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)

DBS Dublin Business School
excellence through learning

Dataset:

Source: Kaggle**(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 shots	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.

Contributors: The assignment was a joint effort wherein all the contibutors provided inputs on each insight and phase. To highlight the domain focuss of the members:

1. Ajit Kumar Shukla (Student No: 10394108) - Data Preparation(including Tableau Visualisation) and Visualisation in PowerBI, Python
2. Damanpreet Kahlon (Student No: 10505907) - Data Visualisation in Python, Tableau dashboard and Report Creation
3. Reema Mascarenhas (Student No: 10402498) - Data Visualisation in Python
4. Ashin Basheer (Student No: 10503646) - Data Visualisation in Python and PowerBI, GitHub Pages Setup

Deployment

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

Interactive Altair Visualizations Deployed at :

https://fifa-players.github.io/Fifa/

Colab Link For Notebook

`[![Google Colab](https://badgen.net/badge/Launch/on%20Google%20Colab/blue?icon=terminal)](https://colab.research.google.com/drive/1v1tkelpBjU33meg5KTH45WJF5kSFvg4S#scrollTo=VcyqzOee05-A)`

Power BI Dashboards:

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

Tableau Dashboard:

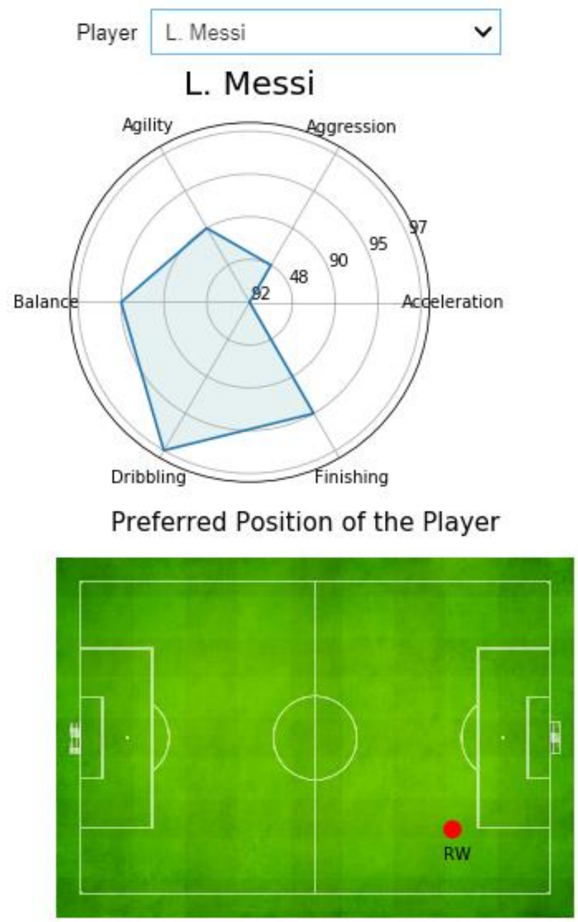
(https://public.tableau.com/profile/ajit.kumar.shukla#!/vizhome/DV_CA02/ClubwiseAnalysis?publish=yes)

Please note, few of the visualizations were not compatible with Colab (like D3) so please check the python notebook for the same.

Player preferred position with other attributes:

This dashboard was created using Altair, Matplotlib, Ipywidgets libraries in Python. The drop down gives the option of selecting players. Their preferred position is being reflected on the football field. Along with that, the radar chart gives the values of various player attributes. This is an interactive and unique way of comparing player performance.

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. Lets select these two players for our comparison.



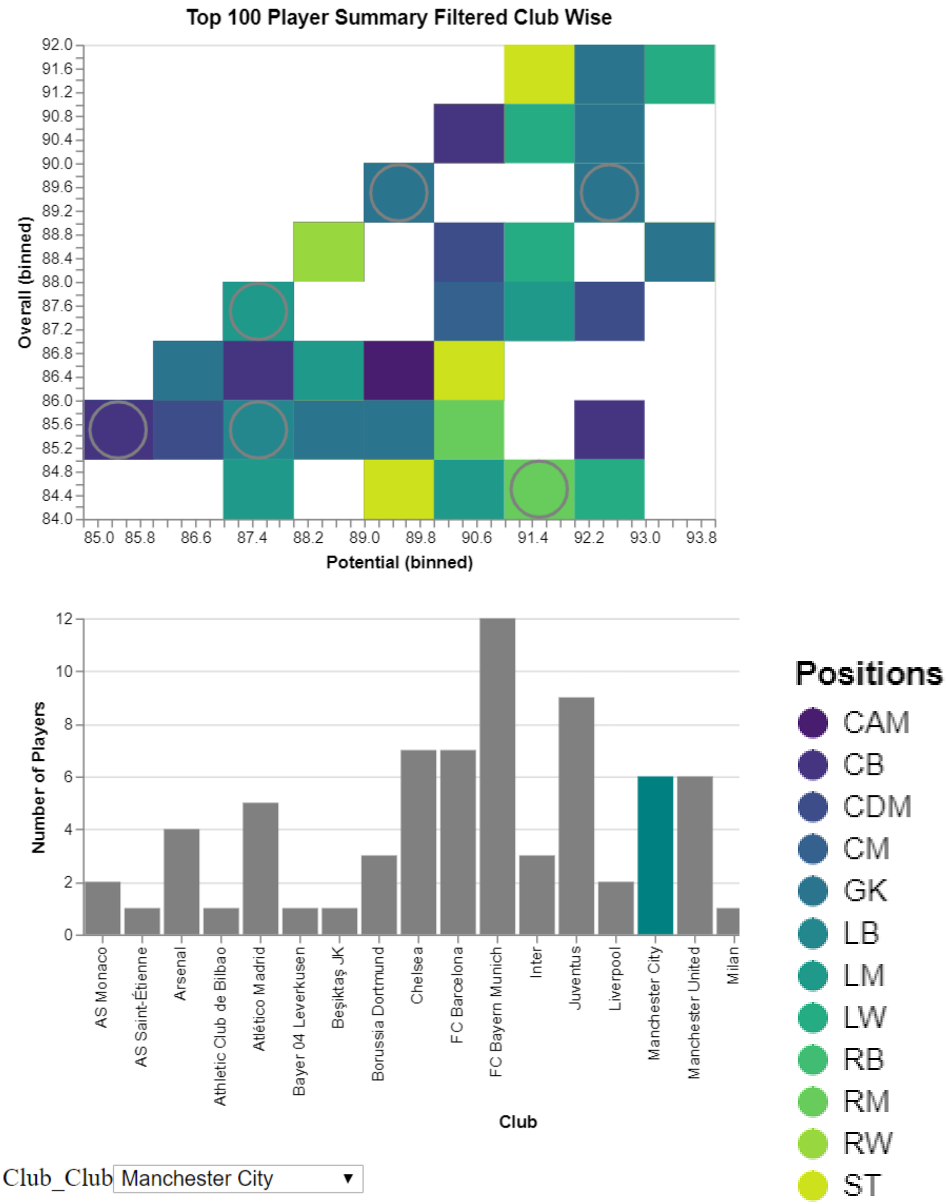
There isn’t much difference in the attributes of both players, although Ronaldo is slightly more aggressive than Messi.

Club wise player analysis for top 100 players

This dashboard was created using Altair with Python. It gives the users an option to select the clubs from the drop down below. On selecting a particular club we get the various attributes of the players that were identified as top 100. This was based on the attribute **Overall** of the players. This visualization can be used to compare the various players of different clubs.

Let’s compare for Chelsea and Manchester City:

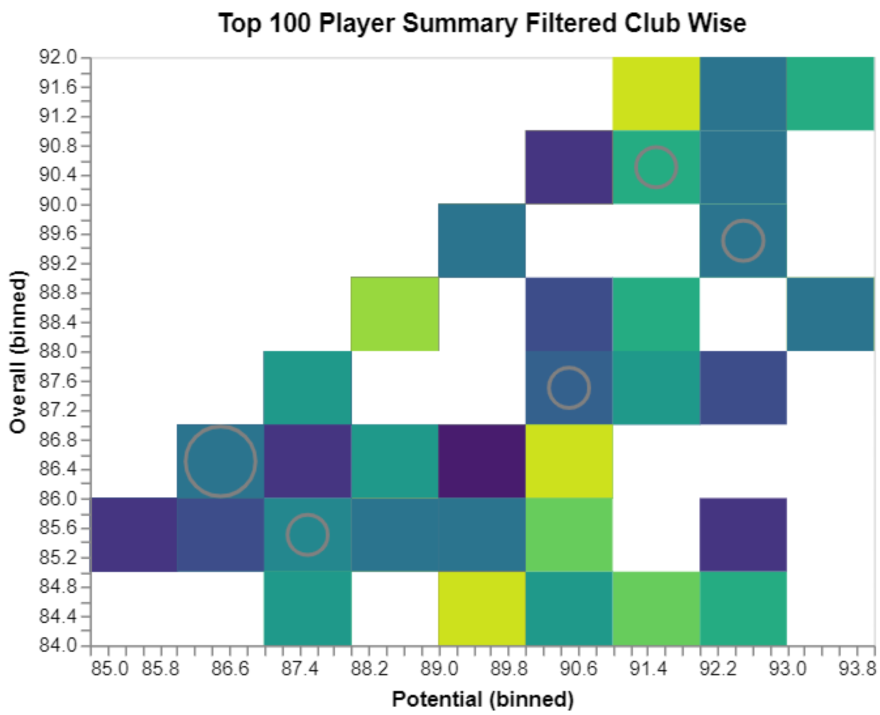
On selecting Manchester City we get 6 faint circles on the Potential vs Overall.

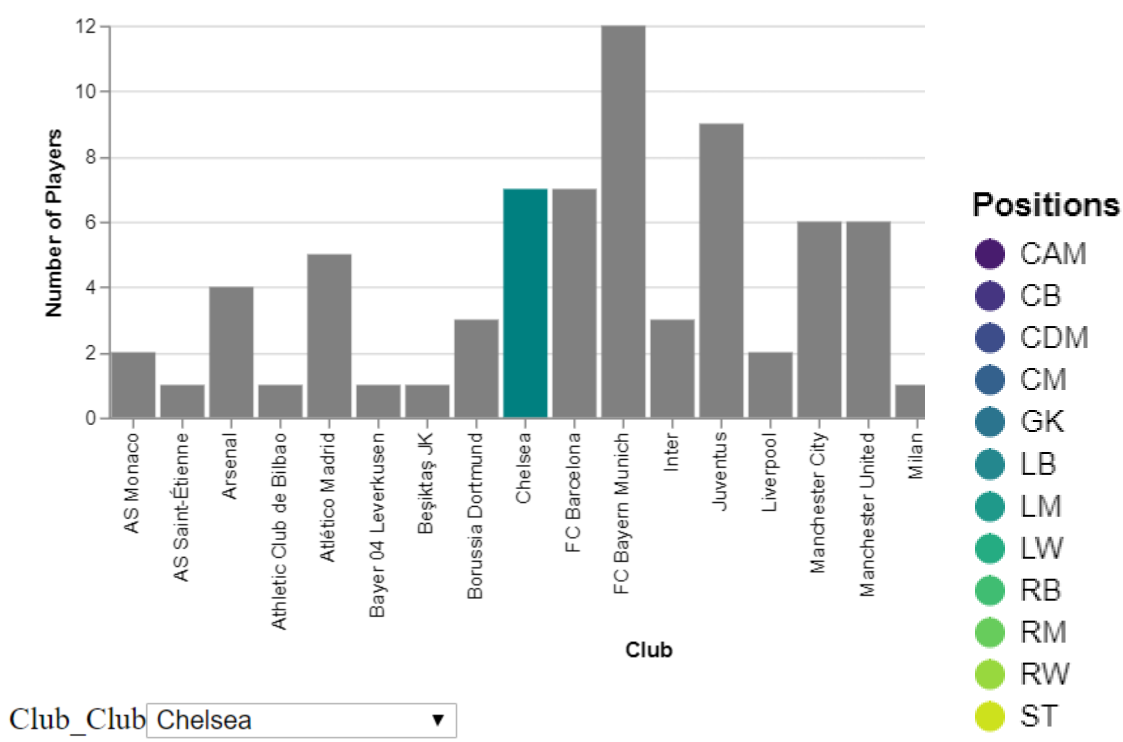


These circles represent the players of the clubs aggregated by the preferred positions.

Size of players indicate the number of players in that group. Two player groups for the goal keeping position have potential and overall more than 85.

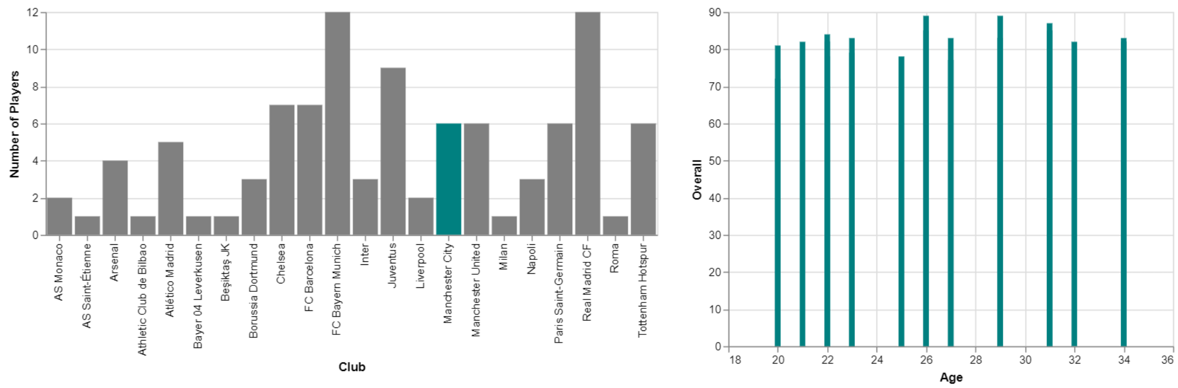
In case of Chelsea:



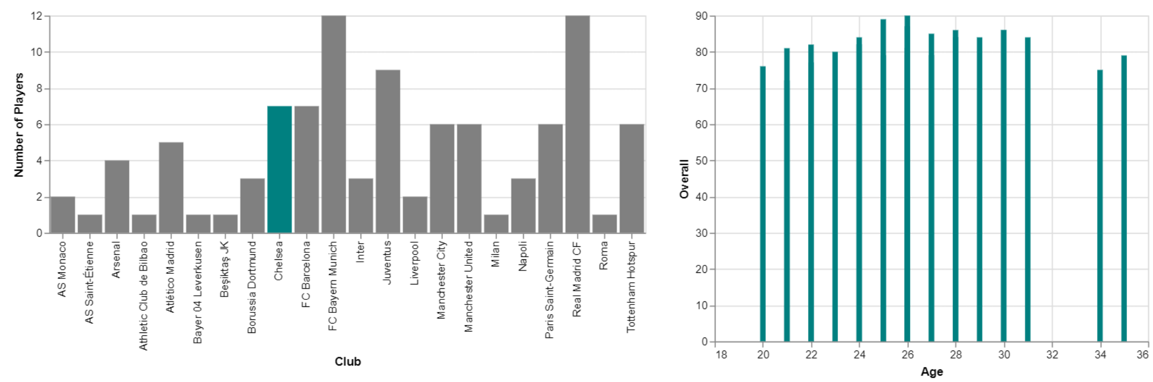


There are 3 player groups for Chelsea, for positions midfielders and running back. The number of players in each group is less for Chelsea compared to Manchester City.

Age of Players for Manchester City and Chelsea comparison.



Club_Club Manchester City ▼

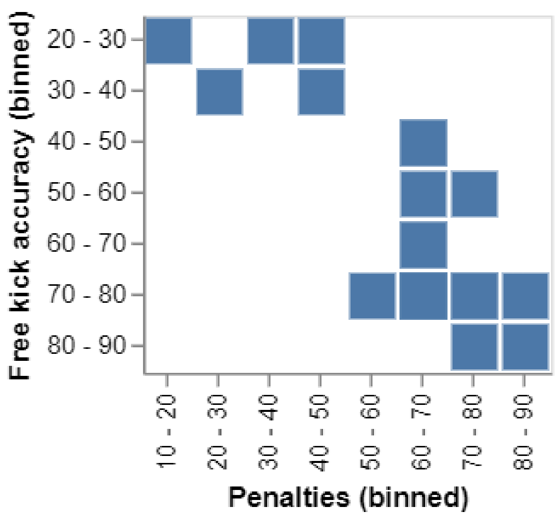


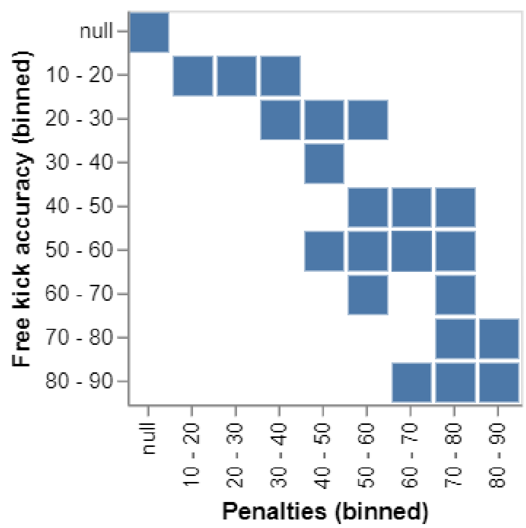
Club_Club Chelsea ▼

This plot helps analyze the age group of the players of different clubs.

Free kick accuracy vs Penalties:

Manchester City (Left) and Chelsea(Right)



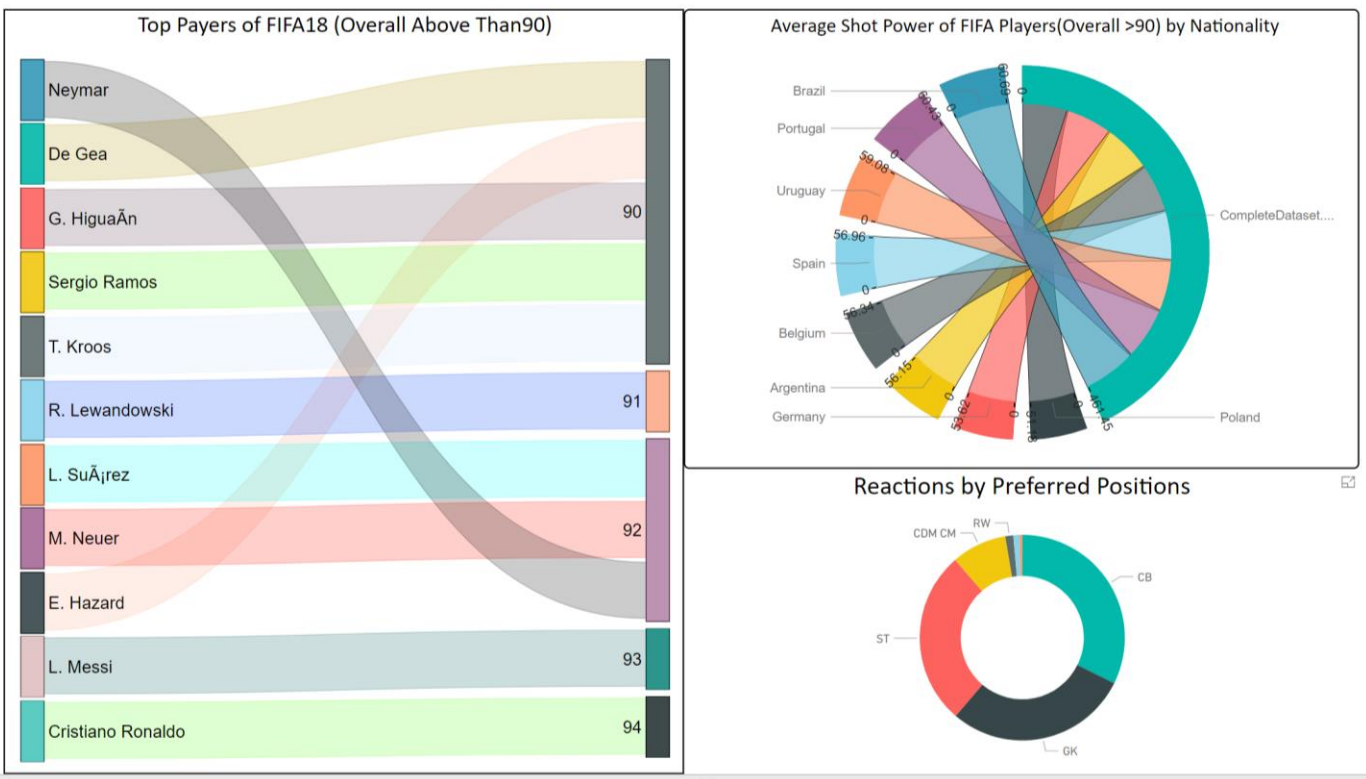


Chelsea clearly has higher number of players in the top 100 and so more number of players that have higher penalties. The free kick accuracy of players is more diverse compared to Manchester City.

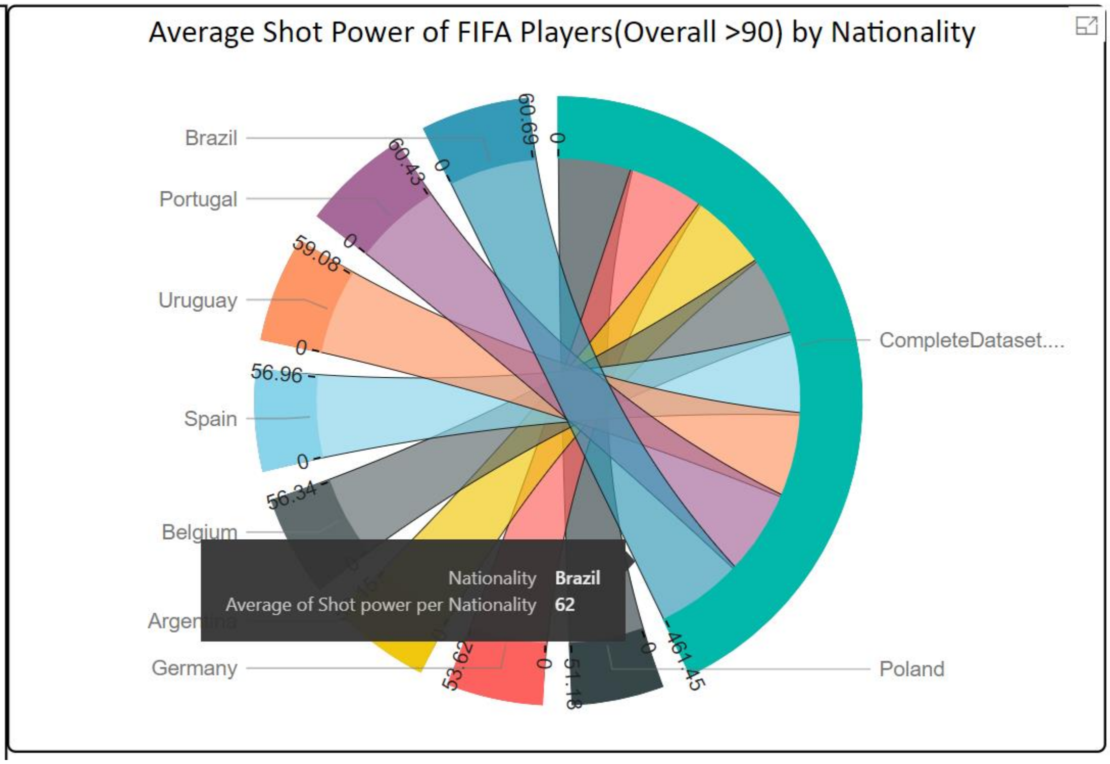
We have also created a word cloud that represents top 100 players:

Player based analysis

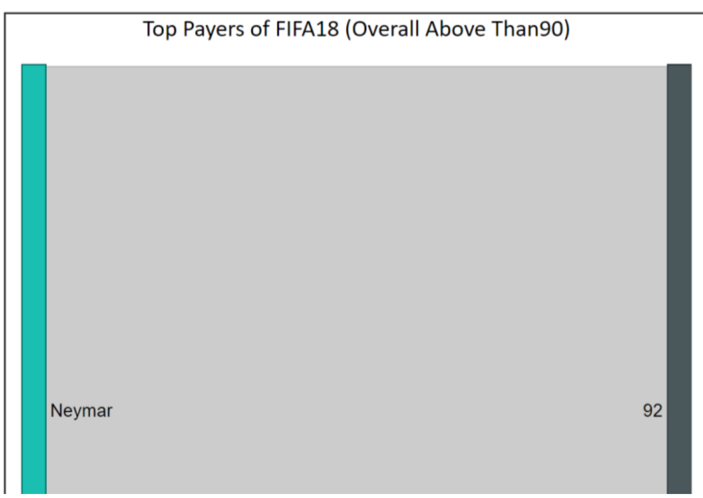
Players having an overall score greater than 90 were identified as top players. In order to further analyze their performance we selected the tool Power BI and created an interactive dashboard. The Sankey diagram on the left displays a list of these identified players along with their overall scores.

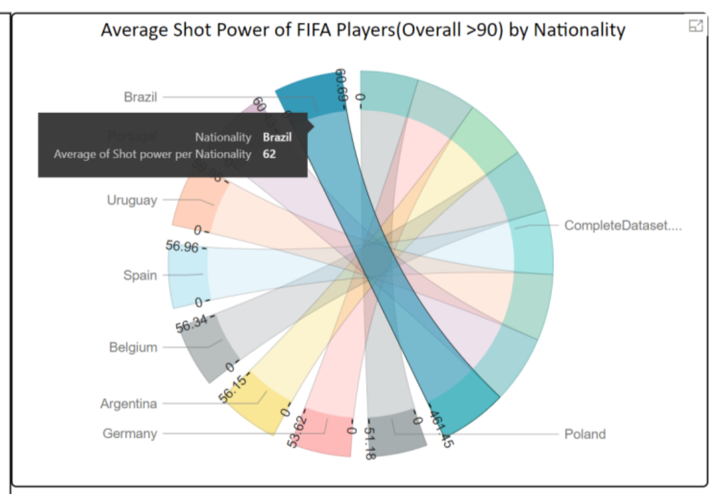


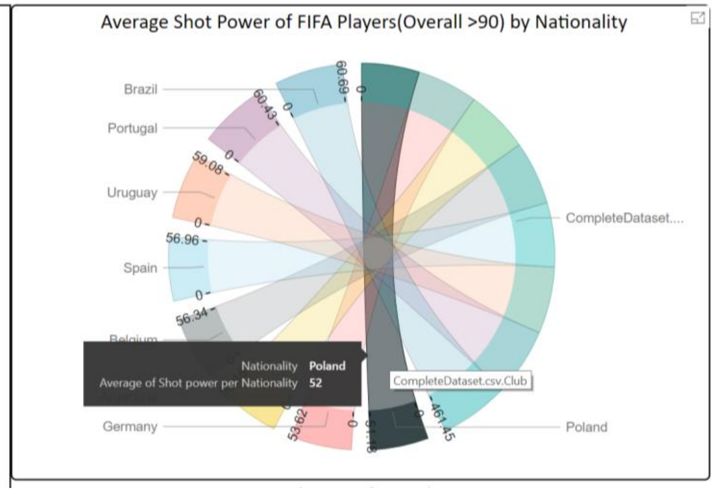
The chord diagram highlights the Shot Power which is essentially the shooting power of these players and highlights their Nationality. This chart is created in comparison with the entire dataset.



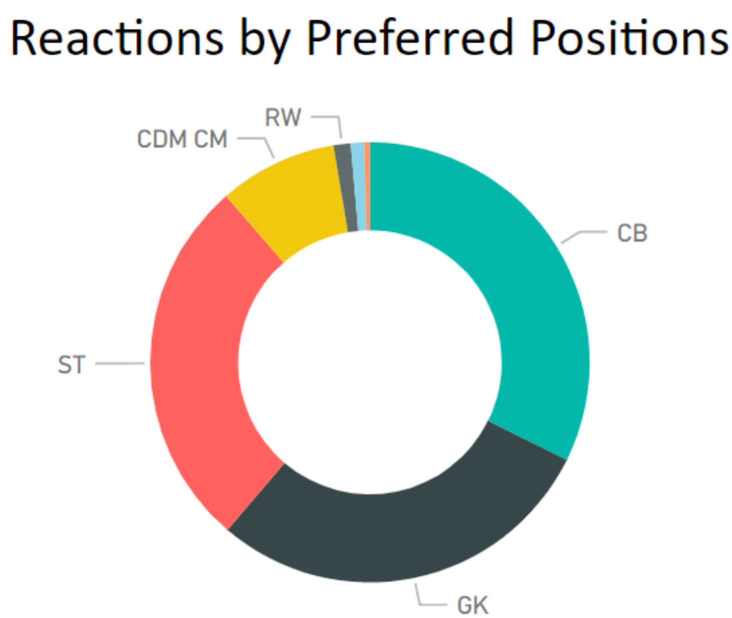
Since these are top players, the difference in shot power is minimal. Neymar from Brazil has the highest shot power of 62 compared to R. Lewandowski from Poland with a shot power of 52. Even among the top players we can deduce that there is a huge gap in the shooting power of these players.



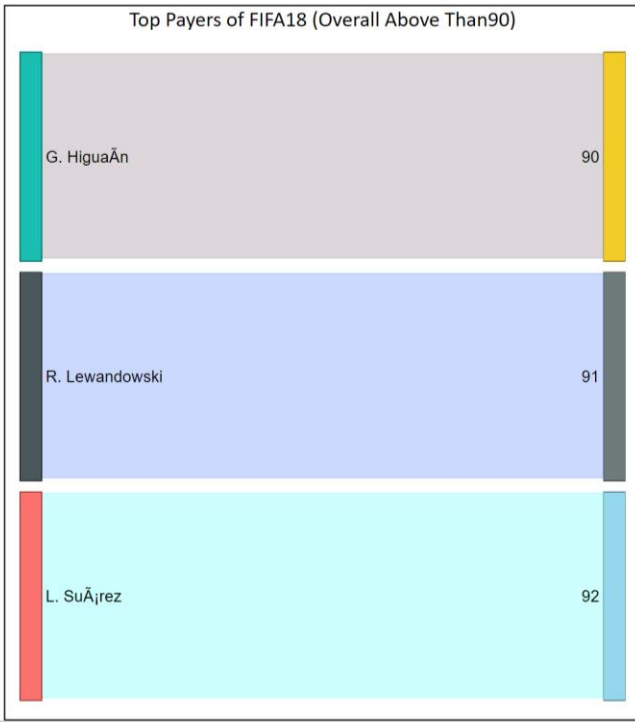


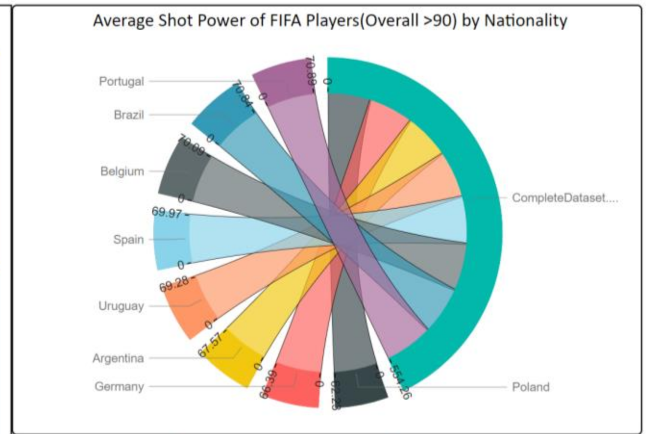


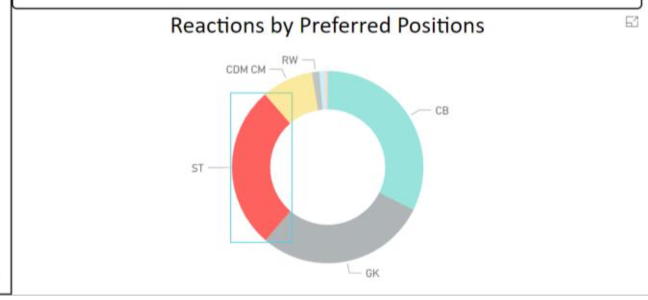
Along with the shooting power, we have also highlighted the preferred positions of these players. These preferred positions are highlighted with the percentage of Reaction. Reaction attribute determines the presence of mind of the players on the field which is a crucial criteria in a competitive sport 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
69.97
69.28
67.57


Reactions by Preferred Positions
RW
CDM CM
CB
ST
GK

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.

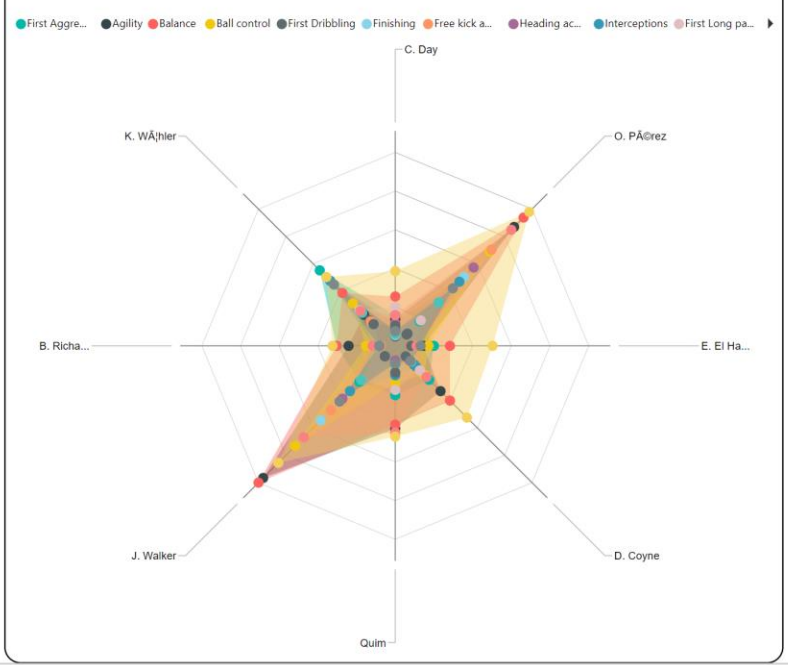
In order to achieve this we have selected Power BI to create an interactive dashboard.

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.

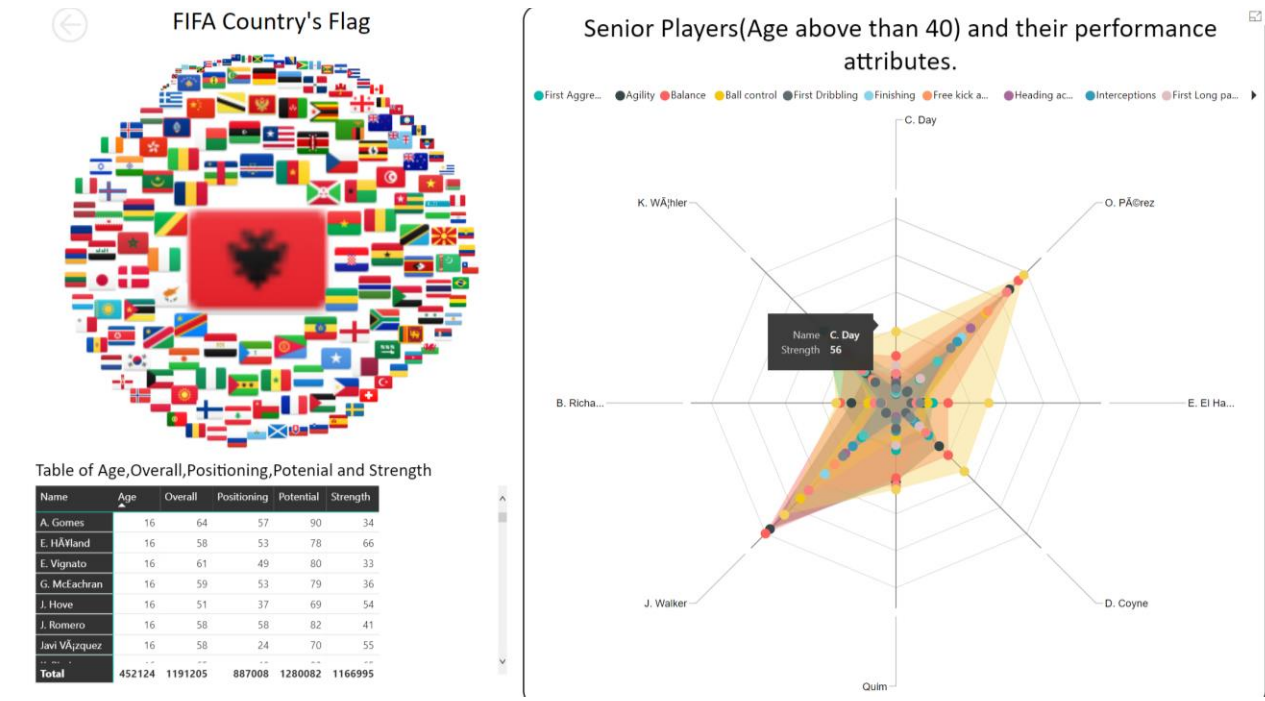


The table provides a list of Players along with their Age, Overall performance measure, their Positioning , Potential marked by their performance and Strength. On selecting an individual row/Player from the table, the flag of the players nationality is highlighted as shown below:

FIFA Country's Flag

Table of Age,Overall,Positioning,Potenial and Strength

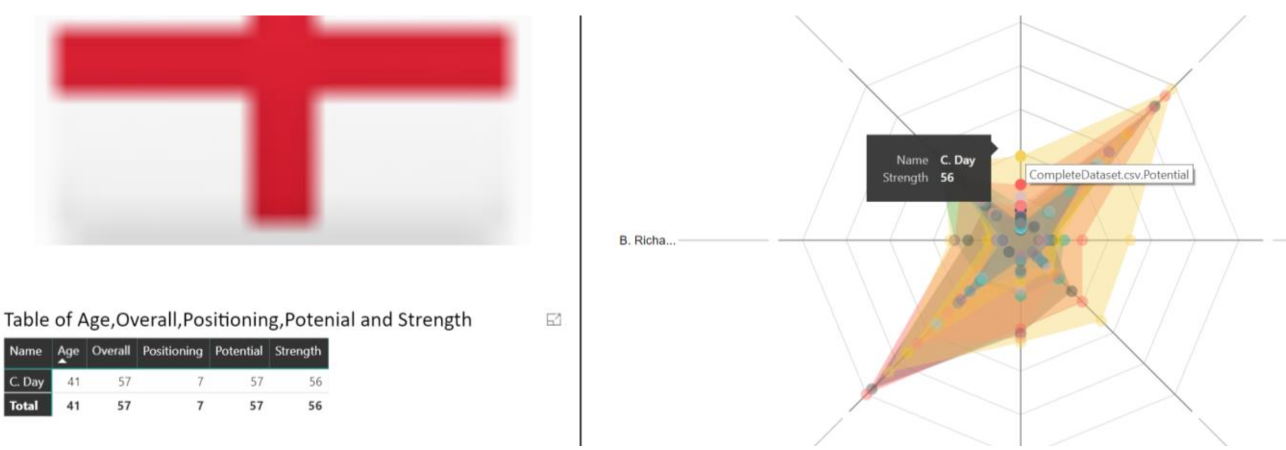
Name	Age	Overall	Positioning	Potential	Strength
A. Gomes	16	64	57	90	34
E. HÃ¥land	16	58	53	78	66
E. Vignato	16	61	49	80	33
G. McEachran	16	59	53	79	36
J. Hove	16	51	37	69	54
J. Romero	16	58	58	82	41
Javi VÃ¡zquez	16	58	24	70	55
Total	452124	1191205	887008	1280082	1166995



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

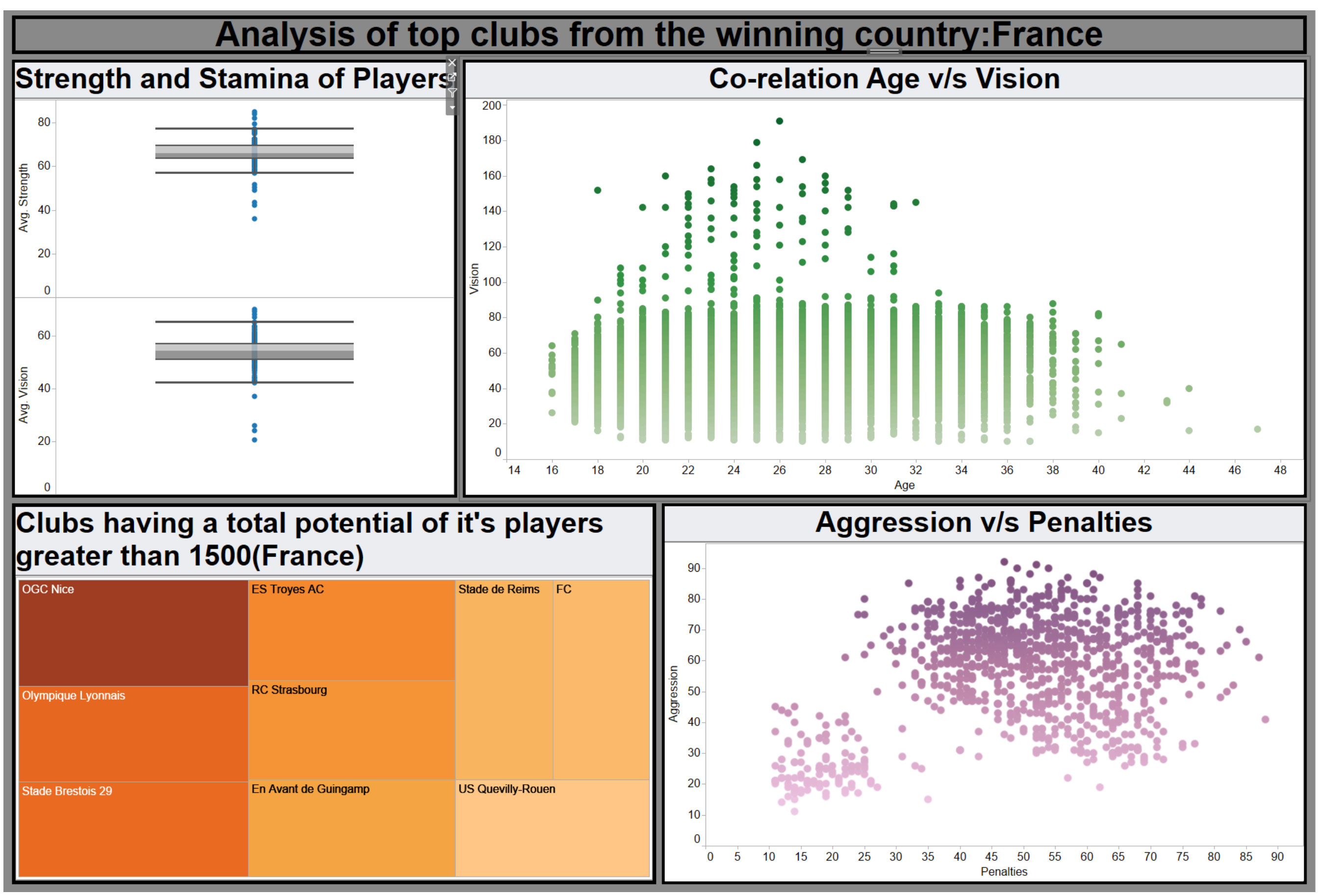


Analysis of top clubs from the winning country: France

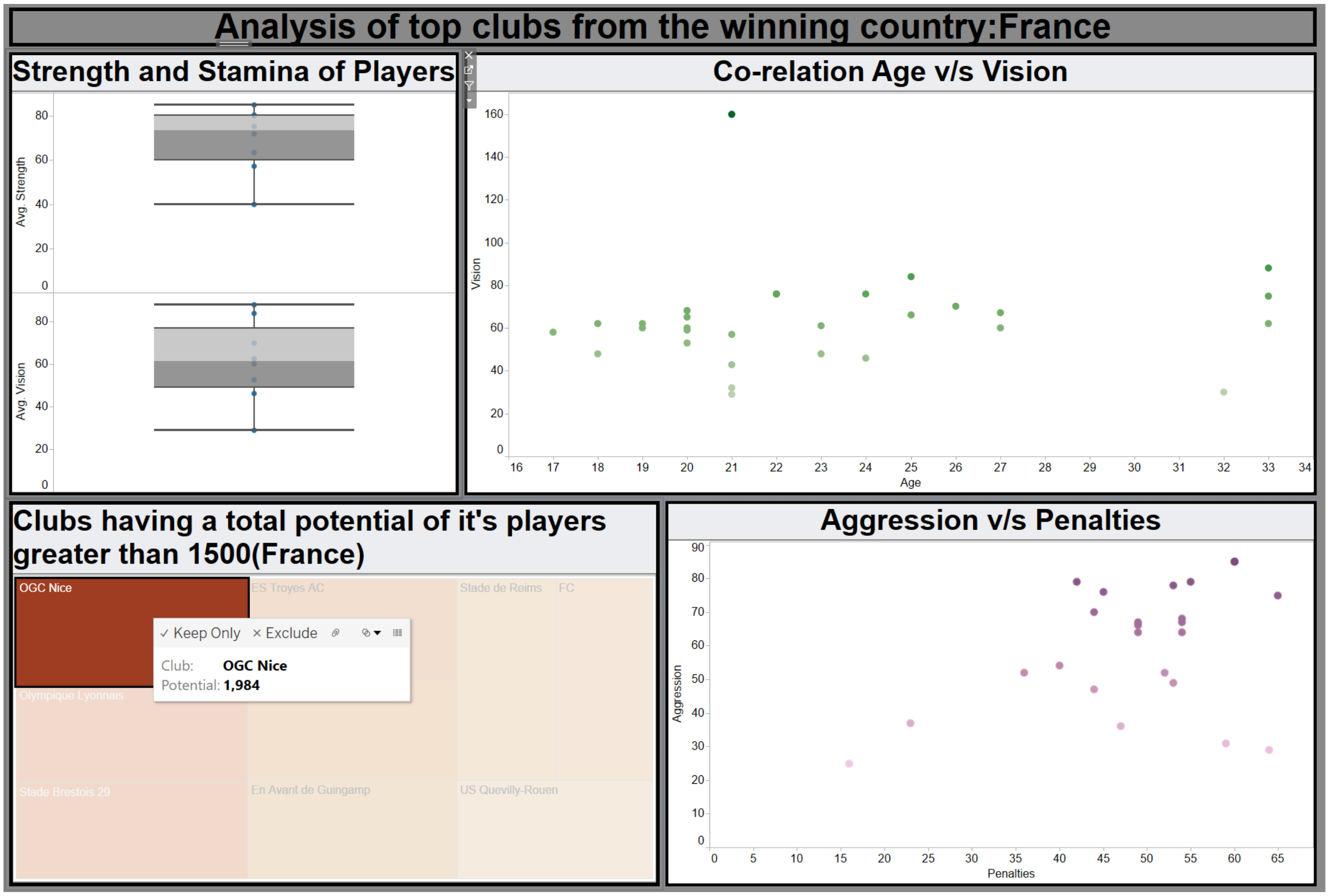
France is the winner of FIFA 2018. We identified clubs having a collective player potential greater than 1500 points from France to understand the different performance measures of their players.

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.



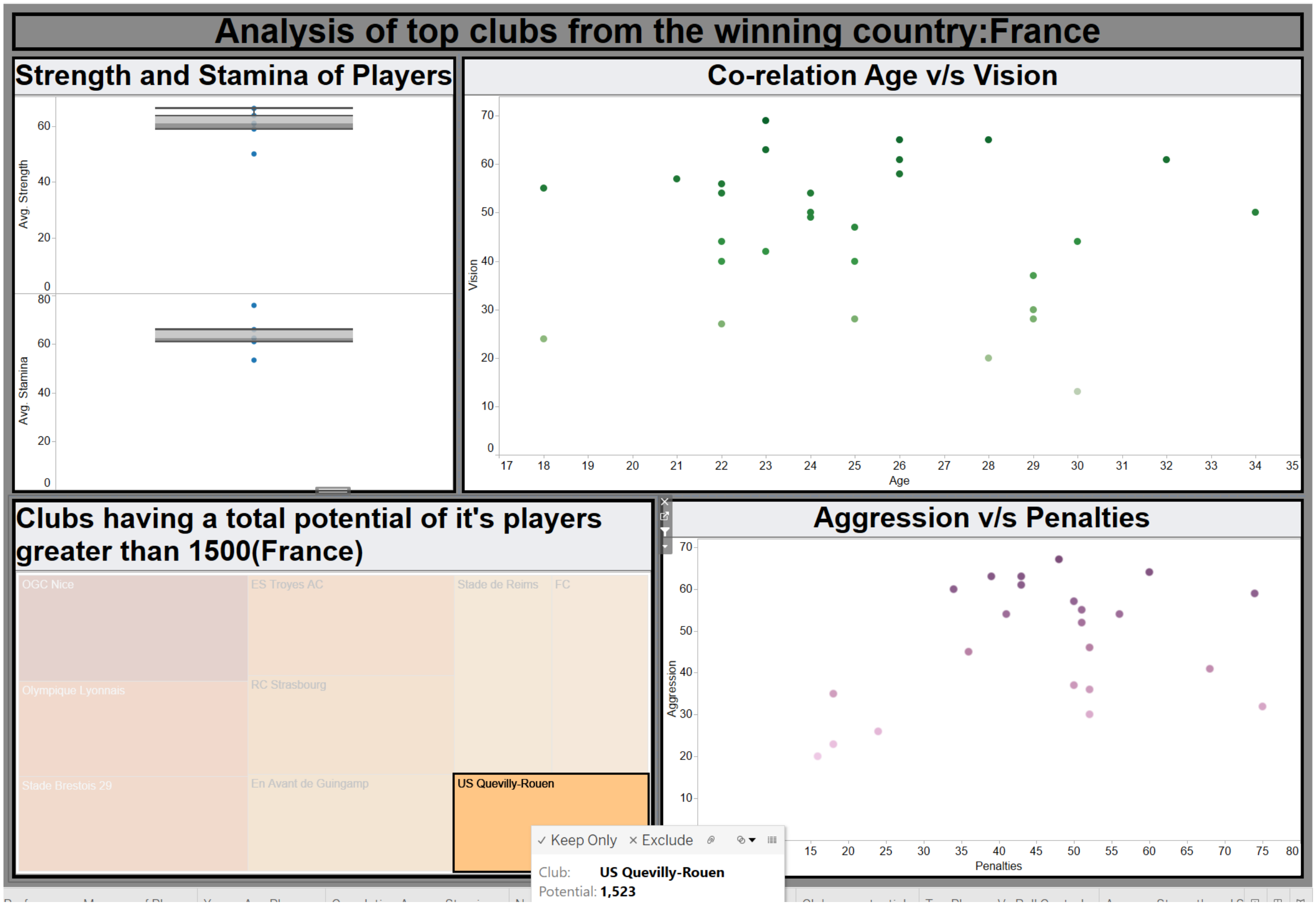
Lets filter for the best club-OGC:Nice



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.



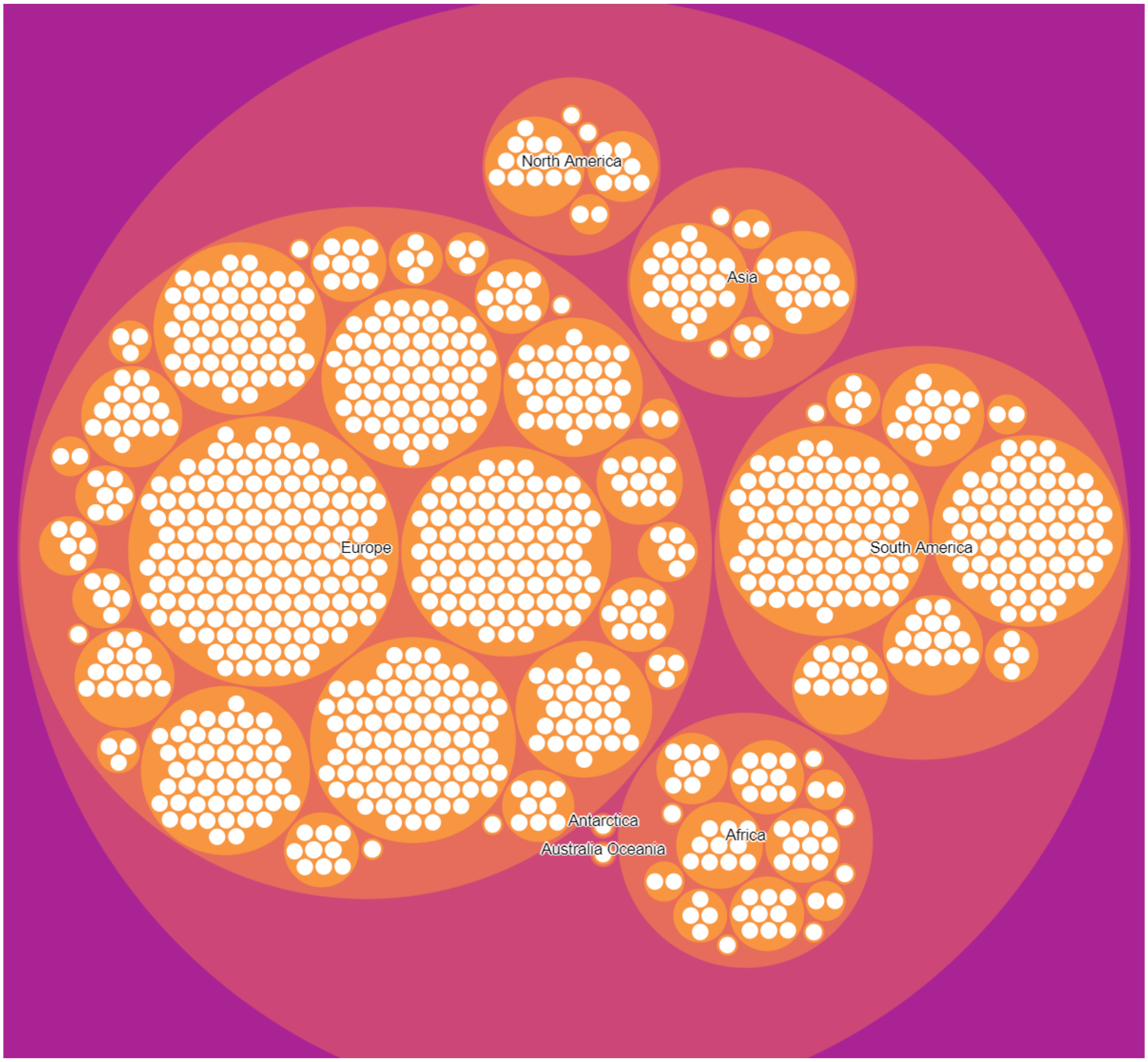
Although the relation between **vision** and **age** is somewhat similar to OGC:Nice, players of US Quevilly-Rouen have a much better vision among players of all ages as compared to the former. The strength and stamina ranges are much lower while aggregation and penalties are more or less same.

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.

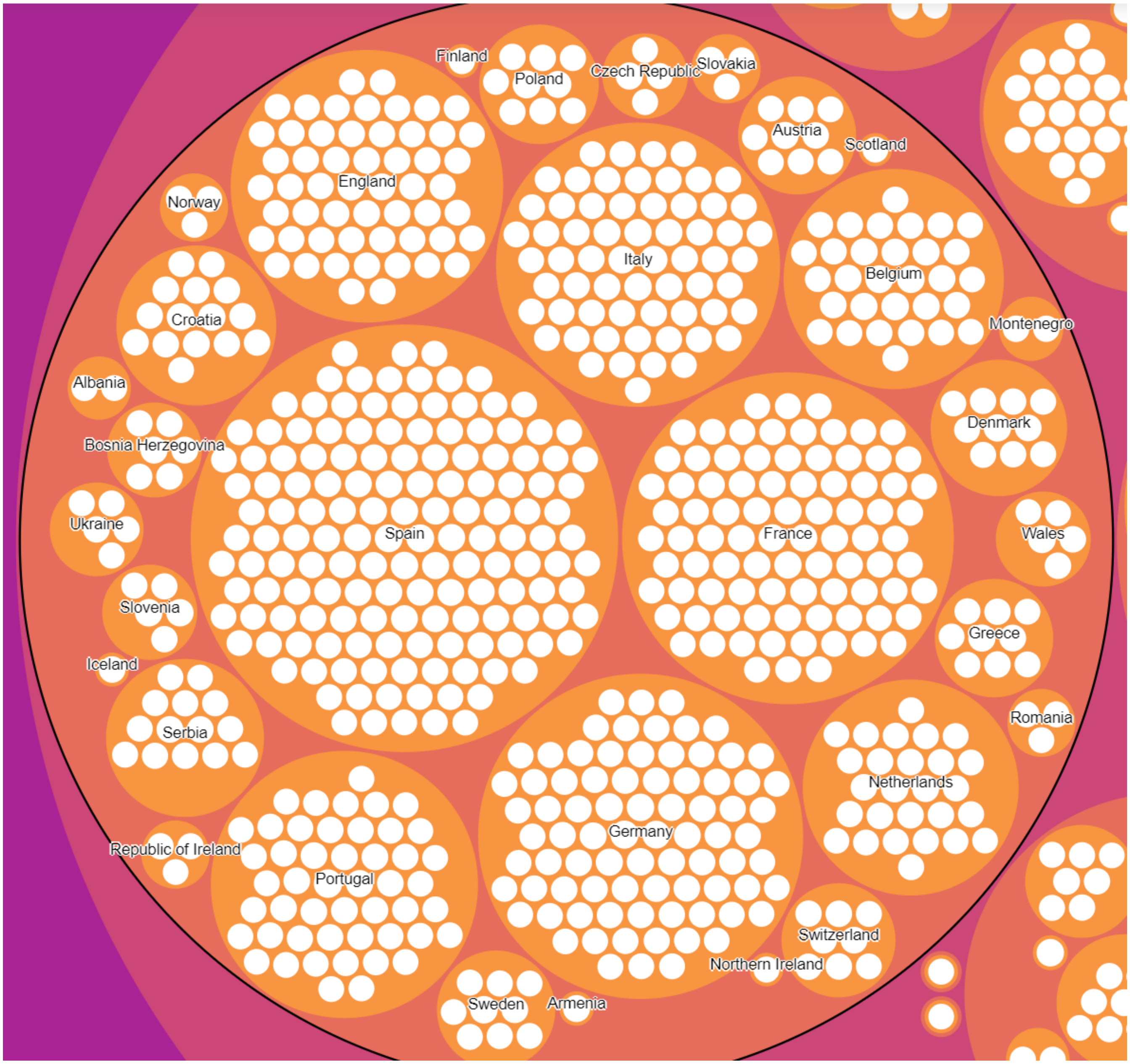
In short this visualization represents the globe and allows us to get the overall geographic location of top 2000 players.

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.



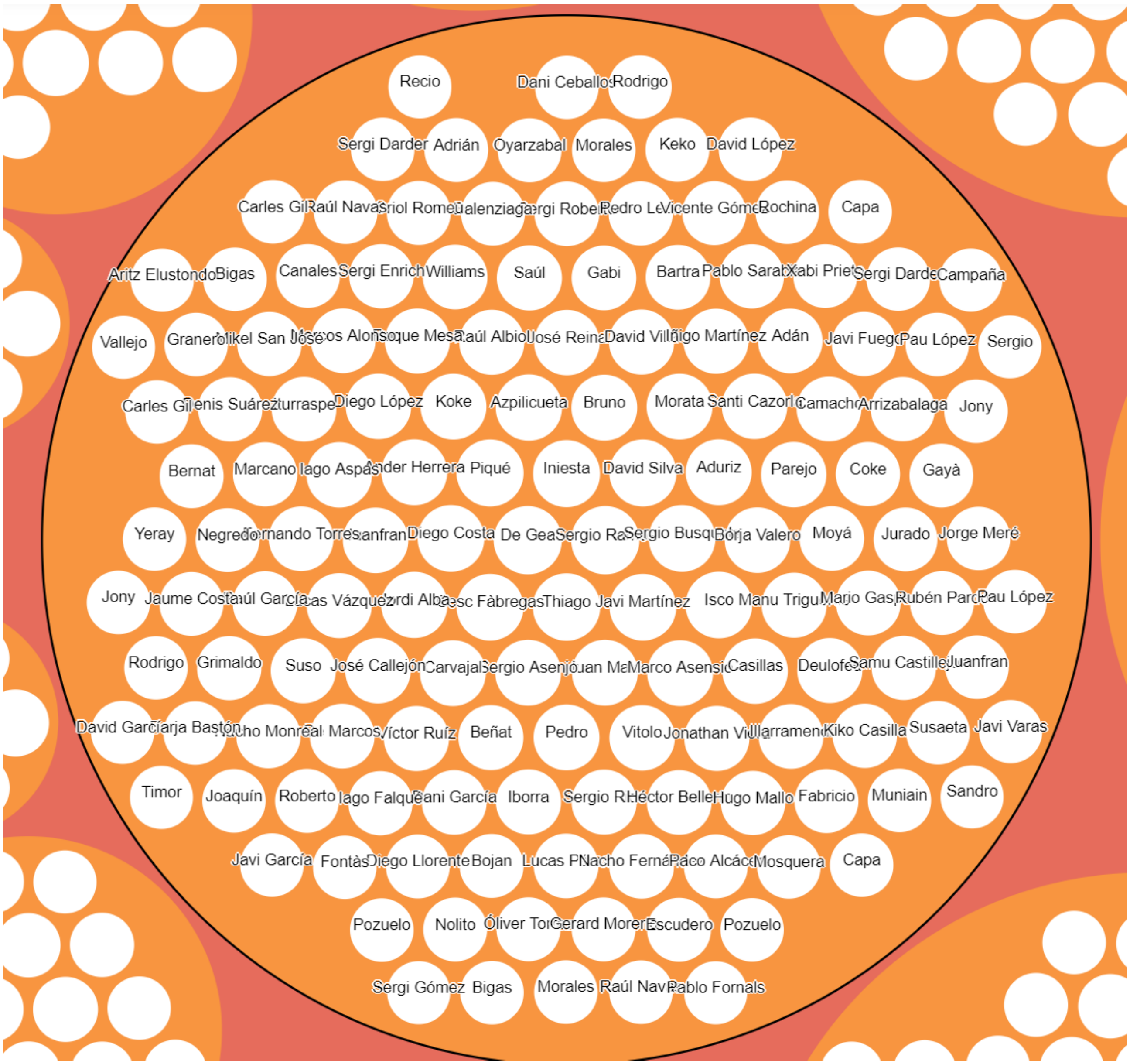
The plot shows the player density on each continent. From the plot, Europe is the leading continent with maximum of the top 2000 players. It is followed by South America, Africa, Asia , North America and then finally Australia and Antarctica being the least popular.

Lets zoom into Europe:



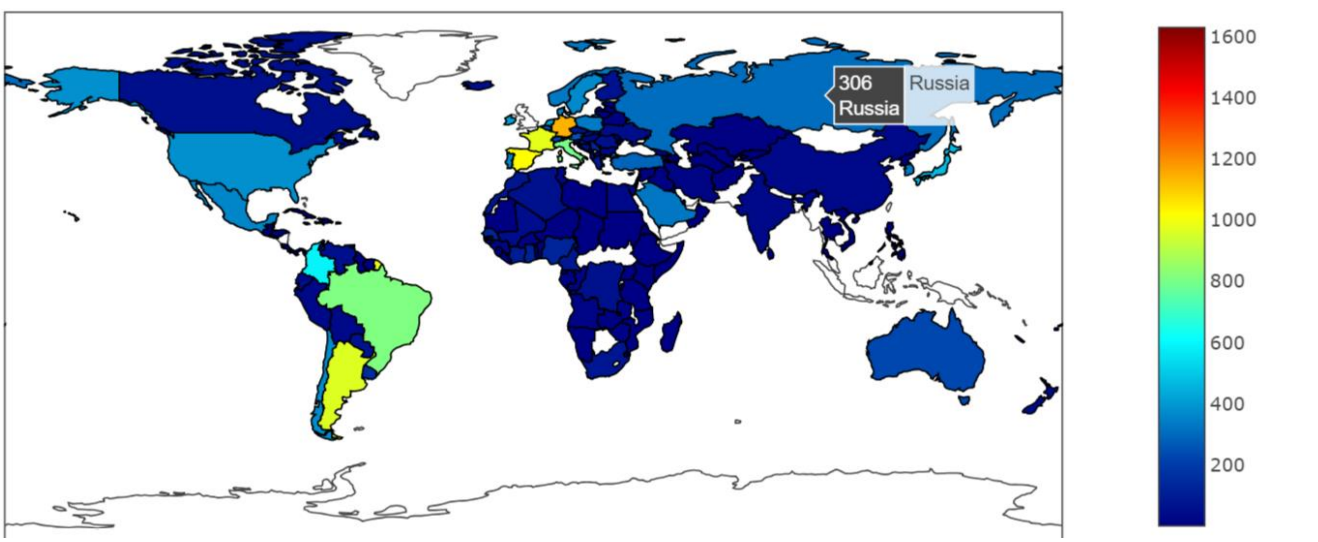
Spain, France, Germany, Italy and England have the maximum number of players. Scotland, Finland etc. having just single players each.

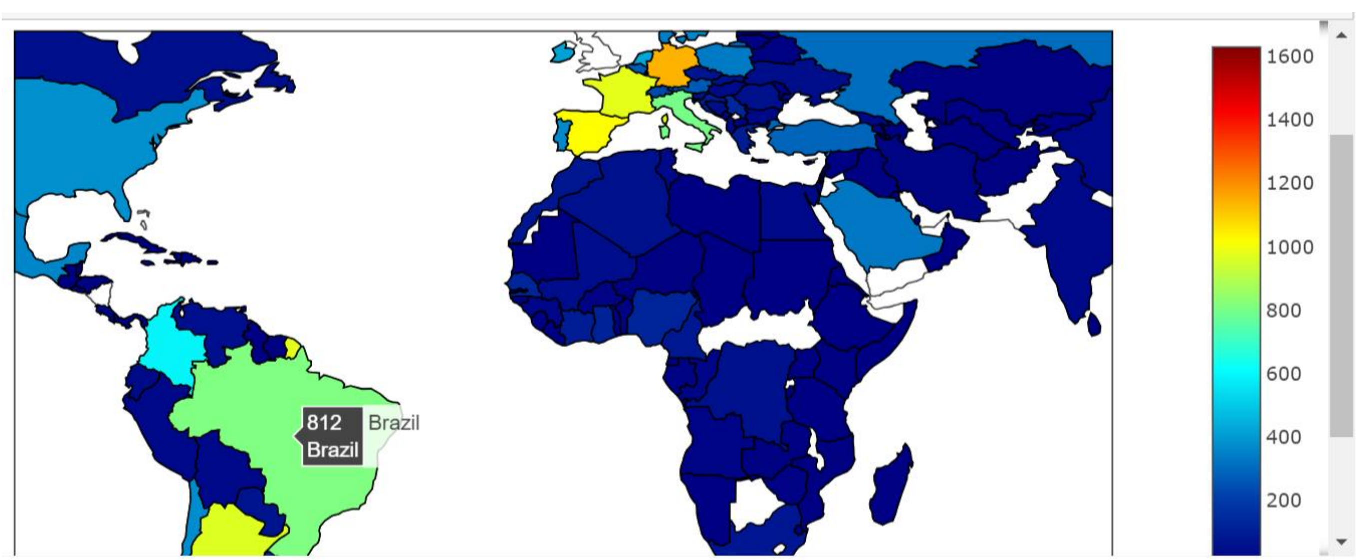
On zooming into Spain we get the names of all the top 2000 players from Spain:



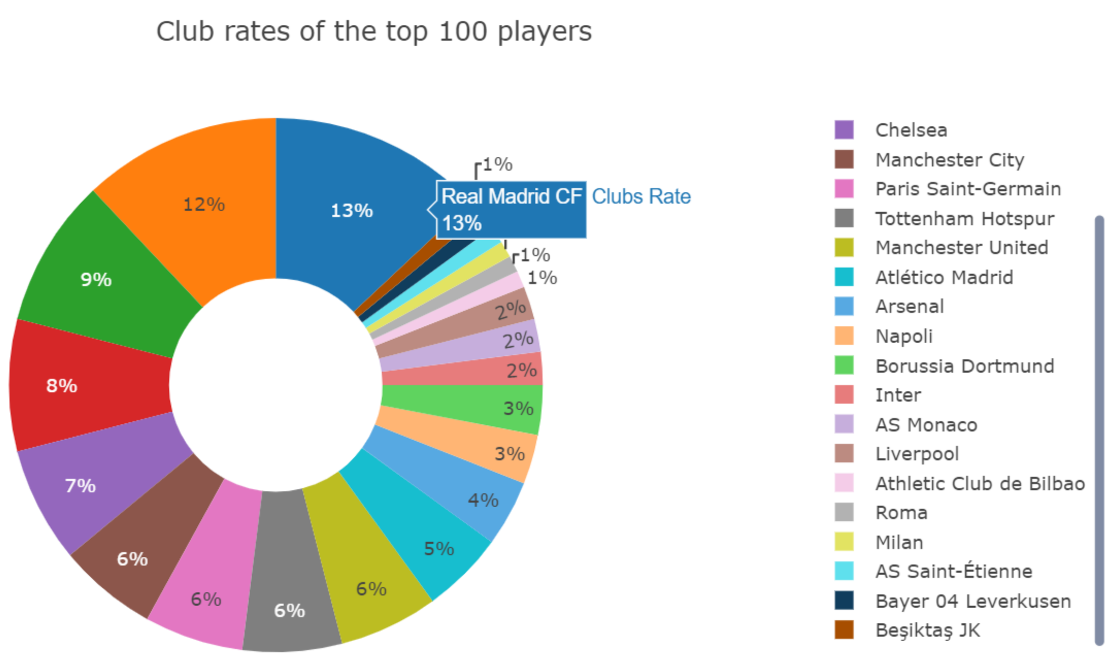
Map Distribution of the Players and How many players from the same Country?

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.

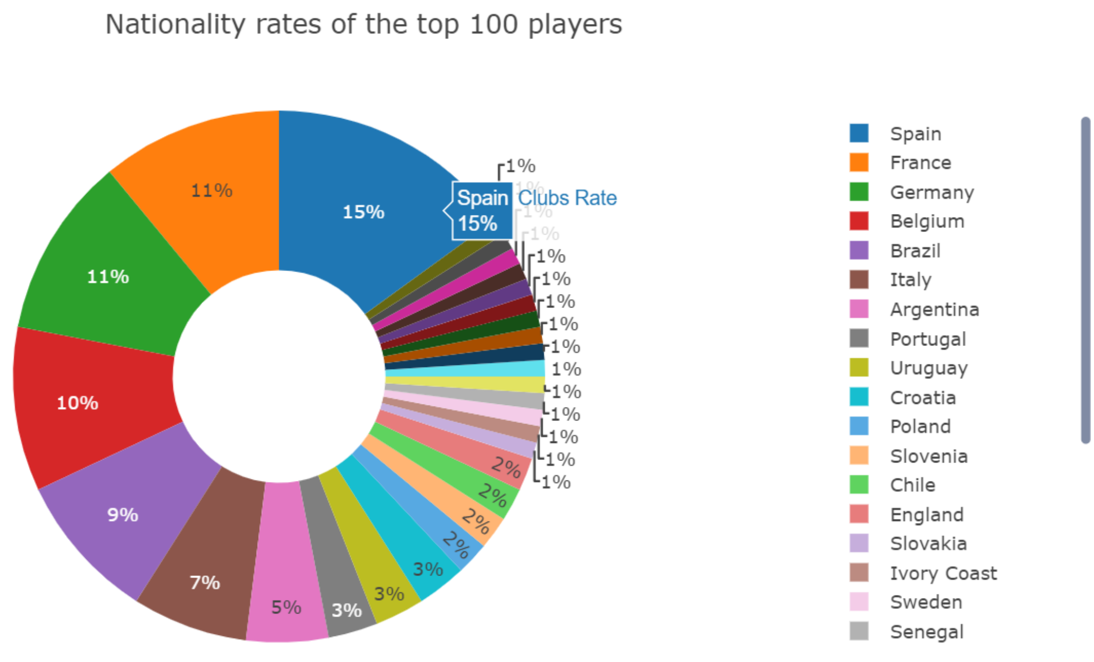




Club rates and Nationality rates of top 100 players



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.

References:

1. **(Shrivastava, 2017)**

Bibliography

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