

¹Data Analysis

Data Information:

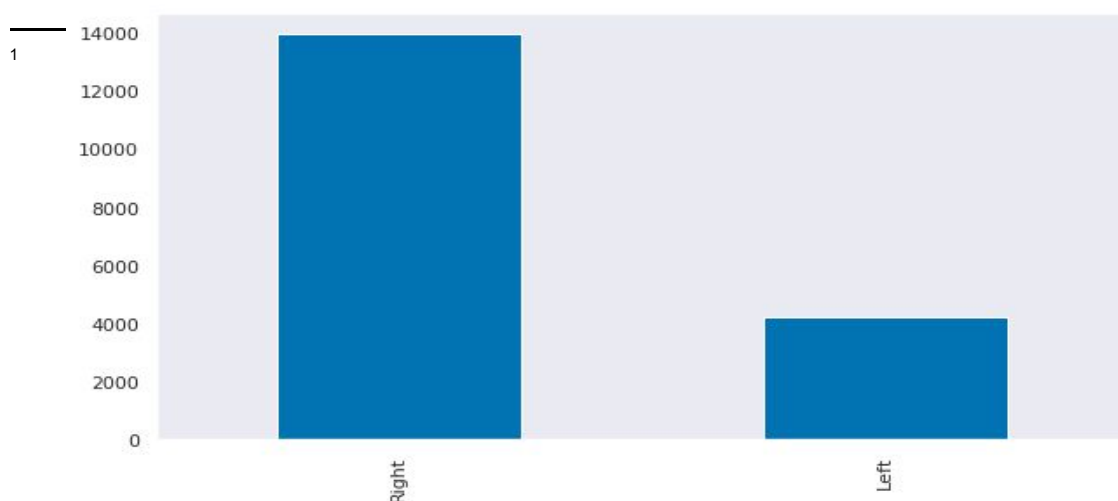
The dataset consists of 89 playing features of 18207 player which explains about the game and the skills of 17194 unique individuals. Further is the Descriptive Analysis of the data provided.

1. It seems that the giving sample holds the majority of the young population as the sample mean here is 25 years, though the Age spread is 16-45 years but the 50% of the sample is 25 years \pm 4 years.

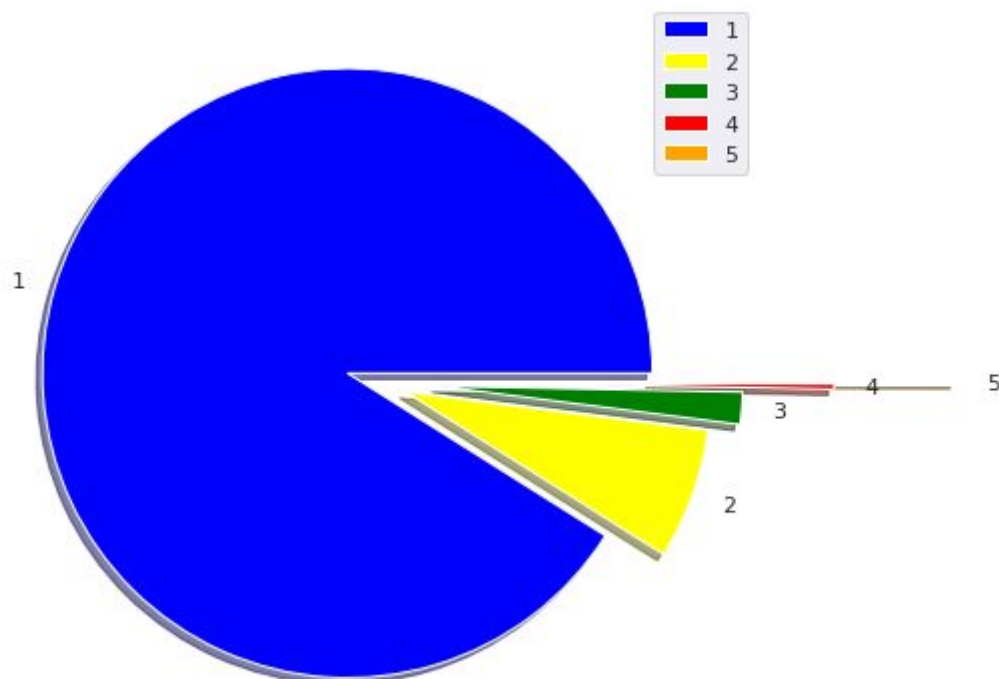
```
[ ] 1 data.describe()
```

	Unnamed: 0	ID	Age
count	18207.000000	18207.000000	18207.000000
mean	9103.000000	214298.338606	25.122206
std	5256.052511	29965.244204	4.669943
min	0.000000	16.000000	16.000000
25%	4551.500000	200315.500000	21.000000
50%	9103.000000	221759.000000	25.000000
75%	13654.500000	236529.500000	28.000000
max	18206.000000	246620.000000	45.000000

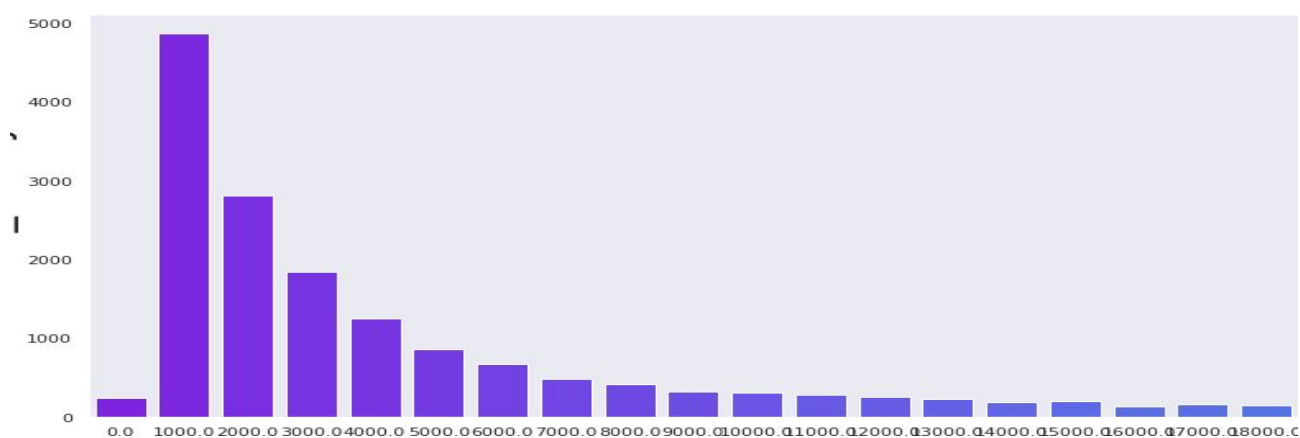
2. The data consisted of the Null values which were replaced by the mean values of respective columns so that the data can be used for the visualization.
3. Further looking the majority(80%) of the players are right-footed players.



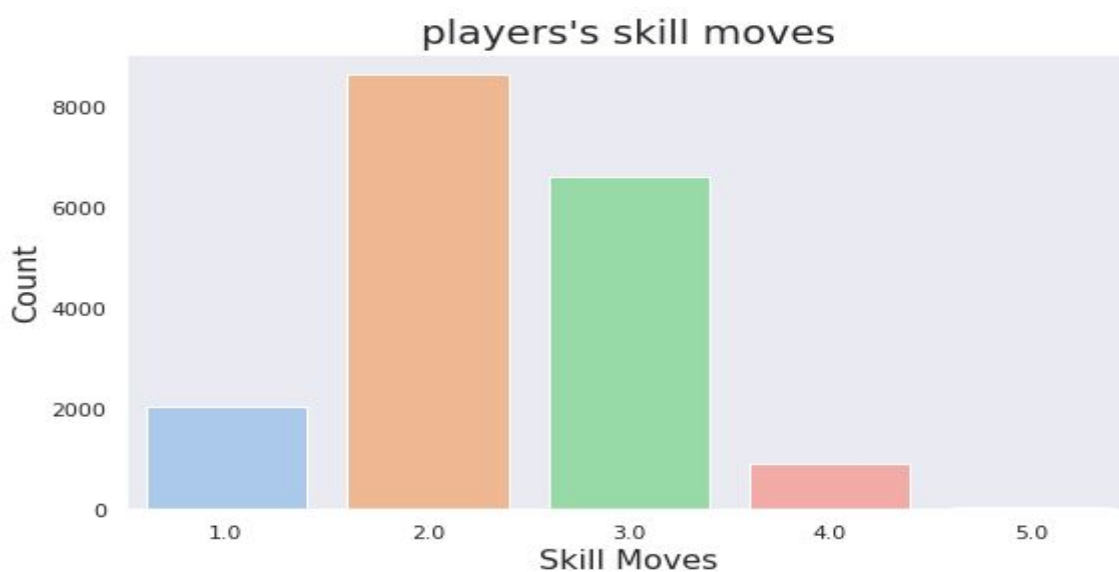
4. Looking into the data that we have got from the international Reputation we can see that the data is very much skewed as less than 1% of players have top ranks.



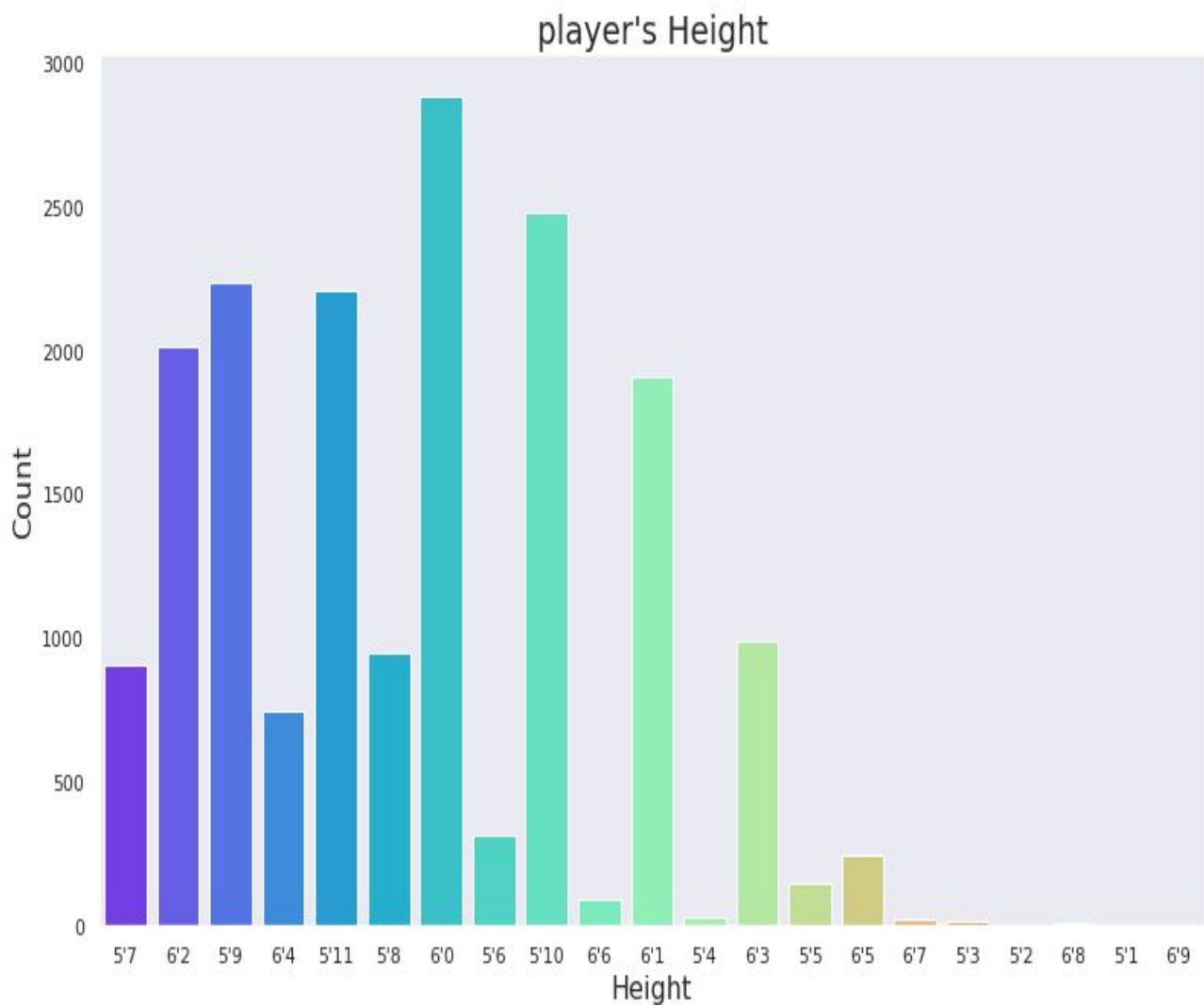
5. Plotting the Wages data in to bar chart shows:



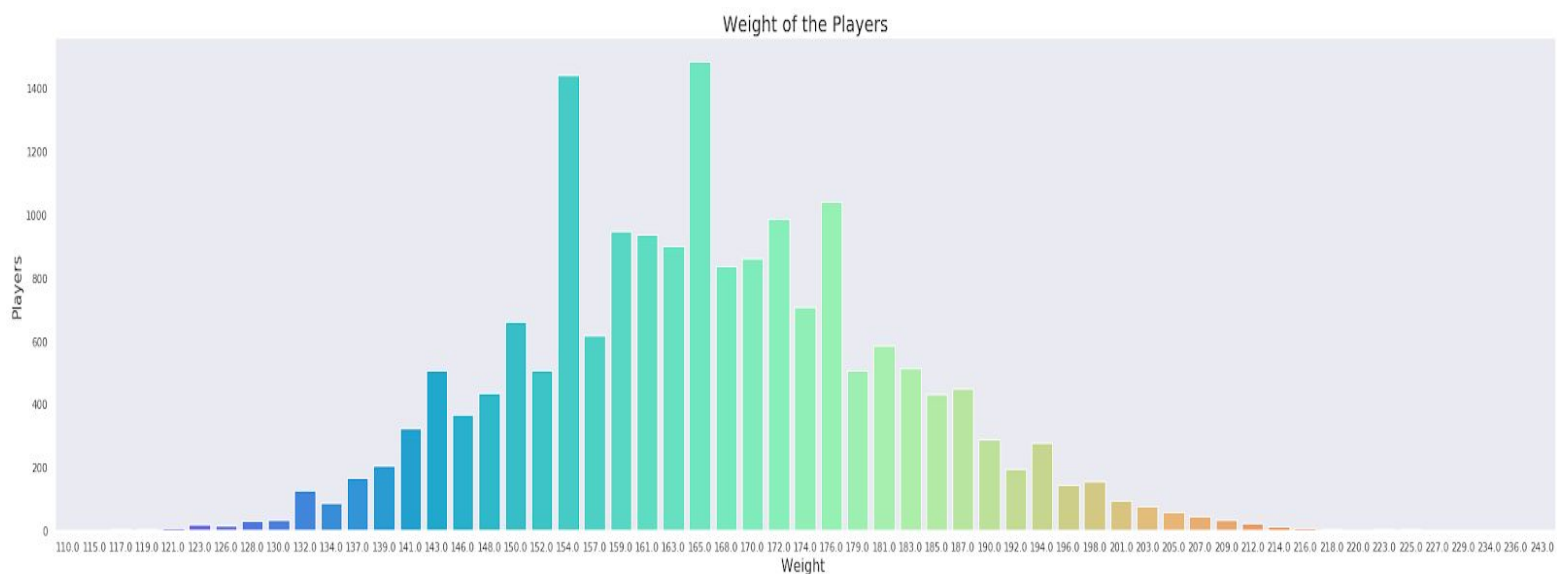
6. looking into the Skill move and having the scale of 1-5 majority of the payers falls as:



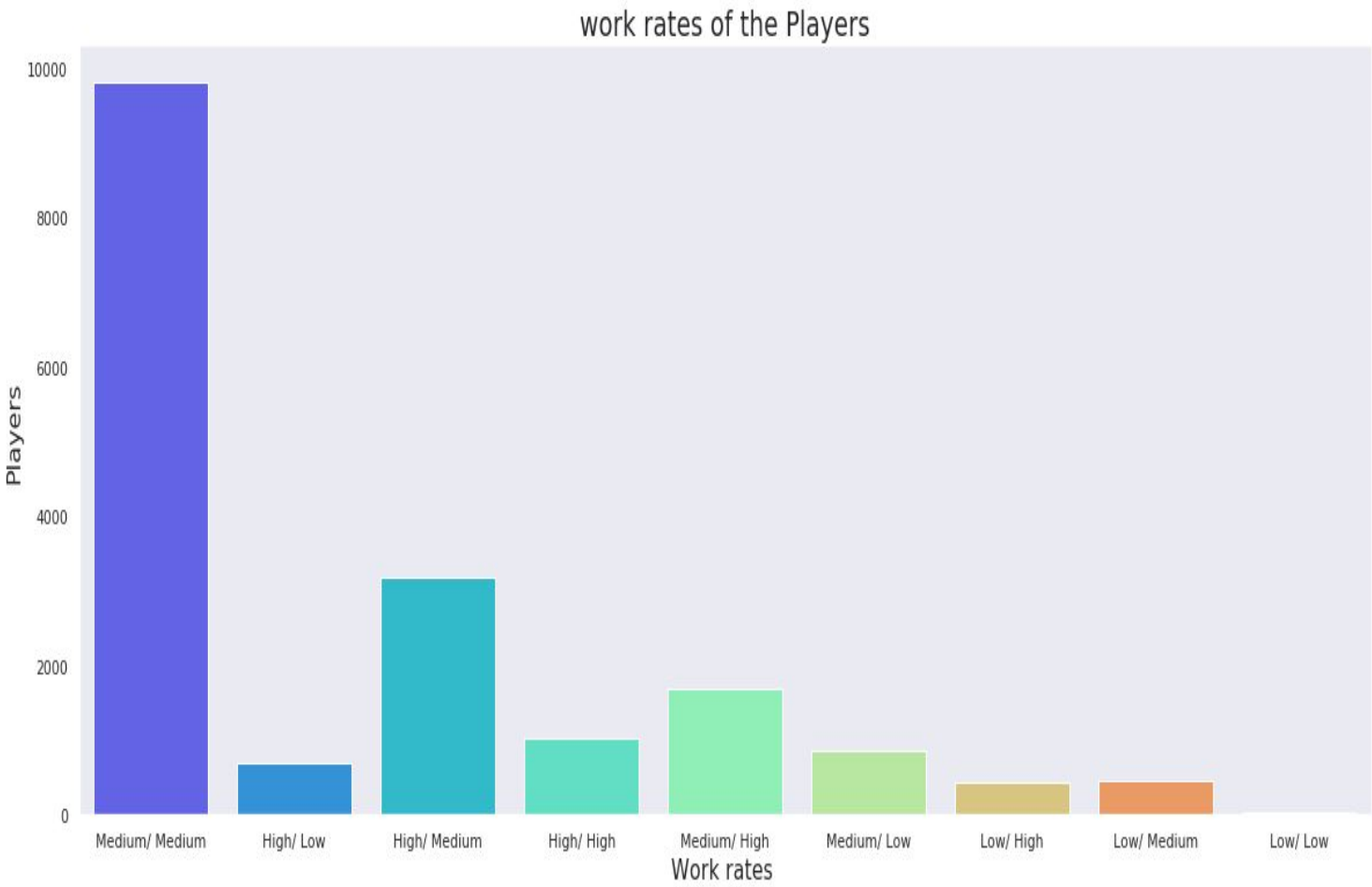
7. height data shows that the majority of players fall under the height ranging 5.7' to 6':



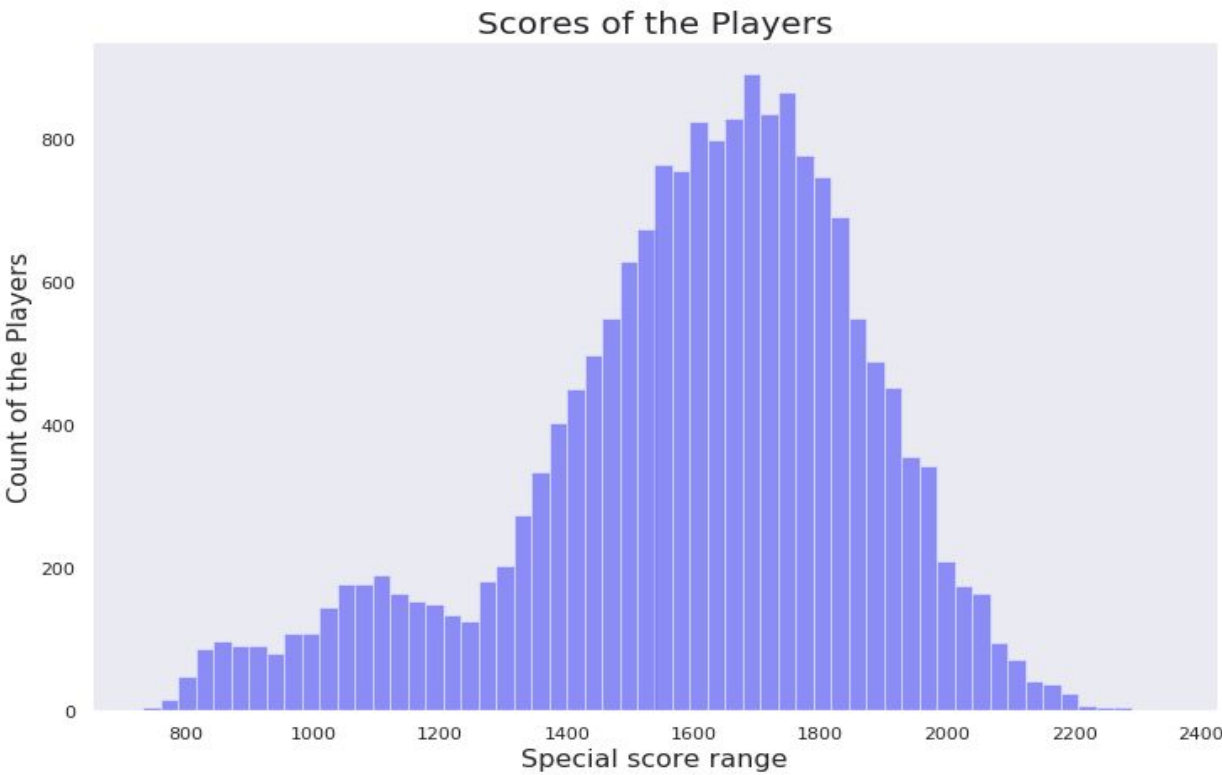
8. weight of the players is well spread across the sample data set as the majority players fall the weight range of 152lbs – 187lbs :



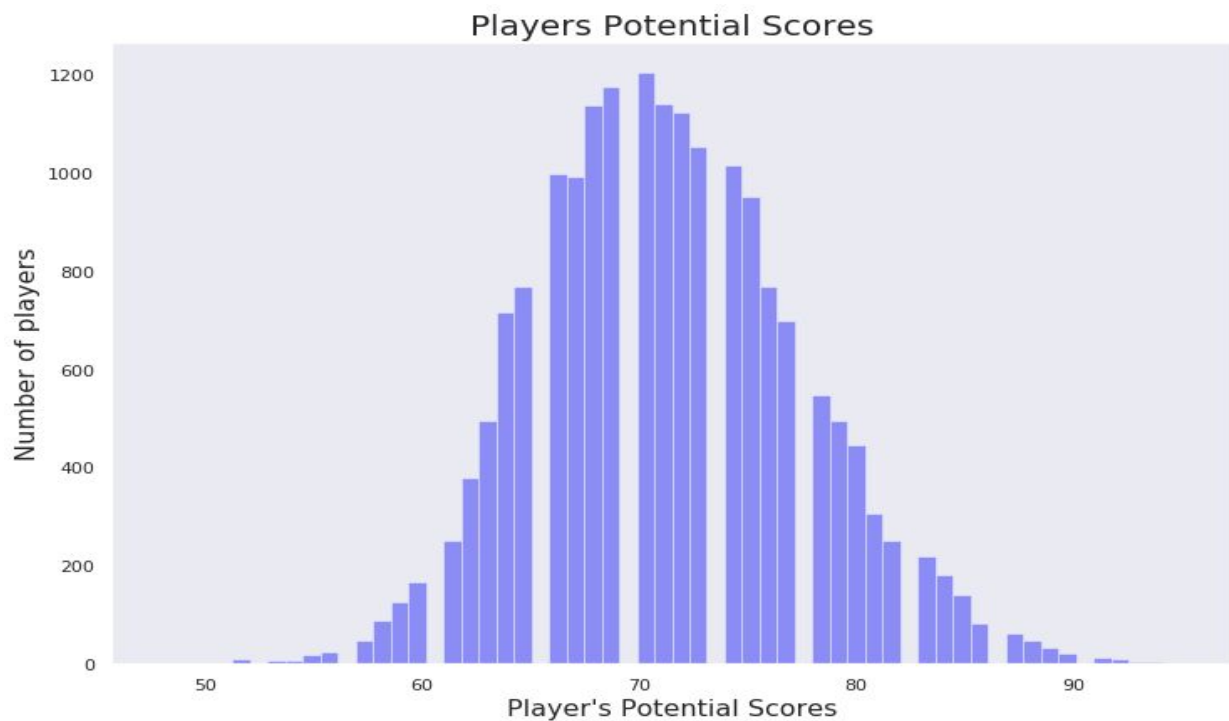
9. The Work rate data shows the nature of the game the player plays (defensive/attack):



10. the below bars shows the score spread of the Players:



11. when we compare the above chart bar chart with the potential score chart we can see that the potential scoring fits the overall scores by 90%:



12. if we plot the nationality data we can see the majority of the players comes from England, Germany, Spain, Argentina & France which makes the top 5 countries having most numbers of football players.

England	1662
Germany	1198
Spain	1072
Argentina	937
France	914
Brazil	827
Italy	702
Colombia	618
Japan	478
Netherlands	453

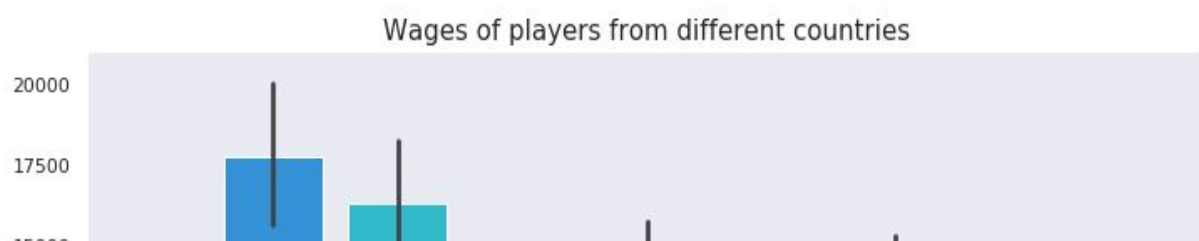
13. Top 15 players per each position based on their overall scores

	Position	Name	Age	Club	Nationality
17	CAM	A. Griezmann	27	Atlético Madrid	France
12	CB	D. Godín	32	Atlético Madrid	Uruguay
20	CDM	Sergio Busquets	29	FC Barcelona	Spain

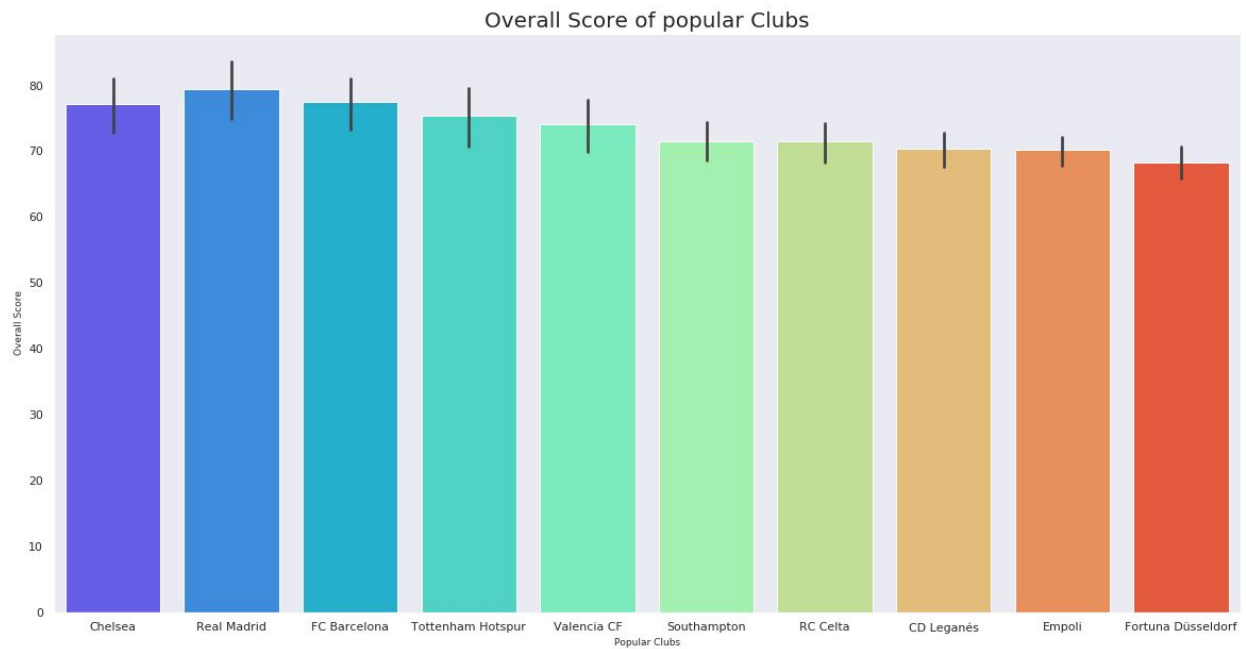
14. top 15 players from each position based on heir potential scores

	Position	Name	Age	Club	Nationality
31	CAM	C. Eriksen	26	Tottenham Hotspur	Denmark
42	CB	S. Umtiti	24	FC Barcelona	France
27	CDM	Casemiro	26	Real Madrid	Brazil
350	CF	A. Milik	24	Napoli	Poland
78	CM	S. Milinković-Savić	23	Lazio	Serbia
3	GK	De Gea	27	Manchester United	Spain
28	LAM	J. Rodríguez	26	FC Bayern München	Colombia
35	LB	Marcelo	30	Real Madrid	Brazil
77	LCB	M. Škriniar	23	Inter	Slovakia
11	LCM	T. Kroos	28	Real Madrid	Germany
14	LDM	N. Kanté	27	Chelsea	France
15	LF	P. Dybala	24	Juventus	Argentina
415	LM	H. Aouar	20	Olympique Lyonnais	France
21	LS	E. Cavani	31	Paris Saint-Germain	Uruguay
2	LW	Neymar Jr	26	Paris Saint-Germain	Brazil

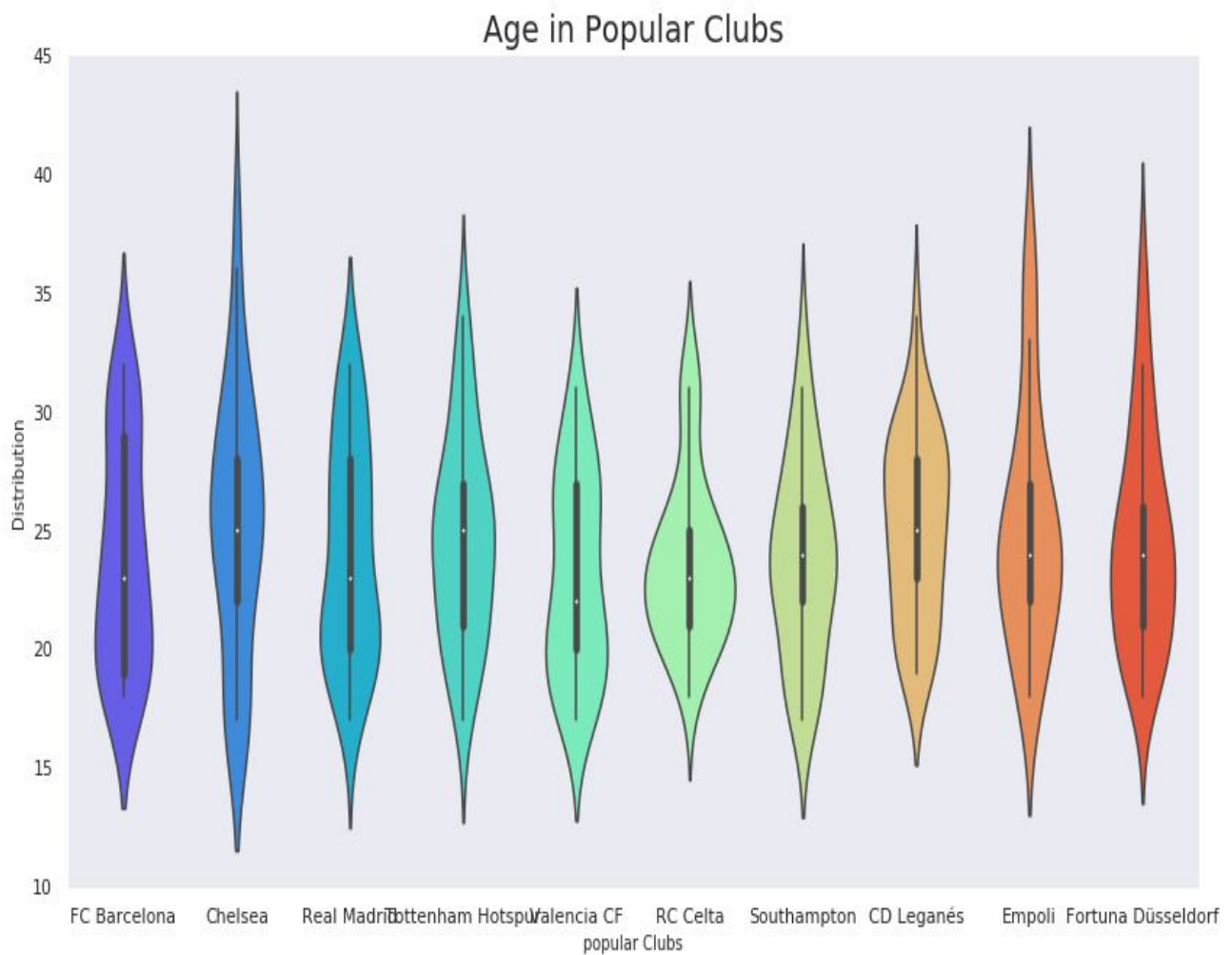
15. When we plot the wages data w.r.t to the nations we get to see that:



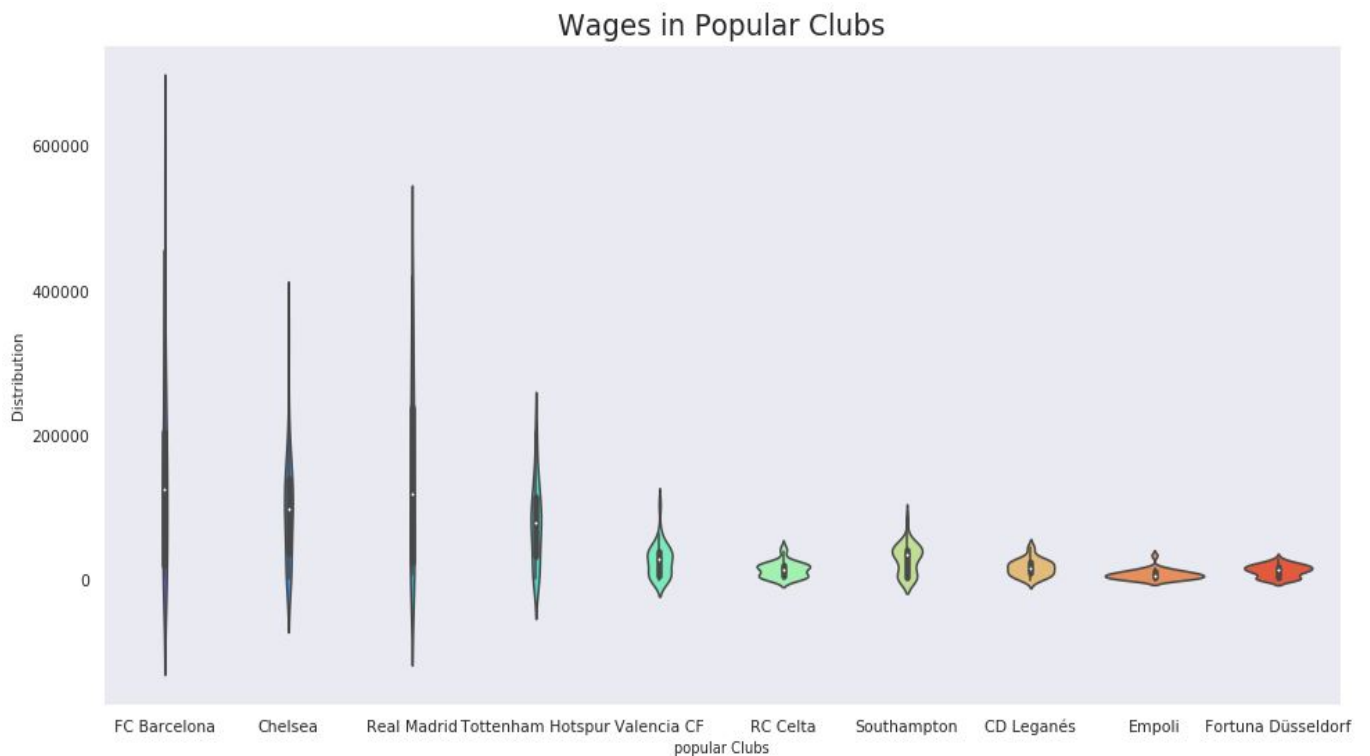
16. if we populate the graph for the clubs w.r.t the scores we :



17. If we graph the Age data of every club we could see:



18. Wages that Clubs pays to their players has a steep spread in the graph:



19. 10 Youngest players and their details:

	Name	Age	Club	Nationality
18206	G. Nugent	16	Tranmere Rovers	England
17743	J. Olstad	16	Sarpsborg 08 FF	Norway
13293	H. Massengo	16	AS Monaco	France
16081	J. Italiano	16	Perth Glory	Australia
18166	N. Ayéva	16	Örebro SK	Sweden
17712	K. Broda	16	Wisła Kraków	Poland
17354	L. D'Arrigo	16	Adelaide United	Australia
13567	Y. Verschaeren	16	RSC Anderlecht	Belgium
16927	B. Nygren	16	IFK Göteborg	Sweden
18018	B. O'Gorman	16	Bray Wanderers	Republic of Ireland

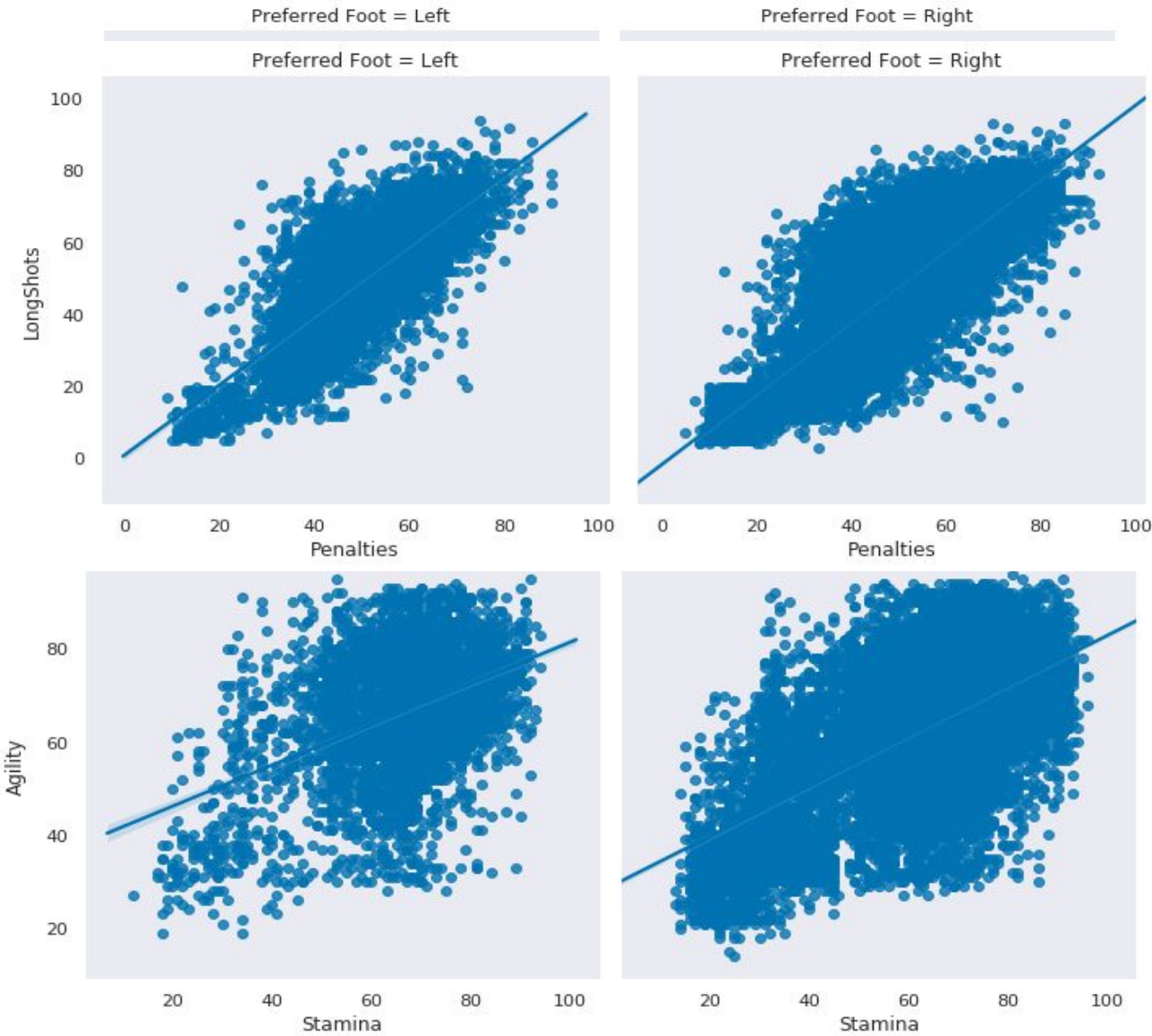
20. 10 Eldest players and their details:

	Name	Age	Club	Nationality
4741	O. Pérez	45	Pachuca	Mexico
18183	K. Pilkington	44	Cambridge United	England
17726	T. Warner	44	Accrington Stanley	Trinidad & Tobago
10545	S. Narazaki	42	Nagoya Grampus	Japan
7225	C. Muñoz	41	CD Universidad de Concepción	Argentina
1120	J. Villar	41	No Club	Paraguay
12192	H. Sulaimani	41	Ohod Club	Saudi Arabia
15426	M. Tyler	41	Peterborough United	England
4228	B. Nivet	41	ESTAC Troyes	France
10356	F. Kippe	40	Lillestrøm SK	Norway

21. Longest club members:

Club member scene		
Name		
O. Pérez	Pachuca	28
M. Al Shalhoub	Al Hilal	21

22. Comparison of skills for the Left-footed and Right footed Players:



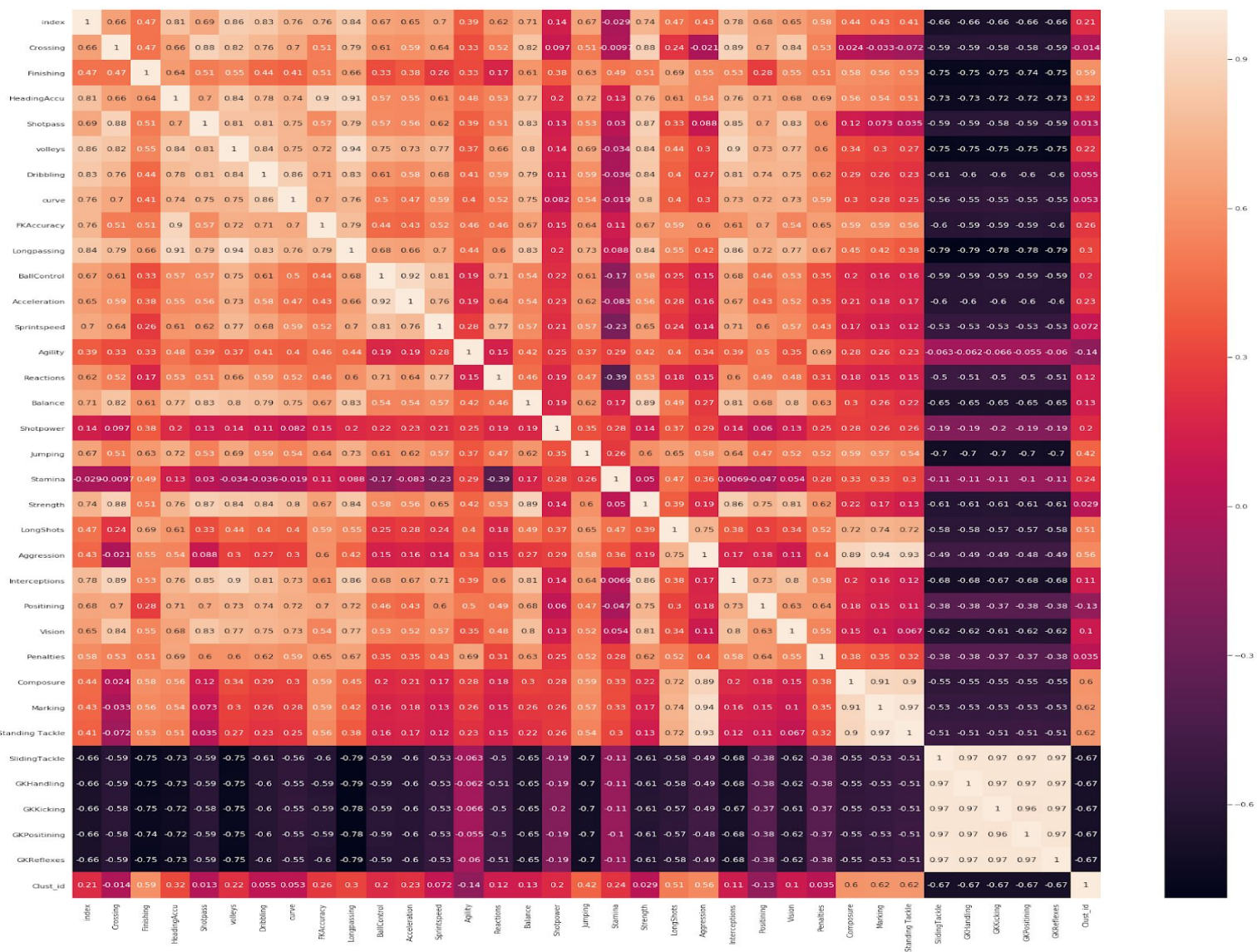
Segmentation of players

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. Here in the dataset, the features that are used to cluster the players are those features that play an important role in the ranking the players.

Features follow are used for clustering the players:

1. Balancing
2. Stamina
3. Aggression
4. Penalties
5. Strength
6. Long Passing
7. Vision
8. Interception
9. Heading Accuracy

The correlation heat map below shows the features' correlation with each other:



looking into the heat map below is the clusters that players can be divided into keeping in mind the number of clusters is 3.