



## DATA VISUALISATION PROJECT

### FIFA DATASET VISUALISATION

**Project by:**  
Surbhi Sanjay Peshwe  
30060567

## Table of Contents

<i>Introduction</i> .....	<b>3</b>
About the Dataset.....	3
Exploration and Narrative .....	4
Target Audience.....	4
<i>Visualisation Design Process</i> .....	4
<i>Five Sheet Design</i> .....	<b>5</b>
Sheet 1: Brain Storm .....	5
Sheet 2: Initial Design 1.....	6
Sheet 3: Initial Design 2.....	7
Sheet 4: Initial Design 3.....	8
Sheet 5: Realization Sheet.....	9
<i>Implementation</i> .....	<b>10</b>
Design.....	10
Libraries Used .....	11
<i>User Guide</i> .....	<b>12</b>
Exploring Visualisation.....	12
<i>Conclusion</i> .....	<b>15</b>
<i>Reflection</i> .....	<b>15</b>
<i>References</i> .....	<b>16</b>

## Introduction

Being a football fan, I thought to explore some FIFA data by analysing various trends in the game such as how various factors are correlated with each other and how do they contribute towards the teams' success and players success. I have used descriptive statistics for the analysis of my objectives. But before that, it is required to filter the data to make it compatible with R, after which then data is cleaned for missing values.

After having done exploratory analysis, it is now time to see the visual understanding of the exploratory analysis. For this purpose, I have used R Shiny along with packages required for the presenting the visualizations.

### About the Dataset

With the above objective in mind, I have chosen a FIFA dataset from the website 'kaggle' (<https://www.kaggle.com/theC03u5/fifa-18-demo-player-dataset>). This dataset contains every players' details who have participated in FIFA, 2018. The dataset has 17981 observation and 70+ attributes which includes attributes such as a players' personal data, players performance and even player position data in the game.

**Player Attribute Data.csv:** This dataset consists players performance details such as player ID, overall, potential, Aggregation etc.

**Player Personal Data.csv:** This dataset contains player personal data such as name, nationality, club, age, wage, value etc.

**Player Playing Position Data.csv:** This dataset contains player positions in game and rating for the allocated positng.

It is a tabular data set which has different types of attributes like spatial and temporal as well as textual attributes.

After integrating and combining the datasets with the required columns and attributes, I have created a "Complete dataset" for further analysis and visualizations.

The attributes used for the purpose of the visualization can be divided into different categories for understanding their meaning:

- About the players: This includes their name, age, nationality, overall and potential scores, their market value, the salary they get paid, etc.
- About their skills: This includes factors contributing towards their success points such as acceleration, aggression, agility, balance, ball control, composure, crossing, curve, dribbling, interceptions, long shots, etc.
- About the positions: These include four key positions of players on the field which are goalkeeper, mid-fielder, defender and forward. Data about these positions is given in the complete dataset.

All these attributes are integrated in the further ahead for narrative visualizations.

### Exploration and Narrative

There were a number of findings based on the exploration of the dataset statistically while giving a visual-explanations about it. From the findings, I wish to combine and integrate them to further analyse the visual appeal and understanding of the FIFA 18 dataset.

Further ahead, the narrative of the visualization is something like the follows:

- ⇒ The first tab introduces the user to the dataset and what all is in it.
- ⇒ The second tab is about the player details which uses the data to display players information in the form of comparisons using visual representations focusing on their market value and players skills.
- ⇒ The third tab is based on the location of top players for all countries present in the dataset.

The aim of this narrative is to give the user views and understanding about the players present in FIFA 18.

### Target Audience

The target audience for this project can be the following:

- different players looking out for their competition
- coaches of teams wanting to strategize games
- fans of teams or players
- news agencies or bloggers wanting to gather information before writing an article or blogs
- game developers like playstation ps4

### Visualisation Design Process

The visualisation process implemented a five design sheet methodology where a number of alternative designs were considered for a narrative visualisations. The design consideration are discussed in the following sections.

# Five Sheet Design

## Sheet 1: Brain Storm

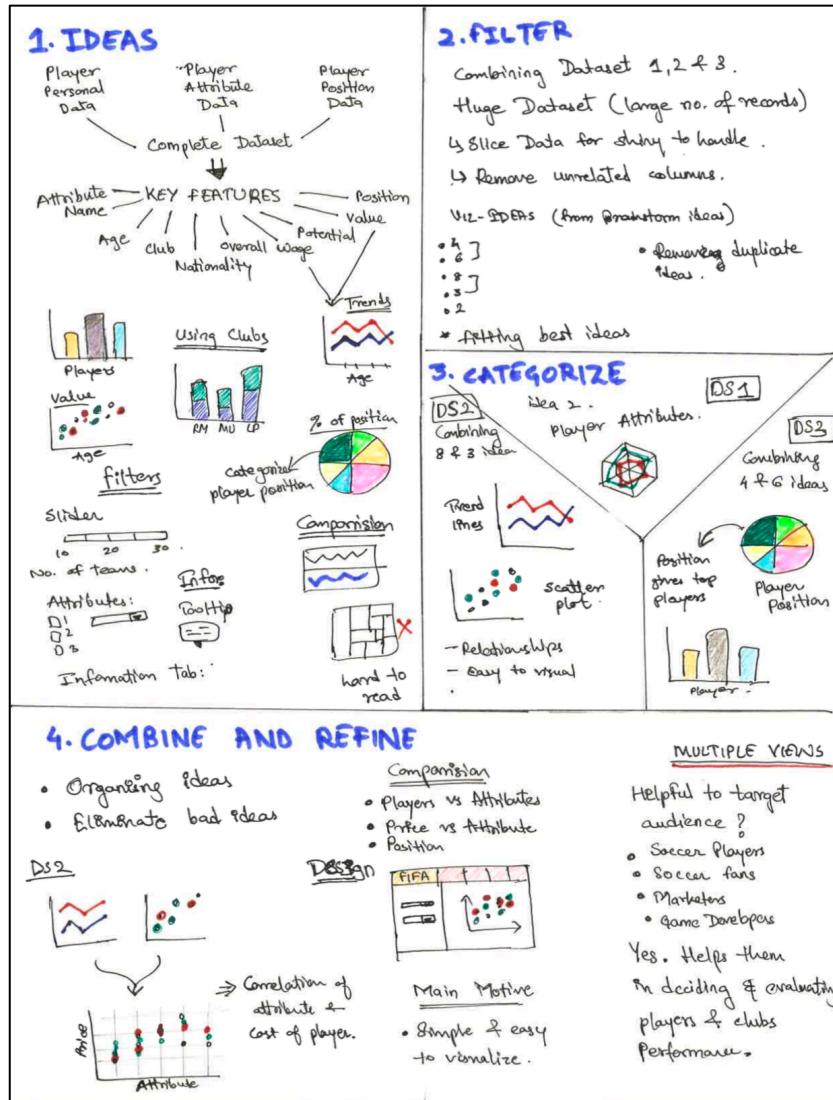


Fig 1: Sheet 1

- This is the first sheet where all the initial ideas have been taken into consideration despite of being feasible or not.
- **Generate Ideas:** From the above sheet, from all the possible ideas few of them are illustrated in the Ideas section. Most of them were simple ideas like a line graph, bar graph and the type of controls to be used and ways to prepare shiny layout.
- **Filter:** In filter section, similar ideas have been grouped together to form a better visualisation where similar ideas/ duplicates have been removed. Out of 10 ideas, ideas 4 and 6 and idea 8 and 2 were grouped and idea 2 which was unique have been considered for visualisation.
- **Categorize:** The ideas which are grouped in filter section are being categorized to check how effective the visualisation turns out to be in combination of two ideas.
- **Combine and Refine:** The final ideas are now combined, and duplicate ideas are removed. The result is a high-level picture of narrative and easily accessible interactive visualisation as seen under Combine and Refine.
- **Question:** In the question section we are checking how it is benefiting our target audience.

## Sheet 2: Initial Design 1

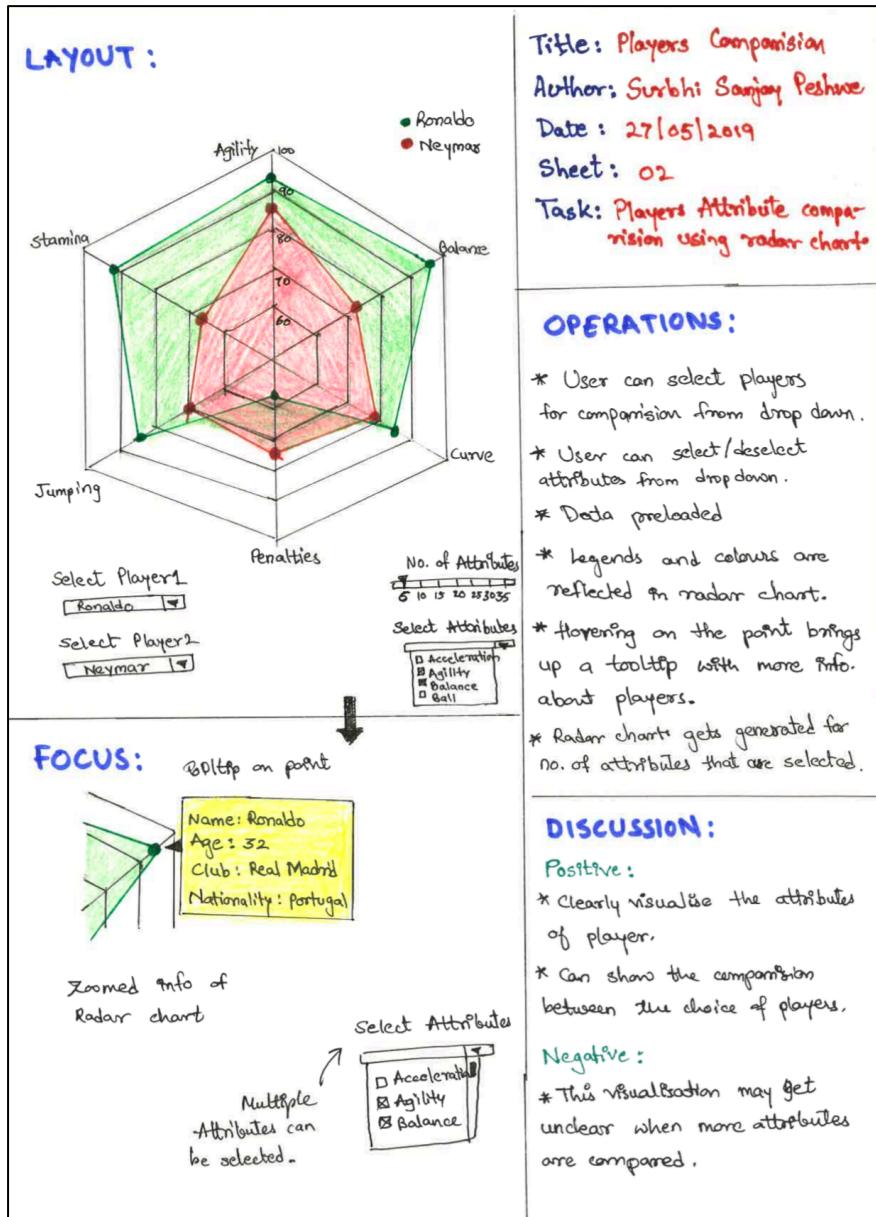


Fig 2 : Sheet 2

- **Layout:** One of the ideas was basic layout using Radar chart to show the comparison between players. This way players attributes can be seen in chart and figure out the attribute score difference between selected players. The players are visualised using different colours.
- **Focus:** From Sheet: 2 As user hovers on one of the attribute web point in radar chart, a tooltip appears showing the players details along with the attribute score.
- **Operations:** There are controls for player and attributes on the UI which enables user to select the number of attributes.
- **Advantages:** It is user-friendly and informative. Everything can be understood by one glance at the visualisation.
- **Disadvantages:** Hard to implement and time taking. For many attributes, visualisation can be unclear.

### Sheet 3: Initial Design 2

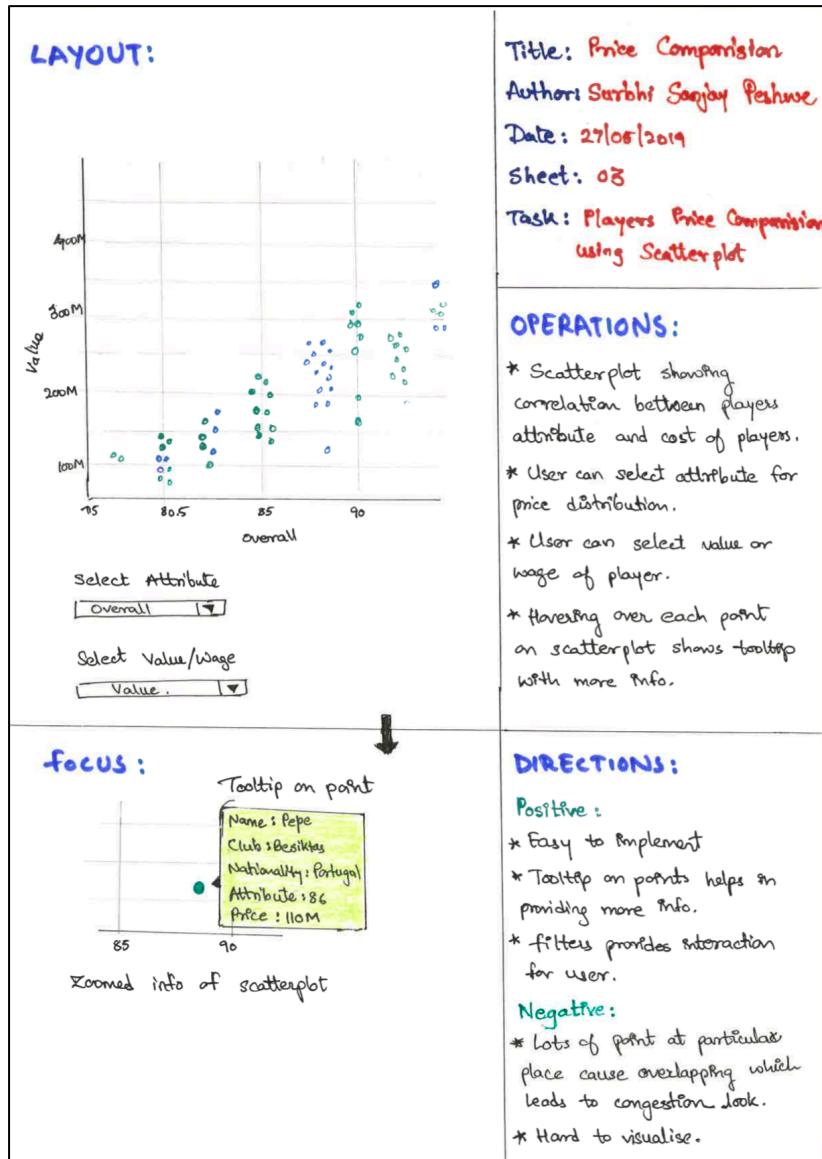


Fig 3: Sheet 3

- **Layout:** Next idea was to design is to plot a scatterplot of attributes vs price of player. Idea is to visualise the correlation between the attributes of value/wage of player. This could be done for all the FIFA players.
- **Focus:** Scatterplot would be the key visualisation as tool tip at each point gives detailed information. (Ex: Player, Club, Nationality, Price etc).
- **Operations:** User can select required Attribute and Price (Value/Wage).
- **Advantages:** It is an easy design to implement using R Shiny comparatively. Growth can be visualised clearly. Tool tip implementation on point would help get required details. This could be considered for final design.
- **Disadvantages:** As we have lot a of Players, it, would be difficult to accommodate them on scatterplot and even if all of the are plotted then there will be problem of overlapping of points which will not be clear for visualisation.

## Sheet 4: Initial Design 3

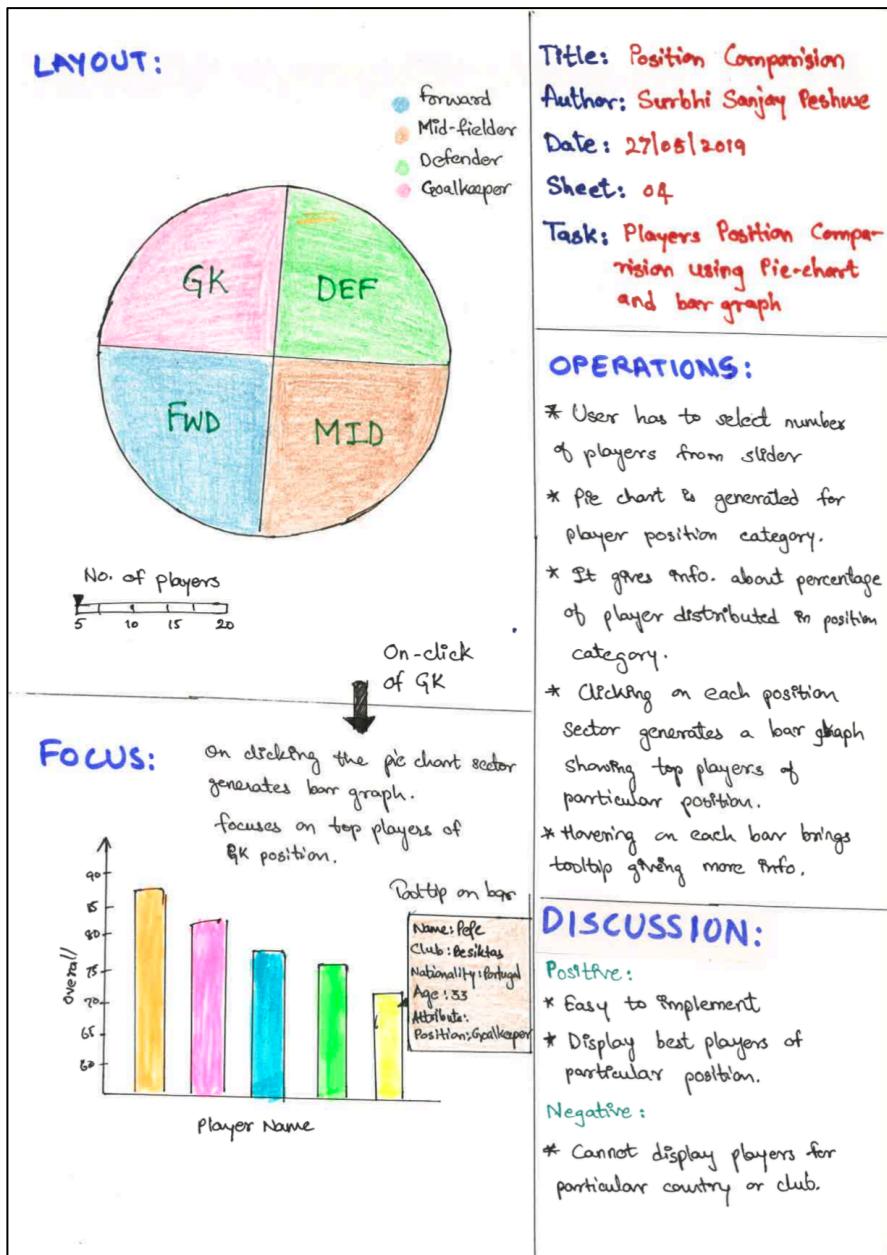


Fig 4: Sheet 4

- **Layout:** Another idea was to show the pie chart for player positions where positions are categorised into 4 categories.
- **Focus:** From sheet 4, as user clicks on one of the positions in pie chart, a bar graph appears showing the count in each player in that position.
- **Operations:** There are controls for number of players on the UI which enables user to select the number of players to be displayed in bar plot. Hovering on brings up the tool tip with more information.
- **Advantages:** Highly interactive visualisation. It is an easy design to implement using R Shiny comparatively.
- **Disadvantages:** Not an interactive visualisation as user cannot select club or country.

## Sheet 5: Realization Sheet

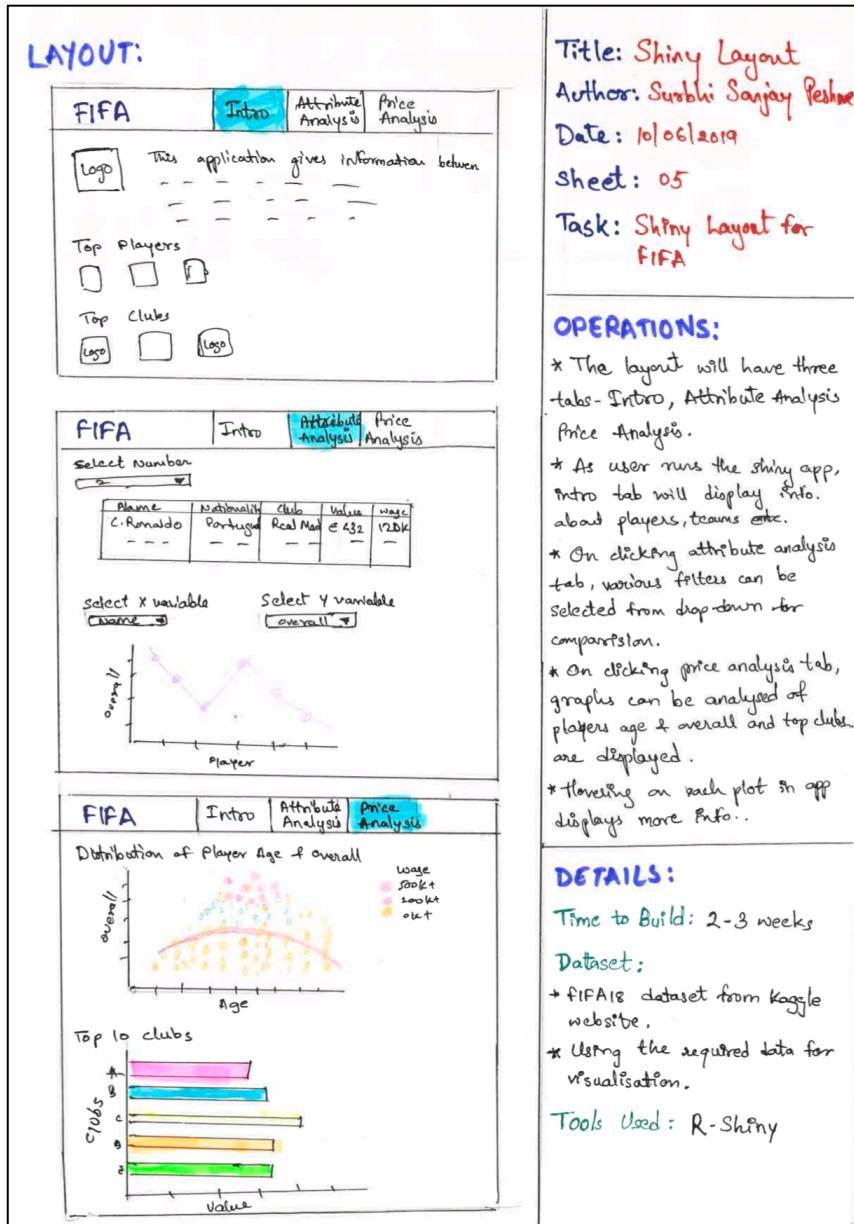


Fig 5: Sheet 5

- Design:** There are three tabs i.e introduction, attribute analysis and price analysis. For visualisation, line plot, scatter plot and horizontal bar graph give a better insight. Few of the initial ideas in final layout were not considered such as design 1 had issue during implementation and design 3 was not included players distribution was shown with other factors.
- Algorithm:** Dataset is cleaned and wrangled and required packages are installed. Data has been explored for visualisation and integrated for final design. Since it is a huge dataset, it takes time to run the application.
- Tools Used:** R Shiny
- Time Required:** 2 to 3 weeks

# Implementation

## Design

Below is the final design layout of my application which provides the narrative visualisation on FIFA.

It has three tabs:

1. **Introduction:** A brief description about FIFA stating the history and record is shown. Upon scrolling down the page, top players based on their overall performance and top clubs based on the value of FIFA 18 are displayed along with their images and club logo.
2. **Attribute Analysis:** This tab has statistical table of players and comparison plot of players attribute. The plot give clear attribute comparison of top 10 players as it has control for choosing attributes.
3. **Price Analysis:** This tab consists of scatter plot and a horizontal bar plot in a single page. Scatter plot depicts the players wages based on their age and overall performance in the game. Horizontal bar plot shows the top 10 valuable clubs.

### **Attribute Analysis:**

At the beginning of the page, the tab has a statistical table of players which includes players details such as Name, Club, Nationality, Value, Wage, Overall and Potential. There is an option for user to choose number of players to be displayed in the table. This table shows the players details in descending order of their overall score. This table helps the viewer to have a look at players details at one glance.

This plot has been plotted for top 10 players since it is not efficient to plot for all the players. There is an option for user to choose the comparison attribute and user can see the comparison of various attributes such as acceleration, balling, jumping, diving etc. for these players. Based on user selection, plot will be generated.

The aim of this plot to give the target audience a view on where these top 10 players attribute varies and comparative who is best at particular attribute. Based on this analysis, they can predict as to which player will score well in the given situation.

### **Price Analysis:**

For the plot which are plotted in this tab required some data preprocessing. For visualisation purpose, I have changed the entries of wage and value into actual

currency value by using a function, where I have removed currency symbol and converted the value into thousands and millions.

A scatterplot has been plotted which depicts the relationship between the overall potential of the individual player with their age keeping legend as wage bracket. User can hover on the point for detailed information such as player name, club, nationality etc. that is shown in Fig 10. This plot plays an important role in getting insight on how an overall attribute of the player affects his wage/value. In addition to that, a smoother has been plotted which shows the decrease of overall value with age, most of the players with age more than 40 has low overall. Practically it shows players efficiency get decrease as they cross age 40.

It can be seen that highest wages are earned by the players whose overall score is above 85 and age is approximately 30. This is justified as player gains experience and improves their skills eventually.

Target audience can get an idea of players health capacity and their performance according to their age. This shows that how age affects the overall performance and simultaneously on their wages.

The bar plot show the top 10 valuable clubs. The players values from each club were added up to calculate the clubs value. Players valuation is very similar to the players wages so it can be seen that top clubs are from the clubs which has players with highest wages.

## Libraries Used

**Below are the libraries that have been used for design implementation:**

**shiny** -> used to build application

**shinythemes** -> to add themes to the application

**shinywidgets** -> to user control

**dplyr** -> for data preprocessing

**tidyR** -> for data preprocessing

**ggplot** -> for plotting the graph

**ggthemes** -> to add themes to ggplots

**plotly** -> for various interactive plots

**DT** -> for displaying stats table.

# User Guide

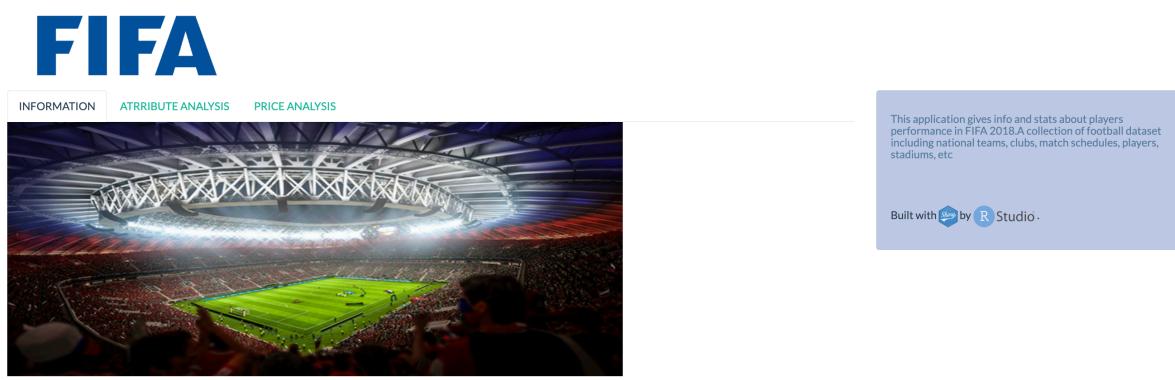
For accessing the visualisation, please follow the below instructions:

## Instructions:

- Make sure all the libraries have been successfully installed.
- Please keep patience as it takes time to load the data because of which app takes time to run.
- Do not interrupt the app before plot appears on screen.
- Please ignore the warnings which appears while running the app.
- For the statistical table please select number greater than 1 i.e minimum one player has to be displayed in the table.
- Please wait for the third tab i.e price analysis tab to load as it takes time. If the plots are in very small size, corned to right and not in readable format then please reload the app.

## Exploring Visualisation

The below image is the homepage of the app displaying information about FIFA and upon scrolling down the page it displays top players and clubs.



The 2018 FIFA World Cup was the 21st FIFA World Cup, an international football tournament contested by the men's national teams of the member associations of FIFA once every four years. It took place in Russia from 14 June to 15 July 2018. It was the first World Cup to be held in Eastern Europe, and the 11th time that it had been held in Europe. At an estimated cost of over \$14.2 billion, it was the most expensive World Cup. In this application we will get deeper insight about the players performance and their wages for FIFA 18 and clubs valuation in FIFA 18

Fig 6: Introduction Tab(Part 1)

### TOP PLAYERS IN FIFA2018



### TOP FOOTBALL CLUBS



Fig 6: Introduction Tab(Part 2)

Next by clicking on attribute comparison takes you to the page where first players statistical table is displayed and later the attribute comparison plot. On clicking the dropdown option of “select number of players”, it displays the top n players in table.

Name	Club	Nationality	Value	Wage	Overall	Potential
Cristiano Ronaldo	Real Madrid CF	Portugal	€95.5M	€565K	94	94
L. Messi	FC Barcelona	Argentina	€105M	€565K	93	93
Neymar	Paris Saint-Germain	Brazil	€123M	€280K	92	94
L. Suárez	FC Barcelona	Uruguay	€97M	€510K	92	92
M. Neuer	FC Bayern Munich	Germany	€61M	€230K	92	92

Fig 8: Attribute Analysis tab – Part 1

Further in attribute analysis tab, by selecting attributes from the dropdown menu it generated the plot of that particular attribute for the top 10 players.

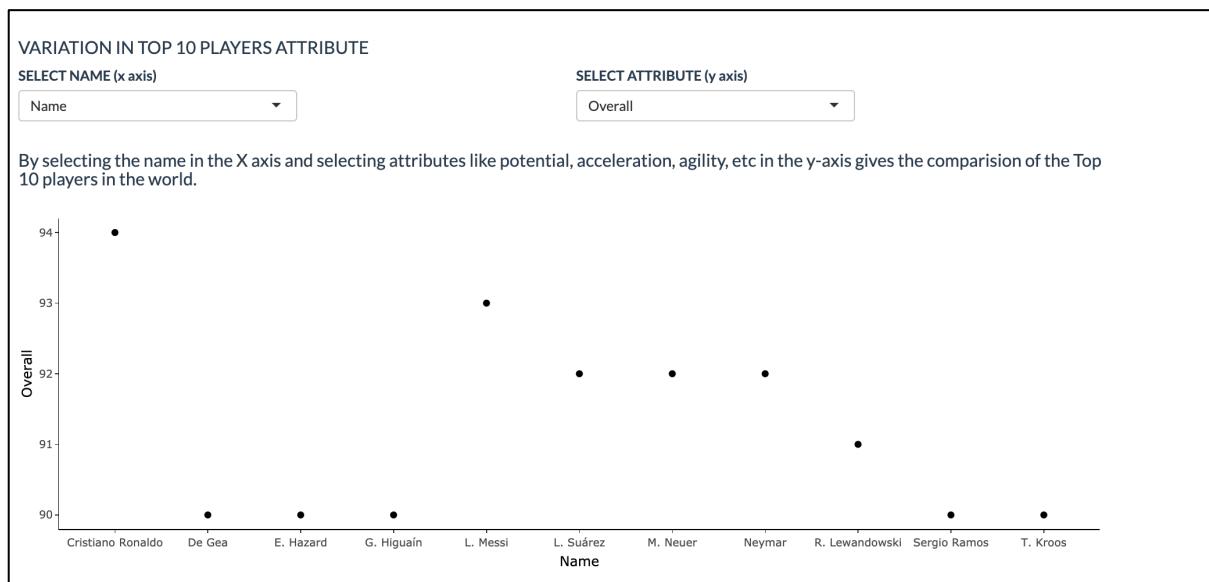


Fig 9: Attribute Analysis Tab – Part 2

Clicking on price analysis tab, it takes you to page where players and club value and wages has been analysed.

The first plot shows the players distribution based on age and overall performance along with wages players are being paid.

Upon hovering over the point players details can be seen. Legend can be referred for the points colour.

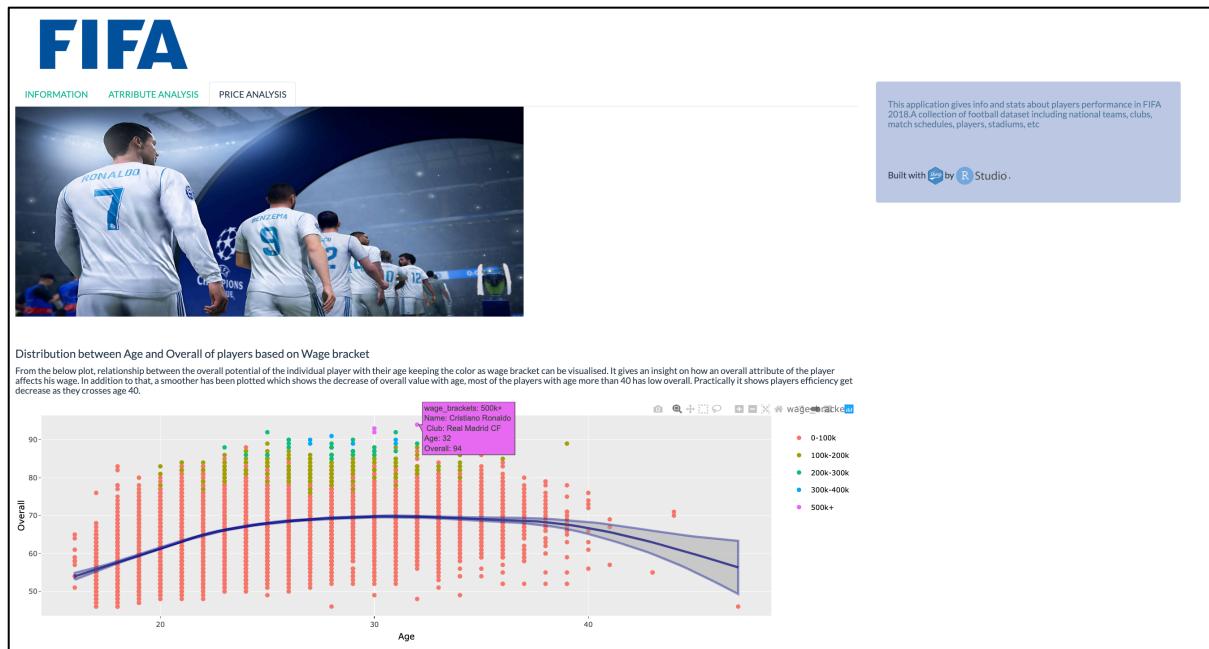


Fig 10 – Price Analysis Tab – Part 1

Further scrolling down the page, top 10 valuable clubs are displayed where by hovering over each bar give the information about the total value of the club and the club name.

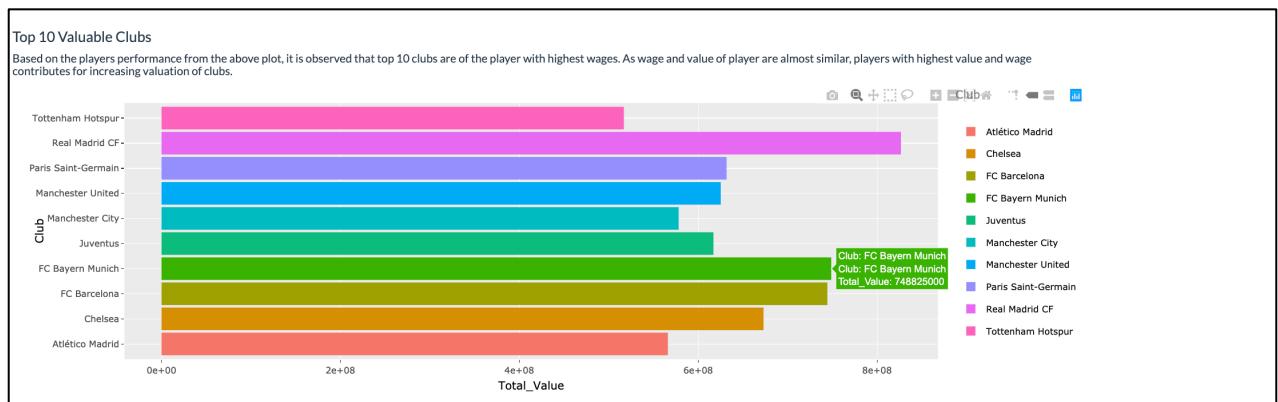


Fig 11 – Price Analysis Tab – Part 2

## Conclusion

This project was build in a narrative visualisation way so that it could be helpful to the target audience for better insight in FIFA. The main objective of my project was to help the target audience where they can analyse the data for their job needs and gain a valuable information. Such as players can compare themselves with others, game developer can build the game according by giving power to the players, news reporters can analyse the visualisations and give reasons for performance etc.

Plots are made for proper visualisation purpose, where trends can be seen and analysed. This can help in further improvement of game for the players and also keeping them aware from their opponents strength and weaknesses.

Common people can look at the app and have a background information about their performances wages, value in market etc.

## Reflection

- From this project, I have learnt a lot about R Shiny and various functions and libraries in R. Working with R Shiny was challenging. For checking the relevancy of visualisation, continues efforts were required as it takes lot of time to test. It was made sure that complete focus
- This project helped me to explore into visualisation where it is noticed that infinite effective visualisations can be made. In this complete process I have learnt many things and here I have mentioned few of them.
- In process of this application, I have learnt this lesson about importance of planning for final interactivity for any visualisation project. Since I have plotted each visual element separately, at the time of integrating all these chunks it was found that there were inconsistencies in variable names and some overlaps in data manipulation.
- Another issue during integration process was about the size of layouts. While integrating all visual elements into final visualisation, I had to resize all the elements. During this process I discovered fluidpage and fluidrow function because of which it fits to the windows screen accordingly. Using this element, I overcame the size problem.
- Learned that Html tags can be used in Shiny various interactive visualisation.

## References

app, R. I. (n.d). *Render Images in Shiny app*. Retrieved from Shiny from RStudio:  
<https://shiny.rstudio.com/articles/images.html>

RStudio, S. f. (n.d). Retrieved from Shiny from RStudio:  
<https://shiny.rstudio.com/reference/shiny/1.0.2/tabPanel.html>

StackOverFlow. (n.d). *StackOverFlow*. Retrieved from  
<https://stackoverflow.com/questions/34605919/formatting-mouse-over-labels-in-plotly-when-using-ggplotly>

FIT5147 Data Exploration and Visualisation modules.