Data Visualization CA02- FIFA 2018

Prof. Basel Magableh

MSc. Data Analytics, Group A

Group Members: Damanpreet Kahlon(10505907)

Ajit Shukla(10394108)

Reema Mascarenhas(10402498)

Ashin Basheer(10503646)

Dataset:

Source: Kaggle

Link: **https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset**

Data Description: The data is scraped from the website https://sofifa.com by extracting the Player personal data and Player Ids and then the playing and style statistics (Shrivastava, 2017).

This dataset consists of 18k records for over 75 attributes. It has more than 50 attributes describing various player performance scores and position details. Along with these performance indicators we also have attributes like player Name, Age, Nationality, Club etc. We will explain the attributes in detail with their visual analysis.

	Name	Age	Photo	Nationalit	Flag	Overall	Potential	Club	Club Logo	Value	Wage	Special	Accelerati	Aggressior	Agility	Balance	Ball contrc	Composur	Crossing	Curve	Dribbling	Finishin
0	Cristiano F	32	https://cd	Portugal	https://cd	94	94	Real Madr	https://cd	â,¬95.5M	â,¬565K	2228	89	63	89	63	93	95	85	81	91	
1	L. Messi	30	https://cd	Argentina	https://cd	93	93	FC Barcelc	https://cd	â,¬105M	â,¬565K	2154	92	48	90	95	95	96	77	89	97	
2	Neymar	25	https://cd	Brazil	https://cd	92	94	Paris Saint	https://cd	â,¬123M	â,¬280K	2100	94	56	96	82	95	92	75	81	96	

Dribbling	Finishing	Free kick a	GK diving	GK handlir	GK kicking	GK positio	GK reflexe	Heading a	Interceptic	Jumping	Long passi	Long shot:	Marking	Penalties	Positionin	Reactions	Short pass	Shot powe	Sliding tac	Sprint spe	Stamina	Standin
91	94	76	7	11	15	14	11	88	29	95	77	92	22	85	95	96	83	94	23	91	92	
97	95	90	6	11	15	14	8	71	22	68	87	88	13	74	93	95	88	85	26	87	73	
96	89	84	9	9	15	15	11	62	36	61	75	77	21	81	90	88	81	80	33	90	78	

Standing t	Strength	Vision	Volleys	CAM	CB	CDM	CF	CM	ID	LAM	LB	LCB	LCM	LDM	LF	LM	LS	LW	LWB	Preferred	RAM	RB
31	80	85	88	89	53	62	91	82	20801	89	61	53	82	62	91	89	92	91	66	ST LW	89	
28	59	90	85	92	45	59	92	84	158023	92	57	45	84	59	92	90	88	91	62	RW	92	
24	53	80	83	88	46	59	88	79	190871	88	59	46	79	59	88	87	84	89	64	LW	88	

RB	RCB	RCM	RDM	RF	RM	RS	RW	RWB	ST
61	53	82	62	91	89	92	91	66	92
57	45	84	59	92	90	88	91	62	88

This dataset had a lot of exploratory potential and so it was fit for us to select this for our assignment.

Tools and Platform used: Power BI, Tableau, Python, R.

Deployment

GitHub Repo: https://github.com/fifa-players/Fifa

Power BI Dashboards:

(https://app.powerbi.com/view?r=eyJrIjoiMTA0MDNkMmItMDFhMi00OWM3LTk1YzctYWYyNjc1NmFjYzEwIiwidCI6IjVkMGFhNmVhLTY2MjAtNDg2My05ZTIxLTllY2IxNDAyMjJiYyIsImMiOjh9)

Tableau Dashboard:

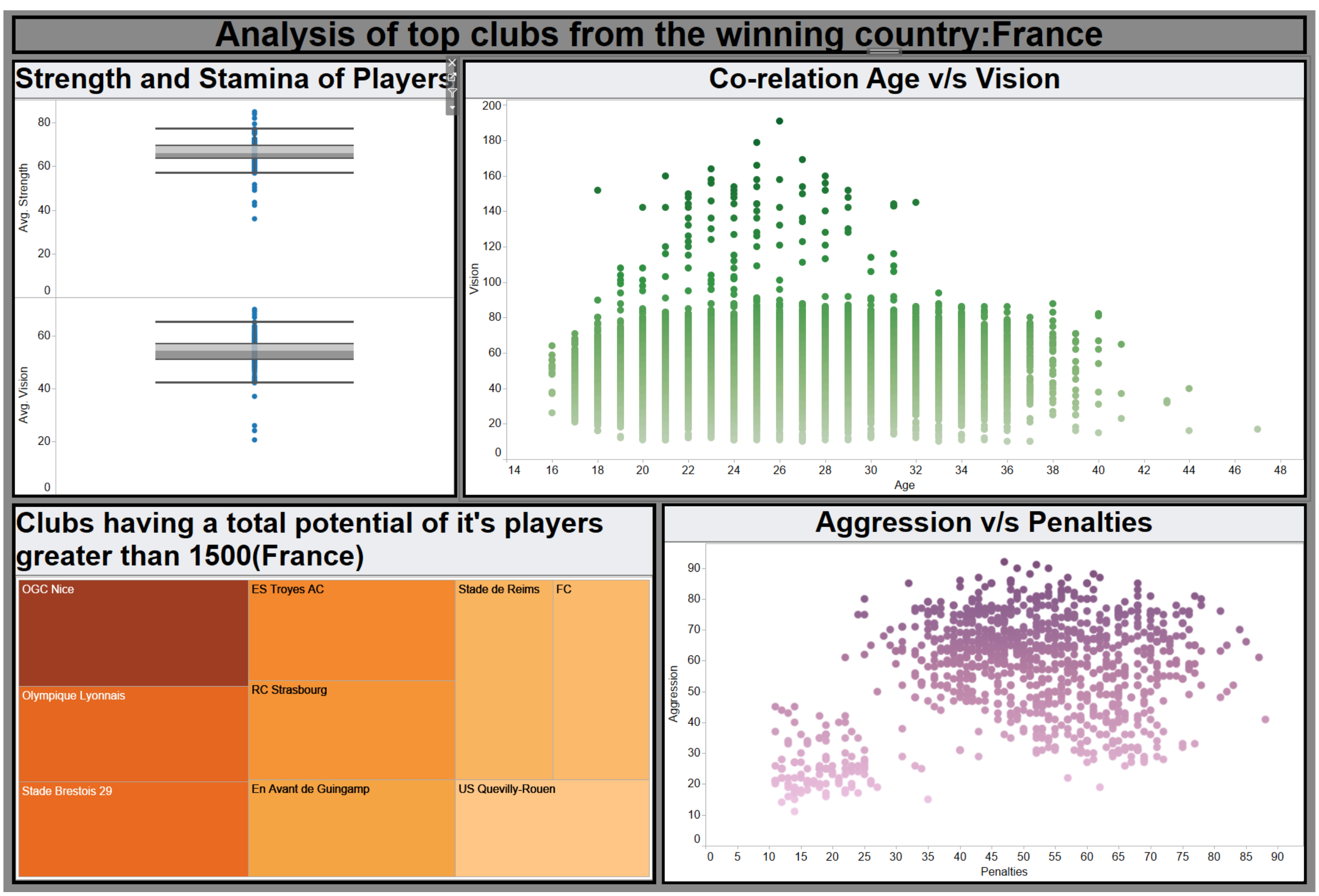
(https://public.tableau.com/profile/ajit.kumar.shukla#!/vizhome/DV_CA02/ClubwiseAnalysis?publish=yes)

Analysis of top clubs from the winning country: France

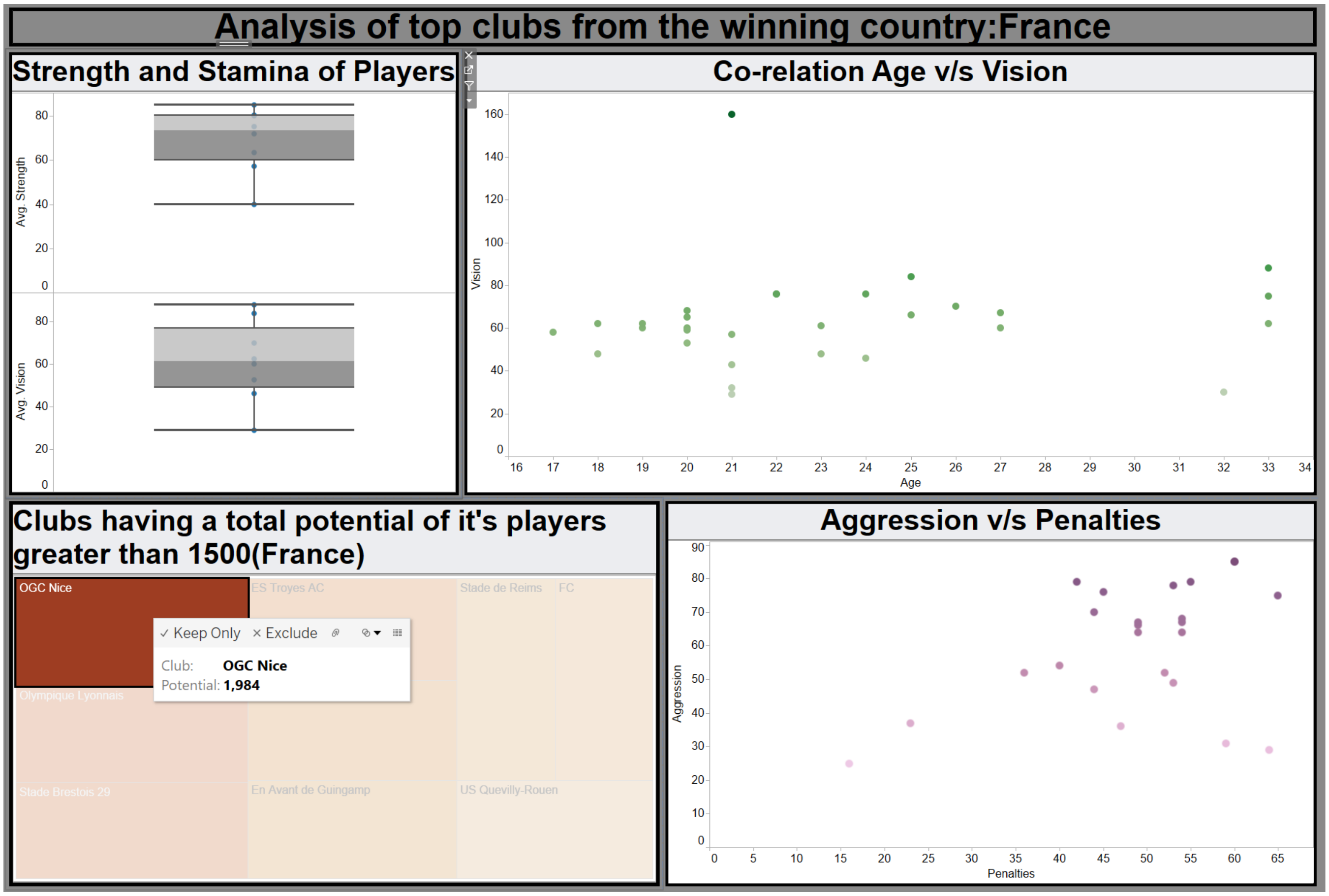
France is the winner of FIFA 2018. We identified clubs having a collective player potential greater than 1500 points from France to understand the different performance measures of their players.

In order to do so, we have created an interactive dashboard on Tableau. This dashboard displays various player attributes.

The list charton the bottom left shows the identified clubs. We can also examine the range of strength and stamina of these players. The co-relation between their Age and Vision powe. We can also analyze how aggressive the players of these teams are along with the number of penalties.



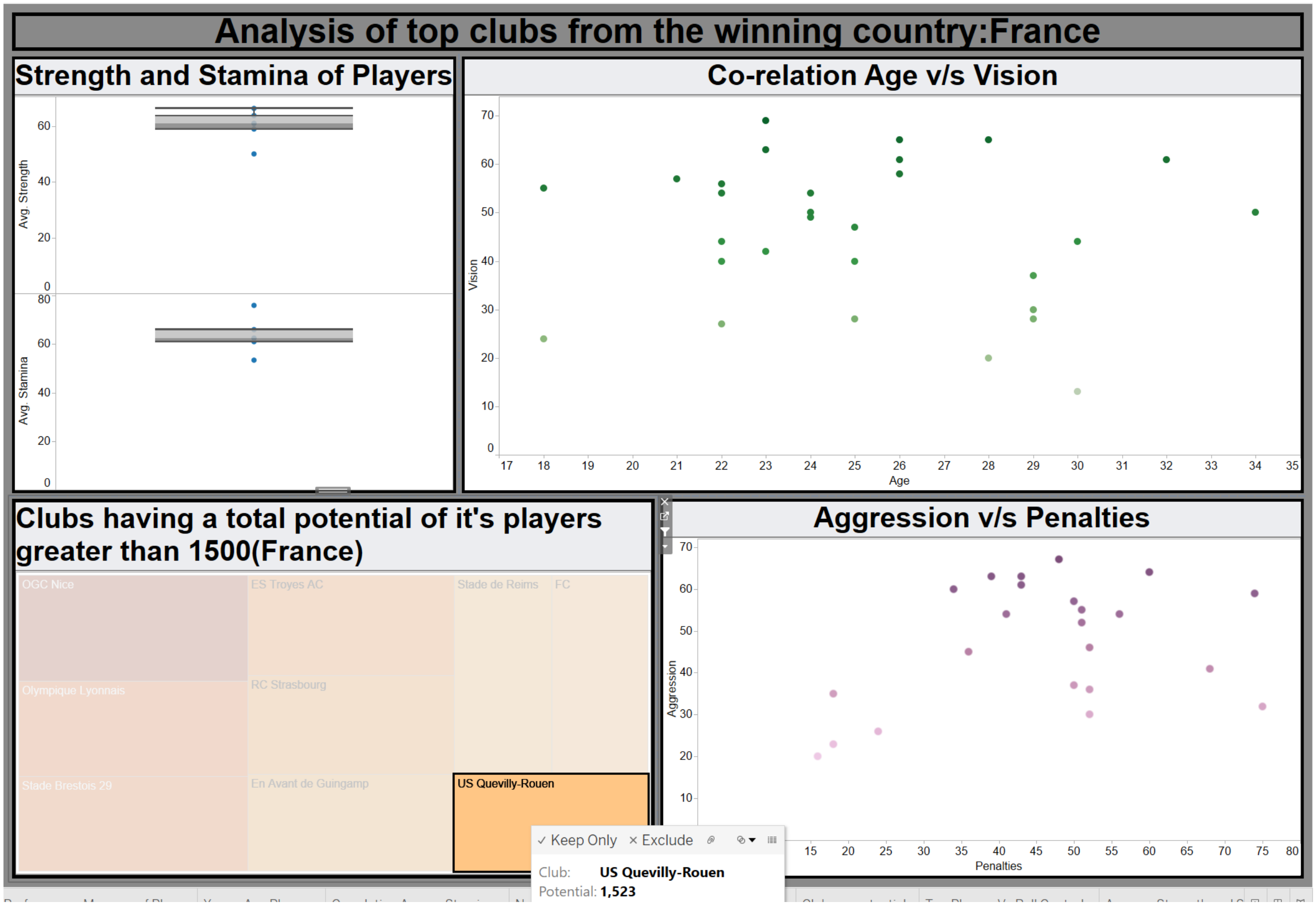
Lets filter for the best club-OGC:Nice



The range for average **strength** of players is between 60-80 with a median of around 74. The range is between 50-70 on average for the **stamina** of OGC: Nice club's players.

The **vision** of the players, although expected to be proportional to **age** is somewhat an indivisual measure. Even at the same ages players have different vision measures and the vision does not decrease in players who are older. Most of the players have more than 60 **penalties** and many may be classified as pretty **aggressive**.

Lets look at the comparatively worst club from our list US Quevilly-Rouen.

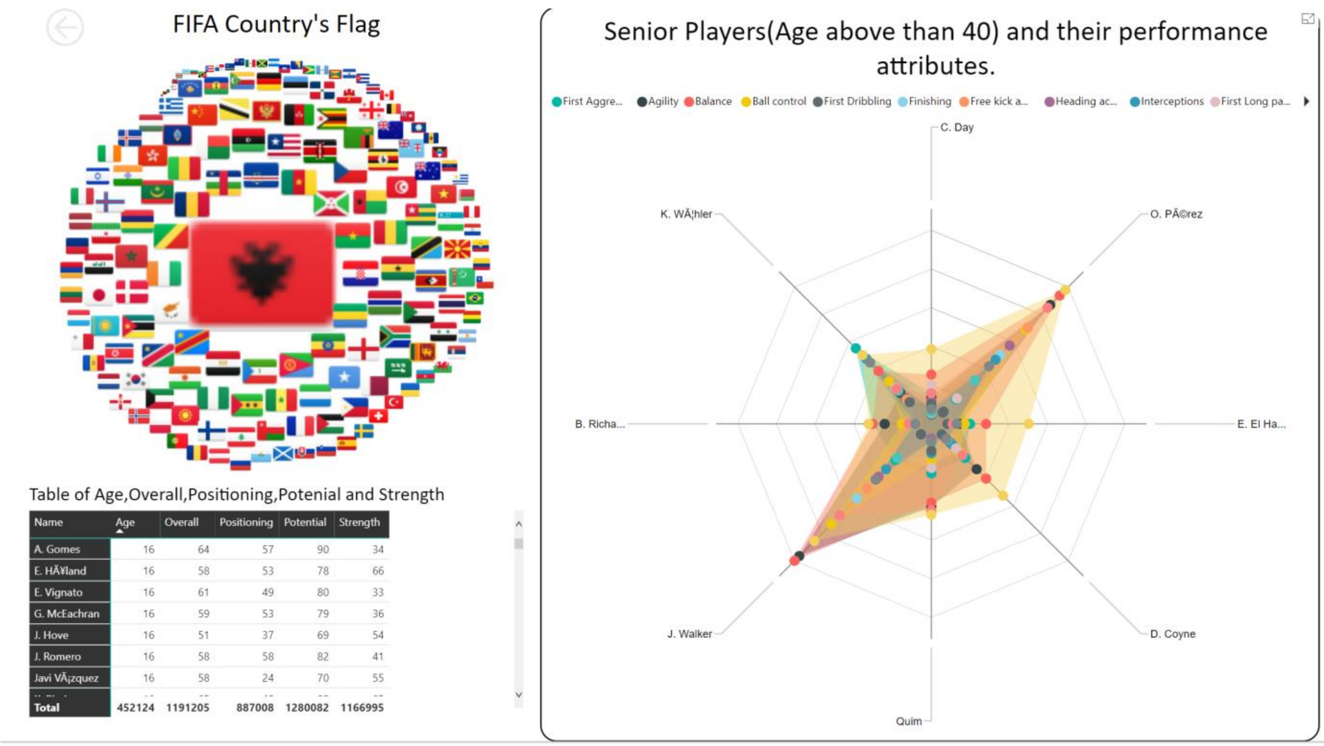


Although the relation between **vision** and **age** is somewhat similar to OGC:Nice, players of US Quevilly-Rouen have a much better vision among players of all ages as compared to the former. The strength and stamina ranges are much lower while aggregation and penalties are more or less same.

Player Age Based Analysis

It is known that football players start their careers at early ages like 16 and retire by 35. We wanted to compare the different attributes of players based on their age. In doing this we also wanted to understand the performance measures of the rare cases of players aged more than 40.

In order to achieve this we have selected Power BI to create an interactive dashboard.



Name	Age	Overall	Positioning	Potential	Strength
A. Gomes	16	64	57	90	34
E. HÃ¥land	16	58	53	78	66
E. Vignato	16	61	49	80	33
G. McEachran	16	59	53	79	36
J. Hove	16	51	37	69	54
J. Romero	16	58	58	82	41
Javi VÃ¡zquez	16	58	24	70	55
Total	452124	1191205	887008	1280082	1166995

The table provides a list of Players along with their Age, Overall performance measure, their Positioning , Potential marked by their performance and Strength. On selecting an individual row/Player from the table, the flag of the players nationality is highlighted as shown below:

FIFA Country's Flag

Table of Age,Overall,Positioning,Potenial and Strength

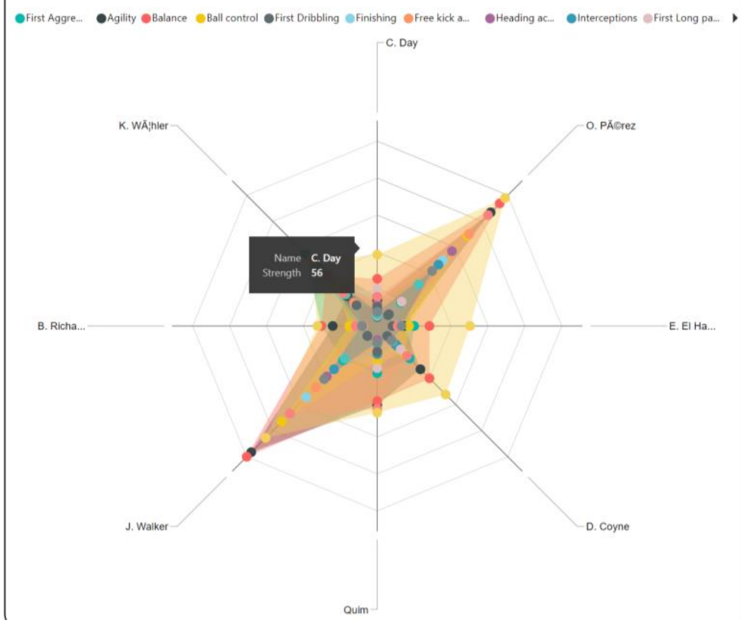
Name	Age	Overall	Positioning	Potential	Strength
A. Gomes	16	64	57	90	34
E. HÃ¥land	16	58	53	78	66
E. Vignato	16	61	49	80	33
G. McEachran	16	59	53	79	36
J. Hove	16	51	37	69	54
J. Romero	16	58	58	82	41
Javi VÃ¡zquez	16	58	24	70	55
Total	452124	1191205	887008	1280082	1166995

FIFA Country's Flag

Table of Age,Overall,Positioning,Potenial and Strength

Name	Age	Overall	Positioning	Potential	Strength
A. Gomes	16	64	57	90	34
E. HÃ¥land	16	58	53	78	66
E. Vignato	16	61	49	80	33
G. McEachran	16	59	53	79	36
J. Hove	16	51	37	69	54
J. Romero	16	58	58	82	41
Javi VÃ¡zquez	16	58	24	70	55
Total	**452124**	**1191205**	**887008**	**1280082**	**1166995**

Senior Players(Age above than 40) and their performance attributes.



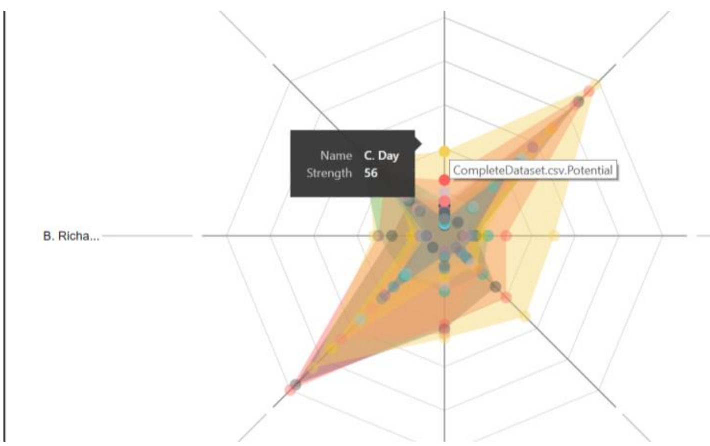
Player C. Day although aged over 40 has a greater strength compared to many of the younger players. Hence it can be deduced that age is not a limiting criteria for performance and strength.

Conversely, on selecting a particular branch from the web diagram, the other attributes are highlighted on the table.

Link: https://app.powerbi.com/view?r=eyJrIjoiMTA0MDNkMmItMDFhMi00OWM3LTk1YzctYWYyNjc1NmFjYzEwIiwidCI6IjVkMGFhNmVhLTY2MjAtNDg2My05ZTIxLTllY2IxNDAyMjJiYyIsImMiOjh9

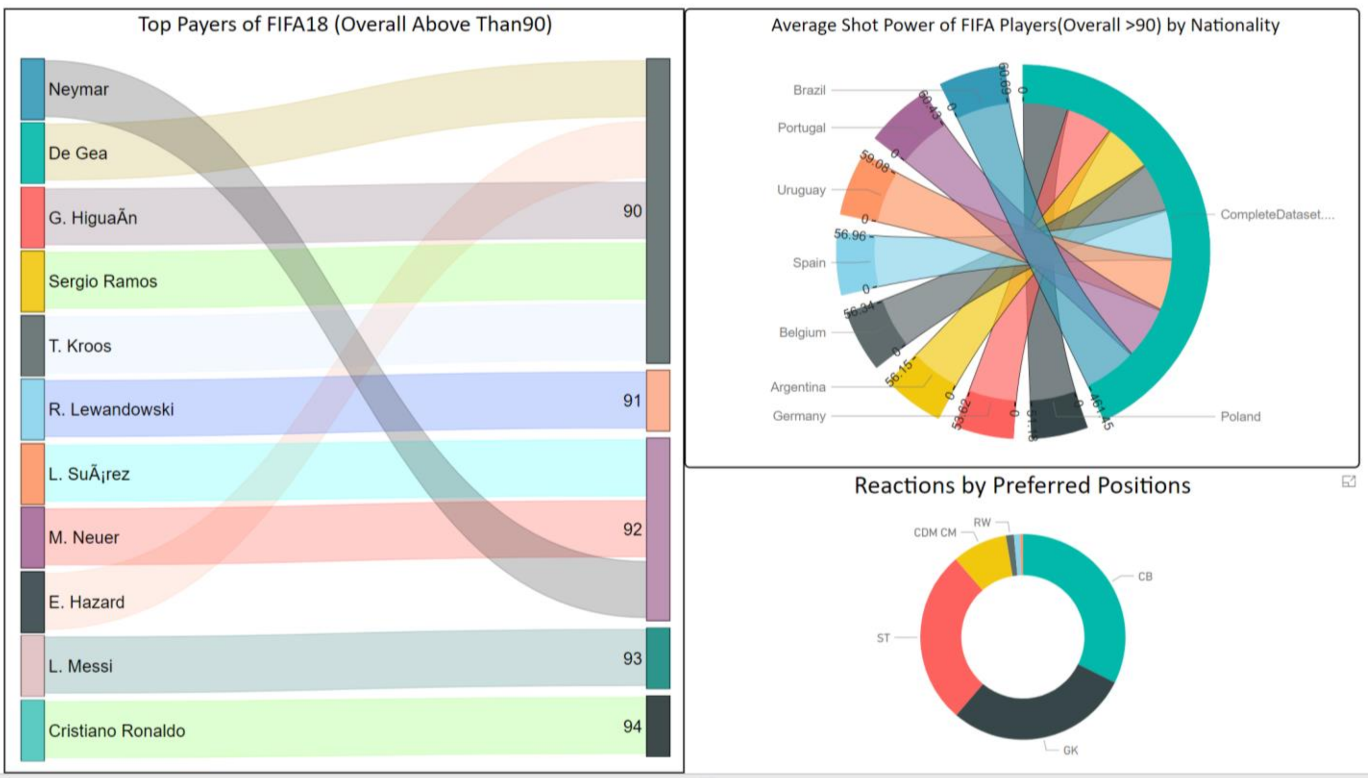
Table of Age,Overall,Positioning,Potenial and Strength

Name	Age	Overall	Positioning	Potential	Strength
C. Day	41	57	7	57	56
Total	**41**	**57**	**7**	**57**	**56**

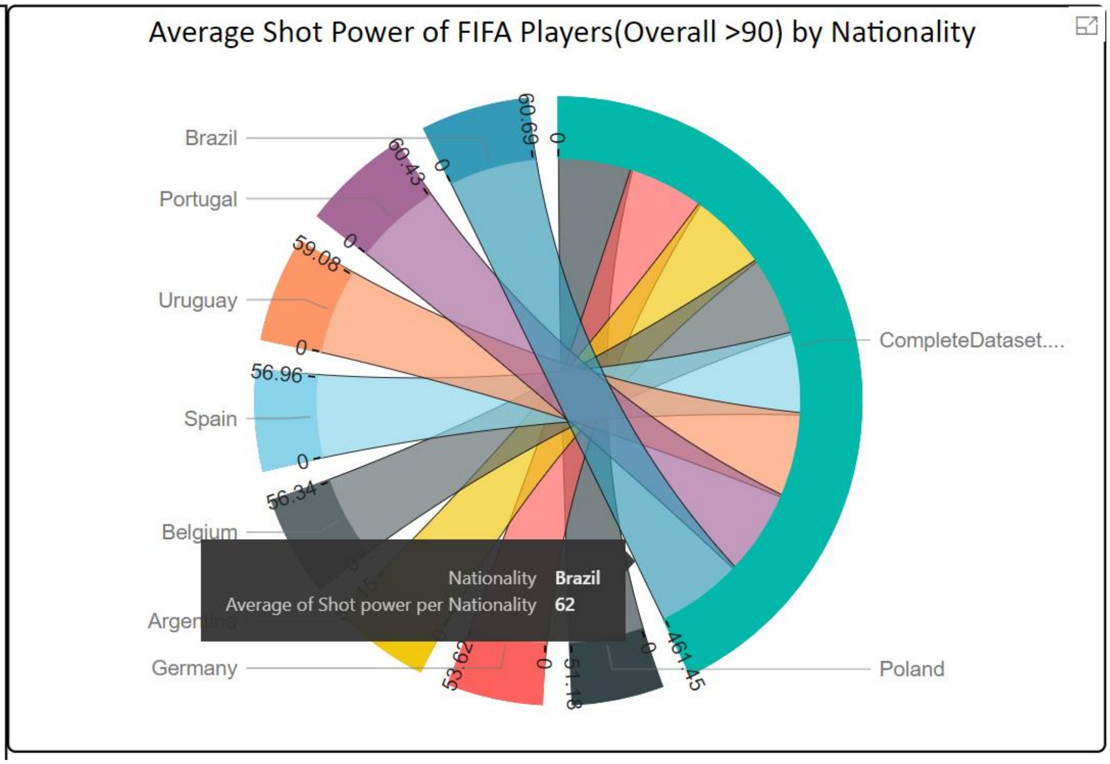


Player based analysis

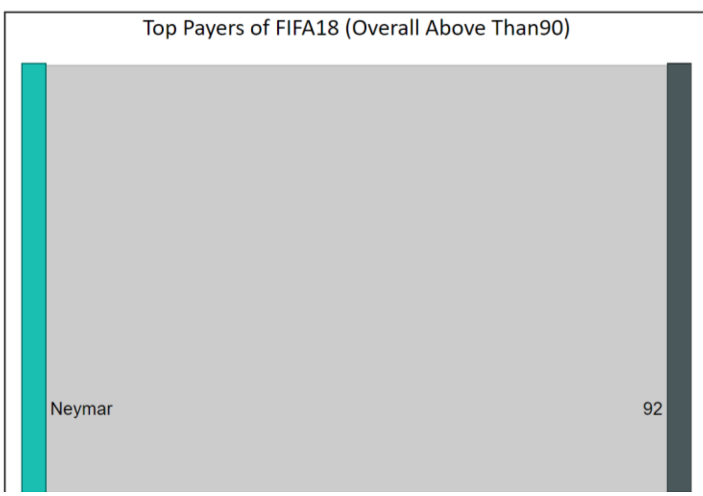
Players having an overall score greater than 90 were identified as top players. In order to further analyze their performance we selected the tool Power BI and created an interactive dashboard. The Sankey diagram on the left displays a list of these identified players along with their overall scores.

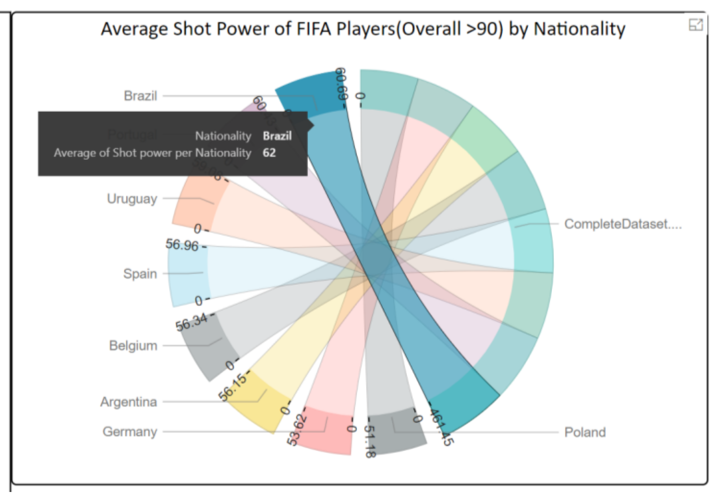


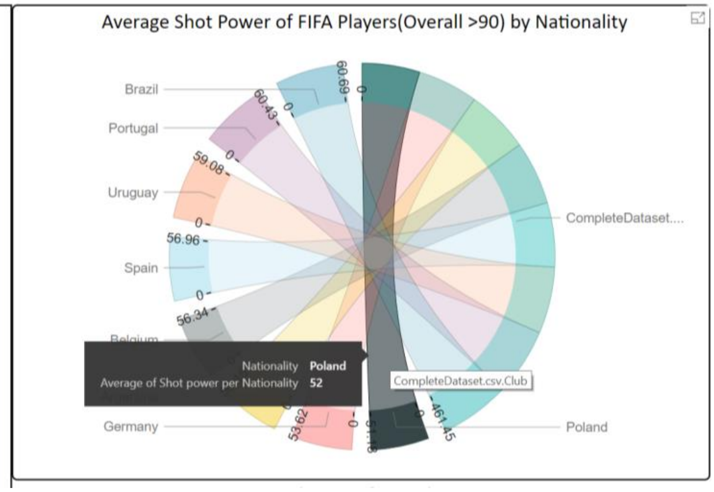
The chord diagram highlights the Shot Power which is essentially the shooting power of these players and highlights their Nationality. This chart is created in comparison with the entire dataset.



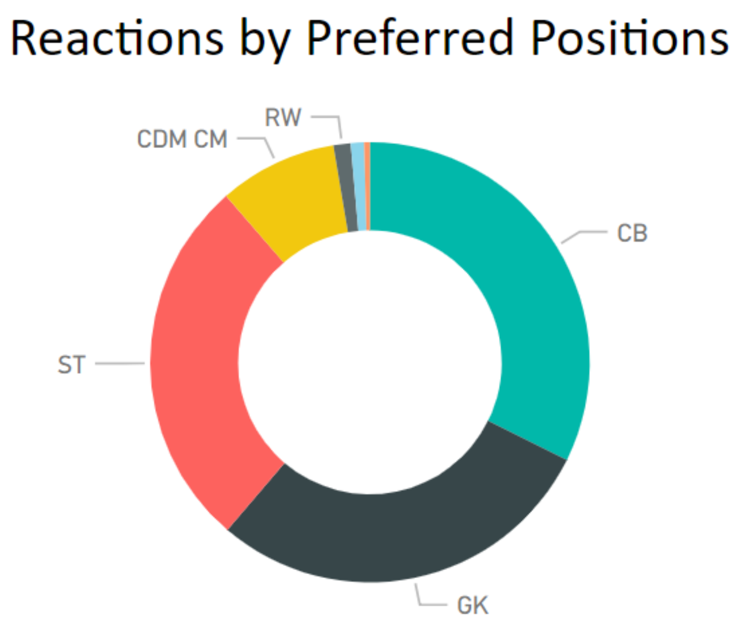
Since these are top players, the difference in shot power is minimal. Neymar from Brazil has the highest shot power of 62 compared to R. Lewandowski from Poland with a shot power of 52. Even among the top players we can deduce that there is a huge gap in the shooting power of these players.



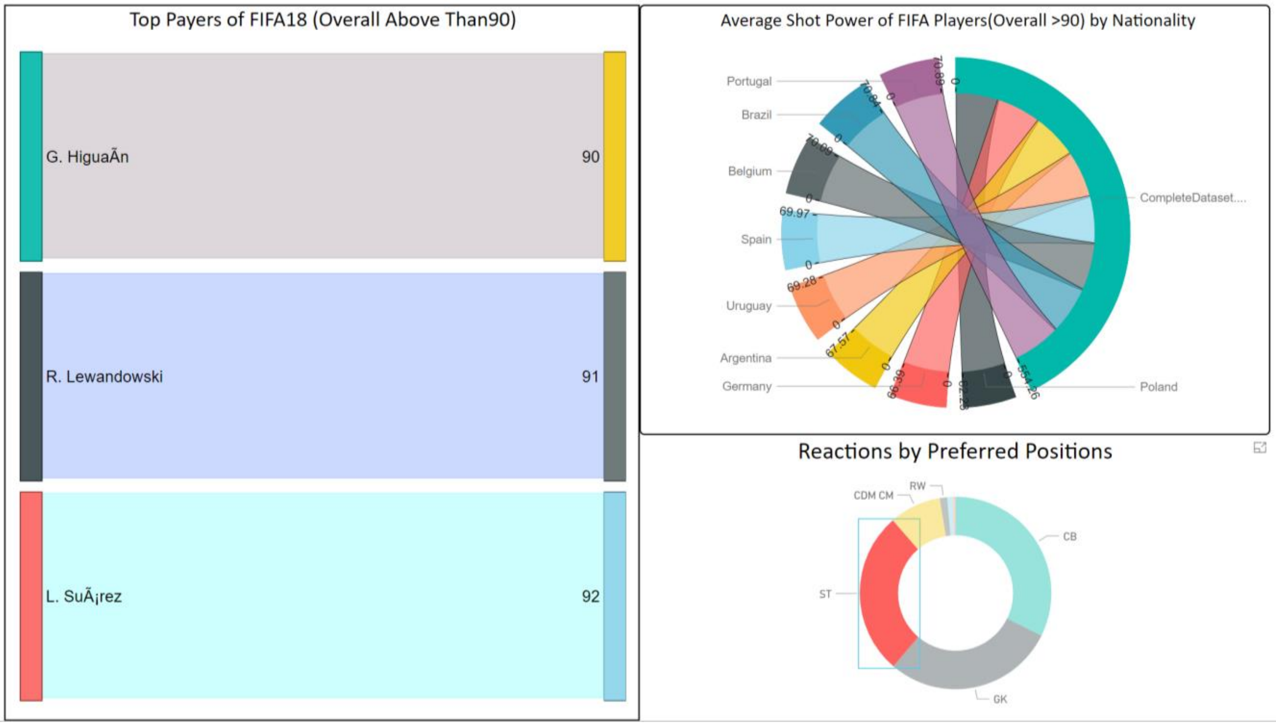




Along with the shooting power, we have also highlighted the preferred positions of these players. These preferred positions are highlighted with the percentage of Reaction. Reaction attribute determines the presence of mind of the players on the field which is a crucial criteria in a competitive sport like football. As expected, most players prefer ST: Special defense teams, GK: Goal Keepers and CB: Center back positions.



On highlighting each position, we get the filtered list of top players.

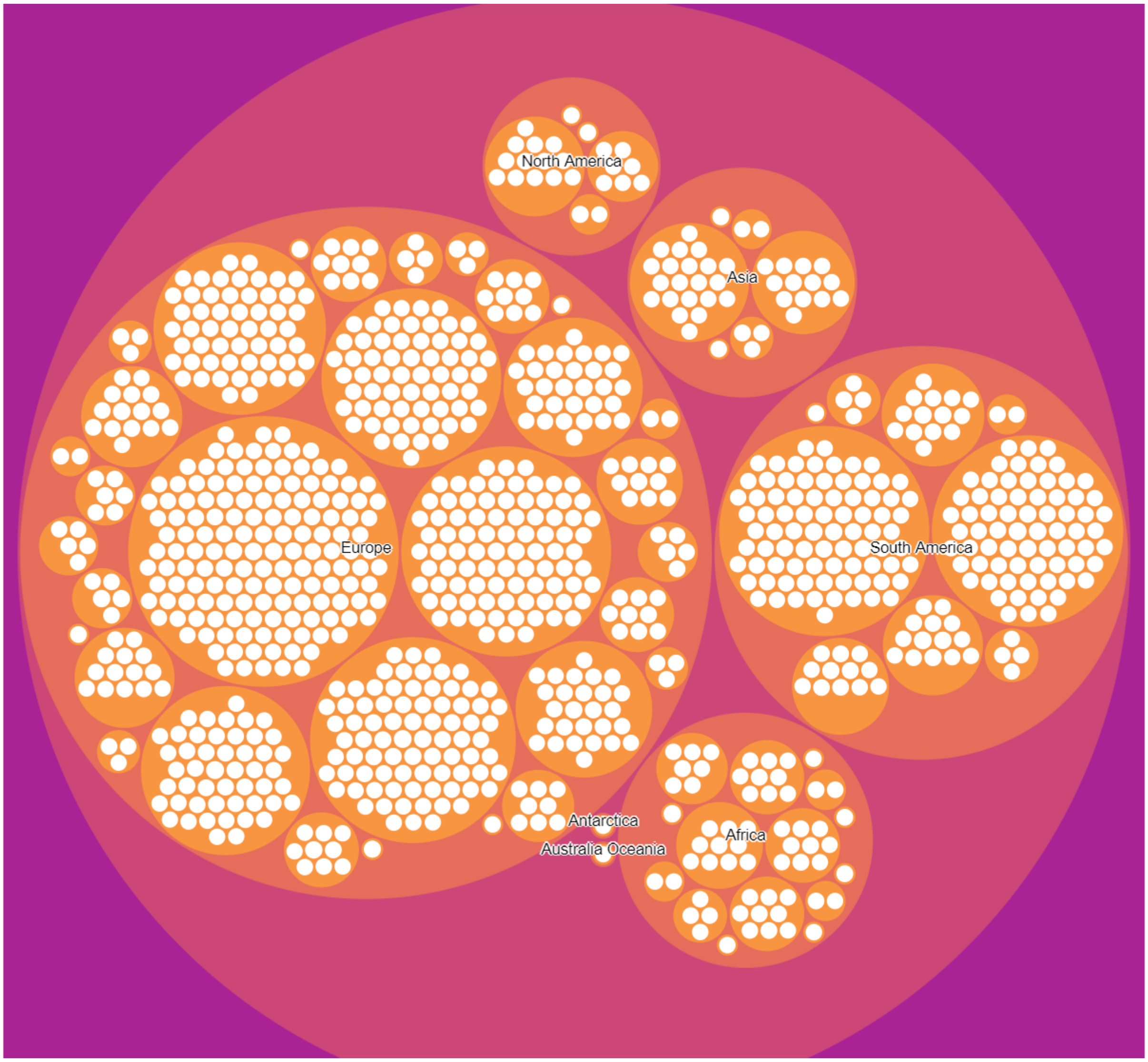

Top Payers of FIFA18 (Overall Above Than90)
G. HiguaÃn
90
R. Lewandowski
91
L. SuÃ¡rez
92
Average Shot Power of FIFA Players(Overall >90) by Nationality
Portugal
Brazil
Belgium
Spain
Uruguay
Argentina
Germany
CompleteDataset....
Poland
70.89
69.97
69.28
67.57
Reactions by Preferred Positions
CDM CM
RW
CB
ST
GK

Geographic distribution of top 2000 players based on overall scores

Circle-packing is the arrangement of circles inside some demarcation so that none of the circles overlap. Circle packing also displays hierarchy where you can get smaller clusters of circles packed within a bigger circle which itself is arranged next to or within other circles. The D3.js plot is interactive and dynamic, where one is able to invoke zoomable animations at different regions and clusters with the click of a mouse button. Each of the player's nationality was mapped to its respective continent. There were 162 distinct Nationality values in the dataset and these countries were mapped to 6 continents: Asia, Europe, Africa, North America, South America and Australia/Oceania. In the plot, the 6 continents will be the parent class (outer circles). We can dive deeper within this class to find the countries (sub-class / sub-circles) and within each country, we will find the players (inner circles). The size of the player circle is determined by the Overall variable. A continent dictionary was created with the names of the continents as the keys and the list of countries as the values for each key. A function was defined to assign the continent for each country. The top 2000 players were chosen based on the overall value. Groupings of the players were hence identified using the Nationality and Continent. This grouping will be displayed with the circle graph plot and is fed into the json file. The data to be displayed is stored in the json file.

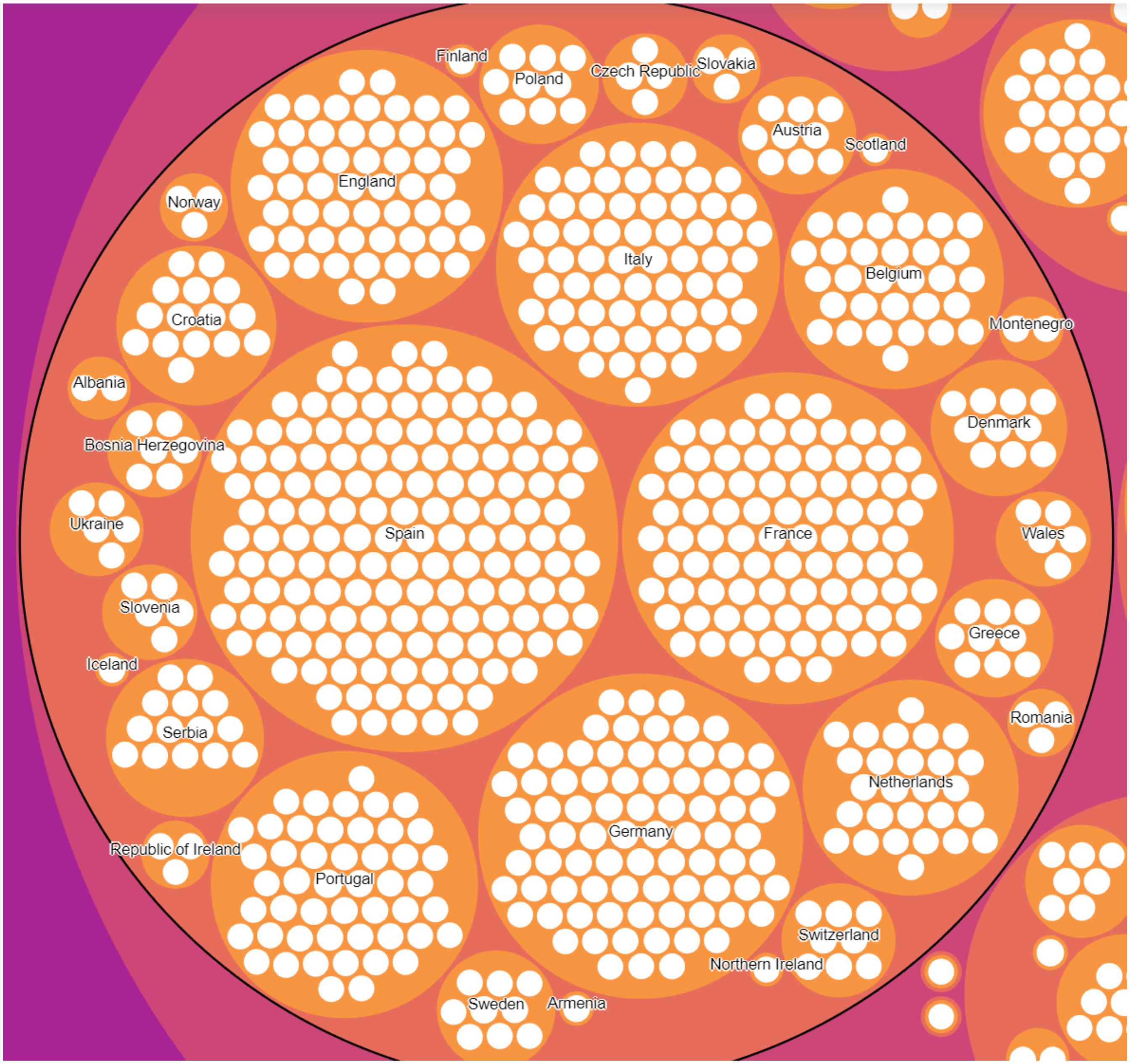
In short this visualization represents the globe and allows us to get the overall geographic location of top 2000 players.

Data Preparation: In order to achieve this we have used Python with numpy, pandas, seaborn, matplotlib, plotly etc. The data was converted to json and plot was created using IPython.



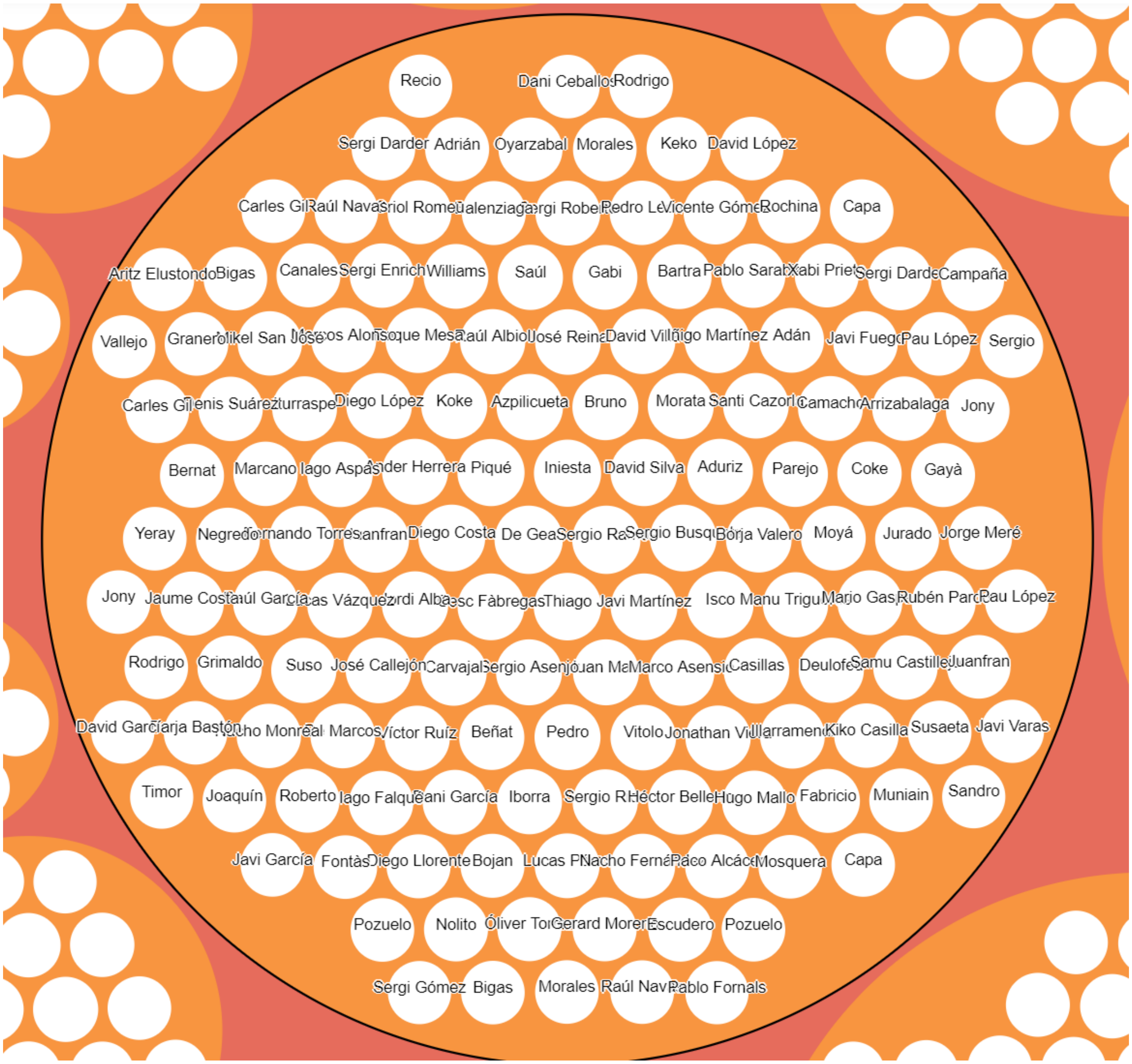
The plot shows the player density on each continent. From the plot, Europe is the leading continent with maximum of the top 2000 players. It is followed by South America, Africa, Asia , North America and then finally Australia and Antarctica being the least popular.

Lets zoom into Europe:



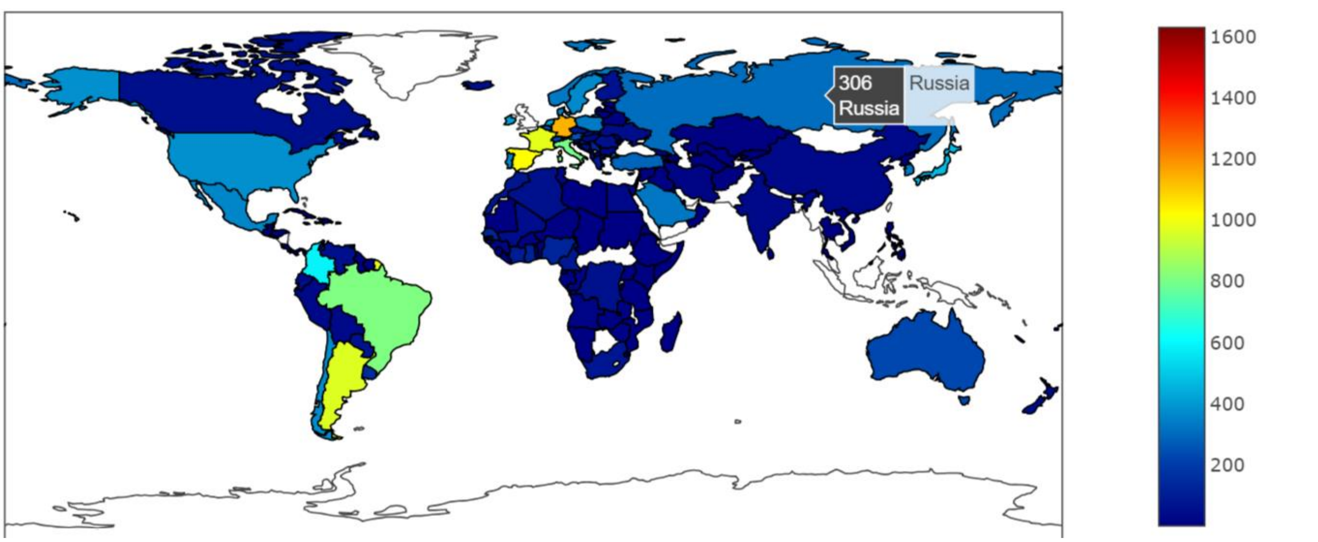
Spain, France, Germany, Italy and England have the maximum number of players. Scotland, Finland etc. having just single players each.

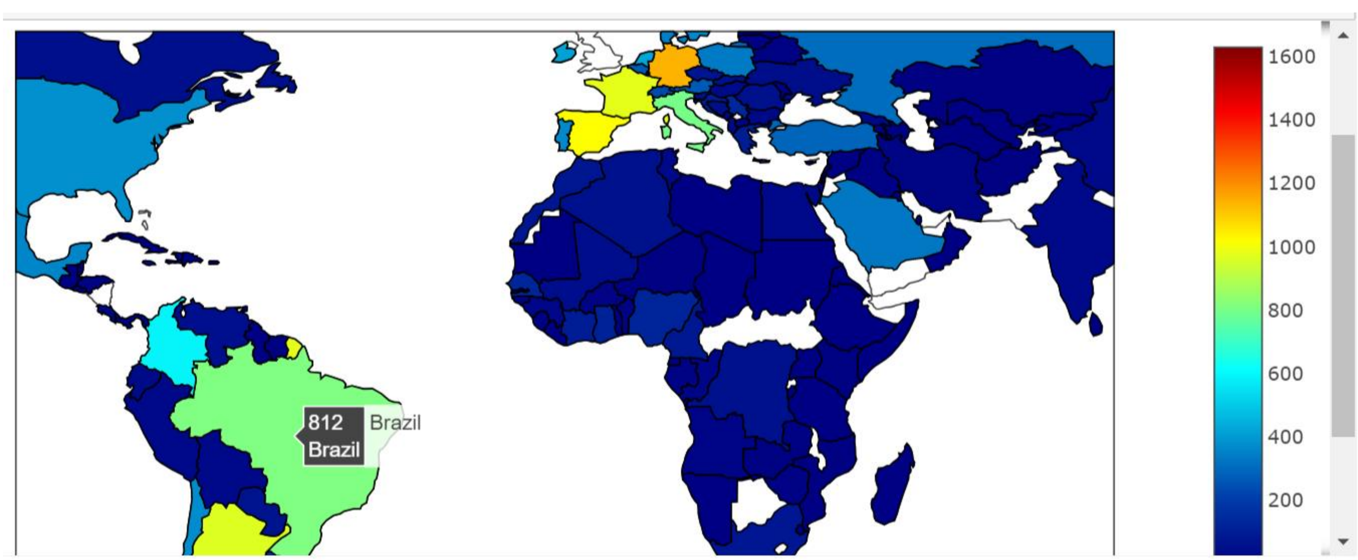
On zooming into Spain we get the names of all the top 2000 players from Spain:



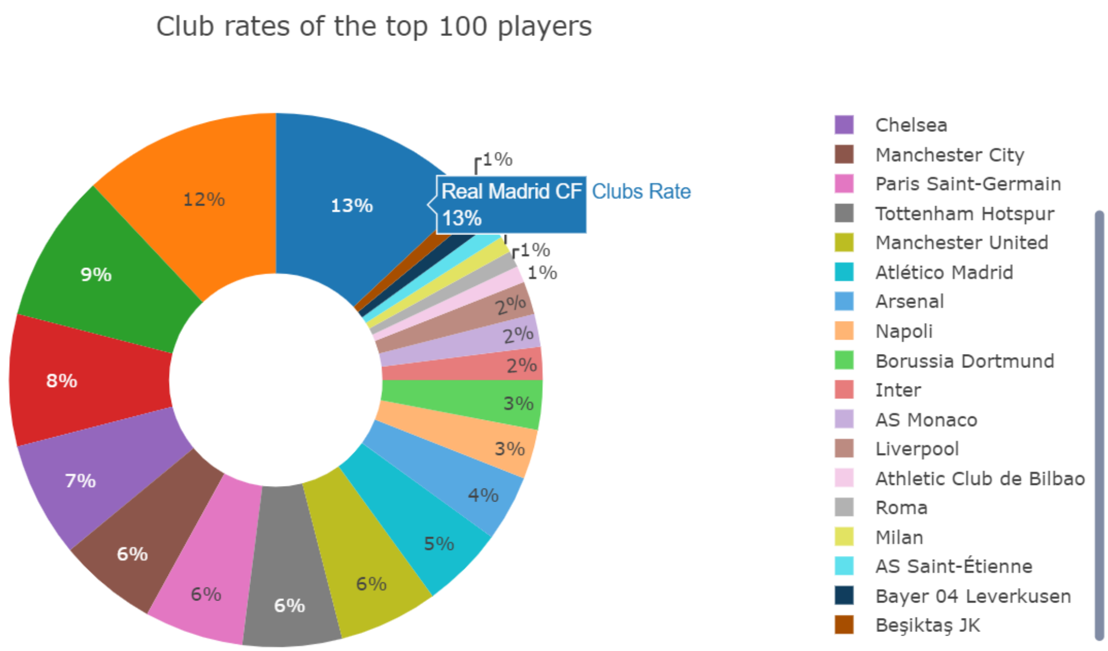
Map Distribution of the Players and How many players from the same Country?

Lets check the player density for all the players now. Similar to the results from the previous plot, maximum players are from European countries like Spain, Germany etc. or Brazil or Africa. To make this interactive we have added zoom in and hovered labels.

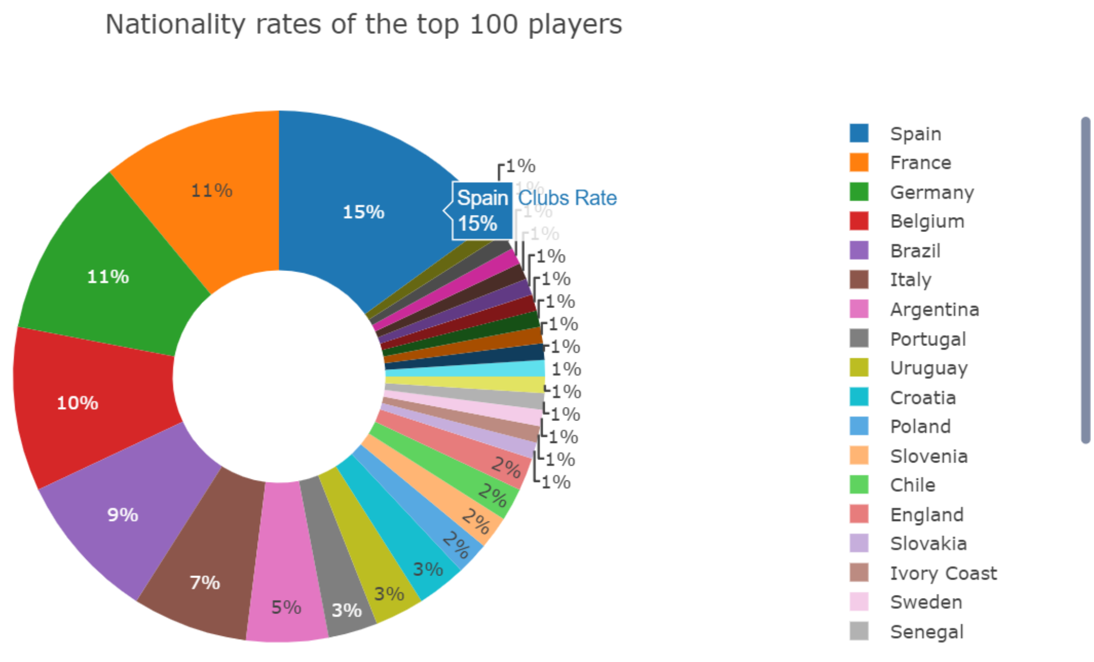




Club rates and Nationality rates of top 100 players



Real Madrid has the highest rates, almost double of its competitors like Chelsea, three times of Arsenal etc. FC Bayern Munich being the second highest. On hovering over each club we get the club names and rate percentage.



Nationality wise, Spain has the highest rates. This can be a reason why Spain has maximum player density. In fact, countries like France, Germany and Brazil which have higher rates have high density of players as well as top performing players.

We can deduce that higher the Nationality rate for a country greater is the density of top performing players in that country.

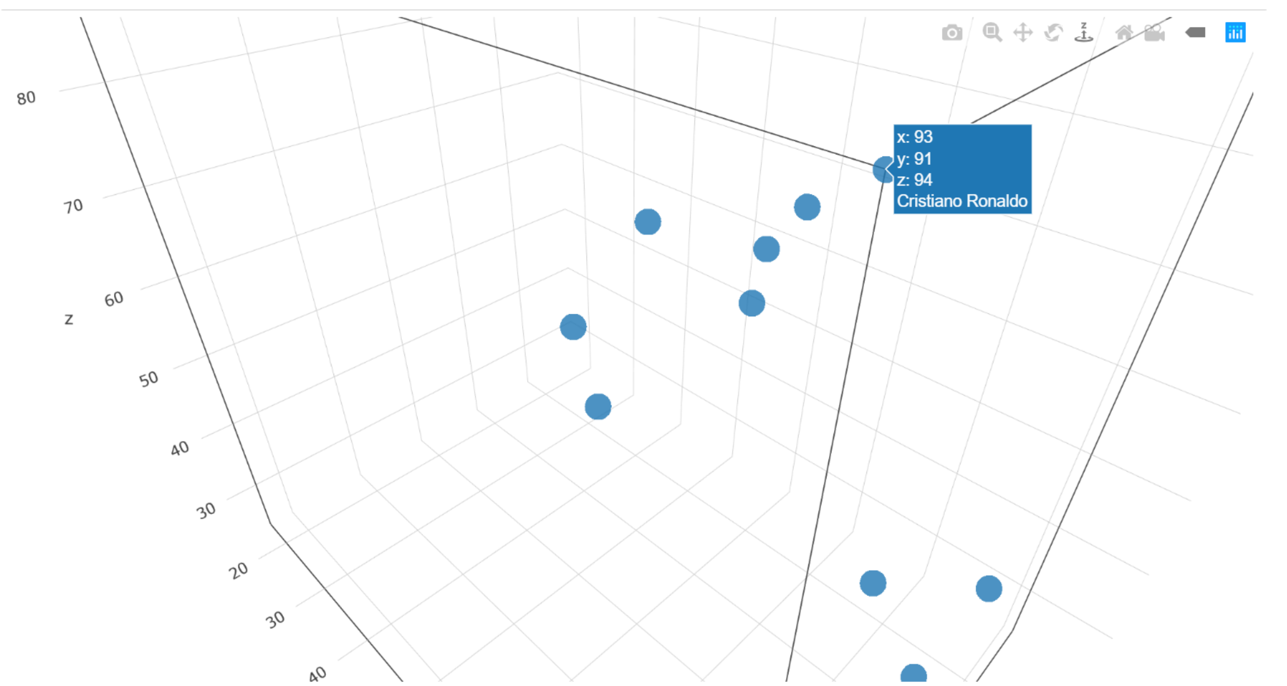
Analysis for Messi and Ronaldo

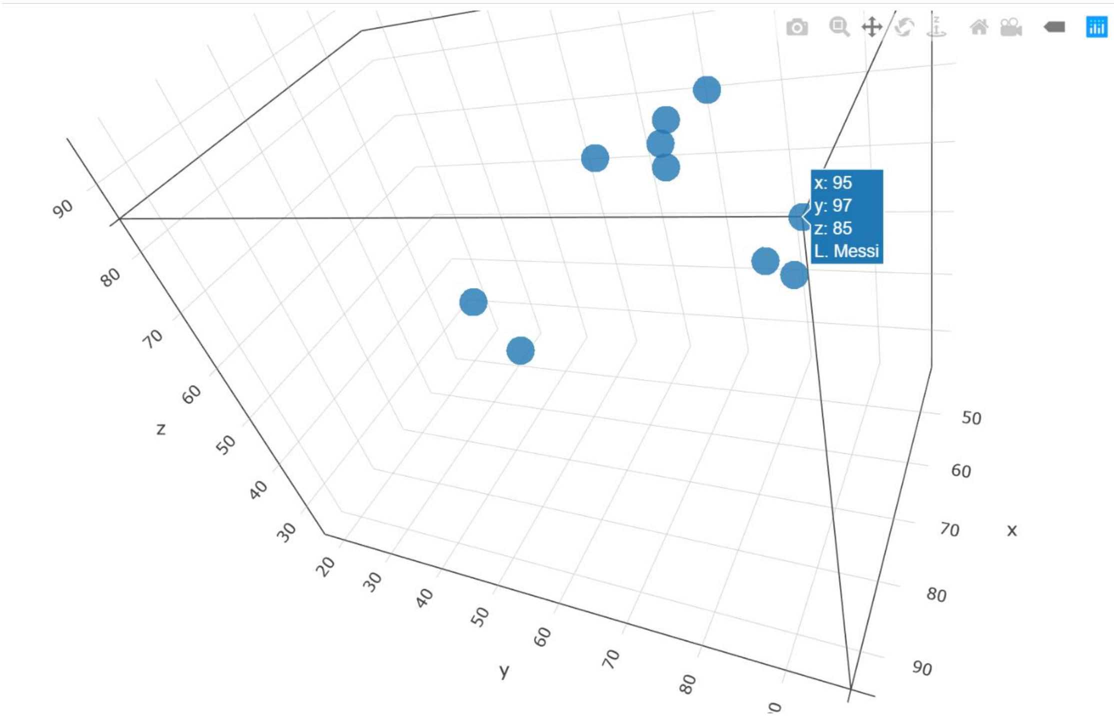
Messi and Ronaldo are probably the first names that come to our mind when we talk football. Even for non-football enthusiasts these names are not alien. Hence we wanted to compare the performance of these two players.

Although there isn't a major difference in most attributes, Ronaldo is physically much stronger than Messi, also being slightly better in defending. There can be no conclusion as to who is better than whom.

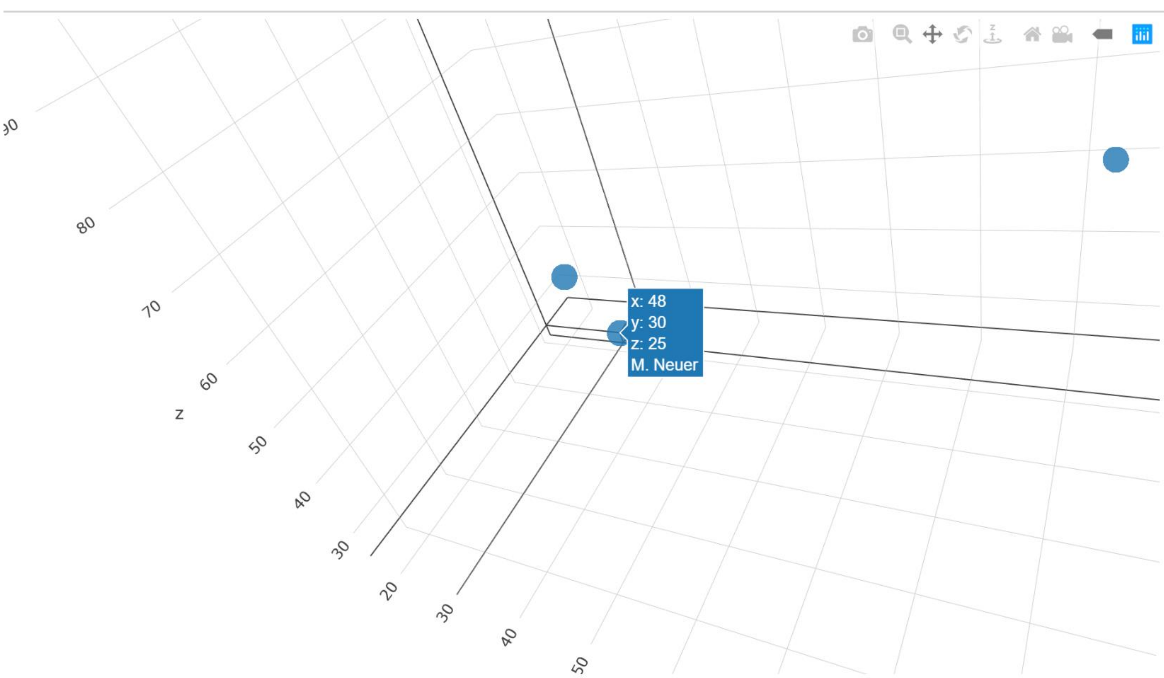
Top 10 Players 'Ball Control', 'Dribbling' and 'Shot Power' Features Comparison with Scatter 3D Plot

The plot below compares the three most important attributes of a football player namely: Ball Control(X), Dribbling(Y) and Shot power(Z) for top 10 players. This 3D plot can allows us to zoom into each player and compare the measures:

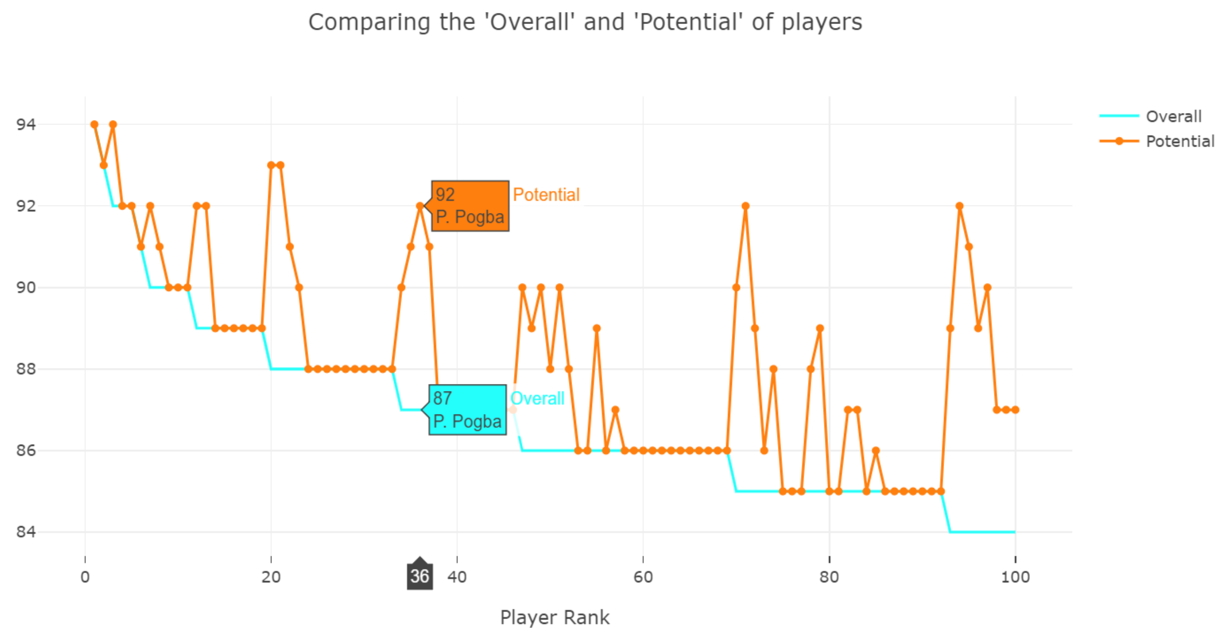




If we hover over the graph, we observe that Messi, Ronaldo, Hazard and Neymar have approximately the same scores of ball control and dribbling.However, in the case of shot power Ronaldo is much better. Furthermore, Neuer and Gea have the worst ball control, dribbling and shot power performance features and hence are observed closest to the origin.

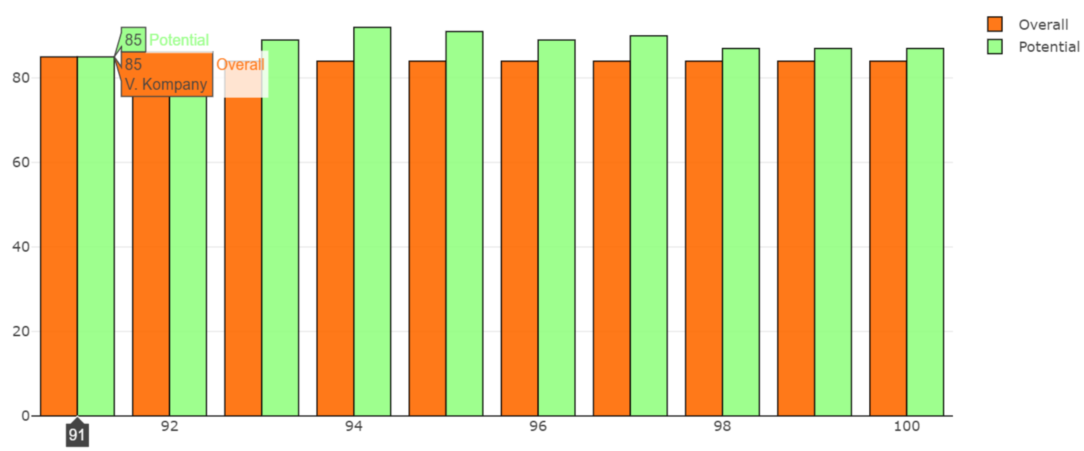


Comparing the 'Overall' and 'Potential' of Players with Scatter Plot for top 100 players

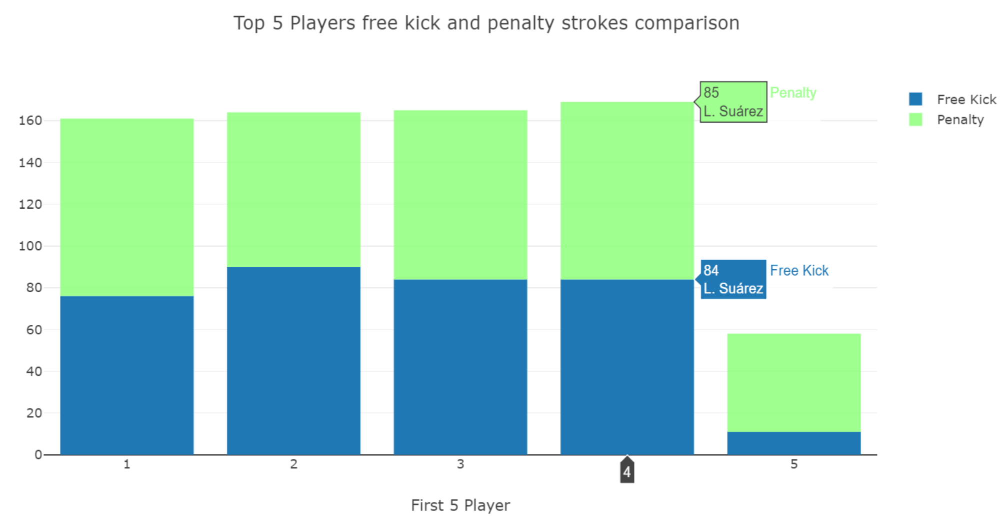


It is interesting to observe that players have higher Potential points as compared to their Overall scores. In some cases there is a huge gap in the two scores indicating that the players are not performing as per their potential and can do much better.

Lets highlight Comparing 'Overall' and 'Potential' Values of Players between 90 and 100 with Bar Plot, potential is still greater for most.

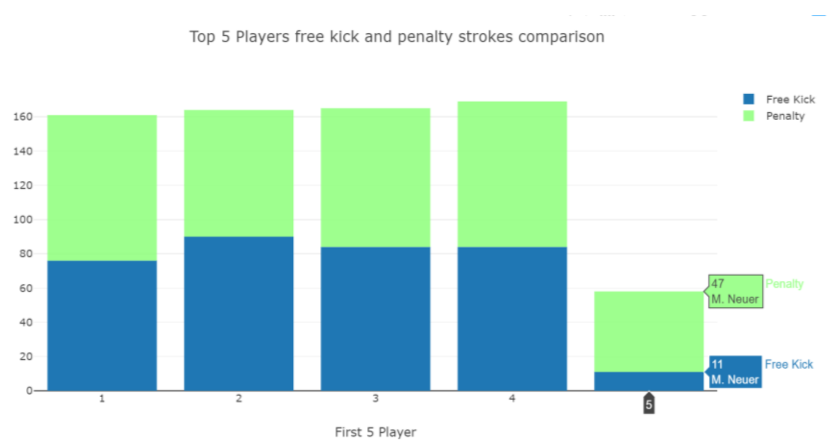


Top 5 Players Free kick and Penalty Strokes Comparison with Bar Plot



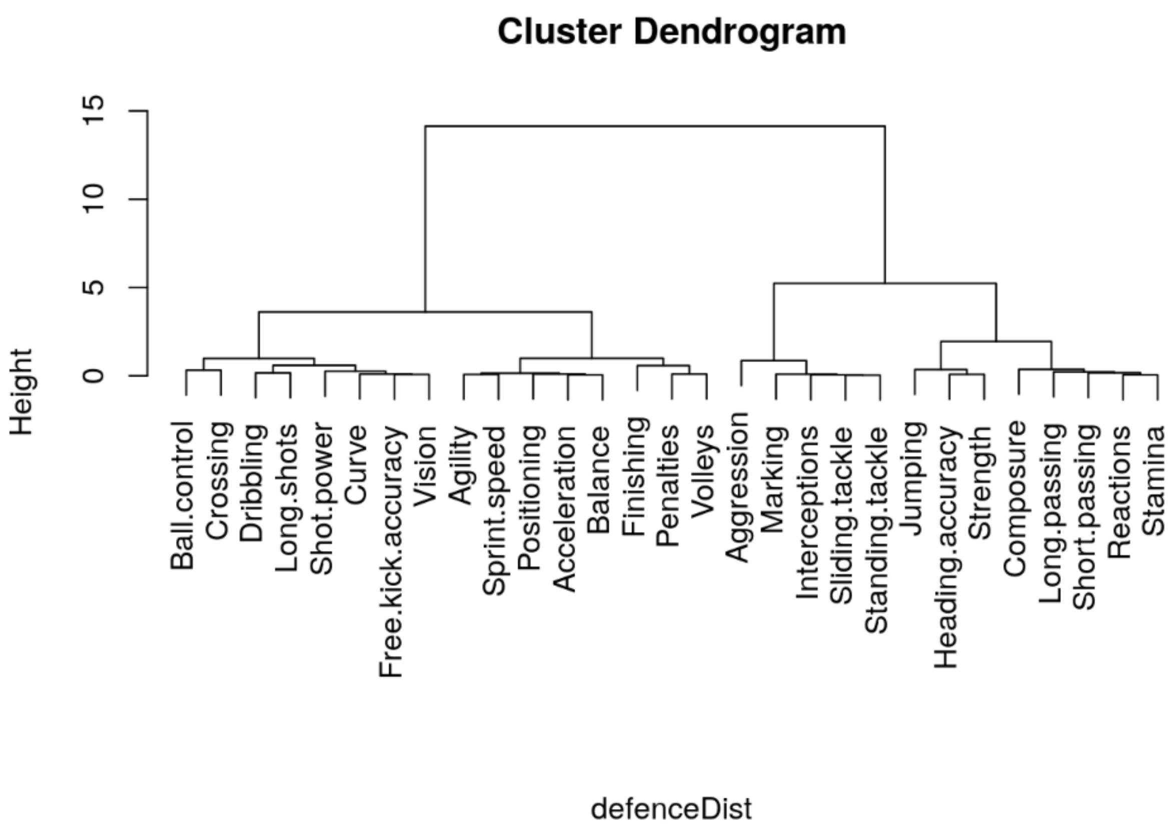
In case of fouls committed by opposing team, players are awarded free kicks and penalty strokes depending on field positioning. These features increase the chance of scoring goals.

Although there isn't much difference in these two values awarded to the top 4 players, The free kicks and penalty strokes of M. Neuer(Rank 5) are quite few compared to the others.



Good Defenders

The data set consists of over 50 attributes that show the various scores assigned to each players based on their performance. We wanted to cluster these attributes and identify the scores needed for specific positions. For clustering we co-relate the player attributes with their defense scores. The results from the clustering are as shown below:



In order to understand these clusters we need to observe the results from the clusters and correlation.

```
> #understanding clusters
> defenceCorr <- mutate(defenceCorr, Attributes = row.names(defenceCorr), Cluster = groups)
> defenceCorr
            CB           LB          LCB         LWB           RB          RCB
1  -0.279416529  0.003611011 -0.279416529  0.10708289  0.003611011 -0.279416529
2   0.730511288  0.638805152  0.730511288  0.60097053  0.638805152  0.730511288
3  -0.251601215  0.035711002 -0.251601215  0.16540137  0.035711002 -0.251601215
4  -0.251147091  0.005845358 -0.251147091  0.10931449  0.005845358 -0.251147091
5   0.050210653  0.288666964  0.050210653  0.43220297  0.288666964  0.050210653
6   0.323641440  0.426948243  0.323641440  0.49944478  0.426948243  0.323641440
7   0.142773708  0.431733428  0.142773708  0.56319448  0.431733428  0.142773708
8  -0.037661405  0.193606330 -0.037661405  0.32710602  0.193606330 -0.037661405
9  -0.177568397  0.105787218 -0.177568397  0.26957516  0.105787218 -0.177568397
10 -0.454848458 -0.301292779 -0.454848458 -0.17033296 -0.301292779 -0.454848458
11  0.018147669  0.198962598  0.018147669  0.30964074  0.198962598  0.018147669
12  0.449566214  0.278668594  0.449566214  0.21430207  0.278668594  0.449566214
13  0.938698558  0.889451341  0.938698558  0.82220212  0.889451341  0.938698558
14  0.309932910  0.260220804  0.309932910  0.23241987  0.260220804  0.309932910
15  0.441731378  0.599145716  0.441731378  0.67871028  0.599145716  0.441731378
16 -0.112829217  0.066711113 -0.112829217  0.19608689  0.066711113 -0.112829217
17  0.930355047  0.850546455  0.930355047  0.76171337  0.850546455  0.930355047
18 -0.269637406 -0.150201332 -0.269637406 -0.04382799 -0.150201332 -0.269637406
19 -0.253693421 -0.049083461 -0.253693421  0.09224552 -0.049083461 -0.253693421
20  0.406205239  0.521104254  0.406205239  0.58789219  0.521104254  0.406205239
21  0.328840022  0.520680425  0.328840022  0.62903713  0.520680425  0.328840022
22 -0.005559316  0.118882216 -0.005559316  0.22258358  0.118882216 -0.005559316
23  0.929143191  0.876752987  0.929143191  0.79014355  0.876752987  0.929143191
24 -0.227072638  0.028422995 -0.227072638  0.11966779  0.028422995 -0.227072638
25  0.402363299  0.553747714  0.402363299  0.60701177  0.553747714  0.402363299
26  0.949810088  0.879563859  0.949810088  0.79323021  0.879563859  0.949810088
27  0.496206720  0.278288880  0.496206720  0.20803346  0.278288880  0.496206720
28  0.002055717  0.220391901  0.002055717  0.35574063  0.220391901  0.002055717
29 -0.272185561 -0.112024965 -0.272185561  0.01336050 -0.112024965 -0.272185561
           RWB        Attributes Cluster
1   0.10708289      Acceleration       1
2   0.60097053        Aggression       2
3   0.16540137           Agility       1
4   0.10931449           Balance       1
5   0.43220297      Ball.control       3
6   0.49944478         Composure       4
7   0.56319448          Crossing       3
8   0.32710602             Curve       3
9   0.26957516         Dribbling       3
10 -0.17033296         Finishing       1
11  0.30964074 Free.kick.accuracy       3
12  0.21430207   Heading.accuracy       4
13  0.82220212      Interceptions       2
14  0.23241987           Jumping       4
```

Based on the results of clusters we identified 4 groups and named them:

Advance: Attributes like Finishing, sprinting, volleys etc. which are never a "Must Have" for a defender.

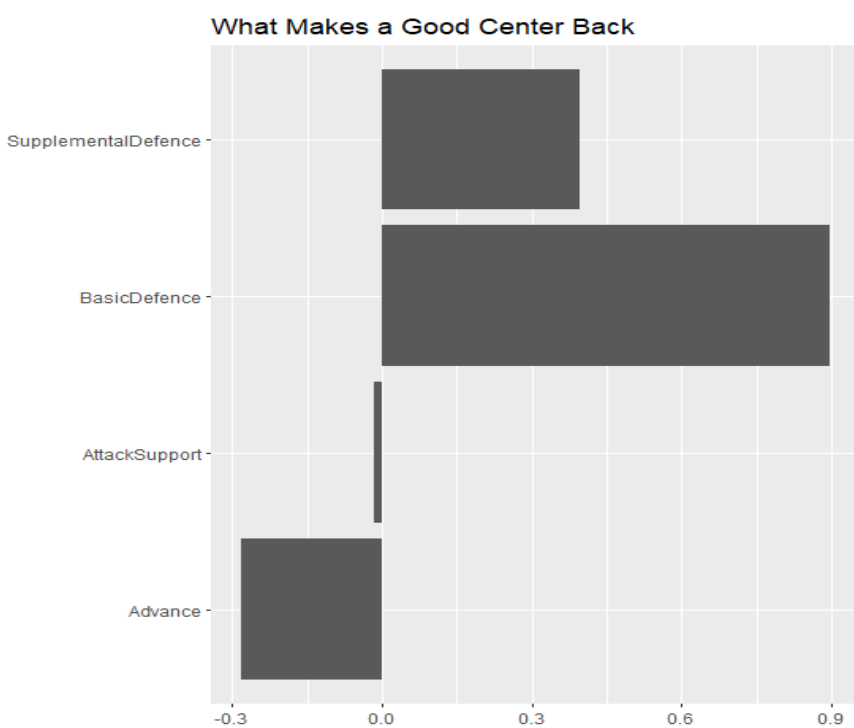
Basic Defence : Sliding tackle, Standing Tackle, Marking etc which are the bread and butter

Attack Support : Attributes like corssing, ball control and curve which are more relevant to supporting attack through the flanks

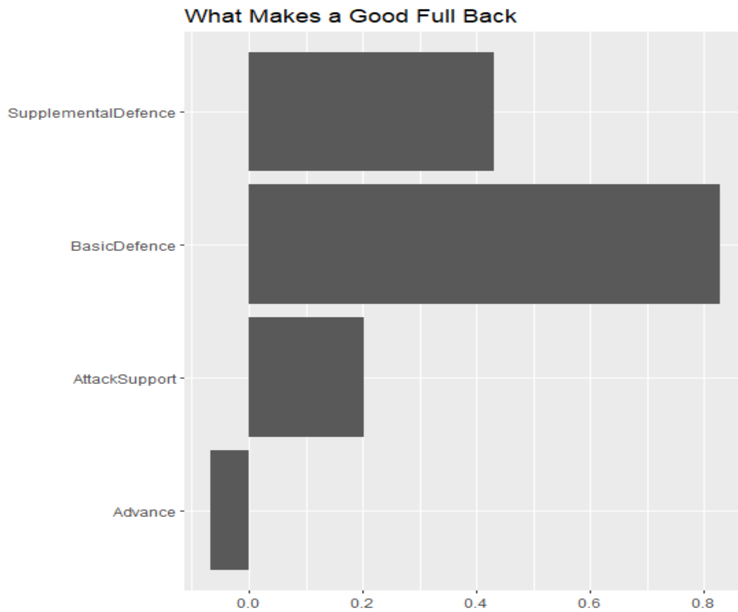
Supplemental Defense : Composure, Reaction, Long passes etc. which differentiate a world class defender from the rest**.**

Based on these groups, on averaging the correlation, following were the results:

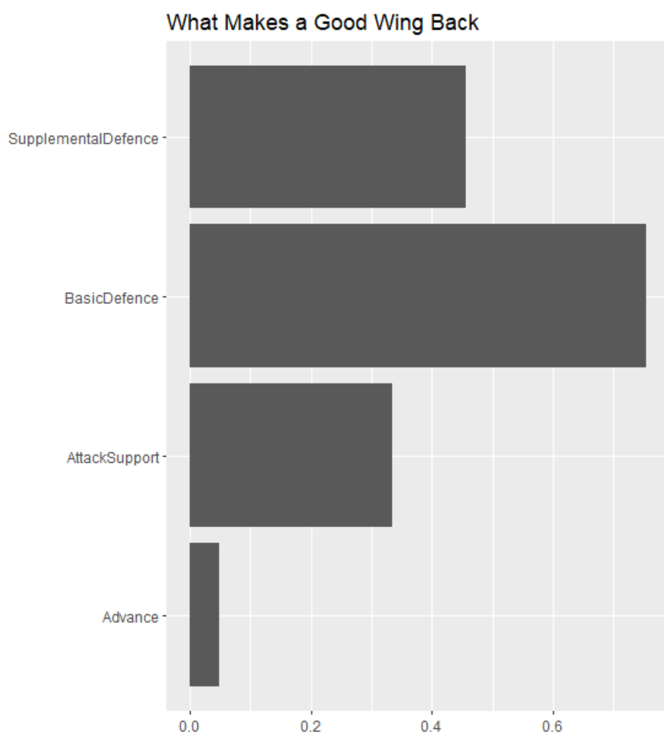
A good center back must have 90% of the basic defense qualities on an average, with at least a few attributes from the Supplemental Defense qualities.



Even though a good Center Back need no have Attack Support qualities, they are necessary for a good full back player. The Supplemental Defense needed is greater in this case whereas the Advance qualities still unnecessary.



Incase of a Good Wing Back, all four characteristics are required with highest attack support. The three can be used to compare the various attributes required.

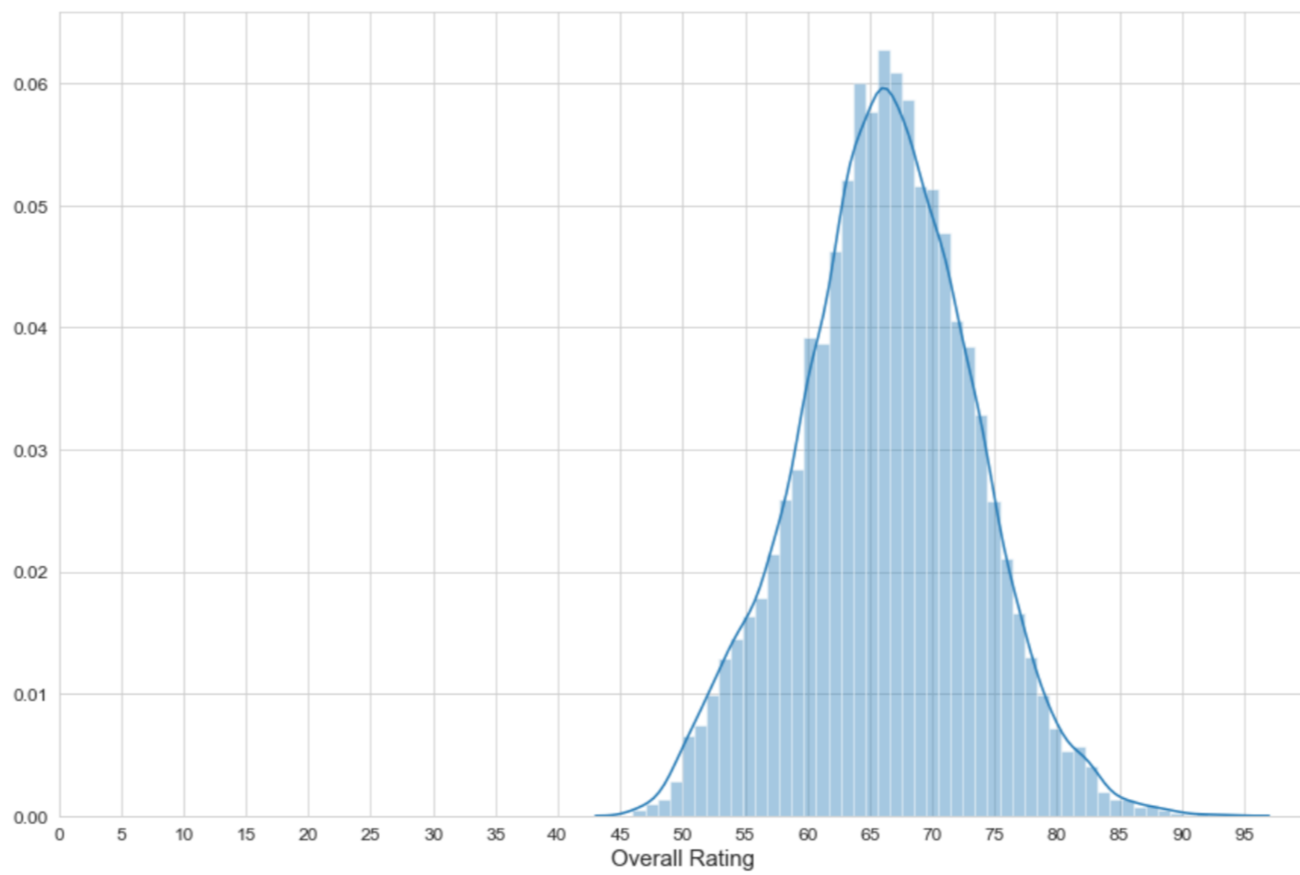


World cloud of Names of top 100 players based on their overall score

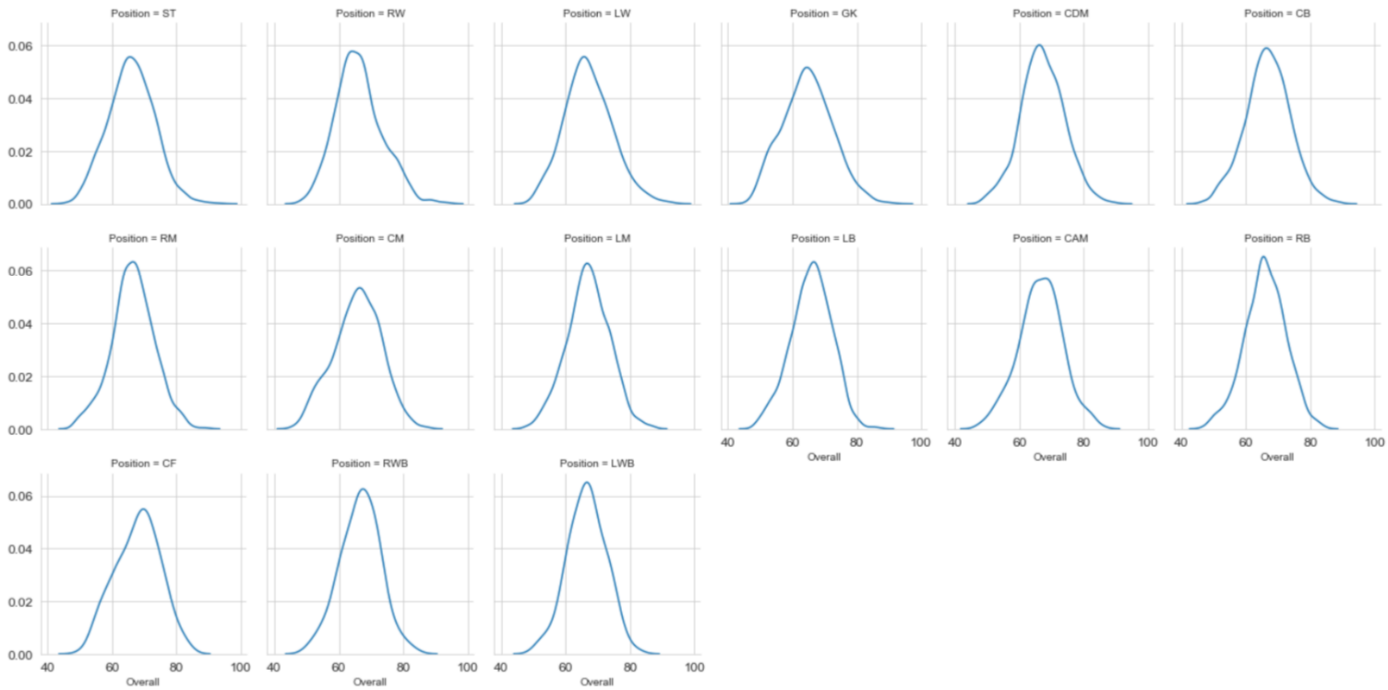
We can identify most of the players as they normally make the headlines. The word cloud shows that Thiago is on the top followed by Silva. Ronaldo, Neymar and Messi.

Statistical Analysis of overall rating of players

The plot of the overall variable shows a normal distribution. On an average the players have an overall potential close to 68. This visualization can be used to judge the potential needed to make a career in football. The potential ranges from 45 to 90. There is a minor skewness towards the right indicating the number of players having potential greater than the mean of 68 is slightly higher.



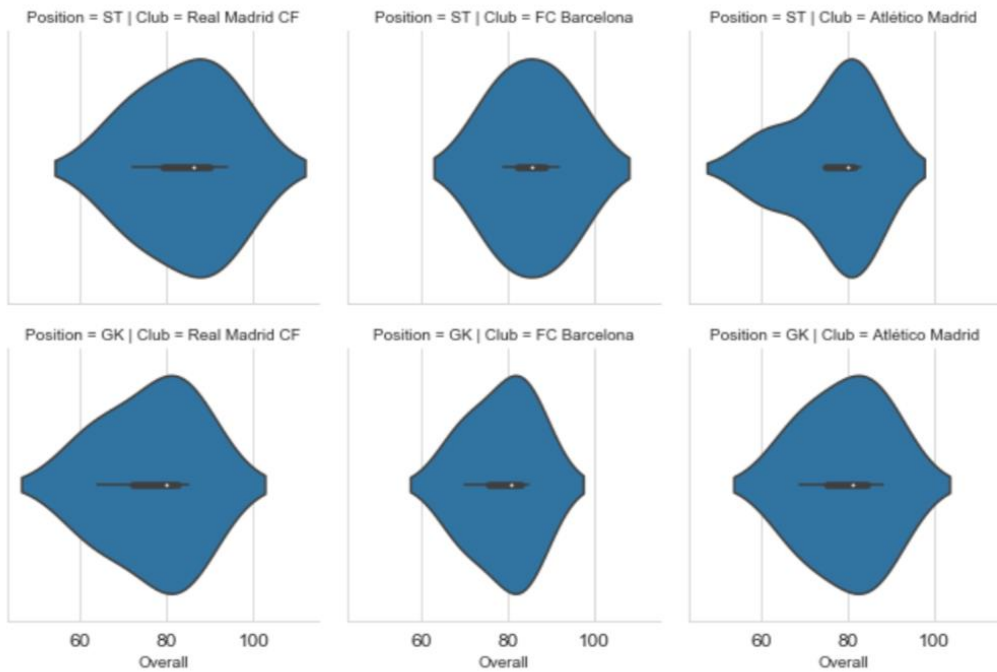
Lets explore the Overall rating based on their positions:



As from the plot it is clear that the overall score is distributed normally having similar properties for various positions. The mean value is around 70.

On comapring the overall scores for three famous clubs:Real Madrid CF, FC Barcelona, Atlético Madrid, for Special Teams and Goal Keeper Positions.

we get the density plot as:



Density of players(ST) of Atlético Madrid with overall score more than 80 is much higher as compared to the other two. Maximum number of Goal Keepers have an overall score of about 80 for all three clubs.

References:

1. **(Shrivastava, 2017)**

Bibliography

Shrivastava, A., 2017. *Kaggle.* [Online]
Available at: https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset