



# THE IMPROVING DYNASTY OF SEVILLA

APPLICATION OF DATA SCIENCE IN SPORTS  
ANALYTICS

By: Soumadeep Bhattacharya

## Introduction-

Data is the new oil in almost every industry, from technology to sports. Sports data analytics refers to analyzing statistical data relating to players' performances to identify their weaknesses and strengths. With management and teams becoming increasingly open to using data to achieve a competitive edge, the global sports sector is also evolving rapidly. The sports data analytics market will likely exceed \$4.5 billion by 2025. This indicates that data analytics will significantly impact the sports industry in the coming years.

Another major reason is that there have been occasions where football analytics was ignored and the team had to pay the price. When sports analytics made its way into the sports industry in the beginning, it was followed by cynicism from a lot of GMs, coaches, players and even former players because they did not believe that a computer could give them information about the game that they did not already know from their years of experience being a part of the football industry.

# A Over View On Sevilla's Performance In LA-LIGA Season -2021-22

**Objective** – Sevilla Has finished 4<sup>th</sup> in LA-LIGA season 2021-22. We have analysed the performance of the team in the this season & drawn up some conclusions for the team for improving their performance for the next season .



# TABLE OF CONTENTS

1

Players Statistics Analysis

2

Season's Points Table Analysis

3

Sevilla's All Match Results Analysis In  
The Season

4

Top 4 Teams Performance  
comparison

5

Simple Linear Regression Modeling On  
League Performance vs Squad Investment

6

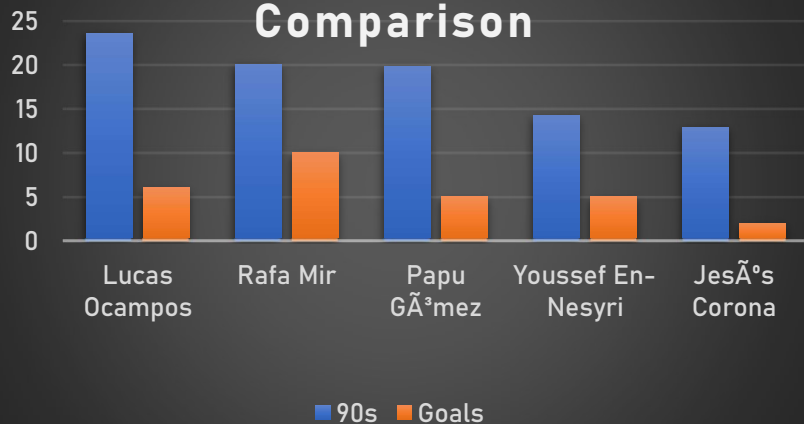
K-Means Clustering On Grouping  
Shot Creators



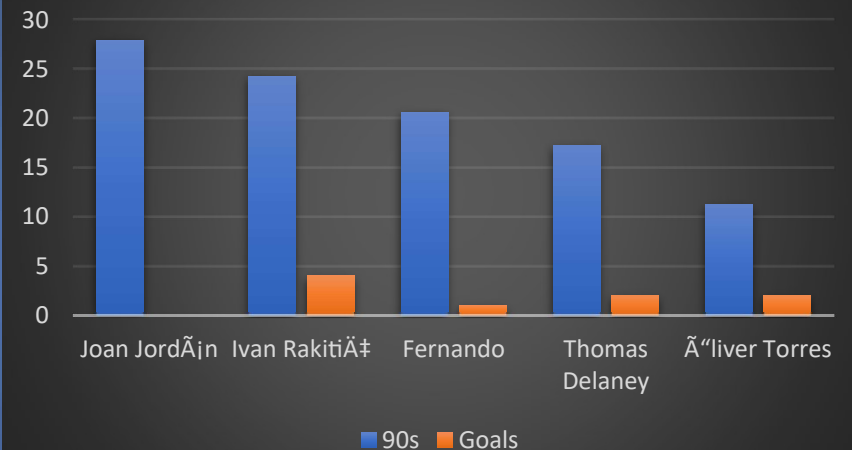
# Player's Performance Statistics Analysis Of The Season 2021-22



## Forward Players Comparison

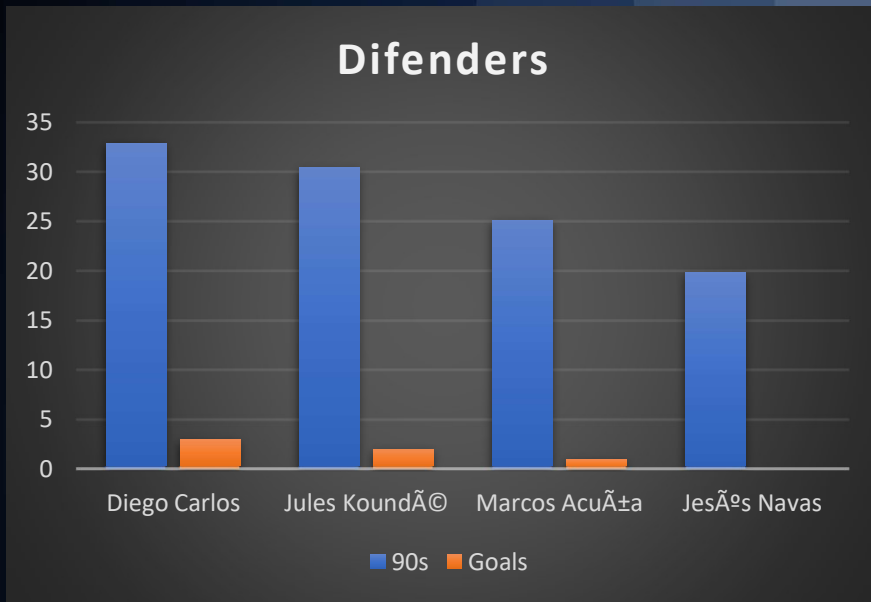


## Midfielders Comparison



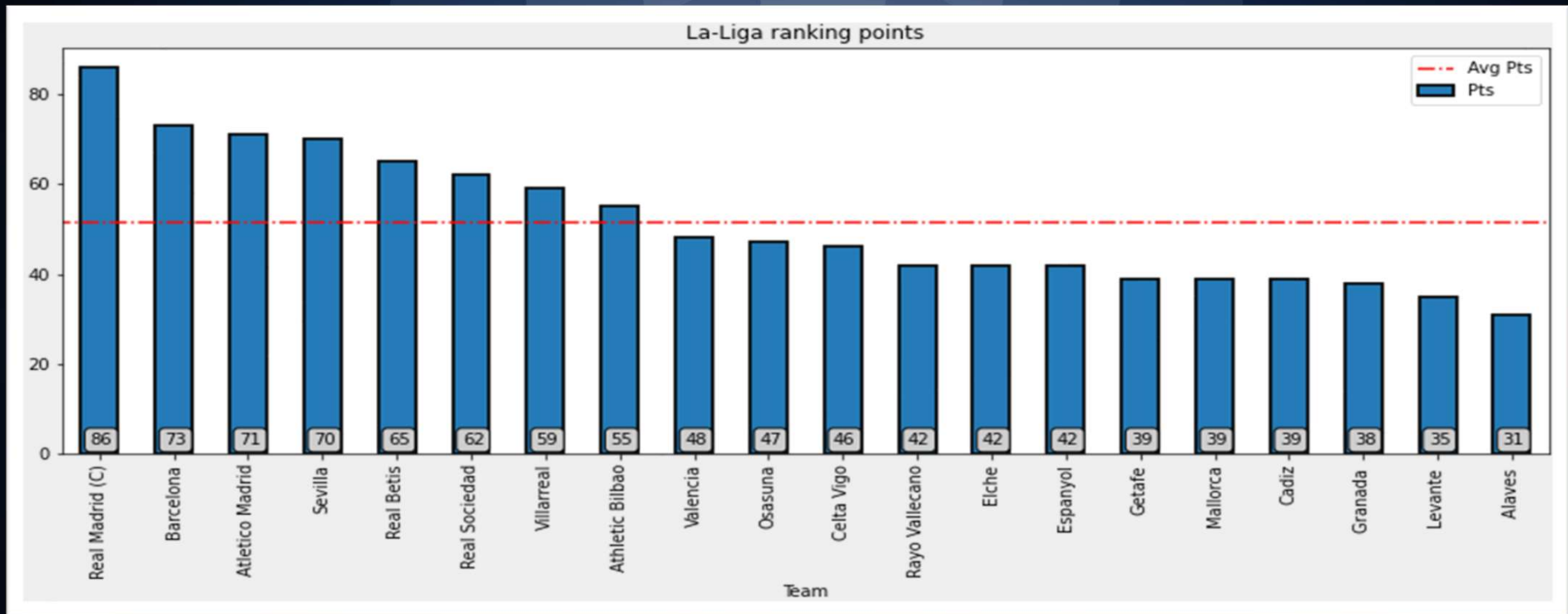
- LUCAS OCAMPOS with avg 23.6 minutes played with 6 goals & RAFA MIR with avg 20.1 mins played with 10 goals are the key players in the forward position.
- Ivan Rakitic with avg 27.9 mins played & Joan Jordan with avg 24.2 mins played & 4 goals are the key players in the mid field.

# Player's Performance Statistics Analysis Of The Season 2021-22



- Diego Carlos having avg playing time 32.9 mins with 3 goals & Jules Koundou having avg playing time 30.4 & 2 goals are the key players in defending position.
- Yassne Bounou is the key player in the Goalkeeping position of Sevilla with avg of 31 mins played..

# LA-LIGA Season 2021-22 Points Table Analysis



- Sevilla Has Scored 20 Points more than the Avg Points .

# Attack & Defence Performance As Per Season Point Table

Ten Best attack

Team

GF

0	Real Madrid (C)	80
1	Barcelona	68
2	Atletico Madrid	65
6	Villarreal	63
4	Real Betis	62
3	Sevilla	53
18	Levante	51
8	Valencia	48
17	Granada	44
10	Celta Vigo	43



Five Best defense

Team

GA

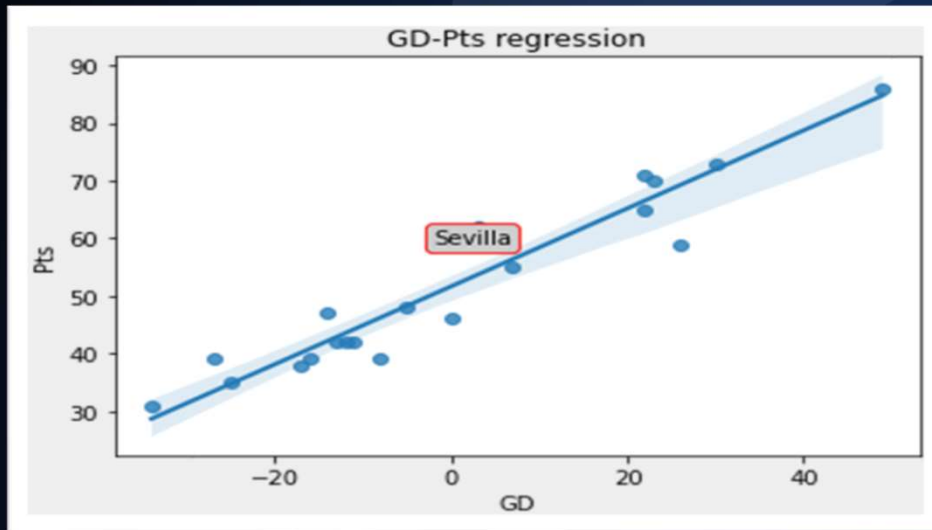
3	Sevilla	30
0	Real Madrid (C)	31
7	Athletic Bilbao	36
5	Real Sociedad	37
6	Villarreal	37

Sevilla has ranked 6th among the teams in Attacking Sills as per their goals for points in La-Liga season 2021-2022.

Sevilla Has The Best Defense Performance among all the teams as per their goals against In La-Liga Season 2021-22.

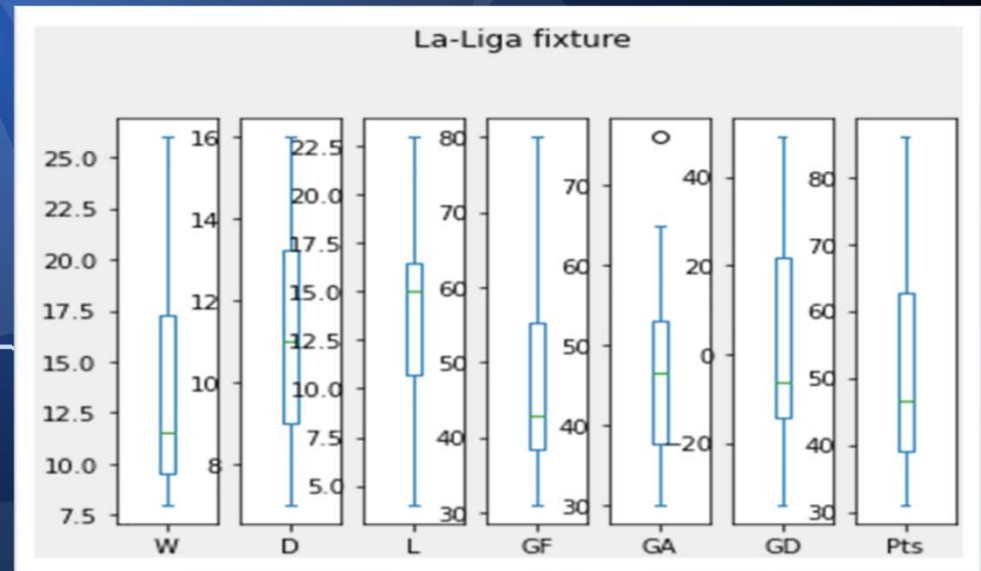


## Team Performance As Per Point Table

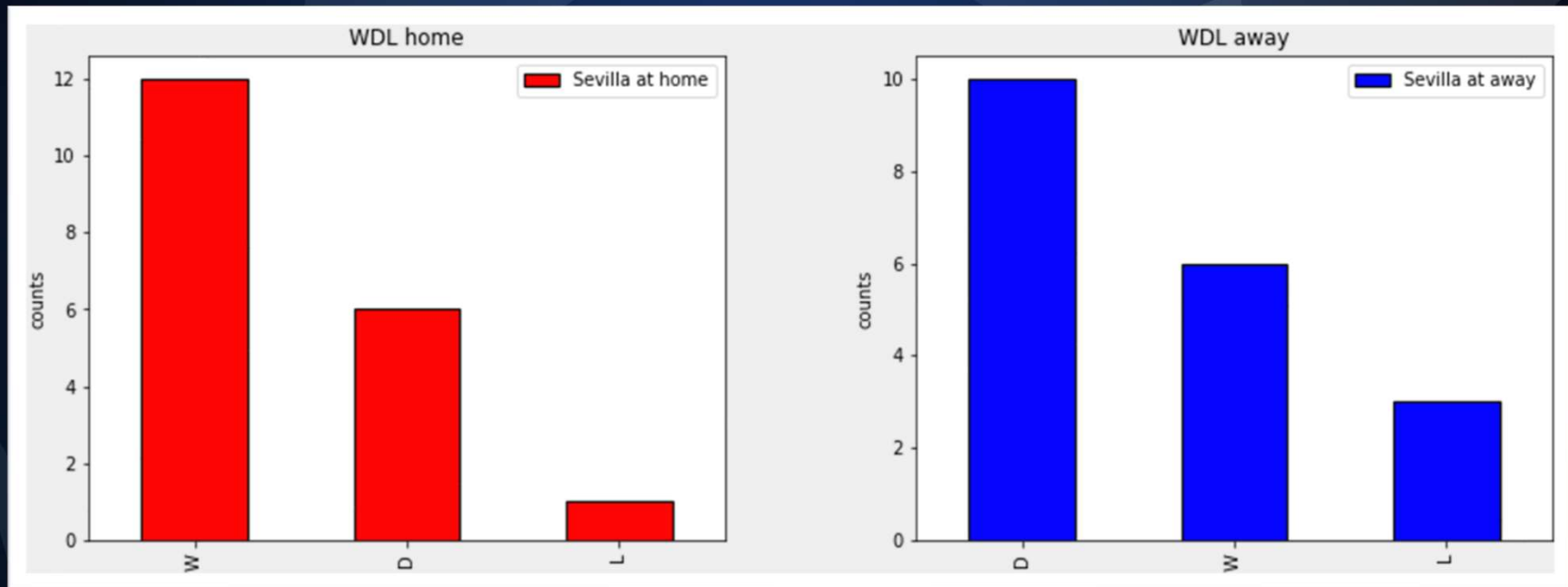


This collinearity plot show that if goal Difference increase then the points also increases. Due to poor Goal difference Sevilla could not take a title.

Seeing the box plot we can assume that Sevilla's performance in La-Liga 2021-2022 is not extraordinary.

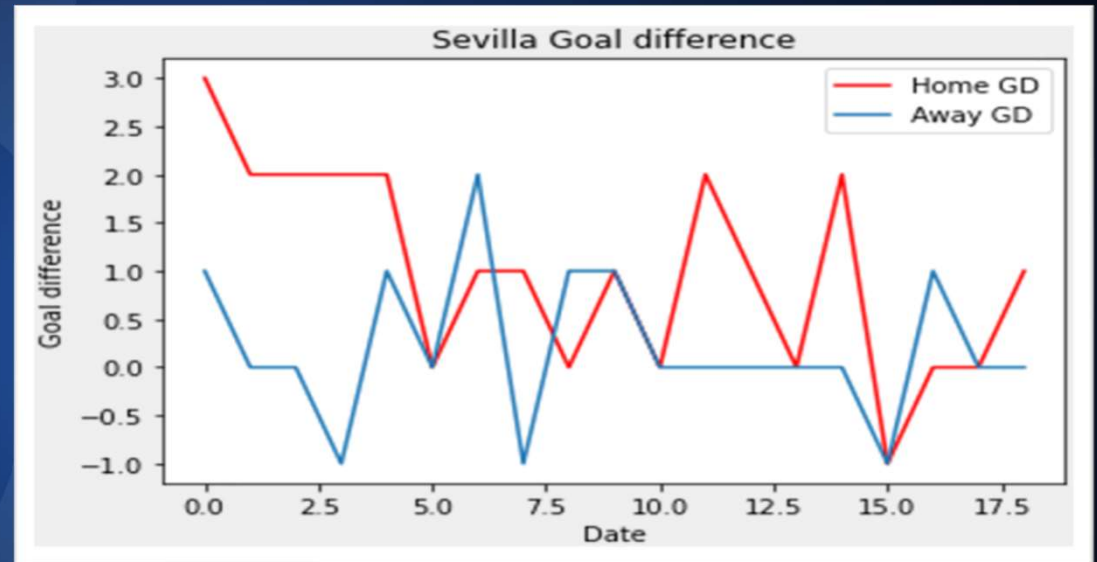


# Sevilla's Win-draw-Loss Comparison Using All Match Result



1. Sevilla has won 12 matches , 6 draws,1 loose in home out of 19 matches .
2. Sevilla has won 10 matches , 6 draws , 3 loose in away out of 19 matches.
3. Sevilla has to improve their game in away match

## As Per All Match Result Sevilla's Goal Scoring & Goal Conceding Comparison



Sevilla's Home Scored avg is - 1.89 , Sevilla's home goal Conceded avg - 0.89 & away scored avg is 0.89 ,conceded average is 0.68 . This chart shows that Sevilla can scored and conceded huge goal .

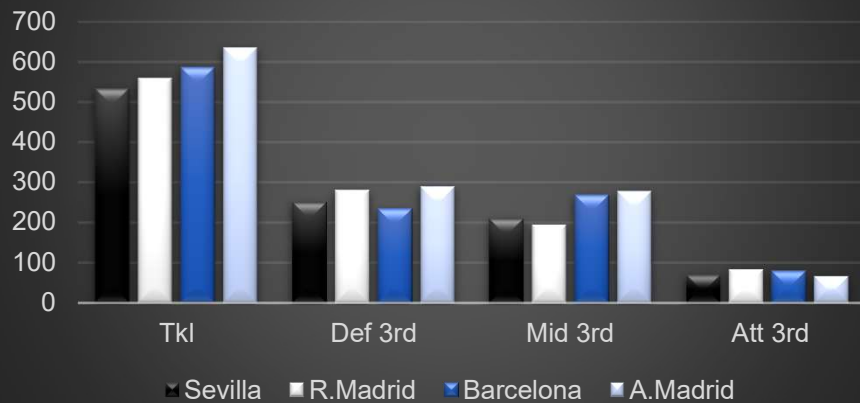
## Conclusions As Per All Match Results & Points Table



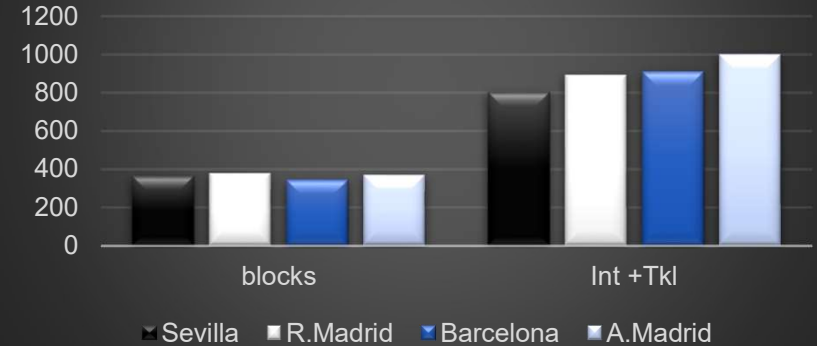
1. Sevilla has ranked 6th among the teams in Attacking Skills as per their goals for points in La-Liga season 2021-2022 .But Sevilla Has The Best Defense Performance among all the teams as per their goals against In La-Liga Season 2021-22.
2. Due to poor Goal difference Sevilla could not take a title.
3. Sevilla has to improve their game in away matches .
4. Sevilla's Home Scored avg is - 1.89 , Sevilla's home goal Conceded avg - 0.89 & away scored avg is 0.89 ,conceded average is 0.68 . Sevilla has scored and conceded huge goals

# Top 4 Teams Defensive Action Data Analysis

## Tackling Data Analysis Of top 4 Teams



## Blocks ,Interception & Tackles Analysis Of 4 Teams



Tkl=number of players Tackled

Def 3rd = Tackles in Defensive 1/3

Mid 3rd = Tackles In Mid field 1/3

Int+Tkl=number of players tackles +number of inter section

Att 3rd= tackles in attacking 1/3

dtkl%= percentage of dribbles tackles

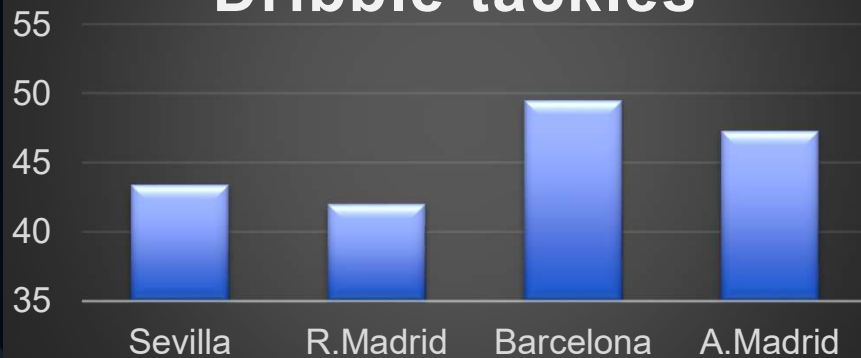
blocks=number of times blocking the ball

Err= mistakes leading to opponent's shot

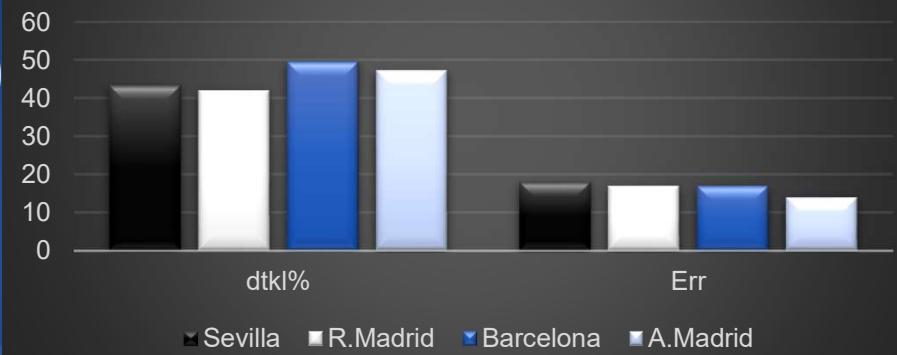
Sevilla's Defensive Performance in successful tackling is average specially the successful tackles in Defensive 1/3 & Attacking 1/3 region . The Defensive performance of team is over all good in Midfield 1/3 region and average number of blocks with respect to the Defensive performance by top 4 teams .

# Top 4 Teams Defensive Action Data Analysis

## Percentage of Dribble tackles



## Dribble tackle percentage & mistakes data analysis



Tkl=number of players Tackled

Def 3rd = Tackles in Defensive 1/3

Mid 3rd = Tackles In Mid field 1/3

Int+Tkl=number of players tackles +number of inter section

Att 3rd= tackles in attacking 1/3

dtkl%= percentage of dribbles tackles

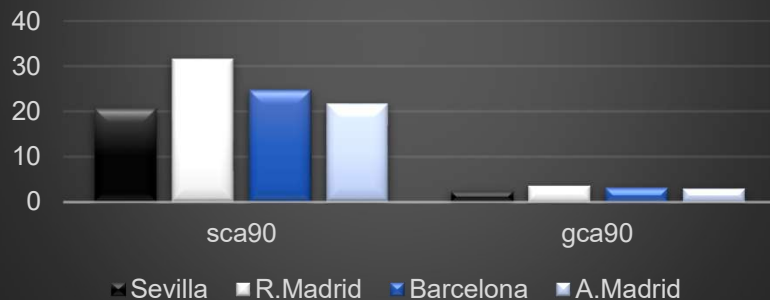
blocks=number of times blocking the ball

Err= mistakes leading to opponent's shot

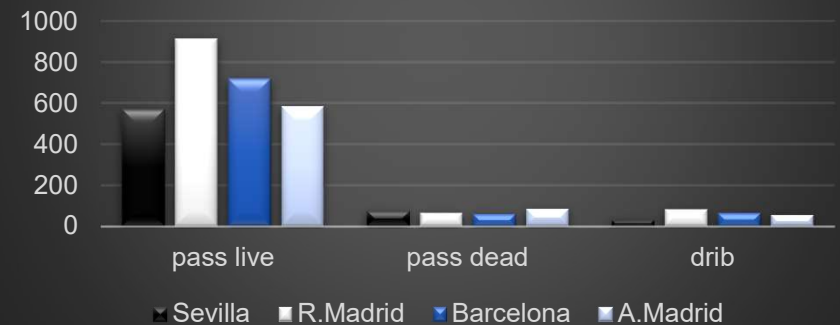
Sevilla's Defensive Performance in successful dribble tackles is average as the number of mistakes in dribble tackling that lead's to shot creation is pretty high as comparison to the other 3 teams.

# Top 4 Team Passes , Goals & Shot Creation Data Analysis

**Per 90 min shot creating action & Goal Creating Action Comparison Of Top 4 Teams**



**Live Ball , Dead Ball & dribble passes that lead to shot attempt comparison of the 4 teams**



sca90 =shot creating actions per 90 min

pass live= completed live ball passes that lead to a shot attempt

pass dead= completed deadball passes that lead to shot attempt

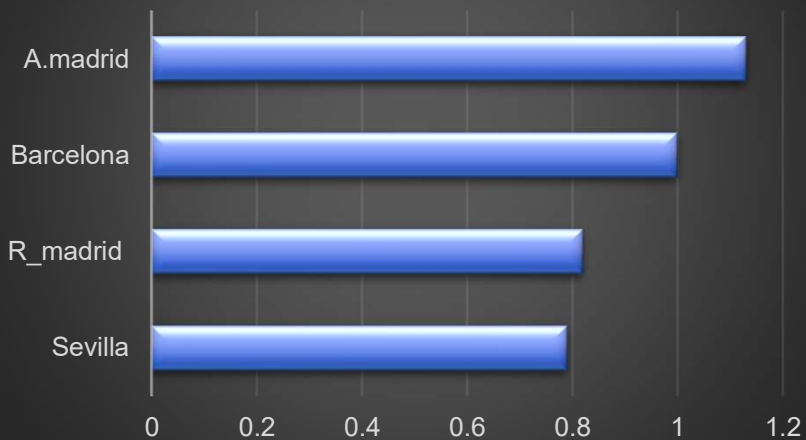
drib= successful dribbles that lead to shot attempts

gca90= goal creating action per 90 min

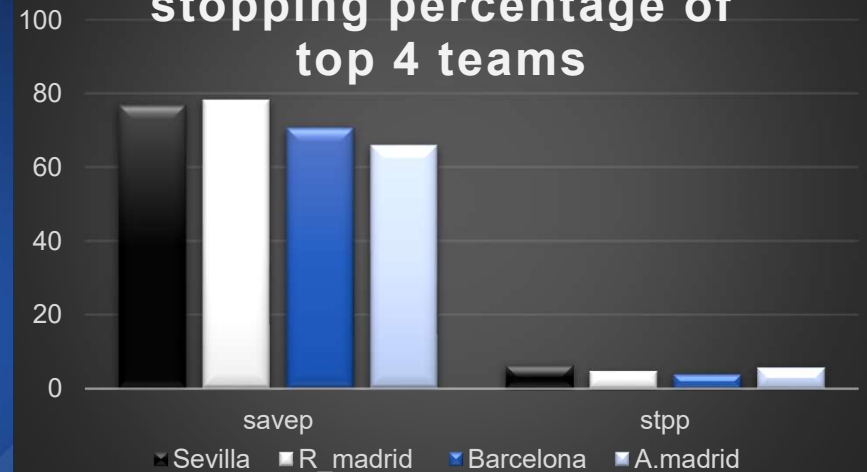
Sevilla's Defensive Performance in success full tackling performance is below average & poorest among the top 4 teams with respect to there per 90 min shot creation , goal creating action , live ball passes ,dead ball passes & dribble passes .

# Top 4 Teams Goal Keeping Data Analysis

## Goal Against Percentage of top 4 teams



## Saving & sweeper stopping percentage of top 4 teams



ga90=Goals Against per 90 min

Stpp=Average crosses into penalty areas that were stopped successfully by goalkeeper

psavep=Saving Percentage

launchp=percentage of passes that were launched.

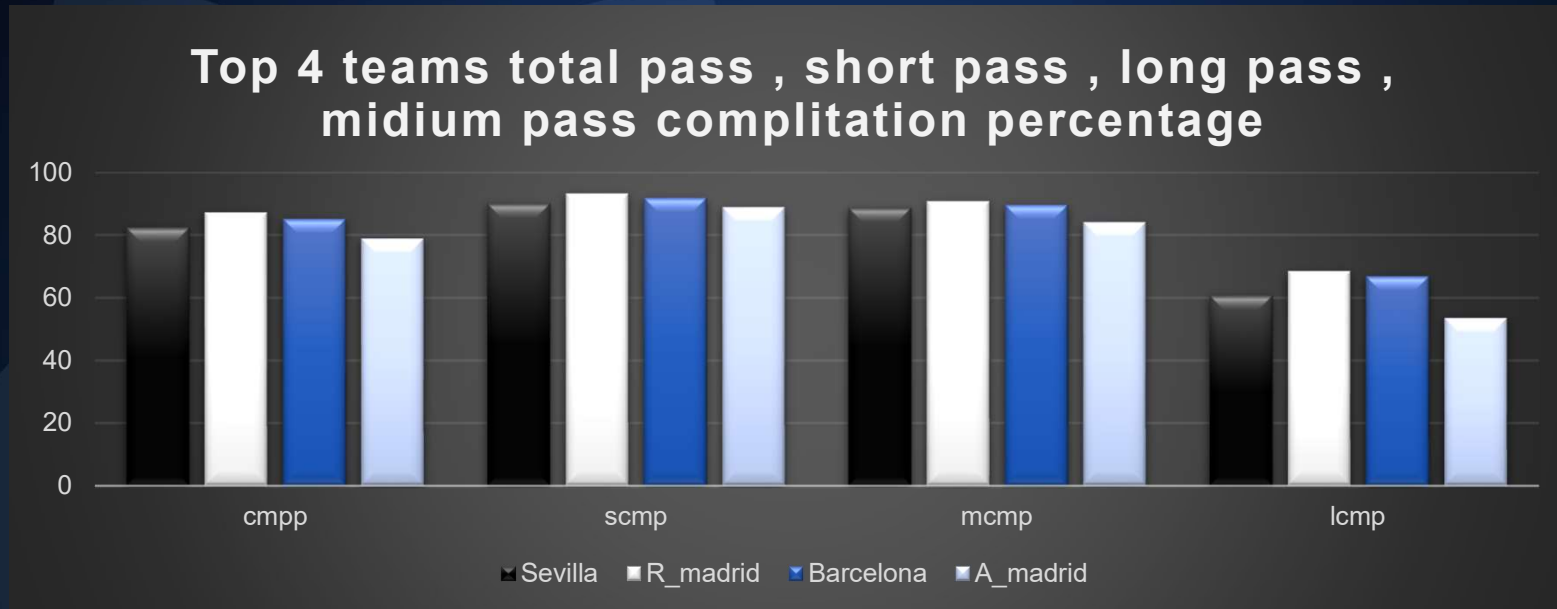
Cmpp= passes completion percentage

avglen=Average length of passes

Sevilla's goal keeping performance is pretty good as compared to the other 3 top team . The goals against percentage is lowest among the 4 teams . Also the goal stopping percentage & sweeper stopping percentage is also high among the 4 teams .



## Top 4 Teams Passes Comparison



cmpp=pass completion percentage	scmp= short pass completion percentage	
pmcmp= medium pass completion percentage	lcmp=long pass completion percentage	

Sevilla's pass giving capabilities as compared to the other 3 teams is below average . Short, long , medium pass completion percentage is also below average as compared to the 4 teams .

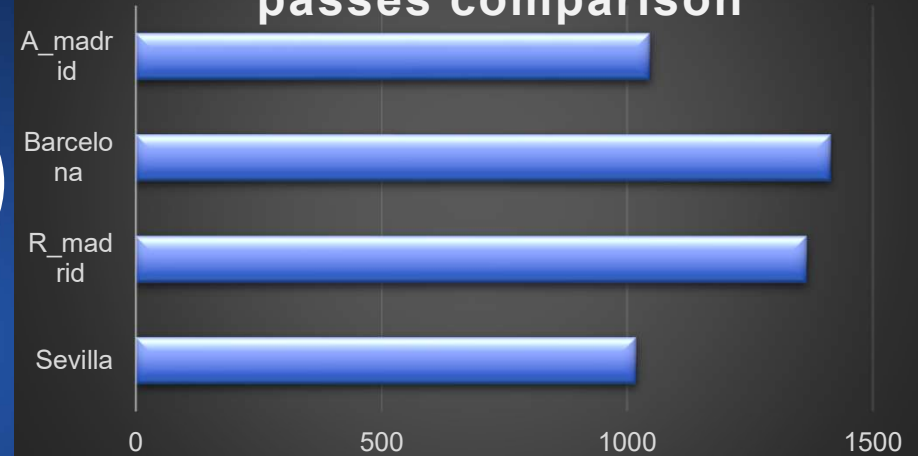
## Top 4 Teams Pass giving comparison



top 4 teams obt, ppa ,  
crspa, passes  
comparison



top 4 teams progressive  
passes comparison



crspa=completed crosses in 18 yards box.

prog =progressive passes ( that move the ball towards opponent's goal bars.)

ppa= completed passes into 18 yards box

obt= completed passes that enters to 1/3 of the pitch closest to the goal

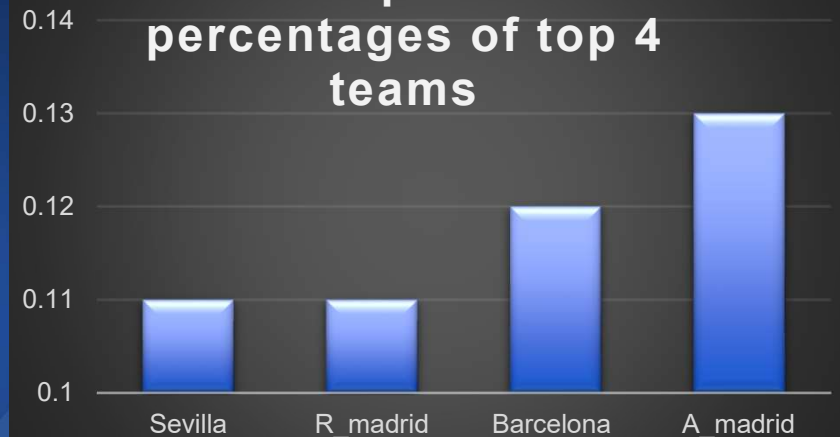
Sevilla's number of completed passes that entered to the 1/3 of the pitch closest to the goal is still average with respect to other 3 teams but team's performance with the other 3 parameters number of complete passes in 18 yards box , number of complete crosses in 18 yards box and number of progressive passes is below average with respect to others

# Top 4 Teams Shoots Comparison

goal scored and total shots  
per 90 mins percentage



Goals per shot  
percentages of top 4  
teams



gls=goals scored

sotp=shots on target percentage

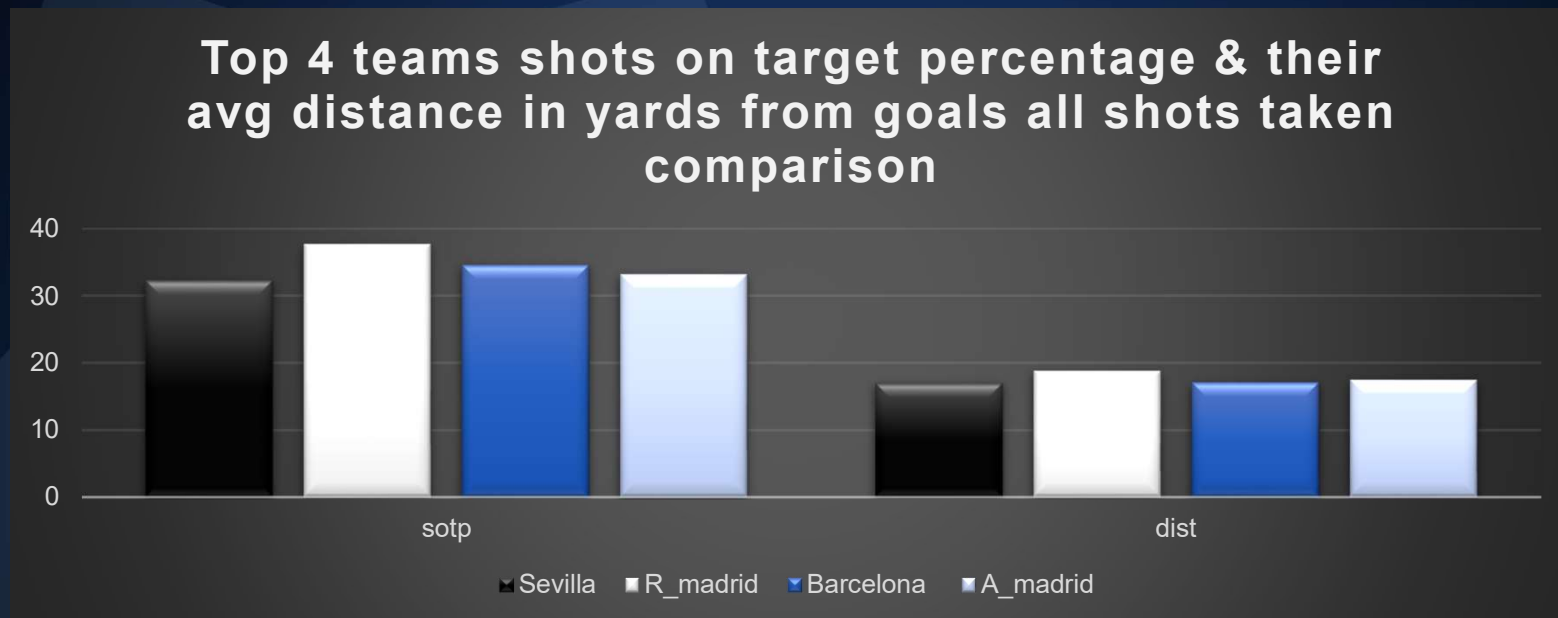
sh90= total shots per 90 mins

gsh= goals per shot

dist = Avg distance in yards from goals all shots taken .

Sevilla's goal scoring , total shots per 90 mins along with their goals per shot percentage is below the average as compared to the other top 3 teams .

## Top 4 Teams Shoots Comparison



gls=goals scored

sotp=shots on target percentage

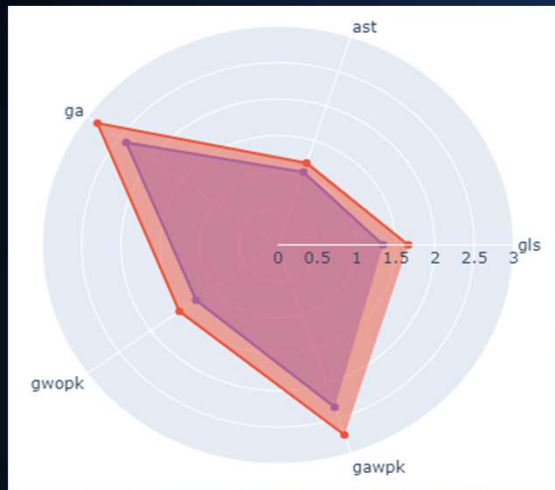
sh90= total shots per 90 mins

gsh= goals per shot

dist = Avg distance in yards from goals all shots taken .

Sevilla's Shot On Target percentage is pretty low on the other hand average distance from goals all shots taken is pretty high as compared to others .

# Overall Performance Comparison Of 3 Teams With Sevilla



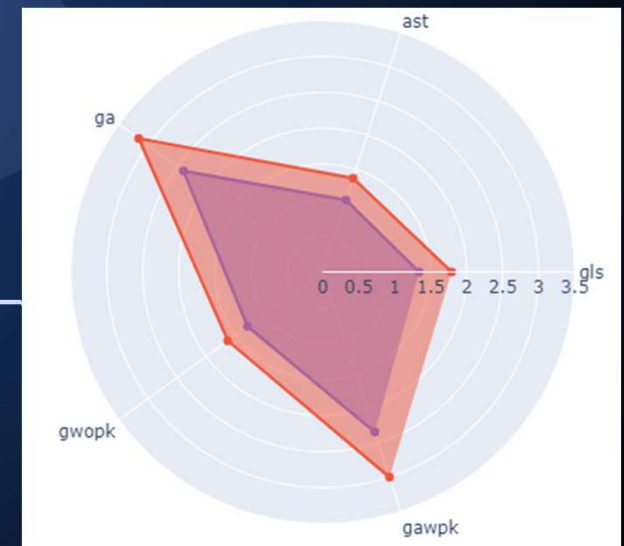
## Sevilla Vs Atletico Madrid

Sevilla's over all performance with Atletico Madrid is quite similar with respect to their performance in assists Per 90 mins .But Atletico Madrid Played far better in other sectors .

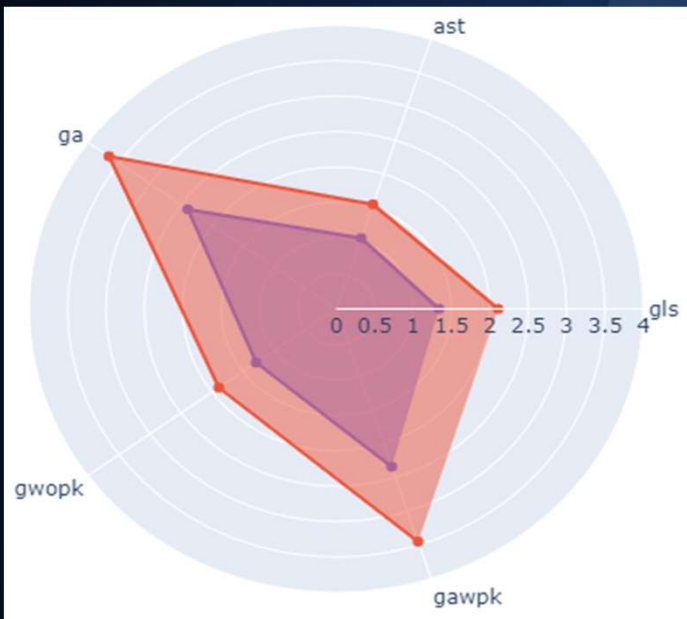
gls= goals scored per 90 min  
ast= assists per 90 min  
gwopk= goals with out penalty kicks per 90 min  
gawpk = assisted goals without penalty kicks

## Sevilla Vs Fc Barcelona

FC Barcelona's performance is way better than Sevilla in all sectors like goals scored per 90 mins ,assists per 90 mins & assisted goals without penalty kicks



# Overall Performance Comparison Of 3 Teams With Sevilla



## Sevilla Vs Real Madrid

Real Madrid's performance is way better than Sevilla in all sectors like goals scored per 90 mins ,assists per 90 mins & assisted goals without penalty kicks

gls= goals scored per 90 min  
ast= assists per 90 min  
gwopk= goals with out penalty kicks per 90 min  
gawpk = assisted goals without penalty kicks

## Overall Performance Report -

Sevilla has to increase their performance in sectors like goals scored per 90 mins ,assists per 90 mins & assisted goals without penalty kicks to play comparatively better against top 3.

## Conclusions Of Overall Performance Comparison Of 3 Teams With Sevilla

1. Sevilla's Defensive Performance in successful tackling is average specially the successful tackles in Defensive 1/3 & Attacking 1/3 region . The Defensive performance of team is over all good in Midfield 1/3 region and average number of blocks with respect to the Defensive performance by top 4 teams .
2. Sevilla's Defensive Performance in successful dribble tackles is average as the number of mistakes in dribble tackling that lead's to shot creation is pretty high as comparison to the other 3 teams.
3. Sevilla's Defensive Performance in success full tackling performance is below average & poorest among the top 4 teams with respect to there per 90 min shot creation , goal creating action , live ball passes ,dead ball passes & dribble passes .
4. Sevilla's goal keeping performance is pretty good as compared to the other 3 top team . The goals against percentage is lowest among the 4 teams . Also the goal stopping percentage & sweeper stopping percentage is also high among the 4 teams .



## Conclusions Of Overall Performance Comparison Of 3 Teams With Sevilla

5. Sevilla's pass giving capabilities as compared to the other 3 teams is below average . Short, long , medium pass completion percentage is also below average as compared to the 4 teams .
6. Sevilla's number of completed passes that entered to the 1/3 of the pitch closest to the goal is still average with respect to other 3 teams but team's performance with the other 3 parameters number of complete passes in 18 yards box , number of complete crosses in 18 yards box and number of progressive passes is below average with respect to others
7. Sevilla's goal scoring , total shots per 90 mins along with their goals per shot percentage is below the average as compared to the other top 3 teams .
8. Sevilla's Shot On Target percentage is pretty low on the other hand average distance from goals all shots taken is pretty high as compared to others .
9. Sevilla has to increase their performance in sectors like goals scored per 90 mins ,assists per 90 mins ,passes & assisted goals without penalty kicks to play comparatively better against top 3.



# Simple Linear Regression Modeling On League Performance vs Squad Investment

We have applied a simple linear regression model to squad value data against performance in the La-Liga. This might help us to see how much a squad might need to invest to avoid relegation, make a data-driven target for our team.

We have collected over 10 years(2011-2021) data of the teams with parameters squad sizes, Avg ages, squad value (in Euros, million) as well as performance data with goal difference, points & position.

Points & goal difference correlate really strongly, as we might expect. Squad value goes up as goal difference and points go up, but as more of a curve than a line. Squad value has increased over time.

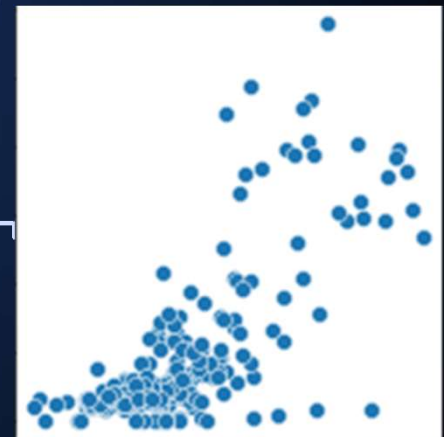


Points vs Goal Difference



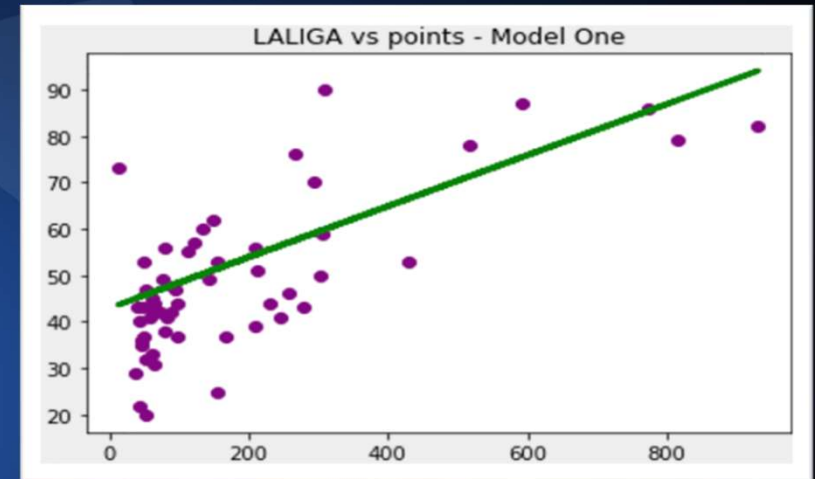
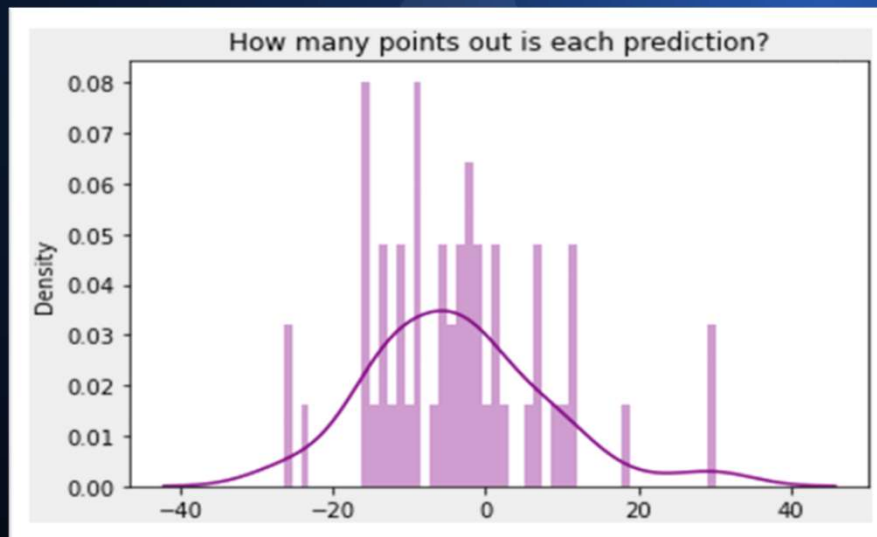
Squad Value Vs Points

Squad Value vs Goal Difference



# Simple Linear Regression Modeling On League Performance vs Squad Investment

Our Regression Coefficient is 0.054 , that means an extra million gets you 0.05 points .



Our Mean Absolute Error is 9.596 in the first model.

## Simple Linear Regression Modeling On League Performance vs Squad Investment

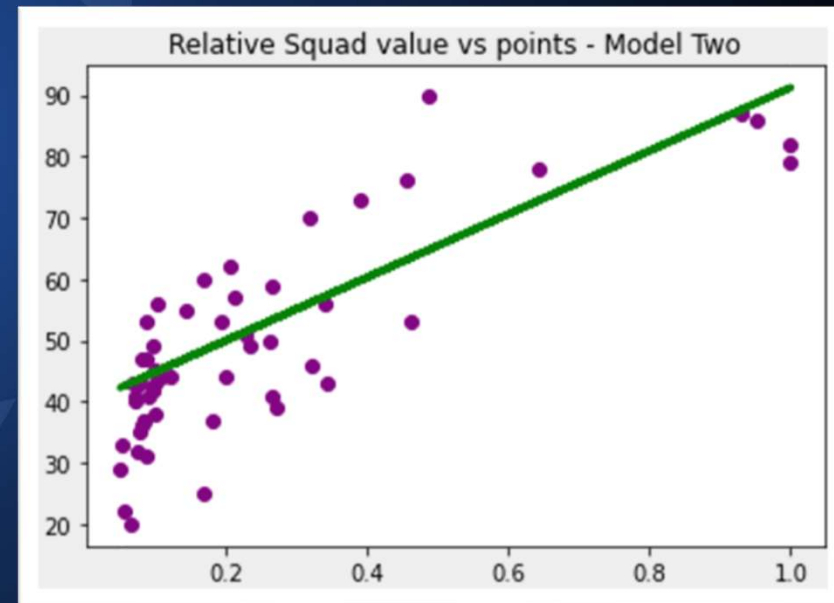
When we took an exploratory look at the data, we found that team values had increased over seasons. As such, comparing a 100m squad in 2011 to a 100m squad in 2021 probably isn't fair.

To counter this, we are going to create a new 'Relative Value' column. This will take each team in a season, and divide it by the highest value in that league. These values will be between 0 & 1 and give a better impression of comparative buying power.

The model predicts just over 5 points. This seems to make sense, as the difference between top and bottom would often range around 50 or so points.

So for every 10% that we are off of the most expensive team, our model suggests that you should expect to drop 5.1 points.

Our Mean Absolute Error is 7.86 . That means our model has improved around 18% over all .



## Conclusions On Simple Linear Regression Modeling On League Performance vs Squad Investment

1. Points & goal difference correlate really strong.
2. Squad value goes up as goal difference and points go up, but as more of a curve than a line. Squad value has increased over time.
3. In the first model our mean absolute error is 9.596 & Regression Coefficient is 0.054 , that means an extra million gets you 0.05 points .
4. After Improving the model for every 10% that we are off of the most expensive team, our model suggests that you should expect to drop 5.1 points.  
Here our mean absolute error is 7.86 . That means our model has improved around 18% over all .



## K-Means Clustering On Grouping Shot Creators



We will look at k-means clustering. We will use the algorithm to put players into different groups based on their shot creating actions.

We have taken out all players data playing in 2021-22 LA-LIGA like Position ,age ,Live Passes , Dead Passes ,Dribbles, Shots ,Foul , minutes play divided by 90, shot creating action , shots creating actions per 90 min.

# K-Means Clustering On Grouping Shot Creators

One more thing we have done is that the effect of playing for a stronger team. As a broad assumption, we can expect players in better teams to create more shots, and players in worse teams to produce fewer. This might produce results that group players based on their production levels, not the styles of their productions. We have added new columns to look at the percentages for each action type. We created a sum column, then divided each column by the sum.

	Player	Nation	Pos	Squad	Age	Born	90s	SCA	SCA90	PassLive	PassDead	Drib	Sh	Fld	Sum	SCA	Pass	SCA	Ratio
0	Luis Abram	pe PER	DF	Granada	25.0	1996.0	6.2	3	0.48	1	0	0	2	0	3				0.333333
1	Marcos Acuna	ar ARG	DF	Sevilla	29.0	1991.0	25.1	75	2.99	48	19	2	3	3	75				0.640000
2	Martin Agirregabiria	es ESP	DF	AlavEs	25.0	1996.0	18.0	33	1.84	25	3	2	0	2	32				0.781250
3	Julen Agirrezabala	es ESP	GK	Athletic Club	20.0	2000.0	4.0	1	0.25	1	0	0	0	0	1				1.000000
4	Sergio Aguiro	ar ARG	FW	Barcelona	33.0	1988.0	1.7	3	1.78	3	0	0	0	0	3				1.000000

## Conclusions After Applying K Means Clustering -

1. Ranking Of First Set Piece Takers In The Sevilla In Season 2021 -2022 – Rafa Mir Ranked 28th this total List.
2. Ranking of Sevilla's player as per Number of dead balls, passes dominate but contributions from other types In the Season 2021-2022 – Ivan Rakitic Ranked 20th in the total list.
3. Ranking Of Sevilla's players as passers in the most part in the season 2021-22- Jesus Corona Ranked 13<sup>th</sup> in the total list.
4. Ranking Of Sevilla's Players who are high on mostly passes but high on dribbles and shots to – Papu Gomez ranked 12th in the total list .
5. Ranking Of Sevilla's Players Who are high on passes but most evenly spread out side of that – Thomas Delaney Ranked 15<sup>th</sup> in the list .





# References

- Fcpython.com - <https://fcpython.com/>
- Wikipedia- [https://en.wikipedia.org/wiki/Main\\_Page](https://en.wikipedia.org/wiki/Main_Page)
- LA-Liga Official Site- <https://www.laliga.com/en-GB>
- ESPN Sports - <https://www.espn.in/>
- Fbref- Football History & Statistics - <https://fbref.com/en/>
- Analytics Vidya - <https://www.analyticsvidhya.com/myfeed>
- NCBI- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992018/>
- ScienceDirect - <https://www.sciencedirect.com/topics/social-sciences/linear-regression-model>
- Towards data science - <https://towardsdatascience.com/statistics-supporting-linear-models-bfc24fb9781f>
- Web pdx- [https://web.pdx.edu/~newsomj/cdaclass/ho\\_glm.pdf](https://web.pdx.edu/~newsomj/cdaclass/ho_glm.pdf)
- IEEE XPLORE - <https://ieeexplore.ieee.org/document/1053964>
- TAYLOR & FRANCIS -  
<https://www.tandfonline.com/doi/abs/10.1080/00949655.2021.1882458>
- OXFORD ACADEMICS - <https://academic.oup.com/comjnl/article-abstract/55/9/1058/400173?redirectedFrom=fulltext>
- AIP Publishing- <https://aip.scitation.org/doi/10.1063/1.1664978>



**Thank  
You**

