CSE 564 PROJECT PRELIM REPORT (Group 26)

Gautham Gunapati - 113274189
Tejeshwar Reddy Gurram - 113271289

1. Abstract

The main objective of our project is to develop an interactive dashboard which shows the analysis of global football players data. We intend to address a few interesting questions on

- Distribution of players over the world
- Distribution of players in a league
- Distribution of players based on attributes, **position**, **age**, **wage**, **rating**.

Inorder to achieve this we are using various interactive plots which are discussed later in the Implementation Section.

Our ultimate goal is to make the user more interactive and find the players of desired characteristics like age, rating, value, position of player (forward, midfielder, Goalkeeper). This can also be used as a talent scout dashboard.

2. Data Preparation

In this section, we discuss the steps we followed in getting the final data which is used in the Dashboard.

2.1. Dataset

For the data visualization we have chosen a time series data over a span of 8 years (2016-2022). There are 8 csv files (each for one year). Our Dashboard visualizes the data of all the 8 years with an option for the user to filter the years.

Source:https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset

2.2. Data Merging

We have merged the data with **player_id** as the key, We have replicated other columns of each year in the master dataset. For example, the overall rating of a player from 2015 to 2022 have different columns. The master dataset columns for overall would look as follows:

Overall_15, Overall_16, Overall_17, Overall_18, Overall_19, Overall_20…..

2.3. Data Cleanup

We have removed the rows, where the basic player details are missing like Nationality and league name are missing. Finally, the total number of rows in the cleaned dataset was 16,200

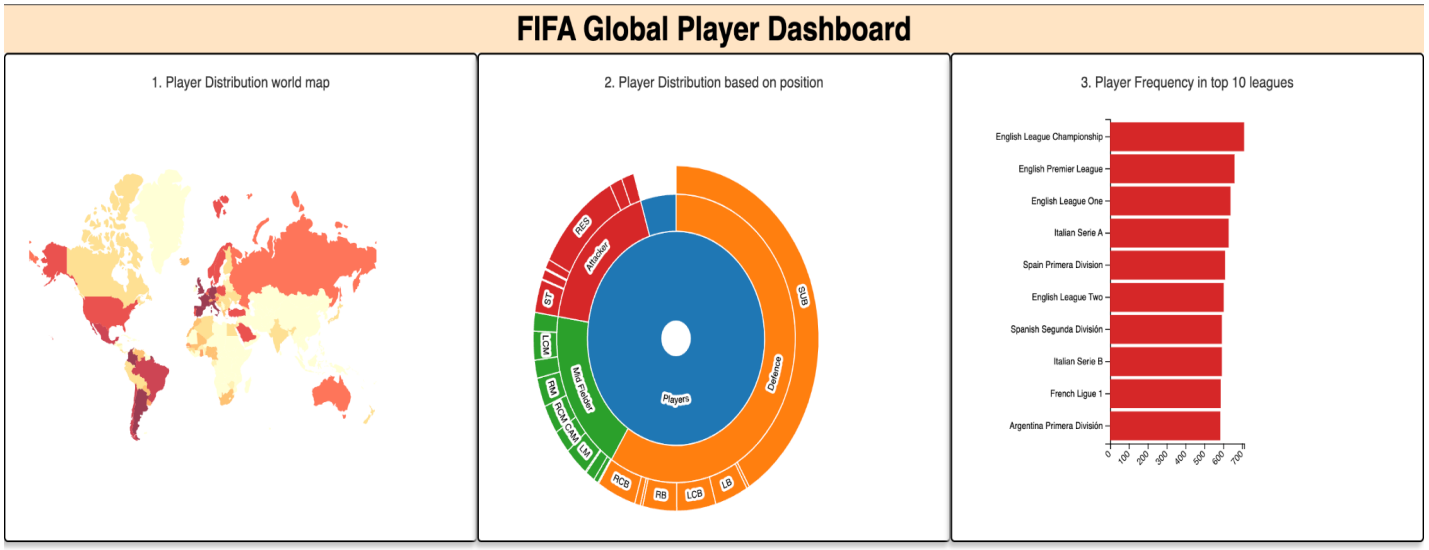
2.4. Data Imputation

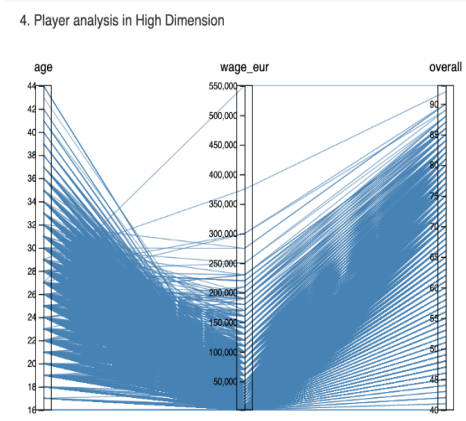
Few attributes like nationality were in a different format than what was required for visualization. We have formatted the country names in such a way that it matches with the data required for visualizing graphs.

2.5. Additional Columns

We have added new columns, which helps the data to organize hierarchically with ease in the sunburst plot. For Example, We have divided the players into Attackers, MidFielders, Defenders, GoalKeepers. This actually was helpful in plotting the sunburst plot and PCP plot

3. Dashboard Implementation



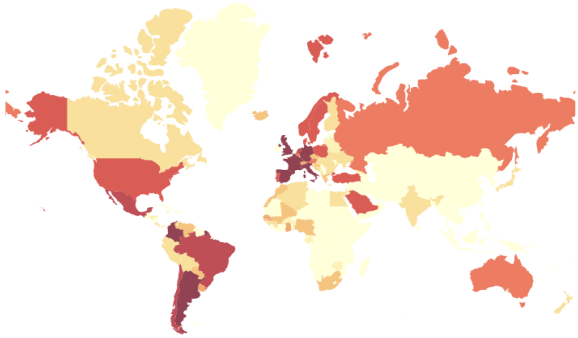


For Implementation of the Web Interface we are using **Javascript** and **D3**. For the backend we are using the **Python Flask server**. Currently, we have implemented 4 graphs. Our Current progress of implementing Dashboard looks as shown in the above Figure. We Discuss our current implementation of each graph in detail in the below sub-sections.

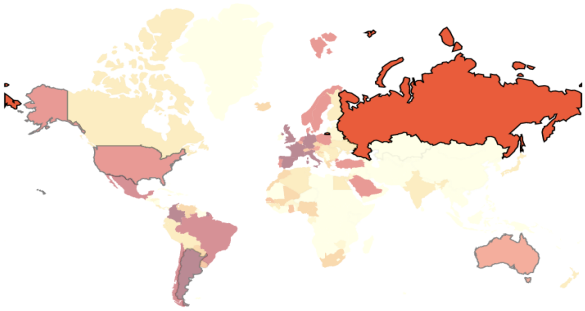
3.1 World Map showing Player Distribution

World map is used to visualize the Frequency of players from each country. We have implemented this by parsing the dataset and binning the frequency of players in each country.

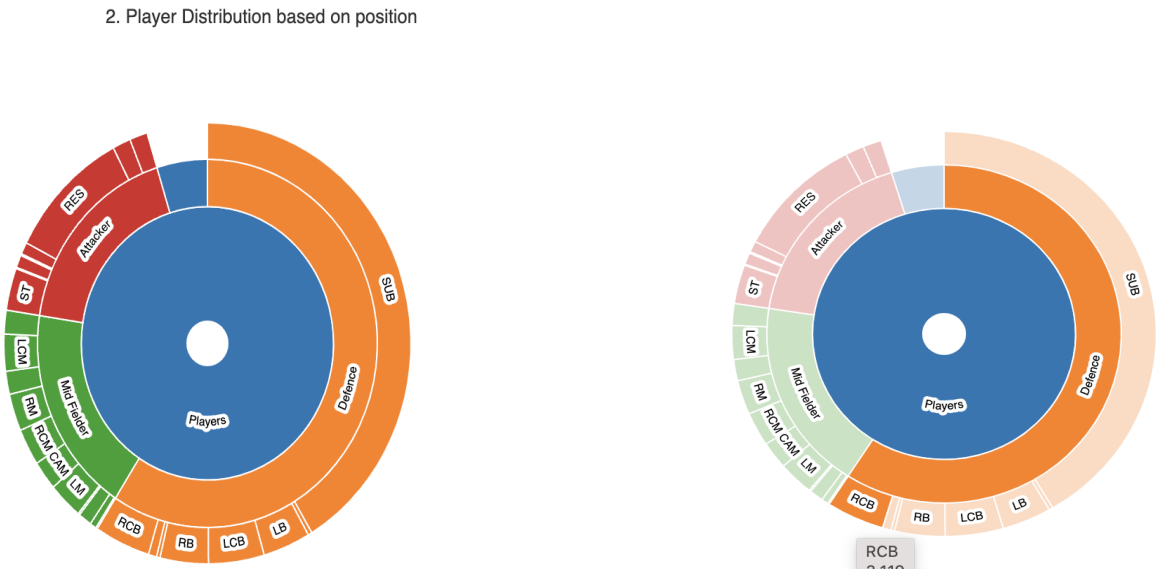
1. Player Distribution world map



We have also added an option to select a country from the world map. The figure below shows the same.



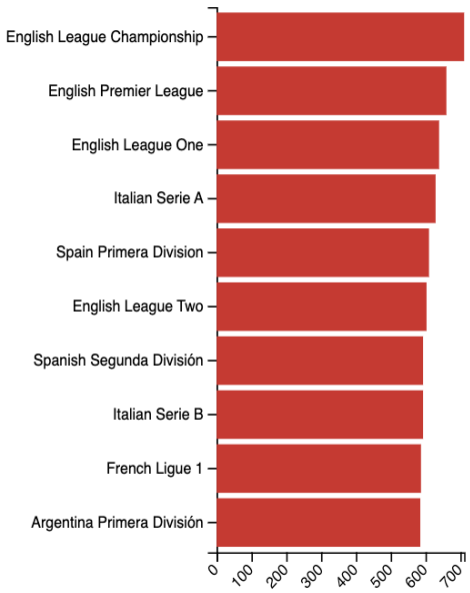
3.2 Expandable Sunburst Plot



We have implemented hierarchical distribution of players based on position. The top level hierarchy has [Attackers, mid-fielders, defenders, Goalkeeper]. Later the first level is further divided into various positions.The plot has following features

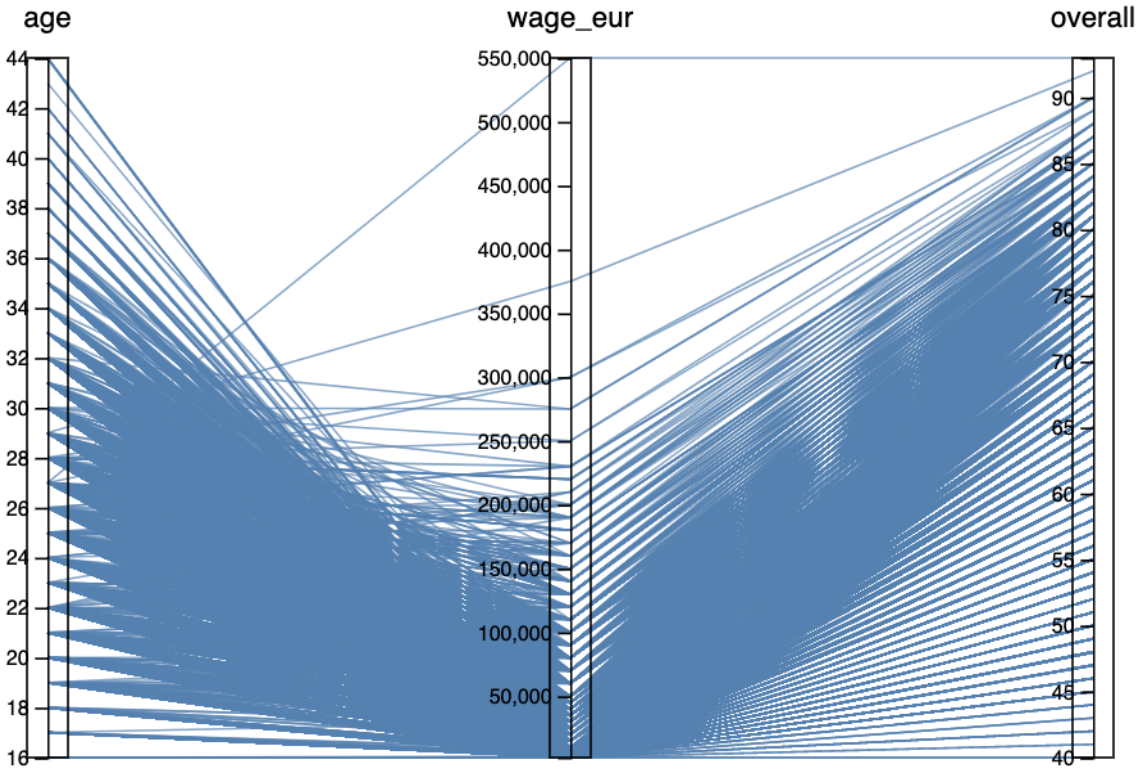
- Expand a sublevel by double clicking it.
- Click in the center to collapse an expanded lower level
- Hover on a sublevel and view its hierarchy

3.3 Bar Chart



We have implemented a basic Bar Graph which shows the top 10 leagues based on player Frequency. FurtherMore, We are planning to add the logos of the leagues in the bar chart.

3.4 PCP (Parallel Coordinates Plot)



We have implemented a partial parallel coordinate plot with attributes like age, salary and rating of the players. Furthermore, we are planning to add more categorical attributes to the plot and enhanced brushing features to the plot

4. Future Work

- Add more graphs to represent data better

We are planning to implement word cloud and stats corner in the upcoming days. Word cloud can be used to visualize the most celebrated player in the chosen time frame. Furthermore, on selecting a player from the word cloud we plan to show the respective stats of the player in the dashboard.

- Add interactions between graphs

After adding word cloud feature, we also plan to add interactions between graphs. Interactions include choosing player position in the sunburst plot and other plots like bar, pcp, word cloud and geomap being updated by the selection made in the sunburst plot. Other interactions include viewing all the players data within a specific time range. Choosing the time range will update all other plots accordingly. We can also select a club from the bar chart and view the respective stats in the other plots. We also plan to filter data based on the data points/ranges chosen from the pcp plot.

- Fix errors and programming bugs

Additionally, we see that there is a lot of scope of error which includes bias while selecting data points from the original dataset. We might also have introduced some programming bugs which we plan to rectify in the future code review cycles. We have already rectified from faulty design conclusions that we made in the proposal document to make it easily scalable to larger data sources. Given that we are handling time series data, there can be multiple ways to represent the data. We are working hard to show correlations in the data by filtering the data point wisely and represent the filtered data using appropriate plots.

- Improve Dashboard Design / Aesthetics.

Finally, we plan to focus on design enhancements and make the dashboard aesthetically pleasing for the viewers by adding good animations and loaders to indicate the data being loaded into the dashboard, since we have a lot of data to parse for every filter applied by the user.

5. References

https://www.kaggle.com

https://www.tableau.com

https://observablehq.com/@d3/zoomable-sunburst

https://d3-graph-gallery.com