

# CSE564 VISUALIZATION - REPORT FOR LAB 1

**NAME – Shreyas Habade**

**SB ID - 115911132**

**How to run the project –** Double-click the start.bat file to run. If it doesn't run, please run it through the Go Live feature in VS Code or please feel free to contact me, I can show it on my machine.

**Dataset - 2021-2022 Top 5 European Leagues Player Statistics**

**Dataset Source:** Kaggle

**Dataset URL:** [2021-2022 Football Player Stats \(kaggle.com\)](https://www.kaggle.com/robervl/2021-2022-football-player-stats)

**Dimension of Dataset:**

1. Number of Columns - 146
2. Number of Rows - 2922

**Description of some of the Columns:**

Attribute	Abbreviation	Description
Nation	Nation	Nationality of the player
Squad	Squad	Team squad
Comp	Competition	Competition the player participated in
Age	Age	Age of the player
Born	Born	Birthyear of the player
MP	Matches Played	Number of matches played
GoalsScored	Goals Scored	Total number of goals scored in season
Starts	Starts	Number of games started by the player
Min	Minutes Played	Total minutes played
90s	90 Minutes Played	Total minutes played divided by 90 (standard match time)
GoalsPer90s	Goals Per 90 Minutes	Average number of goals scored per 90 minutes
Shots	Shots	Total number of shots taken
SoT	Shots on Target	Total number of shots on target

Attribute	Abbreviation	Description
SoT%	Shots on Target Percentage	Percentage of shots that are on target
G/Sh	Goals per Shot	Ratio of goals to shots
G/SoT	Goals per Shot on Target	Ratio of goals to shots on target
TotalSuccessPasses	Total Successful Passes	Total number of successful passes
PasTotCmp	Passes Total Completed	Total number of passes completed
PasTotAtt	Passes Total Attempted	Total number of passes attempted
PasTotCmp%	Passes Total Completion %	Percentage of total passes completed
PasCmp	Passes Completed	Total number of passes completed
CrdY	Yellow Cards	Total number of yellow cards
CrdR	Red Cards	Total number of red cards
2CrdY	Second Yellow Cards	Total number of second yellow cards
FoulsCommitted	Fouls Committed	Total number of fouls committed
Fls	Fouls	Total number of fouls
Fld	Fouled	Total number of times fouled
OG	Own Goals	Total number of own goals

## Why this dataset?

The dataset offers a rich array of attributes vital for thorough player analysis, encompassing diverse performance metrics fundamental for football enthusiasts. As someone deeply passionate about football and with a keen interest in prestigious tournaments such as the La Liga, Premier League, and League 1, delving into intricate player statistics is an endlessly fascinating endeavor.

The dataset provided offers a comprehensive glimpse into player performance metrics, spanning from goals scored and assists to passing accuracy and defensive actions. Through the lens of data visualization, I aim to identify teams that may benefit from player changes, pinpointing areas where improvements are needed based on performance metrics. Additionally, I intend to uncover standout performers, such as players who have provided the most assists leading to goals, shedding light on their pivotal roles within their respective teams. By leveraging interactive charts, graphs, and heatmaps, we can navigate through the complexities of player statistics, offering actionable insights that can inform strategic decisions and enhance team performance in the seasons to come.

The current analysis focuses on presenting comprehensive metrics for each league or all leagues combined, emphasizing goals scored, age distribution, passing efficiency, and disciplinary behavior. Utilizing bar graphs, scatterplots, and histograms, I visually represent the data. Given the nuanced nature of the data attributes, I derived key metrics such as Total Number of Goals Scored, Total Number of Fouls Committed, Goals Per 90 Minutes, and Total Successful Passes. These metrics provide a holistic view for graph visualization. They enable insights into various aspects including the scoring abilities of players across different age groups, the positive correlation between total goals scored and passes completed, and the differences in foul commitment tendencies among teams in different leagues.

One aspect I particularly appreciate about this dataset is its flexibility for customization. With the original dataset comprising around 143 attributes, it offers a vast playground for exploration. I can manipulate the data to my liking, experimenting with various attributes and visualizing different metrics. This versatility allows me to delve into intricate relationships within football, gaining valuable insights and a deeper understanding of the sport.

### **Derived Attributes:**

In the unprocessed dataset, player statistics such as goals scored and fouls committed are originally presented on a per-90-minute basis. To enhance the data for analysis, it was imperative to establish a coherent mapping of their interrelation. This involved a meticulous process of calculation and derivation to transform the raw data into attributes that represent the total goals scored and fouls committed by each player. This refinement was undertaken to facilitate a more comprehensive and meaningful exploration of player performance, transcending the initial per-90-minute metric to provide a holistic perspective on their contributions to the game.

### **Noteworthy Mention:**

In addressing the inherent player-centric nature of the dataset, a pivotal step was taken to aggregate, categorize, and visualize the information at both team and league levels. This necessitated the summation, binning, and categorization of player data based on their respective teams, leagues, and age brackets. By undertaking this process, the aim was to transcend the individual player perspective and offer a comprehensive view of team and league dynamics, thereby enriching the analytical scope. The incorporation of age bins further contributed to a nuanced understanding of performance patterns across different player demographics, infusing a human-centric element into the analytical framework.

I've also made a start.bat file which is a batch file that can be run with double click, which further runs the run\_server.ps1 PowerShell file which starts a python webserver on port 5500 runs the home.html file over and opens it automatically in the browser.

### **Conclusion:**

Utilizing the available dataset, the primary objective is to furnish team scouting and management experts with actionable insights to strategically shape the future trajectory of the team. This involves the identification and recruitment of potential talent, with a specific focus on players exhibiting prolific goal-scoring abilities or the discernment to discern and remove individuals prone to excessive fouls. Additionally, the evaluation extends to goalkeepers, scrutinizing their efficacy in making saves and preventing goals, thereby informing decisions regarding player retention or replacement. The implementation of data visualization techniques serves as a powerful tool, empowering experts to make informed and strategic decisions that contribute to the holistic development and success of the team. This amalgamation of statistical analysis and visual representation forms a nuanced foundation for decision-making in the realm of team management.