Project Report For CS661: BIG DATA VISUAL ANALYTICS

2023-2024 Semester II

Project Title: Visual Analytics System based on FIFA Players Visualization

Group 3: Team Members

Madhav	Shreenaga Tejas	Thota Vishnudatta
(200657@iitk.ac.in)	*(200296@iitk.ac.in)*	*(201051@iitk.ac.in)*
Poojitha	Prashamsha	Ankit Kumar
(201094@iitk.ac.in)	*(200846@iitk.ac.in)*	*(210150@iitk.ac.in)*
Sambhav Agarwal	Vibhanshu Patidar	Aryan Jain
(200852@iitk.ac.in)	*(201106@iitk.ac.in)*	*(200201@iitk.ac.in)*

IIT Kanpur

1. Introduction

In today's football world, leveraging data is crucial for clubs, managers, and scouts to make smart decisions. By analyzing player data, we can uncover insights about where players come from, their leagues, and their attributes. This information helps clubs build better teams, find promising new talents, and stay ahead of the competition.

Our project aims to develop an interactive dashboard for analyzing global football player data. The primary objectives include examining the distribution of players worldwide and within specific leagues and exploring player distributions based on attributes such as position, age, wage, and rating.

As the Proposed Solution Section outlines, we utilize various interactive plots to achieve these goals. Our overarching objective is to enhance user engagement by enabling them to quickly identify players with desired characteristics, such as pace, rating, dribbling, and position (e.g., forward, midfielder, goalkeeper). Additionally, this dashboard can serve as a valuable tool for talent scouting purposes.

Data Description

In this project, we have used the FIFA 22 Complete Player Dataset, a comprehensive collection of data about football players featured in the FIFA 22 video game. The datasets provided include female and male player data for Career Mode from FIFA 15 to FIFA 22. Each player in the dataset possesses 110 attributes, including features such as overall rating, potential, age, nationality, club affiliation, league participation, preferred foot, pace, shooting, and passing abilities.

We scraped the data for this project from the publicly available website sofifa.com. You can find the dataset at the following link:

Dataset link. The number of players' data collected for each year is shown below:

FIFA version	Female Players	Male Players	Columns (Attributes)
FIFA 15	-	16155	110
FIFA 16	248	15623	110
FIFA 17	299	17596	110
FIFA 18	317	17954	110
FIFA 19	299	18085	110
FIFA 20	345	18483	110
FIFA 21	345	18944	110
FIFA 22	391	19238	110

Table 1: Dataset Details

2. Tasks

In this section, we elaborate on the tasks in detail:

- **Dataset preprocessing:**

  In the data processing phase, the dataset underwent several critical operations to prepare it for visualization and analysis. Initially, data were merged using player_id as the primary key, replicating relevant columns across multiple years within the master dataset. Subsequently, data cleanup procedures were implemented to remove rows containing missing basic player details. Data imputation techniques were also applied to ensure consistency in attribute formats, particularly for nationality. We also use geo.json to construct a world map to visualize the nationality distribution of the players.

- Our strategy entails creating an integrated visualization dashboard to depict our dataset accurately. To achieve this goal, we utilize a range of visualization techniques, each customized to emphasize various aspects of the data:

  - **Chloropleth map:** A choropleth map is used to represent data using the color mapping technique, hence providing a way to visualize values over a geographical area. Here, we plot the frequencies of the players worldwide over the map to visualize it.

  - **Bar graph:** A Bar chart is used to depict the distribution of players across different football leagues. This visualization method is selected for its ability to present player distribution patterns clearly and allow for easy comparison between leagues.

  - **Word Cloud:** A word cloud is a visual representation of a text in which the words appear more significant the more often they are mentioned, i.e., higher frequency. Word clouds are great for visualizing unstructured text data and getting insights into trends and patterns. We used word cloud to represent the player's names per their rating. They provide granularity for filtering specific players based on their names.

  - **Parallel coordinates plot:** Parallel coordinates plot displays multivariate data on a set of parallel axes. Each axis represents a different variable, and

each data point is represented by a line that intersects all these axes, allowing for simultaneously visualizing relationships and patterns across multiple dimensions.

– **Sunburst graph:** A Sunburst chart is used to visualize a hierarchical dataset. It shows hierarchy through a series of concentric rings, each corresponding to a level in the hierarchy.

- We enhanced the interactive functionalities of the dashboard by integrating features such as brushing for filtering, zooming, and dragging to improve visual clarity. Additionally, we ensured color consistency across all the plots for a cohesive user experience.

3. Proposed Solution

Year Selection

We propose implementing a button feature to facilitate loading player data from different years, from 2015 to 2022. The buttons positioned at the top right of the interface enable users to select the desired year, thereby loading the corresponding dataset. Additionally, these buttons will reset the plots to display all the data associated with the selected year.


Figure 1: Buttons to select years

Choropleth Map

- **Zooming and Panning:** Zooming and panning functionality are provided for easier selection and viewing, accommodating the varied nature of countries on the map.

- **Color Legend:** A color legend is included based on player frequency, providing visual context for the data represented on the map.

- **Tooltips:** Tooltips display the interaction between player frequency and the number of players for each country, enhancing user understanding and engagement.

- **Interaction with Other Plots:** The map interacts with all other plots based on the selected region.
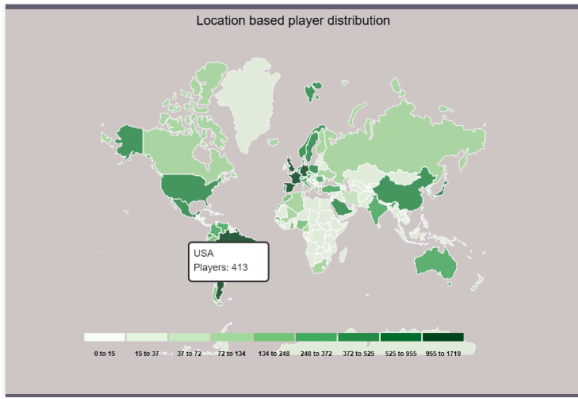
Tooltip Feature
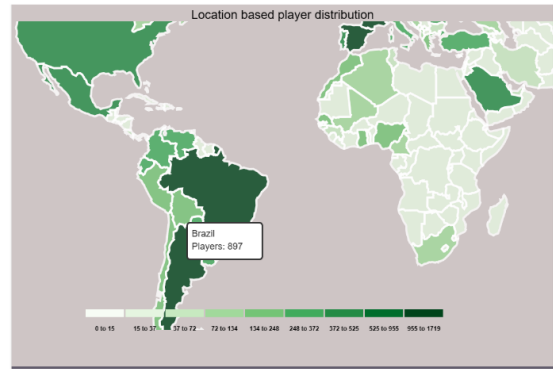
Zooming and Panning Feature

Figure 2: Choropleth map

Bar Graph

- **Opacity Control and Bar Height:** Opacity is adjusted for clarity and bar height to indicate player strength post-selection.

- **Coloring of Bar Charts:** The bars' color is selected from the categories in sunburst plots for visual consistency.

- **ToolTip Functionality:** Hover over league bars to see the number of associated players in a particular league.

- **Interactive Features:** Leagues can be selected or deselected, updating all related plots for dynamic comparisons.
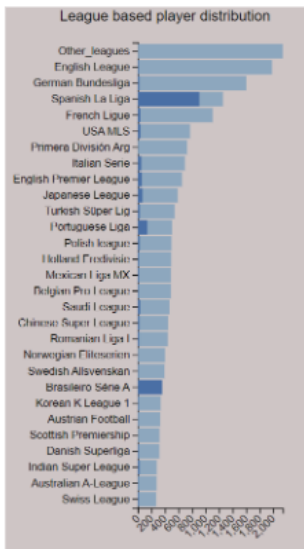
Bar Graph after selecting Spain and Brazil in 2022
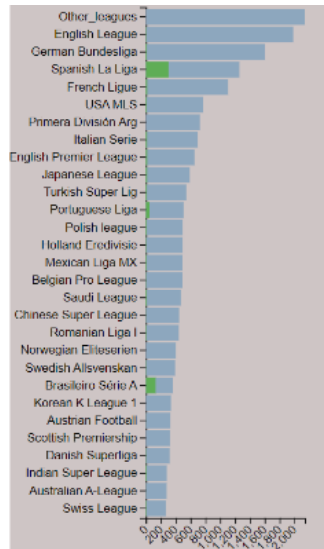
Bar Graph after selecting Defenders for Spain and Brazil
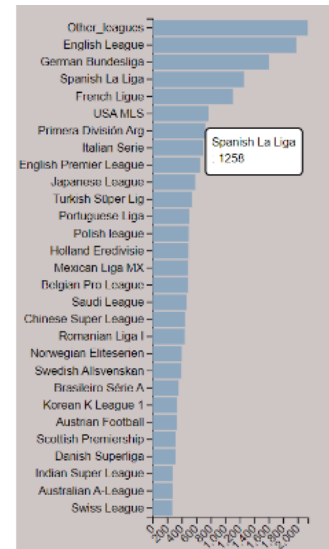
ToolTip Functionality: Hover over league bars to see the numbers

Figure 3: Bar Graph

4

Word Cloud

- **Color Consistency:** Player names are color-coded according to their positions for visual coherence.

- **Size of the Names:** Larger-sized names represent higher-rated players, providing a quick visual indicator of player quality.

- **Interactive Feature:** When clicking upon any player's name, the interactive feature allows users to view the nationality and attributes measured by the Player Characteristic Profile (PCP) for a specific player.

Figure 4: Word Cloud

Sunburst Graph

Figure 5: Sunburst plot

Abbreviation	Full Form
CB	Center Back
RB	Right Back
LB	Left Back
RWB	Right Winger Back
LWB	Left Winger Back
CM	Central Mid Fielder
CDM	Defensive Mid Fielder
CAM	Attacking Mid Fielder
RM	Right Mid Fielder
LM	Left Mid Fielder
ST	Striker
CF	Center Forward
RW	Right Winger
LW	Left Winger

Table 2: Abbreviations and Full Forms

- **Visualization Features:**

  1. Each segment in the sunburst represents a player position, such as Center Back (CB) or Central Midfielder (CM). The size of each segment corresponds to the number of players in that position.

2. Colors differentiate between players' positions and roles (defender, attacker, midfielder, goalkeeper).

- **ToolTip Functionality:** Hover over rings to see the number of associated players, enhancing interactivity.

Color code used for color consistency in the dashboard:

Position of the Player	Color
Attacker	Orange
Midfielder	Sky Blue
Defender	Green
Goalkeeper	Yellow

Table 3: Player Positions and Colors
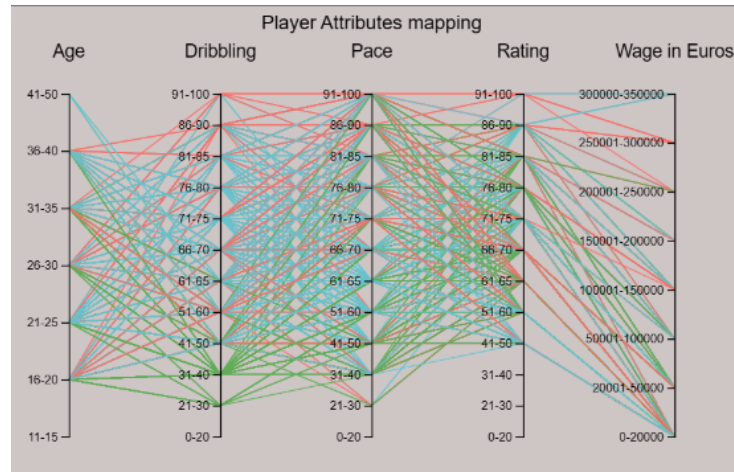
Parallel Coordinates Plot

Figure 6: Parallel Coordinates Plot

- **Visualization Features:**

  1. In the parallel coordinates plot, we measure attributes such as Age, Rating, Wage in Euros, Pace, and Dribbling. This approach provides a comprehensive overview of each player's profile.

  2. The lines in the plot are color-coded based on the player's position on the field (e.g., goalkeeper, defender, midfielder, forward).

- **Brushing Feature:** Brushing feature enables users to filter players based on specific data ranges for each attribute.

- **Dragging Feature:** Brushing feature enhances plot visibility by allowing users to interactively drag the plot horizontally at a particular axis to focus on a specific variable of interest.
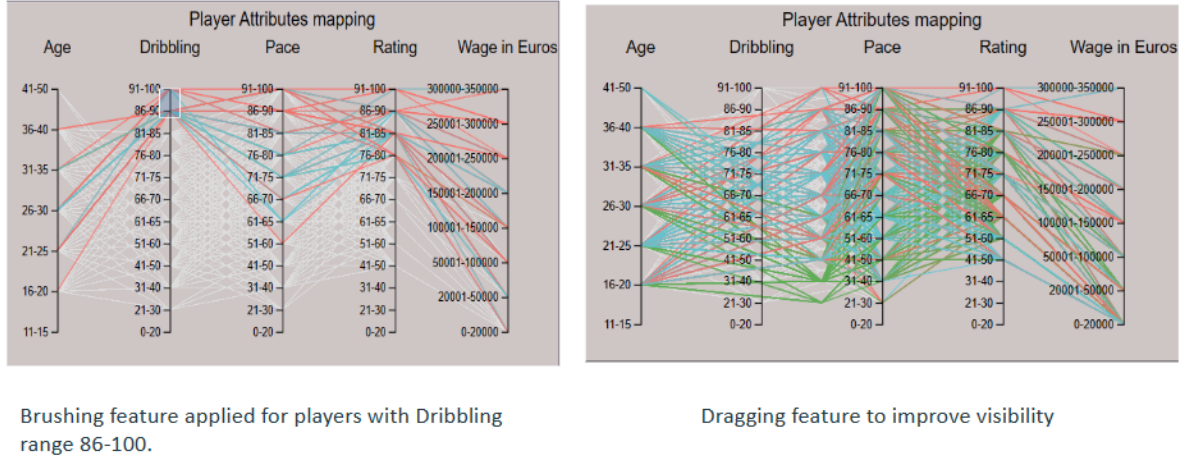
Brushing feature applied for players with Dribbling range 86-100.

Dragging feature to improve visibility

Figure 7: Parallel Coordinates Plot features

Use Cases As a Player Scouting Tool

Our visualization system is a valuable tool for football scouting, offering scouts and analysts insights into player performance, skills, and potential through interactive features and displayed data. Here are some key applications:

- **Identifying Players:** Users can explore players based on age range and skills across various leagues and countries.

- **Analyzing Player Metrics:** Speed, dribbling, shooting, and overall rating can be analyzed using the parallel coordinate plot and sunburst chart to understand a player's strengths and weaknesses.

- **Position-Specific Search:** Users can filter the sunburst chart to find players in specific positions like forwards, midfielders, or defenders.

For instance, users can utilize the brushing feature on the parallel coordinate plot to find defenders with specific attributes, such as a pace between 85-100 and a rating between 86-90. The dashboard filters out relevant players by selecting the "defender" category on the sunburst chart. As a result, users may obtain a narrowed-down list of players, such as João Cancelo and Jordi Alba from Portugal and Spain, respectively, with highlighted age and wages displayed in the dashboard. This information can be instrumental in scouting potential players. The results can be observed in Figure 8.

Figure 8: Scouting for players who are defenders with pace in range 85-100 and rating in range 86-90.

Architecture and Functionality Overview of the code

App.py:

- **Libraries and deployment:** For Implementing the Web Interface, we use Javascript and D3. For the backend, we are using the Python Flask server.

- **Data Fetching and Filtering:** Reads FIFA player data from CSV files based on the selected year and filters the data based on various parameters like league, nationality, position, player name, etc.

- **Routes:**

  - `"/"`: Renders the main HTML template.
  - `"/geo.json"`: Serves a static JSON file.
  - `"/fetchdata"`: Handles POST requests to fetch and process FIFA player data.

- **JSON Response:** Returns a JSON response containing: Player position distribution in a sunburst chart format, Nationality data for visualization on a map, Original unfiltered player data, Filtered player data for a bar chart, Word cloud data representing player names and their attributes, Additional player data for a parallel coordinate plot.
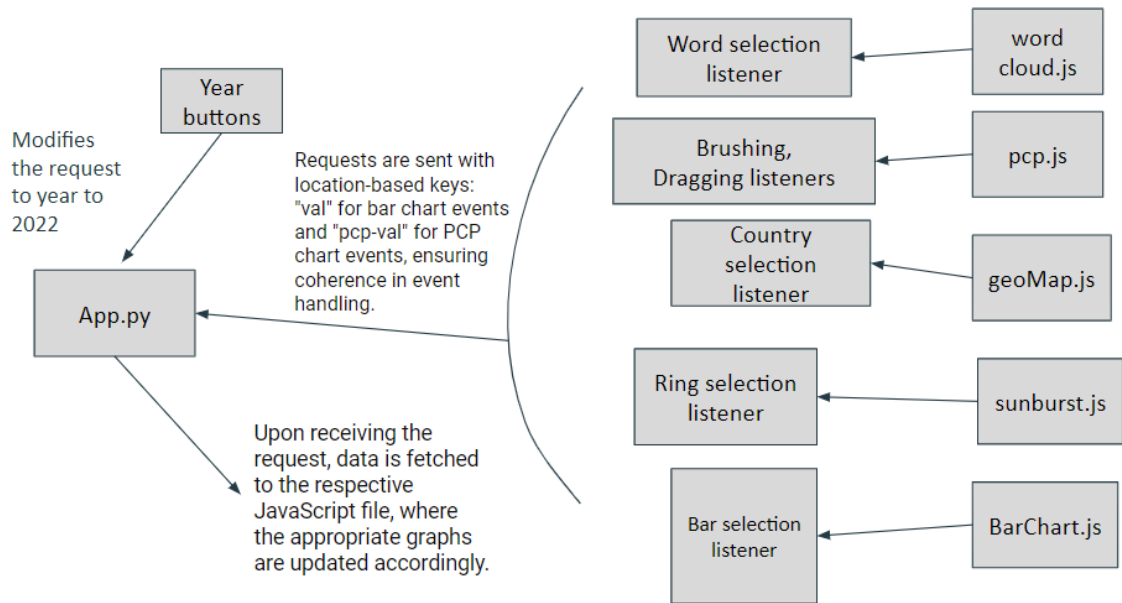
Figure 9: Flowchart overview

Styling

| Bar Chart (.css)	Geomap (.css)
- Controls appearance of the bar chart. - Manages font, size, weight, color. - Defines display properties for text.	- Layout and styling of geo map. - Dimensions & positioning. - Styles for swatches, text, appearance.
Sunburst (.css)	**Landing Page (.css)**
- Appearance of sunburst chart. - Defines styles for slices, arcs, and text. - Ensures readability and clarity.	- Dictates styling of the landing page. - Defines font, size, and alignment. - Sets background color, tooltip appearance.
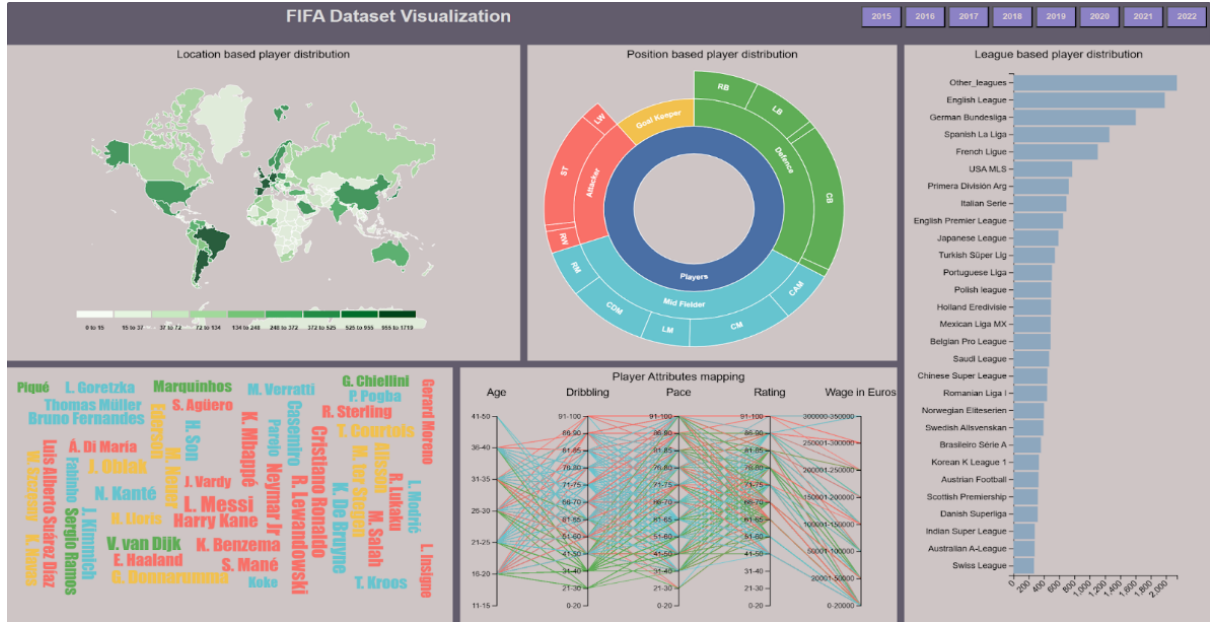
Figure 10: Dashboard landing page

4. Results

Insights drawn from the Data

Comprehensive visualization of the FIFA player dataset enables us to derive several insights, such as:

1. **League Dominance and Wage Distribution**: Leagues with more players tend to dominate in wage distribution compared to leagues with fewer players, indicating a correlation between player quantity and wage earnings.

2. **Regional Strengths**: Specific geographical areas, such as Argentina in Primera Division and Brazil in Brasileiro Série A, show significant player participation, highlighting regional dominance in certain leagues.

3. **Position-based Regional Strengths**: Regions like Brazil and Spain have a notable presence of highly-rated attackers and midfielders, suggesting a correlation between player positions and regional strengths.

4. **Age and Wage Correlation**: Analysis of age ranges, such as 16-20, reveals that players in this bracket typically earn lower wages, indicating a potential correlation between age and wage earnings, with younger players earning less on average.
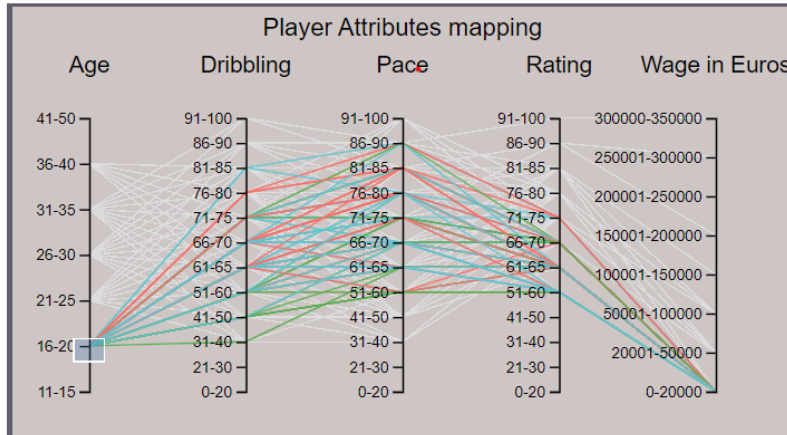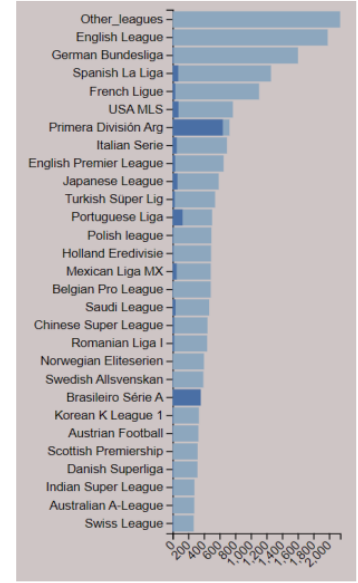
Figure 11: Figure shows the insights drawn from the data.

5. Conclusion

Our project on "Visual Analytics System based on FIFA Players Visualization" offers a robust platform for analyzing global football player data through interactive visualization techniques. By pre-processing the dataset and implementing various visualization methods such as choropleth maps, bar graphs, word clouds, sunburst graphs, and parallel coordinates plots, we provide users with a comprehensive tool for exploring player distribution, attributes, and performance across different leagues and regions.

By harnessing the power of data visualization, we aim to address critical challenges clubs, managers, and scouts face in making strategic decisions regarding player recruitment, team composition, and talent development. Our system enables users to navigate through vast amounts of player data effortlessly, uncovering insights about player distribution, performance metrics, and regional strengths.

Overall, our project underscores the significance of data visualization in understanding complex datasets and extracting meaningful insights.

6. Link to source code:

Code files are updated in the GitHub repository: Link

11

7. Work distribution

Name(Roll number)	Contribution
Shreenaga Tejas Chikoti(200296)	GeoMap graph and handled various integrations
Madhav (200657)	Sunburst graph, layout design & integrations
Poojitha (201094)	Styling & layout
Prashamsha(200846)	Styling & layout
Thota Vishnudatta (201051)	Sunburst graph and buttons for loading different years
Ankit Kumar (210150)	Wordcloud
Sambhav Agarwal (200852)	Parallel Coordinate plot
Vibhanshu Patidar (201106)	Handling data integration and Bar Graph
Aryan Jain (200201)	Handling data integration and Bar Graph

Table 4: Contribution of Team Members

References

https://d3js.org/