MONASH University

UNIT CODE: FIT5147

DATA EXPLORATION AND VISUALIZATION

Project Title: FIFA 19 Data Exploration and Analysis

Name: Anand Rane

Student ID: 29934176

Tutor: Mohammad Haqqani

INDEX:

I. Introduction:

Football is not just a game, its an emotion for many. People follow their favorite Clubs no lesser than their Religion! Great Players are celebrated all over the world. But not everyone knows how much are these players' paid?, what contributes to their Market value?

The Fédération Internationale de Football Association (FIFA) is an organization which describes itself as an international governing body of association football. Using the 2019 Player Dataset by FIFA, we will try to answer some interesting questions:

1) What are the factors responsible for player's overall performance and market value.
2) Why are the young players being paid more than aged and experienced one's?
3) Does the preferred foot affect the positioning of the players, can the left footed players play on the right positions and vice versa? Is the fight for the position easier for the left footed players ?

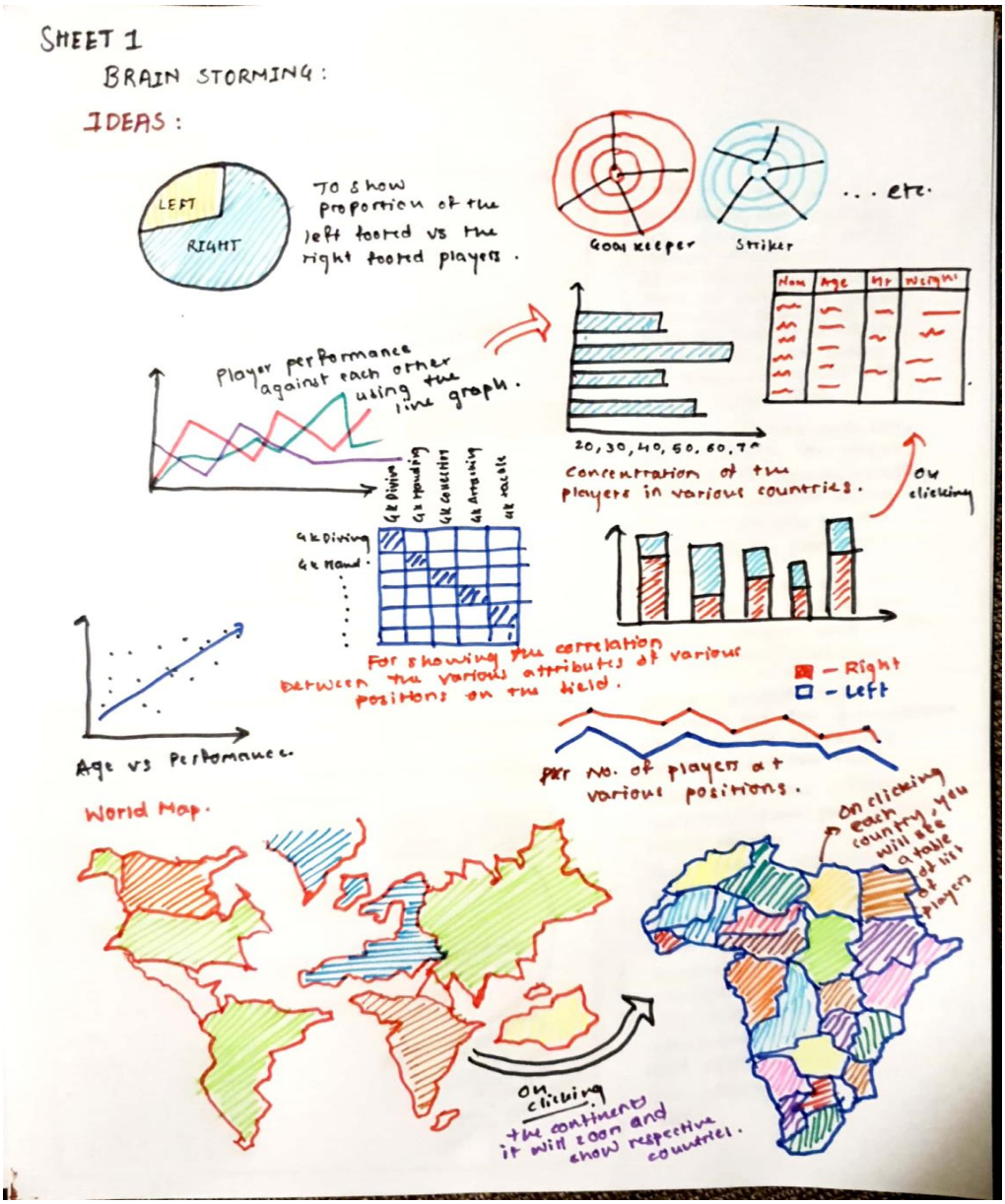
I am myself a Football player and a enthusiast and hence chose this topic for my project, since a very long time I often wondered why is Messi (left footed) better than Ronaldo (right footed), what can be the reasons to that, are the left footed more talented ? And why are the young players like K. Mbappe are being highly paid, since they have just started their career, aren't they supposed to be payed less than the experienced one's? And also what are the factors which are assessed while counting the overall rating of a player, I had the curiosity in me to get the answers to these questions, and which is why I was motivated to take this up as a challenge to analyse FIFA datasets as my Exploration project.

II. DESIGN:

1. Design-Sheet 1:

In the brainstorming process, lot of ideas came from the exploration project which I did previously, I have planned to answer the previous exploration project using visualisations, having that in mind I started brain storming according, I started with my second question about left vs right footed players, and hence I thought of drawing a pie chart to show the proportion of the left and the right footed players from all the countries and clubs, then I moved on to use line graphs to show the performance of each players against other at various attribute levels. I also thought of using the bar graph for the same. Then in the exploration I had found out the attributes responsible for the players value and rating. Hence to visualize it, I planned to show a correlation graph. Moving on, I also thought of showing the stacked bar graph to show the market value of the players. I also wanted to show how the age affects the performance and value of the player.

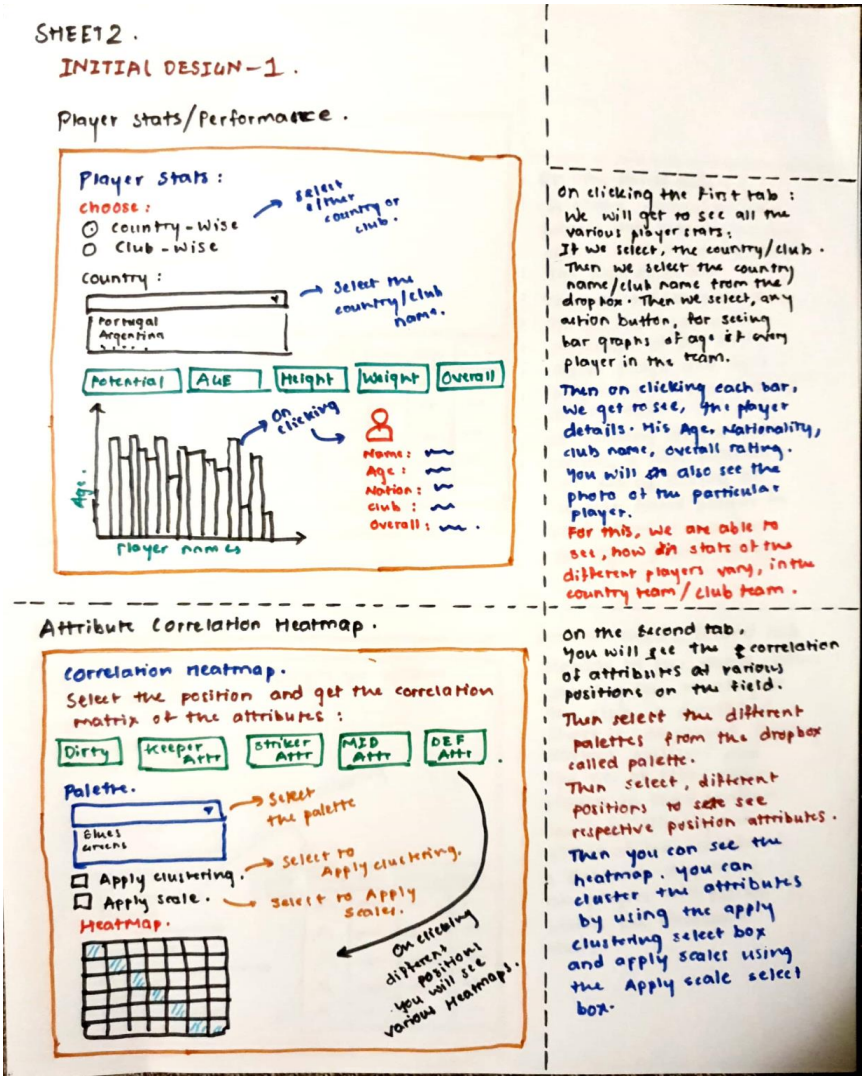
Then again I drew a line graph to show the proportion of the players at various positions on the field hence my first structure sheet simply incorporates my conceptualizing part, in which I came to think about starting phase of perceptions which I can use to clarify my results.



2. Design- Sheet 2:

In the second design sheet, I started with my tab panel, where the second sheet displays my first tab page. Here I have implemented bar graph to show the stats of the players at various attribute level such as Potential, Overall, Age, Height, etc. Firstly we will ask the user to choose the country or the club using the radio button and then selecting the particular club/ country he is after. Then pressing on action button to see how the players compete with each other, represented in the bar graph. You can also see the player information when you click on the particular bar.

The second design sheet also covers the attribute correlations at various positions eg: Goal Keeper, Striker, Mid-fielder, etc. We will use a correlation heatmap to represent how well the attributes correlate to each other. We will also be able to cluster the attributes or apply scale on our correlation heatmap matrix.

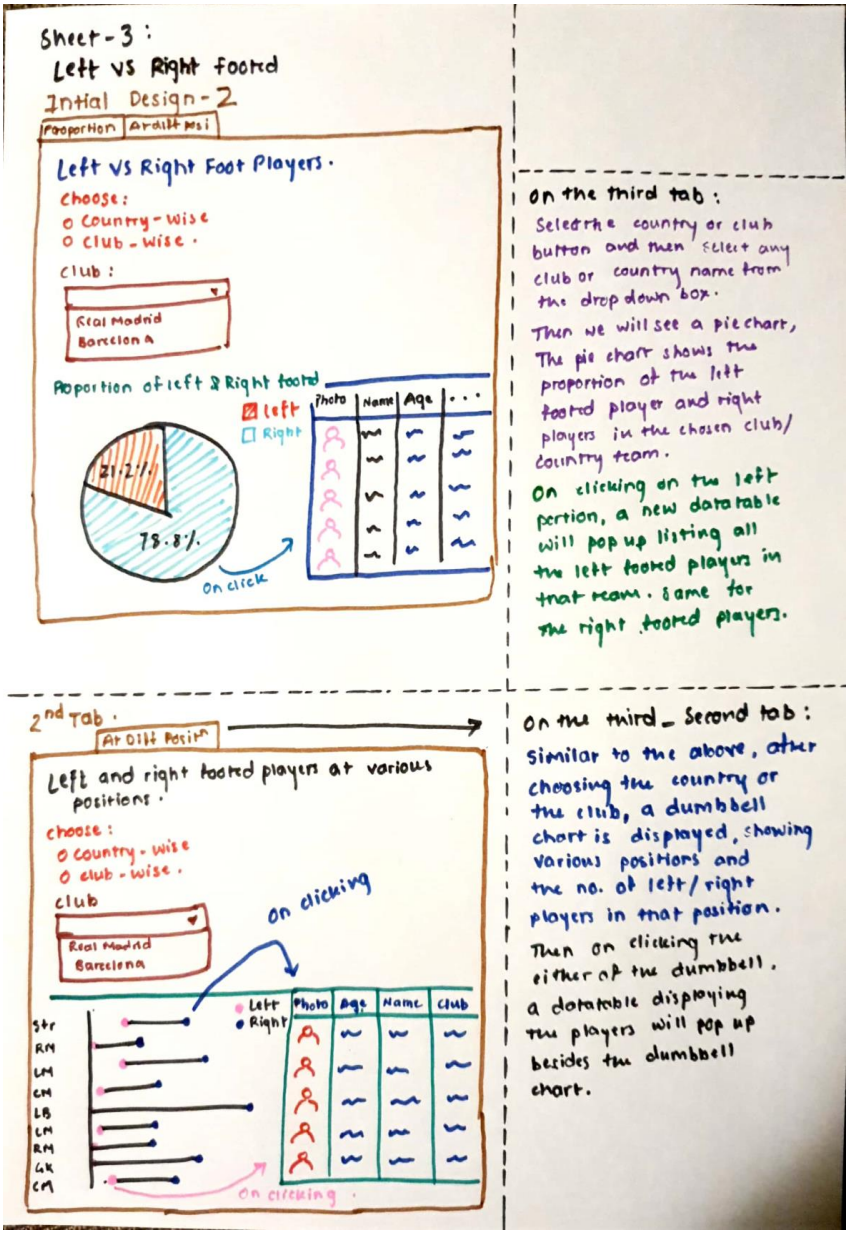


3. Design- Sheet 3 :

The third design sheet consist of the early topic from the first design sheet where we focused on showing the proportion of the left and right footed players all over the world. I have made a interactive pie chart which displays the proportion of the left and right players at different clubs or countries. We will let the user to specify the country name or the club name.

To make it even more informatic, we will also display the information of the left footed players or right footed players in the particular team.

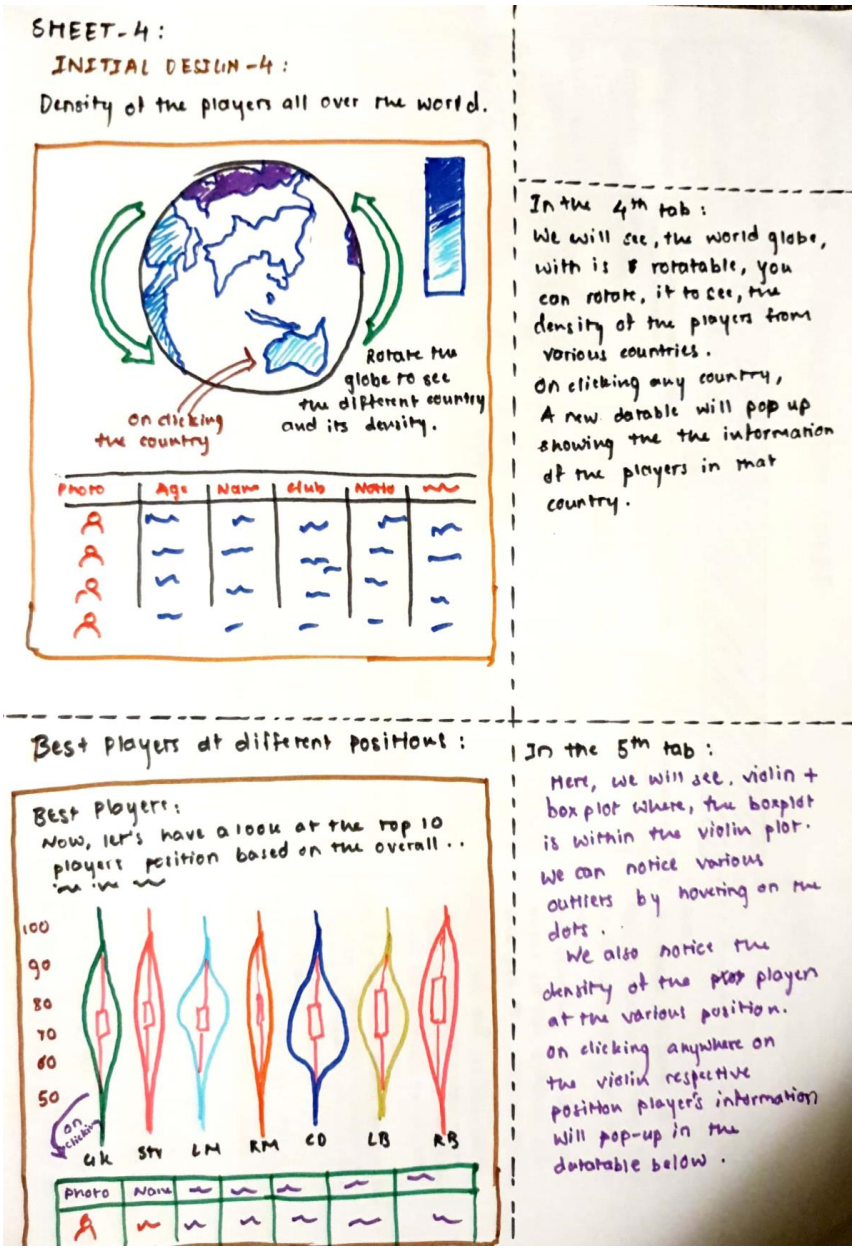
The second part of this sheet is display the dumbbell chart since it was more informative than the line chart previously decided, hence I made use of its categorical property to show the number of left footed and right footed player on each dumbbell at various positions. The user can also see the respective player in the form of a data table.



4. Design-Sheet 4:

In the fourth design sheet, we can see the density globe, this is a rotatable globe displaying the density of the player all over the world, this helps us easily identify the countries which are major contributors to the FIFA world cup and also clubs. The purpose of this map is to give the user independence to look at various countries. The sheet also displays the data table of the players from each and every country.

The second part of our sheet displays a violin plot combined with a box plot. The purpose of this graph is to firstly show the density of the players at various locations and secondly display the outliers at different positions. The design sheet will also display the best 10 players at various positions. In total this particular tab will give 3 important statistical information.



5. Design-Sheet 5 :

The fifth design sheet is nothing but the final realisation sheet or the final product. Our fifth sheet displays the interface of the app. On selecting each of the tabs from the sidebar menu you can see the respective tab page. Where will all the discussed visualisations take place.

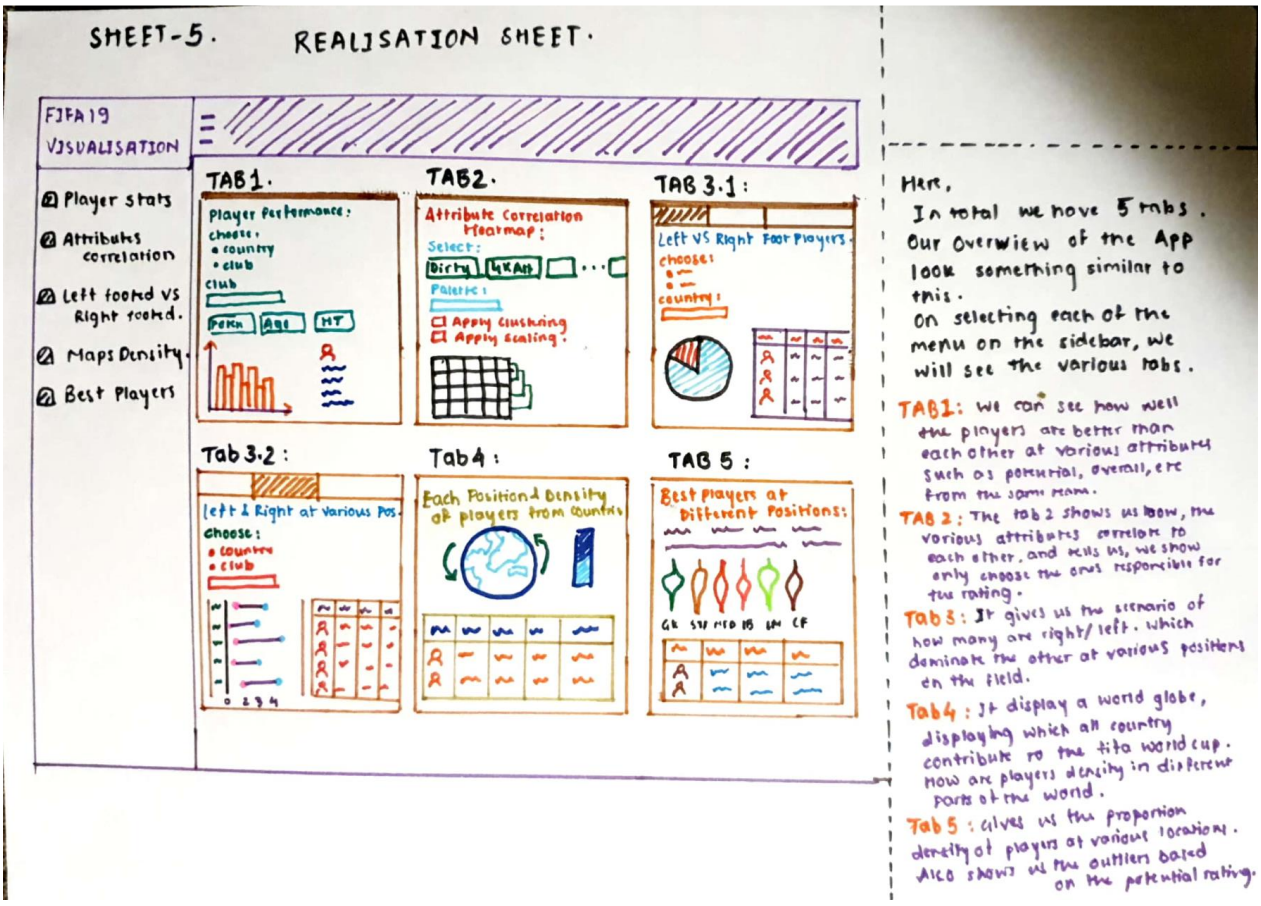
Tab 1: In here we can see how well players perform against each other at different attribute level such as potential, overall, etc.

Tab2: In tab 2, we can see how the attributes correlate with each other, the attributes which correlate well tell us that it belongs to a particular player position. And hence it displays the dirty and also the correlation of the attributes at different positions.

Tab3: In tab 3, the pie chart tell us proportion of the left and right footed players across all the clubs and countries whereas the dumbbell chart tell us whether the left or right footed dominate each other at various positions.

Tab4: In tab 4, the density globe tells us the density of the players all over the country,

Tab5: In tab 5, the violin and boxplot chart gives us the proportion of the players at various positions and also tells us the outliers based on the potential rating. It also displays the top 10 players from every position.



III. Implementation :

After working on five design sheets and watching the dataset intently, I executed the entire undertaking utilizing the R programming language. R helped me to deal with huge dataset immaculately. The primary motivation behind why I picked R is, right off the bat, it accompanies an enormous arrangement of libraries which helped me to chip away at intuitive representations in the most ideal manner; besides, task on huge datasets in R Is quick. Additionally, R is the best software with regards to the graphical portrayal.

Below are the libraries that I used for the implementation:

1) Shinydashboard – It is an interactive dashboard providing a sidebar menu options for adding tabs and overall, it makes the app look more professional and involving, Bored of the same old shiny, I switched to shinydashboard for a try, and I liked it.
2) Shiny – To build the app I included the library.
3) Dplyr – Included this package to play with the dataset.
4) D3Heatmap – To implement the heat map later in the app, I imported this package.
5) Corrgram – Helps to visualize correlation matrix.
6) Countrycode- To convert the country names to 3 letter country code for using the plot_ geo later.

The exploration project gave me a handful of insights for proceeding with the visualization project here.

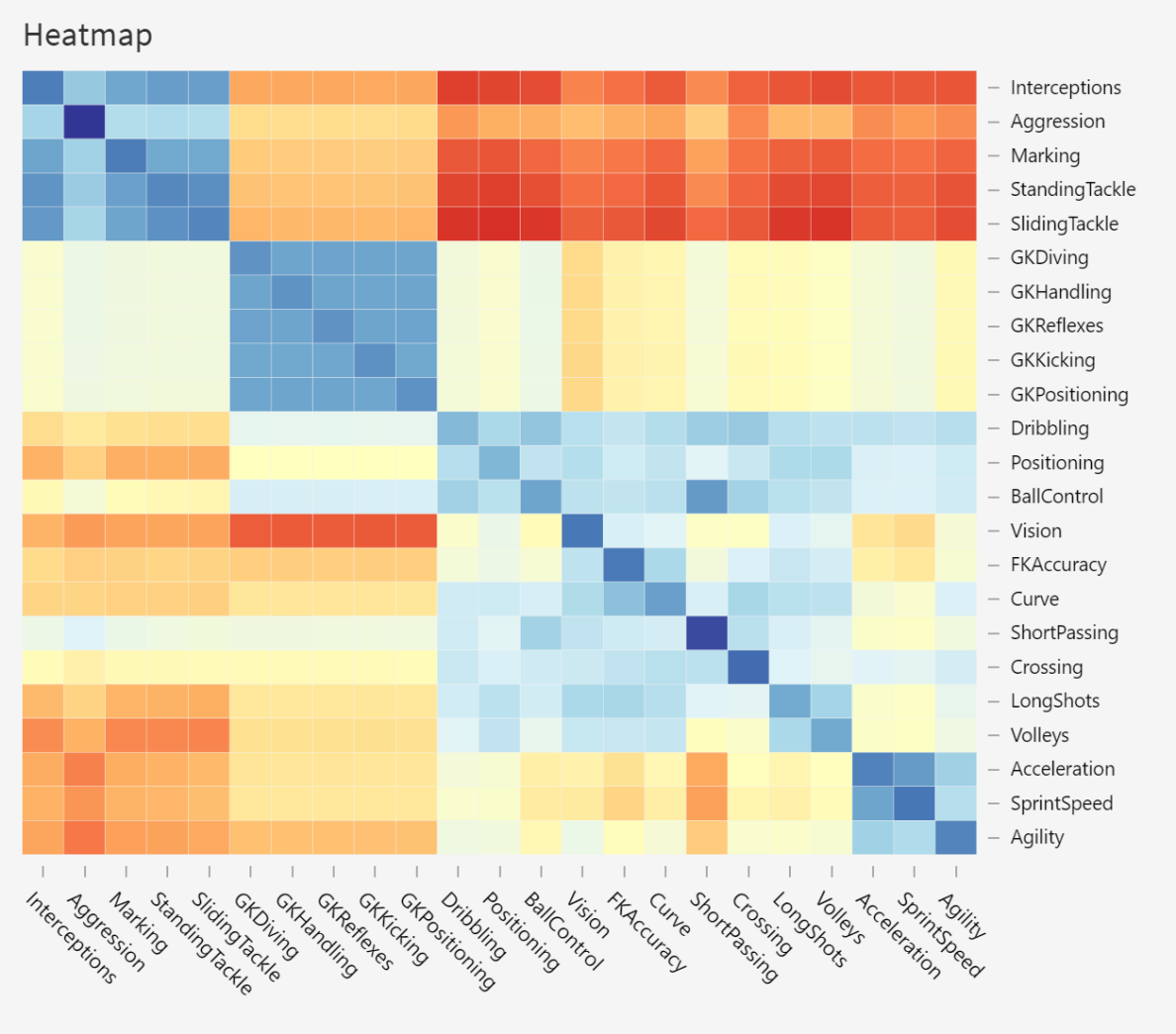
Firstly I read the data by importing and converting it to dataframe using the dplyr package imported above. Then I converted the market value and wage from currency to numbers, removing the euro sign and substrings K and M.

Then I converted all sorts of unnecessary positions to more general positions for the actual 11 positions on the field and then stored the result as a new column in the dataframe.

I started with first creating the dash board body, wherein I included all the tab items which are required for my later cause. Made entry in the dashboard sidebar with the tab names given to my tabs.

My first and foremost tab is Player stats, where in I have radio buttons for the user to choose country-wise or club-wise, on selecting one of those, the user gets a drop down box to select the team he/she wants. There are 5 action buttons provided namely Potential, Overall, Age, etc, on pressing one of these, the bar graph pops up showing the bar graph where the players compete each other based on the selected attribute. On clicking the particular bar, the image and information of the player pops up, for this is used the renderUI and verbatimTextOutput to display the player information.

Moving ahead to the next tab, in the second tab, it gives the correlation of the attributes at various positions on the field. I made use of the heatmap to show the correlation between different attributes. Different action button are given to choose the desired correlation of attributes for that position. Also palette option, applying clustering and scaling to the heatmap is given. This lets us understand which factors or attributes are really responsible for determining the value and rating of the player. For example, we cannot rate a striker based on the goal keeping skills, and hence we need to determine which attributes should be considered. Below is the image from the generated heatmap correlation matrix for the mixed attributes.



For example, we can see that how badly the goal keeper attributes correlate with the other attributes.

For the next tab, that is the third tab, I have performed visualisations to determine, whether the left footed or the right footed players dominate on each other. The third tab consist of subtabs, one is for the pie chart which shows tells us the proportion of the left and right footed players in the country team or the club. On clicking the portion of the pie, a data table will pop up displaying the number of left/right footed player along with their basic information.

In the second tab called At various Positions, I have used a dumbbell chart, since it is much more informative and precised than the basic line chart. I used its categorical property to define the left and right footed players at various positions. I have also added interactivity to the chart. When the user clicks on left or right dumbbell, it will show a data table of the players at that particular position for both left and right.

Then on our fourth tab, I have included a density globe to show the density of the FIFA players across the world. It will help the user to understand which are the countries which contribute to majority in the football world. The density rotates, giving the user flexibility. On clicking a particular country it will show the data table of players which belong to that particular clicked country.

Finally in our fifth tab, I have included a violin plot combined with the boxplot. The violin plot show us the density of the players at that particular position and secondly the box plot shows the outliers at various positions. Instead of using the action buttons to show the list of best players. On clicking the violin plot, the data table of best players at that position pops up.

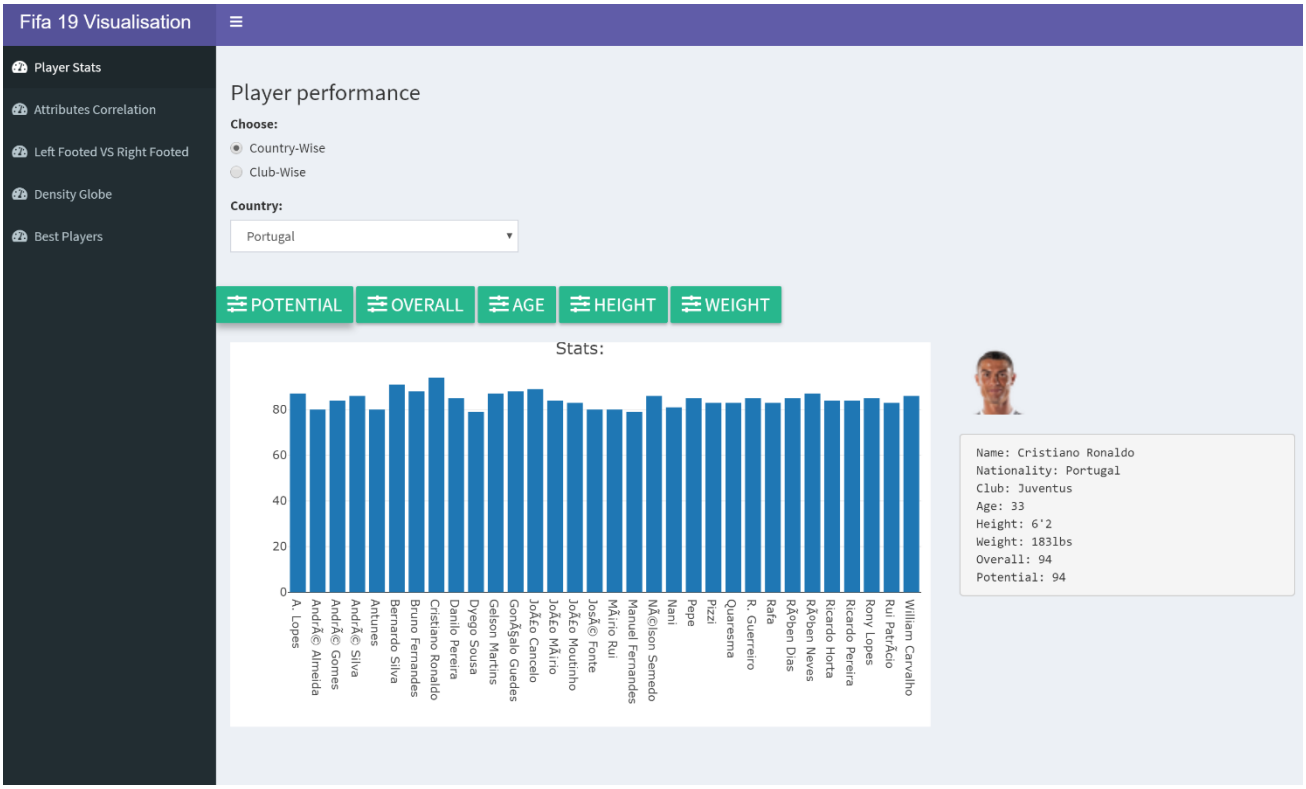
IV. User Guide

To open the app, all you need is to double click on the R script and just make sure that the data file is in the same folder as the R script and remember to set the current working directory.

On opening the app, the user should notice a nicely prepared Shiny dashboard with a purple skin. On the side bar of the app, you can notice that there are three tab items namely:

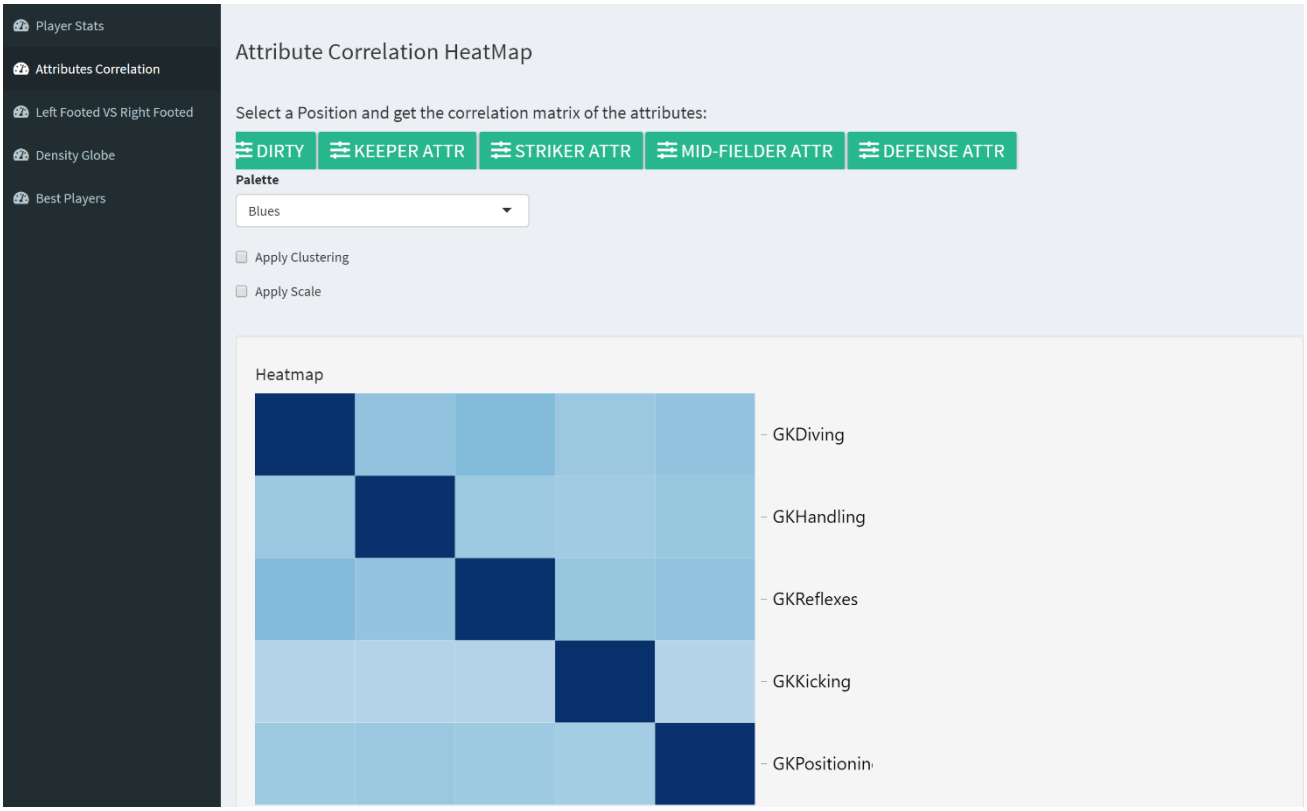
1. Player stats.
2. Attributes correlations.
3. Left footed vs Right footed.
4. Density Globe.
5. Best Players.

There are one or more than one visualisations on each tab. Starting with the player stats tab,



As displayed in the above picture, the user first has to choose the country or club-wise option, then select the concerned team from the dropbox. Once done, the user will select one of the action button to see the bar graph portraying players competing with each other. On hovering the user can see, the name and the attribute value of the respective player. On clicking the bar, it will display the picture of the player along with his basic information like nationality, name, club name, age, etc.

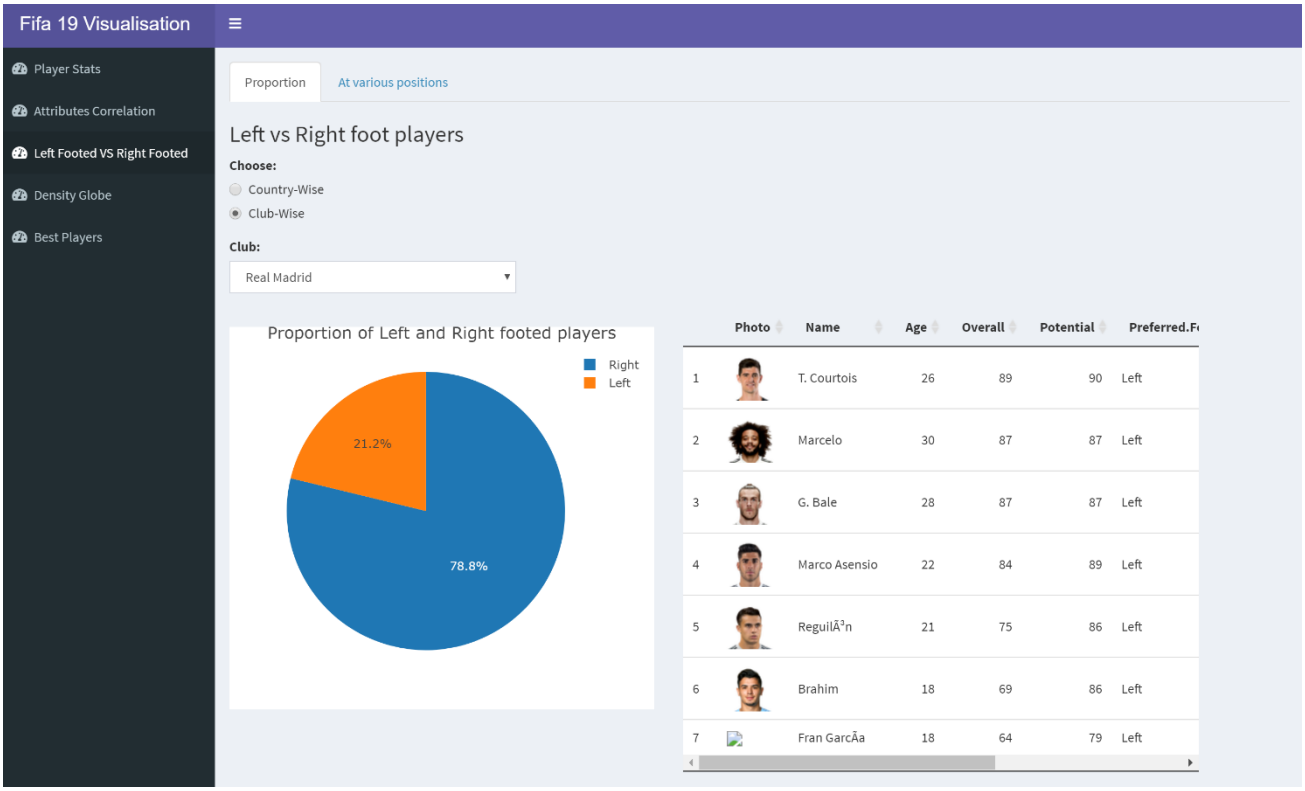
2nd Tab: Attribute Correlations:



In the above tab, we can see that, it displays a heat map of the correlation matrix, there are 5 action buttons namely, Dirty, Keeper Attr, Striker Attr, Mid-fielder Attr and Defense Attr, out of which the Dirty contains all the position attributes, its made available there, so that the user can understand that the unnecessary attributes should be separated and only the ones which correlate need to be kept aside. Except for the Dirty rest all action button gives you highly correlating matrix like the goal keepers as displayed above.

Also you can try different palettes and apply clustering to the heat map and also apply scale to it.

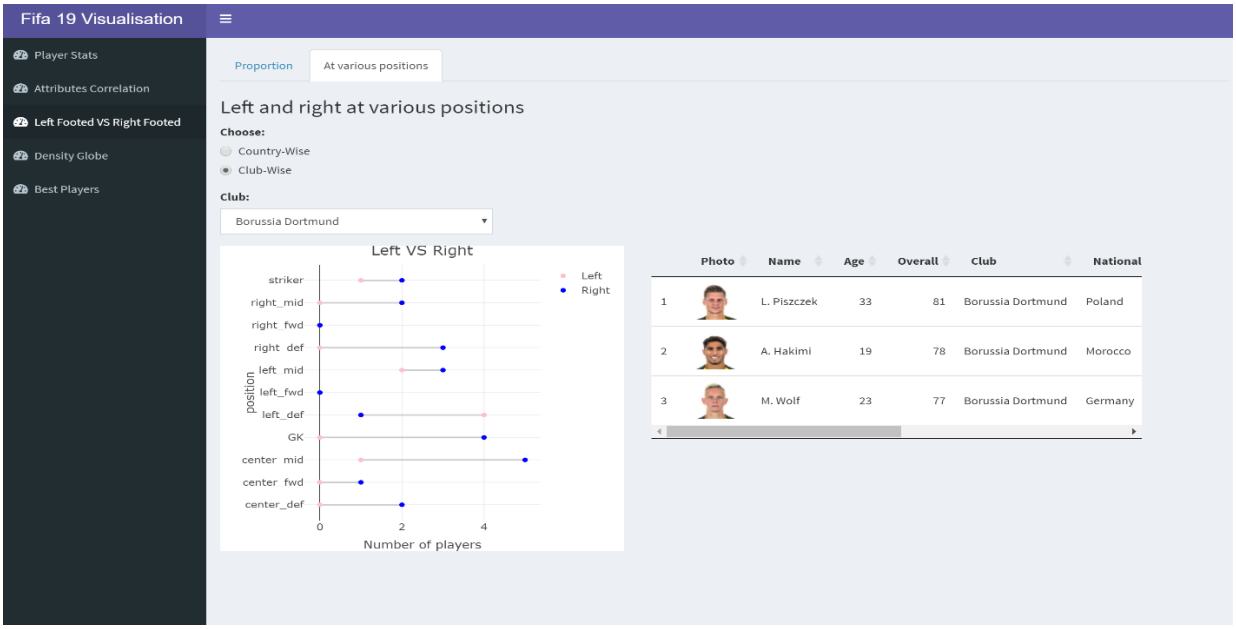
3rd Tab: Left Footed vs Right Footed.



Our third tab consist of 2 sub tabs inside it namely the Proportion and At Various Positions.

The above figures describe the first tab out of it, in here, the user needs to select the country name or the club name, on doing that, a pie chart stating the proportion of the left footed and the right footed players is plotted. On hovering over the pie chart, you can see the percentage of the left/right footed population. On clicking the portion, a data table pops up displaying all the left/right footed players and their few basic attributes.

In the second sub tab,



We can see that there is a dumbbell chart, the same choosing of the country or club is required, then it displays the dumbbell chart, on hovering any of the dumbbells, it displays the number of players at that particular position with the preferred foot. On clicking, a datatable pops up showing the list of players which are concerned with the particular dumbbell dot.

4th Tab: Density Globe

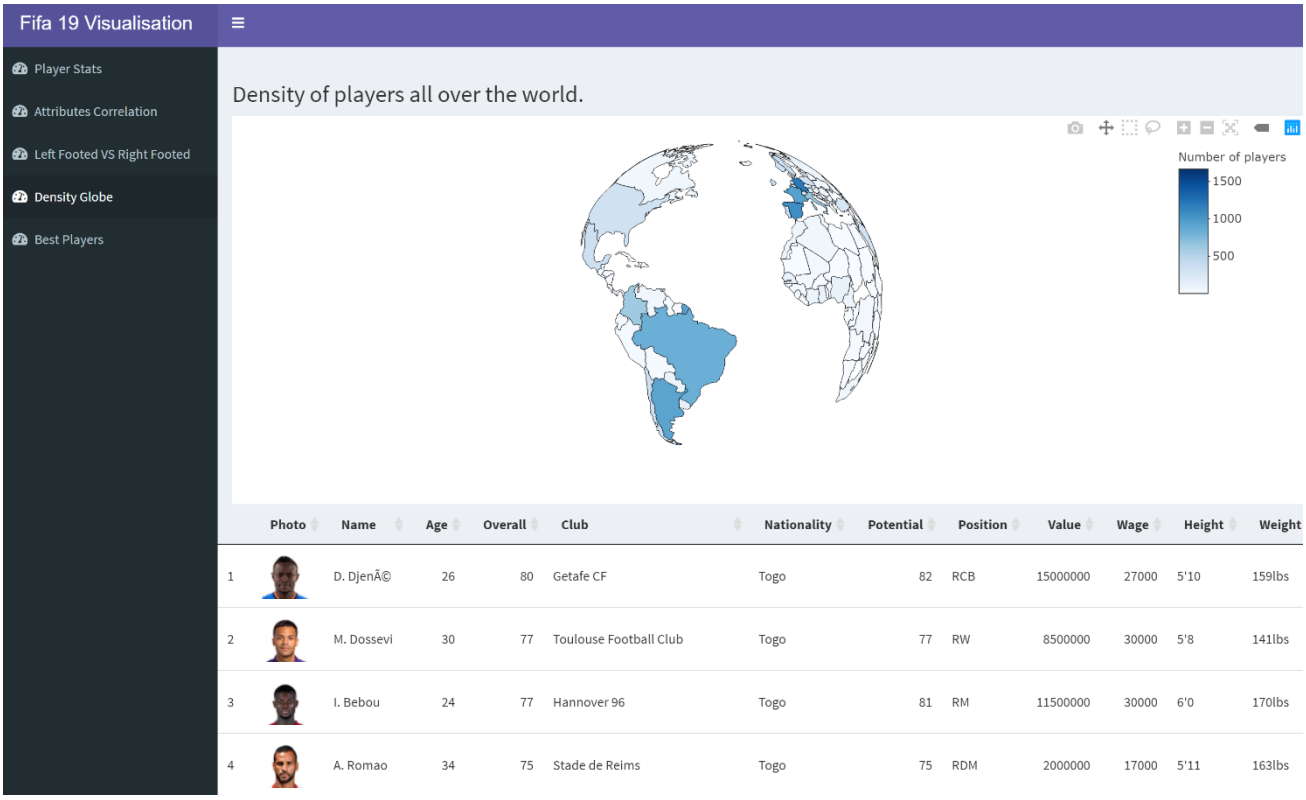
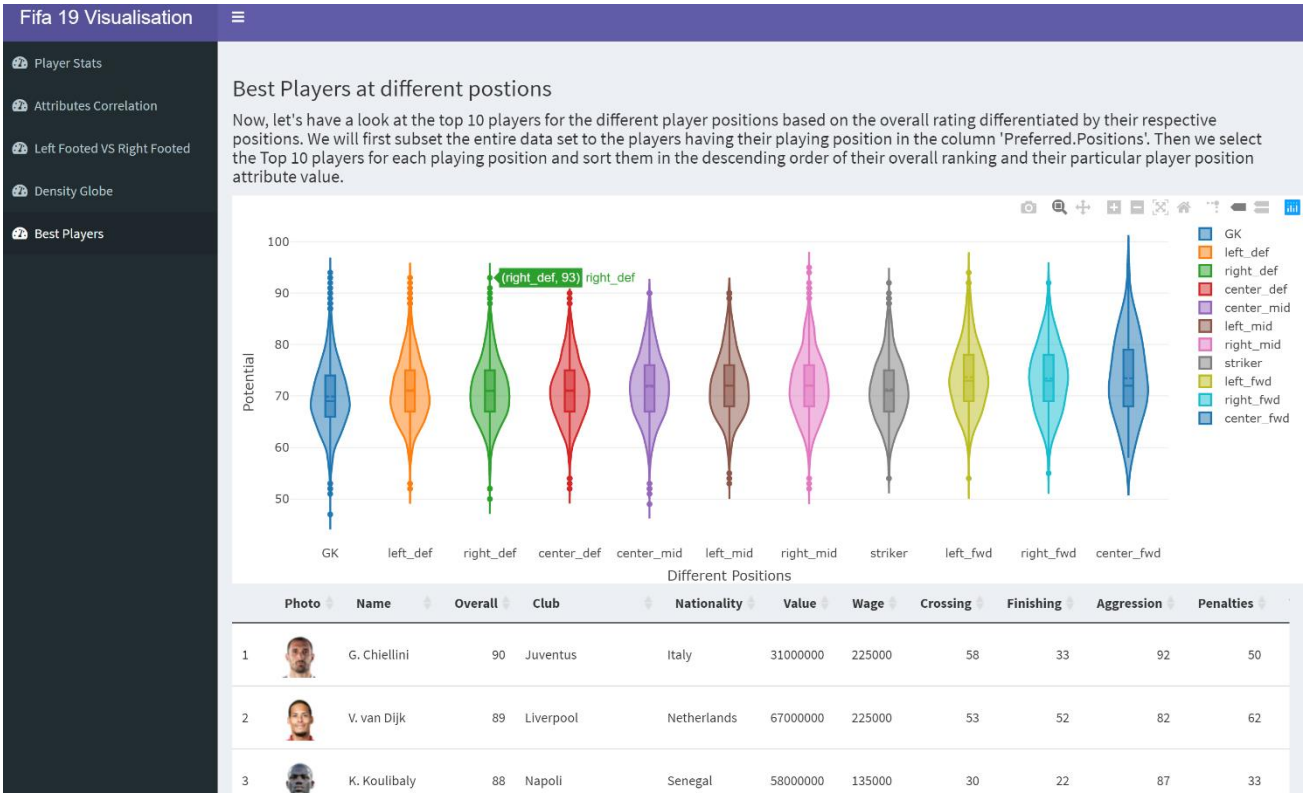


	Photo	Name	Age	Overall	Club	Nationality	Potential	Position	Value	Wage	Height	Weight
1		D. DjenÃ©	26	80	Getafe CF	Togo	82	RCB	15000000	27000	5'10	159lbs
2		M. Dossevi	30	77	Toulouse Football Club	Togo	77	RW	8500000	30000	5'8	141lbs
3		I. Bebou	24	77	Hannover 96	Togo	81	RM	11500000	30000	6'0	170lbs
4		A. Romao	34	75	Stade de Reims	Togo	75	RDM	2000000	17000	5'11	163lbs

By looking at the above image, the user can notice the density globe, it display the density of the football players world-wide. The globe can be rotated by simply dragging it to check the density of different countries. On clicking a particular country, the data table below shows all the players belonging to that country with few attributes.

5th Tab: Best Players



The above image, depicts a violin plot and a boxplot inside it. This visualization gives us three different information here, firstly the violin plot tell us the density of the players at each position, next the boxplot describes the outliers based on potential and thirdly it also displays the top 10 players at various positions.

On hovering on the points of the boxplot, the user can see the outliers like K. mbappe . On clicking the violin plot, it displays the top ten players at that particular position.

V. Conclusion

I would like to conclude the project by saying that, I successfully visualized the data to satisfy the thirst of questions which I wanted the answers to. The facts which came out with exploration were quite astonishing, I was under the impression earlier that left footed players were preferred more over right footed ones since proportion of left footed were less in comparison and had higher demand, the fight for the position was higher for the right footed players. All of these beliefs were rejected by our exploration and visualization. The fight for any position on the field is pretty much the same. And every player has to strive hard for the position, nevertheless the player is a left footed or a right footed.

Another interesting fact which came out is that, every player has skills of an goal keeper or a defender though the player's instinct are that of a strikers. Such skills cannot be assessed to calculate the player

rating, this can turn out biased and inefficient. With our attribute correlation heatmaps we were able to find out the correct skills respective to the player positions. Now the players can be assessed based on the skills they actually possess and which they train hard for.

We can also see which countries the majority of the players contribute to the entire football population, the leading countries are brazil, Argentina, England, etc. Then the using the violin plot and boxplot we could describe the density at various positions and the outliers based on the potential to spot the outliers like K. Mbappe.

VI. Reflections

I learnt a lot of concepts from this project and most importantly I enjoyed the entire process of exploration, since I chose the topic of my interest. This project portrayed how important exploration of the data can be, how deep any dataset can be explored, to be honest, in the this project with all the exploration and visualization I have done, I could barely scratch the surface of this massive dataset. The dataset can be still worked on deeper levels, it brings out the relationships between various factors and variables. I found a issue with the massive dataset that I used, since I had to correlate the attributes with one another, the correlation had huge samples which led to normalizing of the data, I am not sure to what extent that could cause a problem but, I found this point worth making a note of.

Another Good thing is that, I found R programming very easy and efficient with all the packages and libraries it provides, before this I was very afraid of R programming since I had no hands on experience with it. I also got an opportunity to implement statistical measures like Correlation in this project.

VII. References

1) Wikipedia.com
2) TowardsDataScience.com
3) Sofifa.com
4) Previous Exploration Project

VIII. Appendix

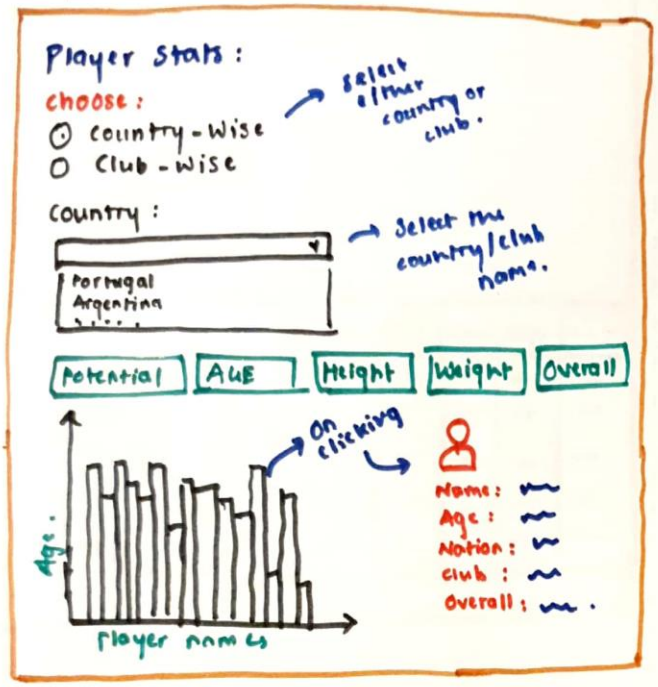
Design Sheet 1 :

Design Sheet 2 :

SHEET 2.

INITIAL DESIGN-1.

Player Stats/Performance.

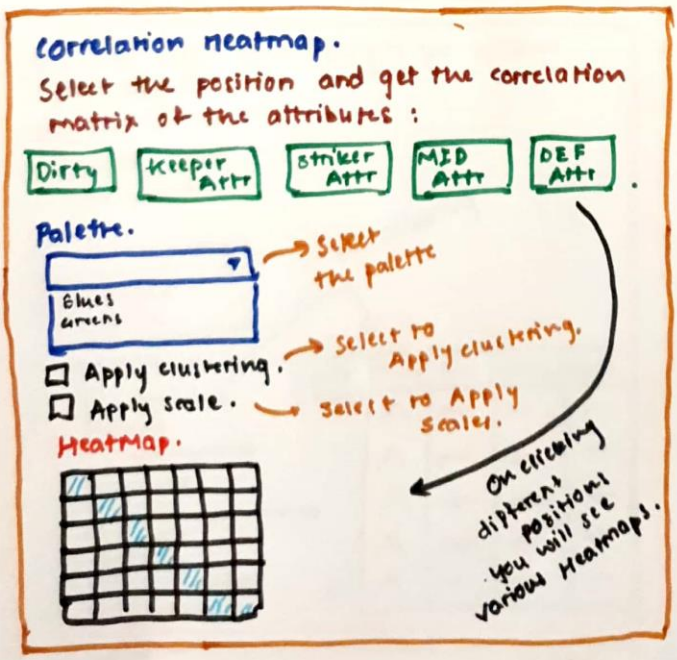


On clicking the first tab : We will get to see all the various player stats. If we select, the country/club. Then we select the country name/club name from the drop box. Then we select, any action button, for seeing bar graphs of age of every player in the team.

Then on clicking each bar, we get to see, the player details. His Age, Nationality, club name, overall rating. You will also see the photo of the particular player.

For this, we are able to see, how the stats of the different players vary, in the country team / club team.

Attribute Correlation Heatmap.



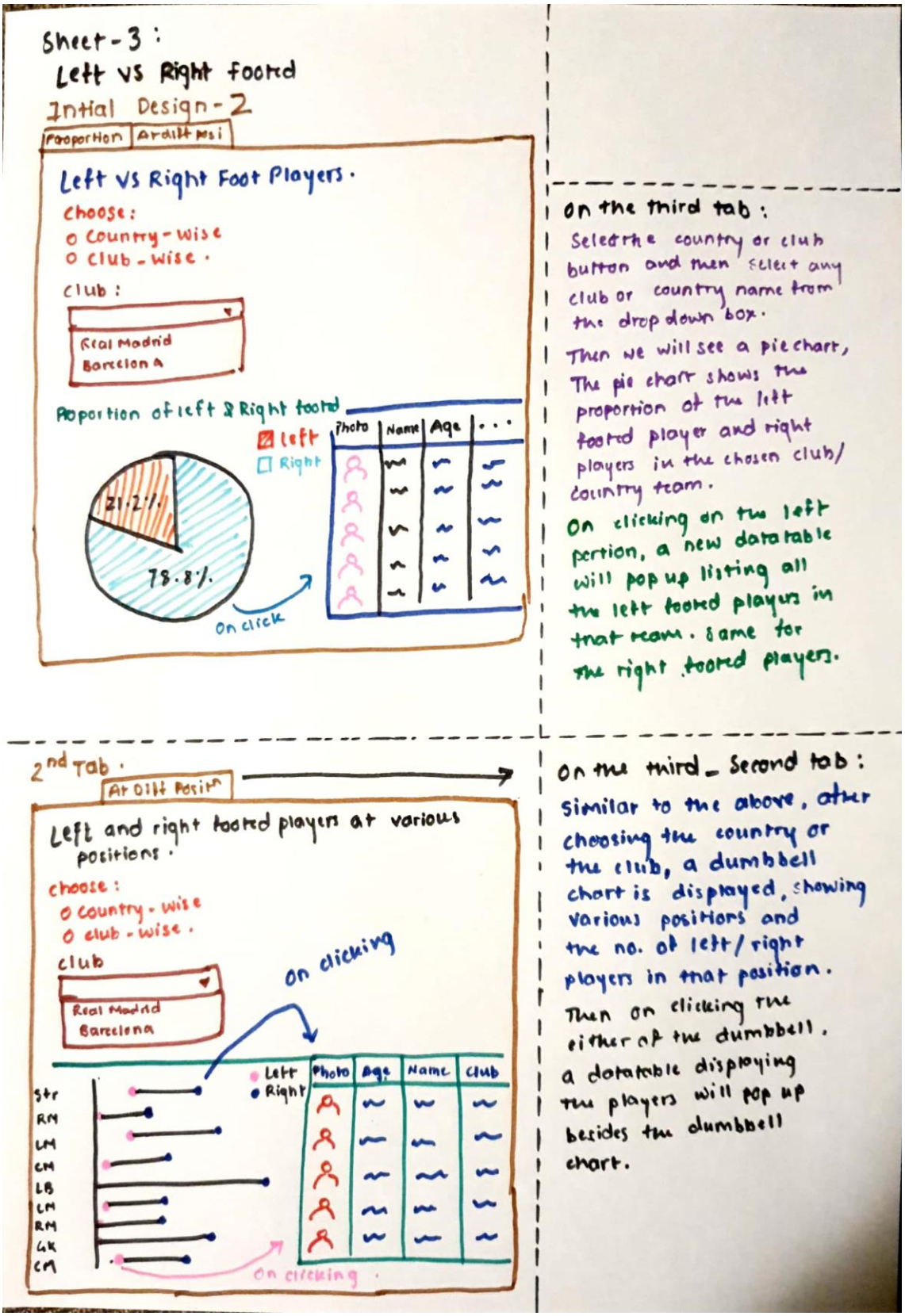
On the second tab, You will see the correlation of attributes at various positions on the field.

Then select the different palettes from the dropbox called palette.

Then select, different positions to see respective position attributes.

Then you can see the heatmap. You can cluster the attributes by using the apply clustering select box and apply scales using the Apply scale select box.

Design Sheet 3 :

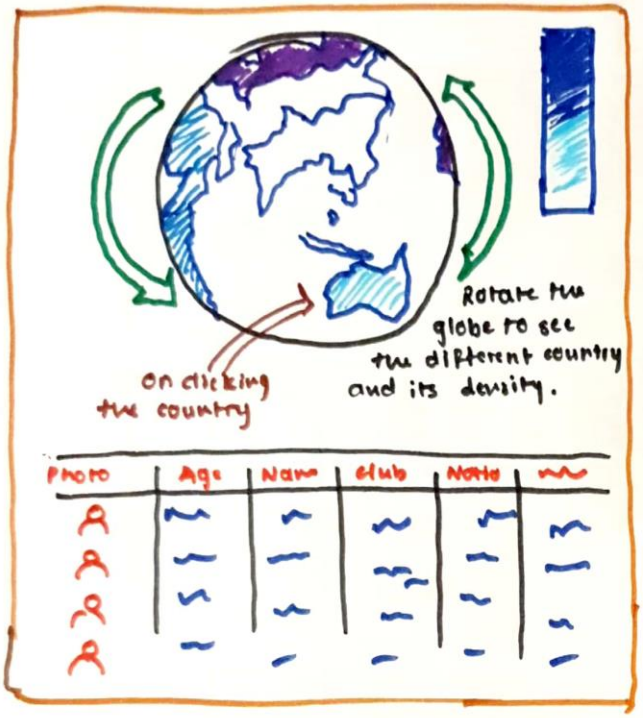


Design Sheet 4 :

SHEET-4 :

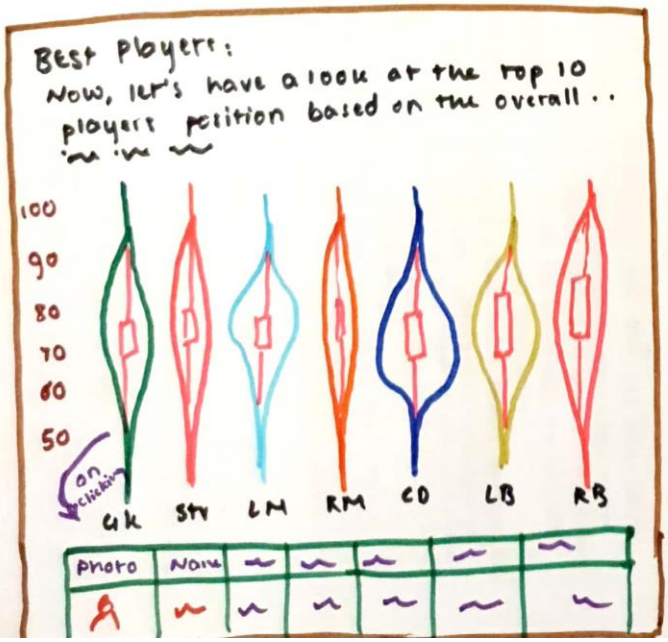
INITIAL DESIGN-4 :

Density of the players all over the world.

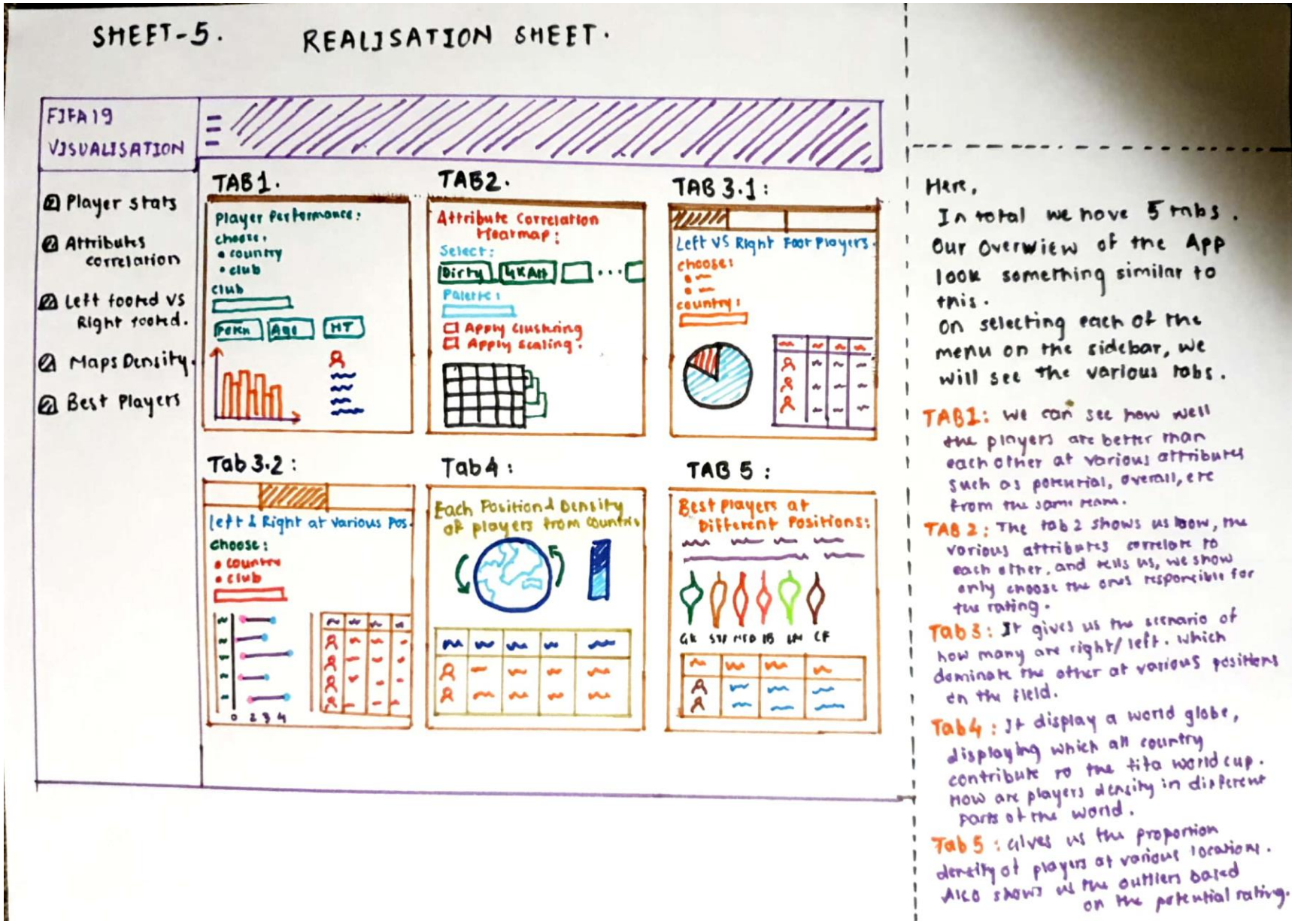


In the 4th tab :
We will see, the world globe, with is rotatable, you can rotate, it to see, the density of the players from various countries.
On clicking any country, A new datable will pop up showing the the information of the players in that country.

Best players at different positions :



In the 5th tab :
Here, we will see, violin + box plot where, the boxplot is within the violin plot.
We can notice various outliers by hovering on the dots.
We also notice the density of the play players at the various position.
On clicking anywhere on the violin respective position player's information will pop-up in the datatable below.



Design Sheet 5 :