DATA VISUALIZATION (CA_TWO)

FIFA 18 COMPLETE PLAYER DATASET

Group Members:

Hima Jyothi Aluri (Student No: 10515222)

Lakshmi Harika Chennamsetty (Student No: 10515217)

Enugu Madan Mohan Reddy (Student No. 10516674)

Dataset: FIFA 18 Complete Player Dataset

Source: The dataset is obtained from 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, Player Ids and the playing and style statistics. This dataset consists of 18k records for 70+ attributes with more than 50 attributes describing various player performance scores and position details. Along with these performance indicators we also have player personal data with attributes like player Name, Age, Nationality, Club, wage etc.

Since the dataset has a lot of exploratory potential, hence we chose this dataset for our assignment. In order to visualize the data, we have used **Python and Tableau** platforms.

Contribution of individual team members: The assignment was completed with the joint effort of all team members. The work was split as follows:

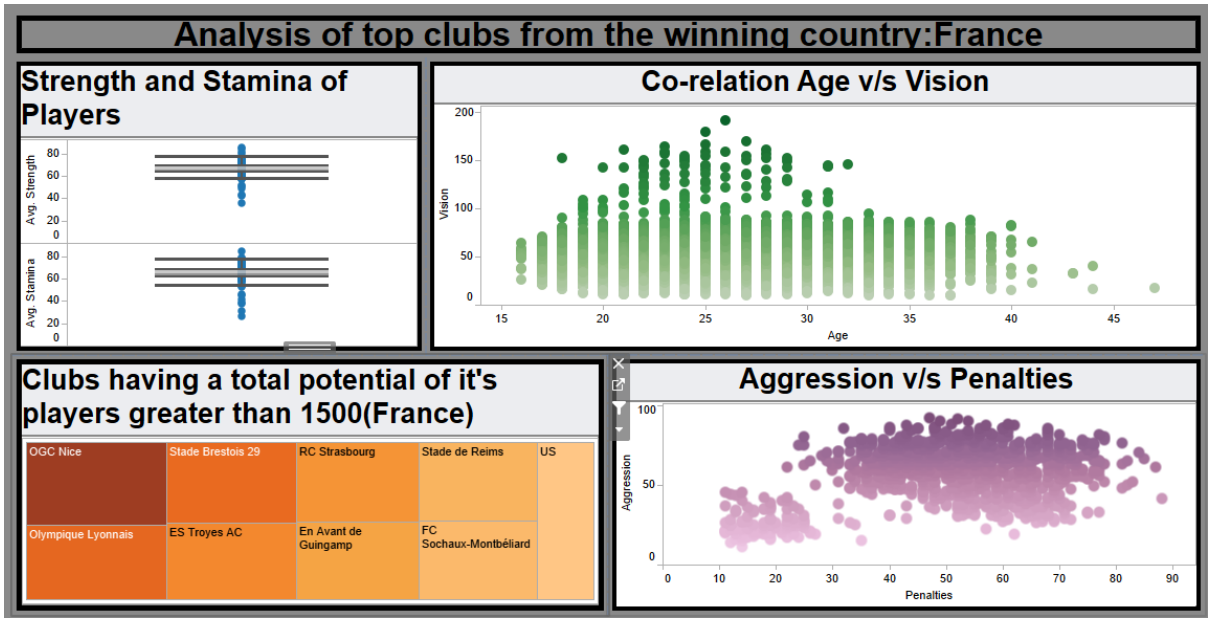
1. Hima Jyothi Aluri (Student No: 10515222) - Data Preparation and Visualisation in Tableau
2. Lakshmi Harika Chennamsetty (Student No: 10515217) - Data Visualisation in Python and Github setup
3. Enugu Madan Mohan Reddy (Student No. 10516674) - Data Visualisation in Python and Report creation

Deployment: The assignment files are hosted on the following Github link:

Visualizations: We have explained the dataset attributes in detail with the help of multiple interactive plots for their visual analysis.

Tableau Visualizations:

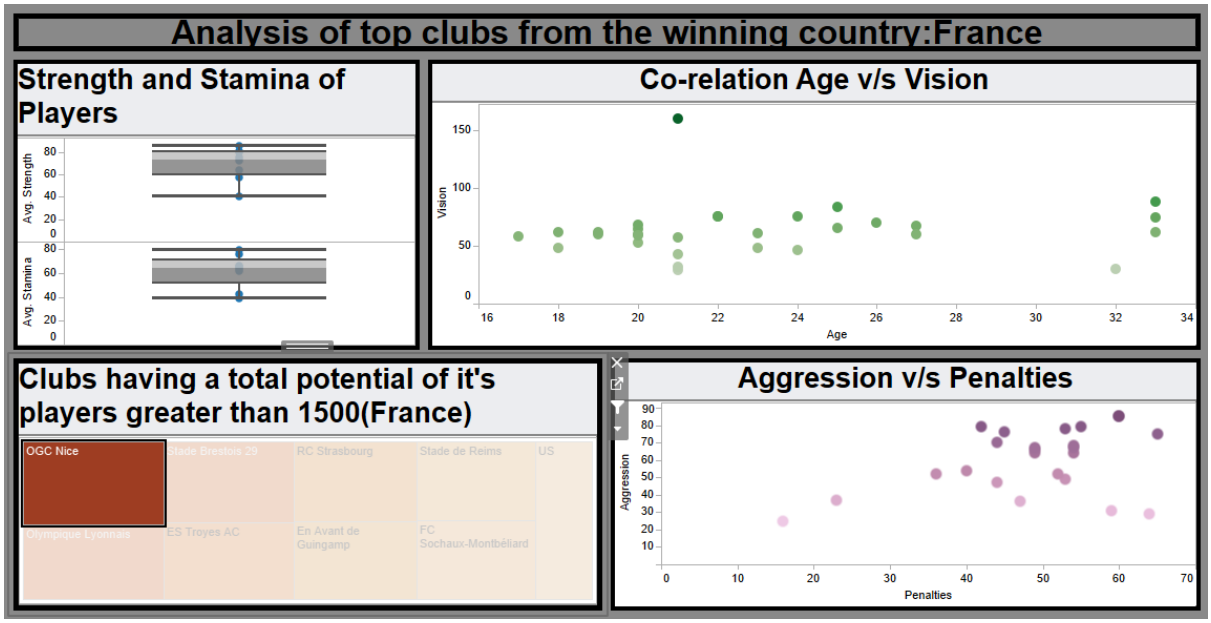
In order to understand the performance measures of the players, we have created an interactive dashboard in Tableau. In order to do so, we have identified clubs which are having players from France with higher potential.



The above dashboard shows the average strength and average stamina of all the players, the co-relation between their age and vision, the aggression level of the players and the number of penalties and the number of clubs belonging to France having players with total potential greater than 1500.

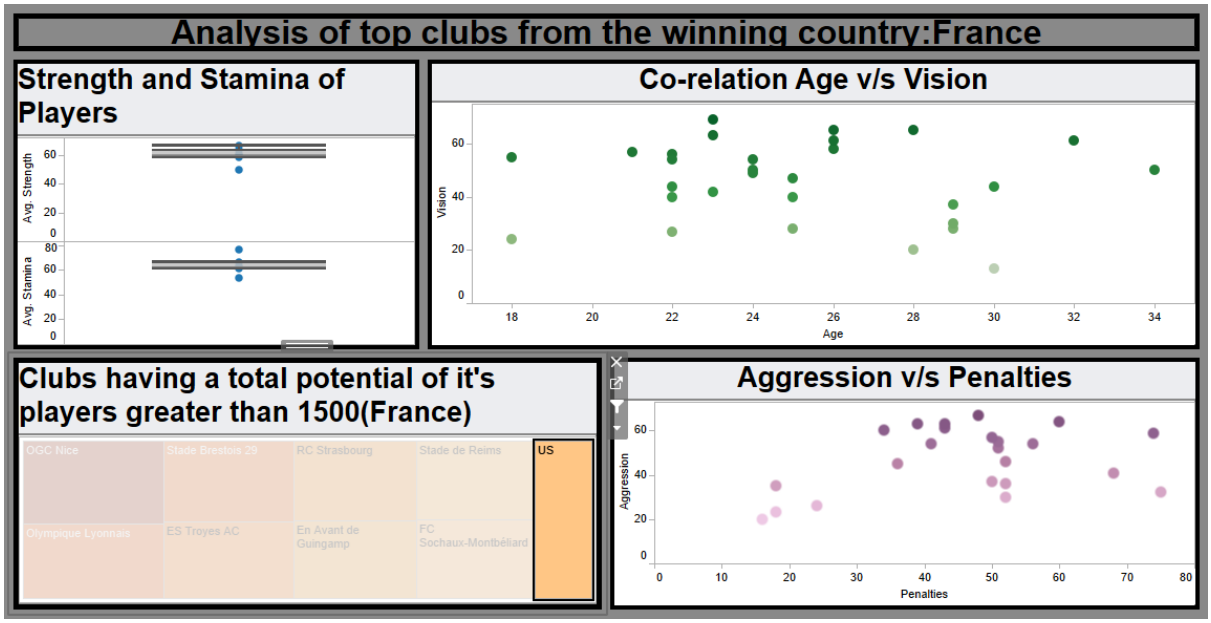
Now let's apply filter and check the performance measures of the players belonging to the best club – OGC Nice.

We see that the range of average strength of the players is between 60 to 80 with a median of 73.5 and the range for average stamina is between 52 to 71 with a median of 64.45. Although the vision of the players is supposed to be proportional to their age, we see that the players belonging to same age have different vision measures and the vision does not decrease in players who are older at age. Most of the players have more than 50 penalties, however many players have aggression level more than 70.

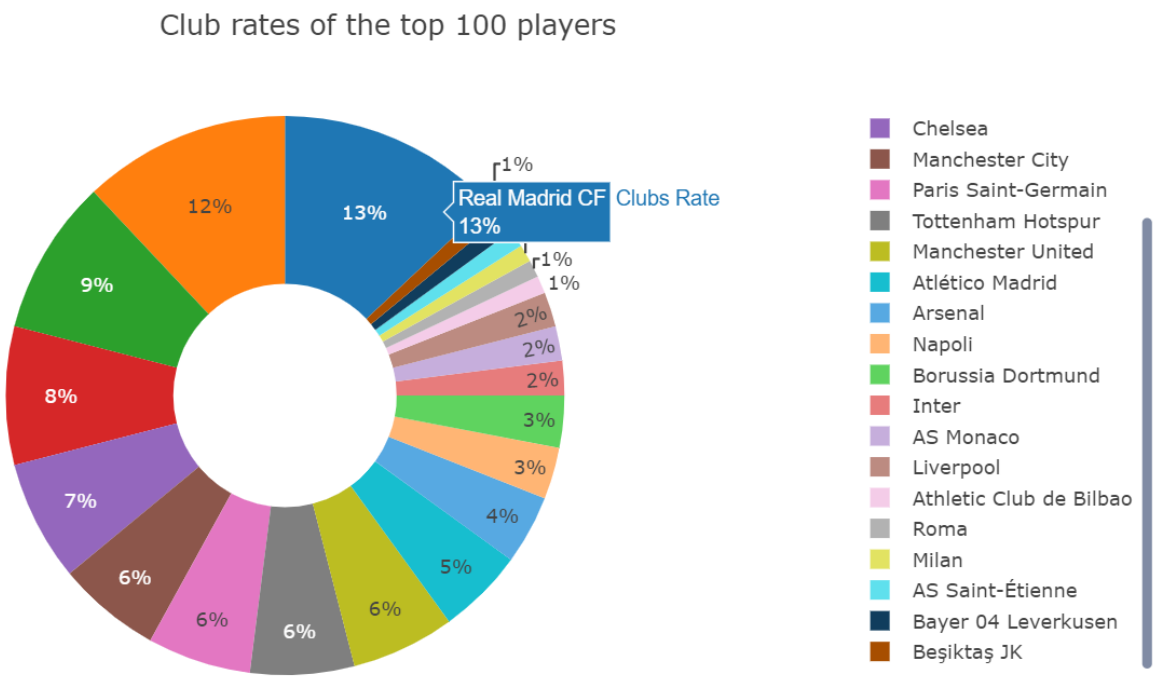


Now let's filter out the worst club - US Quevilly-Rouen and check the performance measures of the players

We see that the average strength and stamina ranges of the players belonging to the club US Quevilly-Rouen are much lower as compared to the players of OGC Nice. The range of average strength of the players is between 59 to 64 with a median of 61 and the range for average stamina is between 61 to 66 with a median of 62.50. However, we see that the players of all ages belonging to the club US Quevilly-Rouen have a much better vision as compared to the players of OGC Nice, whereas the aggregation and penalties are almost the same.



Percentage of top 100 players belong to each club



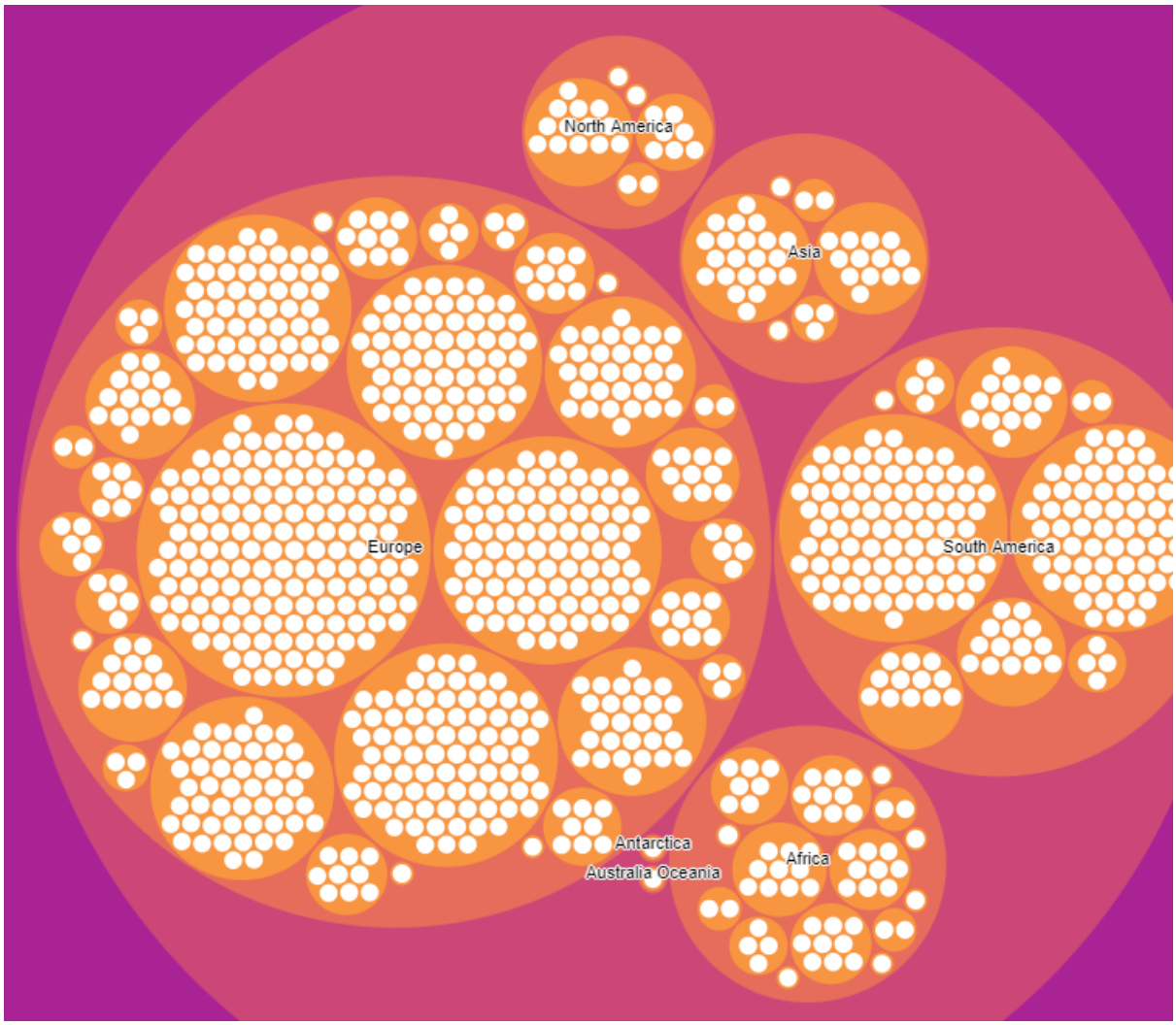
In the above plot, we see that the highest number of top 100 players i.e. 13% belong to Real Madrid club, whereas Chelsea and Liverpool have half the number i.e. 7% and 6% respectively, and Arsenal has just 4%.

Visualizations in Python

Grouping Top 2000 players based on their Nationality

Here we have used Zoomable Circle Packing with the help of D3.js in IPython. Circle Packing is the arrangement of circles of equal or varying sizes on a surface in such a way that none of the circles overlap. In order to do so, we have used Python with its libraries like numpy, pandas, seaborn, matplotlib, plotly etc. The data was converted to json and plot was created using IPython.

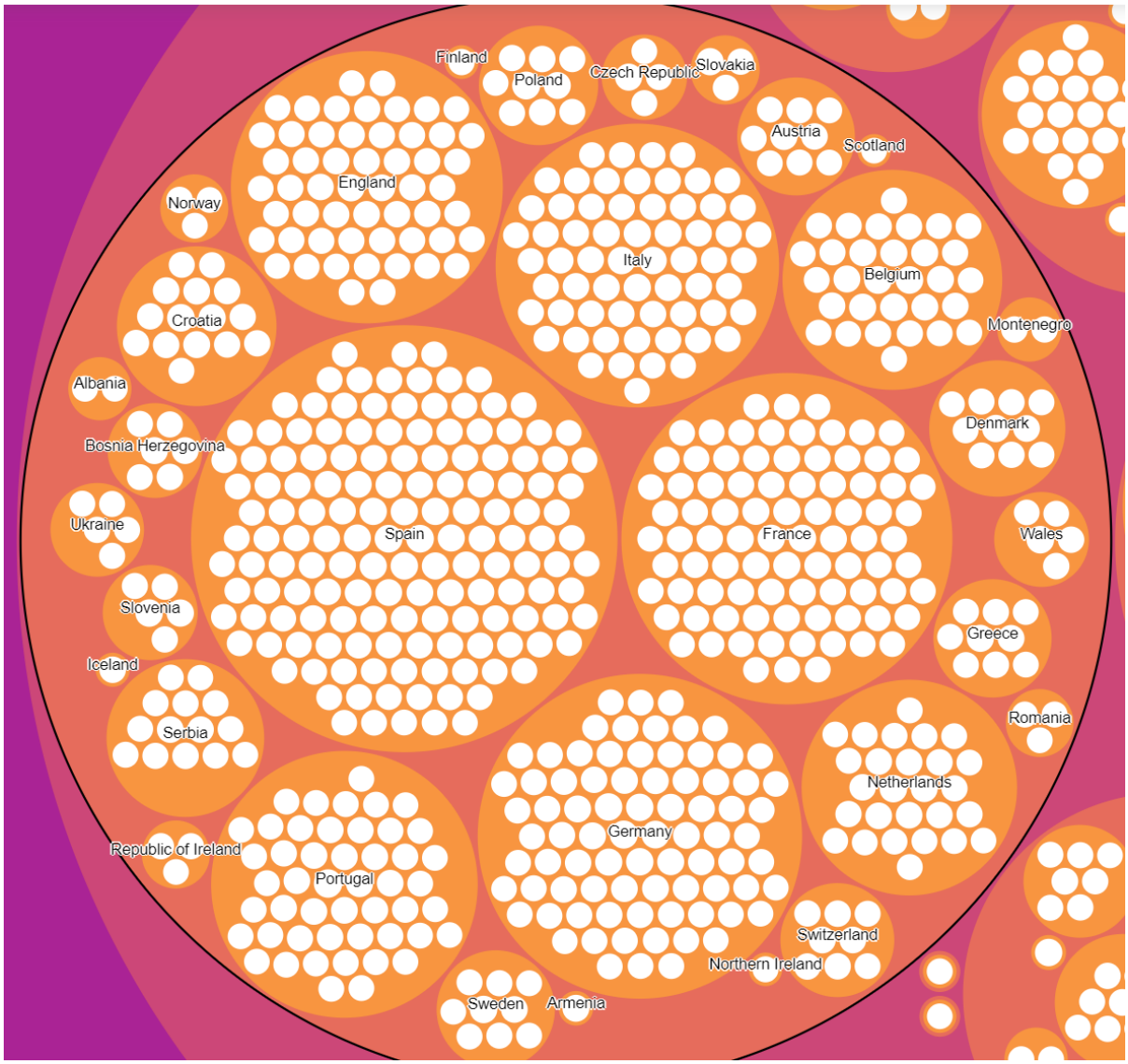
There were 162 countries in the dataset and these countries are mapped to six continents - Asia, Africa, Europe, North America, South America and Australia. The nationality of each player is mapped to its respective continent. In the plot, each continent will be the parent class i.e. the outer circle. We can then dive deeper within the parent class to find the countries with the continent, and within each country, we will find the players i.e. inner circles.



The above plot shows the player density in each of the continent. From the plot, we see that Europe has the maximum of the top 2000 players, followed by South America, Africa, Asia, North America and Australia, with Antarctica being the least popular.

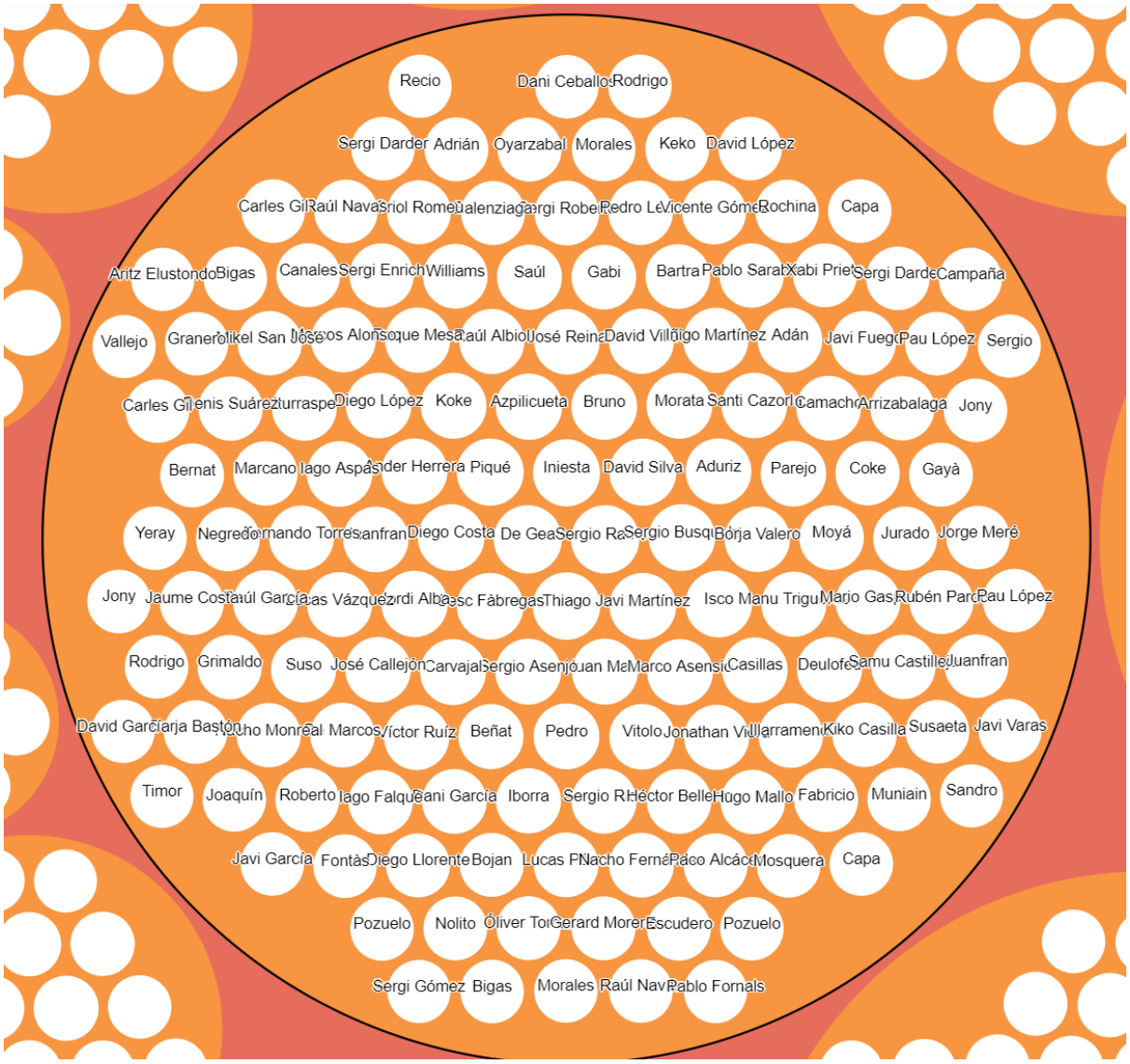
Now let's zoom into Europe

Here we see that Spain, France, Germany, Italy and England have the maximum number of players, while Scotland, Armenia and Finland just have single player each.

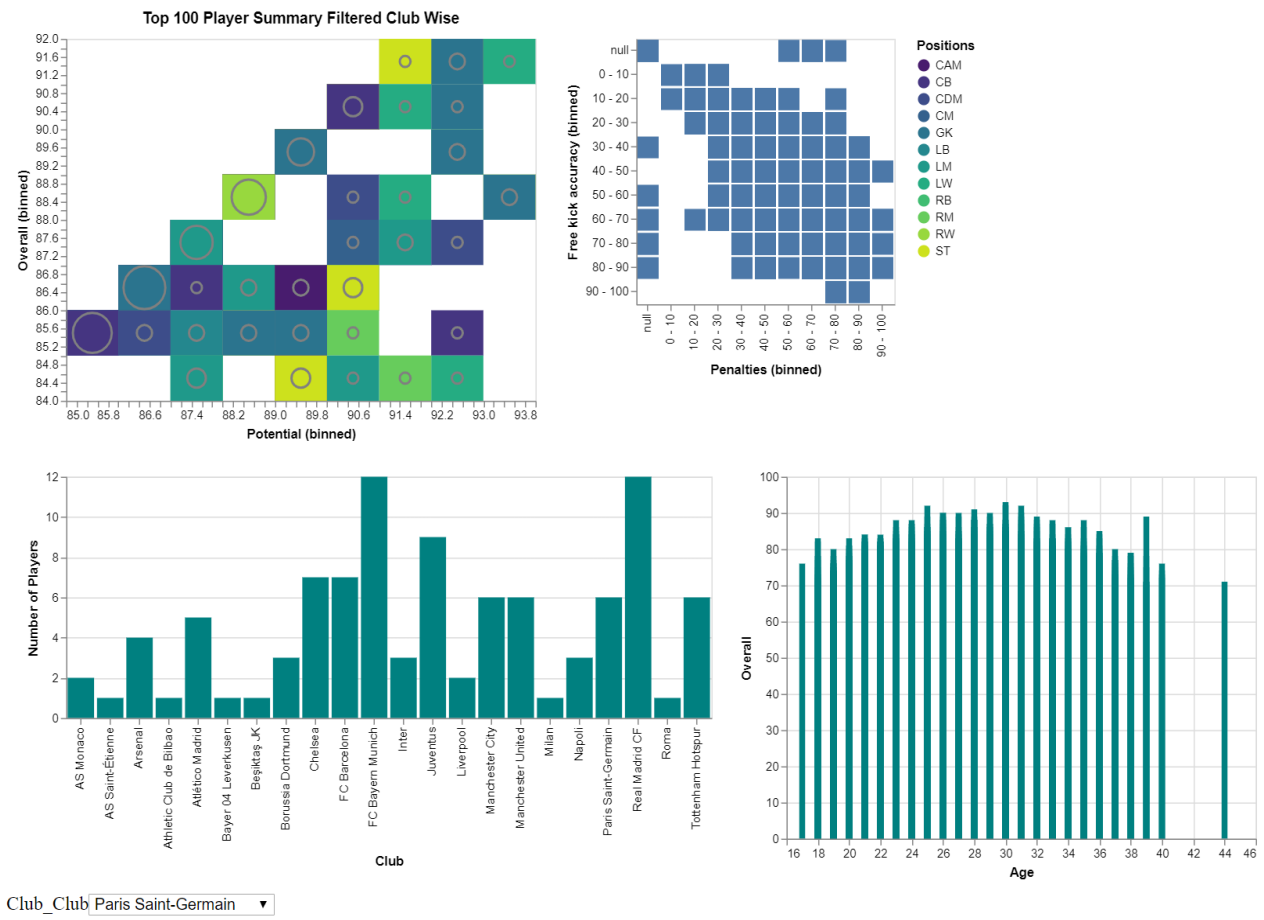


Now let's zoom into Spain

Here, we get the names of all the players from Spain who belong to the top 2000.



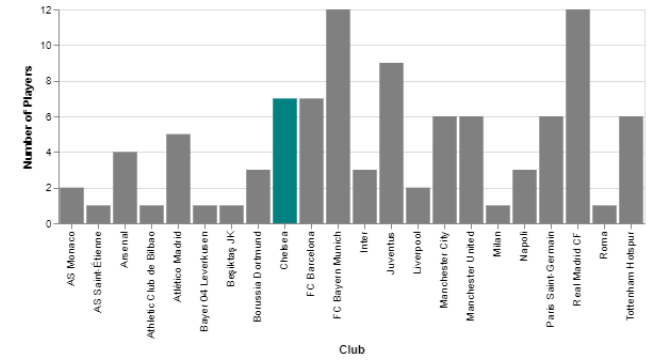
Summary of Top 100 Players filtered Club Wise

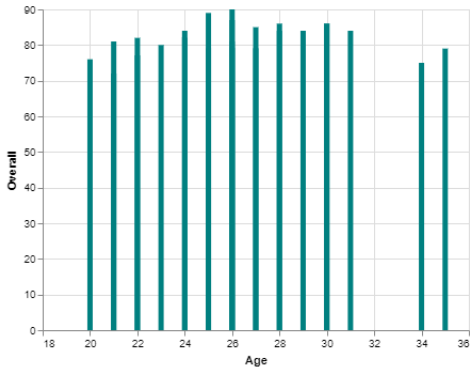


The above dashboard was created using Altair library in Python. There is a dropdown below by which you can select the club and various attributes of the top 100 players belonging to the selected club will be displayed. This is useful in comparing the various attributes of players belonging to different clubs.

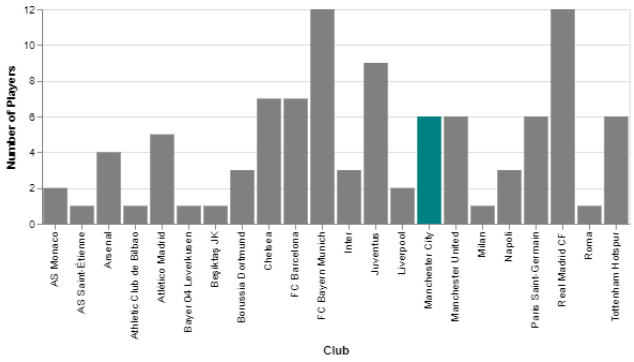
The dashboard shows the player's potential and overall based on their preferred positions, free kick accuracy and penalties of the players, number of players in each club and the overall age of the players.

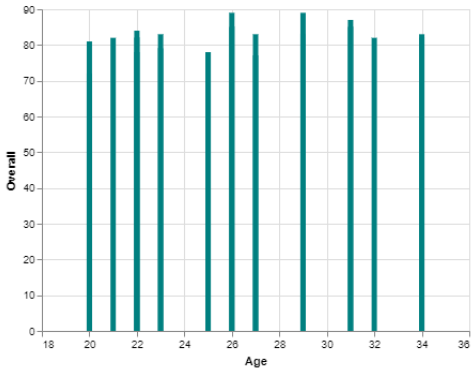
Let’s compare the age of Players belonging to Chelsea and Manchester City





Club_Club Chelsea





Club_Club Manchester City

The above plots help to analyse the age group of the players belonging to Chelsea and Manchester City clubs.