CSCE679: Data Visualization Project: Step 2

Team Members:

1. Gunjan Joshi
2. Kalvakuntla Sreesh Reddy

Goalwise: An Interactive Insight into Team Triumphs & Trials

As mentioned in Step 1, our project aims to generate a website/dashboard containing the live statistics of all the teams in the premier division of Spanish Soccer (Known as La Liga) that provides a comparative assessment of each team's current standing. Through the dashboard, we intend to visualize the measurements or metrics which are usually considered for indicating teams' performance. Some examples of this could be Team Form and Home Advantage.

Data

The data will be primarily sourced from the website: https://www.football-data.org/ which has an in-built API to obtain different datasets of teams and players for different seasons (years) pertaining to our project. The data is stored in a .json file format. The data ranges from 2019 onwards and contains fixtures, standings, results, and player statistics.

As mentioned, the data is stored in a json file. This mainly consists of Key-Value pairs wherein a key is a string that provides a name or a heading for the data, and the value can be of various types, including, strings, numbers, objects, arrays, Booleans or null.

JSON allows for nesting, meaning there can exist objects within objects or arrays within arrays, creating hierarchical and structured data.

The specific data which will be used would the standings of the teams which includes points earned by the team, previous as well as upcoming fixtures. The statistics of the players, primarily goals and assists would also be considered in our analysis and creation of the dashboard. In the fixtures of matches played, the stats considered would be Goals Scored and Conceded by each team and their Clean Sheet status.

Few snippets of the data which can be extracted from the website:

```json
"season": {
    "id": 1577,
    "startDate": "2023-08-13",
    "endDate": "2024-05-26",
    "currentMatchday": 10,
    "winner": null
},
"id": 438563,
"utcDate": "2023-10-07T14:15:00Z",
"status": "FINISHED",
"matchday": 9,
"stage": "REGULAR_SEASON",
"group": null,
"lastUpdated": "2023-10-19T15:21:25Z",
"homeTeam": {
    "id": 86,
    "name": "Real Madrid CF",
    "shortName": "Real Madrid",
    "tla": "RMA",
    "crest": "https://crests.football-data.org/86.png"
},
"awayTeam": {
    "id": 79,
    "name": "CA Osasuna",
    "shortName": "Osasuna",
    "tla": "OSA",
    "crest": "https://crests.football-data.org/79.svg"
},
"score": {
    "winner": "HOME_TEAM",
    "duration": "REGULAR",
    "fullTime": {
        "home": 4,
        "away": 0
    },
    "halfTime": {
        "home": 1,
        "away": 0
    }
}
```

Figure: Latest Fixture of Real Madrid

```json
"scorers": [
    {
        "player": {
            "id": 3218,
            "name": "Lionel Messi",
            "firstName": "Lionel",
            "lastName": "Messi",
            "dateOfBirth": "1987-06-24",
            "nationality": "Argentina",
            "section": "Offence",
            "position": null,
            "shirtNumber": null,
            "lastUpdated": "2023-05-01T19:34:12Z"
        },
        "team": {
            "id": 81,
            "name": "FC Barcelona",
            "shortName": "Barça",
            "tla": "FCB",
            "crest": "https://crests.football-data.org/81.svg",
            "address": "Avenida Arístides Maillol s/n Barcelona 08028",
            "website": "http://www.fcbarcelona.com",
            "founded": 1899,
            "clubColors": "Red / Navy Blue / Orange",
            "venue": "Camp Nou",
            "lastUpdated": "2021-09-16T20:17:51Z"
        },
        "playedMatches": 35,
        "goals": 30,
        "assists": 9,
        "penalties": 3
    },
```

Figure: La-Liga Top Scorer of the 2018-19 Season

Our project relies on an API to fetch football statistics, and this data source is generally considered reliable in nature. The data we access primarily pertains to football matches, teams, and player statistics, and it is regularly updated to reflect the most recent events and outcomes in the sport. Additionally, as football statistics are widely available and documented, it can be cross verified with information from reputable sources such as Google and the official Spanish League website, which adds an extra layer of validation to the data we utilize. The use of a trusted API, along with the easy verification of football-related statistics, contributes to the confidence in the data's accuracy and readiness for use in our project. However, the data is manipulated to better suit our visualizations.

<u>Visualizations:</u>

LINE GRAPH: Line graphs are generally favored for showing trends and changes in data over a continuous or sequential period. The data being plotted for this would be the team rating of the opponent against the months in a season. This would help understand when a team has a big match during the season.

The team rating of a particular team would be calculated using their historical performance. Using the points and the goal difference from previous season, and then normalizing them between 0 and 1, team ratings of 20 teams can be calculated.

BUMP CHART: A bump chart can be used to highlights changes in rank or position of a set of items over a specified period. We'll plot the form, position of each team over each Matchday.

 The team form is a short-term indication of their performance. This can be calculated using the team's goal difference multiplied by the rating of the opponent over a period of time.

MULTI-LINE GRAPH: This type of graph displays multiple lines on the same graph to represent and compare data from different sources, categories, or variables simultaneously. Each line in the graph typically represents a distinct data series, making it easy to visualize and compare trends and patterns across these series. We'll plot the form, position and points of each team over each Matchday.
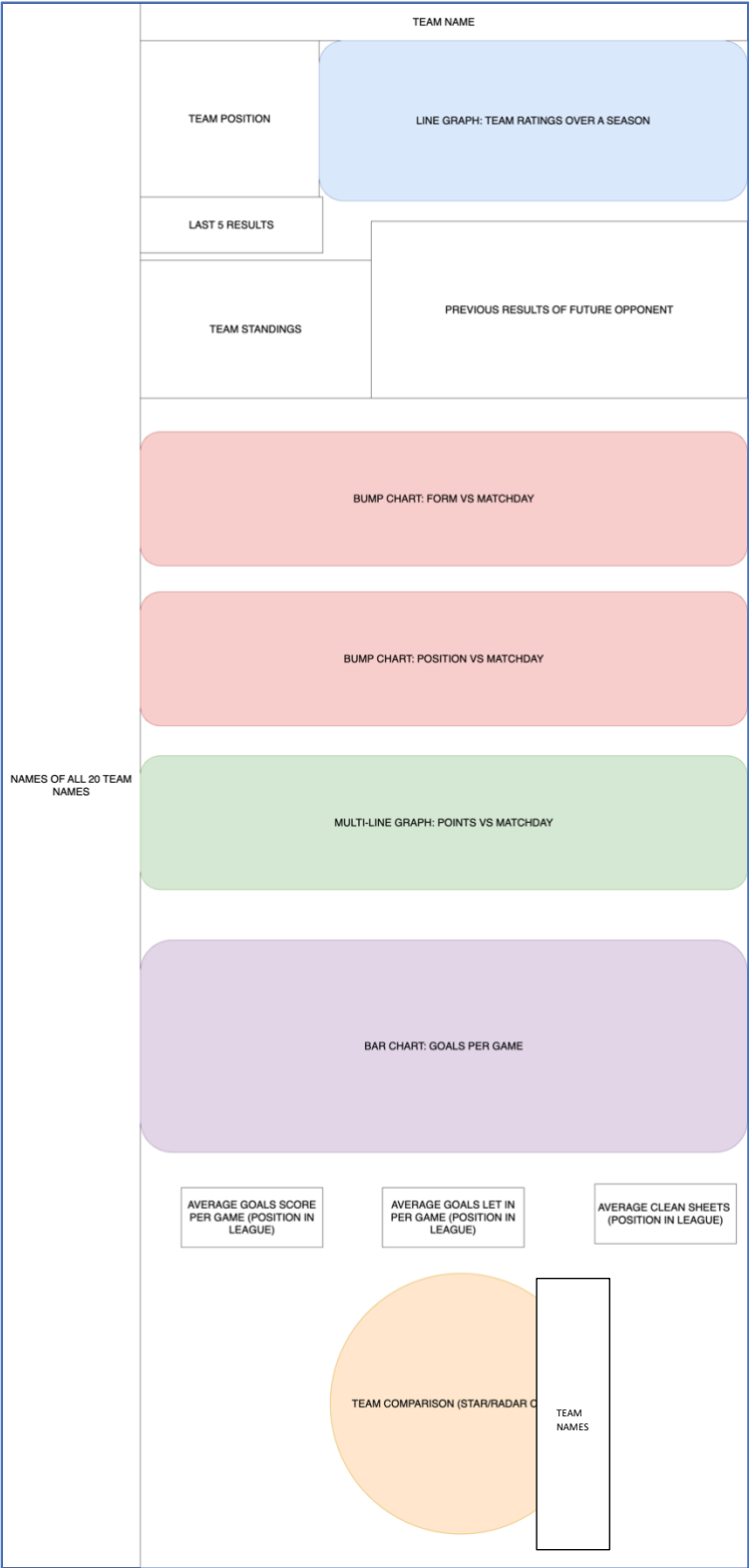
STACKED BAR CHART: Bar graph presents data using rectangular bars of varying lengths. Each bar's length corresponds to the quantity or value it represents, making it easy to compare and contrast different categories or data points. We'd be plotting the goals scored and conceded on every matchday and also have an indicator for a team having clean sheet on that matchday. This would be plotting using the data present in a fixture key-value pair.

BAND CHART: This type of graph depicts a data series within a specific range or band. We use this graph to understand how well a team's attack and defense has done over the course of previous seasons.

STAR CHART: The star chart is particularly useful for comparing multiple variables across different categories or entities. In a star chart, each variable is represented by a spoke or axis radiating outward from a central point, and the length of each spoke corresponds to the magnitude or value of that variable. We use this graph to compare team's attack, defense, consistency, and performance against the big 5 (Usually considered as FC Barcelona, Real Madrid, Atletico Madrid, Valencia, Sevilla)

Consistency can be measure of the games which have consecutive identical games. Or can also be calculated how much their results are deviated from a predetermined average.

The initial prototype of the layout of the visualizations has been displayed below. This might vary as the project moves along.

TEAM NAME

TEAM POSITION

LINE GRAPH: TEAM RATINGS OVER A SEASON

LAST 5 RESULTS

TEAM STANDINGS

PREVIOUS RESULTS OF FUTURE OPPONENT

BUMP CHART: FORM VS MATCHDAY

BUMP CHART: POSITION VS MATCHDAY

NAMES OF ALL 20 TEAM NAMES

MULTI-LINE GRAPH: POINTS VS MATCHDAY

BAR CHART: GOALS PER GAME

AVERAGE GOALS SCORE PER GAME (POSITION IN LEAGUE)

AVERAGE GOALS LET IN PER GAME (POSITION IN LEAGUE)

AVERAGE CLEAN SHEETS (POSITION IN LEAGUE)

TEAM COMPARISON (STAR/RADAR C

TEAM NAMES

Some form of interactions we can see in our visualizations are:

- TOOLTIP: Tooltips are a user interface design element that provides brief, context-sensitive information when a user hovers over or interacts with an element, such as a button, icon, or hyperlink. In our case, whenever we hover over a data point in a graph, a pop up appears which contains a concise description or explanation of the associated element. This helps reduces cognitive load, improve space optimization, and prevents user errors.

Hovering over a data point
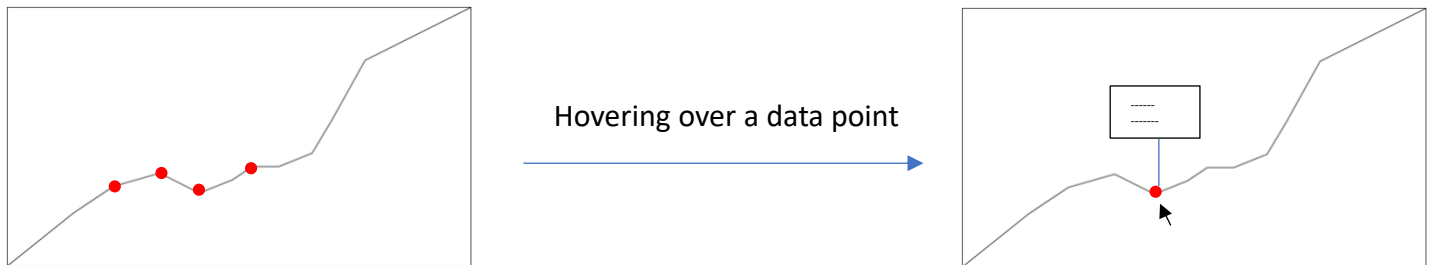
We can see that hovering over the data point gives us even more information about it. This could be the name of the team, the Matchday number, the date at which it was played and the exact value on the y-axis displayed.

- COMPARISON CHECKMARKS: One aspect of our project is to have a comparative analysis between the clubs. Hence, we create comparison checkmarks which act as visual aid used to highlight specific data points or elements for comparison. It can be observed that using checkmarks can enhance the clarity and impact of a data visualization, facilitating quicker and more intuitive analysis of complex information. In our case, we would be using this in our Star chart which compares different metrics of a team.

Selecting a new team
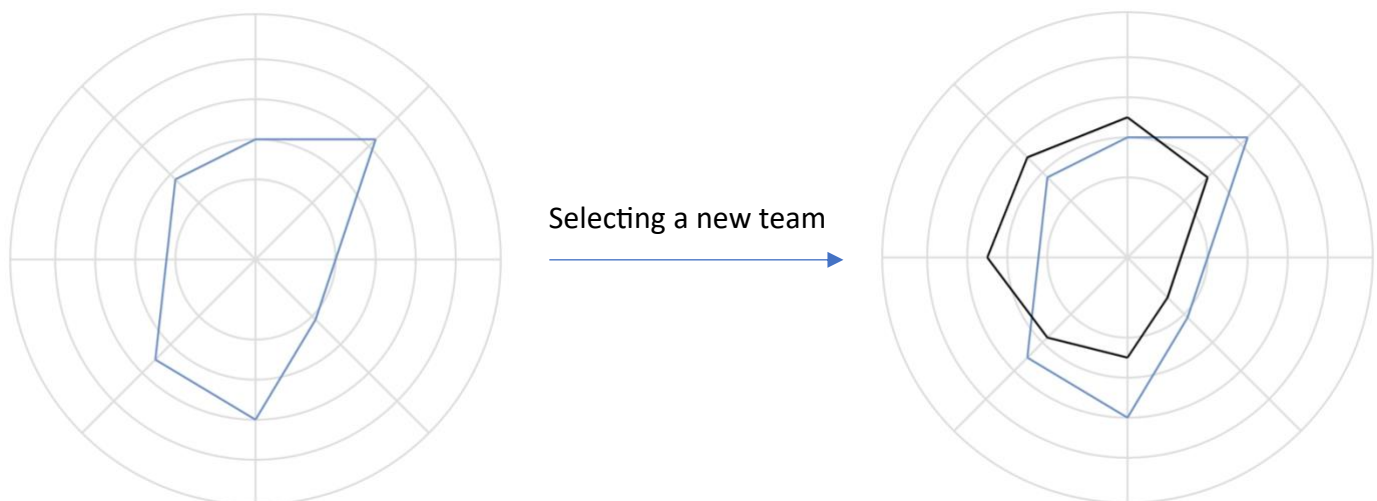
This can be done with any number of teams so that each metric's comparison can be visually inferred since the star chart for both clubs is displayed at the same time.

Revisiting Action-Target pairs:

We revisit the action-target pairs and see how our interactions affect them:

{Filter, Teams}: Users can apply filters to select specific teams they are interested in. This interaction allows them to view and analyze performance metrics, score predictions, and other relevant data pertaining only to the chosen teams, enabling focused analysis.

This would be seen in a line graph when you can view the lines of the teams preferred. This can also be seen by using a navigation bar for each team through which we can navigate to the graph for that particular club.

{Compare, Team Performance}: The audience has the option to compare the performance of two or more teams. This provides insights into relative strengths, weaknesses, and form, aiding users in drawing informed conclusions about upcoming matches or overall team standings.

This would be viewed in the star chart where the performance metrics of different clubs can be compared by selecting the name of the club.

{Explore, Time Series Data}: Users can explore time-series data to observe trends, patterns, and fluctuations in team performance over time. This interaction allows the audience to understand how teams have evolved, adapted, and performed across different seasons or match dates.

This would be viewed In the bump chart where as matchdays in the season progress, the position of all teams can be visually observed and how they vary over time.

{Customize, View Preferences}: The audience can customize their view preferences, adjusting the dashboard layout, choosing which metrics to display, and view tooltips to aide their analysis. This ensures a personalized user experience, catering to individual needs and interests.

One example of this would be including a light and a dark theme to the dashboard as well as having access to tooltips in the graphs to view more information about the teams.

{Navigate, Different Sections}: Users navigate through different sections of the dashboard, such as current standings, score lines and historical data. They can also navigate to teams having stand-alone visualizations.

As mentioned, a navigation bar would be available, which would enable viewing of the same type of graph for different teams with their stats highlighted.

Software and Libraries Used:

We will be using the following Software and Libraries in the programming language for our following project:

- Python: The main programming language used for the project.
  - Pandas: Python library for data manipulation and analysis
  - Plotly: Versatile data visualization library that offers interactive, web-based charts and graphs, ideal for creating dynamic and engaging visualizations.
  - Matplotlib: Widely used Python library for creating static, publication-quality plots and charts
  - Seaborn: Data visualization library built on top of Matplotlib
- Streamlit: User-friendly Python framework that simplifies the process of creating web applications.
- Heroku: Web hosting service to host the dashboard online.
- GitHub: Used for version control as well as managing the code and collaboration.
- Flask/Django: Python web frameworks to create web applications.
- HTML/CSS/JavaScript: Building the frontend of the dashboard.

Workload Division:

The work would be divided mostly equally between 2 members of the team.

Gunjan: Would primarily be responsible for implementing the backend part of the project to have the appropriate API calls to the website and store the data in an usable manner. Would also work on the calculations required for certain visualizations (such as Team Rating and Team Form).

Sreesh: Would primarily be responsible for implementing the frontend part of the project which deals with creating the dashboard in a readable manner as well as deploying it on the web. Would also deal with placement of graphs on the dashboard and rendering the website online.

Both of the members would be involved in implementing the visualizations required for the project and documenting their methods and results for future reports.