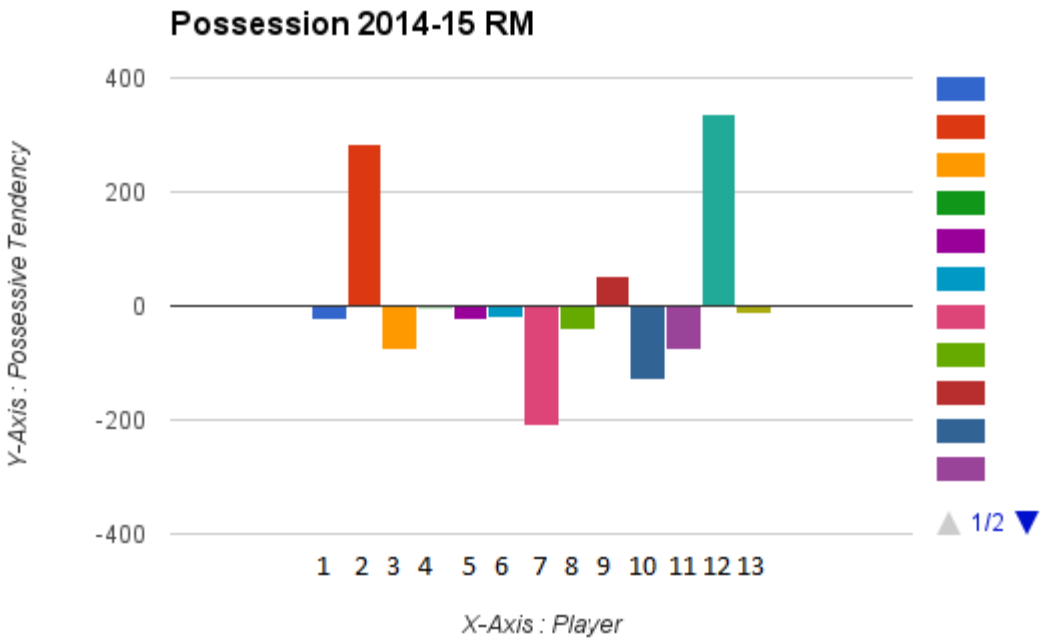
Possession: The possession of ball parameter of each player were collected to compare between the players and determine who could keep the ball from a team. Among them, the best histograms where chosen.

We made the graph by comparing the “possession tendency” of all players. X axis contains Players and Y axis contains the possession unit.

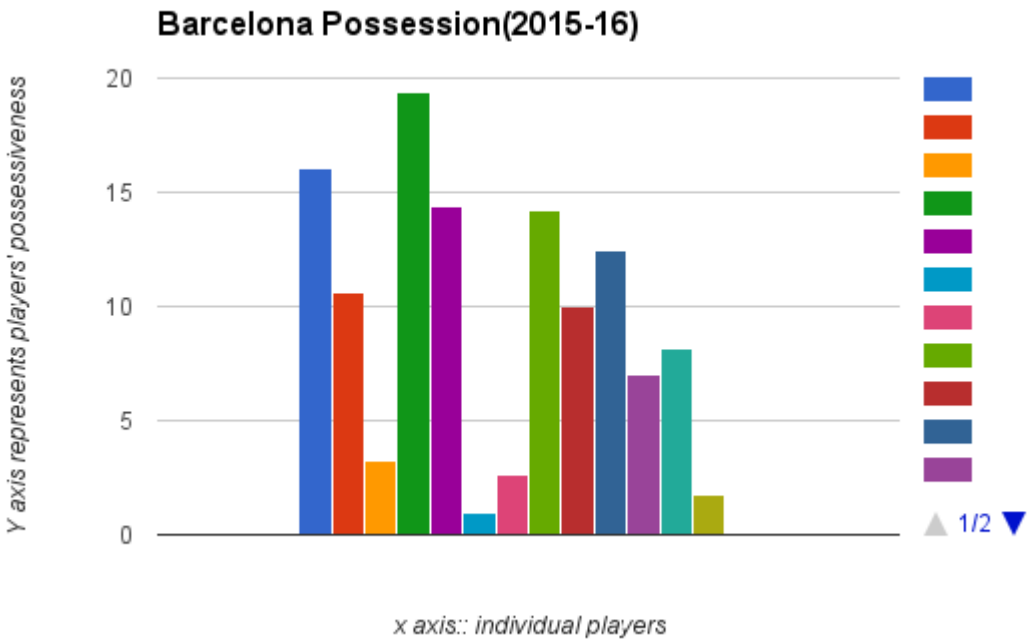
For space and time conviniene we have added one graph from each of the best defencing teams here from last some seasons.Every graph, normally explains in the same way, where x axix denotes the players and y axis denotes their possession.Now every histograms explain in the last some seasons how the team’s players have done in the respect of possesssion. We have included the graph of different teams with their possesssion in the last four years.

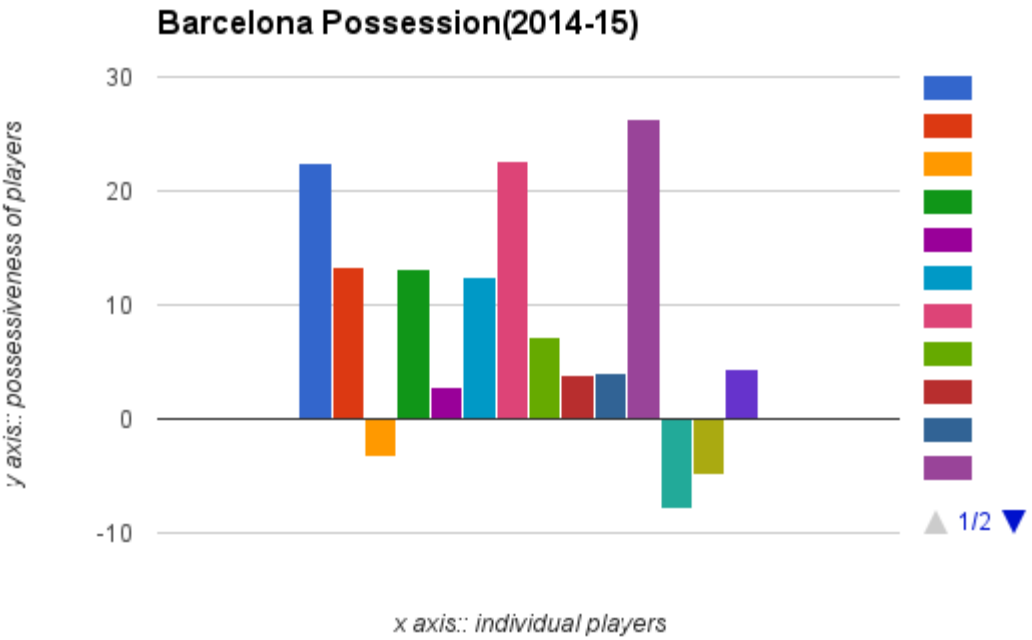
Real Madrid:


Possession 2015-16 RM
420
280
140
0
-140
Y-Axis : Possessive Tendency
1 2 3 4 5 6 7 8 9 10 11 12 13
X-Axis : Player
1/2

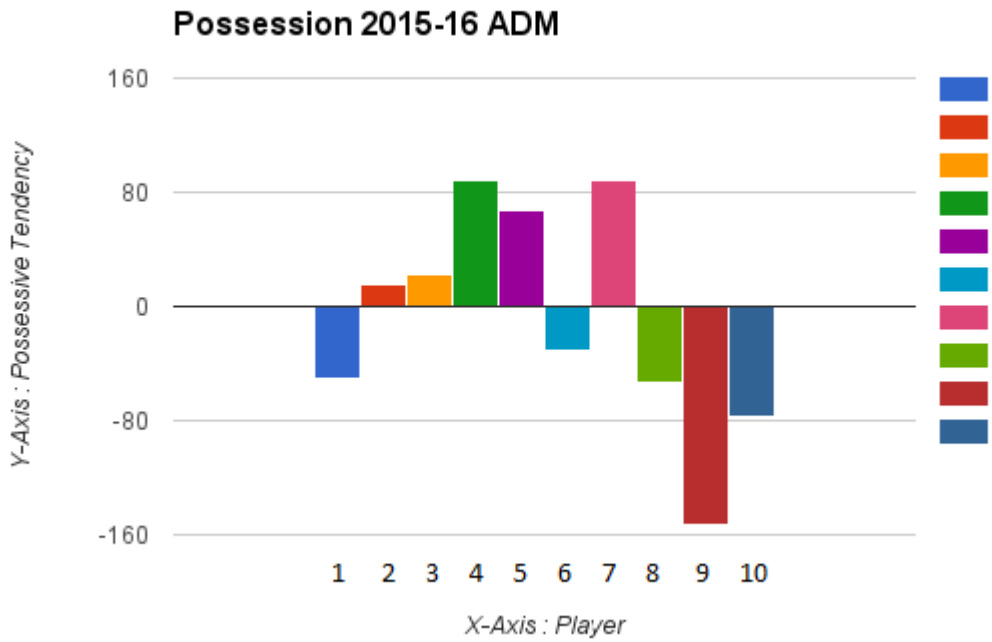

Possession 2014-15 RM
400
200
0
-200
-400
Y-Axis : Possessive Tendency
1 2 3 4 5 6 7 8 9 10 11 12 13
X-Axis : Player
1/2

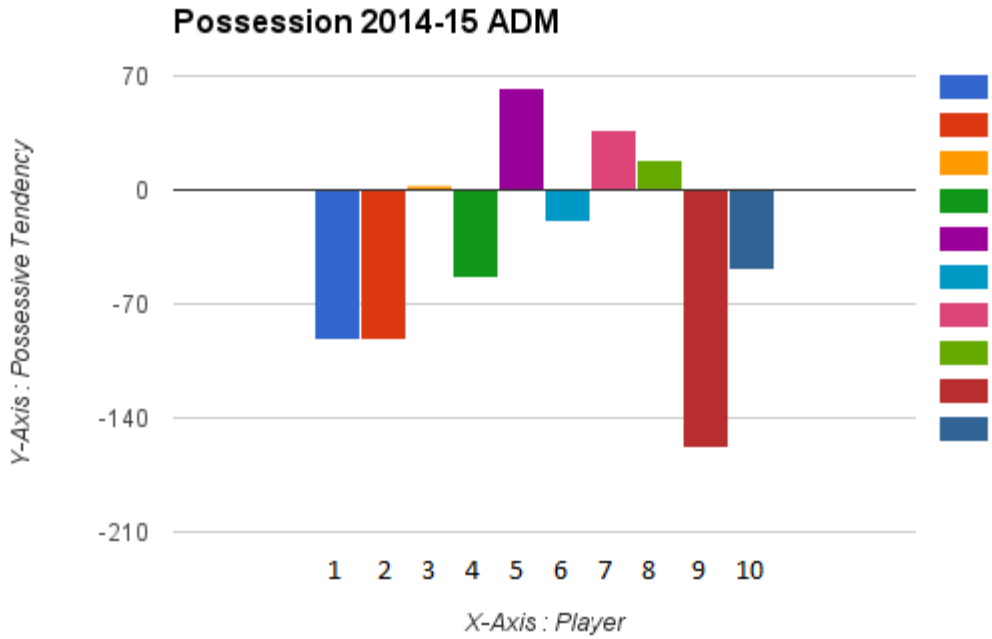
Barcelona


Barcelona Possession(2015-16)
Y axis represents players' possessiveness
20
15
10
5
0
1/2
x axis:: individual players

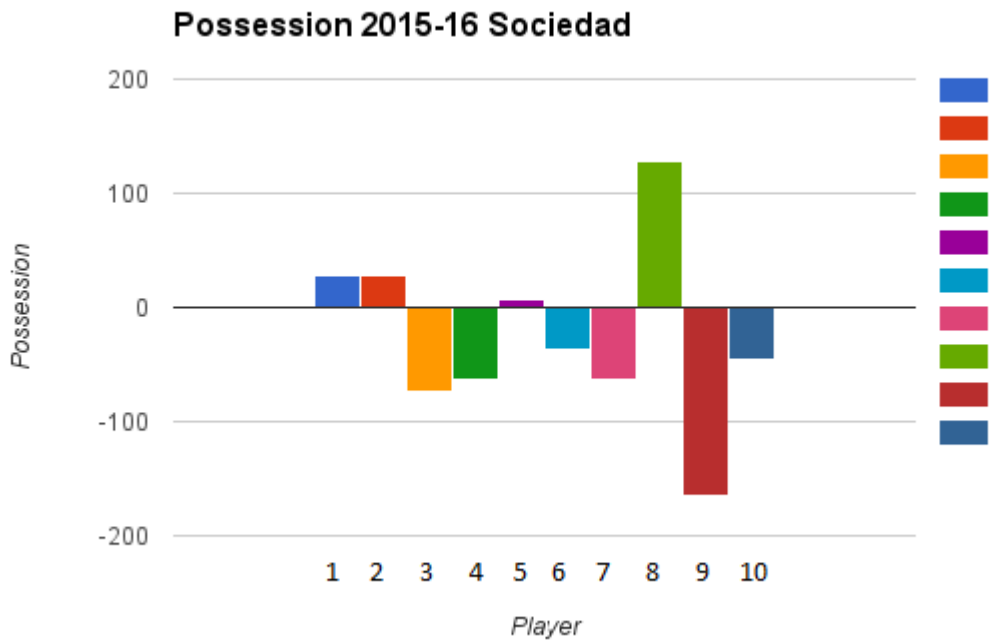

Barcelona Possession(2014-15)
y axis:: possessiveness of players
30
20
10
0
-10
1/2
x axis:: individual players

Athletico De Madrid

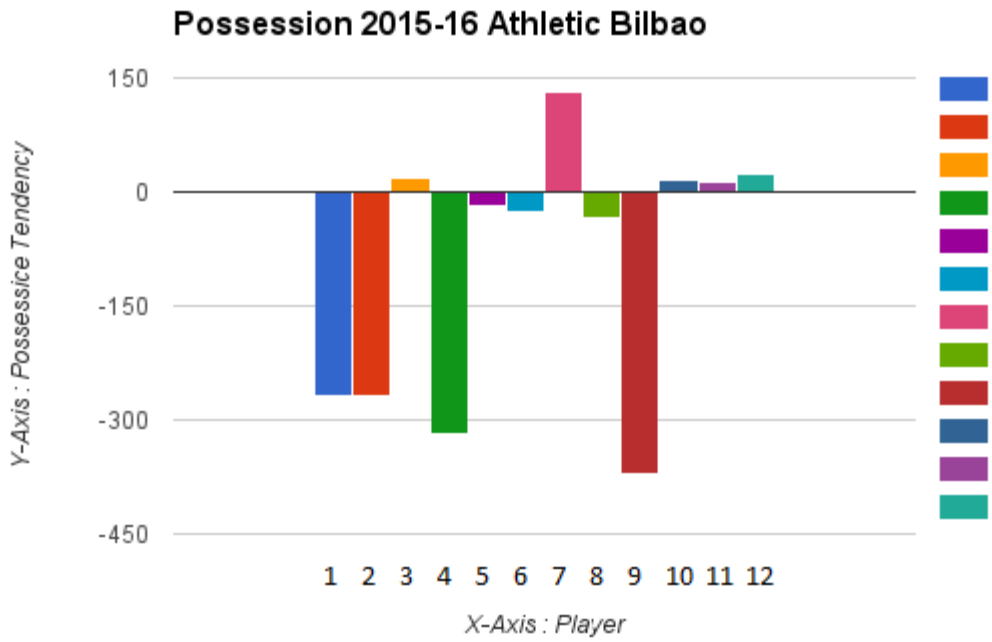




Sociedad



Athletic Bilbao

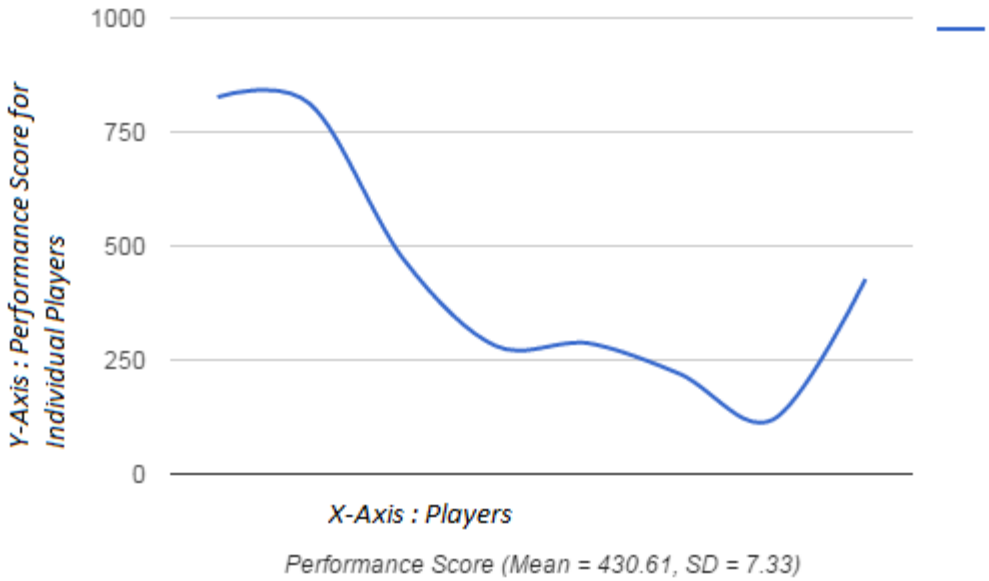


Defense and possession are other very important attributes of the teams to measure their performance on a whole. In general attack, possession, defense all simulate the whole team performance graph. Only attacking team is not good, similarly only possessive team or only defensive also shows nothing alone. The total performance of a whole team sums up what we actually want to measure. If we simulate this things very calculative, we can see, in the attack perspective Real Madrid and Athletico Bilbao shows a very good histograms, due to their performance.
On the while Barcelona had a good defense so also Espanyol and Real Socidad, they also have shown a tremendous performance in defense perspective and possession was also Real Madrid showed a good selection so also Barcelona.

Mathematical Calculations and our Observations:

Best Team Performance:

We have taken our data on Real Madrid, Athletico Bilbao, Barcelona, Athletico Madrid,Celta, Malaga, Espanyol and Real Sociedad. On this 7 teams we have taken data on their player's different type of attributes. And then we formed our statistical calculation to predict which team will play better or may be the best team in the next season.
For this perspective we have taken the season 2015 - 16, 2014 - 15, 2013 - 14 seasons. And then we simulated their player's performance score and formated the mean performance and standard deviation of the teams o this regard and formulated our hypothesis on the "Best Team of 2016 - 17 season".



Now here sample **mean = 430.61** and **standard deviation (sigma) = 7.33** Now we can say the team which have more performance mean then this are better team and who have less than this are worse in comparison to performacne with them. And the lowest comparing to this mean are the most low rated team here.

TeamName	Performance	Prediction
Real Madrid	826.4	Very Good
Barcelona	811.47	very Good
Athletico Madrid	473.48	Good
Athletico Bilbao	281.2424242	Not Good
Celta	286.91	Not Good
Malaga	218.94	Not Good
Espanyol	119.1	Poor
Real Sociedad	427.37	Good
	Mean = 430.61, SD = 7.33	

So, in generel Scenerio "Real Madrid or Barcelona will be the best team in next season."

Best Attacking Team:

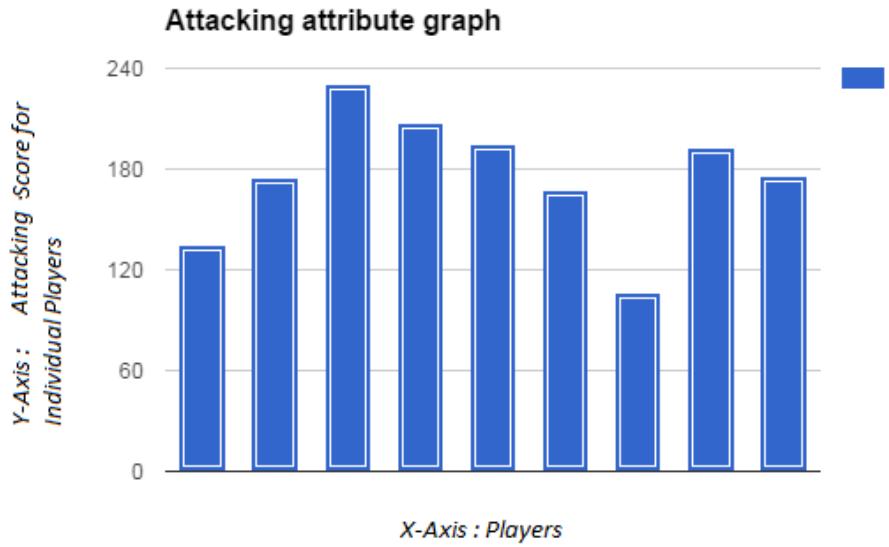
Similarly we have thought on the attributes of “attacking” and given our opinion on the basis of “Best Attacking Team” the team which has more attacking attributes is the best attacking team in our perspective.

TeamName	**Attack Quantity(Cumulative)**	**Prediction(depending on Mean)**
Real Madrid	134.43	Good
Barcelona	175	Good
Athletico Madrid	230.87	Very Good
Athletico Bilbao	206.89	Very Good
Celta	194.33	Very Good
Malaga	167.66	Good
Espanyol	106.138	Not so good
Real Sociedad	192.88	Very Good
	mean = 176.02, SD = 61.95	

Our measuring scenerio:

Very Good	**> 176.02**
Good	**From 114.02 to less than 176.02**
Not So good	**< 114.02**

The following graph compares between Attacking attribute of the best teams, from which we can come up with the desicion about the best team and can conclude with a prediction about who will be the best attacking team for the next season.

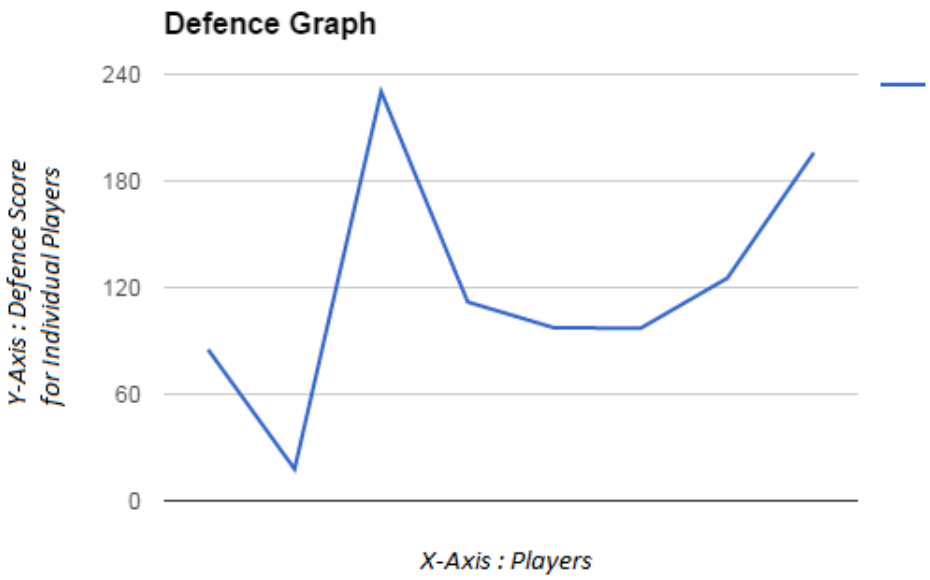


So in our opinion, **"Athletico Madrid is the most attacking team of season 2016 - 2017."**

Best Defensive Team:

Our measuring window was same again and we again gave our openion depending on mean and standard deviation.

TeamName	**Deffensive attribute(Cumulative)**	**Prediction(depending on Mean)**
Real Madrid	195.02	Very Good
Barcelona	17.8	bad
Athletico Madrid	85.02	Not good
Athletico Bilbao	111.87	Not Good
Celta	97.32	Not Good
Malaga	97.14	Not Good
Espanyol	125.29	Good
Real Sociedad	230.24	Very Good
	mean = 176.02,SD = 61.95	



Our measuring window:

Very Good	**> 176.02**
Good	**From 115 to 176.02**
Not Good	**> 54 to < 115**
Bad	**<54**

So, **“Real Socieded is the most defensive team of 2016 - 17 season.”**

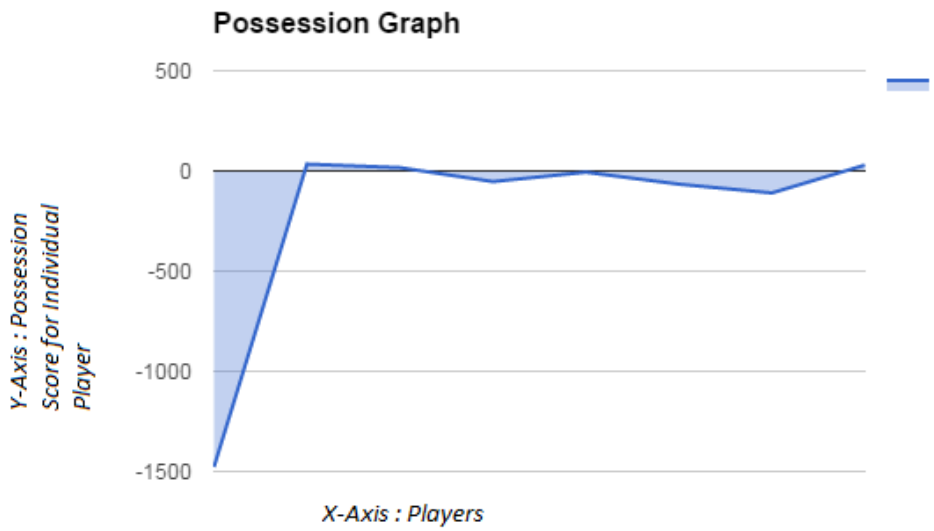
Most Possesive Team:

We have given our prediction on the basis of following data,

Team Name	Possesive Attribute(Cumulative)	Prediction(depending on Mean)
Real Madrid	-1477.76	Not Good
Barcelona	34.31	Best
Athletico Madrid	16.5	Third Best
Athletico Bilbao	-53.1415	Good
Celta	-6.33	Good
Malaga	-66.04	Good
Espanyol	-109.987	Good
Real Sociedad	29.29	Second Best
	mean = -204.15, SD = 483.74	

Our mesuring window

Best	**> -204.15 and which are > 0**
Good	**> -204.15 and < 0**
Not Good	**< - 204.15**



so,"Barcelona is the most possesive team of season 2016 - 17".

Fair Playing Team:
In this field we have thought on the basis of number of red cards, yellow card and on total number of defensive error.

TeamName	Yellow Cards	Red Cards	Total
Real Madrid	199	9	208
Barcelona	116	8	124
Athletico Madrid	181	8	189
Athletico Bilbao	165	12	177
Celta	138	9	147
Malaga	199	9	208
Espanyol	159	10	169
Real Sociedad	159	5	164
	Total = 1316	Total = 70	

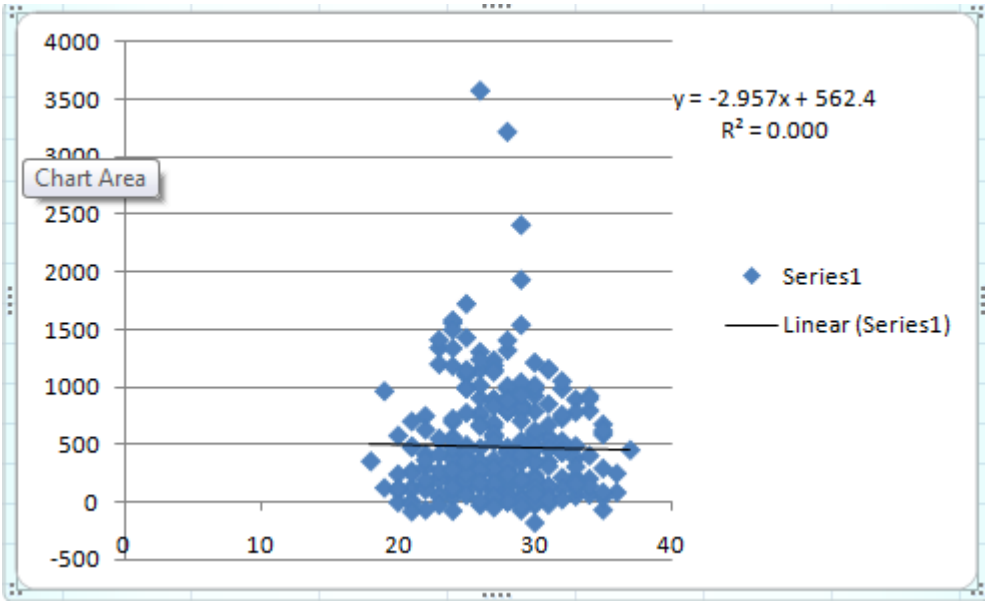
So on the basis of our data collection we can say that, **We have given our prediction that the team which players have got less number of yellow and real cards are fair played team.**
And "Our most fair played team is Barcelona in the season 2016 - 17".

Our Hypothesis: Age Vs Performance

We had a concept that, "how players are doing with their age". We have the following two questions.

1) **Is age a developing parameter for player's performace?**
   We had an ideology that, when a player becomes more experienced he definitely plays good.And experience comes withing age, a player becomes more experienced with his increasing age.
2) **Is age degrading player's performance?**
   Another important terminolgy we thought, as "fitness and speed" is two most important attrbutes in football. Generally this two gets slower with age. Then how a player can do better within age or what can be the best relation between **age and individual performance** of a

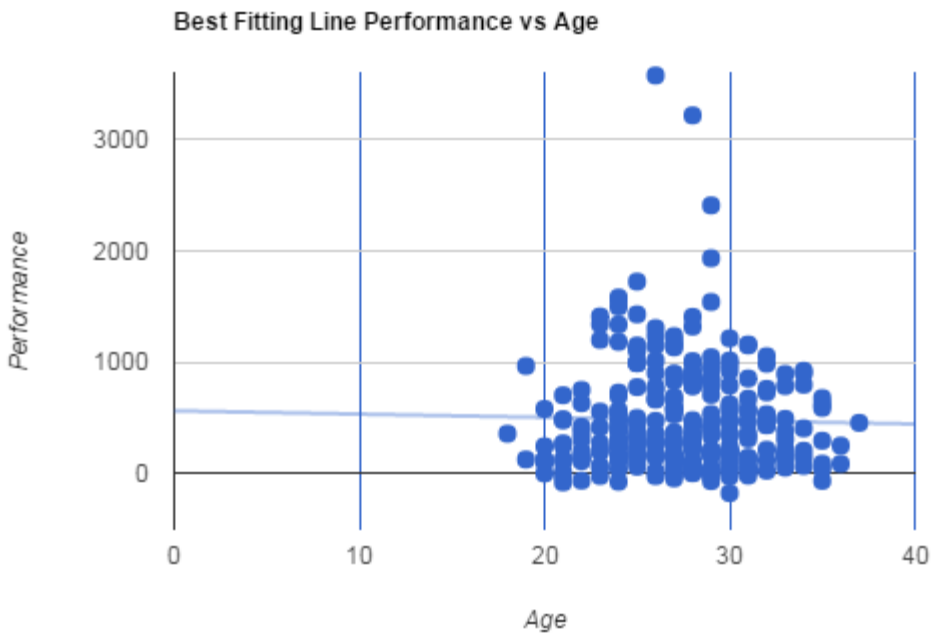
player. So we wanted to find a linear relation between this two.



here y = performance score and x = age of the players. Then we found the best fitting line equation,

y = - 2.957x + 562.4 with regression **R^2 = 0.000**

and the best fitting line

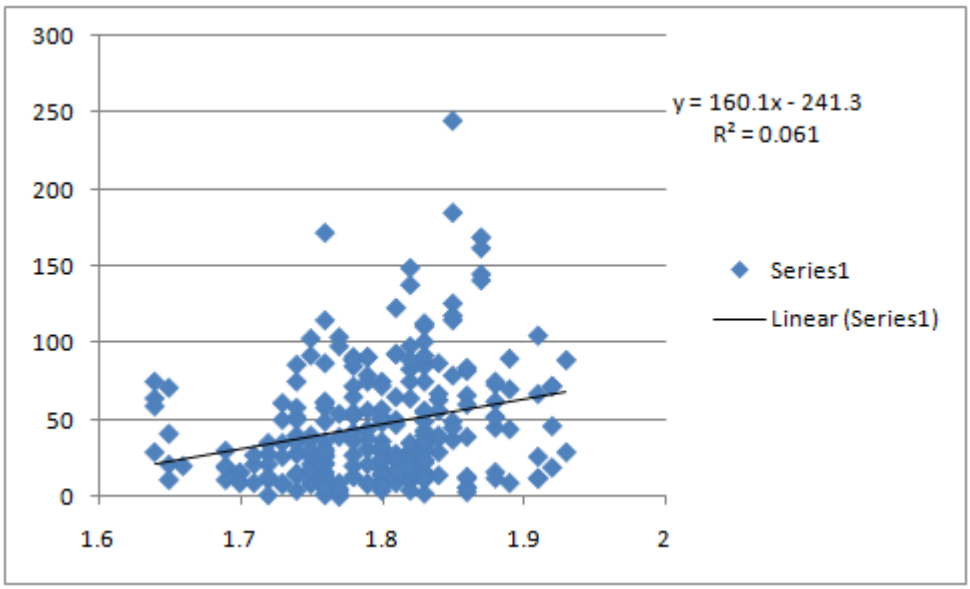


So, **in general we can predict that players generally play good between age 20 and 30. Below 20 there are very few players who have good performance and greater than 30 there are also few players who play well and keep a good fitness.Here we can also see that some players have too much performaance score than others, basically they**

are outliers and generally they are very exceptional players like Christiano Ronaldo , Lionell Messi who are far better than average r normal quality players.

Our Second Hypothesis: Height Vs Number of Heads

Generally we can easily say, “headed” is a attribute of a team in playing game, most possesive team or most attacking team generally show a good number of this type. But, it is a general fact that, the players having a good height gets an extra opportunity to give head shots in matches.So we wanted to find if there is any relation with player’s heights and his number of heads in a match. So we collected the data above two hundred players and the number of heads they gave per season in spanish premier league and got a following graph and realtion.

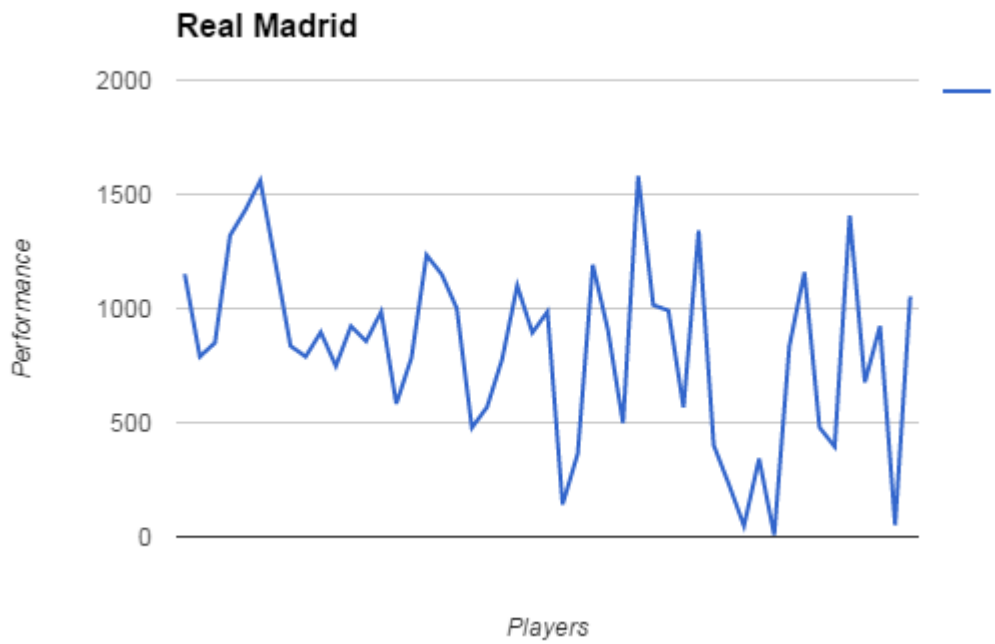


We actually understood that there is a relation among player’s heights and with their giving heads in matches and found a best fitting line **y = 160.1x - 241.3**, and regression factor **R2 = 0.061**. Basically we also found a best fitting line which is shown in the graph. And from there we came to a conclusion that **“player’s height from 1.75 to 1.8 genrally shows a good number of heads in the matches”**

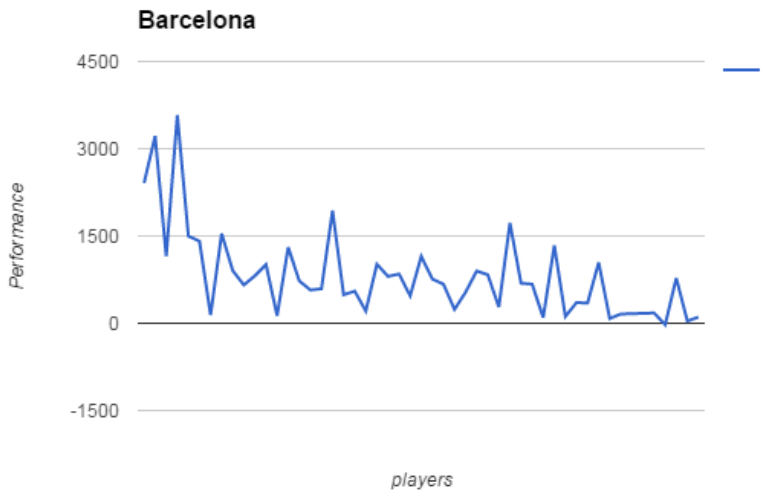
Skewness:

We have calculated the skewness for each of the team on depending on their previous seasons. Here a result of our calculation.

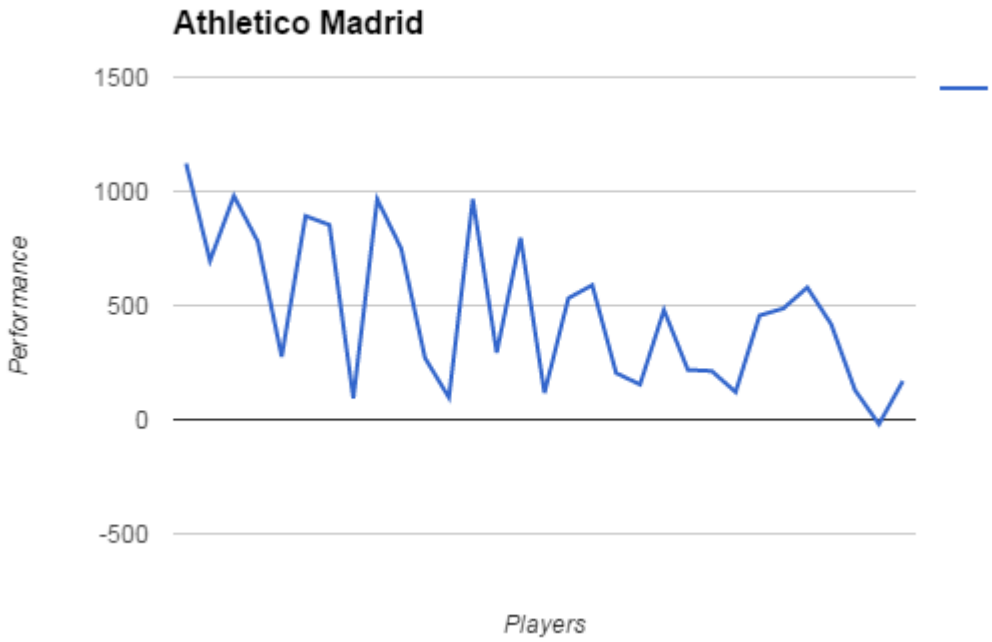
Real Madrid: Skewness for Real Madrid = - 0.26, Real Madris has negative skew, it means most of the players performance are on the left side of the mean.



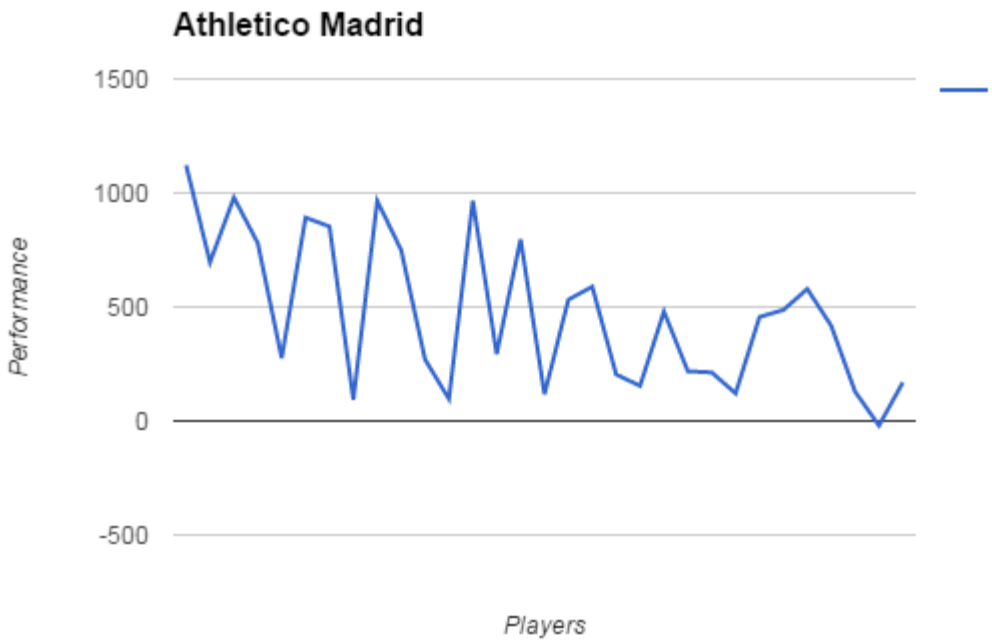
Barcelona: Skewness 1.85, it means most of the player's are on the right side of the mean.



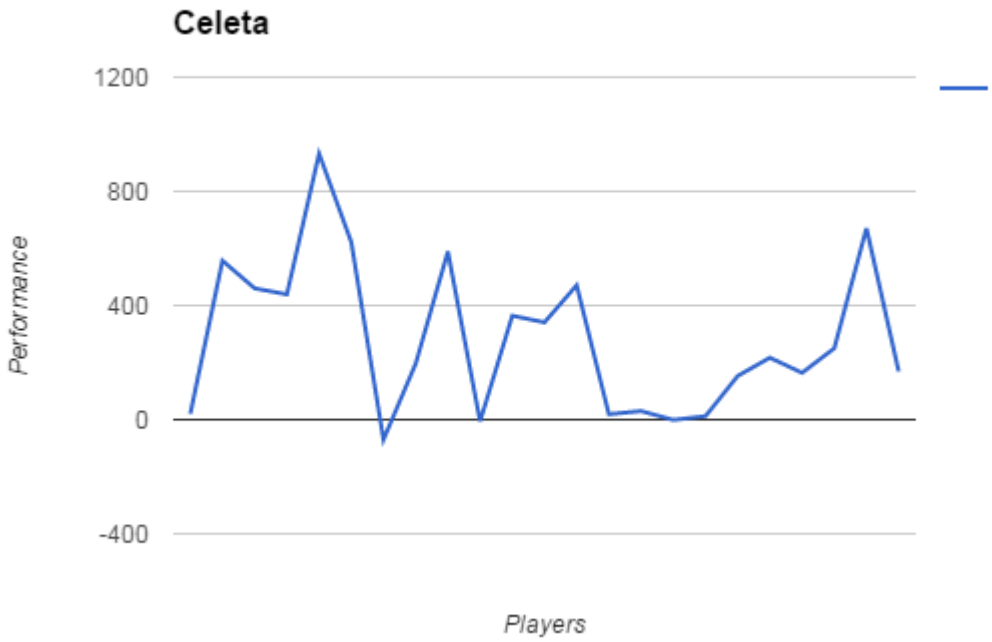
Athletico Madrid: Skewness 0.37, it means most of the players on the right side of mean for this team but they are very close.



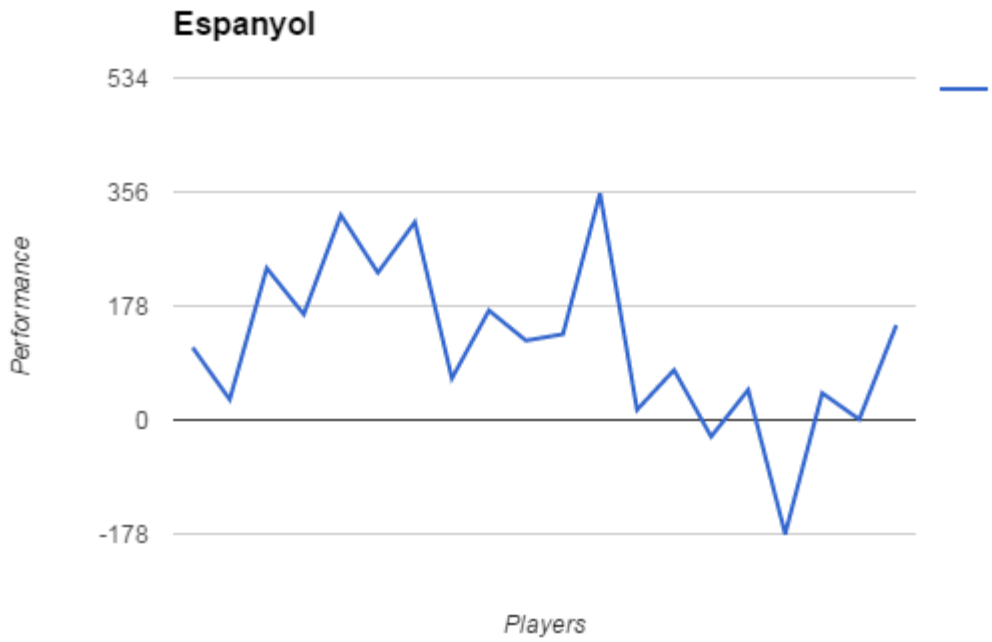
Athletico Bilbao: Skewness 0.022, they are also positively skewed , most of the players are on the right side of the mean and very close to the mean.



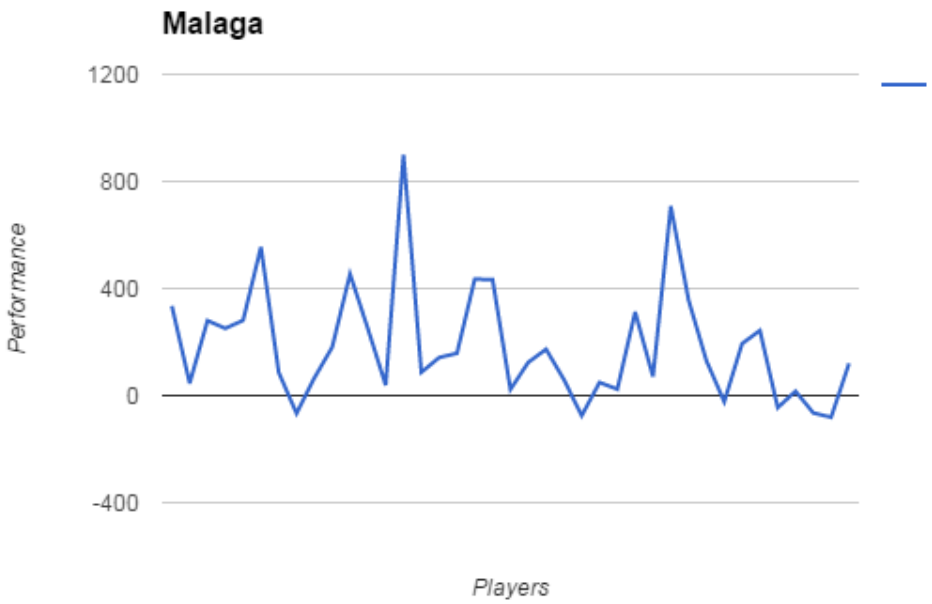
Celeta: Skewness for celeta 0.64, it also indicated also same like previous teams with positive skewness.



Espanyol: Skewness - 0.068, it has negative skewness, most of the players are on the left side of the mean and vey close to the mean.



Malaga: Skewness 1.36. positive skewness and far from mean on the right side.



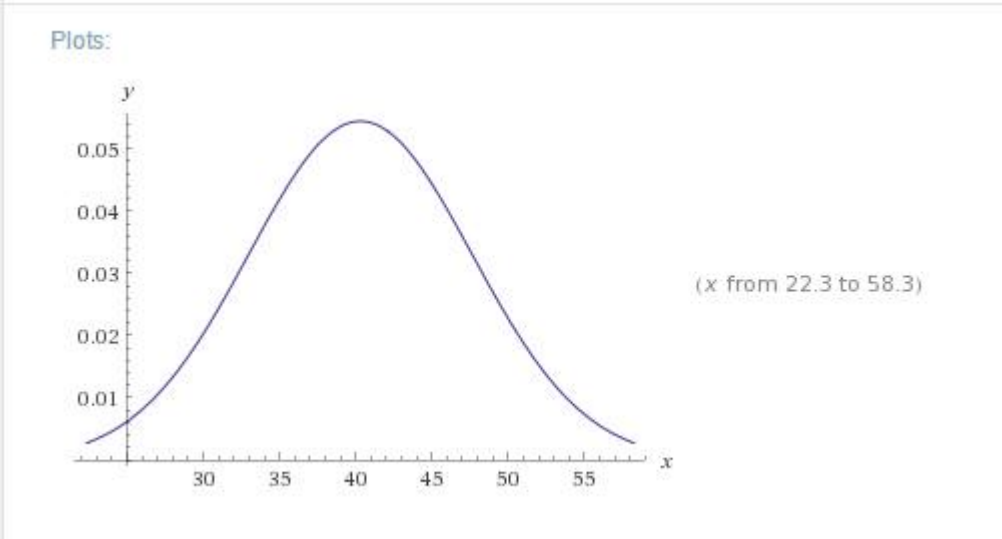
Skewness actually determines the frequencies around the mean. If we think about the data carefully we will see some teams have higher and positive skew where some team have negative skew it totally determines how the team has it's players so that they can generally say that most players of them are up to average or below to average. But there is also a fact that, if all the players are almost close to the mean, they can be said a balaced team but there performance can also be poor. Like suppose, if a team has some players who have a very good performacne score and rest of the players are medium these spacial players can show some really miracle and above teams can do well, where a team's almost every players are close to the mean but there average performacne or cumulative performacne is comparatively poor than them, they may not do well in spite of having a balance team. But this is spacial scenerio, generally balanced and average good performanced team do well.

Distribution:

Actually our data mainly predicts normal distribution. Definitely our data is not properly bale shaped or properly curved or mean in the center. But what we have thought to
Simulate this according to normal distribution and found the following pdf and it's graph

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$\sigma = 7.33$ $\mu = 430.61$



End Credits:

The most important thing, which we learned from our project is how to implement our statistical logic in real life data collection and how to build up statistical thinking on the observation of the collected data. We have collected our information from

1) www.sqwaka.com
2) http://www.footballsquads.co.uk/
3) https://en.wikipedia.org/wiki/Main_Page/
4) www.google/SpreadSheet/
5) www.facebook.com/

Thank you.