CS 6630 - Visualisation for Data Science

PROCESS BOOK
FIFA Viz Play

Project Name : FIFA VizPlay

Aditya Gattu - u1418754@umail.utah.edu [UID : u1418754]

Pranav Madhav Bagepalli - u1471494@umail.utah.edu [UID : u1471494]

Nitish Sai Kumar Kadiyam - u1466435@umail.utah.edu [UID : u1466435]

Project repository : https://github.com/dataviscourse2023/final-project-fifa-vizplay

Background and Motivation. Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

As we all know, Football is one of the most famous sports which is often referred to as the "beautiful game", stands as a universal language, uniting individuals across different corners of the globe. As the sport enjoys an immense following, there exists a substantial group of enthusiasts who are keen to dive deep into the intricacies of the game, exploring beyond just the goals scored and matches won. This project proposal seeks to satisfy this curiosity, offering not just a statistical dive into the world of football but an interactive and engaging visualization of data, enhancing the user's understanding and appreciation of the game.

So our primary motivation behind this project is our genuine interest in the game of football, a sport that brings together people from diverse backgrounds and fosters a sense of community. We as a team thought that we can go beyond traditional metrics and venture into more detailed analyses, thus providing more depth to the narratives surrounding the sport. Inorder to do this we are planning to utilize the vast FIFA player dataset and our aim is to offer users the opportunity to explore and analyze a plethora of variables ranging from player performances, match outcomes, to intricate details like player's physic and skills metrics. The project is also inspired by our intent to show the sports analytics, offering innovative, insightful, and visually appealing interpretations of data that can cater to both avid football enthusiasts and researchers alike.

Project Objectives. Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

The primary objective of our project is to delve deep into the FIFA player dataset to carve out interactive and insightful visualizations that can facilitate a more enriched understanding of the game of football. Our main aim is to address critical questions such as identifying the top players based on various metrics like attack, defense, and shooting scores, and understanding the trends in goals scored across different countries over the years. This project intends to create a dynamic platform where users can explore data interactively, providing a concrete base to football discussions which are often driven by personal opinions.

Furthermore, this visualization venture stands as a beneficial tool for football enthusiasts, researchers, and analysts, offering a data-backed approach to analyzing and appreciating the multifaceted nature of football. By facilitating detailed comparisons and analyses through interactive graphs, it opens up new avenues in the field of sports analytics. Essentially, this project seeks to enhance the community's comprehension of the sport, paving the way for informed, data-driven discourse and fostering a new wave of engagement in the football community.

Data. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

For our project, we are utilizing comprehensive datasets pertaining to the FIFA player database to formulate a rich and detailed analysis. The first dataset harbors an array of information focused on match specifics, encompassing various identifiers such as match_id, team_id, player_id, among others, providing us with an expansive view into the various tournaments and player dynamics within each match. The second dataset, on the other hand, offers a deep dive into individual player statistics and traits, including data columns that outline a player's skills, career trajectory, and market value, facilitating a nuanced analysis of player competencies and standings in the world of football. The combination of these datasets will be instrumental in developing an interactive visualization tool that fosters a deeper understanding and engagement with the sport, by allowing users to explore and analyze player performances and trends in a more immersive manner.

Dataset - https://www.kaggle.com/code/sivsankar/fifa22-recommender-system/input

Data Processing. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

Before we delve into the analysis, we anticipate undertaking a significant data cleanup process to ensure the integrity and reliability of our visualization project. This will involve identifying and handling missing or inconsistent data entries, as well as merging the two

datasets harmoniously by linking relevant columns such as player IDs and team IDs. Moreover, we aim to derive meaningful insights from the data by calculating metrics such as player performance scores based on various attributes, and aggregating goal statistics on a yearly and country-wise basis. The data processing will be implemented using robust data manipulation libraries and tools, allowing us to efficiently filter, transform, and prepare the data for a seamless visualization experience.

Visualization Design :

Potential quantities and insights that we aim to derive from the dataset:

Player Performance Metrics:
- Overall and potential player ratings.
- Player age, height, weight, and nationality distribution.
- Players preferred foot, weak foot, and skill moves.

Player Comparison:
- Comparisons of players based on various attributes (e.g., pace, shooting, passing).
- Clustering or grouping of players with similar skill sets.

National Team Analysis:
- Analysis of players representing their national teams.
- Distribution of players across different national teams.
- Performance metrics of national teams.

Trends and Patterns:
- Changes in player attributes over time.
- Correlations between player attributes and performance.

Visualizations:
- Creating charts, graphs, and visualizations to help convey insights more effectively.

Player and Club Profiles:
- Creating profiles for individual players or clubs that summarize their key statistics and attributes.

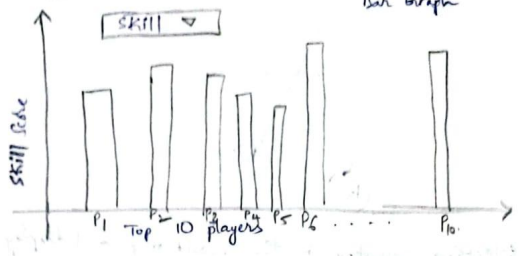
Club-Level Insights:
- Club-wise statistics, such as average player ratings.
- Club budgets based on player values and wages.
- Analysis of club positions in the league.
- Distribution of players across different club positions.

League-Level Insights:
- Analysis of leagues based on the level (e.g., top-tier, second-tier).
- Statistical summaries of leagues, including average player ratings, ages, etc.

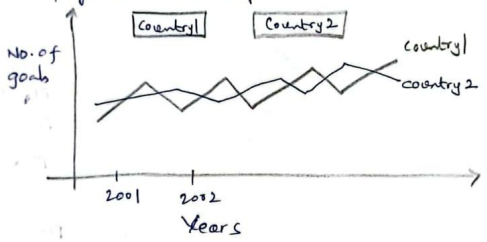
We came up with the initial designs as follows ：

① Top 10 players based on skill worldwide.

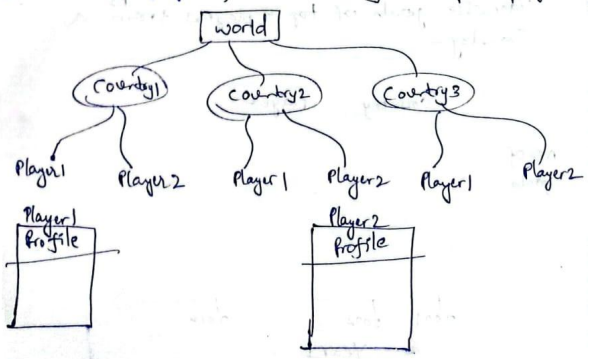
Bar Graph



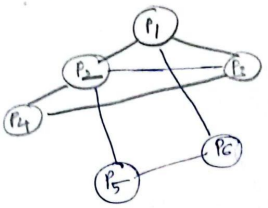
② Comparison of two countries based on the number of goals in worldcups



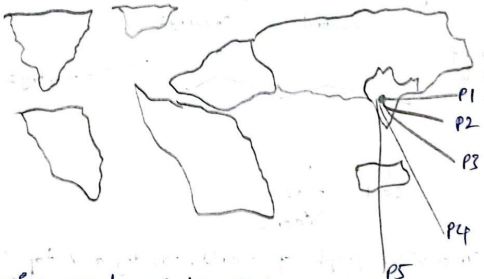
③ Tree view graph of world, country and top 5 players



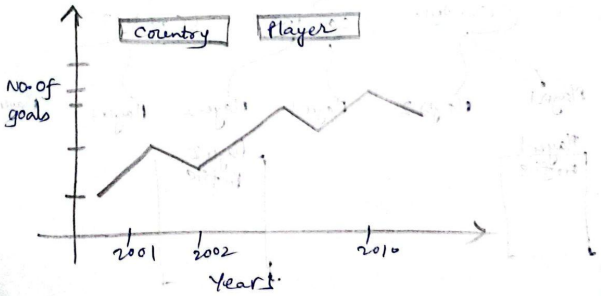
④ Network like graph to study positive and negative interactions between players in a match.



⑤ Selecting country in the world map and display top 5 players from that country :-

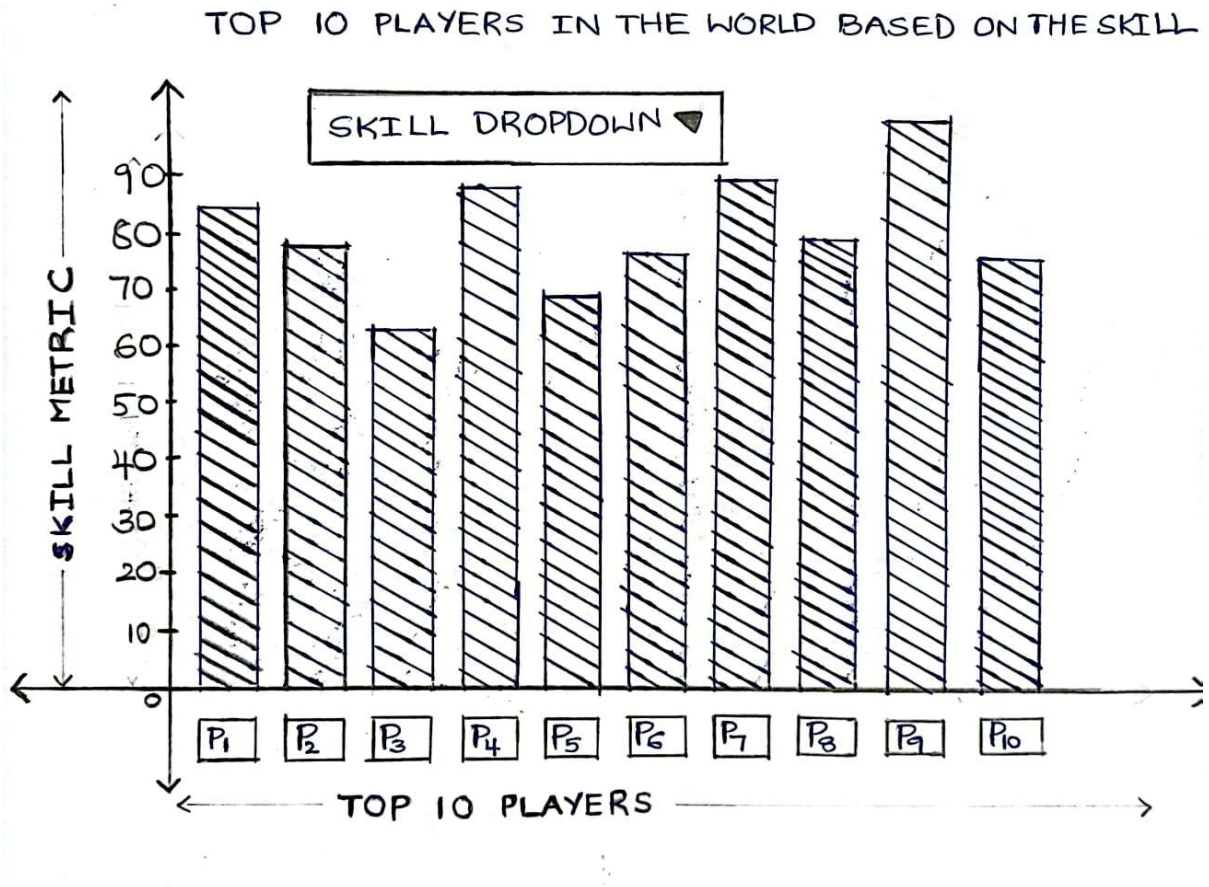


⑥ Yearwise goals of top 5 players from a country :- &

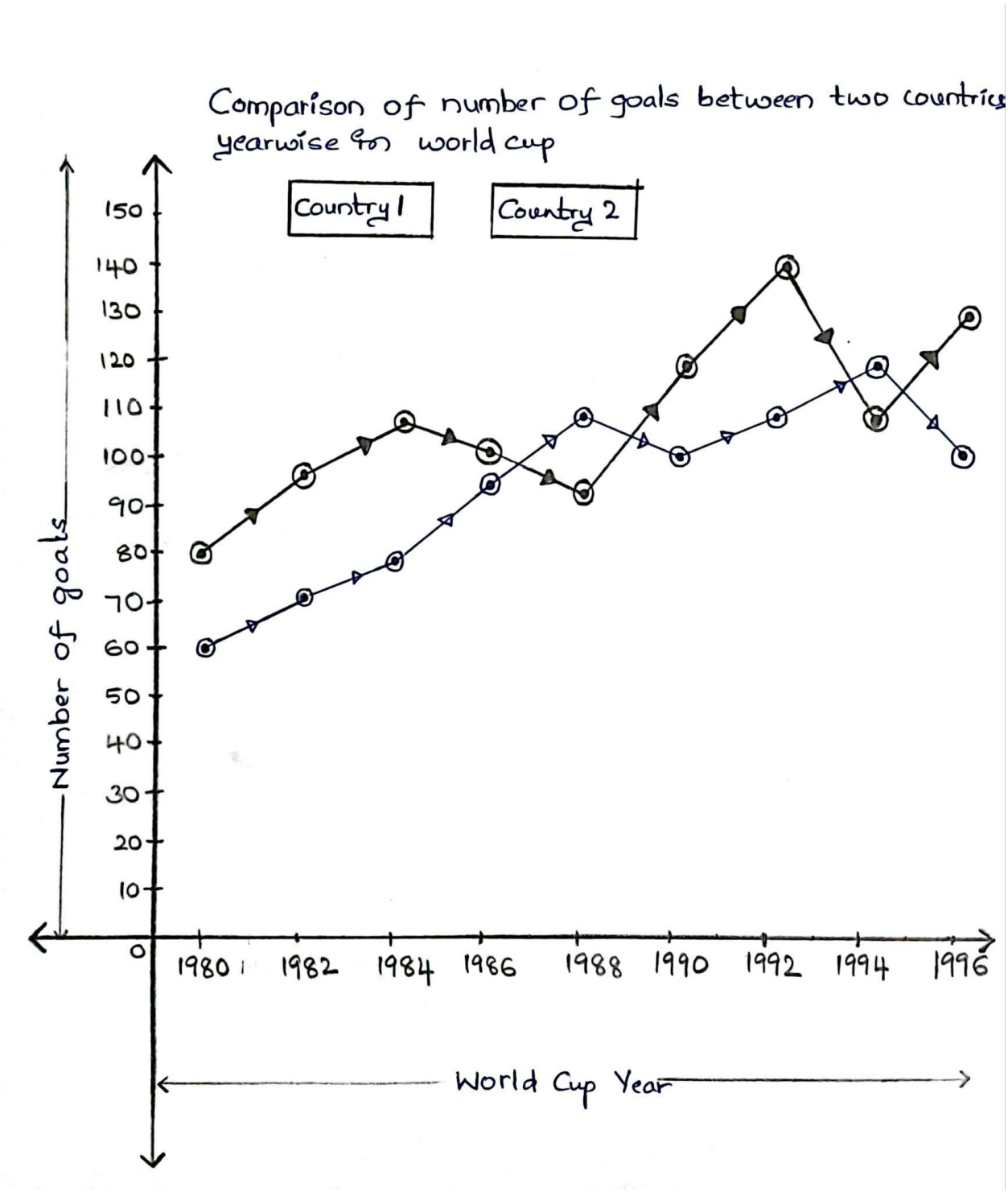


These are our Final Design sketches :

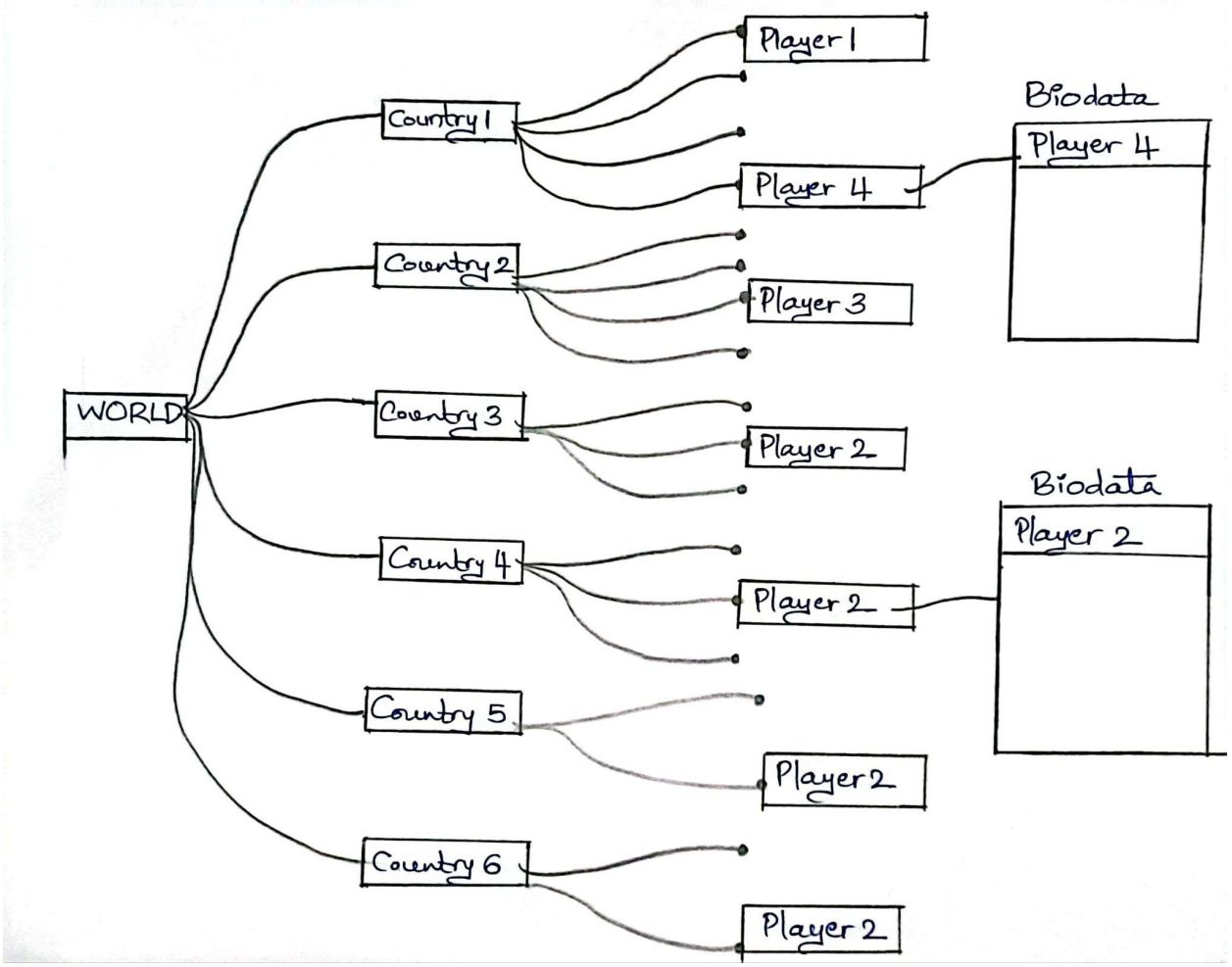
Sketch 1:



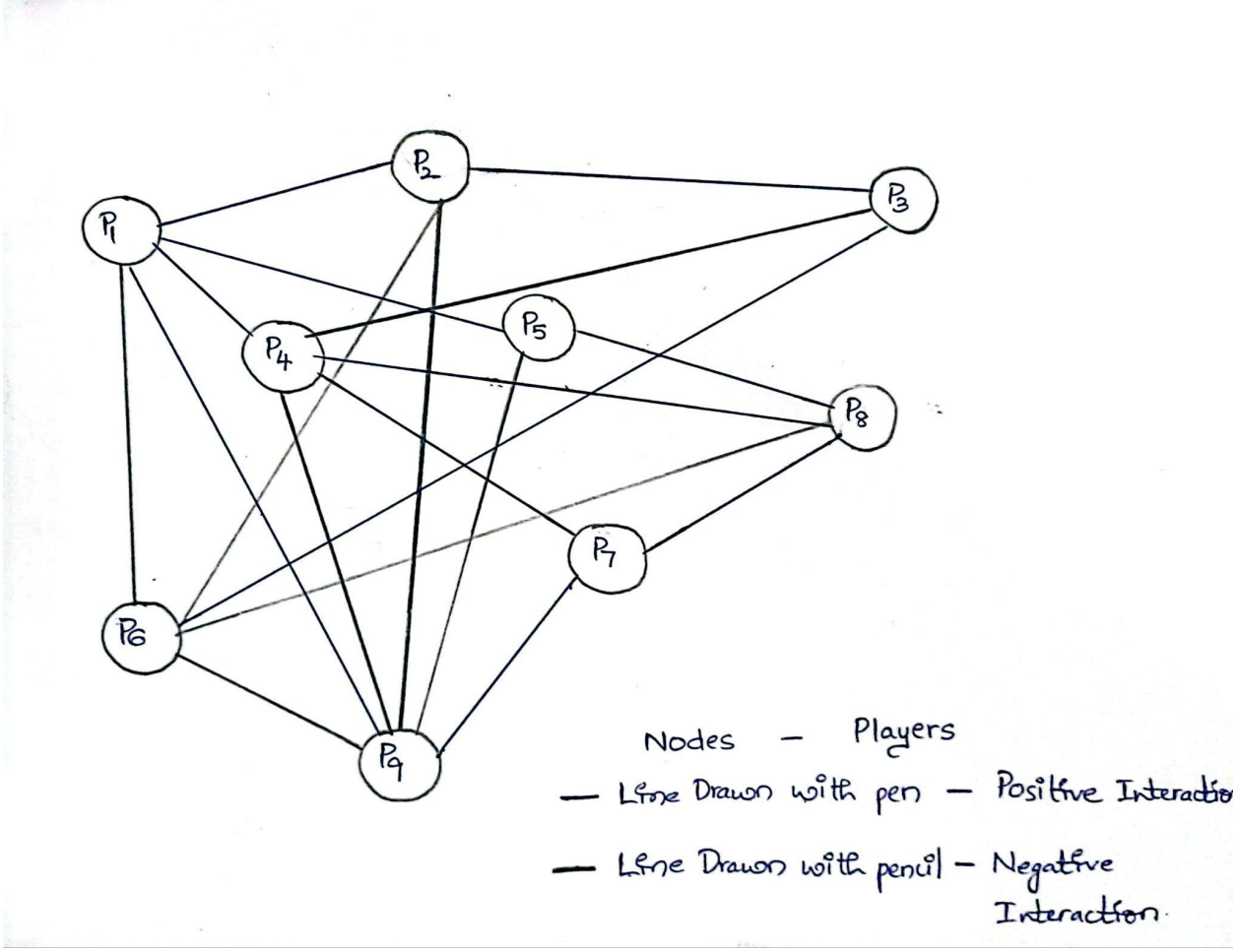
Sketch 2:


Comparison of number of goals between two countries yearwise for world cup
Country 1
Country 2
Number of goals
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1980
1982
1984
1986
1988
1990
1992
1994
1996
World Cup Year

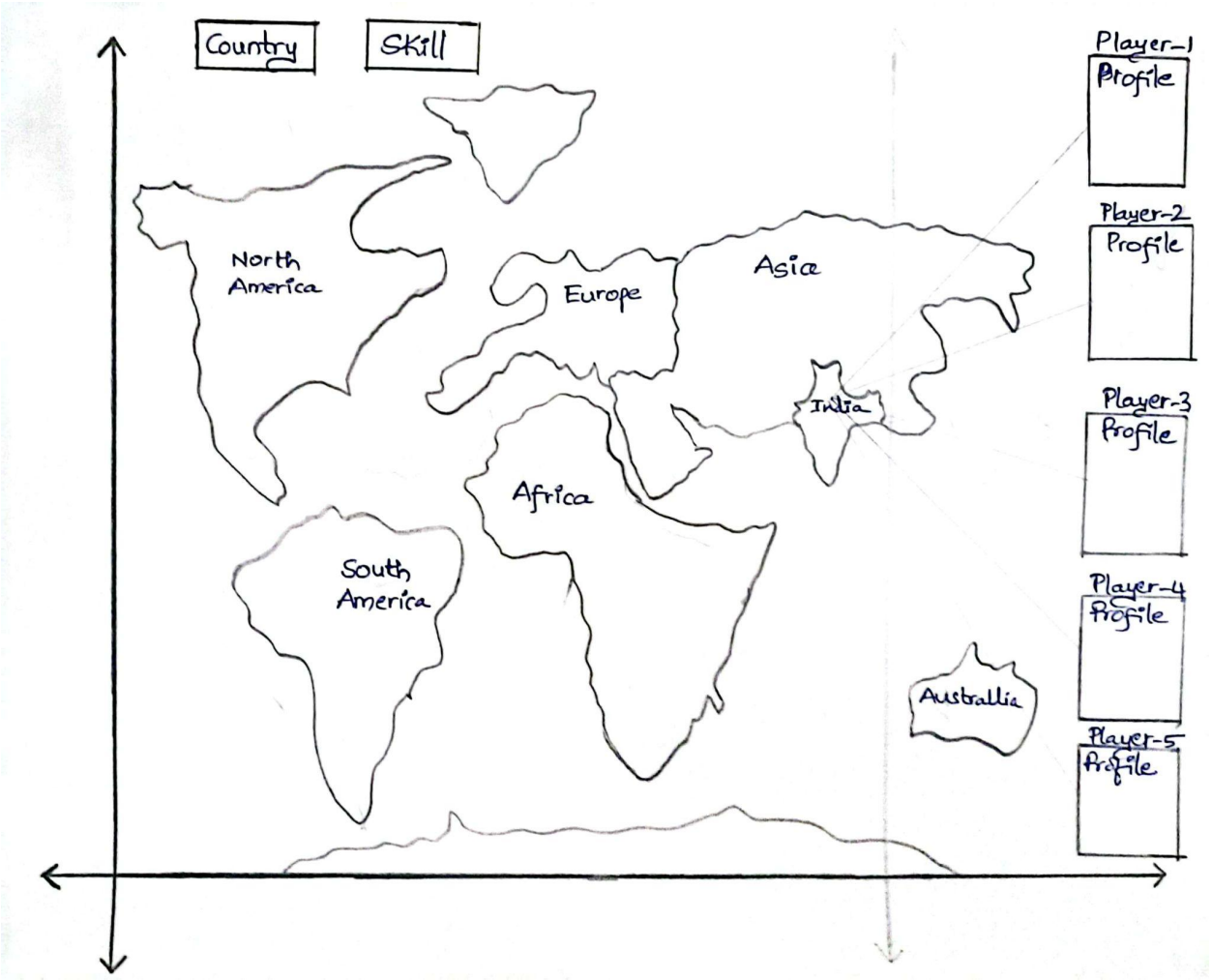
Sketch 3:


WORLD
Country 1
Country 2
Country 3
Country 4
Country 5
Country 6
Player 1
Player 4
Player 3
Player 2
Player 2
Player 2
Player 2
Biodata
Player 4
Biodata
Player 2

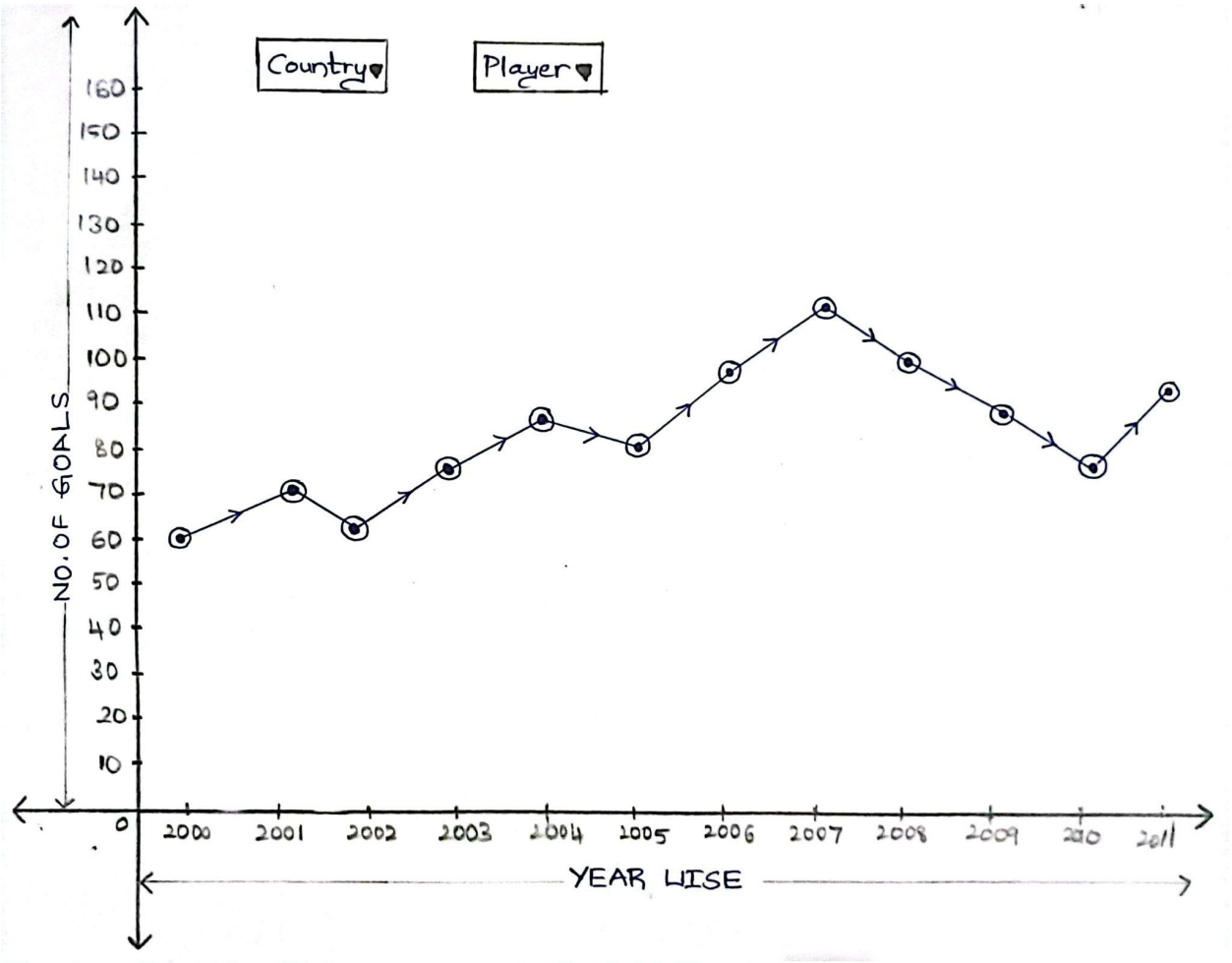
Sketch 4:


P1
P2
P3
P4
P5
P6
P7
P8
P9
Nodes – Players
— Line Drawn with pen – Positive Interactio
— Line Drawn with pencil – Negative Interaction.

Sketch 5:


Country
Skill
North America
Europe
Asia
India
Africa
South America
Australia
Player-1 Profile
Player-2 Profile
Player-3 Profile
Player-4 Profile
Player-5 Profile

Sketch 6:



Steps to first Project Milestone :

Scope of Prototype:

The prototype "FIFA VizPlay" seeks to provide an interactive visualization of the intricate details within the world of football, going beyond the traditional metrics. The scope includes providing insights into player performances, match outcomes, player's physic and skills metrics, and fostering data-driven discussions in the football community.

Data Collected:

Data is sourced from the FIFA player database and comprises two main datasets.
The first dataset focuses on match specifics like match_id, team_id, player_id, etc.
The second dataset delves into individual player statistics, skills, career trajectory, and market value.

Dataset source:
https://www.kaggle.com/code/sivsankar/fifa22-recommender-system/input

Data Still Needed:

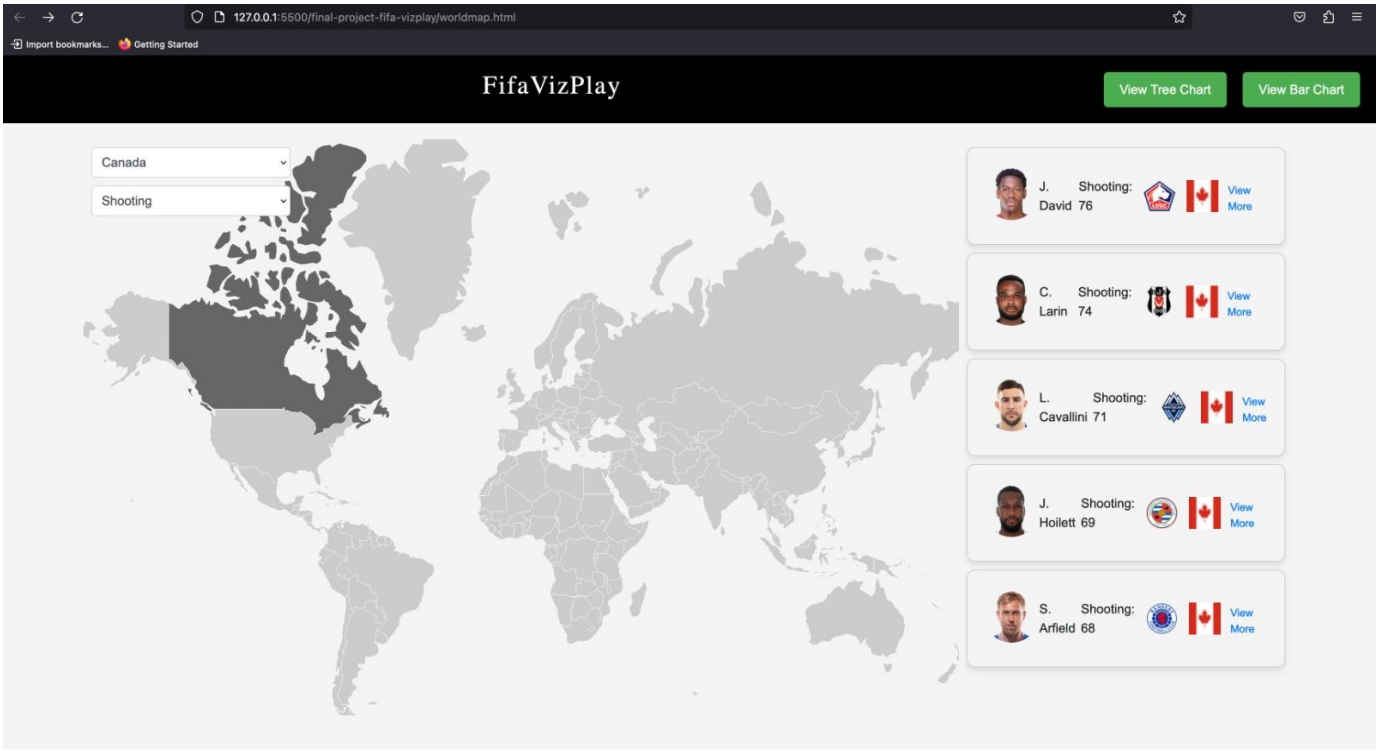
As of now, we have a comprehensive collection of player and match-specific data. However, future iterations may require data on team and player interactions etc., to offer a more holistic view.

Important Data Structures:

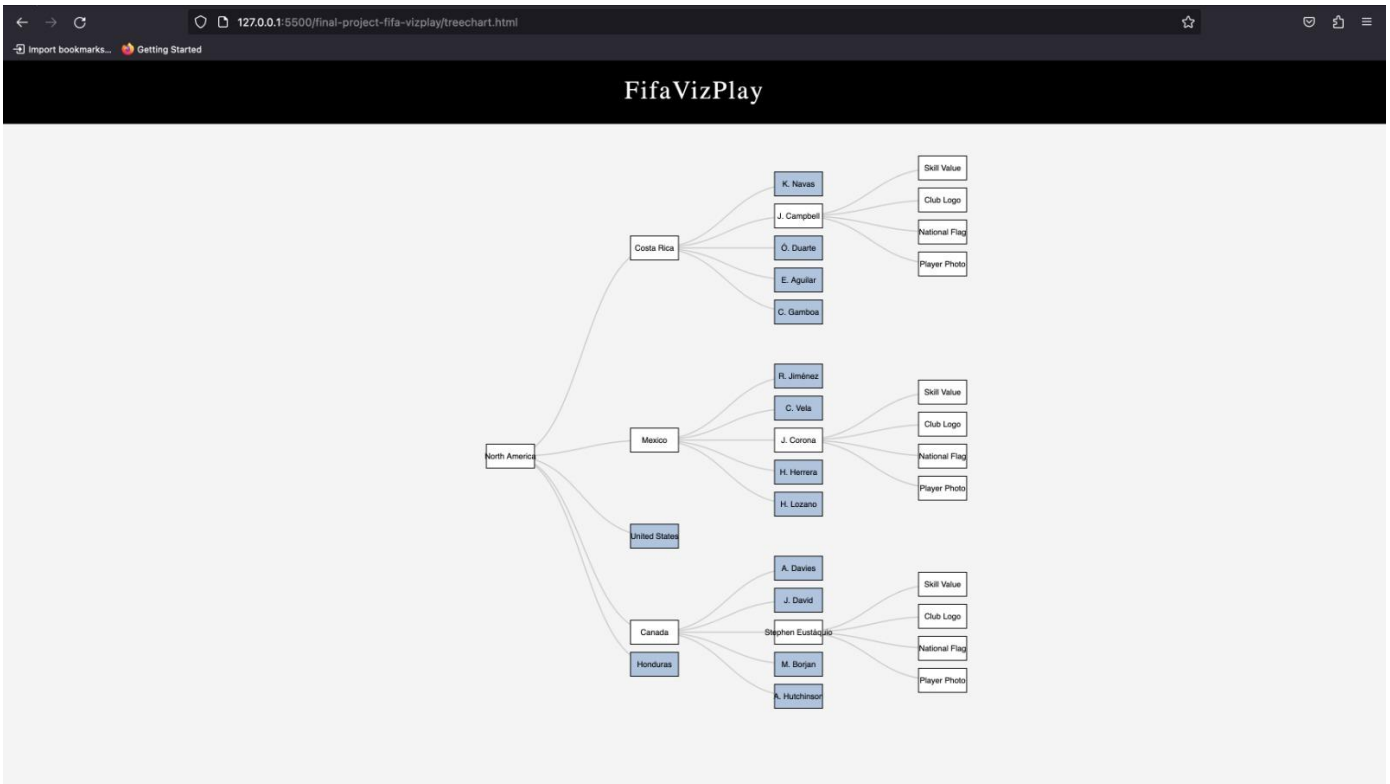
Match specific identifiers such as match_id, team_id, player_id.
Individual player statistics, encompassing player skills, career trajectory, market value, player attributes, and performance metrics.
Aggregated metrics like goal statistics on a yearly and country-wise basis.

Implementation Code Fragments:

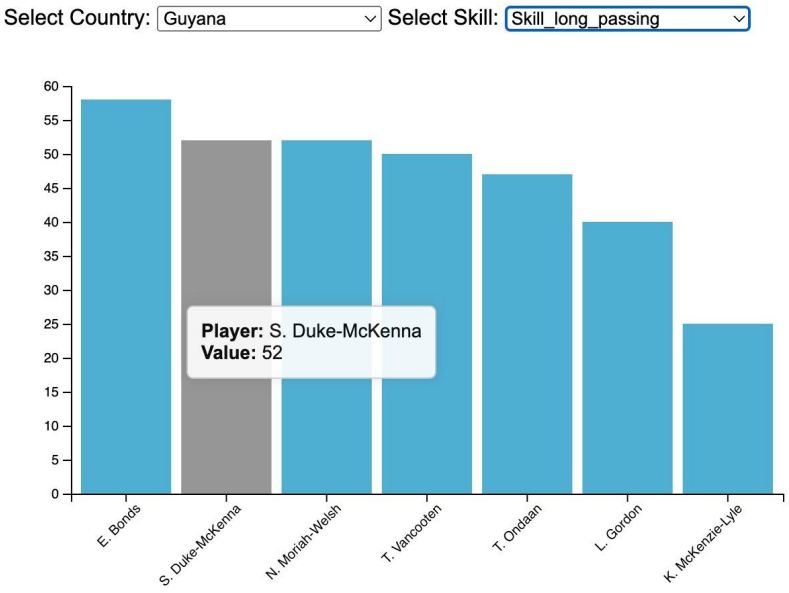
- **Data Cleanup:** Identify and handle missing data, merge datasets by linking relevant columns like player IDs and team IDs, player IDs and skills.
- **Metric Derivation**: Calculate player performance scores based on attributes, aggregate goal statistics.
- **Visualization:** Implement charts, graphs, and visualizations using D3.js and other visualization tools.
- **Geographic and Hierarchical Data -** World and Tree Map Prototype: Preliminary and refined designs illustrate the world map view of top football countries and a tree map breakdown of their premier players. This interactive interface offers a comprehensive insight into the global football landscape, emphasizing both country prominence and individual player excellence.



The Global Football Atlas: Click on any country to unveil its top 5 football stars according to the skill selected from the drop down as shown in the above image, also each with a detailed player card on the right. Dive deep into this dynamic map to witness the spread of world-class talent and get a closer look at each nation's football icons.

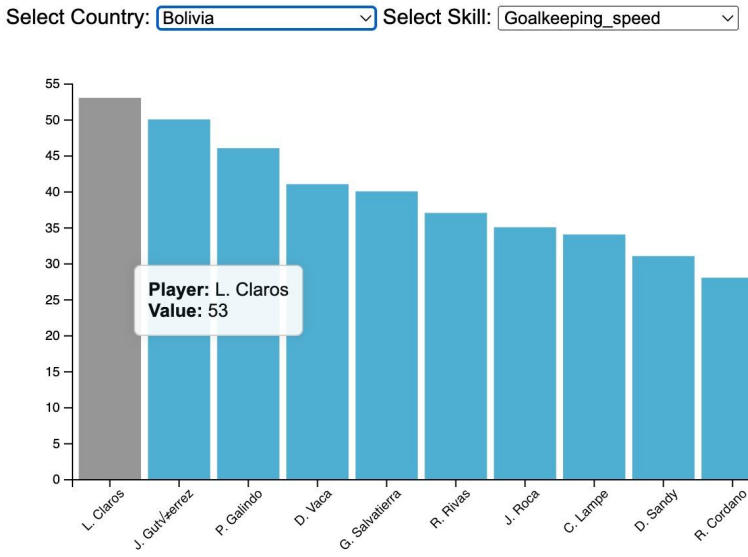


Football Elite Hierarchy: Explore the world's top 5 football nations and delve deeper into their standout players. This chart provides a layered view, first showcasing leading countries in football, and then spotlighting their top 5 players. A visual journey into the heart of global football excellence.



Guyana graph (long passing):

- "Comparison of long passing skills among Guyanese footballers. S. Duke-McKenna stands out with a value of 52."
- "Showcasing Guyana's talents: Long passing abilities of prominent players."



Bolivia graph (goalkeeping speed):

- "Bolivian goalkeepers ranked by their speed. L. Claros leads the pack with a score of 53."
- "A glance at Bolivia's finest: Evaluating goalkeeping speed of top players."

FifaVizPlay Overview and Use Cases:

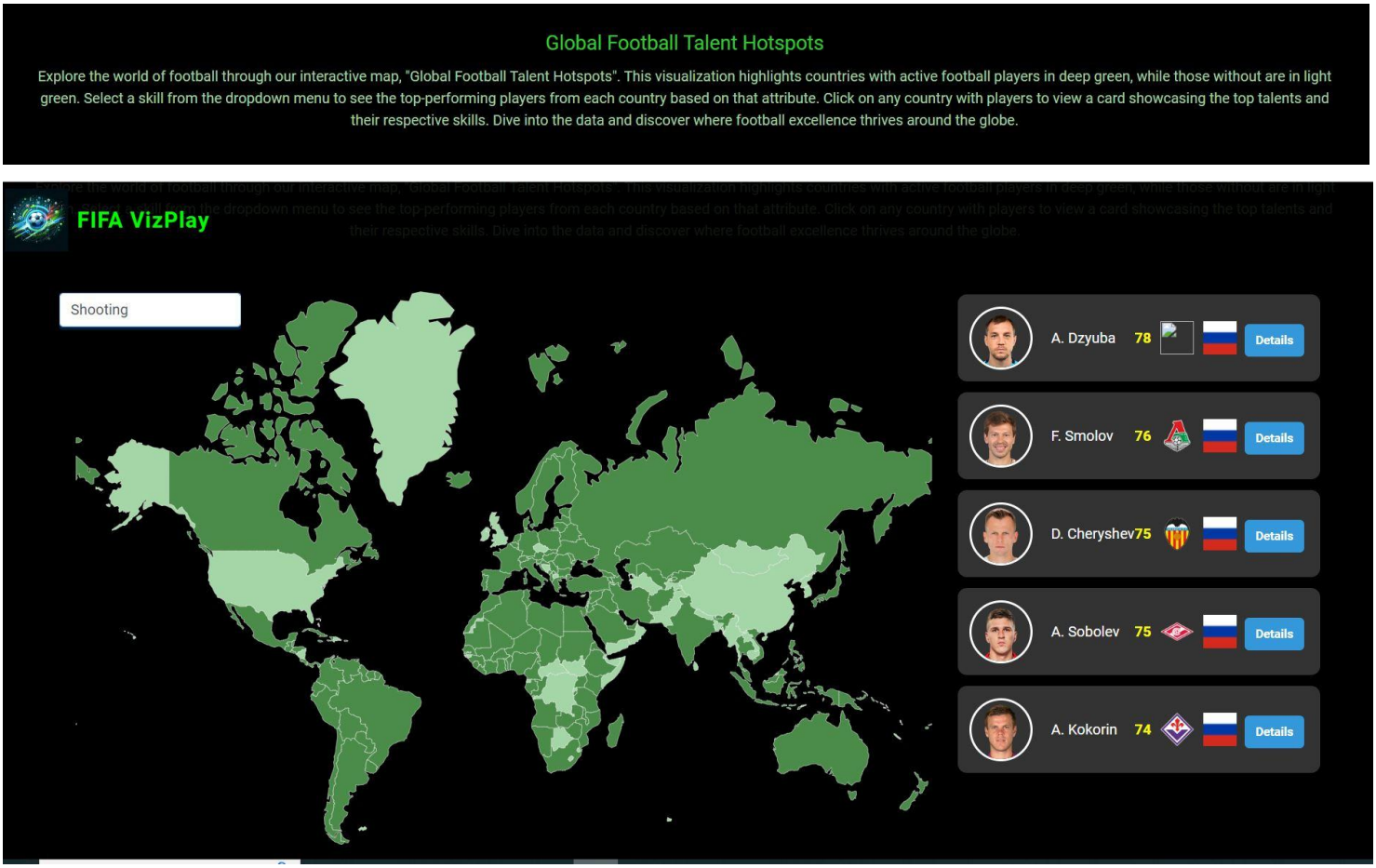
- **Overview:** "FIFA VizPlay" offers a dynamic platform for users to delve deep into football data, fostering informed, data-driven discourse.

Use Cases:

- **Football Enthusiasts:** Dive deep into player statistics, compare players based on attributes, and analyze performance metrics of national teams.
- **Researchers:** Understand trends and patterns, analyze correlations between player attributes and performance.
- **Football Analysts:** Get insights on club-level statistics, understand league-level data and dynamics, and create profiles for individual players or clubs summarizing their key attributes.
- The visual sketches can be included where they fit best, especially under the “Structural Data" and FifaVizPlay Overview and Use Cases sections to give a visual representation of the interface and data structures.

Final Submission :

Graph 1 : TOP 5 Players Based on their Skill in each Country :



The above visualization is an interactive map titled "Global Football Talent Hotspots," ingeniously crafted to provide an insightful overview of the skill distribution among football players globally. It employs a color-coded scheme to indicate the activity level of football talent within each country: countries teeming with football players are shaded a rich, deep green, while those with fewer or no players are represented in a lighter green. This demarcation is further elucidated by tooltips, which reveal the exact number of players—displaying a zero where players are absent. The interface includes a strategically placed dropdown menu, offering users the opportunity to refine their exploration by specific football skills. This feature enriches the user experience by tailoring the displayed information to the user's interests in skill-based performance metrics.

Upon selection of a skill from this menu, the map dynamically adjusts to showcase the relevant data. For instance, clicking on Argentina—a country highlighted for its wealth of players—prompts a detailed overlay to emerge. This display is not just a simple list; it is a meticulously arranged showcase of the country's finest talents, organized by their expertise

in a specific skill, like “Shooting”. Each player's profile is meticulously detailed, featuring their name, skill rating on a centesimal scale, the emblem of their associated club, and a direct link to further details. This interactive map is not just a visual treat; it's an analytical tool that invites the user to delve deep into the heart of football's skill landscape, offering a comparative analysis of top-tier talent across the globe. The map is more than a display; it is an engaging narrative that unfolds the story of football excellence, providing invaluable insights for your process book that encapsulates the journey and findings of this exploration.

Graph 2 : Comparative Visualization of World Cup Goal Statistics and FIFA Rankings :

The above visualization consists of two moving line graphs that provide a comparative analysis of football statistics between two selected countries—England and Germany, as indicated by the dropdown menus.

The first graph, labeled "Goal Stats," is a historical tracker of the countries' goal-scoring performances in World Cup events. The x-axis represents time, marked by years when World Cup tournaments were held, while the y-axis quantifies the goals scored. The plotted lines for each country oscillate to depict the fluctuations in their scoring achievements over time. The use of two different colors for the lines allows for clear differentiation between the countries' performances, making it easy to observe trends, peaks, and troughs in goal-scoring patterns.

The second graph, "Ranking Stats," presents a timeline of FIFA rankings for the two countries, where the y-axis inversely represents the ranking position—the lower the value, the higher the ranking. This graph serves as an indicator of the countries' changing competitive status over

the years. Sharp ascents and descents in the lines can signify significant events or shifts in the football prowess of the nations. Thus, both graphs are interactive and responsive, updating in real-time as different countries are selected from the dropdown menus. This feature not only enhances user engagement but also provides an educational comparison of historical data, allowing for a deeper understanding of each country's football history and global standing.

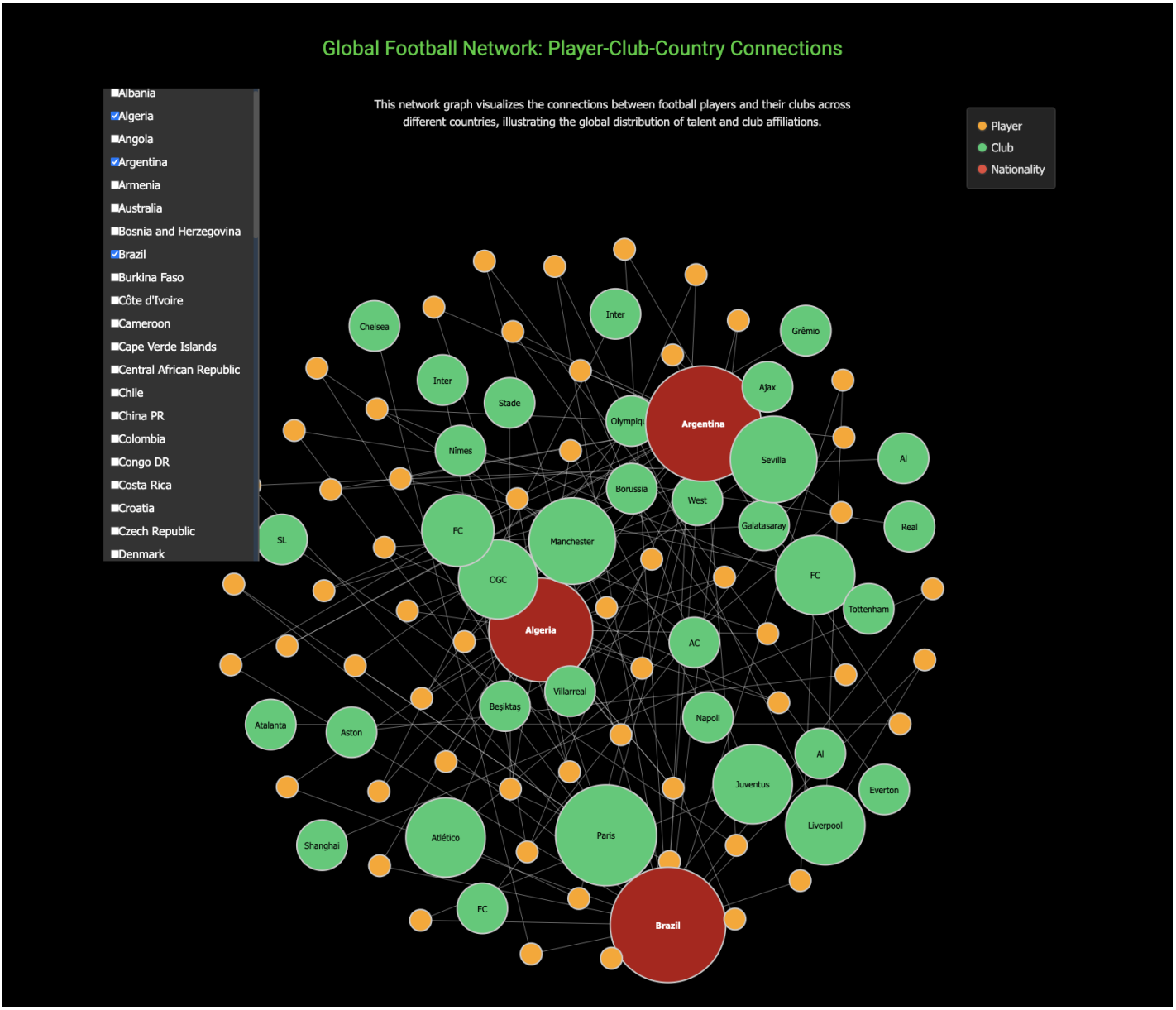
Graph 3 : Interactive Analysis of Football Proficiency and World Cup Dynamics

The first visualization clearly shows "Player Skill Stats," offers a bar chart that ranks players from a selected country based on a specific skill—such as 'Attacking_Crossing' for Kenyan players. This visualization is interactive, allowing users to choose the country and skill from dropdown menus, which then generates a customized bar chart. Each bar represents an individual player and their proficiency in the selected skill, quantified as a 'Skill Value'. This visual format is particularly effective for identifying standout players and making direct comparisons between players' abilities in a chosen skill area.

The second visualization to the right we have "World Cup Stats," tracks a variety of statistics such as goals total, total attendance, teams count, total attendance, average goals, teams qualified related to the World Cup over an adjustable time range, set by the user. This line graph

can display different data points, such as total goals scored across all World Cup tournaments between the years 1930 and 2020. The graph illustrates the ebb and flow of the tournament's statistics over time with a line connected by dots, each representing a World Cup year. This tool allows for historical analysis of the World Cup, providing insights into patterns and trends, such as periods of high scoring, which could correlate with changes in playing styles, rules, or other factors.

Graph 4 : Global Football Network : Player-Club-Country Connections :



The above image shows network visualization titled "Global Football Network: Player-Club-Country Connections." It mainly illustrates the complex web of affiliations between football players, their clubs, and their nationalities, highlighting the vast global distribution of

talent within the sport. In this network graph, individual players, clubs, and countries are represented by nodes, likely color-coded to distinguish between the three categories—players are represented with orange color, clubs with green, and nationalities with red color. The size of a node determines the number of players it has.

The connections between these nodes, possibly represented by lines, demonstrate the relationships and transfers that link players to clubs and the national teams they represent. Such a network graph provides a compelling visual representation of the international nature of football, showing how players from a wide array of national backgrounds come together at clubs around the world, creating a diverse and interconnected community. This type of visualization is particularly valuable for fans, analysts, and professionals within the football industry to explore patterns of player movement, the reach of different clubs in terms of international talent, and the prominence of certain nationalities within the global football landscape.

Graph 5 : Interactive Player Interaction Network :

Interactive Player Interaction Network

Explore the Dynamics: This interactive graph maps the on-field interactions between players, highlighting patterns of play. Red nodes mark frequent negative interactions, potentially indicating strategic weaknesses or Match fixing possibility. Use the filters to view by country or club and uncover insights into team and player connections.

Select Countries

- Albania
- Algeria
- Angola
- Argentina

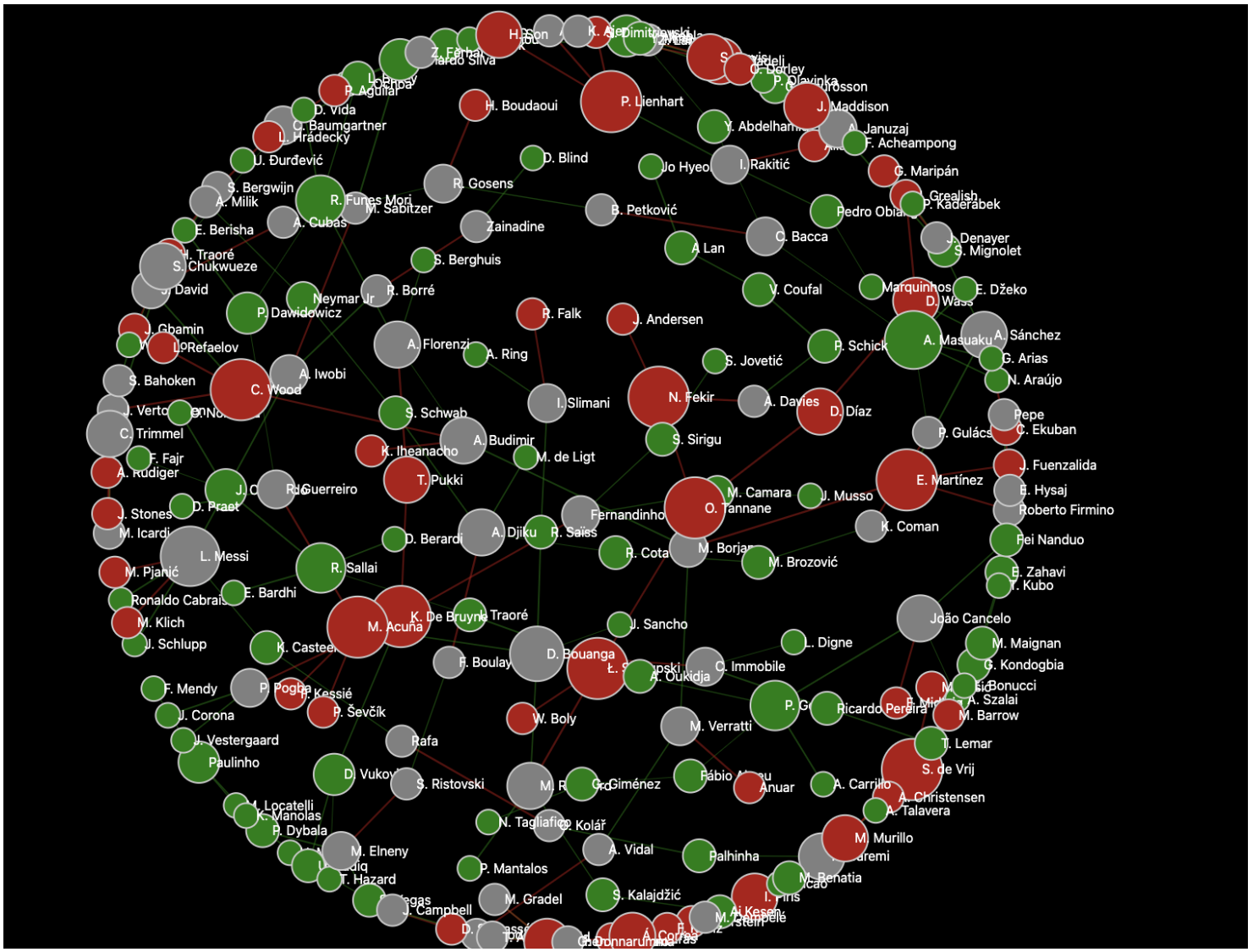
Select Clubs

- 1. FC Köln
- 1. FC Union Berlin
- AC Milan
- AEK Athens

Negative Interaction (-ve): Described as an event where a player's interaction with another player results in a goal for the opposing team. These could be mistakes like lost balls or poor passes leading to counter-attacks.

Negative Impression: This is a metric indicating players who have repeatedly had negative interactions with the same player, possibly suggesting match fixing between those two.

- Red Node: Player with potential **match-fixing** risk
- Green Node: Player with no negative interactions
- Grey Node: Player exhibiting normal behavior
- Red Edge: Frequent negative interactions between those two players
- Green Edge: Very few negative interactions between those two players



The above visualization is all about the "Interactive Player Interaction Network" which mainly maps the intricate web of on-field interactions between football players. This network graph is a sophisticated tool designed to analyze and highlight patterns of play, connections between players and clubs, and the global distribution of talent. Red nodes in the network represent players who frequently engage in negative interactions during a match, such as losing possession or making poor passes that lead to counter-attacks by the opposing team. These nodes could potentially indicate strategic weaknesses or even raise concerns about match-fixing if patterns of negative interactions are consistently associated with the same individuals which is termed as a negative impression.

Also its ability to filter the network graph by country or club, allowing users to focus on specific areas of interest. Negative Impressions, denoted by red edges between nodes, suggest a metric of repeated unfavorable interactions between players, possibly indicating deeper issues within team dynamics or player performance. Conversely, the absence of such impressions, shown by green edges, indicates a neutral or positive interaction history. Nodes without any negative connotations are colored green, while grey nodes represent players exhibiting normal behavior without significant negative events. This visualization not only serves as a means to explore the interconnected nature of the football world but also as a potential investigative instrument to examine the integrity of the sport. It combines the analytical depth with interactive capabilities, making it a powerful feature for stakeholders at various levels of the football ecosystem, from fans and commentators to analysts and governing bodies, to explore and draw conclusions from the data presented.

Visual Encodings:

Color Encoding Used:

- **Global Talent Map:** Different shades of green to indicate the concentration of football talent in countries.
- **Comparative Line Graphs:** Different colors for lines to distinguish between the performances of different countries in World Cup events and FIFA rankings.
- **Player-Club-Country Network Graph:** Nodes color-coded to distinguish players, clubs, and nationalities; red nodes for negative player interactions in the Player Interaction Network.

Position Encoding Used:

- **Goal Stats and Ranking Stats:** The x-axis represents time (years), and the y-axis represents either goals scored or FIFA rankings.
- **Player Skill Bar Chart:** Bars positioned to represent different players' skill levels in a specific area.

Size Encoding Used:

- **Global Football Network Graph:** The size of nodes represents the number of players affiliated with a club or country.
- **Interactive Player Network Graph:** The size of nodes represents the number of negative impressions he had overall.

Evaluation:

Our visualization initiative was designed to decode the complexities of the football world by weaving a detailed network that connects players, clubs, and countries across the globe. It serves as a digital atlas, charting the extensive array of talent and the competitive synergy that defines the sport. More than a mere depiction of connections, our platform is an interactive odyssey that enables users to trace the career pathways of players, the strategic alliances of clubs, and the national affiliations that shape the game's essence. This network does more than just map data; it brings to light the unpredictable nature of football, where unlike the straightforward contests of some sports, the beautiful game presents a kaleidoscope of competition. Our implementation focuses on delivering a user-driven exploration, where every click and selection casts new insights, offering a fresh perspective on the global influence of football.

To craft a user-centric experience, we invested in creating a web of responsive interactions offering users a multitude of ways to engage with the data. Selecting a nation or a club triggers tailored visual responses that illuminate specific player affiliations, encouraging a dynamic and investigative user journey through the global football landscape. Throughout this process, we navigated through a myriad of technical challenges, from data integration to achieving a harmonious balance between rich functionality and fluid user interaction.The graphs provide a clear representation of data related to suspected match-fixing incidents. This clarity is crucial for understanding patterns and anomalies in matches that may suggest manipulation. This demonstrates a deep understanding of the data and the subject matter, going beyond traditional football analytics. While these graphs are informative, they also raise ethical considerations regarding the interpretation of data. It's crucial that such sensitive information is handled responsibly to avoid false accusations or misinterpretation

While the project is comprehensive, challenges in data integration and maintaining a balance between functionality and user interaction were significant. Future iterations could focus on expanding the dataset to include more dynamic aspects of football, like team and player interactions.

Conclusion:

The visualization project successfully encapsulates the global expanse and intricate connections within the world of football, presenting an interactive network that elegantly

maps the relationships between players, clubs, and nationalities. It stands as a testament to the sport's unifying spirit, bridging continents and cultures. Through innovative design and interactive features, the tool not only enlightens enthusiasts and professionals about the fluid dynamics of football affiliations but also celebrates the diversity and interconnectedness that make the sport truly global. The project, overcoming technical hurdles to deliver a seamless user experience, redefines how fans and analysts alike engage with football data, offering new perspectives on the beautiful game's complex ecosystem.

Related Links :

- **Project repository :** https://github.com/dataviscourse2023/final-project-fifa-vizplay
- **Dataset** - https://www.kaggle.com/code/sivsankar/fifa22-recommender-system/input
- **Project Website -** https://adityagattu.github.io/
- **Video -** https://youtu.be/ZxxKwHBvwyA